

Parameter Identification for a Mechanistic Model of Poly- β -hydroxybutyrate Production

Mark A. Pinto* and Charles D. Immanuel[†]

*Centre for Process Systems Engineering, Department of Chemical Engineering,
Imperial College London, South Kensington Campus, London SW7 2AZ, UK
e-mail: c.immanuel@imperial.ac.uk*

Abstract

The use of detailed models for bioprocess design and control has been limited as accurate estimation of model parameters is often difficult. In this paper, the parameter estimation problem for a mechanistic model of the production of a biopolymer, poly- β -hydroxybutyrate (PHB), is examined in detail. Parameter estimation was undertaken using previously published data. Three parameter sets were obtained with large uncertainty in some parameters. Parametric sensitivity, identifiability and estimability analyses indicated that only certain parameters were uniquely identifiable. Experiment design studies indicated that carefully designed experiments could significantly reduce the uncertainty in the parameter estimation problem.

Keywords: biopolymer, parameter estimation, identifiability, experiment design

1. Introduction

The use of systems engineering tools for bioprocess improvement requires a mathematical model that is representative of the biological system over a wide range of operating conditions. Advances in the field of biology are enabling the formulation of such models. A key step in model development is accurate

* Supported by Universities UK through the Overseas Research Students Awards Scheme.

[†] Corresponding author: Charles David Immanuel, e-mail: c.immanuel@imperial.ac.uk, Fax: +44 (0)20 7594 6606.

parameter estimation, which becomes difficult due to the relatively large number of parameters associated with such models as well as limitations in the number of measurable outputs. As a result, it is possible to obtain multiple (non-unique) feasible parameter sets when limited experimental data is used. In this paper, the problem of accurate parameter identification in a mathematical model of the production of the biopolymer, poly- β -hydroxybutyrate (PHB), is studied.

PHB belongs to the important class of biopolymers called polyhydroxyalkanoates (PHAs). PHAs have a wide range of applications and are the subject of much attention within the chemical engineering community. Bacteria synthesise PHAs as a carbon and energy reserve material when their growth is limited due to the unavailability of a nutrient such as nitrogen, sulphur or phosphorous [1]. The polymerisation of the soluble intermediates into insoluble molecules prevents the leakage of these valuable compounds out of the cells. Thus, PHAs are functionally similar to starch in plants and glycogen in animals.

PHB production is normally carried out by limiting cells on ammonia in the presence of excess glucose. Under these conditions, the cells tend to accumulate a large amount of the polymer. However, the bacterial production of PHB on a large scale has been limited, the main problem being the high production cost when glucose is used as the carbon source. Therefore, significant research has been undertaken to improve the productivity of the process. Besides the use of alternate carbon substrates [1], research has also been undertaken on genetically modifying organisms such as *Escherichia coli* to produce PHB, as such genetic modifications could be used to engineer faster growth or easier cell lysis [3]. Detailed systems engineering can be a substantial aid both in the identification of such genetic modifications as well as to optimise and optimally control the process.

In this paper, a mathematical model of PHB production is formulated. Parameter estimation for this model indicates the existence of multiple feasible parameter sets as observed before [2]. The objective of this paper is to utilise key tools from the wealth of systems engineering tools currently available, to study this problem of parameter multiplicity and to attempt to provide a solution.

2. Cybernetic modelling of PHB accumulation

The cybernetic modeling approach [4] was used in formulating the process model. In this approach, cells are construed to be optimal strategists that seek to maximise their growth given the existing environmental conditions. Two cybernetic models of PHB synthesis in microorganisms are available in the literature [5,2]. The first model [5] assumes that cells are composed of two components, namely residual biomass and PHB. Although this model was successful in predicting PHB production in the bacterium *Alcaligenes eutrophus*, it failed to take into consideration the underlying metabolic processes. This deficiency was addressed in the second model [2] which took

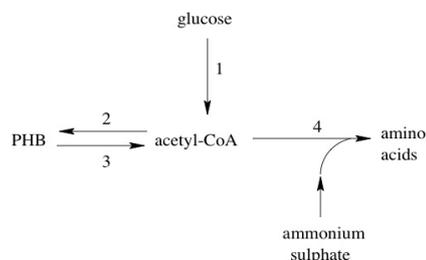


Figure 1. A model of PHB synthesis

into consideration the metabolic pathways leading to cell growth and PHB synthesis.

In this paper, a model of intermediate complexity is formulated accounting for the underlying metabolic processes. The model (see Figure 1) considers four reactions, each representing one of the pathways in the detailed model described previously [2]. The basic regulation is a preferential allotment of the glucose either to aid growth by forming amino acids or to

be stored as PHB for future utilisation. This regulation can be determined by the amount of glucose available and also the amount of ammonium sulphate available to metabolise with the glucose and produce amino acids. Two sets of cybernetic variables are employed in the model. The first set seeks to maximise the production of acetyl-CoA from reactions 1 and 3. The second set of cybernetic variables seeks to maximise the production of PHB and residual biomass from reactions 2 and 4. In defining the corresponding variables, the reaction rates corresponding to glucose and ammonium sulphate assimilation are employed. This strategy was used with the reasoning that, from a biological perspective, the choice of which reaction to maximise is dependent not on the rates of production of PHB and amino acids from acetyl-CoA but on the availability of glucose and ammonium sulphate.

