Physics of Vibrations
1999

Vol. 7. Number 2,
pp. 61 – 74

PREDICTABILITY OF LINEAR AND NONLINEAR AUTOREGRESSIVE MODELS

O.L. Anosov
Department of Mathematical Modeling, Regional Cardio Center, 67 Sudogodski Rd., Vladimir 600023, Russia

O.Ya. Butkovskii
State University, 87 Gorky Street, Vladimir 600029, Russia

Yu.A. Kravtsov
Space Research Institute, Russian Academy of Sciences, 87/32 Profsoyuznaya Street, Moscow 117810 Russia
Space Research Center, Polish Academy of Sciences, 18a Bartycka, Warsaw 00716, Poland

V.A. Protopopescu
Oak Ridge National Laboratory, Computer Science and Mathematics Division, Oak Ridge 2008-6355, TN 37831

(Received May 17, 1999)

We derive bounds for the time of predictability in various nonlinear systems by using linear autoregressive (LAR) and nonlinear autoregressive (NAR) models. Based on the analysis of degree of predictability, defined as a correlator between the observation and the forecast, we relate the predictability time in LAR models to the correlation time of the observed process. This relationship is illustrated by a number of examples for which the degree of predictability can be determined analytically or numerically. In particular, we analyze the ability of LAR models to predict random processes with exponential and Gaussian correlation functions, differentiable and nondifferentiable random processes, discrete maps, and multidimensional continuous dynamical series (Roessler system). In all the cases the predictability time exceeds the correlation time by a factor no larger than 1.1–1.4.

A predictability assessment using NAR algorithms is carried out for three classes: random and continuous chaotic dynamical processes, as well as one-dimensional nonlinear maps in chaotic regime. It is shown that, like in the case of LAR, for processes of random (nondynamical) nature the maximum predictability time provided by NAR models, does not exceed the correlation time. NAR models lead only to an increase of the computation time without noticeably improving the prediction quality. The same is true for multidimensional continuous dynamical processes, like the Roessler system in chaotic regime. In this case, the NAR model also has no noticeable advantages over LAR models, even in noiseless situations.

The only exception is provided by discrete maps in chaotic regimes, where the NAR model provides a predictability time which may significantly exceed the correlation time and brings the former to a limiting "predictability horizon".

PACS: 05.45.+b

1. Introduction

Autoregressive models are widely used for prediction and forecasting of real processes, based on analysis of continuous and discrete time series. Despite their broad utilization, the question of predictability limits guaranteed by autoregressive models is still far from being settled. Noticeable

progress has been achieved only in linear autoregressive (LAR) models [1, 2].

LAR models are frequently used to forecast the outcome of a process over a certain period of time. Based on such models, the predictability time for the process, $\tau_{\text{pred}}^{\text{LAR}}$, is upper-bounded by the correlation time τ_{c} of the process,

$$\tau_{\text{pred}}^{\text{LAR}} < B^{\text{LAR}} \tau_{\text{c}}, \qquad (1)$$

where B^{LAR} is a factor of the order of unity. Although inequality (1) has long been accepted as valid, to date it has no convincing substantiation. In fact, in many cases this inequality is postulated as obvious (see [2]). By developing the approach proposed in [2], we will give more general arguments in support of inequality (1). This general statement will be illustrated by a number of examples indicating that the factor B^{LAR} can hardly exceed 1.1–1.4 (Secs. 2–4).

The question of the quality of forecasting ensured by nonlinear autoregressive (NAR) models for various nonlinear processes is still open. The aim of the second part of this paper (Secs. 6–8) is to assess the limit of predictability of NAR models for following three classes of processes:

(i) processes of truly random (nondynamical) nature;

(ii) continuous chaotic dynamical processes;

(iii) nonlinear maps in the chaotic regime.

Below we outline theoretical and empirical (numerical) arguments to support the argument that for processes (i) and (ii) the predictability time $\tau_{\text{pred}}^{\text{NAR}}$ provided by a NAR model, satisfies a relationship similar to Eq. (1) for LAR,

$$\tau_{\text{pred}}^{\text{NAR}} \approx B^{\text{NAR}} \tau_{\text{c}}, \qquad (2)$$

where the factor B^{NAR} is less than 1.1–1.4 for the overwhelming majority of processes of classes (i) and (ii).

On the other hand, we found that for discrete maps in the chaotic regime the predictability time may reach the limiting value

$$\tau_{\text{lim}} \approx \frac{1}{\lambda^{+}} \ln \frac{A}{\sigma_{\text{f}}}, \qquad (3)$$

where λ^{+} is the largest Lyapunov exponent of the nonlinear map, A is the typical amplitude of the observed time series and σ_{f} is the root-mean-square value of dynamical fluctuations which always exist in any real system. The limiting time of predictability (3) can be identified with the notion of "predictability horizon" [4] for chaotic processes.

2. LAR and NAR models

Let us observe a process $y(t)$, with zero average for a certain time interval preceding the moment t^0. Within the LAR model [1], the forecast $z(t)$ at some future time $t = t^0 + \tau$ is given as a linear combination of values of the observed process $y(t)$ in preceding moments t^0, $t^0 - \tau$, $t^0 - 2\tau, \ldots$,

$$z(t) = a_1 y(t^0) + a_2 y(t^0 - \tau) + \ldots + a_M y\left[t^0 - (M-1)\tau\right], \qquad (4)$$

where M denotes the order of LAR model and $\tau = t - t^0$ is the forecasting time. The coefficients a_1, $a_2, \ldots, a_M$ of the LAR model are usually calculated by minimizing the mean squared forecast error

$$\phi(\tau) = \left\langle \left[z(t) - y(t) \right]^2 \right\rangle. \qquad (5)$$

Angular brackets indicate either "theoretical" averaging on the basis of probabilistic characteristics of the processes $y(t)$ and $z(t)$, or empirical averaging

$$\phi(\tau) = \frac{1}{K} \sum_{k=1}^{K} \left[z(t_k^0 + \tau) - y(t_k^0 + \tau) \right]^2 \qquad (6)$$

on the ensemble of processes $z(t)$ and $y(t)$ sampled at sufficiently many initial moments of time t_k^0, $k = 1, \ldots, K$.

