Data Assimilation in Slow–Fast Systems Using Homogenized Climate Models

Lewis Mitchell and Georg A. Gottwald

School of Mathematics and Statistics, University of Sydney, Sydney, New South Wales, Australia

ABSTRACT

A deterministic multiscale toy model is studied in which a chaotic fast subsystem triggers rare transitions between slow regimes, akin to weather or climate regimes. Using homogenization techniques, a reduced stochastic parameterization model is derived for the slow dynamics. The reliability of this reduced climate model in reproducing the statistics of the slow dynamics of the full deterministic model for finite values of the time-scale separation is numerically established. The statistics, however, are sensitive to uncertainties in the parameters of the stochastic model.

It is investigated whether the stochastic climate model can be beneficial as a forecast model in an ensemble data assimilation setting, in particular in the realistic setting when observations are only available for the slow variables. The main result is that reduced stochastic models can indeed improve the analysis skill when used as forecast models instead of the perfect full deterministic model. The stochastic climate model is far superior at detecting transitions between regimes. The observation intervals for which skill improvement can be obtained are related to the characteristic time scales involved. The reason why stochastic climate models are capable of producing superior skill in an ensemble setting is the finite ensemble size; ensembles obtained from the perfect deterministic forecast model lack sufficient spread even for moderate ensemble sizes. Stochastic climate models provide a natural way to provide sufficient ensemble spread to detect transitions between regimes. This is corroborated with numerical simulations. The conclusion is that stochastic parameterizations are attractive for data assimilation despite their sensitivity to uncertainties in the parameters.

1. Introduction

An area of broad research in the atmospheric sciences in recent times is how to most effectively parameterize subgrid-scale processes. Since the pioneering papers by Leith (1975) and Hasselmann (1976), the stochastic parameterization of processes that cannot be spatially and/or temporally resolved have recently gained popularity across disciplines in the atmospheric, oceanographic, and climate sciences (Palmer 2001). Applications range from the study of atmospheric low-frequency variability (Franzke et al. 2005; Franzke and Majda 2006), deep convection and cloud modeling in GCMs (Lin and Neelin 2000, 2002; Plant and Craig 2008), and decadal climate changes such as El Niño (Kleeman 2008) to paleoclimatic modeling (Ditlevsen 1999; Kwasniok and Lohmann 2009). We refer the reader to the books edited by Imkeller and von Storch (2001) and by Palmer and Williams (2010), which contain excellent overviews of current trends in stochastic climate modeling.

There exists a plethora of different methods to construct stochastic subgrid-scale parameterizations, including phenomenological approaches such as randomization of existing deterministic parameterization schemes (e.g., Buizza et al. 1999), energetic backscattering (e.g., Frederiksen and Davies 1997; Shutts 2005), data-driven techniques such as Markov chains (e.g., Crommelin and Vanden-Eijnden 2008), and systematic approaches using stochastic homogenization (e.g., Majda et al. 1999, 2003). The specific functional forms and parameter values of the respective parameterization schemes can be heuristically postulated on grounds of the particular physics involved (e.g., Lin and Neelin 2000) or estimated using time series analysis (e.g., Wilks 2005). In the case of multiscale dynamics, however, the parameters can be systematically derived using averaging and homogenization techniques (e.g., Majda et al. 1999, 2001, 2008). Our work will be concerned with the latter approach. The general theory of stochastic averaging and homogenization for multiscale dynamical systems goes back to the seminal papers of Khasminsky (1966), Kurtz (1973), and Papanicolaou

Corresponding author address: Georg A. Gottwald, The University of Sydney, School of Mathematics and Statistics, Sydney 2006, Australia.
E-mail: georg.gottwald@sydney.edu.au

(1976). Starting with Majda et al. (1999), these mathematically rigorous ideas have recently been applied in the atmospheric context by Franzke et al. (2005) and Franzke and Majda (2006).

We will introduce a simple deterministic multiscale toy model in which a slow degree of freedom with multiple stable states, resembling slow weather or climate regimes (e.g., Legras and Ghil 1985; Crommelin 2003; Crommelin et al. 2004; Branstator and Berner 2005; Ditlevsen 1999; Kwasniok and Lohmann 2009), is driven by a fast chaotic subsystem that also involves metastable regimes. Although simple, this model involves core phenomena found in realistic applications such as slow and fast metastable states as caricatures of climate and weather regimes, respectively, and rare transitions between them. The model is amenable to the theory of homogenization [for an excellent review see Givon et al. (2004) and Pavliotis and Stuart (2008)], and we will derive a reduced stochastic climate model describing the slow dynamics of the full higher-dimensional model. Whereas for nonsystematic stochastic parameterizations the results are sensitive to details of the assumed processes, we will verify here that the homogenized reduced model faithfully represents the slow dynamics of the full system even for moderate time-scale separation. We have chosen a system amenable to homogenization precisely for the reason that the issue of the appropriate choice of the stochastic parameterization does not arise and the validity of the parameterization is guaranteed by rigorous theorems (at least in the limit of large time-scale separation).

Our main focus here is how well homogenized stochastic climate models perform when used as forecast models in the context of ensemble data assimilation. In data assimilation one attempts to find an optimal estimate of the state of a system combining information of noisy observations and a numerical forecast model exhibiting model error (Kalnay 2003). This procedure is complicated by the nonlinear chaotic dynamics of the system as well as the impossibility of observing the entire system at any one time. On larger time scales, it is often only possible to adequately observe the slow, large-scale degrees of freedom of the system, while the fast, small-scale degrees of freedom in general remain unobservable. We consider one of the state-of-the-art data assimilation methods called the ensemble Kalman filter (EnKF) (Evensen 1994, 2006). In such a filter one evolves an ensemble of state estimates forward in time using a full nonlinear forecast model, and then estimates the forecast (or background) mean and its associated error variance from the ensemble. Together with observations of the system, this covariance is used in a least squares minimization problem to find an optimal state

estimate called the analysis, along with an estimate of the associated analysis error covariance. This filter is optimal for linear systems with Gaussian errors, assumptions that are generically not consistent with real-world systems. Besides their ease of implementation, the attractive feature of ensemble filters is that the forecast error covariance is estimated using the full nonlinear forecast model.

Ensemble-based Kalman filters, however, suffer from the problem of sampling errors due to the insufficient ensemble size. These errors usually underestimate the error covariances that may ultimately lead to filter divergence, when the filter trusts the forecast and ignores the information given by the observations. To avoid filter divergence the concept of covariance inflation was introduced whereby the prior forecast error covariance is increased by an inflation factor (Anderson and Anderson 1999). This is usually done globally and involves careful and computationally expensive tuning of the inflation factor [for recent methods on adaptive estimation of the inflation factor from the innovation statistics, see Anderson (2007, 2009) and Li et al. (2009)].

We will address here the following questions: Under what circumstances can reduced stochastic climate models improve the skill of an ensemble-based data assimilation scheme if used as forecast models? And if improvement happens, why?

Harlim and Majda (2008, 2010) studied ensemble filtering on the Lorenz-96 model (Lorenz 2006) in the fully turbulent regime, where they found skill improvement for a stochastic climate forecast model that was constructed by radically replacing all nonlinear terms by linear stochastic processes whose statistics are estimated by fitting the model to the climatological variance and the decorrelation time. Here we study the performance of systematically derived climate models; however, our results will also shed light on skill improvements seen in other ensemble-based stochastic filtering methods.

Stochastic climate models have often been found to not reproduce the autocorrelation function of the full deterministic system well, in particular for small time-scale separation [see, e.g., Franzke et al. (2005) and Franzke and Majda (2006)]. Here we will see that this inaccuracy in faithfully reproducing the statistics of the slow dynamics does not preclude stochastic reduced models from being beneficial in data assimilation. We will identify a range of observation intervals in which the reduced stochastic climate model actually outperforms the full deterministic model. The range of observation intervals for which skill improvement is observed will be related to the characteristic time scales of the system. In particular we find that the observation interval has to be

larger than the typical time taken for the slow state to switch regimes, and smaller than the decay rate of the autocorrelation function of the slow variable. Hence, stochastic climate models will be beneficial when studying weather or climate systems on time scales that resolve regime switches.

We find that although climate models with appropriately determined drift and diffusion terms faithfully reproduce the slow dynamics of a multiscale deterministic model, the statistics of the climate model is rather sensitive to uncertainties in the drift and diffusion terms. A central result will be that their superior performance in data assimilation is not linked to their accurate approximation of the statistical slow behavior but rather to a controlled increase of the ensemble spread associated with their inherent dynamic stochasticity, with an effect similar to inflating covariances or increasing ensemble size. Hence optimal performance of stochastic reduced climate models will be achieved when the diffusion coefficient is larger than the value that produces the closest match to the full deterministic dynamics.

The remainder of the paper is organized as follows: In section 2 we introduce the model and derive the reduced stochastic climate model for the slow dynamics using homogenization techniques. Estimation of the drift and diffusion terms in the climate model is performed in section 3. In section 4 we examine some of the relevant characteristic time scales of the dynamics. In section 5 we then study the sensitivity of these time scales to uncertainties in the time-scale separation parameter as well as in the diffusion and the drift term of the climate model. After a brief overview of ensemble Kalman filtering in section 6 we present numerical results showing the skill improvement in using the reduced climate model over the full deterministic model in an ensemble Kalman filter in section 7. We conclude with a discussion in section 8.

2. The model

We consider the following deterministic system with slow–fast time-scale separation proposed in Givon et al. (2004):

$$\frac{dx}{dt} = x - x^3 + \frac{4}{90\varepsilon}y_2, \quad (1)$$

$$\frac{dy_1}{dt} = \frac{10}{\varepsilon^2}(y_2 - y_1), \quad (2)$$

$$\frac{dy_2}{dt} = \frac{1}{\varepsilon^2}(28y_1 - y_2 - y_1y_3), \quad (3)$$

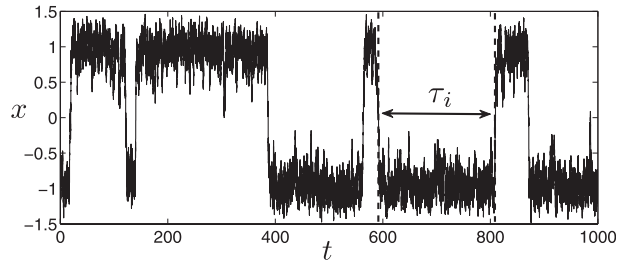$$\frac{dy_3}{dt} = \frac{1}{\varepsilon^2}\left(y_1y_2 - \frac{8}{3}y_3\right). \quad (4)$$

FIG. 1. Sample trajectory of the slow variable x of the full deterministic 4D model (1)–(4). The dynamics appears to randomly switch between two metastable states with randomly distributed sojourn times τ_i.

Here the slow variable x is driven by a fast chaotic Lorenz system (Lorenz 1963). The slow dynamics (1) describes an overdamped degree of freedom in a one-dimensional potential well $V(x) = x^4/4 - x^2/2$ that is being continually "kicked" by the fast chaotic motion of the Lorenz subsystem (2)–(4). In the following we will set $\varepsilon^2 = 0.01$ unless specified otherwise.

In Fig. 1 we show a trajectory of the slow variable x obtained from a simulation of the full deterministic system (1)–(4). The trajectory of the slow variable appears to randomly switch between two metastable states around the minima of the potential $V(x)$ at $x^\star = \pm 1$. In the following section we will derive a reduced stochastic differential equation that describes the effective slow dynamics.

Derivation of the stochastic climate model

The ergodicity of the fast chaotic Lorenz equation (Tucker 1999) and its mixing property (Luzzatto et al. 2005) suggests an application of stochastic model reduction techniques, by which the fast chaotic degrees of freedom are parameterized by a one-dimensional stochastic process. This can be heuristically justified provided the fast processes decorrelate rapidly enough; then the slow variables experience the sum of uncorrelated fast dynamics during one slow time unit. According to the central limit theorem this converges to approximate Gaussian noise in the limit when the time-scale separation becomes infinite. We will formalize this and apply stochastic singular perturbation theory (homogenization) (Khasminsky 1966; Kurtz 1973; Papanicolaou 1976; Givon et al. 2004; Pavliotis and Stuart 2008; Melbourne and Stuart 2011) to deduce the following reduced stochastic 1D climate model for the slow x dynamics:

$$\frac{dX}{dt} = X(1 - X^2) + \sigma\frac{dW}{dt}, \quad (5)$$

with the one-dimensional Wiener process W, where σ is given by the total integral of the autocorrelation function of the fast y_2 variable with

$$\frac{\sigma^2}{2} = \left(\frac{4}{90}\right)^2 \int_0^\infty \left[\lim_{T \to \infty} \frac{1}{T} \int_0^T y_2(s) y_2(t+s)\, ds \right] dt. \quad (6)$$

Rather than studying the system (1)–(4) directly, in stochastic homogenization one considers the associated Fokker–Planck or its adjoint, the backward Kolmogorov equation. Whereas the original ordinary differential equation is nonlinear, the latter ones are linear partial differential equations and can be treated with standard perturbation techniques, expanding in the small parameter ε and studying solvability conditions of the linear equations at the respective orders of ε. The solvability conditions are given by Fredholm alternatives and can be evaluated using the ergodicity of the fast process.

