



# Automated self-optimisation of multi-step reaction and separation processes using machine learning



Adam D. Clayton<sup>a</sup>, Artur M. Schweidtmann<sup>b</sup>, Graeme Clemens<sup>c</sup>, Jamie A. Manson<sup>a</sup>, Connor J. Taylor<sup>a</sup>, Carlos G. Niño<sup>d</sup>, Thomas W. Chamberlain<sup>a</sup>, Nikil Kapur<sup>d</sup>, A. John Blacker<sup>a</sup>, Alexei A. Lapkin<sup>e</sup>, Richard A. Bourne<sup>a,\*</sup>

<sup>a</sup> Institute of Process Research and Development, School of Chemistry & School of Chemical and Process Engineering, University of Leeds, Leeds LS2 9JT, UK

<sup>b</sup> Aachener Verfahrenstechnik – Process Systems Engineering, RWTH Aachen University, Aachen, Germany

<sup>c</sup> Chemical Development, Pharmaceutical Technology & Development, AstraZeneca, Macclesfield Campus, SK10 2NA, UK

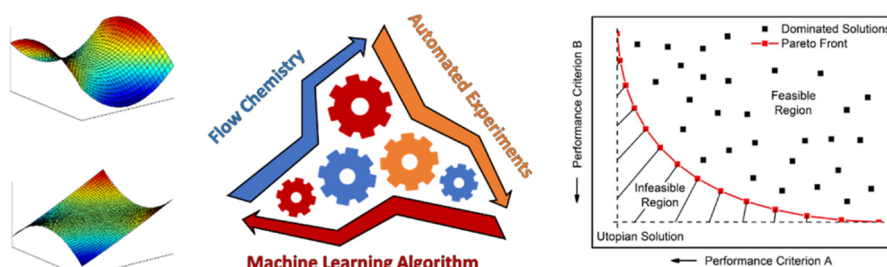
<sup>d</sup> School of Mechanical Engineering, University of Leeds, Leeds LS2 9JT, UK

<sup>e</sup> Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge CB3 0AS, UK

## HIGHLIGHTS

- Multi-objective algorithm applied to the self-optimisation of complex pharmaceutical processes.
- Algorithm targeted automated reactor productivity and downstream purification.
- Pareto front rapidly generated for three competing objectives.

## GRAPHICAL ABSTRACT



## ARTICLE INFO

### Keywords:

Automated flow reactor  
Environmental chemistry  
Machine learning  
Reaction engineering  
Sustainable chemistry

## ABSTRACT

There has been an increasing interest in the use of automated self-optimising continuous flow platforms for the development and manufacture in synthesis in recent years. Such processes include multiple reactive and work-up steps, which need to be efficiently optimised. Here, we report the combination of multi-objective optimisation based on machine learning methods (TSEMO algorithm) with self-optimising platforms for the optimisation of multi-step continuous reaction processes. This is demonstrated for a pharmaceutically relevant Sonogashira reaction. We demonstrate how optimum reaction conditions are re-evaluated with the changing downstream work-up specifications in the active learning process. Furthermore, a Claisen-Schmidt condensation reaction with subsequent liquid-liquid separation was optimised with respect to three-objectives. This approach provides the ability to simultaneously optimise multi-step processes with respect to multiple objectives, and thus has the potential to make substantial savings in time and resources.

## 1. Introduction

Since the industrial revolution, the use of mechanised tools to improve manufacturing processes has been a predominant feature in many

areas of technology. With the goal of transferring the concept of automation to chemical synthesis, closed-loop reaction systems have become more commonplace, providing advantages such as greater reaction control and process safety [1]. Examples of using feedback loops

\* Corresponding author.

E-mail address: [r.a.bourne@leeds.ac.uk](mailto:r.a.bourne@leeds.ac.uk) (R.A. Bourne).

<https://doi.org/10.1016/j.cej.2019.123340>

Received 13 August 2019; Received in revised form 29 October 2019; Accepted 31 October 2019

Available online 02 November 2019

1385-8947/ Crown Copyright © 2019 Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

for the automated optimisation of chemical reactions dates back to the late 1970s, and the pioneering work of Winicov et al. [2]. In this work, reaction parameters such as temperature, stirring and reagent addition were optimised using a Nelder-Mead SIMPLEX algorithm. Since then, developments in the areas of automated laboratory hardware and control software have driven these systems to significantly evolve [3]. More recently, continuous flow systems have been combined with on-line/in-line analytics and optimisation algorithms to automate reaction optimisation (commonly referred to as 'self-optimisation') [4,5]. The ability of self-optimisation to efficiently identify optimal operating conditions in a multivariate parameter space has presented many opportunities for more efficient process development. Such advantages align with the rising interest in continuous flow chemistry towards the 'greener' synthesis of active pharmaceutical ingredients (APIs), and offer the potential to reduce the drug development timeline [6–8].

Process development in the pharmaceutical industry must simultaneously consider multiple performance criteria based on conflicting economic and environmental objectives [9]. This applies to the whole process including non-reactive unit operations, such as work-up. Despite this, the majority of self-optimisation applications to date have focused on single-objective optimisation of single-step reactions, utilising the following algorithms: model-based design of experiments [10–12], SNOBFIT [13–16], Nelder-Mead SIMPLEX or variations thereof [17–23]. Further, these algorithms are not data-efficient as they do not utilise all the available experimental data to build a global surrogate model. Thus, they are not well-suited for expensive-to-evaluate chemical synthesis problems, particularly those involving complex multi-step, and pharmaceutically relevant processes [24].

