Let us assume that α (growth rate constant) and m_S (maintenance coefficient) are unknown and that we have data on substrate and biomass concentrations in the reactor, the dilution rate D and the inlet substrate concentration S_{in} with sampling period T. We further assume that the yield coefficient Y is known. The parameter vector θ therefore equals:

$$\theta = \begin{bmatrix} \alpha \\ m_S \end{bmatrix} \tag{6.45}$$

By approximating the derivative dS/dt by a finite difference $(S_{i+1}\text{-}S_i)/T$ (where i corresponds to the time instant $t = i.T$), we can rewrite the mass balance (6.44) as follows:

$$S_{i+1} - S_i - T D_i S_{in,i} + T D_i S_i = -\frac{1}{Y} S_i X_i T \alpha + X_i T m_S \tag{6.46}$$

In other words, in the formalism of equation (6.30), y_i and ϕ_i correspond to:

$$y_i = S_{i+1} - S_i - T D_i S_{in,i} + T D_i S_i \quad , \quad \phi_i = \begin{bmatrix} -\frac{1}{Y} S_i X_i T \\ X_i T \end{bmatrix}$$

In the simple case of a linear equation with unknown parameters θ, linear regression can be applied, for instance a weighted least squares objective function:

$$J(\theta) = \sum_{i=1}^{N} \beta_i \left(y_i \left(\hat{\theta} \right) - y_i \right)^2$$

The least squares estimator that minimises this criterion, can easily be deduced to be:

$$\hat{\theta} = \left[\sum_{i=1}^{N} \beta_i \phi_i \phi_i^T \right]^{-1} \sum_{i=1}^{N} \beta_i \phi_i \phi_i^T$$

As mentioned before, a possible choice for the weights β_i are the inverses of the variance σ_i^2 of the measurement y_i.

6.6.2 *Nonlinear Parameter Estimation*

Not all parameter estimation problems can be solved as easily as the problems in which parameters appear linearly in the model. In some exceptional cases nonlinear parameters can be determined analytically by solving (a set of) nonlinear equations. The problems, however, turn very quickly mathematically intractable and one needs to try to find the minimum of a nonlinear objective function. Luckily, finding the minimum of a multivariate function f is a common problem in many research fields and the available expertise is substantial.

Overall, the purpose is to find as efficiently as possible values of θ that make $J(\theta)$ minimal. However, typical for nonlinear functions is that the minimum can

FIG. 6.13. Local and global minima in an objective function.

either be global (the lowest function value in the whole parameter space) or local (the lowest function value in a finite neighbourhood), see Figure 6.13. Despite extensive efforts, no perfect minimisation algorithm for nonlinear objective functions exists (so far), and consequently, finding the global minimum for nonlinear problems cannot be guaranteed [203]. Rather we must accept that additional efforts will be needed and care should be taken to maximise the confidence that one is not ending up with a bad local minimum of the objective function, i.e. ending up with sub-optimal parameter estimates. Consequently, one of the important characteristics of minimisation algorithms will clearly be their sensitivity to local minima.

A helpful visualisation of a nonlinear objective function is a landscape with hills and valleys. The minimisation algorithm should search for the lowest point in this landscape, but can eventually end up in a local minimum instead of the global minimum. Minimisation algorithms typically need a set of initial values where they start from on their quest for the parameter set that gives the lowest J. A property of nonlinear function minimisation is that the minimum found by the algorithm (global or local) can be influenced by the choice of the starting values for the parameters $\theta(0)$. Indeed, a search algorithm can get "stuck" in a local minimum when it comes down one way, whereas it may never come near that local minimum when it comes another way. For instance, in Figure 6.13 one could imagine getting stuck in the local minimum when coming from the left, whereas one would end up in the global minimum when $\theta(0)$ was put on the right side of the figure.

The overall procedure of nonlinear parameter estimation is schematised in Figure 6.14. Initially, the model structure, of which we want to estimate certain selected parameters, and the experimental data need to be specified. To start the algorithm, first guesses of the parameters have to be given. The minimisation algo-

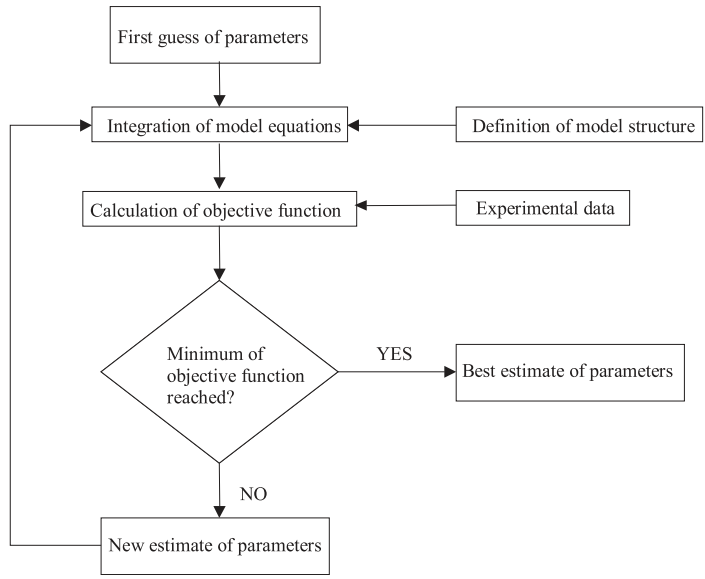

FIG. 6.14. Illustration of parameter estimation routine (modified from [280]).

rithm will then request for model predictions corresponding to this first parameter set. These model predictions are obtained by solving the set of model equations and are passed on to the routine where the objective function is calculated by confronting the predictions with the data. On the basis of rules that are different for each minimisation algorithm either a new proposal for parameters is made and sent to the model solver or, if certain criteria are met, the parameter values are passed on to the user as best estimates. Stopping criteria may be that the maximum number of iterations is reached or that no improvement in objective function is found in recent iterations.

It is important to note here that most calculation time needed to find the best estimates is spent in the box "integration of model equations". Therefore, any approach that can minimise the time spent in this box is very important. For instance, one may aim for the fastest model solution methods. The main gain is, however, obtained by selecting a minimisation algorithm that finds the minimum with the smallest number of iterations through that box.

Schuetze [226] attempted to classify the different minimisation methods in two main groups, local and global minimisation methods. However, he indicated that it may not be possible at all to come up with a strict taxonomy since there are many interrelations between the approaches. For example, most global search methods include at some stage a local procedure in order to refine an approximation of a solution which has been found by the global procedure.

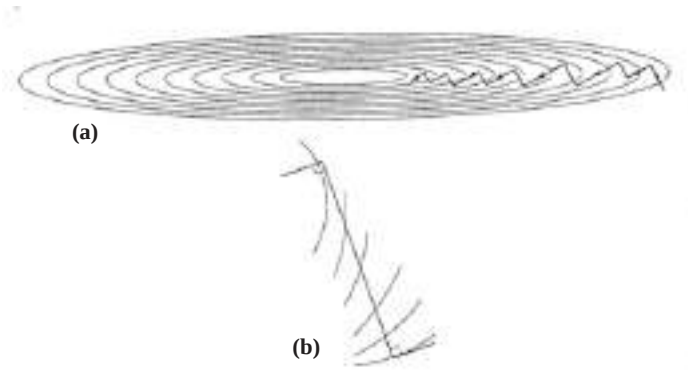

FIG. 6.15. Inefficient steepest decent search along valley.

6.6.3 *Local Minimisation Algorithms Using Derivative Information*

A wide range of minimisation procedures has been developed for locating a local minimum, and many among them make use of information about the gradient of the objective function with respect to the parameters to be estimated. This information is either assumed to be directly available or is computed by numerical approximation.

Steepest descent. The basic idea of these gradient based methods is that, first, a direction is sought in parameter space along which a minimisation of the objective function is pursued. Once the direction is decided upon the step size with which the parameters will be changed is to be determined.

For the steepest descent method, the path of steepest descent is followed as long as J decreases. When the minimum along this direction is reached, a new steepest descent direction is searched for and the parameters are changed according to this new direction. While, theoretically, the method will converge, it may do so in practice with agonising slowness [80] after some rapid initial progress. It happens particularly when the path of steepest descent zigzags slowly down a narrow valley, each iteration bringing only a slight reduction in J (see Figure 6.15).

Gauss-Newton method. A well-known alternative, the Gauss-Newton method, approximates the objective function J locally (around a parameter set θ_i reached at a certain point in the minimisation procedure) by a Taylor series expansion:

$$\begin{aligned} J_i'(\theta) &= J(\theta_i) + \left(\frac{\partial J}{\partial \theta}\Big|_{\theta_i} \right)^T (\theta - \theta_i) + \frac{1}{2} (\theta - \theta_i)^T \frac{\partial^2 J}{\partial \theta^2}\Big|_{\theta_i} (\theta - \theta_i) \\ &= c_i + b_i^T \cdot \delta\theta + \frac{1}{2} \delta\theta^T \cdot A_i \cdot \delta\theta \end{aligned}$$

where $c_i \equiv J(\theta_i)$ $\quad b_i \equiv \nabla J|_{\theta_i}$ $\quad [A_i]_{kl} \equiv \frac{\partial^2 J}{\partial \theta_k \partial \theta_l}\Big|_{\theta_i}$ $\quad \delta\theta = \theta - \theta_i$

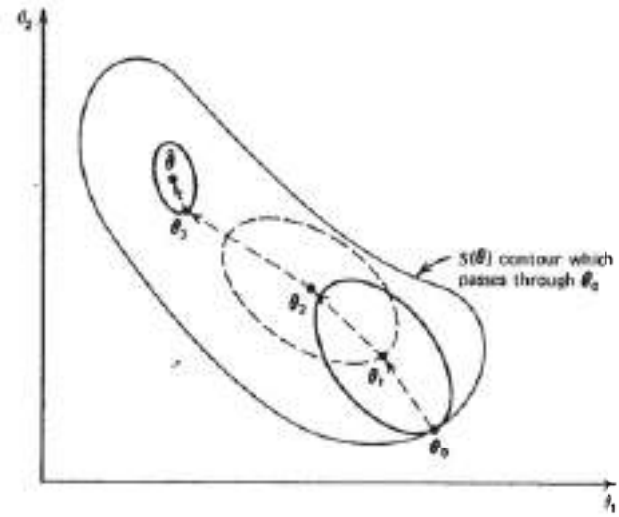

FIG. 6.16. Newton method for parameter estimation by consecutive linearisations of the objective function.

For this approximate linear objective function the minimum can be calculated analytically, i.e. by setting the gradient of J' to zero

$$\frac{\partial J_i'}{\partial \theta} = b_i + A_i \cdot \delta\theta = 0$$

we obtain that the parameter set minimising this approximation of the objective function is found at:

$$\theta_{i+1} = \theta_i - A_i^{-1} b_i$$

Evidently, for a nonlinear problem the minimum obtained in this way is not the true minimum (due to the Taylor series approximation) and the linearisation must be carried out at this point again to make a move to the next minimum, eventually ending in the true minimum. Figure 6.16 shows the procedure for a two-parameter example. Note the nonlinear contour line of J and consecutive linear approximations of J' (ellipsoidal contour lines). The main drawback of the method is that it may converge very slowly (or even diverge) and oscillate around the solution. Our own experience is that the method is not very useful for the type of estimation problems one is encountering in wastewater treatment modelling.

Moreover, the Newton method requires the evaluation of the second derivative matrix A (Hessian) of the objective function at any θ. This is analytically feasible in case the objective function can be written explicitly, but this is often not the case.

Rather, A will need to be approximated numerically, with all the corresponding problems.

When we consider a weighted least squares problem, we can write the elements $[A]_{kl}$ to be:

$$[A]_{kl} \equiv \frac{\partial^2 J}{\partial\theta_k \partial\theta_l} = 2 \sum_{i=1}^{Ndata} \frac{1}{\sigma_i^2} \left[\frac{\partial y_i}{\partial\theta_k} \frac{\partial y_i}{\partial\theta_l} - \left(y_i - \hat{y}_i\,(\theta)\right) \frac{\partial^2 y_i}{\partial\theta_k \partial\theta_l} \right]$$

In the Gauss modification of Newton's method, the second term is neglected because the error $[y_i\text{-}y_i(\theta)]$ multiplying the second derivative should be random and can have either sign. Hence, they tend to cancel out when summed over i [203]. Therefore, we no longer need to evaluate the second derivatives and obtain for the Gauss-Newton methods that

$$\theta_{i+1} = \theta_i + \left(Y_{\theta i} Y_{\theta i}^T\right)^{-1} Y_{\theta i} \varepsilon_i$$

in which $Y\theta_i$ is the matrix of output sensitivities to the parameters and ε_i is the vector of residuals for the parameter set θ_i.

It is to be noted that some "fiddling" with the A-matrix has no effect at all on the final result of the minimisation but will only affect the downhill path that is taken to get there. Indeed, the condition at the minimum that b should be zero is independent of how A is defined.

Levenberg-Marquardt. Probably the best-known modification ("fiddling") of the above two basic methods is the Levenberg-Marquardt algorithm [168] where a compromise is sought between the above inverse Hessian and the steepest-descent method. The idea of Marquardt can be explained briefly as follows. First, both linearisation and steepest-descent methods are "asked" for their optimal direction for a next parameter update step. The Marquardt algorithm then provides a method for interpolating between these two directions and for obtaining a suitable step size as well. We shall not go into any further detail here. This algorithm is good in the sense that it almost always converges and does not slow down as the steepest-descent method often does [80].

Quasi-Newton methods. The basic idea of the quasi-Newton or variable metric methods [203] is to build up, iteratively, a good approximation to the inverse Hessian matrix A^{-1}. Hence, the term "quasi" points to the fact that we only use the current approximation of the inverse Hessian to move forward in the Newton parameter update formula. As explained in Press *et al*. [203] it is found that this approximation often works better than the true Hessian, because it always guarantees in the beginning to move downhill, whereas this guarantee is not given in the Newton method. Of course, close to the minimum the approximation converges to the true Hessian and we can benefit from the quadratic convergence of

Newton's method. The two best known examples of quasi-Newton methods are the BFGS (Broyden, Fletcher, Goldfarb and Shanno) and DFP (Davidon-Fletcher-Powell) methods [87]. These algorithms only differ in details but it has generally become recognised that, empirically, BFGS is superior to DFP. Vanrolleghem and Keesman [264] reported, however, that while these methods converge very quickly, they appear quite sensitive to local minima too.

6.6.4 *Derivative-Free Local Minimisation Algorithms*

In the parameter estimation problems we have discussed in this book on dynamic models, the objective function is defined by a system of ordinary or partial differential equations that are mostly not linear in the system variables and parameters. Consequently, the objective function is evaluated by numerical integration of the dynamic system. Parametric derivatives generally must be found by further integration of a large, derived system (of sensitivity functions), one for each parameter. In situations such as these, derivative-free algorithms are particularly attractive - especially ones that make efficient use of previously computed function values [207]. It is clear that these methods need less preparation for implementation since no calculation of the derivatives must be supported. Also, numerical problems associated with calculating the derivatives may induce convergence problems. In contrast to, for instance, [13], we do not consider methods that use finite differences to approximate the derivative as derivative-free methods. According to us they belong to the first group of local minimisation algorithms.

Rosenbrock method. One of the oldest methods for derivative-free minimisation was introduced by Rosenbrock in 1960 [220]. It was specifically developed for problems in which the objective function is characterised by ridges and valleys and deals with these by rotating the search axes at the end of every stage in the direction of the valley. Riefler *et al*. [217] found this method very insensitive to the local minima which other methods such as BFGS were suffering from.

Brent's algorithm. The direction set method of Powell [202] with refinements proposed by Brent [42] is one of the best derivative-free local minimisation methods [203]. It is based on a repeated combination of one-dimensional searches along a set of various directions [96]. Direction-set methods consist of prescriptions for updating the set of directions as the method proceeds, attempting to come up with a set which either

- includes some very good directions that will take us far along narrow valleys, or else
- includes some number of "non-interfering" directions with the special property that minimisation along one is not "spoiled" by subsequent minimisation along another, so that interminable cycling through the set of directions can be avoided.

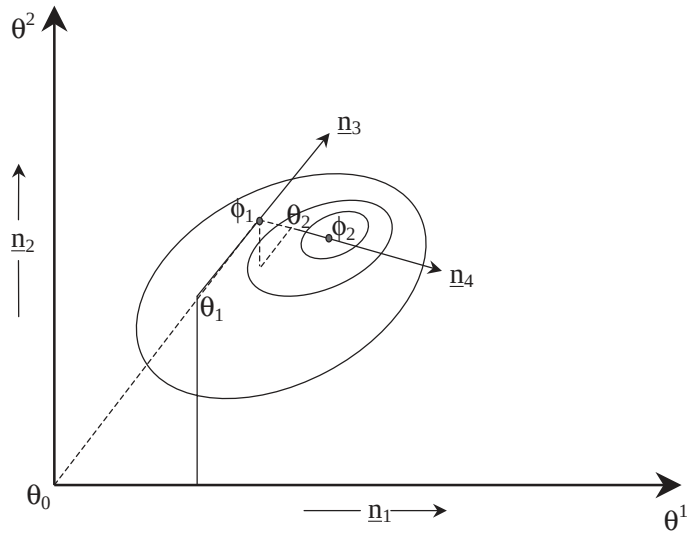

FIG. 6.17. Elementary steps in a direction set algorithm with direction set containing conjugate directions adapting according to Brent [42].

Interesting to mention is that Brent also incorporated some random "jumps" into the procedure to avoid some "local minima" problems that Powell's original algorithm still suffered from.

In Figure 6.17 a typical sequence of searches is presented to illustrate the basic algorithm. We start in an initial parameter guess θ_0 and search sequentially for the minimum along each of the directions in the direction set $\{\underline{n}_1,\underline{n}_2\}$. When all directions are passed through (leading to θ_1), the oldest direction in the set is replaced by a new direction constructed as $\underline{n}_3 = \theta_1\text{-}\theta_0$. This first iteration is ended by searching along this new direction for the minimum ϕ_1. Then a new iteration is started along directions $\{\underline{n}_2,\underline{n}_3\}$ leading to θ_2 and the new direction $\underline{n}_4$ which leads to the minimum ϕ_2. This sequence of steps continues until a stop criterion is reached (e.g. maximum number of allowed iterations or a lack of further decrease in objective function).

Vanrolleghem and Keesman [264] confirmed the statement by Press *et al.* [203] that this method is probably one of the best ones in terms of optimally compromising convergence rate and insensitivity to local minima. Schuetze [226], using the original Powell method, indeed complained about its sensitivity to local minima.

Simplex method. Another well-known local minimisation method, which does not require derivative information, was proposed by Nelder and Mead [184], better known as the simplex method (not to be mixed up with the linear programming simplex method !).

The main concept used by this method is the geometrical concept of a simplex.

A simplex is the geometrical figure consisting, in p dimensions, of $p+1$ points (or vertices) and all their interconnecting line segments, polygonal faces etc. Examples are a triangle ($p=2$) or a tetrahedron ($p=3$). The simplex method only makes use of nondegenerate simplexes, i.e. if any point of a nondegenerate simplex is taken at the origin, then the other p points define vector directions that span the p-dimensional parameter space.

Starting from an initial simplex, elementary steps (Figure 6.18) are developed to find a minimum by evaluating the objective function value at the vertices of the simplex and replacing the vertex with the highest value by another point in p-dimensional space. These steps make sure that the objective function value of each new point of the new simplex is closer to the optimum than the old one. Furthermore, it is ensured that the elementary steps maintain the simplex nondegeneracy property.

The typical progress of iterations is illustrated in Figure 6.19 using a two-dimensional example. Vertices 1,2 and 3 form the initial simplex and increasing numbers indicate the new vertices added at each iteration. Note that vertex 7 has the largest function value for the simplex {4,6,7} but is not reflected immediately since it is the newest vertex in that simplex. When simplex {6,9,10} is reached, vertex 6 has been in the current simplex for four iterations leading to a contraction to the new simplex {6,11,12}. The iteration continues from this simplex and the algorithm will continue to reflect and contract the simplex until the stop criterion has been achieved.

The simplex method is well appreciated for its robustness to local minima, ease of implementation and reasonable convergence rate [203], [264], [226].

Secant or DUD algorithm. The name under which this algorithm was proposed in Ralston and Jennrich [207], DUD (doesn't use derivatives), is clearly indicative of what the authors found to be a main feature of the algorithm. Essentially, the DUD algorithm can be considered a derivative-free Gauss-Newton algorithm. For instance, in the weighted least squares case,

$$J(\theta) = \sum_{i=1}^{Ndata} \frac{1}{\sigma_i^2} \left(y_i - \hat{y}_i(\theta)\right)^2$$

the Gauss-Newton algorithm approximates $\hat{y}_i(\theta)$ by a first order Taylor series expansion about the current value of the parameters θ, whereas the DUD algorithm approximates $\hat{y}_i(\theta)$ by a linear function $F(\theta)$ that exactly agrees with $\hat{y}_i(\theta)$ at $(p+1)$ points. This function describes a secant to the nonlinear function $\hat{y}_i(\theta)$.

The $(p+1)$ points are initially selected as in the Simplex method and an updating mechanism is used to replace one of the $(p+1)$ points. In the weighted least squares case, the objective function calculated with the secant has exactly one minimum which can be easily located. Usually, the point with the highest value of $J(\theta)$ is replaced by this minimum except when it is too close to the best solution found

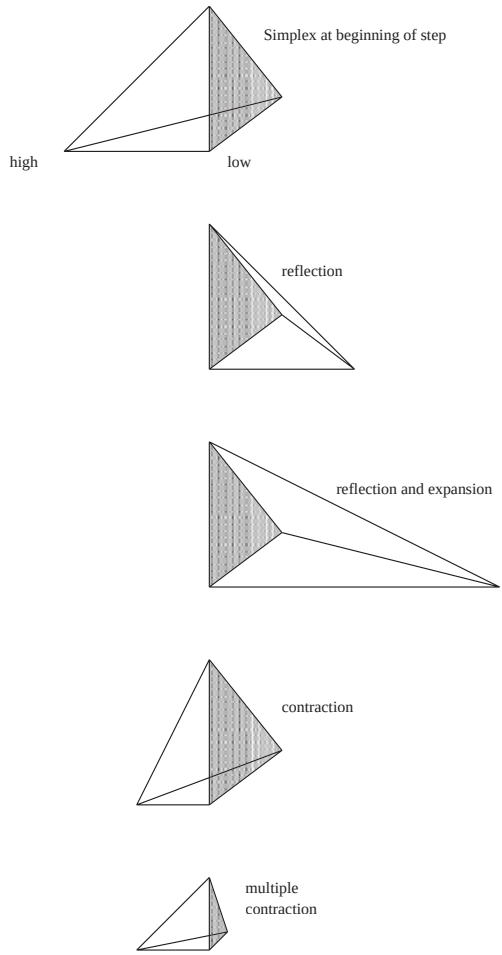

FIG. 6.18. Elementary operations in the simplex minimisation method.

so far. Then, the point with the highest value of J is contracted by a factor of 10 towards the best point to improve the approximation of J by its secant function at the next iteration. If, on the other hand, the minimum found from the secant approximation is larger than any value of the current $(p+1)$ points, convergence is improved by a step reduction mechanism [213].

Reichert [213] extended the original algorithm to take into account individually weighted data and simple inequality constraints. He reported that the parameter transformation approach mentioned in the section on inequality constraints in this chapter could not be used for DUD because the unlimited step size of the

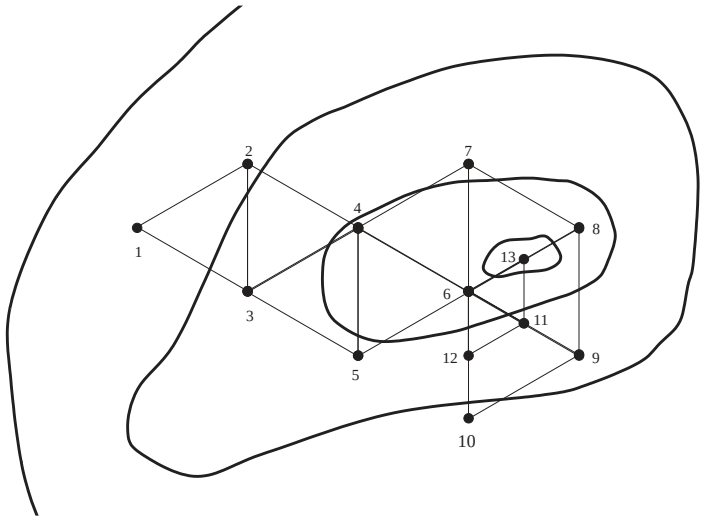

FIG. 6.19. The Simplex algorithm for a two-dimensional optimisation problem.

secant method would lead to rapid divergence of parameters lying on constraints. Rather, the alternative approach to deal with inequality constraints that is based on its translation into an equality constrained problem and minimising an objective function augmented with these constraints and Lagrange multipliers (e.g. [13]) was advocated.

6.6.5 *Global Minimisation Algorithms*

All local minimisation methods suffer from the problem that finding a local minimum does not ensure that a global minimum has been found. Consequently, global minimisation methods have been proposed which do not make use of local properties of the objective function. Thus, the problem of getting stuck in a local minimum is circumvented. However, this advantage of global methods is usually paid for by a higher number of function evaluations necessary for obtaining a solution.

Global optimisation methods can be grouped roughly into two main groups [226], the first of which consists of (purely) deterministic methods, such as gridding. The gridding methods consist of evaluating the objective function at a large number of points predefined on a grid laid in the parameter space. If the number of function evaluations is sufficiently high, there might be some chance to come at least close to minima. This method is clearly very inefficient because it does not allow for any scope of learning from the function evaluations. An obvious refinement is of course to refine the grid after a series of evaluations around the best value found so far.

The second group of global optimisation methods can be termed randomised

methods since random decisions are involved in the search[10]. The degree of randomness varies significantly in different approaches: Uniform random search can be considered as being entirely probabilistic since the objective function is evaluated at every point of a random sample of points in parameter space. Again, as in the gridding approach, no use is being made of information gained in previous evaluations. Methods that do make use of such information are called adaptive random search methods. The main idea here is to perform evaluations of the objective function at points centred around promising points.

Among the first realisations of adaptive random search techniques is the Simulated Annealing technique [203]. The main idea is that the search is not always going towards a candidate solution (eventually ending up in a local minimum) but to allow – occasionally – also a step in a different direction. The probability of such random jumps is slowly decreased during the minimisation process. This method can be regarded as the predecessor of such currently popular methods as the genetic algorithms (GA) and controlled random search (CRS) techniques. These algorithms, which can also be termed "evolutionary" [225], start with an initial population of candidate solutions (a bit similar to the vertices in the simplex method) sampled randomly from the parameter space. The function values of the individuals of this population influence the process of generating new candidate solutions. Various techniques have been proposed for this process. In genetic algorithms [101] new generations of candidate solutions are obtained by imitation of the biological evolution processes of cross-over, mutation and selection of appropriately encoded representations (e.g. bit-strings) of populations of parameters. The definition of the parameters of the algorithm itself is crucial to the success of its application.

In the CRS methods [204], new candidate solutions are generated by reflections on the centre of gravity of the current set of candidate solutions. Duan *et al*. [81] developed the shuffled complex evolution (SCE) algorithm which constitutes an extension of the CRS procedure, to which elements of the simplex method of Nelder and Mead [184] and of competitive evolution were added. In SCE the initial set of candidate solutions is split into different communities ("complexes") which are allowed to develop separately by a combination of CRS and simplex search. From time to time, these complexes are mixed ("shuffled"), thereby passing on information about the parameter space gained independently by each community. It is generally accepted that the SCE method is robust and more efficient than a genetic algorithm procedure [226].

[10]It should be noted that many local minimisation algorithms include random steps to "jump" out of local minima, in this way allowing for increased robustness against this major problem of nonlinear minimisation.

6.6.6 *Comparison of Different Minimisation Algorithms*

Although many studies are certainly available in the literature, only a few are mentioned here for illustrative purposes. In a study by Tseng and Hsu [251] the estimates obtained from three linear transformations of a nonlinear model (Lineweaver & Burk, Eadie & Hofstee and Eisenthal & Cornish-Bowden) were compared with the estimates obtained with their own random search method. They evaluated the quality of fit (expressed as the sum of squared errors) and the computation time for the different methods. A remarkable result was that estimate bias became larger for the transformation-based methods when more data points were included for parameter estimation, again stressing the problems caused by the loss of the required error structure due to these transformations.

In 1992, Fuente and co-workers [94] evaluated a large range of minimisation algorithms for calibration of an activated sludge treatment plant model. They compared steepest descent and the Davidon-Fletcher-Powell methods as examples of unconstrained minimisation and compared different methods for constrained minimisation based on the penalty function method including infinite barriers in addition to log and inverse penalties. The conclusion of the work was that the minimisation algorithms had no difficulties in finding parameter estimates that fitted the data. However, although the fits are similar, the parameter estimates differ substantially which points to identifiability problems.

The study of Vanrolleghem and Keesman [264] compared a series of methods, including BFGS, Simplex, Brent, Random Search and Monte Carlo Set-Membership on identifying Monod-based models with three to six parameters on respirometric data. The results confirmed the relatively high rate of convergence of the BFGS and Brent methods and the robustness to local minima of the Brent, Simplex and the Random Search and Set-Membership methods.

Riefler *et al*. [217] also evaluated the BFGS, steepest descent and Rosenbrock methods on the estimation of biofilm kinetics. The steepest descent method converged too slowly to be practically useful, whereas the BFGS algorithm suffered from sensitivity to local minima as deduced from the sensitivity of its results on the initial parameter guesses. The authors concluded that the Rosenbrock method was most fit for their estimation problem which was characterised by a narrow objective function valley. The steep walls prevented derivative approximations with sufficient accuracy to determine a descent direction on the direction of the valley.

Finally, Schuetze [226] did a comprehensive evaluation of minimisation algorithms, comparing

- the Controlled Random Search (CRS) method,
- a genetic algorithm (GA) according to Caroll and a modification by Krishnakumar, called micro genetic algorithm (μGA), that uses very small population sizes and does not include mutation,

- the original Powell [202] method as presented in [203] without the modifications by [42] since they were considered too computationally demanding,
- a gridding of the parameter space where four values were evaluated in each parameter dimension,
- an approach in which simple substitute models were created for the objective function, i.e. a function was fitted to the objective function values obtained and, given an appropriate choice of the function, its minimum was obtained in an analytical way,
- the Response Surface Methodology (RSM) which requires the building of a quadratic substitute model f over the entire feasible parameter space. The *(p+1)(p+2)/2* coefficients α in p variables (here the parameter values for which we look for a minimum in the objective function)

$$f\left(\theta_1, \cdots \theta_p\right) = \sum_{\substack{i,\, j\, =1 \\ i \geq j}}^{p} \alpha_{ij}\theta_i\theta_j + \sum_{i=1}^{p} \alpha_{i0}\theta_i + \alpha_{00} \tag{6.47}$$

  are determined by least squares regression over all 3^p combinations of the smallest, mean and largest value of the range of each of the p parameters. The parameter estimates are then easily found by minimising f.

The best performance was obtained with the CRS method (giving the lowest objective function within a given maximum number of function evaluations). In this study, the Powell method was found to be quite sensitive to local minima and the methods based on substitute models for the objective function were found to work unsatisfactorily. As expected, allowing a larger number of function evaluations turned out more favourable for the genetic algorithms. As a side remark it was stated that the tangent transformation used to still work with constraints when algorithms for unconstrained minimisation are used, was not very successful in several minimisation runs. Apparently, when the search approaches a region close to the constraint, the tangent transformation (expanding even small differences in the parameters in such region) prevents convergence or further successful search steps of the procedure.

6.7 Evaluation of Parameter Estimation Quality

When a model is calibrated to a data set, it is essential to evaluate the success of this step, even before actual validation of the model is performed. Two aspects need to be evaluated. First, we need to know whether the fitting process itself went well (i.e. did we reach an adequate model fit ?). Second, we want to find out how accurate we have estimated the parameters and conclude whether this accuracy is sufficient for our purpose or whether additional data need to be collected (or a different model built). The first task will, for instance, be based on residuals

analysis whereas the second task will typically focus on the parameter estimation confidence regions, which in most cases will be based on the parameter estimation error covariance matrix. We will, however, also draw attention to alternative methods for parameter accuracy evaluation because the covariance matrix can only be considered an approximation of the true parameter error distribution due to – again – the nonlinearity of the model. These alternative methods will, however, be characterised – again – by high computational demands making their use less popular than the approximate covariance matrix based analysis.

6.7.1 *Residuals Analysis*

Extensive methodology has been developed for investigating whether a calibrated model provides a good description of the data. The methods usually involve examination of the residuals, i.e. the differences between the observed data and the model predictions. It is quite evident that these methods have therefore also been used for structure characterisation (see Chapter 3). Basically, the whole methodology of residuals analysis is based on the feature that an adequate model leaves only measurement error in the residuals. Hence, if the model is appropriate and the model calibration process has been successful, we may expect that the residuals are characterised in the same way as the measurement errors, for instance randomness, homoscedasticity (constant variance), normal distribution. A comparison of their characteristics therefore allows us to conclude whether the calibration was successful.

Analysis of randomness can be based on autocorrelation assessment methods as introduced in Chapter 3 for model selection purposes, i.e. the autocorrelation tests, the runs test, ... A note should be made, however, when correlation analysis is performed on the basis of small data sets. The modeller can indeed be misled when his examination is based on the residuals' (in)dependency since there are only N_{data}-p degrees of freedom left among them after the model identification [208]. On the other hand, the presence of autocorrelation may have statistically serious consequences, more particularly it will lead to an underestimation of the size of the confidence regions, or in other words, we will get too much confidence in our parameter estimates [19]. Thus, investigators should not delude themselves into thinking that more data points are better for least squares estimation of parameters, unless they know that the measurement errors are uncorrelated [218].

Other residuals analysis methods evaluate whether the other assumptions made when choosing the objective function J are fulfilled. For instance, the residuals' homoscedasticity may be evaluated by plotting the residuals against the independent variable (mostly time) or against the measured variable (see for instance Figure 6.2). Trends in the residuals sequence may suggest a switch from an ordinary least squares to a weighted least squares objective function. Testing whether the residuals are indeed normally distributed can be done by performing adequate distribution analysis tests found in statistical handbooks.

6.7.2 *Parameter Estimation Error Covariance Matrix Determination*

Even if the nonlinearity of the model parameters makes it only an approximation and other methods are in principle more correct (see below), the parameter estimation error covariance matrix is the corner stone of parameter estimation accuracy evaluation. Indeed, once this covariance matrix is available, one may use several statistical techniques to evaluate the quality of the estimated parameters. In the next sections some of these techniques are discussed, namely:

- Computation of confidence intervals.
- Determine and draw confidence region ellipses for two parameters.
- Determine whether the zero vector lies within the confidence region.
- Computation and evaluation of correlation values.

Because of its importance, we will therefore review the different methods that are being used to obtain the parameter estimation error covariance matrix.

Linear model. In case we are dealing with a linear parameter estimation problem (see the example in Section 6.6.1), $Y = X\theta$ where Y is the vector of output measurements and X the regressor vector, we obtain that the estimation error covariance matrix can be readily calculated as (under iidN assumption of the residuals):

$$V = \left(X^T X\right)^{-1} \sigma^2$$

where σ^2 is the measurement error, typically estimated as s^2=J_{opt}/(N_{data}-p). In case modelling errors exist, one needs to have an independent measure of the measurement error s_e^2, which can substitute for σ^2. It must be stressed that the matrix inversion that is needed may be troublesome. Indeed, these matrices are often characterised by poor condition numbers due to highly correlated parameters (remember the Modified E criterion for optimal experiment design that could allow this inversion problem to be solved, see Chapter 5).

