

Chaotic-time-series reconstruction by the Bayesian paradigm: Right results by wrong methods

Kevin Judd*

School of Mathematics and Statistics, The University of Western Australia, Nedlands, Western Australia 6009, Australia

(Received 2 July 2002; published 19 February 2003)

Recently, papers have appeared that champion the Bayesian approach to the analysis of experimental data. From reading these papers, the physicist could be forgiven for believing that Bayesian methods reveal deep truths about physical systems and are the correct paradigm for the analysis of all experimental data. This paper makes a contrary argument and is deliberately provocative. It is argued that the Bayesian approach to reconstruction of chaotic time series is fundamentally flawed, and the apparent successes result not from any degree of correctness of the paradigm, but by an accidental and unintended property of an algorithm. We also argue that (non-Bayesian) shadowing techniques provide all the information the erroneous Bayesian methods obtain, but much more efficiently, and also provide a wealth of additional useful information.

DOI: 10.1103/PhysRevE.67.026212

PACS number(s): 05.45.Tp, 02.60.Pn, 02.70.Rr, 06.20.Dk

I. INTRODUCTION

To illustrate our point we consider a recent paper by Meyer and Christensen [1], written in response to a paper by McSharry and Smith [2], concerning the problem of estimation of model parameters from time series of chaotic systems. The Meyer and Christensen paper presents the Bayesian approach as the correct statistical paradigm in this and other problems, claims to provide a correct solution to this particular problem, and makes passing criticism of physicists for developing and applying *ad hoc* methods. The authors apply their Bayesian methods to three problems and obtain apparently good estimates of system parameters. In this paper we argue that Meyer and Christensen make an incorrect Bayesian formulation of the problem and do not solve the problem as posed by McSharry and Smith. We repeat results of Berliner [3] that show that the correct Bayesian formulation of the problem necessarily fails to provide useful results. We then argue that the apparent success of the incorrect Bayesian formulation results from accidental and unintended properties of the methods. Finally we briefly describe established shadowing methods that intentionally exploit dynamical properties of systems (which the Bayesian methods accidentally exploit) to obtain all the information the Bayesian methods obtain, but much more efficiently, and also provide a wealth of additional useful information.

II. THE PROBLEM

The problem, in vague terms, is to determine the “correct” parameters for a *model*, given the inaccurate measurements of a *system*. It is important, at this early stage, to make and keep in mind the distinction between a system and model. The system is the physical experiment, the physical reality. The model is the mathematical representation of the system. These two are not the same, although much mathematical analysis assumes that the perfect model scenario is obtainable, that is, there is a model that is a perfect representation of the system (up to change of coordinates).

The system, in the problem that concerns us here, is assumed to be a smooth, finite-dimensional, discrete time, deterministic, autonomous, nonlinear dynamical system [4], that is, the state of the system at a time $t \in Z$ can be described completely by a point $x_t \in R^d$ and the evolution of the system state is given by $x_{t+1} = g(x_t)$, for some smooth nonlinear function $g: R^d \rightarrow R^d$. It is also assumed that one has a perfect model class, that is, a parametrized family of smooth functions $f(x, a)$, $f: R^d \times R^k \rightarrow R^d$, where there exists $\hat{a} \in R^k$ such that $g(x) = f(x, \hat{a})$.

The state of the system is observed over a period of time to obtain a finite sequence of observations $y_t \in R^d$, $t = 0, \dots, N$. These observations are inaccurate. The error in the observations (that is, the difference $y_t - x_t$) is assumed to be independent and identically distributed, that is, the error of any one observation is random, does not depend on the other observations in any way, nor does it depend on the state of the system. The errors are typically assumed to have a Gaussian distribution $N(0, \Sigma)$, where Σ is the covariance of the errors. We will assume that the errors have a probability density ρ ; for example, for the Gaussian distribution, the density $\rho(u)$ is proportional to $\exp(-u^T \Sigma^{-1} u/2)$.

The aim is to find \hat{a} , given the observations, or more correctly, the probability (density) that a particular value of a is the correct value \hat{a} , given the observations and assumed distribution of observation errors.