In our previous work, we investigated the use of a data efficient Bayesian optimisation algorithm for the self-optimisation of chemical reactions with two competing performance criteria [25]. The Thompson sampling efficient multi-objective (TSEMO) algorithm [26,27] builds machine learning surrogate models, i.e. Gaussian processes (GPs), based on all available data and compares favourably with other algorithms such as EHI [28] and ParEGO [29]. TSEMO was integrated with our automated experimental platform and enabled the simultaneous optimisation of space–time yield (STY) with E-factor or impurity content for two exemplar reactions. Notably, the complete trade-off curve (Pareto front [30]) highlighting the compromise between the objectives was identified in a practical number of experiments, overcoming the issues associated with the scalarisation of multiple objectives [31]. TSEMO was also effectively used in the workflow on solvent selection for optimal reactivity and selectivity [32], and in the context of optimisation of life cycle environmental impacts vs. process economics for computer-aided process design [33]. Based on these results, we hypothesised that the TSEMO algorithm would be well-suited for the optimisation of multi-step continuous reaction processes. Herein, we describe the application of this methodology to: (i) a pharmaceutically relevant Sonogashira reaction; (ii) a multi-step Claisen-Schmidt condensation reaction with in-line liquid–liquid extraction.

## 2. Experimental

### 2.1. TSEMO algorithm

The TSEMO algorithm is designed to solve expensive-to-evaluate black-box multi-objective optimisation problems. The algorithm utilises a Bayesian methodology employing GPs as the surrogate model. The surrogate models are coupled with acquisition functions which are optimised in lieu of the real process to suggest the next point of evaluation. TSEMO makes use of Thompson Sampling to determine the next point of evaluation through random sampling of the posterior distribution of the GPs. The sample is then optimised through use of the NSGA-II algorithm to propose a set of candidate points. NSGA-II is an elitist genetic algorithm which solves multi-objective optimisation

problems using Pareto ranking and crowding distance computations. The candidate which maximises the hypervolume improvement is selected as the next point to sample. This process continues until the algorithm reaches a pre-defined maximum number of process/function evaluations.

The TSEMO algorithm was combined with an automated continuous flow platform to enable the closed-loop multi-objective optimisation of chemical processes. The algorithm was operated in batch-sequential mode, which proposes multiple sampling points at each iteration. In previous work, we demonstrated that there was a minimal reduction in performance when batches of four evaluations were used compared to single evaluations [26]. Therefore, batches of four experiments were used in these case studies, as this enabled faster optimisations by parallelising the start of an experiment with the analysis of the previous experiment.

### 2.2. Self-optimising platform

Reagents were pumped using JASCO PU980, HiTec Zang SyrDos and/or Knauer AZURA HPLC pumps, and were mixed in Swagelok SS-100-3 tee-pieces. Tubular reactors were constructed from Polyfon PTFE tubing (0.1 cm ID) and fitted to a Cambridge Reactor Design Polar Bear Flow Synthesiser. Miniature continuous stirred tank reactor (CSTR) cascades were custom-built and are detailed below. The reactor was maintained under a fixed back pressure using an Upchurch Scientific back pressure regulator. A Zaiput SEP-10 liquid–liquid membrane-based separator fitted with a PTFE membrane (0.5  $\mu\text{m}$  pore size) was used for in-line phase separation when required. Sampling of the organic phase was achieved using a VICI Valco EUDA-CI4W.06 sample loop with a 0.06  $\mu\text{L}$  injection volume. Quantitative analysis was performed on an Agilent 1100 series HPLC instrument (HPLC methods for each case study are provided in the ESI). Steady state was monitored using a Kaiser RxN1 785 nm Raman System when required. The automated reactor was controlled by a custom written MATLAB program, within which the TSEMO algorithm was implemented.

### 2.3. General optimisation procedure

An optimisation program was written in MATLAB that controlled the pump flow rates and reactor temperature, determined steady state, calculated the responses and controlled the inputs and outputs to and from the TSEMO algorithm (TSEMO repository: <https://github.com/Eric-Bradford/TS-EMO>). The reactant flow rates were reduced to a minimum (dead-time conditions) during heating/cooling of the reactor to minimise the amount of material used. The algorithm was operated in batch-sequential mode, such that each iteration included four experiments. The responses for each objective were calculated from the HPLC chromatograms at the end of each iteration, and the results used to update the surrogate models and generate the next set of operating conditions.

### 2.4. Sonogashira reaction

Reactor: tubular reactor (3 mL, 5 bar). Reservoir solutions were prepared by dissolving the desired reagents in solvent under stirring at ambient conditions. Reagent 1 pump: 3,5-dibromopyridine (143.55 g, 0.61 mol, 0.61 mol L<sup>-1</sup>), Pd(PPh<sub>3</sub>)<sub>4</sub> (10.40 g, 0.009 mol, 0.009 mol L<sup>-1</sup>), CuI (11.61 g, 0.061 mol, 0.061 mol L<sup>-1</sup>), pyrrolidine (131.50 g, 1.85 mol, 1.85 mol L<sup>-1</sup>) and biphenyl (18.80 g, 0.122 mol, 0.122 mol L<sup>-1</sup>) in Toluene/MeCN (2:1, 1 L); Reagent 2 pump: 1-hexyne (68.58 g, 0.84 mol, 0.84 mol L<sup>-1</sup>) in PhMe (1 L); Dilution pump: MeCN. Reactions were left for 3.5 reactor volumes to reach steady state. Average quantities per experiment: 3,5-dibromopyridine (718.1 mg, 3.03 mmol), Pd(PPh<sub>3</sub>)<sub>4</sub> (51.7 mg, 0.04 mmol), CuI (57.7 mg, 0.30 mmol), pyrrolidine (653.8 mg, 9.19 mmol) and 1-hexyne (381.6 mg, 4.65 mmol).

