

Semi-empirical Process Modelling with Integration of Commercial Modelling Tools

G. Schopfer^a, O. Kahrs^a, W. Marquardt^{a*}, M. Warncke^b, T. Mrziglod^b and A. Schuppert^b

^a Lehrstuhl für Prozesstechnik, RWTH Aachen,

Turmstr. 46, 52056 Aachen, Germany

^b Bayer Technology Services GmbH,

51368 Leverkusen, Germany

Abstract

The paper presents an approach for semi-empirical (hybrid) modelling of partially unknown processes. Using the integration platform CHEOPS it is possible to integrate process unit models from different modelling tools. The method is applied to the case of an ethylene glycol process model, incorporating known unit models from general purpose modelling tools and a structured hybrid model from the semi-empirical modelling tool HybridTool.

Keywords: semi-empirical modelling, hybrid modelling, chemical process modelling, parameter identification, model reuse

1. Introduction

Semi-empirical modelling, often also referred to as hybrid modelling, aims at providing models for processes that are not fully captured by first principles. Opposed to purely data-driven approaches, hybrid modelling focuses on an optimal model design with respect to accuracy, interpretability and extrapolability by combining data-driven (black-box) models with first-principles (white-box) models (Thompson, et al., 1994). In the so called structured hybrid modelling approach white- and black-box models are combined in a way that allows to reduce the dimension of the data base and achieve extrapolability in a predictable region (Schuppert, 2000).

In this contribution we deal with hybrid modelling of whole production plants, i.e. process models consisting of several interconnected process unit models for the different process steps (e.g. mixing, reaction, separation). As an example we consider a part of an ethylene glycol production process shown in Figure 1. The educts ethylene oxide and water are fed to a mixer that is connected to a reactor where the educts react to ethylene glycol. The remaining educts in the reactor outlet stream are separated from the product in the flash and fed back in a recycle stream to the mixer. In our scenario we assume that the kinetics of the ethylene glycol reaction are unknown. Unknown kinetics are typical in polymerization and fermentation processes, as well as in catalysed

* Correspondence should be addressed to W. Marquardt, marquardt@lpt.rwth-aachen.de

reactions, where the reaction kinetics may be subject to unknown influences such as the contamination of the catalyst during plant operation. It is furthermore assumed that a direct identification of the reaction kinetics is not possible due to lack of the necessary data (i.e. the dependencies of the reaction rates on the concentrations).

The main aim of our approach is to enable the identification of the unknown reaction kinetics by simply using the measurement data available from the operation of the plant, as these are indicated in Figure 1. If one is only interested in a model that describes the relation between the measured inputs and outputs, a purely empirical model can be fitted to this data. However, such a model does not allow any insight into unmeasured process variables. Therefore it is necessary to identify a hybrid process model consisting of both the known unit models (mixer, flash and reactor tank) and an empirical model for the unknown reactor kinetics. A second aim of our approach is to maintain flexibility in choosing the most suitable modelling tool for each process unit. In this sense, reuse of unit models that already exist in different modelling tools should be possible. Such situations give rise to the need for an integration platform for different modelling tools that should also provide all the application specific algorithms for the overall heterogeneous model. To this end, the integration platform CHEOPS has been developed. CHEOPS facilitates the integration of modelling tools and provides flexibility for tailoring solution algorithms to a given problem. It has been already successfully used for steady-state simulation (Schopfer et al., 2004a), parameter estimation (Schopfer and Marquardt, 2004b), and dynamic simulation (Briesen et al., 2004).

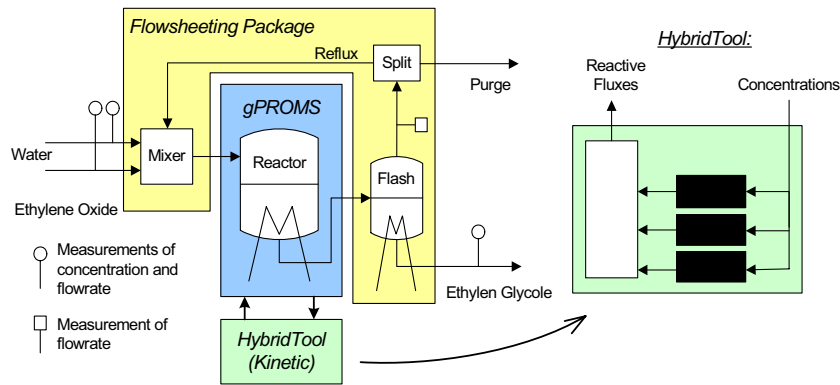


Figure 1: Flowsheet of a section of the ethylene glycol process model with measurement points

