Algorithms for Self-Optimising Chemical Platforms

Jamie A. Manson

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
School of Chemical and Process Engineering

September 2021

The candidate confirms that the work submitted is his/her own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The work in Chapter 1 has appeared in: A. D. Clayton, **J. A. Manson**, C. J. Taylor, T. W. Chamberlain, B. A. Taylor, G. Clemens and R. A. Bourne, React. Chem. Eng., 2019, 4,1545-1554. JAM, ADC and CJT were responsible for preparation of the review. The contribution from other authors was project supervision (TWC, BAT, GC, RAB)

The work in Chapter 2 has appeared in: **J. A. Manson**, T. W. Chamberlain and R. A. Bourne, J. Glob. Optim, 2021, 80, 865-886. JAM was responsible for algorithm design and manuscript preparation. The contribution of other authors was project supervision (TWC, RAB).

The work in Chapter 4 has appeared in: **J. A. Manson**, A. D. Clayton, C. G. Niño, R. Labes, T. W. Chamberlain, A. J. Blacker, N. Kapur and R. A. Bourne, CHIMIA, 2019, 73, 817-822. JAM was responsible for preparation of the manuscript and writing the optimisation algorithm. The contribution from other authors were: ADC was responsible for reactor characterisation and all subsequent experimentation, CGN for construction of the LED light array (CGN), RL for preparation of the manuscript for publication and project supervision (TWC, AJB, NK, RAB).

“A man with new ideas is a madman, until his ideas triumph” – Marcelo Bielsa

Acknowledgements

The development and completion of this thesis could not have happened without the help of several individuals who have helped provide guidance and offer their support along the way. Firstly, I would like to thank my supervisors Richard Bourne and Thomas Chamberlain. The guidance and feedback you have provided throughput the project has been invaluable in achieving the goals set out. The freedom afforded to me, to follow avenues of interest, was greatly appreciated and has allowed me to develop down a pathway which has maximised my potential, for which I am grateful.

I would like to thank everyone I have had the pleasure to collaborate with throughout this body of work. To Adam Clayton, Oliver Kershaw, and Ricardo Labes, thank you for your help throughout with experimental work and providing a sounding board during the project, all this was crucial to the success of the project and helpful in keeping me sane throughout! I would especially like to thank Connor Taylor who made time to make me feel welcome and train me during the initial period of my project. All present and past members of the iPRD lab who have made my time at a pleasure: Calum, Holly, Pia, Tom, Ilias, Claudio, Luke, Brendan, Mary, Rosie, Christaine, Nisha and Micaela.

Thank you to all my family, especially my mum, who’s support throughout has been unwavering, even if you question my choices sometimes.

Finally, to my wonderful partner Alice, without the support you provide continually I would not have been able to do any of this. You have been my rock throughput and allow me to be the best version of myself.

Abstract

The appreciable interest in machine learning has stimulated the development of self-optimising chemical platforms. The power of harnessing computer aided design, coupled with the desire for improved process sustainability and economics, has led to self-optimising systems being applied to the optimisation of reaction screening and chemical synthesis. The algorithms used in these systems have largely been limited to a select few, with little focus paid to the development of optimisation algorithms specifically for chemical systems.

The expanding digitisation of the process development pipeline necessitates the further development of algorithms to tackle the diverse array of chemistries and systems. Improvements and expansion to the available algorithmic portfolio will enable the wider adoption of automated optimisation systems, with novel algorithms required to match the previously unmet domain specific demands and improve upon classical designed experiment procedures which may offer a reduction in optimisation efficiency.

The work in this thesis looks to develop novel approaches, targeting areas currently lacking or underdeveloped in automated chemical system optimisations. This includes development and application of hybrid approaches looking at improving the robustness of optimisation and increasing the users understanding of the optimum region, as well as expanding multi-objective algorithms to the mixed variable domain, enabling the wider application of efficient optimisation and data acquisition methodologies.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

A	Pre-exponential factor
CCC	Central composite circumscribed
CCD	Central composite design
CCF	Central composited face-centred
CCI	Central composite inscribed
CMA-ES	Covariance matrix adaptation evolution strategy
CSTR	Continuous stirred tank reactor
D	Diameter
Da	Damköhler number
DFT	Density functional theory
DIPEA	*N*,*N*-diisopropylethylamine
DMAc	Dimethylacetamide
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DoE	Design of experiments
E_a	Activation energy
EI	Expected improvement
EIM	Expected improvement matrix
Eq.	Equivalents
EtOH	Ethanol
FC	Flow controller
GA	Genetic algorithm
GP	Gaussian process
HPLC	High performance liquid chromatography
HV	Hypervolume
IGD	Inverted generational distance
IGD+	Modified inverted generational distance
iif	If and only if

k	Rate constant
KPI	Key performance indicator
L	Litre
MbDoE	Model-based design of experiments
Me	Methyl
MeCN	Acetonitrile
MFC	Mass flow controller
MINLP	Mixed integer nonlinear programming
MOAL	Multi-objective active learner
MVMOO	Mixed variable multi-objective optimisation
n	Number of moles
NMP	N-methyl-2-pyrrolidone
NSGA	Non-dominated sorting genetic algorithm
ODE	Ordinary differential equation
OVAT	One-variable-at-a-time
P	Pressure
Pd	Palladium
Ph	Phenyl
PI	Probability of improvement
QbD	Quality by design
R	Gas constant
R^2	Coefficient of determination
Re	Reynolds number
RME	Reaction mass efficiency
SL	Sample loop
SMSIM	Super modified simplex
S_NAr	Nucleophilic aromatic substitution
SNOBFIT	Stable noisy optimisation by branch and fit
SS	Sum of squares

STY	Space-time-yield
T	Temperature
THF	Tetrahydrofuran
TOF	Turnover frequency
TON	Turnover number
t_{res}	Residence time
TSEMO	Thompson sampling efficient multi-objective
μ	Dynamic viscosity
V	Volume

Chapter 1 Introduction

1.1 Automated Reaction Platforms

The ever-increasing drive to shorten process development timelines, alongside improving overall process efficiency has led to exponentially rising workloads.[1] To tackle such workloads, it has become a necessity to utilise automated platforms to not only meet demand but to free high value human resources for creative tasks necessary to maintain research output.[2]

1.1.1 Continuous Flow Systems

An enabling factor in the rise of automated chemical system optimisation has been the development and establishment of continuous flow chemistry. Continuous production has been well-established in the oil and gas industry for many decades. This is mainly due to the high production levels required and the fixed nature of the products; meaning process versatility is not a requirement. However, its use for small scale fine chemical applications has only relatively recently seen a rise in published use; in tandem with the advent of small-scale flow reactors.

The expansion in use of small-scale flow systems can be attributed to the advantages they offer in terms of mass and heat transfer for some systems.[3] Flow regimes can be utilised to define the flow patterns within tubular systems and are defined by the Reynolds number:

$$Re = \frac{\rho v D}{\mu} \qquad \textbf{(1-1)}$$

In which ρ is the fluid density, v is the fluid velocity, D is the characteristic diameter of the tube and μ is the dynamic viscosity. Flow regimes for tubular pipes are separated into three regimes: Laminar ($Re < 2000$), transitional ($2000 < Re < 3000$) and turbulent ($Re > 3000$).[4] Due to the small channel size in micro reactors, the flow regime is categorised as laminar. In the laminar regime, the flow is dominated by viscosity effects and due to the lack of shear, the mixing will be mainly driven by diffusion, with the size of the vessel determining the diffusion flux. As the diameter of micro reactors is so small, this leads to the rates of mass transfer often being many orders of magnitude larger than those found in a batch reactor.[5]

The Damköhler number (Da), (1-2), can be used to determine whether the rate of mass transfer by diffusion is sufficient for a given reaction: If Da < 1 the reaction is kinetically limited, for reactions where Da > 1 the reaction rate is faster than mass transfer which can lead to concentration gradients in the system.

$$Da = \frac{kC_0^{n-1}d_t^2}{4D_e} \qquad \textbf{(1-2)}$$

Where k is the rate constant, C_0 is the initial concentration, n is the order of reaction, d_t is the diameter of the tubing and D_e is the diffusion coefficient. This can be a significant issue for reactions with fast consecutive steps and/or competing reactions.[6] Removing or reducing the effects of mass transfer, as small-scale flow reactors can, is extremely beneficial for both reaction optimisation and for kinetic parameter estimation. This ensures the convolution of mass transfer and kinetic rates is reduced, with the kinetic rates being the limiting step, so the user can obtain an accurate representation of the kinetics driving a chemical reaction. Furthermore, due to the high surface area to volume ratios often found in small scale flow systems heat transfer is dramatically improved. This allows for improved operation of highly exothermic reactions in cryogenic conditions, and can also allow for high-temperature operation, that would be infeasible in batch due to safety concerns. In parallel to this, the ability to couple the systems with inline or online analytical techniques enables rapid feedback of process performance metrics making such systems ideal for use in optimisations or kinetic data acquisition.[7–14]

The merging and transformation of small-scale flow chemical platforms with high-throughput technologies presents a key area for future development in the field of flow chemistry and optimisation. Perera *et al*. presented a high-throughput platform utilised in the screening of homogeneous reagents for a Suzuki-Miyaura reaction.[15] The authors exploited the benefits associated with flow chemistry to access conditions unavailable in alternative batch high throughput systems. Employing the system, the authors were able to successfully perform a screening of 5760 possible reactions at a rate of 1500 reactions per 24 hours. This rapid screening application at elevated conditions provides a key niche in which flow chemistry can significantly outperform conventional batch high throughput screening techniques.

1.1.2 Batch Systems

One main drawback to small scale continuous flow systems is their inability to effectively handle solids.[16] Although these issues can often be alleviated through further process design and optimisation during later development stages, during early stage work, where the focus is much more on discovery and data collection than end process optimisation, it is often easier to adopt a batch approach to investigations. Here, automated high-throughput batch systems present a key advantage, providing a time and resource efficient method for screening different combinations of discrete and continuous variables.[17]

Until recently, there has been little in the way of optimisation algorithms applied to batch systems. This has been mainly due to the requirement to perform batches of experiments for each call of the algorithm, and the accessibility of algorithms capable of handling both continuous and discrete variables, which are often investigated utilising automated batch systems. However, recent studies have been performed utilising algorithmically led approaches to guide high throughput based experimentation.[18] Shields *et al.* were able to show the benefits of utilising such approaches for two example reactions, efficiently optimising the reactions within 50 experiments. The use of a batch system allows for the examination of heterogenous systems for which small scale flow systems typically struggle to explore, allowing for the wider adoption of optimisation strategies in process development.

Selection of reactor technology is vital to ensure efficient operation of processes, with both technologies having their advantages and disadvantages. Figure 1-1 details a simplified selection flowchart for determining which reactor technology to select.

1.2 Optimisation

To utilise automated platforms effectively, there is a need for algorithms to act in lieu of the researcher and enable closed loop operation. Such algorithms are a key topic of research, with the need for faster and more efficient optimisations being the main driving force behind their development.

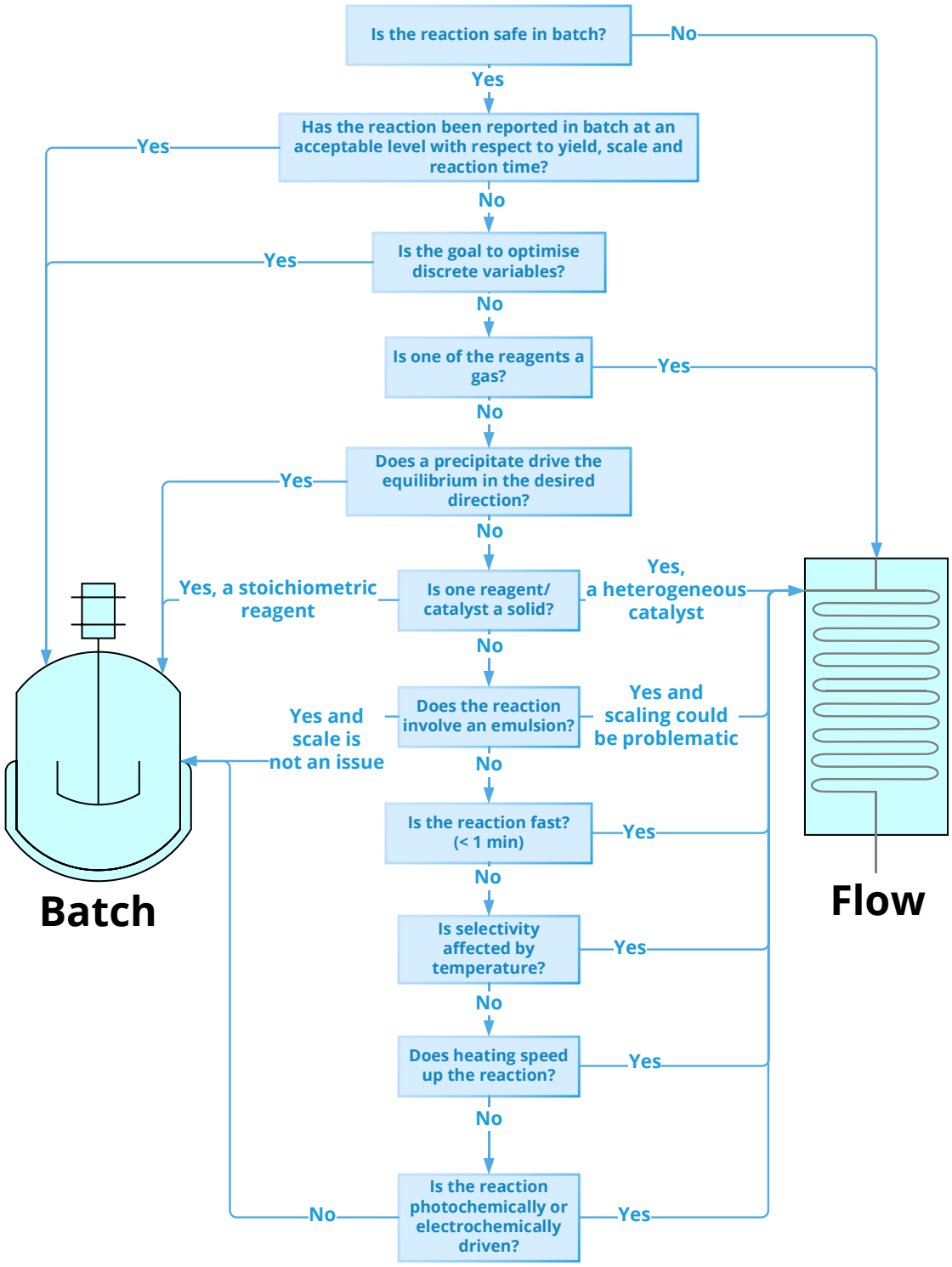


Figure 1-1. A simplified reactor selection decision flowchart. Highlights the high level iterative considerations for when deciding what reactor technology to proceed with.[20]

1.2.1 Automated Optimisation of Chemical Reactions

Automated reaction optimisation or self-optimisation is an approach that combines automated reaction platforms with optimisation algorithms to enable the closed-loop optimisation of chemical systems. The automated reaction platform can take many forms with much of the prior literature focused on the automated optimisation of flow chemical systems.[19] In similar fashion to *in silico* optimisations, experiments are carried out sequentially, in single experiments or batches, until the algorithm achieves a calculated stopping criteria or it reaches a predefined experimental budget.

Table 1-1 provides a summary of published work relating to the self-optimisation of chemical processes between 2007 and June 2021.

Table 1-1. Summary of published work relating to the self-optimisation of chemical systems.

Group, Year	Reaction	Algorithm	Variables Controlled	Objective
deMello, 2007[21]	CdSe nanoparticle synthesis	SNOBFIT	Reactant flowrate, temperature	Wavelength
Jensen, 2008[22]	Benzyl alcohol oxidation	Simplex	Temperature, concentration, residence time	Yield
deMello, 2010[23]	CdSe nanoparticle synthesis	Dynamic simplex	Reactant flowrate, temperature	Wavelength
Jensen, 2010[24]	Heck reaction	Simplex	Equivalence, residence time	Yield
Jensen, 2010[25]	Knoevenagel, Benzyl alcohol oxidation	Simplex, SNOBFIT, Steepest descent method with DoE	Temperature, residence time, inlet concentration, equivalence	Weighted objective function: Yield, throughput Yield
Poliakoff, 2011[26]	Esterification reactions in $scCO_2$	Super modified Simplex (SMSIM)	Temperature, pressure, flowrate	Yield
Poliakoff, 2011[27]	Esterification of 1-pentanol in $scCO_2$	SMSIM	Temperature, pressure, flowrate, equivalence	Yield

Jensen, 2012c	Paal-Knorr reaction	Steepest descent, conjugate gradient, Armijo	Temperature, residence time	$\frac{\text{conversion}}{\text{residence time}}$
Poliakoff, 2012[28]	Esterification of 1-pentanol in $scCO_2$	SMSIM	Temperature, flowrate, pressure, equivalence	Single objective: Yield, Space-time Yield, E factor, E+ factor
Poliakoff, 2013[29]	Esterification of 1-pentanol in $scCO_2$	SMSIM, SNOBFIT	Temperature, flowrate, pressure, equivalence	Yield
Cronin, 2015[30]	Catalytic synthesis of an imine	Modified Simplex	Composition, residence time	Space-time yield
Lapkin, 2015[31]	Emulsion polymerisation	Multi-objective active learner [32]	14 Variables, see reference for details	Multi-objective: Particle size, conversion
Jensen, 2015[33]	Alkylation of 1,2-diaminocyclohexane	Feedback DoE	Discrete: Solvent Continuous: Temperature, concentration, residence time	Yield
Poliakoff, 2015[34]	Catalytic reaction with aniline, DMC and THF in $scCO_2$	SMSIM	Temperature, pressure, flowrate	Yield of various products
Bourne, 2016[35]	Synthesis of EGFR kinase inhibitor AZD9191	SNOBFIT	Flowrate, equivalence, temperature	Yield
Ley, 2016[36]	Appel reaction	Complex Simplex [37]	Concentration, temperature, equivalence, residence time	Weighted-objective function: Throughput, conversion, consumption
Felpin, 2016[38]	Heck-Matsuda reaction	Modified Simplex	Temperature, residence time, equivalence, catalyst loading	Single objective: Yield, throughput, production cost
Jensen, 2016[39]	Suzuki-Miyaura cross-coupling	Feedback DoE	Discrete: Precatalyst scaffold, ligand Continuous: Catalyst loading, temperature, residence time	Constrained single objective: Turnover number (yield >90% of max yield)
Bourne, 2016[40]	Synthesis of N'-nicotinamide	SNOBFIT	Flowrate, equivalence, temperature	Yield
deMello, 2017[41]	Synthesis of o-xylenyl adducts of Buckminsterfullerene	SNOBFIT	Reagent flowrate, temperature	Single constrained objective:

Lapkin, 2017[42]	C-H activation	MbDoE	Residence time, temperature, ratios	Cost, yield
Zare, 2017[43]	Various	Deep reinforcement learning	Concentration, temperature	Relative intensity
Bourne, 2018[44]	Claisen-Schmidt	SNOBFIT, with response surface investigation	Concentration, Equivalents, Residence Time	Yield
Bourne and Lapkin, 2018[45]	S_NAr, N-benzylation	TSEMO	Residence time, equivalence, inlet concentration, temperature	Multi-objective: E-factor, space-time yield
Felpin, 2018[46]	Synthesis of Carpanone	Modified Simplex	Residence time, equivalence, temperature	Single objective: Yield, throughput
Rebrov, 2018[47]	MBY semi hydrogenation reaction	SNOBFIT	Flowrates, pressure, H_2-to-substrate ratio	Weighted-objective function: Selectivity, conversion, and flowrate of reactant
Rueping, 2018[48]	[2 + 2] photocyclization	Modified Simplex	Flowrate, residence time	Conversion
Jamison and Jenson, 2018[49]	Paal-Knorr, Buchwald-Hartwig, HWE Olefination, Suzuki, S_NAr, Photoredox	SNOBFIT	Many, limited to a maximum of 5 variables per optimisation	Yield
Ley, 2018[50]	Synthesis of three active pharmaceutical ingredients	Complex Simplex	Temperature, residence time, equivalence	Weighted-objective function: Throughput, conversion, consumption, energy
Jensen, 2018[51]	Suzuki-Miyaura cross-coupling	Feedback DoE	Discrete: Precatalyst scaffold, ligand Continuous: Catalyst loading, temperature, residence time	Yield constrained catalyst turnover number
Jensen, 2018[52]	Photoredox Iridium-Nickel Dual-Catalysed Decarboxylative Arylation, Cross-coupling	Feedback DoE	Discrete: Base Continuous: Temperature, residence time	Yield constrained productivity
deMello, 2018[53]	Nanocrystal synthesis	Multiparametric Automated Regression Kriging Interpolation and Adaptive Sampling	Reagent concentrations and ratios	Targeted wavelength

Junkers, 2019[54]	Polymer synthesis	Model fitting and function minimisation	Residence time, monomer concentration, control agent concentration	Polymer molecular weight
Felpin, 2019[55]	Lignan synthesis	Modified simplex with screening	Temperature, equivalents, residence time	Yield
Jensen, 2019[56]	C-N cross coupling	Feedback DoE	Discrete: Precatalyst, Base Continuous: Temperature, residence time, equivalents	Yield
Felpin, 2019[57]	Pyridine-oxazoline ligands	Modified simplex	Temperature, equivalents, residence time	Yield
Bourne, 2019[58]	Aerobic oxidation of sp3 C-H bonds	Hybrid SNOBFIT with response surface mapping	Residence time, O_2 equivalents	Yield
Geun Chung, 2020[59]	Synthesis of iron oxide/ gold core-shell nanoparticles	Simplex	Flowrates	Transmission
Cronin, 2020[60]	Gold nanoparticle synthesis	Genetic algorithm	Reagent volume	Shape
Cronin, 2020[61]	Supramolecular compound synthesis	Simplex	Organic inputs, metal salts	Compound discovery
Felpin, 2020[62]	C-H arylation	Modified simplex	Temperature, equivalents, residence time	Yield
Bourne, 2020[63]	Reactive extraction	SNOBFIT	Temperature, residence time, pH, volume ratio	Purity
Bourne, 2020[64]	Sonogashira and Claisen-Schmidt	TSEMO	Temperature, residence time, equivalents	Multi-objective: Space time yield, conversion
Röder, 2020[65]	Imine synthesis	Nelder-Mead simplex	Residence time, temperature, stoichiometric ratio	Yield
Zhu, 2020[66]	Perovskite nanocrystal synthesis	SNOBFIT	Concentration, temperature	Circular dichroism intensity
Malmstadt, 2020[67]	Quantum dot synthesis	Simplex	Reagent driving pressures	Confidence interval of emission maxima
Abolhasani, 2020[68]	Quantum dot synthesis	SNOBFIT, CMA-ES and Neural network with Bayesian optimisation	Precursor flowrates	Target bandgaps

Abolhasani, 2020[69]	Multistep quantum dot synthesis	Neural network with Bayesian optimisation	Precursor flowrate	Weighted-objective function: Peak emission energy and full-width-at-half-maximum
Lapkin, 2020[70]	Aldol condensation	TSEMO	Equivalents, temperature, residence time	Multi-objective: Yield, cost, STY, E-factor
Lapkin, 2020[71]	Mitsunobu reaction	Neural network	Solvent choice	Yield
Berlinguette, 2020[72]	Thin film material discovery	Phoenics [73]	Ratio, time	Hole mobility
Rodrigues, 2020[74]	Various	Random forest	Various	Various (single objective)
Cronin and Lapkin, 2021[75]	Formulation synthesis	TSEMO	Reagent quantity	Multi-objective: Viscosity, turbidity, stability
Hein, 2021[76]	Suzuki-Miyaura coupling	Gryffin (Bayesian optimisation)	Discrete: Ligand Continuous: Ligand ratio, palladium loading, equivalents, temperature	Scalarisation [77]: E-product, Z-product, Pd loading, Equivalents
Doyle, 2021[18]	Mitsunobu, and deoxyfluorination	Bayesian optimisation with DFT descriptors	Discrete: Solvent, Base Continuous: Temperature, concentration, reaction time	Yield
Bourne and Chamberlain, 2021[78]	Gold nanoparticle catalysed nitrophenol reduction	SNOBFIT	Residence time, concentration	Conversion

Details on the optimisation techniques applied in prior published automated reaction optimisation literature, as well as key optimisation techniques are provided below. Methodologies will be detailed in increasing levels of complexity, highlighting the progression of algorithms that have been applied to the optimisation of automated chemical platforms. The methodologies highlighted can be separated into three broad categories: Local optimisation, in which the optimum for the selective objective function is valid for a given region of the input space; Global optimisation, in which the optimum determined is valid across the entire input domain and multi-objective optimisation, in which multiple objective functions are optimised simultaneously.

1.2.2 Local Optimisation

1.2.2.1 One-Variable-at-a-Time

One such strategy is a one-variable-at-a-time (OVAT) approach. In such optimisations each variable is sequentially optimised, while others are held at either predetermined values or at their recently optimised values. An example two variable OVAT optimisation is shown in Figure 1-2. Initially, x_1 is optimised between points A and B with x_2 held at a constant value, A. From this, C is identified as the best value for x_1. Subsequently, x_2 is optimised between D and E, with the apparent optimum of O determined.

The illustrated example highlights some key drawbacks when adopting an OVAT approach to optimisation: (a) optimisation success is largely depended upon starting point; (b) interactions between variables are not considered; (c) there is no exploration of the wider experimental domain. Consequently, as shown below, there is a significant likelihood that the optimum determined by the approach may not be a local optimum. This effect is magnified when the number of variables to be optimised is increased.[79]

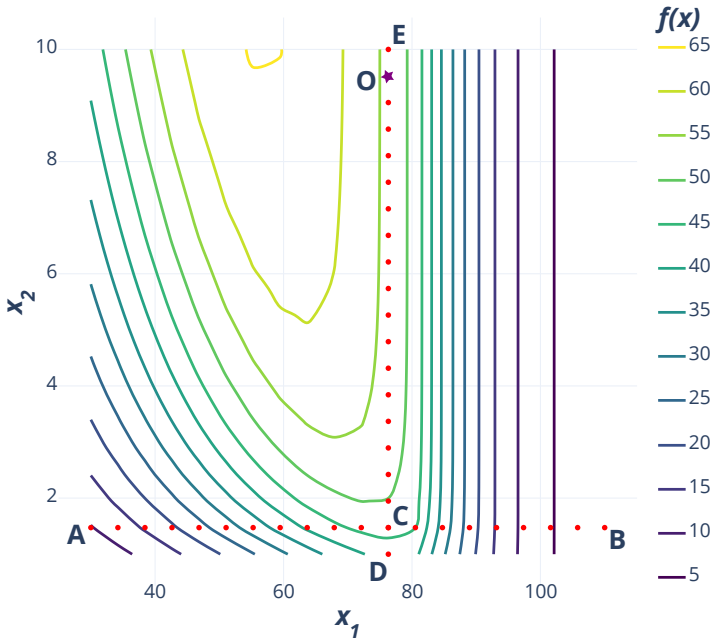


Figure 1-2. A two variable OVAT optimisation of a simulated chemical system. Variables are optimised sequentially, with $\mathbf{x_1}$ initially optimised between points A and B and $\mathbf{x_2}$ held at a constant value, A. From this, C is identified as the best value for $\mathbf{x_1}$. Subsequently, $\mathbf{x_2}$ is optimised between D and E, with the apparent optimum of O.

1.2.2.2 Design of Experiments

Design of experiments (DoE) has been studied and used for chemical process optimisation and screening for many decades.[79] The literature behind the designs is well understood and known throughout chemical industry and academia where it is used readily.[80] The requirement to understand the whole design space and move to a quality by design (QbD) approach have been key factors in the uptake of DoE.[81]

Unlike OVAT approaches, DoE attempts to understand interactions between variables and utilises this to optimise and understand investigated systems.[82] An example of a full factorial design is provided in Figure 1-3.

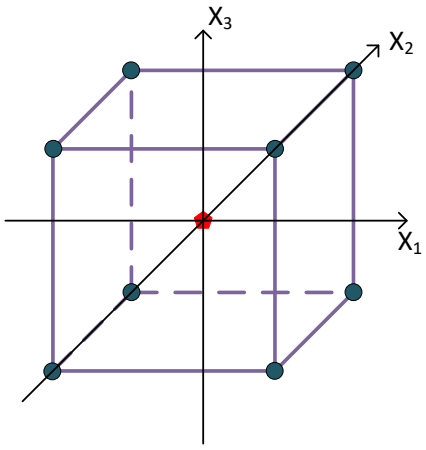


Figure 1-3. Full factorial design for a three-variable system. A central design point has been included. The blue points indicate the experiments to be completed for the factorial design, with the red point highlighting the central design point.

When utilising a structured design, such as a full factorial design, polynomial models can be constructed which empirically describe how a response is affected by input variables, an example of a quadratic model for three variables considering interactions is provided below:

$$y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_{12}X_1X_2 + b_{13}X_1X_3 + b_{23}X_2X_3 + b_{11}X_1^2 + b_{22}X_2^2 + b_{33}X_3^2 + \epsilon \quad \textbf{(1-3)}$$

Where y is the response variable, X_i is an input variable and b_i is the corresponding coefficient in the model. ϵ is the associated error of the model. All postulated models assume that the error is random and normally distributed. Coefficients can be analysed *a posteriori* to determine their significance utilising an analysis of variance approach.

There are many types of DoE designs which can be applied depending upon the requirements of the user. It is beyond the scope of this work to detail all possible DoE designs. Table 1-2 summarises some commonly used designs that can be employed for optimisation tasks.[83] Selecting the correct design for the user's requirements is imperative. Confounding or aliasing can occur when a low-resolution DoE model is selected. This is where some main effects or interactions are biased from higher order interactions.

Table 1-2. Summary of common DoE designs. Figure 1-4 details differences between central composite optimisation designs.

Objective	Examples	No. of experiments for k factors	Advantages/ Disadvantages
Screening	Two level full factorial	2^k	+ Can obtain interaction effects. - Can be inefficient for large design spaces.
	Two level fractional factorial	2^{k-p}, p is the number of generators	+ More efficient than full factorial design. - Depending upon resolution main effect and interactions can be confounded.
	Two level Plackett-Burman	N, where N is a multiple of 4 when k < N-1	+ One of the most experimentally efficient designs. - When interactions are not negligible, they are often confounded with main effects.
Optimisation	CCC	$2^k + 2k + 3$	+ Efficient way for optimising where interactions and quadratic effects are present. - With circumscribed designs the design extends beyond the upper and lower bounds of the design which could be infeasible depending upon the system.
	CCF	$2^k + 2k + 3$	+ Goes fully to the limits set by the user, to check extremes. - Design is not rotatable. As such the variance of the predicted response is not only dependent upon the distance from the design centre point.
	CCI	$2^k + 2k + 3$	+ Rotatable design, which remains within the bounds set by the user. - Does not contain corner points, which can be data rich regions of designs.

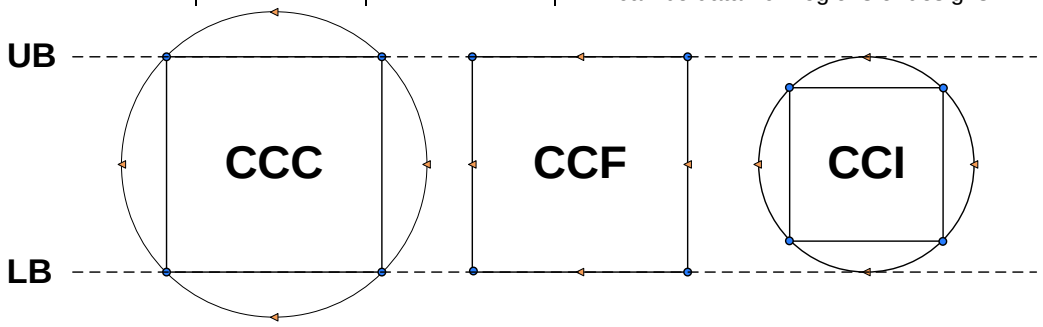
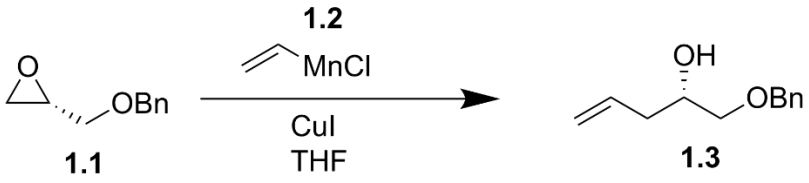


Figure 1-4. Central composite designs for a two-variable optimisation.

A 2^{3-1} fractional factorial design was utilised to examine a copper-catalysed reaction of epoxide with vinyl magnesium, Scheme 1-1.[84]



Scheme 1-1. Copper catalysed epoxide ring opening, studied via DoE methodology.[84]

The authors investigated the robustness of the reaction by varying three parameters, temperature, stoichiometry of **1.2** and addition time of **1.1**. Analysis of the results indicated that temperature was the dominant factor affecting the outcome of the reaction, with the DoE approach providing an efficient methodology of obtaining reaction understanding.

For high dimensional systems, DoE can suffer from the 'curse of dimensionality', in which the number of experiments required to be performed rises exponentially as the number of input variables increases, Figure 1-5. Additionally, as DoE methodologies are fixed methods, there is no dynamic feedback during the execution of experiments. This often means repeat designs must be performed, gradually focusing in on optimal regions. In both cases, the use of dynamic optimisation algorithms can help alleviate these issues.

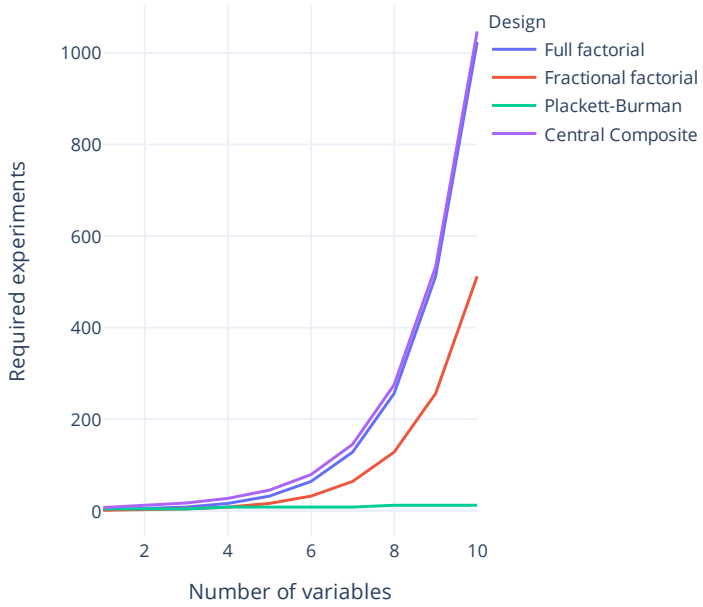


Figure 1-5. Comparison of associated experimental burden for common DoE designs.

1.2.2.3 Sequential Optimal Design

Sequential optimal design builds from a foundation in design of experiments to enable dynamic changes to the optimisation given the

currently available information.[85] Optimal design procedures utilise an underlying model, either an empirical estimate or one with mechanistic foundations, which is then applied in tandem with optimality criteria to guide the experimentation towards a specified optimal goal. Optimality criteria seek to achieve optimal designs for prespecified goals, with D-optimality being one of the most commonly employed. D-optimality seeks to maximise the determinant of the information matrix $X'X$ of the design. This has the effect of minimising the generalised variance of the parameter estimates of the underlying model used. D-optimal designs are often used when resource constraints require that you run fewer experiments or if the design space is constrained.[86] G-optimality represents another category of optimal design criteria. Here, G-optimality seeks to minimise the maximum variance of the predicted values.

As the model is utilised for the calculation of the optimality criteria, each optimal design is model dependent, and the design generated is only optimal for that specific model.

Optimal design methodologies have previously been exploited for the single objective optimisation of mixed variable chemical systems. The Jensen group presented the first applied example in which reaction solvent for an alkylation reaction was optimised alongside continuous reaction variables.[33] A polynomial model based upon an assumed first order kinetic relationship was utilised as the underlying model, with the G-optimality criterion applied to suggest new experiments following an initial data collection. At each iteration, a paired 2-sample t-test was used to determine whether a solvent could be removed from future optimisation, this was done to ensure experiments were not expended on non-optimal solvents.

Although effective at determining process optima in its applied case studies, the requirement for the definition of an underlying model does limit the wider use of the approach. Tailoring the underlying model to the optimised system is a necessity for successful optimisation and therefore could limit the use of the algorithm to more advanced users with specific statistical and domain understanding.

1.2.2.4 Simplex

The simplex algorithm, in its basic form, is a direct search method. This means that it utilises objective function value comparisons to determine the next step of the optimisation. The simplex method was first proposed by Spendley *et al*.[87] and has been subsequently modified over the years to

maximise the efficacy of the algorithm in system optimisations. The method utilises an n+1 geometric polytope to find the local minimum/maximum in an n-dimensional design space. The initially proposed method proceeded towards the optimum through continually reflecting the worse performing point and progressing towards the system optimum. A significant drawback of the original simplex methodology is the algorithms inability to adjust the dimensions of the simplex as the optimisation progresses. As simplex geometry cannot be adjusted, inefficiencies can arise in the algorithm's rate of traversal across the experimental domain, as well as insufficient precision of the final optimisation solution.

Nelder and Mead subsequently suggested modifications to the original design to allow for a more diverse array of transformations illustrated in Figure 1-6.[88]

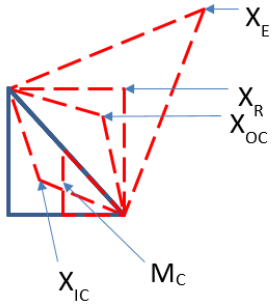


Figure 1-6. Nelder Mead simplex transformations for a 2D problem: reflection (X_R), expansion (X_E), contraction (inside (X_{IC}) and outside (X_{OC})) and multiple contraction (M_C).

The determination of which geometric transformation is applied is carried out through an initial sorting of the current vertex objective function values, followed by a series of logical comparisons. The logic behind the optimisation procedure, and which transformation is performed for each iteration, is detailed in Figure 1-7. An example optimisation of a 2D system provided in Figure 1-8, illustrating the various transformations performed for a simple minimisation problem.

To enhance the efficiency of the Nelder Mead simplex method, various authors have suggested modifications to how and when certain geometric transformations are performed. Routh *et al*. altered the nature of the geometric transformations allowing for variable expansion and contraction of the simplex, utilising a second order polynomial to calculate the expansion/contraction coefficients.[89]

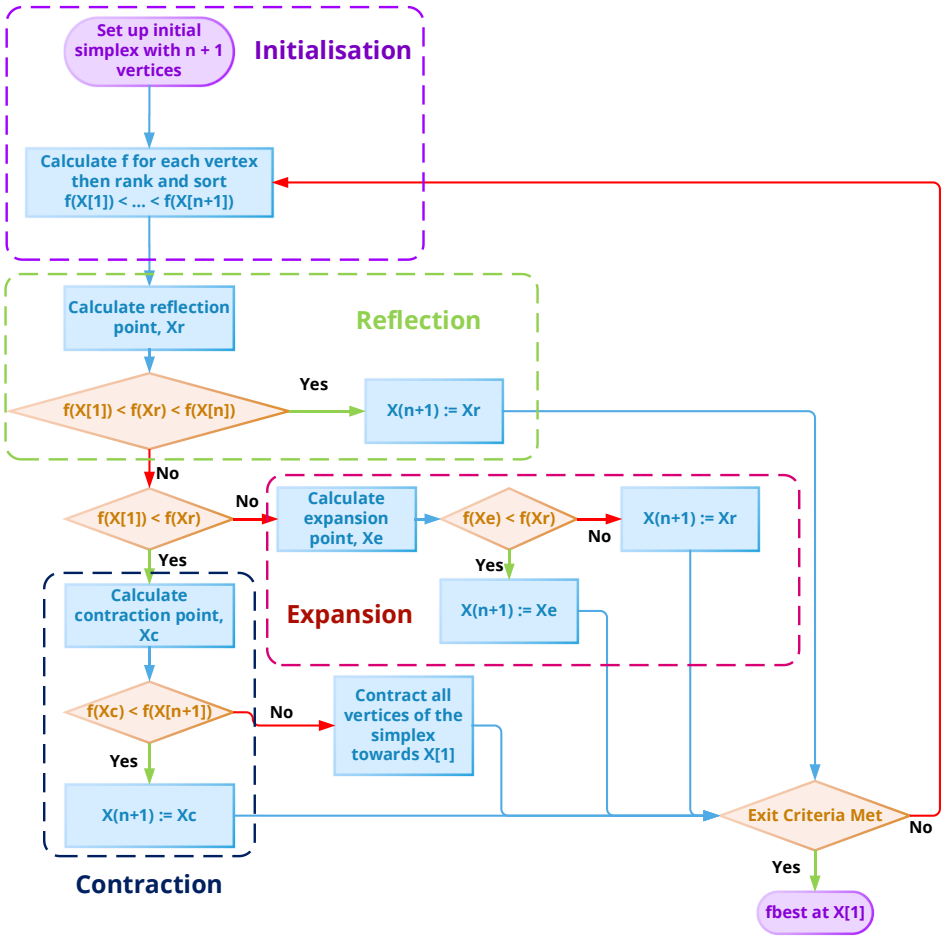


Figure 1-7. Overview of the controlling logic for the Nelder Mead simplex algorithm.

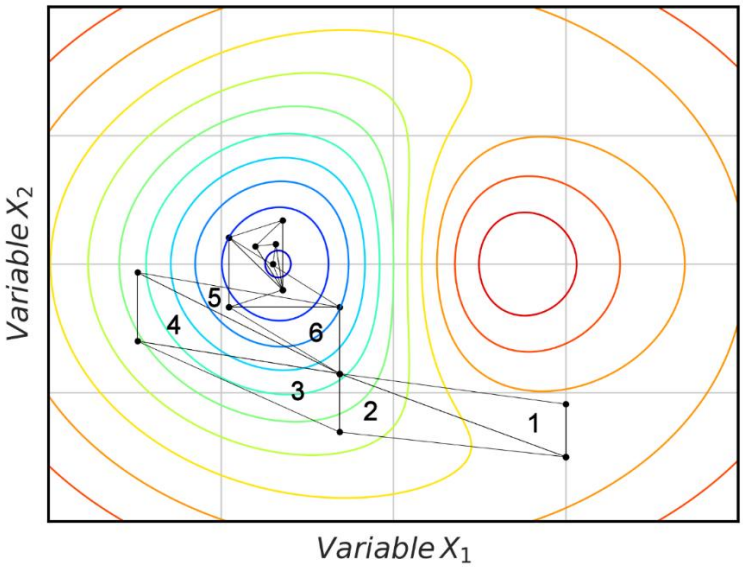
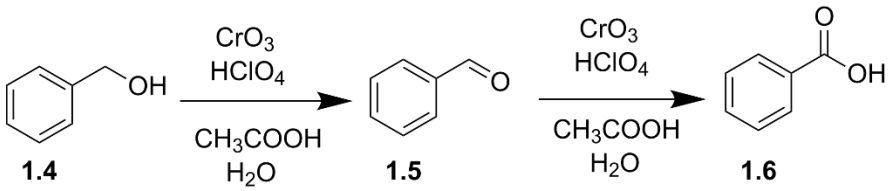


Figure 1-8. Simplified Nelder Mead simplex minimisation optimisation for a 2D system. Contours highlight approximate objective function values with blue indicating an objective function with a lower value. Number triangles indicate the simplex generated for the numbered iteration, up to six iterations. Iterations beyond this point are not numbered for conciseness.

The simplex algorithm and its variations have found extensive use in automated chemical system optimisation literature, with the first reported application on the optimisation of a benzyl alcohol oxidation.[22]

The maximisation of a benzaldehyde (**1.5**) yield was targeted in a four variable optimisation task, Scheme 1-2.



Scheme 1-2. Benzyl alcohol oxidation, utilised as example reaction by McMullen and Jensen.[22]

As one of the first examples of automated algorithmically led experimentation, the optimisation highlighted the efficiency improvements offered versus conventional, designed experimental approaches, with an optimum found in approximately 30 experiments. Notably, as the desired product can over oxidise to an acid, the optimum point lies in the middle of the input domain making the optimisation more challenging than the bound constrained corner point evident in many single objective chemical system optimisations.[34,36,90]

One key consideration before adopting a simplex led approach is the level of noise present in the system. As the algorithm is a direct search method, there is little to no noise handling capabilities and as such the addition of noise can lead to dramatic reductions in the overall optimisation efficiency.[91]

1.2.2.5 Gradient Based

Gradient-based methods form another category of local optimisation algorithms. These methods utilise either the analytical gradient or an estimate of the functions gradient to guide the optimisation towards a local minimum. The simplest implementation of a gradient-based algorithm is the steepest descent method.[92] Considering an example function, $f(x)$, the algorithm computes the next step utilising a given search direction, d_k, where:

$$x_{k+1} = x_k + \alpha_k d_k \qquad \textbf{(1-4)}$$

Step length is given by the following:

$$\alpha_k = \underset{\alpha}{\operatorname{argmin}} f(x_k + \alpha d_k) \qquad \textbf{(1-5)}$$

$$d_k = -\nabla f(x_k) \qquad \textbf{(1-6)}$$

The simplicity of implementation for the method highlights the main advantage, with the main difficulty approximating the gradient if a black box function is being optimised. Due to its simplicity, when applied to real-world problems, it can be quite slow to move towards the optimal region. Given this, various modifications have been suggested to improve the step size calculation. Two key adaptations are the conjugate gradient method[93] and Armijo condition based approaches.[94]

Gradient-based methods have previously been applied to the automated optimisation of chemical reactions, with the Jensen group comparing three gradient based methodologies for the optimisation of a Paal-Knorr reaction.[90] The optimisation was performed with respect to two variables, with the authors finding the Armijo based methodology offering the greatest efficiency in determining the optimum yield for the process. The example highlights the efficiency afforded through application on local methods, with local optimisation methods offering greater optimisation efficiency, for simple convex systems.[95]

As with all local optimisation methods, unless the problem is convex, the solution determined has no guarantee to be the global optimum of the solution. In such instances, where the system behaviour is complex or unknown, use of global optimisation algorithms is recommended.

1.2.3 Global Optimisation

Global optimisation techniques attempt to find the global minima or maxima of a function or system, such that:

$$\begin{aligned} &minimise\ f(\mathbf{x}) \\ &s.t.\quad \mathbf{x} \in D \end{aligned} \qquad \textbf{(1-7)}$$

Where D is a non-empty closed set $D \subset \mathbb{R}^n$, we want to find a point $\mathbf{x}^* \in \mathrm{D}$ that satisfies $f(\mathbf{x}^*) \leq f(\mathbf{x})\ \ \forall \mathbf{x} \in D$.[96]

Unlike local methods, these techniques consider the global nature of the problem to provide confidence that the final solution determined is the global optimum for the system. Often global optimisation methodologies are applied to black-box problems where the underlying behaviour of the system is not fully understood.

1.2.3.1 Genetic Algorithms

Genetic algorithms are population-based optimisation methods that utilise selection, combination, and mutation operations to generate new sample points. They are inspired by nature and rose to prominence as

effective optimisation methods during the early 1970s, remaining to this day a widely utilised technique for optimisation tasks.[97] Although widely employed, given their stochastic nature, genetic algorithms provide no guarantee of convergence upon the global optimum of a solution.

A genetic algorithm is comprised of 4 main stages: evaluation/initialisation, selection, crossover/combination, and mutation. Figure 1-9 provides a graphical overview of this iterative process. Initially, a population of solutions (chromosomes) are randomly generated, each variable of the candidate solution is known as a genotype and the level at which the variable is set is known as a phenotype.

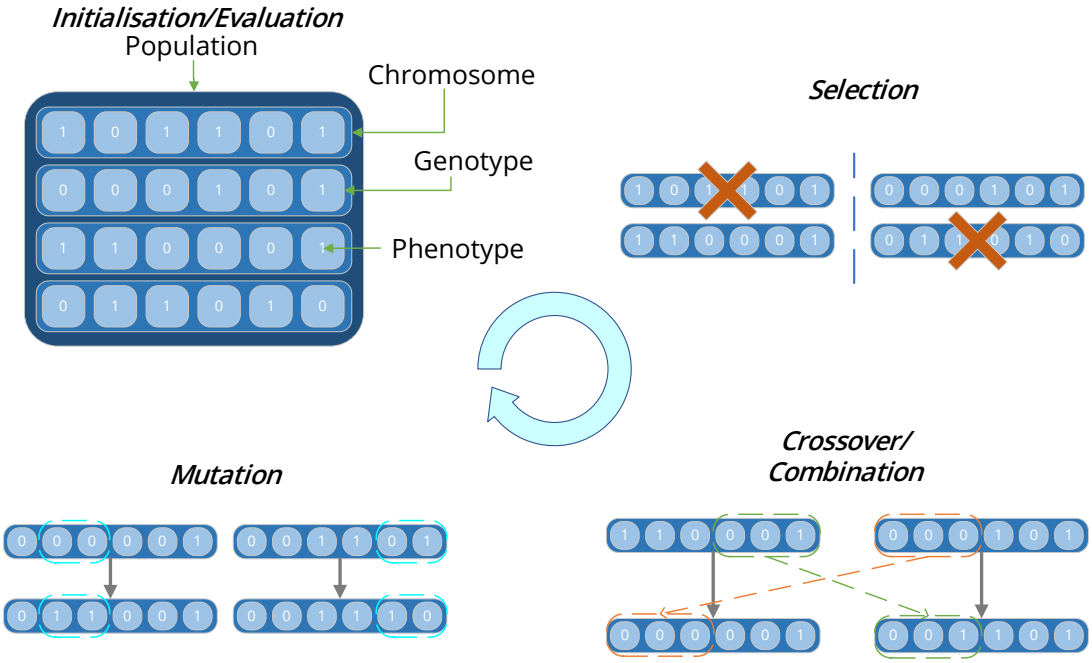


Figure 1-9. Simplified overview of a genetic algorithms iterative process.

Upon generation, the initial population is evaluated with the fitness function values for each chromosome stored. Fitness functions can vary, but often the objective function value is utilised here. Once evaluated, the population undergoes a selection process to determine which chromosomes to take forward for crossover and mutation. Selection can be performed through multiple methods; two commonly applied are tournament and roulette selection:

Tournament: Chromosomes are selected at random for comparison, those with the highest fitness value are taken forward for crossover. Tournament size can be adjusted to increase/decrease selection pressure.

Roulette: The fitness value of a chromosome defines the probability that the chromosome will be selected for crossover. Those with the best fitness value have the largest probability to be taken forward to the next stage.

Crossover is then performed to generate new chromosomes which are the offspring of those taken forward from selection. Crossover can occur at a single point, multiple points or uniformly across the chromosome, in which genotypes are exchanged.

Mutation is utilised to maintain an element of exploration for the genetic algorithm. During this operation, small random changes are made in the chromosome to keep population diversity at an acceptable level. Mutation occurs given a probability, which is usually kept low.

Genetic algorithms have seen a wide range of applications showing excellent capabilities for solving challenging optimisation problems.[98] A key disadvantage however, is the computational overhead often required for running a genetic algorithm optimisation, given the function evaluation budget they require. Due to this, they can be coupled with surrogate models of real world processes to enable optimisations of expensive-to-evaluate objective functions.[99]

As the scale of experiments decreases and capabilities of running parallel experiments becomes more common place, population-based optimisation algorithms such as genetic algorithms may find an increase in their application. The broad exploratory nature of such algorithms lends itself to early stage data acquisition, where examination of the entire experimental domain can yield useful insights. The Cronin group presented the application of a genetic algorithm for the optimisation of nanomaterials shape utilising a custom built liquid handling platform.[60] Shape targets were set utilising the samples UV-Vis spectra, with each generations' experiments performed in parallel to maximise efficiency, Figure 1-10.

Population size was restricted to 15 experiments, with the system allowed to run for 10 generations. This represent quite a restricted budget for a genetic algorithm approach, with optimisations often requiring a significant budget to converge upon an optimum.[98] Adaptations of systems presented by Perera *et al*.,[15] to handle varying the continuous and discrete variable domains, could allow for resource efficient optimisations to still occur, even given the high evaluation demands required when adopting evolutionary techniques.

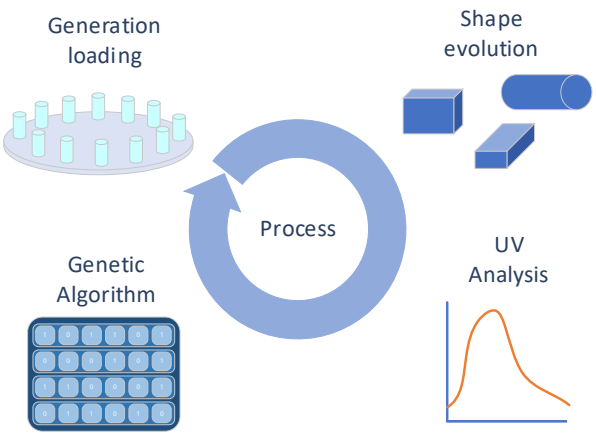


Figure 1-10. Simplified overview of automated shape optimisation. Adapted from Salley *et al.*[60]

1.2.3.2 Stable Noisy Optimisation by Branch and Fit

Stable Noisy Optimisation by Branch and Fit (SNOBFIT) is a global optimisation algorithm for bound constrained noisy optimisation of expensive functions.[100] It is a derivative-free optimisation method, in that it requires no gradient information of the objective function being optimised. The algorithm uses a combination of linear and quadratic surrogate models to determine the optimum point of the system. A basic flow diagram detailing a simplified overview of SNOBFIT is shown in Figure 1-11. Following this logic, the algorithm generates a batch of experiments, with each experiment in the batch falling into five categories:

- Class 1: The point that minimises the local quadratic model around the current best point. It contains at most one point.
- Class 2: Are points that are approximate local minimisers. If there are no local points, then no points in class 2 are generated.
- Class 3: Are points that are approximate non-local minimisers.
- Class 4: Are points in regions that are yet to be explored.
- Class 5: Are points that are randomly generated to fill the design space. They are only generated if the number of evaluated points is less than the number required. The number required is set by the user upon initialisation.

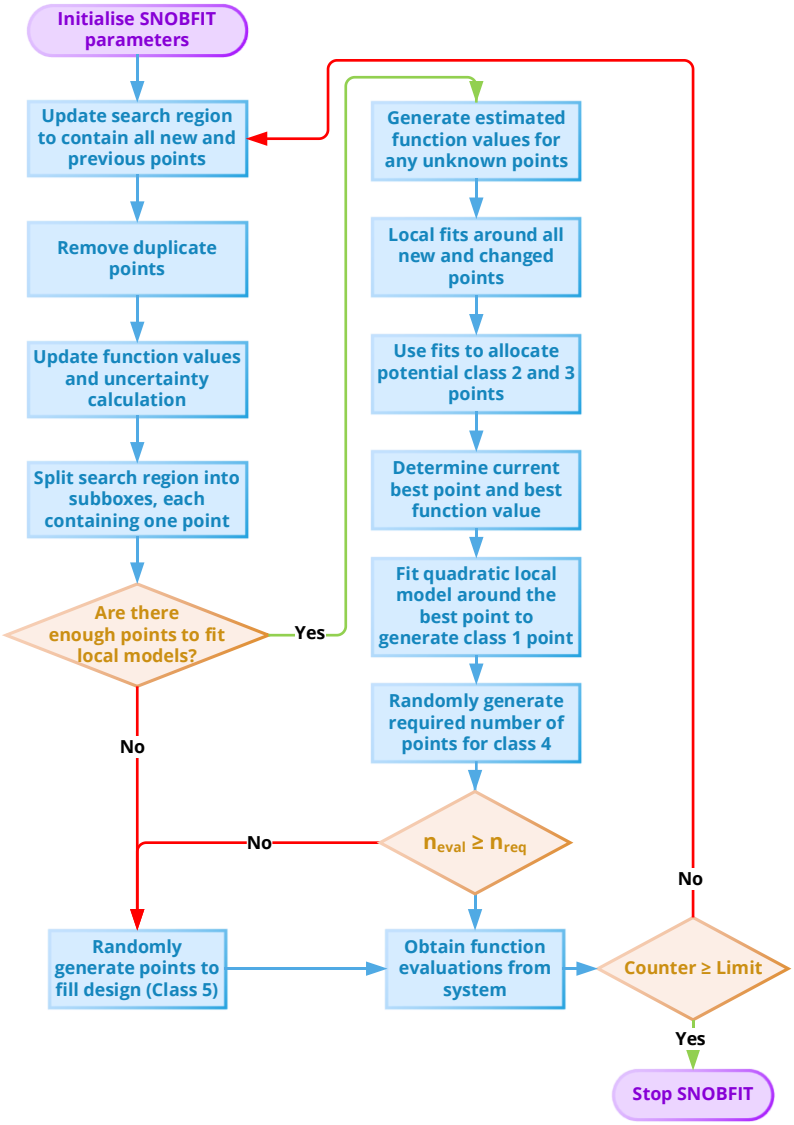


Figure 1-11. Flow diagram for a SNOBFIT optimisation. $\boldsymbol{n_{eval}}$ = number of generated points on this algorithm iteration, $\boldsymbol{n_{req}}$ = number of experiments required per algorithm iteration (set by the user).

The branching aspect of the SNOBFIT algorithm (class 4 points), ensures that the algorithm explores a sufficient portion of the experimental domain. The degree of exploration can be set by the user prior to the optimisation, should the user wish to adopt a greedy approach to the optimisation task. As SNOBFIT includes this exploratory aspect, for simple systems it can take an increased number of function evaluations to determine the system optima.[29] This increased budget, however, is offset by the improved confidence that the determined optima will be the global optimum of that system, although there is no guarantee in this.

Issues can arise when using SNOBFIT for higher dimensional problems, with the algorithms performance and runtime deteriorating when the input domain exceeds nine variables.[101] Nevertheless, when considering

freely available global optimisation algorithms, its performance is among some of the best and is sufficient for most problems.[101] As chemical system optimisations often require the optimisation of a maximum of 4-6 variables, this restriction has not limited its wide use in automated chemical system optimisations.

The first documented use of a self-optimising chemical platform utilised SNOBFIT as the optimisation algorithm.[21] The authors optimised to maximise emission intensity for a target wavelength at the outlet of the reactor. The wavelength selected corresponded to the desired CdSe nanoparticle properties. The authors performed both a 2D and 3D optimisation, initially considering the effects of changing only reagent flowrates, with subsequent expansion to consider reaction temperature upon success of the initial optimisation. In both instances, no *a priori* data was used, with each run started independently and allowed to run for 40 and 100 iterations, respectively.

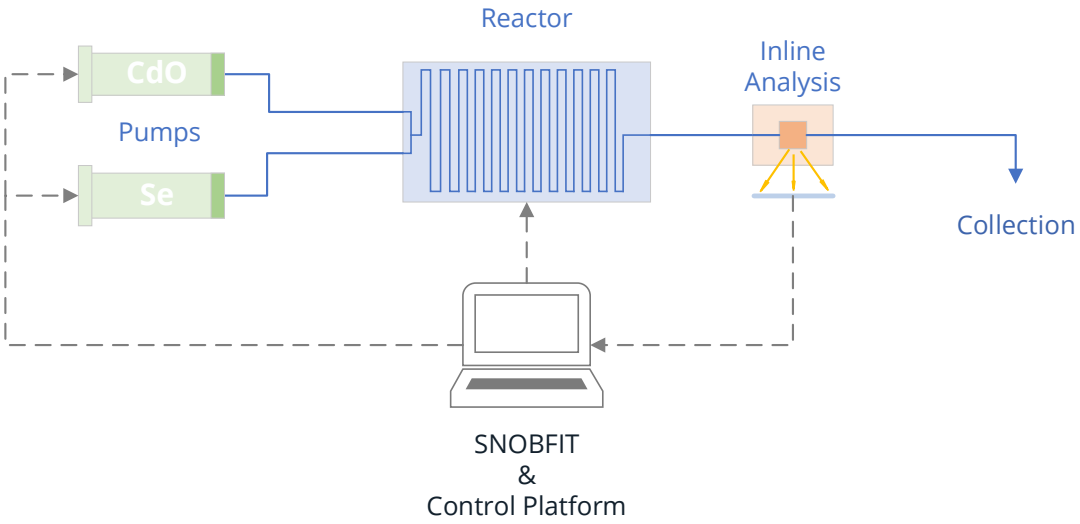


Figure 1-12. Schematic of the automated platform employed by Krishnadasan *et al.* Analysis was provided by laser excitation and a CCD spectrometer which was used to measure an emission spectrum. Adapted from Krisnadasan *et al.*[21]

The algorithm was selected due to its global nature, being able to optimise complex systems such as the ones presented in the paper. Interestingly, the authors elected to utilise substaintial budgets for both optimisations, given the low dimensionality of both problems. This did result in excessive experimentation for the given problem, with the results of the 3D optimisation indicating a near optimum value occuring often approximately 40 iterations; similar behaviour is observed with the 2D optimisation but to a lesser extent. Such behaviour highlights a key disadvatage of budget based

termination criteria, with subsequent experimentation providing little to no improvement upon this value. Utilising calculated termination criteria, especially in global optimisation of black box problems, does present key challenges however, with the risk of premature termination a key factor in their uptake.[102]

1.2.3.3 Bayesian Optimisation

Bayesian optimisation is a derivative free global optimisation method that utilises surrogate models to optimise expensive to evaluate objective functions. It is most often used for systems which have a significant time, monetary or opportunity cost associated with each evaluation. Typically, Bayesian methods are applied to “black box” systems in which there is no derivative information or associated problem structure information for which other optimisation algorithms may be better suited.

The surrogate model is built using sampled data from the process/objective to be optimised. Once constructed, the surrogate model is employed in conjunction with an acquisition function to suggest the next evaluation point. Often, the surrogate model will be constructed using Gaussian process regression (GPR) which is computationally and resource-wise more efficient to evaluate than the actual system.

Gaussian Process Regression (GPR)

A Gaussian process (GP) defines a distribution over all possible functions, $f(x)$, given the observed data. More formally a GP is a collection of random variables, any finite number of which have a joint Gaussian distribution.[103]

A GP is specified by a mean $m(x)$ and covariance $k(x, x')$ as follows:

$$f(x) \sim GP(m(x), k(x, x')) \quad \textbf{(1-8)}$$

$$m(x) = \mathbb{E}[f(x)] \quad \textbf{(1-9)}$$

$$k(x, x') = \mathbb{E}\big[\big(f(x) - m(x)\big)\big(f(x') - m(x')\big)\big] \quad \textbf{(1-10)}$$

In GPR we need to initially define a prior over the distribution, this represents our prior beliefs as to what we expect to observe before seeing any data. It is common procedure to assume a prior distribution with a mean of zero, although this is not necessary.[104] Z-score normalisation of the input data to the GP can ensure this assumption holds true. Where:

$$Z = \frac{x - \mu}{\sigma} \quad \textbf{(1-11)}$$

In defining a prior we must also define a covariance function; this details our assumptions about the function we are attempting to model. The covariance function is used to calculate a similarity measure between two points. The covariance matrix is constructed utilising the covariance function by evaluating the function for each pairwise combination of the input data.

There exist several covariance functions, however, for a covariance function to be valid the resulting covariance matrix they produce must be positive semidefinite (PSD). A symmetric matrix is considered PSD if and only if (iff) all its eigenvalues are non-negative. Examples of commonly used covariance functions are the squared exponential, k_{SE}, and the Matérn, k_M, provided below:

$$k_{SE}(r) = \exp\left(-\frac{r^2}{2l^2}\right) \quad \textbf{(1-12)}$$

$$k_M(r) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}r}{l}\right)^{\nu} K_{\nu}\left(\frac{\sqrt{2\nu}r}{l}\right) \quad \textbf{(1-13)}$$

$K_{\nu}(x)$ is a modified Bessel function, with $\Gamma(x)$ being the gamma function, $(x-1)!$, and is defined as follows:

$$K_{\nu}(x) = \int_0^{\infty} exp\big(-xcosh(t)\big)\, cosh(\nu t)\, dt \quad \textbf{(1-14)}$$

ν is a positive parameter, which controls the smoothness of the function. There exist simplifications of the Matérn function for certain values of ν, such as for $3/2$ and $5/2$, provided below:

$$k_{M:\nu=\frac{3}{2}}(r) = \left(1 + \frac{\sqrt{3}r}{l}\right) exp\left(-\frac{\sqrt{3}r}{l}\right) \quad \textbf{(1-15)}$$

$$k_{M:\nu=\frac{5}{2}}(r) = \left(1 + \frac{\sqrt{3}r}{l} + \frac{5r^2}{3l^2}\right) exp\left(-\frac{\sqrt{5}r}{l}\right) \quad \textbf{(1-16)}$$

These simplifications are commonly used in machine learning applications. As $\nu \to \infty$ the Matérn covariance function converges to the squared exponential covariance function, k_{SE}.

r represents the distance metric of choice, with two common choices being Euclidean, r_{euc}, and Manhattan distance, r_{man},where:

$$r_{euc} = \sqrt{\boldsymbol{x} \cdot \boldsymbol{x}} \quad \textbf{(1-17)}$$

$$r_{man} = \|\boldsymbol{x} - \boldsymbol{x}\|_1 \quad \textbf{(1-18)}$$

l is known as the length scale and describes how far in the input domain you need to move before the function value changes significantly. For example, with larger length scale values, the GPs mean function would vary less over the input domain when compared with smaller values of length scale. This value is often optimised alongside other hyperparameters when initially fitting a GP model. We can draw samples from the prior, when using this covariance function, to generate a vector of random Gaussian points, Figure 1-13.

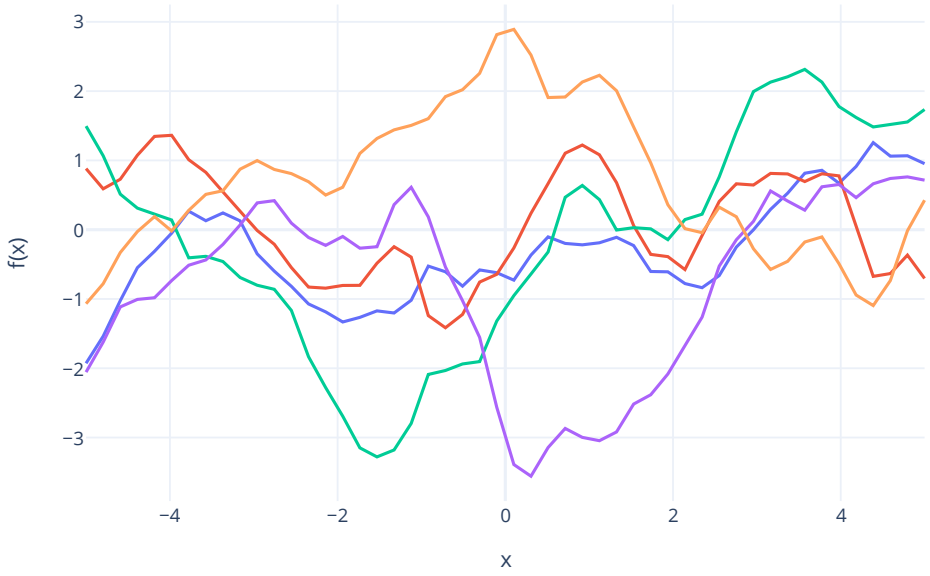


Figure 1-13. Five functions drawn randomly from a GP prior with a mean of zero and using the Matérn$_{3/2}$ covariance function.

We can infer how the inputs map to our function output using the posterior distribution over the possible functions. To do this, we must restrict the prior distribution to functions that agree with the observed data (when considering noiseless data), we can write this as:

$$\mathbf{f}_*|\mathrm{X}_*,\mathrm{X},\mathbf{f}\sim N\big(K(X_*,X)K(X,X)^{-1}\mathbf{f},\\ K(X_*,X_*)-K(X_*,X)K(X,X)^{-1}K(X,X_*)\big) \qquad \textbf{(1-19)}$$

We can understand this as the function values that we wish to calculate, $\mathbf{f}_*$, is dependent upon the input values we wish to determine our function values for, X_*, alongside our training data inputs, X, and outputs, $\mathbf{f}$.

To model noisy systems, Gaussian noise is often added to the covariance function, such that, $cov(\mathbf{y}) = K(X,X) + \sigma_n^2 I$. The degree of noise alongside other hyperparameters such as length scale and signal variance can be determined through maximising the marginal log likelihood. The marginal log likelihood is derived utilising a Bayesian approach considering

the likelihood of the output for the given input data and the selected hyperparameter, formally it is given as follows:

$$\log p(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{\theta}) = -\frac{1}{2}\boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{K}_{\boldsymbol{y}}^{-\mathbf{1}}\boldsymbol{y} - \frac{1}{2}\log|\boldsymbol{K}_{\boldsymbol{y}}| - \frac{n}{2}log2\pi \qquad \textbf{(1-20)}$$

$\boldsymbol{K}_{\boldsymbol{y}}$ is the covariance matrix of noisy response values, $-\frac{1}{2}\boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{K}_{\boldsymbol{y}}^{-\mathbf{1}}\boldsymbol{y}$, assesses how well the model fits the data given the current hyperparameters, $\boldsymbol{\theta}$. $-\frac{1}{2}\log|\boldsymbol{K}_{\boldsymbol{y}}|$ is included to act as a complexity penalty and prevent overfitting of the model. $-\frac{n}{2}log2\pi$ is a normalisation constant. The negative of the log likelihood is normally minimised to determine the optimal hyperparameters. As the function is non-convex, often a local optimum is found, however, in most situations the resulting fit is acceptable. Figure 1-14, details the posterior for an example function.

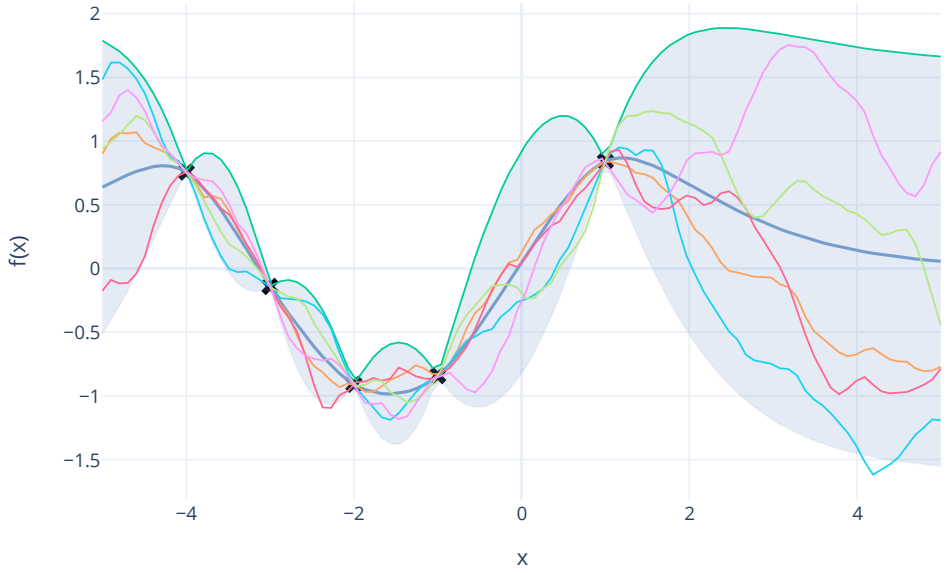


Figure 1-14. Posterior with five functions drawn at random (teal, red, orange, green, pink). Mean (blue) and 95% confidence interval (shaded area) for the data are provided. Noiseless training data (black) was generated using a sinusoid $\boldsymbol{f}(\boldsymbol{x}) = \mathbf{sin}\,(\boldsymbol{x})$.

