

Prognosis of qualitative system behavior by noisy, nonstationary, chaotic time seriesY. I. Molkov,^{1,*} D. N. Mukhin,² E. M. Loskutov,² R. I. Timushev,² and A. M. Feigin²¹*Indiana University–Purdue University Indianapolis*²*Institute of Applied Physics, Russian Academy of Sciences*

(Received 30 June 2011; published 23 September 2011)

An approach to prognosis of qualitative behavior of an unknown dynamical system (DS) from weakly nonstationary chaotic time series (TS) containing significant measurement noise is proposed. The approach is based on construction of a global time-dependent parametrized model of discrete evolution operator (EO) capable of reproducing nonstationary dynamics of a reconstructed DS. A universal model in the form of artificial neural network (ANN) with certain prior limitations is used for the approximation of the EO in the reconstructed phase space. Probabilistic prognosis of the system behavior is performed using Monte Carlo Markov chain (MCMC) analysis of the posterior Bayesian distribution of the model parameters. The classification of qualitatively different regimes is supposed to be dictated by the application, i.e., it is assumed that some classifier function is predefined that maps a point of a model parameter space to a finite set of different behavior types. The ability of the approach to provide prognosis for times comparable to the observation time interval is demonstrated. Some restrictions as well as possible advances of the proposed approach are discussed.

DOI: [10.1103/PhysRevE.84.036215](https://doi.org/10.1103/PhysRevE.84.036215)

PACS number(s): 05.45.Tp, 02.30.Zz, 02.70.Rr

I. INTRODUCTION

The evolution of dynamical characteristics, i.e., nonstationarity of analyzed processes, is a typical phenomenon that manifests itself when one attempts to reconstruct a dynamical system (DS) from data measured in real conditions. Many available methods of DS reconstruction imply that these characteristics are constant (stationary), hence they do not hold for the nonstationary case. In the recent 10–15 years there appeared many papers in which ways of establishing nonstationarity as well as determining the character and measure of nonstationarity of processes of different origin are discussed. A number of methods were developed based on analysis of the distribution of time intervals between neighboring vectors in phase space [1–3], use of the measure of mutual predictability of dynamics between different sections of time series (TS) [4], investigation of the time dependence of probability density of the process and its spectral density [5], etc.

Although the problem of DS reconstruction is much more complicated for the nonstationary situation, nonstationary processes are the most interesting in terms of possible prognosis of qualitative behavior of the system because they carry information about trends of the characteristics that determine dynamical properties of the system. Such a prognosis is traditionally based on analysis of first principles models (FPM) that include a set of equations which, in the opinion of the researcher, correctly describe the observed phenomenon. In other words, the FPM approach presumes that the reconstructed DS is known to us. Clearly, qualitative behavior of the system in the future may be reconstructed only if the structure of the space of parameters of the model corresponds to that of the modeled system. However, bifurcations of the system leading to future changes in characteristics of the observed dynamics are frequently governed by processes that have not yet become apparent by the time of FPM construction and

are thus not taken into account. Consequently, the domain where the space of FPM parameters is adequate is strongly bounded and the use of such models for long-term prognosis is fraught with prognostic error. A typical example of erroneous prognosis using the FPM method is prediction of the ozone layer evolution in the mid 1980s [6] when the Antarctic ozone hole was omitted.

In the current paper we propose an alternative approach to DS reconstruction that is devoid of the above drawback. The approach does not presume any knowledge of the underlying equations of DS evolution and is based on the construction of a universal model of the system by noisy chaotic nonstationary TS. Nonstationary chaotic behavior was revealed in a wide range of real systems in the atmosphere [7] and in the ocean [8], controlling processes in living organisms [9], describing tectonic activity [10], and many others. Efficiency of the proposed approach for prognosis will be demonstrated on an example modeling the situation, most unfavorable in terms of FPM verification, when no bifurcations occur in the system during the observation time, whereas in the future substantial qualitative changes occur in the DS evolution.

II. FORMULATION OF THE PROBLEM

Let us consider a situation when, first, the characteristic time scale of nonstationarity is much more than the characteristic time of DS dynamics and TS duration (in what will follow we will call such nonstationarity weak). Second, we will assume that this nonstationarity is a consequence of slow changes (trends) of control parameters of the observed DS, or in other words, the observed DS is weakly nonautonomous. Such evolution of control parameters of DS may lead to both quantitative and qualitative changes in the observed processes. Of particular interest in this case is the problem of elucidating possible scenarios of system behavior beyond observation time or, in other words, prognosis of future evolution of qualitative behavior of the system.

*ymolkov@iupui.edu

A possibility of reconstructing the bifurcation scenario of unknown DS by *several* TS generated by the system at different *fixed* values of control parameter was demonstrated in the works [11,12]. Polynomial representation [11] and approximation on the basis of an artificial neural network [12] were used in these works as global models of DS evolution operator.

We used analogous parametrized models for developing the method of prognosis of qualitative behavior of an unknown system by weakly nonstationary *noiseless* chaotic TS [13]. We successfully predicted behavior of discrete and flow systems with slowly varying control parameters. However, this method based on least squares estimates is inapplicable in the presence of strong enough measurement noise because it results in systematic bias of estimates due to statistical inconsistency, which distorts the prognosis.

