Exploiting Dynamical Coherence: A Geometric Approach to Parameter Estimation in Nonlinear Models

Leonard A. Smith,* Milena C. Cuéllar, Hailiang Du, and Kevin Judd†

Centre for the Analysis of Time Series.

Department of Statistics. London School of Economics, London WC2A 2AE. UK

(Dated: April 19, 2010)

Parameter estimation in nonlinear models is a common task, and one for which there is no general solution at present. In the case of linear models, the distribution of forecast errors provides a reliable guide to parameter estimation, but in nonlinear models the facts that predictability may vary with location in state space, and that the distribution of forecast errors is expected not to be Normal, means that parameter estimates based on least squares methods will result in systematic errors. A new approach to parameter estimation is presented which focuses on the geometry of trajectories of the model rather than the distribution of distances between model forecast and the observation at a given lead time. Specifically, we test a number of candidate trajectories to determine the duration for which they can shadow the observations, rather than evaluating a forecast error statistic at any specific lead time(s). This yields insights into both the parameters of the dynamical model and those of the observational noise model. The advances reported here are made possible by extracting more information from the dynamical equations, and thus improving the balance between information gleaned from the structural form of the equations and that from the observations. The technique is illustrated for both flows and maps, applied in 2, 3, and 8 dimensional dynamical systems, and shown to be effective in a case of incomplete observation where some components of the state are not observed at all. While the demonstration of effectiveness is strong, there remain fundamental challenges in the problem of estimating model parameters when the system that generated the observations is not a member of the model class. Parameter estimation appears ill defined in this case.

Parameter estimation is often approached in a "statistical" fashion focused on the forecast error distribution for a given lead time. In this letter, we argue that more insight can sometimes be gained by adopting a "geometric" approach, the main distinction being the emphasis placed on the existence of model trajectories which resemble segments in the time series of observations. This reduces the role of traditional summary test statistics using observations and forecasts at a given lead time or set of lead times. Instead we consider, for a given parameter value, the duration over which the model admits a trajectory that shadows the observations. Statistics of the duration of shadowing times are then used to select good parameter values. This enhances the balance between the information contained in the dynamic equations and the information in the observations themselves. The nonlinearity of the problem inextricably links parameter estimation [34], state estimation ("noise reduction" [11]) and specification of the noise model (that is, the statistical properties of the observational noise). The information in the intermediate term model dynamics is complementary to that implicit in the invariant measure, as exploited by McSharry and Smith [22]. Creveling et al (2008) have used synchronisation for parameter estimation.

Candidate parameter values are evaluated upon how well corresponding trajectories (and pseudo-orbits [15, 17]) mimic the observations. Model trajectories close to a segment of observations are found by gradient descent [6, 15, 17, 27, 28] (note, however, reference [18]). For each parameter value we quantify (i) the ability of model trajectories to ι-shadow [7, 32] uncertain observations, (ii) how well model pseudo-orbits approximate relevant trajectories and (iii) the consistency of the distribution of implied-noise with the noise model (when one is known [26, 35]). The results in this letter fall within the perfect model scenario (PMS), in which there exists a parameter set (for the dynamic model and the noise model) which admits the trajectory which did, in fact, generate the observed data. Thus the mathematical structure of the model equations are "correct", but the true parameter values are unknown. It is important to note that while the true parameter values are unknown, they are well defined within PMS; it is unclear how one might define "optimal parameter values" when the model structure is imperfect. Given PMS, our approach could, no doubt, be cast into a proper Bayesian perspective, a point we return to below.

The problem of parameter estimation in the general context of nonlinear dynamical systems is presented in the next section. By moving from the model state-space to a higher dimensional sequence space (in which a sequence of state estimates composes a single point) we can see the relation between trajectories of the model and a time series of observations. For a particular parameter value, pseudo-orbits (and trajectories) are found which resemble both the observations and model trajectories. The strength of that resemblance is then used to evaluate the quality of that candidate parameter value. Specifically, properties of the distribution of shadowing trajectories along with a comparison of the distribution of implied-noise with the known noise model are used to quantify the goodness of fit. Loosely speaking, the best parameter values are those which shadow longest.

The Problem in context

Even if the functional form $\mathbf{F}(\mathbf{x}, \mathbf{a})$ of a model (that is, the model class) is known to match that of the system exactly, perhaps from true first principles, a value for the vector of model parameters, $\mathbf{a}$, must be estimated from observations $\mathbf{s_i}$ of the state variables $\mathbf{x_i}$. We consider the case where both the functional form of the dynamical system and the nature of the observational noise are know exactly. In other words, the noise model is exact and the functional form of the dynamic model is known.

For nonlinear models, the problem is more subtle than might first be imagined. Naively one might hope that, given this perfect model scenario and observations over a sufficiently long duration, it would be possible to "solve" the inverse problem completely: to identify both the parameter values and the "true" trajectory to arbitrary precision. In fact, this is not the case given a chaotic model; even when the parameter values are known exactly, the observations extend into the infinite past and the noise model is known exactly the true state can not be identified (see [15, 17] and references thereof). What then can be extracted from a finite series of observations given an exact observational noise model and the functional form of the system, but no information on the parameter values?

