CHEMICAL ENGINEERING TRANSACTIONS Volume 17, 2009
Editor Sauro Pierucci
 ISBN 978-88-95608-01-3 ISSN 1974-9791
DOI: 10.3303/CET0917216

Parameter identification in the activated sludge process

Päivi Holck, Aki Sorsa and Kauko Leiviskä
Control Engineering Laboratory, University of Oulu
P.O.Box 4300, 90014 Oulun yliopisto, Finland
e-mail: kauko.leiviska@oulu.fi

Numerical optimisation methods are common in nonlinear parameter identification, but they may end up in a local optimum value. Genetic algorithms are able to escape local optima and continue the search towards a global optimum value. This paper shows that they are useful in identifying parameters for the nonlinear model of the activated sludge plant. The number of identified parameters in the model is seven: three coefficients in Haldane's equation, two yields, oxygen kinetic constant, and the oxygen transfer coefficient. The results are good, and show a good correlation with the actual values.

1. Introduction

There are several models for the biological wastewater treatment in the literature. Common features in these models are nonlinear dynamics, high number of parameters describing the kinetics and stoichiometry of the reactions, several operating points, sensitivity to parameter values, and so on.
Dissolved oxygen (DO) control is critical for the good operation of activated sludge processes. Too high DO leads to high energy consumption and bad quality, and too low DO leads to low pollutant removal. Several control methods are in use – starting from conventional PID controls to model predictive control and various adaptive or fuzzy control strategies (Ertunc et al., 2009, Oliveira et al., 2005). More advanced control methods need models for the controller design and tuning and face the problems with parameter identification of nonlinear activated sludge models.
This paper introduces the use of genetic algorithms for the nonlinear parameter identification of the activated sludge process. The data for modelling was produced by a Chemostat model built in Simulink® environment. The model is based on mass balances and Haldane kinetics. The parameters identification uses the values of dissolved oxygen concentration as the system output, because its measurement in practice is easier than other state variables and, as mentioned before, it is also a common control variable.

2. Process model

In this paper, data for identification trials are produced using the model of the aeration basin of the activated sludge process. The activated sludge process is modelled as a Chemostat bioreactor using substrate, biomass, and dissolved oxygen mass balances

Please cite this article as: Holck P., Sorsa A. and Leiviska K., (2009), Parameter identification in the activated sludge process, Chemical Engineering Transactions, 17, 1293-1298 DOI: 10.3303/CET0917216

(Equation 1) and Haldane kinetics (Equation 2) that describe the inhibition of the reaction by high substrate concentration.

$$\begin{aligned}\frac{dS}{dt} &= \frac{Q}{V}\left(S_{in} - S\right) - Y_{X/S}\mu(S)X \\ \frac{dX}{dt} &= \frac{Q}{V}\left(X_{in} - X\right) + \mu(S)X \\ \frac{dO}{dt} &= \frac{Q}{V}\left(O_{in} - O\right) - Y_{X/O}\mu(S)X - k_o X + k_L a\left(O_s - O\right)\end{aligned} \tag{1}$$

$$\mu(S) = \frac{\mu_{\max} S}{K_I^{-1} S^2 + S + K_S}. \tag{2}$$

In the above equations, S, X, and O are the concentrations of substrate, biomass, and dissolved oxygen, the subscript *in* refers to their values in the input flow, Q is the flow rate, V is the liquid volume in the reactor, and Y is the yield.

Table 1 contains the values for eight parameters shown in the above equations: the maximum growth rate (μ_{max}), saturation coefficient (K_s), inhibition constant (K_I), yield coefficients ($Y_{X/S}$, $Y_{X/o}$), oxygen kinetic constant (k_O), volumetric oxygen transfer rate (k_La) and the saturation concentration for dissolved oxygen (O_s). With the chosen parameter set, the system shows nonlinear bi-stable dynamics, i.e. it has two stable operation points (Holck, 2008) (Sorsa and Leiviskä, 2006).

Table 1. Parameter values for Equations 1 and 2 (Holck, 2008).