Minimization of functionals (5)–(6) with respect to (4) leads to the Yule–Walker set of normal equations [3, 5]

$$\sum_{k=1}^{M} a_k r_{|j-k|} = r_j, \qquad j = 1, 2, \ldots, M, \qquad (7)$$

where

$$r(\Delta t) = \frac{R(\Delta t)}{R(0)} = \frac{\left\langle y(t)\, y(t - \Delta t) \right\rangle}{\left\langle y^2(t) \right\rangle} \qquad (8)$$

is the normalized correlation function of the process $y(t)$ and $r_k \equiv r(k\tau)$ is the value $r(\Delta t)$ for $\Delta t = k\tau$, $k = 1, 2, \ldots$

The coefficients a_k are expressed in terms of the correlation coefficients r_k by solving linear algebraic system (7). Thus, for $M = 1$ (first-order autoregression), we have

$$a_1 = r_1 = r(\tau), \tag{9}$$

while for $M = 2$ (second-order autoregression), the coefficients a_1 and a_2 become

$$a_1 = r_1 \frac{1 - r_2}{1 - r_1^2}, \qquad a_2 = \frac{r_2 - r_1^2}{1 - r_1^2}. \tag{10}$$

As will be shown later, the dependence of coefficients a_j on the correlation function determines a fundamental limitation for the horizon of predictability of LAR models and linear evaluation methods.

NAR models differ from LAR model (1) by the presence of nonlinear terms. A polynomial NAR model of order M and power N for prediction $z(t)$ of the process $y(t)$ at the moment $t = t^0 + \tau$ reads

$$z(t) = \sum_{n_j}^{N} \sum_{m_j}^{M} a_{n_1,\ldots,n_k}^{m_1,\ldots,m_k} y^{n_1}(t^0 - m_1\tau) \ldots y^{n_k}(t^0 - m_k\tau), \tag{11}$$

where the maximum sum of indices n_j is the power N of the regression, $\max(n_1 + \ldots + n_k) = N$, and no number $1 + m_k$ can exceed the order M of regression, $1 + m_k \le M$.

3. Degree and time of predictability

The quality of forecast is usually characterized by the squared absolute error or the relative squared error

$$E(\tau) = \frac{\left\langle \left[z(t) - y(t) \right]^2 \right\rangle}{\left\langle z^2(t) \right\rangle + \left\langle y^2(t) \right\rangle}. \tag{12}$$

However, it is more convenient to use here the notion of "degree of predictability", defined as a correlation between the observed process $y(t)$, and the predicted process $z(t)$, [2, 4, 6–8],

$$D(\tau) = \frac{\left\langle y(t)\, z(t) \right\rangle}{\sqrt{\left\langle y^2(t) \right\rangle \left\langle z^2(t) \right\rangle}}, \qquad \tau = t - t^0. \tag{13}$$

The mean values of both $y(t)$ and $z(t)$ are supposed to be zero.

Although $D(\tau)$ can be expressed in terms of the relative error $E(\tau)$,

$$D(\tau) = \frac{\left\langle y^2(t) \right\rangle + \left\langle z^2(t) \right\rangle}{2\sqrt{\left\langle y^2(t) \right\rangle \left\langle z^2(t) \right\rangle}} \left[1 - E(\tau) \right],$$

it has several advantages over $E(\tau)$ that will become clear shortly.

Due to the natural condition $z(t^0) = y(t^0)$, we have $D(0) = 1$. The more the observed process $y(t)$ deviates from the forecast, the more D deviates from unity. Sooner or later, the degree of predictability becomes zero, $D(\tau \to \infty) \to 0$ (recall that the means of processes $y(t)$ and $z(t)$ are zero). Based on the degree of predictability, we can naturally introduce the time τ_{pred} of predictable behavior, defined as the time interval during which the degree of predictability $D(\tau)$ remains higher than a certain threshold value $1-p$. Thus, the time of predictability τ_{pred} becomes a function of the "confidence level" p and satisfies the equation

$$D\left[\tau_{\text{pred}}(p) \right] = 1 - p. \tag{14}$$

Similarly, we can define the correlation time $\tau_c(p)$, as the time interval during which the correlation function $r(\tau)$ remains higher than $1-p$. The correlation time can be calculated from the equation

$$r\left[\tau_c(p) \right] = 1 - p. \tag{15}$$

4. Bounds for predictability times derived from LAR models

The correlation and predictability times have no prescribed natural ordering. The relation between them depends on the confidence level p, in such a way that the best predictability for small p

($\tau_{pred} > \tau_c$) may be close to the worst predictability ($\tau_{pred} < \tau_c$) for greater p This dependence is shown in Fig. 1, where the correlation function $r(\tau)$ is represented together with the degree of predictability $D(\tau)$.

For LAR model (4) $D(\tau)$ is given by

$$D_{LAR}(\tau) = \left(\sum_{k=1}^{M} a_k r_k \right) \left(\sum_{k=1}^{M} \sum_{j=1}^{M} a_k a_j r_{|j-k|} \right)^{-1/2} . \quad (16)$$

Taking into account that all the coefficients of the LAR model are expressed in terms of the correlation function (see (7)), we conclude that the degree of predictability $D(\tau)$ have no other intrinsic time scale, except the correlation time. Therefore, the predictability time τ_{pred}^{LAR} calculated from Eq. (14) with D(τ) replaced by $D_{LAR}(\tau)$, cannot be essentially different from the correlation time calculated in Eq. (15). In other words, the horizon of predictability for LAR models cannot exceed appreciably the correlation time τ_c. Thus, fundamental inequality (1)
which has been long known to be valid as experimental is recovered here by very simple arguments. We illustrate the validity of relationship (1) with some simple examples.

First-order autoregression ($M = 1$). According to (9), $a_1 = r_1 = r(\tau)$, and using (16) we find that the degree of predictability is equal to the correlation function,

$$D(\tau) = r(\tau). \quad (17)$$

Correspondingly, the predictability time is simply equal to the correlation time $\tau_{pred}^{LAR} = \tau_c$. This conclusion is fully corroborated by numerical calculation in the examples given below. The graph of $D(\tau)$, for first-order LAR models, as well as for random series, or for deterministic series, coincides practically with the correlation curve.

Exponential correlation function $r(\tau)=\exp(-|\tau|/\tau_1)$. For this correlation function, all coefficients, except the first one, are zero and $a_1 = r_1 = r(\tau)$. Thus, $D(\tau) = r(\tau)$ and $\tau_{pred}^{LAR} = \tau_c$ for any confidence level p. We note that the quality of forecast is not improved by increasing of the order M of autoregression. The result of numerical experiments averaging 90 realizations of random numbers with the exponential distribution function is represented by the function

$$g(M) = \frac{\tau_{pred}^{LAR}}{\tau_c} . \quad (18)$$

For this example, $g(M)$ is identically equal to unity (see Fig. 2, curve *1*).

