Multi-objective Bayesian optimisation of a two-step synthesis of p-cymene from crude sulphate turpentine

Jorayev, Perman; Russo, Danilo; Tibbetts, Joshua D.; Schweidtmann, Artur M.; Deutsch, Paul; Bull, Steven D.; Lapkin, Alexei A.

DOI
10.1016/j.ces.2021.116938

Publication date
2022

Document Version
Final published version

Published in
Chemical Engineering Science

Chemical Engineering Science 247 (2022) 116938

Contents lists available at ScienceDirect

Chemical Engineering Science

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

Multi-objective Bayesian optimisation of a two-step synthesis of p-cymene from crude sulphate turpentine

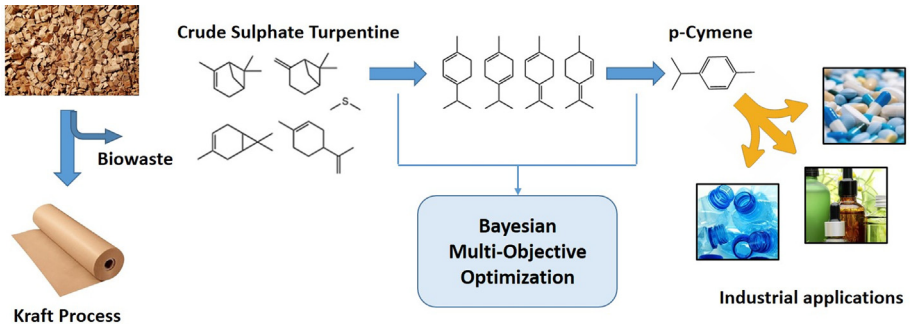
Perman Jorayev [a,b,1], Danilo Russo [a,1,*], Joshua D. Tibbetts [c], Artur M. Schweidtmann [d], Paul Deutsch [e], Steven D. Bull [c], Alexei A. Lapkin [a,b,*]

[a] Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge CB3 0AS, United Kingdom
[b] Cambridge Centre for Advanced Research and Education in Singapore, CARES Ltd. 1 CREATE Way, CREATE Tower #05-05, Singapore 138602, Singapore
[c] Department of Chemistry, University of Bath, Bath BA2 7AY, United Kingdom
[d] Department of Chemical Engineering, Delft University of Technology, Van der Maasweg 9, Delft 2629 HZ, the Netherlands
[e] UCB Pharma S.A. Allée de la Recherche, 60 1070 Brussels, Belgium

HIGHLIGHTS

- Multi-objective surrogate-model-based optimization was applied to a complex reaction.
- Several non-obvious optima were obtained in agreement with physical interpretation.
- Most influential variables were identified.
- Results suggested future investigations to efficiently adopt different feedstocks.
- The method is promising for complex biowaste valorization for circular economy.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 17 February 2021
Received in revised form 12 June 2021
Accepted 6 July 2021
Available online 8 July 2021

Keywords:
Crude sulphate turpentine
Bio-based chemicals
Bayesian optimisation
Reaction development
Circular economy
Biowaste

ABSTRACT

Production of functional molecules from renewable bio-feedstocks and bio-waste has the potential to significantly reduce the greenhouse gas emissions. However, the development of such processes commonly requires invention and scale-up of highly selective and robust chemistry for complex reaction networks in bio-waste mixtures. We demonstrate an approach to optimising a chemical route for multiple objectives starting from a mixture derived from bio-waste. We optimise the recently developed route from a mixture of waste terpenes to *p*-cymene. In the first reaction step it was not feasible to build a detailed kinetic model. A Bayesian multiple objectives optimisation algorithm TS-EMO was used to optimise the first two steps of reaction for maximum conversion and selectivity. The model suggests a set of very different conditions that result in simultaneous high values of the two outputs.

1. Introduction

Development of routes to functional molecules starting from bio feedstocks is one of the strategies to de-carbonise the chemical supply chain: replacement of fossil carbon sources with renewable carbon, preferably from bio-waste sources, allows to significantly

* Corresponding authors at: Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge CB3 0AS, United Kingdom (A. Lapkin).
E-mail addresses: dr473@cam.ac.uk, danilo.russo3@unina.it (D. Russo), aal35@cam.ac.uk (A.A. Lapkin).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.ces.2021.116938
0009-2509/

reduce green-house gases emissions along the chemical supply chain (Guo et al., 2019). We are interested in developing novel routes for valorisation of bio-waste materials which allow access to functional molecules, as drop-in replacements or with novel structures, with quantifiable importance in reaction networks (Jacob et al., 2017; Lapkin et al., 2017; Weber et al., 2019). Most of such routes will start from feedstocks which are complex mixtures: products of de-polymerisation of lignin, of hydrolysis of cellulose, crude glycerol, mixtures of terpenes, etc., and which allow access to a very broad range of functionalised molecules (Guo et al., 2019; Corma et al., 2007).

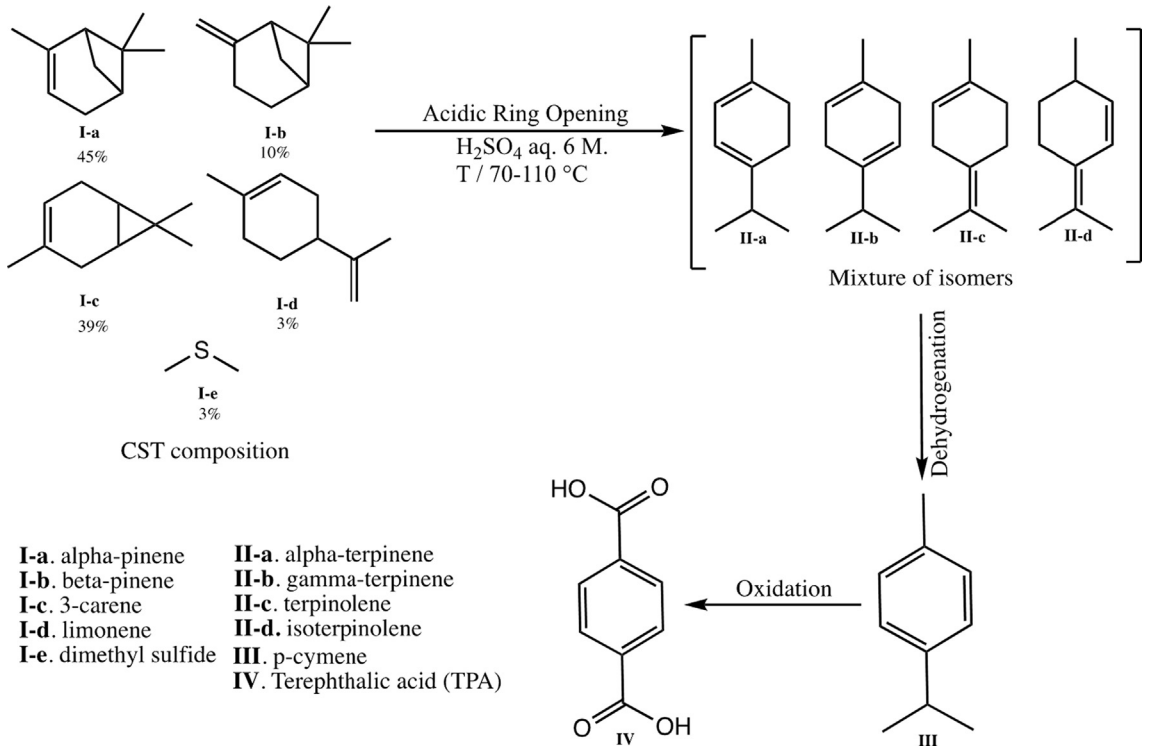
(Bio)chemical conversion of such feedstocks requires either highly selective and robust chemistry, or prior purification. For example, crude sulphate turpentine (CST) is a waste by-product from the pulp and paper industry and can be used to produce *p*-cymene (Scheme 1) (Linnekoski et al., 2014; Zou et al., 2012). *p*-Cymene is used as a solvent for dyes and varnishes and it is the main precursor of p-cresol (Eggersdorfer, 2000). It shows a wide range of antioxidant and biological activity which makes it particularly suited for applications in the food and pharmaceutical industry. Other applications include the synthesis of fragrances and it has been proposed as an alternative precursor of terephthalic acid. The route could start from a purified single terpene, but in our work, the starting point is directly the CST (Tibbetts and Bull, 2021a; Tibbetts and Bull., 2021b; Williams and Whittaker, 1970). In this work a mixture of alpha-pinene, beta-pinene, 3-carene, limonene, and dimethyl sulfide (DMS) was used to mimic CST.

