ACCOUNTABILITY AND INTERNAL CONSISTENCY IN ENSEMBLE FORMATION

Leonard A Smith & Isla Gilmour
Mathematical Institute, University of Oxford
Oxford, England

ABSTRACT

Ensemble forecasting provides the preferred approach for evaluating the uncertainty in forecasts of nonlinear deterministic systems, yet questions remain as to how best to construct ensembles over uncertain initial conditions, and which other uncertainties should be sampled. Forecasts of physical systems are effected by a variety of uncertainties in addition to uncertainty in the initial condition. By considering a perfect model scenario with an exact understanding of observational uncertainty these complications are banished, highlighting the importance of a perfect ensemble. In this paper the formation of ensembles over initial conditions is contrasted in perfect and imperfect model scenarios ranging from the 1963 Lorenz equations, through the thermally driven rotating annulus, to the Earth's atmosphere. Simple tests for consistency between operational constrained ensembles with their methods of formulation are proposed and illustrated. In addition, variational assimilation (which always converges to some result) is contrasted with ι-shadowing (which requires the proposed trajectory to agree with the observations within the observational uncertainty). In a perfect model scenario, a good variational assimilation technique will yield an ι-shadowing trajectory, while in an imperfect model scenario this is not the case; the inability of the model to shadow provides information on model error. Constructing ensembles over uncertainties other than those in the initial condition, such as ensembles over different models, are briefly noted.

1. INTRODUCTION

Ensemble forecasting attempts to quantify the uncertainties inherent in the prediction of nonlinear systems; it is most often invoked against uncertainty in the initial condition. In this case, an ensemble of initial conditions is evolved under a fully nonlinear model with the aim of quantifying some aspect(s) of the uncertainty in the forecast. In this paper, we consider the question of self-consistent ensemble formation in both perfect and imperfect models. Consistency tests can be based both on restrictions imposed by the dynamics of the system, and on restrictions required for the coherent formation of dynamically interesting subspaces (*e.g.* those spanned by singular vectors, or by bred vectors). After a discussion of dynamics in a perfect model state space in section 2, the complications of working in an imperfect model-state space are discussed in section 3. Here a test statistic is introduced which reflects the relevance of the linearized dynamics and which is easily computed whenever twin perturbations are available. A consistency test is illustrated in a laboratory fluid dynamics model, and then applied to operational NWP ensembles. The competing goals of decreasing sampling uncertainty and improving the properties of the population being sampled are discussed in section 4. Section 5 contrasts variational assimilation and ι-shadowing in the imperfect model scenario; near regions where the model can not ι-shadow, variational approaches may needlessly degrade the analysis. We examine perturbations which successfully ι-shadow data from the lab experiments, in particular their projection into the common constrained vector subspaces is presented. The paper concludes with some general discussion in section 6.

2. NONLINEAR DYNAMICAL SYSTEMS

The interpretation of numerical experiments is vastly simplified by working within a perfect model scenario; usually the system and the forecast model are taken to be identical, they evolve in the same state space, and hence it is known *a priori* that there exists a model trajectory which is

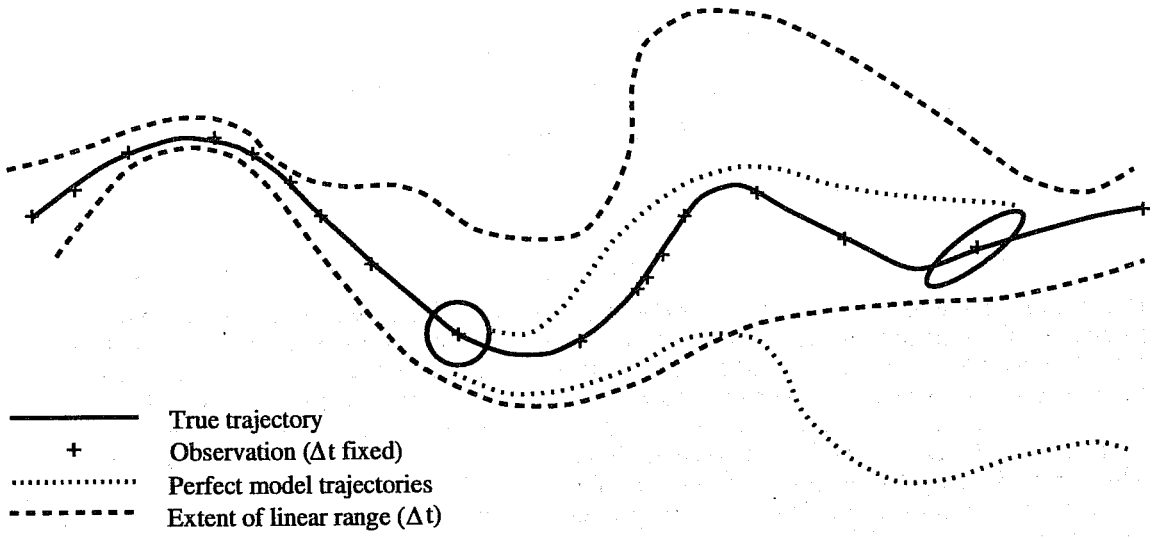

Figure 1. A schematic illustration representing a trajectory of the system (solid line) in its state space. Observations (+) are at equally spaced intervals (Δt) in time. Infinitesimal perturbations will evolve according to the linear propagator, as indicated by the circle evolving into the ellipse. The linear propagator is only approximate for finite uncertainties; an initial magnitude can be defined so that the approximation error is bounded for any given fixed time (in this case Δt). The extent of the linear range (indicated schematically by the dashed line) will vary with location. The two dotted lines are trajectories from "nearby" initial conditions.

indistinguishable from the the particular system trajectory of interest. This framework, also common in theoretical studies of nonlinear dynamical systems, is illustrated in figure 1 which shows a fudicial trajectory of the system and the behaviour of nearby points in the state space. There are, of course, alternative scenarios such as using one model as the system and another model with a somewhat different structure (*e.g.* a lower spatial resolution) as the forecast model. Interpretations of such experiments still suffer both from the obvious fundamental similarities of the two variants, and from the fact that the practitioner is aware of these.

The perfect model scenario can never be realized in the study of any physical system, yet it provides an accessible testbed which will be contrasted with imperfect models below. One nicety of the perfect model scenario is that almost everything is well defined; for example, it makes sense to speak of Lyapunov vectors (defined below) in a perfect model scenario, while they may not be well-defined in other modelling scenarios, as discussed in the next section. A perfect model of the dynamics and an exact understanding of of the observational uncertainty (*e.g.* the observational error covariance matrix) are *not* sufficient to construct an accountable forecast probability density function (PDF). For example, if the dynamics are restricted to a manifold with dimension less than that of the state space[1] then an ensemble distributed only via the covariance matrix will include initial conditions not on the manifold. The initial ensemble will *not* reflect the true initial PDF in state space, and consequently the PDF derived from the ensemble at final time will be incorrect. Thus, as the size of the ensemble increases, the final PDF will not improve accountably: the error will exceed that due to sampling uncertainty. In short, we can compute the probability of an observation $\mathbf{x}_{obs}$ given the true state $\mathbf{x}_{true}$ and the covariance matrix, but we cannot compute the probability that a point $\mathbf{x}$ is the true state given only $\mathbf{x}_{obs}$ and the covariance matrix *unless* we know whether or not the point $\mathbf{x}$ lies on the manifold (see *Smith* 1996 and references therein for additional discussion of this point). No smooth multi-normal distribution can represent a perfect ensemble in this case.

The invariant manifold of a dynamical system describes the subspace of the state space within which the system will evolve. All physically relevant states of the system lie on this manifold. It is common to assume, without justification, that this manifold is the entire state space; this can not be

[1] Or even further restricted to an attractor.

the case whenever the state space includes physically unrealizable states (which can not lie on the manifold). Further, physically realizable states need not be uniformly distributed: some regions of state space may be more densely populated than others and some physically realizable states will lie on transients (never be visited more than once) and hence do not contribute the long term statistics. The probability distribution of the climatology of the system, which lies on this manifold, is often called the natural measure; note that this measure need not be either homogeneous or isotropic on the manifold.

