Quantitative Analyses of Anaerobic Wastewater Treatment Processes: Identifiability and Parameter Estimation

Nelly Noykova,[1] Thorsten G. Müller,[2] Mats Gyllenberg,[1] Jens Timmer[2]

[1]Department of Mathematical Sciences, University of Turku, FIN-20014 Turku, Finland
[2]Freiburg Center for Data Analysis and Modeling, University of Freiburg, 79104 Freiburg, Germany

Received 12 September 2001; accepted 1 November 2001

Abstract: We investigated the problem of identifying the parameters of a nonlinear fifth order model describing the population dynamics of two main bacterial groups in an anaerobic wastewater treatment process. In addition to addressing problems concerning structural and practical identifiability, we also analyzed how mathematical descriptions of bacterial population dynamics can model real data. Using three data sets recorded under different experimental conditions, we estimated important biochemical parameters and demonstrated that our model could describe the data successfully. Parameters, which are simultaneously determined using information from all three experiments, have more reliable estimates. We conclude that, after appropriate estimation, this model can be used for optimization and the control of continuous processes. © 2002 Wiley Periodicals, Inc. *Biotechnol Bioeng* **78**: 89–103, 2002; DOI 10.1002/bit.10179

Keywords: anaerobic digestion; mathematical modeling; theoretical identifiability; practical identifiability; parameter estimation

INTRODUCTION

Anaerobic wastewater purification processes have been increasingly used in the last few decades. These processes are important because they have positive effects: depollution of higher organic loading, which includes low sludge production and high pathogen removal, methane gas production and low energy consumption.

The increased interest in these processes has stimulated mathematical modeling, because it is usually much faster and less expensive to model a system and to simulate its operation than to perform laboratory experiments. The application of sophisticated methods of process control is only possible if mathematical models are available for the system to be optimized (Schürbüscher and Wandrey, 1991).

Correspondence to: N. Noykova

Contract grant sponsor: Graduate School in Computational Biology, Bioinformatics, and Biometry
Contract grant sponsor: Deutsche Forschungsgemeinschaft
Contract grant sponsor: Academy of Finland

The anaerobic degradation of organic matter is a complicated biological process. The conversion of organic matter consists of several independent, consecutive, and parallel reactions in which a close-knit community of bacteria cooperate to form a stable, self-regulating fermentation that transforms organic matter into a mixture of methane and carbon dioxide gases. These processes go through six main stages: hydrolysis of biopolimers (proteins, carbohydrates, lipids) into monomers (aminoacids, sugars, and long-chain fatty acids); fermentation of aminoacids and sugars; anaerobic oxidation of long-chain fatty acids and alcohols; anaerobic oxidation of intermediary products such as volatile fatty acids; conversion of acetate to methane; and the conversion of hydrogen to methane (Jeyaseelan, 1997). Several simulation models of these processes have been proposed (Husain, 1998; Jeyaseelan, 1997; v. Münch, 1999a). Angelidaki et al. (1999) described the hydrolysis of undissolved carbohydrates and the hydrolysis of undissolved proteins as separate paths. Their model included eight bacterial groups, 19 chemical compounds, and a detailed description of pH and temperature characteristics. The specific growth and decay rates can also be presented with differing levels of complexity (Angelidaki et al., 1999; Hill et al., 1983; Möche and Jördening, 1996; Thomas and Nordstedt, 1993).

The models described require the simultaneous solution of mass-balance equations for each individual substrate and bacterial population. Such a treatment is extremely complex, yielding equations with numerous unknown parameters. Therefore, simpler treatments have been developed to predict the dynamic behavior of digesters. The six main groups of bacteria were divided into two major groups: acid producing microorganisms and methane producing microorganisms (Hill and Barth, 1977; Husain, 1998; Jeyaseelan, 1997).

In this study we investigated such a simplified model, which is a modified version of Hill and Barth's model (1977). Although the model is simplified, it still has a large number of unknown parameters, and only a little

experimental data is available, which makes the parameter identification problem difficult to solve.

The main goals of our work were first to investigate the structural and practical identifiability of the model and, second, based on these results, to estimate the most important identifiable parameters for three data sets obtained from laboratory experiments.

MATERIAL AND METHODS

In this study we used three experimental data sets. The first data set and the experimental methods used to obtain this data were published previously (Simeonov et al., 1996). The last two data sets were obtained in the same laboratory.

All experiments were conducted in a 5-L automated stirred fermenter. The working volume of the reactor was 2 L. A fresh digester was started by preparing a mixture of water and cattle dung in a ratio that gave a final total solid concentration of 4.5% for the first data set, 6.3% for the second data set, and 12.65% for third data set. The dry weight at the end of the processes was 1.2%, 2.2% and 5.4%, respectively. The decreasing level also depends on the processing time, which was 51 days for the first two processes, and 58 days for the third.

During all the processes, the digester was maintained at a temperature of 34 ± 0.5°C; that is, the processes were mesophilic. The monitoring of the methane reactor was conducted by a dedicated data acquisition system of on-line sensors, which provided measurements of pH, temperature, rH, and biogas flow rate.

The structural identifiability analysis was made with the symbolic computational tools in MATHEMATICA 3.0. The solution of the differential equations was carried out numerically with a fourth order Runge-Kutta method, realized in SIMULINK toolbox 3.0, MATLAB 5.3. For our parameter estimation, we used a nonlinear constrained optimization method, which is implemented in the program fmincon in OPTIMIZATION toolbox 2.0, MATLAB 5.3. The programs for calculating confidence intervals and Monte Carlo simulations were written by the authors.

MODELING

The Model

In our model the anaerobic digestion is represented as a three-stage process (Ghaly and Pyke, 1991; Hill and Barth, 1977). During the first hydrolytic stage, the hydrolytic bacteria produce extracellular enzymes that hydrolyze the organic compounds into simple soluble compounds. The second stage is the acid-producing stage, in which acid-forming bacteria convert simple organic compounds into volatile acids. During the last, methanogenic stage, methanogenic bacteria convert volatile fatty acids into methane and carbon dioxide.

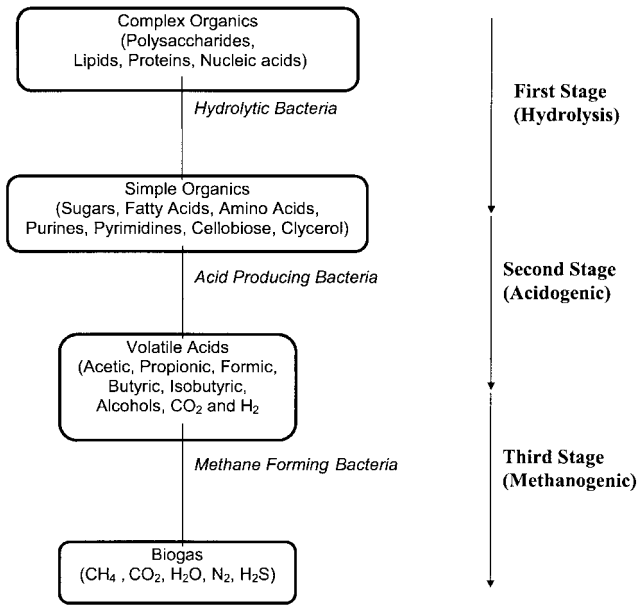

Figure 1. Anaerobic digestion processes.

Schematically these processes are presented in Fig. 1. Our model, corresponding to this three-stage scheme, is presented as follows:

$$\begin{aligned}
\frac{dC_{S_0}}{dt} &= -DC_{S_0} - \beta C_{X_1} C_{S_0} + D Y_P C_{S_{0i}} \\
\frac{dC_{X_1}}{dt} &= (\mu_1 - k_1 - D) C_{X_1} \\
\frac{dC_{S_1}}{dt} &= -DC_{S_1} + \beta C_{X_1} C_{S_0} - \frac{\mu_1 C_{X_1}}{Y_1} \\
\frac{dC_{X_2}}{dt} &= (\mu_2 - k_2 - D) C_{X_2} \\
\frac{dC_{S_2}}{dt} &= -DC_{S_2} + Y_b \mu_1 C_{X_1} - \frac{\mu_2 C_{X_2}}{Y_2} \\
Q &= Y_g \mu_2 C_{X_2}
\end{aligned} \tag{1}$$

where

$$\mu_1 = \frac{\mu_{1\max} C_{S_1}}{k_{S1} + C_{S_1}} \tag{2}$$

$$\mu_2 = \frac{\mu_{2\max} C_{S_2}}{(k_{S_2} + C_{S_2})(1 + \frac{C_{S_2}}{k_i})} \tag{3}$$

The interpretation of all variables and parameters and their dimensions are described under Nomenclature. The growth rate of acidogenic bacteria μ_1 is modeled according to the classical Monod formula Eq. (2). The growth rate of methanogenic bacteria μ_2 is described using the noncompetitive inhibition model Eq. (3). For simplicity we assume that inhibition by volatile fatty

acids occurs only with respect to methanogenic bacteria. These bacteria are the most sensitive to their growing conditions (Forster and Wase, 1990). In our model (1) the dynamic variables are represented by the state vector $\mathbf{x} = [x_j] = [C_{S_0}, C_{X_1}, C_{S_1}, C_{X_2}, C_{S_2}]^T \in \mathrm{R}^5$, the measurable output is $\mathbf{y} = [\mathrm{Q}] \in \mathrm{R}^1$, and the model parameters are represented by the vector $\mathbf{p} = [p_j] = [\mu_{1\max}, k_{S_2}, Y_2, \beta, k_{S_1}, k_1, y_1, \mu_{2\max}, k_i, k_2, Y_b, Y_g]^T \in \mathrm{R}^{12}$.