We do not attempt to answer this question in full, but demonstrate progress within PMS. It is interesting to note that fundamental questions of parameter estimation in nonlinear systems remain open [3, 4, 8, 26, 35], even in the perfect model case.

TECHNICAL STATEMENT OF THE PROBLEM

Suppose the evolution of a system's state variable, $\mathbf{x}_i \in \mathbb{R}^m$, is determined by the map

$$\mathbf{x}_{i+1} = \mathbf{F}(\mathbf{x}_i, \mathbf{a}), \tag{1}$$

where system parameters are contained in the vector $\mathbf{a} \in \mathbb{R}^l$. For $m = 1$, the state x_i is a scalar; assuming additive measurement noise η_i yields observations $s_i = x_i + \eta_i$. In a

Markovian system, knowledge of past states adds no additional information regarding the future if the current state is known; given only an *observation* of the current state however, past observations s_i may hold significant additional information. In this case $l+1$ sequential measurements $\mathbf{s}_i, \mathbf{s}_{i+1}, ..., \mathbf{s}_{i+l}$ would, in general, be sufficient to determine $\mathbf{a}$ in a noise free setting (i.e. $\eta_i = 0 \;\forall\; i$),. With observational noise, the task is somewhat harder.

A pseudo-orbit $\mathbf{U}$ of the map $\mathbf{F}$ is a series of points $\mathbf{u}_i \in \mathrm{I\!R}^m$ in state-space such that $|\mathbf{u_{i+1}} - F(\mathbf{u_i}, \mathbf{a})|$ is small but not zero (if each term was zero, then $\mathbf{U}$, that is the sequence $\mathbf{u}_i$, would be a trajectory segment). For realistic levels of observational uncertainty, the state observations $\mathbf{s}_0, \mathbf{s}_1, ..., \mathbf{s}_N$ form a pseudo-orbit. For a given pseudo-orbit, define the *mismatch* as $(\mathbf{u_{i+1}} - F(\mathbf{u_i}, \mathbf{a}))$; define the implied-noise as $(\mathbf{u_i} - \mathbf{s_i})$. Obviously, we will aim to find pseudo-orbits which remain close to the observations and yet whose mismatches have a distribution significantly narrower than that corresponding to the observations. When the mismatch takes on small values then (i) there are likely to be trajectory segments close to the pseudo-orbit and (ii) we can establish whether or not the distribution of the implied-noise is consistent with the true noise model. A segment of a trajectory is said to ι-shadow a set of observations if the implied-noise series is consistent with the noise model (an operational definition of "consistent" is given below). For each observed state, a set of candidate states are found, each candidate corresponds to a trajectory which will shadow the observations for some period. The shadowing time corresponding to an observed state is the duration of the longest shadowing trajectory segment found among the candidate states tested; the distribution of shadowing times is thus an indication of model quality. One thing that distinguishes our approach from traditional methods based on forecast errors is that statistics of the distribution of shadowing times quantify the time scales over which the dynamics of the model reflect those of the observation. Most traditional forecast statistics, on the other hand, quantify how well the dynamics of the system mimic those of forecasts at a fixed lead time.

States along a pseudo-orbit are used as candidate states from which the distribution of shadowing times is established, and the distribution of shadowing times is then used to identify good parameter values. Perhaps the most accessible set of candidate states come from the pseudo-orbit corresponding to the observations themselves. This approach can outperform parameter estimation based upon linear least squares. As noted by a reviewer, computational constraints will restrict the quality and number of candidates tested. If the computational resource is too small, the results are unlikely to be robust. For simplicity, only forward shadowing times from the pseudo-orbit are considered; in the results presented below, methods for finding longer shadowing orbits will be reported elsewhere.

Gradient Descent.

The extent to which a pseudo-orbit reflects the quality of a parameter set rests on the quality of the minimisation of the mismatch error given some fixed computational resource. For simplicity, we adopt a sequence-space gradient descent (GD) approach to finding pseudo-orbits. For a given segment of length N_a or equivalently a single point in the $m \times N_a$ dimensional sequence-space, the algorithm (see [6, 15, 17, 18, 28] for details) aims to minimise the mismatch cost function

$$C_{\mathrm{MM}}(\mathbf{U}, \mathbf{a}) = \sum_{i=0}^{N_a-2} ||\mathbf{u_{i+1}} - F(\mathbf{u_i}, \mathbf{a})||, \tag{2}$$

by adjusting the components of the pseudo-orbit with **a** fixed. It is critical to note that this is the cost function for the GD algorithm; while a function of **a** it is not the cost function for identifying good parameter values. As noted throughout, good vales of **a** are identified using the distribution of shadowing times for each value of **a**. Also note the only dependence on the observations is in terms of producing starting conditions (see [15] for discussion). To ease comparison with other previous work [6, 15–17, 22, 26, 28, 30, 35] we use the distribution of shadowing times for parameter estimation in the two-dimensional Hénon map [12] in delay coordinates, $F(x_i, x_{i-1}) = 1 - ax_i^2 + bx_{i-1}$, where $a_0 = 1.4$ and $b_0 = 0.3$. (The details of all systems used are given in the appendix.)

