Investigating Gas Chromatography Instrumentation Performance and Developing Data Analysis Strategies for Chemical Analysis Applications

Jungho Ahn

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science in Applied Chemical Science and Technology

University of Washington

2026

Committee:

Robert E. Synovec

Dan Fu

Program Authorized to Offer Degree:

Chemistry

University of Washington

Abstract

Investigating Gas Chromatography Instrumentation Performance and Developing Data Analysis Strategies for Chemical Analysis Applications

Jungho Ahn

Chair of the Supervisory Committee:
Robert E. Synovec
Department of Chemistry

Gas chromatography (GC) is a widely used analytical technique for the separation, detection, and identification of volatile and semi-volatile organic compounds. This thesis investigates two complementary analytical applications of GC-based instrumentation. The first chapter evaluates the analytical performance of the Agilent Intuvo 9000 gas chromatograph equipped with a flame ionization detector (GC-FID), while the second chapter explores the feasibility of using activated carbon strips to recover, preserve, and analyze fragrance residues for forensic fingerprinting applications. In Chapter 1, a complex test mixture was analyzed to determine the limit of detection (*LOD*) of the Intuvo 9000 GC-FID and to investigate compound identification using GC-quadrupole mass spectrometry (GC-qMS). Chromatographic data were preprocessed using baseline correction and analyzed using principal component analysis (PCA) to evaluate analytical reproducibility across multiple concentration levels. The Intuvo 9000 GC demonstrated excellent reproducibility, demonstrating the instrument's ability to detect compounds at low concentrations. Because FID does not provide structural information, GC-

qMS was employed for compound identification for the test mixture. A mass spectrum averaging approach combined with intensity thresholding significantly improved NIST library matching performance, increasing representative match values (MV) from the complex mixture. Overall, this work demonstrates the analytical capabilities of the Agilent Intuvo 9000 GC for sensitive chromatographic detection and highlights the advantages of GC-qMS for compound identification. Chapter 2 investigates the use of activated carbon strips for the recovery, storage, and extraction of fragrance residues as a sample preparation method for GC-MS analysis in forensic fragrance fingerprinting applications. Thirteen commercial fragrances were examined as neat samples, initial carbon strip extracts (I-CSE), and stored carbon strip extracts (S-CSE). Total ion current (TIC) chromatogram-based PCA demonstrated good reproducibility for neat samples but failed to reliably classify I-CSE and S-CSE samples because of analyte loss and increased background noise. To overcome these limitations, a mass spectrum-based workflow incorporating TIC signal intensity thresholding, MV filtering, and concatenated spectral features were developed. Applying a neat TIC peak threshold, mass spectral intensity threshold and a MV criteria substantially improved spectral quality and multivariate classification. The resulting PCA provided improved clustering of fragrance samples, while a target-based peak-counting strategy could distinguish fragrances using persistent chemical fingerprint markers retained after 2.5 months of freezer storage. Therefore, the developed activated carbon strip workflow, combined with mass spectrum-based chemometric analysis, provides a practical approach for preserving and comparing fragrance chemical fingerprints. These findings support the potential forensic application of fragrance residue analysis for sample comparison and chemical fingerprinting following storage and environmental exposure.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Dr. Synovec, for his continuous guidance and support throughout my research journey. His mentorship has been invaluable, and he has consistently guided me in the right direction whenever I felt lost or uncertain. Beyond scientific training, I am sincerely grateful for his encouragement and emotional support, which helped me regain confidence during challenging moments in my research. I have learned not only the fundamentals of analytical chemistry from him, but also deeper insights into scientific thinking, problem-solving, and perseverance. I am truly grateful for the opportunity to continue my PhD studies in his group.

I would also like to sincerely thank all members of the Synovec group for their support and kindness. Wenjing, Rachel, Cassy, Haylee, and Valencia have always been approachable and generous with their time, patiently answering my questions and offering help whenever I needed it. I truly appreciate the collaborative and supportive environment they have created in the lab. I am very grateful for all the discussions, shared experiences, and encouragement, and I look forward to continuing my journey alongside you all.

I would like to extend my sincere appreciation to my committee members, Dr. Fu and Dr. Durfy, for their time, valuable feedback, and support throughout the development of my thesis. Their insightful comments and suggestions have greatly improved the quality of my work, and I am thankful for their willingness to serve on my committee.

Finally, I would like to thank all my friends who have supported me during my graduate studies, both academically and personally. Your encouragement has been invaluable in helping me complete this work.

DEDICATION

To my parents, without your love and support, I would not have been able to complete my master's degree. Thank you for your endless encouragement, sacrifices, and for showing me unconditional love throughout the past 25 years. This achievement would not have been possible without you.

Chapter 1. Basic Principles of Gas Chromatography Through Studies of a Complex Test Mixture using the Intuvo 9000 GC-FID and GC-qMS

1.1 Introduction to Gas Chromatography (GC)

Gas chromatography (GC) is an analytical separation technique used to identify and quantify volatile and semi-volatile compounds in complex mixtures [1]. In GC, a sample is vaporized and transported through a capillary column by an inert carrier gas, while the analyte boiling point and interactions between analytes and the stationary phase cause analyte compounds to elute at different retention times. Separation is achieved based on differences in physicochemical properties such as volatility and affinity for the stationary phase [2-4]. As compounds exit the column, they are detected as peaks in a chromatogram, where retention time provides qualitative information and peak area is related to analyte concentration (Figure 1). Due to its high separation efficiency, sensitivity, and rapid analysis times, GC is widely used in environmental, forensic, pharmaceutical, and chemical analyses [5,6]. The Agilent Intuvo 9000 GC-FID differs from conventional GC systems in several important aspects. Most notably, it incorporates a ferrule-free flow path in Figure 2, which uses Intuvo Smart Connections instead of conventional ferrules and column nuts. This design minimizes dead volume, reduces the likelihood of leaks, shortens maintenance time, and improves the reproducibility of column installation. In addition, the Intuvo employs a low thermal mass heating system that enables rapid heating and cooling, resulting in faster temperature programming and shorter analysis times [7]. This chapter presents the analysis of a complex test mixture and is divided into two parts. The first part focuses on determining the limit of detection (*LOD*) of the Intuvo 9000 GC to evaluate how low concentration can be reliably detected and learn about basic principles of gas chromatography. The overall workflow is shown in Figure 3. The second part of chapter 1, as illustrated in Figure

4, deals with the limitations of the Intuvo 9000 GC in compound identification, therefore GC-qMS is used to further identify the compounds present in the complex test mixture using mass spectrometry.