The goal of this paper is development of the method for constructing prognosis of qualitative behavior of a system by nonstationary noisy chaotic TS that would give unbiased probability estimates of possible events in the future. In particular, in this work the prognosis is aimed at identifying the type of dynamical regime of a system in the future. For such identification the method should give a set of probabilities of possible dynamical regimes for each moment of time. The classification of *qualitatively different* regimes is dictated by the application. We assume that there is a predefined classifier function that maps a regime (or a point in a parameter space of a model) to a finite set of different behavior types. The method proposed in this work is based on the Bayesian approach, which statistically correctly takes into account both data precision and evolution operator (EO) reconstruction error (defect of the model).

III. SOLUTION METHOD

Let the observed TS $\mathbf{X} = \{\mathbf{x}_k = \mathbf{x}(t_k)\}_{k=1}^T$ of duration T be generated by a chaotic dynamical system with time dependent EO $F_t : \mathbb{R}^d \rightarrow \mathbb{R}^d$. We will consider the parametrized model Q approximating unknown evolution operator of system F_t assuming that the observed dynamical variable \mathbf{x} may be described by the following system:

$$\begin{cases} \mathbf{x}_k = \mathbf{u}_k + \boldsymbol{\xi}_k \\ \mathbf{u}_k = Q(\mathbf{u}_{k-1}, \boldsymbol{\mu}, t_k) + \boldsymbol{\eta}_k \end{cases}, \quad (1)$$

where $\boldsymbol{\xi}$ is measurement noise, \mathbf{u} stands for *latent* variables corrupted by measurement noise (true states of the DS), $\boldsymbol{\mu}$ is the vector of model parameters, t_k denotes the moment of recording the value of \mathbf{x}_k , and $\boldsymbol{\eta}$ is the defect of the model (random differences between model and true EO).

Let us pass over to the problem of choosing the EO model functional form $Q(\cdot)$. For the algorithm to be suitable for construction of models of different dynamical systems (high-dimensional systems included) the function Q must allow approximation of any single-valued function of an arbitrary number of variables with any preset accuracy. This universality is inherent in artificial neural networks (ANNs) [14] that enable one to increase accuracy of approximation by merely increasing the number of neurons without changing the

form of the function. In the current work we use a three-layer perceptron [15] specified by the function

$$\mathbf{f}_m(\mathbf{y}, \mathbf{a}, \mathbf{b}, \mathbf{c}) = \left\{ \sum_{i=1}^m a_{ki} \tanh \left(\sum_{j=1}^d b_{ij} y_j + c_i \right) \right\}_{k=1}^d, \quad (2)$$

where d is the dimension of \mathbf{y} and m is the number of neurons.

Apparently, in case of nonstationary TS, for correct description of system dynamics it is necessary to construct a nonautonomous (explicitly depending on time) EO model. It is natural to conjecture that slow changes of control parameters of the system (a trend in the corresponding space of parameters) must be reflected by a slow dependence of parameters of model (1) on time. Bearing this slowness in mind, hereinafter we will use linear expansion of the model in time. It can be readily shown that in the case of approximation (2), the following form of the model is universal:

$$Q(\mathbf{u}, \boldsymbol{\mu}, t) = \mathbf{f}_m(\mathbf{u}, \mathbf{a}_1 + t\mathbf{a}_2, \mathbf{b}, \mathbf{c}), \quad \boldsymbol{\mu} = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{b}, \mathbf{c}). \quad (3)$$

Linear approximation of time dependence of the model will undoubtedly lead to strong discrepancy between prognosis and reality when the model is extrapolated to sufficiently remote times, but it can be expected that qualitative behavior of the model and of the system will be similar in the not too remote future.

In any case, we believe that trends of parameters are slow to an extent that the defect of the model $\boldsymbol{\eta}$ connected with linearity of their approximation is much smaller than measurement noise. Thus we will consider the error of such approximation of nonautonomy to be insignificant within observation time, and the error due to linear *extrapolation* of the model to the future will be taken into account within the framework of the proposed Bayesian procedure (see the text below). As one can infinitely expand the class of functions approximated by the dependence (2) by increasing the number of neurons m , the latter may always be chosen such that the magnitudes of model defect $\boldsymbol{\eta}$ should be much less than the level of measurement noise $\boldsymbol{\xi}$. From these considerations we will further suppose that $\boldsymbol{\eta} = 0$ in the second equation of Eq. (1).

Let us construct a probability density (PD) function for unknown values in system (1) using the Bayesian approach and known form of the joint probability density function of noise $w_\xi(\boldsymbol{\xi}, \nu)$ (ν are distribution parameters that are unknown in a general case) under the condition of the observed process ($\mathbf{X} = \{\mathbf{x}_k\}_{k=1}^T, \mathbf{t} = \{t_k\}_{k=1}^T$). According to the Bayes theorem, the sought PD function is expressed through the data likelihood function $p(\mathbf{X}, \mathbf{t} | \boldsymbol{\mu}, \nu, \mathbf{U})$:

$$p(\boldsymbol{\mu}, \nu, \mathbf{U} | \mathbf{X}, \mathbf{t}) \propto p(\mathbf{X}, \mathbf{t} | \boldsymbol{\mu}, \nu, \mathbf{U}) p_{pr}(\boldsymbol{\mu}, \nu, \mathbf{U}), \quad (4)$$

where $\mathbf{U} = \{\mathbf{u}_k\}$, and $p_{pr}(\boldsymbol{\mu}, \nu, \mathbf{U})$ is *prior* PD of unobserved quantities that accounts for additional prior information about the reconstructed system and the properties of measuring equipment.