2.2 Ensemble formation

Consider a true initial condition $\mathbf{x}(t)$ (the state of the system at time t), and an analysis $\mathbf{A}_{\Delta t}(t)$ (a best guess of $\mathbf{x}(t)$ given all the information available at time $t + \Delta t$ and a particular model). Even if the observational error is nothing more than quantization error, obtaining a perfect ensemble is non-trivial. An observation with quantization error defines $\mathbf{x}(t)$ to be within a hyper-cube of the state space, but gives no further information. How might one generate an ensemble of initial conditions which are not only within the observed hyper-cube, but also on the attractor? One method, suggested by *Lorenz* (1963), is simply to integrate the system of interest and collect exact analogs. Exact, that is, to within our measurement accuracy, thereby obtaining points within the same hyper-cube *and also* on the attractor. The trajectories of these analogues will lie on the attractor, even though they are only *observed* to finite precision. Even this approach assumes successive analogues in time sample the attractor without bias (*i.e.* that they have no statistical correlation whatsoever), an assumption unlikely to be fulfilled by any deterministic system. For operational forecasting models such an approach is infeasible even in the case that the model's climatology is sufficiently similar to that of the system. An alternative is suggested below.

The aim of the perfect ensemble is to forecast an estimate of the correct PDF. The error covariance matrix assigns a relative probability to each member of the ensemble depending on its location relative to the analysis. An alternative aim would be to predict the reliability of the forecast by considering only those members which are likely to contribute significantly to the spread of the ensemble. Here "likely to contribute" reflects a relatively high initial probability of occurrence, and "contribute significantly to the spread" indicates an initial condition whose image at the forecast time is far from that of the analysis. If the distance between the analysis and the true state is sufficiently small (infinitesimal will suffice), one may approximate the dynamics over a time Δt by the linear propagator, $\mathcal{M}(\mathbf{A}_0(t), \Delta t)$. The leading right singular vectors of $\mathcal{M}(\mathbf{A}_0(t), \Delta t)$ indicate the directions along which infinitesimal perturbations which will have grown the most during the interval $[0, \Delta t]$. As indicated in figure 2, the image of finite perturbations evolved under the fully nonlinear flow will, in general, differ from their image under the linear propagator. The issue is whether or not this difference is significant; we shall return to this question in the context of the internal consistency of ensemble formulation below. For the moment, we will define singular vector (SV) ensembles as those which are restricted to a subspace spanned by the leading singular vectors of $\mathcal{M}(\mathbf{A}_0(t), t_{opt})$, where the optimization time, t_{opt}, is fixed.

The global Lyapunov vectors, LV_∞, are defined at a point $\mathbf{x}$ through the eigen-decomposition of $\mathcal{M}(\mathbf{x}, \Delta t)$ in the limit $\Delta t \to -\infty$; obviously $\mathbf{x}$ must lie on the invariant manifold for the LV_∞ to be well defined (a trajectory must exist along which to take the limit). Finite time Lyapunov vectors (LV) are similarly defined for finite $\Delta t < 0$. In the perfect model scenario there will exist a model trajectory arbitrarily close to the fudicial (system) trajectory. In this case the Lyapunov vectors at a point $\mathbf{x}(t)$ of the system are relevant for the analysis $\mathbf{A}_0(t)$ provided that the analysis lies on the manifold (and, of course, assuming infinitesimal uncertainties). It is often argued that the leading LV_∞ represent "the" orientations of sustainable growth; this is true and thus the LV_∞ are often irrelevant for finite time forecasts, since their local orientations are dependent upon events arbitrarily far away in the past/future. For arbitrarily long finite spans of time, perturbations along the first Lyapunov vector (indeed even along the unstable manifold, should one exist) of a chaotic system may *decrease* with time.

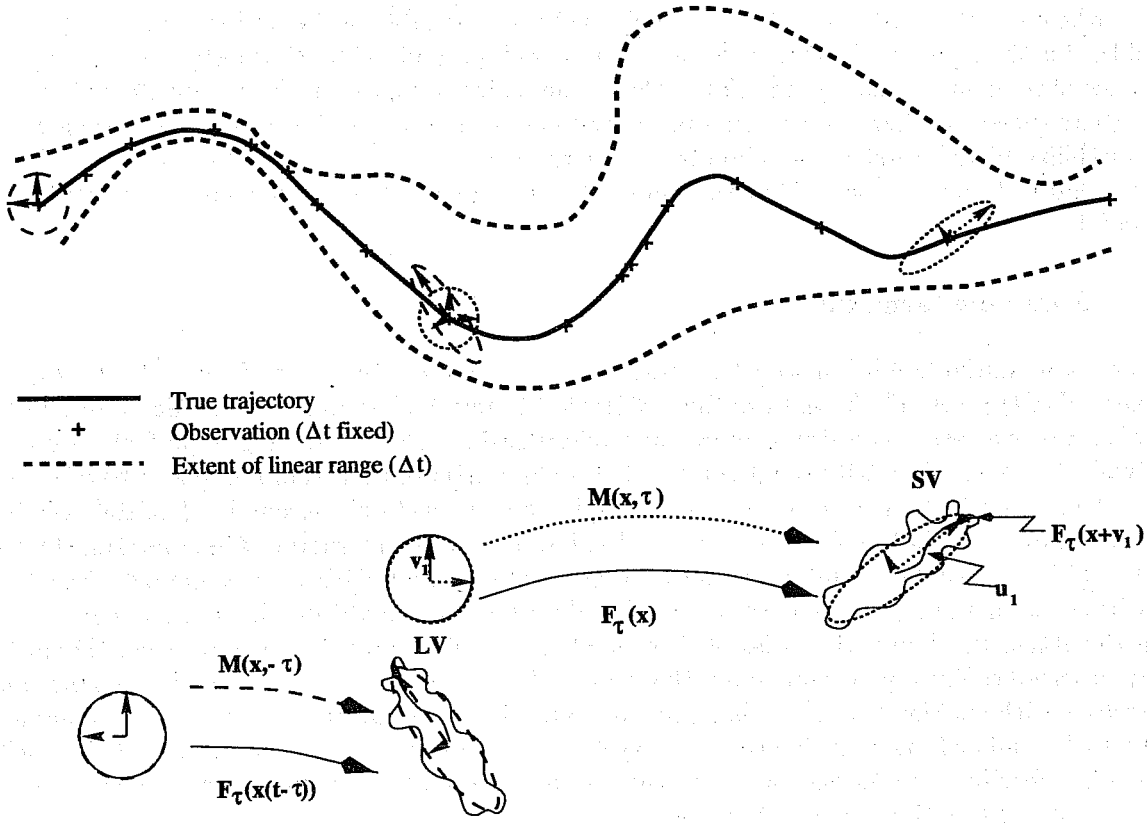

Figure 2. A schematic illustration showing a system trajectory (solid line), observations (+) equally spaced in time and the extent of the linear range (dashed lines), as in figure 1. Evolution under the linear propagator reflects the growth of infinitesimal perturbations (the circle evolves into the ellipse); forward evolution defines the SV. Finite perturbations will evolve nonlinearly under the model, F, (the circle evolves into a 'wavy' closed curve near the ellipse). If the perturbation lies within the linear range, then the evolution of the perturbation under the model is well approximated by the linear propagator. The linear evolution over historical times yields the Lyapunov vectors (LV), defined in the text.

