Electronic Supplementary Material

Decision Support for the Development, Simulation and Optimization of Dynamic Process Models

Norbert Asprion (✉)[1], Roger Böttcher[1], Jan Schwientek[2], Johannes Höller[2], Patrick Schwartz[2], Charlie Vanaret[2], Michael Bortz[2]

1 BASF SE, 67063 Ludwigshafen, Germany
2 Fraunhofer ITWM, 67663 Kaiserslautern, Germany

E-mail: norbert.asprion@basf.com

Appendix A Dynamic simulation in CHEMADIS

In contrast to a steady-state simulation, where a fully defined equation system $0 = h(x)$ has to be solved, in dynamic simulation a differential algebraic system (DAE) is solved.

$$\begin{aligned} \dot{y} &= f(y, x, p, t) \\ 0 &= g(y, x, p, t) \end{aligned} \tag{A1}$$

Herein, y denotes the vector of differential, x algebraic states, f the differential and g the algebraic equations, while p is a parameter vector, which could define controls, design or model parameters. $\dot{y}$ is the time derivative of the differential state. For solving this DAE system in CHEMADIS a single shooting method is used. Hereby different integration methods can be selected and a heuristic time step control can be adjusted to the problem. In the following, first the solution method and then the method to estimate the sensitivities/derivatives is briefly described.

A.1 Temporal integration: procedures and step sizes

CHEMADIS normally uses a 2nd order integration procedure, the trapezoidal rule, for solving the DAE system. This means, that besides the 1st derivative also the 2nd derivative for integration is taken into account. Where this procedure is not applicable, namely at the time zero and after jumps in process variables triggered by events, a 1st order procedure, the implicit Euler method, is used.
It is also possible to select integration solely based on an implicit EULER procedure. Additionally, two single-value-diagonal-implicit Runge-Kutta-methods (SVDIRK 3rd and 4th order) and BDF (Backward Differentiation Formula up to order 6) are implemented and can be chosen.
There is generally only one relatively simple time step size control: a user-defined ‘maximum time step size’ is used for the calculation until a time step does not converge. Then the step size is reduced until the calculation for a time step converges. If convergence for time steps below the lower limit of 10^{-12} seconds could not be achieved, an attempt is made to restart columns in the simulation model. Afterwards the time step size is again gradually increased up to the original value. (One exception is the procedure SVDIRK 4/5, 4th order with a 5th order estimation, where an adaptive step size control is implemented. But there the adaptive time steps will never exceed the given, user-defined 'maximum time step size').
To find a certain event as exactly as possible, the time step size is adjusted by interpolation.
In general, the 'maximum time step size' determines the maximum step size for the integration and is never exceeded. At the beginning and after each event (i.e. a condition in the recipe - a given condition or an implicit such as a phase split or similar) is always started with an ‘initial time step size’ first,to integrate, i.e. in general and in a reasonable way with a smaller step size. If no ‘initial time step size’ is given by the user, it is set to ‘initial time step size’ = min (1 sec, ‘maximum time step size’). If the iteration counter for the solution of the subordinate nonlinear system of equations falls below a user-defined threshold limit, called MINIT, the step size is doubled - until the ‘maximum time step size’ is reached.

A.2 Sensitivities/Derivatives in CHEMADIS

As mentioned already different integrators (Implicit Euler, Trapezoidal, single-value-diagonal-implicit Runge-Kutta 3. and 4. order or BDF (Backwards-Differentiation-Formula up to order 6)) can be used for solving the DAE. This leads from equation system (A1) to a modified right hand side – dependent on the chosen integration method – and an in general non-linear algebraic equation system

$$\begin{aligned} 0 &= \tilde{f}(y, x, p, t) \\ 0 &= g(y, x, p, t) \end{aligned} \tag{A2}$$

which can be solved with Newton's method to estimate the updates $\Delta y, \Delta x$ for the states in each iteration step *for one time point*

$$\begin{pmatrix} -\tilde{f}(y, x, p, t) \\ -g(y, x, p, t) \end{pmatrix} = \tilde{J} \begin{pmatrix} \Delta y \\ \Delta x \end{pmatrix} \text{ with } \tilde{J} = \begin{pmatrix} J_y^{\tilde{f}} & J_x^{\tilde{f}} \\ J_y^{g} & J_x^{g} \end{pmatrix} \tag{A3}$$