In case the measurement errors are not equal (but still random and distributed normal), and the weighting of the errors in the objective function J is achieved via the measurement error covariance matrix Q (which is a diagonal matrix if the measurements are uncorrelated but with different variance σ_i^2):

$$Q = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_{Ndata}^2} \end{bmatrix}$$

the parameter estimation error covariance matrix is again readily calculated, i.e.

$$V = \left(X^T Q X \right)^{-1}$$

In some cases this value of V is multiplied with a term $\chi^2/(N_{data}\text{-}p)$ where $\chi^2$is the minimum value obtained for the objective function J_{opt} in this weighted least squares parameter estimation. Indeed, whereas the above equation gives the parameter variance as a function of measurement errors only, the multiplication with χ^2 leads to more reasonable (larger) estimates of errors in cases in which the standard deviations of the measurements do not take all sources of error into account (e.g. some modelling error). The danger involved in employing this multiplication with χ^2 is that systematic errors may be treated as statistical errors [213]. The multiplication with χ^2 is, however, useful when the σ_i^2 are not given in absolute terms, but merely as relative magnitudes, something which typically happens in case no detailed analysis of the measurement errors is conducted (as for instance in the example of Section 6.4.3.). In that case the χ^2 value obtained as a result of the parameter estimation corrects for this.

From the above, it is clear that calculating the covariance matrix does not involve a lot of calculations as it only requires some matrix manipulations with the available data.

Nonlinear model. For nonlinear parameter estimation problems an approximate parameter error covariance matrix can be calculated by replacing the *X*-terms in the above by the output sensitivity functions with respect to the parameters, i.e.

$$V = \left(\frac{\partial y}{\partial \theta}^T Q \frac{\partial y}{\partial \theta} \right)^{-1}$$

where $\partial y/\partial\theta$ is a vector of the output sensitivities at each of the N_{data} measurement points. Note that again we see the link between the parameter estimation error covariance matrix and the Fisher Information Matrix indicated in Chapter 5.

We immediately can see that the actual evaluation of this covariance matrix will involve many more calculations. Indeed, the evaluation of the sensitivity function either requires a simultaneous solution of a considerable set of differential equations (see the simple example given in Chapter 5). Alternatively, a numerical approximation of the sensitivity functions can be made by performing *p* additional simulations around the nominal trajectory where each parameter is perturbed with a small perturbation $\delta\theta$. The adequate choice of this perturbation parameter $\delta\theta$ is the Achilles' heel of this method (see also the similar Nelder and Mead [184] method below).

Luckily, some minimisation algorithms such as the Levenberg-Marquardt and Newton algorithms need or approximate (quasi-Newton and Gauss-Newton methods) the Hessian, which is the inverse of the covariance matrix. Only some minor modification to these algorithms is therefore needed to get access to this Hessian

matrix which after (careful) inversion leads to the parameter estimation error covariance matrix *V*. Some other algorithms also build up information that leads to an approximation of the Hessian, e.g. Brent's algorithm. However, Van Vooren [269] learnt that this approximation is too crude in some parameter estimations since the method only asymptotically converges to the Hessian.

The method proposed by Spendley *et al*. [237], and extended by Nelder and Mead [184], can also be adopted. It allows the covariance matrix around the minimum for any parameter estimation method to be calculated as it is basically an add-on exercise. The technique is based on the construction of a quadratic surface around the minimum of the cost function *f*. If *(p+1)* points in *p* dimensions are given by P_0, P_1, P_p, then "half-way points" $P_{ij}=(P_i+P_j)/2$, $i\neq j$ are calculated, and a quadratic surface is fitted to this combined set of *(p+1)(p+2)/2* points. The points P_i may be taken as:

$$\begin{aligned} P_0 &= \left(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_p\right) \\ P_1 &= \left(\hat{\theta}_1 + \delta_1, \hat{\theta}_2, \cdots, \hat{\theta}_p\right) \\ P_2 &= \left(\hat{\theta}_1, \hat{\theta}_2 + \delta_2, \cdots, \hat{\theta}_p\right) \\ &\vdots \\ P_p &= \left(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_p + \delta_p\right) \end{aligned}$$

in which $\hat{\theta}_i$ is the estimated optimum parameter value and δ_i is the step size, a user-defined small value or a value automatically chosen as function of the machine precision. In order not to exceed the parameter boundaries (e.g. under constrained optimisation), the step may be chosen positive or negative.

The matrix with the step sizes δ is called the direction matrix *D*.

$$D = \begin{bmatrix} \delta_1 & 0 & \cdots & 0 \\ 0 & \delta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \delta_p \end{bmatrix}$$

A quadratic approximation to the objective function *J* in the neighbourhood of the minimum can be obtained using Taylor series expansion

$$J'(\theta) = J(P) + \sum_i \left.\frac{\partial J}{\partial \theta_i}\right|_P \theta_i + \frac{1}{2}\sum_{i,j} \left.\frac{\partial^2 J}{\partial \theta_i \partial \theta_j}\right|_P \theta_i\theta_j + \cdots \quad \cong c - b\cdot\theta + \frac{1}{2}\theta\cdot A\cdot\theta$$

where $c \equiv J(P)$ $\quad b \equiv -\nabla J|_P$ $\quad [A]_{ij} \equiv \left.\frac{\partial^2 J}{\partial\theta_i\partial\theta_j}\right|_P$

The coefficients of the Hessian matrix A can be estimated as:

$$\begin{aligned} a_{ii} &= 2\,(y_i + y_0 - 2y_{0i}) && i = 1, \cdots, p \\ a_{ij} &= 2\left(y_{ij} + y_0 - y_{0i} - y_{0j}\right) && i = 1, \cdots, p \;\; ; \;\; j = 1, \cdots, p \;\; ; \;\; i \neq j \end{aligned}$$

where y_i are the values of the objective function J at P_i and y_{ij} those at P_{ij}. The Hessian or Fisher Information Matrix in the original coordinate system is given by[11]

$$2\left(D^{-1}\right)^T A D^{-1}$$

so that the parameter estimation error covariance matrix is given by

$$\frac{1}{2} D A^{-1} D^T$$

In many cases, the sum of squares of residuals is minimised, and normal equal-variance independent errors are assumed, thus this matrix must be multiplied by $2\sigma^2$ [184]. As usual σ^2 is estimated by $J_{opt}/(N_{data}\text{-}p)$.

Van Vooren [269] implemented this technique and investigated the effect of the step sizes δ_i. It was found that the final results were not much influenced by different choices of the δ_i's. However, the choice of the step size will depend on the rounding errors, and it is advised in Nelder and Mead [184]) that the step size should be at least 10^3 times that rounding error. A too large step size should be avoided too because then the linear approximation of the objective function may no longer hold (depending on its level of nonlinearity). Hence, a compromise must be sought. A pragmatic approach is to perform the calculation with halved step size and compare the results. If they are sufficiently close, they can be accepted.

Clearly, one of the drawbacks of this method is that it requires a considerable number of additional function evaluations [213]. However, in many parameter estimation cases, the extra calculation time necessary to run this extra algorithm is negligible compared to the overall parameter estimation time.

6.7.3 *Confidence Regions*

In order to give a meaningful value for the estimated parameters, they should be supplied together with a confidence region. A confidence region is a p-dimensional interval in which, with a certain probability (e.g. 68.3 %), the true parameter (a p-dimensional vector) lies. The most common confidence region is the one-dimensional region, the confidence interval. A visualisation of a two-dimensional confidence region and a test whether the (n-dimensional) 0 lies within the n-dimensional confidence region are also given below.

[11] Note that in the original paper Nelder and Mead [184], the factor 2 was erroneously omitted as was corrected in an erratum to this paper in *Comput. J.*, 8, 27, 1965.

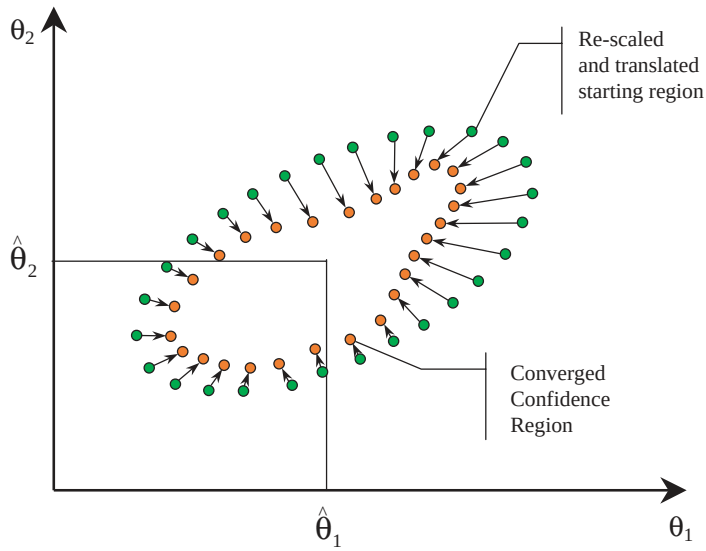

FIG. 6.20. Exact confidence region determination for two parameters θ_1 and θ_2 using the [162] approach.

First, however, we will discuss the actual nature of the confidence region in a nonlinear parameter estimation problem. Let us start from a linear parameter estimation problem in which a weighted least squares objective function was minimised and in which the residuals can be assumed to be independent distributed normal. Under those conditions, the confidence region of the parameter estimates can be exactly calculated from the covariance matrix. As soon as the model is nonlinear in the parameters, this covariance matrix that we have constructed with the methods given above, only allows to approximate the true confidence region. The exact boundaries of a *100(1-α)%* confidence[12] region for the parameter estimate $\hat{\theta}$ are defined as those parameters for which the following equality holds [18]:

$$J_{crit} = J_{opt} \cdot \left(1 + \frac{p}{N_{data} - p} F_{\alpha; p, N_{data}-p}\right) \tag{6.48}$$

where J_{opt} is the minimum value found for the objective function and $F_{\alpha; p, N_{data}-p}$ is the value of the F-distribution with p and $N_{data} - p$ degrees of freedom and a confidence level α.

The *100(1-α)%* confidence region for $\hat{\theta}$ is then the locus of values for θ which result in a value of J below this critical value.

[12]In case we do not know the true residuals' distribution, we are unable to obtain a specified probability level [80]. For residuals with normal distribution, we do have that certainty.

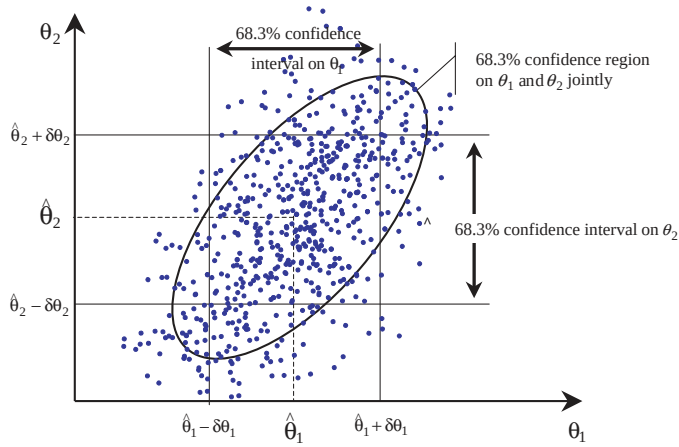

FIG. 6.21. Terminology of confidence intervals in the two-dimensional case.

The problem of finding this boundary may, however, be computation demanding as no easy analytical calculation is feasible. Rather, we have to search for those parameter sets fulfilling the above (in)equalities, i.e. we have to run a considerable number of additional simulations. Typical approaches are to evaluate the parameter space around the parameter estimate $\hat{\theta}$ over a grid or via random sampling using Monte Carlo techniques. However, this method quickly becomes impractical as the number of parameters increases or if the desired resolution increases. Consequently, Lobry *et al.* [162] proposed an elegant, pragmatic method for finding the confidence region boundary with a minimum of additional simulations. The method is introduced in Figure 6.20. First, an initial trial confidence region is proposed that completely encloses the true confidence region, i.e. all $J(\theta)$ should be larger than the value given by the Beale-formula above. Then an algorithm is initiated that makes each of the points converge towards the centre, i.e. where the parameter estimate $\hat{\theta}$ is found. This convergence is stopped when a $J(\theta)$ is found that corresponds to the critical J value given by Beale and a certain α.

Confidence intervals. The difference between a confidence interval in one and two dimensions is illustrated in Figure 6.21. The same fraction of evaluated points lies (i) between the vertical lines, (ii) between the two horizontal lines and (iii) within the ellipse. Overall, a confidence interval of a parameter θ_i is given by $[\theta_i - \delta\theta_i, \theta_i + \delta\theta_i]$ where

$$\delta\theta_i = \sqrt{\Delta\chi_1^2}\sqrt{V_{ii}}$$

V_{ii} is the diagonal element of the covariance matrix corresponding to the i-th parameter. $\Delta\chi_1^2$ will be obtained from distribution data for a given $100(1-\alpha)\%$ confidence and $(N_{data}-p)$ degrees of freedom. For instance, when the 68, 90 and 99%

confidence intervals are to be calculated for one parameter, the $\Delta\chi_1^2$ values to be used are 1.00, 2.71 and 6.63 respectively [203]. For two parameters these $\Delta\chi_1^2$ values would be 2.30, 4.61 and 9.21.

Draper and Smith [80] rightly point to the danger of regarding the square in Figure 6.21 given by the two horizontal and two vertical interval lines, as the joint confidence region. In that case the correlation between the two parameters is completely ignored. The "joint" message of individual confidence intervals should be regarded with caution.

Finally, the marginal confidence intervals of a parameter are given by the projection of the confidence region on the corresponding parameter axis and are therefore bigger than the normal confidence intervals specified above [160].

Confidence region ellipsoids. A confidence region ellipsoid (Fig. 6.22) (two-dimensional confidence region) can be drawn on the basis of the parameter estimation error covariance matrix using the following equation

$$\lambda_i \left(\left[w_{ii} w_{ij} \right] \delta\theta \right)^2 + \lambda_j \left(\left[w_{ji} w_{jj} \right] \delta\theta \right)^2 = \Delta\chi_2^2$$

where $\delta\theta = [\delta\theta_i \quad \delta\theta_j]^T$ and $\Delta\chi_2^2$ are taken for a given *100*(1-α)% confidence region. The ellipsoids will be obtained using

$$\delta\theta_i = \frac{\Delta\chi_2^2}{\lambda_i} w_{ii} cos\,(\phi) - \frac{\Delta\chi_2^2}{\lambda_j} w_{ij} sin\,(\phi)$$
$$\delta\theta_j = \frac{\Delta\chi_2^2}{\lambda_i} w_{ij} cos\,(\phi) - \frac{\Delta\chi_2^2}{\lambda_j} w_{jj} sin\,(\phi)$$

where ϕ varies between 0 and 2π. The w_{ij} and the λ_i are the elements of the eigenvector and eigenvalue decomposition of the covariance matrix V:

$$V = \left[V_{ij} \right] = W \cdot D^{-1} \cdot W^T \qquad W = \left[w_{ij} \right] = \left[\underline{w}_i \right] \qquad D = diag\,(\lambda_i)$$

The centre of the ellipse will be the vector of estimated parameter values.

6.7.4 *Significance of a Single Parameter*

To test whether a parameter estimate is significant, can be done easily by calculating the following t-value [208]

$$t = \frac{\hat{\theta}_i}{\sqrt{V_{ii}}}$$

i.e. the ratio of the parameter estimate $\hat{\theta}_i$ to its standard error, estimated by the square root of the variance of the estimate V_{ii}. This value can then be tested by reference to a Student's t-distribution with *(N_{data}-p)* degrees of freedom and a user-defined confidence level *100(1-α)%*. A high t-value associated with a parameter estimate tends to indicate that the estimate is well determined in the model; conversely, a low t-value tends to indicate that the estimate is poorly determined, although sometimes in multiparameter models, a t-value may be low because of high correlation of the parameter with other parameters in the model.

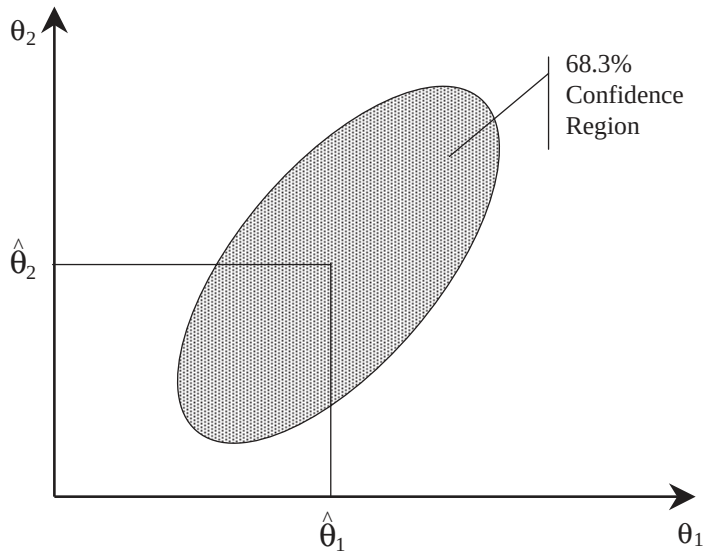

FIG. 6.22. A confidence region ellipsoid for the parameters θ_1 and θ_2.

6.7.5 *Significance of the Parameter Set*

Testing whether *0* lies within the *p*-dimensional *100*(1-α)% confidence region is relevant as it allows assurance that the parameter set found is significantly different from the zero vector. This test can be performed by evaluating whether

$$\Delta\chi_v^2 > \sum_{i=1}^{p} \lambda_i \left(\underline{w}_i \cdot (-\theta)\right)^2$$

Here $\underline{w}_i$ (the columns of *W*) and λ_i are the eigenvectors and the corresponding eigenvalue of *V*. The value $\Delta\chi_v^2$ is obtained for a user-supplied *100(1-α)*% confidence level and a $F_{\alpha;\,p,N_{data}-p}$ distribution. If this inequality holds, *0* will lie within the specified confidence region.

6.7.6 *Correlation Among Parameters*

The correlation coefficient r_{ij} between two parameters θ_i and θ_j gives a measure for the correlation between these two parameters. If r_{ij} is close to *-1* or *1* the parameters are said to be highly correlated. An r_{ij} close to *0* implies a low correlation. The correlation coefficient is given by

$$r_{ij} = \frac{V_{ij}}{\sqrt{V_{ii}V_{jj}}}$$

When a high correlation is encountered this may point to practical identifiability problems that may, for instance, be solved by collecting additional data from an optimally designed experiment (see Chapter 5).

6.8 Conclusions

The purpose of the present chapter was to introduce quite different aspects of parameter estimation from a complete data set. In the next chapter we will focus on how this work can also be done on-line, i.e. with a data set that is continuously growing. To clearly make the difference between both types of estimation, the chapter was introduced with simple, illustrative examples. Next necessary preliminary steps for a parameter estimation problem were introduced, such as the selection of parameters to be estimated, reparametrisation of the model, the guessing of initial values for the parameters. An important section also dealt with the way one deals with constraints on parameter values when using unconstrained minimisation algorithms. Quite some part of this chapter was dedicated to the different aspects of measurement errors since the method that is to be used for parameter estimation depends on the error characteristics. From this and other features, the objective function of the parameter estimation can be selected among the different ones presented in Section 5.5. The real calculation work is then focused upon in Section 6.6, where some linear and nonlinear parameter estimation algorithms are introduced in a rather qualitative manner. Note again that it is important to realise that most of the algorithms presented in this chapter are basically optimisation algorithms that can also be useful in other tasks than parameter estimation, e.g. optimisation of an experiment design. Last but not least, the chapter ends with methods for the assessment of the quality of the parameter estimation performed. It also adresses the important question of the quality of the parameter estimates themselves and the generation of confidence information.

7

Recursive State and Parameter Estimation

7.1 Introduction

This final chapter is dedicated to a question closely related to those already addressed in the preceding chapters. Up to now, we have built models, or we have selected one of them for our specific applications, and we have shown how to calibrate its parameters. In terms of modelling, we can consider that the task is completed. We can use our model(s) for simulating numerically the dynamical behaviour of our processes, possibly for predicting and/or analysing their behaviour. But we can also use the models for monitoring, i.e. to predict the time evolution of the process variables and parameters on-line. The developed monitoring tools can be used to follow the time evolution of variables and/or parameters that are not accessible from on-line measurements; they can also be used for diagnosis about the operation of the plant and help the operator or a supervision system to take appropriate actions to maintain the process in good (ideally, "optimal") operating conditions, diagnose possible process, sensor or actuator failures, or prevent accidents. The supervision can (should) be connected to control loops.

In the following, we shall call these monitoring tools *software sensors*. A software sensor can be defined as an algorithm built from a dynamical model of (part of) the process to estimate on-line unmeasured variables (typically process compo-

nent concentrations like biomass, substrates or products) and/or unknown (or badly known) parameters (e.g. specific reaction rates, or some other kinetic or yield coefficients) from the (few) measurements that are available on-line (typically liquid flow rates, nutrient concentrations (N, P), respiration rate, turbidity, pH, ORP, or gaseous outflow rates). In that sense, these tools can be viewed as "sensors" based on an algorithm (software), i.e. in our nomenclature, as *software sensors*.

In this chapter, we shall introduce two types of software sensors.

1. State observers for reconstructing on-line the time evolution of unmeasured process component concentrations.
2. On-line parameter estimators for unknown or badly known parameters.

The need for monitoring systems and automatic control in order to optimise process operation or detect disturbances in wastewater treatment processes is obvious. Generally speaking, the problems arising from the implementation of these processes are similar to those of more classical industrial processes. Nevertheless, monitoring and automatic control of wastewater treatment processes is clearly developing very slowly. There are at least two main reasons for this:

a) The internal working and dynamics of these processes are as yet badly grasped and many problems of methodology in modelling remain to be solved. It is difficult to develop models taking into account the numerous factors which can influence the specific bacterial growth rate and the yield coefficients which characterise microorganism growth. The modelling effort is often tedious and requires a great number of experiments before producing a reliable model. Reproducibility of experiments is often uncertain due to the difficulty in obtaining the same environmental conditions. Moreover, as these processes involve living organisms, their dynamic behaviour is strongly nonlinear and non-stationary, as we already noted in Chapter 2. Model parameters may not remain constant over a long period[13]. They will vary e.g. due to metabolic variations of biomass or to random and unobservable physiological or genetic modifications. It should also be noted that the lack of accuracy of the measurements often leads to identifiability problems (see Chapter 5).

b) Another essential difficulty lies in the absence, in most cases, of cheap and reliable instrumentation suited to real-time monitoring. To date, the market offers very few sensors capable of providing reliable on-line measurements of the biological and biochemical parameters required to implement high performance automatic control strategies. The main variables, i.e. biomass, substrate and product concentrations, generally need determination through

[13]This may appear ambiguous with the terminology used in this monograph where we consider parameters as being "constant" for all times. However due to the simplicity of the considered models with regard to the complexity of the processes, it appears that different values of the parameter are sometimes necessary to describe satisfactorily the process dynamics in different conditions.

laboratory analyses. The cost and duration of the analyses obviously limit the frequency of the measurements and their timely availability.

To reconstruct the state of the system from the only on-line available measurements and to control [14] [192] biological variables such as the biomass, the substrates or the products, appropriate algorithms have to be developed. The efficiency of any monitoring system highly depends on the design of the monitoring techniques and the care taken in their design. Indeed, monitoring algorithms will prove to be efficient if they are able to incorporate the important well-known information on the process while being able to deal with the missing information (lack of on-line measurements, uncertainty on the process dynamics,...) in a "robust" way, i.e. such that the missing information will not significantly deteriorate the performance of the monitoring system.

In this chapter, we shall show how to incorporate well-known knowledge about the dynamics of wastewater treatment processes (basically, the reaction network and the material balances) in monitoring algorithms which may moreover be capable of dealing with the process uncertainty (in particular on the reaction kinetics) via on-line estimation schemes for the uncertain kinetic parameters.

The chapter is organised as follows. Section 7.2 will concentrate on the notion of state observability as applied to WWTP. The following sections will deal with state observation, i.e. the on-line reconstruction of process components, via different approaches:

1. "classical observers" (extended Luenberger observers, and extended Kalman filter) (Section 7.3);
2. "asymptotic" observers (which are observers independent of the process kinetics) (Section 7.4);
3. an intermediate class of observers, in between classical observers and asymptotic observers, applicable when the reaction rate model structure is known but the model parameters are badly known (Section 7.5).

Section 7.6 will introduce two types of algorithms for recursive estimation of uncertain parameters: a least squares estimation algorithm, and an observer-based estimator. Practical issues and real-life results will be presented in both sections.

7.2 State Observability

Simply speaking, the notion of observability can be defined as the possibility to connect the state variables of a dynamical system to the measured variables via the dynamical model of the system. Essentially, a system is observable if every state variable of the system affects some of the process variables. An important consequence of the observability of a system is the ability to reconstruct the time evolution of the state variables from measured variables in an arbitrary finite time from any initial conditions (see e.g. [151]). This notion is very important in practical applications because it implies that if a system is observable, then it is formally

possible to design an "observer" that is theoretically capable of correctly reconstructing the time evolution of the unmeasured variables, theoretically after a finite arbitrarily chosen time.

Observability is clearly a critical issue in dynamical systems in general and in chemical and biochemical systems in particular. For instance, the implementation of Kalman or Luenberger observers for an application to bioreactors is based on the a priori knowledge of the observability of the process. Because of the nonlinear aspects of their dynamics, the observability analysis is rather complex in biochemical process applications; and the usually large uncertainty in the kinetics of the biochemical reactions and analytical expressions used to describe them makes the approach even more difficult. As a matter of fact, very few works deal with the observability of nonlinear biochemical processes (e.g. [3], [7]) and they are usually concerned with particular process applications.

Let us see how the concept of observability can be formalised mathematically, or more precisely how we can test if a process is observable or not.

For simplicity, we shall consider the notion of observability for *linear* systems (in the state), and for which the matrices (A and B) have constant parameters (i.e. time invariant systems). Let us consider the following dynamical equation system:

$$\frac{dx}{dt} = Ax + Bu \tag{7.1}$$

with dim(x) = N. Consider that the measured variables (the *output*) are denoted y (dim(y) = p) and are related to the state variables x by the following relationship:

$$y = Cx \tag{7.2}$$

where C is a matrix (The simplest obvious case is when the measured variables y are some of the process components (states); then C is a matrix whose entries are "1" or "0"). The linear system (7.1) (7.2) is said to be observable if the following matrix O

$$O = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{N-1} \end{bmatrix} \tag{7.3}$$

is full rank. It is well known (e.g. [151]) that the linear stationary system (7.1), (7.2), is observable if and only if

$$rank(O) = N \tag{7.4}$$

Let us now consider linear approximations of nonlinear models, e.g. of the General Dynamical Model introduced in Chapter 2:

$$\frac{d\xi}{dt} = -D\xi + Y\rho(\xi) + F(\xi) - Q(\xi) \tag{7.5}$$

If we define $f(\xi, D, F, Q)$ as follows:

$$f(\xi, D, F, Q) = -D\xi + Y\rho(\xi) + F(\xi) - Q(\xi) \tag{7.6}$$

the linear approximation of the General Dynamical Model (7.5) around some operating point obtained from the Taylor series expansion of the model equations (also called the *linearised tangent model*, see also Section 2.7.3 in Chapter 2) is written as follows:

$$\frac{dx}{dt} = A(\bar{\xi})x + B(\bar{\xi})u \tag{7.7}$$

where A, B, x are equal to:

$$A(\bar{\xi}) = [\frac{\partial f}{\partial \xi}]_{\xi=\bar{\xi}} = -\bar{D}I_N + Y[\frac{\partial \rho}{\partial \xi}]_{\xi=\bar{\xi}} + [\frac{\partial (F-Q)}{\partial \xi}]_{\xi=\bar{\xi}} \tag{7.8}$$

$$B(\bar{\xi}) = [\frac{\partial f}{\partial u}]_{\xi=\bar{\xi}} \tag{7.9}$$

$$x = \xi - \bar{\xi} \tag{7.10}$$

and u is the input vector which may be constituted by the values varying around the solution $\xi = \bar{\xi}$ of some of the feedrates F_i and/or the dilution rate D.

The observability test can still be applied to the linear model (7.7), but it is worth noting that then the observability test (7.4) is only a sufficient observability condition for the nonlinear system (7.5) (e.g. [152]): if the linearised tangent model (7.7)-(7.10) is observable at $\xi = \bar{\xi}$, then the nonlinear system (7.5) is observable around this point.

7.2.1 *Example #1: Two Step Nitrification Process*

Let us consider the two step nitrification model introduced in Section 2.3.4:

$$S_{NH} \longrightarrow X_1 + S_{NO_2} \tag{7.11}$$

$$S_{NO_2} \longrightarrow X_2 + S_{NO_3} \tag{7.12}$$

characterised by the following matrices and vectors in the General Dynamical Model format:

$$\xi = \begin{bmatrix} S_{NH} \\ S_{NO_2} \\ S_{NO_3} \\ X_1 \\ X_2 \end{bmatrix}, \; Y = \begin{bmatrix} -\frac{1}{Y_1} & 0 \\ Y_3 & -\frac{1}{Y_2} \\ 0 & Y_4 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \; F = \begin{bmatrix} DS_{NH,in} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{7.13}$$

$$\rho = \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix}, \; Q = 0 \tag{7.14}$$

If we assume that the reaction rates are only a function of the components intervening in the reaction (i.e. $\rho_1(S_{NH}, X_1, S_{NO_2})$ and $\rho_2(S_{N0_2}, X_2, S_{NO_3})$), the matrix A of the linearised tangent model is equal to:

$$A(\xi) =$$

$$\begin{bmatrix} -D - \frac{1}{Y_1}\frac{\partial \rho_1}{\partial S_{NH}} & -\frac{1}{Y_1}\frac{\partial \rho_1}{\partial S_{N0_2}} & 0 & -\frac{1}{Y_1}\frac{\partial \rho_1}{\partial X_1} & -\frac{1}{Y_1}\frac{\partial \rho_2}{\partial X_2} \\ Y_3\frac{\partial \rho_1}{\partial S_{NH}} & -D + Y_3\frac{\partial \rho_1}{\partial S_{N0_2}} - \frac{1}{Y_2}\frac{\partial \rho_2}{\partial S_{N0_2}} & -\frac{1}{Y_2}\frac{\partial \rho_2}{\partial S_{N0_3}} & Y_3\frac{\partial \rho_1}{\partial X_1} & -\frac{1}{Y_2}\frac{\partial \rho_2}{\partial X_2} \\ 0 & Y_4\frac{\partial \rho_2}{\partial S_{N0_2}} & -D + Y_4\frac{\partial \rho_2}{\partial S_{N0_3}} & 0 & \frac{\partial \rho_2}{\partial X_2} \\ \frac{\partial \rho_1}{\partial S_{NH}} & \frac{\partial \rho_1}{\partial S_{N0_2}} & 0 & \frac{\partial \rho_1}{\partial X_1} & 0 \\ 0 & \frac{\partial \rho_2}{\partial S_{N0_2}} & \frac{\partial \rho_2}{\partial S_{N0_3}} & 0 & \frac{\partial \rho_2}{\partial X_2} \end{bmatrix}$$

It can be checked that all the states of the linearised tangent model with the above matrix A is observable if S_{NH}, S_{NO_2} and S_{NO_3} are measured, and is not observable if the measured variables are S_{NO_2} and S_{NO_3}, or if these are S_{NH}, X_1 and S_{NO_2} when the reaction rate is independent of S_{NO_3}.

7.2.2 *Simple Local Observability Tests*

Let us now see how to derive a very simple necessary observability condition for the system (7.7), (7.8), (7.9), (7.10). For simplicity, we shall consider that F is independent of ξ and that there is no gaseous outflow rate ($Q = 0$). Let us first define the matrix $\tilde{A}(\bar{\xi})$:

$$\tilde{A}(\bar{\xi}) = Y[\frac{\partial \rho}{\partial \xi}]_{\xi=\bar{\xi}} = Y\rho_\xi \tag{7.15}$$

where ρ_ξ is a more compact notation for $[\frac{\partial \rho}{\partial \xi}]_{\xi=\bar{\xi}}$. Let us denote by $\tilde{O}$, the "observability" matrix computed from $\tilde{A}(\bar{\xi})$. Then we have the following results for the observability of the system (7.7), (7.8), (7.9), (7.10).

Theorem 1: rank(O) $\leq$ min(N, p + R)
with R = rank(Y), N = dim(x), p = dim(y)

Proof: see [71]

The following important consequence can be derived from Theorem 1.

Corollary 1: *A necessary condition for the observability of (7.7), (7.8), (7.9), (7.10) is that:*

$$p + R \geq N \tag{7.16}$$

i.e. the number of measured components + the rank of the yield coefficient matrix must be larger than or equal to the number of process components.

Comment #1: Theorem 1 gives an upper bound for the rank of the observability matrix of system (7.7), (7.8), (7.9), (7.10). This result is obviously local since it applies to the linearised tangent model of (7.5). But it is worth noting that the result is generic in the sense that it is independent of the mathematical structure of the reaction rates $\rho(\xi)$.

Comment #2: the above condition is a necessary condition but not a sufficient one. If it can be useful to detect in a very simple manner the possible lack of observability of the process, the fulfillment of condition p + R $\geq$ N does not guarantee its observability. This will be illustrated in Example #2 here below.

7.2.3 *Example #1: Two Step Nitrification Process (Continued)*

From Theorem 1, we know that system (7.15) will be unobservable if two (or less) components are measured on-line (R + p $\leq 4 < 5$).

7.2.4 *Example #2: Simple Microbial Growth Process*

Let us consider a simple microbial growth process involving one substrate S and one population of microorganisms X, and synthesising a reaction product P. The process can be described by the following reaction scheme:

$$S \longrightarrow X + P \tag{7.17}$$

The dynamical behaviour of the process in a stirred tank reactor is described by the following equations:

$$\frac{d}{dt}\begin{bmatrix} S \\ X \\ P \end{bmatrix} = -D\begin{bmatrix} S \\ X \\ P \end{bmatrix} + \begin{bmatrix} -\frac{1}{Y_1} \\ 1 \\ Y_2 \end{bmatrix}\rho + \begin{bmatrix} DS_{in} \\ 0 \\ 0 \end{bmatrix} \tag{7.18}$$

where Y_1 and Y_2 are the yield coefficients and $\rho = \mu X$. The matrix $\tilde{A}$ of the linearised tangent model is here equal to:

$$\tilde{A}(\bar{S}, \bar{X}, \bar{P}) = \begin{bmatrix} -\frac{1}{Y_1}\rho_S & -\frac{1}{Y_1}\rho_X & -\frac{1}{Y_1}\rho_P \\ \rho_S & \rho_X & \rho_P \\ Y_2\rho_S & Y_2\rho_X & Y_2\rho_P \end{bmatrix} \tag{7.19}$$

with:

$$\rho_i = [\frac{\partial \rho}{\partial i}]_{\xi=\bar{\xi}} \qquad i = S, X, P \tag{7.20}$$

From Theorem 1, we know that the linearised system of (7.18) will be unobservable if only one component is measured on-line (R + p = 2 < 3). As a matter of fact, if, for instance, P is available for on-line measurement, the "observability" matrix $\tilde{O}$ is equal to:

$$\tilde{O} = \begin{bmatrix} 0 & 0 & 1 \\ Y_2\rho_S & Y_2\rho_X & Y_2\rho_P \\ Y_2\rho_S\bar{\rho} & Y_2\rho_X\bar{\rho} & Y_2\rho_P\bar{\rho} \end{bmatrix} \tag{7.21}$$

with $\bar{\rho} = -\frac{1}{Y_1}\rho_S + \rho_X + Y_2\rho_P$.