III. THE “CORRECT” BAYESIAN FORMULATION

The Bayesian formulation of the problem given in this section, and the arguments in the following section, are essentially those of Berliner [3]. At the end of this section we comment on the difference between this correct Bayesian formulation and the incorrect formulation of Meyer and Christensen.

It is useful to first develop the maximum likelihood approach to the problem. If the system trajectory visits a sequence of states $x = (x_0, \dots, x_N)$ and the corresponding observations form a sequence $y = (y_0, \dots, y_N)$, then the probability density of the observation at time t is $\rho(y_t - x_t)$, and since the errors are independent, we have the

*Electronic address: Kevin.Judd@uwa.edu.au

probability density of the observations y , given x , as

$$p(y|\rho, x) = \prod_{t=0}^N \rho(y_t - x_t). \quad (1)$$

The system is deterministic, that is, $x_{t+1} = g(x_t)$, which implies that x is entirely determined by x_0 , and so

$$p(y|\rho, x) = p(y|\rho, g, x_0) = \left(\prod_{t=1}^N \rho(y_t - g^t(x_0)) \right) \rho(y_0 - x_0), \quad (2)$$

where x_t is obtained by the iteration of g , that is, $x_t = g^t(x_0)$. Considering a model, one might construct a sequence of states $z_{t+1} = f(z_t, a)$, and write

$$p(y|\rho, f, a, z_0) = \left(\prod_{t=1}^N \rho(y_t - f^t(z_0, a)) \right) \rho(y_0 - z_0), \quad (3)$$

where a and z_0 are unknowns to be determined. However, this misses an important fact that the distribution of z_0 is not an unknown, because y_0 is known and the observation error distribution is known. So one can eliminate the dependency on z_0 and write

$$\begin{aligned} p(y|\rho, f, a) &= \int p(y_1, \dots, y_N | f, a, w) p(w | y_0) dw \quad (4) \\ &= \int \left(\prod_{t=1}^N \rho(y_t - f^t(w, a)) \right) \rho(y_0 - w) dw. \quad (5) \end{aligned}$$

One therefore obtains the likelihood function $L(a|\rho, f, y) = p(y|\rho, f, a)$. A maximum likelihood estimate of the parameters a maximizes the likelihood function.

The Bayesian approach is developed by considering

$$p(y_1, \dots, y_N | \rho, f, a, w) = \prod_{t=1}^N \rho(y_t - f^t(w, a)), \quad (6)$$

with the two unknowns a and w . The joint posterior density of the unknowns $p(a, w | f, y)$ is proportional to the likelihood (6) multiplied by the priors of the unknowns $p_a(a)$ and $p_w(w)$, that is,

$$p(a, w | \rho, f, y) \propto \left(\prod_{t=1}^N \rho(y_t - f^t(w, a)) \right) p_w(w) p_a(a). \quad (7)$$

An estimate of the probability density of a is obtained from Eq. (7) by generating random samples of w and a according to prior densities $p_a(a)$ and $p_w(w)$. This Monte Carlo Markov chain computation of the distribution of a would seem to be relatively trivial, as there are only two unknowns a and w , but we will see in Sec. IV that this is not the case.

The prior densities are often specified *ad hoc*, the argument being that the posterior density does not depend criti-

cally on the prior, although we shall shortly see that it usually will in this case. Observe that an informed prior density for w is $p_w(w) = \rho(y_0 - w)$. Also note the fact that when using an uninformative uniform prior density for a , the posterior density function of parameters, $p(a|\rho, f, y) = \int p(a, w | \rho, f, y) p_w(w) dw$, is identical (up to normalization) to the likelihood $L(a|\rho, f, y)$ in Eq. (5).

