OPTIMIZATION METHODS FOR INDIVIDUAL-BASED MODEL PARAMETER ESTIMATION IN PREDICTIVE MICROBIOLOGY

C. Prats[1], A. Standaert[2], K. Bernaerts[3], J. Ferrer[1], D. López[1], J. Van Impe[3]
[1]Universitat Politècnica de Catalunya, Barcelona, Spain; [2]VITO-Vlaamse Instelling voor Technologisch Onderzoek, Mol, Belgium; [3]BioTeC - Chemical and Biochemical Process Technology and Control Section, Katholieke Universiteit Leuven, Heverlee, Belgium

Corresponding Author: C. Prats, Dept. Física i Enginyeria Nuclear, Escola Superior d'Agricultura de Barcelona, Universitat Politècnica de Catalunya
Avda. Canal Olímpic 15, 08860 Castelldefels (Barcelona), Spain; clara.prats@upc.edu

Abstract. In the framework of microbiology, Individual-based Models are discrete models in which the main entities are microbes. Their use in simulations as 'virtual experiments' to predict the evolution of populations under specific conditions requires accurate setting of the parameters involved. We adapted and tested two optimization methods for Individual-based Model parameter estimation: the Nelder-Mead Threshold Accepting (NMTA) and the NEWUOA. These methods presented no convergence problems, and the best results in terms of time expenditure were derived with the latter.

1 Introduction

The aim of predictive microbiology is to predict the evolution of a microbial population under specific conditions once the initial state has been determined. In order to build proper mathematical models that describe the systems under study, it is essential to accurately define their variables and precisely delimit the values of their parameters.

In the framework of microbiology, Individual-based Models (IbMs) are discrete models in which the main entities are microbes. These models define rules for the individual cells' behaviour (e.g., uptake, metabolism and cell cycle, among others) and for the processes that take place in the environment (e.g., nutrient diffusion and medium heating, among others). Once they are implemented in a computer, simulations show the evolution of an initial population following individual and local rules. Two examples of IbM simulations that have provided various and interesting results in microbiology are BacSim [4] and INDISIM [3].

The use of IbM simulations as 'virtual experiments' to predict the evolution of a population under specific conditions requires an accurate setting of the parameters involved. Most of them are related to individual characteristics of the microbes (e.g., individual mean growth rate, individual mean uptake rate and individual mean mass to initiate the reproduction cycle, among others). Thus, they are difficult (if not impossible) to assess experimentally. They must be indirectly estimated by fitting the output simulations with the related experimental measurements at the population scale, by means of an objective function that assesses this likelihood. Several methods for parameter estimation have been developed within the framework of continuous modelling. However, they are not usable with IbM because they are generally based on gradient methods, which are not applicable in IbM because of their discrete nature.

The classic method for estimating the parameters of an Individual-based Model is the grid search, which involves great time expenditure. In this study we adapted and tested two optimization methods for an IbM parameter estimation: the Nelder-Mead Threshold Accepting (NMTA) [2,5] and the NEWUOA [6]. Both of them were implemented in Matlab and used to estimate one, two and three parameters of BacSim and INDISIM, providing similar results. We present here the results of INDISIM parameter estimation: (i) the one-parameter estimation (mean individual uptake rate, u_{max}) was performed by means of a grid search, NMTA method and NEWUOA; (ii) the two-parameter estimation (u_{max} and mean individual biomass to initiate the reproduction cycle) was performed by means of a grid search and the NEWUOA method; and (iii) the three-parameter estimation (u_{max} and two parameters of the initial biomass distribution) was tested with NEWUOA.

2 Estimating INDISIM parameters

To estimate the parameters of an IbM, some steps must be taken [7]. First of all, it is necessary to set the values that can be deduced from the literature or induced from experimental data, which are the most arguable sources. The number of parameters to be estimated should be as low as possible to reduce the parameterization time. So this step concludes with the identification of the parameters to be estimated. It is also useful to know some thresholds for these values or at least their order of magnitude. Since we are talking about IbM models of microbiological systems, the parameters to be estimated must be biological or physical values with their units. Therefore, the literature should allow a delimitation of their possible values. In this study we performed three

different trials estimating: (i) one parameter (mean individual glucose uptake rate, u_{max}); (ii) two parameters (u_{max} and mean individual biomass to initiate the reproduction cycle); and (iii) three parameters (u_{max} and two parameters of the initial biomass distribution).

To carry out the estimation of the chosen parameters, experimental information about the system is necessary. That is, the logical estimation process is based on the fitting of the unknown parameters so that simulation results are as similar as possible to an experimental dataset. So, the second step is the choice of the experimental dataset (or datasets) to be used. The IbM simulations must be adapted to generate results that are comparable to these data. In this study, the experimental dataset was taken from Bernaerts [1]. It corresponds to *Escherichia coli* K12 MG1655 growth in Brain Heart Infusion broth at 27.5ºC.