It is obvious that $\tilde{O}$ is not full rank (the last two rows are proportional to each other). Now, if two components are available for on-line measurement (e.g. S and P), the "observability" matrix becomes:

$$\tilde{O} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ -\frac{1}{Y_1}\rho_S & -\frac{1}{Y_1}\rho_X & -\frac{1}{Y_1}\rho_P \\ Y_2\rho_S & Y_2\rho_X & Y_2\rho_P \\ -\frac{1}{Y_1}\rho_S\bar{\rho} & -\frac{1}{Y_1}\rho_X\bar{\rho} & -\frac{1}{Y_1}\rho_P\bar{\rho} \\ Y_2\rho_S\bar{\rho} & Y_2\rho_X\bar{\rho} & Y_2\rho_P\bar{\rho} \end{bmatrix} \tag{7.22}$$

It will be full rank as long as ρ_X is different from zero. In the particular (but widely encountered) situation when the specific growth rate μ is only a function of the limiting substrate S, we can expect the process to be observable whenever $\mu(S)$ is different from zero (which only happens when $S = 0$ with most kinetic models available in the literature (see [14])).

Finally, not any choice of two components will guarantee the observability of the process. Assume for instance in line with many practical situations that the reaction rate ρ is independent of the product concentration P, i.e. $\rho_P = 0$. Consider now that S and X are measured on-line. Then the "observability" matrix $\tilde{O}$ is equal to:

$$\tilde{O} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1}{Y_1}\rho_S & -\frac{1}{Y_1}\rho_X & 0 \\ Y_2\rho_S & Y_2\rho_X & 0 \\ -\frac{1}{Y_1}\rho_S\tilde{\rho} & -\frac{1}{Y_1}\rho_X\tilde{\rho} & 0 \\ Y_2\rho_S\tilde{\rho} & Y_2\rho_X\tilde{\rho} & 0 \end{bmatrix} \tag{7.23}$$

with:

$$\tilde{\rho} = -\frac{1}{Y_1}\rho_S + \rho_X \tag{7.24}$$

It is obvious that since the third column is equal to zero, rank($\tilde{O}$) < 3 and therefore, the linearised system is not observable. However the result is straightforward to obtain by simply looking at the linearised tangent model equations of S, X, P for ρ_P equal to zero and by recalling that an important condition for the observability of the process is that the dynamical equations of the measured variables must incorporate connections with the unmeasured variables. In the example here, the dynamical equations of S and X no longer contain any term related to P if ρ is taken independent of P.

7.3 Classical Observers

7.3.1 *The Basic Structure of a State Observer*

An interesting alternative to on-line measurement for monitoring biomass, reactant and product concentrations that circumvents and exploits the use of a model in conjunction with a limited set of measurements is the use of Luenberger or Kalman observers. In these techniques, a model which includes states that are measured as well as states that are not measured is used in parallel with the process. The model states may then possibly be used for feedback. This configuration may be used to reduce the effect of noise on measurements as well as to reconstruct the states that are not measured. An introduction to these ideas can be found in e.g. Kwakernaak and Sivan[151].

Let us derive the general structure of state observers. Let us consider the following nonlinear state space model:

$$\frac{dx}{dt} = f(x, u) \tag{7.25}$$

The measured variables, denoted y, are related to the process states by the following relation:

$$y = h(x) \tag{7.26}$$

The general structure of a state observer is then written as follows:

$$\frac{d\hat{x}}{dt} = f(\hat{x}, u) + K(\hat{x})(y - \hat{y}) \tag{7.27}$$

where $\hat{x}$ and $\hat{y}$ are the on-line estimations of x and y given by the state observer (7.27) and:

$$\hat{y} = h(\hat{x}) \tag{7.28}$$

and where $K(\hat{x})$ is the "gain" of the observer. The above observer equation can be interpreted as a copy of the process model plus a correction term $K(y - \hat{y})$ proportional to the output observation error $y - \hat{y}$. This term is indeed the driving

term of the observer; it disappears in presence of perfect estimation (obviously, under the assumption that the process model is perfect too). The design of the state observer consists of choosing an appropriate gain $K(\hat{x})$.

The above state observer was originally developed for linear problems. Because of the nonlinear characteristics of bioprocess dynamics, it is of interest to extend these concepts and exploit particular structures for biochemical engineering application problems. The design of the gain matrix K is based on a linearised version (the linearised tangent model) of the process dynamics observation error (computed from a Taylor's series expansion of a state space model around some operating point. If we define the observation error e as follows:

$$e = x - \hat{x} \tag{7.29}$$

the dynamics of the observation error are readily derived from equations (7.25) and (7.27):

$$\frac{de}{dt} = f(\hat{x} + e, u) - f(\hat{x}, u) - K(\hat{x})(h(\hat{x} + e) - h(\hat{x})) \tag{7.30}$$

If we consider the linearisation of the above equation around the observation error $e = 0$, we obtain:

$$\frac{de}{dt} = (A(\hat{x}) - K(\hat{x})C(\hat{x}))e \tag{7.31}$$

where $A(\hat{x})$ and $C(\hat{x})$ are respectively equal to:

$$A(\hat{x}) = [\frac{\partial f}{\partial x}]_{x=\hat{x}},\ C(\hat{x}) = [\frac{\partial h}{\partial x}]_{x=\hat{x}} \tag{7.32}$$

Thus the design problem can be formulated as the choice of the matrix $K(\hat{x})$ such that the linearised error dynamics (7.31) has desired properties. This has resulted in two typical state observation designs: the extended Luenberger observer, and the extended Kalman observer. The word "extended" obviously emphasises that these observers are extensions of the original linear versions to nonlinear systems. These modified observers, particularly the extended Kalman filter (EKF), have found applications in some biochemical processes (e.g. [21], [22], [24] [242], [153], [287], [47]).

7.3.2 *Extended Luenberger Observer*

In the extended Luenberger observer, the objective is to select $K(\hat{x})$ such that the linearised error dynamics (7.31) are asymptotically stable (i.e. the error converges to zero). This is achieved by choosing $K(\hat{x})$ such that (see e.g. [14]):

1. the matrix $A(\hat{x}) - K(\hat{x})C(\hat{x})$ and its time derivative are bounded:

$$\|A(\hat{x}) - K(\hat{x})C(\hat{x})\| \leq C_1, \qquad \forall \hat{x} \tag{7.33}$$

$$\|\frac{d}{dt}(A(\hat{x}) - K(\hat{x})C(\hat{x}))\| \leq C_2, \qquad \forall \hat{x} \tag{7.34}$$

2. the eigenvalues of $A(\hat{x}) - K(\hat{x})C(\hat{x})$ have strictly negative real parts:

$$Re(\lambda_i[A(\hat{x}) - K(\hat{x})C(\hat{x})]) \leq C_3 < 0, \quad \forall \hat{x} \text{ and } i = 1 \; to \; N \tag{7.35}$$

The importance of the requirement of state observability of the linearised system becomes clearer at this point: if the linearised system is not observable, it is then not possible to assign freely the dynamics of the observation error (or in other words, to have perfect estimation after some defined time), i.e. to select arbitrary eigenvalues λ_i. An example will be given in Section 7.3.4 here below.

7.3.3 *Extended Kalman Observer*

Although the Kalman filter has been originally introduced in a stochastic framework, it can also interpreted as the solution of a (deterministic) optimisation problem (see e.g. [14], [66]). Indeed the design of the extended Kalman observer consists of finding the gain matrix $K(\hat{x})$ that minimises the mean square observation error:

$$E = \int_0^t e^T W e \, d\tau \tag{7.36}$$

with the dynamical model (7.31) ("under the constraints of the dynamical model (7.31)", in the usual notations of optimisation theory). W is a weighting matrix that allows different weightings to the different terms of the error e with a view to standardise the error norm, e.g. when the different components of e are not of the same dimension.

The gain matrix $K(\hat{x})$ can be shown to be equal to:

$$K(\hat{x}) = R(\hat{x})C^T \tag{7.37}$$

where the N $\times$ N symmetric matrix $R(\hat{x})$ is a solution of the following dynamical Riccati matrix equation:

$$\frac{dR}{dt} = -RC^T WCR + RA^T(\hat{x}) + A(\hat{x})R, \quad R = R^T, \; R(0) = R_0 = R_0^T \tag{7.38}$$

As we shall see here below, the choice of the intial values of the entries of the matrix R plays an important role in the convergence properties of the extended Kalman observer.

7.3.4 *Application to the General Dynamical Model*

Let us see how to design extended Luenberger and Kalman observers to the General Dynamical Model (7.5). Let us consider that some (exactly, p) process components are accessible for on-line measurement, i.e. the output vector is equal to:

$$y = C\xi \tag{7.39}$$

with C an elementary matrix (i.e. with only "0" and "1" as entries). The matrix $A(\hat{\xi})$ of the linearised tangent model associated to the General Dynamical Model (7.5) is equal to:

$$A(\hat{\xi}) = -DI_N + Y[\frac{\partial \rho}{\partial \xi}]_{\xi=\hat{\xi}} + [\frac{\partial (F-Q)}{\partial \xi}]_{\xi=\hat{\xi}} \tag{7.40}$$

The observer equations specialise as follows:

$$\frac{d\hat{\xi}}{dt} = -D\hat{\xi} + Y\rho(\hat{\xi}) + F - Q(\hat{\xi}) + K(\hat{x})(y - C\hat{\xi}) \tag{7.41}$$

Then the design equations presented precedingly apply.

Example: simple microbial growth process. Let us consider the following example, i.e. a simple microbial growth process in a continuous stirred tank reactor (CSTR)[14]:

$$S \longrightarrow X \tag{7.42}$$

Recall that the model dynamics in a CSTR are given by the following equations:

$$\frac{dS}{dt} = -\frac{1}{Y_1}\mu X + DS_{in} - DS \tag{7.43}$$

$$\frac{dX}{dt} = \mu X - DX \tag{7.44}$$

With $\xi = [S, X]^T$, the matrix A is now equal to:

$$A = \begin{bmatrix} -\frac{1}{Y_1}\hat{\rho}_S - D & -\frac{1}{Y_1}\hat{\rho}_X \\ \hat{\rho}_S & \hat{\rho}_X - D \end{bmatrix} \tag{7.45}$$

Let us consider as a matter of illustration one particular kinetic model for the specific growth rate μ, e.g. the Haldane model:

$$\mu = \frac{\mu_0 S}{K_S + S + \frac{S^2}{K_I}} \tag{7.46}$$

Assume that the substrate concentration S is available for on-line measurement while the biomass concentration X is not. The state observer specialises as follows. The output matrix C and the gain matrix K of the observer are equal to:

$$C = [1 \;\; 0], \;\; K = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} \tag{7.47}$$

[14]The choice of the bioreactor example is arbitrary, yet it will be used throughout the following sections as the illustration support.

And the matrix $A - KC$ is equal to:

$$A - KC = \begin{bmatrix} -\frac{1}{Y_1}\hat{\rho}_S - D - k_1 & -\frac{1}{Y_1}\hat{\rho}_X \\ \hat{\rho}_S - k_2 & \hat{\rho}_X - D \end{bmatrix} \tag{7.48}$$

The observability matrix is equal to:

$$O = \begin{bmatrix} 1 & 0 \\ -\frac{1}{Y_1}\hat{\rho}_S - D & -\frac{1}{Y_1}\hat{\rho}_X \end{bmatrix} \tag{7.49}$$

By computing the determinant of the matrix O, we can conclude that the system is observable if $\hat{\rho}_X$ is different from zero ($\hat{\rho}_X \neq 0$). If $\rho = \mu X$, this means that $\mu + \frac{\partial \mu}{\partial X} X$ must be different from zero. For the Haldane model, this means that μ must be different from zero, which may happen only if $S = 0$. Hence the system is observable from S.

The observer equations are then equal to:

$$\frac{d\hat{S}}{dt} = -\frac{1}{Y_1}\mu_{max}\frac{\hat{S}}{K_S + \hat{S} + \frac{S^2}{K_I}}\hat{X} + DS_{in} - D\hat{S} + k_1(S - \hat{S}) \tag{7.50}$$

$$\frac{d\hat{X}}{dt} = \mu_{max}\frac{\hat{S}}{K_S + \hat{S} + \frac{S^2}{K_I}}\hat{X} - D\hat{X} + k_2(S - \hat{S}) \tag{7.51}$$

The extended Luenberger observer design consists of choosing the eigenvalues of the matrix $A - KC$. In practice, this can be achieved by computing the characteristic polynomial of $A - KC$ (i.e. $\det(\lambda I - A + KC)$) and comparing it with the characteristic polynomial corresponding to the desired observer dynamics. If you select a desired error dynamics with stable real poles, this latter polynomial is here equal to $(\lambda + \lambda_1)(\lambda + \lambda_2)$ with λ_1 and λ_2 strictly positive (we have two poles because the observer is of order two). The above computation formalises as follows:

$$\begin{aligned} det(\lambda I - A + KC) &= \lambda^2 + \lambda(k_1 + 2D + \frac{1}{Y_1}\hat{\rho}_S - \hat{\rho}_X) \\ &\quad + (\frac{1}{Y_1}\hat{\rho}_S + D + k_1)(D - \hat{\rho}_X) + \frac{1}{Y_1}\hat{\rho}_X(\hat{\rho}_S - k_2) \quad (7.52) \\ &= (\lambda + \lambda_1)(\lambda + \lambda_2) \quad (7.53) \end{aligned}$$

By identifying the different terms of the two characteristic polynomials (7.52) (7.53), we obtain the following relationships:

$$k_1 + 2D + \frac{1}{Y_1}\hat{\rho}_S - \hat{\rho}_X = \lambda_1 + \lambda_2 \tag{7.54}$$

$$(\frac{1}{Y_1}\rho_S + D + k_1)(D - \hat{\rho}_X) + \frac{1}{Y_1}\hat{\rho}_X(\hat{\rho}_S - k_2) = \lambda_1\lambda_2 \tag{7.55}$$

This leads to the following values for the gains k_1 and k_2 from a chosen set of poles $-\lambda_1$ and $-\lambda_2$:

$$k_1(\hat{S}, \hat{X}) = -2D - \frac{1}{Y_1}\hat{\rho}_S + \hat{\rho}_X + \lambda_1 + \lambda_2 \tag{7.56}$$

$$k_2(\hat{S}, \hat{X}) = \hat{\rho}_S - \frac{Y_1}{\hat{\rho}_X}[\lambda_1\lambda_2 - (\frac{1}{Y_1}\hat{\rho}_S + D + k_1)(D - \hat{\rho}_X)] \tag{7.57}$$

The extended Luenberger observer is then given by equations (7.50)(7.51) with the values of k_1 and k_2 computed as above. The implementation of the observer requires explicit values for $\hat{\rho}_S$ and $\hat{\rho}_X$ in the above expression. For the Haldane model, these are equal to:

$$\hat{\rho}_S = \frac{\mu_0\hat{X}(K_S - \hat{S}^2/K_I)}{\left(K_S + \hat{S} + \frac{\hat{S}^2}{K_I}\right)^2},\ \hat{\rho}_X = \frac{\mu_0\hat{S}}{K_S + \hat{S} + \frac{\hat{S}^2}{K_I}} \tag{7.58}$$

Note that the values of the observer gains k_1 and k_2 are changing with time as a function of the estimates $\hat{S}$ and $\hat{X}$.

Let us now consider the extended Kalman observer. The matrix R has the following form here:

$$R = \begin{bmatrix} r_1 & r_3 \\ r_3 & r_2 \end{bmatrix} \tag{7.59}$$

Then the extended Kalman observer is given by equations (7.50)(7.51) in which the gains k_1 and k_2 are calcuated according to the following equations:

$$k_1 = r_1 \tag{7.60}$$

$$k_2 = r_3 \tag{7.61}$$

$$\frac{dr_1}{dt} = -r_1^2 - 2(\frac{1}{Y_1}\hat{\rho}_S + D)r_1 - 2\frac{1}{Y_1}\hat{\rho}_X r_3 \tag{7.62}$$

$$\frac{dr_2}{dt} = -r_3^2 + 2(\hat{\rho}_X - D)r_2 + 2\hat{\rho}_S r_3 \tag{7.63}$$

$$\frac{dr_3}{dt} = -r_1 r_3 + \hat{\rho}_S r_1 - \frac{1}{Y_1}\hat{\rho}_X r_2 + (\hat{\rho}_X - \frac{1}{Y_1}\hat{\rho}_S - 2D)r_3 \tag{7.64}$$

The last three equations (7.62)(7.63)(7.64) are the Riccati equations.

Recall that in the preceding section we have mentioned that the lack of observability will not allow the observer dynamics to be assigned. Let us illustrate this.

Let us consider the simple microbial growth process with a product P (Example #2 in Section 7.2). We have seen that the model is not observable with one

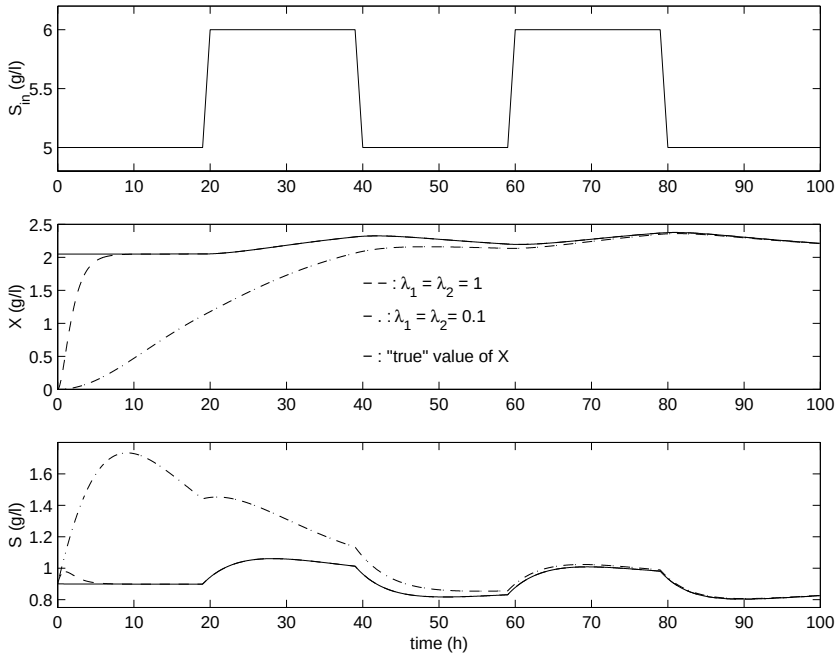

FIG. 7.1. Simulation results of the extended Luenberger observer.

measured component. As in Section 7.2, let us consider that P is the measured component. The matrix $A - KC$ is then equal to:

$$A - KC = \begin{bmatrix} -\frac{1}{Y_1}\hat{\rho}_S - D & -\frac{1}{Y_1}\hat{\rho}_X & -\frac{1}{Y_1}\hat{\rho}_P - k_1 \\ \hat{\rho}_S & \hat{\rho}_X - D & \hat{\rho}_P - k_2 \\ Y_2\hat{\rho}_S & Y_2\hat{\rho}_X & Y_2\hat{\rho}_P - D - k_3 \end{bmatrix} \tag{7.65}$$

It is then routine to check that one of the eigenvalues of the matrix $A - KC$ is equal to $-D$, whatever the values of the gains k_1, k_2 and k_3.

This can be straightforwardly deduced from the characteristic polynomial:

$$det(\lambda I - A + KC) = (\lambda + D)(\lambda^2 + \alpha\lambda + \beta) \tag{7.66}$$

with:

$$\alpha = k_3 + 2D + \frac{1}{Y_1}\hat{\rho}_S - \hat{\rho}_X - Y_2\hat{\rho}_P \tag{7.67}$$

$$\beta = (k_3 + D)(D - \hat{\rho}_X + \frac{1}{Y_1}\hat{\rho}_S) + Y_2(k_1\hat{\rho}_S + k_2\hat{\rho}_X - D\hat{\rho}_P) \tag{7.68}$$

The time response of the observation error is then typically equal to the weighted sum of exponential terms with negative exponents, where one is equal to e^{-Dt}. It

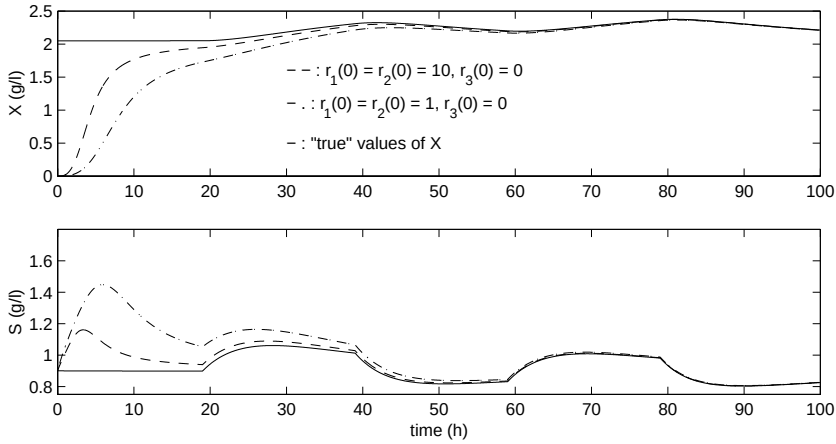

FIG. 7.2. Simulation results of the extended Kalman observer.

is then hopeless to try to look for a convergence rate faster than $1/D$: whatever the values of the observer gains, one pole of the observer will remain equal to $-D$! In the automatic control terminology, the process is not observable (because it is not possible to assign freely the dynamics of the observer) but is detectable (because the dynamics of the observer error is asymptotically stable due to the presence of three poles ($-D$, $-\alpha/2-\sqrt{\alpha^2-4\beta}/2$ and $-\alpha/2+\sqrt{\alpha^2-4\beta}/2$) whose real part is negative).

Simulation results. Let us illustrate the performances of the extended Luenberger observer and of the extended Kalman observer on the above simple microbial growth process with Haldane kinetics, and on the simple microbial growth process with one product P.

The parameters and initial conditions of the simulation model are based on experimental values from a degradation process of lactoserum by *Rhodopseudomonas capsulata* ([275]); they have been set to the following values:

$$\mu_0 = 0.33\ h^{-1},\ K_S = 5\ g/l,\ K_I = 25\ g/l,\ Y_1 = 0.5,\ Y_2 = 0.6$$
$$S(0) = 0.9\ g/l,\ X(0) = 2.05\ g/l,\ P(0) = 1.22\ g/l$$
$$S_{in}(0) = 5\ g/l,\ D(0) = 0.05\ h^{-1}$$

Figures 7.1 and 7.2 illustrate the performances of the extended Luenberger observer (ELO) and of the extended Kalman observer (EKO), respectively with a square variation of the influent substrate S_{in} for different observer gain values. The higher the values of the poles λ_1 and λ_2 of the ELO (from 0.1 to 1, here), the quicker the convergence of the estimated values. Similarly, the higher the initial values of the entries of the matrix R (from 1 to 10), the better the convergence of

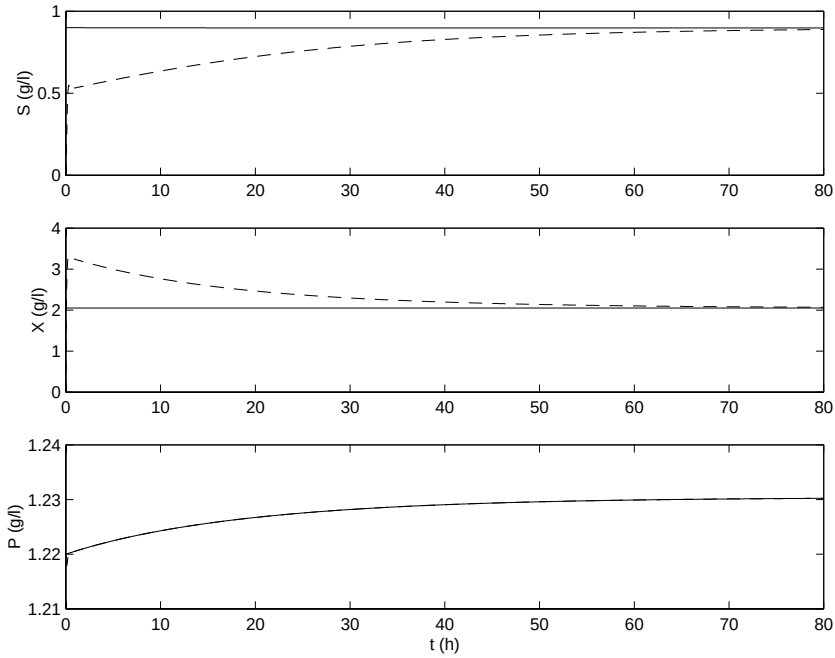

FIG. 7.3. Simulation results of the extended Luenberger observer for an unobservable process.

the EKO. The initial values of the estimates in the observers have been set to the following values: $\hat{X}(0) = 0\ g/l$, $\hat{S}(0) = 0.9\ g/l$ (ELO & EKO) $= 0\ g/l$ (ELO for the unobservable process), $\hat{P}(0) = 1.22\ g/l$ (ELO for the unobservable process).

Figure 7.3 illustrates the inability to assign arbitrarily the dynamics of a classical observer (here, the ELO) when the system is unobservable, as is the case for the simple microbial growth with only one on-line measurement (here, P). The chosen poles λ_1 and λ_2 correspond to much faster dynamics than the residence time $1/D$ ($\lambda_1 = \lambda_2 = 10 >> D = 0.05$). We note on Figure 7.3 that after a fast transient, the dynamics of the observer are dominated by the residence time, i.e. $1/D$ (it takes about three times the residence (60 hours) for the observer to converge).

7.3.5 *Performance of Classical Observers in Presence of Model Uncertainties*

Let us now illustrate the performance of a classical observer when some of the model parameters are badly known. Let us consider here the ELO applied to the estimation of the biomass concentration X from measurements of the substrate

concentration S in a simple microbial growth process[15]. Let us consider Monod kinetics[16]:

$$\mu = \mu_{max} \frac{S}{K_S + S} \tag{7.69}$$

with μ_{max} the maximum specific growth rate (h^{-1}) and K_S the saturation constant (g/l). Assume that the yield coefficient Y_1 is known while the kinetic parameters μ_{max} and K_S may be such that their values may have been determined with some uncertainty. The observer equations are then equal to:

$$\frac{d\hat{S}}{dt} = -\frac{1}{Y_1}\tilde{\mu}_{max}\frac{\hat{S}}{\tilde{K}_S + \hat{S}}\hat{X} + DS_{in} - D\hat{S} + k_1(S - \hat{S}) \tag{7.70}$$

$$\frac{d\hat{X}}{dt} = \tilde{\mu}_{max}\frac{\hat{S}}{\tilde{K}_S + \hat{S}}\hat{X} - D\hat{X} + k_2(S - \hat{S}) \tag{7.71}$$

$\hat{S}$ and $\hat{X}$ represent the estimations of S and X given by the observer. The parameters $\tilde{\mu}_{max}$ and $\tilde{K}_S$ used in the observer computation may be different from their "true" values μ_{max} and K_S.

Recall that in the ELO for this system, the gains k_1 and k_2 are selected as follows:

$$k_1 = \lambda_1 + \lambda_2 - \frac{1}{Y_1}\frac{\tilde{K}_S\tilde{\mu}_{max}\hat{X}}{(K_S + \hat{S})^2} + \frac{\tilde{\mu}_{max}^2\hat{S}}{\tilde{K}_S + \hat{S}} - 2D \tag{7.72}$$

$$k_2 = Y_1\frac{\tilde{K}_S + \hat{S}}{\tilde{\mu}_{max}\hat{S}}[-\lambda_1\lambda_2 + (\lambda_1 + \lambda_2)(D - \frac{\tilde{\mu}_{max}\hat{S}}{\tilde{K}_S + \hat{S}}) - (D - \frac{\tilde{\mu}_{max}\hat{S}}{\tilde{K}_S + \hat{S}})^2 + \frac{1}{Y_1}\frac{\tilde{\mu}_{max}\tilde{K}_S\hat{S}\hat{X}}{(\tilde{K}_S + \hat{S})^3}] \tag{7.73}$$

in order to have observer dynamics assigned to desired values λ_1 and λ_2 (see Section 7.3.1). Let us test the performance of the observer with a wrong value of one of the kinetic parameters. The numerical simulation conditions are the following:

$$Y = 0.5,\ \mu_{max} = 0.33\,h^{-1},\ K_S = 5\,g/l,\ S_{in} = 5\,g/l$$
$$D = 0.05\ h^{-1},\ X(0) = 1\,g/l,\ S(0) = 0.5\,g/l$$

The observer has been initialised with $\hat{S}(0) = S(0)$ and $\hat{X}(0) = 0$. Figure 7.4 shows the estimation results with 10 % error on K_S ($\tilde{K}_S$ = 4.5 g/l instead of 5) for two sets of design parameters k_1 and k_2 (one corresponding to $\lambda_1 = \lambda_2$ = -0.1 ("slow" dynamics), the other for $\lambda_1 = \lambda_2$ = -10 ("fast" dynamics))[17]. The results are quite

[15]The choice of the extended Luenberger is obviously arbitrary and is not based on any priori idea about its advantages and drawbacks.

[16]The choice of the Monod model is obviously also arbitrary. For instance, we could have as well chosen the Haldane model in the preceding section.

[17]Similar results are obtained with an error on μ_{max}.

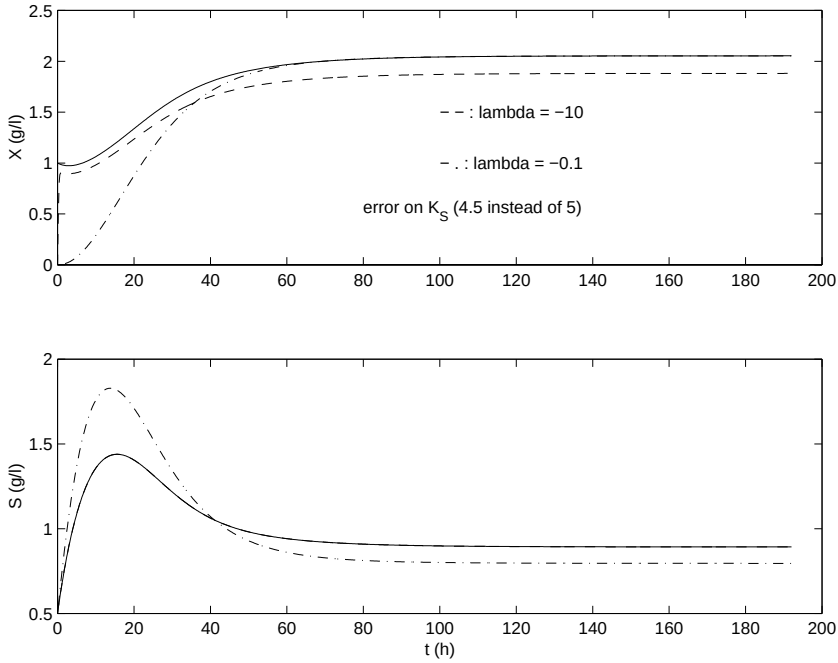

FIG. 7.4. Extended Luenberger observer with a wrong value of K_S (initial transient)(-: "true" simulated values).

typical of classical observers but also of any "high gain" observers for nonlinear systems: the higher the gain, the worse the state estimation results. Indeed high gain (fast dynamics) will result in a good estimation of the measured variable (here S)(it is impossible to distinguish on the figure between S and $\hat{S}$!) while rejecting the parameter uncertainty on the estimated variable (here X). This simple example shows that there is a need to develop observers that can handle parameter uncertainty, a typical situation encountered in (bio)chemical process applications.

7.4 Asymptotic Observers

One of the reasons for the popularity of the EKO/ELO is that it is easy to implement since the algorithm can be derived directly from the state space model. However, since (as the ELO) it is based on a linearised model of the process, the stability and convergence properties are essentially local and valid only around some operating point, and it is rather difficult to guarantee its stability over wide ranges of operation. Ljung [158] shows that the EKO for state and parameter estimation of linear systems may give biased estimates or even diverge if it is not carefully initialised.

One reason for the problem of convergence of EKO/ELO is that, in order to guarantee the (arbitrarily chosen) exponential convergence of the observer, the process must be locally observable, i.e. the linearised tangent model must be observable and fulfill the classical observability rank condition. This condition, as it turns out, is restrictive in many practical situations (as illustrated in Figure 7.3) and may account for the failure of EKO/ELO to find widespread application (e.g. [14], [16], [71]).

Another problem is that the theory for the extended Luenberger and Kalman observers is developed using a perfect knowledge of the system model and parameters, in particular of the process kinetics: it is difficult to develop error bounds and there is often a large uncertainty on these parameters. The performance of ELO/EKO in the presence of badly known parameters has been illustrated in the preceding Section 7.3.5.

It appears from the above remarks that there is a clear incentive to develop alternative methodologies for the on-line estimation of the unmeasured concentration variables in wastewater treatment processes that do not rely on parts of the dynamical model that may be largely uncertain, and that do not require in particular the explicit use of kinetic models. Indeed, the objective of this section is to propose an alternative to EKO/ELO and use process mechanisms in a more direct manner to develop a nonlinear observer applicable to the estimation problem of wastewater treatment processes. The proposed observer is based on the well-known nonlinear model of the process without the knowledge of the process kinetics being necessary. In order to advance the application of this method, we discuss its stability and convergence properties. We would like to emphasise that the presented results are global (i.e. independent of the initial conditions) as opposed to the local properties for EKO/ELO (see e.g.[14]).

This section is organised as follows. We shall first present the general methodology for single tank bioprocesses and discuss its theoretical convergence properties and the practical implementation aspects. Then we shall present a real-life application on an anaerobic digestion process. Finally we shall introduce the extension to fixed bed reactors.

7.4.1 *Asymptotic Observers for Single Tank Bioprocesses*

The derivation of the asymptotic observer equations are based on the Key State Transformation introduced in Section 2.8 and on the following assumptions:

1. p ($\geq$ M (the number of reactions)) components are measured on-line.
2. The feedrates F, the gaseous outflow rates Q and the dilution rate D are known either by measurement or by choice of the user.
3. The yield coefficient matrix Y is known.
4. The reaction rate vector ρ is unknown.
5. The M reactions are irreversible and independent, i.e. rank(Y) = R = M

From assumption 1, we can define the following state partition:

$$\xi = \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} \tag{7.74}$$

where ξ_1 and ξ_2 hold for the measured component concentrations and the unmeasured ones, respectively.