Although there are no local minima in the sequence-space (Footnote: For a proof that the only minima are system trajectories where $C_{\mathrm{MM}}(\mathbf{U}, \mathbf{a}) = 0$, see [15]. All points corresponding to trajectories are minima.), the GD algorithm is run for a pre-determined time and thus a trajectory (a global minima) is not obtained; denote the best pseudo-orbit found (for the moment, take this to be the pseudo-orbit with the lowest values of C_{MM}) as z_i. When the GD algorithm is applied over a series of overlapping windows, then a number of pseudo-orbits will be found for each time i; for simplicity we consider only one pseudo-orbit over the entire duration of the observations.

What is noise, exactly?

The noise model is an important component of every data analysis; ideally a robust methodology for estimating parameters will support, inform, or even reject the *a priori* noise model in the context of the model class. Below we assume a noise model of additive, independent and identically distributed (IID) observational uncertainty; the distribution will be taken to be Normal with mean zero and known standard deviation. In this context "noise level" implies standard deviation. Similarly the "implied-noise level" is defined as the sample standard deviation of the distribution of implied-noise. Note, of course, that in nonlinear systems it may be possible to extract a smaller implied-noise level at the cost of violating the assumed form of the distribution. Inasmuch as we initialise the GD with the pseudo-orbit consisting of the observations and aim to explicitly minimise the mismatch cost function, the final pseudo-orbit obtained is often found to have an implied-noise level less than the true noise level. (That is, the pseudo-orbit found is closer to the observations than expected, given the noise model.) If there is good reason to trust the *a priori* shape of the distribution of the noise, then the apparent improvement reflected in the "lower" noise level is in fact a chimera, to be ignored. Within PMS, there is no dynamic noise unless it is specified within the model. Outside PMS, it is unclear whether or not this type of additive noise model makes any sense.

Shadowing times

Although superficially similar, the question of whether a model shadows a set of observations is a fundamentally different notion from the traditional question of whether or not one mathematical system can shadow the trajectories of another [7, 20, 29, 32]. Traditional shadowing [29] involves two well-defined mathematical systems while our ultimate interest is between a set of numbers (the observations) and a mathematical system (the proposed model). Even within PMS there may be insufficient information in the observations to provide accurate estimates of the parameter values. In the more general case where the model structure is uncertain, we cannot rationally assume that any "true" parameter values even exist. Details of the distinction need not concern us in depth here, where the relevant question is: for a given segment of observations, does the system admit a trajectory such that the residuals defined by the trajectory and the observations are consistent (in distribution and in dynamics) with the noise model. More precisely: what is the largest k such that, for some model state $\mathbf{x}$, the time series $\mathbf{r_i} = \mathbf{s_i} - \mathbf{F}^{(i)}(\mathbf{x}, \mathbf{a}), i = 0, ..., k$ is consistent with the noise model?

For uniform bounded noise this test is straightforward: we might merely compute the number of consecutive residuals, r_i, which are less than the bound. For Gaussian distributed observational noise there are a variety of approaches; we adopt a simple method based on threshold exceedence that, no doubt, could be improved upon. Given that the noise model is unbounded, *any* observation is conceivable; we look for relevant [2] shadows within a certain probability bound. Specifically we test the null hypothesis that the set $(r_i, i = 0, 1, 2, ..., k)$ is consistent in distribution with a draw from a Gaussian distribution of specified variance. In practice, we require both that the 90% isopleth of the residual distribution falls below the 99^{th} percentile of the distributions of .90 isopleths given k draws from a Gaussian distribution, *and* that the median of the residual distribution falls below the corresponding 90^{th} percentile for the median of our noise model. Together this implies that the chance rejection rate is ~ 0.001, which will yield good results as long as the shadowing times we test are below 100 (as they are in the results presented here).

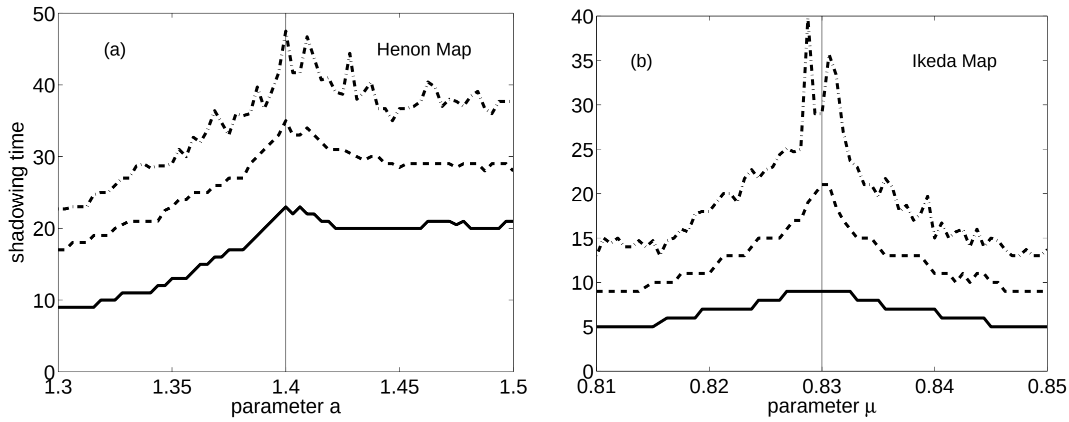

FIG. 1: The median (solid), 90% (dashed) and 99% (dash-dot) isopleths of the distribution of shadowing times (over 512 statistics in sample) for (a) the Hénon map with b=0.3 and (b) the Ikeda map. The vertical line represents the location of the unknown true parameter.

