

Supplementary information

1. MATLAB code for impurity scouting system.

All code for the HPLC method optimization system GUI is available on Github.¹ The datafiles for experiments 1-9 are also available there. Code was written to provide a user-friendly process for automated HPLC method optimization using both single objective and multi-objective Bayesian optimization.

1.1. HPLC method optimization application

HPLC method optimization software has been written that incorporates Bayesian optimization algorithms to autonomously optimize for chromatogram quality. The software handled every element of the method development process: from method creation and execution to data analysis and optimization of method conditions. This makes the system ‘closed loop’ which removes the need for user input during the optimization process. A way to communicate with the HPLC hardware autonomously without any user input was achieved using code. Code was written in MATLAB and ChemStation macros to automate all parts of the HPLC method optimization process, including:

- The writing and execution of HPLC methods.
- Reading chromatogram data and performing analysis to extract the chromatogram quality.
- Machine learning algorithms, specifically Bayesian optimization, to automate the method development.
- A graphical user interface (GUI) to provide a user-friendly application for the user to define parameters for the optimization and analyse the data.

To make the autonomous HPLC method development system user friendly, a graphical user interface (GUI) using MATLAB’s built-in app designer was created, shown in Figure S1. This software built in all the parameters required for the optimization to take place, including:

- The design space, or ‘HPLC space’ for the optimization, which could be modified by selecting the “Configure input variables” button. Here, different variables could be set either to constant values, or to be optimized with upper and lower boundaries to be defined. The more variables that are selected to be optimized, the more dimensionality is added to the optimization and the longer it will take to optimize. Figure S3 shows the how the input variables are configured when the “Configure input variables” button is pressed.
- Selection of an optimization algorithm from the “Algorithm” drop down box, as well as the chosen objective functions. If the algorithm is single objective such as BOAEEI then only one objective function can be selected. If the algorithm is multi-objective such as TS-EMO then multiple objectives can be selected.
- Extra parameter data including:
 - Setting the initial number of experiments to run. The initial experiments generated are based off a LHS design, which is then scaled between the lower and upper bounds defined by each variable.
 - Setting the HPLC directory where the chromatogram data is stored to be read by the app when the HPLC methods finish.
 - Equilibration time to denote how long the app should wait after setting the defined HPLC method conditions before starting the method.
 - The type of detection method used for the chromatograms, either using the DAD detector to select a specific wavelength to monitor (given in the settings the DAD has been enabled for that wavelength range), or a specific VWD channel set in the ChemStation software that needs to be setup before.
 - The option to use built in peak detection with a community made gaussian peak picking function along with its associated parameters, or to use the built in ChemStation detection which the user must predefine in the HPLC method beforehand and configure to save the data as a CSV after the method has finished.
- A Data visualisation tab shown in Figure S2 to show the chromatograms processed when the slider is set to 1 on the left-hand side, or to plot the input and response data in up to five dimensions (x, y, z axis as well as colour and dot size) when the slider is set to 2 on the right-hand side. This allows for most of the data processing to be done by the user within the app.

- The ability to load data from a previous save to either continue an optimization or for data analysis using the “Load” button. When the apply button is pressed all data is saved in a folder called “Results” with the name defined by the user in the “Data file name” text box.
- A reprocess button that allows the user to re-process the objective function results when the built in peak picking parameters have been modified, or if changes to the objective function code are made.
- Start and pause buttons for the optimization. A counter is available labelled “Number of experiments” to define how many optimization algorithm experiments the user wants to run after the initial conditions has finished. Optimization will start automatically if the “Auto start” box is checked.

Define inputs | Data visualisation | Optimisation test

Step 1: Define starting conditions Data file name: test [Load] [Reset]

Configure input variables [Reset]

HPLC Directory: C:\Users\Public\Documents\ChemStation\1\Data\cmdnd\HPLC0170

Number of initial experiments: 10

Equilibration time: 600

Objective functions: RsCrit2, Maximisepeak, Minimise methc

Algorithm: TS-EMO

Calculate all response functions?

Built in detection Chemstation detection

Slope threshold: 0.05 Height cut off: 5

Amp threshold: 15 Area cut off: 50

Smooth width: 1

Peak group: 5

Step 2: Run initial conditions [Start] [Pause] [Progress Indicator]

Step 3: Run optimisation Auto start

Number of experiments: 30 [Start] [Pause] [Progress Indicator]

[Reprocess files] [Reprocess data]

Figure S1: The HPLC optimization GUI inputs tab.

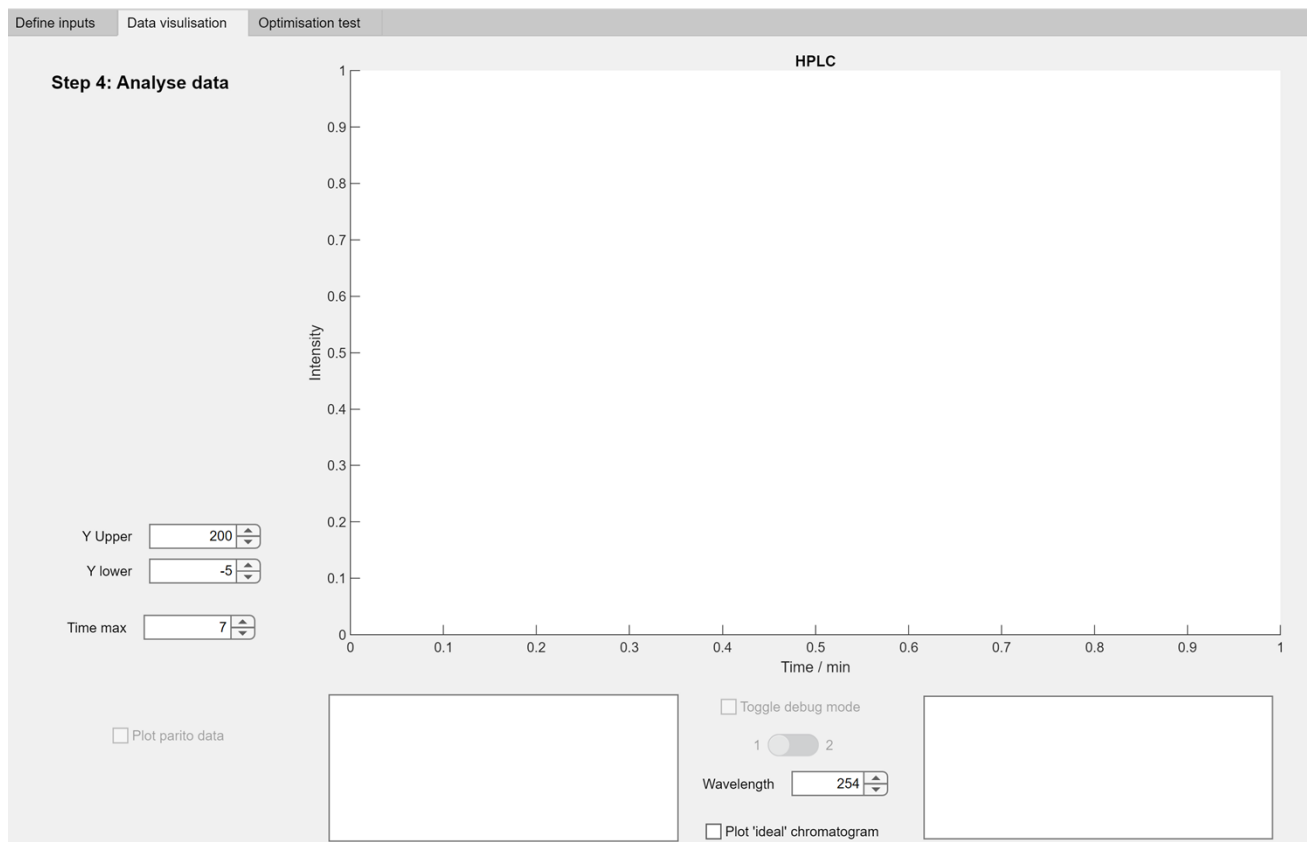


Figure S2: The HPLC optimization GUI Data visualisation tab.

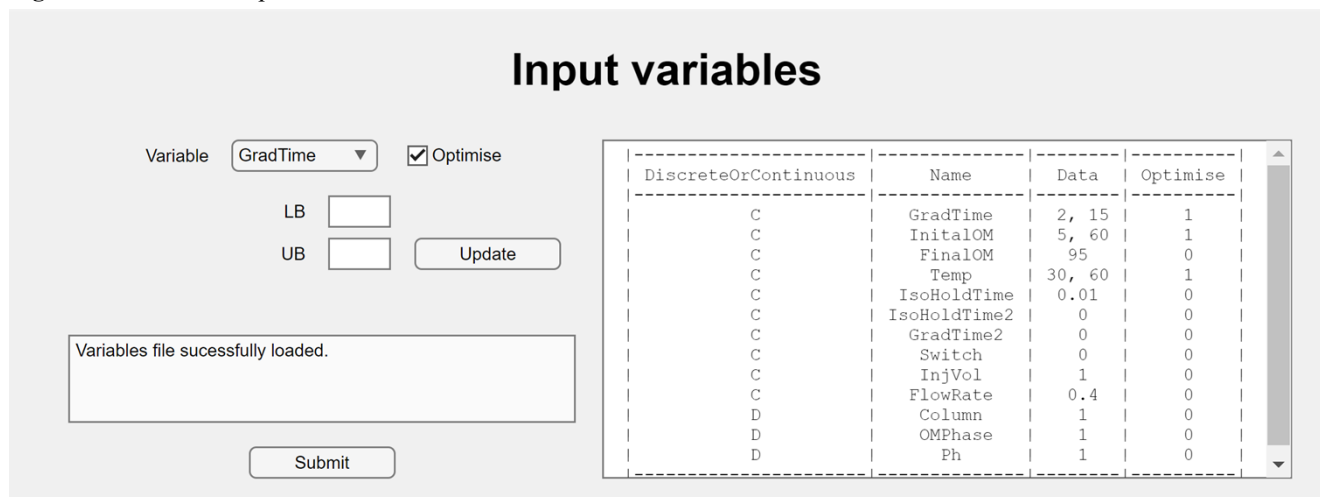


Figure S3: GUI that allows the user to configure the input variables to define the design space for the HPLC method optimization.

