Available online at www.sciencedirect.com

SCIENCE DIRECT®

Physica D 196 (2004) 224–242

PHYSICA D

www.elsevier.com/locate/physd

Indistinguishable states II
The imperfect model scenario

Kevin Judd[a], Leonard A. Smith[b,c,*]

[a] *School of Mathematics, University of Western Australia, Perth 6009, Australia*
[b] *CATS, Statistics Department, LSE, London WC2A 2AE, UK*
[c] *Pembroke College, Oxford OX13LB, UK*

Received 13 August 2001; received in revised form 10 March 2004; accepted 28 March 2004

Communicated by F.H. Busse

Abstract

Given a perfect model of a chaotic system and a set of noisy observations of arbitrary duration, it is not possible to determine the state of this system precisely, rather one must consider a set of states which are indistinguishable from one another given the observations [K. Judd, L.A. Smith, Indistinguishable states I, Physica D 151 (2001)]. Yet the perfect model scenario is a fiction; in practice all models are imperfect. How do the results from the perfect model scenario change under imperfect models? It is shown to be essential to take even small model imperfections into account: failure to do so can systematically degrade state estimation or prediction of nonlinear systems. With an imperfect model, the system state space and model state space are rarely (if ever) equivalent, and so one must consider a projection of the system state into the model state space. Furthermore, it is almost certain that no trajectory of the model is consistent with an infinite series of observations, thus there is no consistent way to estimate the projection of system state using trajectories. There are pseudo-orbits, however, that are consistent with observations and these can be used to estimate the projection of the system state. Using pseudo-orbits one finds that, as in the perfect model scenario, there is a set of states that are indistinguishable from the projection of the system state. Estimation of this set of indistinguishable states and the probability density on these states is discussed. The main conclusions are (i) that there is no state of the model that can be identified with the state of the system; and (ii) that great care must be taken when using an imperfect model to forecast the system, because the initialization of the model state from observations can provide a poor analogue for the system. The forecast may not shadow the future behaviour of the system for very long, even if one were able to obtain a noise-free projection of the system state. The ultimate aims of probability forecasts should be re-examined in light of these results.

Keywords: Model; Noisy observations; Probability; Shadowing; Imperfect models; Forecasting

* Corresponding author.
E-mail address: lenny@maths.ox.ac.uk (L.A. Smith).

0167-2789/$ – see front matter © 2004 Published by Elsevier B.V.
doi:10.1016/j.physd.2004.03.020

1. Introduction

Given a perfect model of a chaotic dynamical system and an arbitrarily long series of noisy observations, it is not possible either to identify the current state of the system [10] or to accurately forecast a unique future state. It is possible, however, to define a probability distribution on a set of indistinguishable states, thereby allowing accountable probability forecasts [23]. In the current paper we consider the more relevant case where no perfect model is at hand, indeed where the class of available models does not contain a perfect model, and investigate the implications for state estimation again with a view towards prediction. We show that, in general, there will be no trajectory of the model which is consistent with the observations, that is, that the set of indistinguishable states is empty, and discuss the implications this holds for modelling a deterministic system when a good model is at hand. In short, our results suggest that one should consider pseudo-orbits of the model, not trajectories; that stochastic models may be required even if the underlying system is deterministic. Weather forecasting provides an example [18,17] where the physics of the system is fairly well understood, but the scale and complexity of the problem alone means that computer models are necessarily imperfect. How should one best initialize the state of the model to obtain a forecast when given noisy observations of the system?

After a brief summary of the perfect model scenario, we introduce the imperfect model scenario along with two particular realizations of model inadequacy [13]: structural inadequacy and ignored-subspace inadequacy. Our aim, of course, is to shed light on the ubiquitous problem in modelling physical systems where the nature of the model error is unknown. The special cases presented below are just that. In Section 3, we explore issues of state estimation in the imperfect model scenario, establishing a set of indistinguishable states for this case requires the introduction of pseudo-orbits of the model. Section 3.4 then considers the construction of indistinguishable sets when the true state is known; this is extended in Section 4 to consider the calculation given only noisy observations, that is, the case in practice. The implications our results hold for forecasting and forecast interpretation are discussed in Section 5, while our conclusions are briefly reviewed in Section 6.

In Ref. [10], we showed that, contrary to what might be expected, collecting more and more data will not provide a continually improving estimate of the true state of a chaotic system, in the sense that the estimate will not converge to the true state. Rather, there is always a set of states (spread along the unstable set of the true state) that are indistinguishable from the true state: the best estimate of the state one can achieve is a probability distribution on these indistinguishable states. Consequently, to forecast the future behaviour of a system one must either evolve a probability density of the indistinguishable states or evolve an ensemble of states drawn from the distribution of the indistinguishable states.

In the perfect model scenario, if the true trajectory of the system is x_t, $t = 0, -1, -2, \ldots$, then the final state x_0 is distinguishable with probability 1 from another state y_0, which is the final state of a trajectory y_t, if $Q_\rho(y_0|x_0) = 0$, where

$$Q_\rho(y_0|x_0) = \prod_{t\le 0} q_\rho(y_t - x_t),$$

$$q_\rho(b) = \frac{q_\rho(b)}{g_\rho(0)}, \tag{1}$$

$$g_\rho(b) = \int \rho(z)\rho(z-b)\,\mathrm{d}z,$$

and ρ is the probability density of the additive observation error. The set of indistinguishable states $H_\rho(x)$ of a state x is the set of states y such that $Q_\rho(y|x) > 0$. In this case $H_\rho(x)$ summarizes our knowledge of the current state of the system given the observations, conditioned on our knowledge of the model and the background knowledge that the model is perfect; clearly x is in $H_\rho(x)$ and if the system is chaotic then $H_\rho(x)$ includes much more. How do these conclusions change when our background knowledge includes the fact that the available model(s) is imperfect?

2. The imperfect model scenario

Outside pure mathematics, the perfect model scenario is a fiction. Arguably, there is no perfect model for any physical dynamical system [1–3,19,24]. The two central questions posed by the current paper are: whether assuming the perfect model scenario (hereafter, PMS) in the presence of model inadequacy can significantly degrade the conclusions drawn and if so, whether more productive alternative approaches exist. We answer "yes" to both questions. Yet while knowledge of the true system is not obtainable with real systems, some knowledge of the true system is required if firm mathematical results are to be established. In this section, we introduce two examples of model inadequacy, cases where both the true system and a class of models are known. We stress we are not attempting to resolve how to best model these particular cases (after all, we know this a priori as we know the true system!). Rather, we consider these model-system pairs in the hope of constructing methods which are relevant to the study of data from physical systems where no perfect model is known (if such a thing even exists).

One form of model inadequacy arises in a structurally incorrect model, where the system dynamics are not known in detail or cannot be expressed in terms of known mathematical functions. Arguably many "Laws of physics" are but useful approximations in restricted circumstances [3]. In an electronic circuit the resistances are assumed to follow Ohm's law, but there is always some nonlinearity, not to mention the unique features of this particular circuit. Consequently, the model is not exact, even though trajectories of the model may be (largely) qualitatively similar to those of the system and even quantitatively similar. Two related forms of model imperfection are (1) where one has the correct model class (that is, the model has the correct form), but the parameters in the model are incorrect, or (2) where one has a phenomenological model not derived from any physical principles which has been fitted to observed data (see [8] and references therein).

Another important type of model inadequacy is found in an ignored-sub space model where a system has a component of its dynamics that is unknown, unobservable, or not included in the model. Perhaps the unknown component is treated effectively as a heatbath [12] or ignored by careful consideration of evolution on a slow manifold [5]. Another example of the ignored-subspace model is where the model involves some course graining, or averaging, for example, a weather model [18,17] where model variables represent some sort of average of a system variable over region or "grid-box". A discussion of the role this type of model inadequacy plays in climate modelling can be found in [26]; the disconnect it causes between spatially distributed pointwise observations and weather models is sometimes called representativeness error [17].