In this paper we analyze only batch processes; that is, from now on we assume D = 0.

In any biologically meaningful model, the concentrations must remain positive and bounded. It is easily verified that for each i $X_i = 0$ implies $\frac{dX_i}{dt} \geq 0$. Therefore, the positive cone $\{X_i \geq 0\}$ remains invariant. Moreover, the relation

$$\frac{d}{dt}\left\{\frac{1}{Y_2 Y_b}(Y_2 C_{S_2} + C_{X_2}) + 2Y_1(C_{S_0} + C_{S_1}) + C_{X_1}\right\} = -k_1 C_{X_1} - \frac{k_2}{Y_2 Y_b} C_{X_2} \leq 0$$

shows that all orbits are bounded.

For the continuous case, when D = 0, there is one equilibrium point. Stability of the equilibrium, as well as the conditions when wash-out can appear, were investigated by Simeonov et al. (1996).

Limitations Caused by the Experimental Conditions

Because we analyzed batch processes only, we were not able to investigate the influence of the parameter Y_p, and we therefore excluded it from the parameter vector. In our model only the biogas production rate Q is measurable, and we suppose that the initial values of the concentrations of the substrates and microorganisms are known. In our experiments the composition of the medium was measurable only once before mixture, but not during the experiments.

The estimation of parameters is much more reliable if there are experimental time series data of the concentrations of all substrates and microorganisms. Although obtaining data about substrate concentrations is possible, measuring of biomass concentration and microbial growth is still difficult. In the literature of wastewater treatment, there is no clear consensus on how microbial kinetics should be measured or how to interpret the results from existing technology (Ahring, 1995; Merkel et al., 1999; Pollard and Greenfield, 1997). The problem of determining biomass concentration, and especially how to incorporate the information from different measuring techniques in the model, is still open.

In the case of continuous cultivation, it is highly desirable to have more measurements for the substrate concentrations. In that case, D = 0 and the model become more complicated because the parameter Y_p has to be estimated as well. Some results from parameter estimation for a similar model in the continuous case with measurable substrates are reviewed by Nopens et al. (2001).

For successful parameter estimation in the continuous case, we have to perform a series of steady-state experiments at different dilution rates, D. From an experimental point of view the batch experiments are far more attractive because they are not so complicated and time-consuming. The main advantages and drawbacks of different experiments (batch, continuous, and fed-batch) for the Monod kinetics are reviewed in Nopens et al. (2001).

Boundary Values for All Parameters

To determine the admissible range of all model parameters, we conducted a review of the literature. The results are presented in Table I. We have not included values obtained from experiments with simulated media (synthetic wastewater) or with relatively pure cultures (e.g., Kalyuzhnyi et al., 2000; Kus and Wiesmann, 1995; Möche and Jördening, 1999). Using a single substrate and a pure culture to describe a reactor that contains a mixed culture and mixed substrates yields limited information, because many interactions inherent in mixed cultures are not taken into account (Hill, 1983).

We have not conducted a temperature correction for the specific growth rates $\mu_{1\max}$ and $\mu_{2\max}$, solubilization rate β, and decay coefficients k_1 and k_2, because the different authors have used different substrates, and there is no information about the involved community of microorganisms.

A temperature correction would be possible for our case if we had experimental data for these rates for different temperatures, but unfortunately this is not the case. As one of our experimental data sets is taken from Simeonov et al. (1996), we take our initial values for these coefficients from there.

THEORETICAL IDENTIFIABILITY

The identification problem is difficult to solve because of the high number (12) of parameters to be estimated, the complexity of the model, and the scarcity of experimental data. Therefore, we did not expect all 12 parameters in our model to be identifiable. Bastin et al. (1982) and Chalon et al. (1982) applied model transformations to determine which parameter combinations were identifiable. Unfortunately, even for their transformed model, it was still unclear whether at least the transformed parameters were identifiable. Furthermore, there was no one-to-one relationship between the original parameter set and the transformed one. Consequently, we have to assume a priori knowledge about most of the parameters from previous studies or from

Table I. Literature values of the model parameters previously used in dynamic anaerobic digestion models (mesophilic or thermophilic conditions).

Parameter [unit]	Value	References	Substrate and microorganisms used	Wastewater	Temperature [°C]
μ_{1max}, [day^{-1}]	**0.4**	Simeonov et al. (1996)	Sugars	Cattle manure	34 ± 0.5
	0.4	Hill and Barth (1977)	Amino acids, sugars	Poultry farming	25
	1	Angelidaki et al. (1993, 1999)	Carboh. enzymatic	Chemical oxygen demand cattle manure with glycerol trioleate or gelatin	55
	5.1	Angelidaki et al. (1993, 1999)	Glucose acidogens		55
	0.3	Kiely et al. (1997)	Glucose acidogens	Codigesting municipal solid waste and pig slurry	36
	0.313	Husain (1998), Hill's model reviewed in v. Münch et al. (1999a)	Biodegradable volatile solids, acidogens	Different types of farming: pig, beef, dairy, poultry	34
	2.5	Tschui (1989)			35
	5	Tschui (1989) and Siergist et al. (1993)	Sugars		35
	6	Bryers (1985)	Amino acids/amino acids and sugars		
	25	Jones et al. (1992)	Amino acids and simple sugars		
	0.55	Siergist et al. (1993)	Biodegradable soluble organics		
k_{S_2}[mg/L]	25	Hill and Barth (1977)	Volatile fatty acids (VFA)	Poultry farming	25
	0.82	Simeonov et al. (1996)		Cattle manure	34 ± 0.5
	120	Angelidaki et al. (1999)	Acetic acid	Chemical oxygen demand cattle manure with glycerol trioleate or gelatin	55
	120	Kiely et al. (1997)	Acetic acid	Codigesting municipal solid waste and pig slurry	36
	3000	Husain (1998), Hill's model	VFA	Different animal wastes: pig, beef, dairy, poultry	34
	64	Masse and Droste (2000)	*Methanosaeta* species	Pig manure	20
	1280	Masse and Droste (2000)	*Methanosarcina*	Pig manure	20
	[11, 421]	reviewed in Jeyaseelan (1997):	Acetic acid/acetate		
	154	Mosey (1983) reviewed in v. Münch et al. (1999a):	Acetic acid		35
	30	Siergist et al. (1993)	Acetic acid		
	80	Tschui (1989)	Acetic acid		35
	500	Bryers (1985)	Acetic acid		
Y_2 [mg/mg]	0.06	Hill and Barth (1977)	VFA, methanogens	Poultry farming	25
	0.0242	Simeonov et al. (1996)	VFA, methanogens	Cattle manure	34 ± 0.5
	0.08	Kiely et al. (1997)	Methanogens, acetic acid	Codigesting municipal solid waste and pig slurry	36
	0.042	Husain (1998), Hill's model	VFA, methanogens	Different animal wastes: pig, beef, dairy, poultry	34
	0.0524	Masse and Droste (2000)	Acetic acid	Pig manure	20
	[0.014, 0.054]	reviewed in Jeyaseelan (1997):	Acetic acid/acetate		
	0.04	Mosey (1983) reviewed in v. Münch et al. (1999a):	Acetic acid		35
	0.025	Tschui (1989) and Siergist et al. (1993)	Acetic acid		35
	0.029	Bryers (1985)	Acetic acid		
β, [day^{-1}]	**0.5**	Simeonov et al. (1996)	Volatile solids	Cattle manure	34 ± 0.5
	0.4	v. Münch et al. (1999b)	Insoluble organics	Raw domestic wastewater and primary sludge in prefermenters of waste water treatment plants in Canada and Australia	18 + 22
	0.15				
	0.15				
	0.05				
	0.092				

Table I. Continued

Parameter [unit]	Value	References	Substrate and microorganisms used	Wastewater	Temperature [°C]
k_{S_i} [mg/L]	150	Hill and Barth (1977)	Amino acids, sugars	Poultry farming	25
	160	Simeonov et al. (1996)	Sugars	Cattle manure	34 ± 0.5
	500	Angelidaki et al. (1999)	Glucose	Codig. cattle manure with glycerol trioleate or gelatin	55
	150	Kiely et al. (1997)	Glucose	Codigesting municipal solid waste and pig slurry	36
	1805	Masse and Droste (2000)	Carbohydrates	Pig manure	20
	[22.5, 630]	reviewed in Jeyaseelan (1977):	Carbohydrates		
	23	Mosey (1983) reviewed in v. Münch et al. (1999a):	Glucose		37
	2.2	Tschui (1989)	Amino acids or sugars		35
	22	Bryers (1985)	Amino acids, sugars		
	50	Siergist et al. (1993)	Amino acids, sugars		
	200	Siergist et al. (1993)	Long-chain fatty acids		
	2000	Tschui (1989)	Long-chain fatty acids		35
k_1, [day^{-1}]	0.025	Hill and Barth (1977)	SO, acidogens	Poultry farming	25
	0.004	Simeonov et al. (1996)	Sugars, acidogens	Cattle manure	34 ± 0.5
	0.05	Angelidaki et al. (1999)	Carbon, enzymatic;	Codig. cattle manure	55
	0.255	Angelidaki et al. (1999)	glucose acidogens	with glycerol trioleate or gelatin	55
	0.006	Masse and Droste (2000)	Acid formers	Pig manure	20
	0.008	Masse and Droste (2000) reviewed in v. Münch et al. (1999a):	Acetogenic butyric acid bacteria	Pig manure	20
	0.43	Tschui (1989)	Amino acids, sugars		35
	1	Siegrist et al. (1993) reviewed in Jeyaseelan (1997):	Amino acids, sugars		
	0.8	Mosey (1983)	Glucose		37
	6.1	Pavlostatis (1991)	Carbohydrates		
Y_1, [mg/mg]	0.2	Hill and Barth (1977)	Amino acids, sugars	Poultry farming	25
	0.0264	Simeonov et al. (1996)	Sugars, acidogens	Cattle manure	34 ± 0.5
	0.188	Kiely et al. (1997)	Glucose acidogens	Codigesting municipal solid waste and pig slurry	36
	0.07	Husain (1998), Hill's model	Sugars, acidogens	Different animal wastes: pig, beef, dairy, poultry	34
	0.228	Masse and Droste (2000)	Carbohydrates	Pig manure	20
	[0.14, 0.17]	reviewed in Jeyaseelan (1997):	Carbohydrates		
	0.173	Mosey (1983) reviewed in v. Münch et al. (1999a):	Glucose		37
	0.036	Bryers (1985)	Amino acids, sugars		
	0.15	Tschui, Siergist et al. (1993)	Amino acids/sugars		35
	0.25	Tschui (1989)	Sugars		35
μ_{2max}, [day^{-1}]	0.4	Hill and Barth (1977)	VFA, methanogens	Poultry farming	25
	0.4	Simeonov et al. (1996)	VFA, methanogens	Cattle manure	34 ± 0.5
	0.6	Angelidaki et al. (1993)	Acetic acid, methanogens	Codig. cattle manure with glycerol trioleate or gelatin	55
	[0.67, 1]	Hansen et al. (1998)	Methanogens	Mixture pig-cattle manure in different ratio	55
	[0.18, 1]	Hansen et al. (1998)	Methanogens	Mixture of pig manure with different ammonia concentrations	55