Let us recall the state transformation ζ introduced in Chapter 2 (Section 2.8):

$$\zeta = C_a \xi_a + C_b \xi_b \tag{7.75}$$

where C_a and C_b are solutions of the matrix equation:

$$C_a Y_a + C_b Y_b = 0 \tag{7.76}$$

and with the following dynamical equations (independent of the reaction rate ρ!):

$$\frac{d\zeta}{dt} = -D\zeta + C_a(F_a - Q_a) + C_b(F_b - Q_b) \tag{7.77}$$

Let us consider one (arbitrarily chosen) transformation ζ defined by (7.75, 7.76). The variable ζ can be rewritten as a linear combination of the measured and unmeasured states ξ_1 and ξ_2, i.e.:

$$\zeta = A_1 \xi_1 + A_2 \xi_2 \tag{7.78}$$

The equations (7.78)(7.77) are the basis for the derivation of the asymptotic observer. The dynamical equations of ζ are used to calculate an estimate of ζ on-line, which is used, via equation (7.78) and the on-line data of ξ_1, to derive an estimate of the unmeasured component ξ_2. Let us further assume that the matrix A_2 is (left) invertible. The following two cases can be differentiated.

Case #1: p = M

In this case, the asymptotic observer is written as follows:

$$\frac{d\hat{\zeta}}{dt} = -D\hat{\zeta} + C_a(F_a - Q_a) + C_b(F_b - Q_b) \tag{7.79}$$

$$\hat{\xi}_2 = A_2^{-1}[\hat{\zeta} - A_1 \xi_1] \tag{7.80}$$

Comment: if we consider the most simple and straightforward choice for the state transformation ζ, i.e. with $\xi_1 = \xi_a$ and $\xi_2 = \xi_b$, and with $C_b = I_{N-p}$, then we have:

$$A_1 = -Y_2 Y_1^{-1}, \; A_2 = I_{N-p} \tag{7.81}$$

And therefore the condition on the invertibility of A_2 is in fact a condition on the invertibility of the submatrix Y_1 (i.e. Y_1 is full rank or rank(Y_1) = M). This condition is indeed fulfilled from assumption 5 (independent reactions) and if the measured variables are independent.

Case #2: $p > M$

Assume that the number of measured components is larger than the number which is strictly necessary than the one needed, i.e. p > M. Then the asymptotic observer is modified as follows:

$$\frac{d\hat{\zeta}}{dt} = -D\hat{\zeta} + C_a(F_a - Q_a) + C_b(F_b - Q_b) \tag{7.82}$$

$$\hat{\xi}_2 = A_2^+[\hat{\zeta} - A_1\xi_1] \tag{7.83}$$

where A_2^+ is a left inverse of A_2 (see the example (case #3) here below).

The observer (7.79)(7.80) or (7.82)(7.83) is completely independent of the process kinetics and can be implemented without the knowledge of the reaction rates $\rho(\xi)$ being required.

Theoretical convergence of the asymptotic observer. The convergence properties can be summarised in the following theorem.

Theorem 2: *If the dilution rate D is a persistently exciting signal, i.e. if there exist positive constants δ and β such that:*

$$\int_t^{t+\beta} D(\tau)d\tau \geq \delta > 0 \tag{7.84}$$

then:

$$\lim_{t\to\infty}(\xi_2 - \hat{\xi}_2) = 0 \tag{7.85}$$

Proof: the proof of the theorem is immediate if one observes that, from (7.77), (7.79), (7.80), (7.83), the dynamics of the estimation error is equal to:

$$\frac{d(\xi_2 - \hat{\xi}_2)}{dt} = -D(\xi_2 - \hat{\xi}_2) \tag{7.86}$$

QED

Remark #1: the persistence-of-excitation condition on D simply requires that D is not equal to zero for too long. This condition is clearly easily fulfilled in fedbatch and continuous reactors.

Remark #2: the general formulation of the estimation algorithm (7.82), (7.83) with the introduction of the left inverse allows for larger flexibility in the use of the asymptotic observer since it permits the possible presence of a number of measured variables larger than M to be taken into account. Let us illustrate this in the following example.

Example: anaerobic digestion process. Let us consider the simplified two step anaerobic digestion reaction network:

$$S_1 \longrightarrow X_1 + S_2 + P_1 \qquad \text{(acidogenesis)} \tag{7.87}$$

$$S_2 \longrightarrow X_2 + P_1 + P_2 \qquad \text{(methanisation)} \tag{7.88}$$

where S_1, S_2, X_1, X_2, P_1 and P_2 represent the organic matter, the volatile fatty acids, the acidogenic bacteria, the methanogenic bacteria, the carbon dioxide, and the methane, respectively. The associated general dynamical model is given by the following vectors and matrices:

$$\xi = \begin{bmatrix} S_1 \\ S_2 \\ X_1 \\ X_2 \\ P_1 \\ P_2 \end{bmatrix}, \; Y = \begin{bmatrix} -\frac{1}{Y_1} & 0 \\ Y_3 & -\frac{1}{Y_2} \\ 1 & 0 \\ 0 & 1 \\ Y_4 & Y_5 \\ 0 & Y_6 \end{bmatrix}, \; F = \begin{bmatrix} DS_{in} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \; Q = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ Q_1 \\ Q_2 \end{bmatrix} \tag{7.89}$$

$$\rho = \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} = \begin{bmatrix} \mu_1 X_1 \\ \mu_2 X_2 \end{bmatrix} \tag{7.90}$$

Let us first define one state transformation ζ, e.g.:

$$\xi_a = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \qquad \xi_b = \begin{bmatrix} S_1 \\ S_2 \\ P_1 \\ P_2 \end{bmatrix} \tag{7.91}$$

with:

$$Y_a = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad Y_b = \begin{bmatrix} -\frac{1}{Y_1} & 0 \\ Y_3 & -\frac{1}{Y_2} \\ Y_4 & Y_5 \\ 0 & Y_6 \end{bmatrix} \tag{7.92}$$

Therefore if C_b is chosen as an identity matrix ($C_b = I_4$), then C_a is equal to:

$$C_a = -Y_b Y_a^{-1} = \begin{bmatrix} \frac{1}{Y_1} & 0 \\ -Y_3 & \frac{1}{Y_2} \\ -Y_4 & -Y_5 \\ 0 & -Y_6 \end{bmatrix} \tag{7.93}$$

The dynamics of ζ are here equal to:

$$\frac{d}{dt} \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \end{bmatrix} = -D \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \end{bmatrix} + \begin{bmatrix} DS_{in} \\ 0 \\ -Q_1 \\ -Q_2 \end{bmatrix} \tag{7.94}$$

Note that since $F_a = Q_a = 0$, the dynamics of ζ are also independent of the yield coefficients. This has been called a *nice partition* in Bastin and Dochain [14].

Case #1: S_2 and P_2 are measured on-line

Then A_1 and A_2 are equal to:

$$A_1 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & \frac{1}{Y_1} & 0 & 0 \\ 0 & -Y_3 & \frac{1}{Y_2} & 0 \\ 0 & -Y_4 & -Y_5 & 1 \\ 0 & 0 & -Y_6 & 0 \end{bmatrix} \tag{7.95}$$

Therefore, S_1, X_1, X_2 and P_1 can be estimated by using the asymptotic observer via the dynamical equation of ζ (7.94) as follows:

$$\hat{S}_1 = \hat{\zeta}_1 + \frac{1}{Y_1 Y_3}(S_2 + \frac{P_2}{Y_6} - \frac{\hat{\zeta}_4}{Y_6} - \hat{\zeta}_2) \tag{7.96}$$

$$\hat{X}_1 = \frac{1}{Y_3}(S_2 + \frac{P_2}{Y_6} - \frac{\hat{\zeta}_4}{Y_6} - \hat{\zeta}_2) \tag{7.97}$$

$$\hat{X}_2 = \frac{1}{Y_6}(P_2 - \hat{\zeta}_4) \tag{7.98}$$

$$\hat{P}_1 = \hat{\zeta}_3 + \frac{Y_4}{Y_3} S_2 - \frac{Y_4}{Y_3}\hat{\zeta}_2 + \frac{Y_4 + Y_3 Y_5}{Y_3 Y_6}(P_2 - \hat{\zeta}_2) \tag{7.99}$$

Case #2: S_1 and X_1 are measured on-line

This choice is interesting because it corresponds to a wrong choice of measured variables, since S_1 and X_1 are not independent. Indeed the matrices A_1 and A_2 are then equal to:

$$A_1 = \begin{bmatrix} \frac{1}{Y_1} & 1 \\ 0 & Y_3 \\ 0 & -Y_4 \\ 0 & 0 \end{bmatrix}, \qquad A_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & \frac{1}{Y_2} & 0 & 0 \\ 0 & -Y_5 & 1 & 0 \\ 0 & -Y_6 & 0 & 1 \end{bmatrix} \tag{7.100}$$

The matrix A_2 is obviously not invertible (the first row is equal to zero!).

This example shows that not any choice of M measured components is valid for the implementation of the asymptotic observer: the submatrix Y_1 must be full rank, i.e. the measured components must be independent or the measured components have to take part in all the reactions (at least one in each reaction) in order to

avoid a submatrix Y_1 with (a) column(s) only filled with zeros.

Case #3: S_2, P_1 and P_2 are measured on-line

Now the number of measured variables is larger than rank(Y). In this case, the vectors ξ_1 and ξ_2 are equal to:

$$\xi_1 = \begin{bmatrix} S_2 \\ P_1 \\ P_2 \end{bmatrix}, \qquad \xi_2 = \begin{bmatrix} S_1 \\ X_1 \\ X_2 \end{bmatrix} \tag{7.101}$$

and the matrices A_1 and A_2 are then defined as follows:

$$A_1 = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad A_2 = \begin{bmatrix} 1 & \frac{1}{Y_1} & 0 \\ 0 & -Y_3 & \frac{1}{Y_2} \\ 0 & -Y_4 & -Y_5 \\ 0 & 0 & -Y_6 \end{bmatrix} \tag{7.102}$$

A_2 is not a square matrix anymore. Then we can choose, as a left inverse, its left pseudo-inverse:

$$A_2^+ = (A_2^T A_2)^{-1} A_2^T \tag{7.103}$$

which takes the following form in our example:

$$A_2^+ = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{7.104}$$

Implementation aspects: choice of the sampling period. The practical computer implementation of the asymptotic observer (7.79)(7.80) or (7.82)(7.83) requires that it be rewritten in a discrete-time form. This can be done simply by replacing the time derivative of ζ by a finite difference (using a first order Euler approximation):

$$\frac{d\hat{\zeta}}{dt} \longrightarrow \frac{\hat{\zeta}_{t+1} - \hat{\zeta}_t}{T} \tag{7.105}$$

where T is the sampling period and t and $t+1$ are time indices. The asymptotic observer is then written as follows for the general case p $\geq$ M:

$$\hat{\zeta}_{t+1} = \hat{\zeta}_t - T D_t \hat{\zeta}_t + T C_a (F_{at} - Q_{at}) + C_b (F_{bt} - Q_{bt}) \tag{7.106}$$

$$\hat{\xi}_{2,t+1} = A_2^+ [\hat{\zeta}_{t+1} - A_1 \xi_{1,t+1}] \tag{7.107}$$

For the discrete-time equation, the value of the sampling period plays a role in the stability. In fact, if the dilution rate D is bounded as follows:

$$0 \leq D(t) \leq D_{max} \tag{7.108}$$

then equation (7.106) will remain stable as long as T is smaller than $2/D_{max}$:

$$T \leq \frac{2}{D_{max}} \tag{7.109}$$

Equation (7.109) gives a condition for the stability of the discrete-time version of the asymptotic observer. But even if the sampling period T is chosen so as to fulfill condition (7.109), large values of T may introduce oscillations in the estimation. As a matter of fact, if we assume that D is constant, the dynamics of ζ are characterised by a (discrete-time) pole equal to (1 - $T\,D$). It will be negative if T is larger than $1/D$ ($T > 1/D$) and then corresponds to oscillating (underdamped) dynamics for ζ. Therefore, in order to avoid (undesirable) oscillations, condition (7.109) can be replaced by:

$$T \leq \frac{1}{D_{max}} \tag{7.110}$$

Remark: note that different sampling periods may be used for the computation of the variables ζ and the calculation of the observation of ξ_2 , e.g. a "fast" computation of the variables ζ and a slower computation of the estimated values of ξ_2. This choice may depend on the measurement sampling interval which may be different from one variable to another.

Application to an anaerobic digestion process. Let us consider the on-line estimation of S_1, X_1 and X_2 from the measurements of methane P_2 and volatile fatty acids S_2. This corresponds to the case #1 hereabove[18].

Note that in practice the methane gas is a low solubility product, and therefore the model reduction approach discussed in Chapter 2 (Section 2.9) applies:

$$P_2 = 0, \;\; \frac{dP_2}{dt} = 0 \tag{7.111}$$

Therefore the asymptotic observer specialises as follows:

$$\frac{d\hat{\zeta}_1}{dt} = -D\hat{\zeta}_1 + DS_{in} \tag{7.112}$$

$$\frac{d\hat{\zeta}_2}{dt} = -D\hat{\zeta}_2 \tag{7.113}$$

[18]We did not consider the on-line estimation of CO_2 (P_1) which is indeed directly obtained either via the gas flow rate and CO_2 percentage measurement, or via the gas flow rate and the CH_4 percentage measurement under the assumption that the outflow gas is composed almost exclusively by CO_2 and CH_4.

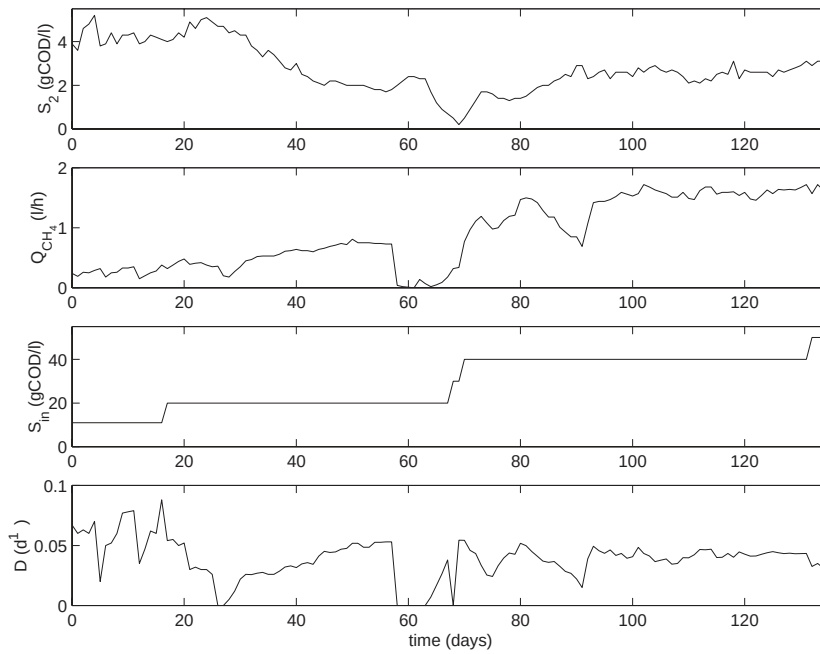

FIG. 7.5. On-line data of the asymptotic observer for the anaerobic digestion process.

$$\frac{d\hat{\zeta}_4}{dt} = -D\hat{\zeta}_4 - Q_2 \tag{7.114}$$

$$\hat{S}_1 = \hat{\zeta}_1 + \frac{1}{Y_1 Y_3}(S_2 - \frac{\hat{\zeta}_4}{Y_6} - \hat{\zeta}_2) \tag{7.115}$$

$$\hat{X}_1 = \frac{1}{Y_3}(S_2 - \frac{\hat{\zeta}_4}{Y_6} - \hat{\zeta}_2) \tag{7.116}$$

$$\hat{X}_2 = -\frac{1}{Y_6}\hat{\zeta}_4 \tag{7.117}$$

The asymptotic observer has been implemented on a 60 liter pilot CSTR of the Unit of Bioengineering, Université Catholique de Louvain, Belgium ([14]). The values of the yield coefficients are equal to:

$$Y_1 = 0.3125, \quad Y_2 = 0.035, \quad Y_3 = 5.7, \quad Y_6 = 27.3 \tag{7.118}$$

Figure 7.5 shows the data used for the on-line estimation (volatile fatty acids S_2, methane gas flow rate Q_2 denoted Q_{CH_4} in the figure), influent organic matter

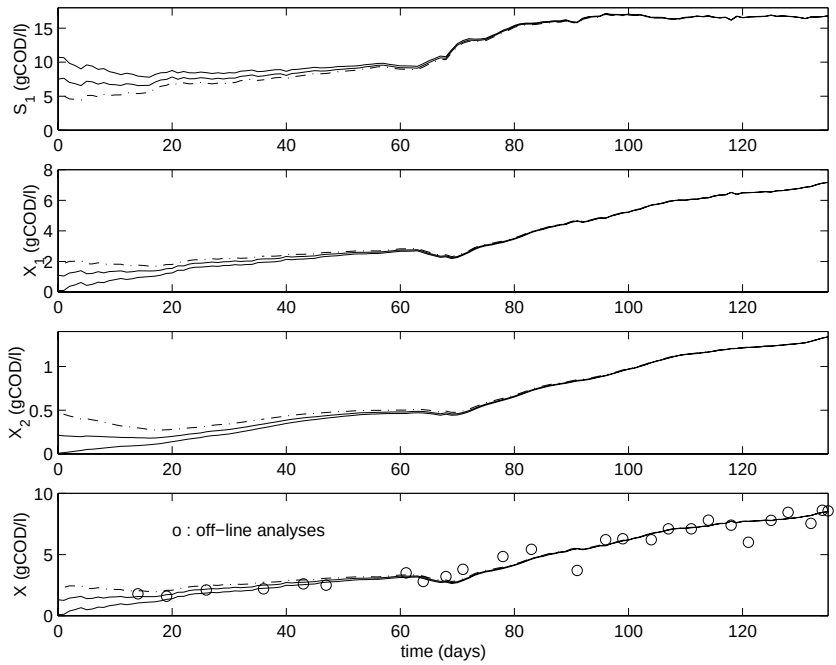

FIG. 7.6. Estimation results of the asymptotic observer of the anaerobic digestion process (-: $\hat{\zeta}_1(0) = 11$, $\hat{\zeta}_2(0) = 4$, $\hat{\zeta}_4(0) = -6$; - -: $\hat{\zeta}_1(0) = 11$, $\hat{\zeta}_2(0) = 3.9$, $\hat{\zeta}_4(0) = 0$; - . -: $\hat{\zeta}_1(0) = 11$, $\hat{\zeta}_2(0) = 6.65$, $\hat{\zeta}_4(0) = -13.65$).

concentration S_{in}, and dilution rate D. The estimation results are given in Figure 7.6. Different initial conditions for the auxiliary variables ζ have been considered:

$$
\begin{aligned}
- &: \hat{\zeta}_1(0) = 11,\ \hat{\zeta}_2(0) = 4,\ \hat{\zeta}_4(0) = -6 \\
- - &: \hat{\zeta}_1(0) = 11,\ \hat{\zeta}_2(0) = 3.9,\ \hat{\zeta}_4(0) = 0 \\
-.- &: \hat{\zeta}_1(0) = 11,\ \hat{\zeta}_2(0) = 6.65,\ \hat{\zeta}_4(0) = -13.65
\end{aligned}
$$

Note that the asymptotic observer converges after an initial transient to the same estimation profile. The convergence time is approximately equal to 60 days, which corresponds to three times the residence time (or equivalently the inverse of the dilution rate, see Figure 7.5).

One of the major difficulties that we had to face here was the validation of the on-line estimates provided by the asymptotic observer, since no measurement, even indirect, of the different (acidogenic and methanogenic) bacterial populations is available. So we have been using COD data for the validation, more precisely the difference between the total COD and the soluble COD, which are known to represent fairly well the active biomass population in the process, i.e. X_1 +

X_2 ([14]). The last figure in Figure 7.6 compares these data with the sum of the estimates $\hat{X}_1 + \hat{X}_2$ for the three different initial conditions.

7.4.2 *Asymptotic Observers for Multi-Tank Reactors*

The design of the above asymptotic observer can be easily extended to multi-reactor processes, since the definition of the transformation ζ is the same. The main question is the stability of the dynamics of the auxiliary variables ζ (see Chapter 2, Section 2.8):

$$\frac{d\zeta}{dt} = -(C_b D_{bb} + C_a D_{ab}) C_b^{-1} \zeta + C_a (F_a - Q_a) + C_b (F_b - Q_b) + [(C_b D_{bb} + C_a D_{ab}) C_b^{-1} - C_b D_{ba} - C_a D_{aa}] \xi_a \tag{7.119}$$

with:

$$D = \begin{bmatrix} D_{aa} & D_{ab} \\ D_{ba} & D_{bb} \end{bmatrix} \tag{7.120}$$

The stability of (7.119) is determined by the matrix:

$$-(C_b D_{bb} + C_a D_{ab}) C_b^{-1} \tag{7.121}$$

We shall not develop this point in detail here (see e.g. [56]). Rather we shall concentrate on one example (the basic model of the activated sludge process), and discuss its stability.

Let us first recall the equation of the basic model of the activated sludge process, written as follows in the General Dynamical Model format:

$$\xi = \begin{bmatrix} S \\ S_O \\ X \\ X_R \end{bmatrix}, \; Y = \begin{bmatrix} -\frac{1}{Y_S} \\ -\frac{1}{Y_O} \\ 1 \\ 0 \end{bmatrix}, \; F = \begin{bmatrix} D_{in} S_{in} \\ D_{in} S_{O,in} + k_L a (S_O^* - S_O) \\ 0 \\ 0 \end{bmatrix} \tag{7.122}$$

$$\rho = \mu X, \; Q = 0, \; D = \begin{bmatrix} D_1 & 0 & 0 & 0 \\ 0 & D_1 & 0 & 0 \\ 0 & 0 & D_1 & -D_2 \\ 0 & 0 & -D_3 & D_4 \end{bmatrix} \tag{7.123}$$

with the following definitions for D_{in}, D_1, D_2, D_3, and D_4:

$$D_{in} = \frac{F_{in}}{V}, \; D_2 = \frac{F_R}{V}, \; D_1 = D_{in} + D_2, \; D_3 = \frac{F_{in} + F_R}{V_S}, \; D_4 = \frac{F_R + F_W}{V_S} \tag{7.124}$$

Let us consider, as in Section 2.8.3, the following state partition:

$$\xi_a = X,\ \xi_b = \begin{bmatrix} S \\ S_O \\ X_R \end{bmatrix} \tag{7.125}$$

This means that the different dilution matrices are equal to:

$$D_{aa} = D_1,\ D_{ab} = \begin{bmatrix} 0 & 0 & -D_2 \end{bmatrix} \tag{7.126}$$

$$D_{ba} = \begin{bmatrix} 0 \\ 0 \\ -D_3 \end{bmatrix},\ D_{bb} = \begin{bmatrix} D_1 & 0 & 0 \\ 0 & D_1 & 0 \\ 0 & 0 & D_4 \end{bmatrix} \tag{7.127}$$

If we select $C_b = I$, the matrix C_a is equal to:

$$C_a = \begin{bmatrix} -\frac{1}{Y_S} \\ -\frac{1}{Y_O} \\ 0 \end{bmatrix} \tag{7.128}$$

and the auxiliary variable ζ is written as follows:

$$\zeta = \begin{bmatrix} S + \frac{1}{Y_S} X \\ S_O + \frac{1}{Y_O} X \\ X_R \end{bmatrix} \tag{7.129}$$

The stability of the dynamics of the auxiliary ζ (and the underlying asymptotic observer) is determined by the matrix (7.121) which specialises here as follows:

$$\begin{bmatrix} -D_1 & 0 & -\frac{D_2}{Y_S} \\ 0 & -D_1 & -\frac{D_2}{Y_O} \\ 0 & 0 & -D_4 \end{bmatrix} \tag{7.130}$$

Its eigenvalues are negative if D_1 and D_4 are positive, i.e. if

$$F_{in} + F_R > 0 \text{ and } F_R + F_W > 0 \tag{7.131}$$

which holds for instance if the sludge is being recycled ($F_R > 0$).

Let us now consider the problem of estimating the concentrations of organic matter S, and of the biomass X and X_R both in the aerator and in the settler, from on-line data of dissolved oxygen S_O. This means that ξ_1 and ξ_2 are defined as follows:

$$\xi_1 = S_O,\ \xi_2 = \begin{bmatrix} S \\ X \\ X_R \end{bmatrix} \tag{7.132}$$

The asymptotic observer is then written as follows:

$$\frac{d\hat{\zeta}_1}{dt} = -\frac{F_{in} + F_R}{V}\hat{\zeta}_1 + \frac{F_{in}}{V} S_{in} + \frac{F_R}{V Y_S}\zeta_3 \tag{7.133}$$

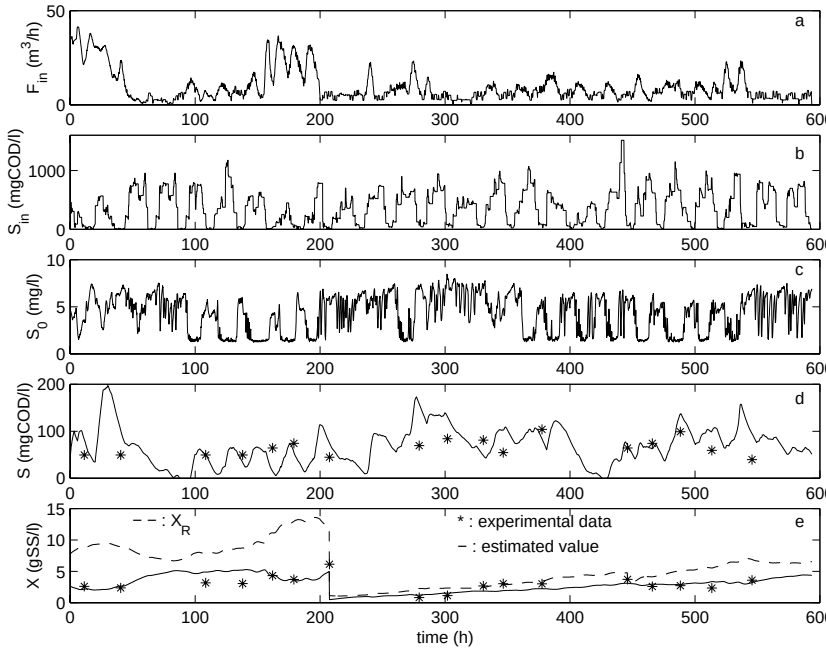

FIG. 7.7. Estimation results of the asymptotic observer of an activated sludge process.

$$\frac{d\hat{\zeta}_2}{dt} = -\frac{F_{in}+F_R}{V}\hat{\zeta}_2 + k_L a(S_O^* - S_O) + \frac{F_R}{VY_O}\hat{\zeta}_3 \tag{7.134}$$

$$\frac{d\hat{\zeta}_3}{dt} = -\frac{F_W+F_R}{V_S}\hat{\zeta}_3 + \frac{F_{in}+F_R}{V_S}(\hat{\zeta}_2 - S_O)Y_O \tag{7.135}$$

$$\hat{S} = (\hat{\zeta}_1 - Y_O\hat{\zeta}_2 + Y_O S_O)Y_S \tag{7.136}$$

$$\hat{X} = (\hat{\zeta}_2 - S_O)Y_O \tag{7.137}$$

$$\hat{X}_R = \hat{\zeta}_3 \tag{7.138}$$

The implementation of the asymptotic observer further requires on-line data of the influent flow rate F_{in}, of the influent organic matter concentration S_{in}, and of the recycle and waste flow rates F_R and F_W. And besides the values of the yield coefficients Y_S and Y_O, we also need the values of the mass transfer coefficient $k_L a$ and of the saturation concentration S_O^*.

Figure 7.7 shows the application of the asymptotic observer to a wastewater treatment plant (the Maria Middelares treatment plant, Gent, Belgium)(see also [76]). The available data were the on-line data of dissolved oxygen and influent COD (obtained from an on-line respirometer), some off-line measurements of the

COD in the effluent and of the suspended solids in the aerator (yet no precise measurement of the recycle flow rate and only time points (t = 208 h and 449 h) when discontinuous sludge wastage was carried out). The state estimation is a challenging question if we consider the measurement constraints, and the oversimplification of the model dynamics. Figure 7.7 shows the estimation results. Figures a, b and c show the data of the influent flow rate F_{in}, the influent substrate concentration S_{in} and the dissolved oxygen S_O, respectively, while Figures d and e compare the estimation results given by the asymptotic observer with the off-line data for S and X, respectively. The computed value of X_R is also shown (in dotted lines) in Figure d. The recycle flow rate F_R was kept constant at the design value of 8 m^3/h. The volumes are: V = 380 m^3, V_S = 265 m^3; the determination of S_O^* and $k_L a$ gave the following result:

$$S_O^* = 10\ mg/l,\ k_L a = 1.7\ h^{-1} \tag{7.139}$$

We note that although the reaction scheme is very simple and that some of the data are not very precise, the reaction scheme is fairly well validated.

7.4.3 *Asymptotic Observers for Tubular Bioreactors*

Let us now discuss the extension and application of the asymptotic observer design to non completely mixed reactors.

Let us modify the first assumption introduced in Section 7.4.1:

1bis. p (= M) components are measured on-line *along the reactor*.

and introduce a sixth one:

6. The axial mass dispersion coefficient D_{am} and the reactor section A are known.

Recall that we had considered in Section 2.8.2 (Chapter 2) $C_b = I$ and that the vector of the fixed component ξ_{fi} has been put in ξ_b:

$$\xi_b = \begin{bmatrix} \xi_{bf} \\ \xi_{fi} \end{bmatrix} \tag{7.140}$$

Then we can rewrite the auxiliary variable ζ as follows:

$$\zeta = \begin{bmatrix} \zeta_{fl} \\ \zeta_{fi} \end{bmatrix} = \begin{bmatrix} \xi_{bf} \\ \xi_{fi} \end{bmatrix} + \begin{bmatrix} C_{af} \\ C_{ae} \end{bmatrix} \xi_a \tag{7.141}$$

The dynamics of ζ can then be written as follows:

$$\frac{\partial \zeta_{fl}}{\partial t} = -\frac{F_{in}}{A}\frac{\partial \zeta_{fl}}{\partial z} + D_{am}\frac{\partial^2 \zeta_{fl}}{\partial z^2} \tag{7.142}$$

$$\frac{\partial \zeta_{fi}}{\partial t} = -\frac{F_{in}}{A} C_{ae} \frac{\partial \xi_a}{\partial z} + D_{am} C_{ae} \frac{\partial^2 \xi_a}{\partial z^2} \tag{7.143}$$

Assume that the measured components are put in the vector ξ_a (i.e. we consider a particular choice for ξ_1 and ξ_2 ($\xi_1 = \xi_a, \xi_2 = \xi_b$); this is done only to simplify the approach but, of course, other choices are possible). Then by using the same arguments as above and the equations (7.141), (7.142) and (7.143), we obtain the following asymptotic observer for tubular bioreactors:

$$\frac{\partial \hat{\zeta}_{fl}}{\partial t} = -\frac{F_{in}}{A} \frac{\partial \hat{\zeta}_{fl}}{\partial z} + D_{am} \frac{\partial^2 \hat{\zeta}_{fl}}{\partial z^2} \tag{7.144}$$

$$\frac{\partial \hat{\zeta}_{fi}}{\partial t} = -\frac{F_{in}}{A} C_{ae} \frac{\partial \xi_a}{\partial z} + D_{am} C_{ae} \frac{\partial^2 \xi_a}{\partial z^2} \tag{7.145}$$

$$\hat{\xi}_2 = \hat{\zeta} - \begin{bmatrix} C_{af} \\ C_{ae} \end{bmatrix} \xi_1 \tag{7.146}$$

There remain two key questions with the above asymptotic observer:

1. The above observer is written under the form of ("infinite dimensional") partial derivative equations (PDEs). These are not very easy or convenient to handle in practical control and monitoring applications. Moreover, it is assumed that p components are available for on-line measurement *along the reactor*. In line with a number of works on the subject (e.g. [97], [136]), we propose to "reduce" the above PDE equations to a finite number of ordinary differential equations (ODEs) at a finite number of positions along the reactor and to consider the reduced equations for the practical application of the observer. This will the object of the next section.

2. The second key question is to know whether the proposed observer is reliable, i.e. under which the conditions it will give estimates that converge to their true values: this will be addressed on page 284.

Practical implementation of the asymptotic observer. Let us reduce the PDEs of the asymptotic observer (7.144) to a finite number of ODEs. We shall not discuss the choice of one reduction method here (see e.g. [277], [211] for this topic); we shall only assume that the user has chosen one method for approximating the PDEs (e.g. finite differences, orthogonal collocation,...) and that the reduced model is a fairly good representation of the PDE asymptotic observer (7.144). Whatever the reduction method, the partial derivatives of the variables ζ_i with respect to the space variable z are approximated by a weighted sum of ζ_i at a finite number of positions z_j (j = 0 to q, where 0 and q hold for the input and output of the reactor, respectively) along the reactor:

$$\frac{\partial^k}{\partial z^k}\begin{bmatrix}\zeta_i(z=z_1)\\ \zeta_i(z=z_2)\\ \vdots\\ \zeta_i(z=z_q)\end{bmatrix} \cong [\tilde{c}_k \mid \tilde{C}_k]\begin{bmatrix}\zeta_i(z=z_0)\\ \zeta_i(z=z_1)\\ \vdots\\ \zeta_i(z=z_q)\end{bmatrix},\ k=1,2 \qquad (7.147)$$

$$dim(\tilde{c}_k) = q \times 1, dim(\tilde{C}_k) = q \times q \qquad (7.148)$$

If we consider a finite difference approximation, the matrices $\tilde{C}_k$ (k = 1, 2) are equal to:

$$\tilde{C}_1 = \frac{1}{\Delta z}\begin{bmatrix}1 & 0 & 0 & \cdots & 0 & 0\\ -1 & 1 & 0 & \cdots & \vdots & \vdots\\ 0 & -1 & 1 & \cdots & \vdots & \vdots\\ 0 & 0 & -1 & \cdots & \vdots & \vdots\\ \vdots & \vdots & \vdots & \cdots & 1 & 0\\ 0 & 0 & 0 & \cdots & -1 & 1\end{bmatrix},\ \tilde{c}_1 = \frac{1}{\Delta z}\begin{bmatrix}1\\ 0\\ \vdots\\ 0\end{bmatrix} \qquad (7.149)$$

$$\tilde{C}_2 = \frac{1}{(\Delta z)^2}\begin{bmatrix}-2 & 1 & 0 & \cdots & 0 & 0\\ 1 & -2 & 1 & \cdots & \vdots & \vdots\\ 0 & 1 & -2 & \cdots & \vdots & \vdots\\ 0 & 0 & 1 & \cdots & \vdots & \vdots\\ \vdots & \vdots & \vdots & \cdots & -2 & 1\\ 0 & 0 & 0 & \cdots & 1 & -2\end{bmatrix},\ \tilde{c}_2 = \frac{1}{(\Delta z)^2}\begin{bmatrix}1\\ 0\\ \vdots\\ 0\end{bmatrix} \qquad (7.150)$$

where Δz is the spatial discretisation step.

