Markov chain Monte Carlo method in Bayesian reconstruction of dynamical systems from noisy chaotic time series

E. M. Loskutov,* Ya. I. Molkov, D. N. Mukhin,† and A. M. Feigin

Institute of Applied Physics, Russian Academy of Sciences, 46, Uljanov Street, Nizhniy Novgorod 603950, Russia

The impossibility to use the MCMC (Markov chain Monte Carlo) methods for long noisy chaotic time series (TS) (due to high computational complexity) is a serious limitation for reconstruction of dynamical systems (DSs). In particular, it does not allow one to use the universal Bayesian approach for reconstruction of a DS in the most interesting case of the unknown evolution operator of the system. We propose a technique that makes it possible to use the MCMC methods for Bayesian reconstruction of a DS from noisy chaotic TS of arbitrary long duration.

I. INTRODUCTION

The presence of a noise component in the time series generated by a dynamical system (DS) results in finite accuracy of reconstruction of the evolution operator (EO): Any model that can be constructed will, generally speaking, differ from the original system. One of the ways to take this fact into consideration is to describe parameters of the model EO as random quantities [1]. Two limiting cases may be distinguished [2]: The "perfect model" scenario when the form of the operator describing the DS evolution is reliably known, and the "imperfect model" scenario when EO is unknown.

The perfect model scenario is realized, for instance, when the problem of hidden information transmission is solved using chaotic time series (TS) generated by the known DS [3]. In this case we are interested in the characteristics of the system ("perfect model") parameters (the most probable value, mean, dispersion, distribution). In investigations of many "natural" (atmospheric-oceanic, tectonic, biological) systems the EO is unknown, which corresponds to the imperfect model scenario. Then, information about model parameters is of no value as the physical meaning of the parameters is unknown, and we are interested in the properties of the model defined by these parameters and reflecting the properties of the reconstructed DS. Nevertheless, in this case too evaluation of statistical characteristics of model parameters is the key element without which the problem of reconstructing a DS generating the original TS cannot be solved.

The mathematical body used to reconstruct the DS is determined by a specific application. For example, for finding the most probable set of model (perfect or imperfect) parameters it suffices to determine the maximum of posterior probability density of model parameters [4–6]. However, it should be remembered that in the presence of noise the inverse problem of reconstruction becomes ill-posed, i.e., it admits an infinite set of solutions. Selection of the "most correct" solutions demands regularization that is a physically justified constraint on admissible values of parameters based solely on *a priori* information [7]. This approach is usually called Bayesian. Within the Bayesian approach, there are methods which make it possible to estimate dispersions and mathematical expectations of the model parameters [8]. A significant advantage of such methods is their relatively low computational resource requirements. However, due to the model nonlinearity, the form of parameter distribution may differ strongly from the normal one. In this case, estimation of expectation and dispersion is insufficient to construct a correct model, and one must use the Bayesian approach in full. In other words, we must construct models which include distributions of parameters and take *a priori* information about the system into account correctly. In this formulation the problem can be solved using the Markov chain Monte Carlo (MCMC) algorithms [9]. However, in the case of reconstruction of DSs by noisy chaotic time series, computational resources required for these algorithms strongly depend on distribution dimensionality. This limits applicability of the algorithms to a narrow class of problems, e.g., the case of short TS [10]. The method proposed in this paper makes it possible to expand the applicability domain of the MCMC algorithms to the case of arbitrarily long TS.

II. FORMULATION OF THE PROBLEM

In what follows, we will assume that the available time series can be used to state the fact of dynamism of the system that has generated it, to determine minimum embedding dimension d [11], and to reconstruct the phase trajectory [12] $\{\mathbf{x}_t\}_{t=0}^{T}, \mathbf{x}_t \in R^d$, and t enumerates moments of the discrete time. Let us assume that the system experimented upon has a set of properties (parameters) $\boldsymbol{\mu}$ which cannot be measured directly. Let us consider, for the sake of certainty, a DS with discrete time

$$\mathbf{x}_t = \mathbf{U}_t + \boldsymbol{\xi}_t, \quad \mathbf{U}_t = \mathbf{f}(\mathbf{U}_{t-1}, \boldsymbol{\mu}) + \boldsymbol{\eta}_t. \quad (1)$$

Here, the vector $\mathbf{x} = \{x_k\}_{k=1}^{d}$, as mentioned above, is an observable quantity; $\mathbf{U} = \{\mathbf{U}_k\}_{k=1}^{d}$ is a latent variable characterizing the true state of the dynamical system in the d-dimensional phase space; $\mathbf{f}(\mathbf{U}, \boldsymbol{\mu})$ is the vector function describing the DS evolution operator; $\boldsymbol{\mu}$ are unobserved parameters; and $\boldsymbol{\xi}_t$ and $\boldsymbol{\eta}_t$ are random quantities ("noises") with distributions w_ξ and w_η. Then, in accordance with the Bayesian theorem [13], the