In a third step, a numerical method to evaluate the soundness of the simulations regarding the experimental data is needed. This numerical evaluation is usually an *objective function* that decreases with the soundness of the fitting. We used the mean square error (MSE) as objective function [8].

Finally, a method to perform the search for the lowest value of the objective function in a systematic manner is needed. We used three different methods: the classic grid search, the Nelder-Mead Threshold Accepting and the NEWUOA.

2.1 Grid search

In a grid search, each parameter is discretized in a certain interval where its best value is assumed to lie, and the resulting grid (different combinations of parameters to be evaluated) is explored to find the best point. This method involves great time expenditure, since it requires an enormous number of simulations.

2.2 Nelder-Mead Threshold Accepting (NMTA)

The Nelder-Mead method [5] is one of the most used algorithms for nonlinear unconstrained optimization. It is a direct search method, since it does not use derivatives. The basic unit in this method is the simplex: a geometric figure in an n-dimensional space that is a convex hull of *n+1* vertexes, with each vertex representing a certain combination of the *n* parameters to be estimated. The value of the objective function is assessed in each simplex vertex, and a new simplex that is closer to the objective function is constructed at each step. The rules for constructing each new simplex are based on a set of geometric operations. The value of the objective function determines whether new vertexes are accepted or rejected. If a new vertex is accepted, a new simplex is constructed by rejecting the worst existing vertex, always according to the objective function values.

The Threshold Accepting algorithm [2] performs a local search that escapes local minima by means of accepting solutions that are not worse than the current one by more than a given threshold, τ. It was incorporated into the Nelder-Mead method in order to escape the possible local minima of the objective function.

2.3 NEWUOA

The NEWUOA algorithm [6] seeks the minimum of an objective function, $F(\vec{x})$, where $\vec{x} \in \Re^n$. The model proposes the use of the objective function values for building a quadratic model, $Q(\vec{x})$, which is built in order that $Q(\vec{x}) \approx F(\vec{x})$ within a trust region. The quadratic model is minimized within the trust region, hopefully yielding a point with a low function value. The radius of the trust region, ρ, is iteratively adjusted and a new quadratic model is built and minimized. In this study, the objective function is $F(\vec{x}) = MSE(\{p_i\})$. That is, the vector $\vec{x}$ contains the different combinations of the *N* parameters to be estimated, $\{p_i\}$.

3 Results

3.1 One-parameter estimation

The one-parameter estimation (u_{max}) was performed with the three methods, providing similar results. The best estimates were $u_{max} = 0.0126$ mol g^{-1} h^{-1} with the grid search (intervals' width of 0.00015 mol·g^{-1}·h^{-1}), $u_{max} = 0.0124$ mol g^{-1} h^{-1} with the NMTA and $u_{max} = 0.0127$ mol g^{-1} h^{-1} with the NEWUOA. Thus, they were similar using any of the presented methods, and the quality of the results was comparable. The classical grid search yielded a lot of information about the *MSE* behaviour in the studied interval, but was slow and tedious. NMTA and NEWUOA showed good convergence for the studied case, although it was necessary to accurately delimit the interval. The NMTA optimization processes lasted 15 times longer than NEWUOA runs. Therefore, the fastest method by far was NEWUOA. Figure 1 shows the grid search and several runs of the NEWUOA estimation.

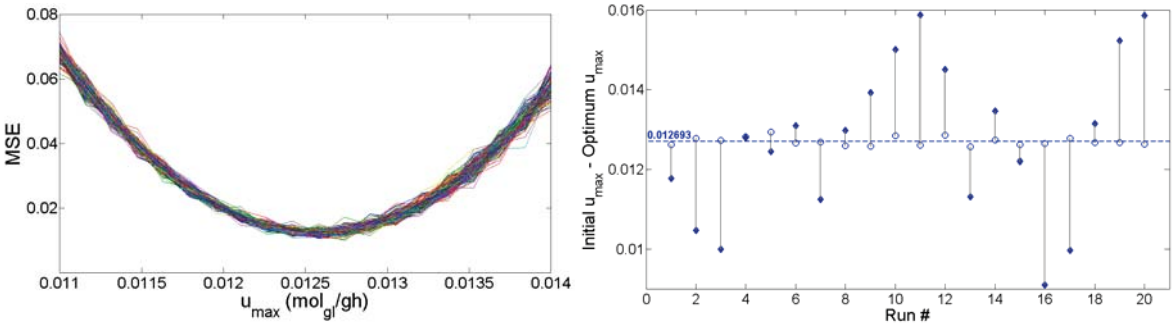

Figure 1. The MSE for a grid search of 20 u_{max} values, for 149 runs executed with non-overlapping random number sequences (left). 20 independent rounds of NEWUOA optimization method for estimating the u_{max} (right): blue full rhombuses are the randomly chosen initial points, and red open circles are the best estimate for each run; the mean of the 20 best estimates is shown with a dashed line.

