Assessment Techniques for Convergence and Degree of Freedom in Biologics Modeling

Corresponging author: taguchi.tomoyuki@chiyodacorp.com

Abstract

To express the cultivation cell line of recombinant proteins in mathematical models, it is necessary to represent cellular functions such as proliferation and death rate, translation systems, and metabolic systems according to the production purpose. In this process, assuming the mathematical models and validating them through experimental results is extremely crucial. The standardization of these trial-and-error approaches and the minimization of the number of experiments are essential technologies for achieving the rapid establishment of a manufacturing system. In this paper, an antibody producing cell line is assumed, utilizing CHO cells and employing a simplified cell biologics model where protein production is connected to energy production as a mock-up. Through the estimation of system parameters from pseudo-experimental results obtained from the biologics model, the existing parameter estimation methods and a newly proposed approach considering the degrees of freedom are compared. Finally, the paper provides comparative results and a discussion of the characteristics of both methods to address the issue of system parameter estimation in biologics modeling.

**Keywords**: “Biologics”, “Modeling and simulation”, “Optimization”, “Process design”, “Parameter identification”, and “Degree of Freedom”.

* 1. Background and purpose

The use of cellular platforms for producing various proteins and substances becomes more popular, especially in the field of biopharmaceutical development (Walsh, 2018). Inside cells, a wide variety of enzymatic reactions and material transport processes take place, modeling the entire system of cell and boundary transportation is extremely challenging. It is crucial to hypothesize models that maintain precision for analyzing the observed phenomena of interest and quickly acquire and analyze experimental data to validate these models (Kitano, 2002).

The requirement for modeling-based system assessment methods is escalated to establish the pharmaceutical manufacturing systems keeping high quality within a short period (Kemmer et. al., 2023). Hence, it is necessary to determine the required minimum experimental data and to identify the parameters in realistic values with the satisfaction of degrees of freedom.

In order to assist in the design of manufacturing systems and the operational procedures, it is advantageous to consider a set of differential-algebraic equations involving critical parameters. This system of equations typically includes simultaneous equations for material transport and Monod-type equations, which depend on state changes and are highly nonlinear (Badr, 2023 and Monteiro, 2023). In other significant aspects, glycolysis and metabolic pathways are crucial for supporting translation work. In recent years, the integration of material and energy production have been considered. The effect of detailed network linking among translation, glycolysis and metabolism further increases the nonlinearity. Therefore, the construction of a model to assess essential cell characteristics from minimal experiments is crucial. (Ramos, 2020).

Parameter identification for such models often falls into the problem of simultaneously determining state variables and parameters multimodality characteristics. Formulating hypothesis-based models and quantitatively verifying them through experiments can significantly contribute to reducing development timelines. Therefore, it is crucial to simplify the above-mentioned nonlinear systems into a linearized form that can be evaluated efficiently.

* 1. Biologics modeling
		1. Simplified biologics model

As a mock-up simulation model of antibody production cell line biologics, the following three basic mathematical models are integrated.

* Cell division and death rate linked to the culture media component change
* Oxygen concentration change in a cell regarding the balance between the respiration and energy consumption rate
* Concentration changes of oxygen, glucose and lactose in cultivation media according to the balance between supply and demand amount

The material correlation is shown in Figure 1.



Figure 1. Conceptual schematic of simplified biologics model for antibody production

Table 1. Symbol definitions for simplified biologics model



* + 1. Constitutive equations, variables and nomenclatures

The simplified biologics model is composed by the equations and variables shown in Table 1. Each symbol is stated with the corresponding nomenclature.

System equations are categorized in three part which are the cell properties, culture media condition and cell internal condition. It is also possible to modify the system modeling to be more complex by adding items to evaluate the behavior of interest. For example, when incorporating more detailed enzyme reaction systems of amino acids or proteins, it is sufficient to add model equations that represent their behavior in the culture medium and inside the cells. The purpose of this study is to identify problems in estimating system parameters from experimental data and its quality. Additionally, as the purpose is to propose a new method for system identification, a preference is given to utilizing as simple a model as possible as minimum.

* + - 1. Typical simulation results

Some measured variables are shown in Figure 2. These are depending on the set of initial conditions and system parameters.



Figure 2. Model outputs used as pseudo-experimental data.

* 1. Methodology

In the conventional approach, the minimization of the squared difference between measured and calculated variables is implemented. However, some problems exist.

* The combination of system parameters tends to generate the multimodality.
* The computational cost tends to increase.
* How much system parameters are effective in the hypothesis model.
* How much experimental data are required to determine the system parameters.
* Experimental data includes some types of error such as systematic, random, human and environmental error, which affects the parameter estimation.

