Check for updates

Reaction Optimization of a Suzuki-Miyaura Cross-Coupling Using Design of Experiments

Jens Bobers[1,*], Lisa Katharina Hahn[1], Tobias Averbeck[1], Andreas Brunschweiger[2], and Norbert Kockmann[1]

DOI: 10.1002/cite.202100194

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

The combination of lab automation and design of experiments for the execution of screening experiments increases productivity and reduces error-prone manual work. A self-developed software tool allows for creating fractional-factorial experimental design (FFED). Application of FFED on the screening of a Suzuki-Miyaura cross-coupling leads to a 93 % reduced design compared to full-factorial design. The resulting regression model qualitatively shows the positive effect of educt concentrations, time, and temperature and reveals the decrease in conversion at high base concentrations.

Keywords: Design of experiments, Lab automation, Reaction optimization, Suzuki-Miyaura reaction

Received: October 31, 2021; *revised:* February 04, 2022; *accepted:* February 21, 2022

1 Introduction

High-throughput screening (HTS) is a well-known industrial and research technique for screening and testing active pharmaceutical ingredients, bioactive molecules or catalysts and combines automation technologies with statistical experimental design [1]. In the (bio-)chemical lab HTS technologies are more difficult to implement, as most experiments are performed manually [2]. This leads to a high workload of error-prone manual tasks, which laboratory automation technology can reduce and provide for reproducible quality and increased productivity through parallel execution of repetitive laboratory tasks. Similar to industrial HTS technology, laboratory automation unfolds all its strengths in combination with statistical design of experiments [1].

According to our experience with design of experiments (DoE) in various applications, e.g., solid-liquid flow and particle precipitation, gas-liquid flow [3], and dispersion in microjets [4], as well as real-time reaction optimization with inline analysis [5], DoE enables a further increase in experimental efficiency by avoiding unnecessary experiments and revealing possible, previously unknown underlying mathematical relationships through statistical analysis [1, 6]. As most chemical reactions in the lab are performed batchwise, the progress of self-optimization is slow due to complex connection between sample preparation, execution of the experiment, and analysis of the reaction mixture [7]. Specifically for chemical reaction screening and reaction optimization, DoE leads to statistically valid results by full factorial experimental designs of an exemplary size of 2^n for n factors investigated at two levels. A reduction in workload can be achieved by creating fractional factorial experimental designs (FFED) with a reduced size of 2^{n-p}, where p is the size of the fraction of a full factorial experimental design, allowing a larger chemical space to be covered [6, 7]. Furthermore, DoE allows for discovering interactions between the influencing parameters, which design size makes it possible to find optima for the chemical process [1].

Nonlinear regression is a tool for visualization of data and deriving a mathematical expression for various input parameters, which can be derived by multidimensional experiment. Understanding the causal relationship between output and several input parameters is the key feature of screening of chemical reactions. [8] Concentrations of multiple reaction components and catalysts with reaction temperature and time combine for a parameter space, which leads to many screening experiments and complex relations between input and output values. For example, Rosso et al. [9] combined automation, DoE, and regression analysis for chemical discovery to show the synergies of a combination of those techniques.

In this work, a newly developed DoE and regression software tool is applied to a Suzuki-Miyaura cross-coupling reaction with five different input parameters and the product concentration as the output parameter. The large

[1]Jens Bobers, Lisa Katharina Hahn, Tobias Averbeck, Prof. Dr.-Ing. Norbert Kockmann
jens.bobers@tu-dortmund.de
TU Dortmund University, Department of Biochemical and Chemical Engineering, Emil-Figge-Straße 68, 44227 Dortmund, Germany.
[2]PD Dr. Andreas Brunschweiger
TU Dortmund University, Department of Chemistry and Chemical Biology, Otto-Hahn-Straße 6, 44227 Dortmund, Germany.