Assuming the noise records $\boldsymbol{\xi}_k$ to be uncorrelated at different moments of time, normally distributed (with unknown

dispersion σ_ξ^2) and centered, we can write the likelihood function for the observable series (\mathbf{X}, \mathbf{t}) as a function of unobservable variables $\boldsymbol{\mu}, \sigma_\xi, \mathbf{U}$. Below we will use the likelihood function corresponding to modified Bayesian approach [18]:

$$p(\mathbf{X}, \mathbf{t} | \mathbf{U}, \boldsymbol{\mu}, \sigma_\xi) \propto \frac{1}{\sigma_\xi^{Td}} \exp \left(-\frac{1}{2\sigma_\xi^2} \sum_{k=0}^{M-1} \sum_{j=0}^w |\mathbf{x}_{k(w+1)+j+1} - Q^j(\mathbf{u}_{k(w+1)+1}, \boldsymbol{\mu}, t_{k(w+1)+1})|^2 \right), \quad (5)$$

where $Q^k(\mathbf{u}_i, \boldsymbol{\mu}, t_i) = Q(Q^{k-1}(\mathbf{u}_i, \boldsymbol{\mu}, t_i), \boldsymbol{\mu}, t_{i+k})$, $Q^0(\mathbf{u}_i, \boldsymbol{\mu}, t_i) \equiv \mathbf{u}_i$ and $M = T/(w + 1)$. Since this function takes into account deterministic coupling between latent variables only within time segments of the duration limited by w , it is fit for analysis even in the case of chaotic TS of long enough duration.¹ Samples of a chaotic TS separated in time by more than w are considered *a priori* independent, thus PDF (5) accounts for prior information about chaoticity of a signal. Parameter w is the technical parameter of the method. The larger w , the more deterministic couplings are taken into account, and, consequently, more information about the underlying system is contained in PDF (5). On the other hand, in the case of a chaotic TS, the number of local maxima of PDF (5) as a function of latent variables grows exponentially with w and, at some value w^* , these maxima become indistinguishable; further increasing of w makes no sense. It is easy to show (please see [18,19] for more detail) that the upper bound of w can be estimated by the formula $w^* = \lambda^{-1} \ln(\sigma_x/\sigma_\xi)$, where σ_x is the standard deviation of the observed TS and λ is the largest Lyapunov exponent of chaotic attractor that in some cases can be estimated directly from the observed TS [20]. In practice it is not necessary. As we have shown in [18], too long segments result in multimodal likelihood. This multimodality leads, first, to more frequent failures during MCMC (Markov chain Monte Carlo) sampling, which are easy to diagnose and, second, the confidence area in a parameter space stops shrinking with an increase of w . Basically, one can just sequentially increase w using the parameter estimates from previous iteration until the results stop improving, i.e., the probabilities of different regimes no longer get more “contrast.”

One of the problems concerned with Eq. (2) is the existence of singular directions in its space of parameters. This degeneration makes numerical study of the function (5) much more complicated. To avoid this difficulty we will use the method of setting a *prior* ensemble for $\boldsymbol{\mu}$ that was described in detail in [21]. Following this method we will introduce *prior* PDF $P_{pr}(\boldsymbol{\mu})$ that enters Eq. (4) in a Gaussian form with zero means:

$$P(\boldsymbol{\mu}) \propto \exp \left[-\sum_{i=1}^m \left(\sum_{k=1}^d \frac{a_{ki}^2}{2\sigma_a^2} + \sum_{j=1}^d \frac{b_{ij}^2}{2\sigma_b^2} + \frac{c_i^2}{2\sigma_c^2} \right) \right]. \quad (6)$$

The choice of dispersions σ_a^2, σ_b^2 , and σ_c^2 of model parameters $\boldsymbol{\mu}$ is determined by *a priori* ideas about statistical properties of dynamic processes generated by the system. Let us use normalization of the observed variable such that $\langle \mathbf{x} \rangle_t = 0$, $\langle \mathbf{x}^2 \rangle_t = 1$ and normalization of model time t setting the time interval of observation to be $[-1; 1]$. It is easy to show that output layer parameters' dispersion should be $\sigma_a^2 = \frac{1}{m}$ to provide generation by the model of a TS with $\langle \mathbf{x}^2 \rangle_t = 1$. The maximum magnitude of the derivatives $\partial Q_i / \partial u_j$ that determines σ_b can be estimated as $e^{\lambda\tau}$, where λ is the largest Lyapunov exponent and τ is a time lag used for model construction.² Note that the value of the largest Lyapunov exponent can be estimated by well-known algorithms (see, for example, [20]) directly from observed TS. Finally, we have $\sigma_c^2 = d\sigma_b^2$.

Another circumstance complicating analysis of the constructed PDF (4) is high dimension of the space of its arguments stipulated by the dependence of the likelihood function (5) on latent variables, the number M of which is proportional to the duration of the studied TS. With allowance for significant non-Gaussian PDF (4), this factor inhibits the use of MCMC techniques, even if they were specially designed for analysis of high-dimensional PDFs. In order to overcome this difficulty we use the method [19] of substantial reduction of the number of parameters of the studied distribution by approximate integration of PDF (5) with respect to latent variables using the Laplace method.

The obtained *posterior* PDF,

$$p_{ps}(\boldsymbol{\mu}, \sigma_\xi^2 | \mathbf{X}, \mathbf{t}) = \int p_{ps}(\boldsymbol{\mu}, \mathbf{U}, \sigma_\xi^2 | \mathbf{X}, \mathbf{t}) d\mathbf{U}, \quad (7)$$

contains all the information about the system extracted from the initial TS. As each vector $\boldsymbol{\mu}$ corresponds to the function of the EO of the system, the ensemble of parameters distributed in conformity with the function (7) specifies a set of possible realizations of dynamical variables of the system (*outside the scope* of the observations). Random parameter spread specified by such a distribution reflects the indefiniteness introduced by measurement noise $\boldsymbol{\xi}$ that is a source of error in EO reconstruction and, as a consequence, of the forecast of future evolution of the system. In addition, the ensemble of σ_ξ^2 values corresponding to Eq. (7) allows one to assess measurement error dispersion that is unknown *a priori*.

