

Chapter 8

Model Equations: Parameter Estimation

Motions and processes observed in nature are extremely diverse and complex. Therefore, opportunities to model them with explicit functions of time are rather restricted. Much greater potential is expected from difference and differential equations (Sects. 3.3, 3.5 and 3.6). Even a simple one-dimensional map with a quadratic maximum is capable of demonstrating chaotic behaviour (Sect. 3.6.2). Such model equations *in contrast to explicit functions of time* describe how a *future state of an object depends on its current state* or how *velocity of the state change depends on the state itself*. However, a technology for the construction of these more sophisticated models, including parameter estimation and selection of approximating functions, is basically the same. A simple example: construction of a one-dimensional map $\eta_{n+1} = f(\eta_n, \mathbf{c})$ differs from obtaining an explicit temporal dependence $\eta = f(t, \mathbf{c})$ only in that one needs to draw a curve through experimental data points on the plane (η_n, η_{n+1}) (Fig. 8.1a–c) rather than on the plane (t, η) (Fig. 7.1). To construct model ODEs $dx/dt = \mathbf{f}(\mathbf{x}, \mathbf{c})$, one may first get time series of the derivatives dx_k/dt ($k = 1, \dots, D$, where D is a model dimension) via numerical differentiation and then approximate a dependence of dx_k/dt on \mathbf{x} in a usual way. Model equations can be multidimensional, which is another difference from the construction of models as explicit functions of time.

For a long time, in empirical modelling of complex processes, one used *linear difference equations* containing noise to allow for irregularity (Sect. 4.4). The idea was first suggested in 1927 (Yule, 1927) and appeared very fruitful so that autoregression and moving average models became a main tool for the description of complex behaviour for the next 50 years.

Only in 1960–1970s, researchers widely realised that simple low-dimensional models in the form of non-linear maps or differential equations can exhibit complex oscillations even without noise influence. It gave a new impulse to the development of empirical modelling techniques, since arousal of powerful and widely accessible computers provided practical implementation of the ideas.

In this chapter, we consider a situation when an observed time series $\eta_i = h(\mathbf{x}(t_i)), i = 1, \dots, N$, is produced by iterations of a map $\mathbf{x}_{n+1} = \mathbf{f}(\mathbf{x}_n, \mathbf{c})$ or integration of an ordinary differential equation $d\mathbf{x}/dt = \mathbf{f}(\mathbf{x}, \mathbf{c})$, whose structure is *completely known*. The problem is *to estimate parameters* \mathbf{c} from the observed

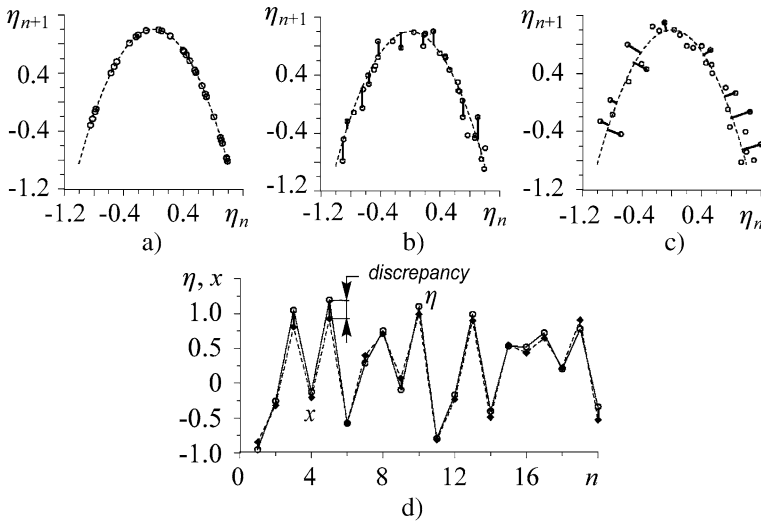


Fig. 8.1 Parameter estimation for the quadratic map (8.1) at $c_0 = 1.85$. Circles show the observed values: (a) no noise; the dashed line is an original parabola; (b) only dynamical noise is present; the dashed line is a model parabola obtained via minimisation of the mean-squared vertical distance (some of those distances are shown by solid lines); (c) only measurement noise is present; the dashed line is a model parabola obtained via minimisation of the mean-squared orthogonal distance; (d) rhombs show a model realisation which is the closest one to an original time series in the sense (8.4)

data. This is called “transparent box” problem (Sect. 5.2). To make the consideration more realistic, we add dynamical and/or measurement noise.

Such a problem setting is encountered in different applications and attracts serious attention. One singles out two main aspects of interest considered below:

- (i) Parameter estimation *with a required accuracy* is important if the parameters cannot be measured directly due to experimental conditions. Then, the modelling procedure plays a role of “measurement device” (Butkovsky et al., 2002; Horbelt and Timmer, 2003; Jaeger and Kanrz, 1996; Judd, 2003; McSharry and Smith, 1999; Pisarenko and Sornette, 2004; Smirnov et al., 2005b) (see Sect. 8.1).
- (ii) Parameter estimation *in the case of data deficit* is even more problematic. Such is a situation when one cannot get time series of all model dynamical variables x_k from the measured values of an observable η , i.e. some variables are “hidden” (Baake et al., 1992; Bezruchko et al., 2006; Breeden and Hubler, 1990; Parlitz, 1996; Voss et al., 2004; (see Sect. 8.2).

8.1 Parameter Estimators and Their Accuracy

Let us consider the estimation of a single parameter in a non-linear map from its noise-corrupted time realisation. The object is the quadratic map in a chaotic regime with unknown parameter $c = c_0$:

$$x_{n+1} = f(x_n, c_0) + \xi_n = 1 - c_0 x_n^2 + \xi_n, \eta_n = x_n + \zeta_n, \quad (8.1)$$

where ξ_n, ζ_n are independent random processes. The first of them is a *dynamical noise* (since it influences the dynamics) and the second one is a *measurement noise* (since it affects only the observations).

If both noises are absent, one has $\eta_n = x_n$ and experimental data points on the plane x_n, x_{n+1} lie exactly on the sought parabola (Fig. 8.1a). Finding the value of c reduces to an algebraic equation whose solution is $\hat{c} = (1 - x_{n+1}) / x_n^2$. Hence, it is sufficient to use any two observations x_n, x_{n+1} with $x_n \neq 0$. As a result, a model coincides with the object up to the computation error.

If noise is present either in dynamics or measurements, one looks for a *parameter estimate* rather than for the precise value of the parameter. The most widely used estimation techniques are described in Sect. 7.1.1. Peculiarities of their application under the considered problem setting are as follows.

8.1.1 Dynamical Noise

Let the dynamical noise ξ_n in Eq. (8.1) be a sequence of statistically independent random quantities and identically distributed with a probability density $p(\xi)$. To estimate the parameter c , one can use the ML technique (Sects. 2.2.1, 7.1.1 and 7.1.2), which is the most efficient one under sufficiently general conditions (Ibragimov and Has'minskii, 1979; Pisarenko and Sornette, 2004). The likelihood function [see also Eqs. (2.26) and (7.10)] reads in this case as

$$\ln L(c) \equiv \ln p_N(\eta_1, \eta_2, \dots, \eta_N | c) \approx \sum_{n=1}^{N-1} \ln p(\eta_{n+1} - f(\eta_n, c)). \quad (8.2)$$

To apply the technique, one must know the distribution law $p(\xi)$, which is rarely the case. Most often, one assumes Gaussian noise so that the maximisation of Eq. (8.2) becomes equivalent to the so-called *ordinary least-squares technique*, i.e. to the minimisation

$$S(c) = \sum_{i=1}^{N-1} (\eta_{i+1} - f(\eta_i, c))^2 \rightarrow \min. \quad (8.3)$$

It means that the plot of a model function on the plane (η_n, η_{n+1}) should go in such a way so as to minimise the sum of the squared *vertical distances* from it to the experimental data points (Fig. 8.1b).

As a rule, the error in the estimator \hat{c} decreases with a time series length N . Under the considered problem setting, both the ML approach and the LS technique give asymptotically unbiased and consistent estimators. It can be shown that the estimator variance decreases as N^{-1} analogous to the examples in Sect. 7.1.2. The reason can be described in the same way: the terms in Eq. (8.3) are stationary with respect to i , i.e. a partial Fisher information is bounded.

