Chemical Engineering Journal 451 (2023) 138443

Contents lists available at ScienceDirect

Chemical Engineering Journal

journal homepage: www.elsevier.com/locate/cej

Machine learning directed multi-objective optimization of mixed variable chemical systems

Oliver J. Kershaw [a], Adam D. Clayton [a], Jamie A. Manson [a], Alexandre Barthelme [b], John Pavey [b], Philip Peach [c], Jason Mustakis [d], Roger M. Howard [d], Thomas W. Chamberlain [a], Nicholas J. Warren [a], Richard A. Bourne [a,*]

[a] Institute of Process Research and Development, School of Chemistry & School of Chemical and Process Engineering, University of Leeds, Leeds, LS2 9JT, UK
[b] UCB Pharma SA, All. de la Recherche 60, 1070 Anderlecht, Belgium
[c] Department of Chemical Research and Development, Pfizer Limited, Ramsgate Road, Sandwich CT13 9NJ, UK
[d] Pfizer Worldwide Research and Development, Groton, CT, United States

ARTICLE INFO

Keywords:
Automated flow reactor
Mixed variable optimization
Multi-objective
Machine learning
Reaction engineering

ABSTRACT

The consideration of discrete variables (e.g. catalyst, ligand, solvent) in experimental self-optimization approaches remains a significant challenge. Herein we report the application of a new mixed variable multi-objective optimization (MVMOO) algorithm for the self-optimization of chemical reactions. Coupling of the MVMOO algorithm with an automated continuous flow platform enabled identification of the trade-off curves for different performance criteria by optimizing the continuous and discrete variables concurrently. This approach utilizes a Bayesian methodology to provide high optimization efficiency, enhances process understanding by considering key interactions between the mixed variables, and requires no prior knowledge of the reaction. Nucleophilic aromatic substitution (S_NAr) and palladium catalyzed Sonogashira reactions were investigated, where the effect of solvent and ligand selection on the regioselectivity and process efficiency were determined respectively whilst simultaneously determining the optimum continuous parameters in each case.

1. Introduction

Machine learning is becoming increasingly pervasive in all chemistry research and development activities, from molecular discovery through to process optimization and chemical manufacturing. [1–3] This has driven the emergence of new and challenging molecular targets with greater three-dimensionality, which requires the exploitation of a wider and more complex reaction toolkit. [4] Subsequent development of these more intricate processes is a challenging and expensive task, often requiring the optimization of both continuous and discrete variables. Traditionally, high-throughput screening (HTS) methodologies are used to screen different combinations of discrete variables (catalyst, ligand, solvent), prior to further optimization of the continuous variables (reaction time, temperature, stoichiometry). [5–8] However, faced with increasingly complex synthetic methodologies, the laborious 'brute-force' approach to HTS (i.e., testing all possible combinations of discrete variables) is becoming increasingly unattractive. In addition, the workflow of optimizing discrete and continuous variables in a sequential manner results in an incomplete process understanding, as key interactions between the mixed variables are not considered (Fig. 1). For example, the effects of varying temperature on the activity of different catalysts would be omitted.

An attractive solution to these limitations would be to simultaneously optimize the mixed variables of chemical reactions algorithmically, by coupling automated reactor platforms with machine learning to intelligently explore the multivariate reaction landscape. [9] This self-optimization technology has been shown to greatly accelerate process development, yet current systems are largely limited to the optimization of continuous variables, owing to a lack of advanced algorithms available for mixed variable problems with expensive-to-evaluate objectives. [10–26] Warren et al. explored the multi-objective optimization of molar mass dispersity and monomer conversion for a series of RAFT polymerizations, using a Thompson sampling-based algorithm (TSEMO). [27] However, individual optimizations were required for each RAFT agent, which resulted in high data density in non-optimal regions of the reaction space. Similarly, Lapkin et al.

* Corresponding author.
E-mail address: r.a.bourne@leeds.ac.uk (R.A. Bourne).

https://doi.org/10.1016/j.cej.2022.138443
Received 6 June 2022; Received in revised form 29 July 2022; Accepted 30 July 2022
Available online 2 August 2022

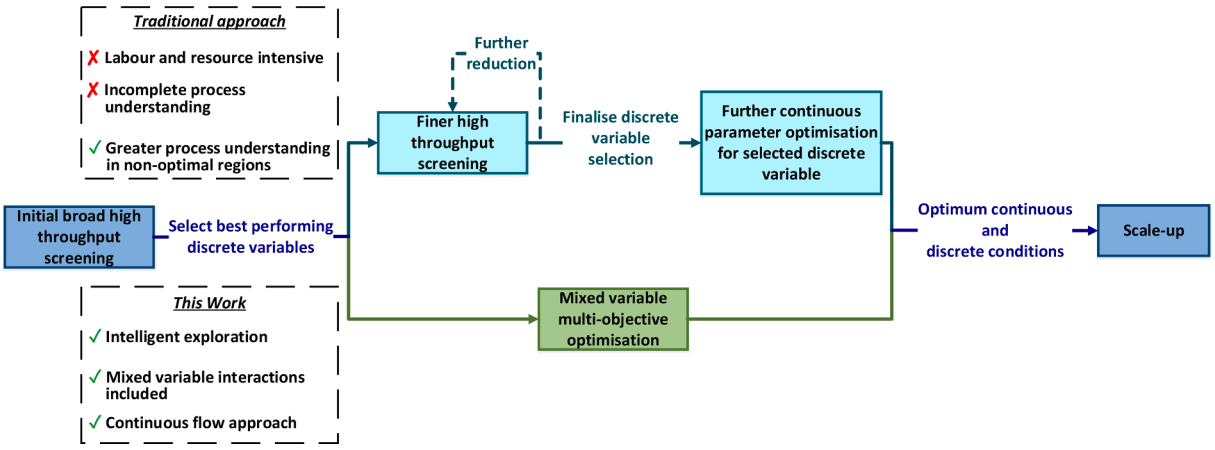
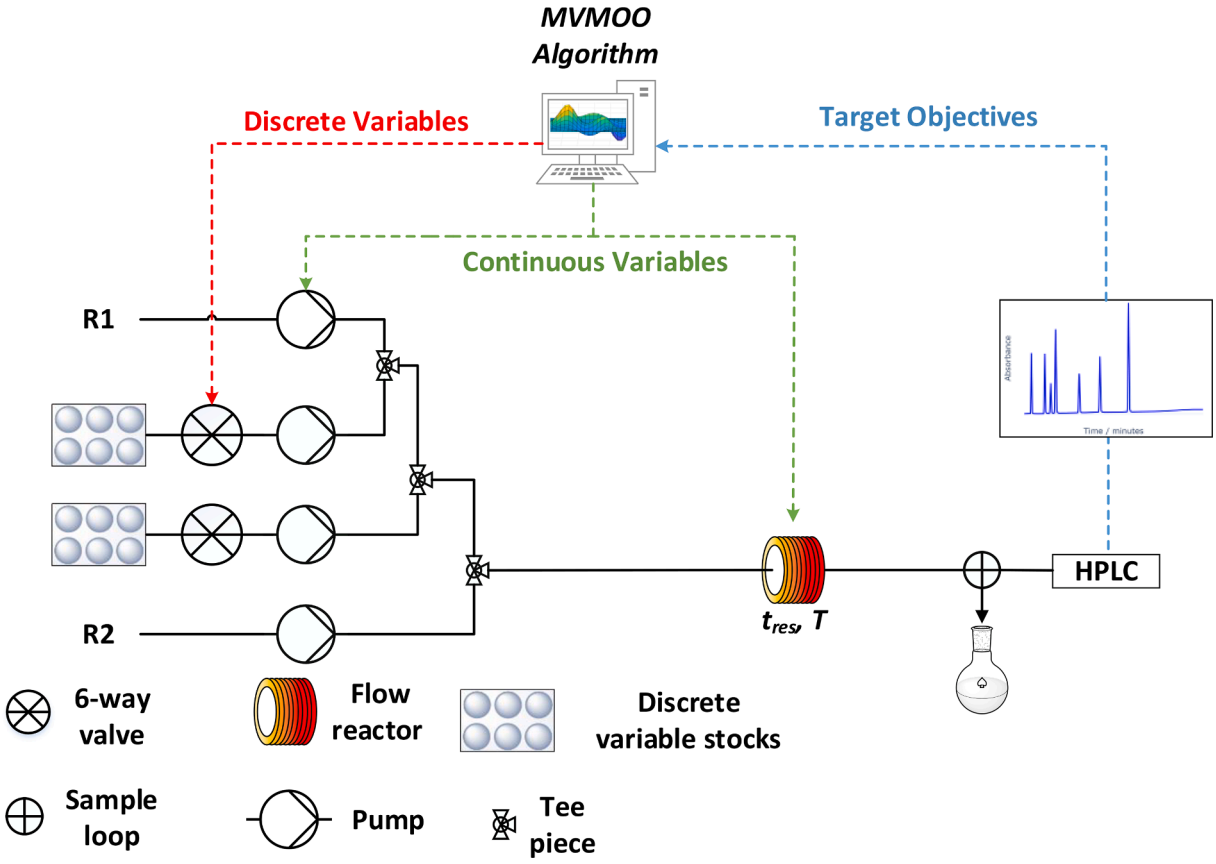