## 2.5. Claisen-Schmidt condensation reaction

Reactor: CSTR cascade (8 mL, 75 psi). Reservoir solutions were prepared by dissolving the desired reagents in solvent under stirring at ambient conditions. Reagent 1 pump: benzaldehyde (25.5 mL, 0.25 mol, 0.50 mol L<sup>-1</sup>) in toluene (500 mL); Reagent 2 pump: acetone (neat); Reagent 3 pump: sodium hydroxide (4.0 g, 0.10 mol, 0.10 mol L<sup>-1</sup>) in deionised water (1 L). Reservoir solutions were replenished when required. Reactions were left for two reactor volumes to reach steady state. Average quantities per experiment: benzaldehyde (307.8 mg, 2.90 mmol), acetone (3.857 g, 66.40 mmol) and sodium hydroxide (43.2 mg, 1.08 mmol).

## 2.6. Miniature CSTR cascade

Each CSTR had a stainless steel base, equipped with a polyacetal lid. The reaction chamber was cylindrical with a 2 mL volume, containing a PTFE coated cross stirrer bar (10 mm diameter) to provide mechanical mixing. A convex glass lens (viewing window) and PTFE gasket were clamped down between the base and lid using three bolts to form a seal. The CSTRs were connected using Polyfon PTFE tubing (1/8" OD, 1/16" ID) to form a cascade of desired length and volume. An aluminium heating mantle was designed to inset four CSTRs (volume = 8 mL). Channels were made in the mantle to provide inlet/outlet flow streams to each CSTR. The mantle was heated using two nickel heating element inserts. The temperature was monitored using a thermocouple and controlled using a Eurotherm temperature controller, which was coded into the control software of the self-optimising system. The compact design ensured that mixing could be achieved in all four CSTRs using a single conventional stirrer plate. Separate thermocouples were placed in the additional inlet of each CSTR, to directly monitor the internal temperature of each reactor using a Pico logger.

## 3. Results and discussion

### 3.1. Towards the synthesis of lanabecestat

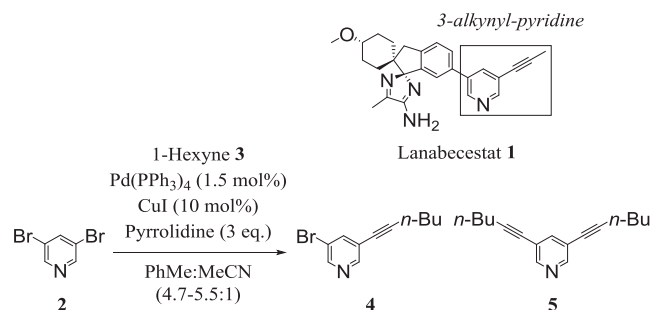
Lanabecestat (AZD3293) is a potent inhibitor of  $\beta$ -site amyloid cleaving enzymes [34], which break down  $\beta$ -amyloid proteins into neurotoxic fragments [35]. Hence, lanabecestat was identified as a potential drug candidate for Alzheimer's disease, reaching Phase III clinical trials. Despite a growing interest in the use of flow chemistry in the pharmaceutical industry, there have been very few reports regarding the self-optimisation of synthetic steps in API production [13].

A key step in the original batch synthesis was the alkylation of 3,5-dibromopyridine with TMS-propyne [36]. The motivations for transferring this step to flow included: (i) TMS-propyne could be safely exchanged for propyne gas, removing the need for additional additives; (ii) downstream lithiation/borylation chemistry is well suited for flow which would enable a telescoped synthesis, providing a significant manufacturing cost saving; (iii) precise control of reaction parameters in flow would give more consistent product quality [37].

To optimise this process, we studied a model Sonogashira reaction between 3,5-dibromopyridine **2** and 1-hexyne **3** (Scheme 1). 1-Hexyne **3** was selected as a model substrate as it is cheaper and easier to handle at room temperature compared to propyne. Due to current difficulties removing **2** during the downstream work-up, the aim of the optimisation was to simultaneously minimise the amount of **2** remaining and maximise the space-time yield (STY) with respect to the mono-alkyne **4** product [Eq. (1)]. The continuous variables which were optimised are: residence time (tR), 1-hexyne **3** equivalents and temperature.

minimise [ $\ln(\%2)$ ,  $-\ln(\text{STY})$ ]

(1)



**Scheme 1.** Model Sonogashira reaction used in the study of the synthesis of the 3-alkynyl-pyridine moiety of lanabecestat **1**. 3,5-dibromopyridine **2** reacts with 1-hexyne **3** to form the desired mono-alkyne **4** and the undesired bis-alkyne **5**. Reservoir concentrations: **2** = 0.61 M, **3** = 0.84 M.