We will analyze the system (1)–(4) in the framework of the backward Kolmogorov equation for the conditional expectation value of some sufficiently smooth observable $\phi(x, y)$ defined as

$$v(x_0, y_0, t) = \mathbb{E}\{\phi[x(t), y(t)] \,|\, x(0) = x_0, y(0) = y_0\}$$

with $y = (y_1, y_2, y_3)$. Here the expectation value is taken with respect to the ergodic measure induced by the fast dynamics of the chaotic Lorenz system. We drop the 0 subscripts for ease of exposition from here on. We study the following Cauchy problem for $t \in [0, \infty)$:

$$\frac{\partial}{\partial t} v(x, y, t) = \mathcal{L} v(x, y, t),$$

$$v(x, y, 0) = \phi(x, y), \quad (7)$$

with the generator

$$\mathcal{L} = \frac{1}{\varepsilon^2}\mathcal{L}_0 + \frac{1}{\varepsilon}\mathcal{L}_1 + \mathcal{L}_2,$$

where

$$\mathcal{L}_0 \equiv g(y) \cdot \mathbf{V}_y, \quad (8)$$

$$\mathcal{L}_1 \equiv \frac{4}{90} y_2 \frac{\partial}{\partial x}, \quad (9)$$

$$\mathcal{L}_2 \equiv x_2(1 - x^2)\frac{\partial}{\partial x}, \quad (10)$$

and $g(y)$ denotes the scaled vector field of the fast Lorenz system $g(y) = [10(y_2 - y_1), (28y_1 - y_2 - y_1 y_3), (y_1 y_2 - 8y_3/3)]$. We remark that equally we could have used the framework of the Fokker–Planck equation.

Pioneered by Kurtz (1973) and Papanicolaou (1976), a perturbation expansion can be made according to

$$v(x, y, t) = v_0 + \varepsilon v_1 + \varepsilon^2 v_2 + \cdots. \quad (11)$$

A recent exposition of the theory of homogenization and their applications is provided in Givon et al. (2004) and Pavliotis and Stuart (2008). Substituting series (11) into the backward Kolmogorov equation (7), we obtain at lowest order $\mathcal{O}(1/\varepsilon^2)$

$$\mathcal{L}_0 v_0 = 0. \quad (12)$$

The chaotic dynamics of the fast Lorenz system, associated with the generator L_0, is ergodic (Tucker 1999). Hence the expectation value does not depend on initial conditions y and we obtain

$$v_0 = v_0(x, t)$$

as the only solution of (12).

Ergodicity implies the existence of a unique invariant density induced by the fast chaotic dynamics given by the unique solution of the associated Fokker–Planck equation

$$\mathcal{L}_0^* \rho = 0,$$

where $\mathcal{L}_0^*$ is the formal L_2 adjoint of the generator $\mathcal{L}_0$. We label this solution as $\rho_\infty(y)$.

At the next order $\mathcal{O}(1/\varepsilon)$, we obtain

$$\mathcal{L}_0 v_1 = -\mathcal{L}_1 v_0. \quad (13)$$

To assure boundedness of v_1 [and thereby of the asymptotic expansion (11)] a solvability condition has to be satisfied prescribed by the Fredholm alternative. Equation (13) is solvable only provided the right-hand side is in the space orthogonal to the (one-dimensional) null space of the adjoint $\mathcal{L}_0^*$; that is, if

$$\langle \mathcal{L}_1 v_0 \rangle_{\rho_\infty} = -\frac{4}{90}\langle y_2 \rangle_{\rho_\infty} \frac{\partial}{\partial x} v_0(x, t) = 0,$$

where we introduced the average of an observable $h(x, y)$ over the fast ergodic density as $\langle h \rangle_{\rho_\infty} := \int h(x, y)\rho_\infty(y)\, dy$. Note that the vanishing of the average of the fast perturbation in the slow equation with respect to the invariant measure induced by the Lorenz system implies that classical averaging would only produce trivial reduced dynamics $\dot{x} = \mathcal{O}(\varepsilon)$. Since $\langle y_2 \rangle_{\rho_\infty} = 0$, there exists a solution of (13), which we can formally write as

$$v_1(x, y, t) = -\mathcal{L}_0^{-1}\mathcal{L}_1 v_0 + R(x), \quad (14)$$

where $R(x)$ lies in the kernel of $\mathcal{L}_0$.

At the next order $\mathcal{O}(1)$, we obtain

$$\mathcal{L}_0 v_2 = \frac{\partial}{\partial t} v_0 - \mathcal{L}_1 v_1 - \mathcal{L}_2 v_0, \quad (15)$$

which yields the desired evolution equation for v_0 as the associated solvability condition

$$\frac{\partial}{\partial t}v_0 = \langle \mathcal{L}_2 v_0 \rangle_{\rho_\infty} + \langle \mathcal{L}_1 v_1(x,t) \rangle_{\rho_\infty}. \qquad (16)$$

The reduced slow backward Kolmogorov equation is then

$$\frac{\partial}{\partial t}v_0 = x(1-x^2)\frac{\partial}{\partial x}v_0 - \left(\frac{4}{90}\right)^2 \langle y_2 \mathcal{L}_0^{-1} y_2 \rangle_{\rho_\infty} \frac{\partial^2}{\partial x^2}v_0. \quad (17)$$

This can be simplified further. For every function h with $\langle h \rangle_{\rho_\infty} = 0$ we define

$$H(x,y) = -\int_0^\infty e^{\mathcal{L}_0 t}h \, dt,$$

which, upon using the fact that $\mathcal{L}_0$ corresponds to an ergodic process, implies that $\mathcal{L}_0 H = h$ (Pavliotis and Stuart 2008). Hence we can evaluate

$$\langle y_2 \mathcal{L}_0^{-1} y_2 \rangle_{\rho_\infty} = -\int_0^\infty \langle y_2 e^{\mathcal{L}_0 t} y_2 \rangle \, dt$$

$$= -\int_0^\infty \left[\lim_{T\to\infty} \frac{1}{T}\int_0^T y_2(s)y_2(t+s)\,ds \right] dt,$$

where the last identity follows from employing Birkhoff's ergodic theorem. The reduced backward Kolmogorov equation (17) can then be written as

$$\frac{\partial}{\partial t}v_0 = x(1-x^2)\frac{\partial}{\partial x}v_0 + \frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2}v_0, \qquad (18)$$

with σ^2 given by (6). Note that $R(x)$ does not contribute to the dynamics. We can therefore choose $R(x) = 0$ in order to assure that $\langle v \rangle_{\rho_\infty} = \langle v_0 \rangle_{\rho_\infty} + \mathcal{O}(\varepsilon^2)$.

The slow reduced Langevin equation associated with the reduced backward Kolmogorov equation (18) is then given by our stochastic 1D climate model (5).

The unique invariant density of the gradient system (5) is readily determined as the unique stationary solution $\hat{\rho}(x)$ of the associated Fokker–Planck equation

$$\frac{\partial}{\partial t}\rho(x) = \frac{\partial}{\partial x}\left(\frac{dV}{dx}\rho\right) + \frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2}\rho. \qquad (19)$$

We find

$$\hat{\rho}(x) = \frac{1}{Z}e^{-(2/\sigma^2)V(x)} \qquad (20)$$

with

$$Z = \int_{-\infty}^\infty e^{-(2/\sigma^2)V(x)}\,dx.$$

Note that the unique invariant density of the full deterministic 4D system (1)–(4) is now approximated by $\rho(x,y) = \rho(x)\rho_\infty(y) + \mathcal{O}(\varepsilon)$.

3. Parameter estimation

The value of the diffusion coefficient σ given by (6) cannot be determined analytically for the reduced stochastic 1D climate model as the unique invariant density $\rho_\infty(y)$ of the fast Lorenz subsystem (2)–(4) is not explicitly known. We therefore need to evaluate the diffusion coefficient numerically from simulations of the full deterministic 4D model (1)–(4) under the assumption that its slow dynamics is well approximated by the stochastic 1D climate model (5). We do so by coarse-graining the full deterministic 4D system and estimating conditional averages (Gardiner 2004; Siegert et al. 1998; Stemler et al. 2007). This method has been used in the meteorological community to study time series in diverse contexts such as synoptic variability of midlatitude sea surface winds, Icelandic meteorological data, planetary waves, and paleoclimatic ice core data (Sura and Barsugli 2002; Sura 2003; Egger and Jónsson 2002; Berner 2005; Ditlevsen 1999). This approach also allows us to determine the drift term of the stochastic 1D climate model, which we write for convenience here as

$$dx = d(x)dt + \sigma dW, \qquad (21)$$

with drift term $d(x) = x(1-x^2)$. Numerically integrating the full deterministic 4D system (1)–(4) with a small time step dt, we create a long time series $x(t_i)$, $i = 0, 1, \ldots, n$ with $t_i = idt$. To estimate the parameters of the Langevin system (21) we subsample the time series $x(t_i)$ at a coarse sampling time $h \gg dt$. We choose h such that during h the finely sampled trajectory $x(t_i)$ performs roughly 3–4 fast (smooth) undulations induced by the fast chaotic dynamics.

We estimate the diffusion coefficient (and the drift term) by partitioning the state space into bins X_i of bin size ΔX. We coarse grain by mapping x into X for $x \in (X, X + \Delta X)$. To estimate the diffusion coefficient we define the average conditioned on the bins $[X, X + \Delta X]$

$$S(X) = \frac{1}{h}\langle (x^{n+1} - x^n)^2 \rangle \Big|_{x^n \in (X, X+\Delta X)}, \qquad (22)$$

where $x^n = x(nh)$. The angle brackets denote a conditional ensemble average over microscopic realizations $x(t_i)$ that fall into a coarse macroscopic bin $[X, X + \Delta X]$. In the limit when the deterministic 4D dynamics can be approximated by the stochastic 1D climate model (21) and for small values of the sampling time h, we can evaluate

$$S(X) = \sigma^2 + X^2(1-X^2)^2 h,$$

where we used $dW = 0$. In the limit $h \to 0$ we estimate $S(X) = \sigma^2$. Similarly the drift $d(x)$ can be estimated by

$$D(X) = \frac{1}{h}\langle(x^{n+1} - x^n)\rangle\Big|_{x^n \in (X, X + \Delta X)}, \qquad (23)$$

with $D(X) \approx d(x)$ in the limit $h \to 0$.

There are several pitfalls one may encounter using this method to estimate the effective drift and diffusion coefficients of chaotic deterministic differential equations related to the choice of the sampling time. [Sampling issues for purely stochastic multiscale systems were already reported in Pavliotis and Stuart (2007).] If h is chosen smaller than the typical time scales on which the fast subsystem decorrelates, the diffusion coefficient does not exist and $S(X)$ will be close to zero in the limit $h \to 0$. On the contrary, if h is chosen too large the estimator for the diffusion coefficient will be swamped by the deterministic part of the dynamics and we will obtain $S(X) \approx X^2(1 - X^2)^2 h$. Similarly there are problems with estimating the drift $D(X)$ for sampling times h taken too large. For such subsampling times h, the coarse-grained dynamics would for large values of h randomly switch from bin to bin according to the invariant density $\hat{\rho}(x)$ given by (20). Since in our case $\hat{\rho}(x)$ is symmetric, we can write the continuous approximation of the estimator for the drift as

$$D(X) = \frac{1}{h}\int(Y - X)\hat{\rho}(Y)\,dY = -\frac{X}{h}.$$

This may lead to erroneous classifications of stochastic processes as Ornstein–Uhlenbeck processes. We remark that if the time series is not sufficiently long, the estimated drift coefficient in our case of a bimodal probability density function will be erroneously estimated as $D(X) = -(X - X_0)/h$ because of the presence of the two metastable states around $X_0 = \pm 1$.