Fig. 1. Comparison of full reaction optimization using traditional HTS methodologies versus newly developed MVMOO self-optimization approach.

explored the multi-objective optimization of a Rhodium catalyzed asymmetric hydrogenation, but where solvent and temperature were simultaneously optimized. [28] This was achieved by using molecular descriptors to define the candidate solvents, thereby converting the discrete variable to the continuous domain. However, selection of the molecular descriptors which effectively account for the observed behavior can require extensive prior knowledge of the system, which is difficult to obtain for novel reactions. For example, correlations between the bite angle of diphosphine complexes and selectivity in catalytic reactions. [29]

In contrast, black-box optimization methods bypass the requirement for physical knowledge of the discrete variables. [30] Jensen et al. utilized an optimal DoE-based algorithm to optimize a series of mixed variable transition metal catalyzed cross-coupling reactions. [31–33] This approach used the adaptive response surface method to iteratively eliminate catalytic species from the optimization. However, as a single-objective optimization method, it does not provide insight into the trade-off (Pareto front) between conflicting performance criteria, which is crucial in the development of viable industrial processes. We have previously demonstrated the importance of identifying the Pareto front between economic and environmental objectives during self-optimization workflows. [34–36] As such, there remains a need to develop efficient multi-objective algorithms which include discrete variables within the optimization domain.



Scheme 1. Generalized flow diagram for automated mixed variable multi-objective reaction optimization. The reactor configuration can be modified for different mixed variable systems via selection of desired 6-way valves/pumps.

Chemical reaction optimizations are inherently expensive-to-evaluate problems due to the requirement of conducting physical experiments, which can be time consuming and resource intensive. In this work, we use our recently developed mixed variable multi-objective optimization (MVMOO) algorithm, which has been shown to perform competitively (i.e. reduced experimental budget) for a series of *in silico* test problems when compared to alternative methods, including a mixed-variable implementation of NSGA-II. [37]

An open-source implementation of MVMOO is available on GitHub (https://github.com/jmanson377/MVMOO). This study identifies key interactions between mixed variables whilst highlighting the complete trade-off between competing objectives, which maximizes process understanding for accelerating the development of chemical processes. Although the simultaneous optimisation of categorical and continuous variables has previously been achieved using HTS batch systems, [38] continuous flow systems offer additional benefits, including access to higher temperatures and hazardous intermediates during these optimizations. Herein, the MVMOO algorithm was combined with an automated continuous reaction system, and evaluated using two case studies: (i) a S_NAr reaction with solvent dependent regioselectivity; [39] (ii) a Sonogashira reaction as a key step in the synthesis of a TRPV1 receptor antagonist. [40]

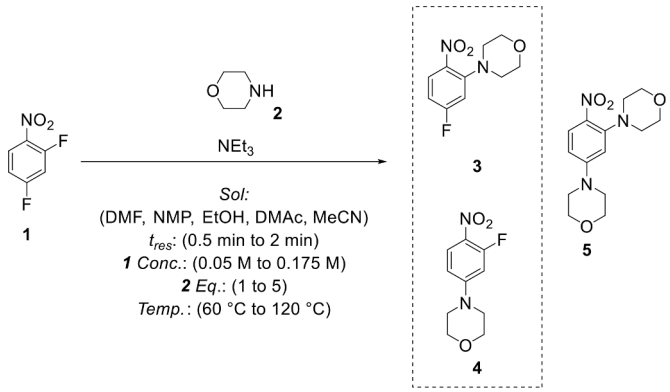
2. Materials and methods

2.1. Chemicals

All of the following compounds were purchased from suppliers and used without further purification. 1-Methyl-2-pyrrolidone (NMP; 99%), dimethylformamide (DMF; Extra Pure), dimethylacetamide (DMAc; 99+%), ethanol (EtOH; 99.8%), acetonitrile (MeCN; HPLC grade), morpholine (99+%) and pyrrolidine (99%) were purchased from Fisher Scientific ltd. 2,4-Difluoronitrobenzene (99%), 2-bromo-4-(trifluoromethyl)benzonitrile (95%), 2-dicyclohexylphosphino-2′-(*N,N*-dimethylamino)biphenyl (DavePhos; 98%), 2-dicyclohexylphosphino-2′.4′.6′-triisopropylbiphenyl (XPhos; 98%), (2-biphenyl)dicyclohexylphosphine (CyJohnPhos; 98%) and 2–dicyclohexylphosphino-2′,6′-dimethoxybiphenyl (SPhos; 98%) were purchased from Fluorochem. Triphenylphosphine (TPP; flake, 99%) was purchased from Alfa Aesar. 3,3-Dimethyl-1-butyne (98%), 1,3,5-trimethoxybenzene (99+%), biphenyl (99.5% GC), copper (I) iodide (98%) and palladium (II) acetate (99+%) were purchased from Merck Life Science UK ltd. Triethylamine (99%) was purchased from Acros Organics.