2.1. Structurally incorrect model inadequacy

In the deterministic case there is a system $x_{t+1} = \Phi(x_t)$, $x_t \in K \subseteq R^d$. An imperfect model of the system will have the form $y_{t+1} = f(y_t)$, $y_t \in K$, where f defines dynamics that are not topologically conjugate to those defined by Φ. We will use the Ikeda [7] system as a simple example of this situation; the system has $x = (u, v) \in K = R^2$, and

$$\Phi(u, v) = \begin{pmatrix} 1 + \mu(u \cos\theta - v \sin\theta) \\ \mu(u \sin\theta + v \cos\theta) \end{pmatrix}, \tag{2}$$

where $\theta = a - b/(1 + u^2 + v^2)$, and $a = 0.4$, $b = 6$, $\mu = 0.83$. An imperfect model is obtained by replacing the trigonometric functions in Φ with truncated power series. The essential point here is that the resulting model is polynomial in u, v and θ, or if θ is eliminated, rational in u and v; models of this class are frequently derived as analytic approximations [19]. We will use the truncations

$$\begin{aligned} \cos\theta &= \cos(w + \pi) \mapsto -w + w^3/6 - w^5/120, \\ \sin\theta &= \sin(w + \pi) \mapsto -1 + w^2/2 - w^4/24, \end{aligned} \tag{3}$$

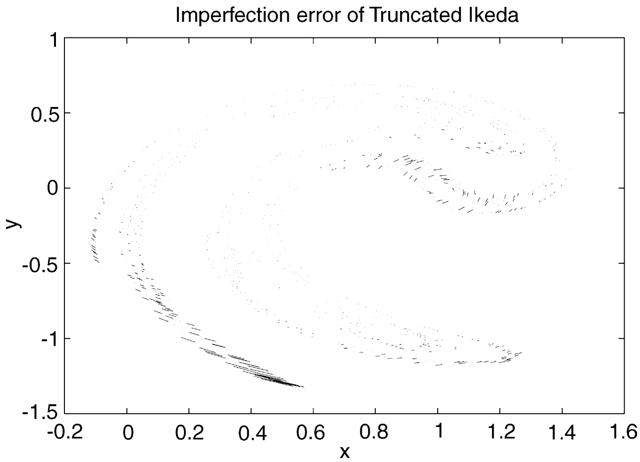

Fig. 1. The one-step prediction errors for the truncated Ikeda map. The small dots are 1000 points on the attractor of the Ikeda map, the lines show the prediction error for 500 points by linking the prediction to the target.

where the change of variable to it w is made since θ has the approximate range -1 to -5.5, and $-\pi$ is conveniently near the middle of this range.

Fig. 1 shows the one-step prediction error between the Ikeda system and the truncated Ikeda model. Generally, the truncated Ikeda model is a good predictor of the Ikeda system, but there are regions where it is not. Numerical investigation indicates that the maximum error is less than 0.15, in agreement with a calculation using the truncation error bound of Taylor's theorem.

2.2. *Ignored-subspace model inadequacy*

Consider a deterministic system with state space $K \times K' \subseteq R^d \times R^{d'}$, where the subspace K is observed and modelled and the subspace K' is either unknown or unobserved[1] and is not modelled. Thus, we have an imperfect model, which reflects only the dynamics on K, of the form $y_{t+1} = f(y_t)$, $y_t \in K$. A more complicated model–system pair introduced by Lorenz [14] has been discussed in the current context [24] and used to illustrate practical issues in weather forecasting [6]. A simple example is coupled Ikeda systems where only one system is modelled. The state space is $K \times K' = R^2 \times R^2$. Define the variables $x = (u, v) \in K$ for the modelled subspace and $x' = (u', v') \in K'$ for the unmodelled subspace. The dynamics are given by, $(x_{t+1}, x'_{t+1}) = F(x_t, x'_t)$, where F: $R^4 \to R^4$,

$$F(u, v, u', v') = \begin{pmatrix} 1 + \mu(u \cos\theta - v \sin\theta) - \gamma' u' \\ \mu(u \sin\theta - v \cos\theta) - \gamma' v' \\ 1 + \mu(u' \cos\theta' - v' \sin\theta') - \gamma u \\ \mu(u' \sin\theta' - v' \cos\theta') - \gamma v \end{pmatrix} \tag{4}$$

[1] Takens' theorem and subsequent generalizations [28] state that generically a time-delay embedding of $x_t \in K$ provides complete knowledge of the dynamics on K'. Theoretically it might seem the present formulation is unnecessary, but merely knowing that a diffeomorphism exists, while comforting, does not provide a perfect model.

where $\theta = a - b/(1 + u^2 + v^2)$, $\theta' = a - b/(1 + u'^2 + v'^2)$, with $a = 0.4$, $b = 6$, $\mu = 0.83$ and $\gamma = \gamma' = 0.02$. The imperfect model will be, $y_{t+1} = f(y_t)$, $y_t \in R^2$, where $F: R^2 \rightarrow R^2$,

$$f(u, v) = \begin{pmatrix} 1 + \mu(u \cos\theta - v \sin\theta) \\ \mu(u \sin\theta - v \cos\theta) \end{pmatrix}, \tag{5}$$

with θ, a, b and μ as above. Note that the coupling $\gamma' = 0.02$ implies that the imperfect model makes around a 2% error at each step. This prediction error does not have a zero mean, however, and a significant bias arises as the expected value of $x' = (u', v')$ is around $(0.67, -0.29)$.

3. A state consistent with observations

Is there a state of an imperfect model that is consistent with observations of the system? In general, no. Clearly there is the immediate problem that the imperfect model is not the same as the system; arguably their state spaces differ even in the case where they have the same dimension and share the same labels. As noted above, the state of the imperfect model can be taken to be a projection of the system state onto a model state. While we believe that this projection operator is important [24], and that much confusion has resulted from taking it to be the identity operator, we shall take it to be the identity operator throughout this paper, noting explicitly where doing so may cause difficulty. In general, the projections of system trajectories will not be trajectories of the model. With structurally incorrect models, the system and model have different dynamics which are not topologically conjugate. In the ignored-subspace models, one model state can represent many system states.

A consequence of the model and system having different dynamics is that no state of the model has a trajectory consistent with observations of the system. To accommodate these difficulties, we will consider pseudo-orbits rather than trajectories; these are sequences of states of the model x_t that at each step differ only slightly from trajectories, that is, $x_{t+1} \neq f(x_t)$; the (hopefully small) difference reflects imperfection error.

3.1. Imperfection error

Before proceeding we need some method to account for differences between the system and a model. Suppose x_t is the projection of a system trajectory into the model state space $K \subseteq R^d$. The model has dynamics $y_{t+1} = f(y_t)$, $y_t \in K$, so allowing for the imperfection of the model we expect that $x_t = f(x_{t-1}) + w_t$, with error terms $w_t \in R^d$. We will refer to the w_t as imperfection errors. If it were possible to obtain a better model one would have done so; for example, given a recurrent system one can, over time, identify systematic model errors and can therefore correct some of the imperfection error [9,25]. Henceforth, we will assume that all imperfection errors have been reduced to the minimum given the available information. By this definition, the actual imperfection errors cannot be known; in practice even statistical information about them (e.g., a bound on their magnitude) may be unavailable. For our theoretical development of indistinguishable states of imperfect models, it will be convenient to assume a distribution η for the imperfection errors w_t. When we come to estimation of indistinguishable states from data, an explicit distribution may need to be assumed; to be tractable, this usually requires either that the duration of the observations is such that the system is recurrent in the model state space (for a discussion, see [26]) or fairly general assumptions about the nature of the errors.

The imperfection error distribution η, and the interpretation of this distribution, will depend on the model scenario and the differences between the system and model. When assuming a structurally incorrect model, η represents the distribution of prediction errors of the model taken over the projection of the system attractor into the model state space as shown in Fig. 1. In an ignored-subspace model η also reflects the fact that there is a multiplicity of system states projecting onto the same model state, the realized trajectory depending on the actual system state.

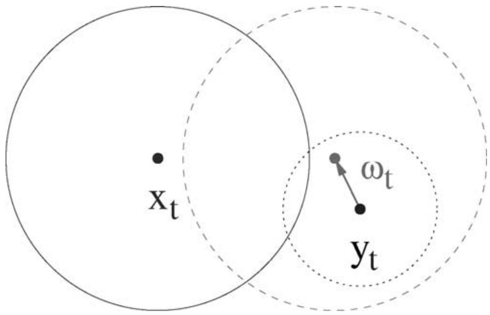

Fig. 2. Let x_t be the projection of the system state into model-state space and the solid circle represent the set of observations that could result from a bounded measurement error with distribution ρ. Let y_t be a model state and w_t represent an imperfection error of a bounded distribution η, represented by the smaller dotted circle. Given the bounded measurement error, the possible observations of the state $y_t + w_t$ lie in the dashed circle. An observation s_t of x_t is consistent with the state y_t if it falls in the overlap of the solid and dashed circle for some w_t. That is, x_t and y_t are indistinguishable on the basis of a single observation that falls in this overlap region for some w_t.