Nondifferentiable random processes. For such processes the correlation function has a derivative

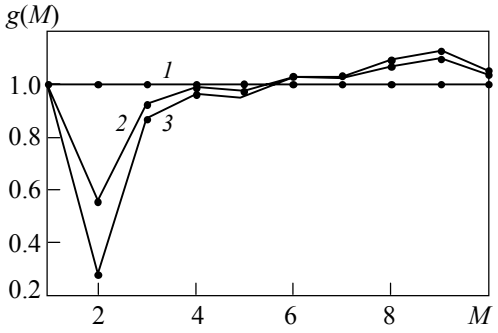

Figure 1. Degree of predictability D and correlation function r as functions of time τ. The predictability time τ_{pred} and the correlation time τ_c are determined from the intersection of curves $D(\tau)$ and $r(\tau)$ with the horizontal line 1–p according to Eqs. (14) and (15).

Figure 2. Dependence of the predictability time on the regression order M. Line *1* corresponds to processes with the exponential correlation function (for any levels of p); line *2* presents processes with the Gaussian correlation function ($p = 0.1$); and line *3* is constructed for the logistic map ($p = 0.1$).

discontinuity at $\tau = 0$, similar to the exponential correlation function $r(\tau) = \exp(-|\tau|/\tau_1)$ [9]. For all processes of this form, the correlation function $r(\tau)$ can be approximated (as $\tau \to 0$) by a "triangle" function, such as $r(\tau) \simeq 1 - |\tau|/\tau_1$, from which the feature mentioned above follows. The forecast does not improve when using higher-order LAR models.

Gaussian correlation function $r(\tau)=\exp(-\tau^2/\tau_2^2)$. This example illustrates that the gain provided by higher-order LAR models is small compared with the first order. When $M = 2$, degree of predictability computed for the Gaussian correlation function according to (16) becomes

$$D_{\mathrm{LAR}}(\tau) = \frac{r(1-2r^4+r^6)}{(1-r^2-2r^4+3r^6-r^8)^{1/2}}. \quad (19)$$

Calculations based on formula (19) using condition (15) show that if the confidence level p exceeds 0.3, then τ_{pred} exceeds τ_{c} only by 1.2. Including the third order ($M = 3$) increases τ_{pred} by at most one percent. Higher-order inclusions lead even to a decrease of the predictability time with respect to τ_{c}. The function $q(M)$ for random series with the Gaussian distribution and correlation functions are shown in Fig. 2 (curve *2*).

Differentiable random processes. As $\tau \to 0$, the Gaussian correlation function $r(\tau)=\exp(-\tau^2/\tau_2^2)$ has quadratic behavior near the origin, namely $r(\tau) \simeq \exp(-\tau^2/\tau_2^2)$. Such a run is characteristic of all differentiable random processes. Thus, for small p the behavior described above for processes with the Gaussian correlation function (i.e., modest improvement and then deterioration of the forecast as a function of the order M of regression) are inherent in a wide class of processes with quadratic run of the correlation function as $\tau \to 0$.

Discrete dynamical series (*logistic map*). The above consideration may be carried over to discrete series with some slight changes. The differences are that the time intervals t, t^0, τ take on only discrete values, say $t = 1, 2, 3, ...$, and one cannot pass to the continuous limit $\tau \to 0$.

We illustrate the features of discrete series on the example of the logistic map

$$y(t+1) = ry(t)\left[1 - y(t)\right], \quad t = 1, 2, \ldots, \quad (20)$$

with parameter $r = 3.82$, which displays fully developed chaotic behavior.

In Fig. 3 the degree of predictability $D_M^{\mathrm{LAR}}(\tau)$, which is in agreement with the first-order LAR model, is represented by a continuous curve, and the degree of predictability for second, fifth, and ninth-order LAR models as points. For $M = 1$, the curve $D_M^{\mathrm{LAR}}(\tau)$ coincides with the correlation function $r(t)$. For $M = 2$, the degree of predictability decreases noticeably, and, as a result, the predictability time $\tau_{\mathrm{pred}}^{\mathrm{LAR}}$ also decreases. With the later increase in the order of regression, the predictability time rises slightly to reach saturation at $M = 9$. The ratio $g \equiv \tau_{\mathrm{pred}}^{\mathrm{LAR}}/\tau_{\mathrm{c}}$ is represented in Fig. 2 by curve *3*. The value g is compared with unity, which is in agreement with general statement (1).

At $M = 2$ we observe a marked decrease of $D(\tau)$ not only for the logistic map but also for other series at times $\tau \geq \tau_{\mathrm{c}}$. A possible reason of this phenomenon may stem from the dominant "sign-variability" of values $y(t)$ at neighboring moments of time: if at a given time t we have $y(t) > y(t-1)$, then at the time $t+1$ it often happens that $y(t+1) < y(t)$, and vice versa.

Multidimensional continuous dynamical series

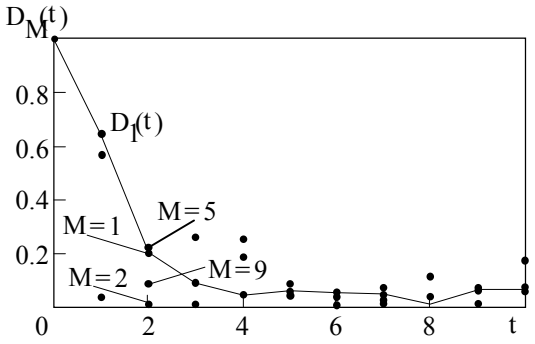

Figure 3. Degree of predictability for the logistic map as derived from LAR models of orders $M = 1, 2, 5, 9$. The first-order LAR model is represented by a line, the other models are represented by points.

(*Roessler system*). For discrete series in chaotic regime, the digitization time as a rule exceeds the correlation time, as may be seen from Fig. 3 for the logistic map.

The same is true for continuous processes, even if we integrate the equations of motion with the digitization step less than the correlation time. To illustrate this, let as consider the Roessler system

$$\begin{aligned} \frac{dy_1}{dt} &= -y_2 - y_3, \\ \frac{dy_2}{dt} &= y_1 + 0.2y_2, \\ \frac{dy_3}{dt} &= 0.2 - 25y_4 + y_1y_3. \end{aligned} \qquad (21)$$

The degree of predictability of the component $y_2(t)$ computed with the LAR model is given in Figs. 4 and 5 and represents the relationship between the parameter $g = \tau_{\text{pred}}^{\text{LAR}}/\tau_{\text{c}}$ and the order M of LAR model. As we see in Fig. 5, the gain in predictability time for this deterministic process is not significantly higher than that for random time series. In all the cases inequality (1) takes place with the factor $B^{\text{LAR}} < 1.4$. We will also see that the same is true for NAR models.