Table I. Continued

Parameter [unit]	Value	References	Substrate and microorganisms used	Wastewater	Temperature [°C]
	[0.21, 1]	Hansen et al. (1998)	Methanogens	Pig manure with different H_2/CO_2 ratio	36
	0.6	Kiely et al. (1997)	Methanogens, acetic acid	Codigesting municipal solid waste and pig slurry	34
	0.313	Husain (1998), Hill's model reviewed in v. Münch et a1. (1999a):	Biodegradable volatile solids, acidogens	Different type wastes: pig, beef, dairy, poultry	
	0.3	Jones et al. (1992)	Acetic acid		35
	0.34	Bryers (1985)	Acetic acid		
	0.36	Tschui (1989)	Acetic acid		
	0.48	Siergist et al. (1993)	Acetic acid		
k_i [g/L]	0.3	Hill and Barth (1977)	Methanogens, VFA	Poultry farming	25
	5	Angelidaki et al. (1999)	Methanogens, LCFA	Codig. cattle manure with glycerol trioleate or gelatin	55
	3,432	Noykova and Gyllenberg (2000)	Methanogens, VFA	Cattle manure	34 ± 0.5
	41.85				
	0.015	Kiely et al. (1997)	Methanogens, acetic acid	Codigesting municipal solid waste and pig slurry	36
	9	Husain (1998), Hill's model	Methanogens, VFA	Different animal wastes: pig, beef, dairy, poultry	34
k_2, [day^{-1}]	0.04	Hill and Barth (1977)	VFA, methanogens	Poultry farming	25
	0.004	Simeonov et al. (1996)	VFA, methanogens	Cattle manure	34 ± 0.5
	0.03	Angelidaki et al. (1999)	Acetic acid, methanogens	Codig. cattle manure with	55
	0.016	Kiely et al. (1997) reviewed in Jeyaseelan (1997):	Methanogens, acetic acid	glycerol trioleate or gelatin	36
	0.019	Mosey (1983)	Acetic acid		35
	6.1	Pavlostatis reviewed in v. Münch et al. (1999a):	Acetic acid/ acetate		
	0.003	Bryers (1985)	Acetic acid	Codigesting municipal solid waste and pig slurry	35
	0.005	Tschui (1989)	Acetic acid		
	0.1	Siergist et al. (1993)	Acetic acid		
Y_b [mg/mg]	**2.45**	Hill and Barth (1977)	SO, amino acids and sugars	poultry farming	25
	45.51	Simeonov et al. (1996)	SO, sugars	Cattle manure	34 ± 0.5
	3.543	Angelidaki et al. (1993, 1999)	Yield of acetate from glucose degraders	Codig. cattle manure with glycerol trioleate or gelatin	55
	0.38	Kiely et al. (1997)	Yield of acetate from glucose degraders	Codigesting municipal solid waste and pig slurry	36
	9	Husain (1998), Hill's model reviewed in v. Münch et al. (1999a):	Yield of acetate from glucose degraders	Different animal wastes: pig, beef, dairy, poultry	34
	26.7	Bryers (1985)	Amino acids, sugars		
	21.22	Tschui (1989)	Long-chain fatty acids		35
	5.66	Siergist et al. (1993)	Amino-acids/sugars		
	3	Tschui (1985)	Sugars		35
Y_g [L^2 CH4/ mg m.o.]	**74.54**	Simeonov et al. (1996)	Yield of methane from VFA for methanogens	Cattle manure	34 ± 0.5
	15.37	Husain (1998), Hill's model reviewed in v. Münch et al. (1999a)	VFA, methanogens	Wastewaters from different types of animal farming: pig, beef, dairy, poultry	34
	19.5	Tschui (1989) and Siergist et al. (1993)	Yield of methane from acetic acid		35
	16.74	Bryers (1985)	Yield of methane from acetic acid		

the literature. In this study we assume only the subset of the parameters $\mathbf{p_2} = [\mu_{1\,\max}, k_{S_2}, Y_2]$ to be unknown. The reason for this choice will be explained later. We also suppose that all initial values of the state variables are known.

First we discuss whether the unknown subset of the model parameters is theoretically identifiable. We have to determine whether identifying every parameter in $\mathbf{p_2}$ from precise and noiseless experimental data is possible (Godfray and DiStefano, 1985; Julien et al., 1998).

There are several approaches to proving structural identifiability:

1. Vajda's approach, based on the local state isomorphism theorem and developed from Vajda et al. (1989) for proving the theoretical identifiability of control systems. Unfortunately, this approach is not applicable in our case for a batch model without external input (Joly-Blanchard and Denis-Vidal, 1998).
2. Ljung's approach for testing theoretical identifiability using differential algebra. The question of global identifiability is reduced to the question as to whether the given model structure can be rearranged as a linear regression (Ljung and Glad, 1994). An improved version of this approach is presented in Denis-Vidal et al. (2001). The computer implementation of Ljung's method, described in Wang (1995), was tried, but unfortunately did not yield any results due to computational problems: The complexity of this method increases rapidly with the size of the problem.
3. The transformation of the nonlinear model into a model linear in the parameters (Dochain et al., 1995). We were not able to find such a transformation for our case.
4. The Taylor series expansion approach (Godfrey and DiStefano, 1985; Holmberg, 1982; Pohjanpalo, 1982), which proves to be successful in our case.

Our model (1) is presented in the following form:

$$M^{\mathbf{p_2}} \begin{cases} \dot{\mathbf{x}}(t,\mathbf{p}) = \mathbf{f}(\mathbf{x}(t,\mathbf{p}),t,\mathbf{p}) \\ \mathbf{y}(t,\mathbf{p}) = \mathbf{g}(\mathbf{x}(t,\mathbf{p}),t,\mathbf{p}), \\ \mathbf{x}_0 = \mathbf{x}(t_0,\mathbf{p}) \end{cases} \tag{4}$$

In model (4) the parameter vector $\mathbf{p}$ is presented as $\mathbf{p} = [\mathbf{p_1}, \mathbf{p_2}]^T$, where $\mathbf{p_1} = [C_{S_0}(0), C_{X_1}(0), C_{S_1}(0), C_{X_2}(0), C_{S_2}(0), \beta, k_{S_1}, k_1, Y_1, \mu_{2\,\max}, k_i, k_2, Y_b, Y_g]$ is the vector of known parameters, $\mathbf{p_2} = [\mu_{1\,\max}, k_{S_2}, Y_2]$ is the vector of unknown parameters, and $\mathbf{f}$ and $\mathbf{g}$ are nonlinear vector functions that define the known coupling parameterized by the parameter vector $\mathbf{p}$, see (1). We assume that $\mathbf{p_2} \in \Omega$, where Ω is an open subset in R^3. The global (cf. local) identifiability at $\mathbf{p_2} \in \Omega$ of the model $\mathbf{M}^{\mathbf{p_2}}$ is defined as follows (Denis-Vidal et al., 200l; Pohjanpalo, 1982): for any $\mathbf{p_2'} \in \Omega$ (cf. there exists an open ε-neighborhood $N(\mathbf{p_2}, \varepsilon) \subset \Omega$, such that for any $\mathbf{p_2'} \in N(\mathbf{p_2}, \varepsilon)$), $\mathbf{p_2} \neq \mathbf{p_2'}$ the systems $M^{\mathbf{p_2}}$ and $M^{\mathbf{p_2'}}$ will yield different outputs. These definitions are generically extended so that $M^{\mathbf{p_2}}$ is said to be globally (locally) theoretically identifiable if it is globally (locally) identifiable at all $\mathbf{p_2} \in \Omega$ except the points of a subset of zero measure in Ω.