The first reaction in the proposed route is an acid-catalysed ring opening of a mixture of terpenes **I**a-d to a mixture of isomers **II**a-c. This reaction exemplifies a highly promising approach to valorisation of bio-waste feedstocks *via* chemical routes that converge a starting mixture of substrates to a single compound, in this case *p*-cymene (**III**), without significant loss of carbon. *p*-Cymene can then be converted to a range of functional molecules, as, for example, terephthalic acid (**IV**) (Tibbetts et al., 2021).

Within this reaction sequence, optimisation of the acid-catalysed ring opening is particularly challenging for traditional process development tools, since many species interconvert into each other and are too similar for most time-resolved sampling techniques. The mechanism of this reaction has been suggested (Linnekoski et al., 2014; Tibbetts and Bull, 2021a; Williams and Whittaker, 1970; Holmen, 2015). It is a multiphase reaction occurring through the reversible addition of DMS to the double bonds of monoterpenes that generate surfactant-like species improving mass transfer between the acidic and the organic phase. Previous studies have demonstrated the preliminary formation of mixtures rich in limonene and terpinolene, converting to α-terpinene, γ-terpinene, and isoterpinolene through equilibrium protonation reactions (Tibbetts and Bull, 2021a). Acid-catalysed polymerizations decrease the selectivity to the products of interest. As a result, the development of a mechanistic kinetic model for rational scale-up of this process is rather challenging. Model-free optimisation approaches are a promising alternative to model-based methods.

In the recent years, a number of optimisation algorithms has been used for the design of experiments (DoE) in the 'self-optimisation' mode of experiments. In self-optimisation, automated experimental systems perform optimisation of a reaction, and potentially also separation, without intervention of a human (Schweidtmann et al., 2018; Fabry et al., 2016). Several algorithms such as Nelder-Mead-Simplex (McMullen et al., 2010a), SNOBFIT (Jeraal et al., 2018), steepest descent (Moore and Jensen, 2012), MOAL (Echtermeyer et al., 2017) and TS-EMO (Schweidtmann et al., 2018) were used for optimising reactions ranging from nanoparticle synthesis (Krishnadasan et al., 2007) to heterogeneous catalytic reactions (Ley et al., 2015). These methods do not require any prior model of the chemical system and focus on the relationships between input and the output variables or objectives (Reizman and Jensen, 2016).

Out of the algorithms mentioned above, Nelder-Mead-Simplex is a local search algorithm that scales poorly to larger systems and is known to have limited convergence guarantees even for



Scheme 1. A proposed route for conversion of crude sulphate turpentine (CST) to terephthalic acid.

convex problems (Lagarias et al., 1998). Steepest descent is an established algorithm but not suitable for experimental systems because derivatives need to be estimated through finite differences, which is expensive and inaccurate for noisy systems. SNOBFIT, is a global search algorithm that has been used successfully used in many self-optimisation studies but has also shown rather slow convergence (McMullen and Jensen, 2010b; Holmes et al., 2016). Bayesian optimisation approaches, such as MOAL, TS-EMO, and Google Vizier (Golovin et al., 2017) construct non-parametric statistical models, Gaussian processes, using a sequential active learning approach (re-training a model when new experimental observations become available). Bayesian optimisation approaches utilise all available data to build statistical models and are thus data efficient. The models provide quantification of uncertainty that is used to solve the inherent exploration–exploitation trade-off within the derivative-free optimisation. The methods can be used to either develop an accurate model that is valid on the whole input domain or they can be used for efficient optimisation. There exist algorithm extensions to multi-objective optimisation that have been used successfully in self-optimisation and solvent selection (Schweidtmann et al., 2018; Amar et al., 2019).

In this work we extend the previously developed TS-EMO algorithm (Schweidtmann et al., 2018; Bradford et al., 2018; Bradford and Schweidtmann, 2018) to improve its efficiency with respect of experimental budget for optimisation of reaction time, and use it for the optimisation of the ring-opening reaction, step 1, and dehydrogenation, step 2 in Scheme 1.

Most of the existing studies on self-optimisation published to date are limited to optimising up to five reaction variables for a single objective (Mateos et al., 2019). Using the modified version of Bayesian optimisation algorithm TS-EMO, this work demonstrates a model-free approach to building an accurate statistical model for a complex reaction by optimising eight continuous variables for multiple objectives. Successful generation of such accurate models for complex reactions paves the way for establishing model-free hierarchical optimisation of multi-step processes, facilitated by the use of Gaussian processes surrogate model, allowing a certain degree of interpretability.

2. Material and methods

2.1. Materials

Reagents: 3-carene, 90%, stabilized; (1S)-(-)-beta-pinene, 98%; (1S)-(-)-alpha-pinene, 98%; dimethyl sulfide, 99+%, extra pure; alpha-terpinene, 90% tech.; gamma-terpinene, 97%, stabilized; ethyl acetate, 99%; cobalt(II) nitrate hexahydrate, 99%, pure; manganese(II) bromide, 99%, anhydrous were purchased from Acros Organics and used as received. (R)-(+)-Limonene, 97% was purchased from Alfa Aesar and used as received. Reagents 1,2,4,5-tetramethylbenzene, 98%; sulfuric acid, ACS reagent, 95.0–98.0%; terpinolene, ≥90% (GC); *p*-cymene 99%; di-*tert*-butyl peroxide 98%; acetic acid, reagent grade, ≥99%, silicone oil, chloroform-d, were all purchased from Sigma Aldrich and used as received. Deionised water was used to prepare acid solutions.

2.2. Procedures

Experiments for the first step of CST conversion were carried out in batch mode using 100 mL glass bottles immersed in a silicone oil bath. The reaction mixture was magnetically stirred (650 rpm). A glass thermometer was immersed into the reaction medium through a hole in the plastic cap to monitor the reaction temperature. The caps were not sealed and no pressure build up was observed. The temperature was kept constant using an IKA magnetic hot plate equipped with Pt sensor and a feedback controller, for precise temperature control. To start the reaction, first, the calculated relative amounts of all five components of crude-sulphate turpentine (CST) were weighed and added into the reaction bottle. Then, 0.032 mol (0.5 M) of 1,2,4,5-tetramethylbenzene (durene) was added as an internal standard. The reaction was then stirred until the mixture was homogeneous and heated to the given reaction temperature (explored range 70 – 110 °C). The volume of the organic phase – five CST components and the internal standard in the reaction mixture was fixed to 50 mL in every experiment, roughly three times the one previously reported in the literature (Tibbetts and Bull, 2021a). The sulfuric acid solution with the given concentration (4.0 to 7.0 mol L^{-1}, depending on the experiment) and volume (7.5 to 17.5 mL, depending on the experiment) was added to the organic phase; the time of addition was taken as t = 0. Sampling of the reaction mixture for ^{1}H NMR analysis was at different time intervals, as suggested by the algorithm, over the approx. 4.5 h of reaction. Collected samples were decanted in the fridge to quench the reaction and allow for phase separation. The lighter organic phase was then dissolved in chloroform-d in NMR tubes. Details of the reaction and analysis are given in the Electronic Supplementary Information (ESI). Repeated experiments under selected conditions allowed to estimate an average RMSE of 1.41 and 1.83, for conversion and yield, respectively.