As there is no noise in the system all posterior samples pass through the data points, this is due to the model variance tending to zero at these points. A widening of the confidence interval is noted in data sparse regions, this represents an increase in uncertainty of the model leading to the diverse nature of the sampled functions in these regions.

Acquisition Functions

Once a GP surrogate is constructed it can be used to suggest points of evaluation in an optimisation. Acquisition functions play a crucial role here, as they can utilise the GP surrogate to guide the optimisation towards its goal.

Probability of Improvement

One of the first acquisition functions to be utilised in Bayesian optimisation was probability of improvement (PI). Suppose we have a current best value, f^*, we want to evaluate our objective function at the point which has the highest likelihood to improve upon our current best.[105] The PI for a given point is calculated as:

$$\begin{aligned} PI(\mathbf{x}) &= P(f(\mathbf{x}) \geq f^*) \\ &= \Phi\left(\frac{\mu(\mathbf{x}) - f^*}{\sigma(\mathbf{x})}\right) \end{aligned} \qquad \textbf{(1-21)}$$

Where $\Phi(\cdot)$ is the Gaussian normal distribution function. The main drawback to utilising this acquisition function in its native form is that it is a purely exploitation algorithm. Given this, a trade-off parameter is often included to enforce a minimum improvement. The minimum improvement parameter attempts to enact an increase in the level of exploration to ensure the optimisation is sufficiently global. This modified version of PI is given as:

$$\begin{aligned} PI(\mathbf{x}) &= P(\hat{\mathrm{y}}(\mathbf{x}) \geq f^* + \varepsilon) \\ &= \Phi\left(\frac{\mu(\mathbf{x}) - f^* - \varepsilon}{\sigma(\mathbf{x})}\right) \end{aligned} \qquad \textbf{(1-22)}$$

The choice of ε can have a high impact on the efficacy of the optimisation and often requires tuning for each specific problem.

Expected Improvement

Expected improvement (EI) expands upon the probability of improvement by considering how much we are likely to improve upon the current best value, (1-23).[106,107]

$$EI(\mathbf{x}) = \left(\mathrm{f}^* - \hat{\mathrm{y}}(\mathbf{x})\right)\Phi(Z(\mathbf{x})) + \sigma(\mathbf{x})\phi(\mathrm{Z}(\mathbf{x})) \qquad \textbf{(1-23)}$$

$$Z(\mathbf{x}) = \frac{\mathrm{f}^* - \hat{\mathrm{y}}(\mathbf{x})}{\sigma(\mathbf{x})} \qquad \textbf{(1-24)}$$

Where $\phi(\cdot)$ is the probability density function. We can consider the EI closed form function in two parts, one considering exploitation, $\left(\mathrm{f}^* - \hat{\mathrm{y}}(\mathbf{x})\right)\Phi(Z(\mathbf{x}))$, and the other considering exploration. The benefit this presents over probability of improvement is that it is not necessary to define a required minimum level of improvement, removing an additional hyperparameter that

may need to be optimised. Moreover, it has been shown to guarantee convergence upon the global optimum for a system, if given a sufficient budget.[108]

A comparison between the native probability of improvement and expected improvement acquisition functions is provided in Figure 1-15. We can clearly see the drawback in using the native probability of improvement, with the progress solely focused around the current best point with no exploration. Conversely, observing the iterations of expected improvement, the algorithm effectively balances exploration and exploitation, building a more complete picture of the underlying process.

Applied examples of Bayesian methodologies for reaction optimisations have been limited. However, there has been an increase in its adoption in recent reaction optimisation publications. The deMello group were among the first to apply Bayesian techniques to automated reaction optimisation. The authors utilised a target adapted PI based methodology for the targeted production of nanocrystals with desired properties, Figure 1-16.[53] From an initial 20 experiments the authors applied their methodology to guide the targeted optimisation.

Furthermore, as a secondary objective, the authors sought to gain an accurate empirical model of the process provided by the applied Gaussian process surrogate. The application highlights the efficiency improvements afforded through use of Bayesian methods, with the algorithm able to determine target regions as well as build an accurate surrogate model. However, as illustrated in Figure 1-15, employing a PI based approach does have limitations in the wider exploration of the system, with there being many alternative acquisition functions that account for the exploration/exploitation trade-off to a greater degree.

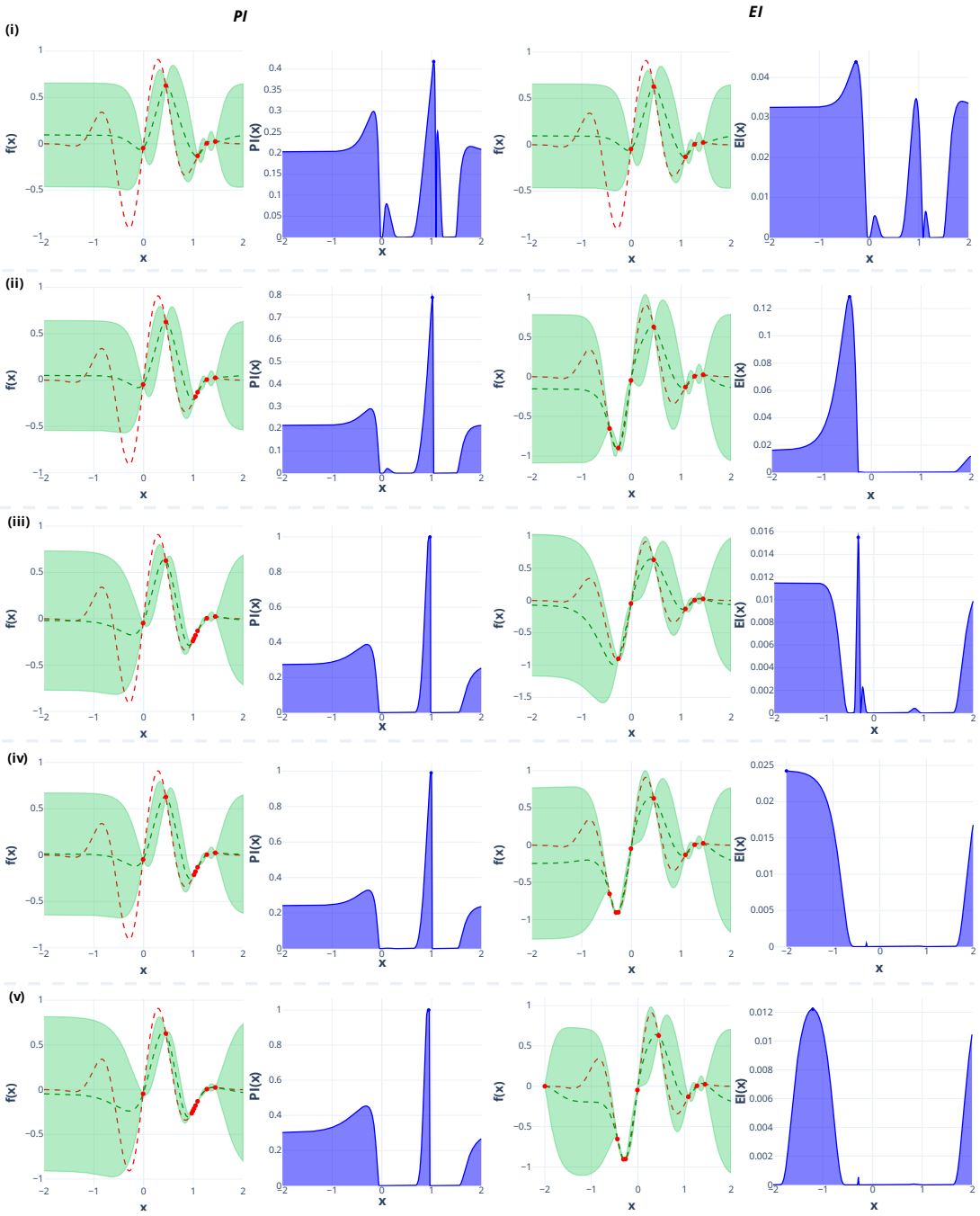


Figure 1-15. Comparison of PI (left) and EI (right) acquisition functions for a single objective 1D problem. Each row represents an iteration (i-v). Both acquisition functions were given identical starting conditions and utilised the same random seed. The true function is shown as the dashed red line, whilst the current mean function for GP surrogate is shown as the green dashed line, with associated shaded confidence interval. The acquisition function for the domain is shown as blue shaded areas with the next sample selected as the value which maximises this function. Observations are shown as red dots.

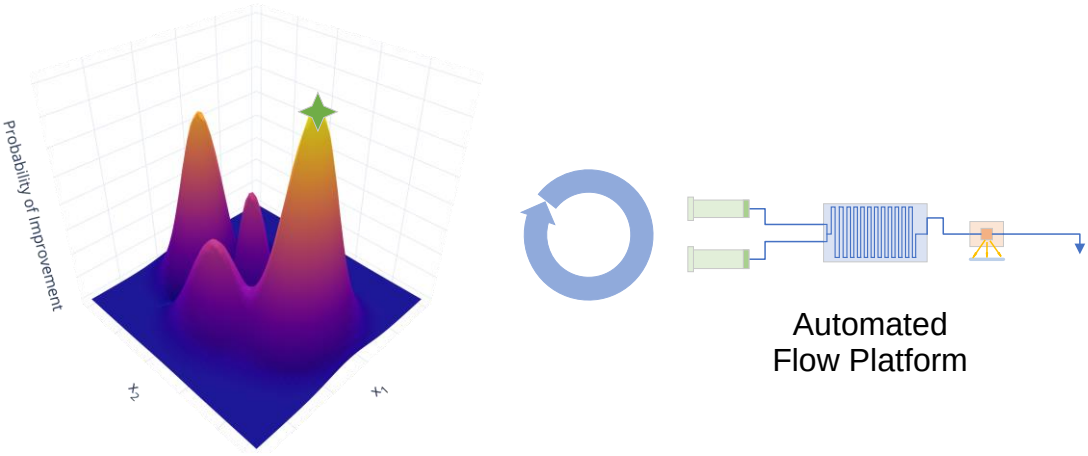


Figure 1-16. Simplified overview of targeted adapted PI based methodology employed by the deMello group.[53] Algorithm iteratively proceeds by evaluating target adapted PI for the entire input domain (LHS), from which a candidate experiment(s) is/are selected and performed using an automated flow platform (RHS).

1.2.4 Multi-Objective Optimisation

Many systems can be considered as multi-objective optimisation problems, in that there may be multiple and conflicting objectives that need to be considered concurrently, such as space-time-yield and E-factor. Productivity and environmental metrics often conflict, with increased throughput not necessarily correlating with improved reaction performance, leading to greater waste generation.[45] As the objectives conflict, there is often no single optimum solution to the problem. Instead, there is an optimal set of solutions highlighting the trade-off between objectives, this is known as the Pareto front, Figure 1-17.[109] In fluid design situations generation of a set of solutions is highly beneficial as it allows the user to not only understand the system to a greater extent but also, should design requirements change, the user can return to the optimisation results and select alternative conditions that meet their new requirements. This is in stark contrast to single objective procedures where changing objective function would likely require a partial or even full recompletion of the optimisation.

In dealing with multiple objectives, Pareto dominance becomes an important concept as a way of describing how efficient a point is in terms of all objectives. Pareto dominance can be defined when considering two objective space vectors $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$. The point $\mathbf{y}^{(1)}$ is said to dominate $\mathbf{y}^{(2)}$ iff:

$$\forall i \in \{1, \dots, m\}: \mathbf{y}_i^{(1)} \leq \mathbf{y}_i^{(2)} \; and \; \exists j \in \{1, \dots, m\}: \mathbf{y}_j^{(1)} \leq \mathbf{y}_j^{(2)} \qquad \textbf{(1-25)}$$

Meaning that $\mathbf{y}^{(1)}$ is not worse than each of the objectives in $\mathbf{y}^{(2)}$ and better in at least one objective. A set of these non-dominated solutions forms the Pareto front for the system.

There are three general approaches for tackling multi-objective optimisation problems: (i) *a priori*, where multiple objectives are combined into a single objective function which is then optimised; (ii) *a posteriori*, where the algorithm operates with a set of candidate solutions aiming to progress the candidate set towards the optimal Pareto set; (iii) interactive, where the user inputs their preferences upon the objective functions as the algorithm progresses.[110]

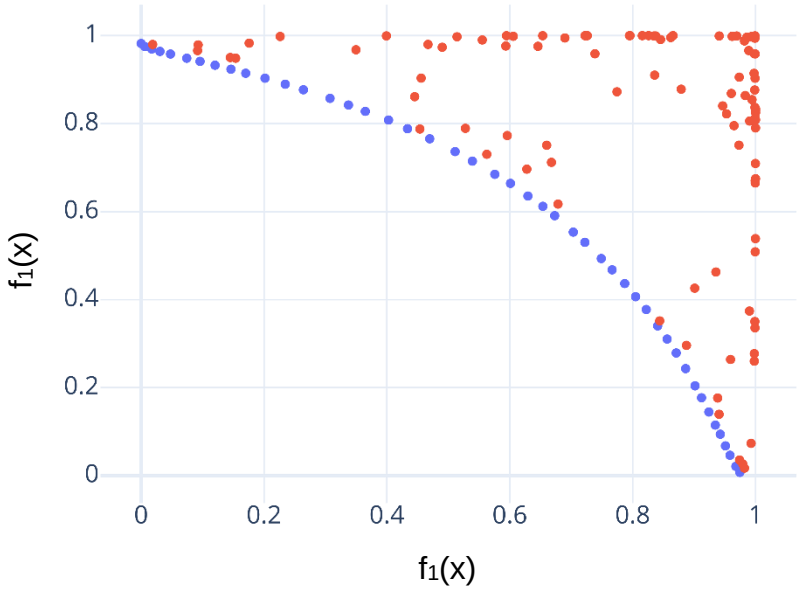


Figure 1-17. Example solution to a bi-objective optimisation problem, in which both objective functions are being minimised. The Pareto front is shown in blue, illustrating the trade-off between the two objectives, with the dominated (non-optimal) solutions shown in red.

More recently Häse *et al*. presented a scalarisation method based on a hierarchal ranking of objectives in order of importance.[77] The authors devised the method to account for systems where experimental budget may be limited, and instead of realising the full Pareto front, a subsection is targeted based upon the users *a priori* ranking. The approach utilises lexicographical methods to ensure while each objective is sequentially optimised, in order of their ranking, that the degradation of prior optimised objective does not fall below an allowable tolerance.

As with other *a priori* methods, a full elucidation of the trade-off between the objectives is sacrificed in lieu of optimisation efficiency towards a pre-defined objective of highest importance. Such methods can prove useful

for efficient optimisation when the user knows their ranking of objectives will not change. However, in many applications of multi-objective algorithms this *a priori* objective ranking may not be desired, with it possibly limiting exploration of alternative interesting competing regions. Additionally, as with more conventional scalarisation methods should the users ranking change, a repeat optimisation would likely be required.

Conversely to *a priori* methods, *a posteriori* approaches result in a set of non-dominated solutions without the user placing any weighting or bias towards any of the objectives. NSGA-II represents a typical algorithm from this classification, which utilises non-dominated sorting and crowding distance to iteratively evolve the population of the genetic algorithm. A simplified overview is given by Algorithm 1-1.

Algorithm 1-1: NSGA-II simplified overview

1: **begin**
2: $P_0 \leftarrow initialise()$
3: **while** $exit == False$ **do**
4: $Q_k \leftarrow GASelectCombineMutate(P_k)$
5: $B_k \leftarrow nonDomSort(P_k \cup Q_k)$
6: $P_k \leftarrow crowdingdistance(B_k)$
7: $k \leftarrow k + 1$
8: **end**

The algorithm begins with the generation of an initial population. Following this, the main algorithm iteration cycle begins, with the population undergoing selection, combination (crossover) and mutation. This stage is common to most genetic algorithms. The modified population and the original population are then merged, and non-dominated sorting is performed. The sorted points then undergo further ranking based on the crowding distance criterion. This is done to ensure an acceptable level of diversity is maintained for the population with respect to the approximate Pareto set.

As with most evolutionary algorithms the function evaluation requirement for an effective optimisation is high. However, NSGA-II still represents an effective algorithm for solving challenging multi-objective optimisation problems where there are no resource constraints.

1.2.4.1 Performance metrics

As the solution of a multi-objective problem is usually an optimum set of points, single point comparison between algorithms does not provide a complete overview of the relative strengths and weaknesses of an approach.

Given this, there has been significant research in developing methodologies for comparing multi-objective algorithm performance, with a wide array of performance metrics available, each with their own merits. An overview of some of the most commonly used metrics is provided below, however, Riquelme *et al*. provide a comprehensive review of all published multi-objective metrics with their associated advantages/disadvantages.[111] Table 1-3 provides a summary of the commonly utilised multi-objective performance metrics, with further discussion on some of the most insightful metrics provided later.

Table 1-3. Summary of commonly utilised multi-objective performance metrics.[111]

Metric	Insight
Hypervolume	Accuracy and diversity
Generational distance	Accuracy
Inverted generational distance	Accuracy and diversity
Summary surface	Qualitative comparison
Delta indicator	Diversity

Hypervolume

Hypervolume is one of the most frequently applied performance metrics when analysing multi-objective optimisation results. The metric relies on a reference point from which the area covered by an approximate set is calculated.[112] Formally, the hypervolume indicator is given as follows for a given set of points, S, and reference point, r:

$$HV(S) = \Lambda\left(\bigcup_{\substack{p \in S \\ p \preceq r}} [p, r]\right) \qquad \textbf{(1-26)}$$

Where $\Lambda(\cdot)$ is the Lebesgue measure and $\cup$ is the union operator. A visualisation of a 2-D hypervolume is provided in Figure 1-18.

The hypervolume indicator has found popularity due to the fact it encapsulates many aspects relating to performance such as accuracy and diversity of the approximate set. For comparative purposes if an approximate set dominates another set, its hypervolume will always be greater than that of the dominated set.

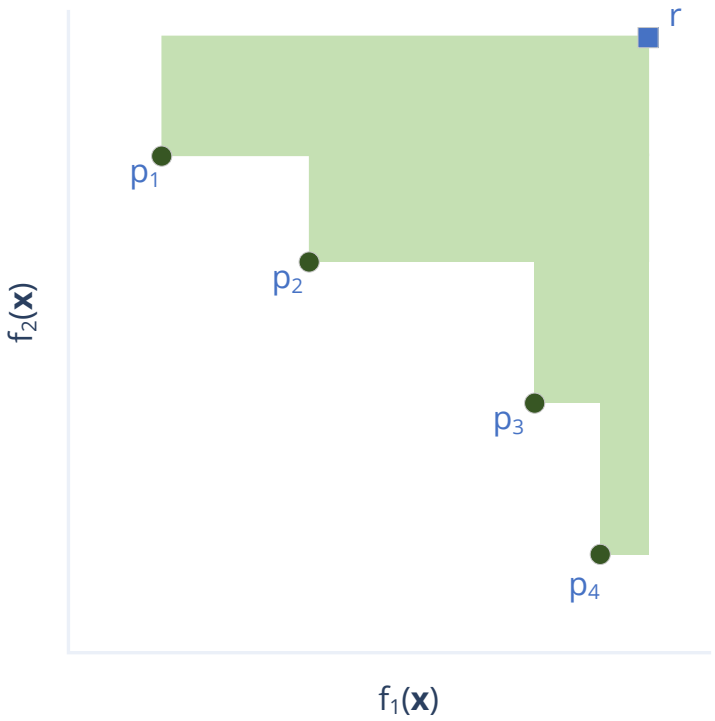


Figure 1-18. Hypervolume visualisation in 2-D objective space. Where p are the current non-dominated points, and r is the reference point from which the hypervolume is calculated from.

Inverted Generational Distance

Originally proposed by Coello and Reyes-Sierra, the inverted generational distance metric was proposed as an improvement upon generational distance to provide a measure for comparing an approximate front to the true Pareto front for a system, (1-27).[113]

$$IGD(R,S) = \frac{1}{|R|}\left(\sum_{p\in S} \min_{r\in R} d(r,p)^q\right)^{\frac{1}{q}}, d(r,p) = \sqrt{\sum_{k=1}^{M}(r_k - p_k)^2} \qquad \textbf{(1-27)}$$

Where S is the approximate set, R is the set of true Pareto points and M is the number of objectives. To enable weak Pareto compliance, which was not achieved in the original metric, the modified Inverted Generational Distance (IGD+) performance metric was proposed by Ishibuchi *et al*..[114] IGD+ is calculated as follows:

$$IGD^+(S) = \frac{1}{|R|}\left(\sum_{i=1}^{|R|} d_i^{+2}\right)^{\frac{1}{2}} \qquad \textbf{(1-28)}$$

Where $d_i^+ = \max\{p_i - r_i, 0\}$, with a_i being a point from the calculated Pareto front S and r_i being a point from the true Pareto front, R.

Worst Attainment Surface

An attainment surface is defined by a set of Pareto points and defines an estimated Pareto front, for a given algorithm, for an experimental run.[115] For test problems, that have been tested multiple times for a given algorithm, each run will have an attainment surface which will provide an overall set of attainments surfaces.

Worst attainment surfaces are defined by a set of Pareto points for a specific algorithm and test function that are dominated by all other points for the specific algorithm on a specific test function. They highlight the divide between dominated and non-dominated points for the given set. This provides insight as to where the various algorithms perform differently and to what extend; thus, it can highlight a specific algorithm's weakness. An example of three summary surfaces and best and worst summary surfaces are highlighted in Figure 1-19 and Figure 1-20, respectively.

Although not a quantitative metric, summary surfaces provide useful qualitative insight into the strengths and weakness of a certain algorithm for a given problem.

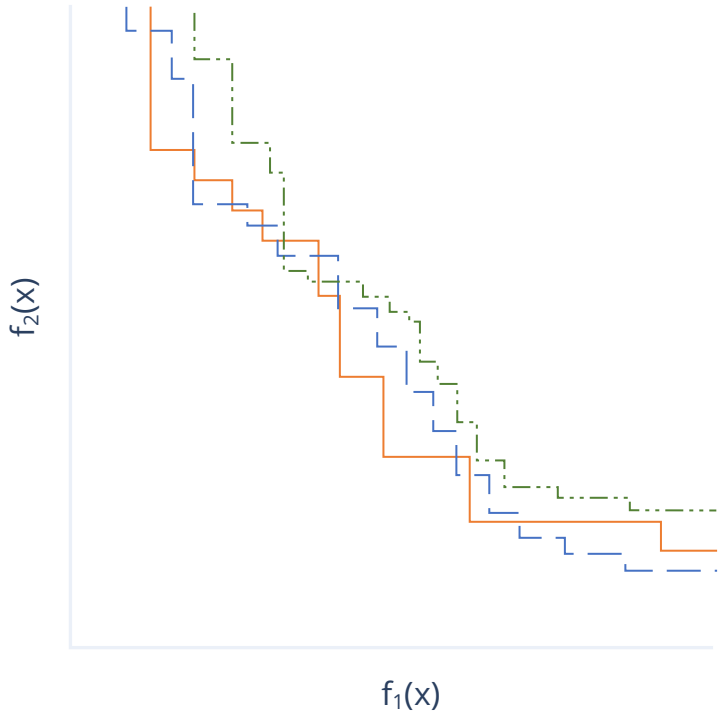


Figure 1-19. An example of three summary surfaces for a 2D problem, each summary surface is shown in a different colour and line style.

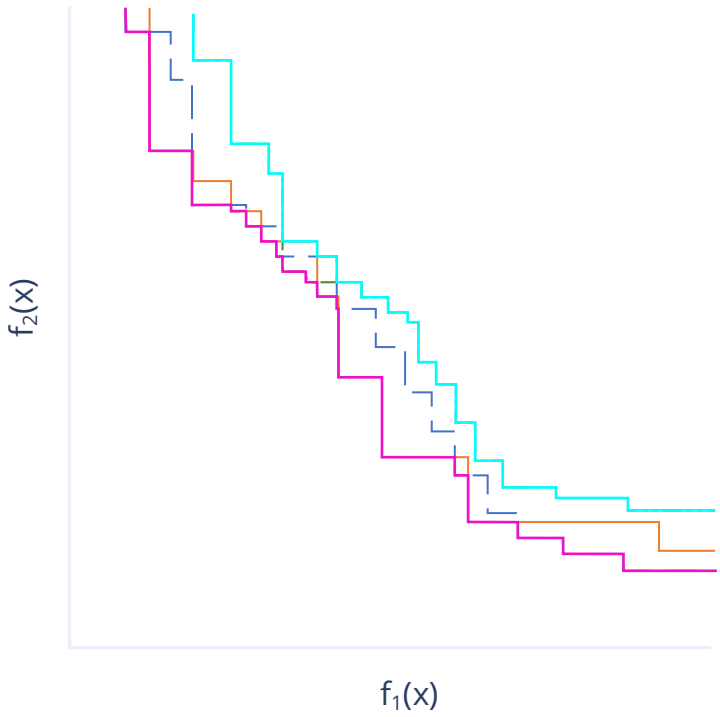
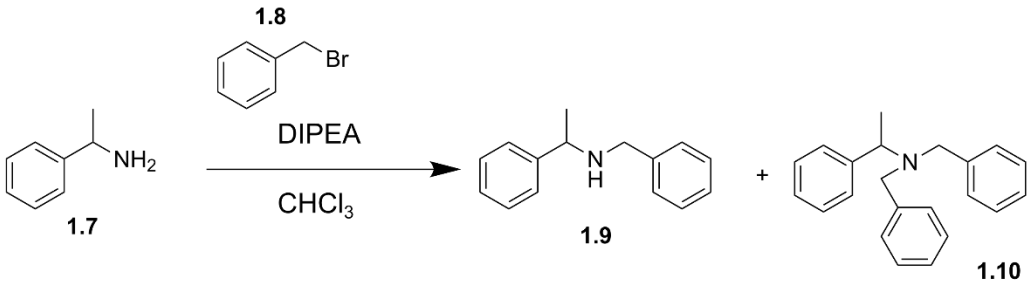


Figure 1-20. Best (shown in pink) and worst (shown in teal) summary attainment surfaces from Figure 1-19. The best summary surface considers all summary surfaces and highlights the best possible Pareto front when considering all points. The worst summary surface illustrates the opposite and highlights the worst possible front when considering all points of the summary surfaces.

Schweidtmann and Clayton detailed the first application of an *a posteriori* based multi-objective algorithm to an automated chemical system optimisation.[45,64] The authors optimised a wide variety of chemistries for both bi-objective and tri-objective optimisation systems. For example, the bi-objective optimisation of a N-benzylation reaction, Scheme 1-3.



Scheme 1-3. N-benzylation reaction for bi-objective optimisation. Impurity (**1.10**) was minimised alongside the maximisation of space-time-yield.[45]

In the optimisation the maximisation of space-time-yield and minimisation of a reaction impurity were simultaneously optimised, resulting in a Pareto front highlighting the trade-off between the two objectives. For all cases, the algorithm was able to converge within a similar timeframe as

previous single objective optimisations providing a set of Pareto points highlighting the trade-off between conflicting environmental and economic factors. As the approach employed an *a posteriori* based method, the authors were able to obtain a diverse Pareto front which could be utilised to inform process development further down the line. The benefits of adopting such approaches are clear, with the need to select objective preference removed, as the algorithm can highlight the global optimum with respect to all objectives, as well as the trade-off curve, when given sufficient budget.

One notable disadvantage of the method employed is the lack of termination criteria, possibly resulting in excessive experimentation for the task at hand. However, in the experimental applications of the TSEMO algorithm,[45,64] Pareto optimal points occurred towards the end of the experimental budget, suggesting the algorithm is continually improving upon the current optimal set.

1.3 Discussion

When selecting an algorithm, it is important to consider the system and the user's requirements. Given the range of possibilities, Figure 1-21 has been created to guide this process based upon the system and requirements.

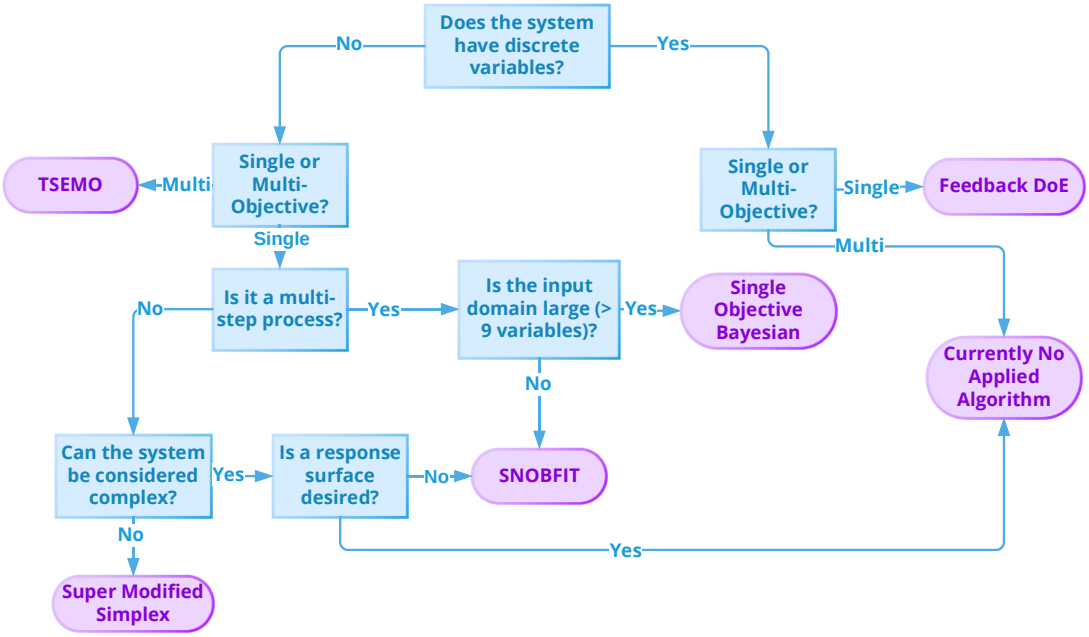


Figure 1-21. Algorithm selection flowchart detailing the recommended algorithm for a given optimisation task.

For discrete variable systems, the algorithm developed by the Jensen group represents the main published example of mixed continuous and discrete variable optimisation for self-optimising systems. The complexity of

the algorithm may perturb users from its use, with the algorithm requiring the assumption of a simple system model in its derivation. There are currently no published examples of multi-objective automated optimisation of chemical systems with discrete variables. Additional development in both the single and multi-objective cases is required to allow for 'black-box' usage with the ability, should the user desire, to ascertain dependencies upon process variables.

For systems where the sole consideration is the optimisation of continuous variables, for the multi-objective case, the TSEMO algorithm presents the only reported instance where *a posteriori* multi-objective optimisation has been performed in a self-optimising system. As such, it represents the only verified option for multi-objective chemical system optimisation. However, as this field is continually developing alternative approaches may offer improvements in terms of optimisation efficiency.

Algorithm selection for single objective systems is dictated by the complexity of the system, and if there are any additional user requirements. For 'simple' systems, where the response is known to be smooth and possibly convex, an efficient local search algorithm will suffice. The Felpin group's modified simplex has been shown to provide this capability, with efficient determination of process local optima. In-built restarts and efficient handling of boundary and linear constraints make it the most capable simplex algorithm applied in self-optimising systems.[62] The super modified simplex can be used in place of the Felpin group's, should availability prove to be an issue, with the super modified algorithm showing excellent results in terms of fast convergence.[26]

For more complex systems, SNOBFIT would be recommended as the algorithm of choice. The algorithm has been successfully applied to a wide range of systems and has shown excellent capabilities in determining the optimum region for the system in an efficient manner.[21,25] Should the user have a system with a large input domain, this could present problems for the SNOBFIT algorithm. Although not readily applied, in such instances a Bayesian approach may prove beneficial with Bayesian algorithms capable of handling larger domain sizes without as significant drop in performance.[101] In such cases expert knowledge can prove to be particularly useful in optimal configuration of a Bayesian algorithm, given the prerequisite to define hyperparameters, such as surrogate model kernel and acquisition function.

1.4 Project Aims

Much of the previous work in the field has utilised similar algorithms; be it local optimisation, or global optimisation. In this body of work there has been little attempt to employ hybrid approaches, coupling local and global algorithms to improve efficiency of optimisation. The only documented use of a hybridised method has been limited to coupling local optimisation algorithms together to improve their efficiency.[90] Therefore, there is clear scope to expand in this relatively unexplored field, either focusing solely on optimisation or coupling with other user requirements such as accurate response surface mapping.

Mixed variable optimisation is optimisation of an objective function which has an input domain containing both qualitative and quantitative variables. Optimisation of this type can be found in areas such as aerospace designs[116] or reaction screening.[39] To date there has been limited work on this subject in the field of chemical reaction self-optimisation. The Jensen group's research in this area represents the sole venture into this domain of self-optimisation research, likely due to the complexity of the optimisation and the requirement for specialised equipment.[33,39,51,52]

Given the lack of further exploration in this area, there is a clear opportunity to capitalise on this and develop the topic of self-optimisation for mixed variable reaction systems, especially concerning system optimisation with multiple objectives. Providing an efficient approach to optimise these systems will prove a key development area, especially for early-stage data acquisition applications.

Chapter 2 initially introduces mixed variable multi-objective optimisation, highlighting development areas from which a novel Bayesian multi-objective algorithm is proposed. The algorithm is subsequently tested, employing three test problems of varying difficulty, with comparison to widely available alternative optimisation techniques.

Chapter 3 provides detail on the experimental application of the previously discussed novel mixed variable multi-objective algorithm. Two example reactions are investigated: the bi-objective optimisation of a S_NAr reaction and the tri-objective optimisation of a Sonogashira reaction.

Chapter 4 describes the development and application of a hybrid optimisation and response surface mapping algorithm. An exemplar photochemical reaction is optimised utilising the novel approach providing the first reported example of an experimental two stage optimisation.

Improvements to the original approach are then suggested with performance characterisation achieved employing data collected from the photochemical reaction.

Chapter 2 Mixed Variable Multi-Objective Optimisation

2.1 Introduction

Many real-world optimisation problems can be composed of both continuous and discrete variables. Continuous variables are variables in which we have access to any real value in the desired optimisation range of the variable. Given their inherent continuous nature, they are often easier to optimise, with a wider array of applicable optimisation techniques. Discrete variables, however, have a limited selection of options and can take the form of integer or categorical values (materials, reaction solvents etc.). These mixed variable problems can often be difficult to solve, requiring specific techniques to overcome the challenges they present.

Commonly, the discrete/qualitative variables are projected to the continuous domain using known relationships or underlying properties of the qualitative variable (polarity index of a solvent or tensile strength of a material).[18,71,117–119] For systems where the relationship between continuous descriptors and the objective space is well understood, this can provide an easy and robust way to approach mixed variable optimisations. Rakshit and Ananthasuresh utilised this technique for the optimisation of trusses when considering both geometry and material selection.[119] The authors employed the underlying design index of a material to enable the conversion to the continuous domain for material selection. A gradient-based optimisation method was then employed for the simultaneous selection of geometry and material, resulting in an optimisation of equal efficiency to methods where material choice was fixed.

However, there is a need for algorithms that can handle mixed data types for systems in early development, or where mapping to continuous variables is unavailable, such as, different equipment for a manufacturing process or novel chemistries where the underlying mechanism is still not understood. In such instances application of conventional continuous variable methodologies is not applicable, with these methods relying on underlying properties of continuous systems, such as differentiability.[120] Furthermore, as the number of discrete variable combinations increase, this can result in vast search domains, which, even for non-expensive objective functions, can result in challenging optimisation problems.

In addition to mixed variable input domains, consideration of multiple objectives is often a requirement, with the objectives frequently conflicting with

additional process constraints. Commonly, in multi-objective problems, there does not exist a shared optimum for all objectives. Instead, the solution to the problem will be a trade-off between the objectives; this non-dominated set is known as the Pareto front.[121] Illustrating the optimal trade-off between the objectives, the Pareto front can be utilised to inform design or process considerations in situations where the design requirements may change.

In general, mixed variable multi-objective optimisation techniques have received less attention than other categories of optimisation. Developments in the field have predominantly focused on the application of meta-heuristics, such as genetic algorithms (GAs), to tackle these problems. Prior empirical or mechanistic modelling is often a prerequisite, when utilising these techniques to solve real world problems, due to the high function evaluation requirement of GAs.[122–128] This approach has been successfully applied to the optimisation of train traction systems in which a derived simulation model was utilised for the bi-objective optimisation of investment cost and energy usage for the simulated train system across an array of track and landscape configurations.[128] An alternative approach was adopted by Brownlee and Wright in which they detail the use of radial basis function networks (RBFN) utilised as a surrogate model coupled with NSGA-II for the bi-objective optimisation of building designs, considering both capital cost and annual energy usage.[122] The RBFN surrogate was trained and validated using expensive building performance simulations which allowed for the high function evaluation overhead associated with NSGA-II. Upon each completed iterations of NSGA-II the true function (the building performance simulation) was evaluated for the top solutions and then the surrogate was retrained and applied again with NSGA-II. Across all test problems presented by the authors the surrogate-based methods offered significant improvements in terms of optimisation efficiency achieving significant reductions in the number of evaluations required to achieve a performance threshold. Although improvements were observed and acceptable for the authors requirements, evaluations remained comparatively high for an expensive optimisation problem, with 1000s of true function evaluations required for the determination of an acceptable solution. This presents a key issue for systems where this high function evaluation demand cannot be achieved.

For all highlighted problems the need for accurate prior modelling work is a necessity to success, for systems where direct optimisation is a requirement (chemical processes) or those that are a black-box (either unknown or too complex to model) such detailed models may not be available

and as such there is a need for algorithms that can effectively optimise these systems without the need for prerequisite work to be performed.

In this chapter, we present a mixed variable multi-objective Bayesian optimisation algorithm, providing three test cases to review the algorithms performance. The algorithm looks to extend Bayesian multi-objective methodologies to the mixed variable domain, which to the best of the authors' knowledge has limited prior work.[129] A review of surrogate modelling methodologies is initially provided, alongside available acquisition functions for multi-objective optimisation problems. Comparison between random sampling and a mixed variable version of NSGA-II provided in jMetalPy[130] is then performed with the relative merits of the algorithm highlighted.

2.2 Algorithm Development

2.2.1 Surrogate Model

Many efficient global optimisation algorithms utilise an underlying surrogate model to reduce the number of evaluations of the actual objective function. These surrogates can take many forms with details of two main categories provided below.

Non-Gaussian methodologies

Random forests are an ensemble method which average the predicted response of a collection of classification and regression tree (CART) models. CART models utilise a series of decisions based on the variable values to best explain the modelled dependent variable. This results in a tree structure where the terminal nodes of the tree provide a prediction of the dependent variable.[131] As the method takes the average of multiple CART models, this dramatically improves the prediction quality when compared to a basic CART model.[132] Issues can arise when utilising random forests for prediction in unseen regions of the search domain.[133] This presents an issue for low budget optimisation task where the training dataset is limited.

The Adaptive COmponent Selection and Shrinkage Operator (ACOSSO) estimate is a smoothing spline ANOVA modelling methodology. It is a modification of the COSSO estimate,[134] which applies calculated weightings to the penalty functions to prevent over smoothing of the splines that can be experienced in standard COSSO estimation.[135] The set of splines are fitted through minimisation of the following:

$$\frac{1}{N}\sum_{n=1}^{N}[y_n - f(\mathbf{w}_n)]^2 + \lambda_1 \sum_{i=1}^{I} v_i \left\{\left[\int_0^1 g_i'(x_i)dx_i\right]^2 + \int_0^1 [g_i''(x_i)]^2 dx_i\right\}^{1/2} + \lambda_2 \sum_{i=1}^{I} w_j \left\{\sum_{z_j}^{b_j} h_j^2(z_j)\right\}^{1/2} \quad \textbf{(2-1)}$$

$$v_i = \left[\int_0^1 \tilde{g}_i^2(x_i)dx_i\right]^{-1}$$

$$w_j = \left(\frac{1}{b_j}\sum_{z_j=1}^{b_j} \tilde{h}_j^2(z_j)\right)^{-1}$$

The terms within the cost function correspond to the following: $\frac{1}{N}\sum_{n=1}^{N}[y_n - f(\mathbf{w}_n)]^2$ ensures there is an adequate fit to the data; $\lambda_1 \sum_{i=1}^{I} v_i \left\{\left[\int_0^1 g_i'(x_i)dx_i\right]^2 + \int_0^1 [g_i''(x_i)]^2 dx_i\right\}^{1/2}$ controls the roughness of the splines estimate and contains a penalty associated with continuous variables; $\lambda_2 \sum_{i=1}^{I} w_j \left\{\sum_{z_j}^{b_j} h_j^2(z_j)\right\}^{1/2}$ is a penalty associated with the discrete variables. The two penalty terms act as variable selection and dimensionality reduction. There are two tuning parameters λ_1 and λ_2 which are set utilising cross validation with the training data. Although providing a good framework for modelling, comparison with Gaussian based methods by Swiler *et al*. indicated that Gaussian based approaches had a more consistent performance across an array of surrogate modelling problems.[136]

As there is a wide precedent for utilising Bayesian approaches for expensive optimisation tasks,[137] employing Gaussian methods for the underlying surrogate model may prove beneficial.

Gaussian methodologies

Gaussian processes have found extensive application in efficient global optimisation tasks. Primarily, this is due to their ability to adequately model in data sparse problems and that they produce a probabilistic prediction which can be utilised to guide further exploration of the input domain.[104]

The difficulty for mixed variable optimisation problems, when using a GP as the surrogate for the response, is formulating a proper correlation structure. As discrete variables can vary in similarity by a large degree, assuming continuous methods can be applied to determine the covariance

matrix is not appropriate and may result in an invalid covariance matrix. The covariance function that characterises a GP is customisable by the user, however, to ensure that it is valid there is a requirement for the covariance function to be symmetric and positive semi-definite; this will be true if its eigenvalues are all non-negative.

A naive approach would be to model each discrete variable combination with an individual gaussian model. A key issue with this approach is the large number of hyperparameters this generates, with each model requiring adequate data to obtain acceptable estimates for these hyperparameters. Additionally, interactions between discrete variable combinations would be completely discounted adopting this approach.

To circumvent this, previous approaches propose modifications to the covariance function; ensuring a valid covariance matrix can be defined for mixed inputs.[138–141] Qian *et al*. were amongst the first to implement GP modelling for systems with mixed variable types.[138] The authors assumed the following covariance structure

$$Cor\big(f(\mathbf{x},\mathbf{d}),\mathbf{f}(\mathbf{x}',\mathbf{d}')\big) = \boldsymbol{\tau}_{\mathbf{d},\mathbf{d}'}\exp\left(-\sum_{i=1}^{|x|}\theta_i(x_i - x_i')^2\right) \qquad \textbf{(2-2)}$$

With $\boldsymbol{\tau}_{\mathbf{d},\mathbf{d}'}$ the correlation between the responses corresponding to discrete levels d and d' and is an element of the correlation matrix $\boldsymbol{\tau}$. The method entails an extensive estimation procedure, requiring the solution of an internal semidefinite programming problem to ensure a valid covariance matrix is computed during model training. Due to the large number of hyperparameters that require fitting, it can be difficult to obtain their optimum value, especially in efficient optimisation applications where the number of evaluations is often restricted. Zhou, Qian and Zhou later simplified the estimation of $\boldsymbol{\tau}$ utilising a hypersphere decomposition to ensure a valid covariance matrix was constructed.[139] Although the authors provided a simplification of the covariance matrix estimate, the method still requires a large number of hyperparameters to be fitted adequately. Given this, alternative methods employing a reduced number of hyperparameters will be considered.

Latent variables

Latent variable GPs (LVGP) usually find application in dimensionality reduction for high dimensional input data. In this area of application, the high

dimensional data is mapped to a set of underlying latent variables which describe the relationship at a lower dimension.[142]

Latent variable approaches have recently been proposed as a method to handle discrete variables. In the method, the differing levels for each discrete variable are mapped to a set of continuous latent variables that effectively describe the discrete relationship.[143] The authors propose a methodology in which each discrete variable and its associated number of levels are mapped to the 2D latent space. Estimation of the latent variables was performed through application of maximum likelihood estimation, alongside the estimation of other hyperparameters. Once optimal values for the latent variable mapping have been determined, the variables can be utilised in unmodified covariance functions. As the authors enforce the first level of the discrete variable to always be mapped to the origin, (0,0), their approach requires the estimation of $2m_j - 3$ parameters for each discrete variable, where m is the number of levels of the j^{th} discrete variable.

Gower similarity

Distance measures provide an insight into the relative similarity between observations with multiple input variables. They have found extensive use, not only in GP regression applications, but also in clustering algorithms for unsupervised learning.[144] For continuous variables, there exists an extensive collection of metrics such as Manhattan and Euclidean distance which have been previously discussed. For observations with a mixture of quantitative and qualitative variables, calculating the similarity between these observations is a non-trivial task.

One such metric capable of handling mixed data types is the Gower similarity.[145] Originally proposed as a similarity measure, further work by Halstrup investigated its use as a distance metric for GP modelling of mixed variable systems.[146] Formally the metric is defined as

$$r_{gow}(\boldsymbol{w}^i, \boldsymbol{w}^j) = \frac{\sum_{k=1}^{k=|\mathbf{x}|} \frac{\left|x_k^i - x_k^j\right|}{\Delta x_k}}{|\mathbf{x}| + |\mathbf{d}|} + \frac{\sum_{m=1}^{m=|\mathbf{d}|} s_{w_m^i w_m^j}}{|\mathbf{x}| + |\mathbf{d}|} \qquad \textbf{(2-3)}$$

where $\mathbf{w} \ni [\mathbf{x}, \mathbf{d}]$ with $\mathbf{x}$ a vector of the quantitative variables and $\mathbf{d}$ a vector of the qualitative variables. The first term considers the quantitative factors and is the weighted Manhattan distance between the two variables, where Δx_k is the range of the k^{th} quantitative variable, $s_{w_m^i w_m^j}$ considers the m^{th} qualitative variable. This is set to 0 if w_m^i is equal to w_m^j and 1 if they are not equal. $|\mathbf{d}|$ and $|\mathbf{x}|$ are the number of qualitative and quantitative variables,

respectively. The distance metric can then be applied in any of the standard covariance functions that are utilised in Gaussian process regression, such as the Matérn 5/2 covariance function.

$$K^{GMat_{\frac{5}{2}}}(\boldsymbol{w^i}, \boldsymbol{w^j}|\boldsymbol{\theta}) = \prod_{i=1}^{k}\left(1+\sqrt{5}\frac{k}{\theta_i}\cdot r_{gow}(\boldsymbol{w}^i,\boldsymbol{w}^j) + \frac{5}{3}\left(\frac{k}{\theta_i}\cdot r_{gow}(\boldsymbol{w}^i,\boldsymbol{w}^j)\right)^2\right) exp\left(-\sqrt{5}\frac{k}{\theta_i}\cdot r_{gow}(\boldsymbol{w}^i,\boldsymbol{w}^j)\right) \quad \textbf{(2-4)}$$

Pelamatti *et al*. compared the Gower similarity based method with other covariance function variations to analyse their performance in optimisation tasks.[116] The authors compared the method to hypersphere decompositions of the kernel as well as separate models for each discrete variable combination. Comparable performance was noted for the Gower method, whilst keeping the number of hyperparameters required to be optimised to a minimum, when compared to the other methods, with only one hyperparameter per discrete variable required. Although beneficial for estimation procedures, the limited number of hyperparameters per discrete variable may present issues in the ability for the GP to accurately model systems with many levels per discrete variable.

2.2.2 Acquisition Function

Expected Hypervolume Improvement

The Expected Hypervolume Improvement (EHI) acquisition function provides an extension of single objective expected improvement to the multi-objective domain.[147] EHI utilises the hypervolume indicator metric coupled with the probability density function to provide the EHI for a given point;

$$\int_{y\in\mathbb{R}^m} I_H(y,P)\cdot PDF(y)dy \quad \textbf{(2-5)}$$

where I_H is the hypervolume improvement of point y with respect to the current Pareto front P.

The algorithm provides good convergence to the true Pareto front, however, requires a high computational budget for computing the infill criteria. Efficiency improvements have been made for the calculation of the metric, however, they are limited to the two and three dimensional cases.[148,149] Alternative acquisition functions and approaches have been suggested,

focusing on reducing the computational burden for acquisition evaluation, as well as improving the efficiency of the optimisations towards an optimal front.

Thompson sampling

Introduced by Thompson, Thompson sampling achieves the trade-off between exploration and exploitation through randomness.[150] Actions are selected according to ones that have the highest probability of leading to a reward.[151] In the context of Bayesian optimisation, this requires samples to be drawn from the posterior, with the arguments for the minimum/maximum of this sample taken as the next point for evaluation of the actual function.

Discrete sampling or approximation methods can be used for the minimisation/maximisation of the drawn sample, with approximate methods offering improved computational efficiency.[152] The acquisition function has shown excellent performance on continuous single objective problems, displaying performance on par or exceeding commonly utilised acquisition functions such as probability of improvement or expected improvement.[153]

Recently, the TSEMO algorithm employing Thompson sampling with an internal NSGA-II optimisation has been proposed.[154] The authors utilise the inherent exploratory nature of Thompson sampling to build Gaussian surrogates, which are subsequently optimised utilising NSGA-II to suggest points that maximise the hypervolume improvement. Approximate analytical functions are drawn from the Gaussian surrogate posterior utilising a process called spectral sampling.[152] Utilising approximate analytical samples allows for inherent randomness to enable wider exploration of the search domain during internal optimisation via the NSGA-II algorithm.

Expected Improvement Matrix

Expected improvement matrix (EIM) is a multi-objective adaptation of expected improvement, with the expected improvement for each candidate and each objective calculated in a matrix.

$$EI_i^j(\mathbf{x}) = \left(f_i^j - \hat{y}_i(\mathbf{x})\right)\Phi(Z(\mathbf{x})) + \sigma_i(\mathbf{x})\phi(\mathrm{Z}(\mathbf{x})) \qquad \textbf{(2-6)}$$

$$Z(\mathbf{x}) = \frac{f_i^j - \hat{y}_i(\mathbf{x})}{\sigma_{\mathrm{i}}(\mathbf{x})} \qquad \textbf{(2-7)}$$

$$\begin{bmatrix} EI_1^1(\mathbf{x}) & \cdots & EI_m^1(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ EI_1^k(\mathbf{x}) & \cdots & EI_m^k(\mathbf{x}) \end{bmatrix} \qquad \textbf{(2-8)}$$

f_i^j are the current best values along the Pareto front, $\hat{y}_i(\boldsymbol{x})$ is the predicted mean of the model, with $\sigma_i(\mathbf{x})$ the standard deviation of the

prediction. $\Phi(\mathbf{x})$ and $\phi(\mathbf{x})$ are the Gaussian cumulative distribution and probability density functions, respectively. Figure 2-1 provides a graphical overview of the EIM calculation for a candidate point.

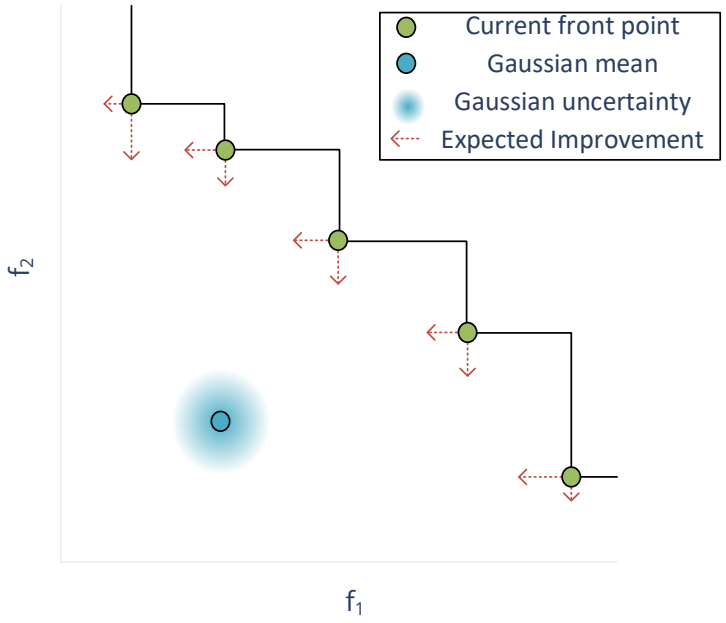


Figure 2-1. Graphical visualisation of a two-dimension EIM calculation for a candidate point. EI is calculated for the candidate point (blue), with respect to all non-dominated points and all objectives, resulting in a matrix of expected improvement values.

This matrix representation of expected improvement resembles the current best solution in multi-objective optimisation, which is also a matrix.

$$\begin{bmatrix} f_1^1(\mathbf{x}) & \cdots & f_m^1(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ f_1^k(\mathbf{x}) & \cdots & f_m^k(\mathbf{x}) \end{bmatrix} \quad \textbf{(2-9)}$$

The matrix representation of expected improvement is then combined into a single value using three possible transformations: Euclidean, hypervolume and min/max distance. All three transformations are detailed as follows.

$$EIM_e(\mathbf{x}) = \min_{j=1,\ldots,k} \sqrt{\sum_{i=1}^{m} \left(EI_i^j(\mathbf{x})\right)^2} \quad \textbf{(2-10)}$$

$$EIM_h(\mathbf{x}) = \min_{j=1,\ldots,k} \left[\prod_{i=1}^{m} \left(r_i + EI_i^j(\mathbf{x}) - f_i^j\right) - \prod_{i=1}^{m} \left(r_i - f_i^j\right) \right] \quad \textbf{(2-11)}$$

$$EIM_m(\mathbf{x}) = \min_{j=1,\ldots,k} [\max_{i=1,\ldots,m} EI_i^j(\mathbf{x})] \quad \textbf{(2-12)}$$

For the test problem detailed we elected to use the Euclidean based transformation as it displayed the best performance across a range of problems in the original paper.[155] As the EI equation is in its closed form it can be rapidly calculated and optimised utilising evolutionary algorithms or multi-start local search methods. This should allow for a global solution for the optimisation of EIM to be found, for little computational cost. EIM has previously been implemented in the form of the EIM-EGO algorithm which was evaluated on a series of test problems.[155]

Comparison

Given the TSEMO algorithm has displayed improved performance over alternative multi-objective methodologies,[154] and has been successfully applied to previous chemical optimisation systems[45,64] it was taken forward for comparison with the EIM based acquisition function.

Initial comparison between Thompson sampling and EIM-based acquisition functions was performed utilising three continuous test problems: (i) vlmop2; (ii) DTLZ5; (iii) DTLZ7.[156,157] The test problems selected were of varying difficulty and nature to highlight the relative strengths and weaknesses of the respective algorithms. Across all test problems the algorithms were only afforded 20 iterations once initialised. This was to somewhat replicate the constrained experimental budget that can be found during real world expensive-to-evaluate applications of these multi-objective algorithms. Figure 2-2 provides comparison between the two acquisition functions across all test problems considering both hypervolume indicator and total CPU time for the optimisation. The box plots highlight the mean and distribution of the overall experimental runs, highlighting the best performing acquisition function in terms of mean value, as well as repeatability, for the three optimisation problems. Hypervolume was selected given its well-known ability to consider both accuracy and diversity of the front.[111]

An example output from the DTLZ5 optimisation problem is shown in Figure 2-3 to provide qualitative comparison between the two acquisition functions. For both algorithms, the Matérn$_{3/2}$ kernel was utilised in the underlying surrogate model, this was to ensure observed differences could be correlated to differences in acquisition function rather than alternative surrogate modelling methodologies.

Figure 2-2. Performance metric comparison between EIM based acquisition function and TSEMO for three test problems. Algorithms were limited to 20 function evaluations with an initial data set of 2n, where n is the number of input variables. Violin plots show end values for 20 repeats for each algorithm and test problem.

Across all compared test problems, the EIM based algorithm significantly outperforms the Thompson sampling based approach both in terms of hypervolume indicator and total CPU time to complete the optimisation. The significant differences in hypervolume, even given such heavily constrained budgets, indicate considerable improvements in convergence towards an optimal front for the EIM based acquisition function. This was hypothesised to be attributed to the less random nature of the EIM based acquisition, with the Thompson sampling based approach inherently more exploratory in the early stages of the optimisation. The stark contrast in CPU time can be attributed to the high computational burden associated with the spectral sampling, requiring a Monte-Carlo based approach for approximation, and internal NSGA-II optimisation process TSEMO utilises.

Additional comparisons for simulated chemical system optimisations has been performed by Clayton.[158] Both the TSEMO algorithm and an implementation of EIM were compared across six simulated case studies of varying difficulty and dimensionality. For all bi-objective cases, utilising EIM with a Euclidean combination function outperformed the TSEMO algorithm, as well as ParEGO,[159] in terms of the hypervolume indicator performance metric. TSEMO was only able to offer improved performance for a tri-objective example, with marginal improvement with respect to final hypervolume across 10 repeats. Similarly, to the cases examined here, improved initial performance of the EIM based approach was noted in the work, with the algorithm displaying an improved rate of change in hypervolume across all problems examined.

Given the targeted application, to budget constrained experimentation, the improved initial performance afforded by the EIM based methodology makes it an ideal candidate for use.

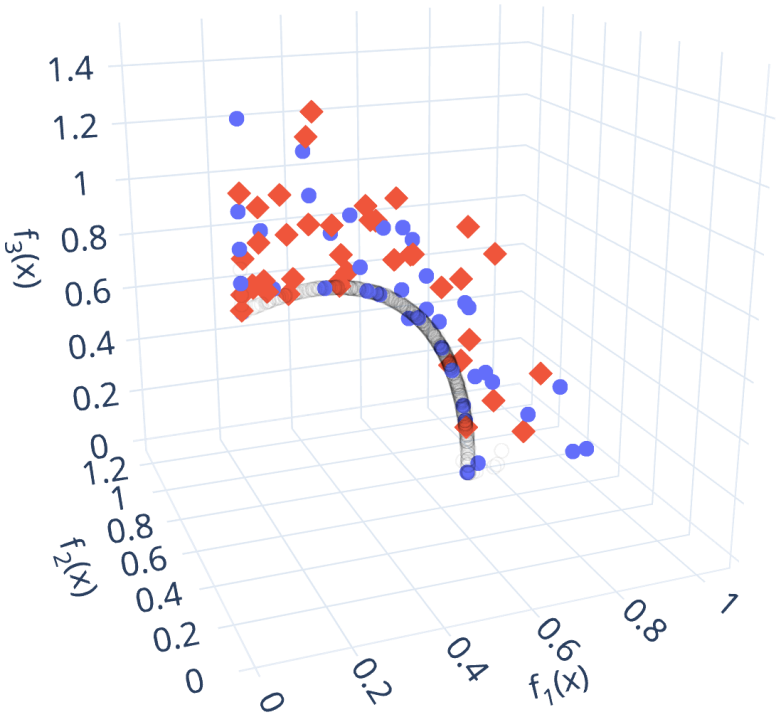


Figure 2-3. Example objective domain output for the DTLZ5 test problem. EIM based acquisition shown in blue, TSEMO shown in red and Pareto front for the test problem shown in black.

2.2.3 Algorithm Overview

The final multi-objective algorithm utilises a Gaussian process surrogate with the Matérn$_{5/2}$ kernel use in conjunction with the Gower distance metric. Expected improvement matrix was selected as the multi-objective acquisition function. Figure 2-4 provides an overview of the optimisation process for the MVMOO algorithm. For the initial space filling design, in all cases, a five sample Latin hypercube for each combination of the discrete variables was performed. It is noted that this can lead to a large initial cost to the optimisations when there are multiple discrete variables or discrete variable levels in the optimisation problem.

Upon completion of the initial data collection, a GP model was constructed for each objective function and hyperparameters optimised using GPflow's internal Adam optimiser. The trained GP models are then used with the EIM acquisition function to determine the next point for evaluation. The optimisation of the acquisition function occurs in two stages; initially a large sample of the acquisition function is taken using a Halton sequence for each discrete variable combination. The leading variable combination was then taken forward with additional local optimisation performed using SciPy's[160] implementation of sequential least squares programming (SLSQP).[161]

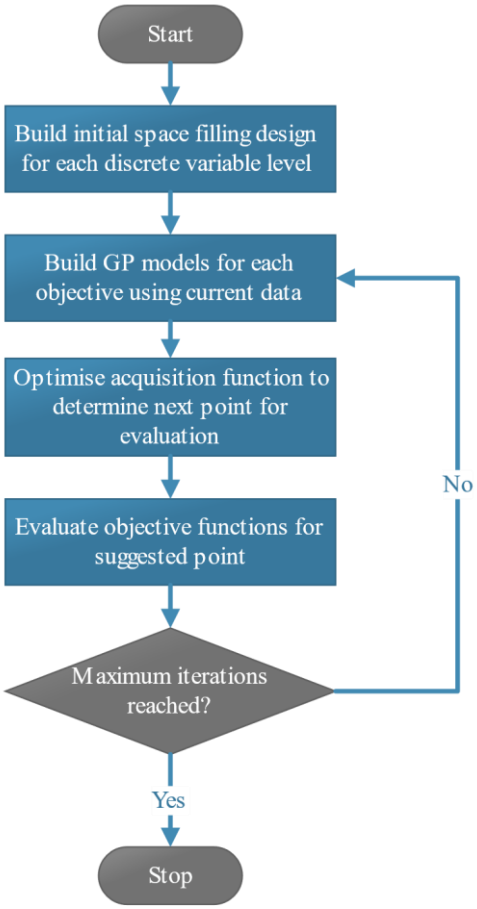


Figure 2-4. Flowchart overview of the iterative cycle of processes for the MVMOO algorithm.

2.3 *In silico* Applications

To test the newly proposed mixed variable multi-objective optimisation algorithm, the approach was applied to three test problems. Comparison between existing mixed variable multi-objective optimisation methods and selected continuous multi-objective optimisation algorithms was performed for benchmarking purposes.

For the first two test problems, the true Pareto set was known, with details provided. As the true front for the final problem was unknown, the Pareto front was determined through extensive evaluation of the underlying objective functions utilising a mixed variable version of the NSGA-II algorithm.[130]

Discrete VLMOP2

VLMOP2 is a well-known bi-objective test problem, which was originally proposed by van Veldhuizen and Lamont as a benchmarking problem for

continuous multi-objective optimisation algorithms.[162] Herein, the original problem has been adapted to create a mixed variable equivalent:

$$\min\left(f_1(\mathbf{x}, d), f_2(\mathbf{x}, d)\right) \quad \textbf{(2-13)}$$

$$x_i \in [-2,2]\ d \in [a, b] \quad \textbf{(2-14)}$$

$$f_1(\mathbf{x}, d) = \begin{cases} 1 - \exp\left(-\sum_{i=1,\dots,n}\left(x_i - \frac{1}{\sqrt{n}}\right)^2\right) & if\ d = a \\ 1.25 - \exp\left(-\sum_{i=1,\dots,n}\left(x_i - \frac{1}{\sqrt{n}}\right)^2\right) & if\ d = b \end{cases} \quad \textbf{(2-15)}$$

$$f_2(\mathbf{x}, d) = \begin{cases} 1 - \exp\left(-\sum_{i=1,\dots,n}\left(x_i + \frac{1}{\sqrt{n}}\right)^2\right) & if\ d = a \\ 0.75 - \exp\left(-\sum_{i=1,\dots,n}\left(x_i + \frac{1}{\sqrt{n}}\right)^2\right) & if\ d = b \end{cases} \quad \textbf{(2-16)}$$

As the modifications are limited to a constant shift in both the first and second objective, the newly proposed discrete VLMOP2 shares similar properties to the original test function. In analogous fashion to the original function, the Pareto front is concave with the Pareto optimal set following a line from $x_i = -\frac{1}{\sqrt{n}}$ to $x_i = \frac{1}{\sqrt{n}}$. The Pareto front and Pareto optimal set for the discrete VLMOP2 are detailed in Figure 2-5.

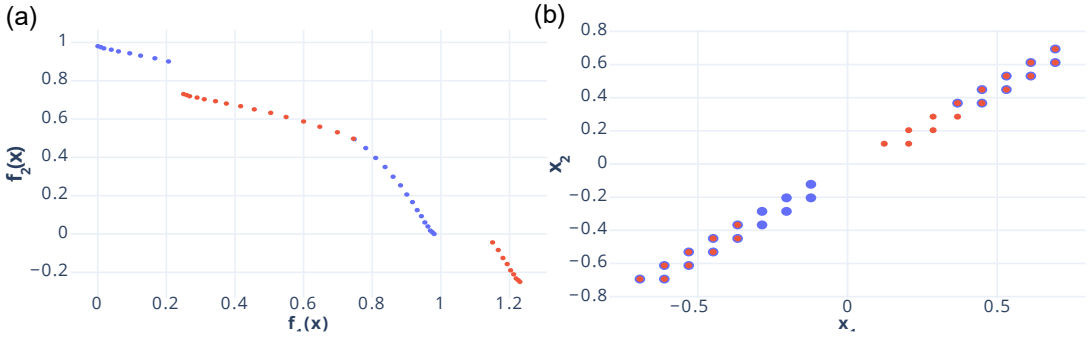


Figure 2-5. Pareto front (a) and Pareto optimal set (b) for the VLMOP2 test problem. Points where d = a are shown in blue and points where d=b are shown in red. Interactive plots can be found at the following links: (a) https://chart-studio.plotly.com/~jmanson377/188, (b) https://chart-studio.plotly.com/~jmanson377/191

As the Pareto front is composed of a mixture of the discrete variable levels, the test problem provides a useful initial test case for the algorithm, verifying whether the underlying surrogate models accurately account for the effects of changing discrete variable levels.

Ordinary Differential Equation (ODE) Catalytic System

An ordinary differential equation involves a function and its derivatives with respect to one independent variable. They are often used in mechanistic modelling to describe the underlying system behaviour with respect to time or space. In chemical reactions, we find ODEs in the form of rate equations. These rate equations describe the rate of reaction for the individual chemical components of the system. Systems of ODEs can be solved analytically or by using ODE solvers to determine how the reaction progresses with respect to time or space. In general, Runge-Kutta methods are used to determine the solution of these chemical rate equations.[163]

The second test problem utilises an ODE system describing a catalytic reaction originally proposed by Baumgartner *et al*..[51] A system of ODEs was used to map the relationship between the four dimensional mixed variable input and the respective objective function. As the original problem was in a single-objective form, it has been adapted to include an additional objective, STY, with the system described in full below:

$$max(Yield, STY) \quad \textbf{(2-17)}$$

$$C_{cat} \in [0.835\ mM, 4.175\ mM]\ T \in [30\ ^oC, 110\ ^oC]\ t_{res} \in [1\ min, 10\ min]$$

$$d \in \{1,2,3,4,5,6,7,8\} \quad \textbf{(2-18)}$$

$$A + B \overset{k}{\rightarrow} P \quad \textbf{(2-19)}$$

$$k = C_{cat}^{\frac{1}{2}} k_0 e^{-\frac{E_{A_R}+E_{A_i}}{RT}} \quad \textbf{(2-20)}$$

$$\frac{dA}{d\tau} = -k[A][B] \quad \textbf{(2-21)}$$

$$\frac{dB}{d\tau} = -k[A][B] \quad \textbf{(2-22)}$$

$$\frac{dP}{d\tau} = k[A][B] \quad \textbf{(2-23)}$$

$$Yield = \frac{[P]_{final}}{[A]_{initial}} \quad \textbf{(2-24)}$$

$$STY = \frac{100 \times [P]_{final}}{t_{res}} \quad \textbf{(2-25)}$$

Where τ is the residence time of the reactor, k_0 is the Arrhenius constant for the reaction ($3.1 \times 10^7 L^{\frac{1}{2}} mol^{-\frac{3}{2}} s^{-1}$), [A] and [B] are the concentrations of A and B in the reactor (Initial values: $[A]_0 = 0.167\ M$ and $[B]_0 = 0.250\ M$) and E_{A_R} is the reaction activation energy ($55\ kJ\ mol^{-1}$). The system of ODEs was solved through application of a pre-existing Runge-Kutta based solver, resulting in the value of $[P]_{final}$ which was used to determine the objective function values. Figure 2-6 details the Pareto front and Pareto set for the ODE catalytic system.

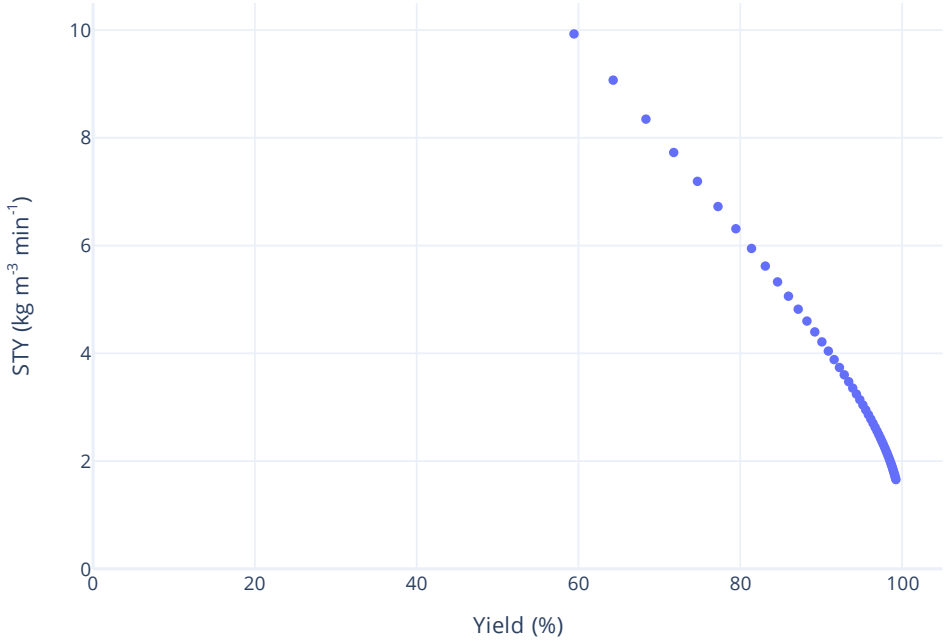


Figure 2-6. Pareto front for the catalytic ODE example. Pareto set at the following conditions: $\boldsymbol{C_{cat} = 4.175\ mM, T = 80\ ^oC, t_{res} \in [1\ min, 10\ min]}$ and $\boldsymbol{d = 1}$. Interactive plot can be found at the following link: https://plotly.com/~jmanson377/193/

As the Pareto front is comprised of a single discrete variable level, this allows for comparison between continuous variable multi-objective algorithms. This will provide insight into the relative efficiency of the newly proposed algorithm, as well as providing a benchmark value that can be achieved through application of a state-of-the-art efficient continuous variable multi-objective algorithm. The recently published TSEMO algorithm[154] was selected for comparative purposes, with it representing an alternative technique for experiment selection, while being at the forefront of the field in its performance.