In the Taylor series approach (Pohjanpalo, 1982), y(t) and its successive time derivatives are evaluated in terms of the unknown model parameters $\mathbf{p_2}$ at a particular time, usually t = 0, that is:

$$y(t,\mathbf{p_2}) = y(0,\mathbf{p_2}) + y^{(1)}(0,\mathbf{p_2})t + y^{(2)}(0,\mathbf{p_2})\frac{t^2}{2} + \ldots + y^{(i)}(0,\mathbf{p_2})\frac{t^i}{i!} + \ldots \tag{5}$$

where $y^{(i)}(0,\mathbf{p_2}) \equiv \frac{d^i y}{dt^i}(0,\mathbf{p_2})$.

Because the measurement vector is unique, all its derivatives are unique. Then the problem of showing theoretical identifiability for model (4) with respect to the parameters $\mathbf{p_2}$ is reduced to determining the number of solutions for $\mathbf{p_2}$ for a set of algebraic equations:

$$g^{(k)}(\mathbf{x}(0),\mathbf{p}) = y^{(k)}(0,\mathbf{p_2})\ k = 0,\ldots,\infty, \tag{6}$$

where $g^{(k)}$ is the k^{th} derivative of the vector function g. The Eq. (6) is in general nonlinear in the parameters.

By definition, the parameter set $\mathbf{p_2}$ is unidentifiable if the set of solutions is uncountable, it is locally identifiable if the set of solutions is countable, and it is globally identifiable if there is a unique solution (Chappel et al., 1990; Godfray and DiStefano, 1985; Pohjanpalo, 1982).

Our theoretical identifiability analysis with respect to the parameter vector $\mathbf{p_2}$ is given in the Appendix. Under some conditions these parameters are locally identifiable with two solutions at most. If there were more solutions for these parameters (in one of the examples of Pohjanpalo [1982] there were 36 different solutions), then the local identifiability would be of very little value.

To find useful identifiable parameter combinations, we also used the results from sensitivity analysis (Noykova and Gyllenberg, 2000). Depending on their influence on the output variable Q, we divided all parameters into three groups: (1) kinetic parameters included in the equation for μ_1, β, and decay coefficient k_1; (2) kinetic parameters included in the equation for μ_2 and the decay coefficient k_2, and (3) all yield coefficients. We had to choose only one parameter from each group because the Taylor series approach does not provide results for theoretical identifiability in cases with more than three parameters. The reason is that computational problems appeared because the expressions are very complicated, and the memory of MATHEMATICA was often exceeded. This clearly shows that existing methods for identifiability analysis have to be improved, especially with respect to their computational implementation.

Table II. Dependence of the noncompetitive substrate inhibition $\mu_2 = \mu_2(C_{S_2})$ on $C_{S_2}(0)$ (or $C_{S_2\text{sum}}$ in some cases), k_{S_2} and k_i.

Possible cases	Specific case approximation	Approximation of the function $\mu_2 = \mu_2(C_{S_2})$	Comments: possible simplifications
Case 1 $k_{S_2} \gg C_{S_2\text{sum}}$	$\frac{C_{S_2}}{k_{s_2}+C_{S_2}} \approx \frac{C_{S_2}}{k_{S_2}}$	$\mu_2 \approx \frac{\mu_{2\max} k_i}{k_{S_2}} \frac{C_{S_2}}{(k_i+C_{S_2})}$	Monod equation with maximal specific growth rate $\alpha = \frac{\mu_{2\max} k_i}{k_{S_2}}$ and saturation constant k_i.
Case 1.1 $k_i \gg C_{S_2\text{sum}}$	$\frac{C_{S_2}}{k_i+C_{S_2}} \approx \frac{C_{S_2}}{k_i}$	$\mu_2 \approx \frac{\mu_{2\max}}{k_{S_2}} C_{S_2}$	Linear dependence $\mu_2 = \mu_2(C_{S_2})$
Case 1.2 $k_i \ll C_{S_2}(0)$	$\frac{C_{S_2}}{(k_i+C_{S_2})} \approx 1$	$\mu_2 \approx \frac{\mu_{2\max} k_i}{k_{S_2}}$	μ_2 is a constant
Case 2 $k_{S_2} \ll C_{S_2}(0)$	$\frac{C_{S_2}}{(k_{s_2}+C_{S_2})} \approx 1$	$\mu_2 \approx \frac{\mu_{2\max} k_i}{k_i+C_{S_2}}$	Remark: $\mu_2 = \mu_2(C_{S_2})$ has a maximum μ_2^* when $C_{S_2}^* = \sqrt{k_i k_{S_2}}$
Case 2.1 $\frac{\sqrt{k_i k_{S_2}}}{C_{S_2}(0)} \geq 1$		$\mu_2 = \begin{cases} 0 \text{ when } C_{S_2}(0) = 0 \\ \mu_{2\max} \text{ when } C_{S_2}(0) > 0 \end{cases}$	$\mu_2 = \mu_2(C_{S_2})$ can be replaced with a more simple discrete model
Case 2.2 $\frac{\sqrt{k_i k_{S_2}}}{C_{S_2}(0)} < 1$		$\mu_2 = \begin{cases} 0 \; C_{S_2}(0) = 0 \\ \mu_2^* \text{ when } 0 < C_{S_2}(0) < \sqrt{k_i k_{S_2}} \\ \mu_2 = \frac{\mu_{2\max} k_i}{k_i + C_{S_2}} \text{ when } C_{S_2}(0) \geq \sqrt{k_i k_{S_2}} \end{cases}$	$\mu_2 = \mu_2(C_{S_2})$ can be replaced with a more simple discrete model
Case 3 k_{S_2} is in the same range as $C_{S_2}(0)$ or $C_{S_2\text{sum}}$			
Case 3.1 $\frac{\sqrt{k_i k_{S_2}}}{C_{S_2}(0)} \geq 1$		$\mu_2 \approx \frac{\mu_{2\max} C_{S_2}}{k_{S_2}+C_{S_2}}$	Monod dependence
Case 3.2 $\frac{\sqrt{k_i k_{S_2}}}{C_{S_2}(0)} < 1$		$\mu_2 = \frac{\mu_{2\max} C_{S_2}}{(k_{S_2}+C_{S_2})\left(1+\frac{C_{S_2}}{k_i}\right)}$	Typical noncompetitive substrate inhibition dependence

PRACTICAL IDENTIFIABILITY IN THE CASE OF SPECIFIC PARAMETER RELATIONSHIPS

In the previous section we demonstrated the structural identifiability of the parameters $\mu_{1\max}$, k_{S_2}, and Y_2 given the model structure and perfect data of the output variable Q. Here we discuss the practical identifiability problems related to specific parameter combinations. After obtaining the parameter estimates, we discuss the practical identifiability problems related to the quality of the experimental data and their informative content. These problems arise in the case of real experimental data, often complicated by unknown noise characteristics (Dochain et al., 1995; Holmberg, 1982; Vanrolleghem and Dochain, 1998).

As Holmberg (1982) demonstrated with the Monod model, it is possible to make model simplifications if the biological parameters are in specific areas of the parameter space.

To generalize this idea, we analyzed the dependence of the noncompetitive substrate inhibition described by μ_2 in Eq. (3) on parameters $\mu_{2\max}$, k_{S_2}, and k_i.

We demonstrate the parameter relationships for which the noncompetitive substrate inhibition can be replaced by simpler mathematical descriptions, in some cases even by a simple linear dependence so that the model could be drastically simplified. We investigated all possible cases.

Later we will be able to check possible model reductions for given $\mu_{2\max}$ and k_i from a priori knowledge and estimated k_{S_2} from experimental data.

The function $\mu_2(C_{S_2})$, Eq. (3), can take on a different shape depending on the ratio between the parameter k_{S_2} and the value of the maximal substrate concentration $C_{S_2\max}$. The substrate concentration C_{S_2} can take values $0 \leq C_{S_2} \leq C_{S_2\max}$. The maximal substrate concentration $C_{S_2\max}$ can take values $C_{S_2}(0) \leq C_{S_2\max} < C_{S_2\text{sum}}$, where $C_{S_2\text{sum}} = C_{S_0}(0) + C_{S_1}(0) + C_{S_2}(0)$ is the sum of all initial substrate concentrations in the beginning of the batch process. To analyze the dependence of μ_2 (C_{S_2}) on the three parameters, we distinguished between three different cases. The main results from the analysis of these cases are summarized in Table II. All possible situations in Case 1 are shown in Figure. 2. This figure

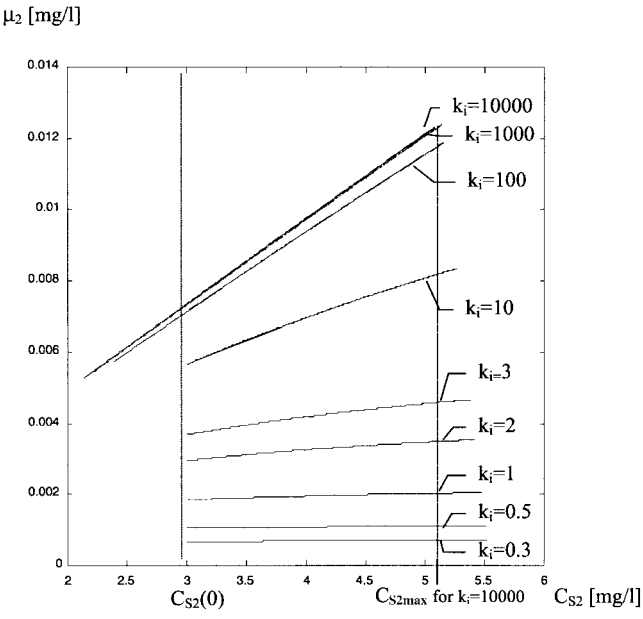

Figure 2. The specific growth rate μ_2 as a function of C_{S_2} for k_{S_2} = 160 (mg/L) and different values of the inhibition coefficient k_i.

