Pseudo-Orbit Data Assimilation. Part II: Assimilation with Imperfect Models

HAILIANG DU AND LEONARD A. SMITH

Centre for the Analysis of Time Series, London School of Economics and Political Science, London, United Kingdom

(Manuscript received 31 January 2013, in final form 4 October 2013)

ABSTRACT

Data assimilation and state estimation for nonlinear models is a challenging task mathematically. Performing this task in real time, as in operational weather forecasting, is even more challenging as the models are imperfect: the mathematical system that generated the observations (if such a thing exists) is not a member of the available model class (i.e., the set of mathematical structures admitted as potential models). To the extent that traditional approaches address structural model error at all, most fail to produce consistent treatments. This results in questionable estimates both of the model state and of its uncertainty. A promising alternative approach is proposed to produce more consistent estimates of the model state and to estimate the (state dependent) model error simultaneously. This alternative consists of pseudo-orbit data assimilation with a stopping criterion. It is argued to be more efficient and more coherent than one alternative variational approach [a version of weak-constraint four-dimensional variational data assimilation (4DVAR)]. Results that demonstrate the pseudo-orbit data assimilation approach can also outperform an ensemble Kalman filter approach are presented. Both comparisons are made in the context of the 18-dimensional Lorenz96 flow and the two-dimensional Ikeda map. Many challenges remain outside the perfect model scenario, both in defining the goals of data assimilation and in achieving high-quality state estimation. The pseudo-orbit data assimilation approach provides a new tool for approaching this open problem.

1. Introduction

Weather forecast models are useful when predicting the weather, and Newton's laws are useful when predicting the motion of (most) planets, but in neither case are the underlying mathematical models perfect. Indeed, there is no scientific reason to believe that a perfect model exists. Generally, the model class from which the particular model equations are drawn does not contain a process that is able to generate[1] the data. This paper focuses on the extension of data assimilation outside the perfect model scenario (PMS) to the situation where the model is structurally imperfect. In this case, not only the observational uncertainty but also model inadequacy (Kennedy and O'Hagan 2001; Smith 2002) needs to be considered when an ensemble of initial conditions is constructed. Assuming the model is perfect is unlikely to produce the optimal results. In a chaotic system it is almost certain that no trajectory of the model is consistent with an infinite series of observations (Judd and Smith 2004), and there appears to be no consistent way to estimate the model states using trajectories since the model's invariant measure is almost certainly a poor prior for the "true" state.[2] There are pseudo orbits that are consistent with observations, however, and these can be used to estimate the model state (Judd and Smith 2004). This paper considers the pseudo-orbit data assimilation (PDA) approach discussed in Du and Smith (2014, hereafter Part I), adding a new stopping criterion to find relevant pseudo orbits outside PMS. The proposed approach is argued to be better suited for the condition encountered in operational state estimation than one version of the variational approach—specifically, the weak-constraint four-dimensional variational assimilation (WC4DVAR) (Miller et al. 1994; Zupanski 1997). Approaches adapting the PDA results to form an ensemble of initial conditions are introduced. By testing the state estimation performance both in the low-dimensional Ikeda system–model pair and in the higher-dimensional Lorenz96 system–model

[1] Produce a trajectory that, given the noise model, is consistent with the observations.

[2] Note even this statement assumes the model states and the true state share the same state space.

Corresponding author address: Hailiang Du, Centre for the Analysis of Time Series, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, United Kingdom.
E-mail: h.l.du@lse.ac.uk

DOI: 10.1175/JAS-D-13-033.1

pair, the PDA approach is demonstrated to be capable of outperforming an ensemble Kalman filter (EnKF) approach (Anderson 2001, 2003) as well.

In section 2, data assimilation outside PMS is defined and alternative approaches are reviewed. The purpose of using a pseudo orbit to account for model inadequacy is explained in section 3. The methodology of PDA with a stopping criterion is presented in section 4. In section 5, differences between the WC4DVAR approach and the PDA approach are discussed. Comparisons between EnKF and PDA for both the two-dimensional Ikeda system–model pair and the 18-dimensional Lorenz96 system–model pair are made in section 6. Section 7 provides a brief summary and conclusions.

2. Imperfect model scenario

Outside pure mathematics, the perfect model scenario is fiction. Arguably, there is no perfect model for any physical dynamical system (Smith 2002; Judd and Smith 2004). In the imperfect model scenario (outside PMS), one may hypothesize a nonlinear system with state space $\mathbb{R}^{\tilde{m}}$; the evolution operator of the system is $\tilde{F}$ [i.e., $\tilde{\mathbf{x}}_{t+1} = \tilde{F}(\tilde{\mathbf{x}}_t)$ where $\tilde{\mathbf{x}}_t \in \mathbb{R}^{\tilde{m}}$ is the state of the system]. The terms $\tilde{F}$, $\tilde{\mathbf{x}}$, and $\tilde{m}$ are unknown. It is often useful to speak as if such a system existed, regardless of whether one actually does exist. What is in hand is a model that approximates the system, with the form $\mathbf{y}_{t+1} = F(\mathbf{y}_t)$, where $\mathbf{y}_t \in \mathbb{R}^m$ ($\mathbb{R}^m$ is the model state space). In reality, it is almost certain that the system state space is different from the model state space. Assume that $\tilde{\mathbf{x}}$ can somehow be projected into the model state space by a projection operator $g(\cdot)$ [i.e., $\mathbf{x} = g(\tilde{\mathbf{x}})$, where $\mathbf{x} \in \mathbb{R}^m$]. In general, the property of this projection operator is unknown and one might question whether $\tilde{\mathbf{x}}$ exists.[3] It is simply assumed that $g(\cdot)$ maps the states of the system into somehow relevant states in the model state space. This operator will be discussed explicitly in each experiment below. A better understanding of $g(\cdot)$ is an important consideration for additional work that lies beyond the scope of this paper. An observation $\mathbf{s}_t$ at time t is defined by $\mathbf{s}_t = h[g(\tilde{\mathbf{x}}_t)] + \boldsymbol{\eta}_t$, where $\mathbf{s}_t \in \mathbb{O}$ and $\boldsymbol{\eta}_t$ represents the observational noise, taken here to be independent and identically distributed (IID) Gaussian, $N(0, \boldsymbol{\sigma}^2)$, for simplicity;[4] $h(\cdot)$ is the observation operator, which projects the model state into the observation space $\mathbb{O}$. For simplicity, $h(\cdot)$ is taken to be the identity operator below. Full observations are made; that is, observations are available for all state variables at every observation time.[5] The goal is to estimate the current state of the model $\mathbf{y}_0$ given the previous and current observations $\mathbf{s}_t$, $t = -n + 1, \ldots, 0$. Results are verified using the projection of the target system state [i.e., $\mathbf{x}_0 = g(\tilde{\mathbf{x}}_0)$].

A review of existing state-estimation approaches (both sequential approaches and variational approaches) can be found in Part I. Outside PMS, adjustments have to be made to account for model inadequacy. For variational approaches, a weak constraint (Miller et al. 1994; Zupanski 1997) is often applied to replace the strong constraint (Courtier et al. 1994; Bennett et al. 1996). For sequential approaches, several methods have been proposed to account for model inadequacy. These include the following: (i) Add stochastic terms in the (deterministic) model equations to alter the nature of model inadequacy by improving the model class (e.g., Buizza et al. 1999; Penland 2003; Leeuwen 2010; Mitchell and Gottwald 2012); one must still deal with inadequacy in the new class, of course. (ii) Add noise to each ensemble member so as to increase the ensemble variance appropriate to model inadequacy (e.g., Mitchell and Houtekamer 2000; Hamill and Whitaker 2005). (iii) Inflate the distance of each ensemble member about their mean (Anderson and Anderson 1999; Hamill et al. 2001). Adopting the weak constraint in variational

[3] It is common to think of the state of a physical system as a vector (or perhaps a continuous field) of real numbers. The simplest interpretation of atmospheric dynamics that would admit $\tilde{\mathbf{x}}$ requires that the continuum hypothesis holds; this is inconsistent with our best knowledge of the physics of fluids. Each of the four steps—the step from reality to partial differential equations, the step from partial differential equations to ordinary differential equations, the step from ordinary differential equations to finite difference equations, and the final step to difference equations on a finite digital grid—is treacherous. Even if these are overcome, 1) the failure of the continuum hypothesis would require $\tilde{\mathbf{x}}$ to be something other than the state of a partial differential equation and 2) even the claim that partial differential equations as simple as the Navier–Stokes equations admit smooth, physically reasonable solutions remains unproven. The difficulty of point 2 may be gauged by the fact that is to a millennium prize problem in modern mathematics (Fefferman 2000). The point of this footnote is to stress that the simple assumption that the state of the atmosphere is mathematically well defined and merely ''uncertain'' is as poorly founded as it is common. Arguably, there simply is no such thing.

