Indistinguishable states II:  
Imperfect model scenarios  

Kevin Judd*       Leonard Smith†  

January 13, 2003

Abstract

Given a perfect model of a chaotic system and a set of noisy observations of arbitrary duration, it is not possible to estimate the state of this system precisely, rather one must consider a set of states which are indistinguishable from one another given the observations (Judd and Smith *Physica D*, 151, 2001). Yet the perfect model scenario is potentially misleading since, in practice, all models are imperfect; how do the results from the perfect model scenario change under imperfect models? In short, it is shown that in any situation of state estimation or prediction of nonlinear systems, it is essential to take even small model imperfections into account. Failure to do so can systematically degrade the results. 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, for imperfect models it is almost certain that no trajectory of the model is consistent with an infinite series of observations, and consequently, 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, just as in the perfect model scenario, that there is a set of states that are indistinguishable from the projection of the system state. Estimation of the set of indistinguishable states and the probability density on these states is discussed. The two main conclusions of this study are first, that there is no state of the model that can be identified with the state of the system. And second, 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 give a model state that is 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 reexamined in light of these results.

1 Introduction

Given a perfect model of a chaotic dynamical system and an arbitrarily long series of noisy observations, it is not possible to identify the current state of the system [10] and thus impossible accurately forecast a unique state at some point in the future. It is possible, however, to define a probability distribution on a set of indistinguishable states, thereby allowing accountable probability forecasts [21]. 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 modeling 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 [15, ?] where the physics of the system

*Centre for Applied Dynamics and Optimization, Department of Mathematics and Statistics, The University of Western Australia, Perth

†Centre for the Analysis of Time Series, Department of Statistics, London School of Economics, London WC2A 2AE, and Oxford Centre for Industrial and Applied Mathematics, Mathematics Institute, Oxford
After a brief summary of the previous study of the perfect model scenario, we introduce the imperfect model scenario along with two particular realizations of model inadequacy [12]: structural inadequacy and ignored subspace inadequacy. Our aim, of course, is to shed light on the ubiquitous problem in modeling 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, with the aim of establishing a set of indistinguishable states for this case; it is shown that doing so requires the introduction of pseudo-orbits of the model, as we expect the set of indistinguishable states based on trajectories to be empty. 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 reference [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 a set of states (spread along the unstable set of the true state) that are always 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 (PMS) 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 one from another state \( y_0 \), which is the final state of a trajectory \( y_t \), if \( Q(\rho(y|x_0)|y_0) = 0 \), where

\[
Q(\rho(y|x_0)|y_0) = \prod_{t \leq 0} q(\rho(y_t - x_t)),
\]

\[
g_\rho(b) = g_\rho(b)/g_\rho(0),
\]

\[
g_\rho(b) = \int \rho(z)\rho(z - b) dz,
\]

and \( \rho \) 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. We now turn to the case were our background knowledge includes the fact that the model is imperfect and all information regarding that imperfection has been exploited.

### 2 The imperfect model scenario

Outside of pure mathematics, the perfect model scenario is a fiction. Arguably, there is no perfect model for any physical dynamical system [1, 2, 16, ?]. 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
There are any number of different varieties of model imperfections that we might consider. One form of model inadequacy thought to be very important arises when we have a \textit{structurally incorrect} model, where the system dynamics are not known in detail or cannot be expressed in terms of known mathematical functions. For example, many “Laws of physics” are not laws at all, 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. Similarly, the response characteristics of semiconductor devices cannot be known exactly and have to be approximated with incomplete theory or determined experimentally and approximated mathematically. 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 accurate to some extent. 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, for example, a radial basis, cylindrical basis, or neural network model, which has been fitted to observed data (see [8] and references therein).

Another important type of model inadequacy is found when one has an \textit{ignored subspace} model where a system has a component of its dynamics that is unknown, unobservable, or not included in the model. One example of this situation is where a system is weakly coupled to another system, but the model only describes the first system and the second system acts as an unknown perturbation. In theory, there may be instances where the unknown component may be treated effectively as a heat bath [11] or where it can be partially avoided by careful consideration of evolution on a slow manifold [5]. These are among the approaches that may improve a model, particularly when a complete mathematical description of the system is in hand. Another example of the ignored subspace model is where the model involves some course graining, or averaging, for example, a weather model [15, 7] 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 modeling can be found in [24]; the disconnect it causes between spatially distributed pointwise observations and weather models is sometimes called representativeness error [7].

