Journal of Statistical Physics, Vol. 101, Nos. 1/2, 2000

Most Probable Histories for Nonlinear Dynamics: Tracking Climate Transitions

G. L. Eyink[1] **and J. M. Restrepo**[1]

Received November 10, 1999; final June 26, 2000

The *effective action* provides an appropriate cost function to determine most probable (or optimal) histories for nonlinear dynamics with strong noise. In such strong-coupling problems, a nonperturbative technique is required to calculate the effective action. We have proposed a *Rayleigh–Ritz variational approximation*, which employs simple moment-closures or intuitive guesses of the statistics to calculate the effective action. We consider here an application to climate dynamics, within a simple "bimodal" Langevin model similar to that proposed by C. Nicolis and G. Nicolis [*Tellus* **33**:225 (1981)]. Capturing climate state transitions even in this simple model is known to present a serious problem for standard methods of data assimilation. In contrast, it is shown that the effective action for the climate history is already well-approximated by a one-moment closure and that the optimal, minimizing history robustly tracks climate change, even with large observation errors. Furthermore, the Hessian of the effective action provides the ensemble variance as a realistic measure of confidence level in the predicted optimal history.

KEY WORDS: Optimal; estimation; Onsager–Machlup; effective action.

1. THE PROBLEM AND THE MODEL

It is a truism of statistical mechanics that most-probable states are obtained as the extremizers of suitable thermodynamic functions. For example, according to the Boltzmann principle, the most probable state of a system in thermodynamic equilibrium is that which maximizes the entropy. Onsager showed that there are similar variational principles as well to determine *most-probable histories*, or time-sequences of states.(1) Thus, in thermodynamic equilibrium, the most probable sequence of fluctuations is

[1] Department of Mathematics and Program in Applied Mathematics, University of Arizona, Tucson, Arizona 85721.

0022-4715/00/1000-0459$18.00/0

the one that minimizes the Onsager–Machlup action, which represents a time-integrated dissipation or entropy-production.[2] However, such variational principles are generally only well-developed and understood for problems with weak noise and thus small probability of large fluctuation events. This is the regime studied in the classic Kramers transition-rate theory of chemical kinetics, where steepest-descent methods may be applied in the limit of vanishing noise. For reviews of the weak-noise theory, see refs. 3–5.

There are many problems, however, where large fluctuations regularly occur and weak noise asymptotics is not relevant. Consider as an example the following nonlinear Langevin model

$$\dot{X}(t) = f(X(t)) + \kappa\eta(t) \tag{1.1}$$

where $\eta(t)$ is white-noise, $\langle \eta(t)\,\eta(t')\rangle = \delta(t-t')$, and $f(x) = -U'(x)$, where $U(x)$ is the double-well potential

$$U(x) = -2x^2 + x^4 \tag{1.2}$$

with minima at $x = \pm 1$. Similar models were proposed by Nicolis and Nicolis[6] as simple climate models exhibiting bimodality (see also refs. 7–10). In their model $X(t)$ represented the average surface temperature of the earth at time t. The function $U(x)$, which was called in ref. 6 the "climate potential," gives the steady-state probability distribution $P_s(x) \propto \exp(-2U(x)/\kappa^2)$ of the event $X(t) = x$ at sufficiently large times t. In the limit $\kappa \to 0$, its minima represent the overwhelmingly most probable states of the climate. Thus, the minimum at $x = +1$ can be thought of as the present climate state, while $x = -1$ represents ice ages, for example. When κ is small, the dynamics consists of small fluctuations about the minima in one of the wells with long residence times and, then, very rarely, large fluctuations leading to transitions into the other well. However, on geological time scales transitions occur frequently, and thus the small κ limit is not realistic and of little relevance to long-time climate dynamics. In their study of the above model, Miller *et al.*[11] employed the value $\kappa \approx 0.5$, which is neither very small nor very large. Those authors checked that with this value of the noise strength the Kramers formula gives rather poor estimates of the mean times between transitions.

Nevertheless, there is great practical interest in calculating most probable histories for such strongly nonlinear systems with large fluctuations. One example is the problem of *data assimilation*, in which the state history of an atmospheric or oceanographic model is estimated, based upon a set of observations with some random errors. Data assimilation is currently used in initializing weather forecast models in modern meteorology and it

is increasingly important in physical oceanography and paleoclimatology. See refs. 12 and 13 for reviews. One method that is now frequently employed is the *extended Kalman filter* (EKF). This method, which makes a linear interpolation of current observations and model forecasts weighted according to their estimated reliability, is a generalization to nonlinear dynamics (see ref. 14) of the original Kalman filter. The latter is proved to be the optimal estimator for linear dynamics with additive Gaussian noise. Its extension to nonlinear dynamics has only been proven rigorously to be correct in the weak-noise limit.[15] Another popular approach, which we will call the *least-squares variational* (LSV) method, is based upon minimizing a least-squares cost functional of time-histories $X(t)$ of the state variable x. For example, in the model (1.1) above, the (dynamical part of the) cost function would be

$$I[X] = \frac{1}{2\kappa^2} \int_{t_i}^{t_f} [\dot{X}(t) - f(X(t))]^2 \, dt \tag{1.3}$$

