Chapter 10
Model Equations: "Black Box" Reconstruction

Black box reconstruction is both the most difficult and the most tempting modelling problem when any prior information about an appropriate model structure is lacking. An intriguing thing is that a model capable of reproducing an observed behaviour or predicting further evolution should be obtained only from an observed time series, i.e. "from nothing" at first sight. Chances for a success are not large. Even more so, a "good" model would become a valuable tool to characterise an object and understand its dynamics. Lack of prior information causes one to utilise *universal model structures*, e.g. artificial neural networks, radial basis functions and algebraic polynomials are included in the right-hand sides of dynamical model equations. Such models are often multi-dimensional and involve quite many free parameters.

Since time series of all variables for such a model must be obtained from observed data, "restoration" of lacking variables gets extremely important. One often calls it "phase orbit reconstruction" or "state vector reconstruction". A theoretical justification is given by celebrated *Takens' theorems* (Sect. 10.1).

Not less important and difficult is the approximation stage, where one fits a dependence of the next state on the current one $\mathbf{x}_{n+1} = \mathbf{f}(\mathbf{x}_n, \mathbf{c})$ or of the phase velocity on the state vector $d\mathbf{x}/dt = \mathbf{f}(\mathbf{x}, \mathbf{c})$. In practice, one usually manages to get a valid model if it appears sufficient to use its moderate dimension, roughly, not greater than 5–6. To construct higher dimensional models, one needs huge amounts of data and deals with approximation of multivariable functions (Sect. 10.2) which is much more difficult than that of one-variable functions (Sects. 7.2, 9.1 and 9.3). Moreover, troubles quickly rise with the model dimension (Kantz and Schreiber, 1997). This is the so-called "curse of dimensionality", the main obstacle in the modelling of multitude of real-world processes.

Yet, successful results have sometimes been obtained for complex real-world objects even under the black box setting. Also, there are several nice theoretical results and many practical algorithms of reconstruction, which appear efficient for prediction and other modelling purposes.

10.1 Reconstruction of Phase Orbit

To get lacking model variables in modelling from a time series $\{\eta(t_1), \eta(t_2), \ldots, \eta(t_N)\}$, one can use subsequent values of η, i.e. a state vector $\mathbf{x}(t_i) = [\eta(t_i), \eta(t_i + \tau), \ldots, \eta(t_i + (D - 1)\tau)]$, where τ is the *time delay*, or successive derivatives, i.e. a state vector $\mathbf{x}(t_i) = [\eta(t_i), \mathrm{d}\eta(t_i)/\mathrm{d}t, \ldots, \mathrm{d}^{D-1}\eta(t_i)/\mathrm{d}t^{D-1}]$. These approaches have been applied for a long time without special justification (Sect. 6.1.2). Thus, the former one is, in fact, used since 1927 for the widely known *autoregression models* (4.12), where a future value of an observable is predicted based on several previous values (Yule, 1927). It seems just reasonable. If there is no other information besides a time series, then one can use only the previous values of an observable or their combinations to make a forecast.

At the beginning of the 1980s, relationships between both mentioned approaches and the theory of dynamical systems were revealed. It was proven that in reconstruction from a scalar time realisation of *a dynamical system* (under some conditions of smoothness), both time delays and successive derivatives assure an equivalent description of the original dynamics if the dimension of the restored vectors D is large enough. Namely, the condition $D > 2d$ should be fulfilled, where d is the dimension of a set M in the phase space of an original system, where a modelled motion occurs.[1] These statements constitute celebrated Takens' theorems (Takens, 1981) as discussed in Sect. 10.1.1. We note that the theorems are related to the case when an object is a *deterministic dynamical system* (Sect. 2.2.1).

In the modelling of real-world objects, one can use the above approaches without referring to the theorems, since it is impossible to check whether the conditions of the theorems are fulfilled and the dimension d is unknown (if one may speak about all that in respect of a real-world object at all). Yet, the value of the theoretical results obtained by Takens is high. Firstly, after their formulation it has become clear that both above approaches are suitable for the modelling of a sufficiently wide class of systems. Thus, the theorems "bless" practical application of the approaches, especially if one has any ideas confirming that the conditions of the theorems are fulfilled in a given situation. Secondly, based on the theory of dynamical systems, one has developed new fruitful approaches to the choice of the reconstruction parameters, such as the time delay τ, the model dimension D and others, as discussed in Sect. 10.1.2.

---

[1] The set M is a compact smooth manifold and the quantity d is its topological dimension (Sect. 10.1.1). There are generalisations of the theorem to the case of non-smooth sets M and fractal dimension d, which are beyond the scope of our discussion (Sauer et al., 1991). We note that the set M mentioned in the theorems *does not inevitably* correspond to the motion on an attractor. For instance, let an attractor be a limit cycle C "reeled" on a torus M. If one is interested only in the description of an established periodic motion on the cycle C, then it is sufficient to use $D > 2$ model variables for reconstruction according to Takens' theorems. If one needs to describe motions on the entire torus M, including transient processes, then it is necessary to use $D > 4$ variables. In practice, one often has a single realisation corresponding to established dynamics. Therefore, one usually speaks of the reconstruction of an attractor.

10.1.1 Takens' Theorems

We start with illustrating the theorems with a simple example and then give their mathematical formulations and discuss some details in a more strict way. Throughout this subsection, we denote the state vector of an original system $\mathbf{y}$ as distinct from the reconstructed vectors $\mathbf{x}$. The notation d is related to the dimension of the set M in the phase space of an *original* system. It is not necessarily the dimension of the entire phase space, i.e. of the vector $\mathbf{y}$. D is the dimension of the reconstructed vectors $\mathbf{x}$ and, hence, of a resulting model.

10.1.1.1   An Illustrative Example

Let an object be a continuous-time three-dimensional dynamical system. Its state vector is $\mathbf{y} = (y_1, y_2, y_3)$. Let a motion to occur on a limit cycle (Fig. 10.1a), i.e. on a set M of the dimension $d = 1$.

If all three variables y_1, y_2, y_3 were observed, one could proceed directly to the approximation of the dependence of $\mathbf{y}(t + \tau)$ on $\mathbf{y}(t)$, which is unique since $\mathbf{y}$ is a state vector. The latter means that whenever a certain value $\mathbf{y} = \mathbf{y}^*$ is observed, a unique future value follows it in a fixed time interval. The same present leads to the same future. If not all the components of the state vector are observed, then the situation is more complicated. One may pose a question: How many variables suffice for an equivalent description of an original dynamics? Which variables are suitable for that and which ones are not?

Since the set M, where the considered motion takes place, is one dimensional ($d = 1$), there should exist such a scalar dynamical variable which is sufficient to describe this motion. For instance, a closed curve M (Fig. 10.1a) can be mapped on
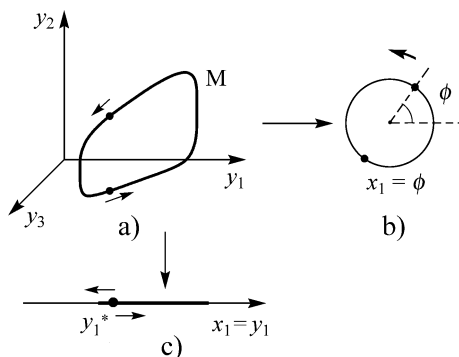

Fig. 10.1 One-dimensional representations of a limit cycle: (**a**) an original limit cycle; (**b**) its mapping on a circle; (**c**) its projection on a coordinate axis. The dimension of an original system is equal to three; the dimension of the set M, which is a closed curve, is $d = 1$; the dimension of the "reconstructed" vectors is $D = 1$ in both cases. Two different states [filled circles in the panel (**a**)] correspond to the two different points on the circle [the panel (**b**)] and to a single point on the axis [the panel (**c**)]. The mapping of a cycle on a circumference is one-to-one and its mapping on a line segment is not one-to-one

a circumference (Fig. 10.1b). It is important that the vectors $\mathbf{y}(t)$ on the cycle M can be related to the angle of rotation $\phi(t)$ of a point around the circumference in a one-to-one way. The variable $\phi(t)$ is the "wrapped" phase of oscillations (Sect. 6.4.3). Due to one-to-oneness, the variable ϕ completely determines the state of the system: The value of the phase ϕ^* corresponds to a unique simultaneous value of the vector $\mathbf{y}^*$. Having an observable ϕ, one can construct a one-dimensional *deterministic dynamical model* ($D = 1$) with $x_1 = \phi$.

However, not any variable is appropriate for a one-dimensional representation. Thus, if one observes just a single component of the vector $\mathbf{y}$, e.g. a coordinate y_1, then a closed curve is mapped on a line segment (a simple projection onto the y_1-axis). This mapping is not one-to-one. Almost any point $y_1^*(t)$ of the segment corresponds to two state vectors $\mathbf{y}(t)$ differing by the direction of the further motion (to the left or to the right along the y_1-axis, see Fig. 10.1a, c). Thus, y_1 does not uniquely determine the state of the system. If one observes some value $y_1 = y_1^*$, then one of the two possible future values can follow. Therefore, a deterministic one-dimensional description of the observed motion with the variable $x_1 = y_1$ is impossible.

In general, if one uses the model dimension D equal to the dimension of the observed motion d, the construction of a dynamical model may appear successful if one is "lucky". However, empirical modelling may fail as well. Both results are typical in the sense that the situation does not change under weak variations of an original system, an observable and parameters of the reconstruction algorithm.

What changes if one uses a two-dimensional description for the above example? The same two situations are typical as illustrated in Fig. 10.2. If the two components of the original state vector y_1 and y_2 are observables, i.e. the model state vector is $\mathbf{x} = (y_1, y_2)$, it corresponds to a projection of the closed curve onto the plane (y_1, y_2). In such a projection, one may get a curve either without self-intersections (Fig. 10.2a) or with them (Fig. 10.2b) depending on the shape of the original curve and its spatial orientation. The former case provides a one-to-one relationship between the original curve and its projection, i.e. the two-dimensional vector $\mathbf{x}$ completely determines the state of the system. The latter case differs, since the self-intersection point y_1^*, y_2^* on the plane (y_1, y_2) in Fig. 10.2b corresponds to two different states of the original system, i.e. the relationship between $\mathbf{x}$ and $\mathbf{y}$ is not one-to-one. Therefore, one cannot uniquely predict the future following the current values y_1^*, y_2^*. Hence, the vector $\mathbf{x}$ is not suitable as the state vector of a global deterministic model. It can be used only locally, far from the self-intersection point.

A similar situation takes place if one uses any two variables instead of y_1 and y_2. For instance, let $\eta = h(\mathbf{y})$ be an observable, where h is an arbitrary smooth function and let the components of $\mathbf{x}$ be the time-delayed values of η: $\mathbf{x}(t) = (\eta(t), \eta(t+\tau))$. Depending on h and τ, one may observe a situation either without self-intersections on the plane (x_1, x_2) as in Fig. 10.2a or with self-intersections as in Fig. 10.2b. Thus, even the number of model variables exceeding the dimension of an observed motion, $D = 2 > d = 1$, does not *assure* the possibility of a deterministic description.
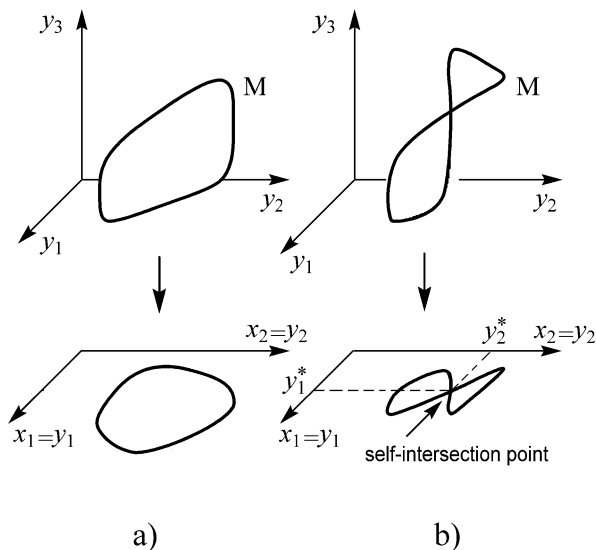
Fig. 10.2 Projections of a one-dimensional manifold from a three-dimensional space onto planes: (**a**) one-to-one mapping; (**b**) many-to-one mapping with a self-intersection point in the projection

Finally, let us consider a three-dimensional representation, e.g. when model state vectors are constructed as $\mathbf{x}(t) = (\eta(t), \eta(t+\tau), \eta(t+2\tau))$. An image of the original closed curve in the three-dimensional space (x_1, x_2, x_3) is also a closed curve, which typically does not exhibit self-intersections, i.e. there is a one-to-one correspondence between $\mathbf{x}$ and $\mathbf{y}$. An original motion on the limit cycle can be equivalently described with the vectors $\mathbf{x}$. In general, a self-intersection of an image curve may be observed in the space x_1, x_2, x_3 as a non-generic situation, i.e. it is eliminated by weak variations in an original system, an observable or reconstruction parameters. Intuitively, one easily agrees that self-intersections of a curve in a three-dimensional space are very unlikely.