[4] Although, the approach introduced in this paper can be applied to deal with non-Gaussian and even noise models that admit nonindependent and nonidentically distributed noise.

[5] As noted in Part I, various generalization to partial observations can be made (Judd et al. 2008; Du 2009; Smith et al. 2010) and the approach could be applied in operational weather forecasting following the approach of Judd et al. (2008). The case of partial observations will be considered elsewhere. In short, a two-pass approach to PDA is taken: the first using background information of the unobserved state variables with the observations frozen, and the second a standard application of the PDA approach described in this paper. While interesting, it is omitted here. Note there is some loss of generality in assuming full observations.

approaches and using method (i) in sequential approaches require prognostic terms in the cost function or model equations. Option (ii) is costly and inefficient, especially for high-dimensional models. Option (iii) aims to account for model inadequacy by adjusting the second moment of the ensemble distribution within the subspace spanned by the ensemble. The PDA approach presented in this paper estimates state-dependent model error and state variables simultaneously without requiring either the prior specification of prognostic terms or a second-moment closure.

Following Judd and Smith (2004), two types of model inadequacy are distinguished. One is structural inadequacy; the other is ignored-subspace inadequacy. For each type of model inadequacy, a system–model pair is designed in order to compare alternative approaches. For structural model inadequacy, where the system dynamics are not known in detail and its mathematical structure (assuming such a thing exists) is different from that of the model, the Ikeda map (Ikeda 1979; Haramel et al. 1985) is treated as the system and a truncated Ikeda map as the model. (Details of all systems, models, and experiments are given in the appendixes.) In this case, the model state and the system state share the same state space and $g(\cdot)$ is the identity naturally. For ignored-subspace model inadequacy some component(s) of the system dynamics is (are) unknown, unobservable, or simply omitted from the model; the Lorenz96 model II flow with both fast and slow variables is treated as the system, while the one layer model I flow excluding the fast variables is taken as the model. (Details of these flows are given in appendix A.) In this system–model pair, the model state space and the system state space differ; $g(\cdot)$ projects the system state into a subspace of the system state space; here, $g(\cdot)$ is a many-to-one projection. In the real atmosphere, of course, many different states of the atmosphere must map into identical model states.

3. Accounting for model inadequacy requires pseudo orbits

When $g(\cdot)$ is one to one, define the pointwise model error to be $g(\tilde{\mathbf{x}}_{t+1}) - F[g(\tilde{\mathbf{x}}_t)]$.[6] While it is sometimes reasonable to assume the observational noise is IID, it is almost certain that the pointwise model error of a nonlinear model varies coherently with $\mathbf{x}$ and will not be well mimicked by any IID process [see Orrell et al. (2001) for relevant evidence in numerical weather prediction]. Figure 1 illustrates how the pointwise model error for the Ikeda system–model pair is spatially correlated:

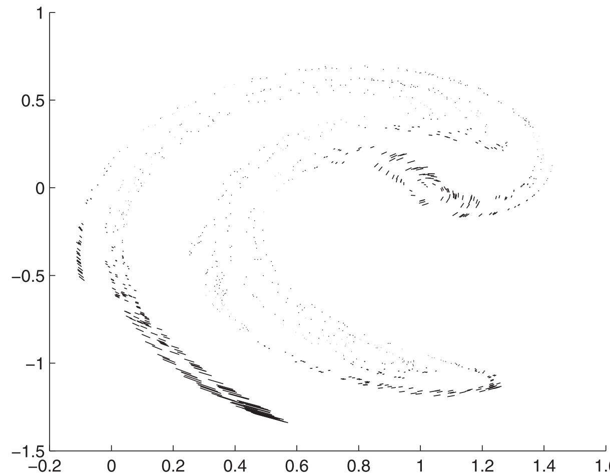

FIG. 1. The pointwise model errors for the truncated Ikeda map. The lines show the pointwise model error for 512 starting points by linking the prediction to the target. This figure provides a reference for Figs. 3 and 4.

there are regions where the pointwise model error is small and regions where it is not. Better understanding the distribution of the pointwise model error could aid in model development. If systematic pointwise model errors are more or less well identified, one may be able to improve the model by correcting some of the errors [for examples in numerical weather prediction with the Navy Operational Global Atmospheric Prediction System (NOGAPS), see Judd et al. (2008)]. While the focus of this paper is on how to better estimate a representative state within the model state space, accounting for model inadequacy is an unavoidable task within the procedure; the proposed approach provides information regarding pointwise model error, which may be of use in improving both the forecast and the model. It is stressed that state-dependent model error information is an output of the proposed approach, whereas both EnKF and WC4DVAR require specifications and/or assumptions as an input. If indeed a viable statistical description of state-dependent model error was available, it could be used to broaden the model class and improve dynamical simulation.

To estimate "the" current state of the model outside PMS, one needs to account for both observational noise and model inadequacy. In the absence of observational noise, the pointwise model error could be derived from the observations directly. In the presence of observational noise, the convolution compounding pointwise model error and observational noise removes the possibility of identifying either precisely.

Recall from Part I that a pseudo orbit, $\mathbf{U} \equiv \{\mathbf{u}_{-n+1}, \ldots, \mathbf{u}_{-1}, \mathbf{u}_0\}$, to is a point in the $m \times n$ dimensional sequence space for which $\mathbf{u}_{t+1} \neq F(\mathbf{u}_t)$ for any component of $\mathbf{U}$. This implies that $\mathbf{U}$ corresponds to a sequence of model

[6] If $g(\cdot)$ is not one to one, the definition must be modified to consider the expectation over all states $\tilde{\mathbf{x}}$ that map into $\mathbf{x}$ via $g(\tilde{\mathbf{x}})$.

states that is not a trajectory of the model. Let the elements of $\mathbf{U}$ corresponding to the model state at a given time be called a "component" of the pseudo orbit. Define the imperfection error of each of the $n - 1$ components of $\mathbf{U}$ to be $\boldsymbol{\omega}_t = \mathbf{u}_{t+1} - F(\mathbf{u}_t)$ for $t = -n + 1, \ldots, -1$. Note the imperfection error will not correspond to the pointwise model error. It is the case, however, that the projection of a system trajectory in the model state space is a pseudo orbit of the model, and in this case the imperfection error does reflect the pointwise model error in the model state space. Arguably,[7] no model trajectory is consistent[8] with an infinite sequence of observations outside PMS. There are pseudo orbits, however, that are consistent with observations and these can be used to provide useful estimates of the projection of the system state into the model state space. The system trajectories projected into the model state space are pseudo orbits of the model in the model state space; these target pseudo orbits, $\{g(\tilde{\mathbf{x}}_{-n+1}), \ldots, g(\tilde{\mathbf{x}}_0)\}$, are both consistent with the observations and their imperfection errors reflect the pointwise model error precisely. Unfortunately, such desirable pseudo orbits cannot be determined precisely outside PMS owing to the confounding of observational noise and the pointwise model error mentioned above. It may still be possible to identify informative pseudo orbits of the model that are consistent with observational noise, however, and the imperfection error of those pseudo orbits may provide information regarding pointwise model error. PDA-based approaches for finding relevant pseudo orbits are introduced in the following section.

