Contents

Handbook of Time Series Analysis. Björn Schelter, Matthias Winterhalder, Jens Timmer

ISBN: ??????????????

8 Nonlinear Dynamical Models from Chaotic Time Series: Methods and Applications

Dmitry A. Smirnov and B. P. Bezruchko

The construction of mathematical models from experimental data is a topical field in mathematical statistics and nonlinear dynamics. It has a long history and still attracts increasing attention. We briefly discuss key problems in nonlinear modeling for typical problem settings ("white," "gray," and "black boxes") and illustrate several contemporary approaches to their solution with simple examples. Finally, we describe a technique for the determination of weak directional coupling between oscillatory systems from short time series based on empirical modeling of their phase dynamics and present its applications to climatic and neurophysiological data.

Author: Please check your address carefully. Please provide complete address (full Christian and family name, institutional affiliation, correct address incl. zip-code, and e-mail) to be inserted into the List of Contributors at the beginning of the book.

Author: Please provide short title for the running head

8.1 Introduction

Ubiquitous use of analog-to-digital converters and fast development of computing power have stimulated considerable interest in methods for modeling discrete sequences of experimental data. The construction of mathematical models from "the first principles" is not always possible. In practice, available information about an object dynamics is often represented in the form of experimental measurements of a scalar or vector quantity η, which is called "observable," at discrete time instants. Such a data set is called "a time series" and denoted by $\{\eta_i\}_{i=1}^{N} \equiv \{\eta_1, \eta_2, \ldots, \eta_N\}$ where $\eta_i = \eta(t_i)$, $t_i = i\Delta t$, Δt is a sampling interval, N is a time series length. Modeling from experimental time series is known as "system identification" in mathematical statistics and automatic control theory [1] or "reconstruction of dynamical systems" in nonlinear dynamics [2].

Dynamical systems' reconstruction has its roots in the problems of *approximation* and *statistical investigation* of dependences. Initially, observed processes were modeled as explicit functions of time which approximated experimental dependences on the plane (t, η). The purpose of modeling was either predicting the future evolution (via extrapolation) or smoothing the data. A significant advance in empirical modeling of complex processes was achieved in the beginning of the twentieth century when *linear stochastic* autoregressive models were introduced [3]. It gave an origin to ARIMA models technology which became

Handbook of Time Series Analysis. Björn Schelter, Matthias Winterhalder, Jens Timmer

ISBN: ??????????????

a predominant approach for half a century (1920s–1970s) and found numerous applications, especially in automatic control [1, 4]. Subsequently, birth of the concept of "deterministic chaos" and fast progress of computational power led to the appearance of a different framework. Currently, empirical modeling is often performed with the use of *nonlinear* difference and differential equations, see pioneering works [5–10]. Such empirical models are demanded in many fields of science and practice such as physics, meteorology, seismology, economy, biomedicine, etc. [11].

In this chapter a brief overview of the problems and techniques for the construction of dynamical models from noisy chaotic time series is given. It supplements existing surveys [12–18] due to the use of a special systematization of the variety of problem settings and methods. Also, we try to provide a clear explanation of the key points with simple examples and illustrate some specific problems with our own results. For the most part, we examine finite-dimensional models in the form of difference equations (maps)

$$\mathbf{x}_{n+1} = \mathbf{f}(\mathbf{x}_n, \mathbf{c}) \tag{8.1}$$

or ordinary differential equations (ODEs)

$$d\mathbf{x}/dt = \mathbf{f}(\mathbf{x}, \mathbf{c}), \tag{8.2}$$

where $\mathbf{x}$ is a D-dimensional state vector, $\mathbf{f}$ is a vector-valued function, $\mathbf{c}$ is a P-dimensional parameter vector, n is the discrete time, and t is the continuous time.

We expose the problems "from simple to complex," as the amount of *a priori* information about an object decreases. We start from a situation where only concrete values of model parameters are to be found ("transparent box" or "white box," Section 8.3). Then, we go via the case where a model structure is partly known ("gray box," Section 8.4) to the case of no *a priori* information ("black box," Section 8.5). Throughout the chapter, we refer to a unified scheme of the empirical modeling process outlined in Section 8.2. Some applications of empirical modeling, in particular, to climatic and neurophysiological data are described in Section 8.6.

8.2 Scheme of the Modeling Process

Despite an infinite number of specific situations, objects, and purposes of modeling, one can single out basic stages of the modeling process and present them using a scheme shown in Fig. 8.1 which generalizes similar schemes given in [1, 4]. It starts with the consideration of available *a priori* information about an object under investigation, formulation of the goals of modeling, acquisition and preliminary analysis of experimental data (stage 1). It ends with a desired application of a constructed model. However, the modeling process typically involves multiple reiterations and a step-by-step approach to a "good" model.

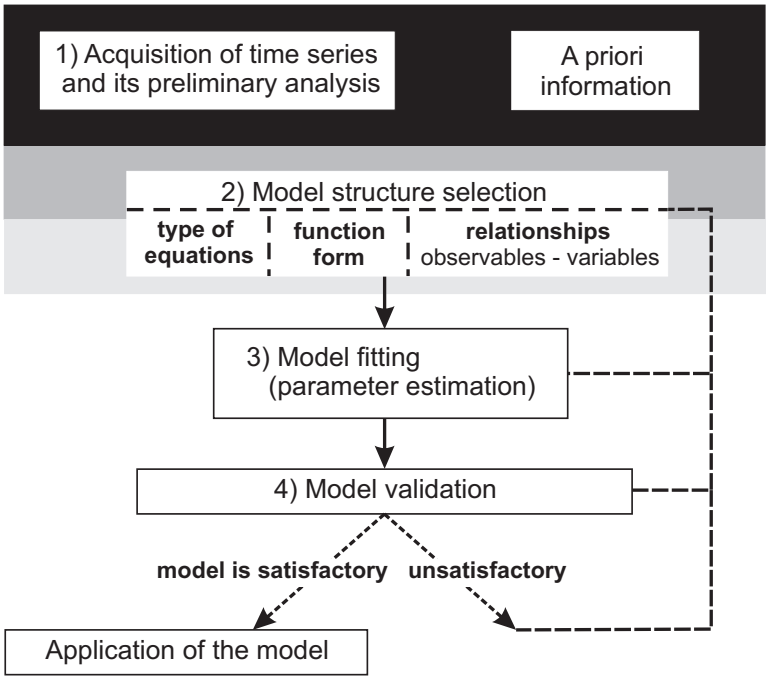

Fig. 8.1: A general scheme of the process of modeling from time series.

At the second stage, a model structure is specified. One chooses the type and number of model equations, the form of functions entering their right-hand sides (components of the function $\mathbf{f}$), and dynamical variables (components of the vector $\mathbf{x}$). As for the latter, one can use just the observable quantities as model variables, but in general the relationship among the observables and dynamical variables may be specified separately. Usually, it takes the form $\eta = h(\mathbf{x})$, where h is called "measurement function." Moreover, the observable values may be corrupted with noise. Stage 2 is often called "structural identification."

At the third stage, the values of the model parameters $\mathbf{c}$ are to be determined. One often speaks of *parameter estimation* or *model fitting*. In the theory of system identification this is a stage of "parametric or nonparametric identification." To perform the estimation, one usually looks for a global extremum of an appropriate *cost function*. For example, the sum of squared deviations of a model time realization from the observed data is often minimized.

Finally, the quality of a model is checked, as a rule, based on a specially reserved test part of a time series. In respect of the final goal of modeling, one can distinguish between two settings: "cognitive identification" (the goal is to obtain an adequate model and to understand better the object behavior) and "practical identification" (a practical goal is to be achieved with the aid of the model, e.g., a forecast). Depending on the setting, one checks either model *adequacy* in respect

of some properties (this step is also called model validation or verification) or model *efficiency* in respect of the practical goal. If a model is found satisfactory (adequate or efficient) then it may be exploited. Otherwise, one must return to one of the previous stages of the scheme.

The background colors in (Fig. 8.1) change from black to white reflecting the degree of *a priori* uncertainty. The worst situation is called "black box" problem: information about an appropriate model structure is completely lacking and one must start the modeling process from the very top of the scheme. The more information about a possible model structure is available, the more probable is the success of modeling: the "box" becomes "gray" and even "transparent" ("white"). In any case, one cannot avoid the stage of parameter estimation. Therefore, we start our consideration with the simplest situation when one knows everything about an object, except for the concrete values of the model parameters. It corresponds to white background color in Fig. 8.1.

8.3 "White Box" Problems

If a model structure is completely known, the problem reduces to the estimation of model parameters $\mathbf{c}$ from the observed data. Such a setting is encountered in different applications and, therefore, attracts considerable attention. There are two basic tasks:

1. to obtain parameter estimates with a desired accuracy; this is especially important if the parameters cannot be measured directly under the conditions of experiment, i.e., the modeling procedure acts as "a measurement device" [19–24];
2. to obtain reasonable parameter estimates when time courses of some model state variables x_k can neither be measured directly nor calculated from the available time series of the observable η, i.e., some model variables are "hidden" [25, 26].

Let us discuss both points in turn.

