**Robust design of experiments for model selection using inverse modeling.**

Moritz Schulze1,2,3, René Schenkendorf1,2,\*

*1 Institute of Energy and Process Systems Engineering (InES), TU Braunschweig, Braunschweig, Germany; Center of Pharmaceutical Engineering (PVZ), TU Braunschweig, Germany; 3 International Max Planck Research School (IMPRS) for Advanced Methods in Process and System Engineering, Magdeburg, Germany*

*\*Corresponding author: r.schenkendorf@tu-braunschweig.de*

**Highlights**

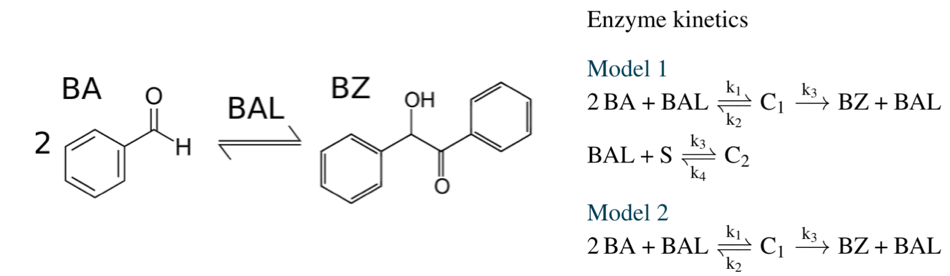
* One-step model selection under consideration of parameter uncertainties.
* Inverse modeling for an efficient solution of the DoE optimization problem.
* Reliable model identification only when uncertainties are included.

**1. Introduction**

In terms of quality by design (QbD), the development and production of active pharmaceutical ingredients benefit from a detailed system understanding. To this end, mathematical models might be a useful tool. A crucial task in modeling of (bio)chemical reaction networks is the identification of the most suitable model candidate from a set of various model hypotheses. One possible strategy for model selection is the design of experiments (DoE) approach, where the challenge of model selection is formulated as an optimization problem. An additional consideration of model parameter uncertainties leads to robust predictions of optimal experimental conditions that are expected to provide highly informative data for a reliable model selection.

**2. Methods**

For an effective solution of the robust model selection problem, we aim at implementing an inverse modeling technique by utilizing the differential flatness concept [2]. With the flatness concept, we avoid simplifying assumptions as linearization and the need for solving differential equations numerically. We consider a biocatalytic step from a carboligation reaction system that forms an essential precursor in the synthesis of natural products and pharmaceuticals [2], see Figure 1.

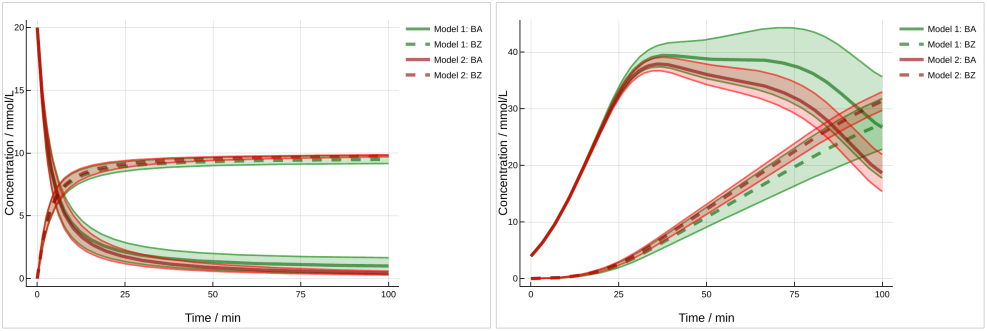
**

**Figure 1.** Model candidates of benzaldehyde lyase (BAL)-catalysed reaction network (BA: benzaldehyde, BZ: benzoine).

The two model candidates differ in the following way. Model candidate 1 suffers from a loss reaction where a second substrate (S) is inhibiting. We identify experimental conditions that maximize the differences of the model responses including the uncertain model parameters. The required statistical moments for robustifying the optimization problem are calculated with the Point Estimate Method (PEM) that guarantees low computational cost and sufficient accuracy [3].

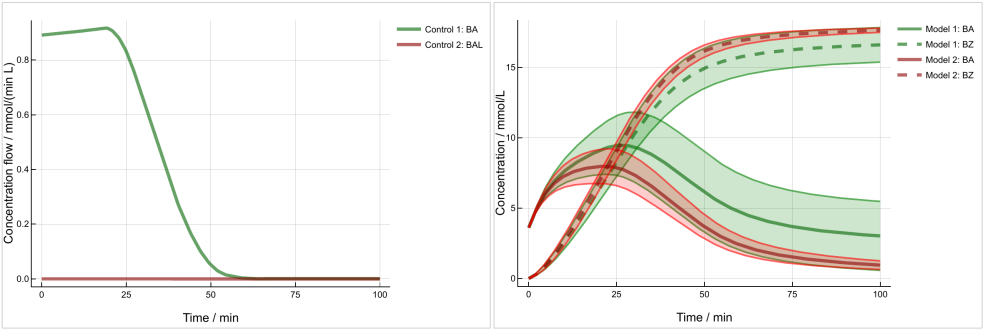
**3. Results and discussion**

In Figure 2, the simulation results of a non-optimized experimental setting are shown on the left. The normalized overlap of the two models’ system states is 1.0 and the expected values, i.e., the thick lines in the middle of the confidence intervals indicate that the models are not distinguishable.



**Figure 2.** Prior to optimization (left) and after non-robust optimization (right).

First, non-robust optimization was performed to increase the difference in the model responses. The results are shown in Figure 2 on the right-hand side. It is visible that the expected values of the system states are driven apart. However, under consideration of the model parameter uncertainties, the normalized overlap is 1.07 and therefore worse than in the non-optimized case. Reliable identification of a particular model is problematic.



**Figure 3.** Robust optimized results: controls (left) and measured states (right).

Next, a robust optimization was conducted, where its results are shown in Figure 3. As before, the expected values are driven apart, but additionally, the normalized overlap could be reduced to 0.76. Here, a one-step model selection after running the experiment in the laboratory with the controls shown on the left-hand side in Figure 3 is more likely.

**4. Conclusions**

We successfully demonstrated how inverse modeling based on flatness could be used for a model selection problem. Furthermore, we showed that the consideration of parameter uncertainties is essential for reliable model identification.

**References**

1. M. Fliess, J. Lévine, P. Martin, P. Rouchon, CR Acad Sci Paris (1992), 619–624.
2. F. Hildebrand et al., Biotechnol. Bioeng. 96 (2007), 835-843.
3. R. Schenkendorf et al., Processes 6 (2018).