The reduction procedure results in a rewriting of the auxiliary variables ζ_f and of the asymptotic observer equations under ODEs at each position z_j (j = 1 to q). Then the dynamics of the auxiliary variables ζ_f and of the asymptotic observer become:

Dynamics of ζ_f:

$$\frac{d\zeta_{rl}}{dt} = [D_{am}C_2 - \frac{F}{A}C_1]\zeta_{rl} + [D_{am}c_2 - \frac{F}{A}c_1](\xi_{2,in} + C_{af}\xi_{1,in}) \qquad (7.151)$$

$$\frac{d\zeta_{ri}}{dt} = [D_{am}C_2 - \frac{F}{A}C_1]C_{ae}\xi_a + [D_{am}c_2 - \frac{F}{A}c_1]C_{ae}\xi_{1,in} \qquad (7.152)$$

Asymptotic Observer:

$$\frac{d\hat{\zeta}_{rl}}{dt} = [D_{am}C_2 - \frac{F}{A}C_1]\hat{\zeta}_{rl} + [D_{am}c_2 - \frac{F}{A}c_1](\xi_{2,in} + C_{af}\xi_{1,in}) \quad (7.153)$$

$$\frac{d\hat{\zeta}_{ri}}{dt} = [D_{am}C_2 - \frac{F}{A}C_1]C_{ae}\xi_a + [D_{am}c_2 - \frac{F}{A}c_1]C_{ae}\xi_{1,in} \quad (7.154)$$

with:

$$\zeta_r = \begin{bmatrix} \zeta_{rl} \\ \zeta_{ri} \end{bmatrix} = \begin{bmatrix} \zeta_{r,1}(z = z_1) \\ \vdots \\ \zeta_{r,1}(z = z_q) \\ \vdots \\ \zeta_{r,R}(z = z_1) \\ \vdots \\ \zeta_{r,R}(z = z_q) \end{bmatrix} \quad (7.155)$$

$$C_k = \begin{bmatrix} \tilde{C}_k & 0 & \cdots & 0 \\ 0 & \tilde{C}_k & \cdots & \vdots \\ \vdots & \vdots & \cdots & 0 \\ 0 & 0 & \cdots & \tilde{C}_k \end{bmatrix}, \; c_k = \begin{bmatrix} \tilde{c}_k \\ \tilde{c}_k \\ \vdots \\ \tilde{c}_k \end{bmatrix}, \; k = 1, 2 \quad (7.156)$$

One important feature of the above equations (7.151)(7.153) is that the tubular reactor is approximated by a stirred multi-tank reactor. In case of a finite difference approximation, the equations represent a cascade of stirred tank reactors; but with other approximation methods (e.g. collocation methods), the model exhibits interconnections between each of the stirred tank reactors of the ODE model since the entries of the matrices $\tilde{C}_k$ (k = 1, 2) are generally different from zero [155].

Stability properties of the asymptotic observer. Let us start by analysing the stability of the PDE asymptotic observer.

If we define the estimation error e on the auxiliary variables ζ as follows:

$$e = \zeta - \hat{\zeta} = \begin{bmatrix} \zeta_{fl} - \hat{\zeta}_{fl} \\ \zeta_{fi} - \hat{\zeta}_{fi} \end{bmatrix} = \begin{bmatrix} e_{fl} \\ e_{fi} \end{bmatrix} \quad (7.157)$$

the dynamics of e are given by the following equations:

$$\frac{\partial e_{fl}}{\partial t} = -\frac{F_{in}}{A}\frac{\partial e_{fl}}{\partial z} + D_{am}\frac{\partial^2 e_{fl}}{\partial z^2}, \quad e_{fl}(t = 0, z) = e_{fl0}(z) \quad (7.158)$$

$$\frac{\partial e_{fi}}{\partial t} = 0, \qquad e_{fi}(t=0,z) = e_{fi0}(z) \quad (7.159)$$

The analysis of the second equation (7.159) shows that under the assumption presented above, any initial bias will remain (this is exactly the same as for the batch reactor in stirred tank reactors).

Let us now analyse the first error equation (7.158). First note that it is a linear equation. If D_{am} is different from zero (parabolic equation), it is characterised (see e.g. [285]) by the following simple, real, negative eigenvalues λ_n, $n \geq 1$:

$$\lambda_n = -\frac{s_n^2 + u^2}{4D_{am}} < -\frac{u^2}{4D_{am}} < 0, \quad \text{for all } n \geq 1\ , \qquad (7.160)$$

where $\{s_n : n \geq 1\}$ is the set of all the solutions to the equation (called the resolvent equation):

$$\tan(\frac{L}{2D_{am}}s) = \frac{2us}{s^2 - u^2}, \quad s > 0\ , \qquad (7.161)$$

with $u = F_{in}/A$ and L the length of the reactor.

If D_{am} is equal to zero (hyperbolic equation, plug flow reactor), the solution of the equation (7.158) is of the following form:

$$\begin{aligned} e_{fl}(t,z) &= e_{fl0}(z-ut), \quad \text{if } t < z/u \\ &= 0 \qquad\qquad\quad \text{if } t \geq z/u \end{aligned}$$

Therefore the equation (7.158) is asymptotically stable, and the asymptotic observer for the components ξ_f is convergent.

Let us concentrate on the stability properties of the reduced form (7.153) of the asymptotic observer, to see in which conditions the above stability result is correctly transferred to the discretised version of the asymptotic observer. For obvious reasons, we shall only concentrate on the error dynamics (7.158). If we define the observation error:

$$e = \zeta_r - \hat{\zeta}_r \qquad (7.162)$$

then the dynamics of the observation error e is readily obtained from (7.151)(7.153):

$$\frac{de}{dt} = [D_{am}C_2 - \frac{F}{A}]e \qquad (7.163)$$

Therefore the stability depends on the state matrix:

$$D_{ma}C_2 - \frac{F}{A}C_1 \qquad (7.164)$$

Because of the diagonal structure of the matrices C_1 and C_2, the stability of the above state matrix (7.164) depends on the stability of each submatrix:

$$D_{ma}\tilde{C}_2 - \frac{F}{A}\tilde{C}_1 \tag{7.165}$$

Therefore it follows that the asymptotic observer (7.153) will be asymptotically stable, i.e.:

$$\lim_{t\to\infty} \hat{\xi}_r = \xi_r \tag{7.166}$$

if the eigenvalues of the matrix $D_{ma}\tilde{C}_2 - \frac{F}{A}\tilde{C}_1$ are stable. Note that the stability of the asymptotic observer only depends on the axial mass transfer and not on the kinetics. In other words, the reactor may be unstable (due to kinetics like the Haldane kinetics, see Section 2.2.6) while the asymptotic observer is asymptotically stable (because of stable hydrodynamics).

Note also that it is routine to check the stability of the matrix (7.165) by using any matrix computation program. An interesting particular case is the finite difference approximation of a fixed bed reactor without dispersion (D_{am} = 0). Indeed, the stability then simply depends on the matrix $\frac{F}{A}C_1$, and it is straightforward to check that the matrix $\frac{F}{A}C_1$ (see equation (7.149)) is stable as long as F is positive. The stability properties of the hydrodynamics approximated via orthogonal collocation are analysed in Lefévre *et al*. [155].

Example. Let us consider a plug flow reactor with fixed biomass and with the following reaction scheme:

$$S \longrightarrow X + P \tag{7.167}$$

$$X \longrightarrow X_d \tag{7.168}$$

with X_d the dead biomass. The dynamics of the process are given by the following equations:

$$\frac{\partial S}{\partial t} = -\frac{F_{in}}{A}\frac{\partial S}{\partial z} - \frac{1}{Y_1}\mu X \tag{7.169}$$

$$\frac{\partial P}{\partial t} = -\frac{F_{in}}{A}\frac{\partial P}{\partial z} + Y_2\mu X \tag{7.170}$$

$$\frac{\partial X}{\partial t} = \mu X - k_d X \tag{7.171}$$

$$\frac{\partial X_d}{\partial t} = -\frac{F_{in}}{A}\frac{\partial X_d}{\partial z} + k_d X \tag{7.172}$$

Assume first that S and X_d are available for on-line measurement (while X and P are not). Assume also that the specific growth rate μ and the death/detachment coefficient k_d are unknown. Then the asymptotic observer can be derived as follows. The auxiliary variables ζ are derived from equation (7.141) and are, for instance, equal to:

$$\zeta_1 = X + X_d + Y_1 S \tag{7.173}$$

$$\zeta_2 = P + Y_1 Y_2 S \tag{7.174}$$

The asymptotic observer equations are then written as follows:

$$\frac{\partial \hat{\zeta}_1}{\partial t} = -\frac{F_{in}}{A}\frac{\partial (X_d + Y_1 S)}{\partial z} \tag{7.175}$$

$$\frac{\partial \hat{\zeta}_2}{\partial t} = -\frac{F_{in}}{A}\frac{\partial \hat{\zeta}_2}{\partial z} \tag{7.176}$$

$$\hat{X} = \hat{\zeta}_1 - X_d - Y_1 S \tag{7.177}$$

$$\hat{P} = \hat{\zeta}_2 - Y_1 Y_2 S \tag{7.178}$$

Another option is when the death coefficient k_d is known. Then only one measurement is necessary, e.g. the substrate concentration S. There are now three entries for the auxiliary variable ζ, e.g.:

$$\zeta_1 = X + Y_1 S \tag{7.179}$$

$$\zeta_2 = P + Y_1 Y_2 S \tag{7.180}$$

$$\zeta_3 = X_d \tag{7.181}$$

Then the asymptotic observer specialises as follows:

$$\frac{\partial \hat{\zeta}_1}{\partial t} = -\frac{F_{in}}{A} Y_1 \frac{\partial S}{\partial z} - k_d(\hat{\zeta}_1 - Y_1 S) \tag{7.182}$$

$$\frac{\partial \hat{\zeta}_2}{\partial t} = -\frac{F_{in}}{A}\frac{\partial \hat{\zeta}_2}{\partial z} \tag{7.183}$$

$$\frac{\partial \hat{\zeta}_3}{\partial t} = -\frac{F_{in}}{A}\frac{\partial \hat{\zeta}_3}{\partial z} + k_d(\hat{\zeta}_1 - Y_1 S) \tag{7.184}$$

$$\hat{X} = \hat{\zeta}_1 - X_d - Y_1 S \tag{7.185}$$

$$\hat{P} = \hat{\zeta}_2 - Y_1 Y_2 S \tag{7.186}$$

$$\hat{X}_d = \hat{\zeta}_3 \tag{7.187}$$

7.4.4 *Asymptotic Observers as a Tool for Model Selection*

Here above we have seen that asymptotic observers are independent of the process kinetics. In this section an approach is suggested on how to use asymptotic observers for model selection, more precisely to model and validate reaction schemes independently of the reaction kinetics [76]. We shall first introduce the modelling concept and procedure based on the asymptotic observer. The results will be illustrated with one process: a detoxification reactor.

Model selection procedure. The basic idea of the approach is to use the asymptotic observer not for on-line estimation but for model building, more precisely for

modelling the reaction network part of the dynamics independently of the kinetics. It is based on the convergence properties of the observer: because of these, we know that the asymptotic observer is able to give a reliable replica of the dynamics of the process components if the reaction scheme is correct (i.e. representative of the key process reactions) and only if the values of the yield coefficients are known (and not the reaction kinetics).

Assume that a set of experimental data of the concentrations and (when appropriate) gaseous outflow rates of the key process components (whose number is equal to N) as well as of the hydrodynamics (mainly, the flow rates) are available. The model selection procedure is the following.

1. Select a plausible reaction network involving the measured components.
2. Build the dynamical model in the format of the General Dynamical Model.
3. Select M (= number of a priori selected reactions) components and build the asymptotic observer to estimate the $N - M$ remaining components.
4. Try to fit the estimated values of the $N-M$ components to their experimental values by selecting appropriate values of the yield coefficients.
5. The final step is usual: either the fitting is satisfactory and you stop, or it is not and you go back to the first step (select another reaction scheme candidate, possibly with another set of process components).

Illustrative case study: a detoxification process. Soil decontamination has become an important matter in wastewater treatment in recent years, resulting for instance in the increasing use and development of biological detoxification processes [125]. For instance, polychlorinated aliphatic compounds [175], [33], recalcitrant compounds, and halogenated aromatic compounds [40] extracted from the contaminated soil may be treated by microbial mixed populations in strict anaerobic conditions. In this context, a packed bed reactor has been developed and successfully applied to the anaerobic destruction of a mixture of toxic and recalcitrant molecules (chlorinated aliphatics) on a laboratory scale [175], [33].

The anaerobic detoxification is a complex biochemical network involving different microorganism populations and biochemical reactions. A four-reaction model is often considered to describe it [178] (see Figure 7.8). Here again the constraints on the available data for modelling are quite hard: only off-line measurements of hydrogen, acetate, toxics and non-toxics at the reactor output, and of the co-substrate in the influent are available. A plausible reaction network is then:

$$S \longrightarrow H_2 + VA_2 \tag{7.188}$$

$$H_2 + T \longrightarrow N_T \tag{7.189}$$

where S, H_2, VA_2, T and N_T represent the co-substrate (ammonium citrate), the hydrogen, the acetate, the toxics and the non-toxics, respectively. Different sets of data corresponding to different experiments are available. Because of the high

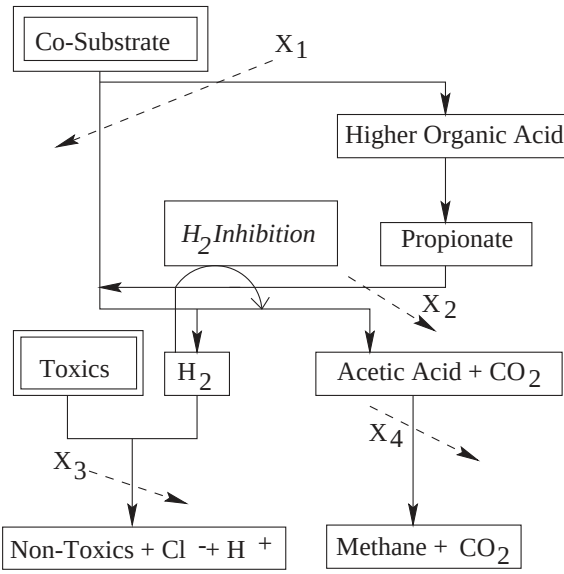

FIG. 7.8. Scheme of the four-reaction model of the anaerobic detoxification process.

recirculation rate (usually 18 d^{-1}), the reactor is run under rather homogeneous conditions, and the transport time delay between the reactor input and its output is negligible with regard to the measurement frequency (the lowest measurement frequency is 3-4 days for the acetate). This validates the use of a stirred tank reactor model. Therefore the dynamics of the process in the General Dynamical Model formalism are as follows:

$$\xi = \begin{bmatrix} S \\ VA_2 \\ H_2 \\ T \\ N_T \end{bmatrix}, \; Y = \begin{bmatrix} -1 & 0 \\ Y_1 & 0 \\ Y_2 & -\frac{1}{Y_3} \\ 0 & -\frac{1}{Y_4} \\ 0 & 1 \end{bmatrix}, \; F = \begin{bmatrix} DS_{in} \\ 0 \\ 0 \\ DT_{in} \\ DN_{in} \end{bmatrix}, \; \rho = \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix}, \; Q = O \tag{7.190}$$

Let us consider the state partition:

$$\xi_1 = \begin{bmatrix} S \\ H_2 \end{bmatrix}, \; \xi_2 = \begin{bmatrix} VA_2 \\ T \\ N_T \end{bmatrix} \tag{7.191}$$

The asymptotic observer equations follow from the above model:

$$\frac{d\hat{\xi}_1}{dt} = -D\hat{\xi}_1 - k_1 DS_{in} \tag{7.192}$$

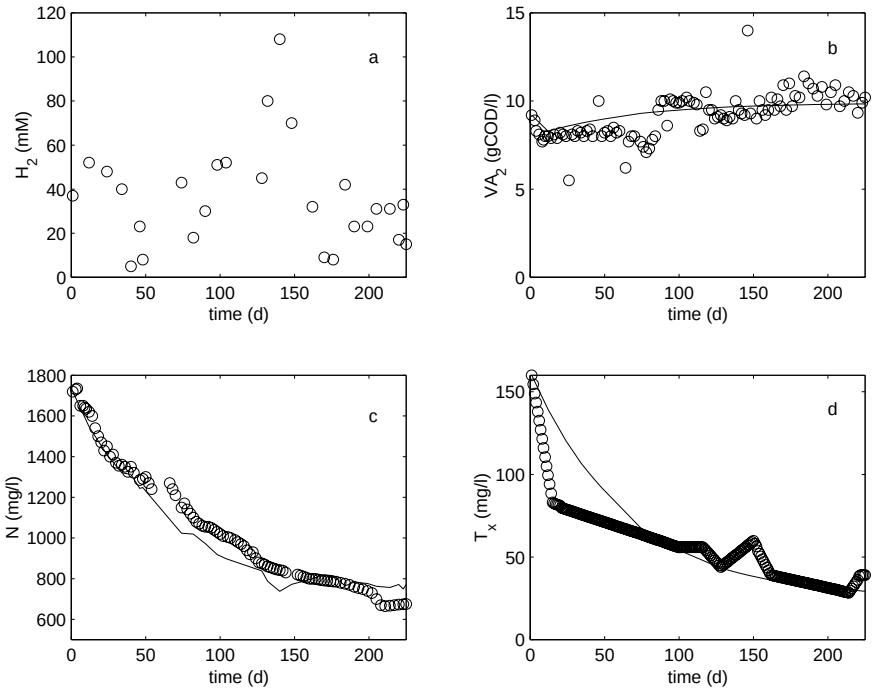

FIG. 7.9. Experimental data and estimation results of the detoxification process (Bioengineering unit, Louvain-la-Neuve, Belgium) (o: experimental data, -: asymptotic observer results).

$$\frac{d\hat{\zeta}_2}{dt} = -D\hat{\zeta}_2 - \frac{Y_2Y_3}{Y_4}DS_{in} + DT_{in} \tag{7.193}$$

$$\frac{d\hat{\zeta}_3}{dt} = -D\hat{\zeta}_3 + Y_2Y_3DS_{in} + DBN_{in} \tag{7.194}$$

$$\hat{VA}_2 = \hat{\zeta}_1 + Y_1S \tag{7.195}$$

$$\hat{T} = \hat{\zeta}_2 + \frac{Y_2Y_3}{Y_4}S + \frac{Y_3}{Y_4}H_2 \tag{7.196}$$

$$\hat{N}_T = \hat{\zeta}_3 - Y_2Y_3S - Y_3H_2 \tag{7.197}$$

The asymptotic observer-based modelling procedure has been applied on three sets of data with similar results. Since the co-substrate is not measured at the reactor output and COD estimation indicates its complete utilisation, its value has been assumed in rough approximation to be negligible. Figure 7.9 shows one set of results for which D = 0.014 d^{-1}, S_{in} = 0.3 gCOD/l, N_{in} = 457 mg/l, T_{in} = 25 mg/l. Figure 7.9.a gives the data of H_2 and Figures 7.9.b, c and d compare the experimental data of VA_2, T and N with their estimates provided by the above

asymptotic observer for the following values of the yield coefficients: $Y_1 = 33$, $Y_2 = 800$, $Y_3 = 1.25$, $Y_4 = 5000$. Note the fairly good validation of the reaction scheme.

7.5 Observers for Processes with Badly Known Kinetics

So far, we have considered (classical) observers when the kinetics are perfectly known, and asymptotic observers when the kinetics are assumed to be unknown. We shall now introduce an intermediate class of observers, i.e. observers of processes for which the structure ("models") of the kinetics are known but with badly or unknown parameter values [77].

7.5.1 *State Observer with the Unknown Parameters as Design Parameters*

The first approach that we propose in this section is based on the following idea: why not use the badly known kinetic parameters as extra design observer parameters in order to guarantee (at least) zero steady state observation errors for the unmeasured variables?

Let us denote the badly known parameters by θ. The process dynamics (7.25) can be rewritten as follows:

$$\frac{dx}{dt} = f(x, u, \theta) \tag{7.198}$$

If we consider that the output vector consists of state variables (as it is often the case in (bio)processes), we can define a state partition with the measured variables ($y = x_1$) and the unmeasured variables, x_2:

$$\frac{dx_1}{dt} = f_1(x, u, \theta) \tag{7.199}$$

$$\frac{dx_2}{dt} = f_2(x, u, \theta) \tag{7.200}$$

The observer design remains basically the same, but now we choose θ such that $x_2 - \hat{x}_2$ is equal to zero in steady state, i.e.:

$$\theta : (x_2(\theta) - \hat{x}_2(\theta))_{ss} = 0 \tag{7.201}$$

Let us illustrate this idea on the simple microbial growth example (Section 7.3.4). The objective is to select one of the kinetic parameters used in the observer such that the estimation error is equal to zero in steady state. Let us first consider that the badly known parameter is the saturation constant K_S (this means that we use $\tilde{K}_S$ and μ_{max} in the observer equations). Then the dynamics of the observation errors is here equal to:

$$\frac{d}{dt}\begin{bmatrix} e_S \\ e_X \end{bmatrix} = \begin{bmatrix} -D - \frac{1}{Y_1}\tilde{\alpha}X - k_1 & -\frac{1}{Y_1}\tilde{\alpha}\hat{S} \\ \tilde{\alpha} - k_2 & \tilde{\alpha}\hat{S} - D \end{bmatrix}\begin{bmatrix} e_S \\ e_X \end{bmatrix}$$

$$+\begin{bmatrix} -\frac{1}{Y_1}SX \\ SX \end{bmatrix} e_\alpha \tag{7.202}$$

with:

$$\tilde{\alpha} = \frac{\mu_{max}}{\tilde{K_S} + \hat{S}},\ e_\alpha = \mu_{max}[\frac{1}{K_S + S} - \frac{1}{\tilde{K_S} + \hat{S}}] \tag{7.203}$$

e_S and e_X are the estimation errors on S and X. As a matter of illustration, if we consider the ELO, the gains k_1 and k_2 would be chosen from equations (7.72)(7.73). Let us set de_S/dt, de_X/dt to zero. Then e_X in steady state, $\bar{e}_X$, is given by the following expression:

$$\bar{e}_X = \frac{(\frac{1}{Y}k_2 + k_1 + D)SXe_\alpha}{(D - \tilde{\alpha}\hat{S})(D + k_1 + \frac{1}{Y_1}\tilde{\alpha}X) + \frac{1}{Y_1}\tilde{\alpha}\hat{S}(\tilde{\alpha}X - k_2)} \tag{7.204}$$

We note that the steady state error $\bar{e}_X$ will be equal to zero if $\frac{1}{Y_1}k_2 + k_1 + D = 0$. This gives the following expression for $\tilde{K_S}$ from (7.72) (7.73):

$$\tilde{K_S} = -\hat{S} + \frac{\mu_{max}D\hat{S}}{D^2 - (\lambda_1 + \lambda_2)D + \lambda_1\lambda_2} \tag{7.205}$$

By using a similar approach, if μ_{max} is assumed to be badly known instead of K_S, the value of $\tilde{\mu}_{max}$ in the observer that guarantees zero steady state error for the estimation of X is given by the following relationship:

$$\tilde{\mu}_{max} = \frac{(D^2 - (\lambda_1 + \lambda_2)D + \lambda_1\lambda_2)(K_S + \hat{S})}{D\hat{S}} \tag{7.206}$$

Figures 7.10 and 7.11 illustrate the behaviour of the ELO with $\tilde{K_S}$ given by equation (7.205) for $\lambda_1 = \lambda_2 = -0.2$. Figure 7.10 shows the convergence of the observer when $\hat{X}(0) = 0$. Figure 7.11 illustrates the performance of the observer in presence of a square wave of the influent concentration S_{in} (between 5 and 6 g/l, variations at time t = 40, 80, 120 and 160 h). The process is assumed to be initially in steady state ($X(0)$ = 2.054 g/l, $S(0)$ = 0.893 g/l); the observer has been initialised with the correct value for $\hat{S}$, and a wrong value for $\hat{X}$ (= 1 g/l). Note that, as expected, the estimate of X (dotted line) converges to the "true" simulated value. This is obviously done at the price of a biased estimate of the measured variable S. Note that in Figure 7.11 the estimate of S rapidly converges to its biased value (around 0.68 g/l).

Our numerical experience with different kinetic models and estimation problems tell us that the gains of the observer have to be selected not too large in order to avoid strange transient behaviour of the observer. So far we have not been able to obtain satisfactorily (i.e. under realistic conditions) theoretical stability and convergence for this approach.

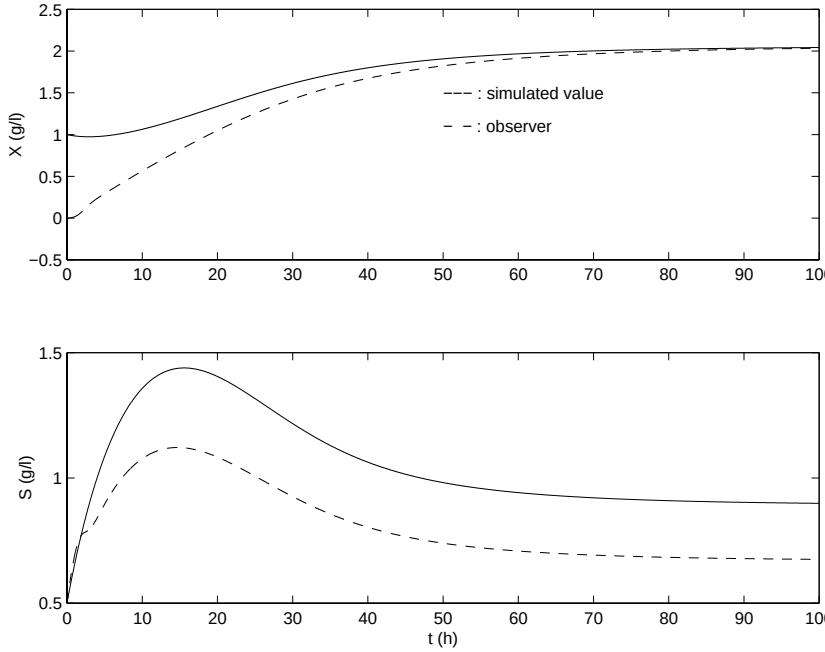

FIG. 7.10. Extended Luenberger observer with K_S computed from (7.205)(initial transient)(-: "true" simulated values; - -: estimates).

7.5.2 *Adaptive State Observer*

The second approach that we propose in order to handle (kinetic) parameter uncertainty in the state observation of (bio)chemical processes is the design of adaptive observers, i.e. observers that also estimate the badly known parameters (e.g. [58], [14]). One of the original features of the present adaptive observers is to consider a nominal (default) process model, i.e. a model with nominal values of the badly known parameters.

For simplicity we consider here that the badly known parameters are such that the process models are linear in these parameters. Then we can write the right hand side of (7.198) as follows:

$$f(x,u,\theta) = \bar{f}(x,u,\bar{\theta}) + b(x,u)\Delta\theta \tag{7.207}$$

From the above equations and the observer equations (7.27), the adaptive observer is readily obtained (by considering the badly known parameters (here $\Delta\theta$) as unmeasured states with dynamics equal to zero):

$$\frac{d\hat{x}}{dt} = \bar{f}(\hat{x},u,\bar{\theta}) + b(\hat{x},u)\widehat{\Delta\theta} + K(y-\hat{y}) \tag{7.208}$$

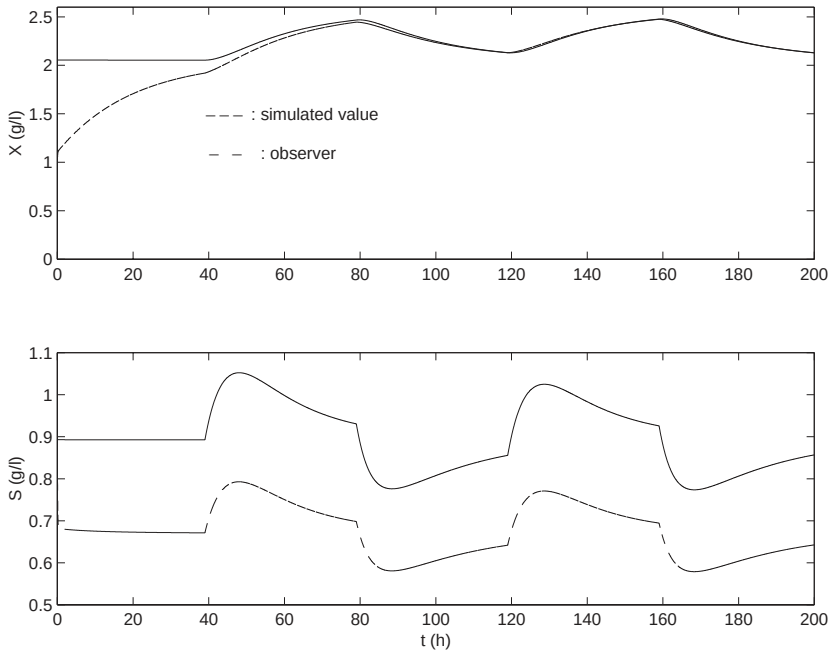

FIG. 7.11. Extended Luenberger observer with K_S computed from (7.205)(square wave S_{in})(-: "true" simulated values; - -: estimates).

$$\frac{d\widehat{\Delta\theta}}{dt} = \Gamma(y - \hat{y}) \tag{7.209}$$

Some important remarks are probably necessary at this point. Although the design of adaptive observers has been a very active research field in the 1980s (see e.g. [150], [70], [183], [167], [14]), the theoretical analysis of adaptive observers may become easily highly complex. Moreover our experience shows that their application to processes is indeed so far quite disappointing because they are very difficult to tune properly in practice. That's why we have limited our study here to the following items:

1. to base the design on a model with nominal values of the badly known parameters;
2. to consider here only models linear in the badly known parameters;
3. to allow only one badly known parameter per measured variable;
4. to consider what is called in the adaptive control nomenclature as the "averaging" (see e.g. [165]) as a key tuning rule for the adaptive observer.

Item 1 is important to increase the flexibility on the observer dynamics. Item 2 is only an a priori choice that simplifies the approach. Item 3 is probably the most

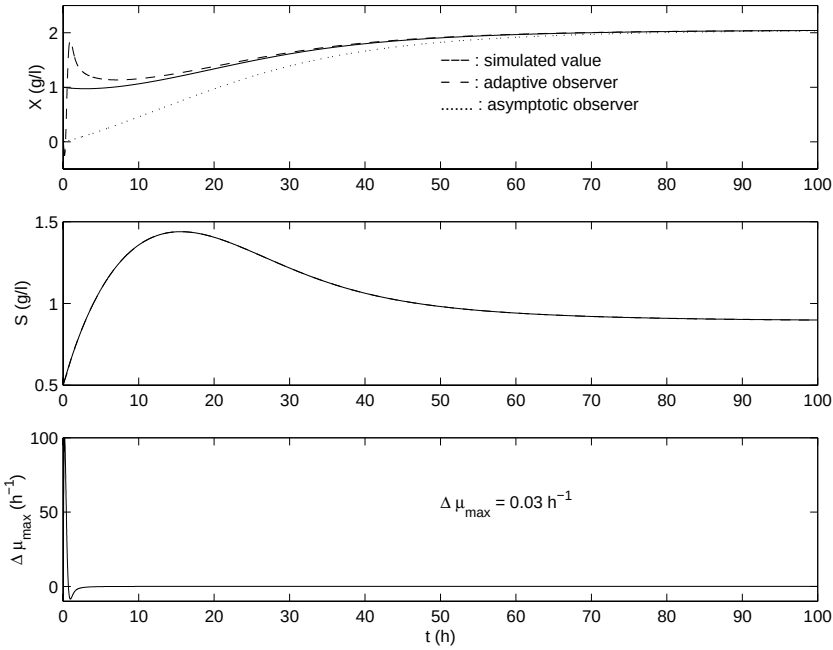

FIG. 7.12. Adaptive observer (initial transient).

essential assumption in order to have some guarantee for the successful implementation of the observer, and is rather easy to understand from a practical viewpoint: how can one expect to handle properly uncertainty on different kinetics and distinguish between them if there is not a sufficient amount of information about them, in particular one independent information per independent unknown. Finally, item 4 is also a key issue in the implementation of the adaptive observer: one important underlying idea of averaging is that the dynamics of the parameter adaptation has to be slower than the dynamics of the rest of the observer.

In the rest of the section we shall consider the same example as before (simple microbial growth, Section 7.3.4) on which we shall perform the theoretical analysis of the adaptive observer and illustrate its performance in numerical simulation.