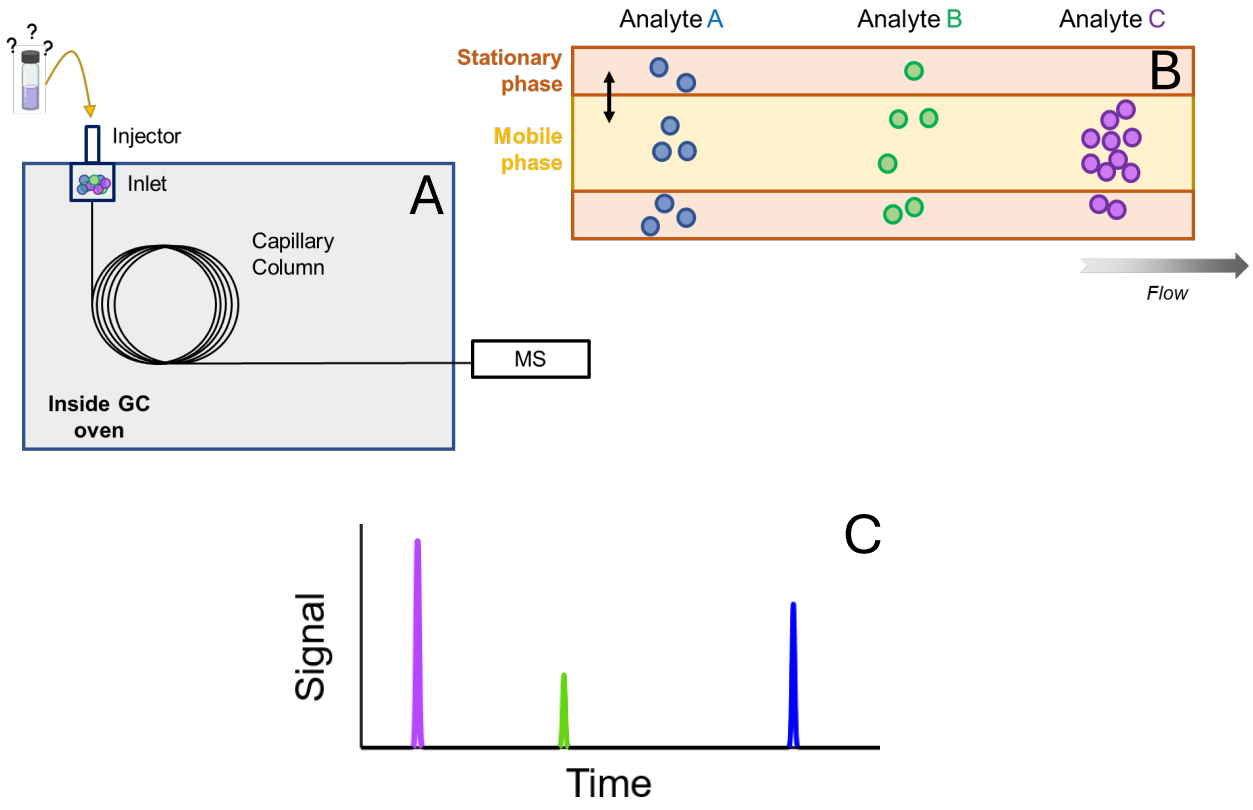


Figure 1. Fundamentals of gas chromatography. (A) Schematic of a conventional GC system showing sample introduction, separation in the capillary column. (B) Illustration of analyte separation within the GC column based on differential interactions with the stationary and mobile phases. (C) Representative TIC chromatogram obtained after chromatographic separation.

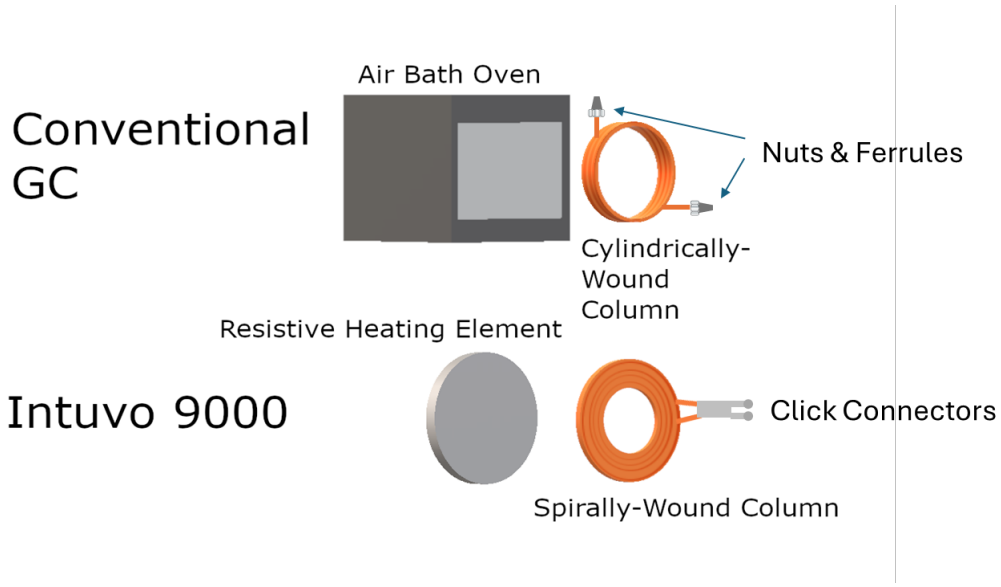


Figure 2. Schematic of conventional GC and Intuvo 9000 GC system. The difference between column heating in conventional GC system and the Intuvo 9000 GC system is illustrated.

