Efficient Generation of Models of Fed-Batch Fermentations for Process Design and Control

Lukas Hebing∗ Tobias Neymann∗∗ Tobias Thüte∗∗∗ Alexander Jockwer∗∗∗ Sebastian Engell∗

∗ Department of Biochemical and Chemical Engineering, TU Dortmund, D-44221 Dortmund, Germany
∗∗ Bayer Technology Services, D-51368 Leverkusen, Germany
∗∗∗ Bayer HealthCare AG, D-42113 Wuppertal, Germany

Abstract: The paper presents a fast and efficient approach to the generation of process models for fed-batch fermentation processes. Elementary Modes (EM) with kinetic terms are used as macro-reactions in the model. A method based upon dynamic metabolic flux analysis (DMFA) is used to find an optimal selection of EM based on concentration measurements instead of reaction rates that must be calculated from the measurements. A multi-objective genetic algorithm is used to investigate the tradeoff between the number of EM and the difference of the estimated concentration profiles from the available measurements. Thus the number of macro-reactions for the model is kept small but the measured data are described sufficiently well by the selected EM. The influence of the process conditions on each macro-reaction rate are identified by a statistical analysis and kinetic expressions are fitted to the estimated evolutions by solving multiple simple optimization problems instead of a multi-dimensional large-scale nonlinear parameter optimization problem. The resulting process model provides a good fit of the measurement data. Provided that the data is generated by a systematic approach, e.g., using a design of experiments strategy, the modelling procedure provides models of adequate complexity that are valid within the so-called design space of the process.

Keywords: Elementary Modes, Macroscopic Modelling, Dynamic Metabolic Flux Analysis, Parameter Estimation

1. INTRODUCTION

With the increasing capabilities of process analytical technology (PAT) and the Quality-By-Design initiative of the FDA (QbD), mathematical models of bioprocesses gain increasing importance (McCurdy, 2011). Shorter process development times and optimized process trajectories as well as online applications as e.g. optimal control are possible benefits of using mathematical models. However, the realization of these benefits depends on the accuracy of the process models and on the effort that is needed to develop them. The models should represent the significant influences on the evolution of the culture but not have a too high complexity so that they can be developed and parameterized from limited data sets. Dynamic models of fermentation processes are commonly used in research and industry. They differ in the level of abstraction. Often formal (external) kinetics are used similar to the modelling of chemical reactions (Leifheit et al., 2007). This approach completely neglects the available knowledge on the biochemical pathways in the cells. On the other hand, detailed biochemical models are too complex to estimate the kinetics of each and every reaction from the usually scarce data that is available. Black-Box approaches as e.g. neural networks have also been used for state estimation and control (Chen et al., 2006), but due to their unstructured nature, they also require a lot of data to fit the parameters which often is not available. Moreover, as all parameters are data-dependent, even small changes of the organisms require a complete re-estimation. A reasonable compromise between formal kinetics and extremely detailed models is the use of macro-reactions. In the definition of the macro-reactions, the biochemical information on the stoichiometry and irreversibility of internal reactions is maintained but the reaction networks are considered as being in internal equilibrium so that only the kinetics of merged internal reactions have to be estimated. Elementary Modes (EM) are the most common choice for macro-reactions, which are consistent with the biological knowledge about the organism (Gao et al., 2007), (Provost, 2006). The main steps in the development of such models are (1) the selection of the macro-reactions and (2) the selection of their kinetics. As the number of possible EM can be very large, even for relatively small metabolic networks, a selection of EM as macro-reactions for the model is needed. This can be done by selection algorithms, which compare cell-specific uptake- or secretion rates with those that are calculated from possible ones of the network (Provost, 2006), (Soons et al., 2010) or compute minimal decomposition on the basis of the network together with uptake- or secretion rates (Jungers et al., 2011). However, to use uptake- or secretion rates is error prone, as these rates have to be estimated from noisy data and different approximation and
estimation techniques may lead to very different estimates. To overcome this problem, a data-driven method for the selection of the EM is proposed here, which uses measured concentration data directly instead of estimated rates (cf. section 3).

