Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Engineering

Automatic discovery and optimization of chemical processes

Claudia Houben and Alexei A Lapkin

This paper presents the first overview of recent developments in techniques and methods that enable closed-loop optimization, also sometimes called 'self optimization', as well as discovery in different areas of molecular sciences. The closed-loop experimental platforms offer tremendous new opportunities by significantly increasing productivity, as well as enabling completely new types of experiments to be performed. Such experiments involve three main enabling technology areas: automated experimental systems, analytical instruments connected to automated chemoinformatics software and optimization or decision-making algorithms. We review the most exciting developments concerning robotic experiments, 3D printed lab-ware, experimental systems with multiple analytical instruments and advanced optimization algorithms based on machine learning approaches. A range of different chemical problems is described, which show the breadth of potential applications of this emerging experimental approach.

Address
Department of Chemical Engineering and Biotechnology, University of Cambridge, Pembroke Street, New Museums Site, Cambridge CB2 3RA, United Kingdom

Corresponding author: Lapkin, Alexei A (aal35@cam.ac.uk)

Current Opinion in Chemical Engineering 2015, **9**:1–7

This review comes from a themed issue on **Reaction engineering and catalysis**

Edited by **Marc-Olivier Coppens** and **Theodore T Tsotsis**

http://dx.doi.org/10.1016/j.coche.2015.07.001

Introduction

Recent advances in laboratory equipment automation and in new analytical methods rapidly transform the way in which experiments in molecular sciences are performed. These changes indicate a remarkable transition, which profoundly affects the research and development processes, and scientific methodologies across many areas of molecular sciences, as well as creating new business opportunities and new business models. Several pioneering studies have recently demonstrated a range of possibilities in automatic optimization of process conditions [1,2,3•,4–6], exploring new chemical structures [7•] and in the search for new bio-active compounds [8].

Technical underpinnings of this transition could be traced to the emergence in recent years of several new capabilities:

- rapid *in situ* or in line analysis of the outcomes of chemical transformations,
- chemoinformatics tools for treating large amounts of analytical data from data-rich experiments,
- miniaturization of experiments to perform reactions in microreactors and in microdroplets,
- automation of experiments and ubiquity of computing, which enable the development of closed loop systems with advanced design-of-experiments algorithms,
- and, finally, the emergence of additive manufacturing as a method of producing reaction ware.

Thus, the transition to the new experimental paradigm is a nexus of several scientific and technical developments that came to maturity at a more or less the same time. Here we give a concise overview of the techniques and methods that underpin the new experimental platforms, and focus on the emerging trends and gaps in our technical capabilities and knowledge. The focus of this paper is on new experimental and software platforms that open completely new opportunities in optimization and discovery. The highly active and exciting area of automation and control of large-scale chemical processes is out of scope of this paper.

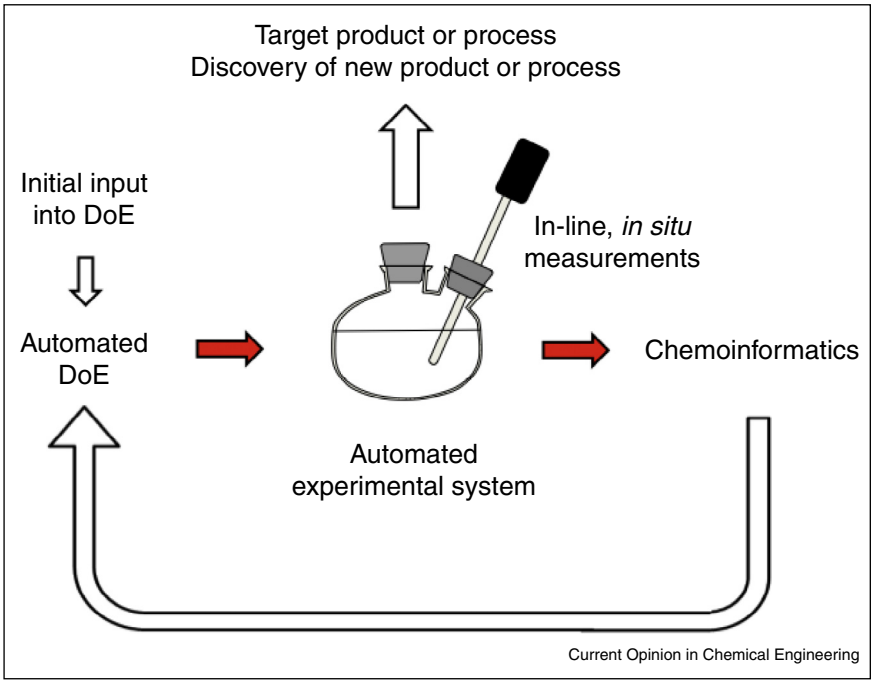
Hardware for self-optimization of closed-loop systems