The second reaction step was carried out feeding compressed air and the liquid phase containing organic substrate to a continuous-flow stainless steel compact reactor with recycle, packed with glass inert spheres to increase the contact surface between the phases. A full description of the device can be found elsewhere (Bavykin et al., 2005). The organic mixture was prepared using *p*-xylene as a solvent, and *tert*-butyl hydroperoxide was added as a radical initiator to increase the concentration of radicals at the start of the reaction. A 9 mL stirred reservoir was used for the liquid phase, equipped with a condenser at −18 °C to reflux the evaporated compounds. Liquid phase was pumped through the continuous flow reactor using a Vapourtec HPLC pump module and continuously recirculated to the reservoir. Air was directly fed to the reactor using a MFC (SmartTrack 100 Sierra). Samples were collected at different reaction times from the reservoir using a sampling port connected to a syringe. Samples were rapidly diluted in acetonitrile (30 μL in 1 mL) and analysed by GC/MS. Detailed reaction system diagram is given in ESI. This reaction was inspired by the one reported in a previous paper (Tibbetts and Bull, 2021b), but with three fundamental differences: (*i*) no DMS was used in the starting mixture, (*ii*) p-xylene was used as solvent, and (*iii*) the experimental set-up was changed from batch to continuous flow with recycle, which does mimic the overall batch behaviour. Repeated experiments under selected conditions allowed to estimate an average RMSE of 2.27 and 1.27, for conversion and yield, respectively.

2.3. Analytical methods

^{1}H NMR spectra were recorded using a Bruker AVANCE III 400 MHz spectrometer with Bruker QNP Cryoprobe. The experimental condition was 32 scans with the receiver gain set to 4, d1 = 2 s, and the total acquisition time of 2.94 s.

For the isoaromatisation, the product composition was analysed using GC–MS (Agilent Technologies 7890B GC, 5977A MSD, CTC PAL autosampler; HP-Innowax Agilent column 19091 N-133, 30 m × 0.25 mm, 0.25 μm; the system was built and supplied by JSB UK and Ireland Ltd). The inlet condition for the sample injection was 300 °C at the septum purge flow of 3 mL min^{-1} with the split ratio of 100:1 and split flow of 300 mL min^{-1}, column flow 3 mL min^{-1}. The initial oven temperature was held at 60 °C for

1 min and then ramped to 85 °C and then to 180 °C at 2 °C min^{-1} and 100 °C min^{-1}, respectively.

2.4. Algorithm development

The original version of TS-EMO was developed for batch-sequential experimental protocol, suggesting a number of experiments to perform which have the highest predicted hypervolume improvement. However, this may result in repeating the same experiment if the algorithm suggested to collect data at the same reaction conditions, but at a different reaction time. To avoid this, we implemented separate sampling for reaction time. Instead of running experiments at the suggested sample points, objective functions at these points were evaluated and the predicted values were added back to the dataset. The algorithm was run again to sample multiple points (with predicted hypervolume improvement), but this time keeping all the reaction variables constant except the reaction time. In other words, the algorithm was run twice with the second sampling exclusively suggesting points for the reaction time under the same values of the other variables. This improved the data efficiency by allowing multiple sampling per reaction and generating more data using the same amount of the reagents.

3. Results and discussion

3.1. Optimisation of the acid catalysed ring opening

3.1.1. Initial exploration of the decision space

As the automated reaction optimisation requires well-defined bounds of the optimisation variables, we conducted an initial exploration of the variables. For this, we designed and carried out eleven batch reactions. The final ranges of the optimisation variables are shown in Table 1. Upper and lower bounds for the molar fractions of single compounds in CST were chosen based on the variability in turpentine composition depending on its geographical origin.

The starting materials were found to polymerise at elevated temperatures, resulting in lower yields. Based on this, and to avoid boiling off of the aqueous layer, the upper boundary for temperature was set to 110 °C. Conversion reached 100% in under 5.5 h for the medium set of reaction conditions (middle of the ranges of temperature and starting concentrations, rows 1–4 in Table 2). Although sulfuric acid is the catalyst and a higher concentration of acidic hydrogen increases conversion, it also leads to an increase in polymerisation and to a formation of insoluble by-products. Results of the study of the effect of the individual reaction parameters on the reactions outcomes are given in Table 2.

The first set of data (rows 1–4, Table 2) shows that the reaction reaches completion within 5.5 h, depending on the adopted conditions. The second set of data (rows 5–7, Table 2) shows the effect of temperature, i.e. conversion increases with the increase in temperature. For the aq./org. phase ratio, we observed higher conversion and yield with the lower amount of the aqueous phase (rows 8,9). Since the reaction is acid catalysed, higher conversions and yields were observed with higher concentrations of sulfuric acid (6 *vs* 5 M), rows 10,11 in Table 2. In terms of time, longer duration of the reaction gives higher conversion. However, the combination of high temperature and high acid concentration (e.g. 6.92 M H_2SO_4) leads to the decrease in yield over time due to polymerisation reactions (rows 12–14, Table 2). Thus, there exist a trade-off between yield and conversion motivating the use of multi-objective optimisation methods (Tibbetts and Bull, 2021a).

3.1.2. The effect of stirring rate

Several experiments were conducted to ensure that the reaction was not mass transfer limited. Reactions were run at 0, 300, 500, and 700 rpm with identical other reaction conditions. It was observed that conversion and yield are independent of the stirring rate between 500 and 700 rpm, see Fig. 1S in ESI.

3.1.3. The effect of DMS

Another factor that affects the reaction outcome is dimethyl sulfide (DMS) concentration, which is part of the CST in the industrially produced waste. Three experiments were conducted with different quantities of DMS, whilst keeping all other reaction parameters constant. It was found that higher amounts of DMS significantly increase reaction rate in the investigated range 3.2 – 17.0 mol % DMS, whilst there was no significant difference in conversion and yield at the end of the reactions with different DMS quantities, see Fig. 2S in ESI.

3.1.4. Initial dataset collection and algorithm-guided reaction optimisation

Following the initial exploratory experiments to set the optimisation variables bounds, the next set of experiments was performed using a space-filling Latin Hypercube sampling (LHS), which provides an initial dataset to initialize the TS-EMO algorithm (McKay et al., 1979; Tang, 1993). 15 reactions were carried out, collecting an average of 6 samples at different times for each experiment. The dataset that combines data collected in experiments used for exploration of the ranges of input variables, together with those designed by LHS, was then used to train the TS-EMO algorithm. The number of initial data points was chosen based on the order of magnitude reported by similar studies (Schweidtmann et al., 2018), and the availability of experimental resources. The optimal number of training data to select depending on the specifics of a chemical system, or more generally – on the shape of function to be discovered – is a research question without a definitive answer as of today. In some cases, it may be feasible to start optimisation with almost no training data, which would result in the significantly larger exploration requirement and the risk of experimental failures.

In order to speed up data collection and to run experiments in parallel, batch-sequential sampling was implemented in the TS-EMO algorithm. This means that the algorithm designs four new reaction conditions in each run and these can be conducted in parallel (Bradford et al., 2018). Within the algorithm, individual GP surrogate models of conversion and yield are trained on the collected dataset to approximate their response surfaces (Williams and Rasmussen, 2006). The TS-EMO algorithm then draws random samples from these GPs using spectral sampling. Then, a multi-objective genetic algorithm is called within TS-EMO and identifies

Table 1
Lower and upper bounds for the input variables in the optimisation. The numbers given for the starting materials indicate their mole fraction. Limonene molar fraction is kept constant at 0.04, which is the commonly observed concentration of limonene in this bio-waste feedstock. Mole fraction ranges for the other starting materials were chosen according to the composition of industrial produced CST waste (Helmdach et al., 2017).

Range	Temp / °C	H_2SO_4 / mol L^{-1}	Aq./org. ratio	α-pinene	3-carene	β-pinene	DMS	Time / min
Lower	70	4.0	0.15	0.40	0.00	0.05	0.009	1
Upper	110	7.0	0.35	0.80	0.35	0.4	0.037	270

Table 2
The effects of the individual reaction parameters on the reaction outcomes.