The user interface breaks down a HPLC optimization into three main steps. The first step involves defining the design space, optimization algorithm, objective functions, peak detection parameters, wavelength monitoring and other experimental parameters. Clicking the “Apply” button once these parameters are set will save the file so that data can be accessed later (via the “Load” button option). It will also generate all the initial experiments to be run. Step 2 involves running each of the generated initial condition HPLC methods. Clicking start will begin the automated closed loop process of setting the HPLC method, running the method and analysing the chromatogram data to obtain the objective function responses. These initial conditions provide a scan of the design space which is used in step three as a starting point for the optimization algorithm. Step 3 will start the optimization, incorporating the selected optimization algorithm to decide the next conditions to be run. The results are updated each time a method has finished, and this data can be visualized in the “Data visualisation” tab shown in Figure S2.

1.2. Bayesian optimization

Figure S4 shows a flow diagram of how a closed-loop HPLC method optimization is achieved.

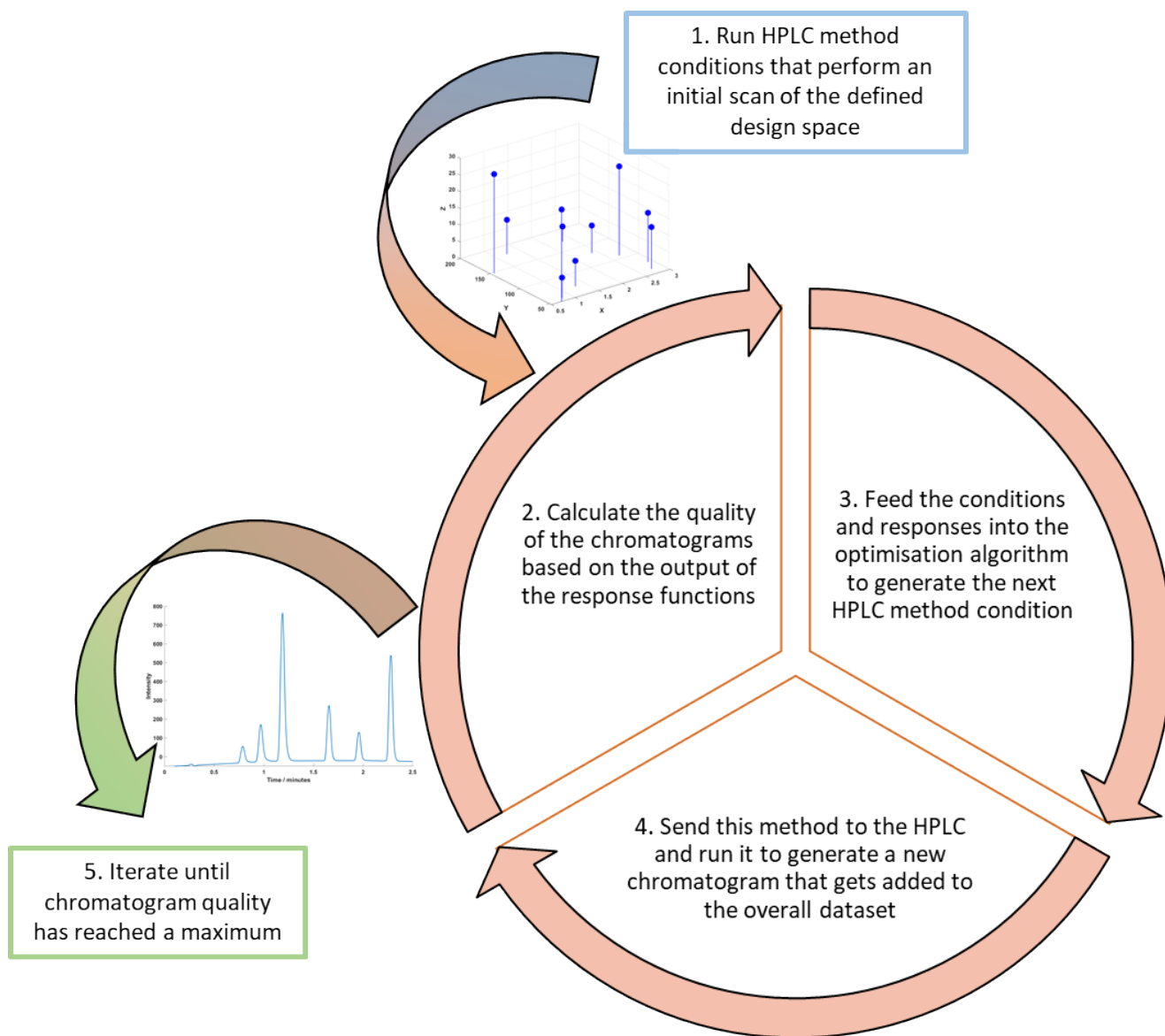


Figure S4: Flow chart detailing how a closed loop-optimization of HPLC methods can be achieved.

Both single and multi-objective Bayesian optimization algorithms were used to optimize HPLC method conditions. These algorithms however require suitable objective functions to be defined that quantify the overall quality of the chromatograms. The key parameters in this work that were used to define chromatogram quality were:

- The time the last peak elutes. This corresponds to the method time. The shorter the method time the more efficient the method.
- The critical resolution ($R_{s,Crit}$), which is the smallest measured resolution between any consecutive pair of peaks. Peaks need to be close enough to ensure that the method time is minimized but not too close to prevent overlap of peaks.
- The number of peaks detected. Maximising the number of peaks will help the algorithm to separate all the different components in the analyte.

By looking at, or a combination of, these different responses, the overall chromatogram quality can be calculated, and can therefore be used as an objective(s) for the chosen optimization algorithm. Four different objective functions were created for optimization. Each one coded required a table containing information about each peak including its retention time, height, and width at half height. Other information about peak cutoffs was also required.

For the single objective Bayesian optimization algorithm BOAEI², a weighted objective function was created that combined the time the last peak eluted, the critical resolution and the number of peaks chromatogram quality factors into one single objective. Code written to do this is available on a GitHub repository under the file name “responceCombFunc1.m”. For the multi-objective Bayesian optimization algorithm TS-EMO^{3, 4}, the time the last peak eluted, the critical resolution and the number of peaks

chromatogram quality factors were written as three separate objective functions under the name “responceMaxTime.m”, “responceRscrit.m” and “responceNumbPeaks.m” respectively.

1.3. Automated HPLC method transfer and execution using ChemStation macros

One step in the automation of HPLC method development involves both autonomously writing and executing HPLC methods to a HPLC system. ChemStation⁵ is used to create and run HPLC methods. The software works by sending commands to the HPLC hardware via serial communication. To access and use these serial communication commands to control the HPLC hardware directly was troublesome due to intellectual property issues and that they were not publicly available. This meant that another method of communication with the HPLC hardware was required. Application program interfaces are also available from companies, which allow for code to be used to communicate with HPLCs and write methods. However generally these can be costly, and so were not suitable for use in this project as a proof of concept for HPLC method optimization.

After further investigation, it was noted that ChemStation offered the ability for macros to be coded. A macro is code that can be used to automate tasks within a software package. Using guides provided by Agilent that were freely available online, macros were written that could modify HPLC method parameters and execute the method. This used Excel as a database, so that desired method conditions could be written to Excel before. Then using a Chemstation macro to read the Excel document, the HPLC method could be automatically modified accordingly before running the method. The macros had the ability to set up a Dynamic Data Link (DDL) with excel. This required a pre-written and open version of excel to transfer data dynamically between excel and ChemStation, details of which can be found in the macro programming guide written by Agilent. This enabled data transfer between software applications as well as autonomous control of the HPLC hardware. The macros written to automate HPLC method writing are available in the GitHub repository in the file “MACROS”.

Three macro function files were written to achieve automation:

- editMeth.MAC– When executed this macro would read and extract HPLC method data from the excel file containing the method information using a DDL. Different cells in the excel document would correspond to different method parameters, such as the timetable for the HPLC method, column temperature and injection volume.
- FullRun.MAC – This macro is executed in the command line of the ChemStation software by some external MATLAB code using a mouse clicker when a method is ready to be run. It ensures that the HPLC is in the “Ready” state, before executing the EditMeth macro described previously. It also creates a DDL with the method excel file to see if the method is ready to be run and obtain a equilibration time. The macro then waits for a set equilibration time before starting the HPLC method.
- updateExcel.MAC – Chemstation has a feature for a macro to be run automatically when a method has finished, defined as a post-run macro. This final macro is used as a post-run macro, and creates a DDL with the excel method file to write the method name and to notify to external MATLAB script that the method has finished running.