*loskutov@appl.sci-nnov.ru
†mukhin@appl.sci-nnov.ru

posterior conditional probability density (PD) $P(\boldsymbol{\mu}, \mathbf{U}|\mathbf{x})$ for the unobserved parameters $\boldsymbol{\mu}$ an unknown (latent) states of the DS are proportional to the product of their prior PD $P(\boldsymbol{\mu}, \mathbf{U})$ and conditional PD for the obtained experimental results $P(\mathbf{x}|\boldsymbol{\mu}, U)$,

$$P(\boldsymbol{\mu}, \mathbf{U}|\mathbf{x}) \propto P(\boldsymbol{\mu}, \mathbf{U}) P(\mathbf{x}|\boldsymbol{\mu}, \mathbf{U}). \qquad (2)$$

In the case of an unknown DS, the evolution operator $\mathbf{f}(\mathbf{U}, \boldsymbol{\mu})$ is a model of the reconstructed system, and the random quantity $\boldsymbol{\eta}_t$ describes the defect of the model, i.e., the discrepancy between the model and the reconstructed system [14]. To write the probability density function (PDF) several assumptions are made. First, the model is supposed to be so "good" that the width of distribution w_ξ is much larger than the width of distribution w_η, so that the latter can be regarded to be the δ function. Second, following [10], we take into account that for the chaotic TS the information coupling between the TS points is known to decrease with increasing time interval between the points. In other words, the system starts to "forget" its initial state with time. Hence, assuming the states of the system separated by large time intervals to be independent we can regard the latent variables to be coupled only in finite time periods ("segments") of length $(w+1)$. As a result, expression (2) for $P(\boldsymbol{\mu}, \mathbf{U}|\mathbf{x})$ can be transformed (see [10] for details) to

$$P(\boldsymbol{\mu}, \mathbf{U}|\mathbf{x}) \propto P(\boldsymbol{\mu}, \mathbf{U}) \prod_{t=0}^{M-1} \prod_{j=0}^{w} w_\xi[\mathbf{x}_{t(w+1)+j+1} - \mathbf{f}^j(\mathbf{U}_{t(w+1)+1}, \boldsymbol{\mu})],$$

$$(3)$$

where T is the total length of the TS; $M = T/(w+1)$ is the number of segments into which the initial TS is split assuming that there is no information coupling between the TS points belonging to different segments; and $\mathbf{f}^j(\cdots)$ is the evolution operator applied j times [15]. Note that for $w = T$ the expression (3) transforms to the classical Bayesian expression which, as was noted above, is inapplicable for reconstruction by noisy chaotic time series. For $w < T$, (3) is, strictly speaking, an approximate expression. Its applicability for reconstruction by a concrete time series was justified in [10] and reduces to a correct choice of w ("the larger, the better"). In Sec. IV we will show that the lower limit on w that ensures correctness of the expression (3) is a sufficient condition of the applicability of the proposed modification of the MCMC method. More detailed information about prior parameter distribution $P(\boldsymbol{\mu}, \mathbf{U})$ will be given in Secs. III and IV.

The obtained PDF (3) includes all information about the system extracted from the TS, and analysis of the ensemble of parameters $\boldsymbol{\mu}$, distributed according to (3), enables one to assess needed characteristics of the system. As this function depends on a large number of arguments, the set goal may be achieved by means of the MCMC analysis [9] that proved to be efficient for investigation of high-dimensional distributions.

In practice, application of these methods is characterized by extreme computational complexity in the cases when the EO is an essentially nonlinear function and the arguments of the distribution function in question are cross correlated,

which leads to long times of self-correlation of the corresponding MCMC process.

Several ways to overcome this difficulty were proposed. One of the most frequently used methods is based on the transition to the principal axes in the space of arguments (see [16] for more detail). Unfortunately, this method does not reach its goals when the dimensionality (number of arguments) of the distribution is great, since this procedure implies estimation of eigenvectors of the covariance matrix of arguments, which *per se* requires large statistic sampling. If the sampling volume is insufficient, pronounced error in determining eigenvectors makes the transition to the principal axes inefficient.

Thus, direct application of the MCMC methods for reconstructing a DS in the case of high-dimensional sought posterior distribution (2) may be absolutely inefficient because of slow convergence to the equilibrium distribution. This very situation takes place in the reconstruction of a dynamical system from the noisy TS generated by it, which is of interest to us: Here, the arguments of the studied distribution are, first, model parameters and, second, unknown (due to the noise) "real" values of phase variables. In the case of not too short TS, the phase variables constitute the major portion of the distribution arguments.

In this paper we propose a method which makes it possible to construct a representative sampling for some part of the whole set of arguments of the initial probability distribution. The concept of the proposed method is approximate integration, by means of the Laplace approximation, of the initial multidimensional distribution with respect to "unnecessary" arguments.

III. SOLUTION METHOD

In the context of the problem in question, i.e., reconstruction of a DS from a noisy TS, when we need representative sampling of model parameters and the search for real ("latent") quantities of the measured characteristic is not an issue, our proposal reduces to "preliminary" (before sampling) integration of the distribution $P(\boldsymbol{\mu}, \mathbf{U}|\mathbf{x})$ over latent variables $\mathbf{U}$. There exists a whole family of techniques that makes such an integration feasible for a stochastic system of the general form (1) (see, e.g., [8,17]). Unfortunately, these methods are inapplicable in the case considered when the system is purely dynamical (deterministic) or, which is the same, at zero $\boldsymbol{\eta}_t$.