In the case of the process depicted in Figure 1 the models of the mixer, the splitter and the flash are supposed to be already available in some flowsheeting package. The reactor model is separated into a tank model and a reaction kinetics model. For the tank model, containing mass balances with the unknown reaction terms, the modelling tool gPROMS is used. For the reaction kinetics a structured hybrid model is already available in HybridTool (Mogk et al., 2002). The software HybridTool is a modular environment that has been developed by Bayer Technology Services for the generation and efficient identification of structured hybrid models, and has been successfully applied in different industrial case studies. Special features are the handling of

incomplete data sets and automatic structure optimization of the utilized black-box models. The HybridTool model for the reaction kinetics shown in Figure 1 consists of one white- and three black-box submodels. The white-box describes the known stoichiometry of the reaction, whereas the black-box models are used for the unknown relations between the reaction rates and the concentrations. For the parameter identification of the overall ethylene glycol process model CHEOPS is used.

2. Semi-empirical modelling of chemical process models

A chemical process is described by mathematical models of the unit operations and a flowsheet that connects the unit models. The unit models may contain unknown parameters or unknown functions. In the case study considered in this paper the three black-boxes of the HybridTool model for the reaction rates are realized by artificial neural networks. Using the connectivity information, the unit models and the neural networks are combined to an implicit flowsheet model $F(y,x,u,p)=0$, where x and y are dependent variables, u are independent variables and p are unknown parameters. Measurements are given for u and y .

The unknown parameters of the flowsheet model are estimated from the M available measurement data sets $[\tilde{u}_k, \tilde{y}_k]_{k=1..M}$ by a feasible path optimization with a weighted least-squares objective function

$$\underset{p}{\text{Min}} \quad \frac{1}{M} \sum_{k=1}^M (y_k(\tilde{u}_k, p) - \tilde{y}_k)^T W (y_k(\tilde{u}_k, p) - \tilde{y}_k). \quad (1)$$

The variables y_k are calculated from the flowsheet model $F(y_k, x_k, \tilde{u}_k, p) = 0$ for given \tilde{u}_k and p .

The solution of some chemical process unit models is only feasible in a limited range, e.g. flash or distillation column models. Furthermore the neural networks are highly nonlinear. Therefore the model F can only be solved, if special initialization strategies are applied. Depending upon which process variables are measured, the model can be initialized by two different strategies. In the *simultaneous approach* the neural networks are replaced by simplified models and equation (1) is solved to obtain estimates of all dependent variables, in particular the inputs and outputs of the neural networks. With these values the neural network parameters can be identified separately with a specialized identification algorithm, which is tailored to the neural network structure. Now all variables and parameters are initialized and equation (1) is solved for the original problem with neural networks. In the *incremental approach* a quite accurate initialization of equation (1) can be obtained if the measured variables suffice to solve the known model equations in F for the unknown input and output variables of the neural networks. If this is possible, the neural networks are trained separately from the flowsheet model. However, in this process numerical errors can be introduced, since the measurement noise might be amplified when solving the known model equations in F . Therefore, the identified neural networks are inserted again into the flowsheet model and equation (1) is solved to yield statistically sound parameter estimates. After the

parameter identification has been completed, the model is validated by calculating confidence intervals for the dependent variables.

3. Semi-empirical modelling with heterogeneous models in CHEOPS

Figure 2 gives an overview of the main elements that are involved in the hybrid modelling of the ethylene glycol process in CHEOPS. The reactor without kinetics is modelled in gPROMS, while the HybridTool is used for the semi-empirical reaction kinetics model. Models available in UnitGen (Geffers et al., 2001) are used for the mixer and the flash. UnitGen is a research prototype from our lab to generate executable models from a Modelica specification.