A summary of each of the events for HPLC method writing and execution is explained below.

- a. MATLAB generates a new HPLC method and it is written to an Excel file which acts as a sever for the current HPLC method information to be run.
- b. The MATLAB software uses an automated mouse and keyboard program to start the method execution macro (FullRun.MAC), by entering the macro command into command prompt in ChemStation and pressing enter to execute it.
- c. The method execution macro begins to switch on the equipment, check for equipment ready state and execute the EditMeth.MAC macro, which consequently will modify the HPLC method information based on what is stored in the method Excel file using a DDL. The macro then waits for an equilibration period before it starts the HPLC method.
- d. Once the HPLC method has finished running, the post-run macro is executed to update the method excel file with the HPLC file name so that it can be read by MATLAB for the optimization process.

1.4. Automated data analysis

Once a HPLC file has finished running, important information from the chromatogram needed to be extracted so that it could be fed into the optimization algorithm. For this work, the number of peaks, the critical resolution, and the time the last peak eluted from the chromatogram were extracted from the chromatogram.

Chemstation can automatically analyse chromatographic data and export it to a CSV file for analysis. However, it is not able to export the raw time-intensity data to a CSV automatically after the HPLC method was run. Therefore, some code was required to read the ChemStation ‘.ch’ files, which are specific to ChemStation to obtain this information. Code written by Dillon *et al.*⁶ was used to convert the raw Agilent ‘.ch’ files into a readable time intensity data array. Figure S5 shows a flow chart of how raw data is processed into objective responses that can then be fed into the optimization algorithms.

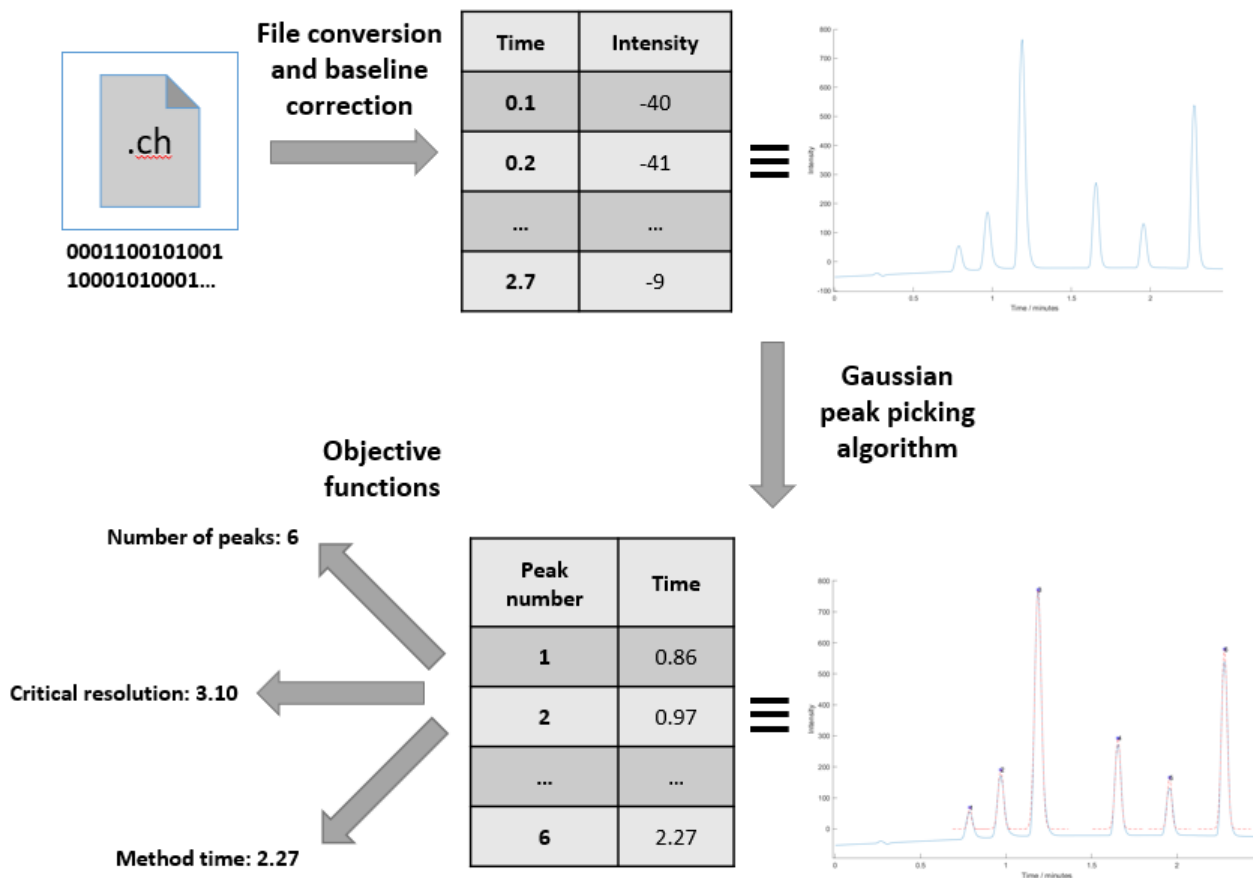


Figure S5: A flow chart detailing the process in which the raw chromatogram data file is processed into objective function responses that describe the quality of the chromatogram.

Peak picking and the associated parameters were extracted using two methods. Firstly, the peak information in the chromatogram was identified using an community made automated gaussian peak fitting algorithm written by O'Haver *et al.*⁷, which allowed for a time-intensity array to be passed into a function and outputted peak information in the form of a table. A function named "autofindpeaks.m" was used to obtain this information, and required the following parameters for peak detection:

- Slope Threshold : Peak detection is achieved by looking for downwards zero-crossings in the first derivative. Slopes that exceed the Slope threshold will be detected and so larger values will result in loss of smaller features.
- Amplitude threshold: Ignores peaks below this amplitude value.
- Smooth width: the width of the smoothed peaks before detection, where a larger number will result in ignoring narrower peaks
- Peak group: The number of data points around the top part of the peak which taken for measurement to fit the peak.
- Smooth type: The type of smoothing algorithm performed on the data, which choices of rectangular (sliding-average), triangular (2 passes of sliding-average) or pseudo-gaussian (3 passes of sliding-average)

By modifying these parameters, the sensitivity of peak detection could be modified to ensure all relevant peaks in the chromatogram were detected and modelled correctly to give accurate peak information.

ChemStation also has built in peak detection with parameters that can be modified, like the one explained above. By modifying the data acquisition method and selecting the setting that exports it to CSV after the method has finished, this approach can be used to automate the data processing instead of using the community made algorithm. This functionality has been built into the application shared within this work, however using ChemStation's built in peak detection was not used for this work.

Table S10. shows the constants for each of the parameters used for the peak picking algorithm for each of the optimizations run in this work.

Variable	Value
Slope Threshold	0.05
Amp Threshold	15
Smooth width	1
Peak groups	5
Smooth type	3

2. Experimental data

Results from Optimization 1 - 9 are shown in Tables S1 – S9 respectively.

2.1. Optimization 1

Table S1. The initial conditions, objectives, and overall weighted objective function (R) results for Optimization 1.

Experiment number	Gradient time / minutes	Initial organic modifier concentration (%)	Initial isocratic hold time / minutes	$R_{s,Crit2}$	Number of peaks	Time last peak eluted	R
1	4.23	58.1	2.77	1.35	6	3.37	0.199
2	1.85	33.3	3.95	2.26	7	6.54	0.030
3	8.84	51.4	5.21	1.52	6	5.63	0.183
4	6.95	43.3	0.26	0.84	7	4.50	0.195
5	6.00	5.5	8.55	6.85	6	14.30	0.151
6	2.57	21.1	9.94	1.15	7	13.16	0.187
7	8.21	16.2	6.90	7.89	6	13.53	0.147
8	10.00	31.6	10.00	2.25	7	15.82	0.072
9	1.00	34.5	9.74	1.96	7	11.66	0.059
10	10.00	33.3	1.55	2.24	7	7.62	0.035
11	1.30	11.0	0.20	0.69	7	2.61	0.208
12	1.01	32.8	2.24	2.28	7	4.29	0.020
13	1.00	36.6	8.49	1.65	7	10.39	0.100
14	1.00	32.3	0.00	1.23	7	2.08	0.125
15	2.62	33.7	8.84	2.18	7	11.76	0.053
16	4.16	33.7	8.51	2.16	7	12.15	0.055
17	1.64	34.7	2.76	1.97	7	5.20	0.028
18	2.70	29.8	4.25	2.19	7	7.41	0.034
19	1.60	29.6	6.54	2.18	7	8.99	0.041
20	9.99	24.6	4.37	1.71	7	11.26	0.095
21	6.37	60.0	10.00	1.28	6	3.00	0.207
22	3.65	31.4	3.05	2.25	7	6.70	0.030
23	5.27	31.6	5.58	2.25	7	9.92	0.045
24	2.32	45.5	10.00	0.66	7	9.60	0.244
25	5.48	21.4	0.00	1.38	7	4.98	0.115
26	5.09	33.3	2.80	2.27	7	7.04	0.032
27	1.00	29.2	2.18	2.12	7	4.26	0.019
28	1.00	39.1	4.06	1.29	7	5.99	0.134
29	1.00	25.8	2.14	1.82	7	4.25	0.046
30	10.00	28.8	1.62	2.08	7	8.17	0.037
31	1.00	29.3	8.12	2.13	7	10.16	0.046
32	9.96	5.0	4.67	8.30	6	12.82	0.144
33	1.01	31.1	1.84	2.21	7	3.92	0.018

34	1.00	16.2	3.43	2.12	6	5.60	0.111
35	1.00	25.2	6.86	1.73	7	8.96	0.081
36	1.01	35.4	1.82	1.85	7	3.85	0.039
37	10.00	29.8	7.36	2.15	7	13.63	0.062
38	1.01	32.5	6.93	2.23	7	8.94	0.041
39	9.37	60.0	0.03	1.26	6	2.60	0.208
40	1.18	31.2	2.25	2.21	7	4.44	0.020

2.2. Optimization 2

Table S2. The initial conditions, objectives, and overall weighted objective function (*R*) results for Optimization 2.

