On-Line Estimation of Parameters in a Respiration Model Using a Continuous-Discrete Type of Recursive Prediction Error Algorithm.

J.D. Stigter*, P.A. Vanrolleghem†, and J.F. Van Impe*

Department of Food and Microbial Technology, Katholieke Universiteit Leuven, Kardinaal Mercierlaan 92, B-3001 Heverlee, Belgium, Fax: +32 16 321960, e-mail: hans.stigter@agr.kuleuven.ac.be

† *Department of Agriculture, BIOMATH, University of Gent, Coupure Links 653, B-9000 Gent, Belgium, Fax: +32 9 2234941, e-mail: peter.vanrolleghem@rug.ac.be*

Keywords : recursive parameter estimation, prediction error methods, Monod kinetics, fault detection

Abstract

Estimates of biokinetic parameters for an activated sludge respiration model are obtained using an 'on-line' combined state/parameter estimation algorithm. The analysis suggests that the continuous-discrete type of recursive prediction error algorithm can be used to (i) obtain improved estimates of certain identifiable parameter combinations based on oxygen data only, and (ii) to evaluate the model 'on-line' through recursive estimates of the steady state Kalman gain matrix K_∞.

1 Introduction

The estimation of biokinetic parameters from dissolved oxygen data *only* is an increasingly topical subject of study and defines one of the challenges in the modelling and control of biological processes [2, 10]. This paper discusses the problem of 'on-line' estimation of biokinetic growth parameters on basis of data collected by a Rapid Oxygen Demand TOXicity device (RODTOX), [9]. The data set used in this paper has been studied earlier with special reference to the structural identifiability of various activated sludge models in an earlier study [1, 2, 10] of which the analysis of the identifiability of the parameters in the Michaelis-Menten kinetics is the most interesting result. The approach to the parameter estimation problem by Vanrolleghem *et al.* [10] is the application of 'off-line' parameter estimation techniques. This study is extended in this contribution using an 'on-line' state/parameter reconstruction algorithm which will be applied to a synthetically generated data set of a Single Monod model only. The 'on-line' or recursive approach provides a good fault detection framework to sense a change of metabolism in the bioreactor since the 'on-line' estimates of the time-invariant parameters can, indeed, reveal sudden shifts relating to an inappropriate candidate model structure. More specifically, the applied recursive algorithm assumes a parameterized steady state Kalman gain K_∞ which weights the innovation $\epsilon(t_k)$ and can be utilized as a fault detection mechanism since deflections in the steady state gain reveal an increase in the uncertainty in the model structure $f(x, u, \alpha; t)$ as a whole in the assumed general model structure

$$\dot{x}(t) = f(x, u, \alpha; t) + \xi(t) \quad (1)$$

$$y(t_k) = h(x, u, \alpha; t_k) + \eta(t_k) \quad (2)$$

where $\xi(t)$ and $\eta(t_k)$ are continuous time and discrete time stochastic noise processes respectively, $x(t)$ is the n-dimensional state vector, α is the p-dimensional (time-invariant) parameter vector, $y(t_k)$ is the m-dimensional observation vector, and $u(t_k)$ is the r-dimensional input vector, while the vector functions $f(x, u, \alpha; t)$ and $h(x, u, \alpha; t_k)$ are the continuous time dynamical model and the discrete time observation function respectively.

The challenge of the RODTOX data set lies in its limitation. It can be shown that reconstruction of all individual parameters in the Single-Monod respiration model (to be defined in section 1.2), i.e. the yield coefficient Y, the saturation constant K_m, the biomass $X(t)$, the substrate concentration $S(t_0)$, and the maximum specific growth rate μ_{max}, based on OUR data *only* is impossible [2]. Combinations of these parameters, however, are identifiable. A second challenge lies in the fact that the respiration model used is non-linear and this introduces another major hurdle for the application of a recursive algorithm since, in general, these algorithms (e.g. the extended Kalman filter) require a linearization of the model at each time step, introducing linearization errors that deteriorate the estimate. Application of a recursive algorithm (the continuous-discrete recursive prediction error algorithm (CDRPE, [7]) will be demonstrated here and help to identify the values of the parameters. It is finally noted that the discussion will focus on the CDRPE results obtained rather than the technicalities involved in the specific case study ([10]) and discussion of the algo-

rithmic development for which we refer to [7, 8].

