

MARKOV CHAIN STOCHASTIC PARAMETRIZATIONS OF ESSENTIAL VARIABLES*

K. NIMSAILA[†] AND I. TIMOFEYEV[†]

Abstract. We analyze the performance of the novel Markov chain stochastic modeling technique for derivation of effective equations for a set of essential variables. This technique is an empirical approach where the right-hand side of the essential variables is modeled by a Markov chain. We demonstrate that the Markov chain modeling approach performs well in a prototype model without scale separation between the essential and the nonessential variables. Moreover, we utilize the truncated Burgers–Hopf model to show that the Markov chain should be properly conditioned on the essential variables to reproduce the structure of two-point statistical quantities. On the other hand, the conditioning can be rather straightforward and unsophisticated.

Key words. Markov chain, effective model, stochastic parametrization

AMS subject classifications. 60J60, 62G99, 62M99

DOI. 10.1137/090770394

1. Introduction. Derivation of efficient parametrizations reproducing statistical features of essential variables has been an active area of research. Most of the research activities concentrated on nonlinear chaotic and turbulent systems, including geophysical models of various complexity. Several semianalytical approaches have been developed for the derivation of effective systems for selected essential degrees of freedom. Examples of the most recently developed semianalytical approaches include techniques based on the Zwanzig–Mori formalism [16], including the optimal prediction framework [6, 4, 5], the stochastic mode-reduction technique for systems with scale separation [24, 25, 26, 27], and course-graining of spin-flip stochastic models [17, 18, 19]. Numerically oriented techniques include methods for the estimation of stochastic differential equations from numerical and/or observational data [9, 8, 29, 2, 3], heterogeneous multiscale methods [11, 14, 13, 12], and the application of hidden Markov models in geophysical fluid dynamics [15].

In this paper we analyze the performance of the Markov chain modeling approach introduced in [10]. This is an empirical approach in which the right-hand side for the essential variables is modeled by a conditional Markov chain. The transition probability matrix is estimated numerically from a single realization of the full model. The main goal of this work is to investigate the performance of this technique in the absence of scale separation between the essential and the nonessential variables. In addition, we also analyze how the conditioning of the Markov chain affects the statistical behavior of the effective model.

The rest of the paper is organized as follows. In section 2 we give a general overview of the Markov chain modeling approach. In section 3 we consider a prototype model without scale separation between the essential and the nonessential variables with nonequilibrium tests discussed in section 3.1. In section 4 we apply the Markov chain approach to the truncated Burgers–Hopf model.

*Received by the editors September 7, 2009; accepted for publication (in revised form) October 21, 2010; published electronically December 21, 2010.

<http://www.siam.org/journals/mms/8-5/77039.html>

[†]Department of Mathematics, University of Houston, Houston, TX 77204 (kawinsk@math.uh.edu, ilya@math.uh.edu). The research of I. Timofeyev is partially supported by the NSF grant DMS-0405944.

2. Discrete-time Markov chain modeling approach. The Markov chain modeling approach is an empirical stochastic parametrization technique where part of the right-hand side in the equation for the essential variables is replaced by a Markov chain. In this paper we consider discrete-time Markov chain modeling which is analogous to the discrete-time Markov chain modeling considered in [10]. Let us consider a dynamical system where dependent variables are decomposed into two sets—essential variables, x , and nonessential variables, y . Therefore, the equations for dependent variables can be schematically written as follows:

$$\begin{aligned}\dot{x} &= f(x) + g(x, y), \\ \dot{y} &= h(x, y).\end{aligned}$$

The main general goal of stochastic modeling is to reduce the dimensionality of the equations by replacing the term $g(x, y)$ by an appropriate stochastic parametrization. In the Markov chain stochastic parametrization approach $g(x, y)$ is replaced by a Markov chain. Since in general $g(x, y)$ depends on the current value of x , the Markov chain has to be conditioned on the essential variables and the effective stochastic model takes the following form:

$$(2.1) \quad \dot{X} = f(X) + m(Z; X),$$

where $m(Z; X)$ is a conditional Markov chain. In this paper we consider discrete Markov chains, which is particularly appropriate for numerical simulations of effective systems. Thus, we consider discrete analogs of (2.1). The simplest discretization is the Euler discretization which can be represented as follows:

$$(2.2) \quad X_{n+1} = X_n + f(X_n)\Delta t + Z_n\Delta t,$$

where Z_n is determined from a conditional Markov chain

$$(2.3) \quad Pr\{Z_n \in I_k | Z_{n-1} \in I_m, X_n \in J_q\} = p_{qmk}.$$

Equation (2.2) is given just as an example. The deterministic part of the effective equation, $\dot{X} = f(X)$, can be discretized with a higher-order scheme and the Euler scheme can be utilized only to discretize stochastic terms. In order to properly condition the transition probability matrix in (2.3) we define a partition $\mathbb{J} = \{J_l, l = 1 \dots M\}$ for the effective variable X . In addition, state space $\mathbb{I} = \{I_k, k = 1 \dots N\}$ for Z_n is also defined in advance. We will comment on particular choice of \mathbb{J} and \mathbb{I} in examples presented below.

Overall, we need to estimate MN^2 coefficients of the transition probability matrix. This is done by a straightforward bin-counting procedure from the discretized time-series of $g(x, y)$. Since the term $g(x, y)$ has in general complex correlation structure, it is best if the time-step for discretizing $g(x, y)$ and the time-step for simulating the effective system in (2.2) are identical. We determined empirically that optimal performance can be achieved if the time-step of simulations, Δt , is approximately $0.1 \dots 0.5$ of the correlation time of $g(x, y)$. We would like to emphasize that in multiscale systems the time-step for simulating the full system in (2.1) is limited by the nonessential variables, y . Therefore, if y variables are faster than the x variables, the effective system can be simulated with a bigger time-step and the stochastic parametrization approach can potentially lead to significant computational savings. Moreover, eliminating the nonessential variables also leads to computational savings due to dimensionality reduction of the problem.

State space of the stochastic process Z_n can be either discrete or continuous. For the discrete state space the value of Z_n in state l can be chosen as an average of the distribution of values of $g(x, y)$ in the interval I_l . Alternatively, for continuous statespace formulation, Z_n can be sampled from a uniform distribution over I_l . Z_n can be sampled from an exponential distribution if I_l is an end-interval involving $\pm\infty$. In this paper we consider discrete state space for Z_n . Results for the continuous statespace formulation are similar and not reported here for the brevity of presentation. Detailed comparison of the discrete state space and the continuous state-space Markov chain modeling can be found in [28].

The Markov chain stochastic parametrization discussed in this paper is conceptually identical to the approach proposed in [10]. The approach to the estimation of the transition probability matrix is slightly different; in [10] the transition probability matrix is estimated as an exponential of the generator of the embedded continuous-time Markov chain. In the present paper we utilize discrete Markov chains and use the time-step for estimating the transition probability matrix which is identical to the time-step of integration of the effective model. Nevertheless, differences between the two approaches are minor and should not affect the performance of the method. Conditioning of the Markov chain stochastic parametrization represents the most crucial empirical issue. Conditioning considered in this paper is less elaborate than in [10]. In particular, in [10] the Markov chain is conditioned on the present value of the essential variable, X_n , as well as the future tendency (value of the essential variable with the stochastic parametrization fixed at the previous value). In this paper we show that such elaborate conditioning may not be always necessary; simpler more practical conditioning schemes perform extremely well in several nonlinear chaotic models.

3. Prototype model: Triad model coupled with two nonessential variables. In this section we consider a prototype low-dimensional model to illustrate the applicability and robustness of the Markov chain stochastic modeling approach. In particular, this model is constructed as an extension of a simple triad model mimicking wave-wave interactions in fluid dynamics. In this model the parametrized term does not depend on the essential variables and we utilize this model to investigate the performance of the Markov chain modeling technique in the absence of strong scale separation.