In the next two subsections we formalize the discussion above in two particular realizations of these types of model inadequacy.

### 2.1 Structurally incorrect model inadequacy

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

\[
\Phi(u, v) = \begin{pmatrix}
1 + \mu(ucos\theta - v\sin\theta) \\
\mu(u\sin\theta + v\cos\theta)
\end{pmatrix},
\]

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 \( \Phi \) with truncated power series of these functions. The essential point of the truncation is that the resulting model is polynomial in \( u, v \) and \( \theta \), or if \( \theta \) is eliminated, rational in \( u \) and \( v \); models of this class are frequently fitted to data or are frequently derived as analytic approximations [16]. We will use the truncations

\[
\cos\theta = \cos(w + \pi) \rightarrow -w + w^3/6 - w^5/120,
\]

\[
\sin\theta = \sin(w + \pi) \rightarrow 1 + w^2/2 - w^4/24,
\]

where the change of variable to \( w \) is made because on the attractor of the Ikeda system \( \theta \) has the approximate range \(-1 \) to \(-5.5 \), and \(-\pi \) is conveniently near the middle of this range.
Figure 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 truncated Ikeda map prediction to the true state.

Figure 1 shows the one-step prediction error between 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 $K \times K' \subseteq \mathbb{R}^d \times \mathbb{R}^{d'}$, where the subspace $K$ is observed and modelled and the subspace $K'$ is either unknown or can not be observed\(^1\) and is not modelled. Thus we have an imperfect model of the system that only models the dynamics on $K$ and that the dynamics on the subspace $K'$ is not modeled. The model will have the form $y_{t+1} = f(y_t)$, $y_t \in K$. A more complicated system/model pair introduced by Lorenz [7] is discussed in the current context in reference [22] and applied to illustrate practical problems in weather forecasting by Hansen and Smith [6].

A simple example of this situation, which we will use for illustration in this paper, is coupled Ikeda [7] systems where only one system is modelled. The state space is $K \times K' = \mathbb{R}^2 \times \mathbb{R}^2$. 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 $F: \mathbb{R}^4 \rightarrow \mathbb{R}^4$,

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

\(^1\)Taken theorem and subsequent generalizations [7] 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. In practice we must contend with the effects of averaging, observation error, the finite duration of the observations and the like.
The imperfect model will be, \( y_{t+1} = f(y_t) \), \( y_t \in \mathbb{R}^2 \), where \( f : \mathbb{R}^2 \to \mathbb{R}^2 \),

\[
f(u, v) = \left( \begin{array}{c}
1 + \mu (u \cos \theta - v \sin \theta) \\
\mu (u \sin \theta + v \cos \theta)
\end{array} \right),
\]

with \( \theta, a, b \) and \( \mu \) as above.

The prediction errors of the model are more difficult to visualize and calculate in this case. To gain some idea recall that the attractor of the Ikeda system for the stated parameter values is contained in a disk approximately of diameter 2 and centred on \((0.67, -0.29)\). Observe that the coupling \( \gamma' = 0.02 \) between the two Ikeda subsystems implies that the imperfect model makes around a 2% error at each step. This prediction error does not have a zero mean, however, and there is a significant bias because the expected value of \( x' = (u', v') \) is around \((0.67, -0.29)\).

### 3 A state consistent with observations

We first address the question of whether there is a state of an imperfect model that is consistent with observations of the system, and show that, in general, there is no such state. 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 that 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 \cite{21}, 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. In case of structurally incorrect models, the system and model have different dynamics which are not be topologically conjugate. In the ignored subspace models, one model state can represent many system states. Here a model state has a unique future according to the model dynamics\(^2\) but there are many system trajectories whose projections arrive at and pass through this model state.

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, \( x_{t+1} \neq f(x_t) \), but the difference is not large.