1.1 An Activated Sludge Respiration Model

Consider a biomass (X) in an aerated tank feeding on an injected substrate of wastewater ($S_1(t)$). Assume that the concentration X is constant over the course of one experiment (in case of the RODTOX device this is a realistic assumption since the duration of one experiment, i.e. recovery to steady state after injection of a unit of wastewater, is usually short (30 mins)). A classical approach for a dynamical description of the reduction of the substrate concentration due biochemical growth of a biomass X is the Michaelis-Menten or Monod model

$$\frac{dS_1}{dt} = -\frac{\mu_{max} X}{Y} \frac{S_1}{K_m + S_1} \tag{3}$$

where time dependency of S_1 on t has been omitted for notational convenience. The oxygen uptake rate (OUR_{ex}) is defined as

$$OUR_{ex} = (1 - Y)\frac{dS_1}{dt} \tag{4}$$

In case of several substrate concentrations $\{S_i(t),\ i = 1, \ldots, k\}$, the total oxygen uptake rate (OUR_{ex}) is simply defined by the sum of the individual contributions, i.e. $OUR_{ex} = \sum_{i=1}^{k}(1 - Y_i)\frac{dS_i}{dt}$. Dochain *et al.* have demonstrated, using the Taylor series expansion method ([6]) for the function $OUR_{ex}(t)$, that in case of OUR_{ex} observations only the parameter combinations

$$\theta_1 = \frac{(1 - Y)\mu_{max} X}{Y} \tag{5}$$

$$\theta_2 = (1 - Y)S(0) \tag{6}$$

$$\theta_3 = (1 - Y)K_m \tag{7}$$

can be uniquely determined. This follows from the observation that all coefficients in the Taylor series expansion for $OUR_{ex}(t)$ can be written as combinations of the set $\{\theta_i, i = 1, 2, 3\}$.

In the RODTOX measurement device the data $\{OUR_{ex}(t_k), k = 0, 1, \ldots, N\}$ are generated from dissolved oxygen data $\{DO(t_k), k = 0, 1, \ldots, N\}$ for which derivatives have to be calculated. Here an unobserved state, including the substrate concentration, will be introduced that is estimated directly. The dynamics of the dissolved oxygen concentration ($x_1(t)$) in the reactor include exogenous respiration $OUR_{ex}(t)$, a re-aeration process (diffusion of oxygen across the air/water interface), and an endogenous respiration which is assumed constant. In summary, the DO process model reads

$$\dot{x}_1(t) = k_1(C_s^e - x_1(t)) - OUR_{ex} \tag{8}$$

where C_s^e is the corrected saturation concentration – the correction is a small factor that substracts the endogenous respiration from the saturation concentration, and k_1 is the re-aeration constant. Equations (4) and (8) constitute the two state model of the respiration process in the bioreactor. Note that only the first state is directly observable, while the second state is unobservable. At this point the previous mentioned identification results can be applied. Since the Taylor series expansion can be developed around *any* time instant t we can estimate the parameters $\{\theta_i, i = 1, \ldots, 3\}$, except that θ_2 should now be replaced by $\theta_2(t) = -(1 - Y)S_1(t)$. The parameters θ_1 and θ_2 are indeed time invariant and will be the same for a Taylor series expansion around any time instant t. Hence, it is natural to redefine the model for the respiration process by

$$\dot{x}_1(t) = k_1(C_s^e - x_1(t)) - \dot{\theta}_2(t) \tag{9}$$

$$\dot{\theta}_2(t) = -\frac{\theta_1 \theta_2(t)}{\theta_3 + \theta_2(t)} \tag{10}$$