Experiment number	Gradient time / minutes	Initial organic modifier concentration (%)	Initial isocratic hold time / minutes	RsCrit2	Number of peaks	Time last peak eluted	R
1	8.84	16.2	7.25	7.75	6	14.21	0.150
2	5.19	38.2	4.68	1.42	7	8.48	0.125
3	6.38	22.3	0.17	1.37	7	5.59	0.120
4	2.29	50.4	9.19	1.63	6	6.17	0.170
5	8.39	53.7	1.88	1.52	6	4.25	0.177
6	1.24	34.5	6.98	2.02	7	9.10	0.041
7	3.87	9.3	3.61	5.01	6	7.88	0.122
8	1.00	33.1	7.44	2.25	7	9.43	0.043
9	2.44	31.8	0.00	1.72	7	3.01	0.055
10	1.00	23.0	3.28	1.43	7	5.41	0.110
11	9.99	27.8	9.92	2.03	7	16.35	0.074
12	6.71	5.0	8.50	7.13	6	14.69	0.153
13	9.99	33.9	9.55	2.15	7	15.02	0.068
14	1.00	37.5	5.01	1.49	7	6.95	0.108
15	1.00	28.7	5.03	2.10	7	7.10	0.032
16	1.01	30.8	7.77	2.21	7	9.80	0.045
17	1.00	27.3	9.43	1.99	7	11.49	0.053
18	1.00	60.0	6.26	1.26	6	2.93	0.210
19	1.00	29.4	2.80	2.14	7	4.88	0.022
20	1.00	28.5	0.00	0.95	7	2.10	0.168
21	1.00	31.6	3.43	2.23	7	5.48	0.025
22	10.00	30.0	3.70	2.17	7	10.05	0.046
23	1.00	31.5	4.82	2.23	7	6.86	0.031
24	1.00	34.1	1.24	2.09	7	3.29	0.015
25	1.00	35.5	0.00	1.60	7	2.05	0.069
26	1.00	32.5	1.73	2.24	7	3.79	0.017
27	9.96	33.8	2.07	2.14	7	8.02	0.036
28	1.00	25.5	7.09	1.81	7	9.19	0.070
29	1.00	34.1	3.14	2.11	7	5.16	0.023
30	1.02	38.0	10.00	1.43	7	11.82	0.139
31	1.00	22.9	10.00	1.41	7	12.11	0.144
32	1.00	15.5	0.00	0.68	7	2.18	0.208

33	10.00	5.0	0.00	0.75	7	8.16	0.224
34	10.00	33.7	5.69	2.21	7	11.45	0.052
35	1.00	10.8	10.00	0.95	7	12.19	0.213
36	1.00	27.2	3.74	1.96	7	5.84	0.032
37	1.00	45.6	0.00	0.66	7	1.96	0.210
38	1.00	30.7	9.99	2.20	6	12.00	0.140
39	10.00	28.2	7.37	2.07	7	13.82	0.063
40	1.00	35.3	2.16	1.86	7	4.18	0.039

2.3. Optimization 3

Table S3. The initial conditions, objectives, and overall weighted objective function (R) results for Optimization 3.

Experiment number	Gradient time / minutes	Initial organic modifier concentration (%)	Initial isocratic hold time / minutes	R_{sCrit2}	Number of peaks	Time last peak eluted	R
1	1.34	29.3	9.34	2.13	7	11.60	0.053
2	8.20	48.0	8.47	0.46	7	7.54	0.266
3	9.46	27.3	6.68	2.00	7	13.04	0.060
4	6.41	18.4	5.28	7.43	6	10.84	0.135
5	4.92	58.7	1.58	1.33	6	3.11	0.200
6	3.50	12.7	0.98	0.73	7	4.96	0.213
7	3.64	43.5	4.06	0.83	7	6.88	0.207
8	9.76	32.6	0.00	2.13	7	6.12	0.028
9	6.95	30.8	0.00	2.00	7	5.23	0.024
10	4.88	31.1	0.00	1.94	7	4.31	0.029
11	10.00	31.4	0.00	2.08	7	6.33	0.029
12	7.11	31.4	0.00	2.04	7	5.25	0.024
13	9.36	31.4	0.00	2.08	7	6.10	0.028
14	5.67	34.2	7.52	2.07	7	11.72	0.053
15	6.76	31.4	0.00	2.03	7	5.11	0.023
16	7.52	31.3	10.00	2.21	7	15.09	0.069
17	1.00	25.4	0.00	0.81	7	2.12	0.188
18	6.49	31.0	3.77	2.21	7	8.69	0.040
19	10.00	29.0	0.00	1.97	7	6.56	0.035
20	9.98	36.4	0.00	1.67	7	5.78	0.076
21	1.00	34.4	0.00	1.45	7	2.06	0.092
22	10.00	5.0	9.99	8.38	5	18.14	0.254
23	4.89	29.0	0.00	1.80	7	4.39	0.050
24	1.00	32.9	9.99	2.24	7	11.95	0.054
25	10.00	22.6	10.00	1.35	7	16.99	0.175
26	10.00	37.4	10.00	1.53	7	14.59	0.136
27	9.94	34.3	0.01	2.06	7	6.01	0.027
28	6.96	33.4	0.00	2.11	7	5.06	0.023
29	10.00	15.5	10.00	7.39	6	17.55	0.166
30	1.00	53.4	0.00	1.51	6	1.86	0.167
31	10.00	28.8	10.00	2.09	6	16.30	0.160

32	2.70	30.5	6.86	2.19	7	9.96	0.045
33	10.00	28.3	3.19	2.05	7	9.74	0.044
34	10.00	33.6	2.34	2.21	7	8.31	0.038
35	10.00	31.4	5.91	2.22	7	12.00	0.055
36	3.01	26.4	6.07	1.89	7	9.47	0.060
37	10.00	24.3	4.28	1.61	7	11.18	0.109
38	8.01	30.5	1.53	2.19	7	7.16	0.033
39	1.07	59.9	9.99	1.28	6	2.95	0.207
40	6.03	28.6	5.30	2.08	7	10.17	0.046

2.4. Optimization 4

Table S4. The initial conditions, objectives, and the distance to Pareto front metric results for Optimization 4.

Experiment number	Gradient time / minutes	Initial organic modifier concentration (%)	Initial isocratic hold time / minutes	$R_{s,Crit2}$	Number of peaks	Time last peak eluted	Distance to Pareto front
1	4.44	12.9	3.22	5.80	6	7.77	1.00
2	7.85	8.8	7.39	7.67	6	14.08	1.00
3	9.44	45.2	4.59	0.64	7	7.74	1.00
4	6.36	31.8	2.40	2.17	7	7.17	0.38
5	1.30	58.8	0.67	1.26	6	2.35	1.00
6	3.36	27.0	7.01	1.92	7	10.54	0.67
7	5.03	41.1	9.84	0.99	7	12.30	1.00
8	9.56	24.7	5.50	1.65	7	12.07	0.82
9	1.47	53.9	1.82	1.42	6	3.50	1.00
10	1.70	20.0	1.73	0.85	7	4.35	1.00
11	2.18	29.5	0.27	1.73	7	3.13	0.06
12	2.59	37.0	1.58	1.49	7	4.48	1.00
13	9.99	28.9	0.98	2.06	7	7.41	0.39
14	1.82	8.9	0.80	0.74	7	3.60	1.00
15	3.84	18.5	0.22	1.06	7	4.29	1.00
16	1.71	49.0	0.31	0.26	7	2.53	1.00
17	1.00	32.3	0.75	2.19	7	2.80	0.02
18	1.00	25.5	3.33	1.78	7	5.42	0.25
19	1.32	5.0	0.57	0.71	7	3.02	1.00
20	1.16	31.6	0.00	1.21	7	2.18	1.00
21	1.34	42.9	1.08	0.83	7	3.21	1.00
22	1.52	34.6	3.46	1.86	7	5.78	0.27
23	1.00	26.3	1.13	1.86	7	3.23	0.06