Although the biologics model in this study is assumed as a simplified mockup, the fourteen system parameters are incorporated. To verify the conventional system parameter estimation method, the pseudo-experimental results are obtained from the biologics model with the 10% error, followed that the system parameter is reversely estimated. The simulation results are found in Figure 3. The curve features are well traced however the system parameters are different from the original set values due to the effect of multimodality and experimental errors.

To comprehend the characteristics of the system equations, a randomized combination of experimental conditions and system parameters is employed in the case study to generate a dataset of experimental results in integral form. Using this dataset, the verification of the degrees of freedom and the transformation of the equations into a more streamlined form for estimating system parameters are simplified.

Translate the original non-linear differential equations to the regression form using the integral values based on random set of experimental conditions ‘$x\_{i}$’ and system parameters ‘$π\_{j}$’.

 $y\~ζ(δx\_{i},δπ\_{j},\left(δx\_{i}\right)^{2},\left(δπ\_{j}\right)^{2},δx\_{i}δπ\_{j},a, b)$ (1)

$δx\_{i}=x\_{i}-\tilde{x\_{i}}$ (2)

$δπ\_{j}=π\_{j}-\tilde{π\_{j}}$ (3)

Symbols ‘a’ and ‘b’ represent the corrected values to regress the experimental results in a linearized form. Symbol ‘$\tilde{x\_{i} }$’ and ‘$\tilde{π\_{j}}$’ serve as representative values to express the typical performance of biologics cell line as a center point in the state space of potential cell activation. The equation (1) transforms the original nonlinear differential equation system into a linearized algebraic system. This purpose is to modify the landscape of the parameter space with multimodality in equation (1) to obtain an optimal solution near the global optimum in the broader parameter space of transformed model.



Figure 3. Fitting results for pseudo-experimental data incorporating random errors

* 1. Results and discussion

Table 2 shows the comparison of estimated results of system parameters. ‘SLVR’ means the conventional system identification based on the minimization of squared difference. On contrary that ‘LPE’ means the linearized parameter estimation.

The following aspects are found.

* ‘SLVR’ results show the different values and large standard deviations. This is the natural features due to the existence of multimodality and strong fitness to the experimental errors.
* ‘LPE’ results show the different values, however the standard deviations are significantly lower. This is presumed to be due to the evaluation of a linearly approximated solution space, which reduces the sensitivity to assigned experimental errors (10%) in parameter estimation.

To harmonize the characteristics of both estimation methods, the loss function of proposed method is incorporated into that of conventional system identification method, then the parameter optimization is carried out again. Results in the hybridization are presented in Table 3. According to the implementation of hybridization methods, the parameter estimation results become closer to original set values, and the standard deviations are found in mediate.

Table 2. System parameter estimation results incorporating random errors



Table 3. System parameter estimation results of hybridization constraints



* 1. Conclusion

To conduct cycles of hypothesis and experimental verification, and to construct a model satisfying the degrees of freedom of the target system, robust evaluation of system parameters is essential. In this paper, a method is proposed for estimating the system parameters of the target system, where the prior learning information of the hypothesis model is replaced with an integral solution featuring algebraic input-output relationships. This approach not only enhances the robustness of the estimation results but also facilitates an understanding of the experimental data volume required to satisfy the degrees of freedom of the system parameters. Furthermore, through hybrid estimation with the original nonlinear differential equation system, an improvement in the accuracy of parameter estimation is also confirmed. It becomes possible to quickly obtain a system model for engineering applications such as the performance evaluation of recombinant protein-producing strains, considerations for scaling up cultivation systems based on these strains, and prediction and monitoring of operating conditions adjusted to manufacturing scales.

References

G. Walsh, 2018. Biopharmaceutical benchmarks 2018. *Nature Biotechnology*, 36(12), 1136-1145.

H. Kitano, 2002. Systems Biology: A Brief Overview. *Science*, 295(5560), 1662-1664.

A. Kemmer, et. al., 2022. Nonlinear state estimation as tool for online monitoring and adaptive feed in high throughput cultivations. *Biotechnology and Bioengineering*.

S. Badr, et. al., 2023. Hybrid modeling and data-driven parameterization of monoclonal antibody cultivation processes: Shifts in cell metabolic behavior. In *Computer Aided Chemical Engineering* (Vol. 52, pp. 985-990), Elsevier.

M. Monteiro, & C. Kontoravdi, 2023. Hybrid dynamic model of monoclonal antibody production using CHO cells. In *Computer Aided Chemical Engineering* (Vol. 52, pp. 375-380), Elsevier.

J. R. Ramos, A.G. Rath, Y. Genzai, V. Sandig & U. Reichi, 2020. A dynamic model linking cell growth to intracellular metabolism and extracellular by-product accumulation. *Biotechnology and Bioengineering*, 117(5), 1533-1553.