 © 2022 The Authors. Chemie Ingenieur Technik published by Wiley-VCH GmbH

parameter space is reduced by designing a fractional-factorial experimental design, and the results are fitted by a multidimensional regression model that provides insight into the effects of each input parameter on the product concentration. Reaction preparation is performed using a custom laboratory robot [10] that executes the preplanned experimental design.

2 Design of Experiments

DoE is a known tool since long-time for various disciplines to gain knowledge about complex processes depending on many influencing parameters. DoE provides statistical methods to design an experimental plan to reduce the number of experiments on the one hand and to get maximum knowledge from the experimental results on the other hand. For the planning of chemical experiments, a DoE software tool was developed using Matlab® and its included statistical toolboxes. The DoE software includes a regression tool, which can calculate multidimensional and nonlinear regression models from the data derived from the preplanned experiments. As the software is focused on screening organic syntheses, the statistics are based on fractional-factorial experimental design for reduction of number of needed experiments.

In contrast to commercially available DoE software, when the user must decide how many experiments to perform, the custom program presented here enables to choose an experimental design in dependence on its design efficiency and number of experiments. This efficiency rating bases on different optimality criteria creating a variance-optimal experimental design in relation to the Fisher matrix. As a starting point, a full-factorial plan is generated as a base for all reduced plans. Based on this plan, an effect matrix is created, transforming the full-factorial experimental design into a matrix consisting only of zeros and ones, with all information preserved. This effects matrix is then orthonormalized by Gram-Schmidt orthonormalization. The orthonormalized effect matrix is used to find D-optimal designs by applying the Fedorov exchange algorithm. For D-optimal designs, the best design is found, when the determinant of the Fisher matrix is maximized in accordance with Eq. (1). p represents the number of input parameters and N_D is the number of trials of the reduced design.

$$\eta_D = 100\frac{1}{N_D\left|(X'X)^{-1}\right|^{1/p}} \tag{1}$$

The selection criterion is the D-optimality for choosing the most efficient design from several thousand generated experimental design. An objective function was defined targeting the number of trials as well as the calculated design efficiency. Eq. (2) shows the selected objective function including the number of trials of the reduced design N_D, the number of trials of the full-factorial design N_{full}, the calculated design efficiency η_D, and the minimum design efficiency of all generated designs η_{min}.

$$\phi = \frac{1}{2}\left(1 - \frac{N_D}{N_{full}} + \frac{\eta_D - \eta_{min}}{100 - \eta_{min}}\right) \cdot 100\% \tag{2}$$

The reduction of experiments and the design efficiency are weighted equally, which can be adjusted for different experiments.

Based on the resulting data from the experimental plan, the regression tool calculates various model equations based on 13 basic mathematical expressions and optimizes these equations along customizable requirements, e.g., root mean square error (RMSE), correlation coefficient R^2, or similar error definitions. Fig. 1 demonstrates all mathematical expression that can be combined by the regression tool to obtain a model equation. Expanding the application range, more expression can be added to the list. For known physical or chemical relations, the user can preselect the equation that should be used by the software, which ranks all obtained model equation regarding different error definitions.

Ensuring good quality of the model, the software follows an algorithm to find suitable initial values for the modelling. First, the algorithm eliminates the columns of the input matrix one by one and searches for identical rows in the remaining columns. The resulting data are used for an initial 2D fit of the eliminated column to determine initial values for the regression. Variables with less than three identical rows found by the algorithm are omitted to avoid errors. If no such rows are found and no initial values could be calculated, the data is rounded up to the next power of ten and the procedure starts again. When no initial values could be found, predefined default values are taken. This is done for each of the mathematical expressions presented above, and