The ordinary LS technique often gives an acceptable estimator accuracy even if the noise is not Gaussian (Sects. 7.1.1 and 7.1.2). Although one may apply other methods, e.g., the least absolute values technique, the LS technique is much easily implemented. An essential technical difficulty arises when the “relief” of the cost function (8.3) exhibits many local minima, which is often the case if f is non-linear with respect to \mathbf{c} . Then, the optimisation problem is solved in an iterative way with some starting guesses for the sought parameters (Dennis and Schnabel, 1983). Whether a global extremum is found depends on how “lucky” are the starting guesses i.e. how close they are to the true parameter values. In the example (8.1), f is linear with respect to c ; therefore, the cost function S is quadratic with respect to c and exhibits the only minimum, which is easily found as a solution to a linear algebraic equation.

We note that if f is linear with respect to x , the model (8.1) is a linear first-order *autoregression model*. More general ARMA models involve a dependence of x_{n+1} on several previous values of x and ξ , see Eq. (4.13) in Sect. 4.4.

8.1.2 Measurement Noise

If only a measurement noise is present ($\eta_n = x_n + \zeta_n$), the estimation problem gets more complicated. This is because one aims at finding a dependence of x_{n+1} on x_n , where x_n is an “independent” variable whose observed values are noise corrupted [a confluent analysis problem, see Eq. (2.28) in Sect. 2.2.1.8].

8.1.2.1 Bias in the Estimator Obtained Via the Ordinary LS Technique

The bias is non-zero for an arbitrarily long time series, since the technique (8.3) is developed under the assumption of only a dynamical noise presence. It can be illustrated with the example (8.1), where one has

$$\begin{aligned} S(c) &= \sum_{i=1}^{N-1} (\eta_{i+1} - f(\eta_i, c))^2 = \sum_{i=1}^{N-1} (x_{i+1} + \zeta_{i+1} - 1 + c(x_i + \zeta_i))^2 = \\ &= \sum_{i=1}^{N-1} (cx_i^2 - c_0x_i^2 + \zeta_{i+1} + 2cx_i\zeta_i + c\zeta_i^2)^2. \end{aligned}$$

Minimum of S over c can be found from the condition $\partial S/\partial c = 0$, which reduces to the form

$$\sum_{i=1}^{N-1} \left(cx_i^2 - c_0x_i^2 + \zeta_{i+1} + 2cx_i\zeta_{i+1} + c\zeta_i^2 \right) \cdot \left(x_i^2 + 2x_i\zeta_{i+1} + \zeta_i^2 \right) = 0.$$

By solving this equation, one gets an estimator (McSharry and Smith, 1999)

$$\hat{c} = \frac{c_0 \left(\sum_{i=1}^{N-1} x_i^4 + 2 \sum_{i=1}^{N-1} x_i^3 \zeta_i + \sum_{i=1}^{N-1} x_i^2 \zeta_i^2 \right) - \sum_{i=1}^{N-1} x_i^2 \zeta_{i+1} - 2 \sum_{i=1}^{N-1} x_i \zeta_i \zeta_{i+1} - \sum_{i=1}^{N-1} \zeta_i^2 \zeta_{i+1}}{\sum_{i=1}^{N-1} x_i^4 + 6 \sum_{i=1}^{N-1} x_i^2 \zeta_i^2 + \sum_{i=1}^{N-1} x_i^4 + 4 \sum_{i=1}^{N-1} x_i^3 \zeta_i + 4 \sum_{i=1}^{N-1} x_i \zeta_i^3}.$$

Under the condition $N \rightarrow \infty$, one can take into account statistical independence ζ_i of ζ_{i+1} and x_i and replace the sums like $\sum_{i=1}^{N-1} x_i^4$ (temporal averaging) by the integrals like $\int_{-\infty}^{\infty} \mu(x, c_0)x^4 dx \equiv \langle x^4 \rangle$ (ensemble averaging). Here, $\mu(x, c_0)$ is an invariant measure for the map (8.1), i.e. a probability density function. At $c_0 = 2$ it can be found analytically: $\mu(x, 2) = 1/\pi (1 - x^2)$, $-1 < x < 1$. Hence, one gets $\langle x^2 \rangle = 1/2$, $\langle x^4 \rangle = 3/8$ and $\langle x^n \rangle = 0$ for uneven n . Finally, at $c_0 = 2$, one comes to the asymptotic expression $\hat{c} = c_0 (4\sigma_\zeta^2 + 3) / (8\langle \zeta^4 \rangle + 24\sigma_\zeta^2 + 3)$. This is a biased estimator. It underestimates the true value c_0 , since the denominator is greater than the numerator in the above expression. Figure 8.2 shows the asymptotic value of \hat{c} versus noise-to-signal ratio for Gaussian noise. It is close to the true value only under the low noise levels; its bias is less than 1% if the noise level is less

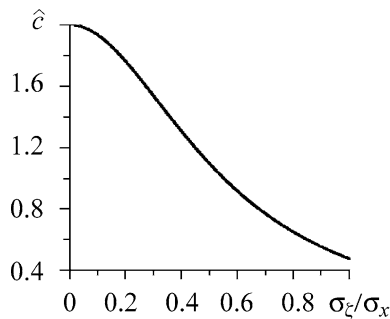


Fig. 8.2 Ordinary LS estimates of the parameter c in Eq. (8.1) versus the noise level at $N = \infty$ and $c_0 = 2$. Noise-to-signal ratio quantified as the ratio of standard deviations is shown versus the abscissa axis (McSharry and Smith, 1999)

than 0.05. The bias rises with the noise level. Analogous properties are observed for other noise distributions (Butkovsky et al., 2002; McSharry and Smith, 1999).

However, since the LS technique is simple for the implementation and can be easily used in the case of many estimated parameters in contrast to other methods, it is often applied in practice with an additional assumption of low measurement noise level.

8.1.2.2 Increasing Estimator Accuracy at High Measurement Noise Level

It is possible in part through the use of the *total least-squares technique* (Jaeger and Kanrz, 1996) when one minimises the sum of the squared orthogonal distances from the data points (η_n, η_{n+1}) to the plot of the function $f(x_n, c)$ (Fig. 8.1c). Thereby, one takes into account that deviations of the observed data points with coordinates (η_n, η_{n+1}) from the plot of the sought function $f(x_n, c_0)$ are induced by the noise influence on both coordinates. Therefore, the deviations may occur in any direction, not only vertically. The use of the orthogonal distances is justified in Jaeger and Kanrz (1996) as an approximate version of the ML approach.

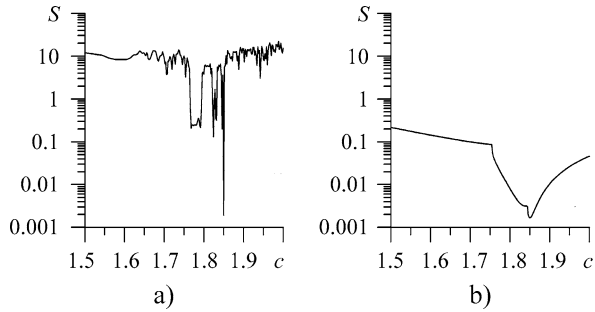
However, a non-zero estimator bias is not fully eliminated under the use of the total LS technique (especially in the case of a very strong noise), since the latter is just an approximation to the ML technique. It may seem that a way out is to write down the likelihood function for the new situation “honestly”, i.e. taking into account how the noise enters the observations. For a Gaussian noise, the problem reduces to the minimisation of the sum of squared deviations of a model *realisation* from an *observed time series* (Fig. 8.1d):

$$S(c, x_1) = \sum_{n=0}^{N-1} \left(\eta_{n+1} - f^{(n)}(x_1, c) \right)^2 \rightarrow \min, \quad (8.4)$$

where $f^{(n)}$ is an n th iterate of the map $x_{n+1} = f(x_n, c)$, $f^{(0)}(x, c) = x$, and the initial model state x_1 is considered as an estimated quantity as well.