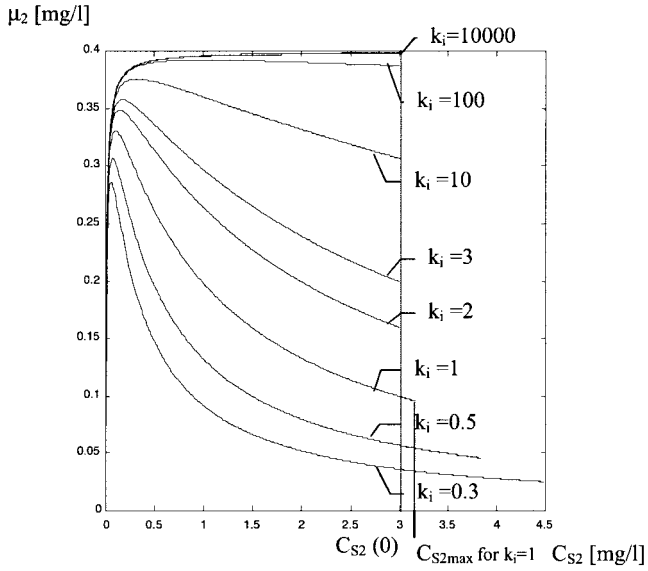

Figure 3. The specific growth rate μ_2 as a function of C_{S_2} for k_{S_2} = 0.01 (mg/L) and different values of the inhibition coefficient k_i.

shows that all curves between the lower and upper ones can be approximated linearly with a high degree of accuracy. All possible situations in Case 2 are shown in Figure 3. The parameter $\mu_{2\max}$ determines an upper bound of the specific growth rate, and the equation $\mu_2 \approx \frac{\mu_{2\max} k_i}{C_{s_2}}$ determines a lower bound of μ_2. The graphical representation of Case 3 is given in Figure 4.

From this analysis, we conclude that if the saturation constant is $k_{S_2} \gg C_{S_2\text{sum}}$ or $k_{S_2} \ll C_{S_2}(0)$, the model Eq. (3) can be reduced to the Monod or simpler equations. In these cases, the Haldane approach to noncompetitive inhibition kinetics is not feasible.

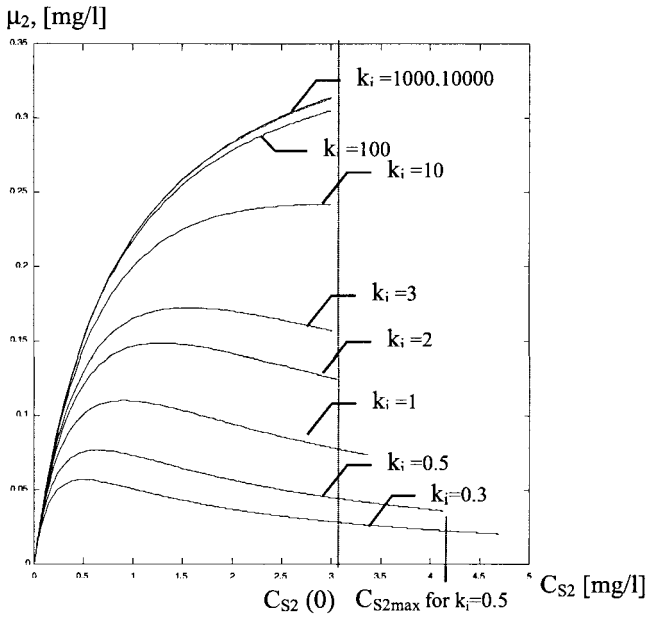

Figure 4. The specific growth rate μ_2 as a function of C_{S_2} for k_{S_2} = 0.82 (mg/L) and different values of the inhibition coefficient k_i.

Note: For our practical identifiability investigation in Cases 2 and 3 (inequality $\frac{\sqrt{k_i k_{S_2}}}{C_{S_2}(0)} \geq 1$) we used the value $C_{S_2}(0)$ instead the maximal substrate concentration value $C_{S_2\max}$. We were not able to predict the value $C_{S_2\max}$ because it is influenced by the dynamics in the first two stages of the model. Consequently, it is possible that some of our predictions for the Monod model are valid only for the concentrations $C_{S_2\max} = C_{S_2}(0)$. Our simulation studies show that, in most cases, this assumption is true, see Figures 3 and 4, where the initial conditions of all substrates are taken from our laboratory experiments.

RESULTS FROM THE PARAMETER ESTIMATION IN THE CASE OF NOISY DATA

Qualitative Analysis of the Results from Parameter Estimation

After investigating local theoretical identifiability for $\mu_{1\max}$, k_{S_2}, and Y_2 we used our experimental data to estimate these biochemically important parameters. Although the data sets were recorded under similar experimental conditions, they describe three different situations: (1) the normal case, (2) a case with strong substrate inhibition, and (3) a case with high organic loading.

Before starting our estimation, we checked whether the conditions for the local structural identifiability held (inequalities [14], [22], [23], and [24] from the Appendix). For this purpose, we assumed $Q^{(0)} = Q^{\text{model}}(0)$, and quite arbitrarily the first two derivatives $Q^{(1)} = Q^{(2)} = 0$. We used MATLAB for our calculations. For all three cases, these inequalities held, and model (1) is locally identifiable with respect to $\mathbf{p_2}$.

We estimated the kinetic parameters $\mu_{1\max}$ and k_{S_2} separately for different data sets because we had no information about the microorganisms involved, and we expected that they would have different characteristics in different data sets. We estimated the yield coefficient Y_2 simultaneously for all data sets because we assumed the same biochemical reactions. We expected smaller confidence intervals for this parameter because we used much more information in the estimation process, namely all three experiments, than for the other parameters.

In particular, we wanted to discover whether the mathematical model could describe all three different experimental situations given the estimated parameters.

To start the estimation procedure, we first needed values for the biological parameters, which we took from the literature. These can be found in Table I, written in boldface. As the second experimental data set shows strong substrate inhibition, i.e., low pH values during the experiment, we assumed a different inhibition coefficient for this data set. For the second data set, we assumed k_i = 3.432 (mg/L) and for first and third data set, k_i = 41.85 (mg/L). Second, we needed starting

Table III. Initial values of the substrate and biomass concentrations for different experiments.

	Initial values of the state variables				
	C_{S_0} (0)	C_{X_1} (0)	$C_{S_1}(0)$	$C_{X_2}(0)$	$C_{S_2}(0)$
Experiment 1	14.24	0.1	7	0.01	3
Experiment 2	14.24	0.1	7	0.01	6
Experiment 3	48	0.1	30	0.01	3

conditions for the state variables. As the first data set was the same as in Simeonov et al. (1996), most of our initial values were the same as the values given there (see Table III).

For our estimation we used OPTIMIZATION toolbox 2.0 in MATLAB 5.3. We used the nonlinear constrained optimization method because we have information about the boundaries for the parameters to be estimated. Our optimization criterion is a sum of the optimization criteria for every data set:

$$CRIT(\mathbf{p_2}) = CRIT_1(\mathbf{p_2}) + CRIT_2(\mathbf{p_2}) + CRIT_3(\mathbf{p_2}), \tag{7}$$

$$CRIT_J(\mathbf{p_2}) = \sum_{i=0}^{N} w_i (Q^i(\mathbf{p_2}) - Q^i_{\text{exp}})^2,\ j = 1, \ldots, 3, \tag{8}$$

where $w_i = \frac{1}{error_i}$ are weighting coefficients, $\mathbf{p_2} = |\mu_{1\max}, k_{S_2}, Y_2|$ is the parameter vector to be estimated, and the number of data points is N = 50.

The numerical results from our estimation are given in Table IV, and the graphical results are shown in Fig. 5, 6, and 7.

The results from the estimation show that the model can fit the experimental data for all three experimental situations.

The estimated values of the parameter $\mu_{1\max}$ for all three experimental sets reveal an almost linear dependence of the Monod curve $\mu_1 = \mu_1(C_{S_i})$. This means we can assume the estimated $\mu_{1\max}$ values are true only if we are sure the values k_{S_1} are determined very accurately. This also follows from Holmberg's (1982) results concerning the practical identifiability of the kinetic coefficients in the Monod model.

With the estimated value of parameter k_{S_2} for the second data set the inequality $\frac{\sqrt{k_i k_{S_2}}}{C_{S_2}(0)} = 0.819 < 1$ still holds, and we have a typical noncompetitive substrate inhibition curve for the second experimental situation. For the other two data sets the inequality $\frac{\sqrt{k_i k_{S_2}}}{C_{S_2}(0)} > 1$ holds, which means we have typical Monod curve. This shows that the obtained results from the estimation procedure are in agreement with the observed experimental phenomena.

Practical Identifiability of the Estimated Parameters

Sensitivity Analysis

Here we investigate the influence of a small deviation in the parameter set on the fit of the model to the data. This means we are interested in the value of the objective functional $CRIT_j$ for a parameter set slightly differently from the optimal one. This expected value is given by Eq. (9) (Vanrolleghem and Dochain, 1998):

$$\begin{aligned} &E[CRIT_J(\mathbf{p_2} + \delta\mathbf{p_2})] \\ &\quad = \delta\mathbf{p_2}^T \left[\sum_{i=0}^{N} \left(\frac{\partial Q}{\partial \mathbf{p_2}}(t_i) \right)^T w_i \left(\frac{\partial Q}{\partial \mathbf{p_2}}(t_i) \right) \right] \delta\mathbf{p_2} \\ &\qquad + \sum_{i=0}^{N} tr(C_i w_i) \end{aligned} \tag{9}$$

in which C_i represents the measurement error covariance matrix. The term between brackets is the Fisher information matrix and expresses the information content of

Table IV. Numerical results from the estimation.