To illustrate the Markov chain stochastic parametrization approach we consider a triad model coupled to two nonessential variables

$$\begin{aligned}
 dx_1 &= B_1 x_2 x_3 dt + y_1 y_2 dt - \beta x_1^3 dt, \\
 dx_2 &= B_2 x_1 x_3 dt - a_2 x_2 dt + b_2 dW_1, \\
 dx_3 &= B_3 x_1 x_2 dt - a_3 x_3 dt + b_3 dW_2, \\
 dy_1 &= -\kappa \gamma y_1 dt + \sqrt{\kappa} \sigma dW_3, \\
 dy_2 &= -\kappa \gamma y_2 dt + \sqrt{\kappa} \sigma dW_4,
 \end{aligned}
 \tag{3.1}$$

where $W_k, k = 1 \dots 4$ are independent Brownian motions and κ controls the separation of time-scales between the essential variables, x_1, x_2, x_3 , and the nonessential variables y_1, y_2 . Parameters in the model (3.1) are

$$\begin{aligned}
 B_1 &= -1.0, & B_2 &= 0.4, & B_3 &= 0.6, \\
 \beta &= 1, & \gamma &= 1, & \sigma &= 2, \\
 a_2 &= 0.4, & a_3 &= 0.5, & b_2 &= 0.6, & b_3 &= 0.7.
 \end{aligned}$$

We consider four values of κ ,

$$\kappa = 10, 2, 1.5, 1,
 \tag{3.2}$$

to study how scale separation between the essential and the nonessential variables affects the Markov chain parametrization. We utilize the Euler scheme for the deterministic and stochastic terms and the Markov chain effective model written in discrete form becomes

$$\begin{aligned}
 (3.3) \quad & X_1(t + \Delta t) = X_1(t) + B_1 X_2(t) X_3(t) \Delta t + \alpha Z(t) \Delta t - \beta X_1^3(t) \Delta t, \\
 & X_2(t + \Delta t) = X_2(t) + B_2 X_1(t) X_3(t) \Delta t - a_2 X_2(t) \Delta t + b_2 \Delta W_1, \\
 & X_3(t + \Delta t) = X_3(t) + B_3 X_1(t) X_2(t) \Delta t - a_3 X_3(t) \Delta t + b_3 \Delta W_2,
 \end{aligned}$$

where $\Delta W_k, k = 1, 2$ are independent Gaussian random variables with mean zero and variance Δt and $Z(t)$ is defined as a discrete Markov chain

$$(3.4) \quad Pr\{Z(t) \in I_k | Z(t - \Delta t) \in I_m\} = p_{mk}.$$

Since in (3.1) the nonessential variables, y_1, y_2 , do not depend on x_1 , the Markov chain does not have to be conditioned on X_1 . Transition probabilities p_{mk} are estimated from a single simulation of the full model in (3.1) with the sampling time-step $\Delta t = 0.02$.

We consider eleven states ($N = 11$) in the Markov chain with the partition

$$\begin{aligned}
 (3.5) \quad \mathbb{I} = \{ & -\infty < \dots < \bar{Y} - 3.0 * sd(Y) < \dots < \bar{Y} - 2.0 * sd(Y) < \dots \\
 & < \bar{Y} - 1.2 * sd(Y) < \dots < \bar{Y} - 0.6 * sd(Y) < \dots < \bar{Y} - 0.1 * sd(Y) < \dots \\
 & < \bar{Y} + 0.1 * sd(Y) < \dots < \bar{Y} + 0.6 * sd(Y) < \dots < \bar{Y} + 1.2 * sd(Y) < \dots \\
 & & < \bar{Y} + 2.0 * sd(Y) < \dots < \bar{Y} + 3.0 * sd(Y) < \dots < \infty \},
 \end{aligned}$$

where $Y = y_1 y_2$ in the equation in (3.1); \bar{Y} denotes the mean of $y_1 y_2$ and $sd(Y)$ denotes standard deviation of $y_1 y_2$. Notation “ \dots ” denotes intervals for each state in the Markov chain. Since y_1 and y_2 are independent, these parameters can be computed explicitly in this model; $\bar{Y} = 0, sd(Y) = 2$. Although partitioning in (3.5) is empirical, we found that, typically, $N = O(10)$ states are required to reproduce statistical features of the essential variables sufficiently accurately and, moreover, partitioning based on the mean and standard deviation of $g(x, y)$ works best in most situations. Studies of how the size of the partition and the number of intervals for conditioning affect the performance of the Markov chain parametrization for prototype triad models are reported in sections 2.3.1 and 2.3.2 of [28]. Moreover, effective Markov chain models with $O(10)$ number of states perform well for the truncated Burgers–Hopf model as well. The intuitive explanation is that the partition should provide a sufficiently accurate approximation of the probability density function of the parametrized term $g(x, y)$ (for the triad model $g(x, y) = y_1 y_2$). Partitions defined using the mean and standard deviation (as in (3.5)) with $O(10)$ number of states seem to achieve this goal.

Correlation times of x_1 and y_1 for four values of κ in (3.1) are presented in Table 3.1. Table 3.1 demonstrates that while weak scale separation is present for $\kappa = 2$, regime with $\kappa = 1$ does not exhibit any scale separation between x_1 and y_1, y_2 variables. On the other hand, $\kappa = 10$ corresponds to strong scale separation between the slow variables, x_1, x_2, x_3 and fast variables y_1, y_2 .

Table 3.2 and Figure 3.1 show statistical agreement between the full model in (3.1) and the effective model in (3.3). Simulations of the bare truncation with $y_1 = y_2 = 0$ are also included for comparison. Since x_1 is the only variable coupled to the nonessential degrees of freedom, agreement between the statistical properties of x_1 in the full and the effective models is crucial in this case. Markov chain stochastic

TABLE 3.1

Correlation times of x_1 and y_1 in the simulations of the full model in (3.1) with different values of κ .

	Corr. Time of x_1	Corr. Time of y_1
$\kappa = 10$	1.35	0.2352
$\kappa = 2$	0.9182	0.4745
$\kappa = 1.5$	0.9406	0.6328
$\kappa = 1$	1.0063	0.9497

TABLE 3.2

Variance and kurtosis of x_k , $k = 1, 2, 3$ in the simulations of the full and effective models with different values of κ . Kurtosis is defined as $K\{x_k\} = E[(x_k - E[x_k])^4]/var[x_k]^2$. Kurtosis is the leading indicator of the departure from the Gaussianity, since $K = 3$ for Gaussian random variables. Statistical behavior of the bare truncation ($y_1 = y_2 = 0$) is also shown for comparison.