In reality η may be a strange (fractal) distribution; usually (hopefully) it will be a bounded distribution. While η might be modelled as time-varying or state-dependent or fractal, we choose to ignore these generalizations for clarity. We will assume η has a density with respect to Lesbegue, which we will also call η. Inasmuch as the actual distribution is, by definition, unknowable we can at best only guess a distribution; this admits these simplifying assumptions. For clarity, we will assume the imperfection errors w_t are independent and identically distributed.

3.2. Determinism and inconsistency

In this subsection, we develop the theory of indistinguishable states for the imperfect model scenario as a direct extension of the perfect model theory and show that it is almost certain that no trajectory of the imperfect model is consistent with observations. In the next section we modify the theory to use pseudo-orbits and show that at least some pseudo-orbits will always be consistent with observations. We stress that our aim is not to rectify the particular model imperfections introduced in the examples below, but to develop an approach which is of value in the ubiquitous case where the model imperfection is not known.

Henceforth, let $x_t \in K$ represent the projection of a system trajectory into the model-state space K. Suppose that at time t we make an observation s_t of x_t and that this observation is affected by observational uncertainty. Assume that $s_t = x_t + \varepsilon_t$, where $\varepsilon_t \in R^d$ and ε_t has density ρ with respect to Lesbegue measure,[2] and that the ε_t are independent and identically distributed. The results below generalize considerably from these assumptions; for example, ρ can be time-varying or state-dependent or fractal.

On the basis of a single noisy observation s_t of $x_t \in K$ there are other states $y_t \in K$ that are indistinguishable from x_t, see Fig. 2. The joint probability density of the projection of the system state x_t and a model state y_t being indistinguishable is given by[3]

$$\int \rho(s_t - x_t)\rho(s_t - y_t - w_t)\eta(w_t)\,\mathrm{d}s_t\,\mathrm{d}w_t. \tag{6}$$

[2] The density should not be singular, but rather more typical such as Gaussian or uniform on a disk.

[3] This analysis implies the model trajectory y_t at least shadows the projection of the system trajectory x_t when we assume that the shadowing error is distributed as η. Shadowing is certainly a necessary condition for consistency, but observational error and imperfection error usually do not have the same distribution. It remains to be seen just how harmful this distortion of meaning is in the present context.

Define

$$g(b) = \int \rho(z)\rho(z - b - w)\eta(w)\,\mathrm{d}z\,\mathrm{d}w, \tag{7}$$

$$q(b) = \frac{g(b)}{g(0)},$$

Observe that the joint probability (6) is $g(y_t - x_t)$ and that the conditional probability that x_t and y_t are indistinguishable, given that x_t is the projection of the system state, is $q(y_t - x_t)$. Also observe that if η is replaced by a delta-function at the origin (an atomic measure of mass one), then one recovers the perfect model scenario. For notational convenience we will frequently identify a trajectory of either the system or model with the state at time $t = 0$, furthermore, we will generally drop the zero subscript of this state and write, for example, x for x_0.

Given a time series of observations s_t, $t = 0, -1, -2, \ldots$, it follows (from the independence of the observational errors) that the probability that a model trajectory y_t is indistinguishable from the projection of the system trajectory x_t is given by,

$$Q(y|x) = \prod_{t\leq 0} q(y_t - x_t). \tag{8}$$

From which immediately follows:

Theorem 1. *Given any time series of observations, extending into the infinite past, of a system trajectory such that the projection of the system trajectory into model-state space terminates at x, and given a trajectory of a model that terminates at y, if* $Q(y|x) = 0$, *then the states x and y are distinguishable with probability* 1.

If $Q(y|x) > 0$, then $Q(y|x)$ is the probability that x and y will not be distinguished, given observations into the infinite past. Define $H(x)$, the set of indistinguishable states, as

$$h(b) = -\log q(b), \tag{9}$$

$$H(x) = \{y \in K : Q(y|x) > 0\} \tag{10}$$

$$H(x) = \left\{ y \in K : \sum_{t\leq 0} h(y_t - x_t) < \infty \right\}. \tag{11}$$

One-dimensional Gaussian error density: when $d = 1$ we have,

$$\rho(z) = \frac{1}{\sqrt{2\pi}\sigma}\,\mathrm{e}^{-z^2/2\sigma^2}, \tag{12}$$

$$\rho(w) = \frac{1}{\sqrt{2\pi}\zeta}\,\mathrm{e}^{-z^2/2\zeta^2}, \tag{13}$$

$$h(b) = \frac{b^2}{4\sigma^2 + 2\zeta^2}, \tag{14}$$

and $H(x)$ consists of all y such that $\sum_{t\leq 0}(y_t - x_t)^2 < \infty$.

Multi-dimensional Gaussian error density: when $d > 1$ and A^{-1} and B^{-1} are the symmetric co-variance matrices of ρ and η, respectively, then

$$\rho(z) = \frac{|A|}{\sqrt{2\pi}}\,\mathrm{e}^{-z^{\mathrm{T}}Az/2}, \tag{15}$$

$$\eta(w) = \frac{|B|}{\sqrt{2\pi}}\,\mathrm{e}^{-w^{\mathrm{T}}Bw/2}, \tag{16}$$

$$h(b) = b^{\mathrm{T}}Sb = b^{\mathrm{T}}(A - A(A+2B)^{-1}A)b, \tag{17}$$

and $H(x)$ consists of all y such that $\sum_{t\le 0}(y_t - x_t)^{\mathrm{T}}S(y_t - x_t) < \infty$, and because S is non-singular, the condition is equivalent to $\sum_{t\le 0}||y_t - x_t||^2 < \infty$. Note that in the case of isotropic error densities $A = I/\sigma^2$ and $B = I/\zeta^2$, the multi-dimensional case is equivalent to the $d = 1$ case.

Just as in the PMS, if $H(x)$ is non-trivial (that is, if it contains states other than x), then the state x cannot be distinguished with certainty from the other states in $H(x)$. But unlike in PMS, there is no necessity that $x \in H(x)$ when the model is imperfect. In fact, $H(x)$ could be empty, which has the embarrassing interpretation that no state of the model is consistent with the observations. This situation can arise even when the model trajectory remains in the proximity of (the observed part of) the system trajectory, that is, even though a proximity condition $||y_t - x_t|| < \varepsilon$, $t \le 0$, is satisfied, an indistinguishability condition $\sum_{t\le 0}||y_t - x_t||^2 < \infty$ is not.

Call the requirement that $x \in H(x)$ *asymptotic consistency*. For imperfect models asymptotic consistency can rarely be satisfied, indeed, $H(x)$ is almost always empty, for the following reason. An indistinguishability condition, for example, $\sum_{t\le 0}||y_t - x_t||^2 < \infty$, implies that the imperfect model has α-limit sets equivalent to those of the real system. This might be possible for systems with a finite number of α-limit sets, but in chaotic systems (defined by a continuous mapping) the attractor contains a dense set of unstable periodic orbits and equivalence of all the α-limit sets would imply, by continuity, that a continuous model is perfect on the attractor [16].

