Subgrid scale parameterization with conditional Markov chains

Daan Crommelin *
CWI, Amsterdam, the Netherlands

Eric Vanden-Eijnden
Courant Institute, New York University, New York, USA

Revised version
To appear in J. Atmos. Sci.

Corresponding author address: CWI, PO Box 94079, 1090 GB, Amsterdam, the Netherlands.
E-mail: Daan.Crommelin@cwi.nl

Abstract

A new approach is proposed for stochastic parameterization of subgrid scale processes in models of atmospheric or oceanic circulation. The new approach relies on two key ingredients. First, the unresolved processes are represented by a Markov chain whose properties depend on the state of the resolved model variables. Second, the properties of this conditional Markov chain are inferred from data. We test the parameterization approach by implementing it in the framework of the Lorenz 96 model. We assess performance of the parameterization scheme by inspecting probability distributions, correlation functions and wave properties, and by carrying out ensemble forecasts. For the Lorenz 96 model, the parameterization algorithm is shown to give good results with a Markov chain with a few states only, and to outperform several other parameterization schemes.

1. Introduction: stochastic parameterization of subgrid scale processes

The parameterization of subgrid scale processes in models of atmospheric flow has drawn a lot of research attention recently. To get beyond the limitations of parameterizations with deterministic functions, the focus of various recent investigations has been on the potential of stochastic methods for the parameterization of processes that cannot be resolved because they fall below the model grid scale (e.g. Majda et al. (1999); Buizza et al. (1999); Lin and Neelin (2000); Palmer (2001); Lin and Neelin (2002); Majda and Khouider (2002); Majda et al. (2003); Khouider et al. (2003); Wilks (2005); Plant and Craig (2007)).

When trying to parameterize unresolved processes stochastically, one is faced with two main issues. The first is to determine the class of models one wants to use for the subgrid processes. Several directions have been proposed and used in the literature, ranging from stochastic differential equations (Majda et al. (1999, 2003); Wilks (2005)) to cellular automata (Shutts (2005), see also Palmer (2001)), multiplicative randomization of deterministic parameterization schemes (Buizza et al. (1999)) and Markov chain models on a discrete state space (Majda and Khouider (2002); Khouider et al. (2003)). This last class of models using Markov chains is the one that we will use in the present paper, as explained below in more detail.

The second issue one faces is how to choose the parameters in the model for the subgrid processes. This can either be done based on physical intuition (like e.g. in Lin and Neelin (2000); Majda and Khouider (2002)) or by using directly the data from the time series for the subgrid processes (e.g. Wilks (2005)). The second approach has the advantage that it typically allows one to make less ad-hoc assumptions about the subgrid processes. This may be less transparent from a physical perspective, but it is also potentially more accurate. The approach proposed in this paper uses the data of the subgrid processes to obtain a stochastic model. The related, more general problem of inferring stochastic models from data is considered in a wide variety of papers; for applications in atmosphere-ocean science, see for example Penland and Matrosova (1994); Egger (2001); Sura (2003); Crommelin (2004); Berner (2005).

Specifically, we propose a strategy for stochastic parameterization which uses Markov processes that are conditional on the state of the resolved variables. This strategy accounts, as all stochastic parameterizations do, for the possibility to have a multiplicity of possible states of the unresolved processes given one fixed state of the resolved variables. It also accounts for the possibility that the properties of the unresolved processes (for example variance, skewness and decorrelation time) vary with the state of the resolved variables. We give an algorithm that translates this dependency into a computationally feasible parameterization scheme. The conditional stochastic processes are represented as Markov chains with a small number of states, making the practical implementation easy. The properties (such as transition probabilities) of the Markov chains are estimated from data in a simple, straightforward way. The algorithm is applied to the Lorenz 96 model (Lorenz (1995)), a frequently used testbed for parameterization strategies (Palmer (2001); Fatkullin and Vanden-Eijnden (2004); Wilks (2005)).

We note that the type of models resulting from the stochastic parameterization developed in this paper have a structure resembling the coupled models proposed by Majda and collaborators in Katsoulakis et al. (2003, 2005, 2006) (in the context of material science) and Majda and Khouider (2002); Khouider et al. (2003) (focussing on tropical convection). There, a systematic subgrid scale parametrization is proposed in which deterministic equations for macroscopic variables are coupled to a stochastic Ising (spin flip) system that represents microscopic phe-

nomena. The spin flip system is a Markov chain which can be coarse-grained using a systematic closure procedure. Similar to our approach, the resulting models consist of deterministic differential equations coupled to Markov chains that are conditional on the state of the macroscopic variables. The main difference is that, in the present paper, the parameterization (or closure) is entirely inferred from data without using any knowledge of the physics or equations that drive the subgrid scales. In the approach by Majda and collaborators, on the other hand, one starts with an explicitly known microscopic model which can be coarse-grained because the equilibrium distribution of the microscopic model is known exactly (under the assumption of detailed balance); there is no attempt to determine the parameters in the model from actual data.

The outline of the remainder of this paper is as follows. In section 2 we give a brief description of the Lorenz 96 model. The stochastic parameterization strategy with data-inferred conditional Markov processes is introduced in section 3. In section 4 practical aspects of the parameterization scheme are discussed, such as data inference and integration scheme. We present numerical results in section 5, where we compare our scheme with other parameterization schemes by inspecting probability distributions, correlation functions, wave statistics and results from ensemble integrations (both ensemble mean and dispersion). We conclude in section 6 with a discussion, where we also address the question how the stochastic parameterization scheme could be used in more realistic models.

2. The Lorenz 96 model: a testbed for parameterization algorithms

The 2-layer Lorenz 96 (L96) model, proposed by Lorenz in 1995 (Lorenz (1995)), has become a popular toy model of the atmosphere to test various concepts and ideas relating to predictability, model error and parameterization (Boffetta et al. (1998); Palmer (2001); Orrell (2003); Fatkullin and Vanden-Eijnden (2004); Wilks (2005)). The model equations read

$$\dot{X}_k = X_{k-1}(X_{k+1} - X_{k-2}) - X_k + F + B_k \tag{1a}$$

$$\dot{Y}_{j,k} = \frac{1}{\varepsilon}\big(Y_{j+1,k}(Y_{j-1,k} - Y_{j+2,k}) - Y_{j,k} + h_y X_k\big) \tag{1b}$$

in which

$$B_k = \frac{h_x}{J}\sum_{j=1}^{J} Y_{j,k}, \tag{2}$$

and $k = 1, \ldots, K;\ j = 1, \ldots, J$. The X_k and $Y_{j,k}$ are interpreted as variables on a circle of constant latitude, where the X_k are "large-scale" variables, each coupled to a collection of J "small-scale" variables $Y_{j,k}$. The indices k and j can be regarded as spatial indices. The periodicity of the spatial domain is reflected in the periodicity of the variables:

$$X_k = X_{k+K} \tag{3a}$$

$$Y_{j,k} = Y_{j,k+K} \tag{3b}$$

$$Y_{j+J,k} = Y_{j,k+1} \tag{3c}$$

In Lorenz (1995), the model is formulated slightly differently from (1), with parameter settings being equivalent to $\varepsilon = 0.1$, $K = 36$, $J = 10$, $F = 10$, $h_x = -1$ and $h_y = 1$. We use the formulation above (from Fatkullin and Vanden-Eijnden (2004)) because it makes the time

scale separation (measured by ε) between the X_k and the $Y_{j,k}$ explicit. If $\varepsilon \ll 1$ the X_k are slow variables and the $Y_{j,k}$ are fast; if $\varepsilon \approx 1$ there is no time scale separation. Studies using the L96 system are often carried out using parameter settings that amount to such time scale separation (Lorenz (1995); Fatkullin and Vanden-Eijnden (2004); Wilks (2005)). In this paper we use $\varepsilon = 0.5$ since (near-)absence of time scale separation between resolved and unresolved processes is both more realistic and more difficult to handle for parameterizations. See Fatkullin and Vanden-Eijnden (2004) and references therein for an overview (and implementation for L96) of recently developed computational strategies to handle the case $\varepsilon \ll 1$.

Using data from a numerical simulation of the full L96 system (1), a scatter plot of B_k versus X_k is shown in figure 1. The parameters used in the simulation are $(\varepsilon, K, J, F, h_x, h_y) = (0.5, 18, 20, 10, -1, 1)$. Because of the translation invariance of the system, this scatter plot has the same statistical properties for all values of k. From the figure, it is clear that for a fixed value of X_k, B_k can take on a range of values. The properties of $\rho(B_k|X_k)$, the probability density function (PDF) of B_k conditional on the value of X_k, are obviously highly dependent on X_k. In figure 2, various PDFs of B_k are shown, estimated from data points with X_k in different intervals.

3. Parameterization with conditional Markov processes

Within the context of the L96 system (1), the aim of a parameterization scheme is to formulate a model for the large-scale variables X_k alone, from which the variables $Y_{j,k}$ have disappeared entirely. The key element is a suitable representation of the quantities B_k. In the full L96 model (1) they depend on the $Y_{j,k}$; in a reduced model for the X_k variables alone, they must be parameterized in terms of the X_k.

One way of parameterizing is deterministic, with a function $\mathbf{B} = \mathbf{G}(\mathbf{X})$. In its most general form, every element B_k of the vector $\mathbf{B}$ is determined by *all* elements X_k of the state vector $\mathbf{X}$. Since the function $\mathbf{G}$ can be nonlinear, this type of parameterization is often too complicated to be of practical use for systems with more than a few degrees of freedom (see however Fatkullin and Vanden-Eijnden (2004) for a on-the-fly computational strategy to handle this problem when $\varepsilon \ll 1$). For complex systems, a simplified function such as $B_k = g(X_k)$ is typically considered. The assumption that B_k is determined by X_k, and not by variables at other gridpoints $k' \neq k$, can be seen as a "locality" assumption.

A more fundamental problem of deterministic parameterizations stems from the chaotic nature of the underlying systems: a function such as $B_k = g(X_k)$ cannot account for the possibility that in the full system, B_k can take on a variety of states given a fixed X_k, rather than one unique state. Put differently: the probability distribution of B_k with X_k fixed is often not a Dirac delta distribution. It is therefore natural to consider stochastic parameterizations, in which the B_k are modeled as stochastic processes. They allow for different realizations of B_k for a fixed value of X_k, consistent with the scatter plot of figure 1.

We propose here a new approach to stochastic parameterization, in which we make the following two key assumptions about the stochastic process that replaces B_k: (i) it is a Markov process; (ii) the process is conditional on $\mathbf{X}$. We infer the properties of the stochastic process from the (X_k, B_k) data of the full L96 model. To make this parameterization scheme computationally as simple as possible, we restrict the conditionality in practice to X_k. Adding conditionality on other gridpoints (for example, the nearest neighbor gridpoints X_{k-1} and X_{k+1}) can be expected to improve the performance of the reduced model; however, it will also make the

model parameterization more complicated. We leave exploration of this possibility to another study.

In a nutshell, we model B_k by numerical (Monte Carlo) simulation of the stochastic process with conditional transition probability

$$P(B_k(t_2) \,|\, X_k(t_2), B_k(t_1), X_k(t_1)) \quad \text{with} \quad t_2 \geq t_1\,, \tag{4}$$

which will be estimated from the (X_k, B_k) data (see section 4 for details). We let $X_k(t_2)$ be determined by integration of (1a), starting from $\mathbf{X}(t_1)$ and with B_k fixed at $B_k(t_1)$. Thus, given $\mathbf{X}(t_1)$ and $\mathbf{B}(t_1)$, first we integrate (1a) to obtain $\mathbf{X}(t_2)$, then we obtain $\mathbf{B}(t_2)$ by Monte Carlo simulations of (4) for each k. More practical aspects of this parameterization scheme are discussed in the next section.