The examples above show that the predictability time of the processes, based on LAR models, is essentially limited by the correlation time. This fundamental limitation is connected to the very essence of the LAR model itself. Formally, the causality principle lies at the basis of this model:

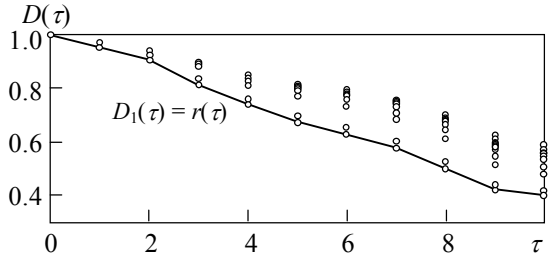

Figure 4. Degree of predictability for the y_2 component of Roessler system (37) as derived from LAR models of orders $M = 1, 2, ..., 10$. The curve shows the predictability for a first-order LAR model, the other data are represented by points.

the value of forecast $z(t)$ is expressed in terms of the process values $y(t-k\tau)$ only at the preceding moments of time.

In this sense, the LAR model is a "linear substitute" for the actual causality constraints. Such a substitute cannot be a serious contender as a long-term forecast tool. Its predictability horizon $\max \tau_{\text{pred}}^{\text{LAR}}$ is comparatively low and, according to (1), does not exceed $B^{\text{LAR}}\tau_{\text{c}}$.

Since only a weak causality principle is used in LAR models, an increase in the regression order M improves predictability only marginally. Indeed, considering higher-order LAR models (4) would only "overload" Yule–Walker set (7), without revealing nonlinear constraints and relationships in the system. This is also quite in agreement with the empirical fact that increasing the linear regression order beyond $M = 2$ practically does not improve the forecast quality and in some cases it even worsens it.

The relationship between the horizon of predictability for LAR models and the correlation time can be explained from a new vantage point. If $y(t)$ is a stationary random process with variance $\langle y^2(t)\rangle = \text{const}$, then it would be natural to expect that the variance $\langle z^2(t)\rangle$ of the forecast $z(t)$ also be constant and equal to $\langle y^2(t)\rangle$. In other words, the following condition should be fulfilled,

$$\langle z^2\rangle = \langle y^2\rangle = R(0),$$

or, equivalently,

$$\sum_{k=1}^{M}\sum_{j=1}^{M} a_k a_j r_{|j-k|} = 1. \qquad (22)$$

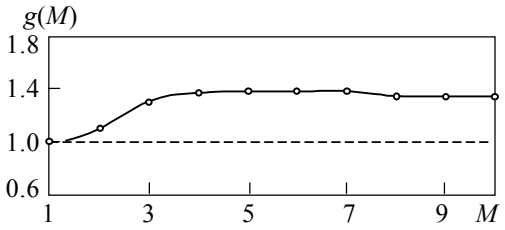

Figure 5. Dependence of the function $g(M)$ (18) on the regression order M for the y_2 component of Roessler system.

Unfortunately, condition (22) cannot be satisfied together with Yule–Walker set (7). In fact, if the coefficients a_k satisfy (7), then the double sum in (22) is equal to

$$\sum_{k=1}^{M}\sum_{j=1}^{M} a_k a_j r_{|j-k|} = \sum_{j=1}^{M} a_j r_j, \qquad (23)$$

and obviously cannot equal unity. Indeed, if we take a trivial example of first-order autoregression ($M = 1$), then $a_1 = r_1$ and in the right-hand side of (22), instead of unity, we have $r_1^2 = r^2(\tau) \le 1$, according to Eq. (16).

The right-hand side of Eq. (22) tends to unity only if $\tau \to 0$. Thus, in spite of natural expectations, the forecast variance $\langle z^2 \rangle$ does not remain constant but tends to zero as $\tau \to 0$. This further limits the applicability of LAR forecasts for times $\tau > \tau_c$.

5. Bounds for predictability times derived from NAR models

5.1. Processes of random (nondynamical) nature

For truly random processes, NAR models present no advantages over LAR models and cannot extend the prediction times beyond the correlations time τ_c of the process $y(t)$ under study. This statement can be motivated in the same way we used for LAR models (4).

Let us write the coefficients $a^{m_1,\ldots,m_k}_{n_1,\ldots,n_k}$ in expansion (11) in a more compact form as a^n_m, where m and n are multi-indices $m = (m_1,\ldots,m_k)$ and $n = (n_1,\ldots,n_k)$. The coefficients a^n_m in NAR model (11) are calculated by minimization of the quadratic functional

$$\Phi = \frac{1}{T}\int_{t^0}^{t^0+T} \left[y(t') - z(t')\right]^2 dt' = \min, \qquad (24)$$

which is a measure of difference between the observed process $y(t)$ and prediction $z(t)$ based on the NAR algorithm. Differentiating Eq. (24) with respect to a^m_n, we derive a set of linear equations for these quantities. The second statistical moments

$$R_{pq}(k\tau) = \left\langle y^p(t')\, y^q(t'+k\tau)\right\rangle, \quad p,q = 1, 2, \ldots, N, \quad k = 1, \ldots, M, \qquad (25)$$

of the observed process $y(t)$ appear in this set as coefficients.

Using nonlinear prediction model (11) in Eq. (12), the degree of predictability $D(\tau)$ can be expressed through statistical moments $R_{pq}(k\tau)$ of the observed process and through coefficients a^m_n, which in turn are expressible through the same moments. Let τ_c be the correlation time of process $y(t)$, i.e., the characteristic time of a second moment is $R_{11}(\tau) = \langle y(t)\, y(t+\tau)\rangle$. In fact, τ_c can be determined from Eq. (15). To fix an idea, we take the parameter p in Eq. (15) to be 0.25, so that the correlation time τ_c satisfies the equation $r(\tau) = 0.75$. There are no reasons to expect that the highest moments $R_{pq}(\tau)$ possess time scales different from τ_c. The same can be said about the predictability time $\tau^{\mathrm{NAR}}_{\mathrm{pred}}$ defined by Eq. (14) and whose value is comparable to the correlation time τ_c.

Therefore, if the observed process is of nondynamical nature and allows only statistical description, then there are no reason to expect that NAR models have advantages over LAR ones. In this case, the predictability time $\tau^{\mathrm{NAR}}_{\mathrm{pred}}$ satisfies condition (2), which is similar to relationship (1) for $\tau^{\mathrm{LAR}}_{\mathrm{pred}}$.

This conclusion can be illustrated by the calculated predictability time $\tau^{\mathrm{NAR}}_{\mathrm{pred}}$ for a process with Gaussian statistics. We assume that the average value of the observed process is zero, $\langle y \rangle = 0$, then all its odd moments turn out to be zero, while all the even moments are expressed through the correlation function $R_{11}(\tau) = \langle y(t)\, y(t+\tau)\rangle \equiv R(\tau)$.