Parameter	μ_{max}	K_S	K_I	$Y_{X/S}$	$Y_{X/O}$	k_O	k_La	O_s
Value	1.04	3.75	60	1.40	1.03	0.006	16.8	6

The model simulator is a Simulink® program. The solver used in the simulations is ode45 that is capable of solving nonstiff differential equations, and uses the Runge-Kutta method.

3. Data

The process has two distinct operation points: low conversion (OP 1) and high conversion area (OP 2). Identification data sets were generated for both points by applying two consecutive step changes of sizes +5 and -10 to the input concentration of the substrate. This approach should also reveal the nonlinearity of the system.

Random disturbances were applied to the input concentration of substrate (S_{in}), and biomass (X_{in}), and to the liquid volume in the reactor (V). The disturbances were taken from the normal distribution and their means and variances are in Table 2. The disturbance to V is very small since it was assumed that the liquid level in the reactor was in control.

Table 2. The means and the variances of disturbances.

Variation in	S_{in}	X_{in}	V
Mean	0	10	100
Variance	1	0.4	1

Table 3 shows the total simulation time (t_{sim}), the steady-state input concentration of the substrate, and the size and time of the step changes in both operation points. The total simulation time for the second operation point is longer than for the first operation point. This is due to working in the low conversion area and with less biomass in the system. Therefore, the reactions occur more slowly.

Table 3. Simulation time, substrate input concentrations and sizes and times for the step changes in both operation points.

			Step change 1		Step change 2	
	t_{sim}	S_{in}	Size	Time	Size	Time
OP1	120	40	5	50	-10	100
OP2	275	75	5	100	-10	200

The identification bases on the concentration of dissolved oxygen. There are two reasons for this: in practice, it is the easiest output variable to measure and secondly, as mentioned in the introduction, several control strategies use the concentration of dissolved oxygen as the controlled variable. Figures 1 and 2 later show the step responses of dissolved oxygen in both operation points.

4. Identification with Genetic Algorithms

The model parameters in this paper are identified by using the real-coded genetic algorithms. Genetic algorithms are an optimization method mimicking evolution. The population consisting of chromosomes evolves towards the global optimum. The possible solutions to the optimization problem are encoded into the chromosomes. Binary or real-valued coding can be used. The link between the chromosomes and the problem is the objective function, the fitness function.

The main stages in genetic optimization are parent selection, fitness evaluation, variation, and population update. These stages are repeated until the optimum has been reached. Typical parent selection mechanisms are the tournament and the roulette wheel mechanisms (Davis, 1991). In tournament mechanism, a certain number of chromosomes are selected randomly to participate in a tournament. The most suitable chromosome is the winner and is selected as a parent. In the roulette wheel method, the fitter chromosomes have a larger slot in the roulette, and have therefore better chances to become a parent. The genetic operators – crossover and mutation – regulate the variation: The crossover operator mates two parents to produce the offspring (1 or 2, depending on the methods). Mutation adds random changes to the population so that the optimization is not trapped into any local optimum.

Genetic algorithms (GA) have been used for both the structure and parameter identification (Gray et al., 1998, Nyarko and Scitovski, 2004, Chang, 2007, Khalik et

al., 2007, Wang et al., 2008). The most significant drawback of conventional methods (such as gradient methods) is the lack of ability to overcome local optima. That problem does not exist with GA since the optimal solutions are searched from multiple directions and, basically, the whole search space is covered. Therefore, GA are more likely to find the global optimum. The drawbacks of GA are that the exact solution may not be found and the optimization is time-consuming with complex systems.

The parameter identification was performed for both operation points by using two different crossover methods: the linear crossover and the arithmetic crossover with crossover probability of 0.9. The mutation method was the uniform mutation with the mutation probability of 0.05. The initial population was taken randomly from the uniform distribution and each generation consisted of 200 chromosomes. Elitism was also included and from each generation the best chromosome was transferred to the next to ensure that the best result will not disappear from the population. The initial population was subjected to genetic operations for 30 generations after which the best chromosome was returned by the algorithm. To validate the results of the parameter identification, the optimization was repeated for both operation points (and with both crossover methods) 500 times.

