Journal of Chromatography A 1730 (2024) 465128

Contents lists available at ScienceDirect

Journal of Chromatography A

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

Analysis and classification of tea varieties using high-performance liquid chromatography and global retention models

P. Peiró-Vila, C. Pérez-Gracia, J.J. Baeza-Baeza, M.C. García-Alvarez-Coque, J.R. Torres-Lapasió *

Department of Analytical Chemistry, Faculty of Chemistry, Universitat de València, C/ Dr. Moliner 50, Burjassot 46100, Spain

ARTICLE INFO

Keywords:
Multi-analyte samples
Medicinal plants
Multi-linear gradient elution
Global retention models
Classification

ABSTRACT

As a result of their metabolic processes, medicinal plants produce bioactive molecules with significant implications for human health, used directly for treatment or for pharmaceutical development. Chromatographic fingerprints with solvent gradients authenticate and categorise medicinal plants by capturing chemical diversity. This work focuses on optimising tea sample analysis in HPLC, using a model-based approach without requiring standards. Predicting the gradient profile effects on full signals was the basis to identify optimal separation conditions. Global models characterised retention and bandwidth for 14 peaks in the chromatograms across varied elution conditions, facilitating resolution optimisation of 63 peaks, covering 99.95 % of total peak area. The identified optimal gradient was applied to classify 40 samples representing six tea varieties. Matrices of baseline-corrected signals, elution bands, and band ratios, were evaluated to select the best dataset. Principal Component Analysis (PCA), *k*-means clustering, and Partial Least Squares-Discriminant Analysis (PLS-DA) assessed classification feasibility. Classification limitations were found reasonable due to tea processing complexities, involving drying and fermentation influenced by environmental conditions.

1. Introduction

Traditional medicine includes medical approaches derived from ancestral beliefs and collective wisdom about the effects of natural substances on health, predating modern medicine [1]. Plants are commonly used because they are more affordable than modern medicines, making them invaluable for healing, especially in non-industrialised societies. *Camellia sinensis*, commonly referred to as tea, is a well-known medicinal plant and one of the most consumed beverages worldwide after water [2]. Its popularity stems from its flavour, stimulating qualities, and therapeutic substances. Tea's use has historical roots in myths and traditions [3,4].

The primary feature distinguishing tea varieties is the leaf fermentation level [5,6]. Preparation steps like drying and crushing also influence oxidation. The most common types are white, green, black, red, and oolong (blue) tea [7]. They can be marketed individually or mixed with other botanicals like ginger, pineapple, or vanilla for enhanced flavour. Green tea is made from partially fermented dried leaves, while black tea results from intensive oxidation. Red tea comes from fermenting *Camellia sinensis assamica* in humid environments, driven by bacteria. White tea is derived from young shoots and leaves quickly dried to inhibit fermentation and requires reduced sunlight to minimise chlorophyll production. Green and white teas have higher bioactive compound concentrations due to milder processing. Oolong tea undergoes partial fermentation, sharing properties of both green and black teas, and offering a range of flavours and aromas. Other plants processed like tea include *Jasonia glutinosa* for rock tea and *Aspalathus linearis* for rooibos tea, which is naturally caffeine-free [8].

The beneficial properties of tea depend on its chemical composition, primarily involving four groups: alkaloids, glycosides, polyphenols, and terpenes [9]. The quality of products derived from medicinal plants is often assessed by combining separation and detection techniques to identify key components. However, determining a limited number of compounds does not fully characterise the product. An alternative approach is using "chromatographic fingerprints" (CFs) [10–13]. A CF is a chromatogram showing the constituents as multiple peaks of varying heights, creating distinct patterns with different degrees of overlap. In a CF, the sample is characterised by the entire pattern of peaks in the chromatogram, not just a few selected peaks. Samples with similar chemical compositions show similar CFs. Their major advantage over

* Corresponding author.
E-mail address: jrtorres@uv.es (J.R. Torres-Lapasió).

https://doi.org/10.1016/j.chroma.2024.465128
Received 7 May 2024; Received in revised form 21 June 2024; Accepted 28 June 2024
Available online 29 June 2024

individual compound characterisation is that they do not require standards or identification of all compounds.

Tea substances are typically analysed by high-performance liquid chromatography (HPLC) [14,15]. Natural samples like tea have a wide array of analytes with diverse chemical properties, including variations in polarity and concentration. Achieving sufficient separation for less retained solutes, while ensuring reasonable retention times for more retained ones is challenging [16]. Therefore, when analysing plant extracts, gradient elution with variable modifier content is essential.

Chromatographic signals in CFs are readily obtained and processed using chemometric classification techniques. These techniques help establish relationships between chemical variables and sample attributes, making CFs effective chemical descriptors for tea. The CF information content depends on both the number of visible peaks (representing sample components) and their separation degree. Thus, identifying experimental conditions that maximise visible peaks and resolution is necessary for classification success.

This work has two key objectives. The primary goal is to develop a practical method for determining the optimal separation gradient for a CF, relying on modelling complete chromatograms using the global modelling approach [17–20]. Once the optimal gradient is identified, it is applied to classify 40 tea samples from the Spanish market representing six tea varieties. This involves constructing datasets (baseline-corrected signals, elution bands, and band ratios) and using pre-processing and classification methods like *k*-means clustering and Partial Least Squares combined with Discriminant Analysis (PLS-DA) to distinguish between tea categories and identify the category of unknown samples. By maximising constituent separation, the optimal gradient is expected to significantly enhance classification outcomes.

2. Methods

2.1. Prediction of retention based on global models

Classical interpretive optimisation methodologies require fitting each compound in the sample to a dedicated retention model based on data from several experiments. Typically, these experiments involve standards, which are injected individually or in groups, under varying elution conditions that influence the migration process. This approach requires a comprehensive knowledge of the nature of all constituents within the sample and the accessibility to standards for model development. However, these pre-requisites are often unmet when dealing with natural samples [21,22].

In prior research, a methodology was introduced for simulating complex chromatograms by creating a common (global) retention model for all components [17–19], based on the Neue-Kuss equation [23]. This model is built using peak data derived from a limited set of selected constituents known as "reference peaks", whose nature may or may not be known. The methodology is then extended to all compounds giving visible peaks in the CFs. In previous work, it was successfully applied and validated using extracts from diverse medicinal plants with different characteristics, including tea (*Camellia sinensis*), linden (*Tilia cordata*), mint (*Mentha piperita*), lemon balm (*Melissa officinalis*), and chamomile (*Matricaria chamomilla*). In this work, we follow the same strategy, which is outlined in the Supplementary material.

2.2. Gradient optimisation

The information content of a CF relies on the number of detected peaks and their resolution. Therefore, optimising both is essential. In this work, we have used the elementary peak purity (p_i) as chromatographic objective function (COF) to optimise resolution [24]. Peak purity is the peak area fraction unaffected by interferences from other peaks. After predicting the chromatogram under specific conditions using suitable models, peak purities are computed, resulting in a vector $\boldsymbol{r} = [p_1, p_2, \ldots p_{ns}]$. These p_i measurements collectively contribute to a global value, representing the overall purity of the entire chromatogram (P) [25]. This resolution metric guided the optimisation algorithm in identifying the most suitable gradient program. More details on the optimisation process can be found in Ref. [20].

The experimental design for generating datasets to fit the global models consisted of seven multi-linear gradients. Each gradient had two consecutive linear segments connected by a node of variable position in a rhomboidal space (Fig. 1) [26]. The optimised parameters were the coordinates (gradient time and organic solvent content) of that node. Though simple enough for a grid search optimisation, we used a program for optimising multi-node problems to ensure sufficient resolution, which was uncertain before the search.

The optimal gradient profile for the classification study was selected using a classical Genetic Algorithm (GA) [20,27]. A random population of 100 gradients was generated within the specified search space (Fig. 1). These gradients competed and evolved across subsequent generations, following rules akin to natural selection (100 % crossing-over probability, 6 % mutation probability, binary encoding, 5 % probability of refreshing with the global best, maximum allowed generations: 500). Each gradient's resolution performance was evaluated and stored, considering global resolution measured as the sum of peak purities and analysis time for each simulated chromatogram.

2.3. Data processing

Considerable interest has been focused on similarity analysis of CFs for identifying and ensuring the quality of herbal samples. These methods are essential for verifying if a sample originates from the expected herb and excluding other sources. Various methods have been used to assess the similarity or difference between CFs [28,29]. Before conducting any classification study, exploring the dataset structure is important to detect patterns, groupings, and potential outliers or experimental errors. Principal Component Analysis (PCA) is commonly used for this exploratory analysis [30,31]. While PCA serves to reduce dimensionality, it is not intended for classification. Therefore, following PCA data analysis, dedicated classification techniques were applied to categorise the samples [32–34].

Two representative classification techniques were utilised: *k*-means clustering [35] as an unsupervised approach and PLS-DA [36,37] as a supervised approach (details in Supplementary material). Chromatographic signals underwent three treatments to create varied data matrices before sample classification:

(i) *Treatment T1* involved creating a chromatographic signal matrix (**S**) after baseline subtraction [38]. To optimise computer processing capabilities, the acquisition frequency was reduced. Consequently, a 40×2772 matrix was constructed, containing absorbance measurements from the 40 samples (including five replicates). The matrix excluded regions below 2 min (dead time) and beyond 39 min.

(ii) *Treatment T2* involved creating an elution band matrix (**B**) focused on specific time domains corresponding to equivalent peaks. This method assumes each band encompasses peak maxima of identical constituents across all samples, accounting for their temporal variations. For each band $\mathbf{B}_i$, the highest signal value within its time domain was considered representative. This approach effectively addresses peak shift impacts.

(iii) *Treatment T3* involved constructing a matrix (**Q**) using ratios of elution bands [39]. It included all possible pair combinations like B_i/B_j and B_j/B_i, excluding unit ratios (B_i/B_i).