Let us suppose the distribution $P(\boldsymbol{\mu}, \mathbf{U})$ is not informative for latent variables $\mathbf{U}$, i.e., $P(\boldsymbol{\mu}, \mathbf{U}) = P(\boldsymbol{\mu})$. For the Gaussian noise $\boldsymbol{\xi}_t \overset{iid}{\sim} N(0, \varepsilon^2)$, the expression for the conditional PD describing probability of generation by the model $\mathbf{f}(\mathbf{U}, \boldsymbol{\mu})$ of the measured data $P(\boldsymbol{\mu}|\mathbf{x}) = \int P(\boldsymbol{\mu}, \mathbf{U}|\mathbf{x}) d\mathbf{U}$ has the following form [see (3)]:

$$P(\boldsymbol{\mu}|\mathbf{x}) \propto P(\boldsymbol{\mu}) \prod_{t=0}^{M-1} \int_{-\infty}^{+\infty} \prod_{j=0}^{w} \exp[-|\mathbf{x}_{t(w+1)+j+1}$$

$$- \mathbf{f}^j(\mathbf{U}_{t(w+1)+1}, \boldsymbol{\mu})|^2/2\varepsilon^2] d\mathbf{U}_{t(w+1)+1}. \qquad (4)$$

It is evident that, since the integrand in (4) has a pronounced

maximum with respect to the variable $\mathbf{U}_t$, it is convenient to use the Laplace method for the integration. Let the function (mismatch)

$$\chi_t^2(\mathbf{x}, \mathbf{U}, \boldsymbol{\mu}) = \sum_{j=0}^{w} [\mathbf{x}_{t+j} - \mathbf{f}^j(\mathbf{U}_t, \boldsymbol{\mu})]^2 \qquad (5)$$

reach its minimum with respect to the latent variable at $\mathbf{U}_t = \overline{\mathbf{U}}_t$. Let us further transform formula (5) in the vicinity of the minimum using the Taylor expansion. Taking into account vanishing of the linear expansion term we obtain

$$\chi_t^2(\mathbf{x}, \mathbf{U}, \boldsymbol{\mu}) \approx \bar{\chi}_t^2 + \frac{1}{2}(\mathbf{U} - \overline{\mathbf{U}}_t)^T \cdot \mathbf{D}_t \cdot (\mathbf{U} - \overline{\mathbf{U}}_t), \qquad (6)$$

where $\bar{\chi}_t^2 = \chi_t^2(\mathbf{x}, \overline{\mathbf{U}}_t, \boldsymbol{\mu})$, $\mathbf{D}_t$ is the matrix of second derivatives of the mismatch (5) with respect to the latent variable, and

$$D_t^{ik} = 2\sum_j \frac{\partial \mathbf{f}^j}{\partial U_t^i} \frac{\partial \mathbf{f}^j}{\partial U_t^k} - 2\sum_j \bar{\chi}_t \frac{\partial^2 \mathbf{f}^j}{\partial U_t^i \, \partial U_t^k}. \qquad (7)$$

Formula (7) becomes simpler if one takes into account the smallness of $\bar{\chi}_t$ and neglects the term which contains it [18],

$$D_t^{ik} = 2\sum_j \frac{\partial \mathbf{f}^j}{\partial U_t^i} \frac{\partial \mathbf{f}^j}{\partial U_t^k}. \qquad (8)$$

Finally, the integral (4) will be reduced to the integration of multidimensional Gaussian distribution,

$$\begin{aligned} P(\boldsymbol{\mu}|\mathbf{x}) &\propto \prod_{t=0}^{M-1} \left[\exp\left(-\frac{\bar{\chi}_t^2}{2\varepsilon^2}\right) \right. \\ &\quad \left. \times \int_{-\infty}^{+\infty} \exp[-(\mathbf{U} - \overline{\mathbf{U}}_t)^T \cdot \mathbf{D}_t \cdot (\mathbf{U} - \overline{\mathbf{U}}_t) d\mathbf{U}_t] \right] \\ &\propto \prod_{t=0}^{M-1} \left[\exp\left(-\frac{\bar{\chi}_t^2}{2\varepsilon^2}\right) \frac{1}{\sqrt{\det \mathbf{D}_t}} \right]. \end{aligned} \qquad (9)$$

The obtained PD depends on the number of variables which is scores of times less than the initial PD (3).

IV. RESULTS

Apparently, the distribution of parameters in various EO models cannot be compared directly from the viewpoint of their adequacy to the EO of the system, when the TS generated by the system is rather noisy. Aiming now at testing the proposed "integrated" distribution of the EO model parameters (9) we compare this distribution with a "fully dimensional" one described by the expression (3). Both of these distributions were constructed using "perfect" TS generated by a computer and corrupted by noise afterward.