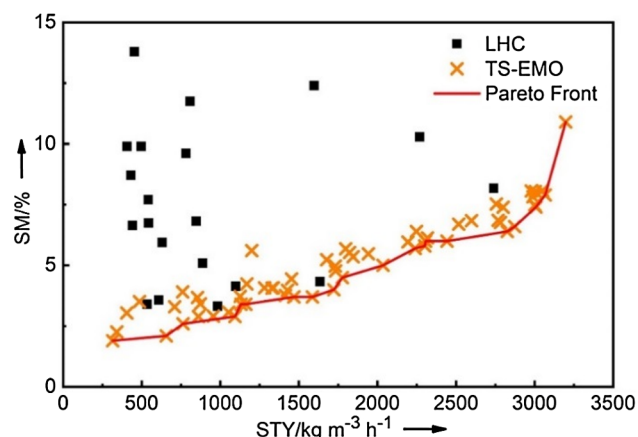
$$tR/\text{min} \in [1, 8]$$

$$\text{subject to: equiv. of } \mathbf{3} \in [1.2, 1.6]$$

$$\text{Temp}/^\circ\text{C} \in [120, 150]$$

The optimisation was initialised with 20 Latin hypercube (LHC) experiments, followed by a subsequent 60 experiments designed by the TSEMO algorithm. The algorithm converged to a Pareto front consisting of 20 solutions (Fig. 1). The maximum STY found was 3198.8 kg m<sup>-3</sup> h<sup>-1</sup> with 10.9% **2** remaining. In contrast, the maximum conversion corresponds to 1.9% **2** remaining and a STY of 315.4 kg m<sup>-3</sup> h<sup>-1</sup>. Between those two edge points, the Pareto front highlights the inherent trade-off between conversion and productivity. Notably, the Pareto front dominates the feasible region. This means that for all points to the left of the Pareto front there exist at least one Pareto point where both objectives are better. *Vice versa*, one objective of the Pareto optimal points cannot be improved without worsening another objective.

The reaction profiles for the % of **2** remaining and STY are shown in Fig. 2a and b respectively. Notably, inspection along the z-axis indicates that varying the temperature between 120 and 150 °C has little effect on either objective. In contrast, reducing the residence time corresponds to a large increase in STY, due to both a reduction in process time and conversion of mono-alkyne **4** to bis-alkyne **5** (Fig. 2c). This correlates with a relatively small increase in **2** remaining. For example, the STY can be increased from 657 to 1586 kg m<sup>-3</sup> h<sup>-1</sup> by reducing the residence time from 4.1 to 1.8 min, whilst only increasing the **2** remaining from 2.1 to 3.7%. The STY can be further increased at the lower limits of residence time by reducing the equivalents of 1-hexyne **3**, which reduces conversion of mono-alkyne **4** to bis-alkyne **5**. However, this correlates with a relatively large increase in % of **2** remaining. For



**Fig. 1.** Results for the multi-objective self-optimisation of the Sonogashira reaction.

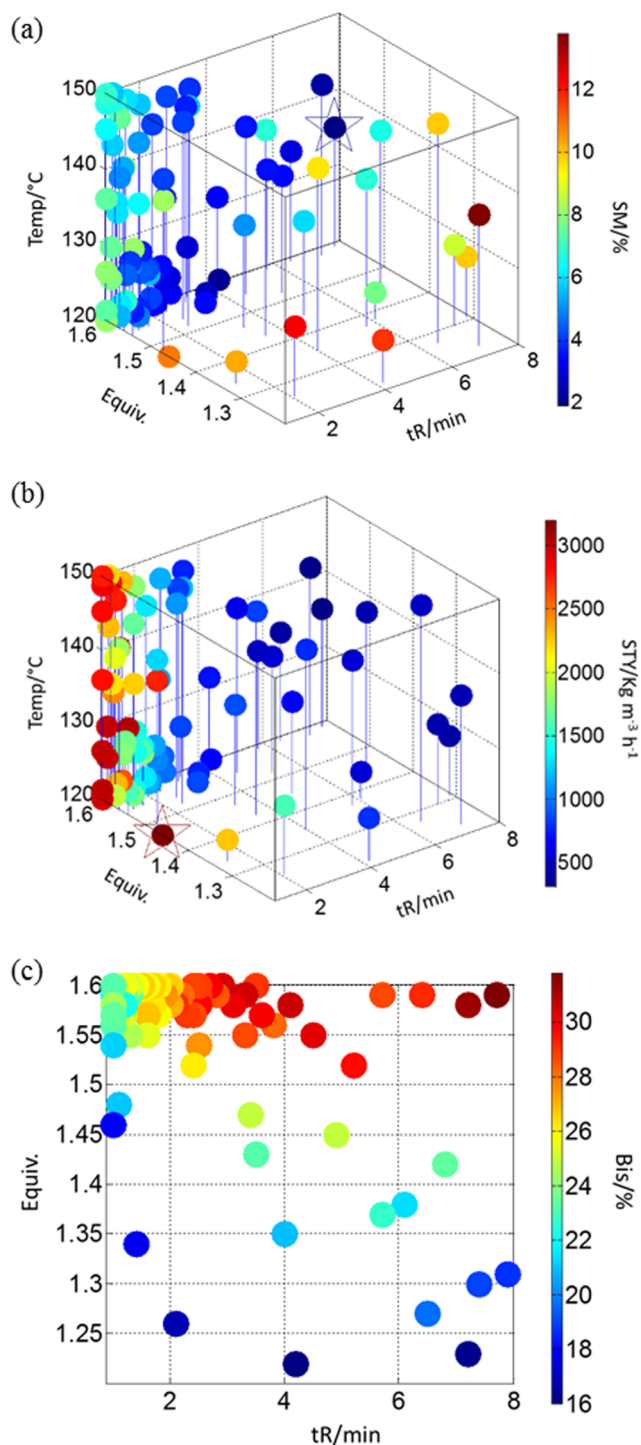


Fig. 2. Plots of experiments performed during the optimisation for different responses: (a) % of 2 remaining, (b) STY with respect to mono-alkyne 4, (c) % of bis-alkyne 5, rotated to depict data as viewed along the z-axis. Equiv. = equivalents of 1-hexyne 3, ☆ = optimum.

example, the STY can be increased from 3009 to 3199  $\text{kg m}^{-3} \text{h}^{-1}$  by reducing the equivalents of 1-hexyne 3 from 1.60 to 1.46 when  $tR = 1$  min, but causes an increase in 2 remaining from 7.4 to 10.9%.