Adding $X_k(t_2)$ to the conditions $X_k(t_1)$ and $B_k(t_1)$ in (4) is natural because the distribution for $B_k(t_2)$ depends strongly on the direction in which X_k is moving. Note that the value of $X_k(t_2))$ information is readily available via integration of (1a) on $t \in [t_1, t_2]$ with $B_k(t) = B_k(t_1)$. As an illustration, we calculated PDFs for B_k on the interval $1.5 < X_k < 2.5$ depending on whether X_k is growing or decreasing in value. More precisely, we calculated the PDFs $\rho(B_k(t) \,|\, 1.5 < X_k(t) < 2.5, X_k(t) > X_k(t - \Delta t))$ and $\rho(B_k(t) \,|\, 1.5 < X_k(t) < 2.5, X_k(t) < X_k(t - \Delta t)))$. They are shown, together with the total PDF $\rho(B_k(t) \,|\, 1.5 < X_k(t) < 2.5)$, in figure 3. The significant differences between these PDFs make clear why a parameterization with the conditional transition probability $P(B_k(t_2) \,|\, B_k(t_1), X_k(t_1))$ instead of (4) would result in a less accurate model in the case of L96. We expect the inclusion of $X_k(t_2)$ as a condition to be less important for models that are more strongly mixing, so that $\rho(B_k(t_2) \,|\, B_k(t_1), X_k(t_1))$ is roughly the same when X_k is growing as it is when X_k is decreasing.

The parameterization we propose has some similarities with the stochastic parameterization for L96 studied in Wilks (2005). There, the B_k are modeled by a deterministic term (a function $g(X_k)$ found by fitting a polynomial to the (X_k, B_k) data) plus a stochastic term coming from an AR(1) process. The AR(1) process itself is not conditional on X_k. However, since the stochastic term is added to a deterministic one, the scheme studied in Wilks (2005) is equivalent to parameterization with an AR(1) process with X_k-dependent mean. The parameterization approach proposed here is more general because the stochastic process we use need not be of AR(1) type, and its conditionality on X_k is not restricted to the mean but extends to, for example, variance, skewness and decorrelation time. In section 5 we will show that our approach outperforms the one proposed in Wilks (2005), at least for the example of L96 in the parameter regime that we consider.

4. A practical algorithm using Markov chains

In this paper, we choose to model the conditional Markov process that replaces B_k as a collection of Markov chains. As a result, B_k becomes a discrete random variable. One can also choose to model the B_k with Markov processes that are continuous in space (e.g., diffusion processes); we use Markov chains here because of their computational ease. Even with a small number of Markov chain states, statistical properties such as variance, skewness and decorrelation time can be captured. Moreover, data inference of Markov chains is much easier to carry out than inference of (possibly non-Gaussian) continuous stochastic processes from data. The conditionality on $X_k(t_1)$ and $X_k(t_2)$ is implemented by letting the properties of the Markov chain change stepwise (not continuously), depending on the intervals in which $X_k(t_1)$

and $X_k(t_2)$ reside. It must be stressed that although the B_k have become discrete variables, and although the Markov chain properties change discretely rather than continuously with X_k, the resulting (reduced) model for the resolved variables (the X_k) is still continuous.

a. Estimation of the Markov chains

For the construction of the Markov chains, all data points (X_k, B_k) are assigned to discrete states (i, n) by partitioning the (X_k, B_k) plane into bins. First, the range of possible values of X_k is divided into N_X non-overlapping intervals $\mathcal{I}_i$ ($i = 1, ..., N_X$). Within each interval $\mathcal{I}_i$, the range of values of B_k is divided into N_B non-overlapping intervals $\mathcal{J}_n^i$ ($n = 1, ..., N_B$). A set of stochastic matrices is constructed by estimating a spatially discrete version of the conditional transition probability (4) from the data:

$$P_{nm}^{(ij)} = P\left(B_k(t+\Delta t) \in \mathcal{J}_m^j \,|\, X_k(t) \in \mathcal{I}_i, B_k(t) \in \mathcal{J}_n^i, X_k(t+\Delta t) \in \mathcal{I}_j\right) \tag{5a}$$

$$= \frac{T_{in|jm}}{\sum_{m=1}^{N_B} T_{in|jm}} \tag{5b}$$

The object $T_{in|jm}$ counts the transitions from (i, n) to (j, m) in the data:

$$T_{in|jm} = \sum_t \mathbf{1}[B_k(t) \in \mathcal{J}_n^i]\, \mathbf{1}[X_k(t) \in \mathcal{I}_i]\, \mathbf{1}[B_k(t+\Delta t) \in \mathcal{J}_m^j]\, \mathbf{1}[X_k(t+\Delta t) \in \mathcal{I}_j] \tag{6}$$

where $\sum_t$ denotes the sum over all discrete times $t = 0, \Delta t, 2\Delta t, \ldots$ in the dataset and $\mathbf{1}[\cdot]$ denotes the indicator function: $\mathbf{1}[B_k(t) \in \mathcal{J}_n^i] = 1$ if $B_k(t) \in \mathcal{J}_n^i$ and $\mathbf{1}[B_k(t) \in \mathcal{J}_n^i] = 0$ if $B_k(t) \notin \mathcal{J}_n^i$, etcetera. For simplicity, we assume that the time interval Δt between two data points is constant throughout the dataset.

For any i and j fixed, the matrix $P^{(ij)}$ satisfies

$$\forall n, m \;:\; P_{nm}^{(ij)} \geq 0, \qquad \forall n \;:\; \sum_{m=1}^{N_B} P_{nm}^{(ij)} = 1, \tag{7}$$

so $P^{(ij)}$ is an $(N_B \times N_B)$ stochastic matrix (note that i and j themselves run from 1 to N_X). The discrete states (i, n) are assigned a value $\mathcal{B}_n^i$ for B_k that is the average of all data points falling in bin $\mathcal{J}_n^i$:

$$\mathcal{B}_n^i = \frac{\sum_t B_k(t)\, \mathbf{1}[B_k(t) \in \mathcal{J}_n^i]\, \mathbf{1}[X_k(t) \in \mathcal{I}_i]}{\sum_t \mathbf{1}[B_k(t) \in \mathcal{J}_n^i] \mathbf{1}[X_k(t) \in \mathcal{I}_i]} \tag{8}$$

Construction of the transition probability matrices $P^{(ij)}$ is not possible if $|i - j|$ is too large. The (X_k, B_k) data do not contain pairs of consecutive points with $X_k(t) \in \mathcal{I}_i$, $X_k(t+\Delta t) \in \mathcal{I}_j$ if i and j are far apart and the time step Δt is small. However, this is not a problem, because transitions between distant intervals $\mathcal{I}_i$ and $\mathcal{I}_j$ are also unlikely to occur during an integration of the reduced L96 model. In practice, the only matrices $P^{(ij)}$ that are really needed are those with $|i - j| \leq 1$. If the (X_k, B_k) data does not contain any transition out of (i, n) into j for some combination of i, j, n, we have $\sum_m T_{in|jm} = 0$. In that case, we set $P_{nm}^{(ij)} = \delta_{nm}$ for that particular triplet i, j, n (where δ_{nm} is the Kronecker delta). This avoids the numerical integration to break off on the rare occasion that a transition occurs for which we have no data. A more sophisticated way of dealing with this problem involves summation over all intermediate steps $(i, n) \to (i', n') \to (j, m)$ with known transition probabilities. We will not do this here.

b. Time step of the Markov chains

The stochastic matrices $P^{(ij)}$ are made up of the transition probabilities (5a) over a fixed time interval Δt. This time interval is the interval at which the (X_k, B_k) data is sampled. However, in practical situations it is possible that the data that is available for the construction of a parameterization scheme is sampled with a time step Δt that is different from the time step δt at which the numerical model will be integrated. For example, Δt can be so large that an integration scheme with $\delta t = \Delta t$ would be numerically unstable.

There are two ways of dealing with this problem of non-matching time steps. One possibility is to use a split integration scheme. Assume for simplicity that Δt is an integer multiple of δt, say $\Delta t = N^t\, \delta t$. We integrate X_k with time step δt and update B_k once every N^t time steps. Of course, the transition probability matrix $P^{(ij)}$ to be used to evolve B_k in time must have i and j such that $X_k(t_1) \in \mathcal{I}_i$ and $X_k(t_2) \in \mathcal{I}_j$ with $t_2 - t_1 = N^t\, \delta t$ and not $t_2 - t_1 = \delta t$.

Another possibility to tackle the problem is by calculating transition probability matrices $\tilde{P}^{(ij)}$ with an associated time step δt, starting from matrices $P^{(ij)}$ with time step Δt. Unfortunately, the straightforward but naive choice $\tilde{P}^{(ij)} = (P^{(ij)})^{\delta t/\Delta t}$ does not, in general, result in a new stochastic matrix (some matrix elements may be negative, or complex) because of the Markov chain embedding problem (see Bladt and Sørensen (2005); Crommelin and Vanden-Eijnden (2006) for a discussion). In order to find a matrix $\tilde{P}^{(ij)}$ that is both a true stochastic matrix, and approximately equal to $(P^{(ij)})^{\delta t/\Delta t}$, one should first construct a so-called generator matrix L from $P^{(ij)}$. L must be such that (i) $\exp(\Delta t\, L)$ resembles $P^{(ij)}$ as closely as possible, and (ii) $\exp(\delta t\, L)$ qualifies as a stochastic matrix for all $\delta t \geq 0$. Two different methods for the numerical construction of generators can be found in Bladt and Sørensen (2005) and Crommelin and Vanden-Eijnden (2006). After L is constructed, the matrix $\tilde{P}^{(ij)} = \exp(\delta t\, L)$ is a true stochastic matrix that approximates $(P^{(ij)})^{\delta t/\Delta t}$ as best as possible. The degree of proximity depends on how close $\exp(\Delta t\, L)$ is to $P^{(ij)}$. The difference between $\exp(\Delta t\, L)$ and $P^{(ij)}$ can be non-negligible for some $P^{(ij)}$, creating a source of error for the parameterization scheme. For these reasons, and because the split integration scheme described before gave good results for L96 (see section 5c), we will not investigate the generator method any further here.

c. Numerical integration

Having the transition matrices $P^{(ij)}$ (or $\tilde{P}^{(ij)}$) for time interval δt available, time integration of the reduced L96 model with conditional Markov chain (CMC) parameterization is done as follows: given the vectors $\mathbf{X}(t)$ and $\mathbf{B}(t)$, the next time iterate $\mathbf{X}(t + \delta t)$ is calculated using the derivative of $\mathbf{X}$ determined by equation (1a), with $\mathbf{B}$ fixed at $\mathbf{B}(t)$. Then the time step for B_k is made by Monte Carlo simulation (independent simulations for different k) of the Markov chain with transition probability matrix $P^{(ij)}$, with i and j such that $X_k(t) \in \mathcal{I}_i$ and $X_k(t + \delta t) \in \mathcal{I}_j$. As mentioned before, the B_k have become discrete variables, because they can only take on a finite number of values (the $\mathcal{B}_n^i$). If B_k is sent from state (i, n) to state (j, m) by the Monte Carlo simulation, it means that the value of B_k to be used in the integration of X_k changes from $\mathcal{B}_n^i$ to $\mathcal{B}_m^j$.

5. Numerical results

We compare the results generated by the reduced L96 model (equation (1a) without (1b)) using three different parameterization schemes for B_k: deterministic (DTM), the AR(1) scheme

as presented in Wilks (2005) (AR1) and the conditional Markov chain scheme (CMC) described in the previous sections. We generate data for X_k and B_k by integrating the full L96 model (1) with parameters $(\varepsilon, K, J, F, h_x, h_y) = (0.5, 18, 20, 10, -1, 1)$. Thus, there is hardly any time scale separation between the resolved and the unresolved variables, making the situation more realistic (and more difficult) than would be the case with the more common choice $\varepsilon = 0.1$. We integrate the full L96 model for 10^3 time units with time step 10^{-3} and store the resulting (X_k, B_k) every 0.01 time unit (i.e., $\Delta t = 0.01$). Thus, the data set contains 10^5 points. The reduced models will be integrated with the same time step, $\delta t = 0.01$. In the last part of this section, we consider the case where $\Delta t > \delta t$, which adds a complication as described in section 4b.

