Chaos, Solitons and Fractals 104 (2017) 327–337

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

journal homepage: www.elsevier.com/locate/chaos

Bayesian optimization of empirical model with state-dependent stochastic forcing

Andrey Gavrilov*, Evgeny Loskutov, Dmitry Mukhin

Institute of Applied Physics of the Russian Academy of Sciences, 46 Ul'yanov Street, Nizhny Novgorod 603950, Russia

ARTICLE INFO

Article history:
Received 31 May 2017
Revised 1 August 2017
Accepted 29 August 2017

Keywords:
Time series analysis
Bayesian methods
Stochastic processes
Artificial neural networks

ABSTRACT

A method for optimal data simulation using random evolution operator is proposed. We consider a discrete data-driven model of the evolution operator that is a superposition of deterministic function and stochastic forcing, both parameterized with artificial neural networks (particularly, three-layer perceptrons). An important property of the model is its data-adaptive state-dependent (i.e. inhomogeneous over phase space) stochastic part. The Bayesian framework is applied to model construction and explained in detail. Particularly, the Bayesian criterion of model optimality is adopted to determine both the model dimension and the number of parameters (neurons) in the deterministic as well as in the stochastic parts on the base of statistical analysis of the data sample under consideration. On an example of data generated by the stochastic Lorenz-63 system we investigate this criterion and show that it allows to find a stochastic model which adequately reproduces invariant measure and other statistical properties of the system. Also, we demonstrate that the state-dependent stochastic part is optimal only for large enough duration of the time series.

1. Introduction

A traditional method of modeling the observed processes generated by a nonlinear dynamical system is the reconstruction of the underlying evolution operator [1,2]. The first-principle models (FPMs), i.e. models based on axiomatic physical or other laws and rules of system operation, are hard to apply for modeling various natural physical phenomena, biological, socio-economic processes, and so on mainly because of the extreme generality of these models: the complexity and spatial distribution of the system under consideration frequently leads to cumbersome models, hence, to a large scatter of qualitative prognostic estimates because of their strong sensitivity to subgrid parameterizations. As a result, the FPMs are, in the majority of cases, redundant for description of the dynamics of particular phenomena. Reduction of such models to simpler forms based on selecting variables with definite time scales enables qualitative description of the mechanisms underlying the studied phenomena (for instance, there exist climatic models of intermediate complexity [3–5], basic dynamic models of atmospheric photochemistry [6,7], conceptual models of natural phenomena [8,9], etc.). However, in such procedures the models often lose their connection to the experimentally measured time series and, hence, their ability of quantitative forecasting. Alternatively to FPMs, empirical models are aimed at extraction of information about the dynamics of the system directly from the time series: the most statistically justified (in terms of certain empirical criterion of validity, see below) model of evolution operator is chosen, while the FPM equations are not brought into play. In fact, the empirical approach implies a model sufficient for reconstruction of the system's dynamical properties manifested in the observed dynamics at the timescales of interest. In the past 30 years (see [2] and the literature therein), the empirical models, by virtue of their universality and independence of different physical assumptions, became very popular in forecasting climate dynamics [10–12], financial time series [13], investigation of living systems [14], and so on.

A typical situation in the construction of an empirical model is when there is no one-to-one coupling between the consecutive states in the space of the phase variables reconstructed from the time series, i.e., the problem of reconstructing an evolution operator as a single-valued function is incorrect. In the context of real complex systems, this circumstance is a consequence of high dimension of the system that has generated the time series: the number of dynamic variables of the model with which it is reasonable to work is limited by finite size of available data sample. There are a lot of special studies devoted to the problem of appropriate empirical reduction of observed high-dimensional data: from the standard principal component analysis [15–18] and more

* Corresponding author.
E-mail addresses: gavrilov@ipfran.ru (A. Gavrilov), loskutov@ipfran.ru (E. Loskutov), mukhin@ipfran.ru (D. Mukhin).

http://dx.doi.org/10.1016/j.chaos.2017.08.032
0960-0779/

advanced linear techniques [19–27] to nonlinear reduction methods [28–36]. Anyway, when working with real data we have to construct a model in a certain low-dimensional subspace of the full phase space of the system, where, generally speaking, there is no single-valued evolution operator. In model examples, such a situation always occurs if interactive noise is added to the data-generating equations. Such an ambiguity is traditionally regarded to be a defect of the model and is described as a random (Gaussian) error additive to the evolution operator function. For example, in the works [37,38] artificial neural network (ANN) is used to approximate the evolution operator of the low-dimensional deterministic system assuming that the defect of the model is Gaussian noise with constant dispersion.

On the other hand, the described ambiguity of the evolution operator is not, obviously, homogeneous over the phase space of the model due to the nonlinearity of the studied system. In the present work we use the ANN-based model modified according to the following idea: it is possible to take this inhomogeneity into consideration; namely, in the regions of phase space with small scatter in the dependence of the current state on the previous state, the evolution operator may be reconstructed more accurately, thus decreasing systematic approximation bias. A more general structure of the model taking this inhomogeneity into account is proposed in [12,39], where a random component of the model is supposed to depend on phase variables, i.e. to be state-dependent, with both the stochastic and the deterministic components being approximated by the ANN-based function in the form of a three-layer perceptron [40]. We have chosen ANN functions because they are known to be effective approximators of unknown dependencies (see [40]), their efficiency in constructing prediction models by time series was shown in many works (e.g., [13,39,41–44]). The convenience of this approximation resides on the following: firstly, the obtained functions are bounded, which allows avoiding worsening of the approximation accuracy at the edges of the definition domain. Secondly, it is possible to efficiently increase dimension and complexity of the model by increasing a small number of parameters. The drawback is degeneracy of the ANN parameter space that may be overcome by introducing regularizing restrictions (see Section 2.2.3).