The reaction rates of the selected macro-reactions which can be combinations of hundreds of reactions which are assumed to be in internal equilibrium are described by simple kinetic equations that depend on the values of measured variables. These equations can be formal kinetics that are chosen based upon biochemical insight (e.g. limitations of substrates described by Monod-kinetics) or additional terms from inhibitions from by-products or pH-dependencies. For the choice and the parameterization of these kinetic relationships, a data-driven approach is presented here. The evolution of the (estimated) reaction rates is used for a statistical analysis to reveal correlations between the measured process data and the reaction rates. A method for the fast design and testing of kinetics based on these correlations is presented in section 4. Finally all model parameters are fitted to the data. This usually is a difficult task because of the existence of many local minima, but in our approach it is simplified as prior estimates of the kinetic parameters are available and the structure of the kinetics has already been validated.

These steps lead to a systematic modelling strategy in which basic knowledge about the organism is used together with data and statistical methods to generate grey-box models of a fermentation process. Of course, the resulting model will only be reliable in the operating region where sufficient experimental data has been collected.

2. MODEL STRUCTURE

The metabolism of the cell is described by a set of macro-reactions that connect the external metabolites. All macro-reactions of the model are captured in the stoichiometric matrix \( L \). The elementary balance of a component (metabolite or substrate) \( i \) is:

\[
\frac{dc_i}{dt} = D \cdot (c_{i,in} - c_i) + L_i \cdot \mathcal{R} \cdot X_v
\]

where \( D \) denotes the Dilution rate, \( c_{i,in} \) the concentration of component \( i \) in the feed, \( L_i \) the \( i \)-th row of the stoichiometric matrix \( L \), \( \mathcal{R} \) the vector of cell-specific reaction-rates and \( X_v \) the concentration of vital cells. If cellular growth is part of the metabolic network (e.g. as a lumped reaction of cell-internal substrates (Nolan and Lee, 2011)), the balance of vital cells can also be represented by (1).

For gaseous components, the mass transport from the gaseous phase is added:

\[
\frac{dc_i}{dt} = D \cdot (c_{i,in} - c_i) + L_i \cdot \mathcal{R} \cdot X_v + k_i a \cdot (c^*_{i} - c_i)
\]

where \( c^*_i \) is the liquid concentration which is in equilibrium with the concentration in the gaseous phase and \( k_i a \) is the mass-transfer coefficient.

The goal of the modelling procedure which is presented in this paper is to find a suitable set of reactions, captured in the stoichiometric matrix \( L \), and of kinetic rate expressions that enable the calculation of \( \mathcal{R} \) in dependence on the measurable variables of the process as e.g. concentrations of components in the media, temperature, pH value or others.

In our approach, elementary modes (EM) are used as possible macro-reactions. As described by Provost (2006) and Jungers et al. (2011), the stoichiometry of the external reactions of the EM (\( K \)) can be calculated by multiplying the (external) stoichiometric matrix \( P \) with the EM-matrix \( E \) that describes the possible stationary combinations of the internal reactions.

\[
K = P \cdot E
\]

Each column of \( K \) represents a possible macro-reaction in the model. As mentioned above, the number of columns of \( K \) is potentially very large. The Columns of \( L \) are chosen from those of \( K \).

3. CHOICE OF THE EM

The selection of a set of EM that are used as macro-reactions in the model (which provides the matrix \( L \)) is carried out in two consecutive steps: 1. Pre-reduction of the matrix \( K \) (see section 3.1) and 2. Selection of the set of EM (see section 3.3). The set of EM that is used in the model (1) should be able to capture the metabolic activity of the cells over time sufficiently well. Methods to quantify the quality of the fit are discussed in section 3.2. The number of reactions (EM) that are selected determines the model complexity as kinetic expressions for every EM have to be defined and to be parameterized from the available data. The selection of EM in section 3.3 is therefore carried out as a multi-objective optimization with two objectives: correctly capturing the metabolic activity and minimizing the number of selected EM.

