Reaction Chemistry & Engineering

MINI REVIEW

View Article Online
View Journal | View Issue

Cite this: *React. Chem. Eng.*, 2019, 4, 1536

Automated platforms for reaction self-optimization in flow†

Carlos Mateos, * María José Nieves-Remacha * and Juan A. Rincón

Received 14th March 2019,
Accepted 26th April 2019

DOI: 10.1039/c9re00116f

rsc.li/reaction-engineering

Automation is a reality in everyday life, including in the chemical synthesis environment. It relegates time-consuming and repetitive tasks to machines, saving time for other activities that provide more value to scientists. The development of intelligent feedback controls applying novel algorithms with real-time reaction analysis and their combination with automation bring new opportunities for discovery, process optimization and work-up processes, that until now have been conducted using traditional manual procedures. Furthermore, the current trend is to increase the autonomy of automated platforms that can guide the optimization process with minimal human intervention (self-optimization). This article highlights recent progress in continuous flow self-optimizing platforms. Monitoring techniques, intelligent algorithms, as well as the autonomous platforms utilized are discussed.

Introduction

Synthetic chemistry and Industry 4.0

The digitization of manufacturing processes, within the so-called Industry 4.0,[1] involves the leverage of smart factories where systems are connected and communicate with each other to make decisions without human intervention. In principle, as machines receive access to more and more data, they can evolve and become much more efficient and productive. Computerization and simple process automation was considered disruptive in Industry 3.0, now it is time to optimize or refine this by the use of intelligent cyber-physical systems that connect production with logistics, based on data collection, storage and machine learning (ML).[2]

The innovative chemical industry is no exception and has adopted many of these principles in areas such as smart supply chain, safety in production and logistics.[3] In other areas such as the research and development space, automation in chemical synthesis can be applied in several domains. For synthetic route planning, chemists together with engineers and IT scientists have developed computer-assisted solutions[4] and the field is rapidly growing.[5] In the case of optimization of synthetic conditions for a particular chemical transformation, this remains a largely analog process but new approaches to build smart systems are being developed, specifically with the aid of powerful automated continuous flow process tools.[6] Governmental bodies such as DARPA in the US (through several initiatives, including Make-It program[7] and Battlefield Medicine programs)[8] or the EPSRC *via* the Dial-a-molecule Grand Challenge in the UK[9] are examples of collaborative efforts towards developing artificial intelligence approaches to plan and optimize molecule synthesis.

Self-optimizing continuous flow platforms

Continuous flow chemistry is an ideal technology for the computer-controlled optimization of chemical reactions due to its inherent benefits, such as precise control of reaction times, temperatures and composition. The setup of different process conditions can be programmed upfront, the output of each one automatically analysed and the protocol iteratively repeated (Scheme 1).

To the best of our knowledge, existing reviews[10] focus on the monitoring aspects of the self-optimization techniques in flow. In this review, we aim to provide an overview of the monitoring methods, algorithms, optimization functions, as well as bibliography of the state-of-the-art platforms.[11]

Monitoring techniques

A great number of automated analytical approaches have been developed in order to meet the requirements of science in industry and academia. They speed up routine operations and frees human resources. These approaches have also aided the fast automatic synthesis of molecules by providing quick analysis for monitoring processes and characterization.[12] The progress in the field of the implementation of inline or online[13] monitoring tools for self-optimizing platforms has been reviewed at the end of 2015.[10b] Only more recent examples will be mentioned below.

Online HPLC. Online HPLC is by far the most used technique due to its high versatility, low implementation time

Centro de Investigación Lilly S.A., Avda. de la Industria, 30, Alcobendas-Madrid 28108, Spain. E-mail: c.mateos@lilly.com, nieves_maria_jose@lilly.com

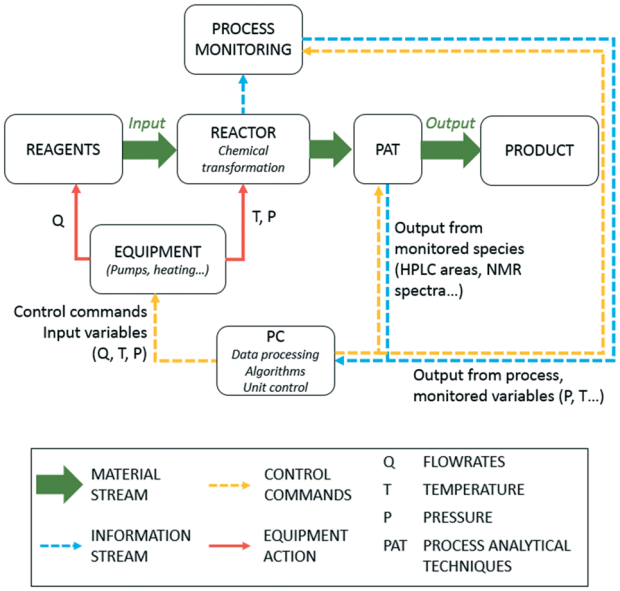
† Electronic supplementary information (ESI) available: Table with references on self-optimization processes in flow to date. See DOI: 10.1039/c9re00116f

and readily availability of the offline version in almost every synthetic laboratory. An interesting recent example from Bourne and coworkers describes the self-optimization of the last step of the synthesis of Osimertinib (AZD9291),[14] an irreversible epidermal growth factor receptor (EGFR) kinase inhibitor using the SNOBFIT algorithm. (Scheme 2). A filtered aliquot of the reaction mixture is sent to the HPLC through a sample loop without quenching or dilution.

Other processes that use HPLC have also been reported by Bourne *et al.*, such as Claisen condensation,[15] or S_NAr reactions with simultaneous optimization of multiple objectives including productivity and environmental impact (E-factor) or % impurity.[16] Jensen *et al.* have integrated an online LC-MS analysis with the on-demand creation of droplets, with interchangeable reagents and catalysts, which reacted in a fully automated microfluidic system[17] (Scheme 3). They incorporated a refractive index sensor to timely sample and injected a quenched aliquot through a loop into the HPLC. It is interestingly useful to control the time for sample collection and analysis based on another detection method, especially in a flow setup where steady-state conditions are targeted.