#### 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 \( K \subseteq \mathbb{R}^d \). 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 \( x_t = f(x_{t-1}) + \omega_t \), with error terms \( \omega_t \in \mathbb{R}^d \). We will refer to the \( \omega_t \) as imperfection errors. If one could 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 \cite{9, 23}. Henceforth we will assume that all imperfection errors have been reduced to the minimum given the available information. Consequently, by this definition, the actual imperfection errors cannot be known; in practice even statistical information about them (e.g. a bound on their magnitude) is unavailable. For our theoretical development of indistinguishable states of imperfect models it will be convenient to assume a distribution \( \eta \) for the imperfection errors \( \omega_t \). When we come to estimation of indistinguishable states from data, an explicit distribution may need to be assumed; to be tractable, this usually requires that the duration of the observations is such that the system is recurrent in the model state space (for a discussion, see [24]), or we are prepared to accept fairly general assumptions about the nature of the errors.

\(^2\)And if the model is invertible, then it has a unique prehistory as well.
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. This development will 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 $s_t = x_t + \epsilon_t$, where $\epsilon_t \in \mathbb{R}^d$ and $\epsilon_t$ has density $\rho$ with respect to Lebesgue measure, and that the $\epsilon_t$ are independent and identically distributed. The results below generalize considerably from these assumptions; for example, $\rho$ can be time varying or state dependent or fractal, but we choose to ignore these generalizations for clarity.

On the basis of a single observation $s_t$ of $x_t \in K$ there are other states $y_t \in K$ that are indistinguishable from $x_t$, because the model has an unknown imperfection and because there is observational uncertainty; see figure 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

$$\int \rho(s_t - x_t) \rho(s_t - y_t - \omega_t) \eta(\omega_t) \, ds_t \, d\omega_t. \quad (6)$$

Define

$$g(b) = \int \rho(z) \rho(z - b - w) \eta(w) \, dz \, dw,$$

$$q(b) = g(b)/g(0). \quad (7)$$

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 $\eta$ 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 $x_0$.

---

3The density should not be singular, but rather more typical such as Gaussian or uniform on a disk.

4This 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 $\eta$. Shadowing is certainly a necessary condition for consistency, but observational error and imperfection error usually do not have the same distribution. In practice, much is often known about the distribution of the observational error, but issues of representativeness and projection [15] and the fact that the distribution $\eta$ is unknown and somewhat arbitrary, suggest that this assumption may be allowed. It remains to be seen just how harmful this distortion of meaning is in the present context.
Figure 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 $\rho$. Let $y_t$ be a model state and $\omega_t$ represent a possible imperfection error of a bounded distribution $\eta$, which is represented by the smaller dotted circle. Given the bounded measurement error the possible observations of the state $y_t + \omega_t$ lie in the dashed circle. An observation $s_t$ of $x_t$ will be consistent with the state $y_t$ if it falls in the overlap of the solid and dashed circle for some $\omega_t$. That is, $x_t$ and $y_t$ will be indistinguishable on the basis of a single observation that falls in this overlap region for some $\omega_t$.

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,

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

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 one.

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

$$h(b) = -\log q(b),$$

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

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

We now illustrate the case of Gaussian errors.

**One dimensional Gaussian error density:** When $d = 1$ we have,

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

$$\eta(w) = \frac{1}{\sqrt{2\pi}\xi} e^{-w^2/2\xi^2},$$

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

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

7
\[ \rho(z) = \frac{|A|}{\sqrt{2\pi}} e^{-z^T A z / 2}, \]  
\[ \eta(w) = \frac{|B|}{\sqrt{2\pi}} e^{-w^T B w / 2}, \]  
\[ h(b) = b^T S b = b^T (A - A(A + 2B)^{-1} A) b, \]
and \( H(x) \) consists of all \( y \) such that \( \sum_{t \leq 0} (y_t - x_t)^T S (y_t - x_t) < \infty \), and because \( S \) is non-singular, the condition is equivalent to \( \sum_{t \leq 0} ||y_t - x_t||^2 < \infty \). Note that in the case of isotopic error densities \( A = I / \sigma^2 \) and \( B = I / \kappa^2 \), the multi-dimensional case is equivalent to the \( d = 1 \) case.