As was mentioned above, the source of another significant error of prognosis is the error associated with linear extrapolation of the trend of model parameters unaccounted for in Eq. (7). Assuming that the system of interest is weakly nonautonomous we can assess extrapolation error as an error of the linear trend of parameters that is quadratic with respect to time. Estimation of this correction within the framework of the Bayesian approach must be based on explicit allowance for the quadratic term in the trend of EO parameters:

$$Q(\mathbf{u}, \boldsymbol{\mu}, t) = \mathbf{f}_m(\mathbf{u}, \mathbf{a}_1 + t\mathbf{a}_2 + t^2\boldsymbol{\gamma}, \mathbf{b}, \mathbf{c}), \quad \boldsymbol{\mu} = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{b}, \mathbf{c}, \boldsymbol{\gamma}). \quad (8)$$

¹Note that there arise difficulties in using the classical (unmodified) Bayesian approach in the case of chaotic TS (please see [16,17]).

²For instance, if a Poincaré section is used for reconstruction of a map in the form (1), the time lag τ may be taken to be 1.

Let us interpret γ as a random vector with zero mean. Then the problem of assessing extrapolation error reduces to estimation of the covariance matrix \mathbf{C}_γ of quantities γ that may be obtained within the framework of the general approach described above. Here we will restrict ourselves to the *upper* estimate for which we can use an ensemble of parameters distributed according to Eq. (7). For this we assume that the rms error $\langle \delta \mathbf{a}^T(t) \delta \mathbf{a}(t) \rangle_{a,t}$ (averaged over the obtained ensemble and time) in the estimate of the trend $\mathbf{a}(t) = \mathbf{a}_1 + t \mathbf{a}_2$ is due only to the quadratic correction or, in other words, $\langle \delta \mathbf{a}^T(t) \delta \mathbf{a}(t) \rangle_{a,t} = \mathbf{C}_\gamma \langle t^4 \rangle$. If the quadratic correction is not taken into account, this quantity may be estimated to be³

$$\langle (\delta \mathbf{a}_1 + t \delta \mathbf{a}_2)^T (\delta \mathbf{a}_1 + t \delta \mathbf{a}_2) \rangle_{a,t} = \mathbf{C}_1 + \mathbf{C}_2 \langle t^2 \rangle, \quad (9)$$

where the matrices $\mathbf{C}_k = \langle \delta \mathbf{a}_k^T \delta \mathbf{a}_k \rangle$ may be calculated by the ensemble of parameters generated by the distribution (7). The resulting estimate of the covariance matrix of extrapolating parameter \mathbf{a} is

$$\mathbf{C}_\gamma = \frac{\mathbf{C}_1 + \mathbf{C}_2 \langle t^2 \rangle}{\langle t^4 \rangle}. \quad (10)$$

Finally, let us enumerate the main steps which we recommend for practical realization of the approach. Let us have TS $\{y(t_k)\}$ generated by a flow system. We suggest to analyze it using the following procedure:

(i) Reconstructing phase variables of the system from this TS. Since the choice of variables must provide embedding of the attractor underlying the observed behavior, we use the Takens delay method, i.e., we construct state vector $\mathbf{x}(t_k) = \{y(t_k), y(t_{k+\ell}), \dots, y(t_{k+(d-1)\ell})\}$, where ℓ is a value of delay (usually it is the first minimum of mutual information function of the TS) and d is embedding dimension used for representation of an attractor.

(ii) Obtaining a series $\{\mathbf{x}_j\}$ appropriate for construction of a discrete model in the form (1). For this we select the Poincaré section in the reconstructed phase space and register its intersections with the phase trajectory. Then, we obtain a sequence of vectors $\mathbf{x}_j \in \mathbb{R}^{d-1}$, where j is the number of the intersection. Note that, in a general case, the obtained TS $\{\mathbf{x}_j\}$ includes an additive random component due to both observation noise and nonzero value of time lag $t_{k+1} - t_k$ in the observed TS $\{y(t_k)\}$.

(iii) Obtaining an ensemble of model parameters corresponding to PDF (7) by MCMC sampling. This procedure consists of two stages. First, the rough maximum of PDF (4) with likelihood (5) should be found to initiate the Markov process of the MCMC method. For instance, in the examples given in the current paper (see the next section) we use the variable metric (quasi-Newtonian) method [22] for optimization of the function of multiple arguments. Second, we apply the method based on the Metropolis-Hasting MCMC technique that was described in detail in [19]. Following this

method, generation of each element $(\boldsymbol{\mu}_k, \sigma_{\xi_k})$ of the sought ensemble consists of the following:

(a) Laplace integration of PDF (4) over latent variables \mathbf{u} to calculate PDF (7) value (see [19] and the next section for more detail);

(b) the Metropolis-Hasting step to PDF (7) according to [23].

(iv) Calculating matrix \mathbf{C}_γ using Eq. (10) and matrices \mathbf{C}_1 and \mathbf{C}_2 estimated from the ensemble $\{\boldsymbol{\mu}_k\}$ generated at the previous step. Thus we will get a joint ensemble $\{\boldsymbol{\mu}_k, \gamma_k\}$.