24	2.03	27.4	0.31	1.46	7	3.10	1.00
25	5.00	25.7	0.00	1.57	7	4.51	0.19
26	1.98	23.7	0.22	1.07	7	3.02	1.00
27	1.19	31.1	1.38	2.16	7	3.55	0.07
28	1.00	36.1	6.60	1.63	7	8.51	0.52
29	1.25	41.6	3.51	0.94	7	5.51	1.00
30	2.93	28.5	1.37	2.05	7	4.68	0.16
31	1.00	34.4	0.00	1.46	7	2.04	1.00
32	1.11	32.9	0.89	2.20	7	3.00	0.03
33	1.00	14.5	3.46	2.08	6	5.63	1.00
34	1.17	33.9	0.73	2.04	7	2.88	0.02
35	3.58	29.8	1.90	2.14	7	5.53	0.23
36	8.11	21.9	0.89	1.22	7	7.09	1.00
37	1.00	40.4	0.00	1.04	7	1.99	1.00
38	4.16	5.8	1.81	0.76	7	6.31	1.00
39	5.86	32.5	0.83	2.19	7	5.42	0.23
40	1.00	36.0	0.82	1.65	7	2.83	0.04
41	7.22	38.3	0.77	1.32	7	5.41	1.00
42	1.05	39.7	4.43	1.14	7	6.34	1.00
43	1.00	33.7	1.03	2.06	7	3.06	0.02
44	1.09	30.6	1.57	2.14	7	3.69	0.08
45	1.41	35.1	0.96	1.80	7	3.25	0.07
46	1.43	26.8	1.71	1.89	7	4.09	0.13
47	1.72	25.0	0.00	1.06	7	2.60	1.00
48	1.00	32.6	1.24	2.17	7	3.28	0.05
49	2.89	34.3	1.85	1.94	7	4.99	0.20
50	1.00	25.8	10.00	1.78	7	12.06	0.81

2.5. Optimization 5

Table S5. The initial conditions, objectives, and the distance to Pareto front metric results for Optimization 5.

Experiment number	Gradient time / minutes	Initial organic modifier concentration (%)	Initial isocratic hold time / minutes	$R_{s,Crit2}$	Number of peaks	Time last peak eluted	Distance to Pareto front
1	6.14	31.3	0.30	2.01	7	5.10	0.20
2	9.68	7.8	5.51	0.65	7	13.29	1.00
3	2.58	45.1	1.92	0.66	7	4.48	1.00
4	8.18	41.0	8.65	0.99	7	11.88	1.00

5	4.73	25.6	8.16	1.67	7	12.47	0.85
6	1.74	14.9	6.23	3.62	6	8.92	1.00
7	6.01	56.3	4.09	1.36	6	3.71	1.00
8	2.44	41.7	1.97	0.94	7	4.64	1.00
9	2.79	22.6	4.72	1.24	7	8.05	1.00
10	8.57	60.0	6.99	1.21	6	2.87	1.00
11	4.26	43.0	0.08	0.82	7	3.53	1.00
12	2.89	34.0	0.33	1.96	7	3.52	0.07
13	4.46	34.6	9.71	1.84	7	13.26	0.91
14	1.00	36.7	0.25	1.49	7	2.27	1.00
15	1.00	34.8	0.00	1.48	7	2.04	1.00
16	1.00	19.4	0.56	0.58	7	2.71	1.00
17	1.35	24.6	0.00	0.91	7	2.35	1.00
18	1.18	57.4	5.13	1.32	6	3.42	1.00
19	1.00	55.0	0.85	1.39	6	2.56	1.00
20	2.93	16.2	0.00	1.00	7	3.53	1.00
21	1.22	32.0	0.00	1.29	7	2.21	1.00
22	9.65	35.5	0.60	1.71	7	6.26	0.33
23	3.50	49.8	2.83	1.46	6	5.02	1.00
24	1.36	40.1	10.00	1.09	7	11.82	1.00
25	2.10	5.0	0.00	0.89	7	3.03	1.00
26	3.21	36.7	1.41	1.53	7	4.64	0.20
27	2.07	29.7	0.00	1.41	7	2.79	1.00
28	1.31	38.3	0.00	1.28	7	2.21	1.00
29	2.00	35.3	4.84	1.72	7	7.41	0.42
30	1.58	26.5	2.39	1.69	7	4.87	0.21
31	1.13	33.0	0.54	2.00	7	2.68	0.00
32	1.84	33.0	0.23	1.93	7	2.83	0.02
33	7.08	35.2	2.29	1.74	7	7.07	0.39
34	1.00	27.2	0.00	0.83	7	2.10	1.00
35	4.98	34.6	3.14	1.84	7	7.17	0.39
36	1.21	29.9	4.60	1.95	7	6.78	0.35
37	4.49	37.1	0.26	1.47	7	4.07	1.00
38	10.00	12.4	0.02	0.71	7	7.73	1.00

39	1.35	29.1	1.18	1.93	7	3.49	0.08
40	2.67	33.2	1.49	2.05	7	4.55	0.15
41	7.12	7.6	0.00	0.88	7	6.33	1.00
42	1.33	30.4	0.81	1.98	7	3.11	0.04
43	3.82	24.2	6.07	1.47	7	9.97	1.00
44	2.20	33.5	0.73	2.03	7	3.53	0.07
45	1.00	31.4	1.56	2.00	7	3.61	0.08
46	1.00	39.4	1.13	1.17	7	3.10	1.00
47	2.02	30.6	1.53	2.00	7	4.26	0.13
48	3.39	31.2	0.94	2.02	7	4.44	0.14
49	1.50	29.4	0.83	1.93	7	3.25	0.05
50	2.00	32.7	1.39	2.03	7	4.08	0.11

2.6. Optimization 6

Table S6. The initial conditions, objectives, and the distance to Pareto front metric results for Optimization 6.

Experiment number	Gradient time / minutes	Initial organic modifier concentration (%)	Initial isocratic hold time / minutes	$R_{s,Crit2}$	Number of peaks	Time last peak eluted	Distance to Pareto front
1	8.24	11.5	0.05	0.74	7	6.87	1.00
2	4.74	24.9	1.92	1.60	7	6.32	0.34
3	8.96	29.9	2.98	2.03	7	8.90	0.52
4	2.91	44.4	4.85	0.70	7	7.22	1.00
5	5.44	13.4	5.92	0.43	7	11.07	1.00
6	1.47	36.5	9.34	1.56	7	11.46	0.78
7	7.27	58.1	7.45	1.28	6	3.23	1.00
8	3.06	30.6	4.38	2.04	7	7.68	0.42
9	2.10	29.1	8.57	1.98	7	11.31	0.73
10	4.54	27.1	0.42	1.66	7	4.65	0.20
11	9.77	27.7	8.07	1.89	7	14.39	1.00
12	4.89	30.1	0.00	1.77	7	4.30	0.16
13	6.84	47.9	0.80	0.39	7	4.30	1.00
14	1.91	28.2	0.62	1.86	7	3.33	0.07
15	1.00	5.1	4.77	1.76	6	6.99	1.00
16	1.32	8.0	0.72	0.66	7	3.15	1.00
17	1.00	41.5	0.44	0.96	7	2.41	1.00
18	1.00	39.0	1.15	1.21	7	3.13	1.00
19	1.00	18.2	0.00	0.69	7	2.15	1.00

20	1.18	33.9	0.00	1.44	7	2.17	1.00
21	1.00	30.8	0.00	1.04	7	2.07	1.00
22	1.19	32.6	1.07	2.08	7	3.24	0.04
23	7.35	26.5	3.50	1.80	7	9.05	0.55
24	1.57	23.8	0.12	0.94	7	2.63	1.00
25	5.41	27.9	0.00	1.69	7	4.64	0.19
26	6.37	32.6	7.22	2.11	6	11.79	1.00
27	9.41	55.6	5.28	1.37	6	3.89	1.00
28	2.79	54.4	6.44	1.37	6	4.24	1.00
29	1.50	26.8	1.79	1.84	7	4.22	0.15
30	1.02	56.3	0.13	1.35	6	1.92	1.00
31	1.00	38.1	5.65	1.32	7	7.54	1.00
32	1.04	34.5	0.34	1.84	7	2.41	0.00
33	3.92	29.3	0.44	1.94	7	4.29	0.14
34	1.00	35.2	0.88	1.76	7	2.81	0.03
35	1.00	26.2	4.72	1.80	7	6.80	0.36
36	10.00	43.6	0.90	0.76	7	5.51	1.00
37	1.66	36.0	0.40	1.61	7	2.85	0.05
38	1.00	21.1	1.46	0.87	7	3.60	1.00
39	1.06	29.1	4.76	2.01	7	6.85	0.35
40	2.64	34.8	1.41	1.82	7	4.42	0.16
41	1.42	30.2	3.86	2.08	7	6.19	0.29
42	1.17	27.6	1.12	1.91	7	3.32	0.06
43	1.00	19.3	2.47	0.68	7	4.62	1.00
44	1.00	28.0	1.79	1.93	7	3.87	0.11
45	1.00	36.8	0.47	1.51	7	2.49	0.02
46	1.00	25.1	0.00	0.78	7	2.11	1.00
47	1.78	26.0	0.65	1.46	7	3.29	1.00
48	1.93	29.7	1.09	2.04	7	3.78	0.09
49	2.06	32.6	0.85	2.11	7	3.59	0.07
50	1.00	37.9	1.30	1.35	7	3.29	1.00

2.7. Optimization 7

Table S7. The initial conditions, objectives, and the distance to Pareto front metric results for Optimization 7.