2.2. MVMOO algorithm

The MVMOO algorithm was integrated with the automated flow reactor to enable closed-loop mixed variable multi-objective optimization of chemical reactions (Scheme 1). [37] The MVMOO algorithm is initialized with a space filling design, using five identical Latin hypercube (LHC) sampling points per discrete variable, to provide sufficient exploratory information for use with the iterative process models. [41] Within the algorithm, individual Gaussian processes (GPs) are utilized as surrogate models for each objective. To enable modelling of mixed variables, the GP surrogates use an internal distance metric based on Gower similarity, which permits the use of pre-existing covariance functions such as the mixed Matérn 5/2 kernel. [42] The hyperparameters are optimized by maximizing the log marginal likelihood of the current data using GPflow's internal Adam optimizer. [43] The kernel lengthscales are learned hyperparameters which correlate to the relative importance of each input variable, where lower values indicate a greater contribution to the objective. The next set of reaction conditions are determined via internal optimization of the expected improvement matrix (EIM) acquisition function with a Euclidean based transformation. [44] The optimization of the EIM occurs in two stages: (i) a large sample of the EIM is taken using a Halton sequence for each discrete variable combination; (ii) the leading variable combination undergoes additional local optimized using SciPy's implementation of sequential least squares programming (SLSQP). [45] After conducting the suggested experiment, the GPs are updated, and the process is repeated iteratively for a desired number of function evaluations. This approach balances exploration and exploitation to identify the global Pareto front.



Scheme 2. S_NAr reaction between 2,4-difluoronitrobenzene **1** and morpholine **2**, forming *ortho*-**3** and *para*-**4** regioisomers, and bis adduct **5**. Includes optimization parameters.

2.3. Automated reaction platform

Reagents were pumped using JASCO PU-2080 dual piston HPLC pumps and flows were mixed using Swagelok SS-100–3 tee-pieces. Reactors of a desired volume were made from PTFE tubing (1/16″ OD, 1/32″ ID), which were fitted to a cylindrical aluminium block and heated with a Eurotherm 3200 temperature controller. Solvent and ligand selection was achieved using a JASCO CO4062 column oven module as a 6-way switching valve. Sampling was achieved using a VICI Valco EUDA-CI4W sample loop (4-port) with 0.5 μL injection volume. The reactor was maintained under the desired fixed back pressure using an Upchurch Scientific back pressure regulator (250 psi). Quantitative analysis was performed on an Agilent 1260 Infinity II series HPLC instrument fitted with an Agilent Poroshell 120 EC-C18 reverse phase column (5 cm length, 4.6 mm ID and 2.7 μm particle size). The automated system was controlled using a custom written MATLAB program, and the MVMOO algorithm was written and implemented in Python.

2.4. Optimization procedure

A MATLAB script controlled the pump flow rates, valve positions, reactor temperature and sampling. For each iteration valve positions were set to the corresponding discrete variable; the reactor was allowed to stabilize at the desired operating temperature; the pumps were set to the required flow rates and left for three reactor volumes to reach steady state; and finally, the sampling valve was triggered alongside HPLC analysis. To minimize the duration and material consumption per iteration: (i) pump flow rates were reduced to a minimum during heating/cooling of the reactor; (ii) initial LHC experiments were sorted in order of increasing temperature; (iii) sequential LHC experiments were started whilst analysis of the previous experiment was running. Responses for each objective were calculated from HPLC chromatograms obtained after each iteration and used to update the surrogate models and generate the next set of reaction conditions using the MVMOO algorithm. In each case, the hypervolume was monitored after 60 experiments, and the optimizations terminated when a significant plateau was observed (see ESI for details).

Table 1
Values of common solvent polarity metrics for evaluation as chemical descriptors.

Solvent	Polarity Index[a]	Dipole Moment	Dielectric Constant
DMF	6.4	3.86	36.71
NMP	6.7	4.09	32.20
EtOH	5.2	1.66	24.55
DMAc	6.5	3.72	37.78
MeCN	5.8	3.44	37.50

[a] Taken from Snyder *et al.* [46].

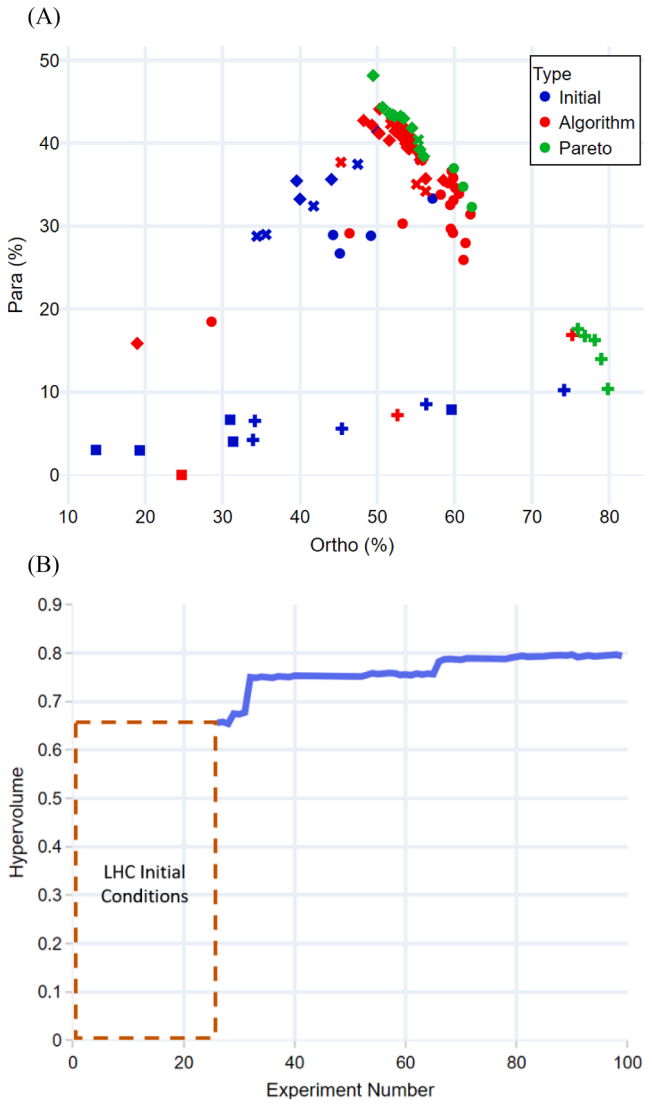


Fig. 2. (A). Results of the five-parameter mixed variable multi-objective optimization of the S_NAr reaction. For each solvent, 5 initial LHC points were collected. The MVMOO algorithm conducted 74 additional experiments, 20 of which formed a Pareto front highlighting the trade-off in regioselectivity between the *ortho*-**3** and *para*-**4** products. Solvent shapes represent: ● – DMF, ◆ – NMP, ■ – EtOH, + – MeCN, × – DMAc (B). Calculated hypervolume vs experiment number graph to visualize the hypervolume increase over the optimization period for the S_NAr.

3. Results and discussion

3.1. Nucleophilic aromatic substitution (S_NAr)

Nucleophilic aromatic substitution (S_NAr) reactions provide a convenient method for forming aromatic carbon-heteroatom bonds. The regioselectivity of S_NAr reactions using 2,4-dihaloaromatic compounds are known to be greatly influenced by solvent effects. In addition, the ability to selectively synthesize different regioisomers is key for the preparation of desired compounds, as well as their corresponding by-products for impurity profiling. Therefore, we decided to investigate the S_NAr reaction between 2,4-difluoronitrobenzene **1** and morpholine **2**, simultaneously optimizing for the formation of both the *ortho*-isomer **3** and *para*-isomer **4** (Scheme 2). In contrast to previous S_NAr self-optimizations, solvent choice was included as a discrete variable, alongside continuous variables comprising residence time, concentration, equivalents, and temperature. Solvents were selected based on their ability to provide a homogeneous reaction mixture, as well as their different solvent polarities (see Table 1), which is a factor known to influence the regioselectivity of the reaction. The MVMOO algorithm was initialized with five LHC experiments per solvent (25 experiments), and then allowed to run for 74 sequential iterations. Of these experiments, there were 20 non-dominated solutions identified, which highlight the trade-off between formation of the *ortho*-**3** and *para*-**4** regioisomers (Fig. 2). The optimal *ortho*-**3** yield was 80% with a corresponding *para*-**4** yield of 10%. Conversely, the optimal *para*-**4** yield was 48% with a corresponding *ortho*-**3** yield of 49%.