8.3.1 Parameter Estimates and Their Accuracy

As a basic test example, we consider parameter estimation in a nonlinear map from its time series. The object is a quadratic map in a chaotic regime

$$x_{n+1} = f(x_n, c) + \xi_n, \quad \eta_n = x_n + \zeta_n\,, \tag{8.3}$$

where $f(x_n, c) = 1 - cx_n^2$, the only parameter c is considered unknown, ξ_n, ζ_n are random processes. The process ξ_n is called "dynamical noise" since it affects the evolution of the system, while ζ_n is referred to as "measurement noise" since it corrupts only the observations. In the absence of any noise, one has $\eta_n = x_n$ so that all experimental data points on the plane (η_n, η_{n+1}) lie exactly on

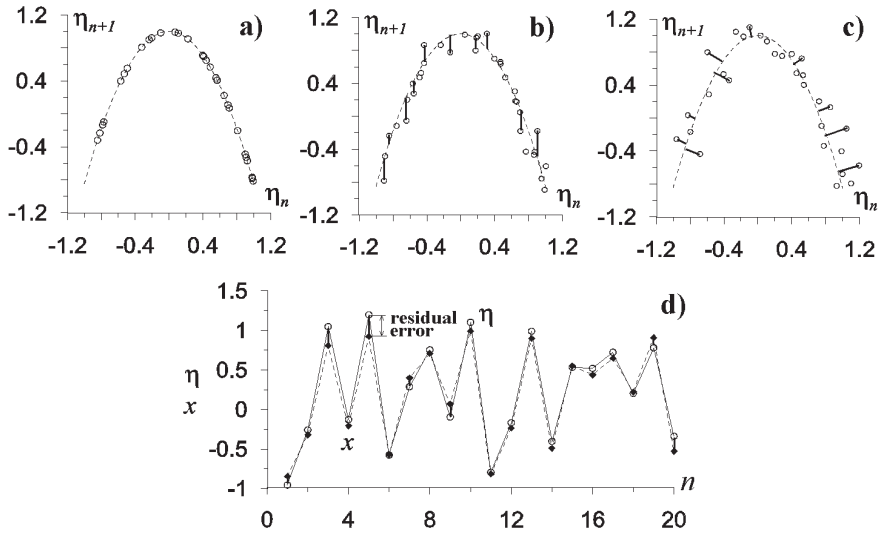

Fig. 8.2: Parameter estimation in the quadratic map (8.3); the true value is c = 1.85. Open circles denote observed data. (a) Noise-free case, the dashed line is an original parabola. (b) Uniformly distributed dynamical noise. The dashed line is a model parabola obtained via minimization of the vertical distances. (c) Gaussian measurement noise. The dashed line is a model parabola obtained via minimization of the orthogonal distances. (d) Gaussian measurement noise. Rhombs indicate a model time realization which is the closest one to the observed data in the least-squares sense.

the quadratic parabola (Fig. 8.2(a)). The value of c can be determined from an algebraic equation whose solution takes the form $\hat{c} = (1-\eta_{n+1})/\eta_n^2$ (throughout the paper, a "hat" denotes quantities calculated from a time series). It is sufficient to use any pair of successive observed values with $\eta_n \neq 0$. As a result, the model is practically ideal.

In the presence of any noise, one must speak of statistical estimates instead of precise calculation of the parameter value. There are various estimation techniques [27]. Below, we describe several of them, which are most widespread.

Maximum Likelihood Approach

The maximum likelihood (ML) approach is the most efficient under

quite general conditions [27]. It is most often announced as a method of choice. However, additional assumptions about the properties of an object and noise are typically accepted in practice reducing the ML approach to a version of the least-squares (LS) technique.

Let us start with the simplest situation when only dynamical noise is present in the system, Eq. (8.3). Let ξ_n be a sequence of independent identically dis-

tributed random values whose one-dimensional probability density function is $p_\xi(z)$. Then, an ML estimate is such a value of c which maximizes logarithmic likelihood function

$$\ln L(c) \equiv \ln p(\eta_1, \ldots, \eta_N|c) \approx \sum_{n=1}^{N-1} \ln p_\xi\big(\eta_{n+1} - f(\eta_n, c)\big)\,, \tag{8.4}$$

which is, roughly speaking, a logarithm of a conditional probability to observe the available time series $\{\eta_1, \ldots, \eta_N\}$ at a given c. To apply the ML method, one needs to know the distribution law $p_\xi(z)$ *a priori*. This is rarely the case, therefore, Gaussian distribution is often assumed. It is not always the best idea but it is reasonable both from theoretical (central limit theorem) and practical (successful results) points of view.

Dynamical Noise: Ordinary Least-Squares Technique

For Gaussian noise, the ML estimation, Eq. (8.4), reduces to the "ordinary" LS (OLS) technique. The LS method is the most popular estimation technique due to the relative simplicity of implementation, bulk of available theoretical knowledge about the properties of the LS estimates, and many satisfactory practical results. The OLS technique consists in the minimization of the sum of squared deviations

$$S(c) = \sum_{n=1}^{N-1} \big(\eta_{n+1} - f(\eta_n, c)\big)^2 \to \min\,. \tag{8.5}$$

Geometrically, it means that a curve of a specified functional form is drawn on the plane (η_n, η_{n+1}) in such a way that the sum of squared *vertical distances* from experimental data points to this curve is minimized (Fig. 8.2(b)). The OLS technique often gives acceptable accuracy of the estimates even if noise is not Gaussian, which is justified by the robust estimation theory, see e.g., [28]. Therefore, it is valuable on its own, apart from being a particular case of the ML approach.

A technical problem in the application of the ML and the OLS estimation arises if a "relief" of the cost function to be optimized exhibits multiple local extrema. It may be the case for the problem, Eq. (8.5), if f is nonlinear in parameter c. Then, the optimization problem is solved with the aid of iterative techniques which require a starting guess for the estimated parameter. Whether a global extremum will be found depends typically on the closeness of the starting guess to the true value of the parameter. The function f is linear in c for the example, Eq. (8.3), therefore the cost function S is quadratic in c and has the only minimum which is easily found via the solution of a linear algebraic equation. Such a simplicity of the LS problem solution is a reason for the widespread use of the models which are linear in parameters, the so-called *pseudo-linear* models, see also Section 8.5.

The error in the estimate $\hat{c}$ decreases with the time series length. Namely, for the dynamical noise case, both ML and OLS techniques give asymptotically

unbiased and consistent estimates, i.e., error in the estimate vanishes as $N \to \infty$. Moreover, it can be shown that the variance of the estimates decreases as N^{-1} [27, 28].

Measurement Noise: Monotonically the Total Least-Squares Technique and Others

If only measurement noise is present, the estimation problem becomes more difficult. The OLS technique, Eq. (8.5), provides biased estimates for arbitrary long time series, since it is developed under the assumption of the dynamical noise. However, it is simple in implementation and still may be used sometimes to get a crude approximation. Roughly speaking, if the measurement noise level is not high, namely up to 1%, then the OLS estimates are reasonably good [20]. Throughout the chapter, we define the noise level as the ratio of the noise root-mean-squared value to the signal root-mean-squared value.

At a higher noise level, to enhance the accuracy of the estimates is partly possible with the aid of the total LS (TLS) method [19] where the sum of squared *orthogonal distances* is minimized, see Fig. 8.2(c). But this is only a partial solution since the bias in the estimates is not completely eliminated. A more radical approach is to write the "honest" likelihood function taking into account the effect of measurement noise. To accomplish it, one must include an initial condition of a model map into the set of estimated quantities. Thus, for Gaussian measurement noise the problem reduces to a version of the LS technique where a model *time realization* is made as close to *the observed time series* as possible (Fig. 8.2(d))

$$S(c, x_1) = \sum_{n=0}^{N-1} \left(\eta_{n+1} - f^{(n)}(x_1, c)\right)^2 \to \min\,, \tag{8.6}$$

where $f^{(n)}$ stands for the nth iteration of the map $x_{n+1} = f(x_n, c)$, $f^{(0)}(x, c) \equiv x$.

As an orbit of a chaotic system is highly sensitive to initial conditions and parameters, the variance of such an estimate decreases very quickly with time series length N, even exponentially for specific examples [22, 23]. But it holds true only if a global minimum of the cost function, Eq. (8.6), is guaranteed to be found. However, the graph of the cost function S becomes so "jagged" for a large N that it appears practically impossible to find its global minimum (see Fig. 8.3(a)) because it would require unrealistically lucky starting guesses for c and x_1. It is also difficult to speak of the asymptotic properties of such estimates since the cost function, Eq. (8.6), is no longer smooth in the limit $N \to \infty$. Therefore, modifications of the direct ML approach have been developed for this problem setting [20, 21, 23, 24].

In particular, it was suggested to divide an original time series into segments of moderate length L, minimize Eq. (8.6) for each segment separately, and average the segment estimates (a piecewise approach). This is a practically reasonable technique but the resulting estimate may remain asymptotically biased. Its vari-

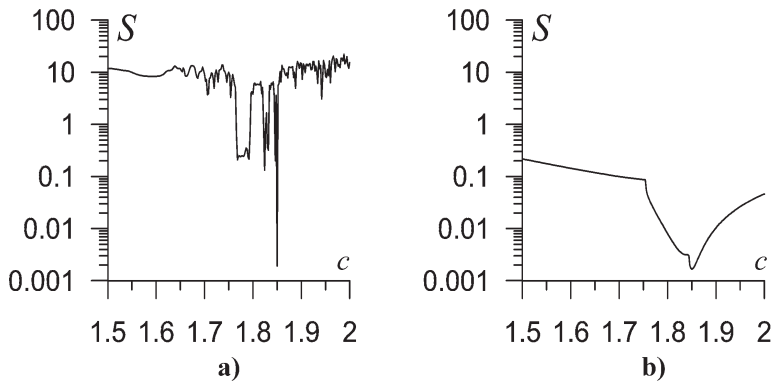

Fig. 8.3: Cost functions for the example of the quadratic map (8.3) at $N = 20$ and true values $c = 1.85$, $x_1 = 0.3$. (a) for the forward iteration approach, Eq. (8.6), (b) for the backward iterations, Eq. (8.7). Trial values of x_1 and x_N are kept equal to their true values for illustration purposes.

ance decreases again only as N^{-1}. Several tricks to enhance the accuracy of the estimates are described below (Section 8.3.2). Here, we would like to note a specific version of the LS technique suggested in [24] for one-dimensional maps. It relies upon the property that the only Lyapunov exponent of a one-dimensional map becomes negative under the time reversal so that a "reverse-time" orbit is no longer highly sensitive to parameters and an "initial" condition. Therefore, one minimizes

$$S(c, x_N) = \sum_{n=0}^{N-1} \left(\eta_{N-n} - f^{(-n)}(x_N, c)\right)^2 \to \min\,, \tag{8.7}$$

where $f^{(-n)}$ is the nth backward iteration of the map. The graph of this cost function looks rather smooth and gradually changing (as in Fig. 8.3(b)) even for arbitrary long time series so that its global minimum can be readily found. At low and moderate noise levels (up to 5–15 %), the error in the estimates obtained via Eq. (8.7) turns out less than for the piecewise approach. Moreover, for sufficiently low noise levels the backward iteration technique gives asymptotically unbiased estimates whose variance decreases generically as N^{-2}. The latter property is determined by close returns of the map orbit to an arbitrary small vicinity of the extrema of the function f [24].