Fuel Injector Design

This problem was first published by Burachik *et al*. in which they utilised response surface modelling to develop empirical models for four objectives relating to the design of a fuel injector.[164] Originally a continuous variable system, the authors adapted the problem to impose an integer constraint on one of the design variables.

$$min\big(f_1(\mathbf{x}), f_2(\mathbf{x}), f_3(\mathbf{x}), f_4(\mathbf{x})\big) \quad \textbf{(2-26)}$$

$$x_{2-4} \in [-2,2]\ \bar{x}_1 \in \{0,1,2,3\}$$
$$x_1 = 0.2\bar{x}_1 \quad \textbf{(2-27)}$$

$$\begin{aligned} f_1(\mathbf{x}) = {} & 0.692 + 0.4771x_1 - 0.687x_4 - 0.08x_3 - 0.065x_2 \\ & - 0.167x_1^2 - 0.0129x_1x_4 + 0.0796x_4^2 - 0.0634x_1x_3 \\ & - 0.0257x_3x_4 + 0.0877x_3^2 - 0.0521x_1x_2 \\ & + 0.00156x_2x_4 + 0.00198x_2x_3 + 0.0184x_2^2 \end{aligned} \quad \textbf{(2-28)}$$

$$\begin{aligned} f_2(\mathbf{x}) = {} & 0.37 - 0.205x_1 + 0.0307x_4 + 0.108x_3 + 1.019x_2 \\ & - 0.135x_1^2 + 0.0141x_1x_4 + 0.0998x_4^2 + 0.208x_1x_3 \\ & - 0.0301x_3x_4 - 0.226x_3^2 + 0.353x_1x_2 \\ & - 0.0497x_2x_3 - 0.423x_2^2 + 0.202x_1^2x_4 - 0.281x_1^2x_3 \\ & - 0.342x_1x_4^2 - 0.245x_3x_4^2 + 0.281x_3^2x_4 \\ & - 0.184x_1x_2^2 + 0.281x_1x_3x_4 \end{aligned} \quad \textbf{(2-29)}$$

$$\begin{aligned} f_3(\mathbf{x}) = {} & 0.153 - 0.322x_1 + 0.396x_4 + 0.424x_3 + 0.0226x_2 \\ & + 0.175x_1^2 + 0.0185x_1x_4 - 0.0701x_4^2 - 0.251x_1x_3 \\ & + 0.179x_3x_4 + 0.015x_3^2 + 0.0134x_1x_2 \\ & + 0.0296x_2x_4 + 0.0752x_2x_3 + 0.0192x_2^2 \end{aligned} \quad \textbf{(2-30)}$$

$$\begin{aligned} f_4(\mathbf{x}) = {} & 0.758 + 0.358x_1 - 0.807x_4 + 0.0925x_3 - 0.0468x_2 \\ & - 0.172x_1^2 + 0.0106x_1x_4 + 0.0697x_4^2 - 0.146x_1x_3 \\ & - 0.0416x_3x_4 + 0.102x_3^2 - 0.0694x_1x_2 \\ & - 0.00503x_2x_4 + 0.0151x_2x_3 + 0.0173x_2^2 \end{aligned} \quad \textbf{(2-31)}$$

A slice of the four-dimensional Pareto front is highlighted in Figure 2-7. As with the discrete VLMOP2 example, the Pareto front is composed of a combination of discrete variable levels. This, coupled with the increased objective space should provide a challenging performance test for the algorithm.

Due to issues relating to numerical instability when performing matrix inversion, min-max normalisation of the input domain, alongside z-score

standardisation of the objective domain was implemented. This was done to improve performance and stability of the algorithm. Testing across all problems indicated a significant improvement when normalisation was performed. This was linked to the varying magnitude of the continuous inputs, prior to normalisation, leading to the covariance matrix moving towards regions of numerical instability. Given this, for all test problems normalisation and standardisation was performed as standard when computing the next point for evaluation.

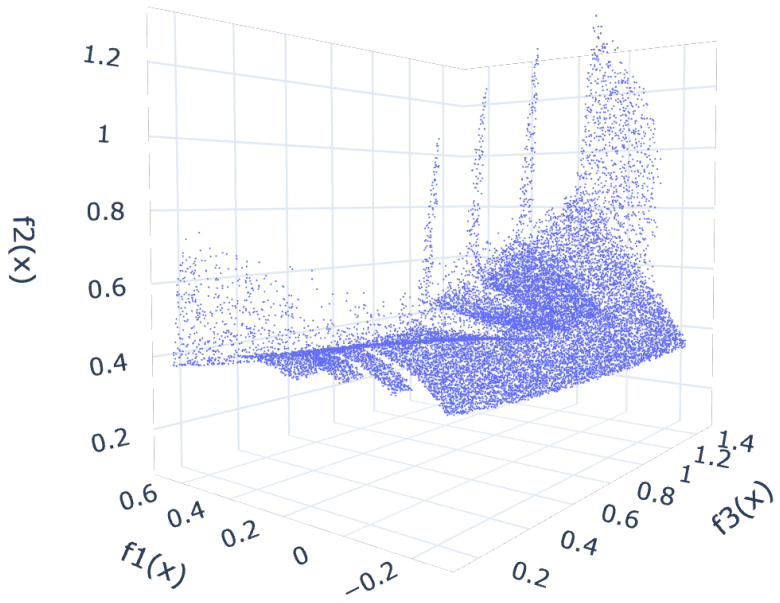


Figure 2-7. Fuel Injector Pareto front displayed for the first three objective functions. Interactive plot can be found at the following link: https://chart-studio.plotly.com/~jmanson377/195

2.3.1 Results and Discussion

Performance characterisation was achieved through use of three indicators: hypervolume, modified inverted generational distance and worst attainment surface. All three metrics highlight key areas of performance which will provide insight into the effectiveness of the proposed method.

Details on the initial dataset size, algorithm budgets and settings for the mixed variable NSGA-II are provided in Table 4 and Table 5. Termination criteria for NSGA-II was set to be once the predefined number of generations had been evaluated. The additional budget afforded to NSGA-II across all problems was to provide comparison to a reasonable budget provided to a readily available evolutionary approach. This allowed for efficiency comparisons to be made and enables the relative strengths and weaknesses to be drawn from alternative methods. Full results for all optimisation problems can be found in section 6.2.1.

Table 4. Algorithm budgets for test problems.

Parameter	Discrete VLMOP2	Ordinary Differential Equation Catalytic System	Fuel Injector Design
MVMOO Initial Dataset Size	10	40	20
MVMOO, LHC and Random Function Evaluation Budget	40	125	100
NSGA-II Generations	20	85	80
NSGA-II Population Size	40	125	100

Table 5. NSGA-II settings, where $\boldsymbol{n}$ is the problem input dimension. Recommended settings from the original paper were utilised for all test problems.[165]

Setting	Operator	Probability	Distribution Index
Selection Function	Binary Tournament		
Recombination Function	SBX Crossover Integer SBX Crossover	0.9	20
Mutation Function	Polynomial Mutation Integer Polynomial Mutation	$\frac{1}{n}$	20

Discrete VLMOP2

Firstly, we will consider the results of the optimisation of the discrete VLMOP2 test problem. Summary plots for all performance metrics, as well as iterative summary plots, are shown in Figure 2-8. A reference point of $R = [1.25,1.0]$ was used for the hypervolume indicator.

Overall, the results indicate that the novel MVMOO algorithm provides leading performance for both hypervolume and IGD+, when compared to the alternative approaches. Across both metrics MVMOO offers improved performance in terms of optimal values achieved and the reproducibility of the results, indicated by the narrow spread of the final box plot values. Iterative results for all algorithms, Figure 2-8 (a-b), suggest that premature termination has occurred with limited plateauing of the performance indicators, which given the restricted evaluation budget is an expected behaviour of the problem. Comparing the MVMOO algorithm to both random algorithms, we can clearly see the benefit the Bayesian approach affords, significantly increasing the rate of improvement for both indicator-based metrics. Comparison with the NSGA-II algorithm shows a 21-fold and 36-fold improvement in efficiency for hypervolume and IGD+ respectively, reinforcing

the benefits of adopting the proposed novel approach to mixed variable multi-objective problems.

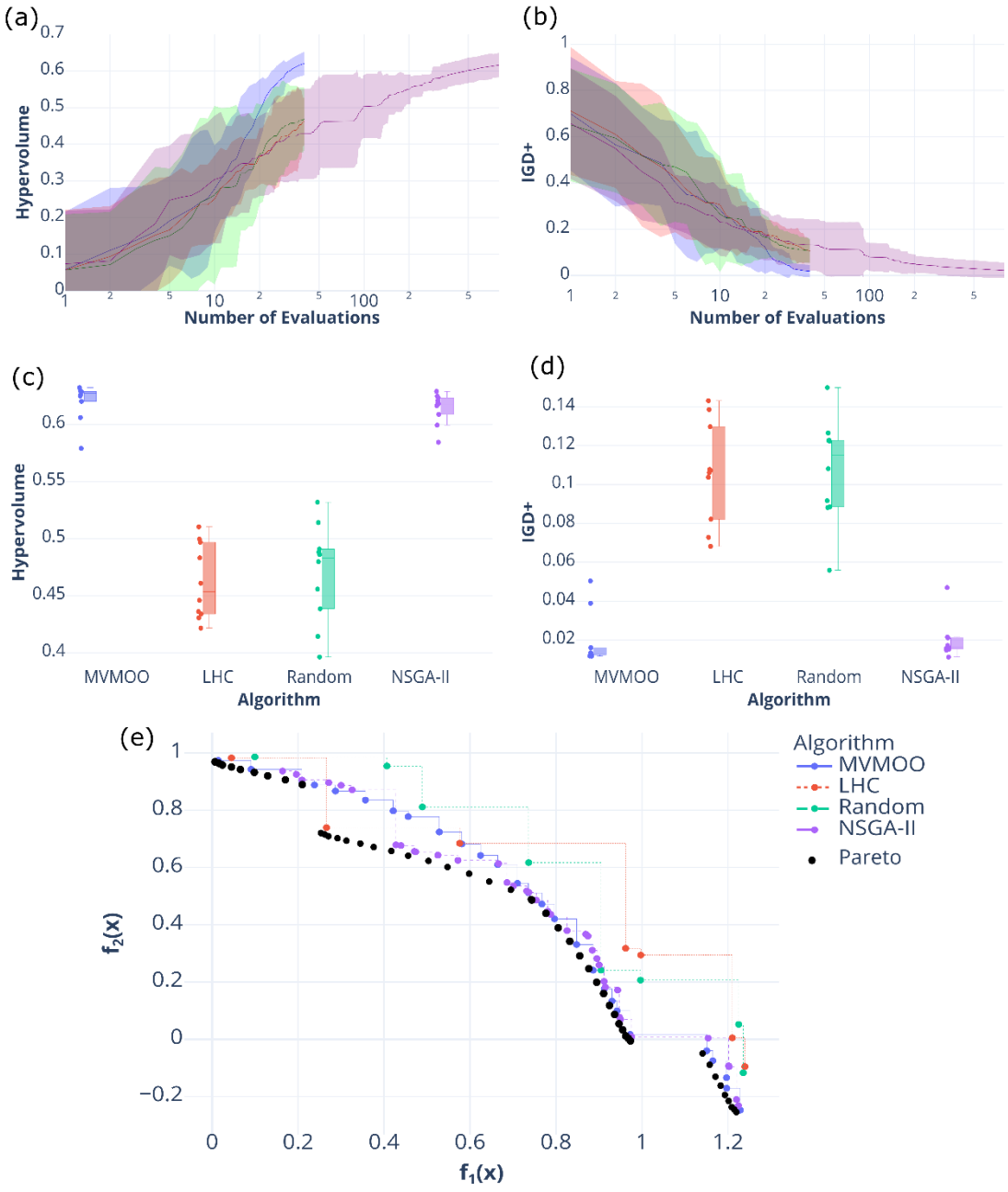


Figure 2-8. Optimisation results for the discrete VLMOP2 test problem for 10 runs. Algorithms are shown as follows: — MVMOO, — LHC, — Random and — NSGA-II. For (a) and (b) 95% confidence interval (CI) for each algorithm indicated by the shaded areas. For (a-b) NSGA-II was afforded a total of 800 function evaluations per run (20 generations with a population size of 40). For (e) the ground truth Pareto front is plotted in black for reference. Interactive plots can be found at the following links:(a) https://plotly.com/~jmanson377/207/,(b) https://plotly.com/~jmanson377/218/,(c) https://plotly.com/~jmanson377/139/,(d) https://plotly.com/~jmanson377/135/ and (e) https://plotly.com/~jmanson377/137/

Generally, this test problem highlights the underlying GP surrogate model(s) ability to effectively model this mixed variable system. The surrogate model(s) can successfully map the effect of switching discrete variable across the Pareto front, where sufficiently information rich training data has been supplied. The worst attainment plot, Figure 2-8 (e), highlights a key issue with the algorithm, with this plot indicating that there are worst case instances where the algorithm was unable to switch between the two discrete variable levels, resulting in a sub-optimal Pareto front. This is likely due to the low number of initialisation points used for the problem and the deliberately restrictive maximum evaluation limit. In these instances, where the initial points have limited global information; the algorithm can become stuck in a locally optimal Pareto front and will require additional evaluations to escape from this. Maximising information gained from initial data collection is a key aspect, especially for limited budget optimisations. This highlights the key importance of selecting the correct experimental design for initial data collection.

ODE Catalytic System

Summary plots for all performance metrics, as well as iterative summary plots for the ODE catalytic system test problem are shown in Figure 2-9. For the hypervolume performance indicator a reference point of $R = [0,0]$ was used. As the optimum Pareto front is composed of only one discrete variable level, comparison with the continuous multi-objective algorithm TSEMO has also been performed.

Overall, the MVMOO algorithm offers comparative performance when compared to both the TSEMO and NSGA-II algorithms. Algorithmically led approaches offer a clear improvement when compared to both random sampling implementations. TSEMO and NSGA-II exceed MVMOO's performance in terms of final hypervolume and IGD+ values with respect to both optimum value and reproducibility, with NSGA-II offering superior performance across all metrics. Analysing both iterative plots, Figure 2-9 (a-b), we see there is a significant stall in the iterative process for both performance metrics. This observed effect is due to the initial data gathering procedure for the MVMOO algorithm, using a five sample Latin hypercube for each discrete variable level. This type of sampling may prove effective when the Pareto front consists of multiple discrete variable levels, however, in systems where this is not the case this can lead to a large portion of the evaluation budget being exhausted on unproductive conditions.

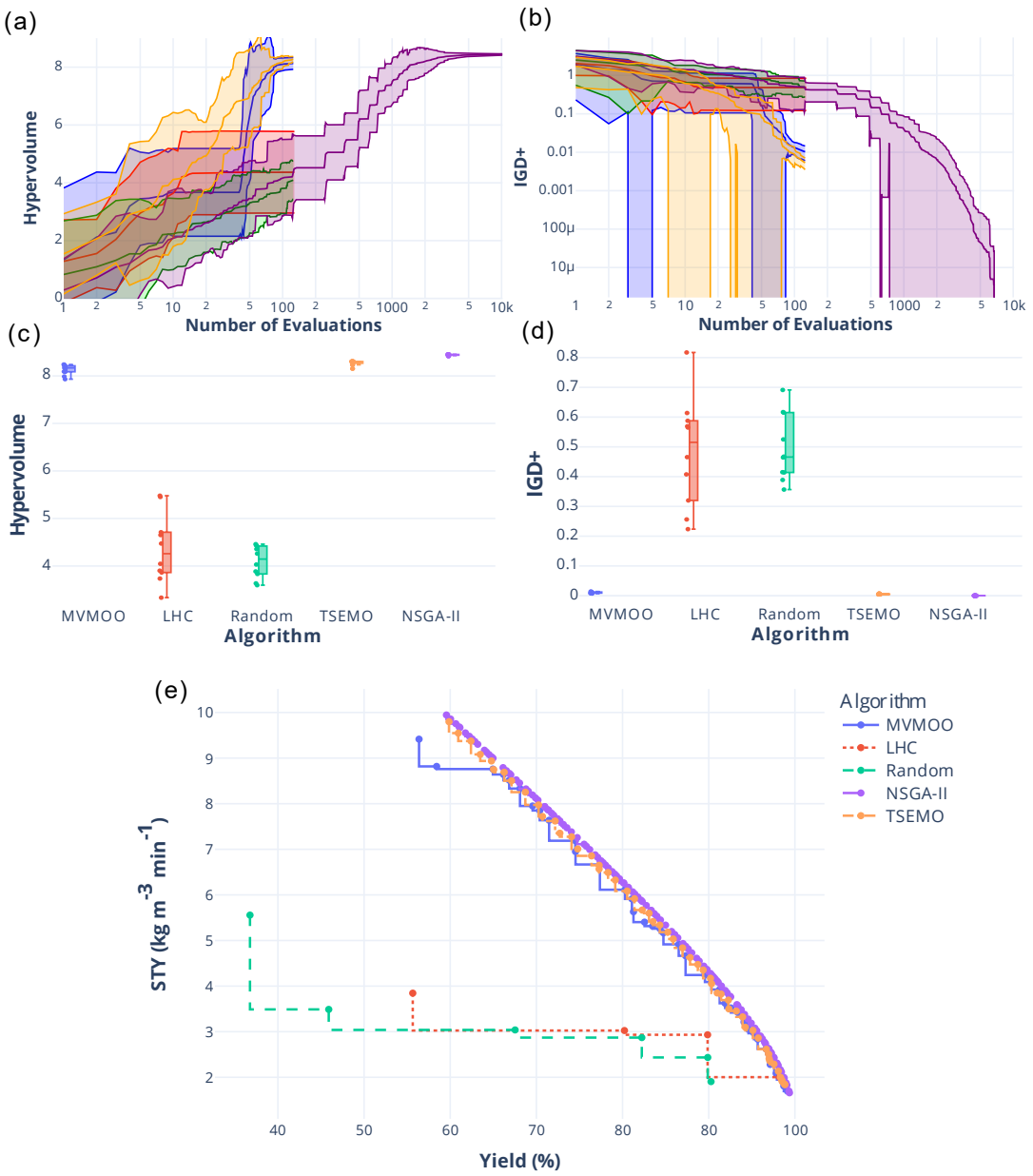


Figure 2-9. Optimisation results for the ODE catalytic system test problem for 10 runs. Algorithms are shown as follows: — MVMOO, — LHC, — Random, — TSEMO and — NSGA-II. For (a) and (b) 95% CI for each algorithm indicated by the shaded areas. NSGA-II was afforded a total of 10200 function evaluations per run (85 generations with a population size of 120). N.B. The y axis for (b) is plotted using a log scale to allow discernment between algorithms. The large jumps in CI are where the CI tends to zero. Interactive plots for all subfigures are given as follows: (a) https://plotly.com/~jmanson377/224/, (b) https://plotly.com/~jmanson377/226/,(c) https://plotly.com/~jmanson377/146/,(d) https://plotly.com/~jmanson377/148/ and (e) https://plotly.com/~jmanson377/153/.

Additionally, we see how the MVMOO algorithm significantly plateaus in terms of hypervolume. This effect is seen to a lesser extent for the IGD+

metric, which suggests that the algorithm may be focusing new evaluation points on a limited area of the objective domain. The worst attainment surface plot, Figure 2-9(e), offers insight into explanation for this observed behaviour, with the algorithm struggling to achieve the extremities of the Pareto front with respect to STY. It is also noted that both TSEMO and NSGA-II have a greater density of Pareto points which leads to improvements in relation to the final IGD+ values for these methods.

Although not achieving the same level of peak performance, the MVMOO algorithm still offers an efficient methodology for optimising this system with the rate of change in terms of both hypervolume and IGD+ exceeding that of both TSEMO and NSGA-II during the initial algorithmically led period. Comparatively to NSGA-II, the MVMOO algorithm can offer the same level of performance in approximately 7% of the iterations to that required by NSGA-II, offering significant efficiency improvements.

Fuel Injector Design

Summary plots for all performance metrics, as well as iterative summary plots for the fuel injector design test problem are shown in Figure 2-10 - Figure 2-12. For the hypervolume performance indicator a reference point of $R = [0.8,1.4,1.7,1.0]$ was used.

Overall, the results highlighted in Figure 2-10 indicate the MVMOO algorithm offers significant improvements with respect to both hypervolume and IGD+. As with the first test problem, the improvement afforded is both greater in terms of final optimum value and the reproducibility of the results. Figure 2-10 (a-b) illustrates the dramatic improvement in efficiency achieved when employing the MVMOO algorithm, with the algorithm achieving a 145-fold and 190-fold improvement in terms of iteration efficiency for hypervolume and IGD+ respectively. For both hypervolume and IGD+ minor plateauing is observed, suggesting that, should additional iterations be provided, we would see an improvement for both performance metrics. This observation is contrary to the previous example, where significant plateauing was evident suggesting the algorithm had achieved near best performance for the problem within the given budget.

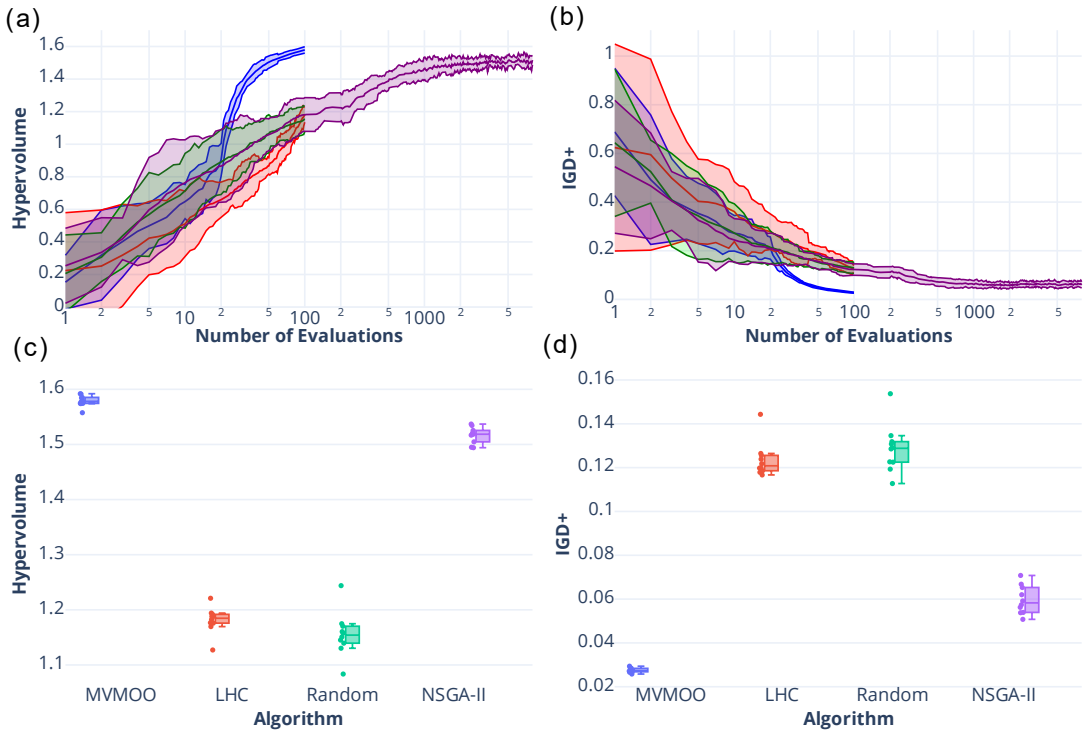


Figure 2-10. Optimisation results for the fuel injector design test problem for 10 runs. Algorithms are shown as follows: — MVMOO, — LHC, — Random and — NSGA-II. For (a) and (b) 95% CI for each algorithm indicated by the shaded areas. NSGA-II was afforded a total of 8000 function evaluations per run (80 generations with a population size of 100). Interactive plots for all subfigures are given as follows: (a) https://plotly.com/~jmanson377/234/, (b) https://plotly.com/~jmanson377/237/, (c) https://plotly.com/~jmanson377/165/, and (d) https://plotly.com/~jmanson377/167/

Figure 2-11 and Figure 2-12 provide insight into the observed differences for both performance metrics, with MVMOO providing a more uniform spread of points compared to the clustering evident with NSGA-II. The MVMOO algorithm is able to obtain a much-improved approximation of the bulk of the front, with this likely explaining the significant improvement in terms of hypervolume and IGD+ when compared to the other methods. The greater degree of plateauing evident for MVMOO in terms of IGD+ is likely linked to the algorithm struggling to achieve the extremities of the Pareto front.

Considering all problems, the newly proposed algorithm offers significant improvements in terms of efficiency for all problems, and efficacy, for two out of the three problems. The performance of the algorithm for the multi discrete level Pareto fronts highlight the algorithm's ability to proficiently switch between discrete variable levels, where necessary, and indicates that the underlying surrogate model can effectively understand and model these relationships.

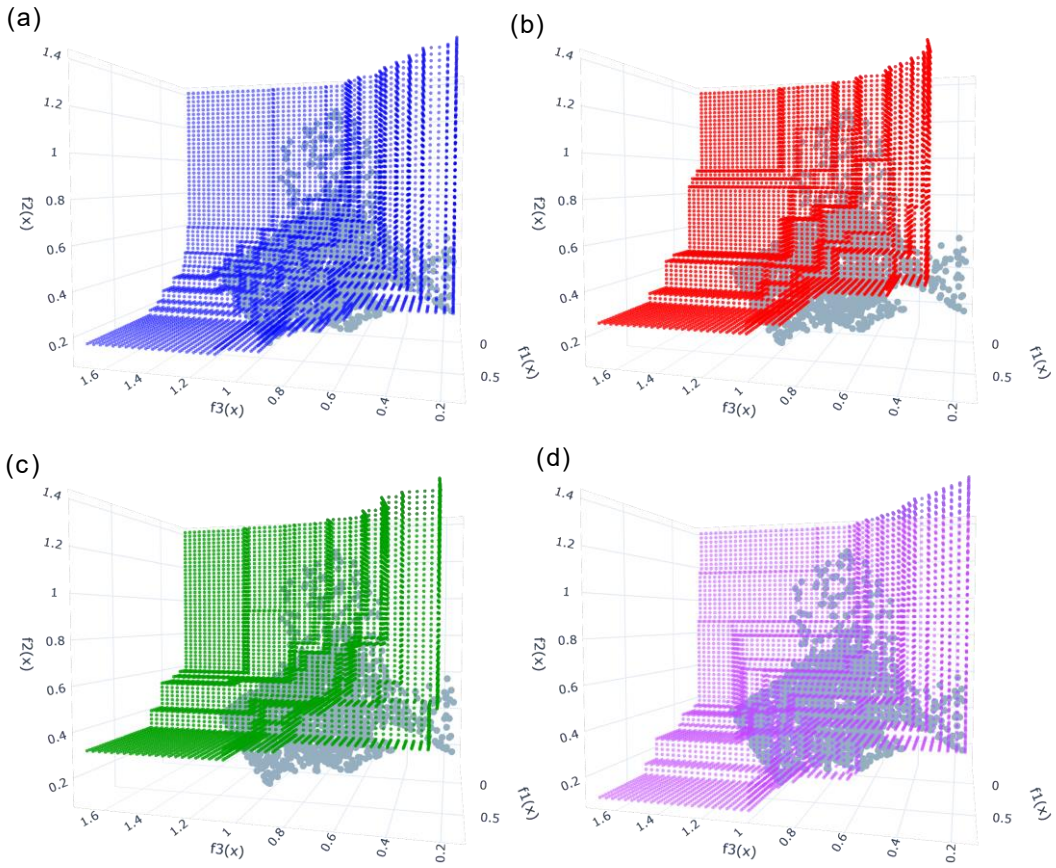


Figure 2-11. Worst attainment summary surface for the fuel injector design test problem. Reference surface shown as grey points, with the plots corresponding to the following algorithms: (a) MVMOO, (b) LHC, (c) Random and (d) NSGA-II. Link to interactive plot: https://chart-studio.plotly.com/~jmanson377/201

The narrow quartile ranges evident across all performance metrics demonstrate the robustness of the algorithm with respect to changing initial conditions. However, ensuring an efficient strategy for initial data collection plays a vital role in a successful optimisation, especially for limited budget optimisation problems. As evidenced by both the discrete VLMOP2 and the ODE catalytic system test problems, strategies that can effectively balance information gain without wasting experimental budget are key. Further work examining novel initial data collection strategies for mixed variable systems will prove beneficial, not only for optimisation tasks but also with respect to situations requiring efficient data acquisition. Adopting an iterative experimental design approach, utilising optimality criterion, to maximise the information gain of the initial data collection may provide an efficient route.[51]

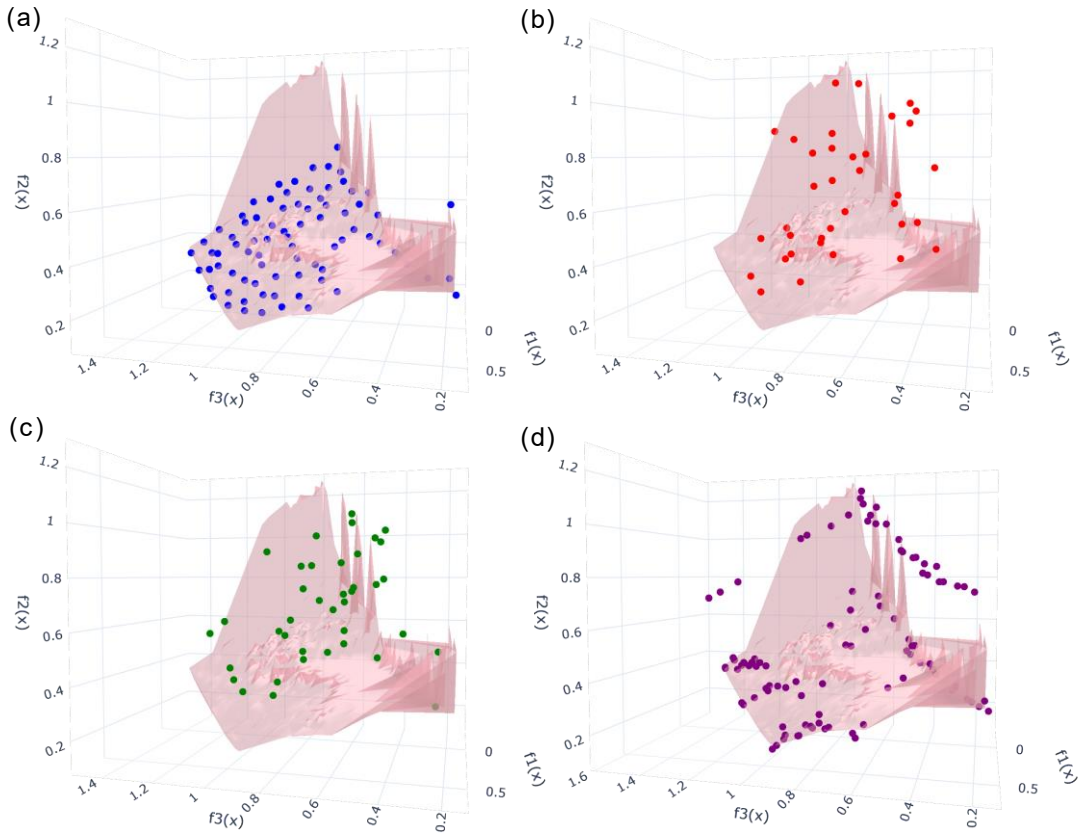


Figure 2-12. Worst attainment scatter plots for the fuel injector design test problem. Reference surface shown as a pink mesh, with the plots corresponding to the following algorithms: (a) MVMOO, (b) LHC, (c) Random and (d) NSGA-II. Link to interactive plot: https://plotly.com/~jmanson377/163/

However, issues arise in the requirement of an underlying model for use with optimality criteria, which may be difficult to produce, especially in black-box optimisation tasks. Given this, minimal efficient designs coupled with acquisition functions that allow for adequate exploration of the experimental domain are a necessity. Alternatively, utilising an approach similar to latent variable surrogate modelling, where more hyperparameters are utilised could allow for more information to be drawn from lower quality experiments. However, this does significantly increase the required number of initial experiments for an effective model fit, with current hyperparameter estimation procedures proving computationally expensive for a larger number of discrete variable levels.[143]

2.4 Conclusions

Many real-world optimisation problems are often composed of mixed variable types and consider multiple competing objectives.[137,166] When qualitative variables cannot be mapped to a continuous representation, having

algorithms that can handle such mixed variable systems and optimise them in an efficient manner becomes a necessity. In this work, a novel mixed variable multi objective optimisation algorithm was proposed and applied to three simulated optimisation problems. Comparison between NSGA-II and two random sampling techniques was performed for all problems, with the addition of continuous multi-objective algorithm comparison where applicable. Across all test problems, the need for an efficient yet informative initial sampling strategy was highlighted as a key prerequisite for a successful optimisation. Examination of optimal sampling strategies for mixed variable systems will prove beneficial to improving the outcome of expensive to evaluate optimisation problems. For two out of the three test problems, the algorithm outperformed NSGA-II across all performance metrics with a significantly reduced function evaluation budget. For the simulated catalytic reaction test problem, the algorithm compared well with NSGA-II as well as performing similarly to a continuous variable multi objective optimisation algorithm, TSEMO. Overall, the algorithm performed competitively when compared to readily available alternatives, providing a viable efficient option when optimising expensive mixed variable multi objective optimisation problems.

Chapter 3 Mixed Variable Optimisation of Chemical Systems

3.1 Introduction

Early-stage reaction development utilises high throughput reaction screening as an efficient method for exploring the reaction landscape. Such methodologies provide an excellent way to screen many different combinations of discrete variables in a resource efficient manner.[17] Although efficient when screening discrete variables, native high throughput batch experiments often overlook the interactions between discrete and continuous parameters. This is often due to rigid experimental designs suffering from the 'curse of dimensionality' when considering a large number of factors.[167] In many cases, to circumvent this, experimental designs are reduced to two levels for continuous factors, significantly reducing the resolution of the design and therefore the amount of information that can be obtained from it.[168]

A typical reaction optimisation may initially consider a broad screening, evaluating many discrete parameter combinations. This can subsequently be repeated, with discrete combinations fathomed until an optimal set of discrete parameters are determined. Upon completion, further optimisation of the continuous variables affecting the reaction can be performed, leading to the final reaction conditions. This workflow of optimising discrete then continuous variables, in a sequential manner, results in incomplete process understanding, as key interactions between the mixed variable types may not be considered. For example, the effect of temperature on the activity of different catalysts would not be considered at multiple levels.[169]

Adopting algorithmically led approaches can help circumvent these issues, with the optimisation algorithms actively guiding the process towards optimum regions of interest. There are two main approaches when optimising chemical reactions with mixed variable domains: (i) utilising chemical descriptors to convert discrete variables to the continuous domain, (ii) maintaining the discrete variable domain.

When utilising chemical descriptors, the effect of changing discrete variables is instead correlated with the associated change in chemical descriptors for the optimised discrete variables. For solvents, this could be polarity or dipole moment with the choice of descriptors paramount to the success of the algorithm. Lapkin *et al*. adopted this methodology for the multi-objective optimisation of a rhodium catalysed asymmetric hydrogenation,

where solvent selection was solely optimised.[170] The authors employed a library of 459 solvents with 17 molecular descriptors associated with each solvent. An initial set of 25 solvents were then utilised alongside various dimensionality reduction techniques to provide features for a Gaussian process model utilised by the TSEMO algorithm to determine optimally performing solvents for fixed continuous experimental conditions.

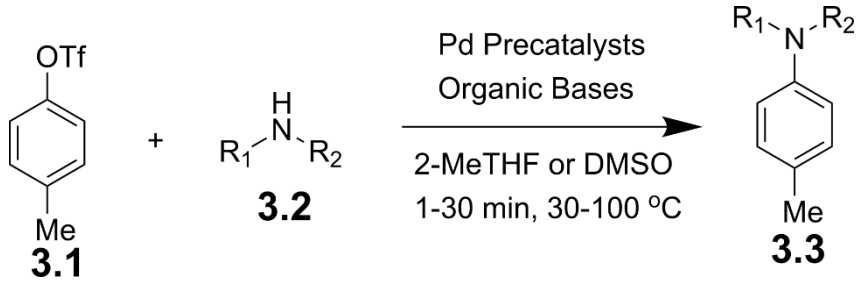
Once selected, the promising solvents underwent further optimisation focusing on continuous parameters, such as solvent ratios and the temperature of the reaction. Utilising this methodology, the authors were able to identify optimal discrete and continuous conditions, within an efficient timeframe. The optimisation was additionally benefitted by the production of an accurate statistical model that could be utilised to estimate objective function values for unexplored regions of the search space.

More recently Doyle *et al*. presented two case studies highlighting the application of Bayesian single objective methodologies for the optimisation of mixed variable systems.[18] In analogous fashion to the work by the Lapkin group, the authors utilised chemical descriptors to convert discrete parameters such as solvent, catalyst and base to the continuous domain where they could be handled utilising standard Bayesian optimisation techniques. Chemical descriptors were generated *a priori* via density functional theory (DFT) calculations, providing upwards of 1000 descriptors relating to the discrete variables to be optimised. Employing a high throughput automated batch system, the authors were able to efficiently determine optimum operating conditions for Mitsunobu and deoxofluoronation reactions, requiring only 40 and 50 experiments, respectively.

A key issue both methodologies present, is the conversion of discrete input variables to associated continuous descriptors. Understanding which descriptors to select presents a difficult challenge for surrogate modelling. The assumption that the selected molecular descriptors can effectively account for and describe observed behaviour may not always be the case. Selection of descriptors may require extensive preliminary data collection to deduce descriptors relevant to the observed system. Additionally, in systems with more than one discrete variable type, the additional continuous descriptors will significantly increase the dimensionality of the problem. Although, this can be somewhat alleviated through use of dimensionality reduction techniques, such as principal component analysis. Alternatively, as in the Doyle example, applying a prior to the surrogate model you are employing can allow for a user to utilise a large descriptor domain, without leading to overfitting of the

underlying surrogate model. Although with such large input domains surrogate model fitting can present a challenge and may require usage of specifically designed high dimensional Bayesian optimisation approaches.[171]

Given the potential complexities associated with descriptor selection, there remains a place for optimisations in which discrete variables are kept in the optimisation domain. To this end, the Jensen group have investigated a series of single objective mixed variable chemical reaction optimisations.[33,39,51,52,56] The proposed methodology assumes first order kinetics to obtain correct scaling of the continuous variables in the underlying model. The surrogate model is then employed, alongside a mixed integer non-linear optimisation algorithm, to optimise the continuous and discrete variables for the system. The authors have applied the algorithm to optimisations containing multiple discrete variables, recently optimising both base and palladium precatalyst for a C-N coupling reaction, Scheme 3-1; employing a slug oscillatory reactor with a liquid handler for slug generation.[56]



Scheme 3-1. Pd-catalysed reaction explored in a single objective mixed variable optimisation.[56]

Across four substrate examples the algorithm was successfully able to determine the process optimum, fathoming poorly performing precatalysts and bases as the optimisation progressed. Although effective, the methodology does require expert domain knowledge to operate, with kinetic model definition a necessity for accurate surrogate modelling. Given this, application to complex or unknown chemistries, without prior data collection, may not be the best use case for the approach.

In many cases system optimisation may not be limited to one performance metric or objective, with there being cases where simultaneous consideration is a requirement. Often these objective functions are conflicting with the solution of such problems being a set of optimal points. In such instances, multi-objective algorithms can provide an efficient approach to investigating such problems. In expanding multi-objective algorithms to the mixed variable chemical domain, it is envisaged that efficient improvements

in process development can be afforded, streamlining the development process, Figure 3-1.

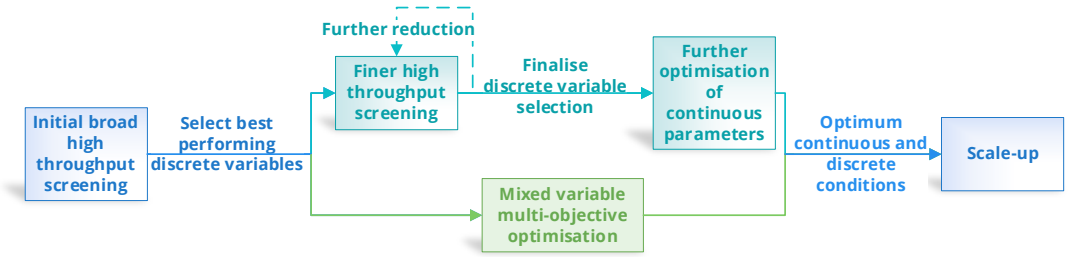


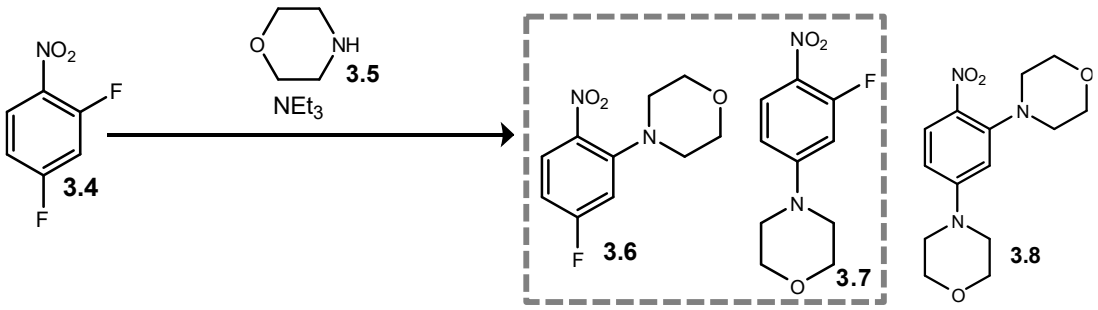
Figure 3-1. Simplified comparison of the process of a full reaction optimisation using current HTE methodologies versus the newly reported optimisation technique.

The objective functions to be optimised can be tailored to promote wider reaction understanding, such as setting the algorithm to explore trade-offs between competing reaction products. In such instances the algorithm provides a framework for efficient process investigation, which cannot be achieved to such a large extent, utilising conventional designed experiment procedures, especially in high dimensional cases.[172]

Therefore, in this work the application of a mixed variable multi-objective optimisation algorithm is described for two example reactions: (a) a S_NAr reaction where solvent is optimised alongside continuous variables; (b) a Sonogashira reaction where palladium ligand complex is optimised alongside continuous variables.

3.2 S_NAr Mixed Variable Optimisation

The S_NAr reaction between 2,4-difluoronitrobenzene (**3.4**) and morpholine (**3.5**) was selected as an initial case study, Scheme 3-2.[45]



Scheme 3-2. S_NAr reaction between 2,4-difluoronitrobenzene (**3.4**) and morpholine (**3.5**).

As the reaction is known to be effected by changing solvent, with solvent affecting the rate of reactions with respect to both *ortho* and *para* product, the trade-off between *ortho* (**3.6**) and *para* (**3.7**) product was selected as the target for the optimisation.[173,174] The reaction was hypothesised to have a diverse Pareto front, in terms of changing solvent along the front, and should provide a challenging test problem for the MVMOO algorithm; given the addition of experimental noise and the potential for overreaction to the *bis* product. As such, the optimisation was formulated as follows:

$$
\begin{aligned}
&\min[-ortho, -para] \\
&s.t. \\
&Solvent \in \{DMF, NMP, EtOH, DMAc, MeCN\} \\
&t_{res}(mins) \in [0.5, 2.0] \\
\mathbf{3.4}\ &Conc.(M) \in [0.05, 0.175] \\
\mathbf{2}\ &Eq. \in [1, 5] \\
&T(°C) \in [60, 120]
\end{aligned}
\qquad \textbf{(3-1)}
$$

Given the MVMOO algorithm is a minimisation algorithm, the negative of the *ortho* and *para* yields were used as the objective functions, resulting in their maximisation. The bounds for concentration of 2,4-difluoronitrobenzene **(3.4**) were adjusted significantly from the original paper for the final optimisation run (0.1 M to 0.5 M originally proposed). This was primarily due to solid formation occurring for reactions at elevated temperatures and equivalents in N-Methyl-2-pyrrolidone (NMP), noted during preliminary runs. The solid formed was hypothesised to be the triethylamine salt formed as a side product of the reaction.[175] Decreasing concentrations to the levels noted in (3-1) led to the elimination of equipment failure inducing salt formation. However, as a precaution a mixture of DMSO and water (5%v/v water in DMSO) was utilised as a flush between experiments to remove any solid that had accumulated. The addition of water was to prevent the reservoir solution freezing when left overnight in the lab. A schematic of the reaction setup is provided in Figure 3-2, this was utilised for both described examples.

Initialisation of the algorithm was performed using a five sample Latin hypercube for each solvent. This resulted in 25 initial experiments, with the algorithm afforded 75 subsequent experiments to determine the optimal Pareto set for the system. Results of the optimisation are summarised in Figure 3-3 and Figure 3-4. Full results for the optimisation can be found in section 6.3.3.

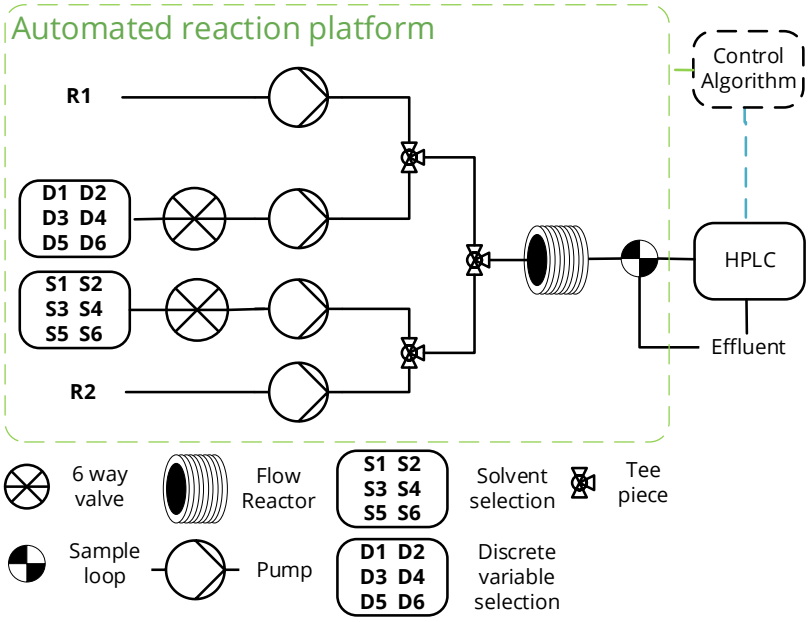


Figure 3-2. Overview of reaction set-up for S_NAr and Sonogashira case studies. Stock solutions of discrete variable options are stored in up to six different reservoirs with each reservoir connected to a port of the six-way valve. For the S_NAr study the following solvents were utilised: S1 – dimethylformamide (DMF), S2 – N-Methyl-2-pyrrolidone (NMP), S3 – Ethanol (EtOH), S4 – Dimethylacetamide (DMAc) and S5 – Acetonitrile (MeCN). S6 was utilised as the solvent flush channel. D1-D5 corresponded to (**3.4**) in the solvent mirroring the S1-S5 configuration. R1 was not utilised, with R2 containing a solution of (**3.5**) in NEt_3. For the Sonogashira study, solvent selection was limited to a single selection, with D1-D5 containing the selection of ligands. Full detail on reservoir composition can be found in section 6.3.2.

Observing Figure 3-3, the effect of solvent on the outcome of the reaction is evident, with the algorithm effectively switching between solvents in the process of obtaining Pareto optimal values. The algorithm was able to produce 20 Pareto points when considering the entire dataset, with good spacing between each of these Pareto values. The effect of the continuous input parameters on the final reaction outcome is highlighted in Figure 3-4. Interestingly, there is a notable difference in relative optimal conditions for each solvent, with acetonitrile generally favouring more forcing conditions (higher temperature and concentrations) when compared to other solvents. The absence of ethanol from the Pareto front is noted, with the algorithm only attempting one additional ethanol experiment in the entire run. This observed result is likely due to the overall poor results for the initial conditions of ethanol when compared to acetonitrile, with both solvents occupying a similar region in the objective space.

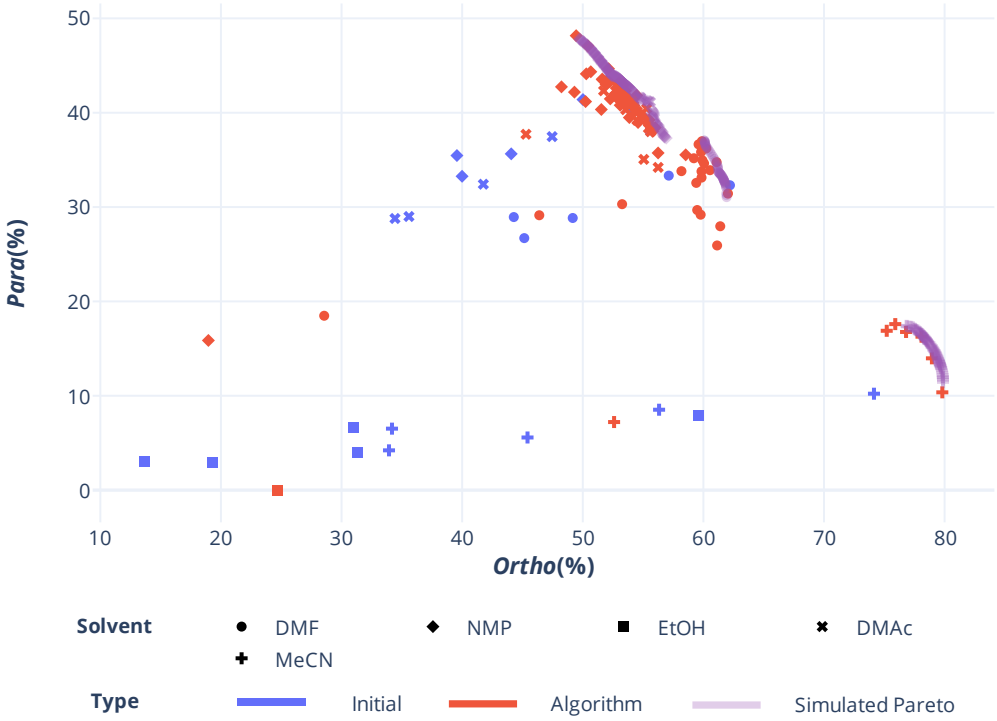


Figure 3-3. Pareto plot for experimental data, with additional simulated Pareto front, from the multi objective optimisation of the SNAr case study.

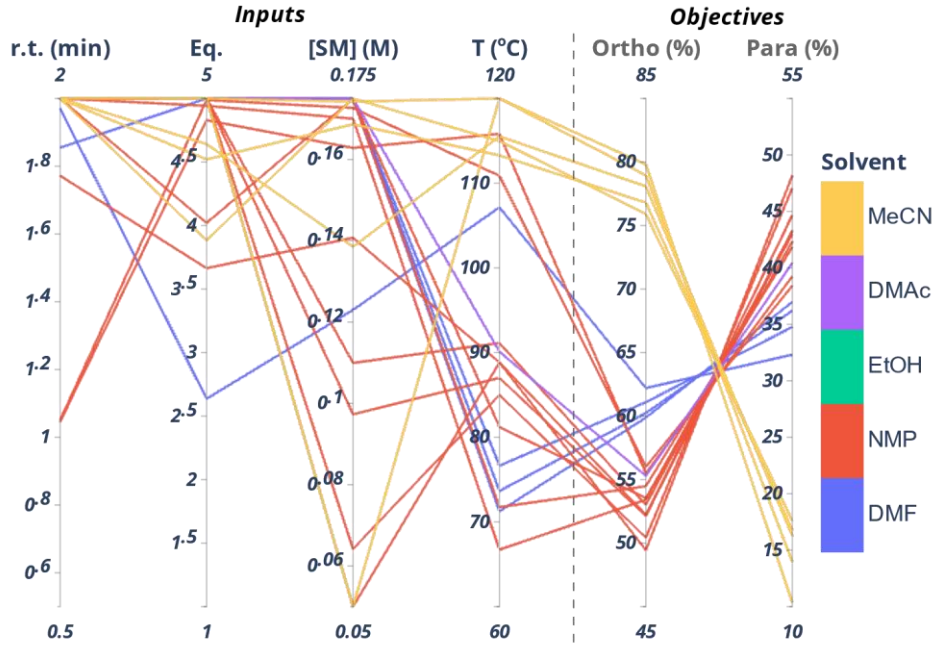


Figure 3-4. Parallel coordinates plot for the SNAr case study showing the effect of both discrete and continuous parameters on the objective functions for the Pareto points. Each line represents a single experiment.

Here, *a posteriori* multi-objective optimisation techniques provide a key advantage, exploring the entire Pareto front without any imposed bias on favoured objectives. Given the diverse data available after the optimisation, analysis on what underlying chemical descriptors effectively model the

observed behaviour can be investigated. With prior literature highlighting the important role the polarity of the solvent has on the outcome of the reaction, three common polarity metrics were examined: (i) polarity index, a measure of a solvents' ability to interact with various polar test solutes; (ii) dipole moment, the product of the magnitude of separated charges and the distance between them; (iii) dielectric constant, a measure of a substance's ability to insulate charges from each other.

Utilising the full optimisation data set, backward stepwise regression was performed separately for *ortho* and *para* products considering each descriptor individually. In backwards stepwise regression, the regression begins with a full model, a full quadratic model in this instance, and for each step, the variable with the largest p-value is removed if it is over a predefined significance value. This process continues until all variables are below the predefined p-value.

The quality of the three descriptor-based modelling approaches was evaluated using the coefficient of determination. The coefficient of determination (R^2), measures the proportion of variance in the data that is explained by the underlying model.[176] If we have a vector of observed responses $\mathbf{y} = [y_1, \ldots, y_n]$ and a vector of predicted responses $\mathbf{f} = [f_1, \ldots, f_n]$, R^2 can be calculated as follows:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \textbf{(3-2)}$$

$$SS_{tol} = \sum_{i=1}^{n} (y_i - \bar{y})^2 \qquad \textbf{(3-3)}$$

$$SS_{res} = \sum_{i=1}^{n} (y_i - f_i)^2 = \sum_{i=1}^{n} e_i^2 \qquad \textbf{(3-4)}$$

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \qquad \textbf{(3-5)}$$

In the best case the residual sum of squares SS_{res} will be zero and R^2 will equal 1. One issue presented by the coefficient of determination, is that it can increase when extra variables are added. This can lead to an increase in the value of R^2 and thus should not be used to compare models with a different number of variables. Given this, Ezekiel proposed the adjusted R^2 metric.[177] The adjusted R^2 regulates the R^2 metric based on the number of variables in the model relative to the number of data points. As such, it is more applicable for comparing alternative models that may have differing numbers of model terms, with it given as follows:

$$R^2_{adj} = 1 - (1 - R^2)\frac{n}{n-p} \qquad \textbf{(3-6)}$$

Where p is the number of variables in the fitted model and n is the number of observations. Fitting data for the 6 polynomial models is shown in Figure 3-5, Figure 3-6 and Figure 3-7.

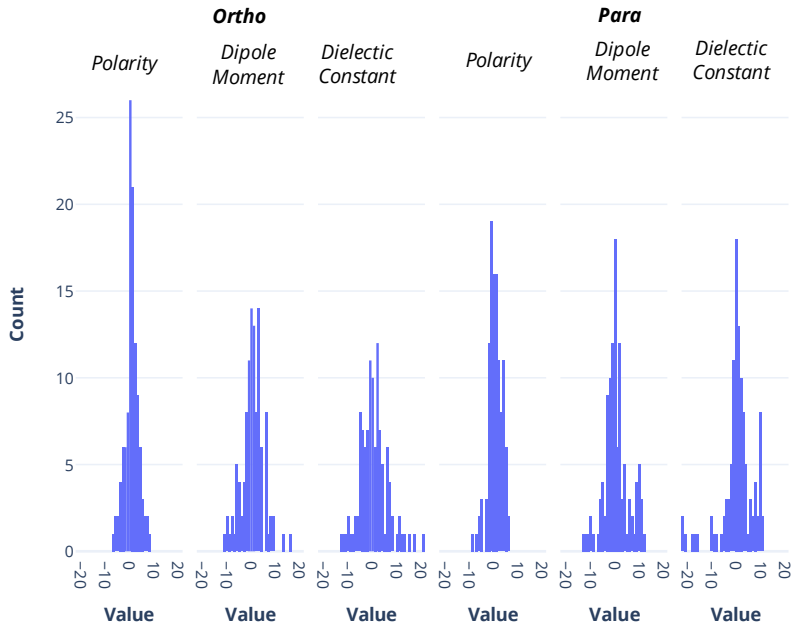


Figure 3-5. Residual histograms for each model.

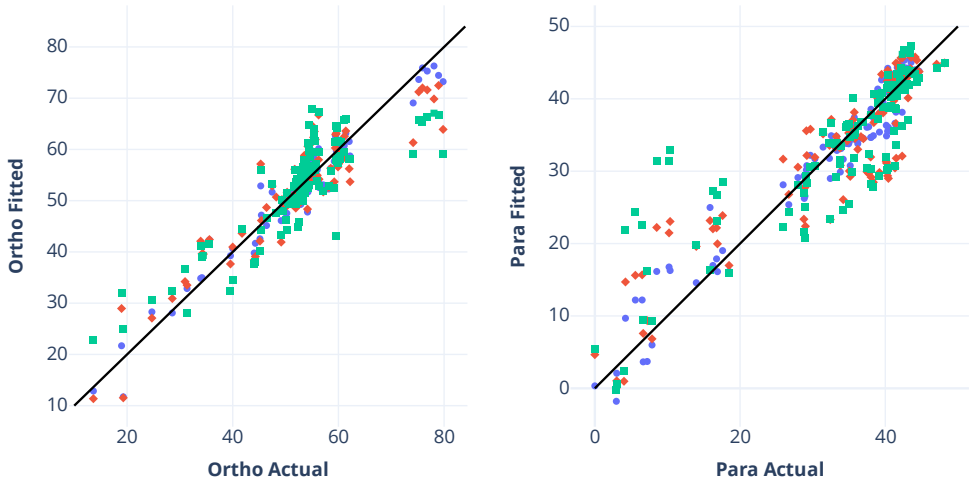


Figure 3-6. Actual vs. fitted *ortho* and *para* product yields for all fitted polynomial models: Polarity index - blue circle (ortho: $\boldsymbol{R^2_{adj} = 0.9259}$, para: $\boldsymbol{R^2_{adj} = 0.9388}$); dipole moment - red diamond (ortho: $\boldsymbol{R^2_{adj} = 0.8178}$, para: $\boldsymbol{R^2_{adj} = 0.8055}$); dielectric constant - green square (ortho: $\boldsymbol{R^2_{adj} = 0.7084}$, para: $\boldsymbol{R^2_{adj} = 0.6643}$).

The residual histograms generally suggest that the error in the models is normally distributed, suggesting our underlying assumption of normally

distributed error is correct. For both *ortho* and *para* polynomial models, polarity index delivers models with the best fit to the experimental data achieving R^2_{adj} values of 0.9259 and 0.9388, respectively.

Figure 3-7. Normalised weights for both *ortho* and *para* models utilising different chemical descriptors.

Analysis of the normalised weights for the *ortho* polarity index based model indicates that with the given data there exists an optimal value for polarity index in maximising *ortho* product. This is indicated by the magnitude of the coefficient for polarity index and its square term. The prediction of an optimal polarity index level for *ortho* product generation is hypothesised to be due to the imposed residence time constraint. Likely, should residence time be increased, ethanol may achieve higher levels of *ortho* product. The lower levels of *ortho* product observed in ethanol are likely due to a reduction in overall reaction rate, due to the polarity of the solvent.[178] The link between polarity index of a solvent and a reduction in reaction rate is further verified with acetonitrile favouring more forcing conditions when compared with solvents with a higher polarity index such as NMP and DMAc.

The relationship deduced by the model, in which increasing the polarity index of the solvent results in an increase in *para* product formation, correlates well with existing literature studies.[173] The findings highlight the importance that chemical descriptor selection can have on a models ability to accurately map the input domain to the objective space. Selecting either dipole moment or dielectric constant fails to completely describe the observed relationship, with dielectric constant predicting MeCN to behave similarly to DMAc, which was not observed in the data.

One solution around predefining selective descriptors would be to include a wide array of descriptors and perform subsequent optimisations utilising this large set. A key disadvantage to this approach is the significant increase in the optimisations input dimensions which can have a dramatic effect on the efficacy of many optimisation algorithms.[101] Principal component analysis, alongside other dimensionality reduction techniques, can be utilised here, however, this still assumes that the selected descriptors accurately map any observed relationship which may not be the case for novel systems. Moreover, as discrete variables are now represented by their descriptors, certain descriptor levels may be in infeasible regions of the input space. This is likely to present issues for continuous optimisation algorithms should the algorithm continually suggest points that cannot be run. This problem can be circumvented through adopting a grid-based optimisation and limiting the evaluated points to a predefined grid, this approach was adopted by Doyle *et al*.[18] However, in limiting the continuous parameters to only grid based values, the search domain is limited, with the requirement to run finer grids for a more accurate optimisation. The reduction of grid step size can create extremely

large search domains which can become computationally challenging even when utilising surrogate model based approaches.

It is hypothesised that to bridge the discontinuous Pareto front observed in Figure 3-3, solvents with polarity indices between 5.8 and 6.4 would be recommended. This hypothesis was analysed utilising the experimental data and the polynomial models for both *ortho* and *para* products. A simulated study was performed in which the empirical models were optimised utilising NSGA-II. This enables full exploration of the estimated behaviour of the system with respect to the continuous inputs and polarity index. Figure 3-8 details the outcome of the simulated study.

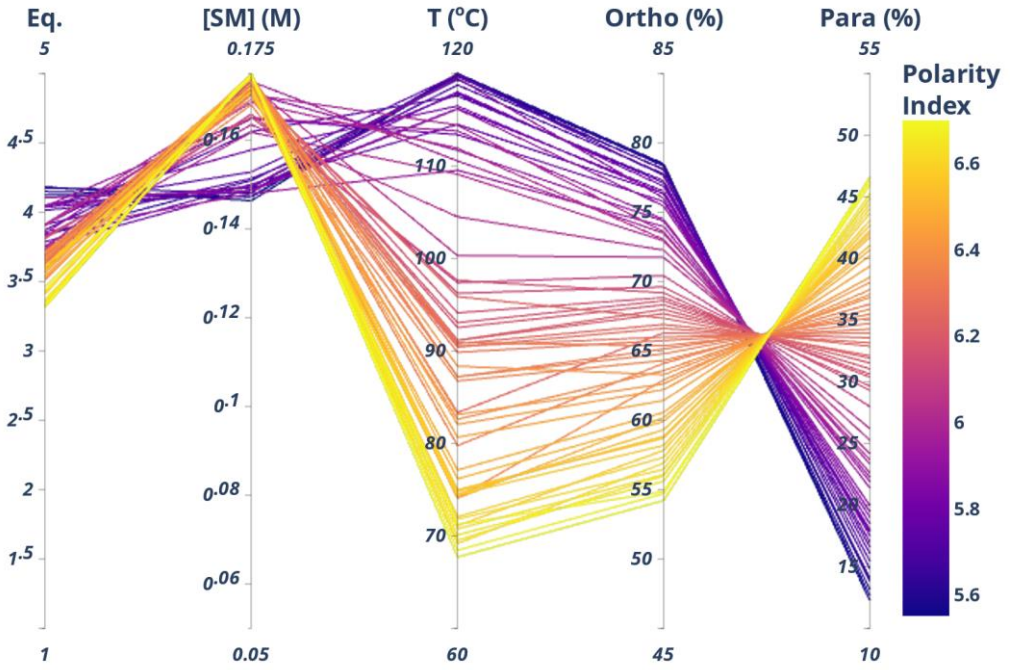


Figure 3-8. Parallel coordinates plot for the simulated Pareto front utilising polarity index as an input variable. Residence time for all Pareto points was 2 minutes and therefore has been excluded from the figure.

The results of the simulated study further correlate with prior literature, highlighting the estimated effect of polarity index, coupled with the reaction conditions, on the outcome of the reaction. The use of such simulated data is useful to highlight a target solvent, should the user require a specific reaction outcome. In this case study, this form of investigation has little use, however, it highlights the potential of additional data extraction when adopting an algorithmically led approach, with the ability to suggest alternative conditions not performed within a study based upon relationships determined with this form of data interrogation. This type of smart data acquisition illustrates a key area of application for the algorithm, with potential for use not only in process

optimisations but for exploring a novel reaction domain. It should be noted that utilising solvents in the suggested polarity index range would require solubility studies to be performed prior to the optimisation, with the model giving no consideration for this.

Analysing the progress of the algorithm utilising the hypervolume metric, Figure 3-9, indicates the improvements gained through application of an algorithmically led approach. The plot provides insight into how well the algorithm searches the domain, with large stall sections indicating the algorithm potentially gets stuck in multiple locally optimal fronts and requires a significant portion of experimental budget to escape. The small change observed in hypervolume can be attributed to the discontinuous nature of the front, with a large area of the ideal continuous Pareto front, Figure 3-8, unable to be accessed with the solvent choices available to the algorithm.

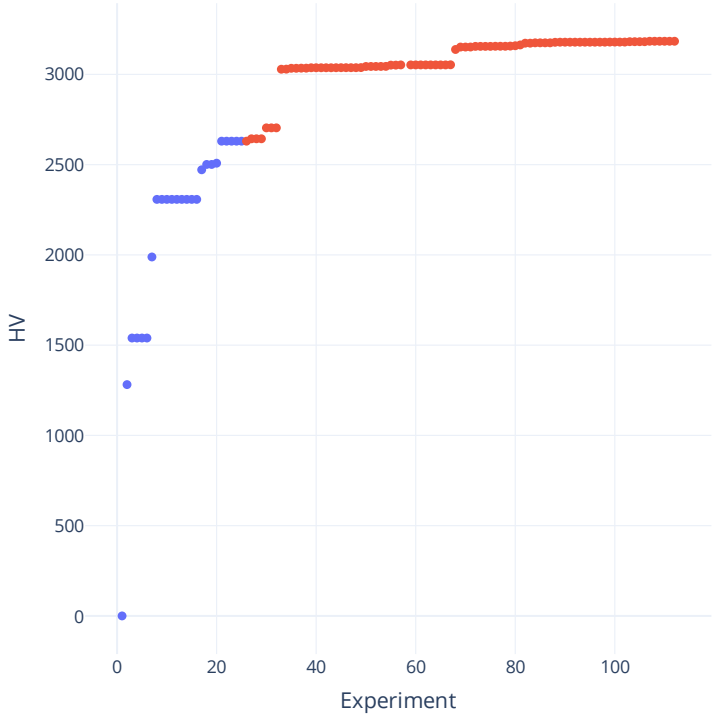


Figure 3-9. Algorithm progression in terms of hypervolume (HV) for the S_NAr reaction. A reference point of $\boldsymbol{R} = [\mathbf{0}, \mathbf{0}]$ was used for the calculation.

Given no prior knowledge the algorithm has been able to elucidate a diverse Pareto front, from which analysis of the experimental data and fitting can be performed *a posteriori.* As there is no need for prior data collection to build continuous descriptors the algorithm can provide an informed methodology of efficient data acquisition in the presence of limited data.

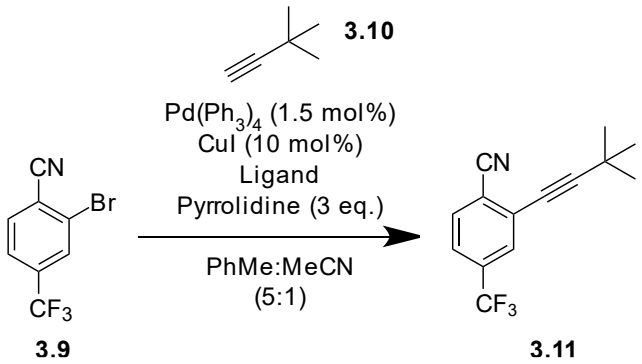
Given the successful application to a well understood experimental system, the algorithm was taken forward to a catalytic example. The inclusion

of a catalytic example was hypothesised to provide greater complexity in optimisation, with catalytic systems presenting a key potential application in screening designs given the wide use of catalysts in the pharmaceutical industry.[179]

3.3 Sonogashira Mixed Variable Optimisation

For the second case study a Sonogashira coupling reaction was selected, Scheme 3-3. The below reaction forms an initial stage in the synthesis of a $TRPV_1$ receptor antagonist, that provides an non-opiate approach to pain management.[180]

The reaction was modified for use with a flow system due to the initial conditions leading to a heterogenous solution not compatible with flow-based optimisations. Given this, an aryl bromide was utilised in lieu of the originally proposed aryl chloride, alongside use of an alternative catalyst, base, and solvent system.



Scheme 3-3. Sonogashira reaction between 2-bromo-4-(trifluoromethyl)benzonitrile (**3.9**) and 3,3-dimethyl-1-butyne (**3.10**). Five ligand options were utilised for the discrete variable optimisation: DavePhos, XPhos, CyJohnPhos, tricyclohexyphosphine and triphenylphosphine.

An initial 25 experiments, five per palladium-ligand complex, were performed, the results of which were employed to select optimisation targets that would result in a challenging multi-objective optimisation for the algorithm. The selection of competing objectives is not a requirement for an effective multi-objective optimisation, with the algorithm capable of handling objectives with shared optima. However, ensuring objectives compete allows for a more robust test of the algorithm for applied problems during the proofing stage of algorithm development. Commonly utilised optimisation targets were utilised, Table 3-1,with Figure 3-10 summarising the results of the initial dataset objective comparison. The figure was used to determine the objective

functions with the greatest trade-off to allow for full testing of the multi-objective algorithm.

Table 3-1. Commonly utilised economic and environmental[181,182] metrics for evaluating a chemical process. $\boldsymbol{m}$ – mass, $\boldsymbol{V}$ – volume, $\boldsymbol{t_{res}}$ – residence time, $\boldsymbol{n}$ – moles, $\boldsymbol{C_i}$ – cost of component i. $\boldsymbol{m_{in}}$ is the total mass in including reaction solvents.