Experiment number	Gradient time / minutes	Initial organic modifier concentration (%)	Initial isocratic hold time / minutes	R _{s,Crit2}	Number of peaks	Time last peak eluted	Distance to Pareto front
-------------------	-------------------------	--	---------------------------------------	----------------------	-----------------	-----------------------	--------------------------

1	9.76	51.9	3.13	0.63	11	5.06	1.00
2	5.43	38.0	9.97	1.20	11	13.34	1.00
3	4.10	13.3	1.64	0.95	11	5.98	1.00
4	6.19	57.9	5.22	0.28	10	3.31	1.00
5	3.14	8.5	6.66	0.99	11	10.43	1.00
6	8.18	28.0	7.90	1.37	11	13.59	1.00
7	1.80	34.4	0.64	0.30	11	3.19	1.00
8	9.31	18.8	1.68	1.56	11	8.65	0.57
9	9.59	28.8	0.29	1.64	11	6.63	0.17
10	7.23	43.5	1.20	0.97	11	5.24	1.00
11	1.81	21.0	3.57	0.71	11	6.27	1.00
12	9.81	26.4	4.33	1.50	11	10.89	1.00
13	3.09	53.9	0.00	0.55	10	2.52	1.00
14	8.07	17.3	1.60	1.53	11	8.05	0.49
15	9.49	45.9	4.75	0.89	11	7.70	1.00
16	2.17	41.7	0.37	0.54	11	3.01	1.00
17	4.17	39.3	0.19	1.01	11	3.79	1.00
18	4.93	30.8	0.56	1.40	11	4.85	1.00
19	9.83	8.5	0.00	1.56	11	7.85	0.43
20	10.00	37.0	0.44	1.38	11	6.04	1.00
21	6.79	32.5	0.27	1.58	11	5.26	0.00
22	9.85	17.1	0.52	1.60	11	7.88	0.42
23	6.49	27.0	0.00	1.54	11	5.19	0.05
24	8.06	31.3	0.00	1.66	11	5.57	0.05
25	6.44	5.0	3.03	1.34	10	9.04	1.00
26	7.11	11.7	0.74	1.46	11	6.93	1.00
27	5.87	22.0	0.37	1.44	11	5.49	1.00
28	2.14	48.3	0.00	0.75	10	2.47	1.00
29	1.00	15.0	0.00	0.28	10	2.18	1.00
30	6.66	31.4	0.17	1.58	11	5.18	0.01
31	1.00	48.8	1.78	0.74	11	3.56	1.00
32	2.03	60.0	1.89	0.56	9	2.85	1.00
33	5.18	17.4	6.20	0.64	11	3.98	1.00
34	4.04	41.3	3.16	1.10	11	6.35	1.00

35	4.54	24.2	0.92	1.28	11	5.25	1.00
36	5.52	39.9	1.11	1.19	11	5.07	1.00
37	5.40	34.0	5.47	1.35	11	9.62	1.00
38	4.88	43.0	0.00	0.97	11	3.65	1.00
39	9.12	22.3	7.97	1.46	10	14.56	1.00
40	5.80	26.9	2.71	1.42	11	7.54	1.00
41	7.46	33.5	0.19	1.62	11	5.37	0.00
42	7.86	28.7	1.01	1.47	11	6.66	1.00
43	3.51	21.2	0.03	0.65	11	3.86	1.00
44	5.97	30.1	0.80	1.42	11	5.59	1.00
45	4.99	49.3	1.47	0.73	11	4.35	1.00
46	1.12	49.2	5.40	0.72	11	6.52	1.00
47	7.23	24.8	0.31	1.47	11	5.96	1.00
48	5.15	25.6	5.79	1.26	10	10.34	1.00
49	9.30	33.0	2.40	1.49	11	8.14	1.00
50	9.83	41.3	4.27	1.12	11	8.64	1.00

2.8. Optimization 8

Table S8. The initial conditions, objectives, and the distance to Pareto front metric results for Optimization 8.

Experiment number	Gradient time / minutes	Initial organic modifier concentration (%)	Initial isocratic hold time / minutes	$R_{s,Crit2}$	Number of peaks	Time last peak eluted	Distance to Pareto front
1	2.76	34.8	5.65	0.97	10	8.61	1.00
2	5.77	43.0	6.92	0.97	11	9.64	1.00
3	6.49	47.6	2.13	0.77	11	5.26	1.00
4	9.72	14.1	1.16	1.25	11	8.61	1.00
5	1.52	53.7	9.16	0.56	11	4.71	1.00
6	7.98	21.5	8.18	1.19	9	14.30	1.00
7	4.28	7.4	2.86	0.55	11	7.41	1.00
8	4.31	48.3	9.87	0.74	10	7.04	1.00
9	10.00	51.8	1.51	0.64	11	4.54	1.00
10	8.92	28.5	1.57	1.34	11	7.65	1.00
11	6.09	27.2	6.77	1.19	10	11.66	1.00
12	10.00	33.6	1.30	1.38	11	7.24	1.00
13	1.36	29.5	2.46	0.60	11	4.76	1.00
14	2.03	20.2	0.14	0.82	10	3.01	1.00
15	1.73	41.7	0.53	0.30	11	2.93	1.00

16	3.86	54.1	0.00	0.52	10	2.72	1.00
17	9.31	53.8	0.58	0.55	11	3.68	1.00
18	3.43	16.9	0.89	0.64	11	4.73	1.00
19	3.05	28.0	1.93	1.05	11	5.32	1.00
20	5.65	27.0	0.44	1.31	11	5.22	1.00
21	7.42	60.0	2.57	1.17	9	2.96	1.00
22	6.23	60.0	0.00	0.95	9	2.46	1.00
23	2.06	54.7	2.31	0.50	11	3.93	1.00
24	5.96	50.5	0.21	0.67	11	3.46	1.00
25	4.00	38.7	0.00	0.80	11	3.56	1.00
26	1.00	27.0	0.06	0.78	9	2.16	1.00
27	2.14	13.3	0.19	0.78	10	3.18	1.00
28	3.04	21.6	1.27	0.71	11	4.78	1.00
29	7.64	38.4	2.45	1.25	11	7.04	1.00
30	1.49	60.0	0.43	1.05	9	2.21	1.00
31	1.50	50.7	3.20	0.66	11	4.85	1.00
32	6.94	58.8	7.24	0.17	10	3.21	1.00
33	7.90	35.8	0.45	1.42	11	5.56	1.00
34	8.53	38.3	4.71	1.26	11	9.32	1.00
35	1.66	33.5	0.00	0.85	10	2.49	1.00
36	6.38	15.6	0.09	1.31	11	5.70	1.00
37	1.84	46.8	4.77	0.80	11	6.60	1.00
38	3.65	7.1	8.55	0.50	10	12.67	1.00
39	7.07	29.6	1.15	1.43	11	6.40	1.00
40	7.20	45.8	0.29	0.86	11	4.21	1.00
41	4.31	30.9	0.00	0.92	11	4.00	1.00
42	1.42	5.0	1.14	0.58	9	3.64	1.00
43	4.52	25.8	0.21	0.99	11	4.48	1.00
44	7.96	33.1	0.17	1.48	11	5.54	1.00
45	5.38	39.2	0.65	1.20	11	4.65	1.00
46	5.30	18.2	1.20	1.29	11	6.13	1.00
47	9.18	49.8	0.77	0.70	11	4.30	1.00
48	6.62	26.6	0.09	1.42	11	5.35	1.00
49	3.91	32.2	1.11	1.24	11	4.84	1.00

50	8.15	49.0	0.29	0.72	11	3.98	1.00
----	------	------	------	------	----	------	------

2.9. Optimization 9

Table S9. The initial conditions, objectives, and the distance to Pareto front metric results for Optimization 9.

Experiment number	Gradient time / minutes	Initial organic modifier concentration (%)	Initial isocratic hold time / minutes	R _{s,Crit2}	Number of peaks	Time last peak eluted	Distance to Pareto front
1	9.94	53.9	6.20	0.55	11	4.52	1
2	4.54	6.5	2.88	0.34	11	7.62	1
3	2.67	29.8	1.96	1.05	11	5.09	1
4	6.95	25.4	0.24	1.44	11	5.72	1
5	1.45	19.1	5.64	0.72	10	8.09	1
6	8.68	39.1	9.93	1.19	11	13.61	1
7	5.64	46.0	7.53	0.84	10	8.69	1
8	10.00	19.9	3.00	1.45	11	10.20	1
9	9.15	22.9	1.37	1.48	11	7.98	1
10	9.78	32.7	0.15	1.58	11	6.18	0.136517
11	8.28	43.3	0.00	0.98	11	4.48	1
12	2.18	25.6	0.87	0.30	11	3.78	1
13	3.46	59.1	0.47	0.12	10	2.56	1
14	1.31	40.0	0.86	0.31	11	3.04	1
15	1.00	60.0	5.36	1.03	9	2.93	1
16	7.74	38.6	1.43	1.24	11	6.14	1
17	8.63	60.0	8.87	1.16	9	2.93	1
18	9.44	55.4	0.79	0.41	11	3.51	1
19	8.95	53.9	0.31	0.52	11	3.48	1
20	2.01	57.6	3.55	0.29	10	3.45	1
21	4.07	47.8	3.91	0.78	11	6.19	1
22	2.12	52.1	8.58	0.61	11	5.26	1
23	9.74	21.7	0.01	1.52	11	6.99	0.311462
24	10.00	50.1	0.70	0.68	11	4.29	1
25	10.00	41.3	0.28	1.09	11	5.31	1
26	8.47	54.3	5.71	0.51	11	4.36	1
27	2.29	41.8	4.53	1.07	11	6.96	1
28	5.63	33.3	1.49	1.41	11	5.91	1
29	10.00	54.7	10.00	0.49	11	4.24	1