3.2 Two-parameter estimation

The parameters chosen to be estimated were the mass at division, m_d, and the mean maximum uptake rate, u_{max}. According to the above-mentioned results regarding the convergence and time expenditure of the optimization methods, we only performed a grid search and several NEWUOA optimization runs. The obtained best estimates for these parameters were $u_{max} = 0.0124$ mol g^{-1} h^{-1} and $m_d = 0.40\ pg$ with the grid search (Figure 2), and $u_{max} = 0.0125$ mol g^{-1} h^{-1} and $m_d = 0.47\ pg$ with the NEWUOA (Figure 3). These best m_d estimates were similar to the value used in Section 3.1 (0.43 pg). The two-parameter NEWUOA optimization processes lasted just 1.2 times longer than the one-parameter NEWUOA runs, and no convergence problems were detected.

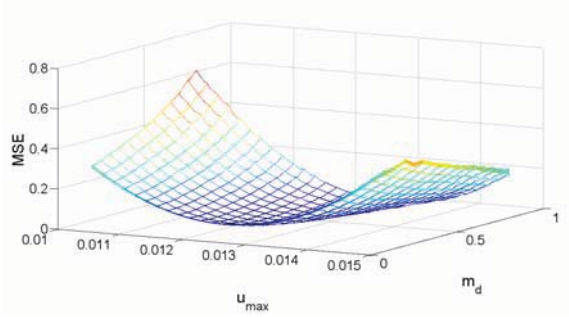

Figure 2. Grid search for mass at division (m_d) and maximum uptake rate (u_{max}).

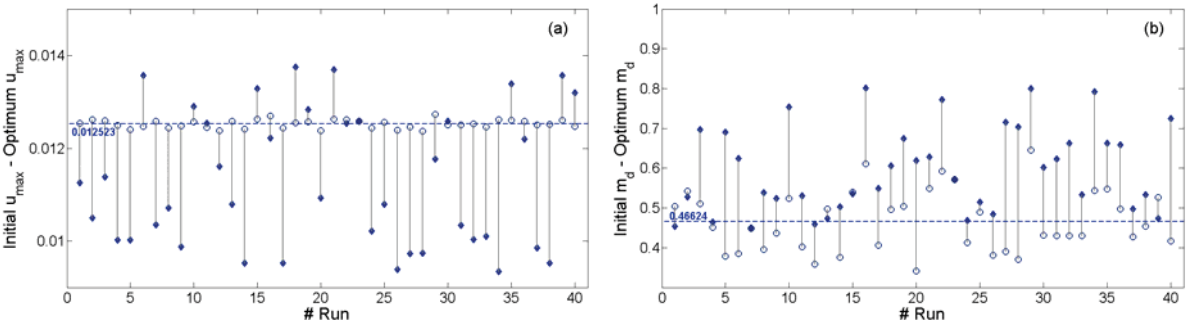

Figure 3. NEWUOA results for optimizing (a) maximum uptake rate (u_{max}) and (b) mass at division (m_d). Full circles are the initial points, and open circles are the best estimates for each run. The dashed line indicates the mean value of the best estimates.

3.3 Three-parameter estimation

A third test was carried out to evaluate the usefulness of the NEWUOA method for the estimation of more than one parameter. A three-parameter evaluation was performed, and chosen parameters were the u_{max} again, and the two Weibull parameters for setting the inoculum biomass distribution (A and B). The m_d was taken from the bibliography. The mean of the 40 independent NEWUOA runs resulted in the best estimates $u_{max} = 0.0124$ mol g^{-1} h^{-1}, $A = 0.10\ pg$ and $B = 2.99$. The convergence was acceptable for u_{max} and A, but it was not as good for B due to a lack of information in the experimental data. The three-parameter optimization processes lasted between 1.2 and 2.4 times longer than the one-parameter NEWUOA runs.