Labware

The new experimental platforms for optimization and discovery require very different lab-ware from the conventional round bottom flasks or even simple flow microreactors. Such experimental systems combine automation of reaction ware and of analytical instruments to create a closed-loop control system, schematically shown in Figure 1. The control algorithm in this case is implementing a specific strategy of experimental design (DoE).

An example of a highly automated and instrument-rich reaction ware is the Automatic Continuous Online Monitoring of Polymerization Reactions (ACOMP) system, which included a recirculation loop with automated sampling and sample preparation (dilution) to allow measurements of refractive index, UV–vis, viscosity, multi-angle light scattering and gel-permeation chromatography (GPC), all connected in series [9••]. This reactor system was made for close to real-time monitoring of

Figure 1



A schematic diagramme of elements of a closed-loop experimental system.

polymerization reactions without human interaction and was able to operate in batch, semi-batch and continuous experiments [10–12].

The important feature of such a system is a combination of analytical techniques that provide complementary information for a complex chemical system and largely include fast analytical methods based on optical and spectroscopic measurements, but also include more slow methods, such as GPC. ACOMP represents an example of an automated reaction system with several in-line analytical instruments, but it lacks two other necessary components for a closed-loop optimization: chemoinformatics and design of experiments.

The ACOMP system was built using a conventional approach to laboratory automation, using signal conditioning on data acquisition National Instruments platform and LabVIEW software interface. A similar approach has been followed in most systems reported to date, in which self-optimization has been implemented. These systems cover different areas of chemistry: polymerization reaction [13•,14••], heterogeneous catalytic processes [1,2,5], droplet formation [15,16] assembly of nanoparticles [17] and even multi-step syntheses [4,18–21]. As we shall show below, this approach to automation is readily extendable to include chemoinformatics and design of experiments, using data exchange standards, such as OPC. This brings the developed automated optimization and discovery systems very close to industrial production systems with respect to control hardware architectures.

A significant departure from this approach is the use of dedicated robotic experimental systems, which could either involve parallel batch or flow experiments. Such systems are developed for high-throughput experiments and enable rapid discovery of new reactions [22,23,24•], or optimization of process conditions of known reactions [25,26]. A flow array system has been used to discover a new inorganic cluster [7•]. A recent example from Merck & Co. Inc. shows that high throughput robotic experiments could be performed with very small quantities of reagents, making this technology highly appealing as a platform for chemical discovery [27].

Another significant recent development is the ability to reproduce labware through additive manufacturing, such as 3D printing [28••]. This enables not only rapid development of ideas for new reactor types, but also sharing of ideas across many laboratories through universal access to identical labware. The uptake of this technology strongly depends on the continuing decrease in the cost of 3D printing equipment, the ability to print chemically resistant and thermally stable devices, and on the development of user resources and user communities with an open innovation philosophy.

On-line, *in situ* and in-line analytical methods

Spectroscopic methods are providing fast and non-destructive chemical and process information about reacting systems, which is critical for the implementation of self-optimization experiments. The use of in-line optical spectroscopy methods for real time sensing and control has been studied intensively over the last decade [2,29••,30•,31••,32]. Major drawbacks of most spectroscopic analytical methods include relatively low sensitivity (excluding fluorescence spectroscopy) and the need for calibration models for multi-component reaction systems.

The problem of low sensitivity is particularly acute for tasks requiring quantification of minor components, such as optimization of medicinal syntheses. It is less of an issue when large variations in the measured variables are monitored, such as moisture (e.g. to enable real-time optimization of drying), pH, conductivity, etc. The latter methods carry less 'chemical' information compared to spectroscopic methods that provide evidence of molecular structure and concentration. The significant advantage of in-line or *in situ* spectroscopic methods over off-line methods is revealed in the case of unstable samples, when degradation of the substance of interest makes off-line analysis highly unreliable [33]. This highlights the importance of sampling time in the range of applications of interest for the new reaction platforms, and allows to further differentiate from the conventional applications of sensors in real-time control.