The same group also reported the generation of droplets (~15 μL) in a segmented flow reactor for a self-optimizing platform of a photoredox reaction[18] (Scheme 4).

Very recently, Jensen, Jamison and co-workers have leveraged the system to a reconfigurable, multi-purpose platform capable of self-optimizing a plethora of chemical reactions[19] including Buchwald–Hartwig cross coupling, Horner–Wadsworth–Emmons olefination, reductive amination,



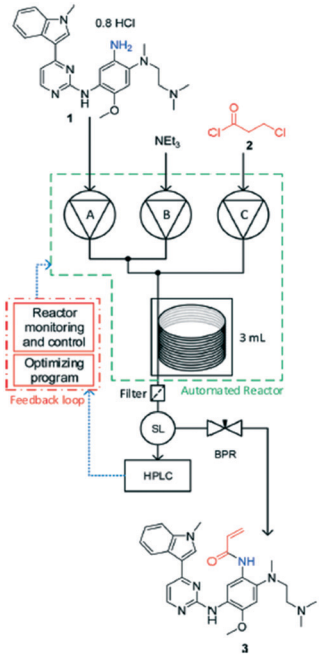
Scheme 1 Typical elements of a closed-loop reaction optimization platform. A material stream with reagents and solvents are pumped into the reactor, where a chemical transformation occurs. The output of the reactor is a crude reaction mixture stream that is analysed by a process analytical technique (PAT). Sensors attached to the reactor and PAT generate data that is stored and processed by a computer, which also sends commands to the equipment according to the self-optimization algorithm output that selects the next operating conditions.

Juan A. Rincón, María José Nieves-Remacha, Carlos Mateos

Dr. Carlos Mateos obtained his degree in Organic Chemistry, in 1998, in the University of Oviedo (Spain). Then, he moved to Leverkusen, (Germany) to enjoy an industrial internship in Bayer AG working on the industrial development of azo-dyes. In 1999 he moved back to Spain to pursue a PhD on natural product synthesis, in the University of Oviedo under the supervision of Professor José Barluenga. After receiving his PhD in 2004, Carlos joined the CRO Galchimia (Santiago de Compostela, Spain) as a project manager. In 2006, he joined Lilly where he is currently Research Scientist of the flow chemistry group in the R&D Lilly site in Alcobendas (Madrid). His main research interests include medicinal chemistry, process development, continuous processing and external R&D collaboration.

Dr. María José Nieves-Remacha, is a Chemical Engineer Ph.D. (2014) and MS.CEP. (2009) by Massachusetts Institute of Technology and B.S. by Universidad Complutense de Madrid. She focused her doctoral studies on the scale-up of multiphase continuous flow chemistries under the supervision of Professor Klavs F. Jensen, conducting both experiments and computational fluid dynamic simulations of transient immiscible systems. After that, she joined the Dow Chemical Company to work as a Senior Engineer in the Reaction Engineering group within the Core R&D Department at Freeport, TX. In 2016 she joined Lilly in Alcobendas, Spain, where she currently works as a Senior Research Scientist within Lilly Research Laboratories. Her research interests are: reaction engineering, automation, flow technologies, and autonomous systems to aid process optimization.

Dr. Juan A. Rincón (Valladolid, Spain) studied Chemistry in the University of Valladolid and received his doctorate in Organic Chemistry in 1999 in the same university under the supervision of Prof. Francisco J. Pulido, working in organotin and cuprates chemistry. During his thesis, he made two predoctoral stays at the University of Cambridge (UK) under the supervision of Prof. Ian Fleming. In 1999, he joined the Organic Chemistry Institute (CSIC, Madrid) as a postdoctoral researcher to work in the field of catalysis supported on new materials, under the supervision of Prof. Félix Sánchez. In 2002, he joined Lilly where he is currently Research Scientist of the flow chemistry group in the R&D Lilly site in Alcobendas (Madrid). His main research interests include continuous processing, route development, process safety and synthetic and medicinal chemistry.

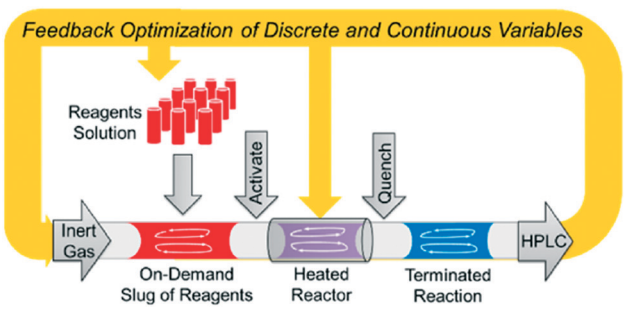


Scheme 2 Automated flow reactor with adaptive feedback control and optimization algorithm. Reproduced from ref. 14 with permission from the Royal Society of Chemistry

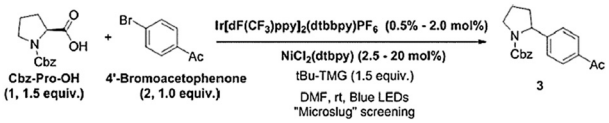
Suzuki–Miyaura cross-coupling, nucleophilic aromatic substitution (S_NAr), and a visible light photoredox reaction. Remarkably, they have used this platform for a two-step process (ketene generation followed by alkene cycloaddition) where the product of the first reaction is of limited stability (Scheme 5).

Gas chromatography. Gas chromatography (GC) has been employed in cases where the high volatility of target molecules does not permit the use of LC methods. In the case of the work by Felpin *et al.* for the optimization of the Heck–Matsuda reaction, an offline GC-MS analysis was carried out after collecting samples in a fraction collector[20] (Scheme 6).

In addition to offline GC, Cherkasov, Rebrov *et al.* described the use of online GC analysis in several applications of their OpenFlowChem platform for the self-optimization of different hydrogenation reactions, using a standard



Scheme 3 Concept diagram for on-demand preparation, reaction, analysis and feedback in automated reaction flow screening. Reproduced from ref. 17 with permission from the American Chemical Society.