SV ensembles avoid this dependence on irrelevant events in the far future by considering the linear evolution over a fixed optimization time t_{opt}; but if rate of uncertainty growth changes with location, a global value of t_{opt} is not desirable: in regions of rapid uncertainty growth, the nonlinearities may become important at times less than t_{opt}. In low-dimensional dynamical systems, the first singular value, σ_1, of $\mathcal{M}(\mathbf{x}(t), t_{opt})$ often varies greatly with $\mathbf{x}$; if, for some $\mathbf{x}$, σ_1 is so large that nonlinearities are important[2] after a time $\Delta t' \ll t_{opt}$, then the singular vectors of $\mathcal{M}(\mathbf{x}(t), t_{opt})$ need not reflect the directions of interest. Nightmare vectors (NV) are defined by setting some threshold on the allowed linear growth, say a factor of 4; if this threshold is not exceeded within $0 \leq t \leq t_{opt}$ then the NV are identical to the SV. However, if the threshold is exceed at time $\Delta t' < t_{opt}$, then the NV are the right singular vectors of the linear propagator evaluated over this (shorter) time. The behaviour of a variety of ensembles within the perfect model scenario is discussed by *Smith, Ziehmann and Fraedrich* (1997) where the importance of these effects in the Lorenz 1963 and Moore-Spiegel systems is illustrated; these low-dimensional systems allow verification via large perfect ensembles formed by finding analogues from a long integration. In some cases, the NV ensembles out-perform two member perfect ensembles detecting regions of rapid uncertainty growth in (much larger) perfect ensembles. While all methods perform relatively well in the Lorenz case, the importance of locating an initial condition on the manifold is apparent in the Moore-Spiegel system, where two-member perfect ensembles consistently out-perform all the dynamically constrained ensembles in reflecting the true increase in the spread with time. More detail (and figures) are given in the reference.

[2]Clearly this will depend upon the initial magnitude of the perturbations.

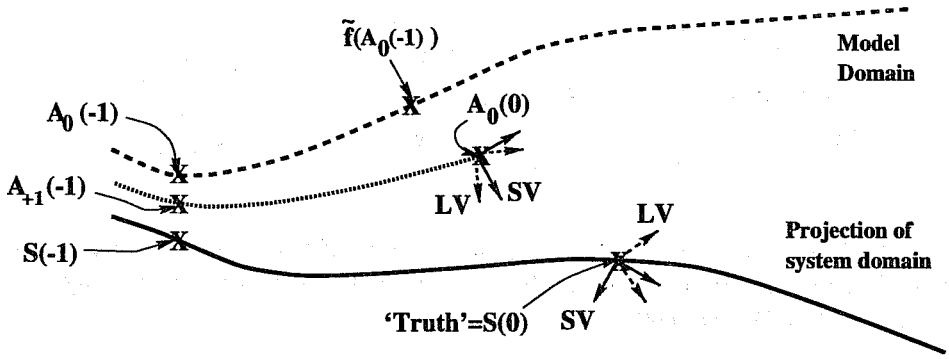

Figure 3. Schematic illustration of the projection of a system trajectory into the imperfect model state space (solid line) with projections of true states at time t denoted by $\mathbf{S}(t)$. The analysis (defined in text) at time $t = -1$ is updated from $\mathbf{A}_0(-1)$ to $\mathbf{A}_{+1}(-1)$ when information at $t = 0$ becomes available, giving an improved estimate of the projected system state, $\mathbf{S}(-1)$. Evolving the model from $\mathbf{A}_0(-1)$ gives a model trajectory (dashed line) distinct from the projected system trajectory. An alternative trajectory (dotted line), *e.g.* provided using a variational assimilation technique, is shown; it is assumed that this trajectory extends continuously through analysis values $\mathbf{A}_t(-t)$ as $t \to -\infty$, enabling the definition of LV_∞ at $\mathbf{A}_0(0)$. The projections of the system LV and SV at $\mathbf{S}(0)$ into the model state space (assumed here to exist) are shown, as are the SV at $\mathbf{A}_0(0)$.

3. IMPERFECT NONLINEAR DYNAMICAL MODELS

In general, of course, the model is not the system. A few of the complications this introduces are illustrated in figure 3. One crucial difference is that the model state space is no longer easily identified with the system state space. Schematically, figure 3 shows the true system trajectory "projected" into the model state space, but this projection need not be one-to-one, and the projected trajectory need not be a trajectory of either the system or the model (just as the path of an ensemble mean need not be a trajectory of either the model or the system).

The projection of the true state at time t into the model state space is denoted by $\mathbf{S}(t)$. The best estimate of the optimal model state[3] at time t available at time $t+\Delta t$, is taken as the analysis $\mathbf{A}_{\Delta t}(t)$. As additional information becomes available, the analysis is updated, *e.g.* the analysis $\mathbf{A}_0(t = -1)$ available at $t = -1$ is updated at $t = 0$ to $\mathbf{A}_1(-1)$ by incorporating additional information; this later analysis is indicated as being nearer to the projected true state $\mathbf{S}(-1)$. If the assimilation used to determine the analyses can provide a continuous trajectory through each $\mathbf{A}_t(-t)$ as $t \to \infty$, then the LV_∞ about $\mathbf{A}_0(0)$ are well defined[4]. Otherwise, since the model is imperfect, the LV of the model may not exist even in cases where LV are well-defined in the system state space. Similarly, the projection of the SV and LV of the system into the model state space may not exist (the system may not be equivalent to *any* system of differential equations; and then even if the system SV and LV exist, the projection may not be meaningful). The SV depend only upon the forward integration of the model, and thus can be defined at $\mathbf{A}_0(0)$, as can the NV.

Ideally, the projected system SV at $\mathbf{S}(0)$ exist and are well approximated by the model SV at $\mathbf{A}_0(0)$; the question of whether this is actually the case is unanswerable. If we are able to find a model trajectory consistent with each observation (that is, within the observational uncertainty) for $t \in [0, \tau_\iota]$, then the model ι-shadows the system over a time τ_ι. This approach may be taken to define a set of analysis values, $\mathbf{A}_{\tau_\iota - t}(t), t \in [0, \tau_\iota]$; including $\mathbf{A}_{\tau_\iota}(0)$ in the ensemble constructed at $t = 0$ yields a forecast consistent with all the observations up to (at least) $t = \tau_\iota$. For a sufficiently large τ_ι, $\mathbf{A}_{\tau_\iota}(0)$

[3] In this context, "optimal" will be a function of the problem at hand.

[4] Often, $\mathbf{A}_0(0)$ does not lie on the invariant manifold of the model and hence no such continuous trajectory exists: running freely, the model would never return to within a small (finite) neighbourhood of $\mathbf{A}_0(0)$. If this is the case, then global Lyapunov vectors are not defined at this location.

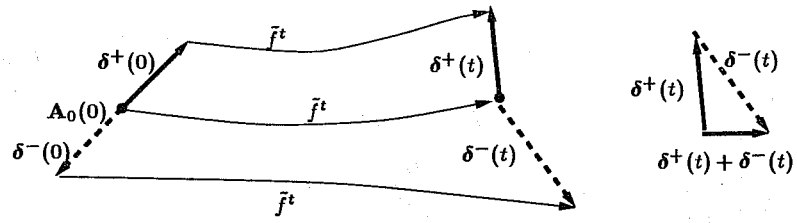

Figure 4. Defining Θ: Equal and opposite perturbations at $t = 0$, $\boldsymbol{\delta}^{\pm}(0)$, evolve so as to no longer symmetric at time t: the error in assuming linear dynamics, $||\boldsymbol{\delta}^+(t) + \boldsymbol{\delta}^-(t)||$, is scaled by the average magnitude of the evolved perturbations to give the test statistic Θ.

defines an optimal[5] model state at $t = 0$; we may now require that the SV evaluated at $\mathbf{x} = \mathbf{A}_0(0)$ be similar to the SV at $\mathbf{x} = \mathbf{A}_{+\tau_t}(0)$ for the model SV available at time $t = 0$ to be considered relevant: such similarity will clearly be a function of the optimization time, the observational uncertainties and the local structure of the model state space. If the SVD about these two points differ significantly, then the SV ensembles will not reflect the relevant subspace, even for infinitesimal perturbations at $t = t_{opt}$. This test of relevance, however, requires both knowledge of the optimal model state (which is known only after the fact) and additional model integrations. Below, we turn to a test of the internal consistency of the SV ensemble which avoids both of these constraints.

