Mariano Nicolas Cruz Bournazou, Tilman Barz, D. B. Nickel, Diana Carolina Lopez Cárdenas, Florian Glauche, Andreas Knepper, Peter Neubauer

Online optimal experimental re-design in robotic parallel fed-batch cultivation facilities

Online Optimal Experimental Re-Design in Robotic Parallel Fed-Batch Cultivation Facilities for Validation of Macro-Kinetic Growth Models using *E. coli* as an Example

M.N. Cruz Bournazou, T.Barz, D. Nickel, D. Lopez Cárdenas, F. Glauche, A. Knepper, P. Neubauer

corresponding author:

Dr. M. Nicolas Cruz Bournazou
Laboratory of Bioprocess Engineering
Department of Biotechnology
Technische Universität Berlin (TU Berlin)
Ackerstr. 76, ACK24
D-13355 Berlin, Germany
Tel. direct: +49 30 314 72527
Secretary: +49 30 314 72573
Fax: +49 30 314 27577
www.bioprocess.tu-berlin.de

Abstract

We present an integrated framework for the online optimal experimental re-design of parallel nonlinear dynamic processes that aims to precisely estimate the parameter set of macro kinetic growth models with minimal experimental effort. This provides a systematic solution for rapid validation of a specific model to new strains, mutants, or products. In biosciences, this is especially important as model identification is a long and laborious process which is continuing to limit the use of mathematical modeling in this field.

The strength of this approach is demonstrated by fitting a macro-kinetic differential equation model for *Escherichia coli* fed-batch processes after six hours of cultivation. The system includes two fully-automated liquid handling robots; one containing eight mini-bioreactors and another used for automated at-line analyses, which allows for the immediate use of the available data in the modeling environment. As a result, the experiment can be continually re-designed while the cultivations are running using the information generated by periodical parameter estimations.

The advantages of an online re-computation of the optimal experiment are proven by a fifty-fold lower average variation coefficient on the parameter estimates compared to the sequential method (4.83% instead of 235.86%). The success obtained in such a complex system is a further step towards a more efficient computer aided bioprocess development.

Introduction

Today, mathematical modeling in biotechnology is not hampered by computer capacity or by insufficient understanding of the microbial systems, but rather by the lack of fast, cheap, and informative experiments – especially in the case of dynamic processes such as cultivations describing industrial conditions. When properly validated, simplified mathematical models are capable of describing the dynamics of complex systems (Cruz Bournazou et al. 2012; Kokkalis et al. 2014; Legmann et al. 2009; Rosen et al. 2006; Zavrel et al. 2008). In literature, we can find models that successfully describe biological systems at all scales (Brunk et al. 2012; Neubauer and Junne 2010), including models for optimization of dynamic processes (Cruz Bournazou et al. 2013; Delgado San Martín et al. 2014; Hidalgo-Bastida et al. 2012; Lu et al. 2013; Oddone et al. 2007; Venkata Mohan et al. 2005) and processes with mixtures of microbial populations with low quality data (Junker and Wang 2006; Su et al. 2005). However, contrary to other industries where processes are typically predicted using existing equations and literature values (thermodynamics, chemical reactions, combustion, etc.), model building in biotechnology is inevitably coupled with iterative and laborious experimental validation for each process-product pair (Ataman and Hatzimanikatis 2015; Medema et al. 2012). Regardless of its level of complexity, a model needs to be constantly re-fitted against observations in order to adapt its outputs to variations in the environment or in the microorganism itself (e.g mutations). Therefore, creating an experimental facility that can generate the data required to fit a specific model with high cost-time efficiency is certainly relevant for research and industry. This implies that the experimental set-ups should: i) run fast and cheap, ii) emulate process relevant conditions, and iii) consider the evolution of the system over time.

These requirements have pushed the development in automation, miniaturization, and parallelization of experimental facilities developing automatic Liquid Handling Stations (LHS) for High Throughput Screening (HTS) and High Throughput Bioprocess Design (HTBD), (Long et al. 2014; Puskeiler et al. 2005; Schäpper et al. 2009) as well as sensor and Process Analytical Technologies (PAT) for a better and less invasive insight into biological systems (Neubauer et al. 2013). The rapid growth of these technologies brings additional challenges, such as the correct handling and analysis of very large data sets (Kozak 2014; Shockley 2015) and the design of extremely complex experiments. The design of multiple parallel experiments can be increased by using methods for design of experiments (Glauche et al. 2016; Lutz et al. 1996) and neural networks (Glassey et al. 1994) among others. Nevertheless, we need to go beyond "endpoint" or "steady state" experiments to uncover the dynamics of the process and predict its evolution over time. This means that the design of the dynamic experiment cannot be done exclusively with statistical tools. This time dependent behavior is better described by nonlinear differential equations. Methods for Optimal Experimental Design (OED), also called Model Based Design of Experiments (MBDoE), which design the experiments based on nonlinear differential equation systems, have been developed (Franceschini and Macchietto 2007a; Körkel et al. 2004) and specific applications in biotechnology exist (Baltes et al. 1994; Takors et al. 1997). In order to maximize the precision of the parameter estimates, the experimenter needs to find the optimal combination of input variables (e.g. feeding strategy) and sampling setup (type and time) within the feasible region. In addition, each sampling consists of a number of "actions" (pipetting, mixing, incubating, measuring, etc.) and requires different "resources" (1-, 8-, or 96-chanel pipette, shaker, photometer, flow cytometer, reaction

vessels, plates, etc.) which have to be optimally coordinated in an efficient schedule. Unfortunately, tools to plan and perform dynamic experiments exploiting full LHS capacity are not available. Despite various publications using OED in biotechnology (Franceschini and Macchietto 2007b; Galvanin et al. 2011; Gernaey et al. 2002; Kreutz and Timmer 2009; López Cárdenas et al. 2013; Skanda and Lebiedz 2010) and design of experiments applied in HTS, to our knowledge, there are no publications dealing with the online design of nonlinear dynamic parallel experiments as they are needed for HTBD.

Furthermore, in order to minimize the experimental effort and increase the robustness of the experiment, the model should be re-fitted after each sample and the experiment re-designed as data is being generated. This requires an efficient solution of a fairly large nonlinear and possibly ill-conditioned optimization program.

Works dealing with online experimental re-design (Galvanin et al. 2012; Stigter et al. 2006; Zhu and Huang 2011) show that the information generated in each experiment is significantly higher compared to experiments that are planned in sequence (offline). It may appear that the extension to parallel experiments is fairly straight forward from the theoretical point of view (Cruz Bournazou et al. 2014a); nevertheless, former studies are applied on simple systems and its application to a real set of parallel cultivations presents many challenges. This includes the complexity of the biological system, the operation of the experimental facilities, the typically long delays and the limited information content of the analyses, and the scheduling of all actions considering resource availability (Mayer and Raisch 2004). Finally, a robust and cheap computation of both optimization programs, namely the Parameter Estimation (PE) and the Experimental Design (ED) problem, is

critical in order to handle data and compute the optimal strategy at the speed dictated by the running experiment.

Problem statement

The goal of this work is to achieve an online computation of optimal parallel dynamic experiments together with its automatic implementation in a LHS for fed-batch cultivations of the bacterium E. *coli* (Figure 1) using a Sliding Window Optimal experimental Re-Design (SWORD). We regularize the singularity problem by an appropriate selection of the parameter subset to assure a well-conditioned parameter estimation even with reduced experimental information, and use a moving horizon approach to reduce the computational burden of the optimization. Ultimately, it is possible to re-design the running experiment after each measurement and take full advantage of the present state of information to plan the following steps.