Scheme 4 Photoredox iridium–nickel dual-catalyzed decarboxylative arylation cross-coupling. Reproduced from ref. 18 with permission from the American Chemical Society.

autosampler with a modified flow-through vial for sampling.[21]

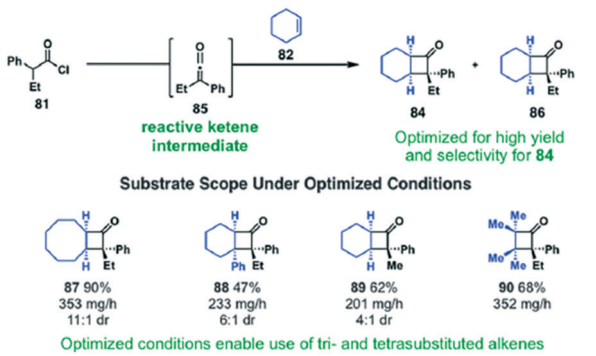
Lapkin *et al.* utilized an inline UV cell to detect flow segments subsequently triggering online GC analysis of the aziridine products synthesized[22] using model-based design of experiments and self-optimization approaches in flow (Scheme 7).

Online mass spectrometry (MS). Recent developments in small-footprint mass spectrometers have enabled the monitoring of ion abundances of starting materials, intermediates or products. Ley *et al.* have developed an internet-based reaction monitoring system for the heterogeneous hydration of 3-cyanopyridine over manganese dioxide[23] (Scheme 8).

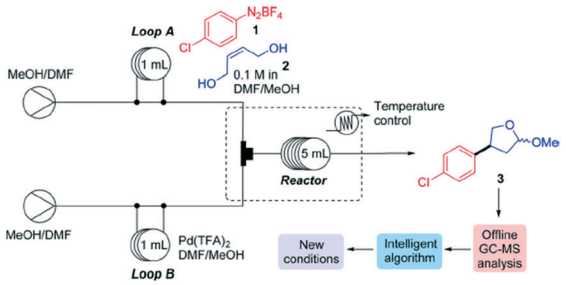
Zare *et al.* have used the output of a mass spectrometer as the optimization objective to develop their Deep Reaction Optimizer by Reinforcement Learning.[24] Four different type of reactions have been optimized following this approach (Scheme 9).

Inline FTIR. Despite the great uptake of the use of infrared radiation (IR) in the self-optimization space, given the reported examples,[10b] this technique has had very limited use in regard to self-optimization since 2015. Rueping *et al.* described a self-optimizing reactor system for continuous-flow photochemical Paternò–Büchi reactions using an FTIR flow cell[25] (Scheme 10).

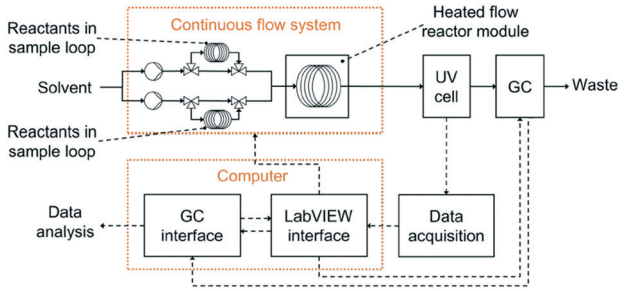
Other monitoring techniques. Junkers *et al.* used **online SEC** (size exclusion chromatography) for the analysis of final product streams in the preparation of polymers with predefined molecular weights in high accuracy.[26] Felpin *et al.* described the use of either online HPLC or **inline benchtop NMR** for monitoring purposes depending on several parameters discussed in their work, with an interesting comparison of pros and cons of these two techniques.[27] A self-optimized



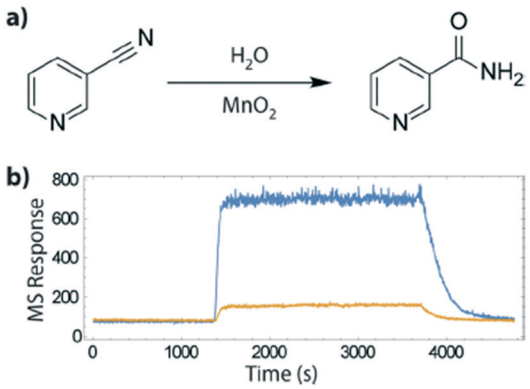
Scheme 5 Substrate scope of the auto-optimized ketene generation and [2 + 2] cycloaddition. Reproduced from ref. 19 with permission from American Association for the Advancement of Science.



Scheme 6 Optimization of a Heck–Matsuda reaction using a Nelder–Mead simplex method. Reproduced from ref. 20 with permission from the American Chemical Society.



Scheme 7 Schematics of the automated continuous-flow system used for model development and "black-box" sequential optimization. Reproduced from ref. 22 with permission from Beilstein-Institut.

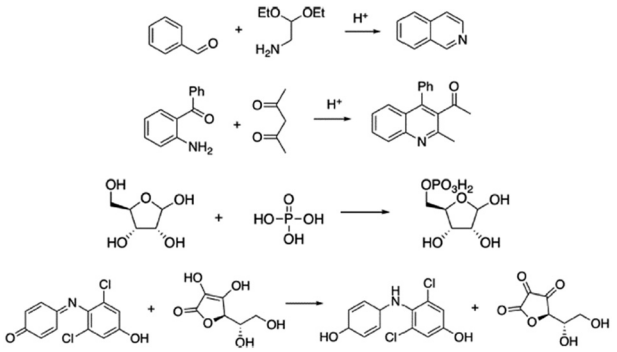


Scheme 8 a) 3-Cyanopyridine was converted to its amide over MnO_2 as a demonstration of simple experiment automation, b) the mass spectrometer response collected by LeyLab when a single plug of reagent solution was passed through the packed column (the blue line corresponds to the starting material, while the orange line represents the product). Reproduced from ref. 23 with permission from the American Chemical Society.

[3,3]-Claisen rearrangement platform for the synthesis of an intermediate of the natural product carpanone is depicted in Scheme 11.