As noted above, in order to formulate a SV ensemble in an internally consistent manner, the linearized dynamics must provide a good approximation of the model dynamics up until t_{opt} for the operational perturbations: the ensemble must be consistent with the assumptions under which it is formulated. If the evolution of operational perturbations becomes highly nonlinear for some $t < t_{opt}$, then the SV defined on $[0, t_{opt}]$ need not reflect which perturbations will have grown the most at $t = t_{opt}$ (or which will first feel the nonlinearity). The extent to which the linearized dynamics are a relevant approximation may be computed from the results of twin perturbations already present in operational ensembles, thus avoiding additional model integrations.

If the control trajectory, initiated from $\mathbf{A}_0(0)$, is taken as the fudicial trajectory, and the deviations of a positive (negative) perturbation from the control in evolution are denoted by $\boldsymbol{\delta}^+(t)$ ($\boldsymbol{\delta}^-(t)$), then a test statistic for linearity, Θ, may be defined as

$$\Theta(\hat{\boldsymbol{\delta}}, ||\boldsymbol{\delta}||, t) = \frac{||\boldsymbol{\delta}^+(t) + \boldsymbol{\delta}^-(t)||}{0.5\{||\boldsymbol{\delta}^+(t)|| + ||\boldsymbol{\delta}^-(t)||\}} \tag{1}$$

(where $||\cdot||$ is an appropriate norm); Θ will, of course, vary with the $\mathbf{A}_0(0)$ of initialization. Θ reflects the error which would be made in assuming linear dynamics; as long as the perturbations are growing linearly, $\boldsymbol{\delta}^+ = -\boldsymbol{\delta}^-$ and $\Theta = 0$. Θ quantifies the error made in assuming this linear equality as a fraction of the average magnitude of the evolved perturbations. Thus $\Theta = 0$ implies that the linearized dynamics *may*[6] be exact; $\Theta = 0.5$ implies that an error made will be at least 50% of the average magnitude of the evolved perturbations. The definition of Θ is reflected by the schematic in figure 4.

How might this information be used? Figure 5 shows the fraction of initial conditions for which $\Theta < 0.2$ as a function of time for SV perturbations from a model of the thermally driven rotating annulus (see *Smith* 1992, *Read et al.* 1991, *Smith* 1996). Results are shown for three optimization times ($t_{opt} = 2, 4, 8$ time steps), and two initial magnitudes (0.01°C, 0.1°C). Note that the data separates into two regions: for a perturbation size of 0.1°C (the lower set of curves) the loss of linearity is extremely rapid, the fraction of initial conditions for which the linear approximation holds at the 20% level (*i.e.* $\Theta < 0.2$) is about 10-15% of the initial conditions at $t = 4$ and less than

[5] In terms of prediction.

[6] It is crucial to remember that $\Theta = 0$ is only a necessary condition from linearity; one can contrive examples (*e.g.* cubic nonlinearities) where $\Theta = 0$ for some perturbations and yet the linear approximation is wildly inaccurate.

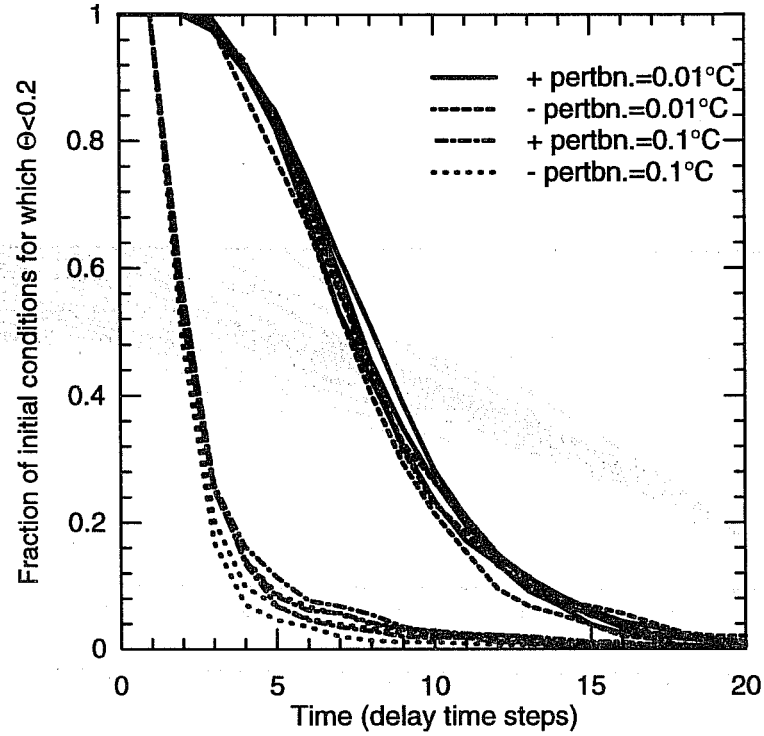

Figure 5. Twin experiment linearity results of SV perturbations from a model of the thermally driven rotating annulus. The fraction of initial conditions for which $\Theta < 0.2$ is shown as a function of time for 2 different initial perturbation magnitudes. At each magnitude there are 3 sets of results, obtained by varying the optimization time ($t_{opt} = 2, 4, 8$ time steps); there is little difference between these sets.

5% at $t = 8$. Perturbations of this magnitude are comparable with the true experimental uncertainty in the annulus and thus figure 5 indicates that adopting an optimization time of 4 or 8 time steps is inadvisable: the growth of realistically sized perturbations is nonlinear before such optimization times. The upper curves correspond to initial perturbations of magnitude 0.01°C; in this case the linearized dynamics dominates for over 90% of the perturbations at $t = 4$. Thus, if the uncertainty in the analysis was of order 0.01°C, an optimization time of $t_{opt} = 4$ would be reasonable; in general the larger the analysis uncertainty the shorter a justifiable t_{opt} for a given system.

For any operational forecast system using twinned perturbations one may compute $\Theta(t)$ as a test of internal consistency. Results from the 25 twins in a single SV ensemble forecast from ECMWF ($t_{opt} = 48$ hours) of 500 mb height are shown in figure 6. In this single case, $\Theta(t_{opt})$ ranges from 0.5 to > 1.0 (see panel 1, fig. 8); the error in assuming linearity at the optimization time ranges from 50% to more than 100% of the mean magnitude of the evolved perturbations, which seems rather large. The third panel in figure 6 shows Buizza's correlation ℓ statistic (*Buizza* 1995)

$$\ell = -\frac{(\boldsymbol{\delta}^+(t), \boldsymbol{\delta}^-(t))}{||\boldsymbol{\delta}^+(t)|| \cdot ||\boldsymbol{\delta}^-(t)||}. \tag{2}$$

The line representing a correlation coefficient of 0.7 is indicated since this is taken as the 'threshold' value, below which nonlinearities are considered to dominate (*Buizza* 1994); here it is crossed between 36 and 84 hours. Note however, that $\ell = 0.7$ corresponds to the perturbations deviating by $\sim 45°$ from anti-parallel; the corresponding error in assuming linearity is $\sim 75\%$ of the mean magnitude of the evolved perturbations (*i.e.* $\Theta > 0.75$). There is no inconsistency here since both statistics are necessary conditions for linearity; ℓ reflects only the orientation of the evolved perturbations, and is blind to their relative magnitudes. The difference between the magnitudes, as a fraction of their mean magnitude, is shown on the second panel of figure 6: note that at 48 hours the magnitudes in one case differ by $> 20\%$ (in this case, the corresponding value of $\ell = 0.5$ while $\Theta > 1.0$).