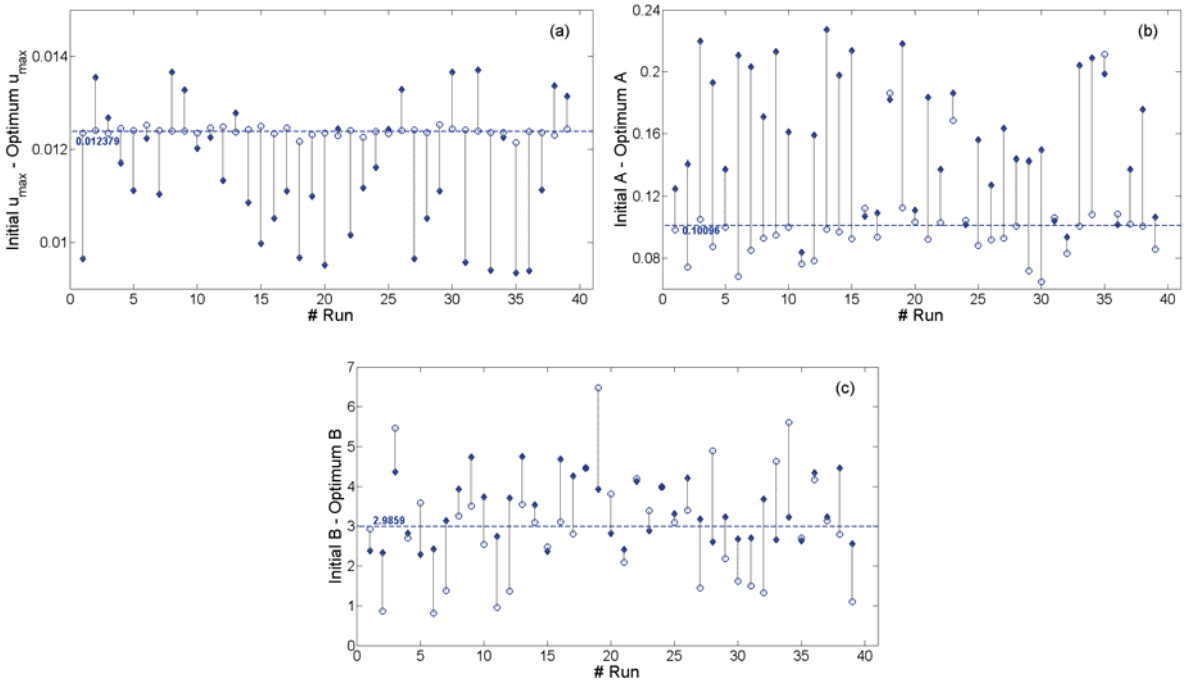

Figure 4. NEWUOA results for optimizing (a) the maximum uptake rate (u_{max}), (b) the Weibull distribution constant for the inoculum A and (c) the Weibull distribution constant for the inoculum B. Full rhombuses are the initial points, and open circles are the best estimates for each run. The dashed line indicates the mean value of the best estimates.

4 Conclusions

Two methods were adapted and tested for IbM parameter estimation. The NMTA method, which was considerably faster than the classic grid search, showed a good convergence for one parameter but some difficulties for the programming. NEWUOA was proved to be a useful tool for IbM parameterization, although it required a minimum knowledge of the parameters to be estimated in order to set the trust region radius. It was the fastest method, and the results had sufficient precision. For further and better applications of the NEWUOA method to IbM parameter estimation, the objective function should be slightly modified to incorporate a penalty function that avoids the algorithm of leaving the real trust region.

5 Acknowledgements

We gratefully acknowledge the financial support of the Plan Nacional I+D+i of the Ministerio de Educación y Ciencia CGL2007-65142/BOS. K.Bernaerts is a Postdoctoral Fellow with the Fund for Scientific Research – Flanders (F.W.O-Vlaanderen). The authors also thank Matthieu Guilbert for kindly providing the NEWUOA program and his assistance in its integration with the IbM models.

6 References

[1] Bernaerts, K.: *Dynamic experiment design and validation in predictive microbiology: optimal estimation of two-parameter microbial growth kinetics as function of temperature.* PhD thesis (2002). Katholieke Universiteit Leuven.

[2] Dueck, G. and Scheuer, T.: *Threshold accepting: A general purpose optimization algorithm appearing superior to simulated annealing.* Journal of Computational Physics, 90 (1990), 161 - 175.

[3] Ginovart, M., López, D. and Valls, J.: *INDISIM, and individual based discrete simulation model to study bacterial cultures*. Journal of Theoretical Biology, 214 (2002), 305 - 319.

[4] Kreft, J.U., Booth, G. and Wimpenny, J.W.T.: *BacSim, a simulator for individual-based modelling of bacterial colony growth.* Microbiology, 144 (1998), 3275 - 3287.

[5] Nelder, J.A and Mead, R.: *A simplex method for function minimization.* The Computer Journal, 7 (1965), 308 - 313.

[6] Powell M.: *The NEWUOA software for unconstrained optimization without derivatives.* In: Nonconvex Optimization and Its Applications, vol. 83, 2006, Springer US, 255-297.

[7] Prats, C.: Individual-based Modelling of bacterial cultures in the study of the lag phase. PhD Thesis (2008). Universitat Politècnica de Catalunya, Spain (http://www.tdx.cat/TDX-0701108-105132).

[8] Standaert, A.R.: From macroscopic to microscopic modelling methodologies in predictive microbiology. PhD thesis (2007). Katholieke Universiteit Leuven, Belgium.