Materials and Methods

The most relevant aspects of the experiment are presented here, for a more detailed description of the experimental setting, the reader is referred to the Appendix in the supplementary material and to (Nickel et al. submitted). The experiments were carried out using two different LHSs. The mini-bioreactor system bioREACTOR 48 (2mag AG, Munich, Germany) was integrated in a TECAN Freedom Evo LHS (Tecan, Crailsheim, Germany) (Figure 2) to automate all liquid handling steps. General process data (temperature, stirrer speed, pH, dissolved oxygen) was stored in the iLab-Bio database (infoteam, Bubenreuth, Germany). For pH control of the vessels, a LabVIEW (National Instruments, Munich, Germany) based pH controller was used. Worklists were generated of titrant volumes for the Tecan Gemini software. Samples were analyzed on a Hamilton

Microlab Star platform (Hamilton Robotics, Bonaduz, Switzerland), which is equipped with a Biotek Synergy MX microplate reader (BioTek Instruments GmbH, Bad Friedrichshall. Germany). The data was stored in the iLab-Bio database for further processing.

Bacterial strain and cultivation medium

The experiments were performed using *E. coli* W3110 (DSM No. 5911) stored at – 80 °C in LB medium containing 20% glycerol. Cultivations were carried out in EnPresso media (BioSilta Ltd., Cambridge, UK): the seed cultures were grown in EnPresso B medium and the main culture was performed in EnPresso B defined medium. Unless otherwise stated, media were prepared according to the manufacturer's instructions.

Seed and main cultures

The seed culture was performed in a 125 mL UltraYield Flask (Thompson Instrument Company, Oceanside, USA) containing 25 mL Enpresso B, which was inoculated with 24 µL of cryo culture. For glucose release, 1.5 U L^{-1} Reagent A (BioSilta) were added. The shake flask was covered with AirOtop enhanced shake flask seals (Thompson Instrument Company) and incubated at 30°C and 220 rpm in a shaking incubator with 50 mm offset until the mid-log phase was reached (4 – 6 h). Then the main culture was prepared in 100 mL of EnPresso B defined medium containing 10 g L^{-1} glucose with an initial OD_{620} of 0.01. For the initial batch phase, pre-calibrated baffled 500 mL sensor flasks for online pH and DO measurements were used (PreSens GmbH, Regensburg, Germany). The flasks were incubated overnight at 30°C and 200 rpm with 25 mm offset. At 20 h, the 11 mL of culture were transferred into the 8 mini-bioreactors.

The cultures were incubated at 30°C with an aeration rate of 0.1 L min^{-1} and a stirrer speed of 2600 rpm until they reach glucose limitation. This was indicated by a sudden increase of

DO. Then 6 UL^{-1} of Reagent A were added to initiate the glucose limited fed-batch phase. The pH was adjusted to 7.0 before starting the SWORD experiment by the addition of 6.25 % (v/v) ammonia solution via the Freedom Evo LHS.

SWORD experiment

Deck layout of Freedom Evo LHS

For addition of Reagent A, a 3000 U L^{-1} Reagent A solution was placed on a cooled plate carrier at 4 °C. The EnPresso B defined medium, the 40 gL^{-1} acetate additive solution and the glucose additive solution were pipetted individually into 24 deep well plates (Ritter, Schwabmünchen, Germany). Samples for analysis were collected into a V-bottom plate. The pH was adjusted by adding 6.25 % (v/v) ammonia solution prepared in a sterile container. Sterilization of the metallic pipetting tips was done in a washing station using 70 % (v/v) ethanol (EtOH).

Sequential arrangement of additions during SWORD

The SWORD experiment was divided into cycles of 20 minutes with a total experimental time of 6 h. One cycle comprised 5 subroutines each of which lasted 4 minutes (Figure 3). During each 4-minute subroutine, all steel needles were firstly sterilized. Then, either a MATLAB work list for the addition of a specific pulse was loaded into the Gemini environment and executed or the LHS took 300 µL samples from each mini reactor into the V-bottomed plate.

Analytics

Sampling was done every 20 minutes. The samples were pipetted into plates containing anhydrous NaOH and mixed thoroughly to a final NaOH concentration of 0.1 M.

Afterwards, 20 µL of sample were used for optical density measurements. The remaining sample volumes were centrifuged at 3000 rpm and 4 °C for 10 min to separate the cells. Duplicates of 75 µL supernatant were pipetted for glucose and acetate analyses. Samples from four successive time points were cooled at 4 °C and analyzed together.

The OD_{620} samples were diluted 15 fold in flat bottom plates and were measured manually with a PHOmo microplate reader (Anthos, Krefeld, Germany). Samples were shaken inside the microplate reader before the measurement. The OD_{620} values were corrected against the blank, correlated to OD_{600} in a 1 cm cuvette by multiplication with the factor 2.62. The cell dry weight (CDW) was approximated by dividing the OD values by 3.3 as obtained from dry weight calibrations. OD-measurements were carried out every 20 min.

The extracellular glucose concentration was measured with the enzymatic Glucose Hexokinase FS kit (DiaSys Diagnostic Systems, Holzheim, Germany) on the Hamilton LHS as described previously (Knepper et al. 2014). The extracellular acetic acid concentration was measured with the enzymatic Enzytec fluid™ kit (R-Biopharm, Darmstadt, Germany). Both enzymatic assays were performed on the Microlab Star LHS.

Software

All numeric computations were performed using MATLAB Release 2013b. Model and parameter sensitivity equations were integrated using CVODES solver from SundialsTB Toolbox (Hindmarsh et al. 2005).

The computed parameter sensitivities, normalized with respect to initial parameter guesses and measurements, were used to accurately calculate the gradients of the PE problem and the Fisher Information Matrix (FIM). The PE problems were solved by single shooting and using TOMLABs clsSolve solver with the Wang, Li, Qi Structured MBFGS method and

user defined gradients (Holmström et al. 2003). The experimental design (ED) problems were solved using the TOMLAB implementation of SNOPT (Gill et al. 2002). The objective was computed from the FIM. The gradients were computed with finite differences with four threads in parallel.

E. coli kinetic model

The macro-kinetic model used in this work describes E. *coli* substrate limited growth and a Matlab® version is available online (see Appendix). The validity of the model is constrained to a narrow operability region, namely i) glucose concentrations lower than 0.2 g/l, ii) at least 20% of dissolved oxygen tension, and iii) overflow phases must be shorter than 60 seconds. In addition to substrate feeding, the model considers glucose release by the EnBase system (Panula-Perälä et al. 2008). The model consists of 6 state variables, namely biomass concentration (X) in [g/l], substrate concentration (S) in [g/l], dissolved oxygen tension (DOT) in [% of saturation], acetic acid (A) in [g/l], liquid volume (V) in [l], and glucose release enzyme concentration (E) in [Units/l]. These are represented by an Ordinary Differential Equation (ODE) system as given in the appendix in equations A1 to A6. The structure of the model is based on the work of (Lin et al. 2001). The model was modified to add recent discoveries in the acetate production mechanism (Valgepea et al. 2010). As a result, acetate production and consumption are always active and at equilibrium with a net acetate production of 0.

SWORD

We present a brief introduction to the basics of OED. The reader is referred to (Barz et al. 2012; Cruz Bournazou et al. 2014b; Franceschini and Macchietto 2007a) for further details.

Optimal experimental design

Let us consider a model consisting of a nonlinear implicit differential equation system that describes the dynamic behavior of a process (e.g. an *E. coli* fed-batch cultivation). This model will contain a set of unknown parameters that can be varied to fit the outputs of the model against observations of the real process. OED aims at finding experimental setups such that the statistical uncertainty of estimates of the unknown model parameters is minimized (Korkel 2002). Generally, the main factors that affect the identifiability of a model (i.e. the size of its confidence intervals) are: i) the structure of the model, ii) the quality of the measurements (type, frequency, accuracy), and iii) the design of the experiment used to generate the data (inputs, conditions, measurmenents, etc.). In other words, the computed optimal conditions aim to maximize the information content of the measurements so that the parameters are determined most accurately.