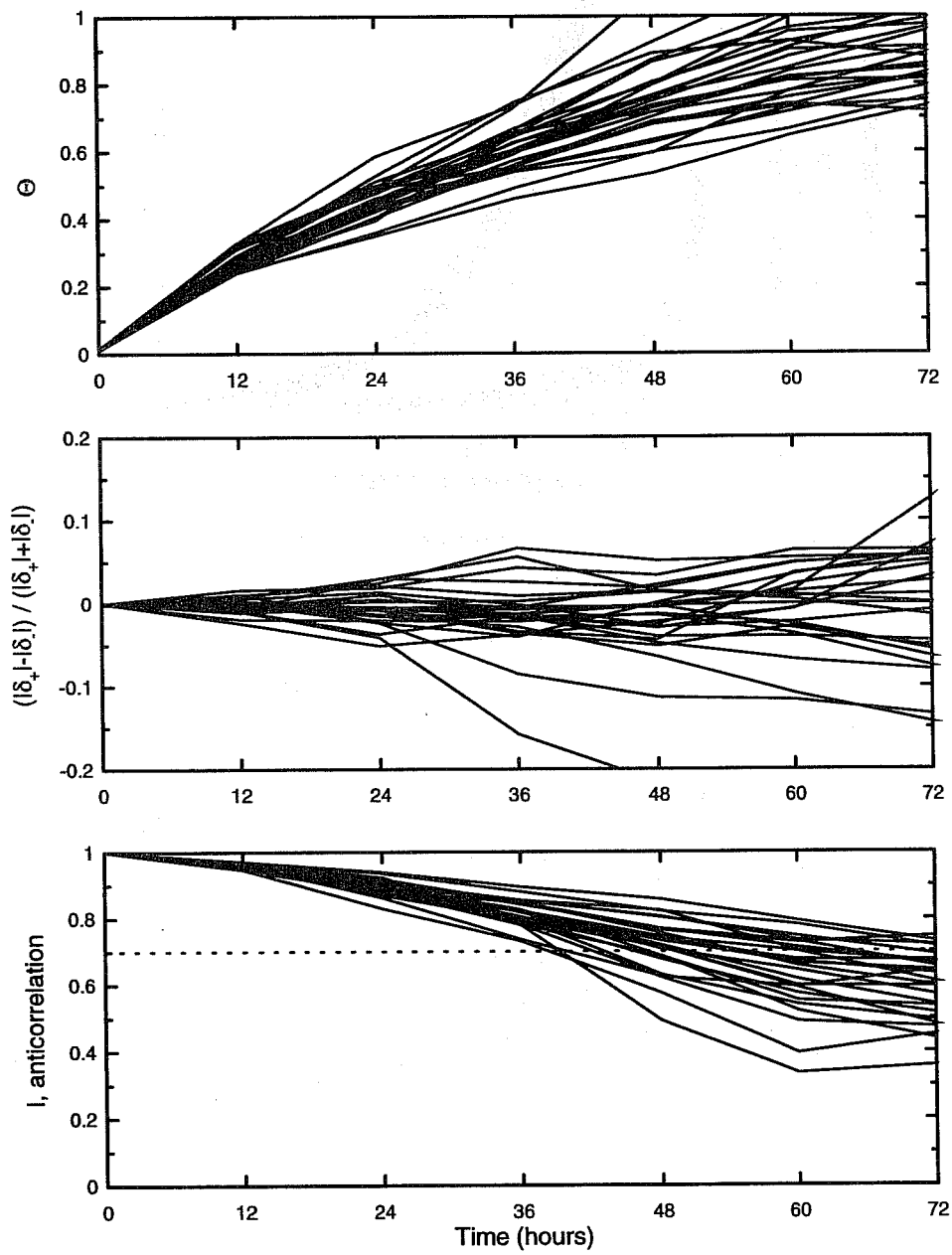

Figure 6. Linearity results of 25 twin SV perturbations (t_{opt}=48 hours) for the 500 mb geopotential height from the operational ECMWF model. The perturbations were initiated at 00h on December 12th, 1996 and the norms are taken over the Northern hemisphere excluding the Tropics (22.5° N - 90° N). The panels show, from top to bottom, Θ (see eqn. 1), the difference in magnitude of the positive and negative evolved perturbations as a fraction of their sum, and Buizza's ℓ statistic (see eqn. 2), all as functions of time.

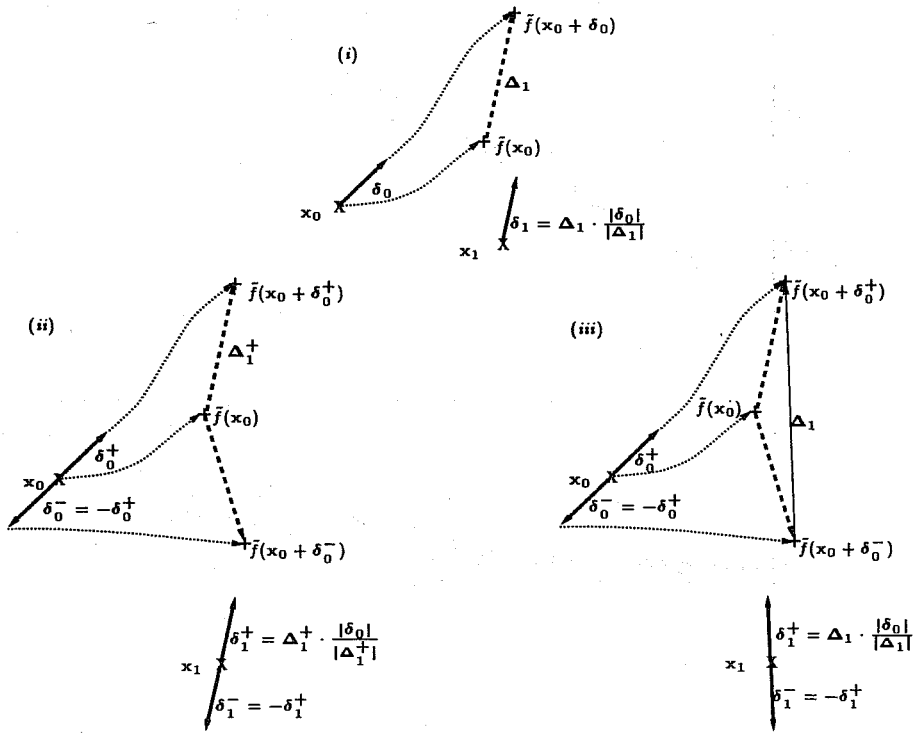

Figure 7. Schematics of breeding. (*i*) The ideas of breeding vectors: A perturbation $\boldsymbol{\delta}_0$ (solid line at $t = 0$) is added to the analysis $\mathbf{x}_0$ at initial time and both are iterated forward under the model (dotted lines). The evolved perturbation $\boldsymbol{\Delta}_1$ (dashed line) is then rescaled to the magnitude $|\boldsymbol{\delta}_0|$ of the initial perturbation to give the new BV $\boldsymbol{\delta}_1$ (solid line at $t = 1$) which is used as the perturbation from the new analysis $\mathbf{x}_1$. If the dynamics are linear for magnitude $|\boldsymbol{\delta}_0|$ of perturbation, then twins of BV may be generated by considering $\pm\boldsymbol{\delta}_0$ which will evolve to $\pm\boldsymbol{\Delta}_1$ and be rescaled to $\pm\boldsymbol{\delta}_1$ .... If the dynamics are not linear there is a choice of how to generate BV twins. (*ii*) Method a: Chose one of the evolved twins (here the positive) over the other, rescale and introduce its symmetric image. (*iii*) Method b: Alternatively the difference between the evolved twins may be rescaled and its symmetric image introduced.