With J̃ the Jacobian of the modified equation system with respect to y and x. To solve the DAE system for fixed parameters p only first derivatives with respect to y and x are required. But in optimization problems like for example optimal control or optimal design of experiments in general also first derivatives with respect to the parameters are required. One possibility is to apply the chain rule to the DAE system (1), which leads to a second DAE system, which has to be solved. Another possibility is to extend the equation system to include the parameters and to the extract this information from the solved equation system (similar as in the steady-state case, see supporting information of [18], cf. reference in main publication). In CHEMADIS this second alternative is pursued. This is realized by deriving the integration methods and results in

$$\begin{pmatrix} -\tilde{f}_p(y, x, p, t) \\ -g_p(y, x, p, t) \end{pmatrix} = \tilde{J} \begin{pmatrix} y_p \\ x_p \end{pmatrix} \tag{A4}$$

For solving these equations with Newton's method, the parameters are included in the algebraic states *x* and equations which assign values to these parameters are added to the equation system.

Special attention has to be made, when a schedule or recipe triggers changes in the DAE system. These so-called event-triggered changes might result in jumps in the derivatives as discussed by Galán et al. [22] (cf. reference in main publication).

Appendix B Further Results of the MVA example

In this Appendix further results of the model adjustment to batch distillation measurements of Nad and Spiegel (Reference please see main document) are shown: These include results for the liquid composition in the still in Fig. B1, temperatures at top and bottom of the column in Fig. B2, the vapor flow ate the top of the column and the reflux flow in Fig B3 and finally the compositions and amounts of the different fraction in Fig. B4.

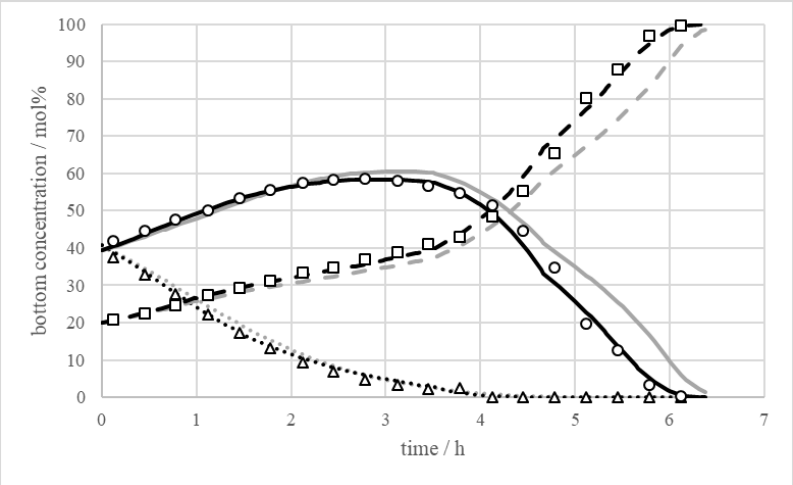

Fig. B1: Comparison of the dynamic simulation based on the initial (gray lines) and the optimized values (black lines) with the measured liquid concentration in the still (symbols): ⋯, △ Cyclohexane, —, ○ n-Heptane, – –, □ Toluene