A. Comments

The preceding discussion provides the correct Bayesian formulation of the problem [3]. The formulation of Meyer and Christensen is not at all like the one just discussed. Their formulation is different, and incorrect, because they assume a stochastic system, or model, or both; it is not entirely clear which. Meyer and Christensen change the problem to suit their method. They assert that “a proper statistical paradigm requires treating the system states as stochastic instead of deterministic;” this is false as the preceding discussion demonstrates. Meyer and Christensen give only *ad hoc* justification for treating the deterministic problem like a stochastic problem, for example, that the stochastic formulation is more “realistic,” and that the deterministic case can be treated as just “small” dynamic noise. Such justifications are not sufficient on a technical level, and are disputable even on an informal level, because, for example, the reality is that many physical systems are indistinguishable from deterministic systems, there is no apparent small dynamic noise, and what is often attributed as such is in fact model error. (One could even argue that when a stochastic effect appears to be present, such as, say, *thermal noise*, it is just complex high-dimensional deterministic dynamics and is therefore really model error.) On the technical level, although the deterministic problem is a subcase of the stochastic problem, in the sense of the limit as stochastic perturbations go to zero, it does not follow that a solution to a stochastic problem provides a solution to the deterministic problem, because the solution spaces are quite different.

Meyer and Christensen appear to presume that their method is correct because the method appears to give correct estimates in the examples. In the following section we demonstrate that the correct *deterministic formulation* fails to provide useful estimates, and in Sec. V we explain why the incorrect *stochastic formulation* of Meyer and Christensen apparently gives correct results.

IV. FAILURE OF CORRECT BAYESIAN SOLUTION

Having developed in Sec. III the correct Bayesian approach to the problem, we demonstrate that this correct approach almost always fails. Our argument is essentially that of Berliner [3]. The reason for the failure is the very low probability of sampling near the true initial condition, that is, the need for an impossibly narrow and well-placed prior density.

Consider a very simple case of a ten-point trajectory of the logistic equation $g(x) = 1 - \hat{a}x^2$ with $\hat{a} = 1.85$, $x_0 = 0.3$, observations with Gaussian errors of standard deviation $\sigma = 0.1$, and the perfect model class $f(x, a) = 1 - ax^2$. Figure 1

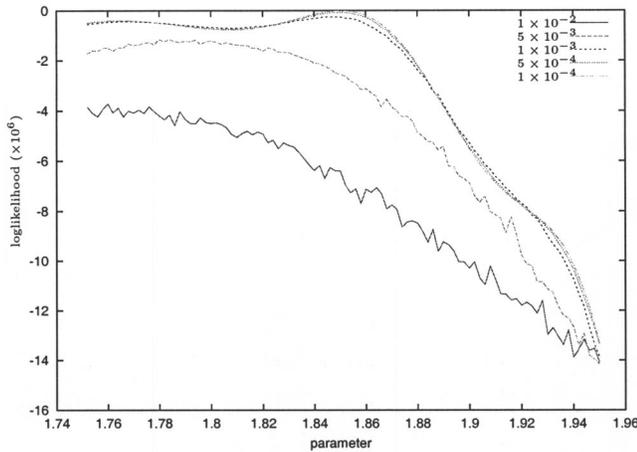


FIG. 1. Variation of the posterior density of parameter a for different prior densities on x_0 for a ten-point trajectory of the logistic equation $g(x) = 1 - \hat{a}x^2$ with $\hat{a} = 1.85$, $x_0 = 0.3$, observations with Gaussian errors of standard deviation $\sigma = 0.1$, and the perfect model class $f(x, a) = 1 - ax^2$. The prior density for x_0 is uniform, centered on the true value 0.3 with radii varying from 10^{-2} down to 10^{-4} as indicated. The logarithm of the likelihood is the natural logarithm. The densities are calculated from a 10 000-member Monte Carlo sample. Observe that unless the prior density is very narrow (two orders of magnitude less than the observational error, equivalent to almost knowing the initial condition), the true parameter value is not revealed. Alternatively, a prohibitively large sample is required to obtain a correct posterior density.

shows how the posterior probability density (likelihood) of the parameter a varies for various prior sampling distributions $p_w(w)$ that are uniform, centered on the true initial condition, and have different diameters. Observe that unless one samples very close to the correct initial condition, in this case two orders of magnitude less spread than the observation error ρ , the posterior density does not reveal the true parameter value. The reason is simply that unless the sampled initial condition is close to the true initial condition, the trajectory will diverge too rapidly from the true trajectory and give a trajectory too far from the observations.