	$Var\{x_1\}$	$Var\{x_2\}$	$Var\{x_3\}$	$K\{x_1\}$	$K\{x_2\}$	$K\{x_3\}$
Full $\kappa = 10$	0.2973	0.4346	0.4755	2.3172	3.0655	3.0794
Eff. $\kappa = 10$	0.3098	0.4179	0.4609	2.2215	3.0032	2.9770
Full $\kappa = 2$	0.5139	0.48	0.5179	2.4575	3.0286	3.0181
Eff. $\kappa = 2$	0.5043	0.4672	0.5245	2.4537	2.9431	2.9838
Full $\kappa = 1.5$	0.5551	0.4934	0.5347	2.4426	3.0296	3.0351
Eff. $\kappa = 1.5$	0.5337	0.4744	0.5331	2.4255	2.9478	2.9881
Full $\kappa = 1$	0.6036	0.5172	0.5638	2.3999	3.0508	3.0885
Eff. $\kappa = 1$	0.5702	0.4881	0.5521	2.3777	2.9666	3.0102
Truncation	0.1446	0.4011	0.4332	2.3244	3.0884	3.0708

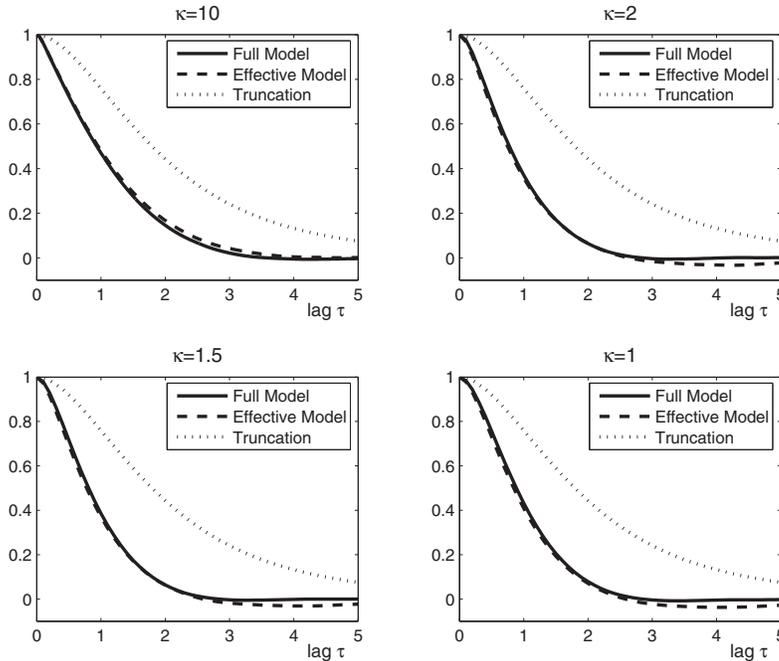


FIG. 3.1. Comparison of the correlation function $E[x_1(t)x_1(t + \tau)]$ in the simulations of the full model and corresponding effective model in the regime with $\kappa = 10, 2, 1.5, 1$; correlation function from the bare truncation ($y_1 = y_2 = 0$) is also plotted for comparison.

parametrization performs extremely well with $(\kappa = 10, 2)$ and without $(\kappa = 1.5, 1)$ scale separation between the essential and the nonessential variables. In particular, the Markov chain model reproduces the non-Gaussian features of x_1 very well. Values of the kurtosis and its overall trend for different values of κ are reproduced extremely accurately by the effective model. Moreover, the correlation function of x_1 is also reproduced extremely well by the effective model, even in the regime without the scale separation. Statistical features of x_2 and x_3 are also reproduced with comparable accuracy and correlation functions of these two variables are not presented here only for the brevity of presentation.

Numerical results for the bare truncation show that stochastic modeling is essential for reproducing the statistical behavior of x_1 . In particular, the bare truncation severely underestimates the variance of x_1 (Table 3.2) and overestimates the decay of correlations (Figure 3.1).

Simulations of the full model with $k = 2, 1.5, 1$ and corresponding stochastic models show that the Markov chain modeling approach is applicable in the regimes without the scale separation between the essential and the nonessential variables. Markov chain parameterizations also perform well in the regime with clear scale separation. We would like to mention that other techniques are also available for the derivation of reduced models (e.g. [24, 25, 7, 4]) for systems with scale separation. Nevertheless, the Markov chain modeling approach does not seem to be sensitive to this particular issue.

3.1. Comparison of the nonequilibrium statistical behavior of the full and the reduced dynamics. We also compare the nonequilibrium behavior of the triad model in (3.1) and the corresponding effective Markov chains model. In particular, we consider a regime of weak scale separation between the essential and the nonessential variables with $\kappa = 2$ and investigate the behavior of Gaussian ensembles of initial conditions centered on particular values of the reduced variables. Since x_1 is most sensitive to changes in y_1 and y_2 we concentrate on the nonequilibrium behavior of x_1 . We depict only the time-dependent mean and variance of x_1 in the ensemble simulations. We checked the behavior of the time-dependent density of x_1 and verified that it follows a similar trend.

One difficulty arises since the ensemble behavior of the full model in (3.1) depends on the values of the nonessential variables y_1 and y_2 . Therefore, we consider the ensemble of sub-ensembles to investigate the nonequilibrium behavior of the full model. Initial conditions for the essential variables in each sub-ensemble are sampled from a Gaussian distribution with

$$(3.6) \quad \text{mean}\{x_1, x_2, x_3\} = (0.75, 0.9, -0.8), \quad \text{var} = 0.2,$$

but sub-ensembles differ by centering initial conditions for the nonessential variables y_1 and y_2 on different values drawn from a Gaussian distribution with mean zero and variance two, since this distribution corresponds to the stationary distribution of the nonessential variables. We consider 20 sub-ensembles with 200,000 individual trajectories each. The variance of $y_1(0)$ and $y_2(0)$ in each sub-ensemble is also equal to 0.2. Since the effective model does not depend on the nonessential variables, we compute the nonequilibrium behavior of the reduced model by considering a single Gaussian ensemble of initial conditions with (3.6) and 600,000 members. We average the behavior of the full model over all sub-ensembles and depict the comparison between the averaged mean and variance with the mean and variance in the ensemble simulation of the reduced model in Figures 3.2 and 3.3. Error bars correspond to the 95% con-

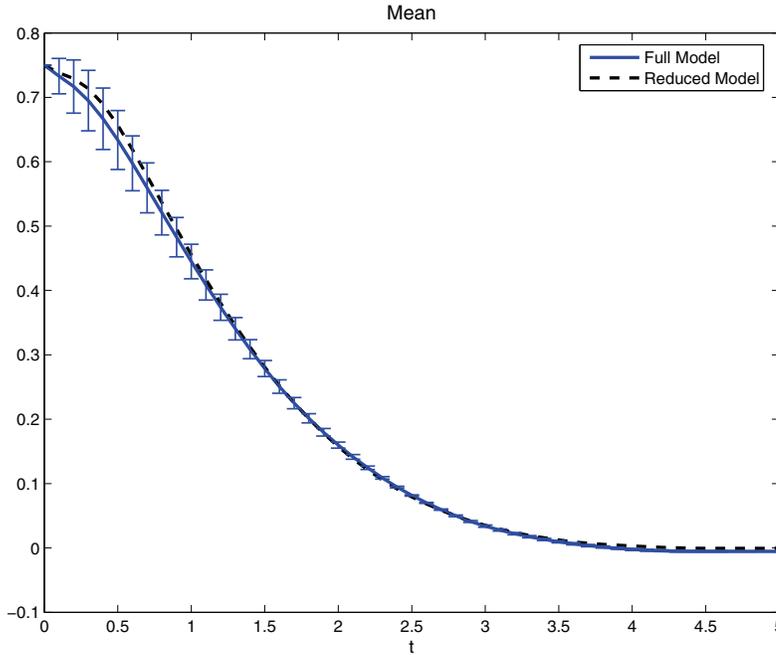


FIG. 3.2. Mean in the ensemble simulations of the full and reduced models.

fidence intervals for the (time-dependent) mean and variance in each sub-ensemble. The reduced model reproduces the nonequilibrium behavior of the full model quite well on average.