This procedure can be readily extended to study multidimensional stochastic processes where the bins would be hypercubes; the interpretation of the results in higher dimensions and the deducing functional dependencies may, however, be difficult. We remark that one may also use Kalman filters to estimate the parameters (Cornick et al. 2009) or estimate the generator directly (Crommelin and Vanden-Eijnden 2006).

In Fig. 2 we show results from the parameter estimation procedure. We choose a coarse-grain bin size of $\Delta X = 0.05$ and sample the slow variable x at constant sampling intervals $h = 0.005$ corresponding to four fast oscillations. We used a trajectory of the full deterministic 4D system (1)–(4) with $\varepsilon^2 = 0.0005$ and evolved until
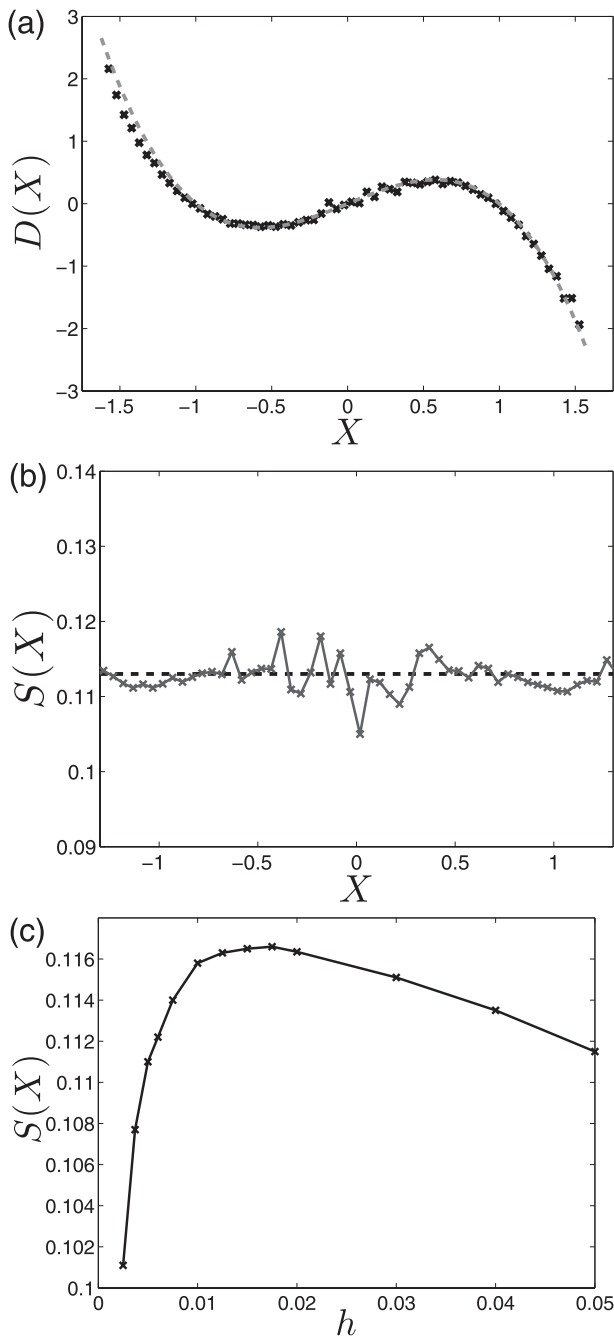

FIG. 2. Parameter estimation for the (a) drift term $D(X)$ and (b) diffusion coefficient $S(X)$ from a long trajectory of the full deterministic 4D system (1)–(4) with $\epsilon^2 = 0.0005$ and sample interval $h = 0.005$. The dashed line depicts the theoretically expected drift from the stochastic 1D climate model (5) in (a), and the constant diffusion coefficient $\sigma^2 = 0.113$ in (b). (c) Diffusion coefficient as a function of the sampling time.

$T = 3 \times 10^4$ time units. We confirm the 1D climate model (5) by estimating the drift coefficient clearly as

$$D(X) = X(1 - X^2),$$

and estimating a constant diffusion coefficient $S(X) = \sigma^2 = 0.113$. We note that this is slightly smaller than the value $\sigma^2 = 0.126$ found by Givon et al. (2004) using a different method.

In Fig. 2c we show how the diffusion coefficient depends on the sampling time as discussed above. It is clearly seen that the diffusion coefficient cannot be unambiguously determined (at least not using this method). We will investigate in section 5 how sensitive the statistics of the 1D climate model (5) is to this degree of uncertainty in the estimation of the diffusion coefficient.

4. Time scales

An important parameter in data assimilation is the observation interval Δt_{obs}. We will see in section 7 that skill improvement of the analysis is dependent on Δt_{obs}. We therefore now analyze several characteristic time scales of the deterministic 4D toy model (1)–(4). In particular we will focus on the slow variable x.

a. Autocorrelation time

We first estimate the decorrelation time τ_{corr}, which is linked to the decay rate of the autocorrelation function of the slow variable

$$C(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_0^T x(s)x(\tau + s)\, ds. \quad (24)$$

The autocorrelation time is estimated as the e-folding time of $C(\tau)$. The autocorrelation function $C(\tau)$ can be estimated from a long trajectory of x obtained from a simulation of the full deterministic 4D model (1)–(4). For small values of τ it decays approximately exponentially with a measured decay rate of 0.004 81, corresponding to an e-folding time of $\tau_{\text{corr}} = 208$ time units.

b. Sojourn and exit times

As a second time scale we consider the mean sojourn time

$$\bar{\tau} = \frac{1}{M} \sum_{i=1}^{M} \tau_i, \quad (25)$$

where the sojourn times τ_i measure the times spent in one of the slow metastable states around $x^\star = \pm 1$ as depicted in Fig. 1. We expect this time scale to strongly correlate with the autocorrelation time scale as it is random transitions between the slow regimes that cause decorrelation of the slow variable. We numerically estimate the mean sojourn time using two separate methods. First, we estimate the mean sojourn time directly from a long trajectory of the slow variables x using (25) as $\bar{\tau} \approx 218.2$ time units. Second, the succession of rapid
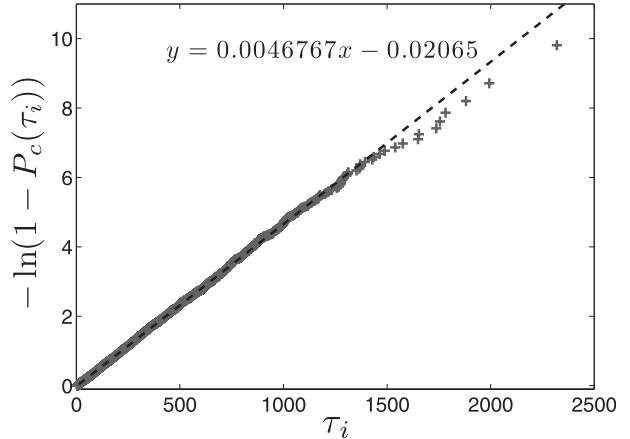

FIG. 3. Log plot of the normalized histogram of sojourn times τ_i of the slow variable x in each metastable state around $x^\star = \pm 1$ for the full deterministic 4D system (1)–(4). The dashed line shows a least squares fit. Using (26) the mean sojourn time can be estimated via the inverse of the slope of the least squares fit as $\bar{\tau} \approx 213.8$.

transitions and relatively long residence times in the potential wells suggests that successive jumps between metastable states can be treated as independent random events, and that the sojourn times are a Poisson process with cumulative probability distribution function

$$P_c(\tau_i) = 1 - \exp\left(\frac{\tau_i}{\tau}\right). \quad (26)$$

In Fig. 3 we show the empirical ranked histogram $P_c(\tau_i)$ for the sojourn times measured from a long trajectory $x(t)$ of the full deterministic 4D system (1)–(4), allowing us to determine the mean sojourn time $\bar{\tau} \approx 213.8$ time units. Calculating the mean sojourn time by either the average of individual sojourn times (25) or via the Poisson process approximation (26) yields results differing by only 2%, indicating that for $\varepsilon^2 = 0.01$ the fast chaotic Lorenz subsystem has almost fully decorrelated and that the rare transitions between the metastable states are approximately a Poisson process. In the following we will use the average of the two obtained values of the mean sojourn time and set $\bar{\tau} \approx 216.0$ time units.

These numerically obtained results for the full deterministic 4D system (1)–(4) compare well with the value calculated analytically for the stochastic 1D climate model (5) using Kramers' theory (Kramers 1940) to calculate the first exit time

$$\tau_e = \frac{1}{2}\bar{\tau}.$$

We define the first exit time τ_e to be the average time it takes from the potential well at $x = \pm 1$ to reach the saddle of $V(x)$ at $x = 0$. To calculate the exit time we solve the following Cauchy problem (see, e.g., Zwanzig 2001):

$$L_{\text{clim}}\tau_e = -1, \qquad (27)$$

where

$$L_{\text{clim}} = -V'\partial_x + \frac{\sigma^2}{2}\partial_{xx}$$

is the generator of the stochastic 1D climate model (5), with the boundary conditions $\partial_x\tau_e(x = \pm 1) = 0$ and $\tau_e(0) = 0$. Note that the first exit times for the symmetric potential out of the potential wells at $x = -1$ or $x = 1$ are identical. Upon using the boundary conditions, we solve (27) to obtain

$$\tau_e = \beta\int_{-1}^{0}dy\, e^{\beta V(y)}\int_{-1}^{y}e^{-\beta V(z)}\,dz,$$

with $\beta = 2/\sigma^2$, which can be numerically evaluated to yield $\tau_e = 117.8$ for $\sigma^2 = 0.113$. The exact analytical value calculated from the reduced stochastic 1D climate model compares well with the numerically estimated values of the first exit times of the full deterministic 4D model $\tau_e = 108.0$, confirming the accuracy of the homogenized model to faithfully reproduce the statistics of the full parent model.

c. Transit time

As a third characteristic time scale, we study the mean transit time τ_t, which is the average over the shortest transit time $\tau_{t,i}$ between the two slow metastable states around $x^\star = \pm 1$. We define this time to be the mean of

$$\tau_{t,i} = \min[(t_i - t_j)\,|\,x(t_i) = \pm 1 \wedge x(t_j) = \mp 1]. \quad (28)$$

For a long trajectory containing 5000 transitions between the slow metastable states we numerically estimate $\tau_t = 5.90$ for the full deterministic 4D model (1)–(4). We found that the transit time does not vary greatly with the value of ε.

As for the mean sojourn time, the transit time can be analytically calculated for the stochastic 1D climate model (see, e.g., Gardiner 2004) as

$$\tau_t = 4\frac{\Pi_{+1}(0) - \Pi_{-1}(0)}{\sigma^2\hat{\rho}_{-1}(0)}, \qquad (29)$$

where

$$\Pi_{+1}(x) = \hat{\rho}_{+1}(x)\int_{x}^{1}dx'\hat{\rho}^{-1}(x')\int_{-1}^{x'}\hat{\rho}_{-1}(z)\hat{\rho}(z)\,dz,$$

$$\Pi_{-1}(x) = \hat{\rho}_{-1}(x)\int_{-1}^{x}dx'\hat{\rho}^{-1}(x')\int_{-1}^{x'}\hat{\rho}_{-1}(z)\hat{\rho}(z)\,dz,$$

with the splitting probabilities

$$\rho_{-1}(x) = \frac{\displaystyle\int_{x}^{1}dz\hat{\rho}^{-1}(z)}{\displaystyle\int_{-1}^{1}dz\hat{\rho}^{-1}(z)},$$

$$\rho_{+1}(x) = \frac{\displaystyle\int_{-1}^{x}dz\hat{\rho}^{-1}(z)}{\displaystyle\int_{-1}^{1}dz\hat{\rho}^{-1}(z)}.$$

For the stochastic 1D climate model with $\sigma^2 = 0.113$ we evaluate this numerically to be $\tau_t = 5.66$, which again compares well with the numerically obtained value of the transit time for the full deterministic 4D system.

