Systems & Control Transactions

Research Article - Peer Reviewed Conference Proceeding
ESCAPE 35 - European Symposium on Computer Aided Process Engineering
Ghent, Belgium. 6-9 July 2025
Jan F.M. Van Impe, Grégoire Léonard, Satyajeet S. Bhonsale,
Monika E. Polańska, Filip Logist (Eds.)

PSE PRESS

Digital Twin supported Model-based Design of Experiments and Quality by Design

Amy Koch[a,b*], Jessica Ebert[a], Isabell Viedt[a,b], Andreas Bamberg[d], Leon Urbas[a,b,c]

[a] Technische Universität Dresden, Process-to-Order Group, Process Systems Engineering
[b] Technische Universität Dresden, Process-to-Order Group, Process-to-Order Lab Learning Factory
[c] Technische Universität Dresden, Process-to-Order Group, Chair of Process Control Systems
[d] Merck Electronics KGaA
* Corresponding Author: amy.koch@tu-dresden.de.

ABSTRACT

The pharmaceutical and specialty chemical industries are challenged with the requirement of faster time-to-process to meet market demands. Here, Modular Plants made up of predesigned process equipment assemblies (PEAs) are advantageous. Moreover, equipment-based Digital Twins of these modules can further reduce the time-to-process when combined with methods such as Quality by Design (QbD) and model-based design of experiments (MBDoE) to reduce uncertainty. This paper presents a lab scale-based workflow using an equipment-based Digital Twin, which applies QbD and MbDoE methods to identify the Design Space in the lab scale which can be transferred to production scale equipment as part of a Digital Twin based workflow for scale-up in Modular Plants.

Keywords: Digital Twins, Model-based Design of Experiments, Quality by Design, Scale-up

MOTIVATION

In the specialty chemical and pharmaceutical industries, faster time-to-process is a significant measure of success. Here the use of predesigned modules according to VDI 2776 [1] are advantageous. Additional benefit can be derived by creating equipment-based Digital Twins of these modular process equipment assemblies which can be combined with multiple product descriptions. This further supports and expedites the process development phase in multipurpose plants by providing Digital Twin supported prior knowledge of equipment capabilities with multiple product descriptions. In addition to these Digital Twins, another method to support faster time-to-process is reducing experimental efforts in the process development phase using model-based methods. Here, Digital Twin workflows which incorporate methods such as Quality by Design (QbD), Global System Analysis, model-based design of experiments (MBDoE), and the identification of the Design Space as well as leveraging prior knowledge of the equipment capabilities can be utilized to reduce the experimental load [2]. Prior knowledge of equipment capabilities and limits is supplied by the Digital Twin and is used to pre-screen combinations of Critical Process Parameters (CPP), Critical Material Attributes (CMA), and model parameters to identify suitable parameter combinations for inclusion in the MBDoE optimization problem [3].

In this paper, the relevant methodology and background is presented. Then, a Digital Twin workflow in the using a Capability Model of a Digital Twin of the lab scale equipment is introduced, explained, and applied using the process simulation tool gPROMS® for the specific example of an esterification process in a stirred tank reactor module. The use case highlights the benefits of combining these methodologies such as improved reduced experimental effort compared to traditional DoE and increased knowledge of Critical Process Parameters. We conclude this work with a critical evaluation of the applied methods and an outlook to future work.

METHODOLOGY

Digital Twins in Modular Plants

Modular Plants (MPs) are defined in the VDI 2776 [1] and VDI 2658 [4] set of specifications and consist of Functional Equipment Assemblies (FEAs) grouped into reusable and adaptable Process Equipment Assemblies (PEAs). The functional capabilities of the PEAs are described by services, which can be orchestrated into recipes in the process orchestration layer (POL)[5].

For the context of this paper, a Digital Twin is defined as a semantically linked collection of digital artifacts [6]. These digital artifacts can be organized into partial models following a Product-Process-Resource-Model [7]. For Modular Plants, the Digital Twin is created from knowledge that is shared between PEA manufacturers and owner/ operator (O/O). The PEA manufacturer knowledge is a **Capability Model** describing the equipment capabilities independent of the product. The owner/operator knowledge is a **Transformation Model**, which is equipment independent and product specific. The Digital Twin combines these two models to form an **Operation Model**, which is used for recipes and process optimization. Similar separation-of-concern approaches are discussed in the domain of discrete manufacturing with a focus on standardized description of tasks and skills of machinery, e.g. [8], [9], [10]. For this purpose of this work, the Digital Twin is an equipment-based Digital Twin of a PEA which should be understood as a well characterized **Capability Model** that can be flexibly combined with a **Transformation Model** and uncertainties of a new product.