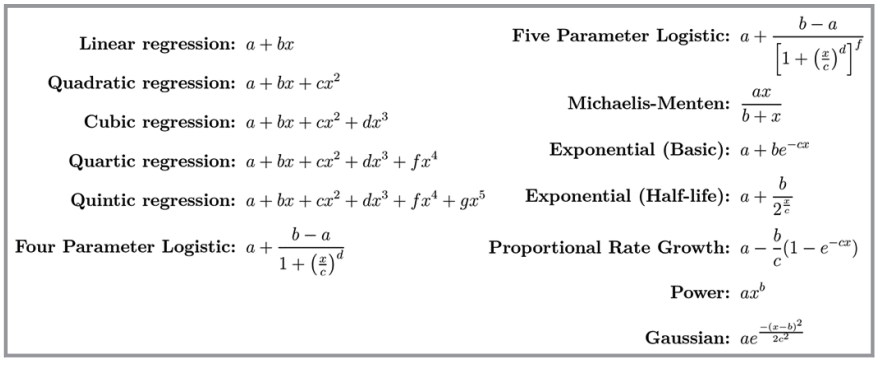


Figure 1. 13 basic mathematical expressions used by the custom nonlinear regression tool to gain a combined regression equation.

 © 2022 The Authors. Chemie Ingenieur Technik published by Wiley-VCH GmbH

15222640, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cite.202100194 by Technische Universitaet Dortmund Dezernat Finanzen und Beschaffung, Wiley Online Library on [28/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the parameters are evaluated based on their specific RMSE. In the next step, all possible combinations of the 13 mathematical expressions are calculated and classified according to the preselected error definition.

The validity of the obtained equations is checked by k-fold cross validation, where k is normally 10. The k-fold cross validation splits the data randomly in k subparts of similar size. The model parameters are estimated with k–1 subsets, while the remaining subset is utilized to validate the obtained parameters. The process is repeated k times as each subset is used for both validation and parameter estimation. A cross-validated model error (e.g., RMSE) indicates the validity of the derived model parameters. Here, the models were ranked by its cross-validated coefficient of correlation R^2. For chemical screening purposes, it is important to get a qualitative trend where the reaction optimization should head to. Therefore, the experimental plan for the reaction investigated in this contribution is reduced with the intention to show how to gain knowledge from a complex reaction with a close to minimum number of experiments. For example, although the relationship between concentration and temperature is already known, no specific equation was preselected to gain experience with how the software would process the generated data without the prior knowledge provided by the user.

3 Experimental Setup

The experiments presented in this contribution were prepared with the automated dosage system (ADoS) [10], an in-house developed lab robot capable of pipetting and mixing several reagents into a 96-well microtiter plate together with performing preplanned experimental routines to fulfill DoE requirements. As shown in Fig. 2, the ADoS is based on the open-source design of a 3D printer and consists of an aluminum frame and working platform executing movement in the z-dimension. The printer head was replaced by an automated injection unit (AIU), which is responsible of the process of loading and releasing reagents into lab equipment placed onto the working platform of the system and is moved in the X-Y plane. Modularization of this platform allows for inserting up to six operation plates (OP) into customized templates located on the platform, which fulfills various tasks for standard lab work. Here, four OPs are used. A 96-well microtiter plate acts as the reaction vessel, while the cleaning station continuously provides distilled water and acetonitrile for cleaning the syringe between pipetting steps. The OP with Schott bottles stores the reagents for the performed reaction. The OP with Petri dishes is used for dilution and fast discontinuous cleaning of the needle of the syringe. The reaction mixtures were placed on a thermal shaker plate (VWR Thermal Shake lite, VWR International, Radnor, USA) for the preplanned reaction time and at the desired reaction temperature.

The reaction scheme of the investigated Suzuki-Miyaura cross-coupling reaction is presented in Fig. 3 in combination with the varied as well as constant process parameters including their variation range. The reaction was performed in a 4:1 mixture of tetrahydrofuran (THF) (VWR Chemicals, Radnor, USA) and water. The desired product, biphenyl, is synthesized by a reaction of phenyl iodide (98 % purity, Alfa Aesar, Haverhill, USA) and phenylboronic acid (95 % purity, Aldrich Chemistry, St. Louis, USA) catalyzed by tris(dibenzylideneacetone)dipalladium (97 % purity, Aldrich Chemistry, St. Louis, USA) and tricyclohexylphosphane (> 88 % purity, Alfa Aesar, Haverhill, USA) as ligand. As described in literature [11] potassium fluoride (99 % purity, Acros Organics B.V.B.A, Fair Lawn, USA) was added to the reaction mixture as a base.