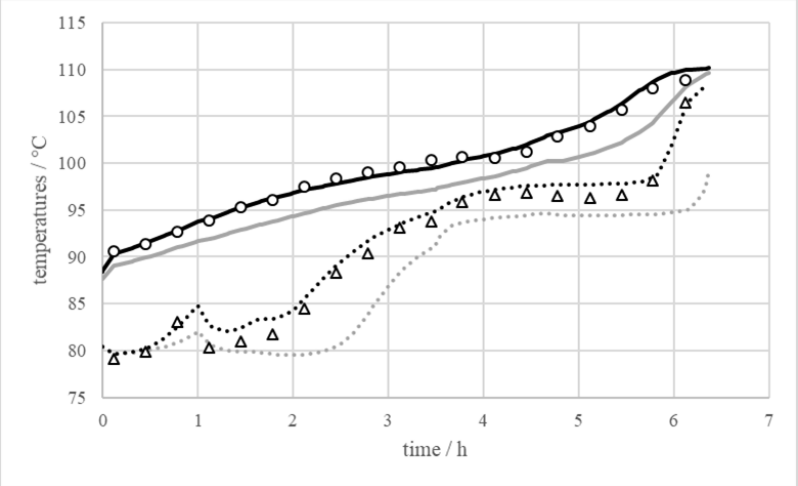

Fig. B2: Comparison of the dynamic simulation based on the initial (gray lines) and the optimized values (black lines) with the measured temperatures in the column (symbols): ⋯, △ top of the column, —, ○ bottom of the column

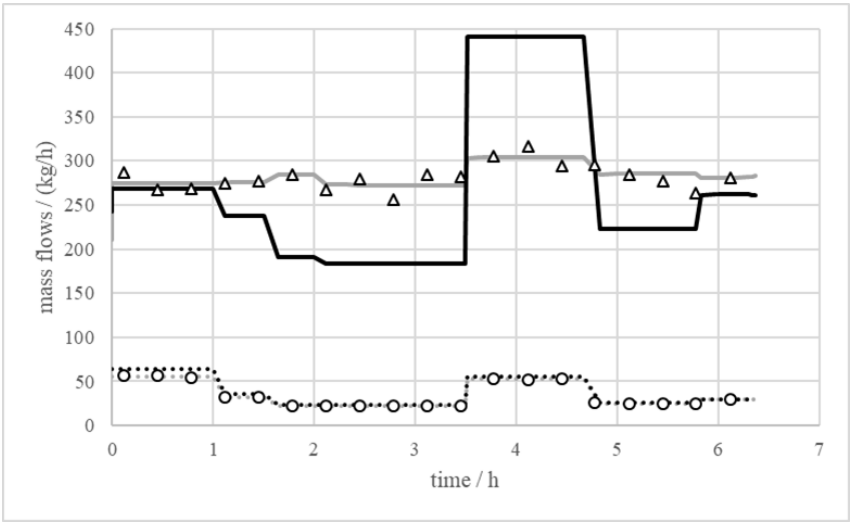

Fig. B3: Comparison of the dynamic simulation based on the initial (gray lines) and the optimized values (black lines) with the measured mass flows (symbols): —, △ vapor flow at the top of the column, ⋯,○ liquid flow of the distillate

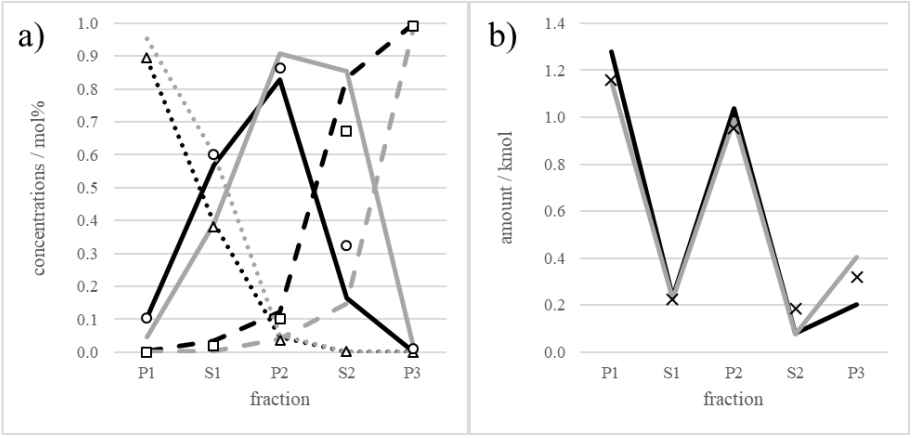

Fig. B4: Comparison of the dynamic simulation based on the initial (gray lines) and the optimized values (black lines) with the single time measurements (symbols) for the concentrations and amounts of the fractions (P1,S1, P2, S2) and the final product in the still (P3): a): ⋯, △ Cyclohexane, —, ○ n-Heptane, – –, □ Toluene, b): —, × amount

