Journal of Chromatography A 1741 (2025) 465593

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Method development of comprehensive two-dimensional liquid chromatography: A new metric for preliminary column selection

Monica Tirapelle, Fanyi Duanmu, Dian Ning Chia, Maximilian O. Besenhard, Luca Mazzei, Eva Sorensen *

Department of Chemical Engineering, University College London, Torrington Place, London, WC1E 7JE, UK

ARTICLE INFO

Keywords:
Two-dimensional liquid chromatography
In-silico method development
Column orthogonality
Multi-objective optimization
Reversed-phase liquid chromatography

ABSTRACT

To elevate the separation performance, two-dimensional liquid chromatography (2D-LC) uses two chromatographic columns with different stationary phases to diversify solute interactions with the resin, hence providing a second "dimension" to solute-specific separation. Developing methods for 2D-LC starts therefore with preliminary column selection. Selecting columns that yield (metaphorically) orthogonal dimensions is of utmost importance, but remains challenging. Although several metrics exist to quantify column orthogonality, currently there is no established methodology, and none of the existing methods accounts for the non-homogeneity of peak band broadening across each separation dimension.

In this work, we propose a new approach to select columns a priori. This approach is based on critical resolution distribution statistics and implicitly accounts for local peak crowding and peak band broadening. Furthermore, we assess the importance of preliminary column selection during *in-silico* method development and multi-objective optimization of comprehensive 2D-LC. The comparison of the multi-objective Pareto fronts revealed that column pairs selected with our approach provide better separation quality and reduce analysis time compared to column selections via the most established metrics in the literature. Our results prove the importance of preliminary column selection for method development and optimization of 2D-LC systems, and they also show that choosing the right orthogonality metric (such as that proposed here) is crucial.

1. Introduction

Two-dimensional liquid chromatography (2D-LC) represents a valuable extension to single-column separation to resolve complex mixtures within a short time. By *complex mixtures* we refer to both heterogeneous mixtures containing hundreds of compounds and mixtures containing one or more pairs or groups of compounds that are chemically homogeneous [1].

2D-LC involves two consecutive chromatographic separation stages that can be performed offline (manual intervention required) or online (automated). The four separation modes of online 2D-LC are the single heart-cutting [2], the multiple heart-cutting [3], the full comprehensive [4] and the selective comprehensive [5] modes. The single heart-cutting (LC-LC) and multiple heart-cutting (mLC-LC) modes involve transferring a *single fraction* or *several fractions* of the ^{1}D column effluent to the ^{2}D column for additional separation, respectively [6,7]. Unlike the heart-cutting modes, the full comprehensive (LCxLC) and selective comprehensive (sLCxLC) modes are holistic approaches that consist in sampling either *the whole* or *a limited portion* of the ^{1}D effluent in narrow volumes that are sent in a regular fashion to the ^{2}D separation [7,8]. While heart-cutting modes are extremely useful when only a few target solutes have to be separated, comprehensive modes are preferred when many components have to be resolved. Method development and implementation of comprehensive modes, however, are more challenging since all components must elute from the ^{2}D column within the modulation time of the interface valve [9].

Proper method development is crucial to fully exploit the potential of comprehensive multidimensional liquid chromatography [6,8,10]. Performing LCxLC method development from scratch involves two primary steps. As described below, the first step entails identifying two columns with dissimilar selectivity from a selection of available columns. The choice depends strongly on the nature of the sample under analysis. To prevent compatibility issues related to solvent mismatch, the two-column separation types must be compatible. Examples of possible combinations of separation modes with compatible separation mechanisms are ion-exchange chromatography with reversed-phase liquid chromatography (IEXxRPLC), hydrophilic interaction liquid chromatography with reversed-phase liquid chromatography (HILICxRPLC), and reversed-phase with reversed-phase

---

* Corresponding author.
E-mail address: e.sorensen@ucl.ac.uk (E. Sorensen).

https://doi.org/10.1016/j.chroma.2024.465593
Received 16 August 2024; Received in revised form 8 December 2024; Accepted 9 December 2024

liquid chromatography (RPLCxRPLC). The last combination is usually employed to separate large biomolecules, such as peptides and proteins, because of the high separation efficiency, the miscibility of the mobile phases, and the high compatibility with MS detectors [1,11]. The second step of the LCxLC method development involves determining optimal separation conditions (e.g., modulation time, flow rate, temperature and pH) and optimal column design (e.g., column length and diameter). The optimal conditions depend on the objective function considered, which is determined by the analyst [9].

Selecting the first and second dimension columns is the most crucial and challenging choice of LCxLC method development as this will determine the separation capacity of the system [8]. The selection of the two columns is often driven by the principle of retention orthogonality, where columns are deemed orthogonal if they are able to spread the individual component peaks across the separation space, which can be visualized as a 2D space with each dimension representing a separation [12]. Orthogonality does not depend only on the column separation mechanism, but also on sample mixture and separation conditions [13,14]. The higher the column orthogonality, the more the separation space is accessible [15].

In the literature, the orthogonality of separation is evaluated by considering different approaches. Commonly used are the geometric approach to factor analysis [13], information theory-based indices [16,17], multivariate data analysis techniques [18–20], visual interpretation of OPTICS color maps and weighted-average-linkage dendrograms [21–23], geometric criteria characterizing the coverage of the 2D separation space such as bin counting methods [24–29] and convex-hulls [30–32], statistical tests and correlation coefficients [31,33], nearest neighbor distances (NND) approaches [34,35] and the asterisk equations [36]. In the case of RPLCxRPLC systems, a common method for evaluation of separation orthogonality is the F_s factor of column similarity, which is based on the Snyder–Dolan hydrophobic-subtraction model (HSM) [37,38]. A summary of the different approaches, along with their relative metrics, is outlined in Table 1. For detailed overviews of the various approaches, we refer the reader to Guiochon et al. [39], Al Bakain et al. [31] and Agrawal et al. [40].

Each of the methods described so far has its drawbacks. Factor analysis and information theory-based indices consider peak spreading on the separation space (long-length effect) but do not account for peak clustering (short-length effects). Information theory and box-counting methods require discretization of the separation space into bins, i.e., a suitable choice of the number of bins is required. Correlation coefficients are unsuited because they were originally designed to determine correlation rather than data scattering. Multivariate data analysis techniques are not always accessible because they require visual interpretation of the results. The asterisk equations are not optimal for a limited number of components. The F_s factor, on the other hand, is designed to account only for differences in column resins while ignoring the mixture under analysis. Last but not least, all methods described so far do not consider the fact that peak broadening is not uniform across the separation space, even though peak broadening affects the resolution. To the best of our knowledge, the only attempts to account for peak band broadening were made by Davis et al. [41] and Nowik et al. [34]. However, their approach considers an *average* peak width for each dimension, rather than the *actual* width of each peak. Since peak width across each separation dimension may vary significantly, especially under isocratic elution conditions, improvements would be needed.

Given the multitude of methods available for evaluating the degree of orthogonality of 2D-LC chromatographic systems and their corresponding drawbacks, several comparative studies have been undertaken to identify the most suitable method [14,31,42–47]. These studies show that some of the metrics are mathematically related and that some may be preferred over others for specific sets of components and/or because of their utility and easy applicability. The large number of metrics and the corresponding comparison studies arise from the fact that orthogonality does not inherently ensure complementary columns, and the metrics are often biased toward particular scenarios, such as the presence of outliers or the influence of normalization. Although some orthogonality evaluation methods are considered well-established, these metrics are often less than ideal, and no single generally accepted metric currently exists [29,46]. Instead, method developers choose which metric to employ, and whether to employ multiple metrics in combination for a better understanding of the system performance, based only on their personal preferences.

Resolution or scoring approaches are likely to be more effective than the currently established orthogonality metrics. Lindsey et al. [48] developed a scoring function that yields column pairs that on average resolve 14%–50% more analytes than competing methods. Nevertheless (1) their scoring function accounts for peak width solely in the second dimension, and (2) they assume that an analyte does never elute into multiple samples taken from the first dimension, even if, in practice, a single component could span consecutive bins.

In this work, we propose a new resolution approach for preliminary column selection. Our approach is based on the definition of 2D resolution by Schure [49], and requires the knowledge of the retention behavior (i.e., retention factor or retention time) of the components under analysis in a set of chromatographic systems. This knowledge can be gained from experimental campaigns or predicted with appropriate retention models such as linear solvation energy relationships (LSER) [50] or quantitative structure retention relationships (QSRR) [51]. Our new approach for preliminary column selection accounts for the short-length scale effects and peak band broadening, and can be tailored for heart-cutting and comprehensive 2D-LC modes. In this work, we compare this approach with eight different metrics of orthogonality that are well-established in the literature (i.e., Pearson, Spearman and Kendall's correlation coefficient, geometric surface coverage, minimum convex hull, column similarity factor, arithmetic and harmonic mean of the nearest neighbor distances). Furthermore, we assess the impact of the preliminary column selection on the optimization results, when both method development and optimization of comprehensive 2D-LC are considered.

This work is structured as follows. Section 2 provides the theory behind the hydrophobic-subtraction model and outlines the most well-established metrics for evaluating column orthogonality. Section 3 introduces a new approach for preliminary column selection based on summary statistics (e.g., mean, median and 25th percentile) of the critical 2D resolutions. This approach is compared with other metrics in Section 4, where the influence of preliminary column selection on the optimal results is also assessed. Conclusions are then drawn in Section 5. Throughout the manuscript, we will adopt the following notation to refer to first- and second-dimension properties, respectively: ^{1}property and ^{2}property. On the other hand, 1D and 2D are used to indicate either one-dimensional or two-dimensional systems.