In the classification study, the three datasets (**S, B**, and **Q** matrices) underwent: (i) centring based on the mean, and (ii) autoscaling (z transformation). This ensured comparability either of means alone (centred data) or both dispersion and means (autoscaled data) across all variables. PCA was applied to each dataset as pre-processing method to

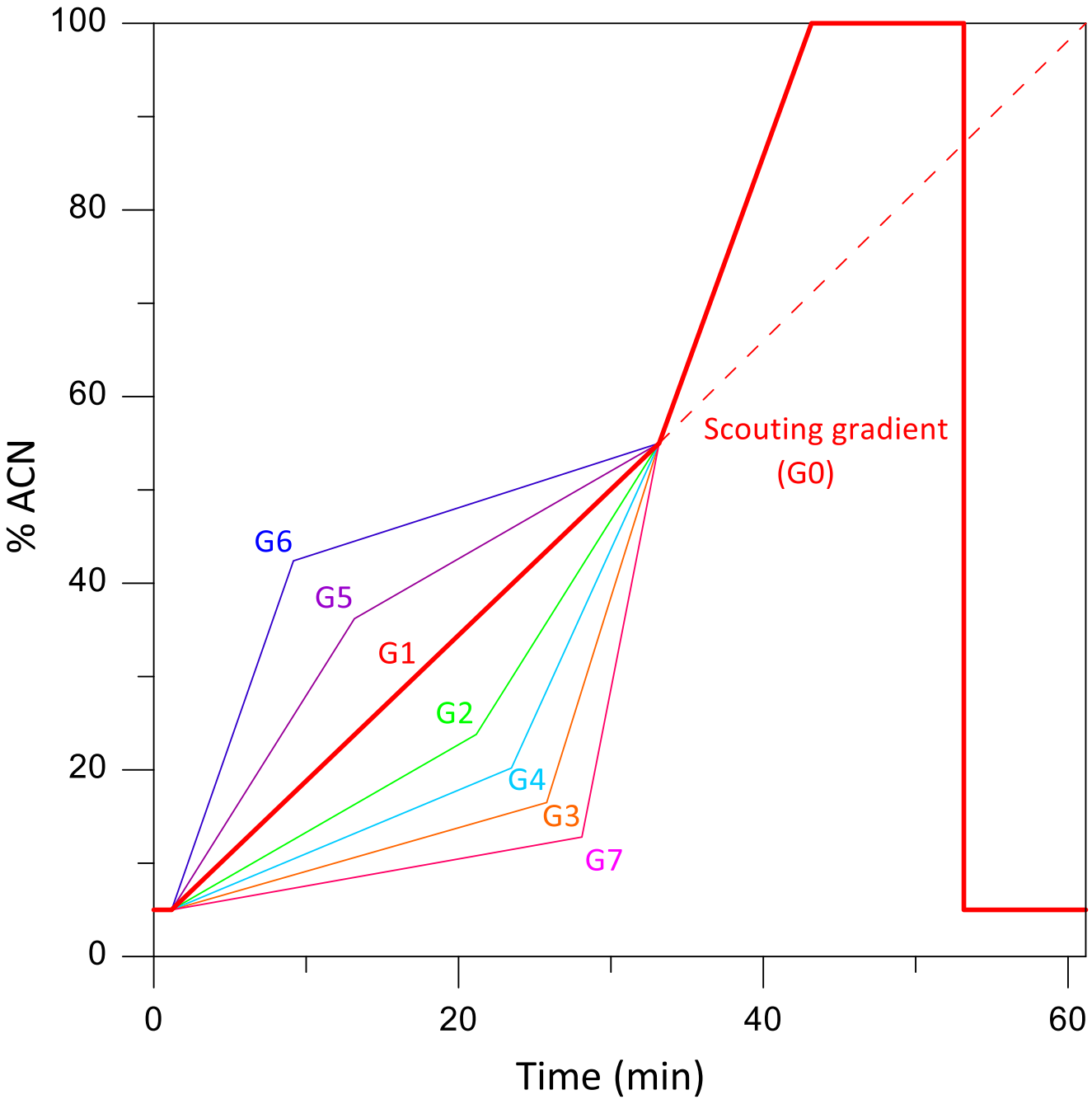

Fig. 1. Training design featuring gradients with one node at a variable position used to model the retention of tea components. The design evolved from a scouting gradient incorporating a node whose position was changed along a diagonal path, resulting in two consecutive intermediate linear segments with different slopes. After the ending node of the bilinear sector at 32 min, the concentration was raised to 100 % (*v/v*) acetonitrile, and kept constant for 10 min to elute highly hydrophobic compounds. The acetonitrile concentration was then returned to 5 % (*v/v*) and the column re-equilibrated for 25 min.

reveal the data's underlying structure and identify associations between samples and variables. PCA also helped rank the effectiveness of treatments and datasets, aiding in the selection of the most suitable for classification purposes.

3. Experimental and software

3.1. Sample and extraction protocol

The samples included common brands from local Spanish markets, covering green, red, black, white, oolong (or blue), and rooibos teas. These teas comprised crushed, dried botanicals, often in individual infusion sachets. Table 1 lists the analysed samples, categorised into six groups. Obtaining diverse samples, especially for oolong and rooibos teas, was challenging due to limited availability. Some samples were duplicated to check reproducibility.

Several extraction conditions for tea components were explored through trial and error, guided by previous studies [40]. Initially, 1 g of samples was extracted using 30 % methanol (HPLC-grade from Scharlab, Barcelona, Spain). The largest signals were too high, indicating a need to reduce the sample size. Reducing the sample to 0.2 yielded signals of 3200 mAU. Increasing methanol to 70 % did not significantly improve leaching under the applied sonication time and temperature. The extraction protocol is detailed in the Supplementary material.

Table 1
Tea samples classified according to their varieties.

Type	Brand	Code[a]	Purity, %	Additives
White	Carrefour (×2)	w_1	100	
	Hornimans	w_2	<100	Vanilla scent
	Herboristería Navarro	w_3	100	
Green	Consum (×2)	g_1	100	
	Ship	g_2	-[b]	
	Hacendado	g_3	-[b]	
	Hornimans	g_4	<100	Lemon
	Matcha	g_5	100	
	Hornimans, Slim	g_6	< 100	Guaraná, pineapple and matcha
	Chai	g_7	80	Peppermint
	Gourmet	g_8	100	
	Lord Nelson	g_9	-[b]	
	Pompadour [c]	g_{10}	100	
	Westminster	g_{11}	<100	Vanilla
	Herboristería Navarro	g_{12}	100	
Oolong	Artemis Bio	o_1	100	
	High Mountain Ten Ren Tea	o_2	<100	Cinnamon
Rooibos	Codonyer & March	s_1	100	
	Herbalism	s_2	-[b]	
	Hornimans	s_3	<100	Raspberry
Black	Consum (×2)	b_1	100	
	Ship (×2)	b_2	-[b]	
	Darjeeling	b_3	-[b]	
	PG Tips	b_4	-[b]	
	Gourmet	b_5	100	
	Pompadour	b_6	100	
	Chai	b_7	35	Zingibere, cinnamon and scents
	Lord Nelson	b_8	-[b]	
	Herboristería Navarro	b_9	100	
Red	Hacendado	r_1	-[b]	
	Consum	r_2	100	
	Hornimans (×2)	r_3	100	
	Pompadour	r_4	< 100	Pineapple and lemon
	Gourmet	r_5	100	
	Herboristería Navarro	r_6	100	

[a] Colours correspond to the sample labels in Figs. 5 to 8.
[b] The manufacturer does not explicitly indicate if it is a mixture or a pure product.
[c] Sample taken as representative to optimise the separation, since it presents most peaks that can be found in other samples.

3.2. Apparatus, column, and chromatographic conditions

The chromatographic signals were acquired with an instrument from Agilent Technologies (Waldbronn, Germany) with these modules: quaternary pump (1260 Series), autosampler (1200 Series) with 2 mL vials, temperature column-control compartment (1200 Series), and variable wavelength UV–visible detector (1260 Series). The system was governed by an Agilent OpenLAB CDS LC ChemStation (version C.01.08). UV detection was chosen for detecting the primary constituents of tea. A scan in 10 nm increments within the 210–280 nm range was carried out.

A Zorbax Eclipse XDB-C18 column (150 mm × 4.6 mm, 5 µm particle size, from Agilent) was used for chromatographic separation, protected with a Spherisorb C18 pre-column (30 mm × 4.6 mm, 5 µm particles, from Scharlab). Injections were conducted at 25 °C with a 10 µL injection volume and a flow rate of 1.0 mL/min.

The analyses used acetonitrile gradients generated by mixing 5 % acetonitrile (HPLC-grade from Panreac, Barcelona) with pure organic solvent, both containing 0.1 % formic acid (Acros Organics, Fair Lawn, NJ) to achieve a pH near 3.0. Both eluents were vacuum-filtered through 0.45 µm Nylon membranes. KBr and uracil (Acros Organics, Geel, Belgium) were used to determine dead and extra column times. An acetone gradient (Scharlab) determined dwell time. Nanopure water was obtained using an Adrona B30 Trace purification system (Burladingen, Germany).

3.3. Software

To measure peak properties like retention times and half-widths, perform data treatment, and predict chromatograms, in-house functions (Chromscan Automatic Analysis) were created using Matlab 2022a (The MathWorks Inc., Natick, MA) [40]. Baseline subtraction was done using the assisted BEADS algorithm [38]. Quantitative information for all detected peaks was stored for further processing.

The *k*-means function used the Statistics and Machine Learning Toolbox in Matlab 2020a, and PLS-DA was implemented with the PLSDAGUI tool in Matlab 2020a [41].

4. Results and discussion

4.1. Selection of a representative sample and modelling of retention

To achieve optimal classification, elution conditions maximising chemical information are essential. Initial tests were conducted to identify the optimal experimental separation using a multi-linear gradient and detection conditions. Pompadour green tea (sample g_{10} in Table 1), for its minimal processing which resulted in the highest number of peaks among all samples, was selected as representative for sampling the presence of most peaks found in other tea samples.