At a given observation in time t, we are most interested in the longest shadowing trajectory that the model admits for a given set of parameter values. To obtain the results below, we consider only a finite set of candidate trajectory segments for each observation and define the shadowing time $\tau_{s_t} = \max_x \tau_s(x, t)$ where the maximum is taken over all $\mathbf{x}$ values tested (Here $\tau_s(x, t)$ is the duration over which the trajectory from the state x shadows the series of observations starting at time t). In the results presented below, only three candidates per observation were tested: the corresponding state on the pseudo-orbit, the image under F of the previous state on the pseudo-orbit, and the state midway between these two. Given a

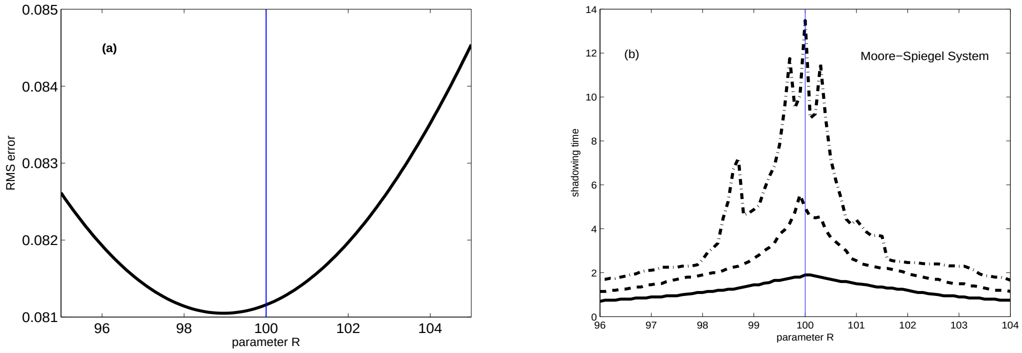

FIG. 2: (a) Least Square estimates for Moore-Spiegel Flow R=100 and (b) Isopleths of the distribution of shadowing times as in Fig 1. (Noise level=0.05)

shadowing time at each state on the trajectory, we have a distribution of shadowing times for a given parameter set. Once a range of empirically reasonable parameter values is in hand, one may wish to employ more complicated methods to identify shadowing trajectories [7, 10]. Simultaneous estimation of two parameter values is demonstrated in the next section.

Panels in the first three figures show median, 90 and 99 percent isopleths of the distribution of observed shadowing times, the true parameter value is denoted by a vertical line. These figures establish that our approach can be effective in 2-dimensional chaotic maps, 3-dimensional chaotic flows and partially observed 8-dimensional chaotic flows. (Details for each system are given in the Appendix.) Before discussing these individually, note that in each case the vicinity of the true parameter value is clearly indicated. The choice of isopleth is not critical, although sampling variations will of course, become an issue for extreme values of the distribution. Thresholds will vary with the size of the data set and the noise model; a straightforward bootstrap re-sampling approach can identify how robustly a given isopleth can be estimated. For both the Hénon system (panel A) and the Ikeda system (panel B) in Figure 1, the median and 90% contours provide good parameter estimates, while the 99% contour suffers from small-sample effects.

Panel B of Figure 2 shows isopleths for the Moore-Spiegel [25] third order ODE, again good parameter estimates are shown. Panel A of the same figure reflects a systematic bias in least squares approaches to parameter estimation. The ability to outperform least squares comes as no surprise, due to the nonlinearity of the systems [22] and the variation in predictability of even infinitesimal uncertainties [31, 36]

Figure 3 considers the case of higher dimensional systems where the state vector is not completely observed. In such cases the shadowing-time is determined without placing any constraints whatsoever on the value taken by the unobserved component(s). In panel A seven of the eight components of the state vector are observed, in panel B only 5 of the eight components are observed. In both cases, the "true" parameter value is clearly indicated.