3.1 Pre-reduction of the set of EM

Depending on the metabolic network, the number of possible EM can be very large. But for a macroscopic representation only a very small subset of EM is sufficient. Prior to the selection of the final set of EM for the model via multi-objective optimization (see section 3.3) a data-independent pre-reduction procedure is used to discard very similar EM.

The angle between the columns of \( K \) is used as the selection criterion. Beginning with a randomly selected elementary mode, all other modes with a cosine-similarity higher than a specified value are deleted from the matrix \( K \). This procedure is repeated with another of the remaining EM until no cosine similarity of two EM is above the specified value.

The validity of the reduction can be tested based upon the cell-specific uptake or secretion rates over time \( g(t) \). By a nonnegative-least-squares optimization, an optimal value of \( \omega \) can be calculated for every time-step \( t_k \), such that:

\[
\min \left( \| K \cdot \omega(t_k) - q(t_k) \|_2^2 \right)
\]

s.t.

\[
\omega(t_k) \geq 0
\]

Thus, \( \omega(t_k) \) is a (time varying) decomposition of the flux \( q(t) \) into the reduced set of EM. For each reduction step \( \omega \) of the matrix \( K \), the sum of squared residuals (SSR\( q \)) determines whether the physiologically important region of the process is still part of the solution space.
As discussed later (see section 3.2), the drawback of this evaluation is the usage of cell-specific rates, which are only available with significant uncertainty. However, for a first quantitative evaluation of the efficiency of the pre-reduction procedure and the choice of the maximum allowed value for the cosine similarity the above formula is useful. The main selection procedure directly uses measurement data (see section 3.3).

The outcome of the pre-reduction step is a matrix $K$ with a significantly reduced number of EM but with nearly the same spanning capacity as the original matrix. This means that the behaviour of the cells in all physiologically meaningful regions can still be represented by linear combinations of the remaining modes. The computational effort of the following selection algorithms is significantly lower, if the reduced matrix is used instead of the original matrix $K$.

### 3.2 Quality of a selected set of reactions

The ability of a selected set of EM to capture the metabolic behaviour has previously been tested based upon cell-specific uptake rates of media components $q$. Soons et al. (2010) determine the reaction rate vector $\mathbf{\omega}$ for a rate-vector $\mathbf{q}$ by solving (4) with a nonnegative-least-squares algorithm. This method can also be used for large sets of EM. The resulting rate vector is used for ranking and selecting the EM in $L$. However, the use of specific rates requires the numeric differentiation of noisy measurement data. Depending on the method used for approximate differentiation, the results can vary significantly and this may lead to wrong results. To overcome this problem, a new method is used here in which the evolution of the concentrations is approximated and not the cell-specific rates. The best-fit concentration profile for every component can be calculated for a specific set of EM. The sum of squared residuals between this concentration profile and the measurements can then be used as a criterion which describes the quality of the selection.

The method of Dynamic Metabolic Flux Analysis (DMFA) by Leighty and Antoniewicz (2011) provides a method to calculate an optimal concentration profile of media components for a batch process by varying reversible volumetric reactions (the so-called free fluxes) at inflection points. This problem can be formulated as a linear optimization problem. With this method, the flux distribution within a network can easily be calculated using measurement data from batch experiments.

For a estimation of the reaction rates of irreversible EM over time in a fed-batch process, the method of Leighty and Antoniewicz (2011) cannot be used. It is therefore extended by the following elements: By introducing lower bounds to the linear optimization problem, the concept is also applicable to irreversible macro-reactions (like EM). The data obtained from fed-batch experiments is corrected: Changes of concentrations which are not caused by the metabolic activity of the cells are compensated. The resulting concentrations are called shifted concentrations.