2. Theory and background

This section outlines the Snyder–Dolan hydrophobic-subtraction model (HSM) [37,38], i.e., the retention model employed in this work to characterize solute-column interactions in reversed-phase liquid chromatography (RPLC) columns, as well as some of the criteria previously proposed in the literature for column selection and evaluation of column orthogonality. These criteria will be used as benchmarks to evaluate the performance of our new approach.

2.1. The hydrophobic-subtraction model

In RPLC, the main interactions between solutes and columns are hydrophobic interactions, steric resistance, hydrogen bonding between basic solutes and acidic groups, hydrogen bonding between acidic solutes and basic groups, and cation-exchange [37,38]. These five

Table 1

Overview of different approaches proposed to investigate column orthogonality with relative metrics. (AAMRT = Auto-Associative Multivariate Regression Trees; GPCM = Generalized pairwise correlation method; HSM = Hydrophobic-Subtraction Model; NND = Nearest Neighbor Distances; PCA = Principal Component Analysis.)

Authors	Approach	Metrics
Liu et al. [13]	Geometric approach to factor analysis	Peak spreading angles Practical peak capacity
Slonecker et al. [16]	Information theory	Informational similarity
Detroyer et al. [18]	Chemometric	PCA Cluster analysis Sequential Projection Pursuit
Van Gyseghem et al. [21,22,23]	Chemometric	Visual interpretation of color maps and weighted-average-linkage dendrograms of: PCA, Pearson's correlation coefficient, different hierarchical clustering techniques, the Kennard and Stone algorithm, AAMRT and GPCM with McNemar's statistical test
Snyder et al. [38]	HSM based approach	F_s factor of column similarity
Van Gyseghem et al. [19]	Chemometric	Pareto optimality PCA Derringer's desirability functions
Gilar et al. [24]	Bin counting method	Orthogonality
Forlay-Frick et al. [33]	Statistical tests Correlation coefficients	GPCM with different statistical tests Pearson's r Sperman's ρ Kendall's τ Orthogonality ratio
Fornal et al. [20]	Chemometric	PCA
Watson et al. [25]	Bin counting method	Fractional coverage Percentage coverage Correction of Gilar's orthogonality [24]
Davis et al. [41] Semard et al. [30]	Statistical Overlap Theory Geometric approach	Effective peak capacity Minimum convex hull
Al Bakain et al. [31]	Geometric approach Statistical tests	Area covered by the confidence ellipse Pearson χ^2 statistic tests of independence Likelihood-ratio statistic
Schure [52]	Fractal analysis	Fractal dimension
Pourhaghighi et al. [17]	Information theory	Percentage orthogonality based on conditional information entropy
Rutan et al. [32]	Geometric approach	Fractional coverage metrics based on ecological home range
Nowik et al. [34,35]	Nearest neighbor distances approach	Arithmetic mean of NNDs Geometric mean of NNDs Harmonic mean of NNDs Optimality coefficient
Zeng et al. [26]	Bin counting method combined with statistic analysis	Marriott's orthogonality metric
Camenzuli and Schoenmakers [36]	Asterisk equations	Z parameters A_0 value
Mani-Varnosfaderani and Ghaemmaghami [27]	Bin counting method combined with statistical analysis	Correction of Marriott's orthogonality metric [26]
Leonhardt et al. [28]	Bin counting method combined with statistical analysis	Effective separation space coverage
Mommers and Wal [29]	Bin counting method Statistical analysis	%BIN %FIT

contributions are accounted for in the semi-empirical hydrophobic-subtraction model by Wilson et al. [37] and Snyder et al. [38]:

$$\log\alpha \equiv \log\left(\frac{k}{k_{ref}}\right) = \eta' H - \sigma' S^* + \beta' A + \alpha' B + \kappa' C \qquad (1)$$

where α is the separation factor, k is the retention factor of the solute under consideration, and k_{ref} is the retention factor of a non-polar reference solute (e.g., ethyl-benzene). The normalization of k with k_{ref} minimizes the effect of differences in column porosity, column surface area, and column ligand concentration [37]. In Eq. (1), the solute-specific parameters η', σ', β', α' and κ' represent, in order, hydrophobicity, bulkiness, hydrogen-bond basicity, hydrogen-bond acidity, and solute ionization state. Similarly, the column selectivity parameters H, S^*, A, B and C denote hydrophobicity, stationary phase resistance to solute penetration, hydrogen-bond acidity, hydrogen-bond basicity, and cation exchange capacity, respectively. The solute-specific parameters represent the solutes' molecular interaction at a given temperature and mobile phase. On the other hand, column selectivity parameters are almost insensitive to small variations in operating conditions, except for C, which strongly depends on pH [38,53].

The HSM retention model has been employing by many authors in a similar context. As an example, Liu et al. [54–56] uses the HSM model in their framework for evaluating and comparing the performance of single and tandem column separations; while Lindsey et al. [48] uses the HSM model as part of their scoring approach for column selection in comprehensive two-dimensional liquid chromatography.

Employing the HSM retention model is advantageous because of its prediction accuracy (k is estimated with an accuracy of ±0.7% [37]), and because of the large online repositories of column selectivity parameters that exist [57,58], which allows almost 300,000 pairs of RP columns to be investigated. The reader may want to use other retention models or experimental retention data available from within their organization, which is of course also possible.

2.2. A short overview of the orthogonality criteria

Eight orthogonality criteria are selected for comparison with our new approach: three correlation coefficients, two geometric criteria, a column similarity factor, and arithmetic and harmonic mean of the nearest neighbor distances. Among these, the geometric criteria characterizing the amount of surface coverage and the nearest neighbor distances approaches require that the retention factors ${}^{x}k_i$ be normalized in the interval [0, 1]:

$$ {}^{x}k_{i,norm} = \frac{{}^{x}k_i - {}^{x}k_{min}}{{}^{x}k_{max} - {}^{x}k_{min}} \tag{2} $$

where ${}^{x}k_{min}$ and ${}^{x}k_{max}$ are the retention factors of the least and most retained components (among those in the mixture) in the column under consideration, respectively. We do not consider for comparison older methods, such as the geometric approach to factor analysis and information theory-based indices, nor do we consider multivariate data analysis techniques, since selecting a proper set of columns from chemometric results (e.g., reading the loading plot and the score plot of a PCA) is far from straightforward and requires visual inspection of the data.

Correlation coefficients. Let us consider a sample made of n different components to be separated in an LCxLC system. The two columns are characterized by two sets of independent retention factors, ${}^{1}\mathbf{k} = \{{}^{1}k_1 \dots {}^{1}k_n\}$ and ${}^{2}\mathbf{k} = \{{}^{2}k_1 \dots {}^{2}k_n\}$, that can be derived from Eq. (1). The Pearson's r, Spearman's ρ, and Kendall's τ correlation coefficients between ${}^{1}\mathbf{k}$ and ${}^{2}\mathbf{k}$ have been used for column orthogonality evaluation [33]. A correlation coefficient equal to 1 in absolute value means a perfect correlation between ${}^{1}\mathbf{k}$ and ${}^{2}\mathbf{k}$ (i.e., the peaks cluster on one diagonal of the 2D separation space, thus, the same separation can be achieved with a 1D-LC column). On the other hand, a correlation coefficient equal to 0 indicates no correlation between ${}^{1}\mathbf{k}$ and ${}^{2}\mathbf{k}$ (i.e., the chromatographic peaks will spread over the separation space, thus, most of the area is utilized for separation) [40]. Note that, Kendall's τ and Spearman's ρ are to be preferred over Pearson's r when the sets of retention factors are not normally distributed and contain outliers [31].

Geometric criteria. The most popular geometric criteria are the bin counting method and the minimum convex hull method. The bin counting method involves (1) discretizing the normalized 2D separation space into a number of bins approximately equal to the number of components n within the samples; (2) counting the number of bins occupied by the components ($\sum$bins); and (3) calculating the orthogonality of separation according to one of the different metrics available. The original orthogonality metric O, defined by Gilar et al. [45] for square 2D separation space, and holding true only for $P \to \infty$, where P is the one-dimensional peak capacity, reads:

$$ O = \frac{\sum \text{bins} - P}{0.63P^2} \tag{3} $$

Later, Watson et al. [25] corrected this formulation to make it valid for all P:

$$ O_W = \frac{\sum \text{bins} - P}{0.63P^2 - P} \tag{4} $$

The column pair with the highest degree of orthogonality is characterized by the highest O (O_W). This work also considers another geometric criterion, the minimum convex hull method by Semard et al. [30]. This method (1) considers the smallest convex polygon (having angles no greater than 180 degrees) that contains all the normalized retention data; and (2) determines the convex hull area. Since the convex hull area provides information on the surface coverage, the best column pair is characterized by a minimum convex hull with the broadest area. Unlike the bin counting method, the minimum convex hull method does not require discretization; however, it is sensitive to outlying components, and it includes areas that are not utilized for separation [45].

