Journal of Chromatography A 1726 (2024) 464941

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Enhancing LC×LC separations through multi-task Bayesian optimization

Jim Boelrijk [a,b,1], Stef R.A. Molenaar [c,d,1], Tijmen S. Bos [c,d,1], Tina A. Dahlseid [e], Bernd Ensing [a,f], Dwight R. Stoll [e], Patrick Forré [a,b], Bob W.J. Pirok [a,d,*]

[a] AI4Science Lab, Informatics Institute, University of Amsterdam, Amsterdam, Science Park 904, 1098 XH, The Netherlands
[b] AMLab, Informatics Institute, University of Amsterdam, Amsterdam, Science Park 904, 1098 XH, The Netherlands
[c] Division of Bioanalytical Chemistry, Amsterdam Institute of Molecular and Life Sciences, De Boelelaan 1085, Amsterdam, 1081 HV, The Netherlands
[d] Analytical Chemistry Group, Van 't Hoff Institute for Molecular Sciences, University of Amsterdam, Amsterdam, Science Park 904, 1098 XH, The Netherlands
[e] Department of Chemistry, Gustavus Adolphus College, Saint Peter, MN 56082, United States
[f] Computational Chemistry Group, Van 't Hoff Institute for Molecular Sciences, University of Amsterdam, Amsterdam, Science Park 904, 1098 XH, The Netherlands

ARTICLE INFO

Keywords:
Bayesian optimization
2D-LC
Closed-loop method development
Machine learning
Shifting gradients

ABSTRACT

Method development in comprehensive two-dimensional liquid chromatography (LC×LC) is a challenging process. The interdependencies between the two dimensions and the possibility of incorporating complex gradient profiles, such as multi-segmented gradients or shifting gradients, make trial-and-error method development time-consuming and highly dependent on user experience. Retention modeling and Bayesian optimization (BO) have been proposed as solutions to mitigate these issues. However, both approaches have their strengths and weaknesses. On the one hand, retention modeling, which approximates true retention behavior, depends on effective peak tracking and accurate retention time and width predictions, which are increasingly challenging for complex samples and advanced gradient assemblies. On the other hand, Bayesian optimization may require many experiments when dealing with many adjustable parameters, as in LC×LC. Therefore, in this work, we investigate the use of multi-task Bayesian optimization (MTBO), a method that can combine information from both retention modeling and experimental measurements. The algorithm was first tested and compared with BO using a synthetic retention modeling test case, where it was shown that MTBO finds better optima with fewer method-development iterations than conventional BO. Next, the algorithm was tested on the optimization of a method for a pesticide sample and we found that the algorithm was able to improve upon the initial scanning experiments. Multi-task Bayesian optimization is a promising technique in situations where modeling retention is challenging, and the high number of adjustable parameters and/or limited optimization budget makes traditional Bayesian optimization impractical.

1. Introduction

Comprehensive two-dimensional liquid chromatography (LC×LC) is a powerful technique for separating complex mixtures. When applied adequately, introducing a second dimension can substantially enhance peak capacity and resolution [1,2]. Therefore, LC×LC methods are developed for analyzing diverse samples, such as small molecules, polymers, proteins, oils, and bioactive compounds, among others [3–9]. The continual advancements in the LC×LC domain, providing enhanced capabilities for analyzing increasingly complex samples, contribute to the growing challenge of method development. To begin, analysts must select an appropriate retention mechanism for the first dimension (^{1}D) separation and then identify a complementary second dimension (^{2}D) retention mechanism. Once these and other system parameters have been set, the method can be refined using different gradient programs [10]. However, optimizing gradient settings through a "trial & error" method can be time-consuming.

Therefore, research has focused on computational methods to accelerate these design steps. Here, a distinction can be made between two kinds of optimization that we will refer to as direct and simulation-aided optimization. In direct optimization, an objective function is directly applied to an experimental measurement, and then black-box optimization techniques such as simplex methods [11,12], evolutionary algorithms [13,14], reinforcement learning [15], and Bayesian optimization [16,17] can be used to guide optimization. However, these methods generally require an increasing number of iterations as the

* Corresponding author at: Analytical Chemistry Group, Van 't Hoff Institute for Molecular Sciences, University of Amsterdam, Amsterdam, Science Park 904, 1098 XH, The Netherlands.
E-mail addresses: jim.boelrijk@gmail.com (J. Boelrijk), bob.pirok@uva.nl (B.W.J. Pirok).
[1] Equal contribution.

https://doi.org/10.1016/j.chroma.2024.464941
Received 16 February 2024; Received in revised form 17 April 2024; Accepted 22 April 2024
Available online 3 May 2024

number of adjustable parameters increases, and experimental measurements are both time-consuming and costly. Therefore, these approaches have generally been restricted to one-dimensional LC and/or GC with a handful of adjustable parameters.

In simulation-aided approaches, first, a set of experiments is performed based on scanning gradients [18–20] or on the Design of Experiment strategies (DoE) [21]. Next, the chromatographic peaks are detected and each unique compound is tracked over these measurements [22]. Then using this tracking information, regression models can be built. These models can be as flexible as neural networks [21], or can have a lower-dimensional fixed functional form where for instance only 2–3 parameters describe the retention behavior as a function of the mobile phase composition [23–26]. Here the latter approaches generally require significantly fewer initial experiments to build a reliable retention model.

Once this retention model is built, it can be used to simulate the separation under a broad range of conditions to optimize a predefined separation objective *in silico* using, for instance, a grid search [27], evolutionary algorithms [28,29], Bayesian optimization [16], or gradient descent-based methods [30]. If fully automated, this can then result in a bilevel optimization loop, where the optimal separation conditions found in the retention model can be performed as a real experiment, then with data from this new experiment, the retention model can be updated and optimized again, etc. By utilizing the retention model, satisfactory optima can generally be found with fewer experiments than in direct optimization settings. Therefore, these approaches are generally more suitable in situations where there are many adjustable parameters.

Although retention modeling has been successfully applied in both LC [28,30,31] and LC×LC [32], several challenges persist. The first challenge is peak-tracking; any inaccuracies in this step could lead to incorrectly predicted retention times. Also, compounds that have not been successfully resolved in the initial experiments might not be tracked and hence might be missing from the retention model.

Another challenge is to accurately predict the retention times and peak widths using the retention model itself. This is especially true in LC×LC when advanced gradient assemblies (e.g., shifting gradients) are used in the second dimension, as the typically slow speed of ^{2}D separation compared to ^{1}D peak widths results in a much lower number of data points in the first dimension compared to the second dimension. A prediction error in the ^{1}D retention time could lead to predicting the ^{2}D retention time in a wrongly assigned modulation, thereby propagating the error even further. In addition, as ^{2}D gradients are generally fast, gradient deformation may occur which can lead to less accurate retention time predictions [33,34]. This has been observed in LC×LC [32], where for a complex antibody sample (tryptic digest of a monoclonal antibody) the mean absolute percentage error in the retention time predictions were in the range of 5%–15% in the first dimension and 12%–35% in the second dimension, where the error decreased with more iterations. Mean absolute percentage errors of the peak widths were around 40% in the first dimension and 40%–60% in the second dimension. The combination of incorrectly tracked peaks and prediction errors in the retention times and peak widths introduces a bias in the retention model, where an optimum in the retention model may not necessarily lead to an optimum in the experimental separation. Despite the simulator bias, the proposed optimum of the retention model represents a significant improvement over initial experiments, as highlighted in [30,32]; however, there could be potential for further improvements.

Biased or lower fidelity simulators find widespread use throughout science and industry. In materials science for instance, proposed materials can be simulated with relatively inexpensive computer simulations (e.g., molecular dynamics) compared to significantly more expensive synthesis and characterization in a laboratory. Although these simulations are generally approximate, they still manage to significantly narrow down the search space of potential materials [35]. Likewise, when tuning internet services (e.g., adding a new feature, or testing a new ranking model), offline simulators can provide estimates of human interactions and offer higher throughput than deploying and testing all designs online [36]. Another example is automated machine learning, where the goal is to tune the hyperparameters of a model by minimizing the validation error after training on a large training set. While training on the full dataset can be costly, model performance could be estimated with lower fidelity on a subset of the training set, or by terminating training earlier, thereby finding optimal hyperparameters faster and/or at a reduced cost [37].

A technique that can combine information between these lower-fidelity simulators and the real problem at hand is multi-task Bayesian optimization (MTBO). Here the underlying idea is that if performance on various tasks (for instance the objective in the retention model and the real system) is correlated as a function of the optimizable parameters, we may accelerate optimization by transferring information between tasks [35,36,38]. At the heart of this method is a multi-task Gaussian process (MTGP) which can model the response surface jointly across tasks, and can be fitted to a variable number of data points on each task. This allows the MTGP to be fit on larger amounts of lower fidelity data than data on the expensive task at hand, or on historical data that enables a "warm start" to the optimization problem rather than starting from scratch. Because of this, MTBO generally is more scalable to a higher number of optimizable parameters and is more sample-efficient than conventional BO when tasks are correlated, making it a potentially suitable approach for method optimization in LC×LC. Especially given the fact that retention models generally are less accurate, and the number of adjustable parameters and timescale of measurements might make conventional BO unpractical.

Therefore, in this work, we explore the applicability of multi-task Bayesian optimization (MTBO) for tuning ^{1}D and ^{2}D gradient parameters in LC×LC with ^{2}D shifting gradients. We describe the experimental details and theory regarding the methods in Sections 2, and 3. In Section 4.1 we first set up a simulation framework using retention modeling to test our MTBO framework and benchmark it with single-task BO by optimizing a simulated 12-parameter problem containing retention parameters for 187 compounds determined from an IgG1 monoclonal antibody [39]. Code related to this section is shared open-source. Finally, in Section 4.2, we apply our BO and MTBO framework in the real world to the optimization of a LC×LC method for analysis of a complex pesticide sample by optimizing seven gradient parameters in a closed-loop, automated, and unsupervised fashion and show significant improvements over the initial experiments in under 20 method development iterations (MDIs).