Intelligent algorithms

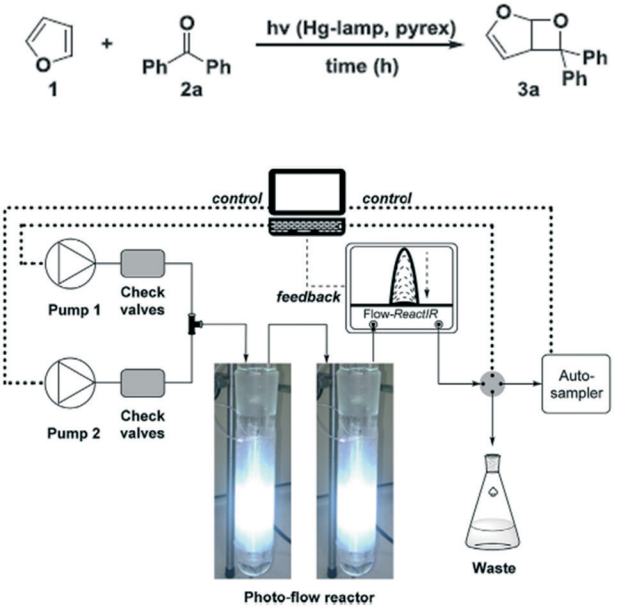
Today's automated platforms tend to incorporate intelligent algorithms to guide the optimization process and reduce the



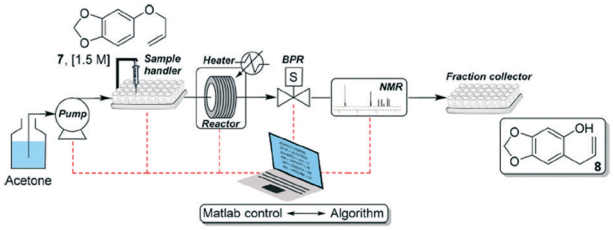
Scheme 9 Optimization of different chemical reactions with deep reinforcement learning. Reproduced from ref. 24 with permission from the American Chemical Society.

level of human intervention. This means that automation is not only performed at a control level of the equipment and monitoring of process variables; rather, closed loops and feedback controls are incorporated to decide the next experimental conditions for the reaction to optimize. This approach enables reduction of human bias in the process of reaction optimization and, ideally, reduction of the number of experiments required to reach the optimum by following the most efficient decision path.

A variety of algorithms have been used in reaction self-optimization examples. Depending on the algorithm, either a local or a global optimal solution can be found. While a global optimal solution is a feasible solution with a value for the objective function that is as good or better than any other feasible solutions, a local optimum is a feasible solution that is better than neighboring solutions, but does not guarantee



Scheme 10 Self-optimizing photo-flow Paternò–Büchi reaction setup with two HPLC pumps and on-line ReactIR. Reproduced from ref. 25 with permission from Elsevier.



Scheme 11 Reproduced from ref. 27 with permission from the American Chemical Society.

to be the best within the entire design space (Fig. 1). Frequently, a target value for the objective function is specified and when the algorithm finds conditions that yield this value the optimization process terminates. Alternatively, termination is defined when two consecutive solutions are closer than a predefined tolerance, or when a maximum number of iterations is reached. It is also important to consider how the experimental error affects the efficiency of the algorithm. That is, if there is noise in the data or if there is an outlier, will the algorithm deviate from the local optimum, or rather, will the algorithm self-correct its path to yield the local optimum.

Classical optimization methods base the search in a point-by-point approach (one solution is modified to a different one in each iteration).[28] In the gradient-free direct (or pattern) search method, only the objective function and constrained values guide the search. However, gradient-based methods use first- and/or second-order derivatives of the objective function and/or constraints. While gradient-based methods offer rapid convergence, gradient-free methods are widely applicable to both continuous and discrete functions. Nelder–Mead simplex is a gradient-free method, while the steepest descent (or ascent) and the conjugate gradient are gradient-based.[29]

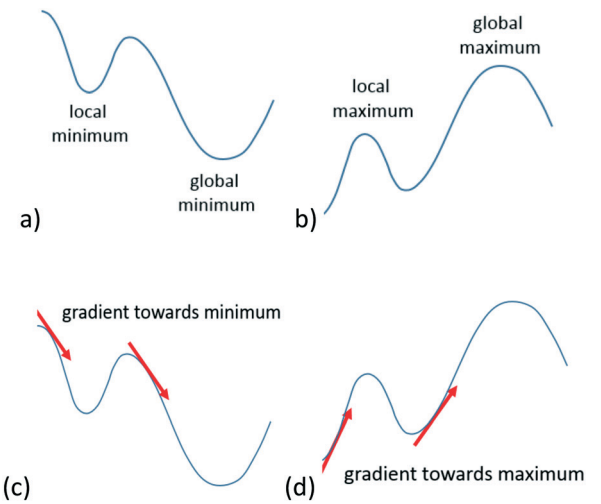


Fig. 1 Global *versus* local optimum: a) minimum, b) maximum. Gradient vector points in the direction of maximum function value change: c) steepest descent, or d) steepest ascent.

Nelder–Mead simplex (Nonlinear Simplex)[30] – and its different versions – is probably the most used gradient-free method in reaction self-optimization together with the SNOBFIT algorithm. The algorithm forms a simplex with n + 1 (ref. 32) vertices and reshapes at each iteration depending on the new function evaluation value after ordering the vertices (Fig. 2). It is efficient for problems with fewer than ten design variables. This algorithm leads to local optima, but coupled with multiple point restarts can lead to the global optimum. It is also able to self-correct the path to a local optimum even if there are outliers present in the data.[31]

Ley *et al.* demonstrated the use of a modified version of the simplex algorithm (called Complex) in several reaction examples.[23] The system required 12 experiments to find optimal conditions in a 3D[32] problem and 30 experiments in a 5D problem. Felpin *et al.* modified the simplex algorithm adapting it to constrained optimizations in multidimensional space.[20] They demonstrated the flexibility of the algorithm by applying it to different objective functions; 14 experiments were required to maximize yield (Fig. 3), 13 to maximize throughput, and 18 experiments to minimize production cost. Another publication from the same author includes, in addition to the simplex method, a golden section search to handle 1D boundary searches where the Nelder–Mead algorithm fails.[27] They demonstrated its use in a four-step synthesis where all discrete steps were optimized for product yield within 66 experiments. Rueping *et al.* showed the robustness of the Simplex algorithm, which was able to continue the optimization process and achieve convergence in 25 experiments, regardless of casual experimental errors that occurred.[25]

SNOBFIT (Stable Noisy Optimization by Branch and FIT), developed by Huyer and Neumaier, combines global and local search by branching and local fits.[33] While sampling/direct search methods work by generating a sequence of points, modelling methods try to approximate the function over a region by a model function. The algorithm generates points widely distributed across the defined chemical space to increase the chances of finding a global optimum.