In our simulations we verified that the nonequilibrium behavior of the full model depends on the initial conditions for the nonessential variables. Since the scale separation is weak, the nonessential variables do not relax sufficiently quickly to their equilibrium values (*eq.* $mean\{y_1, y_2\} = (0, 0)$) and affect the behavior of the trajectories for short times. There is a considerable variation in the statistical behavior of different sub-ensembles. The reduced model approximates best the statistical behavior of sub-ensembles with $mean\{y_1, y_2\} \approx (0, 0)$. This is not surprising, since the transition probability matrix for the Markov chain in the reduced model is estimated from a single stationary trajectory generated by a long numerical simulation. Therefore, typical values of y_1 and y_2 sample the Gaussian distribution with mean zero and variance two. We illustrate the statistical behavior of two sub-ensembles with different initial distributions for $y_1(0)$ and $y_2(0)$. The behavior of the mean and variance for the first sub-ensemble with the nontypical (large) values of

$$(3.7) \quad mean\{y_1(0), y_2(0)\} = (1.3, 1.4)$$

is depicted in Figure 3.4. There is a considerable discrepancy between the $mean\{x_1\}(t)$ in the ensemble simulations of the full model with (3.7) and the reduced model. Discrepancies in the variance are also possible for other nontypical values of the initial conditions for y_1 and y_2 . Nevertheless, the relaxation time-scale for both mean and variance is reproduced quite accurately by the reduced model. The statistical behavior of $x_1(t)$ in ensemble simulations with typical (small) values of

$$(3.8) \quad mean\{y_1(0), y_2(0)\} = (0.25, -0.4)$$

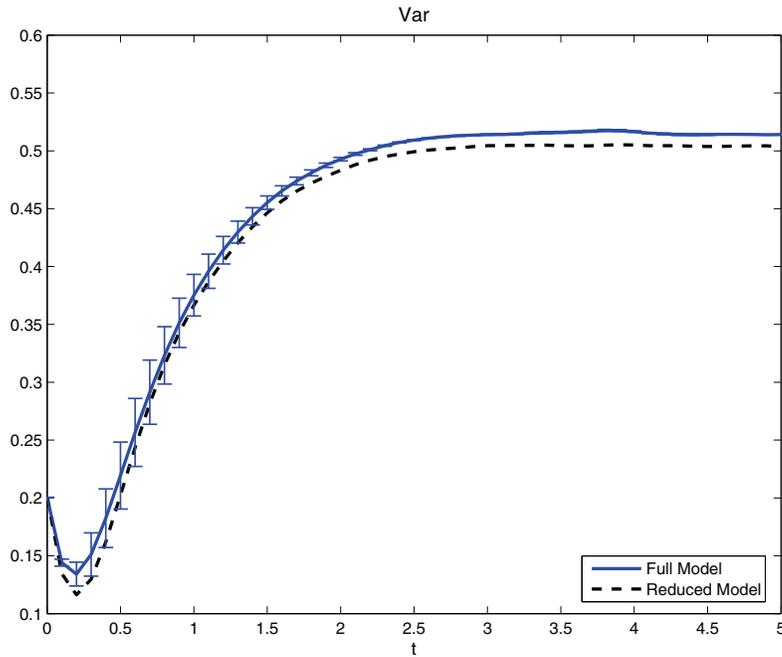


FIG. 3.3. Variance in the ensemble simulations of the full and reduced models.

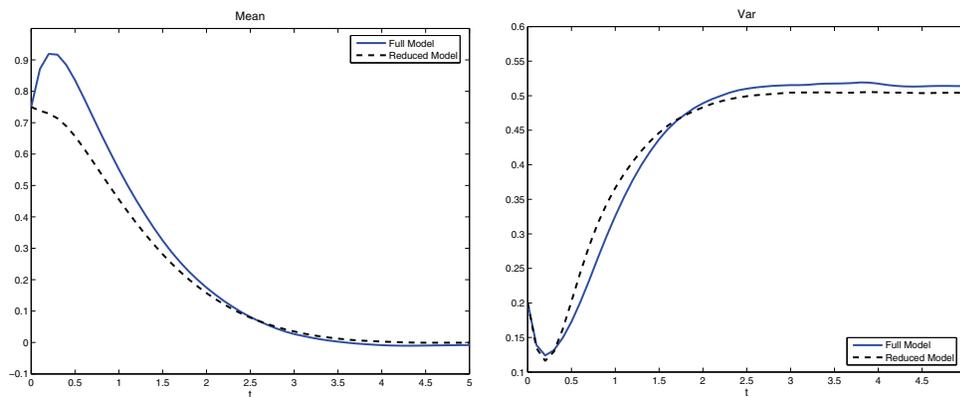


FIG. 3.4. Time-dependent mean (left) and variance (right) in the ensemble simulations of the full (solid line) and reduced (dashed line) models with nontypical initial conditions in (3.7).

is depicted in Figure 3.5. In this ensemble simulation the agreement between the ensemble simulations of the full and reduced models is nearly perfect.

Therefore, the reduced model provides a very good approximation of the nonequilibrium behavior of the full dynamics with typical (small) initial values of the nonessential variables. On the other hand, the relaxation time for the mean and variance is reproduced quite accurately even for nontypical (large) initial values of $y_1(0)$ and $y_2(0)$ (Figure 3.4). We also considered several other values for the means of the essential variables in the statistical ensemble. The nonequilibrium behavior described above appears to be generic; the nonequilibrium behavior of the full model

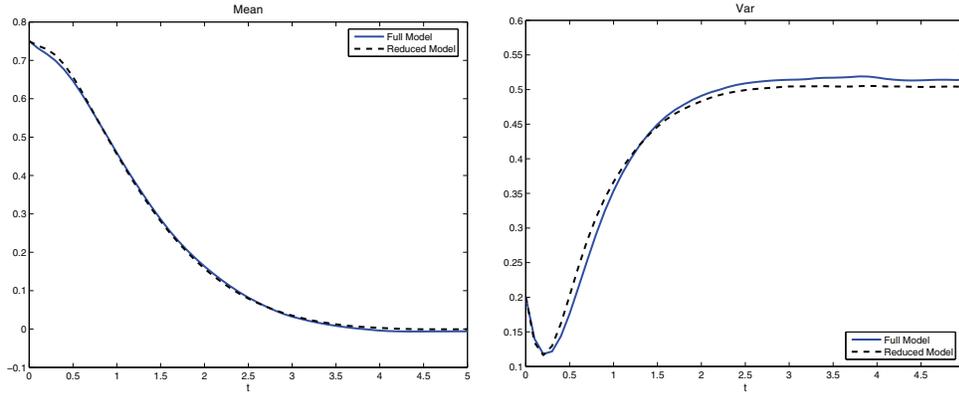


FIG. 3.5. Time-dependent mean (left) and variance (right) in the ensemble simulations of the full (solid line) and reduced (dashed line) models with typical initial conditions in (3.8).

is always approximated extremely well for sub-ensembles with small mean values of $y_1(0), y_2(0)$; on the other hand, various discrepancies often appear for sub-ensembles centered on large mean values of $y_1(0), y_2(0)$.