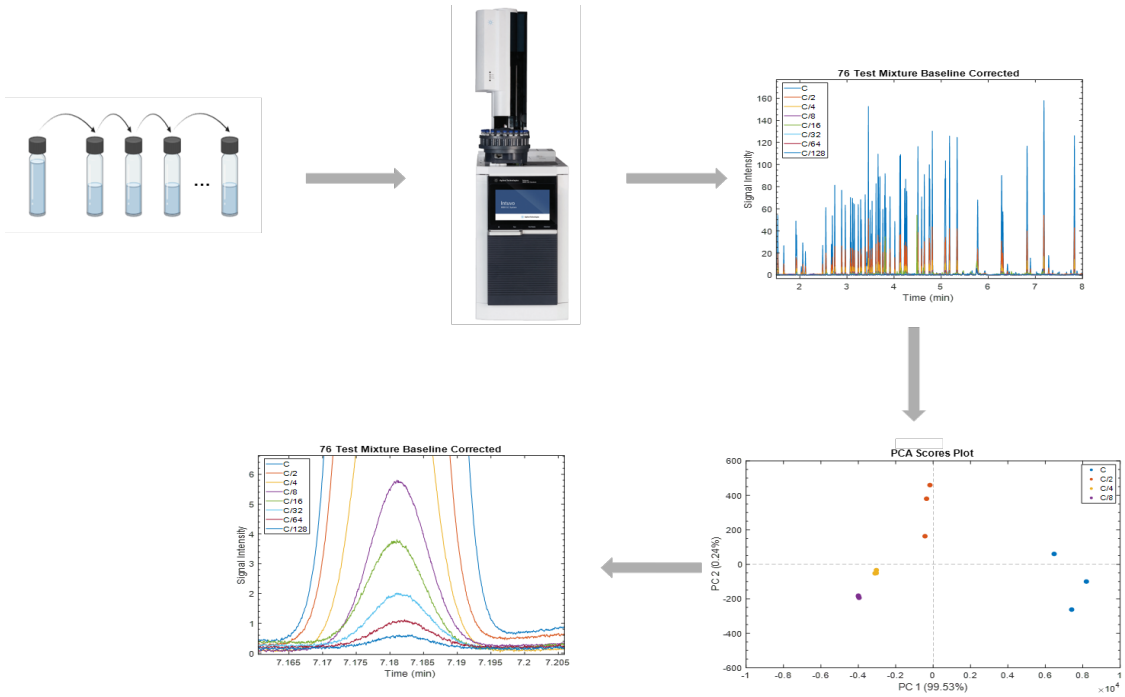


Figure 3. Overview of the study using a test mixture analyzed with the Intuvo 9000 GC. Serial dilutions (1000 ppm to 7.8 ppm) were prepared and analyzed. TIC chromatograms were used for PCA to evaluate reproducibility across concentration levels, and LOD were determined to assess the sensitivity of the Intuvo 9000 GC.

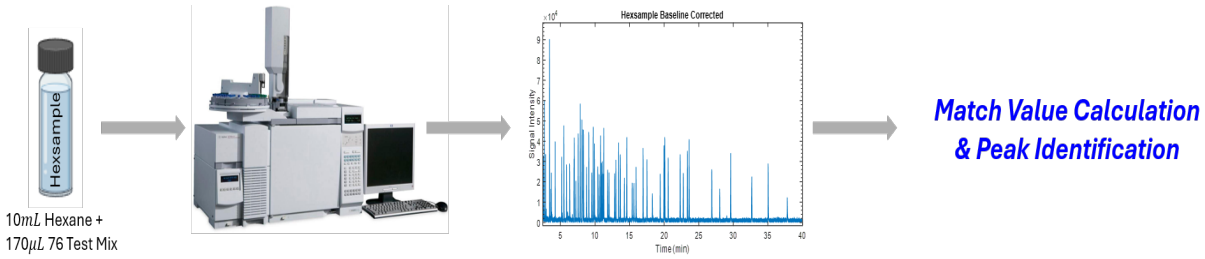


Figure 4. Overview of the study using the test mixture diluted in hexane analyzed by GC-qMS. A hexane-diluted complex test mixture (10 mL hexane + 170 µL of the complex test mixture) was analyzed using GC-qMS. Mass spectra were acquired for chromatographic peaks and compared with the NIST mass spectral library for MV calculation and compound identification.

1.2 Experimental

1.2.1 Sample Preparation

A complex test mixture prepared to contain 74 compounds (Table 1) was prepared in methanol (MeOH). The initial concentration of the complex test mixture was ~1000 ppm (referred to as concentration level = C). Two-fold serial dilutions were performed using MeOH in Figure 5 to get concentrations down to 7.8 ppm (C/128). For GC-qMS analysis, the original complex test mixture was diluted in hexane by adding 170 µL of the complex test mixture to 10 mL of hexane. The solution was vortex-mixed for 30 s prior to analysis.

Table 1. Complex test mix list of compounds.

Alkanes	Conc (ppm in 25mL)	Terpenes	Conc (ppm in 25mL)
heptane	1000.2	(-)-alpha-pinene	1002.1
octane	1001.1	(-)-linalool	1002.1
decane	998.6	**Alcohols**	
undecane	1000.5	1-butanol	1002.8
dodecane	1002	1-pentanol	1000.4
tetradecane	1002.4	2-pentanol	1002.8
hexadecane	1001.8	1-hexanol	1006.5
pristane	1002.2	2-heptanol	1008.6
octadecane	1000.8	1-octanol	1005.6

Alkenes		1-nonanol	995.4
1-hexene	998	2-ethyl-1-hexanol	999.9
1-undecene	1002	tert-amyl alcohol	997.6
Alkynes		**Ketones**	
1-hexyne	1001	2-hexanone	999.2
1-heptyne	998.1	2-heptanone	998.4
5-decyne	998.9	3-octanone	999.6
Cycloalkanes		2-decanone	996.6
cyclooctane	1000.8	2-undecanone	1001.4
butylcyclohexane	1001.2	2-dodecanone	997.1
bicyclohexyl	1002.2	acetophenone	997
Aromatics		butyrophenone	1000.6
benzene	1002.1	**Esters**	
toluene	998.4	dibutyl phthalate	999.6
3-ethyltoluene	1003.4	diethyl phthalate	1003.5
4-ethyltoluene	991.9	ethyl formate	997.7
mesitylene	1002.2	methyl decanoate	996.4
butylbenzene	996.4	methyl caprylate	1003.3
tert-butyl benzene	998.7	methyl salicylate	1001.5
1-propylbenzene	999.9	ethyl salicylate	1003.4
1-ethylnapthalene	999.9	methyl laurate	1002.2
cyclohexylbenzene	997.7	methyl caproate	998.3
diphenylmethane	1001.9	**Amines / Nitrogen Compounds**	
sec-butylbenzene	1001.1	o-toluidine	999.9
1,2,3,4-tetrahydronapthalene	1001		
Sulfur Compounds		**Halogenated Hydrocarbons**	
1-decanethiol	1000.5	1,5-dichloropentane	1003.2
2-thiophenemethylamine	997.1	1-chlorohexane	998.5
2-propylthiophene	1000	1-bromohexane	1000.6
2-(methylthio)-pyridine	1227.5	1-bromoheptane	1003.2
1,4-Oxathiane	998.1	1-bromooctane	1001.7
1-dodecanethiol	1000.5	carbon tetrachloride	1004.9
Phosphorus Compounds		1-iodo-2-methylpropane	1000.5
dimethyl methyl phosphonate	998.4		
diethyl methyl phosphonate	1003.4		