Roughly speaking, the inputs should excite the system so as to let the experimenter observe its dynamics by sampling in the most sensitive points of the experiment. To find this optimal setup, mixed second order sensitivities of the least square functional with regard to parameters and states have to be computed.

Mathematical formulation

A simplified description of the problem is presented, the reader is referred to Appendix for further details. We consider a twice continuously differentiable nonlinear Ordinary Differential Equation (ODE) system:

$$\dot{x}(t) = f\big(x(t), \theta, u(t)\big);\ y(t) = h\big(x(t)\big);\ x(t_0) = x_0 \qquad 1$$

where $t \in [t_o, t_{end}] \subseteq \mathbb{R}$ is the time, $x(t) \in \mathbb{R}^{n_x}$ are dependent state variables, $u(t) \in \mathbb{R}^{n_u}$ are the time-varying input or experimental design variables and $\theta \in \mathbb{R}^{n_p}$ the unknown

parameter vector and initial conditions are given by x_0. The vector $y(t) \in \mathbb{R}^{n_y}$ contains the predictions of the observed responses. In Eq. (1) we consider a set of $\mathcal{R} \coloneqq \{r_1, \cdots, r_{n_r}\}$ parallel experiments, with n_r being the number of reactors. The robot simultaneously feeds all mini-reactors at discrete feeding time instances $\mathcal{T}^f \coloneqq \{t_1, \cdots, t_{n_f}\}$, with n_f being the total number of injections into one reactor. Each reactor has its corresponding input signals or experimental design variables $u_r(t) \in \mathbb{R}^{n_u}$, with $t \in \mathcal{T}^f$ and $r \in \mathcal{R}$. For all reactors, the experimental design variables are collected in the vector

$$U \coloneqq \left[u_{r=1}(t_{f=1})^T, \cdots, u_{r=1}(t_{n_f})^T\right]^T, \ldots, \left[u_{n_r}(t_{f=1})^T, \cdots, u_{n_r}(t_{n_f})^T\right]^T \in \mathbb{R}^{n_u \cdot n_f \cdot n_r}$$

with the number of possible individual species to be injected n_u, the total number of injections n_f and the number of reactors n_r. The robot simultaneously takes samples from all mini-reactors at discrete measurement time instances $\mathcal{T}^s \coloneqq \{t_1, \cdots, t_{n_s}\}$, with n_s being the number of samples taken by the robot from one reactor. The measurements are defined by

$$Y^m \coloneqq \left[y^m_{r=1}(t_{s=1})^T, \cdots, y^m_{r=1}(t_{n_s})^T\right]^T, \ldots, \left[y^m_{n_r}(t_{s=1})^T, \cdots, y^m_{n_r}(t_{n_s})^T\right]^T \in \mathbb{R}^{n_y \cdot n_s \cdot n_r}$$

with the number of predicted responses n_y, the number of samplings taken by the robot n_s and the number of reactors n_r. In the same way, corresponding predicted response variables are defined as $y_r(t; u_r, \theta) \in \mathbb{R}^{n_y}$, with $t \in \mathcal{T}^s$ and $r \in \mathcal{R}$ and collected in the vector

$$Y(U, \theta) \in \mathbb{R}^{n_y \cdot n_s \cdot n_r} \qquad 2$$

Note that the predicted responses are obtained from the solution of Eq. (1) for each reactor r and therefore depend on u_r and θ.

Model parameters θ are estimated by maximum likelihood estimation minimizing the quadratic residual (Bard 1974):

$$\hat{\theta} := \arg\min_{\theta} \frac{1}{2} (Y(U,\theta) - Y^m)^T \, (C^m)^{-1} \, (Y(U,\theta) - Y^m) \qquad 3$$

with the weighting matrix $C^m \in \mathbb{R}^{n_y \cdot n_s \cdot n_r \times n_y \cdot n_s \cdot n_r}$ being the variance-covariance matrix of the measurement errors in the data. C^m is assumed to be a diagonal matrix with the variance σ_i^2 of each measurement i in its diagonal entries. The precision of a maximum-likelihood estimate $\hat{\theta}$ can be analysed by evaluation of the so called Fisher Information Matrix (FIM) $\mathrm{F}(U,\hat{\theta})$ or its inverse (lower bound on the covariance matrix $C^{\theta}(U,\hat{\theta})$) (Bard 1974). The design of an optimal experiment for improving parameter precision therefore minimizes some metric of $C^{\theta}(U,\hat{\theta})$ by optimally choosing the inputs or experiment design variables U subject to equality and inequality constraints. In this contribution we apply the so called A-optimal criterion ϕ^A to formulate the optimal experiment design objective function:

$$\phi^A(u,\hat{\theta}) := \frac{1}{n_p} \operatorname{tr}\left(F^{-1}(U,\hat{\theta})\right) \qquad 4$$

Online optimal experimental re-design

The real information content of an ED depends on the accuracy of the assumed parameter values (initial parameter guess). Unfortunately, uncertainties in the initial parameter guess and initial conditions cause a mismatch between computed and experimental outputs leading to suboptimal ED. This mismatch can be reduced by an adaptive online ED, where the experiment is iteratively re-designed as information is generated. This idea has been discussed in previous publications and is described in detail by (Barz et al. 2012).

In sequential OED, the identifiability of the parameter estimation problem can be studied and solved offline. In contrast, when performing a recursive fit of the model and a design of each sampling setup, an efficient solution of nonlinear and possibly ill-conditioned problems is required. (Barz et al. 2012) propose the addition of a local parameter identifiability

analysis method to select the parameters such that well-conditioned optimization problems are guaranteed. Therefore, it is possible to redesign the running experiment in order to take full advantage of the present state of information and plan the following step (optimal input $u(t)$, sampling time t_i, sampling setup) after each measurement. This is, to the best of our knowledge, the first attempt to carry out an online parallel design of dynamic experiments using real automated cultivations systems.

The F matrix needs to be well conditioned as it is used to approximate the covariance matrix of the parameters C_θ by inversion. If this is not the case, e.g. due to insufficient measurement data or correlations in the parameters, a regularization method is applied to approximate F by a well-conditioned matrix. The regularization is based on the Subset Selection (SsS) method which is very effective when applied to ED problems with ill-conditioned matrices (López Cárdenas et al. 2015). Moreover, the SsS method also proposes a reduction of the parameter space (and the PE and ED problem) by finding the parameters which cannot be identified for the existing measurement data (identifiability analysis).

Experimental implementation of SWORD

The duration of the SWORD experiment is 6 h and is divided in 18 cycles of 20 min. A detailed description of one cycle is given in Figure 3. The initial experimental design is carried out up to minute 140 (first re-design is available). The first at-line data is available after 100 minutes. Once new at-line data is available, i) the identifiability of the parameters is checked, ii) a PE is solved for max. 20 min, and iii) the parameter estimates obtained are passed to re-compute the optimal experiment (max. time 20 min). This procedure is repeated three more times until 300 min where the last re-design is calculated. The experiment run is depicted in Figure 4.