Both educts, PBA and PI, and PF are varied in concentrations in the range from 1 mM to 5 mM in 1 mM increments. Three variations of the reaction temperature as well as for the reaction time were investigated. The concentration of the catalyst and the ligand were kept constant for all performed experiments. As shown in Fig. 3c, UV/Vis spectroscopy as analytical method provides reproduceable results for the desired product biphenyl in the expected concentration range. The calibration was performed in triplicates with pure biphenyl dissolved in THF/water mixture.

A full-factorial experimental plan consists of all combinations of the influencing parameters and includes 1125 experiments. With help of the self-designed

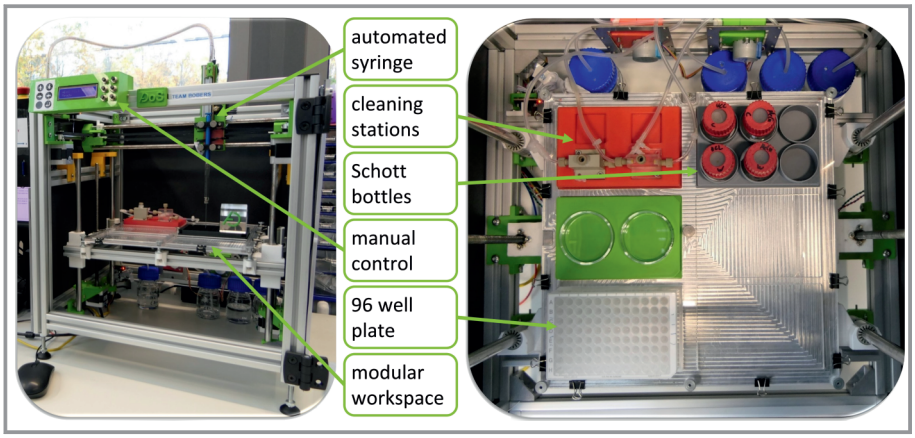
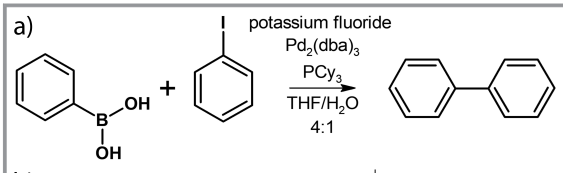


Figure 2. Experimental setup for the batchwise execution of a Suzuki-Miyaura cross-coupling reaction. Left side: in-house developed lab robot automated dosage system (ADoS), right side: modular workspace of the ADoS including a 96-well microtiter plate as reaction vessels, continuous cleaning for the syringe, Schott bottles for storage of reagents, and Petri dishes for dilution and discontinuous cleaning of the syringe.

 © 2022 The Authors. Chemie Ingenieur Technik published by Wiley-VCH GmbH

15222640, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cite.202100194 by Technische Universitaet Dortmund Dezernat Finanzen und Beschaffung, Wiley Online Library on [28/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

a)



b)

parameters	range
phenyl iodide (PI)	1,2,3,4,5 mM
phenylboronic acid (PBA)	1,2,3,4,5 mM
potassium fluoride (PF)	1,2,3,4,5 mM
reaction time	1,3,5 h
temperature	20,30,40 °C
tris(dibenzylideneacetone)dipalladium	0.156 mM
tricyclohexylphosphane	0.156 mM

c)