Quality by Design

The Quality by Design (QbD) methodology is outlined in the ICH Q8 - 10. The QbD methodology combines different scientific methods such as Design Space identification, design of experiments, and risk management with the result of guaranteed product quality [11]. The two most relevant sub-concepts in QbD methodology are statistical DoE and the Design Space (DS). Statistical DoE is used to identity which experiments must be run to collect data required to describe the Knowledge Space. From the Knowledge Space, the Design Space can be identified. The Design Space describes the relationship between Critical Process Parameters (CPPs) and Critical Material Attributes (CMAs) which are inputs to a process and the acceptable ranges of Critical Quality Attributes (CQAs) of the product.

Model-based Design of Experiments

As an alternative to statistical design of experiments, Model-based Design of Experiments (MBDoE) is a methodology which integrates differential and algebraic equation (DAE) type models and Design of Experiments (DoE). Here, the goal is to select the experiment variables with the highest sensitivities to the model parameters, yielding improved information from experiments [12], [13]. Prior knowledge (model structure from the Capability Model and initial parameter estimates from the Transformation Model) are used to design an experiment [14]. The result is quick identification of optimum process conditions based on the Critical Process Parameters for experiments, this aims to reduce experimental load as the information obtained from experiments increases. This is advantageous in process development workflows where time-to-process is critical and with additional benefit of supporting situations where experiment budget is limited.

Digital Twin Workflow

The application of QbD and MBDoE methodologies into smart PEAs to reduce uncertainty in the process qualification phase has been explored by [3], [15] . The workflow presented in Figure 1 extends the workflow from [15] and is intended for scale-up based on a Digital Twin as shown in [2], [16]. This work focuses on the pre-

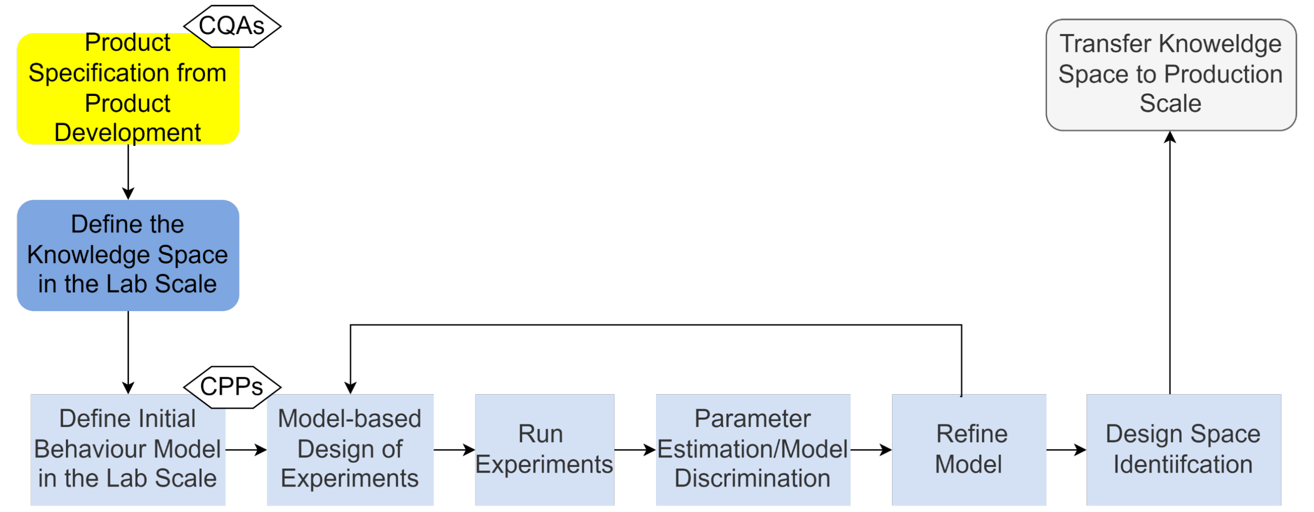

Figure 1: Digital Twin based workflow using the equipment-based Digital Twin in the lab scale with input from product specification and CQAs from product development.