(v) Iterating model (8) at each parameter vector from ensemble $\{\boldsymbol{\mu}_k, \gamma_k\}$ at certain time moments in the future and classifying the obtained dynamic regimes (further, in the next section we will use periodicity of the solution as a classification sign of dynamic regime). This results in an ensemble of qualitatively different regimes of the system behavior, which can be observed in the future. Having such a statistical ensemble we calculate the probability to observe every possible regime at each time moment and this is a solution to the problem.

IV. RESULTS

We will illustrate the proposed approach on examples of TS generated by two different dynamical systems with a slowly varying control parameter. As the first example we will use the Rössler system that is a third-order system of ordinary differential equations:

$$\begin{cases} \dot{x} = -y - z \\ \dot{y} = x + 0.2y \\ \dot{z} = 0.2 + z(x - c) \end{cases}. \quad (11)$$

The scalar TS $y(t_k)$ generated by system (11) has duration $T = 7500$ [in the time units of system (11)], with the control parameter c linearly changing from 5.13 to 4.41 on this time interval. The system (11) exhibits chaotic behavior for all c from this range.

We used for the reconstruction the TS containing the noise component arising because of large discretization steps $\Delta t = 0.1, \Delta t = 0.2$, and $\Delta t = 0.5$. To extract from such a series sample data fit for constructing a discrete model of the form (3) we proceeded as follows. The phase trajectory $\{\mathbf{y}(t_k)\}_{t_k=0}^T$, $\mathbf{y} \in \mathbb{R}^d$ was reconstructed by the scalar TS in the space of dimension $d = 3$ by the method of delays (for the series under study the delays were chosen to be $\tau = 5$, $\tau = 2.5$, and $\tau = 1$). Then, we registered intersections of this trajectory with the secant specified by the equation $y_3 = 0$. The system was reconstructed using the sequence of vectors

$$\mathbf{y}_i = \frac{1}{2} \begin{pmatrix} (y_1(t_{k_i}) + y_1(t_{k_i+1})) \\ (y_2(t_{k_i}) + y_2(t_{k_i+1})) \end{pmatrix}, \quad \begin{matrix} y_3(t_{k_i}) > 0 \\ y_3(t_{k_i+1}) < 0 \end{matrix}$$

(here i is the number of the intersection of phase trajectory and secant). Clearly, the magnitude of the random component in the series $\{\mathbf{y}_i\}$ depends on the step $\Delta t = t_{k+1} - t_k$: the larger

³We center the time variable t in the middle of the observed TS, hence we obtain $\langle t \rangle = 0$.

Δt , the greater is the error of registering these intersections. The sequence of the first component of vector \mathbf{y} is given by blue dots in Fig. 1. The red dots in the same figure stand for the “future” TS.

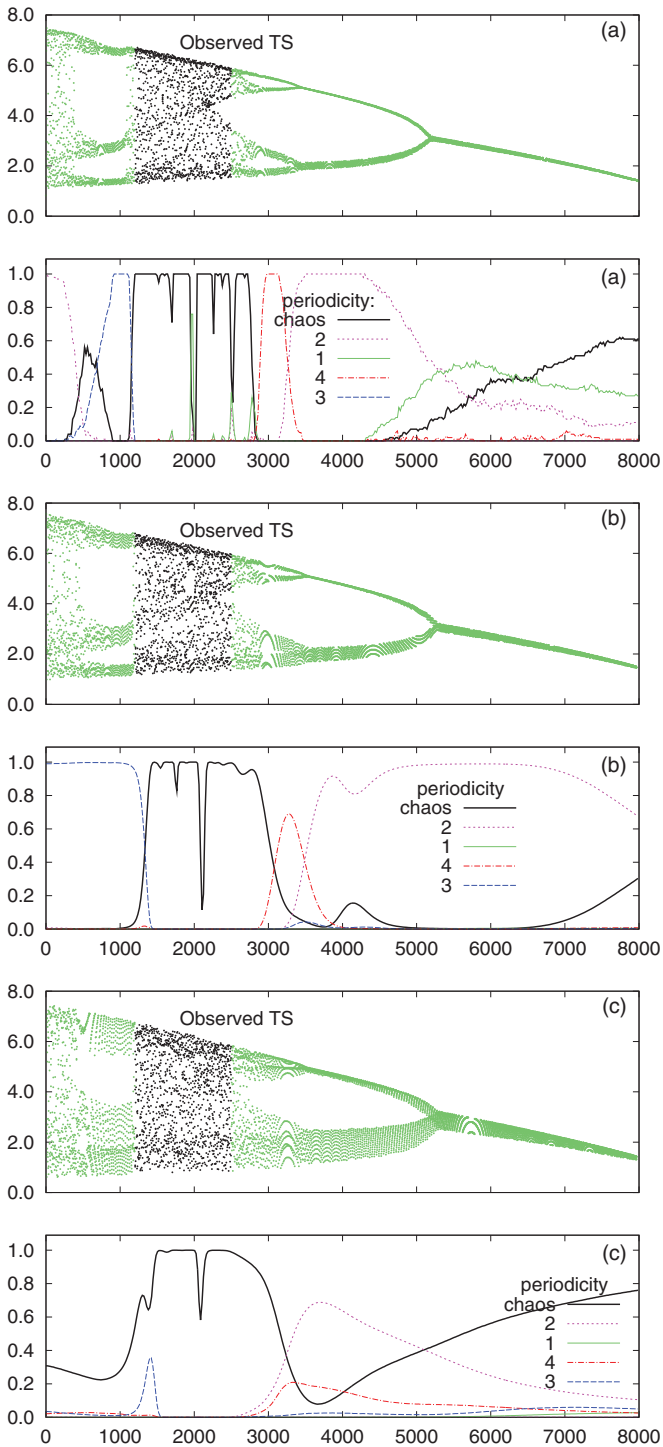


FIG. 1. (Color online) Prognosis of Rössler system behavior for different noise levels: (a) for $\Delta t = 0.1$, (b) for $\Delta t = 0.2$, and (c) for $\Delta t = 0.5$ each represented by a pair of panels. The upper panel shows a sequence of y_1 variable versus time index. Black dots are the observed time series; gray dots (green online) are the time series to be predicted. The lower panel shows time dependencies of calculated probabilities of observed and predicted dynamic regimes.