As a measure of the correspondence between the model EO to the EO of the modeled system we use the function $\nu(\mathbf{Y}, \boldsymbol{\mu})$,
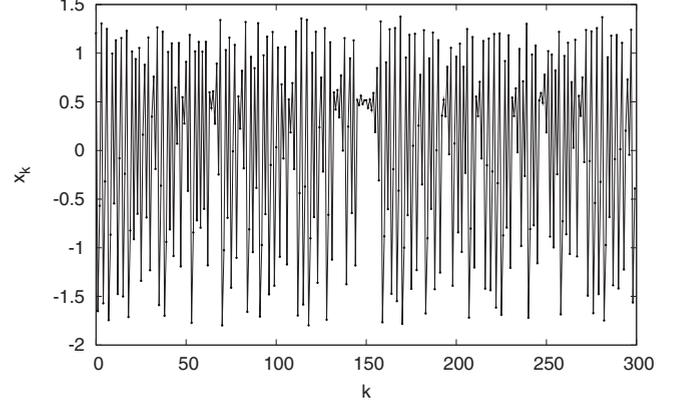
FIG. 1. Normalized chaotic time series (logistic map $x_{k+1} = 1 - hx_k^2, h = 1.85$) corrupted by additive white noise with $\varepsilon = 0.05$.

$$\nu(\mathbf{Y}, \boldsymbol{\mu}) = \frac{\sum_{t=1}^{T-1} [\mathbf{Y}_{t+1} - \mathbf{f}(\mathbf{Y}_t, \boldsymbol{\mu})]^2}{T - 1}, \qquad (10)$$

where $\mathbf{Y}$ is the true state of the system, and $\boldsymbol{\mu}$ is the vector of parameters of the model constructed by the noisy data. Function (10) [which is, evidently, the "defect measure" for the model $\mathbf{f}(\mathbf{U}, \boldsymbol{\mu})$] allows direct comparison between the models including various number of variables from the viewpoint of their best match with the underlying system [19].

As was mentioned above, an important factor restricting application of the Bayesian models for reconstructing DSs from chaotic TS of arbitrary length is slow convergence of the MCMC procedure of generation of an ensemble of parameters of the EO model $\{\boldsymbol{\mu}_n\}_{n=1}^N$ and latent variables $\{\mathbf{U}^n\}_{n=1}^N$ corresponding to the posterior distribution (3). Integration with respect to latent variables leading to the PD (9) reduces this constraint substantially thanks to multiple reduction of the dimension of the sought distribution. Clearly, convergence of the procedure of generating the ensemble $\{\boldsymbol{\mu}_n\}_{n=1}^N$ means convergence of the corresponding ensemble of measure $\{\nu(\mathbf{Y}, \boldsymbol{\mu}_n)\}_{n=1}^N$ [see (10)] that depends on these parameters. The rate of convergence of measure (10) was estimated by approximating the ensemble $\{\nu(\mathbf{Y}, \boldsymbol{\mu}_n)\}_{n=1}^N$ by the function

$$g(n) = \alpha + \beta e^{(-n/\tau)}, \qquad (11)$$

which is characterized by the time scale τ.

The testing was performed using a chaotic TS, 300-points long, generated by the logistic map $x_{k+1} = 1 - hx_k^2$, corrupted by additive noise $\boldsymbol{\xi}_t \overset{iid}{\sim} N(0, \varepsilon^2)$. Figure 1 shows an example of the chaotic TS generated by this system for the parameter $h = 1.85$ corrupted by additive white noise with $\varepsilon = 0.05$.

To simulate the situation (unknown DS and high-quality model) described by Eqs. (3), (4), and (9) we used a simple artificial neural network (ANN) as a model of the evolution operator. The choice of ANN as a model was determined by the existence of the theorem which states that any continuous function of an arbitrary number of variables can be approxi-
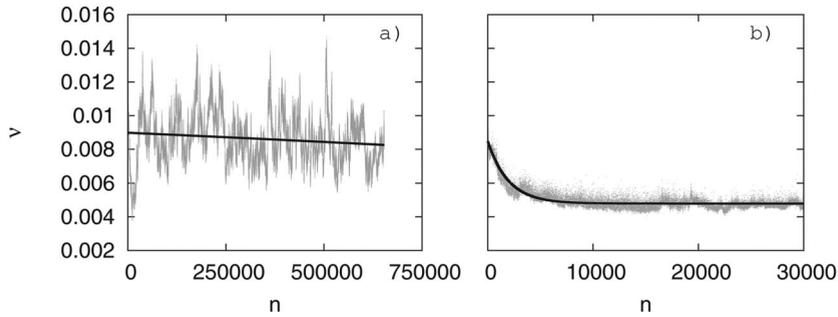
FIG. 2. Dependence of $\nu(\mathbf{Y}, \boldsymbol{\mu}_n)$ on the number n of iterations of the sampling procedure (gray curve) and function (11) which approximates it (black curve) for the "fully dimensional" (a) and "integrated" (b) PD; $w=4$.

mated by an ANN having a sufficient number of neurons to any preset accuracy [20]. In the current paper we used a three-layer perceptron set by the function

$$\mathbf{f}_m(\mathbf{x}, \boldsymbol{\mu}) = \left[\sum_{i=1}^{m} a_{ki} \tanh\left(\sum_{j=1}^{d} b_{ij} x_j + c_i \right) \right]_{k=1}^{d}, \quad (12)$$

where m is the number of neurons in the hidden layer, and $\boldsymbol{\mu} = (\mathbf{a}, \mathbf{b}, \mathbf{c})$ are the parameters. The initial training of the network was performed using the supervised learning method which is typical for the problems of data approximation (see [21] for more detail). The model (12) used to reconstruct the EO included 15 parameters ($m=5, d=1$).