For simplicity we restrict ourselves to a NAR model of second order ($M = 2$) and the second power ($N = 2$),

$$z(t) = a_0 + a_1 y_0 + a_2 y_1 + a_3 y_0^2 + a_4 y_0 y_1 + a_5 y_1^2, \qquad (26)$$

where we denote $y_0 \equiv y(t^0)$ and $y_1 \equiv y(t^0 - \tau)$. Using extremum condition (24) and denoting $R_0 = R(0)$, $R_1 = R(\tau)$, $R_2 = R(2\tau)$, we get a set of six linear equations for six coefficients $a_0, a_1, \ldots, a_5$, which may be separated into two independent subsystems: four homogeneous equations for the coefficients a_0, a_3, a_4, and a_5,

$$\begin{aligned} &a_0 + R_0 a_3 + R_2 a_4 + R_0 a_5 = 0, \\ &R_2 a_0 + 3R_0^2 a_3 + 3R_0 R_1 a_4 + (R_0^2 + 2R_1^2) a_5 = 0, \\ &R_2 a_0 + 3R_0 R_1 a_3 + (R_0^2 + 2R_1^2) a_4 + 3R_0 R_1 a_5 = 0, \\ &R_0 a_0 + (R_0^2 + 2R_0^2) a_3 + 3R_0 R_1 a_4 + 3R_0^2 a_5 = 0, \end{aligned} \tag{27}$$

and two inhomogeneous equations for coefficients a_1 and a_2,

$$\begin{aligned} R_0 a_1 + R_1 a_2 &= R_1, \\ R_1 a_1 + R_0 a_2 &= R_2. \end{aligned} \tag{28}$$

The determinant of homogeneous system (27) is nonzero, when implying $a_0 = a_3 = a_4 = a_5 = 0$. At the same time it follows from Eq. (28) that the coefficients a_1 and a_2 at linear terms in Eq. (26) are exactly the same as for LAR model (10).

Thus, adding quadratic terms to the LAR model yields no advantage, so that the predictability time given by a NAR model satisfies relation (2), similar to relation (1) for LAR model. This conclusion remains essentially valid for autoregression models of the higher order ($M > 2$) and power ($N > 2$) and is corroborated numerically for various random processes. In all these cases, the NAR models are not more successful than LAR ones, regardless of the model order M and power N: the predictability time $\tau_{\text{pred}}^{\text{NAR}}$ for processes of class (i) always obeys estimate (2). Thus, the NAR models only brings about an increase of the computation time without actually improving the prediction quality.

5.2. Discrete one-dimensional maps

Let us now apply NAR models to time series generated by nonlinear one-dimensional maps. Let $y(\nu)$ be a sequence generated by the map $G(x)$,

$$y(\nu + 1) = G\big[y(\nu)\big] + f(\nu), \tag{29}$$

where the discrete time ν takes integer values 1, 2, 3, ... For generality we include in Eq. (29) the fluctuation term $f(\nu)$, which in fact determines a horizon of predictability. Expanding the function $G(\nu)$ in Taylor series and truncating at the Nth power of y, we get

$$G\big[y(\nu)\big] = g_0 + g_1 y(\nu) + g_2 y^2(\nu) + \ldots + g_N y^N(\nu). \tag{30}$$

We try to describe a sequence generated by such a map with the help of NAR model of the first-order ($M = 1$) and Nth power. For a discrete time NAR algorithm (11) of the first order and of Nth power takes on a form

$$z(\nu + 1) = a_0 + a_1 y(\nu) + \ldots + a_N y^N(\nu). \tag{31}$$

Minimizing error functional (24), one can readily ensure that in the absence of fluctuations, the model coefficients a_n coincide with those of expansion (30),

$$a_0 = g_0, \quad a_n = g_n, \quad n = 1, 2, \ldots, N. \tag{32}$$

This implies that fitting NAR model (31) to the time series $y(\nu)$ generated by noiseless map $y(\nu+1) = G[y(\nu)]$ allows us to reconstruct this map (at least to the order N). In the presence of noise ($f \neq 0$), equalities (32) become approximate. It is evident that the differences $\delta a_n = a_n - g_n$, characterizing the accuracy of the map reconstruction, depend on the noise level σ_f.

The role of the noise $f(\nu)$ can be illustrated by the example of a logistic map

$$G(x) = ry(1 - y), \tag{33}$$

for which $g_0 = 0$, $g_1 = r$, and $g_2 = -r$. NAR model (31) is chosen in the form of polynomial of the fifth power ($N = 5$). The minimization of error functional (24) yields the coefficients a_n which are close to the coefficients of original map (33), that is $a_0 \simeq 0$, $a_1 \simeq r$, $a_2 \simeq -r$, $a_3 \simeq a_4 \simeq a_5 \simeq 0$ in (31). The difference between actual coefficients g_n and the reconstructed a_n is usually of the order σ_f.

2	–3.821	–3.8201	–3.9459
3	0	0.0019	0.2380
4	0	0.0019	–0.2108
5	0	–0.0007	0.0703

Figure 6. Degree of predictability $D(\tau)$ based on NAR model (31) for logistic map (33) with noise intensities corresponding to $s \geq 2, 3, 5, 7$, and 10. For reference the autocorrelation function $r(\tau)$ is shown as characterized by rather short correlation time $\tau_c \simeq 1$.

After constructing nonlinear map (31) with a_n close to g_n, we can calculate the degree of predictability

$$D(\mu) = \frac{\langle y(\nu)\, z(\nu)\rangle}{\left[\langle y^2(\nu)\rangle \langle z^2(\nu)\rangle\right]^{1/2}}, \quad \mu = \nu - \nu^0, \quad (34)$$

of the process $y(\nu)$ based on (31). Five curves in Fig. 6 show degree of predictability (34) for different noise intensities σ_f,

$$\sigma_f = \frac{1}{2\sqrt{3}} \cdot 10^{-s}, \quad s = 2, 3, 5, 7, 10. \quad (35)$$

Fluctuations $f(\nu)$ are assumed to be uniformly distributed in the interval $(-10^{-s}, 10^{s})$.

The calculations are performed for the control parameter $r = 3.821$ which, according to [10], corresponds to a regime of developed chaos. The optimum parameters a_k are found by minimizing functional (24) for 500 realizations of the process. The values $a_0, a_1, \ldots, a_5$ are rather close to the coefficients $g_0, g_1, \ldots, g_5$ of original polynomial (30). Indeed, for $r = 3.821$ and for two noise levels $\sigma_f = 10^{-5}/2\sqrt{3}$ and $\sigma_f = 10^{-3}/2\sqrt{3}$ the coefficients g_k and a_k are as shown in Table 1.

Table 1. Initial coefficients g_i and restored coefficients a_j for the logistic map.

i	g_i	a_i ($s = 5$)	a_i ($s = 3$)
0	0	0	–0.0025
1	3.821	3.8208	3.8508

According to Fig. 6, the greater the noise intensity σ_f^2, the shorter the predictability time $\mu_{\text{pred}}^{\text{NAR}}$. The values $\mu_{\text{pred}}^{\text{NAR}}$ are in good agreement with horizon of predictability (3). The curves of Fig. 6 are reminiscent of the results of a similar calculation, performed in [3] for the map $G(y) = 2y \pmod 1$ (see also [7, 8]).