An orbit of a chaotic system is very sensitive to initial conditions and parameters. Therefore, the variance of the estimator obtained from Eq. (8.4) for a chaotic orbit rapidly decreases with N , sometimes even exponentially (Horbelt and Timmer, 2003; Pisarenko and Sornette, 2004). This is a desirable property, but it is achieved only if one manages to find the global minimum of Eq. (8.4). In practice, even for moderate values of N , the “relief” of S for a chaotic system becomes strongly “jagged” (Fig. 8.3a) so that it gets almost impossible to find the global minimum numerically (Dennis and Schnabel, 1983). To do it, one would need very “lucky” starting guesses for c and x_1 . It is also difficult to speak of asymptotic properties of the estimators since the cost function gets non-smooth in the limit $N \rightarrow \infty$. Therefore, one develops modifications of the ML technique in application to the parameter estimation from a chaotic time series (Pisarenko and Sornette, 2004; Smirnov et al., 2005b).

Fig. 8.3 Cost functions for the quadratic map (8.1) at $c_0 = 1.85$ and $N = 20$. The left panel shows the cost function for the direct iterates (8.4), where $x_1 = 0.3$; the right one for the reverse iterates (8.5), where $x_N = f^{(N-1)}(0.3, c_0)$



Thus, according to a piecewise technique, one divides an original time series into segments of a moderate length so that it is possible to find the global minimum of Eq. (8.4) for each of them and averages the obtained estimates. This is a reasonable approach, but a resulting estimator may remain asymptotically biased. Its variance decreases again as N^{-1} . Some ways to improve the estimator properties are described in Sect. 8.2.

Here, we note only an approach specific to one-dimensional maps (Smirnov et al., 2005b). It is based on the property that the only Lyapunov exponent of a one-dimensional chaotic map becomes negative under the time reversal so that an orbit gets much less sensitive to parameters and a “final condition”. Therefore, one minimises a quantity

$$S(c, x_N) = \sum_{n=0}^{N-1} \left(\eta_{N-n} - f^{(-n)}(x_N, c) \right)^2 \rightarrow \min, \tag{8.5}$$

where $f^{(-n)}$ is an n th iterate of the map $x_{n+1} = f(x_n, c)$ in reverse time, in particular $f^{(-1)}$ is an inverse function for f with respect to x . The plot of the cost function (8.5) looks sufficiently smooth for an arbitrarily long time series (Fig. 8.3b) so that it is not difficult to find its global minimum. At low and moderate noise levels (σ_ζ/σ_x up to 0.05–0.15), the error in the estimator (8.5) appears less than for the piecewise technique. Moreover, the expression (8.5) gives asymptotically unbiased estimates, whose variance typically scales as N^{-2} at weak noise. The high rate of error decrease is determined by the returns of a chaotic orbit to a small vicinity of the extremum of the function f (Smirnov et al., 2005b).

8.2 Hidden Variables

If the measurement noise level is considerable, the state variable x is often regarded “hidden” since its true values are, in fact, unknown. Variables are “even more hidden” if even their noise-corrupted values cannot be measured directly or computed from the observed data. This is often the case in practice. In such a situation, the

parameter estimation is much more problematic than in the cases considered in Sect. 8.1. However, if one manages to solve the problem, then an additional opportunity to restore time series of hidden variables appears as a by-product. Then, a modelling procedure serves as a measurement device with respect to dynamical variables as well.

8.2.1 Measurement Noise

We illustrate the techniques with the example of the parameter estimation in ordinary differential equations without a dynamical noise. An object is the classical chaotic system – Lorenz system:

$$\begin{aligned} dx_1/dt &= c_1(x_2 - x_1), \\ dx_2/dt &= -x_2 + x_1(c_3 - x_3), \\ dx_3/dt &= c_2x_3 + x_1x_2, \end{aligned} \tag{8.6}$$

for a “canonical” set of parameter values $c_1 = 10$, $c_2 = 8/3$, $c_3 = 46$. A noise-corrupted realisation of x_1 is considered as observed data, i.e. $\eta_n = x_1(t_n) + \zeta_n$, while the variables x_2 and x_3 are hidden. A model is constructed in the form (8.6) where all the three parameters c_k are regarded unknown.

8.2.1.1 Initial Value Approach

All the estimation techniques are based to a certain extent on the ideas like Eq. (8.4), i.e. one chooses such initial conditions and parameter values to provide maximal closeness of a model realisation to an observed time series in the least-squares sense. The direct solution to the problem like (8.4) is called an initial value approach (Horbelt, 2001; Voss et al., 2004). As indicated in Sect. 8.1.2, it is not applicable to a long chaotic time series. Improving the approach is not straightforward. Thus, a simple division of a time series into segments with subsequent averaging of the respective estimates gives a low accuracy of the resulting estimator. The reverse-time iterations are not suitable in the case of a multidimensional dissipative system.

8.2.1.2 Multiple Shooting Approach

The difficulties can be overcome in part with Bock’s algorithm (Baake et al., 1992; Bock, 1981). It is also called a multiple shooting approach since one replaces the Cauchy initial value problem for an entire observation interval with a set of boundary value problems. Namely, one divides an original time series $\{\eta_1, \eta_2, \dots, \eta_N\}$ into L non-overlapping segments of length M and considers model initial states $\mathbf{x}^{(i)}$ for each of them (i.e. at time instants $t_{(i-1)M+1}$, $i = 1, \dots, L$) as estimated quantities, but *not as free parameters*. One solves the problem of conditional minimisation,

which reads in the case of a scalar observable $\eta = h(\mathbf{x}) + \zeta$ ($\eta = x_1 + \zeta$ in our example) as

$$\begin{aligned}
 S(\mathbf{c}, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(L)}) &= \\
 &= \sum_{i=1}^L \sum_{n=1}^M \left[\eta(t_{(i-1)M+n}) - h(\mathbf{x}(t_{(i-1)M+n} - t_{(i-1)M+1}, \mathbf{x}^{(i)}, \mathbf{c})) \right]^2 \rightarrow \min, \\
 \mathbf{x}(t_{iM+1} - t_{(i-1)M+1}, \mathbf{x}^{(i)}, \mathbf{c}) &= \mathbf{x}^{(i+1)}, \quad i = 1, 2, \dots, L - 1.
 \end{aligned}
 \tag{8.7}$$

The quantity $\mathbf{x}(t, \mathbf{x}^{(i)}, \mathbf{c})$ denotes a model realisation (a solution to model equations), i.e. a model state \mathbf{x} at a time instant t for an initial state $\mathbf{x}^{(i)}$ and a parameter value \mathbf{c} . The first equation in Eq. (8.7) means minimisation of the deviations of a model realisation from the observed series over the entire observation interval. The second line provides “matching” of the segments to get finally a *continuous* model orbit over the entire observation interval. This matching imposes the conditions of the equality type on the L sought vectors $\mathbf{x}^{(i)}$, i.e. only one of the vectors can be regarded as a free parameter of the problem.

Next, one solves the problem with ordinary numerical iterative techniques using some *starting guesses* for the sought quantities $\mathbf{c}, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(L)}$. The starting guesses correspond, as a rule, to L non-matching pieces of a model orbit (Fig. 8.4b,

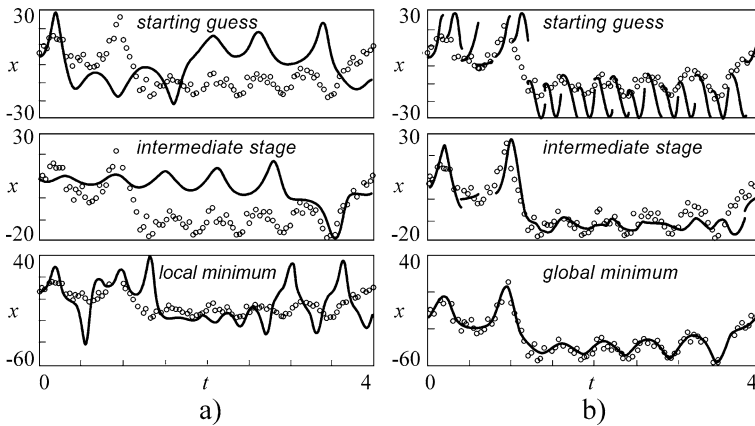


Fig. 8.4 Parameter estimation from a chaotic realisation of the coordinate $x = x_1$ of the Lorenz system (8.6), $N = 100$ data points, the sampling interval is 0.04, measurement noise is Gaussian and white with standard deviation of 0.2 of the standard deviation of the noise-free signal (Horbelt, 2001). (a) An initial value approach. Realisations of an observable (circles) and a corresponding model variable. The fitting process converges to a local minimum, where a model orbit and parameter estimates strongly differ from the true ones. (b) A multiple shooting approach. The fitting process converges to the global minimum, where a model orbit and parameter estimates are close to the true ones

upper panel). The situation is similar for intermediate values of the sought quantities during the iterative minimisation procedure, but the “discrepancies” should get smaller and smaller if the procedure converges at given starting guesses. Such a temporary admission of the model orbit discontinuity distinguishes Bock’s algorithm (Fig. 8.4b) from the initial value approach (Fig. 8.4a) and provides greater flexibility of the former.