The results from this optimisation enable process chemists to visualise the trade-off between conversion and productivity. In this case, the best STY can readily be selected under the current work-up limitations with respect to 2. Unlike targeted or weighted objective optimisation, the TSEMO algorithm identifies the complete trade-off, such

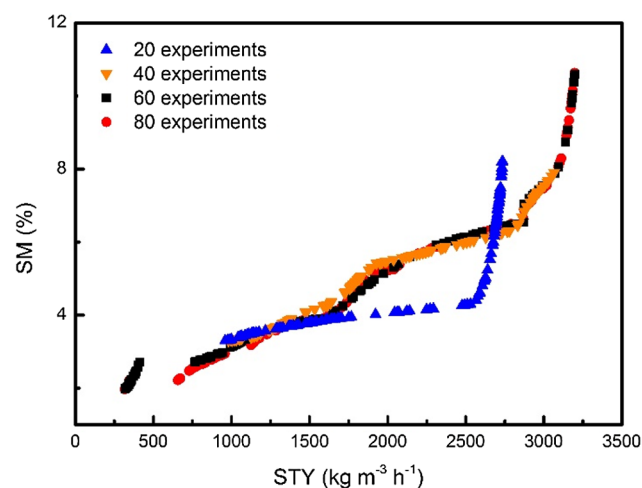


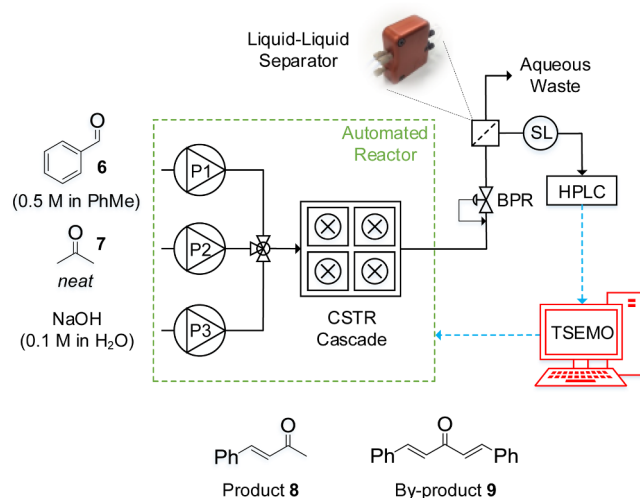
Fig. 3. Pareto fronts found by optimisation using the NSGA-II algorithm of the GP surrogate model predictions after 20, 40, 60 and 80 executed experiments.

that the data can be re-evaluated without further experimentation if process parameters are altered. This is particularly beneficial in the pharmaceutical industry, where the specifications of the downstream work-up are dynamic during process development. Therefore, with most pharmaceutical processes having competing objectives, the use of such algorithms is clearly a beneficial tool for flexible project scenario planning.

One of the challenges associated with expensive-to-evaluate black-box optimisation problems is determining an appropriate termination criterion. Too few experiments can yield suboptimal solutions and inaccurate surrogate models, whereas too many experiments are wasteful in terms of time and materials. In terms of multi-objective optimisation, one could envisage automatically terminating the optimisation once the hypervolume improvement between successive iterations falls below a pre-defined level, where the hypervolume is the volume in the objective space between the current Pareto front and a reference point. However, as there is no guarantee that the hypervolume will improve between each iteration, there is a risk of premature termination of the optimisation (see ESI). Thus, we monitored the progress of the optimisation *via* visual inspection of changes to the predicted Pareto front (Fig. 3). The predicted Pareto front is found by performing a multi-objective optimisation of the GP surrogate model predictions using the NSGA-II algorithm [38]. It is evident that the initial 20 LHC experiments are insufficient for creating GP surrogate models that accurately describe the final Pareto front. The shape of the Pareto front is significantly changed after the initial exploration by the TSEMO algorithm and the subsequent updating of the GP surrogate models with new data. The optimisation was terminated after 80 experiments, as we were satisfied that there were no significant changes to the GP surrogate models between experiments 60–80.

### 3.2. Multi-step self-optimisation

In the previous example, we demonstrated that a reaction in a continually developing process can be optimised whilst considering subsequent work-up operations. An alternative approach would be to optimise the reaction and work-up steps simultaneously. This would reduce the overall development time and enable the impact of downstream processes on economic and environmental objectives to be simultaneously considered. To test this hypothesis, we selected a multi-step process involving the Claisen-Schmidt condensation reaction between benzaldehyde 6 and acetone 7 (Scheme 2), and its subsequent work-up step, targeting benzylideneacetone 8 as the desired product [39]. To the best knowledge of the authors, this is the first report of



**Scheme 2.** Reactor set-up for the Claisen-Schmidt condensation reaction between benzaldehyde **6** and acetone **7** to form the desired benzylideneacetone **8** and the undesired dibenzylideneacetone **9**. P = pump, BPR = back pressure regulator, SL = sample loop.

simultaneous automated optimisation of a multi-step reaction-separation sequence.