3.3. Pseudo-orbits and consistency

Next we modify the theory of indistinguishable states to use pseudo-orbits, rather than trajectories, and show that with this modification asymptotic consistency with observations is always assured. There are three obvious ways to deal with asymptotic inconsistency and empty $H(x)$. One is to weaken the notion of indistinguishability, that is, modify the definition of g and q in Eq. (7), so that proximity is a sufficient condition for indistinguishability. This approach is difficult to implement and seems to lead to a cumbersome, perhaps intractable theory. A second approach is to note that in practice one does not have access to infinite past data and so cannot calculate $Q(y|x)$ and hence determine $H(x)$. Instead one can calculate finite-time approximations

$$Q_T(y|x) = \prod_{-T\le t\le 0} q(y_t - x_t), \tag{18}$$

$$H_T(x) = \{y \in K : Q_T(y|x) > 0\}. \tag{19}$$

Clearly, as $T \to \infty$, $Q_T(y|x) \to Q(y|x)$ point-wise and $H(x) = \cap_{T\le 0} H_T(x)$. The function $Q_T(y|x)$ and the set $H_T(x)$ represent knowledge about the indistinguishable states when only finite information is available. The set $H_T(x)$ could still be empty, however, which represents the discovery that no state of the model is consistent with a finite number of observations. In general, reducing T increases the probability that $H_T(x)$ is non-empty, but only $H_0(x)$ can be guaranteed to be non-empty. This approach is unsatisfactory because the set $H_T(x)$ may be small either because the a state x is very predictable or because it is not very consistent with the observations.[4]

The third approach, developed below, considers pseudo-orbits of the imperfect model rather than trajectories. Observe that when a system trajectory is projected into the model state space K, the sequence of states visited $x_t \in K$ forms a pseudo-orbit of the model f, that is, one can write $x_t = f(x_{t-1}) + w_t$, and by assumption w_t has a

[4] This statement has a familiar ring to it: imperfect models often make confident predictions that are entirely wrong.

density η. Call the series x_t, which is a projection of the system trajectory into K, the true pseudo-orbit. In order to differentiate notationally between trajectories and pseudo-orbits, let $\tilde{z}$ denote a pseudo-orbit z_t that arrives at z at $t = 0$; $\tilde{x}$ can denote the true pseudo-orbit. Using pseudo-orbits is mathematically equivalent to replacing the deterministic imperfect model, $y_t = f(y_{t-1})$, with a stochastic model

$$z_t = f(z_{t-1}) + w_t, \tag{20}$$

where w_t has a density η. Observe, however, that the system is still considered deterministic: only the model becomes stochastic in order to overcome the imperfection of f when used as a deterministic model. Ideally one should improve the model f, but we have assumed we have taken all practical steps to do so, and our only option is to cope with the imperfect model. Note this is quite different from the common approach that assumes the system must be stochastic too.

In this formulation the joint probability density that a pseudo-orbit state z_t is indistinguishable from the true pseudo-orbit state x_t given the preceding pseudo-orbit state z_{t-1}, is given by,[5]

$$g(x_t, z_t|z_{t-1}) = \left(\int \rho(s_t - x_t)\rho(s_t - z_t)ds_t\right)\eta(z_t - f(z_{t-1})), \tag{21}$$

from which we can define

$$\begin{aligned} q_{\rho,\eta}(x_t, z_t|z_{t-1}) &= \frac{g(x_t, z_t|z_{t-1})}{g(x_t, x_t|x_{t-1})} = q_\rho(z_t - x_t)\frac{\eta(z_t - f(z_{t-1}))}{\eta(x_t - f(x_{t-1}))}, \\ Q_{\rho,\eta}(\tilde{z}|\tilde{x}) &= \prod_{t\leq 0} q_{\rho,\eta}(x_t, z_t|z_{t-1}), \end{aligned} \tag{22}$$

where $q_\rho(b)$ is the conditional probability that occurs in the PMS (see Eq. (1)), while $q_{\rho,\eta}(.)$ and $Q_{\rho,\eta}(\tilde{z}|\tilde{x})$ are the corresponding probabilities in the pseudo-orbit case of the imperfect model scenario. The following theorem is a result of these definitions.

Theorem 2. *Let $\tilde{x}$ be a true pseudo-orbit of a system, where $\tilde{x}$ extends into the infinite past and terminates at x. Let $\tilde{z}$ be a pseudo-orbit of an imperfect model with an imperfection error density η, where $\tilde{z}$ extends into the infinite past and terminates at z. If $Q_{\rho,\eta}(\tilde{z}|\tilde{x}) = 0$, then for observations of $\tilde{x}$ with an observational error density ρ the states x and z are distinguishable with probability* 1.

Let $H_{\rho,\eta}(x)$ be the set of states z for which there exists a pseudo-orbit $\tilde{z}$ with $Q_{\rho,\eta}(\tilde{z}|\tilde{x}) > 0$, that is, the set of states accessible by pseudo-orbits indistinguishable from the true pseudo-orbit. Returning to the example of Gaussian error densities it is seen in the $d = 1$ case, or in isotropic cases, that $H_{\rho,\eta}(x)$ is the set of pseudo-orbits such that

$$\sum_{t\leq 0}\left(\frac{1}{4\sigma^2}||x_t - z_t||^2 + \frac{1}{2\zeta^2}||z_t - f(z_{t-1})||^2 - \frac{1}{2\zeta^2}||x_t - f(x_{t-1})||^2\right) < \infty. \tag{23}$$

Clearly in this case $x \in H_{\rho,\eta}(x)$ and so we have asymptotic consistency. The terms of the above condition have a nice interpretation. The first term is identical to the indistinguishability condition in the PMS, see [10]. The last two terms represent the total square deviation of pseudo-orbits from trajectories of the imperfect model; these terms effectively require that for a pseudo-orbit $\tilde{z}$ to be indistinguishable from the true pseudo-orbit $\tilde{x}$, the deviation of $\tilde{z}$ from a trajectory of the model should be on average no worse than the deviation of the true pseudo-orbit. Also note that the expected value of $||x_t - f(x_{t-1})||^2$ is $\mu^2 + \zeta^2$, where μ and ζ^2 are the bias (mean) and variance of η. Thus, the expected value of the partial sum over $-T < t \leq 0$ of the last term of the inequality (23) is $(\mu^2 + \zeta^2)T$.

[5] Note that unlike Eq. (6), η here is precisely the imperfection error density.

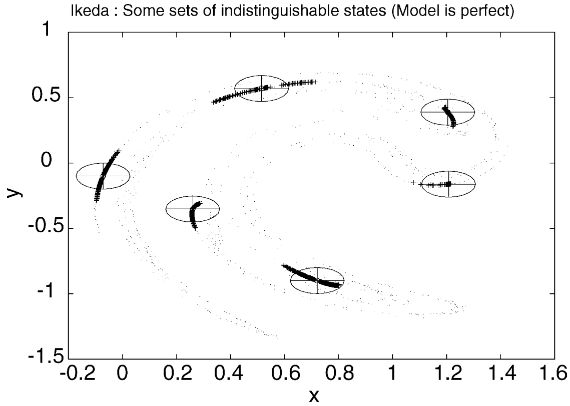

Fig. 3. Indistinguishable sets for six separate states in the perfect model scenario. The background of dots is the attractor of Ikeda system (2). The cross-hairs in large circles mark the true system states. The observation error is Gaussian with a $\sigma = 0.1$, which is the radius of the large circles with cross-hairs. The indistinguishable states are marked with plus signs.

When the imperfection error is non-Gaussian and bounded, asymptotic consistency holds provided the model has a pseudo-orbit remaining in the proximity of the true pseudo-orbit.[6] For uniform bounded observation error and uniform bounded imperfection error,

$$\rho(z) = \begin{cases} \dfrac{1}{V(R,d)}, & ||z|| \leq R, \\ 0, & \text{otherwise}, \end{cases} \tag{24}$$

$$\eta(w) = \begin{cases} \dfrac{1}{V(r,d)}, & ||w|| \leq r, \\ 0, & \text{otherwise}, \end{cases} \tag{25}$$

where $V(\cdot, d)$ is the volume of a d-dimensional ball, then $H_{\rho,\eta}(x)$ is the set of states z for which there are pseudo-orbits $\tilde{z}$ such that

$$\sup_{t\leq 0} ||x_t - f(x_{t-1})|| \leq r, \tag{26}$$

$$\sup_{t\leq 0} ||z_t - f(z_{t-1})|| \leq r, \tag{27}$$

$$\sup_{t\leq 0} ||x_t - z_t|| \leq R, \tag{28}$$

$$\sum_{t\leq 0} ||x_t - z_t|| < \infty. \tag{29}$$

Note that the first two inequalities ensure that the pseudo-orbits are consistent with η, and the last two inequalities are identical to those of the PMS with uniform bounded observational error [10].

It is instructive to contrast the indistinguishable sets of perfect and imperfect models. Fig. 3 illustrates indistinguishable sets for six different states in the perfect model scenario, and Fig. 4 illustrates the Q-density (22) for six

[6] Loosely speaking, one could say that the pseudo-orbit shadows the true pseudo-orbit; in doing so, however, care must be taken when interpreting the error bound. For ι-shadowing, this bound is based on the observational uncertainty as this interpretation is complicated by the projection from observation space to model-state space.