Thus, in our example, an equivalent description of the dynamics is achieved *for sure* only if the state vectors are reconstructed in the space of the dimension $D > 2d$.[2] This is the main contents of Takens' theorems. We stress that this is a *sufficient* condition. Sometimes, an equivalent description is possible even for $D = d$ as illustrated above. In practical modelling, Takens' theorems serve just as a psychological support, because they state that there is a finite model dimension D at which deterministic modelling should be appropriate. Technically, one tries different values of D, starting from small ones, and aims at obtaining a "good" model with as low dimension as possible to avoid difficulties related to the above-mentioned "curse of dimensionality".

---

[2] An equivalent description of a motion on a limit cycle is assured for $D = 3$ even if the cycle "lives" in an infinite-dimensional phase space.

10.1.1.2 Mathematical Details

To formulate the theorems in a more rigorous way, let us introduce some notations. Let an object be a dynamical system

$$
\begin{aligned}
\mathbf{y}(t_0 + t) &= \Phi_t(\mathbf{y}(t_0)), \\
\boldsymbol{\eta}(t) &= \mathbf{h}(\mathbf{y}(t)),
\end{aligned}
\tag{10.1}
$$

where $\mathbf{y}$ is a state vector, Φ_t is an evolution operator and $\mathbf{h}$ is a measurement function.[3] The vector of observables is finite-dimensional: $\boldsymbol{\eta} \in R^m$. We discuss further only the case of a scalar time series $\eta(t_i)$, i.e. $m = 1$. It is the most widespread situation, which is also the most difficult for modelling.

Manifold

Let the motion of the system occur at some manifold M of the finite dimension d that can be observed even for infinite-dimensional systems. *Manifold* is a generalisation of the concept of a smooth surface in the Euclidean space (Gliklikh, 1998; Makarenko, 2002; Malinetsky and Potapov, 2000; Sauer et al., 1991). Roughly speaking, a d-dimensional manifold M is a surface which can be locally parameterised with d Euclidean coordinates in the vicinity of any of its points. In other words, any point $p \in M$ together with its local neighbourhood $U(p)$ can be mapped on a d-dimensional fragment (e.g. a ball) of the space R^d in a one-to-one and continuous way. A corresponding image $\Psi : U \to \Psi(U)$ is called a *chart* of the neighbourhood. The continuous map Ψ is called a *homeomorphism*. Examples of two-dimensional manifolds in a three-dimensional Euclidean space are a sphere, a torus, a bottle with a handle, etc. (Fig. 10.3a), but not a (double) cone (Fig. 10.3b).
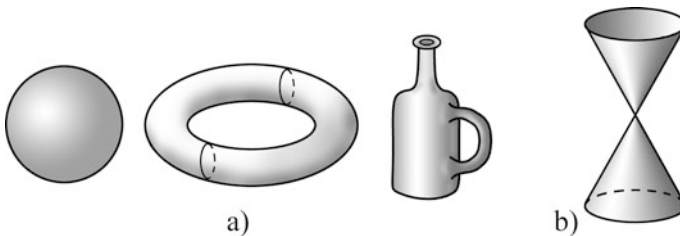

Fig. 10.3 Examples of the sets which (**a**) are manifolds and (**b**) is not a manifold

---

[3] If an object is a map $\mathbf{y}(t_{n+1}) = \mathbf{F}(\mathbf{y}(t_n))$, then an evolution operator $\Phi_{\Delta t}(\mathbf{y}(t_0))$ is just the function $\mathbf{F}$. If an object is a set of ODEs $d\mathbf{y}/dt = \mathbf{F}(\mathbf{y}(t))$, then the function $\Phi_t(\mathbf{y}(t_0))$ is the result of the integration of the ODEs over a time interval of length t. If an original system is given by a partial differential equation $\partial\mathbf{y}/\partial t = \mathbf{F}(\mathbf{y}, \partial\mathbf{y}/\partial\mathbf{r}, \partial^2\mathbf{y}/\partial\mathbf{r}^2, \ldots)$, where $\mathbf{r}$ is a spatial coordinate, then $\mathbf{y}$ is a vector belonging to an *infinite-dimensional* space of functions $\mathbf{y}(\mathbf{r})$ and Φ_t is an operator acting in that space.

If Ψ is an n times differentiable mapping with the n times differentiable inverse, then one says that M belongs to the class C^n. If $n \geq 1$, the mapping Ψ is called a *diffeomorphism*. If the manifold M is mapped on a manifold $S \in R^D$, $D \geq d$, via a diffeomorphism, then M and S are called *diffeomorphic* to each other. One says that S is an *embedding* of the manifold M into the Euclidean space R^D. Below, we speak of a bounded and closed M. Boundedness means that M can be included into a ball of a finite radius. Closedness means that all limit points of M belong to M. Such a manifold in a finite-dimensional space is called *compact*.

The Question and Notations

Each phase orbit of the system (10.1) $\mathbf{y}(t), 0 \leq t < \infty$, on a manifold M corresponds to a time realisation of an observable η: $\eta(t) = h(\mathbf{y}(t)), 0 \leq t < \infty$. The vector $\mathbf{y}(t_0)$ determines the entire future behaviour of the system (10.1), in particular, the entire realisation $\eta(t)$, $t \geq t_0$. Is it possible to determine a state on the manifold M at a time instant t_0 and, hence, the entire future evolution from a segment of the realisation $\eta(t)$ around t_0? In other words, can one "restore" a state of the system from the values of $\eta(t)$ on a finite-time interval? This is a key question and Takens' theorems give a positive answer under some conditions.

Let us introduce some notations necessary to formulate rigorously the time-delay embedding theorem. A vector $\mathbf{y}(t)$ corresponds to a D-dimensional vector $\mathbf{x}(t) = [\eta(t), \eta(t + \tau), \ldots, \eta(t + (D - 1)\tau)]$. Dependence of $\mathbf{x}$ on a simultaneous value of $\mathbf{y}$ is given by a *unique* mapping $\Psi : M \rightarrow R^D$ expressed via the evolution operator $\Phi_t : M \rightarrow M$ and the measurement function $h : M \rightarrow R$ as

$$\mathbf{x}(t) = \Psi(\mathbf{y}(t)) \equiv \begin{bmatrix} \Psi_1(\mathbf{y}(t)) \\ \Psi_2(\mathbf{y}(t)) \\ \cdots \\ \Psi_D(\mathbf{y}(t)) \end{bmatrix} = \begin{bmatrix} h(\mathbf{y}(t)) \\ h(\Phi_\tau(\mathbf{y}(t))) \\ \cdots \\ h(\Phi_{(D-1)\tau}(\mathbf{y}(t))) \end{bmatrix}. \qquad (10.2)$$

Smoothness of Ψ (continuity, differentiability, existence and continuity of the higher order derivatives) is determined by the smoothness of Φ_τ and h. An image of the manifold M under the mapping Ψ is a certain set $S \subset R^D$.

The above question can now be formulated as follows: Is Ψ a diffeomorphism? If yes, then S is an embedding of M and each vector $\mathbf{x}$ on S corresponds to a single vector $\mathbf{y}$ on M.[4] Then, $\mathbf{x}(t)$ can be used as a state vector to describe the dynamics on M and Eq. (10.1) can be rewritten as

$$\mathbf{x}(t_0 + t) = \varphi_t(\mathbf{x}(t_0)), \qquad (10.3)$$

---

[4] It means that for the same segments $[\eta(t), \eta(t+\tau), \ldots, \eta(t+(D-1)\tau)]$ encountered at different time instants t, one observes the same continuation (i.e. the same future). It gives a justification to the predictive *method of analogues* applied already by E. Lorenz. The method is based on the search of the time series segments, which "resemble" a current segment, in the past and subsequent usage of a combination of their "futures" as a forecast. In a modern formulation, it is realised with local models (Sect. 10.2.1).

where a new evolution operator is $\varphi_t(\mathbf{x}) = \Psi(\Phi_t(\Psi^{-1}(\mathbf{x})))$. Due to the diffeomorphism, local properties of the dynamics such as stability and types of fixed points and others are preserved. Each phase orbit $\mathbf{y}(t)$ on M corresponds to an orbit $\mathbf{x}(t)$ on S in a one-to-one way. If a system (10.1) has an attractor in M, then a system (10.3) has an attractor in S. Such characteristics as fractal dimension and Lyapunov exponents coincide for both attractors. In other words, the system (10.3) on the manifold S and the system (10.1) on the manifold M can be considered as two representations of the same dynamical system.

Obviously, the mapping Ψ (10.2) is not always a diffeomorphism. Thus, Fig. 10.2b gives an example where a smooth mapping Ψ has a non-unique inverse Ψ^{-1}. Another undesirable situation is encountered if Ψ^{-1} is unique but *non-differentiable* (Fig. 10.4). The latter property takes place at the *return point* on the set S. In its neighbourhood, the two-dimensional vector (y_1, y_2) cannot be used to describe the dynamics with a set of ODEs, since the return point would be a fixed point so that S could not be a limit cycle. Here, the differentiability properties of M and S differ due to non-differentiability of Ψ^{-1}.

Formulation of the Time-Delay Embedding Theorem

Coming back to the system (10.1) and the mapping (10.2), one can say that any one of the above mentioned situations can be encountered for some Φ, M, h, d, D and τ. Sets of self-intersections and return points on $S = \Psi(M)$ can be vast, which is very undesirable. However, one can also meet a "good" situation of embedding (Fig. 10.2a). The result formulated below was first obtained rigorously by Dutch mathematician Floris Takens (1981) and then generalised in Sauer et al. (1991). It shows under what conditions an embedding of an original compact d-dimensional manifold M in the space R^D is assured with the mapping (10.2). Takens' theorem
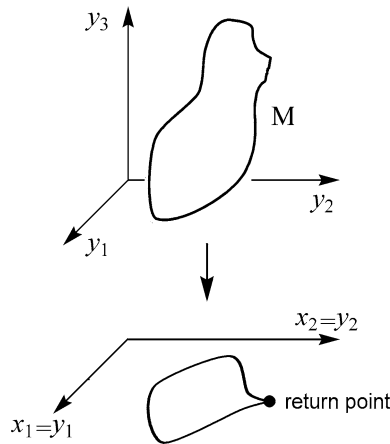

Fig. 10.4 The situation when a projection of a one-dimensional manifold M exhibits a return point on a plane

is related to the Whitney's embedding theorem (from the courses of differential geometry), which concerns arbitrary mappings. Takens' statement differs in that it concerns the special case of the mappings (10.2) determined by an evolution operator of a dynamical system.

Theorem 1 *Let M be a compact d-dimensional C^2 manifold. For almost any pair of functions Φ_t and h, which are twice continuously differentiable on M, the mapping $\Psi : M \rightarrow R^D$ given by the formula (10.2) is a diffeomorphism for almost any $\tau > 0$ and $D > 2d$.*

Comments

Diffeomorphism implies that an image of M under the mapping (10.2) is its embedding. The space R^D containing the image $S = \Psi(M)$ is called *embedding space*. The term "almost any pair" is understood by Takens in the sense of *genericity*. For instance, if for some Φ_t the mapping (10.2) does not provide an embedding, then there exists such an arbitrarily weak variation $\Phi_t + \delta\Phi_t$ that an embedding is achieved. More rigorously, generic properties are fulfilled on an intersection of open and everywhere dense sets. A metric analogue to genericity is *prevalence* (Sauer et al., 1991). "Almost any τ" should be understood in a similar way. In particular, if a limit cycle exists within M, the value of τ should not be equal to the period of that cycle, see Sauer et al. (1991) for more detail.

Discussion

Thus, if the dimension of the time-delay vector **x** (10.2) is high enough, one typically gets an embedding of the manifold M and can use **x** as a state vector of a deterministic model. It is possible to interpret the condition $D > 2d$ vividly as follows (Malinetsky and Potapov, 2000). To establish possible non-uniqueness of the mapping Ψ^{-1}, one must find such vectors $\mathbf{y}_1$ and $\mathbf{y}_2$ on M so that $\Psi(\mathbf{y}_1) = \Psi(\mathbf{y}_2)$. The latter equality is a set of D equations with $2d$ variables (d components for the two vectors $\mathbf{y}_1$ and $\mathbf{y}_2$ specifying their location on M). Roughly speaking, this set of equations has *typically* no solutions if the number of equations is greater than the number of variables, i.e. if $D > 2d$. This is the contents of Takens' theorem.