8.3.2 Hidden Variables

If the measurement noise level is considerable, the state variable x can be treated as "hidden" since its true values are not known. But even "more hidden" are those variables whose values can neither be measured directly nor calculated

from the observed time series. The latter case is encountered in practice very often. To estimate model parameters is much more problematic in such a situation than for the settings considered in Section 8.3.1. However, if one succeeds, there appears a possibility of getting time courses of the hidden variables as a by-product of the estimation procedure. Hence, a modeling procedure acts as a measurement device in respect of dynamical variables.

Let us briefly mention available techniques. To a significant extent, all of them rely on the idea, Eq. (8.6), i.e., one looks for initial conditions and parameters of a model which provide the least deviation of a model time realization from the observed data. The naive solution of the problem, Eq. (8.6), directly is called "initial value approach" [18]. As we already mentioned, such a method is inapplicable already for moderately long chaotic time series, while simple division of the time series into segments decreases the accuracy of the estimates and the backward iterations are not appropriate for multidimensional dissipative systems.

To overcome the difficulties and exploit longer time series (than allowed by the initial value approach) is partly possible with the aid of Bock's algorithm [18, 25]. It is often called "multiple shooting approach" since it replaces the Cauchy problem with a set of boundary-value problems to get a model orbit. Namely, the idea is to divide the time series into shorter segments of the length L and consider "initial conditions" of the model on each of them as additional quantities to be estimated. Optimization problems, Eq. (8.6), are solved for each segment *while keeping* model parameter values $\mathbf{c}$ *the same* for all segments *and* imposing constraints of "sewing the segments together" to finally obtain a model orbit which is continuous over the entire observation period. Thus, the number of free parameters ("independent" estimated quantities) remain the same as in the initial value approach but intermediate trial values for all estimated quantities may pass through a domain which corresponds to a discontinuous model orbit and is, therefore, forbidden for the initial value approach. The latter property provides higher flexibility of Bock's algorithm [25].

The multiple shooting approach softens the demands to the choice of starting guesses for the estimated quantities. However, for a longer time series it can also become inefficient since the requirement of closeness of a chaotic model orbit to the observed time series over the entire observation interval can appear very strict. One can overcome some difficulties if final discontinuity of a model orbit at some fixed time instants within the observation period is allowed. It increases the number of free parameters and, hence, leads to the growth of the variance of their estimates, but simultaneously the probability of finding a global minimum of the cost function increases. Such a modification allows the use of arbitrary long chaotic time series. The undesirable "side effect" is that a model with inadequate structure can sometimes be regarded "good" due to its ability to reproduce only short segments of a time series. Therefore, one must avoid the use of too short continuity segments [18].

We note that there exist and are currently developed several methods for

parameters and hidden variables' estimation which are suitable even for the case of simultaneous presence of dynamical and measurement noise. They are based on the Bayesian approach [29] and Kalman filtering [18, 30]. But that broad field of research is beyond the scope of this chapter.

Model validation for the "white box" problems can be performed via one of the two basic lines: (1) analysis of residual model errors, i.e., checking the agreement among their statistical properties and expected theoretical properties of the noise (typically, Gaussianity and temporal uncorrelatedness) [4]; (2) comparison of dynamical, geometrical, and topological characteristics of a model attractor with the corresponding properties of an object [2].

8.3.3 What Do We Get from Successful and Unsuccessful Modeling Attempts?

Success of the methods described above provides both estimates of model parameters and time courses of hidden variables. It promises exciting applications such as validation of the "physical" ideas underlying a specified model structure, "indirect measurement" of quantities inaccessible for a device of an experimentalist, and restoration of the lost or distorted segments of an observed time realization. However, unsuccessful modeling attempts also give useful information. Let us elaborate.

In practice, one never encounters a purely "white box" problem. A researcher may only have faith that a trial model structure is adequate to an object. Therefore, the result of modeling may well appear negative, i.e., reveal an impossibility to get an adequate model with the specified structure. In such a case, a researcher has to claim falseness of his/her ideas about underlying mechanisms of the investigated process and return to the stage of structural identification.

If there are several alternative model structures, then the results of time series modeling may reveal the most adequate among them. In other words, a modeling procedure provides opportunity to falsify or verify (or, possibly, make more accurate) substantial notions about the dynamics of an object. An impressive example of such a modeling process and substantial conclusions about the mechanism underlying a biochemical signaling process in cells is given in [31]. In a similar way, Horbelt and co-authors validated concepts about a gas laser behavior and reconstructed interdependences among transition rates and pumping current which are difficult to measure directly [32]. However, despite these and some other successful practical attempts, an estimation problem can often appear technically unsolvable: the more hidden variables and unknown parameters involved, the weaker are the chances for the success and the lower is the accuracy of the obtained estimates.

8.4 "Gray Box" Problems

From our point of view, the most promising line of research in the field of dynamical systems' reconstruction is related to the "gray box" problems when one knows a lot about an appropriate model structure except for some components of the function $\mathbf{f}$ in Eqs. (8.1) or (8.2). These components are, in general, nonlinear functions which can often be meaningfully interpreted as *equivalent characteristics* of certain elements of an object under investigation.

One has to choose some *approximating* functions for the characteristics. In this section we focus on the approximation of univariate dependences. Such a case is much simpler than multivariate approximation addressed in Section (8.5). Despite models deduced from physical considerations most often take the form of differential equations, let us consider a model map as the first illustration for the sake of clarity.

8.4.1 Approximation and "Overlearning" Problem

Let the object be a one-dimensional map $x_{n+1} = F(x_n)$. We pretend that the form of the function F is unknown. Let the observable coincide with the dynamical variable x: $\eta_n = x_n$. One has to build a one-dimensional model map $x_{n+1} = f(x_n, \mathbf{c})$. The problem reduces to the selection of a model function $f(x, \mathbf{c})$ and its parameters $\mathbf{c}$ so that it could approximate F to the best possible accuracy. It is the matter of agreement to attribute this problem setting to the "gray box" class. We do so since the knowledge that *one-dimensional* model is appropriate can be considered as an important *a priori* information.

Usually, the OLS technique, Eq. (8.5), is used to calculate parameter values. However, the interpretation of the results differs. Now, one speaks of approximation and its mean-squared error rather than of the estimates and noise. Typically, an individual model parameter is not physically meaningful, only the entire model function $f(x, \hat{\mathbf{c}})$ can make sense as a nonlinear characteristic. A key question is how to choose the form of the model function f.

One may choose it intuitively via looking at the experimental data points on the plane (η_n, η_{n+1}). However, this way is not always possible. Thus, it is practically excluded if an unknown univariate function is only a component of a multidimensional model. A more general and widespread approach is to use a functional basis for approximation. For example, the celebrated Weierstrass theorems state that any continuous function over a finite interval can be uniformly approximated to arbitrary high accuracy with an algebraic polynomial (or a trigonometric polynomial under an additional condition). An algebraic polynomial $f(x, \mathbf{c}) = c_1 + c_2 x + \cdots + c_{K+1} x^K$ is one of the most efficient constructions for approximation of smooth univariate dependences. Therefore, we use it below for illustration.

Theoretically, any smooth function can be accurately approximated with a polynomial of *sufficiently high* order K. What value of the order must be chosen

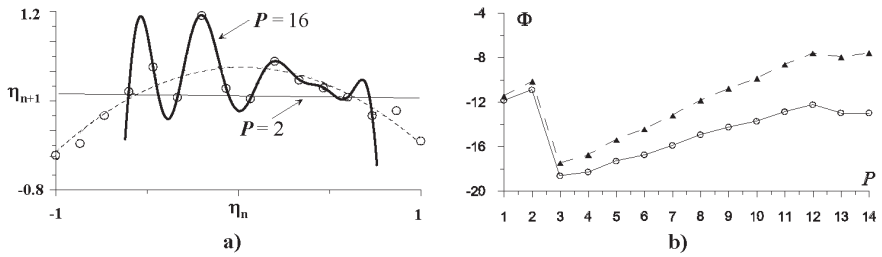

Fig. 8.4: Approximation based on the noisy quadratic map data. (a) Observed data points are shown with circles. Graphs of model polynomials of different orders K are presented. The dashed line for $K = 2$, the thin line for $K = 1$, and the thick line corresponds to $K = 15$. (b) Different cost functions (Eq. (8.8)) versus model size: Circles for the Akaike criterion and triangles for the Schwartz criterion. Both cost functions indicate an optimal model size $P = 3$ corresponding to the true polynomial order $K = 2$.

in practice given a time series of the finite length N, i.e., $N - 1$ data points on the plane (η_n, η_{n+1})? It is a bad idea to specify a very small polynomial order since a model function could not reasonably reproduce an observed nonlinearity (Fig. 8.4(a), the thin line). It is a bad idea to choose very big order as well: e.g., at $K = N - 2$ the graph of the model polynomial on the plane (η_n, η_{n+1}) can pass through all the experimental data points *exactly*, but typically it would extremely badly predict additional (test) observations. In the latter case, the model is said to be *overlearned* or *overtrained* [28]. It does not generalize, rather it just reproduces the observed $N - 1$ data points (Fig. 8.4(a), the thick line).

In practice, one often tries different polynomial orders, starting from a very small one and successively increasing it. One stops when a model gives more or less satisfactory description of an object dynamics and/or the results of approximation saturate. This is a subjective criterion, but it is the only one which is generally applicable, since any "automatic" approach to the order selection is based on a specific well-formalized practical requirement and may not recognize the most adequate model. Such automatic criteria were developed, e.g., in the framework of the information theory. They are obtained from different considerations, but formally reduce to the minimization of a cost function

$$\Phi(P) = (\text{model error}) + (\text{model size}) \to \min . \tag{8.8}$$

Here, the model error rises monotonically with the mean-squared approximation error $\varepsilon^2 = S/(N - 1)$. The model size is an increasing function of the number of model parameters P. Thus, the first term in the sum, Eq. (8.8), may be very large for small polynomial orders, while the second term dominates for big orders. One often observes a minimum of the cost function, Eq. (8.8), for an intermediate K. The minimum corresponds to an optimal model size. The cost function $\Phi(P) = (N/2) \ln \varepsilon^2(\hat{\mathbf{c}}) + P$ is called the Akaike criterion, $\Phi(P) = (N/2) \ln \varepsilon^2(\hat{\mathbf{c}}) + P \ln N/2$

is the Schwartz criterion, and $\Phi(P) = \ln \varepsilon^2(\hat{c}) + P$ is a model entropy [5]. More "cumbersome" is a formula for a cost function named *description length* [33]. Description length minimization is currently the most widely used approach to the model size selection, e.g., [34]. It is based on the ideas of optimal information compression, the Schwartz criterion is an asymptotic expression for the description length. In Fig. 8.4(b) we present an example of a polynomial order selection for approximation of quadratic function from a short time series of the quadratic map, Eq. (8.3), with dynamical noise.