One would like to construct a similar consistency test for operational BV ensembles. Figure 8 shows the statistics of figure 6 for a NCEP BV ensemble forecast. The construction of BV does not require the linearization hold, *per se*; yet in order for the BV subspace not to depend sensitively upon the breeding perturbation size, perturbations of that magnitude should evolve linearly over the breeding cycle. The generation of bred vectors is illustrated in figure 7. Ideally twin perturbations about the control are evolved for one breeding cycle, at the end of which each is rescaled (usually shrunk) to the original magnitude; *if* the linearized dynamics describes this evolution *then* the rescaled perturbations remain symmetric about the evolved control. *If not* then one must either (*i*) chose one of the twins over the other and introduce its symmetric image (method *a*), (*ii*) double the size of the ensemble by introducing symmetric images of both perturbations, or (*iii*) combine the two in some way, *e.g.* use the difference between the evolved twins and its symmetric image (method *b*). Operationally, NCEP performs the third of these options. If the dynamics are linear then this is equivalent to method *a*, otherwise the operational definition is sensitive to nonlinearities. Data is not available at the breeding cycle time of 6 hours, however examining $\Theta(t = 12)$ suggests that nonlinearities will often produce ambiguities at the 10-15% level at each 6 hour time step. Applying a rescaling mask would introduce additional ambiguities.

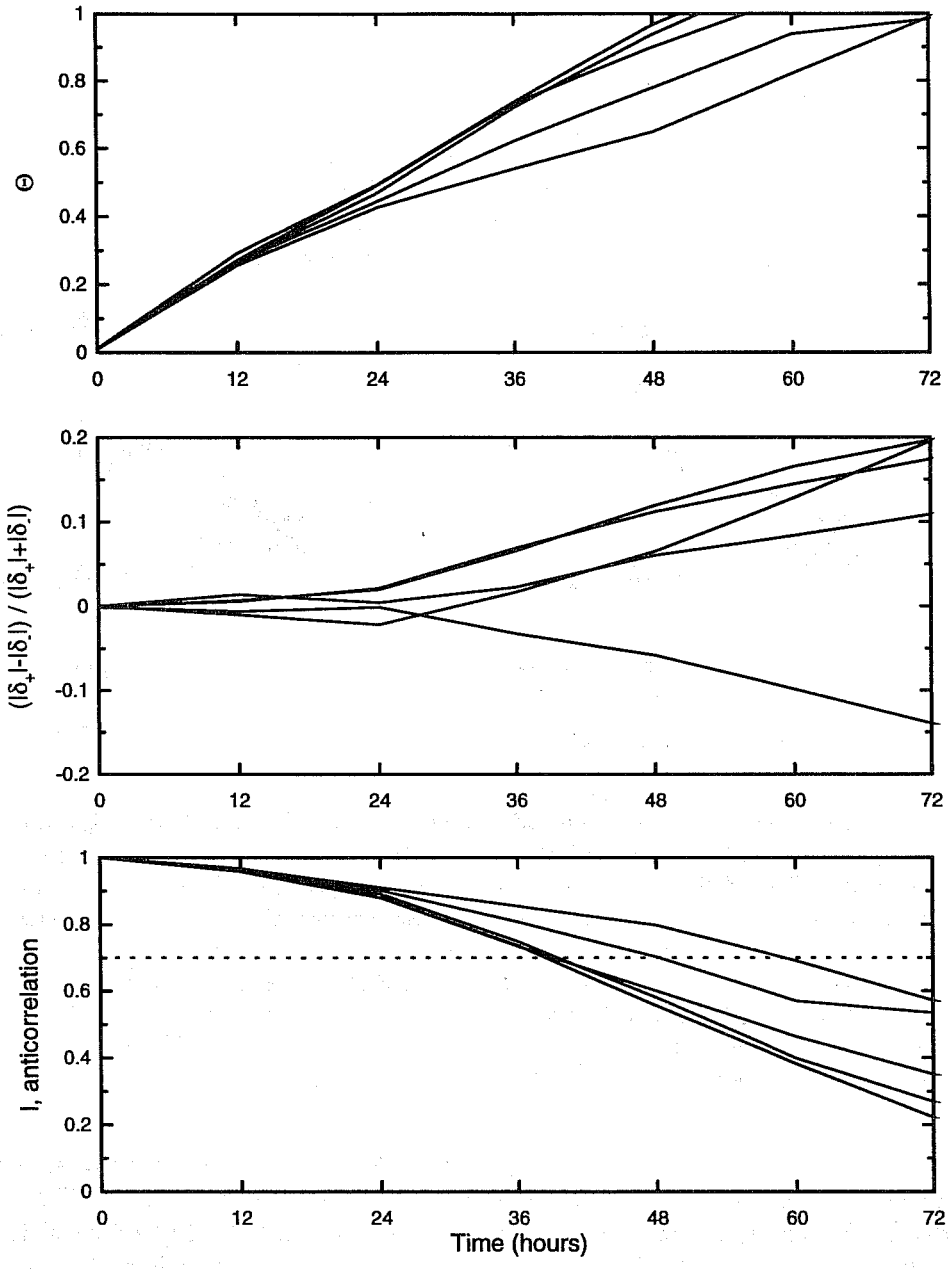

Figure 8. Linearity results of 5 twin operational BV perturbations (generated by NCEP) for the 500 mb geopotential height from the operational NCEP model. The perturbations were initiated at 00h on October 10th, 1997 and the norms are taken over the Northern hemisphere excluding the Tropics (22.5°N-90°N). The panels show, from top to bottom, Θ (see eqn. 1), the difference in magnitude of the positive and negative evolved perturbations as a fraction of their sum, and ℓ (see eqn. 2), all as a functions of time.

4. ON THE INTERPRETATION OF NON–PERFECT ENSEMBLES

4.1 Almost all Perturbations Lie "Off the Attractor"

The interpretation of forecast probability density functions based upon perfect ensembles is straightforward. When the observational uncertainty is due only to quantization error as discussed in Section 2.2, then each member of the ensemble is equally likely; the distribution of the ensemble members at forecast time may be interpreted as an unbiased estimate of the true forecast PDF. This forecast PDF will be *accountable*: its only shortcoming will be due to sampling uncertainty. Under a perfect model, as the number of ensemble members increases, the distribution of ensemble members will

converge to the true system PDF arising from the observational uncertainty. This is what makes such ensembles "perfect". If the observational uncertainty is more complex than quantization error, then the members of the ensemble will no-longer carry equal weight; those "closer" to the analysis usually receiving greater weight. As long as the statistics of the observational uncertainty are known (*i.e.* if the error covariance matrix accurately reflects the observational uncertainty), then again the forecast PDF will suffer only from sampling uncertainties which will decrease (accountably) as the number of ensemble members increases.

Current operational ensemble formation schemes are not accountable for at least three reasons: (*i*) in general, members are not weighted with respect to their likelihood given the analysis[7], (*ii*) the members are often chosen in a low dimensional subspace of the state space (as defined by SV, BV, LV, *etc.*) and (*iii*) the members are not weighted with respect to the invariant measure of the system (and so they are both (1) "off the attractor" and (2) ignore the variations in the initial probability density on the attractor, or manifold, even when the manifold intersects the constrained subspace).

Consider the Lorenz attractor; while the true initial condition is on the attractor, the relative probabilities assigned by the error covariance matrix correspond to ellipses centered about the observation. The major axis of these ellipses is determined by the accuracy with which we observe and need *not* coincide with the orientation of the local structure of the attractor itself.

In general an unconstrained ensemble, sampling with respect to the covariance matrix for example, will select points in the entire state space and will not respect the fact that the true initial condition lies on the attractor. Constrained subspace ensembles (SV, BV, LV, *etc.*), will sample a subspace of the full state space, but in general this subspace will not be confined to the attractor either; indeed subspaces based on the linearized dynamics might have only a small intersection with the attractor. This is the origin of the claim that they may point "off the attractor"; in fact almost any Euclidean (*i.e.* flat) subspace will have a very limited intersection with the attractor. Thus constrained ensembles will represent a different initial PDF than an ensemble composed solely of points on the attractor. As the constrained ensemble increases in size, this difference will become apparent since the sampling uncertainty of these forecast PDFs will not decrease accountably. No constrained ensemble of this kind can be perfect. Unconstrained ensembles, in contrast, sample the entire state space: these cannot be accountable either if the true initial conditions lie on a manifold of lower dimension.