The re-aeration constant k_1 can readily be determined after all substrate has been consumed, for example by fitting an analytical solution of the differential equation

$$\dot{x}_1(t) = k_1(C_s^e - x_1(t)) \tag{11}$$

to the final part of the data record – when all substrate has been consumed and re-aeration is the only active process in the above DO balance. Subsequently, the time invariant parameters θ_1 and θ_3 can be determined, together with the unobservable state $\theta_2(t)$. This will be done using the CDRPE in the next section. In case of two substrate concentrations the estimation can be solved backwards, i.e. first estimate k_1 from the final part of the data record, then estimate the parameters $\{\theta_i, i = 4, 5, 6\}$ of the second substrate (after the first substrate has been consumed), and finally estimate $\{\theta_i, \ldots, i = 1, 2, 3\}$ from the first part of the data record when both substrates are active. Strictly speaking such an approach is not 'on-line' in the sense we used it earlier but the data set is still processed recursively in such an approach. The only thing that changes is the fact that the data set is seperated into three parts which can be treated in reverse order. More elegant and certainly worth further investigation, is to develop a filter that processes the dataset backwards in time. Such a backwards filter could then be used to estimate first k_1 in reverse time until a structural collapse is observered through a deflection of the gain (K_∞) parameters. Then the re-aeration parameter can be fixed, i.e. $\forall k:\ P_{k_1 k_1}(t_k) = 0$, and the parameters θ_4, θ_5, and θ_6 can be estimated using the backwards filter. Having established these values a final (backwards) run of such a filter can be used to estimate $\{\theta_i, i = 1, 2, 3\}$. Finally, it is noted that, using the above recursive approach, 'on-line' estimates of $(1 - Y)S(t)$ become available so that the substrate concentration can be estimated once the yield is known.

2 Some Results and Discussion

Synthetic data were generated using $\theta_1 = 2.0$ and $\theta_3 = 1.5$ and measurement noise with an intensity (variance-covariance matrix) $R = 0.03$ was added to the observations. A typical 'a priori' example of the simulated and the updated estimates generated by the CDRPE is shown in figures 1 and 2, where $\theta_1(t_0) = 1.2$, and $\theta_3(t_0) = 3.0$ for the first run. In the second and third run (figures 3–6) the initial values $\theta_1(t_0)$ and $\theta_3(t_0)$ are set to the final values $\theta_1(t_N)$ and $\theta_3(t_N)$ from the previous run. In this case it was observed that the parameter estimates for the two parameters converge very close to the true values $\theta_1 = 2.0$ and $\theta_3 = 1.5$ which is a good result given the non-linear dynamics of the Michaelis-Menten model. It should be noted though that in some cases the initial estimates did not converge to the true values. In figure 4 the parameter estimates for the three runs are depicted in the same figure and one can see clearly the improvement in the obtained estimates.

An important note is that the (parametrized) steady state gain K_∞ was fixed to zero in the above example in order to maximize the changes in the parameters θ_1 and θ_3. The motivation for such an approach is that we want to use the recursive algorithm to correctly estimate the parameters θ_1 and θ_3 while the model structure itself is assumed given and known. In case of model structure identification, i.e. both the model and the parameters are only partly known 'a priori', it is normal to increase the 'a priori' uncertainty on the Kalman gain parameter. Some examples are presented in [7].

Metaphorically, one can think of the model structure $f(x, u, \alpha; t)$ as a plastic or other flexible structure which is continuously under stress while the CDRPE processes the data. Those parts of the flexible structure that have most elasticity will change first while other, more rigid parts in the structure resist these changes [8]. We also note that the flexibility of the individual components in the model and, more specifically, the 'a priori' uncertainty in the parameter estimates $\theta_1(t_0), \theta_3(t_0)$, and $K_\infty(t_0)$, can be specified through the associated leading diagonal elements. Thus, by specifying the variance-covariance element associated with $K_\infty(t_0)$ equal to zero we allow to let the data *only* massage, as it were, the model components associated with the two parameters θ_1 and θ_3. The values for $\theta_i(t_N)$ were in both cases considerably improved when compared to the initial estimates $\theta_1(t_0)$ and $\theta_3(t_0)$. Also, the uncertainty bounds for θ_3 associated with the Monod constant K_m hardly decreased while processing the data, contrary to the uncertainty bounds on θ_1 associated with the μ_{max} parameter.