Here the aim is to either determine if a reaction outcome satisfies the criteria set out for a new process (in the process optimization/discovery scenario) or for a new product function (in the product discovery scenario). In this case, one can imagine that experimental system may be kept in steady state or idle till analysis results are available, prior to a new set of input conditions is established. This is rather different from real-time process control, when sampling time must be commensurate with the system's response time to allow predictive control.

More recently the use of in-line NMR spectroscopy in flow chemistry has been demonstrated [34••,35•,36]. This represents a significant step forward, since NMR is a direct technique and significantly simpler calibration methods are required. At present, low sensitivity of bench-top NMR instruments is a significant limitation.

Besides optical and NMR spectroscopy, other analytical techniques were implemented as online analytical methods in new experimental set-ups, such as online HPLC [3•], GC [2], or MS [37•,38]. Key challenges for these techniques are the speed of sampling, as well as robustness and reliability during the online monitoring. However, significant reductions in sampling time could be attained with multiplexing samples even for the traditionally slow methods, such as liquid chromatography. For LC–MS the sampling time could be reduced to 5–22 s per sample, using multiple injections in a single run technique [27].

A summary of different analytical techniques currently applied for in-line and *in situ* analysis in closed-loop optimization/discovery applications with the corresponding typical issues is shown in Table 1. Here, the stated characteristics of the techniques do not span all their capabilities as off-line methods!

Most studies on closed-loop optimization published to date use rather primitive search or optimization algorithms, that require significant numbers of experiments and are only suitable for cheap experiments and simple problems with very few input variables, that is, low dimensionality of experimental space. The recently emerged techniques of machine learning offer a tremendous opportunity to develop highly efficient self-optimization experimental systems [39•].

Software for self-optimization of closed-loop system

Automated design of experiment, or a decision-making algorithm, is the third necessary component of a closed-loop optimization system. Conventional DoE algorithms familiar to most experimental scientists are based on the ideas of factorial design of experiments and linear optimization. For a number of input variables a range of values to be tested is defined, a matrix of experiments is generated and the complete matrix of results is analyzed after experiments are performed either in parallel or one-by-one. This leads to a problem of explosion in the number of required experiments when the number of input parameters that must be tested is large.

Self-optimization instrumentation and experimental philosophy differs from such classical DoE in the sense that there is an opportunity to *learn* from each experiment and update a DoE model in a sequential fashion. In this case a DoE algorithm may take the form of an optimization algorithm. This strategy has been implemented in several studies. Thus, simple linear algorithms have been used for optimization of flow and temperature in catalytic reactions [3•,4,5]. In the case of the more complex problems with a large number of independent input variables such linear algorithms, for example a simplex algorithm [40,41], would lead to an unreasonably large number of experiments.

A more advanced approach to sequential design of experiments is to use machine learning optimization algorithms that allow taking advantage of the results of the previous experiments. The class of algorithms that are suitable for this task are either global or target optimization algorithms. One of the most prominent algorithms for such sequential optimization is the

Table 1

A summary of sensors and analytical techniques used in closed-loop optimization

Technique	Type of information	Sensitivity	Speed of acquisition	Limitations
Mid-IR	Chemical identity, concentration, gas, liquid or solid samples	$\sim 10^{-1}$ mol%	~1 s	Short fibres. Intolerant to water
Near-IR	Chemical identity, concentration	$\sim 10^{-1}$ mol%	~1 s	Less informative than Mid-IR, tolerant to water
Raman	Chemical identity, crystal structure, concentration. Solid and liquid samples.	$\sim 10^{-1}$ mol% Potentially, to individual molecules in the case of SER(R)S	~1–100 s	Fluorescence masking Raman signal.
UV–vis	Chemical identity, concentration	10^{-4} mol%	<1 s	Limited number of species
NMR	Molecular structure, identification of unknown compounds, concentration	$\sim 10^{-3}$ mol%	~10 s	At present flow method is limited in sensitivity and resolution due to low field.
GC	Concentrations	$> 10^{-6}$ mol%	10–1500 s	Typically – slow. Cannot identify unknown compounds. Difficult to automate.
HPLC	Concentrations	$> 10^{-6}$ mol%	200–1500 s	Long method development times. Must be combined with MS for proof of molecular identity.
MS MS/MS	Concentration Chemical identity	$> 10^{-8}$ mol%	~5–20 s	Requires chemoinformatics expertise. MS/MS is more informative, but few process instruments on the market. Difficult method development on more advanced methods.
Process sensors:				
		Pressure		
		Temperature		
		pH		Sensor fouling
		Conductivity		Sensor fouling
		Viscosity		Requires specific in-line cell
		Dynamic light scattering		Difficult for in-line, requires dilution
		Ultrasound		Presently used only for level. Underdeveloped.