Column C reads the real time. Column A shows the tasks performed by the first computer, namely the PE (green; cell [4,Q]) and SsS and identifiability analysis (dark green; cell [5,Q])). Column B depicts the performance of the second computer which was responsible for the ED-SH optimizations of the short horizon (dark red; cell [6,Q])) and DE-LH random seed global optimizations computing the long horizon (red; cell [7,Q])), and the data that is available is depicted by bars (black; cell [10,Q]). The time points of the samples and designs can be seen in row 1 (columns D to V). For example: at 100 min (cell [7,C]) the data from sampling at 0 and 20 min is stored as depicted in the bar in cells [7,D- E] and the comp1 initiates the parameter estimation (cell [7,A]). We can also see the three optimization horizons (orange; cell [8,Q]), which were implemented at 140 min, 240 min, and 340 and finally the results of the long term horizon (blue; cell [9,Q]), which were available at 0 min, 220 min and 320 min. The run of the experiment is depicted at the example of biomass for reactor 1 in Figure 5.

Results and discussion

SWORD was able to fit the model despite the fact that the enzymatic assay for acetate quantification was inaccurate, proving the advantages of an online re-design approach Figure 6. The program could obtain reliable parameter estimates by re-designing the experiment taking the unexpectedly high variance in acetate into account. Further screenshots of the monitoring station are presented in the Appendix to show the development of the experiment and the inputs selected over time. The reactor replicates proved to be statistically equal according to the hypothesis test and confidence intervals (Montgomery and Runger 2010) carried out on the cell dry weight data over the complete experiment, see Figure 7.

In Figure 8, the normal probability density function of the whole parameter vector centered at the normalized parameter estimate $\hat{\theta}$ with variance σ^2 are displayed. The size of the variability of the parameters with large uncertainty can be observed (i.e., Yam and Yofm). 'qSmax' has the lowest relative standard deviation $\sigma_r = 0.56\,\%$ whereas 'Yam' and 'Yofm' have the highest ones with 341% and 822 % respectively.

The identifiability list for this problem can be found in Table 1. It should be stressed that due to parameter correlations, the ranking between variance, sensitivity and SsS are not equal. Still parameter with the largest effect on the outputs is also 'qSmax' and the parameter with the smallest effect on the measurable variables after all orthogonal projections is 'Yofm' showing consistency with the precision results above mentioned. In addition, the sensitivity matrix is not full-rank but 23, i.e. only 23 parameters are identifiable. The unidentifiable parameters being 'Yam', and 'Yofm'. Additionally, the sensitivity ranking list built based on the overall output-parameter sensitivity information is displayed. The metric used is the sensitivity measure $\delta = \frac{1}{\mathrm{N_y},N_m,N_e} * \sum s_{ij}^2$ with $s_{ij}^2 = \frac{\partial y_i}{\partial \theta_j}$ which considers the norm of the sensitivity of the whole predicted response variables to a change in the parameter θ_j.

Table 1. Parameter ranking list according to: parameter variance σ^2, relative standard deviation σ_r and mean sensitivity δ_j, of the predicted response variables i to a change in the parameter j.

Now let us show the advantages of the re-design of the experiment. We compare the results obtained in the real experiment against a simulation of the initial design with the final parameter values. The A-criterion, which represents the average coefficient of variation (normed standard deviation) of the parameter estimates is almost 50 times lower (235.86% vs. 4.83%) with SWORD considering all 23 identifiable parameters.

Furthermore, the relative variance σ_r of the parameters at the estimated vector θ_f is depicted in Figure 9 in ascending order (light green bars). A comparison between the parameter

precision at the initial guess θ_0 and the estimated vector θ_f shows the superiority of the estimated parameter vector θ_f over θ_0. The identifiability increased from only 13 parameters with a variance lower than 10% with the original design (dark brown bars) to 23 due to the continues re-design of the inputs. This proves that the final experiment cannot be computed beforehand since, due to wrong initial estimates, the predicted outputs are in reality suboptimal.

Conclusions

The online re-design of optimal experiments for 4x2 parallel fed-batch cultivations is possible using a moving horizon approach which reduces the experimental effort compared to sequential designs. In this case study, one experimental run generated sufficient information to fit a macro-kinetic model of *E. coli* W3110, significantly reducing time and costs of model validation. This is achieved by re-designing the optimal strategy of the parallel cultivations using the existing data as it is generated during the experiment.

The results show that the SWORD program can overcome complications related with ill-posed optimization problems and highly nonlinear designs in parallel bioreactors. This is a first step towards new model based tools that can fully exploit the advantages of LHS and speed up the development from screening to production in bioprocesses. Faster, cheaper, and more efficient experiments will encourage the use of mechanistic models, allowing for the application of computer-aided tools in design, monitoring, control, and optimization of bioprocesses.

Acknowledgments

The authors acknowledge financial support by the German Federal Ministry of Education and Research (BMBF) within the Framework Concept ”Research for Tomorrow’s Production”

(project no. 02PJ1150, AUTOBIO project) managed by the Project Management Agency Karlsruhe (PTKA).

We would like to thank Sven Olof Enfors for fruitful discussions during the development of the model, Tiffany Wong, and students Michael Heisser, Lorenz Thielert, Sebastian Hans, and Terrance Wilms.

References

Ataman M, Hatzimanikatis V. 2015. Heading in the right direction: thermodynamics-based network analysis and pathway engineering. Current opinion in biotechnology 36:176-182.

Baltes M, Schneider R, Sturm C, Reuss M. 1994. Optimal experimental design for parameter estimation in unstructured growth models. Biotechnology Progress 10(5):480-488.

Bard Y. 1974. Nonlinear parameter estimation. New York: Academic Press New York.

Barz T, López Cárdenas DC, Arellano-Garcia H, Wozny G. 2012. Experimental evaluation of an approach to online redesign of experiments for parameter determination. AIChE Journal 59(6):1981-1995.

Brunk E, Neri M, Tavernelli I, Hatzimanikatis V, Rothlisberger U. 2012. Integrating computational methods to retrofit enzymes to synthetic pathways. Biotechnology and Bioengineering 109(2):572-582.

Cruz Bournazou MN, Arellano-Garcia H, Wozny G, Lyberatos G, Kravaris C. 2012. ASM3 extended for two-step nitrification–denitrification: a model reduction for sequencing batch reactors. Journal of Chemical Technology and Biotechnology 87(7):887-896.

Cruz Bournazou MN, Barz T, Nickel D, Neubauer P. 2014a. Sliding-Window Optimal Experimental Re-Design in Parallel Microbioreactors. Chemie Ingenieur Technik 86(9):1379-1380.

Cruz Bournazou MN, Hooshiar K, Arellano-Garcia H, Wozny G, Lyberatos G. 2013. Model Based Optimization of the Intermittent Aeration Profile for SBRs under Partial Nitrification. Water Research 47(10):3399-3410.

Cruz Bournazou MN, Junne S, Neubauer P, Barz T, Arellano-Garcia H, Kravaris C. 2014b. An approach to mechanistic event recognition applied on monitoring organic matter depletion in SBRs. AIChE Journal 60(10):3460-3472.

Delgado San Martín JA, Cruz Bournazou MN, Neubauer P, Barz T. 2014. Mixed integer optimal control of an intermittently aerated sequencing batch reactor for wastewater treatment. Computers & Chemical Engineering 71:298-306.

Franceschini G, Macchietto S. 2007a. Model-based design of experiments for parameter precision: State of the art. Chem Eng Sci 63(19):4846–4872.

Franceschini G, Macchietto S. 2007b. Validation of a model for biodiesel production through model-based experiment design. Ind Eng Chem Res 46(1):220-232.

Galvanin F, Barolo M, Macchietto S, Bezzo F. 2011. Optimal design of clinical tests for the identification of physiological models of type 1 diabetes in the presence of model mismatch. Medical and Biological Engineering and Computing 49(3):263-277.