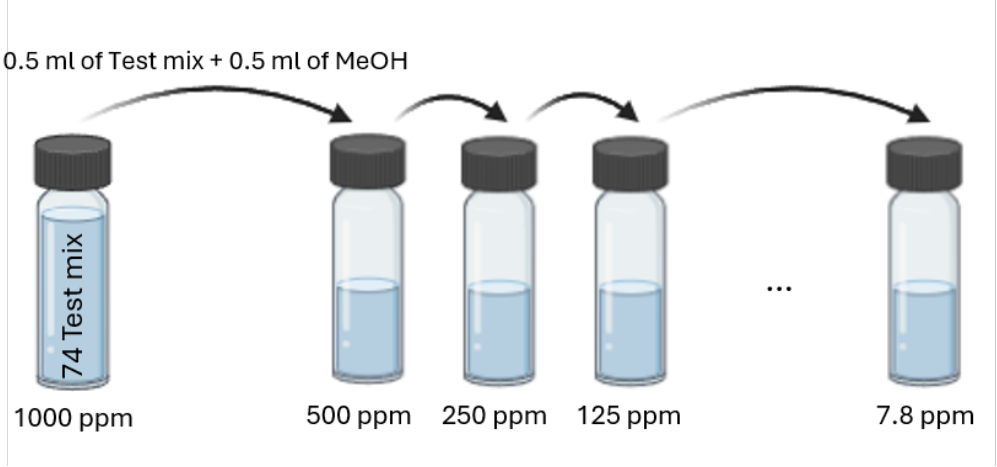


Figure 5. Serial dilution scheme used for the experiment. The original complex test mixture was serially diluted by a factor of two at each step to produce concentration levels ranging from 1000 ppm (C) to 7.8 ppm (C/128) for the analyses.

1.2.2 Instrumental

Samples were analyzed using a GC instrument of an Agilent Intuvo 9000 GC system and Agilent 7693 Series Automatic Liquid Sampler (ALS). All samples were injected 0.5 µL. The Intuvo GC instrument was fit with an Agilent HP-5MS UI column (part #19091S-433UI-INT) that is 30 m in length, an i.d. of 250 µm and film thickness of 0.25 µm. The GC oven temperature program began at 45°C (held for 1 min), followed by a non-linear ramp of 35°C/min, 2°C/min, 60°C/min and 10°C/min to a final temperature of 300°C over an 11-min period.

Table 2. Intuvo 9000 GC operating condition for complex test mixture separation.

Injector	
Temperature	250 °C
Split ratio	1:20
Sample size	0.5 µL
Column	
Dimension	30 m × 0.25 mm × 0.25 µm
Stationary phase	HP-5MS UI
Flowrate	3.0 mL/min
Column Temperature Program	
Initial temperature	45 °C
First program rate Second program rate Third program rate Fourth program rate	35 °C/min 2 °C/min 60 °C/min 10 °C/min
Final temperature	300 °C

For the GC-qMS analysis, sample was analyzed using a GC-MS instrument of an Agilent (Model 6890) gas chromatography with a Model 5973–mass selective detector and an Agilent 7683 Automatic Liquid Sampler (ALS) autoinjector. All samples were injected 0.2 µL. The GC instrument was fit with an Restek Rxi-1 MS column (catalog #13323) that is 30 m in length, an i.d. of 250 µm and film thickness of 0.25 µm. The GC oven temperature program began at 40°C, followed by a linear ramp of 4°C/min to a final temperature of 240°C over a 50-min period. The scan rate was 20 Hz over a *m/z* range of 40 to 280.

Table 3. GC-qMS operating condition for complex test mixture separation.

Injector	
Temperature	250 °C
Split ratio	1:10
Sample size	0.2 µL
Column	
Dimension	30 m × 0.25 mm × 0.25 µm
Stationary phase	Rxi-1MS
Flowrate	1.0 mL/min
Column Temperature Program	
Initial temperature	45 °C
Final temperature	240 °C

1.2.3 Data Analysis

All data acquired from the Intuvo 9000 GC and GC-qMS were processed using MATLAB R2024a (MathWorks, Natick, MA, USA). Raw chromatographic data were baseline-corrected using a rolling-ball minimum algorithm [8]. To evaluate reproducibility across concentration levels, PCA was performed on four sample concentrations ranging from 1000 ppm (C) to 125 ppm (C/8).

For multivariate analysis, the baseline-corrected chromatograms were first truncated to the region of interest (1.5–8.0 min) and organized into a data matrix consisting of 12 samples (four concentrations with three replicates each). Prior to PCA, the data was mean-centered by subtracting the variable-wise mean from each column to ensure equal weighting of all variables.

The processed data matrix was then subjected to singular value decomposition (SVD), yielding orthogonal score and loading matrices. The variance explained by each principal component was calculated from the squared singular values and normalized to obtain

percentage explained variance. Loadings were derived from the singular vectors scaled by the singular values, representing the contribution of each variable to the corresponding principal components.

To determine the concentration *LOD* of the Intuvo 9000 GC for a typical analyte, samples ranging from 62.5 ppm (C/16) to 7.8 ppm (C/128) were analyzed in a single replicate. The *LOD* was determined using the *S/N* approach, where baseline noise was measured from a chromatographic region adjacent to the analyte peak and the standard deviation of the baseline noise was calculated.

For GC–qMS analysis, baseline-corrected data was used for compound identification. Each peak position defined in the peak list was used as a reference center, and a scan window of ±25 scans was selected around each chromatographic peak. Within this window, mass spectrum intensities were averaged across scans to generate a representative spectrum for each peak.

This averaging approach was applied because single-scan mass spectra at the peak maximum were insufficiently stable in this data, which resulted in reduced library matching performance with low MVs. Integrating multiple scans across the peak profile provided a more robust and representative mass spectrum.

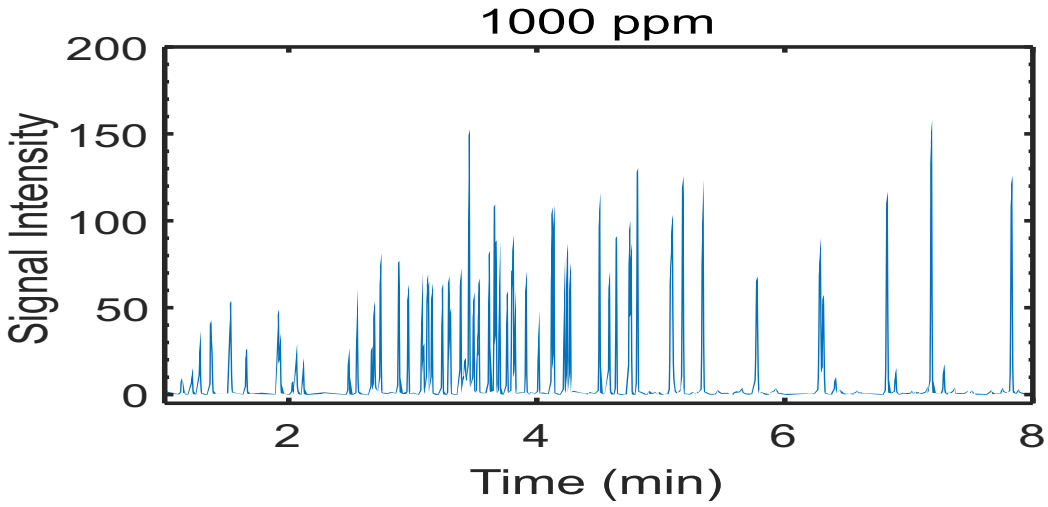
The resulting spectra were then normalized to the base peak intensity, and signals below 1% or 2% relative abundance (with respect to the normalized base peak) were removed to reduce background noise and low-intensity artifacts. The processed spectra were stored for each detected peak and subsequently used as input for library-based compound identification using a modified MATLAB implementation of the NIST matching algorithm.