2. Experimental

2.1. Chemicals

Acetonitrile was obtained from Sigma Aldrich (St. Louis, MO). Formic acid was supplied by Honeywell Research chemical (Muskegon, MI). HPLC-grade water was obtained from an in-house Milli-Q system (Burlington, MA). The studied sample was prepared from LC Multiresidue Pesticide Standard #4 (63 components) obtained from Restek. The standard was diluted from 100 μg mL^{-1} to 10 μg mL^{-1} with acetonitrile (ACN) and held at −20 °C until the time of analysis. The analytical sample was prepared by diluting a portion of the frozen standard to 1 μg mL^{-1} with 50:50 water:ethanol.

2.2. Chromatographic system

2.2.1. LC×LC-MS

The LC instrument used in the study was the Agilent Infinity II two-dimensional (2D)-LC system, consisting of two binary pumps (G7120A) with Jet Weaver V35 mixers (G7120-68135), an autosampler

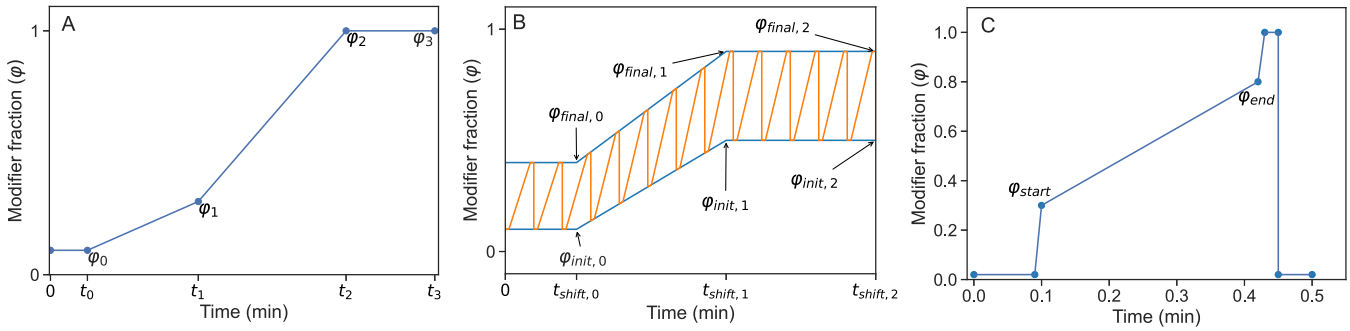

Fig. 1. Outline of the adjustable parameters of our 2D-LC setup (A-C). **(A)** Example of a gradient program in the first dimension consisting of two gradient steps. **(B)** Example of a ^{2}D shifting gradient. **(C)** Example of the gradient program of a single modulation in the second dimension. In every modulation, there is a fixed isocratic hold at the start and end of the program.

(G4226A), and two column ovens (G7116B, and G1316A). To connect the two dimensions, the active solvent modulation (ASM) valve interface (part number: 5067–4266) was configured with two 40 µL sample loops and a restriction capillary (170 × 0.12 mm, 1.9 µL), achieving an ASM factor of 3. Dwell volumes were estimated to be 0.225 mL in the first dimension and 0.070 mL plus the loop volume of 0.040 mL in the second dimension. Control of the 2D-LC instrument was performed using Agilent OpenLab CDS ChemStation Edition (C.01.10 [287]) with a 2D-LC add-on (rev. A.01.04 [033]). The mass spectrometer utilized was the Agilent Technologies quadrupole time-of-flight (Q-TOF) instrument (G6545XT) equipped with a Dual Agilent Jet Stream Electrospray Ionization (AJS ESI) source. Mass analyzer calibration was achieved using a standard tuning compound mixture (Agilent, part number: G1969-85000). A reference mass compound, hexakis (1H,1H,3H-perfluoropropoxy) phosphazene (mass-to-charge-ratio (m/z) 922.0098), was continuously sprayed into the electrospray source through a secondary reference nebulizer. Control of the Q-TOF was conducted using Agilent MassHunter Workstation Data Acquisition version 11.

2.2.2. LC columns

In the first dimension, an Agilent Zorbax Bonus-RP (2.1 × 100 mm, 3.5 µm) was used, whereas in the second dimension, we used an Agilent Zorbax SB-C18 (2.1 × 30 mm, 3.5 µm). The column dead volumes were estimated to be 0.243 mL in the first dimension and 0.069 mL in the second dimension, based on a total porosity of 0.7 and respective column dimensions.

2.3. Chromatographic conditions

2.3.1. First dimension

Gradient elution was performed using 0.1% formic acid in water (Solvent A) and ACN (Solvent B). Four scanning gradients were employed, with a linear gradient profile of 2%–100% B. However, the gradient time (t_G) varied across three different durations: 30 min, 40 min, 50 min, and 60 min. The temperature of the ^{1}D column was set to 40 °C, and the flow rate in the first dimension was 0.08 mL min^{-1}. Each analysis involved injecting 4 µL of pesticide mix at a concentration of 1 µg mL^{-1}.

2.3.2. Second dimension

Gradient elution was performed using 0.1% formic acid in water (Solvent A) and ACN (Solvent B). Four scanning gradients were utilized, maintaining a constant gradient profile of 2%–100% B. However, the gradient time (t_G) was varied across three durations: 9 s, 12 s, 15 s, and 19.2 s. At the start of each ^{2}D cycle, the mobile phase was held at 2% B for 5.4 s to serve as the diluent during ASM [40]. After 25.2 s the mobile phase was held at 100%B until 27 s, after which the column reequilibrated to 2% B until the end of the modulation time at 30 s. All other gradient profiles were computed using the algorithm described in earlier work [32] and uploaded to the LC×LC system using an in-house C++ script. The temperature of the ^{2}D column was maintained at 60 °C, and the flow rate for the ^{2}D column was set to 2 mL min^{-1}. To achieve the desired split ratio, the ^{2}D flow entering the MS nebulizer was reduced to approximately 0.3 mL min^{-1} using a simple tee split and short, narrow restriction capillaries (75 µm i.d.).

2.4. MS instrument and conditions

The Q-TOF mass spectrometer was operated in positive ion mode. The following MS settings were used: gas temperature, 225 °C; drying gas, 12 L min^{-1}; nebulizer, 35 psi; sheath gas temperature, 350 °C; sheath gas flow, 11 L min^{-1}; VCap, 3500 V; nozzle voltage, 500 V; mass range, 50 – 1200 m/z; acquisition rate, 10 spectra s^{-1}.

2.5. Software

2.5.1. In silico case study

In the *in silico* case study in Section 4.1, we used an in-house Python (version 3.8.13) implementation of multi-linear gradient retention modeling using the linear solvent strength (LSS) model that is described in [17]. For this work, it was extended to support shifting gradients. It utilizes the peak-compression model of Hao et al. [41] to predict peak widths. Retention parameters are based on the tryptic digest of a monoclonal antibody sample used in Molenaar et al. [32], and can be found in the Supplementary Information, including other specified modeling parameters. All Bayesian optimization code was implemented in Ax (version 0.2.5.1) and BoTorch (version 0.6.4) [42] and was run on Intel(R) Xeon(R) CPU E5-2640 CPUv4 with 64 GB RAM.

2.5.2. Closed-loop platform

For most components we reused earlier developed software described in [32] and its details are reiterated where deemed necessary. Peak tracking was performed with previously developed algorithms [22,43]. Multi-linear gradient retention modeling was performed with the LSS model and was written in-house in Matlab (R2021b). Retention parameters were estimated using the *multistart* function in conjunction with *fmincon* with an optimality tolerance of 10^{-6} and a maximum of 3000 function evaluation. Retention model optima were calculated using the *ga* function. We used an in-house version of the peak compression model of Hao et al. [41] to predict peak widths. Individual plate numbers per compound were estimated using the *fminsearch* function. We established overall system interactions between the MS, LC×LC, and Matlab algorithms via Python (version 3.8.12). The methods applied to the LC×LC were initiated using C++ in Visual Studio 2022, similar to earlier research [17,30,32]. The Bayesian optimization framework was implemented in the Python package of Ax [42] (version 0.6.4). All Bayesian optimization code was run on

CPU and details related to the hardware can be found in previous work ([30], system A). Raw MS data were converted to .mz5 format using ProteoWizard 3.0.22144 64-bit [44], using a threshold count of 2000. This counts only the 2000 most intense peaks. In addition, as some consistent background was present, some background masses were excluded, as described in the Supplementary materials.

3. Theory & methods

3.1. Shifting gradients in two-dimensional chromatography

In conventional LC×LC, the second dimension elution conditions remain constant throughout the analysis. This is reasonable if the two separation mechanisms are orthogonal; however, this is generally not the case, especially when reversed-phase separations are used in both dimensions. In this situation, the ^{2}D separation might benefit from an advanced assembly of gradients that shift as a function of the ^{1}D mobile-phase composition program to increase the utilization of the separation space [39,45]. Although powerful, this will result in more complex method optimization compared to conventional LC×LC and thus is avoided in some cases [46], leading to an under-optimized method.

Fig. 1 outlines the parameters of the LC×LC gradient program. In the first dimension (See Fig. 1A), the gradient nodes (φ_i, t_i), describe the transition points between consecutive linear segments. Gradient nodes can be added or removed depending on the desired flexibility/complexity of the ^{1}D program. Throughout this work, we ensure through inequality constraints that $\varphi_i \leq \varphi_{i+1}$ and $t_i \leq t_{i+1}$ for all i in $\mathcal{I}$ gradient nodes. In addition, we kept the start point fixed at $\varphi_0 = 0.02$ at $t_0 = t_{\text{init}}$, and the end point fixed at $\varphi_I = 1$ at $t_I = t_{\text{max}}$, where t_{max} is some pre-specified maximum measurement time. This is done to ensure that all analytes elute.