For the deterministic (DTM) parameterization scheme, a 5th order polynomial is fitted to the (X_k, B_k) data. The resulting function $B_k = g(X_k)$, shown in figure 4, is used as parameterization of B_k.

The AR1 scheme from Wilks (2005) consists of representing $B_k(t)$ by a deterministic term plus a stochastic term:

$$B_k(t) \approx g(X_k(t)) + \xi(t) \tag{9}$$

For $g(X_k)$ we take the polynomial function also used for the DTM scheme. The stochastic term ξ is generated by an AR(1) process, whose parameters are obtained by fitting the process to the timeseries of $B_k(t) - g(X_k(t))$. The data in this timeseries is sampled at a time interval $\Delta t = 0.01$, and leads to an AR(1) process with mean zero, standard deviation 0.88 and e-folding (decorrelation) time 4.3. Thus, the AR(1) process can be regarded as red noise.

For the CMC scheme, we have to choose the intervals $\mathcal{I}_i$ and $\mathcal{J}_n^i$. Once this is done, we construct the transition matrices $P^{(ij)}$ and the discrete values $\mathcal{B}_n^i$ for B_k from the (X_k, B_k) data, using expressions (5b),(6) and (8). For X_k, we use 16 intervals of width 1, centered at integer values of X_k. At the beginning and end of the domain, we use unbounded intervals. Thus, $\mathcal{I}_1 = (-\infty, -4.5]$, $\mathcal{I}_2 = (-4.5, -3.5]$, $\mathcal{I}_3 = (-3.5, -2.5]$, ... , $\mathcal{I}_{16} = (9.5, +\infty)$. Within each interval $\mathcal{I}^i$, the bins $\mathcal{J}_n^i$ are determined such that each bin contains (approximately) an equal number of data points. We set the number of bins to $N_B = 4$ for all calculations. The resulting values of $\mathcal{B}_n^i$ are plotted in the lower panel of figure 4. Using $N_B = 8$ instead of $N_B = 4$ gave no significant improvement in the results and are therefore not shown.

a. PDFs, correlations and wave statistics

We integrate the reduced L96 models with the various parameterization schemes, and calculate several quantities from the data of the reduced models as well as from the data of the full L96 model. The timeseries are each 10^5 datapoints long with sampling interval $\Delta t = 0.01$. We calculate the following quantities:

- The **PDF** for X_k.
- The **autocorrelation function** (ACF) for X_k.
- The **cross-correlation function** (CCF) for X_k and X_{k+1}.
- By Fourier transforming the state vector $\mathbf{X}$ at every datapoint, a timeseries for the wave number vector $\mathbf{u}$ is obtained. From this timeseries we calculate the **wave variance** $\langle |u_m - \langle u_m \rangle|^2 \rangle$ for every wave number $0 \leq m \leq K/2$, where $\langle . \rangle$ denotes time average.

- The **mean wave amplitude** $\langle |u_m| \rangle$.

For calculation of the PDFs, ACFs and CCFs we take the average over all values of k. The PDF for X_k of the full L96 model is well reproduced by all parameterization schemes (figure 5). Differences between the schemes are quite small here. The ACF (figure 6) and CCF (figure 7), both strongly oscillatory, are more accurately reproduced by the CMC scheme than the other schemes. All schemes give oscillatory ACFs and CCFs, but with the DTM and AR1 schemes, the amplitude of the oscillation is too low, and a phase shift can be seen. The wave variances and mean wave amplitudes of the full L96 model are also better reproduced with the CMC scheme than the other schemes (figure 8). The CMC scheme gives a correct peak in the variance at wavenumber $m = 3$; the other schemes show a lower, broader peak spread out over wavenumbers $m = 3$ and $m = 4$. A similar effect can be seen in the mean wave amplitudes.

In Wilks (2005), the PDF of X_k is also rather well reproduced by various parameterization schemes (DTM, AR1 with white noise and red noise), similar to what we find. Correlation functions and wave statistics, like our figures 6, 7 and 8, are not calculated in Wilks (2005).

b. Ensemble tests

We carry out ensemble integrations for further testing of the CMC parameterization. For each ensemble, we calculate the mean trajectory and compare it with the true trajectory from the full L96 model. This is done by calculating the Root Mean Square Error (RMSE) and the Anomaly Correlation (ANCR) using results from many different ensembles. The number of ensemble members in each ensemble is 1, 5 or 20. For the 20-member ensembles, we also calculate rank histograms to show the ensemble dispersion.

The set-up is as follows. Let $\mathbf{X}_t^{\text{full}}$ be a timeseries for $\mathbf{X}$ of the full L96 model. A number of N_{init} initial states from this timeseries is selected:

$$\mathbf{X}^{\text{init},s} = \mathbf{X}_{t_s}^{\text{full}}, \qquad s = 1, ..., N_{\text{init}} \tag{10}$$

We take one data point every 10 time units from $\mathbf{X}_t^{\text{full}}$, so $t_{s+1} - t_s = 10$. For every initial state, we do N_{ens} integrations of the reduced model over T time units (using the CMC, DTM or AR1 parameterization), starting from $\mathbf{X}^{\text{init},s}$ plus a small random perturbation $\xi^{s,n}$ ($n = 1, ..., N_{\text{ens}}$). We use random perturbations that are drawn from a Gaussian distribution with mean 0 and standard deviation 0.15 (for comparison: the standard deviation of X_k is about 3.5 with the parameter settings we use for the full L96 model). We make no attempt to sample the unstable directions of phase space more heavily than other directions, or to generate ensembles using fastest growing perturbations from singular value decomposition. For the purpose of comparing different parameterization approaches, the simple generation of ensemble members described above is sufficient.

The different integrations from one ensemble are averaged, resulting in a mean trajectory $\mathbf{X}_t^{\text{mean},s}$ (where $t \in [0, T]$ and $s = 1, ..., N_{\text{init}}$). The Root Mean Square Error (RMSE) measures the average difference between the mean ensemble trajectory and the trajectory from the full L96 model:

$$\text{RMSE}\,(\tau) = \left(\frac{1}{N_{\text{init}}} \sum_s |\mathbf{X}_\tau^{\text{mean},s} - \mathbf{X}_{t_s+\tau}^{\text{full}}|^2 \right)^{1/2} \tag{11}$$

For the Anomaly Correlation, we need the anomalies of the full L96 model and the ensemble mean:

$$\mathbf{a}_\tau^{\text{full},s} = \mathbf{X}_{t_s+\tau}^{\text{full}} - \langle \mathbf{X}^{\text{full}} \rangle, \tag{12}$$

$$\mathbf{a}_\tau^{\text{mean},s} = \mathbf{X}_{\tau}^{\text{mean},s} - \langle \mathbf{X}^{\text{full}} \rangle, \tag{13}$$

where $\langle \mathbf{X}^{\text{full}} \rangle$ is the time mean of $\mathbf{X}_t^{\text{full}}$. The Anomaly Correlation (ANCR) is

$$\text{ANCR}\,(\tau) = \frac{1}{N_{\text{init}}} \sum_s \frac{\mathbf{a}_\tau^{\text{full},s} \cdot \mathbf{a}_\tau^{\text{mean},s}}{\sqrt{|\mathbf{a}_\tau^{\text{full},s}|^2\,|\mathbf{a}_\tau^{\text{mean},s}|^2}}, \tag{14}$$

where $\mathbf{a}\cdot\mathbf{b} = \sum_k a_k b_k$ and $|\mathbf{a}|^2 = \mathbf{a}\cdot\mathbf{a}$. The Anomaly Correlation shows the average correlation after time τ between the "true" trajectories (those of the full L96 model) and the mean "forecast" trajectories (the mean ensemble trajectories of the reduced model).

The rank histograms are calculated from ensemble integrations at lead time $\tau = 2$. For each gridpoint k, we rank the $N_{\text{ens}} + 1$ values for X_k from the ensemble members and the full L96 model. Ideally, the distribution of the position of the truth (the value from the full L96 model) in this ranking should approach a uniform distribution (Hamill (2001)); in that case the rank histogram is (nearly) flat. If the ensemble is under-dispersed, the truth occupies the extremes (locations at or near 1 and $N_{\text{ens}} + 1$) too often, leading to a U-shaped rank histogram. In the reverse situation (over-dispersion), the extremes are occupied too rarely, which gives a histogram with the shape of an inverted U. For the rank histograms, we combine the results from all gridpoints k.

The ensemble integrations were carried out using reduced models with CMC, DTM and AR1 parameterization schemes, each starting from $N_{\text{init}} = 1000$ different initial states. For the calculations we took ensemble sizes $N_{\text{ens}} = 1$, $N_{\text{ens}} = 5$ and $N_{\text{ens}} = 20$. Results for RMSE and ANCR are shown in figures 9, 10 and 11. The CMC scheme clearly performs better than the other two schemes. The forecast lead time τ at which the anomaly correlation drops below 0.6 is extended with about 20% when changing from the DTM to the CMC scheme in case $N_{\text{ens}} = 1$, and with about 40% if $N_{\text{ens}} = 20$. From the DTM scheme with $N_{\text{ens}} = 1$ to the CMC scheme with $N_{\text{ens}} = 20$, this extension is about 65%. The CMC scheme with $N_{\text{ens}} = 5$ performs better than the DTM or AR1 scheme with $N_{\text{ens}} = 20$.

Figure 12 shows the rank histograms for $N_{\text{ens}} = 20$. The CMC scheme has a positive effect on the ensemble spread: the corresponding rank histogram is nearly flat. The AR1 scheme gives under-dispersed ensembles; the DTM scheme leads to strong under-dispersion. The rank histograms of the ensembles with AR1 and DTM scheme can be made flatter by increasing the amplitude of the initial state perturbations $\xi^{s,n}$; the ensembles with CMC scheme become over-dispersed if the perturbation amplitude is increased, as could be expected.

It is interesting to note that the AR1 scheme does not perform better than the DTM scheme in our tests. As mentioned, the properties of the AR(1) stochastic process for the AR1 scheme were estimated using the data at the shortest available time interval, $\Delta t = 0.01$. If we carry out the estimation at longer time intervals, resulting in AR(1) processes with shorter e-folding times, the performance of the AR1 parameterization scheme does not improve (results not shown). In Wilks (2005), results are reported where the AR1 scheme performs better than the DTM scheme in ensemble tests. Two important differences with the tests in our study are (i) the parameter settings of the full L96 model used in Wilks (2005) (equivalent to $\varepsilon = 0.1, K = 8, J = 32, h_x = -3.2, h_y = 1$ and $F = 18$ or $F = 20$) and (ii) the set-up of the ensemble tests (different ways to

generate ensemble members). Rerunning our own tests using the parameter settings from Wilks (2005) gave quite similar results for the DTM and AR1 schemes, both in the non-ensemble tests (PDFs, correlation functions, wave statistics) and the ensemble tests. The difference of our findings with those reported by Wilks (2005) can be due to differences in the ensemble test set-up (but note the consistency of our own ensemble and non-ensemble test results) or to small yet apparently important differences in the way the schemes are implemented.

c. The case $\Delta t > \delta t$

As already discussed in section 4b, practical circumstances may be such that the integration time step δt of the numerical model is different from the sampling time step Δt of the available data. Two ways of dealing with this problem were described in section 4b: (i) using a split integration scheme (update $\mathbf{X}$ every δt time units, but $\mathbf{B}$ only every Δt time units), and (ii) calculating transition probability matrices $\tilde{P}^{(ij)}$ with time step δt, starting from matrices $P^{(ij)}$ with time step Δt (the "generator method"). The latter possibility is more complicated to implement, and gives less accurate results (for reasons discussed in section 4b) than the split integration scheme in our testing set-up. Therefore, we only present results here using the split integration scheme. However, we note that the generator method can still be of practical use if model output is desired at a shorter sampling interval than the sampling interval of the available data. Moreover, we speculate that the performance using the generator method will improve for models that are more strongly mixing than the L96 model.