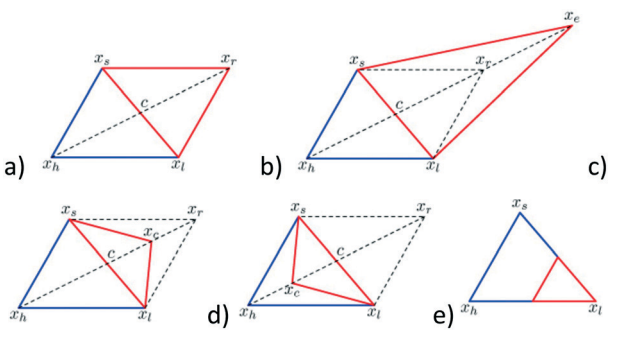


Fig. 2 Graphical representation of the simplex algorithm: a) reflection, b) expansion, c) outside contraction, d) inside contraction, e) shrinking. The blue line represents the original simplex. Nomenclature of vertices: x_h, worst; x_l, best; x_s, second best; x_r, reflected; x_c, contracted; x_e, expanded; c, centroid.[30c]

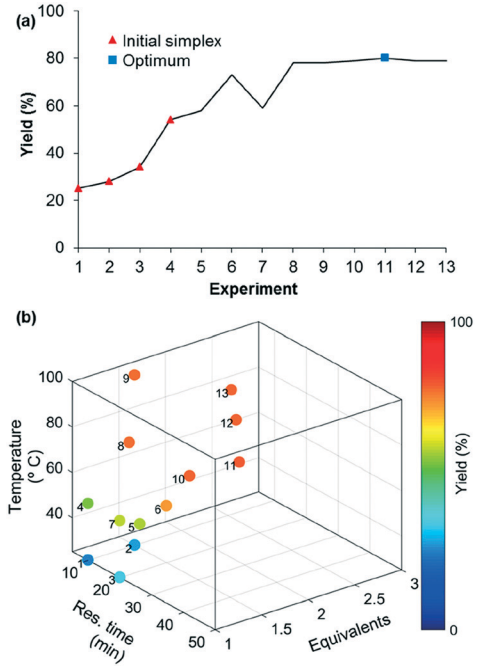


Fig. 3 3D optimization using the simplex algorithm. a) Evolution of response with iteration number; b) 3D representation of variables optimized to maximize yield. Reproduced from ref. 27 with permission of the American Chemical Society.

Using the SNOBFIT algorithm Bourne *et al.* were able to optimize a multistep process within 42 experiments in 26 h run for 4 variables (temperature and flowrate of three streams) to maximize yield calculated as the ratio of percent area of product to percent area of an internal standard (Fig. 4).[14] Bourne *et al.* used the same algorithm to optimize a minor product yield, changing values of four variables in 70 experiments, and subsequently fitted the data to surface response models, which allowed them to obtain different reaction metrics without further experimentation.[15] Jensen and Jamison *et al.* applied the SNOBFIT algorithm to different reaction chemistries and reached convergence within

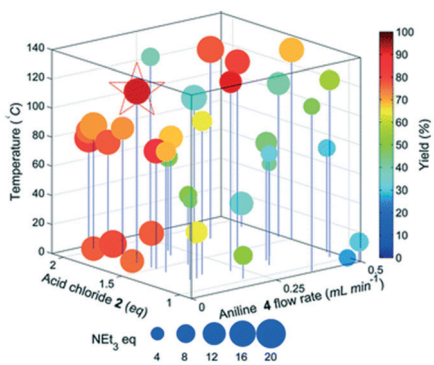


Fig. 4 4D optimization using the SNOBFIT algorithm. The size of the point represents the molar eq. of NEt_3, and the colour represents the product yield. The optimal conditions are highlighted by the star. Reproduced from ref. 14 with permission of the Royal Society of Chemistry.

30–45 experiments for 3–5 manipulated variables (temperatures, flowrates and catalyst percent) to optimize yield.[19] Cherkasov, Rebrov *et al.* performed 3D optimization, minimizing an objective function that incorporated squared product yield (calculated through selectivity to product and substrate conversion) and substrate flowrate, reaching convergence in 61 experiments.[21]

The majority of self-optimizing platforms target a single-objective function, usually embedding several variables throughout an equation and adding penalty terms to account for multiple performance criteria. For example, Jensen, Robinson *et al.* introduced a residence time term in the definition of productivity, calculated as product yield/residence time to penalize long reaction times.[18] However, the single-objective approach does not identify trade-offs between opposing performance variables and the stability of the solution may depend on the equation used. Moreover, the determination of weight factors for each term in the objective function is difficult and often requires additional experimentation. For instance, Ley *et al.* compared two different objective functions for a 5D optimization that included terms for throughput, conversion and reagent consumption.[23] The first objective function combined conversion and consumption into the same term, whereas the second considered three terms with different weight factors. The two-term objective function produced an oscillatory response, whereas separation of the evaluation function into three terms produced a more stable response that led to an optimal solution in 30 iterations.