	Simultaneous estimation of Y_2^*					
	First data set		Second data set		Third data set	
$\mu^*_{1\max}$	0.8313	+0.08 −0.07	0.27	±0.08	0.4263	±0.06
$k^*_{S_2}$	5.0858	+0.4 −0.3	7.0305	+2.2 −1.7	22.5625	±0.7
Y_2^*	0.0127			+0.0003	−0.0004	
CRIT	670.3802					

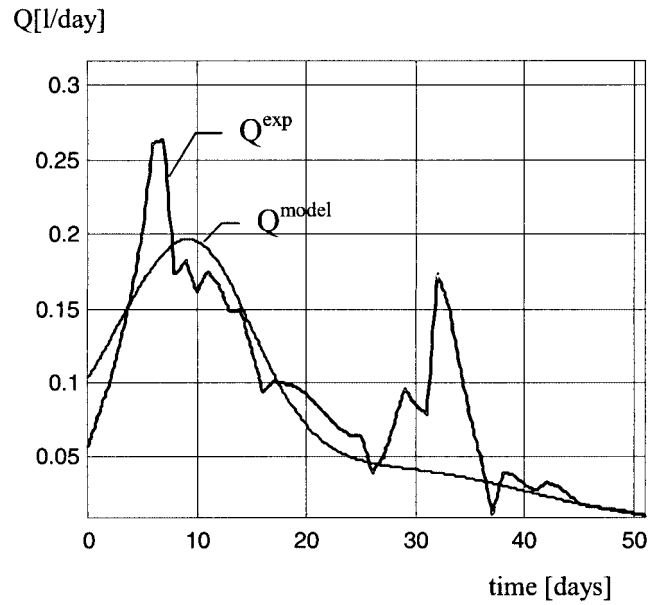

Figure 5. Results from parameter estimation for first data set. Q^{model} is very close to Q^{exp} for the points with high weighting factors.

the experimental data (Vanrolleghem and Dochain, 1998):

$$\mathbf{F} = \sum_{i=0}^{N} \left(\frac{\partial Q}{\partial \mathbf{p_2}}(t_i) \right)^T w_i \left(\frac{\partial Q}{\partial \mathbf{p_2}}(t_i) \right) \tag{10}$$

The terms $\frac{\partial Q}{\partial \mathbf{p_2}}$ are the sensitivity functions of the parameters $\mathbf{p_2}$ with respect to the output measurable variable Q. The sensitivity analysis is a central task in the practical identifiability study (Holmberg, 1982; Noykova and Gyllenberg, 2000; Vanrolleghem and Dochain,

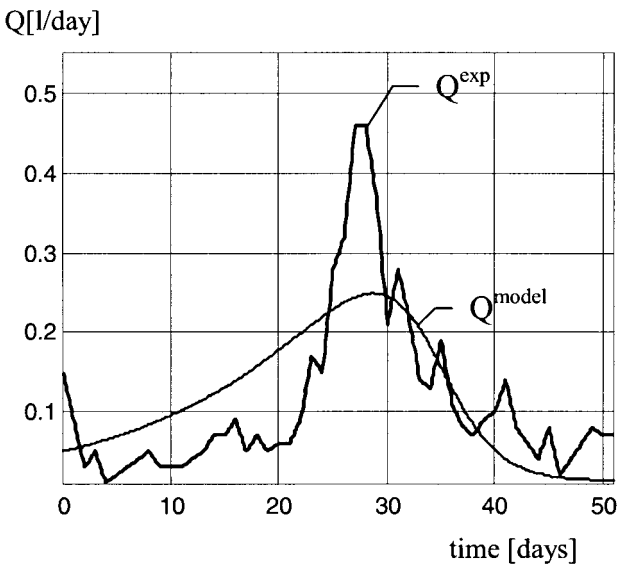

Figure 6. Results from parameter estimation for second data set. Q^{model} present the main trend in the measurable behavior Q^{exp}.

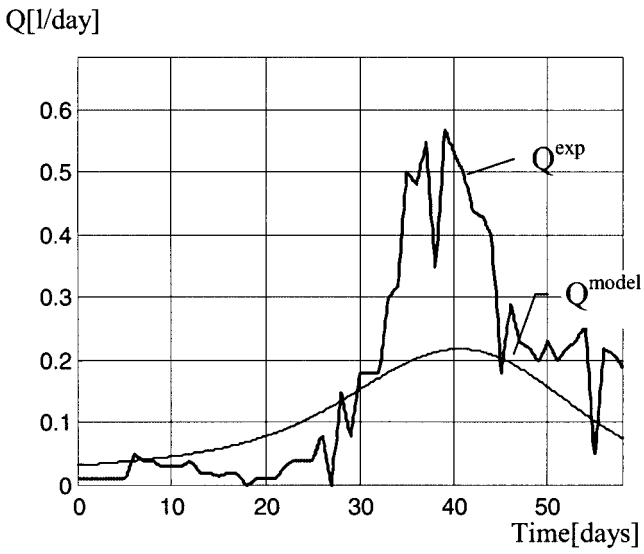

Figure 7. Results from parameter estimation for third data set. Q^{model} present the main trend in the measurable behavior $Q^{\text{exp.}}$

1998; Yordanova and Noykova, 1996). Here we use the relative sensitivity functions, logarithmic sensitivity functions, because they are nondimensional and allow us to compare the results for different parameters and variables. These functions are defined as:

$$T_{Qi} = \frac{\partial \ln Q}{\partial \ln p_i}, \quad i = 1, 2, 3. \tag{11}$$

Vanrolleghem and Dochain (1998) suggested a useful test for practical identifiability. If the sensitivity functions are linearly dependent, the model is not practically identifiable. Stronger evidence can be obtained by calculating of the rank of the Fisher information matrix. If no linear dependence exists, it should be full rank. According to our results, we have rank(F) = 3, i.e., the Fisher matrix has full rank, for all data sets. This means our model is practically identifiable with respect to the parameters $\mu_{1\max}$, k_{S_2}, and Y_2 for all three data sets.

Confidence Region of the Parameter Estimates

We can measure the quality of the estimates quantitatively by calculating confidence regions. Unfortunately, in nonlinear systems an analytical description of the probability distribution of the parameters exists only in the limit of large number of data points, namely using the asymptotic normality properties of the maximum likelihood estimator (Cox and Hinkley, 1974). In this rare case the Hessian matrix, i.e., the second derivative of the score function with respect to the parameters, can be used to obtain confidence intervals. In experiments with small sample sizes, it is better to rely on the Monte Carlo simulation (MCS) (Press et al., 1992). The idea of the MCS is that the probability distribution of the pa-

rameters at the convergence point does not differ substantially from the probability distribution of the true values, because we assume that our estimate is close to the true parameters. Therefore, we can use our estimate to simulate many data sets closely related to the original data set. We can obtain an approximation of the probability distribution for every parameter and hence can calculate 95% confidence intervals for all parameters.

One of the advantages of this method is that we use the original nonlinear model instead a linear approximation in the neighborhood of the best parameter estimates.

The results after applying the MCS method are displayed in Table IV. Due to the small number of data and the nonlinearity of the model, confidence intervals are not symmetric and the computation of the confidence intervals with the Hessian matrix approach would yield inappropriate results. The parameter Y_2, which was estimated simultaneously, could be determined most reliably. This is a clear sign that simultaneous estimation using many experiments is superior to the analysis of single experiments.

To summarize, the model is able to describe different experimental situations, producing reliable estimates with relatively low confidence intervals.

CONCLUSIONS

We investigated a modified fifth order nonlinear Hill's and Barth's model, conducting both structural and practical identifiability analyses. The results show that the model is locally structurally identifiable. Practical identifiability analysis was carried out using several approaches — qualitative analysis of the kinetic parameters for methanogenic growth, sensitivity analysis, and calculating confidence regions using Monte Carlo simulations. We conclude that the model is practically identifiable and the parameter estimates are reliable.

In addition, we have provided a review of the literature concerning the possible parameter values. These values show the possible parameter boundaries, which can assist the work of other researchers in this area, too.

Finally, three main important parameters were estimated. One important feature of the estimation procedure is the simultaneous estimation of the parameters, which make the parameter estimates more reliable.

The results from the parameter estimation show that the model can describe different experimental phenomena. We conclude that, after appropriate estimation, this model can be used for optimization and control of continuous processes, which is the subject of our further work.