The results obtained by using the split integration scheme are shown in figures 13 – 16 (RMSE is omitted here, since it gives a similar picture as the anomaly correlation). For the construction of the stochastic matrices $P^{(ij)}$ we used data from the full L96 model with sampling interval $\Delta t = 0.1$. The integration time step of the reduced model with CMC parameterization is $\delta t = 0.01$. Thus, during the integration the B_k are updated once every 10 time steps. As can be seen, this version of the CMC scheme for the case $\Delta t \gg \delta t$ performs well, although it remains slightly less accurate than the CMC method using $\Delta t = \delta t$ presented in sections 5a and 5b (in the figures, those results are added for comparison). The fact that the B_k are updated only once every 10 time steps speeds up the computation significantly. This computational advantage can be such that using the split scheme may be attractive even if data with $\Delta t = \delta t$ is available. The increase in speed can outweigh the decrease in performance (which is rather small in our results) in some circumstances, in particular if B_k does not evolve fast, but on a similar timescale as X_k.

6. Discussion

The purpose of this study was to present a new approach to stochastic parameterization, and to test this approach by implementing it in the framework of the Lorenz 96 model. The new parameterization scheme we have proposed represents unresolved processes as stochastic (Markov) processes whose properties are conditional on the state of the resolved variables. In our numerical algorithm, these conditional Markov processes took the form of a collection of Markov chains, making practical implementation easy. The Markov chains are inferred from data in a very simple way, using only binning and counting.

We have compared the conditional Markov chain (CMC) scheme with two other parameterization schemes, a standard deterministic one (DTM) and the stochastic scheme proposed in Wilks (2005) (AR1). Several tests were carried out to assess the performance of the various

schemes, comparing probability distributions (PDFs), correlation functions (ACFs and CCFs), wave statistics and ensemble forecasts. The CMC scheme performed better than the DTM and AR1 schemes, even though the number of states of the Markov chains was small ($N_B = 4$). To test the robustness of these results, we made a brief exploration of other parameter settings for the L96 model. With $\varepsilon = 0.5$, the CMC scheme performed substantially better than the other two schemes. In the presence of clear time scale separation between the resolved and the unresolved variables ($\varepsilon = 0.1$), differences in performance between the schemes were rather minor (results not shown).

For a better understanding of the good performance of the CMC scheme it should be pointed out that with the parameter settings for L96 used in this study, the motion of phase points in the (X_k, B_k) plane tends to follow a roughly clockwise path. When X_k increases, B_k typically takes a (noisy) path through the upper part of the cloud of points shown in figure 1. During a decrease of X_k, B_k is more likely to be in the lower part of the cloud. The imprint of this "loop" (somewhat reminiscent of a hysteresis loop) in otherwise noisy behavior can be captured with the CMC scheme but not with the DTM or AR1 schemes. With the AR1 scheme driven by red noise, sustained trajectories through the lower or upper part of the data cloud are also possible; however, with the AR1 scheme those trajectories are equally likely to follow the loop as they are to go against it. The CMC scheme is better equipped, by design, to capture such structures.

Altogether, the results from the proposed CMC parameterization scheme are encouraging. Clearly, the L96 model is an idealized toy model, making it a suitable testing ground for a new parameterization approach. However, care should be taken when extrapolating results obtained within the L96 framework to other model situations. As a next step, the CMC scheme will have to be implemented and tested in a more realistic modeling environment.

In our testing, we already took into account two issues that can be expected to be of importance in more comprehensive model set-ups. First, we used parameter settings for the full L96 model that give little to no time scale separation between the resolved and unresolved variables. Previous studies of the L96 system were often carried out with parameters such that the unresolved variables (the $Y_{j,k}$) evolve on a faster time scale than the resolved variables. This time scale separation allows for the use of designated mathematical results and techniques (see e.g. Fatkullin and Vanden-Eijnden (2004); Vanden-Eijnden (2003)), but can be unrealistic.

A second issue is the sampling interval of the data that is available to base the parameterization on. The sampling interval can be different (longer) than the time step at which the numerical model with parameterization scheme is integrated. This issue may arise when dealing, for example, with data stemming from observations that do not have high temporal resolution. We discussed two potential solutions to this problem (see sections 4b and 5c). The split integration scheme solution gave the best results here, with only a minor decrease in performance (and significant reduction in computation time) compared to the case where both time steps are equal.

We conclude by discussing a number of further issues that will be of relevance for the application of the CMC scheme in more realistic models:

- The X_k variables of the L96 model are usually interpreted as lying on a circle of constant latitude, see section 2. Thus, the L96 model can be regarded, somewhat loosely, as a model with one spatial dimension (the x, or E-W direction), discretized on a number of gridpoints. The y and z directions (N-S and vertical) are absent in L96. In more realistic models, all three spatial directions are present; gridpoint discretization leads to a number of gridpoints that is orders of magnitude higher than in the L96 model. From

a computational point of view, this increase need not be problematic. The CMC scheme runs independently for each gridpoint and is therefore suitable for parallelization. The computational cost at each gridpoint is very small: only a small Markov chain must be evolved in time.

- The conditionality of the Markov chain at each gridpoint (the X_k dependence) was limited by using a "locality" assumption: the Markov chain at gridpoint k depends only on the resolved variable (X_k) at that same gridpoint, not on variables at other gridpoints ($X_{k'}$ with $k' \neq k$). If the number of gridpoints increases because one considers models with 2 or 3 spatial dimensions, this locality assumption will keep the Markov chain conditionality equivalently simple. Thus, the complexity of the CMC scheme at each gridpoint is the same for a model with $\mathcal{O}(10^1)$ gridpoints as it is for a model with $\mathcal{O}(10^6)$ gridpoints.

- The main challenge for applying the CMC scheme to realistic models concerns the number of different variables at each gridpoint. In the L96 model, there is one quantity to be parameterized (B_k) and one resolved variable (X_k) at each gridpoint k. However, the adiabatic core of an atmosphere model based on the primitive equations, for example, uses 5 resolved variables at each gridpoint. Making the Markov chain conditional on several resolved variables can lead to an intractable scheme if done naively. For example, dividing the range of each resolved variable in 16 intervals (similar to the 16 intervals $\mathcal{I}_i$ for X_k, see section 4) gives $16^5 \approx 10^6$ possible bins in which the vector of the 5 resolved variables at a single gridpoint can be at any moment. Keeping the number of these bins limited is necessary, both to limit computer memory demands of the CMC scheme and to keep estimation of the Markov chains from data feasible. Limiting the Markov chain conditionality to one or two judiciously chosen resolved variables, or making the chain conditional on a (linear) combination of variables, are possible solutions.

  The number of quantities for which a parameterization is needed at each gridpoint is usually also larger than one. This can be dealt with by using separate Markov chains for separate quantities. Correlations between these quantities are partly accounted for through the dependence on resolved variables. If strong correlations or physical requirements (e.g. conservation properties) do not permit the use of separate Markov chains at a single gridpoint, one collective Markov chain must be constructed in which each Markov chain state corresponds to a particular combination of values for the different parameterized quantities at the gridpoint.

- The properties of the conditional Markov chain were inferred from data. For the L96 model, such data was easily generated by integrating the full L96 model. For more realistic models, there is no "meta-model" that can be run easily and cheaply to produce the necessary data. Instead, one can use data from two sources. Data from observations (or reanalysis data) is the first option. The other option is data produced by high-resolution models of limited spatial extent (e.g. a single GCM gridbox) and sufficient physical sophistication.

  Using the same conditional Markov chain (i.e., the same set of transition matrices) at every gridpoint, as we did for the L96 model, will be too simple for some purposes (depending on the parameterized quantity). Geographical location of gridpoints can be expected to play a role in realistic models. It may be impractical to construct separate Markov

chains for each gridpoint; instead one can define a few groups of gridpoints (e.g., tropics or midlatitudes, boundary layer or above, gridpoints over land or sea) and construct Markov chains for each group separately.

1) *

Acknowledgments.

We thank Andrew Majda for interesting discussions. D.C. is financially supported by the NWO research cluster NDNS+. E.V.-E. is supported in part by NSF grants DMS02-09959 and DMS02-39625, and by ONR grant N00014-04-1-0565.

References

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

Bladt, M. and M. Sørensen, 2005: Statistical inference for discretely observed Markov jump processes. *J. R. Statist. Soc. B*, **67**, 395–410.

Boffetta, G., P. Giuliani, G. Paladin, and A. Vulpiani, 1998: An extension of the Lyapunov analysis for the predictability problem. *J. Atmos. Sci.*, **55**, 3409–3416.

Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainty in the ECMWF Ensemble Prediction System. *Q. J. R. Meteorol. Soc.*, **125**, 2887–2908.

Crommelin, D. T., 2004: Observed nondiffusive dynamics in large-scale atmospheric flow. *J. Atmos. Sci.*, **61**, 2384–2396.

Crommelin, D. T. and E. Vanden-Eijnden, 2006: Fitting timeseries by continuous-time Markov chains: A quadratic programming approach. *J. Comp. Phys.*, **217**, 782–805.

Egger, J., 2001: Master equations for climatic parameter sets. *Clim. Dyn.*, **18**, 169–177.

Fatkullin, I. and E. Vanden-Eijnden, 2004: A computational strategy for multiscale systems with applications to Lorenz 96 model. *J. Comp. Phys.*, **200**, 605–638.

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

Katsoulakis, M. A., A. J. Majda, and A. Sopasakis, 2005: Multiscale couplings in prototype hybrid deterministic/stochastic systems: Part ii, stochastic closures. *Comm. Math. Sci.*, **3**, 453–478.

———, 2006: Intermittency, metastability and coarse graining for coupled deterministic-stochastic lattice systems. *Nonlinearity*, **19**, 1021–1047.

Katsoulakis, M. A., A. J. Majda, and D. G. Vlachos, 2003: Coarse-grained stochastic processes for microscopic lattice systems. *Proc. Natl. Acad. Sci.*, **100**, 782–787.

Khouider, B., A. J. Majda, and M. Katsoulakis, 2003: Coarse grained stochastic models for tropical convection and climate. *Proc. Natl. Acad. Sci.*, **100**, 11 941–11 946.

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

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

Lorenz, E. N., 1995: Predictability - a problem partly solved. *Proceedings of the 1995 ECMWF seminar on Predictability*, ECMWF, Reading, UK, 1–18.

Majda, A. J. and B. Khouider, 2002: Stochastic and mesoscopic models for tropical convection. *Proc. Natl. Acad. Sci.*, **99**, 1123–1128.

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

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

Orrell, D., 2003: Model error and predictability over different timescales in the Lorenz '96 systems. *J. Atmos. Sci.*, **60**, 2219–2228.

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

Penland, C. and L. Matrosova, 1994: A balance condition for stochastic numerical models with application to the El Niño-Southern Oscillation. *J. Climate*, **7**, 1352–1372.

Plant, R. S. and G. C. Craig, 2007: A stochastic parameterization for deep convection based on equilibrium statistics. *J. Atmos. Sci.*, to appear.

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

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

Vanden-Eijnden, E., 2003: Numerical techniques for multi-scale dynamical systems with stochastic effects. *Comm. Math. Sci.*, **1**, 385–391.

Wilks, D. S., 2005: Effects of stochastic parameterizations in the Lorenz '96 system. *Q. J. R. Meteorol. Soc.*, **131**, 389–407.