Only very recent and few publications include multi-objective self-optimization. In 2017 Lapkin *et al.* applied a black-box sequential optimization using a Multi-Objective Active Learner (MOAL) algorithm to optimize a C–H activation in flow.[22] With cost and yield as evaluation functions, the process required 11 iterations, five of which comprised the training set. In 2018 Lapkin, Bourne *et al.* used the Thompson sampling efficient multi-objective (TSEMO) algorithm to self-optimize the conflicting productivity and environmental objectives simultaneously with fewer iterations than genetic algorithms,[28] generating a Pareto front for the trade-off of objectives (Fig. 5).[16]

Many research publications deal with optimization of continuous or quantitative variables, but very few include discrete or qualitative variables. Reizman and Jensen published in 2016 a strategy to optimize simultaneously both discrete and continuous variables.[17] They use a DoE-based adaptive response surface algorithm that removes poor-performing discrete variables as experiments progress and use a G-optimal strategy for the continuous variables. Based on a fractional factorial design, the solution for the best solvent converged in 67 experiments, validated by a gradient-based quasi-Newton method for the remaining variables. Later, Jensen, Robinson, *et al.*[18] followed a similar approach: starting with a minimal D-optimal set of experiments, creating quadratic-based response surface models for each discrete variable, and following a G-optimal strategy for the remaining continuous variables. The algorithm reached

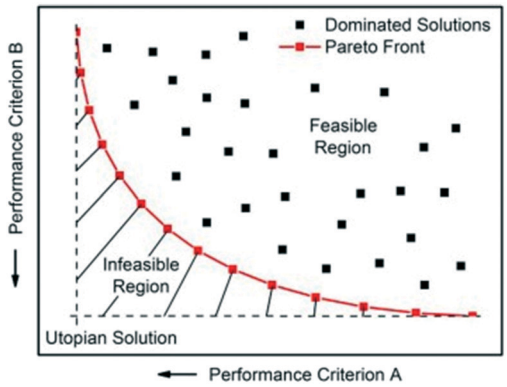


Fig. 5 An example of a system with two competing minimization performance criteria A and B. It is infeasible to find the utopian point where both A and B are at their optimal values. The points on the Pareto front are non-dominated solutions, as A or B cannot be improved without having a detrimental effect on the other. Reproduced from ref. 16 under the Creative Commons Attribution License.[34]

convergence after 22 experiments (13 for D-optimal and 9 for G-optimal strategies).

More recently, Zare *et al.* applied deep reinforcement learning[35] to self-optimize chemical reactions by combining a policy gradient model with an efficient exploration strategy based on probability distributions.[24] This algorithm learns from past experience and the model is updated with the new experimentally-generated data. The authors compare its performance through simulations with the Nelder–Mead simplex, SNOBFIT, and covariance-matrix adaptation-evolution strategy (CMA-ES)[36] showing a 71% reduction in iterations.

Autonomous platforms

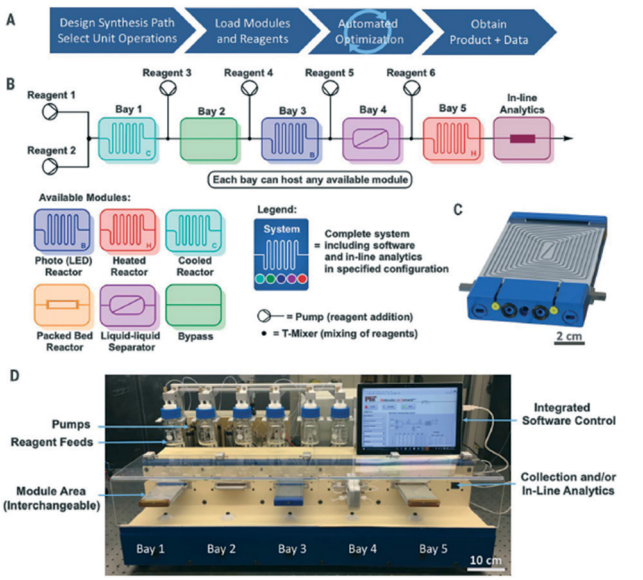
To date, many of the platforms have been developed using commercial software such as LabVIEW[37] for automation of equipment and data management, and Matlab®,[38] most frequently to implement optimization algorithms. Both platforms are integrated to create feedback loops between process analytics and the optimization algorithm, which generates new conditions for evaluation. This is the case of Reizman and Jensen,[17] Jensen, Robinson *et al.*,[18] and Rueping *et al.*[25] Felpin *et al.* used Matlab® to develop an adapted Nelder–Mead algorithm but do not report integrated feedback control in this publication.[20] In very few cases a single programming platform, such as Matlab®, has been used to both automate equipment control and implement the optimization algorithm.[14–16,27]

Most automated self-optimization platforms are designed and developed for a specific application. Creating general, fully automated systems integrating equipment from different vendors to operate in an autonomous way reducing the amount of human intervention is a challenge. Frequently, the equipment we have in laboratories use different communication protocols for its control, the monitoring devices log data in different formats, and in many cases both are restricted to the vendor's software for their control from a computer. There is an increasing effort to create modular or plug-and-play systems not only from the hardware viewpoint but also from the software viewpoint. This includes not only equipment control, but also data management from collection through monitoring devices to storage, processing, and visualization in a reliable and consistent manner. Having access to platforms that facilitate automation integration of different hardware components from different vendors in a modular and generalized manner reduces the time for process development and lowers the barriers for unexperienced users that otherwise would require additional expertise on the field.

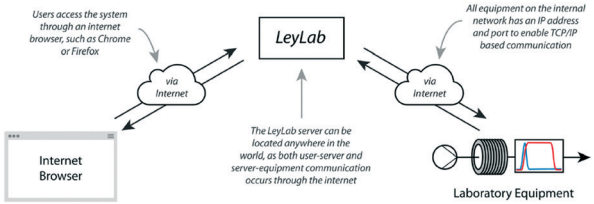
In 2018 Jensen and Jamison *et al.* presented a reconfigurable system with modules for heating, cooling, mixing and photochemical capabilities, compatible with different monitoring systems and able to perform self-optimization of different reactions (Scheme 12). Automation is based on LabVIEW integrated with Matlab® and uses the SNOBFIT algorithm for reaction self-optimization.[19]

Today, there is a growing trend to use open-source alternatives and cloud-based systems enabling remote control and monitoring. Often, authors share the code within the community through platforms such as GitHub.[39]

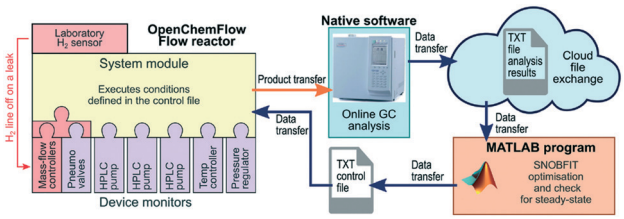
In 2016 Ley *et al.* developed an Internet-based modular software system (called *LeyLab*) with remote control and monitoring, self-optimization based on the simplex algorithm, and communication between server-equipment and server–server based on the TCP/IP protocol (Scheme 13).[23] The system is comprised of four modules: graphical interface



Scheme 12 Plug-and-play, reconfigurable, continuous-flow chemical synthesis system. (A) General four-step protocol for using the system. (B) Representative configuration of the components in the system. (C) CAD (computer-aided design) representation of the LED reactor; shown is a view of the end that attaches to a universal bay on the system. (D) Schematic representation of the configuration shown in (B) and available modules. Reproduced from ref. 19 with permission from American Association for the Advancement of Science.