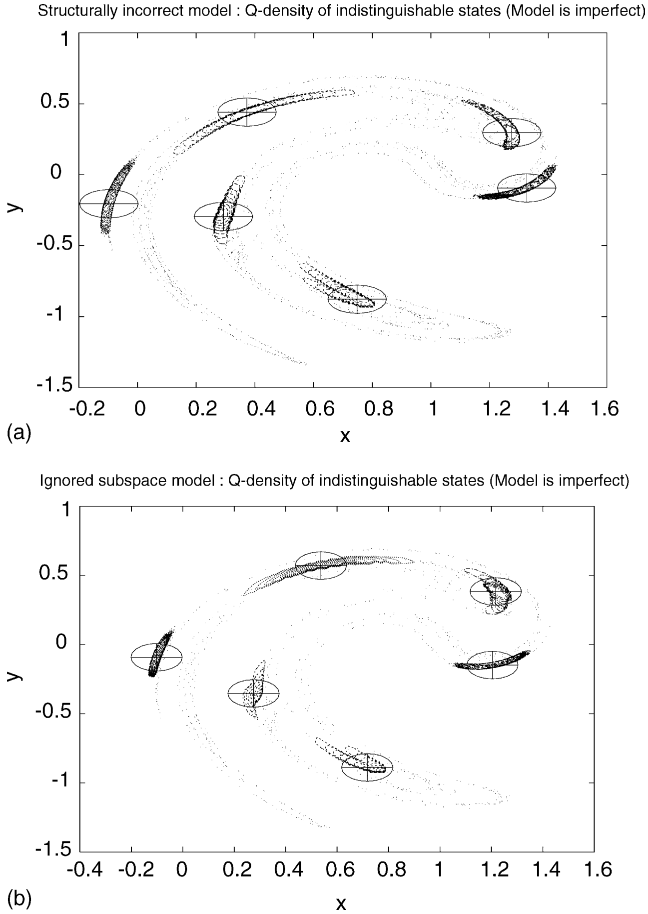

Fig. 4. The Q-density (22) for six separate states of (a) the structurally incorrect model (3) and (b) the ignored-subspace model (5). The background of dots is the attractor of the system. The cross-hairs in large circles mark the projections of system states. The observation error is Gaussian with standard deviation $\sigma = 0.1$, as reflected by the radius of the large circles with cross-hairs. The Q-densities of each state are represented by 10 equally spaced contour levels of probability. Note that projections of the system states are not the same in the two panels. Contrast these densities with those in figure (3).

similar states for the two imperfect models considered above. In PMS the indistinguishable sets are one dimensional and so the Q-density (1) is difficult to represent (see Ref. [10] for plots these densities). The character of the indistinguishable sets of comparable states can vary markedly between the perfect and imperfect scenarios. Generally, in the imperfect model scenario the sets become fatter (with a perfect model the sets are ultimately restricted to the unstable set of the true state). Comparing the indistinguishable sets of the perfect and imperfect model scenarios is not entirely justified, because there are states visited in the imperfect scenario that lie off of the attractor of the perfect model (that is, the model gives asymptotically zero probability of visiting a neighbourhood of the relevant state); such a situation is shown at the far left of Fig. 4(b).[7]

[7] In fact, the common confusion of the system state space with the model-state space [23], inadvertently promoted by the use of the identity as the projection operator between system and model, disallows this comparison.

3.4. Estimating indistinguishable sets for known states

In principle, calculating the indistinguishable states in the imperfect model scenario is no more difficult than in the perfect model scenario; in practice some details need to be appreciated. In PMS the indistinguishable states can be calculated from a sufficiently long segment of the true trajectory by making small (say uniform or Gaussian) perturbations of the initial state of the segment and calculating the resulting trajectory. The indistinguishable states are the final states of trajectory segments for which $Q_T(y|x)$ exceeds some threshold. The length of trajectory T and variance of initial perturbation are easily determined by trial: the length is determined by observing the convergence of Q_T and the variance of the perturbation by the magnitude and spread of Q. Once we know the maximum value of Q we can impose a significance threshold (so that $Q > 10^{-2}$, say) and then "cover" the most significant range of values of Q.

In the imperfect model scenario one has to calculate potential pseudo-orbits around the true pseudo-orbit,[8] that is, find solutions z_t to the stochastic model Eq. (20), and then calculate the finite-time approximation

$$Q_{\rho,\eta,T}(z|x) = \prod_{-T\leq t\leq 0} q_{\rho,\eta}(x_t, z_t|z_{t-1}). \tag{30}$$

Calculating random pseudo-orbits is not efficient because most pseudo-orbits rapidly diverge from the true pseudo-orbit, resulting in very small $Q_{\rho,\eta}$ probabilities. In other words, this simple approach may require computing a very, very large number of pseudo-orbits, most of which are immediately discarded. A more efficient approach is to perturb the true pseudo-orbit. If the true pseudo-orbit $\tilde{x}$ has deviations from being a trajectory $w_t = x_{t+1} - f(x_t)$ then construct another pseudo-orbit $\tilde{z}$ with deviations $z_{t+1} - f(z_t) = w_t + \kappa_t$. If the perturbations κ_t are chosen carefully, then potential pseudo-orbits can be generated more efficiently. For example, a method for a Gaussian distribution η of the imperfection errors w_t is to use Gaussian perturbations κ_t with standard deviations $\sigma_t = \sigma_0\, e^{\lambda t}$, $t = 0, \ldots, p$, where σ_0 and λ are chosen as follows. To determine σ_0 initially set $\kappa_t = 0$ for $t > 0$ and find a value for σ_0 so that the variance of the perturbation κ_0 results in pseudo-orbits whose Q values "cover" the most significant range of values of Q, as in PMS. Then select λ so that $\sigma_p = \sigma_0\, e^{\lambda p} = \zeta$, that is, the final imperfection has the same variance as η. The effect of this choice is that perturbations κ_t are initially small but grow exponentially in size; this ensures that the pseudo-orbits do not diverge too far from truth, but far enough to cover the significant part the indistinguishable set. For Fig. 4 we used a bounded uniform distribution η, but the alteration to the stated method is clear.

If a sample of states drawn from $Q_{\rho,\eta,T}$ is required, then a Monte Carlo Markov Chain (MCMC) method [4] could be used to generate them. The method described above provides an efficient generator for a Metropolis-Hastings implementation. With MCMC calculations there is always a probability normalizing factor that is conveniently avoided, and so it is here. For example, with Gaussian imperfection error, the last term of Eq. (23), which derives from the factors $\eta(z_t - f(z_{t-1}))$ in Eq. (22), is effectively a constant with an expected value $\mu^2 + \zeta^2$. When calculating $Q_{\rho,\eta,T}$ from a finite product, however, these factors contribute an indeterminate scaling factor. Consequently, it is convenient either to set this term to its expected value or ignore it entirely.

4. Maximum likelihood states

In the perfect model scenario we can obtain ensemble estimates of the true state by finding a maximum likelihood estimate of the state of a system and then obtaining an ensemble estimate of the indistinguishable set of this maximum likelihood estimate of the state [10]. Unfortunately, there are several difficulties in extending this procedure to the

[8] By "true pseudo-orbit" we mean the observations after they are assimilated into the model-state space; for simplicity we assume the identity operator as the projection operator throughout this paper.

imperfect model scenario. First, there is no meaningful maximum likelihood state of an imperfect model, only a maximum likelihood projection of the system state into model-state space. Second, there is an unavoidable confounding of observational uncertainty and model imperfection error, which has the consequence that maximum likelihood estimates of the projection of the system state can have a significant, and unavoidable, bias that is dependent on properties of the trajectory of the system. Third, there is no reason that the forecast initiated from the maximum likelihood model-state, however it is defined, will be particularly skillful.

4.1. The gradient descent algorithm

Within PMS a maximum likelihood estimate of a true state x can be found by minimizing a cost function and via gradient descent. The generalization of the cost function and optimization to imperfect models is as follows. For a finite sequence of observations, s_t, $t = 1, \ldots, p + 1$, define

$$e_t = s_{t+1} - \delta_{t+1} - \omega_{t+1} - f(s_t - \delta_t) \tag{31}$$

and

$$L(\delta, \omega) = \frac{1}{2}\sum_{t=1}^{p} e_t^{\mathrm{T}} e_t. \tag{32}$$

To find a suitable pseudo-orbit from the observations, solve

$$\min_{\delta,\omega} L(\delta, \omega), \tag{33}$$

by gradient descent [10], that is, solve the differential equations

$$\dot{\delta} = -\frac{\partial L}{\partial \delta}, \qquad \dot{\omega} = -\frac{\partial L}{\partial \omega}, \tag{34}$$

to compute the asymptotic values of (δ, ω) when starting from the initial values $(\delta, \omega) = 0$.