The shifting gradient program in the second dimension, shown in Fig. 1B, is specified by a lower shift bound denoted with $\varphi_{\text{init},i}$ and an upper shift bound denoted by $\varphi_{\text{final},i}$ at time points $t_{\text{shift},i}$. Each modulation (shown in Fig. 1C) follows these lower and upper bounds depending on the ^{1}D time which is programmed in φ_{start} and φ_{end}. Although the shifting times of the lower and upper bound do not necessarily have to be at the same point in time, we do fix this to reduce the number of adjustable parameters. Therefore, each additional shifting step introduces at most three additional parameters. Unless stated otherwise, we keep the time of the start of the lower bound and the end of the upper bound of the shifting gradient fixed, so that: $t_{\text{shift},0} = t_{\text{init}}$, and $t_{\text{shift},2} = t_{\text{max}}$. In the real-world experiments described in Section 4.2, we set $t_{\text{init}} = 1.44$ min, and $t_{\text{max}} = 62$ min and set $\varphi_{\text{init},1} = \varphi_{\text{init},2}$, and $\varphi_{\text{final},1} = \varphi_{\text{final},2}$. In addition, we set inequality constraints so that consecutive lower and upper bound points and timepoints are always increasing, i.e, $\varphi_{\text{init},i} \leq \varphi_{\text{init},i+1}$, $\varphi_{\text{final},i} \leq \varphi_{\text{final},i+1}$ and $t_{\text{shift},i} \leq t_{\text{shift},i+1}$ for all i in $\mathcal{I}$ shifting steps. Lastly, the lower and upper bounds are not allowed to cross by ensuring that $\varphi_{\text{init},i} < \varphi_{\text{final},i}$ for all $\mathcal{I}$ shifting steps.

In the experiments in Section 4.2, ASM was used. Therefore, we kept an initial hold of 0.09 min at $\varphi = 0.02$ at each ^{2}D modulation before starting the programmed shifting gradient from 0.1 to 0.45 min followed by a fixed endpoint at $\varphi = 1.0$, as shown in Fig. 1C. Not maintaining the isocratic hold during the ASM time caused considerable peak broadening and less stable retention times.

To do retention modeling with shifting gradients, we follow the same approach as in our previous work [32] and refer the reader to the detailed description there. In short, retention modeling in the second dimension requires the determination of the ^{1}D retention time so that the ^{2}D modulation number can be determined and accompanying lower and upper bounds of the gradient program of that modulation. If the ^{1}D retention time was predicted to be after $t_{\text{shift},\mathcal{I}}$, the final ^{2}D conditions were used.

3.2. Objective function

Many objective functions, or chromatographic response functions (CRFs) have been developed to assess the quality of separation of a chromatogram in both LC and 2D-LC [47]. However, most CRFs are designed to either assess the quality of a separation predicted based on a retention model or that of an experimental measurement, but not both. For example, some CRFs include terms such as the number of observed peaks, which is useful for experimental measurements, but not for retention modeling. As with retention modeling, once the retention model is built, the number of compounds in the retention model is effectively constant. In some cases, a resolution score is normalized by the number of observed peaks, which is sensible in the retention modeling case, but not in the case of experimental measurements, as the number of observed peaks might fluctuate over measurements, and hence scores are not comparable over measurements. In our setting, we need an objective function, or CRF that should be applicable to both experimental separations, and separations predicted from retention models, and should be as correlated as possible. Therefore, as a separation criterion, we use the two-dimensional resolution between two peaks i and j:

$$R^*_{S_{i,j}} = \sqrt{\frac{(t^1_{R,i} - t^1_{R,j})^2}{\left[2\left(\sigma^1_i + \sigma^1_j\right)\right]^2} + \frac{(t^2_{R,i} - t^2_{R,j})^2}{\left[2\left(\sigma^2_i + \sigma^2_j\right)\right]^2}} \tag{1}$$

where $t^1_{R,i}$ and $t^2_{R,i}$ are the retention times in the first- and second-dimension. σ^1_i, and σ^2_i are the corresponding standard deviations of the peaks. These resolutions are normalized as follows:

$$Rs_{i,j} = \begin{cases} \frac{R^*_{S_{i,j}}}{2.0} \text{ if } R^*_{S_{i,j}} < 2.0 \\ 1 \text{ if } R^*_{S_{i,j}} \geq 2.0 \end{cases} \tag{2}$$

We then take the sum of the resolution of all nearest neighbor pairs in the chromatogram as the final objective:

$$\sum_i^N \min_{j \neq i} Rs_{i,j} \tag{3}$$

This CRF is designed to be applicable to both simulated and experimental chromatograms. Its emphasis lies in maximizing the number of resolved peak pairs within the specified separation time.

3.3. Gaussian processes

For single-task Bayesian optimization, we use a Gaussian process (GP) as the probabilistic regression model due to its proven performance in the small data regime [37,48]. A GP is a distribution over functions, which is parameterized by a mean function $\mu(\mathbf{x})$ and a covariance function $k(\mathbf{x}, \mathbf{x}')$ which generally is called the kernel function. Now given a regression problem with N pairs of observations $\left\{\left(\mathbf{x}_i, y_i\right)\right\}_{i=1}^N$, where for simplicity we consider these to be noiseless so that we have outputs $\boldsymbol{y} = \left[y(\mathbf{x}_1), y(\mathbf{x}_2), \ldots, y(\mathbf{x}_n)\right]^{\mathrm{T}}$, and inputs $\mathbf{X} = \left[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\right]^{\mathrm{T}}$, the Gaussian process for $\mathbf{y}$ can be described as:

$$\mathbf{y} = \begin{bmatrix} y\left(\mathbf{x}_1\right) \\ \vdots \\ y\left(\mathbf{x}_N\right) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu\left(\mathbf{x}_1\right) \\ \vdots \\ \mu\left(\mathbf{x}_N\right) \end{bmatrix}, \begin{bmatrix} \kappa\left(\mathbf{x}_1, \mathbf{x}_1\right) & \ldots & \kappa\left(\mathbf{x}_1, \mathbf{x}_N\right) \\ \vdots & \ddots & \vdots \\ \kappa\left(\mathbf{x}_N, \mathbf{x}_1\right) & \ldots & \kappa\left(\mathbf{x}_N, \mathbf{x}_N\right) \end{bmatrix}\right) \tag{4}$$

Or in more compact notation:

$$\mathbf{y} \sim \mathcal{N}\left(\boldsymbol{\mu}(\mathbf{X}), K(\mathbf{X}, \mathbf{X})\right) \tag{5}$$

Here $\mathcal{N}$ is the normal distribution, and $K(\mathbf{X}, \mathbf{X})$ is the Gram matrix (i.e., the right-hand side of the normal distribution in Eq. (4)). As we normalize the inputs (between 0 and 1), and standardize the outputs

to zero mean and unit variance, we can set $\boldsymbol{\mu}(\mathbf{X}) = \mathbf{0}$ so that the GP is entirely described by the kernel function.

As a kernel function, we use the Matérn-5/2 kernel with automatic relevance determination (ARD), which is defined as:

$$k(\mathbf{x}, \mathbf{x}') = \theta_0 \left(1 + \sqrt{5}r + \frac{5r^2}{3}\right) \exp\left(-\sqrt{5}r\right) \tag{6}$$

where r is a weighted distance:

$$r = \sqrt{\sum_{d=1}^{D} \left(\frac{x_d - x'_d}{\theta_d}\right)^2} \tag{7}$$

Here θ_0 is a scaling factor controlling the horizontal scale, and $\theta_{1,\ldots,D}$ are length scale parameters that govern the smoothness of the functions.

The parameters $\boldsymbol{\theta}$ can be inferred by maximizing the log marginal likelihood, which is the probability that the model predicts the training outputs given the inputs and kernel parameters and is defined as:

$$\ln p(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\theta}) = -\frac{1}{2}\mathbf{y}^T \left[K(\mathbf{X}, \mathbf{X})\right]^{-1} \mathbf{y} - \frac{1}{2} \ln |K(\mathbf{X}, \mathbf{X})| - \frac{N}{2} \ln 2\pi \tag{8}$$

Where the first term is a data-fit term, the second term is a complexity penalty, which favors longer length scales over shorter ones (i.e., smooth over oscillating functions) and hence regularizes overfitting. The third parameter is a constant originating from the normalizing constant of the normal distribution.

To make predictions for unseen test inputs $\mathbf{X}^*$ we can define the joint distribution of both the previous observations and the test inputs, so that:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}(\mathbf{X}) \\ \boldsymbol{\mu}(\mathbf{X}^*) \end{bmatrix}, \begin{bmatrix} \mathbf{K}(\mathbf{X}, \mathbf{X}) & \mathbf{K}(\mathbf{X}, \mathbf{X}^*) \\ \mathbf{K}(\mathbf{X}^*, \mathbf{X}) & \mathbf{K}(\mathbf{X}^*, \mathbf{X}^*) \end{bmatrix}\right) \tag{9}$$

Then the conditioning properties for Gaussians allow for the computation of the posterior predictive distribution in closed form:

$$p\left(\mathbf{y}^* \mid \mathbf{X}^*, \mathbf{X}, \mathbf{y}\right) = \mathcal{N}\left(\mathbf{y}^* \mid \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*\right) \tag{10}$$

with

$$\boldsymbol{\mu}^* = \boldsymbol{\mu}\left(\mathbf{X}^*\right) + \mathbf{K}(\mathbf{X}^*, \mathbf{X})^T \left[\mathbf{K}(\mathbf{X}, \mathbf{X})\right]^{-1} \left(\mathbf{y} - \boldsymbol{\mu}\left(\mathbf{X}\right)\right) \tag{11}$$

and

$$\boldsymbol{\Sigma}^* = \mathbf{K}(\mathbf{X}^*, \mathbf{X}^*) - \mathbf{K}(\mathbf{X}^*, \mathbf{X}) \left[\mathbf{K}(\mathbf{X}, \mathbf{X})\right]^{-1} \mathbf{K}(\mathbf{X}, \mathbf{X}^*) \tag{12}$$

For a more detailed description of GPs we refer the reader to Rasmussen and Williams [49].