For construction of prognosis of system behavior we used the EO model in the form (3) for the number of neurons $m = 5$. Numerical analysis of function (7) constructed for such a model was done by the Metropolis-Hasting (MH) MCMC technique [23] within the framework of which a random sequence of values of μ and σ_ξ converging to the distribution (7) was constructed. It was found that for the given TS an optimal (for the considerations presented above) length of the segment w entering Eq. (5) is 4. In each iteration of this technique, we found values of latent variables \mathbf{u} corresponding to the maximum of function (5) [and hence to PDF (4)], calculated a quadratic matrix approximating the exponent of this function by the Laplace method, calculated the integral (7), and generated new values of μ and σ_ξ by the MH method. The first 50 000 iterations of the process were neglected to exclude the transition regime, and the next 10 000 steps were used to construct an ensemble of parameters μ of the model. These MCMC simulations were the most time consuming: four hours on Intel Core2Duo 2.4-GHz processor. To initiate the MCMC process we took the values of μ and σ_ξ obtained as a result of search of the maximum of function (4) by the quasi-Newtonian method [22] that was employed also for seeking the maximum of Eq. (5) by \mathbf{u} at each MH iteration. The error of linear extrapolation of the trend of parameters a was taken into account in accord with the algorithm described above. Further, by iterating functions (8) corresponding to the obtained values of μ and by extrapolating them outside the observation interval we generated an ensemble of 10 000 TS. Then ratios of TS with a certain type of behavior were calculated for each time instant that gives us estimations of probabilities of corresponding dynamical regimes.

The time dependence of the probability of different dynamical regimes obtained by the TS described above generated by model (8) is plotted in Fig. 1. It is seen that the model correctly detects the observed behavior of the system and the sequence of behavior changes in the future ($t > 2500$) for $\Delta t = 0.1$ and $\Delta t = 0.2$ [Figs. 1(a) and 1(b)]. In particular, bifurcations resulting in the transition from chaos to the regime of period 2 are predicted with confidence. For stronger noise corresponding to $\Delta t = 0.5$ [Fig. 1(c)] the model only reproduces the behavior within the observation interval and apparently fails to predict anything. Even at the lowest level of noise we have sufficiently low probability of the most remote regime of period 1 [or probability close to zero at higher noise level; Fig. 1(b)] as well as “false” prognosis of chaotic regime at $t > 4500$ and period-2 regime at $t > 5100$ [Fig. 1(a)]; such an uncertainty is a consequence of the parameter extrapolation error (10), which becomes appreciable in this time region. In the opposite direction (in the “past”), destruction of chaos and further detection of a period-3 regime have been successfully predicted [Figs. 1(a) and 1(b)], but we cannot forecast for sure the following cascade of period doubling resulting in chaos at $t < 800$. This failure is caused by impossibility of prognosis of more complex system behavior (at $t < 800$) by a fragment of more simple observed behavior ($1300 < t < 2500$) due to lack of information about the system phase space contained in the observed TS.

As the second example we will use the model of the mesoscopic photochemical system [24–26]. The model simulates