4.2 On Increasing N: Distinguishing Sampling Uncertainty and Sampling Error

When increasing the size of an ensemble, one may target either sampling uncertainty, sampling error, or both. Sampling uncertainty comes about whenever a finite sample of N members is drawn from a larger, perhaps infinite, population. As N increases the distribution obtained from the sample converges toward that of the population. In addition, the larger value of N provides a better resolved sample PDF. Alternatively, sampling error will arise when the population which is sampled differs from the population which is actually of interest. Increasing N will always decrease the sampling uncertainty, but this may well result in a better approximation of the wrong PDF! Thus when increasing N, one is forced to divide resources between the options of improving the sampling density and improving the population being sampled.

An operational example is provided by the increase of the ECMWF SV ensemble from 33 to 51 members. The first method discussed above would have implied drawing the additional 18 members from the same 16-dimensional subspace as the first 33. This would have decreased the sampling uncertainty and the operational ensemble would have described the PDF of this 16-dimensional subspace with greater resolution. In practice, the second method was chosen: the 51 members are

[7]Of course, if they are chosen uniformly with respect to the distribution of observational uncertainty this weighting is implicit.

taken from a higher dimensional subspace, improving the relevance of the population from which elements are drawn.

5. VARIATIONAL ASSIMILATION AND SHADOWING

Given a perfect model and an infinite series of good observations, variational data assimilation techniques (see *Talagrand and Courtier* 1987 and references thereof) allow the construction of a model trajectory which is consistent with the true trajectory to within the observational uncertainty. As argued by *Pires et al.* (1996), it will sometimes be difficult to reduce uncertainties which lie along the stable manifold in this case. In contrast to Pires *et al.* we note that over any finite period of time, displacements in the (asymptotically) stable directions may increase, allowing uncertainties along these directions to be reduced as well.

Given finite data and an imperfect model, variational assimilation will still locate the minimum of any given cost function, but the trajectory thereby identified need not resemble either the true trajectory or the data. Alternatively, one may seek a model trajectory which is consistent with the observations to within the observational uncertainty; such a trajectory is said to ι-shadow the observed true trajectory (see *Gilmour and Smith* 1997). In general, there will exist regions of the state space through which the model cannot shadow the observations due to model error. When the trajectory passes through such a region, variational techniques will tend to corrupt portions of the trajectory which could have been shadowed in order to decrease the cost function in a region which can *not* be made consistent with the observations. Rather than forcing an unphysical "best fit", shadowing trajectories simply terminate in such regions, allowing the resulting trajectory to remain consistent with the observations where this is possible, and illuminating regions of state space where this is not possible for further study.

An example for the annulus data is given in figure 9; the solid line reflects the observations. A series of ensemble forecasts are reflected by the dots scatters about solid curve, while a large circle denotes the initiation of a new ensemble (at $t = 11, 17, 21, 29$). In this case, the ensembles are unconstrained; they loose calibration after about 4 steps, and spread is on the order of a quarter the range of the data after 6 steps. The dashed trajectory reveals, however, that there exists a model trajectory which shadows the observations to within 0.2°C of each observation for 26 steps - significantly longer than the prediction time.

Is there some method by which we could have increased the probability of including this particular initial condition in the relevant ensemble? Figure 10 is a scatterplot of the dot products between this "dream perturbation" (DP) (the unit vector from the observation to such an initial condition) and the leading BV (method *a*), the leading SV, and the leading LV. Restricting attention to only those initial conditions where at least one of the constrained subspaces has a large projection upon the DP (say, CV $\bullet$ DP > 0.8), shows that the BV consistently outperform other methods; the reason for this is not clear. Work is underway to construct a constrained subspace consistent with local distribution of the data, in an attempt to determine whether the DP reflects the need to restrict perturbations to be "near the attractor."

6. CONCLUSIONS

Ensemble forecasts provide a powerful tool for prediction by providing probabilistic information regarding the occurrence of an event; in order to utilize the information available from ensemble forecasts, they must be interpreted using knowledge both of the method of formulation of the ensemble and of the model representation of the system. Further, assumptions made in formulation must hold operationally for the ensemble to be internally consistent.

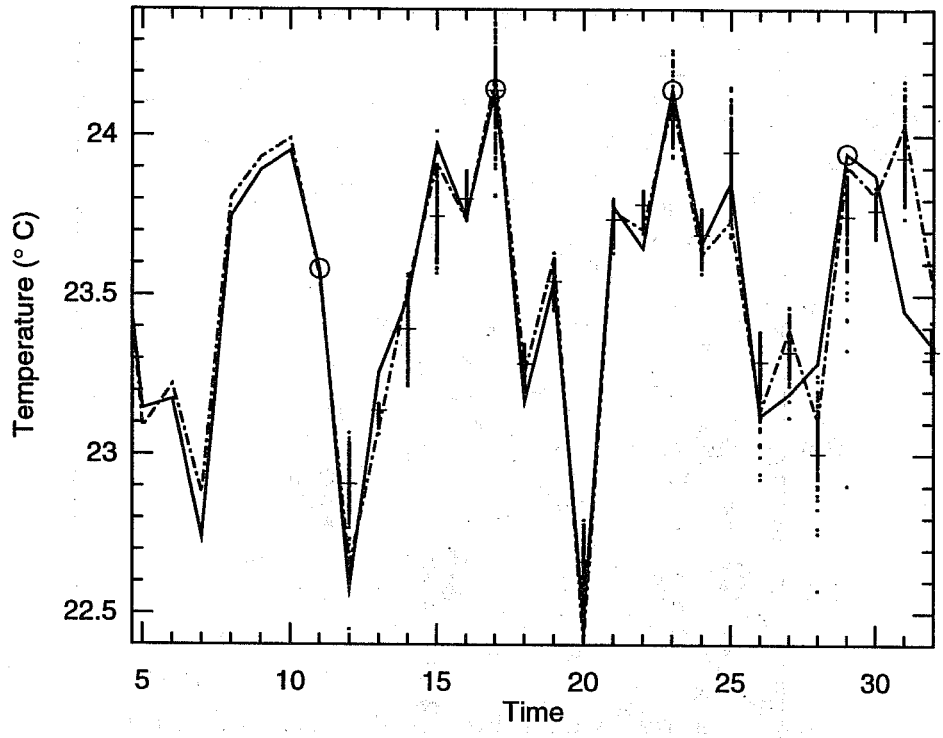

Figure 9. An observed temperature time series (solid line) from the annulus and an ι-shadowing trajectory (dot-dashed line) from a RBF model which stays 'close' to (within 0.2°C of) the observed trajectory for 26 time steps. Four 6–step ensemble predictions are shown: an ensemble of 128 points, normally distributed within 0.075°C of (each component of) the initial observation, is initiated at times 11, 17, 23 and 29 (circles) and iterated under the model to give a distribution after 1,2,3,4,5 and 6 steps (dots), the mean of which is denoted (+).

The perfect ensemble can be formulated in a perfect model scenario with an exact understanding of the observational uncertainty and knowledge of the invariant manifold of the system. It is the easiest to interpret, since the PDF of the ensemble is an accountable representation of the true system PDF; the only error is that due to sampling uncertainty. In physical systems, where models are imperfect, perfect ensembles are not realizable, and long return times may make the formation of perfect ensembles impractical.

Combined with restrictions on the size of ensemble which can be afforded operationally, ensembles are often no longer formulated to give an approximation to the true PDF of the physical system, but may be constrained to subspaces with aims of capturing the spread. While dynamically constrained ensembles can be consistently used to increase the spread, this is usually done at the cost of knowing the relative probability to assign these faster growing initial conditions. Ideally, the CV computed about an analysis value should be a good approximation to the projection of CV (which may or may not exist) about the related system value. An additional assumption made in formulating SV ensembles is that the linearized dynamics are a good approximation to the full nonlinear model dynamics out to the optimization time for the magnitude of perturbation employed in the ensemble. This assumption is fundamental to the use of the SV subspace, as if it holds then then the SV are the directions which will have grown the most over $[0, t_{opt}]$.