In this case, the Pareto front is divided into three discontinuous sections as a result of different solvent effects. MeCN and NMP afford the highest yields of *ortho*-**3** and *para*-**4** respectively, whereas DMF provides a moderate compromise. Despite the structural similarity between DMF and DMAc, DMAc gave higher *para*-**4** selectivity, but was dominated by NMP in all but one experiment. In contrast, EtOH gives high *ortho*-**3** selectivity, but suffers from significantly lower overall yields in the optimization domain. This resulted in the MVMOO algorithm only suggesting one EtOH-based experiment for exploration beyond the initial LHC.

Notably, the MVMOO algorithm was able to map the relative importance of the variables at each discrete level across the Pareto front (Fig. 3). When using NMP, a relatively high *para*-**4** yield can be achieved across a wide range of continuous variable combinations, indicating that the solvent effect is the dominant factor. In contrast, when using MeCN, a high residence time, equivalents and temperature are also required to maximize the *ortho*-**3** yield, whereas concentration has lesser importance. Similarly, the moderate compromise achieved using DMF requires a high temperature, equivalents and concentration, combined with moderate temperatures. Importantly, traditional optimization workflows, where discrete and continuous variables are optimized sequentially, would have falsely assumed that the interactions between the continuous variables and each solvent were the same. In contrast, the MVMOO algorithm used in this approach successfully optimized the mixed variables simultaneously, thus providing greater understanding of the trade-off in regioselectivity. In addition, the optimization was completed in just over 18 h with no human intervention, representing an increase in efficiency compared to iterative HTS campaigns, or individual optimizations for each discrete variable. [27]

Given the diverse dataset provided by the optimization, an analysis of which underlying chemical descriptors effectively model the observed behavior was investigated (see ESI for details). Three common solvent polarity metrics (Table 1) were examined: (i) polarity index, a measure

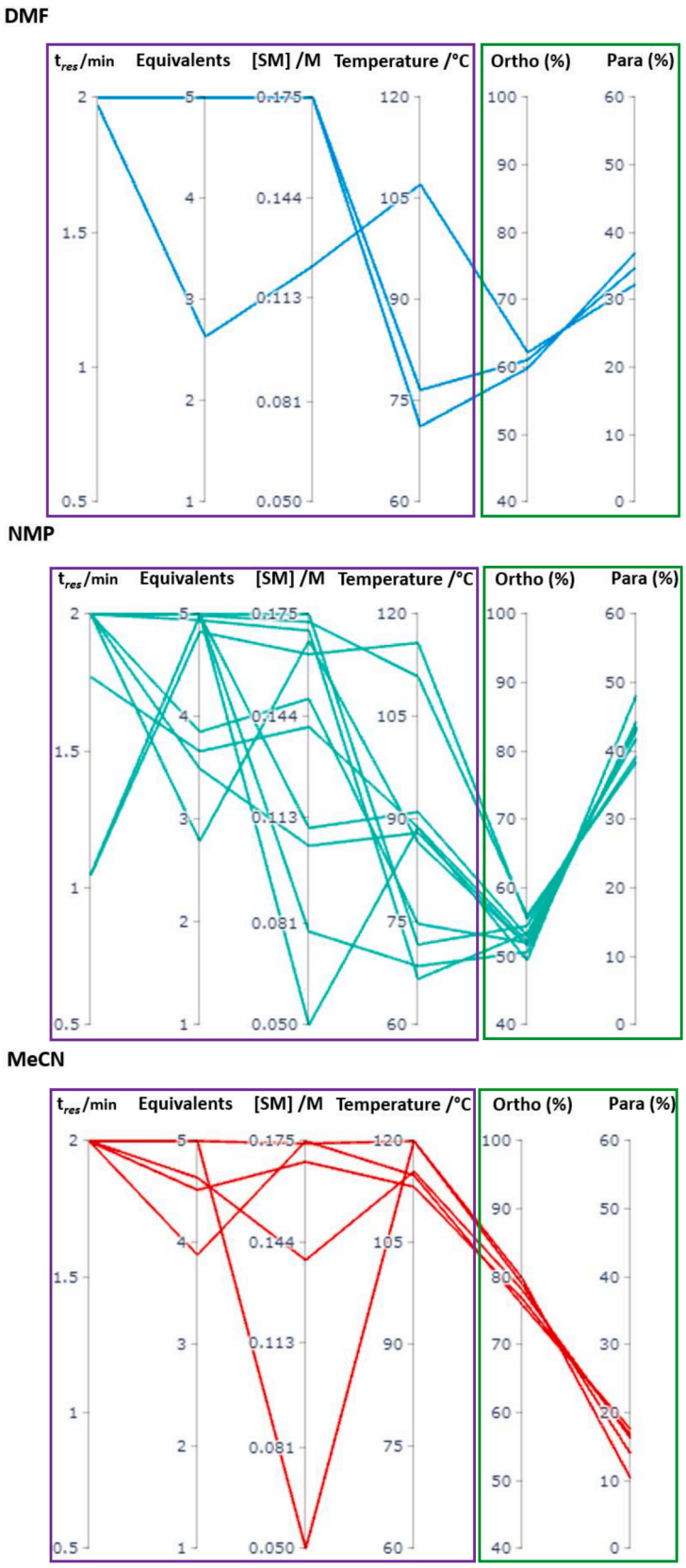


Fig. 3. Parallel coordinate plots showing the interactions between the variables for each Pareto optimal solution of the S_NAr reaction. Each line represents a single non-dominated solution for the corresponding solvent. DMAc also achieved a Pareto point (see ESI) which achieved an *ortho* yield of 55.3% and *para* yield of 40.4% requiring: t_{res} = 2 min, equivalents = 5, SM concentration = 0.175 M and Temperature = 90 °C.

of a solvents' ability to interact with various polar test solutes; (ii) dipole moment, the product of the magnitude of separated charges and the distance between them; (iii) dielectric constant, a measure of a substance's ability to insulate charges from each other. Overall, a bias towards *ortho*-substitution was observed, likely caused by hydrogen bonding between the nucleophile and the neighboring nitro group. [47] The relationship deduced by the models, in which increasing the polarity index of the solvent results in an increase in *para*-**4** yield, correlates well with previous studies. [48] In contrast, selecting either dipole moment or dielectric constant fails to completely describe the observed relationship, with dielectric constant predicting MeCN to behave similarly to DMAc, which was not observed experimentally. These findings highlight the importance that chemical descriptor selection can have on a model's ability to accurately map the input domain to the objective space. It is worth noting that, although this method enabled the identification of optimal solvents from a small selection, a significantly larger dataset would be required to truly study the key solvent descriptors in more detail.