 We stress again that the condition $D > 2d$ is sufficient, but not necessary. If it is fulfilled, a diffeomorphism is assured. However, if one is "lucky", a good reconstruction can be obtained for lower D as in Fig. 10.1a, b, where an embedding of a one-dimensional manifold M is achieved at $D = 1$ and is not a degenerate case.

 What are those non-generic cases when the theorem is invalid? Let us indicate two examples (Malinetsky and Potapov, 2000):

(1) A measurement function is constant: $h(\mathbf{y}) = a$. This is a smooth function, but it maps the entire dynamics to a single point. This situation is almost surely eliminated via a weak variation in the measurement function, i.e. via adding an almost arbitrary "small" function of **y** to a.
(2) A system consisting of two unidirectionally coupled subsystems $d\mathbf{y}_1/dt = \mathbf{F}(\mathbf{y}_1, \mathbf{y}_2)$, $d\mathbf{y}_2/dt = \mathbf{G}(\mathbf{y}_2)$ when only the driving subsystem is observed,

i.e. $\eta = h(\mathbf{y}_2)$. In a non-synchronous regime, such an observable does not carry complete information about the driven subsystem $\mathbf{y}_1$. Therefore, an embedding of the original dynamics is not achieved. This situation is eliminated almost surely if an arbitrarily weak dependence on $\mathbf{y}_1$ is introduced into η.

Similar Theorems

A more general version of the theorem 1 is proven in Sauer et al. (1991). It concerns the *filtered embedding*, where coordinates of $\mathbf{x}$ are not just subsequent values of an observable but their linear combinations, which can be considered as outputs of a linear non-recursive filter.

Moreover, Takens proved a similar theorem for successive derivatives used as components of a state vector:

$$\mathbf{x}(t) = \begin{bmatrix} \eta(t) \\ \mathrm{d}\eta(t)\big/\mathrm{d}t \\ \dots \\ \mathrm{d}^{D-1}\eta(t)\big/\mathrm{d}t^{D-1} \end{bmatrix}, \tag{10.4}$$

where $D > 2d$. The theorem is formulated in the same way as theorem 1, but with stricter requirements to the smoothness of Φ_t and h. Namely, one demands continuous derivatives of the Dth order for each of these functions to assure the existence of the derivatives entering Eq. (10.4). If the latter derivatives are approximated with finite differences, then the relationship (10.4) becomes a particular case of the filtered embedding (Gibson et al., 1992).

In practice, one must always cope with noises. Takens' theorems are not directly related to such a case, although there are some generalisations (Casdagli et al., 1991; Stark et al., 1997). Nevertheless, the theorems are of significant value for practical modelling as discussed at the beginning of Sect. 10.1.

10.1.2 Practical Reconstruction Algorithms

10.1.2.1 Time-Delay Technique

This is the most popular reconstruction technique. One gets the vectors $\{\mathbf{x}_i = (\eta_i, \eta_{i+l}, \dots, \eta_{i+(D-1)l})\}_{i=1}^{N-(D-1)l}$ from an observed scalar time series $\{\eta_i = \eta(t_i)\}_{i=1}^{N}$, $t_i = i\Delta t$. Theoretically, the value of the time delay $\tau = l\Delta t$ can be almost arbitrary, but in practice one avoids both too small l, giving strongly correlated components[5] of the state vector, and too large l, introducing considerable complications into the geometrical structure of the reconstructed attractor.

---

[5] For $l = 1$ and a small sampling interval Δt, a reconstructed phase orbit stretches along the main diagonal, since it appears that $x_1 \approx x_2 \approx \dots \approx x_D$.

Therefore, it was suggested to choose the value of τ equal to the first zero of the autocorrelation function (Gibson et al., 1992), first minimum of the mutual information function (Fraser and Swinney, 1986) and so on (Liebert and Schuster, 1989). One also uses a *non-uniform embedding*, where time intervals between subsequent components of **x** are not the same, which is relevant for the dynamics with several characteristic timescales (Eckmann and Ruelle, 1985; Judd and Mees, 1998). For the dynamics representing alternating intervals of almost periodic and very complicated behaviour, one has developed the *variable embedding*, where a set of time delays depends on the location of **x** in the state space (Judd and Mees, 1998). Each of the ideas is appropriate for a specific kind of systems and does not assure successful results in general (Malinetsky and Potapov, 2000).

How to choose the model dimension D based on the analysis of an observed time series? There are different approaches including the false nearest neighbour technique (Kennel et al., 1992), the principal component analysis (Broomhead and King, 1986), the Grassberger and Procaccia method (Grassberger and Procaccia, 1983) and the "well-suited basis" approach (Landa and Rosenblum, 1989). Moreover, one should often try different values of D and construct model equations for each trial value until a "good" model is obtained. Then, the selection of D and even of the time delays can be a part of a united modelling procedure, rather than an isolated first stage.

10.1.2.2 False Nearest Neighbour Technique

It gives an integer-valued estimate of the attractor dimension. It is based on checking the property that a phase orbit reconstructed in the space of the sufficient dimension must not exhibit self-intersections. Let us illustrate the technique with a simple example of reconstruction from a time realisation of a sinusoid $\eta(t) = \sin t$, Fig. 10.5a.

At $D = 1$, i.e. $\mathbf{x}(t) = \eta(t)$, the reconstructed set lies on a line segment, Fig. 10.5b. Then, a data point at the instant t_k has the data points at the instants t_s and t_l as its close neighbours. However, the latter two states of an original system differ by the sign of the derivative of $\eta(t)$. In a two-dimensional space with $\mathbf{x}(t) = [\eta(t), \eta(t + \tau)]$, all the points go away from each other. However, the points at the instants t_k and t_l get *weakly* more distant, while the points at the instants t_k and t_s become very far from each other, Fig. 10.5c. Accordingly, one calls the neighbours at t_k and t_l "true" and the neighbours at t_k and t_s "false".

One of the version of the algorithm is as follows. At a trial dimension D, one finds a single nearest neighbour for each vector $\mathbf{x}_k$. After increasing D by 1, one determines which neighbours appear false and which ones are true. Then, one computes the ratio of the number of the false neighbours to the total number of the reconstructed vectors. This ratio is plotted versus D as in Fig. 10.5d. If this relative number of self-intersections reduces to zero at some value $D = D^*$, the latter is the dimension of the space, where an embedding of the original phase orbit is achieved. In practice, the number of the false neighbours becomes sufficiently small, starting from some "correct" value D^*, but does not decrease to zero due to noises and other
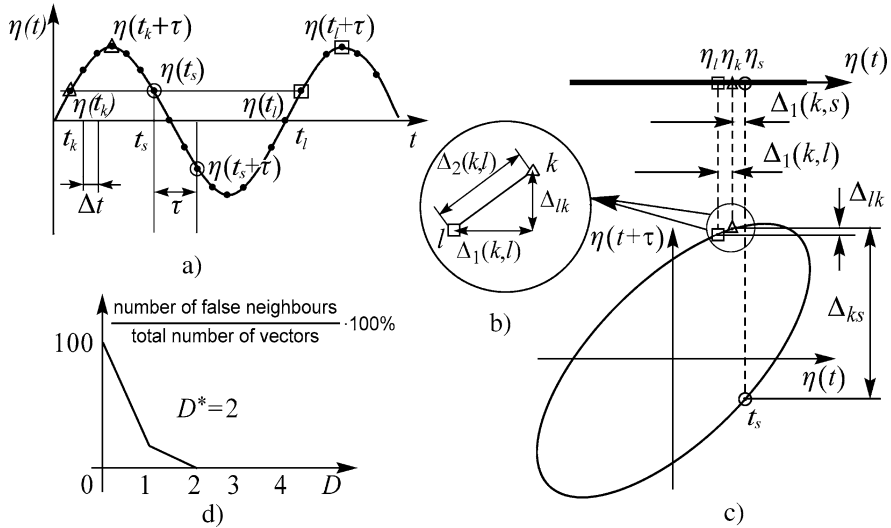
Fig. 10.5 An illustration to the false nearest neighbour technique: (**a**) a time realisation $\eta(t)$, where symbols indicate the data points $\eta(t_k)$, $\eta(t_s)$, $\eta(t_l)$, and the close values of η together with the points shifted by $\tau = 3\Delta t$; (**b**) an orbit reconstructed in a one-dimensional space; (**c**) an orbit reconstructed in a two-dimensional space; (**d**) the number of the false nearest neighbours divided by the total number of the reconstructed vectors in a time series versus the trial dimension of the reconstructed vectors D

factors. Then, D^* can be taken as a trial model dimension. It equals 2 for the example illustrated in Fig. 10.5d (see, e.g., Malinetsky and Potapov, 2000 for details).

10.1.2.3 Principal Component Analysis

It can be used both for the dimension estimation and for the reconstruction of state vectors. The technique is used in different fields and has many names. Its application to the reconstruction was suggested in Broomhead and King (1986). The idea is to rotate coordinate axes in a multidimensional space and choose a small subset of directions, along which the motion mainly develops.

For simplicity of notations, let the mean value of η be zero. The vectors $\mathbf{w}(t_i) = (\eta_i, \eta_{i+1}, \ldots, \eta_{i+k-1})$ of a sufficiently high dimension k are constructed. Components of these vectors are strongly correlated if the sampling interval is small. Figure 10.6 illustrates the case of a sinusoidal signal and the reconstruction of the phase orbit in a three-dimensional space ($k = 3$).

One performs a rotation in this space so that the directions of new axes (e.g. $\{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3\}$ in Fig. 10.6b, c) correspond to the directions of the most intensive motions in the descending order. Quantitatively, the characteristic directions and the extensions of an orbit along them are determined from the covariance matrix $\mathbf{\Theta}$ of the vector $\mathbf{w}$, which is a square matrix of the order k:
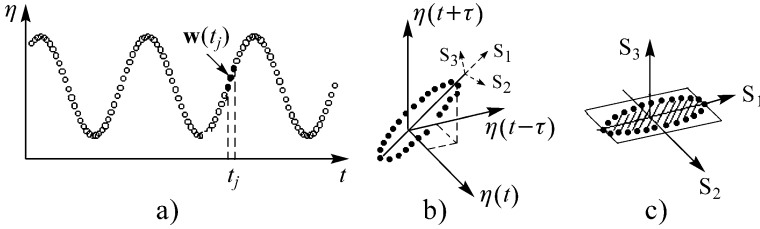
Fig. 10.6 Noise-free harmonic oscillations: (**a**) reconstruction of the time-delay vectors **w** of the dimension $k = 3$ from a scalar time series; (**b**) a reconstructed orbit is an ellipse stretched along the main diagonal of the space R^k; (**c**) a reconstructed orbit in a new coordinate system (after the rotation), where the component of the reconstructed vectors along the direction $\mathbf{s}_3$ is zero

$$\Theta_{i,j} = \sum_{n=0}^{N-k} \eta_{i+n}\eta_{j+n}, \, i, \, j = 1, \ldots, k$$

It is symmetric, real valued and positive definite. Hence, its eigenvectors constitute a complete orthonormal basis of the space R^k. Its eigenvalues are non-negative. Let us denote them as $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$ in the non-ascending order and the corresponding eigenvectors as $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_k$. The transformation to the basis $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_k$ can be performed via the coordinate change $\mathbf{x}'(t_i) = \mathbf{S}^{\mathrm{T}} \cdot \mathbf{w}(t_i)$, where $\mathbf{S}$ is a matrix with the columns $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_k$ and T means transposition. This is known in the theory of information as the Karhunen and Loeve transform. One can easily show that the covariance matrix of the components of the vector $\mathbf{x}'$ is diagonal:

$$\mathbf{\Theta}' = \mathbf{S}^{\mathrm{T}}\mathbf{\Theta}\mathbf{S} = \begin{bmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \sigma_k^2 \end{bmatrix}$$

i.e. the components of $\mathbf{x}'$ are uncorrelated, which is a sign of a "good" reconstruction. Each diagonal element σ_i^2 is the mean-squared value of the projection of $\mathbf{w}(t_i)$ onto the coordinate axis $\mathbf{s}_i$. The values σ_i^2 determine the extensions of the orbit along the respective directions. Rank of the matrix $\mathbf{\Theta}$ equals the number of non-zero eigenvalues (these are σ_1^2 and σ_2^2 for the situation shown in Fig. 10.6b, c) and the dimension of the subspace, where the motion occurs.