The optimal wavelength was selected to maximise peak visibility, ensure a clean baseline for easy removal, and maintain an acceptable noise level. Given these conflicting criteria, a compromise was necessary in wavelength selection. Across the studied range (210–280 nm), fewer peaks were observed at higher wavelengths, while shorter wavelengths showed increased baseline noise, drifts, and noticeable discontinuities. Fig. 2 shows chromatograms of the Pompadour sample at eight recorded wavelengths using the scouting gradient (G1 in Fig. 1). Peak intensity notably decreased beyond 250 nm, and shorter wavelengths exhibited decreased signal stability with increased noise. A compromise wavelength of 230 nm was chosen, effectively preserving signals while facilitating baseline removal and minimal noise.

The steps involved to create the global model (see Section 2.1 and Supplementary Material) were as follows:

(i) Selection of 14 reference peaks traceable with the seven training multi-linear gradients outlined in Fig. 1.
(ii) Measurement of peak properties (retention times, half-widths and peak areas) for the reference compounds, using the Chromscan Automatic Analysis function developed in our laboratory [40].
(iii) Fitting the retention data of reference peaks to the restricted global model.
(iv) Creation of an extended global model covering 158 peaks detected with the slowest gradient.
(v) Optimisation by GA to enhance resolution for the 63 most intense peaks (representing 99.95 % of total peak area) within feasible analysis times and the search space restricted to the limits of the experimental design.

Model parameters were derived through two methods: (i) fitting each solute individually to a dedicated model, and (ii) fitting all solutes collectively to a global model. Initial parameter estimates for individual fittings were $\log k_0 = 1$, $b = 50$, and $c = 5$. For the global model, initial estimates used $\log k_0$ values from individual fittings and median values of b and c across all solutes. The search was constrained with limits: [−1, 10] for $\log k_0$, [0, 300] for b, and [−1, 10] for c, without additional hyper-parameters. Table S1 in the Supplementary material presents regression parameters for models describing the retention of reference peaks using all gradients in the experimental design. Mean relative prediction errors with respect to the average retention were 0.27 % for the individual models and 0.80 % for the global models. This demonstrates the reliability of chromatogram simulation using the global model approach.

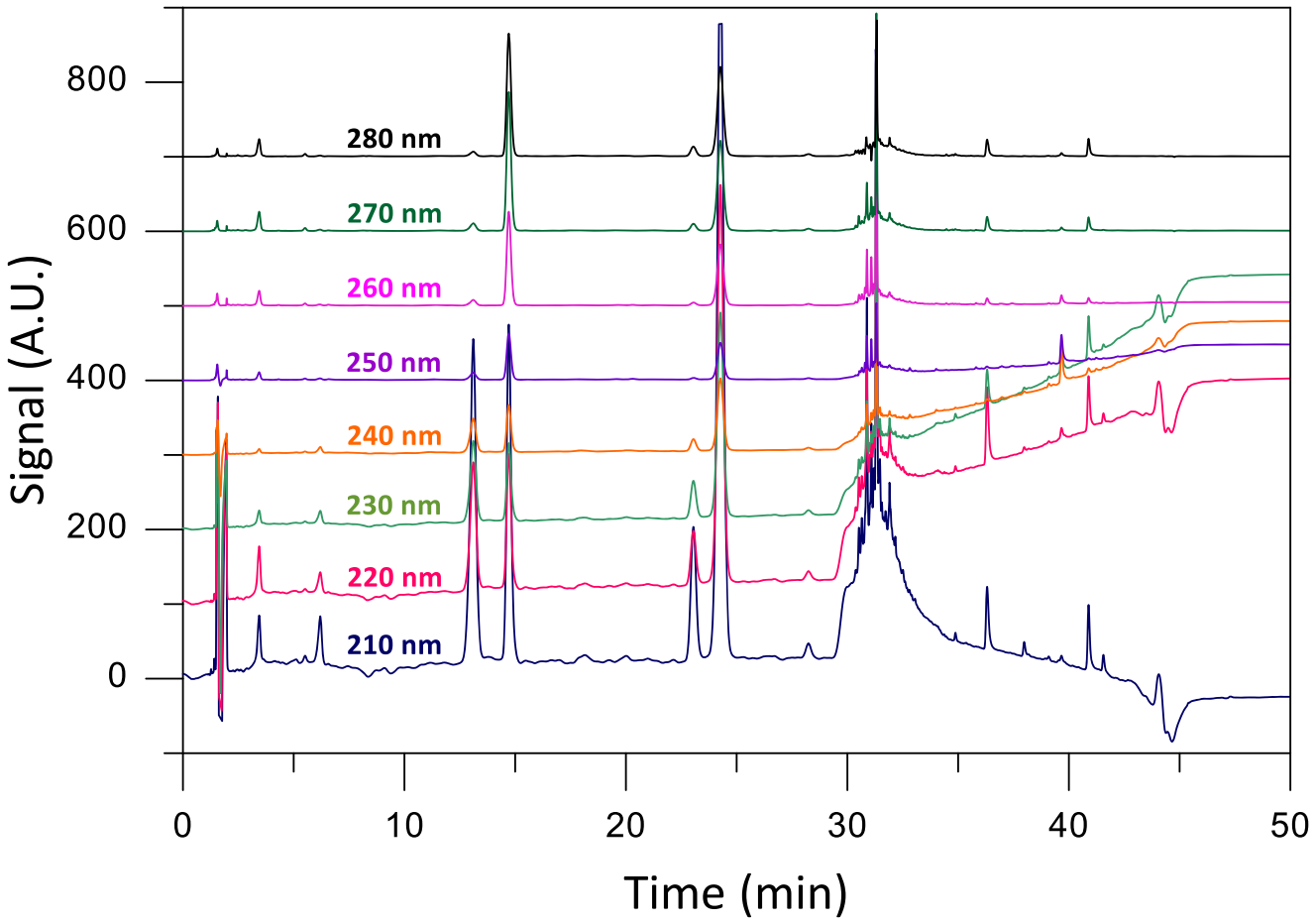

Fig. 2. Chromatogram of an extract of green tea (sample g_{10} in Table 1) at varying detection wavelengths. The sample underwent elution using the scouting gradient (5–100 % (*v/v*) acetonitrile in 60 min, see Fig. 1).

4.2. Search of the optimal gradient

Fig. 3 displays the results of the GAs search as a Pareto plot, where each point illustrates the performance of a gradient within the experimental design space, throughout the search process. Gradients whose resolution cannot be improved without compromising the analysis time, forming the Pareto front [27], are highlighted in red. A specific gradient marked with a yellow dot denotes near-maximal resolution and an appropriate analysis time. This selected gradient was used for analysing all tea samples.

Fig. 4 depicts the chromatogram predicted by the global model for the optimal gradient in Fig. 3 (highlighted in green). The experimental chromatogram obtained with the same gradient is displayed in blue, demonstrating excellent agreement between the predicted and actual separations.

4.3. Analysis of the extracts of the whole set of samples

After identifying the optimal gradient (see Section 4.2) for sample g_{10}, we analysed the remaining samples listed in Table 1 using the same gradient (see Fig. 4). Fig. 5 shows baseline-subtracted chromatograms for the samples categorised in Table 1, organised by their respective classes. All samples underwent identical extraction conditions (see Section 3.1).

The samples of the six tea varieties were classified using the three treatments outlined in Section 2.3 (T1 to T3). Fig. 5 displays grey rectangles representing elution bands surrounding the 39 most significant peaks. Each rectangle is strategically defined to encompass all observed peak shifts across samples, ensuring preservation of constituent information despite incidental shifts. Regions with no useful classification information, such as those with zero signals, are excluded. A similar approach was applied to several peaks at the chromatogram end (marked with asterisks in Fig. 5), likely due to injector contamination by bacteria from previous analyses using citric buffer. These peaks could potentially be eliminated following a manufacturer-provided protocol for injector cleaning, but we found it interesting to show them. To enhance visibility of less intense peaks, chromatograms were vertically shifted and some signal maxima were clipped.

Certain peaks in chromatograms are characteristic of each tea variety, consistently appearing in all samples of that type. Others appear sporadically, present in some samples of a specific variety and occasionally in other types of tea. Furthermore, some peaks are shared across several varieties, while others are unique to specific types of tea. However, exceptions exist, limiting generalisations.

Acknowledging the variability in chromatograms among extracts of the same tea type is important, as it tends to attenuate differences between classes, making classification studies complex. Several factors contribute to this variability:

(i) Chromatographic profiles are influenced by tea origin, variety, and chemical treatment, considering factors, such as different plants, soils, cultivation practices, storage conditions, and processing procedures, among others.
(ii) Injections of the extracts in the HPLC instrument were conducted over several weeks, necessitating fresh solution preparation and extractions under consistent conditions to ensure stability of both the instrument and chemical components in the solutions and mobile phases.