If an approximating function is defined in a closed form for the entire range of the argument (e.g., an algebraic polynomial) then the approximation and the model are called *global* [9]. An alternative approach is a *local* (piecewise) approximation where a model function is defined through a simple formula whose parameters' values differ for different small domains within the range of the argument [7, 9]. The most popular examples of the latter approach are piecewise-constant functions, piecewise-linear functions, and cubic splines. Local models are superior for the description of "complicated" nonlinear dependences (strongly fluctuating dependences, dependences with knees and discontinuities, etc.), but they are less robust to noise influence and require larger amount of data than global models of moderate size.

8.4.2 Model Structure Selection

As a rule, one needs to supplement a procedure for model size selection with a technique to search for an optimal model of a specified size. Thus, according to the technique described above the polynomial order is increased starting from zero and the procedure is stopped at a certain value of K, i.e., the terms are added to a model structure in a predefined order. Therefore, a final model inevitably comprises all power of x up to K, inclusively. However, some of the low-order terms might be "superfluous." Hence, it would be much better to exclude them from the model. Different approaches have been suggested to realize a more flexible way of the model structure selection. They are based either on successive selective complication of a model [34] or its selective simplification starting from the biggest size [16, 35–37], see also [38]. Let us describe briefly a version of the latter strategy [36].

One of the efficient principles to recognize "superfluous" model terms is to look at the behavior of the corresponding coefficient estimates when reconstruction is performed from different segments of a time series, i.e., from the sets of data points occupying different domains in the model state space. Typically, it is realized in the most efficient way of a time series corresponding to a transient process is used. The idea is that the parameter values of an adequate global model of a *dynamically stationary* system must not depend on the reconstruction segment. However, the estimates of parameters corresponding to superfluous terms may exhibit significant changes when a reconstruction segment is moved

along a time series. A procedure for model structure selection can be based on successive removal of the terms whose coefficients are the least stable being estimated from different segments. In [36] the degree of instability of a coefficient is defined as the ratio of its standard deviation to its empirical mean. Removal is stopped, e.g., when model ability to reproduce an object behavior in a wide domain of state space starts to worsen.

8.4.3 Reconstruction of Regularly Driven Systems

In many cases uncertainty in a model structure can be reduced if *a priori* knowledge about object properties is taken into account. We illustrate it with an example of systems under regular (periodic or quasiperiodic) driving. Indication to the presence of external driving can be often seen in the power spectrum which typically exhibits pronounced discrete peaks for regularly driven systems, even though it is neither a necessary nor a sufficient sign. Having the hypothesis about the presence of external regular driving, one can incorporate functions explicitly depending on time into the model structure to describe the assumed driving. For the first time, it was done for nonlinear two-dimensional oscillators under sinusoidal driving in [39]. In the same work, the successful reconstruction of nonlinear dynamical characteristics of a capacitor with ferroelectric was demonstrated.

In a more general setting, the reconstruction of regularly driven systems was considered in [40, 41]. For harmonical additive driving, it is reasonable to construct a model in the form

$$d^D x/dt^D = f(x, dx/dt, \ldots, d^{D-1}x/dt^{D-1}, \boldsymbol{c}) + a\cos\omega t + b\sin\omega t\,, \quad (8.9)$$

where f is an algebraic polynomial and the number of variables D is less than for a corresponding *standard* model by 2 (see Section 8.5 about the standard structure).

In the case of arbitrary additive regular driving (either complex periodic or quasiperiodic one), it is convenient to use the model form

$$d^D x/dt^D = f(x, dx/dt, \ldots, d^{D-1}x/dt^{D-1}, \boldsymbol{c}) + g(t, \boldsymbol{c})\,, \quad (8.10)$$

where the function g describes driving and also depends on unknown parameters. It may take the form of a sum of trigonometric polynomials [41]

$$g(t, \boldsymbol{c}) = \sum_{i=1}^{k} \sum_{j=1}^{K_i} c_{i,j} \cos(j\omega_i t + \varphi_{i,j})\,. \quad (8.11)$$

We note that adequate models with trigonometric polynomials can be obtained even for a very large number of involved harmonics (K_i of the order of hundreds), while the use of a high-order algebraic polynomial K leads typically to model orbits diverging to infinity.

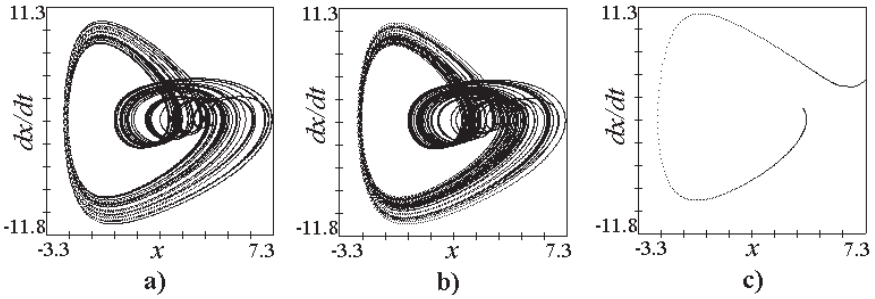

Fig. 8.5: The reconstruction of the driven Toda oscillator $\mathrm{d}^2\mathrm{x}/\,\mathrm{dt}^2 = -0.45\,\mathrm{dx}/\,\mathrm{dt} + (5+4\cos\mathrm{t})(e^{-\mathrm{x}}-1)+7\sin\mathrm{t}$. (a) an attractor of the original system; (b) an attractor of a model of the type, Eq. (8.9), with $\mathrm{D}=2,\mathrm{K}=9$, and sinusoidal dependence of time introduced into all polynomial coefficients, (c) a diverging phase orbit of a standard model, Eqs. (8.13) and (8.14) with $\mathrm{D}=4,\mathrm{K}=6$.

Besides, the explicit time dependence can be introduced into all the coefficients of the algebraic polynomial f to allow the description of not just additive driving [40], Fig. 8.5. Efficiency of all these approaches was shown in numerical experiments with the reconstruction of equations of exemplary oscillators from their noise-corrupted chaotic time series for pulse periodic, periodic with subharmonics, and quasiperiodic driving.

8.5 "Black Box" Problems

If nothing is known about an appropriate model structure, one must appeal to universal constructions. They usually involve huge number of parameters that do not allow the use of majority of the estimation techniques described in Section 8.3. In particular, the hidden variables problem is unsolvable in such a case. Therefore, time series of all dynamical variables must be either measured directly or calculated from the observed data. The latter is called "reconstruction of state vectors." Then, one constructs a multidimensional model of the form, Eq. (8.1) or Eq. (8.2), where the multivariate function $\mathbf{f}$ takes one of the universal forms comprising many parameters. In practice, to estimate these parameters is reasonable with the aid of the OLS technique. To simplify the problem further, it is desirable to choose functions $\mathbf{f}$ which are linear in parameters $\mathbf{c}$ (pseudo-linear models). Considerable efforts of many researchers were devoted to the development of such techniques.

8.5.1 Universal Structures of Model Equations

A theoretical background for different approaches to the reconstruction of model state variables from a scalar observable time realization is the celebrated Takens

theorem [42]. One of them states that for almost any deterministic dynamical system of the form, Eq. (8.1) or (8.2), with a sufficiently smooth function on the right-hand side, its dynamics on an m-dimensional smooth manifold can be topologically equivalently described in terms of vectors constructed as D-plets of successive values of almost any observable $\eta = h(\mathbf{x})$ separated with an almost arbitrary fixed time interval τ. The equivalent description is (almost) guaranteed if dimensionality of these vectors is high enough, namely, $D > 2m$. One says that the original manifold is *embedded* into the new state space which is often called "embedding space." Rigorous formulations, detailed discussions, and generalizations of the theorems can be found in [43–45].

Thus, the vectors $(\eta_n, \eta_{n+\tau}, \ldots, \eta_{n+(D-1)\tau})$, where τ is a time delay, can serve as state vectors. This approach is very popular since it does not involve any transformation of the observed time series. It is usually employed for the construction of model maps in the form

$$\eta_n = f(\eta_{n-\tau}, \ldots, \eta_{n-D\tau}, \mathbf{c}) \,. \tag{8.12}$$

Theoretically, the value of τ may be almost arbitrary. However, in practice it is undesirable to use both very small delays (to avoid strong correlations among the state vector components) and very big ones (to avoid complication of the structure of the reconstructed attractor). Therefore, an optimal choice of τ is possible. There are several recipes such as to take the first zero of the autocorrelation function of the time series [46], the first minimum of the mutual information function [47], etc. [48]. It was also suggested to use a nonuniform embedding where time intervals separating successive components of a state vector are not the same [49, 50]. Finally, a variable embedding is possible where the set of time delays and even dimensionality of a state vector depends on the location in state space [50].

Since the value of m is not known *a priori*, it is not clear what value of model dimension to specify. There are several approaches which can give a hint: false nearest neighbors technique [51], correlation dimension estimation [52], or principal component analysis [53]. However, in practice one usually tries different model dimensions, starting from a very small value and successively increasing it until a satisfactory model is obtained or the results saturate. Therefore, the choice of the model dimension and even of the time delays may become an integral part of a monolithic modeling process, rather than a separate first stage.

Different approaches have been suggested to choose the form of the function f in Eq. (8.12). Algebraic polynomials perform extremely badly already for the approximation of bivariate functions [16, 40], while for the "black box" problem one must often exploit the value of D in the range $5 \div 10$. Therefore, algebraic polynomials are rarely used in practice. They represent an example of *weak approximation* technique [34] since their number of parameters and errors rise very quickly with model dimension D. Weak approximation techniques also involve trigonometric polynomials and wavelets.

Author: Do you mean 5–10 instead of $5 \div 10$?

Much attention has been paid to the search for *strong approximation* techniques which behave almost equally well for small and rather big model dimensions. They involve, in particular, local methods [7, 9, 10, 54]. Strong global approximation can be achieved using radial, cylindrical, and elliptic basis functions [34, 50, 55], and artificial neural networks [8]. See also [56] for examples of different approaches. We do not discuss them in details but note that these constructions involve many parameters and the problem of model structure selection (Section 8.4.2) is especially important here.