Galvanin F, Barolo M, Pannocchia G, Bezzo F. 2012. Online model-based redesign of experiments with erratic models: a disturbance estimation approach. Computers & Chemical Engineering 42:138-151.

Gernaey K, Petersen B, Dochain D, Vanrolleghem PA. 2002. Modeling aerobic carbon source degradation processes using titrimetric data and combined respirometric–titrimetric data: Structural and practical identifiability. Biotechnology and Bioengineering 79(7):754-767.

Gill PE, Murray W, Saunders MA. 2002. SNOPT: An SQP algorithm for large-scale constrained optimization. SIAM Journal on Optimization 12(4):979-1006.

Glassey J, Montague G, Ward A, Kara B. 1994. Artificial neural network based experimental design procedure for enhancing fermentation development. Biotechnology and Bioengineering 44(4):397-405.

Glauche F, Pilarek M, Cruz Bournazou MN, Grunzel P, Neubauer P. 2016. Design of Experiments based high-throughput strategy for development and optimization of efficient cell disruption protocols. Engineering in Life Sciences:accepted.

Hidalgo-Bastida LA, Thirunavukkarasu S, Griffiths S, Cartmell SH, Naire S. 2012. Modeling and design of optimal flow perfusion bioreactors for tissue engineering applications. Biotechnology and Bioengineering 109(4):1095-1099.

Hindmarsh AC, Brown PN, Grant KE, Lee SL, Serban R, Shumaker DE, Woodward CS. 2005. SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. ACM Transactions on Mathematical Software (TOMS) 31(3):363-396.

Holmström K, Edvall MM, Göran AO. TOMLAB for large-scale robust optimization; 2003. p 218-222.

Junker BH, Wang HY. 2006. Bioprocess monitoring and computer control: Key roots of the current PAT initiative. Biotechnology and Bioengineering 95(2):226-261.

Knepper A, Heiser M, Glauche F, Neubauer P. 2014. Robotic Platform for Parallelized Cultivation and Monitoring of Microbial Growth Parameters in Microwell Plates. Journal of laboratory automation 19(6):593-601.

Kokkalis A, Thygesen UH, Nielsen A, Andersen KH. 2014. Limits to the reliability of size-based fishing status estimation for data-poor stocks. Fisheries Research 171:4-11.

Korkel DMS. 2002. Numerische Methoden fuer Optimale Versuchsplanungsprobleme bei nichtlinearen DAE-Modellen: Ruprecht-Karls-Universitaet Heidelberg.

Körkel S, Kostina E, Bock HG, Schlöder JP. 2004. Numerical methods for optimal control problems in design of robust optimal experiments for nonlinear dynamic processes. Optim Method Softw 19(3-4):327-338.

Kozak K. 2014. Large scale data handling in biology: Bookboon.

Kreutz C, Timmer J. 2009. Systems biology: experimental design. FEBS Journal 276(4):923-942.

Legmann R, Schreyer HB, Combs RG, McCormick EL, Russo AP, Rodgers ST. 2009. A predictive high-throughput scale-down model of monoclonal antibody production in CHO cells. Biotechnology and Bioengineering 104(6):1107-1120.

Lin HY, Mathiszik B, Xu B, Enfors SO, Neubauer P. 2001. Determination of the maximum specific uptake capacities for glucose and oxygen in glucose-limited fed-batch cultivations of Escherichia coli. Biotechnol Bioeng 73(5):347-57.

Long Q, Liu X, Yang Y, Li L, Harvey L, McNeil B, Bai Z. 2014. The development and application of high throughput cultivation technology in bioprocess development. Journal of Biotechnology 192:323-338.

López Cárdenas DC, Barz T, Körkel S, Wozny G. 2015. Nonlinear ill-posed problem analysis in model-based parameter estimation and experimental design. Computers & Chemical Engineering 77:24-42.

López Cárdenas DC, Barz T, Peñuela M, Villegas A, Ochoa S, Wozny G. 2013. Model-based identifiable parameter determination applied to a simultaneous

saccharification and fermentation process model for bio-ethanol production. Biotechnology Progress 29(4):1064-1082.

Lu F, Toh PC, Burnett I, Li F, Hudson T, Amanullah A, Li J. 2013. Automated dynamic fed-batch process and media optimization for high productivity cell culture process development. Biotechnology and Bioengineering 110(1):191-205.

Lutz MW, Menius JA, Choi TD, Laskody RG, Domanico PL, Goetz AS, Saussy DL. 1996. Experimental design for high-throughput screening. Drug discovery today 1(7):277-286.

Mayer E, Raisch J. 2004. Time-optimal scheduling for high throughput screening processes using cyclic discrete event models. Mathematics and Computers in Simulation 66(2):181-191.

Medema MH, van Raaphorst R, Takano E, Breitling R. 2012. Computational tools for the synthetic design of biochemical pathways. Nature Reviews Microbiology 10(3):191-202.

Montgomery DC, Runger GC. 2010. Applied statistics and probability for engineers: John Wiley & Sons.

Neubauer P, Junne S. 2010. Scale-down simulators for metabolic analysis of large-scale bioprocesses. Current opinion in biotechnology 21(1):114-21.

Nickel DB, Cruz Bournazou MN, Wilms T, Neubauer P, Knepper A. submitted. Online bioprocess data generation, analysis and optimization for parallel fed-batch fermentations at mL scale.

Oddone GM, Lan CQ, Rawsthorne H, Mills DA, Block DE. 2007. Optimization of fed-batch production of the model recombinant protein GFP in Lactococcus lactis. Biotechnology and Bioengineering 96(6):1127-1138.

Panula-Perälä J, Siurkus J, Vasala A, Wilmanowski R, Casteleijn MG, Neubauer P. 2008. Enzyme controlled glucose auto-delivery for high cell density cultivations in microplates and shake flasks. Microb Cell Fact 7(3):1.

Puskeiler R, Kaufmann K, Weuster-Botz D. 2005. Development, parallelization, and automation of a gas-inducing milliliter-scale bioreactor for high-throughput bioprocess design (HTBD). Biotechnology and Bioengineering 89(5):512-523.

Rosen C, Vrecko D, Gernaey K, Pons M-N, Jeppsson U. 2006. Implementing ADM1 for plant-wide benchmark simulations in Matlab/Simulink. Water Science and Technology 54(4):11-19.

Schäpper D, Alam MNHZ, Szita N, Lantz AE, Gernaey KV. 2009. Application of microbioreactors in fermentation process development: a review. Analytical and bioanalytical chemistry 395(3):679-695.

Shockley KR. 2015. Quantitative high-throughput screening data analysis: challenges and recent advances. Drug discovery today 20(3):296-300.

Skanda D, Lebiedz D. 2010. An optimal experimental design approach to model discrimination in dynamic biochemical systems. Bioinformatics 26(7):939-945.

Stigter J, Vries D, Keesman K. 2006. On adaptive optimal input design: a bioreactor case study. AIChE Journal 52(9):3290-3296.

Su WW, Liu B, Lu W-B, Xu N-S, Du G-C, Tan J-L. 2005. Observer-based online compensation of inner filter effect in monitoring fluorescence of GFP-expressing plant cell cultures. Biotechnology and Bioengineering 91(2):213-226.

Takors R, Wiechert W, Weuster-Botz D. 1997. Experimental design for the identification of macrokinetic models and model discrimination. Biotechnology and Bioengineering 56(5):564-576.

Valgepea K, Adamberg K, Nahku R, Lahtvee P-J, Arike L, Vilu R. 2010. Systems biology approach reveals that overflow metabolism of acetate in Escherichia coli is triggered by carbon catabolite repression of acetyl-CoA synthetase. BMC Systems Biology 4(1):166.