Just as in the perfect model scenario, 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 \( H(x) \). Unlike in PMS, however, 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 \( ||y_t - x_t|| < \epsilon, t \leq 0 \), is satisfied, an indistinguishability condition \( \sum_{t \leq 0} ||y_t - x_t||^2 < \infty \) 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, \( \sum_{t \leq 0} ||y_t - x_t||^2 < \infty \), implies that the imperfect model has \( \alpha \)-limit sets equivalent to those of the real system. This might be possible for systems with a finite number of \( \alpha \)-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 \( \alpha \)-limit sets would imply, by continuity, that a continuous model is perfect on the attractor [?].

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 equation (7), so that proximity is a sufficient condition for indistinguishability. This approach is difficult to implement and seems to lead to a cumbersome or intractable theory; furthermore, the third option to come is a better approach.

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 cannot determine \( H(x) \). Instead one can calculate finite-time approximations

\[ Q_T(y|x) = \prod_{-T \leq t \leq 0} q(y_t - x_t), \]

\[ H_T(x) = \{ y \in K: Q_T(y|x) > 0 \}. \]

Clearly, as \( T \to \infty \), \( Q_T(y|x) \to Q(y|x) \) point-wise and \( H(x) = \bigcap_{T \leq 0} H_T(x) \). The function \( Q_T(y|x) \) and the set \( H_T(x) \) represent knowledge about the indistinguishable states when only finite information is available. \( H_T(x) \) could still be empty, however, which represents the discovery that no state of the model is consistent with a finite number of observations, that is, finite time inconsistency is still a possibility. 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 not very satisfactory 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\(^5\).

\(^5\)This statement has a familiar ring to it. That imperfect models based on limited data often make confident predictions that are entirely wrong.
rather than trajectories. In some senses this approach combines the motivations of the previous two approaches by requiring proximity but accepting that only \( H_0(x) \) is guaranteed to be non-empty. 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 \( x_t = f(x_{t-1}) + \omega_t \), and by assumption \( \omega_t \) has a density \( \eta \).

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 \( \tilde{x} \) denote a pseudo-orbit \( z_t \) that arrives at \( z \) at \( t = 0 \), for example, \( \tilde{x} \) 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

\[
z_t = f(z_{t-1}) + \omega_t, \tag{20}
\]

where \( \omega_t \) has a density \( \eta \). 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, which should only be done if there is compelling reason to believe the system is not deterministic.

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:\(^6\)

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

from which we can define

\[
q_{\rho, \eta}(x_t, z_t | z_{t-1}) = g(x_t, z_t | z_{t-1}) / g(x_t, x_t | x_{t-1})
\]

\[
= q_\rho(z_{t} - x_t) \frac{\eta(z_t - f(z_{t-1}))}{\eta(x_t - f(x_{t-1}))},
\tag{22}
\]

\[
Q_{\rho, \eta}(\tilde{z} | \tilde{x}) = \prod_{t \leq 0} q_{\rho, \eta}(x_t, z_t | z_{t-1}),
\]

where \( q_\rho(b) \) is the conditional probability that occurs in the perfect model scenario (see equations (1)), while \( q_{\rho, \eta}(\cdot) \) and \( Q_{\rho, \eta}(\cdot | \cdot) \) 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 \( \eta \), 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 \( \rho \) the states \( x \) and \( z \) are distinguishable with probability one.

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 isotopic cases that \( H_{\rho, \eta}(x) \) is the set of pseudo-orbits such that

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

Clearly in this case \( x \in H_{\rho, \eta}(x) \) and so asymptotic consistency is satisfied. The terms of the above condition have a nice interpretation. The first term is identical to the indistinguishability condition in the perfect model scenario as derived [10]. The last two terms represent the total square deviation

\(^6\)Note that unlike equation 6, \( \eta \) here is precisely the imperfection error density.
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 there are some details that need to be

7Loosely 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 $i$-shadowing, this bound is based on the observational uncertainty, but in general, this interpretation is complicated by the projection from observation space to model state-space. A general discussion of the issue is beyond the scope of this paper, which avoids the issue by using the identity as the projection operator.