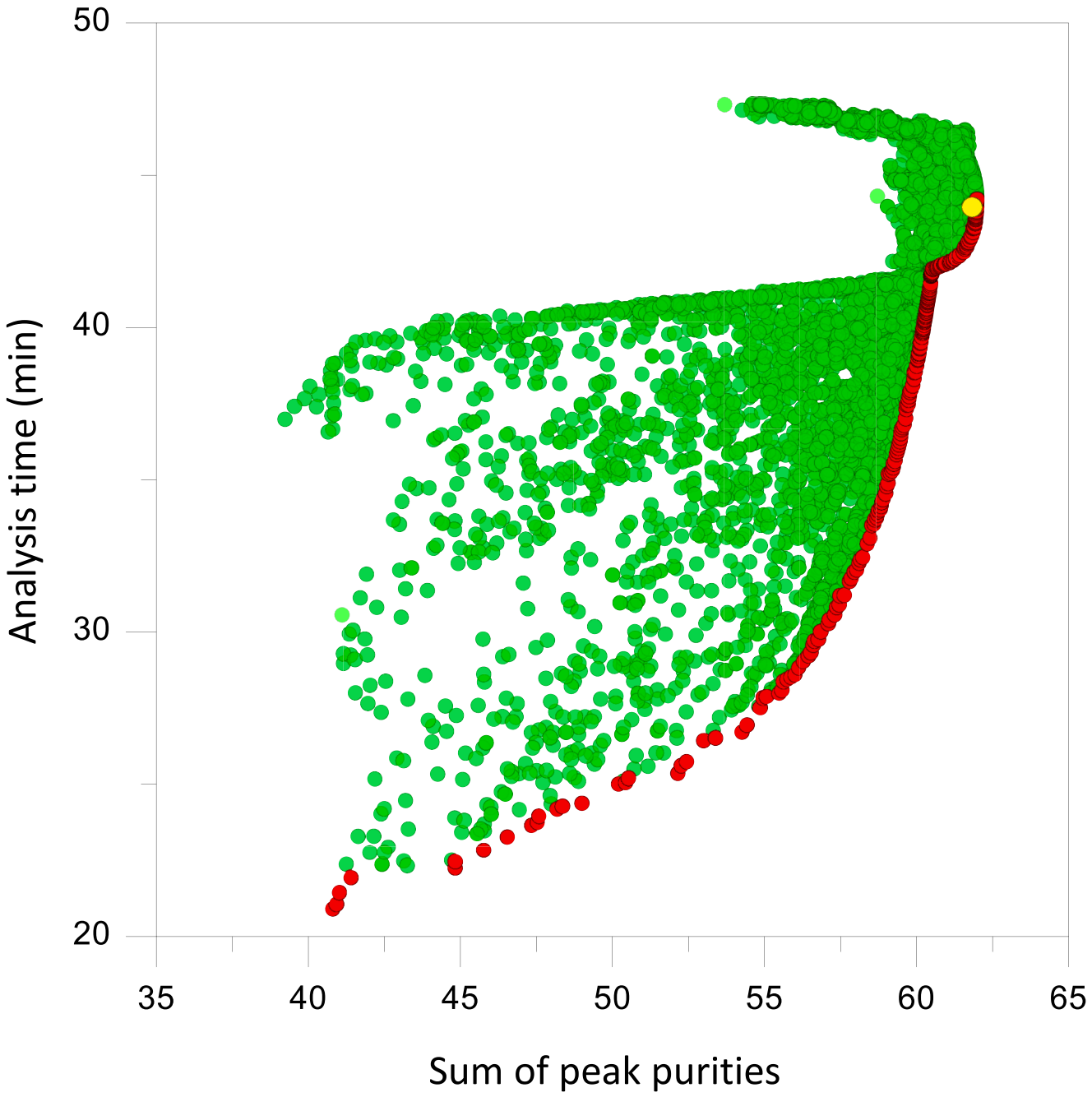

Fig. 3. Pareto plot displaying the results of resolution optimisation for the sample g_{10}. Multi-linear gradients with one node, positioned within the rhomboid space in Fig. 1, were considered. The data include all gradients examined during the GAs search. Gradients on the Pareto front are in red, and the selected optimal gradient is in yellow. The optimisation enhanced the resolution of 63 chromatographic peaks representing over 99.95 % of the total signal area. The sum of elemental resolutions reached 61.8, which is 98.1 % of the maximum resolution (63). The retention time of the last peak in the optimal chromatogram was 44.0 min.

(iii) Despite using a relatively small sample size (0.2 g), significant signal intensities were observed for major peaks. Initially testing 1 g samples led to detector saturation. It is important to note that a portion of material above 1 g was carefully crushed and homogenised, before taking the 0.2 g sample used for analysis.

(iv) Chromatograms of tea samples can show peaks from additives such as vanilla, ginger, cinnamon, or pineapple aromas, which may not appear in other samples of the same variety, complicating classification.

(v) Besides the type of sample, extraction efficiency depends on factors like sample grinding level, humidity, and storage conditions, as well as controllable variables such as solvent mixture and temperature.

(vi) Fluctuations in peak position can occur due to random variations in equipment pressure or flow-rate, and other factors. Ideally, equivalent peaks should align across different chromatograms.

Among the three pre-processing treatments, T1 (direct chromatograms) may appear as most favourable, preserving original information despite introducing baseline distortions, but bands treatment (T2) is recommended to mitigate non-coincident signals from random position fluctuations. This approach also eliminates chromatographic regions lacking distinctive sample information, which can be detrimental for

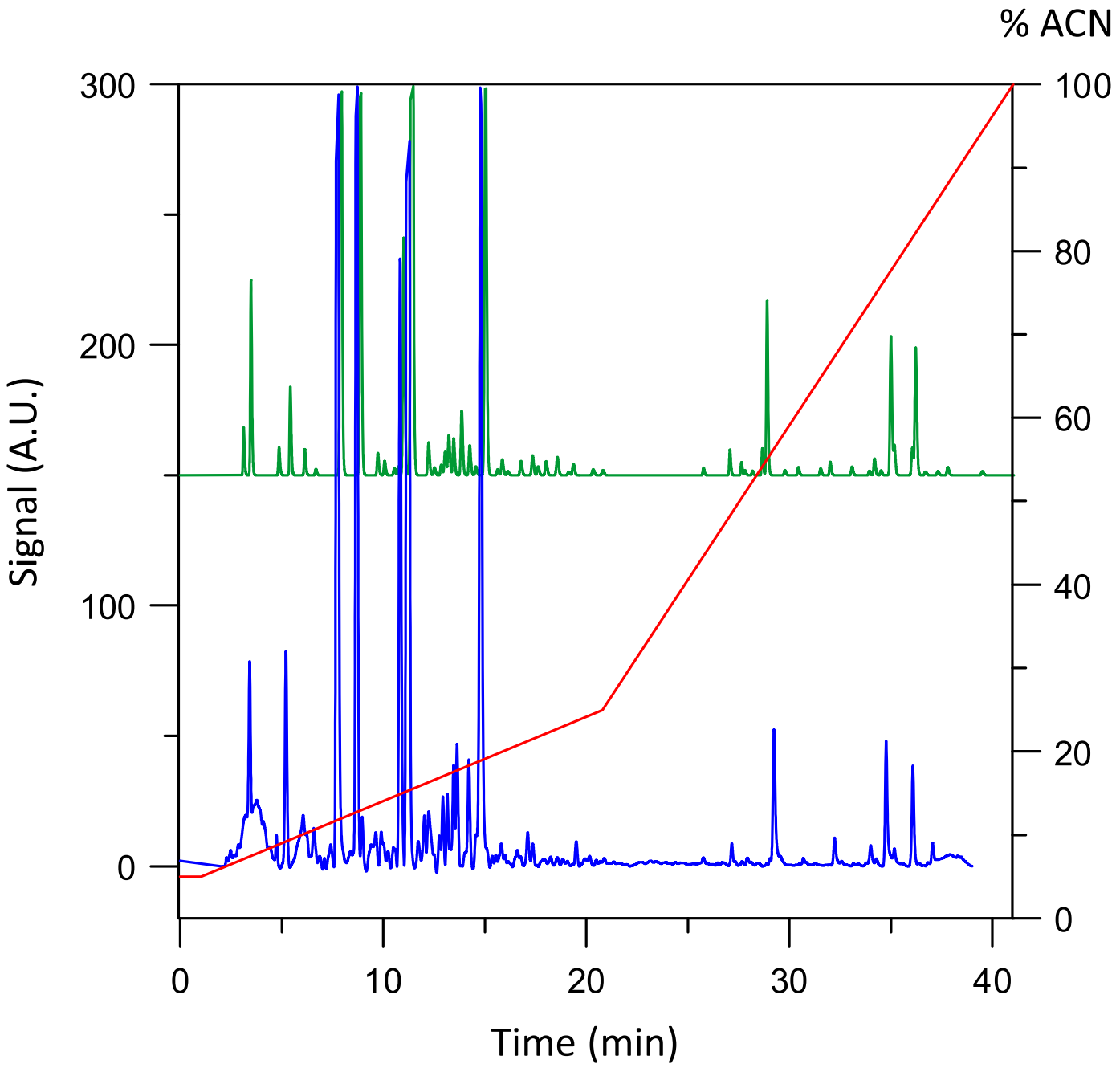

Fig. 4. Comparison between the optimal predicted chromatogram (green) for sample g_{10}, and the baseline-corrected experimental chromatogram (dark blue). The optimal gradient program is overlaid in red, corresponding to the yellow circle in the Pareto plot in Fig. 3.

classification purposes. Treatment T3 addresses variations in extraction yields by computing band ratios relative to an arbitrary peak, akin to using an internal standard for correction. However, distinguishing whether peak intensity variations stem from extraction differences or sample characteristics becomes challenging when diverse samples are included in the study. Hence, exploring all possible combinations of band ratios is crucial for identifying the most suitable peak as a correction standard.

Initially, all 39 bands shown in Fig. 5 were considered, but due to the limited size of the eight final bands and potential interference of contamination peaks, only bands below 32 min (which were free of interference) were processed, resulting in a total of 31 bands. Ultimately, the data matrices for classification included the following elements: $\mathbf{S} = 40\times2772$ (T1), $\mathbf{B} = 40\times31$ (T2), and $\mathbf{Q} = 40\times930$ (T3).

4.4. Classification assays

Matrices of baseline-corrected signals, selected elution bands, and band ratios were evaluated to identify the most suitable dataset. The results from applying PCA and classification methods like *k*-means and PLS-DA are presented next to assess the feasibility of classifications.