3.4. Multi-task Gaussian processes

In MTGP regression, besides input pairs, we have an output for each task, so that for N inputs $\mathbf{X} = \left[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\right]^{\mathrm{T}}$ we have a set of responses for T tasks as $\mathbf{y} = (y_{11}, \ldots, y_{N1}, \ldots, y_{12}, \ldots, y_{N2}, \ldots, y_{1T}, \ldots, y_{NT})^{\mathrm{T}}$, where $y_{i,t}$ is the response for task t on input $\mathbf{x}_i$. In the simplest case, we can treat each task as a separate regression task and can use an independent GP for each task following the description above. However, we want to exploit the correlation between each task. This can be done by introducing a kernel that incorporates both tasks and inputs, i.e., $k((\mathbf{x}, t), (\mathbf{x}', t'))$. One such approach is the intrinsic model of coregionalization (ICM), which decouples the task intercorrelation and the data domain and is defined as follows:

$$k_{\mathrm{ICM}} = k_{\mathrm{T}}(t, t') \otimes k(\mathbf{x}, \mathbf{x}') \tag{13}$$

where $\otimes$ is the Kronecker product [50]. Here $k_{\mathrm{T}}(t, t')$ can be interpreted as a $T \times T$ matrix of trainable intertask correlation parameters. Given N inputs for both tasks, the resulting Gram matrix will have dimensions $TN \times TN$, and will thus induce some additional compute over a conventional GP. Then using k_{ICM} we again have a similar structure as a conventional GP and may use the equations above to fit kernel parameters and do predictions.

3.5. Bayesian optimization

In Bayesian Optimization we aim to solve the optimization problem:

$$\mathbf{x}^* \in \underset{\mathbf{x} \in \mathcal{X}}{\operatorname{argmax}}\, y(\mathbf{x}) \tag{14}$$

where $y(\mathbf{x})$ is the objective function score we observe via experiments $\mathbf{x}$. BO utilizes the GP model (or the MTGP model in case of MTGP) and an acquisition function to guide optimization. The acquisition function uses the mean prediction and their variance from the GP model to find points that are likely to improve upon the previously performed experiments by trading off exploration (high variance predictions) and exploitation (high mean predictions).

In the context of multi-fidelity BO, the acquisition function can also incorporate terms regarding the cost of the queried task, where simulator queries are significantly less expensive than performing an experiment in the laboratory. In this way, an optimum can be found at the lowest possible cost by alternating between the simulations and experiments.

Throughout this work we use the Noisy Expected Improvement acquisition function (NEI) [36,51] as it handles measurement noise robustly, and supports batch optimization. Following the approach of Letham et al. [36], we exclusively apply NEI to the real task of the MTGP (i.e., the part of the MTGP predicting the objective function for the real measurement). This enables us to generate a batch of promising method parameters. Subsequently, we assess these parameters on the simulator, recording their corresponding CRF scores and refitting the MTGP with these new observations. Next, we reevaluate the generated batch on the real task using the updated model, selecting the method parameters with the highest CRF score to use in a real experiment. Thus, the real task remains the primary driver of the optimization process. At the same time, the simulator serves to reduce predictive uncertainty and aids in filtering out predictions with high variance and low mean. The entire MTBO loop can be outlined as follows:

1. Fit MTGP to scanning experiments and randomly drawn method parameters evaluated using the simulator.
2. Generate candidate method parameters using NEI on the real task.
3. Evaluate method parameters using the simulator and observe their objective function score.
4. Refit the model with the new simulator observations.
5. Evaluate the generated candidate parameters on the real task and select the best-performing candidate to use in an experiment.
6. Update the model with the experimental results and repeat from step 3.

4. Results & discussion

4.1. In silico case study

4.1.1. Setting up a simulator pair

To test and evaluate the MTBO framework and compare it with single-task BO, we set up a simulator pair that allows us to run many *in silico* BO experiments. The simulator pair consists of a "ground-truth" simulator and a "biased" simulator. Here in a real-world setting, the ground truth simulator would describe the laboratory experiments, and the biased simulator would use the accompanying retention model. The ground-truth simulator contains retention parameters for 187 peptides found in a tryptic digest of an IgG1 monoclonal antibody characterized using LC×LC [32]. The biased simulator is created by introducing normally distributed random noise to the predicted retention times and peak widths and by randomly removing 30 compounds. This approach attempts to mimic the error that would be encountered in a realistic setting, where not all compounds can be sufficiently tracked and there

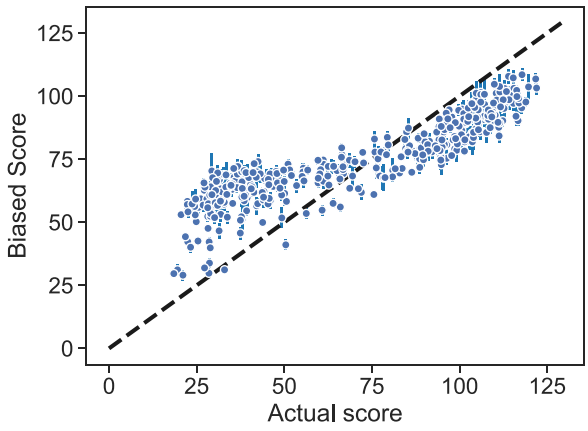

Fig. 2. Biased simulator observations compared to the ground truth simulator consisting of 400 measurements. Each dot represents a set of parameters that was evaluated in both simulators. For the biased simulator, the outcome of each parameter set is stochastic due to the randomly drawn noise and was therefore computed three times, and the respective standard deviations are shown by the error bars. Although the biased simulator scores are correlated with the ground truth simulator, using the biased simulator to optimize the problem directly is challenging.

are prediction errors in the retention model. Details regarding the added noise and other retention parameters can be found in Section S1 of the Supplementary Information. We note that we do not include peaks in the retention model, that are not present in the true system (i.e., false positives), which could in reality be another source of bias, we cover this in more detail, along with other bias variations, in Section S2 of the Supplementary Information.

To assess the introduced simulator bias, we evaluate both simulators on 400 randomly drawn sets of method parameters of a shifting gradient program with 12 adjustable parameters: $\mathbf{x} = [\varphi_1, \varphi_2, t_1, t_2, \varphi_{\text{init},1}, \varphi_{\text{init},2}, \varphi_{\text{init},3}, \varphi_{\text{final},0}, \varphi_{\text{final},1}, \varphi_{\text{final},2}, t_{\text{shift},2}, t_{\text{shift},3}]$ describing a two-step ^{1}D gradient (as in Fig. 1A) and a two-step ^{2}D shifting gradient (as in Fig. 1B, but with one additional shifting step). A further description of the parameters is given in Section 3.1. The resulting data are shown in Fig. 2, with the ground truth results shown on the x-axis and the biased results on the y-axis. It can be seen the introduction of noise and removal of compounds led to consistent overconfidence around low ground truth scores, potentially because the biased simulator observes more separated peaks that are not separated in the real system due to the incorporation of more compounds. At high ground truth scores, the biased simulator is generally underconfident because it is missing compounds, leading to a lower number of separated compounds and a lower score. While there still is a clear correlation between the two simulators, it is clear that an optimum in the biased simulator does not necessarily coincide with an optimum in the ground truth simulator. Therefore it provides an interesting test case for our multi-task Bayesian optimization framework, but also from a chromatographic method development standpoint.

4.1.2. Bayesian optimization & multi-task Bayesian optimization

BO (described in detail in Section 3.5) relies on two main components: a probabilistic model, and an acquisition function. The probabilistic model is trained on the input parameters and the observed metric of previously performed experiments and then acts as a surrogate model of the real system. We use a GP (described in detail in Section 3.3) as the probabilistic model due to its proven performance in the small data regime [48]. In MTBO, the GP model is replaced by a multi-task GP (described in detail in Section 3.4). A multi-task GP can predict outcomes for several tasks and can learn the correlations between different tasks to boost predictive capabilities. In our setting, the tasks are predicting the performance metric (CRF) of the real experiment and that of the retention model described in Section 3.2.

To compare the performance of a single-task GP with a multi-task GP, we perform leave-one-out cross-validation on predicting the outcome of 24 observations of the ground-truth simulator described in Section 4.1.1 using the same 12 adjustable parameters. This involves fitting the model to 23 observations and evaluating the prediction accuracy for the left-out observation for all possible permutations. The results for the single-task GP are shown in Fig. 3A, where it is shown that given the relatively low number of observations and the relatively high number of adjustable parameters, the GP cannot capture the response surface accurately. The multi-task GP was also fitted to 120 observations of the biased simulator and then used to do leave-one-out cross-validation on the 24 observations from the ground-truth simulator. By incorporating the biased observations, the predictive accuracy of the multi-task GP (see Fig. 3B) improves dramatically upon the single-task GP. The accuracy of the probabilistic model is imperative in Bayesian optimization, and the single-task GP will likely need more than 24 observations to become accurate.

The second core component of BO, the acquisition function, queries the GP (or the MTGP in case of MTBO) at new test locations and makes a trade-off between exploration (high variance predictions of the GP) or exploitation (high mean predictions of the GP) and finds the experiment that is most likely to improve upon the previously performed experiments. We describe details regarding the acquisition function in Section 3.5.

Given the two core components the conventional BO loop proceeds by an initial fitting of the GP model to initial experiments, followed by optimizing the acquisition function to propose new method parameters, then performing the measurement and retraining the model on the updated dataset, and is repeated until some convergence criterion or optimization budget is met.