screening using Global System Analysis, MBDoE, and identification of the Design Space in the lab scale using a Digital Twin of a PEA. This Digital Twin in this work combines a Capability Model [7] of the PEA with a Transformation Model of the product yielding an Operation Model. In the workflow in Figure 1, first Critical Quality Attributes are identified based on a product specification provided by product development. Then the Digital Twin of the lab scale is utilized. Using this Operation Model of the lab scale equipment, first Global System Analysis and variance-based sensitivities are used identify Critical Process Parameters and Critical Material Attributes based on the identified Critical Quality Attributes from the product specification. Critical model parameters for the MBDoE optimization problem are also identified.

Next, model-based design of experiments combines the prior knowledge of the CPPs from the Capability Model to identify the ideal set of process conditions to yield the maximum amount of information from experiments. The model parameters can be divided according to the PPR model [7], where the product specific parameters (i.e. activation energy, k_∞, heat of reaction) are described by the Transformation Model and resource specific parameters (i.e. heat transfer coefficients, mixing time) are described by the Capability Model. Here, it is assumed that the uncertainty regarding the Transformation Model is significantly higher than model parameters which are described by the Capability Model, which has been characterized and validated from characterization experiments and previous processes run in the equipment. Following the MBDoE, experiments are run, and the experimental data is used to improve the parameter estimates in the model. These improved parameter estimates are then implemented into the model and validated with experimental data which has not been used for the parameter estimation. If the predictive capabilities of the model are sufficiently accurate based on the validation data, the model can be used to define the Knowledge Space. The final step of the workflow in Figure 1 in the lab scale is to use the CQAs to identify the Design Space from the Knowledge Space. Based on the identified Design Space, the relevant model parameters and uncertainty from the Transformation Model are transferred from the Operation Model of the lab reactor to the Capability Model of the production equipment. This yields an Operation Model of the production equipment which is then used to characterize the Knowledge Space and subsequently identify the Design Space in the production scale.

CASE STUDY

Application Example

To illustrate the methods of the workflow presented in Figure 1, an esterification reaction in a 2 L stirrer tank reactor was selected. The butyl acetate synthesis [17] is a homogeneously catalysed reaction. The catalyst in this example is sulfuric acid. The CQA for the product specification is the concentration $C_{butyl\ acetate} \geq 0.5\ mol/L$ of the target product butyl acetate the outlet of the reactor.

$$butanol + acetic\ acid\ \leftrightarrow butyl\ acetate + water \quad (1)$$

The kinetic behavior of the process can be described with the following set of equations:

$$r = c_{sulfuric\ acid}\,(k_1 * c_{butanol} * c_{acetic\ acid} - k_2 * c_{butyl\ acetate} * c_{water}) \quad (2)$$

$$K_{eq} = k_1/k_2 \quad (3)$$

$$k_i = k_{i,\infty} \exp\left(\frac{-E_{A,i}}{R\,T}\right) \quad (4)$$

To mimic the real conditions in process development, initial estimates for the kinetic parameters and uncertainties was estimated by fitting real experimental data from [17] to Arrhenius' law given by Equation 4.

$$E_{A,1} = 36281\,\frac{J}{mol};\ k_{1,\infty} = 31.5\frac{L^2}{min \cdot mol^2}$$

$$E_{A,2} = -478.34\,\frac{J}{mol};\ k_{2,\infty} = 1.6E^{-5}\,\frac{L^2}{min \cdot mol^2}$$

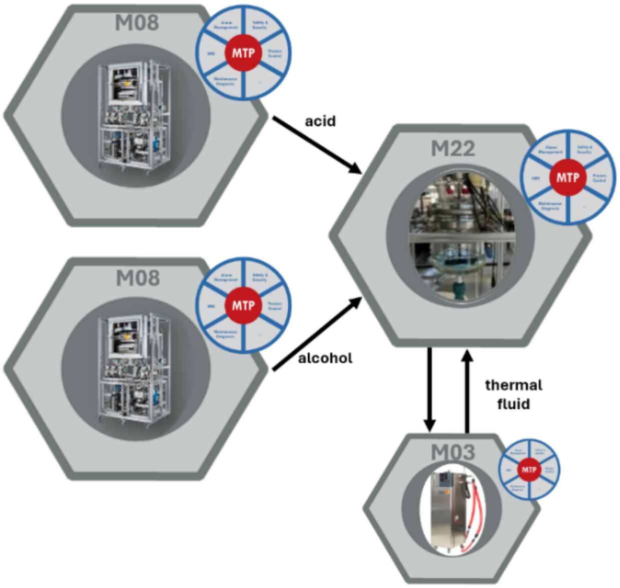

Figure 2. Modular plant configuration for butyl acetate synthesis.