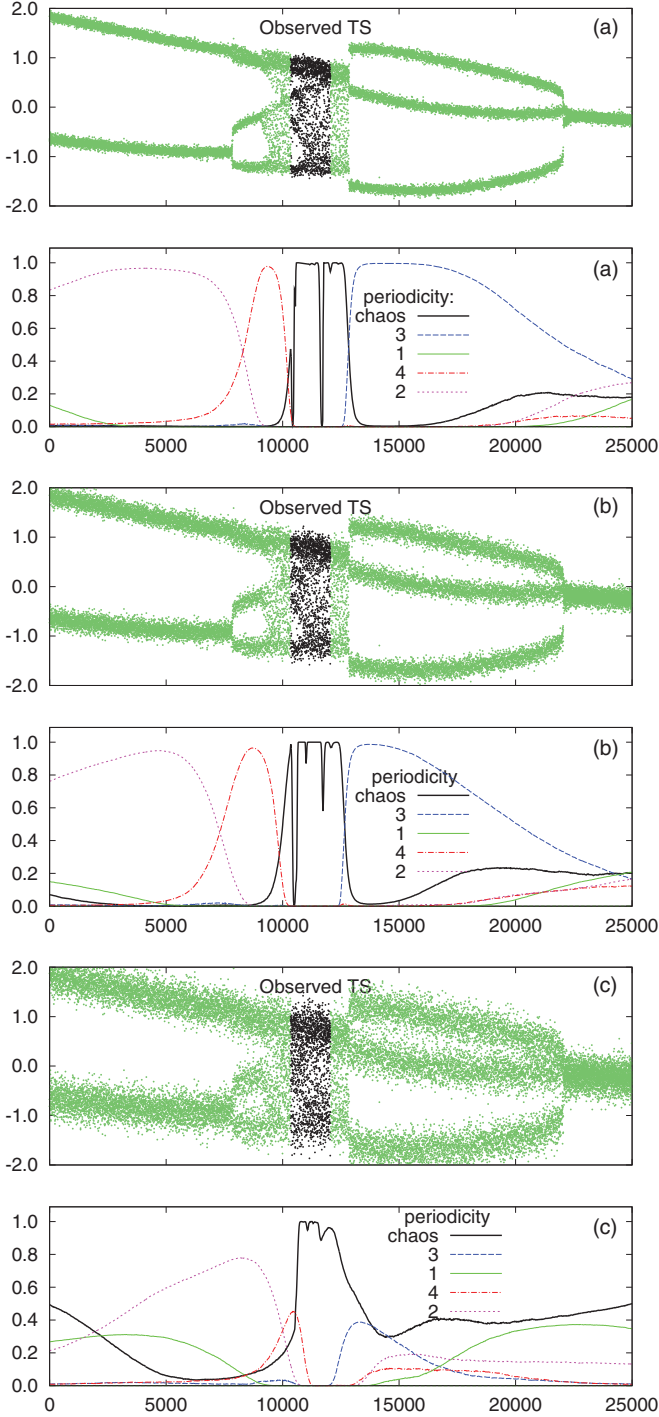


FIG. 2. (Color online) Prognosis of the mesospheric photochemical system behavior for three different noise levels: (a) $\sigma_\xi = 0.05$, (b) $\sigma_\xi = 0.1$, and (c) $\sigma_\xi = 0.2$. The upper panels show the sequence of y_2 variable versus time index. Black dots represent the observed time series; gray (green online) dots are to be predicted. The lower panels depict time dependencies of calculated probabilities of the predicted dynamic regimes on time index.

the behavior of five chemical components: O, H, O₃, OH, and HO₂. The concentrations of these constituents are denoted below as x_1, x_2, x_3, x_4 , and x_5 , respectively. The model includes the following set of differential equations describing dynamics

of the photochemical system:

$$\begin{cases} \dot{x}_1 = -(a_9 + 2a_{11}x_1 + a_{10}x_3 + a_4x_4 + a_5x_5)x_1 \\ \quad + a_1x_2x_5 + a_{15}x_4^2 + a_{10}s(t)x_3 + 2a_8s(t) \\ \dot{x}_2 = -[a_6 + 2a_{12}x_3 + (a_1 + a_2 + a_{14})x_5]x_2 \\ \quad + a_4x_1x_4 + a_7s(t)r \\ \dot{x}_3 = -[a_{10}x_1 + a_{12}x_2 + a_{13}x_4 + a_{16}s(t)]x_3 + a_9x_1 \\ \dot{x}_4 = -(a_4x_1 + 2a_{15}x_4 + a_3x_5 + a_{13}x_3)x_4 \\ \quad + a_5x_1x_5 + a_{12}x_2x_3 + 2a_{14}x_2x_5 + a_7s(t)r \\ \dot{x}_5 = -[a_5x_1 + a_3x_4 + (a_1 + a_2 + a_{14})x_2]x_5 \\ \quad + a_6x_2 + a_{13}x_3x_4 \end{cases} \quad (12)$$

Here a_1 – a_{16} stand for coefficients of the chemical reactions; r is a control parameter of the model which denotes a mixing ratio of H₂O. The function $s(t)$ parametrizes periodic forcing of the system resulting from diurnal variations of the photolysis rate [24]. We took for analysis three scalar TS $x_2(t_k)$ of duration $T = 1700$ days each with different noise levels σ_ξ and linearly changing r from 4.93 to 4.65.

Two-dimensional phase space was reconstructed for each TS by delay coordinates, then the Poincaré section was used, which is defined in time domain by equation $t = c \pmod{24}$ h. Here c is a constant: $0 \leq c < 24$. Resulting TSs consisting of 1700 points each are shown in Figs. 2(a)–2(c) (upper panels). Red color marks the behavior of the system outside the interval of observation, which is to be predicted from “blue” TS where only chaotic behavior was observed.

For the approximation (3) of the obtained two-dimensional ($d = 2$) return map on Poincaré section we used $m = 3$ neurons. For the analysis of the posterior PDF (5) the same approach was used as in the first example. To address different noise levels three different time series were generated for $\sigma_\xi = 0.05$, $\sigma_\xi = 0.1$, and $\sigma_\xi = 0.2$. The optimal segment lengths were chosen $w = 7$, $w = 6$, and $w = 5$, respectively.

From Fig. 2 one can see that “correct” regimes are predicted with the probability close to unity for certain values of time up to 18 000 for noise level $\sigma_\xi = 0.05$ [Fig. 2(a)] and up to 16 000 for noise level $\sigma_\xi = 0.1$ [Fig. 2(b)]. Correspondingly, the sequence of those regimes is also predicted with very high confidence. At the same time we failed to predict the remotest bifurcation when the regime of period 3 is replaced by the regime of period 1 occurring in the original system at $t \approx 22\,500$ since for such a remote futurity the probability of the latter does not exceed any significant level.

At higher noise level $\sigma_\xi = 0.2$ the model confidently reproduces the behavior of the system (chaos) within observation interval only except the regime of period 2 in the interval of $6000 \leq t \leq 9000$ [Fig. 2(c)] accepting the probability 0.5 as a confidence level. This failure is concerned with sharp growth of extrapolation error.