The reaction was conducted in a toluene/acetone/water solvent mixture, resulting in a biphasic organic-aqueous reaction medium. Multiphasic reactions require effective mixing to overcome mass transfer limitations. This can be achieved in tubular reactors by using static inserts (passive mixing), however relatively high flow rates and large reactor volumes are required [40], which are unsuitable for expensive-to-evaluate optimisations. In contrast, miniature CSTRs provide active mixing which decouples mixing performance and flow rates [41], thus enabling laboratory-scale optimisation of multiphasic reactions with varying residence times. Given this, a temperature-controlled version of our previously reported laboratory-scale CSTR [42] was selected as an appropriate reactor for this reaction.

The subsequent separation of the organic and aqueous phases was facilitated using an in-line liquid-liquid membrane-based separator [43]. The process was optimised by simultaneously maximising purity, STY and reaction mass efficiency (RME) with respect to benzylideneacetone **8** in the organic phase [Eq. (2)]. The optimisation was conducted in terms of flow rates ( $\nu$ ) and ratios,

where aq = aqueous and org = organic [ $\nu(6) + \nu(7)$ ]. Changing flow rates in this system affects the process in two ways: (i) the reaction in terms of residence time and acetone/sodium hydroxide equivalents; (ii) the solvent ratio and, therefore, the separation in terms of partitioning between the organic and the aqueous phases.

$$\text{minimise}[-\ln(\text{purity}), -\ln(\text{STY}), -\ln(\text{RME})] \quad (2)$$

$$\begin{aligned} \nu(6)/\text{mL min}^{-1} &\in [0.2, 0.4] \\ \nu(7): \nu(6) &\in [0.5, 1.1] \\ \text{subject to: } \nu(\text{aq}): \nu(\text{org}) &\in [0.5, 1.2] \\ \text{Temp}/^{\circ}\text{C} &\in [50, 110] \end{aligned}$$

The optimisation was initialised with 20 LHC experiments, followed by a subsequent 89 experiments designed by the TSEMO algorithm. Of the 109 experiments conducted, 18 Pareto-optimal solutions were identified. A comparison of the responses at the optimum for each function identified conflicts between all three objectives (Table 1). In this case, the purity was  $\approx 10\%$  lower, the STY was  $\approx 2.5 \times$  lower and the RME was  $\approx 1.5 \times$  lower at the optima of the other objectives compared to their own. Inspection of the optimum conditions for each objective showed that they were located at three different corners of the experimental space (see ESI for plots). In general, all of the objectives

**Table 1**

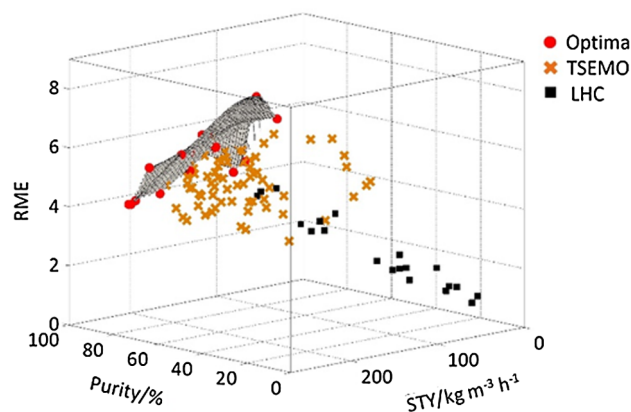
Comparison of the three optimised functions (row headings) and the yield, STY and RME at these optimised conditions (column headings). Figures in bold are the values for the function that have been optimised.

Objective	Purity/%	STY/kg m <sup>-3</sup> h <sup>-1</sup>	RME
Purity	<b>87.9</b>	98	4.74
STY	76.1	<b>260</b>	4.42
RME	78.8	108	<b>7.16</b>

favoured high temperatures, as formation of the dibenzylideneacetone **9** by-product was negligible ( $< 1\%$ ) in the presence of a large excess of acetone **7** [44]. Although the optimum conditions for each objective were identified, this method provided limited process knowledge regarding the influence of the variables on each individual step. This was due to confounding between the reaction and work-up steps, caused by monitoring of the multi-step process using a single downstream analytical source. Therefore, systems optimised using this method are best treated as a black-box, where the inputs and outputs are described but without knowledge of how they are related. In cases where a higher degree of process understanding is required, additional process analytical technologies should be integrated downstream of each individual step [45].

A surface was fitted to the non-dominated solutions to provide a visual representation of the Pareto front (See Fig. 4). This successfully highlighted the complete trade-off between all three objectives, which could be used to aid decision making during process design. For example, the Pareto front indicates that a purity of 82.8%, STY of 166 kg m<sup>-3</sup> h<sup>-1</sup> and RME of 5.66 can be obtained, which represents an approximately equal compromise between all three objectives. Alternatively, a greater importance could be placed on one or more of the objectives, and the process conditions selected accordingly post-optimisation. In this case, the reagents were all cheap and readily available. Hence, the termination criterion was based on practical time limitations, and the optimisation stopped after 65 h. Although the number of experiments was relatively high ( $> 100$ ), a two-step process was successfully optimised with respect to three-objectives without human intervention, and the trade-off between the objectives was identified to aid decision making during process design. Therefore, the efficiency of this approach is favourable when considered against other optimisation strategies.