Factor of column similarity. Another criterion that can be employed for RPLC column comparison, and thus for preliminary column selection during RPLCxRPLC method development, is the F_s similarity factor by Snyder et al. [38]. The F_s similarity factor, which is based on the Snyder–Dolan HSM model (see Section 2.1), consists in evaluating the Euclidean distance in the five-dimensional space of two columns as:

$$ F_s = \left\{ w_H \left[H_2 - H_1\right]^2 + w_S \left[S_2^* - S_1^*\right]^2 + w_A \left[A_2 - A_1\right]^2 + w_B \left[B_2 - B_1\right]^2 + w_C \left[C_2 - C_1\right]^2 \right\}^{1/2} \tag{5} $$

where H_i, S_i^*, A_i, B_i and C_i are the column selectivity parameters ($i = [1, 2]$), and w_H, w_S, w_A, w_B and w_C are weighting factors. The default values for the weighting factors were empirically determined by Gilroy et al. [59] considering a 67 small molecule sample and are $w_H = 12.5$, $w_S = 100$, $w_A = 30$, $w_B = 143$ and $w_C = 83$ [38]. Low values of F_s indicate high similarity between two columns, while high values of F_s indicate that the two columns are highly uncorrelated, and thus better suited for RPLCxRPLC systems. The main disadvantage of the F_s factor is that it accounts for column selectivity parameters only (i.e., solute-specific parameters do not appear in Eq. (5)) while ignoring the sample composition.

Nearest Neighbor Distances. The nearest neighbor distances approach consists in connecting each of the n peaks in the normalized 2D separation space with their closest neighbor [35]. The result is a minimum spanning tree characterized by $n - 1$ Euclidian distances di. The arithmetic mean of the distances:

$$ \overline{A} = \frac{\sum_{i=1}^{n-1} di}{n-1} \tag{6} $$

conveys information on peak spreading into the 2D separation space (long-length effect), while the harmonic mean of the distances:

$$ \overline{H} = \frac{n-1}{\sum_{i=1}^{n-1} di^{-1}} \tag{7} $$

offers insight into dispersion homogeneity and clustering (short-length effect). The best column pair is the one with maximum $\overline{A}$ and $\overline{H}$. Note that the nearest neighbor distances approach has been used for orthogonality assessment by many authors [34–36,46,47].

3. Methodology

In most articles available in the literature, resolution is used as a performance metric after chromatography is run. Our approach for preliminary column selection (see Section 3.2) is unique in that it embeds the definition of 2D resolution (see Section 3.1) into the preliminary stages of column selection and method development. Although we developed this approach for preliminary column selection of comprehensive procedures, it can easily be adaptable to aid the method development for heart-cutting 2D-LC.

3.1. The 2D resolution

In comprehensive 2D-LC, the aim is to maximize the likelihood that all solutes within the sample will be resolved [40]. Thus, we want to increase the orthogonality of separation by choosing a pair of uncorrelated columns that are able to spread the peaks across the separation space while reducing clustering. To overcome the limitations of the currently available metrics, we propose an alternative approach based on the definition of 2D resolution that implicitly accounts for both local effects (i.e., it considers the critical neighbor of each solute) as well as band broadening. Note that equal differences in retention time do not necessarily indicate equal differences in resolution. Thus, proper metrics for quantifying orthogonality should include resolution.

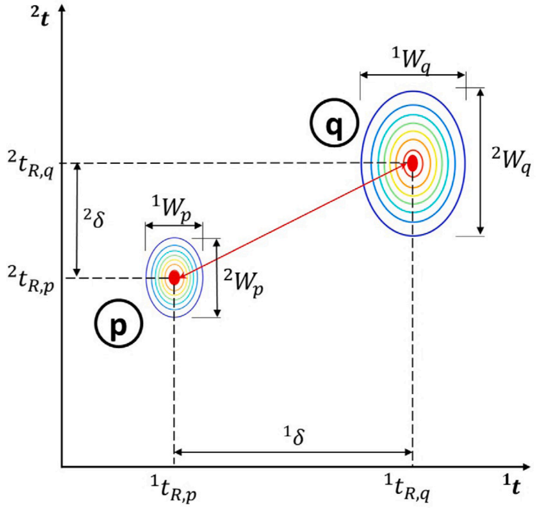


Fig. 1. The bivariate Gaussian peaks (with zero covariance) of two generic components p and q are represented in the 2D separation space, with ^1t and ^2t being the x- and y-axis, respectively. The Cartesian coordinates of the peak centres are $(^1t_{R,p}, {}^2t_{R,p})$ and $(^1t_{R,q}, {}^2t_{R,q})$. The Euclidian distance between the peak centres is represented by the red arrow. The contours represent the peak band broadening. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We consider the chromatographic peaks of 2D-LC separation to be bi-variate Gaussian with zero covariance in the 2D separation space. This allows to define the 2D resolution between two generic components p and q, R_{pq}, as an Euclidean norm [49,60]:

$$R_{pq} = \sqrt{^1R_{pq}^2 + {}^2R_{pq}^2} = \sqrt{\frac{^1\delta^2}{\left[\frac{1}{2}\left(^1W_p + {}^1W_q\right)\right]^2} + \frac{^2\delta^2}{\left[\frac{1}{2}\left(^2W_p + {}^2W_q\right)\right]^2}} \tag{8}$$

As can be seen from Fig. 1, the distances $^1\delta$ and $^2\delta$ can be defined as:

$$^x\delta = {}^xt_{R,q} - {}^xt_{R,p} \tag{9}$$

while ^xW_p and ^xW_q correspond to the peak width at the peak base (in time unit) for peaks p and q, respectively, and along direction x. The peak width at the peak base of Gaussian peaks is fourfold the standard deviation [39,61–63]:

$$^xW_p = 4\,{}^x\sigma_p = 4\frac{^xt_{R,p}}{\sqrt{^xN}} \tag{10}$$

$$^xW_q = 4\,{}^x\sigma_q = 4\frac{^xt_{R,q}}{\sqrt{^xN}} \tag{11}$$

In Eqs. (9) to (11), ^xN is the number of equivalent theoretical plates, while $^xt_{R,p}$ and $^xt_{R,q}$ denote the retention times of components p and q in the dimension x. For isocratic elution, these retention times are linked to the retention factors ^xk_p and ^xk_q, and are proportional to $^xt_0$, i.e., the elution time of a non-retained and non-penetrating component [61]:

$$^xt_{R,p} = {}^xt_0(^xk_p + 1) \tag{12}$$

$$^xt_{R,q} = {}^xt_0(^xk_q + 1) \tag{13}$$

Incorporating Eqs. (9) to (11) into Eq. (8), and expressing the resolution in terms of retention factors rather than retention times (see Eqs. (12) and (13)), yields:

$$R_{pq} = \frac{1}{2}\sqrt{^1N\frac{\left(^1k_q - {}^1k_p\right)^2}{\left(2 + {}^1k_p + {}^1k_q\right)^2} + {}^2N\frac{\left(^2k_q - {}^2k_p\right)^2}{\left(2 + {}^2k_p + {}^2k_q\right)^2}} \tag{14}$$

At the column selection stage of method development, column design (i.e., described by the column length and diameter) and column efficiency (i.e., described by the theoretical height equivalent to a theoretical plate) are not yet decided, thus ^1N and ^2N are unknown. To solve Eq. (14), we can either estimate ^1N and ^2N or assume that the two columns are characterized by the same number of equivalent theoretical plates:

$$^1N = {}^2N = N \tag{15}$$

If this is the case, i.e., for isocratic elution conditions only and assuming small injection volumes, the 2D resolution can be rewritten as:

$$R_{pq} \propto \sqrt{\frac{\left(^1k_q - {}^1k_p\right)^2}{\left(2 + {}^1k_p + {}^1k_q\right)^2} + \frac{\left(^2k_q - {}^2k_p\right)^2}{\left(2 + {}^2k_p + {}^2k_q\right)^2}} \tag{16}$$

Although considering isocratic elution conditions in LCxLC might seem a significant limitation, it might be advantageous in practice: it eliminates compatibility issues due to solvent mismatch and viscous fingering, it eliminates the need to perform 2D column re-equilibration after a gradient is run, and thus, it increases the time available for the 2D separation [9,64].

3.2. A new approach for preliminary column selection

Let us consider a sample made of n different components to be separated and a collection of m chromatography columns to choose from. To evaluate the most suitable pair of columns for an LCxLC system, the 2D resolution between all component pairs p and q in all possible combinations of columns is evaluated according to Eq. (14). The resolution $\mathbf{R}$ is a multidimensional array of shape $m \times m \times n \times n$:

$$\mathbf{R}^{m\times m\times n\times n} = \begin{bmatrix} \begin{bmatrix} R_{1111} & \cdots & R_{111n} \\ \vdots & \ddots & \vdots \\ R_{11n1} & \cdots & R_{11nn} \end{bmatrix} & \cdots & \begin{bmatrix} R_{1m11} & \cdots & R_{1m1n} \\ \vdots & \ddots & \vdots \\ R_{1mn1} & \cdots & R_{1mnn} \end{bmatrix} \\ \vdots & \ddots & \vdots \\ \begin{bmatrix} R_{m111} & \cdots & R_{m11n} \\ \vdots & \ddots & \vdots \\ R_{m1n1} & \cdots & R_{m1nn} \end{bmatrix} & \cdots & \begin{bmatrix} R_{mm11} & \cdots & R_{mm1n} \\ \vdots & \ddots & \vdots \\ R_{mmn1} & \cdots & R_{mmnn} \end{bmatrix} \end{bmatrix} \tag{17}$$

where the $R(i, j, p, q)$ element represents the two-dimensional resolution between components p and q when considering an LCxLC system made by coupling columns i and j (see Eq. (16)). For each pair of columns, we are interested in the *critical resolution* of each component, which can be defined as the 2D resolution of each component with its closest neighbor. In mathematical terms, we are interested in finding $\min_{q\neq p} R(i, j, p, :)\ \forall p \in [1, n]$. The multidimensional array of the critical resolution, $\mathbf{CR}$, has shape $m \times m \times n$ and reads:

$$\mathbf{CR}^{m\times m\times n} = \begin{bmatrix} \begin{bmatrix} CR_{111} \\ \vdots \\ CR_{11n} \end{bmatrix} & \cdots & \begin{bmatrix} CR_{1m1} \\ \vdots \\ CR_{1mn} \end{bmatrix} \\ \vdots & \ddots & \vdots \\ \begin{bmatrix} CR_{m11} \\ \vdots \\ CR_{m1n} \end{bmatrix} & \cdots & \begin{bmatrix} CR_{mm1} \\ \vdots \\ CR_{mmn} \end{bmatrix} \end{bmatrix} \tag{18}$$