The value of the shifted concentration of component $i$ is:

$$c_{i,s}(t) = c_{i,0} - \int_0^t D \cdot (c_{i,im} - c_{i}(\tau)) d\tau$$

For gaseous components, the concentration changes due to the mass-transfer from the gaseous phase to the liquid phase are considered as well. The shifted concentrations can be calculated as pseudo batch measurements in the DMFA method. When the DMFA method with the above mentioned extensions is used for a selection of the EM, the sum of squared residuals ($SSR$) between the best possible concentration profile $\hat{c}_{i,s,DMFA}$ and the shifted pseudo-batch data $c_{i,s}$ together with the profile of the piecewise linear volumetric reaction rates for the EM ($R$) of the selection is calculated by solving a linear optimization problem, as each predicted concentration $\hat{c}_{i,s,DMFA}$ is a linear function of $R$:

$$\min_{\mathbf{R}} \sum_{k=1}^{N_t} \sum_{i,s} (\hat{c}_{i,s,DMFA}(t_k, R) - c_{i,s}(t_k))^2$$

The cell-specific reaction rates $\mathbf{q}$ are obtained by dividing $R$ by the number of vital cells at each point in time. The final SSR value is used for the quantification of the quality of fit and describes the ability of a subset of EM of capturing the metabolic behaviour of the cells. Typically, this problem can be solved within <1s of computation time. The uncertainty of the cell-specific rates $\mathbf{q}(t)$ is also computed. The temporal evolution of $\mathbf{q}(t)$ and $\mathbf{q}(t)$ are used in step 4 as prior estimates of the cell-specific reaction rates and its respective uncertainty.

### 3.3 Selection of a set of EM

To find a suitable set of EM, a multi-objective genetic algorithm is used. The degrees of freedom are binary variables which indicate whether an EM of the reduced matrix $K$ is part of the subset $L$ or not. Two objective values are taken into account: 1. The number of selected EM and 2. The final SSR value for the optimized prediction of the shifted concentrations ($\hat{c}_{i,s,DMFA}$).

The result of this multi-objective optimization is a Pareto-front with SSR values over the number of EM. Each point of the front represents an optimal selection of EM as macro-reactions and is a candidate for the final choice of EM for the model ($L$). However, as the complexity of the model increases with the number of reactions, a combination with a low number of EM but still acceptable SSR value may be preferred over models with higher numbers of EM and lower SSR values. Experience shows that when a critical number of EM is used, no significant improvement in the SSR value occurs for larger numbers of EM (see section 6.2).

If no combination of EM provides a sufficient fit to the data, an expansion or redesign of the network should be considered. The fit of the final model cannot be better than the fit of the optimized profile $\hat{c}_{i,s,DMFA}(t)$ to the shifted process data.

### 4. CHOICE OF KINETIC EXPRESSIONS

Kinetic rate expressions quantify the influences of the process conditions on the metabolic behaviour of the
production organism. In the model structure used here, EM are used as macro-reactions. Each EM represents a combination of internal reactions. The kinetics of the EM are not necessarily linked to a particular enzyme that controls the bottleneck in the network but rather quantify the combined effect of multiple limitations by several cell internal reactions or regulation cascades. Each reaction rate \( j \) is described by a kinetic expression of the following structure:

\[
r_j = r_{j,\text{max}} \prod_{i=1}^{N_l} \tilde{r}_{j,i} (c_i, T, pH, \ldots)
\]

where \( N_l \) is the number of limitations and \( \tilde{r}_{j,i} \) are kinetic terms that represent limitations. The values of \( \tilde{r}_{j,i} \) are between zero and one and depend on the process conditions. For the design of the kinetics, the limitations have to be identified and the mathematical expressions for each limitation have to be defined, parameterized, and validated. Provost (2006) and Gao et al. (2007) suggest that each substrate of an EM is also a possible limitation. This meaningful assumption can directly be derived from the reaction stoichiometry. However, additional influences e.g. from inhibiting by-products or the pH of the media cannot be identified from the stoichiometry. To overcome this lack of information, the prior estimate of the reaction rates \( r_j(t) \) can be used to detect additional influences on the reaction rates of the EM and to select appropriate expressions for the limitations.