List of Figures

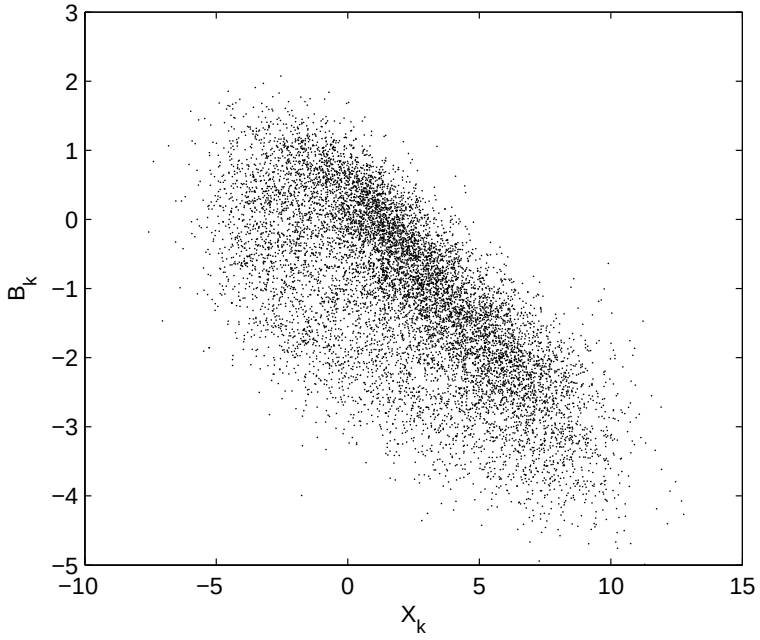

FIG. 1. Scatter plot of $B_k(t)$ versus $X_k(t)$ for the full L96 model (1). Parameter settings are $(\varepsilon, K, J, F, h_x, h_y) = (0.5, 18, 20, 10, -1, 1)$.

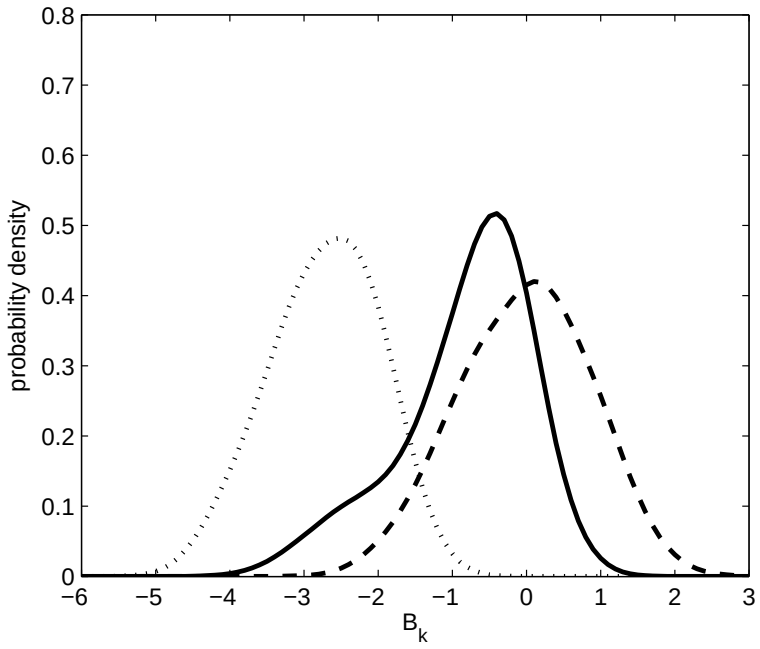

FIG. 2. Probability density functions $\rho(B_k)$ for B_k in various intervals of X_k. Dashed line: $\rho(B_k(t) \,|\, -4.5 < X_k(t) < -3.5)$. Solid line: $\rho(B_k(t) \,|\, 1.5 < X_k(t) < 2.5)$. Dotted line: $\rho(B_k(t) \,|\, 7.5 < X_k(t) < 8.5)$.

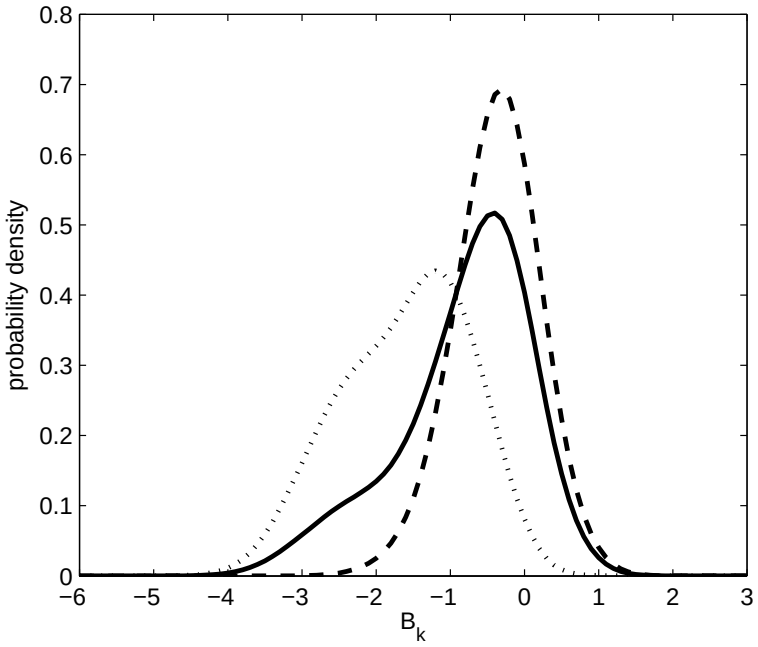

FIG. 3. Probability density functions $\rho(B_k)$ for B_k in the interval $1.5 < X_k < 2.5$, conditional on the sign of $X_k(t) - X_k(t - \Delta t)$. Dashed line: $\rho(B_k(t) \,|\, 1.5 < X_k(t) < 2.5, X_k(t) > X_k(t - \Delta t))$. Dotted line: $\rho(B_k(t) \,|\, 1.5 < X_k(t) < 2.5, X_k(t) < X_k(t - \Delta t))$. Solid line: total PDF $\rho(B_k(t) \,|\, 1.5 < X_k(t) < 2.5)$.

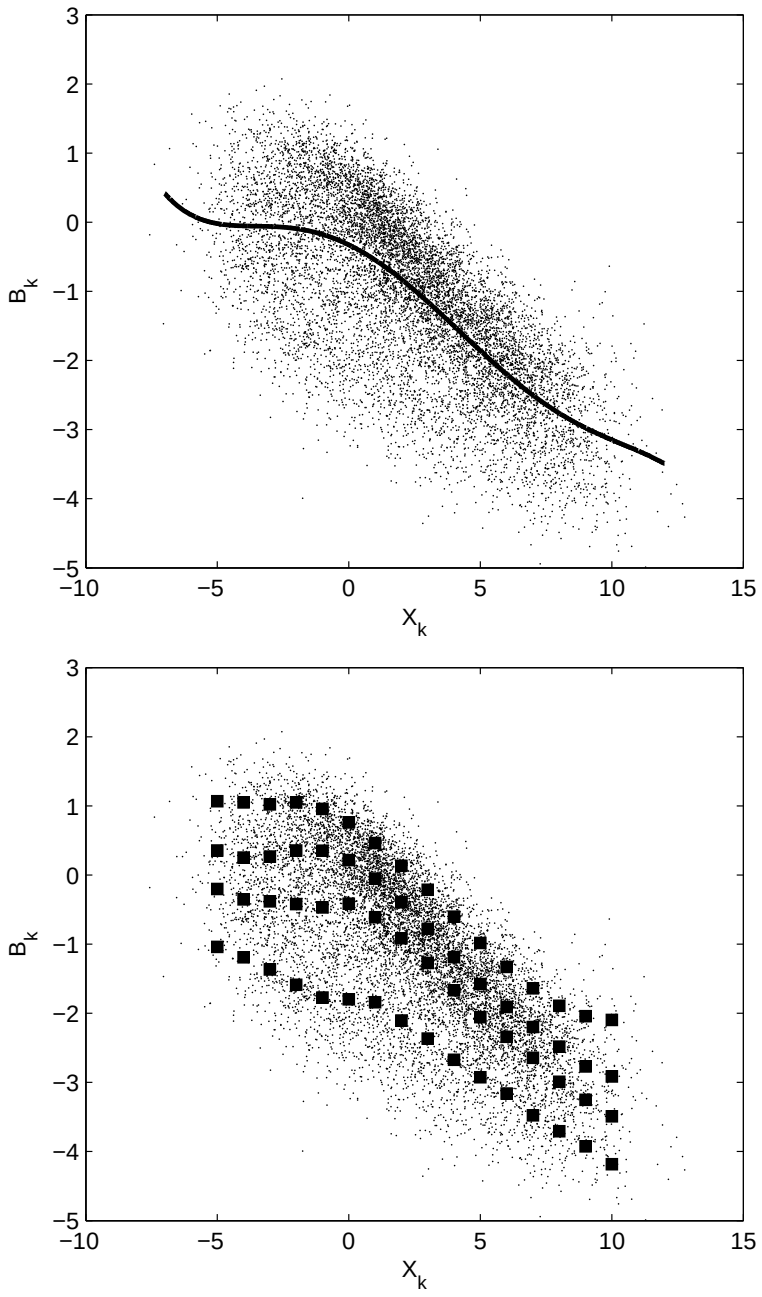

FIG. 4. Top: The solid curve is the 5th order polynomial fit used for the DTM parameterization of B_k. Bottom: The black squares denote the values of the $\mathcal{B}_n^i$ used for the CMC parameterization with $N_B = 4$ (see text). In both panels, the dots are the scatter plot of B_k versus X_k for the full L96 model (1).

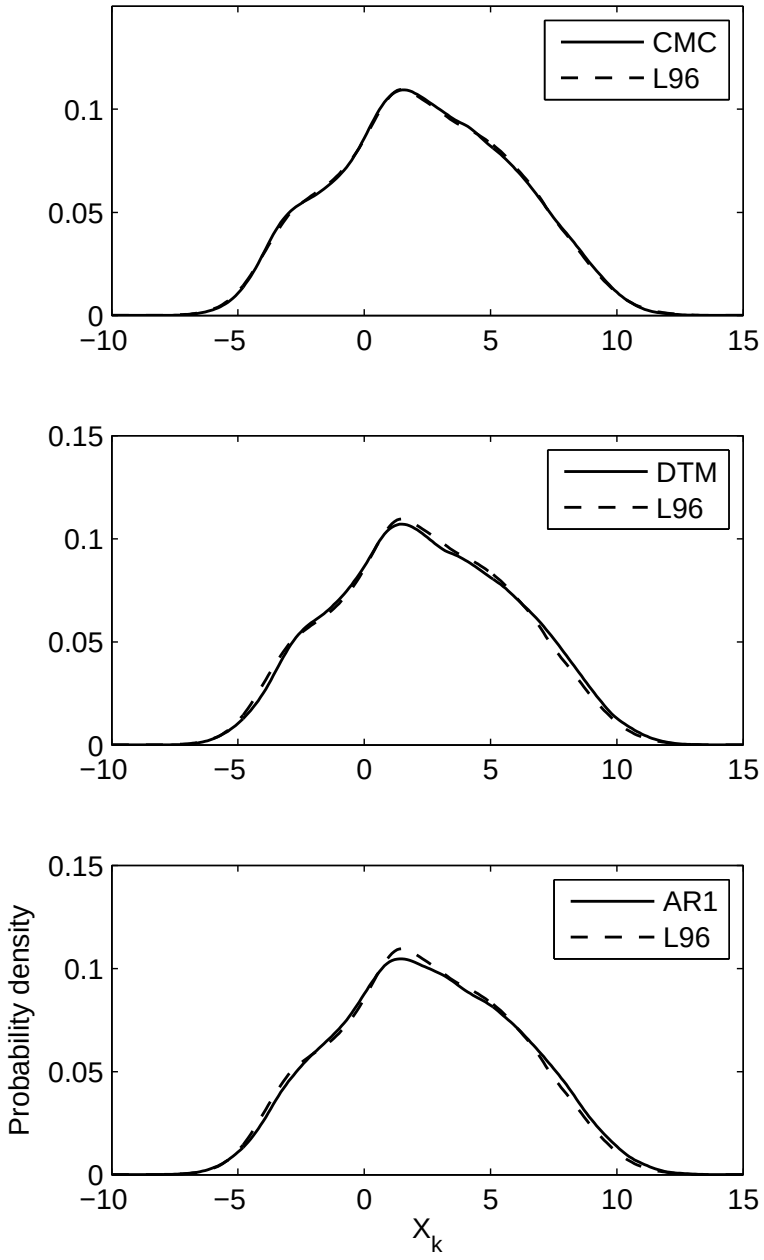

FIG. 5. PDFs of X_k, produced by the reduced models with various parameterization schemes. Top: conditional Markov chain (CMC) scheme with $N_B = 4$. Middle: deterministic (DTM) scheme. Bottom: AR1 scheme from Wilks (2005). The PDF produced by the full L96 model (1) is added in each panel for comparison.