Figure 2 provides an overview of the modular configuration which is used to carry out the esterification process. Here, three different modules are used:

- M08 is a doubling dosing module which is used to store and supply the reactants to the reactor

- M22 is a 2 L stirred tank reactor which is used to carry out the reaction
- M03 is a thermostat module used to supply mono ethylene glycol to the M22 reactor jacket.

We can describe our system with the following set of variables:

- **x** describes the state of the system at time (t)
- $\mathbf{u}_{CMA}$ describes independent inputs related to CMAs
    - Purity of the reactants $(x_{Acid,1}, x_{Alcohol,2})$
- $\mathbf{u}_{CPP}$ describes independent inputs related to CPPs
    - Reactor temperature (T) controlled by the Temperature of the heating medium T_{HM} with the Service *Tempering*
    - Feed ratio of reactants $(\dot{n}_{Alcohol}, \dot{n}_{Acid})$ controlled by the *Dosing* Services
    - Mixing speed $n_{Stirrer}$ controlled by the *Stirring* Service
- $\boldsymbol{\theta}_{TM}$ describes the parameter values related to the Transformation Model of the product:
    - Activation energy $(E_{A,1}, E_{A,2})$
    - Pre-exponential factor $(k_{inf,1}, k_{inf,2})$
- $\boldsymbol{\theta}_{CM}$ describes the parameters contained in the Capability Model of the PEA:
    - Heat transfer coefficient (k_D)

In the system described by the variable set, the significant uncertainty in the model is the Transformation Model parameters $(\boldsymbol{\theta}_{TM})$ as outlined in the methodology.

To better understand the effect that the uncertainty of the estimated model parameters in the Transformation Model $\boldsymbol{\theta}_{TM}$ has on the predictive capabilities of the Digital Twin, a Global System Analysis is carried out with respect to $\boldsymbol{u}_{CPP}$ the described by the resources model and $\boldsymbol{u}_{CMA}$ described by the product model. The parameter screening uses the GSA tool in gPROMS®, which combines a Monte-Carlo-Simulation and Sensitivity analysis based on variance-based indices. The result of the pre-screening is a series of scatter plots that serve two purposes: (1) identify relationships between CPPs u_{CPP} and CMAs u_{CMA} and model parameters $\boldsymbol{\theta}_{TM}$ and (2) identify relationship between model parameters $\boldsymbol{\theta}_{TM}$ and the CQAs. A selection of two GSA scatter plots is shown in Figure 3. The effects of model parameter from the Transformation Model $\theta_{TM}(E_{A,1})$ and CPP from the resource model $u_{CPP}(T_{HM})$ are shown. Here, it is evident that the estimated value of model parameter of the activation energy of the forward reaction $E_{A,1}$ has a significant influence on the target CQA, while the CPP of the temperature of the heating medium T_{HM} does not have a clear effect.

MBDoE Problem

To describe the Knowledge Space prior knowledge from the equipment capabilities from the PEA manufacturer is combined with initial parameter estimates provided by the owner/operator. These initial parameter estimates can be obtained with short cut models or group contribution estimation methods. This yields a Digital Twin built from combined knowledge of a characterized equipment model with less characterized product model.