Another Takens theorem considers continuous-time dynamical systems, Eq. (8.2), with much smoother functions on their right-hand side. It states that one can perform embedding into the space of successive derivatives of the observable, i.e., state vectors can be constructed as $\eta, d\eta/dt, \ldots, d^{D-1}\eta/dt^{D-1}$. This approach does not involve a parameter τ which is an advantage. However, it is more difficult to realize in practice since even weak measurement noise is a serious obstacle in the calculation of high-order derivatives. Sometimes, this problem can be solved with the aid of filtering, e.g., Savitsky–Golay filter, but for a sufficiently strong noise it becomes unsolvable. In practice, it is realistic to use the values of $D = 2 \div 3$, ; rare successes are reported for $D = 5$ [16]. In combination with these state vectors, one constructs usually a model ODE in the form

Author: Do you mean 2–3 instead of $2 \div 3$?

$$d^D\eta/dt^D = f(\eta, d\eta/dt, \ldots, d^{D-1}\eta/dt^{D-1}, \boldsymbol{c}) . \tag{8.13}$$

The situation with the choice of approximating function is the same as discussed above for the model, Eq. (8.12). However, when using the successive derivatives, there are more chances to observe a gradually varying experimental dependence, Eq. (8.13). Therefore, additional reasons to use algebraic polynomials appear. So, in Eq. (8.13) f often takes the form

$$f(x_1, x_2, \ldots, x_D, \boldsymbol{c}) = \sum_{l_1, l_2, \ldots, l_D = 0}^{K} c_{l_1, l_2, \ldots, l_D} \prod_{j=1}^{D} x_j^{l_j}, \quad \sum_{j=1}^{D} l_j \leqslant K . \tag{8.14}$$

The structure, Eq. (8.13), with algebraic polynomial, Eq. (8.14), or rational function on the right-hand side is even called *standard* [57] since, theoretically, any smooth dynamical system can be transformed into such a form for a sufficiently large D and K. The values of coefficients in both Eq. (8.12) and Eq. (8.13) are estimated with the aid of the OLS technique. This is valid for a sufficiently low measurement noise level.

Successful results of constructing a model in the form (8.12) can be found, e.g., in [50, 54, 56]. Examples of successful modeling with the aid of Eq. (8.13), we are aware of, are even more rare [16]. As a rule, the structure, Eqs. (8.13) and (8.14), leads to very cumbersome equations tending to exhibit orbits diverging to infinity. It is especially inefficient in the case of multidimensional models. We stress that all the approaches described in this section are rigorously justified only in the case of absence of both measurement and dynamical noise. Their generalizations to the noisy cases are quite problematic [58].

8.5.2 Choice of Dynamical Variables

Let us pay more attention to the important problem of the choice of dynamical (state) variables, i.e., components of the state vectors $\mathbf{x}$. There are very many techniques to obtain time series of state variables from an observable η. Having only a scalar observable, one can use either successive differentiation or time delay embedding (Section 8.5.1). Besides, there are techniques of weighted summation [59] and integration [60] appropriate for strongly nonuniform signals. Further, one can restore a phase of the signal as an additional variable using the analytic signal approach implemented either via the Hilbert transform or the complex wavelet transform [61]. It is also possible to use combinations of all the techniques, e.g., to obtain several variables with the time-delay embedding, several others with integration, and the rest with differentiation [59]. If one observes more than one quantity characterizing a process under investigation, then it is possible to obtain dynamical variables from a time realization of each observable using any combination of the mentioned techniques so that the number of variants rises extremely quickly, see also [62]. It may appear possible that some of the observables should better be ignored in modeling. For example, it may well happen that a better model can be constructed with successive derivatives of the only observable if it turns out easy to find an appropriate approximating function f in Eq. (8.13) for such a choice.

After the reconstruction of state vectors $\{\mathbf{x}(t_i)\}$, an experimental time series of "left-hand sides" of model equations $\{\mathbf{y}(t_i)\}$ is obtained from the time series $\{\mathbf{x}(t_i)\}$ via the numerical differentiation of $\{\mathbf{x}(t_i)\}$ for model ODEs, Eq. (8.2), or the time shift of $\{\mathbf{x}(t_i)\}$ for model maps, Eq. (8.1). "Unlucky" choice of dynamical variables can make the approximation of the model dependence $\mathbf{y}(\mathbf{x})$ with a smooth function more difficult, or even impossible if the relationship among $\mathbf{y}$ and $\mathbf{x}$ appears nonunique.

Taking into account the importance of the stage of the state variables selection [63, 64] and multiple alternatives available, an actual problem is to look for the best (or, at least, for a reasonable) set of state variables. It is, of course, possible just to try different variants and look for the best model in each case. However, this procedure would be too time consuming. Moreover, it may remain unclear why a good model is not achieved for a given set of dynamical variables: Whether it is due to inappropriate model function or due to inappropriate state variables.

A procedure suggested in [65] allows us to test different sets of dynamical variables and select variants which are more promising for the global modeling purposes. It is based on the ideas of [66, 67] and consists in a nonparametric test of an approximated dependence $y(\mathbf{x})$ for uniqueness and continuity. A domain V comprising the set of vectors $\{\mathbf{x}(t_i)\}$ is divided into "hypercubic" boxes of the size δ (Fig. 8.6(a)). Then, all the boxes $s_1, s_2, \ldots, s_M$ comprising at least two vectors are selected. The difference between maximal and minimal values of the "left-hand side" variable y within a box s_k is called *a local variation* ε_k. Maximal local

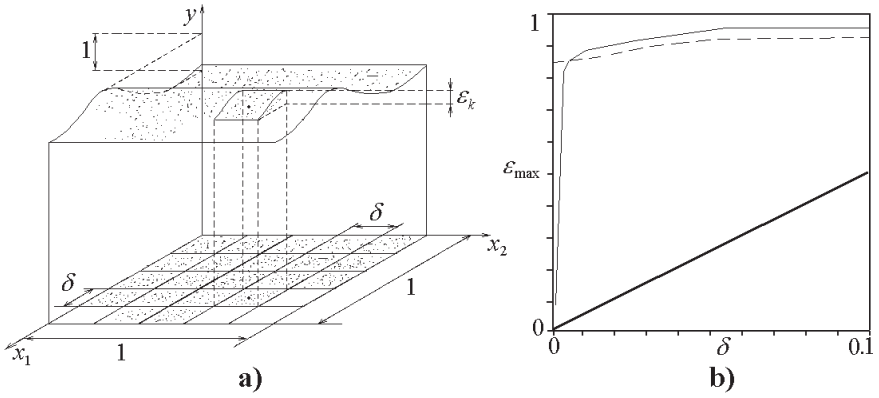

Fig. 8.6: (a) An illustration for the technique of testing a dependence $y(\mathbf{x})$ for uniqueness and continuity, $D = 2$. (b) The plots $\varepsilon_{\max}(\delta)$ for different sets of dynamical variables. The thick line corresponds to the best variant, the dashed line to the worst one (nonuniqueness).

variation $\varepsilon_{\max} = \max\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_M\}$ and the plot $\varepsilon_{\max}(\delta)$ are used as the main characteristics of the investigated dependence $y(\mathbf{x})$. Suitability of the considered quantities $\mathbf{x}$ and y for global modeling is estimated as follows. One must choose the variables so that the plot $\varepsilon_{\max}(\delta)$ tend to the origin gradually, without "knees" (Fig. 8.6(b), the lowest curve) for each of the approximated dependences $y(\mathbf{x})$.

8.6 Applications of Empirical Models

Probably, the most famous application is a forecast of the future evolution based on the available time series. This intriguing task is considered, e.g., in [4, 7, 9–11, 54–56]. Weather and climate forecasts, prediction of earthquakes, currency exchange rates and stock prices are often in the center of attention. Up to now, empirical models of the type described here are rarely useful to predict such complex processes due to "the curse of dimensionality" (difficulties in modeling quickly grow with dimensionality of the investigated dynamics), deficit of

experimental data, and noise. But chances for a successful forecast are higher in simpler situations.

An adequate empirical model may provide a deeper insight into mechanisms underlying the process under investigation [5, 16]. A positive result of model construction (high model quality) may validate physical ideas underlying the model structure. Such a conclusion is of an all-sufficient basic value and may inspire later practical applications.

Below, we consider other applications of empirical models. Namely, we focus on the problem of determination of a directional coupling between oscillators from short time series (Section 8.6.1) and present its applications to climatic sig-

nals (Section 8.6.2) and electroencephalograms (Section 8.6.3). Finally, we mention different practical applications and give references for further reading (Section 8.6.4).

8.6.1 Method to Reveal Weak Directional Coupling Between Oscillatory Systems from Short Time Series

One can extract different useful information from the estimates of model parameters. Thus, a sensitive approach to the determination of directionality of coupling between two oscillatory systems solely from their bivariate time series, a problem which is important in many practical and scientific fields, was suggested recently in [68]. It is based on the construction of model equations for *the phase dynamics* of the systems. Its main idea is to estimate how strong future evolution of the first system's phase depends on the second system's phase and vice versa. A detailed discussion can be found in the chapter written by M. Rosenblum (Chap. 7 in this volume). We describe only several points necessary to explain our modification of the method for the case of short time series and its applications.

First, one restores time series of the oscillations phases $\{\phi_1(t_1), \phi_1(t_2), \ldots, \phi_1(t_N)\}$ and $\{\phi_2(t_1), \phi_2(t_2), \ldots, \phi_2(t_N)\}$ from the original signals $\{x_1(t_1), x_1(t_2), \ldots, x_1(t_N)\}$ and $\{x_2(t_1), x_2(t_2), \ldots, x_2(t_N)\}$. We do it below with the analytic signal approach implemented via complex wavelet transform [61]. Given a signal $X(t)$, one defines signal $W(t)$ as

$$W(t) = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} X(t')\psi^*\big((t-t')/s\big)\,dt', \tag{8.15}$$

where $\psi(\eta) = \pi^{-1/4}\exp(-j\omega_0\eta)\exp(-\eta^2/2)$ is Morlet wavelet, s is a fixed time scale. For $\omega_0 = 6$ used below, $\mathrm{Re}W(t)$ can be regarded as $X(t)$ band-pass filtered around the frequency $f \approx 1/s$ with the relative bandwidth of $1/8$. The phase is defined as $\phi(t) = \arg W(t)$. It is the angle of rotation of the radius vector on the plane $(\mathrm{Re}W, \mathrm{Im}W)$ which increases by 2π after each complete evolution. To avoid edge effects while estimating Eq. (8.15) from a time series, we ignore segments of the length 1.4 s at each edge after the phase calculation.