efficient global optimization (EGO) algorithm [42]. Since its introduction, the algorithm has been adapted for different types of optimization problems, including target optimization. One idea is to make use of the concepts of desirability [43] and virtual observations [44] to construct an algorithm capable of identifying and, with each iteration, improving on a cluster of solutions that best associate with target values [45]. Even though the algorithm undoubtedly explores globally throughout the search, it is not designed to actively search for solutions that would allow one to gain the most information about the underlying process (i.e. solutions optimal in terms of experimental design).

Machine learning algorithms require construction of a statistical surrogate model, which is then used to predict the outcomes of the future experiments, the decision making process. Gaussian Process is a popular surrogate model type as it provides a principled way of assessing uncertainty of the model and has successfully been used in many optimization problems, including chemistry-related problems [46–48]. The advantage of using Gaussian Processes is in the ability to deal efficiently with both demands on the sampling criterion: exploration and exploitation. A similar approach is to use a surrogate based online evolutionary algorithm [39•,42,49•,50•,51–54]. The attractive features of this algorithm are that (*i*) it has an evolutionary algorithm at the core, capable of solving multi-dimensional multimodal problems, and (*ii*) attempts to strike a balance between the need to reduce the amount of expensive evaluations and the need to improve on the quality of the surrogate model. Although unfamiliar to most practicing chemists, such advanced optimization methods have already been applied in several chemical processes. Thus, Gaussian Processes has been used in prediction of quality of polypropylene [55•], in simulation of catalytic batch etherification reaction [56], in real-time prediction of properties for industrial rubber mixing processes [57] and in screening of new additives for a Friedel–Crafts catalyst [58••].

Recently a new combination of Gaussian Processes, mutual information and a genetic algorithm for multi-target optimization has been published — a multi-objective active learner (MOAL) algorithm [39•]. This algorithm outperformed another published algorithm, the surrogate

based on-line evolutionary algorithm (SOEA) [59,60•, 61,62], compared on the basis of *in silico* tests with different mathematical functions. Due to its objective of multi-target optimization, the MOAL algorithm can be used as a decision making software in discovery of products with specified properties or new chemical reaction routes with a specified target.

MOAL algorithm was recently applied as a decision making element of a closed-loop system in a process of experimental discovery of recipes for semi-batch emulsion co-polymerization [63]. For this the algorithm was extended to incorporate a Gaussian Process (GP) binary classification model [64]. The task of the classification model is to learn the regions of feasible experiments, assuming that the experiments are done for a problem with unknown bounds of experimental input variables. The classification was performed in reduced space via application of a dimensionality reduction technique. Normally this can be done by principal component analysis (PCA) [65] or multidimensional sealing (MDS) [66] techniques. These techniques are limited by their global linearity. To use classification model in a highly dimensional nonlinear case, manifold learning techniques, such as locally linear embedding (LLE) [67] and Isomap [68•] can be used. In the MOAL algorithm the SIsomap [69] was incorporated as a classification model to speed up the process of identification of the feasible experimental range. This system was able to find a feasible recipe for a two-monomer semi-batch co-polymerization within 20 experiments. The result is important in the context of the large number of input variables — in this case 14 input variables were treated as independent and used in the optimization.

Outlook

Closed-loop experimental systems have recently emerged as a powerful tool for optimization of process conditions and as a potential discovery platform. This platform thus far was only developed for sequential experiments, where advantage is taken of the knowledge obtained in previous experiments. Here, the more sophisticated statistical algorithms are capable to design experiments and minimize the number of experiments required for optimization or discovery. The area of automated experimental discovery is still in its nascent state. However, the range of the already demonstrated applications shows the potential significance of this approach: optimization of heterogeneous and homogeneous catalytic reactions, synthesis of new materials, synthesis of new bioactives, discovery of new process recipes for semi-batch reactions.

Current state of technology is characterized by several limitations, mainly in its analytical and decision making components. At present a narrow range of analytical techniques can be used in automated experiments. Development of fast and cheap in-line LC–MS and NMR techniques, as well as methods of observing product properties (particulates properties, for example) in real time and under reaction conditions, would significantly broaden the range of applications. The robotic experimental platforms available on the market are already better than what can at present be realistically used in closed-loop systems due to limitations of both the analytical instrumentation and in the design of experiments algorithms. A serious current limitation of the decision-making algorithms is the lack of use of *a priori* knowledge of the chemical systems in design of experiments. These current limitations are all being addressed simultaneously and, undoubtedly, the technology of automated optimization and discovery in molecular systems will see rapid development over the next few years.