In the weak-noise limit, this is the Onsager–Machlup action functional. Thus, minimizing this functional in the weak-noise limit is proved to select the statistically most probable history, conditioned on available information. If observations r_m of the state x of the system are taken at times t_m, assumed to be imperfect with Gaussian observation errors of variance R_m, for $m = 1,..., M$, then the additional information may be incorporated into another term to the cost function, yielding

$$I_*[X] = \frac{1}{2\kappa^2} \int_{t_i}^{t_f} [\dot{X}(t) - f(X(t))]^2 \, dt + \sum_{m=1}^{M} \frac{1}{2R_m} [X(t_m) - r_m]^2 \tag{1.4}$$

as the final functional to be minimized.

However, as emphasized in ref. 11, both of these standard methods —EKF and LSV—are known to fail frequently when applied to strongly nonlinear dynamics. There may be different causes and manifestations of the failure. Deterministic chaos and exponential instability of trajectories pose one set of difficulties. Another problem has to do with "multimodality," in which the statistics of the system exhibit multiple regimes, with occasional transitions between them occurring either randomly or quasi-periodically. In a simple context, the phenomenon is found in the model (1.1), which shows a bimodality between two climate states, "normal" and "ice age." As discussed in detail in ref. 11 the standard methods such as EKF and LSV fail to track transitions between regimes, unless the observational sampling rate is high or the observations are assumed to be

extremely accurate. For example, in the model (1.1) above, climate transitions were found in the estimated history when observed in the data at times separated by unity and with observation error variance $R = 0.01$, i.e., errors 10% of the mean. However, with observation error variances estimated as $R = 0.04$, or errors 20% of the mean, both EKF and LSV estimators failed to show a transition using the same data at unit intervals exhibiting a clear transition. Associated with this failure was an unrealistic confidence level in the faulty estimate. With data showing the variable $X(t_m)$ in the opposite well as the estimator, the EKF and LSV methods gave as their estimated error variances only ≈ 0.016, or estimated error 12.5% of the mean. The overconfidence in reliability of these estimators and their sensitivity to the precise size of the observation error and the frequency of observations is an important limitation in their practical application. Uncertainties inherent in meteorological and oceanographic observations are often hard to assess accurately. Additionally, the data may be sparse or poorly distributed. Yet we see that with even a change from 10% to 20% estimated error in the observations, the standard assimilation methods may either indicate or not indicate the presence of a transition in the past climate.

In our judgement, these failures are due to the use of estimation methods without any statistical justification for strongly nonlinear, noisy systems. Recently, one of us has advocated the use of the so-called *effective action* as the proper cost function for optimal estimation of nonlinear dynamics.[16] This functional is a central object in the ensemble theory of fluctuations for empirical N-sample averages. It thus has a statistical justification for use in data assimilation of chaotic or stochastic nonlinear dynamics. This paper is a preliminary report on work of Eyink and Restrepo[17] validating the method on the two models advanced in ref. 11 as crucial tests of proposed assimilation schemes. One of these test models is the chaotic 3-mode dynamical system of Lorenz.[18] The other is the bimodal climate model of Nicolis and Nicolis,[6] in the precise form of equation (1.1). Here we shall present some of the essential results for the model (1.1), deferring full details of this work end treatment of the Lorenz model to the forthcoming paper.[17]

2. THE EFFECTIVE ACTION AND RAYLEIGH–RITZ METHOD

We must first give a quick summary of the effective action, its basic definition, properties, and calculational schemes. Because these subjects have been discussed in detail in earlier publications,[21–24] here we shall be brief. What we call the *effective action* is the Cramér large-deviation rate function[25] (see also ref. 26, Section 8.6.4) for empirical N-sample averages

of time-series. Thus, in the context of the model (1.1), the effective action $\Gamma[X]$ is a functional of histories $\{X(t): t \in [t_i, t_f]\}$ of the model which determines their "cost" to occur as spontaneous fluctuations of the N-sample average:

$$\bar{X}_N(t) := \frac{1}{N} \sum_{n=1}^{N} X_n(t) \tag{2.1}$$

Here, $X_n(t)$, $n = 1,..., N$ are N independent sample outcomes of the system (1.1). More precisely, the probability for the event $\{\bar{X}_N(t) \approx X(t): t \in [t_i, t_f]\}$ to occur in the limit as $N \to \infty$ is exponentially small, given as

$$P(\bar{X}_N(t) \approx X(t): t \in [t_i, t_f]) \sim \exp(-N \cdot \Gamma[X]) \tag{2.2}$$