In our MTBO framework, we draw inspiration from Letham et al. [36] In this framework, after the initial fitting of the MTGP model, we optimized the acquisition function on the task describing the ground truth simulated response surface to find ten promising experiments. Next, we evaluated these method parameters on the biased simulator and updated the MTGP with these new observations. We then reevaluated the proposed experiment on the ground truth task of the updated model, selecting the best-performing candidate to use in a real experiment, and iteratively repeat. Thus, the ground-truth task remains the primary driver of the optimization process, while the biased simulator serves to reduce predictive uncertainty and aids in filtering out predictions with high variance and low mean that would not be a good use of resources.

Fig. 4 describes a single- and multi-task BO loop for the setup discussed in Section 4.1.1 on a 12-parameter optimization problem, using the CRF described in Section 3.2. Both algorithms were allowed to do 25 experiments on the ground truth simulator. The MTGP could perform four biased simulator experiments per iteration and was initialized with 30 random biased simulator experiments. Both algorithms were initialized with five random ground truth experiments. All experiments were repeated for eight trials. It can be seen that the MTBO found reasonable optima with far fewer iterations, ascribing its applicability to the problem setting of LC×LC. We provide additional experiments with alternative simulator biases, that include a different number of removed compounds, but also a case where we have added compounds to resemble bias introduced by the tracking of noise peaks, in Section S2 of the Supplementary Information. In addition, we provide another retention modeling example of a synthetic sample comprised of two distributions in Section S3 of the Supplementary Information and draw similar conclusions.

4.2. Case study: Separation of a pesticide mix

To test our framework in a real-world setting, we employed our single-task and multi-task BO schemes to develop a method for the separation of a pesticide mix implemented within the automated closed-loop workflow we developed in earlier work [32]. We considered the

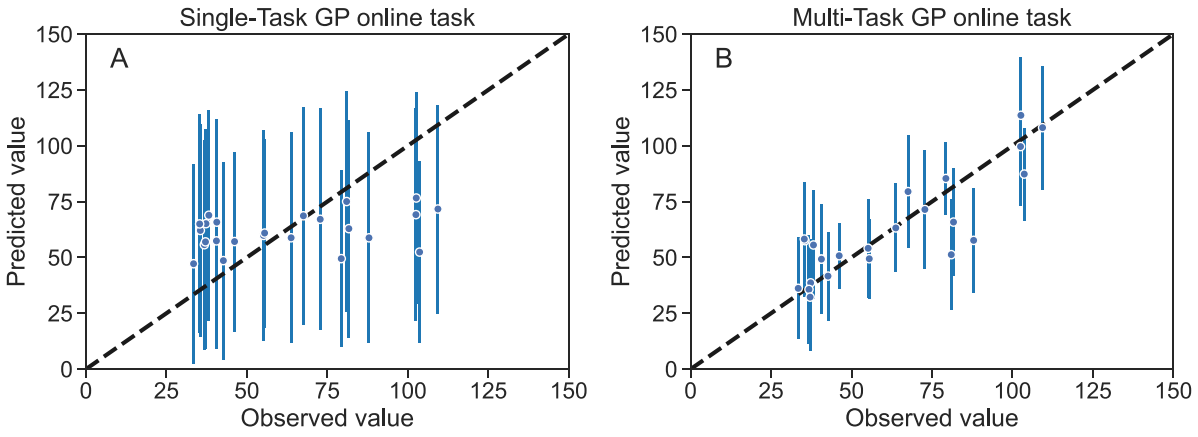

Fig. 3. Leave-one-out cross-validation predicting the outcome (i.e., the CRF value) of 24 observations in a 12-parameter shifting gradient program. **(A)** A single-task GP fitted only to 24 observations of the true simulator. **(B)** A multi-task GP fit to both the 24 observations of the true simulator and 120 biased simulator observations. Due to the high dimensionality of the optimization problem and a low number of observations, the single-task model does not describe the observations well enough to make useful predictions. Incorporating biased simulator observations increases the accuracy of the multi-task model significantly.

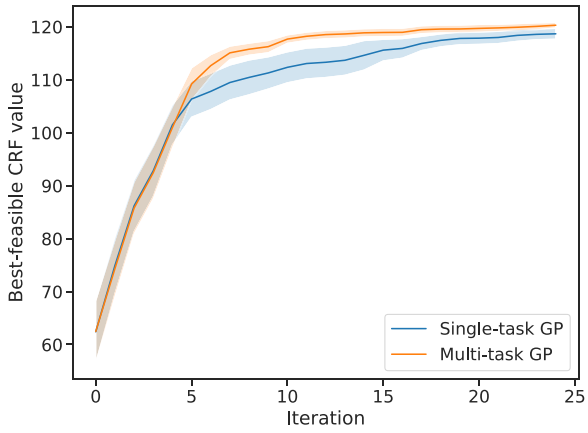

Fig. 4. Comparison of the performance of single-task and multi-task Bayesian optimization on the simulated optimization problem of a 12-parameter shifting gradient program. The mean and standard deviation are reported over 25 trials. It is observed that MTBO results find better optima at earlier iterations than single-task BO.

development of a two-step ^{1}D gradient program (as in Fig. 1A) with a fixed starting point at ($t_0 = 2, \varphi_0 = 0.02$) and a fixed end point at ($t_2 = 62, \varphi_2 = 1$), effectively leading to one adjustable gradient node (t_1, φ_1) in the first dimension. In the second dimension, we considered a shifting gradient program (as in Fig. 1B). This led to the following seven adjustable parameters: $\mathbf{x} = [\varphi_1, t_1, \varphi_{\text{init},0}, \varphi_{\text{init},1}, \varphi_{\text{final},0}, \varphi_{\text{final},1}, t_{\text{shift},1}]$. Other details regarding bounds and inequality constraints are discussed in Section 3.1. We used the CRF described in Section 3.2 to guide optimization, focusing on having as many separated peak pairs as possible within the given separation time. For both optimization cases, we allowed for a budget of 20 laboratory experiments including the four initial scanning gradients, which we refer to as method development iterations (MDIs).

4.2.1. *Multi-task Bayesian optimization*

Besides the four initial scanning experiments, the MTBO algorithm was initialized with 20 random sets of method parameters evaluated using the retention model (RM). At each iteration, the MTBO algorithm proposed 10 candidates which were first tested using the RM, after which the MTGP was updated and then the best-performing candidate in the updated model was carried out in a real laboratory experiment (see Section 3.4 for details regarding the MTBO loop). Fig. 5A shows the observed CRF values for each performed measurement (in blue) and the CRF values observed in the retention model (RM) (in orange). Here it is observed that method parameters are proposed that improve upon the initial scanning experiments both in terms of the CRF evaluated using the RM prediction, but also for the laboratory experiment, indicating successful optimization. Fig. 5B, shows all CRF values of the method parameters tested with the retention model in blue, with the orange dots indicating what measurements were then picked to be tested in a laboratory experiment. Here it is seen that method parameters were explored that would give low CRF values with the RM model, which were in turn not picked for use in laboratory experiments, because the additional data allowed for a reduction of uncertainty in the MTGP, thereby focusing on interesting areas and preventing poor uses of precious resources. The dashed line in Fig. 5A shows the optimum in the retention model which was computed *post-hoc* using BO and the retention model with a more exhaustive budget and which was cross-referenced with the genetic algorithm optimization approach proposed earlier [32]. Although the RM optimum is not met, within only 214 evaluations of the RM, the highest queried value is close to the RM optimum and was also selected at MDI 17, and in the process, other promising method parameters were evaluated, which would not have been attempted using RM to drive optimization.

Interestingly, although MDI 17 has the highest RM score, it is not the highest observed score from the real measurements. This is shown in Fig. 5A where it can be seen that the observed scores based on the RM and on the real measurement are generally in the same range but do differ, indicating some bias. Likewise, MDI 7 has the highest score on the real measurement, but not on the RM score. In terms of method parameters (shown in Fig. 5C), MDI 7 and MDI 17, both are improvements upon the longest scanning experiment (MDI 4). Judging from the chromatogram of MDI 4, shown in Fig. 6A, the method would benefit from a slightly shallower ^{1}D gradient and a narrower shift, starting at higher modifier concentrations, as most compounds eluted at relatively high modifier concentrations in the second dimension. This is indeed what is proposed in MDI 7 and MDI 17, where in both cases the ^{1}D program is more shallow than in the scan, with MDI 7 being the most shallow. In the case of MDI 7, shown in Fig. 6B, this led to more separation in the first dimension between 20 and 50 min than in MDI 17 (Fig. 6C), but arguably led to more compression around 60–70 min. In the second dimension, MDI 7 set a slightly reduced lower bound than MDI 17 and started the upper bound at a lower modifier fraction (See Fig. 5C), arguably leading to more separation in the second dimension. Note that there is the fixed isocratic hold of 5.4 s at 2% modifier in addition to the dwell time (2.1 s) associated with the pump, column dead time (2.07 s), and gradient delay time associated with the loop (1.2 s), which is why compounds can only elute under isocratic