Venkata Mohan S, Chandrasekhara Rao N, Krishna Prasad K, Murali Krishna P, Sreenivas Rao R, Sarma PN. 2005. Anaerobic treatment of complex chemical wastewater in a sequencing batch biofilm reactor: Process optimization and evaluation of factor interactions using the Taguchi dynamic DOE methodology. Biotechnology and Bioengineering 90(6):732-745.

Zavrel M, Schmidt T, Michalik C, Ansorge-Schumacher M, Marquardt W, Büchs J, Spiess AC. 2008. Mechanistic kinetic model for symmetric carboligations using benzaldehyde lyase. Biotechnology and Bioengineering 101(1):27-38.

Zhu Y, Huang B. 2011. Constrained receding-horizon experiment design and parameter estimation in the presence of poor initial conditions. AIChE Journal 57(10):2808-2820.

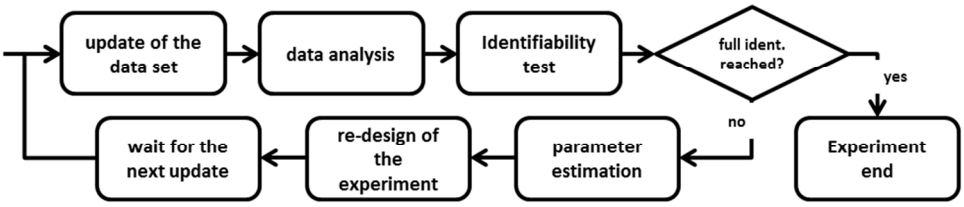

Figure 1: simplified flow diagram of the online re-design of experiments procedure

172x39mm (150 x 150 DPI)

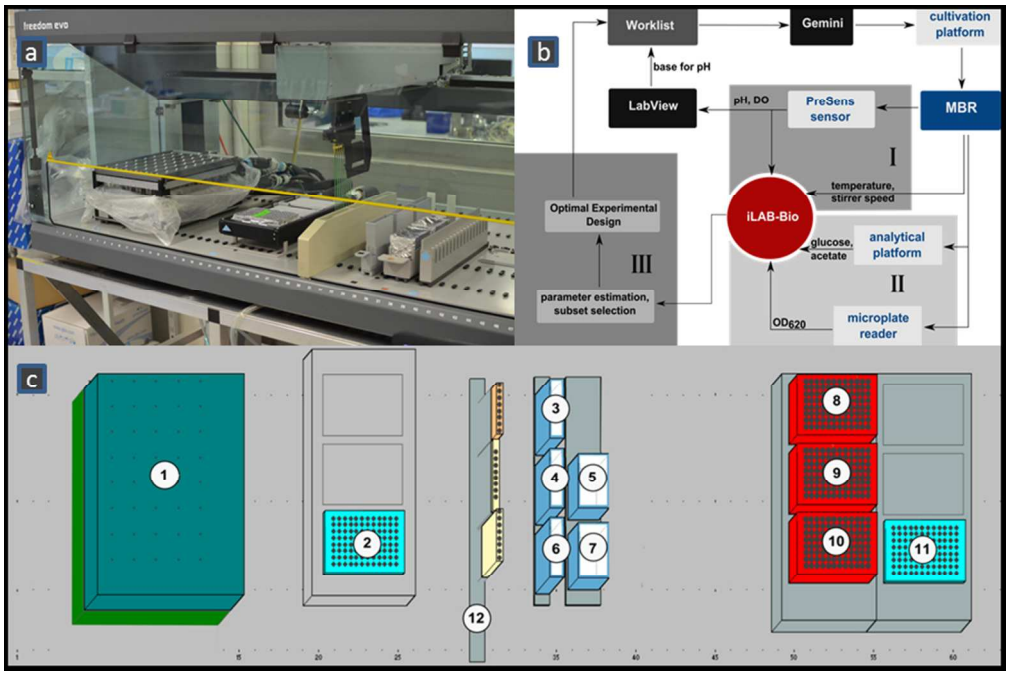

Figure 2 (a) bioREACTOR 48 integrated into the Freedom Evo LHS. (b) Data flow concept of SWORD. On-line signals from the mini-bioreator system (MBR, I) and measured data from samples (II) is stored in a database, which can be accessed for the sliding window re-design (III). The LHS was controlled via MATLAB. (c): Deck layout of the Freedom Evo LHS: bioREACTOR 48 (1), Additives: Reagent A (2), EnPresso B defined (8), 40 g L-1 sodium acetate (9), glucose (10); sample plate (11).The steel needles were washed in deionised water (3), twice in EtOH (4, 6) and in sterile deionised water (7) pH adjustment was done with 6,25 % ammonia (5).

152x101mm (150 x 150 DPI)

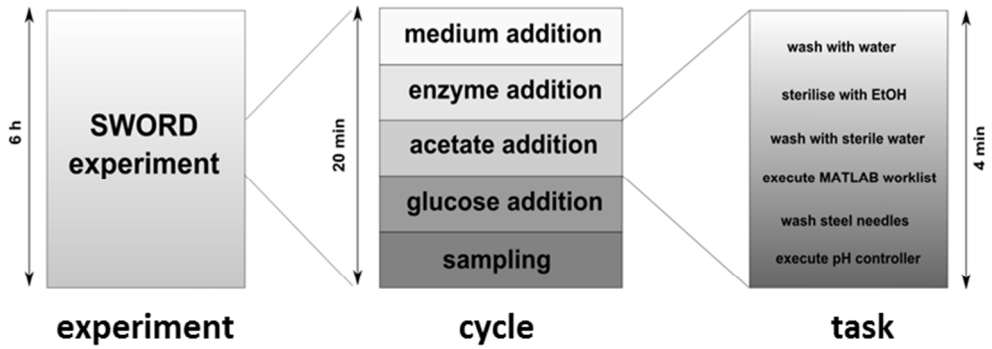

Figure 3: Scheduling of the SWORD experiment. Cycles of 20 min were repeated for 6h. During these intervals, in cycles of 4 min EnPresso B defined medium, Reagent A, acetate and glucose were added to the cultivations. Afterwards, 300 µL samples were taken from each reactor. Each addition routine consisted of several subroutines (right).

133x51mm (150 x 150 DPI)

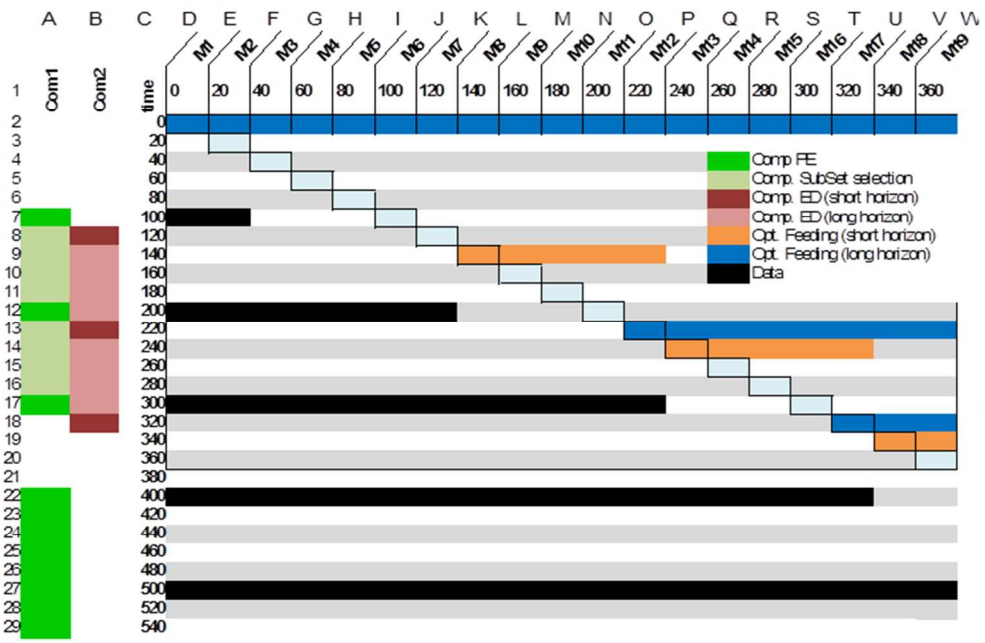

Figure 4: Time schedule of the experiment. Following time column C, the development of the experimental setup is to be seen (starting at the top). The activity of the computers (computer1 PE in column A, and computer2 ED in column B). Availability of at line measurement (biomass, glucose, and acetate) are represented by black bars in relation to measurement number and sampling time, row 1.