To find a maximum likelihood pseudo-orbit one could solve

$$\min_{\delta,\omega} L(\delta, \omega) + a\left(\frac{\delta^{\mathrm{T}}\delta}{\sigma^2} + \frac{\omega^{\mathrm{T}}\omega}{\zeta^2}\right), \tag{35}$$

for $a \to 0$, or iteratively solve the above while alternating $a = 0$ and $a > 0$.

Our algorithm differs from the perfect model case by introducing perturbations ω_t that are intended to allow for the imperfection of the model. It is clear that there is a confounding of the perturbations δ_t and ω_t, that is, it is not possible to determine whether a model prediction is incorrect as a result observation error δ_t or model error ω_t and thus it is not clear how the total error should be distributed between the two sources of error.[9] Thus, the maximum likelihood estimate of the projection of the system state can have a significant bias, depending on details of the true pseudo-orbit and how we choose to distribute error between the two sources; we return to this point later. Simply ignoring the ω_t terms is equivalent to assuming that the model is perfect; this is ill-advised.

Note that the additional terms ω_t play an important role. It is necessarily the case that $L(\delta, \omega)$ attains a minimum of zero. If a minimum of L occurs at $(\hat{\delta}, \hat{\omega})$, then it follows that $L(\hat{\delta}, 0) = (a/\zeta^2)\sum_t \hat{\omega}^{\mathrm{T}}\hat{\omega}$, that is, the amount by which the $s_t - \hat{\delta}_t$ fail to be a trajectory of the model is precisely $\hat{\omega}_t$. This would seem to imply that one could just minimize $L(\delta, 0)$ ignoring the ω terms. Doing so, however, is equivalent to assuming the model is perfect. This yields biased estimates of the projection of the system state, because it forces the solution to give a trajectory of the

[9] This problem might vanish if the model admits ι-shadowing trajectories over the duration of the observations; then one might arguably be within the perfect model scenario. We know of no dynamic physical examples [24].

model whenever one can be found,[10] but since the model is imperfect one should look for a solution in the broader class of pseudo-orbits. By including the terms ω_t in L, the gradient descent minimization of L arrives at a nearby pseudo-orbit that is less biased than the one obtained by gradient descent in the restricted sub-space where $\omega = 0$. In fact, when there is no trajectory of the model consistent with the data (or perhaps better said, only model trajectories with negligible probability given the data), then estimates of the state by gradient descent without ω terms can be wildly "inaccurate".

To better understand the confounding imperfection and observation error, consider the analysis of noise-free observations of a true pseudo-orbit. Ideally such a pseudo-orbit is unaltered, but it is more likely that the algorithm will find a "more probable" pseudo-orbit, attributing some of the deviation from a trajectory to be observational error; typically the larger imperfection errors will be altered more. This altering of noise-free pseudo-orbits results in biased estimates of the projection of the system state. The bias is dependent on particulars of the pseudo-orbit and is in general unavoidable, although other state estimation schemes may have better performance than our algorithm.

Figs. 5 and 6 show maximum likelihood estimates of the projection of the system states based on observations of six different true pseudo-orbits: firstly, under the assumption that the model is perfect, that is, $\omega = 0$ (panel (a)) and secondly under the assumption of an imperfect model (panel (b)). Recall within PMS the gradient descent algorithm applied to noisy observations provided maximum likelihood estimates of the state which lay in the indistinguishable set H_ρ of the true state. These estimates were distributed more or less as Q_ρ. Figs. 5(a) and 6(a) reveal a significant bias under the perfect model assumption. When presented with the noise-free true pseudo-orbit, the perfect-model gradient descent algorithm converges to a nearby trajectory. The noisy observations converge to states in the indistinguishable set H_ρ (that is, in the perfect model sense) of the noise-free solution. This is exactly as should be expected: the estimates lie in the indistinguishable set of a meaningless "true state".

It is important to note that there are cases in Figs. 5(a) and 6(a) where the maximum likelihood estimates of the state are far from the relevant projection of the true state. Thus, ensembles constructed under the perfect model assumption may have zero probability of containing the projection of the true state. Figs. 5(a) and 6(a) show that even under assumptions of an imperfect model the minimization of $L(\delta, \omega)$ can lead to biased estimates, that is, although the size and shape of the set of state estimates in Figs. 5(a) and 6(a) are similar to the densities of indistinguishable states of Fig. 4, they are sometimes displaced from their correct locations. Comparing panels (a) and (b) of Figs. 5 and 6 shows the bias under the imperfect model assumption is less than half the bias that occurred under the perfect model assumption. When presented with the noise-free true pseudo-orbit, the imperfect-model gradient descent algorithm can converge to a different pseudo-orbit, as there can be pseudo-orbits which in the presence of observational error are more likely than the true pseudo-orbit. We do not claim that the this approach is optimal. While this kind of bias is unavoidable,[11] other methods may be able to obtain better results than those obtained here.

5. Forecasting with imperfect models

Forecasting with an imperfect model is a dubious endeavor, yet all real-world forecasts are made in this context. Forecasts using an imperfect model under the assumption it is perfect are unnecessarily suboptimal. We postpone a number of issues regarding forecasting with imperfect models for the moment, and assume both sufficient wisdom to recognize a model is imperfect and sufficient knowledge to at least guess a distribution η for the imperfection errors that provides some form of an upper bound on the actual imperfection errors. Specifically, we try to avoid systematically underestimating imperfection errors. Given these assumptions, cautious ensemble forecasting with an imperfect model is described below.

[10] A similar approach, called 4DVAR, is used in operational weather forecastsing, see [18].

[11] There is a fundamental question of identifiability, perhaps even definition, which is not pursued here.

Clearly, forecasting with an imperfect model cannot be a better situation than forecasting with a perfect model. In the perfect model scenario we suggested [10] first obtaining a maximum likelihood state estimate $\hat{x}$, then basing a forecast ensemble $\varepsilon(\hat{x})$ on states from the set of indistinguishable states $H_\rho(\hat{x})$. Within PMS a symmetry exists in that x is the true state and $\hat{x}$ a maximum likelihood estimate of x, then $x \in H_\rho(\hat{x})$ and $\hat{x} \in H_\rho(x)$. In fact we found that within an ensemble $\varepsilon(\hat{x})$ we expect to find the true state x, or states close to it, with the same probability as obtaining the maximum likelihood estimate $\hat{x}$ given the true state x. In general, a desirable property of an ensemble is that it contain the true state in the following sense.

Definition 1. An ensemble ε is said to contain the state x, if x lies within the convex hull (or bounding box) of the ensemble ε.

Making a probability forecast by assuming an imperfect model is perfect is a downright dangerous practice, the forecasts are at best misleading. We have already seen in Figs. 5 and 6 how this perfect model assumption leads