The majority of existing automated optimisation techniques are focused on single objective optimisation (e.g. Nelder-Mead SIMPLEX, SNOBFIT etc.), where global optimisers require a larger number of experiments per objective compared to this approach. For example, the SNOBFIT optimisation of two steps with respect to three objectives and four variables would require between 252 and 420 experiments based



**Fig. 4.** Results for the multi-objective self-optimisation of the Claisen-Schmidt condensation reaction.

on previous reports [13,39], compared to 109 experiments in this study. Furthermore, single objective optimisation does not provide the trade-off curve between conflicting performance criteria. In contrast, the multi-objective TSEMO algorithm identifies the Pareto front, notably within a more practical number of experiments. An additional drawback of single objective optimisation is that the data is not suitable for reevaluation in cases where specifications are dynamic during process development, such as continuous processes with downstream unit operations. In these examples, identification of the trade-off curve was shown to be essential for the active learning process during multi-step process design. Although this work optimises only continuous variables, TSEMO has also been used for discrete variables, i.e., solvent selection for asymmetric chemistry [32] and, the extension to mixed-integer decisions (e.g. catalysts, bases, solvents and simultaneous reaction conditions) is an ongoing area of research in our laboratories.

#### 4. Conclusions

In conclusion, we have successfully combined multi-objective optimisation based on machine learning methods with self-optimising platforms for the optimisation of multi-step continuous reaction processes. The benefits of this approach compared to other methods include the ability to reevaluate data post-optimisation during dynamic process development, and the identification of a trade-off curve to aid the active learning process during multi-step process design. Application of this method enabled the optimisation of a pharmaceutically relevant Sonogashira reaction in just 13 h, removing the need to conduct further experiments if the downstream work-up process specifications were changed. This corresponded to a significant reduction in the quantity of high-value catalytic reagents required during process development. Furthermore, a multi-step reaction and work-up process were simultaneously optimised with respect to three-objectives in just 65 h, which would conventionally require six separate optimisations carried out over multiple weeks with no guarantee that the trade-off would be obtained. Hence, this method provides a highly efficient and sustainable approach for the development of multi-step continuous flow processes, and aligns with the increasing utilisation of multi-step flow chemistry for the synthesis of active pharmaceutical ingredients.

#### 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.

#### Acknowledgments

ADC, CJT and JAM thank the EPSRC (EP/R513258/1, EP/N509681/1) and University of Leeds for funding. ADC and CJT also thank AstraZeneca for CASE student funding. RAB was supported by the Royal Academy of Engineering under the Research Chairs and Senior Research Fellowships scheme.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cej.2019.123340>.

#### References

- [1] M. Trobe, M.D. Burke, The molecular industrial revolution: automated synthesis of small molecules, *Angew. Chem. Int. Ed.* 57 (2018) 4192–4214.
- [2] H. Winicov, J. Schainbaum, J. Buckley, G. Longino, J. Hill, C.E. Berkoff, Chemical process optimization by computer – A self-directed chemical synthesis system, *Anal. Chim. Acta* 103 (1978) 469–476.
- [3] T. Dimitrov, C. Kreisbeck, J.S. Becker, A. Aspuru-Guzik, S.K. Saikin, Autonomous molecular design: then and now, *ACS Appl. Mater. Interfaces* 11 (2019) 24825–24836.
- [4] V. Sans, L. Cronin, Towards dial-a-molecule by integrating continuous flow, analytics and self-optimisation, *Chem. Soc. Rev.* 45 (2016) 2032–2043.
- [5] D.C. Fabry, E. Sugiono, M. Rueping, Online monitoring and analysis for autonomous continuous flow self-optimizing reactor systems, *React. Chem. Eng.* 1 (2016) 129–133.
- [6] J.C. McWilliams, A.D. Allian, S.M. Opalka, S.A. May, M. Journet, T.M. Braden, The evolving state of continuous processing in pharmaceutical API manufacturing: a survey of pharmaceutical companies and contract manufacturing organizations, *Org. Process Res. Dev.* 22 (2018) 1143–1166.
- [7] R. Porta, M. Benaglia, A. Puglisi, Flow chemistry: recent developments in the synthesis of pharmaceutical products, *Org. Process Res. Dev.* 20 (2016) 2–25.
- [8] M. Baumann, I.R. Baxendale, The synthesis of active pharmaceutical ingredients (APIs) using continuous flow chemistry, *Beilstein J. Org. Chem.* 11 (2015) 1194–1219.
- [9] C. Houben, N. Peremzhney, A. Zubov, J. Kosek, A.A. Lapkin, Closed-loop multi-target optimization for discovery of new emulsion polymerization recipes, *Org. Process Res. Dev.* 19 (2015) 1049–1053.
- [10] 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.
- [11] A. Echtermeyer, Y. Amar, J. Zakrzewski, A. Lapkin, Self-optimisation and model-based design of experiments for developing a C-H activation flow process, *Beilstein J. Org. Chem.* 13 (2017) 150–163.
- [12] 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.
- [13] N. Holmes, G.R. Akien, A.J. Blacker, R.L. Woodward, R.E. Meadows, R.A. Bourne, Self-optimisation of the final stage in the synthesis of EGFR kinase inhibitor AZD9291 using an automated flow reactor, *React. Chem. Eng.* 1 (2016) 366–371.
- [14] 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.
- [15] R.A. Skilton, A.J. Parrott, M.W. George, M. Poliakoff, R.A. Bourne, Real-time feedback control using online attenuated total reflection Fourier Transform (ATR-FT-IR) spectroscopy for continuous Flow optimization and process knowledge, *Appl. Spectrosc.* 67 (2013) 1127–1131.
- [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.
- [17] A.J. Parrott, R.A. Bourne, G.R. Akien, D.J. Irvine, M. Poliakoff, Self-optimizing continuous reactions in supercritical carbon dioxide, *Angew. Chem. Int. Ed.* 50 (2011) 3788–3792.
- [18] 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 CO<sub>2</sub>, *J. Flow. Chem.* 1 (2012) 24–27.
- [19] V. Sans, L. Porwol, V. Dragone, L. Cronin, A self optimizing synthetic organic reactor system using real-time in-line NMR spectroscopy, *Chem. Sci.* 6 (2015) 1258–1264.
- [20] 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. Chem. Int. Ed.* 49 (2010) 7076–7080.
- [21] R.A. Bourne, R.A. Skilton, A.J. Parrott, D.J. Irvine, M. Poliakoff, Adaptive process optimization for continuous methylation of alcohols in supercritical carbon dioxide, *Org. Process Res. Dev.* 15 (2011) 932–938.
- [22] 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.
- [23] 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.L. Grogne, F.-X. Felpin, An Autonomous self-optimizing flow reactor for the synthesis of natural product carpanone, *J. Org. Chem.* 83 (2018) 14286–14299.
- [24] A. Adamo, R.L. Beingsner, M. Behnam, J. Chen, T.F. Jamison, K.F. Jensen, J.-C.M. Monbaliu, A.S. Myerson, E.M. Revalor, D.R. Snead, T. Stelzer, N. Weeranoppanant, S.Y. Wong, P. Zhang, On-demand continuous-flow production of pharmaceuticals in a compact, reconfigurable system, *Science* 352 (2016) 61–67.
- [25] 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.
- [26] E. Bradford, A.M. Schweidtmann, A. Lapkin, Efficient multiobjective optimization employing Gaussian processes, spectral sampling and a genetic algorithm, *J. Glob. Optim.* 71 (2018) 407–438.
- [27] E. Bradford, A. M. Schweidtmann, A. Lapkin, TS-EMO, GitHub, 2018.
- [28] M. Emmerich, Single and Multi-Objective Evolutionary Design Optimization Assisted by Gaussian Random Field Metamodels, University of Dortmund, 2005.
- [29] J. Knowles, ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems, *IEEE Trans. Evol. Comput.* 10 (2006) 50–66.
- [30] K. Deb, K. Sindhya, J. Hakanen, *Decision Sciences: Theory and Practice*, CRC Press, 2016.
- [31] 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.