127x81mm (150 x 150 DPI)

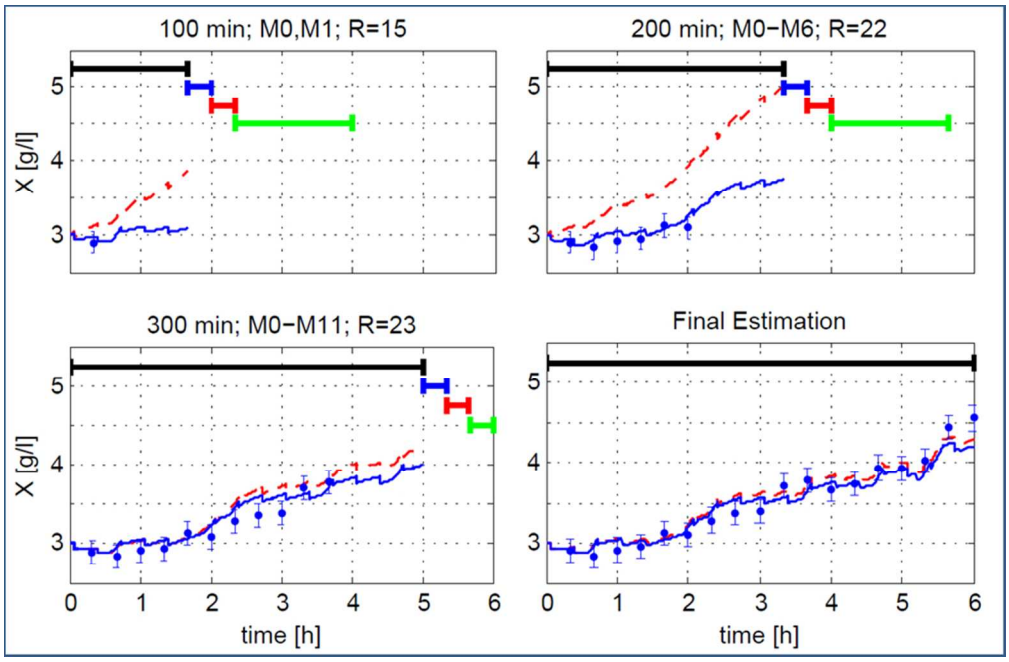

Figure 5: Evolution of SWORD experiment at the example of cell dry weight of the firs reactor (from plot A to D). Red dashed and blue lines represent the simulation results before and after parameter estimation respectively. The bars represent the horizon of available online and at-line data (black), the time required to compute the parameter estimation (blue), the time for ED optimization (red), and horizon length of the experimental re-design (green).

165x107mm (150 x 150 DPI)

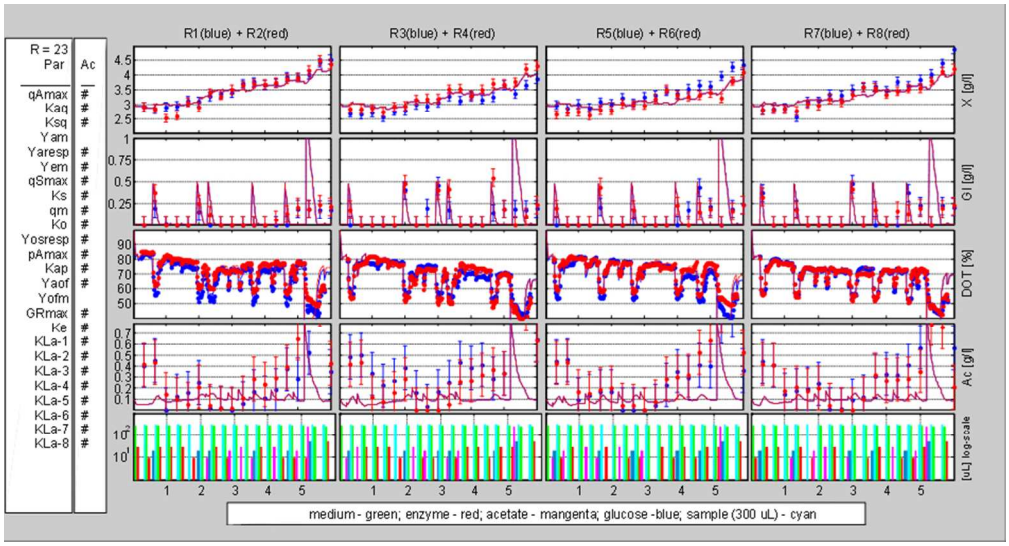

Figure 6: Screenshot of the monitoring station. Data collected over the complete experiment. Continuous lines represent the simulations. The parameters are shown in the left side of the Monitor. Active parameters (considered in the PE) are marked with "#" on the left side. Measurements vs. simulations of all four reactor pairs (first reactor blue second reactor red) are depicted in a 5x4 format. Column-wise beginning from the upper graph, cell dry weight (X) in grams per liter is first plotted followed by glucose concentration (Gl) in grams per liter, dissolved oxygen tension (DOT) in % of saturation, and acetate (Ac) in grams per liter in the fourth graph. In the semi-log graph at the bottom of each column, the volumes in micro liters added every 4 min (or extracted for sampling) are depicted with bars. Following the order of addition in each cycle, medium additions are represented by green, enzyme by red, acetate by magenta, glucose by blue and samples (fixed at 300 µl) by cyan bars. While this screenshot was taken, the additions between time 140 and 240 min were re-designed with the new parameter values.

160x85mm (150 x 150 DPI)