4.4.1. Exploration using principal component analysis

As mentioned, prior to the classification study, PCA was applied to each dataset (T1-T3 in Section 2.3). Fig. 6 depicts the projections of sample data (scores) onto the orthogonal axis system (loadings are not shown to avoid overcrowding):

(a) *Treatment T1: Baseline-corrected chromatograms*

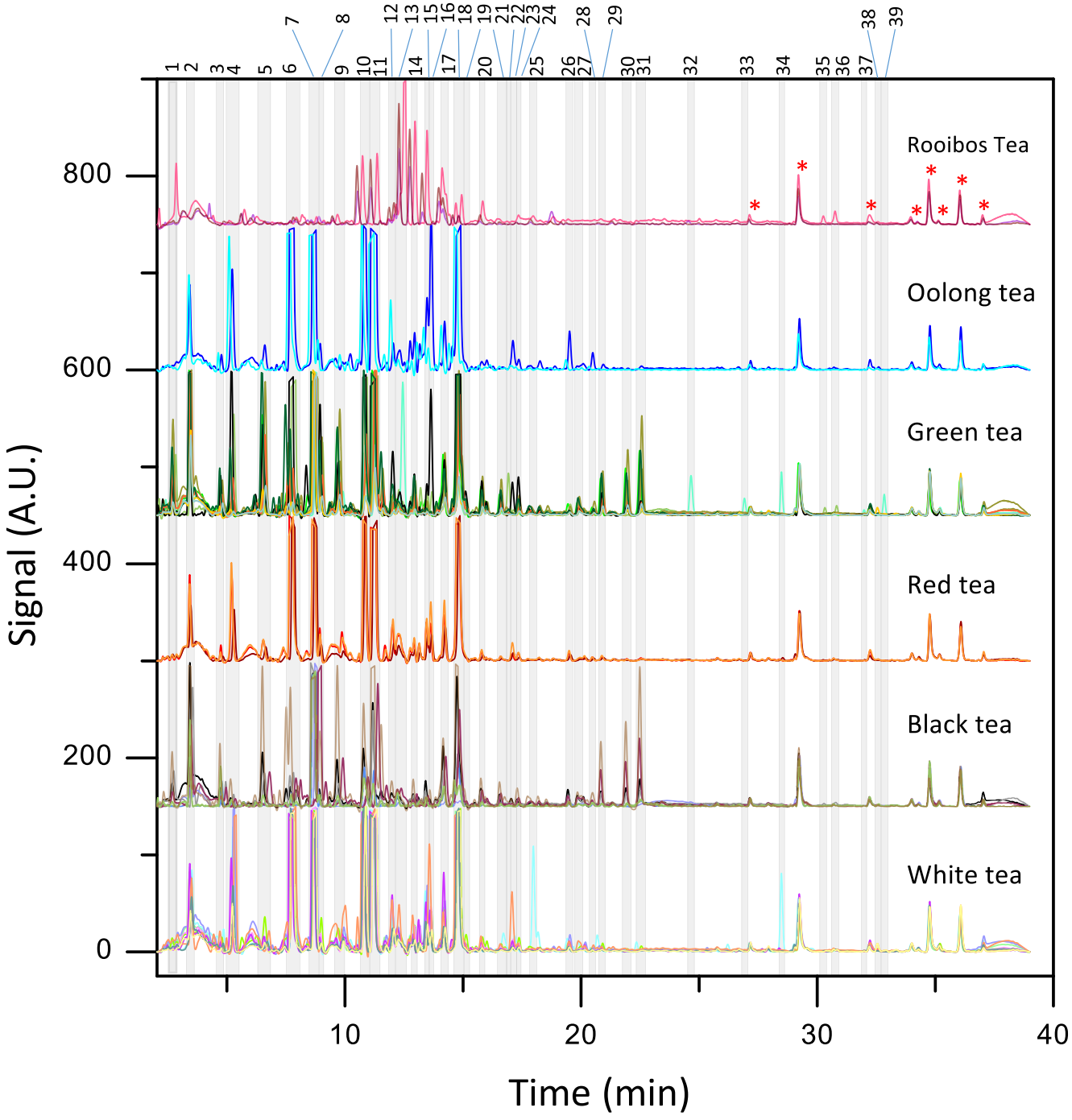

Fig. 5. Chromatograms of the tea extracts illustrating the 39 elution bands (codes above) used to generate the 40×39 **B** matrix, gathering the maximal absorbance within each band for each sample. The extracts include white ($\mathbf{w}_i$), green ($\mathbf{g}_i$), oolong ($\mathbf{o}_i$), rooibos ($\mathbf{s}_i$), black ($\mathbf{b}_i$), and red ($\mathbf{r}_i$) teas. Five samples were replicated (see Table 1).

Fig. 6a and d display the projections of the 40 tea samples (including some replicates, see Table 1) onto the first two principal PCs. In Fig. 6a, where the data were mean-centred, a clear separation is visible between two groups: a cohesive cluster consisting of red, black, and rooibos teas, and a more scattered arrangement comprising green, white, and oolong teas, with two distinct outliers (w_3 and b_3). In contrast, when the data are autoscaled (Fig. 6d), a less structured arrangement is observed without discernible separation.

(b) *Treatment T2: Elution bands*

This treatment notably improves the differentiation of distinct groups, as shown in Fig. 6b and e. PCA analysis on the **B** data matrix confirms the trends observed in Fig. 6a, but with enhancements: rooibos tea is clearly separated at the upper end of a vertical arrangement alongside black and red teas, indicating emerging differentiation. The other samples exhibit more scattering, with less pronounced class distinctions. Autoscaling the data (Fig. 6e) improves class separation, although it is still somewhat insufficient. Importantly, both plots highlight two samples (w_3 and b_3) from different classes that are completely segregated from the others, as noted in Fig. 6a. Band processing consistently yields superior results compared to original baseline-

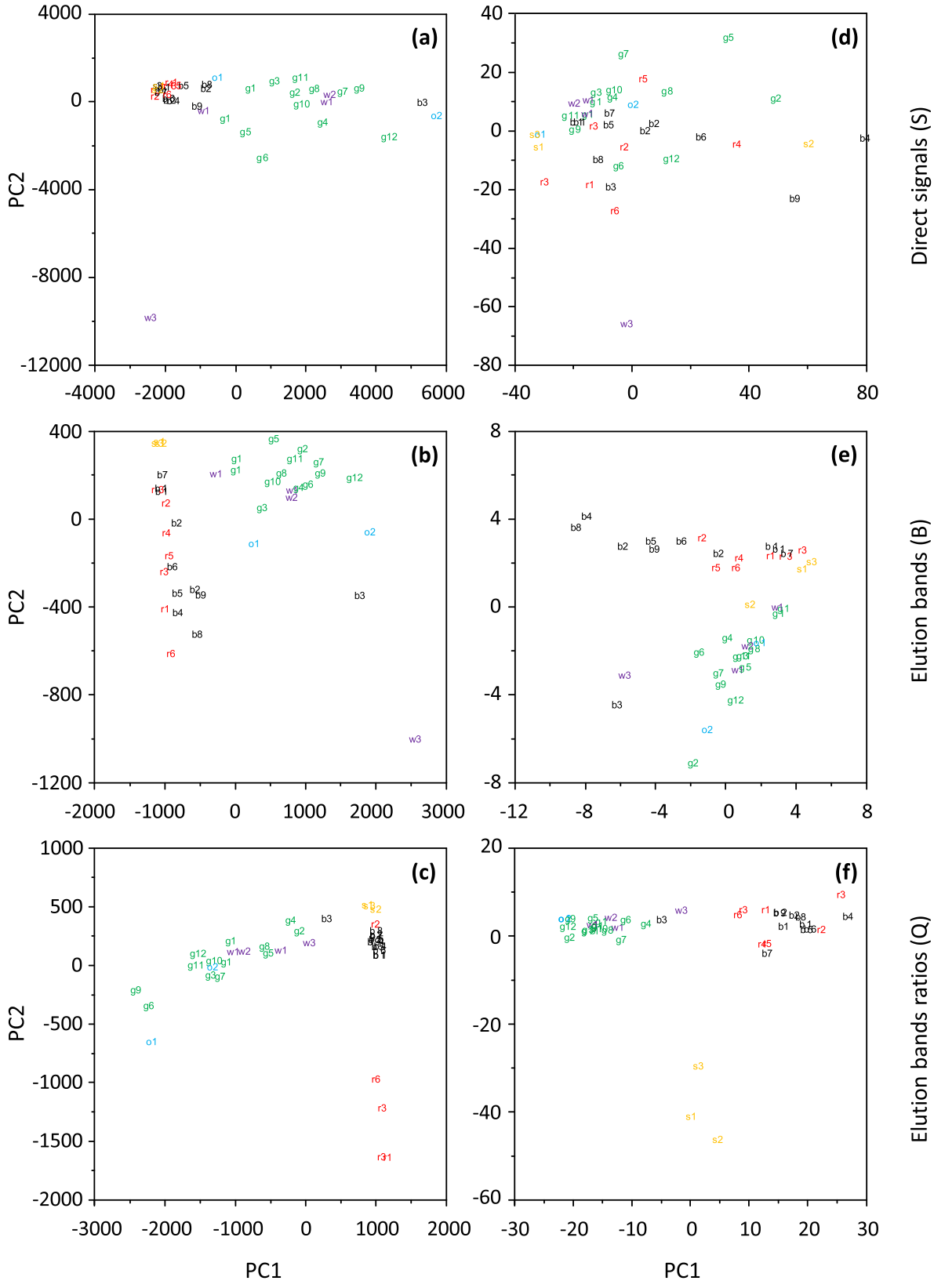

Fig. 6. PC1 and PC2 scores for the 40 tea samples (including five replicates), using the following data matrices: (a,d) **S** (signals after baseline subtraction), (b,e) **B** (elution bands), and (c,f) **Q** (elution band ratios). Pre-processing steps included data centring (a–c), and column-wise autoscaling (d–f). Refer to Table 1 for sample identification.

corrected signals for both mean-centred and autoscaled data.

(c) *Treatment T3: Band ratios*

Similar to previous treatments, distinct groups are evident, with one comprising white and green tea samples and the other consisting of black and red tea samples (Fig. 6c and f). However, the cohesion within the first group is more pronounced than in the second, particularly noticeable in Fig. 6c when considering PC2. Here, black tea samples are closely clustered, unlike in Fig. 6f where no separation from red teas is apparent. Oolong tea samples align closely with the group consisting of white and green teas in both plots, resulting in overlap with no clear distinction between them. Notably, the complete differentiation of rooibos tea from the other groups in Fig. 6c and f is noteworthy.