8In fact, the common confusion of the system state space with the model state space (inadvertently promoted in the current paper by the use of the identity as the projection operator between system and model) disallows this comparison. For a discussion, see Smith [21]
Figure 3: Indistinguishable sets for six separate states in the perfect model scenario, where the system and model are the identical, in this case the Ikeda system (2). The background of dots is the attractor of the system. The states marked with cross-hairs in large circles are the true system states. The observation error is Gaussian with standard deviation $\sigma = 0.1$; the radius of the large circles with cross-hairs represent this standard deviation. The states marked with plus signs are states indistinguishable from the true system states.

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 $Q_T(y|x)$ is larger than a chosen threshold. The required 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 $Q$, that is, we know the maximum value of $Q$ is one and can require a significance threshold so that $Q > 10^{-2}$, say. In this way we can “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 \(^9\), that is, find solutions $z_i$ to the stochastic model equation (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, however, 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 then immediately discarded. A more computationally efficient approach is to perturb the true pseudo-orbit. If the true pseudo-orbit $\bar{x}$ has deviations from being a trajectory $\omega_t = x_{t+1} - f(x_t)$, then construct another pseudo-orbit $\tilde{x}$ with deviations $z_{t+1} - f(z_t) = \omega_t + \kappa_t$. If the perturbations $\kappa_t$ are chosen carefully, then potential pseudo-orbits can be generated more efficiently. For example, a method for a Gaussian distribution $\eta$ of the imperfection errors $\omega_t$ is to use Gaussian perturbations $\kappa_t$ with standard deviations $\sigma_t = \sigma_0 e^{\lambda t}$, $t = 0, \ldots, p$, where $\sigma_0$ and $\lambda$ are chosen as follows. To determine $\sigma_0$ initially set $\kappa_t = 0$ for $t > 0$ and find a value for $\sigma_0$ so that the variance of the perturbation $\sigma_0$

\(^9\)By the 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 in the examples in this paper.
perfect model scenario. Then select \( \lambda \) so that \( \sigma_p = \sigma_0 e^{\lambda \nu} = \zeta \), that is, the final imperfection has the same variance as \( \eta \). The effect of this choice is that perturbations \( \kappa_i \) 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 figure 4 we used a bounded uniform distribution \( \eta \), 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 equation (23), which derives from the factors \( \eta(z_i - f(z_{i-1})) \) in equation (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.
Figure 4: The density of indistinguishable states, that is, the $Q$-density (22), for six separate states of (a) the structurally incorrect model (3) of the Ikeda system (2) and (b) the ignored subspace model (5). The background of dots is the attractor of the system. The states marked with cross-hairs in large circles are the projections of system states. The observation error is Gaussian with standard deviation $\sigma = 0.1$; the radius of the large circles with cross-hairs represent this standard deviation. The $Q$-densities of each state are represented by ten equally spaced contour levels of probability. Note that projections of the system states are not necessarily the same in the two panels, but they have been chosen to be close for comparison. Also compare these densities with the perfect model scenario in figure (3), and note in these examples of the imperfect model scenario the projection of the system states can lie off of the attractor of the model, whereas this does not happen in the perfect model case; note, for example, the state at the far left of panel (b).
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 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 to believe that the forecast initiated from the maximum likelihood model-state, however it is defined, will be particularly skillful.

4.1 The gradient descent algorithm

In the perfect model scenario a maximum likelihood estimate of a true state \( x \) could be found by minimizing a cost function. The nature of the cost function meant that gradient descent would be a suitable optimization technique. 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

\[
e_t = s_{t+1} - \delta_{t+1} - \omega_{t+1} - f(s_t - \delta_t)
\]

and

\[
L(\delta, \omega) = \frac{1}{2} \sum_{t=1}^{p} e_t^T e_t.
\]

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

\[
\min_{\delta, \omega} L(\delta, \omega),
\]

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

\[
\dot{\delta} = -\frac{\partial L}{\partial \delta}, \quad \dot{\omega} = -\frac{\partial L}{\partial \omega},
\]

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

To find a maximum likelihood pseudo-orbit one could solve

\[
\min_{\delta, \omega} L(\delta, \omega) + a(\delta^T \delta/\sigma^2 + \omega^T \omega/\zeta^2),
\]

for \( a \to 0 \), or iteratively solve the above while alternating \( a = 0 \) and \( a > 0 \) using the solution at each step as the initial value for the next step.