30	3.05	32.5	6.62	1.06	11	9.80	1
31	7.64	52.6	0.72	0.59	11	3.78	1
32	6.61	37.9	1.78	1.30	11	6.19	1
33	1.12	36.4	1.00	0.75	10	3.09	1
34	8.33	28.7	0.34	1.49	11	6.19	1
35	1.00	26.7	5.30	0.73	10	7.37	1
36	7.79	31.4	0.12	1.49	11	5.56	1
37	1.00	53.0	9.20	0.57	11	4.86	1
38	2.69	51.7	5.82	0.63	11	5.40	1
39	1.00	54.6	0.83	0.41	11	2.57	1
40	1.38	24.8	0.00	0.59	10	2.37	1
41	1.30	8.5	7.19	0.50	10	9.59	1
42	3.66	30.9	0.00	0.74	11	3.67	1
43	6.34	50.9	0.04	0.59	11	3.33	1
44	2.52	31.8	0.09	0.31	11	3.12	1
45	8.84	27.6	0.00	1.46	11	6.14	1
46	4.32	32.8	1.25	1.28	11	5.14	1
47	1.00	48.2	1.77	0.74	11	3.57	1
48	1.00	13.9	0.31	0.68	9	2.48	1
49	1.09	52.4	3.95	0.61	11	5.08	1
50	1.00	33.2	2.27	0.52	11	4.29	1

3. Distance to Pareto front metric code

All the code used to compute the distance to Pareto front is available via GitHub.⁸

The code calculated the distance of each point in the trade-off graph with an $R_{sCrit} \geq 1.5$ to a line that predicted the path of the Pareto front. The Pareto front lines were fitted with linear and quadratic equations. Code was written that calculated the shortest distance from the Pareto front line to each point. Finally, all the points were normalized between 0 and 1.

4. Optimization traces

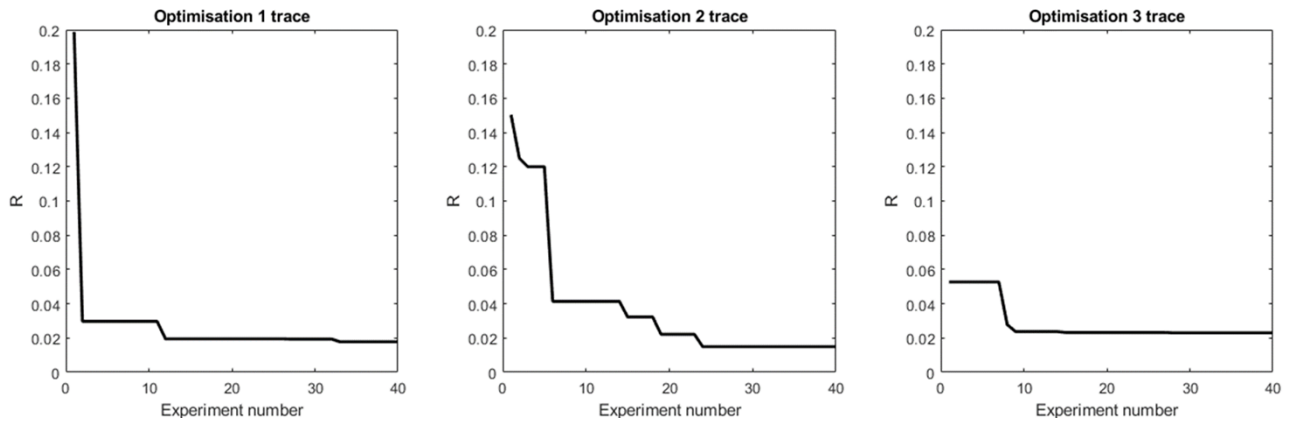


Figure S6. Traces for Optimizations 1-3 from the single-objective optimization algorithm BOAEI, showing the smallest overall value of R after each experiment.

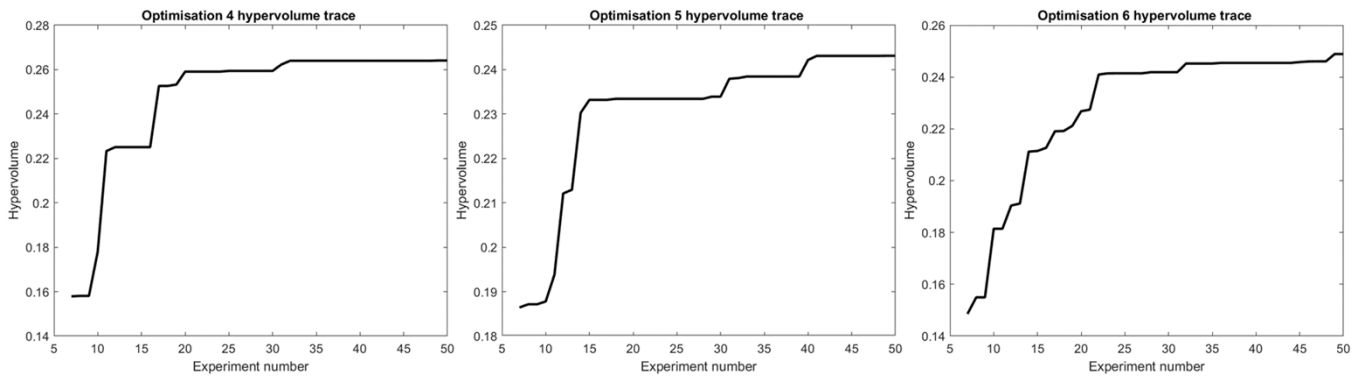


Figure S7. Traces for Optimisations 4-6 from the multi-objective optimization algorithm TS-EMO, showing the maximum hypervolume improvement after each experiment.

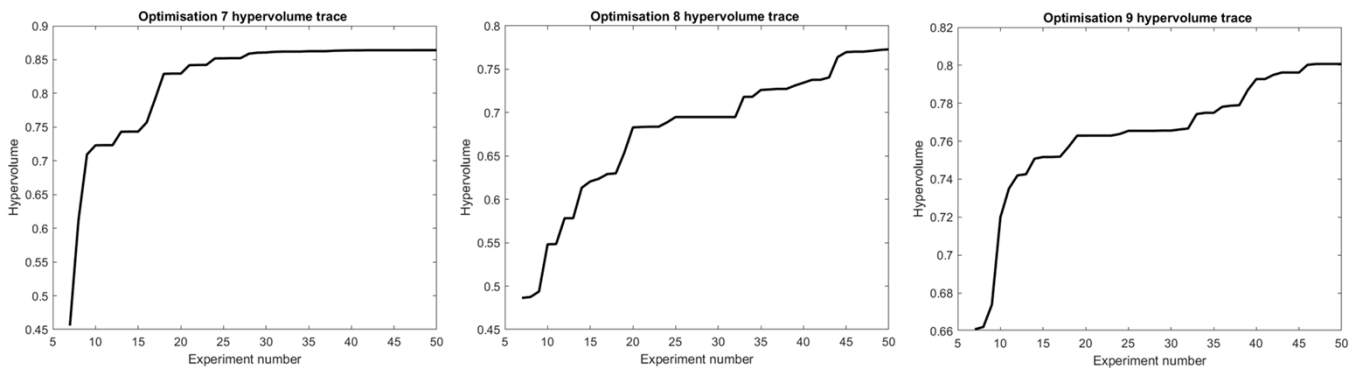


Figure S8. Traces for Optimizations 7-9 from the multi-objective optimization algorithm TS-EMO, showing the maximum hypervolume improvement after each experiment.

Hypervolume improvement was calculated using code written in MATLAB and is available via GitHub.⁹ The utopian and anti-utopian point for the objective space of each test problem were selected by creating a superset of the non-dominated solutions from all runs across all algorithms. The reference point for the objective space was then defined as the anti-utopian point shifted by 0.01 of the difference between the utopian and anti-utopian point.^{10, 11}