The fitness function was the sum of the squared prediction error, SSE that compared the predicted value of dissolved oxygen with its actual value.

Seven parameters were identified from the step response data, i.e. all parameters in Table 1 except O_s. The feasible range allotted for the parameters is very important in genetic optimization, and the ranges for the seven parameters are in Table 4.

Table 4. Parameter ranges for genetic optimization.

Parameter	μ_{max}	K_S	K_I	$Y_{X/S}$	$Y_{X/O}$	k_O	k_La
Range	0-2	0-10	30-90	0-5	0-20	0-0.1	10-40

5. Results and discussion

Table 5 shows the best parameter sets for both operation points and crossover methods. The table shows also differences between the actual parameters and the identified parameters in percentages. Table 6 shows the statistical values of the fit: SSE, Mean error (EM), Standard deviation (SD), and correlation (R2) between the predicted and actual responses. It also shows the number of iterations (NI), where the best parameter sets were achieved.

Figures 1 and 2 together with the correlation coefficients in Table 6 show that genetic algorithms perform well in parameter identification. There is also only a little difference between the used crossover methods even though some of the identified parameter values differ a lot from the actual values. There are two reasons for the big differences: the random disturbances applied to the system inputs in data sets generation and also the fact that the parameters are identified for both operation points separately. The parameter that differs the most from the actual parameter values is k_O regardless of the crossover method used. Its value is very small and its search area is much larger than the relative search areas for the other parameters.

Table 5. The best set of parameter values and their differences from the actual ones for both crossover methods (CM) and operation points (OP). L means Linear Crossover and A means Arithmetic Crossover. Difference is in percentages.

Parameter	OP	CM	μ_{max}	K_S	K_I	$Y_{X/S}$	$Y_{X/O}$	k_O	k_La
Value	1	L	0.91	7.16	79.2	1.65	1.38	0.016	18.6
Difference			-12.5	90.9	32.1	17.9	34.0	167	10.7
Value		A	0.97	6.16	72.5	1.55	2.11	0.05	33.1
Difference			-6.7	64.3	20.8	10.7	105	733	97.0
Value	2	L	0.74	7.13	33.1	0.007	1.46	0.013	25.2
Difference			-28.8	90.1	-44.8	-99.5	-41.7	117	49.9
Value		A	0.60	6.73	81.6	1.82	2.43	0.045	32.9
Difference			-42.3	79.5	35.9	30.0	136	651	96.1

Table 6. Statistical values for the model validation.

OP	CM	SSE*10^{-4}	EM*10^{-3}	SD*10^{-3}	R2	NI
1	L	2.37	2.45	4.48	0.998	462
	A	2.96	7.08	5.96	0.998	397
2	L	2.40	1.94	2.36	0.991	142
	A	2.92	2.81	1.77	0.990	105

In the future, the parameter search areas should be studied more to ensure that the parameter ranges do not include impossible values for the real process. This would help to cut out unrealistic parameter sets. It would also be essential to perform sensitivity analysis for the model parameters. SSE is only one of the many possible fitness functions. Other functions should be tested to find out whether different fitness functions have a significant influence to the identification results. During the optimization, the fitness function is evaluated hundreds of times. Cutting down the runtime of the whole genetic algorithm is important to improve the efficiency of the solution.