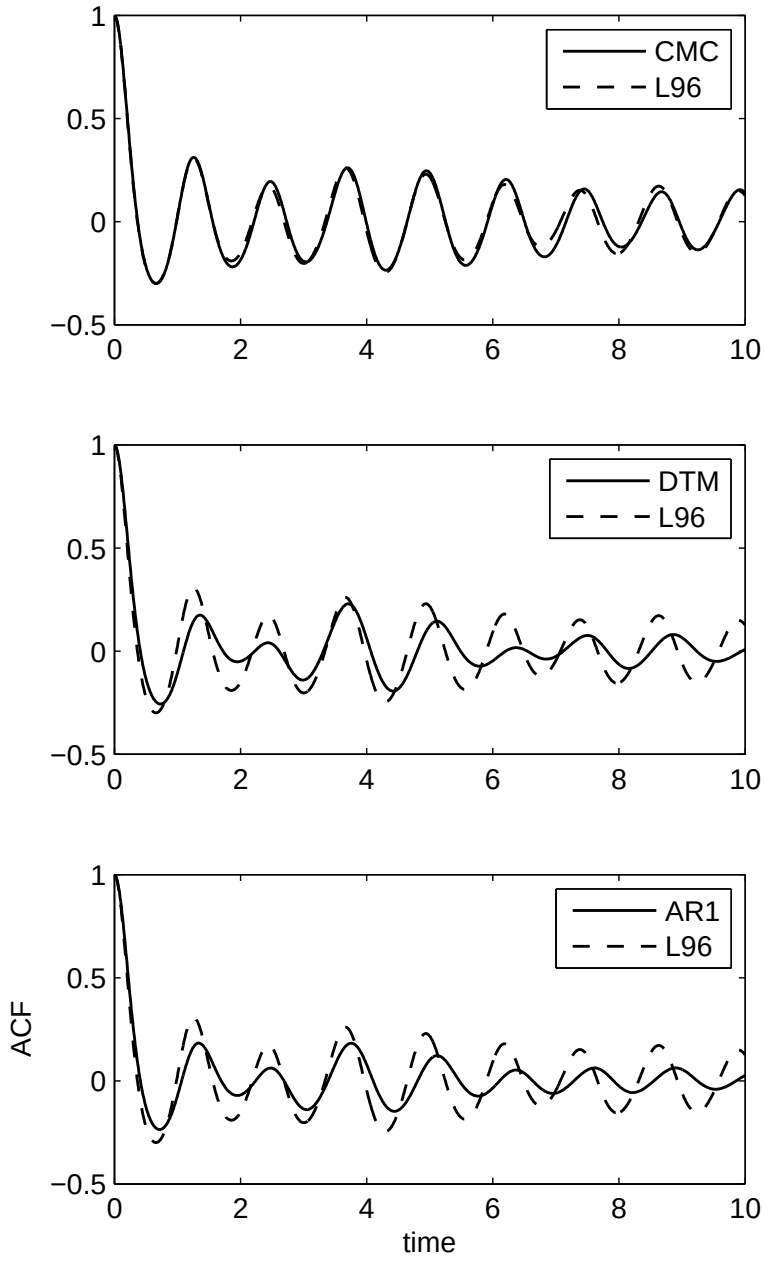

FIG. 6. Autocorrelation functions (ACF) of X_k. Results from reduced models with different parameterization schemes (CMC, DTM, AR1) are shown, together with the ACF from the full L96 model (L96).

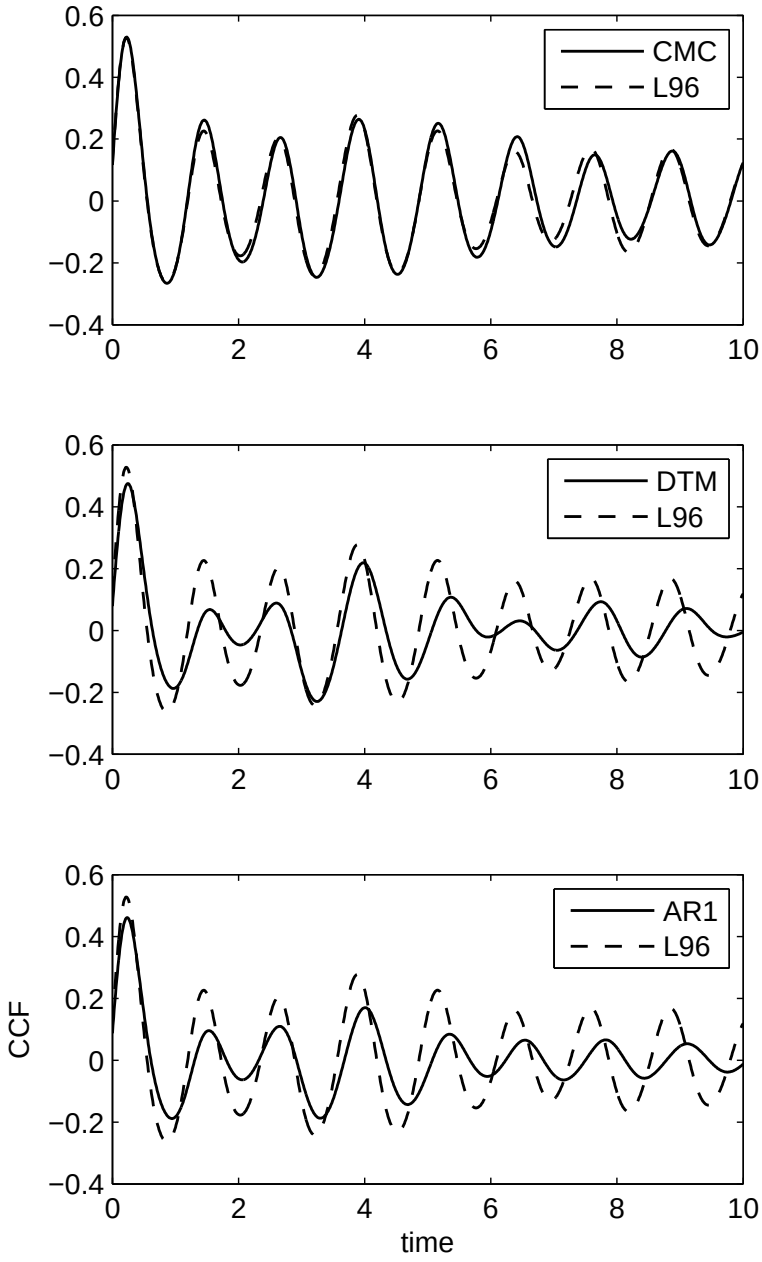

FIG. 7. Cross correlation functions (CCF) of X_k and X_{k+1}. Results from reduced models with different parameterization schemes (CMC, DTM, AR1) are shown, together with the CCF from the full L96 model (L96).

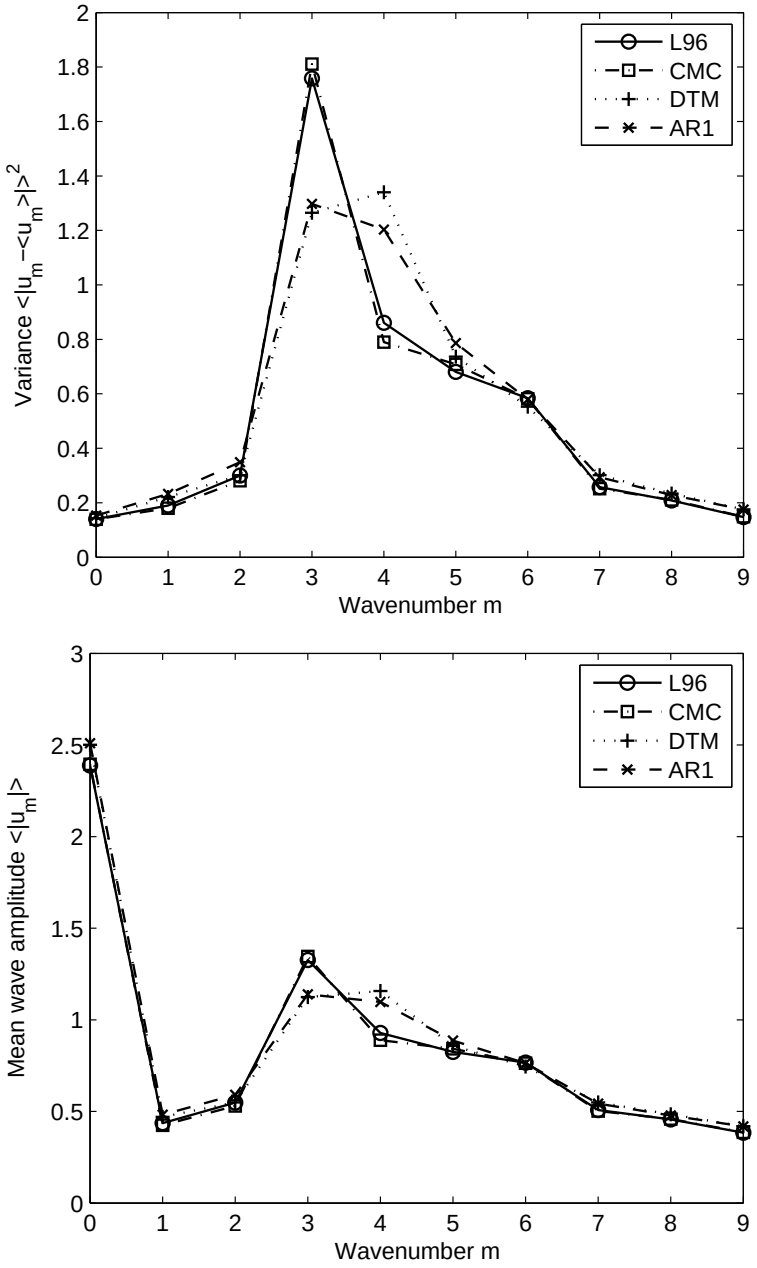

FIG. 8. Top: Wave variances $\langle |u_m - \langle u_m \rangle|^2 \rangle$. Bottom: Mean wave amplitudes $\langle |u_m| \rangle$. Time-series for the wave number vector $\mathbf{u}$ are obtained by Fourier transforming the state vector $\mathbf{X}$ at every datapoint. The Fourier mode with wavenumber m has variance $\langle |u_m - \langle u_m \rangle|^2 \rangle$ and mean amplitude $\langle |u_m| \rangle$ (where $\langle . \rangle$ denotes time average). Results from the full L96 model (L96) and from reduced models with different parameterization schemes (CMC, DTM, AR1) are shown.