For the EO model in the form of a three-layer perceptron, we used prior distribution $P(\boldsymbol{\mu})$ of the form

$$P(\boldsymbol{\mu}) = \exp\left[-\sum_{i=1}^{m} \left(\sum_{k=1}^{d} \frac{a_{ki}^2}{2\sigma_a^2} + \sum_{j=1}^{d} \frac{b_{ij}^2}{2\sigma_b^2} + \frac{c_i^2}{2\sigma_c^2} \right) \right], \quad (13)$$

where σ_a^2, σ_b^2, and σ_c^2 are dispersions of the corresponding parameters the magnitude of which determines rigidity of prior restrictions imposed on the solution of the inverse problem of reconstruction. A detailed discussion of the choice of values of these parameters can be found, for instance, in [22].

The estimated convergence rate of the measure (10) shows that for the "integrated" PD (9) and various values of the number of EO reiterations w, the characteristic times τ ranged from 2000 to 3500, whereas in the case of the "fully dimensional" PD (3), the value of τ ranged from 10^6 to 10^8 [23].

An example illustrating this difference in the convergence rate is shown in Fig. 2. It is seen that the process of sampling PD (9) can be regarded to be fully converged after about 7000 iterations [see Fig. 2(b)], whereas the sampling PD (3) does not converge, even if the number of iterations is 2 orders of magnitude more [see Fig. 2(a)]. Thus, one can conclude that convergence takes place when the "integrated" PD (9) is used and does not occur in the case of the "fully dimensional" PD (3) [24].

Figure 3 shows the measure (10) as a function of w for PDs (3) and (9) in the case of a fixed level of measurement noise. They are plotted for an equal (sufficiently large) number of iterations used to construct the corresponding distributions of parameters.

It is clearly seen that the model with the parameters distributed in accordance with PD (9) demonstrates better correspondence with the initial system (smaller average defect

and dispersion) for all values of w and for all considered noise levels.

It is also seen from the presented plots that there is an optimum along w in both of the cases. The cause of decreasing measure (10) as w grows is common for both "fully dimensional" and integrated PD. The model is improving when it is getting closer to the form which is "perfect" in the framework of the Bayesian approach [when the segment length ($w+1$) is equal to the total length of the TS]. In other words, the greater w, the better the model $\mathbf{f}(\mathbf{U}, \boldsymbol{\mu})$ captures dynamical properties of the modeled system.

The increase of measure (10) in the case of an "excessive" growth of w is also easy to explain. For integrated PD finding of the global maximum of $P(\boldsymbol{\mu}, \mathbf{U}|\mathbf{x})$ [or, which is the same for normal distribution of measurement noise, of the global minimum of mismatch (5)] eventually becomes im-
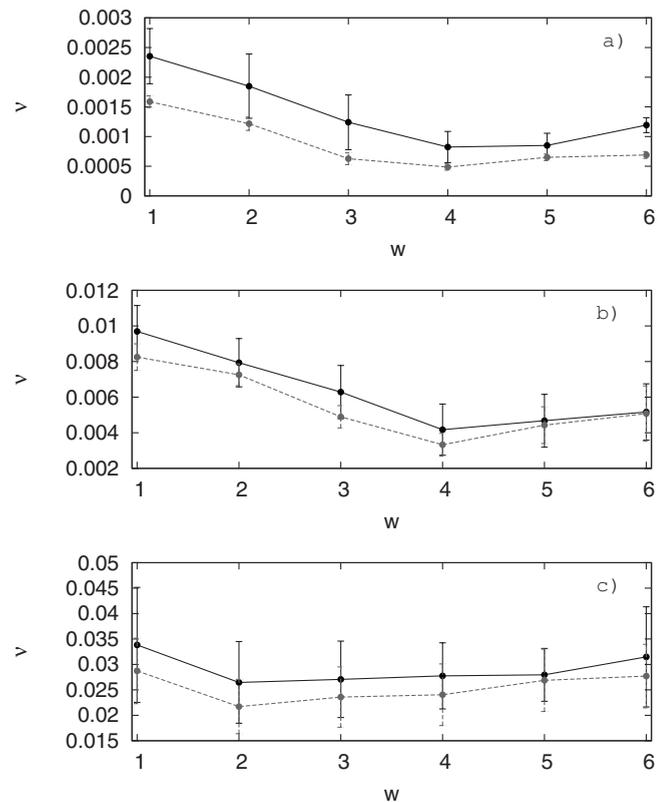
FIG. 3. Measure (10) versus w: Average and standard deviations of the model's "defect measure" for the "fully dimensional" PD (solid curve) and the PD integrated over latent variables (dashed curve). Noise level is $\varepsilon=0.05$ (a); $\varepsilon=0.1$ (b); $\varepsilon=0.2$ (c).
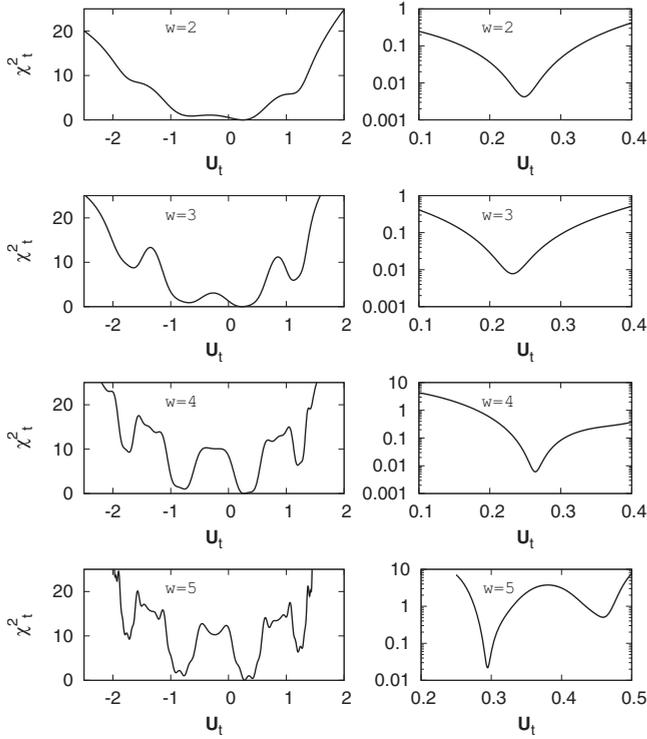
FIG. 4. Dependence of $\chi_t^2(\mathbf{U}_t)|_{\mathbf{x},\boldsymbol{\mu}}$ on $\mathbf{U}_t$ for w from 2 to 5 in normal (left-hand column) and logarithmic (right-hand column) scales.