In this approach, each column pair i and j is characterized by a distribution of critical resolutions, offering insights into whether the peaks overlap and on the probability of the overlapping occurrence. For the first time, we suggest using these distributions to discriminate column pair performance. And the simplest way to compare distributions is by using summary statistics, such as mean and quartiles. The use of summary statistics on the distributions reduce the dimensionality of $\mathbf{CR}$ while preserving the amount of information:

$$\mathbf{Q}^{m\times m} = \begin{bmatrix} q_{11} & \cdots & q_{1m} \\ \vdots & \ddots & \vdots \\ q_{m1} & \cdots & q_{mm} \end{bmatrix} \tag{19}$$

In what follows, we will consider three different summary statistics to reduce the dimensionality of $\mathbf{CR}$ and obtain $\mathbf{Q}$, namely the 25th percentile, the 50th percentile, and the mean. However, other statistics could also be used. Eq. (19) can be used to evaluate the performance of each column pair, compare and rank different pairs, and select the best column pair for comprehensive 2D-LC. The best pair of columns is the one with the maximum value of the q_{ij} statistics.

A similar procedure can be applied when we are interested in separating only one or a few target components, for which the heart-cutting procedures may be the most appropriate separation techniques. For these cases, we need to find a pair of columns suitable for separating the n^* components of interest ($n^* < n$), ensuring that the peaks of the target molecules are as pure as possible and do not overlap with the peaks of the impurities. **R** in Eq. (17) and **CR** in Eq. (18) then reduce to a $m \times m \times n \times n^*$ and a $m \times m \times n^*$ matrix, respectively. Note that, when $n^* = 1$, i.e., we are only interested in a single component, no statistics are needed, and the best column pair is the one with maximum CR_{ij1}.

The main advantages of this new approach are that: (1) it implicitly accounts for local effects (i.e., local crowding) and band broadening; (2) it can be tailored to a specific sample from which it is possible to obtain the maximum amount of relevant information; (3) it is not influenced by outlying components (i.e., data outliers in the 2D separation space); (4) it does not rely on adjustable parameters; (5) there is no need for normalization; and (6) it can easily be implemented.

4. Results and discussion

In this work, we examine the validity of our new approach for two different scenarios. In the first scenario, we considered a sample made of the standard 16 test components used in Wilson et al. [37] and Snyder et al. [38] studies: Acetophenone, Benzonitrile, Anisole, Toluene, Ethylbenzene, 4-Nitrophenol, 5-Phenylpentanol, 5,5-Diphenylhydantoin, cis-Chalcone, trans-Chalcone, N,N-dimethylacetamide, N,N-diethylacetamide, 4-n-Butylbenzoic acid, Mefenamic acid, Nortriptyline, and Amitriptyline. The solute-specific parameters for these components are given at 35 °C, at pH 2.8, and with a mobile phase of 50/50 acetonitrile and water [38]. To illustrate the effectiveness of our approach when considering much more complex samples, a second scenario was considered. In the second scenario, the sample was composed of 25 artificial components with solute-specific parameters that were synthetically generated as random variables and drawn from a uniform distribution. The range of variability for each solute-specific parameter was established by considering the minimum and maximum values of the corresponding solute-specific parameters in the first scenario, i.e., $\eta' \in [-1.90, 0.05]$, $\sigma' \in [-0.22, 0.92]$, $\beta' \in [-0.05, 0.99]$, $\alpha' \in [-0.29, 1.12]$ and $\kappa' \in [-0.05, 0.85]$. Note that, synthetically-generated analytes are also used in [48,54–56].

4.1. Comparison of metrics for determining column orthogonality

Given the database of column selectivity parameters [57] (which lists more than 700 RPLC columns) and the HSM, we evaluated almost 300,000 pairs of RPLC columns, ranked the column pairs, and selected the most suitable pair according to the different metrics used to describe column orthogonality summarized in Table 2. The new metrics proposed in this work are the 25th percentile, the 50th percentile, and the mean of the critical resolution distributions. The other metrics are instead well established in the literature (see Section 2.2).

Table 2
Overview of metrics used to describe column orthogonality that has been considered in the two scenarios. The first three metrics are the critical resolution (CR) statistics proposed in this work, while the others are well-established in the literature.

Orthogonality metric	Symbol	Equation	Reference
CR statistic: the 25th percentile	q_1	–	This work
CR statistic: the 50th percentile	q_2	–	This work
CR statistic: the mean	$\overline{CR}$	–	This work
Pearson's correlation coefficient	r	–	Forlay-Frick et al. [33]
Spearman's correlation coefficient	ρ	–	Forlay-Frick et al. [33]
Kendall's correlation coefficient	τ	–	Forlay-Frick et al. [33]
Watson's Orthogonality metric	O_W	Eq. (4)	Watson et al. [25]
Minimum convex hull	CH	–	Semard et al. [30]
Column similarity factor	F_s	Eq. (5)	Snyder et al. [38]
NND: Arithmetic mean	$\overline{A}$	Eq. (6)	Nowik et al. [35]
NND: Harmonic mean	$\overline{H}$	Eq. (7)	Nowik et al. [35]

We used Python 3.11 [65] and the `DataFrame.corr` method of the pandas library to compute the Pearson, Spearman and Kendall correlation coefficients [66], and the `spatial.ConvexHull` method of the Scipy library to determine the minimum convex hull and its surface area [67]. For the other metrics, dedicated Python functions were coded.

Figs. 2 and 3 show the normalized retention factors (see Eq. (2)) spread across the 2D separation space when considering the best RPLC column pair as provided by each of the considered metrics, for both the first and the second scenario, respectively. The normalization allows us to compare the retention factor of each component in each dimension on the same scale. In all panels, the x-axis is relative to the first column dimension, while the y-axis is relative to the second column dimension. The axis labels report the names of the selected columns and the relative identification numbers as they appear in Stoll [57]. The order of the columns is chosen according to the rule of thumb that says that, when using RP columns in both dimensions, the ^{1}D column must be less retentive than the ^{2}D column [10]. In other words, each subplot is a simplified 2D space (simplified as only dots are plotted) for the two columns selected with each method.

Fig. 2 shows that the best column pairs are those selected by the 2D critical resolution statistics proposed in this work, i.e., 25th percentile (or lower quartile), 50th percentile (or median) and mean of the CRs; and the arithmetic mean and harmonic mean of the nearest neighbor distances. The peaks are well-scattered and unevenly distributed in both directions of separation. Furthermore, the 2D critical resolution statistics allow more space around the most retained compounds, which are characterized by higher peak band broadening (i.e., components on the top right corner of the 2D separation space are more sparse), compared to the less retained components in the same plot, which are characterized by narrower peaks. On the other hand, the column pairs suggested by the other metrics do not seem to be good candidates for the scenario considered. In the case of the Pearson's correlation coefficient and the F_s similarity factor, the peak apexes are roughly parallel to the x-axis, and all but one or two components cluster within a very narrow elution window in the second-dimension separation. In the case of the Spearman and Kendall correlation coefficients, the two columns are very low correlated; however, a big portion of the separation space is unused. Finally, the spread of the components over the 2D separation space is rather weak also if we consider the orthogonality metric by Watson and the minimum convex hull.

The panels in the second scenario, Fig. 3, are more challenging to interpret, given the higher number of components in the sample. The only conclusion that we can draw is that the orthogonality metric by Watson and F_s similarity factors advise pairs of columns that are clearly not optimal as they do not utilize properly the separation space, with peaks clustering on the diagonal and the axes, respectively.

The plots in Figs. 2 and 3 have been included to facilitate a visual inspection of the peak distribution in the normalized 2D separation space. Although such plots are often used for ease of visualization of column orthogonality information, they do not provide any quantitative information. A quantitative comparison between the selected column pairs is possible if we consider the metrics reported in Tables 3 and 4. The results in Table 3 show that, although Nucleodur Polartec C18 (ID 582) and Primesep A (ID 664) are the two most different columns, with $F_s \approx 512$, they are not recommended for the mixture under consideration. The corresponding O_W is indeed only 0.164. Very low values of $\overline{H}$ suggest low dispersion homogeneity and clustering when correlation coefficients and F_s similarity factor are employed for preliminary column selection. Surface coverage techniques (i.e., O_w and CH) are not suitable either. Because of the very low critical