1.3 Results and Discussion

1.3.1 Intuvo 9000 Gas Chromatography Limit of Detection Study

This study evaluated the analytical performance of the Agilent Intuvo 9000 GC and GC-qMS using the complex test mixture across multiple concentration levels. Chromatographic reproducibility, PCA and compound identification were investigated using preprocessed chromatograms and mass spectral analysis.

Representative chromatograms obtained from the complex test mixture showed peak separation across the concentration range. Figure 6 shows the TIC chromatograms for serially diluted samples from 1000 ppm to 125 ppm. Most compounds were consistently found within the selected retention time region (1.5 – 8.0 min). Figure 7 shows overlaid baseline-corrected TIC chromatograms of the 76-component test mixture at four concentration levels (C to C/8), each measured in triplicate. The results demonstrate good reproducibility, with consistent peak patterns observed across both concentration levels and replicate injections. Retention times remain well aligned across all runs, indicating stable chromatographic performance and effective baseline correction. Zoomed in region Fig. 7B further shows consistent peak shape and retention time across replicates.



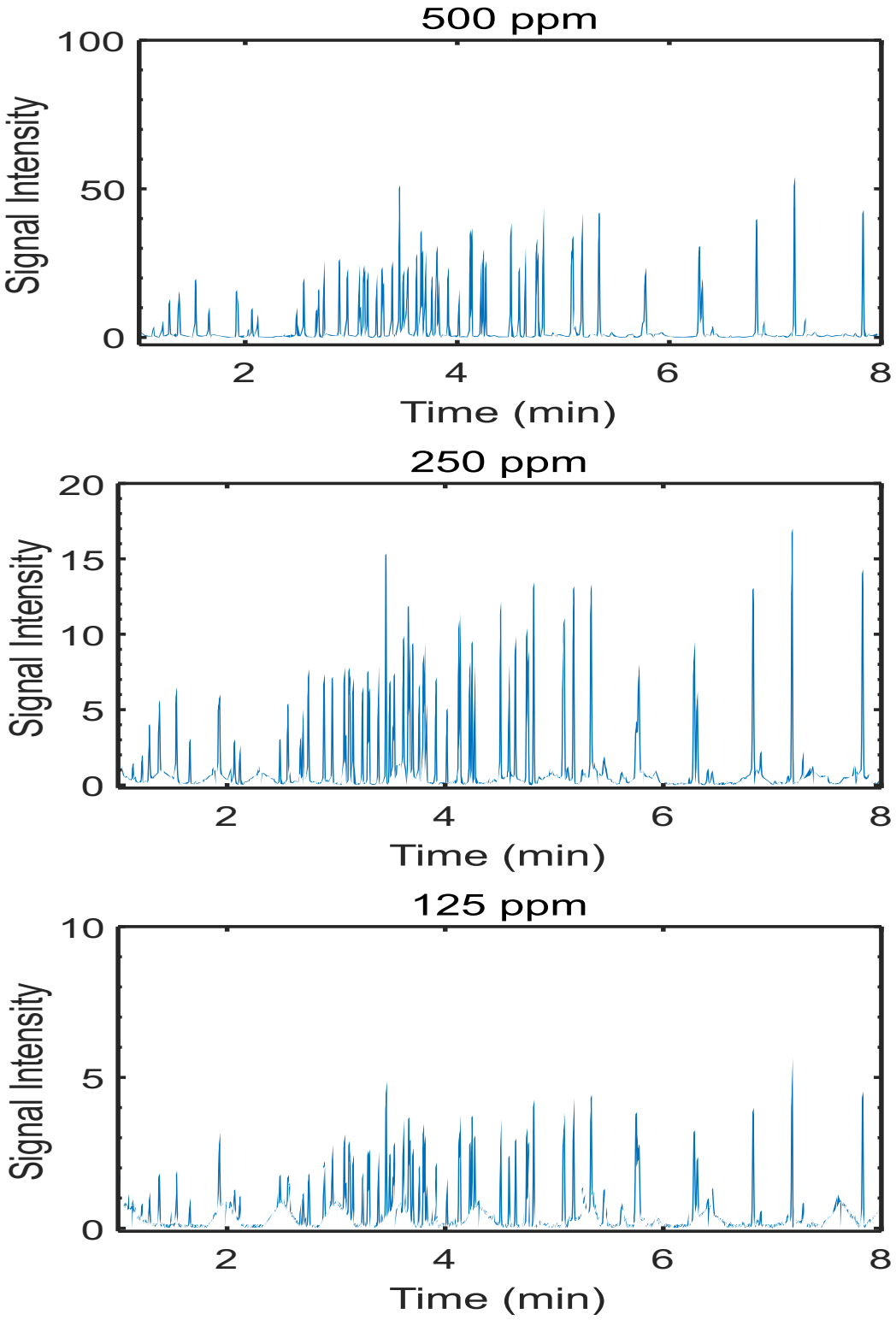


Figure 6. TIC chromatograms of the complex test mixture. Chromatograms are shown for four concentration levels ranging from 1000 ppm (C) to 125 ppm (C/8).