The functional $\Gamma[X]$ is a nonnegative, convex functional of the histories $X(t)$. The most important property for our purposes is that its minimizer $X_*(t)$ subject to a set of constraints $\mathscr{R}$ is the most probable value of the N-sample average in the subensemble $\langle \cdot \rangle_*$ conditioned upon the event that $\mathscr{R}$ occur. For example, the absolute minimizer $\bar{X}(t)$ with minimum $\Gamma[\bar{X}] = 0$ is the unconditional ensemble average $\bar{X}(t) := \langle X(t) \rangle$. It is generally accepted that the optimal solution of the estimation problem for nonlinear dynamics is provided by such conditional statistics, e.g., see refs. 19 and 20. Thus, the effective action has a statistical basis to be used as a cost function for estimation of the state history $X(t)$. Furthermore, the effective action is a generating functional for (irreducible) multitime correlations of X. In particular, the two-time correlation $C_*(t, t') := \langle X(t) X(t') \rangle_*$, $-\langle X(t) \rangle_* \langle X(t') \rangle_*$ is given by the inverse Hessian of the effective action:

$$C_*(t, t') = \left(\frac{\delta^2 \Gamma}{\delta X(t) \, \delta X(t')} [X_*] \right)^{-1} \tag{2.3}$$

The diagonal $C_*(t, t)$ gives the ensemble variance $\mathrm{Var}_*(t) := \langle [X(t) - X_*(t)]^2 \rangle_*$.

There is no closed form expression for $\Gamma[X]$ in general. However, there are various schemes by which it may be calculated. It was shown in ref. 21 that there is a variational characterization which is a useful basis for calculation. Let $\mathscr{L}$ be the Liouville or Fokker–Planck operator appearing in the (forward) evolution equation $\partial_t P(x, t) = \mathscr{L} P(x, t)$ for the probability density $P(x, t)$ of the event $X(t) = x$. For example, in the model (1.1),

$$\partial_t P(x, t) = \mathscr{L} P(x, t) := -\frac{\partial}{\partial x} [f(x) \, P(x, t)] + \frac{1}{2} \kappa^2 \frac{\partial^2}{\partial x^2} P(x, t) \tag{2.4}$$

Then, it is shown in ref. 21 that $\Gamma[X]$ is obtained by a constrained variation

$$\Gamma[X] = \text{st.pt.}_{A,\, P\in\mathscr{C}_X} \int_{t_i}^{t_f} dt \int dx\, A(x,t)\cdot(\partial_t - \mathscr{L})\, P(x,t) \tag{2.5}$$

Here the variation is over both normalizable probability distributions $P(t)\in L^1$ and over bounded observables $A(t)\in L^\infty$. This variation is subject to the set of constraints $\mathscr{C}_X$: unit normalization $\int A(x,t)\, P(x,t)\, dx = 1$ and fixed expectation $\int x\cdot A(x,t)\, P(x,t)\, dx = X(t)$ both for $t\in[t_i, t_f]$. Here, $X(t)$ is the particular history at which the value of $\Gamma[X]$ is desired. This variational problem may be solved by the use of Lagrange multipliers $h(t)$ to implement the second constraint. As in ref. 24, this leads to a pair of Euler–Lagrange equations, the forward equation solved with initial data $P(t_i) = P_0$:

$$\partial_t P(x,t) = \mathscr{L}P(x,t) + h(t)[x - \langle x\rangle_t]\, P(x,t) \tag{2.6}$$

and the backward equation for the adjoint operator $\mathscr{L}^*$ solved with final data $A(t_f)\equiv 1$:

$$\partial_t A(x,t) + \mathscr{L}^* A(x,t) + h(t)[x - \langle x\rangle_t]\, A(x,t) = 0 \tag{2.7}$$

Here $\langle x\rangle_t := \int xP(x,t)\, dx$ and $h(t)$ appears as a "control field" used to steer the system of equations so that constraint $\int x\cdot A(x,t)\, P(x,t)\, dx = X(t)$ is satisfied by the solution. The equations (2.6), (2.7) are a practical means to calculate the effective action $\Gamma[X]$ for simple low-dimensional dynamical systems. Such results for the model (1.1) will be presented in the longer work.[17]