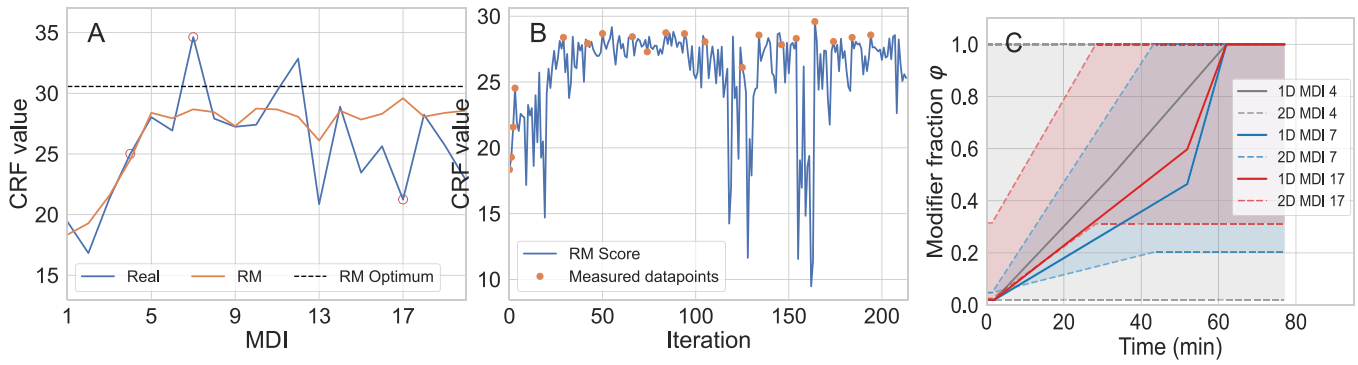

Fig. 5. Panel showing the results of the MTBO run. (A) Overview of observed CRF values for each MDI. The CRF computed on the real measurement is shown in blue, whereas the CRF computed using the retention model (RM) prediction is shown in orange. The black dashed line corresponds to the best optimum found using the RM model. (B) CRF scores for all sets of method parameters evaluated using the RM during the MTBO run, the orange points indicate which iterations were carried out using the real system. (C) Gradient programs used in MDI 4, 7, and 17, which are the encircled points in A, and correspond to the longest scanning experiment, the optimum according to the CRF computed using the real measurement, and the optimum according to the CRF computed on the retention model (RM) prediction, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

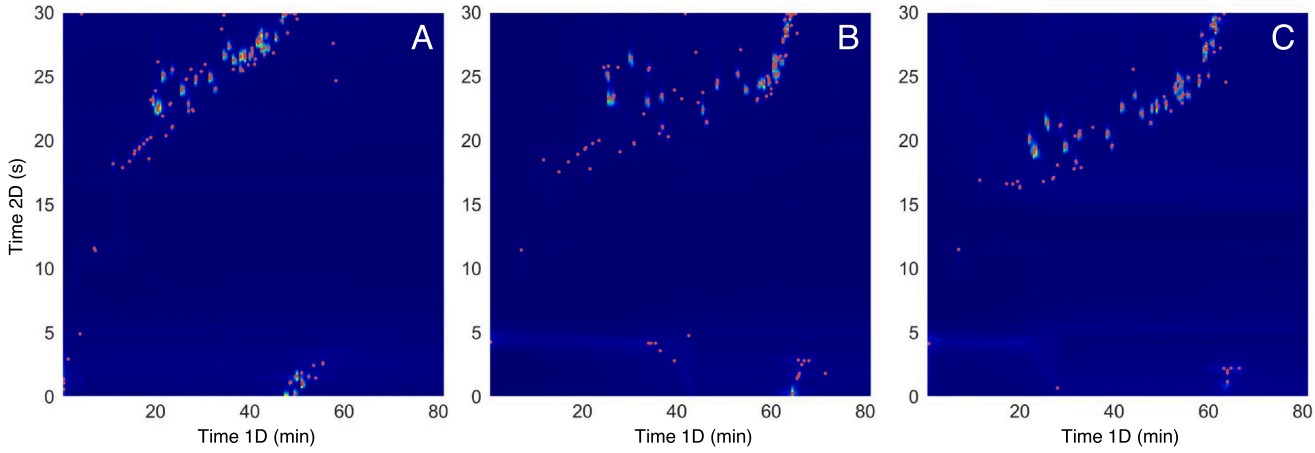

Fig. 6. LC×LC-MS total ion chromatograms (TICs) of MDIs during the MTBO run. Red dots indicate the peaks detected by the peak detection algorithm. **(A)** Chromatogram of MDI 4, which is the longest initial scanning experiment. **(B)** Chromatogram of MDI 7, which was the best measurement according to the CRF computed using the real measurement. **(B)** chromatogram of MDI 17, which was the best measurement according to the CRF computed using the RM prediction.

conditions before ~10.7 s. Compounds in a modulation can elute during the combined dwell and dead time (5.28 s) of the next modulation. It is difficult to make a quantitative determination as to why the CRF scores determined using the real measurement differ, although it could be caused by improved separation (for instance around 20 min in the first dimension and 20 s in the second dimension). It could also be due to detected noise (around 40 min in the first dimension and 5 s in the second dimension) in MDI 7. In Section S4, we compare the RM prediction of MDI 17 with the true measurement of MDI 17 and describe the differences and similarities there.

4.2.2. Single-task Bayesian optimization

Fig. 7A describes the optimization campaign of the single-task BO loop. Unfortunately, these measurements were performed at a different point in time and MS detector sensitivity was lower, indicated by the CRF scores of MTBO (dashed green line), and the CRF scores of BO (blue line) of MDI 1–4, which are the same initial scanning measurements and have consistently lower scores. This is also supported by the RM scores (orange lines), which were computed *post hoc* at the measured method parameters, using the retention model developed in Section 4.2.1, which are consistently higher than the CRF scores of the real measurements. This unfortunately does not allow for a quantitative comparison of the MTBO and BO runs. However, despite the lowered sensitivity, the BO algorithm is seen to provide method parameters that improve upon the scanning experiments. Here MDI 11 is the best measurement according to the CRF evaluated using the real measurement, and MDI 6 is best according to the RM, which are both shown in Fig. 8. Note that the BO algorithm did not use the RM to guide optimization, and this is solely used for discussion.

The method parameters for MDI 6 and MDI 11 are shown in Fig. 7B, where it is seen that both MDIs, similar to those from the MTBO, have a shallower ^{1}D gradient that increases the separation in the first dimension. However, the ^{2}D parameters are not varied as much as in the MTBO run. This is especially true for MDI 6, where the ^{2}D parameters are the same as in the scanning experiments. Note, however, that this leads to a score that is close to the optimum in the RM, indicating that much improvement can be gained from tuning the ^{1}D parameters, according to the retention model. This is also supported by Section S4, where it is seen that ^{2}D peak widths are generally overpredicted, and therefore these peaks are hard to effectively separate in the second dimensions according to the RM. MDI 11, slightly lowers the upper bound of the shifting gradient but leaves the lower bound of the shifting gradient untouched compared to the scanning gradients. While this leads to less separation than was proposed in the MTBO run, it does lead to a better separation of the cluster of peaks eluting around 60 min, which was not observed in the MTBO run. However, from a qualitative perspective, the MTBO obtained more reasonable method parameters than what was found in the BO run, and BO will likely need more MDIs to find better optima, which was also demonstrated more broadly in the *in silico* experiments in Section 4.1.2 and in sections S2 and S3 of the Supplementary Information.

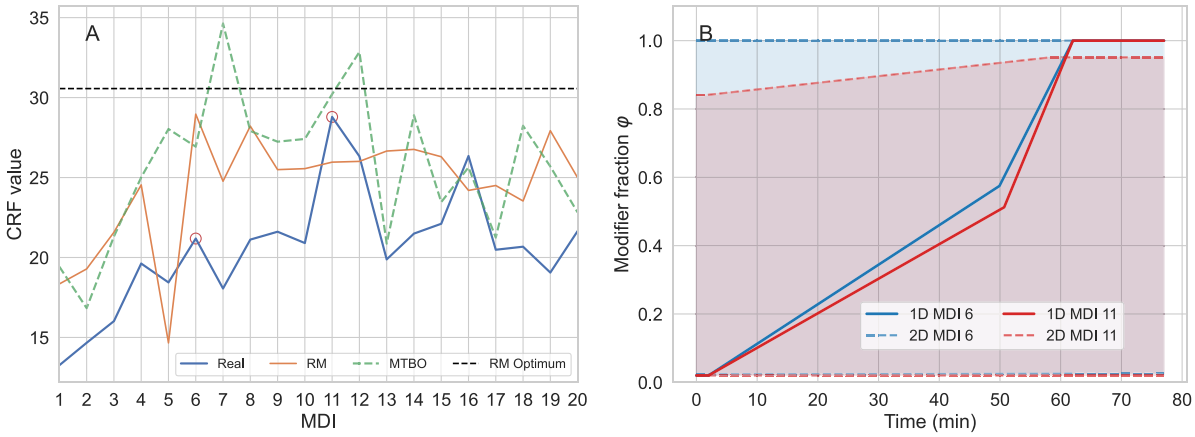

Fig. 7. Panel showing the results of the BO run. **(A)** Overview of observed CRF values as a function of MDIs. The CRF computed using the real measurement is shown in blue, whereas the CRF computed using the RM prediction is shown in orange; note that this is not used during the optimization. The black dashed line corresponds to the best optimum found using the RM model. The dashed green line shows the MTBO values shown in Figure 5 for comparison. **(B)** Gradient programs used in MDI 6 and 11, which are the encircled measurements in A, and correspond to the optimum according to the CRF computed on the RM prediction, and the optimum according to the CRF computed on the real measurement, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

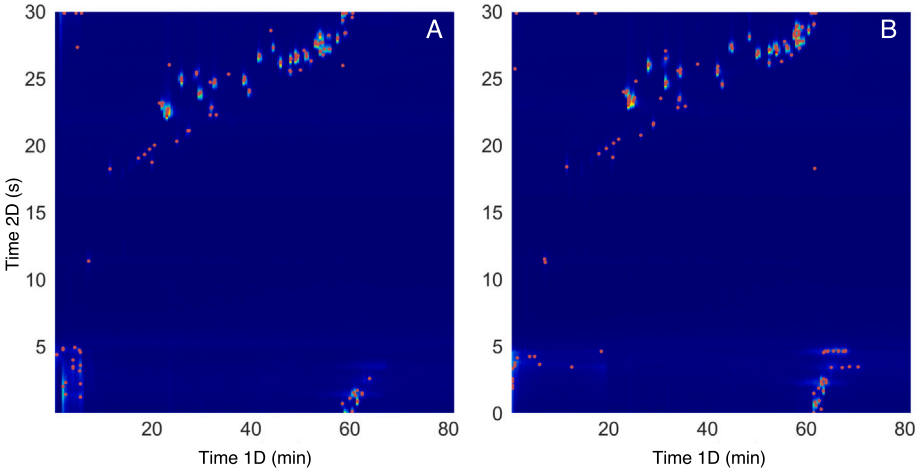

Fig. 8. LC×LC-MS TICs of MDIs during the BO run. Red dots indicate the peaks detected by the peak detection algorithm. **(A)** Chromatogram of MDI 6, which was the best measurement according to the CRF computed using the RM model. **(B)** Chromatogram of MDI 11, which was the best measurement according to the CRF computed using the real measurement.