Using this behavior model of the Digital Twin, MBDoE methods [14] are applied for parameter estimation and for model discrimination.

$$y_i = m(u_i, \boldsymbol{\theta}_{TM}) + \epsilon_i \quad (5)$$

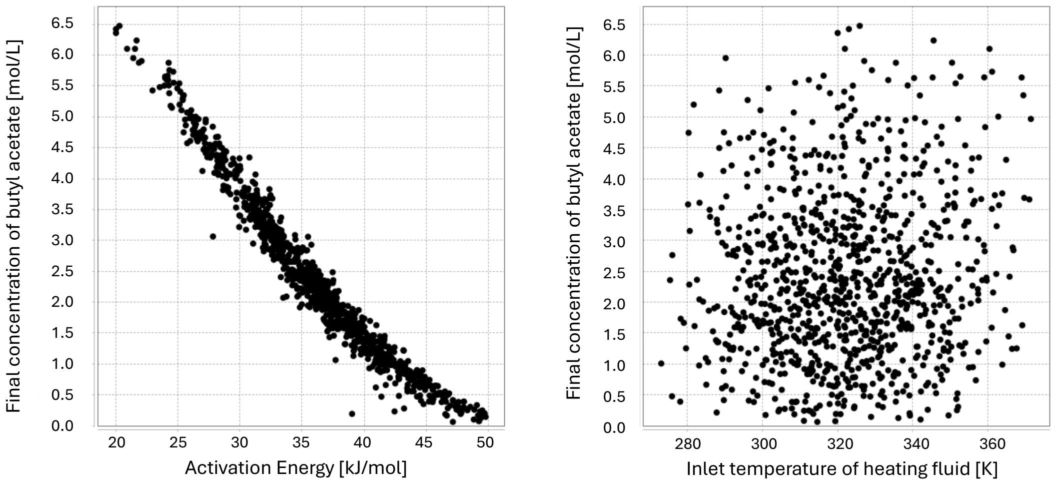

Figure 3: Influence of the Transformation Model parameter of Activation Energy $\theta_{TM}(E_{A,1})$ on the CQA $(C_{butyl\ acetate})$ (left). Influence of the CPP of temperature of the heating medium $u_{CPP}(T_{HM})$ on the the CQA $(C_{butyl\ acetate})$ (right).

Here, y_i are the measurements (concentration samples), m is the DAE model, θ_{TM} is the set of model parameters to be optimized ($E_{A,1}$, $k_{1,\infty}$), u_i are the control variables based on the service specification and parameterization of the module which can be manipulated in the optimization problem. For the MBDoE optimization problem, the experimental budget was determined to be six experiments: five for model-based experiments and one for model validation. The data collected during these experiments is used to estimate parameters in the Transformation Model $\boldsymbol{\theta}_{\mathrm{TM}}$ including uncertainty estimates, applying e.g., the maximum likelihood method [14].

Identification of the Design Space from the Knowledge Space

Once the model has been revised with the parameter estimates from the experiment data. The design spec can be identified. To describe the Knowledge Space a system of nonlinear algebraic equations (NAEs) based on the previously described set of variables is proposed:

$$\mathbf{0} = \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{u}_{\mathrm{CMA}}, \boldsymbol{u}_{\mathrm{CPP}}, \boldsymbol{\theta}_{\mathrm{TM}}, \boldsymbol{\theta}_{\mathrm{CM}}) \; (6)$$

To identify the Knowledge Space, prior knowledge is required. The prior knowledge of CPPs is based on the resource model, specifically the operational limits of the model and the service specification described by the Capability Model of the Digital Twin. Then, the model according to Equation (1) is revised with the improved estimates of the model parameters. To identification of the Design Space from a Knowledge Space described in Equation (6), an Extended Flexibility Analysis approach such as described by [18] and illustrated in [3], [15] is applied. In the Flexibility Analysis, the range of the CQAs are described by the feasibility function

$$\boldsymbol{g_j}(\boldsymbol{x}, \boldsymbol{u}_{\mathrm{CMA}}, \boldsymbol{u}_{\mathrm{CPP}}, \boldsymbol{\theta}_{\mathrm{TM}}, \boldsymbol{\theta}_{\mathrm{CM}}) \leq 0 \; (7)$$

An example of the Knowledge Space according to Equation 6 as well as the identification of the Design Space from the Knowledge Space as part of the flexibility analysis applying Equation 7 is shown in Figure 4.