T / °C	H_2SO_4 / M	Phase ratio (aq./org.)	α-pinene	3-carene	β-pinene	DMS	Time / min	Conversion / %	Yield / %
90	6	0.2	0.45	0.35	0.12	0.04	**60**	81	66
90	6	0.2	0.45	0.35	0.12	0.04	**120**	93	65
90	6	0.2	0.45	0.35	0.12	0.04	**195**	97	63
90	6	0.2	0.45	0.35	0.12	0.04	**330**	100	56
70	6	0.2	0.45	0.35	0.12	0.04	60	53	41
90	6	0.2	0.45	0.35	0.12	0.04	60	81	66
110	6	0.2	0.45	0.35	0.12	0.04	60	86	65
90	6	**0.2**	0.45	0.35	0.12	0.04	60	81	66
90	6	**0.33**	0.45	0.35	0.12	0.04	60	71	55
90	**6**	0.2	0.45	0.35	0.12	0.04	60	81	66
90	**5**	0.2	0.45	0.35	0.12	0.04	60	25	15
109	**6.92**	0.263	0.45	0.35	0.12	0.04	65	100	**38**
109	**6.92**	0.263	0.45	0.35	0.12	0.04	105	100	**5**
109	**6.92**	0.263	0.45	0.35	0.12	0.04	180	100	**0**

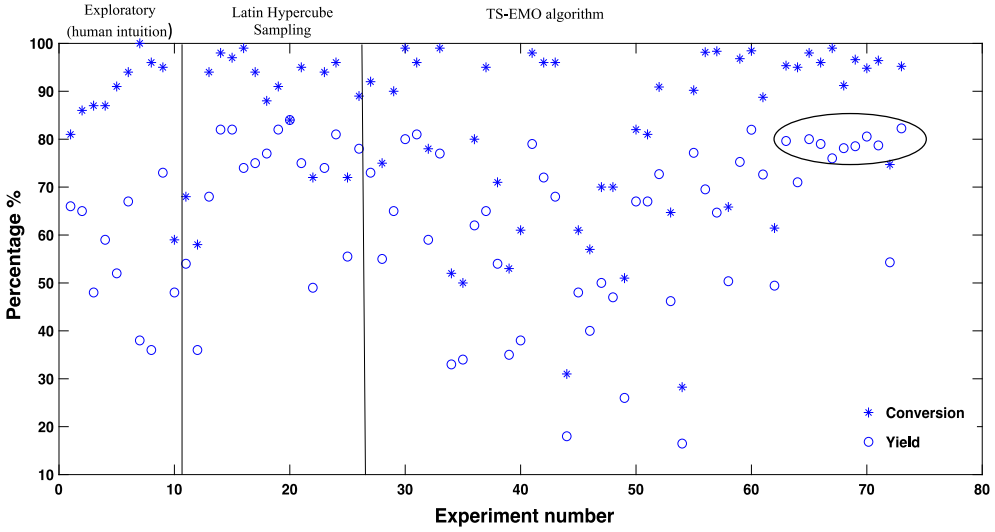


Fig. 1. Results of the optimisation driven by a statistical algorithm in the absence of a physical process model. Dataset is split in a way as human intuition *vs* LHS (initial dataset) *vs* algorithm generated reaction conditions by TS-EMO.

the Pareto front of the random samples (Schweidtmann et al., 2018). Finally, TS-EMO identifies a set of experiments from that Pareto front (of the random GP samples), which aim to improve the hypervolume of the actual Pareto front (of the experiments conducted). TS-EMO algorithm is described in detail in the provided references. Briefly, after the first training of the algorithm, at each iteration, new experimental conditions were suggested, tested in the laboratory, and the results were fed back to the algorithm, until satisfactory outputs were obtained. For better clarity and replicability, the reader can refer to pseudocode reported in the ESI. Selected parameters for the algorithm are: maximum number of function evaluation = 35; number of spectral sampling points for each objective = 1000; Matern type = 1 (Bradford et al., 2018); function evaluations by direct algorithm per input dimension for each objective = 200. At each iteration, the computational time required for new suggestions was of the order of magnitude of hundreds of seconds, and in any case, completely negligible compared to the time required to run the experiments.

As one can see in Fig. 1, the initial training dataset includes conditions with both high and low outputs for both objectives. The algorithm suggested conditions, which start from experiment 26, also include conditions resulting in both high and low objectives. This illustrates the behaviour of TS-EMO algorithm, balancing the trade-off between exploration (developing a good model, specifically targeting experiments to reduce uncertainty of the model) and exploitation (finding conditions that give optimal objective values – high yield and conversion in this case). The exploration of the algorithm can be observed mainly in the beginning (experiments from 26 to 54, Fig. 1), where also regions with comparable low objective values are explored. This is the result of the intrinsic behaviour of the adopted algorithm, aiming to maximize objectives and, at the same time, reduce the uncertainty of GP model predictions by exploring areas of the input space distant from the found local optima. After a certain amount of exploration the algorithm was mainly suggesting conditions that achieved high objective values. It is important to point out that the suggested conditions significantly differ from each other in terms of the process parameters that lead to high objectives, confirming that the algorithm is not stuck at a local. For instance, very similar outcomes were observed for significantly different values of temperature, e.g. 77 *vs* 105 °C (rows 3, 5, Table 3). Also note that the algorithm is never relying on pure exploration or exploitation but is always solving an exploration–exploitation trade-off.

A commonly used rule of thumb to decide the number of experiments to reach the termination criteria in an optimisation process is to set a target for the objectives (e.g. 99% for conversion and 80% for yield) and stop when the target is reached (McMullen et al., 2010a). Another approach is to see if the newly achieved objectives

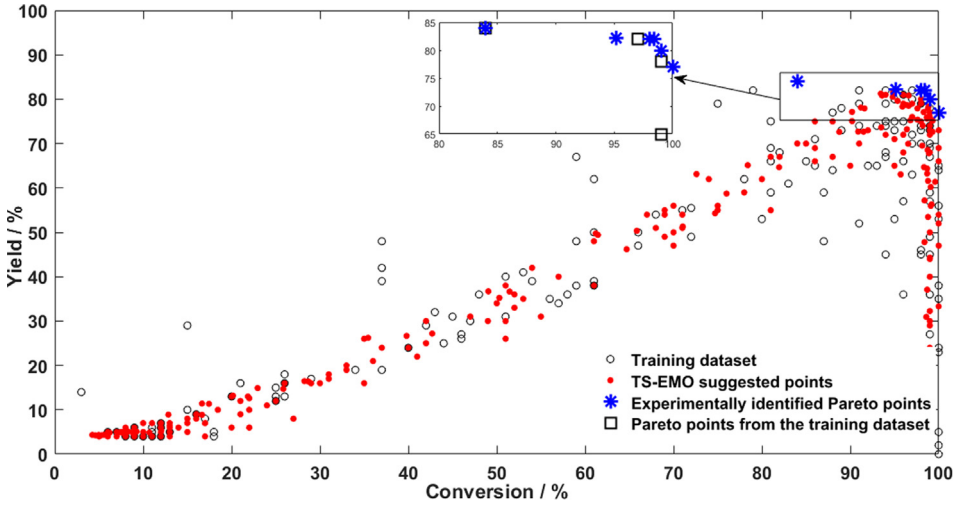


Fig. 2. All the sample points for conversion and yield in acid catalysed ring opening reaction. Empty circles indicate the training dataset, red full circles indicate the result from TS-EMO suggested conditions, and the blue asterisks indicate Pareto points, including top five solutions given in Table 3; empty squares indicate the Pareto point of the original training dataset. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Experimentally validated reaction conditions and composition for the best solutions selected from the final Pareto front.