Scheme 13 Scheme for LeyLab communication system. Reproduced from ref. 23 with permission from the American Chemical Society.



Scheme 14 Scheme of the OpenFlowChem automation system with self-optimization. Reproduced from ref. 21 with permission from the Royal Society of Chemistry.

(accessed through the Internet), database, equipment communication module, and equipment command module.

Cherkasov and Rebrov *et al.* published an open-source flexible platform (called *OpenFlowChem*) based on LabVIEW and cloud-based data transfer interacting with Matlab for SNOBFIT optimization (Scheme 14).[21] Using this platform, new automated systems can be created or modified with minimal efforts within hours.

ChemOS is a modular software package developed by Aspuru-Guzik *et al.* that includes different structured layers to operate an autonomous laboratory: interaction with researchers, databases, robots, characterization, analysis, and artificial-intelligence-based learning procedures for autonomous operation.[40]

Although not specific to self-optimization in flow, it is worth mentioning the efforts in the field of open-source software for automation by Van der Made,[41] S. Ley[42] and R. Ingham,[43] and the recent modular robotic platform developed by Cronin *et al.*[44]

Summary and future outlook

Automation is an attractive field for discovery chemistry, process development and optimization as demonstrated by the latest advancements published in the literature. While in the past the automated platforms followed a deterministic approach with scheduled tasks and protocols, the trend today is to increase the level of autonomy through intelligent algorithms that provide these systems with the ability to learn, make decisions and take actions depending on the different scenarios that arise.

In the development of autonomous platforms, there is an increasing focus on multi-variate and multi-objective optimization algorithms, together with the design of modular, flexible and connected systems that facilitate components integration with the aim of reducing process development times.

We envision that an ideal fully automated system would take as input the molecule to synthesize and the system would perform all the subsequent necessary actions from discovery to process development and scale-up. Although we have focused this mini-review on flow systems, application of automation to batch systems has already been widely implemented as high-throughput screening platforms. The current tendency is also to combine it with artificial intelligence algorithms that are able to learn from past data to guide the experimentation.

The vast majority of self-optimization work published has been developed in academia to date. It is obvious that industry is also working in this area and in the near future, it is expected a further increase in industry uptake of these methodologies driven by market demand for greater innovation speed and cost-reduction initiatives.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors would like to thank Dr. Graham R. Cumming for helpful discussions during the manuscript preparation.

Notes and references

1 A. Rojko, Industry 4.0 Concept: Background and Overview, *International Journal of Interactive Mobile Technologies (iJIM)*, 2017, **11**(5), 77–90, DOI: 10.3991/ijim.v11i5.7072.
2 For recent reviews about ML in drug discovery see: (*a*) J. Panteleeva, H. Gaoa and L. Jiab, *Bioorg. Med. Chem. Lett.*, 2018, **28**, 2807–2815; (*b*) Y.-C. Lo, S. E. Rensi, W. Torng and R. B. Altman, *Drug Discovery Today*, 2018, **23**, 1538.
3 A. Reitze, N. Jürgensmeyer, S. Lier, M. Kohnke, J. Riese and M. Grünewald, *Angew. Chem., Int. Ed.*, 2018, **57**, 4242–4247.
4 For reviews see: (*a*) S. Szymkuć, E. P. Gajewska, T. Klucznik, K. Molga, P. Dittwald, M. Startek, M. Bajczyk and B. A. Grzybowski, *Angew. Chem., Int. Ed.*, 2016, **55**, 5904–5937; (*b*) F. Feng, L. Lai and J. Pei, *Front. Chem.*, 2018, **6**(199), 1–10.
5 J. M. Perkel, *Nature*, 2017, **542**, 125–126.
6 Excellent example from Pfizer:D. Perera, J. W. Tucker, S. Brahmbhatt, C. J. Helal, A. Chong, W. Farrell, P. Richardson and N. W. Sach, *Science*, 2018, **359**(6374), 429–434.
7 The Make-It program in the Defense Advanced Research Projects Agency DARPA Defense Sciences Office (DSO) is soliciting innovative research proposals in the area of research related to the development of an automated chemical synthesizer that can produce, purify, characterize and scale a wide range of small molecules. http://go.usa.gov/3Pzww.
8 J. J. Lewin, E. J. Choi and G. Ling, *Am. J. Health-Syst. Pharm.*, 2016, **73**, 45–54.