V. CONCLUSION

Let us make a few comments concerning the problem of prognosis of DS behavior by observable dynamics. A “too great” difference between dimension d of the used model and topological dimension of the chaotic attractor determining the evolution of the DS during observation may narrow substantially regions of “likeness” of the model and the

system in the corresponding spaces of parameters and thus limit possible prognosis of system behavior. This situation is typical for reconstruction of DS by scalar TS. Indeed, as follows from the Takens theorem, when an attractor of a DS is reconstructed by a scalar series, the dimension of the phase space ensuring embedding may achieve $(2D + 1)$, where D is the dimension of the system underlying the dynamics. Therefore the dimension of the model of EO needed for reconstruction may, generally speaking, be more than twice higher than the attractor dimension. Apparently, the larger this difference, the less coarse is the model in terms of extrapolation of the system behavior beyond the initial TS. This is because the reconstructed initially “low-dimensional” phase trajectory becomes less and less informative for the multidimensional phase space of the model. In other words, the probability of constructing in the model parameter space of a trend corresponding to the system behavior decreases. Besides, for a “too high-dimensional” model, there may appear the effect described in the work [27], according to which nonstationarity of the TS generated by a relatively low-dimensional system manifests itself in natural (autonomous) dynamics of the model, which leads to unpredicted model behavior outside the observation interval. Detailed investigation of the dependence of prognosis quality (and horizon) on the relationship between embedding dimension and topological dimension of the attractor will be considered elsewhere.

Another interesting feature of the proposed approach concerns the case of catastrophic bifurcation in a future system behavior, that is, hard birth of a new dynamic regime. Since in this situation the phase trajectory leaves the region of phase space that was used for model construction, we will be able to predict the time moment of such bifurcation, but, in the general case, we will say nothing about further behavior of the system. Thus prediction of catastrophes could be one of the important applications of the approach.

Note that the results presented above were obtained neglecting in the reconstruction procedure *qualitative* prior restrictions on the “strength” of nonautonomy of the reconstructed system. At the same time, it may be expected that inclusion into the Bayesian approach of prior information about the measure of process nonstationarity retrieved from the TS, for example, by the methods described in [1–5] will improve accuracy of prognosis. Development of the technique of effective model extrapolation, including allowance for information about the measure of nonstationarity of the studied process will be considered elsewhere.

ACKNOWLEDGMENT

This work was supported by Federal Target Program “Scientific and research and educational staff of innovative Russia.”

-
- [1] M. B. Kennel, *Phys. Rev. E* **56**, 316 (1997).
 - [2] J. B. Gao, *Phys. Rev. Lett.* **83**, 3178 (1999).
 - [3] C. Rieke, K. Sternickel, R. G. Andrzejak, C. E. Elger, P. David, and K. Lehnertz, *Phys. Rev. Lett.* **88**, 244102 (2002).
 - [4] T. Schreiber, *Phys. Rev. Lett.* **78**, 843 (1997).
 - [5] A. Witt, J. Kurths, and A. Pikovsky, *Phys. Rev. E* **58**, 1800 (1998).
 - [6] P. Yang, G. Brasseur, and J. Gille, *Physica D* **76**, 331 (1994).
 - [7] I. Li, P. Biswas, and S. Islam, *Atmos. Environ.* **28**, 1707 (1994).
 - [8] B. Wang, A. Barcilon, and Z. Fang, *J. Atmos. Sci.* **56**, 5 (1999).
 - [9] G. Frank, T. Lookman, and M. Nerenberg, *Physica D* **46**, 427 (1990).
 - [10] H. Srivastava, S. Bhattacharya, and K. Sinha Ray, *Geophys. Res. Lett.* **23**, 3519 (1996).
 - [11] E. Bagarinao, K. Pakdaman, T. Nomura, and S. Sato, *Phys. Rev. E* **60**, 1073 (1999).
 - [12] E. Bagarinao, K. Pakdaman, T. Nomura, and S. Sato, *Physica D* **130**, 211 (1999).
 - [13] A. Feigin, Y. Molkov, D. Mukhin, and E. Loskutov, *Radiophys. Quantum Electron.* **44**, 348 (2001).
 - [14] K. Hornik, M. Stinchcombe, and H. White, *Neural Networks* **2**, 359 (1989).
 - [15] *The Handbook of Brain Theory and Neural Networks*, edited by M. A. Arbib (MIT, Cambridge, MA, 1995).
 - [16] K. Judd, *Phys. Rev. E* **67**, 026212 (2003).
 - [17] R. Meyer and N. Christensen, *Phys. Rev. E* **62**, 3535 (2000).
 - [18] D. N. Mukhin, A. M. Feigin, E. M. Loskutov, and Y. I. Molkov, *Phys. Rev. E* **73**, 036211 (2006).
 - [19] E. M. Loskutov, Y. Molkov, D. N. Mukhin, and A. M. Feigin, *Phys. Rev. E* **77**, 066214 (2008).
 - [20] R. Hegger, H. Kantz, and T. Schreiber, *Chaos* **9**, 413 (1999).
 - [21] R. M. Neal, Technical Report No. CRG-TR-93-1, Department of Computer Science, University of Toronto, 1993.
 - [22] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C* (Cambridge University Press, Cambridge, England, 1992).
 - [23] W. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in Practice* (Chapman and Hall, London, 1996).
 - [24] A. Feigin, I. Kononov, and Y. Molkov, *J. Geophys. Res.* **103**, 25447 (1998).
 - [25] G. Sonnemann, A. Feigin, and Y. Molkov, *J. Geophys. Res.* **104**, 30591 (1999).
 - [26] G. Sonnemann and A. M. Feigin, *Phys. Rev. E* **59**, 1719 (1999).
 - [27] R. Hegger, H. Kantz, L. Matassini, and T. Schreiber, *Phys. Rev. Lett.* **84**, 4092 (2000).