However, naive application of the above prescription to large-scale, spatially-extended systems with many degrees of freedom becomes computationally intractable. A variational method based on (2.5), known as the *Rayleigh–Ritz approximation scheme*,[21–24] addresses this difficulty. In such a scheme, finitely-parameterized guesses $\tilde{P}(x,t;\boldsymbol{\mu})$ and $\tilde{A}(x,t;\boldsymbol{\alpha},\boldsymbol{\mu})$ are inserted into the action functional as trial states, and variation effected over the $2p$ parameters α_i, μ_i, $i = 1,..., p$. It is convenient to take the left trial function as a linear superposition of p moment-functions $\psi_i(x)$:

$$\tilde{A}(x,t;\boldsymbol{\alpha},\boldsymbol{\mu}) = 1 + \sum_{i=1}^{p} \alpha_i(t)[\psi_i(x) - \mu_i(t)] \tag{2.8}$$

where

$$\int \psi_i(x)\, \tilde{P}(x,t;\boldsymbol{\mu})\, dx = \mu_i(t), \qquad i = 1,..., p \tag{2.9}$$

are the moment-averages in the guess $\tilde{P}(x, t; \boldsymbol{\mu})$ for the probability density. These trial functions yield a reduced form of the action functional

$$\tilde{\Gamma}[\boldsymbol{\alpha}, \boldsymbol{\mu}] := \int_{t_i}^{t_f} \boldsymbol{\alpha}^{\top}(t)[\dot{\boldsymbol{\mu}}(t) - \mathbf{V}(t))]\, dt \tag{2.10}$$

where $V_i(\boldsymbol{\mu}) := \langle \mathscr{L}^* \psi_i \rangle_{\boldsymbol{\mu}}$, with $\langle \cdot \rangle_{\boldsymbol{\mu}} := \int (\cdot)\, P(x, t; \boldsymbol{\mu})\, dx$. An unconstrained variation over the parameters yields the *moment-closure equations* $\dot{\boldsymbol{\mu}} = \mathbf{V}(\boldsymbol{\mu})$ and an associated set of adjoint equations for $\boldsymbol{\alpha}(t)$. The equations (2.8), (2.9) automatically implement the normalization constraint. The expectation constraint becomes

$$\xi(\boldsymbol{\mu}(t)) + \boldsymbol{\alpha}^{\top}(t)\, \mathbf{C}(\boldsymbol{\mu}(t)) = X(t) \tag{2.11}$$

where $\xi(\boldsymbol{\mu}) := \langle x \rangle_{\boldsymbol{\mu}}$ and $C_i(\boldsymbol{\mu}) := \langle x \psi_i(x) \rangle_{\boldsymbol{\mu}} - \langle x \rangle_{\boldsymbol{\mu}} \langle \psi_i(x) \rangle_{\boldsymbol{\mu}}$, $i = 1, \ldots, p$. This constraint may be implemented by the Lagrange multiplier $h(t)$. The reduced Euler–Lagrange equations are the forward equation solved with initial data $\boldsymbol{\mu}(t_i) = \boldsymbol{\mu}_0$:

$$\dot{\boldsymbol{\mu}} = \mathbf{V}(\boldsymbol{\mu}) + h(t)\, \mathbf{C}(\boldsymbol{\mu}) := \mathbf{V}_X(\boldsymbol{\mu}, h(t)) \tag{2.12}$$

and the backward equation solved with the final condition $\boldsymbol{\alpha}(t_f) = \mathbf{0}$:

$$\dot{\boldsymbol{\alpha}} + \left(\frac{\partial \mathbf{V}_X}{\partial \boldsymbol{\mu}}\right)^{\top} (\boldsymbol{\mu}, h(t))\, \boldsymbol{\alpha} + \left(\frac{\partial \xi}{\partial \boldsymbol{\mu}}\right)^{\top} (\boldsymbol{\mu})\, h(t) = 0 \tag{2.13}$$

Substituting the solutions $\boldsymbol{\alpha}(t)$, $\boldsymbol{\mu}(t)$ into the reduced action (2.10) yields the Rayleigh–Ritz approximation $\tilde{\Gamma}[X]$ to the effective action. In this way, moment-closure methods that have traditionally been used in the calculation of ensemble averages may also be employed in the approximate evaluation of optimal histories. The ability of the Rayleigh–Ritz scheme to predict high Reynolds number turbulent decay has already been tested, using standard $K-\varepsilon$ closure methods.[27] In our longer work[17] closure ideas developed by Nicolis and collaborators for the chaotic Lorenz model[28, 29] will be applied to the data assimilation problem in that model.

It is illuminating to compare our proposed estimation method, based upon the Rayleigh–Ritz approximation of the effective action, with another recently developed approximate estimation scheme, the Ensemble Kalman Filter.[30–32] The two are closely related conceptually, since they both attempt to calculate the statistics in ensembles conditioned upon observations.