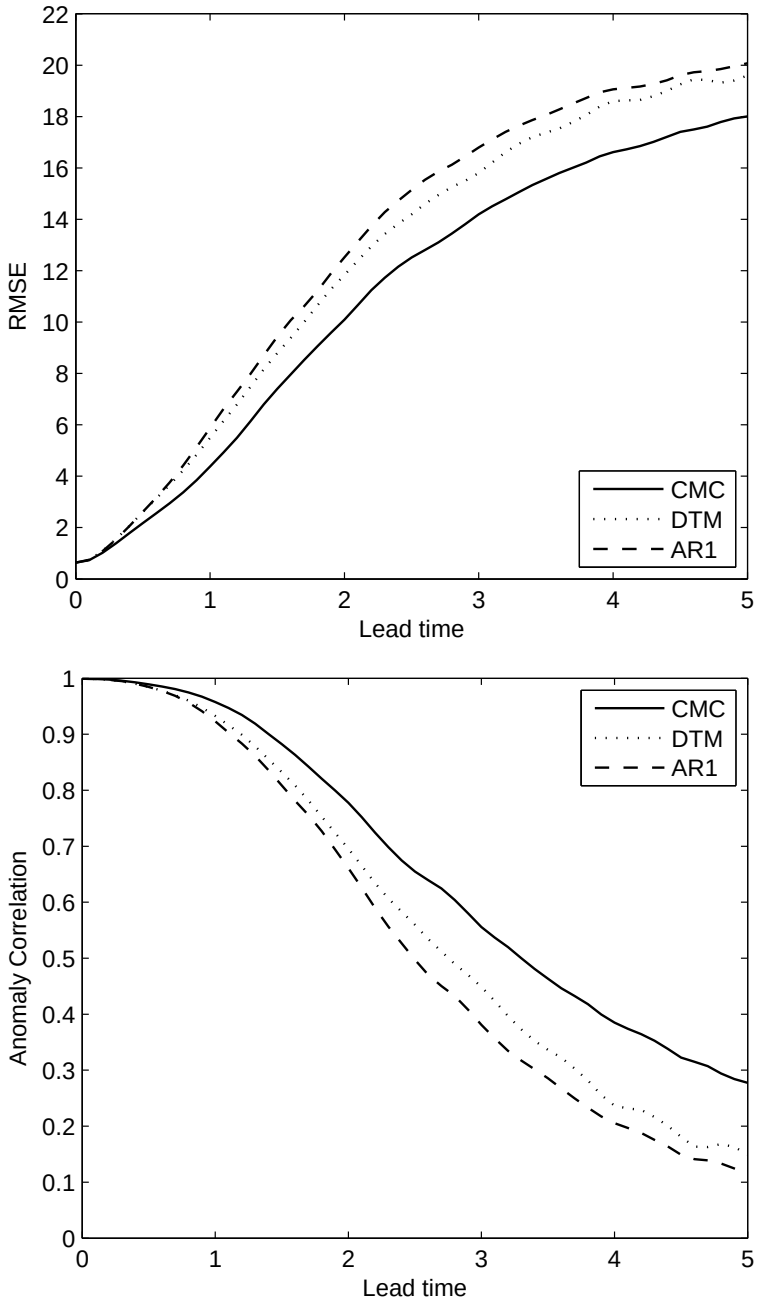

FIG. 9. Results from ensemble integrations with reduced models using the CMC parameterization scheme ($N_B = 4$), the DTM scheme and the AR1 scheme. Top: Root Mean Square Error (RMSE) versus lead time (τ); bottom: Anomaly Correlation versus lead time. The number of ensemble members $N_{\rm ens} = 1$; the number of initial states $N_{\rm init} = 1000$.

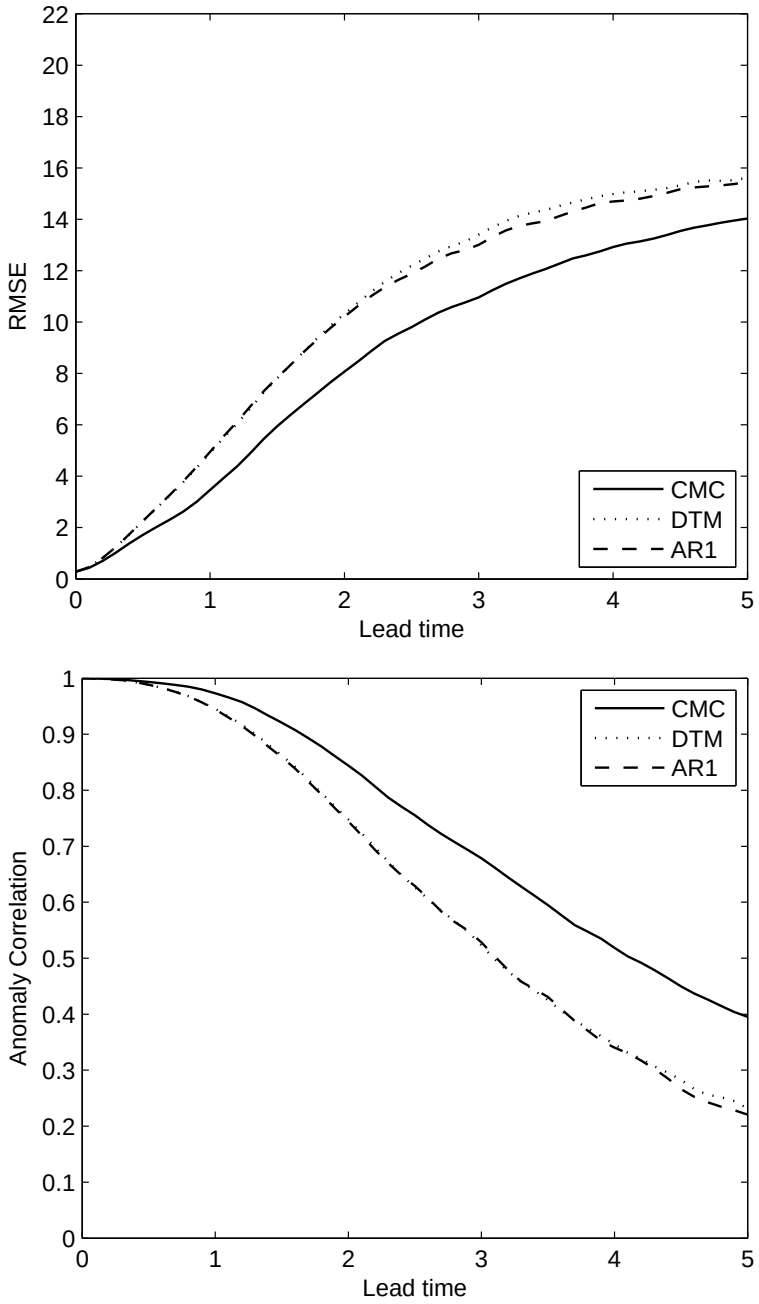

FIG. 10. Results from ensemble integrations with reduced models using the CMC parameterization scheme ($N_B = 4$), the DTM scheme and the AR1 scheme. Top: Root Mean Square Error (RMSE) versus lead time (τ); bottom: Anomaly Correlation versus lead time. The number of ensemble members $N_{\rm ens} = 5$; the number of initial states $N_{\rm init} = 1000$.

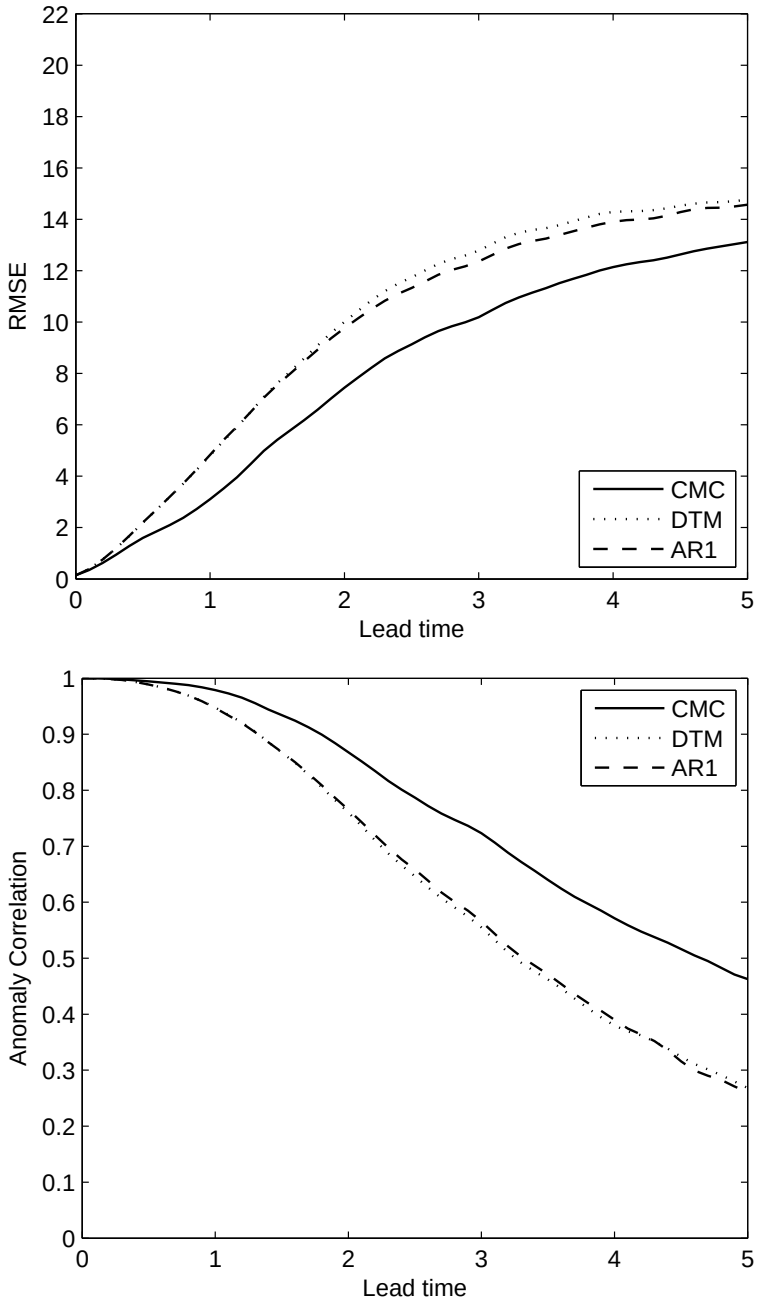

FIG. 11. Results from ensemble integrations with reduced models using the CMC parameterization scheme ($N_B = 4$), the DTM scheme and the AR1 scheme. Top: Root Mean Square Error (RMSE) versus lead time (τ); bottom: Anomaly Correlation versus lead time. The number of ensemble members $N_{\rm ens} = 20$; the number of initial states $N_{\rm init} = 1000$.

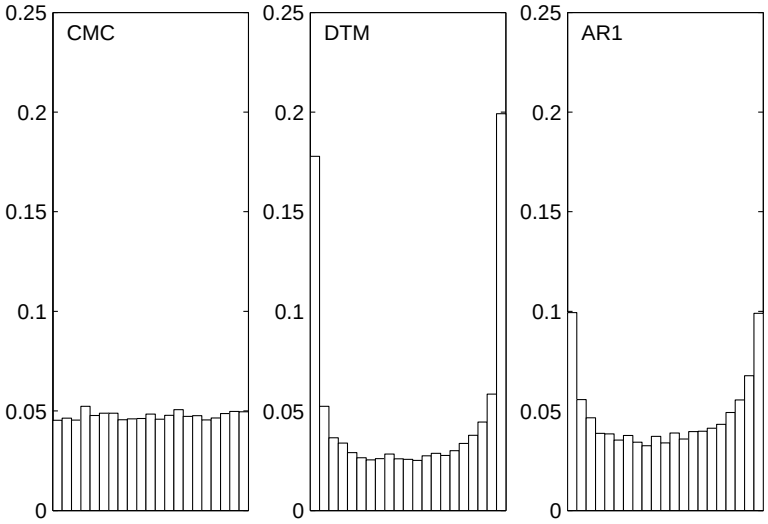

FIG. 12. Rank histograms resulting from ensemble integrations with reduced models using the CMC parameterization scheme ($N_B = 4$), the DTM scheme and the AR1 scheme. Lead time is $\tau = 2$; ensemble size $N_{\rm ens} = 20$. The CMC scheme gives a near uniform distribution; the DTM and AR1 schemes lead to under-dispersed ensembles, visible as U-shaped histograms.