If a measurement noise is present, then all σ_i^2 are non-zero, since noise contributes to the directions, which are not explored by the deterministic component of an orbit. In such a case, the dimension can be estimated as the number D of considerable eigenvalues as illustrated in Fig. 10.7. Projections of $\mathbf{w}(t_i)$ onto the corresponding directions (i.e. the first D components of the vector $\mathbf{x}'$) are called its *principal components*. The remaining eigenvalues constitute the so-called noise floor and the respective components can be ignored. Thus, one gets D-dimensional vectors $\mathbf{x}(t_i)$ with coordinates $x_k(t_i) = \mathbf{s}_k \cdot \mathbf{w}(t_i), \, k = 1, \ldots, D$.
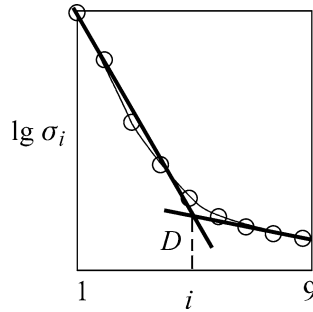
Fig. 10.7 Eigenvalues of the covariance matrix versus their order number: a qualitative illustration for $k = 9$. The "break point" D at the plot is an estimate of the dimension of an observed motion

If there is no characteristic break at the plot, then one increases a trial dimension k until the break emerges. The dimension estimate D is more reliable if the break is observed at the same value of D under the increase in k.

The principal component analysis is a particular case of the filtered embedding. It is very useful in the case of a considerable measurement noise, since it allows to filter the noise out to a significant extent: a realisation of $x_1(t)$ is "smoother" than that of the observable $\eta(t)$.

10.1.2.4 *Successive Derivatives* and Other Techniques

The usage of the reconstructed vectors (10.4) is attractive due to the clear physical meaning of their components. Many processes are described with a higher order model ODE (9.4), which involves successive derivatives of a single variable (Sect. 9.1). Some ODEs can be rewritten in such a form *analytically*, e.g. the *Roessler* system (see Sect. 10.2.2). However, an essential shortcoming in exploiting the vectors (10.4) is high sensitivity of the approach to the measurement noise, since the derivatives must be computed numerically (Sect. 7.4.2).

To summarise, there are many techniques to reconstruct a phase orbit. Having only a scalar time series, one can use successive derivatives or time delays. At that, several parameters can be selected in different ways, e.g. a time delay and a numerical differentiation scheme. Besides, one can use weighted summation (Brown et al., 1994; Sauer et al., 1991); and integration (Janson et al., 1998), which is advantageous for strongly non-uniform signals. One often exploits principal components, empirical modes, conjugated signal and phase (Sect. 6.4.3). It is possible to use combinations of all the techniques, e.g. to get some components via time delays, additional ones via integration and the rest via differentiation (Brown et al., 1994). In the case of a vector observable, one can restore variables from each of its components with any combination of the above techniques. Hence, the number of possible variants strongly increases (Cao et al., 1998; Celucci et al., 2003).

10.1.2.5 Choice of Dynamical Variables

Which of the state vector versions should be preferred? This question is important and attracts considerable attention (Letellier and Aguirre, 2002; Letellier et al., 1998b; Small and Tse, 2004). Trying all possible variants in turn and approximating a dependence $d\mathbf{x}/dt = \mathbf{f}(\mathbf{x},\mathbf{c})$ or $\mathbf{x}_{n+1} = \mathbf{f}(\mathbf{x}_n, \mathbf{c})$ for each of them is unfeasible, since solving the approximation problem often requires significant computational efforts and special approaches. Therefore, one should select a small number of reasonable sets of dynamical variables in advance. It can be done based on the preliminary analysis of experimental dependencies to be approximated (Rulkov et al., 1995; Smirnov et al., 2002). The respective procedures exploit an obvious circumstance that one needs such set of variables which would provide uniqueness and continuity of the dependencies $d\mathbf{x}/dt(\mathbf{x})$ or $\mathbf{x}_{n+1}(\mathbf{x}_n)$, where components of $\mathbf{x}$ are either observed or computed from the observed data.
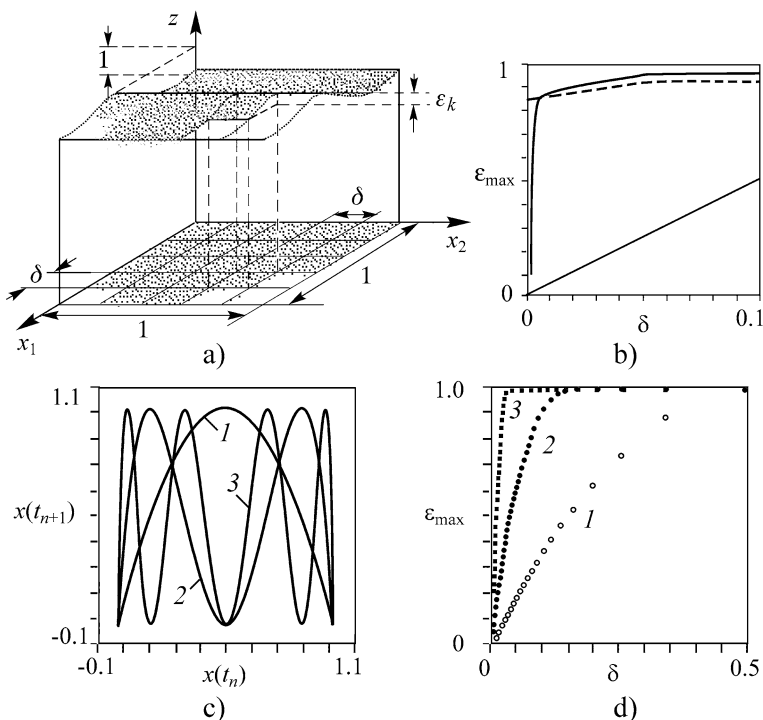

Fig. 10.8 Checking uniqueness and continuity of a dependence $z(\mathbf{x})$: (**a**) an illustration for $D = 2$; (**b**) typical plots $\varepsilon_{max}(\delta)$ for different choices of variables; the *straight line* is the best case, the *dashed line* corresponds to non-uniqueness or discontinuity of $z(\mathbf{x})$, the *broken line* corresponds to a complicated dependence $z(\mathbf{x})$ with the domains of fast and slow variations; (**c**) the plots of the first, the second and the third iterates of a quadratic map; (**d**) the plots $\varepsilon_{max}(\delta)$ for the dependence of $x(t_{n+1})$ on $x(t_n)$ in the three cases shown in panel (**c**)

Let us denote the left-hand side of model equations as $\mathbf{z}$: $\mathbf{z}(t) = \mathrm{d}\mathbf{x}(t)/\mathrm{d}t$ for a set of ODEs $\mathrm{d}\mathbf{x}(t)/\mathrm{d}t = \mathbf{f}(\mathbf{x}(t))$ and $\mathbf{z}(t_n) = \mathbf{x}(t_{n+1})$ for a map $\mathbf{x}(t_{n+1}) = \mathbf{f}(\mathbf{x}(t_n), \mathbf{c})$. After the reconstruction of the vectors $\mathbf{x}$ from an observable η, one should get a time series $\{\mathbf{z}(t_i)\}$. It is achieved via the numerical differentiation of the series $\{\mathbf{x}(t_i)\}$ for a set of ODEs and via the time shift of $\{\mathbf{x}(t_i)\}$ for a map. Further, it is necessary to check whether close vectors $\mathbf{x}(t_1)$ and $\mathbf{x}(t_2)$ correspond to close simultaneous vectors $\mathbf{z}(t_1)$ and $\mathbf{z}(t_2)$. A possible procedure is as follows (Smirnov et al., 2002).

A domain V containing the set of vectors $\{\mathbf{x}(t_i)\}$ is divided into equal hypercubic cells with the side δ (Fig. 10.8a). One selects all cells $s_1, \ldots, s_M$ such that each s_k contains more than one vector $\mathbf{x}(t_i)$. Thus, the cell s_k corresponds also to more than one vector $\mathbf{z}(t_i)$. The difference between the maximal and the minimal value of z (one of the components of the vector $\mathbf{z}$) over the cell s_k is called *local scattering* ε_k. Suitability of the quantities $\mathbf{x}$ and z for the global modelling is assessed from the maximal local scattering $\varepsilon_{\max} = \max\limits_{1 \leq k \leq M} \varepsilon_k$ and the plot $\varepsilon_{\max}(\delta)$. To construct a global model, one should choose variables such that the plot $\varepsilon_{\max}(\delta)$ gradually tend to the origin (Fig. 10.8b, straight line) for each of the approximated dependencies $z_k(\mathbf{x})$, $k = 1, \ldots, D$.

Moreover, it is desirable to provide the least slope of the plot $\varepsilon_{\max}(\delta)$, since one needs then a simpler approximating function, e.g. a low-order polynomial. This is illustrated in Fig. 10.8c, d, where the next value of an observable is shown versus the previous one and an observable is generated by the first, the second or the third iterate of the quadratic map $x(t_{n+1}) = \lambda - x^2(t_n)$. The plot for the first iterate is the "least oscillating" and, therefore, the slope of $\varepsilon_{\max}(\delta)$ is the smallest. In this case, one can get a "good" model most easily, since it requires the usage of only the second-order polynomial. At that, the eighth-order polynomial is necessary to describe the third iterate of the map. These three cases are even more different in respect of the reconstruction difficulties in the presence of noise. Additional details are given in Smirnov et al. (2002).

10.2 Multivariable Function Approximation

10.2.1 Model Maps

The time-delay embedding is typically used to construct multidimensional model maps

$$x_n = f(x_{n-D}, x_{n-D+1}, \ldots, x_{n-1}, \mathbf{c}), \qquad (10.5)$$

where the variable x corresponds to an observable and the time delay is set equal to $l = 1$ for the simplicity of notations. Various choices of the function f in Eq. (10.5) are possible. One says that the function f, which is specified in a closed form (Sect. 3.5.1) in the *entire* phase space, provides a *global approximation*. Then, one also speaks of a *global model* and a *global reconstruction*. Alternatively, one can

use a *local approximation*, i.e. the function f with its own set of parameter values for each small domain of the phase space. Then, one speaks of a *local model*.

In practice, a global approximation with algebraic polynomials often performs badly already for two-variable functions (Bezruchko and Smirnov, 2001; Casdagli, 1989; Judd and Mees, 1995; Pavlov et al., 1997. A pronounced feature is that the number of model parameters and the model prediction errors rise *quickly* with the model dimension D. The techniques with such a property are characterised as *weak approximation*. They also include trigonometric polynomials and wavelets. In practical black box modelling, one often has to use D at least as large as 5–6. Therefore, algebraic polynomials are not widely used.

Much efforts of researchers have been spent to *strong approximation* approaches, i.e. the approaches which are relatively insensitive to the rise in D. They include local techniques with low-order polynomials (Casdagli, 1989; Abarbanel et al., 1989; Farmer and Sidorowich, 1987; Kugiumtzis et al., 1998; Sauer, 1993; Schroer et al., 1998), radial, cylindrical, and elliptical basis functions (Giona et al., 1991; Judd and Mees, 1995, 1998; Judd and Small, 2000; Small and Judd, 1998; Small et al., 2002; Smith, 1992) and artificial neural networks (Broomhead and Lowe, 1988; Makarenko, 2003; Wan, 1993). All these functions usually contain many parameters so that a careful selection of the model structure and the model size is especially important to avoid overfitting (see Sects. 7.2.3 and 9.2).

10.2.1.1 A Generalised Polynomial

To construct a global model (10.5), one selects the form of f and estimates its parameters via the ordinary LS technique:

$$S(\mathbf{c}) = \sum_{i=D+1}^{N} (\eta_i - f(\eta_{i-D}, \eta_{i-D+1}, \ldots, \eta_{i-1}, \mathbf{c}))^2 \to \min. \qquad (10.6)$$

To simplify computations, it is desirable to select the function f, which is linear in its parameters $\mathbf{c}$. This is the case for a function

$$f(\mathbf{x}) = \sum_{k=1}^{P} c_k f_k(\mathbf{x}) \qquad (10.7)$$

which is called a *generalised polynomial* with respect to a *set of basis functions* $f_1, f_2, \ldots, f_P$. Then, the problem (10.6) is linear so that the local minima problem is avoided. A particular case of such an approach is represented by an algebraic polynomial. A trial polynomial order is increased until an appropriate model is obtained or another condition is fulfilled as discussed in Sect. 7.2.3.