The operational use of twin (equal and opposite) pairs in ensembles enables this linearity assumption to be tested. Results from the annulus show that the onset of the breakdown of linearity varies with the magnitude of the perturbation. Results from operational ECMWF SV ensemble 500 mb heights show that linearity is not a good approximation for the majority of cases at the operational optimization time of 2 days. Similar timing in the loss of linearity is found for the NCEP BV 500 mb heights; while the motivation of BV ensembles does not explicitly depend on linearity providing

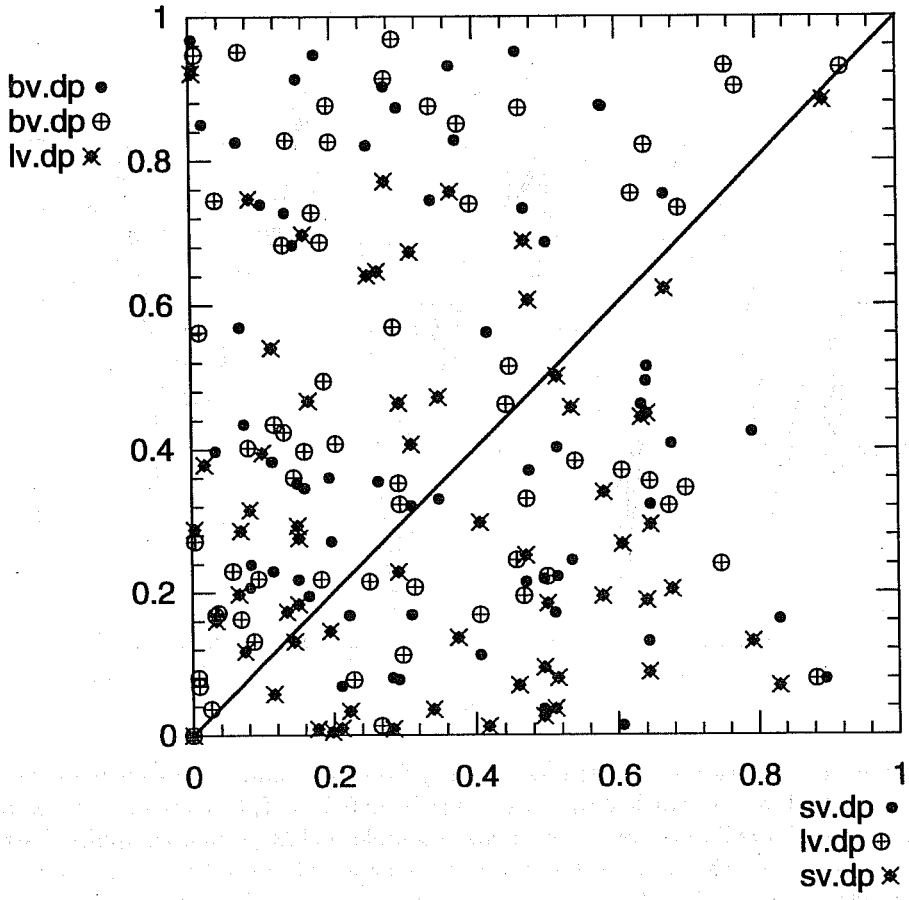

Figure 10. Scatterplot of the relative projections of all pairs of (SV, LV, BV) onto the dream perturbation: BV vs. SV (·), BV vs. LV (circle plus) and LV vs. SV (diamond cross). The line $y = x$ is also shown.

a good approximation, the operational definition of BV is sensitive to nonlinearities. Further work is needed to determine the usefulness of NV ensembles as an alternative to SV ensembles, and to investigate the effect of nonlinearities on the various possible definitions of BV.

ι-shadowing combines observations with the model dynamics to find an initial condition from which the model trajectory agrees with the observations (to within observational uncertainty) for the longest time; such an initial condition may be considered to be the optimal initial condition and the best approximation to the model representation of the true system state. Results of how well the different CV project onto the DP (the vector from the observation to such an optimal initial condition) are shown for the annulus; further work is needed to determine if the DP reflects the need to restrict perturbations to be "near the attractor".

In the perfect model scenario, it is natural to focus on ensembles over different initial conditions. In practice, additional uncertainties may also be accounted for (see *Draper* 1995). Forecasts of physical systems will be effected not only by uncertainty in the initial condition, but also by uncertainty in (*i*) model structure, (*ii*) model parameters (within a given structure), (*iii*) external forcing and (*iv*) uncertainty in the verification conditions. Experience with the annulus shows ensembles over small variations in parameters has thus far been of limited use; in this case large model errors tend to be highly correlated between models. This may reflect shortcomings in model structure which cannot be accounted for by small changes in parameter values. Alternatively, ensembles over several models with different structure significantly reduces the correlation between forecast errors of the models. The question of whether better ensemble forecasts can be obtained by somehow combining these

models or by maintaining an explicit ensemble over different models remains open. Ensembles over different models provide immediate access to state-dependent model error which appears to be of significant value.

ACKNOWLEDGEMENTS

The results reported in this paper could not have been obtained without the assistance of T. Palmer and Z. Toth in discussion of the operational ensembles and in obtaining the data. We have also benefited from discussions with J. Anderson, D. Broomhead, R. Buizza, J. Hansen, M. Harrison and C. Ziehmann. We are grateful for the assistance of W. Ebisuzaki in processing the NWP forecasts.

References

1. R. Buizza. Optimal perturbation time evolution and sensitivity of ensemble prediction to perturbation amplitude. *Q.J.R. Meteorol. Soc.*, 121:1705–1738, 1995.
2. D. Draper. Assessment and propagation of model uncertainty (with discussion). *J. Roy. Stat. Soc. B*, 57:45–97, 1995.
3. I. Gilmour and L.A. Smith. Enlightenment in shadows. In J.B. Kadtke and A. Bulsara, editors, *Applied nonlinear dynamics and stochastic systems near the millenium*, pages 335–340, New York, 1997. AIP.
4. E. N. Lorenz. Deterministic nonperiodic flow. *J. Atmos. Sci.*, 20:130–141, 1963.
5. F. Molteni, R. Buizza, T.N. Palmer, and T. Petroliagis. The ECMWF ensemble prediction system: methodology and validation. *Q. J. R. Meteorol. Soc.*, 122:73–120, 1996.
6. C. Pires, R. Vautard, and O. Talagrand. On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus A*, 48(1):96–121, 1996.
7. P. Read, M. J. Bell, D. W. Johnson, and R. M. Small. Quasi-periodic and chaotic flow regimes in a thermally driven, rotating fluid annulus. *J. Fluid Mech.*, 238:599–632, 1992.
8. L. A. Smith. Identification and prediction of low-dimensional dynamics. *Physica D*, 58:50–76, 1992.
9. L. A. Smith. The Maintenance of Uncertainty. In G. Cini, editor, *Nonlinearity in Geophysics and Astrophysics*, volume CXXXIII of *International School of Physics "Enrico Fermi"*, pages 177–246, Amsterdam, 1997. IOS Press.
10. L.A. Smith. Local optimal prediction: Exploiting strangeness and the variation of sensitivity to initial condition. *Phil Trans R Soc Lond A*, 348(1688):371–381, 1994.
11. L.A. Smith. Accountability and Error in Ensemble Forecasting. In *Predictability*, ECMWF Seminar Proceedings, pages 351–369, Available from ECMWF, Shinfield Park, Reading RG2 9AX, England, 1996. ECMWF.
12. O. Talagrand and P. Courtier. Variational assimilation of meteorological observations with the adjoint vorticity equation - part I. *Q.J.R. Meteorol. Soc.*, 113:1311–1328, 1987.
13. Z. Toth and E. Kalnay. Ensemble forecasting at the NMC: The generation of perturbations. *Bull. Amer. Meteorol. Soc.*, 74(12):2317–2330, 1993.