4. PDA with a stopping criterion

In Part I, Du and Smith applied the PDA approach introduced by Judd and Smith (2001) for state estimation in PMS by minimizing the mismatch cost function given by

$$C(\mathbf{U}) = \sum |\mathbf{u}_{t+1} - F(\mathbf{u}_t)|^2. \tag{1}$$

PDA minimizes the mismatch cost function in an $m \times n$ dimensional sequence space using a gradient descent (GD) algorithm. In practice, the minimization is initialized with the observation-based pseudo orbit (i.e., ${}^0\mathbf{U} = \{\mathbf{s}_{-n+1}, \ldots, \mathbf{s}_0\}$). The pseudo-orbit is updated on every iteration of the GD minimization. Let the result of the GD minimization be ${}^\alpha\mathbf{U}$, where α indicates algorithmic time in GD (see Part I for additional discussion).

In a misstep, Judd and Smith (2004) adjusted the approach by adding the imperfection error term in the mismatch cost function to account for model inadequacy (to be clear, this approach is not recommended). Specifically, Judd and Smith (2004) adjusted cost function to be

$$C^*(\mathbf{U}, \boldsymbol{\omega}) = \sum |\mathbf{u}_{t+1} - \boldsymbol{\omega}_{t+1} - F(\mathbf{u}_t)|^2. \tag{2}$$

By minimizing $C^*(\mathbf{U}, \boldsymbol{\omega})$, one obtains a pseudo orbit $\mathbf{u}_t$ and the corresponding imperfection error $\boldsymbol{\omega}_t$. Du (2009) shows that the results of such an approach are inconsistent[9] with the observational noise and the pointwise model error; minimizing $C^*(\mathbf{U}, \boldsymbol{\omega})$ to obtain pseudo orbits is not recommended.

The alternative presented here is to minimize $C(\mathbf{U})$ with a stopping criterion, thereby obtaining more consistent (less biased) pseudo orbits. Notice that minimization of $C(\mathbf{U})$ is actually minimizing the imperfection error of the pseudo orbit $\mathbf{u}$. The imperfection error is treated as an estimate of the pointwise model error, which is known to exist when the model is imperfect. Note the aim is not minimizing the imperfection error but producing better (more consistent) estimates of the model states and the corresponding pointwise model error. Let $\mathbf{s}_t - h(\mathbf{u}_t)$ be the implied noise. It is desirable to obtain pseudo orbits whose implied noise and imperfection error are consistent with the observational noise and the pointwise model error, respectively. Inasmuch as all available information on model inadequacy will have been included in refining the model, information is only available regarding the observational noise, and that information is statistical. One obvious goal is to match the statistics of the implied noise with that of the observational noise, so that the implied noise is consistent with the noise model. Figure 2 shows that the statistics of implied noise, imperfection error, and pseudo orbits change as the minimization runs deeper and deeper (α increases). Both the higher-dimensional Lorenz96 system–model pair experiment (Fig. 2, left) and the low-dimensional Ikeda system–model pair experiment (Fig. 2, right) are shown.

As the GD minimization advances (as α increases), the standard deviation of the implied noise tends to increase beyond that of the observational noise; this

[7] This is expected when the model is chaotic (Judd and Smith 2004).

[8] Specifically, there is no trajectory $\{\mathbf{y}_0, \mathbf{y}_1, \mathbf{y}_2, \ldots\}$ such that the series $\mathbf{s}_t - \mathbf{y}_t$ is consistent with the noise model.

[9] The obtained pseudo orbits tend to have the implied noise (defined in the following paragraph) much smaller than the observational noise and the imperfection error much larger than the pointwise model error.

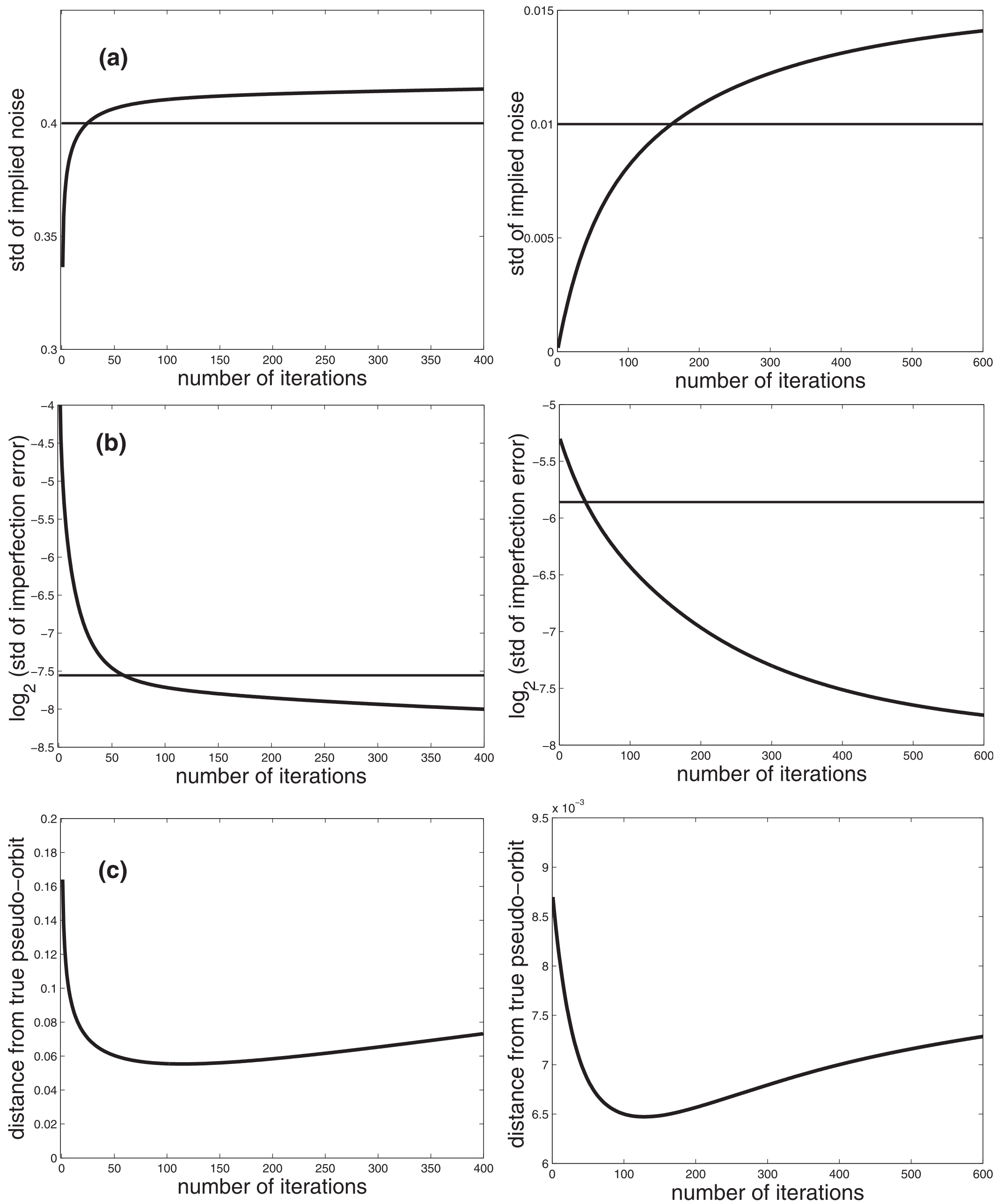

FIG. 2. Statistics of the pseudo orbit as a function of the number of gradient descent iterations for both the (left) higher-dimensional Lorenz96 system–model pair experiment and (right) low-dimensional Ikeda system–model pair experiment. (a) Standard deviation of the implied noise (the flat line is the standard deviation of the noise model); (b) standard deviation of the model imperfection error (the flat line is the sample standard deviation of the pointwise model error); and (c) RMS distance between the pseudo orbit and the target pseudo orbit.