DISCUSSION

The Digital Twin workflow based on the Capability Model of the lab scale equipment presented in the methodology and illustrated using the application example of the esterification process shows how application of MBDoE methods reduces the required number of experiments required to estimate new parameter values to improve the prediction accuracy of the lab scale model. In the example here, an experimental budget of six experiments was selected, whereas statistical plans to describe the Knowledge Space require a minimum of nine or more experiments. Although real experimental data was used to generate initial estimates for the model parameters, the experimental data for the MbDoE optimization problem was simulated using models based on the real equipment geometry and capabilities. The incorporation of real experimental data into MBDoE step of the workflow would strength the results of the use case. However, the benefit of the workflow is still illustrated by the presented results. Moreover, metrics based on heuristics are needed to compare the results of this method both in terms of time saved as well as information gained.

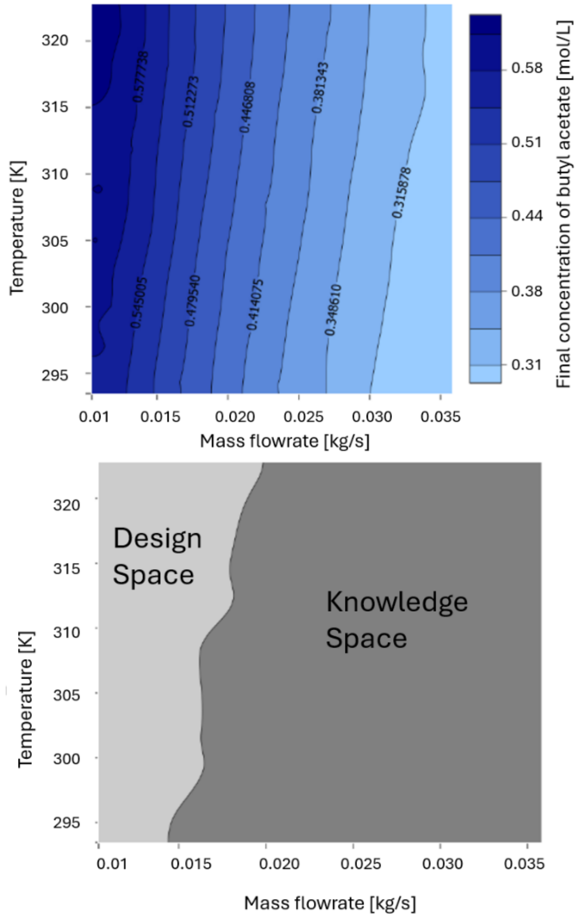

Figure 4. Identification of the Knowledge Space (top), identification of the Design Space from the Knowledge Space based on the CQA criteria from the product model (bottom).

CONCLUSION AND OUTLOOK

This paper presented a Digital Twin based workflow using a Capability Model of a lab scale PEA incorporating MBDoE and QbD methodology for simultaneously increasing information about the Knowledge Space and resulting Design Space, while using model-based methods to reduce experimental load in the process development phase resulting in a faster time-to process. In a next step, the workflow should be illustrated using real experimental data. Future work will also focus on the specific challenges in scaling-up between dimensions as well as the benefits of scale-up in Modular Plants where the advantages of process intensification can be retained. A specific challenge is to integrate the identified parameter values from Transformation Model along with uncertainty factors into the Capability Model of the production scale equipment. These Capability Models are for equipment that is functionally similar, but of different scales and different PEA Manufacturers.

ACKNOWLEDGEMENTS

The authors would like to thank our partners Merck Electronics KGaA and Siemens AG in the industrial research project Lab2Plant for the fruitful discussions as well as P2O-Lab Learning Factory at TU Dresden's Process-to Order Group for the information about their modules M02, M08, and M22.