10.2.1.2 Radial Basis Functions

These are functions $\phi_k(\mathbf{x}) = \phi(\|\mathbf{x} - \mathbf{a}_k\| / r_k)$, where $\| \cdot \|$ denotes a vector norm, a "mother" function ϕ is usually represented by a well-localised function,
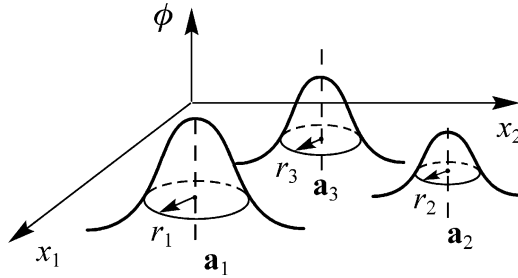
Fig. 10.9 The plots of two-variable radial basis functions (qualitative outlook): three "Gaussian hills"

e.g., $\phi(y) = \exp(-y^2/2)$, the quantities $\mathbf{a}_k$ are called "centres" and r_k are "radii". The model function f is taken to be a generalised polynomial with respect to the set of functions ϕ_k : $f(\mathbf{x,c}) = \sum_k c_k \phi_k(\mathbf{x})$. Each term essentially differs from zero only within the distance about r_k from the centre $\mathbf{a}_k$ (Fig. 10.9). Intuitively, one can see that such a superposition can approximate a very complicated smooth relief. Radial basis functions possess many attractive properties and are often used in the approximation practice. However, we stop their discussion here and describe in more detail two approaches, which are even more widespread.

10.2.1.3 Artificial Neural Networks

Models with ANNs (Sect. 3.8) are successfully used to solve many tasks. Their right-hand side is represented by a *composition* of basis functions, rather than by their sum. In contrast to the generalised polynomial (10.7), ANNs are almost always non-linear with respect to the estimated parameters. This is the "most universal" way of the multivariable function approximation in the sense that along with a firm theoretical justification, it successfully performs in practice.

Let us introduce an ANN formally (in addition to the discussion of Sect. 3.8) with an example of a multilayer perceptron. Let $\mathbf{x} = (x_1, \ldots, x_D)$ be an argument of a multivariable function f. Let us consider the set of functions $f_j^{(1)}(\mathbf{x})$:

$$f_j^{(1)}(\mathbf{x}) = \phi \left(\sum_{i=1}^{D} w_{j,i}^{(0)} \cdot x_i - \upsilon_j^{(0)} \right), \qquad (10.8)$$

where $j = 1, \ldots, K_1$, the constants $w_{j,i}^{(0)}$ are called weights, $\upsilon_j^{(0)}$ are thresholds, ϕ is an activation function. The function ϕ is usually non-linear and has a step-like plot. One often uses the *classical sigmoid*: $\phi(x) = 1/(1 - e^{-x})$. Let us say that each function $f_j^{(1)}$ represents an output of *a standard formal neuron* with an order number j, whose input is the vector $\mathbf{x}$. Indeed, a living neuron sums up external stimuli and reacts to them in a threshold way that determines the properties of the
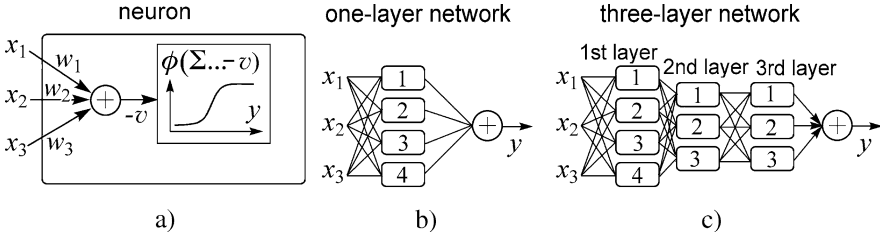
Fig. 10.10 Illustrations to artificial neural networks: (**a**) a standard formal neuron; (**b**) a scheme for a one-layer ANN with a single output; a single rectangle denotes a single neuron; (**c**) a scheme for a *multi-layer ANN* with a single output

function ϕ (Fig. 10.10a). The set of functions $f_1^{(1)}, \ldots, f_{K_1}^{(1)}$ is called the set of first-layer neurons (Fig. 10.10b). The values of $f_j^{(1)}$ are the outputs of the first-layer neurons. Let us denote them as vector $\mathbf{y}^{(1)}$ with components $y_j^{(1)} = f_j^{(1)}(\mathbf{x})$.

By defining the function f as a linear combination of $f_j^{(1)}$, one gets a *one-layer ANN* model

$$f(\mathbf{x}) = \sum_{j=1}^{K_1} w_j^{(1)} y_j^{(1)} - \upsilon^{(1)} \equiv \sum_{j=1}^{K_1} w_j^{(1)} \phi \left(\sum_{i=1}^{D} w_{j,i}^{(0)} x_i - \upsilon_i^{(0)} \right) - \upsilon^{(1)}, \quad (10.9)$$

where $w_j^{(1)}$, $\upsilon^{(1)}$ are additional weights and a threshold, respectively. The number of free parameters is $P = K_1(D+1)+1$. This representation resembles the generalised polynomial (10.7), but the ANN depends on $w_{j,i}^{(0)}$ and $\upsilon_j^{(0)}$ in a non-linear way.

By induction, let us consider a set of K_1-variable functions $f_k^{(2)}$, $k = 1, \ldots, K_2$, of the form (10.9). These are second-layer neurons, whose input is the output $\mathbf{y}^{(1)}$ of the first-layer neurons (Fig. 10.10c). Let us denote their output values as a vector $\mathbf{y}^{(2)}$ of the dimension K_2 and define the function f as a linear combination of the output values of the second-layer neurons:

$$f(\mathbf{x}) = \sum_{j_2=1}^{K_2} w_{j_2}^{(2)} \phi \left(\sum_{j_1=1}^{K_1} w_{j_2,j_1}^{(1)} \phi \left(\sum_{i=1}^{D} w_{j_1,i}^{(0)} x_i - \upsilon_{j_1}^{(0)} \right) - \upsilon_{j_2}^{(1)} \right) - \upsilon^{(2)}. \quad (10.10)$$

This is a *two-layer ANN* which involves compositions of functions. The latter circumstance makes it essentially different from the *pseudo-linear model* (10.7). Increasing the number of layers is straightforward.

To solve the approximation problems, one most often uses two-layer ANNs (10.10) and sometimes three-layer ones (Malinetsky and Potapov, 2000). The increase in the number of layers does not lead to a significant improvement. Improvements can be more often achieved via the increase in the number of neurons in each layer K_1, K_2. A theoretical base underlying the usage of the ANNs is

the *generalised approximation theorem* (*Weierstrass' theorems* are its partial cases), which states that any continuous function can be arbitrarily accurately uniformly approximated with an ANN. A rigorous exposition is given, e.g., in Gorban' (1998).

The procedure for the estimation of parameters in an ANN via the minimisation (10.6) is called *learning* of an ANN. This is a problem of multidimensional non-linear optimisation. There are special "technologies" for its solution including backward error propagation algorithm, scheduled learning, learning with noise, stochastic learning (genetic algorithms and simulated annealing), etc. An ANN may contain many superfluous elements so that it is very desirable to make the structure of such a model (i.e. a network architecture) "more compact". For that, one excludes from a network those neurons whose weights and thresholds remain almost unchanged during the learning process.

If several alternative ANNs with different architectures are obtained from a training time series, then the best of them is usually selected according to the least test error (Sect. 7.2.3). To get an "honest" indicator of its predictive ability, one uses one more data set (not the training one and not the test one, since both of them are used to get the model), which is called a validation time series.

An advantage of an ANN over other constructions in empirical modelling is not easy to understand (Malinetsky and Potapov, 2000). If one gets an ANN, which performs well, it is usually unclear why this model is so good. It is the problem of the "network transparency"; a model of a black box is also a black box in a certain sense. Yet, even such a model can be investigated numerically and used to generate predictions.

10.2.1.4 Local Models

Local models are constructed so as to minimise the sum of squares like Eq. (10.6) over a local domain of the phase space. Thus, to predict the value η_{i+D}, which follows a current state $\mathbf{x}_i = [\eta_i, \eta_{i+1}, \ldots, \eta_{i+D-1}]$, one uses the following procedure. One finds k *nearest neighbours* of the vector $\mathbf{x}_i$ among all the vectors in the training time series (in the past). These are vectors with time indices n_j, whose distance to $\mathbf{x}_i$ are smallest:

$$\left\| \mathbf{x}_{n_j} - \mathbf{x}_i \right\| \leq \left\| \mathbf{x}_l - \mathbf{x}_i \right\|, \, j = 1, \ldots, k, l \neq i, l \neq n_j. \qquad (10.11)$$

They are also called the *analogues* of $\mathbf{x}_i$, see Figs. 10.11 and 10.12.

The values of an observable, which followed the neighbours $\mathbf{x}_{n_j}$ in the past, are known. Hence, one can construct the model (10.5) from those data. For that, one typically uses a simple function $f(\mathbf{x},\mathbf{c})$, whose parameters are found with the ordinary LS technique (Sect. 8.1.1), although more sophisticated estimation techniques are available (Kugiumtzis et al., 1998). An obtained function $f(\mathbf{x}, \hat{\mathbf{c}}_i)$ is used to generate a prediction of the value η_{i+D} according to the formula $\hat{\eta}_{i+D} = f(\mathbf{x}_i, \hat{\mathbf{c}}_i)$, see Fig. 10.12. The vector $\hat{\mathbf{c}}_i$ has a subscript i, since it corresponds only to the vicinity of the vector $\mathbf{x}_i$. According to the so-called *iterative forecast* (Sect. 10.3), one predicts the next value η_{i+D+1} by repeating the same procedure of the neighbour search and
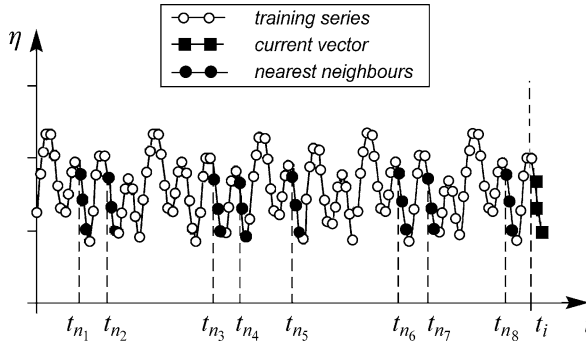
Fig. 10.11 Illustration for a three-dimensional local model: nearest neighbours (*filled circle*) of a vector $\mathbf{x}_i$ (*filled squares*) found in a training time series
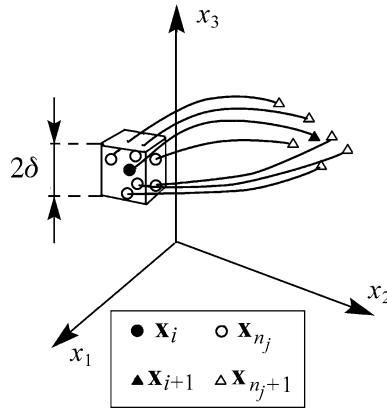

Fig. 10.12 Nearest neighbours (*open circles*) of a vector $\mathbf{x}_i$ (*filled circles*) and the vectors following them in time (*open triangles*). The latter are used to predict the vector $\mathbf{x}_{i+1}$ (*filled triangle*)

parameter estimation for the model state vector $\hat{\mathbf{x}}_{i+1} = (\eta_{i+1}, \ldots, \eta_{i+D-1}, \hat{\eta}_{i+D})$. Thereby, one gets a new forecast $\hat{\eta}_{i+D+1} = f(\hat{\mathbf{x}}_{i+1}, \hat{\mathbf{c}}_{i+1})$ and so on.

Relying on the Taylor polynomial expansion theorem, one uses such approximating functions as the constant $f(\mathbf{x}, \mathbf{c}) = c_1$, the linear function $f(\mathbf{x}, \mathbf{c}) = c_1 + \sum_{j=1}^{D} c_{j+1}x_j$ and the polynomials of a higher order K. On the one hand, an approximation error is smaller if the neighbours are closer to the current vector. Therefore, it should decrease with an increasing time series length, since closer returns to the vicinity of each vectors would occur. On the other hand, one should use a greater number of neighbours k to reduce the noise influence. Thus, a trade-off is necessary: one cannot use *too distant* "neighbours" to keep an error of approximation with a low-order polynomial small, but one cannot take *too a small number* of the nearest neighbours as well.

Local constant models are less requiring to the amount of data and more robust to noise since they contain a single free parameter for each small domain. *Local linear models* are superior for weak noise and sufficiently long time series: The concrete values depend on the necessary value of D. To construct a local linear model, one must use at least $k = D+1$ neighbours, since a model contains $D+1$ free parameters for each "cell". Its approximation error scales as λ^2 for a very long time series and "clean" data, where λ is a characteristic distance between the nearest neighbours in the time series. Local models with higher order polynomials are rarely used.

For the above local models, the function f is usually discontinuous, since different "pieces" of local approximation are not matched with each other. Sometimes, it leads to undesirable peculiarities of the model dynamics, which are not observed for the original system. One can avoid the discontinuity via triangulation (Small and Judd, 1998). At that, a model acquires some properties of the global approximation (f becomes continuous) and is described as a global-local model. However, the triangulation step makes the modelling procedure much more complicated.