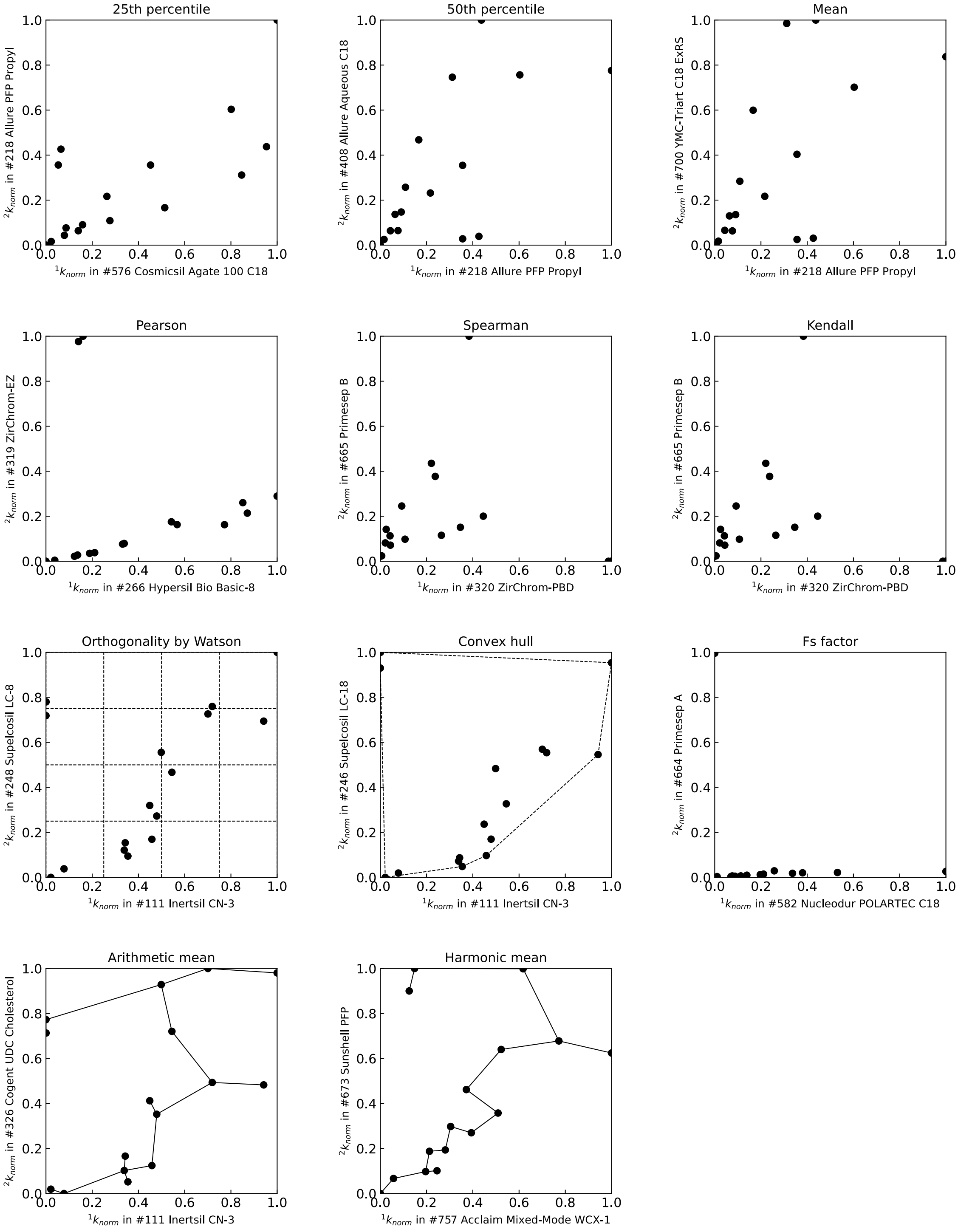


Fig. 2. Scenario 1 (16 components in the sample): Retention factors spread across the normalized 2D separation space when considering the best column pair as provided by the different metrics. The first three metrics (three panels on top) are the critical resolution (CR) statistics proposed in this work, while the other metrics are well-established in the literature.

resolution statistic, there is a high probability of peak overlapping. Overall, the CR statistics and the NNDs statistics seem to be the most appropriate metrics, providing a good compromise between data scattering and distribution homogeneity. The situation is slightly different if we consider the second scenario. In Table 4 the low CR statistics associated with CH, $\overline{A}$ and $\overline{H}$ indicate intense overlapping, while the low O_W values associated with the correlation coefficients implies an inefficient utilization of the separation space. Although the CR statistics exhibit some dispersion, they seem to outperform all the other metrics when it comes to considering column orthogonality while avoiding data clustering.

4.2. Impact of different metrics for determining column orthogonality on optimization

Our orthogonality metric is based on critical resolution statistics, where the resolution is defined by Eq. (14). Eq. (14) is a function of retention factors only, i.e., is independent of column design and some

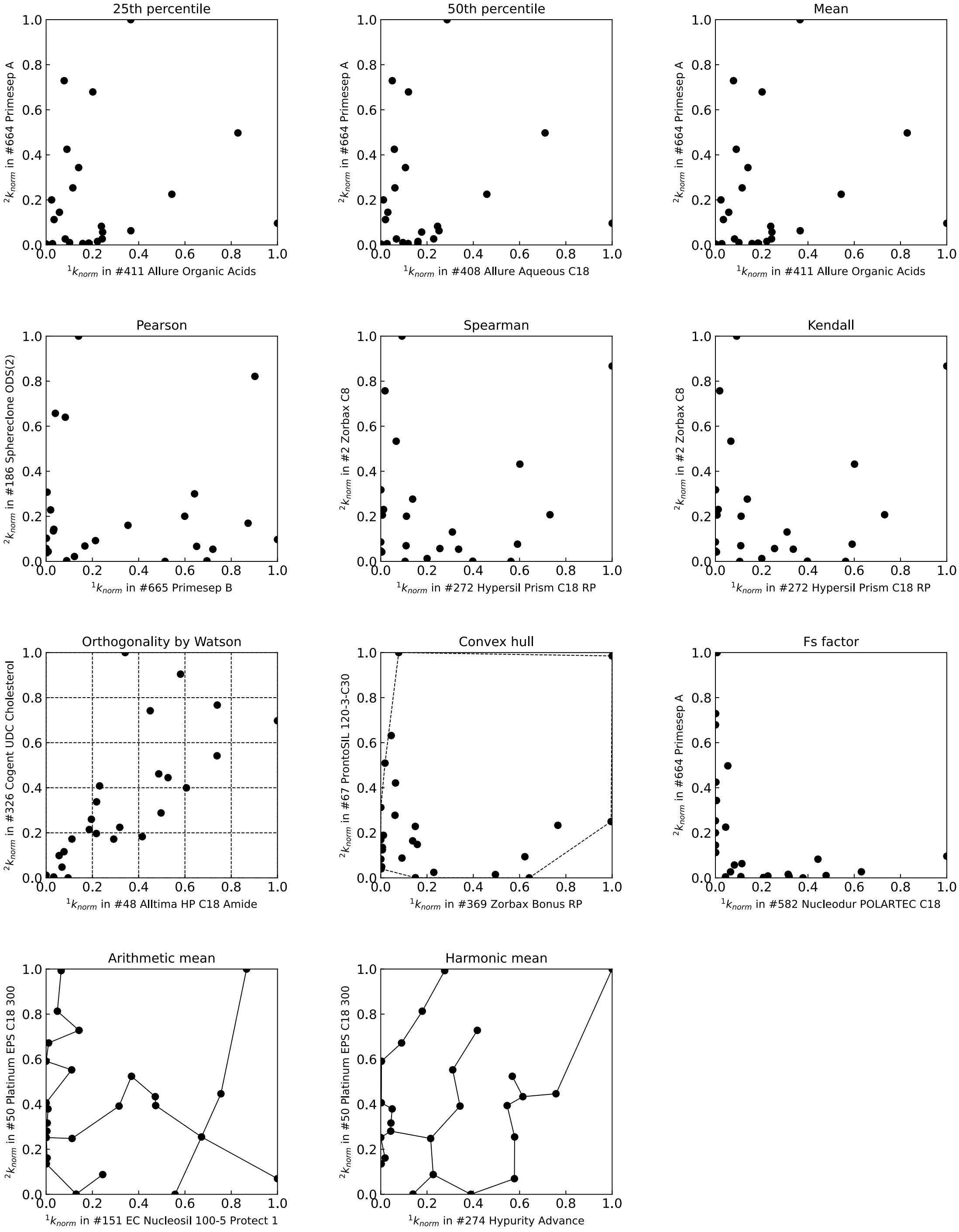


Fig. 3. Scenario 2 (25 components in the sample): Retention factors spread across the normalized 2D separation space when considering the best column pair as provided by the different metrics. The first three metrics (three panels on top) are the critical resolution (CR) statistics proposed in this work, while the other metrics are well-established in the literature.

operating conditions. Similar considerations can be drawn for the other orthogonality metrics available. Thus, the preliminary column selection step of the 2D-LC method development provides the most suitable column pair for the considered mixture only. Once selected, proper optimization of column design (e.g., column diameter and length) and operating conditions (e.g., flow rate into the column and pH) is required to maximize separation efficiency while minimizing time and costs.

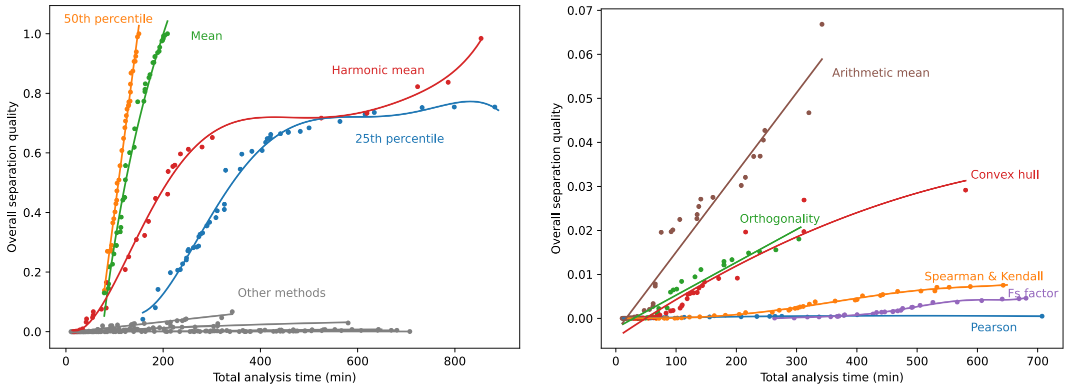
Given the best pair of columns identified in Section 4.1, we next assess the impact of considering different orthogonality metrics for preliminary column selection on column optimization. We determine the optimal operating conditions and column design of the RPLCxRPLC system via a new shortcut model for in-silico prediction of retention time and peak width. The shortcut model was proposed by Tirapelle et al. [9] based on the following assumptions: isocratic elution conditions are considered in both dimensions; all peaks are characterized by a specific peak width; and all peaks are assumed bi-variate Gaussian in

Table 3
Scenario 1 (16 components in the sample): Best column candidates according to different orthogonality metrics. For cross-comparison, the value of all metrics for each candidate pair is also reported.