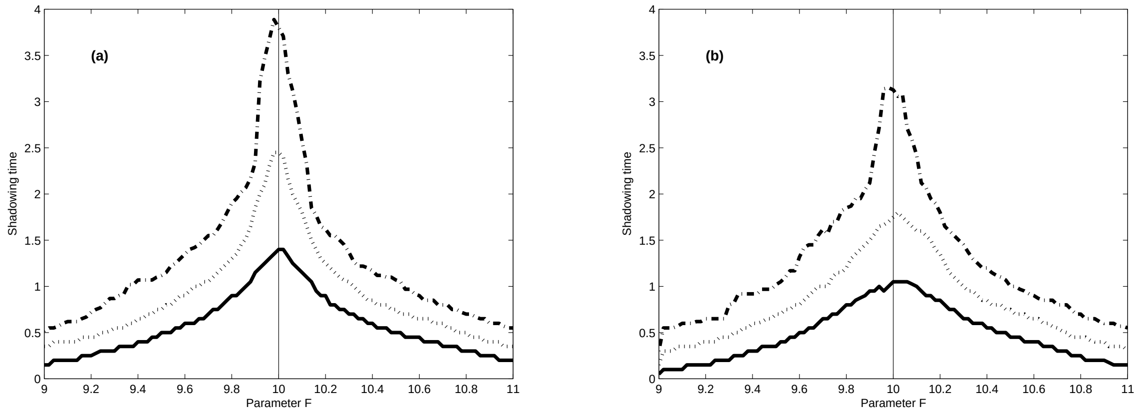

FIG. 3: Shadowing time isopleths as in Fig 1. for 8-D Lorenz96 given only partial observations, a) the 8th component of the state vector is not observed b) none of the 2nd, 5th or 8th variables are observed only the other five components. In this case the noise level is 0.2.

DISCUSSION

The problem of parameter estimation in nonlinear systems differs fundamentally from its well-understood counterpart in linear systems. Variations in the shadowing time distribution with parameter may yield more insight than maps of root-mean-square forecast error, or other linear statistics which have well-known shortcomings in nonlinear models (see [22] and references therein). It would be interesting to extend the work presented above by locating the regions in state space where the model can not shadow and contrasting them with regions identified as having inconsistent dynamics by other means (as in McSharry and Smith [23]).

Communicating the uncertainty in estimated parameters of nonlinear models remains challenging. In linear stochastic systems, nearby parameter values result in similar short term dynamics and similar (Gaussian) asymptotic distributions (invariant measures). In that context, the statement $a_0 \pm 0.1234$ is a useful characterisation of the uncertainty in a parameter a. In nonlinear deterministic systems in general, and structurally unstable systems [9] in particular, this is no longer the case. In short, the idea of "goodness" being related to the difference between the proposed parameter set and "true" parameter values must be rethought: if the goal is to identify parameters that might have actually generated the observed data, then each candidate parameter set must be assessed on its own merits.

The shadowing approach may prove relevant to environmental modelling. Even in systems as unwieldy as multi-million-dimensional operational climate models, variations in parameters over the relevant range of uncertainties yield demonstrably nonlinear effects [33] in the most basic summary statistics (i.e. climate sensitivity); we note that the gradient descent methods above have been used on models of this level of complication [18]. Outside PMS there may be no single "optimal" parameter, of course. In this case shadowing times have the advantage of providing information on likely lead times at which a forecast will have utility. We also note that the timescales on which the dynamics of the model are consistent with the noise model and the observations can be of use in setting the window of observations to be used, and the effectiveness of, variational approaches to data assimilation [18]. If no model trajectory is expected to be consistent, then the output of variational methods can be expected to be misleading.

Within the perfect model scenario, our approach might be profitably recast in a Bayesian framework. Our computationally less expensive shadowing cost function might, for example, be used as an importance sampler in place of a more computationally expensive invariant measure cost function. Difficulties in traditional likelihood based approaches in this context are nicely clarified by Berliner [1]. Coherent Bayesian formulations condition the probabilities extracted upon all information available [30]. For the systems considered here, that information includes the fact that the data were generated by a deterministic system of known mathematical form. To assume some stochastic model for the sake of applying a so-called "Bayesian technique" is to fail to grasp the fundamentals of Bayesian analysis. The dangers are illustrated elsewhere (see [16]), we merely note that the common stochastic version of both the Logistic map and the Hénon map are unstable: with probability one, trajectories under the stochastic version of either map will diverge towards attractors at infinity, making attempts to match either long term dynamics or invariant measures somewhat dubious.

Our approach provides information in addition to the shadowing times. Panels of Figure 4