5. Conclusion

We developed and investigated the use of multi-task Bayesian optimization (MTBO) applied to the development of methods for comprehensive two-dimensional liquid chromatography (LC×LC) separations. We first compared the performance of MTBO with conventional Bayesian optimization (BO) in an *in silico* test case and showed that MTBO, with its ability to incorporate information from both retention modeling and real experimental data, finds better optima in fewer iterations than conventional BO. This was also demonstrated in a real test case where we compared the performance of MTBO and BO for optimizing a method developed for separation of a complex pesticide sample. Here it was shown that although both methods improved upon the performance of initial scanning experiments, from a chromatographers perspective, MTBO made better use of both the first- and second-dimension parameters than BO. This makes MTBO a promising method for method development over conventional BO when retention modeling is challenging, and the number of adjustable parameters and/or limited optimization budget makes conventional BO impractical. We note that this is not necessarily only the case in LC×LC, but MTBO could also be used in challenging applications in LC, GC, GC×GC, and other separation techniques where the true objective to be optimized is costly, and there is access to a cheaper, less accurate source of data.

It remains an open question as to when and where MTBO should be used over retention modeling. When retention modeling can accurately predict retention times and peak widths for most compounds in the sample under study, it is arguably more sensible to directly use the retention model to guide optimization. However, one only finds out if this is the case during the optimization process itself. If the retention model is too biased to effectively guide optimization, from that point on the MTBO could be initialized with the previously acquired data and could take over optimization from there. In addition, there may be use cases where only some of the adjustable parameters are described by a retention model or other sources of information, and others are not; in these cases, MTBO could still use this information and provide a "warm start" to the optimization, rather than starting from scratch with conventional BO. This could be a promising area for MTBO.

Since the MTBO framework required an objective that was both computable and applicable to both real measurements and predictions made using a retention model, we utilized a resolution-based CRF that focused on maximizing the sum of the resolution between closest neighbor pairs. This is a relatively uncommon scenario, and we did not investigate whether other CRFs could potentially work better; this could be investigated in future work. Another vital component

for successfully building an accurate retention model, but also for robustly assessing the CRF for a real measurement is the signal processing pipeline, from background correction to peak detection and peak tracking. The performance of the algorithm is likely to improve with improvements in this pipeline. For instance, background correction artifacts could introduce anomalies in the retention model, but could also lead to over/underconfident CRF scores for the real measurement, impacting the algorithm's performance. This is equally true for the peak detection and tracking. Future work could focus on improving and/or benchmarking these methods, but could also focus on CRFs that are more robust to these anomalies.

Finally, in the MTBO framework, some design decisions had to be made. For instance, the number of times the retention model can be queried before a real measurement is made was set to 4–10 times in this study. Since a retention model prediction (order of seconds) and real measurements (order of hours) occur on very different time scales, this number could potentially be much higher, which could provide benefits but could also introduce higher MTGP fitting costs or instabilities; this could also be further optimized in future work. Another interesting future research direction could focus on using data from other (similar) separations to initialize the MTGP which might further boost its performance and reduce the overall number of experimental iterations required, thereby enhancing the efficiency of the optimization process.

CRediT authorship contribution statement

Jim Boelrijk: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Stef R.A. Molenaar:** Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Tijmen S. Bos:** Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Tina A. Dahlseid:** Writing – review & editing, Validation, Methodology, Investigation. **Bernd Ensing:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Dwight R. Stoll:** Writing – review & editing, Validation, Resources, Project administration, Methodology, Funding acquisition. **Patrick Forré:** Writing – review & editing, Visualization, Supervision, Project administration, Methodology, Funding acquisition. **Bob W.J. Pirok:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

Saer Samanipour is kindly acknowledged for providing the pesticide mixture.

SM and TB acknowledge the UNMATCHED project, which is supported by BASF, Envalior, and Nouryon, and receives funding from the Dutch Research Council (NWO), The Netherlands in the framework of the Innovation Fund for Chemistry (CHIPP Project 731.017.303) and from the Ministry of Economic Affairs, The Netherlands in the framework of the "PPS-toeslagregeling". TB acknowledges the PARADISE project, The Netherlands (ENPPS.TA.019.001) and received funding from the Dutch Research Council (NWO), The Netherlands in the framework of the Science PPP Fund for the top sectors and from the Ministry of Economic Affairs of the Netherlands in the framework of the "PPS Toeslagregeling". BP acknowledges the TTW VENI research program, The Netherlands (Project 19173, "Unleashing the Potential of Separation Technology to Achieve Innovation in Research and Society (UPSTAIRS)"), which is financed by the Dutch Research Council (NWO), The Netherlands. All of the 2D-LC instrumentation used in this work was provided to DS through the Agilent Technologies Thought Leader Award program. DS and TD were supported by a grant from the United States National Science Foundation (CHE-2003734).

This work was performed in the context of the Chemometrics and Advanced Separations Team (CAST) within the Centre for Analytical Sciences Amsterdam (CASA). The valuable contributions of the CAST members are gratefully acknowledged.

Supplementary materials

Additional data including code used in Section 4.1 and excluded masses can be found at: https://github.com/Jimbo994/AutoLC-MTBO. All other data is available upon request.

References

[1] M.M. Bushey, J.W. Jorgenson, Automated instrumentation for comprehensive two-dimensional high-performance liquid chromatography of proteins, Anal. Chem. 62 (2) (1990) 161–167.

[2] F. Erni, R.W. Frei, Two-dimensional column liquid chromatographic technique for resolution of complex mixtures, J. Chromatogr. A 149 (C) (1978) 561–569.

[3] E. Uliyanchenko, P.J.C.H. Cools, S. van der Wal, P.J. Schoenmakers, Comprehensive two-dimensional ultrahigh-pressure liquid chromatography for separations of polymers, Anal. Chem. 84 (18) (2012) 7802–7809.

[4] X. Wang, D.R. Stoll, A.P. Schellinger, P.W. Carr, Peak capacity optimization of peptide separations in reversed-phase gradient elution chromatography: Fixed column format, Anal. Chem. 78 (10) (2006) 3406–3416.

[5] J. Beens, J. Blomberg, P.J. Schoenmakers, Proper tuning of comprehensive two-dimensional gas chromatography (GC×GC) to optimize the separation of complex oil fractions, J. High Resolut. Chromatogr. 23 (3) (2000) 182–188.

[6] M. Sorensen, D.C. Harmes, D.R. Stoll, G.O. Staples, S. Fekete, D. Guillarme, A. Beck, Comparison of originator and biosimilar therapeutic monoclonal antibodies using comprehensive two-dimensional liquid chromatography coupled with time-of-flight mass spectrometry, mAbs 8 (7) (2016) 1224–1234.

[7] F. Cacciola, D. Giuffrida, M. Utczas, D. Mangraviti, P. Dugo, D. Menchaca, E. Murillo, L. Mondello, Application of comprehensive two-dimensional liquid chromatography for carotenoid analysis in red mamey (*Pouteria sapote*) fruit, Food Anal. Methods 9 (8) (2016) 2335–2341.

[8] L. Montero, E. Ibáñez, M. Russo, R. di Sanzo, L. Rastrelli, A.L. Piccinelli, R. Celano, A. Cifuentes, M. Herrero, Metabolite profiling of licorice (*Glycyrrhiza glabra*) from different locations using comprehensive two-dimensional liquid chromatography coupled to diode array and tandem mass spectrometry detection, Anal. Chim. Acta 913 (2016) 145–159.

[9] M. Muller, A.G. Tredoux, A. de Villiers, Predictive kinetic optimisation of hydrophilic interaction chromatography×reversed phase liquid chromatography separations: Experimental verification and application to phenolic analysis, J. Chromatogr. A 1571 (2018) 107–120.

[10] D.R. Stoll, X. Li, X. Wang, P.W. Carr, S.E. Porter, S.C. Rutan, Fast, comprehensive two-dimensional liquid chromatography, J. Chromatogr. A 1168 (1–2) (2007) 3–43.

[11] J.C. Berridge, Unattended optimisation of reversed-phase high-performance liquid chromatographic separations using the modified simplex algorithm, J. Chromatogr. A 244 (1) (1982) 1–14.

[12] J.C. Berridge, Simplex optimization of high-performance liquid chromatographic separations, J. Chromatogr. A 485 (C) (1989) 3–14.

[13] S. O'Hagan, W.B. Dunn, M. Brown, J.D. Knowles, D.B. Kell, Closed-loop, multiobjective optimization of analytical instrumentation: Gas chromatography/time-of-flight mass spectrometry of the metabolomes of human serum and of yeast fermentations, Anal. Chem. 77 (1) (2005) 290–303.

[14] J. Bradbury, G. Genta-Jouve, J.W. Allwood, W.B. Dunn, R. Goodacre, J.D. Knowles, S. He, M.R. Viant, MUSCLE: automated multi-objective evolutionary optimization of targeted LC-MS/MS analysis, Bioinformatics (Oxford, England) 31 (6) (2014) 975–977.