9 K. J. Kilpin and R. J. Whitby, Chemistry Central Journal theme issue: Dial-a-molecule, *Chem. Cent. J.*, 2015, **9**, 43, DOI: 10.1186/s13065-015-0122-3.
10 Reviews covering the field to Oct 2015: (*a*) V. Sans and L. Cronin, *Chem. Soc. Rev.*, 2016, **45**, 2032; (*b*) D. C. Fabry, E. Sugiono and M. Rueping, *React. Chem. Eng.*, 2016, **1**, 129–133.
11 See ESI† for a table containing references on self-optimization processes in flow.
12 G. Rakesh, D. Prabhu and P. L. Urban, *TrAC, Trends Anal. Chem.*, 2017, **88**, 41–52.
13 For a discussion of the difference between inline and online methods and early example of online mass-spec monitoring see:D. L. Browne, S. Wright, B. J. Deadman, S. Dunnage, I. R. Baxendale, R. M. Turner and S. V. Ley, *Rapid Commun. Mass Spectrom.*, 2012, **26**, 1999–2010.
14 N. Holmes, G. R. Akien, A. J. Blacker, R. L. Woodward, R. E. Meadows and R. A. Bourne, *React. Chem. Eng.*, 2016, **1**, 366–371.
15 M. I. Jeraal, N. Holmes, G. R. Akien and R. A. Bourne, *Tetrahedron*, 2018, **74**, 3158–3164.
16 A. M. Schweidtmann, A. D. Clayton, N. Holmes, E. Bradford, R. A. Bourne and A. A. Lapkin, *Chem. Eng. J.*, 2018, **352**, 277–282.
17 B. J. Reizman and K. F. Jensen, *Acc. Chem. Res.*, 2016, **49**, 1786–1796.
18 H.-W. Hsieh, C. W. Coley, L. M. Baumgartner, K. F. Jensen and R. I. Robinson, *Org. Process Res. Dev.*, 2018, **22**, 542–550.
19 A.-C. Bédard, A. Adamo, K. C. Aroh, M. G. Russell, A. A. Bedermann, J. Torosian, B. Yue, K. F. Jensen and T. F. Jamison, *Science*, 2018, **361**, 1220–1225.
20 D. Cortés-Borda, K. V. Kutonova, C. Jamet, M. E. Trusova, F. Zammattio, C. Truchet, M. Rodriguez-Zubiri and F.-X. Felpin, *Org. Process Res. Dev.*, 2016, **20**, 1979–1987.
21 N. Cherkasov, Y. Bai, A. J. Expósito and E. V. Rebrov, *React. Chem. Eng.*, 2018, **3**, 769–780.
22 A. Echtermeyer, Y. Amar, J. Zakrzewski and A. Lapkin, *Beilstein J. Org. Chem.*, 2017, **13**, 150–163.
23 D. E. Fitzpatrick, C. Battilocchio and S. V. Ley, *Org. Process Res. Dev.*, 2016, **20**, 386–394.
24 Z. Zhou, X. Li and R. N. Zare, *ACS Cent. Sci.*, 2017, **3**, 1337–1344.
25 K. Poscharny, D. Fabry, S. Heddrich, E. Sugiono, M. Liauw and M. Rueping, *Tetrahedron*, 2018, **74**, 3171–3175.
26 M. Rubens, J. H. Vrijsen, J. Laun and T. Junkers, *Angew. Chem., Int. Ed.*, 2019, **57**, 3183–3187.
27 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 and F.-X. Felpin, *J. Org. Chem.*, 2018, **83**(23), 14286–14299.
28 K. Deb, *Multi-objective optimization using evolutionary algorithms*, John Wiley & Sons, Ltd. England, UK, 2001.
29 Steepest descent or gradient-descent: is a gradient-based method that moves the point to a new one located along the line of maximum gradient. This, coupled with surface response models from DoE-based experiments, enables the model to propose new conditions to reach the optimal solution. Conjugate gradient is a gradient-based method that accelerates the rate of steepest descent without the need of computing the Hessian, which is required in the faster Newton's method.
30 (*a*) J. A. Nelder and R. Mead, *Comput. J.*, 1965, **7**, 308–313; (*b*) J. C. Lagarias, J. A. Reeds, M. H. Wright and P. E. Wright, *SIAM J. Control Optim.*, 1998, **9**(1), 112–147; (*c*) S. Singer and J. Nelder, *Scholarpedia*, 2009, **4**(7), 2928, Website: http://www.scholarpedia.org/article/Nelder-Mead_algorithm (Accessed on 26/04/2019).
31 P. Milavec, A. Podgornik, R. Stravs and T. Koloini, *Bioprocess Biosyst. Eng.*, 2002, **25**, 69–78.
32 Throughout this article, "*n*" refers to an *n*-dimensional (*n*D) problem with "*n*" being an integer and representing the number of variables manipulated to optimize a response.
33 W. Huyer and A. Neumaier, *ACM Trans. Math. Softw.*, 2008, **35**, 1–25.
34 Creative Commons Attribution License website, https://creativecommons.org/licenses/by/4.0/ (Accessed on 8/3/2019).
35 Deep reinforcement learning is an area of Machine Learning within the broader Artificial Intelligence framework that concerns with how software agents with capacity to act choose actions in an environment that influence the agent's future state to maximize a future reward. DeepMind is pioneer in this field: https://deepmind.com/blog/deep-reinforcement-learning/ (Accessed on 22/02/2019).
36 Covariance matrix adaptation-evolution strategy (CMA-ES) is a stochastic, evolutionary method for real-valued optimization.
37 (*a*) LabVIEW (Laboratory Virtual Instrument Engineering Workbench) is a graphical programming language from National Instruments for automation control and data acquisition: www.ni.com (Accessed on 22/02/2019); (*b*) C. Elliott, V. Vijayakumar, W. Zink and R. Hansen, National Instruments LabVIEW: A Programming Environment for Laboratory Automation and Measurement, *JALA*, 2007, **12**(1), 17–24.
38 Matlab® https://www.mathworks.com/products/matlab.html (Accessed on 22/02/2019).
39 GitHub is a code development platform: https://github.com/ (Accessed on 22/02/2019).
40 L. M. Roch, F. Häse, C. Kreisbeck, T. Tamayo-Mendoza, L. P. E. Yunker, J. E. Hein and A. Aspuru-Guzik, 2018, DOI: 10.26434/chemrxiv.5953606.v1, ChemRxiv. Preprint.
41 M. O'Brien, A. Hall, J. Schrauwen and J. van der Made, *Tetrahedron*, 2018, **74**(25), 3152–3157.
42 R. J. Ingham, C. Battilocchio, J. M. Hawkins and S. V. Ley, *Beilstein J. Org. Chem.*, 2014, **10**, 641–652.
43 (*a*) R. J. Ingham, *PhD Thesis*, Department of Chemistry, 2014, Ley Group: https://core.ac.uk/download/pdf/77406786.pdf (Accessed on 25/04/2019); (*b*) Open-source software can be accessed at the following address: https://github.com/richardingham/octopus (Accessed on 25/04/2019).
44 S. Steiner, J. Wolf, S. Glatzel, A. Andreou, J. M. Granda, G. Keenan, T. Hinkley, G. Aragon-Camarasa, P. J. Kitson, D. Angelone and L. Cronin, *Science*, 2019, **363**, 1–8.