CHEOPS supports the construction of process models from unit models coming from different tools (Schopfer et al., 2004a) and allows to realize tailored solution algorithms (Schopfer et al., 2004b). It consists of *model-based application modules* that implement solution algorithms and a *neutral model representation* that provides the required mathematical information. A *tool wrapper* integrates a specific modelling tool into the neutral model representation. Finally, the communication between the tool wrappers and the modelling tools is established via an appropriate communication channel. For gPROMS, which provides a CORBA interface, this is the CORBA object bus, whereas UnitGen and the HybridTool are linked to CHEOPS as libraries.

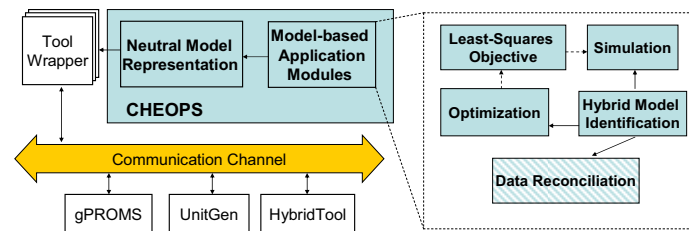


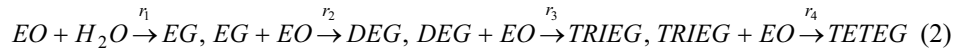
Figure 2: Hybrid modelling in CHEOPS for the ethylene glycol process

All model-based applications provide the same interface and are realized in a modular fashion so that they can be used as building blocks of the environment. This allows realizing aggregated model-based applications by combining these building blocks. In the box on the right of Figure 2 the model-based application modules used for hybrid model identification are shown together with their interactions. In order to allow for the flexibility required for hybrid model identification, the *Hybrid Model Identification Module* realizes the simultaneous and the incremental identification approaches described in the previous section. For this purpose it builds suitable submodels of the overall flowsheet model and applies the *Simulation Module* and the *Optimization Module*. For example, in the first step of the incremental approach, the *Simulation Module* is applied to the known partial model for the calculation of the inputs and outputs of the neural networks. The *Optimization Module* is used for the parameter identification in the first and the second step of the approach. It receives an implementation of the *Least-Squares Objective* function as stated in equation (1). For determining the variables y_k , the least-squares function has access to the *Simulation Module*. Note that instead of simply solving the known equations of F with the

Simulation Module in the first step of the incremental approach, a *Data Reconciliation Module* should be used instead, which explores overspecified problems and enforces physical constraints on the solution. This is not yet implemented in the tool set.

4. Case study: Ethylene Glycol (EG) Production Process

The EG process is modelled by material and energy balances. For the thermodynamical calculations the physical property package IKCAPE is used. The flash is modelled by a vapour-liquid equilibrium and is assumed to operate at a constant temperature of 305 K and a constant pressure of 0.3 bar. The reactor is cooled by a cooling jacket and operates at a constant temperature of 295 K. The production of mono-, di-, tri- and tetra-ethylene glycol (EG, DEG, TRIEG, TETEG) from ethylene oxide (EO) and water (H₂O) takes place in four parallel reactions



and can be described by the elementary reaction kinetics

$$r_1 = k_1 C_{EO} C_{H_2O}, r_2 = k_2 C_{EO} C_{EG}, r_3 = k_3 C_{EO} C_{DEG}, r_4 = k_4 C_{EO} C_{TRIEG} \quad (3)$$

with $k_1=60 \cdot N \cdot \exp(30.163-10583/T)$, $k_2=2.1 \cdot k_1$, $k_3=2.2 \cdot k_1$ and the normality $N=0.0873$ according to Parker and Prados (1964). The production of TETEG is comparably small and is therefore neglected in the following. In this case study the reaction rates r_1 , r_2 and r_3 are assumed to be unknown and are represented by the three black-boxes depicted in Figure 1. Each black-box is modelled by a neural network with two hidden nodes with a nonlinear transfer function, which results in a total of 27 unknown parameters. Since the temperature is assumed to be constant, the inputs of each black-box are two concentrations according to the reactions (3). The white-box has as inputs the three reaction rates, i.e. the outputs of the three black-boxes, and, according to the known stoichiometry, its outputs are the reaction fluxes of the five components.