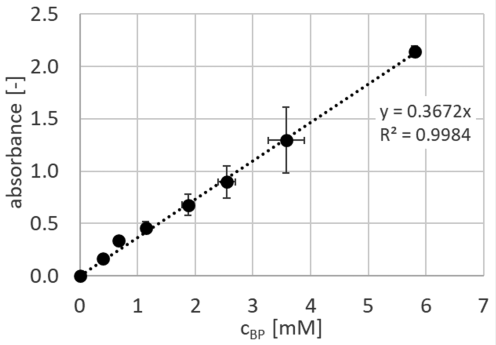


Figure 3. a) Reaction scheme for the performed Suzuki-Miyaura cross-coupling reaction, b) investigated reaction parameters regarding concentration of educts and catalyst, reaction temperature, and reaction time, c) calibration curve determined in triplicates for biphenyl in 4:1 THF/water mixture with R^2 = 0.9984 measured via UV/Vis spectroscopy at a wavelength of 249 nm.

software tool, the experimental plan was reduced to 75 experiments, which is a reduction of more than 93 %. Fig. 4 shows the results of the DoE tool, which represent different FFED with their calculated objective function. For this work, the optimum of 98 % and 75 experiments was chosen. As the objective function includes reduction of experiments and design efficiency equally weighted, the high percentage might result from the reduction of 93 % of the full-factorial experimental design. Defining the objective function as in Eq. (2), a high reduction of experiments might hide a low design efficiency of the reduced design. A distinction between number of trials and design efficiency for the reduced designs as objectives could be an alternative.

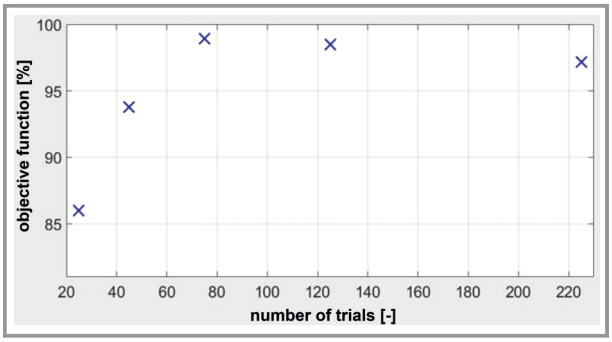


Figure 4. Calculated objective function in accordance with Eq. (2) for different fractional-factorial experimental designs in dependence on their number of trials, optimum for the objective function at 75 experiments and 98 %.

4 Results and Discussion

A first overview of the obtained results is shown in Fig. 5 with a comparison of the measured biphenyl concentration with the biphenyl concentration calculated by the model shown in Fig. 6a.

It is possible to preselect certain mathematical equations for the regression model to fit the data to known constants. But as mentioned before, the experiments should show whether an approximately minimal number of experiments could be sufficient to obtain reasonable and statistically valid results. For this reason, the composition of the model equation was chosen using the software tool to find a minimum model error. Comparison of the concentrations shows good agreement with the modeled data. The distinction of the data according to the reaction temperature reveals an underestimation by the model for high biphenyl concentrations for 30 °C and especially for 40 °C, whereas

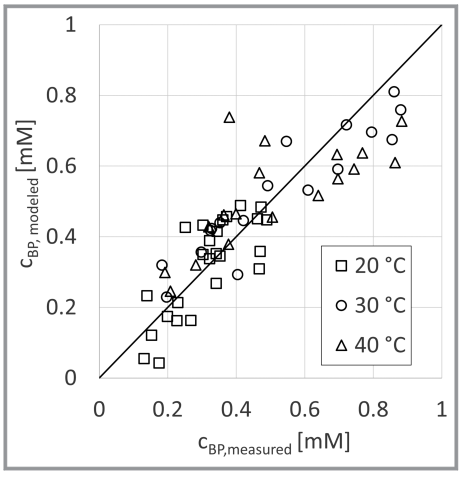


Figure 5. Parity plot for comparison of measured biphenyl concentrations with the calculated biphenyl concentrations by the DoE-obtained model for 20 °C, 30 °C, and 40 °C; bisector indicates perfect fit of the modeled to the measured data; obtained R^2 = 0.73.