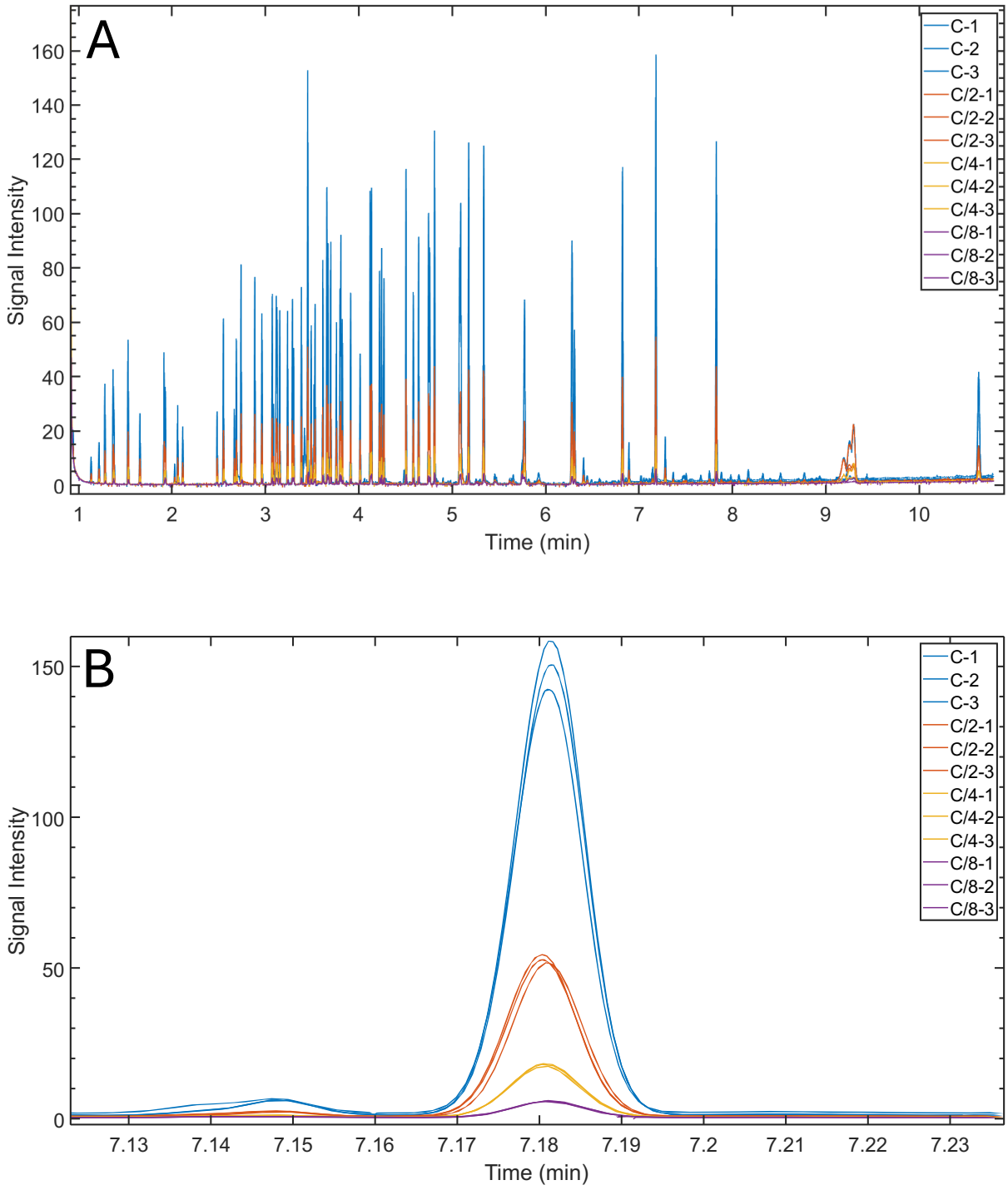


Figure 7. Overlay baseline-corrected TIC chromatograms of the test mixture analyzed at four concentration levels (1000 to 125 ppm, i.e., C to C/8), with three replicate injections per concentration.

PCA was performed to visually assess reproducibility of the four concentration levels, based upon the clustering behavior and analytical reproducibility across replicate injections. PCA scores showed clear clustering according to dilution level in Figure 8. PC1 accounted for 99.53% of the total variance and provided complete separation of the concentration groups. Samples shifted progressively toward lower PC1 scores with increasing dilution, indicating that concentration was the dominant source of variation in the data. PC2 explained only 0.24% of the variance and contributed minimally to sample discrimination. Replicate samples clustered closely within each concentration level, demonstrating good analytical reproducibility. The absence of overlap between groups indicates that the selected variables effectively distinguished the four dilution levels.

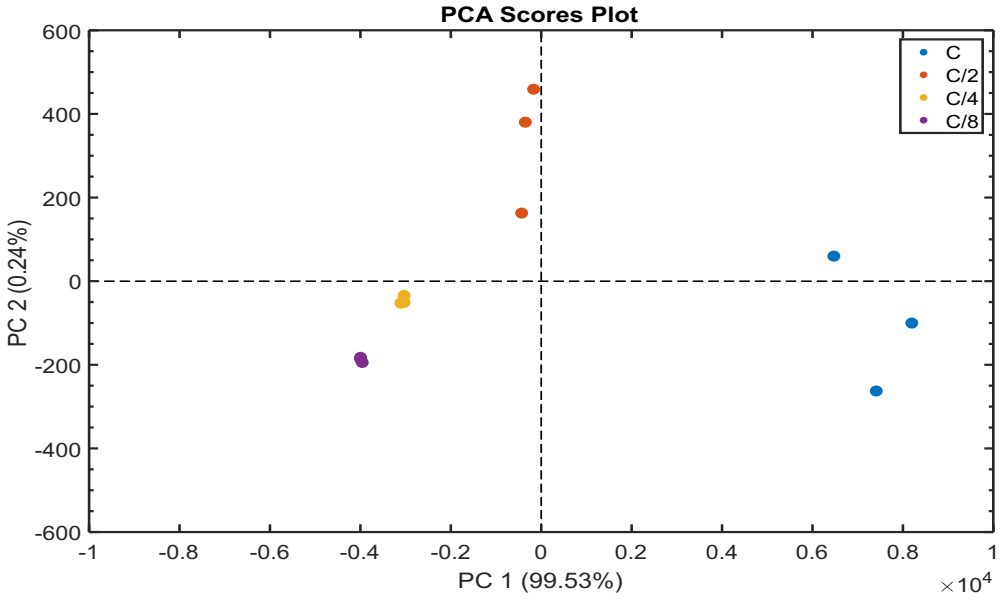


Figure 8. PCA scores plot for reproducibility study. Baseline-corrected total ion chromatogram (TIC) data of the complex test mixture across four concentrations (1000 to 125 ppm, i.e., C to C/8), with three replicates. % variance is 99.53% captured in PC1 and 0.24% captured in PC2.

A PCA loadings plot shows the contribution of each variable to a specific principal component and identifies the variables responsible for the sample clustering and separation observed in the corresponding scores plot. In Figure 9A, the loadings plot for PC1 showed

that only a subset of variables exhibited large positive loading values, indicating that these variables were primarily responsible for the concentration-dependent separation observed in the PCA scores plot (Figure 8). Samples with higher concentrations exhibited higher PC1 scores due to greater contributions from these variables, whereas progressive dilution resulted in lower PC1 scores. In contrast, in Figure 9B, PC2 explained only 0.24% of the total variance and displayed comparatively small loading values distributed around zero, suggesting that it captured minor residual variation such as background noise rather than systematic concentration-dependent effects.