Local models are often exploited for practical predictions. There are various algorithms taking into account delicate details. In essence, this is a contemporary version of the predictive "method of analogues" (Fig. 10.11).

10.2.1.5 Nearest Neighbour Search

It can take much time if a training time series is long. Thus, if one naively computes distances from a current vector to each vector in the time series and selects the least ones, the number of operations scales as N^2. Below, an efficient search algorithm based on the preliminary partition of the training set into cells (Kantz and Schreiber, 1997) is described.

The above local models are characterised by *fixed number of neighbours*. Let us consider another (but similar) version: local models with *fixed neighbourhood size*. The difference is that one looks for the neighbours of a vector $\mathbf{x}_i$, which are separated from $\mathbf{x}_i$ by a distance not greater than δ (Fig. 10.12):

$$\left\| \mathbf{x}_{n_j} - \mathbf{x}_i \right\| \leq \delta. \tag{10.12}$$

The number of neighbours may differ for different $\mathbf{x}_i$, but it should not be less than $D+1$. If there are too a few neighbours, one should increase the neighbourhood size δ. Under a fixed time series length, an optimal neighbourhood size rises with the noise level and the model dimension. An optimal δ is selected via trials and errors. One can use any norm of a vector in Eq. (10.11) or (10.12). The most convenient one is $\|\mathbf{x}\| = \max\{|x_1|, |x_2|, \ldots, |x_D|\}$, since it is quickly computed. In such a case, the neighbourhood (10.12) is a cube with the side of length 2δ.

Computation of the distances from a reference vector $\mathbf{x}_i$ to all the vectors in the training time series would require a lot of time. It is desirable to skip the vectors, which deliberately cannot be close neighbours of $\mathbf{x}_i$. For that, one preliminarily *sorts* all the vectors based on the first and the last of their D coordinates. Let $\eta_{\min}$ and $\eta_{\max}$ be the minimal and maximal values, respectively, of an observable over
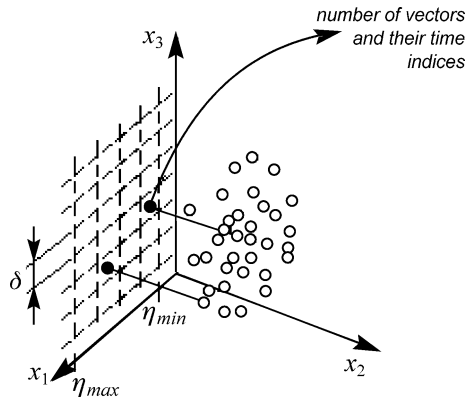
Fig. 10.13 Vectors of a training time series are sorted based on their first and last coordinates: one creates a square array whose elements contain information about the number of vectors in the respective cell and their time indices

the training series. Then, the corresponding orbit on the plane (x_1, x_D) lies within the square with the sides belonging to the straight lines defined by the equations $x_1 = \eta_{\min}, x_1 = \eta_{\max}, x_D = \eta_{\min}, x_D = \eta_{\max}$ (Fig. 10.13). The square is divided into square cells of size δ. One determines to which cell each vector falls and creates an array, whose elements correspond to the cells. Each element contains time indices of the vectors falling into the respective cell. To find the nearest neighbours of a vector **x**, one checks into which cell it falls and computes distances from **x** to the vectors belonging to the same cell or the cells having a common vertex with it. In total, one must check at most nine cells. This algorithm speeds up the process of neighbour search and requires the order of N operations if there are no too densely and too rarely "populated" domains in the reconstructed phase space.

10.2.1.6 A Real-World Example

Chaotic dynamics of a laser (Fig. 10.14) was suggested as a test data set for the *competition in time series prediction* at the conference in Santa-Fe in 1993 (Gerschenfeld and Weigend, 1993). Competitors had to provide a continuation of the time series, namely to predict the next 100 data points based on 1000 given data points. A winner was Eric Wan, who used a feed-forward ANN-based model of the form (10.5) (Wan, 1993).
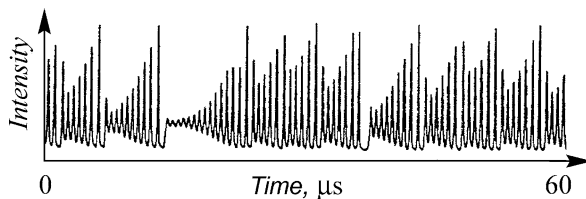

Fig. 10.14 Data from a ring laser in a chaotic regime (Hubner et al., 1993), $\Delta t = 40\,\text{ns}$

Figure 10.15a shows the observed time series (thin lines) and predictions (thick lines) with the ANN-based model for different starting instants (Wan, 1993). Similar prediction accuracy is provided by a local linear model (Sauer, 1993), see Fig. 10.15b. Accuracy of predictions for different intervals depends on how accurately the models predict an instant of switching from the high-amplitude oscillations to the low-amplitude ones. Thus, the local linear model performs worse than the ANN for the starting instants 1000 and 2180 and better for the three others. Local linear models appear to reproduce better some dynamical characteristics of the process and its long-term behaviour (Gerschenfeld and Weigend, 1993); the top panel in Fig. 10.15b shows an iterative forecast over 400 steps, which agrees well with the experimental data. The long-term behaviour is described a bit worse with the ANN (Sauer, 1993; Wan, 1993).

It is interesting to note that Eric Wan used ANN with 1105 free parameters trained on only 1000 data points. The number of parameters was even greater than the number of data points that usually leads to an ill-posed problem in statistics.
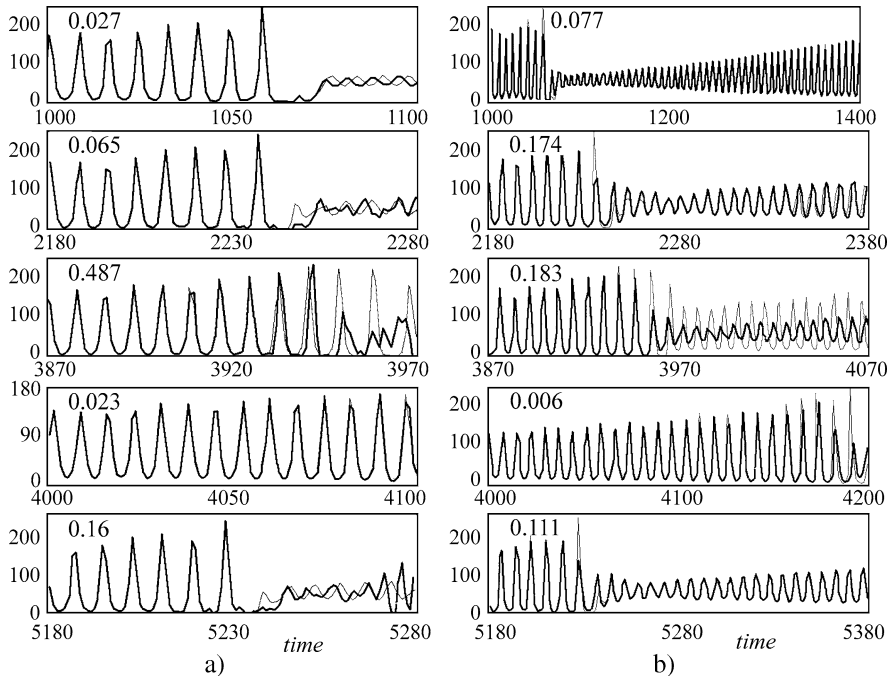
Fig. 10.15 Forecast of the laser intensity: (**a**) an ANN-based model (Wan, 1993); (**b**) a local linear model (Sauer, 1993). Laser intensity is shown along the vertical axis in arbitrary units. Time is shown along the horizontal axis in the units of sampling interval. The *thin lines* are the observed values and the *thick lines* are the predictions. Different panels correspond to predictions starting from different time instants: 1000, 2,180, 3,870, 4,000, and 5,180. The number at the top left corner of each panel is the normalised mean-squared prediction error over the first 100 data points of the respective data segment. The *top panels* show the segment proposed for the competition in Santa-Fe: the data points 1001–1100 were to be predicted

However, an ANN is a highly non-linear model function so that the number of "effective degrees of freedom" ("effectively free" parameters) is not equal to the full number of estimated parameters. There are constraints imposed on the possible model behaviour by the topology of the ANN. The author performed cross-validation (Sect. 7.2.3) by using only 900 data points as the training time series and 100 data points as the test one. He stated that there were no signs of overfitting when the size of the network was changed. Still, he noted an indirect sign of overfitting: after a good short-term forecast over several dozens of time steps, the ANN-based model exhibited "noisier" long-term behaviour than it is observed in the original data (Fig. 10.14). This can also be the reason why the local linear model of Tim Sauer appeared superior for the description of the long-term behaviour. An overall judgement seems to be that an ANN and a local linear model are approximately equally good in the example considered.

A number of applications of local models to predictions can be found, e.g., in Farmer and Sidorowich (1987); Kantz and Schreiber (1997); Kugiumtzis et al. (1998). ANN-based models are used probably more often, since they are less demanding with respect to the time series length and noise level. There are examples of their successful applications even to geophysical and financial predictions (Makarenko, 2003). Forecasts with other models of the form (10.5) are described, e.g., in Judd and Small (2000); Small and Judd (1998).

10.2.2 Model Differential Equations

In construction of model ODEs from a scalar time series, one often gets state vectors with successive derivatives $[\eta, \mathrm{d}\eta/\mathrm{d}t, \ldots, \mathrm{d}^{D-1}\eta/\mathrm{d}t^{D-1}]$ and uses the *standard form of model equations* (Sect. 3.5.3):

$$\mathrm{d}^{D}x\Big/\mathrm{d}t^{D} = f(x, \mathrm{d}x/\mathrm{d}t, \ldots, \mathrm{d}^{D-1}x\Big/\mathrm{d}t^{D-1}, \mathbf{c}) \qquad (10.13)$$

where $\eta = x$. The approximating function f is selected in the same way as described above for the models (10.5). Here, one more often observes "smoother" dependencies to be approximated and uses algebraic polynomials

$$f(x_1, x_2, ..., x_D, \mathbf{c}) = \sum_{l_1, l_2, ..., l_D = 0}^{K} c_{l_1, l_2, ..., l_D} \prod_{j=1}^{D} x_j^{l_j}, \quad \sum_{j=1}^{D} l_j \leq K. \qquad (10.14)$$

Model ODEs (10.13) with ANNs and other functions are rarely used (Small et al., 2002).

Some systems can be rewritten in the standard form (10.13) analytically. Thus, the *Roessler system,* which is a paradigmatic chaotic system, reads as

$$\begin{aligned}
\mathrm{d}x\big/\mathrm{d}t &= -y - z, \\
\mathrm{d}y\big/\mathrm{d}t &= x + C_1 y, \\
\mathrm{d}z\big/\mathrm{d}t &= C_2 - C_3 z + xz.
\end{aligned} \tag{10.15}$$

One can show that it can be reduced to a three-dimensional system with successive derivatives of y and the second-order polynomial in the right-hand side:

$$\begin{aligned}
\mathrm{d}x_1\big/\mathrm{d}t &= x_2, \\
\mathrm{d}x_2\big/\mathrm{d}t &= x_3, \\
\mathrm{d}x_3\big/\mathrm{d}t &= -C_2 - C_3 x_1 + (C_1 C_3 - 1)x_2 + (C_1 - C_3)x_3 - C_1 x_1^2 + \\
&\quad + (C_1^2 + 1)x_1 x_2 - C_1 x_1 x_3 - C_1 x_2^2 + x_2 x_3,
\end{aligned} \tag{10.16}$$

where $x_1 = y$. With successive derivatives of x and z, one gets the equations similar to Eq. (10.16), but with rational functions in the right-hand side (Gouesbet and Letellier, 1994).

The standard models are used in practice (see, e.g., Gouesbet, 1991; Gouesbet and Letellier, 1994; Gouesbet et al., 2003b Gribkov et al., 1994a, 1994b; Letellier et al., 1995, 1997, 1998a; Pavlov et al., 1999), but successful results are quite rare. The usage of the model structure (10.13) with an algebraic polynomial (10.14) often leads to quite cumbersome equations.

In construction of ODEs from a vector time series, everything is similar, but one looks for D scalar-valued functions rather than for a single one (analogous to the "grey box" case described in Sect. 9.1).