The influences on the reaction rates are determined by Partial Least Squares (PLS), which is a well-known method for multivariate data-analysis. Correlations between the profile of reaction rates \( r_j(t) \) of multiple experiments that were performed under different conditions and possible influences are determined. Concentrations of measured components in the media (substrates, products, by-products etc.) and process conditions (temperature, pH, etc.) are normalized with respect to their mean value to compensate different scales and interpolated between measurement points if a measurement frequency is too low. A positive correlation between the reaction \( r_j(t) \) and any concentration \( c_i(t) \) can be represented by a Monod type limitation \( \tilde{r}_{j,i} \) (9), a negative correlation by a Haldane type limitation \( \tilde{r}_{j,i} \) (10).

\[
\tilde{r}_{j,i} = \left( \frac{c_i}{K_{m,j,i} + c_i} \right)^{n_{j,i}} \quad (9)
\]

\[
\tilde{r}_{j,i} = \left( \frac{K_{l,j,i} + c_i}{K_{l,j,i} + c_i} \right)^{n_{j,i}} \quad (10)
\]

Both limitation types can also be defined with a threshold-value for the concentration of the substrate or inhibitor under which the limitation is completely active (\( \tilde{r}_{j,i} = 0 \), for Monod type) or inactive (\( \tilde{r}_{j,i} = 1 \), for Haldane type). A possible limitation for the pH dependency can be expressed by a modified Gaussian type function:

\[
\tilde{r}_{j,pH} = \exp \left( -\frac{(pH - pH_{0,j})^2}{2\sigma_j^2} \right) \quad (11)
\]

### 4.1 Fast Evaluation of Kinetics and Generation of Initial Values

Prior to the final parameter estimation the parameters of the kinetic rate expressions can be determined by a simple optimization procedure. Instead of solving the ODE-system using a numerical scheme, the difference of the calculated reaction rate based on kinetics and interpolated measurements \( \hat{c}^{\text{int}} \) and the prior estimate of the rate \( r_j(t) \) is minimized. In this step, only the parameters of the respective kinetics \( p_j \) are varied. The differences are weighted by the uncertainty \( \sigma_j(t) \) of the prior estimate.

\[
\min_{\hat{c}} \sum_{i=1}^{N_l} \left( \frac{\hat{r}_j(t_k), \tilde{r}_j(t_k), pH(t_k), p_j, \ldots - r_j(t_k)}{\sigma_j(t_k)} \right)^2
\]

This optimization problem is usually easily solvable e.g. with a conventional Gau-Newton solver. Our experience is that the obtained parameter values do not depend on the initial guess, so the solver seems to usually converge to a unique global optimum.

This parameter estimation method is used for fast screening and validation of the kinetic expressions and additionally generates good initial values for the full scale parameter estimation which is described in the next section.

### 5. PARAMETER ESTIMATION

In the final parameter fitting step, all model parameters \( p \) are estimated at once by comparing the simulated concentration profiles \( \hat{c}(t) \) with the measured values \( c^{\text{obs}}(t) \), weighted by their measurement uncertainty \( \sigma_c \).

\[
\min_{\hat{c}} \sum_{i=1}^{N_l} \sum_{k=1}^{N_t} \left( \frac{\hat{c}_i(t_k, \hat{p}, \hat{c}_{i,0}) - c^{\text{obs}}_i(t_k)}{\sigma_{c_i}(t_k)} \right)^2
\]

The complete model with all reaction kinetics and a kinetic term for apoptosis is solved by a conventional implicit ODE-solver for stiff systems using backward differentiation formulas.