However, they differ in the approximations employed.[2] There are two major approximations in each method. In our method, the first approximation is upon the conditioning set $\mathscr{R}$, which involves conditions upon the N-sample average (2.1) rather than upon individual sample elements. For further discussions of this point, see ref. 16. It is shown there that the minimizer of the effective action gives the conditional statistics in a larger ensemble than the correctly conditioned ensemble and is thus suboptimal in a precise sense. In ref. 16 this property of the variational estimator is called "mean-optimality." However, it is definitely an approximation whose effect, while hopefully small, is essentially uncontrolled. The second approximation in our scheme is the Rayleigh–Ritz method itself, that is, the use of moment-closure equations to evolve the statistics in time rather than by means of the exact Kolmorogov or Liouville equations for the probability distributions, as in (2.6), (2.7). The two approximations used in the Ensemble Kalman Filter are quite different. First, as in the Kalman filtering method generally, it is assumed that the estimate at measurement times can be obtained by making a linear interpolation between the measured and forecast values, the so-called "analysis" step. The appropriate weighting is determined from a "Kalman gain matrix," which is calculated from model forecast and measurement error covariances. However, the optimal nonlinear methods[19, 20] do not lead in general to such a linear interpolation scheme which is, again, an approximation of unknown validity. The second approximation in the ensemble method is the use of an actual emsemble of N samples evolving under the nonlinear dynamics in order to propagate statistics in time between measurements. In principle, this approximation is systematic, converging as number of samples N is increased toward infinity. However, Monte Carlo errors vanish only proportionately to $1/\sqrt{N}$ and, thus, achieving even a 10% accuracy in statistical evolution would require no less than $N = 100$ samples. It will generally not be practical in real applications to deal with ensembles of more than a hundred or so samples. Thus, another essential approximation is made here. It is hoped that, because our variational scheme and the Ensemble Kalman Filter scheme attempt to calculate the same conditional statistics, but with quite different approximations, that the two methods will prove complementary. Each of them can be used to validate (or to call into question) the results of the other.

[2] It must be emphasized that some approximation is certainly necessary. For the same reasons which make (2.6), (2.7) impractical to solve for spatially extended systems, the fully optimal methodes of refs. 19 and 20 also cannot be applied to such large-scale "distributed" systems: see ref. 33 and also ref. 11 for discussions of this point.

3. RESULTS OF A ONE-MOMENT CLOSURE

In this work we will just illustrate our method using a simple one-moment closure for the bimodal Langevin model (1.1). It is natural to take as the moment function $\psi_1(x) = x$, the state variable itself. As a guess for the statistics we will take the model PDF

$$P(x;\mu) = \left(\frac{1-\mu}{2}\right)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x+1)^2}{2\sigma^2}\right] + \left(\frac{1+\mu}{2}\right)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x-1)^2}{2\sigma^2}\right], \qquad |\mu|<1 \tag{3.1}$$

and

$$P(x;\mu) = \frac{1}{\sqrt{\pi\sigma^2}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \qquad |\mu|>1 \tag{3.2}$$

It is easy to check that $\int xP(x;\mu)\,dx = \mu$. The value $\sigma^2 = \kappa^2/16$ is chosen so that, when $\mu = 0$, the guess reduces to the steepest descent approximation of the stationary distribution $P_s(x)$, as considered in ref. 6. This consists of a sum of two Gaussian distributions centered at ± 1 with standard deviations σ. It gives a reasonable representation of the exact stationary distribution, but it is more convenient for calculation. As μ moves away from 0 toward one of the minima ± 1, the corresponding peak grows while the opposite one diminishes. When $\mu = \pm 1$, only the one peak remains and the opposite has completely disappeared. Then as μ moves beyond the minimum to $|\mu| > 1$, the single Gaussian peak is simply shifted to the new center μ with the same height and variance. This gives a reasonable model of the steady-state statistical distribution of $X(t)$ at large times t conditioned upon the event $\bar{X}_N(t) = \mu$.

It is easy to derive from this statistical model the moment-closure equation

$$\dot{\mu} = -12\sigma^2\mu - 4\mu\max\{\mu^2 - 1, 0\} := V(\mu) \tag{3.3}$$

It is important to observe that the origin 0, which is unstable for the deterministic dynamics $\dot{X}(t) = f(X(t))$, is now the global stable fixed point of the closure dynamics (3.3). This corresponds to the stability of the Fokker–Planck equation (2.4), which, in the absence of any additional information from observations, implies a relaxation of the time-dependent distribution $P(x, t)$ to the double-peaked stationary distribution $P_s(x)$. In the steady-state there is complete loss of information as to which regime, + or −, the

system is in and the mean is $\mu = 0$. This *statistical stability* inherent to chaotic or stochastic nonlinear dynamics is an important ingredient of the present method. The other elements of the closure in addition to $V(\mu)$ required for the calculation of the effective action are $\xi(\mu) = \mu$, and the covariance