3.2. Sonogashira cross-coupling

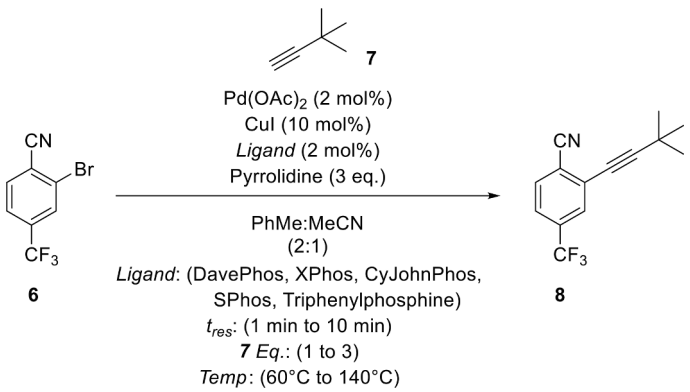
Sonogashira cross-coupling reactions provide a convenient method for the formation of carbon–carbon bonds between a vinyl or aryl halide and a terminal alkyne, generally performed using a palladium catalyst, copper co-catalyst, phosphine ligand and amine base. There are a large number of factors which can influence the performance of catalytic reactions (e.g. sterics, solvent, ligand, heteroatoms, base, temperature, metal source), resulting in substantial optimization often being needed for different pairs of substrates. The Jensen group have previously optimized Suzuki-Miyaura cross-coupling reactions using a droplet-flow microfluidic system combined with a MINLP optimization approach. However, this method was limited to the optimization of a single objective, which restricts the amount of process understanding that can be achieved. [30,31] As such, an efficient experimental approach for multi-objective optimization of mixed variable catalytic systems on a substrate-by-substrate basis would be desirable. Therefore, we decided to investigate a pharmaceutically relevant Sonogashira cross-coupling reaction, utilizing the developed MVMOO self-optimization approach. The Sonogashira cross-coupling reaction between aryl bromide **6** and terminal alkyne **7** to form aryl alkyne **8** was investigated (Scheme 3).

This is a modified step in the synthesis of a TPRV1 receptor antagonist, used primarily for the management and treatment of pain. [40,49] The following modifications were made to the original reported synthesis to enable optimization using our continuous flow platform: (i) the aryl chloride was replaced with the corresponding aryl bromide to increase the rate of reaction; (ii) Pd_2dba_3 was replaced with the more stable $PdOAc_2$ precatalyst; (iii) NEt_3 was replaced with a PhMe:MeCN (2:1) solvent mixture and pyrrolidine base, providing a homogeneous reaction mixture. In addition, reservoirs were prepared and stored under nitrogen for the duration of the optimization, overcoming the experimental challenge of metal–ligand complex stabilization.

As this represents an industrially relevant process, the productivity (space–time yield, STY) and environmental impact (reaction mass efficiency, RME) were selected as optimization objectives to identify viable operating conditions. Owing to the crucial role of ligands in catalyzed reactions, the choice of phosphine ligand was included as a discrete variable, alongside residence time, equivalents, and temperature. A range of monodentate phosphine ligands were selected for the optimization due to their high activity in palladium-catalyzed cross-coupling reactions, good solubility in organic solvents and commercial availability. In addition, ligands with different cone angles were chosen to assess the effect of steric bulk on the outcome of the reaction. The MVMOO algorithm was initialized with five LHC experiments per ligand (25 experiments), and then allowed to run for 44 sequential iterations. Of these experiments, there were 12 non-dominated solutions identified, which highlight the trade-off between STY and RME (Fig. 4). The optimal STY was 322.0 kg m^{-3} h^{-1} with a corresponding RME of 51.5. Conversely, the optimal RME was 68.2 with a corresponding STY of 32.31 kg m^{-3} h^{-1}. In this case, the Pareto front highlights a steep linear trade-off, where the STY can be significantly improved whilst having little detrimental effect on the RME. Notably, an excellent in-situ yield of 90% was achieved at the conditions corresponding to the optimum RME.

The results from the initial LHC were in the region of objective space corresponding to a range of RMEs with low STYs. Notably, for each LHC experiment, triphenylphosphine was found to give the highest STY and RME compared to all other ligands, resulting in the MVMOO algorithm predominantly suggesting triphenylphosphine-based experiments for subsequent iterations. In general, Sonogashira reactions involving deactivated aryl halides favor the use of sterically demanding and electron-rich phosphine ligands. [50] The cone angle can be used as a measure of the steric bulk of ligands. Although values reported in literature vary, the general trend in cone angles for the monodentate phosphine ligands used in this work is: TPP < DavePhos, XPhos < CyJohnPhos < SPhos. Therefore, the results from this optimization are counterintuitive based on current chemical understanding. This highlights the usefulness of efficient experimental optimization, particularly for complex reactions involving novel substrate pairs, where interactions are not fully understood or easily predicted.

Despite localization of the LHC on regions of low STY, the MVMOO algorithm was successful in identifying new regions of objective space corresponding to more productive reactions. The MVMOO algorithm demonstrated its exploratory ability by conducting a small number of SPhos-based experiments, which was identified as the second most favorable ligand during the initial LHC. Notably, the optimum STY was achieved with SPhos at low residence times, high equivalents and high temperatures (Fig. 5). Subsequently, the MVMOO algorithm conducted a direct comparison of these continuous conditions using triphenylphosphine, which resulted in an inferior STY and RME as a result of a



Scheme 3. Sonogashira cross-coupling reaction between aryl bromide **6** and terminal alkyne **7** to form aryl alkyne **8**. Includes optimization parameters.

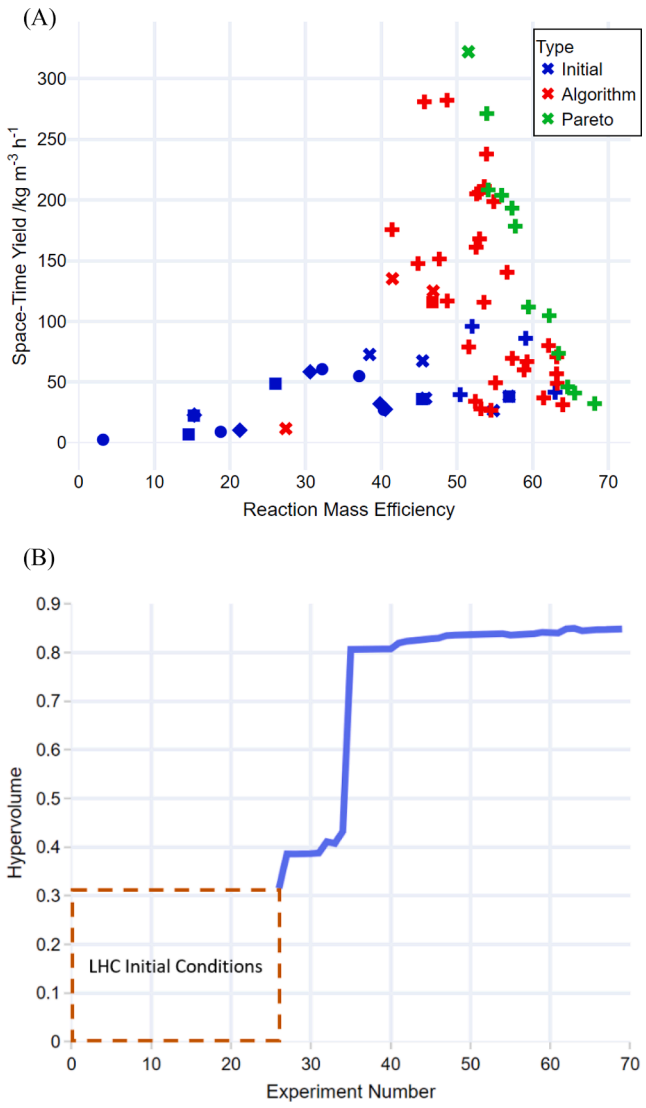


Fig. 4. (A). Results of the four-parameter mixed variable multi-objective optimization of the Sonogashira cross-coupling reaction. For each ligand, 5 initial LHC points were collected. The MVMOO algorithm conducted 44 additional experiments, 12 of which formed a Pareto front highlighting the trade-off in STY and RME. Ligand shapes represent: ● – DavePhos, ◆ – XPhos, ■ – CyJohnPhos, × – SPhos, + – TPP. (B). Calculated hypervolume vs experiment number for the Sonogashira optimization.