5. Validity of the stochastic climate model

In this section we numerically investigate to what degree the stochastic 1D climate model (5) faithfully reproduces the slow dynamics of the full deterministic 4D system (1)–(4). It is clear that such a correspondence can only be of a statistical nature. The close correspondence of the time scales of the full deterministic 4D system and of the stochastic 1D climate model for $\sigma^2 = 0.113$ and $\varepsilon^2 = 0.01$ already indicates the validity of the stochastic 1D climate model in reproducing statistical aspects of the slow dynamics. The rigorous theory by Kurtz (1973) and Papanicolaou (1976) assures weak convergence of the stochastic climate model in the limit of $\varepsilon \to 0$. Here we discuss the effect of finite time-scale separation ε. Furthermore we study the sensitivity of the climate model to uncertainties in the diffusion coefficient σ^2 as we have encountered in section 3. In the appendix we will also investigate the sensitivity to possible uncertainties in the drift term.

a. Probability density function

We first examine how the probability density function varies for different values of the time-scale separation parameter ε. In Fig. 4 we show empirical densities for the slow variable x obtained from a long simulation of the full deterministic 4D system (1)–(4) for different values of the time-scale separation. The obtained estimates of the empirical invariant densities exhibit a bimodal structure reflecting the two slow metastable regimes near $x^\star = \pm 1$. The probability density functions are reasonably insensitive to changes in ε^2, with the maxima of the densities at $x^\star = \pm 1$ differing for $\varepsilon^2 = 0.0005$ and $\varepsilon^2 = 0.01$ by less than 1%, and by approximately 5% for $\varepsilon^2 = 0.01$ and $\varepsilon^2 = 0.1$. This numerically demonstrates the weak convergence of solutions of the homogenized
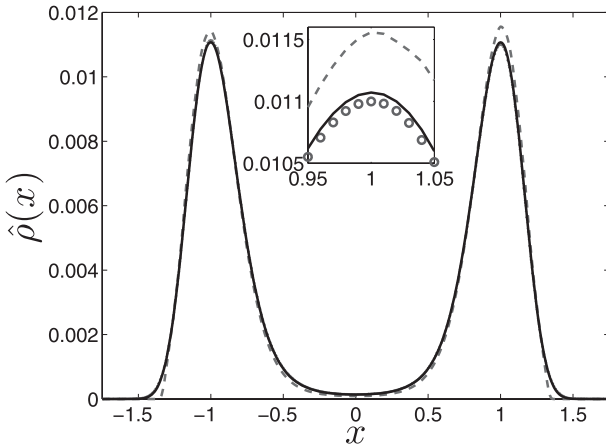
FIG. 4. Normalized empirical densities for the slow variable x of the full deterministic 4D model (1)–(4) with $\varepsilon^2 = 0.1$ (dashed line), 0.01 (solid line), and 0.0005 (circles). The empirical densities were obtained from long trajectories integrated to time $T = 10^9$ time units.
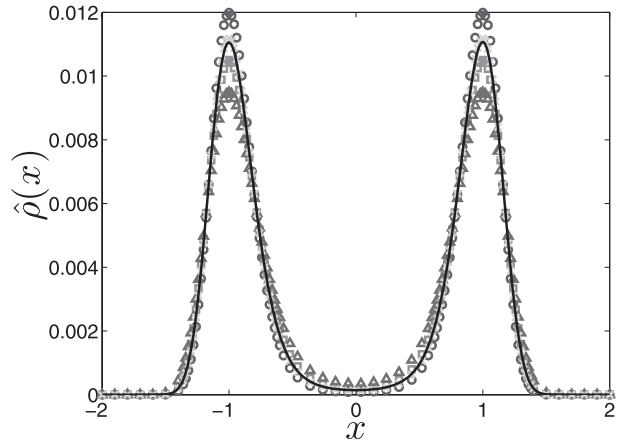

FIG. 5. Invariant density (20) for the stochastic 1D climate model (5), with $\sigma^2 = 0.1$ (circles), 0.113 (crosses), 0.126 (squares), and 0.15 (triangles), and empirical density for the slow variable x of the full deterministic 4D model (1)–(4) with $\varepsilon^2 = 0.01$ (solid line).

stochastic 1D climate model to solutions of the full deterministic slow dynamics.

Next we show in Fig. 5 how sensitive the empirical density of the stochastic 1D climate model (5) is to changes in the diffusion coefficient. Unlike for changes in the time-scale parameter, the 1D climate model is quite sensitive to changes in the diffusion coefficient. The invariant density of the full deterministic 4D model is best estimated by the stochastic 1D climate model with $\sigma^2 = 0.113$.

b. Time scales

Differences in the probability density function imply different statistics of the dynamics. In particular, we will now examine how the characteristic time scales of the full deterministic 4D model (1)–(4) change with the time-scale separation and how they change for the stochastic 1D climate model (5) with varying diffusion coefficients.

Consistent with our earlier observation that the probability density function does not vary greatly with the time-scale separation parameter (provided it is sufficiently small), we confirm here that the time scales are insensitive to a decrease in ε for $\varepsilon^2 < 0.01$. In particular the transit time τ_t exhibits very little variation with ε.

We find that with the exception of the transit time τ_t, the characteristic time scales of section 4 are very sensitive to uncertainties in the diffusion coefficient. For example, in Fig. 6 the autocorrelation function $C(\tau)$ is shown as estimated from a long trajectory of x obtained from a simulation of the full deterministic 4D model (1)–(4), and from simulations of the stochastic 1D climate model (5) with different values of the diffusion coefficient.

Using $\sigma^2 = 0.113$ in (5) produces the best fit to the shape of $C(\tau)$ for the full deterministic 4D model (1)–(4).

Table 1 lists the values of the characteristic time scales for the stochastic 1D climate model (5) for various values of the diffusion coefficient. We show the percentage difference of each value with that for the full deterministic 4D model (1)–(4) with $\varepsilon^2 = 0.01$ in brackets. All time scales vary by approximately a factor of 5 over the range of values indicated except for the transit time, which varies by just 13% and approximates the transit time of the full deterministic model better as σ^2 decreases. The transit time is determined by the rate of divergence associated with the unstable fix point at $x = 0$ of the deterministic drift term, making it relatively insensitive to changes in the diffusion coefficient. The transit times of
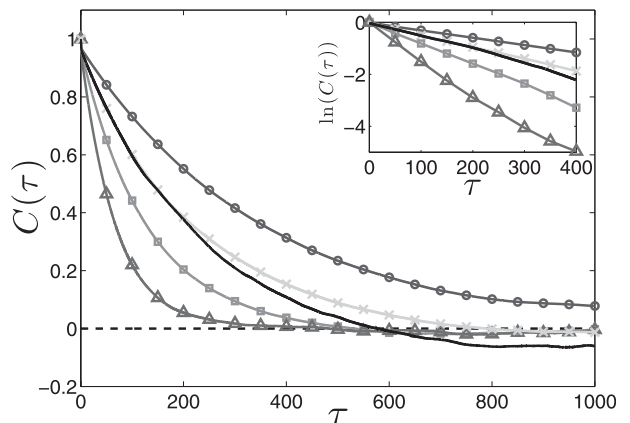

FIG. 6. Autocorrelation function $C(\tau)$ as a function of the time lag τ for the full deterministic 4D model (1)–(4) (solid line) and for the stochastic 1D climate model (5) with $\sigma^2 = 0.1$ (circles), 0.113 (crosses), 0.126 (squares), and 0.15 (triangles).

TABLE 1. Characteristic time scales of the stochastic 1D climate model (5) for different values of the diffusion coefficient σ^2. In parentheses we indicate the error when compared to values obtained from the full deterministic 4D model (1)–(4) with $\varepsilon^2 = 0.01$.

	$\sigma^2 = 0.1$	$\sigma^2 = 0.113$	$\sigma^2 = 0.126$	$\sigma^2 = 0.15$
τ_{corr}	353.9 (70.1%)	221.7 (6.6%)	129.0 (61.2%)	70.5 (195%)
τ_e	205.7 (90.5%)	117.8 (9.1%)	75.6 (42.6%)	40.8 (164.7%)
τ_t	5.86 (0.07%)	5.66 (4.2%)	5.48 (7.7%)	5.17 (14.1%)

the full deterministic model, however, are slightly overestimated, since definition (28) does not take into account the skewness of the probability density function (see Fig. 4), which implies that the trajectory spends more time within the range $x \in [-1, 1]$ than outside this range. The table indicates that all time scales (except the transit time τ_t) are best approximated by the stochastic 1D climate model with $\sigma^2 = 0.113$, consistent with the better approximation of the probability density function for that value of σ^2. However, we will see in section 7 that in the context of ensemble data assimilation it is not necessarily ideal to use this value for σ^2; in fact, it is preferable to use models with larger diffusion to create a more reliable ensemble with superior analysis skill.

6. Ensemble Kalman filtering

We briefly introduce how data assimilation is performed in an ensemble filter framework. Given an N-dimensional dynamical system

$$\dot{\mathbf{z}}_t = \mathbf{F}(\mathbf{z}_t), \qquad (30)$$

which is observed at discrete times $t_i = i\Delta t_{\text{obs}}$, data assimilation aims at producing the best estimate of the current state given a typically chaotic, possibly inaccurate model $\dot{\mathbf{z}} = f(\mathbf{z})$ and noisy observations of the true state $\mathbf{z}_t$ (Kalnay 2003).

Observations $\mathbf{y}_o \in \mathbb{R}^n$ are expressed as a perturbed truth according to

$$\mathbf{y}_o(t_i) = \mathbf{H}\mathbf{z}_t(t_i) + \mathbf{r}_o,$$

where the observation operator $\mathbf{H}$: $\mathbb{R}^N \rightarrow \mathbb{R}^n$ maps from the whole space into observation space, and $\mathbf{r}_o \in \mathbb{R}^n$ is assumed to be independent and identically distributed observational Gaussian noise with associated error covariance matrix $\mathbf{R}_o$.

In an ensemble Kalman filter (Evensen 1994, 2006) an ensemble with k members $\mathbf{z}_k$

$$\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k] \in \mathbb{R}^{N \times k}$$

is propagated by the dynamics (30) according to

$$\dot{\mathbf{Z}} = f(\mathbf{Z}), \quad f(\mathbf{Z}) = [f(\mathbf{z}_1), f(\mathbf{z}_2), \ldots, f(\mathbf{z}_k)] \in \mathbb{R}^{N \times k}. \qquad (31)$$

This forecast ensemble is split into its mean $\bar{\mathbf{z}}_f$ and ensemble deviation matrix $\mathbf{Z}'_f$. The ensemble deviation matrix can be used to approximate the ensemble forecast error covariance matrix via

$$\mathbf{P}_f(t) = \frac{1}{k-1}\mathbf{Z}'(t)[\mathbf{Z}'(t)]^{\text{T}} \in \mathbb{R}^{N \times N}. \qquad (32)$$

Note that $\mathbf{P}_f(t)$ is rank-deficient for $k < N$, which is the typical situation in numerical weather prediction where N is on the order of 10^9 and k on the order of 100.

Given the forecast mean, the forecast error covariance, and an observation, an analysis mean is produced by minimizing the cost function

$$J(\mathbf{z}) = \frac{1}{2}(\mathbf{z} - \bar{\mathbf{z}}_f)^{\text{T}}\mathbf{P}_f^{-1}(\mathbf{z} - \bar{\mathbf{z}}_f)$$
$$+ \frac{1}{2}(\mathbf{y}_o - \mathbf{H}\mathbf{z})^{\text{T}}\mathbf{R}_o^{-1}(\mathbf{y}_o - \mathbf{H}\mathbf{z}),$$

which penalizes distance from both the forecast mean and the observations with weights given by the inverse of the forecast error covariance and the observational noise covariance, respectively. The analysis mean is readily calculated as the critical point of this cost function with

$$\bar{\mathbf{z}}_a = \bar{\mathbf{z}}_f + \mathbf{K}_o(\mathbf{y}_o - \mathbf{H}\bar{\mathbf{z}}_f), \qquad (33)$$

where

$$\mathbf{K}_o = \mathbf{P}_f\mathbf{H}^{\text{T}}(\mathbf{H}\mathbf{P}_f\mathbf{H}^{\text{T}} + \mathbf{R}_o)^{-1},$$
$$\mathbf{P}_a = (\mathbf{I} - \mathbf{K}_o\mathbf{H})\mathbf{P}_f. \qquad (34)$$

Using the Shermann–Morrison–Woodbury formula (Golub and Loan 1996) we may rewrite the Kalman gain matrix and the analysis covariance matrix in the more familiar—though computationally more complex—form as $\mathbf{K}_o = \mathbf{P}_a\mathbf{H}^{\text{T}}\mathbf{R}_o^{-1}$ and

$$\mathbf{P}_a = (\mathbf{P}_f^{-1} + \mathbf{H}^{\text{T}}\mathbf{R}_o^{-1}\mathbf{H})^{-1}. \qquad (35)$$