REFERENCES

1. VDI 2776 Blatt 1, Process engineering plants - Modular plants - Fundamentals and planning modular plants. 2020.
2. A. Koch, J. Mädler, A. Bamberg, and L. Urbas, "Digital Twins for Scale-Up in Modular Plants: Requirements, Concept, and Roadmap," in Computer Aided Chemical Engineering, vol. 52, Elsevier, 2023, pp. 2063–2068. doi: 10.1016/B978-0-443-15274-0.50328-0.
3. J. Mädler, A. Koch, A. Bamberg, and L. Urbas, "Process Validation in Modular Plants considering the Scale-Up and Prior Knowledge," in Computer Aided Chemical Engineering, vol. 52, A. C. Kokossis, M. C. Georgiadis, and E. Pistikopoulos, Eds., Elsevier, 2023, pp. 2057–2062. doi: 10.1016/B978-0-443-15274-0.50327-9.
4. VDI/VDE/NAMUR 2658-1:2019, Automation engineering of modular systems in the process industry - General concept and interfaces, Oct. 2019.
5. A. Klose et al., "Anforderungen an die Orchestrierung modularer Prozessanlagen," 2019, doi: 10.25532/OPARA-29.
6. R. Rosen, J. Fischer, and S. Boschert, "Next Generation Digital Twin: an Ecosystem for Mechatronic Systems?," IFAC-Pap., vol. 52, no. 15, pp. 265–270, Jan. 2019, doi: 10.1016/j.ifacol.2019.11.685.
7. A. Bamberg, L. Urbas, S. Bröcker, M. Bortz, and N. Kockmann, "The Digital Twin – Your Ingenious Companion for Process Engineering and Smart Production," Chem. Eng. Technol., vol. 44, no. 6, pp. 954–961, 2021, doi: 10.1002/ceat.202000562.
8. M. Schleipen, A. Lüder, O. Sauer, H. Flatt, and J. Jasperneite, "Requirements and concept for Plug-and-Work: Adaptivity in the context of Industry 4.0," - Autom., vol. 63, no. 10, pp. 801–820, Oct. 2015, doi: 10.1515/auto-2015-0015.
9. J. Pfrommer, D. Stogl, K. Aleksandrov, S. Escaida Navarro, B. Hein, and J. Beyerer, "Plug & produce by modelling skills and service-oriented orchestration of reconfigurable manufacturing systems," - Autom., vol. 63, no. 10, pp. 790–800, Oct. 2015, doi: 10.1515/auto-2014-1157.
10. J. Pfrommer et al., "An ontology for remanufacturing systems," - Autom., vol. 70, no. 6, pp. 534–541, Jun. 2022, doi: 10.1515/auto-2021-0156.
11. S. Chatterjee, C. M. V. Moore, and M. M. Nasr, "An Overview of the Role of Mathematical Models in Implementation of Quality by Design Paradigm for Drug Development and Manufacture," in Comprehensive Quality by Design for Pharmaceutical Product Development and Manufacture, 2017, pp. 9–24. doi: 10.1002/9781119356189.ch2.
12. D. M. Bates and D. G. Watts, Eds., Nonlinear Regression Analysis and Its Applications. in Wiley Series in Probability and Statistics. Hoboken, NJ, USA: John Wiley & Sons, Inc., 1988. doi: 10.1002/9780470316757.
13. N. Asprion et al., "Implementation and Application of Model-Based Design of Experiments in a Flowsheet Simulator," J. Chem. Eng. Data, vol. 65, no. 3, pp. 1135–1145, 2020a, doi: 10.1021/acs.jced.9b00494.
14. G. Franceschini and S. Macchietto, "Macchietto, S.: Model-based design of experiments for parameter precision: State of the art. Chemical Engineering Science 63(19), 4846-4872," Chem. Eng. Sci., vol. 63, pp. 4846–4872, Oct. 2008, doi: 10.1016/j.ces.2007.11.034.
15. J. Mädler, Smarte Process Equipment Assemblies für die Unterstützung der Prozessvalidierung in modularen Anlagen, Als Manuskript gedruckt. in VDI Fortschritt-Berichte Reihe 20, Rechnerunterstützte Verfahren, no. Nr. 480. Düsseldorf: VDI Verlag GmbH, 2023.
16. A. Koch, J. Madler, A. Bamberg, and L. Urbas, "Towards PEA Matching from Simulation as Part of a Digital Twin Concept for Scale-Up in Modular Plants," in Computer Aided Chemical Engineering, vol. 53, Elsevier, 2024, pp. 595–600. doi: 10.1016/B978-0-443-28824-1.50100-9.
17. R. Soni and G. Jyoti, "Kinetics of esterification of acetic acid with n-butanol over different homogeneous acid catalysts," Int. J. Chem. Kinet., vol. 56, no. 2, pp. 96–104, Feb. 2024, doi: 10.1002/kin.21694.

18. M. P. Ochoa, S. García-Muñoz, S. Stamatis, and I. E. Grossmann, “Novel flexibility index formulations for the selection of the operating range within a design space,” Comput. Chem. Eng., vol. 149, p. 107284, 2021, doi: https://doi.org/10.1016/j.compchemeng.2021.107284.