4. Truncated Burgers–Hopf model. As an application of the Markov chain stochastic parametrization we consider spectral projection of the inviscid Burgers–Hopf model in periodic geometry. In particular, we consider the truncated Burgers–Hopf (TBH) model

$$(4.1) \quad \partial_t U_\Lambda(x, t) + \frac{1}{2} \partial_x P_\Lambda U_\Lambda^2(x, t) = 0,$$

where $x \in [0, 2\pi]$, P_Λ is the projection operator in Fourier space

$$(4.2) \quad P_\Lambda u(x, t) = \sum_{|k| \leq \Lambda} u_k(t) e^{ikx},$$

and U_Λ is a finite dimensional projection

$$(4.3) \quad U_\Lambda(x, t) = \sum_{|k| \leq \Lambda} u_k(t) e^{ikx}.$$

We supplement (4.3) with the reality condition $u_{-k}(t) = u_k^*(t)$. The equation in (4.1) can be recast as a 2Λ -dimensional system of ordinary differential equations

$$(4.4) \quad \frac{d}{dt} u_k = -\frac{ik}{2} \sum_{\substack{k+p+q=0 \\ |p|, |q| \leq \Lambda}} u_p^* u_q^*, \quad |k| \leq \Lambda.$$

The equations in (4.4) conserve energy

$$E = \frac{1}{4\pi} \int U_\Lambda^2 dx = \sum_{k=1}^\Lambda |u_k|^2$$

and Hamiltonian

$$H = \frac{1}{12\pi} \int_0^{2\pi} U_\Lambda^3 dx.$$

Since the equation for u_0 is trivial, we can assume $u_0(0) = 0$ without the loss of generality. Statistical properties of this model were studied in detail in [22, 23, 1]. Statistical behavior of this model indicates that the TBH model is ergodic on the fixed energy-Hamiltonian surface. In particular, it was demonstrated that generic initial conditions correspond to $H \approx 0$. Moreover, for generic initial conditions Fourier coefficients achieve equipartition and equilibrium statistical properties of Fourier coefficients follow a joint Gaussian distribution

$$(4.5) \quad \pi(u_1, u_2, \dots, u_\Lambda) = Ce^{-\beta E},$$

where C is a normalization constant. Nevertheless, it was demonstrated in [1] that Hamiltonian influences the statistical behavior of the model. In particular, for non-generic initial conditions with large values of H the TBH model in (4.4) exhibits a tilt in the spectrum and oscillatory correlation functions.

The TBH model in (4.4) is not relevant for shock formation and studies of hyperbolic conservation laws in general. Since the TBH model is projected onto a finite number of Fourier modes, the energy cannot transfer to infinity in Fourier space. Instead, the energy in the truncation is redistributed between the high and the low wavenumbers resulting in an approximate equipartition of energy. The TBH model in (4.4) has been utilized in several studies of various statistical theories [23, 20, 25, 27, 1, 21]. We will utilize this equation to test the performance of the Markov chain modeling approach in a model with nonlinear features of more complicated turbulence systems.

5. Markov chain modeling for the truncated Burgers–Hopf. In the dimension reduction framework, we are typically interested in deriving a reduced stochastic model for a first few Fourier modes. Let $\lambda < \Lambda$ be the number of modes of interest. We call u_k , $k = 1, 2, \dots, \lambda$ as the essential variables and u_k , $k = \lambda + 1, \dots, \Lambda$ as the nonessential variables. From (4.4), we can decompose the right-hand side of the equation as a sum of two terms as follows:

$$(5.1) \quad \frac{du_k}{dt} = -\frac{ik}{2} \sum_{\substack{k+p+q=0 \\ |p|, |q| \leq \lambda}} u_p^* u_q^* - ik \left(\sum_{p=\lambda-k+1}^{\Lambda-k} u_p^* u_{p+k} \right).$$

The first term on the right-hand side of (5.1) represents interactions of the essential variables with themselves. The second term represents interactions with the nonessential variables. It involves two types of interactions—(i) the essential with the nonessential modes and (ii) the nonessential variables with themselves. Since we are interested in deriving an effective model for the essential variables, the second term in (5.1) is modeled by a Markov chain. Thus, the dynamics of the essential variables can be written as follows:

$$(5.2) \quad \frac{du_k}{dt} = -\frac{ik}{2} \sum_{\substack{k+p+q=0 \\ |p|, |q| \leq \lambda}} u_p^* u_q^* + m_k(u_k).$$

Since the second term on the right-hand side of (5.1) depends on the essential variables, Markov chains in the effective model should be conditioned on u_k , $k = 1 \dots \lambda$. It is computationally infeasible to condition Markov chains on all essential variables; it requires estimating a transition probability matrix of a very high dimensionality. Therefore, in (5.2) we introduced a much simpler conditioning where the Markov

chain for the essential variable u_k is conditioned only on u_k itself. Typically, such simplifications are necessary to reduce the computational complexity of estimating the transition probability matrix in the conditional Markov chain.

Although reduced models in (5.2) do not conserve energy or Hamiltonian, we will show that they reproduce the averaged behavior of solutions quite accurately. This can be expected, since Markov chain reduced models are trained on the solution which evolves on the appropriate energy-Hamiltonian surface.

Statistical behavior of the full model is crucial for selecting the appropriate Markov chain modeling framework. In particular, imposing the proper conditioning in the Markov chain is highly nontrivial and model dependent. Nevertheless, models considered here provide general guidelines for the conditioning of the Markov chains and demonstrate that this conditioning can be rather straightforward.

5.1. $\lambda = 1$; u_1 is the essential variable. $\lambda = 1$ corresponds to u_1 being the only essential variable. Therefore, there are no self-interactions of the essential variables with themselves; the first term on the right-hand side of (5.1) is not present.

As discussed in [1], Hamiltonian plays an important role in the statistical behavior of the full model. In particular, when $H \approx 0$, cross-correlations between real and imaginary components of the same mode are extremely weak. Therefore, for $H \approx 0$ the effective model written in real notation becomes

$$(5.3) \quad \begin{aligned} \frac{d}{dt}u_1^{re} &= m_1^{re}(u_1^{re}), \\ \frac{d}{dt}u_1^{im} &= m_1^{im}(u_1^{im}). \end{aligned}$$

Here we introduced further simplifications for the conditional Markov chains; the Markov chain m_1^{re} depends only on the real component, u_1^{re} , and the Markov chain m_1^{im} depends only on the imaginary component, u_1^{im} . Therefore, equations in the effective model in (5.3) are completely decoupled.

The partition for u_1^{re} and u_1^{im} is defined with $M = 6$ intervals

$$(5.4) \quad P_u = \{-\infty < \dots < \bar{u} - 1.4sd(u) < \dots < \bar{u} - 0.6sd(u) < \dots < \bar{u} < \dots < \bar{u} + 0.6sd(u) < \dots < \bar{u} + 1.4sd(u) < \dots < \infty\}$$

with $u = u_1^{re}$, $u = u_1^{im}$. We utilize $N = 7$ states in the Markov chains m_1^{re} and m_1^{im} with partitions defined as

$$(5.5) \quad P_m = \{-\infty < \dots < \bar{m} - 2.0sd(m) < \dots < \bar{m} - 1.0sd(m) < \dots < \bar{m} - 0.3sd(m) < \dots < \bar{m} + 0.3sd(m) < \dots < \bar{m} + 1.0sd(m) < \dots < \bar{m} + 2.0sd(m) < \dots < \infty\}$$

with $m = RHS(u_1^{re})$, $RHS(u_1^{im})$. The time-step for the estimation of the Markov chain model is $\Delta t = 0.1$. Therefore, $\Delta t = 0.1$ is also the time-step for the numerical integration of the effective Markov chain model.

We utilize the following parameters in the simulation.

$$\Lambda = 20, \quad \beta = 50.$$

Transition probabilities for the Markov chains are estimated with the time-step $\Delta t = 0.1$ for $T_{MC} = 20,000$. Statistical properties of the TBH model are computed from

a single microcanonical simulation with $T = 200,000$. We utilize initial conditions generated on the surface $H = 0$ and, therefore, Hamiltonian remains zero for the total length of the simulation.