If the trajectory is longer than ten points, the initial condition has to be sampled even more closely; for example once the trajectory exceeds about 70 points, the sampling accuracy must be better than 10^{-17} , the machine precision of the computer we used. Hence the posterior density of the initial condition is exceedingly narrow, and unless the prior density also reflects this, the posterior density of the parameter a is unreliable.

We conclude that in systems with sensitive dependence on initial conditions (chaotic), the posterior density of parameters will be unreliable, unless the initial condition prior density is narrow and close to the true initial condition. However, the most informed prior is the sampling error $p_w(w) = \rho(y_0 - w)$. This means a Bayesian approach will require an exponentially large Monte Carlo sampling as the trajectory length increases in order to have any chance of sampling near the correct initial condition.

V. WHY IS A WRONG METHOD APPARENTLY RIGHT?

We now reach an interesting juncture. On one hand, Meyer and Christensen use an incorrect stochastic Bayesian formulation of the problem that appears to provide accurate estimates of the parameters for a number of simple examples. On the other hand, the correct deterministic Bayesian formulation fails, in general, and certainly for the same examples. Should one then use the wrong formulation because it appears to work? The answer to such questions should always be no, it is better to understand what the Bayesian black box is doing and why it should appear to work. We argue that the success of the incorrect stochastic Bayesian formulation is the fortuitous consequence of Bayesian sampling, being a (very inefficient) method of finding shadowing pseudo-orbits.

There are several variant definitions of the concepts of pseudo-orbits and shadowing, very broad definitions are sufficient here. Pseudo-orbits are sequences of states of a model or system that are close to being trajectories, that is, for a system with dynamics given by $g: \mathbb{R}^d \rightarrow \mathbb{R}^d$, a sequence of states $x_t \in \mathbb{R}^d$, $t = 0, \dots, N$, should have the property that the *indeterminism* $L(x) = \sum_{i=1}^N \|x_i - g(x_{i-1})\|^2$ should be small. Observe that $L(x) = 0$ if the sequence of states is a trajectory. A sequence of states $x_t \in \mathbb{R}^d$, $t = 0, \dots, N$, is said to *shadow* another sequence of states $y_t \in \mathbb{R}^d$, $t = 0, \dots, N$, if $\sum_{i=0}^N \|y_i - x_i\|^2$ is small or minimal in some sense. Shadowing orbits and pseudo-orbits play an important role in the study of nonlinear dynamical systems and the study of imperfect models.

Consider the shadowing of a trajectory of a system $g(x) = f(x, \hat{a})$ by trajectories and pseudo-orbits of a smooth model family $f(x, a)$. It can be shown that if $a \neq \hat{a}$, then the model cannot have trajectories that shadow a trajectory of the system for arbitrary long times except under very strict conditions. In general, the further a is from \hat{a} , the shorter the period of time is for which shadowing of system trajectories by model trajectories can be sustained.

We note the following jumble of facts that will be used to explain the apparent success of the stochastic Bayesian formulation. In general, the closer a is to \hat{a} , the easier it is to find trajectories of the model that shadow a short sequence of the trajectory of the system, or a short sequence of observations of a trajectory. We note that the observations of the trajectory of a system form a pseudo-orbit of the system, with the indeterminism dependent on the observation error distribution. A solution of the stochastic equation $z_{t+1} = f(z_t, a) + v_t$, where v_t is $N(0, \tau^2)$, is a pseudo-orbit of the deterministic model. By construction, when τ is small enough, these pseudo-orbits shadow trajectories of the model for short periods. Observe that if pseudo-orbits obtained from the stochastic equation shadow segments of trajectories of the model, then there will exist pseudo-orbits that also shadow the segments of a trajectory of the system and the pseudo-orbit of observations of the trajectory. It is easier to find such simultaneous shadowing pseudo-orbits when a is close to \hat{a} .