The division into two main groups: white, green, and oolong tea samples on one side, and red and black tea samples on the other, is rational considering the manufacturing processes:

(i) White, green and oolong teas involve natural leaf drying to prevent fermentation and withering, while white tea specifically uses younger leaves.
(ii) Black and red teas, in contrast, share significant similarity due to undergoing fermentation processes that lead to higher oxidation levels.

Rooibos tea distinguishes itself from other teas, as it does not originate from *Camellia sinensis*, although undergoes similar fermentation processes to red tea.

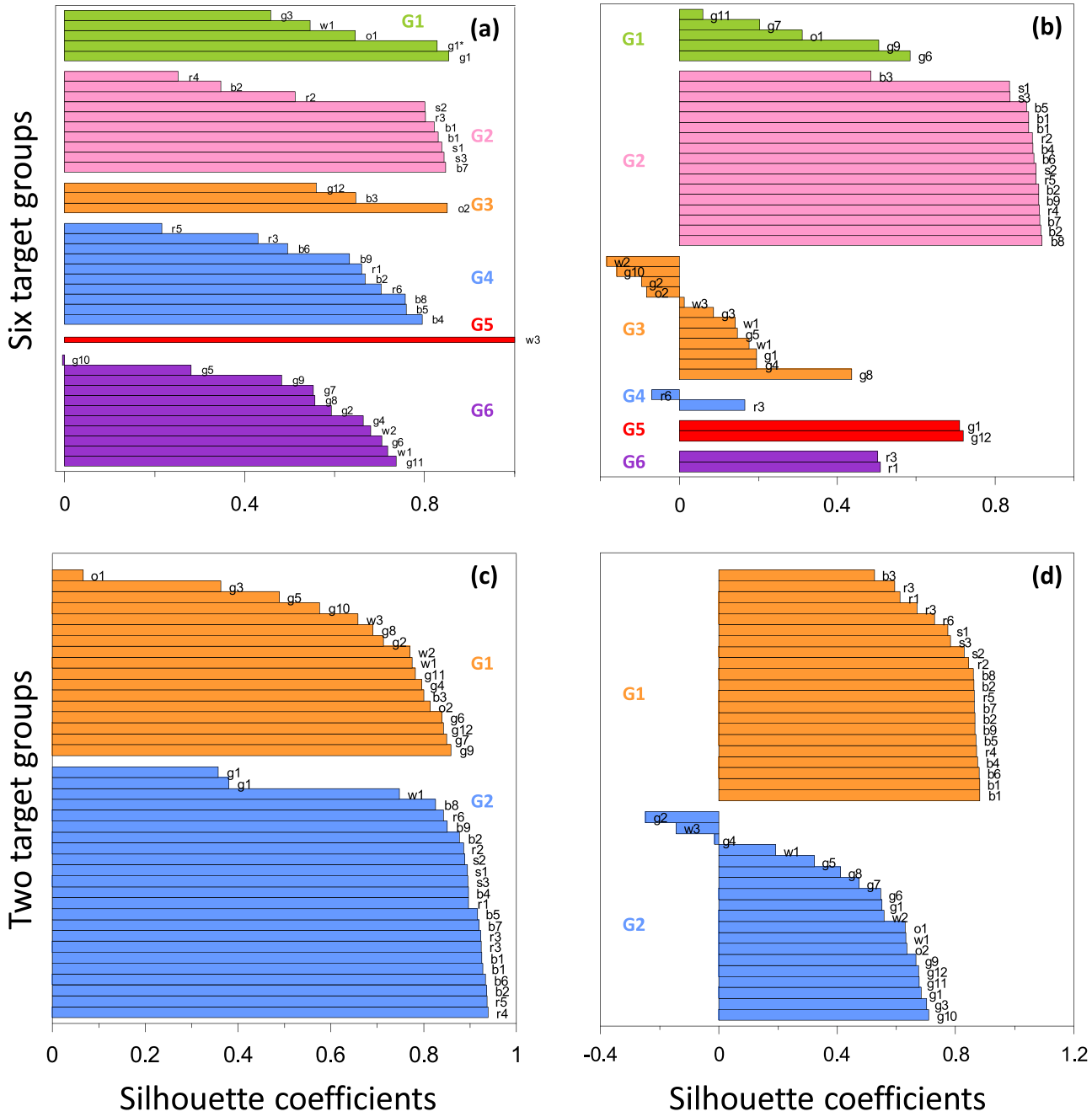

Fig. 7. Silhouette plots for the 40 tea samples, targeting: (a,b) six groups, and (c,d) two groups. The processed data matrices were: (a,c) $\mathbf{B_C}$, and (b,d) $\mathbf{Q_C}$ (see Table 1 for sample identification).

4.4.2. Classification using the k-means algorithm

PCA clearly distinguishes between two major groups of samples: one composed of white, green, and oolong tea varieties, and the other comprising black, red, and rooibos teas. This differentiation is particularly evident in PCA using centred elution bands ($\mathbf{B_C}$) and centred band ratios ($\mathbf{Q_C}$) as datasets (Fig. 6b and e for $\mathbf{B_C}$, and 6c and f for $\mathbf{Q_C}$).

Next, the 40 tea samples (including five replicates) underwent *k*-means clustering using the $\mathbf{B_C}$ and $\mathbf{Q_C}$ data matrices. Initially, six centroids ($k = 6$) were tested corresponding to the six tea varieties. Fig. 7 illustrates the outcomes of the *k*-means clustering through silhouette plots (Section 2.3), which quantifies the success of classifying each sample into predefined groups. The silhouette size (abscissa) measures similarity, offering insights into each sample consistency within its assigned group and its distinction from other groups. Silhouette values range from −1 to +1, with higher positive values indicating better agreement within the sample group and greater dissimilarity from neighbouring groups.

(a) *Creation of six clusters based on the k-means algorithm*

Fig. 7a shows results from processing elution bands (T2 in Section 2.3). Black, red, and rooibos tea samples cluster into two groups (G2 and G4) with silhouette values ranging from 0.6 to 0.8, indicating well-defined groups with strong coherence among samples. Nevertheless, particular samples such as r_2, b_2, r_4 in G2, and b_6, r_3, r_5 in G4 display silhouette values below 0.5, indicating significant differentiation from their respective groups.

The formation of group G5, comprising a single white tea sample (w_3), and group G3, consisting of three distinct samples (o_2, b_3, and g_{12}), is noteworthy. Among these, sample b_3 particularly stands out as significantly different from the others, potentially qualifying as an outlier. White, green, and oolong tea samples organise into two groups, with slightly lower silhouette values (average values: $\overline{x}_{G1} = 0.67$ and $\overline{x}_{G6} = 0.54$). Consistently low silhouette values across multiple observations suggest that either too few or too many groups have been defined. This outcome indicates the need to explore configurations with different numbers of centres. Additionally, the presence of samples with silhouette values below 0.5 suggests that these samples might better align with other groups.

Fig. 7b shows the results from processing centred band ratios ($\mathbf{Q_C}$). A well-defined group (G2) comprising samples of black, red, and rooibos tea is identified, with $\overline{x}_{G2} \pm s_{G2} = 0.87 \pm 0.10$ and a silhouette size close to 1, indicating strong coherence among samples within the group. Additionally, two distinct groups of red tea samples (G4 and G6) are observed, but with silhouette values below 0.5, suggesting the possibility of their inclusion in other groups, particularly G2, or being on the borderline between groups.

Moreover, three groups have been identified consisting of samples of white, green, and oolong tea: G1, G3, and G5. Among these, the G3 group contains the most significant number of samples with $\overline{x}_{G3} \pm s_{G3} = 0.072 \pm 0.18$, although with relatively low silhouette values. Therefore, the results obtained with the two data matrices $\mathbf{B_C}$ and $\mathbf{Q_C}$ do not differ significantly, indicating that similar types of samples have been grouped in both cases.

(b) *Creation of two clusters based on the k-means algorithm*

When two groups using elution bands were sought (Fig. 7c), the *k*-means algorithm identified two significant groups containing the same type of tea samples: G1 ($\overline{x} \pm s = 0.69 \pm 0.21$) and G2 ($\overline{x} \pm s = 0.85 \pm 0.16$), aligning with the PCA results. Group G1 exhibits greater scattering than G2, with lower silhouette values below 0.5 (particularly for samples g_5, g_3, and o_1; even as low as 0.06 for o_1). For group G2, all observations have high silhouette values (> 0.8), except for sample g_1 (< 0.4), indicating it belongs to a different tea variety.

Using band ratios (Fig. 7d) and aiming for two groups, the samples exhibited similar associations. In group G1 ($\bar{x} \pm s = 0.80 \pm 0.11$), silhouette values surpassed 0.6 in most cases, indicating strong coherence among different samples. However, consistent with PCA findings, silhouette values for r_6, r_3, r_1 and b_3 were smaller due to the distance of these samples from the centroid that groups the others. Group G2 ($\overline{x} \pm s = 0.46 \pm 0.30$) encompassed all white, green, and oolong teas, but with lower silhouette values than G1, signifying greater scattering in the samples, particularly for g_4, w_3 and g_2.

4.4.3. Classification using PLS-DA

The final classification technique, utilising both elution bands and band ratios, was PLS-DA. Rooibos and oolong tea samples were excluded due to insufficient representation (three and two samples, respectively) for proper validation. The outcomes of the classification using hard PLS-DA are illustrated in Fig. 8. Fig. 8a and b depict the projection of the remaining 35 samples (green, white, red, and black teas) onto the discriminant PC1-PC2 plane, using $\mathbf{B_C}$ and $\mathbf{Q_C}$ as input data, respectively. A clearer distinction between varieties is observed when $\mathbf{Q_C}$ is used. The varieties are reasonably well differentiated, except for white tea, which projects close to the origin. In constructing the PLS-DA models, no samples were misclassified using the $\mathbf{Q_C}$ matrix, whereas one sample (w_1) was wrongly assigned using the $\mathbf{B_C}$ matrix.