Assume that μ_{max} is the badly known parameter with:

$$\mu_{max} = \bar{\mu}_{max} + \Delta\mu_{max} \tag{7.210}$$

where $\bar{\mu}_{max}$ is a nominal value of μ_{max}. Then equation (7.207) specialises as follows:

$$\frac{dS}{dt} = -\frac{1}{Y_1}\bar{\mu}_{max}\frac{SX}{K_S + S} + DS_{in} - DS - \frac{1}{Y_1}\frac{SX}{K_S + S}\Delta\mu_{max} \tag{7.211}$$

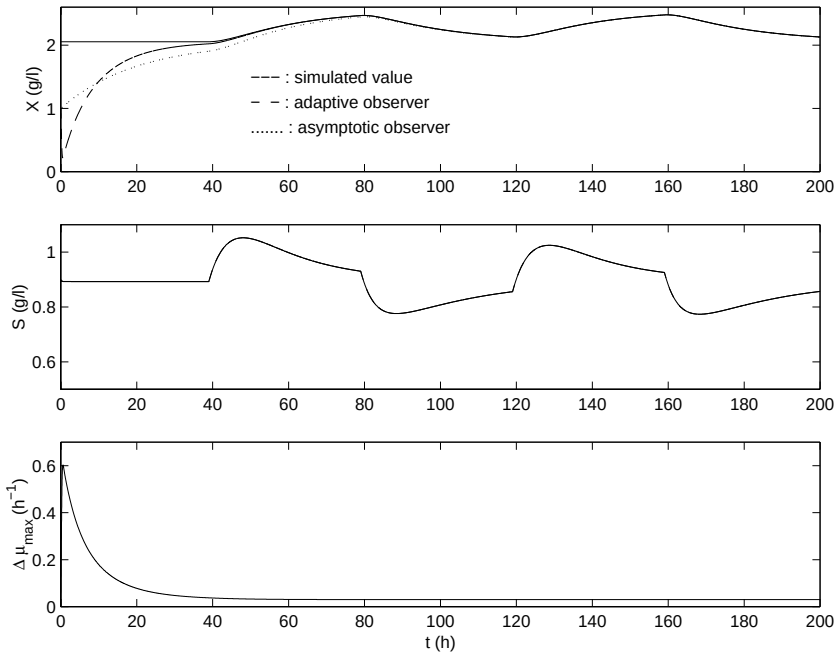

FIG. 7.13. Adaptive observer (square wave S_{in}).

$$\frac{dX}{dt} = \bar{\mu}_{max}\frac{SX}{K_S + S} - DX + \frac{SX}{K_S + S}\Delta\mu_{max} \tag{7.212}$$

At this point we consider an asymptotic observer for estimating the value of X that appears in the regressor (i.e. the multiplicative term) of $\Delta\mu_{max}$. Its introduction is fully motivated by the theoretical analysis that will be performed below. The asymptotic observer is based on the state transformation: $\zeta = S + \frac{1}{Y}X$. Its dynamical equations are readily derived from the model equations (7.43),(7.44):

$$\frac{d\zeta}{dt} = -D\zeta + DS_{in} \tag{7.213}$$

$$X_e = Y_1(\zeta - S) \tag{7.214}$$

where X_e denotes the estimate of X given by the asymptotic observer (we use a different notation in order to avoid the confusion with the estimate of X ($\hat{X}$) given by the adaptive observer). The adaptive observer is then equal to:

$$\frac{d\hat{S}}{dt} = -\frac{1}{Y_1}\bar{\mu}_{max}\frac{S\hat{X}}{K_S + S} + DS_{in} - DS - \frac{1}{Y_1}\frac{SX_e}{K_S + S}\widehat{\Delta\mu_{max}} + K_1(S - \hat{S}) \tag{7.215}$$

$$\frac{d\hat{X}}{dt} = \bar{\mu}_{max}\frac{S\hat{X}}{K_S + S} - D\hat{X} + \frac{SX_e}{K_S + S}\widehat{\Delta\mu_{max}} - K_2(S - \hat{S}) \quad (7.216)$$

$$\frac{d\widehat{\Delta\mu_{max}}}{dt} = -\gamma(S - \hat{S}) \quad (7.217)$$

The analysis of the theoretical stability and convergence properties of the adaptive observer is based on the estimation error dynamics. If we define:

$$e_S = S - \hat{S},\ e_X = X - \hat{X},\ e_\mu = \Delta\mu_{max} - \widehat{\Delta\mu_{max}} \quad (7.218)$$

these are written as follows:

$$\frac{de}{dt} = Ae + Be_{Xe} \quad (7.219)$$

with:

$$A = \begin{bmatrix} -K_1 & -\frac{1}{Y_1}\frac{\bar{\mu}_{max}S}{K_S+S} & -\frac{1}{Y_1}\frac{SX_e}{K_S+S} \\ K_2 & \frac{\bar{\mu}_{max}S}{K_S+S} & \frac{SX_e}{K_S+S} - D \\ \gamma & 0 & 0 \end{bmatrix}$$

$$e = \begin{bmatrix} e_S \\ e_X \\ e_\mu \end{bmatrix},\ B = \begin{bmatrix} -\frac{1}{Y_1}\frac{\Delta\mu_{max}S}{K_S+S} \\ \frac{\Delta\mu_{max}S}{K_S+S} \\ 0 \end{bmatrix},\ e_{Xe} = X - X_e$$

From the theory of the asymptotic observers, we know that e_{Xe} will tend asymptotically to zero, $\lim_{t\to\infty} e_{Xe} = 0$. We also know (see e.g. [14]) that the matrix B is bounded. Similarly, if K_1, K_2 and γ are bounded, then A is bounded. Therefore the error dynamics are that of a time-varying system with an input, e_{Xe}, going asymptotically to zero. If the state matrix A is asymptotically stable, we can use a classical stability result (e.g. [284], p.55) to state that the estimation errors will tend asymptotically to zero. Let us check now that A is a stable matrix. Its characteristic polynomial $\det(\lambda I - A)$ is equal to:

$$\lambda^3 + \lambda^2(K_1 + D - \frac{\bar{\mu}_{max}S}{K_S + S}) + \lambda(K_1(D - \frac{\bar{\mu}_{max}S}{K_S + S}) + K_2\frac{1}{Y}\frac{\bar{\mu}_{max}S}{K_S + S} + \gamma\frac{1}{Y}\frac{SX_e}{K_S + S}) + \gamma\frac{1}{Y}D\frac{SX_e}{K_S + S}$$

It is then routine to check that A is asymptotically stable via a proper choice of the gains K_1, K_2, and γ, e.g. if we assign the dynamics of the observer with the three eigenvalues λ_1, λ_2, λ_3. Figure 7.12 illustrates the performance of the observer under the same initial and operating conditions as in Figure 7.10, but with a 10% error on μ_{max}. The gains have been chosen primarily to assign the dynamics

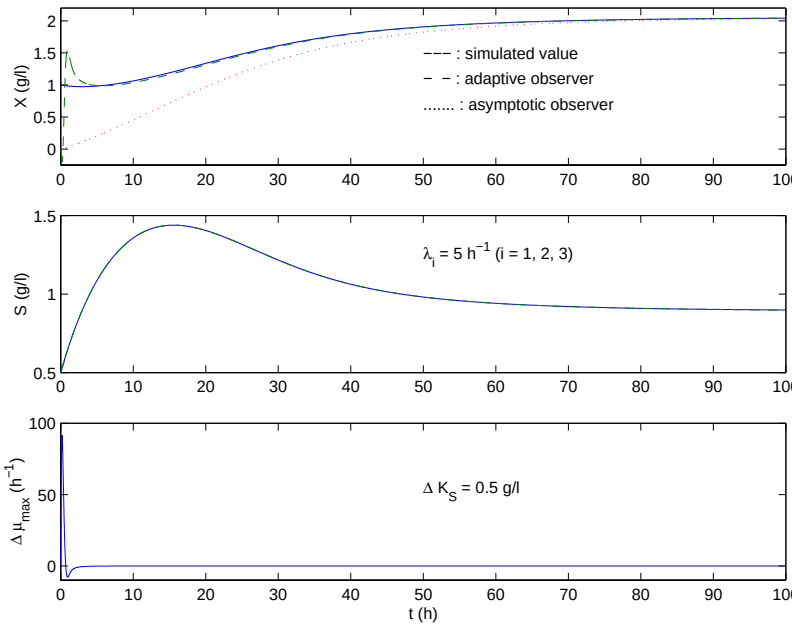

FIG. 7.14. Adaptive observer (10% error on K_S).

with $\lambda_1 = \lambda_2 = \lambda_3 = 5\ \mathrm{h}^{-1}$, then γ has been reduced by a factor 20 to follow the averaging recommendations (the factor is somewhat arbitrary (10 would also have been good too for instance), it just gives a rough idea of a typical value for γ). Figure 7.13 illustrates the performance of the adaptive observer in the presence of a square wave influent substrate concentration.

Note that the adaptive observer (which assumes the knowledge of the kinetics model) is able to converge faster than the asymptotic observer (which ignores the kinetics model).

It is also worth noting that the adaptive observer (7.215)(7.216)(7.217) is capable of handling nonlinear uncertainties like those on K_S, as it illustrated in Figure 7.14, which presents a numerical simulation performed under the same conditions as in Figure 7.12, but with a 10% error on K_S. The adaptive observer compensates the error on K_S by a bias on the estimate of μ_{max}.

7.5.3 *Generalisation*

The generalisation of both proposed observers is based on the General Dynamical Model:

$$\frac{d\xi}{dt} = -D\xi + Y\rho + F - Q \tag{7.220}$$

If we consider that the measured outputs y are process components, then the part of the General Dynamical Model associated to y is written as follows:

$$\frac{dy}{dt} = -Dy + Y_y\rho + F_y - Q_y \tag{7.221}$$

where Y_y, F_y and Q_y gather the rows of Y, F and Q related to y. The key assumptions at this point are to consider that the number of measured components is (at least) equal to the number of uncertain kinetics, and that the measured components are independent (this basically requires that the submatrix Y_y is full rank). The generalisation consists then of introducing the following state transformation (see also [191]):

$$\Psi = Y_y^{-1}y \tag{7.222}$$

This transformation decouples all the kinetic terms and associates each of them to only one variable Ψ. Indeed equation (7.221) can be rewritten as follows:

$$\frac{d\Psi}{dt} = -D\Psi + \rho + Y_y^{-1}(F_y - Q_y) \tag{7.223}$$

It is then straightforward to extend the design of both observers introduced in the above sections to each decoupled system (Ψ_i, ρ_i).

7.6 On-line Parameter Estimation

So far we have considered the estimation of state variables, i.e. state observation. In this section, we address the problem of estimating the reaction rates from on-line knowledge of the state variables (knowledge available either from measurements or from state estimation). It has been decided to restrict the section to the development, analysis and implementation of estimators of kinetic parameters. Our motivation for doing so comes basically from our experience that shows the kinetic parameters are in practice difficult to evaluate precisely via kinetic models and are often the primary and key source of uncertainty of bioprocess models (these difficulties has also motivated the writing of the preceding chapters, particularly Chapter 5 dedicated to experiment design for parameter estimation). It is also important to note that the approaches presented here do not exclusively apply to the estimation of kinetic parameters, but can also be used for the estimation of any other process parameters, like yield coefficients or transfer coefficients. Examples of estimation of such parameters can be found e.g. [14].

Two approaches will be considered: the observer-based estimator and the recursive least squares estimator.

7.6.1 *The Observer-Based Estimator*

The underlying idea of the observer-based estimator is (as its name suggests) to consider the classical observer structure already introduced in Section 7.3. Yet here

we go a step further by assimilating the unknown parameter as unknown states for which the dynamics are equal to zero. This means that if we consider that the dynamics of the process are given by the following equations:

$$\frac{dx}{dt} = f(x, u, \theta) \tag{7.224}$$

we assume that the unknown parameters θ have the following "dynamics":

$$\frac{d\theta}{dt} = 0 \tag{7.225}$$

and we consider both equations (7.224)(7.225) for the design of the observer-based estimator.

Let us now proceed further in the design of the estimator. Let us first consider the General Dynamical Model:

$$\frac{d\xi}{dt} = Y\rho(\xi) - D\xi - Q + F \tag{7.226}$$

We assume that:

a) the matrix Y of yield coefficients is known,
b) the dilution rate D, the feed rates F and the gaseous outflow rates Q are measured on-line,
c) the vector of state variables ξ is known either by measurement or by estimation using an asymptotic observer (as described in Section 7.4).

We further assume that the vector $\rho(\xi)$ of reaction rates is partially unknown and written as follows:

$$\rho(\xi) = \begin{bmatrix} H(\xi)r(\xi) \\ h(\xi) \end{bmatrix} \tag{7.227}$$

where $H(\xi)$ is a diagonal matrix of known functions of the state and $r(\xi)$ a vector of unknown functions of ξ with $\dim r(\xi) = n_u$.

The known reaction rates are given by vector $h(\xi)$ with $\dim h(\xi) = M - n_u$.

Using equation (7.227), the general dynamical model is rewritten as:

$$\frac{d\xi}{dt} = Y_u H(\xi)r(\xi) + Y_k h(\xi) - D\xi - Q + F \tag{7.228}$$

where Y_u and Y_k are matrices of yield coefficients associated with the unknown and known reaction rates respectively.

The observer-based estimator is written as follows:

$$\frac{d\hat{\xi}}{dt} = Y_u H(\xi)\hat{r} + Y_k h(\xi) - D\xi - Q + F - \Omega(\xi - \hat{\xi}) \tag{7.229}$$

$$\frac{d\hat{r}}{dt} = [Y_u H(\xi)]^T \Gamma(\xi - \hat{\xi}) \tag{7.230}$$

As in the classical observer, the update of the parameter vector $\hat{r}$ is driven by the deviation term $(\xi - \hat{\xi})$ which reflects the mismatch between $\hat{r}$ and r.

The matrices Ω and Γ are tuning parameters for adjusting the rate of convergence of the algorithm [14]. A common choice is:

$$\Omega = diag(-\omega_i), \;\; \Gamma = diag(-\gamma_j), \;\; \omega_i, \gamma_j > 0 \tag{7.231}$$

With this choice, the stability of the estimator is satisfied.

The tuning procedure may be simplified if the state equations are first decoupled using the transformation (7.222) already considered in Section 7.5.3 [191]:

$$\Psi = Y_u^{-1}\xi \tag{7.232}$$

The dynamical equations of the new variables Ψ are readily obtained from (7.228):

$$\frac{d\Psi}{dt} = H(\xi)r(\xi) + Y_u^{-1}Y_k h(\xi) - D\Psi + Y_u^{-1}(F - Q) \tag{7.233}$$

Note that each entry Ψ_i of Ψ is a linear combination of the entries ξ_i of ξ. Applying the estimation algorithm to the transformed state equations yields:

$$\frac{d\hat{\Psi}}{dt} = H\hat{r} + Y_u^{-1}Y_k h - D\Psi + Y_u^{-1}(F - Q) - \Omega(\Psi - \hat{\Psi}) \tag{7.234}$$

$$\frac{d\hat{r}}{dt} = H\Gamma(\Psi - \hat{\Psi}) \tag{7.235}$$

Recall that H is a diagonal matrix. Let us gather each variable Ψ_i with its related parameter r_i and rearrange the entries of the vector $[\Psi, r]^T$ in the following order in a vector ζ:

$$\zeta = \begin{bmatrix} \Psi_1 \\ r_1 \\ \Psi_2 \\ r_2 \\ \dots \\ \Psi_p \\ r_p \end{bmatrix} \tag{7.236}$$

Basic tuning rule. Let us first define the estimation error:

$$e = \zeta - \hat{\zeta} \tag{7.237}$$

The estimation error dynamics are readily derived from equations (7.233) (7.234) and (7.235):

$$\frac{de}{dt} = Ae + b \tag{7.238}$$

with a block diagonal matrix A with 2×2 blocks:

$$A = diag\{A_i\},\ A_i = \begin{bmatrix} -\omega_i & h_i(\xi) \\ -\gamma_i & 0 \end{bmatrix},\ i = 1\ to\ p \tag{7.239}$$

and b equal to:

$$b = \begin{bmatrix} 0 \\ \frac{dr_1}{dt} \\ 0 \\ \frac{dr_2}{dt} \\ \cdots \\ 0 \\ \frac{dr_p}{dt} \end{bmatrix} \tag{7.240}$$

The characteristic equation of the matrix A, $\det(\lambda I - A)$, is equal to:

$$det(\lambda I - A) = \prod_{i=1}^{p} (\lambda^2 + \omega_i \lambda + \gamma_i h_i(\xi)) \tag{7.241}$$

The key idea of the tuning rule consists of choosing each γ_i inversely proportional to the corresponding term $h_i(\xi)$:

$$\gamma_i = \frac{\bar{\gamma}_i}{h_i(\xi)},\ \bar{\gamma}_i > 0,\ i = 1\ to\ p \tag{7.242}$$

With the choice above, the characteristic equation (7.241) is rewritten as follows:

$$det(\lambda I - A) = \prod_{i=1}^{p} (\lambda^2 + \omega_i \lambda + \bar{\gamma}_i) \tag{7.243}$$

and the observer-based estimator dynamics are now independent of the state variables. Such a choice corresponds to a Lyapunov transformation (see [192]). It is obviously valid for values of $h_i(\xi) \neq 0$: this condition is usually met easily in (bio)process applications, as it will be illustrated in the following section.

The values of the design parameters can then be set to arbitrarily fix the estimator's dynamics for each unknown parameter r_i. Since the estimator reduces, via the transformations, to a set of independent second-order linear systems, the classical rules for assigning the dynamics of second-order linear systems apply straightforwardly here. The reader is therefore referred to the classical automatic control textbooks for further information on the subject. However the following basic guidelines are suggested.

One important guideline is to choose real poles:

$$\omega_i^2 - 4\bar{\gamma}_i \geq 0 \tag{7.244}$$

The objective of this is to avoid oscillations in the estimation of the parameters that do not correspond to any physical phenomenon related to the estimated reaction rates.

Pomerleau and Perrier ([198]) suggest choosing double poles, i.e.:

$$\bar{\gamma}_i = \frac{\omega_i^2}{4} \tag{7.245}$$

The tuning of the estimation algorithm reduces to the choice of one design parameter, ω_i, per estimated parameter. This allows having a design procedure that has the double advantage of being simple (one design parameter) and flexible (each parameter estimation can be tuned differently if needed, e.g. if the time variations of the parameters are different).

As an alternative, Oliveira *et al.* ([186]) propose to choose complex poles with a damping factor equal to 0.7 in order to increase the speed of convergence of the estimator with a reduced overshoot. (Generally speaking, the damping factor can be freely chosen; the choice may then depend on the type and nature of the application, of the time variations of the parameter to be estimated, and of the noise on the measured data. This means that then there are, in this approach, two design parameters per estimated parameter).

So far, we have suggested that it is possible to assign arbitrarily the dynamics of the estimator. However in presence of noisy data, it appears that indeed a compromise has to be made between a fast estimator convergence and a good noise rejection. A detailed and somewhat involved study is performed in Bastin and Dochain [14] (pp. 162-72) to analyse the performance of the observer-based estimator both in theory and in numerical simulation in the presence of bounded noisy data in the particular case of the estimation of the specific growth rate of a simple microbial growth process. The theoretical optimisation analysis is based on the evaluation of the asymptotic properties of the estimator and results in the following optimal value for ω_1:

$$\omega_{1,opt} = 2\sqrt{\frac{M_1}{\alpha(M_2^2 + Y_1 S_{max} M_2)}}, \qquad 0 < \alpha < 1 \tag{7.246}$$

where M_1 and M_2 are the upper bounds on the time derivative of μ and on the measurement noise, respectively, and S_{max} the maximum value of the influent substrate concentration S_{in}.

This result is probably rather conservative because it is based on upper bounds for the measurement noise, the time variation of μ and the influent substrate concentration, but it is qualitatively confirmed by numerical simulation studies which

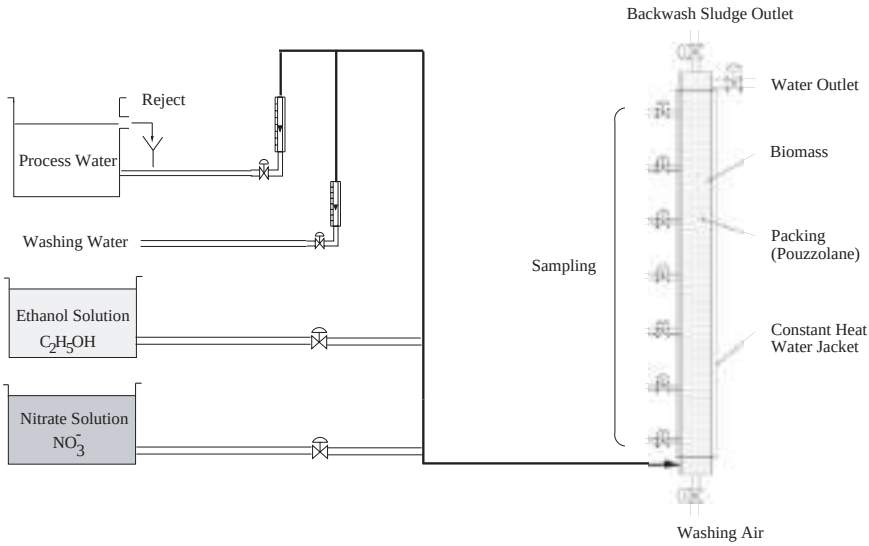

FIG. 7.15. Experimental set-up of the denitrification process.

also give a value of the design parameters that minimises the estimation error. Because it is conservative, the theoretical optimum has to be interpreted with care, but since it is qualitatively correct, our suggestion (also based on our practical experience), in presence of noisy measurements, is to perform numerical simulations with a plausible reaction rate model and noise in order to get a first initial guess for the design parameter values. These can then be adjusted when applied to the real process.

Application to a denitrification process. Let us consider the application to the observer-based estimator to a denitrification process. The process considered here is a pilot-scale denitrifying biofilter [34], [36]. The denitrification process under study is a submerged granular fixed bed biofilter (see Figure 7.15); it is packed with pouzzolane, a very porous volcanic material which retains a big amount of microorganisms even after back-washing. Thus the initial biomass concentration is often very close to the maximum active biomass concentration. The substrates to be removed can move freely along the reactor. The inside temperature is kept almost constant via a double insulation jacket. An ethanol solution is mixed with the feeding nitrate in a tank at the inlet of the reactor. Eight measurement points are distributed at every thirty centimeters along the column for the measurement of nitrate and nitrite concentrations.

The denitrification process is defined as being the reduction of nitrates (electron acceptor) NO_3^- into gaseous nitrogen N_2 by using organic carbon (electron

donor) with the production of intermediate compounds, namely nitrites NO_2^-. In the denitrification, two sequential biological reactions are considered in the model, denitratation (1) and denitritation (2):

$$NO3^- \xrightarrow{(1)} NO2^- \xrightarrow{(2)} N_2$$

If we assume that the liquid and solid mediums are in completely mixed conditions and that the biomass is fixed, the dynamics of the process are given by the following equations:

$$\frac{dS_{NO_3}}{dt} = -DS_{NO_3} + DS_{NO_{3,in}} - \frac{1}{Y_1}\mu_1 X \tag{7.247}$$

$$\frac{dS_{NO_2}}{dt} = -DS_{NO_2} + Y_3\mu_1 X - \frac{1}{Y_2}\mu_2 X \tag{7.248}$$

$$\frac{dS_C}{dt} = -DS_C - DS_{C,in} - \frac{1}{Y_4}\mu_1 X - \frac{1}{Y_5}\mu_2 X \tag{7.249}$$

$$\frac{dX}{dt} = (\mu_1 + \mu_2 - k_d)\, X \tag{7.250}$$

where S_C is the (external) source of carbon (ethanol) for both biochemical reactions, and k_d is the biomass death coefficient.

As a matter of illustration of the flexibility of the algorithm, let us consider that we consider the following rewriting of the specific growth rates μ_1 and μ_2:

$$\mu_1 = r_1 S_{NO_3},\ \mu_2 = r_2 S_{NO_2} \tag{7.251}$$

which simply expresses an explicit dependence between the specific growth rates and one of their limiting substrates. Then:

$$r = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix},\ H = \begin{bmatrix} S_{NO_3} X & 0 \\ 0 & S_{NO_2} X \end{bmatrix} \tag{7.252}$$

In the present experiments, the concentrations of nitrate S_{NO_3} and nitrite S_{NO_2} are accessible for on-line measurement. The auxiliary variables Ψ are obtained from the inverse of the yield coefficient matrix Y associated to these two variables:

$$Y = \begin{bmatrix} -\dfrac{1}{Y_1} & 0 \\ Y_3 & -\dfrac{1}{Y_2} \end{bmatrix} \tag{7.253}$$

The auxiliary variables are then defined as follows:

$$\Psi_1 = -Y_1 S_{NO_3},\ \Psi_2 = -Y_2 S_{NO_2} - Y_1 Y_2 Y_3 S_{NO_3} \tag{7.254}$$

The observer-based estimator is written as follows:

$$\frac{d\hat{\Psi}_1}{dt} = -D\Psi_1 - DY_1 S_{NO_{3,in}} + \hat{r}_1 S_{NO_3} X + \omega_1(\Psi_1 - \hat{\Psi}_1) \tag{7.255}$$

$$\frac{d\hat{r}_1}{dt} = \frac{\omega_1^2}{4S_{NO_3} X}(\Psi_1 - \hat{\Psi}_1) \tag{7.256}$$

$$\frac{d\hat{\Psi}_2}{dt} = -D\Psi_2 - DY_2(Y_1 Y_3 S_{NO_{3,in}} + S_{NO_{2,in}}) + \hat{r}_2 S_{NO_2} X + \omega_2(\Psi_2 - \hat{\Psi}_2) \tag{7.257}$$

$$\frac{d\hat{r}_2}{dt} = \frac{\omega_2^2}{4S_{NO_2} X}(\Psi_2 - \hat{\Psi}_2) \tag{7.258}$$

Figure 7.16 shows the estimation results of r_1 and r_2 based on real-life data. The (known) parameters of the model are equal to:

$$Y_1 = 1.05,\ Y_2 = 0.67,\ Y_3 = 1.05,\ Y_4 = 0.33,\ Y_5 = 0.22,\ k_d = 0.01h^{-1}$$

The flow rate was equal to 32.5 l/h. The influent concentrations are given in Figure 7.16. The observer-based estimator was initialised as follows:

$$\omega_1 = 0.5\,h^{-1},\ \omega_2 = 0.5\,h^{-1},\ \hat{\Psi}_1(0) = \Psi(0),\ \hat{\Psi}_2(0) = \Psi_2(0) \tag{7.259}$$

$$\hat{r}_1(0) = 0.0001\,m^3/g/h,\ \hat{r}_2(0) = 0.0002\,m^3/g/h \tag{7.260}$$

Validation of the estimation has been performed on data of ethanol S_C by comparing real-life data with values computed on the basis of the estimated values of r_1 and r_2 and the mass balance equation of S_C, i.e.:

$$\frac{d\hat{S}_C}{dt} = -D\hat{S}_C - DS_{C,in} - \frac{1}{Y_4}\hat{r}_1 S_{NO_3} X - \frac{1}{Y_5}\hat{r}_2 S_{NO_2} X \tag{7.261}$$

The values are given in Figure 7.16 (bottom right).

7.6.2 *The Recursive Least Squares Estimator*

Since the model is linear in the unknown parameters, an alternative estimation algorithm may be the least squares algorithm in recursive form, or the recursive least squares (RLS) scheme. We have presented in Chapter 6 the non recursive version of the least squares estimation algorithm.

Let us consider the following discrete-time equation linear in the (unknown) parameter θ:

$$y_t = \phi_t^T \theta_t \tag{7.262}$$

where t is the time index. The RLS scheme with forgetting factor γ based on the above equation (7.262) is indeed written as follows (see e.g. [151] for a detailed approach to derive the recursive form from the batch ("off-line") version):

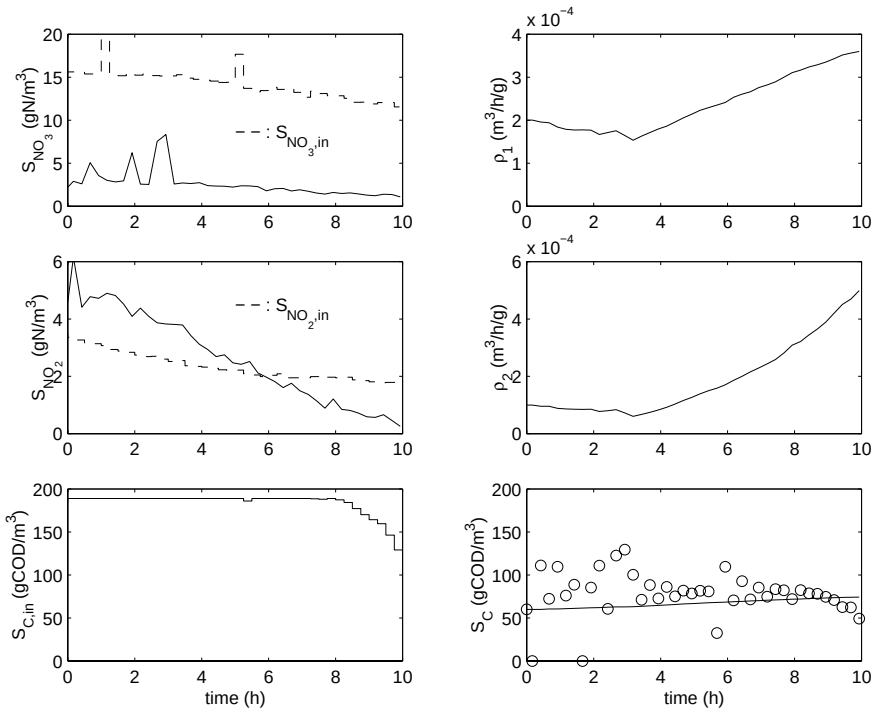

FIG. 7.16. Observer-based estimator for a denitrifying biofilter

$$\hat{\theta}_{t+1} = \hat{\theta}_t + g_t \phi_t (y_t - \phi_t^T \hat{\theta}_t) \tag{7.263}$$

$$g_t = \frac{g_{t-1}}{\gamma} \left[I - \phi_t (\gamma I + \phi_t^T g_{t-1} \phi_t)^{-1} \phi_t^T g_{t-1} \right] \qquad g_0 > 0,\ 0 < \gamma \leq 1 \tag{7.264}$$

where I is the identity matrix, and g_t is the gain of the estimator. The role of the forgetting factor is to follow possible time variations of the unknown parameter θ by giving a weight to recent data that is higher than that given to old data. $\gamma = 1$ means that an equal weight is given to all data, and there is then no forgetting effect. The forgetting factor is typically (but not necessarily) chosen between 0.9 and 0.99 in practical applications.

Let us see how to apply the RLS scheme to the estimation of kinetic parameters on the basis of the General Dynamical Model. If we consider an Euler approximation for the time derivative:

$$\frac{d\xi}{dt} = \frac{\xi_{t+1} - \xi_t}{\Delta t} \tag{7.265}$$

with Δt the sampling period, the version used for the observer-based estimator (7.228) is written in discrete-time as follows:

$$\xi_{t+1} = \xi_t + \Delta t(Y_u H(\xi_t) r(\xi_t) + Y_k h(\xi_t) - D_t \xi_t - Q_t + F_t) \tag{7.266}$$

The RLS scheme can be immediately applied by considering that:

$$y_t = \xi_{t+1} - \xi_t - \Delta t(Y_k h(\xi_t) - D_t \xi_t - Q_t + F_t) \tag{7.267}$$

$$\phi_t^T = \Delta t Y_u H(\xi_t),\ \theta_t = r(\xi_t) \tag{7.268}$$

The selection of the initial gain g_0 is typically an important choice for starting correctly the RLS estimator. Obviously it must be positive definite. Usually the literature suggests starting with a diagonal matrix with large values. However in practice too large values may result in very large transients in the starting phase. Our recommendation for RLS estimation with forgetting factor is to select initial values for entries of the gain matrix that have the same order of magnitude as those reached in steady state.

This can be easily done when estimating only one parameter. Indeed the gain matrix g_t (as well the regressor ϕ_t) reduces to a scalar and is written as follows:

$$g_t = \frac{g_{t-1}}{\gamma + g_{t-1}\phi_t^2} \tag{7.269}$$

In steady state , $g_t = g_{t-1}$, and therefore we can draw the following expression for g_t:

$$g_t = \frac{1-\gamma}{\phi_t^2} \tag{7.270}$$

Therefore our recommendation is then to select the initial value g_0 close to $\dfrac{1-\gamma}{\phi_0^2}$.

Application to an anaerobic digestion process. Let us consider the anaerobic digestion process already introduced in Section 7.4 and Figures 7.5 and 7.6.

Here again, methane P_2 and volatile fatty acids S_2 are available measurements. Let us use them to estimate the specific growth rates μ_1 and μ_2 of the acidogenesis and of the methanisation, respectively.

In discrete-time, and by considering the low solubility assumption on methane, the mass balance equations for methane and volatile fatty acids are written as follows:

$$Q_{2t} = Y_6 \mu_{2t} X_{2t} \tag{7.271}$$

$$S_{2,t+1} = S_{2t} - \Delta t D_t S_{2t} + \frac{1}{Y_2}\mu_{1t}\Delta t X_{1t} - Y_3 \mu_{2t} \Delta t X_{2t} \tag{7.272}$$

Because μ_2 appears in both equations while μ_1 appears only in the second, we have here designed the estimators in cascade. At each time step, we proceed as follows:

1. estimation of μ_2 on the basis of equation (7.271);

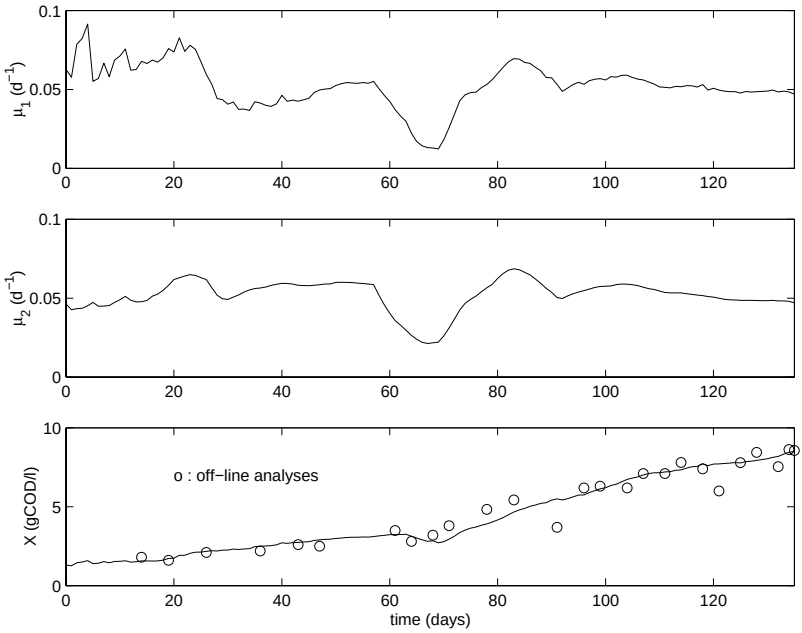

FIG. 7.17. On-line estimation of the specific growth rates in an anaerobic digestion process with a recursive least squares algorithm.

2. estimation of μ_1 on the basis of equation (7.272) and with the estimate of μ_2 given at step 1

The estimation results are given in Figure 7.17. The values of the biomass concentrations X_1 and X_2 are provided by the asymptotic observer of Section 7.4. The RLS estimators have been initialised as follows:

$$\text{estimation of } \mu_2 : y_t = Q_{2t},\ \phi_t = Y_6 X_{2t},\ \gamma = 0.9,\ g_0 = 0.01 \quad (7.273)$$

$$\text{estimation of } \mu_1 : y_t = S_{2,t+1} - S_{2t} + \Delta t D_t S_{2t} + Y_3 \hat{\mu}_{2t} \Delta t X_{2t} \quad (7.274)$$

$$\phi_t = \frac{1}{Y_2} \Delta t X_{1t},\ \gamma = 0.9,\ g_0 = 0.01 \quad (7.275)$$

The estimates $\hat{\mu}_1$ and $\hat{\mu}_2$ have been validated by considering, as in Section 7.4 and Figure 7.6, the off-line data of total and soluble COD to calculate the biomass concentration. These have been compared to values computed on the basis of the estimates of μ_1 and μ_2 with the mass balance equation for $X_1 + X_2$:

$$\frac{d\widehat{(X_1 + X_2)}}{dt} = -D\widehat{(X_1 + X_2)} + \hat{\mu}_1 X_1 + \hat{\mu}_2 X_2 \quad (7.276)$$

7.7 Conclusions

This chapter has been dedicated to the design of software sensors to estimate concentrations (state observers) and kinetic parameters (parameter estimation). We have started the chapter with theoretical notions of state observability, which appeared to be important to evaluate the possibility to apply and design state observers. The state observers have been presented in three sections: classical observers, asymptotic observers, and "intermediate" observers. The last section has been concerned with the design of recursive parameter estimation algorithms. Two approaches have been presented: the observer-based estimator, and the recursive least squares estimator. All the approaches proposed in the present chapter have been mathematically analysed. Their performances have been illustrated via numerical simulations and/or real-life results. And we have proposed several guidelines to help the user to implement them.

At this point, we would also like to draw attention to the following conclusions concerning the application of the software sensors introduced here.

1. As it has been already explained, the parameter estimation section has been restricted to the estimation of kinetic parameters. Our motivation for doing so comes basically from our experience that shows that kinetic parameters are in practice difficult to evaluate precisely via kinetic models and are often the primary and key source of uncertainty of bioprocess models (these difficulties have also motivated the writing of the preceding chapters, particularly Chapter 5 dedicated to experiment design for parameter estimation). It is also important to note that the approaches presented here do not exclusively apply to the estimation of kinetic parameters, but can also be used for the estimation of any other process parameters, like yield coefficients or transfer coefficients. Examples of estimations of such parameters can be found e.g. [14].
2. Today the estimation of states and parameters in nonlinear systems remains generally speaking an open question. This explains why it is an active research area. In the present chapter, we have followed a line in which we have privileged three approaches (classical observers, asymptotic observers, and "intermediate" observers). The reason for doing so is twofold. First, this sequence had the pedagogical advantage to present approaches that are dedicated to solve three types of problems (state observation with known model parameters, with unknown model structure, and with badly known parameters, respectively). Secondly, the proposed approaches present a high degree of reliability due to our experience in implementing them (as illustrated in the numerical and real-life examples). But this choice does not mean that we consider that other approaches must be rejected or are generally speaking worse than those presented here. We even have the feeling that we are giving here an instantaneous photograph of the estimation of

state and parameter estimation in WWTP, and that the situation may change rapidly with possibly new emerging approaches.