The preceding facts work together to explain the success of the incorrect stochastic Bayesian formulation. The argument is this: The stochastic equations generate pseudo-orbits. The closer a is to the true \hat{a} , the greater is the chance that segments of these pseudo-orbits shadow segments of the ob-

servations. The better a pseudo-orbit shadows the observations (that is, the longer and closer it shadows), the higher its posterior density probability. Hence, the simulated posterior density probability ought to take a maximum when $a = \hat{a}$, because here there is the best chance of generating pseudo-orbits that partially shadow.

We observe that the sampling approach does not set out to find shadowing pseudo-orbits—it is quite inefficient at doing so—it just accidentally finds pseudo-orbits that partially shadow, which score highly in the posterior density, and hence reveal the correct parameter value.

Consequently, the stochastic Bayesian formulation is successful for estimating parameters of deterministic systems, not because it is the correct paradigm (it is an incorrect and irrelevant paradigm in this context), but because it fortuitously exploits shadowing properties of nonlinear dynamical systems.

VI. SHADOWING METHODS ARE A BETTER ALTERNATIVE

If the stochastic Bayesian formulation only fortuitously exploits shadowing properties of nonlinear dynamical systems, then it is appropriate to ask whether a direct approach that deliberately exploits shadowing properties is available and better. Shadowing methods do exist and we now argue that they can provide all the information Bayesian sampling provides, at considerably less cost, and provide closely shadowing pseudo-orbits that give a wealth of additional important information about the system dynamics and the model class.

Methods for finding shadowing pseudo-orbits have existed for some time (Davies [5], Farmer and Sidorowich [6], Grebogi *et al* [7], Hammel [8], Kostelich and Schreiber [9], Ridout and Judd [10]). Only a brief description and illustration is given here, for details see Ridout and Judd and citations therein.

One method for finding shadowing trajectories is the gradient descent method that takes an initial pseudo-orbit $x = y$, where y is the sequence of observations, then makes simultaneous adjustments of all the states of the pseudo-orbit x so as to move down the gradient of the indeterminism function $L(x)$. One achieves this by solving the differential equation

$$\frac{d}{d\tau}x(\tau) = -\frac{\partial}{\partial x}L(x(\tau)), \quad x(0) = y, \quad (8)$$

and finding the shadowing orbit from $\lim_{\tau \rightarrow \infty} x(\tau)$. In simple cases it is possible to solve the differential equation using a stiff integration routine. This algorithm has provable convergence properties and other benefits, for details see Ridout and Judd [10] and Judd [11].

We illustrate shadowing methods with the logistic example again. It should be noted that the gradient descent method used here does not depend on the covariance of the noise, so the problem of finding shadowing pseudo-orbits involves only model parameters, with the noise variance only

entering into the computation of the likelihood of a given shadowing pseudo-orbit. In this example we consider a 100-point trajectory of the logistic equation $g(x) = 1 - \hat{a}x^2$ with $\hat{a} = 1.85$, $x_0 = 0.3$, with Gaussian errors of standard deviation $\sigma = 0.1$, and the perfect model class $f(x, a) = 1 - ax^2$. Shadowing pseudo-orbits were obtained by integrating the descent equation (8) until $\tau = 1000$, which is not always sufficient to obtain a shadowing trajectory when the parameter a is far from the true value.

Figure 2 shows the variation of various quantities with the parameter a in this shadowing pseudo-orbit experiment. Observe from panel (b) that for half of the parameter values in the range shown, the gradient descent algorithm has converged to a shadowing *trajectory* to within machine precision. Observe from panel (c) that the pseudo-orbits obtained shadow the true trajectory very closely. In fact, when the initial x_0 of one of the shadowing pseudo-orbits is iterated, the trajectory it provides shadows the true trajectory nearly as closely as the pseudo-orbit itself for between 50 and 60 steps, which is close to the limits of machine precision. The likelihood shown in panel (d) of the parameter value a , given the shadowing pseudo-orbit and observations, is consistent with the stochastic Bayesian formulation calculations of Meyer and Christensen. The position of the maximum likelihood value of the parameter varies with the realization of the noise, as should be expected, and the variation is similar to that obtained by the stochastic Bayesian formulation calculations.