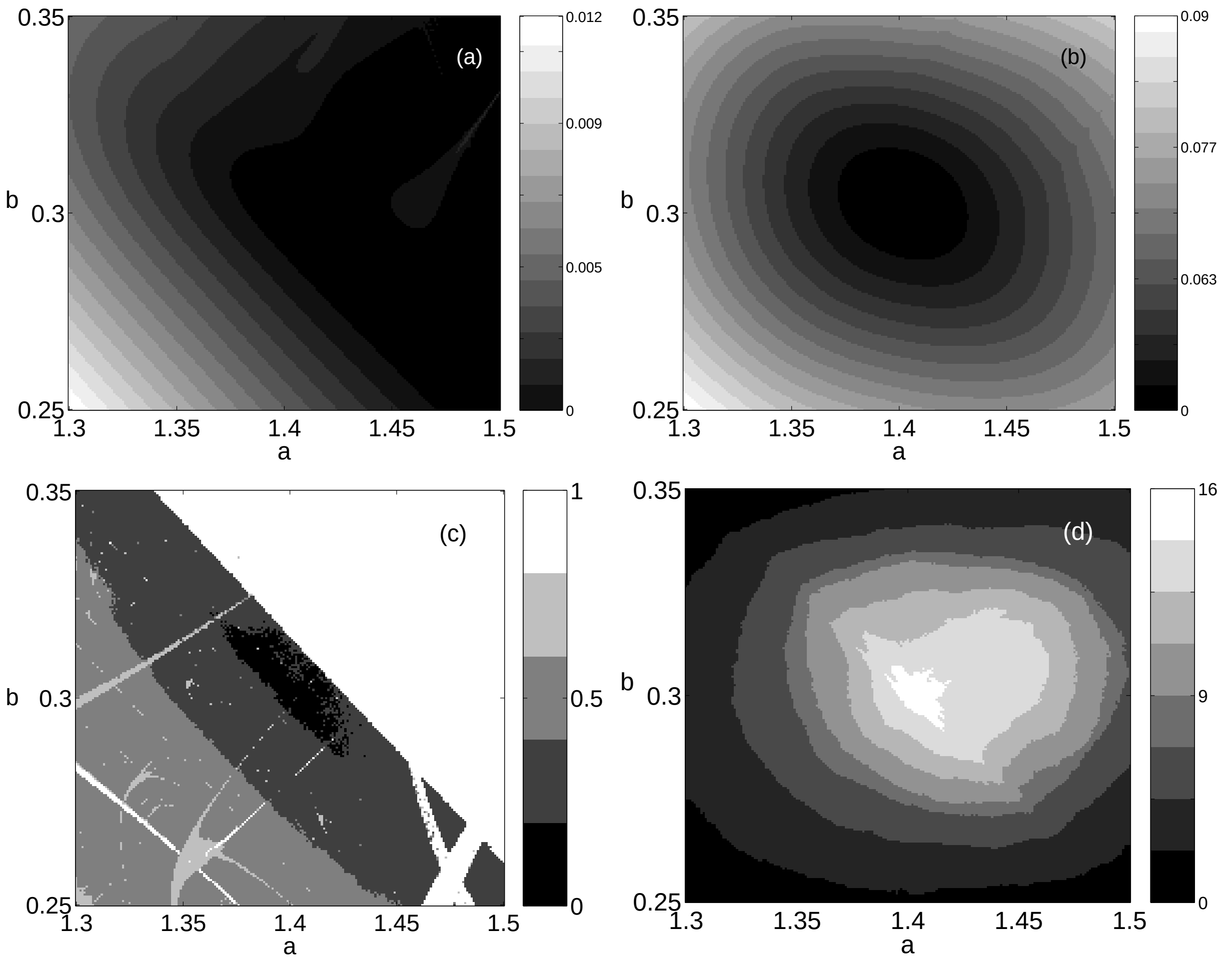

FIG. 4: Information from a pseudo-orbit determined via gradient descent applied to a 1024 observations of the Hénon map with a noise level of 0.05. (a) standard deviation of the mismatch, (b) standard deviation of the implied-noise, (c) a cost function based on the model's invariant measure (after Fig.4(b) of ref [22]), (d) median of shadowing time distribution.

show the variation with the two parameters in the Hénon map of (a) the standard deviation of the mismatch; (b) the standard deviation of the implied-noise; (c) a cost-function based on the invariant measure (after [22]) and (d) the median of the distribution of shadowing times. Contrasting the properties of the implied-noise with those expected from the noise model provides an after the fact consistency check on the procedure.

The fine structure ("tongues") in Fig. 4 (c) arises due to sensitivity to the parameters, nevertheless its minima are in the relevant regions. Contrasting panels (c) and (d) of figure 4 reveals that shadowing times provide information complimentary to that obtained by estimating the invariant measure (the C_{ML} of ref [22]). Figure 5 shows both the duration corresponding to the 90% isopleth of the distribution of shadowing times and a circle within which the implied-noise level is less than or equal to the true noise level. Maps of shadowing times provide complimentary information quantifying the time scales on which the model dynamics reflects the observed behaviour. We conclude that better parameter estimates can be obtained by quantifying the realism of both the short term dynamics (shadowing trajectories) and the long term dynamics (invariant measure). Ultimately, questions of which is "right" are mute due to the finite duration of any data set.

Figures 4 and 5 summarise the main results of this letter: statistics of the distribution of shadowing times provide an unambiguous indication of the "true" parameter values. Only in the perfect model case can we be certain that some single ideal balance between the information in the dynamic equations and the information in the observations even exists. When the model class is inadequate we no longer expect the invariant measure to be informative; considering shadowing times, however, can identify parameter values which can mimic the dynamics, quantify the time scales on which they can shadow, and extract information for improving the model class itself. This is a significant step forward. It will be interesting to examine the performance of this approach in a variety of systems, model classes and noise levels.