In Fig. B5 for the vapor-liquid equilibrium (VLE) data of the systems Cyclohexane + n-Heptane and n-Heptane + Toluene a comparison is shown between the experimental data at 1 bar and the prediction of the original thermodynamic model of Nad and Spiegel as well as of the one using an adjusted vapor pressure for n-Heptane. As can be seen the vapor pressure coefficient estimated from the batch distillation experiment significantly improves the agreement with the experimental data. This is just to show the reliability of the optimization. Usually, as mentioned before, the thermodynamic model is recommended to be parametrized based on the VLE data and then just used in simulation. This example is only a showcase for situations where reliable VLE data is not available or obtainable with reasonable effort as discussed in the introduction.

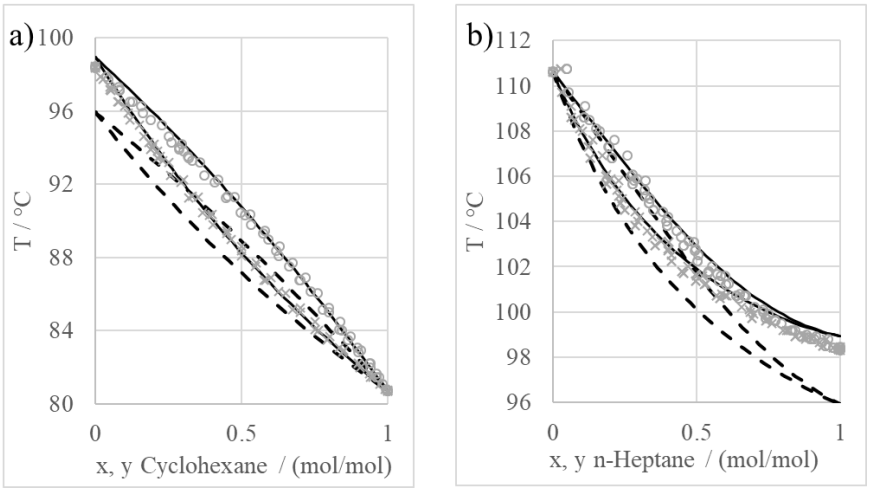

Fig. B5: Vapor-liquid equilibrium data of the systems a) Cyclohexane + n-Heptane and b) n-Heptane + Toluene at 1 bar: Comparison of experimental data (symbols): ○ vapor, × liquid with the original model of Nad and Spiegel (dashed lines) and the model with adjusted n-Heptane vapor pressure coefficient (full lines)

Appendix C Semi-batch Williams-Otto reactor

The model equations of the semi-batch Williams-Otto reactor are the following

$$\dot{x}_A = -\frac{x_A \cdot F_B}{1000 \cdot V_R} - k_1 \xi_1 x_A x_B$$

$$\dot{x}_B = \frac{(1 - x_B) \cdot F_B}{1000 \cdot V_R} - k_1 \xi_1 x_A x_B - k_2 \xi_2 x_B x_C$$

$$\dot{x}_C = -\frac{x_C \cdot F_B}{1000 \cdot V_R} + k_7 \xi_1 x_A x_B - k_3 \xi_2 x_B x_C - k_6 \xi_3 x_C x_P$$

$$\dot{x}_P = -\frac{x_P \cdot F_B}{1000 \cdot V_R} + k_2 \xi_2 x_B x_C - k_4 \xi_3 x_C x_P$$

$$\dot{x}_E = -\frac{x_E \cdot F_B}{1000 \cdot V_R} + k_3 \xi_2 x_B x_C$$

$$\dot{x}_G = -\frac{x_G \cdot F_B}{1000 \cdot V_R} + k_5 \xi_3 x_C x_P$$

$$\dot{T}_R = -\frac{(T_{in} - T_R) \cdot F_B}{1000 \cdot V_R} + k_8 \xi_1 x_A x_B + k_9 \xi_2 x_B x_C + k_{10} \xi_3 x_C x_P - l_1 \cdot (T_R - T_W)$$