lower yield. Rather, triphenylphosphine favored the use of low to moderate equivalents, combined with high residence times and low temperatures to achieve high RMEs, or low residence times and high temperatures to achieve high STYs. Similar to the S_NAr case study, the interactions between the continuous variables and each ligand were found to be significantly different, highlighting simultaneous optimization of mixed variables to be a superior approach for finding the true optima compared to traditional sequential optimization.

The ability to optimize complex mixed variable catalytic reactions for different substrate pairs offers an opportunity to explore new regions of chemical space, provided that the method is not experimentally laborious. In this case, the developed MVMOO self-optimization approach was able to autonomously optimize a four parameter, mixed variable Sonogashira cross-coupling reaction, with respect to two objectives in just 69 experiments over a 22-hour period, with no human intervention beyond the preparation of reagent solutions. This is comparable to previously reported experimental optimizations of single objective mixed variable systems and multi-objective continuous variable systems. [31,35] As such, we envisage that this approach will be functional in expanding the reaction toolkit available for synthetic and process chemists.

4. Conclusion

In conclusion, we have demonstrated the application of a mixed variable multi-objective optimization algorithm for the autonomous development of chemical reactions. This approach was used to optimize the regioselectivity of a S_NAr reaction, and the efficiency of a pharmaceutically relevant Sonogashira reaction, including the solvent and ligand effects respectively. For the S_NAr reaction, the trade-off between the *ortho* and *para* regioisomers was identified, and the corresponding variable effects elucidated. Notably, the polarity index of the solvent was found to be a useful chemical descriptor for S_NAr reactions between 2,4-dihalonitrobenzenes and amines. For the Sonogashira reaction, the trade-off between the productivity and environmental impact was identified, which included screening a selection of monodentate phosphine ligands. In contrast to current chemical understanding, the least sterically demanding ligand, triphenylphosphine, was found to be the best performing ligand. This highlights that the results of complex catalytic reactions with different and/or novel substrate pairs can be nonintuitive, and therefore warrant experimental optimization. The MVMOO algorithm utilized in this work provides an efficient method for achieving this, optimizing the two mixed variable reactions with respect

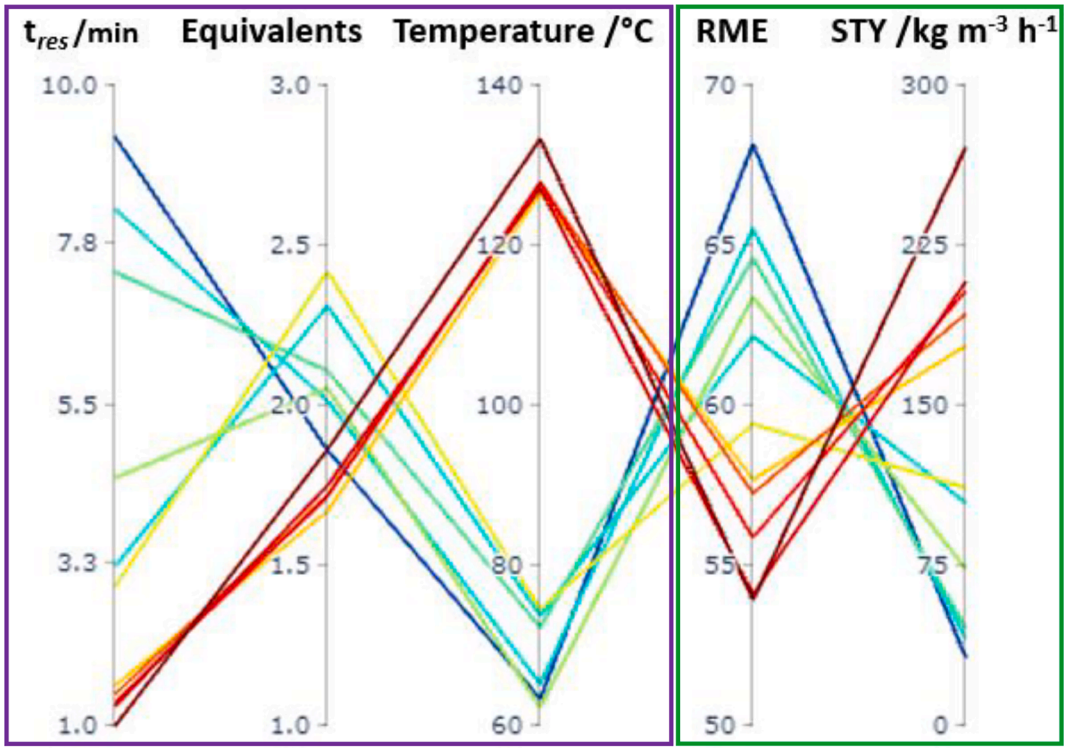


Fig. 5. Parallel coordinate plots showing the interactions between the variables for each Pareto optimal solution of the Sonogashira reaction. Each line represents a single non-dominated solution for the TPP ligand points. Line colour is scaled with regards to STY weighting to aid in visualization (high STY/low RME = –, low STY/high RME = –). SPhos achieved a Pareto point (see ESI) which achieved a STY of 322 kg m^{-3} h^{-1} and RME of 68.19 requiring: t_{res} = 1 min, equivalents = 3 and Temperature = 140 °C.

to two objectives each in a combined time of just 40 h. In addition, this approach simultaneously optimizes discrete and continuous variables, which enables the identification of key interactions that would otherwise be missed with traditional sequential optimization. In both case studies herein, the interactions between the continuous variables and each discrete variable were found to be significantly different i.e., different solvents and ligands favored different combinations of continuous conditions. Integration of this optimization approach with a segmented flow microfluidic system for more resource efficient process development is currently under investigation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

We have shared a link to the algorithm code within the manuscript

Acknowledgments

OK thanks EPSRC, Pfizer Inc. and University of Leeds for funding. ADC thanks UCB and University of Leeds for funding. RAB was supported by the Royal Academy of Engineering under the Research Chairs and Senior Research Fellowships scheme. This work was funded, in part, by the EPSRC project "Cognitive Chemical Manufacturing" EP/R032807/1.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cej.2022.138443.