APPENDIX

Theoretical Identifiability Analysis with Respect to the Parameters μ_{1max}, K_{S_2} and Y_2

If we denote $X^{(i)} = \frac{d^i X(0,\mathbf{p})}{dt^i}$ and $Q^{(i)} = \frac{d^i Q(0,\mathbf{p})}{dt^i}$, model (1) can be written as:

$$C_{S_0}^{(1)} = -DC_{S_0}^{(0)} - \beta C_{X_1}^{(0)} C_{S_0}^{(0)},$$

$$C_{x_1}^{(1)} = \left(\frac{\mu_{1\max} C_{S_1}^{(0)}}{k_{S_1} + C_{S_1}^{(0)}} - k_1 \right) C_{X_1}^{(0)},$$

$$C_{S_i}^{(1)} = \beta C_{x_1}^{(0)} C_{S_0}^{(0)} - \frac{\mu_{1\max} C_{S_i}^{(0)} C_{X_1}^{(0)}}{(k_{S_1} + C_{S_1}^{(0)}) Y_1},$$

$$C_{X_2}^{(1)} = \left(\frac{\mu_{2\max} C_{S_2}^{(0)}}{(k_{S_2}^{(0)} + C_{S_2}^{(0)}) \left(1 + \frac{C_{S_2}^{(0)}}{k_i}\right)} - k_2 \right) C_{X_2}^{(0)}, \qquad (12)$$

$$C_{S_2}^{(1)} = Y_b \frac{\mu_{1\max} C_{S_1}^{(0)}}{k_{S_1} + C_S^{(0)}} C_{x_1}^{(0)} - \frac{\mu_{2\max} C_{S_2}^{(0)}}{(k_{s_2} + C_{S_2}^{(0)}) \left(1 + \frac{C_{S_2}^{(0)}}{k_1}\right)} - \frac{C_{x_2}^{(0)}}{Y_2},$$

$$Q^{(0)} = Y_g \frac{\mu_{2\max} C_{S_2}^{(0)}}{(k_{S_2} + C_{S_2}^{(0)}) \left(1 + \frac{C_{S_2}^{(0)}}{k_i}\right)} C_{X_2}^{(0)},$$

Determining K_{s_2}

We can determine parameter k_{S_2} directly from the last equation from model (12) because the initial values of all variables, as well as the values of the parameters μ_{2max} and k_i are known:

$$k_{S_2} = \frac{C_{S_2}^{(0)} (k_i \mu_{2\max} C_{X_2}^{(0)} Y_g - k_i Q^{(0)} - Q^{(0)} C_{S_2}^{(0)})}{Q^{(0)} (k_i + C_{S_2}^{(0)})} \qquad (13)$$

As $k_{S_2} > 0$, one necessarily has

$$k_i \mu_{2\max} C_{X_2}^{(0)} Y_g > (k_i Q^{(0)} + Q^{(0)} C_{S_2}^{(0)}). \qquad (14)$$

Determining the μ_{1max} and Y_2 Values

These values have been determined in two steps: *Determining the values $C_{S_2}^{(1)}$ and $C_{S_2}^{(2)}$*. We have to find the first two derivatives for all state variables in model (12).

For this purpose we use the symbolic computational tools in MATHEMATICA 3.0. Using the information for the known values $\mathbf{p_1}$, $C_{X_2}(0), C_{s_2}(0)$, $\mathrm{Q}(0), Q^{(1)}$, and $Q^{(2)}$ we can find exact expressions for the derivatives $Q^{(1)}$, $Q^{(2)}$, $C_{X_2}^{(2)}$, and $C_{S_2}^{(2)}$.

To calculating $C_{X_2}^{(1)}$ directly from the fourth equation of model (12), we assume that all parameters and initial values in this equation are known.

For the first derivative $Q^{(1)}$ we obtain the expression: Determining $C_{S_2}^{(1)}$ from the Eq. (15):

$$Q^{(1)} = \frac{k_i \mu_{2\max}\left(C_{X_2}^{(1)} S_2^{(0)}\left(k_{S_2} + C_{S_2}^{(0)}\right)(k_i + C_{S_2}^{(0)}) + C_{S_2}^{(1)}\left(k_{S_2} k_i - (C_{S_2}^{(0)})^2\right) C_{X_2}^{(0)} Y_g\right)}{\left(k_{S_2} + C_{X_2}^{(0)}\right)^2 \left(k_i + C_{X_2}^{(0)}\right)^2} \tag{15}$$

$$C_{s_2}^{(1)} = \frac{(k_{S_2} + C_{S_2}^{(0)})^2 (k_i + C_{S_2}^{(0)})^2 \left(Q^{(0)} - \frac{C_{x_2}^{(1)} k_i \mu_{2\max} C_{S_2}^{(0)} Y_g}{(k_{S_2} + C_{S_2}^{(0)})(k_i + C_{S_2}^{(0)})}\right)}{k_1 \mu_{2\max} (k_{S_2} k_i - C_{S_2}^{(0)^2}) C_{X_2}^{(0)} Y_g} \tag{16}$$

For the second derivative $C_{X_2}^{(2)}$, we obtain:

$$C_{X_2}^{(2)} = C_{X_2}^{(1)}\left(-k_2 + \frac{k_i \mu_{2\max} C_{S_2}^{(0)}}{(k_{S_2} + C_{S_2}^{(0)})(k_i + C_{S_2}^{(0)})}\right) + \frac{C_{S_2}^{(1)} k_i \mu_{2\max} (k_i k_{S_2} - (C_{S_2}^{(0)})^2) C_{X_2}^{(0)}}{(k_{S_2} + C_{S_2}^{(0)})^2 (k_i + C_{S_2}^{(0)})^2} \tag{17}$$

Determining $C_{S_2}^{(2)}$ from the equation for the second derivative $Q^{(2)}$:
In Eq. (18) only the value $C_{S_2}^{(2)}$ is unknown.

$$\begin{aligned} Q^{(2)} = {} & \frac{Y_g k_i \mu_{2\max} (2 C_{X_2}^{(1)} C_{S_2}^{(1)} (k_i + C_{S_2}^{(0)})(k_{S_2} + C_{S_2}^{(0)})(k_i k_{S_2} - C_{S_2}^{(0)^2}))}{(k_{S_2} + C_{S_2}^{(0)})^3 (k_1 + C_{S_2}^{(0)})^3} \\ & - \frac{Y_g k_i \mu_{2\max} (2 (C_{S_2}^{(1)})^2 (k_i k_{S_2} (k_i + k_{S_2}) + 3 k_i k_{S_2} C_{S_2}^{(0)} - C_{S_2}^{(0)^3}) C_{X_2}^{(0)})}{(k_{S_2} + C_{S_2}^{(0)})^3 (k_1 + C_{S_2}^{(0)})^3} \\ & + \frac{Y_g k_i \mu_{2\max} C_{X_2}^{(2)} C_{S_2}^{(0)}}{(k_{S_2} + C_{S_2}^{(0)})(k_i + C_{S_2}^{(0)})} + \frac{Y_g k_i \mu_{2\max} C_{X_2}^{(0)} C_{S_2}^{(0)} (k_i k_{S_2} - (C_{S_2}^{(0)})^2)}{(k_{S_2} + C_{S_2}^{(0)})^3 (k_i + C_{S_2}^{(0)})^3}. \end{aligned} \tag{18}$$

Constructing and solving the set of algebraic equations in respect $\mu_{1\max}$ and Y_2. The values $C_{S_2}^{(1)}$ and $C_{S_2}^{(2)}$ have been determined. Then we can express these values in terms of derivatives $C_{X_1}^{(1)}$, $C_{X_1}^{(2)}$, $C_{S_0}^{(1)}$, $C_{S_1}^{(1)}$, $C_{S_1}^{(2)}$, and the value $C_{X_2}^{(1)}$ (Eq. [12]). Thus, we can write the following system of equation:

$$C_{S_2}^{(1)} = A_1 \mu_{1\max} + \frac{A_2}{Y_2} \tag{19}$$

$$C_{S_2}^{(2)} = \mu_{1\max}(A_3 + A_4 \mu_{1\max}) + \frac{A_5}{Y_2}, \tag{20}$$

where A_1, A_2, A_3, A_4, and A_5 are known constants because they are functions of known variables:

$$A_1 = \frac{C_{X_1}^{(0)} C_{S_1}^{(0)} Y_b}{k_{S_1} + C_{S_1}^{(0)}};$$

$$A_2 = -\frac{k_i \mu_{2\max} C_{S_2}^{(0)} C_{X_2}^{(0)}}{(k_{S_2} + S_2^0)(k_i + S_2^0)}$$

$$A_3 = \frac{Y_b C_{X_1}^{(0)} (-k_i C_{S_1}^{(0)} (k_{S_i} + C_{S_1}^{(0)}) + k_{S_1} C_{S_0}^{(1)})}{(k_{S_1} + C_{S_1}^{(0)})^2};$$

$$A_4 = \frac{Y_b C_{X_1}^{(0)} C_{S_1}^{(0)} \left(C_{S_1}^{(0)} - \frac{k_{S_1} C_{X_1}^{(0)}}{(k_{S_1} + C_{S_1}^{(0)}) Y_1}\right)}{(k_{S_1} + C_{S_1}^{(0)^2})}$$

$$A_5 = \frac{k_1 \mu_{2\max} C_{X2}^{(1)} C_{S_2}^{(0)} C_{X2}^{(0)}}{(k_{S_2} + C_{S_2}^{(0)})(k_i + C_{S_2}^{(0)})} - \frac{k_1 \mu_{2\max} C_{X2}^{(0)} C_{S_2}^{(1)} \left(-k_i k_{S_2} + (C_{S_2}^{(0)})^2\right)}{(k_{S_2} + C_{S_2}^{(0)})^2 (\mathrm{k_i} + C_{S_2}^{(0)})^2}$$

Because Eq. (20) is quadratic in respect to $\mu_{1\max}$, it is impossible to find a unique solution in respect to this parameter. After applying the symbolic computation in MATHEMATICA, it is possible to observe that there are, at most, two solutions for the coefficients $\mu_{1\max}$ and $\mathrm{Y_2}$:

$$\mu_{1\max}^{1,2} = -\frac{1}{2 A_2 A_4}(A_2 A_3 - A_1 A_5 \pm \sqrt{Discr})$$

$$Y_2^{1,2} = -\frac{A_2}{A_1 \mu_{1\max}^{1,2} - C_{S_2}^{(1)}} \tag{21}$$

As the specific growth rate μ_{1max} has positive real values, the following conditions should be satisfied to guarantee at least one solution for this parameter:

$$Discr = A_1^2 A_5^2 - 2A_2A_5(A_1A_3 + 2A_4C_{S_2}^{(1)}) + A_2^2(A_3^2 + 4A_4C_{S_2}^{(2)}) > 0 \text{ or} \sqrt{Discr} \approx 0 \quad (22)$$

$$\frac{1}{2A_2A_4}(A_2A_3 + A_1A_5) > 0 \quad (23)$$

Analogously, the parameter Y_2 will be positive if the following condition is satisfied:

$$A_1\mu_{1\,max}^{1,2} - C_{S_2}^{(1)} > 0. \quad (24)$$

Our conclusion from the theoretical identifiability analysis is that if the inequalities (14), (22), (23), and (24) hold, the model parameters μ_{1max} and Y_2 are locally identifiable with at most two solutions, whereas k_{S_2} is uniquely identifiable.