$$\dot{V}_R = \frac{F_B}{1000}$$

with

$$\xi_1 = \exp(-1000 \cdot b_2 / (T_R / °C + 273.15))$$

$$\xi_2 = \exp(-8333.3 / (T_R / °C + 273.15))$$

$$\xi_3 = \exp(-11111 / (T_R / °C + 273.15))$$

The optimization is subject to the following path constraints

$$60 \leq T_R/°C \leq 90$$

$$V_R/\mathrm{m}^3 \leq 5$$

and bounds of the control variables

$$0 \leq F_B/(\frac{\mathrm{m}^3}{\mathrm{s}}) \leq 5.784$$

$$20 \leq T_W/°C \leq 100$$

The initial values (subscript *0* for time dependent variables) are

$$T_{in} = 35.°C, \quad T_{R,0} = 65\,°C, \quad x_{A,0} = 1, \quad V_{R,0} = 2\ \mathrm{m}^3$$

and for the model parameters the following values are used

$$k_1 = 1659900., \quad k_2 = 721170000., \quad k_3 = 1442340000.$$

$$k_4 = 1.33725 \cdot 10^{12}, \quad k_5 = 4.01175 \cdot 10^{12}, \quad k_6 = 2.6745 \cdot 10^{12}$$

$$k_7 = 3319800., \quad k_8 = 104656218.9, \quad k_9 = 27285184270.$$

$$k_{10} = 1.4465568 \cdot 10^{14}, \quad k_{11} = 1.8160503 \cdot 10^{11}, \quad l_1 = 2.4345469 \cdot 10^4$$

$$b_2 = 6.6667$$

The competing objectives are to maximize the amount of P and to maximize the amount of E

$$m_P = x_P V_R$$

$$m_E = x_E V_R$$

Appendix D Lysine semi-batch reactor

The model equations for the lysine semi-batch reactor are the following

$$\dot{X} = \mu \cdot X$$

$$\dot{S} = -\sigma \cdot X + u \cdot C_{S,F}$$

$$\dot{P} = \pi \cdot X$$

$$\dot{V} = u$$

Herein are X, S, P the masses of biomass, substrate and lysine (product) and V is the volume of the mixture in the reactor. $C_{S,F}$ is the concentration of the substrate in the feed and u is the flow rate of the feed, which is the control variable. The specific growth μ, substrate consumption σ and the production rate π are given by

$$\mu = \mu_p \frac{S}{V}$$

$$\sigma = \frac{\mu}{\sigma_p}$$

$$\pi = -\pi_{p,1} \cdot \mu^2 + \pi_{p,2} \cdot \mu$$

The following constraints for the optimization are given

$$5\,\mathrm{l} \leq V \leq 20\,\mathrm{l}$$

$$P \geq 400\,\mathrm{g}$$

$$(V - V_0) C_{S,F} \geq 20\,\mathrm{g}$$

$$20\,\mathrm{h} \leq t_f \leq 40\,\mathrm{h}$$

and bounds for the control variable u are specified

$$0\,\mathrm{l/h} \leq u \leq 2\,\mathrm{l/h}$$

The initial values (subscript *0* for time dependent variables) are

$$X_0 = 0.1\,\mathrm{g}, \qquad P_0 = 0\,\mathrm{g}, \qquad S_0 = V_0 C_{S,F} = 14\,\mathrm{g}, \qquad V_0 = 5\,\mathrm{l}, \qquad C_{S,F} = 2.8\ \mathrm{g/l}$$

and the model parameters have the following values

$$\mu_p = 0.125\ \mathrm{l/(gh)}, \qquad \sigma_p = 0.135\ \mathrm{1/h}$$

$$\pi_{p,1} = 134, \qquad \pi_{p,1} = 384\,\mathrm{h}$$

The objectives are the reactor productivity J_1 and the lysine yield J_2

$$J_1 = P/t_f$$

$$J_2 = \frac{P}{(V - V_0) C_{S,F}}$$