Acknowledgements

This work was in part funded by EPSRC project 'Closed Loop Optimization for Sustainable Chemical Manufacture' [EP/L003309/1].

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Bourne RA, Skilton RA, Parrott AJ, Irvine DJ, Poliakoff M: **Adaptive process optimization for continuous methylation of alcohols in supercritical carbon dioxide**. *Org Process Res Dev* 2011, **15**:932-938.

2. Skilton RA, Parrott AJ, George MW, Poliakoff M, Bourne RA: **Real-time feedback control using online attenuated total reflection Fourier transform infrared (ATR FT-IR) spectroscopy for continuous flow optimization and process knowledge**. *Appl Spectrosc* 2013, **67**:1127-1131.

3. • McMullen JP, Stone MT, Buchwald SL, Jensen KF: **An integrated microreactor system for self-optimization of a Heck reaction: from micro- to mesoscale flow systems**. *Angew Chem Int Ed Engl* 2010, **49**:7076-7080.

This is one of the first papers in which a microreactor was used in a self-optimized, closed-loop experimental system.

4. Smith CJ, Nikbin N, Ley SV, Lange H, Baxendale IR: **A fully automated, multistep flow synthesis of 5-amino-4-cyano-1,2,3-triazoles**. *Org Biomol Chem* 2011, **9**:1938-1947.

5. Parrott AJ, Bourne RA, Akien GR, Irvine DJ, Poliakoff M: **Self-optimizing continuous reactions in supercritical carbon dioxide**. *Angew Chem Int Ed Engl* 2011, **50**:3788-3792.

6. Stevens JG, Bourne RA, Twigg MV, Poliakoff M: **Real-time product switching using a twin catalyst system for the hydrogenation of furfural in supercritical CO_2**. *Angew Chem Int Ed Engl* 2010, **49**:8856-8859.

7. • Richmond CJ, Miras HN, de la Oliva AR, Zang H, Sans V, Paramonov L, Makatsoris C, Inglis R, Brechin EK, Long D-L *et al.*: **A flow-system array for the discovery and scale up of inorganic clusters**. *Nat Chem* 2012, **4**:1037-1043.

This is one of the first paper in which a closed-loop experimental system was used for discovery of new inorganic materials.

8. Czechtizky W, Dedio J, Desai B, Dixon K, Farrant E, Feng Q, Morgan T, Parry DM, Ramjee MK, Selway CN *et al.*: **Integrated synthesis and testing of substituted xanthine based DPP4 inhibitors: application to drug discovery**. *ACS Med Chem Lett* 2013, **4**:768-772.

9. •• Alb AM, Drenski MF, Reed WF: **Automatic continuous online monitoring of polymerization reactions (ACOMP)**. *Polymer Int* 2008, **57**:390-396.
This is the most instrument-rich automated experimental system with in-line monitoring which provides many sensor outputs for development of process understanding in a highly complex chemical system.

10. Kreft T, Reed WF: **Predictive control and verification of conversion kinetics and polymer molecular weight in semi-batch free radical homopolymer reactions**. *Eur Polym J* 2009, **45**:2288-2303.

11. McFaul CA, Drenski MF, Reed WF: **Online, continuous monitoring of the sensitivity of the LCST of NIPAM-Am copolymers to discrete and broad composition distributions**. *Polymer* 2014, **55**:4899-4907.

12. Reed WF, Alb AM: *Monitoring Polymerization Reactions*. Hoboken, NJ: Wiley-VCH Verlag GmbH & Co. KGaA; 2013.

13. • Vieira RAM, Sayer C, Lima EL, Pinto JC: **Closed-loop composition and molecular weight control of a copolymer latex using near-infrared spectroscopy**. *Ind Eng Chem Res* 2002, **41**:2915-2930.

14. •• Nogueira ES, Borges CP, Pinto JC: **In-line monitoring and control of conversion and weight-average molecular weight of polyurethanes in solution step-growth polymerization based on near infrared spectroscopy and torquemetry**. *Macromol Mater Eng* 2005, **290**:272-282.

15. Miller E, Rotea M, Rothstein JP: **Microfluidic device incorporating closed loop feedback control for uniform and tunable production of micro-droplets**. *Lab Chip* 2010, **10**:1293-1301.