10.3 Forecast with Various Models

The novel techniques developed within the non-linear dynamics framework and discussed above are often the most efficient ones to predict complex real-world processes. Especially, this is the case when it appears sufficient to use a low-dimensional model. Multiple non-linear dynamics techniques can be distinguished according to different aspects: iterative forecast versus direct forecast; model maps of different kinds, e.g., global models versus local models; model maps versus model ODEs. Below, advantages and disadvantages of different approaches are briefly discussed and compared. To run ahead, the most efficient tool is usually a model map (global or local one depending on the amount of available data and the necessary model dimension) with an iterative, direct or combined prediction technique (depending on the required advance time). However, let us start the comparison with "older" approaches.

10.3.1 Techniques Which Are not Based on Non-linear Dynamics Ideas

For very simple signals, good predictions can be achieved even with explicit functions of time (Chap. 7). For stationary irregular signals without signs of

non-linearity, the most appropriate tool is linear ARMA models (Sects. 4.4 and 8.1) despite their possibilities being quite limited. Thus, one can show that prediction with a linear ARMA model can be reasonably accurate only over an interval of the order of the process correlation time τ_{cor} (Anosov et al., 1995), i.e. a characteristic time of the autocorrelation function decay (see Sect. 2.3, Fig. 2.8).

For a chaotic time series, τ_{cor} can be quite small making possible only quite short-term predictions with an ARMA model. Although a chaotic process cannot be accurately predicted far in the future in principle, the prediction time can be much greater than τ_{cor} for non-linear models. It can be roughly estimated with the formula (2.34): $\tau_{\mathrm{pred}} = \left(1/2\Lambda_1\right) \ln \left(\sigma_x^2 / \left(\sigma_v^2 + \sigma_\mu^2 + \sigma_{\Delta M}^2\right)\right)$. If noises and model errors are not large, then τ_{pred} can strongly exceed the correlation time roughly estimated as $\tau_{\mathrm{cor}} \sim 1/\Lambda_1$ (see Sect. 2.4 and Kravtsov, 1989; Smith, 1997).

10.3.2 Iterative, Direct and Combined Predictors

One can predict the values of an observable following the last value in a time series η_N with a model (10.5) via the above-mentioned (Sect. 10.2.1) *iterative* way:

(i) One-step-ahead prediction is generated as

$$\hat{\eta}_{N+1} = f(\mathbf{x}_{N-D+1}, \hat{\mathbf{c}}) = f(\eta_{N-D+1}, \eta_{N-D+2}, \dots, \eta_N, \hat{\mathbf{c}}). \qquad (10.17)$$

(ii) The predicted value $\hat{x}_{N+1}$ is considered as the last coordinate of the new state vector $\hat{\mathbf{x}}_{N-D+2} = (\eta_{N-D+2}, \eta_{N-D+3}, \dots, \eta_N, \hat{\eta}_{N+1})$;

(iii) The vector $\hat{\mathbf{x}}_{N-D+2}$ is used as an argument of the function f to generate a new forecast $\hat{\eta}_{N+2} = f(\hat{\mathbf{x}}_{N-D+2}, \hat{\mathbf{c}})$ and so on.

Thus, one gets a forecast $\hat{\eta}_{N+l}$ over any number of steps l ahead.

Alternatively, one can get an *l*-step-ahead forecast by constructing a model which *directly* approximates dependence of η_{i+l} on $(\eta_{i-D+1}, \eta_{i-D+2}, \dots, \eta_i)$ instead of making l iterations of the map (10.5). The form of such a dependence can be rather complicated for the chaotic dynamics and large l due to high sensitivity of the future behaviour η_{i+l} to the initial conditions $\mathbf{x}_{i-D+1}$. As a result, for a very large l, one gets an approximately constant model function $f \approx \langle \eta \rangle$ and, hence, a low prediction accuracy. However, the direct approach can be advantageous for moderate values of l.

How does the prediction time for both approaches depend on the time series length N and other factors? This question can be answered theoretically for a local model with a polynomial of an order K. According to the estimates of Casdagli (1989) and Farmer and Sidorowich (1987), the prediction error grows with l as $\sigma_{\Delta M} \cdot e^{\Lambda_1 l \Delta t}$ for the iterative technique and as $\sigma_{\Delta M} \cdot e^{(K+1)Hl\Delta t}$ for the direct technique, where H is the sum of the positive Lyapunov exponents. Thus, the error grows faster for the direct approach. The reason is mentioned above: it is difficult to approximate the dependence of the *far future* on the present state. However, this superiority of

the iterative technique takes place only if the model (10.5) gives very accurate one-step-ahead predictions, which are achieved only for a very long training time series and a very low noise level. Otherwise, the direct technique can give more accurate *l*-step-ahead predictions for *l* greater than 1 but less than a characteristic time of the divergence of initially nearby orbits (Judd and Small, 2000). This is because a one-step-ahead predictor (10.5) can exhibit systematic errors (due to an inappropriate approximating function or an insufficient model dimension), whose accumulation over iterations can be more substantial than the approximation error of the direct approach.

To improve predictions for a moderate advance time *l*, a combined "predictor – corrector" approach is developed in Judd and Small (2000), which is as follows. One generates a forecast via a direct or an iterative way with an existing model. Let us call it a "base predictor". Then, one "corrects" the predictions with an additional model map (the so-called "corrector") which is also constructed from a training time series and describes the dependence of an *l*-step-ahead prediction error of the base predictor on the prediction itself. A functional form of the corrector is taken much simpler than that for the base predictor. The combination "*predictor – corrector*" can give essentially more accurate forecasts in comparison with the "pure" direct and iterative approaches.

Finally, we note an important conceptual distinction between the dynamical models (10.5) and explicit functions of time (Sect. 7.4.1). In contrast to explicit *extrapolation* of a temporal dependence, the model (10.5) relies on *interpolation in the phase space* and, therefore, appears much more efficient. Indeed, a current value of the state vector $\mathbf{x}_i$ used to generate a prediction typically lies "between" many vectors of the training time series, which are used to construct the model (Fig. 10.12). If the model state vector $\mathbf{x}$ under iterations of the model map goes out of the domain populated by the vectors of the training time series, the usage of the model to generate further predictions is tantamount to *extrapolation in the phase space*. Then, the forecasts get much less reliable and a model orbit may behave quite differently from the observed process, e.g. diverge to infinity. The latter often occurs if a model involves an algebraic polynomial, which usually quite badly extrapolates.

10.3.3 Different Kinds of Model Maps

Let us compare predictive abilities of the models (10.5) for different forms of the function *f*.

Algebraic polynomials of a moderate order *K* are quite efficient to approximate gradually varying one-variable functions without discontinuities and breaks. Only cubic or higher order splines are even better in this case (Johnson and Riess, 1982; Kalitkin, 1978; Press et al., 1988; Samarsky, 1982; Stoer and Bulirsch, 1980). The greater the necessary model dimension *D* and the polynomial order *K*, the less the probability of successful modelling results with algebraic polynomials.

Rational functions are efficient under the same conditions but can better describe dependencies with the domains of fast variations. Trigonometric polynomials and

wavelets (Sect. 6.4.2) are also kinds of weak approximation. They perform well for dependencies with specific properties described in Sect. 7.2.4.

Radial basis functions (Judd and Mees, 1995) are much superior to the mentioned approaches in the case of a large model dimension (roughly speaking, the dimension greater than 3). Artificial neural networks exhibit similar performance. According to some indications (Casdagli, 1989; Wan, 1993), the ANNs approximate complicated dependencies even better. Requirements to the amount of data and noise level are not very strict for all the listed models, since all of them are global.

Local linear models are very efficient for moderate model dimensions (less than a moderate number depending on the length of a training time series), long time series (giving a considerable number of the close neighbours for each state vector) and low levels of the measurement noise. Requirements to the amount of data and noise level are quite strict. Local constant models are better than local linear ones for higher noise levels and shorter time series.

In any case, all the approaches suffer from the curse of dimensionality. In practice, very high-dimensional motions (roughly, with the dimensions about 10 or greater) typically are not successfully described with the above empirical models.

10.3.4 Model Maps Versus Model ODEs

In general, model maps give better short-term forecasts than do ODEs (Small et al., 2002). It can be understood in analogy with the situation, where the iterative predictions are less accurate than the direct ones due to significant errors of the one-step-ahead predictor (10.5). Model ODEs are constructed so as to approximate a dependence of the phase velocity $d\mathbf{x}/dt$ on $\mathbf{x}$ (9.3), i.e. to provide better forecasts over short time intervals: $\mathbf{x}(t_{i+1}) \approx \mathbf{x}(t_i) + \left(d\mathbf{x}(t_i)/dt\right)\Delta t$. Then, the integration of ODEs to predict distant future values resembles an iterative forecast. It can be less precise if a systematic error is present in the model ODEs.

Empirical model maps are often superior even for long-term description of the observed dynamics (Small et al., 2002). Besides, their construction and exploitation are simpler: One does not need numerical differentiation of the signals and numerical integration of the equations.

Model ODEs are a good choice if they are "relatives" to an object, i.e. if an original dynamics yields almost exactly to a set of ODEs with some known structure. Such a case is more typical for the "transparent box" or the "grey box" settings and almost improbable without detailed prior information.

Yet, many authors deal with the construction of model ODEs even under the "black box" setting. It is related in part to the problem of the model "transparency". If one gets a "good" model, it is desirable to understand how it "works" and to interpret its variables and parameters from the physical viewpoint. One may hope for such physical interpretations when model ODEs with algebraic polynomials are used, since asymptotic models of many real-world processes take such a form, e.g., in chemical kinetics and laser physics. For the same reason, one can use model ODEs with the successive derivatives (10.4) rather than with the time-delay vectors:

derivatives can be easily interpreted as velocity, acceleration, etc. However, the hope for physical interpretations does not usually prove to be correct: If one does not include physical ideas into the structure of model ODEs in advance (Bezruchko and Smirnov, 2001; Smirnov and Bezruchko, 2006), it gets almost impossible to extract physical sense from an algebraic polynomial (10.14) or a similar universal construction a posteriori.

10.4 Model Validation

Though it is relevant to perform the *residual error analysis* (see Sect. 7.3), if a dynamical noise is assumed to influence an original dynamics, one typically computes model characteristics popular in the theory of dynamical systems and compares them to experimental estimates to validate a dynamical model.

  (i) For a deterministic model, *predictability time* can be theoretically estimated as $\tau_{\text{pred}} = (1/2\Lambda_1)\ln\left(\sigma_x^2/(\sigma_\nu^2 + \sigma_\mu^2 + \sigma_{\Delta M}^2)\right)$. For an adequate model, this estimate must coincide with the corresponding empirical estimate.
 (ii) *Qualitative similarity of the phase orbits projected onto different planes.* This subjective criterion seems important. It is directed to the assessment of the similarity between *essential features* of an object and model dynamics. Its quantitative formulations lead to several ways of model validation, which are mentioned below following the review (Gouesbet et al., 2003).
(iii) *Comparison of invariant measures* (probability distribution densities for a state vector) or their projections (marginal probability distributions). The approach applies to stochastic models as well.
(iv) *Comparison of the largest Lyapunov exponent* of a model attractor with its estimate obtained from an observed time series.
 (v) *Comparison of fractal dimensions and entropies* of a model attractor with their estimates obtained from an observed time series.
(vi) *Comparison of topological properties*. This delicate approach is based on the search and analysis of unstable periodic orbits embedded into an attractor and the determination of their mutual location in the phase space. Strictly speaking, it applies only to deterministic systems, whose dimensionality is not greater than three, and represents a very strict test for a model. If a model reproduces a major part of unstable orbits found from an observed time series, it is a strong evidence in favour of its validity.
(vii) *Comparison of the Poincare maps.* It is easily achieved for one-dimensional Poincare maps. As a rule, one analyses a dependence of the next local maximum value of an observable on its previous local maximum. The approach relates to the analysis of the topological properties of attractors and is often used in combination with the latter.
(viii) *Synchronisation of model dynamics by an original signal.* A model is regarded valid if its dynamics synchronise up to a given accuracy by an observed time

series, used as a driving, under a moderate driving intensity (Brown et al., 1994). This approach applies if model parameters were not estimated via synchronisation-based technique (Sect. 8.2.1) from the same time series.

(ix) It has been suggested to check whether a model has the same number of attractors of the same type as an object; whether the attractors of a model and an object are located in the same domains of the phase space and whether their basins of attraction coincide. These are very strict tests and in practice no empirical model can usually pass them.

Finally, we note that modelling of spatially extended systems with partial differential equations and other tools is actively studied for the last years (Bar et al., 1999; Parlitz and Mayer-Kress, 1995; Parlitz and Merkwirth, 2000; Sitz et al., 2003; Voss et al., 1998) which is not discussed here. Also, we have only briefly touched upon stochastic model equations (Sitz et al., 2002; Timmer, 2000; Tong, 1990). Diverse useful information on those and adjacent subjects can be found, in particular, at the websites mentioned in the Preface.