$$C(\mu) = \sigma^2 + \max\{1 - \mu^2, 0\}. \tag{3.4}$$

These are substituted into (2.12), (2.13) to yield the forward and backward equations that are numerically integrated to obtain the approximate effective action $\tilde{\Gamma}[x]$ and its gradient $\tilde{h}[t; X] = (\delta\tilde{\Gamma}/\delta X(t))[X]$. It is convenient to note that, by Legendre duality, the latter coincides with the control field or Lagrange multiplier $h(t)$ that appears in (2.12), (2.13). Hence it is easy to carry out a conjugate-gradient or variable-metric minimization of the full cost function

$$\tilde{\Gamma}_*[X] = \tilde{\Gamma}[X] + \frac{1}{2R}\sum_{m=1}^{M}[X(t_m) - r_m]^2 \tag{3.5}$$

The latter is the joint effective action $\tilde{\Gamma}[X, r]$ of the state-history $X(t)$ and the measurements r_m, $m = 1,..., M$, if the latter are assumed to have Gaussian errors with common variances R. We have carried out this minimization for the model problem with $\kappa = 0.5$, obtaining an optimal history $X_*(t)$. We were particularly interested to test the ability of this estimator to track climate transitions observed in an empirical dataset. Here we shall illustrate this with an artifical dataset of $M = 10$ observations taken at unit times $t_m = m$, $m = 1,..., 10$. This is the same spacing of measurements as in ref. 11. As a simple example with a transition we take $r_m = -1$ for $m = 1,..., 5$ and $r_m = +1$ for $m = 6,..., 10$. We consider the four choices of observation error variance $R = 0.04, 0.16, 0.36$ and 0.64. In addition to the optimal history $X_*(t)$, we also calculate the ensemble variance $\mathrm{Var}_*(t) = \langle (X(t) - X_*(t))^2 \rangle_*$ by a finite-difference approximation to the Hessian of the effective action. This gives the intrinsic dispersion that would be observed in an N-sample ensemble experiment and it provides a reasonable measure of the confidence level in the predicted optimal history. It is a consequence of the large fluctuations naturally occurring in this system that the variance $\mathrm{Var}_*(t)$ need not be small. The details of our calculations and numerical methods, along with a treatment of a more realistic dataset and a more complete discussion of the results, can be found in ref. 17.

For the above artifical dataset the results are shown in Fig. 1(a)–(d), which plot both the optimal history $X_*(t)$ and the ensemble deviation $\sigma_*(t) := \sqrt{\mathrm{Var}_*(t)}$ at each of the values $R = 0.04$–0.64. It may be seen that,

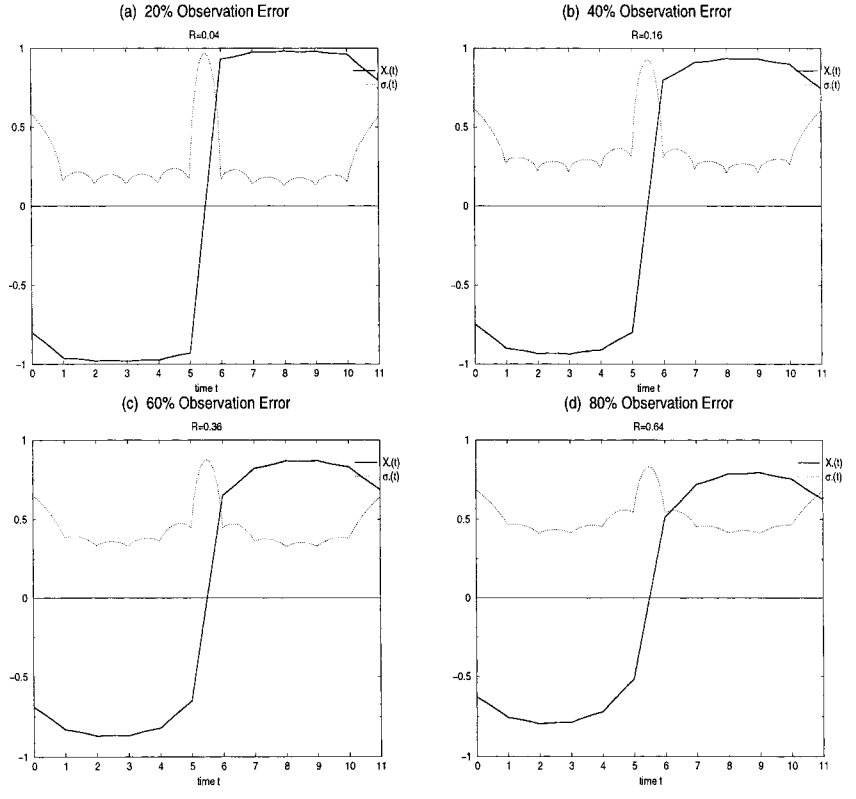

Fig. 1. Optimal histories and standard deviations. Graphed are the mean $X_*(t)$ and standard deviation $\sigma_*(t)$ in the ensembles conditioned on the imperfect observations, as approximated by the one-moment closure. Gaussian observation errors were assumed with variances (a) $R = 0.04$, (b) $R = 0.16$, (c) $R = 0.36$, and (d) $R = 0.64$.