- [32] 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.
- [33] D. Helmdach, P. Yaseneva, P.K. Heer, A.M. Schweidtmann, A.A. Lapkin, A Multiobjective optimization including results of life cycle assessment in developing biorenewables-based processes, *ChemSusChem* 10 (2017) 3632–3643.
- [34] G. Cebers, R.C. Alexander, S.B. Haeberlein, D. Han, R. Goldwater, L. Ereshefsky, T. Olsson, N. Ye, L. Rosen, M. Russell, J. Maltby, S. Eketjäll, A.R. Kugler, AZD3293: pharmacokinetic and pharmacodynamic effects in healthy subjects and patients with Alzheimer's disease, *Alzheimer's Dis.* 55 (2016) 1039–1053.
- [35] R. Yan, R. Vassar, Targeting the  $\beta$  secretase BACE1 for Alzheimer's disease therapy, *Lancet Neurol.* 13 (2014) 319–329.
- [36] G. Csajnyik, S. Karlstroem, A. Kers, K. Kolmodin, M. Nyloef, L. Oehberg, L. Rakos, L. Sandberg, F. Sehgelmeble, P. Soederman, B.-M. Swahn, B. Von, *Compounds and Their Use as BACE Inhibitors*, 2012.
- [37] D. Znidar, C.A. Hone, P. Inglesby, A. Boyd, C.O. Kappe, Development of a continuous-flow sonogashira cross-coupling protocol using propyne gas under process intensified conditions, *Org. Process Res. Dev.* 21 (2017) 878–884.
- [38] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Trans. Evol. Comput.* 6 (2002) 182–197.
- [39] M.I. Jeraal, N. Holmes, G.R. Akien, R.A. Bourne, Enhanced process development using automated continuous reactors by self-optimisation algorithms and statistical empirical modelling, *Tetrahedron* 74 (2018) 3158–3164.
- [40] E. Paul, V. Atiemo-Obeng, *Industrial Mixing: Science and Practice*, John Wiley & Sons, New York, 2003.
- [41] Y. Mo, H. Lin, K.F. Jensen, High-performance miniature CSTR for biphasic C-C bond forming reactions, *Chem. Eng. J.* 335 (2018) 936–944.
- [42] M.R. Chapman, M.H.T. Kwan, G. King, K.E. Jolley, M. Hussain, S. Hussain, I.E. Salama, C.G. Niño, L.A. Thompson, M.E. Bayana, A.D. Clayton, B.N. Nguyen, N.J. Turner, N. Kapur, A.J. Blacker, Simple and versatile laboratory scale CSTR for multiphasic continuous-flow chemistry and long residence times, *Org. Process Res. Dev.* 21 (2017) 1294–1301.
- [43] A. Adamo, P.L. Heider, N. Weeranoppanant, K.F. Jensen, Membrane-based, liquid-liquid separator with integrated pressure control, *Ind. Eng. Chem. Res.* 52 (2013) 10802–10808.
- [44] A.F.M.M. Rahman, R. Ali, Y. Jahng, A.A. Kadi, A facile solvent free claisen-schmidt reaction: synthesis of  $\alpha$ ,  $\alpha'$ -bis-(Substituted-benzylidene) cycloalkanones and  $\alpha$ ,  $\alpha'$ -bis-(Substituted-alkylidene)cycloalkanones, *Molecules* 17 (2012) 571–583.
- [45] P. Sagmeister, J.D. Williams, C.A. Hone, C.O. Kappe, Laboratory of the future: a modular flow platform with multiple integrated PAT tools for multistep reactions, *React. Chem. Eng.* 4 (2019) 1571–1578.