Solving the optimization problem with a gradient-based optimization routine requires many evaluations of the objective function. For each evaluation the ODE-system has to be solved. Depending on the initial values of the parameters, the obtained parameters sometimes vary, indicating the existence of local minima. Finding the best parameter values under these circumstances can be a laborious task. However, as good initial values for the model parameters

---

**Fig. 1.** Error between estimated cell-specific rates \( q \) and possible cell-specific rates of the geometrically reduced set of EM (\( SSR_R \)). The maximum allowed cosine-similarity between two EM is reduced from 1 (>40,000 EM) to 0.951 (>180 EM).
Fig. 2. Pareto-front of the GA, consisting of different selections of EM and their corresponding error between the optimized shifted concentration profile and the shifted concentration measurements are available and the model structure (reactions and kinetics) has already been validated in the prior estimation step, the optimization routine starts in the proximity of a good local optimum. To reduce the computational effort, a sensitivity analysis is performed to identify the most important parameters. All other parameters are fixed at their values from the prior parameter estimation step.

6. APPLICATION TO A CHO FED-BATCH FERMENTATION

The modelling concept is applied to fermentation data from a design of experiments (DoE) study of fed-batch fermentations with Chinese-Hamster-Ovary Cells (CHO). The DoE study with six experiments contains variations of the culture pH, the feeding strategy of liquid and gaseous substrates and the initial cell density. Measurements of 10 different media components are available with sampling times from 12 to 24 h.

6.1 Pre-reduction of the matrix $K$

The simplified metabolic network, published by Nolan and Lee (2011) was extended and used for the calculation of possible EM. Over 40,000 EM were determined. The geometrical reduction based on the cosine similarity was used to reduce the number of EM. The reduction was tested by calculating the $SSR_q$ value for different choices of the maximum allowed similarity (cf. section 6.1). Figure

Fig. 3. Calculated limitations $\tilde{r}_{j,i}$ of one macro-reaction for three different experiments (Exp. 1, Exp. 2 and Exp. 3). Two components with a negative correlation to the reaction rate and the pH-value are considered as main influences. The selected equations for the limitations are of Haldane type ($-f/\cdots$) and the modified Gaussian distribution ($\cdots$). A limitation of Monod type is considered for all substrates of the reaction, but is not shown in the graph.

1 indicates that a reduction to <900 EM is possible without a loss of physiological meaningful reactions. The corresponding maximal cosine similarity was 0.991. As not all external components of the resulting network were measured, the corresponding macro-reactions also contain substrates and products about which no quantitative information from the process is available. This means that these components are considered to be available or being produced during the process but do not have any influence on the metabolic behaviour of the organism. Limitations from unmeasured components obviously cannot be taken into account by this approach.

6.2 Selection of a set of reactions

The reaction set was obtained from the reduced matrix $K$ with the remaining EM from the pre-reduction step. A multi-objective genetic algorithm was used to generate combinations of EM which provide a sufficiently good fit to the shifted concentration measurements. The results in Figure 2 indicate that eight EM are sufficient to capture the cellular activity of the process. These eight reactions were chosen as the final set of macro-reactions for the model (defining the matrix $L$). Additionally to the $SSR$ value, the prior estimate of all eight cell-specific reaction rates $\tilde{r}(t)$ and the corresponding uncertainty $\sigma(t)$ for all six experiments were estimated in the selection step.

6.3 Choice of Kinetic Expressions

The influences of process conditions on each EM were firstly analysed statistically with the PLS method. Each time-step is regarded as one sample. The scores of three main components are indicating correlations between all measured, interpolated and normalized media concentrations, the normalized pH value and the profiles of reaction rates $\tilde{r}(t)$.

The kinetic rate expressions in this case consist of Monod-type (9), Haldane-type (10) and Gaussian-type (11) limitations. The parameter values of the possible combinations of limitations that were chosen from the detected correlations were fitted to $\tilde{r}(t)$ (see 12) and compared with respect to their sum of squared residuals and the number of parameters. Figure 3 shows the evolution of the chosen limitations of one EM over time. Figure 4 shows the calculated reaction rate $\hat{r} = f(q, pH)$ over time in comparison to the prior estimate of the reaction rate $\tilde{r}(t)$.\n
Fig. 4. Comparison of the prior estimate of a reaction rate $\tilde{r}(t)$ (dashed line) and calculated reaction rates $\hat{r} = f(q, pH)$ (solid line), based on the kinetics of one macro-reaction for three different experiments (Exp. 1, Exp. 2 and Exp. 3).
Parameter estimation

The complete model with eight reactions and additional kinetics for apoptosis contains 81 free parameters. A sensitivity analysis identified 25 of the parameters as important. The values of these parameters were fitted with a Levenberg-Marquardt optimization routine. The final fit of the model to the data is shown in Figure 5 (Model 1). For intellectual property reasons, the stoichiometry of the macro-reactions and the parameter values of the kinetics cannot be shown.