To determine an ensemble $\mathbf{Z}_a$ that is consistent with the analysis error covariance $\mathbf{P}_a$ and that satisfies

$$\mathbf{P}_a = \frac{1}{k-1}\mathbf{Z}'_a[\mathbf{Z}'_a]^{\text{T}},$$

where the prime denotes the deviations from the analysis mean, we use the method of ensemble square root filters (Simon 2006). In particular, we use the method proposed in Tippett et al. (2003) and Wang et al. (2004),

the so-called ensemble transform Kalman filter (ETKF), which seeks a transformation $\mathbf{S} \in \mathbb{R}^{k \times k}$ such that the analysis deviation ensemble $\mathbf{Z}'_a$ is given as a deterministic perturbation of the forecast ensemble $\mathbf{Z}_f$ via

$$\mathbf{Z}'_a = \mathbf{Z}'_f \mathbf{S}. \tag{36}$$

Alternatively one could choose the ensemble adjustment filter (Anderson 2001) in which the ensemble deviation matrix $\mathbf{Z}'_f$ is premultiplied with an appropriately determined matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$. Note that the matrix $\mathbf{S}$ is not uniquely determined for $k < N$. The transformation matrix $\mathbf{S}$ can be obtained by (Wang et al. 2004)

$$\mathbf{S} = \overline{\mathbf{C}}(\mathbf{I}_k + \overline{\mathbf{\Gamma}})^{-1/2}\overline{\mathbf{C}}^{\mathrm{T}}.$$

Here $\mathbf{C}\mathbf{\Gamma}\mathbf{C}^{\mathrm{T}}$ is the singular value decomposition of

$$\mathbf{U} = \frac{1}{k-1}\mathbf{Z}'^{\mathrm{T}}_f\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}_o\mathbf{H}\mathbf{Z}'_f.$$

The matrix $\overline{\mathbf{C}} \in \mathbb{R}^{k \times (k-1)}$ is obtained by erasing the last zero column from $\mathbf{C} \in \mathbb{R}^{k \times k}$, and $\overline{\mathbf{\Gamma}} \in \mathbb{R}^{(k-1) \times (k-1)}$ is the upper left $(k-1) \times (k-1)$ block of the diagonal matrix $\mathbf{\Gamma} \in \mathbb{R}^{k \times k}$. The deletion of the zero eigenvalue and the associated columns in $\mathbf{C}$ assure that $\mathbf{Z}'_a = \mathbf{Z}'_a\mathbf{S}$ and therefore that the analysis mean is given by $\overline{\mathbf{z}}_a$. This method assures that the mean is preserved under the transformation (Wang et al. 2004), which is not necessarily true for general square root filters. We note that the continuous Kalman–Bucy filter can be used to calculate $\mathbf{z}'_a$ without using any computations of matrix inverses, which may be advantageous in high-dimensional systems (Bergemann et al. 2009).

A new forecast is then obtained by propagating $\mathbf{Z}_a$ with the nonlinear forecast model to the next time of observation, where a new analysis cycle will be started.

In the next section we perform data assimilation for the toy model introduced in section 2. We examine when and why the stochastic 1D climate model (5) can be used as a forecast model to improve the analysis skill. In particular, we will relate the range of validity of the 1D climate model to the relation between the observation interval and the characteristic time scales introduced in section 4.

7. Stochastic climate model as forecast model in data assimilation

We investigate how the ETKF as described in the previous section performs when either the full 4D deterministic system (1)–(4) or the reduced stochastic 1D climate model (5) is used as a forecast model, when the truth evolves according to the full deterministic 4D
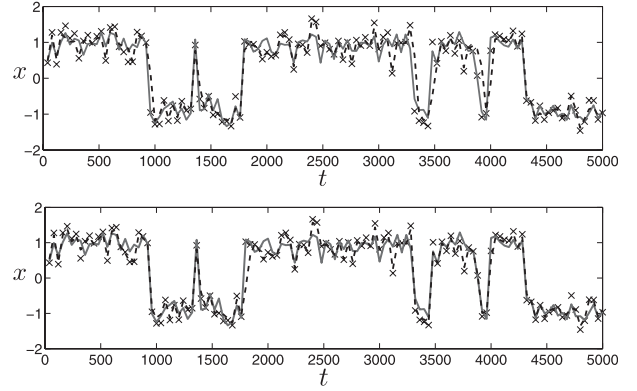

FIG. 7. Sample truth (solid line) and ETKF analyses (dashed line) for the observed x-component using (top) the full deterministic 4D model (1)–(4) and (bottom) the reduced stochastic 1D climate model (5) as the forecast model. The crosses are observations with error variance $R_{\mathrm{obs}} = 0.5\sigma^2$ ($\sigma^2 = 0.126$). Here we have used an observation interval $\Delta t_{\mathrm{obs}} = 50$ and $k = 15$ ensemble members.

model. We are concerned here with the following questions: Can the dimension reduced stochastic climate model produce a better analysis than the full deterministic model? And if so, under what circumstances and why?

We generate a truth $\mathbf{z}_t = (x_t, y_{1t}, y_{2t}, y_{3t})$ by integrating (1)–(4) using a fourth-order Runge–Kutta scheme with time step $dt = \varepsilon^2/20$. Observations of the slow variables only are then generated at equidistant times separated by the observation interval Δt_{obs} according to $x_{\mathrm{obs}} = x_t + \eta$ where $\eta \sim \mathcal{N}(0, \sqrt{R_{\mathrm{obs}}})$, with $R_{\mathrm{obs}} = 0.5\sigma^2$ and $\sigma^2 = 0.126$. Note that this value is slightly larger than the optimal value $\sigma^2 = 0.113$, which was found in the previous section to best approximate the full deterministic 4D model.

To integrate the forecast model we use a fourth-order Runge–Kutta scheme with $dt = \varepsilon^2/20$ for the full deterministic 4D model with $\varepsilon^2 = 0.01$, and an Euler–Maruyama scheme for the reduced stochastic 1D climate model using the same time step. Unless stated otherwise we use $\sigma^2 = 0.126$ for the stochastic 1D climate model. We include a spinup time of 100 analysis cycles. We perform twin experiments, where both forecast models are started from the same initial conditions, and are then subsequently assimilated using the same truth and observations over 5000 time units. To avoid rare occasions of filter divergence due to an underestimation of the forecast covariances we employ a 2% variance inflation (Anderson and Anderson 1999).

Figure 7 shows two sample analysis time series created by assimilating the same truth and observations at regularly spaced intervals of $\Delta t_{\mathrm{obs}} = 50$ where the stochastic 1D climate model and the full deterministic 4D model are used as the respective forecast models. Both forecast
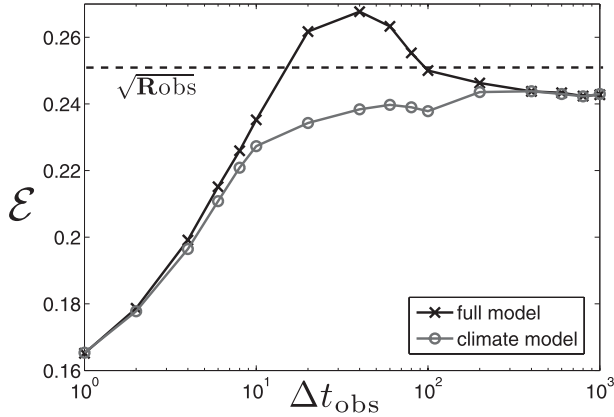
FIG. 8. RMS errors $\mathcal{E}$ of the analysis using the full deterministic 4D model (1)–(4) (crosses) and the stochastic 1D climate model (5) with $\sigma^2 = 0.126$ (circles) as forecast models, as a function of the observation interval Δt_{obs}. The horizontal dashed line indicates the observational error $\sqrt{R_{\mathrm{obs}}} = \sigma/\sqrt{2} = 0.251$ ($\sigma^2 = 0.126$). The results are averaged over 1000 realizations.

FIG. 9. Proportional skill improvement $\mathcal{S}$ of the stochastic 1D climate model (5) over the full deterministic 4D model (1)–(4) as a function of the observation interval Δt_{obs}. We show $\mathcal{S}$ as calculated over all analyses (circles), the metastable regimes only (crosses), and the transitions only (squares). Dashed vertical lines indicate the characteristic time scales of the slow dynamics: the decorrelation time τ_{corr}, first exit time τ_e, and transit time τ_t. Parameters are as in Fig. 8.

models track the truth well. However, the stochastic 1D climate model tracks the transitions between the slow metastable states around $x^\star = \pm 1$ more accurately.

We quantify the improvement made using the reduced stochastic 1D climate model by measuring the RMS error $\mathcal{E}$ between the observed truth x_t and the analysis mean $\bar{x}_a$:

$$\mathcal{E} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[\bar{x}_a(t_i) - x_t(t_i)]^2},$$

where N is the total number of analysis cycles. In Fig. 8 we show $\mathcal{E}$ as averaged over 1000 realizations, where we have used both the full deterministic 4D model and the reduced stochastic 1D climate model as forecast models, as a function of the observation interval Δt_{obs}. Both filters exhibit the same accuracy for small observation intervals (<10), when the system is frequently observed, and for very large observational times (>500), when the forecast is much larger than all the characteristic time scales we discussed in section 4. At very large observation intervals (not shown), when the ensemble of both forecast models will have explored the state space of the slow variable with climatological variance $\sigma_{\mathrm{clim}}^2 \approx 1$, the resulting analysis error covariance can be estimated using the Kirchoff-type addition rule of covariances (35) as $\mathbf{P}_a^{-1} = \sigma_{\mathrm{clim}}^{-2} + R_{\mathrm{obs}}^{-1}$ as $P_a = 0.244^2$ for $R_{\mathrm{obs}} = 0.5\sigma^2$ and $\sigma^2 = 0.126$, which corresponds well with the measured asymptote of the RMS error in Fig. 8. Interestingly, for moderate observation intervals, the stochastic 1D climate model performs considerably better than the full deterministic 4D model. Moreover, the analysis using the
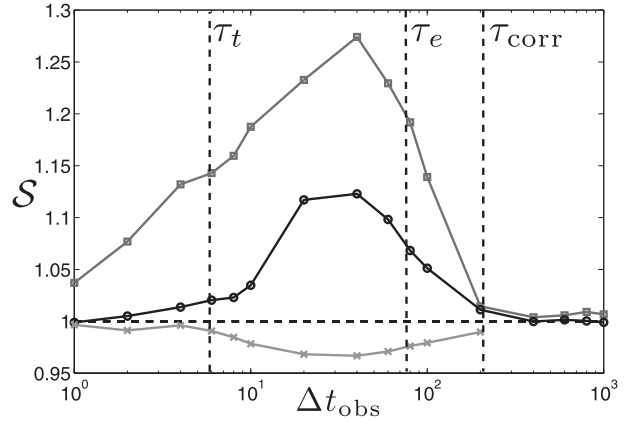
full deterministic 4D forecast model exhibits RMS errors worse than the observational error of $\sqrt{R_{\mathrm{obs}}} = 0.251$. In contrast, when the stochastic 1D climate model is used as a forecast model for the filter, the analysis error is always below the observational error (i.e., performing data assimilation is desirable over just trusting the observations, albeit with an improvement in RMS error of only around 5%).

We introduce the proportional improvement in the RMS error or skill

$$\mathcal{S} = \frac{\mathcal{E}_{\mathrm{full}}}{\mathcal{E}_{\mathrm{clim}}},$$

where $\mathcal{E}_{\mathrm{full}}$ and $\mathcal{E}_{\mathrm{clim}}$ are the RMS errors when the full deterministic or the reduced stochastic model is used as forecast model, respectively. Values of $\mathcal{S} \geq 1$ imply that it is beneficial to use the reduced stochastic 1D climate model (note that when $\mathcal{S} = 1$ it is still beneficial to use the stochastic 1D climate model because of the reduced computational expense). Figure 9 shows $\mathcal{S}$ as a function of Δt_{obs}, averaged again over 1000 realizations, with the time scales τ_t, τ_e, and τ_{corr} superimposed. Skill improvements occur only for observation intervals larger than the mean transit time τ_t, with the maximum skill improvement over all forecasts occurring for Δt_{obs} slightly smaller than the mean exit time. If forecasts are longer than τ_{corr}, the full and climate models are essentially indistinguishable with $\mathcal{S} = 1$.