5. 2D matrix plots for alternative data visualisation

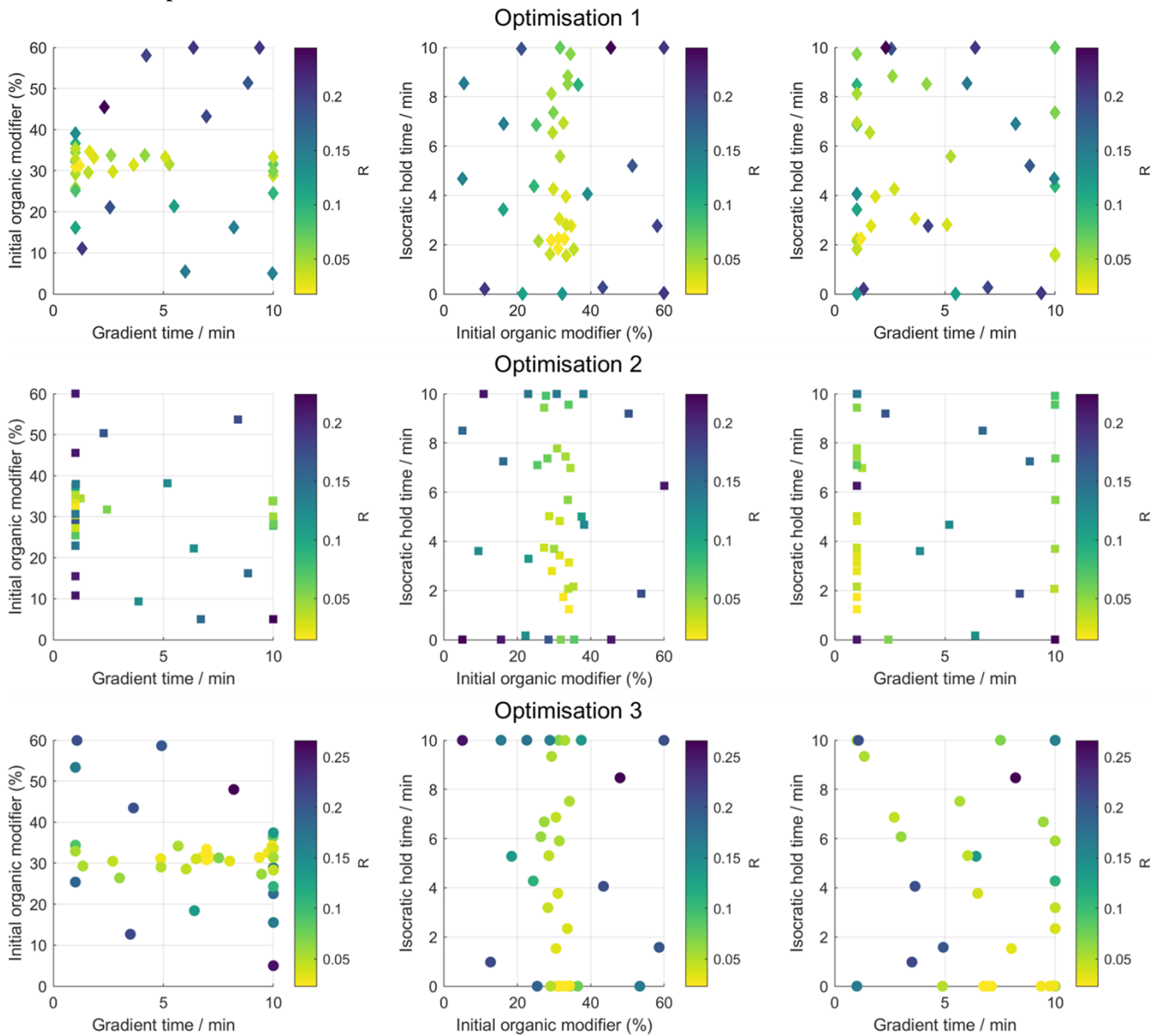


Figure S9. 2D matrix plots for the design spaces for Optimizations 1-3 The color scale represents the value of the weighted objective function R .

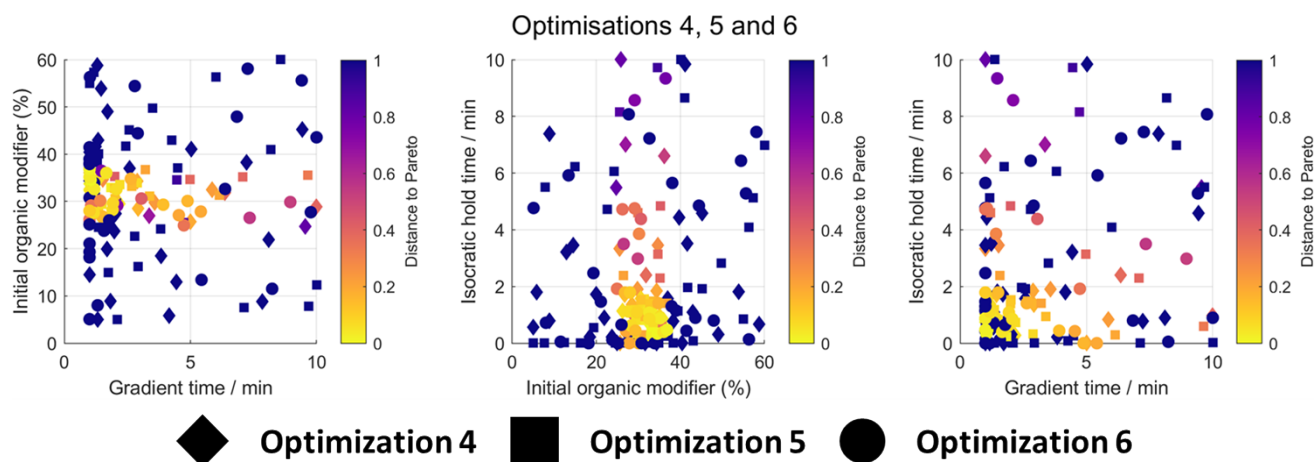


Figure S10. 2D matrix plots for the design spaces for Optimizations 4-6. The color scale represents the Distance to Pareto front when $R_{sCrit} \geq 1.5$,

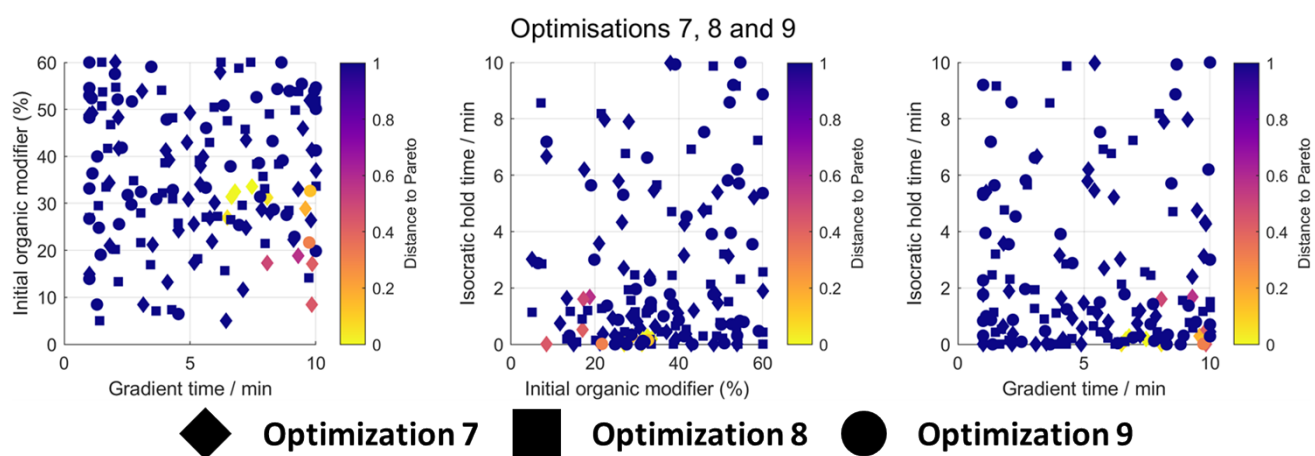


Figure S11. 2D matrix plots for the design spaces for Optimizations 1-3 The color scale represents the value of the weighted objective function R.

6. Graph code

Code used to make the graphs is available via GitHub and is written in Python.¹²

7. References

Reference List

1. T. Dixon, HPLCMethodOptimisationGUI, <https://github.com/Bourne-Group/HPLCMethodOptimisationGUI>, (accessed 10th January, 2024).
2. A. D. Clayton, E. O. Pyzer-Knapp, M. Purdie, M. F. Jones, A. Barthelme, J. Pavey, N. Kapur, T. W. Chamberlain, A. J. Blacker and R. A. Bourne, *Angewandte Chemie International Edition*, 2023, **62**, e202214511.
3. E. Bradford, A. M. Schweidtmann and A. Lapkin, *Journal of Global Optimization*, 2018, **71**, 407-438.
4. E. Bradford and A. Schweidtmann, Thompson sampling efficient multiobjective optimization, <https://github.com/Eric-Bradford/TS-EMO>, (accessed 10th August, 2021).

5. Agilent, OpenLab ChemStation, <https://www.agilent.com/en/product/software-informatics/analytical-software-suite/chromatography-data-systems/openlab-chemstation>, (accessed 18th August, 2021).
6. J. Dillon, Chemplexity, <https://github.com/chemplexity/chromatography>, (accessed 10th August, 2021).
7. T. O'Haver, Peak Finding and Measurement, <https://terpconnect.umd.edu/~toh/spectrum/PeakFindingandMeasurement.htm>, (accessed 10th August, 2021).
8. T. Dixon, HPLCMethodOptimisationMATLABParetoDistance, <https://github.com/Bourne-Group/HPLCMethodOptimisationMATLABParetoDistance>, (accessed 10th January, 2024).
9. T. Dixon, HPLCMethodOptimisationHypervolumeImprovement, <https://github.com/Bourne-Group/HPLCMethodOptimisationHypervolumeImprovement>, (accessed 9th May, 2024).
10. J. Knowles, *IEEE Transactions on Evolutionary Computation*, 2006, **10**, 50-66.
11. P. Müller, A. D. Clayton, J. Manson, S. Riley, O. S. May, N. Govan, S. Notman, S. V. Ley, T. W. Chamberlain and R. A. Bourne, *Reaction Chemistry & Engineering*, 2022, **7**, 987-993.
12. T. Dixon, HPLCMethodOptimisationPythonCode, <https://github.com/Bourne-Group/HPLCMethodOptimisationPythonCode>, (accessed 10th January, 2024).