An example of the application of the two techniques to the parameter estimation for the system (8.6) is shown in Fig. 8.4. The multiple shooting technique “finds” the *global minimum*, while the initial value approach stops at a local one. This is quite a typical situation. However, the multiple shooting technique does not assure finding the global minimum. It only softens requirements to the “goodness” of starting guesses for the sought quantities (Bezruchko et al., 2006). For even longer chaotic time series, it also gets inefficient since the basic principle of closeness of a chaotic model orbit to the observed time series over a long time interval again leads to very strict requirements to starting guesses.

8.2.1.3 Modification of the Multiple Shooting Technique

As shown in Bezruchko et al. (2006), one can avoid some difficulties via allowing discontinuity of the resulting model orbit at several time instants within the observation interval, i.e. via ignoring several equalities in the last line of Eq. (8.7). In such a way it is easier to find the global minimum of the cost function S in Eq. (8.7).

Such a modification allows to use arbitrarily long chaotic time series, but the requital is that sometimes a model with an inadequate structure may be accepted as a “good” one due to its ability to reproduce short segments of the time series. Therefore, one should carefully select the number and the size of the continuity segments for the model orbit.

There is also an additional difficulty in the situation with hidden variables. Apart from lucky starting guesses for the parameters \mathbf{c} , it appears important to generate lucky starting guesses for the hidden variables (components of vectors $\mathbf{x}^{(i)}$) in contrast to very optimistic early statements (Baake et al., 1992). Quite often, one has to proceed intuitively or via blind trials and errors. However, useful information may be obtained sometimes through a preliminary study of the properties of model time realisations at several trial parameter values (Bezruchko et al., 2006).

8.2.1.4 Synchronisation-Based Parameter Estimation

A further improvement of the technique is based on the idea of synchronising a model by the observed time series. It was suggested in Parlitz (1996) and further elaborated in many works. Here, we briefly describe its main points, advantages and difficulties following the works Parlitz (1996) and Parlitz et al. (1996).

Let us consider a system of ODEs $\mathbf{dy}/dt = \mathbf{f}(\mathbf{y}, \mathbf{c}_0)$ with a state vector $\mathbf{y} \in R^D$ and a parameter vector $\mathbf{c}_0 \in R^P$ as an object of modelling. Parameter values \mathbf{c}_0 are unknown. An observable vector is $\boldsymbol{\eta} = \mathbf{h}(\mathbf{y})$. It may have a lower dimension than the state vector \mathbf{y} (the case of hidden variables). A model is given by the equation

$d\mathbf{x}/dt = \mathbf{f}(\mathbf{x}, \mathbf{c})$. Let us assume that there is a unidirectional coupling scheme using the available signal $\eta(t)$, which enables asymptotically stable synchronisation of the model (\mathbf{x}) by the object (\mathbf{y}), i.e. $\mathbf{x} \rightarrow \mathbf{y}$ as $t \rightarrow \infty$ if $\mathbf{c} = \mathbf{c}_0$. Thus, one can vary model parameters \mathbf{c} and integrate model equations with somehow introduced input signal $\eta(t)$ at each value of \mathbf{c} . If at some value $\mathbf{c} = \hat{\mathbf{c}}$ identical synchronisation between $\eta(t)$ and the corresponding model realisation $\mathbf{h}(\mathbf{x}(t))$ (i.e. the regime $\eta(t) = \mathbf{h}(\mathbf{x}(t))$) is achieved after some transient process, then the value $\hat{\mathbf{c}}$ should be equal to \mathbf{c}_0 . If only an approximate relationship $\eta(t) \approx \mathbf{h}(\mathbf{x}(t))$ holds true, then $\hat{\mathbf{c}}$ can be taken as an estimate of \mathbf{c}_0 .

There are different implementations of the idea. One can introduce a measure of the discrepancy between $\eta(t)$ and $\mathbf{h}(\mathbf{x}(t))$, e.g. their mean-squared difference after some transient, and minimise it as a function of \mathbf{c} (Parlitz et al., 1996). This is a complicated problem of non-linear optimisation similar to those encountered under the multiple shooting approach. An advantage of the synchronisation-based estimation is that the minimised function often changes quite gradually with \mathbf{c} and has a pronounced minimum at $\mathbf{c} = \mathbf{c}_0$ with a broad “basin of attraction”, i.e. a starting guess for \mathbf{c} does not have to be so “lucky” as under the multiple shooting approach. This is explained by the following property of many non-linear systems: If \mathbf{c} is not equal to \mathbf{c}_0 but reasonably close to it, the identical synchronisation is impossible but there often occurs the generalised synchronisation (Sect. 6.4.5), where \mathbf{x} is a function of \mathbf{y} not very much different from $\mathbf{x} = \mathbf{y}$. Then, the discrepancy between $\eta(t)$ and $\mathbf{h}(\mathbf{x}(t))$ changes smoothly in the vicinity of $\mathbf{c} = \mathbf{c}_0$.

It can be even more convenient to avoid minimisation of a complicated cost function by considering the parameters \mathbf{c} as additional variables in model ODEs and update their values depending on the current mismatch between $\eta(t)$ and $\mathbf{h}(\mathbf{x}(t))$ in the course of integration of the model ODEs (Parlitz, 1996). An example is again the chaotic Lorenz system

$$\begin{aligned} dy_1/dt &= \sigma(y_2 - y_1), \\ dy_2/dt &= c_{1,0}y_1 - c_{2,0}y_2 - y_1y_3 + c_{3,0}, \\ dy_3/dt &= y_1y_2 - by_3, \end{aligned} \tag{8.8}$$

with the parameters $c_{1,0} = 28$, $c_{2,0} = 1$, $c_{3,0} = 0$, $\sigma = 10$, $b = 8/3$ and an observable $\eta = h(\mathbf{y}) = y_2$. The following unidirectional coupling scheme and equations for the parameter updates were considered:

$$\begin{aligned} dx_1/dt &= \sigma(\eta - x_1), \\ dx_2/dt &= c_1x_1 - c_2x_2 - x_1x_3 + c_3, \\ dx_3/dt &= x_1x_2 - bx_3, \\ dc_1/dt &= (\eta - x_2)x_1, \\ dc_2/dt &= -(\eta - x_2)x_2, \\ dc_3/dt &= \eta - x_2. \end{aligned} \tag{8.9}$$

Using a global Lyapunov function, the author has shown that at correct parameter values $\mathbf{c} = \mathbf{c}_0$, the model is synchronised by the signal $\eta(t)$ at all initial conditions. The system (8.9) tends to the regime $\mathbf{x}(t) = \mathbf{y}(t)$ and $\mathbf{c}(t) = \mathbf{c}_0$ at any values of $x_1(0), x_2(0), x_3(0), c_1(0), c_3(0)$ and any positive $c_2(0)$. Parameter estimates in this example appeared not very sensitive to the mismatch in the parameter σ . More general recommendations on the choice of the coupling scheme were also suggested from geometrical considerations in Parlitz (1996). Rather different equations for the parameter updates were suggested by several authors, e.g., Konnur (2003).