Metric	Equation	
Yield	$\frac{100 \times m_{product}}{m_{theoretical}}$	**(3-7)**
Space-time-yield (STY)	$\frac{m_{product}}{V \times t_{res}}$	**(3-8)**
Reaction mass efficiency (RME)	$\frac{100 \times m_{product}}{m_{reactants}}$	**(3-9)**
E-factor	$\frac{m_{in} - m_{product}}{m_{product}}$	**(3-10)**
Turnover number (TON)	$\frac{n_{reacted}}{n_{catalyst}}$	**(3-11)**
Turnover frequency (TOF)	$\frac{TON}{t_{res}}$	**(3-12)**
Cost	$\frac{\sum C_i \times m_i}{m_{product}}$	**(3-13)**

Yield is extensively utilised as a metric when evaluating the performance of a reaction, with it providing a measure on the conversion of the limiting reagent to the desired product. Although widely used, the metric provides little insight into the efficiency of the process in terms of raw material utilisation[183] or productivity.

Turnover frequency provides a measure of the number of revolutions of the catalytic cycle per unit time.[184] The metric provides an insight into the activity of the catalyst at the current conditions and is widely used to measure the performance of a given catalyst. Similar to yield, however, it provides no insight into the overall efficiency of the process in terms of waste produced, and therefore should be coupled with more environmentally focused metrics such as RME or E-factor. It should be noted that neither RME or E-factor consider the hazards or environmental risk of the waste produced from the process. To capture this completely the use of life cycle assessment is recommended, however, where this is unavailable metrics combining RME

and/or E-factor with hazard rankings for waste[185] may produce metrics that estimate the environmental impact of a process to a greater extent.

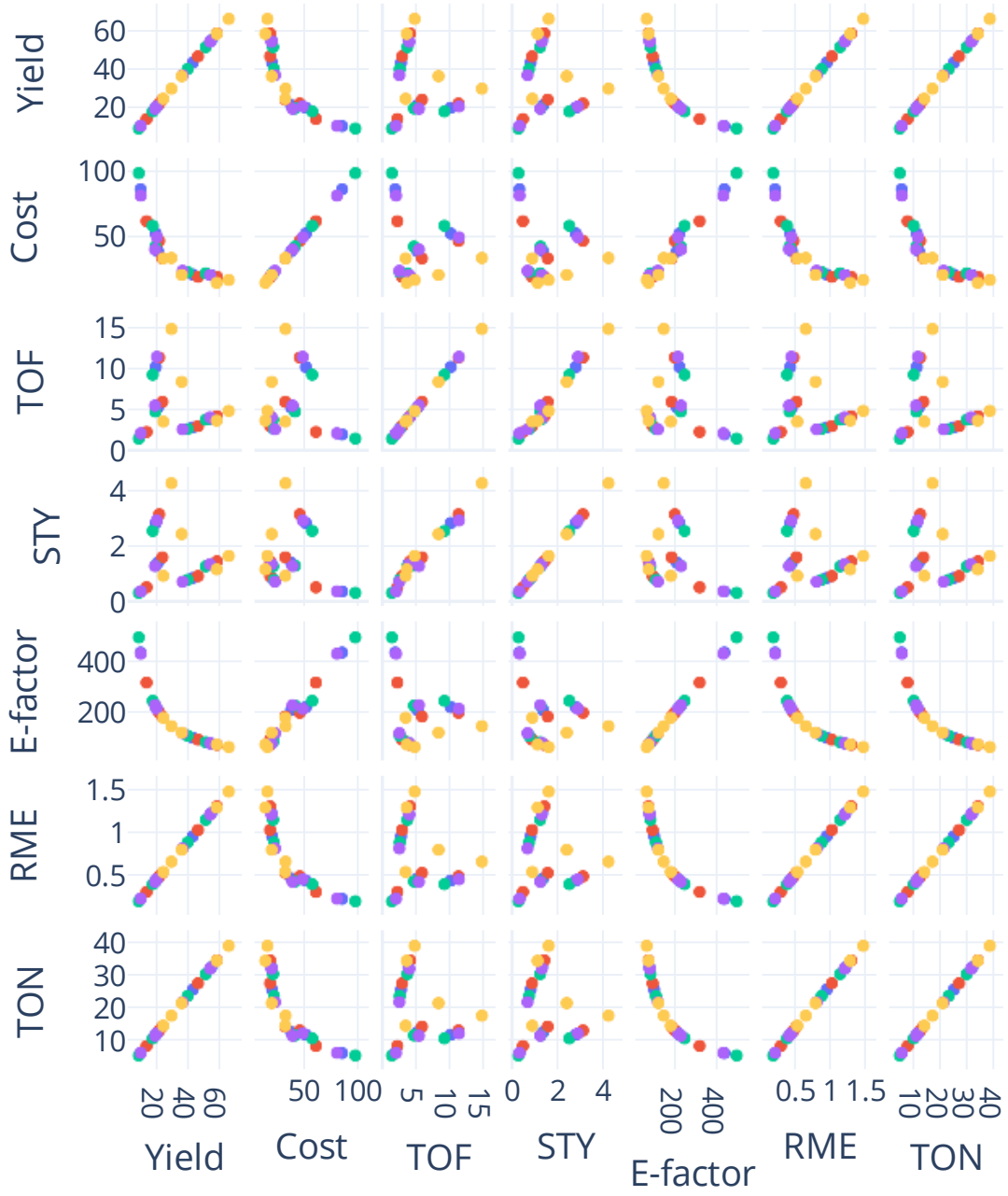


Figure 3-10. Scatter matrix comparing candidate objective functions for Sonogashira optimisation. Colours correspond to the following: blue – DavePhos, red – XPhos, green – CyJohnPhos, purple – tricyclohexyphosphine and yellow – triphenylphosphine.

Cost is a key driver for any process and can be argued to be one of the most important metrics when evaluating a process.[186] Utilising cost provides an effective way of selecting not only reagents/components that lead to an efficient process but also a cost effective one, allowing for the substitution of complex, expensive components for simple and more cost-effective ones. A

key downside to the sole consideration of process costs is the lack of inclusion of costs relating to extraction of the raw materials as well as non-monetary costs associated with use and production of material. These other costs can be encapsulated somewhat through the use of additional environmental metrics, however, often a full life cycle assessment is required to understand the full impact of a production process.[187]

It was hypothesised that the more expensive Buchwald ligand options would lead to improved yields at the sacrifice of increasing the overall cost of the reaction, leading to the possibility of a mixed ligand Pareto front.

Analysis of the initial conditions led to the optimisation being formulated according to (3-14). Yield, turnover frequency, and cost were selected due to their apparent trade-off in the initial data.

$$max[Yield, TOF, -Cost/g]$$

$$Ligand \in \{DavePhos;\ XPhos;\ CyJohnPhos;\ Tricyclohexyphosphine;\ Triphenylphosphine\}$$

$$t_{res}(mins) \in [1,10] \qquad \textbf{(3-14)}$$

$$\textbf{7}\ Eq. \in [1,3]$$

$$Ligand\ Conc.(mol\%) \in [1,5]$$

$$T(^{o}C) \in [60,140]$$

Following initial data collection, the algorithm was afforded an additional 64 experiments to determine the systems Pareto front. Figure 3-11 details the results of the multi-objective optimisation with respect to the objective space. Full results for the optimisation can be found in section 6.3.3.

In contrast to our hypothesis the results indicate that triphenylphosphine clearly provides the optimal performance with respect to all objectives. Given the Pareto front consist entirely of a single ligand choice, there is minor variation with respect to cost, with it following a linear trend with respect to yield and both objectives sharing optimal values. This relationship is highlighted when the results are displayed in scatter matrix form, Figure 3-12, with the green Pareto points clearly showing a linear relationship between objectives.

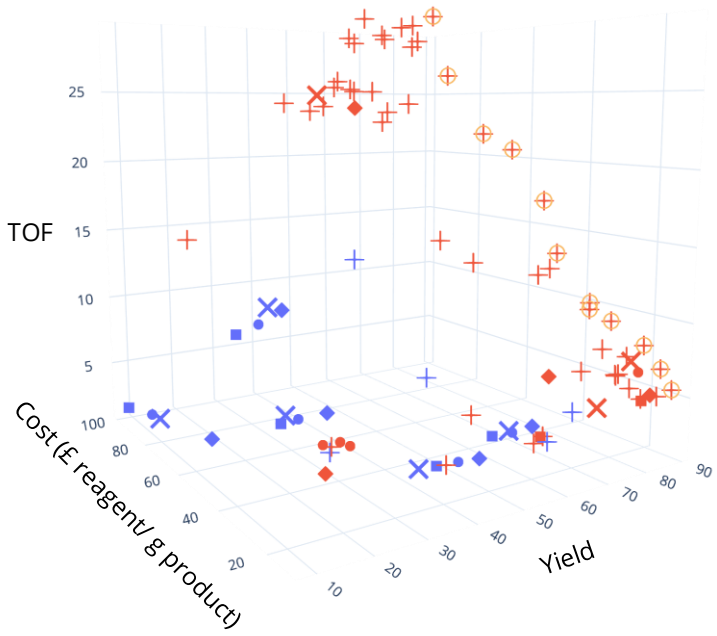


Figure 3-11. Pareto plot for the Sonogashira case study optimisation. Colours correspond to the following: blue – initial conditions, red – algorithm suggestions. Shapes correspond to the following ligands: • - DavePhos, ♦ - XPhos, ▪ - CyJohnPhos, x – tricyclohexyphosphine and + triphenylphosphine. Pareto points are circled in orange.

The benefits of utilising an algorithmically led approach are evident with the significant increase in calculated hypervolume from the initial conditions to the final results, Figure 3-13. The initial rapid increase in hypervolume indicates how the algorithm rapidly moves towards an optimal region of the objective domain, with the latter stages of the algorithm focused on minor improvements once a close approximate for the systems Pareto front has been determined. The large degree of plateauing suggests the algorithm struggles to find any additional improvements; this behaviour potentially suggests that the algorithm could have been terminated earlier with minimal loss to the overall results. Premature termination of the optimisation does offer benefits in terms of resource savings, however, in black box optimisations it is difficult to determine whether the algorithm is temporarily trapped in a local optimum. Full utilisation of a sufficiently large experimental budget, with an algorithm with prior successful optimisation applications, can provide confidence to the user that the front determined is a close approximation for the true Pareto front of the investigated system.

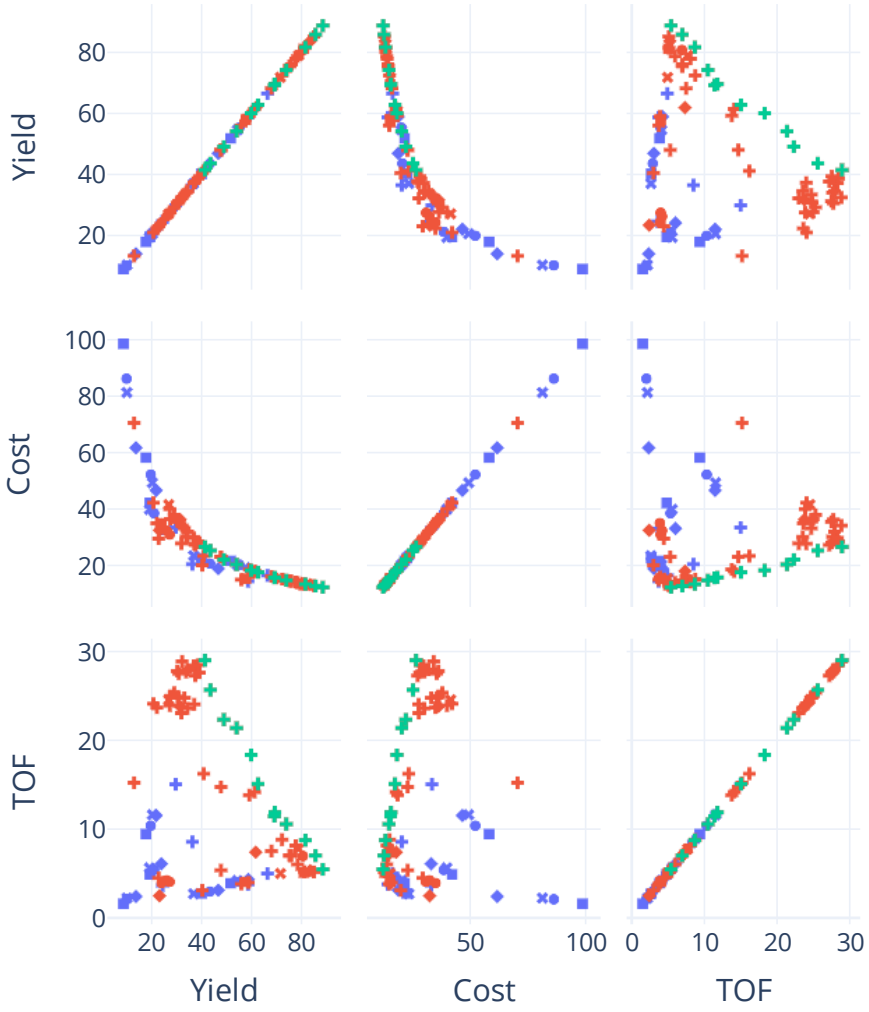


Figure 3-12. Scatter matrix for the objective domain of the Sonogashira reaction optimisation. Colours correspond to the following: blue – initial conditions, red – algorithm suggestions, green – algorithm Pareto points. Shapes correspond to the following ligands: • - DavePhos, ♦ - XPhos, ▪ - CyJohnPhos, x – tricyclohexyphosphine and + triphenylphosphine.

The decision to enact an early termination strategy is very much problem and resource specific, with full budget utilisation likely acceptable in smart data acquisition tasks compared to processes in need of rapid optimisation, given this will often be the limiting factor due to resource constraints placed upon the optimisation. This point is reinforced through observation of Figure 3-9 in which significant stalls in hypervolume are experienced. The use of terminal criteria for black box systems will likely require multiple termination criteria to be utilised in an ensemble type methodology, combining criteria and/or performance metrics to either automatically trigger termination or provide the user with sufficient data to make an informed choice.

Alternatively, improvements in automated platform reagent efficiency can likely offset the need for any termination criteria, with reductions in material consumption per experiment allowing for a greater budget being afforded to the algorithm. Recent examples of high reagent efficiency platforms have illustrated how significant experimental budgets can be offered whilst minimising reagent consumption.[15,188,189] Notably, Gesmundo *et al.* present a nanomole scale high-throughput batch platform, requiring less than 0.05 mg of substrate per reaction.[189] This allows for the potential of alternative algorithmically led approaches to be adopted such as a genetic algorithm based approach, which offers lower per experiment efficiency, but, given its random nature, allows for extensive exploration of the experimental domain.[60,190] Although not applicable to all systems, the wider accessibility of reagent efficient platforms allows for the study of complex expensive systems that may have previously been limited.[191] For chemical system optimisations, the selection of approach needs to be done on a case-by-case basis, with selected operation likely combining both options to allow for adequate domain exploration.

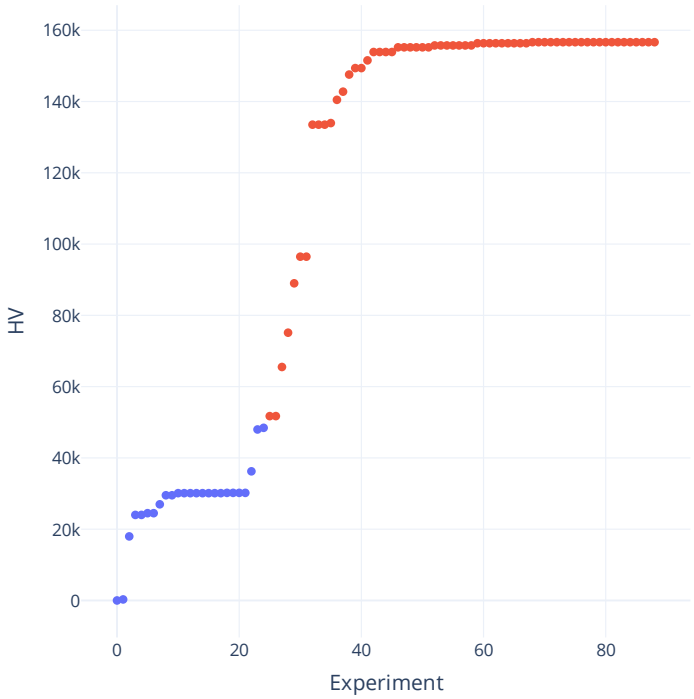


Figure 3-13. Iterative hypervolume for the Sonogashira reaction optimisation. Reference point of $\boldsymbol{R} = [\mathbf{0}, \mathbf{100}, \mathbf{0}]$ was used for the calculation.

Analysis of the input domain reveals the algorithm predominantly favours triphenylphosphine across the entire run, with the algorithm selecting the optimal ligand in 80% of the suggested experiments, Figure 3-14. Observation of the initial conditions indicates the reason for this behaviour,

with triphenylphosphine demonstrating optimal performance from the five initial conditions. Although efficient, this behaviour may not always be desirable. Wider examination of alternative combinations, even if poorly performing, may be advantageous to serve as a verification process for the optimisation outcome or when the algorithm is being employed in a smart exploration type application.

Analysis of the Pareto front input data revealed that residence time was exploited to control the trade-off between the objectives, Figure 3-15. The wide range of values for ligand mole percent indicate that this parameter has little to no effect on the outcome. This observation is expected, with triphenylphosphine being utilised in the precomplexed form, when compared to the alternative ligands. This has likely contributed to the optimal performance observed, with there being no requirement for ligand exchange when compared to the alternative ligand options. Here, the algorithm has highlighted the cheapest and simplest option to implement as the optimal choice, with the selected complex readily available in its precomplexed form, without the need for an additional stage, as necessary for the other options.

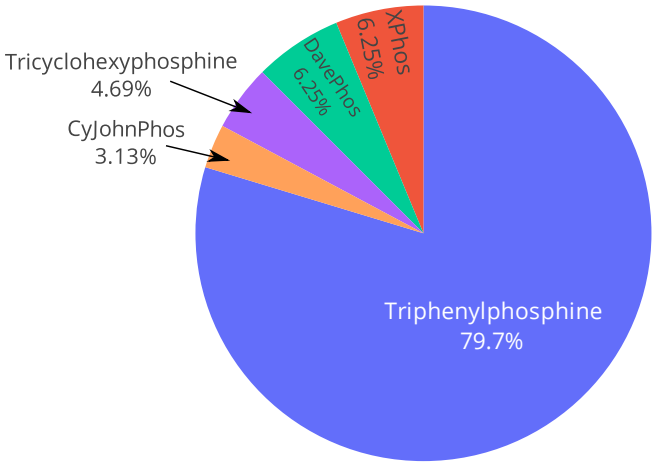


Figure 3-14. Algorithm ligand selection across the entire optimisation.

For both equivalents of alkyne and temperature the algorithm favours values towards the upper bounds of the optimisation moving the reaction to a more forcing regime. This was hypothesised to be due to the residence time constraints placed upon the system, with the reaction needing such conditions to sufficiently maximise the rate of reaction. Increasing the residence time bounds to a wider range could potentially introduce more variability with respect to the input domain of the Pareto points, moving the optimal reaction space to one which is not solely controlled by varying residence time. However, further experimentation would be needed to validate this hypothesis.

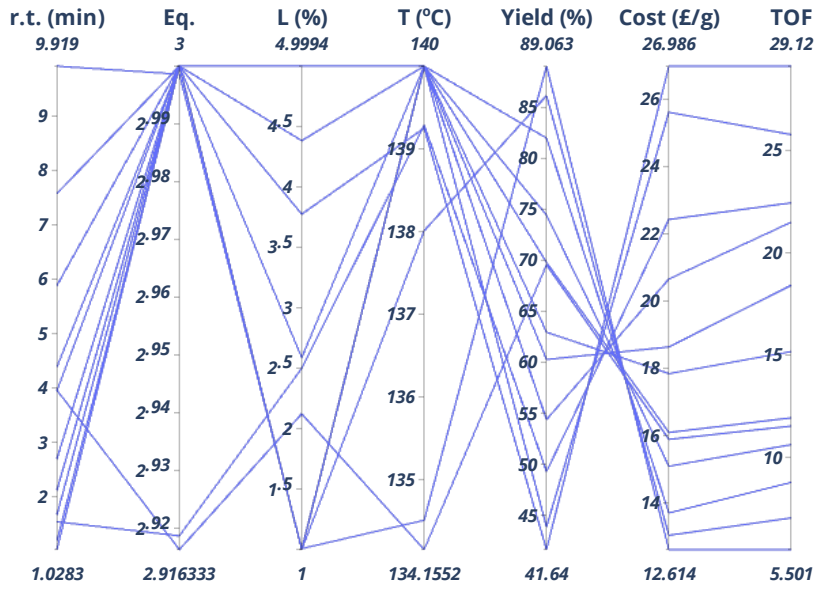


Figure 3-15. Parallel coordinates plot for Pareto points of the Sonogashira reaction optimisation.

Utilising the surrogate GP model and a mixed variable implementation of NSGA-II, a simulated Pareto front was generated. The simulated front can be used to interrogate the models understanding of the system and provide information of the dominant effects provided by the input domain, Figure 3-16.

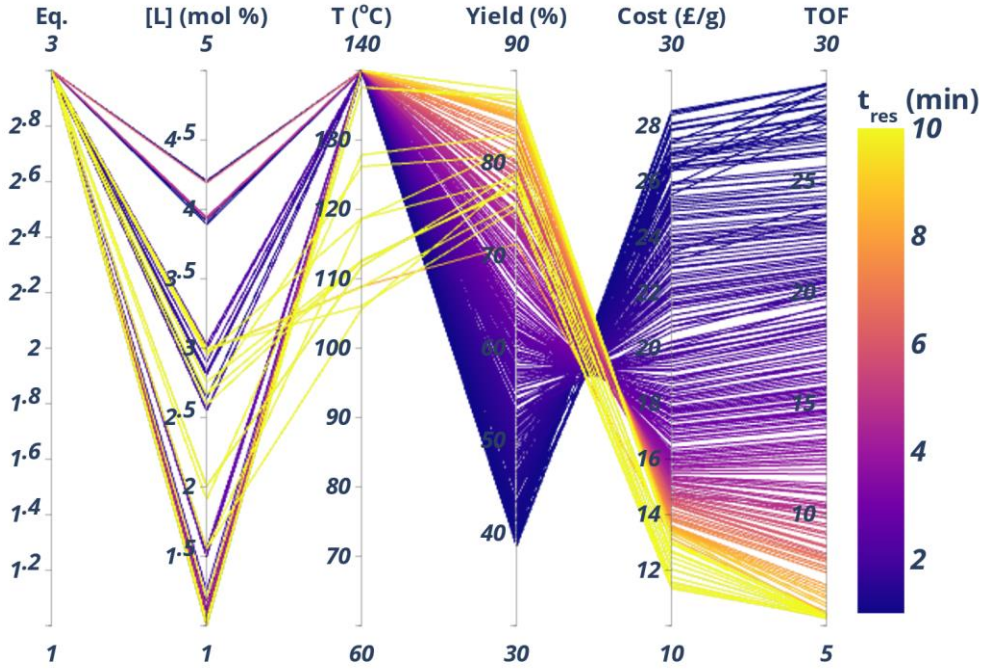


Figure 3-16. Parallel coordinates plot for the Sonogashira reaction optimisation simulated Pareto front.

From the simulation results, the dominant effect of residence time that is predicted by the model is clearly evident. In similar fashion to the true data,

ligand concentration varies somewhat randomly across the simulated Pareto front, suggesting it has little to no effect on the predicted reaction outcome. Reagent equivalents and temperature both favour values towards the upper end of the bounds, suggesting that the progression of the reaction is benefited by the more forcing conditions offered at these values.

Optimisation without the need for *a priori* data was successfully performed for both case studies. In both instances the algorithm moved efficiently towards a Pareto optimal set considering both continuous and discrete parameters. The ability to optimise systems without prior work in generating continuous descriptors and determining relevant parameters presents a streamlined approach to mixed variable multi-objective optimisations without transferring workload to expensive computations. Given the successful application of the algorithm in the two case studies, combination of the proposed methodology with more advanced automated screening platforms, such as the ones presented by Perera *et al.*[15] and Shields *et al.*[18] would prove beneficial. The expansion to a large set of discrete variable choices, as well as offering more efficient material usage throughout the optimisation, would allow for application in exploring challenging and expensive chemistries, aligning the approach as an alternative to HTE screening approaches, for which the currently illustrated platform may prove too inefficient.

3.4 Conclusions

Early-stage chemical reaction screening represents a complex challenge with the requirement to balance information gain with minimising material consumption. Where systems are sufficiently challenging, utilising mixed variable optimisation methodologies to guide experimentation can provide a simpler route than conventional design experiment routes. Having methodologies capable of handling such instances is vitally important, especially in novel development work.

In this work, a mixed variable multi-objective optimisation algorithm has been applied to the automated optimisation of two reactions: a S_NAr reaction and a Sonogashira reaction. A bi-objective optimisation was performed for the S_NAr reaction in which the effect on reaction solvent, alongside four other continuous reaction variables, on the trade-off between *ortho* and *para* product formation was investigated. Subsequent optimisation data analysis highlighted that the polarity index of the solvents was a key factor in determining the ratio of *ortho* to *para* product with high polarity index solvents

favouring more *para* product formation. For the Sonogashira reaction, yield, cost, and turnover frequency were optimised simultaneously with respect to palladium-ligand complex and four continuous parameters. Analysis of the results highlighted a single optimum palladium-ligand complex for all objectives, with the triphenylphosphine complex dominating the Pareto front for the system.

This example demonstrates one of the first reported applications of an *a posteriori* based mixed variable multi-objective optimisation algorithm to an experimental system. Both case studies clearly highlight the benefit of adopting an algorithmically led approach in mixed variable tasks that require optimisation of multiple competing variables. The improved efficiency offered by such techniques lends them to early-stage development applications, where material consumption must be kept to a minimum.

Chapter 4 Hybridised Optimisation Algorithm

4.1 Introduction

The coupling of algorithms and/or methodologies can lead to enhanced capability and performance when compared to use of a singular algorithm. In regions where the problem is smooth and convex, local optimisation algorithms can outperform their global counterparts in both speed and precision.[95]

Determining the optimum for a process is essential to enable efficient operation. However, it is important to understand the response surface around the optimum, given the dynamic nature of chemical processes. Full knowledge of the response surface around an optimum can inform the user of how changes in process inputs affect the response variable for the system. With single point optimisation, should no further response work be performed, the user lacks a complete understanding of the full reaction landscape surrounding the optimum. Full understanding of this area can highlight whether the optimum determined is near a performance "cliff face" and how far this cliff edge extends into the surrounding reaction landscape or whether there is a larger unexplored region of process stability. Clayton *et al*. recently reported the presence of a cliff edge or cliff face type response surface, for the selective extraction of amine mixtures in a continuous flow system, Figure 4-1.[63]

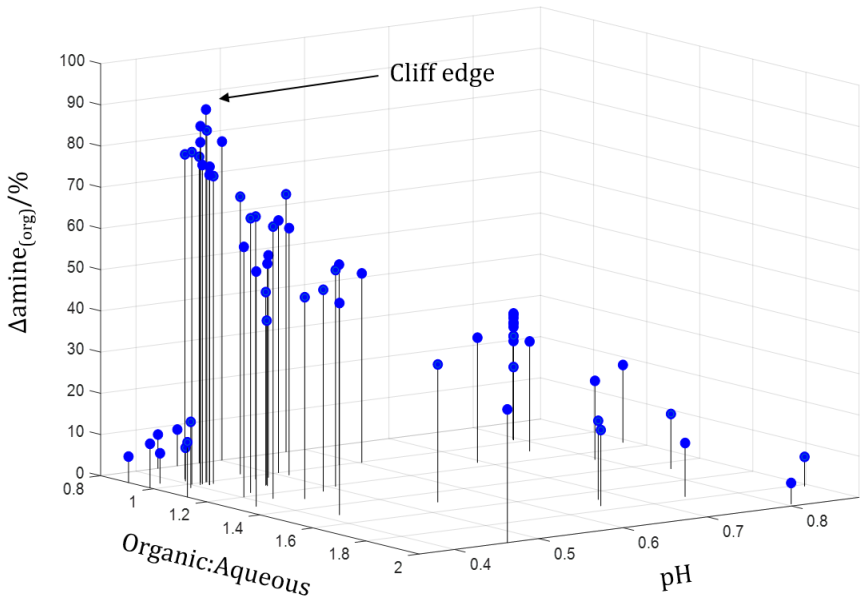


Figure 4-1. Cliff edge reactive extraction. Reprinted from Clayton.[158]

The example highlights the need for not only efficient optimisation, but accurate modelling of the cliff face region for greater process understanding. Response surface work around the optimum allows the user to ascertain this knowledge, further informing the continued design process.

Prior work by Wyvratt *et al.* has looked at the elucidation of a response surface model through designed dynamic experimentation.[192] Applying a sinusoidal ramp to the process inputs, the methodology is able to efficiently explore the input domain to determine the underlying response surface, Figure 4-2. The authors applied the methodology to a Knoevenagel condensation reaction, observing the effects of base equivalents and residence time on the concentration of the product exiting the flow reactor. As there is no feedback aspect to the methodology, samples were collected at fixed interval during the dynamic experimentation with a fraction collector utilised for offline analysis.

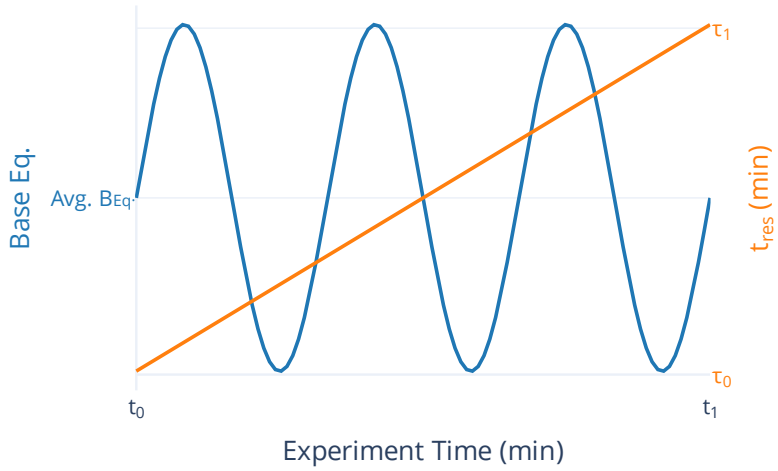


Figure 4-2. Non-linear dynamic experimentation proposed by Wyvratt *et al.*. Adapted from Wyvratt *et al.*[192] Sinusoidal input utilises the following for its calculation: $\boldsymbol{Base\ Eq.} = \boldsymbol{Avg.B_{Eq}} + \mathbf{0.1sin}\left(\frac{\boldsymbol{\pi t}}{\mathbf{5}}\right)$

As this methodology searches the whole domain, for systems with a larger input domain, this could result in the requirement for excessive experimentation and an overall reduced efficiency in data acquisition. Additionally, as the generation of the sinusoidal wave for process inputs requires precise and continuous change in the equipment controlling the parameters, such as pumps, errors can be introduced and intensified as the dynamic experimentation proceeds. Application to systems with multiple variables may become infeasible, without precise monitoring of true flowrates, due to the number of additional pumps; with each pump adding additional sources of error. This can be mitigated with use of independent process condition monitoring, however, the need for additional instrumentation is likely to offset any benefits gained through utilising this dynamic approach. Although

this will be heavily dependent on the application. In such instances, or where reaction times are significantly long, it is likely more appropriate to resort to methods that utilise batch or steady state reactions in their data collection. A similar outcome could be achieved through application of design of experiments, with iterative response surface designs able to determine the process optimum, as well as provide information on the reaction landscape.[193] The requirement for iterative designs is a necessity for complex reaction landscapes, with a single design unable to effectively map detailed changes over a wide domain.[194] Although effective for a small number of process inputs, such approaches can be affected by the 'curse of dimensionality' requiring excessive experimentation for high dimensional search spaces, with experiments scaling exponentially with increasing input domain size.

The development of response surface models around an optimum can provide key insights into the underlying system behaviours. In systems where full mechanistic understanding is yet to be obtained, this information can play a key role in accelerating the progress along the process development pipeline.

Herein, this work describes the development of a hybrid optimisation and response surface mapping algorithm, initially using a local optimisation approach, followed by subsequent application of global optimisation methods with comparison provided utilising a simulated chemical reaction. The algorithm is then applied to a photochemical reaction in which an optimal operating region is determined. Based on limitations observed in the experimental application of the algorithm, subsequent algorithm improvements are suggested, with Bayesian optimisation techniques applied, and comparison to the original methodology is performed utilising a surrogate model of the photochemical reaction and the initial simulated chemical reaction.

4.2 Hybrid Optimisation Algorithm Development

In using an optimisation algorithm to determine the area of interest; it was hypothesised that the later response surface mapping could be more efficient and have a greater information density around the optimum, in comparison to iterative design of experiment studies. Additionally, where models do exist, the proposed methodology can be employed in a similar manner to that of Quaglio *et al.*, in which a utility function is employed to compare model estimates with the actual process to highlight regions of model validity.[195] In this instance the utility function would be utilised as the objective

function for the algorithm with a limit set on its value to define where the approximate mechanistic model moves to a region of invalidity. An illustrative summary of the hybrid optimisation approach alongside a conceptual overview are provided in Figure 4-3 and Figure 4-4.

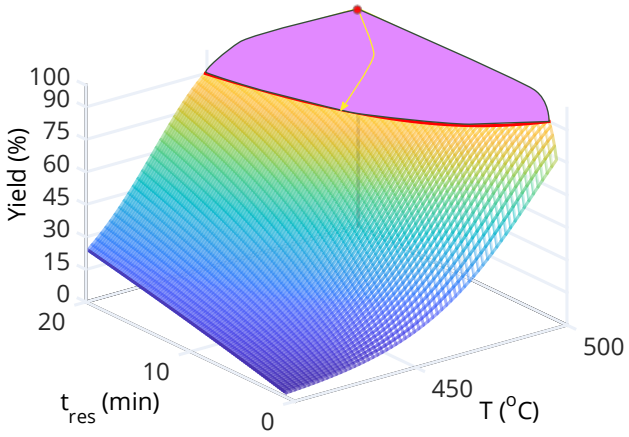


Figure 4-3. Overview of the algorithm goal for a simulated reaction. The optimum is shown in red, with the mapped response surface indicated in pink. The yellow arrow indicates the search direction of the algorithm from the optimum point.

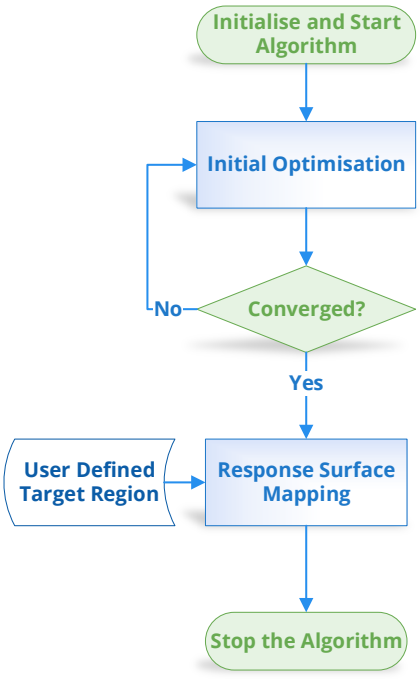


Figure 4-4. Conceptual overview of the hybrid algorithm.

4.2.1 Local Search and Screening

To develop the concept and structure for the algorithm a local search methodology was initially utilised. A bounded variant of the Nelder Mead simplex algorithm was selected as the primary optimisation algorithm. The

Nelder Mead simplex and its variants have seen extensive use in chemical system optimisation and are known for their optimisation efficiency in convex objective domains.[22,24,48,196]

The initial optimisation procedure, once a process optimum has been determined, is subsequently followed by the mapping portion of the algorithm. The algorithm was designed to map the area surrounding the optimum, in terms of objective space change, with the size of the area desired selected by the user prior to performing the optimisation.

To determine this area surrounding the optimum, a sequential screening design was utilised with the concept to expand from the optimum, as depicted in Figure 4-3. This iterative expansion aims to map the area around the optimum and determine regions matching the user's reduction criteria. The reduction criterion is utilised by the screening portion of the algorithm to define the limits of the area of interest for the user. Given these limits with respect to the input domain are unknown prior to the optimisation, an iterative procedure is employed to determine the mapping limits. The screening design utilises a quadratic model as a surrogate model. Surrogate modelling was selected to reduce the evaluation burden required in the screening design. An example of a quadratic model is provided:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \epsilon \qquad \textbf{(4-1)}$$

Multiple linear regression was employed to estimate the model terms, which when utilising an ordinary least squares approach, can be given by:

$$y = X\beta + \epsilon \qquad \textbf{(4-2)}$$

$$\hat{\beta} = (X^T X)^{-1} X^T y \qquad \textbf{(4-3)}$$

Where y_i is the dependent variable/response, x_i are factors and β_k are weightings for each factor. x_i can be process variables or can be transformations of process variables such as their squared term or a multiplicative combination of factors. In matrix form, X is a matrix of process variables and their selected transformations, with β a vector of the calculated weights for the variables.

Utilising this surrogate model, the upper and lower bounds of a Plackett-Burman screening design were optimised. The Plackett-Burman design is a low experimental burden design of experiments methodology and

is often utilised in dimensionality reduction of experimental search spaces.[197] Given this, the Plackett-Burman design was selected with the efficiency of the design and the nature of the screening task only requiring search limits to be identified the main reasons behind its selection.

The goal of the internal optimisation was to minimise the squared difference between the outputs of the surrogate and the user defined target for all screening design points. This theoretically should iteratively lead to the design spreading from the optimal point, determining the regions of interest for the user. Algorithm 1-1 provides an overview of the screening portion of the algorithm. Upon each iteration of the screening design, the internal optimisation bounds are set to ensure the local algorithm expands from the optimum point. Bounds are kept within their global limits with a constraint check.

Algorithm 4-1: Iterative screening design

begin
$lb \leftarrow optimum * 0.95$
$ub \leftarrow optimum * 1.05$
$allow_{level} \leftarrow optimum - limit_{user}$
$exit \leftarrow FALSE$
while $exit == False$ **then**
    $lb, ub \leftarrow checkConstraints(lb, ub, bounds_{global})$
    $conditions \leftarrow PBDesign(lb, ub)$
    $y \leftarrow f(conditions)$
    if $y_i \approx allow_{level}\ \forall y$ **then**
        $exit \leftarrow TRUE$
    $mdl \leftarrow surrogate(X_{all}, y_{all})$
    $f_{temp}(x) \leftarrow (mdl.predict(x) - allowable)^2$
    $bounds_{optim} \leftarrow setBounds(lb, ub)$
    $lb, ub \leftarrow fmincon(f_{temp}, bounds_{optim})$
    if $\left(lb_i == b_i\ \forall lb\ \forall bounds_{global}$ **and** $ub_i == b_i\ \forall ub\ \forall bounds_{global}\right)$ **or** $its == maxIteration$ **then**
        $exit \leftarrow TRUE$
end

Internal optimisation bounds are set to ensure the lower and upper bounds do not cross during the internal optimisation. This was done to ensure the bounds do not move to a common area of the experimental domain. Once the upper and lower bounds were defined, they were utilised to perform a final

CCF design around the optimum. This ensures the area is well mapped and understood for an accurate response surface model.

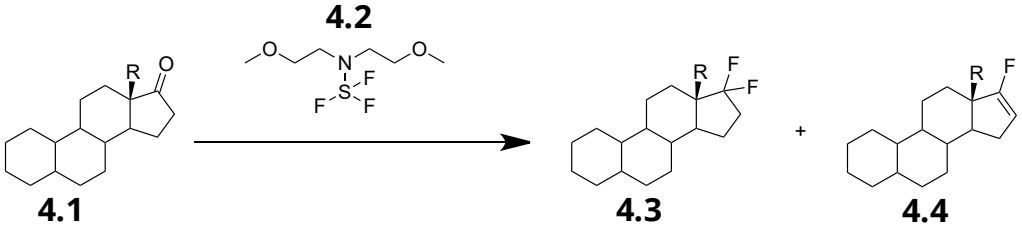
Preliminary algorithm testing was performed employing a simulated chemical reaction, with ordinary differential equations utilised to model the concentration profile for a given set of inputs.

$$\frac{dC_i}{dt} = r = kC_A^n C_B^m \qquad \textbf{(4-4)}$$

$$k = k_0 \exp\left(-E_A/RT\right) \qquad \textbf{(4-5)}$$

E_A and k_0 are the activation energy and pre-exponential factor for the modelled reaction and can be determined experimentally.

For simple systems, the system of ordinary differential equations can be solved analytically, however, more often numerical methods are applied to obtain approximate solutions for the modelled system.[198] In this instance, a previously investigated deoxofluorination reaction was used as a test problem for algorithm development, Scheme 4-1.[199]



Scheme 4-1. Deoxofluorination of a steroid

Full model equations are provided as follows.

$$\frac{d[\mathbf{4.1}]}{dt} = -k_1[\mathbf{4.1}][\mathbf{4.2}] - k_2[\mathbf{4.1}][\mathbf{4.2}] \qquad \textbf{(4-6)}$$

$$\frac{d[\mathbf{4.2}]}{dt} = -k_1[\mathbf{4.1}][\mathbf{4.2}] - k_2[\mathbf{4.1}][\mathbf{4.2}] - k_3[\mathbf{4.1}]^2 \qquad \textbf{(4-7)}$$

$$\frac{d[\mathbf{4.3}]}{dt} = k_1[\mathbf{4.1}][\mathbf{4.2}] \qquad \textbf{(4-8)}$$

$$\frac{d[\mathbf{4.4}]}{dt} = k_2[\mathbf{4.1}][\mathbf{4.2}] \qquad \textbf{(4-9)}$$

$$\frac{d[byproduct]}{dt} = k_3[\mathbf{4.2}]^2 \qquad \textbf{(4-10)}$$

k values for all equations were calculated utilising the Arrhenius equation with the constants provided in Table 4-1.

Table 4-1. Kinetic constants for deoxofluorination of a steroid.

Reaction	$k_0\ (M^{-1}\,min^{-1})$	$E_A\ (kJ\,mol^{-1})$

1	1.85×10^{-4}	62.67
2	1.85×10^{-4}	66.90
3	1.85×10^{-4}	69.59

The target for all the simulated studies was to maximise the yield of product **4.3**. This was then followed by the screening portion of the algorithm which maps the area around the optimum; a 10% reduction from the optimum value was selected as the target for the secondary screening stage. The optimisation was performed with respect to three variables: reaction time, temperature, and concentration of the deoxofluor **4.2**. Bounds for the variables are provided in Table 4-2.

Table 4-2. Optimisation variable bounds for simulated case study**.**

Global Limits	Reaction Time (min)	Temperature (°C)	[B] (M)
Lower	1	80	1
Upper	20	140	3

Results for the simplex and polynomial based hybrid algorithm are provided in Figure 4-5.

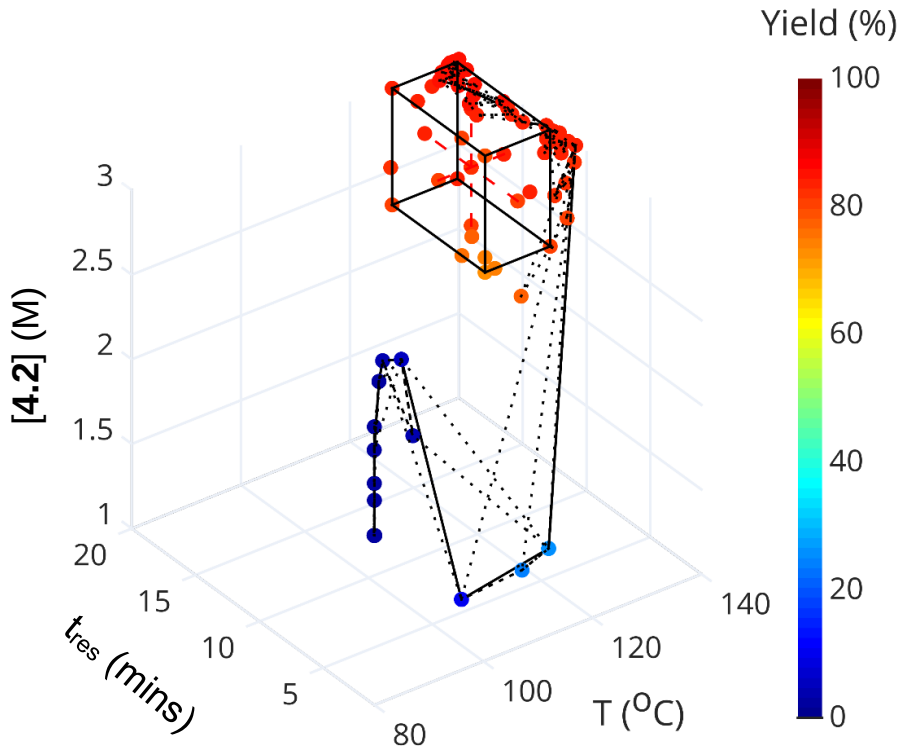


Figure 4-5. Hybrid simplex-DoE optimisation results for the simulated reaction.

Analysing the initial stage of the hybrid algorithm, the algorithm initially moves well through the search domain, however, upon arrival at the optimum region dramatically slows in its progress. This is an inherent feature of the standard implementation of the Nelder-Mead simplex with the algorithm being slow to converge when at/near the optimum region. In such regions, the simplex undergoes multiple contractions until the exit criteria or function evaluation limit are reached. Although modifications exist for the Nelder-Mead simplex, which aim at improving its efficiency, given the inherently local nature of the simplex algorithm effective surrogate modelling of the entire domain may not be possible. In view of this, adopting a global approach to the optimisation was hypothesised to produce a more accurate global surrogate model, as well as providing greater confidence in the optimum that is determined.

For the first development stage, a quadratic surrogate was utilised to model the underlying process. When adopting a global quadratic model, there will be limitations as to the surface the model can effectively map. Higher order models can improve the fit for complex surfaces; however, such models require an increase in the training dataset to produce accurate estimates for the model parameters. Zhou *et al*. have previously investigated the effectiveness of different surrogate modelling approaches for use in optimisation algorithms.[200] The authors compared the effectiveness of different surrogate modelling strategies with varying resource constraints. Gaussian processes were found to be effective for systems which are expensive to evaluate with a limited training dataset. Generally, performing experiments with chemical systems can be considered an expensive to evaluate problem, given the time to run the experiments or the cost of the reagents. Therefore, utilising Gaussian processes as the surrogate model for the screening stage was pursued.

4.2.2 Global Search and Screening

As the simplex search is inherently local in nature, utilising the data for a surrogate model can lead to an inaccurate estimation of the underlying response surface. This is primarily due to the sparseness and non-uniformity of the data across the domain. When utilising models designed for interpolation, having large data sparse regions can lead to unexpected predictions from the underlying surrogate model.

Jeraal *et al*. were able to utilise the data acquired during a global optimisation run to build empirical response surface models of the objective space.[44] The models displayed good agreement with the collected data and

provided a basis to predict performance in previously unexplored regions. The hybrid approach looks to build further upon this by utilising an initial global optimisation stage to later inform targeted response surface modelling around the optimum.

Although not implemented in the published form of the SNOBFIT algorithm, Huyer and Neumaier provide suggestions for stopping criteria based on the algorithms estimate for the current best point. The authors suggest that, should the predicted best value fail to improve for a set number of calls, then the optimisation can be terminated.[100] As with any termination criteria, but especially when utilised with global optimisation algorithms, there is always a risk of premature termination.

As with the local implementation of the optimisation and screening algorithm, a surrogate model was applied for the sequential optimisation of the screening design. In this second iteration, a GP model was used in place of the original polynomial model. The selection of an alternative surrogate model was to enable an improved fit, given the limited dataset that may be present after the global optimisation.[201]

The proposed changes were tested with the aforementioned simulated reaction, the results of which are highlighted in Figure 4-6.

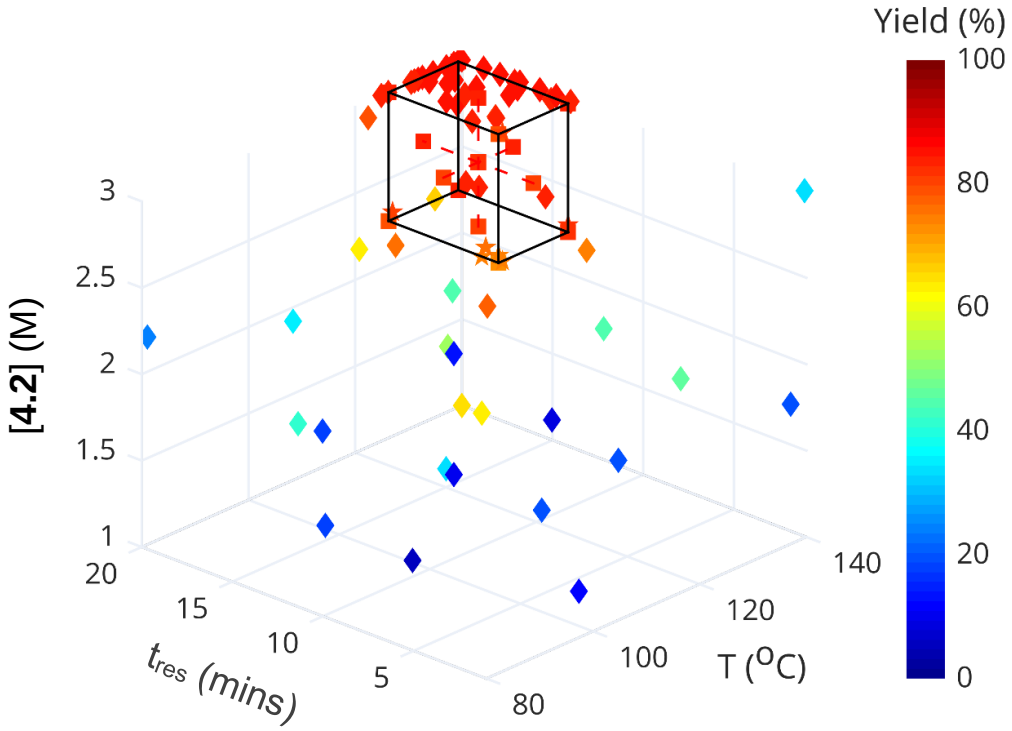


Figure 4-6. Hybrid SNOBFIT-DoE results for the simulated reaction.

Analysing both Figure 4-5 and Figure 4-6 we can see that both algorithms successfully determine the optimum for the system, with both successfully performing the secondary screening task to the same level of precision. One key difference is the improved uniformity of the data across the entire domain, for the second algorithm iteration. Here, the algorithm successfully explores a wider region, not only providing confidence in our optimisation results but also increasing our understanding of the entire experimental domain. This improved uniformity was hypothesised to play a key part in improving the global fit of the secondary surrogate model for the screening portion of the algorithm. Analysis of this hypothesis was performed by comparing the root mean squared error (RMSE) of the two approaches for a ten-sample test set, the results of which are indicated in Table 4-3. The RMSE can be calculated as follows:

$$RMSE = \sqrt{\frac{\sum_i^n \left(y_{pred} - y_{act}\right)^2}{n}} \qquad \textbf{(4-11)}$$

Table 4-3. RMSE comparison between optimisation and surrogate modelling methodologies.

Method	RMSE
Simplex	44.22
SNOBFIT	1.11

The RMSE values clearly indicate a dramatic improvement in the ability of the surrogate model to predict unseen points in the input domain, indicating the efficacy of the GP-SNOBFIT coupling for the screening portion of the algorithm. Efficiency of the optimisation saw a dramatic improvement with the secondary method, with the initial optimisation only requiring 61 experiments compared to the 150 for the initial simplex method. It should be noted that this implementation of simplex was not designed for efficient optimisation and alternatives exist for improved efficiency. However, given the diversity of the points produced by the SNOBFIT algorithm, the advantages of adopting this approach over a local optimisation method outweigh any improvements gained from utilising a more efficient local search.

4.2.3 Algorithm Overview

The finalised hybrid optimisation procedure is summarised in, Figure 4-7. With the noted improvements for the simulated test problem, the

algorithm was applied to an experimental case study, as a further proof of concept.

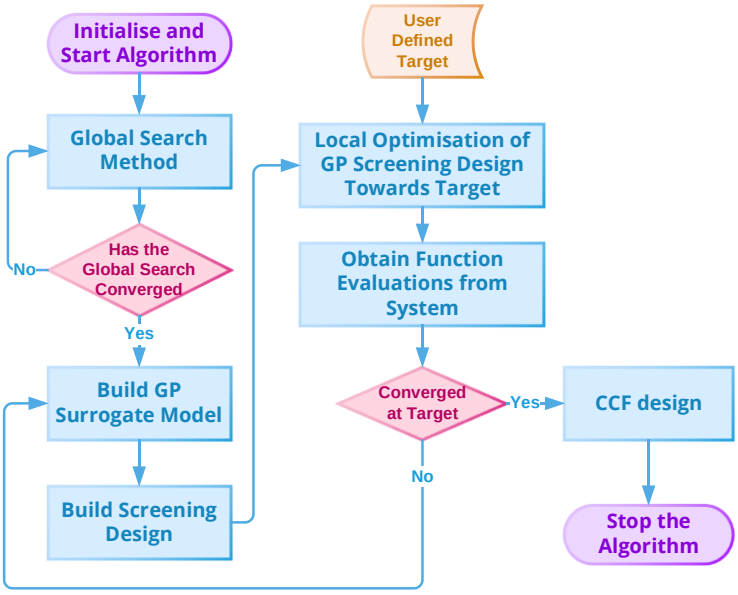
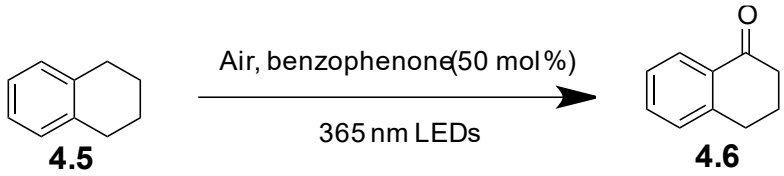


Figure 4-7. Final optimisation procedure for the hybrid algorithm.

4.3 Aerobic Oxidation of Tetralin

The algorithm was applied to the photo-oxidation of tetralin **4.5** to tetralone **4.6**, Scheme 4-2, performed in a series of miniature continuous stirred tank reactors.[202]



Scheme 4-2. Benzophenone catalysed aerobic oxidation of tetralin to tetralone.

The reaction was adapted from the original paper[203] to utilise benzophenone and air at atmospheric pressure (Figure 4-8), in place of the original requirement for a tetra-*n*-butylammonium decatungstate catalyst and oxygen, due to economic and safety concerns. An inline membrane separator (Zaiput[204]) was utilised to separate the biphasic gas-liquid stream prior to sampling. This was performed to ensure only the liquid phase was sampled, with no risk of air entering the HPLC, see section 6.4.2 for further details.

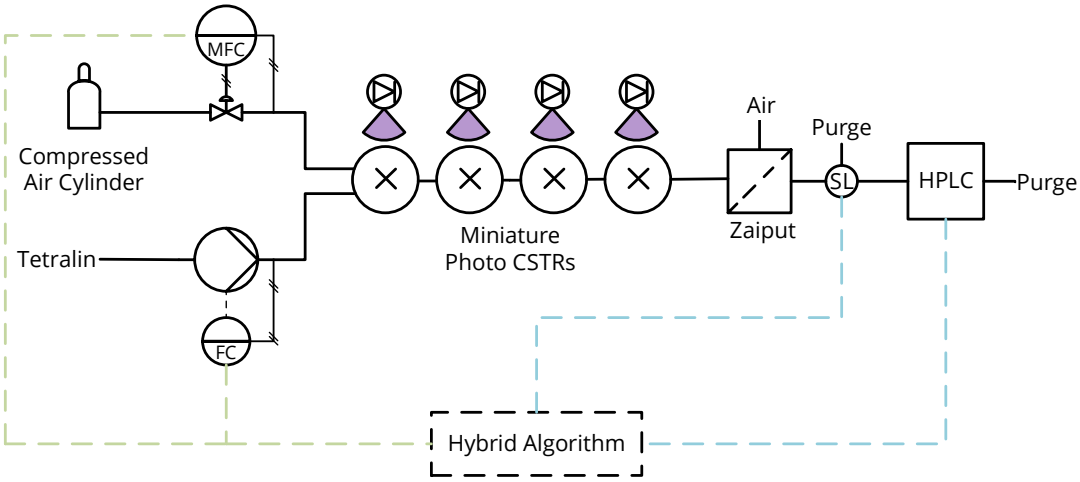


Figure 4-8. Process and instrumentation diagram for the aerobic oxidation case study.

Tetralone yield was set as the maximisation target for the hybrid optimisation, with the bounds of the two variable optimisation indicated in Table 4-4.

Table 4-4. Input domain bounds for automated optimisation of aerobic oxidation reaction.

Global Limits	Residence Time (min)	Oxygen Equivalents
Lower	1	1
Upper	30	5

Due to the bi-phasic nature of the system, residence time calculations were estimated assuming ideal behaviour of the gas phase.

$$V_g = \frac{nRT}{P} \qquad \textbf{(4-12)}$$

The volume of one mole of gas was used to calculate the required volumetric flowrate of gas for the reactor system at given conditions.

$$A{:}F = \frac{V_g * Eq_{O2} * C_{Tetralin}}{\theta_{O2}} \qquad \textbf{(4-13)}$$

$$Q_{air} = \frac{Q_T * A{:}F}{A{:}F + 1} \qquad \textbf{(4-14)}$$

Where θ_{O2} is the fraction of oxygen in the air supply, Eq_{O2} is the required equivalents of oxygen and A:F is the calculated air to fluid ratio required. $Q_T = V_{reactor}/t_{res}$ and is the total volumetric flowrate required to achieve the required residence time. Where assumptions are made relating to the behaviour of the gas it is more appropriate to refer to the calculated time to process a reactor volume being space time, with residence time being the mean time a particle spends in the reactor.

The automated chemical reaction platform was initialised in the morning and terminated the following evening, running continuously overnight. The hybrid algorithm automatically terminated when the final central composite experimental design had been completed; with a total of 61 experiments performed: 38 SNOBFIT experiments; 12 screening experiments and 11 experiments for the final experimental design. The results of each stage are summarised in Figure 4-9, Figure 4-10 and Figure 4-11. Full results can be found in section 6.4.3.

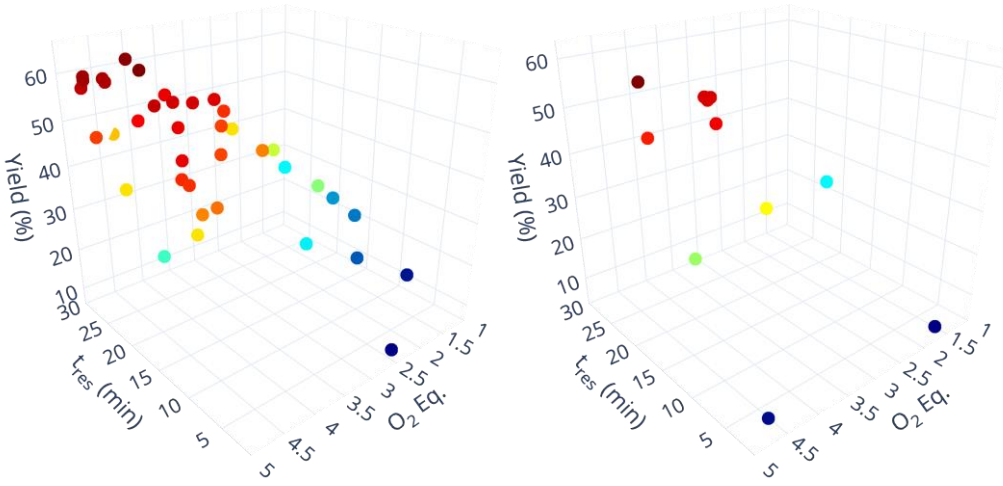


Figure 4-9. Initial optimisation (left) and screening (right) stage results for the tetralin case study.

As GPs provide a value for both the mean function and its associated variance, model confidence intervals can be plotted and utilised. Figure 4-11 highlights the final response surface model for the system, with the associated 95% confidence interval. Furthermore, as the underlying GP surrogate was constructed utilising automatic relevance detection, where each input variable has its own associated length scale, examination of the relative importance of each variable in relation to the objective space can be carried out *a posteriori,* Table 4-5.

The indicated strong dependence on oxygen by the model is verified through examination of the final results. This, coupled with the GP model can be used to inform the user of the underlying behaviours of the system and guide further process design, by offering qualitative and quantitative analysis relating to the system.

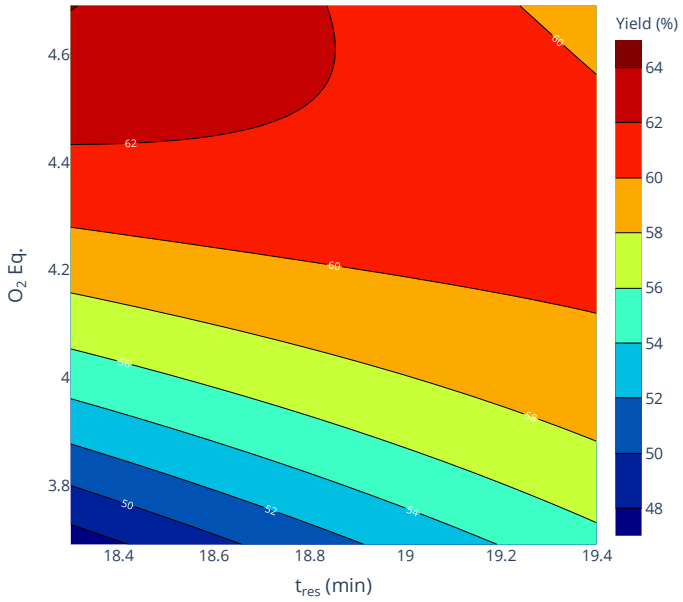


Figure 4-10. Final CCF results for the tetralin case study.

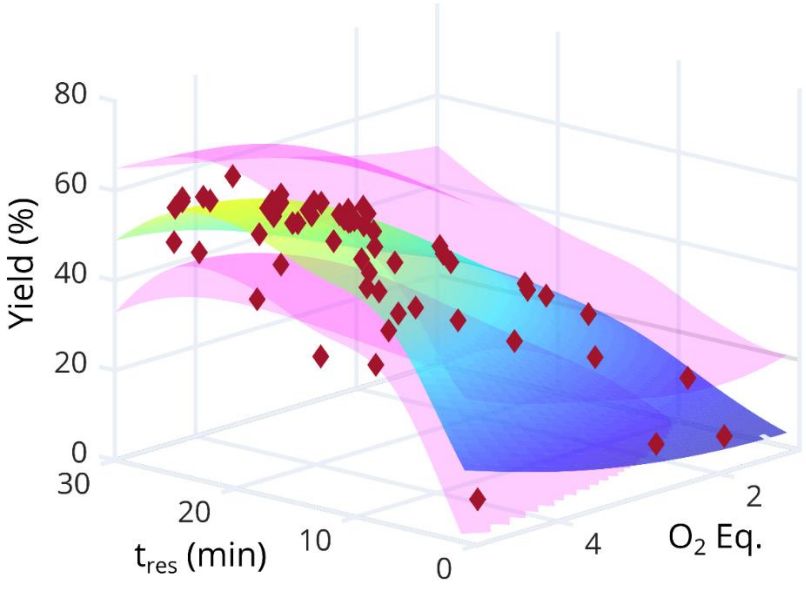


Figure 4-11. Overall GP model utilising entire data set.

Table 4-5. Length scales for each variable taken from the final GP model. A lower value indicates the variable has a large effect on the objective function.

Variable	Length scale
Length scale residence time	13.72
Length scale oxygen equivalents	4.03

For longer residence times the model and the data both indicate a reduction in final tetralone yield. This phenomenon is hypothesised to be due to the overoxidation of tetralone, with HPLC chromatograms also indicating the presence of impurities for longer residence times, Figure 4-12 and Figure 4-13.

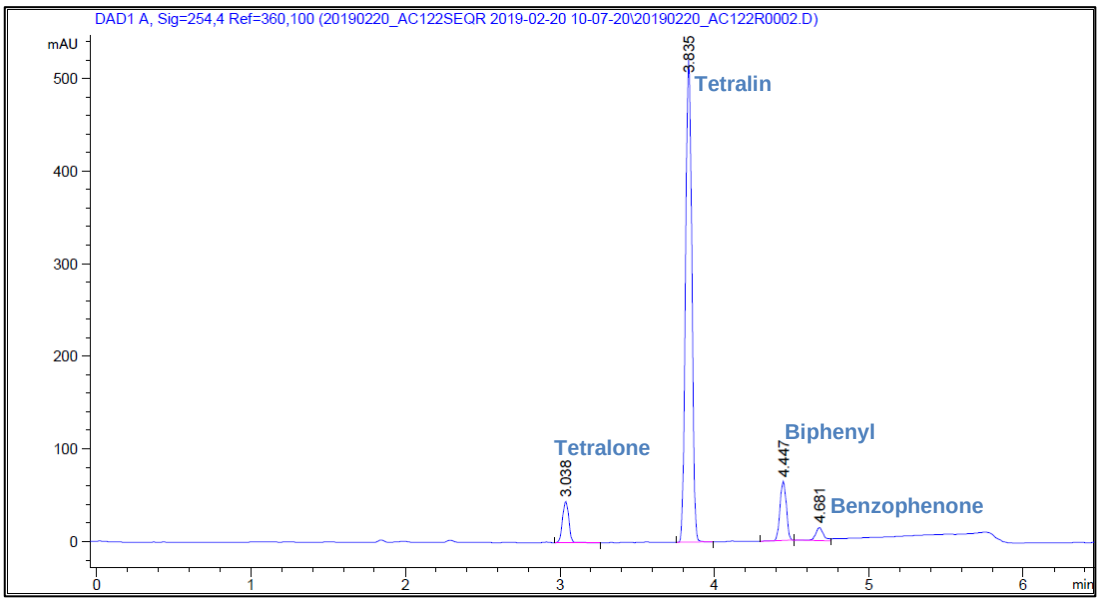


Figure 4-12. Chromatogram, at 254 nm, of the product with the following input conditions (residence time = 8.3 minutes, O_2 eq. = 1.14). Tetralone retention time is approximately 3 minutes.

Overoxidation of the product had previously been observed in the original paper, with its occurrence confirmed for this system through offline analysis by Clayton.[158] Calibration for the overoxidation product was not performed due to its degradation upon separation, however, given the mass balance for the reaction was maintained above 90%, this indicates only small quantities of side product formation.

The benefits of adopting a global optimisation approach are clearly evident, with the diverse array of experiments performed by SNOBFIT. Due to this spread, the underlying surrogate model was able to effectively model the search domain. The use of optimisation algorithms in place of conventional DoE provides greater confidence in the determined optimum region, with corresponding experimental values for the optimum achieved without additional user input.

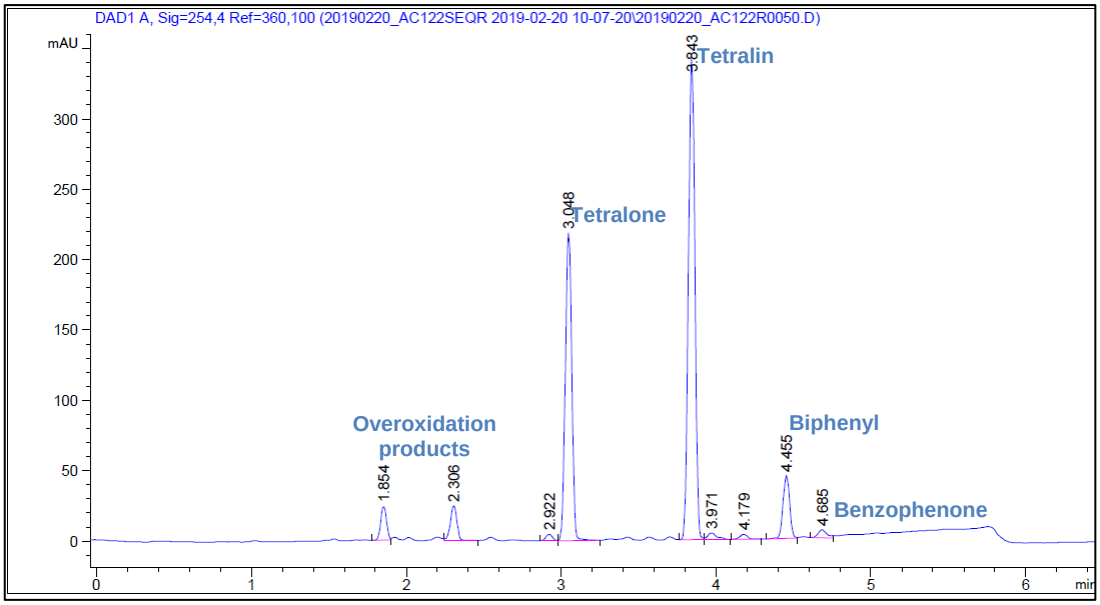


Figure 4-13. Chromatogram, at 254 nm, of the product with the following input conditions (residence time = 19.4 minutes, O_2 eq. = 3.69). Tetralone retention time is approximately 3 minutes.

It is noted that not all points pass through the mean function of the global system model, this was due to the kernel including a noise term. The inclusion of a noise term in the underlying covariance function allows for an estimate to be made relating to the system noise, providing useful feedback in terms of system characterisation. For the photochemical reaction, the underlying model estimates a noise level of approximately 5%.

The final optimal mapped region generated by the algorithm provides the end user with an input domain window that needs to be maintained to ensure the process is kept within their desired operating window ($t_{res} \in [18.35, 19.4], O_2\ Eq. \in [3.7, 4.7]$). Development of a defined operating window ensures effects such as input variability are well understood and can be directly mapped an effect on process outcome; enabling understanding on how this could affect downstream processing units. This highlights a key use case for the approach; in retrospective optimisation and response surface mapping tasks, with the methodology not limited to continuous flow operations.

4.4 Algorithm Improvements

For the experimental case study, difficulties in converging upon the optimum region were observed for the SNOBFIT algorithm. As such it may prove beneficial to investigate alternative global search heuristics to

implement in the initial stage of the hybrid algorithm. Given the system is currently focused on continuous, single objective optimisation, and there is extensive prior literature on their use for expensive optimisation problems, adopting a Bayesian approach is hypothesised to prove beneficial in terms of algorithm efficiency. As such, the proposed modifications utilise the expected improvement algorithm to perform the initial optimisation stage. Termination of the initial optimisation stage proceeds in the same manner as the aforementioned hybrid approach, with termination being triggered by successive iterations without improvement in the current best value. Although this approach can lead to premature termination, if the initial stage has terminated close to that of the global optimum region, the nature of the screening portion of the algorithm will enable the global optimum to be determined.