 © 2022 The Authors. Chemie Ingenieur Technik published by Wiley-VCH GmbH

15222640, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cite.202100194 by Technische Universitaet Dortmund Dezernat Finanzen und Beschaffung, Wiley Online Library on [28/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the data points for 20 °C and 30 °C at lower concentrations are distributed equally around the angle bisector.

This might be caused by the fact that more data points are located at lower concentrations, which leads to a better fit of the model and a higher model error for higher biphenyl concentrations. The poorer fit of the data leads to the obtained coefficient of correlation of 0.73. The poorer fit at higher biphenyl concentration can also be caused by biphenyl produced by homocoupling of the educts instead of the intended C–C coupling. A more detailed insight into the data gives Fig. 6, in which contour plots are shown with the dependence of each process parameter in relation to the PBA concentration and the modeled biphenyl concentration. Since the data shown are derived from the two-dimensional analysis of the equation shown in Fig. 6a and the massive reduction of experiments, the results are interpreted more qualitatively. For visualization, the qualitative trend of the biphenyl concentration is indicated in every single graph. In general, it can be stated that increasing PI concentration, reaction time, and reaction temperature affect the product concentration positively, which corresponds with the Arrhenius equation and the reaction scheme.

For increasing PBA concentration, the positive effect of these three parameters is reduced, which is also reasonable as increasing PBA concentration accelerates the reaction rate and improves the conversion. The effect of PF differs in contrast to the other parameters when the biphenyl concentration drops with rising PF concentration. But this effect is compensated by an increased PBA concentration. The negative effect of a high base concentration was described by Lima et al. [12] before, who found an increased reactivity of the boronic acid at lower base concentration. As the model is only based on the results of 75 experiments, quantitative statements about the reaction mechanism or similar are not possible. The sensitivity of the results is high for small uncertainties in the input parameters due to the limited number of experiments, which causes high uncertainties in the obtained model. Especially the effect of time and temperature was not displayed accurately, because the variation range of these parameters was chosen too small. Nevertheless, qualitative trends as the negative effect of the base concentration could be identified, which gained important knowledge for further investigations and screenings.

5 Conclusion and Outlook

In this contribution, a newly developed DoE and regression software tool was applied on a Suzuki-Miyaura cross-coupling reaction to investigate the influence of the educt concentration, the base concentration, and the effect of time and temperature. With the intention to verify the software as a tool for fast screening experiments, a maximum reduction of experiments of more than 93 % was reached leading to 75 experiments. The reaction mixtures were prepared in a self-built robotic system in accordance with the experimental design.

The results show good agreement with the expected behavior of the reaction and provide a qualitative insight into the effects of the parameters studied. Increased residence time, temperature, and reactant concentrations lead to an accelerated reaction rate and higher conversion. In contrast to these effects, an increase in base concentration leads to a decrease in product concentration, as previously reported in the literature. The limitation of this approach was shown by the fact that the effects of time and temperature could not be correctly determined due to the small range of variation. Hence,