For reference the normalized autocorrelation function $r(\mu)$ is shown in Fig. 6, which displays very rapid decrease with a characteristic time of the order of unity, $\mu_c \simeq 1$. According to the figure, the predictability time may considerably exceed the correlation time, $\mu_{\text{pred}}^{\text{NAR}} >> \mu_c$. This means that the behavior of discrete nonlinear systems like (29) can be predicted for much longer times than the correlation time which typically is of the order of unity.

Next we determine how the autoregression power N affects the predictability of a noisy logistic map. The results of numerical modeling of the degree of predictability $D(\mu)$ are brought in the form of cubic spline in Fig. 7*a* for the noise with rms $\sigma_f = 10^{-5}/2\sqrt{3}$, and in Fig. 7*b* for the greater $\sigma_f = 10^{-3}/2\sqrt{3}$. At $N = 1$, the NAR model coincides with the LAR model of the same order. The degree of predictability $D_1(\mu)$ in Fig. 7*a* practically coincides with the autocorrelation function $r(\mu)$, so that the predictability time coincides with the correlation time, which is close to unity, $\mu_{\text{pred}}(N = 1) = \mu_c \sim 1$. Thus, on average the predictability power of the first power autoregression model ($N = 1$) does not extend beyond one step of the map. The same is true also for the larger noise with $\sigma_f = 10^{-3}/2\sqrt{3}$. The curves $D_1(\mu)$ and $r(\mu)$ practically merge together in Fig. 7*b* like in Fig. 7*a*.

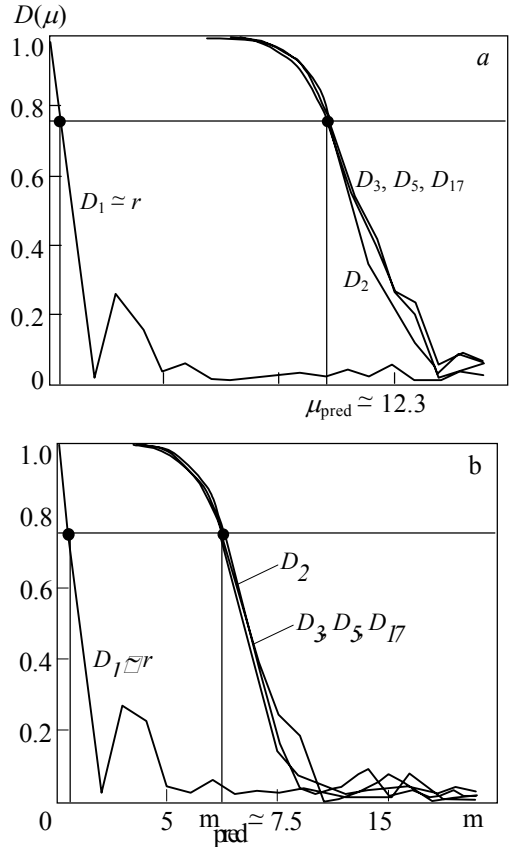

Figure 7. Degree of predictability $D_N(\tau)$ for logistic map (33) with two noise intensities (*a*) $s = 5$, (*b*) $s = 3$ and for different powers N of NAR model (31). Curves $D_{1,2,3,5,17}(\tau)$ correspond to orders $N = 1, 2, 3, 5, 17$. The curve $D_1(\tau)$ practically coincides with the normalized correlation function $r(\tau)$, whereas curves $D_{1,2,3,5,17}(\tau)$ in (*a*) and (*b*) illustrate the saturation of the forecasting power at $N \geq 3$.

While the NAR model power increases to $N = 2$ (curves $D_2(\mu)$ in Fig. 7) the predictability rises sharply to $\mu_{\text{pred}}^{\text{NAR}} = 12.3$ for $s = 5$ (Fig. 7*a*) and to $\mu_{\text{pred}}^{\text{NAR}} = 7.5$ for $s = 3$ (Fig. 7*b*). Both these values are close to the corresponding horizons of predictability (3). Indeed, taking into account that the amplitude A is approximately 0.8 and the Lyapunov index λ_+ is of the order 0.6, we conclude, that the predictability horizon is estimated as $\mu_{\text{lim}} \approx 19$ for $s = 5$ and as $\mu_{\text{lim}} \approx 11$ for $s = 3$.

Increasing further the polynomial power N in (31) saturates the degree of predictability: curves D_2, D_3, D_5, and D_{17} in Figs. 7*a* and *b* are quite close to each other. One can even notice a certain deterioration of the predictability at large N. This can be explained by the fact that the additional terms with the powers $N > 2$ (in the case of logistic map (33)) become an unnecessary burden to worsen the accuracy of recovering the map $G(y)$ and increasing the computation time.

Similar results have been found also for the cubic map

$$G\left[y(\nu)\right] = y^3(\nu) - 2.83y(\nu) - 0.01. \qquad (36)$$

The results of numerical modeling of the degree of predictability $D(\mu)$ with the noise rms $\sigma_f = 10^{-3}/2\sqrt{3}$, are presented in Fig. 8. The curve D_1 shows the predictability for the first-order LAR model. This curve is close to the autocorrelation function $r(\mu)$: the predictability and correlation times are practically alike and occur near 0.75, i.e., less than one step. A certain divergence between $D_1(\mu)$ and $r(\mu)$ comes under $\mu \geq 2$. In contrast to the square-law logistic map, a sudden increase of the predictability time occurs for $N = 3$. When $N \geq 3$ the degree of predictability saturates: as is seen from Fig. 8, plots of $D(\mu)$ for $N = 3, 5, 7$, and 17 practically overlap. All of them correspond to almost the same time of predictability $\mu_{\text{pred}}^{\text{NAR}} \simeq 8.2$ and are close to its horizon.

5.3. Continuous dynamical processes

Real dynamical processes are seldom controlled by one-dimensional maps which provide predictability times close to predictability horizon (3). In the case of continuous dynamic processes, NAR models turn out to be much less effective. Numerical calculations show that the time of predictability for such processes is comparable to the correlation time and never approaches limiting time of predictability (3). As an illustration we consider a dynamic process generated by the Roessler system

$$\begin{aligned} \frac{dy_1}{dt} &= -y_2 - y_3 + f_1, \\ \frac{dy_2}{dt} &= y_1 + 0.2y_2 + f_2, \\ \frac{dy_3}{dt} &= 0.2 - 10y_3 + y_1y_2 + f_3 \end{aligned} \qquad (37)$$

in the chaotic regime.