Figure 9 also shows how the skill is distributed over the whole set of the analyses including metastable states

and transitions between them, as well as only over those analyses which remain within a metastable state near either $x^\star = -1$ or $x^\star = 1$, and those which fall in the transitions between the metastable states. Note that for large observation intervals $\Delta t_{\mathrm{obs}} > \tau_{\mathrm{corr}}$ half of all forecasts switch between regimes, so the notion of analyses that remain within one metastable regime becomes obsolete. As such, we do not plot skill as calculated over the metastable regimes for $\tau > \tau_{\mathrm{corr}}$ in this case.

Figure 9 illustrates that the major contribution to the skill improvement is from the stochastic 1D climate model better detecting the transitions between slow metastable states. While there is actually a minor degradation in skill of around 3% for forecasts with $\tau_t < \Delta t_{\mathrm{obs}} < \tau_{\mathrm{corr}}$ made for analyses that remain within metastable states, skill improvements of over 25% are made for those analyses which fall into the transitions between the metastable states. This is because analyses that detect transitions between metastable states incorrectly contribute large RMS errors $O(1)$, compared to the small errors $O(R_{\mathrm{obs}})$ made by incorrect analyses within the metastable states. These large errors have a major impact upon the total skill. There is also skill improvement as calculated over the transition periods for small values of the observation interval. However, there is no corresponding overall skill improvement since for small observation intervals the number of instances where the truth remains within a metastable regime is overwhelmingly larger, thereby swamping the skill improvement obtained over the rare transitions.

We note that the skill is insensitive to the observational noise level, provided R_{obs} is sufficiently small that observations do not misrepresent the actual metastable state or regime in which the truth resides. The skill is roughly the same for $R_{\mathrm{obs}} = 0.1\sigma^2$ and $R_{\mathrm{obs}} = 0.25\sigma^2$, and we observe no skill improvement $\mathcal{S} \approx 1$ when we take very poor observations with $R_{\mathrm{obs}} = 0.5$, approximately half the climatological variance of the full slow system.

a. Sensitivity of analysis skill to uncertainties in the diffusion coefficient

As for the time scales in section 5, we examine how the analysis skill depends on the diffusion coefficient (see the appendix for dependence on the drift term). Figure 10 shows the skill as a function of the observation interval for several values of σ^2. For very small diffusion and for very large diffusion (not shown) the skill is actually smaller than 1 for all observation intervals, implying that the stochastic 1D climate model performs worse than the full deterministic 4D model. There exists a range of values of σ^2 for which the stochastic climate model outperforms the full deterministic model with roughly similar skill values $\mathcal{S}(\Delta t_{\mathrm{obs}}) > 1$ for $\tau_t < \Delta t_{\mathrm{obs}} < \tau_{\mathrm{corr}}$.
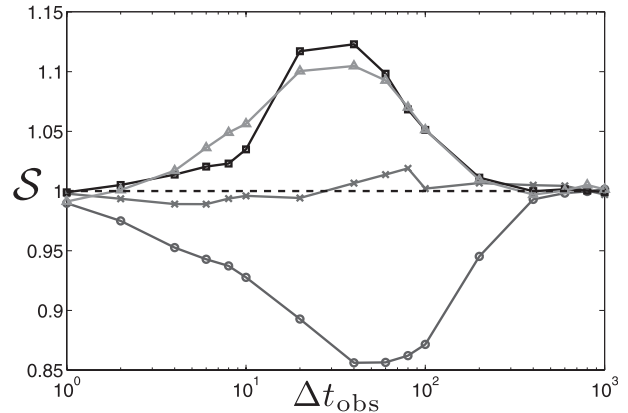

FIG. 10. Proportional skill improvement $\mathcal{S}$ of the stochastic 1D climate model (5) over the full deterministic 4D model (1)–(4) as a function of the observation interval Δt_{obs} with $\sigma^2 = 0.1$ (circles), 0.113 (crosses), 0.126 (squares), and 0.15 (triangles). All other parameters are as in Fig. 8.

The dependence of analysis skill upon the diffusion coefficient can be understood as follows: For small diffusion ($\sigma^2 = 0.1$) the stochastic 1D climate model is less diffusive than the truth and the deterministic 4D forecast model, and its associated forecast is likely to remain in a metastable state. This produces small ensemble spread and therefore causes more instances of the forecast not detecting a transition between regimes. On the other extreme, for large diffusion ($\sigma^2 \gg 0.15$), the stochastic 1D climate model exhibits regime switches too frequently, thereby producing a large forecast error that contaminates the analysis. For moderate values of the diffusion coefficient, the increase in spread of the forecast ensemble helps the detection of regime switches.

b. Effect of ensemble size and covariance inflation

The results from the previous subsection suggest that the observed skill improvement is linked to the increased forecast ensemble spread of the stochastic climate model, rather than to the accuracy of the stochastic parameterization. This is exemplified by the fact that for $\sigma^2 = 0.113$ for which the stochastic 1D climate model approximates the statistics of the full deterministic 4D model best, there is no skill improvement $\mathcal{S} \approx 1$, and skill improvement $\mathcal{S} > 1$ is given for $\sigma^2 > 0.113$ where transitions between the metastable states are better captured.

The lack of sufficient ensemble spread when using the full deterministic system as a forecast model is caused by the finite size of the ensemble (Ehrendorfer 2007). In the data assimilation community a common approach to account for insufficient spread is covariance inflation (Anderson and Anderson 1999), whereby the forecast variance $\mathbf{P}_f$ is artificially inflated by multiplication with a factor $\delta > 1$. We now investigate how the ensemble
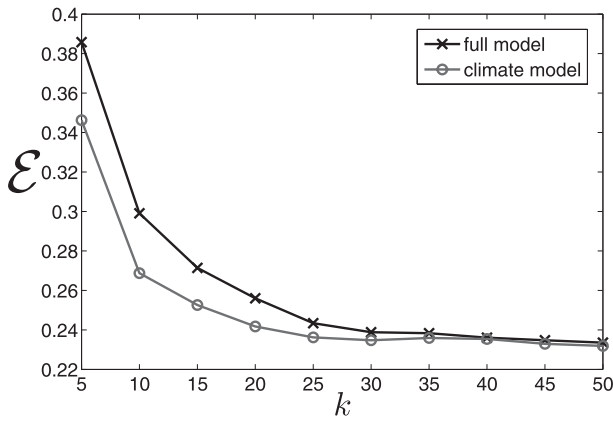
FIG. 11. RMS error $\mathcal{E}$ as a function of ensemble size k, using the full deterministic 4D model (1)–(4) (crosses) and the stochastic 1D climate model (5) with $\sigma^2 = 0.126$ (circles) as forecast models. We used $\Delta t_{obs} = 40$; all other parameters are as in Fig. 8.
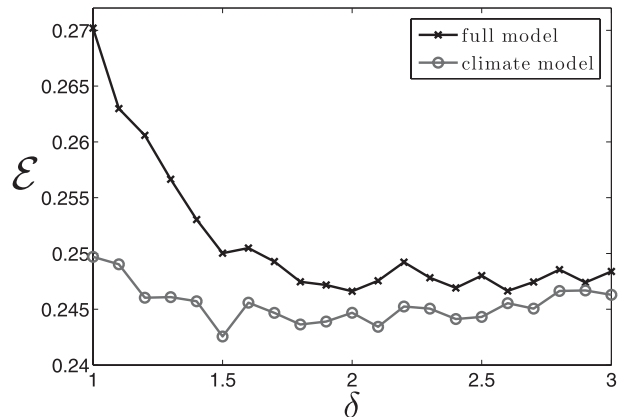

FIG. 12. RMS error $\mathcal{E}$ as a function of inflation factor δ, using the full deterministic 4D model (1)–(4) (crosses) and the stochastic 1D climate model (5) with $\sigma^2 = 0.126$ (circles) as forecast models. We used $\Delta t_{obs} = 40$; all other parameters are as in Fig. 8.

size k and inflation factor δ affect the skill. In particular we show that the RMS error using the full deterministic model can be reduced by either increasing the ensemble size or by using larger inflation factors.

Figure 11 shows $\mathcal{E}$ as a function of the ensemble size for $\Delta t_{obs} = 40$ (which is close to the observation interval yielding maximal skill improvement; see Fig. 9), $\sigma^2 = 0.126$, and averaged over 200 realizations. For large ensembles with $k \geq 35$ the overdiffusive stochastic 1D climate model and the full deterministic 4D model perform as well as forecast models. We have checked that when using a large ensemble ($k = 50$) the stochastic climate and full deterministic models perform equally well for the range of observation intervals used in Fig. 8. For smaller ensemble sizes the stochastic 1D climate model outperforms the full deterministic 4D model. This illustrates that the skill improvement is linked to the insufficient ensemble spread of the deterministic 4D model, which is compensated in the stochastic 1D climate model by a slightly increased diffusion coefficient.

Similarly, for fixed ensemble size, the ensemble spread can be artificially increased by inflating the forecast covariance with a constant factor. Figure 12 shows $\mathcal{E}$ as a function of the inflation factor, for $1 \leq \delta \leq 3$. It is seen that for unrealistically large values of δ the RMS error decreases when the full deterministic 4D model is used as a forecast model, whereas it is relatively insensitive to changes in δ for the stochastic 1D climate model. For even larger values of δ the RMS errors for the two forecast models eventually become equal (not shown).

c. Reliability and ranked probability diagrams

Another way of understanding why the ensemble variance of the stochastic 1D climate model produces skill better than that of the full deterministic 4D model is through ranked probability histograms (also known as Talagrand diagrams; Anderson 1996; Hamill and Colucci 1997; Talagrand et al. 1997). To create probability density histograms we sort the forecast ensemble $\mathbf{x}_f = [x_{f,1}, x_{f,2}, \ldots, x_{f,k}]$ and create bins $(-\infty, x_{f,1}]$, $(x_{f,1}, x_{f,2}], \ldots, (x_{f,k}, \infty)$ at each forecast step. We then increment whichever bin the truth falls into at each forecast step to produce a histogram of probabilities P_i of the truth being in bin i. A reliable ensemble is considered to be one where the truth and the ensemble members can be viewed as drawn from the same distribution. A flat probability density histogram therefore is seen as indicating a reliable ensemble for which each ensemble member has equal probability of being nearest to the truth. A convex probability density histogram indicates a lack of spread of the ensemble, while a concave diagram indicates an ensemble that exhibits too large spread. We direct the reader to Wilks (2006) for a detailed discussion.

Hamill (2001) suggested that a single ranked probability histogram is not sufficient for dynamical systems with several dynamical regimes as in our case; rather, one must look at the variability of the ensemble in the different dynamical regions individually. We therefore construct ranked probability histograms (for both forecast models) over the metastable regime with characteristic time $\bar{\tau}$ only and over the complementary regime of transitions with characteristic time τ_t only, as done for the distribution of the skill over those regimes (see Fig. 9).

In Fig. 13, we plot probability density histograms of the forecast for the full deterministic 4D model (1)–(4) and the stochastic 1D climate model (5) with various values of the diffusion coefficient, (a) over all analyses,
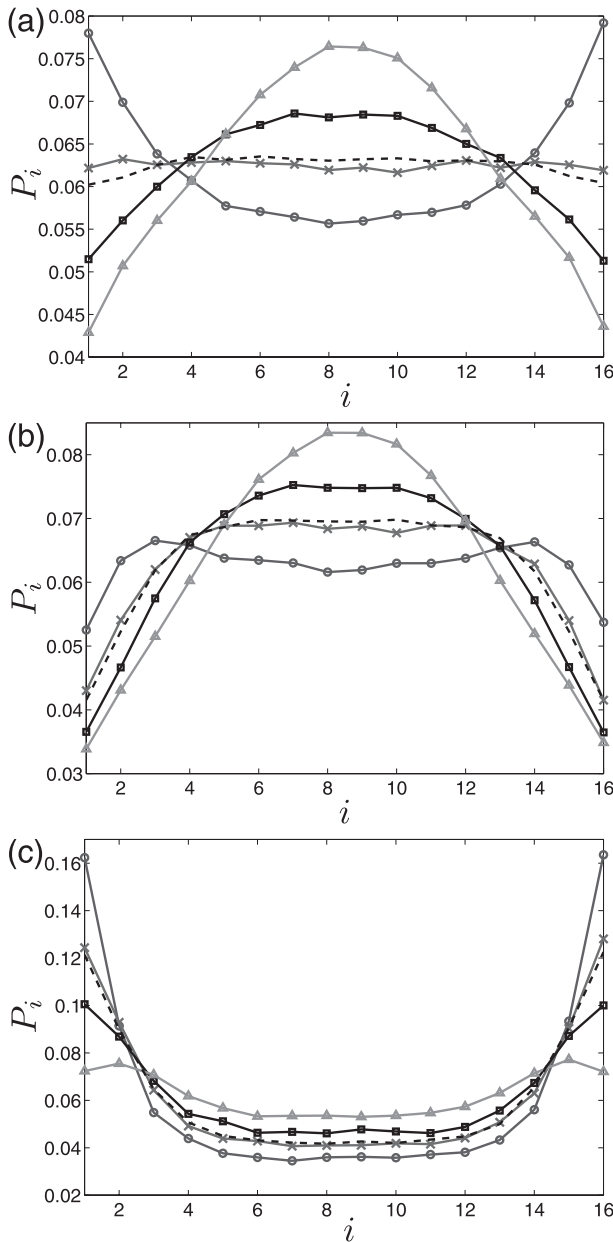
FIG. 13. Ranked probability histograms for the full deterministic 4D model (1)–(4) (dashed line) and the stochastic 1D climate model (5) with $\sigma^2 = 0.1$ (circles), 0.113 (crosses), 0.126 (squares), and 0.15 (triangles) over (a) all forecasts, as well as (b) only those forecasts that fall near the metastable states $x^\star = \pm 1$ and (c) those that fall in the transitions between metastable states.