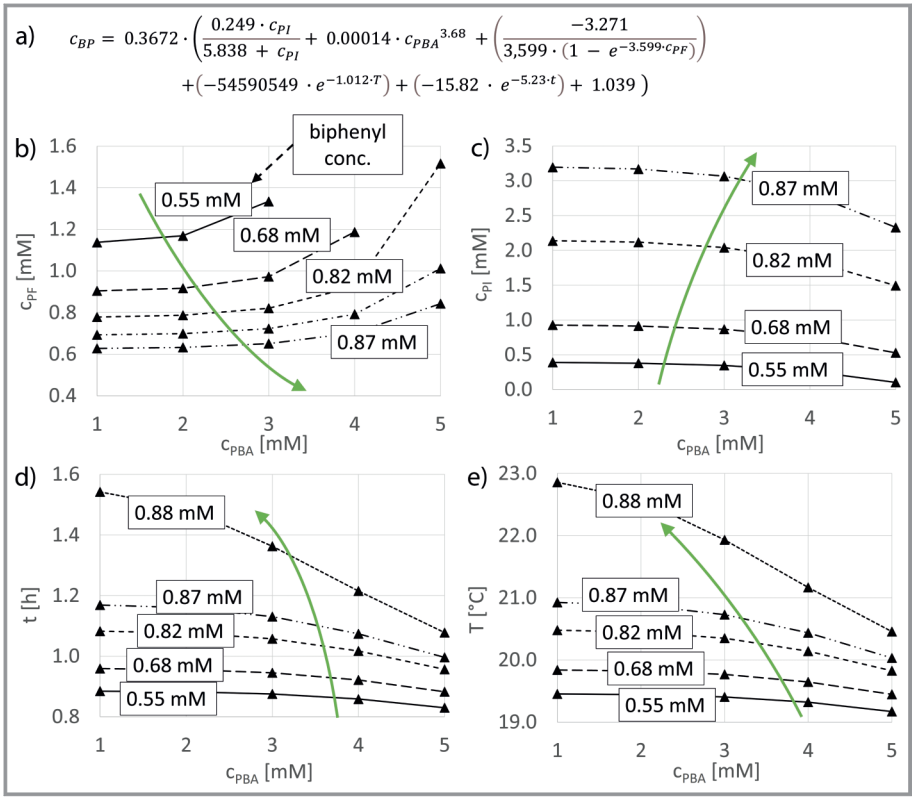


Figure 6. Two-dimensional contour plots for the derived multidimensional model for every varied process parameter related to the concentration of PBA as educt: a) derived model equation, b) PF concentration, c) phenyl iodide, d) reaction time, e) reaction temperature; values in boxes indicate the calculated resulting biphenyl concentration; arrows demonstrate the qualitative trend of the resulting biphenyl concentration calculated by the model.

 © 2022 The Authors. Chemie Ingenieur Technik published by Wiley-VCH GmbH

15222640, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cite.202100194 by Technische Universitaet Dortmund Dezernat Finanzen und Beschaffung, Wiley Online Library on [28/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

quantitative statements were not possible as the uncertainty of the derived model is too high regarding the limited data input and the impossibility to distinguish between biphenyl produced by homocoupling of the reactants and intended C–C coupling. For further investigations, reactants with suitable sidechains should be used to implement an analytical method for differentiation between side and main product.

Nevertheless, it was shown that screening of chemical reaction with the help of the software tool and in combination with lab automation can be rapidly performed to gain first insight into influencing effects. As an outlook, the software tool will be applied on a continuous automated screening system [13] compatible with the presented lab robot and a closed loop system with self-optimization should lead to a faster reaction optimization. Experiments with more data points should show the possibility to gain quantitative knowledge about kinetic data of chemical reactions.

The authors thank Marco Potowski of the Department of Chemistry and Chemical Biology for support and advice regarding the Suzuki-Miyaura reaction and Carsten Schrömges for technical support. Open access funding enabled and organized by Projekt DEAL.

Symbols used

c	[mM]	concentration
N	[–]	number of trials
p	[–]	number of input parameters
R^2	[–]	measure of determination
t	[h]	time
T	[°C]	temperature

Greek symbols

η	[%]	design efficiency
ϕ	[%]	objective function

Sub- and Superscripts

BP	biphenyl
D	D-optimal design
full	full-factorial design
measured	measured in performed experiments
min	minimum
modeled	calculated with regression model
PBA	phenylboronic acid
PF	potassium fluoride
PI	phenyl iodide