Leave-one-out cross-validation was employed to assess the predictive capability of the PLS-DA model, with results summarised in Table 2. The $\mathbf{Q_C}$ matrix generally provided accurate predictions, though some misclassifications occurred for samples w_1, w_2, r_2, r_4, and r_5. The $\mathbf{B_C}$ matrix misclassified the same samples along with w_3. These results reveal that both $\mathbf{B_C}$ and $\mathbf{Q_C}$ models exhibit limitations in accurately distinguishing white and red teas from other varieties. In the $\mathbf{B_C}$ model (Table 2), three white tea samples were incorrectly classified as green tea and one as black tea. Conversely, the $\mathbf{Q_C}$ model incorrectly classified three white tea samples as green tea. For red tea, four out of seven samples were correctly classified in both models, while the remaining three were misclassified as black tea. Overall statistics (sensitivity, specificity, and efficiency) were 83 % for $\mathbf{Q_C}$ and 77 % for $\mathbf{B_C}$, confirming these observations.

The limited sample size for certain classes posed challenges in accurately defining them. Moreover, classes with insufficient informative content led to ambiguous classification outcomes. Improving these results would require enhancing the analytical methodology, either by increasing the sample size to better define classes and observe a wider range of constituents, or by incorporating different types of chemical information (e.g., from alternative chromatographic modes like Hydrophilic Interaction Liquid Chromatography and Supercritical Fluid Chromatography).

5. Conclusions

One major advantage of CF-based analysis over targeted compound determination is its independence from standards for sample characterisation, enabling more flexible and efficient analyses. Similarly, global models do not necessarily require standards to accurately simulate and predict full chromatograms. Combining both approaches simplifies method development for complex samples like CFs. This work illustrates that global models not only comprehensively describe retention but they are also useful to optimise CFs to enhance classifications. Global models make optimisations as straightforward as conventional methodologies using individual models for simpler samples.

The global-model approach is applied for the first time to a large sample set, including 40 tea samples across six plant species, highlighting its broad applicability and reliability. This underscores its potential for diverse fields dealing with complex samples. Initial assessments indicated that green tea showed the majority of characteristic peaks. Hence, a representative sample from this category (g_{10}) was chosen to determine the optimal gradient. Using a conventional genetic algorithm (GA), the best gradient was selected from the optimal Pareto

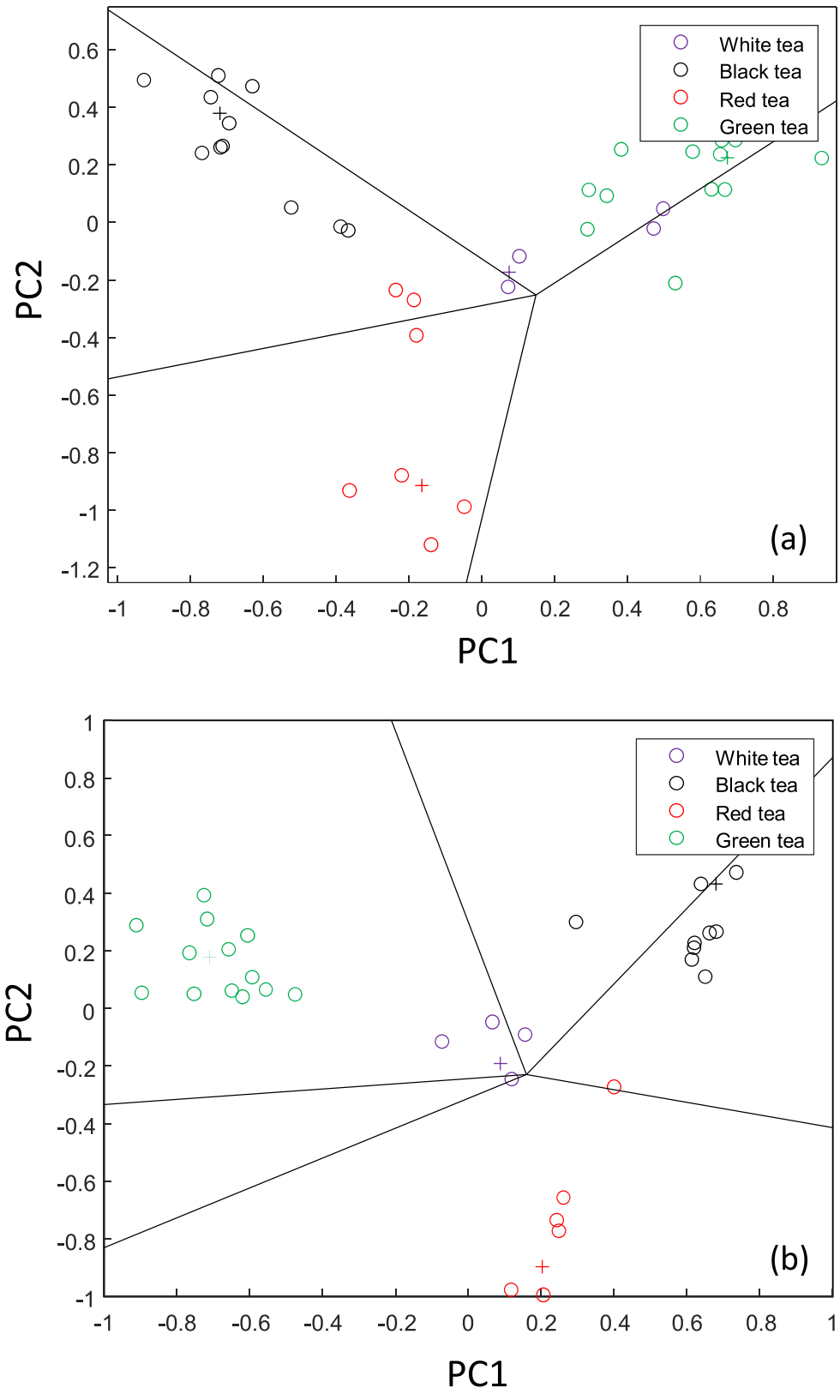

Fig. 8. Classification plot using 35 samples (excluding rooibos and oolong teas with insufficient samples) as the training set and hard PLS-DA. The processed data matrices were: (a) $\mathbf{B}_C$, and (b) $\mathbf{Q}_C$, both on the PC1-PC2 projection plane.

Table 2
Results from leave-one-out cross-validation obtained by PLS-DA models, constructed using both elution bands and band ratios as input data, together with some classification statistics [41]. The analysis encompassed samples from white, black, red, and green teas (due to limited sample availability for validation, oolong and rooibos classes were excluded from consideration).

Figures of merit	Elution bands ($\mathbf{B_C}$)				Band ratios ($\mathbf{Q_C}$)			
Confusion matrices[a]	White	Black	Red	Green	White	Black	Red	Green
White	0	1	0	3	1	0	0	3
Black	0	11	0	0	0	11	0	0
Red	0	3	4	0	0	3	4	0
Green	0	1	0	12	0	0	0	13
True positive	0	11	4	12	1	11	4	13
False positive	0	5	0	3	0	3	0	3
Class statistics (%)								
Sensitivity[b]	0	100	100	92	25	100	57	100
Specificity[c]	100	79	79	86	100	88	100	86
Efficiency[d]	0	89	89	89	50	94	76	93
Total statistics (%)[e]	77				83			

[a] Rows indicate the number of samples within each class, and columns the assignation of the samples to each of the categories according to the PLS-DA model.
[b] Class sensitivity is the percentage of samples of each target class correctly recognised as members of the class.
[c] Class specificity is defined, for each target class, as the percentage of samples from other classes, which are correctly attributed as inconsistent with the target class.
[d] Class efficiency is the geometric mean of sensitivity and specificity.
[e] Total statistics characterising the performance of the entire PLS-DA model.

front, balancing resolution, analysis time, and peak distribution. Validation confirmed its efficacy through comparison with the experimental chromatograms.

PCA initially categorised tea samples and identified the optimal dataset (direct signals, elution bands, or band ratios) and pre-processing treatments (centring or autoscaling). It revealed clear differentiation between two main groups: white, green, and oolong teas formed a cohesive cluster, while red, black, and rooibos teas constituted another distinct group, which was confirmed by *k*-means clustering. The preferred classification method was PLS-DA using band-centred ratios. Classification limitations were anticipated due to the complex nature of tea processing, involving drying and fermentation. These manufacturing factors influence chemical composition, resulting in chromatograms with variable information content influenced by environmental conditions.

A challenge affecting classification success was variability in peak positions due to chromatographic acquisitions spanning several weeks, resulting in less accurate raw signal classifications. To mitigate these fluctuations, time intervals (elution bands) were defined to cover observed peak positions across all samples. This approach compensated for time fluctuations and excluded time domains lacking significant chemical information. Additionally, band ratios were explored to address variability in extraction performance by computing ratios between elution bands. This method ensured each peak acted as an internal standard for others, maintaining equal relevance for both the least and most intense signals in subsequent classifications.

The applied analytical method has inherent limitations. Sampling requires maximising sample quantity while avoiding detector and column saturation. Obtaining representative samples for each category is challenging due to variables such as storage time, origin, variety, soil and growing conditions, and additives. Chemical characterisation is particularly challenging for varieties with minimal peaks, such as rooibos and oolong, for which the number of available samples was smaller. Furthermore, tea manufacturing procedures impact chemical composition, complicating the classification of tea categories. These factors highlight the complexities that can hinder successful classification efforts.