3. Yet, our experience shows that today the asymptotic observer represents the best state observers that can be applied to bioprocesses in general and to WWTP in particular. This argument is largely based to the large number of applications that we have seen over the last 20 years. The deciding factor in favour of the asymptotic observers is the following. The asymptotic observer is indeed a reliable (because of its theoretical stability and convergence properties) replica of the reaction scheme and of the hydraulics that have been chosen by the user to represent the dynamics of the process. Therefore, if for any reason the asymptotic observer does not work well, the diagnosis is usually easy: it is either due to a bad selection of the reaction scheme or of the hydrodynamics, or to wrong values of the yield coefficients, or finally to on-line measurements that may appear to be unreliable.

APPENDIX A
Glossary

A.1 Model Constituents

constant: model constituent, whose value is constant throughout all possible applications of the model.

forcing function: function used as model input.

input/output model: model that describes system behavior as being a function of only present input and past inputs and outputs.

model: abstraction of reality.

model structure: the relations between inputs, outputs and eventually states formulated as equations.

observation equation: equation in state-space model, that relates the state variables to the outputs (sometimes also denoted as output equation).

parameter: model constituent, whose value needs to be determined for each specific application of the model.

state: present situation of the system as described by the model.

state variable: model constituent in state-space models, acting as mediator between inputs and outputs and used for a descriptive representation of the system.

state-space model: model that includes a descriptive representation of the system by means of an additional set of state variables.

state-transition equation: function that relates the future state of the system to the present state and inputs.

transfer function: same as input/output function.

variables: inputs, outputs and eventually state variables in model equations.

A.2 Model Attributes

adaptive: model that interacts with the real system and changes the values of its inputs or state variables depending on past output values.

aggregated: model that contains state variables that represent functional classes of different constituents (e.g. organisms) or that simplifies the spatial configuration of a system by lumping it together (cf. segregated).

black-box: model that describes the observed behaviour of the corresponding subsystem without being based on the mechanisms of this subsystem (cf. white-box).

complex: a relative attribute that values whether the model contains more state variables, parameters, forcing functions, etc. or an attribute that qualifies that (irrespective of the number of variables) there exists chaotic solutions of the model equations (cf. simple).

continuous space: the model resolves the spatial domain of the system continuously (cf. discrete space).

continuous time: the model resolves the time axis continuously; the time evolution is usually described by differential equations (cf. discrete time).

conceptual: a model that contains a description of the ideas/hypothesis on system behaviour without giving a mathematical formulation.

deterministic: the time evolution of the model solution is uniquely determined by the initial state (for state-space models) and the time evolution of inputs (cf. stochastic).

discrete space: the model approximates the spatial domain of the system by a number of mixed compartments (cf. continuous space).

discrete time: the model divides the time axis into periods of finite length and the output or the state of the model in the next period is given as an algebraic equation depending on the old inputs or states (cf. continuous time).

distributed parameter: model with more than one independent variable, i.e. model behaviour is governed by partial differential equation in time and space.

dynamic: the model describes the time evolution of a system; a solution of the model may anyway be in steady-state (cf. static, steady-state).

empirical: the model equations are not based on generally accepted laws but are just of a descriptive nature (cf. phenomenological, mechanistic).

grey-box: the model consists of submodels that partly are based on mechanistic, partly on phenomenological descriptions (cf. black-box, white-box).

heuristic: model not based on rigorous development but on rules of thumb, feeling, qualitative reasoning.

linear: model equations are linear in input variables (for input-output models) or in state variables (for state-space models).

lumped: equal to aggregated.

mechanistic: model with the goal of describing the mechanisms that lead to the observed behaviour (cf. phenomenological).

nonlinear: model equations are nonlinear in input variables (for input-output models) or in input and state variables (for state-space models).

phenomenological: describing the observed phenomena without representing the mechanisms governing the behaviour (cf. mechanistic).

physical: model that is based on a description with physical, chemical or biological laws; sometimes also used for small scale reproductions of a system made in physical materials.

reduced order: model of reduced complexity obtained by direct deduction (e.g. by aggregation/lumping) from a more complex basic model.

reductionist: hierarchical description of a system by resolving it in subsystems that again are resolved in sub-subsystems until a description level is reached at which a satisfying description is possible without empirical assumptions (in the ideal case down to a description that is based on natural laws).

segregated: model that separates variables in more functional classes (cf. aggregated).

semi-physical: equal to grey-box .

simple: relative attribute that describes that the model equations contain only few state variables, parameters, forcing functions, etc. and the solutions show simple behaviour (periodic or quasi-periodic; cf. complex).

static: the model only describes the steady-state solution of a system (cf. dynamic).

steady state: the model only describes the steady-state solution of a system (cf. dynamic).

stochastic: the time evolution of the model contains random elements (cf. deterministic).

time-invariant: the way the model processes input to output does not change with time.

transparent: equal to white-box.

white-box: model that describes a system by one or several submodels (white-boxes) that describe the observed behaviour of the corresponding subsystem by describing the relevant mechanisms of this subsystem (cf. black-box).

A.3 Terms of Model Building

calibration: the same as parameter estimation but not necessarily by using statistical methods.

corroboration: the same as validation; the term introduced by Popper [200] makes it clearer that the correctness of the model cannot be proved and that each successful test only increases the belief that the model is correct (cf. confirmation, falsification, validation, verification).

confirmation: the same as validation; the term makes it clearer that the correctness of the model cannot be proved and that each successful test only increases the belief that the model is correct (cf. corroboration, falsification, validation, verification).

falsification: demonstrating the invalidity of a model by showing that the model results deviate significantly from the measurements confirmation, corroboration, validation and verification are fail trials of falsification (cf. confirmation, corroboration, validation, verification).

frame definition: selection of which components of a system are to be described and specification of classes of models to be included in the model structure selection process and specification of the experimental conditions for use of the model (experimental frame).

identifiability analysis (structural, practical): evaluation of the uniqueness of the estimates of model parameters from measured data. Structural (theoretical, a priori) identifiability analysis assesses the uniqueness of parameter estimates from ideal data for a given experimental frame, practical (a posteriori) identifiability analysis assesses the accuracy with which parameters can be estimated with a given data set. In the latter case identifiability is not an objective property, but it depends on the required accuracy.

model building: the process of finding an adequate model of a system by (iteratively) processing the following model building steps: Problem formulation, prior knowledge collection, system identification and model testing (see Figure 1.1).

model reduction: simplification of an existing model in order to improve its identifiability without loosing the description of the most important phenomena.

model (structure) selection: selecting out of a given set of model structures the structure that makes an optimal (as simple as possible but as complicated as required for the intended purpose) description of measured data possible.

optimal experimental design: using a (preliminary) model of a system in order to plan an experiment that maximises the possible gain in information.

parameter estimation (batch, recursive): process of finding parameter values that lead to an optimal agreement of model results with measured data by using statistical methods. Time series of data can be used as a whole (batch estimation) or data points from within a moving data window can be used. In

the latter case the parameters become time-dependent and the algorithm can be implemented to modify the previous estimate by considering the omitted and the new data points (recursive estimation).

regression (linear, nonlinear): the same as parameter estimation, however usually used for the special case of algebraically given linear or nonlinear model equations and using the (weighted) least squares technique goodness-of-fit criterion.

simulation (interactive, real-time, Monte Carlo): calculating the solution of a model for given values of the parameters, inputs and intial values (usually numerically). Interactive simulations are processed on a computer which allows the user to interact with the program (stop, change parameter values, etc.). Monte Carlo simulation is a method to propagate probability distributions of parameters, inputs and initial values to probability distributions of model results by performing a lot of simulations with parameter values sampled randomly from the probability distribution of the parameters, inputs and initial values.

system identification: finding a model to solve a given problem by (iteratively) processing the following identification steps: Frame definition, model structure selection, parameter estimation, model diagnosis. (see Figure 1.1).

structure characterisation: the same as model (structure) selection.

uncertainty analysis: estimating the uncertainty of model predictions and analysing the sources of uncertainty.

validation: test of a model with a data set not used for identification; note that such tests only increase the belief in the correctness of the model, it is not possible to prove that the model is correct (cf. confirmation, corroboration, falsification, verification).

verification: the same as validation (cf. confirmation, corroboration, falsification, validation).

APPENDIX B
Nomenclature

a_i, i = 1 to n	: Coefficients of the polynomial A(q)	
a	: Inhibition constant (logistic model)	[l/g]
a, b	: Constants (weir flow rate model)	
a, b, c	: Constants (μ(pH))	
A	: State matrix	
A	: Reactor cross-section	[m^2]
A_L	: Liquid phase cross-section	[m^2]
A_S	: Solid phase cross-section	[m^2]
A_{se}	: Settler cross-section	[m^2]
$A(q)$	: Polynomial in q (denominator of a transfer function)	
b	: Death coefficient	[/h]
b_i, i = 1 to m	: Coefficients of the polynomial B(q)	
B	: Input matrix	
$B(q)$	: Polynomial in q (numerator of a transfer function)	
C	: Component concentration	[g/l]
C	: Joint experimental design criterion for SC and PE	
C_i	: Measurement error covariance matrix	
D	: Dilution rate	[/h]
D	: Discriminative power of an experiment	
D_{ma}	: Axial mass dispersion (diffusion) coefficient	[m^2/s]

DO	: Dissolved oxygen	
e	: Observation or estimation error	
E	: Parameter estimation accuracy of an experiment	
E	: Activation energy	[J/mol]
E	: Expectation operator	
f	: State function	
F	: Hydraulic flow rate	[l/h]
F	: Feed rate vector	[g/l/h]
F	: Fisher Information Matrix	
F_l	: Solid flux	[g/m^2/h]
g	: Gravity constant	[m/s^2]
h	: Output function	
$H(s)$	: Transfer function	
h, H	: Height	[m]
J	: Objective functional	
J_{opt}	: Minimal value of the objective functional	
k_0	: Kinetic constant	
k_r, k_s	: Rate constants	[/h]
$k_L a$	: Mass transfer coefficient	[/h]
K	: Gain matrix (of an observer)	
K_c	: Contois model constant	
K_i	: Inhibition constant	[g/l]
K_p	: Product inhibition constant	[g/l]
K_s	: Affinity (or saturation, or Michaelis-Menten) constant	[g/l]
L	: Length	[m]
L	: Output matrix	
L_f	: Lie derivative along the vector field f	
m_s	: Maintenance coefficient	[/h]
n	: Expansion index	
n	: Number of data in moving window regression	
N	: Number of weirs	
N	: Number of measured data	
O	: Observability matrix	
OUR	: Oxygen uptake rate	[mgO_2/l.min]
OTR	: Oxygen transfer rate	[mgO_2/l.min]
p_i	: Number of parameters of model i	
p_L	: Liquid phase pressure	[N/m^2]
p_S	: Solid phase pressure	[N/m^2]
P	: Product concentration	[g/l]
Q	: Gaseous outflow rate (vector)	[g/l/h]
Q_i	: Weighting matrix	
$r(f)$	: Reliability of an inflection point f	

$r_\epsilon(\tau)$	: Covariance with lag τ	
R	: Ideal gas constant	[J/mol/K]
R	: Riccati matrix	
s	: Laplace variable	
s^2	: Residual mean square	
S	: Substrate concentration	[g/l]
S_O	: Oxygen concentration	[g/l]
S_O^*	: Oxygen saturation concentration	[g/l]
t	: Time	[min]
T	: Sampling period	[min]
T	: Temperature	[K]
t_{puls}	: Time of pulse addition	[min]
u	: Fluid superficial velocity	[m/h]
$u(t)$	: Input vector	
u_0	: Fluid superficial velocity in absence of solid particles	[m/h]
u_S	: Particles' velocity	[m/h]
U_T	: Terminal settling velocity of the particles	[m/h]
V	: Parameter estimation covariance matrix	
V	: Volume	[m^3]
W	: Weighing matrix (Kalman observer)	
w_D	: Weight attributed to model discrimination	
w_E	: Weight attributed to parameter estimation	
x	: State vector	
X	: Concentration of biomass	[g/l]
X_d	: Concentration of dead/detached biomass	[g/l]
Y	: Yield coefficient (matrix)	
y	: Measurement vector	
$\hat{y}_i(\theta)$	: Model prediction vector for model i	
z	: Space	[m]

B.1 Greek Letters

α_i	: Reaction order	
δ	: Small parameter (singular perturbation)	
ϵ	: Void fraction	
λ	: Local state isomorphism (structural identifiability)	
λ_{max}	: Largest eigenvalue of F (in absolute values)	
λ	: Eigenvalue	
λ_{min}	: Smallest eigenvalue of F (in absolute values)	
Γ	: Gain matrix (observer-based estimator)	
μ_{max}	: Maximum specific growth rate	[/h]
μ	: Specific growth rate	[/h]

ν	: Specific production rate	[/h]
ν	: Settling velocity	[/h]
ν	: Maximum settling velocity	[/h]
η	: Non-growth associated specific production rate	[/h]
ρ	: Reaction rate	[/h]
ρ_L	: Liquid phase density	[g/m^3]
ρ_S	: Solid phase density	[g/m^3]
ϕ	: Conversion rate	[/h]
Ψ	: State transformation (state obsever and observer-based estimator)	
$\sigma(\theta_i)$	: Standard error of parameter i	
θ_i	: Parameter i	
χ_i	: Experimental conditions for experiment i	
ξ	: Vector of the bioprocess component concentrations	[g/l]
Ω	: Gain matrix (observer-based estimator)	

B.2 Indices

B, A	: Autotrophic bacteria
B, H	: Heterotrophic bacteria
end	: Endogenous
ex	: Exogenous
in	: Influent
min	: Minimum
max	: Maximum
ND	: Organic nitrogen
NH	: Ammonia nitrogen
NO	: Nitrate nitrogen
out	: Effluent
R	: Recycle
W	: Waste

B.3 Abbreviations

AIC	: Akaike's Information Criterion
$BFGS$	: Broyden, Fletcher, Goldfarb and Shanno
$CSTR$	: Continuous stirred tank reactor
det	: Determinant (of a matrix)
DFP	: Davidson-Fletcher-Powell
DUD	: doesn't use derivatives
GA	: Genetic algorithm
GIC	: General Information Criterion

$iidN$	: independent and identically distributed normally
MAP	: Maximum a posteriori
$MCSM$	: Monte Carlo set-membership
ODE	: Ordinary differential equation
OED/PE	: Optimal experimental design for parameter estimation
OED/SC	: Optimal experimental design for structure characterization
OLS	: Ordinary least squares
PDE	: Partial differential equation
PDF	: Probability density function
PE	: Parameter estimation
$PRBS$	: Pseudo random binary signal
RSM	: Response surface methodology
SC	: Structure characterization
SCE	: Shuffled complex evolution
SSR_i	: Sum of squared residuals of model i
tr	: Trace (of a matrix)
WLS	: Weighted least squares
$WWTP$	: Wastewater treatment process (or plant)

References

[1] U. Abeling and C.F. Seyfried C.F. Anaerobic-aerobic treatment of high-strength ammonium wastewater - nitrogen removal via nitrite. *Wat. Sci. Tech.*, 26(5-6):1007–1015, 1992.

[2] S. Aborhey and D. Williamson. State and parameter estimation of microbial gowth processes. *Automatica*, 14:493–498, 1978.

[3] P. Agrawal and H.C. Lim. Analyses of various control schemes for continuous bioreactors. *Adv. Biochem. Eng./Biotechnol., A. Fiechter (Ed.)*, 30:61–90, 1984.

[4] S. Aiba, M. Shoda, and M. Nagatani. Kinetics of product inhibition in alcohol fermentation. *Biotechnol. Bioeng.*, 10:40–50, 1968.

[5] A. Aivasidis, H. Hochscherf, G. Rottmann, T. Hagen, M.T. Mertens, G. Reiners, and C. Wandrey. Neuere konzepte zur prozessaberwachung und -regelung bei der biologischen stickstoffelimination. *AWT*, 5:48–55, 1992.

[6] H. Akaike. A new look at the statistical model identification. *IEEE Trans. Aut. Control*, 19:716–723, 1974.

[7] J.G. Alvarez and J.G. Alvarez. Analysis and control of fermentation processes by optimal and geometric methods. *Proc. ACC*, 2:1112–1117, 1988.

[8] J.F. Andrews. A mathematical model for the continuous culture of microorganisms utilizing inhibiting substrates. *Biotechnol. Bioeng.*, 10:707–723, 1968.

[9] G.L. Atkins. The use of non-parametric methods for fitting equations to biological data. In G.C. Vansteenkiste and P.C. Young, editors, *Modelling and Data Analysis in Biotechnology and Medical Engineering*, pages 209–217. North-Holland, Amsterdam, 1983.

[10] D. Baetens, L.H. Hosten, B. Petersen, S. Van Volsem, and P.A. Vanrolleghem. Optimal experimental design for the calibration of models of phosphorous removing activated sludge systems. *Journal A*, 41(3):65–73, 2000.

[11] B. Balmelle, K.M. Nguyen, B. Capdeville, J.C. Cornier, and A. Deguin. Study of factors controlling nitrite build-up in biological processes for water nitrification. *Wat. Sci. Tech.*, 26(5-6):1017–1025, 1992.

[12] M. Baltes, R. Schneider, C. Sturm, and M. Reuss. Optimal experimental design for parameter estimation in unstructured growth models. *Biotechnol. Prog.*, 10:480–488, 1994.

[13] Y. Bard. *Nonlinear Parameter Estimation*. Academic Press, London, 1974.

[14] G. Bastin and D. Dochain. *On-line Estimation and Adaptive Control of Bioreactors*. Elsevier, Amsterdam, 1990.

[15] G. Bastin, D. Dochain, M. Haest, M. Installé, and Ph. Opdenacker. Modelling and adaptive control of a continuous anaerobic fermentation process. In A. Halme, editor, *Modelling and Control of Biotechnical Processes*, pages 299–306. Pergamon Press, Oxford, 1982.

[16] G. Bastin and J. Levine. On state reachability of reaction systems. *Proc. 29th CDC*, pages 2819–2824, 1990.

[17] D.M. Bates and D.G. Watts. Relative curvature measures of nonlinearity. *J. Roy. Stat. Soc.*, B42:1–25, 1980.

[18] E.M.L. Beale. Confidence regions in nonlinear estimation. *J. Roy. Stat. Soc.*, B22:41–88, 1960.

[19] J.V. Beck and K.J. Arnold. *Parameter Estimation in Engineering and Science*. John Wiley, New York, 1977.

[20] M.B. Beck. System identification and control. In G.G. Patry and D. Chapman, editors, *Dynamical Modelling and Expert Systems in Wastewater Engineering*, pages 261–323. Lewis Publishers, Chelsea, Michigan, 1989.

[21] M.B. Beck. Uncertainty, system identification and the prediction of water quality. In M.B. Beck and G. van Straten, editors, *Uncertainty and Forecasting of Water Quality*, pages 3–68. Springer Verlag, Berlin, 1983.

[22] M.B. Beck. Identification, estimation and control of biological waste-water treatment processes. *IEE Proc. (Part D)*, 133(5):254–264, 1986.

[23] M.B. Beck. Water quality modeling: A review of the analysis of uncertainty. *Water Res. Res.*, 23:1393–1442, 1987.

[24] M.B. Beck and P.C. Young. An introduction to system identification, parameter and state estimation. *Computer Appl. in Ferment. Technol.: Modelling and Control Biotechnol. Proc., Fish N., Fox R. and Thornhill N. (Eds)*, Elsevier, London:129–158, 1989.

[25] C.M. Bender and S.A. Orszag. *Advanced Mathematical Methods for Scientists and Engineers*. McGraw-Hill, Auckland, 1984.

[26] O. Bernard, Z. Hadj-Sadok, and D. Dochain. Advanced monitoring and control of anaerobic treatment plants: II - Dynamical model development and

identification. *Proc. Watermatex 2000*, 3.57–3.64, 2000.

[27] O. Bernard, Z. Hadj-Sadok, and D. Dochain. Dynamical model development and parameter identification for an anaerobic wastewater treatment process. *Biotechnol. Bioeng.*, to appear, 2001.

[28] K. Beven and A. Binley. The future of distributed models: Model calibration and uncertainty prediction. *Hydrol. Proc.*, 6:279–298, 1992.

[29] R. Binot, T. Bol, H. Naveau, and E.J. Nyns. Biomethanation by immobilized fluidised cells. *Wat. Sci. Tech.*, 15:103–115, 1983.

[30] B. Bonnet, D. Dochain, and J.Ph. Steyer. Dynamical modelling of an anaerobic digestion fluidised bed reactor. *Wat. Sci. Tech.*, 36(5):285–292, 1997.

[31] C. Bonvillain, D. Benyamina, M. Schaegger, A. Pauss, O. Bernard, and D. Dochain. Modelling of an extensive wastewater treatment plant (lagoon), based on two-year intensive follow-up. *Proc. Advanced Wastewater Treatment, Recycling and Reuse (AWT98)*, pages 535–541, 1998.

[32] B. Bouaziz and D. Dochain. Control analysis of fixed bed reactors: a singular perturbation approach. *Proc. ECC'93*, pages 1741–1745, 1993.

[33] J.B. Boucquey, P. Renard, P. Amerlynck, P. Modesto Filho, S.N. Agathos, H.P. Naveau, and E.J. Nyns. High-rate continuous biodegradation of concentrated chlorinated aliphatics by a durable enrichment of methanogenic origin under carrier-dependent conditions. *Biotechnol. Bioeng.*, 47:298–307, 1995.

[34] S. Bourrel. *Estimation et Commande d'un Procédé à Paramètres Répartis Utilisé pour le Traitement Biologique de l'Eau à Potabiliser*. PhD thesis, Université Paul Sabatier, Toulouse, France, 1996.

[35] S. V. Bourrel, J.P. Babary, S. Julien, M.T. Nihtila, and D. Dochain. Modelling and identification of a fixed-bed denitrification bioreactor. *Syst. Anal. Model. Simul.*, 30:289–309, 1998.

[36] S. V. Bourrel, D. Dochain, J.P. Babary, and I. Queinnec. Modelling, identification and control of a denitrifying biofilter. *J. Process Control*, 10:73–91, 2000.

[37] J.R. Bowen, A. Acrivos, and A.K. Oppenheim. Singular perturbation refinement to quasi-steady-state approximation in chemical kinetics. *Chem. Eng. Sci.*, 18:177–181, 1963.

[38] G. Box and G. Jenkins. *Time Series Analysis, Forecasting and Control*. Holden-Day, 1976.

[39] G.E.P. Box and W.J. Hill. Discrimination among mechanistic models. *Technometrics*, 9:57–71, 1967.

[40] S.A. Boyd and D.R. Shelton. Anaerobic biodegradation of chlorophenols in fresh and acclimated sludge. *Appl. Environ. Microbiol.*, 47:272–277, 1984.

[41] W.C. Boyle and P.M. Berthouex. Biological wastewater treatment model building: fits and misfits. *Biotechnol. Bioeng.*, 16:1139–1159, 1974.

[42] R.P. Brent. *Algorithms for Minimization without Derivatives*. Prentice-Hall, Englewood Cliffs, NJ, 1973.

[43] G.E. Briggs and J.B.S. Haldane. *Biochem. J.*, 242:3973, 1925.

[44] H. Brouwer, A. Klapwijk, and K.J. Keesman. Identification of activated sludge and wastewater characteristics using respirometric batch experiments. *Wat. Res.*, 32:1240–1254, 1998.

[45] C.G. Broyden. A new method of solving nonlinear simultaneous equations. *Comput. J.*, 12:94–99, 1969.

[46] K.F. Cacossa and D.A. Vaccari. Calibration of a compressive gravity thickening model from a single batch settling curve. *Wat. Sci. Tech.*, 30(8):107–116, 1994.

[47] G. Caminal, F.J. Lafuente, J. Lopez-Santin, M. Poch, and C. Sola. Application of the extended Kalman filter to identification of enzymatic deactivation. *Biotechnol. Bioeng.*, 24:366–369, 1987.

[48] C.W. Carroll. The created response surface technique for optimizing nonlinear, restrained, systems. *Operations Res.*, 9:169–184, 1961.

[49] J. Carstensen, P. Harremoës, and R. Strube. Software sensors based on the grey-box modelling approach. *Wat. Sci. Tech.*, 33(1):117–126, 1996.

[50] J. Carstensen, H. Madsen, N.K. Poulsen, and M.K. Nielsen. Identification of wastewater treatment processes for nutrient removal on a full-scale WWTP by statistical methods. *Wat. Res*, 28:2055–2066, 1994.

[51] J. Carstensen, P.A. Vanrolleghem, W. Rauch, and P. Reichert. Terminology and methodology in modelling for water quality management - a review. *Wat. Sci. Tech.*, 36(5):157–168, 1997.

[52] J.L. Casti. *Dynamical Systems and Their Application - Linear Theory*. Academic Press, New York, 1977.

[53] H. Caswell. The validation problem. In B.C. Patten, editor, *System Analysis and Simulation in Ecology, Vol.IV*, pages 313–325. Academic Press, New York, 1976.

[54] H.T. Chang and B.E. Rittmann. Predicting bed dynamics in three-phase, fluidized-bed biofilm reactors. *Wat. Sci. Tech.*, 29(10-11):231–241, 1994.

[55] M.J. Chappell, K.R. Godfrey, and S. Vajda. Global identifiability of the parameters of nonlinear systems with specified inputs: a comparison of methods. *Math. Biosci.*, 102:41–73, 1990.

[56] L. Chen. *Modelling, Identifiability and Control of Complex Biotechnological Processes*. PhD thesis, Université Catholique de Louvain, Belgium, 1992.

[57] L. Chen and G. Bastin. Structural identifiability of the yield coefficients in bioprocess models when the reaction rates are unknown. *Math. Biosci.*, 132:35–67, 1996.

[58] Y.H. Chen. Adaptive robust observers for non-linear uncertain systems. *Int. J. Systems Sci.*, 21:803–814, 1990.

[59] S.H. Cho, F. Colin, M. Sardin, and C. Prost. Settling velocity model of activated sludge. *Wat. Res.*, 27:1237–1242, 1993.

[60] C.K. Chui and G. Chen. *Linear Systems and Optimal Control*. Springer Verlag,

Berlin, 1988.

[61] D. Contois. Kinetics of bacterial growth relationship between population density and specific growth rate of continuous culture. *J. Gen. Microbiol.*, 21:845–864, 1959.

[62] A. Cornish-Bowden. Parameter estimating procedures for the michaelis-menten model: Reply to Tseng and Hsu. *J. Theor. Biol.*, 153:437–330, 1991.

[63] P.V. Danckwerts. Continuous flow systems. Distribution of residence times. *Chem. Eng. Sci.*, 2:1–13, 1953.

[64] B. De Clercq, F. Coen, B. Vanderhaegen and P.A. Vanrolleghem. Calibrating simple models for mixing and flow propagation in waste water treatment plants. *Wat. Sci. Tech.*, 39(4):61–69, 1999.

[65] J.G. de Gooijer, B. Abraham, A. Gould, and L. Robinson. Methods for determining the order of an autoregressive-moving average process: A survey. *Int. Statist. Rev.*, 53:301–329, 1985.

[66] P. de Larminat and Y. Thomas. *Automatique des Systèmes Linéaires. 2. Identification*. Flammarion Sciences, Paris, 1977.

[67] D. Defour, D. Derycke, J. Liessens, and P. Pipyn. Field experience with different systems for biomass accumulation in anaerobic reactor technology. *Wat. Sci. Tech.*, 30(12):181–191, 1994.

[68] B. De heyder, P.A. Vanrolleghem, H. Van Langenhove, and W. Verstraete. Kinetic characterisation of mass transfer limited degradation of a poorly water soluble gas in batch experiments. *Biotechnol. Bioeng.*, 55:511–519, 1997.

[69] R.I. Dick and B.B. Ewing. Evaluation of the activated sludge thickening theories. *J. Sanit. Engng. Div. Am. Soc. Civ. Eng.*, 93:9–29, 1967.

[70] D. Dochain. *On-line Parameter Estimation, Adaptive State Estimation and Adaptive Control of Fermentation Processes*. PhD. Thesis, Université Catholique de Louvain, Belgium, 1986.

[71] D. Dochain and L. Chen. Local observability and controllability of stirred tank reactors. *J. Process Control*, 2:139–144, 1992.

[72] D. Dochain and B. Bouaziz. Approximation of the dynamical model of fixed bed reactors via a singular perturbation approach. *Proc. IMACS Int. Symp. MIM-S2'93*, pages 34–39, 1993.

[73] D. Dochain. *Contribution to the Analysis and Control of Distributed Parameter Systems with Application to (Bio)chemical Processes and Robotics*. Thèse d'Aggrégation de l'Enseignement Supérieur, Université Catholique de Louvain, Belgium, 1994.

[74] D. Dochain, P.A. Vanrolleghem, and M. Van Daele. Structural identifiability of biokinetic models of activated sludge respiration. *Wat. Res.*, 29:2571–2578, 1995.

[75] D. Dochain. Modelling the dynamics of settlers: a basic problem. *Proc. FAB'97, Med. Fac. Landbouww. Univ. Gent*, 62:1617–1624, 1997.

[76] D. Dochain, S. Agathos, and P. Vanrolleghem. Asymptotic observers as a tool for

modelling process dynamics. *Wat. Sci. Tech.*, 36(5):259–268, 1997.

[77] D. Dochain and M. Pengov. State observers for processes with uncertain kinetics. *Proc. ADCHEM2000*, pages 171–176, 2000.

[78] D. Dochain. *Automatique des Bioprocédés*. Traité IC2, Section Systèmes Automatisés, Hermès, Paris, 2001.

[79] D. Draper. Assessment and propagation of model uncertainty (with discussion). *J. Roy. Stat. Soc.*, B57(1):45–97, 1995.

[80] N.R. Draper and H. Smith. *Applied Regression Analysis*. John Wiley, New York, 1981.

[81] Q. Duan, S. Sorooshian, and V. Gupta. Effective and efficient global optimisation for conceptual rainfall-runoff models. *Water Res. Res.*, 28:1015–1031, 1992.

[82] F. Ehlinger, Y. Escoffier, J.P. Couderc, J.P. Leyris, and R. Moletta. Development of an automatic control system for monitoring an anaerobic fluidized-bed. *Wat. Sci. Tech.*, 29(10-11):289–295, 1994.

[83] G.A. Ekama, J.L. Barnard, F.W. Günthert, P. Krebs, J.A. McCorquodale, D.S. Parker, and E.J. Wahlberg. *Secondary Settling Tanks: Theory, Modelling, Design and Operation*. IWA (Scientific and Technical Report n6), London, 1997.

[84] B. Eramo, R. Gavasci, A. Misiti, and P. Viotti. Validation of a multisubstrate mathematical model for the simulation of the denitrification process in fluidized bed biofilm reactors. *Wat. Sci. Tech.*, 29(10-11):401–408, 1994.

[85] S. Feyo de Azevedo, M.A. Romero-Ogawa, and A.P. Wardle. Modelling of tubular fixed-bed catalytic reactors: a brief review. *Trans. I.Chem.E.*, 68(Part A):2–8, 1990.

[86] M. Fjeld, O.A. Asbjornsen, and K.J. Aström. Reaction invariants and their importance in the analysis of eigenvectors, state observability and controllability of the continuous stirred tank reactor. *Chem. Eng. Sci.*, 29:1917–1926, 1974.

[87] R. Fletcher. *Practical Methods of Optimization*. John Wiley, New York, 1987.

[88] H.S. Fogler. *Elements of Chemical Reaction Engineering*. Prentice-Hall, Englewood Cliffs, NJ, 1986.

[89] P.U. Foscolo and L.G. Gibilaro. Fluid dynamic stability of fluidised suspensions: the particle bed model. *Chem. Eng. Sci.*, 42:1489–1500, 1987.

[90] G.F. Franklin, J.D. Powell, and A. Emami-Naeini. *Feedback Control of Dynamic Systems (3rd ed.)*. Addison-Wesley, Reading, Mass., 1995.

[91] A.G. Frederickson and H.M. Tsuchiya. Microbial kinetics and dynamics. In L. Lapidus and N.R. Amundson, editors, *Chemical Reactor Theory*, chapter 7. Prentice-Hall, Englewood Cliffs, NJ, 1977.

[92] J.L. Fripiat, T. Bol, R. Binot, H. Naveau, and E.J. Nyns. A strategy for the evaluation of methane production from different types of substrate biomass. *In R. Buvet, M.F. Fox and D.J. Picker (eds.), Biomethane, Production and Uses*,

Roger Bowskill Ltd, Exeter, UK:95–105, 1984.

[93] G.F. Froment and K.B. Bischoff. *Chemical Reactor Analysis and Design*. John Wiley, New York, 1990.

[94] M.J. Fuente, C. De Prada, and P. Vega. Optimization tools in a continuous dynamical simulation language. In W. Krug and A. Lehmann, editors, *Simulation and AI in Computer-aided Techniques*, pages 475–479. Society for Computer Simulation, San Diego, 1992.

[95] G.R. Gavalas. *Nonlinear Differential Equations of Chemically Reacting Systems*. Springer Verlag, Berlin, 1968.

[96] K.R. Gegenfurtner. Praxis: Brent's algorithm for function minimization. *Behavior Research Methods, Instruments & Computers*, 24:560–564, 1992.

[97] C. Georgakis, R. Aris, and R. Amundson. Studies in the control of tubular reactors - I. General considerations. *Chem. Eng. Sci.*, 32:1359–1369, 1977.

[98] K. Gernaey, P. Vanrolleghem, and W. Verstraete. On-line estimation of kinetic parameters of NH_4^+ oxidizing bacteria in activated sludge samples using titration in-sensor-experiments. *Wat. Res.*, 32:71–80, 1998.

[99] K.R. Godfrey and J.J. Di Stefano III. Identifiability of model parameters. In P. Young, editor, *Identification and System Parameter Estimation*, pages 89–114. Pergamon Press, Oxford, 1985.

[100] K.R. Godfrey and J.J. Di Stefano III. Identifiability of model parameters. In E. Walter, editor, *Identifiability of Parametric Models*, pages 1–19. Pergamon Press, Oxford, 1987.

[101] D.E. Goldberg. *Genetic algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Menlo Park, California, 1989.

[102] R.F. Goncalves, L. Le Grand, and T. Rogalla. Biological phosphorus uptake in submerged biofilters with nitrogen removal. *Wat. Sci. Tech.*, 29(10-11):135–143, 1994.

[103] G.C. Goodwin. Identification: Experiment design. In M. Singh, editor, *Systems and Control Encyclopedia*, Volume 4, pages 2257–2264. Pergamon Press, Oxford, 1987.

[104] G.C. Goodwin, M. Gevers, and B. Ninness. Quantifying the error in estimated transfer functions with application to model order selection. *IEEE Trans. Autom. Control*, 37:913–928, 1992.

[105] L.A. Gould. *Chemical Process Control: Theory and Applications*. Addison-Wesley, Reading, Mass., 1969.

[106] S. Grégoire, D. Dochain, A. Pauss, and M. Schaegger. Identification of a dynamical model of a waste stabilisation pond. *Proc. Watermatex 2000*, pages 1.51–1.58, 2000.

[107] J.B.S. Haldane. *Enzymes*. Longmans, London, 1930.