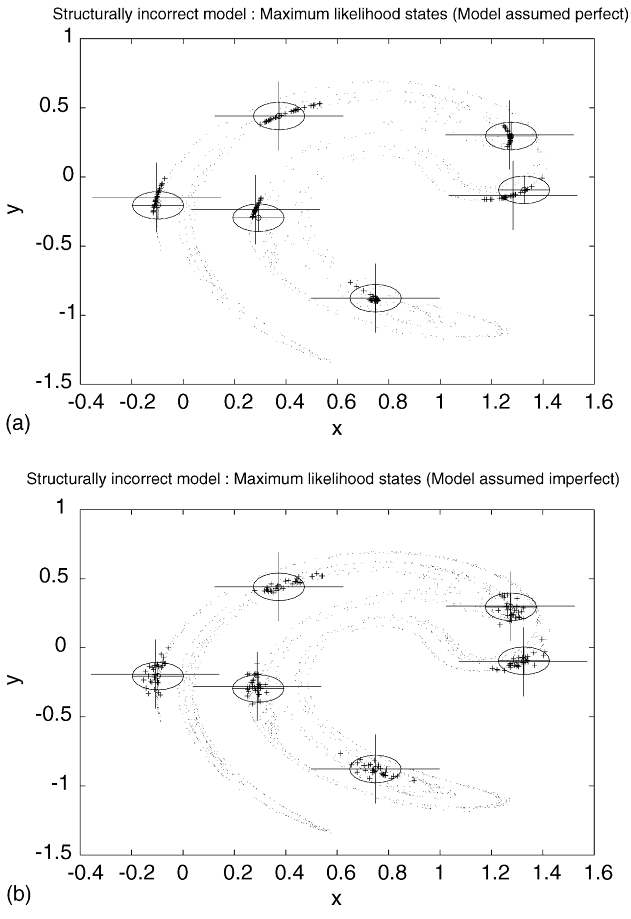

Fig. 5. Maximum likelihood estimates of states using the structurally incorrect model, obtained by using the model (2) with the truncation (3), under assumptions of (a) a perfect model and (b) an imperfect model. The cross-hairs in a large circle locate the six states. The observation error was Gaussian with mean zero and standard deviation $\sigma = 0.1$. The radius of the large circles corresponds to the standard deviation of the observation error. The background of dots is the attractor of the model. Maximum likelihood estimates were calculated for 30 different observations (noise realizations) of the same pseudo-orbit segments of 16 steps terminating on the marked states. The plus signs locate the state estimates obtained. The large cross-hairs not in circles locate the maximum likelihood estimate when the estimate was calculated from the noise-free pseudo-orbit. The misalignment of the circled and uncircled cross-hairs shows the bias of the state estimates.

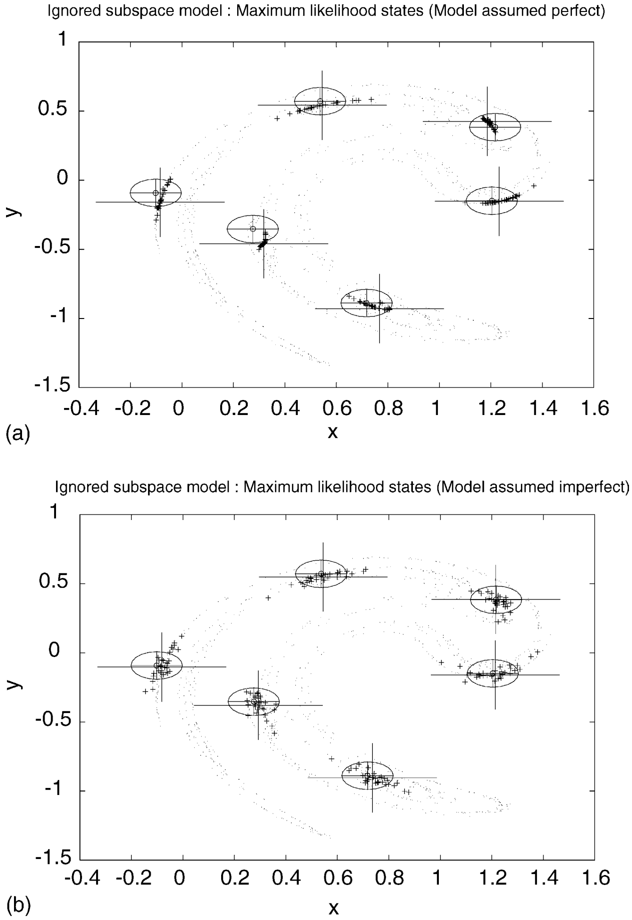

Fig. 6. Maximum likelihood estimates of model states using the ignored-subspace model (5) under assumptions of (a) a perfect model and (b) an imperfect model. The cross-hairs in a large circle locate the six states. The observation error was Gaussian with mean zero and standard deviation $\sigma = 0.1$. The radius of the large circles corresponds to the standard deviation of the observation error. The background of dots is the attractor of the model. Maximum likelihood estimates were calculated for 30 different observations (noise realizations) of the same pseudo-orbit segments of 16 steps terminating on the marked states. The plus signs locate the state estimates obtained. The large cross-hairs not in circles locate the maximum likelihood estimate when the estimate was calculated from the noise-free pseudo-orbit. The misalignment of the circled and uncircled cross-hairs shows the bias of the state estimates.

to an increased bias in the estimate of the maximum likelihood state $\hat{x}$; this is expected to result in a predictive disadvantage. It should be clear that constructing $H_\rho(\hat{x})$ for any of the states $\hat{x}$ in Figs. 5(a) and 6(a) will not overlap the projection of the true state x, especially when $\hat{x}$ lies far from the attractor of the system. This is, in fact, just a resurfacing of the problem of asymptotic consistency. One consequence of this is that an ensemble of states $\varepsilon(\hat{x})$ selected from $H_\rho(\hat{x})$ has zero probability of *containing* (in the above sense) the projection of the true state x, and is therefore of little forecast value. If the imperfections of a model are taken into account, then more useful ensembles can be obtained, and the same approach to finding a maximum likelihood estimate $\hat{x}$ of the projection of the system state x and an ensemble of states $\varepsilon(\hat{x}) \subset H_{\rho,\eta}(\hat{x})$ can be applied.

Referring again to Figs. 5 and 6, we observe that failing to take into account an imperfection error of less than 2% results in wildly inaccurate maximum likelihood estimates of the projection of the system state, even with noise-free observations (panel (a)). Of course, maximum likelihood estimates can be just as far from the projection of the system state when model imperfection is taken into account, but in this case, selecting the ensembles from $H_{\rho,\eta}(\hat{x})$ accounts for the variation, whereas selecting from $H_{\rho}(\hat{x})$ does not. Recall that Fig. 4 shows what typical densities of indistinguishable states look like. It should be clear that when ensembles are selected according to densities like these for any of the state estimates shown in Figs. 5(a) and 6(a), then there is a non-zero probability that they contain the projection of the true state x. When compared to their counterparts under the perfect model assumption, the state estimates $\hat{x}$ are less biased and the indistinguishable states are more spread out, particularly away from the unstable set of the state estimate.

5.1. Weighted ensembles

Often one would like to associate a weight with each ensemble member reflecting, for example, either the likelihood that it represents the projection of the true state or the relative value of its inclusion in an ensemble. Constructing a weighted-ensemble estimate of the projection of the system state using an imperfect model can proceed along the same lines used when the model is perfect. With a perfect model we found the maximum likelihood estimate of the state $\hat{x}$, and then selected members for the ensemble from the indistinguishable set of the maximum likelihood state $H_{\rho}(\hat{x})$. The weight of ensemble member y is $Q_{\rho}(y|\hat{x})$; in practice, a finite-time approximation Q_T is used. For an imperfect model the method is similar. A maximum likelihood estimate of the projection of the system state $\hat{x}$ is found as above. Elements in the indistinguishable set of $\hat{x}$ are found using the method described in Section 3.4 and weights are again assigned with $Q_{\rho,\eta}(y|\hat{x})$ or its finite-time approximation. An alternative to weighted ensembles is to select states from $Q_{\rho,\eta}(\hat{x})$ according to $Q_{\rho,\eta}(y|\hat{x})$ using a MCMC approach.

While methods of constructing an ensemble in the imperfect model scenario look reasonable at first sight, there is a crucial problem that they assume knowledge of the imperfection error distribution η, which we now recall is unknowable. To construct (and evolve) an ensemble in the imperfect model scenario requires guessing a reasonable stand-in for the distribution η. Given bounds for the errors (or their standard deviation), an appropriate uniform on a disk (or Gaussian) density might supply a guess for η. Usually this information is not immediately available, however, the gradient descent algorithm described in Section 4.1 provides useful information since it estimates the errors δ_t and ω_t, which are estimates of the observational error and imperfection error, respectively. We found that for all the 180 pseudo-orbits calculated for Fig. 6(b) the standard deviation of the observation error, estimated from the δ_t and ω_t, fell in the range from 0.070 and 0.17, and the standard deviation of the imperfection error fell in the range from 0.012 and 0.033. These numbers are in general agreement with the actual observational error of 0.1 and the actual imperfection error of around 0.02. The accuracy of both estimates give confidence in using the imperfection errors as information to obtain a guess of the imperfection error distribution η. Further analysis of this point is underway.