The synchronisation-based approach is theoretically justified for the noise-free case. Still, numerical experiments have shown its good performance when a moderate measurement noise is present. As it has been already mentioned, the technique can be readily used in the case of hidden variables. However, despite several important advantages of the approach, it may encounter its own significant difficulties. Firstly, an asymptotically stable synchronisation may not be achieved for any observable $\boldsymbol{\eta} = \mathbf{h}(\mathbf{y})$. This possibility depends on the system under study and the coupling scheme. Secondly, it is not always clear what coupling scheme should be used to assure synchronisation at $\mathbf{c} = \mathbf{c}_0$. Finally, it may be very important to select appropriate initial conditions in Eqs. (8.8) and (8.9), in other words, to select starting guesses for the model parameters and hidden variables. Further details can be found in Chen and Kurths (2007); Freitas et al. (2005); Hu et al. (2007); Huang (2004); Konnur (2003); Marino and Miguez (2005); Maybhate and Amritkar (1999); Parlitz (1996); Parlitz et al. (1996); Tao et al. (2004).

8.2.2 Dynamical and Measurement Noise

Estimating model parameters in the case of simultaneous presence of the dynamical and measurement noise is a more complicated task. However, there have been recently developed corresponding sophisticated techniques such as Kalman filtering-based methods (Sitz et al., 2002, 2004; Voss et al., 2004) and Bayesian approaches (Bremer and Kaplan, 2001; Davies, 1994; Meyer and Christensen, 2000). Below, we describe in some detail a recently suggested technique (Sitz et al., 2002) called *unscented Kalman filtering*.

Kalman filtering is a general idea which was originally developed for the state estimation in linear systems from observed noise-corrupted data (Kalman and Bucy, 1961). It is widely used, e.g., in data assimilation (Sect. 5.1) as well as in many other applications (Bar-Shalom and Fortmann, 1988). The idea has been generalised for the estimation of model parameters together with model states. It has been recently further generalised for the estimation of parameters and state vectors in non-linear systems (Sitz et al., 2002).

In the linear case, the model is assumed to be of the form

$$\begin{aligned}\mathbf{x}_i &= \mathbf{A} \cdot \mathbf{x}_{i-1} + \boldsymbol{\xi}_i, \\ \boldsymbol{\eta}_i &= \mathbf{B} \cdot \mathbf{x}_i + \boldsymbol{\zeta}_i,\end{aligned}\tag{8.10}$$

where \mathbf{x} is a state vector of the dimension D , i is the discrete time, $\boldsymbol{\eta}$ is the vector of observables whose dimension may be different from that of \mathbf{x} , \mathbf{A} and \mathbf{B} are constant matrices, and $\boldsymbol{\xi}_n$ and $\boldsymbol{\zeta}_n$ are independent zero-mean Gaussian white noises with diagonal covariance matrices. The problem is to estimate the state \mathbf{x} at a time instant n having observations $\boldsymbol{\eta}_i$ up to time n inclusively, i.e. a set $\mathbf{H}_n = \{\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \dots, \boldsymbol{\eta}_n\}$. Let us denote such an estimator as $\hat{\mathbf{x}}(n|n)$. Kalman filter provides an optimal linear estimate, i.e. unbiased and with the smallest variance. Formally, the estimation procedure consists of the predictor and corrector steps. At the predictor step, one estimates the value $\hat{\mathbf{x}}(n|n-1)$, i.e. takes into account only the observations \mathbf{H}_{n-1} up to time instant $n-1$. The optimal solution to such a problem is $\hat{\mathbf{x}}(n|n-1) = E[\mathbf{x}_n | \mathbf{H}_{n-1}]$. It can be written as

$$\begin{aligned}\hat{\mathbf{x}}(n|n-1) &= E[\mathbf{A} \cdot \mathbf{x}_{n-1} | \mathbf{H}_{n-1}] = \mathbf{A} \cdot E[\mathbf{x}_{n-1} | \mathbf{H}_{n-1}] = \mathbf{A} \cdot \hat{\mathbf{x}}(n-1|n-1), \\ \hat{\boldsymbol{\eta}}(n|n-1) &= E[\mathbf{B} \cdot \mathbf{x}_n | \mathbf{H}_{n-1}] = \mathbf{B} \cdot \mathbf{A} \cdot \hat{\mathbf{x}}(n-1|n-1).\end{aligned}\quad (8.11)$$

The second equation in Eq. (8.11) will be used at the correction step below. The point estimators must be equipped with the confidence bands (see Sect. 2.2.1). It is known that due to the linearity of the system, the estimators are Gaussian distributed. Thus, their confidence bands are simply expressed via their covariance matrices whose optimal estimates read as

$$\begin{aligned}\mathbf{P}(n|n-1) &= E[(\mathbf{x}_n - \hat{\mathbf{x}}(n|n-1)) \cdot (\mathbf{x}_n - \hat{\mathbf{x}}(n|n-1))^T | \mathbf{H}_{n-1}], \\ \mathbf{P}_{\boldsymbol{\eta}\boldsymbol{\eta}}(n|n-1) &= E[(\boldsymbol{\eta}_n - \hat{\boldsymbol{\eta}}(n|n-1)) \cdot (\boldsymbol{\eta}_n - \hat{\boldsymbol{\eta}}(n|n-1))^T | \mathbf{H}_{n-1}], \\ \mathbf{P}_{\mathbf{x}\boldsymbol{\eta}}(n|n-1) &= E[(\mathbf{x}_n - \hat{\mathbf{x}}(n|n-1)) \cdot (\boldsymbol{\eta}_n - \hat{\boldsymbol{\eta}}(n|n-1))^T | \mathbf{H}_{n-1}].\end{aligned}\quad (8.12)$$

These matrices can be expressed via their previous estimates $\mathbf{P}(n-1|n-1)$, $\mathbf{P}_{\boldsymbol{\eta}\boldsymbol{\eta}}(n-1|n-1)$, $\mathbf{P}_{\mathbf{x}\boldsymbol{\eta}}(n-1|n-1)$ analytically for the linear system.

Now, the corrector step updates the predictor-step estimators taking into account the last observation $\boldsymbol{\eta}_n$ as follows:

$$\begin{aligned}\hat{\mathbf{x}}(n|n) &= \hat{\mathbf{x}}(n|n-1) + \mathbf{K}_n \cdot (\boldsymbol{\eta}_n - \hat{\boldsymbol{\eta}}(n|n-1)), \\ \mathbf{P}(n|n) &= \mathbf{P}(n|n-1) - \mathbf{K}_n \cdot \mathbf{P}_{\boldsymbol{\eta}\boldsymbol{\eta}}(n|n-1) \cdot \mathbf{K}_n^T, \\ \mathbf{K}_n &= \mathbf{P}_{\mathbf{x}\boldsymbol{\eta}}(n|n-1) \cdot \mathbf{P}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1}(n|n-1).\end{aligned}\quad (8.13)$$

Thus, the corrections represent discrepancy between the predictor-step estimates and actual observations multiplied by the so-called Kalman gain matrix \mathbf{K}_n . Having Eqs. (8.11), (8.12) and (8.13), one can start from initial guesses $\hat{\mathbf{x}}(1|1)$ and $\mathbf{P}(1|1)$ and recursively get optimal state estimates for all subsequent time instants taking into account subsequent observations. Due to Gaussianity of the distributions, a 95% confidence band for the j th component of the vector \mathbf{x}_n is given by $\hat{x}_j(n|n) \pm 1.96\sqrt{\mathbf{P}_{jj}(n|n)}$.