already for the value $R = 0.04$ at which the EKF and LSV assimilation methods fail to track the transition, the optimal estimator from the effective action clearly follows the climate change indicated in the data. More impressively, the transition persists with very little qualitative change in the estimate $X_*(t)$ as R is increased up to the highest value $R = 0.64$. The transition continues to be followed, but with a quantitative reduction in amplitude. This robustness of the estimator to different confidence levels assumed for the reliability of measurements is, obviously, a very attractive feature. It may furthermore be observed that the ensemble standard deviations obtained from the effective action are closely correlated during the measurement interval (1, 10) to the size $\approx \sqrt{R}$ of the observation errors, except at the time of the transition itself near $t = 5.5$. At that time $\sigma_*(t)$

grows to near 1, reflecting 100% uncertainty in the precise state of the system as it goes through the transition. Before and after the measurement interval the ensemble variance $\sigma_*(t)$ also grows, reflecting the intrinsic loss of predictability of the system as time goes forward (and backward) from the measurements.

The scrupulous honesty of the method in admitting the inherent impossibility of predicting with precision in such circumstances is, in fact, the primary reason it succeeds to track the transition. On the contrary, the EKF and LSV methods follow the deterministic dynamics $\dot{X}(t) = f(X(t))$ between measurements into the wells at stable fixed points $X = \pm 1$. There they show a single-peaked Gaussian distribution of X with variance $\sigma^2 = 1/64$. Unless the measurement error variance R is smaller than this value, the standard EKF and LSV methods will fail to track the transition to the other well even when it is observed in the data. In the effective action method the statistics evolve naturally in the absence of data to give estimate $X_* = 0$, with a corresponding growth of the error variance to $\sigma_*^2 \approx 1$ or 100%. The estimator is then quite ready to accept information from measurements, even those with rather large errors.

We have thus demonstrated that, in this simple example, the proposed variational method succeeds where EKF and LSV fail. Of course, some modifications of the latter methods have been shown in ref. 11 to succeed as well (for the harder 3-mode Lorenz model rather than the double-well system). Thus, it is worth comparing our approach briefly with those alternatives. In ref. 11 a "higher-order EKF" method was tested, in which the forecast error covariance equation was evolved within a moment-truncation closure including moments up to 4th-order. While this worked well for the 3-variable system, the authors noted themselves that it would be impractical for more realistic geophysical models. In such more realistic applications, even 2nd-order closure schemes are not very practical. It is therefore very important that the method developed by us here can succeed using only a *first-order* moment-closure. Another modification considered in ref. 11 was a "stochastic approximation," in which the model noise term was replaced by an empirical noise model representing effects of the nonlinear chaotic dynamics. In realistic applications, such a term would arise as an "eddy noise" induced, for example, by subgrid-scale turbulence. As discussed in ref. 22, the method that we have employed here based upon the effective action takes into account such chaos-induced stochasticity. In fact, the Rayleigh–Ritz effective action—for small departures from the absolute minimum—reduces just to an Onsager–Machlup functional for the linearized *closure* dynamics (not the microscopic dynamics) and with model noise replaced by an effective, nonlinearly-generated noise. A main difference from the "stochastic approximation" considered by ref. 11 is that

the effective noise is calculated within the Rayleigh–Ritz approximation from the dynamics and not modelled empirically.

ACKNOWLEDGMENTS

We thank the organizers for their invitation to participate in the Brussels meeting, "Nonlinear Science: Dynamics and Stochasticity," in honor of the 60th birthday of Grégoire Nicolis. The influence of Prof. Nicolis upon our subject is evident and we are pleased to offer this contribution as a testament to his work. G.E. thanks the DOE for partial support through grant LDRD-ER 2000047 at the Los Alamos National Laboratory.