16. Moyle TM, Walker LM, Anna SL: **Controlling thread formation during tipstreaming through an active feedback control loop**. *Lab Chip* 2013, **13**:4534-4541.

17. Sans V, Glatzel S, Douglas FJ, Maclaren D, Lapkin A, Cronin L: **Non-equilibrium dynamic control of gold nanoparticle and hyper-branched nanogold assemblies**. *Chem Sci* 2014, **5**:1153-1157.

18. Ingham RJ, Battilocchio C, Hawkins JM, Ley SV: **Integration of enabling methods for the automated flow preparation of piperazine-2-carboxamide**. *Beilstein J Org Chem* 2014, **10**:641-652.

19. Lange H, Carter CF, Hopkin MD, Burke A, Goode JG, Baxendale IR, Ley SV: **A breakthrough method for the accurate addition of reagents in multi-step segmented flow processing**. *Chem Sci* 2011, **2**:765-769.

20. Guetzoyan L, Ingham RJ, Nikbin N, Rossignol J, Wolling M, Baumert M, Burgess-Brown N, Strain-Damerell CM, Shrestha L, Brennan PE *et al.*: **Machine-assisted synthesis of modulators of the histone reader BRD9 using flow methods of chemistry and frontal affinity chromatography**. *Med Chem Comm* 2014, **5**:540-546.

21. Ingham RJ, Battilocchio C, Fitzpatrick DE, Sliwinski E, Hawkins JM, Ley SV: **A systems approach towards an intelligent and self-controlling platform for integrated continuous reaction sequences**. *Angew Chem Int Ed Engl* 2015, **54**:144-148.

22. Friedfeld MR, Shevlin M, Hoyt JM, Krska SW, Tudge MT, Chirik PJ: **Cobalt precursors for high-throughput alkene hydrogenation catalysts**. *Science* 2013, **342**:1076-1080.

23. Dirocco D, Dykstra K, Krska S, Vachal P, Conway DV, Tudge M: **Late-stage functionalization of biologically active heterocycles through photoredox catalysis**. *Angew Chem Int Ed Engl* 2014, **53**:4802-4806.

24. • Collins KD, Gensch T, Glorius F: **Contemporary screening approaches to reaction discovery and development**. *Nat Chem* 2014, **6**:859-871.

25. Bellomo A, Celebi-Olcum N, Bu X, Rivera N, Ruck RT, Welch CJ, Houk KN, Dreher SD: **Rapid catalyst identification for the synthesis of the pyrimidinone core of HIV integrase inhibitors**. *Angew Chem Int Ed Engl* 2012, **51**:6912-6915.

26. Schmink JR, Bellomo A, Berritt S: **Scientist-led high-throughput experimentation (HTE) and its utility in academia and industry**. *Aldrichim Acta* 2013, **46**:71-76.

27. Santanilla BA, Regalado EL, Pereira T, Shevlin M, Bateman K, Campeau L-C, Schneeweis J, Berritt S, Shi Z-C, Nantermet P *et al.*: **Nanomole-scale high-throughput chemistry for the synthesis of complex molecules**. *Science* 2015, **347**:49-53.

28. •• Kitson PJ, Marshall RJ, Long D, Forgan RS, Cronin L: **3D printed high-throughput hydrothermal reactionware for discovery, optimization, and scale-up**. *Angew Chem Int Ed Engl* 2014, **53**:12723-12728.
This is the first paper on 3D printing of labware. Most notably the group of Cronin developed polypropylene printed devices for improved chemical resistance.

29. •• Mozharov S, Nordon A, Littlejohn D, Wiles C, Watts P, Dallin P, Girkin JM: **Improved method for kinetic studies in microreactors using flow manipulation and noninvasive raman spectrometry**. *J Am Chem Soc* 2011, **133**:3601-3608.
The 'push-out' method of collecting kinetic information allows to generate large amounts of data in a flow experiment, making it closer to conventional batch kinetic experiment. This contributes to the debate whether flow or batch experiments are better for collecting the intrinsic kinetic information. The difficulty of the 'push-out' method is in the impact of specific residence time distribution of the device used on observed reaction rates. The assumption of plug-flow is very difficult to attain in conventional 'flow' reactors used by most groups. Hence, the obtained kinetic data is not the 'true' kinetics, but must be treated mathematically having the experimental residence time distribution.