Table 5.1 and Figures 5.2 and 5.1 show statistical behavior of u_1^{re} in the full TBH model with $H = 0$ and in the corresponding effective equation. Results for u_1^{im} are similar and not presented here only for the brevity of presentation. Overall, there is a very good agreement for the one-point statistical quantities (Table 5.1 and Figure 5.1), including higher-order moments, since marginal probability densities of u_1^{re} in the full and the effective equations depicted in Figure 5.1 are nearly identical. The structure of the correlation function of u_1^{re} from the effective model has a slight discrepancy compared with the full model, but the overall decay rate is reproduced sufficiently accurately. It is unrealistic to expect that any effective models would be capable of reproducing two-point statistical quantities exactly since correlation functions are complex objects determined by the full nonlinear dynamics of the original equations. Detailed structure of nonlinear interactions is lost in any effective model and small discrepancies in statistical quantities for the full and effective models are inevitable. Similar discrepancies were also observed in the analytical approach for the derivation of the reduced equation for the TBH described in [27].

TABLE 5.1

Means and variances of u_1^{re} , and u_1^{im} , from the full model with $H = 0$ and the effective Markov chain model for u_1 in (5.3).

	Theory	Full Model	Effective Model
Mean of u_1^{re}	0.0	5.702×10^{-4}	-1.572×10^{-4}
Mean of u_1^{im}	0.0	2.104×10^{-3}	3.209×10^{-3}
Variance of u_1^{re}	0.01	1.014×10^{-2}	1.034×10^{-2}
Variance of u_1^{im}	0.01	1.034×10^{-2}	1.019×10^{-2}

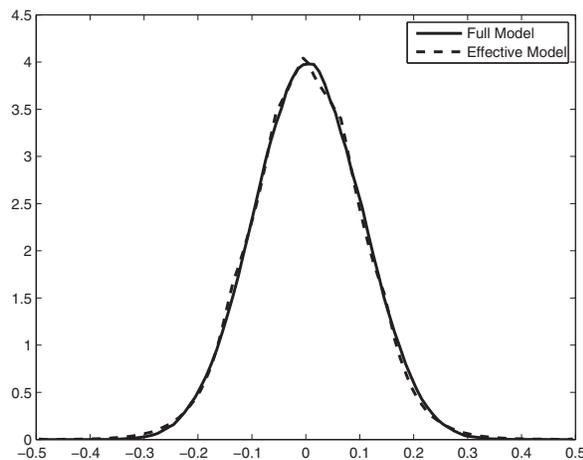


FIG. 5.1. Marginal probability density of u_1^{re} from the full model with $H = 0$ (solid line) and the corresponding effective Markov chain model in (5.3) (dashed line).

Statistical behavior of the TBH model changes considerably when $H \neq 0$. In particular, one major difference is the occurrence of the cross-correlation between u_1^{re} and u_1^{im} (depicted in the top right part of Figure 5.3). Thus, the effective model

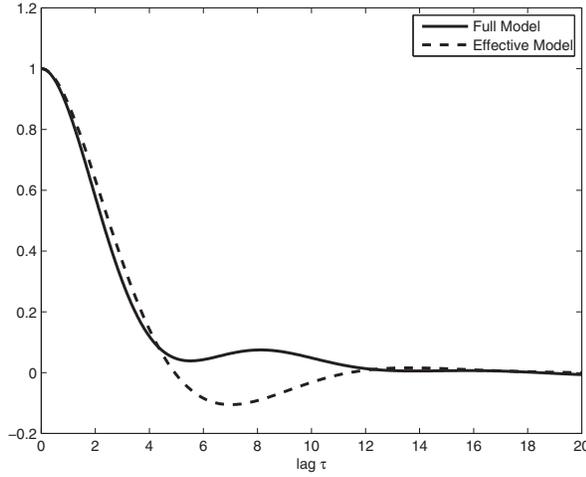


FIG. 5.2. Correlation functions of u_1^{re} from the full model with $H = 0$ (solid line) and the corresponding effective Markov chain model in (5.3) (dashed line).

should take the interaction between u_1^{re} and u_1^{im} into account in this case. Therefore, we consider the effective model with Markov chains depending on both u_1^{re} and u_1^{im} :

$$(5.6) \quad \begin{aligned} \frac{d}{dt}u_1^{re} &= m_1^{re}(u_1^{re}, u_1^{im}), \\ \frac{d}{dt}u_1^{im} &= m_1^{im}(u_1^{re}, u_1^{im}). \end{aligned}$$

The partition for the primary variable in the Markov chain is identical to the partition in (5.4). We utilize a very simple partition for the secondary variable,

$$(5.7) \quad P_v = \{-\infty < \dots < \bar{v} < \dots < \infty\}.$$

Conditioning in (5.7) consists of only two intervals $(-\infty, \bar{v})$ and (\bar{v}, ∞) . For instance, to condition m_1^{re} we utilize the partition (5.4) for u_1^{re} and the partition (5.7) for u_1^{im} . The roles of u_1^{re} and u_1^{im} are reversed in the conditioning of m_1^{im} .

Statistical agreement between the full TBH model with $H \neq 0$ and the corresponding effective model in (5.6) is presented in Figure 5.3. In particular, the marginal probability density function of u_1^{re} presented in the left top part of Figure 5.3 demonstrates good agreement for the one-point statistics of u_1^{re} . Results for u_1^{im} are similar and not presented here only for the brevity of presentation. The nontrivial cross-correlation between u_1^{re} and u_1^{im} is presented in the right top part of Figure 5.3. The structure of the cross-correlation is reproduced well by the effective model. There is still a small discrepancy between the correlation functions computed from the full TBH model and the effective model. Nevertheless, the overall structure of correlation functions and correlation times of u_1^{re} and u_1^{im} is reproduced reasonably well by the effective model.

Proper conditioning of the Markov chains in the effective model solves the problem of the cross-correlation between effective variables. We would like to emphasize that very simple additional conditioning presented in (5.7) is sufficient to represent the structure of the cross-correlation between the effective variables. Conditioning on the secondary variable with the partition in (5.7) can be implemented for more than two

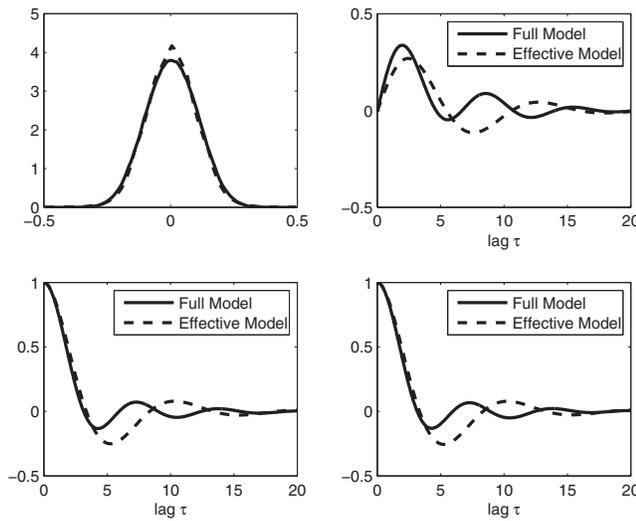


FIG. 5.3. Comparison of the statistical behavior of the full model (solid line) and the effective model (dashed line) in (5.6) with $H \neq 0$. Top left - marginal probability density function of u_1^{re} . Top right - cross-correlation between u_1^{re} and u_1^{im} . Bottom left, bottom right - correlation functions of u_1^{re} and u_1^{im} , respectively.

effective variables. Thus, we expect that the Markov chain modeling approach can be generalized to situations with many essential variables. This will be discussed in the context of a more realistic geophysical example in a subsequent paper.