The main question arising with such a modification of the model is the following: is it justified to increase the number of parameters by introducing an additional function into the model or, in other words, is the resulting model more economic in terms of description of the time series? To answer this question we need a criterion for comparing models of different complexity or, which is the same (without loss of generality), with different numbers of parameters. In this work we use a very general optimality criterion based on Bayesian evidence [45]. This criterion is transparently interpreted in the framework of the Bayesian paradigm: if all the considered models are equiprobable *a priori*, then the best model is the one reproducing the observed data with the highest probability. From the information theory point of view, this criterion is equivalent to the minimum description length principle [37], i.e. the best model is required to provide the most informationally-compressed representation of the observed data. We also developed a numerical method for estimating the Bayesian evidence of the described ANN-based stochastic model, including the procedure of Bayesian regularization of ANN.

The paper is organized in the following way. In Section 2 the structure of the evolution operator model is described (Section 2.1), the measure of model optimality is introduced (Section 2.2), and the method of its numerical assessment is presented (Section 2.3). In Section 3 the proposed approach is used for reconstructing the stochastic Lorenz system by its scalar time series having different duration and generated at different noise level. Also, the properties of the Bayesian evidence with respect to the proposed stochastic model are investigated in detail in this section. It is shown that the stochastic model admitting a state-dependent random term is, generally, more preferable than the model with constant noise and gives a better description of the dynamic properties determining the observed behavior. Finally, the concluding remarks are formulated in Section 4.

2. Method of optimal stochastic model construction

2.1. Stochastic model of evolution operator

Let us consider discrete time series $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$, $\mathbf{x}_n \in \mathbb{R}^D$ of characteristics $\mathbf{x}$ of an unknown dynamical system at N equidistant time moments, for which we want to construct an evolution operator model. Evolution operator, by definition, is a function acting in a phase space of the system. However, in a general case analysis of the observed data does not allow us to make a conclusion about finiteness of the dimension of the phase space of the dynamical system that has generated the time series $\mathbf{X}$. Therefore, we have to construct a model in a certain subspace of dimension d, whose elements must be derived from observables $\mathbf{X}$ (see the particular realizations below) and will be denoted by $\mathbf{u}_n \in \mathbb{R}^d$. Obviously, in general case there is no single-valued evolution operator $\mathbf{Q}: \mathbf{u}_n \to \mathbf{u}_{n+1}$ in such a subspace. The arising ambiguities of this map may be described stochastically by defining the evolution operator $\mathbf{Q}$ as a random dynamical system [46]. In practical applications, it can be simplified (see [39] for details) to the following form:

$$\mathbf{u}_{n+1} = \mathbf{f}(\mathbf{u}_n) + \hat{\mathbf{g}}(\mathbf{u}_n) \cdot \boldsymbol{\xi}_n. \tag{1}$$

Here the mapping $\mathbf{f}: \mathbb{R}^d \to \mathbb{R}^d$ represents the "deterministic" part of the model; $\boldsymbol{\xi}_n \in \mathbb{R}^d$ is a normal delta-correlated random process, with vector components being mutually independent and having zero mean and unity variance; the matrix function $\hat{\mathbf{g}}: \mathbb{R}^d \to \mathbb{R}^{d\times d}$ maps model state $\mathbf{u}_n$ to the matrix having dimension $d \times d$ and represents, together with $\boldsymbol{\xi}_n$, the state-dependent stochastic perturbation term. The model (1) proved to be efficient as applied to reconstructing systems of different complexity [11,39] because the strong Gaussianity assumption about $\boldsymbol{\xi}_n$ is complemented here by the state-dependence (i.e. inhomogeneity over the phase space) of the covariance matrix.

One of the most widely used ways to specify the phase variables $\mathbf{u}$ via the data $\mathbf{X}$ in case of small dimension D is the Takens delay method and its modifications [47], when the phase variable is obtained by successively shifting the time series $\mathbf{X}$ by an arbitrary time lag. In practice, different values of the time lag lead to different (more or less adequate) structures of the phase space projections and, eventually, to different results of modeling. Therefore, usually the lag is taken close to the time scale of interest. For example, for the lag equal to one time step the Takens method gives:

$$\mathbf{u}_k = \left(\mathbf{x}_k, \mathbf{x}_{k-1}, \ldots, \mathbf{x}_{k-(l-1)}\right), \tag{2}$$

Here l is the number of delays. Otherwise, if the dimension D is high, the empirical dimensionality reduction methods (see Section 1) can be applied as a preprocessing step before the delaying. In this work, for notation and demonstration simplicity, we restrict our consideration to the case of the scalar time series $X = (x_1, \ldots, x_N)$, $x_n \in \mathbb{R}$ and the reconstruction of the phase space by the Takens delay method with the lag equal to one time step (with such a restriction, any lag can be set by a proper resampling of the time series). Obviously, in this case the dimension d is equal to l. The form of the model (1) in this case will apparently become much simpler due to the presence of $d-1$ trivial couplings