indicates the tendency of pseudo orbits to eventually move far from the observations. Comparing the standard deviation of the implied noise with that of the actual noise model (flat line) in the first panels of Fig. 2 reveals that at the beginning of the GD minimization, the observational noise has a larger standard deviation than the implied noise does since the pseudo orbit is close to the observations; this simply reflects the fact that the minimization algorithm is initialized at the observations. As the minimization proceeds, the standard deviation of the implied noise grows (approaching that of the observational noise), and the pseudo orbit gets closer to the target pseudo orbit, as shown in Fig. 2c. At some point, however, the standard deviation of the

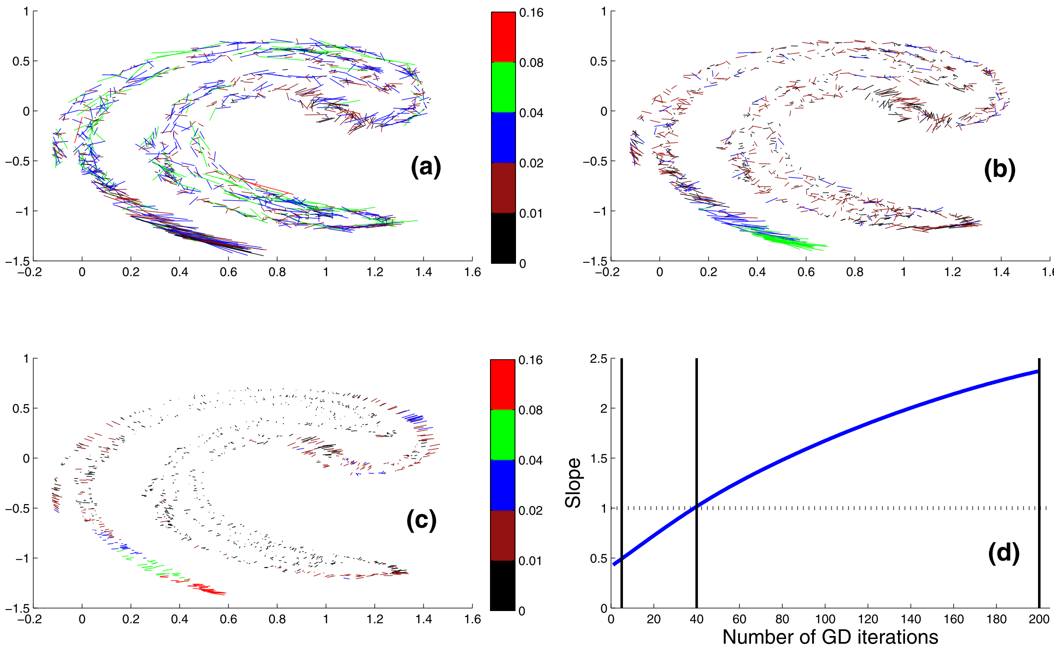

FIG. 3. Snapshots of the imperfection error (magnified by a factor of 3 in length) in state space during the gradient descent runs for the Ikeda map experiment with noise model $N(0, 0.01^2)$: after (a) 5, (b) 40, and (c) 200 GD iterations. The color reflects the difference between the imperfection error and the corresponding pointwise model error. (d) The evolution of the slope of the best-fit line relating the imperfection error and the corresponding pointwise model error as a function of the number of GD iterations. The vertical lines show the locations of $\alpha = 5, 40$, and 200; these values of α correspond to (a),(b), and (c), respectively.

implied noise exceeds that of the observational noise, and the distance between the pseudo orbit and the target pseudo orbit grows larger still. This results from model inadequacy: When the imperfection error of the pseudo orbit becomes smaller than the actual pointwise model error (see Fig. 2b), the implied noise compensates for the imperfection error to account for the effects of model inadequacy. This makes the implied noise distribution too wide, and the pseudo orbits become inconsistent with the observations. Thereafter, the minimization of $C(\mathbf{U})$ reduces the imperfection error without improving the pseudo orbit—indeed, while degrading it. This problem calls for some sort of stopping criterion.

In cases where the pointwise model error distribution is neither IID nor Gaussian, the extent to which the imperfection error mimics the pointwise model error is incompletely reflected by the second-moment statistics. The ability of PDA to cope with such conditions will be presented elsewhere. Figure 1 indicates the extent to which the pointwise model error is spatially correlated. As an estimate of the pointwise model error, the imperfection error is expected to have similar spatial correlations to those of the pointwise model error. Evidence that this expectation is fulfilled is provided in Fig. 3, which plots the imperfection error[10] in the state space of the Ikeda map for different numbers of GD iterations. Compare Fig. 3 with the actual pointwise model error plotted in Fig. 1. A quantitative comparison of model error and imperfection error is obtained from the slope l of the regression line of model error against imperfection error. Figure 3d shows l as a function of α, which is how the regression evolves as the GD algorithm advances. At the beginning of the minimization, the imperfection errors tend to be larger than the pointwise model error in most places; early in the minimization (α is small), the imperfection error contains both the observational noise and the pointwise model error. Similarly in the asymptotic regime (α is large), the imperfection error has evolved to be much too small and the slope of the regression line grows much larger than 1. This can be seen in Fig. 3d. The imperfection error loses the spatial information available at smaller α. Less spatial correlation of the imperfection error can be seen in Fig. 3c. With $\alpha \approx 40$ (Fig. 3b), the imperfection error provides a better estimate of the pointwise model error. Here, the

[10] Note that the starting points of the imperfection error are not the same in Figs 3a–c.

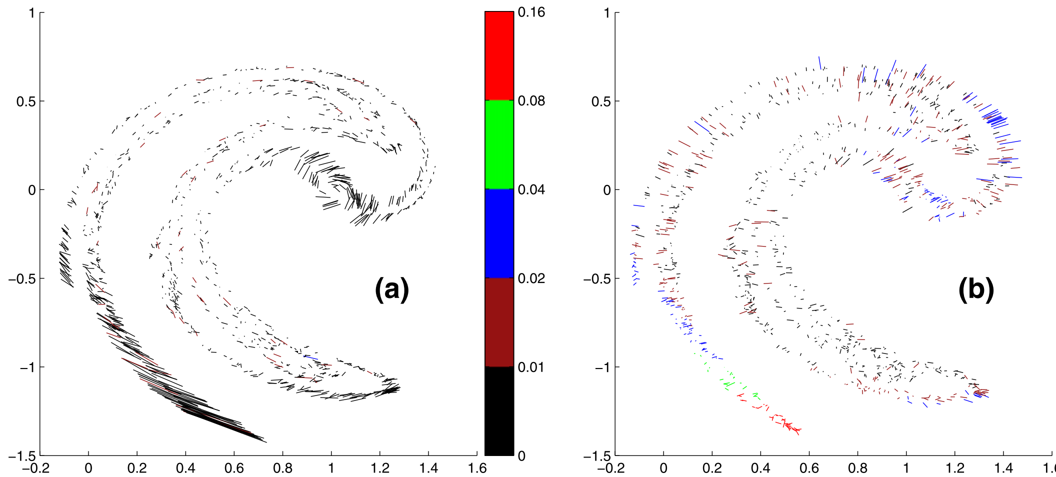

FIG. 4. Imperfection errors (magnified by a factor of 3 in length) after intermediate gradient descent runs for the Ikeda system–model pair are plotted in the state space. The color reflects the difference between the imperfection error and the corresponding pointwise model error. (a) $\sigma = 0.002$ and (b) $\sigma = 0.05$, where σ is the standard deviation of the Gaussian noise model.

patterns in Figs. 3b and 1 are very similar, and the differences between the imperfection error and the corresponding pointwise model error are relatively small.