Analysis of Figure 4-9 indicates the screening portion of the algorithm fails to effectively investigate a diverse region of the objective space. Alongside this, all four points in each sequential screening design gradually converge towards the same point, leading to early termination of the screening portion of the algorithm. Therefore, utilising a rigid screening design structure may not be the best way to investigate non-linear objective domains, with the designs having reduced flexibility to explore the experimental domain.

In selecting the required area in terms of the objective function value, this assumes that the user has ran some initial experiments and is able to select an appropriate mapping area. This knowledge may not always be present prior to performing the optimisation; in lieu of this, it may be more appropriate to define an expected input variable system noise. An approach akin to that adopted by Wyvratt *et al*.[192] may prove a more appropriate methodology for the secondary response surface mapping stage. Utilising this exploratory approach upon a more targeted area will not only improve the efficiency of data collection but lead to a significantly more robust model around the process optimum.

An alternative strategy would be to utilise a purely exploratory Bayesian approach; given that a GP is employed as a surrogate model. This would require an expected noise, relative to the input domain, to be defined with the algorithm sampling in areas where it has the lowest information, or where the model's variance is at its maximum.

$$\mathbf{x}_{next} = argmax_{\mathbf{x}\in\chi}\sigma(\mathbf{x}) \qquad \textbf{(4-15)}$$

Conversely, the original objective function reduction methodology could remain, with the expected improvement acquisition function utilised to guide the search towards the vector of optimal values surrounding the system optimum. For this application, the expected improvement acquisition function was utilised with a trade-off parameter, ζ, which controls the degree of exploration/exploitation:

$$EI(\mathbf{x}) = \left(\mathrm{f}^* - \hat{\mathrm{y}}(\mathbf{x})\right)\Phi(Z(\mathbf{x})) + \sigma(\mathbf{x})\phi(\mathrm{Z}(\mathbf{x})) \quad \textbf{(4-16)}$$

$$Z(\mathbf{x}) = \frac{\mathrm{f}^* - \hat{\mathrm{y}}(\mathbf{x}) - \zeta}{\sigma(\mathbf{x})} \quad \textbf{(4-17)}$$

ζ, was set to a value of 1.5 to ensure adequate exploration of the wider domain. Like the original algorithm, this secondary stage was performed by remapping the objective space to target a user defined reduction. The algorithm utilises the squared difference between the target and the experimental values. Algorithm 4-2 describes the optimisation procedure for both suggested improvements.

The bounds of the sub optimisation for the noise-based approach were calculated as follows:

$$lb_{adj} = \max\left(X_{optim} - X_{optim} * T, lb_{global}\right) \quad \textbf{(4-18)}$$

$$ub_{adj} = \min(X_{optim} + X_{optim} * T, ub_{global}) \quad \textbf{(4-19)}$$

Examination of the methodologies is provided below for the original simulated system. Figure 4-14 and Figure 4-15 highlight example outputs for the noise-based methodology and the targeted expected improvement based methodology, respectively.

Observation of the input noise based results, Figure 4-14, indicate how the optimisation stage of the algorithm efficiently determines the optimal region within minimal experimentation. Upon successful optimisation, determined by the convergence criteria utilised for the SNOBFIT based algorithm, the algorithm proceeds to the exploration stage, in which the bounds of this sub-optimisation were set to be +/- 15% input noise from the optimal input vector. In this stage, the algorithm then expends the remaining experimental budget on minimising the variance within this subdomain, resulting in a hypercube of evaluated points evenly exploring the domain.

Algorithm 4-2: Improved hybrid optimisation and response surface mapping algorithms.

Input: bounds: (lb, ub), objective function: f, mode: M, noise/target value: T, dimension $nvar$, max iterations Max_{it}
begin
$X \leftarrow lhc(nvar, bounds)$
$y \leftarrow f(X)$
$stage \leftarrow optimise$
while iteration $< Max_{it}$ **do**
  if $stage$ **is** $optimise$ **then**
    $mdl \leftarrow fitgp(X, y)$
    $x_{next}, EI_{val} = acquisition(\min y, mdl)$
    $stage \leftarrow convergence\_check(y)$
  else
    if M **is** $noise$ **then**
      $mdl \leftarrow fitgp(X, y)$
      $bounds_{adj} \leftarrow setbounds(bounds, \min y, T)$
      $x_{next}, EX_{val} \leftarrow exploratory(bounds_{adj}, mdl)$
    else M **is** target **then**
      $y_{adj} \leftarrow remap(y, T)$
      $mdl \leftarrow fitgp(X, y_{adj})$
      $x_{next}, EI_{val} = acquisition(\min y_{adj}, mdl)$
  end
  $y_{next} \leftarrow f(x_{next})$
  $X \cup \{x_{next}\}$
  $y \cup \{y_{next}\}$
end
end

Conversely, for the targeted approach, see Figure 4-15, a much wider level of exploration is observed, with the algorithm free of limitations to explore the wider domain in the secondary stage. Like the first algorithm iteration, a target of a 10% reduction from the optimal value was set, which results in this wide spread of data from the optimum value. The use of the expected improvement acquisition function, with additional exploration, enables this wider domain search, allowing for increased understanding of the wider domain, whilst understanding the targeted region around the optimum.

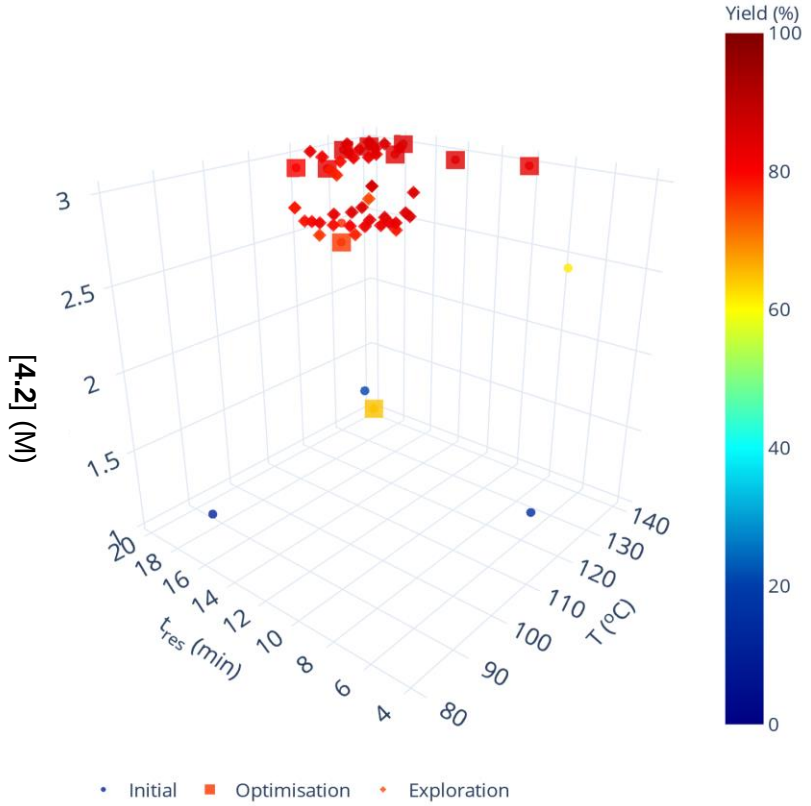


Figure 4-14. Example experimental outcome for the simulated case study utilising expected improvement and maximum standard deviation search.

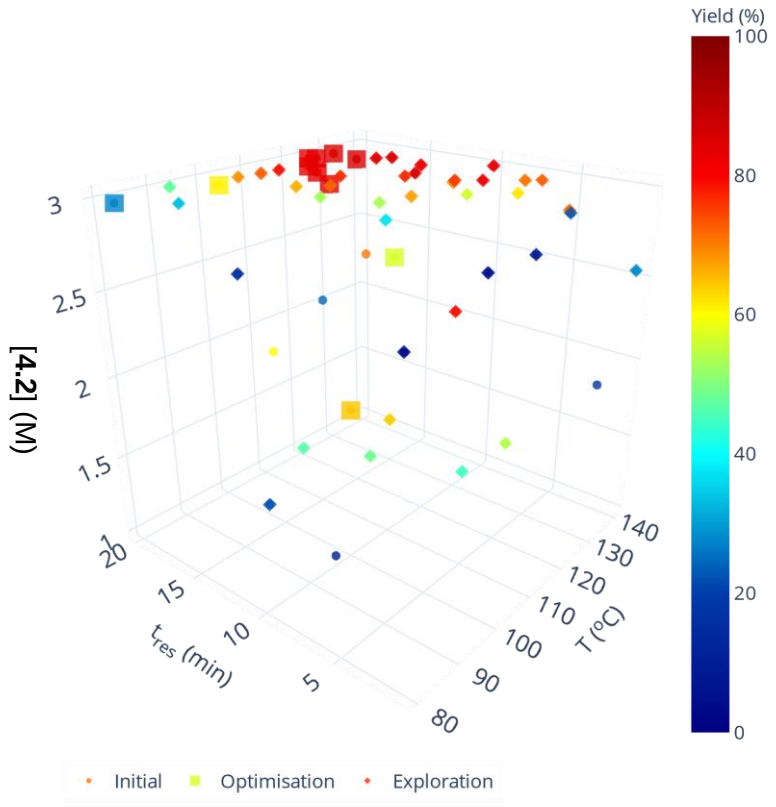


Figure 4-15. Example experimental outcome for the simulated case study utilising expected improvement and targeted expected improvement.

Quantitative comparison between the methodologies was accomplished through two performance tests. Firstly, the performance within the objective space optimal region was verified, Figure 4-16, followed by performance within the average input noise deviation optimal region ($T \in [125,140], t_{res} \in [17,20], [B] \in [2.5,3]$). This allows for comparison of the two newly proposed methodologies with the existing SNOBFIT-DoE based approach. A point weighted RMSE was utilised for the comparison between approaches (4-20) with points weighting applied to account for the different final dataset sizes.

$$RMSE_n = n \times RMSE \qquad \textbf{(4-20)}$$

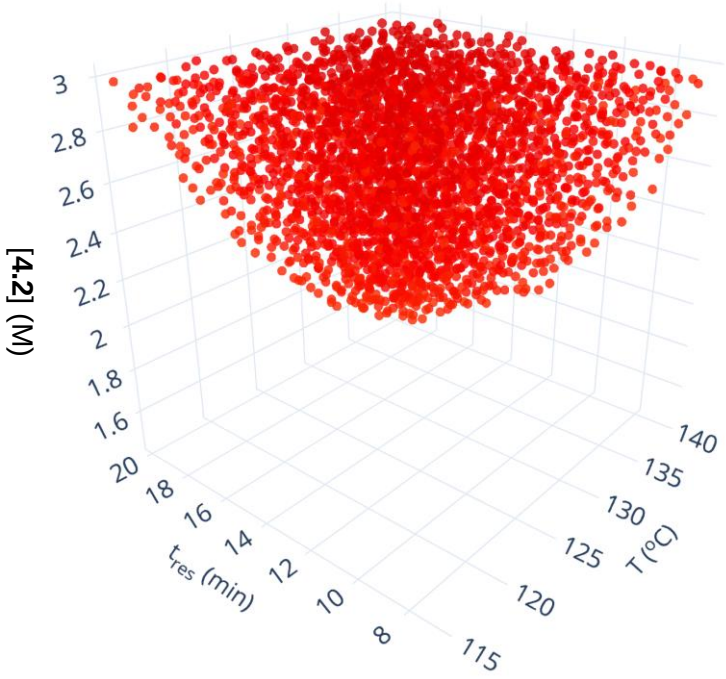


Figure 4-16. Optimal mapping region for the simulated case study, defined as follows: $\mathbf{x} = \mathbf{arg}\,(\boldsymbol{f}(\mathbf{x}) > \boldsymbol{f_{opt}} - \mathbf{10}\%)$. Where $\boldsymbol{f}(\cdot)$ is the output of the ODE function and $\boldsymbol{f_{opt}}$ is the optimum value for the ODE system.

For each region 100 test points were selected with the $RMSE_n$ computed for each run. In total 10 runs were performed, with the results summarised in box plot form, Figure 4-17 and Figure 4-18.

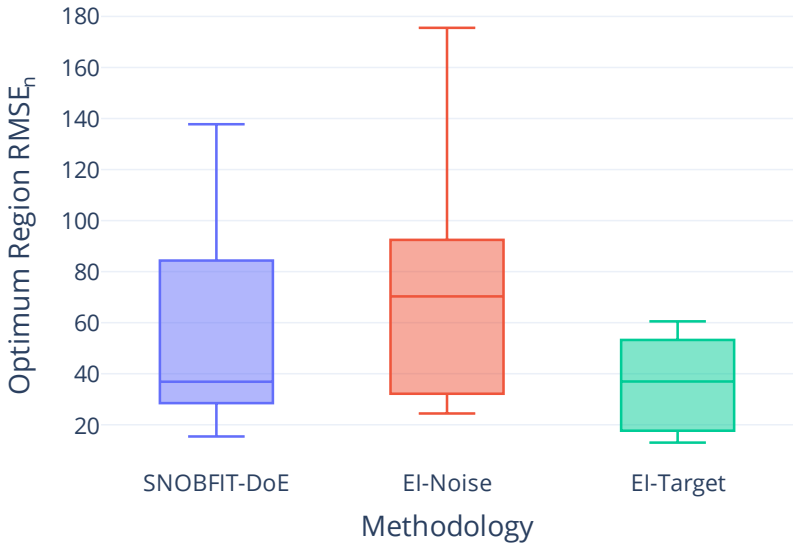


Figure 4-17. Box plot comparison of $\boldsymbol{RMSE_n}$ for the optimum region between different methodologies utilising the deoxofluorination simulated case study.

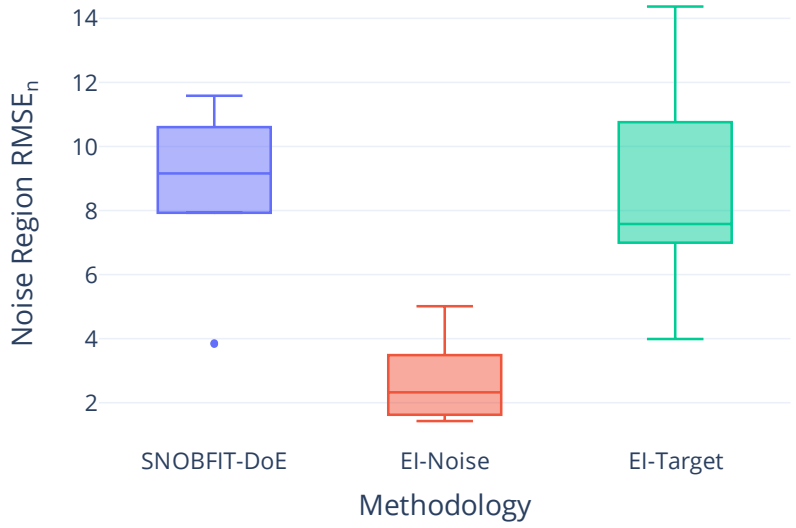


Figure 4-18. Box plot comparison of $\boldsymbol{RMSE_n}$ for the noise region between different methodologies utilising the deoxofluorination simulated case study.

Both figures indicate the proposed methodologies either match or exceed the performance of the originally proposed methodology. In domain specific tasks both methodologies offer improvement in terms of a reduction in the deviation of results, suggesting the algorithms offer robust optimisations and region mapping regardless of the input conditions. The differences

between the two newly proposed methodologies correlate to the intrinsic goals of the algorithms highlighting the potential use cases for the methodologies.

Given the promising results indicated on the simulated system, further comparison with the original algorithm was performed utilising the underlying Gaussian model developed from the applied experimental photochemical example. For this comparison, the underlying model was utilised as the target function to optimise, with the addition of 5% random normally distributed noise to reflect potential experimental noise that may be encountered. Results of the studies are highlighted in Figure 4-19 and Figure 4-20. For both cases, the algorithm was afforded 55 experiments in total, five of which were an initial Latin hypercube. It should be noted that contrary to the results of the experimental optimisation, the Gaussian surrogate predicts two optimum regions at 18- and 24-minutes residence time with the oxygen equivalents set to their upper bound limit.

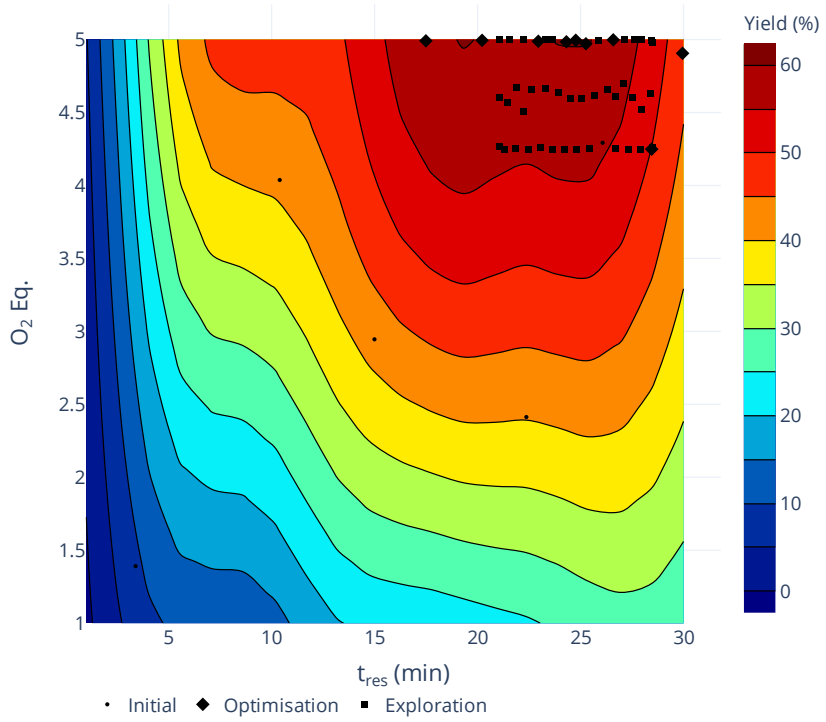


Figure 4-19. Gaussian process surrogate model optimisation of the photochemical system utilising expected improvement and maximum standard deviation search. The contour surface shows the GP surrogate output without the presence of noise.

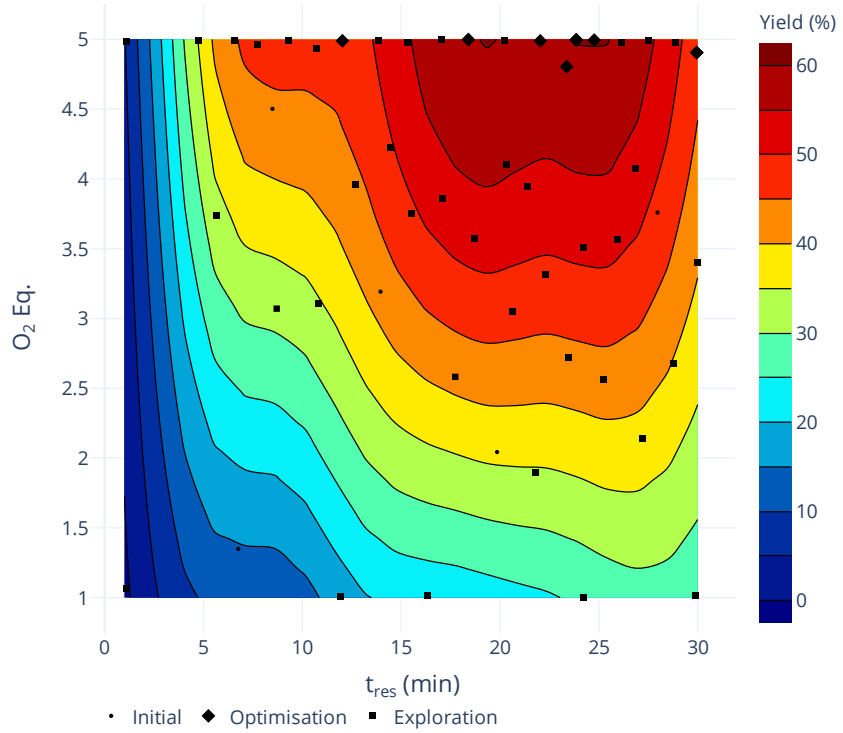


Figure 4-20. Gaussian process surrogate model optimisation of the photochemical system utilising expected improvement and targeted expected improvement. The contour surface shows the GP surrogate output without the presence of noise.

For both suggested improvements, the optimisation was repeated 10 times, with Figure 4-19 and Figure 4-20 highlighting an example output from one of the repeats. In all 10 cases the optimisation algorithms were able to identify the optimum regions within 20 algorithm iterations (25 total experiments); the remaining experiments were utilised for the differing exploration strategies. It was noted that premature termination of the optimisation stage did occur for some runs, however, as the noise bound limits or target objective function value are recalculated for each iteration, this led to the effect of the algorithm correcting for these instances and the exploratory stage determining the process optimum in all cases.

Clearly both suggested alterations offer differing advantages, with the maximum standard deviation approach providing an excellent strategy for input related mapping around the optimum. Adoption of this approach is likely most applicable to the optimisation and investigation of processes in the latter stages of development, where understanding of input noise effects on process key performance indicators (KPIs) is more of a requirement. Alternatively, the targeted expected improvement modification clearly offers superior global search characteristics and may be more suited to investigatory applications in

which wider understanding of the effects of continuous variables on the reaction outcome is required. The suggested approaches provide significant improvement upon the first iteration of the hybrid algorithm, both in terms of initial optimisation efficiency and screening efficiency. As both methodologies present a strategy for non-dynamic experiment mapping, they provide an alternative route to that presented by Wyvratt *et al*.; enabling efficient determination of the optimum and subsequent exploration of the surrounding domain where equipment or reactions do not have the flexibility for dynamic methods.

4.5 Conclusions

When optimising chemical systems, it is not always sufficient to solely determine a point optimum. Understanding of the response surface surrounding an optimum provides an insight into the robustness of the determined optimum when the effects of deviations in inputs are considered.

In this chapter, a novel hybrid algorithm has been developed and applied to a simulated case study, with further application to the optimisation and investigation of the photochemical aerobic oxidation of tetralin. In both cases, the algorithm was able to move efficiently towards the optimum and utilise a surrogate screening methodology to efficiently explore the area surrounding the optimum. The exploratory nature of the global optimisation algorithm, SNOBFIT, was utilised to construct a global surrogate model to inform the screening stage of the algorithm. This ensured maximal efficiency was maintained throughout the screening portion of the algorithm.

Subsequent analysis of the experimental case study highlighted potential efficiency improvements in replacing the global optimisation and screening portions of the algorithm with a Bayesian approach. Given this, two modifications were suggested: one utilising an input-based noise approach and the other adopting an objective function deviation approach, in a similar vein to the original algorithm. Investigation of these two proposed methodologies was performed *in silico* utilising the data collected from the photochemical experimental case study and a previously employed simulated case study. The two potential Bayesian approaches were found to offer significant improvements over the originally proposed algorithm, achieving improvements in both the optimisation and screening stages. This was correlated to improvements in mean and deviation values of the points-weighted root mean squared error for the simulated system, with further qualitative verification utilising a surrogate model of the previously

investigated photochemical reaction. These newly proposed approaches offer differing screening strategies which can be utilised, across a wide array of equipment and reactions, depending upon the current stage of process development.

Chapter 5 Conclusions and Future Work

The presented work in this thesis has focused on investigating the application of optimisation algorithms coupled with automated reaction platforms. Analysis of existing literature highlighted key areas for algorithmic development that would allow for improvements in the efficiency and efficacy of automated reaction optimisation. Given this, the body of work presented has focused on the development of novel algorithms tackling the following issues: 1, optimisation with robust analysis of the objective space surrounding the optimum and, 2, the development of efficient multiobjective optimisation algorithms capable of handling mixed variable inputs (inputs composed of both continuous and discrete variables).

For many applications there is a need to optimise both continuous (temperature, residence time) and discrete (catalyst, ligand, solvent) variables simultaneously to consider the interactive effects between all variable levels. In addition, it is often the case that multiple objectives are required to be optimised. Often, these objectives are conflicting and have no single globally optimum solution. Chapter 2 described the development of a novel mixed variable multi-objective optimisation algorithm, MVMOO. The MVMOO algorithm utilises a novel distance metric, based on Gower similarity, to enable standard multi-objective Bayesian techniques to be applied to mixed variable domains. The algorithm was tested utilising three *in silico* test problems, with comparison to NSGA-II and other readily available optimisation techniques. The proposed algorithm offered notable improvements in terms of efficiency across all three test problems when compared to alternative techniques, with the algorithm achieving comparable results with a significantly reduced experimental budget. The methodology was highlighted to lead to some degree of plateauing with respect to both hypervolume and IGD+ for one of the examined test problems. This could indicate that the algorithm may struggle to escape from locally optimum solutions. Future work examining more exploratory strategies and modifications to the EIM based acquisition may prove beneficial in improving algorithm performance and robustness in the face of poorly defined initial conditions. This could utilise an adaptive expected improvement based approach in a similar manner to the single objective version suggested by Jasrasaria and Pyzer-Knapp.[205] The suggested approach utilises an estimation procedure to balance exploration and exploitation of the optimisation algorithm for each iteration. However, verification on the applicability of the approach to the multi-objective case would be necessary.

Often, in early-stage reaction development, multiple discrete and continuous variables will be screened simultaneously. Normally, designed experiments are used to efficiently screen variables, however, the designs can often overlook interactions between continuous and discrete variables across a range of levels. Algorithmically led approaches can overcome these limitations, while additionally looking at the trade-off between multiple objectives. Chapter 3 details the application of the previously developed MVMOO algorithm on two example case studies. Initially, the bi-objective optimisation of an S_NAr reaction was investigated, with the algorithm set to maximise both *ortho* and *para* products. The optimisation was performed with respect to four continuous parameters and solvent choice. The algorithm was successfully able to identify the trade-off curve between the two competing products highlighting the underlying effect that solvent polarity index had on the outcome of the reaction. The application provided a key example of the effectiveness of multi-objective algorithms in smart data acquisition application for exploration and understanding of an investigated reaction. Following this, a tri objective Sonogashira reaction optimisation was performed, with the effect of palladium ligand complex investigated alongside four continuous parameters. The algorithm successfully identified the optimal trade-off curve for the system, with the front consisting of a single palladium ligand complex. The example highlights a key use case of the algorithm in optimal reaction condition selection for such mixed variable systems in presence of multiple conflicting objectives.

Material consumption is a key criterion for most optimisations, but especially in screening applications, where novel materials with minimal inventory can be investigated. In such situations, it is vital that as much information is maximised for a given reagent quantity. Given this, future work should focus on the application of mixed variable multi-objective algorithms with nanomole-scale high throughput equipment; ensuring minimal material consumption per experiment and maximising the information gain per gram of material used.

Although forming a key aspect of full process operation, the application of algorithms to the automated optimisation of separation processes presents a relatively underdeveloped field. Application of mixed variable multi-objective optimisation algorithms to multi-stage synthesis and separation processes, considering the multifarious effects of both stages, to provide greater process understanding will prove a key avenue for future investigation. Understanding the synergistic effect between reaction variables and downstream processing

units can allow for improved streamlining of the entire process, reducing the need for dramatic changes in the overall process operating conditions when moving from one stage to another.

In optimisation tasks, ensuring robustness in the determined optimum represents a key criterion, especially in chemical system optimisations. Understanding of the objective space surrounding the optimum and how variation in the input variables can affect the stability of the determined process optimum is vital for robust process development and plant scale operation. Chapter 4 detailed the development and application of a novel hybrid optimisation procedure, combining the extensively utilised SNOBFIT algorithm with a custom response surface mapping procedure. The detailed methodology was able to efficiently determine the process optimum whilst accurately mapping the surround area to not only improve user confidence in the determined optimum, but also provide greater understanding of the effects that deviations in input variables can have on the optimised objective. Application of the methodology to a photochemical reaction optimisation provided an experimental case study for the procedure, highlighting the benefits of the approach, with the optimum and surrounding region determined and mapped in an efficient manner. Further improvements to the optimisation procedure, utilising fully Bayesian approaches, were suggested with the improvements applied to a surrogate model of the previous experimental work as well as a simulated case study. For both suggestions, targeted expected improvement, and an input noise based exploration approach, improvements against the original SNOBFIT and DoE approaches were observed, with reduction in the points weighted root mean squared error, even with the addition of 5% normally distributed noise to the surrogate model. Future work looking at experimental applications of the newly proposed procedures will further validate the improvements they provide. Application to downstream process where input variability can have a large effect on separation efficiency would likely greatly improve understanding of the robustness of such processes.[63]

To summarise, algorithms provide an excellent methodology to direct experimentation in an efficient and intelligent manner towards the users end goal. This work has focused on identifying areas in current automated chemical platform optimisation that are lacking in development and proposing novel algorithmically led approaches to tackle these issues. Throughout this work, future developmental areas have been highlighted, which include: (i) development and application of nanomole scale high throughput equipment,

combined with algorithmically led approaches; (ii) application of mixed variable multi-objective methods to multi-stage synthesis, where the synergistic effects of linked stages can be considered. To enable the wider adoption of algorithm-based methodologies, developers must ensure sufficient ease of use and documentation relating to methodology implementation. This not only enables wider adoption of newly developed methodologies for non-expert users, but allows for rapid integration into existing systems, enabling researchers to access improved and more efficient optimisation methodologies. Utilisation of extensively data driven approaches acts as an enabling technology, reducing costs and timelines associated with process development, allowing for wider and cheaper access to resources leading to greater benefit for the end users of products.

Chapter 6 Experimental

6.1 Reactor Platform

A photograph of the automated reactor platform utilised throughout the mixed variable work is shown in Figure 6-1. Reagents were pumped using JASCO PU2080 pumps, with flows combined using Swagelok tee-pieces. A custom reactor block coiled with PTFE tubing (1/16” OD, 1/32” ID) was used in conjunction with a Eurotherm 3200 temperature controller for reaction temperature control. Reagent, ligand, and solvent switching was achieved using a JASCO CO4062 column oven module. The reaction was performed under the desired fixed back pressure provided by an Upchurch Scientific back pressure regulator (BPR). Sampling of the reaction mixture was performed using a VICI Valco 4 port sample loop. Quantitative analysis was provided by an Agilent 1100 series HPLC instrument fitted with a Sigma Ascentis Express C18 reverse phase column (5 cm length, 4.6 mm ID and 2.7 μm particle size).

Figure 6-1. Photograph of the continuous flow reaction platform utilised during the mixed variable optimisation portion of this work.

A custom written MATLAB program was used to control the automated reactor, determine steady state, calculate the responses, and control the inputs and outputs from the optimisation algorithms, Figure 6-2. The automated platform was designed to minimise the consumption of material during the optimisation. This was done through setting flowrates to a minimum

during HPLC analysis and when the reactor was changing temperature set points. For each optimisation point the reactor was allowed a user defined number of reactor volumes to reach steady state, this was set on a case-by-case basis. For batches of experiments, the next experiment was started whilst the analysis of the previous experiment was completed. This minimised the waiting time for analysis between experiments.

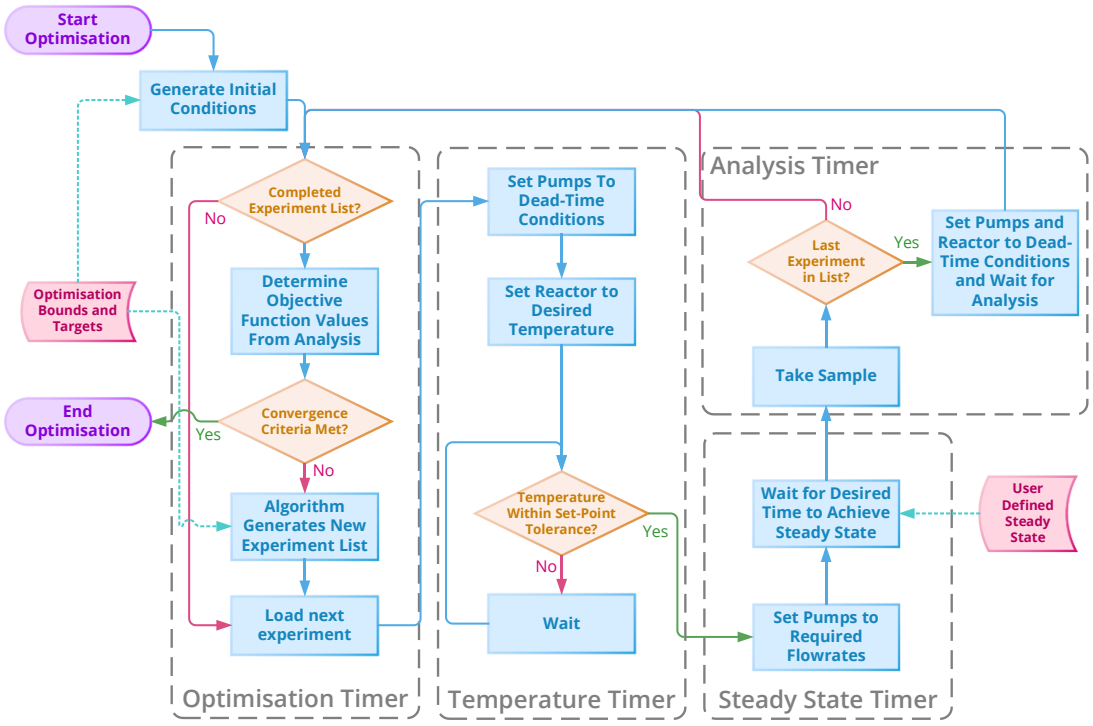


Figure 6-2. Overview of automated platform control code.

6.2 Chapter 2 Materials and Methods

A Python implementation of the algorithm can be found on GitHub: https://github.com/jmanson377/MVMOO.

6.2.1 Optimisation Results

Discrete VLMOP2

Table 6-1. Hypervolume results for the discrete VLMOP2 test problem.

Algorithm	MVMOO		LHS		Random		NSGA-II (Generation)	
Experiment	Mean	STD	Mean	STD	Mean	STD	Mean	STD
1	0.0582	0.0804	0.0573	0.0836	0.0572	0.0774	0.4286	0.0626
2	0.1110	0.0867	0.0940	0.0701	0.0717	0.0733	0.4622	0.0649
3	0.1386	0.0740	0.1272	0.0758	0.1156	0.0840	0.5047	0.0448
4	0.1615	0.0668	0.1523	0.0792	0.1382	0.1023	0.5346	0.0273
5	0.1912	0.0624	0.1665	0.0741	0.1505	0.0934	0.5497	0.0205
6	0.2082	0.0772	0.1949	0.0716	0.1706	0.1017	0.5588	0.0148
7	0.2243	0.0974	0.2202	0.0512	0.1996	0.1152	0.5683	0.0116
8	0.2406	0.1001	0.2353	0.0556	0.2415	0.1178	0.5798	0.0167
9	0.2528	0.0953	0.2355	0.0557	0.2504	0.1120	0.5856	0.0177
10	0.2783	0.0605	0.2507	0.0602	0.2615	0.1239	0.5913	0.0168
11	0.3045	0.0553	0.2793	0.0566	0.2811	0.1110	0.5947	0.0167
12	0.3240	0.0646	0.2934	0.0455	0.2814	0.1111	0.5987	0.0173
13	0.3524	0.0623	0.3062	0.0508	0.2829	0.1115	0.6021	0.0172
14	0.3606	0.0650	0.3251	0.0456	0.2875	0.1123	0.6040	0.0178
15	0.3926	0.0701	0.3284	0.0471	0.3208	0.0793	0.6070	0.0168
16	0.4180	0.0628	0.3430	0.0359	0.3208	0.0793	0.6092	0.0175
17	0.4330	0.0638	0.3455	0.0353	0.3310	0.0785	0.6109	0.0178
18	0.4471	0.0596	0.3456	0.0353	0.3310	0.0785	0.6122	0.0174
19	0.4732	0.0479	0.3471	0.0339	0.3562	0.0794	0.6140	0.0163
20	0.4921	0.0429	0.3669	0.0391	0.3735	0.0691	0.6157	0.0168
21	0.5083	0.0378	0.3689	0.0388	0.3765	0.0681		
22	0.5293	0.0244	0.3753	0.0441	0.3988	0.0618		
23	0.5417	0.0229	0.3836	0.0483	0.4103	0.0605		
24	0.5526	0.0239	0.3836	0.0483	0.4163	0.0654		
25	0.5575	0.0237	0.4012	0.0631	0.4226	0.0689		
26	0.5661	0.0241	0.4068	0.0637	0.4226	0.0688		
27	0.5731	0.0228	0.4089	0.0633	0.4333	0.0625		

28	0.5780	0.0235	0.4209	0.0553	0.4437	0.0646		
29	0.5823	0.0239	0.4212	0.0555	0.4453	0.0655		
30	0.5933	0.0153	0.4221	0.0559	0.4521	0.0580		
31	0.5980	0.0152	0.4238	0.0558	0.4523	0.0581		
32	0.6002	0.0157	0.4275	0.0545	0.4569	0.0520		
33	0.6054	0.0149	0.4352	0.0581	0.4574	0.0518		
34	0.6081	0.0153	0.4376	0.0566	0.4582	0.0516		
35	0.6112	0.0150	0.4379	0.0567	0.4582	0.0516		
36	0.6132	0.0153	0.4420	0.0594	0.4590	0.0499		
37	0.6157	0.0159	0.4530	0.0483	0.4650	0.0450		
38	0.6174	0.0162	0.4530	0.0483	0.4663	0.0453		
39	0.6189	0.0163	0.4610	0.0340	0.4663	0.0453		
40	0.6203	0.0163	0.4619	0.0329	0.4695	0.0429		

Table 6-2. IGD+ results for the discrete VLMOP2 test problem.

Algorithm	**MVMOO**		**LHS**		**Random**		**NSGA-II (Generation)**	
Experiment	Mean	STD	Mean	STD	Mean	STD	Mean	STD
1	0.6997	0.1252	0.7116	0.1408	0.6507	0.1240	0.1327	0.0565
2	0.5650	0.1357	0.6102	0.1176	0.5939	0.1197	0.1137	0.0600
3	0.5204	0.1061	0.5207	0.1571	0.5182	0.1101	0.0799	0.0313
4	0.4879	0.0952	0.4711	0.1551	0.4809	0.1358	0.0597	0.0240
5	0.4028	0.1350	0.4252	0.1211	0.4695	0.1207	0.0504	0.0209
6	0.3498	0.1378	0.3712	0.1098	0.4458	0.1137	0.0472	0.0180
7	0.3429	0.1445	0.3358	0.0925	0.3840	0.1405	0.0439	0.0179
8	0.3196	0.1141	0.3155	0.0994	0.3264	0.1429	0.0386	0.0191
9	0.3101	0.1088	0.3155	0.0994	0.3004	0.1231	0.0363	0.0191
10	0.2854	0.0858	0.3065	0.0956	0.2657	0.1273	0.0338	0.0184
11	0.2532	0.0717	0.2604	0.0783	0.2495	0.0955	0.0323	0.0186
12	0.2460	0.0739	0.2452	0.0703	0.2484	0.0974	0.0304	0.0182
13	0.2169	0.0818	0.2344	0.0765	0.2467	0.1001	0.0291	0.0180
14	0.2131	0.0817	0.2117	0.0555	0.2416	0.1024	0.0284	0.0183
15	0.1918	0.0778	0.2094	0.0565	0.2060	0.0781	0.0270	0.0179
16	0.1796	0.0736	0.1929	0.0326	0.2060	0.0781	0.0260	0.0179
17	0.1684	0.0748	0.1901	0.0340	0.1960	0.0666	0.0248	0.0179
18	0.1560	0.0722	0.1901	0.0340	0.1949	0.0659	0.0243	0.0175
19	0.1348	0.0595	0.1899	0.0341	0.1715	0.0493	0.0232	0.0170
20	0.1244	0.0594	0.1786	0.0338	0.1662	0.0475	0.0222	0.0169

21	0.0975	0.0429	0.1762	0.0321	0.1643	0.0451		
22	0.0780	0.0294	0.1725	0.0339	0.1536	0.0398		
23	0.0684	0.0272	0.1693	0.0319	0.1464	0.0337		
24	0.0631	0.0276	0.1693	0.0319	0.1434	0.0384		
25	0.0616	0.0276	0.1548	0.0345	0.1314	0.0348		
26	0.0578	0.0287	0.1518	0.0317	0.1314	0.0347		
27	0.0522	0.0285	0.1472	0.0332	0.1254	0.0338		
28	0.0455	0.0272	0.1387	0.0318	0.1216	0.0361		
29	0.0438	0.0270	0.1387	0.0318	0.1196	0.0351		
30	0.0354	0.0161	0.1365	0.0348	0.1172	0.0329		
31	0.0291	0.0126	0.1344	0.0339	0.1171	0.0329		
32	0.0275	0.0131	0.1331	0.0349	0.1152	0.0316		
33	0.0255	0.0123	0.1294	0.0397	0.1149	0.0315		
34	0.0240	0.0127	0.1254	0.0345	0.1119	0.0286		
35	0.0226	0.0132	0.1202	0.0327	0.1119	0.0286		
36	0.0219	0.0133	0.1178	0.0340	0.1115	0.0283		
37	0.0209	0.0135	0.1117	0.0279	0.1094	0.0274		
38	0.0204	0.0136	0.1117	0.0279	0.1090	0.0277		
39	0.0198	0.0136	0.1073	0.0265	0.1090	0.0277		
40	0.0193	0.0137	0.1059	0.0260	0.1076	0.0268		

Table 6-3. Function values corresponding to the worst summary surface for the discrete VLMOP2 test problem.

MVMOO		**LHC**		**Random**		**NSGA-II**	
$f_1(x)$	$f_2(x)$	$f_1(x)$	$f_2(x)$	$f_1(x)$	$f_2(x)$	$f_1(x)$	$f_2(x)$
0.014	0.974	0.045	0.983	0.099	0.986	0.244	0.978
0.091	0.943	0.266	0.739	0.407	0.954	0.244	0.978
0.209	0.905	0.576	0.684	0.489	0.811	0.247	0.885
0.238	0.888	0.962	0.317	0.736	0.617	0.300	0.860
0.287	0.867	0.997	0.294	0.905	0.241	0.369	0.827
0.357	0.835	1.210	0.005	0.997	0.206	0.399	0.812
0.421	0.797	1.239	-0.096	1.225	0.052	0.431	0.791
0.456	0.777			1.236	-0.117	0.432	0.791
0.528	0.724					0.482	0.772
0.581	0.682					0.485	0.732
0.625	0.642					0.555	0.726
0.664	0.610					0.555	0.726
0.710	0.545					0.563	0.604

0.736	0.511					0.631	0.570
0.767	0.472					0.681	0.541
0.796	0.420					0.683	0.540
0.848	0.331					0.728	0.535
0.886	0.241					0.739	0.516
0.912	0.176					0.773	0.458
0.930	0.133					0.802	0.411
0.942	0.099					0.816	0.388
0.973	0.017					0.841	0.341
1.151	-0.040					0.842	0.339
1.165	-0.075					0.878	0.279
1.197	-0.134					0.878	0.279
1.197	-0.171					0.891	0.239
1.229	-0.248					0.913	0.231
						0.943	0.171
						0.943	0.171
						0.955	0.129
						0.962	0.085
						0.967	0.038
						0.987	0.010
						1.155	-0.004
						1.157	-0.013
						1.167	-0.049
						1.215	-0.067
						1.223	-0.162
						1.229	-0.203
						1.237	-0.233

ODE Catalytic System

Table 6-4. Hypervolume results for the ODE catalytic test problem.

Algorithm	**MVMOO**		**LHS**		**Random**		**NSGA-II (Generation)**		**TSEMO**	
Experiment	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD
1	1.333	1.270	1.283	0.731	0.829	0.946	4.192	0.671	1.541	0.709
2	2.123	1.147	1.554	0.602	1.095	0.910	4.520	0.563	2.065	0.646
3	2.286	1.053	2.043	0.899	1.320	1.053	5.054	0.486	2.419	0.608
4	3.193	1.023	2.588	0.836	1.544	0.963	5.485	0.473	2.901	1.248

5	3.447	0.823	2.950	0.899	1.544	0.963	6.207	0.504	3.022	1.226
6	3.465	0.795	3.125	0.889	1.793	0.787	6.689	0.483	3.077	1.207
7	3.511	0.819	3.207	0.887	1.861	0.652	7.179	0.401	3.382	1.214
8	3.539	0.818	3.241	0.929	2.207	0.612	7.372	0.380	3.882	1.210
9	3.645	0.722	3.539	0.842	2.224	0.636	7.564	0.397	3.909	1.229
10	3.669	0.774	3.638	0.782	2.235	0.632	7.763	0.389	4.115	1.184
11	3.669	0.774	3.839	0.644	2.278	0.599	7.900	0.344	4.138	1.157
12	3.669	0.774	4.204	0.799	2.278	0.599	8.026	0.240	4.162	1.162
13	3.669	0.774	4.232	0.786	2.292	0.595	8.086	0.266	4.274	1.144
14	3.669	0.774	4.297	0.740	2.425	0.538	8.112	0.256	4.530	1.016
15	3.669	0.774	4.328	0.735	2.582	0.593	8.155	0.254	4.702	0.880
16	3.669	0.774	4.330	0.732	2.595	0.573	8.175	0.250	4.714	0.853
17	3.669	0.774	4.330	0.732	2.603	0.571	8.201	0.212	4.837	0.679
18	3.669	0.774	4.332	0.734	2.606	0.568	8.235	0.195	4.849	0.674
19	3.669	0.774	4.332	0.734	2.637	0.532	8.246	0.185	4.912	0.603
20	3.669	0.774	4.332	0.734	2.637	0.532	8.265	0.165	4.932	0.595
21	3.669	0.774	4.332	0.734	2.671	0.536	8.294	0.142	5.036	0.653
22	3.669	0.774	4.332	0.734	2.672	0.534	8.326	0.109	5.228	0.716
23	3.669	0.774	4.343	0.739	2.715	0.586	8.335	0.098	5.287	0.711
24	3.669	0.774	4.343	0.739	2.766	0.542	8.342	0.092	5.421	0.696
25	3.669	0.774	4.343	0.739	2.766	0.542	8.346	0.087	5.507	0.727
26	3.669	0.774	4.343	0.739	2.766	0.542	8.359	0.065	5.729	0.733
27	3.669	0.774	4.363	0.722	2.774	0.553	8.367	0.065	5.859	0.768
28	3.669	0.774	4.363	0.722	2.934	0.499	8.376	0.052	5.971	0.848
29	3.669	0.774	4.365	0.720	2.934	0.499	8.378	0.051	6.256	0.650
30	3.669	0.774	4.365	0.720	2.968	0.510	8.381	0.048	6.301	0.661
31	3.669	0.774	4.365	0.720	2.993	0.542	8.385	0.044	6.339	0.675
32	3.669	0.774	4.365	0.720	2.993	0.542	8.388	0.043	6.380	0.651
33	3.669	0.774	4.365	0.720	3.009	0.535	8.393	0.040	6.417	0.662
34	3.669	0.774	4.365	0.720	3.017	0.533	8.395	0.038	6.534	0.696
35	3.669	0.774	4.365	0.720	3.065	0.597	8.399	0.034	6.588	0.689
36	3.669	0.774	4.365	0.720	3.103	0.595	8.401	0.034	6.719	0.801
37	3.669	0.774	4.365	0.720	3.134	0.567	8.405	0.029	6.761	0.836
38	3.669	0.774	4.365	0.720	3.134	0.567	8.407	0.028	6.813	0.849
39	3.669	0.774	4.365	0.720	3.146	0.564	8.408	0.028	6.819	0.853
40	3.669	0.774	4.365	0.720	3.146	0.564	8.409	0.027	6.832	0.856
41	3.781	0.793	4.365	0.720	3.162	0.537	8.410	0.026	6.846	0.847
42	4.005	0.975	4.365	0.720	3.182	0.559	8.410	0.026	6.861	0.839

43	4.321	0.984	4.365	0.720	3.182	0.559	8.411	0.026	6.864	0.838
44	4.881	1.376	4.365	0.720	3.191	0.551	8.412	0.025	6.868	0.837
45	5.125	1.302	4.365	0.720	3.191	0.551	8.413	0.024	6.878	0.841
46	5.350	1.312	4.365	0.720	3.191	0.551	8.413	0.024	6.892	0.852
47	5.764	1.025	4.365	0.720	3.191	0.551	8.414	0.023	6.897	0.852
48	5.950	1.043	4.365	0.720	3.254	0.562	8.414	0.023	6.905	0.857
49	6.246	1.175	4.365	0.720	3.254	0.562	8.415	0.023	6.930	0.819
50	6.402	1.208	4.365	0.720	3.308	0.534	8.415	0.022	7.092	0.787
51	6.504	1.199	4.365	0.720	3.308	0.534	8.416	0.021	7.107	0.790
52	6.645	1.064	4.365	0.720	3.324	0.552	8.417	0.018	7.170	0.782
53	6.800	0.906	4.365	0.720	3.336	0.552	8.418	0.018	7.293	0.726
54	6.985	0.786	4.365	0.720	3.336	0.552	8.418	0.017	7.295	0.727
55	7.021	0.773	4.365	0.720	3.338	0.548	8.418	0.016	7.360	0.734
56	7.060	0.770	4.365	0.720	3.342	0.544	8.419	0.015	7.369	0.739
57	7.156	0.790	4.365	0.720	3.342	0.544	8.420	0.014	7.401	0.761
58	7.188	0.786	4.365	0.720	3.347	0.538	8.421	0.014	7.416	0.764
59	7.219	0.776	4.365	0.720	3.351	0.541	8.421	0.014	7.461	0.788
60	7.246	0.772	4.365	0.720	3.351	0.541	8.422	0.013	7.471	0.792
61	7.273	0.768	4.365	0.720	3.427	0.503	8.422	0.013	7.539	0.797
62	7.305	0.741	4.365	0.720	3.427	0.503	8.422	0.013	7.544	0.800
63	7.432	0.630	4.365	0.720	3.442	0.479	8.422	0.013	7.553	0.804
64	7.532	0.640	4.365	0.720	3.445	0.481	8.422	0.013	7.787	0.418
65	7.553	0.633	4.365	0.720	3.450	0.475	8.422	0.013	7.814	0.425
66	7.564	0.624	4.365	0.720	3.450	0.475	8.423	0.013	7.823	0.430
67	7.585	0.622	4.365	0.720	3.476	0.512	8.423	0.013	7.831	0.431
68	7.597	0.618	4.365	0.720	3.516	0.449	8.423	0.012	7.836	0.433
69	7.606	0.618	4.365	0.720	3.516	0.449	8.423	0.012	7.858	0.382
70	7.688	0.628	4.365	0.720	3.516	0.449	8.423	0.012	7.864	0.382
71	7.739	0.638	4.365	0.720	3.517	0.450	8.423	0.012	7.868	0.383
72	7.790	0.645	4.365	0.720	3.600	0.382	8.424	0.012	7.871	0.385
73	7.796	0.646	4.365	0.720	3.603	0.380	8.424	0.012	7.877	0.386
74	7.799	0.647	4.365	0.720	3.603	0.380	8.424	0.013	7.881	0.386
75	7.805	0.640	4.365	0.720	3.603	0.380	8.424	0.012	7.883	0.387
76	7.813	0.629	4.365	0.720	3.603	0.380	8.424	0.012	7.936	0.254
77	7.822	0.616	4.365	0.720	3.633	0.417	8.424	0.012	7.980	0.181
78	7.838	0.586	4.365	0.720	3.637	0.412	8.424	0.012	8.026	0.160
79	7.870	0.494	4.365	0.720	3.643	0.419	8.424	0.012	8.027	0.160
80	7.890	0.443	4.365	0.720	3.644	0.418	8.424	0.012	8.030	0.159

81	7.904	0.410	4.365	0.720	3.644	0.418	8.424	0.012	8.054	0.144
82	7.911	0.397	4.365	0.720	3.645	0.418	8.424	0.012	8.056	0.143
83	8.003	0.152	4.365	0.720	3.645	0.418	8.424	0.012	8.059	0.143
84	8.011	0.143	4.365	0.720	3.645	0.417	8.424	0.012	8.062	0.144
85	8.022	0.131	4.365	0.720	3.647	0.416	8.425	0.012	8.090	0.119
86	8.027	0.130	4.365	0.720	3.647	0.416			8.093	0.116
87	8.032	0.129	4.365	0.720	3.649	0.417			8.096	0.115
88	8.068	0.122	4.365	0.720	3.649	0.417			8.099	0.114
89	8.072	0.121	4.365	0.720	3.670	0.401			8.129	0.112
90	8.077	0.120	4.365	0.720	3.670	0.401			8.131	0.111
91	8.079	0.119	4.365	0.720	3.670	0.401			8.134	0.112
92	8.081	0.119	4.365	0.720	3.670	0.401			8.136	0.112
93	8.085	0.120	4.365	0.720	3.671	0.400			8.138	0.113
94	8.087	0.118	4.365	0.720	3.689	0.403			8.140	0.114
95	8.091	0.117	4.365	0.720	3.689	0.403			8.141	0.114
96	8.094	0.114	4.365	0.720	3.791	0.433			8.143	0.114
97	8.095	0.113	4.365	0.720	3.802	0.444			8.161	0.119
98	8.100	0.111	4.365	0.720	3.802	0.444			8.162	0.119
99	8.102	0.109	4.365	0.720	3.807	0.448			8.164	0.118
100	8.103	0.108	4.365	0.720	3.807	0.448			8.165	0.117
101	8.107	0.108	4.365	0.720	3.807	0.448			8.186	0.087
102	8.110	0.108	4.365	0.720	3.807	0.448			8.199	0.088
103	8.111	0.107	4.365	0.720	3.814	0.454			8.218	0.085
104	8.112	0.107	4.365	0.720	3.814	0.454			8.219	0.085
105	8.112	0.107	4.365	0.720	3.814	0.454			8.221	0.085
106	8.114	0.108	4.365	0.720	3.814	0.454			8.222	0.085
107	8.115	0.108	4.365	0.720	3.814	0.454			8.223	0.085
108	8.117	0.108	4.365	0.720	3.814	0.454			8.224	0.085
109	8.118	0.108	4.365	0.720	3.898	0.399			8.225	0.085
110	8.118	0.108	4.365	0.720	3.898	0.399			8.226	0.085
111	8.119	0.108	4.365	0.720	4.020	0.346			8.227	0.084
112	8.121	0.106	4.365	0.720	4.025	0.349			8.228	0.084
113	8.121	0.106	4.365	0.720	4.025	0.349			8.230	0.083
114	8.123	0.105	4.365	0.720	4.032	0.335			8.247	0.059
115	8.123	0.106	4.365	0.720	4.032	0.335			8.248	0.059
116	8.124	0.105	4.365	0.720	4.046	0.350			8.249	0.059
117	8.125	0.105	4.365	0.720	4.046	0.350			8.250	0.059
118	8.126	0.105	4.365	0.720	4.062	0.361			8.251	0.060

119	8.128	0.104	4.365	0.720	4.070	0.367			8.252	0.059
120	8.129	0.105	4.365	0.720	4.070	0.367			8.254	0.058
121	8.130	0.105	4.365	0.720	4.075	0.362			8.255	0.058
122	8.130	0.105	4.365	0.720	4.079	0.359			8.265	0.050
123	8.131	0.105	4.365	0.720	4.091	0.338			8.266	0.050
124	8.131	0.105	4.365	0.720	4.091	0.338			8.273	0.048
125	8.131	0.105	4.365	0.720	4.091	0.338			8.274	0.048

Table 6-5. IGD+ results for the ODE catalytic test problem.

Algorithm	**MVMOO**		**LHS**		**Random**		**NSGA-II (Generation)**		**TSEMO**	
Experiment	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD
1	1.984	0.895	2.042	0.534	2.512	0.998	0.508	0.193	1.725	0.634
2	1.437	0.705	1.792	0.417	2.087	0.928	0.421	0.110	1.424	0.510
3	1.344	0.628	1.456	0.501	1.900	0.918	0.328	0.094	1.221	0.399
4	0.811	0.453	1.148	0.477	1.792	0.804	0.242	0.079	1.105	0.514
5	0.713	0.301	0.977	0.450	1.792	0.804	0.145	0.062	1.049	0.414
6	0.706	0.285	0.875	0.341	1.545	0.568	0.098	0.050	1.011	0.374
7	0.694	0.294	0.857	0.334	1.478	0.387	0.057	0.031	0.897	0.391
8	0.693	0.294	0.849	0.343	1.251	0.297	0.045	0.025	0.693	0.379
9	0.636	0.256	0.725	0.298	1.247	0.305	0.035	0.025	0.688	0.380
10	0.628	0.266	0.679	0.273	1.247	0.305	0.025	0.023	0.615	0.366
11	0.628	0.266	0.608	0.206	1.240	0.298	0.019	0.018	0.592	0.338
12	0.628	0.266	0.516	0.213	1.240	0.298	0.013	0.011	0.586	0.341
13	0.628	0.266	0.506	0.205	1.225	0.291	0.011	0.011	0.568	0.341
14	0.628	0.266	0.493	0.188	1.165	0.280	0.010	0.011	0.472	0.286
15	0.628	0.266	0.487	0.185	1.072	0.281	0.008	0.010	0.421	0.250
16	0.628	0.266	0.487	0.185	1.056	0.263	0.008	0.010	0.410	0.220
17	0.628	0.266	0.487	0.185	1.056	0.262	0.006	0.008	0.373	0.140
18	0.628	0.266	0.487	0.185	1.056	0.262	0.005	0.007	0.369	0.139
19	0.628	0.266	0.487	0.185	1.049	0.256	0.005	0.006	0.345	0.106
20	0.628	0.266	0.487	0.185	1.049	0.256	0.004	0.005	0.341	0.105
21	0.628	0.266	0.487	0.185	1.044	0.258	0.003	0.004	0.319	0.114
22	0.628	0.266	0.487	0.185	1.044	0.258	0.002	0.003	0.290	0.126
23	0.628	0.266	0.486	0.186	1.035	0.269	0.002	0.003	0.281	0.126
24	0.628	0.266	0.486	0.186	1.015	0.251	0.002	0.002	0.262	0.112
25	0.628	0.266	0.486	0.186	1.015	0.251	0.001	0.002	0.253	0.116
26	0.628	0.266	0.486	0.186	1.015	0.251	0.001	0.001	0.225	0.111

27	0.628	0.266	0.483	0.185	1.014	0.252	0.001	0.001	0.214	0.117
28	0.628	0.266	0.483	0.185	0.949	0.259	0.001	0.001	0.205	0.123
29	0.628	0.266	0.483	0.184	0.949	0.259	0.001	0.001	0.158	0.072
30	0.628	0.266	0.483	0.184	0.936	0.263	0.001	0.001	0.152	0.073
31	0.628	0.266	0.483	0.184	0.926	0.273	0.001	0.001	0.147	0.075
32	0.628	0.266	0.483	0.184	0.926	0.273	0.001	0.001	0.141	0.073
33	0.628	0.266	0.483	0.184	0.924	0.273	0.000	0.001	0.137	0.073
34	0.628	0.266	0.483	0.184	0.922	0.273	0.000	0.000	0.126	0.076
35	0.628	0.266	0.483	0.184	0.903	0.299	0.000	0.000	0.120	0.076
36	0.628	0.266	0.483	0.184	0.880	0.300	0.000	0.000	0.112	0.082
37	0.628	0.266	0.483	0.184	0.859	0.282	0.000	0.000	0.111	0.083
38	0.628	0.266	0.483	0.184	0.859	0.282	0.000	0.000	0.107	0.083
39	0.628	0.266	0.483	0.184	0.858	0.282	0.000	0.000	0.107	0.083
40	0.628	0.266	0.483	0.184	0.858	0.282	0.000	0.000	0.106	0.083
41	0.606	0.274	0.483	0.184	0.847	0.263	0.000	0.000	0.104	0.083
42	0.561	0.292	0.483	0.184	0.845	0.267	0.000	0.000	0.103	0.083
43	0.485	0.305	0.483	0.184	0.845	0.267	0.000	0.000	0.103	0.083
44	0.414	0.342	0.483	0.184	0.836	0.257	0.000	0.000	0.102	0.083
45	0.341	0.257	0.483	0.184	0.836	0.257	0.000	0.000	0.101	0.083
46	0.310	0.243	0.483	0.184	0.836	0.257	0.000	0.000	0.100	0.083
47	0.209	0.126	0.483	0.184	0.836	0.257	0.000	0.000	0.099	0.084
48	0.197	0.128	0.483	0.184	0.812	0.262	0.000	0.000	0.098	0.084
49	0.174	0.138	0.483	0.184	0.812	0.262	0.000	0.000	0.094	0.075
50	0.159	0.139	0.483	0.184	0.780	0.246	0.000	0.000	0.080	0.073
51	0.150	0.137	0.483	0.184	0.780	0.246	0.000	0.000	0.079	0.073
52	0.124	0.096	0.483	0.184	0.775	0.252	0.000	0.000	0.074	0.073
53	0.104	0.071	0.483	0.184	0.775	0.252	0.000	0.000	0.063	0.070
54	0.086	0.056	0.483	0.184	0.775	0.252	0.000	0.000	0.063	0.070
55	0.084	0.055	0.483	0.184	0.774	0.250	0.000	0.000	0.059	0.071
56	0.079	0.052	0.483	0.184	0.774	0.250	0.000	0.000	0.059	0.071
57	0.072	0.051	0.483	0.184	0.774	0.250	0.000	0.000	0.058	0.072
58	0.068	0.050	0.483	0.184	0.773	0.249	0.000	0.000	0.057	0.072
59	0.066	0.048	0.483	0.184	0.773	0.250	0.000	0.000	0.055	0.073
60	0.063	0.049	0.483	0.184	0.773	0.250	0.000	0.000	0.055	0.073
61	0.060	0.049	0.483	0.184	0.731	0.227	0.000	0.000	0.052	0.074
62	0.058	0.047	0.483	0.184	0.731	0.227	0.000	0.000	0.051	0.074
63	0.050	0.040	0.483	0.184	0.717	0.203	0.000	0.000	0.051	0.074
64	0.046	0.041	0.483	0.184	0.716	0.204	0.000	0.000	0.028	0.024

65	0.045	0.041	0.483	0.184	0.714	0.201	0.000	0.000	0.027	0.025
66	0.044	0.040	0.483	0.184	0.714	0.201	0.000	0.000	0.026	0.025
67	0.043	0.040	0.483	0.184	0.708	0.208	0.000	0.000	0.026	0.025
68	0.042	0.039	0.483	0.184	0.688	0.176	0.000	0.000	0.025	0.025
69	0.041	0.039	0.483	0.184	0.688	0.176	0.000	0.000	0.023	0.019
70	0.038	0.040	0.483	0.184	0.688	0.176	0.000	0.000	0.023	0.019
71	0.036	0.040	0.483	0.184	0.688	0.176	0.000	0.000	0.023	0.020
72	0.034	0.041	0.483	0.184	0.645	0.158	0.000	0.000	0.022	0.020
73	0.034	0.040	0.483	0.184	0.645	0.158	0.000	0.000	0.022	0.020
74	0.033	0.041	0.483	0.184	0.645	0.158	0.000	0.000	0.022	0.020
75	0.032	0.039	0.483	0.184	0.645	0.158	0.000	0.000	0.022	0.020
76	0.031	0.038	0.483	0.184	0.645	0.158	0.000	0.000	0.018	0.008
77	0.031	0.036	0.483	0.184	0.639	0.164	0.000	0.000	0.016	0.004
78	0.030	0.034	0.483	0.184	0.639	0.163	0.000	0.000	0.014	0.004
79	0.026	0.024	0.483	0.184	0.639	0.164	0.000	0.000	0.014	0.004
80	0.025	0.022	0.483	0.184	0.639	0.163	0.000	0.000	0.014	0.003
81	0.024	0.020	0.483	0.184	0.639	0.163	0.000	0.000	0.013	0.003
82	0.024	0.019	0.483	0.184	0.639	0.163	0.000	0.000	0.013	0.003
83	0.019	0.006	0.483	0.184	0.639	0.163	0.000	0.000	0.013	0.003
84	0.019	0.006	0.483	0.184	0.639	0.163	0.000	0.000	0.012	0.003
85	0.018	0.005	0.483	0.184	0.638	0.163	0.000	0.000	0.012	0.003
86	0.018	0.005	0.483	0.184	0.638	0.163			0.011	0.002
87	0.017	0.004	0.483	0.184	0.638	0.163			0.011	0.002
88	0.016	0.004	0.483	0.184	0.638	0.163			0.011	0.002
89	0.016	0.004	0.483	0.184	0.637	0.162			0.010	0.002
90	0.015	0.004	0.483	0.184	0.637	0.162			0.010	0.002
91	0.015	0.004	0.483	0.184	0.637	0.162			0.010	0.002
92	0.015	0.004	0.483	0.184	0.637	0.162			0.010	0.002
93	0.015	0.003	0.483	0.184	0.637	0.162			0.010	0.003
94	0.014	0.003	0.483	0.184	0.634	0.162			0.010	0.002
95	0.014	0.003	0.483	0.184	0.634	0.162			0.010	0.002
96	0.014	0.003	0.483	0.184	0.600	0.165			0.010	0.002
97	0.014	0.003	0.483	0.184	0.599	0.167			0.009	0.003
98	0.014	0.003	0.483	0.184	0.599	0.167			0.009	0.002
99	0.013	0.002	0.483	0.184	0.599	0.167			0.009	0.002
100	0.013	0.002	0.483	0.184	0.599	0.167			0.009	0.002
101	0.013	0.002	0.483	0.184	0.599	0.167			0.008	0.002
102	0.013	0.002	0.483	0.184	0.599	0.167			0.008	0.002

103	0.013	0.003	0.483	0.184	0.598	0.167			0.008	0.001
104	0.013	0.003	0.483	0.184	0.598	0.167			0.008	0.001
105	0.013	0.003	0.483	0.184	0.598	0.167			0.007	0.001
106	0.012	0.003	0.483	0.184	0.598	0.167			0.007	0.001
107	0.012	0.003	0.483	0.184	0.598	0.167			0.007	0.001
108	0.012	0.003	0.483	0.184	0.598	0.167			0.007	0.001
109	0.012	0.003	0.483	0.184	0.556	0.138			0.007	0.001
110	0.012	0.003	0.483	0.184	0.556	0.138			0.007	0.001
111	0.012	0.003	0.483	0.184	0.509	0.113			0.007	0.001
112	0.012	0.002	0.483	0.184	0.509	0.113			0.007	0.001
113	0.011	0.002	0.483	0.184	0.509	0.113			0.007	0.001
114	0.011	0.002	0.483	0.184	0.508	0.111			0.006	0.001
115	0.011	0.002	0.483	0.184	0.508	0.111			0.006	0.001
116	0.011	0.002	0.483	0.184	0.506	0.113			0.006	0.001
117	0.011	0.002	0.483	0.184	0.506	0.113			0.006	0.001
118	0.011	0.002	0.483	0.184	0.500	0.117			0.006	0.001
119	0.011	0.002	0.483	0.184	0.499	0.118			0.006	0.001
120	0.011	0.002	0.483	0.184	0.499	0.118			0.006	0.001
121	0.011	0.002	0.483	0.184	0.498	0.117			0.006	0.001
122	0.010	0.002	0.483	0.184	0.497	0.116			0.006	0.001
123	0.010	0.002	0.483	0.184	0.496	0.113			0.006	0.001
124	0.010	0.002	0.483	0.184	0.495	0.112			0.005	0.001
125	0.010	0.002	0.483	0.184	0.495	0.112			0.005	0.001

Table 6-6. Function values corresponding to the worst summary surface for the ODE catalytic test problem.

MVMOO		**LHC**		**Random**		**NSGA-II**		**TSEMO**	
Yield	STY	Yield	STY	Yield	STY	Yield	STY	Yield	STY
0.564	9.415	0.557	3.846	0.368	5.557	0.596	9.946	0.599	9.799
0.584	8.818	0.802	3.026	0.459	3.489	0.600	9.857	0.609	9.548
0.649	8.759	0.899	2.932	0.675	3.039	0.607	9.754	0.624	9.374
0.662	8.641	0.982	2.003	0.822	2.870	0.610	9.685	0.635	9.081
0.668	8.515			0.899	2.435	0.611	9.668	0.648	8.937
0.681	8.332			0.903	1.905	0.618	9.553	0.650	8.748
0.696	7.950					0.622	9.479	0.663	8.689
0.704	7.850					0.625	9.420	0.671	8.502
0.715	7.637					0.629	9.364	0.687	8.253
0.745	7.188					0.632	9.303	0.703	7.973

0.745	6.950					0.640	9.173	0.707	7.720
0.774	6.667					0.642	9.133	0.722	7.620
0.803	6.113					0.645	9.078	0.727	7.353
0.811	5.912					0.648	9.027	0.741	7.275
0.813	5.631					0.650	8.999	0.748	7.007
0.826	5.403					0.661	8.791	0.764	6.856
0.835	5.315					0.665	8.724	0.773	6.650
0.843	5.259					0.669	8.664	0.773	6.567
0.847	5.168					0.670	8.634	0.783	6.491
0.865	4.915					0.677	8.523	0.792	6.325
0.873	4.664					0.680	8.453	0.805	6.085
0.896	4.243					0.688	8.324	0.814	5.917
0.903	4.089					0.691	8.273	0.823	5.668
0.913	3.925					0.695	8.191	0.831	5.598
0.919	3.624					0.699	8.126	0.835	5.415
0.926	3.521					0.702	8.078	0.843	5.338
0.934	3.419					0.710	7.936	0.852	5.183
0.939	3.350					0.710	7.936	0.859	5.034
0.941	3.209					0.713	7.872	0.870	4.837
0.945	3.144					0.714	7.857	0.878	4.627
0.946	3.051					0.718	7.791	0.887	4.475
0.953	2.966					0.720	7.746	0.894	4.353
0.957	2.836					0.725	7.671	0.902	4.171
0.966	2.623					0.727	7.623	0.903	4.052
0.970	2.513					0.731	7.544	0.910	3.849
0.971	2.281					0.735	7.473	0.914	3.825
0.976	2.279					0.741	7.383	0.923	3.693
0.978	2.086					0.747	7.257	0.923	3.500
0.979	2.051					0.755	7.112	0.932	3.452
0.985	1.950					0.758	7.066	0.939	3.325
0.986	1.910					0.762	6.995	0.943	3.097
0.987	1.808					0.768	6.879	0.951	3.029
0.991	1.697					0.771	6.823	0.957	2.862
						0.775	6.748	0.967	2.615
						0.777	6.706	0.970	2.507
						0.780	6.654	0.970	2.409
						0.783	6.593	0.971	2.361
						0.788	6.511	0.976	2.290

						0.791	6.456	0.981	2.134
						0.793	6.416	0.983	1.999
						0.796	6.354	0.985	1.940
						0.801	6.261	0.985	1.905
						0.801	6.261	0.988	1.852
						0.806	6.169	0.988	1.839
						0.807	6.158		
						0.812	6.067		
						0.814	6.019		
						0.818	5.955		
						0.821	5.885		
						0.823	5.856		
						0.828	5.768		
						0.833	5.662		
						0.833	5.658		
						0.837	5.588		
						0.840	5.533		
						0.844	5.451		
						0.850	5.338		
						0.857	5.198		
						0.860	5.133		
						0.861	5.123		
						0.864	5.058		
						0.870	4.934		
						0.873	4.874		
						0.876	4.815		
						0.880	4.728		
						0.886	4.611		
						0.889	4.548		
						0.894	4.438		
						0.898	4.364		
						0.898	4.363		
						0.902	4.280		
						0.903	4.247		
						0.905	4.204		
						0.907	4.166		
						0.909	4.118		
						0.911	4.084		

						0.915	4.003		
						0.919	3.904		
						0.921	3.854		
						0.925	3.779		
						0.925	3.762		
						0.933	3.595		
						0.934	3.570		
						0.937	3.482		
						0.938	3.477		
						0.942	3.381		
						0.946	3.279		
						0.949	3.188		
						0.954	3.079		
						0.956	3.020		
						0.960	2.906		
						0.962	2.844		
						0.965	2.769		
						0.966	2.737		
						0.969	2.635		
						0.970	2.617		
						0.972	2.545		
						0.973	2.530		
						0.975	2.443		
						0.976	2.426		
						0.979	2.312		
						0.981	2.261		
						0.983	2.165		
						0.984	2.117		
						0.987	2.000		
						0.988	1.980		
						0.990	1.880		
						0.991	1.828		
						0.993	1.698		
						0.994	1.659		

Fuel Injector

Table 6-7. Hypervolume results for the fuel injector design test problem.