(b) over the analyses which fall near the metastable states $x^\star = \pm 1$, and (c) over the analyses that fall in the transitions between metastable states. We use a long analysis cycle containing 250 000 forecasts. Although the full deterministic model produces the most reliable ensemble when averaged over all forecasts, it overestimates the variance in the metastable regions and

underestimates the variance across the transitions. For the stochastic 1D climate model, the more under-diffusive model with $\sigma^2 = 0.1$ produces the most reliable ensemble in the metastable regions while the more diffusive models with $\sigma^2 = 0.126$ and $\sigma^2 = 0.15$ both produce more uniform ranked probability histograms than the full model does in the transition regions. The reduced stochastic 1D climate model with optimal diffusion coefficient $\sigma^2 = 0.113$ behaves similarly to the full deterministic 4D model as expected. It is thus more desirable to favor the more diffusive stochastic climate models because a comparatively large improvement in RMS error can be made in the transition regions compared to in the metastable regions. This agrees with our earlier observation that the skill improvement is entirely due to the stochastic climate model being more accurate in the transition regions where large errors can be accrued by misclassifying the regime (see Fig. 9).

8. Discussion

We have investigated a homogenized stochastic climate model of a bistable multiscale system with a chaotic fast subsystem, and its usage as a forecast model in ensemble Kalman filtering. The reduced stochastic climate model was numerically shown to faithfully reproduce the statistics of the slow dynamics of the full deterministic model even in the case of finite time-scale separation where the theorems underpinning homogenization fail. Homogenized stochastic climate models replace in a controlled fashion the accumulative effect of the fast chaotic dynamics on the slow dynamics by white noise. We estimated the analytical expressions of the drift and diffusion terms of the reduced stochastic model using coarse-graining and quadratic variations, and found that the diffusion coefficient is very sensitive to the applied undersampling time. The statistics of the stochastic climate model (i.e., the probability density function and averaged characteristic time scales) were found to be very sensitive to the diffusion and drift terms used. Despite this sensitivity and the implied uncertainties associated with the "correct" stochastic climate model, we found that stochastic climate models can be beneficial as forecast models for data assimilation in an ensemble filter setting. We have shown numerically that using such a stochastic climate model has the advantage of 1) being computationally more efficient to run due to the reduced dimensionality and avoidance of stiff dynamics and 2) producing a superior analysis mean for a range of observation intervals. These skill improvements occur for observation intervals larger than the time taken to switch between slow metastable states τ_t but less than the decorrelation time τ_{corr}.

This skill improvement is due to the associated larger ensemble variance of the stochastic climate model, which allows the ensemble to explore the full state space of the slow variable better than the full deterministic model does. Forecasts made by the stochastic climate model will generally contain more ensemble members that cross the potential barrier between the slow metastable states (weather or climate regimes) than full model forecasts will. Consequently, there is a greater chance during the forecast of one or more climate model ensemble members crossing the potential barrier to the opposite metastable state than in the full deterministic model ensemble. If an observation appears near the other metastable state from where the forecast began, the Kalman filter analysis equation (33) trusts the observation more than the forecast, correctly producing an analysis that changes between the regimes. Thus, the stochastic climate model with large enough diffusion better detects transitions between slow metastable states and so has lower RMS error near these transitions in which large errors can be accrued.

We remark that strict time-scale separation is not necessary for the existence of metastable regimes; we have checked that for $\epsilon = 1$ dynamics similar to those depicted in Fig. 1 can be observed for increased coupling strength. In this case we can still obtain skill for observation intervals between the transit time and the decorrelation time when using a stochastic Langevin equation of the form (5), albeit with decreased diffusion whose dynamically optimal value is no longer approximated by homogenization.

Similarly, we remark that multimodal probability density functions are not necessary for the existence of metastable regimes, as was pointed out by Wirth (2001) and Majda et al. (2006). In cyclostationary systems (i.e., systems that involve time-periodic coefficients) or in aperiodic systems in which "regimes" occur intermittently, the probability density function may be unimodal whereas if restricted to time windows focusing on the regimes the "hidden" regimes can be identified. This has important implications for interpreting atmospheric data (Majda et al. 2006; Franzke et al. 2008, 2009) where one needs to reconcile the apparent paradox of multimodal probability density functions of planetary waves obtained from (relatively short time) observational data (Kimoto and Ghil 1993; Cheng and Wallace 1993; Corti et al. 1999) and unimodal but non-Gaussian probability density functions obtained from long-time simulations of atmospheric general circulation models (Kondrashov et al. 2004; Branstator and Berner 2005; Franzke and Majda 2006; Berner and Branstator 2007) [for a critical account on the analysis of short climatic data see, e.g., Hsu and Zwiers (2001); Stephenson et al. (2004); Ambaum (2008)].

However, it is pertinent to mention that the results found here are not dependent on the existence of metastable states but only require dynamics with rapid large amplitude excursions as, for example, in intermittent systems. In Harlim and Majda (2008, 2010) a system without multiple equilibria was investigated and superior filter performance was found for stochastic parameterizations (in this case by radically replacing all nonlinear terms by linear Ornstein–Uhlenbeck processes in Fourier space). We believe that their observed increased analysis skill can be explained by the increased ensemble spread counteracting finite sampling errors of the underlying ensemble Kalman filter approach as we did. The advantage when homogenization can be used is that the slow dynamics and its mean are still well resolved and reproduced and therefore the ensemble spread increasing stochasticity is of a more controlled nature. The use of homogenized stochastic climate models with adjusted larger diffusivity therefore amounts to incorporating deliberate but controlled model error into the data assimilation scheme.

We found that the same skill can be obtained using the full deterministic model either by using a larger ensemble or by increasing the covariance inflation factor. The first is undesirable for large models because of the computational difficulty of simulating large ensembles and their covariances, while the second is undesirable because it introduces unphysical inflation (here a multiplicative factor of $\delta > 2$ was needed) and requires expensive empirical tuning of the inflation factor.

Using stochastic reduced models for multiscale systems has computational advantages. First of all, the dimensional reduction allows for a gain in computational speed. Furthermore, by replacing a stiff ordinary differential equation by a nonstiff stochastic differential equation means that coarser integration time steps can be used that otherwise would lead to numerical instabilities when simulating the stiff multiscale system.

Homogenized stochastic climate models therefore provide a cheap way of simulating large ensembles, and a more natural way of incorporating forecast covariance inflation into the data assimilation algorithm. In particular, we find that the stochastic climate model produces a more reliable ensemble (as characterized by the ranked probability histogram) that better detects the transitions between slow metastable states. This leads to a far superior analysis in these regimes than that produced by the full deterministic model.
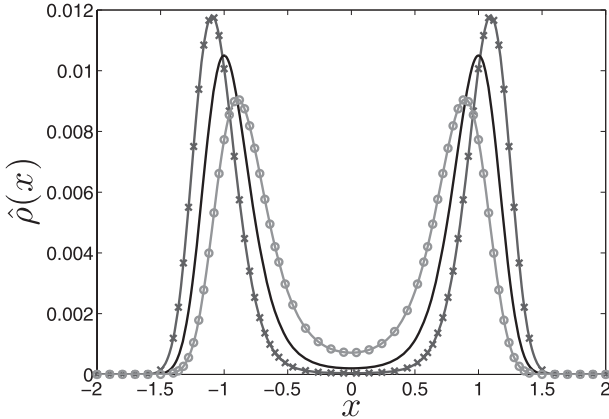
FIG. A1. Invariant density (20) for the stochastic 1D climate model (5), with $\sigma^2 = 0.126$ and $(a, b) = (1, 1)$ (solid line), $(1, 0.8)$ (circles), and $(1, 1.2)$ (crosses).

APPENDIX

Sensitivity of the Stochastic Climate Model to Uncertainties in the Drift Term

In this appendix we examine how the statistics of the stochastic 1D climate model (5) depends on uncertainties in the estimation of the drift term

$$d(X) = ax(b - x^2).$$

The parameter b controls the location of the metastable states near $x^\star = \pm\sqrt{b}$ and their separation. The height of the potential barrier $\Delta V(x) = ab^2/4$ is controlled by a and b.

In Fig. A1 we show the invariant density (20) for the stochastic 1D climate model with $\sigma^2 = 0.126$ and several combinations of the drift term parameters a and b. The invariant density is more sensitive to changes in b than in a (not shown) since b simultaneously affects the distance of the minima of the potential and the height of the potential barrier. The location of the maxima of the probability density function changes by approximately 22% and the actual values of the maxima change by approximately 27% over the range of b values shown. Varying a produces similar changes in $\hat{\rho}(x)$, but as expected the locations of the maxima of the probability density function are not shifted.

As for the diffusion coefficient, differences in the approximation of the empirical density imply a sensitivity of the statistics to changes in the drift term parameters. In Table A1 we show how the characteristic time scales

TABLE A1. Characteristic time scales of the stochastic 1D climate model (5) with $\sigma^2 = 0.126$ for different values of the drift coefficients a and b.

	$a = 0.8$, $b = 1$	$a = 1.2$, $b = 1$	$a = 1$, $b = 0.8$	$a = 1$, $b = 1.2$
τ_{corr}	77.3	256.3	41.1	645.7
τ_e	43.7	136.1	31.1	212.9
τ_t	6.4	4.8	7.2	4.5

change for varying drift term parameters. Note that now the transit time τ_t varies because by varying a we modify the rate of divergence of the unstable fixpoint.

We now investigate how uncertainties in the drift term parameters a and b may affect the skill $\mathcal{S}$ in an ETKF data assimilation procedure. Figure A2 shows skill curves for fixed value of the diffusion coefficient $\sigma^2 = 0.126$, for different combinations of a and b. As with the characteristic time scales, the skill is more sensitive to changes in b than in a. The sensitivity to changes in the drift term is of a similar nature to the one for diffusion. Whereas decreasing a or b by 20% produces little change in skill, increasing a or b by 20% makes the stochastic 1D climate model again less skillful than the full deterministic 4D model. This is readily explained by noticing that increasing b increases the difference between the two metastable states and also the height of the potential barrier; therefore, for fixed diffusion, this inhibits transitions between the metastable states, thereby decreasing the spread of the ensemble. Similarly, decreasing a (while keeping b fixed) increases the
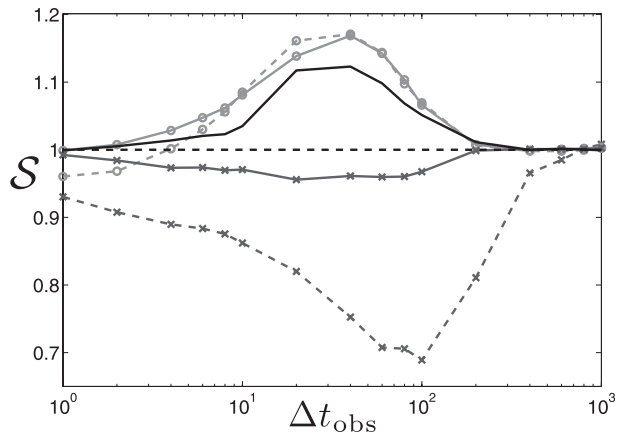

FIG. A2. Proportional skill improvement $\mathcal{S}$ of the stochastic 1D climate model (5) over the full deterministic 4D model (1)–(4) as a function of the observation interval Δt_{obs} for $\sigma^2 = 0.126$, with $(a, b) = (1, 1)$ (black solid line), $(0.8, 1)$ (solid, circles), $(1.2, 1)$ (solid, crosses), $(1, 0.8)$ (dashed, circles), and $(1, 1.2)$ (dashed, crosses). All other parameters are as in Fig. 8.

potential barrier, inhibiting transitions and thereby decreasing the ensemble spread.