References

[1] D. Caramelli, J.M. Granda, S.H.M. Mehr, D. Cambié, A.B. Henson, L. Cronin, Discovering New Chemistry with an Autonomous Robotic Platform Driven by a Reactivity-Seeking Neural Network, ACS Cent. Sci. 7 (2021) 1821–1830, https://doi.org/10.1021/acscentsci.1c00435.
[2] D.E. Graff, E.I. Shakhnovich, C.W. Coley, Accelerating high-throughput virtual screening through molecular pool-based active learning, Chem. Sci. 12 (2021) 7866–7881, https://doi.org/10.1039/d0sc06805e.
[3] C.W. Coley, W.H. Green, K.F. Jensen, Machine Learning in Computer-Aided Synthesis Planning, Acc. Chem. Res. 51 (2018) 1281–1289, https://doi.org/10.1021/acs.accounts.8b00087.
[4] S. Chow, S. Liver, A. Nelson, Streamlining bioactive molecular discovery through integration and automation, Nat. Rev. Chem. 2 (2018) 174–183, https://doi.org/10.1038/s41570-018-0025-7.
[5] D. Perera, J.W. Tucker, S. Brahmbhatt, C.J. Helal, A. Chong, W. Farrell, P. Richardson, N.W. Sach, A platform for automated nanomole-scale reaction screening and micromole-scale synthesis in flow, Science 359 (6374) (2018) 429–434.
[6] R.R. Steimbach, P. Kollmus, M. Santagostino, A. Validated, "Pool and Split" Approach to Screening and Optimization of Copper-Catalyzed C-N Cross-Coupling Reactions, J. Org. Chem. (2020), https://doi.org/10.1021/acs.joc.0c02392.
[7] E. Brouzes, M. Medkova, N. Savenelli, D. Marran, M. Twardowski, J.B. Hutchison, J.M. Rothberg, D.R. Link, N. Perrimon, M.L. Samuels, Droplet microfluidic technology for single-cell high-throughput screening, Proc. Natl. Acad. Sci. U.S.A. 106 (2009) 14195–14200, https://doi.org/10.1073/pnas.0903542106.
[8] C.S. Shultz, S.W. Krska, Unlocking the potential of asymmetric hydrogenation at Merck, Acc. Chem. Res. 40 (2007) 1320–1326, https://doi.org/10.1021/ar700141v.
[9] A.D. Clayton, J.A. Manson, C.J. Taylor, T.W. Chamberlain, B.A. Taylor, G. Clemens, R.A. Bourne, Algorithms for the self-optimisation of chemical reactions, React Chem. Eng. 4 (2019) 1545–1554, https://doi.org/10.1039/c9re00209j.

[10] A.D. Clayton, L.A. Power, W.R. Reynolds, C. Ainsworth, D.R.J. Hose, M.F. Jones, T. W. Chamberlain, A.J. Blacker, R.A. Bourne, Self-optimising reactive extractions: towards the efficient development of multi-step continuous flow processes, J. Flow Chem. 10 (2020) 199–206, https://doi.org/10.1007/s41981-020-00086-6.

[11] J.A. Manson, A.D. Clayton, C.G. Niño, R. Labes, T.W. Chamberlain, A.J. Blacker, N. Kapur, R.A. Bourne, A hybridised optimisation of an automated photochemical continuous flow reactor, Chimia (Aarau). 73 (2019) 817–822, https://doi.org/10.2533/chimia.2019.817.

[12] D.N. Jumbam, R.A. Skilton, A.J. Parrott, R.A. Bourne, M. Poliakoff, The effect of self-optimisation targets on the methylation of alcohols using dimethyl carbonate in supercritical CO2, J. Flow Chem. 2 (2012) 24–27, https://doi.org/10.1556/jfchem.2012.00019.

[13] B.E. Walker, J.H. Bannock, A.M. Nightingale, J.C. Demello, Tuning reaction products by constrained optimisation, React. Chem. Eng. 2 (2017) 785–798, https://doi.org/10.1039/c7re00123a.

[14] C.A. Hone, N. Holmes, G.R. Akien, R.A. Bourne, F.L. Muller, Rapid multistep kinetic model generation from transient flow data, React. Chem. Eng. 2 (2017) 103–108, https://doi.org/10.1039/c6re00109b.

[15] N. Holmes, G.R. Akien, R.J.D. Savage, C. Stanetty, I.R. Baxendale, A.J. Blacker, B. A. Taylor, R.L. Woodward, R.E. Meadows, R.A. Bourne, Online quantitative mass spectrometry for the rapid adaptive optimisation of automated flow reactors, React. Chem. Eng. 1 (2016) 96–100, https://doi.org/10.1039/c5re00083a.

[16] S. Krishnadasan, R.J.C. Brown, A.J. DeMello, J.C. DeMello, Intelligent routes to the controlled synthesis of nanoparticles, Lab Chip 7 (2007) 1434–1441, https://doi.org/10.1039/b711412e.

[17] N. Cherkasov, Y. Bai, A.J. Expósito, E.V. Rebrov, OpenFlowChem-a platform for quick, robust and flexible automation and self-optimisation of flow chemistry, React. Chem. Eng. 3 (2018) 769–780, https://doi.org/10.1039/c8re00046h.

[18] K. Poscharny, D.C. Fabry, S. Heddrich, E. Sugiono, M.A. Liauw, M. Rueping, Machine assisted reaction optimization: A self-optimizing reactor system for continuous-flow photochemical reactions, Tetrahedron 74 (2018) 3171–3175, https://doi.org/10.1016/j.tet.2018.04.019.

[19] P. Sagmeister, F.F. Ort, C.E. Jusner, D. Hebrault, T. Tampone, F.G. Buono, J. D. Williams, C.O. Kappe, Autonomous Multi-Step and Multi-Objective Optimization Facilitated by Real-Time Process Analytics, Adv. Sci. 2105547 (2022) 2105547, https://doi.org/10.1002/advs.202105547.

[20] E. Morgan, K.W. Burton, G. Nickless, Optimization using the super-modified simplex method, Chemom. Intell. Lab. Syst. 8 (1990) 97–107, https://doi.org/10.1016/0169-7439(90)80127-R.

[21] D. Cortés-Borda, K.V. Kutonova, C. Jamet, M.E. Trusova, F. Zammattio, C. Truchet, M. Rodriguez-Zubiri, F.X. Felpin, Optimizing the Heck-Matsuda Reaction in Flow with a Constraint-Adapted Direct Search Algorithm, Org. Process Res. Dev. 20 (2016) 1979–1987, https://doi.org/10.1021/acs.oprd.6b00310.

[22] D. Cortés-Borda, E. Wimmer, B. Gouilleux, E. Barré, N. Oger, L. Goulamaly, L. Peault, B. Charrier, C. Truchet, P. Giraudeau, M. Rodriguez-Zubiri, E. Le Grognec, F.X. Felpin, An Autonomous Self-Optimizing Flow Reactor for the Synthesis of Natural Product Carpanone, J. Org. Chem. (2018), https://doi.org/10.1021/acs.joc.8b01821.

[23] J.P. McMullen, M.T. Stone, S.L. Buchwald, K.F. Jensen, An integrated microreactor system for self-optimization of a heck reaction: From micro-to mesoscale flow systems, Angew. Chemie – Int. Ed. 49 (2010) 7076–7080, https://doi.org/10.1002/anie.201002590.

[24] J.S. Moore, K.F. Jensen, Automated multitrajectory method for reaction optimization in a microfluidic system using online IR analysis, Org. Process Res. Dev. 16 (2012) 1409–1415, https://doi.org/10.1021/op300099x.

[25] D.E. Fitzpatrick, T. Maujean, A.C. Evans, S.V. Ley, Across-the-World Automated Optimization and Continuous-Flow Synthesis of Pharmaceutical Agents Operating Through a Cloud-Based Server, Angew. Chemie – Int. Ed. 57 (2018) 15128–15132, https://doi.org/10.1002/anie.201809080.