Algorithm	MVMOO		LHS		Random		NSGA-II (Generation)	
Experiment	Mean	STD	Mean	STD	Mean	STD	Mean	STD
1	0.153	0.084	0.224	0.181	0.206	0.121	1.182	0.053
2	0.319	0.141	0.252	0.176	0.307	0.076	1.227	0.045
3	0.402	0.111	0.319	0.160	0.437	0.101	1.293	0.039
4	0.465	0.087	0.372	0.127	0.521	0.106	1.347	0.040
5	0.499	0.071	0.423	0.115	0.568	0.131	1.387	0.029
6	0.526	0.077	0.433	0.115	0.612	0.101	1.416	0.032
7	0.551	0.083	0.446	0.106	0.642	0.102	1.437	0.028
8	0.593	0.086	0.468	0.112	0.663	0.107	1.443	0.029
9	0.628	0.070	0.497	0.115	0.684	0.099	1.457	0.031
10	0.648	0.053	0.511	0.102	0.714	0.117	1.465	0.035
11	0.670	0.080	0.550	0.104	0.749	0.109	1.472	0.032
12	0.696	0.077	0.565	0.106	0.773	0.118	1.479	0.033
13	0.718	0.079	0.593	0.104	0.791	0.096	1.484	0.026
14	0.728	0.078	0.603	0.088	0.821	0.096	1.486	0.022
15	0.776	0.059	0.619	0.077	0.831	0.098	1.489	0.023
16	0.812	0.087	0.632	0.072	0.850	0.096	1.488	0.030
17	0.830	0.082	0.635	0.071	0.862	0.096	1.491	0.030
18	0.836	0.078	0.646	0.062	0.873	0.096	1.495	0.023
19	0.861	0.065	0.652	0.058	0.883	0.099	1.499	0.022
20	0.904	0.051	0.659	0.055	0.891	0.102	1.496	0.021
21	1.020	0.064	0.668	0.054	0.902	0.098	1.503	0.023
22	1.090	0.052	0.677	0.051	0.910	0.104	1.501	0.018
23	1.124	0.043	0.686	0.050	0.920	0.090	1.502	0.018
24	1.171	0.045	0.700	0.043	0.921	0.090	1.497	0.015
25	1.208	0.043	0.703	0.039	0.931	0.080	1.501	0.020
26	1.247	0.034	0.721	0.058	0.935	0.081	1.500	0.017
27	1.282	0.039	0.727	0.060	0.939	0.083	1.504	0.020
28	1.298	0.037	0.741	0.063	0.944	0.080	1.501	0.019
29	1.322	0.030	0.741	0.063	0.952	0.078	1.506	0.024
30	1.345	0.022	0.757	0.069	0.961	0.069	1.510	0.022
31	1.362	0.022	0.768	0.063	0.965	0.069	1.510	0.017
32	1.380	0.021	0.778	0.064	0.971	0.074	1.509	0.020

33	1.394	0.020	0.778	0.064	0.982	0.068	1.511	0.016
34	1.403	0.022	0.789	0.064	0.993	0.069	1.510	0.017
35	1.415	0.024	0.792	0.059	1.007	0.060	1.516	0.012
36	1.427	0.026	0.803	0.063	1.008	0.058	1.518	0.016
37	1.434	0.027	0.808	0.065	1.011	0.060	1.517	0.014
38	1.440	0.027	0.811	0.065	1.014	0.062	1.513	0.018
39	1.448	0.022	0.817	0.056	1.014	0.062	1.510	0.014
40	1.454	0.023	0.819	0.056	1.016	0.062	1.513	0.012
41	1.461	0.021	0.826	0.051	1.017	0.063	1.513	0.014
42	1.467	0.023	0.836	0.033	1.020	0.059	1.511	0.014
43	1.472	0.022	0.837	0.033	1.022	0.060	1.512	0.014
44	1.477	0.022	0.839	0.031	1.028	0.057	1.508	0.014
45	1.484	0.016	0.847	0.031	1.030	0.058	1.510	0.012
46	1.487	0.017	0.850	0.032	1.036	0.061	1.513	0.014
47	1.493	0.021	0.851	0.030	1.047	0.058	1.515	0.014
48	1.496	0.021	0.856	0.026	1.059	0.059	1.511	0.017
49	1.500	0.019	0.865	0.029	1.062	0.060	1.509	0.023
50	1.504	0.019	0.867	0.026	1.064	0.060	1.514	0.017
51	1.509	0.020	0.882	0.036	1.073	0.052	1.517	0.018
52	1.512	0.018	0.884	0.035	1.078	0.045	1.516	0.014
53	1.513	0.017	0.889	0.033	1.078	0.045	1.514	0.014
54	1.515	0.017	0.894	0.035	1.080	0.046	1.508	0.017
55	1.517	0.016	0.906	0.042	1.080	0.046	1.512	0.021
56	1.519	0.016	0.909	0.041	1.081	0.047	1.510	0.018
57	1.522	0.016	0.910	0.041	1.081	0.047	1.513	0.021
58	1.524	0.016	0.913	0.040	1.088	0.044	1.516	0.011
59	1.526	0.016	0.933	0.056	1.088	0.044	1.514	0.009
60	1.530	0.010	0.946	0.049	1.092	0.041	1.517	0.013
61	1.533	0.010	0.958	0.047	1.095	0.041	1.524	0.018
62	1.534	0.010	0.961	0.042	1.098	0.038	1.520	0.014
63	1.536	0.010	0.969	0.050	1.099	0.039	1.522	0.016
64	1.538	0.010	0.977	0.044	1.099	0.039	1.523	0.013
65	1.538	0.010	0.979	0.042	1.100	0.038	1.521	0.020
66	1.541	0.010	0.980	0.043	1.101	0.039	1.519	0.018
67	1.543	0.011	0.982	0.043	1.102	0.039	1.518	0.021
68	1.544	0.011	0.990	0.041	1.107	0.044	1.517	0.025
69	1.545	0.011	0.992	0.044	1.111	0.043	1.522	0.020
70	1.547	0.011	1.001	0.037	1.112	0.043	1.519	0.020

71	1.549	0.010	1.008	0.038	1.113	0.044	1.518	0.017
72	1.550	0.011	1.010	0.037	1.113	0.044	1.519	0.020
73	1.551	0.011	1.010	0.038	1.116	0.050	1.518	0.015
74	1.552	0.011	1.016	0.038	1.118	0.050	1.518	0.015
75	1.553	0.011	1.017	0.039	1.118	0.051	1.512	0.019
76	1.554	0.011	1.037	0.046	1.119	0.051	1.512	0.019
77	1.556	0.010	1.044	0.030	1.119	0.052	1.515	0.019
78	1.558	0.009	1.051	0.029	1.123	0.048	1.513	0.017
79	1.558	0.009	1.060	0.027	1.124	0.046	1.516	0.015
80	1.560	0.009	1.063	0.026	1.127	0.046	1.517	0.015
81	1.561	0.009	1.065	0.022	1.127	0.046		
82	1.562	0.009	1.077	0.034	1.129	0.046		
83	1.563	0.009	1.085	0.032	1.130	0.045		
84	1.564	0.009	1.092	0.027	1.131	0.044		
85	1.565	0.009	1.101	0.036	1.133	0.044		
86	1.566	0.010	1.103	0.037	1.136	0.045		
87	1.566	0.010	1.103	0.037	1.136	0.045		
88	1.567	0.010	1.108	0.036	1.137	0.045		
89	1.568	0.010	1.116	0.031	1.137	0.045		
90	1.569	0.010	1.121	0.031	1.138	0.045		
91	1.571	0.010	1.129	0.032	1.140	0.044		
92	1.572	0.010	1.139	0.035	1.140	0.044		
93	1.573	0.010	1.145	0.032	1.140	0.044		
94	1.574	0.010	1.146	0.033	1.140	0.044		
95	1.575	0.010	1.156	0.031	1.143	0.043		
96	1.576	0.010	1.163	0.037	1.144	0.043		
97	1.577	0.010	1.164	0.037	1.148	0.045		
98	1.577	0.010	1.170	0.030	1.148	0.045		
99	1.578	0.010	1.175	0.027	1.153	0.042		
100	1.579	0.010	1.181	0.024	1.156	0.040		

Table 6-8. IGD+ results for the fuel injector design test problem.

Algorithm	**MVMOO**		**LHS**		**Random**		**NSGA-II (Generation)**	
Experiment	Mean	STD	Mean	STD	Mean	STD	Mean	STD
1	0.688	0.134	0.624	0.217	0.644	0.155	0.122	0.012
2	0.492	0.136	0.595	0.200	0.525	0.066	0.111	0.012

3	0.408	0.087	0.498	0.140	0.409	0.097	0.094	0.013
4	0.376	0.066	0.446	0.103	0.365	0.094	0.080	0.010
5	0.353	0.063	0.403	0.088	0.343	0.091	0.072	0.009
6	0.339	0.063	0.400	0.088	0.320	0.078	0.067	0.011
7	0.321	0.059	0.394	0.082	0.307	0.078	0.063	0.009
8	0.296	0.056	0.385	0.083	0.298	0.073	0.064	0.008
9	0.278	0.041	0.372	0.083	0.285	0.059	0.062	0.008
10	0.272	0.037	0.361	0.063	0.272	0.061	0.062	0.008
11	0.269	0.040	0.337	0.066	0.261	0.057	0.059	0.007
12	0.264	0.040	0.331	0.068	0.253	0.051	0.059	0.008
13	0.256	0.038	0.314	0.069	0.243	0.043	0.058	0.006
14	0.254	0.039	0.304	0.056	0.230	0.040	0.058	0.005
15	0.238	0.031	0.298	0.053	0.230	0.040	0.057	0.006
16	0.233	0.035	0.292	0.051	0.225	0.039	0.058	0.008
17	0.228	0.033	0.291	0.051	0.224	0.039	0.057	0.008
18	0.226	0.028	0.284	0.042	0.222	0.037	0.057	0.006
19	0.214	0.023	0.281	0.040	0.219	0.040	0.056	0.006
20	0.211	0.023	0.278	0.037	0.214	0.034	0.056	0.005
21	0.183	0.016	0.272	0.037	0.210	0.031	0.056	0.006
22	0.166	0.016	0.267	0.036	0.207	0.031	0.057	0.005
23	0.153	0.017	0.262	0.032	0.205	0.028	0.058	0.005
24	0.140	0.014	0.255	0.034	0.205	0.029	0.060	0.005
25	0.130	0.007	0.254	0.032	0.199	0.024	0.061	0.007
26	0.122	0.008	0.246	0.037	0.197	0.025	0.062	0.007
27	0.113	0.007	0.245	0.037	0.196	0.025	0.061	0.007
28	0.106	0.005	0.240	0.039	0.193	0.027	0.061	0.005
29	0.101	0.005	0.240	0.039	0.190	0.022	0.060	0.007
30	0.097	0.004	0.234	0.042	0.185	0.022	0.060	0.007
31	0.092	0.004	0.229	0.040	0.185	0.022	0.060	0.006
32	0.087	0.004	0.225	0.040	0.183	0.021	0.060	0.005
33	0.083	0.004	0.225	0.040	0.181	0.023	0.060	0.006
34	0.079	0.004	0.219	0.038	0.178	0.023	0.061	0.006
35	0.075	0.004	0.218	0.037	0.170	0.015	0.061	0.006
36	0.073	0.003	0.215	0.038	0.170	0.014	0.060	0.006
37	0.070	0.003	0.214	0.038	0.168	0.015	0.059	0.006
38	0.067	0.002	0.212	0.037	0.166	0.017	0.060	0.006
39	0.064	0.002	0.209	0.034	0.166	0.017	0.061	0.007
40	0.062	0.002	0.209	0.034	0.165	0.016	0.061	0.007

41	0.060	0.003	0.208	0.033	0.165	0.016	0.060	0.008
42	0.058	0.002	0.199	0.017	0.165	0.016	0.061	0.007
43	0.057	0.003	0.199	0.017	0.165	0.016	0.061	0.006
44	0.056	0.003	0.199	0.017	0.162	0.016	0.062	0.006
45	0.054	0.003	0.196	0.015	0.161	0.017	0.062	0.006
46	0.053	0.002	0.196	0.015	0.160	0.017	0.061	0.007
47	0.052	0.002	0.195	0.014	0.156	0.014	0.059	0.008
48	0.050	0.002	0.194	0.013	0.154	0.014	0.059	0.006
49	0.049	0.002	0.192	0.015	0.153	0.015	0.059	0.008
50	0.048	0.002	0.191	0.015	0.152	0.015	0.058	0.007
51	0.047	0.002	0.190	0.014	0.150	0.014	0.058	0.007
52	0.046	0.002	0.190	0.014	0.149	0.013	0.057	0.006
53	0.045	0.002	0.189	0.013	0.149	0.013	0.058	0.007
54	0.045	0.002	0.188	0.014	0.147	0.010	0.060	0.007
55	0.044	0.002	0.184	0.017	0.147	0.010	0.060	0.008
56	0.043	0.002	0.183	0.017	0.147	0.011	0.062	0.007
57	0.042	0.002	0.183	0.017	0.147	0.011	0.061	0.009
58	0.042	0.002	0.182	0.017	0.145	0.011	0.060	0.006
59	0.041	0.002	0.177	0.016	0.145	0.011	0.061	0.006
60	0.041	0.002	0.174	0.013	0.145	0.011	0.061	0.006
61	0.040	0.002	0.171	0.015	0.144	0.010	0.060	0.007
62	0.039	0.001	0.169	0.014	0.144	0.010	0.060	0.006
63	0.039	0.001	0.166	0.017	0.144	0.010	0.060	0.008
64	0.038	0.001	0.164	0.018	0.144	0.010	0.059	0.007
65	0.038	0.001	0.164	0.017	0.144	0.010	0.059	0.005
66	0.037	0.001	0.164	0.018	0.144	0.010	0.059	0.005
67	0.037	0.001	0.163	0.018	0.143	0.010	0.059	0.005
68	0.036	0.002	0.161	0.017	0.142	0.010	0.059	0.007
69	0.036	0.001	0.160	0.017	0.140	0.010	0.060	0.007
70	0.036	0.001	0.158	0.015	0.140	0.010	0.059	0.007
71	0.035	0.002	0.156	0.016	0.140	0.010	0.060	0.006
72	0.035	0.002	0.156	0.016	0.140	0.010	0.060	0.007
73	0.034	0.002	0.156	0.016	0.138	0.012	0.061	0.006
74	0.034	0.002	0.154	0.016	0.138	0.012	0.060	0.005
75	0.034	0.002	0.154	0.017	0.138	0.012	0.062	0.005
76	0.033	0.002	0.153	0.017	0.138	0.012	0.061	0.006
77	0.033	0.001	0.152	0.017	0.138	0.013	0.061	0.006
78	0.033	0.001	0.150	0.015	0.137	0.012	0.061	0.007

79	0.033	0.001	0.148	0.016	0.136	0.011	0.060	0.007
80	0.032	0.001	0.148	0.015	0.135	0.011	0.060	0.006
81	0.032	0.001	0.148	0.015	0.135	0.011		
82	0.032	0.001	0.145	0.015	0.135	0.011		
83	0.031	0.001	0.144	0.015	0.134	0.011		
84	0.031	0.001	0.142	0.015	0.134	0.011		
85	0.031	0.001	0.140	0.015	0.134	0.011		
86	0.030	0.001	0.140	0.015	0.133	0.012		
87	0.030	0.001	0.140	0.015	0.133	0.012		
88	0.030	0.001	0.139	0.015	0.133	0.012		
89	0.030	0.001	0.136	0.015	0.133	0.012		
90	0.030	0.001	0.135	0.014	0.133	0.012		
91	0.029	0.001	0.133	0.013	0.133	0.012		
92	0.029	0.001	0.133	0.014	0.133	0.012		
93	0.029	0.001	0.132	0.014	0.133	0.012		
94	0.029	0.001	0.132	0.014	0.133	0.012		
95	0.029	0.001	0.130	0.014	0.132	0.012		
96	0.028	0.001	0.129	0.015	0.131	0.012		
97	0.028	0.001	0.129	0.015	0.130	0.012		
98	0.028	0.001	0.126	0.009	0.130	0.012		
99	0.028	0.001	0.125	0.009	0.130	0.012		
100	0.028	0.001	0.123	0.008	0.129	0.011		

Table 6-9. Function values corresponding to the worst summary surface for the fuel injector design problem.

MVMOO				**LHS**				**Random**				**NSGA-II**			
$f_1(x)$	$f_2(x)$	$f_3(x)$	$f_4(x)$	$f_1(x)$	$f_2(x)$	$f_3(x)$	$f_4(x)$	$f_1(x)$	$f_2(x)$	$f_3(x)$	$f_4(x)$	$f_1(x)$	$f_2(x)$	$f_3(x)$	$f_4(x)$
-0.266	0.575	1.150	-0.154	-0.276	0.928	1.168	-0.158	-0.217	0.580	1.288	-0.043	-0.279	1.012	1.042	-0.197
-0.261	0.665	1.016	-0.184	-0.231	0.886	1.054	-0.123	-0.163	0.923	1.048	-0.073	-0.279	0.995	1.074	-0.188
-0.256	0.748	0.841	-0.229	-0.204	1.152	0.838	-0.146	-0.142	0.545	0.982	-0.068	-0.275	1.055	0.914	-0.229
-0.254	0.503	1.206	-0.123	-0.169	0.912	0.846	-0.094	-0.125	0.748	0.859	-0.075	-0.267	1.187	0.764	-0.261
-0.248	0.431	1.329	-0.083	-0.163	0.819	0.940	-0.057	-0.106	0.991	0.792	-0.018	-0.267	1.222	0.759	-0.262
-0.240	0.681	0.929	-0.178	-0.112	0.463	1.053	0.045	-0.084	0.855	0.809	-0.027	-0.265	1.159	0.750	-0.263
-0.235	0.593	1.052	-0.136	-0.102	0.841	0.830	-0.049	-0.082	0.852	0.858	0.037	-0.261	1.106	0.823	-0.235
-0.234	0.847	0.716	-0.236	-0.091	1.133	0.704	-0.023	-0.067	0.533	1.305	0.186	-0.258	0.775	1.411	-0.070
-0.234	0.626	0.967	-0.159	-0.059	0.801	0.723	-0.020	-0.063	0.697	0.768	-0.014	-0.253	0.721	1.509	-0.039
-0.221	0.368	1.402	-0.024	-0.040	0.668	0.900	0.108	-0.014	0.609	0.891	0.095	-0.246	0.689	1.571	-0.014
-0.208	0.755	0.755	-0.187	-0.015	0.741	0.683	0.074	-0.007	0.865	0.657	0.080	-0.241	0.379	1.433	-0.046
-0.202	0.476	1.173	-0.054	0.013	0.814	0.650	0.019	0.014	1.076	0.607	0.009	-0.221	0.370	1.418	-0.020
-0.191	0.315	1.459	0.035	0.057	0.975	0.531	0.006	0.029	1.036	0.604	0.074	-0.216	0.731	0.792	-0.185
-0.184	0.648	0.840	-0.125	0.077	0.436	1.115	0.242	0.037	0.742	0.602	0.038	-0.206	0.333	1.463	0.012
-0.174	0.554	0.969	-0.071	0.125	0.461	0.961	0.226	0.048	0.902	0.576	0.090	-0.206	0.327	1.462	0.012
-0.167	0.405	1.228	0.011	0.125	0.454	0.821	0.320	0.063	0.546	0.895	0.257	-0.193	1.068	0.713	-0.169
-0.164	0.506	1.102	-0.018	0.141	0.436	0.821	0.296	0.083	0.664	0.675	0.169	-0.176	0.387	1.289	0.015
-0.155	0.732	0.685	-0.132	0.159	0.703	0.766	0.225	0.087	0.396	1.144	0.366	-0.172	0.362	1.342	0.033

-0.149	0.577	0.888	-0.059	0.173	0.283	1.127	0.485	0.088	0.698	0.625	0.118	-0.170	0.353	1.357	0.040
-0.139	0.659	0.740	-0.092	0.200	0.615	0.498	0.235	0.088	0.481	0.807	0.130	-0.167	0.331	1.378	0.051
-0.134	0.330	1.322	0.085	0.205	0.578	0.701	0.312	0.104	0.483	0.694	0.119	-0.149	1.092	0.692	-0.157
-0.117	0.476	1.007	0.023	0.226	0.375	0.952	0.462	0.109	0.449	0.799	0.158	-0.147	1.128	0.664	-0.162
-0.115	0.608	0.772	-0.046	0.226	0.651	0.481	0.228	0.199	0.364	1.085	0.514	-0.139	1.050	0.665	-0.151
-0.113	0.379	1.188	0.080	0.262	0.403	0.915	0.553	0.207	0.959	0.481	0.205	-0.134	0.644	0.785	-0.114
-0.108	0.690	0.663	-0.114	0.283	0.542	0.457	0.376	0.209	0.316	1.040	0.463	-0.129	0.711	0.650	-0.150
-0.099	0.560	0.829	-0.005	0.288	1.008	0.428	0.344	0.211	0.587	0.603	0.248	-0.125	0.568	0.880	-0.074
-0.083	0.404	1.088	0.096	0.305	0.931	0.426	0.367	0.233	0.796	0.442	0.227	-0.113	0.375	1.323	0.062
-0.071	0.582	0.736	0.006	0.317	0.417	0.724	0.438	0.243	0.986	0.436	0.242	-0.090	0.381	1.271	0.083
-0.065	0.255	1.361	0.197	0.334	0.304	0.854	0.634	0.317	0.545	0.585	0.516	-0.080	0.668	0.633	-0.085
-0.060	0.648	0.616	-0.061	0.352	0.557	0.391	0.450	0.337	0.738	0.584	0.469	-0.073	0.367	1.279	0.111
-0.060	0.340	1.262	0.180	0.353	0.259	1.015	0.716	0.339	0.776	0.458	0.425	-0.055	0.387	1.216	0.122
-0.051	0.419	1.011	0.120	0.375	0.528	0.720	0.523	0.339	0.936	0.442	0.464	-0.048	0.369	1.232	0.136
-0.039	0.495	0.849	-0.002	0.381	0.971	0.395	0.427	0.339	0.388	0.848	0.597	-0.038	1.039	0.600	-0.068
-0.037	0.285	1.251	0.209	0.456	0.765	0.321	0.502	0.347	0.764	0.543	0.469	-0.024	0.562	0.691	-0.026
-0.035	0.359	1.099	0.171	0.458	0.436	0.439	0.581	0.437	0.354	0.836	0.737	-0.017	0.614	0.559	-0.059
-0.027	0.529	0.754	0.075	0.561	0.491	0.306	0.641	0.503	0.520	0.432	0.608	0.017	0.494	0.774	0.058
-0.019	0.462	0.921	0.049					0.505	0.607	0.342	0.615	0.031	0.386	1.154	0.183
-0.007	0.273	1.283	0.263					0.582	0.557	0.220	0.636	0.032	0.293	1.136	0.280
-0.003	0.439	0.899	0.156					0.638	0.386	0.225	0.714	0.037	0.492	0.746	0.078
-0.003	0.555	0.637	-0.014									0.050	0.209	1.269	0.342
0.019	0.597	0.571	-0.003									0.051	0.964	0.531	-0.002
0.023	0.648	0.530	-0.010									0.061	0.916	0.520	0.008
0.042	0.359	1.022	0.261									0.066	0.240	1.199	0.347
0.053	0.273	1.145	0.312									0.075	0.207	1.250	0.374
0.062	0.407	0.860	0.239									0.078	0.910	0.512	0.029
0.068	0.491	0.585	0.032									0.083	0.533	0.481	0.019
0.075	0.201	1.244	0.372									0.104	0.486	0.487	0.052
0.081	0.558	0.482	0.018									0.135	0.476	0.468	0.089
0.086	0.324	0.995	0.315									0.139	0.301	1.115	0.381
0.101	0.450	0.630	0.093									0.155	0.317	1.054	0.387
0.121	0.470	0.511	0.084									0.161	0.295	1.098	0.408
0.128	0.260	1.060	0.397									0.188	0.881	0.455	0.168
0.129	0.178	1.207	0.442									0.189	0.883	0.443	0.166
0.153	0.206	1.126	0.452									0.209	0.520	0.454	0.229
0.154	0.371	0.814	0.357									0.211	0.522	0.442	0.227
0.168	0.430	0.650	0.196									0.216	0.453	0.420	0.188
0.183	0.259	0.993	0.457									0.223	0.454	0.410	0.194
0.190	0.453	0.458	0.164									0.247	0.357	0.941	0.442
0.206	0.299	0.894	0.460									0.263	0.818	0.408	0.259
0.222	0.167	1.102	0.547									0.274	0.335	0.978	0.492
0.253	0.367	0.712	0.466									0.281	0.305	0.907	0.527
0.257	0.197	1.017	0.572									0.284	0.854	0.400	0.287
0.274	0.433	0.389	0.258									0.294	0.810	0.386	0.296
0.274	0.324	0.776	0.518									0.311	0.463	0.372	0.306
0.276	0.235	0.927	0.572									0.319	0.442	0.360	0.312
0.303	0.171	0.999	0.634									0.335	0.182	0.972	0.674
0.314	0.359	0.670	0.539									0.356	0.373	0.796	0.515
0.323	0.245	0.864	0.620									0.366	0.157	0.951	0.711
0.372	0.186	0.902	0.704									0.381	0.144	0.952	0.733
0.378	0.408	0.335	0.387									0.387	0.117	0.982	0.751
0.442	0.344	0.645	0.675									0.390	0.102	0.998	0.762
0.512	0.386	0.255	0.545									0.390	0.102	0.998	0.762
0.524	0.298	0.711	0.777									0.424	0.422	0.463	0.578
0.538	0.258	0.772	0.817									0.431	0.842	0.342	0.554
0.585	0.336	0.558	0.812									0.440	0.444	0.340	0.551
0.599	0.276	0.658	0.868									0.448	0.785	0.330	0.534
0.645	0.652	0.179	0.711									0.450	0.785	0.316	0.530
0.645	0.413	0.180	0.704									0.466	0.389	0.515	0.660
0.691	0.371	0.159	0.759									0.468	0.393	0.280	0.492
												0.468	0.230	0.835	0.778
												0.482	0.230	0.810	0.790
												0.491	0.191	0.859	0.820
												0.495	0.227	0.800	0.806

												0.502	0.195	0.852	0.834
												0.516	0.285	0.768	0.783
												0.525	0.417	0.265	0.604
												0.542	0.772	0.261	0.640
												0.548	0.770	0.238	0.601
												0.555	0.262	0.758	0.837
												0.574	0.383	0.223	0.620
												0.581	0.249	0.728	0.866
												0.591	0.258	0.712	0.874
												0.601	0.380	0.208	0.652
												0.612	0.753	0.203	0.678
												0.639	0.373	0.183	0.696
												0.666	0.294	0.569	0.895
												0.683	0.266	0.604	0.932
												0.688	0.402	0.166	0.759
												0.691	0.370	0.154	0.757
												0.699	0.254	0.592	0.952

6.3 Chapter 3 Materials and Methods

6.3.1 Chemicals

All compounds were used as received. 1-Methyl-2-pyrrolidone (NMP; 99%), dimethylformamide (DMF; Extra Pure), dimethylacetamide (DMAc; 99+%), ethanol (EtOH; 99.8%), acetonitrile (MeCN; HPLC grade), morpholine (99+%) were purchased from Fisher Scientific Ltd. 2,4-Difluoronitrobenzene (99%), DavePhos (98%), XPhos (98%), CyJohnPhos (99%), 3,3-dimethyl-1-butyne (98%), $Pd(PPh_3)_4$ (98%), pyrrolidine (98%) were purchased from Fluorochem. Toluene (HPLC grade) was purchased from VWR. Triethylamine (99.5%), tricyclohexylphosphine (>94%), CuI (>99.5%) were purchased from Merck. 2-Bromo-4-(trifluoromethyl)benzonitrile (99.5%) was purchased from Apollo scientific.

6.3.2 Experimental Set-Up

S_NAr Reaction

Reservoir solutions were prepared by dissolving the desired reagents in solvent under stirring at ambient conditions. Reagent 1: 2,4-Difluoronitrobenzene (1.015 mol L^{-1}), and biphenyl (0.05 mol L^{-1}) in the desired solvent; Reagent 2: Morpholine (2.085 mol L^{-1}) in triethylamine; Solvent reservoirs (NMP, DMF, DMAc, MeCN, ethanol).

HPLC mobile phases were A H_2O (18.2 MΩ), B MeCN, both buffered with 0.1% TFA. The method used was 40% B 5 mins, 40% to 95% B 2.5 mins, 95% to 40 % B 0.1 min, 40% B 1 min, flow rate 1.5 mL min^{-1}, column temperature 20 °C.

Reaction optimisation was performed with respect to five variables: residence time, morpholine equivalents, 2,4-difluoronitrobenzene concentration, temperature, and solvent choice. Variable bounds are provided in Table 6-10. The target of the optimisation was the simultaneous maximisation of the *ortho* and *para* products of the reaction.

Table 6-10. Optimisation bounds for S_NAr case study.

	Variables				
Limits	t_{res} (min)	Morph equiv.	2,4-difluoronitrobenzene concentration (M)	Temperature (°C)	Solvent
Lower	0.5	1	0.05	60	NMP, DMF, DMAc, MeCN, Ethanol
Upper	2	5	0.175	120	

Sonogashira Reaction

Reservoir solutions were prepared by dissolving the desired reagents in solvent under stirring at ambient conditions. Reagent 1: 2-Bromo-4-(trifluoromethyl)benzonitrile (0.3 mol L^{-1}), $Pd(PPh_3)_4$ (0.0051 mol L^{-1}), CuI (0.06 mol L^{-1}), pyrrolidine (0.9 mol L^{-1}) and biphenyl (0.075 mol L^{-1}) in Toluene/MeCN (5:1); Reagent 2: 3,3-Dimethyl-1-butyne (1 mol L^{-1}) in Toluene/MeCN (5:1); Ligands: DavePhos (0.015 mol L^{-1}) in Toluene/MeCN (5:1), XPhos (0.015 mol L^{-1}) in Toluene/MeCN (5:1), CyJohnPhos (0.015 mol L^{-1}) in Toluene/MeCN (5:1), tricyclohexyphosphine (0.015 mol L^{-1}) in Toluene/MeCN (5:1) and Toluene/MeCN (5:1); Solvent: Toluene/MeCN (5:1).

HPLC mobile phases were A H_2O (18.2 MΩ), B MeCN, both buffered with 0.1% TFA. The method used was 5% to 95% B 5 mins, 95% to 5% B 0.1 min, 5% B 1 min, flow rate 1.5 mL min-1, column temperature 20 °C.

Reaction optimisation was performed with respect to five variables: residence time, butyne equivalents, ligand concentration, temperature, and ligand choice. Variable bounds are provided in Table 6-11. The target of the optimisation was the simultaneous maximisation of the Yield and TOF and the minimisation of cost for the reaction.

Table 6-11. Sonogashira optimisation bounds.

	Variables				
Limits	t_{res} (min)	Butyne equiv.	Ligand concentration (mol %)	Temperature (°C)	Ligand
Lower	1	1	1	60	DavePhos, XPhos, CyJohnPhos, tricyclohexylphosphine, triphenylphosphine
Upper	10	3	5	140	

6.3.3 Optimisation Results

$$\max\,[ortho, para]$$

S_NAr Reaction

Table 6-12. Optimisation results for the S_NAr case study. PI – Polarity index, DM – Dipole moment, DC – Dielectric constant.

Exp.	Type	t_{res} (min)	Eq.	[SM] (M)	T (°C)	Ortho (%)	Para (%)	Solvent	PI	DM	DC
1	Initial	1.19	2.26	0.15	62.66	44.28	28.94	DMF	6.4	3.86	36.71
2	Initial	1.19	2.26	0.15	62.66	39.56	35.46	NMP	6.7	4.09	32.2
3	Initial	1.19	2.26	0.15	62.66	13.62	3.01	EtOH	5.2	1.66	24.55
4	Initial	1.19	2.26	0.15	62.66	34.43	28.79	DMAc	6.5	3.72	37.78
5	Initial	1.19	2.26	0.15	62.66	34.18	6.51	MeCN	5.8	3.44	37.5
6	Initial	1.65	4.96	0.06	79.27	57.13	33.33	DMF	6.4	3.86	36.71
7	Initial	1.65	4.96	0.06	79.27	50.02	41.40	NMP	6.7	4.09	32.2
8	Initial	1.65	4.96	0.06	79.27	30.98	6.66	EtOH	5.2	1.66	24.55
9	Initial	1.65	4.96	0.06	79.27	47.46	37.47	DMAc	6.5	3.72	37.78
10	Initial	1.65	4.96	0.06	79.27	56.33	8.52	MeCN	5.8	3.44	37.5
11	Initial	0.56	3.98	0.10	90.78	45.15	26.71	DMF	6.4	3.86	36.71
12	Initial	0.56	3.98	0.10	90.78	39.99	33.26	NMP	6.7	4.09	32.2
13	Initial	0.56	3.98	0.10	90.78	19.28	2.96	EtOH	5.2	1.66	24.55
14	Initial	0.56	3.98	0.10	90.78	35.59	29.00	DMAc	6.5	3.72	37.78
15	Initial	0.56	3.98	0.10	90.78	33.93	4.22	MeCN	5.8	3.44	37.5
16	Initial	1.97	2.64	0.12	107.14	62.20	32.31	DMF	6.4	3.86	36.71
17	Initial	1.97	2.64	0.12	107.14	54.15	40.48	NMP	6.7	4.09	32.2
18	Initial	1.97	2.64	0.12	107.14	59.58	7.88	EtOH	5.2	1.66	24.55
19	Initial	1.97	2.64	0.12	107.14	55.62	38.05	DMAc	6.5	3.72	37.78
20	Initial	1.97	2.64	0.12	107.14	74.14	10.22	MeCN	5.8	3.44	37.5
21	Initial	0.83	1.65	0.16	112.06	49.17	28.84	DMF	6.4	3.86	36.71
22	Initial	0.83	1.65	0.16	112.06	44.07	35.64	NMP	6.7	4.09	32.2
23	Initial	0.83	1.65	0.16	112.06	31.35	4.01	EtOH	5.2	1.66	24.55
24	Initial	0.83	1.65	0.16	112.06	41.76	32.42	DMAc	6.5	3.72	37.78
25	Initial	0.83	1.65	0.16	112.06	45.42	5.58	MeCN	5.8	3.44	37.5
26	Algorithm	2.00	2.73	0.15	107.11	52.85	41.60	NMP	6.7	4.09	32.2
27	Algorithm	2.00	2.77	0.18	107.61	52.93	41.47	NMP	6.7	4.09	32.2

28	Algorithm	1.67	3.65	0.18	109.05	53.86	39.49	NMP	6.7	4.09	32.2
29	Algorithm	2.00	5.00	0.05	120.00	79.81	10.36	MeCN	5.8	3.44	37.5
30	Algorithm	2.00	3.73	0.05	117.08	53.56	40.32	DMAc	6.5	3.72	37.78
31	Algorithm	2.00	5.00	0.05	120.00	53.10	40.80	NMP	6.7	4.09	32.2
32	Algorithm	2.00	5.00	0.05	88.83	49.45	48.16	NMP	6.7	4.09	32.2
33	Algorithm	1.81	4.47	0.16	94.87	54.37	40.26	NMP	6.7	4.09	32.2
34	Algorithm	1.77	3.66	0.14	88.83	52.23	43.29	NMP	6.7	4.09	32.2
35	Algorithm	1.78	4.49	0.05	88.13	48.23	42.74	NMP	6.7	4.09	32.2
36	Algorithm	1.98	4.33	0.18	87.97	53.39	41.87	NMP	6.7	4.09	32.2
37	Algorithm	2.00	5.00	0.05	120.00	62.04	31.43	DMF	6.4	3.86	36.71
38	Algorithm	2.00	5.00	0.18	91.31	59.85	33.12	DMF	6.4	3.86	36.71
39	Algorithm	2.00	5.00	0.05	96.81	51.73	42.32	DMAc	6.5	3.72	37.78
40	Algorithm	2.00	1.00	0.05	85.90	28.56	18.48	DMF	6.4	3.86	36.71
41	Algorithm	1.95	5.00	0.18	108.20	59.77	29.19	DMF	6.4	3.86	36.71
42	Algorithm	2.00	1.01	0.05	120.00	18.96	15.86	NMP	6.7	4.09	32.2
43	Algorithm	2.00	3.28	0.18	98.52	52.93	42.15	NMP	6.7	4.09	32.2
44	Algorithm	2.00	5.00	0.10	113.46	54.50	39.19	DMAc	6.5	3.72	37.78
45	Algorithm	2.00	5.00	0.05	103.90	49.31	42.18	NMP	6.7	4.09	32.2
46	Algorithm	2.00	3.63	0.05	106.15	45.29	37.72	DMAc	6.5	3.72	37.78
47	Algorithm	1.85	3.56	0.18	120.00	59.49	29.69	DMF	6.4	3.86	36.71
48	Algorithm	1.52	5.00	0.18	120.00	56.25	34.21	DMAc	6.5	3.72	37.78
49	Algorithm	0.96	5.00	0.18	120.00	58.54	35.54	NMP	6.7	4.09	32.2
50	Algorithm	2.00	5.00	0.18	60.83	53.58	40.93	NMP	6.7	4.09	32.2
51	Algorithm	2.00	2.81	0.18	117.83	55.62	37.99	DMAc	6.5	3.72	37.78
52	Algorithm	1.39	5.00	0.18	119.00	56.25	35.73	NMP	6.7	4.09	32.2
53	Algorithm	2.00	4.02	0.17	94.07	53.74	41.64	NMP	6.7	4.09	32.2
54	Algorithm	2.00	5.00	0.18	66.21	59.60	36.65	DMF	6.4	3.86	36.71
55	Algorithm	1.46	2.72	0.18	104.20	52.75	42.06	NMP	6.7	4.09	32.2
56	Algorithm	2.00	5.00	0.18	71.75	54.47	41.83	NMP	6.7	4.09	32.2
57	Algorithm	0.72	5.00	0.05	120.00	53.26	30.31	DMF	6.4	3.86	36.71
59	Algorithm	2.00	3.42	0.18	105.35	59.41	32.56	DMF	6.4	3.86	36.71
60	Algorithm	0.80	5.00	0.05	116.29	50.23	41.18	NMP	6.7	4.09	32.2
61	Algorithm	0.50	5.00	0.05	120.00	24.70	0.00	EtOH	5.2	1.66	24.55
62	Algorithm	1.92	4.57	0.06	119.84	54.23	41.34	DMAc	6.5	3.72	37.78
63	Algorithm	1.56	5.00	0.18	119.16	61.40	27.97	DMF	6.4	3.86	36.71

64	Algorithm	2.00	4.44	0.13	120.00	55.06	35.06	DMAc	6.5	3.72	37.78
65	Algorithm	2.00	1.57	0.06	119.33	52.60	7.22	MeCN	5.8	3.44	37.5
66	Algorithm	1.03	4.83	0.08	120.00	52.92	43.14	NMP	6.7	4.09	32.2
67	Algorithm	2.00	5.00	0.18	113.32	75.20	16.89	MeCN	5.8	3.44	37.5
68	Algorithm	2.00	5.00	0.17	120.00	78.95	13.97	MeCN	5.8	3.44	37.5
69	Algorithm	2.00	2.51	0.16	112.78	54.35	40.17	DMAc	6.5	3.72	37.78
70	Algorithm	2.00	5.00	0.11	91.12	53.00	43.25	NMP	6.7	4.09	32.2
71	Algorithm	2.00	5.00	0.18	76.63	61.10	34.76	DMF	6.4	3.86	36.71
72	Algorithm	1.76	2.73	0.18	111.05	54.52	40.14	DMAc	6.5	3.72	37.78
73	Algorithm	2.00	2.29	0.18	105.66	53.35	41.50	DMAc	6.5	3.72	37.78
74	Algorithm	1.98	2.96	0.05	112.64	58.18	33.81	DMF	6.4	3.86	36.71
75	Algorithm	2.00	2.79	0.17	86.66	51.61	43.55	NMP	6.7	4.09	32.2
76	Algorithm	1.83	2.41	0.17	97.96	51.93	43.01	NMP	6.7	4.09	32.2
77	Algorithm	2.00	3.85	0.15	74.79	52.00	43.40	NMP	6.7	4.09	32.2
78	Algorithm	2.00	5.00	0.08	68.61	50.67	44.34	NMP	6.7	4.09	32.2
79	Algorithm	2.00	5.00	0.18	90.10	55.29	40.43	DMAc	6.5	3.72	37.78
80	Algorithm	2.00	4.52	0.17	113.28	76.81	16.76	MeCN	5.8	3.44	37.5
81	Algorithm	2.00	3.88	0.17	114.97	75.91	17.59	MeCN	5.8	3.44	37.5
82	Algorithm	2.00	5.00	0.14	77.37	53.74	41.45	NMP	6.7	4.09	32.2
83	Algorithm	2.00	5.00	0.18	71.25	59.88	36.98	DMF	6.4	3.86	36.71
84	Algorithm	2.00	5.00	0.08	93.64	51.54	40.34	NMP	6.7	4.09	32.2
85	Algorithm	2.00	3.50	0.10	88.06	51.65	43.49	NMP	6.7	4.09	32.2
86	Algorithm	1.86	2.64	0.17	102.45	52.76	42.09	NMP	6.7	4.09	32.2
87	Algorithm	2.00	4.64	0.14	115.52	78.09	16.26	MeCN	5.8	3.44	37.5
88	Algorithm	1.05	4.99	0.17	110.87	56.00	38.42	NMP	6.7	4.09	32.2
89	Algorithm	1.40	5.00	0.18	109.60	55.78	38.02	NMP	6.7	4.09	32.2
90	Algorithm	2.00	5.00	0.18	99.14	55.44	38.36	DMAc	6.5	3.72	37.78
91	Algorithm	2.00	4.32	0.10	67.75	50.29	44.12	NMP	6.7	4.09	32.2
92	Algorithm	1.46	5.00	0.18	68.63	59.21	35.17	DMF	6.4	3.86	36.71
93	Algorithm	1.40	5.00	0.18	64.86	52.28	41.47	NMP	6.7	4.09	32.2
94	Algorithm	2.00	5.00	0.18	120.00	61.14	25.93	DMF	6.4	3.86	36.71
95	Algorithm	1.55	3.19	0.13	110.20	60.55	33.91	DMF	6.4	3.86	36.71
96	Algorithm	1.05	4.83	0.16	115.82	55.50	39.24	NMP	6.7	4.09	32.2
97	Algorithm	1.73	5.00	0.05	60.07	46.39	29.13	DMF	6.4	3.86	36.71
98	Algorithm	2.00	4.94	0.17	66.73	53.41	42.98	NMP	6.7	4.09	32.2

99	Algorithm	1.84	2.68	0.18	109.41	60.05	34.63	DMF	6.4	3.86	36.71
100	Algorithm	1.71	4.93	0.05	115.29	59.81	35.83	DMF	6.4	3.86	36.71
101	Algorithm	1.53	4.04	0.06	116.03	59.84	33.78	DMF	6.4	3.86	36.71
102	Algorithm	2.00	5.00	0.10	87.01	52.14	44.66	NMP	6.7	4.09	32.2
103	Algorithm	1.48	5.00	0.18	101.39	54.93	39.70	NMP	6.7	4.09	32.2
104	Algorithm	2.00	5.00	0.17	90.33	54.24	40.78	NMP	6.7	4.09	32.2
105	Algorithm	0.96	5.00	0.15	119.70	55.38	38.94	NMP	6.7	4.09	32.2
106	Algorithm	2.00	5.00	0.06	85.01	50.43	47.07	NMP	6.7	4.09	32.2
107	Algorithm	1.45	5.00	0.11	115.48	54.58	38.96	NMP	6.7	4.09	32.2
108	Algorithm	1.37	3.98	0.16	105.03	54.35	40.44	NMP	6.7	4.09	32.2
109	Algorithm	2.00	4.40	0.05	114.50	59.92	34.95	DMF	6.4	3.86	36.71
110	Algorithm	2.00	3.15	0.18	90.81	52.78	42.61	NMP	6.7	4.09	32.2
111	Algorithm	2.00	4.02	0.17	81.21	53.54	42.33	NMP	6.7	4.09	32.2
112	Algorithm	1.85	5.00	0.18	73.65	60.24	36.21	DMF	6.4	3.86	36.71

Sonogashira Reaction

Table 6-13. Results for the Sonogashira reaction optimisation

Exp.	Type	t_{res} (min)	Eq.	L (mol %)	T (°C)	Ligand	Yield (%)	Cost (£/g)	TOF
0	Initial	5.11	1.63	4.08	63.50	DavePhos	10.25	86.74	2.12
1	Initial	7.93	2.98	1.38	85.70	DavePhos	55.52	20.54	4.07
2	Initial	1.39	2.49	2.54	101.00	DavePhos	19.99	52.64	10.44
3	Initial	9.82	1.82	3.34	122.90	DavePhos	43.73	21.00	3.00
4	Initial	2.95	1.32	4.43	129.40	DavePhos	21.28	38.92	5.45
5	Initial	5.11	1.63	4.08	63.50	XPhos	14.04	62.17	2.44
6	Initial	7.93	2.98	1.38	85.70	XPhos	59.15	19.18	4.36
7	Initial	1.39	2.49	2.54	101.00	XPhos	22.16	47.03	11.60
8	Initial	9.82	1.82	3.34	122.90	XPhos	47.08	19.23	3.15
9	Initial	2.95	1.32	4.43	129.40	XPhos	24.21	33.48	6.14
10	Initial	5.11	1.63	4.08	63.50	CyJohnPhos	9.03	99.18	1.61
11	Initial	7.93	2.98	1.38	85.70	CyJohnPhos	52.02	21.96	3.94
12	Initial	1.39	2.49	2.54	101.00	CyJohnPhos	18.00	58.67	9.48
13	Initial	9.82	1.82	3.34	122.90	CyJohnPhos	40.35	22.90	2.79
14	Initial	2.95	1.32	4.43	129.40	CyJohnPhos	19.62	42.57	4.93
15	Initial	5.11	1.63	4.08	63.50	Tricyclohexylphosphine	10.37	81.82	2.26
16	Initial	7.93	2.98	1.38	85.70	Tricyclohexylphosphine	54.76	20.57	4.23
17	Initial	1.39	2.49	2.54	101.00	Tricyclohexylphosphine	20.63	49.77	11.68
18	Initial	9.82	1.82	3.34	122.90	Tricyclohexylphosphine	37.16	23.81	2.75
19	Initial	2.95	1.32	4.43	129.40	Tricyclohexylphosphine	19.46	40.27	5.67
20	Initial	5.11	1.63	4.08	63.50	Triphenylphosphine	24.70	33.47	3.69
21	Initial	7.93	2.98	1.38	85.70	Triphenylphosphine	66.80	16.75	5.03
22	Initial	1.39	2.49	2.54	101.00	Triphenylphosphine	29.98	33.80	15.11
23	Initial	9.82	1.82	3.34	122.90	Triphenylphosphine	58.97	14.71	3.83

24	Initial	2.95	1.32	4.43	129.40	Triphenylphosphine	36.64	20.74	8.61
25	Algorithm	9.14	2.24	1.98	86.76	Triphenylphosphine	59.17	16.19	4.02
26	Algorithm	2.07	3.00	1.00	114.60	Triphenylphosphine	48.18	23.32	14.82
27	Algorithm	10.00	3.00	1.00	140.00	Triphenylphosphine	85.43	13.15	5.23
28	Algorithm	4.39	3.00	1.00	140.00	Triphenylphosphine	74.45	15.09	10.63
29	Algorithm	2.70	3.00	1.00	140.00	Triphenylphosphine	62.97	17.85	15.18
30	Algorithm	5.50	1.67	1.80	71.43	DavePhos	26.35	32.71	4.28
31	Algorithm	1.00	3.00	1.00	140.00	Triphenylphosphine	38.81	28.95	28.61
32	Algorithm	5.50	1.67	1.80	71.43	DavePhos	24.33	35.42	3.97
33	Algorithm	5.50	1.67	1.80	71.43	DavePhos	27.56	31.28	4.09
34	Algorithm	1.78	2.33	1.04	121.85	Triphenylphosphine	41.28	23.69	16.32
35	Algorithm	2.12	3.00	1.00	139.99	Triphenylphosphine	60.29	18.64	18.42
36	Algorithm	9.92	3.00	1.01	134.51	Triphenylphosphine	89.06	12.61	5.50
37	Algorithm	1.67	3.00	1.00	140.00	Triphenylphosphine	54.40	20.66	21.50
38	Algorithm	7.58	3.00	1.00	138.01	Triphenylphosphine	86.17	13.04	7.05
39	Algorithm	10.00	2.75	1.00	140.00	CyJohnPhos	81.44	13.33	5.15
40	Algorithm	1.19	3.00	5.00	140.00	Triphenylphosphine	43.87	25.62	25.78
41	Algorithm	1.03	3.00	3.78	139.25	Triphenylphosphine	41.64	26.99	29.12
42	Algorithm	9.36	1.64	4.46	76.56	Triphenylphosphine	40.66	20.36	3.12
43	Algorithm	1.02	2.64	2.01	137.38	Triphenylphosphine	37.83	27.63	27.38
44	Algorithm	1.05	2.18	4.42	77.67	Triphenylphosphine	13.32	70.97	15.30
45	Algorithm	5.88	3.00	4.38	140.00	Triphenylphosphine	82.01	13.70	8.79
46	Algorithm	1.00	3.00	1.47	116.89	Triphenylphosphine	31.25	35.96	24.89
47	Algorithm	6.68	3.00	1.97	112.34	Triphenylphosphine	75.67	14.85	7.11
48	Algorithm	1.06	3.00	5.00	134.80	Triphenylphosphine	39.49	28.46	27.72
49	Algorithm	1.14	2.71	1.45	125.02	Triphenylphosphine	34.73	30.54	23.70
50	Algorithm	5.26	3.00	3.53	120.62	Triphenylphosphine	72.73	15.45	8.88
51	Algorithm	1.54	2.92	2.50	139.29	Triphenylphosphine	49.31	22.43	22.44
52	Algorithm	1.08	3.00	5.00	120.88	Triphenylphosphine	33.49	33.55	24.87
53	Algorithm	1.00	3.00	1.00	133.42	Triphenylphosphine	37.29	30.14	28.50
54	Algorithm	6.23	3.00	4.03	75.50	Triphenylphosphine	48.15	23.34	5.42
55	Algorithm	10.00	1.13	4.63	140.00	XPhos	23.46	32.84	2.52
56	Algorithm	1.07	2.98	3.98	139.21	XPhos	31.79	36.64	23.92
57	Algorithm	10.00	1.78	1.46	102.19	Triphenylphosphine	56.26	15.25	3.86
58	Algorithm	3.96	2.92	2.12	134.16	Triphenylphosphine	69.56	15.89	11.54
59	Algorithm	7.50	3.00	1.00	140.00	DavePhos	81.09	14.05	7.01
60	Algorithm	10.00	3.00	1.00	129.16	XPhos	84.00	13.51	5.36
61	Algorithm	1.14	2.01	2.46	135.02	Triphenylphosphine	32.31	28.16	23.16
62	Algorithm	1.03	2.96	4.90	111.56	Triphenylphosphine	27.59	40.42	24.03
63	Algorithm	1.23	2.67	3.71	133.03	Triphenylphosphine	37.47	28.08	24.17
64	Algorithm	6.66	3.00	1.00	140.00	Tricyclohexylphosphine	79.26	14.25	7.82
65	Algorithm	6.26	3.00	1.00	140.00	Triphenylphosphine	78.14	14.38	8.20
66	Algorithm	1.02	3.00	1.00	129.87	Triphenylphosphine	34.74	32.35	28.08
67	Algorithm	3.95	3.00	2.59	140.00	Triphenylphosphine	69.81	16.10	11.94
68	Algorithm	10.00	3.00	1.00	116.52	Triphenylphosphine	81.67	13.76	5.24
69	Algorithm	10.00	1.94	1.00	111.73	CyJohnPhos	58.16	15.66	4.09
70	Algorithm	1.08	3.00	1.00	120.21	Triphenylphosphine	30.80	36.49	25.01

71	Algorithm	1.00	3.00	5.00	125.15	Triphenylphosphine	31.30	35.90	27.64
72	Algorithm	7.20	3.00	4.40	132.41	Triphenylphosphine	76.35	14.72	7.11
73	Algorithm	8.66	3.00	1.00	131.64	Triphenylphosphine	78.96	14.23	6.07
74	Algorithm	7.22	3.00	1.00	126.75	Triphenylphosphine	75.81	14.82	7.03
75	Algorithm	1.05	3.00	5.00	116.23	Triphenylphosphine	29.32	38.33	25.47
76	Algorithm	1.07	2.64	1.99	137.39	Triphenylphosphine	34.20	30.59	27.81
77	Algorithm	1.00	1.48	5.00	137.97	Triphenylphosphine	22.43	35.37	23.77
78	Algorithm	1.00	3.00	5.00	131.34	Triphenylphosphine	32.55	34.52	29.02
79	Algorithm	1.06	2.55	2.80	122.84	Triphenylphosphine	27.82	36.85	25.11
80	Algorithm	5.38	1.00	1.00	92.15	Triphenylphosphine	23.10	29.86	4.51
81	Algorithm	3.08	3.00	4.33	140.00	Triphenylphosphine	61.67	18.22	14.26
82	Algorithm	6.32	3.00	3.16	136.23	Triphenylphosphine	68.42	16.43	7.59
83	Algorithm	1.00	3.00	5.00	127.78	Triphenylphosphine	30.67	36.64	27.93
84	Algorithm	6.09	3.00	1.00	135.97	XPhos	62.11	18.28	7.44
85	Algorithm	1.08	2.98	3.98	139.21	Tricyclohexylphosphine	27.19	41.97	24.67
86	Algorithm	10.00	3.00	1.00	140.00	Tricyclohexylphosphine	72.06	15.67	5.05
87	Algorithm	3.13	3.00	1.38	140.00	Triphenylphosphine	59.51	18.88	13.91
88	Algorithm	1.00	1.96	1.00	127.36	Triphenylphosphine	21.08	42.58	24.22

6.4 Chapter 4 Materials and Methods

6.4.1 Chemicals

All compounds were used as received. Tetralin (≥97%, Fisher), benzophenone (≥99.0%, Fluka), α-tetralone (97%, Aldrich) and ethyl acetate (VWR).

6.4.2 Experimental Set-Up

Four miniature CSTR's (known as fReactors) were used in series as the reaction vessels. Each CSTR had an internal volume of 2 mL and contained a PTFE cross magnetic stirrer bar to provide active mixing during the reaction. A long stage IKA RT 5 stirrer plate was used to control the magnetic stirrer bars. Each reactor had a viewing window through which blue light was irradiated utilising Engin LZ4-44UV00-0000 LEDs (365 nm, 2.9 W) that were positioned above each window. Both active and passive cooling of the LEDs was provided by Delta Electronics FFB0412SHN Server Square Fans (12 V, 0.6 A) and heat sinks respectively. Air flow rate regulated by a Bronkhorst EL-FLOW prestige mass flow controller. Gas-liquid separation was provided by an inline Zaiput SEP-10 separator using PTFE membrane with a pore size of 0.5 µm.

Reservoir solutions were prepared by dissolving the desired reagents in solvent under stirring at ambient conditions. Pump 1: Tetralin (0.04 mol L^{-1}),

benzophenone (50 mol%) and biphenyl (0.0025 mol L^{-1}) in ethyl acetate. The residence times were calculated based on the total flow rate of liquid and gas, where the flow rate of gas was corrected for in the reactor using the ideal gas equation. However, the effect of liquid holdup in the CSTRs, and the resultant gas bypassing, was not accounted for.

HPLC mobile phases were A H_2O (18.2 MΩ), B MeCN, both buffered with 0.1% TFA. The method used was 5% to 95% B 5 mins, 95% to 5% B 0.1 min, 5% B 1 min, flow rate 1.5 mL min^{-1}, column temperature 20 °C.

6.4.3 Optimisation Results

The goal of the optimisation was the maximisation of tetralone yield and subsequent response surface mapping around the optimum yield.

Table 6-14. Global bounds for the two variable self-optimisation.

	Variables	
Limits	t_{res}/min	O_2 equiv.
Lower	2	1
Upper	30	5

Table 6-15. List of results from the hybrid self-optimisation. SNOBFIT experiments are highlighted in green, screening experiments are highlighted in orange and CCF experiments are highlighted in blue.

Entry	t_{res} (min)	O_2 equiv.	Conversion (%)	Yield (%)
1	24.8	2.69	59.4	43.3
2	2.1	2.46	8.7	7.7
3	13.4	4.93	41.5	33.2
4	8.3	1.14	17.4	11.5
5	18.5	1.37	37.2	24.6
6	29.9	4.28	64.7	44.8
7	6	4.86	57.5	48.4
8	21.1	4.54	62.4	53.5
9	18.9	3.58	63.7	55.9
10	25	5	70.9	59.4
11	8.9	5	63.8	55.1
12	24.9	3.81	67.6	54.6
13	20.7	3.25	63.7	51.5
14	23.7	4.76	71	60.5
15	18.3	4.69	72.4	64.9
16	9.1	5	60	51.7
17	9.3	3.14	39.7	29.7
18	15	1.37	33.9	22.7
19	18.2	5	55.9	43.3
20	6.5	5	55.4	47.5
21	23	5	61.9	50.7
22	12.9	2.47	45.8	36.7
23	13.6	3.38	54.1	47.7

24	18.3	3.54	60	50.6
25	7.3	5	52.1	43.2
26	20.7	2.36	53.3	39.5
27	19.1	4.19	65	57.1
28	24.4	4.74	73.1	60.8
29	8.1	5	60.7	51.4
30	9.3	2.14	31.1	20.8
31	15.6	4.44	59.9	54.9
32	18.2	3.98	64.5	56.8
33	24.4	5	72.8	61.9
34	24.2	1.7	50.8	30.2
35	17.7	4.57	64.9	58.3
36	24.4	5	71.5	61.2
37	24.1	4.42	74.9	63.9
38	12.6	4.13	58.2	50.4
39	2	1.63	5.7	5.1
40	2	4.69	10.3	7.3
41	18.3	1.63	42.9	27.4
42	18.3	4.69	60.7	49.3
43	8.5	3.96	48.1	39.2
44	8.5	4.98	46.5	34.7
45	15.4	3.96	59.5	51.2
46	15.4	4.98	66.5	61.9
47	18.9	3.69	64.4	53.7
48	18.9	3.74	64.4	53.5
49	19.4	3.69	63.6	52.9
50	19.4	3.74	65.6	53.7
51	18.3	3.69	59.6	45.2
52	18.3	4.69	68.6	63.1
53	19.4	3.69	63.1	55.5
54	19.4	4.69	67.7	61.2
55	18.3	4.19	67.2	60.4
56	19.4	4.19	67.3	58.1
57	18.9	3.69	64.5	53.5
58	18.9	4.69	68.6	59.7
59	18.9	4.19	66.6	60.3

Surrogate modelling was performed utilising MATLAB's `fitrgp` function. The underlying surrogate model had the following hyperparameters.

Table 6-16 - Hyperparameters and parameters for Gaussian process model. ARD = automatic relevance detection.

Parameter	Value
Kernel Function	ARDMatern52
Basis Function	Constant
Fit Method	Exact
Predict Method	Exact
Length Scale (t_{res})	13.7184

Length Scale (O_2 equiv.)	4.0281
SigmaF	22.8659
Sigma	5.7856

Model can be used for prediction of mean and variance using the following algorithmic procedure.[103]

Algorithm 6-1: Prediction of mean, variance for Gaussian process regression. As suggested by Rasmussen *et al*.

27: **inputs**: $X\ (inputs), y\ (targets), k\ (covariance\ function), \sigma_n^2\ (noise\ level)$

28: $L \leftarrow cholesky(K + \sigma_n^2 I)$

29: $\alpha \leftarrow L^T (L \backslash y)$

30: $\widehat{f_*} \leftarrow k_*^T \alpha$

31: $v \leftarrow L \backslash k_*$

32: $V[f_*] \leftarrow k(x_*, x_*) - v^T v$

33: **inputs**: $\widehat{f_*}\ (mean),\ V[f_*]\ (variance)$

6.4.4 Algorithm Improvement Results

Table 6-17. Results for SNOBFIT-DoE optimisation of the simulated deoxofluorination.