Second, one constructs a global model relating phase increments over a time interval τ to the phases. Similarly to [37, 68], we use the form

$$\begin{aligned} \phi_1(t+\tau) - \phi_1(t) &= F_1\big(\phi_1(t), \phi_2(t+\Delta_1)\big) + \xi_1(t)\,, \\ \phi_2(t+\tau) - \phi_2(t) &= F_2\big(\phi_2(t), \phi_1(t+\Delta_2)\big) + \xi_2(t)\,, \end{aligned} \tag{8.16}$$

where $\xi_{1,2}$ are zero-mean random processes, $\Delta_{1,2}$ stand for possible time delays in coupling, F_1 is a trigonometric polynomial

$$F_1 = \sum_{m,n} [a_{m,n}\cos(m\phi_1 + n\phi_2) + b_{m,n}\sin(m\phi_1 + n\phi_2)]\,, \tag{8.17}$$

F_2 is defined analogously. The strength of the influence of system 2 on system 1 ($2 \to 1$) is quantified as

$$\begin{aligned} c_1^2 &= \frac{1}{2\pi^2} \int_0^{2\pi} \int_0^{2\pi} (\partial F_1/\partial \phi_2)^2 \, d\phi_1 \, d\phi_2 \\ &= \sum_{m,n} n^2 \left(a_{m,n}^2 + b_{m,n}^2\right). \end{aligned} \tag{8.18}$$

The influence $1 \to 2$ is quantified "symmetrically" (c_2^2). We use the third-order polynomials for $F_{1,2}$ and set τ equal to a basic oscillation period.

Given a time series, one estimates the coefficients $a_{m,n}, b_{m,n}$ via the OLS technique. Then, one can get the estimate of $\hat{c}_1^2$ by replacing the true values of $a_{m,n}$, $b_{m,n}$ in Eq. (8.18) with their estimates. A reliable detection of the weak directional coupling can only be achieved in nonsynchronous regimes. The latter can be diagnosed if the mean phase coherence

$$\rho(\Delta) = \sqrt{\langle \cos\big(\phi_1(t) - \phi_2(t+\Delta)\big)\rangle_t^2 + \langle \sin\big(\phi_1(t) - \phi_2(t+\Delta)\big)\rangle_t^2} \tag{8.19}$$

[69] is much less than 1.

The estimators $\hat{c}_1$ and $\hat{c}_2$ are quite precise only for long signals (about 1000 basic periods for moderate noise levels). However, in practice one must often deal with much shorter signals of about several dozens of basic periods. Thus, to analyze a nonstationary time series (e.g., in physiology) one must divide it into relatively short segments and estimate coupling characteristics from each segment separately. An attempt to apply the technique without modifications to such short series leads to biased estimates. Unbiased estimators γ_1 and γ_2 have been proposed in [70] instead of $\hat{c}_1^2$ and $\hat{c}_2^2$, respectively, and an index $\delta = \gamma_2 - \gamma_1$ is used to characterize coupling directionality. Expressions for their 95% confidence bands have also been derived. The latter allows us to trace significance of the estimates obtained from each short segment. (We do not show the formulas here since they are rather cumbersome.) For moderate coupling strength and phase nonlinearity, γ_1 and γ_2 guarantee the probability of erroneous conclusions about the presence of coupling less than 0.025 [71]. Additional tests with exemplary oscillators show that $\gamma_1(\Delta_1)$ and $\gamma_2(\Delta_2)$ are applicable for a time series as short as 20 basic periods if $\rho(\Delta) < 0.4$. The latter condition excludes synchronous-like signals. Other available techniques for coupling direction identification and conditions for superiority of the described technique are reported in [72].

8.6.2 Application to Climatic Data

Using the above technique, we investigated the dynamics of the North Atlantic oscillation (NAO) and El Niño/Southern oscillation (ENSO) processes for the second half of the twentieth century. ENSO and NAO represent the leading modes of interannual climate variability for the globe and Northern Hemisphere (NH),

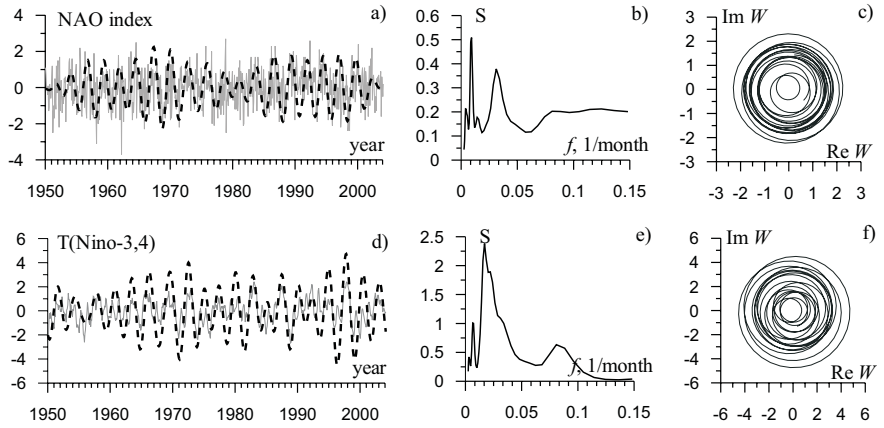

Fig. 8.7: Individual characteristics of the NAO index and T(Niño-3,4). (a) NAO index (the gray line) and ReW for s = 32 months (the dashed line). (b) Global wavelet spectrum of the NAO index. (c) An orbit $W(t)$ for the NAO index, s = 32 months. (d)–(f) The same as (a)–(c), but for T(Niño-3,4).

respectively [73, 74]. Different tools have been used for the analysis of their interaction, in particular, cross-correlation function and Fourier and wavelet coherence, e.g., [75]. However, all the climatic signals are rather short that has made confident inference about the character of interaction between those processes difficult.

Here, we present the results only for a specific pair of climatic indices. The first one is NAO index `http://www.ncep.noaa.gov` defined as the leading decomposition mode of the field of 500 hPa geopotential height in NH based on the "rotated principal component analysis" [76]. The second one is T(Niño-3,4) which characterizes sea surface temperature in an equatorial region of the Pacific Ocean (5°N–5°S, 170°W–120°W) [77]. These time series cover the period 1950–2004 (660 monthly values).

Figure 8.7 demonstrates individual characteristics of the NAO index (Fig. 8.7(a)) and T(Niño-3,4) (Fig. 8.7(d)). Global wavelet spectra of the NAO index and T(Niño-3,4) exhibit several peaks (Figs. 8.7(b)and (e)). One can assume that the peaks correspond to some oscillatory processes for which the phase can be adequately introduced. To extract phases of "different rhythms" in NAO and ENSO, we tried several values of s in Eq. (8.15) corresponding to the different spectral peaks. We estimated coupling between all the rhythms pairwise. The only case when substantial conclusions about the presence of coupling are inferred is the "rhythm" with s = 32 months for both signals, see the dashed lines in Figs. 8.7(a) and 8.7(d). The phases of 32-month rhythms in both signals are well defined since clear rotation of the orbits around the origin on the complex plane takes place (Figs. 8.7(c) and 8.7(f)).

The results of the phase dynamics modeling are shown in Fig. 8.8 for s = 32

and model, Eq. (8.16), with $\tau = 32$. Figure 8.8(a) shows that the technique is applicable only for $\Delta_1 > -30$ where $\rho < 0.4$. The influence ENSO $\rightarrow$ NAO is pointwise significant for $-30 \leqslant \Delta_1 \leqslant 0$ and maximal for $\Delta_1 = -24$ months (Fig. 8.8(b)). Apart from the pointwise p-level, one can infer the presence of the influence ENSO $\rightarrow$ NAO as follows. Probability of a random erroneous conclusion about coupling presence based only on a pointwise significant γ_1 for a specific Δ_1 is 0.025. Taking into account that the values of $\gamma_1(\Delta_1)$ separated with Δ_1 less than τ are strongly correlated, one can consider as "statistically independent" the values of γ_1 from the two groups: $-30 \leqslant \Delta_1 \leqslant 0$ and $0 < \Delta_1 \leqslant 32$. Then, the probability of erroneous conclusion based on pointwise significant γ_1 at least in one of the two groups as observed in Fig. 8.8(b) is approximately twice as large and, hence, equal to 0.05. Thus, we conclude with confidence probability of 0.95 that the influence ENSO $\rightarrow$ NAO is present. Most probably, it is delayed by 24 months. However, the latter conclusion is not so reliable. No signs of the influence NAO $\rightarrow$ ENSO are detected (Fig. 8.8(c)).

We note that large ρ for $\Delta < -30$ do not indicate strong coupling. For such short time series and close basic frequencies of oscillators, the probability to get $\rho > 0.4$ for uncoupled processes is greater than 0.5 as observed in numerical experiments with exemplary oscillators. More details can be found in [78].

We stress that the conclusion about the presence of the influence ENSO $\rightarrow$ NAO is quite reliable here. Confidence probability 0.95 was not accessible for traditional techniques. It can be attributed to high sensitivity of the phases to weak coupling.

8.6.3 Application to Electroencephalogram Data

Here, we present an application of the estimators to analyze a two-channel human intracranial epileptic electroencephalogram (EEG) recording with the purpose of epileptic focus localization.