References

Abarbanel, H.D.I., Brown, R., Kadtke, J.B.: Prediction and system identification in chaotic nonlinear systems: time series with broadband spectra. Phys. Lett. A. **138**, 401–408 (1989)

Anosov, O.L., Butkovsky, O.Ya., Kravtsov, Yu.A.: Limits of predictability for linear autoregression models. J. Communications Technol. Electron. **40**(12), 1866–1873, (in Russian) (1995)

Bär, M., Hegger, R., Kantz, H.: Fitting partial differential equations to space-time dynamics. Phys. Rev. E. **59**, 337–343 (1999)

Bezruchko, B.P., Smirnov, D.A.: Constructing nonautonomous differential equations from a time series. Phys. Rev. E. **63**, 016207, (2001)

Broomhead, D.S., King, G.P.: Extracting qualitative dynamics from experimental data. Physica D. **20**, 217–236 (1986)

Broomhead, D.S., Lowe, D.: Multivariable functional interpolation and adaptive networks. Complex Syst. **2**, 321–355 (1988)

Brown, R., Rulkov, N.F., Tracy, E.R.: Modeling and synchronizing chaotic systems from experimental data. Phys. Lett. A. **194**, 71–76 (1994)

Cao, L., Mees, A.I., Judd, K.: Dynamics from multivariate time series. Phys. D. **121**, 75–88 (1998)

Casdagli, M.: Nonlinear prediction of chaotic time series. Phys. D. **35**, 335–356 (1989)

Casdagli, M., Eubank, S., Farmer, J.D., Gibson, J.: State space reconstruction in presence of noise. Phys. D. **51**, 52–98 (1991)

Cellucci, C.J., Albano, A.M., Rapp, P.E.: Comparative study of embedding methods. Phys. Rev. E. **67**, 066210 (2003)

Eckmann, J.P., Ruelle, D.: Ergodic theory of chaos and strange attractors. Rev. Mod. Phys. **57**, 617–656 (1985)

Farmer, J.D., Sidorowich, J.J.: Predicting chaotic time series. Phys. Rev. Lett. **59**, 845–848 (1987)

Fraser, A.M., Swinney, H.L.: Independent coordinates for strange attractors from mutual information. Phys. Rev. A. **33**, 1131–1140 (1986)

Gerschenfeld, N.A., Weigend, A.S. (eds.): Time Series Prediction: Forecasting the Future and Understanding the Past. SFI Studies in the Science of Complexity, Proc. V. XV. Addison-Wesley, New York (1993)

Gibson, J.F., Farmer, J.D., Casdagli, M., Eubank, S.: An analytic approach to practical state space reconstruction. Phys. D. **57**, 1–30 (1992)

Giona, M., Lentini, F., Cimagalli, V.: Functional reconstruction and local prediction of chaotic time series. Phys. Rev. E. **44**, 3496–3502 (1991)

Gliklikh Yu.E.: What a smooth manifold is. Soros Educ. J. **11**, 155–159, (in Russian) (1998)

Gorban' A.N.: Multivariable functions and neural networks. Soros Educational J. No. 12, 105–112, (in Russian) (1998)

Gouesbet, G.: Reconstruction of the vector fields of continuous dynamical systems from scalar time series. Phys. Rev. A. **43**, 5321–5331 (1991)

Gouesbet, G., Letellier, C.: Global vector-field approximation by using a multivariate polynomial L_2 approximation on nets. Phys. Rev. E. **49**, 4955–4972 (1994)

Gouesbet, G., Meunier-Guttin-Cluzel, S., Ménard, O.: Global reconstructions of equations of motion from data series, and validation techniques, a review. In: Gouesbet, G., Meunier-Guttin-Cluzel, S., Ménard, O. (eds.) Chaos and Its Reconstructions, pp. 1–160. Nova Science Publishers, New York, (2003)

Grassberger, P., Procaccia, I.: Measuring the strangeness of strange attractors. Phys. D. **9**, 189–208 (1983)

Gribkov, D.A., Gribkova, V.V., Kravtsov Yu.A., et al. Reconstructing structure of dynamical system from time series. J. Commun. Technol. Electron. **39**(2), 269–277, (in Russian) (1994)

Hubner, U., Weiss, C.-O., Abraham, N.B., Tang, D.: Lorenz-like chaos in $NH_3 - FIR$ lasers (data set A). In: Gerschenfeld, N.A., Weigend, A.S. (eds.) Time Series Prediction: Forecasting the Future and Understanding the Past. SFI Studies in the Science of Complexity, Proc. V. XV, pp. 73–104. Addison-Wesley, Reading, MA (1993)

Janson, N.B., Pavlov, A.N., Anishchenko, V.S.: One method for restoring inhomogeneous attractors. Int. J. Bif. Chaos. **8**, 825–833 (1998)

Johnson, L.W., Riess, R.D.: Numerical Analysis, 2nd edn. Addison-Wesley, Reading, MA (1982)

Judd, K., Mees, A.I. On selecting models for nonlinear time series. Phys. D. **82**, 426–444 (1995)

Judd, K., Mees, A.I.: Embedding as a modeling problem. Phys. D. **120**, 273–286 (1998)

Judd, K., Small, M.: Towards long-term prediction. Phys. D. **136**, 31–44 (2000)

Kalitkin, N.N.: Numerical Methods. Nauka, Moscow, (in Russian) (1978)

Kantz, H., Schreiber, T.: Nonlinear Time Series Analysis. Cambridge University Press, Cambridge (1997)

Kartsev, V.P.: Adventures of Great Equations. Znaniye, Moscow, (in Russian) (1976)

Kennel, M.B., Brown, R., Abarbanel, H.D.I.: Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys. Rev. A. **45**, 3403–3411 (1992)

Kravtsov Yu.A.: Randomness, determinacy, predictability. Phys. Uspekhi. **158**(1), 93–115, (in Russian) (1989)

Kugiumtzis, D., Lingjaerde, O.C., Christophersen, N.: Regularized local linear prediction of chaotic time series. Phys. D. **112**, 344–360 (1998)

Landa, P.S., Rosenblum, M.G.: Comparison of methods for construction of phase space and determination of attractor dimension from experimental data. Tech. Phys. **59**(11), 1–8, (in Russian) (1989)

Letellier, C., Aguirre, L.A.: Investigating nonlinear dynamics from time series: The influence of symmetries and the choice of observables. Chaos. **12**, 549–558 (2002)

Letellier, C., Le Sceller, L., Gouesbet, G., et al.: Recovering deterministic behavior from experimental time series in mixing reactor. AIChE J. **43**(9), 2194–2202 (1997)

Letellier, C., Le Sceller, L., Maréchal, E., et al.: Global vector field reconstruction from a chaotic experimental signal in copper electrodissolution. Phys. Rev. E. **51**, 4262–4266 (1995)

Letellier, C., Maquet, J., Labro, H., et al. Analyzing chaotic behavior in a Belousov-Zhabotinskyi reaction by using a global vector field reconstruction. J. Phys. Chem. **102**, 10265–10273 (1998a)

Letellier, C., Macquet, J., Le Sceller, L., et al. On the non-equivalence of observables in phase space reconstructions from recorded time series. J. Phys. A: Math. Gen. **31**, 7913–7927 (1998b)

Liebert, W., Schuster, H.G.: Proper choice the of time delay for the analysis of chaotic time series. Phys. Lett. A. **142**, 107–111 (1989)

Makarenko, N.G.: Fractals, attractors, neural networks and so forth. Procs. IV All-Russian Conf. "Neuroinformatics-2002". Part 2, pp. 121–169. Moscow, (in Russian) (2002)

Makarenko, N.G.: Embedology and Neuro-Prediction. Procs. V All-Russian Conf. "Neuroinformatics-2003". Part 1, pp. 86–148. Moscow, (in Russian) (2003)

Malinetsky, G.G., Potapov, A.B.: Contemporary Problems of Nonlinear Dynamics. Editorial URSS, Moscow, (in Russian) (2000)

Parlitz, U., Mayer-Kress, G.: Predicting low-dimensional spatiotemporal dynamics using discrete wavelet transforms. Phys. Rev. E. **51**, R2709–R271 (1995)

Parlitz, U., Merkwirth, C.: Prediction of spatiotemporal time series based on reconstructed local states. Phys. Rev. Lett. **84**, 1890–1893 (2000)

Pavlov, A.N., Janson, N.B., Anishchenko, V.S.: Application of statistical methods to solve the problem of global reconstruction. Tech. Phys. Lett. **23**(4), 297–299 (1997)

Pavlov, A.N., Janson, N.B., Anishchenko, V.S.: Reconstruction of dynamical systems. J. Commun. Technol. Electron. **44**(9), 999–1014 (1999)

Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T.: Numerical Recipes in C. Cambridge University Press, Cambridge (1988)

Rulkov, N.F., Sushchik, M.M., Tsimring, L.S., Abarbanel, H.D.I.: Generalized synchronization of chaos in directionally coupled chaotic systems. Phys. Rev. E. **51**, 980–994 (1995)

Samarsky, A.A.: Introduction to Numerical Methods. Nauka, Moscow, (in Russian) (1982)

Sauer, T., Yorke, J.A., Casdagli, M.: Embedology. J. Stat. Phys. **65**(3–4), 579–616 (1991)

Sauer, T.: Time series prediction by using delay coordinate embedding. in Time Series Prediction: Forecasting the Future and Understanding the Past. Eds. N.A. Gerschenfeld, A.S. Weigend. SFI Studies in the Science of Complexity, Proc. V. XV, pp. 175–193. Addison-Wesley (1993)

Schroer, C., Sauer, T., Ott, E., Yorke, J.: Predicting chaos most of the time from embeddings with self-intersections. Phys. Rev. Lett. **80**, 1410–1413 (1998)

Sitz, A., Kurths, J., Voss, H.U.: Identification of nonlinear spatiotemporal systems via partitioned filtering. Phys. Rev. E. **68**, 016202 (2003)

Sitz, A., Schwartz, U., Kurths, J., Voss, H.U.: Estimation of parameters and unobserved components for nonlinear systems from noisy time series. Phys. Rev. E. **66**, 016210 (2002)

Small, M., Judd, K., Mees, A.: Modelling continuous processes from data. Phys. Rev. E. **65**, 046704 (2002)

Small, M., Judd, K.: Comparisons of new nonlinear modeling techniques with applications to infant respiration. Physica, D. **117**, 283–298 (1998)

Small, M., Tse, C.K.: Optimal embedding: A modelling paradigm. Phys. D. **194**, 283–296 (2004)

Smirnov, D.A., Bezruchko, B.P., Seleznev Ye.P.: Choice of dynamical variables for global reconstruction of model equations from time series. Phys. Rev. E. **65**, 026205 (2002)

Smirnov, D.A., Bezruchko, B.P.: Nonlinear dynamical models from chaotic time series: methods and applications. In: Winterhalder, M., Schelter, B., Timmer, J. (eds.) Handbook of Time Series Analysis, pp. 181–212. Wiley-VCH, Berlin (2006)

Smith, L.A.: Identification and prediction of low-dimensional dynamics. Phys. D. **58**, 50–76 (1992)

Smith, L.A.: Maintenance of uncertainty. Proc. Int. School of Physics "Enrico Fermi", Course CXXXIII, pp. 177–246. Italian Physical Society, Bologna, (1997). Available at http://www.maths.ox.ac.uk/~lenny

Stark, J., Broomhead, D.S., Davies, M., Huke, J.: Takens embedding theorem for forced and stochastic systems. Nonlinear Analysis. Theory, Methods, and Applications. Proc. 2nd Congress on Nonlinear Analysis. Elsevier Science Ltd., **30**(8), 5303–5314 (1997)

Stoer, J., Bulirsch, R.: Introduction to Numerical Analysis. Springer, New York (1980)

Takens, F.: Detecting strange attractors in turbulence. Lect. Notes Math. **898**, 366–381 (1981)

Timmer, J.: Parameter estimation in nonlinear stochastic differential equations. Chaos, Solitons Fractals **11**, 2571–2578 (2000)

Tong, H.: Nonlinear Time Series Analysis: a Dynamical System Approach. Oxford University Press, Oxford (1990)

Voss, H.U., Bünner, M., Abel, M.: Identification of continuous, spatiotemporal systems. Phys. Rev. E. **57**, 2820–2823 (1998)

Wan, E.A.: Time series prediction by using a connectionist network with internal delay lines. In: Gerschenfeld, N.A., Weigend, A.S. (eds.) Time Series Prediction: Forecasting the Future and Understanding the Past. SFI Studies in the Science of Complexity, Proc. V. XV, pp. 195–217. Addison-Wesley (1993)

Yule, G.U.: On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. Phil. Trans. R. Soc. Lond. A. **226**, 267–298 (1927)