Finally, the sensitivity coefficients that also become available 'on-line' during the filtering process are shown in figure 3. The graph shows that, indeed, the parameter μ_{max} can best be identified at the very beginning (when the growth is large) and the middle part of the data record, while K_S can best be identified in the middle of the data record. This confirms the well-known fact that, indeed, the substrate constant K_m is most difficult to estimate. Further, both the sensitivities coefficients return to their steady state values after the injection of wastewater has been processed by the RODTOX device, meaning that the magnitude of the dynamics described by the model $f(x, u, \alpha; t)$ are close to zero at the end of the data record.

3 Summary and Further Research

A continuous-discrete type of recursive prediction error algorithm has been applied to a (synthetically generated) respirogram. On basis of 'on-line' dissolved oxygen measurements only it was found that the parameter combinations θ_1, θ_2, and θ_3 can indeed be identified as predicted by the results of Dochain *et al.*. The CDRPE algorithm is demonstrated to be a good parameter estimator and this, in fact, is due to an improvement of the coupling between the parameter and the state vector that is omited in, for example, the familiar extended Kalman filter [5], still very frequently used in the literature as a parameter estimator for (non-linear) systems (see also [7]). The results of the CDRPE may further be improved using local smoothing techniques that improve the estimate at time t_k and, basically, alternate prediction and smoothing formulaes as to improve the current estimate [3, 4].

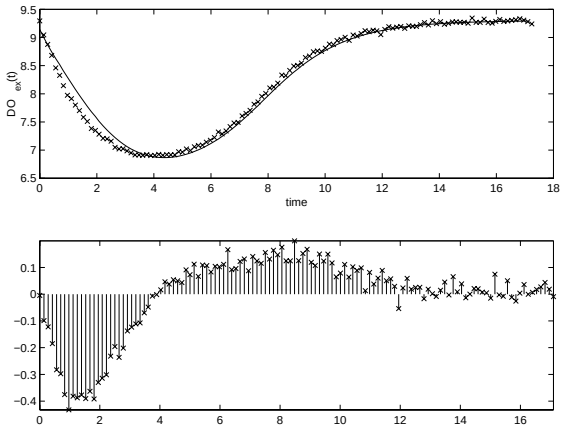

Figure 1: Dissolved Oxygen estimates generated by the CDRPE – first run.

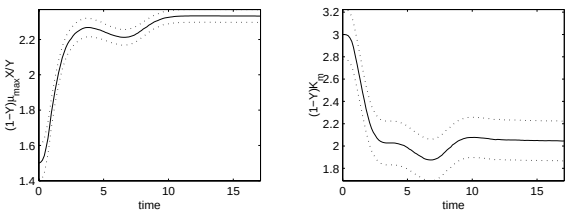

Figure 2: Parameter estimates for θ_1 and θ_3. Note the change in uncertainty for θ_1 when compared to θ_3.