[15] A. Kensert, P. Libin, G. Desmet, D. Cabooter, Deep reinforcement learning for the direct optimization of gradient separations in liquid chromatography, J. Chromatogr. A 1720 (2024).

[16] J. Boelrijk, B. Pirok, B. Ensing, P. Forré, Bayesian optimization of comprehensive two-dimensional liquid chromatography separations, J. Chromatogr. A 1659 (2021) 462628.
[17] J. Boelrijk, B. Ensing, P. Forré, B.W. Pirok, Closed-loop automatic gradient design for liquid chromatography using Bayesian optimization, Anal. Chim. Acta 1242 (2023).
[18] M.J. den Uijl, P.J. Schoenmakers, B.W. Pirok, M.R. van Bommel, Recent applications of retention modelling in liquid chromatography, J. Sep. Sci. 44 (1) (2021) 88–114.
[19] J. Navarro-Huerta, A. Gisbert-Alonso, J. Torres-Lapasió, M. García-Alvarez-Coque, Testing experimental designs in liquid chromatography (I): Development and validation of a method for the comprehensive inspection of experimental designs, J. Chromatogr. A 1624 (2020) 461180.
[20] G. Vivo-Truyols, J. Torres-Lapasio, M. Garcıia-Alvarez-Coque, Error analysis and performance of different retention models in the transference of data from/to isocratic/gradient elution, J. Chromatogr. A 1018 (2) (2003) 169–181.
[21] E. Marengo, V. Gianotti, S. Angioi, M.C. Gennaro, Optimization by experimental design and artificial neural networks of the ion-interaction reversed-phase liquid chromatographic separation of twenty cosmetic preservatives, J. Chromatogr. A 1029 (1–2) (2004) 57–65.
[22] S.R. Molenaar, T.A. Dahlseid, G.M. Leme, D.R. Stoll, P.J. Schoenmakers, B.W. Pirok, Peak-tracking algorithm for use in comprehensive two-dimensional liquid chromatography – Application to monoclonal-antibody peptides, J. Chromatogr. A 1639 (2021) 461922.
[23] P.J. Schoenmakers, H.A. Billiet, R. Tussen, L. De Galan, Gradient selection in reversed-phase liquid chromatography, J. Chromatogr. A 149 (C) (1978) 519–537.
[24] P. Jandera, M. Holčapek, L. Kolářová, Retention mechanism, isocratic and gradient-elution separation and characterization of (co)polymers in normal-phase and reversed-phase high-performance liquid chromatography, J. Chromatogr. A 869 (1–2) (2000) 65–84.
[25] C.M. Roth, K.K. Unger, A.M. Lenhoff, Mechanistic model of retention in protein ion-exchange chromatography, J. Chromatogr. A 726 (1–2) (1996) 45–56.
[26] B.W. Pirok, S.R. Molenaar, R.E. van Outersterp, P.J. Schoenmakers, Applicability of retention modelling in hydrophilic-interaction liquid chromatography for algorithmic optimization programs with gradient-scanning techniques, J. Chromatogr. A 1530 (2017) 104–111.
[27] B.W. Pirok, S. Pous-Torres, C. Ortiz-Bolsico, G. Vivó-Truyols, P.J. Schoenmakers, Program for the interpretive optimization of two-dimensional resolution, J. Chromatogr. A 1450 (2016) 29–37.
[28] W. Hao, B. Li, Y. Deng, Q. Chen, L. Liu, Q. Shen, Computer aided optimization of multilinear gradient elution in liquid chromatography, J. Chromatogr. A 1635 (2021) 461754.
[29] B. Huygens, K. Efthymiadis, A. Nowé, G. Desmet, Application of evolutionary algorithms to optimise one- and two-dimensional gradient chromatographic separations, J. Chromatogr. A 1628 (2020) 461435.
[30] T.S. Bos, J. Boelrijk, S.R. Molenaar, B. Van 'T Veer, L.E. Niezen, D. Van Herwerden, S. Samanipour, D.R. Stoll, P. Forré, B. Ensing, G.W. Somsen, B.W. Pirok, Chemometric strategies for fully automated interpretive method development in liquid chromatography, Anal. Chem. 94 (46) (2022) 16060–16068.
[31] P. Nikitas, A. Pappa-Louisi, P. Agrafiotou, Multilinear gradient elution optimisation in reversed-phase liquid chromatography using genetic algorithms, J. Chromatogr. A 1120 (1–2) (2006) 299–307.
[32] S.R. Molenaar, T.S. Bos, J. Boelrijk, T.A. Dahlseid, D.R. Stoll, B.W. Pirok, Computer-driven optimization of complex gradients in comprehensive two-dimensional liquid chromatography, J. Chromatogr. A 1707 (2023) 464306.
[33] T.S. Bos, L.E. Niezen, M.J. den Uijl, S.R.A. Molenaar, S. Lege, P.J. Schoenmakers, G.W. Somsen, B.W.J. Pirok, Reducing the influence of geometry-induced gradient deformation in liquid chromatographic retention modelling, J. Chromatogr. A 1635 (2021) 461714.
[34] L.E. Niezen, T.S. Bos, P.J. Schoenmakers, G.W. Somsen, B.W. Pirok, Capacitively coupled contactless conductivity detection to account for system-induced gradient deformation in liquid chromatography, Anal. Chim. Acta 1271 (2023).
[35] R. Garnett, Bayesian Optimization, Cambridge University Press, 2022.
[36] B. Letham, E. Bakshy, Bayesian optimization for policy search via online-offline experimentation, J. Mach. Learn. Res. 20 (0000) (2019).
[37] J. Wu, S. Toscano-Palmerin, P.I. Frazier, A.G. Wilson, Practical multi-fidelity Bayesian optimization for hyperparameter tuning, in: 35th Conference on Uncertainty in Artificial Intelligence, UAI 2019, 2019.
[38] C.J. Taylor, K.C. Felton, D. Wigh, M.I. Jeraal, R. Grainger, G. Chessari, C.N. Johnson, A.A. Lapkin, Accelerated chemical reaction optimization using multi-task learning, ACS Central Sci. 9 (5) (2023) 957–968.
[39] D.R. Stoll, H.R. Lhotka, D.C. Harmes, B. Madigan, J.J. Hsiao, G.O. Staples, High resolution two-dimensional liquid chromatography coupled with mass spectrometry for robust and sensitive characterization of therapeutic antibodies at the peptide level, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 1134–1135 (2019).
[40] D.R. Stoll, K. Shoykhet, P. Petersson, S. Buckenmaier, Active solvent modulation: A valve-based approach to improve separation compatibility in two-dimensional liquid chromatography, Anal. Chem. 89 (17) (2017) 9260–9267.
[41] W. Hao, K. Wang, B. Yue, Q. Chen, Y. Huang, J. Yu, D. Li, Peak compression in linear gradient elution liquid chromatography, J. Chromatogr. A 1619 (2020) 460908.
[42] M. Balandat, B. Karrer, D.R. Jiang, S. Daulton, B. Letham, A.G. Wilson, E. Bakshy, BoTorch: A framework for efficient Monte-Carlo Bayesian optimization, Adv. Neural Inf. Process. Syst. 33 (2020).
[43] S.R. Molenaar, J.H. Mommers, D.R. Stoll, S. Ngxangxa, A.J. de Villiers, P.J. Schoenmakers, B.W. Pirok, Algorithm for tracking peaks amongst numerous datasets in comprehensive two-dimensional chromatography to enhance data analysis and interpretation, J. Chromatogr. A 1705 (2023) 464223.
[44] M.C. Chambers, B. Maclean, R. Burke, D. Amodei, D.L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egertson, K. Hoff, D. Kessner, N. Tasman, N. Shulman, B. Frewen, T.A. Baker, M.-Y. Brusniak, C. Paulse, D. Creasy, L. Flashner, K. Kani, C. Moulding, S.L. Seymour, L.M. Nuwaysir, B. Lefebvre, F. Kuhlmann, J. Roark, P. Rainer, S. Detlev, T. Hemenway, A. Huhmer, J. Langridge, B. Connolly, T. Chadick, K. Holly, J. Eckels, E.W. Deutsch, R.L. Moritz, J.E. Katz, D.B. Agus, M. MacCoss, D.L. Tabb, P. Mallick, A cross-platform toolkit for mass spectrometry and proteomics, Nature Biotechnol. 30 (10) (2012) 918–920.
[45] D.R. Stoll, P.W. Carr, Multi-dimensional liquid chromatography: Principles, practice, and applications, in: D.R. Stoll, P.W. Carr (Eds.), Multi-Dimensional Liquid Chromatography: Principles, Practice, and Applications, first ed., CRC Press, Boca Raton, 2022, pp. 1–401.
[46] S. Chapel, F. Rouvière, S. Heinisch, Sense and nonsense of shifting gradients in on-line comprehensive reversed-phase LC × reversed-phase LC, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 1212 (2022) 123512.
[47] J.T.V. Matos, R.M.B.O. Duarte, A.C. Duarte, Chromatographic response functions in 1D and 2D chromatography as tools for assessing chemical complexity, Trends Anal. Chem. 45 (2013) 14–23.
[48] J. Snoek, H. Larochelle, R.P. Adams, Practical Bayesian optimization of machine learning algorithms, Adv. Neural Inf. Process. Syst. 4 (2012) 2951–2959.
[49] C.E. Rasmussen, Gaussian Processes in Machine Learning, vol. 3176, Springer Verlag, 2004, pp. 63–71.
[50] H.V. Henderson, F. Pukelsheim, S.R. Searle, On the history of the kronecker product, Linear Multilinear Algebra 14 (2) (1983) 113–120.
[51] B. Letham, B. Karrer, G. Ottoni, E. Bakshy, Constrained Bayesian optimization with noisy experiments, 2018.