VII. CONCLUSIONS AND DISCUSSION

Recent papers have presented the Bayesian paradigm as the one correct method of data analysis, and presented certain algorithms as one-size-fits-all general purpose algorithms. One should have been wary of such claims because it is just too good to be true. We have argued that at least in the application to parameter estimation from chaotic time series this is demonstrably false, that is, the correct Bayesian formalism of the problem provides no useful information in practice. We have argued that the apparent success of a clearly inapplicable stochastic formulation is the consequence of subtle properties of chaotic systems, and is not evidence of some deep truth of the paradigm.

This is not to say that the Bayesian paradigm is not useful; we merely point out that there are situations where it is not, and that the physicist is right in developing more narrowly applicable methodologies that are more efficient and useful in situations where the Bayesian paradigm fails.

Physicists should be wary of mathematical and statistical methods presented as deep truths, just as they are used to insisting on experimental verification of physical theories. It is harder to apply such tests to the mathematical and statistical methods they are encouraged to use. In many cases, the computer experiments used to illustrate the methods are idealized and are not verification. At the very least it is important to probe the assumptions behind an impressive theory, and treat results as questionable when obtained in situations where assumptions do not hold. This is the case in the sto-

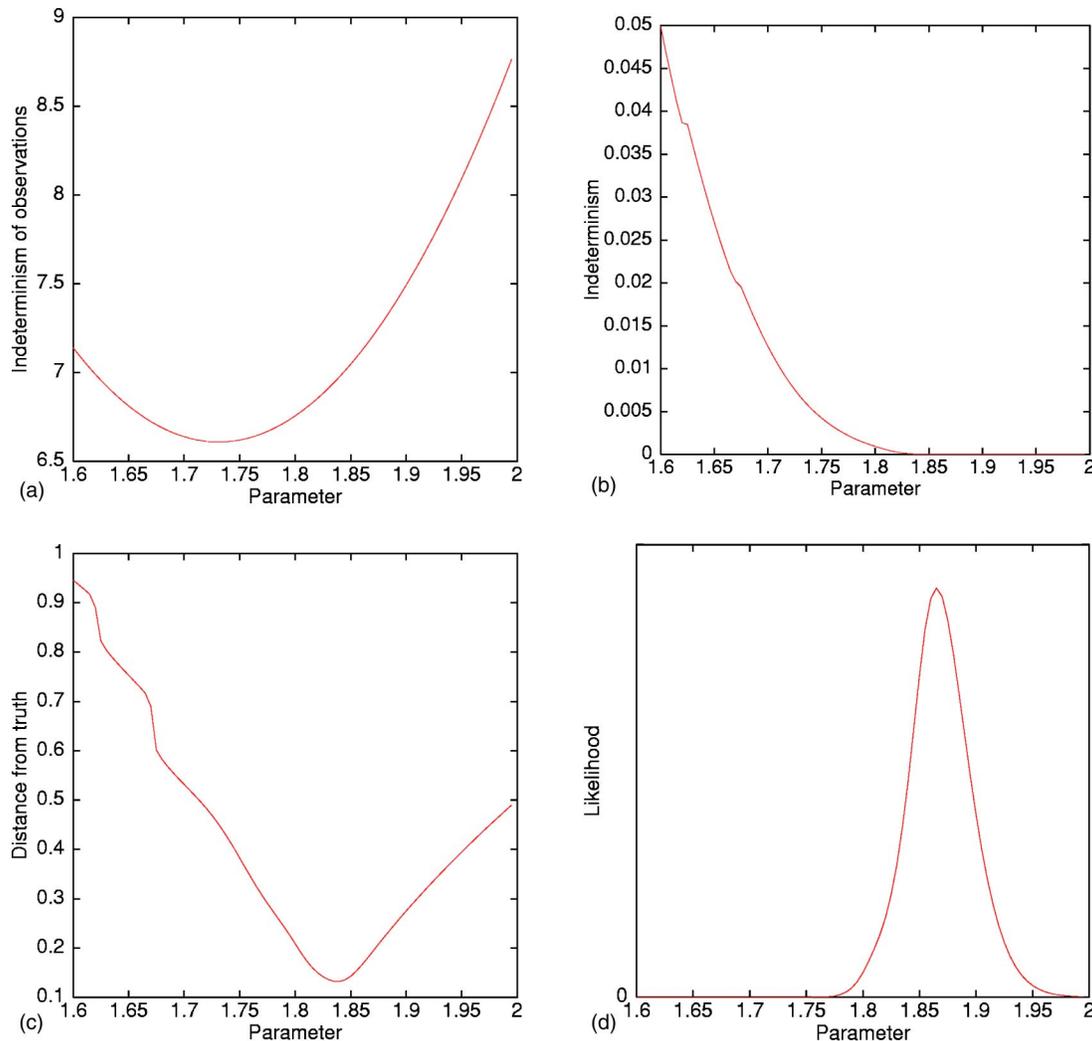


FIG. 2. Variation of various quantities with the parameter a for a 100-point trajectory of the logistic equation $g(x) = 1 - \hat{a}x^2$ with $\hat{a} = 1.85$, $x_0 = 0.3$, observations with Gaussian errors of standard deviation $\sigma = 0.1$, and the perfect model class $f(x, a) = 1 - ax^2$. (a) Shows the variation of the indeterminism of the observation pseudo-orbit with the model. (b) Shows the variation of the indeterminism of the shadowing pseudo-orbit obtained by gradient descent from the observations given a model. (c) Shows the distance $\sqrt{\sum_{t=0}^N \|\hat{x}_t - x_t\|^2}$ between the true trajectory of the system \hat{x}_t and the shadowing pseudo-orbit x_t . (d) Shows the likelihood of the shadowing pseudo-orbit, given the observations, assuming Gaussian errors of standard deviation $\sigma = 0.1$.