[26] D.E. Fitzpatrick, C. Battilocchio, S.V. Ley, A Novel Internet-Based Reaction Monitoring, Control and Autonomous Self-Optimization Platform for Chemical Synthesis, Org. Process Res. Dev. 20 (2016) 386–394, https://doi.org/10.1021/acs.oprd.5b00313.

[27] S.T. Knox, S.J. Parkinson, C.Y.P. Wilding, R.A. Bourne, N.J. Warren, Autonomous polymer synthesis delivered by multi-objective closed-loop optimisation, Polym. Chem. 13 (2022) 1576–1585, https://doi.org/10.1039/d2py00040g.

[28] Y. Amar, A.M. Schweidtmann, P. Deutsch, L. Cao, A. Lapkin, Machine learning and molecular descriptors enable rational solvent selection in asymmetric catalysis, Chem. Sci. 10 (2019) 6697–6706, https://doi.org/10.1039/c9sc01844a.

[29] P.W.N.M. Van Leeuwen, P.C.J. Kamer, J.N.H. Reek, P. Dierkes, Ligand bite angle effects in metal-catalyzed C-C bond formation, Chem. Rev. 100 (2000) 2741–2769, https://doi.org/10.1021/cr9902704.

[30] B.J. Reizman, K.F. Jensen, Simultaneous solvent screening and reaction optimization in microliter slugs, Chem. Commun. 51 (2015) 13290–13293, https://doi.org/10.1039/c5cc03651h.

[31] B.J. Reizman, Y.M. Wang, S.L. Buchwald, K.F. Jensen, Suzuki-Miyaura cross-coupling optimization enabled by automated feedback, React Chem. Eng. 1 (2016) 658–666, https://doi.org/10.1039/c6re00153j.

[32] L.M. Baumgartner, C.W. Coley, B.J. Reizman, K.W. Gao, K.F. Jensen, Optimum catalyst selection over continuous and discrete process variables with a single droplet microfluidic reaction platform, React Chem. Eng. 3 (2018) 301–311, https://doi.org/10.1039/c8re00032h.

[33] H.W. Hsieh, C.W. Coley, L.M. Baumgartner, K.F. Jensen, R.I. Robinson, Photoredox Iridium-Nickel Dual-Catalyzed Decarboxylative Arylation Cross-Coupling: From Batch to Continuous Flow via Self-Optimizing Segmented Flow Reactor, Org. Process Res. Dev. 22 (2018) 542–550, https://doi.org/10.1021/acs.oprd.8b00018.

[34] P. Müller, A.D. Clayton, J. Manson, S. Riley, O.S. May, N. Govan, S. Notman, S. V. Ley, T.W. Chamberlain, R.A. Bourne, Automated multi-objective reaction optimisation: which algorithm should I use? React. Chem. Eng. (2022) https://doi.org/10.1039/d1re00549a.

[35] A.M. Schweidtmann, A.D. Clayton, N. Holmes, E. Bradford, R.A. Bourne, A. A. Lapkin, Machine learning meets continuous flow chemistry: Automated optimization towards the Pareto front of multiple objectives, Chem. Eng. J. 352 (2018) 277–282, https://doi.org/10.1016/j.cej.2018.07.031.

[36] A.D. Clayton, A.M. Schweidtmann, G. Clemens, J.A. Manson, C.J. Taylor, C. G. Niño, T.W. Chamberlain, N. Kapur, A.J. Blacker, A.A. Lapkin, R.A. Bourne, Automated self-optimisation of multi-step reaction and separation processes using machine learning, Chem. Eng. J. 384 (2020), 123340, https://doi.org/10.1016/j.cej.2019.123340.

[37] J.A. Manson, T.W. Chamberlain, R.A. Bourne, MVMOO: Mixed variable multi-objective optimisation, J. Glob. Optim. 80 (4) (2021) 865–886.

[38] M. Christensen, L.P.E. Yunker, F. Adedeji, F. Häse, L.M. Roch, T. Gensch, G. dos Passos Gomes, T. Zepel, M.S. Sigman, A. Aspuru-Guzik, J.E. Hein, Data-science driven autonomous process optimization, Commun. Chem. 4 (2021) 1–12, https://doi.org/10.1038/s42004-021-00550-x.

[39] A. Valvi, S. Tiwari, Solvent-Controlled Regioselectivity in Nucleophilic Substitution Reactions of 1-X-2,4-Difluorobenzenes with Morpholine Using Deep Eutectic Solvents, ChemistrySelect 6 (2021) 249–254, https://doi.org/10.1002/slct.202002806.

[40] S. Yu, A. Haight, B. Kotecki, L. Wang, K. Lukin, D.R. Hill, Synthesis of a TRPV1 receptor antagonist, J. Org. Chem. 74 (2009) 9539–9542, https://doi.org/10.1021/jo901943s.

[41] V.R. Joseph, H. Ying, Orthogonal-maximin latin hypercube designs, Stat. Sin. 18 (2008) 171–186.

[42] M. Halstrup, Black-box optimization of mixed discrete-continuous optimization problems, Tech. Univ. Dortmund. (2016). https://d-nb.info/112468123X/34.

[43] C.E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning), The MIT Press, 2005.

[44] D. Zhan, Y. Cheng, J. Liu, Expected improvement matrix-based infill criteria for expensive multiobjective optimization, IEEE Trans. Evol. Comput. 21 (2017) 956–975, https://doi.org/10.1109/TEVC.2017.2697503.

[45] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, A. Vijaykumar, A. Pietro Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C.N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D.A. Nicholson, D.R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G.A. Price, G.L. Ingold, G.E. Allen, G.R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J.T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J.V. de Miranda Cardoso, J. Reimer, J. Harrington, J.L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N.J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P.A. Brodtkorb, P. Lee, R.T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T.J. Pingel, T.P. Robitaille, T. Spura, T.R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y.O. Halchenko, Y. Vázquez-Baeza, SciPy 1.0: fundamental algorithms for scientific computing in Python, Nat. Methods 17 (2020) 261–272, https://doi.org/10.1038/s41592-019-0686-2.

[46] L.R. Snyder, J.J. Kirkland, J.L. Glajch, Appendix II: Properties of Solvents Used in HPLC, in: Pract. HPLC Method Dev., John Wiley & Sons, Inc., Hoboken, NJ, USA, 2012: pp. 721–728. 10.1002/9781118592014.app2.

[47] W. Greizerstein, J.A. Brieux, The ortho: Para Ratio in the Activation of the Nucleophilic Aromatic Substitution by the Nitro Group, J. Am. Chem. Soc. 84 (1962) 1032–1036, https://doi.org/10.1021/ja00865a032.

[48] N. Chéron, L. El Kaïm, L. Grimaud, P. Fleurat-Lessard, Evidences for the key role of hydrogen bonds in nucleophilic aromatic substitution reactions, Chem. - A Eur. J. 17 (2011) 14929–14934, https://doi.org/10.1002/chem.201102463.

[49] R. Brito, S. Sheth, D. Mukherjea, L. Rybak, V. Ramkumar, TRPV1: A Potential Drug Target for Treating Various Diseases, Cells. 3 (2014) 517–545, https://doi.org/10.3390/CELLS3020517.

[50] M. An Der Heiden, H. Plenio, The effect of steric bulk in Sonogashira coupling reactions, Chem. Commun. (2008) 972–974, https://doi.org/10.1039/b616608c.