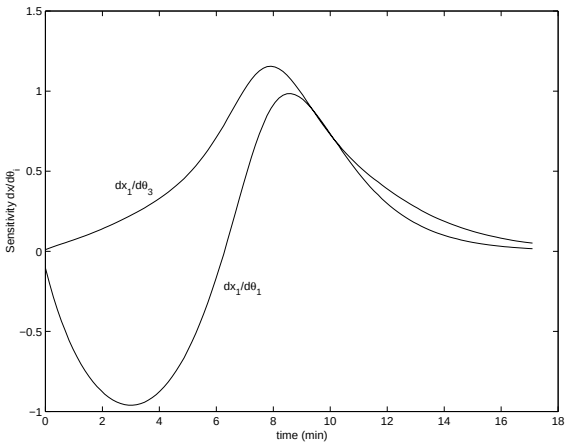

Figure 3: On-Line sensitivity estimates for θ_1 and θ_3

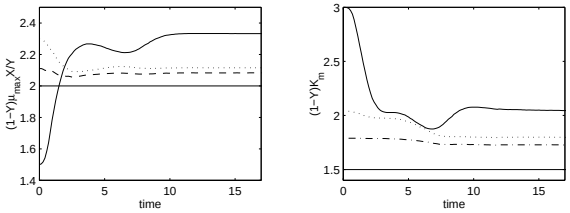

Figure 4: Parameter estimates developing over three runs when $\theta_{i+1}(t_0) = \theta_i(t_N),\ i = 1, 2$

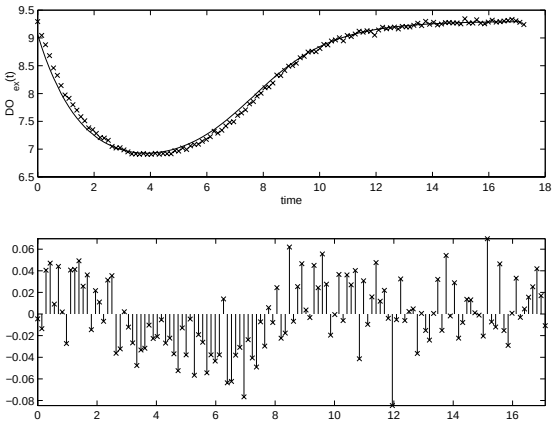

Figure 5: Dissolved Oxygen estimates generated by the CDRPE – final run.

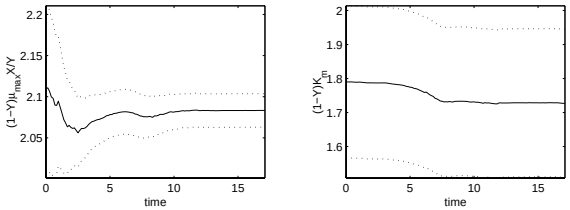

Figure 6: Parameter estimates for θ_1 and θ_3. Note the difference in uncertainties between the two parameters θ_1 and θ_3.

References

[1] Bastin, G., Dochain, D., "On-Line Estimation of Adaptive Control of Bioreactors", Elsevier, Amsterdam, (1990).

[2] Dochain D., Vanrolleghem, P.A., Van Daele, M., "Structural Identifiability of Biokinetic Models of Activated Sludge Respiration", Water Research, **29(11)**, 2571–2578, (1995).

[3] Gelb A., "Applied Optimal Estimation", M.I.T. Press, Massachusetts, Cambridge, (1974).

[4] Jazwinski, A.H., Stochastic Processes and Filtering Theory, **64**, Math. in Science and Engineering, Academic Press, New York, (1970).

[5] Ljung L., "Asymptotic Behaviour of the Extended Kalman Filter as a Parameter Estimator for Linear Systems", IEEE Transactions on Automatic Control, **AC-24**, 36–50, (1979)

[6] Pohjanpalo H., "System Identifiability Based on the Power Series Expansion of the Solution", Mathematical Biosciences, **Vol. 41**, 21–33, (1978)

[7] Stigter J.D., "The Development and Application of a Continuous-Discrete Recursive Prediction Error Algorithm in Environmental Systems Analysis", PhD Thesis, University of Georgia, (1997)

[8] Stigter J.D., Beck M.B., "Model Structure Identification: Development and Assessment of a Recursive Prediction Error Algorithm", IIASA Working Paper WP-95-105, International Institute for Applied Systems Analysis, Laxenburg, Austria, (1995)

[9] Vanrolleghem P.A., Van Daele M., "Optimal Experimental Design for Structure Characterization of Biodegradation Models: On-line Implementation in a Respirographic Biosensor", Water Science and Technology, **30(4)**, 243–253,(1994)

[10] Vanrolleghem P.A., Van Daele M., Dochain D.", "Practical Identifiability of a Biokinetic Model of Activated Sludge Respiration", Water Research, **29(11)**, 2561–2570, (1995).,