REFERENCES

1. L. Onsager, Reciprocal relations in irreversible processes, I, II, *Phys. Rev.* **37**:405–426, 2265–2279 (1931).
2. L. Onsager and S. Machlup, Fluctuations and irreversible processes, *Phys. Rev.* **91**: 1505–1512 (1953).
3. M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems* (Springer, New York, 1984).
4. R. Graham, Onset of cooperative behavior in nonequilibrium steady states, in *Order and Fluctuations in Equilibrium and Nonequilibrium Statistical Mechanics*, G. Nicolis, G. Dewel, and J. W. Turner, eds. (Wiley, New York, 1981).
5. R. Maier and D. Stein, A scaling theory of bifurcations in the symmetric weak-noise escape problem, *J. Stat. Phys.* **83**:291–357 (1996).
6. C. Nicolis and G. Nicolis, Stochastic aspects of climatic transitions—additive fluctuations, *Tellus* **33**:225–234 (1981).
7. R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani, Stochastic resonance in climatic change, *Tellus* **34**:10–16 (1982).
8. R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani, A theory of stochastic resonance in climatic change, *SIAM J. Appl. Math.* **43**:565–578 (1983).
9. A. Sutera, Stochatic perturbation of a pure convective motion, *J. Atmos. Sci.* **37**:246–249 (1980).
10. A. Sutera, On stochastic perturbation and long-term climate behavior, *Quart. J. Roy. Meteorol. Soc.* **107**:137–152 (1981).
11. R. N. Miller, M. Ghil, and V. Gauthiez, Advanced data assimilation in strongly nonlinear dynamical systems, *J. Atmos. Sci.* **51**:1037–1056 (1994).
12. M. Ghil and P. Malanotte-Rizzoli, Data assimilation in meteorology and oceanography, *Adv. in Geophys.* **33**:141–266 (1991).
13. O. Talagrand, Assimilation of observations, an introduction, *J. Met. Soc. Jap.* **75**:191–209 (1997).
14. A. H. Jazwinski, *Stochastic Processes and Filtering Theory* (Academic Press, New York, 1970).
15. Y. Steinberg, B. Z. Bobrovsky, and Z. Schuss, Fixed-point smoothing of scalar diffusions, I: An asymptotically optimal smoother, *SIAM J. Appl. Math.* **54**:833–853 (1994).
16. G. L. Eyink, A variational formulation of optimal nonlinear estimation, preprint.

17. G. L. Eyink and J. Restrepo, Optimal variational assimilation in strongly nonlinear dynamical systems, to be submitted to *J. Atmos. Sci.*
18. E. N. Lorenz, Deterministic non-periodic flow, *J. Atmos. Sci.* **20**:130–141 (1963).
19. H. J. Kushner, Dynamical equations for optimal nonlinear filtering, *J. Diff. Eq.* **3**:179–190 (1967).
20. E. Pardoux, Non-linear filtering, prediction and smoothing, in *Stochastc Systems: The Mathematics of Filtering and Identification and Applications*, M. Hazelwinkel and J. C. Willems (D. Reidel Publishing, Dordrecht, 1981).
21. G. L. Eyink, Action principle in nonequilibrium statistical dynamics, *Phys. Rev. E* **54**:3419–3435 (1996).
22. G. L. Eyink, Linear stochastic models of nonlinear dynamical systems, *Phys. Rev. E* **58**:6975–6991 (1998).
23. G. L. Eyink, Action principle in statistical dynamics, *Prog. Theor. Phys. Suppl.* **130**:77–86 (1998).
24. G. L. Eyink, Fluctuation-response relations for multi-time correlations, to appear in *Phys. Rev. E* **62**:210–220 (2000).
25. H. Cramér, Sur un noveau theorème-limite de la théorie des probabilités, *Actualités Scientifiques et Industrialles* **736**:5–23 (1938).
26. U. Frisch, *Turbulence. The Legacy of A. N. Kolmogorov* (Cambridge University Press, Cambridge, 1995).
27. G. L. Eyink and A. Wray, Evaluation of the statistical Rayleigh–Ritz method in isotropic turbulence decay, in *Proceedings of the 1998 CTR Summer Program* (Center for Turbulence Research, NASA Ames & Stanford University, 1998). Available at: http://ctr.stanford.edu/SP98.html.
28. C. Nicolis and G. Nicolis, Closing the hierarchy of moment equations in nonlinear dynamical systems, *Phys. Rev. E* **58**:4391–4400 (1998).
29. K. E. Kürten and G. Nicolis, Moment equations and closure schemes in chaotic dynamics, *J. Phys. A: Math. Gen.* **31**:7331–7340 (1998).
30. P. J. van Leeuwen and G. Evensen, Data assimilation and inverse methods in terms of a probabilistic formulation, *Mon. Wea. Rev.* **124**:2898–2912 (1996).
31. G. Evensen, Advanced data assimilation for strongly nonlinear dynamics, *Mon. Wea. Rev.* **125**:1342–1354 (1996).
32. G. Burgers, P. J. van Leeuwen, and G. Evensen, Analysis scheme in the ensemble Kalman filter, *Mon. Wea. Rev.* **126**:1719–1724 (1998).
33. H. J. Kushner, Approximations to optimal nonlinear filters, *IEEE Trans. Auto. Cont.* **AC-12**:546–556 (1967).