CRediT authorship contribution statement

P. Peiró-Vila: Writing – original draft, Software, Investigation, Formal analysis. **C. Pérez-Gracia:** Writing – original draft, Investigation, Formal analysis. **J.J. Baeza-Baeza:** Writing – original draft, Methodology, Investigation, Formal analysis. **M.C. García-Alvarez-Coque:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Investigation, Funding acquisition. **J. R. Torres-Lapasió:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

M.C. García-ALvarez-Coque reports financial support was provided by Spain Ministry of Science and Innovation. Other 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

Work supported by Grant PID2019–106708GB-I00 funded by MCIN (Ministery of Science and Innovation) of Spain/AEI/10.13039/501100011033; Grant PID2022–148935NB-I00 funded by MCIU (Ministery of Science, Innovation and Universities) of Spain/AEI/10.13039/501100011033; and Grant GVRTE/2023/4534963 funded by Generalitat Valenciana (Spain). Pau Peiró-Vila thanks the pre-doctoral grant ACIF 2022 (Generalitat Valenciana).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.chroma.2024.465128.

References

[1] H. Siddique, M. Sarwat, Herbal Medicines: a Boon for Healthy Human Life, Academic Press, Cambridge, MA, 2022.
[2] K. Hayat, H. Iqbal, U. Malik, U. Bilal, S. Mushtaq, Tea and its consumption: benefits and risks, Crit. Rev. Food Sci. Nutr. 55 (2015) 939–954.
[3] E.G. Wheelwright, Medicinal Plants and Their History, Dover Publications, Mineola, New York, 1974.
[4] A.B. Sharangi, Medicinal and therapeutic potentialities of tea (Camellia sinensis L.): a review, Food Res. Intl. 42 (2009) 529–535.

[5] T. Yi, L. Zhu, W.L. Peng, X.C. He, H.L. Chen, J. Lie, T. Yu, Z.T. Liang, Z.Z. Zhao, H. B. Chen, Comparison of ten major constituents in seven types of processed tea using HPLC-DAD-MS followed by principal component and hierarchical cluster analysis, LWT Food Sci. Technol 62 (2015) 194–201.
[6] M. Wong, S. Sirisena, K. Ng, Phytochemical profile of differently processed tea: a review, J. Food Sci. 87 (2022) 1925–1942.
[7] J. Pettigrew, B. Richardson, The New Tea Companion: A Guide to Teas throughout the World, 3rd ed., Benjamin Press, Humble, TX, 2015.
[8] L. Bramati, F. Aquilano, P. Pietta, Unfermented rooibos tea: quantitative characterization of flavonoids by HPLC-UV and determination of the total antioxidant activity, J. Agric. Food Chem. 51 (2003) 7472–7474.
[9] V.V. Chopade, A.A. Phatak, A.B. Upaganlawar, A.A. Tankar, Green tea (Camellia sinensis): chemistry, traditional, medicinal uses and its pharmacological activities: a review, Phcog. Rev. 2 (2008) 157–162.
[10] C.J. Xu, Y.Z. Liang, F.T. Chau, Y. Vander Heyden, Pretreatments of chromatographic fingerprints for quality control of herbal medicines, J. Chromatogr. A 1134 (2006) 253–259.
[11] X.K. Zhong, D.C. Li, J.G. Jiang, Identification and quality control of Chinese medicine based on the fingerprint techniques, Curr. Med. Chem. 16 (2009) 3064–3075.
[12] C. Tistaert, B. Dejaegher, Y. Vander Heyden, Chromatographic separation techniques and data handling methods for herbal fingerprints: a review, Anal. Chim. Acta 690 (2011) 148–161.
[13] C.S. Funari, R.L. Carneiro, A.M. Andrade, E.F. Hilde, A.J. Cavalheiro, Green chromatographic fingerprinting: an environmentally friendly approach for the development of separation methods for fingerprinting complex matrices, J. Sep. Sci. 37 (2014) 37–44.
[14] J. Pons, A. Bedmar, N. Núñez, J. Saurina, O. Núñez, Tea and chicory extract characterization, classification and authentication by non-targeted HPLC-UV-FLD fingerprinting and chemometrics, Foods 10 (2021) 2935–2951.
[15] M.S. El-Shahawi, A. Hamza, S.O. Bahaffi, A.A. Al-Sibaai, T.N. Abduljabbar, Analysis of some selected catechins and caffeine in green tea by high performance liquid chromatography, Food Chem. 134 (2021) 2268–2275.
[16] M.C. García-Alvarez-Coque, J.J. Baeza-Baeza, G. Ramis-Ramos, Reversed phase liquid chromatography, in: J.L. Anderson, A. Berthod, V. Pino, A. Stalcup (Eds.), Analytical Separation Science Series, Vol. 1, 2015, pp. 159–167.
[17] A. Gisbert-Alonso, J.A. Navarro-Huerta, J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, Global retention models and their application to the prediction of chromatographic fingerprints, J. Chromatogr. A 1637 (2021) 461845.
[18] A. Gisbert-Alonso, S. López-Ureña, J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, Chromatographic fingerprint-based analysis of extracts of green tea, lemon balm and linden: I. Development of global retention models without the use of standards, J. Chromatogr. A 1672 (2022) 463060.
[19] A. Gisbert-Alonso, A. Navarro-Martínez, J.A. Navarro-Huerta, J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, Chromatographic fingerprint-based analysis of extracts of green tea, lemon balm and linden: II. Simulation of chromatograms using global models, J. Chromatogr. A 1684 (2022) 463561.
[20] P. Peiró-Vila, M.D. Villamonte, I. Luján-Roca, J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, Performance of global retention models in the optimisation of the chromatographic separation (I): simple multi-analyte samples, J. Chromatogr. A 1689 (2023) 463756.
[21] J. Jing, H.S. Parekh, M. Wei, W.C. Ren, S.B. Chen, Advances in analytical technologies to evaluate the quality of traditional Chinese medicines, Trends Anal. Chem. 44 (2013) 39–45.
[22] H. Wang, M. Chen, J. Li, N. Chen, Y. Chang, Z. Dou, Y. Zhang, P. Zhuang, Z. Yang, Quality consistency evaluation of Kudiezi injection based on multivariate statistical analysis of the multidimensional chromatographic fingerprint, J. Pharm. Biomed. Anal. 177 (2020) 112868.
[23] U.D. Neue, H.J. Kuss, Improved reversed-phase gradient retention modeling, J. Chromatogr. A 1217 (2010) 3794–3803.
[24] J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, Levels in the interpretive optimisation of selectivity in high-performance liquid chromatography: a magical mystery tour, J. Chromatogr. A 1120 (2006) 308–321.
[25] J.A. Navarro-Huerta, T. Alvarez-Segura, J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, Study of the performance of a resolution criterion to characterise complex chromatograms with unknowns or without standards, Anal. Methods 9 (2017) 4293–4303.
[26] A. Gisbert-Alonso, J.A. Navarro-Huerta, J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, Testing experimental designs in liquid chromatography (II): influence of the design geometry on the prediction performance of retention models, J. Chromatogr. A 1654 (2021) 462458.
[27] J. Horn, N. Nafpliotis, D.E. Goldberg, in: A niched Pareto genetic algorithm for multiobjective optimization, 1, 1994, pp. 82–87.
[28] M. Goodarzi, P.J. Russell, Y. Vander Heyden, Similarity analyses of chromatographic herbal fingerprints: a review, Anal. Chim. Acta 804 (2013) 16–28.
[29] G. Alaerts, J. Van Erps, S. Pieters, M. Dumarey, A.M. van Nederkassel, M. Goodarzi, J. Smeyers-Verbeke, Y. Vander Heyden, Similarity analyses of chromatographic fingerprints as tools for identification and quality control of green tea, J. Chromatogr. B 910 (2021) 61–70.
[30] D.L. Massart, Handbook of Chemometrics and Qualimetrics, Data Handling in Science and Technology, Elsevier, Amsterdam, 1997.
[31] I.T. Jolliffe, Principal Component Analysis, Springer, Berlin, 2002.
[32] A.K. Jain, R.P.W. Duin, J. Mao, Statistical pattern recognition: a review, IEEE Trans. Pattern Anal. Mach. Intell. 22 (2000) 4–37.
[33] D. Maltoni, D. Maio, A.K. Jain, S. Prabhakar, Handbook of Fingerprint Recognition, Springer Science & Business Media, Berlin, 2009.
[34] M. Alloghani, D. Al-Jumeily, J. Mustafina, A. Hussain, A.J. Aljaaf, A systematic review on supervised and unsupervised machine learning algorithms for data science, in: M. Berry, A. Mohammed, B. Yap (Eds.), Supervised and Unsupervised Learning For Data Science, Springer, Berlin, 2020, pp. 3–21.
[35] S.P. Lloyd, Least squares quantization in PCM, IEEE Trans. Inf. Theory 28 (1982) 129–137.
[36] D. Ballabio, V. Consonni, Classification tools in Chemistry. Part 1: linear models. PLS-DA, Anal. Methods 5 (2013) 3790–3798.
[37] A.L. Pomerantsev, O. Ye Rodionova, Multiclass partial least squares discriminant analysis: taking the right way. A critical tutorial, J. Chemom. 32 (2018) e3030.
[38] J.A. Navarro-Huerta, J.R. Torres-Lapasió, S. López-Ureña, M.C. García-Alvarez-Coque, Assisted baseline subtraction in complex chromatograms using the BEADS algorithm, J. Chromatogr. A 1507 (2017) 1–10.
[39] M.J. Lerma-García, R. Lusardi, E. Chiavaro, L. Cerretani, A. Bendini, G. Ramis-Ramos, E.F. Simó-Alfonso, Use of triacylglycerol profiles established by high performance liquid chromatography with ultraviolet-visible detection to predict the botanical origin of vegetable oils, J. Chromatogr. A 1218 (2011) 7521–7527.
[40] T. Alvarez-Segura, E. Cabo-Calvet, J.R. Torres-Lapasió, M.C. García-Alvarez-Coque, An approach to evaluate the information in chromatographic fingerprints: application to the optimisation of the extraction and conservation conditions of medicinal herbs, J. Chromatogr. A 1422 (2015) 178–185.
[41] Y.V. Zontov, O. Ye Rodionova, S.V. Kucheryavskiy, A.L. Pomerantsev, PLS-DA: a MATLAB GUI tool for hard and soft approaches to partial least square discriminant analysis, Chemometr. Intell. Lab. Syst. 203 (2020) 104064.