Run	1				2				3				4				5			
Exp.	T (°C)	t_{res} (min)	[B] (M)	Yield (%)	T (°C)	t_{res} (min)	[B] (M)	Yield (%)	T (°C)	t_{res} (min)	[B] (M)	Yield (%)	T (°C)	t_{res} (min)	[B] (M)	Yield (%)	T (°C)	t_{res} (min)	[B] (M)	Yield (%)
1	97.21	2.79	2.06	8.86	90.59	19.06	2.68	39.50	139.67	6.11	2.03	65.48	112.70	14.41	2.88	66.76	130.98	10.97	2.13	70.95
2	139.45	19.89	1.36	75.47	139.85	1.62	1.24	21.17	80.11	1.15	2.26	1.49	80.21	1.19	1.51	1.13	80.65	1.81	1.71	2.18
3	122.18	1.35	1.25	9.00	118.25	19.95	1.13	50.11	107.70	19.55	2.18	59.22	139.63	1.38	1.95	27.25	103.89	19.20	2.42	56.35
4	82.02	19.89	1.05	12.36	104.27	1.02	1.78	4.26	118.89	1.83	1.39	11.68	91.48	19.75	2.83	43.57	111.93	2.12	1.98	13.59
5	112.40	20.00	1.20	45.42	83.20	1.50	2.30	2.58	80.30	19.80	2.10	21.14	131.60	19.90	2.90	85.06	82.20	19.90	2.60	28.42
6	139.80	1.50	2.30	33.46	137.20	19.90	1.20	69.30	100.20	2.50	1.40	6.23	98.30	1.30	1.50	3.49	95.40	4.00	1.20	6.51
7	81.10	2.00	2.60	3.59	121.80	1.60	1.80	14.83	126.50	18.90	2.60	81.86	122.60	3.10	2.90	38.21	118.30	17.40	1.60	58.72
8	97.30	18.20	1.00	22.51	103.90	16.50	1.10	29.36	93.80	16.30	1.40	24.02	80.20	14.60	2.80	21.33	123.90	1.20	2.80	18.53
9	125.90	16.70	1.40	61.97	127.50	12.10	2.40	73.58	137.60	18.80	1.40	74.04	110.50	2.60	1.70	13.25	139.50	19.20	1.40	75.61
10	110.30	7.60	1.40	26.05	113.00	9.50	2.30	48.73	128.90	8.20	1.40	48.44	101.20	14.00	2.20	41.05	137.40	1.10	2.50	26.02
11	88.60	10.20	1.00	9.50	94.10	6.10	1.70	12.68	109.40	8.00	1.20	23.22	139.30	12.10	2.20	80.24	93.50	15.60	1.40	23.02
12	130.80	7.80	1.20	44.19	81.80	12.50	1.80	13.82	89.10	6.60	1.80	11.32	89.40	7.10	1.70	11.68	104.20	9.20	2.70	40.49
13	140.00	20.00	2.30	84.08	140.00	20.00	3.00	85.16	140.00	20.00	2.50	84.64	140.00	18.90	3.00	85.00	140.00	20.00	2.80	84.73
14	140.00	20.00	1.00	64.31	140.00	17.60	3.00	85.62	140.00	20.00	3.00	85.42	140.00	17.50	1.90	81.30	135.80	15.90	2.60	83.25
15	140.00	20.00	1.90	82.59	140.00	20.00	1.80	82.09	140.00	20.00	2.00	82.87	140.00	20.00	2.20	84.43	140.00	20.00	2.10	83.49
16	121.80	7.50	2.70	59.48	125.10	14.70	2.90	79.47	140.00	12.80	2.40	82.15	116.10	12.30	1.60	47.55	130.70	20.00	2.20	81.74
17	140.00	16.10	2.50	84.04	106.00	14.80	2.70	56.18	102.00	10.10	2.10	32.56	140.00	20.00	2.60	85.39	121.10	8.80	1.60	45.21
18	137.10	20.00	2.00	82.18	139.00	18.40	3.00	85.36	140.00	1.80	3.00	46.13	124.40	20.00	3.00	83.40	140.00	19.10	2.40	84.61
19	130.80	13.80	3.00	82.42	131.60	19.40	3.00	84.83	131.60	20.00	3.00	84.83	136.00	20.00	3.00	85.29	136.00	20.00	3.00	85.33
20	130.90	10.50	2.30	72.41	132.60	16.30	3.00	84.25	140.00	16.00	2.90	85.08	140.00	14.40	2.60	83.68	133.90	20.00	2.60	84.81

Run	6				7				8				9				10			
Exp.	T (°C)	t_{res} (min)	[B] (M)	Yield (%)	T (°C)	t_{res} (min)	[B] (M)	Yield (%)	T (°C)	t_{res} (min)	[B] (M)	Yield (%)	T (°C)	t_{res} (min)	[B] (M)	Yield (%)	T (°C)	t_{res} (min)	[B] (M)	Yield (%)
1	112.14	2.56	1.16	9.73	84.75	1.18	1.98	1.91	138.12	6.59	1.35	51.63	89.91	11.73	2.05	21.67	128.96	19.59	1.38	68.01
2	80.65	18.01	2.74	25.43	139.43	18.50	1.54	77.73	80.14	16.53	2.27	19.57	139.91	8.44	2.83	80.14	80.95	4.85	2.23	6.60
3	139.17	18.12	1.49	76.53	114.50	2.67	1.72	16.34	108.02	3.09	1.64	13.11	115.14	19.66	2.69	74.68	107.79	1.81	1.92	9.47
4	92.50	1.76	2.56	5.35	98.06	18.63	2.38	46.94	118.27	19.18	1.14	49.92	107.96	1.43	1.49	5.80	139.84	1.05	1.02	12.35
5	119.60	19.70	2.30	74.72	118.90	19.80	1.60	63.15	99.10	19.10	1.60	36.67	124.50	2.60	2.80	34.80	95.50	18.10	2.10	38.34
6	100.80	18.50	2.20	47.67	132.40	2.30	1.10	19.42	90.40	1.60	1.90	3.14	130.30	18.70	2.00	78.89	111.80	19.70	1.00	38.93
7	131.50	2.10	2.10	30.41	81.50	17.20	2.60	24.14	123.00	1.70	1.80	16.16	82.00	1.10	2.10	1.79	124.10	2.50	1.50	20.02
8	80.40	4.90	2.20	6.46	99.70	2.80	2.60	12.28	137.30	20.00	1.30	72.59	103.50	14.30	1.90	39.98	81.20	19.20	1.00	11.05
9	121.90	8.40	1.70	47.11	107.70	11.70	1.50	33.94	128.30	13.00	1.80	67.17	80.10	19.60	2.30	22.50	94.20	4.70	2.80	16.45
10	111.30	13.10	1.00	30.05	124.90	10.50	1.70	55.94	90.20	12.40	1.10	13.17	97.00	1.20	2.70	5.04	139.90	16.40	2.30	83.31
11	101.80	8.30	2.60	33.19	91.90	9.80	1.30	13.80	108.40	14.10	2.90	60.19	115.90	9.40	1.10	30.64	115.40	9.70	2.70	57.76
12	89.30	13.30	1.90	21.60	129.60	19.70	2.20	81.04	82.30	6.80	2.70	11.80	95.00	19.80	2.30	42.92	104.50	12.20	2.00	38.94
13	140.00	20.00	2.60	84.70	140.00	20.00	2.20	84.41	140.00	19.70	2.90	85.01	95.00	19.80	1.10	23.43	140.00	12.70	3.00	84.60
14	134.80	20.00	3.00	85.79	140.00	20.00	2.80	84.97	140.00	18.60	2.40	84.42	140.00	20.00	2.50	84.76	140.00	20.00	3.00	85.31
15	140.00	20.00	2.00	83.41	130.50	15.20	2.20	77.83	140.00	20.00	2.10	83.66	140.00	14.10	3.00	85.26	127.70	14.70	3.00	81.90
16	130.90	12.00	2.40	75.95	132.40	8.40	2.40	70.81	140.00	13.30	2.10	80.87	127.80	20.00	3.00	83.89	97.70	12.40	2.50	36.54
17	140.00	17.30	3.00	84.83	140.00	19.40	2.50	84.50	124.20	17.10	2.60	78.79	119.10	12.60	2.10	60.82	140.00	16.40	3.00	85.34
18	129.80	20.00	3.00	84.49	131.40	14.20	3.00	82.89	140.00	17.70	2.80	84.98	140.00	4.20	1.10	36.73	132.30	20.00	3.00	84.71
19	131.60	16.90	3.00	84.04	132.60	20.00	2.80	84.62	131.60	20.00	3.00	85.11	140.00	20.00	3.00	85.16	140.00	18.50	3.00	85.16
20	124.00	8.60	2.70	65.17	105.70	12.10	2.70	50.37	140.00	14.90	2.50	83.51	137.90	20.00	2.50	84.54	102.10	6.70	1.30	15.93

21	140.00	17.20	2.20	83.85	140.00	19.40	2.40	84.55	110.00	7.20	2.50	39.64	98.10	16.10	1.30	26.52	137.00	20.00	2.10	83.42	137.40	19.90	2.90	85.79	137.30	19.50	2.80	85.56	125.70	8.20	2.20	60.01	125.20	20.00	3.00	83.57	138.30	17.30	3.00	85.12
22	140.00	20.00	3.00	85.06	85.30	11.50	2.40	19.65	140.00	18.30	2.80	85.21	140.00	18.60	2.70	85.42	102.40	11.00	1.70	29.76	119.20	20.00	3.00	80.47	133.20	11.50	3.00	81.72	135.70	20.00	3.00	85.60	129.60	9.10	1.80	60.56	128.20	20.00	3.00	84.19
23	138.20	19.40	2.20	83.38	134.90	20.00	2.70	84.72	126.00	20.00	3.00	83.87	122.20	20.00	3.00	82.16	132.00	19.60	3.00	84.73	140.00	20.00	2.50	85.17	133.20	17.10	3.00	84.26	127.50	20.00	3.00	84.77	140.00	17.20	2.90	85.14	135.10	20.00	3.00	85.25
24	105.50	7.80	2.00	29.98	140.00	18.00	3.00	85.65	133.10	20.00	3.00	85.45	140.00	20.00	3.00	85.58	140.00	18.80	2.90	85.83	115.00	16.60	1.20	44.42	132.40	8.10	1.40	52.14	130.60	14.10	2.70	80.55	140.00	12.00	3.00	84.39	140.00	14.40	3.00	84.84
25	137.10	18.60	2.90	85.14	135.80	20.00	3.00	85.56	140.00	14.90	2.90	84.25	140.00	15.80	3.00	85.43	140.00	20.00	2.50	85.38	136.40	20.00	3.00	85.77	136.60	20.00	2.70	85.45	140.00	15.70	3.00	84.71	139.10	20.00	2.90	85.32	130.70	8.70	2.00	63.80
26	133.40	9.50	2.90	78.56	119.50	11.90	1.60	51.10	113.00	16.40	1.60	51.20	90.60	17.00	2.00	28.98	133.00	20.00	2.60	84.31	125.60	20.00	3.00	84.03	137.40	20.00	3.00	85.20	132.80	6.20	2.60	66.21	131.20	20.00	3.00	84.61	140.00	17.60	2.80	84.92
27	138.70	20.00	2.20	84.02	140.00	17.70	2.70	84.94	139.20	20.00	2.90	84.96	140.00	20.00	2.70	84.99	92.10	11.70	2.40	27.03	140.00	20.00	2.20	84.18	131.40	16.80	3.00	84.01	135.70	18.50	3.00	85.10	126.90	14.60	2.50	77.42	136.70	20.00	3.00	85.43
28	122.30	7.80	2.00	50.91	134.40	19.00	3.00	85.25	127.40	20.00	2.00	78.18	132.20	20.00	3.00	85.00	140.00	18.10	1.40	75.46	90.90	9.60	2.00	18.57	97.70	8.80	2.30	26.57	125.00	16.80	3.00	82.19	137.40	20.00	2.90	85.60	140.00	14.60	3.00	85.08
29	140.00	17.20	3.00	84.73	140.00	19.00	2.40	84.61	140.00	17.90	3.00	85.27	140.00	18.60	2.40	84.50	134.90	20.00	3.00	85.34	135.60	20.00	2.80	84.66	137.90	17.10	3.00	84.70	129.30	20.00	3.00	84.46	138.90	15.10	3.00	85.26	104.20	16.30	1.40	35.82
30	139.00	19.30	2.20	83.57	131.20	15.40	3.00	83.53	110.80	7.20	1.80	31.28	140.00	16.90	3.00	85.47	136.20	17.60	3.00	84.56	125.00	20.00	3.00	83.80	135.70	19.10	3.00	84.98	123.10	3.60	2.80	41.43	134.80	10.70	3.00	81.81	140.00	20.00	2.30	85.10
31	137.00	20.00	2.50	84.68	111.20	7.60	1.40	27.01	136.60	20.00	2.90	84.93	121.90	10.00	2.00	57.08	91.20	8.70	1.40	12.73	133.50	16.90	2.10	80.43	140.00	20.00	2.60	84.94	135.70	20.00	2.70	84.58	140.00	17.90	3.00	85.47	140.00	16.10	3.00	85.26
32	100.90	9.80	2.90	39.34	140.00	18.10	3.00	85.66	140.00	17.30	3.00	85.08	140.00	20.00	2.80	85.23	133.70	16.80	3.00	85.11	139.40	19.60	2.30	84.64	130.50	16.30	3.00	83.15	137.90	20.00	3.00	84.96	100.10	10.30	2.50	35.72	138.60	20.00	3.00	85.11
33	140.00	19.50	2.70	84.99	133.20	20.00	3.00	84.98	140.00	12.50	1.50	71.22	140.00	20.00	2.40	85.04	134.50	20.00	3.00	85.25	124.80	15.70	1.70	66.33	103.80	15.90	1.30	32.78	137.70	10.60	3.00	82.56	136.40	18.30	3.00	85.18	91.10	3.90	1.50	6.41
34	137.40	19.30	3.00	85.41	140.00	18.80	2.40	84.74	127.10	17.90	3.00	83.63	140.00	17.30	3.00	85.68	140.00	20.00	2.60	85.00	125.43	13.88	3.00	79.21	137.50	19.70	2.70	85.16	133.60	15.70	3.00	84.59	140.00	16.00	3.00	84.88	136.90	17.40	2.30	83.17
35	131.10	14.10	3.00	83.23	135.60	8.20	2.80	77.33	84.60	9.90	2.40	16.50	128.10	16.10	1.30	60.64	130.60	11.00	1.30	54.90	125.43	20.00	2.41	80.15	137.00	18.90	3.00	85.33	99.00	17.10	2.60	48.45	140.00	9.70	3.00	82.41	138.60	20.00	2.70	85.05
36	129.70	11.90	2.90	79.62	140.00	19.80	2.50	84.99	134.30	17.90	3.00	84.70	136.00	14.60	3.00	83.96	140.00	17.60	3.00	85.66	134.80	13.88	2.41	80.67	136.50	20.00	2.40	84.51	140.00	20.00	2.60	85.04	130.10	20.00	3.00	84.36	140.00	15.80	3.00	85.38
37	92.80	10.00	2.20	23.02	134.40	20.00	3.00	85.56	138.80	10.60	2.90	82.48	130.40	20.00	2.10	80.59	135.40	20.00	3.00	85.17	134.80	20.00	3.00	85.31	86.00	14.90	1.60	17.56	140.00	14.20	3.00	84.85	90.10	13.60	2.80	31.68	137.10	20.00	3.00	85.08
38	110.00	19.60	2.90	71.21	140.00	17.10	3.00	85.44	135.50	20.00	3.00	85.39	139.90	18.20	3.00	85.67	132.50	17.80	3.00	84.47	125.44	13.42	2.43	74.08	139.40	17.10	2.60	84.22	134.80	20.00	3.00	85.19	134.80	20.00	2.70	85.19	134.50	15.50	1.20	62.76
39	137.20	20.00	3.00	85.45	105.60	6.70	2.70	34.08	113.60	16.40	2.60	67.84	130.20	20.00	3.00	84.61	92.20	4.10	2.40	11.44	125.44	13.90	2.34	73.91	136.30	20.00	2.90	84.96	136.30	17.40	3.00	85.08	136.70	20.00	3.00	84.96	135.00	20.00	2.30	83.49
40	118.00	20.00	3.00	79.53	140.00	18.80	3.00	85.91	138.70	17.90	1.40	73.75	117.00	11.30	1.10	35.34	139.50	18.40	3.00	85.47	127.52	13.42	2.34	75.01	132.30	16.70	3.00	84.48	108.20	10.10	2.10	40.54	137.70	14.50	2.80	84.37	135.70	20.00	2.70	84.60
41	134.30	16.70	3.00	84.41	140.00	19.30	2.70	85.33	140.00	11.30	2.90	83.69	136.00	17.40	1.90	80.15	134.90	19.90	2.70	85.32	127.52	13.90	2.43	76.32	131.70	20.00	2.60	84.26	137.60	20.00	2.80	85.25	128.00	14.70	1.30	58.73	140.00	15.60	3.00	85.36
42	138.40	20.00	2.60	84.90	132.50	20.00	3.00	84.87	120.10	20.00	3.00	81.14	131.30	17.70	3.00	84.51	136.70	20.00	2.50	85.20	125.43	13.42	2.34	72.97	121.70	4.70	2.70	45.63	139.80	13.40	3.00	84.41	135.90	20.00	2.80	85.37	135.40	4.90	2.50	62.28
43	95.00	17.40	2.80	45.46	134.20	11.90	1.60	67.36	137.30	20.00	3.00	85.71	140.00	18.30	3.00	85.08	139.80	20.00	2.20	84.18	125.43	13.42	3.00	79.23	124.20	19.80	3.00	83.01	131.90	12.20	3.00	81.55	129.00	15.30	3.00	82.43	138.60	19.20	2.30	84.09
44	135.40	20.00	3.00	85.75	139.20	19.10	2.50	84.87	86.90	18.10	1.60	21.68	140.00	15.50	3.00	84.90	126.60	11.20	2.50	71.87	125.43	20.00	2.34	80.12	127.43	14.41	2.80	80.22	99.40	7.50	2.50	27.48	138.20	6.90	3.00	77.56	130.40	10.30	3.00	78.23
45	133.60	16.40	3.00	84.18	140.00	19.20	3.00	85.02	127.65	13.46	3.00	80.17	97.60	8.80	2.40	27.54	129.80	12.63	2.90	80.34	125.43	20.00	3.00	83.53	127.43	19.50	2.29	80.28	127.03	13.90	3.00	80.51	87.50	13.10	1.30	13.98	135.90	19.60	3.00	85.38
46	138.80	19.10	2.60	84.96	140.00	18.40	2.70	84.73	127.65	20.00	2.50	81.80	136.00	18.80	2.70	85.23	129.80	18.80	2.37	81.27	134.80	13.42	2.34	80.62	137.30	14.41	2.29	81.70	127.03	20.00	2.44	81.82	137.80	18.30	2.90	85.04	111.00	12.50	2.50	55.79

47	131.20	11.50	1.80	67.13	129.90	4.60	2.40	52.38	133.10	13.46	2.50	80.30	137.90	18.30	3.00	85.09	140.00	12.63	2.37	82.34	134.80	13.42	3.00	83.73	137.30	19.50	2.80	85.62	135.70	13.90	2.44	81.50	124.80	16.00	3.00	81.18	127.58	12.59	3.00	79.63
48	134.10	19.60	3.00	85.29	139.20	19.00	3.00	85.69	133.10	20.00	3.00	85.06	131.20	19.70	3.00	84.40	140.00	18.80	2.90	85.29	134.80	20.00	2.34	84.10	127.40	11.81	2.53	74.10	135.70	20.00	3.00	85.05	135.90	19.90	2.70	84.48	127.58	20.00	2.41	81.43
49	133.40	16.00	3.00	84.66	136.60	12.90	3.00	83.93	127.53	12.93	2.42	74.86	88.80	6.50	2.70	15.96	125.35	12.63	2.45	72.75	134.80	20.00	3.00	85.40	127.40	14.41	2.28	75.75	127.03	12.99	2.44	75.06	132.30	3.70	2.20	46.77	136.70	12.59	2.41	80.22
50	124.60	20.00	3.00	83.65	140.00	19.90	2.70	85.43	127.53	13.26	2.30	74.10	138.20	18.80	2.30	83.93	125.35	15.92	2.38	76.69	125.43	16.71	2.67	80.07	128.22	11.81	2.28	72.11	127.03	13.90	2.39	75.49	125.86	13.67	2.90	79.23	136.70	20.00	3.00	85.58
51	87.50	18.70	2.50	32.50	132.10	20.00	3.00	84.60	129.83	12.93	2.30	76.00	138.90	18.20	3.00	85.42	129.80	12.63	2.38	75.93	134.80	16.71	2.67	84.34	128.22	14.41	2.53	78.41	128.80	12.99	2.39	75.50	125.86	20.00	2.51	81.75	127.59	12.40	2.42	74.66
52	134.20	20.00	3.00	85.09	130.50	7.50	1.30	45.61	129.83	13.26	2.42	77.45	132.80	18.40	3.00	84.29	129.80	15.92	2.45	79.87	130.11	13.42	2.67	79.56	127.40	11.81	2.28	72.04	128.80	13.90	2.44	77.28	137.40	13.67	2.51	82.76	127.59	12.60	2.38	74.81
53	101.50	20.00	3.00	61.51	139.20	19.60	2.60	84.77	127.53	12.93	2.30	73.89	89.10	10.00	1.10	10.44	125.35	12.63	2.37	72.23	130.11	20.00	2.67	83.54	127.40	11.81	2.80	77.11	127.03	12.99	2.39	74.66	137.40	20.00	2.90	85.69	128.50	12.40	2.38	75.22
54	135.30	13.80	3.00	83.81	140.00	18.70	3.00	84.82	127.53	12.93	3.00	79.98	128.32	12.99	3.00	80.70	125.35	12.63	2.90	77.05	130.11	16.71	2.34	80.21	127.40	19.50	2.28	80.00	127.03	12.99	3.00	79.74	125.86	13.67	2.50	75.46	128.50	12.60	2.42	75.04
55	132.70	16.00	3.00	84.17	133.50	18.20	3.00	84.65	127.53	20.00	2.30	80.30	128.32	17.30	2.31	79.17	125.35	18.80	2.37	79.30	130.11	16.71	3.00	83.59	127.40	19.50	2.80	83.43	127.03	20.00	2.39	81.55	125.86	14.05	2.37	74.95	127.58	12.40	2.38	74.06
56	90.50	7.50	2.70	19.77	137.10	12.30	3.00	83.43	127.53	20.00	3.00	84.54	140.00	12.99	2.31	82.25	125.35	18.80	2.90	82.37	130.11	16.71	2.67	82.08	137.30	11.81	2.28	79.35	127.03	20.00	3.00	84.50	127.09	13.67	2.37	75.64	127.58	12.40	3.00	79.50
57	135.60	19.10	3.00	84.75	110.80	17.40	2.40	63.10	133.10	12.93	2.30	78.23	140.00	17.30	3.00	85.36	140.00	12.63	2.37	82.00	130.11	16.71	2.67	82.20	137.30	11.81	2.80	82.38	135.70	12.99	2.39	81.06	127.09	14.05	2.50	77.16	127.58	20.00	2.38	81.39
58	134.40	14.80	2.80	83.74	139.50	18.10	2.70	85.29	133.10	12.93	3.00	82.28	127.87	13.00	2.35	74.54	140.00	12.63	2.90	83.58	130.11	16.71	2.67	82.43	137.30	19.50	2.28	83.88	135.70	12.99	3.00	83.66	125.86	13.67	2.37	74.25	127.58	20.00	3.00	84.20
59	129.30	20.00	3.00	84.91	140.00	19.00	3.00	85.22	133.10	20.00	2.30	82.78	127.87	13.07	2.33	74.52	140.00	18.80	2.37	84.90	130.11	16.71	2.67	82.19	137.30	19.50	2.80	85.31	135.70	20.00	2.39	84.05	125.86	13.67	2.90	79.46	136.70	12.40	2.38	79.88
60	129.20	3.30	2.00	36.96	118.00	20.00	3.00	80.21	133.10	20.00	3.00	85.39	128.32	13.00	2.33	74.73	140.00	18.80	2.90	84.69	130.11	16.71	2.67	82.41	127.40	15.66	2.54	79.37	135.70	20.00	3.00	85.41	125.86	20.00	2.37	80.21	136.70	12.40	3.00	83.57
61	134.60	18.60	2.50	83.91	111.10	13.10	1.30	36.58	127.53	16.47	2.65	80.69	128.32	13.07	2.35	75.02	125.35	15.71	2.64	78.84	130.11	16.71	2.67	82.91	137.30	15.66	2.54	83.84	127.03	16.49	2.70	80.60	125.86	20.00	2.90	83.38	136.70	20.00	2.38	84.29
62	134.20	14.20	3.00	83.44	131.17	11.44	3.00	80.70	133.10	16.47	2.65	83.16	127.87	12.99	2.31	74.41	140.00	15.71	2.64	84.24	130.11	16.71	2.67	82.72	132.35	11.81	2.54	78.79	135.70	16.49	2.70	84.04	137.40	13.67	2.37	82.27	136.70	20.00	3.00	85.13
63	139.10	20.00	2.90	85.29	131.17	18.80	2.41	82.61	130.32	12.93	2.65	78.93	127.87	12.99	3.00	80.60	132.67	12.63	2.64	80.62	130.11	16.71	2.67	82.02	132.35	19.50	2.54	84.39	131.36	12.99	2.70	80.64	137.40	13.67	2.90	83.63	127.58	16.20	2.69	81.28
64	121.20	2.20	2.50	24.89	140.00	11.44	2.41	81.44	130.32	20.00	2.65	83.42	127.87	17.30	2.31	79.52	132.67	18.80	2.64	84.15	130.11	16.71	2.67	82.39	132.35	15.66	2.28	80.30	131.36	20.00	2.70	83.87	137.40	20.00	2.37	84.69	136.70	16.20	2.69	84.43
65	125.68	14.50	3.00	80.20	140.00	18.80	3.00	85.16	130.32	16.47	2.30	80.22	127.87	17.30	3.00	82.89	132.67	15.71	2.37	80.95	130.11	16.71	2.67	82.65	132.35	15.66	2.80	83.00	131.36	16.49	2.39	81.08	137.40	20.00	2.90	85.38	132.14	12.40	2.69	80.12
66	125.68	20.00	2.44	80.96	128.96	13.08	2.38	75.70	130.32	16.47	3.00	83.89	140.00	12.99	2.31	82.06	132.67	15.71	2.90	84.09					132.35	15.66	2.54	82.02	131.36	16.49	3.00	84.28	125.86	16.84	2.64	79.97	132.14	20.00	2.69	84.15
67	135.40	14.50	2.44	81.38	128.96	13.47	2.23	75.16	130.32	16.47	2.65	82.44	140.00	12.99	3.00	84.60	132.67	15.71	2.64	82.85					132.35	15.66	2.54	81.84	131.36	16.49	2.70	83.06	137.40	16.84	2.64	84.33	132.14	16.20	2.38	81.32
68	135.40	20.00	3.00	85.75	130.40	13.08	2.23	75.73	130.32	16.47	2.65	81.87	140.00	17.30	2.31	84.27	132.67	15.71	2.64	82.59					132.35	15.66	2.54	82.03	131.36	16.49	2.70	82.96	131.63	13.67	2.64	80.94	132.14	16.20	3.00	83.96
69	125.69	13.43	2.47	74.75	130.40	13.47	2.38	77.80	130.32	16.47	2.65	81.83	140.00	17.30	3.00	85.06	132.67	15.71	2.64	82.97					132.35	15.66	2.54	81.68	131.36	16.49	2.70	82.97	131.63	20.00	2.64	84.00	132.14	16.20	2.69	83.47
70	125.69	14.52	2.37	75.27	128.96	11.44	2.23	71.71	130.32	16.47	2.65	81.83	127.87	15.14	2.66	80.15	132.67	15.71	2.64	82.43					132.35	15.66	2.54	82.21	131.36	16.49	2.70	83.16	131.63	16.84	2.37	81.44	132.14	16.20	2.69	83.32
71	128.95	13.43	2.37	76.61	128.96	11.44	3.00	79.11	130.32	16.47	2.65	82.40	140.00	15.14	2.66	83.92	132.67	15.71	2.64	82.71					132.35	15.66	2.54	82.25	131.36	16.49	2.70	82.55	131.63	16.84	2.90	84.22	132.14	16.20	2.69	83.32
72	128.95	14.52	2.47	79.13	128.96	18.80	2.23	80.54	130.32	16.47	2.65	82.40	133.94	12.99	2.66	81.65	132.67	15.71	2.64	82.23					132.35	15.66	2.54	81.71	131.36	16.49	2.70	82.81	131.63	16.84	2.64	82.86	132.14	16.20	2.69	82.52

73	125.68	13.43	2.37	73.92	128.96	18.80	3.00	83.73	130.32	16.47	2.65	81.66	133.94	17.30	2.66	83.23	132.67	15.71	2.64	82.86					132.35	15.66	2.54	81.63	131.36	16.49	2.70	82.70	131.63	16.84	2.64	82.83	132.14	16.20	2.69	82.54
74	125.68	13.43	3.00	79.08	140.00	11.44	2.23	79.82	130.32	16.47	2.65	82.57	133.94	15.14	2.31	80.94	132.67	15.71	2.64	82.34					132.35	15.66	2.54	81.79	131.36	16.49	2.70	82.99	131.63	16.84	2.64	83.36	132.14	16.20	2.69	83.07
75	125.68	20.00	2.37	79.86	140.00	11.44	3.00	83.57	130.32	16.47	2.65	82.41	133.94	15.14	3.00	84.24	132.67	15.71	2.64	82.47					132.35	15.66	2.54	82.07	131.36	16.49	2.70	82.80	131.63	16.84	2.64	82.73	132.14	16.20	2.69	83.01
76	125.68	20.00	3.00	83.82	140.00	18.80	2.23	83.71	130.32	16.47	2.65	81.88	133.94	15.14	2.66	82.53	132.67	15.71	2.64	82.40									131.36	16.49	2.70	82.43	131.63	16.84	2.64	83.12	132.14	16.20	2.69	82.65
77	135.40	13.43	2.37	80.61	140.00	18.80	3.00	85.91					133.94	15.14	2.66	82.79																	131.63	16.84	2.64	82.79	132.14	16.20	2.69	83.19
78	135.40	13.43	3.00	83.83	128.96	15.12	2.62	80.84					133.94	15.14	2.66	82.92																	131.63	16.84	2.64	82.84	132.14	16.20	2.69	83.30
79	135.40	20.00	2.37	84.03	140.00	15.12	2.62	83.80					133.94	15.14	2.66	82.97																	131.63	16.84	2.64	82.81				
80	135.40	20.00	3.00	85.19	134.48	11.44	2.62	79.79					133.94	15.14	2.66	83.06																	131.63	16.84	2.64	82.70				
81	125.68	16.72	2.68	80.14	134.48	18.80	2.62	84.38					133.94	15.14	2.66	83.02																	131.63	16.84	2.64	82.88				
82	135.40	16.72	2.68	84.04	134.48	15.12	2.23	80.77					133.94	15.14	2.66	82.70																								
83	130.54	13.43	2.68	80.62	134.48	15.12	3.00	84.10					133.94	15.14	2.66	83.24																								
84	130.54	20.00	2.68	83.92	134.48	15.12	2.62	83.07					133.94	15.14	2.66	82.95																								
85	130.54	16.72	2.37	80.98	134.48	15.12	2.62	83.10					133.94	15.14	2.66	83.01																								
86	130.54	16.72	3.00	84.44	134.48	15.12	2.62	82.58																																
87	130.54	16.72	2.68	82.06	134.48	15.12	2.62	82.85																																
88	130.54	16.72	2.68	82.49	134.48	15.12	2.62	83.43																																
89	130.54	16.72	2.68	82.39	134.48	15.12	2.62	82.70																																
90	130.54	16.72	2.68	82.73	134.48	15.12	2.62	83.15																																
91	130.54	16.72	2.68	82.56	134.48	15.12	2.62	82.61																																
92	130.54	16.72	2.68	82.94	134.48	15.12	2.62	82.89																																
93	130.54	16.72	2.68	82.41	134.48	15.12	2.62	82.72																																
94	130.54	16.72	2.68	82.35																																				
95	130.54	16.72	2.68	82.82																																				
96	130.54	16.72	2.68	82.32																																				

Table 6-18. Results for the noise-based Bayesian optimisation of the simulated deoxofluorination.

Run	1				2				3				4				5				6				7				8				9				10			
Exp	T (oC)	tres (min)	[B] (M)	Yield (%)	T (oC)	tres (min)	[B] (M)	Yield (%)	T (oC)	tres (min)	[B] (M)	Yield (%)	T (oC)	tres (min)	[B] (M)	Yield (%)	T (oC)	tres (min)	[B] (M)	Yield (%)	T (oC)	tres (min)	[B] (M)	Yield (%)	T (oC)	tres (min)	[B] (M)	Yield (%)	T (oC)	tres (min)	[B] (M)	Yield (%)	T (oC)	tres (min)	[B] (M)	Yield (%)	T (oC)	tres (min)	[B] (M)	Yield (%)
1	80.8	1.97	2.11	2.63	107.92	9.85	1.3	27.15	84.96	13.29	2.75	14.19	80.11	3.27	1.34	82.97	125.97	18.93	2.53	72.33	105.26	18.3	2.4	26.11	93.69	2.72	1.79	70.68	92.9	15.67	1.16	12.26	108.32	19.65	2.05	22.02	122.11	2.54	2.44	6.83
2	108.58	8.86	2.43	42.16	127.71	4.4	2.09	43.74	138.83	1.31	1.91	24.41	105.43	5.35	2.32	2.66	106.54	10.14	1.46	80.90	95.6	9.96	2.67	57.08	137.72	12.36	1.14	6.28	131.58	10.67	2.31	18.88	80.27	1.03	1.27	57.98	115.42	13.8	2.64	28.31
3	127.96	7.17	1.58	47.80	103.61	5.18	2.8	26.76	112.21	9	2.33	25.03	126.43	12.62	1.47	25.28	101.73	13.52	2.94	29.00	80.31	2.47	2.02	30.45	80.6	6.73	2.46	59.35	108.23	1.57	2.64	73.68	129.41	4.95	2.29	0.81	128.95	19.11	1.6	66.26
4	134.01	18.2	2.61	84.02	89.24	12.98	2.35	25.55	120.04	18.51	1.11	46.29	98.88	19.01	1.88	57.77	90.07	3.67	1.37	49.94	138.72	14.01	1.11	3.05	112.2	14.6	2.82	9.77	125.49	19.03	1.69	11.27	117.15	13.87	2.81	52.06	92.73	11.86	1.82	72.02
5	95.43	14.72	1.37	23.33	130.56	19.59	1.43	70.33	120.7	19.95	2.98	49.84	139.9	19.55	1.65	41.02	139.53	19.85	2.08	5.31	138.82	13.66	1.46	61.79	118.99	19.9	2.7	65.71	139.9	19.55	1.65	70.56	122.72	19.87	2.18	70.39	118.99	19.9	2.7	22.35
6	139.9	19.55	1.65	80.23	139.53	19.85	2.08	83.72	125.41	19.94	2.65	81.38	139.96	5.25	2.8	80.23	134.94	18.08	2.73	83.72	138.59	19.97	2.89	71.08	115.25	19.92	1.13	78.30	139.99	12.4	2	80.23	120.82	18.08	2.22	76.20	115.25	19.92	1.13	78.30
7	139.53	19.85	2.08	83.72	138.9	19.79	1.96	82.87	130.19	19.86	2.78	82.00	139.9	19.55	1.65	71.90	138.59	19.97	2.89	84.51	138.59	19.97	2.89	85.35	130.19	19.86	2.78	46.67	135.7	19.08	2.98	78.81	138.59	19.97	2.89	73.24	138.59	19.97	2.89	46.67
8	135.8	19.92	2.54	84.63	139.72	19.56	1.18	70.38	131.43	19.82	2.97	84.19	139.41	14.29	2.87	80.23	131.43	19.82	2.97	85.35	139.97	17.76	2.99	85.35	131.43	19.82	2.97	84.19	133.93	19.89	1.35	85.21	135.76	19.97	1.07	85.35	139.53	19.85	2.08	85.35
9	131.43	19.82	2.97	84.86	139.53	19.85	2.08	83.72	134.82	19.96	2.48	84.86	138.42	13.45	1.03	84.51	80.86	19.83	2.56	84.86	138.59	19.97	2.89	85.29	131.43	19.82	2.97	84.86	139.97	17.76	2.99	71.40	138.59	19.97	2.89	64.51	139.53	19.85	2.08	83.72
10	133.27	19.8	2.73	84.65	139.53	19.85	2.08	83.72	138.59	19.97	2.89	84.30	139.53	19.85	2.08	57.70	135.8	19.92	2.54	25.99	110	10.5	2	85.35	131.43	19.82	2.97	84.86	139.97	17.76	2.99	85.29	138.59	19.97	2.89	85.35	139.53	19.85	2.08	83.72
11	134.82	19.96	2.48	84.30	139.53	19.85	2.08	83.72	138.59	19.97	2.89	85.35	139.53	19.85	2.08	83.72	139.97	17.76	2.99	84.63	139.55	10.98	2.92	42.63	131.43	19.82	2.97	84.86	139.97	17.76	2.99	85.29	138.59	19.97	2.89	85.35	110	10.5	2	83.72
12	134.26	19.52	2.92	85.07	118.61	19.82	2.39	74.98	138.59	19.97	2.89	85.35	139.53	19.85	2.08	83.72	137.6	18.5	2.93	85.29	128.9	18.49	2.73	83.18	138.59	19.97	2.89	84.86	136.68	16.97	2.97	85.29	139.97	17.76	2.99	85.35	125	5.75	2.5	42.63
13	138.59	19.97	2.89	85.35	118.69	19.99	1.96	69.51	138.59	19.97	2.89	85.35	139.66	19.26	2.77	83.72	117.8	19.82	3	85.21	118.12	19.97	2.88	83.10	124.27	19.73	2.93	85.35	139.23	1.92	2.98	84.97	130.24	16.99	2.98	85.29	124.85	19.99	2.98	53.39
14	121.65	19.99	2.87	81.29	129.3	18.44	2.08	78.69	138.59	19.97	2.89	85.35	139.66	19.26	2.77	85.23	120.56	17	2.93	79.56	123.24	19.98	2.51	79.10	117.69	3.67	2.97	82.76	137.14	19.23	2.96	46.43	117.8	19.82	3	83.83	139.91	16.99	2.88	83.30
15	138.59	19.97	2.89	85.35	123.97	16.88	2.36	76.38	107.15	19.79	2.96	85.35	139.66	19.26	2.77	85.23	129.55	16.99	2.5	78.72	117.8	19.82	3	79.85	135.76	19.97	1.07	36.97	119.04	15.18	2.82	85.29	117.89	19.99	2.63	79.56	133.69	16.99	2.92	85.11
16	138.59	19.97	2.89	85.35	139.92	16.89	2.25	83.59	139.91	16.99	2.88	68.48	139.34	19.65	2.71	85.23	121.78	19.95	2.48	81.28	133.37	17.04	2.97	79.56	131.43	19.82	2.97	64.51	119.06	19.98	2.68	74.48	135.35	17	2.99	76.88	122.67	16.98	2.53	84.45
17	126.74	19.79	2.87	83.49	133.92	16.89	2.3	81.92	128.9	18.49	2.73	85.11	139.34	19.65	2.71	85.19	117.84	17.4	2.5	78.60	121.25	16.99	3	84.52	139.97	17.76	2.99	84.86	127.09	15.15	2.99	78.32	131.53	19.98	2.98	84.86	127.82	17	2.97	77.10
18	138.59	19.97	2.89	85.35	118.65	16.97	2.02	66.71	117.94	17.07	2.92	83.10	139.66	19.26	2.77	85.19	124.07	16.98	2.52	72.85	132.35	19.92	2.46	79.66	139.96	19.93	2.63	85.29	129.36	19.97	2.95	81.58	125.4	16.99	2.87	84.94	131.53	19.98	2.98	82.98
19	117.87	17.02	2.79	75.35	125.06	19.99	2.33	79.55	134.45	17.73	2.87	76.55	118.79	19.99	2.55	85.23	128.54	19.97	2.5	78.07	125.52	17.01	2.51	83.64	128.9	18.49	2.73	85.17	123.47	19.98	3	84.49	127.18	19.96	2.54	81.42	117.94	17.07	2.92	84.94
20	117.89	19.99	2.63	76.88	134.39	19.96	2.37	83.79	123.35	19.24	2.6	84.66	129.35	18.19	2.68	76.96	123.78	19.85	2.99	82.53	118.06	17.18	2.48	79.04	118.26	17.01	2.97	83.10	132.17	15.1	2.97	82.83	117.84	17.4	2.5	82.18	135.35	19.84	2.47	76.55
21	139.91	16.99	2.88	85.11	114.24	19.82	2.72	74.24	124.48	17.02	2.96	80.06	118.77	16.43	2.75	82.89	126.89	17.08	2.97	82.87	127.31	19.94	2.99	72.58	134.51	16.98	2.79	77.17	122.9	15.13	2.58	83.50	134.89	19.95	2.99	72.85	137.3	16.98	2.52	84.32
22	124.07	16.98	2.52	78.07	114.31	17.01	2.45	67.77	117.89	19.99	2.63	81.42	124.13	20	2.83	75.12	132.35	17.09	3	82.66	136.79	17.1	2.92	84.12	121.78	17.11	2.98	84.29	121.07	17.84	3	75.71	139.96	19.93	2.63	85.27	121.78	19.95	2.48	84.05

23	129.09	17	2.93	83.30	122.01	19.99	2.65	80.16	131.07	17.06	2.51	76.88	134.68	17.28	2.84	82.35	138.41	17.01	2.46	84.40	130.24	16.99	2.98	84.93	134.45	17.73	2.87	80.00	130.14	17.27	2.56	80.25	132.43	16.98	2.47	85.17	117.8	19.82	3	78.60
24	136.88	17.03	2.5	83.93	114.33	19.99	2.23	68.77	126.12	18.11	2.8	82.11	132.82	20	2.54	84.52	119.28	20	2.46	84.13	139.9	19.93	2.51	83.83	117.89	19.99	2.63	84.66	139.77	19.91	2.93	82.11	122.8	17.06	2.99	82.43	130.66	17.03	2.49	79.56
25	121.25	16.99	3	79.66	114.24	19.82	2.72	74.24	137.3	16.98	2.52	82.09	124.03	19.09	2.52	84.07	133.37	19.88	2.49	76.50	133.9	19.99	2.47	85.01	139.98	17.03	2.65	76.88	125.84	17.99	3	85.39	120.44	19.89	2.98	80.64	127.84	19.97	2.99	81.81
26	131.73	16.98	2.96	84.14	130.17	19.98	2.7	83.96	139.9	19.93	2.51	84.05	126.69	17.73	2.76	79.77	136.19	17.03	2.99	83.99	117.94	17.07	2.92	84.09	131.17	17	2.61	84.81	132.26	19.93	2.97	82.83	138.41	17.01	2.46	81.22	125.52	17.01	2.51	84.25
27	124.26	19.93	2.47	80.14	110.65	19.82	3	73.09	121.25	16.99	3	85.01	137.34	19.55	2.43	81.92	120.01	18.66	3	84.96	120.96	19.77	2.99	76.55	139.96	19.93	2.63	82.66	118.84	17.36	2.53	85.00	124.85	19.99	2.98	84.13	139.91	16.99	2.88	79.04
28	120.02	18.49	2.46	75.71	110.74	17.03	2.73	66.76	127.82	17	2.97	79.66	118.79	19.99	2.55	84.50	122.24	17.98	2.46	80.19	123.37	16.98	2.98	81.49	126.37	17.01	2.99	85.17	133.94	17	2.5	74.12	127.92	17.03	2.5	83.30	120.15	17.04	2.98	85.11
29	129.36	19.94	2.49	82.81	127.09	17	2.98	82.73	127.84	19.97	2.99	82.98	118.77	16.43	2.75	76.96	129.15	18.66	2.99	77.07	130.05	19.95	2.95	80.86	117.8	19.82	3	82.47	126.42	19.96	2.96	83.09	119.74	17.05	2.96	80.50	139.99	19.64	3	78.81
30	126.37	17.01	2.99	82.47	136.98	17.03	2.94	84.97	138.55	18.36	2.46	84.25	118.79	19.99	2.55	75.12	139.96	19.93	2.63	84.17	117.89	19.99	2.63	84.62	126.8	19.96	2.46	79.56	121.15	19.96	2.54	83.77	129.81	18.98	2.47	78.39	123.02	18.44	2.99	85.41
31	139.99	19.64	3	85.41	120.1	14.5	3	76.07	136.42	18.66	2.97	84.50	122.02	16.39	3	76.96	126.94	20	2.91	85.17	136.55	19.63	2.46	76.88	137.3	16.98	2.52	81.53	123.18	17.46	2.51	78.68	139.98	17.03	2.65	82.43	135.86	18.29	2.99	81.74
32	134.65	16.78	2.96	84.65	139.91	14.49	2.88	84.67	119.24	18.73	3	85.20	132.02	16.82	2.6	79.65	131.12	18.39	2.46	83.76	139.91	16.99	2.88	84.48	135.09	19.97	2.97	84.05	127.89	17.07	2.5	77.75	133.27	18.71	2.47	84.81	133.11	18.62	2.55	85.13
33	119	19.81	3	80.37	130.48	14.49	2.91	82.28	132.89	17.02	2.98	79.71	121.37	18.64	2.92	82.81	133.92	17.02	2.49	82.58	127.82	17	2.97	85.11	121.78	19.95	2.48	85.26	138.49	16.96	2.5	80.47	139.98	17.03	2.65	83.49	129.74	18.54	2.99	83.75
34	127.49	16.72	2.57	80.47	116.51	14.55	2.81	70.75	132.72	18.73	2.47	84.45	122.7	17.51	2.56	80.55	118.26	17.01	2.97	83.06	125.08	19.95	2.47	82.98	119.85	17.84	2.47	78.60	132.24	17.92	3	84.23	127.88	18.54	3	84.81	124.26	16.71	2.98	84.28
35	138.1	16.7	2.98	85.06	135.51	14.5	2.52	82.43	137.06	19.98	2.49	83.33	133.35	19.32	2.88	77.85	117.89	19.99	2.63	77.17	125.08	18.36	2.99	80.65	123.86	18.01	2.46	74.93	134.44	19.96	2.98	84.62	136.68	18.57	2.48	83.79	138.63	18.21	2.55	81.19
36	119.38	16.76	2.62	74.84	125.97	14.51	2.52	76.92	123.08	17.02	2.48	84.69	128.57	19.97	2.38	84.87	122.8	17.06	2.99	76.88	138.55	18.36	2.46	82.68	128.65	17.01	2.54	78.36	118.88	18.9	2.99	85.23	125.07	18.56	2.47	84.30	133.57	20	2.97	84.69
37	124	18.6	2.99	82.34	139.91	14.49	2.88	84.67	125.73	17.04	2.47	76.99	138.67	16.6	2.72	81.83	125.08	19.95	2.47	80.64	135.35	17	2.99	84.50	133.26	19.98	2.53	81.12	128.43	18.52	3	79.55	117.89	19.99	2.63	79.66	126.42	18.59	2.98	85.15
38	132.83	16.71	2.56	82.87	123.53	14.54	2.96	78.65	119.46	17.04	2.53	78.85	118.77	16.43	2.75	84.69	139.99	17.18	2.46	80.65	120.44	17.05	2.48	84.86	138.55	18.36	2.46	84.14	124.79	16.93	2.55	83.94	122.23	18.93	3	76.88	139.92	16.71	2.9	83.28
39	130.47	18	2.55	82.62	131.48	17.63	2.99	84.37	139.95	18.69	2.95	74.30	118.79	19.99	2.55	75.12	124.52	18.33	2.99	84.43	132.43	16.98	2.47	74.70	139.96	19.93	2.63	84.50	138.5	20	2.51	78.79	128.91	19.91	3	81.69	119.07	19.99	2.69	85.09
40	122.32	16.8	2.56	76.89	120.96	17.17	2.99	79.55	133.21	20	2.96	85.33	129.23	19.98	2.93	76.96	133.01	18.6	3	82.42	121.78	19.95	2.48	82.43	117.85	18.3	2.51	85.17	130.97	19.07	2.52	84.90	139.26	18.69	2.46	84.50	135.17	16.85	2.56	78.36
41	137.09	18.41	2.99	85.23	132.96	14.58	2.97	83.50	135.77	16.99	2.49	85.11	128.73	16.37	2.89	84.40	127.87	18.35	2.46	84.89	139.9	19.93	2.51	78.60	127.84	19.97	2.99	74.00	127.76	19.96	2.57	83.13	119.06	18.85	2.49	84.66	120.44	18.86	2.97	83.64
42	128.5	19.96	2.99	84.40	139.99	18.98	3	85.38	117.85	18.3	2.51	83.59	118.77	16.43	2.75	82.67	134.89	19.95	2.99	81.11	134.89	19.95	2.99	85.01	139.91	16.99	2.88	84.25	136.31	19.97	2.53	82.58	131.66	18.17	2.99	75.48	132.17	16.7	2.97	80.47
43	139.99	16.92	2.55	84.59	125.66	19.99	2.98	83.60	130.71	18.07	3	74.00	118.79	19.99	2.55	75.12	119.38	17.19	2.47	85.27	132.15	18.53	2.98	85.27	136.75	18.79	2.98	85.11	120.39	16.99	2.56	84.69	136.83	19.97	2.99	84.55	137	19.97	2.99	84.17
44	127.3	18.5	2.55	81.49	136.57	19.97	2.95	85.33	122.55	19.99	3	84.36	128.02	18.41	2.39	76.96	139.96	19.93	2.63	73.75	117.89	19.99	2.63	84.71	132.48	18.19	2.49	85.25	120.01	18.83	2.52	75.40	116.32	17.45	2.95	85.37	126.32	20	2.6	85.38
45	120.39	20	2.55	78.22	133.24	19.99	3	85.15	121.9	18.5	3	82.43	125.36	17.05	2.96	80.73	125.73	17.04	2.47	85.17	122.58	18.22	2.49	76.88	126.68	18.18	2.46	83.11	138.6	18.58	2.51	76.60	116.4	19.99	2.68	75.80	119.15	18.03	2.57	82.14
46	125.33	19.99	2.96	83.40	119.83	19.92	2.99	80.89	129.55	16.99	2.5	81.23	131.88	19.98	2.98	81.90	130.38	19.9	2.46	78.85	119.26	18.78	2.48	77.79	117.87	17.02	2.79	80.32	124.34	18.59	2.99	84.67	133.69	17.08	2.97	76.09	129.62	19.99	2.62	75.52
47	124.26	16.71	2.98	81.19	127.64	20	2.92	84.02	126.8	19.96	2.46	81.28	136.01	18.87	2.64	84.98	118.21	18.73	2.47	83.00	128.08	19.98	2.48	75.54	132.89	17.02	2.98	75.35	122.45	18.8	2.5	82.49	137.29	17.1	2.99	84.59	124.38	18.65	2.55	83.50
48	138.63	18.21	2.55	84.69	139.88	19.85	2.95	85.40	138.52	17.33	3	81.53	118.77	16.43	2.75	84.65	130.58	16.99	2.96	74.40	134.71	18.33	2.47	82.24	117.89	19.99	2.63	84.45	133.58	18.69	2.5	78.33	116.46	18.81	2.55	85.07	138.1	16.7	2.98	79.95

49	130.33	16.73	2.56	81.90	123.41	19.99	3	82.81	119.28	20	2.46	85.19	130.68	17.96	2.47	75.12	136.83	19.97	2.99	83.84	128.9	17.08	2.46	83.73	139.96	19.93	2.63	76.88	135.76	17.74	2.49	83.70	133.05	19.96	2.6	73.58	129.52	16.76	2.98	85.06
50	122.07	18.49	2.98	81.24	135.14	19.95	2.99	85.29	121.78	19.95	2.48	76.50	120.04	19.77	2.8	82.26	116.31	19.82	3	85.37	124.85	19.99	2.98	80.77	122.55	19.99	3	85.17	129.13	16.98	3	83.89	123.83	17.16	2.56	84.29	123.22	19.97	2.57	83.49
51	119.21	18.28	2.99	79.25	119.2	19.97	2.89	79.93	135.95	19.9	2.97	78.60	120.7	18.07	2.9	79.72	116.38	17.02	2.82	78.44	138	16.98	2.97	83.30	120.87	18.74	2.99	82.43	118.81	19.82	3	83.53	121.51	16.99	2.6	78.43	131.59	18.45	3	80.33
52	123.49	19.98	3	82.84	129.18	16.93	3	83.52	124.63	18.28	2.47	85.31	131.35	19.89	2.58	79.55	116.36	18.3	2.58	74.21	127.02	18.32	2.47	85.09	124.25	17.1	2.99	80.74	139.99	17.13	2.49	80.25	126.33	18.83	2.98	76.81	121.25	17.9	2.56	84.64
53	135.99	16.71	2.57	83.85	137.59	17	2.99	85.08	133.65	17.05	2.46	79.21	139.99	16.61	2.35	83.80	120.91	19.9	2.97	73.32	130.63	18.51	2.47	80.69	130.71	18.07	3	81.53	122.1	16.94	3	84.50	123.77	18.78	3	83.36	136.4	16.76	2.99	77.12
54	125.45	18.12	2.56	80.28	119.17	17.23	3	78.34	128.61	19.86	2.47	82.86	126.56	19.92	2.98	83.95	139.21	18.69	2.54	81.40	117.94	17.07	2.92	82.48	137.06	19.98	2.49	84.36	131.92	17.1	2.51	80.18	134.89	18.59	2.98	82.35	121.61	16.72	2.94	84.93
55	132.13	18.36	2.56	83.40	131.95	19.98	2.99	85.00	128.61	19.86	2.47	82.31	126.56	19.92	2.98	83.86	139.21	18.69	2.54	84.83	117.94	17.07	2.92	76.55	137.06	19.98	2.49	84.69	131.92	17.1	2.51	82.47	134.89	18.59	2.98	85.08	121.61	16.72	2.94	79.30

Table 6-19. Results for the targeted EI Bayesian optimisation of the simulated deoxofluorination.

Run	**1**				**2**				**3**				**4**				**5**				**6**				**7**				**8**				**9**				**10**			
Exp	T (oC)	tres (min)	[B] (M)	Yield (%)	T (oC)	tres (min)	[B] (M)	Yield (%)	T (oC)	tres (min)	[B] (M)	Yield (%)	T (oC)	tres (min)	[B] (M)	Yield (%)	T (oC)	tres (min)	[B] (M)	Yield (%)	T (oC)	tres (min)	[B] (M)	Yield (%)	T (oC)	tres (min)	[B] (M)	Yield (%)	T (oC)	tres (min)	[B] (M)	Yield (%)	T (oC)	tres (min)	[B] (M)	Yield (%)	T (oC)	tres (min)	[B] (M)	Yield (%)
1	97.72	14.17	2.51	40.40	132.45	4.05	2.30	50.52	82.62	9.55	1.36	8.58	136.21	9.12	1.67	64.84	106.97	3.18	1.62	12.76	119.47	5.69	1.20	25.90	105.61	19.45	1.35	40.50	92.28	12.88	1.16	15.71	124.32	5.48	1.17	29.20	94.87	3.06	2.25	9.28
2	109.86	9.28	1.06	23.44	114.35	10.04	1.16	31.34	120.51	17.45	1.42	57.43	83.70	13.42	1.82	16.13	82.09	14.36	2.28	19.23	105.66	3.53	2.68	20.61	100.08	12.69	2.11	35.86	109.32	5.63	2.14	28.63	88.18	9.95	2.70	22.11	81.57	12.64	1.12	8.72
3	133.58	16.29	1.70	74.73	103.48	13.00	2.07	40.38	100.20	1.20	2.04	4.61	122.27	2.60	2.70	31.43	136.48	7.51	1.17	48.38	81.46	9.03	1.73	9.62	133.76	10.02	1.72	65.46	131.58	18.02	1.75	75.71	102.69	14.28	1.65	35.15	121.24	10.41	1.94	56.10
4	91.49	3.50	1.84	7.29	84.81	6.42	1.65	8.05	114.51	13.31	2.51	62.58	97.19	6.16	1.29	11.59	123.77	11.57	1.91	61.46	134.42	14.93	2.31	80.93	81.49	2.60	2.47	4.19	84.46	10.94	2.53	18.88	128.33	18.10	2.44	81.03	106.41	19.58	1.57	46.32
5	116.52	8.14	2.79	55.27	119.18	19.53	2.80	78.95	135.87	7.01	2.66	72.86	108.63	18.82	2.33	61.52	95.96	17.94	3.00	49.70	101.33	17.59	1.85	41.90	120.62	6.36	2.89	55.14	122.87	2.35	2.97	32.19	106.53	2.96	2.05	14.58	133.06	7.54	2.73	72.52
6	139.90	19.55	1.65	80.23	139.53	19.85	2.08	83.72	139.66	19.26	2.77	85.23	136.34	9.34	1.52	62.30	123.95	11.50	2.20	66.15	134.20	14.79	2.32	80.78	139.95	1.37	1.70	24.49	139.53	19.85	2.08	83.72	138.59	19.97	2.89	85.35	130.43	10.00	2.66	75.11
7	139.90	19.55	1.65	80.23	139.53	19.85	2.08	83.72	139.90	19.55	1.65	80.23	133.96	6.25	2.80	69.91	125.64	14.24	2.43	75.40	131.43	19.82	2.97	84.86	128.23	19.93	2.12	79.56	139.22	19.75	1.02	64.77	132.94	19.99	1.67	76.94	139.90	19.55	1.65	80.23
8	139.53	19.85	2.08	83.72	110.00	10.50	2.00	42.63	139.90	19.55	1.65	80.23	121.65	19.99	2.87	81.29	125.41	19.94	2.65	82.00	138.59	19.97	2.89	85.35	134.82	19.96	2.48	84.30	139.53	19.85	2.08	83.72	139.53	19.85	2.08	83.72	133.27	19.80	2.73	84.65
9	139.53	19.85	2.08	83.72	130.19	19.86	2.78	84.19	139.51	15.50	2.95	84.94	138.59	19.97	2.89	85.35	131.43	19.82	2.97	84.86	135.76	19.97	1.07	64.51	139.22	19.75	1.02	64.77	139.53	19.85	2.08	83.72	121.65	19.99	2.87	81.29	132.06	19.90	1.10	62.44
10	126.18	19.67	2.97	83.60	134.82	19.96	2.48	84.30	135.70	19.08	2.98	85.21	138.59	19.97	2.89	85.35	130.19	19.86	2.78	84.19	139.66	19.26	2.77	85.23	131.43	19.82	2.97	84.86	139.53	19.85	2.08	83.72	139.97	17.76	2.99	85.29	139.97	17.76	2.99	85.29
11	135.80	19.92	2.54	84.63	125.41	19.94	2.65	82.00	137.60	18.50	2.93	85.21	110.00	10.50	2.00	42.63	138.59	19.97	2.89	85.35	139.66	19.26	2.77	85.23	132.87	18.22	2.94	84.69	139.53	19.85	2.08	83.72	126.18	19.67	2.97	83.60	139.97	17.76	2.99	85.29
12	138.59	19.97	2.89	85.35	139.51	15.50	2.95	84.94	137.22	10.25	1.02	51.15	138.59	19.97	2.89	85.35	134.82	19.96	2.48	84.30	139.97	17.76	2.99	85.29	133.27	19.80	2.73	84.65	139.96	5.25	2.80	71.90	139.95	1.37	1.70	24.49	139.97	17.76	2.99	85.29
13	131.43	19.82	2.97	84.86	139.97	17.76	2.99	85.29	133.14	1.76	2.96	36.71	125.00	5.75	2.50	53.39	138.59	19.97	2.89	85.35	124.27	19.73	2.93	82.76	134.26	19.52	2.92	85.07	120.70	19.95	2.98	81.38	80.23	19.93	1.62	16.97	139.97	17.76	2.99	85.29

14	118.99	19.90	2.70	78.30	136.68	16.97	2.97	84.97	135.55	7.75	2.90	76.65	95.00	15.25	1.50	25.39	138.59	19.97	2.89	85.35	133.54	14.99	2.95	83.77	129.99	16.09	3.00	83.40	139.87	11.80	2.81	83.42	113.18	19.98	1.97	63.29	139.97	17.76	2.99	85.29
15	129.71	1.33	2.96	26.63	137.14	19.23	2.96	85.29	108.47	19.93	2.21	61.03	128.90	11.70	2.99	79.21	138.59	19.97	2.89	85.35	135.56	13.76	1.10	58.30	130.97	17.56	2.96	84.16	110.35	19.96	1.31	45.90	80.19	1.30	2.22	1.77	139.97	17.76	2.99	85.29
16	129.56	19.78	1.16	62.28	134.70	19.36	2.67	84.64	95.29	19.98	2.46	45.43	134.62	8.71	2.94	78.26	138.59	19.97	2.89	85.35	133.92	15.42	2.14	79.66	133.33	15.65	2.99	84.11	139.95	1.37	1.70	24.49	98.12	19.96	2.81	53.99	138.59	19.97	2.89	85.35
17	123.35	15.29	2.98	79.43	138.59	19.97	2.89	85.35	124.60	9.78	2.99	72.25	135.55	7.75	2.90	76.65	127.05	11.11	2.91	76.43	134.33	15.97	2.49	82.67	131.66	16.45	2.99	84.02	134.48	9.32	2.80	78.17	135.56	13.76	1.10	58.30	131.69	10.83	2.96	79.73
18	120.62	13.71	2.94	75.02	139.87	11.80	2.81	83.42	126.74	19.79	2.87	83.49	139.77	1.07	2.54	28.19	109.65	19.89	2.77	69.71	127.29	19.98	1.46	68.51	133.80	17.26	2.90	84.51	125.86	15.08	1.06	49.64	107.15	19.79	2.96	68.48	130.37	9.51	2.94	76.56
19	139.68	17.49	1.03	63.66	135.70	19.08	2.98	85.21	80.23	19.93	1.62	16.97	129.56	19.78	1.16	62.28	139.77	1.07	2.54	28.19	133.92	15.42	2.14	79.66	128.65	13.61	1.02	49.01	126.74	19.79	2.87	83.49	118.07	13.92	2.91	72.53	139.77	1.07	2.54	28.19
20	120.70	19.95	2.98	81.38	131.43	19.82	2.97	84.86	117.65	1.01	2.86	12.53	115.25	19.92	1.13	46.67	117.12	19.88	2.94	78.71	133.92	15.42	2.14	79.66	126.74	19.79	2.87	83.49	139.85	8.23	1.18	54.13	123.01	15.96	1.04	47.46	120.70	19.95	2.98	81.38
21	131.66	16.45	2.99	84.02	135.24	18.65	2.88	84.98	139.89	1.67	2.35	36.62	121.99	13.85	2.99	76.83	139.92	13.59	2.59	83.59	133.92	15.42	2.14	79.66	134.91	11.26	2.96	81.87	135.47	15.28	2.89	84.19	112.29	14.60	2.97	67.45	123.29	1.03	1.57	9.31
22	118.10	19.59	2.88	78.80	139.28	18.86	2.93	85.31	129.37	14.26	2.94	81.78	102.82	19.95	1.48	40.30	100.94	19.98	1.22	32.56	133.92	15.42	2.14	79.66	129.06	19.39	2.91	84.14	134.82	19.96	2.48	84.30	128.90	11.70	2.99	79.21	114.99	15.81	2.99	72.77
23	134.07	17.50	2.28	82.19	117.12	19.88	2.94	78.71	114.50	19.84	2.76	74.84	126.00	15.57	1.01	48.66	125.07	1.61	2.91	25.73	133.92	15.42	2.14	79.66	134.02	12.73	3.00	82.86	80.19	1.30	2.22	1.77	139.60	10.09	2.99	82.77	115.25	19.92	1.13	46.67
24	121.10	11.22	2.99	71.61	120.70	19.95	2.98	81.38	102.83	14.89	2.98	54.87	134.26	19.52	2.92	85.07	81.88	1.04	2.82	1.99	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	80.23	19.93	1.62	16.97	89.27	19.83	2.92	41.23	125.26	15.80	2.88	80.46
25	139.72	19.56	1.18	70.38	118.99	19.90	2.70	78.30	100.94	19.98	1.22	32.56	80.19	1.30	2.22	1.77	115.56	14.16	2.97	70.69	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	116.79	13.81	2.82	70.03	134.62	8.71	2.94	78.26	95.29	19.98	2.46	45.43
26	121.99	13.85	2.99	76.83	118.10	19.59	2.88	78.80	108.59	14.21	2.98	61.99	80.23	19.93	1.62	16.97	134.43	13.52	2.97	83.38	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	116.37	19.80	1.06	45.88	117.12	19.88	2.94	78.71	139.88	10.61	1.36	64.44
27	80.23	19.93	1.62	16.97	118.99	19.90	2.70	78.30	128.91	8.19	1.01	38.14	115.56	14.16	2.97	70.69	89.27	19.83	2.92	41.23	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	101.57	19.74	2.85	59.46	133.33	15.65	2.99	84.11	107.24	10.86	2.98	52.39
28	104.70	19.99	2.67	62.14	118.99	19.90	2.70	78.30	121.99	13.85	2.99	76.83	139.89	15.68	2.86	84.88	108.59	14.21	2.98	61.99	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	106.29	14.67	2.92	58.88	116.29	8.88	2.98	59.63	99.98	15.53	2.99	51.81
29	95.29	19.98	2.46	45.43	118.99	19.90	2.70	78.30	95.21	13.50	2.96	40.25	126.18	19.67	2.97	83.60	94.59	10.41	2.99	32.99	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	128.16	10.17	2.96	76.15	104.33	19.78	1.01	31.26	136.85	16.02	1.01	59.07
30	139.96	5.25	2.80	71.90	118.99	19.90	2.70	78.30	120.70	19.95	2.98	81.38	118.94	9.18	2.91	63.21	139.68	17.49	1.03	63.66	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	121.10	11.22	2.99	71.61	122.94	11.82	2.91	73.95	80.86	19.83	2.56	25.99
31	139.60	10.09	2.99	82.77	118.99	19.90	2.70	78.30	139.56	8.02	2.85	79.55	109.22	1.36	2.98	11.60	120.00	16.65	3.00	78.44	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	98.07	1.26	2.85	5.95	131.43	19.82	2.97	84.86	126.74	19.79	2.87	83.49
32	111.24	19.65	1.12	41.67	118.99	19.90	2.70	78.30	89.27	19.83	2.92	41.23	105.85	15.12	2.89	58.73	117.45	8.50	2.95	59.63	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	92.84	19.91	2.99	47.63	114.29	18.38	2.98	75.14	113.96	11.75	1.08	32.37
33	139.95	1.37	1.70	24.49	118.99	19.90	2.70	78.30	132.06	19.90	1.10	62.44	129.83	7.84	1.08	40.03	80.42	19.63	2.65	26.04	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	111.39	13.10	2.99	63.68	137.71	13.98	2.97	84.35	111.18	19.35	2.96	72.81
34	133.70	9.58	2.98	79.36	118.99	19.90	2.70	78.30	80.19	1.30	2.22	1.77	133.54	14.99	2.95	83.77	105.69	19.64	2.98	66.59	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	129.37	14.26	2.94	81.78	111.18	19.35	2.96	72.81	80.19	1.30	2.22	1.77
35	117.65	1.01	2.86	12.53	118.99	19.90	2.70	78.30	118.94	9.18	2.91	63.21	139.60	10.09	2.99	82.77	102.16	13.95	2.98	51.90	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	107.15	19.79	2.96	68.48	94.78	14.21	2.96	40.97	120.14	14.45	2.92	75.40
36	80.19	1.30	2.22	1.77	118.99	19.90	2.70	78.30	107.98	8.82	3.00	47.71	129.71	1.33	2.96	26.63	139.36	9.00	2.93	81.32	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	132.68	4.03	2.97	58.99	95.06	1.06	2.91	4.38	102.13	19.57	2.95	61.11
37	127.07	9.96	1.09	42.72	118.99	19.90	2.70	78.30	117.82	15.20	2.98	74.74	124.60	9.78	2.99	72.25	121.10	11.22	2.99	71.61	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	114.29	18.38	2.98	75.14	129.71	1.33	2.96	26.63	107.08	1.13	2.89	8.59
38	100.94	19.98	1.22	32.56	118.99	19.90	2.70	78.30	103.72	11.38	1.02	20.89	110.94	15.59	2.91	66.84	129.99	16.09	3.00	83.40	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	116.29	8.88	2.98	59.63	108.90	10.29	2.96	53.06	125.68	10.68	2.97	75.01
39	87.83	14.65	2.98	31.95	118.99	19.90	2.70	78.30	105.69	19.64	2.98	66.59	139.22	19.75	1.02	64.77	132.45	8.85	2.95	76.96	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	137.99	8.27	2.99	79.94	117.65	1.01	2.86	12.53	89.27	19.83	2.92	41.23

40	112.29	14.60	2.97	67.45	118.99	19.90	2.70	78.30	134.43	13.52	2.97	83.38	89.27	19.83	2.92	41.23	112.01	9.94	2.99	56.93	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	121.91	16.04	2.92	78.71	139.85	14.78	1.11	63.58	108.07	15.38	2.88	62.31
41	105.85	15.12	2.89	58.73	118.99	19.90	2.70	78.30	111.41	18.07	2.89	70.80	129.99	16.09	3.00	83.40	112.85	18.53	2.95	73.57	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	99.98	15.53	2.99	51.81	80.22	13.22	2.96	20.51	137.99	8.27	2.99	79.94
42	137.69	7.99	1.03	46.65	118.99	19.90	2.70	78.30	113.41	8.57	2.89	53.61	118.56	17.92	2.97	78.35	121.79	19.89	1.27	57.88	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	139.89	15.68	2.86	84.88	126.63	15.15	2.93	81.03	129.99	16.09	3.00	83.40
43	110.02	19.44	2.99	71.90	118.99	19.90	2.70	78.30	89.81	5.15	2.98	14.88	111.18	19.35	2.96	72.81	137.37	17.03	2.94	85.01	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	128.38	5.64	2.93	62.39	137.69	7.99	1.03	46.65	91.98	13.29	3.00	35.54
44	89.27	19.83	2.92	41.23	118.99	19.90	2.70	78.30	126.00	15.57	1.01	48.66	98.12	19.96	2.81	53.99	95.06	1.06	2.91	4.38	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	103.71	10.14	2.98	45.16	129.20	8.06	2.96	72.35	136.09	9.72	1.01	48.53
45	103.15	1.21	2.95	7.68	118.99	19.90	2.70	78.30	130.97	17.56	2.96	84.16	139.78	6.14	1.04	43.29	132.68	4.03	2.97	58.99	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	131.43	19.82	2.97	84.86	108.07	15.38	2.88	62.31	139.08	13.24	2.90	84.19
46	136.28	14.11	2.93	84.02	118.99	19.90	2.70	78.30	126.02	4.91	2.99	56.22	105.69	19.64	2.98	66.59	127.10	6.98	2.86	65.60	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	131.09	8.09	1.14	43.81	99.67	9.69	2.97	38.06	117.45	8.50	2.95	59.63
47	116.22	12.19	2.93	67.40	118.99	19.90	2.70	78.30	139.87	11.80	2.81	83.42	136.85	16.02	1.01	59.07	132.70	17.05	1.03	57.50	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	87.13	19.33	2.94	37.47	136.39	18.29	1.06	62.99	117.12	19.88	2.94	78.71
48	135.54	5.50	2.95	70.06	118.99	19.90	2.70	78.30	123.81	17.78	2.97	81.61	103.71	10.14	2.98	45.16	120.70	19.95	2.98	81.38	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	124.12	7.29	2.94	64.02	133.30	11.77	2.97	81.67	100.99	10.78	3.00	43.07
49	127.76	12.36	2.98	79.24	118.99	19.90	2.70	78.30	131.97	6.58	3.00	71.03	99.98	15.53	2.99	51.81	111.57	15.13	1.03	33.73	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	124.31	12.57	2.97	76.88	120.00	16.65	3.00	78.44	132.92	3.01	2.90	50.33
50	80.12	10.54	2.82	16.10	118.99	19.90	2.70	78.30	123.53	19.95	1.05	52.92	114.89	7.89	2.99	54.38	104.63	9.27	2.94	43.64	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	110.48	17.25	2.95	69.18	115.08	11.97	2.96	65.83	123.53	19.95	1.05	52.92
51	124.12	7.29	2.94	64.02	118.99	19.90	2.70	78.30	80.22	13.22	2.96	20.51	89.34	1.09	2.99	3.39	137.23	11.41	2.98	82.98	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	113.62	1.12	2.94	11.73	119.91	10.98	2.95	69.49	139.83	6.44	2.89	76.55
52	129.20	8.06	2.96	72.35	118.99	19.90	2.70	78.30	129.68	10.32	2.97	77.68	117.12	19.88	2.94	78.71	113.99	1.56	3.00	16.36	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	97.45	19.56	1.03	24.51	125.40	9.30	2.99	71.97	124.12	7.29	2.94	64.02
53	103.71	10.14	2.98	45.16	118.99	19.90	2.70	78.30	87.83	14.65	2.98	31.95	131.43	19.82	2.97	84.86	122.75	8.04	2.88	64.23	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	123.53	19.95	1.05	52.92	100.96	17.00	2.95	55.47	107.15	19.79	2.96	68.48
54	100.46	17.72	2.95	55.85	118.99	19.90	2.70	78.30	135.06	16.59	1.09	61.19	136.55	13.42	2.84	83.50	136.84	5.34	2.99	71.08	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	136.84	5.34	2.99	71.08	129.56	19.78	1.16	62.28	104.03	15.46	2.96	57.46
55	123.53	19.95	1.05	52.92	118.99	19.90	2.70	78.30	99.06	19.94	2.91	56.66	118.96	12.79	1.09	39.78	99.06	19.94	2.91	56.66	133.92	15.42	2.14	79.66	128.23	19.93	2.12	79.56	132.74	12.25	2.87	81.36	139.83	6.44	2.89	76.55	128.38	5.64	2.93	62.39

Table 6-20. $RMSE_n$ results for the deoxofluorination simulated case study.

Optimum Region $RMSE_n$	Noise Region $RMSE_n$	Methodology
51.36072448	10.60359378	SNOBFIT-DoE
15.41256609	10.30828514	SNOBFIT-DoE
31.4021659	11.58368661	SNOBFIT-DoE
84.3441399	9.67246762	SNOBFIT-DoE
137.764197	7.929553445	SNOBFIT-DoE
89.41699502	3.831319812	SNOBFIT-DoE
28.47810596	8.644667915	SNOBFIT-DoE
31.0634344	11.46238161	SNOBFIT-DoE
17.41028494	3.8622197	SNOBFIT-DoE
42.32373715	8.468055642	SNOBFIT-DoE
78.73774119	2.420825313	EI-Noise
76.0603868	3.404580344	EI-Noise
211.6670574	4.030473452	EI-Noise
34.38771366	1.624941411	EI-Noise
37.50290879	1.42787295	EI-Noise
25.44757654	1.439058029	EI-Noise
175.5368941	3.486451823	EI-Noise
70.29880729	1.686762177	EI-Noise
133.5842593	5.012800342	EI-Noise
24.41790321	2.228339909	EI-Noise
17.6659198	3.98897242	EI-Target
52.05655347	8.0485144	EI-Target
19.47892806	6.996673596	EI-Target
60.52354799	14.36807077	EI-Target
55.74710934	5.538282851	EI-Target
53.23412157	7.015413939	EI-Target
50.23408458	9.056531931	EI-Target
13.00770469	12.27160407	EI-Target
23.67310791	7.11173691	EI-Target
14.4096846	10.75831396	EI-Target

Table 6-21. Results for the noise based exploration for the simulated photochemical system. Stages of algorithm differentiated by colour: green – initial optimisation and blue – exploratory stage.