5.2. $\lambda = 2$; u_1 and u_2 are the essential variables. In this section we consider the first two modes, u_1 and u_2 , as the essential variables. The main difference between this case (two essential modes) and the previous section (only one essential mode) is that there is a nontrivial self-interaction between modes u_1 and u_2 which have to be included in the effective model. The TBH model in (4.4) can be rewritten to emphasize the self-interaction between the essential variables as follows:

$$\begin{aligned}
 \frac{d}{dt}u_1 &= -iu_2u_1^* + \{TBH\}, \\
 \frac{d}{dt}u_2 &= -iu_1^2 + \{TBH\}, \\
 \frac{d}{dt}u_k &= \{TBH\},
 \end{aligned}
 \tag{5.8}$$

where $\{TBH\}$ denotes the TBH terms introduced in (4.4). Therefore, rewritten through the real and imaginary components the effective model for u_1 and u_2 becomes

$$\begin{aligned}
 \frac{d}{dt}u_1^{re} &= u_1^{re}u_2^{im} - u_1^{im}u_2^{re} + m_1, \\
 \frac{d}{dt}u_1^{im} &= -u_1^{re}u_2^{re} - u_1^{im}u_2^{im} + m_2, \\
 \frac{d}{dt}u_2^{re} &= 2u_1^{re}u_1^{im} + 2m_3, \\
 \frac{d}{dt}u_2^{im} &= (u_1^{im})^2 - (u_1^{re})^2 + 2m_4.
 \end{aligned}
 \tag{5.9}$$

We assume that Markov chains can be conditioned only on the corresponding essential variable, i.e.,

$$m_1 = m_1(u_1^{re}), \quad m_2 = m_2(u_1^{im}),$$

$$m_3 = m_3(u_2^{re}), \quad m_4 = m_4(u_2^{im}).$$

We utilize the identical partition for conditioning of Markov chains on $u_{1,2}^{re,im}$ defined in (5.4). Partitions for the states of Markov chains are also the same as in the previous section and are defined in (5.5). The time-step, Δt , for the estimation of the discrete Markov chains and integration of the effective model is $\Delta t = 0.1$.

Comparison of means and variances is presented in Table 5.2 and comparison of marginal densities is depicted in Figure 5.4. Similar to the previous section,

TABLE 5.2

Means and variances of u_1 and u_2 in the simulations of the full model (5.8) and the effective model (5.9).

	Theory	Full Model	Effective Model
Mean of u_1^{re}	0.0	-6.169×10^{-4}	-2.935×10^{-3}
Mean of u_1^{im}	0.0	-7.013×10^{-4}	-1.449×10^{-3}
Mean of u_2^{re}	0.0	-3.351×10^{-4}	1.204×10^{-3}
Mean of u_2^{im}	0.0	-2.225×10^{-4}	-4.538×10^{-4}
Variance of u_1^{re}	0.01	9.854×10^{-3}	1.006×10^{-2}
Variance of u_1^{im}	0.01	9.845×10^{-3}	9.690×10^{-3}
Variance of u_2^{re}	0.01	9.789×10^{-3}	9.959×10^{-3}
Variance of u_2^{im}	0.01	9.720×10^{-3}	1.002×10^{-2}

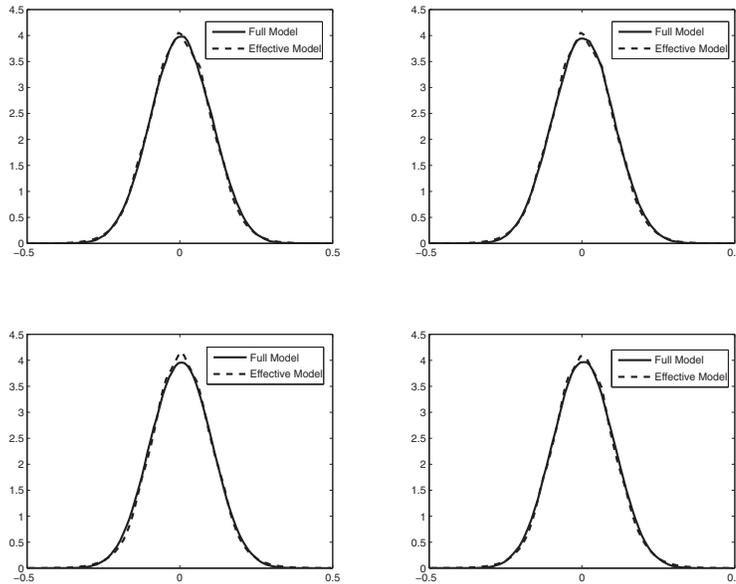


FIG. 5.4. Comparison of marginal probability densities of the essential variables in the simulations of the full model (5.8) and the effective Markov chain model (5.9). Top left - u_1^{re} . Top right - u_1^{im} . Bottom left - u_2^{re} . Bottom right - u_2^{im} .

one-point statistics, including higher moments, is reproduced very well by the effective model.

Comparison of the correlation functions for u_1 and u_2 is presented in Figure 5.5. Although discrepancies between the full TBH model and the effective model are larger for u_2 , correlation times are still reproduced well in this case.

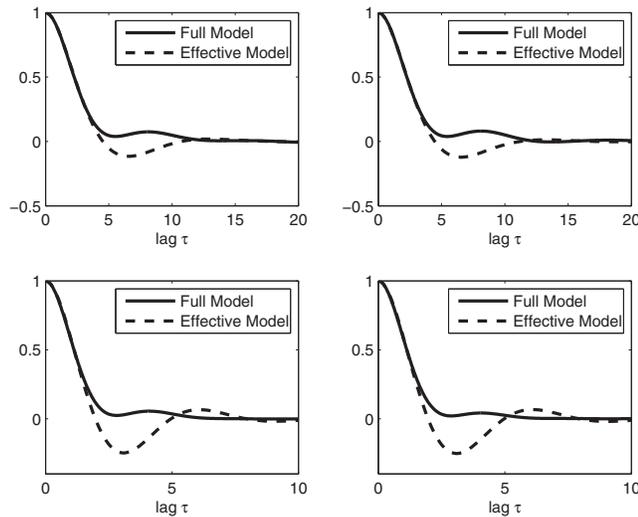


FIG. 5.5. Comparison of correlation functions of the essential variables in the simulations of the full model (5.8) and the effective Markov chain model (5.9). Top left - u_1^e . Top right - u_1^m . Bottom left - u_2^e . Bottom right - u_2^m .

6. Conclusions. In this paper we consider the Markov chain modeling approach for derivation of effective models for the essential degrees of freedom. Models of this type can potentially be utilized in stochastic parametrizations of unresolved processes in geophysical fluid dynamics. One of the main advantages of the Markov chain modeling approach is a straightforward estimation of the effective model. Therefore, this approach can potentially be utilized in various interdisciplinary models of very high complexity. On the other hand, there is an empirical component of the approach consisting of conditioning of the Markov chain which cannot be predicted a priori. This problem is easily overcome in models presented here, since very simple conditioning yields considerable improvement of the effective model.

We demonstrate that the Markov chain modeling approach can be potentially utilized in systems without strong scale separation. The Coupled triad model considered in section 3 lacks the scale separation between the essential and the nonessential variables in some parameter regimes. Nevertheless, the effective Markov chain model performs extremely well with and without the scale separation. Therefore, we expect that the Markov chain modeling approach can be successful in systems without strong scale separation, including models from fluid dynamics.

Nonequilibrium tests of the triad model in section 3.1 show that the reduced model reproduces the mean and variance of the full model extremely well for ensembles centered on typical (small) values of the nonessential variables, but various discrepancies arise for ensembles centered on larger values of nonessential variables. Nevertheless, the relaxation time for the mean and variance is reproduced well in all

cases. We expect that the nonequilibrium comparison between the full and Markov chain reduced models described in section 3.1 is generic, since the transition probability matrix for the Markov chain is estimated from a single long equilibrium time-series of the nonessential variables. Therefore, the Markov chain parametrization reproduces the behavior of the full model best when the nonessential variables are close to their equilibrium mean.