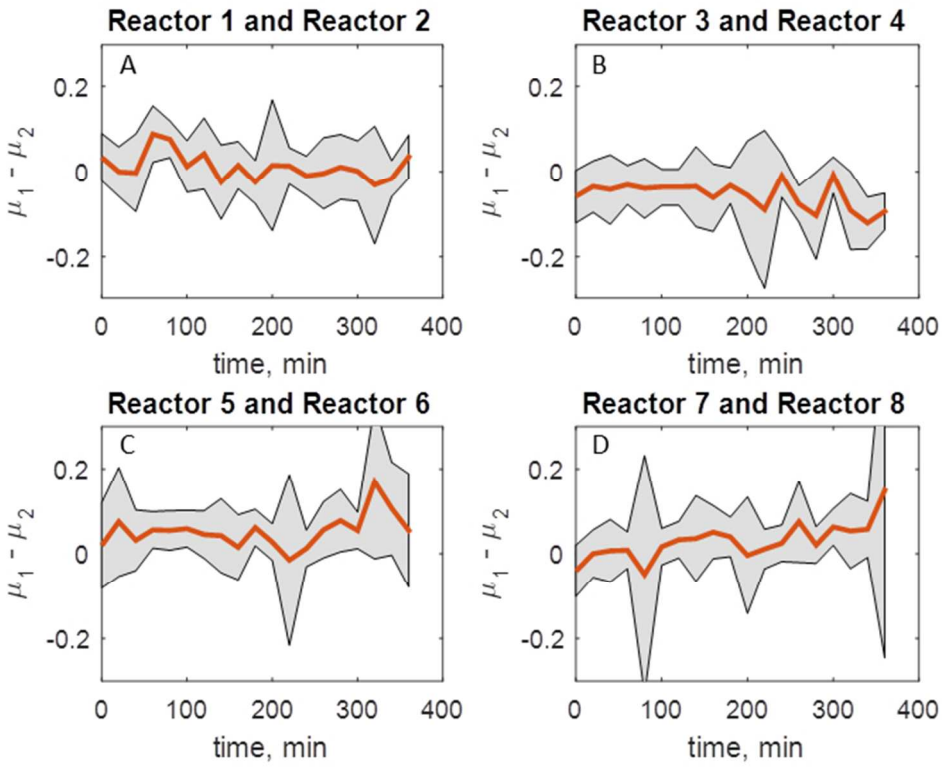

Figure 7: Graphical representation of the 95% confidence intervals (shaded area) of the difference in the means μ_1-μ_2 (solid line) for 4 experimental conditions (2 reactor replicates) over the whole experimental run. μ_1 and μ_2 represent the means of the dry biomass of each reactor in a replicate. A confidence interval containing the zero-value suggests no difference in the means at the evaluated experimental point. The percentages of accepted measurements are: A=89%, B=79%, C=68%, and D=95%.

120x96mm (150 x 150 DPI)

Parameter	σ^2	σ_r	δ	Ranking			Parameter	σ^2	σ_r	δ	Ranking		
				Variance	Sensitivity	SsS					Variance	Sensitivity	SsS
'qSmax'	3.13E-05	1%	19.543	1	1	1	'Kaq'	1.02E-03	3%	7.272	14	19	16
'qAmax'	3.97E-04	2%	16.100	2	16	10	'Ks'	1.50E-03	4%	6.933	15	21	17
'KLa-4'	5.01E-04	2%	11.309	3	5	4	'Ke'	2.48E-03	5%	6.889	16	17	19
'KLa-3'	5.01E-04	2%	9.825	4	7	5	'Yem'	6.05E-03	8%	5.334	17	3	3
'KLa-1'	5.02E-04	2%	9.427	5	12	9	'Kap'	7.24E-03	9%	4.220	18	18	18
'KLa-2'	5.03E-04	2%	9.358	6	14	14	'Yaof'	9.73E-03	10%	3.610	19	15	15
'KLa-7'	5.04E-04	2%	9.236	7	10	7	'Ksq'	1.16E-02	11%	1.461	20	23	22
'KLa-8'	5.04E-04	2%	9.229	8	11	8	'Yosresp'	1.26E-02	11%	1.213	21	6	20
'KLa-5'	5.07E-04	2%	9.229	9	9	6	'Ko'	1.74E-02	13%	0.709	22	22	21
'KLa-6'	5.08E-04	2%	8.997	10	13	11	'qm'	7.90E-02	28%	0.626	23	20	23
'pAmax'	5.62E-04	2%	8.932	11	4	13	'Yam'	1.16E+01	341%	0.218	24	24	24
'Yaresp'	7.11E-04	3%	8.629	12	2	2	'Yofm'	6.75E+01	822%	0.111	25	25	25
'GRmax'	8.14E-04	3%	8.029	13	8	12							

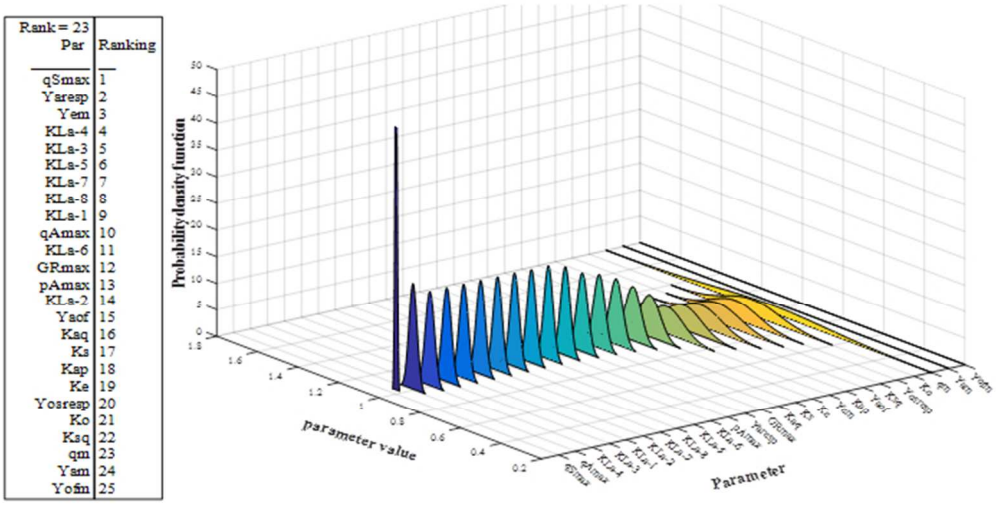

Figure 8: Estimator analysis through normal probability density functions of the parameters.

118x62mm (150 x 150 DPI)

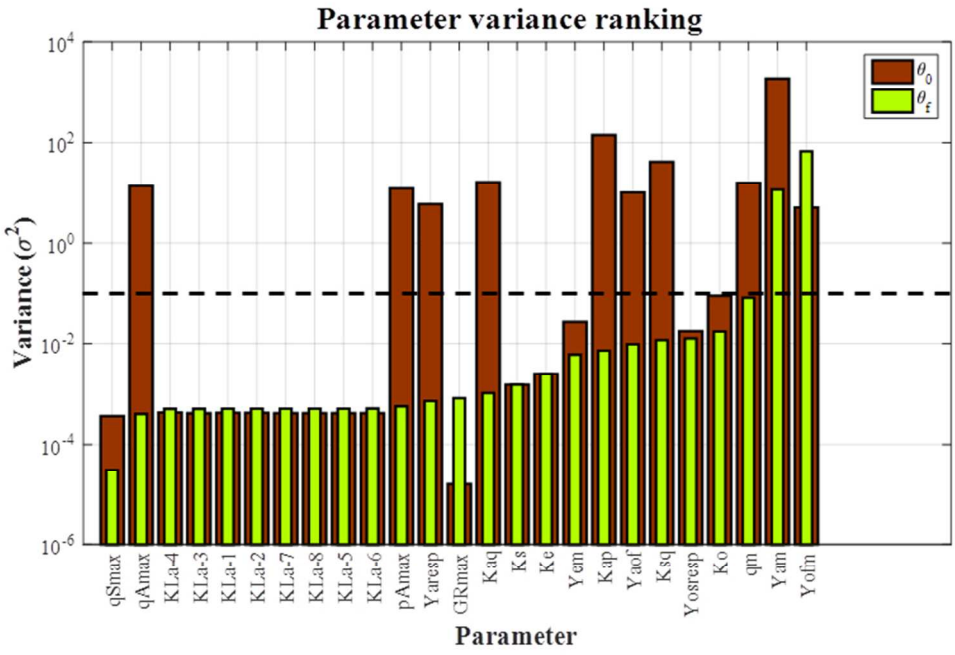

Figure 9: Estimator analysis with parameter precision ranking in terms of relative standard deviation. Comparison between the variance obtained with the original design (brown bars) and with 3 re-designs (green bars).

133x89mm (150 x 150 DPI)