Method	# 1D	# 2D	$q1$	$q2$	$\overline{CR}$	r	ρ	τ	O_W	CH	F_s	$\overline{A}$	$\overline{H}$
q_1	576	218	**0.138**	0.161	0.177	0.742	0.668	0.583	0.658	0.485	102.131	0.160	0.079
q_2	218	408	0.136	**0.195**	0.182	0.673	0.676	0.583	0.658	0.496	107.941	0.163	0.078
$\overline{CR}$	218	700	0.136	0.166	**0.188**	0.631	0.638	0.533	0.658	0.551	124.606	0.171	0.075
r	266	319	0.016	0.019	0.029	**0.000**	0.632	0.600	0.329	0.518	155.606	0.136	0.035
ρ	320	665	0.023	0.046	0.064	0.020	**0.174**	0.217	0.164	0.504	349.615	0.143	0.035
τ	320	665	0.023	0.046	0.064	0.020	0.174	**0.217**	0.164	0.504	349.615	0.143	0.035
O_W	111	248	0.038	0.047	0.057	0.501	0.471	0.483	**1.151**	0.595	308.977	0.159	0.080
CH	111	246	0.038	0.073	0.078	0.216	0.353	0.417	0.658	**0.752**	349.384	0.170	0.063
F_s	582	664	0.052	0.066	0.089	0.302	0.306	0.433	0.164	0.498	**512.248**	0.133	0.017
$\overline{A}$	111	326	0.036	0.044	0.054	0.418	0.488	0.383	1.151	0.703	280.645	**0.198**	0.120
$\overline{H}$	757	673	0.052	0.080	0.082	0.424	0.500	0.483	0.658	0.571	113.258	0.174	**0.121**

Table 4
Scenario 2 (25 components in the sample): Best column candidates according to different orthogonality metrics. For cross-comparison, the value of all metrics for each candidate pair is also reported.

Method	# 1D	# 2D	$q1$	$q2$	$\overline{CR}$	r	ρ	τ	O_W	CH	F_s	$\overline{A}$	$\overline{H}$
q_1	411	664	**0.159**	0.175	0.209	0.198	0.395	0.307	0.465	0.632	211.618	0.116	0.028
q_2	408	664	0.128	**0.181**	0.206	0.137	0.348	0.280	0.372	0.582	227.312	0.110	0.025
$\overline{CR}$	411	664	0.159	0.175	**0.209**	0.198	0.395	0.307	0.465	0.632	211.618	0.116	0.028
r	665	186	0.040	0.060	0.072	**0.001**	0.022	0.033	0.558	0.822	244.255	0.130	0.043
ρ	272	2	0.026	0.070	0.080	0.155	**0.000**	0.000	0.465	0.758	315.169	0.127	0.023
τ	272	2	0.026	0.070	0.080	0.155	0.000	**0.000**	0.465	0.758	315.169	0.127	0.023
O_w	48	326	0.057	0.074	0.079	0.727	0.840	0.653	**0.930**	0.449	67.478	0.110	0.071
CH	369	67	0.033	0.048	0.060	0.151	0.025	0.007	0.558	**0.917**	297.694	0.132	0.047
F_s	582	664	0.059	0.079	0.089	0.424	0.633	0.393	0.372	0.521	**512.248**	0.091	0.049
$\overline{A}$	151	50	0.026	0.029	0.045	0.046	0.026	0.007	0.930	0.897	290.424	**0.158**	0.082
$\overline{H}$	274	50	0.028	0.057	0.061	0.282	0.238	0.173	0.930	0.700	135.104	0.153	**0.100**



(a) Better-performing metrics and other metrics

(b) Magnification of the other worse-performing metrics

Fig. 4. Scenario 1 (16 components in the sample): Fitted Pareto fronts for the solutions obtained from multi-objective optimization of the LCxLC separation with overall separation efficiency and total analysis time as the objective functions of (a) the better-performing metrics (i.e., 25th percentile of the CRs, 50th percentile of the CRs, mean of the CRs, and harmonic mean of NNDs) and other metrics considered, and (b) the other worse-performing metrics at a magnified scale (notice that the overall separation quality is only up to 0.07). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the 2D separation space. The shortcut model considers constraints on modulation time and maximum allowable pressure drop [9].

The shortcut model is implemented in gPROMS ModelBuilder [68], while optimization is performed through an external stochastic optimizer developed in-house [69] that is based on the Fast and Elitist Non-dominated Sorting Genetic Algorithm (NSGA-II) [70]. Multi-objective optimization is performed considering total analysis time and overall separation quality as the two objective functions, and with a minimum baseline resolution of $R_{min} \geq 1.2$. More details of the optimization procedure are provided by Tirapelle et al. [9].

Fig. 4 shows the Pareto fronts, with their corresponding fitting curves, of the solutions obtained from multi-objective optimization (maximizing both overall separation efficiency and minimizing total analysis time) of the LCxLC separation of a mixture of 16 components. Fig. 4(a) clearly displayed that the metrics can be categorized into (1) the better-performing metrics, which is the focus of this figure, and (2) the worse-performing metrics, which are magnified and illustrated in Fig. 4(b). As we can see, the 50th percentile and the mean of the CRs (i.e., two of the metrics proposed in this work) outperform all the metrics traditionally used for 2D-LC column selection, with the 50th percentile of the CRs showing a slightly better performance. Both metrics have close Pareto fronts, and complete separation (i.e., overall separation quality equals one) of 16 components can be realized within 200 min. The harmonic mean of the nearest neighbor distance metric is the next best-performing metric. Compared to the 50th percentile and mean of the CRs, this requires a much longer time to achieve the same separation performance (i.e., same overall separation quality values), and the time difference becomes more significant towards complete separation. The 25th percentile (another metric based on critical resolution statistics) also shows a good performance (shown in Fig. 4(a)). The Pareto fronts indicate that no available column designs

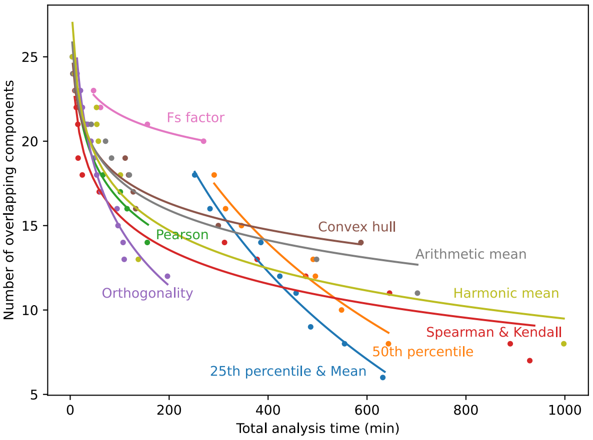


Fig. 5. Scenario 2 (25 components in the sample): Fitted Pareto fronts for the solutions obtained from the multi-objective optimization of the LCxLC separation with the number of overlapping components and total analysis time as the objective functions of all the metrics studied. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

could achieve complete separation within 1000 min given the optimization constraints. If the constraint on total analysis time is relaxed, the 25th percentile of the CRs can also achieve a complete separation (not shown). It is worth noting that constraints such as pressure drop and total analysis time are not taken into account by the column selection metrics in the preliminary column selection stage (see Fig. 2); however, these constraints are considered during optimization of column design and operating conditions. This explains the "different" conclusions that can be drawn from Figs. 2 and 4. The difference between metrics and optimization is the most obvious for the arithmetic mean of the nearest neighbor distances, where the metric shows great potential for complete separation, but the optimization results indicate overall bad performance. Similar to the 25th percentile of the CRs, a complete separation can also be achieved for this metric by removing the time constraint (not shown). Other metrics, such as correlation coefficients, convex hull and Watson's Orthogonality metric, performed equally badly in terms of both the total analysis time and the overall separation quality, and will not be discussed further.

Moving on to the second scenario, considering 25 in-silico components, the multi-objective optimization was originally performed with the same objective functions; however, the optimization became unsuitable for this more complex mixture. The overall separation quality is defined as the product of a separation indicator, namely the normalized resolution values (more details in Tirapelle et al. [9]), which has a drawback for a complex mixture. To illustrate this drawback, imagine a system where 23 out of 25 components can be completely separated, but two components almost overlap (i.e., are very difficult to separate). The 23 easy components will have a separation indicator of 1, but the two difficult components may have values of 0.01, hence the final overall separation quality is 0.0001. If the two difficult components are almost impossible to separate regardless of the column pairs, then all the metrics will end up with indistinguishable overall separation quality; thus, the overall separation quality fails to represent the actual performance of the columns. Instead, for the LCxLC separation of the 25-component mixture, the overall separation quality objective function is replaced with an objective function based on the number of overlapping components, which is a straightforward indication of the separation performances of the column pairs.