Proper conditioning of Markov chains is one of the main practical issues since it is difficult to determine dependencies between the essential variables a priori. On the other hand, time-series of the essential variables are often available and simple data-analysis can be utilized to determine cross-correlations between the essential variables. This situation is typical in the geophysical fluid dynamics. Although it is infeasible to utilize detailed partitions for conditioning of Markov chains on all essential variables, simple conditioning significantly improves the performance of the effective model. For instance, effective Markov chain models developed for the TBH equation demonstrate that Markov chains must be properly conditioned to capture statistical behavior of the essential variables. On the other hand, we also demonstrate that Markov chains can be conditioned rather straightforwardly. In particular, a very simple partition (5.7) consisting of only two intervals for the secondary variable drastically improves the performance of the effective model. Thus, simple conditioning should be sufficient to reproduce the statistical behavior of the essential variables. Therefore, we expect that the Markov chain parametrization approach should be applicable in complex systems and, in particular, in geophysical fluid dynamics. We will address this issue in a more realistic geophysical model in a subsequent paper.

Acknowledgment. The authors thank E. Vanden-Eijnden for his comments and suggestions.

REFERENCES

- [1] R. ABRAMOV, G. KOVACIC, AND A. J. MAJDA, *Hamiltonian structure and statistically relevant conserved quantities for the truncated Burgers-Hopf equation*, *Comm. Pure Appl. Math.*, 56 (2003), pp. 1–46.
- [2] R. AZENCOTT, A. BERI, AND I. TIMOFEYEV, *Adaptive sub-sampling for parametric estimation of Gaussian diffusions*, *J. Stat. Phys.*, 139 (2010), pp. 1066–1089.
- [3] R. AZENCOTT, A. BERI, AND I. TIMOFEYEV, *Sub-sampling in parametric estimation of stochastic differential equations from discrete data*, *Multiscale Model. Simul.*, 2010, submitted.
- [4] A. J. CHORIN, O. H. HALD, AND R. KUPFERMAN, *Optimal prediction and the Mori-Zwanzig representation of irreversible processes*, *Proc. Natl. Acad. Sci. USA*, 97 (2000), pp. 2968–2973.
- [5] A. J. CHORIN, O. H. HALD, AND R. KUPFERMAN, *Optimal prediction with memory*, *Phys. D*, 166 (2002), pp. 239–257.
- [6] A. J. CHORIN, A. P. KAST, AND R. KUPFERMAN, *Optimal prediction of underresolved dynamics*, *Proc. Natl. Acad. Sci. USA*, 95 (1998), pp. 4094–4098.
- [7] A. J. CHORIN, A. P. KAST, AND R. KUPFERMAN, *Unresolved computation and optimal prediction*, *Comm. Pure Appl. Math.*, 52 (1999), pp. 1231–1254.
- [8] D. CROMMELIN AND E. VANDEN-EIJNDEN, *Fitting timeseries by continuous-time Markov chains: A quadratic programming approach*, *J. Comput. Phys.*, 217 (2006), pp. 782–805.
- [9] D. CROMMELIN AND E. VANDEN-EIJNDEN, *Reconstruction of diffusions using spectral data from time-series*, *Commun. Math. Sci.*, 4 (2006), pp. 651–668.
- [10] D. CROMMELIN AND E. VANDEN-EIJNDEN, *Subgrid-scale parameterization with conditional Markov chains*, *J. Atmospheric Sci.*, 65 (2008), pp. 2661–2675.
- [11] W. E AND B. ENGQUIST, *The heterogeneous multiscale methods*, *Commun. Math. Sci.*, 1 (2003), pp. 87–132.
- [12] W. E, B. ENGQUIST, X. LI, W. REN, AND E. VANDEN-EIJNDEN, *Heterogeneous multiscale methods: A review*, *Commun. Comput. Phys.*, 2 (2007), pp. 367–450.

- [13] W. E. D. LIU, AND E. VANDEN-EIJNDEN, *Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates*, J. Chem. Phys., 123 (2005), 194107.
- [14] E. VANDEN-EIJNDEN E W., D. LIU, *Analysis of multiscale methods for stochastic differential equations*, Comm. Pure Appl. Math., 58 (2005), pp. 1544–1585.
- [15] C. FRANZKE, D. CROMMELIN, A. FISCHER, AND A. MAJDA, *A hidden Markov model perspective on regimes and metastability in atmospheric flows*, J. Climate, 21 (2008), pp. 1740–1757.
- [16] C. HIJON, P. ESPANOL, E. VANDEN-EIJNDEN, AND R. DELGADO-BUSCALIONI, *Mori-zwanzig formalism as a practical computational tool*, Faraday Discuss., 144 (2010), pp. 301–322.
- [17] M. A. KATSOUKAKIS, A. J. MAJDA, AND A. SOPASAKIS, *Multiscale couplings in prototype hybrid deterministic/stochastic systems: Part 1, deterministic closures*, Commun. Math. Sci., 2 (2004), pp. 255–294.
- [18] M. A. KATSOUKAKIS, A. J. MAJDA, AND A. SOPASAKIS, *Multiscale couplings in prototype hybrid deterministic/stochastic systems: Part 2, stochastic closures*, Commun. Math. Sci., 3 (2005), pp. 453–478.
- [19] M. A. KATSOUKAKIS, A. J. MAJDA, AND A. SOPASAKIS, *Intermittency, metastability and coarse graining for coupled deterministic-stochastic lattice systems*, Nonlinearity, 19 (2006), pp. 1027–1047.
- [20] R. KLEEMAN, A. J. MAJDA, AND I. TIMOFEYEV, *Quantifying predictability in a model with statistical features of the atmosphere*, Proc. Natl. Acad. Sci. USA, 99 (2002), pp. 15291–15296.
- [21] X. SAN LIANG AND R. KLEEMAN, *A rigorous formalism of information transfer between dynamical system components. II. Continuous flow*, Phys. D, 227 (2007), pp. 173–182.
- [22] A. J. MAJDA AND I. TIMOFEYEV, *Remarkable statistical behavior for truncated Burgers-Hopf dynamics*, Proc. Natl. Acad. Sci. USA, 97 (2000), pp. 12413–12417.
- [23] A. J. MAJDA AND I. TIMOFEYEV, *Statistical mechanics for truncations of the Burgers-Hopf equation: A model for intrinsic stochastic behavior with scaling*, Milan J. Math., 70 (2002), pp. 39–96.
- [24] A. J. MAJDA, I. TIMOFEYEV, AND E. VANDEN-EIJNDEN, *A mathematics framework for stochastic climate models*, Comm. Pure Appl. Math., 54 (2001), pp. 891–974.
- [25] A. J. MAJDA, I. TIMOFEYEV, AND E. VANDEN-EIJNDEN, *A priori tests of a stochastic mode reduction strategy*, Phys. D, 170 (2002), pp. 206–252.
- [26] A. J. MAJDA, I. TIMOFEYEV, AND E. VANDEN-EIJNDEN, *Systematic strategies for stochastic mode reduction in climate*, J. Atmospheric Sci., 60 (2003), pp. 1705–1722.
- [27] A. J. MAJDA, I. TIMOFEYEV, AND E. VANDEN-EIJNDEN, *Stochastic models for selected slow variables in large deterministic systems*, Nonlinearity, 19 (2006), pp. 769–794.
- [28] K. NIMSAILA, *Markov Chain and Time-Delay Reduced Modeling of Nonlinear Systems*, Ph.D. dissertation, University of Houston, Houston, TX, 2009.
- [29] G. A. PAVLIOTIS AND A. STUART, *Parameter estimation for multiscale diffusions*, J. Stat. Phys., 127 (2007), pp. 741–781.