To apply the idea to a non-linear system $\mathbf{x}_{i+1} = \mathbf{f}(\mathbf{x}_i) + \boldsymbol{\xi}_i$ and $\boldsymbol{\eta}_i = \mathbf{h}(\mathbf{x}_i) + \boldsymbol{\zeta}_i$, one can either approximate non-linear functions with Taylor expansion or simulate

the distribution of states and compute many model orbits to get an estimator $\hat{\mathbf{x}}(n|n)$ and its covariance matrix $\mathbf{P}(n|n)$. The latter idea appeared more fruitful in practice (Sitz et al., 2002). Its fast and convenient implementation includes the selection of the so-called sigma points $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(2D)}$ specifying the distribution of states at a time instant $n - 1$:

$$\begin{aligned}\mathbf{x}^{(j)}(n-1|n-1) &= \hat{\mathbf{x}}(n-1|n-1) + \left[\sqrt{D \cdot \mathbf{P}(n-1|n-1)} \right]_j, \\ \mathbf{x}^{(j+D)}(n-1|n-1) &= \hat{\mathbf{x}}(n-1|n-1) - \left[\sqrt{D \cdot \mathbf{P}(n-1|n-1)} \right]_j,\end{aligned}\quad (8.14)$$

where $j = 1, 2, \dots, D$ and $[\sqrt{\cdot}]_j$ means j th column of the matrix square root. The sigma points are propagated through the non-linear systems giving

$$\begin{aligned}\mathbf{x}^{(j)}(n|n-1) &= \mathbf{f}(\mathbf{x}^{(j)}(n-1|n-1)), \\ \mathbf{y}^{(j)}(n|n-1) &= \mathbf{h}(\mathbf{x}^{(j)}(n|n-1)),\end{aligned}\quad (8.15)$$

where $j = 1, \dots, 2D$. Now, their sample means and covariances define the predictor estimates as follows:

$$\begin{aligned}\hat{\mathbf{x}}(n|n-1) &= \frac{1}{2D} \sum_{j=1}^{2D} \mathbf{x}^{(j)}(n|n-1), \\ \hat{\mathbf{y}}(n|n-1) &= \frac{1}{2D} \sum_{j=1}^{2D} \mathbf{y}^{(j)}(n|n-1), \\ \mathbf{P}_{\eta\eta}(n|n-1) &= \frac{1}{2D} \sum_{j=1}^{2D} (\mathbf{y}^{(j)}(n|n-1) - \hat{\mathbf{y}}(n|n-1)) \cdot (\mathbf{y}^{(j)}(n|n-1) - \hat{\mathbf{y}}(n|n-1))^T, \\ \mathbf{P}_{\mathbf{x}\eta}(n|n-1) &= \frac{1}{2D} \sum_{j=1}^{2D} (\mathbf{x}^{(j)}(n|n-1) - \hat{\mathbf{x}}(n|n-1)) \cdot (\mathbf{y}^{(j)}(n|n-1) - \hat{\mathbf{y}}(n|n-1))^T, \\ \mathbf{P}(n|n-1) &= \frac{1}{2D} \sum_{j=1}^{2D} (\mathbf{x}^{(j)}(n|n-1) - \hat{\mathbf{x}}(n|n-1)) \cdot (\mathbf{x}^{(j)}(n|n-1) - \hat{\mathbf{x}}(n|n-1))^T.\end{aligned}\quad (8.16)$$

The estimates (8.16) are updated via the usual Kalman formulas (8.13). This procedure is called unscented Kalman filtering (Sitz et al., 2002).

If parameter \mathbf{a} of a system $\mathbf{x}_{i+1} = \mathbf{f}(\mathbf{x}_i, \mathbf{a}) + \boldsymbol{\xi}_i$ is unknown, then it can be formally considered as an additional state component. Moreover, it is convenient to consider noises $\boldsymbol{\xi}_i$ and $\boldsymbol{\zeta}_i$ as components of a joint state vector. Thus, joint equations of motion read as

$$\begin{aligned}
\mathbf{x}_{i+1} &= \mathbf{f}(\mathbf{x}_i, \mathbf{a}_i) + \boldsymbol{\xi}_i, \\
\mathbf{a}_{i+1} &= \mathbf{a}_i, \\
\boldsymbol{\xi}_{i+1} &= \boldsymbol{\xi}_i, \\
\boldsymbol{\zeta}_{i+1} &= \boldsymbol{\zeta}_i,
\end{aligned}
\tag{8.17}$$

where the last three equations do not alter starting values of the additional state components. However, the estimates of these components change in time due to the correction step (8.13), which takes into account a new observation.

The technique can be easily generalised to the case of ordinary and stochastic differential equations, where numerical integration scheme would enter the first equation of Eq. (8.17) instead of a simple function \mathbf{f} .

We note that the procedure can be used in the case of hidden variables, which correspond to the situation where the dimension of the observable vector $\boldsymbol{\eta}$ is less than D . Examples with a scalar observable from two- and three-dimensional dynamical systems are considered in Sitz et al. (2002), where efficiency of the approach is illustrated. Thus, even in the case of the deterministically chaotic Lorenz system, the unscented Kalman filtering allowed accurate estimation of the three parameters from a scalar time realisation. Concerning this example, we note two things. Firstly, the successful application of a statistical method (Kalman filtering has its roots in mathematical statistics and the theory of random processes) to estimate parameters in a deterministic system illustrates again a close interaction between deterministic and stochastic approaches to modelling discussed in Chap. 2 (see, e.g., Sect. 2.6). Secondly, the unscented Kalman filtering seems to be more efficient than the multiple shooting approach (Sect. 8.2.1) in many cases since the former technique does not require a continuous model orbit over the entire observation interval. In this respect, the unscented Kalman filtering is similar to the modified multiple shooting approach which allows several discontinuity points. However, the unscented Kalman filtering is easier in implementation, since it does not require to solve optimisation problem for the parameter estimation. On the other hand, the multiple shooting technique would certainly give more accurate parameter estimates for a deterministically chaotic system if one manages to find good starting guesses for the parameters and find the global minimum of the cost function (8.7).

The Kalman filtering-based approach also resembles the synchronisation-based approach: It also involves parameter updates depending on the current model state. However, the difference is that the Kalman filtering does not require any synchronisation between a model and the observed data and is appropriate when a dynamical noise is present in contrast to the synchronisation-based technique developed for deterministic systems.

A proper selection of starting guesses for the model parameters and the hidden variables is of importance for the unscented Kalman filtering as well, since an arbitrary starting guess does not assure convergence of the non-linear recursive procedure [Eqs. (8.16) and (8.13)] to the true parameter values. More detailed discussion

of the unscented Kalman filtering can be found in Sitz et al. (2002, 2004); Voss et al. (2004).

Even more general techniques are based on the Bayesian approach (Bremer and Kaplan, 2001; Davies, 1994; Meyer and Christensen, 2000), where the state and parameter estimators are calculated based on the entire available set of observations rather than only on the previous ones. However, full Bayesian estimation is more difficult to use and implement, since it requires to solve a complicated numerical problem of sampling from multidimensional probability distributions.

8.2.2.1 Concluding Remark

Model validation for the considered “transparent box” problems is performed along two main lines. The first one is the analysis of the model residuals (Box and Jenkins, 1970), i.e. the check for their correspondence to the assumed noise properties (see Sect. 7.3). The second one is the computation of dynamical, geometrical and topological characteristics of a model attractor and their comparison with the respective properties of an original (Gouesbet et al., 2003b) (see Sect. 10.4).

8.3 What One Can Learn from Modelling Successes and Failures

Successful realisation of the above techniques (Sect. 8.2) promises an opportunity to obtain parameter estimates and time courses of hidden variables. It would allow several useful applications such as validation of model ideas, “measurement” of quantities inaccessible to measurement devices and restoration of lost or distorted segments of data. Let us comment it in more detail.

In studying of real-world objects, a researcher never meets a purely “transparent box” setting. He/she can only believe subjectively that a trial model structure is adequate to an original. Therefore, even with a perfect realisation of the procedures corresponding to the final modelling stages (Fig. 5.1), the result may appear negative, i.e. one may not get a valid model with a given structure. Then, a researcher should declare incorrectness of his/her ideas about the process under investigation and return to the stage of the model structure selection. If there are several alternative mathematical constructions, then modelling from time series may reveal the most adequate among them. Thus, the modelling procedure gives an opportunity to reject or confirm (possibly, to make more accurate) some substantial ideas about the object under investigation.

There are a number of practical examples of successful application of the approaches described above. Thus, in Horbelt et al. (2001) the authors confirm validity of their ideas about gas laser functioning and get directly immeasurable parameters of the transition rates between energetic levels depending on the pumping current. In Swameye et al. (2003) the authors are able to make substantial conclusions about the mechanism underlying a biochemical signalling process in cells which is described below.

8.3.1 An Example from Cell Biology

In many applications it is necessary to find out which cell properties determine an undesirable process in the cells most strongly and how one can purposefully affect those properties.¹ To answer such questions, it may be sufficient to get an adequate mathematical model as demonstrated in Swameye et al. (2003).