As for the diffusion coefficient, the values of the parameters of the drift term that are closest to the actual values $a = b = 1$ for which the stochastic climate model best approximates the slow dynamics of the full deterministic 4D model are not the optimal values in terms of skill improvement. Values that increase the ensemble spread are more favorable.

REFERENCES

Ambaum, M., 2008: Unimodality of wave amplitude in the Northern Hemisphere. *J. Atmos. Sci.,* **65,** 1077–1086.

Anderson, J. L., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate,* **9,** 1518–1530.

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

——, 2007: An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus,* **59A,** 210–224.

——, 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus,* **61A,** 72–83.

——, 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.

Bergemann, K., G. A. Gottwald, and S. Reich, 2009: Ensemble propagation and continuous matrix factorization algorithms. *Quart. J. Roy. Meteor. Soc.,* **135,** 1560–1572.

Berner, J., 2005: Linking nonlinearity and non-Gaussianity of planetary wave behavior by the Fokker–Planck equation. *J. Atmos. Sci.,* **62,** 2098–2117.

——, and G. Branstator, 2007: Linear and nonlinear signatures in the planetary wave dynamics of an AGCM: Probability density functions. *J. Atmos. Sci.,* **64,** 117–136.

Branstator, G., and J. Berner, 2005: Linear and nonlinear signatures in the planetary wave dynamics of an AGCM: Phase space tendencies. *J. Atmos. Sci.,* **62,** 1792–1811.

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.

Cheng, X., and J. M. Wallace, 1993: Cluster analysis of the Northern Hemisphere wintertime 500-hPa height field: Spatial patterns. *J. Atmos. Sci.,* **50,** 2674–2696.

Cornick, M., B. Hunt, E. Ott, H. Kurtuldu, and M. F. Schatz, 2009: State and parameter estimation of spatiotemporally chaotic systems illustrated by an application to Rayleigh–Bénard convection. *Chaos,* **19,** 013108, doi:10.1063/1.3072780.

Corti, S., F. Molteni, and T. Palmer, 1999: Signature of recent climate change in frequencies of natural atmospheric circulation regimes. *Nature,* **398,** 799–802.

Crommelin, D. T., 2003: Regime transitions and heteroclinic connections in a barotropic atmosphere. *J. Atmos. Sci.,* **60,** 229–246.

——, and E. Vanden-Eijnden, 2006: Reconstruction of diffusions using spectral data from time series. *Commun. Math. Sci.,* **4,** 651–668.

——, and ——, 2008: Subgrid-scale parameterization with conditional Markov chains. *J. Atmos. Sci.,* **65,** 2661–2675.

——, J. D. Opsteegh, and F. Verhulst, 2004: A mechanism for atmospheric regime behavior. *J. Atmos. Sci.,* **61,** 1406–1419.

Ditlevsen, P. D., 1999: Observation of α-stable noise induced millennial climate changes from an ice-core record. *Geophys. Res. Lett.,* **26,** 1441–1444.

Egger, J., and T. Jónsson, 2002: Dynamic models for Icelandic meteorological data sets. *Tellus,* **54A,** 1–13.

Ehrendorfer, M., 2007: A review of issues in ensemble-based Kalman filtering. *Meteor. Z.,* **16,** 795–818.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99** (C5), 10 143–10 162.

——, 2006: *Data Assimilation: The Ensemble Kalman Filter.* 2nd ed. Springer, 330 pp.

Franzke, C., and A. J. Majda, 2006: Low-order stochastic mode reduction for a prototype atmospheric GCM. *J. Atmos. Sci.,* **63,** 457–479.

——, ——, and E. Vanden-Eijnden, 2005: Low-order stochastic mode reduction for a realistic barotropic model climate. *J. Atmos. Sci.,* **62,** 1722–1745.

——, D. T. Crommelin, A. Fischer, and A. J. Majda, 2008: A hidden Markov model perspective on regimes and metastability in atmospheric flows. *J. Climate,* **21,** 1740–1757.

——, I. Horenko, A. J. Majda, and R. Klein, 2009: Systematic metastable atmospheric regime identification in an AGCM. *J. Atmos. Sci.,* **66,** 1997–2012.

Frederiksen, J. S., and A. G. Davies, 1997: Eddy viscosity and stochastic backscatter parameterizations on the sphere for atmospheric circulation models. *J. Atmos. Sci.,* **54,** 2475–2492.

Gardiner, C. W., 2004: *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences.* 3rd ed. Springer, 414 pp.

Givon, D., R. Kupferman, and A. Stuart, 2004: Extracting macroscopic dynamics: Model problems and algorithms. *Nonlinearity,* **17,** R55, doi:10.1088/0951-7715/17/6/R01.

Golub, G. H., and C. F. V. Loan, 1996: *Matrix Computations.* 3rd ed. The Johns Hopkins University Press, 694 pp.

Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.,* **129,** 550–560.

——, and S. J. Colucci, 1997: Verification of Eta/RSM short-range ensemble forecasts. *Mon. Wea. Rev.,* **125,** 1312–1327.

Harlim, J., and A. J. Majda, 2008: Filtering nonlinear dynamical systems with linear stochastic models. *Nonlinearity,* **21,** 1281–1306.

——, and ——, 2010: Catastrophic filter divergence in filtering nonlinear dissipative systems. *Commun. Math. Sci.,* **8,** 27–43.

Hasselmann, K., 1976: Stochastic climate models. Part 1: Theory. *Tellus,* **28,** 473–485.

Hsu, C. J., and F. Zwiers, 2001: Climate change in recurrent regimes and modes of Northern Hemisphere atmospheric variability. *J. Geophys. Res.,* **106** (D17), 20 145–20 159.

Imkeller, P., and J.-S. von Storch, Eds., 2001: *Stochastic Climate Models.* Birkhauser Verlag, 398 pp.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability.* Cambridge University Press, 341 pp.

Khasminsky, R. Z., 1966: On stochastic processes defined by differential equations with a small parameter. *Theory Prob. Appl.,* **11,** 211–228.

Kimoto, M., and M. Ghil, 1993: Multiple flow regimes in the Northern Hemisphere winter. Part I: Methodology and hemispheric regimes. *J. Atmos. Sci.,* **50,** 2625–2644.

Kleeman, R., 2008: Stochastic theories for the irregularity of ENSO. *Philos. Trans. Roy. Soc.,* **366A,** 2509–2524.

Kondrashov, D., K. Ide, and M. Ghil, 2004: Weather regimes and preferred transition paths in a three-level quasigeostrophic model. *J. Atmos. Sci.,* **61,** 568–587.

Kramers, H. A., 1940: Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica,* **7,** 284–304.

Kurtz, T. G., 1973: A limit theorem for perturbed operator semigroups with applications to random evolutions. *J. Funct. Anal.,* **12,** 55–67.

Kwasniok, F., and G. Lohmann, 2009: Deriving dynamical models from paleoclimatic records: Application to glacial millennial-scale climate variability. *Phys. Rev. E,* **80,** 066104, doi:10.1103/PhysRevE.80.066104.

Legras, B., and M. Ghil, 1985: Persistent anomalies, blocking and variations in atmospheric predictability. *J. Atmos. Sci.,* **42,** 433–471.

Leith, C. E., 1975: Climate response and fluctuation dissipation. *J. Atmos. Sci.,* **32,** 2022–2026.

Li, H., E. Kalnay, and T. Miyoshi, 2009: Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.,* **135,** 523–533.

Lin, J. W.-B., and J. D. Neelin, 2000: Influence of a stochastic moist convective parametrization on tropical climate variability. *Geophys. Res. Lett.,* **27,** 3691–3694.

——, and ——, 2002: Considerations for stochastic convective parameterization. *J. Atmos. Sci.,* **59,** 959–975.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.,* **20,** 130–141.

——, 2006: Predictability—A problem partly solved. *Predictability,* T. Palmer, Ed., Cambridge University Press, 40–58.

Luzzatto, S., I. Melbourne, and F. Paccaut, 2005: The Lorenz attractor is mixing. *Commun. Math. Phys.,* **260,** 393–401.

Majda, A. J., I. Timofeyev, and E. Vanden Eijnden, 1999: Models for stochastic climate prediction. *Proc. Natl. Acad. Sci. USA,* **96,** 14 687–14 691.

——, ——, and ——, 2001: A mathematical framework for stochastic climate models. *Commun. Pure Appl. Math.,* **54,** 891–974.

——, ——, and E. Vanden-Eijnden, 2003: Systematic strategies for stochastic mode reduction in climate. *J. Atmos. Sci.,* **60,** 1705–1722.

——, C. Franzke, A. Fischer, and D. T. Crommelin, 2006: Distinct metastable atmospheric regimes despite nearly Gaussian statistics: A paradigm model. *Proc. Natl. Acad. Sci. USA,* **103,** 8309–8314.

——, ——, and B. Khouider, 2008: An applied mathematics perspective on stochastic modelling for climate. *Philos. Trans. Roy. Soc.,* **366A,** 2427–2453.

Melbourne, I., and A. M. Stuart, 2011: A note on diffusion limits of chaotic skew product flows. *Nonlinearity,* **24,** 1361–1367.

Palmer, T. N., 2001: A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Quart. J. Roy. Meteor. Soc.,* **127,** 279–304.

——, and P. Williams, Eds., 2010: *Stochastic Physics and Climate Modelling.* Cambridge University Press, 480 pp.

Papanicolaou, G. C., 1976: Some probabilistic problems and methods in singular perturbations. *Rocky Mt. J. Math.,* **6,** 653–674.

Pavliotis, G. A., and A. M. Stuart, 2007: Parameter estimation for multiscale diffusions. *J. Stat. Phys.,* **127,** 741–781.

——, and ——, 2008: *Multiscale Methods: Averaging and Homogenization.* Springer, 307 pp.

Plant, R. S., and G. C. Craig, 2008: A stochastic parametrization for deep convection based on equilibrium statistics. *J. Atmos. Sci.,* **65,** 87–105.

Shutts, G. J., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.,* **131,** 3079–3102.

Siegert, S., R. Friedrich, and J. Peinke, 1998: Analysis of data sets of stochastic systems. *Phys. Lett. A,* **243,** 275–280.

Simon, D. J., 2006: *Optimal State Estimation.* John Wiley & Sons, 526 pp.

Stemler, T., J. P. Werner, H. Benner, and W. Just, 2007: Stochastic modeling of experimental chaotic time series. *Phys. Rev. Lett.,* **98,** 044102, doi:10.1103/PhysRevLett.98.044102.

Stephenson, D., A. Hannachi, and A. O'Neill, 2004: On the existence of multiple climate regimes. *Quart. J. Roy. Meteor. Soc.,* **130,** 583–605.

Sura, P., 2003: Stochastic analysis of Southern and Pacific Ocean sea surface winds. *J. Atmos. Sci.,* **60,** 654–666.

——, and J. Barsugli, 2002: A note on estimating drift and diffusion parameters from time series. *Phys. Lett. A,* **305,** 304–311.

Talagrand, O., R. Vautard, and B. Strauss, 1997: Evaluation of probabilistic prediction systems. *Proc. ECMWF Workshop on Predictability,* Reading, United Kingdom, ECMWF, 1–25.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.,* **131,** 1485–1490.

Tucker, W., 1999: The Lorenz attractor exists. *C. R. Acad. Sci.,* **328,** 1197–1202.

Wang, X., C. H. Bishop, and S. J. Julier, 2004: Which is better, an ensemble of positive–negative pairs or a centered spherical simplex ensemble? *Mon. Wea. Rev.,* **132,** 1590–1605.

Wilks, D. S., 2005: Effects of stochastic parametrizations in the Lorenz '96 system. *Quart. J. Roy. Meteor. Soc.,* **131,** 389–407.

——, 2006: *Statistical Methods in the Atmospheric Sciences.* 2nd ed. Elsevier, 627 pp.

Wirth, V., 2001: Detection of hidden regimes in stochastic cyclostationary time series. *Phys. Rev. E,* **64,** 016136, doi:10.1103/PhysRevE.64.016136.

Zwanzig, R., 2001: *Nonequilibrium Statistical Mechanics.* Oxford University Press, 222 pp.