5.2. Forecasting

Interpreting ensemble forecasts with an imperfect model is no longer the simple matter of calculating the trajectories of the ensemble members as in the perfect model scenario. When the model is imperfect, a reasonable ensemble forecast will only be obtained by calculating pseudo-orbits. Ideally one would evolve the density of indistinguishable states using the stochastic Eq. (20), but it is impossible to represent this distribution analytically in practice (and would not yield the desired PDF even in principle). Given a weighed ensemble ε_0 at $t = 0$, one could generate a forecast weighted ensemble ε_t representing the forecast t steps into the future. One could construct ε_{t+1} by generating for each $z_t \in \varepsilon_t$ a number of forecast states $z_{t+1} = f(z_t) + \omega_t$ where the weight of each z_{t+1} is the product of the weight of z_t and $\eta(\omega_t)$. In practice some pruning of the ensemble ε_{t+1} may be necessary to prevent the ensemble growing too large. Pruning in the imperfect model scenario requires care, however, as it is not clear whether forecasts which would be considered “unlikely” in the perfect model scenario should be thinned or encouraged. Nevertheless, one might want to consider variants of the particle filter or SIR filter [21].

Without a perfect model and a perfect ensemble, an ensemble forecast will not be accountable [22], it will suffer from more than the effects expected from finite counting statistics. While there are a number of skill scores used in practice, there is no accepted optimal method for scoring such probability forecasts. We would recommend using proper skill scores [29,15], such as the ignorance score (see [20] and references thereof). It is not clear that one should interpret these ensembles as probability forecasts at all. Other aims, such as a ensemble that contains the verification within its convex hull, or its bounding box, provide alternative targets (see [24] and references thereof).

Once the unrealizable goal of finding a perfect model has been discarded, scientific motives for determining the 'best' model become clouded;[12] the value of using multiple models needs to be carefully re-examined. The use of multiple models may well provide new methods for estimating η as well as the best operational defense against model inadequacy, particularly when the various models have similar over-all prediction skill and very different structural assumptions. There are additional complications, including projections between the various model-state spaces and complications in interpreting the results [24], but this approach holds the prospect of increasing our understanding of the phenomena by easing the identification of shortcomings in our various models, in addition to providing more useful forecasts.

6. Conclusions

We have extended the concept of indistinguishable states and methods for calculating them to the case of imperfect models. In order to maintain consistency between observations of a system and an imperfect model it is necessary to study pseudo-orbits rather than trajectories. The theory applies to a wide variety of model imperfections. It is not, of course, our aim to solve the two particular examples of model error above (where we knew a perfect model and merely proceeded as if we did not); rather we aim to develop tools which can be applied when no perfect model is known (if such a thing even exists). It is, of course, impossible to prove anything about the relation of the system to the model in this scenario; nevertheless this is arguably the case with every physical system.

Treating an imperfect model as a perfect model yields an inaccurate estimate of the projection of the system state and an incoherent ensemble results. Consequently, it is essential to take model imperfections into account, failure to do so will result in degraded forecasts. These results are also relevant to parameter estimation, where there are two rather distinct situations: (i) the perfect model class, where there are parameter values that realize a perfect model, and (ii) imperfect model classes, which contain no perfect model for any parameter set. Work in progress addresses the question of simultaneously estimating the state and the parameter values of the model, where again we employ the gradient descent algorithm to good effect. The cost of obtaining gradient information in high (say, 10^7) dimensional models is not trivial even when a relevant adjoint is available [18]; we are also investigating methods of gradient-free descent [11].

To conclude we note that the very concept of "uncertainty in the initial condition" is brought into question in the imperfect model scenario. If there is no initial state in the whole of the model-state space which will ι-shadow the observations [27,24] over the forecast period of interest, then model inadequacy will limit accurate probability forecasts in a manner analogous to the way that uncertainty in the initial state limits accurate best-first-guess forecasts of chaotic systems in the perfect model scenario. In the imperfect model scenario, there need be no initial state to be uncertain of. How then, should one identify a good model? or progress towards a better understanding the underlying physical system?

References

[1] K. Beven, On explanatory depth and predictive power, Hydrol. Process. 15 (2001) 3069–3072.
[2] K. Beven, Towards a coherent philosophy for modelling the environment, Proc. Roy. Soc. A 458 (2002) 2465–2484.
[3] N. Cartwright, How the Laws of Physics Lie, Oxford University Press, 1983.

[12] By issues of metric dependence, if nothing else.

[4] R. Fitzgerald, Numerical Bayesian Methods Applied to Signal Processing, Springer, 1997.
[5] A.C. Fowler, M.C. Mackey, Relaxation oscillations in a class of delay differential equations, SIAM J. Appl. Math. 63 (1) (2002) 299–323.
[6] J.A. Hansen, L.A. Smith, The role of operational constraints in selecting supplementary observations, J. Atmos. Sci. 57 (17) (2000) 2859–2871.
[7] K. Ikeda, Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system, Opt. Commun. 30 (1979) 257.
[8] K. Judd, A.I. Mees, Embedding as a modeling problem, Physica D 120 (1998) 273–286.
[9] K. Judd, M. Small, Towards long-term prediction, Physica D 136 (1999) 31–44.
[10] K. Judd, L.A. Smith, Indistinguishable states I, Physica D 151 (2001) 125.
[11] K. Judd, L.A. Smith, Gradient free descent, Physica D 190 (2004) 153–166.
[12] W. Just, H. Kantz, C. Rodenbeck, M. Helm, Stochastic modelling: replacing fast degrees of freedom by noise, J. Phys. 34 (15) (2001) 3199–3213.
[13] M.C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, J. Roy. Stat. Soc. Ser. B 63 (2001) 425–464.
[14] E.N. Lorenz, Predictability—a problem partly solved, in: Proceedings of the ECMWF Workshop on Predictability, vol. 1, Shinfield Park, Reading, UK, 1996 (ECMWF).
[15] A.H. Murphy, R.L. Winkler, A general framework for forecast verification, Monthly Weather Rev. 115 (1987) 1330–1338.
[16] E. Ott, Chaos in Dynamical Systems, CUP, Cambridge, 1993.
[17] T.N. Palmer, R. Buizza, F. Molteni, Y.-C. Chen, S. Corti, Singular vectors and the predictability of weather and climate, Philos. Trans. R. Soc. Lond. A 348 (1688) (1994) 459–475.
[18] T.N. Palmer, Predicting uncertainty in forecasts of weather and climate, Rep. Prog. Phys. 63 (2000) 71–116.
[19] P.E. Rapp, T.I. Schmah, A.I. Mees, Models of knowing and the investigation of dynamical systems, Physica D 132 (1999) 133–149.
[20] M.S. Roulston, L.A. Smith, Evaluating probabilistic forecasts using information theory, Monthly Weather Rev. 130 (6) (2002) 1653–1660.
[21] D.B. Rubin, et al., Using the SIR algorithm to simulate posterior distributions, in: J. Bernardo (Ed.), Bayesian Statistics 3, Oxford Unversity Press, 1988.
[22] L.A. Smith, Accountability and error in ensemble forecasting, in: Proceedings of the ECMWF Seminar on Predictability, vol. 1, Shinfield Park, Reading, 1995, pp. 351–369 (ECMWF).
[23] L.A. Smith, The maintenance of uncertainty, in: Proceedings of the International School of Physics "Enrico Fermi", vol. CXXXIII, Bologna, Italy, 1997, pp. 177–246 (Societa Italiana di Fisica).
[24] L.A. Smith, Disentangling uncertainty and error: on the predictability of nonlinear systems, in: Alistair I. Mees (Ed.), Nonlinear Dynamics and Statistics, Birkhauser, Boston, 2000, pp. 31–64.
[25] L.A. Smith, Limits to predictability in 2000 and 2100, in: S. Haykin (Ed.), Proceedings of IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium, IEEE, Piscataway, 2000, pp. 129–134.
[26] L.A. Smith, What might we learn from climate forecasts? Proc. Natl. Acad. Sci. 4 (99) (2002) 2487–2492.
[27] L.A. Smith, C. Ziehmann-Schlumbohm, K. Fraedrich, Uncertainty dynamics and predictability in chaotic systems, Q. J. R. Meteorol. Soc. 125 (1999) 2855–2886.
[28] J. Stark, D. Broomhead, M. Davies, J. Huke, Delay embeddings for forced systems, J. Nonlinear Sci. 13 (2003) 519–577.
[29] R.L. Winker, A.H. Murphy, "Good" probability assessors, J. Appl. Meteor. 7 (1968) 751–758.