The extent to which pointwise model error can be identified robustly depends strongly on the relative magnitudes of the observational noise and pointwise model error. Figure 4 plots the imperfection error in the state space at two different noise levels. (In each case, the GD stopped when the standard deviation of the implied noise first exceeded the standard deviation of the observational noise.) When the observational noise is much smaller than the pointwise model error, the latter can be well estimated by the imperfection error (as in Fig. 4a). When the observational noise is significantly bigger than the pointwise model error, the imperfection error appears more random (as in Fig. 4b).

These experiments suggest that PDA with a stopping criterion based on the statistics of the implied noise produces pseudo orbits closer to the target than those with significantly smaller α or significantly larger α. When should one stop the GD minimization in order to obtain the most relevant pseudo orbit? The answer to this question will vary with the definition of ''most relevant'' (or ''best'') pseudo orbit. If, for example, ''better'' means a pseudo orbit more consistent with the observations, then the stopping criterion can be based upon consistency between implied noise and the noise model;[11] alternatively, if better means a preferred candidate for an ensemble forecast (evaluated at a certain forecast lead time), then the stopping criterion can be determined based on past forecast performance.[12] Generally, the number of iterations α is a tuning parameter and need not be specified a priori. Denote PDA with a stopping criterion as PDA^c. In the experiments presented in this paper, the stopping criterion targeted state-estimation performance. That is, α was tuned to maximize the skill (in terms of ignorance defined below) of the state estimation at $t = 0$. Evaluation criteria that lead to the same result inside PMS (where a perfect ensemble is well defined) are expected to lead to different results outside PMS. While the simple criterion above is adequate for our purpose, it no doubt could be improved upon. The key take-home point here is that even this simple stopping criterion provides more consistent state-estimation results than the alternative approaches considered. Furthermore, outside PMS, PDA need not pursue a pseudo orbit all the way to its asymptotic approach to a model trajectory. This means the cost of obtaining useful pseudo orbits is reduced substantially, which in turn makes the approach more attractive for use in operational prediction models (e.g., Judd et al. 2008).

The PDA^c approach can also play a role in forming ensembles of initial conditions. To capture the uncertainty in the nowcast and forecast in numerical weather prediction, some approaches (Leutbecher and Palmer

[11] Of course, higher-resolution measurements (i.e., small noise observations) available over a limited duration might also be used as a target to optimize the stopping criterion for state estimation.

[12] Note that as the model is not perfect, one should expect the characteristic of the ideal initial condition ensemble to vary with lead time.

2008) sample the model's more rapidly growing directions at the ($t = 0$) analysis. Such sampling, however, is unlikely to produce states that are consistent with the long-term model dynamics (Smith 1996).

An alternative approach is to apply PDA^c to perturbed observations. To form an ensemble in this case, the observations are perturbed using the inverse of the observational noise distribution. More explicitly, given the observational noise model, one adds random draws from the inverse of the observational noise model to the observation. For Gaussian observational noise, the perturbed observations are statistically equivalent to increase the noise level by $\sqrt{2}$. PDA^c is then applied to the perturbed observations. Last, the final component of each pseudo orbit produced by the PDA^c approach is taken as an ensemble member. Each ensemble member from perturbed observation is treated as equally likely. This approach is used to generate the experiments results in the paper. The alternative approach of adding noise to the optimized pseudo orbit obtained from the observations is not considered here. Given that the model is acknowledged to be imperfect, it is not clear that these forecasts should be used as probability forecasts however they are formed.

5. Contrasting WC4DVAR with PDA^c

In the traditional 4DVAR approach (Lorenc 1986; Talagrand and Courtier 1987; Courtier et al. 1994), the model dynamics are interpreted as a strong constraint (Courtier et al. 1994; Bennett et al. 1996), which effectively assumes that the model is perfect (i.e., model trajectories consistent with the observations are targeted). Comparison between 4DVAR and PDA in PMS in Part I suggests that the 4DVAR approach suffers from the multiple local minima when applied to long windows (Miller et al. 1994; Pires et al. 1996). The PDA approach, on the other hand, can benefit from the additional dynamical information contained in a larger window. To account for model inadequacy outside PMS, one might apply the model as some form of weak constraint (Miller et al. 1994; Zupanski 1997) rather than as a strong constraint (Sasaki 1970). Bennett et al. (1993, 1996) show that some applications of the model dynamics as a weak constraint in a 4DVAR approach can outperform the approach using it as a strong constraint.

Although rarely phrased in this way, the WC4DVAR approach can be viewed either as a search for pseudo orbits of a deterministic model given a specified dynamical noise model or as a search for trajectories of a fully specified stochastic model. Following Lorenc (1986), the version of WC4DVAR (Zupanski 1997; Judd 2008) considered below is derived with the assumption of Gaussian IID observational noise[13] and Gaussian IID pointwise model error. While this may prove preferable to assuming that the deterministic model is perfect, it is long known that in numerical weather prediction the pointwise model error is not IID (e.g., Orrell et al. 2001). Making a model stochastic does not remove the basic challenges posed by model inadequacy, unless doing so makes the model perfect.

Following the maximum likelihood principle, the probability of $\mathbf{y}_{-n+1}, \ldots, \mathbf{y}_0$ given $\mathbf{s}_{-n+1}, \ldots, \mathbf{s}_0$, that is, $p(\mathbf{y}_{-n+1}, \ldots, \mathbf{y}_0 \,|\, \mathbf{s}_{-n+1}, \ldots, \mathbf{s}_0)$ is proportional to

$$\exp\left\{-\frac{1}{2}\sum_{t=-n+1}^{0}[h(\mathbf{y}_t)-\mathbf{s}_t]^{\mathrm{T}}\boldsymbol{\Gamma}_t^{-1}[h(\mathbf{y}_t)-\mathbf{s}_t]\right\} \times \exp\left\{-\frac{1}{2}\sum_{t=-n+2}^{0}[\mathbf{y}_t-F(\mathbf{y}_{t-1})]^{\mathrm{T}}\mathbf{Q}_t^{-1}[\mathbf{y}_t-F(\mathbf{y}_{t-1})]\right\}, \tag{3}$$

where $\boldsymbol{\Gamma}$ and $\mathbf{Q}$ are the observational error and pointwise model error covariance matrices, respectively. The WC4DVAR cost function is derived by taking the logarithm of the above equation; that is,

$$C_{\mathrm{wc}} = \frac{1}{2}\sum_{t=-n+1}^{0}[h(\mathbf{y}_t)-\mathbf{s}_t]^{\mathrm{T}}\boldsymbol{\Gamma}_t^{-1}[h(\mathbf{y}_t)-\mathbf{s}_t] + \frac{1}{2}\sum_{t=-n+2}^{0}[\mathbf{y}_t-F(\mathbf{y}_{t-1})]^{\mathrm{T}}\mathbf{Q}_t^{-1}[\mathbf{y}_t-F(\mathbf{y}_{t-1})]. \tag{4}$$

In practice, an additional background term may be used to take account of the information either from previous estimates or from any available prior distribution of the initial state. Note that although the expression of the first term in the cost function is the same as the term in the 4DVAR cost function, they are different in the sense that in the original 4DVAR case the estimate of the model states $\mathbf{y}_{-n+1}, \ldots, \mathbf{y}_0$ are states along a single trajectory of the model [i.e., $\mathbf{y}_t - F(\mathbf{y}_{t-1}) = 0$], while in the WC4DVAR case those states form a pseudo orbit [i.e., $\mathbf{y}_t - F(\mathbf{y}_{t-1}) \neq 0$]. It is also assumed that the difference between $\mathbf{y}_t$ and $F(\mathbf{y}_{t-1})$ is well described by an IID Gaussian process with $\mathbf{Q}$; $\mathbf{Q}$ must be specified a priori if Eq. (4) is to be evaluated. The difference between $\mathbf{y}_t$ and $F(\mathbf{y}_{t-1})$ is the imperfection error of the pseudo orbit $\mathbf{y}_{-n+1}, \ldots, \mathbf{y}_0$, which is expected to be minimized by the second term of the cost function. In practice, WC4DVAR is often implemented so as to obtain pseudo orbits of the model by maintaining the balance that such

[13] In practice, meteorological observations commonly contain systematic errors correlated in time.

a pseudo orbit stays close to the observations but with small imperfection errors; justification of the initial specification of $\mathbf{Q}$ in practice remains a challenge. A similar challenge has long plagued the specification of $\boldsymbol{\Gamma}$ as well.