possible because of the exponentially fast diverging of the initially close phase trajectories. As a result of the divergence an increase in w leads to decreasing characteristic size of the regions of values of the latent variables and of model parameters that ensure passage of the model phase trajectory in the noise-specified neighborhood of the trajectory reconstructed by the noisy TS.

Accordingly, the mismatch (5) as a function of its arguments takes on a multimodal ("jagged") form (see Fig. 4). Hence, for excessively large w, the PD (9) leads to a model that is less correct than at smaller w. In the case of the "fully dimensional" PD, redundant increase of w makes correct sampling of the distribution (3) impossible because of its too high dimensionality. As a consequence, in both of the cases, the measure (10) starts to increase when a certain optimal value is exceeded by w.

To illustrate this statement in relation to integrated PD, we will consider the simplest case when the latent variable $\mathbf{U}_t$ is unique. In other words, we will consider the function $\mathbf{P}(\boldsymbol{\mu}, \mathbf{U}_t|x_t)$ corresponding to one segment of (4) and compare the integrals

$$I(\mu_k) = \int P\left(\boldsymbol{\mu}, \mathbf{U}_t|\mathbf{x}_t\right)|_{\boldsymbol{\mu}^*} d\mathbf{U}_t \qquad (14)$$

of this function for fixed values of the other parameters $\boldsymbol{\mu}^* = (\mu_1, \ldots, \mu_{k-1}, \mu_{k+1}, \ldots, \mu_n)$, obtained by precise numerical integration and using the approximate integration by the Laplace method (Fig. 5). It is seen from the plots in Fig. 5 that, when w exceeds a certain value $w_{\max}$, the distribution
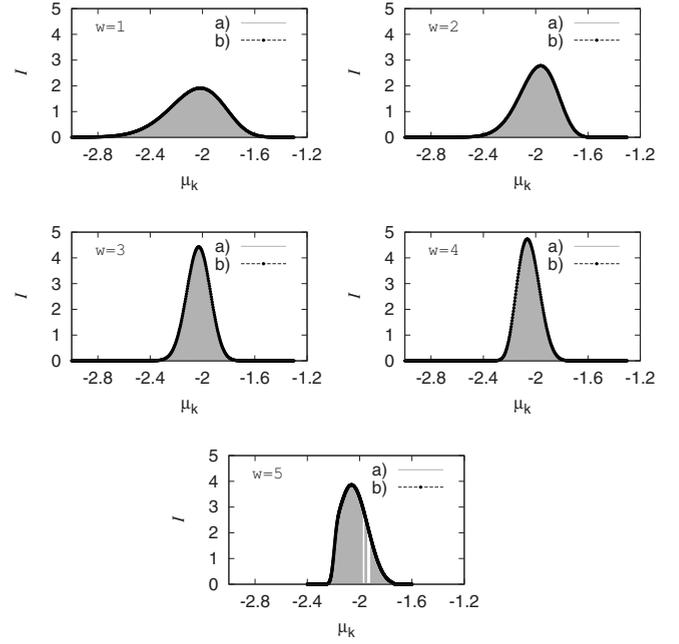
FIG. 5. $I(\boldsymbol{\mu}_k)$ [see (14)] as a function of w calculated by the Laplace method (gray) and precisely (black). Noise level is $\varepsilon = 0.05$.

integrated by the method differs significantly from that calculated precisely (by the method of trapezoids with appropriate step). This difference is the consequence of the error in searching the global minimum in the profile of the mismatch $\chi_t^2(\mathbf{U}_t)|_{\mathbf{x},\boldsymbol{\mu}}$ when using the Laplace integration method.