First we choose 500 random initial values y_{10}, y_{20}, y_{30} and numerically integrate set (37) for a sufficiently long time with zero noise. As a result we find a set of 500 initial values, sufficiently close to

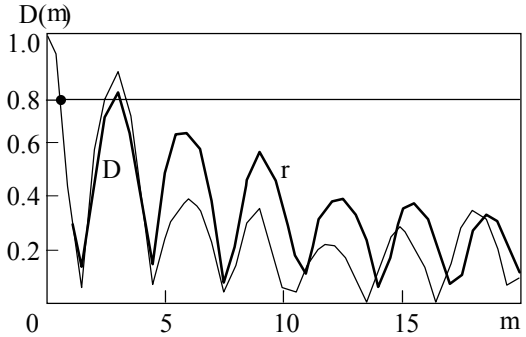

Figure 9. Degree of predictability $D(\tau)$ and normalized correlation function $r(\tau)$ for Roessler system (37). The predictability time is comparable with the correlation time τ_c.

the attractor, for which transient effects were practically eliminated. The initial conditions derived by this procedure are uniformly distributed over the whole area of the Roessler attractor.

NAR models are built for each of the components of this process and calculations are carried out for noise intensities 0.01 % and 0.1 % of the variance of corresponding component. The final value of N is chosen to maximize the degree of predictability. The optimum values of the power N range between 5 and 7. The increase in the degree of predictability with N turns out to be rather modest, by 10–15 %.

The numerical calculations show that for none of components $\{y_1, y_2, y_3\}$ of the process the predictability time exceeds the correlation time by more than a few percent. This is illustrated in Fig. 8, where the correlation function $r(\tau)$ and the degree of predictability $D(\tau)$ are plotted for the y_1 component of a Roessler process with 0.1 % noise. For small τ, the degree of predictability for the polynomial model behaves like the correlation function and their values are practically identical. The time of predictability at the first "half-period" of curve $D(\tau)$ is about 0.7.

On the following half-period, $D(\tau)$ also exceeds the level 0.75, but for definiteness we choose the shorter interval, $\tau \approx 0.7$. Otherwise, the interval of satisfactory prediction will be broken in two parts by the intervals of extremely unsatisfactory forecast at $\tau \approx 2$.

It is important to understand, why the NAR models cannot satisfactorily forecast continuous processes of chaotic type. Qualitatively, discrete NAR models are badly matched with differential equations of chaotic processes. To better illustrate these considerations we turn to the simplest dynamic system of the third order,

$$\begin{aligned} \frac{dy_1}{dt} &= F^{(1)}(y_1, y_2, y_3) + f_1(t), \\ \frac{dy_2}{dt} &= F^{(2)}(y_1, y_2, y_3) + f_2(t), \\ \frac{dy_3}{dt} &= F^{(3)}(y_1, y_2, y_3) + f_3(t), \end{aligned} \qquad (38)$$

which admits a chaotic behavior and where $f_i(t)$, $i = 1, 2, 3$, are random (fluctuation) forces.

To pass from the set of differential equations (38) to a set of difference equations, for instance for the variable $y_1(\tau)$, we integrate (38) from t to $t+\tau$ and get the integral equations

$$y_1(t+\tau) = y_1(t) + \int_t^{t+\tau} F^{(1)}\left[y_1(t'), y_2(t'), y_3(t')\right] dt' + \varphi_1(t), \tag{39}$$

$$y_2(t+\tau) = y_2(t) + \int_t^{t+\tau} F^{(2)}\left[y_1(t'), y_2(t'), y_3(t')\right] dt' + \varphi_2(t), \tag{40}$$

$$y_3(t+\tau) = y_3(t) + \int_t^{t+\tau} F^{(3)}\left[y_1(t'), y_2(t'), y_3(t')\right] dt' + \varphi_3(t). \tag{41}$$

In equations (39)–(41) the values $\varphi_{1,2,3}(t)$ represent the fluctuation forces integrated over the time interval from t to $t+\tau$,

$$\varphi_{1,2,3}(t) = \int_t^{t+\tau} f_{1,2,3}(t')\, dt'. \tag{42}$$

We express the integrals in equations (39)–(41) through the values of variables at preceding moments $t-\tau$, $t-2\tau$, ..., restricting ourselves to the second-order quadrature formula

$$\int_t^{t+\tau} F(t')\, dt' \cong \left[\left(2-\frac{1}{12}\right)F(t) - \frac{4}{3}F(t-\tau) + \frac{5}{12}F(t-2\tau)\right]\tau. \tag{43}$$

Then Eq. (39) becomes

$$y_1(t+\tau) = y_1(t) + \varphi(t) + \tau\left\{\left(2-\frac{1}{2}\right)F^{(1)}\left[y_1(t), y_2(t), y_3(t)\right]\right.$$
$$\left. -\frac{4}{3}F^{(1)}\left[y_1(t+\tau), y_2(t+\tau), y_3(t+\tau)\right] + \frac{5}{12}F^{(1)}\left[y_1(t+2\tau), y_2(t+2\tau), y_3(t+2\tau)\right]\right\}. \tag{44}$$

Equations (40) and (41) yield similar expressions. Equation (43) together with the similar equations for $y_2(t+\tau)$ and $y_3(t+\tau)$ are now playing the role of equations of motion for the delayed variables $y_k(t)$, $y_k(t-\tau)$, $y_k(t-2\tau)$, which are referred to as Takens' variables.

Let us represent the functions $F^{(1,2,3)}$ as polynomials of degree S,

$$F^j(y_1, y_2, y_3) = \sum_{\alpha_1,\alpha_2,\alpha_3} g^{(j)}_{\alpha_1,\alpha_2,\alpha_3} y_1^{\alpha_1} y_2^{\alpha_2} y_3^{\alpha_3}, \tag{45}$$

where $\alpha_{1,2,3} = 0, 1, \ldots$ and the indices' sum $\alpha_1+\alpha_2+\alpha_3$ does not exceed S. Then, Eq. (43) contains the powers of variables y_1, y_2, and y_3 at the moments t, $t-\tau$, and $t-2\tau$. When restricting in Eq. (44) to terms of the second order only, each of three functions $F^{(j)}$ has nine terms,

$$F^{(j)} = g^{(j)}_{100}\, y_1 + g^{(j)}_{010}\, y_2 + g^{(j)}_{001}\, y_2 + g^{(j)}_{200}\, y_1^2 + g^{(j)}_{020}\, y_2^2 + g^{(j)}_{002}\, y_3^2 + g^{(j)}_{110}\, y_1 y_2 + g^{(j)}_{101}\, y_1 y_3 + g^{(j)}_{011}\, y_2 y_3\,. \tag{46}$$

In total, set (39)–(41) will contain 27 factors.