T / °C	H_2SO_4 / M	Phase ratio (aq./org.)	α-pinene	3-carene	β-pinene	Time / min	Conversion / %	Yield / %
89	6.7	0.21	0.75	0.06	0.13	61	97	82
83	6.1	0.24	0.54	0.11	0.29	134	94	82
77	5.5	0.31	0.53	0.05	0.36	140	96	81
101	5.3	0.22	0.48	0.11	0.35	102	96	81
105	6.9	0.34	0.63	0.006	0.29	30	99	80

differ by less than a pre-defined values, say by 3%, from the previously obtained result (McMullen et al., 2010a). As shown in Table 3, 32 experiments were enough to surpass both of these criteria. Loeppky *et al.* reported a study on choosing the sample size for a computer experiment and used the $10 \times d$ rule, d being the number of variables involved in a reaction (Loeppky et al., 2009). In another study, based on TS-EMO algorithm, 68 and 78 experiments were performed to optimise four continuous variables in an S_NAr and N-benzylation reactions, respectively (Schweidtman et al., 2018). In this work, we see that after 60 experiments the TS-EMO algorithm suggests a dozen of consecutive optimal solutions, Fig. 1. A further stopping criterium is to reach a stable plateau in the hypervolume change as a function of the number of experiments, as highlighted in Fig. 3S in the ESI for both steps of reaction considered in this work.

In the earlier studies we have used TS-EMO to optimise automated reaction and separation systems in continuous flow (Schweidtman et al., 2018; Jeraal et al., 2020) or batch experiments with a fixed batch time (Amar et al., 2019). In this study, the batch time is considered as a continuous degree of freedom. However, the batch time needs to be handled individually during optimisation because it is cheap to withdraw several samples from the same batch at different batch times. Thus, we extend the TS-EMO algorithm accordingly to a two-step optimisation procedure: First, we run the algorithm on the full set of optimisation variables (including batch time). This gives us a suggested experiment, i.e., reaction conditions, and a suggested batch time. Second, we fix the reaction conditions and re-run the algorithm to suggest additional batch times. This gives us a suggested experiment with a set of batch times. The number of suggested batch times can be adapted through a batch sequential approach that has been described in our previous work (Bradford et al., 2018) and is available as an option in the open-source Matlab implementation (Bradford and Schweidtmann, 2018).

The results of the optimisation of acid catalysed ring opening step are given in Fig. 2. The results include a range of high and low values for the objectives, which are useful to efficiently explore the input variable space, and reduce the model uncertainty. Although the objectives increase simultaneously for most of the sample points, one can see that maximum conversion could be achieved at the lowest value for yield. This indicates a non-obvious correlation between the objectives. The blue asterisks are experimentally identified Pareto points. One can see a cluster of optimal solutions explored by the TS-EMO algorithm leading up to the Pareto points. In Fig. 2 we also highlighted the Pareto points of the initial training data set, comprising the expert-guided experiments and the LHS points. As one can see, two of these points can also be found in the Pareto front of the final optimization (97% conversion and 82% yield, and 84% conversion and 84% yield). However, the rest of the points found at the end of the optimization procedure were significantly higher than the ones only based on the training dataset and, most importantly, the outcomes were obtained under a variety of conditions, some of which are particularly relevant from an industrial point of view, i.e. lower temperature and lower acid concentration.

Finally, it is worth stressing that the optimized solutions given in Table 3 are in agreement with general chemistry-based observations previously reported in the literature (Linnekoski et al., 2014; Tibbetts and Bull, 2021a; Williams and Whittaker, 1970). In particular, low 3-carene content is crucial to obtain fast conversion and low occurrence of polymerisation. This supports the previously reported suggestion of blending turpentine feedstocks or distilling carene away from pinenes mixtures to ensure good yields. Also, with more forcing conditions, i.e. higher acid concentration and

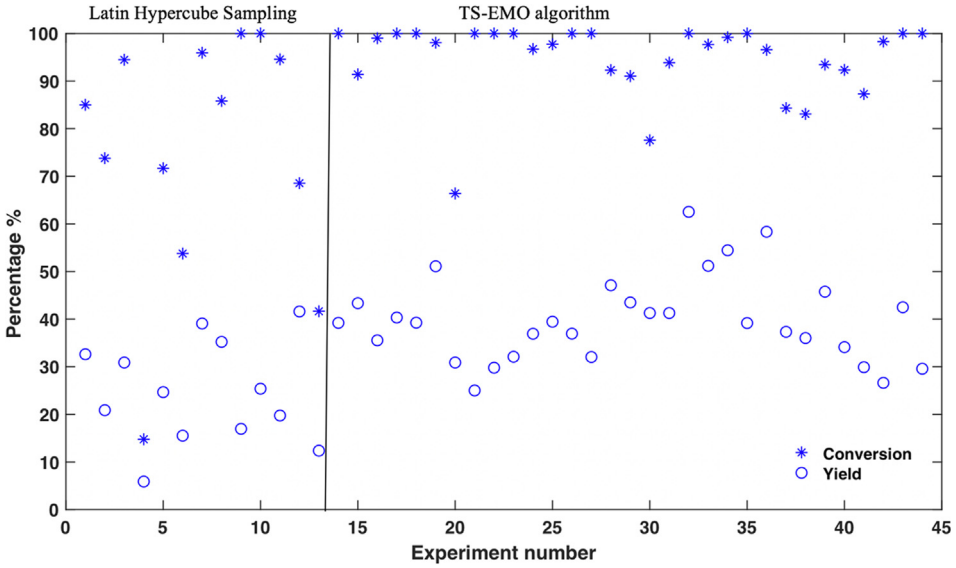


Fig. 3. Results of the optimisation of reaction of *p*-cymene synthesis driven by TS-EMO algorithm. The data is split into the experiments suggested by LHS (initial dataset), and the algorithm-generated reaction conditions.

temperature, the reaction can be stopped sooner and still achieve good yields compared to milder conditions which take longer and would favour polymerisation of the alkene bonds. The best solutions summarized in Table 3 were generally higher than the ones previously reported for similar systems, i.e. ~ 70%. Higher isolated yields were reported (~90%) for systems with no 3-carene and higher contents of DMS, even though the results are not directly comparable because of the different upper limit for DMS concentration of 3.7% used in this work (Tibbetts and Bull, 2021a). It is important to highlight that high outputs values were also obtained under a variety of conditions that were previously unexplored, with reaction times reduced down to 30 min for systems with a volume three times larger than the one previously reported in the literature (Tibbetts and Bull, 2021a). A final remark concerns the attainment of high values of yield and conversion, 99% and 72%, respectively, even in the presence of high contents of 3-carene, i.e. 0.35 M fraction (exp. 64, Table S2 in the ESI). In fact, depending on its origin, CST can contain high amounts of 3-carene, whose separation adds significant costs to the process. Although yield maximization in the presence of high 3-carene content was not the aim of our optimization, the results of explorative experiments, guided by the surrogate model-based approach, suggested new interesting conditions that will be further investigated in future research.

3.2. Optimisation of the radical dehydrogenation reaction

A similar procedure was followed for the second reaction step in order to maximize conversion of terpinene isomers **II**a-c and the selectivity to *p*-cymene. However, four important differences must be highlighted: (*i*) a high accuracy analytical method was used to determine the product concentration, (*ii*) a sequential sampling was adopted, since the experimental setup only allowed to carry out one reaction at a time, (*iii*) constraints of input variables were only based on human intuition and limitations of the experimental setup, and (*iv*) the algorithm was only trained using a set of experiments suggested by LHS.

Eight input variables were selected: temperature, flow rates of the air and the liquid phases, volumetric fractions of the reactants and the radical initiator in the liquid phase, and time. The ranges are summarized in Table 4. The experiment was performed in a flow system with a recirculation, effectively resulting in a batch-like observed temporal response. However, using the compact reactor with internal structure and embedded heat exchangers allows to enhance mass transfer between the phases and operate at a very broad range of temperatures, whilst not compromising on safety. The internal packing of the reactor is an efficient radical scavenger, whereas the micro-heat exchanger has previously been shown to be highly efficient in elevated temperature selective oxidation reactions (Bavykin et al., 2005). We hypothesised that the addition of a radical initiator may enhance the reaction rate, especially at short residence times; this is based on our previous work on aerobic oxidation in the liquid phase (Aworinde et al., 2018).