6.4 Parameter estimation

The work of Lukas Hebing was supported by BAYER TECHNOLOGY SERVICES GmbH, Leverkusen, Germany. Measurements were provided by BAYER HEALTH CARE AG, Wuppertal, Germany.

ACKNOWLEDGEMENTS

The complete model with eight reactions and additional kinetics for apoptosis contains 81 free parameters. A sensitivity analysis identified 25 of the parameters as important. The values of these parameters were fitted with a Levenberg-Marquardt optimization routine. The final fit of the model to the data is shown in Figure 5 (Model 1). For intellectual property reasons, the stoichiometry of the macro-reactions and the parameter values of the kinetics cannot be shown.

7. CONCLUSIONS

With the presented systematic modelling approach it is possible to derive a mathematical model of a complex fermentation process that includes a variety of influences. Six fed-batch runs with strongly varying process conditions where ten components are measured can be reproduced by a relatively simple metabolic CHO model consisting of only eight dominating macro-reactions. The presented procedure was helpful to identify and to quantify process influences on the metabolic activities of the cells. As the pareto front in Figure 2 indicated, it is expected that the fit to the data will not improve significantly, when more than eight EM are chosen from the matrix $K$. Figure 5 also shows the fit of a model with 16 EM and 167 parameters (Model 2) which was generated with the same method. The behaviour of the two models differ, but the overall quality-of-fit is comparable. To prevent overfitting, the model with 8 EM should be preferred over the one with 16 EM. The resulting model is suitable for applications in offline optimization, state estimation or model based design of experiments.

REFERENCES

Chen, L.Z., Chen, X.D., and Nguang, S.K. (eds.) (2006). Modelling and optimization of biotechnological processes: Artificial intelligence approaches, volume 15 of Studies in Computational Intelligence. Springer, Berlin [u.a.].

Gao, J., Gorenflo, V.M., Scharer, J.M., and Budman, H.M. (2007). Dynamic metabolic modeling for a mab bioprocess. Biotechnology progress, 23(1), 168–181.

Jungers, R.M., Zamorano, F., Blondel, V.D., Wouwer, A.V., and Bastin, G. (2011). Fast computation of minimal elementary decompositions of metabolic flux vectors. Automatica, 47(6), 1255–1259.

Leifheit, J., Heine, T., Kawohl, M., and King, R. (2007). Rechergestützte halbautomatische modellierung biotechnologischer prosesse (semiautomatic modeling of biotechnical processes). at – Automatisierungstechnik, 55(5).

Leighty, R.W. and Antoniewicz, M.R. (2011). Dynamic metabolic flux analysis (dmfa): A framework for determining fluxes at metabolic non-steady state. Metabolic Engineering, 13(6), 745–755.

McCurdy, V. (2011). Quality by design. In I. Houson (ed.), Process understanding: For scale-up and manufacture of active ingredients, 1–15. John Wiley & Sons.

Nolan, R.P. and Lee, K. (2011). Dynamic model of cho cell metabolism. Metabolic Engineering, 13(1), 108–124.

Provost, A. (2006). Metabolic Design of Dynamic Bioreaction Models. Ph.D. thesis, Universite catholique de Louvain, Louvain.

Soons, Z.I., Rocha, E.C., and Ferreira, I. (2010). Selection of elementary modes for bioprocess control. 11th IFAC Symposium on Computer Applications in Biotechnology, July 7-9, 2010, Leuven, Belgium.