The optimization results (i.e., Pareto fronts and fitted curves) using the new objective functions (minimizing the number of overlapping components and minimizing the total analysis time) are shown in Fig. 5. It is clear that the metrics based on critical resolution statistics proposed in this work have different trends/patterns of the Pareto fronts compared to the other traditional metrics from the literature. In particular, the metrics based on critical resolution statistics are more suited when a high number of compounds have to be separated. It is noted that no metrics can achieve complete separation for the second scenario. However, the metrics based on critical resolution statistics (i.e., 25th percentile and mean) still perform the best (e.g., at most, 19 out of 25 components can be separated) thus satisfying the main purpose of 2D-LC, which is to separate multiple components. Comparing the metrics based on critical resolution statistics proposed in this work and the second-best metrics, the Spearman and Kendall correlation coefficients (which both yielded the same column pair), in order to achieve the same separation performance (e.g., eight overlapping components), the total analysis time is increased from around 550 min for the critical resolution metrics to 900 min for the other two methods.

In conclusion, the comparison of the Pareto fronts for the 16- and 25-component systems indicates that the most suitable approach for preliminary column selection is the new approach proposed in this work. Indeed, when using the columns recommended by our approach in a 2D-LC system, a more complete (i.e., separate more components from the mixture) and faster separation (under the same separation performance) is always achieved, at least for the two scenarios considered. Three metrics based on critical resolution statistics (i.e., 25th percentile, 50th percentile, mean) were studied, but there was no clear indication of which of these is better. We have seen that no single summary statistic is universally better than the others. However, we recommend employing the mean when the main focus is to minimize the chromatography running time, and the 25th percentile when the primary goal is to maximize the number of resolved components. Compared to the median and mean, the 25th percentile is a stricter summary statistic, yielding the identification of the best column pair, though requiring a longer runtime. It is also important to remember that, the mean is influenced by outliers, while the median is not. If there are a lot of anomalies and outliers, we recommend comparing the medians instead of the means. Note that, with the shortcut method [9], a single multi-objective optimization, which requires around 100,000 simulation runs, takes only about 2 min (the simulations are performed in parallel using 54 threads on a workstation equipped with Intel Xeon Gold 6226R CPUs featuring 32 cores and 64 threads, 2.91 GHz clock speed and 192 GB RAM (3200 MHz)), so different metrics can easily be considered and optimized.

5. Conclusions

Two-dimensional liquid chromatography (2D-LC) can be used to tackle complex separation tasks efficiently by combining the separation performance of two different columns. However, 2D-LC is rarely implemented in industry because of the challenges encountered in developing the method, requiring a two-step process involving preliminary column selection followed by system optimization. In this work, we have focused mainly on the preliminary column selection step and we have proposed a new approach for the evaluation of column orthogonality. Our proposed approach, which can be tailored for both heartcutting and comprehensive 2D procedures, is based on the definition of 2D resolution, and so it implicitly accounts for both local effects and peak band broadening, which has so far not been considered in the literature. Furthermore, the approach is easily deployable, consisting in evaluating summary statistics of critical resolution distributions. The two main outcomes of this work are: (1) as shown by multi-objective optimization and the Pareto fronts, preliminary column selection is a key step of *in-silico* method development of comprehensive 2D-LC chromatography, where the use of an appropriate approach is essential, and (2) our proposed approach outperforms traditional methods for evaluation of column orthogonality, proving the best-performing LC × LC system for the scenarios considered. Note that the applicability of the method is not limited to predicting the best combinations of chromatographic columns for the optimization of 2D-LC systems. It can also be used, for instance, to evaluate the separation performance of a multi-dimensional chromatographic system, or to select a second independent column (for an initial 1D separation) to increase peak purity to meet regulatory requirements. These extensions are, however, not considered in this work.

CRediT authorship contribution statement

Monica Tirapelle: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Fanyi Duanmu:** Writing – original draft, Visualization, Software, Investigation. **Dian Ning Chia:** Writing – original draft, Visualization, Software, Investigation. **Maximilian O. Besenhard:** Writing – review & editing. **Luca Mazzei:** Writing – review & editing. **Eva Sorensen:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Monica Tirapelle reports financial support was provided by Eli Lilly and Company.

Fanyi Duanmu and Dian Ning Chia report financial support was provided by Engineering and Physical Sciences Research Council.

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

Acknowledgments

The authors wish to acknowledge the financial support given to this research project by Eli Lilly and Company, and the Engineering and Physical Sciences Research Council (EPSRC), grant code EP/T005556/1.

Data availability

Data will be made available on request.

References

[1] D.R. Stoll, Introduction to two-dimensional liquid chromatography—theory and practice, in: Handbook of Advanced Chromatography/Mass Spectrometry Techniques, Elsevier, 2017, pp. 227–286.
[2] F. Erni, R. Frei, Two-dimensional column liquid chromatographic technique for resolution of complex mixtures, J. Chromatogr. A 149 (1978) 561–569.
[3] K. Zhang, Y. Li, M. Tsang, N.P. Chetwyn, Analysis of pharmaceutical impurities using multi-heartcutting 2D LC coupled with UV-charged aerosol MS detection, J. Sep. Sci. 36 (18) (2013) 2986–2992.
[4] 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.
[5] S.R. Groskreutz, M.M. Swenson, L.B. Secor, D.R. Stoll, Selective comprehensive multi-dimensional separation for resolution enhancement in high performance liquid chromatography. Part I: Principles and instrumentation, J. Chromatogr. A 1228 (2012) 31–40.
[6] S.A. Cohen, M.R. Schure, Multidimensional Liquid Chromatography: Theory and Applications in Industrial Chemistry and the Life Sciences, John Wiley & Sons, 2008.
[7] D.R. Stoll, P.W. Carr, Two-dimensional liquid chromatography: A state of the art tutorial, Anal. Chem. 89 (1) (2016) 519–531.
[8] O. Jones, Two-Dimensional Liquid Chromatography: Principles and Practical Applications, Springer Nature, 2020.
[9] M. Tirapelle, D.N. Chia, F. Duanmu, M.O. Besenhard, L. Mazzei, E. Sorensen, In-silico method development and optimization of on-line comprehensive two-dimensional liquid chromatography via a shortcut model, J. Chromatogr. A 1721 (2024) 464818.
[10] S. Kromidas, Optimization in HPLC. Concepts and Strategies, WILEY-VCH, Weinheim, Germany, 2021.
[11] M. Iguiniz, S. Heinisch, Two-dimensional liquid chromatography in pharmaceutical analysis. Instrumental aspects, trends and applications, J. Pharm. Biomed. Anal. 145 (2017) 482–503.
[12] F. Bedani, P.J. Schoenmakers, H.-G. Janssen, Theories to support method development in comprehensive two-dimensional liquid chromatography–A review, J. Sep. Sci. 35 (14) (2012) 1697–1711.
[13] Z. Liu, D.G. Patterson, M.L. Lee, Geometric approach to factor analysis for the estimation of orthogonality and practical peak capacity in comprehensive two-dimensional separations, Anal. Chem. 67 (21) (1995) 3840–3845.
[14] M.J. Gray, G.R. Dennis, P.J. Slonecker, R.A. Shalliker, Evaluation of the two-dimensional reversed-phase-reversed-phase separations of low-molecular mass polystyrenes, J. Chromatogr. A 1015 (2003) 89–98.
[15] I. François, K. Sandra, P. Sandra, Comprehensive liquid chromatography: fundamental aspects and practical considerations—a review, Anal. Chim. Acta 641 (1–2) (2009) 14–31.
[16] P.J. Slonecker, X. Li, T.H. Ridgway, J.G. Dorsey, Informational orthogonality of two-dimensional chromatographic separations, Anal. Chem. 68 (1996) 682–689.
[17] M.R. Pourhaghighi, M. Karzand, H.H. Girault, Orthogonality of two-dimensional separations based on conditional entropy, Anal. Chem. 83 (2011) 7676–7681.
[18] A. Detroyer, V. Schoonjans, F. Questier, Y.V. Heyden, A.P. Borosy, Q. Guo, D.L. Massart, Exploratory chemometric analysis of the classification of pharmaceutical substances based on chromatographic data, J. Chromatogr. A 897 (2000) 23–36.
[19] E. Van Gyseghem, M. Jimidar, R. Sneyers, D. Redlich, E. Verhoeven, D.L. Massart, Y.V. Heyden, Selection of reversed-phase liquid chromatographic columns with diverse selectivity towards the potential separation of impurities in drugs, J. Chromatogr. A 1042 (2004) 69–80.
[20] E. Fornal, P. Borman, C. Luscombe, Determination of selectivity differences for basic compounds in gradient reverse phase high performance liquid chromatography under high pH conditions by partial least squares modelling, Anal. Chim. Acta 570 (2006) 267–276.
[21] E. Van Gyseghem, S. Van Hemelryck, M. Daszykowski, F. Questier, D. Massart, Y. Vander Heyden, Determining orthogonal chromatographic systems prior to the development of methods to characterise impurities in drug substances, J. Chromatogr. A 988 (1) (2003) 77–93.
[22] E. Van Gyseghem, I. Crosiers, S. Gourvénec, D.L. Massart, Y.V. Heyden, Determining orthogonal and similar chromatographic systems from the injection of mixtures in liquid chromatography–diode array detection and the interpretation of correlation coefficients color maps, J. Chromatogr. A 1026 (2004) 117–128.
[23] E. Van Gyseghem, B. Dejaegher, R. Put, P. Forlay-Frick, A. Elkihel, M. Daszykowski, K. Héberger, D.L. Massart, Y.V. Heyden, Evaluation of chemometric techniques to select orthogonal chromatographic systems, J. Pharm. Biomed. Anal. 41 (2006) 141–151.
[24] M. Gilar, P. Olivova, A.E. Daly, J.C. Gebler, Orthogonality of separation in two-dimensional liquid chromatography, Anal. Chem. 77 (2005) 6426–6434.
[25] N.E. Watson, J.M. Davis, R.E. Synovec, Observations on "orthogonality" in comprehensive two-dimensional separations, Anal. Chem. 79 (2007) 7924–7927.
[26] Z.D. Zeng, H.M. Hugel, P.J. Marriott, A modeling approach for orthogonality of comprehensive two-dimensional separations, Anal. Chem. 85 (2013) 6356–6363.
[27] A. Mani-Varnosfaderani, M. Ghaemmaghami, Assessment of the orthogonality in two-dimensional separation systems using criteria defined by the maximal information coefficient, J. Chromatogr. A 1415 (2015) 108–114.
[28] J. Leonhardt, T. Teutenberg, G. Buschmann, O. Gassner, T.C. Schmidt, A new method for the determination of peak distribution across a two-dimensional separation space for the identification of optimal column combinations, Anal. Bioanal. Chem. 408 (2016) 8079–8088.
[29] J. Mommers, S.V.D. Wal, Two metrics for measuring orthogonality for two-dimensional chromatography, J. Chromatogr. A 1586 (2019) 101–105.
[30] G. Semard, V. Peulon-Agasse, A. Bruchet, J.P. Bouillon, P. Cardinaël, Convex hull: A new method to determine the separation space used and to optimize operating conditions for comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1217 (2010) 5449–5454.
[31] R. Al Bakain, I. Rivals, P. Sassiat, D. Thiébaut, M.C. Hennion, G. Euvrard, J. Vial, Comparison of different statistical approaches to evaluate the orthogonality of chromatographic separations: Application to reverse phase systems, J. Chromatogr. A 1218 (2011) 2963–2975.
[32] S.C. Rutan, J.M. Davis, P.W. Carr, Fractional coverage metrics based on ecological home range for calculation of the effective peak capacity in comprehensive two-dimensional separations, J. Chromatogr. A 1255 (2012) 267–276.
[33] P. Forlay-Frick, E. Van Gyseghem, K. Héberger, Y.V. Heyden, Selection of orthogonal chromatographic systems based on parametric and non-parametric statistical tests, Anal. Chim. Acta 539 (2005) 1–10.
[34] W. Nowik, M. Bonose, S. Héron, M. Nowik, A. Tchapla, Assessment of two-dimensional separative systems using the nearest neighbor distances approach. Part 2: Separation quality aspects, Anal. Chem. 85 (2013) 9459–9468.
[35] W. Nowik, S. Héron, M. Bonose, M. Nowik, A. Tchapla, Assessment of two-dimensional separative systems using nearest-neighbor distances approach. Part 1: Orthogonality aspects, Anal. Chem. 85 (2013) 9449–9458.