chastic formulation of the essentially deterministic problem discussed here, but there are deeper problems too. From the outset we have restricted attention to the *perfect model class scenario*. It is questionable whether such a situation can ever be realized, and this is a topic of active discussion among philosophers of science. The more likely scenario is the *imperfect model class scenario*, where there is no parameter value that gives a correct model. Indeed, the deterministic imperfect model scenario is arguably a more accurate model of most situations than the “realistic” stochastic model as-

serted by Meyer and Christensen; what is often attributed to noise is really model error.

Acceptance of the imperfect model class scenario leads to a number of profound consequences. In the current context we merely state, for the time being, that the Bayesian posterior probability, the probability that a parameter value is correct given the model class and observations, is usually a meaningless concept if the model class is imperfect. Going into this in any more detail takes us beyond the scope of this paper, but we intend to take up the theme in the future.

- [1] R. Meyer and N. Christensen, Phys. Rev. E **62**, 3535 (2000).
 [2] P. McSharry and L.A. Smith, Phys. Rev. Lett. **83**, 4285 (1999).
 [3] M.L. Berliner, J. Am. Stat. Assoc. **86**, 939 (1991).

- [4] In the paper of McSharry and Smith all these conditions were not explicitly stated, but they are implied. For smoothness it is sufficient to take differentiable conditions.

- [5] M. Davies, *Physica D* **79**, 174 (1994).
- [6] J.D. Farmer and J.J. Sidorowich, *Physica D* **47**, 373 (1991).
- [7] C. Grebogi, S.M. Hammel, J.A. Yorke, and T. Sauer, *Phys. Rev. Lett.* **65**, 1527 (1990).
- [8] S.M. Hammel, *Phys. Lett. A* **148**, 421 (1990).
- [9] E.J. Kostelich and T. Schreiber, *Phys. Rev. E* **48**, 1752 (1993).
- [10] D. Ridout and K. Judd, *Physica D* **165**, 27 (2001).
- [11] K. Judd (unpublished).