[108] R.P. Hamalainen, A. Halme, and A. Gyllenberg. A control model for activated sludge wastewater treatment process. *Proc. 6^{th} IFAC World Congress, Boston, Paper 61:6*, 1975.

[109] J. Hamilton, R. Jain, P. Antoniou, S.A. Svoronos, B. Koopman, and G. Lyberatos. Modeling and pilot-scale verification for predenitrification process. *J. Environ. Eng.*, 118:38–55, 1992.

[110] E.J. Hannan. The estimation of the order of an ARMA process. *Ann. Statist.*, 8:1071–1081, 1980.

[111] P. Harremoës, A.G. Capodaglio, B.G. Hellstrom, M. Henze, K.N. Jensen, A. Lynggaard-Jensen, R. Otterpohl, and H. Soeberg. Wastewater treatment plants under transient loading - performance, modelling and control. *Wat. Sci. Tech.*, 27(12):71–115, 1993.

[112] P. Harremoës and J. Carstensen. Deterministic versus stochastic interpretation of continuously monitored sewer systems. *European Water Pollution Control*, 4(5):117–126, 1994.

[113] A. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Reading, Mass., 1989.

[114] F.G. Heineken, H.M. Tsuchiya, and R. Aris. On the mathematical status of the pseudo-steady state hypothesis of biochemical kinetics. *Math. Biosci.*, 1:95–113, 1967.

[115] A. Heitzer, H.-P..E. Kohler, Reichert P., and G. Hamer. Utility of phenomenological models for describing temperature dependence of bacterial growth. *Appl. Environ. Microbiol.*, 57:2656–2665, 1991.

[116] C. Hellinga, P.A. Vanrolleghem, M.C.M. van Loosdrecht, and J.J. Heijnen. The potentials of off-gas analysis for monitoring and control of waste water treatment plants. *Wat. Sci. Tech.*, 33(1):13–23, 1996.

[117] C. Hellinga, M.C.M. van Loosdrecht and J.J. Heijnen. Model based design of a novel process for ammonia removal from concentrated flows. *Math. Comp. Mod.Dyn. Syst.*, 5:351–371, 1999.

[118] M. Henze. Characterization of wastewater for modelling of activated sludge processes. *Wat. Sci. Tech.*, 25(6):1–15, 1992.

[119] M. Henze, R. Dupont, P. Grau, and A. de la Sota. Rising sludge in secondary settlers due to denitrification. *Wat. Res.*, 27:231–236, 1993.

[120] M. Henze, C.P.L. Grady Jr, W. Gujer, G.v.R. Marais, and T. Matsuo. Activated sludge model no. 1. Technical report, IAWPRC, Scientific and Technical Reports No. 1, 1986.

[121] M. Henze, W. Mino, W. Gujer, T. Matsuo, M.C. Wentzel, G.v.R. Marais, and T. Matsuo. Activated sludge model no. 2. Technical report, IAWQ Scientific and Technical Reports No. 3, IAWQ, London, 1995.

[122] H. Hertz, A. Krogh, and R. Palmer. *Introduction to the Theory of Neural Computation*. Addison-Wesley, Reading, Mass., 1991.

[123] P.D.H. Hill. A review of experimental design procedures for regression model discrimination. *Technometrics*, 20:15–21, 1978.

[124] W.J. Hill, W.G. Hunter, and D.W. Wichern. A joint design criterion for the dual problem of model discrimination and parameter estimation. *Technometrics*,

10:145–160, 1968.

[125] C. Holliger, A.J.M. Stams, and A.J.B. Zehnder. Anaerobic degradation of recalcitrant compounds. *Anaerobic Digestion, E.R. Hall and P.N. Hobson (eds)*. Pergamon Press, Oxford, 1988.

[126] A. Holmberg. On the practical identifiability of microbial growth models incorporating Michaelis-Menten type nonlinearities. *Math. Biosci.*, 62:23–43, 1982.

[127] A. Holmberg and J. Ranta. Procedures for parameter and state estimation of microbial growth process models. *Automatica*, 18:181–193, 1982.

[128] J. Holst, U. Holst, H. Madsen, and H. Melgaard. Validation of grey box models. *Proc. IFAC Symposium on Adaptive Control and Signal Processing*, pages 407–414, 1992.

[129] G.M. Hornberger and R.C. Spear. An approach to the analysis of behavior and sensitivity in environmental systems. In M.B. Beck and G. van Straten, editors, *Uncertainty and Forecasting of Water Quality*, pages 101–116. Springer-Verlag, Heidelberg, 1983.

[130] W.G. Hunter and A.M. Reiner. Designs for discriminating between two rival models. *Technometrics*, 7:307–323, 1965.

[131] K.F. Janning, P. Harremoës, and M. Nielsen. Evaluating and modelling the kinetics in a full scale submerged denitrification filter. *Wat. Sci. Tech.*, 32(8):115–123, 1995.

[132] H.H. Jean and L.H. Fan. On the model equations of Gibilaro and Foscolo with corrected buoyancy force. *Powder Technology*, 72:201–205, 1992.

[133] U. Jeppsson. *Modelling Aspects of Wastewater Treatment Processes*. PhD thesis, Lund Institute of Technology, Sweden, 1996.

[134] U. Jeppsson and G. Olsson. Reduced order models for on-line parameter identification of the activated sludge process. *Wat. Sci. Tech.*, 28(11-12):173–183, 1993.

[135] D.B. Johnson and P.M. Berthouex. Using multiresponse data to estimate biokinetic parameters. *Biotechnol. Bioeng.*, 17:571–583, 1975.

[136] S.B. Jorgensen. Fixed bed reactor dynamics and control - A review. *Proc. IFAC Control of Distillation Columns and Chemical Reactors*. Pergamon, pages 11–24, 1986.

[137] S.E. Jørgensen. *Integration of Ecosystem Theories: A Pattern*. Kluwer, Dordrecht, 1992.

[138] S. Julien, J.P. Babary, and P. Lessard. Theoretical and practical identifiability of a reduced order model in an activated sludge process doing nitrification and denitrification. *Wat. Sci. Tech.*, 37(12):309–368, 1998.

[139] S. Julien, J.P. Babary, and P. Lessard. Identifiability and identification of an activated sludge process model. *Syst. Anal. Model. Sim.*, 37:481–499, 2000.

[140] R.L. Kashyap. Inconsistency of the AIC rule for estimating the order of autoregressive models. *IEEE Trans. Autom. Control*, 25:996–998, 1980.

[141] K.J. Keesman. Membership-set estimation using random-scanning and principal component analysis. *Math. Comp. Simul.*, 32:535–543, 1990.

[142] K.J. Keesman, H. Spanjers, and G. van Straten. Analysis of endogenous process behavior in activated sludge. *Biotechnol. Bioeng.*, 57:155–163, 1998.

[143] K.J. Keesman and G. van Straten. Modified set theoretic identification of an ill-defined water quality system from poor data. In B. Beck, editor, *System Analysis in Water Quality Management*, pages 297–308. Pergamon Press, Oxford, 1987.

[144] K.J. Keesman and G. van Straten. Set-membership approach to identification and prediction of lake eutrophication. *Water Res. Res.*, 26:2643–2652, 1990.

[145] R. Khanna and J.H. Seinfeld. Mathematical modelling of packed bed reactors: numerical solutions and control model development. *Adv. Chem. Eng.*, 13:113–191, 1987.

[146] A.B. Koehler and E.S. Murphree. A comparison of the Akaike and Schwarz criteria for selecting model order. *Appl. Statist.*, 37:187–195, 1988.

[147] H.-P. Kohler, A. Heitzer, and G. Hamer. Improved unstructured model describing temperature dependence of bacterial maximum specific growth rates. *Proceedings EERO/EFB International Symposium Environmental Biotechnology ISEB. Ostend, Belgium, April 22-25 1991*, Part II:511–514, 1991.

[148] P. Kokotovic, H.K. Khalil, and J. O'Reilly. *Singular Perturbation Methods in Control: Analysis and Design*. Academic Press, London, 1986.

[149] Z. Kong, P.A. Vanrolleghem, and W. Verstraete. Automated respiration inhibition kinetics analysis (ARIKA) with a respirographic biosensor. *Wat. Sci. Tech.*, 30(4):275–284, 1994.

[150] G. Kresseilmeier. Adaptive observers with exponential rate of convergence. *IEEE Trans. Autom. Control*, 22:2–8, 1977.

[151] H. Kwakernaak and R. Sivan. *Linear Optimal Control Systems*. John Wiley, New York, 1972.

[152] E.B. Lee and L. Markus. *Foundations of Optimal Control Theory*. John Wiley, New York, 1967.

[153] S.H. Lee, P. Tsobanakis, J.A. Phillips, and C. Georgakis. Issues in the optimization, estimation and control of fed-batch bioreactors using tendency models. *Proc. ICCAFT 5/IFAC-BIO 2, Keystone, Colorado*, 1992.

[154] T.T. Lee, F.Y. Wang, and R.B. Newell. Distributed parameter approach to the dynamics of complex biological processes. *AIChE J.*, 45:2245–2268, 1999.

[155] L. Lefèvre, D. Dochain, S. Feyo de Azevedo, and A. Magnus. Analysis of the orthogonal collocation method when applied to the numerical integration of chemical reactor models. *Computers and Chemical Engineering*, 24 (12):2571–2588, 2000.

[156] J.A. Lennox, Z. Yuan and J. Harmand. A systematic approach to error isolation in computerized wastewater simulation models. *Wat. Sci. Tech.*, 43(7):367-376, 2001.

[157] H. Linhart and W. Zucchini. *Model Selection*. John Wiley, New York, 1986.

[158] L. Ljung. Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems. *IEEE Trans. Autom. Control*, 24:36–50, 1979.

[159] L. Ljung. *System Identification - Theory for the User*. Prentice-Hall, Englewood Cliffs, NJ, 1999.

[160] J.R. Lobry. *Re-évaluation du Modèle de Croissance de Monod. Effet des Antibiotiques sur l'Energie de Maintenance*. PhD thesis, Université Claude Bernard, Lyon, France, 1991.

[161] J.R. Lobry and J.P. Flandrois. Comparison of estimates of Monod's growth model parameters from the same data set. *Binary*, 3:20–23, 1991.

[162] J.R. Lobry, L. Rosso, and J.P. Flandrois. A Fortran subroutine for the determination of parameter confidence limits in non-linear models. *Binary*, 3:86–93, 1991.

[163] A. Lübbert. Characterization of bioreactors. In K. Schügerl, editor, *Biotechnology, A Multi-Volume Comprehensive Treatise, Vol.4, Measuring, Modelling and Control*, pages 107–148. VCH, Weinheim, 1991.

[164] R. Luedeking and E. Piret. A kinetic study of the lactic acid fermentation batch process at controlled pH. *J. Bioch. Microb. Technol. Eng.*, I (4):393–412, 1959.

[165] I. Mareels and J.W. Polderman. *Adaptive Systems. An Introduction*. Birkhauser, Boston, 1996.

[166] A.T. Marino, J.J. Di Stefano III, and E.M. Landaw. Dimsum: An expert system for multiexponential model discrimination. *Am. J. Physiol.*, 262:E546–E556, 1992.

[167] R. Marino. Adaptive observers for single output nonlinear systems. *IEEE Trans. Autom. Control*, 35:1054–1058, 1990.

[168] D.W. Marquardt. An algorithm for least squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.*, 11:431–441, 1963.

[169] S. Marsili-Libelli. Optimal control of the activated sludge process. *Trans. Inst. Meas. Control*, 6:146–152, 1984.

[170] S. Marsili-Libelli. Modelling, identification and control of the activated sludge process. *Adv. Biochem. Eng./Biotechn.*, 38:90–148, 1989.

[171] S. Marsili-Libelli. Parameter estimation of ecological models. *Ecol. Modelling*, 62:233–258, 1992.

[172] H. Melcer, W.J. Parker, and B.E. Rittmann. Modeling of volatile organic contaminants in trickling filter systems. *Wat. Sci. Tech.*, 31(1):95–104, 1995.

[173] L. Michaelis and M.L. Menten. Die Kinetic der Invertinwirkung. *Biochemische Zeitschrift*, 49:334–369, 1913.

[174] M. Milanese and A. Vicino. Optimal estimation theory for dynamic systems with set membership uncertainty: an overview. *Automatica*, 27:997–1009, 1991.

[175] P. Modesto Filho, P. Amerlynck, E.J. Nyns, and H.P. Naveau. Acclimation of methanogenic consortium to polychlorinated compounds in a fixed-film sta-

tionary bed reactor. *Appl. Microb. Biot.*, 1992.

[176] J. Monod. *Recherches sur la Croissance des Cultures Bactériennes*. Hermann, Paris, 1942.

[177] A. Moser. *Bioprocess Technology*. Springer-Verlag, New York, 1988.

[178] F.E. Mosey. Mathematical modelling of the anaerobic digestion process: regulatory mechanisms for the formation of short-chain volatile acids from glucose. *Wat. Sci. Tech.*, 15(8/9):209–232, 1983.

[179] A. Munack. Optimal feeding strategy for identification of Monod-type models by fed-batch experiments. *Proc. Comp. Appl. in Ferm. Techn.: Mod. and Control of Biotech. Proc.*, pages 195–204, 1989.

[180] A. Munack. Optimization of sampling. In K. Schugerl, editor, *Biotechnology, a Multi-volume Comprehensive Treatise*, volume 4, pages 251–264. VCH, Weinheim, 1991.

[181] A. Munack. Some improvements in the identification of bioprocesses. In M.N. Karim and G. Stephanopoulos, editors, *Modelling and Control of Biotechnical Processes*, pages 89–94. Pergamon Press, Oxford, 1992.

[182] A. Munack and C. Posten. Design of optimal dynamical experiments for parameter estimation. *Proc. ACC*, pages 2010–2016, 1989.

[183] K.S. Narendra and A.M. Annaswamy. *Stable Adaptive Systems*. Prentice-Hall, Englewood Cliffs, NJ, 1989.

[184] J.A. Nelder and R. Mead. A simplex method for function minimization. *Comp. J.*, 7:308–313, 1964.

[185] J.P. Norton. *An Introduction to Identification*. Academic Press, London, 1986.

[186] R. Oliveira, E.C. Ferreira, F. Oliveira, and S. Feyo de Azevedo. A study on the convergence of observer-based kinetics estimators in stirred tank bioreactors. *J. Process Control*, 6:367–371, 1996.

[187] G. Olsson and B. Newell. *Wastewater Treatment Systems. Modelling, Diagnosis and Control*. IWA Publishing, London, 1999.

[188] G. Olsson and J.P. Stephenson. The propagation of hydraulic disturbances and flow rate reconstruction in activated sludge plants. *Env. Technol. Letters*, 6:536–545, 1985.

[189] R.V. O'Neill and R.H. Gardner. Sources of uncertainty in ecological models. In G.S. Innis and R.V. O'Neill, editors, *Methodology in Systems Modelling and Simulation*, pages 447–463. North-Holland, Amsterdam, 1979.

[190] S.L. Ong. A comparison of estimates of kinetic constants for a suspended growth treatment system from various linear transformations. *J. Water Pollut. Control Fed.*, 62:894–900, 1990.

[191] M. Perrier, S. Feyo de Azevedo, E.C. Ferreira, and D. Dochain. Tuning of observer-based estimators: theory and application to the on-line estimation of kinetic parameters. *Control Eng. Practice*, 8 (4):377–388, 2000.

[192] M. Perrier and D. Dochain. Evaluation of control strategies for anaerobic digestion processes. *Int. J. Adaptive Cont. Signal Proc.*, 7:309–321, 1993.

[193] B. Petersen. *Calibration, Identifiability and Optimal Experimental Design of Activated Sludge Models*. PhD thesis, Ghent University, Belgium, 2000.

[194] B. Petersen, K. Gernaey, and P.A. Vanrolleghem. Improved theoretical identifiability of model parameters by combined respirometric-titrimetric measurements. A generalization of results. *Proc. 3rd IMACS Symposium on Mathematical Modelling MATHMOD, I. Troch and F. Breitenecker (Eds)*, 2:639–642, 2000.

[195] B. Petersen, K. Gernaey, and P.A. Vanrolleghem. Practical identifiability of model parameters by combined respirometric-titrimetric measurements. *Wat. Sci. Tech.*, 43(7):347–356, 2001.

[196] E.E. Petersen. *Chemical Reaction Analysis*. Prentice-Hall, Englewood Cliffs, NJ, 1965.

[197] H. Pohjanpalo. System identifiability based on the power series expansion of the solution. *Math. Biosci.*, 41:21–33, 1978.

[198] Y. Pomerleau and M. Perrier. Estimation of multiple specific growth rates in bioprocesses. *AIChE J.*, 36:207–215, 1990.

[199] D. Poncelet, H. Naveau, E.J. Nyns, and D. Dochain. Transient response of a solid-liquid model biological fluidised bed to a step change in fluid superficial velocity. *J. Chem. Tech. Biotechnol.*, 48:439–452, 1990.

[200] K.R. Popper. *The Logic of Scientific Discovery*. Hutchinson, London, 1980.

[201] C. Posten and A. Munack. On-line application of parameter estimation accuracy to biotechnical processes. *Proc. ACC*, pages 2181–2186, 1990.

[202] M.J.D. Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Comp. J.*, 7:155–162, 1964.

[203] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge, UK, 1986.

[204] W.L. Price. A controlled random search procedure for global optimization. *Comp. J.*, 20:367–370, 1979.

[205] I. Queinnec and D. Dochain. Modelling and simulation of the steady-state of secondary settlers in wastewater treatment plants. *Wat. Sci. Tech.*, 43(7):39-46, 2001.

[206] A. Raksanyi, Y. Lecourtier, E. Walter, and A. Venot. Identifiability and distinguishability testing via computer algebra. *Math. Biosci.*, 77:245–266, 1985.

[207] M.L. Ralston and R.I. Jennrich. DUD, a derivative-free algorithm for nonlinear least squares. *Technometrics*, 20:7–14, 1978.

[208] D.A. Ratkowsky. *Nonlinear Regression Modeling - A Unified Practical Approach*. Marcel Dekker, Basel, Switzerland, 1983.

[209] D.A. Ratkowsky. A suitable parameterization of the Michaelis-Menten enzyme reaction. *Biochem. J.*, 240:357–360, 1986.

[210] D.A. Ratkowsky, T. Ross, T.A. McMeekin, and J. Olley. Comparison of Arrhenius-type and Belehradek-type models for prediction of bacterial

growth in foods. *J. Appl. Bacteriol.*, 71:452–459, 1991.

[211] W.H. Ray and D.G. Lainiotis. *Distributed Parameter Systems, Identification, Estimation and Control*. Marcel Dekker, New York, 1978.

[212] K.H. Reckhow and S.C. Chapra. Confirmation of water quality models. *Ecol. Modelling*, 20:113–133, 1983.

[213] P. Reichert. *Concepts Underlying a Computer Program for the Identification and Simulation of Aquatic Systems*. PhD thesis, Swiss Federal Institute of Environmental Science and Technology (EAWAG), Dübendorf, Switzerland, 1994.

[214] P. Reichert and M. Omlin. On the usefulness of overparameterized ecological models. *Ecol. Modelling*, 95:289–299, 1997.

[215] P. Reichert and P.A. Vanrolleghem. Identifiability and uncertainty analysis of the river water quality model no. 1 (RWQM1). *Wat. Sci. Tech.* 43(7):329–338, 2001.

[216] J.F. Richardson and W.N. Zaki. Sedimentation and fluidisation. Part I. *Trans. Inst. Chem. Eng.*, 32:35–53, 1954.

[217] R.G. Riefler, D.P. Ahlfeld, and B.F. Smets. Respirometric assay for biofilm kinetics estimation: Parameter identifiability and retrievability. *Biotechnol. Bioeng.*, 57:35–45, 1998.

[218] J.A. Robinson. Determining microbial parameters using nonlinear regression analysis: Advantages and limitations in microbial ecology. *Adv. Microb. Ecol.*, 8:61–114, 1985.

[219] M.R. Rose. *Quantitative Ecological Theory - An Introduction to Basic Models*. Croom Helm, London, 1987.

[220] H.H. Rosenbrock. An automated method for finding the greatest or least value of a function. *Comput. J.*, 3:175–184, 1960.

[221] P.J. Rousseeuw and A.M. Leroy. *Robust Regression and Outlier Detection*. John Wiley, New York, 1987.

[222] A. Rozzi. Modelling and control of anaerobic digestion processes. *Trans. Inst. Meas. Control*, 6:153–159, 1984.

[223] A. Rozzi. Anaerobic process control by bicarbonate monitoring. *Env. Technol. Letters*, 6:594–601, 1985.

[224] E. Sarner. Removal of dissolved and particulate organic matter in high-rate trickling filters. *Wat. Res.*, 15:671–678, 1981.

[225] K. Schmidt and S.H. Isaacs. An evolutionary algorithm for initial state and parameter estimation in complex biochemical models. In *Preprints 6th IFAC Conference on Computer Applications in Biotechnology - CAB6. Garmisch-Partenkirchen, Germany, May 14-17 1995*, pages 239–242. 1995.

[226] M. Schuetze. *Integrated simulation and optimum control of the urban wastewater system*. PhD thesis, Imperial College, London, UK, 1998.

[227] K. Schugerl. Biofluidization: application of the fluidization technique in biotechnology. *Can. J. Chem. Eng.*, 67:178–184, 1989.

[228] G. Schwarz. Estimating the dimension of a model. *Ann. Statist.*, 6:461–464, 1978.

[229] Z. Shi and K. Shimizu. Neuro-fuzzy control of bioreactor systems with pattern recognition. *J. Ferment. Bioeng.*, 74:39–45, 1992.

[230] H. Siegrist, P. Krebs, R. Bahler, I. Purtschert I., C. Rtck, and R. Rufer. Denitrification in secondary clarifiers. *Wat. Sci. Tech.*, 31(2):205–214, 1995.

[231] J. Singer and D.C.G. Lewis. *Water and Water Engineering*, pages 105–111, 1966.

[232] T. Söderström and P. Stoica. *Model validation and model structure determination*. Prentice-Hall, Englewood Cliffs, NJ, 1989.

[233] U. Sollfrank and W. Gujer. Characterisation of domestic wastewater for mathematical modelling of the activated sludge process. *Wat. Sci. Tech.*, 23(4-6):1057–1066, 1991.

[234] E.D. Sontag. *Mathematical Control Theory. Deterministic Finite Dimensional Systems (2nd ed.)*. Texts in Applied Mathematics, 6, Springer Verlag, New York, 1998.

[235] S. Sorooshian, H.V. Gupta, and L.A. Bastidas. Calibration of hydrologic models using multi-objectives and visualization techniques. Final report project EAR-9418147. Technical report, University of Arizona, Tucson, USA, 1998.

[236] H. Spanjers and K. Keesman. Identification of wastewater biodegradation kinetics. In *Proceedings 3rd IEEE Conference on Control Applications, Session on Environmental Process Control. Glasgow, UK, August 24-26 1994*, pages 1011–1016, 1994.

[237] W. Spendley, G.R. Hext, and F.R. Himswokin. Sequential application of simplex designs in optimisation and evolutionary operation. *Technometrics*, 4:441–461, 1962.

[238] J.A. Spriet. Structure characterization - An overview. In H.A. Barker and P.C. Young, editors, *Identification and System Parameter Estimation*, pages 749–756. Pergamon Press, Oxford, 1985.

[239] J.A. Spriet and P. Herman. Simulation study of structure characterisation methods. *IMACS 1983*, pages 452–459, 1983.

[240] J.A. Spriet and G.C. Vansteenkiste. *Computer-aided Modelling and Simulation*. Academic Press, London, 1982.

[241] H. Stehfest. An operational dynamic model for the final clarifier. *Trans. Inst. Meas. Control*, 6:160–164, 1984.

[242] G. Stephanopoulos and K.-Y. San. Studies on on-line bioreactor identification. *Biotechnol. Bioeng.*, 26:1176–1188, 1984.

[243] P.M. Sutton and P.N. Mishra. Activated carbon based biological fluidized beds for contaminated water and wastewater treatment: A state-of-the-art review. *Wat. Sci. Tech.*, 29(10-11):309–317, 1994.

[244] L. Szalai, P. Krebs, and W. Rodi. Simulation of flow in circular clarifiers with and without swirl. *J. Hydraulic Eng.*, 120:4–17, 1993.

[245] I. Takács, G.G. Patry, and D. Nolasco. A dynamic model of the clarification-thickening process. *Wat. Res.*, 25(10):1263–1271, 1991.

[246] H. te Braake, R. Babuska, E. van Can, and C. Hellinga. Predictive control in biotechnology using fuzzy and neural models. In J. VanImpe, P.A. Vanrolleghem, and D. Iserentant, editors, *Advanced Instrumentation, Data Interpretation and Control of Biotechnological Processes*. Kluwer, Dordrecht, 437–464, 1998.

[247] G. Tessier. Croissance des populations bactériennes et quantités d'aliments disponibles. *Rev. Sci.*, 80–209, 1942.

[248] M.L. Thompson and M.A. Kramer. Modelling chemical processes using prior knowledge and neural networks. *AIChE J.*, 40:1328–1340, 1994.

[249] L. Tijhuis, W.A.J. van Benthum, M.C.M. van Loosdrecht, and J.J. Heijnen. Solids retention time in spherical biofilms in a biofilm airlift suspension reactor. *Biotechnol. Bioeng.*, 44:867–879, 1994.

[250] H. Topiwala and C.G. Sinclair. Temperature relationship in continuous culture. *Biotechnol. Bioeng.*, 13:795–813, 1971.

[251] S. Tseng and J.-P. Hsu. A comparison of the parameter estimating procedures for the Michaelis-Menten model. *J. Theor. Biol.*, 143:457–464, 1990.

[252] H. Tulleken. *Grey-box Modelling and Identification Topics*. PhD thesis, Delft University of Technology, Holland, 1992.

[253] A. Urrutikoetxea and J.L. Garcia-Heras. Secondary settling in activated sludge: a lab-scale dynamic model of thickening. *Proc. FAB, Med. Fac. Landbouww. Univ. Gent*, 59/4a:2025–2036, 1994.

[254] S. Vajda, K.R. Godfrey, and H. Rabitz. Similarity transformation approach to identifiability analysis of nonlinear compartmental models. *Math. Biosci.*, 93:217–248, 1989.

[255] V. Van Breusegem and G. Bastin. Reduced order modelling of reaction systems: a singular perturbation approach. *Proc. 30th IEEE Conf. Decision and Control*, 1049–1054, 1992.

[256] H.L. Vangheluwe, F. Claeys, S. Kops, F. Coen, and G.C. Vansteenkiste. A modelling and simulation environment for waste-water treatment plant design. *Proc. European Simulation Symposium 1996*, Genoa, Italy:90–97, 1996.

[257] P. Van Overschee and B. De Moor. N4SID: Two subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, 30:75–93, 1994.

[258] P.A. Vanrolleghem, M. Van Daele, and D. Dochain. Practical identifiability of a biokinetic model of activated sludge respiration. *Wat. Res.*, 29:2561–2570, 1995.

[259] P.A. Vanrolleghem, M. Van Daele, P. Van Overschee, and G.C. Vansteenkiste. Model structure characterization of nonlinear wastewater treatment systems. *Proc. 10th IFAC Conf. Syst. Ident.*, 1:279–284, 1994.

[260] P.A. Vanrolleghem, A. Vanderhasselt, P. Krebs, and P. Reichert. Identification of a second-order one-dimensional clarifier model from on-line data. *Technical report, Ghent University*, 1997.

[261] P.A. Vanrolleghem and M. Van Daele. Optimal experimental design for structure characterization of biodegradation models: On-line implementation in a respirographic biosensor. *Wat. Sci. Tech.*, 30(4):243–253, 1994.

[262] P.A. Vanrolleghem and J.F. Van Impe. On the use of structured compartment models for the activated sludge biodegradation process. *Proceedings Workshop Modelling, Monitoring and Control of the Activated Sludge Process. Med. Fac. Landbouww. Rijksuniv. Gent*, 57:2215–2228, 1992.

[263] P.A. Vanrolleghem, J.F. Van Impe, J. Vandewalle, and W. Verstraete. Advanced monitoring and control of the activated sludge process: On-line estimation of crucial biological variables in a structured model with the RODTOX biosensor. In M.N. Karim and G. Stephanopoulos, editors, *Modelling and Control of Biotechnical Processes*, pages 355–358. Pergamon Press, Oxford, 1992.

[264] P.A. Vanrolleghem and K.J. Keesman. Identification of biodegradation models under model and data uncertainty. *Wat. Sci. Tech.*, 33(2):91–105, 1996.

[265] P.A. Vanrolleghem and W. Verstraete. On-line monitoring equipment for wastewater treatment processes: State of the art. *Proc. TI-KVIV Studiedag Optimalisatie van Waterzuiveringsinstallaties door Proceskontrole en -sturing*, pages 1–22, 1993.

[266] P.A. Vanrolleghem and W. Verstraete. Simultaneous biokinetic characterization of heterotrophic and nitrifying populations of activated sludge with an on-line respirographic biosensor. *Wat. Sci. Tech.*, 28(11-12):377–387, 1993.

[267] G.C. Vansteenkiste and J.A. Spriet. Modelling ill-defined systems. In F.E. Cellier, editor, *Progress in Modelling and Simulation*, pages 11–38. Academic Press, London, 1982.

[268] G. van Straten. Analytical methods for parameter-space delimitation and application to shallow-lake phytoplankton-dynamics modeling. *Appl. Math. Comput.*, 17:459–482, 1985.

[269] L. Van Vooren. *Buffer capacity based multipurpose hard- and software sensor for environmental applications*. PhD thesis, Ghent University, Belgium, 2000.

[270] A. Varma and R. Aris. Stirred pots and empty tubes. In L. Lapidus and N.R. Amundson, editors, *Chemical Reactor Theory: A Review.*, pages 79–154. Prentice-Hall, Englewood Cliffs, NJ, 1977.

[271] R. Verhulst. Notice sur la loi que la population suit dans son accroissement. *Corr. Math. et Phys., A. Quetelet (Ed.)*, t. X:113, 1838.

[272] L.K. Vermeersch, B. Kroes, and P.A. Vanrolleghem. Feature based model identification of nonlinear biotechnological processes. *Ecol. Modelling*, 75:629–640, 1994.

[273] K.J. Versyck, J.E. Claes, and J.F. Van Impe. Practical identification of un-

structured growth kinetics by application of optimal experimental design. *Biotechnol. Prog.*, 13:524–531, 1997.

[274] K.J. Versyck and J.F. Van Impe. Optimal design of system identification experiments for bioprocesses. *Journal A*, 41(2):25–34, 2000.

[275] C. Vialas. *Modélisation et Contribution à la Conception d'un Procédé Biotechnologique*. PhD thesis, ENSIEG, Grenoble, France, 1984.

[276] C. Vialas, A. Cheruy, and S. Gentil. An experimental approach to improve the Monod model identification. In A. Johnson, editor, *Modelling and Control of Biotechnological Processes*, pages 155–160. Pergamon, Oxford, 1986.

[277] J. Villadsen and M.L. Michelsen. *Solution of Differential Equation Models by Polynomial Approximation*. Prentice-Hall, Englewood Cliffs, NJ, 1978.

[278] Z.Z. Vitasovic. *An Integrated Control Strategy for the Activated Sludge Process*. PhD thesis, Rice University, Houston, USA, 1986.

[279] E. Walter. *Identifiability of State Space Models*. Springer-Verlag, Berlin, 1982.

[280] O. Wanner, J. Kappeler, and W. Gujer. Calibration of an activated sludge model based on human expertise and on a mathematical optimization technique - A comparison. *Wat. Sci. Tech.*, 25(6):141–148, 1992.

[281] R.W. Watts, S.A. Svoronos, and B. Koopman. One-dimensional modeling of secondary clarifiers using a concentration and feed velocity-dependent dispersion coefficient. *Wat. Res.*, 30:2112–2124, 1996.

[282] S.R. Weijers and P.A. Vanrolleghem. A procedure for selecting the most important parameters in calibrating the activated sludge model No.1 with full-scale plant data. *Wat. Sci. Tech.*, 36(5):69–79, 1997.

[283] T. Wik, A. Mattsson, E. Hansson, and C. Niklasson. Nitrification in a tertiary trickling filter at high hydraulic loads - pilot plant operation and mathematical modelling. *Wat. Sci. Tech.*, 32(8):185–192, 1995.

[284] J.L. Willems. *Stability Theory of Dynamical Systems*. Nelson, London, 1970.

[285] J. Winkin, D. Dochain, and Ph. Ligarius. Dynamical analysis of distributed parameter tubular reactors. *Automatica*, 36:349–361, 2000.

[286] L. Yang and J.E. Alleman. Investigation of batchwise nitrite build-up by an enriched nitrification culture. *Wat. Sci. Tech.*, 26(5-6):997–1005, 1992.

[287] Y.J. Yoo, J. Hong, and R.T. Hatch. Sequential estimation of states and kinetic parameters and optimization of fermentation processes. *Proc. ACC*, 2:866–871, 1985.

[288] P.C. Young. The validity of models for badly defined systems. In M.B. Beck and G. van Straten, editors, *Uncertainty and Forecasting of Water Quality*, pages 69–100. Springer-Verlag, Heidelberg, 1983.

[289] I. Zambettakis. *Simplification des Systèmes à Paramétres Distribués Multiéchelles. Application à la Commande*. PhD thesis, Université des Sciences et Techniques de Lille, France, 1987.

[290] E. Zauderer. *Partial Differential Equations of Applied Mathematics*. John Wiley, New York, 1989.

[291] M.H. Zwietering, J.T. De Koos, B.E. Hasenack, J.C. De Wit, and K. Van 'T Riet. Modelling of bacterial growth as a function of temperature. *Appl. Environ. Microbiol.*, 57:1094–1101, 1991.

Environmental quality is becoming an increasing concern in our society. In that context, waste and wastewater treatment, and more specifically biological wastewater treatment processes play an important role. *Dynamical Modelling and Estimation in Wastewater Treatment Processes* concentrates on the mathematical modelling of these processes. The main purpose is to provide the increasing number of professionals who are using models to design, optimise and control wastewater treatment processes with the necessary background for their activities of model building, selection and calibration.

Dynamical Modelling and Estimation in Wastewater Treatment Processes deals specifically with dynamic models because they allow us to describe the behaviour of treatment plants under the highly dynamic conditions that we want them to operate (for example Sequencing Batch Reactors) or we have to operate them (for example storm conditions and spills). Further extension is provided to reactor systems for which partial differential equation descriptions are necessary to account for their distributed parameter nature (settlers and fixed bed reactors for example).

The model building exercise is introduced as a step-wise activity that, in this book, starts from mass balancing principles. In many cases, different hypotheses and their corresponding models can be proposed for a particular process. It is therefore essential to be able to select from these candidate models in an objective manner. To this end, structure characterisation methods are introduced. Important sections of the book deal with the collection of high quality data using optimal experimental design, parameter estimation techniques for calibration and the on-line use of models in state and parameter estimators.

All material is illustrated with actual data collected by the authors.

Dynamical Modelling and Estimation in Wastewater Treatment Processes will be an invaluable text for practitioners, researchers and students concerned with the design, operation and control of biological wastewater treatment processes and plants.

Professor Denis Dochain is Honorary Research Director at Université Catholique de Louvain, Belgium. **Professor Peter A. Vanrolleghem** is Associate Professor in Bioprocess Control at Ghent University, Belgium.

ISBN: 1 900222 50 7