There are some similarities between PDA^c and WC4DVAR: (i) both can be applied to an assimilation window to produce an estimate of model states (analysis) and (ii) the time series of analyses produced by both approaches is a pseudo orbit of the deterministic model, each with its own corresponding sequence of imperfection errors.

There are also fundamental differences between PDA^c and WC4DVAR. Despite there being multiple realization of the WC4DVAR approach, each of them requires some similar input assumptions—specifically, that the pointwise model error is IID Gaussian. The PDA^c approach does not require this assumption; indeed, this was stressed as a strength of the PDA approach as early as Judd and Smith (2004). The assumption that the pointwise model error is IID Gaussian is known to be unrealistic on both theoretical and empirical grounds—a fact reflected by the practice of dropping this term in the forecast model. Implementing the WC4DVAR approach forces the imperfection error toward this assumption. Arguably, this weak constraint is an improper constraint (also see Judd 2008).

In the PDA^c approach, however, no such assumption regarding the pointwise model error is ever made. From Fig. 1, it is obvious that the pointwise model error is not IID in the state space. Results in the previous section show that the PDA^c approach can, in practice, produce informative imperfection errors. For models recurrent in the model state space, the imperfection errors can be used in forecast mode [as in Smith (1992)]. And it has been shown that, as the imperfection errors in European Centre for Medium-Range Weather Forecasts (ECMWF) operational models (T42 and T63) vary slowly in time, they can be used to reduce the forecast RMS error (Orrell et al. 2001).

The WC4DVAR approach also assumes observational noise to be IID, while meteorological observations often have systematic observational errors (Lu and Browning 1998). The PDA^c approach requires no assumption of the observational noise model. In practice, the WC4DVAR approach [Eq. (4)] suffers from local minima as does 4DVAR. It is shown in Du (2009) that the performance of the WC4DVAR approach deteriorates as the length of the assimilation window increases; for PDA^c, this is not the case. Most strikingly, WC4DVAR requires rather large error covariance matrices to be specified as a priori where PDA^c does not. Indeed, PDA^c provides information on the pointwise model error.

More generally, the variational approaches see data assimilation as a statistical problem: one is looking for an optimal trajectory. PDA^c, on the other hand, sees data assimilation as more similar to a control problem: one is looking for the control perturbations required to keep a pseudo orbit of the model close to the observations.

6. Ensemble Kalman filter versus PDA^c

The sequential approach used here is the ensemble adjustment Kalman filter (Anderson 2001, 2003) with covariance inflation in order to account for model inadequacy. The comparison is made first in the lower-dimensional case in order to ease visualization of the evidence. Both PDA^c and EnKF are applied to the two-dimensional Ikeda model–system pair and the ensemble results in the state space are plotted. Four examples of the estimated states are shown in Fig. 5. Whether the state estimations lie on the system attractor may prove irrelevant outside PMS. Sampling from (near) the models attractor can again be more efficient than sampling the full volume in the model state space. In Fig. 5, the ensemble produced by the PDA^c approach is visibly relatively closer to the target state. In Fig. 5 (top), the EnKF ensemble manages to cover the target state, while in Fig. 5 (bottom), the EnKF ensemble members are far from the target state. Note that inflation increases the spread of the ensemble, but it does not change the subspace spanned by the ensemble (Hamill 2006). For example, in the bottom-right panel, the EnKF ensemble members are almost lying along the line parallel to the y axis; in such a case, inflating the ensemble will not move any of the members toward the target state in the x direction.

To measure the difference between these two approaches quantitatively, the initial condition ensemble is translated into a predictive distribution function by standard kernel dressing (Brocker and Smith 2008). Each ensemble member is replaced by a Gaussian kernel centered on that member; this makes a continuous distribution (a non-Gaussian sum of Gaussian kernels). The width of each kernel (the standard deviation of the Gaussian, called the ''kernel width'') is determined by optimizing the ignorance score in a training set. (The training set is then discarded; it is not used in the evaluation below.)

The performance of a state-estimation technique can be evaluated with the ''log p score'' or ignorance score (Good 1952; Roulston and Smith 2002). The ignorance score is the only proper local score for continuous variables (Bernardo 1979; Raftery et al. 2005; Brocker and Smith 2007). Although there are also nonlocal proper scores, the authors prefer using ignorance as 1) it has a clear interpretation in terms of information theory,

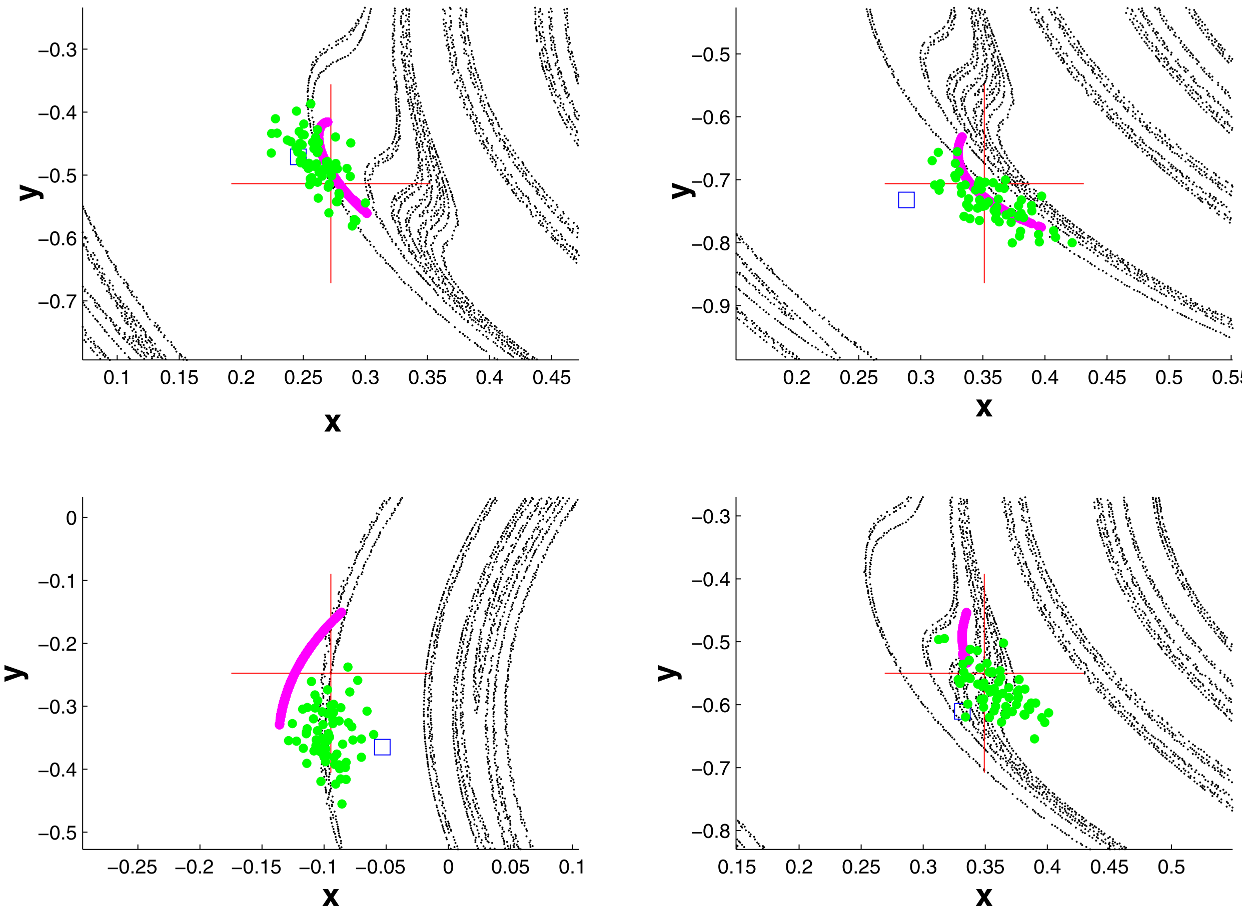

FIG. 5. Ensemble results from both EnKF and PDA^c for the Ikeda system–model pair. The target state is centered in each panel (large cross); the square is the corresponding observation, and the background dots indicate samples from the Ikeda map attractor. The EnKF ensemble is depicted by 512 magenta dots. The PDA^c ensemble is depicted by 512 green dots. Each panel is an example of state estimation.