[36] M. Camenzuli, P.J. Schoenmakers, A new measure of orthogonality for multi-dimensional chromatography, Anal. Chim. Acta 838 (2014) 93–101.

[37] N. Wilson, M. Nelson, J. Dolan, L. Snyder, R. Wolcott, P. Carr, Column selectivity in reversed-phase liquid chromatography: I. A general quantitative relationship, J. Chromatogr. A 961 (2) (2002) 171–193.

[38] L. Snyder, J. Dolan, P. Carr, The hydrophobic-subtraction model of reversed-phase column selectivity, J. Chromatogr. A 1060 (1–2) (2004) 77–116.

[39] G. Guiochon, N. Marchetti, K. Mriziq, R.A. Shalliker, Implementations of two-dimensional liquid chromatography, J. Chromatogr. A 1189 (2008) 109–168.

[40] R. Agrawal, S. Belemkar, C. Bonde, Orthogonal separations in reversed-phase chromatography, Chromatographia 81 (4) (2018) 565–573.

[41] J.M. Davis, D.R. Stoll, P.W. Carr, Dependence of effective peak capacity in comprehensive two-dimensional separations on the distribution of peak capacity between the two dimensions, Anal. Chem. 80 (2008) 8122–8134.

[42] M. Gray, G.R. Dennis, P. Wormell, R.A. Shalliker, P. Slonecker, Two dimensional reversed-phase-reversed-phase separations Isomeric separations incorporating C and carbon clad zirconia 18 stationary phases, J. Chromatogr. A 975 (2002) 285–297.

[43] M. Dumarey, R. Put, E.V. Gyseghem, Y.V. Heyden, Dissimilar or orthogonal reversed-phase chromatographic systems: A comparison of selection techniques, Anal. Chim. Acta 609 (2008) 223–234.

[44] M. Dumarey, Y.V. Heyden, S.C. Rutan, Evaluation of the identification power of RPLC analyses in the screening for drug compounds, Anal. Chem. 82 (2010) 6056–6065.

[45] M. Gilar, J. Fridrich, M.R. Schure, A. Jaworski, Comparison of orthogonality estimation methods for the two-dimensional separations of peptides, Anal. Chem. 84 (20) (2012) 8722–8732.

[46] M.R. Schure, J.M. Davis, Orthogonal separations: Comparison of orthogonality metrics by statistical analysis, J. Chromatogr. A 1414 (2015) 60–76.

[47] V. Cuzuel, A. Sizun, G. Cognon, I. Rivals, F.O. Heulard, D. Thiébaut, J. Vial, Human odor and forensics. Optimization of a comprehensive two-dimensional gas chromatography method based on orthogonality: How not to choose between criteria, J. Chromatogr. A 1536 (2018) 58–66.

[48] R.K. Lindsey, B.L. Eggimann, D.R. Stoll, P.W. Carr, M.R. Schure, J.I. Siepmann, Column selection for comprehensive two-dimensional liquid chromatography using the hydrophobic subtraction model, J. Chromatogr. A 1589 (2019) 47–55.

[49] M.R. Schure, Quantification of resolution for two-dimensional separations, J. Microcolumn Sep. 9 (3) (1997) 169–176.

[50] R.W. Taft, J.-L.M. Abboud, M.J. Kamlet, M.H. Abraham, Linear solvation energy relations, J. Solut. Chem. 14 (1985) 153–186.

[51] R. Kaliszan, QSRR: quantitative structure-(chromatographic) retention relationships, Chem. Rev. 107 (7) (2007) 3212–3246.

[52] M.R. Schure, The dimensionality of chromatographic separations, J. Chromatogr. A 1218 (2011) 293–302.

[53] N. Wilson, M. Nelson, J. Dolan, L. Snyder, P. Carr, Column selectivity in reversed-phase liquid chromatography: II. Effect of a change in conditions, J. Chromatogr. A 961 (2) (2002) 195–215.

[54] Z. Liu, J.P. Foley, Are two liquid chromatography columns in tandem better than one?: Answers from the hydrophobic subtraction model, J. Chromatogr. A 1668 (2022) 462890.

[55] Z. Liu, Y. Zhou, Q. Wang, J.P. Foley, D.R. Stoll, J.G. Shackman, Development of tandem-column liquid chromatographic methods for pharmaceutical compounds using simulations based on hydrophobic subtraction model parameters, J. Chromatogr. A 1695 (2023) 463925.

[56] Z. Liu, J.P. Foley, J.G. Shackman, Q. Wang, Y. Zhou, A simulation-guided approach to the selection of RPLC columns for platform small molecule separations, J. Chromatogr. A 1730 (2024) 465131.

[57] D. Stoll, HPLC column selectivity database, 2022, URL: http://www.hplccolumns.org/database/index.php.

[58] U.S. Pharmacopoeial Convection, Column selectivity database, 2022, URL: https://apps.usp.org/app/USPNF/columns.html.

[59] J.J. Gilroy, J.W. Dolan, L.R. Snyder, Column selectivity in reversed-phase liquid chromatography: IV. Type-B alkyl-silica columns, J. Chromatogr. A 1000 (1–2) (2003) 757–778.

[60] J.C. Giddings, Use of multiple dimensions in analytical separations, in: Multidimensional Chromatography, CRC Press, 2020, pp. 1–28.

[61] H. Schmidt-Traub, M. Schulte, A. Seidel-Morgenstern, H. Schmidt-Traub, Preparative Chromatography, Wiley Online Library, 2012.

[62] A. Staby, A.S. Rathore, S. Ahuja, Preparative Chromatography for Separation of Proteins, John Wiley & Sons, 2017.

[63] S.C. Moldoveanu, V. David, Essentials in Modern HPLC Separations, Elsevier, 2022.

[64] B.W. Pirok, A.F. Gargano, P.J. Schoenmakers, Optimizing separations in online comprehensive two-dimensional liquid chromatography, J. Sep. Sci. 41 (1) (2018) 68–98.

[65] Python Software Foundation, Python documentation, 2023, URL: https://docs.python.org/3.11/.

[66] Pandas development team, pandas.DataFrame.corr, 2023, URL: https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.corr.html.

[67] SciPy Developers, scipy.spatial.ConvexHull, 2022, URL: https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.ConvexHull.html.

[68] Process Systems Enterprise, gPROMS ModelBuilder version 7.1, 2022.

[69] F. Duanmu, D.N. Chia, E. Sorensen, A novel stochastic optimization software for the optimal design of chemical processes modeled in commercial simulation software, in: Foundation of Computer-Aided Process Operations and Chemical Process Control, Texas, US, 2023.

[70] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Trans. Evol. Comput. 6 (2) (2002) 182–197.