30. • McGill CA, Nordon A, Littlejohn D: **Comparison of in-line NIR, Raman and UV–visible spectrometries, and at-line NMR spectrometry for the monitoring of an esterification reaction**. *Analyst* 2002, **127**:287-292.

31. •• Carter CF, Lange H, Ley SV, Baxendale IR, Wittkamp B, Goode JG, Gaunt NL: **ReactIR flow cell: a new analytical tool for continuous flow chemical processing**. *Org Process Res Dev* 2010, **14**:393-404.
The reactIR system by Mettler Toledo has been the work-horse of process chemists for at least a decade. But, access to IR data from miniaturised flow experiments was always a problem. The only other alternative, for a long time, was Golden Gate ATR IR cell, which allowed measurements in a ca. 30 μL volume flow cell at up to 250 °C and 100 bar. The group of S. Ley developed a miniature flow-IR system for in-line measurements, which is now commercially available from Mettler Toledo.

32. Moore JS, Jensen KF: **''Batch'' kinetics in flow: online IR analysis and continuous control**. *Angew Chem Int Ed Engl* 2014, **53**:470-473.

33. Chen Y, Zhou GX, Brown N, Wang T, Ge Z: **Study of lactol activation by trifluoroacetic anhydride via in situ Fourier transform infrared spectroscopy**. *Anal Chim Acta* 2003, **497**:155-164.

34. •• Danieli E, Perlo J, Duchateau LL, Verzijl GKM, Litvinov VM, Blümich B, Casanova F: **On-line monitoring of chemical reactions by using bench-top nuclear magnetic resonance spectroscopy**. *ChemPhysChem* 2014, **15**:3060-3066.
This is the first paper in which on-line monitoring of chemical reactions by bench-top NMR was presented. The following paper (34) also shows how to include flow NMR measurements into a closed-loop experimental system.

35. • Sans V, Porwol L, Dragone V, Cronin L: **A self optimizing synthetic organic reactor system using real-time in-line NMR spectroscopy**. *Chem Sci* 2015, **6**:1258-1264.

36. Danieli E, Perlo J, Blümich B, Casanova F: **Small magnets for portable NMR spectrometers**. *Angew Chem Int Ed Engl* 2010, **49**:4133-4135.

37. • Browne DL, Wright S, Deadman BJ, Dunnage S, Baxendale IR, Turner RM, Ley SV: **Continuous flow reaction monitoring using an on-line miniature mass spectrometer**. *Rapid Commun Mass Spectrom* 2012, **26**:1999-2010.
This is one of the first papers in which a miniaturised MS was used in flow chemistry.

38. Silva BV, Violante FA, Pinto AC, Santos LS: **The mechanism of Sandmeyer's cyclization reaction by electrospray ionization**

mass spectrometry. *Rapid Commun Mass Spectrom* 2011, **25**:423-428.

39. • Peremezhney N, Hines E, Lapkin AA, Connaughton C: **Combining Gaussian processes, mutual information and a genetic algorithm for multi-target optimization of expensive-to-evaluate functions**. *Eng Optim* 2014, **46**:1593-1607.

40. Nelder JA, Mead R: **A simplex method for function minimization**. *Comput J* 1965, **7**:308-313.

41. Routh MW, Swartz PA, Denton MB: **Performance of the super modified simplex**. *Anal Chem* 1977, **49**:1422-1428.

42. Jones DR, Schonlau M, Welch WJ: **Efficient global optimization of expensive black-box functions**. *J Glob Optim* 1998, **13**:455-492.

43. Harrington ECJ: **The desireability function**. *Ind Qual Control* 1965, **10**:494-498.

44. Cox DD, John S: **SDO: a statistical method for global optimization**. In *Multidisciplinary Design Optimization: State of the Art.* Edited by Alexandrow NMH. Philadelphia: SIAM; 1997:315-329.

45. Bock H-H, Gaul W, Vichi M, Arabie P, Baier D, Critchley F, Decker R, Diday E, Greenacre M, Lauro C *et al.*: *Classification as a Tool for Research*. Berlin/Heidelberg: Springer; 2010.

46. Xia W, Luo B, Liao X: **An enhanced global optimization method based on Gaussian process and its application of warpage control in injection molding**. *2010 IEEE Int. Conf. Inf. Autom.* 2010 http://dx.doi.org/10.1109/ICINFA.2010.5512143.

47. Villemonteix J, Vazquez E, Walter E: **An informational approach to the global optimization of expensive-to-evaluate functions**. *J Glob Optim* 2009, **44**:509-534.