The results with the highest yields, and conversions above 96% are reported in Table 5, whereas the whole sequence of experiments is reported in Fig. 3.

It is worth highlighting several differences in this optimisation with respect to the previous case. For the second step of the overall route to p-cymene, the optimum was found in 44 experiments and the best results were all obtained under similar conditions, e.g. temperature and reaction time are in a relatively narrow range, 132 – 138 °C and 210 – 240 min, respectively. Any attempt of exploration outside of this narrow area of input variables resulted in a significant decrease in the target outputs. This indicates a much simpler solution space with a single optimum, compared to multiple optimal solutions in the first reaction. The entire dataset in the output space is reported in Fig. 4, highlighting the existence of a single optimal solution, in this specific case. As for the first step of reaction, also in this case we reported the Pareto points of the training data set, demonstrating that the optimization procedure led to a significant improvement of the desired targets.

We also show that in this case the sequential (as opposed to batch-sequential) optimisation strategy results in a more efficient knowledge acquisition from the algorithmic point of view. In this case, each suggested experiment takes into account the results of the previous ones, whereas in a batch-sequential optimisation a certain number of experiments is suggested and some of the suggested experiments may not be as informative as each of the experiments in the sequential optimisation. Batch-sequential optimisation may suit better the problems where both discrete and continuous reaction variables need to be optimised simultaneously, as well as when balancing algorithmic efficiency and

Table 4
Lower and upper bounds for the input variables in the optimisation of the second reaction. The numbers given for the starting materials indicate their volumetric fraction.

Range	Temp / °C	Q_{liq} / mL min^{-1}	Q_{gas} / mL min^{-1}	α-terpinene	γ-terpinene	Terpinolene	Tert-butyl hydroperoxide	Time / min
Lower	80	0.1	5	0.00	0.00	0.00	0.00	0
Upper	150	5.0	120	0.22	0.22	0.22	0.22	240

Table 5
Experimentally validated reaction conditions and composition for best solutions; the first line represents a single optimal solution.

Temp / °C	Q_{liq} / mL min^{-1}	Q_{gas} / mL min^{-1}	α-terpinene	γ-terpinene	Terpinolene	Tert-butyl hydroperoxide	Time / min	Conversion (%)	Yield (%)
138	2.63	98.5	0.059	0.142	0.0068	0.220	234	100	62.5
132	3.48	119.0	0.073	0.175	0.0220	0.146	240	96.6	58.3
136	2.85	105.0	0.096	0.205	0.0296	0.180	230	99.2	54.5
138	4.48	45.6	0.054	0.166	0.0198	0.119	235	97.7	51.2
136	4.70	75.6	0.106	0.165	0.0222	0.098	210	98.0	51.1

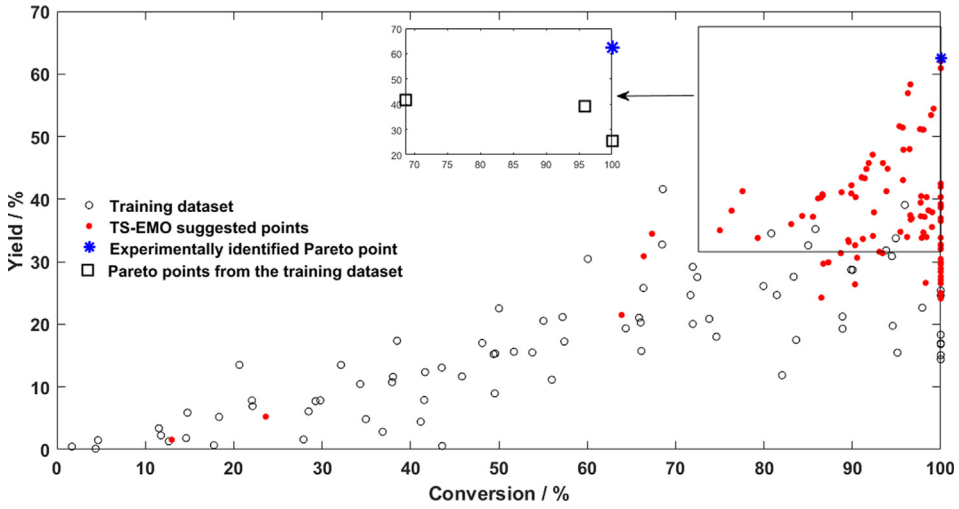


Fig. 4. All the sample points for conversion and yield in p-cymene synthesis. Empty circles indicate the training dataset, red full circles indicate the result from TS-EMO suggested conditions, and the blue asterisk indicate the identified Pareto point; empty squares indicate the Pareto front of the initial training dataset. TS-EMO suggested cluster of optimal solutions leading up to the Pareto point. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

experimental budget (for example, when it is cheap to realise batch experiments).

Some of the best obtained results are in line with the observations reported elsewhere (Bi et al., 2011; Colonna et al., 2011; Iwamuro et al., 1978; Asikainen et al., 2013). In particular, the highest yields reported in the literature are obtained when γ-terpinene is the major component (Asikainen et al., 2013). Thus, it could be possible to further optimise the overall reaction sequence by modifying the feed into the second reaction step (Scheme 1) after the first reaction. In this case direct comparison with previously reported yields is not possible, since the investigated system is substantially different for the following reasons: (i) no DMS was used in the starting mixture, (ii) p-xylene was used as solvent, and (iii) the experimental set-up was changed from batch to continuous flow with recycle

3.3. Sensitivity of developed statistical models to input variables

Within the TS-EMO algorithm, Gaussian process surrogate models are trained on the experimental data. The Gaussian processes contain a number of hyperparameters that can be used to analyse sensitivity of model to input variables, which also implies the relevance or significance of the specific input variables to the objectives of the optimisation. In particular, *lengthscale* hyperparameters describe the influence of input variables on the reaction output, a concept known as *automatic relevance determination* (Williams and Rasmussen, 2006). A lower value of a lengthscale, θ_i, indicates a greater contribution of the corresponding input variable to the objective. In our study of the first step of CST conversion, we can see that temperature, reaction time and acid concentration have strong influence on conversion and yield, Table 6. It is important to point out that the acid concentration has a relatively stronger contribution to conversion than it does to yield. Temperature, on the other hand, has a relatively stronger contribution to yield than it does to conversion. This shows that a higher acid concentration has stronger impact on formation of by-products (e.g. polymers) than temperature. Compared to α-pinene, 3-carene and β-pinene concentrations have smaller contributions to conversion, which matches with the reaction composition where α-pinene exists in a larger quantity and reacts significantly faster than 3-carene leading to a greater contribution to the objectives (Table 1, ESI). Even though β-pinene is quickly consumed, its lower quantity in the starting CST sample, compared to α-pinene, decreases its contribution to conversion and yield.

Besides the lengthscales, hyperparameters σ_f and σ_{noise} correspond to the output variance and noise hyperparameter, respectively. Low output variance, σ_f, indicates that the model is not sensitive to small fluctuations (noise) in the input data and can be used to make accurate predictions on a new dataset. High output variance, on the other hand, implies that the model does not

Table 6
Hyperparameters of GP surrogate models for the acid catalysed ring opening reaction.

Hyperparameter	GP1 (Conversion)	GP2 (Yield)
$\theta_{temperature}$	3.27	2.29
$\theta_{phase\ ratio}$	11.69	10.98
$\theta_{conc.\ H2SO4}$	1.46	2.23
$\theta_{alpha\text{-}pinene}$	3.85	7.00
$\theta_{3\text{-}carene}$	22.74	14.00
$\theta_{beta\text{-}pinene}$	31.62	4.04
θ_{time}	1.48	2.13
σ_f	1.57	2.443
σ_{noise}	$6.14 \cdot 10^{-6}$	$6.14 \cdot 10^{-6}$

Table 7
Hyperparameters of GP surrogate models for the synthesis of *p*-cymene.