We are grateful to Ivan Simeonov for allowing us to use the experimental data sets obtained at the Central Laboratory of Bioinstrumentation and Automation, BAS, Sofia.

NOMENCLATURE

$C_{S_{0i}}$	volatile solids in the influent (g/L)
C_{S_0}, C_{S_1}, C_{S_2}	concentrations of volatile solids, soluble volatile solids, and volatile fatty acids (mg/L)
C_{X_1}, C_{X_2}	concentrations of acidogenic and methanogenic bacteria (mg/L)
D	dilution rate (day^{-1})
k_1, k_2	decay coefficients for acidogenic and methanogenic bacteria (day^{-1})
k_{S_1}, k_{S_2}	saturation constants for acidogenic and methanogenic bacteria (mg/L)
k_i	inhibition coefficient for methanogenic bacteria (mg/L)
p	parameter vector
Q	biogas production rate (L/day)
x	vector of state variables
y	output
Y_1, Y_2	yield coefficients for acidogenic (mg organism/mg soluble organics) and methanogenic (mg organism/mg volatile acids) bacteria
Y_b	yield coefficient for the yield of volatile acids from soluble organics (mg volatile acids/mg organism)
Y_g	yield coefficient with respect to the gaseous output (L^2mg^{-1})
Y_p	fraction of volatile solids in the influent that can be solublized (mg/mg)
β	solubilization rate per unit of acidogenic biomass (L/mg day)
μ_1, μ_2	specific growth rate of acidogenic and methanogenic bacteria (day^{-1})
μ_{1max}, μ_{2max}	maximum specific growth rate for acidogenic and methanogenic bacteria (day^{-1})

References

Ahring BK. 1995. Methanogenesis in thermophilic biogas reactors. Antonie van Leeuwenhoek 67:91–102.

Ainsworth S. 1977. Steady state enzyme kinetics. London: Macmillan Press.

Angelidaki I, Ellagard L, Ahring BK. 1993. A mathematical model for dynamic simulation of anaerobic digestion of complex substrates: focusing on ammonia inhibition. Biotechnol Bioeng 42: 159–166.

Angelidaki I, Ellagard L, Ahring BK. 1999. A comprehensive model of anaerobic bioconversion of complex substrates to biogas. Biotechnol Bioeng 63:363–372.

Bastin G, Dochain D, Haest M, Installe M, Opdenacker Ph. 1982. Modelling and adaptive control of a continuous anaerobic fermentation process. In: Holm A, editor. Proceedings of IFAC symposium modeling and control of biotechnical processes. Helsinki, Finland:299–306.

Chalon A, Bastin G, Installe M. 1982. Identification of a biomethanization process: a case study. In: IFAC Symp on identification and system parameter estimation. Washington DC USA: p 409–413.

Chappell MJ, Godfrey KR, Vajda S. 1990. Global identifiability of the parameters of nonlinear systems with special inputs: a comparison of methods. Math Biosci 102:41–73.

Cox DR, Hinkley CV. 1974. Theoretical statistics. In: Bekeg GA, Soredis GU, editors. Chapman & Hall.

Denis-Vidal L, Joly-Blanchard Gh, Noiret C. 2001. Some effective approaches to check the identifiability of uncontrolled nonlinear systems. 57:35–44.

Dochain D, Vanrolleghem PA, van Daele M. 1995. Structural identifiability of biokinetic models of activated sludge respiration. Wat Res 11:2571–2578.

Forster C, Wase D. 1990. Environmental biotechnology. Chichester. Ellis Horwood Limited.

Ghaly AE, Pyke JB. 1991. Amelioration of methane yield in cheese whey fermentation by controlling the pH of the methanogenic stage. Appl Biochem Biotechnol 27:217–237.

Godfray KR, DiStefano JJ. 1985. Identifiability of model parameters. In: identification and system parameter estimation. Oxford: Pergamon Press. p 89–l14.

Hansen KH, Angelidaki I, Ahring BK. 1998. Anaerobic digestion of swine manure: inhibition by ammonia. Wat Res 32:5–12.

Hill DT, Barth CL. 1977. A dynamical model for simulation of animal waste digestion. J Water Pollution Contr Fed 10:2129–2143.

Hill DT, Tollner EW, Holmberg RD. 1983. The kinetics of inhibition in methane fermentation of swine manure. Agric Wastes 5:105–123.

Holmberg A. 1982. On the practical identifiability of microbial growth models incorporating Michaelis-Menten type nonlinearities. Math Biosci 62:23–43.

Husain A. 1998. Mathematical models of the kinetics of anaerobic digestion—a selected review. Biomass Bioenergy 14:561–571.

Jeyaseelan S. 1997. A simple mathematical model for anaerobic digestion process. Wat Sci Tech 35:185–191.

Joly-Blanchard G, Denis-Vidal L. 1998. Some remarks about an identifiability results of nonlinear systems. Automatica 34:1151–1152.

Julien S, Babary JP, Lessard P. 1998. Theoretical and practical identifiability of a reduced order model in an activated sludge process doing nitrification and denitrification. Wat Sci Technol 37:309–316.

Kalyuzhnyi S, Veeken A, Hamelers B. 2000. Two-particle model of anaerobic solid state fermentation. Wat Sci Technol 41:43–50.

Kiely G, Tayfur G, Dolan C, Tanji K. 1997. Physical and mathematical modelling of anaerobic digestion of organic wastes. Wat Res 31:534–541.

Kus F, Wiesmann V. 1995. Degradation kinetics of acetat mixed cultures. Wat Res 29:1427–1443.

Ljung L, Glad T. 1994. On global identifiability for arbitrary model parametrization. Automatica 30:265–276.

Masse DI, Droste RL. 2000. Comprehensive model of anaerobic digestion of swine manure slurry in a sequencing batch reactor. Wat Res 34:3087–3106.

Merkel W, Manz W, Szewzyk U, Krauth K. 1999. Population dynamics in anaerobic wastewater reactors: modeling and in situ characterization. Wat Res 33:2392–2402.

Möche M, Jördening HJ. 1999. Comparison of different models of substrate and product inhibition in anaerobic digestion. Wat Res 33:2545–2554.

Nopens I, Hopkins LN, Vanrolleghem PA. 2001. An overview of the posters, presented at Watermatex 2000. III: Model selection and calibration/optimal experimental design. Wat Sci Technol 43:387–389.

Noykova N, Gyllenberg M. 2000. Sensitivity analysis and parameter estimation in a model of anaerobic waste water treatment processes with substrate inhibition. Bioproc Eng 23:343–349.

Pohjanpalo H. 1982. Identifiability of deterministic differential models in state space. An implementation for a computer. (Research report 56). Espoo: Technical Research Center of Finland.

Pollard PC, Greenfield PF. 1997. Measuring in situ bacterial specific growth rates and population dynamics in wastewaters. Wat Res 31:1074–1082.

Press W, Teukolsky S, Flannery B, Wetterling W. 1992. Numerical recipes. Cambridge, UK: Cambridge University Press.

Schürbüscher D, Wandrey C. 1991. Anaerobic waste water process models. In: Schügerl K., editor. Biotechnology, a multi-volume compehensive treatise, vol. 4: Measuring, modelling and control. Weinheim: VCH. p 445–484.

Simeonov I, Momchev V, Grancharov D. 1996. Dynamic modelling of mesophilic anaerobic digestion of animal waste. Wat Res 30:1087–1094.

Thomas MV, Nordstedt RA. 1993. Generic anaerobic digestion model for the simulation of various reactor types and substrates. Trans ASAE 36:537–544.

Vajda S, Godfrey KR, Rabitz H. 1989. Similarity transformation approach to identifiability analysis of nonlinear compartmental models. Math Biosci 93:217–248.

Vanrolleghem PA, Dochain D. 1998. Bioprocess Model Identification. In: Van Impe JFM, Vanrolleghem PA, Iserentant D, editors. Advanced instrumentation, data interpretation, and control of biotechnological processes. Dordrecht: Kluwer. p 251–318.

v. Münch E, Keller J, Lant P, Newell R. 1999a. Mathematical modelling of prefermenters—I. Model development and verification. Wat Res 33:2757–2768.

v. Münch E, Keller J, Lant P, Newell R. 1999b. Mathematical modelling of prefermenters—II. Model applications. Wat Res 33:2844–2854.

Wang DM. 1995. An implementation of the characteristic set method in MAPLE. In: Pfalzgraf J, Wang D, editors. Automated practical reasoning: algebraic approaches. Springer. New York: p 187–201.

Yordanova ST, Noykova NA. 1996. Influence of perturbations on the waste water treatment process. Chem Biochem Eng Q 10:9–14.