This algorithm differs from the perfect model case by introducing perturbations \( \omega_t \) that are intended to allow for the imperfection of the model. It is clear that there is a confounding of the perturbations \( \delta_t \) and \( \omega_t \), that is, it is not possible to determine whether a model prediction is incorrect as a result observation error \( \delta_t \) or model error \( \omega_t \) and thus it is not clear how the total error should be distributed between the two sources of error\(^{10}\). One of the consequences of this is that 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 \( \omega_t \) terms is equivalent to assuming that the model is perfect; this is ill-advised.

\(^{10}\) Of course, this problem might vanish if the model admits \( \iota \)-shadowing trajectories over the entire duration of the available data, then one might arguably be within the perfect model scenario. We know of no dynamic physical systems where this has been shown to be the case [22].
that $L(\delta, \omega)$ attains a minimum of zero. If a minimum of $L$ occurs at $(\hat{\delta}, \hat{\omega})$, then it follows that $L(\hat{\delta}, 0) = (a/\epsilon^2) \sum_i \hat{\omega}_i^2$, that is, the amount by which the $s_i - \hat{s}_i$ fail to be a trajectory of the model is precisely $\hat{\omega}_i$. This would seem to imply that one could just minimize $L(\delta, 0)$ ignoring the $\omega$ terms; this, however, is equivalent to assuming the model is perfect. This approach gives biased estimates of the projection of the system state, because it forces the solution to give a trajectory of the model whenever one can be found\(^{11}\), but since the model is imperfect one should look for a solution in the broader class of pseudo-orbits. By including the terms $\omega_i$ in $L$, the gradient descent minimization of $L$ arrives at a near-by 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 $\omega$ terms can be wildly "inaccurate”.

In order to better understand the response of the proposed algorithm to the problem of confounded imperfection and observation error, it is useful to consider what happens when a noise-free observation of the true pseudo-orbit is presented to the algorithm. Ideally such a pseudo-orbit is unaltered, but it is more likely that the algorithm will find a different and "more probable” pseudo-orbit by attributing some of the deviation from a trajectory to be observational error; typically the larger imperfection errors will be altered more. We will shortly see that 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, we cannot rule out the possibility that other state estimation schemes may have better performance than the algorithm described here.

Figures 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 in the perfect model scenario that when the gradient descent algorithm was applied to noisy observations of a system trajectory to find maximum likelihood estimates of the state, then the estimates were found to lie in the indistinguishable set $H_\rho$ of the true state. Also, it was found that these estimates were distributed more or less as $Q_\rho$. In figures 5(a) and 6(a) the significant bias under the perfect model assumption is clearly seen. When presented with the noise-free true pseudo-orbit, the perfect-model gradient descent algorithm converges to a near-by trajectory. The noisy observations converge to states in the indistinguishable set $H_\rho$ (that is, in the perfect model sense) of the noise-free solution. This is exactly as should be expected, that is, the estimates lie in the indistinguishable set of the erroneous "true state”.

It is important to note that there are cases in figures 5(a) and 6(a) where the maximum likelihood estimates of the state are far from the relevant projection of the true state. The most important consequence of this, as we explain in the next section, is that ensembles constructed around these state estimates, under the perfect model assumption, have zero probability of containing the projection of the true state.

Figures 5(b) and 6(b) 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 figures 5(b) and 6(b) are similar to the densities of indistinguishable states of figure 4, they are sometimes displaced from their correct locations. Comparing panels (a) and (b) of figures 5 and 6 shows the bias under 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, because 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\(^{12}\), other methods may be able to obtain better results than those obtained here.

---

\(^{11}\)A similar approach, called 4DVAR, is used in operational weather forecasting; see [15] and references therein.