The data were recorded from intracranial depth electrodes implanted in a patient with medically refractory temporal lobe epilepsy as part of routine clinical investigations to determine candidacy for epilepsy surgery (provided by Dr. Richard Wennberg, Toronto Western Hospital). The recordings included several left temporal neocortical $\rightarrow$ hippocampal seizures that occurred over the course of a long partial status epilepticus, see an example in Figs. 8.9(a) and (b). Two channels were analyzed: the first channel situated in the left hippocampus, and the second channel in the left temporal neocortex, where the "interictal" activity between seizures at the time comprised pseudoperiodic epileptiform discharges. The visual analysis of the interictal–ictal transitions (shown with vertical dashed lines) determined that the seizures all started first in the neocortex, with an independent seizure subsequently beginning at the ipsilateral hippocampus. We analyzed four recordings, but here we present the results for only one of

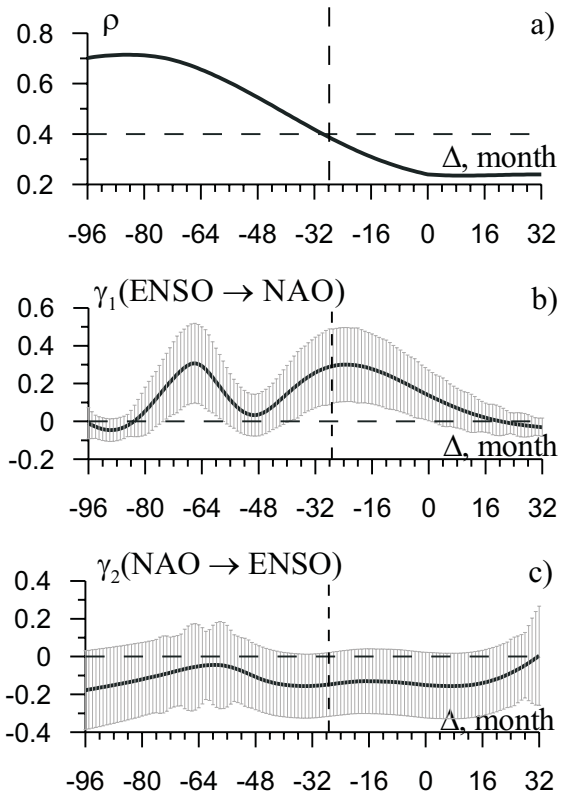

Fig. 8.8: Analysis of coupling from the NAO index and T(Niño-3,4). (a) Mean phase coherence. (b, c) The estimators of the strength of the influence ENSO → NAO (Δ means Δ_1) and NAO → ENSO (Δ means Δ_2), respectively, with their 95% confidence bands.

them for the sake of brevity, as an illustration of application of the method to a nonstationary real-world system.

The time series of Figs. 8.9(a) and (b) contains 4.5 min of depth electrode EEG (referential recording to scalp vertex electrode) recorded at a sampling frequency of 250 Hz. There are more or less significant peaks in power spectra for both channels (not shown). For the hippocampal channel: at frequency 3.2 Hz before the seizure (starting approximately at the 100th second and finishing approximately at the 220th second), 2.3 Hz after the seizure, and 7.1 Hz during the seizure. For the neocortex channel: at frequency 1.4 Hz before the seizure, 1.6 Hz after the seizure, and 7.1 Hz during the seizure. We have computed coupling characteristics in a running window. The length of running window was changed from $N = 10^3$ data points to $N = 10^4$ data points. Time delays $\Delta_{1,2}$ were set equal to zero. The phases were determined using Eq. (8.15) with $\omega_0 = 2$ and different

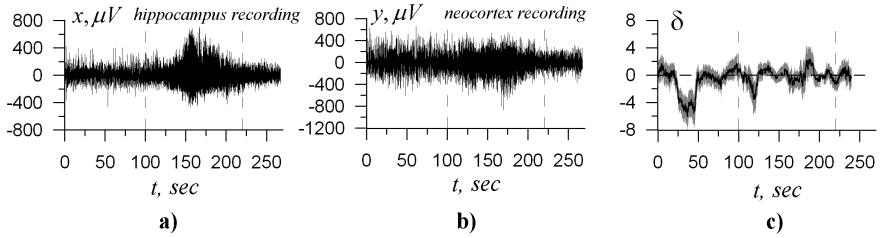

Fig. 8.9: Intracranial EEG recordings: (a) from the hippocampus, (b) from the temporal lobe of the neocortex. (c) Coupling directionality index δ with its 95% confidence band (the gray train), t is the starting time instant of a running window of the length of 8 s. Negative values of δ indicate influence of the neocortex on the hippocampus. The vertical dashed lines indicate a seizure onset and offset. Index δ is significantly less than zero during a period of 25–55 s before the seizure.

time scales s. In particular, we tried the time scales corresponding to the main peak of the scalogram for each signal which is $s = 0.14$ s for the hippocampal signal, and $s = 0.19$ s for the neocortex signal, see Fig. 8.9(c) (where $\tau = 33\Delta t$).

We present only one set of results in Fig. 8.9(c) (gray tail denotes 95% confidence bands) obtained for $N = 6000$. Coupling is regarded as significant if the confidence band does not include zero, e.g., gray tail does not intersect the abscissa axis. The preliminary results seem promising for the localization of the epileptic focus, because a long interval (30 s length for the example shown) of significant predominant coupling direction neocortex $\rightarrow$ hippocampus is observed before the seizure. It can be considered as an indication that epileptic focus is located near the neocortex channel that agrees with *a priori* clinical information. Despite we presented only one example, we note that the results are sufficiently robust and are observed for a significant range of values of the above-mentioned window lengths and parameters.

Similar results are observed for the three of the four analyzed recordings and not observed for one of them. Right now, we do not draw any definite conclusions about the applicability of the method to localize epileptic focus. This is only the first attempt and, of course, more EEG recordings should be processed to quantify the method's sensitivity and specificity. This is a subject of ongoing research. Therefore, the results presented here should not be overestimated, being rather an illustration of the way how to apply the method in practice and what kind of information one can expect from it.

8.6.4 Other Applications

We should mention several other useful applications of the reconstruction methods. They include detection of quasistationary segments in nonstationary signals [79–82], prediction of bifurcations in weakly nonautonomous systems [83],

multichannel confidential transmission of information [84, 85], signal classification [86], testing for nonlinearity and determinism [87], and adaptive nonlinear noise reduction [88–90]. Among the very interesting applications, we stress again on the reconstruction of characteristics of nonlinear elements in electric circuits and other systems with the aid of a modeling procedure in the "gray box" setting when such characteristics may not be accessible to direct measurements. This approach is successfully brought about during the investigation of dynamical properties of a ferroelectric [39], semiconductor diodes [91], and optical fiber ring [92].

8.7 Conclusions

Seemingly, mathematical modeling will always remain an art to a significant extent. However, there may be developed some general principles and particular recipes increasing our chances to obtain a "good" model. Some results of this type related to the time series modeling are discussed in this chapter. Besides, we systematized many available techniques based on the scheme of Fig. 8.1 whose different items were illustrated with different problem settings: from "white box" via "gray box" to "black box" problems. We outlined different techniques which were tested in numerical experiments with the reconstruction of exemplary equations from their noise-corrupted solutions. Many of the techniques were already successfully applied to the investigation of laboratory and real-world systems such as nonlinear electric circuits, climatic processes, functional systems of living organisms, etc. In particular, we reported the results of the analysis of the interaction between complex processes in climatology and neurophysiology based on their empirical modeling.

We have not discussed modeling of spatially distributed systems, even though it attracts considerable attention [93–97]. As well, we have omitted discussion of time-delay systems [92, 98, 99] and only briefly touched on stochastic nonlinear models [29, 100]. Many methods for the construction of finite-dimensional deterministic models are also just mentioned. Instead, we have tried to give simple illustrations of some key points and provide multiple references to the works comprising more detailed discussion for the further reading. Therefore, this survey is only an "excursus into ...," rather than an irrefragable treatment of the empirical modeling problems.

Acknowledgements

We acknowledge fruitful collaboration with our colleagues Ye. P. Seleznev, V. I. Ponomarenko, M. D. Prokhorov, T. V. Dikanev, M. B. Bodrov, I. V. Sysoev, A. S. Karavaev, V. S. Vlaskin, R. A. Wennberg, J.-L. Perez Velazquez, and I. I. Mokhov. Our research in the field of time series modeling was supported by the Russian Foundation for Basic Research (grant 05-02-16305), the President of Russia (MK-

1067.2004.2), Program "Basic Sciences for Medicine" of the Presidium of Russian Academy of Sciences, Program BRHE of the American Civilian Research and Development Foundation and Russian Ministry of Education (REC-006), and Russian Science Support Foundation.

Author: Please update [78]

References

[1] L. Ljung. *System Identification. Theory for the User*. Prentice-Hall, New Jersey, 1987.

[2] G. Gouesbet, S. Meunier-Guttin-Cluzel, and O. Menard, editors. *Chaos and Its Reconstructions*. Nova, New York, 2003.

[3] G. U. Yule. *Phil. Trans. R. Soc. London A*, 226:267, 1927.

[4] G. Box and G. Jenkins. *Time Series Analysis. Forecasting and Control*. Holden-Day, San-Francisco, 1970.

[5] J. P. Crutchfield and B. S. McNamara. *Complex Syst.*, 1:417, 1987.

[6] J. Cremers and A. Hubler. *Z. Naturforschung A*, 42:797, 1987.

[7] J. D. Farmer and J. J. Sidorowich. *Phys. Rev. Lett.*, 59:845, 1987.

[8] D. S. Broomhead and D. Lowe. *Complex Syst.*, 2:321, 1988.

[9] M. Casdagli. *Physica D*, 35:335, 1989.

[10] H. D. I. Abarbanel, R. Brown, and J. B. Kadtke. *Phys. Lett. A*, 138:401, 1989.

[11] J. B. Kadtke and Yu. A. Kravtsov, editors. *Predictability of Complex Dynamical Systems*. Springer, Berlin, 1996.

[12] H. D. I. Abarbanel, R. Brown, J. J. Sidorowich, and L. S. Tsimring. *Rev. Mod. Phys.*, 65:1331, 1993.

[13] H. D. I. Abarbanel. *Analysis of Observed Chaotic Data*. Springer, New York, 1996.

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

[15] P. E. Rapp, T. I. Schmah, and A. I. Mees. *Physica D*, 132:133, 1999.

[16] G. Gouesbet, S. Meunier-Guttin-Cluzel, and O. Menard, editors. *Chaos and Its Reconstructions*, pages 1–160. Nova, New York, 2003.

[17] C. R. Shalizi. *arXiv:nlin.AO/*, 0307015:3, 2003. URL `http://www.arxiv.org/abs/nlin.AO/0307015`.

[18] H. U. Voss, J. Timmer, and J. Kurths. *Int. J. Bif. Chaos*, 14:1905, 2004.

[19] L. Jaeger and H. Kantz. *Chaos*, 6:440, 1996.

[20] P. E. McSharry and L. A. Smith. *Phys. Rev. Lett.*, 83:4285, 1999.