2) it is local, and 3) it can be easily communicated in terms of effective interest returns (Good 1952; Roulston and Smith 2002; Hagedorn and Smith 2009). The ignorance score is defined by

$$S[p(y), Y] = -\log_2[p(Y)], \tag{5}$$

where Y is the outcome and $p(Y)$ is the probability of event Y. In practice, given K forecast–outcome pairs (p_t, Y_t, $t = 1, \ldots, K$), the empirical average ignorance skill score is

$$S_{\text{Emp}}[p(y), Y] = \frac{1}{K}\sum_{i=1}^{K} -\log_2[p_i(Y_i)]. \tag{6}$$

The PDA^c approach is compared with the EnKF approach in both the lower-dimensional Ikeda system–model pair and higher-dimensional Lorenz96 system–model pair. In both cases, the state-estimation performance is evaluated by empirical ignorance.

Table 1 shows the comparison between EnKF and PDA^c (details of the experiments are given in appendix B) using the ignorance score (the optimized kernel width is

TABLE 1. Ignorance score and kernel width of initial-condition ensemble for the Ikeda model–system pair and Lorenz96 model–system pair; the noise model is $N(0, 0.05^2)$ and $N(0, 0.5^2)$, respectively. The 512-member ensembles generated by PDA^c and EnKF are compared. Lower and upper are the 90% bootstrap resampling bounds of the ignorance score. The statistics are calculated based on 8192 assimilations and 4096 bootstrap samples are used to determine the resampling bounds.

	Ignorance		Lower		Upper		Kernel width	
Systems	EnKF	PDA^c	EnKF	PDA^c	EnKF	PDA^c	EnKF	PDA^c
Ikeda	−2.66	−3.67	−2.76	−3.70	−2.51	−3.64	0.056	0.0023
Lorenz96	−3.32	−4.08	−3.40	−4.12	−3.23	−4.04	0.45	0.26

also presented). From the table, it is clear that the ensemble generated by the PDA^c approach outperforms significantly the ensemble generated by EnKF in both experiments. Relative ignorance between the two approaches is found to be around 1 bit in the Ikeda experiment and 0.75 bits in the Lorenz96 experiment, which can be interpreted as the PDA^c approach placing, on average, 100% (and 68%) more probability density near the outcome than the EnKF approach. The much smaller kernel width for the PDA^c ensemble also indicates that the PDA^c ensemble members are more concentrated around the (projection of) target state than the EnKF ensemble.

7. Conclusions

In this paper, the problem of estimating the current state(s) of a model outside PMS is addressed. In practice, the assumption of a perfect model is unrealistic as is the assumption that model inadequacy is IID in time. A PDA approach with a stopping criterion, PDA^c, allows state estimation without the assumption of a perfect model. The PDA^c approach is shown to produce pseudo orbits that are consistent with the observations and yield imperfection error as an output that reflects state-dependent model error in the examples considered. The differences between the WC4DVAR approach (Miller et al. 1994; Zupanski 1997) and the PDA^c approach are stressed, and the fact that WC4DVAR requires an additional assumption about the dynamics of model error is noted.

Comparisons between the PDA^c approach and the EnKF approach have been made both in the lower-dimensional Ikeda map and in the higher-dimensional Lorenz96 model. By looking at the ensemble results in the model state space and statistically evaluating the ensemble using ignorance, it is demonstrated that the proposed approach systematically outperforms the EnKF approach considered (Anderson 2001, 2003) in both cases. Additional comparisons on the same data with more sophisticated filters would be welcome.

The reasons that PDA^c outperforms EnKF and WC4DVAR are easily understood. Given the illustration by Judd et al. (2008) that PDA^c is deployable on large-scale models, its evaluation in a true operational context is hoped to contribute to the improvement of operational state estimation, data assimilation, and forecasting.

Acknowledgments. This research was supported by the LSE's Grantham Research Institute on Climate Change and the Environment and the ESRC Centre for Climate Change Economics and Policy, funded by the Economic and Social Research Council and Munich Re. L. A. Smith gratefully acknowledges support from Pembroke College, Oxford.

APPENDIX A

Details of System–Model Pairs

a. Ikeda system–model pair

The Ikeda map was introduced by Haramel et al. (1985) based on Ikeda's model of laser pulses in an optical cavity (Ikeda 1979). It is a common testbed in data assimilation studies (e.g., Hansen and Smith 2001; Judd and Smith 2004). With real variables, it has the form

$$X_{n+1} = \gamma + u(X_n \cos\phi - Y_n \sin\phi), \tag{A1}$$

$$Y_{n+1} = u(X_n \sin\phi + Y_n \cos\phi), \tag{A2}$$

where $\phi = \beta - \alpha/(1 + X_n^2 + Y_n^2)$. With the parameters $\alpha = 6, \beta = 0.4, \gamma = 1$, and $u = 0.83$, the system is believed to be chaotic. The imperfect Ikeda model is obtained by using the truncated polynomial to replace the trigonometric function in $\tilde{F}$; that is,

$$\cos\theta = \cos(\omega + \pi) \mapsto -\omega + \omega^3/6 - \omega^5/120, \tag{A3}$$

$$\sin\theta = \sin(\omega + \pi) \mapsto -1 + \omega^2/2 - \omega^4/24, \tag{A4}$$

where the change of variable to ω was suggested by Judd and Smith (2004) since θ has the approximate range from -1 to -5.5, and $-\pi$ is conveniently near the middle of this range. In this case, the model state and the system state share the same state space. Generally, the truncated Ikeda model is a good approximation to the Ikeda system.

b. Lorenz96 system–model pair

A system of nonlinear ordinary differential equations (Lorenz96 system) was introduced by Lorenz (1995) and called model II. It is a common testbed in data assimilation studies (e.g., Fertig et al. 2007; Leeuwen 2010). The variables involved in the system are analogous to some atmospheric variables regionally distributed around the Earth. The mathematical functions of the system are

$$\frac{dx_i}{dt} = -x_{i-2}x_{i-1} + x_{i-1}x_{i+1} - x_i + F - \frac{h_x c}{b}\sum_{j=1}^{n} y_{i,j}, \tag{A5}$$

$$\frac{dy_{j,i}}{dt} = cby_{j+1,i}(y_{j-1,i} - y_{j+2,i}) - cy_{j,i} + -\frac{h_y c}{b}x_i \tag{A6}$$

for $i = 1, \ldots, n$, and $j = 1, \ldots, m$. The system used in experiments presented in this paper contains $n = 18$ variables $x_1, \ldots, x_{18}$ with cyclic boundary conditions (where $x_{n+1} = x_1$). Like the large-scale variables x_i, the

TABLE B1. Details of the PDA implementation.

	Ikeda	Lorenz96
Observational noise	$N(0, 0.05^2)$	$N(0, 0.5^2)$
Window length	$16\Delta t$	$32\Delta t$
No. of GD iterations	40	60
GD step size	0.4	0.4
No. of ensemble members	512	512

small-scale variables $y_{j,i}$ have the cyclic boundary conditions as well (that is, $y_{m+j,i} = y_{j,i+1}$). In experiments considered in this paper, $m = 5$. The coefficients used are $b = c = F = 10$, for which the small-scale variables tend to fluctuate 10 times more rapidly but with 10 times smaller magnitude than the large-scale variables. For more information, see Lorenz (1995) and Orrell et al. (2001).