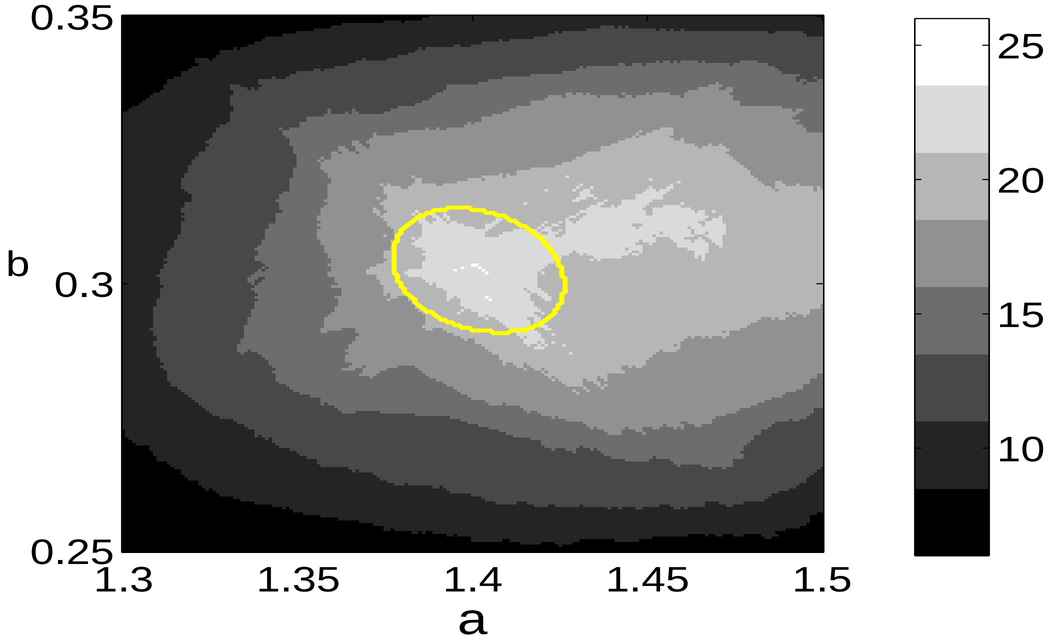

FIG. 5: Duration corresponding to the 90% isopleth of shadowing time for Hénon map, the contour circle is where the implied-noise is equal to the noise level. See text for a discussion of why we expect the target parameter values to be within this contour of the implied-noise level.

Appendix: Systems considered

The four dynamical systems considered are defined below; further information can be found in the references for the Hénon Map [12], Ikeda Map [13], Moore-Spiegel system [25] and Lorenz 1996 system [24].

Hénon Map is defined by:

$$X_{n+1} = 1 - aX_n^2 + Y_n \tag{3}$$

$$Y_{n+1} = bX_n. \tag{4}$$

The parameter values used were $a = 1.4, b = 0.3$. Observations of the state included additive IID observation noise, Normally distributed with $\sigma = 0.05$.

Ikeda Map is defined by:

$$X_{n+1} = \gamma + u(X_n \cos\phi - Y_n \sin\phi) \tag{5}$$

$$Y_{n+1} = u(X_n \sin\phi + Y_n \cos\phi), \tag{6}$$

where $\phi = \beta - \alpha/(1 + X_n^2 + Y_n^2)$ and $\alpha = 6, \beta = 0.4, \gamma = 1, u = 0.9$. Observations of the state included additive IID observation noise, Normally distributed with $\sigma = 0.01$.

Moore-Spiegel system is defined by:

$$dx/dy = y \tag{7}$$

$$dy/dt = z \tag{8}$$

$$dz/dt = -z - (T - R + Rx^2)y - Tx. \tag{9}$$

The parameter values used were $T = 36, R = 100$. We use the forth order Runge Kutta scheme with fixed integration time step of 0.01. Observations are recorded after each 0.05 time unit, and included additive IID observation noise, Normally distributed with $\sigma = 0.05$.

Lorenz96 system is defined by:

$$\frac{d\tilde{x}_i}{dt} = -\tilde{x}_{i-2}\tilde{x}_{i-1} + \tilde{x}_{i-1}\tilde{x}_{i+1} - \tilde{x}_i + F, \tag{10}$$

for $i = 1, ..., m$. The system used in the experiment containing m=8 variables $x_1, ..., x_m$ with cyclic boundary conditions (where $x_{m+1} = x_1$). The parameter value used is $F = 10$. We use the forth order Runge Kutta scheme with integration time step of 0.005. The observations are recorded after each 0.05 time unit. In the first partial observation experiment, we treat the 8th variable x_8 unobserved. In the second experiment, we treat x_2, x_5 and x_8 unobserved. The observed components included additive IID observation noise, Normally distributed with $\sigma = 0.2$.

---

* Also at Pembroke College. University of Oxford. UK

† Also at University of Western Australia, Perth, Australia

[1] M.L. Berliner, Likelihood and Bayesian Prediction for Chaotic Systems, *J. Am. Stat. Assoc.* **86**, 938-952 (1991).

[2] E. Borel, *Probability and Certainty* Walker, New York, (1963).

[3] C.L.Bremer and D.T.Kaplan, Markov chain Monte Carlo estimation of nonlinear dynamics from time series, *Physica D* **160** 116–126 (2001).

[4] M. Casdagli, S. Eubank, J. D. Farmer, and J. Gibson, State space reconstruction in the presence of noise, *Physica D*, **51**, 52 (1991).