[21] K. Judd. *Phys. Rev. E*, 67:026212, 2003.

[22] W. Horbelt and J. Timmer. *Phys. Lett. A*, 310:269, 2003.

[23] V. F. Pisarenko and D. Sornette. *Phys. Rev. E*, 69:036122, 2004.

[24] D. A. Smirnov, V. S. Vlaskin, and V. I. Ponomarenko. *Phys. Lett. A*, 336:448, 2005.

[25] E. Baake, M. Baake, H. J. Bock, and K. M. Briggs. *Phys. Rev. A*, 45:5524, 1992.

[26] U. Parlitz. *Phys. Rev. Lett.*, 76:1232, 1996.

[27] I. A. Ibragimov and R. Z. Has'minskii. *Asymptotic Theory of Estimation*. Nauka, Moscow, 1979. In Russian.

[28] V. N. Vapnik. *Estimation of Dependencies Based on Empirical Data*. Springer, Berlin, Heidelberg, 1982.

[29] C. L. Bremer and D. T. Kaplan. *Physica D*, 160:116, 2001.

[30] A. Sitz, U. Schwartz, J. Kurths, and H. U. Voss. *Phys. Rev. E*, 66:016210, 2002.

[31] I. Swameye, T. G. Muller, J. Timmer, O. Sandra, and U. Klingmuller. *Proc. Natl. Acad. Sci. USA*, 100:1028, 2003.

[32] W. Horbelt, J. Timmer, M. J. Bunner, R. Meucci, and M. Ciofini. *Phys. Rev. E*, 64:016222, 2001.

[33] J. Rissanen. *Stochastic Complexity in Statistical Inquiry*. World Scientific, Singapore, 1989.

[34] K. Judd and A. I. Mees. *Physica D*, 82:426, 1995.

[35] L. A. Aguirre, U. S. Freitas, C. Letellier, and J. Maquet. *Physica D*, 158:1, 2001.

[36] B. P. Bezruchko, T. V. Dikanev, and D. A. Smirnov. *Phys. Rev. E*, 64:036210, 2001.

[37] L. Cimponeriu, M. Rosenblum, and A. Pikovsky. *Phys. Rev. E*, 70:046213, 2004.

[38] T. Nakamura, D. Kilminster, and K. Judd. *Int. J. Bif. Chaos*, 14:1129, 2004.

[39] R. Hegger, H. Kantz, F. Schmuser, M. Diestelhorst, R.-P. Kapsch, and H. Beige. *Chaos*, 8:727, 1998.

[40] B. P. Bezruchko and D. A. Smirnov. *Phys. Rev. E*, 63:016207, 2001.

[41] B. P. Bezruchko, Ye. P. Seleznev, D. A. Smirnov, and I. V. Sysoev. *Sov. Tech. Phys. Lett.*, 29:69, 2003.

[42] F. Takens. *Lec. Notes Math.*, 898:366, 1981.

[43] T. Sauer, J. A. Yorke, and M. Casdagli. *J. Stat. Phys.*, 65:579, 1991.

[44] M. Casdagli, S. Eubank, J. D. Farmer, and J. Gibson. *Physica D*, 51:52, 1991.

[45] C. J. Cellucci, A. M. Albano, and P. E. Rapp. *Phys. Rev. E*, 67:066210, 2003.

[46] J. F. Gibson, J. D. Farmer, M. Casdagli, and S. Eubank. *Physica D*, 57:1, 1992.

[47] A. M. Fraser and H. L. Swinney. *Phys. Rev. A*, 33:1131, 1986.

[48] W. Liebert and H. G. Schuster. *Phys. Lett. A*, 142:107, 1989.

[49] J. P. Eckmann and D. Ruelle. *Rev. Mod. Phys.*, 57:617, 1985.

[50] K. Judd and A. I. Mees. *Physica D*, 120:273, 1998.

[51] M. B. Kennel, R. Brown, and H. D. I. Abarbanel. *Phys. Rev. A*, 45:3403, 1992.

[52] P. Grassberger and I. Procaccia. *Physica D*, 9:189, 1983.

[53] D. S. Broomhead and G. P. King. *Physica D*, 20:217, 1986.

[54] D. Kugiumtzis, O. C. Lingjaerde, and N. Christophersen. *Physica D*, 112: 344, 1998.

[55] K. Judd and M. Small. *Physica D*, 136:31, 2000.

[56] N. A. Gerschenfeld and A. S. Weigend, editors. *Time Series Prediction: Forecasting the Future and Understanding the Past*, volume XV of *SFI Studies in the Science of Complexity*. Addison-Wesley, Reading, MA, 1993.

[57] G. Gouesbet and C. Letellier. *Phys. Rev. E*, 49:4955, 1994.

[58] J. Stark, D. S. Broomhead, M. Davies, and J. Huke. *Nonlinear Analysis. Theory, Methods and Applications*. Elsevier, The Netherlands, 1997.

[59] R. Brown, N. F. Rulkov, and E. R. Tracy. *Phys. Rev. E*, 49:3784, 1994.

[60] N. B. Janson, A. N. Pavlov, and V. S. Anishchenko. *Int. J. Bif. Chaos*, 8:825, 1998.

[61] J. P. Lachaux, E. Rodriguez, M. Le Van Quyen, A. Lutz, J. Martienerie, and F. J. Varela. *Int. J. Bif. Chaos*, 10:2429, 2000.

[62] L. Cao, A. I. Mees, and K. Judd. *Physica D*, 121:75, 1998.

[63] C. Letellier, J. Macquet, L. Le Sceller, G. Gouesbet, and L. A. Aguirre. *J. Phys. A: Math. Gen.*, 31:7913, 1998.

[64] C. Letellier and L. A. Aguirre. *Chaos*, 12:549, 2002.

[65] D. A. Smirnov, B. P. Bezruchko, and Ye. P. Seleznev. *Phys. Rev. E*, 65:026205, 2002.

[66] D. T. Kaplan. *Physica D*, 73:738, 1994.

[67] N. F. Rulkov, M. M. Sushchik, L. S. Tsimring, and H. D. I. Abarbanel. *Phys. Rev. E*, 51:980, 1995.

[68] M. G. Rosenblum and A. S. Pikovsky. *Phys. Rev. E*, 64:R045202, 2001.

[69] F. Mormann, K. Lehnertz, P. David, and C. E. Elger. *Physica D*, 144:358, 2000.

[70] D. Smirnov and B. Bezruchko. *Phys. Rev. E*, 68:046209, 2003.

[71] D. A. Smirnov, M. B. Bodrov, J. L. Perez Velazquez, R. A. Wennberg, and B. P. Bezruchko. *Chaos*, 15:024102, 2005.

[72] D. A. Smirnov and R. G. Andrzejak. *Phys. Rev. E*, 71:036207, 2005.

[73] Clivar Initial Implementation Plan. Technical report, WCRP No. 103. WMO/TD No.869. ICPO No.14, 1998. URL `http://www.clivar.dkrz.de/hp.html`.

[74] J. T. Houghton, Y. Ding, D. J. Griggs, and M. Noguer et al., editors. *Climate Change 2001: The Scientific Basis. Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, 2001.

[75] S. Jevrejeva, J. Moore, and A. Grinsted. *J. Geophys. Res.*, 108:4677, 2003.

[76] A. G. Barnston and R. E. Livezey. *Mon. Wea. Rev.*, 115:1083, 1987.

[77] K. Arpe, L. Bengtsson, G. S. Golitsyn, I. I. Mokhov, V. A. Semenov, and P. V. Sporyshev. *Geophys. Res. Lett.*, 27:2693, 2000.

[78] D. A. Smirnov and I. I. Mokhov. *Geophys. Res. Lett.*, 2005. submitted.

[79] T. Schreiber. *Phys. Rev. Lett.*, 78:843, 1997.

[80] T. Schreiber. *Phys. Rep.*, 308:3082, 1999.

[81] D. Gribkov and V. Gribkova. *Phys. Rev. E*, 61:6538, 2000.

[82] T. Dikanev, D. Smirnov, R. Wennberg, J. L. Perez Velazquez, and B. Bezruchko. *Clin. Neurophysiol.*, 116:1796, 2005.

[83] A. M. Feigin, Y. I. Molkov, D. N. Mukhin, and E. M. Loskutov. *Faraday Discussions*, 120:105, 2002.

[84] V. S. Anishchenko and A. N. Pavlov. *Phys. Rev. E*, 57:2455, 1998.

[85] V. I. Ponomarenko and M. D. Prokhorov. *Phys. Rev. E*, 66:026215, 2002.

[86] M. Kremliovsky, J. Kadtke, M. Inchiosa, and P. Moore. *Int. J. Bif. Chaos*, 8: 813, 1998.

[87] M. Small, K. Judd, and A. I. Mees. *Stat. Comp.*, 11:257, 2001.

[88] J. D. Farmer and J. J. Sidorowich. *Physica D*, 47:373, 1991.

[89] E. J. Kostelich and T. Schreiber. *Phys. Rev. E*, 48:1752, 1993.

[90] M. E. Davies. *Physica D*, 79:174, 1994.

[91] I. V. Sysoev, D. A. Smirnov, Ye. P. Seleznev, and B. P. Bezruchko. *Proc. 2nd IEEE International Conference on Circuits and Systems for Communications*, volume 140. Moscow, Russia, 2004.

[92] H. U. Voss, A. Schwache, J. Kurths, and F. Mitschke. *Phys. Lett. A*, 256:47, 1999.

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

[94] H. U. Voss, M. Bunner, and M. Abel. *Phys. Rev. E*, 57:2820, 1998.

[95] M. Baer, R. Hegger, and H. Kantz. *Phys. Rev. E*, 59:337, 1999.

[96] U. Parlitz and C. Merkwirth. *Phys. Rev. Lett.*, 84:1890, 2000.

[97] A. Sitz, J. Kurths, and H. U. Voss. *Phys. Rev. E*, 68:016202, 2003.

[98] M. J. Bunner, M. Ciofini, A. Giaquinta, R. Hegger, H. Kantz, R. Meucci, and A. Politi. *Eur. Phys. J. D*, 10:165, 2000.

[99] M. D. Prokhorov, V. I. Ponomarenko, A. S. Karavaev, and B. P. Bezruchko. *Physica D*, 203:209, 2005.

[100] J. Timmer. *Chaos, Solit. Fract.*, 11:2571, 2000.

Index