Based on the above, one can estimate the value of $w_{\max}$ analytically. With increasing w (i.e., with increasing number of EO reiterations), the number of the minima of function $\chi_t^2(\mathbf{U}_t)|_{\mathbf{x},\boldsymbol{\mu}}$ grows approximately by the law $e^{\lambda w}$ (here λ is the largest Lyapunov exponent for the chaotic attractor under consideration). Correspondingly, the distance l between adjacent minima can be estimated by $e^{-\lambda w}$. If the uncertainty of setting initial conditions in terms of latent variables, which is determined by the noise level ε, exceeds l, the probability of entrapment of the initial values of latent variables into the attraction domain of the local (and not global) minimum increases strongly. Thus, the maximum value of w can be estimated from the following relationship:

$$e^{-\lambda w_{\max}} \geq \varepsilon. \qquad (15)$$

The calculation of the largest Lyapunov exponent λ using the TISEAN software [25] shows that, for the logistic map $x_{k+1} = 1 - hx_k^2$ at $h = 1.85$, we have $\lambda \approx 0.5$. From Eq. (15) for the noise level $\varepsilon = 0.05$, we estimate the maximum number of EO reiterations to be $w_{\max} \lesssim 5$, which corresponds to the results of the calculations shown in Fig. 3.

To conclude this section we note that the distribution function (4) constructed assuming normal noise distribution may be used for noises with a different unimodal distribution with limited dispersion. Consider by way of example the case when δ-correlated noise entering data of measuring $\mathbf{x}$ has uniform distribution. The function of posterior PD inte-

grated over latent variables that is correct for this case has the following form:

$$P(\boldsymbol{\mu}|\mathbf{x}) \propto \prod_{t=0}^{M-1} \int_{-\infty}^{+\infty} \prod_{j=0}^{w} H[\mathbf{x}_{t(w+1)+j+1}$$

$$- \mathbf{f}^j(\mathbf{U}_{t(w+1)+1}, \boldsymbol{\mu})]d\mathbf{U}_{t(w+1)+1},$$

$$H(y) = \begin{cases} 1/(b-a), & a \le y \le b, \\ 0, & \text{else,} \end{cases} \qquad (16)$$

where a and b are boundaries of the uniform distribution. Comparison of the dependence of the integrals (14) determined by the tth factor of the function (16) and by the corresponding factor of the function (4) ("incorrect" for uniformly distributed noise) on one of the model parameters is given in Fig. 6 for different segment lengths w. These cofactors were calculated by means of exact numerical integration over the unique latent variable.

One can see good agreement between the plotted dependences. The confidence intervals of the parameter corresponding to root-mean-square deviations overlap by more than 90% for all the used w.

V. CONCLUSION

The paper proposes an approach that makes it possible to apply MCMC algorithms for solution of inverse problems in the framework of the Bayesian approach when a great number of variables of posterior probability density do not allow their direct application. This approach can be used when the dependence of the probability density on part of the variables allows integration over those variables using the Laplace method (in other words, when the probability density is "quasi-Gaussian"). In many cases, it allows a significant reduction of the number of variables used in subsequent sampling. As shown in the paper, one of such applications is the problem of reconstruction of dynamical systems from noisy chaotic time series: In this situation, the dependence of the probability density on latent variables (noiseless states of the system) is close, in the aforesaid sense, to the Gaussian one. Efficiency of this approach is demonstrated on an example of such a system, and the limits for its applicability are determined. Specifically, it is shown that as the length of the se-
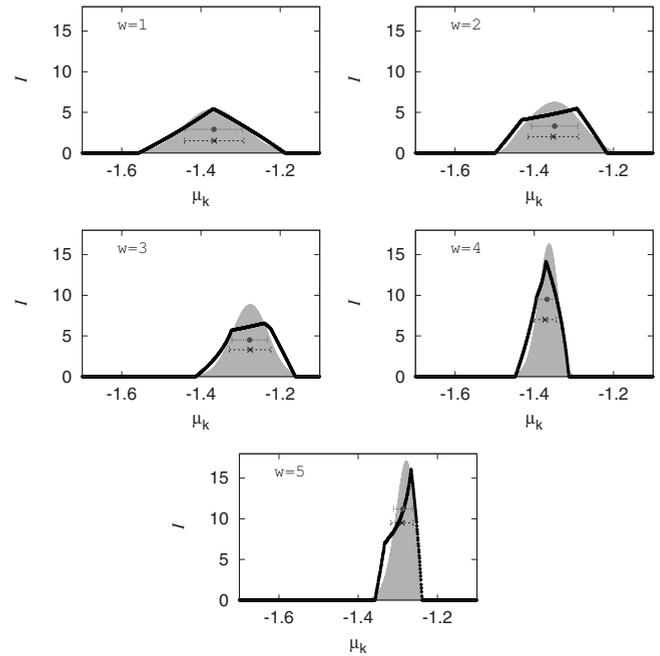
FIG. 6. Dependence of the integral (14) determined by the tth factor of the function (16) on one of the parameters of the model is shown by the black curves; analogous dependences determined by the function (4) are shown with gray shading. Also given are mean values (crosses for the black curves and closed circles for gray shading) and confidence intervals corresponding to root-mean-square deviations of the plotted dependences. Noise level is $\varepsilon = 0.05$.

ries grows, direct application of MCMC becomes impossible very rapidly, whereas the proposed approach makes it possible to exclude the dependence of the rate of convergence of the corresponding iteration procedure on the TS length.

