Process Biochemistry xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Process Biochemistry

journal homepage: www.elsevier.com/locate/procbio

Workflow to set up substantial target-oriented mechanistic process models in bioprocess engineering

Paul Kroll[a,b], Alexandra Hofer[a], Ines V. Stelzer[a,b], Christoph Herwig[a,b,*]

[a] Research Area Biochemical Engineering, Institute of Chemical Engineering, TU Wien, Vienna, Austria
[b] Christian Doppler Laboratory for Mechanistic and Physiological Methods for Improved Bioprocesses, TU Wien, Vienna, Austria

ARTICLE INFO

Keywords:
Modeling workflow
Substantial target-oriented mechanistic models
Mechanistic links
Practical identifiability
Bioprocess engineering,

ABSTRACT

A multitude of new applications in bioprocess technology strongly depend on model-based methods as they feature prediction and control capabilities. The critical path is usually the availability of suitable models. In this work a workflow for the generation of substantial target-oriented mechanistic process models is presented. This workflow is based on backpropagation starting from a material balance for a certain target variable. Iteratively, necessary states as well as mechanistic links are included in the model using a model library reducing the computational effort. The parameters of these links are estimated using a simplex algorithm whose objective function depends on the target variable only. Practical identifiability analysis is used for the assessment of the need of further iterations and for validating the mechanistic model.

To demonstrate the workflow, a model describing a mammalian cell culture process aiming at modeling viable cell count is used as an example. The generated model satisfies the predefined requirements and is very simple, consisting of three states and seven model parameters only. The presented workflow is simple, generic, transparent, so that also applications in a regulatory environment should be possible. It also provides additional process knowledge that can be used in bioprocess development and optimization.

1. Introduction

The maximization of product formation under certain process boundary conditions is one of the major aims of industrial bioprocesses. Thereby, biopharmaceutical production processes represent a particular challenge due to their highly complex structure resulting from media, cell behavior and process parameters. In order to ensure consistent high product quality, the Quality by Design (QbD) and the Process Analytical Technology (PAT) initiative have become more and more important in recent years, where the latter suggests to monitor and to control critical process parameters (CPP) [1–4]. In this context, a variety of methods and technologies can be found in literature for process monitoring and control. A multitude of them, such as soft sensors for example [5–7], depend on mathematical models describing partial aspects of the process. Soft sensors are a powerful tool for the estimation of difficult-to-measure or not measureable variables of interest basing upon other measurements and process models. The success of such methods depends strongly on the quality of the model used, i.e. the model used has to reflect the investigated process adequately. For that purpose, validated models have to be set up. According to [8] model validation means that the model reflects the system behavior of the modeling target variable with a satisfactory range of accuracy. More precisely, a detailed description and valuation of: (i) the model in- and outputs, (ii) the model parameters and (iii) the validity space, that should largely overlap with the operational space of the process, is needed. As mentioned before, the modeling of bioprocesses is challenging due to the high number of biochemical reactions and the dynamic behavior of biological systems that have to be taken into account. On the other hand, in bioprocess engineering a good process is by definition as simple as possible and defined boundaries should be controlled as accurate as necessary which should be reflected by the process model of course. This leads to a number of simplifications reducing the quality of the model and its area of validity.

In order to generate adequate process models various workflows exist in literature. Most of these workflows are designed for data-driven models. The description of a very detailed workflow can be found in the handbook of good modeling practice (GMP) [9]. Depending on the aim and the application of the model, different methodologies of validation exist. As data-driven models describe an input-output relation they can be validated using statistical methods such as cross-validations or

Abbreviations: CHO, Chinese hamster ovary; DCC, dead cell count; DoE, design of experiment; NRMSE, normalized root-mean-square error; VCC, viable cell count; ANN, artificial neural network; Glu, glutamine; Glc, glucose; Lac, lactate; NH_4, ammonia; LS, limiting substrate; ODE, ordinary differential equations; QbD, quality by design

* Corresponding author at: TU Wien, Institute of Chemical Engineering, Research Area Biochemical Engineering, Gumpendorfer Straße 1a, 1060 Vienna, Austria.
E-mail address: christoph.herwig@tuwien.ac.at (C. Herwig).

http://dx.doi.org/10.1016/j.procbio.2017.07.017
Received 24 February 2017; Received in revised form 3 July 2017; Accepted 22 July 2017

analysis of variance (ANOVA). In order to model bioprocesses, mechanistic models, which are based on interpretable equations, become more and more popular [10]. They describe the sum of relevant individual occurring processes or mechanisms in a system [11]. The validation of mechanistic models includes the validation of the model structure and the model parameters. The enormous variety of possible relevant physiological variables and parameters as well as their interactions inevitably would result in a limitation of resources, which are process knowledge, information and computation power. Therefore, modelers have to restrict themselves by a subjective choice of a model structure. Single publications are trying to limit this freedom by providing a model library for certain applications and systems [12]. Generally, the model structure used for describing a process is usually predefined in a certain way. Model libraries consisting of several process models are often used by modelers leading to two main limitations: (i) insufficient computational power to estimate the best model structure based on systematic investigation of all possible models being available in the model library using a grid search, (ii) limited degrees of freedom due to a limited model library. These two limitations are negatively proportional to each other. This means that the grid search becomes increasingly ineffective with a larger model library. Within the framework of data to information to knowledge, a variety of methods can be used to estimate potential correlations in an unknown system [13–15] reducing the number of potential model structures. Herold and King [16] shown for example how process events can be used within a bootstrap analysis in order to setup hypotheses of overall model structures. Nevertheless, the challenge of restricting variables and parameters still exists. Furthermore, model parameters have to be validated. For complex models containing many parameters, this is a very complex and computationally expensive procedure resulting from their cross-correlations.