[5] D.R. Creveling, P.E. Gill and H.D.I. Abarbanel, State and Parameter Estimation in Nonlinear Systems as an Optimal Tracking Problem, *Physics Letters A*, **372**, 2640-2644 (2008).

[6] M.E.Davies, Noise reduction by gradient descent, *Int. J. Bifurcation Chaos* **3**, 113-118, (1992).

[7] I. Gilmour, Nonlinear model evaluation: ι-shadowing, probabilistic prediction and weather forecasting, Ph.D. thesis, University of Oxford, (1998).

[8] P. Grassberger, T. Schreiber, and C. Schaffrath, Non–linear time sequence analysis, *Int. J. Bifurcation and Chaos* **1**, 521 (1991).

[9] J. Guckenheimer, J. Moser and S. Newhouse, Dynamical Systems, CIME Lectures, Birkhauser, 289pp., (1980).

[10] J.A.Hansen, Adaptive observations in spatialy-extended nonlinear dynamical systems, Ph.D. thesis, University of Oxford, (1998).

[11] J.A.Hansen and L.A.Smith, Probabilistic Noise Reduction, *Tellus* 53 A (5):585-598, (2001).

[12] M. Hénon, A two-dimensional mapping with a strange attractor, *Commun. Math. Phys.* **50** 69-77 (1976).

[13] K.Ikeda, Multiple valued stationarity state and its instability of the transmitted light by a ring cavity system, *Optical Communications*, **30** 257-261 (1979).

[14] L. Jaeger and H. Kantz, Unbiased reconstruction of the dynamics underlying a noisy chaotic time series, *Chaos* **6**, 440 (1996).

[15] K. Judd and L.A. Smith, Indistinguishable States I: The Perfect Model Scenario, *Physica D* **151**: 125-141, (2001).

[16] K. Judd, Chaotic-time-series reconstruction by the Bayesian paradigm: Right results by wrong methods, *Physics Review Letters*, **67**, 026212,(2003).

[17] K. Judd and L.A. Smith, Indistinguishable States II: The Imperfect Model Scenario, *Physica D* **196**: 224-242, (2004).

[18] K. Judd and L.A. Smith and A. Weisheimer, Gradient free descent: shadowing and state estimation with limited derivative information, *Physica D*, **190**, 153–166 (2004).

[19] H. Kantz and L. Jaeger, Homoclinic tangencies and non-normal Jacobians - effects of noise in non-hyperbolic systems, *Physica D* **105**, 79-96 (1997).

[20] E.J. Kostelich and J.A. Yorke, Noise Reduction in Dynamical Systems, *Phys Rev A* **38**, 1649, (1988).

[21] E.J. Kostelich, Problems in estimating dynamics from data, *Physica D* **58**, 138 (1992).

[22] P.E. McSharry and L.A. Smith, Better nonlinear models from noisy data: Attractors with maximum likelihood, *Phys. Rev. Lett* **83**, (21): 4285-4288, (1999).

[23] P.E. McSharry and L.A. Smith, Consistent Nonlinear Dynamics: identifying model inadequacy, *Physica D* 192: 1-22, (2004)

[24] E.N. Lorenz, Predictability A problem partly solved, in: ECMWF Seminar Proceedings on Predictability, Reading, United Kingdom, ECMWF, (1995).

[25] W. D.Moore and E. A.Spiegel, A thermally excited nonlinear oscillator. *The Astrophysical Journal.* Volume **143**, pp. 871-887, (1966).

[26] V.F. Pisarenko and D. Sornette, Statistical methods of parameter estimation for deterministically chaotic time series, *Phys Rev E*, **69** (2004).

[27] W. H. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes in C* (CUP, Cambridge, 1992).

[28] D.Ridout and K.Judd, Convergence properties of gradient descent noise reduction. *Physica D* **165** 27-48, (2001).

[29] T. Sauer, J.A. Yorke, M. Casdagli, Embedology, *J. Stat. Phys.* **65**, 579, (1991).

[30] D.S. Sivia, Data analysis - a Bayesian tutorial, Clarendon Press, Oxford University Press, (1997).

[31] L.A. Smith, Local Optimal Prediction: Exploiting strangeness and the variation of sensitivity to initial condition. *Phil. Trans. Royal Soc. Lond. A*, **348** (1688): 371-381, (1994).

[32] L.A. Smith, Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems. Nonlinear Dynamics and Statistics, A.I.Mees, Ed., Birkhauser, 31-64 (2000).

[33] D.A. Stainforth, M.R. Allen, E.R. Tredger and L.A. Smith, Confidence, uncertainty and decision-support relevance in climate predictions, *Phil. Trans. R. Soc. A*, **365**, 2145-2161, (2007).

[34] A. Tarantola, Inverse Problem Theory and Model Parameter Estimation. Philadelphia. PA: SIAM, (2004).

[35] J. Timmer, Modeling noisy time series: physiological tremor, *Int. J. Bifurcation and Chaos* **8**, 1505C1516 (1998).

[36] C. Ziehmann, L.A Smith and J. Kurths, Localized Lyapunov Exponents and the Prediction of Predictability, *Phys. Lett. A*, **271** (4): 237-251, (2000).