Published experimental data [5] was used in finding the values of the 11 kinetic parameters. Three locally optimal solutions were obtained (data not shown), the first using gPROMS (Process Systems Enterprise) and the others

using the NAG Fortran Library routines (Numerical Algorithms Group). In all cases, sequential quadratic programming, a local optimisation technique, was used to minimise a least squares objective function. Figure 2 shows a comparison of the model predictions of the three parameter sets with the data used for parameter estimation. All three solutions are in excellent agreement with the data even though large differences were present

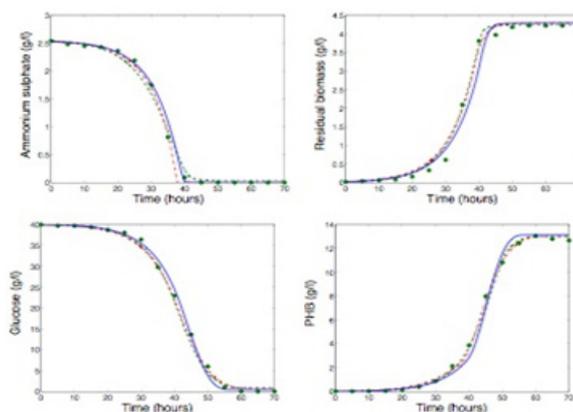


Figure 2. Comparison of the simulation results of the three parameter sets with the data used for parameter estimation [5]. — Solution 1, - - - Solution 2, - . - . - Solution 3.

among the sets in the values of some parameters.

As an attempt to evaluate the qualitative similarity in the model predictions with these three parameter sets, bifurcation analyses were performed for a continuous stirred tank reactor. The results were very different (figures not shown). Therefore, in order to improve the parameter estimates, sensitivity and estimability studies were undertaken as described in the next section.

3. Analysis of parametric uncertainty

Figure 3 shows simulation-based sensitivity analysis results when the parameters were perturbed by $\pm 50\%$ from their estimated values. Most

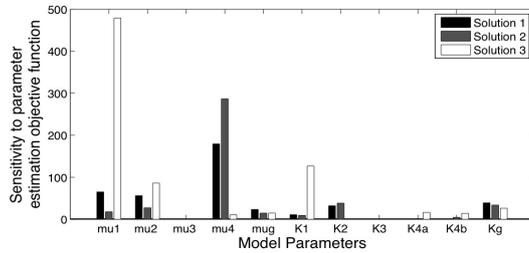


Figure 3. Simulation-based sensitivities of the three parameter sets to the parameter estimation objective.

parameters were found to be considerably sensitive, the exceptions being the Monod constants μ_3 and K_3 that correspond to PHB degradation (reaction 3). As the experimental data used corresponded to conditions of PHB accumulation (and minimal PHB degradation), this result was reasonable. However, as the remaining parameters appeared to show

some sensitivity, it was hypothesised that correlations among some of the parameters could be responsible for the uncertainty in the parameters.

In order to confirm this hypothesis, parameter estimability studies were undertaken. For a given set of parameters and experimental data, parameter estimability involves the determination of the subset of parameters that can be accurately and uniquely determined. The parameter estimability method used here [6] is based on an analysis of the sensitivity coefficients, which are the scaled first-order partial derivatives of the outputs with respect to the parameters, $\partial\eta/\partial\theta$ (Jacobian), at each sampling time. Table 1 lists the subset of

Table 1. Parameters estimable from experimental data

Rank	Solution 1	Solution 2	Solution 3
1	μ_1	μ_1	μ_1
2	μ_4	μ_4	μ_g
3	μ_2	K_g	K_g
4	K_1	μ_3	K_1

Further, although the estimable parameters from each set vary, the differences are not substantial.

In order to confirm these results, the following parameter identifiability problem [7] was solved:

estimable parameters obtained. A comparison of Table 1 with Figure 3 indicates that in each parameter set, the parameters that were estimable were not necessarily those that were most sensitive to the parameter estimation objective function.

$$\text{P1: } \max_{\theta, \theta^*} \Phi = (\theta - \theta^*)^T W_{\theta} (\theta - \theta^*) \quad (1)$$

$$\text{s.t. } \sum_{i=1}^{N_s} (y_i(\theta) - y_i(\theta^*))^T W_y (y_i(\theta) - y_i(\theta^*)) < \varepsilon_y \quad (2)$$

$$\dot{x} = f(x, u, \varphi) \quad y = g(x) \quad \varphi = \theta, \theta^* \quad (3)$$

$$\sum_{i=1}^{N_s} (y_i(\theta) - y_i^{exp})^T W_y (y_i(\theta) - y_i^{exp}) < \varepsilon_{exp} \quad (4)$$

Here y is the vector of outputs, N_s is the number of sampling points and W_{θ} and W_y are weighting matrices. The solution to the parameter identifiability problem gives the largest distance between two parameter sets, θ and θ^* , that give