Using the equations for $y_2(t+\tau)$ and $y_3(t+\tau)$, which are similar to Eq. (43), we express the variables y_2 and y_3 in Eq. (43) through the values of all variables $\{y_1, y_2, y_3\}$ at the preceding moments of time.

Repeating this procedure P times, we shift the variables y_2 and y_3 in time for P steps. After that we can single out in Eq. (43) the terms with delays $m_1, \ldots, m_k$, which do not exceed P,

$$y_1(t+\tau) = y_1(t) + \sum_{\beta,m} \gamma_{\beta_1,\ldots,\beta_k}^{m_1,\ldots,m_k} (t-m_1\tau)\ldots y^{\beta_k}(t-m_k\tau) + Q_P\left\{y_1, y_2, y_3, \varphi^{(j)}\right\}, \quad (47)$$

whereas the terms with larger delays and all the fluctuative terms $\varphi^{(j)}$ are included in the function Q_P.

In spite of the formal similarity, the right-hand parts of Eq. (47) and NAR model (23) are essentially different. The main difference consists of the following: initial systems (37) of three first-order differential equations is equivalent to the third-order differential equation. On the other hand, expression (47) preserves only the information on the first derivative in the form of the first-order difference $y_1(t+\tau)-y_1(t)$, while the differences of the second and third order are absent in this equation, at least in the explicit form. As a result, part of the dynamic structure of system (38) is already lost after the first iteration, since an important information is relegated to the term Q_P, which contains an infinite number of components. By discarding this term, we limit the order of autoregression, thereby losing information on the system dynamics. In other words, fitting decomposition (47) (without Q_P) to experimental time series rather worsens the quality of approximation due to local instability of chaotic systems.

Thus, NAR model (23) reflects nonlinear properties of multidimensional dynamical systems in a very truncated form. In fact, the NAR model works within the framework of first-order LAR model only. This explains the very short interval of prediction τ_{pred}, comparable with τ_c.

The inability of NAR models to reveal the dynamics of multidimensional nonlinear systems does not mean that such dynamics cannot be brought to light by other methods. The issue is related to an inverse problem in nonlinear dynamics, namely to the reconstruction of dynamical equations from the experimental time series. The essence of the reconstruction method consists in assuming the structure of nonlinear functions (say, polynomial) entering the differential equations and then determining the coefficients of polynomials from the best fit to the experimental data. This procedure was analyzed extensively in [11–14].

In fact, this reconstruction procedure is nothing but a version of nonlinear autoregression analysis [1, 15]. The difference lies in the fact that the reconstruction implies determination of the polynomial coefficients in governing equations, whereas the standard autoregression deals with the coefficients of polynomial approximation (23) of the signal itself. It comes then as no surprise that reconstruction methods have a stronger prognosis power. In particular, these are able to reach horizon of predictability (3), although to do so they require more complex and time-consuming algorithms.

Of course, as the dimensionality of the system increases, the efficiency of reconstruction algorithms decreases accordingly due to numerical calculations of high-order derivatives from noisy time series. Under conditions of multidimensional noisy processes the advantages of the method for recovering dynamic equations against the standard NAR procedure are not too evident.

6. Discussion

We have confirmed numerically that the predictability time for nonlinear processes, derived from LAR models, cannot significantly exceed the correlation time of these processes.

We also have shown that the predictability of NAR models depends strongly on the nature of underlying process. For random processes, i.e., processes of a nondynamical origin, the time of forecast based on NAR models does not exceed that based on LAR models and is comparable with a correlation time in the observed process. NAR models apply well to time sequences generated by one-dimensional maps. For such sequences the predictability time may approach predictability horizon (3). We also have noticed a saturation effect of the predictability as the power of NAR model increases.

For continuous dynamical processes, the efficiency of NAR models does not exceed the efficiency of LAR ones. The time of predictability for LAR and NAR models is comparable with the correlation time. These models are considerably less powerful than reconstruction methods which ensure predictability times close to limit (3).

These results show that the only class of processes where the NAR models perform better than the LAR ones are the discrete maps. For all other signals, the efficiency of NAR models is comparable to that of first-order LAR models, i.e., the predictability time computed by former models does not exceed the correlation time.

O.L.A., O.Ya.B., and Yu.A.K. would like to thank INTAS for partial support of this work under Grant 96-0305 and the Russian Ministry of Higher Education for partial support under the Russian Federal Program "Integration" (Project A-0029/98). V.A.P. acknowledges partial support from the U.S. Department of Energy, Office of Basic Energy Sciences, under Contract AC05-96OR22464 with the Lockheed Martin Energy Research Corporation.

References

1. Pankratz A. Forecasting with Dynamic Regression Models. N.Y.: Wiley, 1991.
2. Anosov O.L., Butkovskii O.Ya., and Kravtsov Ya.A. *J. Comm. Technol. Electron* 1995, **40**(16), 61.
3. Ljung L. System Identification: Theory for the User. Prentice-Hall, Englewood Cliffs, 1987.
4. Kravtsov Yu.A. *Sov. Phys.-Usp*. 1989, **32**(5), 434.
5. Lloyd E. and Ledermann W. Handbook of Applicable Mathematics. Vol. 2. N.Y.: Wiley, 1984.
6. Kravtsov Yu.A. and Etkin V.S. *Radiophys. Quantum Electron*. 1981, **24**(8).
7. Ilkova L.S., Kravtsov Yu.A., et al. *Ibid*. 1985, **28**(7).
8. Kravtsov Yu.A. In: Limits of Predictability. Ed. Yu.A. Kravtsov. Berlin: Springer, 1993, p. 173.
9. Rytov S.M., Kravtsov Yu.A., and Tatarskii V.I. Principles of Statistical Radiophysics, Vol. 2: Correlation Theory of Random Processes. Berlin, Heidelberg: Springer, 1987.
10. Schuster H.G. Deterministic Chaos, an Introduction. Weinheim: Physik-Verlag GmbH, 1984.
11. Anosov O.L., Butkovskii O.Ya., and Kravtsov Yu.A. In: Predictability of Complex Dynamical System. Eds. Yu.A. Kravtsov and J.B. Kadtke. Berlin: Springer, 1996, p. 105.
12. Mees A.I. and Judd K. *Ibid*., p. 123.
13. Predictability of Complex Dynamical System. Eds. Yu.A. Kravtsov, J.B. Kadtke. Berlin: Springer, 1996.
14. Anosov O.L., Butkovskii O.Ya., and Kravtsov Yu.A. *Math. Models Methods Appl. Sci*. 1997, **7**(1), 49.
15. Gershenfeld N.A. and Weigend A.S. Pattern Formation in the Physical and Biological Sciences. Eds. H.F. Nijhout, L. Nadel, and D.L. Stein. Reading, MA: Addison–Wesley, 1997.