\(^{12}\)There is a fundamental question of identifiability, perhaps even definition, which we will not pursue further here.
Figure 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.
Figure 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.
Forecasting with an imperfect model is a dubious endeavor, but all real world forecasts are done this way. We present evidence below that suggests forecasts using an imperfect model under the assumption it is perfect almost certainly suboptimal. There are many issues regarding forecasting with imperfect models that should be addressed, but we postpone discussion of many of these issues to a subsequent paper. For the present we will assume sufficient wisdom to recognize a model is imperfect, and sufficient knowledge to at least guess a distribution \( \eta \) for the imperfection errors that provides some form of an upper bound on the actual imperfection errors. Specifically, we attempt not to systematically under estimate imperfection errors. Given these assumptions cautious ensemble forecasting with an imperfect model may be possible as described below.

Clearly, forecasting with an imperfect model cannot be a better situation than forecasting with a perfect model, where, as a consequence of indistinguishable states, forecasting requires creating an ensemble of model states which in turn provide a probabilistic forecast. In the perfect model scenario we suggested [10] first obtaining a maximum likelihood state estimate \( \hat{x} \), then selecting an ensemble \( \mathcal{E}(\hat{x}) \) of states from the set of indistinguishable states \( H_\rho(\hat{x}) \). Within PMS a symmetry exists in that if \( x \) 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 \( \mathcal{E}(\hat{x}) \) 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 \( x \). Note that statements about the probability of the true state \( x \) being in an ensemble \( \mathcal{E}(\hat{x}) \) only make sense when the indistinguishable states form a one-dimensional set. In general, a desirable property of an ensemble is that it contain the true state is the following sense.

**Definition 1** An ensemble \( \mathcal{E} \) is said to contain the state \( x \), if \( x \) lies within the convex hull (or bounding box) of the ensemble \( \mathcal{E} \).

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 figures 5 and 6 how this perfect model assumption leads to an increased bias in the estimate of the maximum likelihood state \( \hat{x} \); this is expected to result in a predictive disadvantage. Recall also that the indistinguishable states \( H_\rho(\hat{x}) \) will lie along the unstable set of \( \hat{x} \), that is, they will look something like those shown in figure 3. It should be clear that constructing \( H_\rho(\hat{x}) \) for any of the states \( \hat{x} \) in figures 5(a) and 6(a) will not overlap the projection of the true state \( x \), this is particularly obvious in cases where \( \hat{x} \) lies off of the attractor of the system. This is, in fact, just a resurfacing of the problem of asymptotic consistency. The consequence of this is that an ensemble of states \( \mathcal{E}(\hat{x}) \) 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; the forecasts are at the very best unrepresentative, at worst misleading.

If the imperfections of a model are taken into account, then useful ensembles can be obtained. In the imperfect model scenario the same approach of finding a maximum likelihood estimate \( \hat{x} \) of the projection of the system state \( x \) and an ensemble of states \( \mathcal{E}(\hat{x}) \subset H_{\rho \eta}(\hat{x}) \) can be applied.

Referring to figures 5 and 6 again, we observe that although the two imperfect models had a less than 2% imperfection error, if this is not taken into account, as in panel (a), then one may obtain wildly inaccurate maximum likelihood estimates of the projection of the system state, even with noise free observations. 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 variance, whereas selecting from \( H_\rho(\hat{x}) \) does not. Recall that figure 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 figures 5(b) and 6(b), then there is a non-zero probability that they contain the projection of the true state \( x \), because when compared to 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.
Sometimes confidence that an ensemble contains the projection of the true state is not enough, and 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 that aims to bound the verification.

Constructing a weighted ensemble estimate of the projection of the system state using an imperfect model can proceed along the same lines as that used when the model is perfect. With a perfect model a weighted ensemble estimate of the state can be constructed by first finding the maximum likelihood estimate of the state \( \hat{x} \), then selecting members for the ensemble by selecting elements of the indistinguishable set of the maximum likelihood state \( H_p(\hat{x}) \). The weighting of a member \( y \) of the ensemble should be \( Q_p(y|\hat{x}) \); in practice, a finite time approximation \( \tilde{Q}_T \) 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. The weighting of an ensemble member \( y \) should be \( Q_{p,T}(y|\hat{x}) \) or its finite time approximation. An alternative to weighted ensembles is to select states from \( H_{p,T}(\hat{x}) \) according to \( Q_{p,T}(y|\hat{x}) \). This would require using a MCMC approach as mentioned in section 3.4.