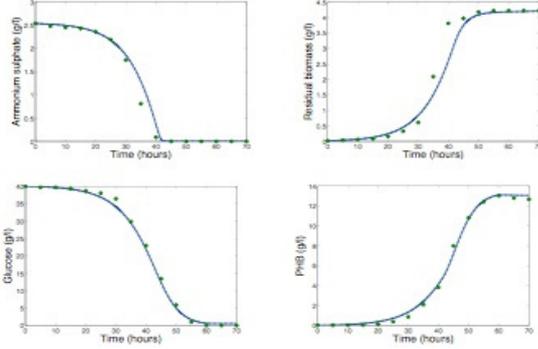


Figure 4. Solution to the parameter identifiability problem P1. — θ , - - - θ^* .

similar predictions within a set tolerance limit, ε_y . In order to consider the data used for parameter estimation, Equation (4) was added as an additional constraint. Figure 4 depicts the solution obtained by solving problem P1 with $\varepsilon_y=10^{-3}$ and $\varepsilon_{exp}=1.0$. It can be seen that the solutions give almost identical predictions even though large differences were present in about six of the model parameters. The problem P1 was then solved

with only the estimable parameters as decision variables. It was found that when the inestimable parameters were neglected, the remaining parameters could be estimated with very little uncertainty (data not shown), thus corroborating the findings of the parameter estimability studies.

Finally, the following parameter distinguishability problem P2 [7] was solved to design additional optimal experiments to minimise the parametric uncertainty. The solution to this problem is an experiment that maximises the difference between the outputs of two models (parameter sets) thereby providing a mechanism to discriminate between the parameter sets. This problem can be expressed mathematically as follows:

$$\text{P2: } \max_u \Phi = \int_0^{t_f} \left[\frac{y(\theta) - y(\theta^*)}{\max(y(\theta), y(\theta^*))} \right]^T \left[\frac{y(\theta) - y(\theta^*)}{\max(y(\theta), y(\theta^*))} \right] dt \quad (5)$$

$$\text{s.t. } \dot{x} = f(x, u, \varphi) \quad y = g(x) \quad \varphi = \theta, \theta^* \quad u_L \leq u(t) \leq u_U \quad (6)$$

Here $u(t)$ is the vector of experiment design variables including the initial conditions, and substrate feed rates; u_L and u_U are lower and upper bounds respectively on these quantities.

The parameter distinguishability problem was solved with all combinations of the three solutions previously obtained. Figure 5 shows the solution obtained

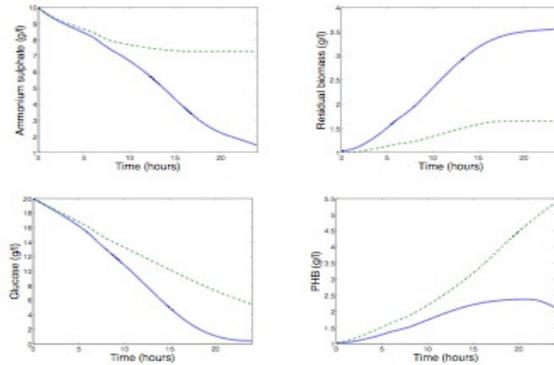


Figure 5. Solution to the parameter distinguishability problem P2. — Solution 1, - - - Solution 3.

when parameter sets 1 and 3 were employed. As can be seen, the fed-batch experiment, obtained as optimal from problem P2, predicts very different concentration profiles when the two parameter sets are employed. Thus, it can be concluded that the use of carefully designed experiments to help distinguish between these two parameter sets could minimise the uncertainty in the parameter estimation problem.

4. Conclusions

In this paper, the problem of accurate parameter identification in a model of poly- β -hydroxybutyrate production was discussed. It was found that the experimental data used for parameter estimation was not sufficient both in quantity and excitation to enable an accurate estimation of all model parameters, resulting in significant parametric uncertainty. Parameter estimability and identifiability studies indicated that with the experimental data used, only a subset of the model parameters could be accurately estimated. Finally, simple experiment design studies were undertaken which indicated to carefully designed experiments that could potentially provide more accurate parameter estimates.

References

1. AJ Anderson, EA Dawes. *Microbiol. Rev.* 54 (1990) 450
2. KG Gadkar, FJ Doyle III, TJ Crowley, JD Varner. *Biotech. Prog.* 19 (2003) 1487
3. LL Madison, GW Huisman. *Microbiol. Mol. Biol. Rev.* 63 (1999) 21
4. DS Kompala, D Ramkrishna, NB Jansen, GT Tsao. *Biotech. Bioeng.* 28 (1986) 1044
5. S Yoo, W-S Kim. *Biotech. Bioeng.* 43 (1994) 1043
6. B Kou, KB McAuley, CC Hsu, DW Bacon, KZ Yao. *Ind. Eng. Chem. Res.* 44 (2005) 2428
7. SP Asprey, S Macchietto. *Comp. Chem. Eng.* 24 (2000) 1261