48. Quayle OR: **The parachors of organic compounds. An interpretation and catalogue**. *Chem Rev* 1953, **53**:439-589.

49. • El-Beltagy MA, Keane AJ: **Evolutionary Optimization for Computationally expensive problems using Gaussian Processes**. In *Proceedings of International Conference on Artificial Intelligence*. CSREA Press; 2001:708-714.

50. • Osborne MA, Garnett R, Roberts SJ: **Gaussian processes for global optimization**. In *Proceedings of 3rd International Conference on Learning and Intelligent Optimization*. 2009:1-15.

51. Huang D, Allen TT, Notz WI, Zeng N: **Global optimization of stochastic black-box systems via sequential kriging meta-models**. *J Glob Optim* 2006, **34**:441-466.

52. Frean M, Boyle P: **Using Gaussian processes to optimize expensive functions**. In In *Proceedings of the 21st Australasian Joint Conference on Articial Intelligence: Advances in Articial Intelligence (AI '08)*. Edited by Wobcke W, Zhang M. *Proceedings of the 21st Australasian Joint Conference on Articial Intelligence: Advances in Articial Intelligence (AI '08)* Berlin, Heidelberg: Springer-Verlag; 2008:258-267.

53. Knowles J: **ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems**. *IEEE Trans Evol Comput* 2006, **10**:50-66.

54. Srinivas N, Krause A, Seeger M: **Gaussian process optimization in the bandit setting: no regret and experimental design**. In *Proceedings of the 27th International Conference on Machine Learning*; *Haifa, Israel: 2010:1015-1022. also published in IEEE Transactions on Information Theory. 2012, 58: 3250–3265*.

55. • Ge Z, Chen T, Song Z: **Quality prediction for polypropylene production process based on CLGPR model**. *Control Eng Pract* 2011, **19**:423-432.

56. Chang JS, Liu WL, Tang JK, Lin WC, Chiou YJ: **Expected improvement in efficient experimental design supported by a global optimizer**. *J Taiwan Inst Chem Eng* 2014, **45**:1369-1388.

57. Liu Y, Gao Z: **Real-time property prediction for an industrial rubber-mixing process with probabilistic ensemble gaussian process regression models**. *J Appl Polym Sci* 2015, **132**:1-9.

58. •• Omata K: **Screening of new additives of active-carbon-supported heteropoly acid catalyst for friedel-crafts reaction by Gaussian process regression**. *Ind Eng Chem Res* 2011, **50**:10948-10954.
This is one of the first papers which exemplifies the use of advanced statistical methods that could be used for discovery of new chemical systems. The number of improtant variables was reduced from 16 to 5 using a PCA method and then a Gaussian Process surrogate model was trained on the available experimental data and the reduced number of input variables to predict the likely new active catalyst.

59. Ackley DH: *A Connectionist Machine for Genetic Hillclimbing*. Norwell, MA: Kluwer Academic Publishers; 1987.

60. • Jamil M, Yang XS: **A literature survey of benchmark functions for global optimisation problems**. *Int J Math Model Numer Optim* 2013, **4**:150-194.

61. Laguna M, Martí R: **Experimental testing of advanced scatter search designs for global optimization of multimodal functions**. *J Glob Optim* 2005, **33**:235-255.

62. Dixon LCW, Price RC: **Truncated Newton method for sparse unconstrained optimization using automatic differentiation**. *J Optim Theory Appl* 1989, **60**:261-275.

63. Houben C, Peremezhney N, Zubov A, Kosek J, Lapkin AA: *Closed-loop Multi-target Optimisation for Discovery of New Emulsion Polymerisation Recipes*. 2015:. (under review).

64. Rasmussen CE, Williams CKI: *Gaussian Processes for Machine Learning*. MIT Press; 2006.

65. Jolliffe IT: *Principal Component Analysis*. New York: Springer; 2002.

66. Borg I, Groenen P: *Modern Multidimensional Scaling: Theory and Application*. New York: Springer; 2005.

67. Roweis ST, Saul LK: **Nonlinear dimensionality reduction by locally linear embedding**. *Science* 2000, **290**:2323-2326.

68. • Tenenbaum JB, Silva V, De Langford JC: **A global geometric framework for nonlinear dimensionality reduction**. *Science* 2000, **290**:2319-2323.

69. Geng X, Zhan D-C, Zhou Z-H: **Supervised nonlinear dimensionality reduction for visualization and classification**. *IEEE Trans Syst Man Cybern B* 2005, **35**:1098-1107.