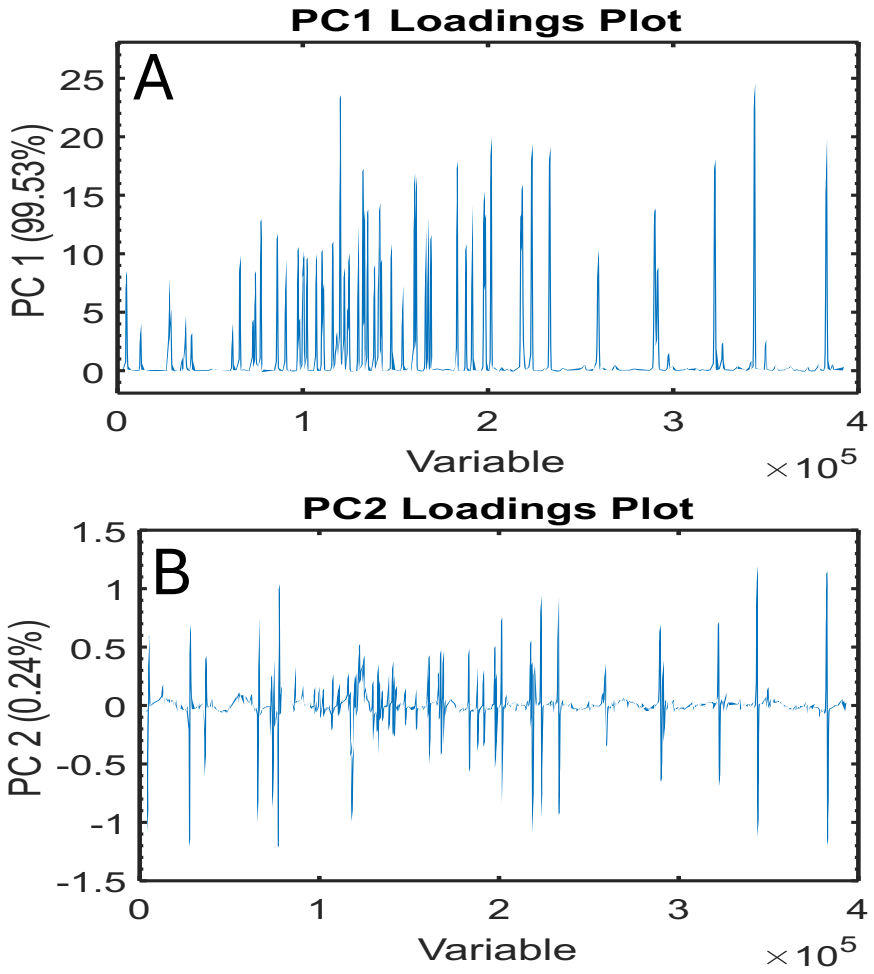


Figure 9. PCA loading plot for reproducibility study. (A) PC1 loadings plot with 99.53% variance captured. (B) PC2 loadings plot with 0.24% variance captured only background noise.

The *LOD* of the Intuvo 9000 GC was evaluated using the *S/N* ratio method. Figure 10 shows the overlay chromatogram of serially diluted samples from C to C/128 (1000 ppm to 7.8 ppm). The analyte peak at ~7.18 min remained clearly detectable at C/128, corresponding to a concentration of 7.8 ppm, and was still distinguishable from the baseline noise.

The *S/N* was calculated as:

$$S/N = \frac{\text{Peak Height}}{S_{\text{Noise}}}$$

where S_{Noise} is the standard deviation of the baseline noise. The *LOD* was then estimated:

$$LOD \text{ define as when } S/N = 3$$

$$LOD \text{ is when the peak height} = 3 \times S_{\text{Noise}}$$

The *LOD* of the peak at 7.18 min:

$$S_{\text{Noise}} = 0.021$$

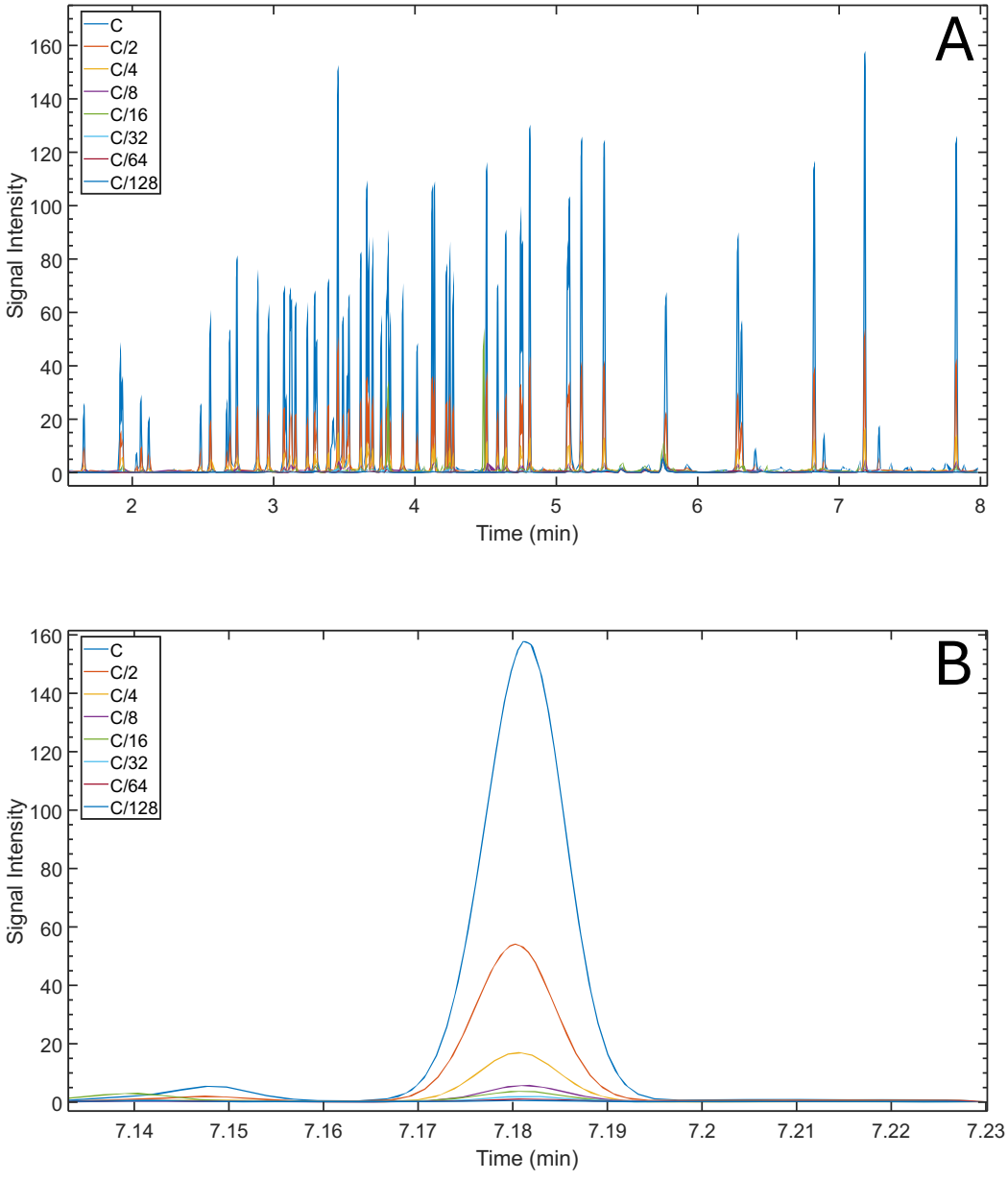
$$LOD \text{ of the peak height} = 0.063$$