The authors investigate one of many intracellular signalling pathways, which provide a cell with an opportunity to produce necessary substances in response to variations in surroundings. In particular, such pathways provide reproduction, differentiation and survival of cells. The authors consider the so-called signalling pathway JAK-STAT, which transforms an external chemical signal into activation of a respective gene transcription in a cell nucleus (Fig. 8.5). One of the simplest mathematical models of the process can be written down based on the law of active mass (a usual approach in chemical kinetics) and reads as

$$\begin{aligned}
 dx_1/dt &= -k_1x_1E(t), \\
 dx_2/dt &= -k_2x_2^2 + k_1x_1E(t), \\
 dx_3/dt &= -k_3x_3 + k_2x_2^2/2, \\
 dx_4/dt &= k_3x_3.
 \end{aligned}
 \tag{8.18}$$

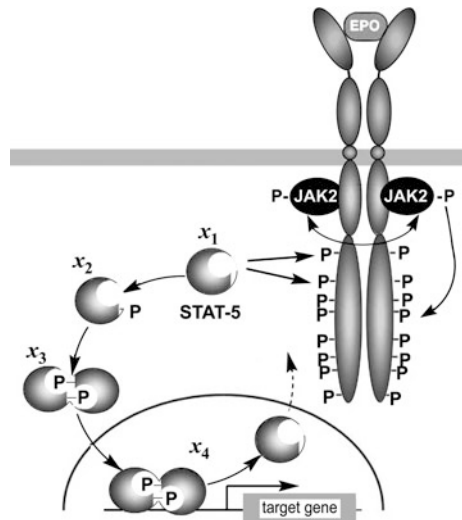


Fig. 8.5 A scheme of a biochemical signalling process in a cell (Swameye et al., 2003)

¹ Thus, growth of cancer cells is determined by the fact that they produce substances “inadequately” to the surrounding situation. A method of struggle against the disease, which is currently only hypothetical, could rely on the empirical modelling similar to that described in Swameye et al. (2003).

Here, k_i are reaction rates, $E(t)$ is the concentration of erythropoietin in the extracellular space (denoted Epo in Fig. 8.5), whose variations lead to the activation of respective receptors on a cell membrane. The receptors are bound to tyrosine kinase of the type JAK-2 existing in a cell cytoplasm. Tyrosine kinase reacts with molecules of the substance STAT5, whose concentration is denoted x_1 . As a result of the reaction, the latter are phosphorylated. There arise monomeric tyrosine phosphorylated molecules STAT5, whose concentration is denoted x_2 . This reaction leads to a decrease in x_1 [the first line in Eq. (8.18)] and an increase in x_2 [a positive term in the second line of Eq. (8.18)]. The monomeric molecules dimerise when they meet each other. Concentration of the dimeric molecules is denoted x_3 . This reaction leads to a decrease in x_2 [a negative term in the second line of Eq. (8.18)] and an increase in x_3 [a positive term in the third line of Eq. (8.18)]. The dimeric molecules penetrate into the nucleus, where their concentration is denoted x_4 . This process leads to a decrease in x_3 [a negative term in the third line of Eq. (8.18)] and an increase in x_4 [the fourth line in Eq. (8.18)]. The dimeric molecules activate the transcription of a target gene. As a result, a specific protein is produced. At that, the dimeric molecules dissociate into the monomeric ones, which *degrade inside the nucleus according to the hypothesis underlying the model* (8.18).

However, there is another hypothesis according to which the monomeric molecules STAT5 are relocated from the cell nucleus to the cytoplasm after a certain delay time. Under such an assumption, the mathematical model slightly changes and takes the form

$$\begin{aligned} dx_1(t)/dt &= -k_1x_1(t)E(t) + 2k_4x_3(t - \tau), \\ dx_2(t)/dt &= -k_2x_2^2(t) + k_1x_1(t)E(t), \\ dx_3(t)/dt &= -k_3x_3(t) + k_2x_2^2(t)/2, \\ dx_4(t)/dt &= -k_4x_3(t - \tau) + k_3x_3(t), \end{aligned} \tag{8.19}$$

where additional time-delayed terms appear in the first and the fourth lines and τ is the delay time. Which of the two models (i.e. which of the two hypotheses) is valid is unknown. Opportunities of the observations are quite limited, an equipment is quite expensive and a measurement process is quite complicated (Swameye et al., 2003). In an experiment, the authors could measure only the total mass of the phosphorylated STAT5 in the cytoplasm η_1 and the total mass of STAT5 in the cytoplasm η_2 up to constant multipliers:

$$\begin{aligned} \eta_1 &= k_5(x_2 + 2x_3), \\ \eta_2 &= k_6(x_1 + x_2 + 2x_3), \end{aligned} \tag{8.20}$$

where k_5, k_6 are unknown proportionality coefficients. Along with those two observables, the concentration $E(t)$ is measured up to a proportionality coefficient (Fig. 8.6a). We stress that all model dynamical variables x_k are hidden. There are

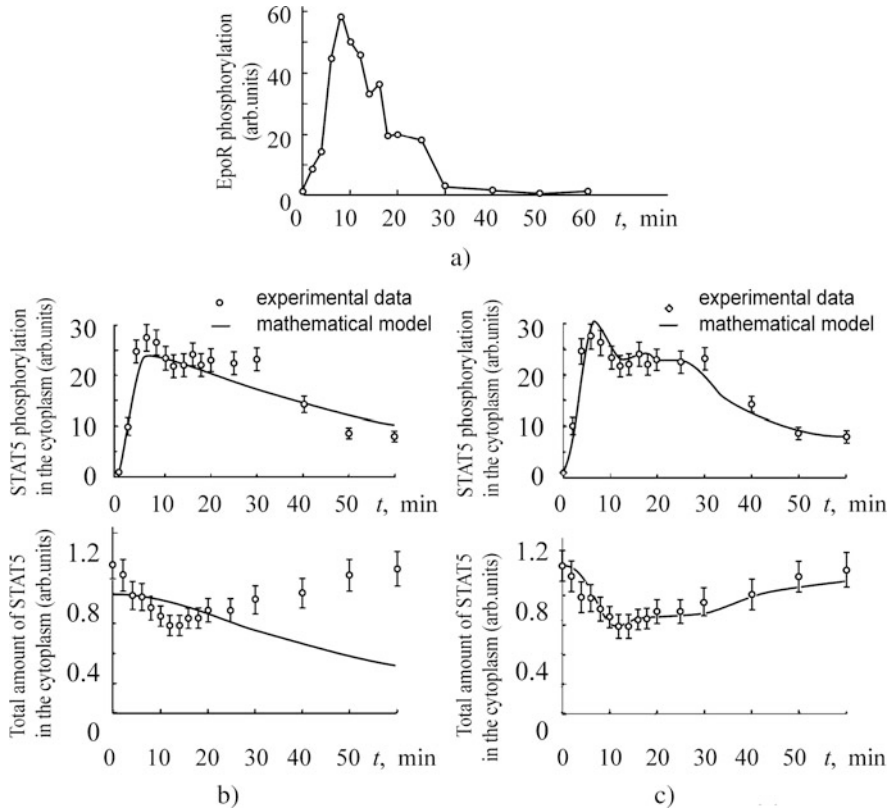


Fig. 8.6 Modelling of a biochemical cell signalling process (Swameye et al., 2003): (a) a time series of an external driving, variations in erythropoietin concentration; (b), (c) the results of an empirical model construction in the form (8.18) and (8.19), respectively

only two observables, which are related to the four dynamical variables in the known way (8.20).

The parameters in both models (8.18) and (8.19) are estimated in Swameye et al. (2003) from the described experimental data. The authors have shown invalidity of the former model and a good correspondence between the experiments and the latter model. Namely, the time series were measured in three independent experiments and one of them is shown in Fig. 8.6. Model parameters were estimated from all the available data via the initial value approach (Sect. 8.2.1). This approach was appropriate since the time series were rather short. They contained only 16 data points per experiment session and represented responses to “pulse” changes in the erythropoietin concentration (Fig. 8.6a). Thus, one should not expect problems with finding the global minimum of the cost function. The model (8.18) appeared incapable of reproducing the observed time series (Fig. 8.6b), while the model (8.19) was adequate in this respect (Fig. 8.6c). The estimate of the delay time was $\tau \approx 6$ min,

which agreed by the order of magnitude with the results of other authors obtained with different techniques for similar objects.