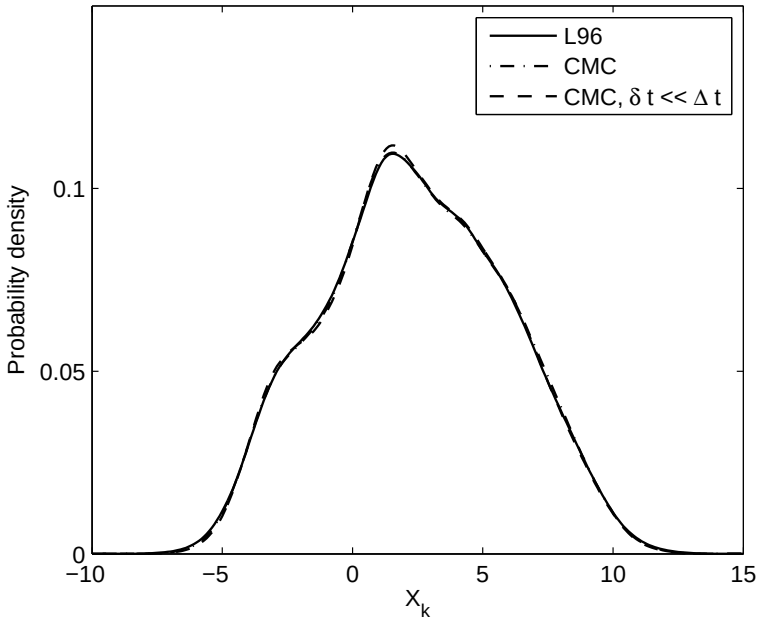

FIG. 13. PDFs of X_k, as produced by the full L96 model, the reduced model with CMC scheme based on data with $\Delta t = \delta t$ and the reduced model with CMC scheme based on data with $\Delta t \gg \delta t$ using the split integration scheme.

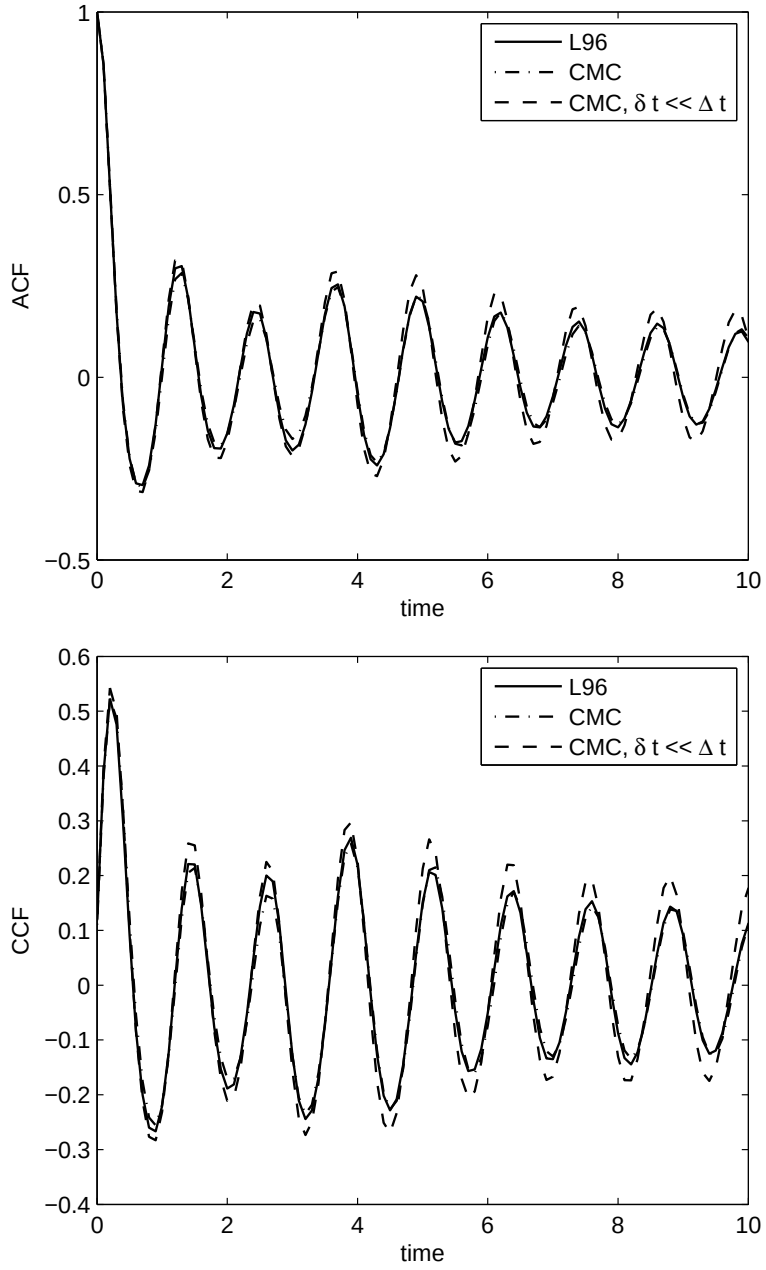

FIG. 14. Top: Autocorrelation functions (ACF) of X_k. Bottom: Cross correlation functions (CCF) of X_k and X_{k+1}. Results are from the full L96 model, the reduced model with CMC scheme based on data with $\Delta t = \delta t$ and the reduced model with CMC scheme based on data with $\Delta t \gg \delta t$ using the split integration scheme.

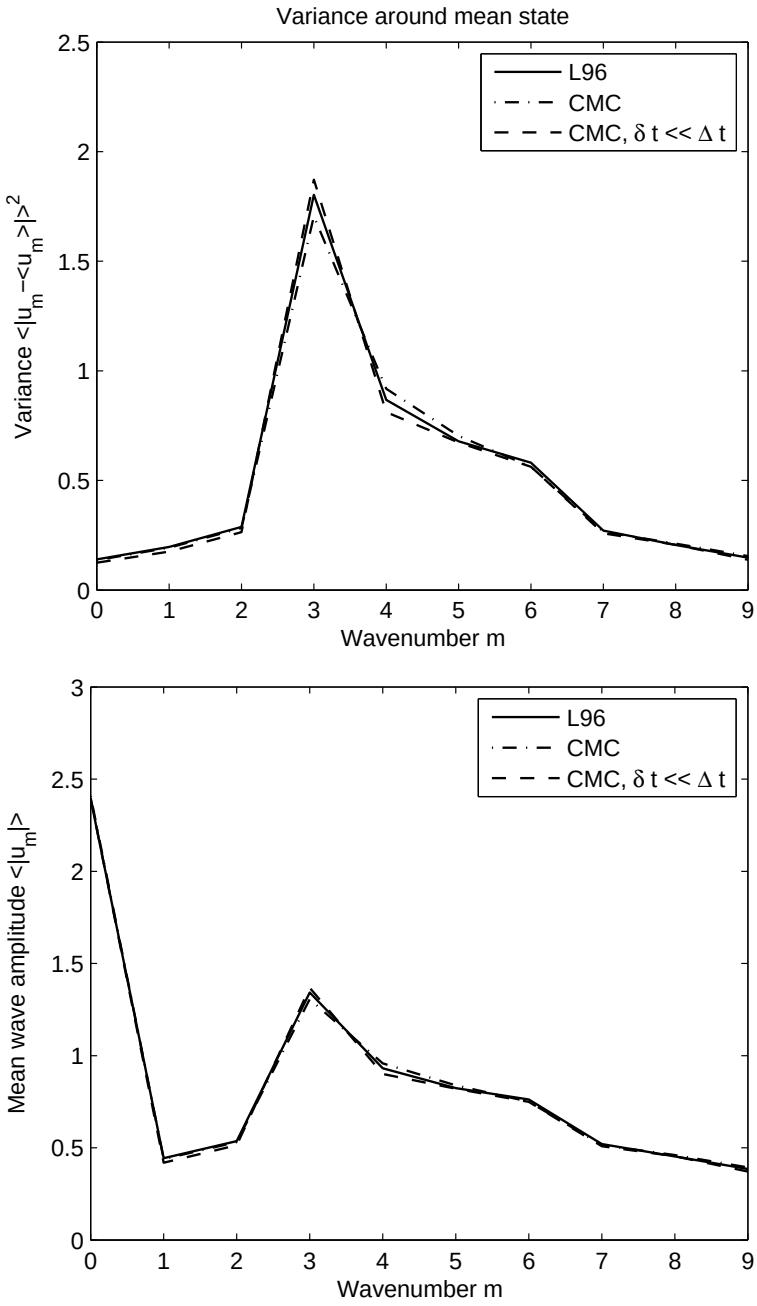

FIG. 15. Top: Wave variances $\langle |u_m - \langle u_m \rangle|^2 \rangle$. Bottom: Mean wave amplitudes $\langle |u_m| \rangle$. Results are from the full L96 model, the reduced model with CMC scheme based on data with $\Delta t = \delta t$ and the reduced model with CMC scheme based on data with $\Delta t \gg \delta t$ using the split integration scheme.

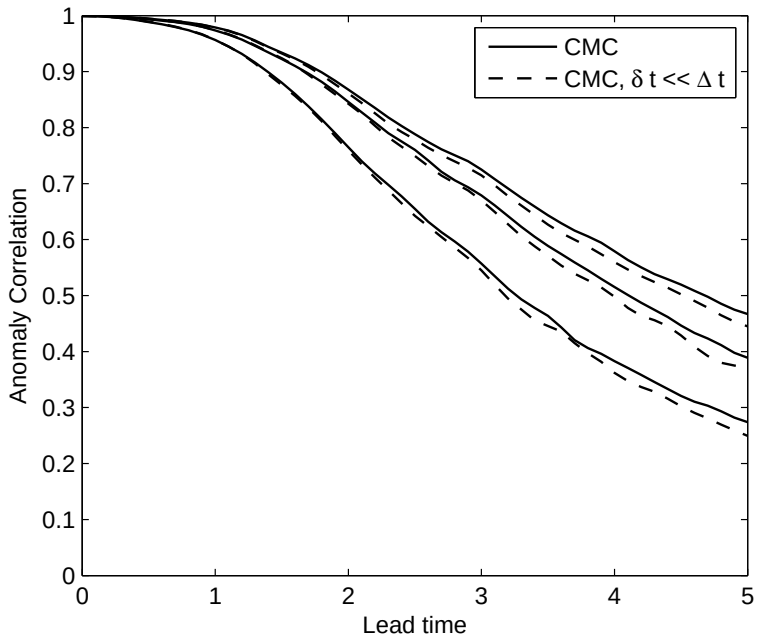

FIG. 16. Anomaly Correlation versus lead time. Results are from the reduced model with CMC scheme based on data with $\Delta t = \delta t$ (solid lines) and the reduced model with CMC scheme based on data with $\Delta t \gg \delta t$ using the split integration scheme. The various curves show results using $N_{\rm ens} = 1$, $N_{\rm ens} = 5$ and $N_{\rm ens} = 20$. At any value of the lead time, the larger $N_{\rm ens}$, the higher the correlation.