[1] R. Meyer and N. Christensen, Phys. Rev. E **62**, 3535 (2000).

[2] K. Judd, Phys. Rev. E **67**, 026212 (2003).

[3] V. S. Anishchenko and A. N. Pavlov, Phys. Rev. E **57**, 2455 (1998).

[4] P. E. McSharry and L. A. Smith, Phys. Rev. Lett. **83**, 4285 (1999).

[5] V. F. Pisarenko and D. Sornette, Phys. Rev. E **69**, 036122 (2004).

[6] J. Timmer, H. Rust, W. Horbelt, and H. U. Voss, Phys. Lett. A **274**, 123 (2000).

[7] C. D. Rodgers, *Inverse Methods for Atmospheric Sounding* (World Scientific, Singapore, 2000).

[8] A. Sitz, U. Schwarz, J. Kurths, and H. U. Voss, Phys. Rev. E **66**, 016210 (2002).

[9] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice* (Chapman and Hall, London, 1996).

[10] D. N. Mukhin, A. M. Feigin, E. M. Loskutov, and Y. I. Molkov, Phys. Rev. E **73**, 036211 (2006).

[11] H. D. I. Abarbanel, *Analysis of Observed Chaotic Data* (Springer-Verlag, New York, 1997).

[12] The phase space can be reconstructed from scalar TS using the delay-coordinate method [11,26].

[13] Y. I. Molkov and A. M. Feigin, *Nonlinear Waves* (Institute of

Applied Physics, Nizhny Novgorod, 2002), pp. 34–52 (in Russian).

[14] O. Y. Butkovsky, Y. A. Kravtsov, and M. Y. Logunov, Radiophys. Quantum Electron. **45**, 48 (2002).

[15] Note that first, when deriving (3) it was assumed that the "measurement noise" ξ_t is δ correlated and, second, if the noisy chaotic TS is generated by a known DS, the model is "good" *a forteriori* in the mentioned sense and (3) is also applicable in this case (for more detail, see [10]).

[16] S. E. Dosso, J. Acoust. Soc. Am. **111**, 129 (2002).

[17] R. Meyer and N. Christensen, Phys. Rev. E **65**, 016206 (2001).

[18] W. H. Press, B. P. Flannery, and S. A. Teukolsky, *Numerical Recipes. The Art of Scientific Computing* (Cambridge University Press, Cambridge, 1990).

[19] Note that we use function (10) only to test the proposed model. Naturally, its application to real data is impossible for lack of information about true (noiseless) values of the observed variable.

[20] K. Hornik, M. Stinchcombe, and H. White, Neural Netw. **2**, 359 (1989).

[21] *The Handbook of Brain Theory and Neural Networks*, edited by M. A. Arbib (MIT Press, Cambridge, MA, 1995).

[22] R. M. Neal, University of Toronto, Technical Report No. CRG-TR-93-1, 1993 (unpublished).

[23] We employed the adaptive Metropolis algorithm with normal proposal distribution and continuously adapting first and second moments [27,28].

[24] Note that the MCMC literature contains various diagnostics to assess the convergence of MCMC algorithms (see [29,30]). However, the use of measure (10) for assessing convergence of MCMC when solving the problem of DS reconstruction from TS has two important advantages. First, measure (10) permits one to estimate the characteristic time of algorithm convergence, even in the case when the procedure has insufficient time to converge during a reasonable number of iterations, as is the case of sampling of the ensemble corresponding to the distribution (3). Second, measure (10) shows (see Fig. 3) that MCMC sampling of the integrated posterior distribution (9) not only converges incomparably faster than at sampling (3), but also provides lower discrepancy measure, i.e., construction of a model that better corresponds to the modeled system.

[25] R. Hegger, H. Kantz, and T. Schreiber, Chaos **9**, 413 (1999).

[26] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, 1997).

[27] H. Haario, M. Laine, M. Lehtinen, E. Saksman, and J. Tamminen, J. R. Stat. Soc. Ser. B (Stat. Methodol.) **66**, 591 (2004).

[28] H. Haario, E. Saksman, and J. Tamminen, Bernoulli **7**, 223 (2001).

[29] L. Robert, M. Robert, and C. Guihenneuc-Jouyaux, citeseer.comp.nus.edu.sg/78250.html

[30] N. Best, M. Cowles, and K. Vines, CODA: Convergence diagnosis and output analysis software for gibbs sampling output, version 0.30, 1995, citeseer.ist.psu.edu/best97coda.html