This paper followed the hypothesis that it is more difficult to reduce complex models than setting up substantial models which are designed exactly for the modeling target. The aim is to present a new workflow for the generation of substantial, transparent and comprehensible target-oriented mechanistic process models. These models are restricted in their complexity due to the modeling objectives and are set up using objective criteria. A special feature of the developed workflow is the possibility to set up bioprocess models without predefining model states and equations focusing on the target variable only. Furthermore, instead of using a model library containing complete process models, a smaller model library consisting of so-called “mechanistic links” is used. Mechanistic links represent single equation terms describing a single mechanism like a kinetic or stoichiometric relationship between indirectly determinable physiological variables and directly determinable system states. In contrast to data-driven artificial neural networks (ANN), where the inputs for single neurons in the hidden layer before fitting are unknown, the inputs for mechanistic links can be calculated from existing data. A significant challenge in the development of ANNs is the estimation of the big amount of model parameters. Therefore, different algorithms such as the backpropagation and Marquardt algorithm have been developed [17,18]. In this work, the basic idea of the backpropagation algorithm is applied to mechanistic models setting up the model iteratively starting inversely from a material balance for the target variable. This means a stepwise parameter fit of single mechanistic links in order to prevent cross correlations between model parameters of different investigated mechanistic links. Potential mechanistic links are chosen from the model library due to objective criteria based on an identifiability analysis. This allows the step-by-step setup and validation of single mechanistic links leading to a model consisting of necessary state variables and parameters only. Thus, the setup und validation of a whole process model consisting of predefined states is reduced to the setup and the validation of simple single mechanistic links. For estimating the parameters of these links a standard simplex algorithm with objective function, only depending on the target variable, is used. This reduces the number of considered parameters drastically in every iteration step. Thus, the basic hypothesis of this paper can be formulated as follows: “The necessary condition for a valid process model is the validity of the single mechanistic links.” This idea bases on the idea of the incremental identification of kinetic models by Brendel et al. [19].

In the following the developed workflow is described from the theoretical point of view step-by-step. After that we illustrate the workflow by an example, setting up an adequate model for an industrial mammalian cell culture process using experimental data. As target variable of the model the viable cell count (VCC) is considered. The results show good agreement of experimental data with the model.

2. Material and methods

2.1. Investigated bioprocess

In order to test the modeling workflow using experimental data a fed-batch process was performed. An industrial CHO cell line was cultivated in chemically defined medium provided by the industrial partner. The fed-batch process was carried out in a 3.6 L bioreactor system (Labfors 5, Infors, Switzerland) with an initial working volume of 2 L. Closed-loop controlled process parameters were the temperature (37 °C), the pH-value (6.8), the dissolved oxygen tension (40%) and the partial pressure of carbon dioxide (125 mbar). The experiment was performed using three different feeds, namely a glucose feed, a glutamine feed and a feed with limiting components. The glucose feed and the glutamine feed were controlled in a way that a minimum supply was ensured in accordance with a specific glucose uptake rate of 0.025 mM/109cells/h and a specific glutamine uptake rate of 0.01 mM/109cells/h respectively. Under these conditions glycolysis is a limiting flux with respect to the tricarboxylic acid cycle [20]. Samples were taken every 12 h. Viable cell count (VCC) and dead cell count (DCC) were measured by an automated image analyzer (Cedex HiRes Analyzer, Roche, Mannheim, Germany). Both measurements were performed in triplicates. As lysis is a well-known phenomenon in cell culture processes, it was taken into account for the determination of dead cells according to Klein et al. [21]. The concentration of glucose (Glc), glutamine (Glu), lactate (Lac) and ammonia (NH_4) were measured by an enzymatic analyzer (Cedex BioHT, Roche, Mannheim, Germany). Limiting substrate (LS) concentration was measured by an HPLC (Thermo Fisher Scientific, United States).

2.2. Calculation of specific rates

All specific rates appearing in the investigated mechanistic links were calculated by using a simple material balance equation for the corresponding state. Basic assumption for the calculation is that the specific rate is constant between two measurement points. Solving the material balance equation using an ordinary differential equation (ODE) solver (MATLAB 2015b: *ode23*) as well as measurements generated by the experiment, the specific rates were calculated checking whether the material balance closes in each measured point. The computation was performed within a loop for each measurement interval. Firstly, the specific growth and death rate have to be determined in order to provide a time resolved active biomass, which is necessary for the calculation of the additional specific rates as input. Thereafter, the other specific rates are calculated according to the material balance. The advantage of this approach is that the specific rates can be used directly for the calculation of the states. Because each specific rate has a defined time window, different sampling frequencies can be easily compensated by this approach. In addition the assumption of a linear trend between two measurement points is not necessary. This approach without smoothing can result in noisy specific rates.