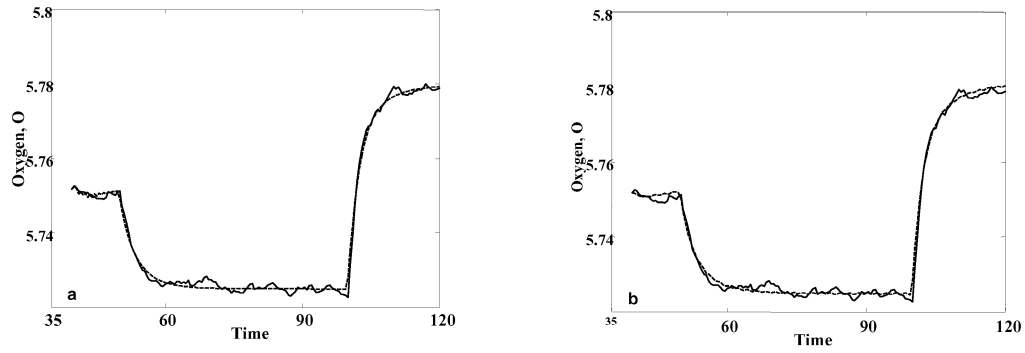

Figure 1. The actual and predicted concentration of dissolved oxygen for Operation Point 1. Linear Crossover (a) and Arithmetic Crossover (b). Denotations: --- predicted, — actual.

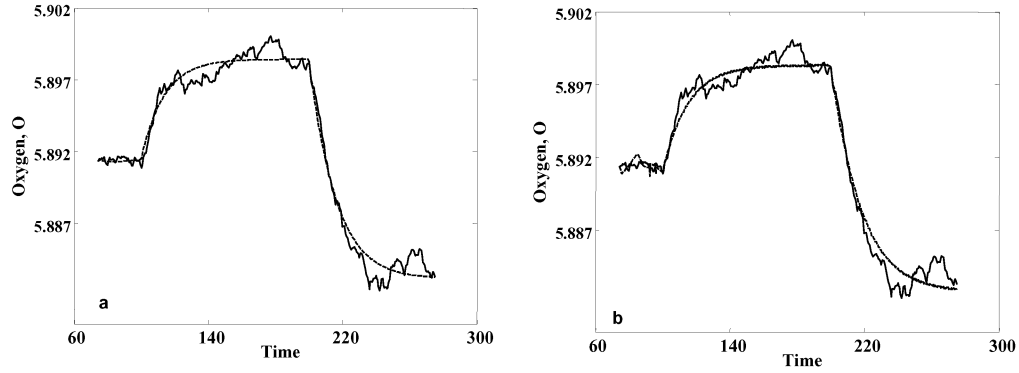

Figure 2. The actual and predicted concentration of dissolved oxygen for Operation Point 2. Linear Crossover (a) and Arithmetic Crossover (b). Denotations: --- predicted, — actual.

References

Sorsa A. and Leiviskä K., 2006, State detection in the biological water treatment process. Oulu, Control Engineering Laboratory. 53. Report A. 31. ISBN 951-42-8273-6/ISSN 1238-9390.

Holck P., 2008, Parameter identification with genetic algorithms for biological wastewater treatment process. Master's thesis, Control Engineering Laboratory, University of Oulu, Finland.

Gray G.J., Murray-Smith D.J., Li Y., Sharman K.C and Weinbrenner T., 1998, Nonlinear model structure identification using genetic programming, Control Engineering Practice 6, 11, 1341-1352.

Chang W.-D., 2007, Nonlinear system identification and control using a real-coded genetic algorithm, Applied Mathematical Modelling 31, 3, 541-550.

Wang G., Feng E. and Xiu Z., 2008, Modeling and parameter identification of microbial bioconversion in fed-batch cultures, Journal of Process Control 18, 12, 1323-1326.

Khalik M.A., Sherif M., Saraya S. and Areed G., 2007, Parameter identification problem: Real-coded GA approach, Applied Mathematics and Computation 187, 2, 1495-1501.

Nyarko E.K. and Scitovski R., 2004, Solving the parameter identification problem of mathematical models using genetic algorithms, Applied Mathematics and Computation 153, 3, 651-658.

Davis L., 1991, Handbook of Genetic Algorithms. Van Nostrand Reinhold, New York.

Ertunc S., Akay B., Boyacioglu H. and Hapoglu H., 2009, Self-tuning control of dissolved oxygen concentration in a batch bioreactor, Food and Bioproducts Processing (in press).

Oliveira R., Clemente J.J. Cunha A.E. and Carrondo, M.J.T., 2005, Adaptive dissolved oxygen control through the glycerol feeding in a recombinant *Pichia pastoris* cultivation in conditions of oxygen transfer limitation, Journal of Biotechnology 116, 35–50.