The methods of constructing an ensemble in the imperfect model scenario just described look reasonable at first sight, but there is a crucial problem that the constructions imply knowing the imperfection error distribution \( \eta \), 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 \( \eta \). If one could guess bounds for the errors, or their standard deviation, then an appropriate uniform on a disk, or Gaussian, density might supply a guess for \( \eta \). 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 \( \delta_t \) and \( \omega_t \), which are estimates of the observational error and imperfection error respectively. For example, figure 6(b) involved gradient descent calculations for trajectory segments of 16 points. For each of the maximum likelihood pseudo-orbits obtained we estimated standard deviations of the observational error and imperfection error from the obtained sixteen \( \delta_t \) and \( \omega_t \). We found that for all the 180 pseudo-orbits calculated for figure 6(b) the estimated standard deviation of the observation error 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, which are in agreement with the actual observational error of 0.1 and the actual imperfection error of around 0.02. Only the imperfection error estimate is needed for ensemble creation, but the accuracy of both estimates give confidence in using the imperfection errors as information to obtain a guess of the imperfection error distribution \( \eta \). Further analysis of this point is underway.

5.2 Forecasting

Finally, we can consider making ensemble forecasts with an imperfect model, noting that this is no longer a 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 equation (20), but it is impossible to represent this distribution analytically in practice. Given a weighed ensemble \( \mathcal{E}_0 \) at \( t = 0 \), one could generate a forecast weighted-ensemble \( \mathcal{E}_1 \), representing the forecast \( t \) steps into the future. One could construct \( \mathcal{E}_{t+1} \) by generating for each \( z_t \in \mathcal{E}_t \) 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 \( \mathcal{E}_{t+1} \) may be necessary to prevent the ensemble growing too large, however, pruning in the imperfect model scenario requires consideration, as it is not clear whether forecasts which would be considered “unlikely” in the perfect model scenario should be thinned or encouraged. One might want to consider variants of the particle filter or SIR filter [19, 4].

Without a perfect model and a perfect ensemble, an ensemble forecast will not be accountable[20], 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
6 Conclusions

In this paper 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, provided a distribution or bound can be found to describe the imperfection errors. It is not, of course, our aim to solve the two particular examples of model error studies above (where in we knew a perfect model and merely proceeded as if we did not), rather we aim to develop tools which can be applied when the perfect model (if such a thing exists) is not known. It is, of course, impossible to prove anything about the relation of the system to the model in this scenario; nevertheless this is the case we are in when working with data from any physical system.

An imperfect model treated as a perfect model yields an inaccurate estimate of the projection of the system state and incoherent ensemble results. Consequently, in any situation of state estimation or prediction of nonlinear systems it is essential to take model imperfections into account, failure to do so will result in degraded forecasts.

There are a many issues still to be resolved. We have not taken into account the fact that often one cannot observe all the dynamical variables of a model, that is, one is only a partial observer. A more detailed analysis of requirements and procedures for generating ensembles is needed, such as, more efficient and exact methods of generating ensembles or samples using Monte Carlo Markov Chain methods. Similarly, a more detail analysis of the behaviour of ensembles when they are evolved forward to provide forecasts, for example, questions of a suitable size of the ensemble need to be addressed, that is, as an ensemble is evolved the probability that it contains the projection of the true state decreases, which can be offset by making the initial ensemble larger. Most importantly the question of when it is appropriate to use an imperfect model for forecasting a system needs to be addressed in detail. We will address these questions in a subsequent paper, both in the simplified context presented here and in the case of operational forecasting models [14].

Further work also in progress is a study of parameterized model classes. There are two rather distinct situations here, the perfect model class, where there are parameter values that realize a perfect model, and imperfect model classes, which contain no parameter for any parameter set. Work in progress addresses the question of simultaneously estimating the state and the parameter values of the model, here 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 [15]; we are also investigating methods of gradient-free descent.

In concluding, 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 $\nu$-shadow the observations [25, 22] over the forecast period of interest, then

---

13By issues of metric dependence, if nothing else.
in the initial state limits accurate best 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 in understanding the underlying physical nonlinear system?

References