The Lorenz96 model I is

$$\frac{dx_i}{dt} = -x_{i-2}x_{i-1} + x_{i-1}x_{i+1} - x_i + F. \qquad \text{(A7)}$$

Small dynamical variables y in the system equations [Eqs. (A5) and (A6)] are not included in the Lorenz96 imperfect model. The magnitude of error made by the imperfect model depends on the coupling parameter h_x, h_y, and in experiments presented in this paper, both h_x and h_y are set to be 1. In this system and model pair setting, the model state space and the system state space are different.

APPENDIX B

Experiments' Details

Details of the experiments discussed in the paper are given here. Table B1 provides specific experimental details of the PDA implementation conducted in the paper. The ensemble adjustment Kalman filter (Anderson 2001, 2003) is applied to produce an ensemble of initial conditions. Large ensemble sizes (512 members) have been considered in this case so as to avoid some of the complications required in operational implementations (i.e., ensemble covariance localization). Covariance inflation is adopted to improve (perhaps artificially) the appearance of EnKF data assimilation results. For each experiment, the inflation parameter value is properly tuned in order to achieve best ignorance score. The inflation parameter values are 1.04 for the Ikeda experiment and 1.07 for the Lorenz96 experiment.

REFERENCES

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.,* **129,** 2884–2903.

——, 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.,* **131,** 634–642.

——, and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.,* **127,** 2741–2758.

Bennett, A. F., L. M. Leslie, C. R. Hagelberg, and P. E. Powers, 1993: Tropical cyclone prediction using a barotropic model initialized by a generalized inverse method. *Mon. Wea. Rev.,* **121,** 1714–1729.

——, B. S. Chua, and L. M. Leslie, 1996: Generalized inversion of a global numerical weather prediction model. *Meteor. Atmos. Phys.,* **60,** 165–178.

Bernardo, J. M., 1979: Expected information as expected utility. *Ann. Stat.,* **7,** 686–690.

Brocker, J., and L. A. Smith, 2007: Scoring probabilistic forecasts: On the importance of being proper. *Wea. Forecasting,* **22,** 382–388.

——, and ——, 2008: From ensemble forecasts to predictive distribution functions. *Tellus,* **60,** 663–678.

Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.,* **125,** 2887–2908.

Courtier, P., J. N. Thepaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4DVAR, using an incremental approach. *Quart. J. Roy. Meteor. Soc.,* **120,** 1367–1387.

Du, H., 2009: Combining statistical methods with dynamical insight to improve nonlinear estimation. Ph.D. dissertation, London School of Economics and Political Science, 190 pp.

——, and L. A. Smith, 2014: Pseudo-orbit data assimilation. Part I: The perfect model scenario. *J. Atmos. Sci.,* **71,** 469–482.

Fefferman, C. L., 2000: Existence and smoothness of the Navier–Stokes equation. Princeton University Department of Mathematics Tech. Rep. 6 pp.

Fertig, E. J., J. Harlim, and B. R. Hunt, 2007: A comparative study of 4D-VAR and a 4D ensemble Kalman filter: Perfect model simulations with Lorenz-96. *Tellus,* **59A,** 96–100.

Good, I. J., 1952: Rational decisions. *J. Roy. Stat. Soc.,* **16A,** 107–114.

Hagedorn, R., and L. A. Smith, 2009: Communicating the value of probabilistic forecasts with weather roulette. *Meteor. Appl.,* **16,** 143–155.

Hamill, T. M., 2006: Ensemble-based atmospheric data assimilation. *Predictability of Weather and Climate,* T. Palmer and R. Hagedorn, Eds., Cambridge University Press, 124–156.

——, and J. S. Whitaker, 2005: Accounting for the error due to unresolved scales in ensemble data assimilation: A comparison of different approaches. *Mon. Wea. Rev.,* **133,** 3132–3147.

——, ——, and C. Snyder, 2001: Distance-dependent filtering of background-error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.,* **129,** 2776–2790.

Hansen, J. A., and L. A. Smith, 2001: Probabilistic noise reduction. *Tellus,* **53A,** 585–598.

Haramel, S. M., C. K. R. T. Jones, and J. V. Moloney, 1985: Global dynamical behavior of the optical field in a ring cavity. *J. Opt. Soc. Amer.,* **2B,** 552–564.

Ikeda, K., 1979: Multiple valued stationarity state and its instability of the transmitted light by a ring cavity system. *Opt. Commun.,* **30,** 257–261.

Judd, K., 2008: Forecasting with imperfect models, dynamically constrained inverse problems, and gradient descent algorithms. *Physica D,* **237,** 216–232.

——, and L. A. Smith, 2001: Indistinguishable states I: The perfect model scenario. *Physica D,* **151,** 125–141.

——, and ——, 2004: Indistinguishable states II: The imperfect model scenario. *Physica D,* **196,** 224–242.

——, C. A. Reynolds, T. E. Rosmond, and L. A. Smith, 2008: The geometry of model error. *J. Atmos. Sci.,* **65,** 1749–1772.

Kennedy, M., and A. O'Hagan, 2001: Bayesian calibration of computer models. *J. Roy. Stat. Soc.,* **63B,** 425–464.

Leeuwen, P. J. V., 2010: Nonlinear data assimilation in geosciences: An extremely efficient particle filter. *Quart. J. Roy. Meteor. Soc.,* **136,** 1991–1999.

Leutbecher, M., and T. N. Palmer, 2008: Ensemble forecasting. *J. Comput. Phys.,* **227,** 3515–3539.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.,* **112,** 1177–1194.

Lorenz, E. N., 1995: Predictability: A problem partly solved. *Proc. Seminar on Predictability,* Shinfield Park, United Kingdom, ECMWF, 40–58.

Lu, C. G., and G. L. Browning, 1998: The impact of observational and model errors on four-dimensional variational data assimilation. *J. Atmos. Sci.,* **55,** 995–1011.

Miller, R. N., M. Ghil, and F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.,* **51,** 1037–1056.

Mitchell, H. L., and P. L. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.,* **128,** 416–433.

Mitchell, L., and G. A. Gottwald, 2012: Data assimilation in slow–fast systems using homogenized climate models. *J. Atmos. Sci.,* **69,** 1359–1377.

Orrell, D., L. A. Smith, T. Palmer, and J. Barkmeijer, 2001: Model error in weather forecasting. *Nonlinear Processes Geophys.,* **8,** 357–371.

Penland, C., 2003: A stochastic approach to nonlinear dynamics: A review. *Bull. Amer. Meteor. Soc.,* **84,** 921–925.

Pires, C., R. Vautard, and O. Talagrand, 1996: On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus,* **48A,** 96–121.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.,* **133,** 1155–1174.

Roulston, M. S., and L. A. Smith, 2002: Evaluating probabilistic forecasts using information theory. *Mon. Wea. Rev.,* **130,** 1653–1660.

Sasaki, Y., 1970: Some basic formalisms on numerical variational analysis. *Mon. Wea. Rev.,* **98,** 875–883.

Smith, L. A., 1992: Identification and prediction of low dimensional dynamics. *Physica D,* **58,** 50–76.

——, 1996: Accountability and error in ensemble forecasting. *Proc. Seminar on Predictability,* Shinfield Park, United Kingdom, ECMWF, 351–368.

——, 2002: What might we learn from climate forecasts? *Proc. Natl. Acad. Sci. USA,* **4,** 2487–2492.

——, M. C. Cuellar, H. Du, and K. Judd, 2010: Exploiting dynamical coherence: A geometric approach to parameter estimation in nonlinear models. *Phys. Lett.,* **374A,** 2618–2623.

Talagrand, O., and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Quart. J. Roy. Meteor. Soc.,* **113,** 1311–1328.

Zupanski, D., 1997: A general weak constraint applicable to operational 4DVAR data assimilation systems. *Mon. Wea. Rev.,* **123,** 1112–1127.