Thus, only the modelling from time series allowed the authors to make a non-trivial conclusion that relocation of STAT5 molecules to the cytoplasm plays a significant role in the process under study. They found out some details of the process, which cannot be observed directly, for instance, the stay of the STAT5 molecules in the nucleus approximately for 6 min.

Moreover, by studying the model (8.19), one can predict what happens if some parameters of the process are varied. For example, the authors studied how the total mass of the protein produced by a cell (which is proportional to a total number of STAT5 molecules participating in the process) changes under variations in different model parameters. This quantity appeared to depend very weakly on k_1 , k_2 and quite strongly on k_3 , k_4 , τ . In other words, the processes in the nucleus play a major role. According to the model (8.19), decreasing k_4 down to zero (inhibition of a nuclear export) leads to the decrease in the produced protein mass by 55%. In experiments with leptomycin B, the authors inhibited a nuclear export (an analogue to the parameter k_4) only by 60%. According to the model, it should lead to the reduction in the amount of activated STAT5 by 40%. In experiment, the reduction by 37% was observed. Thus, a model prediction was finely confirmed, which further increases one's confidence to the model and allows to use it for a more detailed study of the process and its control. Having such achievements, one could think about opportunities to use empirical modelling for medical purposes.

8.3.2 Concluding Remarks

Despite the successes mentioned above, the problem of modelling may often appear technically unsolvable even under the relatively simple “transparent box” setting and for an adequate model structure. To date, examples of successful modelling are observed when the difference between the model dimension and the number of observables is not greater than two or three and the number of estimated parameters is not greater than 3–5. These are only rough figures to give an impression about typical practical opportunities. In each concrete case, modelling results depend on specific non-linearities involved in the model equations. In any case, the greater the number of hidden variables and unknown model parameters, the weaker the chances for a successful modelling and the lower the accuracy of parameter estimators.

References

- Baake, E., Baake, M., Bock, H.J., Briggs, K.M.: Fitting ordinary differential equations to chaotic data. *Phys. Rev. A* **45**, 5524–5529 (1992)
- Bar-Shalom, Y., Fortmann, T.E.: *Tracking and Data Association*. Academic Press, Orlando (1988)
- Bezruchko, B.P., Smirnov, D.A., Sysoev, I.V.: Identification of chaotic systems with hidden variables (modified Bock's algorithm). *Chaos, Solitons Fractals* **29**, 82–90 (2006)

- Bock, H.G.: Numerical treatment of inverse problems in chemical reaction kinetics. In: Ebert, K.H., Deuffhard, P., Jaeger, W., et al. (eds.) *Modelling of Chemical Reaction Systems*, pp. 102–125 Springer, New York (1981)
- Box, G.E.P., Jenkins, G.M.: *Time Series Analysis. Forecasting and Control*. Holden-Day, San Francisco (1970)
- Breeden, J.L., Hubler, A.: Reconstructing equations of motion from experimental data with unobserved variables. *Phys. Rev. A* **42**, 5817–5826 (1990)
- Bremer, C.L., Kaplan, D.T.: Markov chain Monte Carlo estimation of nonlinear dynamics from time series. *Phys. D* **160**, 116–126 (2001)
- Butkovsky, O.Ya., Kravtsov Yu.A., Logunov, M.Yu. Analysis of a nonlinear map parameter estimation error from noisy chaotic time series. *Radiophys. Quantum Electron* **45**(1), 55–66, (in Russian) (2002)
- Chen, M., Kurths, J.: Chaos synchronization and parameter estimation from a scalar output signal. *Phys. Rev. E* **76**, 027203 (2007)
- Davies, M.E.: Noise reduction schemes for chaotic time series. *Physica D* **79**, 174–192 (1994)
- Dennis, J., Schnabel, R.: *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Engle Wood Cliffs, NJ (1983)
- Freitas, U.S., Macau, E.E.N., Grebogi, C.: Using geometric control and chaotic synchronization to estimate an unknown model parameter. *Phys. Rev. E* **71**, 047203 (2005)
- 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, (2003b)
- Horbelt, W., Timmer, J., Büchner, M.J., et al. Identifying physical properties of a CO_2 laser by dynamical modeling of measured time series. *Phys. Rev. E* **64**, 016222 (2001)
- Horbelt, W., Timmer, J.: Asymptotic scaling laws for precision of parameter estimates in dynamical systems. *Phys. Lett. A* **310**, 269–280 (2003)
- Horbelt, W.: *Maximum Likelihood Estimation in Dynamical Systems: PhD Thesis*. University of Freiburg, Freiburg. Available at <http://webber.physik.uni-freiburg.de/~horbelt/diss>. (2001)
- Horbelt, W.: *Maximum Likelihood Estimation in Dynamical Systems: PhD thesis*. University of Freiburg, Freiburg. Available at <http://webber.physik.uni-freiburg.de/~horbelt/diss> (2001)
- Hu, M., Xu, Z., Zhang, R., Hu, A.: Parameters identification and adaptive full state hybrid projective synchronization of chaotic (hyper-chaotic) systems. *Phys. Lett. A* **361**, 231–237 (2007)
- Huang, D.: Synchronization-based estimation of all parameters of chaotic systems from time series. *Phys. Rev. E* **69**, 067201 (2004)
- Ibragimov, I.A., Has'minskii R.Z.: *Asymptotic Theory of Estimation*. Nauka, Moscow (1979). Translated into English Under the Title *Statistical Estimation*: Springer, New York (1981)
- Jaeger, L., Kanz, H.: Unbiased reconstruction of the dynamics underlying a noisy chaotic time series. *Chaos* **6**, 440–450 (1996)
- Judd, K.: Chaotic time series reconstruction by the Bayesian paradigm: Right results by wrong methods?. *Phys. Rev. E* **67**, 026212 (2003)
- Kalman, R.E., Bucy, R.S.: New results in linear filtering and prediction theory. *J. Basic Eng. ASME* **83**, 95–108 (1961)
- Konnur, R.: Synchronization-based approach for estimating all model parameters of chaotic systems. *Phys. Rev. E* **67**, 027204 (2003)
- Marino, I.P., Miguez, J.: Adaptive approximation method for joint parameter estimation and identical synchronization of chaotic systems. *Phys. Rev. E* **72**, 057202 (2005)
- Maybhat, A., Amritkar, R.E.: Use of synchronization and adaptive control in parameter estimation from a time series. *Phys. Rev. E* **59**, 284–293 (1999)
- McSharry, P.E., Smith, L.A.: Better nonlinear models from noisy data: attractors with maximum likelihood. *Phys. Rev. Lett.* **83**, 4285–4288 (1999)
- Meyer, R., Christensen, N.: Bayesian reconstruction of chaotic dynamical systems. *Phys. Rev. E* **62**, 3535–3542 (2000)

- Parlitz, U., Junge, L., Kocarev, L.: Synchronization-based parameter estimation from time series. *Phys. Rev. E* **54**, 6253–6259 (1996)
- Parlitz, U.: Estimating model parameters from time series by auto-synchronization. *Phys. Rev. Lett.* **76**, 1232–1235 (1996)
- Pisarenko, V.F., Sornette, D.: Statistical methods of parameter estimation for deterministically chaotic time series. *Phys. Rev. E* **69**, 036122 (2004)
- 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)
- Sitz, A., Schwarz, U., Kurths, J.: The unscented Kalman filter, a powerful tool for data analysis. *Int. J. Bifurc. Chaos* **14**, 2093–2105 (2004)
- Smirnov, D.A., Vlaskin, V.S., Ponomarenko, V.I.: Estimation of parameters in one-dimensional maps from noisy chaotic time series. *Phys. Lett. A* **336**, 448–458 (2005)
- Swameye, I., Muller, T.G., Timmer, J., et al.: Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by data based modeling. *Proc. Natl. Acad. Sci. USA* **100**, 1028–1033 (2003)
- Tao, C., Zhang Yu., Du, G., Jiang, J.J.: Estimating model parameters by chaos synchronization. *Phys. Rev. E* **69**, 036204 (2004)
- Voss, H.U., Timmer, J., Kurths, J.: Nonlinear dynamical system identification from uncertain and indirect measurements. *Int. J. Bif. Chaos* **14**, 1905–1933 (2004)
- Yule, G.U.: On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Phil. Trans. R. Soc. London A* **226**, 267–298 (1927)