Since real measurement data has not been available for this case study, the data necessary for model identification was created by simply simulating the model with the reaction rates from Parker and Prados (1964). The flow rate and the concentrations at the product outlet of the flash, as well as the flow rate in the recycle, are considered as measured variables. The flow rate at the mixer inlet is assumed to be 3000 kmol/h. 50 randomly distributed data points have been generated by varying the EO concentration at the mixer inlet within the range [0.15, 0.4].

Parameter identification, based on the incremental approach, is carried out in two steps, as described in Section 2. The measurements described above are sufficient for solving the partial process model consisting of the mixer and the flash model for the inputs and outputs of the reactor. Using these values and only the reactor model, the reactor kinetics is then identified within 42 iterations in the first step. The identified parameters are then used as initial values for the identification of the overall process model in the second step. In the second step a good fit was obtained after 10 iterations.

The graphs in Figure 3 show the EG concentration and the flow rates at the reactor outlet for the real model, the model identified in the first step of our approach and the model identified in the second step, as functions of the EO concentration at the mixer inlet. As expected, the model identified in the second step is closer to the real model than the one identified in the first step. Although the objective function did not explicitly optimize the prediction of the variables shown in the figure, the identification results after the second step are good inside the given range of measurements, i.e. EO concentration in [0.15, 0.4], but are also satisfactory outside.

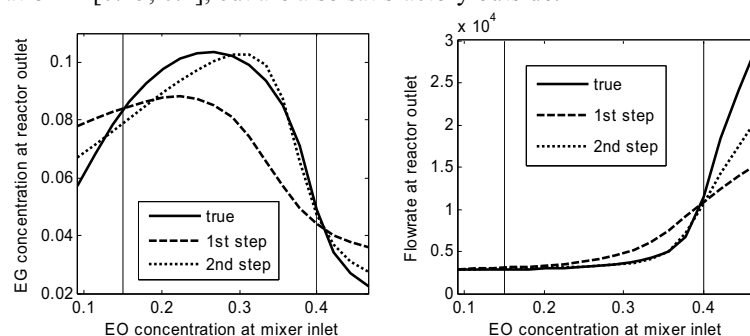


Figure 3: EG concentration (left) and flow rate (right) at the reactor outlet as functions of EO concentration at the mixer inlet.

References

- Briesen, H., R. Grosch, V. Kulikov, L. von Wedel, A. Yang and W. Marquardt, 2004, Coupled simulation of particulate process flowsheets by means of integration of different simulation tools, *Chemie Ingenieur Technik*, 76(6), 714-718.
- Geffers, W., L. von Wedel, J. Wyes and W. Marquardt, 2001, Integration of declarative and executable models in open simulation environments, In: Panreck K. and F. Dörrscheidt (Eds.), 15. Symposium Simulationstechnik, 579-586, Paderborn, Germany, September 11-14, 2001.
- Mogk, G., T. Mrziglod and A. Schuppert, 2002, Application of hybrid models in chemical industry, *European Symposium on Computer Aided Process Engineering - 12*, Elsevier.
- Parker, W. A. and J.W. Prados, 1964, Analog computer design of an ethylene glycol system, *Chemical Engineering Progress*, 60(6), 74-78.
- Schopfer, G., A. Yang, L. von Wedel and W. Marquardt, 2004a, CHEOPS: A tool-integration platform for chemical process modelling and simulation, Accepted for publication in: *International Journal on Software Tools for Technology Transfer*.
- Schopfer, G. and W. Marquardt, 2004b, A tool integration framework for general model-based applications, *Proceedings of ISC 2004 Annual Industrial Simulation Conference*, Malaga, Spain, June 7-9, 2004.
- Schuppert, A., 2000, Extrapolability of structured hybrid models: a key to optimization of complex processes, *Proceedings of EquaDiff '99*, Fiedler, B., K. Gröger and J. Sprekels (Eds.), 1135-1151.
- Thompson, M.L. and M.A. Kramer, 1994, Modelling chemical processes using prior knowledge and neural networks, *AIChE J.*, 40(8), 1328-1340.

Acknowledgements

The financial support of the Bundesministerium für Bildung und Forschung and the implementation work of W. Wontgaard (GUI) and C. Koch (error handling) is greatly appreciated.