Hyperparameter	GP1 (Conversion)	GP2 (Yield)
$\theta_{temperature}$	5.11	31.49
θ_{Qliq}	18.73	11.91
θ_{Qgas}	2.42	7.53
$\theta_{alpha\text{-}terpinene}$	14.72	5.75
$\theta_{gamma\text{-}terpinene}$	23.21	11.77
$\theta_{terpinolene}$	7.05	4.37
θ_{tBuOOH}	31.62	2.21
θ_{time}	2.30	3.95
σ_f	1.08	1.79
σ_{noise}	$1.43 \cdot 10^{-2}$	$6.80 \cdot 10^{-6}$

perform well on the dataset it is not trained on, indicating overfitting of the input data. Overfitting happens when noise in the input data interferes with the signal and causes the algorithm to model noise when it is trained on a noisy dataset. Low values for hyperparameters σ_{noise} and σ_f indicate the quality of the input data and model accuracy to make good predictions on a new dataset, respectively.

Hyperparameters for the model of the second reaction step are listed in Table 7. In this case, the lowest hyperparameter values are associated with reaction time for both objectives. This is in agreement with the experimental data. The decrease in yield cannot be ascribed to consecutive reactions involving the desired product, since yield monotonously increases with reaction time. High gas flow rates appear to be associated with the higher conversion, suggesting an influence of mass transfer limitations on the reaction. Concentration of radical initiator does not affect the rate of the reaction significantly, but it is highly correlated with selectivity to the product of interest.

4. Conclusions

We have demonstrated multi-objective reaction development on previously reported crude sulphate turpentine (CST) conversion to *p*-cymene (Tibbetts and Bull, 2021a; Tibbetts and Bull, 2021b), in the absence of a prior kinetic model using an extended version of Bayesian optimisation algorithm TS-EMO. Eight continuous variables were optimised for acid catalysed ring opening reaction (step 1) in batch, and radical dehydrogenation reaction (step 2) in flow to maximise conversion and yield, without including any prior physico-chemical mechanistic information. Optimisation results showed that the algorithm suggested experimental points that include low and high values for the objectives. This was necessary to reduce the model uncertainty by balancing the exploration–ex ploitation trade-off.

For step 1, the algorithm was able to suggest a group of optimal conditions after 60 experiments. Optimal solutions were achieved under very different range of input variables values, indicating that the algorithm was not stuck at a local optima. However, this was not the case for step 2, where the optimal conditions were achieved in 44 experiments under similar values for temperature and reaction time. The narrow range for these input variables indicates simpler response surface for step 2 with a single optimum. In both reactions, the model suggested cluster of optimal conditions around the experimentally identified Pareto points, which revealed the trade-off between the objectives.

The hyperparameters of the models revealed further information about model sensitivity to input variables. For instance, in step 1, the low values for hyperparameters of temperature, acid concentration, and reaction time indicated strong contribution to conversion and yield than other input variables. Low values for hyperparameters σ_{noise} and σ_f indicate the quality of the input data and model accuracy to make good predictions on a new dataset, respectively.

In summary, Bayesian optimisation algorithms are data efficient tools to develop accurate reaction models where *a priori* knowledge is not available, the number of input variables is large, and the objectives are competing. The developed models for individual steps could be used for potential process design and scale-up.

This is particularly relevant for the development of bio-waste routes to functional molecules, considering the chemical complexity and the high number of variables usually associated with such organic matrices. As a test case, in this paper we show that the application of hybrid approaches for the optimisation of reactions allows to speed up the identification of the best conditions to obtain aromatic components of highly valuable products, like p-cymene, from bio-based side-stream of traditional processes, representing a further step in the de-carbonisation of the chemical supply chain.

CRediT authorship contribution statement

Perman Jorayev: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualization. **Danilo Russo:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualization, Supervision. **Joshua Tibbetts:** Methodology, Formal analysis, Investigation. **Artur M. Schweidtmann:** Methodology, Software. **Paul Deutsch:** Conceptualization, Resources. **Steven D. Bull:** Resources, Supervision. **Alexei A. Lapkin:** Conceptualization, Resources. **:** Supervision, Project administration, Funding acquisition.

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.

Acknowledgements

PJ is grateful to UCB Pharma for PhD scholarship. This project was co-funded by the UKRI project "Combining Chemical Robotics and Statistical Methods to Discover Complex Functional Products" (EP/R009902/1), UKRI project "Terpene-based Manufacturing for Sustainable Chemical Feedstocks" EP/K014889 and the National Research Foundation, Prime Minister's Office, Singapore under its CREATE programme, project "Cambridge Centre for Carbon Reductions in Chemical Technology". AMS is supported by the TU Delft AI Labs Programme.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ces.2021.116938.

References

Amar, Y., Schweidtmann, Artur M., Deutsch, P., Cao, L., Lapkin, A., 2019. Machine learning and molecular descriptors enable rational solvent selection in asymmetric catalysis. Chemical Science 10 (27), 6697–6706. https://doi.org/10.1039/C9SC01844A.

Asikainen, M., Jauhiainen, O., Aaltonen, O., Harlin, A., 2013. Continuous catalyst-free aromatization of γ-terpinene using air as an oxidant. Green Chemistry 15 (11), 3230–3235. https://doi.org/10.1039/C3GC41224E.

Aworinde, S.M., Wang, K., Lapkin, A.A., 2018. Borate-assisted liquid-phase selective oxidation of n-pentane. Applied Catalysis A: General 563, 28–42. https://doi.org/10.1016/j.apcata.2018.06.023.

Bavykin, D.V., Lapkin, A.A., Kolaczkowski, S.T., Plucinski, P.K., 2005. Selective oxidation of alcohols in a continuous multifunctional reactor: ruthenium oxide catalysed oxidation of benzyl alcohol. Applied Catalysis A: General 288, 165–174. https://doi.org/10.1016/j.apcata.2005.04.042.

Bi, L.W., Zhang, Q.G., Wang, P., Zhao, Z.D., Li, D.W., Chen, Y.X., Li, D.M., Gu, Y., Wang, J., Liu, X.Z., 2011. Study on Gas-Phase Catalytic Conversion of Turpentine-Based Dipentene (TBDP) by Pd/C Catalysts. Advanced Materials Research 236–238, 27–34. https://doi.org/10.4028/www.scientific.net/AMR.236-238.27.

Bradford, E., Schweidtmann, A.M., Lapkin, A., 2018a. Efficient multiobjective optimization employing Gaussian processes, spectral sampling and a genetic algorithm. Journal of Global Optimization 71 (2), 407–438. https://doi.org/10.1007/s10898-018-0609-2.

Bradford, E., Schweidtmann, A. M., TS-EMO GitHub, 2018. https://github.com/Eric-Bradford/TS-EMO.

Colonna, M., Berti, C., Fiorini, M., Binassi, E., Mazzacurati, M., Vannini, M., Karanam, S., 2011. Synthesis and radiocarbon evidence of terephthalate polyesters completely prepared from renewable resources. Green Chemistry 13 (9). https://doi.org/10.1039/C1GC15400A.

Corma, A., Iborra, S., Velty, A., 2007. Chemical routes for the transformation of biomass into chemicals. Chemical Reviews 107, 2411–2502. https://doi.org/10.1021/cr050989d.

Echtermeyer, A., Amar, Y., Zakrzewski, J., Lapkin, A., 2017. Self-optimisation and model-based design of experiments for developing a C-H activation flow process. Beilstein Journal of Organic Chemistry 13, 150–163. https://doi.org/10.3762/bjoc.13.18.

Eggersdorfer, M., Terpenes. Ullmann's Encyclopedia of Industrial Chemistry 2000.