Run	1			2			3			4			5			6			7			8			9			10		
Exp	t_{res} (min)	O_2 Eq.	Yield (%)	t_{res} (min)	O_2 Eq.	Yield (%)	t_{res} (min)	O_2 Eq.	Yield (%)	t_{res} (min)	O_2 Eq.	Yield (%)	t_{res} (min)	O_2 Eq.	Yield (%)	t_{res} (min)	O_2 Eq.	Yield (%)	t_{res} (min)	O_2 Eq.	Yield (%)	t_{res} (min)	O_2 Eq.	Yield (%)	t_{res} (min)	O_2 Eq.	Yield (%)	t_{res} (min)	O_2 Eq.	Yield (%)
1	8.60	.41	7.55	7.00	2.57	41.57	8.81	.59	58.55	7.35	3.05	30.26	17.66	3.95	53.91	15.70	3.43	47.21	7.49	4.58	44.55	24.03	3.44	50.38	26.95	4.60	56.29	9.29	3.77	38.29
2	.56	.90	5.95	2.93	4.13	54.92	.40	.08	40.05	15.51	4.85	54.91	7.15	4.43	42.31	1.54	2.23	3.97	1.66	2.47	4.64	16.26	2.09	34.82	1.16	1.93	0.35	20.38	2.00	35.18
3	7.75	.02	6.56	5.04	2.84	41.23	.41	.26	4.57	5.08	1.98	16.97	19.31	1.52	29.18	11.17	1.15	16.99	14.93	2.61	38.36	8.70	4.65	45.26	15.29	1.06	21.74	5.45	4.89	39.56
4	.93	.43	4.90	.49	1.63	16.23	6.28	.83	32.04	23.31	1.45	29.65	29.88	2.87	37.79	20.57	4.96	59.06	23.03	1.12	26.66	1.50	2.97	5.04	8.50	3.96	39.53	28.81	2.94	41.18
5	5.43	.82	1.48	1.17	4.52	44.82	29.73	2.67	37.23	29.16	3.97	46.64	2.63	2.07	8.35	27.33	3.21	46.40	25.76	4.15	55.39	29.47	1.44	30.28	18.89	3.19	49.01	13.28	1.37	23.68
6	9.98	.08	7.00	1.13	5.00	59.25	22.95	4.99	59.39	22.55	4.98	58.96	29.95	4.91	47.11	21.28	1.21	26.41	29.95	4.91	47.21	22.04	4.99	59.77	29.95	4.91	46.28	29.95	4.91	46.86
7	9.95	.91	6.70	3.86	5.00	60.49	22.04	4.99	59.27	20.22	4.99	59.80	29.95	4.91	47.01	29.77	4.97	48.13	29.95	4.91	46.49	20.22	4.99	60.06	29.95	4.91	46.76	29.95	4.91	46.87
8	9.95	.91	6.02	9.77	4.97	47.36	18.40	5.00	60.02	21.13	5.00	59.34	29.95	4.91	47.06	23.86	5.00	60.14	29.95	4.91	46.38	25.68	4.99	59.64	25.68	4.99	59.62	27.95	4.99	54.33
9	9.95	.91	6.90	2.95	4.99	59.53	19.84	4.98	59.52	19.31	4.99	60.78	13.85	4.99	51.40	17.49	4.99	58.76	29.95	4.91	46.92	17.03	5.00	58.64	22.04	4.99	59.47	24.77	5.00	60.00
10	9.95	.91	7.65	4.77	5.00	60.53	25.68	4.99	59.69	18.40	5.00	59.50	23.45	4.98	59.58	25.68	4.99	59.58	27.95	4.99	54.91	23.86	5.00	59.93	23.86	5.00	59.91	26.59	5.00	58.25
11	5.68	.99	9.99	4.31	4.98	59.93	24.77	5.00	60.36	25.68	4.99	60.36	22.95	4.99	60.09	14.76	4.99	52.74	21.92	4.77	58.29	19.31	4.99	60.33	18.40	5.00	59.90	22.95	4.99	59.85
12	6.16	.98	8.54	8.40	5.00	60.17	12.94	5.00	49.01	24.77	5.00	60.54	22.95	4.99	60.17	24.77	5.00	60.80	18.66	5.00	59.99	29.32	4.99	50.31	19.84	4.98	59.63	23.86	5.00	60.92
13	2.95	.99	9.50	5.68	4.99	59.55	24.31	4.98	60.05	1.11	4.98	5.88	22.95	4.99	59.49	19.31	4.99	60.15	15.88	5.00	56.09	18.40	5.00	59.67	24.77	5.00	60.58	17.03	5.00	58.66
14	4.77	.00	0.18	9.31	4.99	60.54	18.94	4.97	59.49	26.59	5.00	59.12	19.31	4.99	59.92	24.31	4.98	60.79	20.08	4.25	56.69	24.77	5.00	60.42	15.67	4.99	55.67	19.31	4.99	60.28
15	7.03	.00	8.00	6.58	4.98	56.99	17.49	4.99	59.24	16.96	4.24	54.95	17.49	4.99	58.38	18.40	5.00	59.62	17.27	4.25	55.42	13.85	4.99	51.07	24.77	5.00	60.13	20.67	4.99	59.29

16	4.77	.00	0.59	9.84	4.98	60.16	24.70	3.69	52.67	20.59	4.24	56.31	21.13	5.00	60.02	22.04	4.99	59.24	20.62	5.00	59.73	21.13	5.00	59.28	24.77	5.00	60.02	18.40	5.00	60.25
17	4.77	.00	0.31	8.94	4.97	60.40	28.47	4.98	52.89	16.80	5.00	57.92	25.68	4.99	60.46	24.77	5.00	60.89	15.88	4.26	53.16	22.95	4.99	59.21	24.77	5.00	60.48	13.85	4.99	50.52
18	4.77	.00	0.80	.11	4.98	4.68	27.35	4.25	53.16	22.18	4.24	55.57	24.77	5.00	60.53	24.77	5.00	60.49	21.44	4.25	55.83	19.84	4.98	60.13	24.77	5.00	60.20	27.43	4.26	53.02
19	4.31	.98	0.13	5.07	4.95	60.13	21.29	4.25	55.90	18.78	4.24	56.77	25.26	4.97	60.39	24.77	5.00	61.09	17.11	5.00	58.41	24.31	4.98	59.92	25.02	4.25	56.53	23.65	4.25	56.19
20	3.85	.99	0.88	7.68	4.24	55.88	22.92	4.25	55.78	17.86	4.24	56.07	24.77	5.00	60.36	24.77	5.00	60.60	18.67	4.25	56.76	27.81	4.25	51.90	28.45	4.25	49.92	21.63	4.25	55.73
21	0.22	.99	0.32	0.32	4.24	56.50	27.14	5.00	57.15	16.43	4.25	54.17	26.59	5.00	58.17	24.77	5.00	60.76	19.57	5.00	60.18	21.08	4.26	56.08	28.31	5.00	53.53	25.67	4.25	56.14
22	9.31	.99	9.59	6.43	4.25	54.17	28.45	4.25	49.92	19.68	4.24	56.82	20.22	4.99	59.89	24.77	5.00	59.88	22.50	4.26	55.73	27.61	5.00	55.83	21.29	4.25	55.90	20.30	4.26	56.62
23	1.13	.00	9.27	1.77	4.24	55.63	26.06	4.25	55.68	21.46	4.24	55.78	25.26	4.97	59.83	24.77	5.00	60.72	22.50	4.98	59.22	26.41	4.25	55.14	22.92	4.25	55.78	21.88	5.00	59.21
24	8.28	.25	0.48	8.96	4.24	56.83	21.08	5.00	59.47	17.58	4.99	59.01	28.11	4.25	51.03	24.77	5.00	60.50	17.89	5.00	59.37	22.68	4.25	55.69	26.18	4.25	55.52	24.59	4.25	56.62
25	2.22	.25	5.60	7.40	4.99	58.78	23.92	4.25	56.39	21.89	5.00	59.20	24.20	4.25	56.47	27.96	5.00	54.72	16.65	4.26	54.60	25.13	4.25	56.50	21.08	5.00	59.47	22.52	4.25	55.64
26	5.02	.25	6.53	2.07	5.00	59.20	25.14	4.28	56.72	16.41	4.96	57.10	21.75	4.25	55.68	23.15	4.25	55.89	21.36	5.00	59.34	28.48	4.26	49.91	24.04	4.25	56.43	25.70	5.00	59.89
27	7.96	.00	4.72	1.05	.24	5.99	2.05	.25	5.64	8.25	.57	8.12	6.18	.25	5.52	1.08	.26	6.08	9.40	.25	6.93	8.42	.99	3.10	7.58	.25	2.55	6.57	.25	4.89
28	1.08	.26	6.08	9.59	.24	6.85	7.84	.00	5.09	7.32	.60	7.23	8.42	.99	3.10	8.48	.26	9.91	7.92	.25	6.20	3.85	.25	6.30	7.14	.00	7.15	7.27	.00	6.81
29	6.88	.25	4.26	7.02	.25	5.07	6.44	.00	8.75	0.74	.67	8.40	2.92	.25	5.78	5.25	.25	6.44	0.75	.25	6.29	6.68	.00	8.26	2.95	.00	9.50	0.95	.25	6.11
30	3.62	.25	6.17	0.51	.99	9.77	3.65	.00	9.98	9.88	.67	8.88	1.08	.26	6.08	6.79	.26	4.50	6.66	.00	7.67	1.83	.25	5.70	2.05	.25	5.64	1.27	.99	9.35
31	2.01	.00	9.20	8.31	.24	6.49	8.47	.98	2.89	9.12	.56	8.54	7.16	.25	3.64	6.91	.00	7.73	0.08	.00	0.01	7.11	.25	3.73	6.34	.99	8.92	5.17	.56	8.28
32	8.47	.98	2.89	2.20	.50	7.02	6.70	.25	4.67	6.41	.96	7.10	7.50	.00	6.14	4.20	.25	6.47	1.98	.25	5.65	4.52	.25	6.58	8.48	.26	9.91	3.11	.26	5.93
33	7.14	.00	7.15	7.91	.70	8.35	8.47	.98	2.89	1.77	.59	7.48	5.13	.25	6.50	2.22	.25	5.60	9.14	.00	0.17	5.76	.25	6.09	1.08	.26	6.08	7.43	.98	6.30
34	5.95	.25	5.84	1.54	.66	7.92	4.53	.32	7.01	8.85	.99	0.07	2.01	.00	9.20	2.95	.00	9.50	2.00	.00	9.20	3.27	.25	5.96	7.73	.00	5.43	0.28	.96	9.78

35	2.92	.25	5.78	6.93	.70	7.10	7.93	.25	1.56	6.68	.60	6.30	8.48	.26	9.91	8.47	.98	2.89	8.27	.63	8.40	6.18	.72	8.02	5.59	.26	6.28	6.12	.60	7.55
36	7.58	.25	2.55	0.71	.60	8.11	1.60	.65	7.86	0.17	.25	6.62	3.60	.25	6.18	7.62	.25	2.46	7.48	.66	7.72	1.52	.65	7.87	3.48	.25	6.10	4.22	.62	8.60
37	4.32	.25	6.52	8.65	.60	8.54	2.49	.62	7.66	2.20	.70	7.95	4.03	.00	0.22	1.31	.00	9.36	1.02	.64	8.08	8.47	.98	2.89	6.88	.25	4.26	2.39	.73	8.16
38	3.65	.00	9.98	9.99	.57	8.41	3.38	.53	7.63	6.43	.25	4.17	2.37	.25	5.66	6.06	.25	5.68	9.75	.58	8.56	2.40	.65	7.79	8.47	.98	2.89	3.39	.70	8.50
39	8.48	.26	9.91	6.41	.64	6.01	1.06	.60	7.88	6.44	.00	7.26	5.68	.25	6.15	6.21	.00	9.17	6.64	.60	6.23	7.14	.72	6.00	1.55	.00	9.28	6.98	.58	5.77
40	1.63	.25	5.75	9.31	.57	8.60	7.54	.64	4.57	1.19	.56	7.63	6.65	.25	4.72	1.63	.25	5.75	9.03	.25	6.91	8.07	.64	2.96	4.55	.25	6.56	1.34	.59	7.66
41	6.68	.00	8.26	1.62	.00	9.24	5.63	.54	7.83	8.31	.24	6.49	7.62	.25	2.46	3.62	.25	6.17	2.50	.56	7.34	5.27	.69	8.88	3.41	.00	9.81	0.61	.59	8.14
42	2.48	.00	9.28	7.89	.00	9.38	8.22	.65	2.49	9.74	.99	0.13	7.96	.00	4.72	7.38	.00	6.50	0.39	.58	8.22	3.41	.72	8.58	2.48	.00	9.28	5.11	.25	6.51
43	5.48	.25	6.31	7.40	.54	7.08	6.29	.61	7.32	8.01	.00	9.49	4.69	.26	6.63	4.75	.25	6.60	8.32	.00	9.76	4.15	.61	8.49	2.45	.25	5.62	7.43	.98	6.30
44	1.55	.00	9.28	2.18	.24	5.57	5.23	.98	0.21	9.24	.25	6.92	7.03	.00	7.44	2.68	.25	5.69	8.29	.25	6.52	1.05	.65	8.10	7.93	.25	1.56	2.08	.25	5.63
45	6.41	.25	5.14	6.98	.00	8.20	6.88	.65	6.31	7.44	.25	5.64	6.09	.00	9.36	2.48	.00	9.28	7.53	.99	8.95	2.48	.00	9.28	6.68	.00	8.26	5.25	.00	0.27
46	1.08	.26	6.08	1.43	.25	5.82	4.01	.67	8.73	1.05	.24	5.99	1.55	.00	9.28	3.41	.00	9.81	8.99	.64	8.86	6.70	.56	6.29	5.42	.00	0.15	4.10	.25	6.43
47	7.61	.00	5.83	1.14	.60	7.81	1.55	.00	9.28	0.65	.99	9.70	2.41	.00	9.25	8.48	.26	9.91	1.00	.00	9.51	2.22	.25	5.60	1.75	.25	5.68	2.44	.00	9.27
48	8.47	.98	2.89	0.32	.71	8.82	2.48	.00	9.28	1.52	.00	9.29	1.06	.60	7.88	5.68	.25	6.15	1.06	.25	6.05	8.47	.98	2.89	4.16	.69	8.90	7.02	.25	3.95
49	3.27	.25	5.96	6.44	.00	7.26	4.91	.64	8.76	7.76	.63	7.93	8.48	.70	1.80	7.21	.26	3.57	2.50	.98	9.22	4.76	.61	8.62	1.05	.65	8.10	4.35	.00	0.37
50	3.30	.00	9.73	0.68	.25	6.33	3.38	.25	6.02	7.16	.00	8.48	4.27	.65	8.78	1.08	.00	9.47	7.56	.25	5.79	8.47	.98	2.89	6.00	.63	7.86	0.31	.00	9.90
51	7.30	.25	3.29	9.87	.24	6.76	2.45	.25	5.62	8.67	.64	8.70	2.04	.58	7.42	8.47	.98	2.89	1.56	.62	7.71	7.56	.55	4.13	6.56	.26	4.93	6.15	.00	9.28
52	4.55	.25	6.56	2.20	.98	9.15	5.53	.25	6.29	0.32	.61	8.39	3.32	.56	7.74	7.99	.25	1.39	9.77	.25	6.84	6.21	.00	9.17	3.15	.60	7.85	6.12	.25	5.62
53	5.42	.00	0.15	8.05	.24	6.25	6.09	.00	9.36	1.86	.24	5.60	6.39	.63	7.23	1.08	.26	6.08	6.91	.25	4.94	1.55	.00	9.28	7.91	.66	3.55	2.85	.59	7.65

54	2.56	.25	5.66	6.69	.24	4.53	1.07	.27	6.16	6.43	.25	4.17	7.70	.63	4.06	6.45	.61	7.00	0.44	.26	6.51	2.93	.58	7.61	5.23	.58	8.36	1.84	.60	7.55
55	1.08	.00	9.47	8.26	.62	8.35	8.47	.98	2.89	2.20	.98	9.15	1.40	.25	5.83	1.74	.68	7.91	1.63	.00	9.25	8.47	.98	2.89	7.30	.25	3.29	7.42	.63	4.85

Table 6-22. Results for the targeted expected improvement for the simulated photochemical system. Stages of algorithm differentiated by colour: green – initial optimisation and blue – exploratory stage.

Run	1			2			3			4			5			6			7			8			9			10		
Exp.	t_{res} (min)	O_2 Eq.	Yield (%)	t_{res} (min)	O_2 Eq.	Yield (%)	t_{res} (min)	O_2 Eq.	Yield (%)	t_{res} (min)	O_2 Eq.	Yield (%)	t_{res} (min)	O_2 Eq.	Yield (%)	t_{res} (min)	O_2 Eq.	Yield (%)	t_{res} (min)	O_2 Eq.	Yield (%)	t_{res} (min)	O_2 Eq.	Yield (%)	t_{res} (min)	O_2 Eq.	Yield (%)	t_{res} (min)	O_2 Eq.	Yield (%)
1	.51	.27	1.84	6.06	.10	8.28	8.74	.95	4.87	2.85	.26	5.75	.52	.18	2.17	7.75	.92	5.78	.39	.58	3.85	9.67	.87	3.93	3.76	.23	8.80	0.07	.97	4.97
2	6.02	.41	1.20	3.24	.53	7.66	.90	.71	.42	5.96	.40	1.17	6.70	.60	6.48	5.09	.25	1.27	.63	.14	1.03	.63	.04	1.26	7.32	.46	8.34	9.97	.10	7.12
3	2.80	.09	4.77	.95	.50	.68	.06	.89	6.65	0.02	.05	3.71	7.74	.46	0.27	.87	.34	4.67	2.92	.52	4.31	2.72	.20	6.95	.51	.74	8.35	7.93	.67	2.65
4	.81	.14	3.13	1.97	.34	8.59	5.55	.02	7.48	0.89	.54	1.30	9.95	.25	6.38	.62	.18	3.47	1.26	.04	4.62	3.02	.23	0.22	5.37	.53	3.26	.00	.90	0.02
5	2.81	.41	6.32	0.39	.99	4.89	8.01	.49	8.44	.58	.05	0.06	.28	.02	5.77	0.41	.87	5.34	9.49	.13	0.72	.71	.63	6.08	.91	.72	2.59	.15	.70	2.98
6	8.40	.00	0.79	8.40	.00	9.42	2.95	.99	9.50	9.95	.91	6.35	9.95	.91	6.43	9.98	.15	6.23	9.95	.91	6.50	9.95	.91	6.97	8.40	.00	0.12	1.65	.97	9.04
7	0.22	.99	9.98	7.49	.99	9.23	2.04	.99	9.39	9.95	.91	6.89	9.95	.91	7.76	8.40	.00	0.08	3.45	.98	9.82	9.95	.91	7.22	9.95	.91	6.76	3.45	.98	9.30
8	7.03	.00	8.76	1.13	.00	9.86	7.49	.99	8.88	9.77	.97	8.11	9.95	.91	6.96	1.13	.00	9.14	8.40	.00	0.06	9.95	.91	6.74	0.22	.99	9.53	2.55	.98	9.35
9	8.94	.97	0.04	0.22	.99	9.64	9.31	.99	9.99	1.13	.00	9.15	9.95	.91	6.71	9.84	.98	9.99	1.13	.00	8.86	9.77	.97	7.19	4.77	.00	0.17	2.76	.76	2.48
10	9.77	.97	7.23	2.95	.99	9.41	9.77	.97	8.16	8.40	.00	0.18	9.95	.91	6.49	9.77	.97	7.90	7.03	.00	8.18	6.58	.98	7.15	2.95	.99	9.29	0.22	.99	0.08
11	3.86	.00	9.67	2.04	.99	9.04	3.85	.99	0.27	.11	.98	.36	8.61	.34	4.55	9.31	.99	0.78	4.77	.00	0.74	8.40	.00	9.54	1.13	.00	9.07	9.31	.99	0.39
12	9.98	.08	6.59	9.77	.97	8.03	4.77	.00	0.83	4.77	.00	9.90	5.03	.96	3.90	4.77	.00	0.76	5.68	.99	0.61	0.22	.99	9.76	5.68	.99	9.45	6.59	.00	8.08

13	8.29	.02	2.67	4.77	.00	0.63	5.68	.99	9.54	3.45	.98	9.73	3.23	.40	3.66	3.86	.00	9.78	5.26	.97	0.59	9.31	.99	0.45	9.31	.99	0.17	8.40	.00	9.52
14	.76	.99	7.11	4.31	.98	0.46	4.31	.98	9.99	7.03	.00	7.89	7.49	.99	9.19	5.68	.99	0.27	.11	.98	.88	9.31	.99	0.35	4.31	.98	0.29	9.77	.97	8.16
15	7.07	.98	7.33	5.68	.99	9.83	0.67	.99	9.79	9.31	.99	9.94	9.98	.08	8.32	8.94	.97	0.24	9.31	.99	0.35	0.22	.99	0.01	3.86	.00	0.35	4.77	.00	9.83
16	3.48	.03	5.04	4.77	.00	9.95	8.40	.00	0.39	2.94	.00	8.29	8.41	.99	3.89	7.03	.00	7.95	4.31	.98	0.01	9.31	.99	0.15	7.49	.99	9.34	9.84	.98	1.00
17	4.76	.99	2.95	4.77	.00	1.33	4.96	.95	0.86	7.95	.99	5.17	2.95	.99	9.63	5.67	.99	5.77	1.12	.00	6.59	9.31	.99	0.11	8.94	.97	9.80	4.76	.99	2.97
18	.57	.99	6.19	4.77	.00	0.23	5.26	.97	0.02	9.89	.02	5.28	0.37	.97	7.24	6.37	.15	3.68	4.20	.00	6.94	9.31	.99	9.87	2.94	.00	9.08	9.84	.98	0.05
19	.11	.98	.16	4.77	.00	0.72	2.04	.99	8.99	1.93	.01	4.10	5.19	.12	8.59	8.08	.60	1.98	7.70	.02	1.74	5.67	.99	5.64	3.86	.00	9.97	9.84	.98	0.60
20	4.76	.62	9.11	4.77	.00	9.63	4.77	.00	0.38	4.76	.99	3.22	0.22	.99	0.49	7.95	.99	4.39	9.98	.08	7.45	5.67	.99	5.85	4.31	.98	9.99	9.84	.98	0.48
21	8.36	.49	6.24	4.77	.00	0.61	4.77	.00	0.94	.01	.97	7.35	2.94	.00	8.46	1.12	.00	7.63	3.98	.98	1.56	5.67	.99	6.01	4.76	.99	2.94	9.31	.99	0.22
22	0.21	.31	0.08	4.77	.00	0.12	4.77	.00	0.41	7.08	.49	9.90	.11	.97	6.87	6.47	.00	1.77	.37	.37	5.18	5.67	.99	5.32	9.98	.08	6.78	1.93	.01	4.89
23	7.26	.39	9.07	6.47	.00	1.73	4.77	.00	0.06	6.36	.15	4.89	.11	.98	.56	9.98	.08	7.41	7.95	.99	5.11	5.68	.99	9.93	7.95	.99	4.39	3.74	.01	0.47
24	5.22	.93	4.04	.30	.99	7.39	4.77	.00	0.36	.57	.99	4.96	2.15	.35	9.54	3.83	.98	6.13	.01	.97	7.66	3.40	.99	9.42	4.65	.00	1.51	1.12	.00	7.39
25	.11	.97	6.78	9.98	.08	6.67	4.77	.00	0.64	3.84	.49	0.91	.19	.00	2.05	.48	.00	6.81	6.60	.12	7.21	2.95	.99	9.54	.48	.00	6.68	.39	.00	6.94
26	.09	.06	1.08	2.51	.05	4.84	4.77	.00	0.37	1.12	.00	7.12	5.75	.91	3.40	.11	.98	.45	.66	.99	1.80	7.50	.99	6.31	.11	.98	.28	6.48	.88	5.93
27	2.83	.01	9.09	.66	.99	1.38	9.98	.08	7.25	8.52	.66	6.59	.30	.98	3.79	3.40	.99	9.48	5.55	.56	8.09	9.98	.08	7.62	.91	.98	7.44	8.39	.74	7.97
28	0.00	.40	0.53	.11	.98	.32	0.10	.00	3.26	5.13	.19	4.69	4.77	.00	9.88	1.84	.90	4.78	3.78	.37	9.99	4.43	.97	1.57	9.96	.03	4.01	.09	.06	0.77
29	.37	.33	2.17	.28	.01	2.04	.66	.99	1.84	9.02	.28	9.08	7.18	.02	7.74	.66	.99	1.19	8.59	.46	0.79	1.12	.00	7.80	6.28	.03	8.09	.93	.96	2.13
30	0.83	.48	7.67	8.00	.85	8.57	2.08	.05	5.01	6.92	.01	2.06	4.09	.58	1.44	9.19	.50	4.09	0.99	.01	5.50	1.65	.97	9.45	.40	.97	9.86	6.47	.01	7.99
31	1.41	.95	7.20	1.41	.95	7.63	.11	.98	.23	2.31	.94	2.85	9.78	.56	1.82	6.02	.44	0.22	7.38	.01	7.68	6.48	.51	9.21	1.26	.42	0.21	6.05	.70	0.17

32	.75	.99	5.30	5.33	.98	4.72	1.41	.95	6.68	5.93	.27	3.78	3.86	.68	5.15	2.44	.68	1.78	5.33	.98	4.25	6.35	.02	1.51	7.80	.19	7.37	3.08	.92	4.36
33	7.38	.01	8.19	.48	.00	7.12	8.04	.60	7.15	.09	.06	.45	0.00	.40	1.35	.30	.99	7.01	2.55	.18	4.99	5.83	.17	2.50	2.05	.36	9.31	0.00	.40	0.42
34	2.04	.99	9.28	3.78	.37	9.54	5.17	.18	5.04	.75	.99	6.22	3.21	.07	5.87	.09	.06	0.89	.09	.06	1.07	7.19	.17	4.37	9.39	.42	3.01	4.53	.03	6.93
35	5.68	.93	0.89	7.95	.99	4.66	7.95	.99	3.97	7.38	.01	8.45	6.92	.01	1.47	1.31	.02	4.51	4.54	.02	0.81	.06	.13	0.03	5.11	.45	1.01	2.94	.00	8.34
36	1.78	.71	3.18	6.48	.51	9.43	.89	.44	4.39	.96	.65	8.26	6.02	.44	0.52	.10	.47	5.72	.83	.53	2.51	.48	.00	6.74	4.07	.58	5.07	6.58	.98	6.98
37	5.68	.99	9.41	8.64	.97	5.96	.48	.00	7.05	0.00	.40	0.05	1.72	.91	2.21	5.74	.04	8.31	7.08	.49	9.97	6.41	.57	2.90	9.54	.59	2.14	0.62	.32	5.84
38	8.41	.99	2.87	0.00	.40	1.00	2.88	.36	9.99	3.28	.07	1.37	6.37	.15	3.67	1.01	.01	5.55	7.98	.73	8.68	.11	.98	.49	3.13	.71	1.98	7.38	.01	1.63
39	3.94	.91	5.99	4.43	.10	2.68	5.67	.99	5.92	4.65	.01	5.85	1.45	.39	6.48	4.24	.68	2.49	5.20	.55	1.26	0.05	.00	6.48	0.48	.02	4.59	.03	.08	7.08
40	.91	.03	0.25	.44	.51	3.61	2.71	.40	0.36	.91	.03	1.10	8.77	.33	7.57	4.12	.84	9.19	.84	.99	8.67	3.95	.42	9.88	1.41	.95	7.09	8.41	.99	2.66
41	8.72	.57	1.97	5.79	.66	2.47	0.66	.30	9.74	5.15	.67	1.92	.09	.06	1.16	.32	.77	6.11	0.52	.88	0.18	.30	.99	6.62	.31	.73	5.08	.12	.11	9.76
42	6.86	.63	0.31	1.99	.36	8.95	0.00	.40	0.36	9.66	.95	5.51	.50	.98	3.72	7.44	.80	0.74	2.40	.06	6.36	1.30	.03	6.20	6.10	.05	4.61	7.77	.94	4.80
43	.72	.70	7.73	.54	.59	6.81	4.65	.00	1.60	.96	.24	1.92	.46	.91	0.08	8.84	.18	7.26	4.13	.71	6.36	.66	.99	1.57	.09	.06	0.38	0.02	.40	0.41
44	5.74	.03	1.88	2.83	.01	9.06	7.38	.01	8.10	6.59	.00	8.22	7.35	.27	6.69	2.58	.11	4.56	9.97	.52	1.95	1.91	.45	9.78	8.17	.03	5.90	5.59	.99	4.36
45	1.28	.02	4.13	8.17	.03	5.78	6.90	.57	9.81	7.39	.88	4.07	.79	.78	4.03	0.00	.40	9.70	1.30	.03	6.41	7.38	.01	7.74	7.32	.13	3.24	1.44	.20	7.88
46	4.09	.83	6.12	0.13	.14	7.96	6.22	.61	0.99	4.12	.02	7.67	0.95	.76	3.76	2.55	.98	9.40	6.76	.33	4.78	7.83	.44	6.88	2.51	.05	4.60	4.60	.26	3.43
47	0.86	.02	4.19	.10	.09	6.52	.55	.00	1.27	.15	.46	3.99	1.72	.97	7.11	6.47	.88	3.82	5.35	.23	9.42	0.28	.62	7.61	.87	.02	2.00	.76	.99	7.39
48	3.40	.99	9.25	7.95	.94	4.18	0.25	.68	7.84	2.54	.72	4.06	8.80	.68	2.65	0.51	.03	0.94	2.45	.31	5.93	9.36	.04	3.38	0.31	.78	9.53	8.31	.24	7.44
49	8.59	.20	6.88	6.90	.33	4.74	4.46	.53	1.29	1.65	.41	8.40	5.15	.94	9.95	1.25	.79	2.68	9.00	.40	9.06	.62	.39	0.08	.84	.99	9.86	7.21	.03	2.90
50	2.70	.96	7.96	4.12	.96	1.06	.84	.99	8.94	8.12	.02	5.03	9.06	.27	8.32	6.91	.95	7.58	.37	.01	1.77	8.47	.29	9.61	.69	.96	6.87	3.23	.33	9.27

51	9.27	.43	0.31	5.15	.06	9.62	4.40	.10	9.11	9.98	.15	6.53	5.21	.05	5.74	2.12	.54	1.90	8.34	.37	9.11	5.77	.78	3.13	6.86	.06	3.77	1.01	.01	5.38
52	5.86	.96	6.51	4.84	.03	6.63	1.68	.97	4.12	8.82	.30	7.34	.30	.99	6.79	5.23	.56	2.75	9.29	.24	8.26	4.70	.01	1.88	6.18	.72	3.77	.50	.98	3.87
53	6.36	.65	1.36	9.22	.01	5.68	9.40	.55	1.47	7.95	.55	1.06	2.99	.03	4.07	4.43	.97	2.01	2.62	.98	9.08	4.86	.91	6.30	8.82	.44	9.02	8.89	.55	1.90
54	4.07	.96	4.76	9.06	.27	7.46	2.62	.98	8.38	2.19	.97	8.97	8.75	.07	4.91	3.74	.01	0.45	6.11	.44	8.32	4.31	.98	0.58	6.57	.25	4.91	.91	.03	0.66
55	5.37	.02	7.80	9.05	.02	2.85	.99	.58	9.08	.75	.96	6.84	3.30	.70	3.61	.35	.05	6.13	.50	.79	8.21	.64	.01	0.01	6.23	.97	6.90	.98	.97	0.07

List of References

1. Selekman, J. A. *et al.* High-Throughput Automation in Chemical Process Development. *Annu. Rev. Chem. Biomol. Eng.* **8**, 525–547 (2017).
2. Sanderson, K. Automation: Chemistry shoots for the Moon. *Nature* **568**, 577–579 (2019).
3. Hartman, R. L., McMullen, J. P. & Jensen, K. F. Deciding whether to go with the flow: Evaluating the merits of flow reactors for synthesis. *Angew. Chemie - Int. Ed.* **50**, 7502–7519 (2011).
4. Chhabra Raj, S. V. 3.3.2.2 Resistance to Flow in Pipes. *Coulson and Richardson's Chemical Engineering, Volume 1A - Fluid Flow - Fundamentals and Applications (7th Edition)* (2018).
5. Elvira, K. S., i Solvas, X. C., Wootton, R. C. R. & DeMello, A. J. The past, present and potential for microfluidic reactor technology in chemical synthesis. *Nat. Chem.* **5**, 905–915 (2013).
6. Plutschack, M. B., Pieber, B., Gilmore, K. & Seeberger, P. H. The Hitchhiker's Guide to Flow Chemistry. *Chem. Rev.* **117**, 11796–11893 (2017).
7. Aroh, K. C. & Jensen, K. F. Efficient kinetic experiments in continuous flow microreactors. *React. Chem. Eng.* **3**, 94–101 (2018).
8. Schaber, S. D., Born, S. C., Jensen, K. F. & Barton, P. I. Design, Execution, and Analysis of Time-Varying Experiments for Model Discrimination and Parameter Estimation in Microreactors. *Org. Process Res. Dev.* **18**, 1461–1467 (2014).
9. Mathpati, A. C., Kalghatgi, S. G., Mathpati, C. S. & Bhanage, B. M. Immobilized lipase catalyzed synthesis of n-amyl acetate: parameter optimization, heterogeneous kinetics, continuous flow operation and reactor modeling. *J. Chem. Technol. Biotechnol.* **93**, 2906–2916 (2018).
10. Pedersen, M. J. *et al.* Optimization of Grignard Addition to Esters: Kinetic and Mechanistic Study of Model Phthalide Using Flow Chemistry. *Ind. Eng. Chem. Res.* **57**, 4859–4866 (2018).
11. Hone, C. A., Holmes, N., Akien, G. R., Bourne, R. A. & Muller, F. L. Rapid multistep kinetic model generation from transient flow data. *React. Chem. Eng.* **2**, 103–108 (2017).
12. Heublein, N., Moore, J. S., Smith, C. D. & Jensen, K. F. Investigation of Petasis and Ugi reactions in series in an automated microreactor system. *RSC Adv.* **4**, 63627–63631 (2014).
13. Reizman, B. J. & Jensen, K. F. An Automated Continuous-Flow Platform for the Estimation of Multistep Reaction Kinetics. *Org. Process Res. Dev.* **16**, 1770–1782 (2012).
14. McMullen, J. P. & Jensen, K. F. Rapid Determination of Reaction Kinetics with an Automated Microfluidic System. *Org. Process Res. Dev.* **15**, 398–407 (2011).

15. Perera, D. *et al.* A platform for automated nanomole-scale reaction screening and micromole-scale synthesis in flow. *Science (80-.).* **359**, 429–434 (2018).

16. Hartman, R. L. Flow chemistry remains an opportunity for chemists and chemical engineers. *Curr. Opin. Chem. Eng.* **29**, 42–50 (2020).

17. Steimbach, R. R., Kollmus, P. & Santagostino, M. A Validated “Pool and Split” Approach to Screening and Optimization of Copper-Catalyzed C–N Cross-Coupling Reactions. *J. Org. Chem.* (2020) doi:10.1021/acs.joc.0c02392.

18. Shields, B. J. *et al.* Bayesian reaction optimization as a tool for chemical synthesis. *Nature* **590**, 89–96 (2021).

19. Mateos, C., Nieves-Remacha, M. J. & Rincón, J. A. Automated platforms for reaction self-optimization in flow. *React. Chem. Eng.* **4**, 1536–1544 (2019).

20. Plutschack, M. B., Pieber, B., Gilmore, K. & Seeberger, P. H. The Hitchhiker’s Guide to Flow Chemistry. *Chem. Rev.* **117**, 11796–11893 (2017).

21. Krishnadasan, S., Brown, R. J. C., DeMello, A. J. & DeMello, J. C. Intelligent routes to the controlled synthesis of nanoparticles. *Lab Chip* **7**, 1434–1441 (2007).

22. Mcmullen, J. P. & Jensen, K. F. An intelligent microreactor system for real-time optimization of a chemical reaction. in *Twelfth International Conference on Miniaturized Systems for Chemistry and Life Sciences* 1907–1909 (Proc. Int. Conf. Miniat. Syst. Chem. Life Sci {(μTAS),} 12th, 2008).

23. Krishnadasan, S., Yashina, A., DeMello, A. J. & DeMello, J. C. Microfluidic Reactors for Nanomaterial Synthesis. in *Advances in Chemical Engineering* (ed. Schouten, J. C.) 195–231 (Academic Press, 2010). doi:10.1016/S0065-2377(10)38004-5.

24. McMullen, J. P., Stone, M. T., Buchwald, S. L. & Jensen, K. F. An Integrated Microreactor System for Self-Optimization of a Heck Reaction: From Micro- to Mesoscale Flow Systems. *Angew. Chemie Int Ed* **49**, 7076–7080 (2010).

25. McMullen, J. P. & Jensen, K. F. An Automated Microfluidic System for Online Optimization in Chemical Synthesis. *Org. Process Res. Dev.* **14**, 1169–1176 (2010).

26. Parrott, A. J., Bourne, R. A., Akien, G. R., Irvine, D. J. & Poliakoff, M. Self-optimizing continuous reactions in supercritical carbon dioxide. *Angew. Chemie - Int. Ed.* **50**, 3788–3792 (2011).

27. Bourne, R. A., Skilton, R. A., Parrott, A. J., Irvine, D. J. & Poliakoff, M. Adaptive process optimization for continuous methylation of alcohols in supercritical carbon dioxide. *Org. Process Res. Dev.* **15**, 932–938 (2011).

28. Jumbam, D. N., Skilton, R. A., Parrott, A. J., Bourne, R. A. & Poliakoff, M. The Effect of Self-Optimisation Targets on the Methylation of

Alcohols Using Dimethyl Carbonate in Supercritical {CO2}. *J Flow Chem* **2**, 24–27 (2012).

29. Skilton, R. A., Parrott, A. J., George, M. W., Poliakoff, M. & Bourne, R. A. Real-Time Feedback Control Using Online Attenuated Total Reflection Fourier Transform Infrared (ATR FT-IR) Spectroscopy for Continuous Flow Optimization and Process Knowledge. *Appl. Spectrosc.* **67**, 1127–1131 (2013).

30. Sans, V., Porwol, L., Dragone, V. & Cronin, L. A self optimizing synthetic organic reactor system using real-time in-line NMR spectroscopy. *Chem. Sci.* **6**, 1258–1264 (2015).

31. Houben, C., Peremezhney, N., Zubov, A., Kosek, J. & Lapkin, A. A. Closed-Loop Multitarget Optimization for Discovery of New Emulsion Polymerization Recipes. *Org. Process Res. Dev.* **19**, 1049–1053 (2015).

32. Peremezhney, N., Hines, E., Lapkin, A. & Connaughton, C. Combining Gaussian processes, mutual information and a genetic algorithm for multi-target optimization of expensive-to-evaluate functions. *Eng. Optim.* **46**, 1593–1607 (2014).

33. Reizman, B. J. & Jensen, K. F. Simultaneous solvent screening and reaction optimization in microliter slugs. *Chem. Commun.* **51**, 13290–13293 (2015).

34. Amara, Z. *et al.* Automated Serendipity with Self-Optimizing Continuous-Flow Reactors. *European J. Org. Chem.* **2015**, 6141–6145 (2015).

35. Holmes, N. *et al.* Self-optimisation of the final stage in the synthesis of EGFR kinase inhibitor AZD9291 using an automated flow reactor. *React. Chem. Eng.* **1**, 366–371 (2016).

36. Fitzpatrick, D. E., Battilocchio, C. & Ley, S. V. A Novel Internet-Based Reaction Monitoring, Control and Autonomous Self-Optimization Platform for Chemical Synthesis. *Org. Process Res. Dev.* **20**, 386–394 (2016).

37. Kazmierczak Jr., R. F. Optimizing Complex Bioeconomic Simulations Using an Efficient Search Heuristic. *DAE Research Report No. 704C61* (1997) doi:10.2139/ssrn.15071.

38. Cortés-Borda, D. *et al.* Optimizing the Heck-Matsuda Reaction in Flow with a Constraint-Adapted Direct Search Algorithm. *Org. Process Res. Dev.* **20**, 1979–1987 (2016).

39. Reizman, B. J., Wang, Y.-M. M., Buchwald, S. L. & Jensen, K. F. Suzuki-Miyaura cross-coupling optimization enabled by automated feedback. *React Chem Eng* **1**, 658–666 (2016).

40. Holmes, N. *et al.* Online quantitative mass spectrometry for the rapid adaptive optimisation of automated flow reactors. *React. Chem. Eng.* **1**, 96–100 (2016).

41. Walker, B. E., Bannock, J. H., Nightingale, A. M. & DeMello, J. C. Tuning reaction products by constrained optimisation. *React. Chem. Eng.* **2**,

785–798 (2017).

42. Echtermeyer, A., Amar, Y., Zakrzewski, J. & Lapkin, A. A. Self-optimisation and model-based design of experiments for developing a C–H activation flow process. *Beilstein J. Org. Chem.* **13**, 150–163 (2017).

43. Zhou, Z., Li, X. & Zare, R. N. Optimizing Chemical Reactions with Deep Reinforcement Learning. *ACS Cent. Sci.* **3**, 1337–1344 (2017).

44. Jeraal, M. I., Holmes, N., Akien, G. R. & Bourne, R. A. Enhanced process development using automated continuous reactors by self-optimisation algorithms and statistical empirical modelling. *Tetrahedron* **74**, 3158–3164 (2018).

45. Schweidtmann, A. M. *et al.* Machine learning meets continuous flow chemistry: Automated optimization towards the Pareto front of multiple objectives. *Chem. Eng. J.* **352**, 277–282 (2018).

46. Cortés-Borda, D. *et al.* An Autonomous Self-Optimizing Flow Reactor for the Synthesis of Natural Product Carpanone. *J. Org. Chem.* **83**, 14286–14299 (2018).

47. Cherkasov, N., Bai, Y., Expósito, A. J. & Rebrov, E. V. OpenFlowChem – a platform for quick, robust and flexible automation and self-optimisation of flow chemistry. *React. Chem. Eng.* **3**, 769–780 (2018).

48. Poscharny, K. *et al.* Machine assisted reaction optimization: A self-optimizing reactor system for continuous-flow photochemical reactions. *Tetrahedron* **74**, 3171–3175 (2018).

49. Bédard, A.-C. *et al.* Reconfigurable system for automated optimization of diverse chemical reactions. *Science (80-.).* **361**, 1220–1225 (2018).

50. Fitzpatrick, D. E., Maujean, T., Evans, A. C. & Ley, S. V. Across-the-World Automated Optimization and Continuous-Flow Synthesis of Pharmaceutical Agents Operating Through a Cloud-Based Server. *Angew. Chemie - Int. Ed.* **57**, 15128–15132 (2018).

51. Baumgartner, L. M., Coley, C. W., Reizman, B. J., Gao, K. W. & Jensen, K. F. Optimum catalyst selection over continuous and discrete process variables with a single droplet microfluidic reaction platform. *React. Chem. Eng.* **3**, 301–311 (2018).

52. Hsieh, H.-W., Coley, C. W., Baumgartner, L. M., Jensen, K. F. & Robinson, R. I. 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**, 542–550 (2018).

53. Bezinge, L., Maceiczyk, R. M., Lignos, I., Kovalenko, M. V & deMello, A. J. Pick a Color MARIA: Adaptive Sampling Enables the Rapid Identification of Complex Perovskite Nanocrystal Compositions with Defined Emission Characteristics. *ACS Appl. Mater. Interfaces* **10**, 18869–18878 (2018).

54. Rubens, M., Vrijsen, J. H., Laun, J. & Junkers, T. Precise Polymer Synthesis by Autonomous Self-Optimizing Flow Reactors. *Angew.*

Chemie Int. Ed. **58**, 3183–3187 (2019).

55. Aka, E. C. *et al.* Reconfigurable Flow Platform for Automated Reagent Screening and Autonomous Optimization for Bioinspired Lignans Synthesis. *J. Org. Chem.* **84**, 14101–14112 (2019).

56. Baumgartner, L. M., Dennis, J. M., White, N. A., Buchwald, S. L. & Jensen, K. F. Use of a Droplet Platform To Optimize Pd-Catalyzed C–N Coupling Reactions Promoted by Organic Bases. *Org. Process Res. Dev.* **23**, 1594–1601 (2019).

57. Wimmer, E. *et al.* An autonomous self-optimizing flow machine for the synthesis of pyridine–oxazoline (PyOX) ligands. *React. Chem. Eng.* **4**, 1608–1615 (2019).

58. Manson, J. A. *et al.* A hybridised optimisation of an automated photochemical continuous flow reactor. *Chimia (Aarau).* **73**, 817–822 (2019).

59. Ahrberg, C. D., Wook Choi, J. & Geun Chung, B. Automated droplet reactor for the synthesis of iron oxide/gold core-shell nanoparticles. *Sci. Rep.* **10**, 1737 (2020).

60. Salley, D. *et al.* A nanomaterials discovery robot for the Darwinian evolution of shape programmable gold nanoparticles. *Nat. Commun.* **11**, 2771 (2020).

61. Porwol, L. *et al.* An Autonomous Chemical Robot Discovers the Rules of Inorganic Coordination Chemistry without Prior Knowledge. *Angew. Chemie Int. Ed.* **59**, 11256–11261 (2020).

62. Vasudevan, N. *et al.* Direct C−H Arylation of Indole-3-Acetic Acid Derivatives Enabled by an Autonomous Self-Optimizing Flow Reactor. *Adv. Synth. Catal.* adsc.202001217 (2020) doi:10.1002/adsc.202001217.

63. Clayton, A. D. *et al.* Self-optimising reactive extractions: towards the efficient development of multi-step continuous flow processes. *J. Flow Chem.* **10**, 199–206 (2020).

64. Clayton, A. D. *et al.* Automated self-optimisation of multi-step reaction and separation processes using machine learning. *Chem. Eng. J.* **384**, 123340 (2020).

65. Fath, V., Kockmann, N., Otto, J. & Röder, T. Self-optimising processes and real-time-optimisation of organic syntheses in a microreactor system using Nelder–Mead and design of experiments. *React. Chem. Eng.* **5**, 1281–1299 (2020).

66. Li, J. *et al.* Autonomous discovery of optically active chiral inorganic perovskite nanocrystals through an intelligent cloud lab. *Nat. Commun.* **11**, 2046 (2020).

67. Wang, L., Karadaghi, L. R., Brutchey, R. L. & Malmstadt, N. Self-optimizing parallel millifluidic reactor for scaling nanoparticle synthesis. *Chem. Commun.* **56**, 3745–3748 (2020).

68. Epps, R. W. *et al.* Artificial Chemist: An Autonomous Quantum Dot Synthesis Bot. *Adv. Mater.* **32**, 2001626 (2020).

69. Abdel-Latif, K. *et al.* Self-Driven Multistep Quantum Dot Synthesis Enabled by Autonomous Robotic Experimentation in Flow. *Adv. Intell. Syst.* **n/a**, 2000245 (2020).

70. Jeraal, M. I., Sung, S. & Lapkin, A. A. A Machine Learning-Enabled Autonomous Flow Chemistry Platform for Process Optimization of Multiple Reaction Metrics. *Chemistry–Methods* **1**, 71–77 (2021).

71. Zhang, C., Amar, Y., Cao, L. & Lapkin, A. A. Solvent Selection for Mitsunobu Reaction Driven by an Active Learning Surrogate Model. *Org. Process Res. Dev.* **24**, 2864–2873 (2020).

72. MacLeod, B. P. *et al.* Self-driving laboratory for accelerated discovery of thin-film materials. *Sci. Adv.* **6**, (2020).

73. Häse, F., Roch, L. M., Kreisbeck, C. & Aspuru-Guzik, A. Phoenics: A Bayesian Optimizer for Chemistry. *ACS Cent. Sci.* **4**, 1134–1145 (2018).

74. Reker, D., Hoyt, E. A., Bernardes, G. J. L. & Rodrigues, T. Adaptive Optimization of Chemical Reactions with Minimal Experimental Information. *Cell Reports Phys. Sci.* **1**, 100247 (2020).

75. Cao, L. *et al.* Optimization of Formulations Using Robotic Experiments Driven by Machine Learning DoE. *Cell Reports Phys. Sci.* **2**, 100295 (2021).

76. Christensen, M. *et al.* Data-Science Driven Autonomous Process Optimization. (2021) doi:10.21203/rs.3.rs-116245/v1.

77. Häse, F., Roch, L. M. & Aspuru-Guzik, A. Chimera: enabling hierarchy based multi-objective optimization for self-driving laboratories. *Chem. Sci.* **9**, 7642–7655 (2018).

78. Hall, B. L. *et al.* Autonomous optimisation of a nanoparticle catalysed reduction reaction in continuous flow. *Chem. Commun.* **57**, 4926–4929 (2021).

79. Weissman, S. A. & Anderson, N. G. Design of Experiments (DoE) and Process Optimization. A Review of Recent Publications. *Org. Process Res. Dev.* **19**, 1605–1633 (2015).

80. Weissman, S. A. & Anderson, N. G. Design of Experiments (DoE) and Process Optimization. A Review of Recent Publications. *Org. Process Res. Dev.* **19**, 1605–1633 (2014).

81. Yu, L. X. *et al.* Understanding Pharmaceutical Quality by Design. *AAPS J.* **16**, 771–783 (2014).

82. Zitrom, V. C. & Czitrom, V. One-Factor-at-a-Time versus Designed Experiments. *Am. Stat.* **53**, 126–131 (1999).

83. Rodrigues, M. . & Iemma, A. . Experimental Strategy for Fractional Factorial and Central Composite Rotational Designs. in *Experimental Design and Process Optimization* (CRC Press, 2015).

84. Alam, M., Wise, C., Baxter, C. A., Cleator, E. & Walkinshaw, A. Development of a Robust Procedure for the Copper-catalyzed Ring-Opening of Epoxides with Grignard Reagents. *Org. Process Res. Dev.* **16**, 435–441 (2012).

85. Walter, E. & Pronzato, L. Qualitative and quantitative experiment design for phenomenological models—A survey. *Automatica* **26**, 195–213 (1990).

86. Goos, P. & Jones, B. *Optimal Design of Experiments*. (John Wiley & Sons, Ltd, 2011). doi:10.1002/9781119974017.

87. Spendley, W., Hext, G. R. & Himsworth, F. R. Sequential Application of Simplex Designs in Optimisation and Evolutionary Operation. *Technometrics* **4**, 441–461 (1962).

88. Nelder, J. A. & Mead, R. A Simplex Method for Function Minimization. *Comput. J.* **7**, 308–313 (1965).

89. Routh, M. W., Swartz, P. A. & Denton, M. B. Performance of the Super Modified Simplex. *Anal Chem* **49**, 1422–1428 (1977).

90. Moore, J. S. & Jensen, K. F. Automated multitrajectory method for reaction optimization in a microfluidic system using online IR analysis. *Org. Process Res. Dev.* **16**, 1409–1415 (2012).

91. Huang, X., Corbett, J., Safranek, J. & Wu, J. An algorithm for online optimization of accelerators. *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.* **726**, 77–83 (2013).

92. Cauchy, M. A. Méthode générale pour la résolution des systèmes d'équations simultanées. *C. R. Hebd. Seances Acad. Sci.* **25**, 536–538 (1847).

93. Hestenes, M. R. & Stiefel, E. Methods of conjugate gradients for solving linear systems. *J. Res. Natl. Bur. Stand. (1934).* **49**, 409 (1952).

94. Armijo, L. Minimization of functions having Lipschitz continuous first partial derivatives. *Pacific J. Math.* **16**, 1–3 (1966).

95. Xu, P. A hybrid global optimization method: The multi-dimensional case. *J. Comput. Appl. Math.* **155**, 423–446 (2003).

96. Horst, R. & Tuy, H. Some Important Classes of Global Optimization Problems. in *Global Optimization: Deterministic Approaches* 3–50 (Springer Berlin Heidelberg, 1990). doi:10.1007/978-3-662-02598-7_1.

97. Holland, J. H. *Adaptation in natural and artificial systems : an introductory analysis with applications to biology, control, and artificial intelligence*. *Ann Arbor University of Michigan Press 1975* (1975).

98. Katoch, S., Chauhan, S. S. & Kumar, V. A review on genetic algorithm: past, present, and future. *Multimed. Tools Appl.* **80**, 8091–8126 (2021).

99. Ong, Y. S., Nair, P. B. & Keane, A. J. Evolutionary Optimization of Computationally Expensive Problems via Surrogate Modeling. *AIAA J.* **41**, 687–696 (2003).

100. Huyer, W. & Neumaier, A. SNOBFIT -- Stable Noisy Optimization by Branch and Fit. *ACM Trans. Math. Softw.* **35**, 1–25 (2008).

101. Rios, L. M. & Sahinidis, N. V. Derivative-free optimization: a review of algorithms and comparison of software implementations. *J. Glob. Optim.* **56**, 1247–1293 (2013).

102. Zielinski, K. & Laur, R. Stopping Criteria for Differential Evolution in

Constrained Single-Objective Optimization BT - Advances in Differential Evolution. in (ed. Chakraborty, U. K.) 111–138 (Springer Berlin Heidelberg, 2008). doi:10.1007/978-3-540-68830-3_4.

103. Rasmussen, C. E. & Williams, C. K. I. Regression. in *Gaussian Processes for Machine Learning* 7–32 (MIT Press, 2006).

104. Schulz, E., Speekenbrink, M. & Krause, A. A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *J. Math. Psychol.* **85**, 1–16 (2018).

105. Kushner, H. J. A New Method of Locating the Maximum Point of an Arbitrary Multipeak Curve in the Presence of Noise. *J. Basic Eng.* **86**, 97–106 (1964).

106. Močkus, J. On bayesian methods for seeking the extremum. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 400–404 (1975). doi:10.1007/3-540-07165-2_55.

107. Jones, D. R., Schonlau, M. & Welsh, W. J. Efficient global optimization of expensive black-box functions. *J. Glob. Optim.* **13**, 455–492 (1998).

108. Locatelli, M. Bayesian Algorithms for One-Dimensional Global Optimization. *J. Glob. Optim.* **10**, 57–76 (1997).

109. Emmerich, M. T. M. & Deutz, A. H. A tutorial on multiobjective optimization: fundamentals and evolutionary methods. *Nat. Comput.* **17**, 585–609 (2018).

110. Miettinen, K. *Nonlinear Multiobjective Optimization*. vol. 12 (Springer US, 1998).

111. Riquelme, N., Von Lucken, C. & Baran, B. Performance metrics in multi-objective optimization. in *2015 Latin American Computing Conference (CLEI)* 1–11 (IEEE, 2015). doi:10.1109/CLEI.2015.7360024.

112. Zitzler, E. & Thiele, L. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. *IEEE Trans. Evol. Comput.* **3**, 257–271 (1999).

113. Coello Coello, C. A. & Reyes Sierra, M. A Study of the Parallelization of a Coevolutionary Multi-objective Evolutionary Algorithm. in (eds. Monroy, R., Arroyo-Figueroa, G., Sucar, L. E. & Sossa, H.) 688–697 (Springer Berlin Heidelberg, 2004). doi:10.1007/978-3-540-24694-7_71.

114. Ishibuchi, H., Masuda, H., Tanigaki, Y. & Nojima, Y. Modified Distance Calculation in Generational Distance and Inverted Generational Distance. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 110–125 (2015). doi:10.1007/978-3-319-15892-1_8.

115. Fonseca, C. M. & Fleming, P. J. On the performance assessment and comparison of stochastic multiobjective optimizers. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (1996). doi:10.1007/3-540-61723-X_1022.

116. Pelamatti, J., Brevault, L., Balesdent, M., Talbi, E.-G. & Guerin, Y. Efficient global optimization of constrained mixed variable problems. *J. Glob. Optim.* **73**, 583–613 (2019).

117. Hildebrandt, T. & Branke, J. On Using Surrogates with Genetic Programming. *Evol. Comput.* **23**, 343–367 (2015).

118. Nguyen, S., Zhang, M., Johnston, M. & Tan, K. C. Selection Schemes in Surrogate-Assisted Genetic Programming for Job Shop Scheduling BT - Simulated Evolution and Learning. in (eds. Dick, G. et al.) 656–667 (Springer International Publishing, 2014).

119. Rakshit, S. & Ananthasuresh, G. K. Simultaneous material selection and geometry design of statically determinate trusses using continuous optimization. *Struct. Multidiscip. Optim.* **35**, 55–68 (2008).

120. Pelamatti, J., Brevault, L., Balesdent, M., Talbi, E.-G. & Guerin, Y. How to Deal with Mixed-Variable Optimization Problems: An Overview of Algorithms and Formulations. in *Advances in Structural and Multidisciplinary Optimization* (eds. Schumacher, A., Vietor, T., Fiebig, S., Bletzinger, K.-U. & Maute, K.) 64–82 (Springer International Publishing, 2018).

121. Deb, K. Multi-objective Optimisation Using Evolutionary Algorithms: An Introduction. in *Multi-objective Evolutionary Optimisation for Product Design and Manufacturing* (eds. Wang, L., Ng, A. H. C. & Deb, K.) 3–34 (Springer London, 2011). doi:10.1007/978-0-85729-652-8_1.

122. Brownlee, A. E. I. & Wright, J. A. Constrained, mixed-integer and multi-objective optimisation of building designs by NSGA-II with fitness approximation. *Appl. Soft Comput.* **33**, 114–126 (2015).

123. Zhou, T., Zhou, Y. & Sundmacher, K. A hybrid stochastic–deterministic optimization approach for integrated solvent and process design. *Chem. Eng. Sci.* **159**, 207–216 (2017).

124. Tamilselvi, S. & Baskar, S. Modified parameter optimization of distribution transformer design using covariance matrix adaptation evolution strategy. *Int. J. Electr. Power Energy Syst.* **61**, 208–218 (2014).

125. Ahmadi, M., Arabi, M., Hoag, D. L. & Engel, B. A. A mixed discrete-continuous variable multiobjective genetic algorithm for targeted implementation of nonpoint source pollution control practices. *Water Resour. Res.* **49**, 8344–8356 (2013).

126. Fazlollahi, S., Mandel, P., Becker, G. & Maréchal, F. Methods for multi-objective investment and operating optimization of complex energy systems. *Energy* **45**, 12–22 (2012).

127. Lin, M., Chin, K. S., Ma, L. & Tsui, K. L. A comprehensive multi-objective mixed integer nonlinear programming model for an integrated elderly care service districting problem. *Ann. Oper. Res.* 1–31 (2018) doi:10.1007/s10479-018-3078-6.

128. Dullinger, C., Struckl, W. & Kozek, M. Simulation-based multi-objective system optimization of train traction systems. *Simul. Model. Pract. Theory* **72**, 104–117 (2017).

129. Zhan, D. & Xing, H. Expected improvement for expensive optimization: a review. *J. Glob. Optim.* **78**, 507–544 (2020).

130. Benítez-Hidalgo, A., Nebro, A. J., García-Nieto, J., Oregi, I. & Del Ser, J. jMetalPy: A Python framework for multi-objective optimization with metaheuristics. *Swarm Evol. Comput.* **51**, 100598 (2019).

131. Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A. & Brown, S. D. An introduction to decision tree modeling. *Journal of Chemometrics* vol. 18 275–285 (2004).

132. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).

133. Bliek, L., Guijt, A., Verwer, S. & de Weerdt, M. Black-box Mixed-Variable Optimisation using a Surrogate Model that Satisfies Integer Constraints. (2020) doi:10.1007/s10472-020-09712-4.

134. Lin, Y. & Zhang, H. H. Component selection and smoothing in multivariate nonparametric regression. *Ann. Stat.* **34**, 2272–2297 (2006).

135. Storlie, C. B., Bondell, H. D., Reich, B. J. & Zhang, H. H. Surface Estimation, Variable Selection, and the Nonparametric Oracle Property. *Stat. Sin.* **21**, 679–705 (2011).

136. Swiler, L. P. *et al.* Surrogate Models for Mixed Discrete-Continuous Variables. in *Constraint Programming and Decision Making* (eds. Ceberio, M. & Kreinovich, V.) 181–202 (Springer International Publishing, 2014). doi:10.1007/978-3-319-04280-0_21.

137. Bartz-Beielstein, T. & Zaefferer, M. Model-based methods for continuous and discrete global optimization. *Appl. Soft Comput.* **55**, 154–167 (2017).

138. Qian, P. Z. G., Wu, H. & Wu, C. F. J. Gaussian process models for computer experiments with qualitative and quantitative factors. *Technometrics* **50**, 383–396 (2008).

139. Zhou, Q., Qian, P. Z. G. & Zhou, S. A simple approach to emulation for computer models with qualitative and quantitative factors. *Technometrics* **53**, 266–273 (2011).

140. Zhang, Y. & Notz, W. I. Computer Experiments with Qualitative and Quantitative Variables: A Review and Reexamination. *Qual. Eng.* **27**, 2–13 (2015).

141. Deng, X., Lin, C. D., Liu, K. W. & Rowe, R. K. Additive Gaussian Process for Computer Models With Qualitative and Quantitative Factors. *Technometrics* **59**, 283–292 (2017).

142. Li, P. & Chen, S. A review on Gaussian Process Latent Variable Models. *CAAI Trans. Intell. Technol.* **1**, 366–376 (2016).

143. Zhang, Y., Tao, S., Chen, W. & Apley, D. W. A Latent Variable Approach to Gaussian Process Modeling with Qualitative and Quantitative Factors. *Technometrics* **62**, 291–302 (2020).

144. Xu, D. & Tian, Y. A Comprehensive Survey of Clustering Algorithms. *Ann. Data Sci.* **2**, 165–193 (2015).

145. Gower, J. C. A General Coefficient of Similarity and Some of Its Properties. *Biometrics* **27**, 857–871 (1971).

146. Halstrup, M. Black-box optimization of mixed discrete-continuous optimization problems. *Tech. Univ. Dortmund* (2016) doi:10.17877/DE290R-17800.

147. Emmerich, M. T. M., Giannakoglou, K. C. & Naujoks, B. Single- and multiobjective evolutionary optimization assisted by Gaussian random field metamodels. *IEEE Trans. Evol. Comput.* **10**, 421–439 (2006).

148. Emmerich, M. T. M., Deutz, A. H. & Klinkenberg, J. W. Hypervolume-based expected improvement: Monotonicity properties and exact computation. in *2011 IEEE Congress of Evolutionary Computation (CEC)* 2147–2154 (2011). doi:10.1109/CEC.2011.5949880.

149. Yang, K., Emmerich, M., Deutz, A. & Fonseca, C. M. Computing 3-D Expected Hypervolume Improvement and Related Integrals in Asymptotically Optimal Time. in *Evolutionary Multi-Criterion Optimization* (eds. Trautmann, H. et al.) 685–700 (Springer International Publishing, 2017). doi:10.1007/978-3-319-54157-0_46.

150. Thompson, W. R. On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples. *Biometrika* **25**, 285–294 (1933).

151. Russo, D. J., Roy, B., Kazerouni, A., Osband, I. & Wen, Z. A Tutorial on Thompson Sampling. *Found. Trends Mach. Learn.* **11**, 1–96 (2018).

152. Hernández-Lobato, J. M., Hoffman, M. W. & Ghahramani, Z. Predictive Entropy Search for Efficient Global Optimization of Black-box Functions. *Adv. Neural Inf. Process. Syst. 27* **1**, 918–926 (2014).

153. Shahriari, B., Wang, Z., Hoffman, M. W., Bouchard-Côté, A. & de Freitas, N. An Entropy Search Portfolio for Bayesian Optimization. (2014).

154. Bradford, E., Schweidtmann, A. M. & Lapkin, A. A. Efficient multiobjective optimization employing Gaussian processes, spectral sampling and a genetic algorithm. *J. Glob. Optim.* **71**, 407–438 (2018).

155. Zhan, D., Cheng, Y. & Liu, J. Expected Improvement Matrix-Based Infill Criteria for Expensive Multiobjective Optimization. *IEEE Trans. Evol. Comput.* **21**, 956–975 (2017).

156. Deb, K., Thiele, L., Laumanns, M. & Zitzler, E. Scalable Test Problems for Evolutionary Multiobjective Optimization. in *Evolutionary Multiobjective Optimization* (2005). doi:10.1007/1-84628-137-7_6.

157. Van Veldhuizen, D. A. Multiobjective Evolutionary Algorithms: Classifications, Analyses, and New Innovations. (Air Force Institute of Technology, 1999).

158. Clayton, A. D. Multi-Objective, Multiphasic and Multi-Step Self-Optimising Continuous Flow Systems. (University of Leeds, 2020).

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

160. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

161. Kraft, D. A Software Package for Sequential Quadratic Programming. *DFVLR-FB* **88**–**28**, (1988).

162. van Veldhuizen, D. A. & Lamont, G. B. Multiobjective evolutionary algorithm test suites. in *Proceedings of the 1999 ACM symposium on Applied computing - SAC '99* 351–357 (ACM Press, 1999). doi:10.1145/298151.298382.

163. Dormand, J. R. & Prince, P. J. A family of embedded Runge-Kutta formulae. *J. Comput. Appl. Math.* **6**, 19–26 (1980).

164. Burachik, R. S., Kaya, C. Y. & Rizvi, M. M. Algorithms for Generating Pareto Fronts of Multi-objective Integer and Mixed-Integer Programming Problems. *arXiv:1903.07041v1* (2019).

165. Deb, K., Pratap, A., Agarwal, S. & Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **6**, 182–197 (2002).

166. Pelamatti, J., Brevault, L., Balesdent, M., Talbi, E.-G. & Guerin, Y. Overview and Comparison of Gaussian Process-Based Surrogate Models for Mixed Continuous and Discrete Variables: Application on Aerospace Design Problems. in (eds. Bartz-Beielstein, T., Filipič, B., Korošec, P. & Talbi, E.-G.) 189–224 (Springer International Publishing, 2020). doi:10.1007/978-3-030-18764-4_9.

167. Murray, P. M., Tyler, S. N. G. & Moseley, J. D. Beyond the Numbers: Charting Chemical Reaction Space. *Org. Process Res. Dev.* **17**, 40–46 (2013).

168. Lutz, M. W. *et al.* Experimental design for high-throughput screening. *Drug Discov. Today* **1**, 277–286 (1996).

169. Shevlin, M. Practical High-Throughput Experimentation for Chemists. *ACS Med. Chem. Lett.* **8**, 601–607 (2017).

170. Amar, Y., Schweidtmann, A. M., Deutsch, P., Cao, L. & Lapkin, A. A. Machine learning and molecular descriptors enable rational solvent selection in asymmetric catalysis. *Chem. Sci.* **10**, 6697–6706 (2019).

171. Moriconi, R., Deisenroth, M. P. & Sesh Kumar, K. S. High-dimensional Bayesian optimization using low-dimensional feature spaces. *Mach. Learn.* **109**, 1925–1943 (2020).

172. Cawse, J. N., Gazzola, G. & Packard, N. Efficient discovery and optimization of complex high-throughput experiments. *Catal. Today* **159**, 55–63 (2011).

173. Wendt, M. D. & Kunzer, A. R. Ortho selectivity in SNAr substitutions of 2,4-dihaloaromatic compounds. Reactions with piperidine. *Tetrahedron Lett.* **51**, 641–644 (2010).

174. Chéron, N., El Kaïm, L., Grimaud, L. & Fleurat-Lessard, P. Evidences for the Key Role of Hydrogen Bonds in Nucleophilic Aromatic Substitution Reactions. *Chem. – A Eur. J.* **17**, 14929–14934 (2011).

175. Gopalakrishnan, J. Analysis of Organic Volatile Amines in Samples by

Headspace Gas Chromatographic Technique Based on their Physicochemical Properties. (VIT University, 2018).

176. Miles, J. R -Squared, Adjusted R -Squared. in *Encyclopedia of Statistics in Behavioral Science* (John Wiley & Sons, Ltd, 2005). doi:10.1002/0470013192.bsa526.

177. Ezekiel, M. The Application of the Theory of Error to Multiple and Curvilinear Correlation. *J. Am. Stat. Assoc.* **24**, 99 (1929).

178. Greizerstein, W. & Brieux, J. A. The ortho:para Ratio in the Activation of the Nucleophilic Aromatic Substitution by the Nitro Group. *J. Am. Chem. Soc.* **84**, 1032–1036 (1962).

179. Mennen, S. M. *et al.* The Evolution of High-Throughput Experimentation in Pharmaceutical Development and Perspectives on the Future. *Org. Process Res. Dev.* **23**, 1213–1242 (2019).

180. Yu, S. *et al.* Synthesis of a TRPV1 Receptor Antagonist. *J. Org. Chem.* **74**, 9539–9542 (2009).

181. Sheldon, R. A. Metrics of Green Chemistry and Sustainability: Past, Present, and Future. *ACS Sustain. Chem. Eng.* **6**, 32–48 (2018).

182. Sheldon, R. A. The: E factor 25 years on: The rise of green chemistry and sustainability. *Green Chem.* **19**, 18–43 (2017).

183. Sheldon, R. A. Fundamentals of green chemistry: efficiency in reaction design. *Chem. Soc. Rev.* **41**, 1437–1451 (2012).

184. Boudart, M. Turnover Rates in Heterogeneous Catalysis. *Chem. Rev.* **95**, 661–666 (1995).

185. Byrne, F. P. *et al.* Tools and techniques for solvent selection: green solvent selection guides. *Sustain. Chem. Process.* **4**, 7 (2016).

186. Constable, D. J. C., Jimenez-Gonzalez, C. & Lapkin, A. Process Metrics. *Green Chemistry Metrics* 228–247 (2008) doi:https://doi.org/10.1002/9781444305432.ch6.

187. Kralisch, D., Ott, D. & Gericke, D. Rules and benefits of Life Cycle Assessment in green chemical process and synthesis design: a tutorial review. *Green Chem.* **17**, 123–145 (2015).

188. Buitrago Santanilla, A. *et al.* Nanomole-scale high-throughput chemistry for the synthesis of complex molecules. *Science (80-.).* **347**, 49–53 (2015).

189. Gesmundo, N. J. *et al.* Nanoscale synthesis and affinity ranking. *Nature* **557**, 228–232 (2018).

190. Corma, A. *et al.* Optimisation of olefin epoxidation catalysts with the application of high-throughput and genetic algorithms assisted by artificial neural networks (softcomputing techniques). *J. Catal.* **229**, 513–524 (2005).

191. Isbrandt, E. S., Sullivan, R. J. & Newman, S. G. High Throughput Strategies for the Discovery and Optimization of Catalytic Reactions. *Angew. Chemie Int. Ed.* **58**, 7180–7191 (2019).

192. Wyvratt, B. M., McMullen, J. P. & Grosser, S. T. Multidimensional

dynamic experiments for data-rich process development of reactions in flow. *React. Chem. Eng.* (2019) doi:10.1039/c9re00078j.

193. Kuethe, J. T., Tellers, D. M., Weissman, S. A. & Yasuda, N. Development of a Sequential Tetrahydropyran and Tertiary Butyl Deprotection: High-Throughput Experimentation, Mechanistic Analysis, and DOE Optimization. *Org. Process Res. Dev.* **13**, 471–477 (2009).

194. Lucay, F. A., Sales-Cruz, M., Gálvez, E. D. & Cisternas, L. A. Modeling of the Complex Behavior through an Improved Response Surface Methodology. *Miner. Process. Extr. Metall. Rev.* **42**, 285–311 (2021).

195. Quaglio, M., Fraga, E. S., Cao, E., Gavriilidis, A. & Galvanin, F. A model-based data mining approach for determining the domain of validity of approximated models. *Chemom. Intell. Lab. Syst.* **172**, 58–67 (2018).

196. Cortés-Borda, D. *et al.* Optimizing the Heck–Matsuda Reaction in Flow with a Constraint-Adapted Direct Search Algorithm. *Org. Process Res. Dev.* **20**, 1979–1987 (2016).

197. Vanaja, K. & Shobha Rani, R. H. Design of Experiments: Concept and Applications of Plackett Burman Design. *Clin. Res. Regul. Aff.* **24**, 1–23 (2007).

198. Davis, M. E. *Numerical Methods and Modeling for Chemical Engineers*. (John Wiley & Sons, Inc., 1984).

199. Negi, D. S. *et al.* Kinetics and Process Development for Deoxofluorination of a Steroid. *Org. Process Res. Dev.* **12**, 345–348 (2008).

200. Zongzhao Zhou, Yew Soon Ong, My Hanh Nguyen & Dudy Lim. A Study on Polynomial Regression and Gaussian Process Global Surrogate Model in Hierarchical Surrogate-Assisted Evolutionary Algorithm. *2005 IEEE Congr. Evol. Comput.* **3**, 2832–2839 (2005).

201. Zongzhao Zhou, Yew Soon Ong, My Hanh Nguyen & Dudy Lim. A Study on Polynomial Regression and Gaussian Process Global Surrogate Model in Hierarchical Surrogate-Assisted Evolutionary Algorithm. in *2005 IEEE Congress on Evolutionary Computation* vol. 3 2832–2839 (IEEE, 2005).

202. Chapman, M. R. *et al.* Simple and Versatile Laboratory Scale CSTR for Multiphasic Continuous-Flow Chemistry and Long Residence Times. *Org. Process Res. Dev.* **21**, 1294–1301 (2017).

203. Laudadio, G. *et al.* Selective C(sp3)−H Aerobic Oxidation Enabled by Decatungstate Photocatalysis in Flow. *Angew. Chemie - Int. Ed.* **57**, 4078–4082 (2018).

204. Adamo, A., Heider, P. L., Weeranoppanant, N. & Jensen, K. F. Membrane-Based, Liquid–Liquid Separator with Integrated Pressure Control. *Ind. Eng. Chem. Res.* **52**, 10802–10808 (2013).

205. Jasrasaria, D. & Pyzer-Knapp, E. O. Dynamic Control of Explore/Exploit Trade-Off in Bayesian Optimization. in *SAI 2018: Intelligent Computing* (eds. Arai, K., Kapoor, S. & Bhatia, R.) 1–15 (Springer International Publishing, 2019). doi:10.1007/978-3-030-01174-1_1.