Abbreviations

ADoS	automated dosage system
AIU	automated injection unit
BP	biphenyl
DoE	design of experiments
OP	operation plates
PBA	phenylboronic acid
PCy_3	tricyclohexylphosphane
$Pd_2(dba)_3$	tris(dibenzylideneacetone)dipalladium
PF	potassium fluoride
PI	phenyl iodide
RMSE	root mean square error
THF	tetrahydrofuran

References

[1] J. A. Selekman, J. Qiu, K. Tran, J. Stevens, V. Rosso, E. Simmons, Y. Xiao, J. Janey, *Annu. Rev. Chem. Biomol. Eng.* **2017**, *8*, 525–547. DOI: https://doi.org/10.1146/annurev-chembioeng-060816-101411

[2] S. A. Pereira, P. J. Dyson, M. L. M. Saraiva, *TrAC, Trends Anal. Chem.* **2020**, *126*, 115862. DOI: https://doi.org/10.1016/j.trac.2020.115862

[3] S. K. Kurt, M. Akhtar, K. D. P. Nigam, N. Kockmann, *Ind. Eng. Chem. Res.* **2017**, *56 (39)*, 11320–11335. DOI: https://doi.org/10.1021/acs.iecr.7b02240

[4] F. Reichmann, F. Varel, N. Kockmann, *Processes* **2017**, *5 (4)*, 57. DOI: https://doi.org/10.3390/pr5040057

[5] V. Fath, N. Kockmann, J. Otto, T. Röder, *React. Chem. Eng.* **2020**, *5 (7)*, 1281–1299. DOI: https://doi.org/10.1039/D0RE00081G

[6] T. M. Gagliardi, R. Chelikani, Y. Yang, G. Tuozzolo, H. Yuan, *Biotechnol. Prog.* **2019**, *35 (4)*, e2811. DOI: https://doi.org/10.1002/btpr.2811

[7] C. Nunn, A. DiPietro, N. Hodnett, P. Sun, K. M. Wells, *Org. Process Res. Dev.* **2018**, *22 (1)*, 54–61. DOI: https://doi.org/10.1021/acs.oprd.7b00295

[8] S. S. Henley, R. M. Golden, T. M. Kashner, *Biostat. Epidemiol.* **2020**, *4 (1)*, 105–139. DOI: https://doi.org/10.1080/24709360.2019.1618653

[9] V. Rosso, J. Albrecht, F. Roberts, J. M. Janey, *React. Chem. Eng.* **2019**, *4 (9)*, 1646–1657. DOI: https://doi.org/10.1039/C9RE00079H

[10] J. Bobers, M. K. Škopić, R. Dinter, P. Sakthithasan, L. Neukirch, C. Gramse, R. Weberskirch, A. Brunschweiger, N. Kockmann, *ACS Comb. Sci* **2020**, *22 (3)*, 101–108. DOI: https://doi.org/10.1021/acscombsci.9b00207

[11] A. F. Littke, G. C. Fu, *Angew. Chem.* **2002**, *114 (22)*, 4350–4386. DOI: https://doi.org/10.1002/1521-3757(20021115)114:22<4350:AID-ANGE4350>3.0.CO;2-0

[12] C. F. R. A. C. Lima, A. S. M. C. Rodrigues, V. L. M. Silva, A. M. S. Silva, L. M. N. B. F. Santos, *ChemCatChem* **2014**, *6 (5)*, 1291–1302. DOI: https://doi.org/10.1002/cctc.201301080

[13] J. Bobers, J. Grühn, S. Höving, T. Pyka, N. Kockmann, *Org. Process Res. Dev.* **2020**, *24 (10)*, 2094–2104. DOI: https://doi.org/10.1021/acs.oprd.0c00152

 © 2022 The Authors. Chemie Ingenieur Technik published by Wiley-VCH GmbH

15222640, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cite.202100194 by Technische Universitaet Dortmund Dezernat Finanzen und Beschaffung, Wiley Online Library on [28/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License