Fabry, D.C., Sugiono, E., Rueping, M., 2016. Online monitoring and analysis for autonomous continuous flow self-optimizing reactor systems. Reaction Chemistry & Engineering 1 (2), 129–133. https://doi.org/10.1039/C5RE00038F.

Golovin, D., Solnik, B., Moitra, S., Kochanski, G., Karro, J., Sculley, D., 2017. Google Vizier. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining -, 1487–1495.

Guo, Z., Yan, N., Lapkin, A.A., 2019. Towards circular economy: integration of bio-waste into chemical supply chain. Current Opinion in Chemical Engineering 26, 148–156. https://doi.org/10.1016/j.coche.2019.09.010.

Helmdach, D., Yaseneva, P., Heer, P.K., Schweidtmann, A.M., Lapkin, A.A., 2017. A Multiobjective Optimization Including Results of Life Cycle Assessment in Developing Biorenewables-Based Processes. ChemSusChem 10 (18), 3632–3643. https://doi.org/10.1002/cssc.201700927.

Holmen, AB,. WO/2015/023225/A1, 2015.

Holmes, N., Akien, G.R., Savage, R.J.D., Stanetty, C., Baxendale, I.R., Blacker, A.J., Taylor, B.A., Woodward, R.L., Meadows, R.E., Bourne, R.A., 2016. Online quantitative mass spectrometry for the rapid adaptive optimisation of automated flow reactors. Reaction Chemistry & Engineering 1 (1), 96–100. https://doi.org/10.1039/C5RE00083A.

Iwamuro, H., Ohshio, T., Matsubara, Y., 1978. Novel synthesis of p-methylacetophenone. Nippon Kagaku Kaishi 6, 909–911. https://doi.org/10.1246/nikkashi.1978.909.

Jacob, P.-M., Yamin, P., Perez-Storey, C., Hopgood, M., Lapkin, A.A., 2017. Towards automation of chemical process route selection based on data mining. Green Chemistry 19, 140–152. https://doi.org/10.1039/C6GC02482C.

Jeraal, M.I., Holmes, N., Akien, G.R., Bourne, R.A., 2018. Enhanced process development using automated continuous reactors by self-optimisation algorithms and statistical empirical modelling. Tetrahedron 74 (25), 3158–3164. https://doi.org/10.1016/j.tet.2018.02.061.

Jeraal, M.I., Sung, S., Lapkin, A.A., 2020. A Machine Learning-Enabled Autonomous Flow Chemistry Platform for Process Optimization of Multiple Reaction Metrics. Chemistry-Methods 1 (1), 71–77. https://doi.org/10.1002/cmtd.202000044.

Krishnadasan, S., Brown, R.J.C., deMello, A.J., deMello, J., 2007. C., Intelligent routes to the controlled synthesis of nanoparticles. Lab on a Chip 7 (11). https://doi.org/10.1039/B711412E.

Lagarias, J.C., Reeds, J.A., Wright, M.H., Wright, P.E., 1998. Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions. SIAM Journal on Optimization 9 (1), 112–147. https://doi.org/10.1137/S1052623496303470.

Lapkin, A.A., Heer, P.K., Jacob, P.-M., Hutchby, M., Cunningham, W., Bull, S., Davidson, M.G., 2017. Automation of route identification and optimisation based on datamining and chemical intuition. Faraday Discussions 202, 483–496. https://doi.org/10.1039/C7FD00073A.

Ley, S.V., Fitzpatrick, D.E., Ingham, R.J., Myers, R.M., 2015. Organic Synthesis: March of the Machines. Angewandte Chemie International Edition 54 (11), 3449–3464. https://doi.org/10.1002/anie.201410744.

Linnekoski, J.A., Asikainen, M., Heikkinen, H., Kaila, R.K., Räsänen, J., Laitinen, A., Harlin, A., 2014. Production of p-Cymene from Crude Sulphate Turpentine with Commercial Zeolite Catalyst Using a Continuous Fixed Bed Reactor. Organic Process Research & Development 18 (11), 1468–1475. https://doi.org/10.1021/op500160f.

Loeppky, J.L., Sacks, J., Welch, W.J., 2009. Choosing the Sample Size of a Computer Experiment: A Practical Guide. Technometrics 51 (4), 366–376. https://doi.org/10.1198/TECH.2009.08040.

Mateos, C., Nieves-Remacha, M.J., Rincón, J.A., 2019. Automated platforms for reaction self-optimization in flow. Reaction Chemistry & Engineering 4 (9), 1536–1544. https://doi.org/10.1039/C9RE00116F.

McKay, M.D., Beckman, R.J., Conover, W.J., 1979. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics 21 (2). https://doi.org/10.1080/00401706.1979.10489755.

McMullen, J.P., Stone, M.T., Buchwald, S.L., Jensen, K.F., 2010. An Integrated Microreactor System for Self-Optimization of a Heck Reaction: From Micro- to Mesoscale Flow Systems. Angewandte Chemie International Edition 49 (39), 7076–7080. https://doi.org/10.1002/anie.201002590.

McMullen, J.P., Jensen, K.F., 2010. An Automated Microfluidic System for Online Optimization in Chemical Synthesis. Organic Process Research & Development 14 (5), 1169–1176. https://doi.org/10.1021/op100123e.

Moore, J.S., Jensen, K.F., 2012. Automated Multitrajectory Method for Reaction Optimization in a Microfluidic System using Online IR Analysis. Organic Process Research & Development 16 (8), 1409–1415. https://doi.org/10.1021/op300099x.

Reizman, B.J., Jensen, K.F., 2016. Feedback in Flow for Accelerated Reaction Development. Accounts of Chemical Research 49 (9), 1786–1796. https://doi.org/10.1021/acs.accounts.6b00261.

Schweidtmann, A.M., Clayton, A.D., Holmes, N., Bradford, E., Bourne, R.A., Lapkin, A. A., 2018. Machine learning meets continuous flow chemistry: Automated optimization towards the Pareto front of multiple objectives. Chemical Engineering Journal 352, 277–282. https://doi.org/10.1016/j.cej.2018.07.031.

Tang, B., 1993. Orthogonal Array-Based Latin Hypercubes. Journal of the American Statistical Association 88 (424). https://doi.org/10.2307/2291282.

Tibbetts, J.D., Bull, S.D., 2021a. Dimethyl sulfide facilitates acid catalysed ring oppening of the bicyclic monoterpenes in crude sulfate turpentine to afford p-menthadienes in good yield. Green Chemistry 23 (1), 597–610. https://doi.org/10.1039/d0gc03452e.

Tibbetts, J.D., Bull, S., 2021b,. D., p-Menthadienes as Biorenewable Feedstocks for a Monoterpene-Based Biorefinery. Advanced Sustainable Systems, 2000292. https://doi.org/10.1002/adsu.202000292.

Tibbetts, J.D., Russo, D., Lapkin, A. A., Bull, S. D., 2021, Catalytic Oxidation of bio-p-Cymene into p-Methylacetophenone, p-Toluic Acid and Terephthalic Acid. ACS Sustaibanle Chemistry & Engineering 9, 25, 8642–8652. doi: https://doi.org/10.1021/acssuschemeng.1c02605.

Weber, J.M., Ló, P., Lapkin, A.A., 2019. Identification of strategic molecules for future circular supply chains using large reaction networks. Reaction Chemistry & Engineering. https://doi.org/10.1039/C9RE00213H.

Williams, C.M., Whittaker, D., 1970. Evidence for intimate ion-pair formation in the addition of acids to olefins. Journal of the Chemical Society D: Chemical Communications 15. https://doi.org/10.1039/C29700000960.

Williams, C.K., Rasmussen, C.E., 2006. Gaussian processes for machine learning. MIT press Cambridge, MA 2.

Zou, J.-J., Chang, N., Zhang, X., Wang, L., 2012. Isomerization and Dimerization of Pinene using Al-Incorporated MCM-41 Mesoporous Materials. ChemCatChem 4 (9), 1289–1297. https://doi.org/10.1002/cctc.201200106.