$$\text{Peak height at 7.8 ppm} = 0.42$$

$$LOD \approx 1.2 \text{ ppm}$$

The *LOD* was estimated using a *S/N* criterion of 3. For the peak eluting at 7.18 min, the standard deviation of the baseline noise was measured as 0.021 units, corresponding to a minimum detectable peak height of 0.063 units. At the lowest concentration analyzed (7.8 ppm), the peak height was 0.42 units, which was substantially above the detection threshold and more than six times greater than the minimum detectable signal. Assuming a linear relationship between concentration and peak height, the *LOD* for this compound was estimated to be approximately 1.2 ppm. This result demonstrates that the Intuvo 9000 GC system is capable of detecting the analyte at low ppm concentrations and that the peak remained readily detectable at the lowest concentration evaluated in this study. Overall,

Figure 11 shows illustration of the *LOD* determination using the lowest concentration peak at 7.18 minutes.



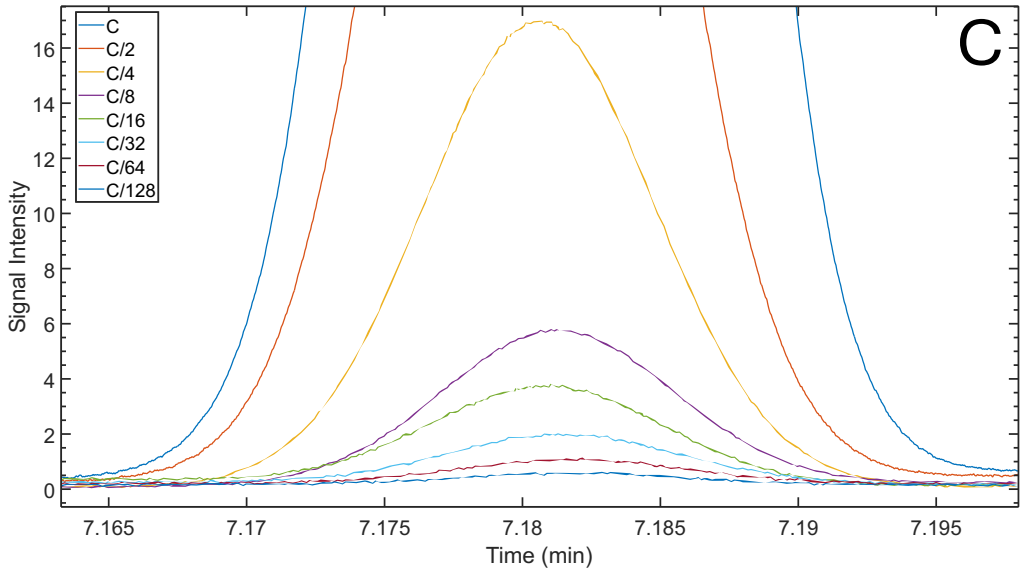


Figure 10. Overlay chromatogram with lower concentration from 1000 ppm to 7.8 ppm (C to C/128). (A) Overlaid chromatogram from 1 to 8 min. (B) Zoomed in peak at 7.18 min. (C) Further zoomed in peak at 7.18 min to see the lowest concentration of complex test mixture.

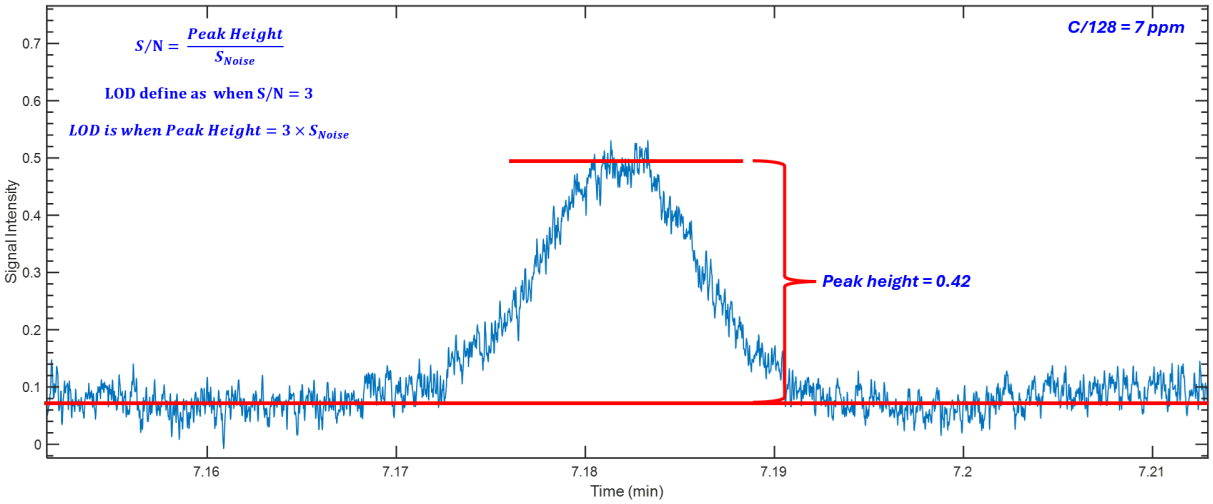


Figure 11. Illustration of the *LOD* determination. Using the lowest concentration (7.8 ppm) peak at 7.18 min.

The Intuvo 9000 GC was able to detect compounds at very low concentrations. However, the Intuvo 9000 GC system used in this study was equipped with a flame ionization detector (FID), which has limitations for compound identification. Although a list of the complex test

mixture was available, the FID could not provide structural information, molecular weight information, or library matching for compound confirmation. Compound identification mainly relies on retention time, which can make confirmation difficult when compounds co-elute or have similar retention behavior. In addition, unknown compounds could not be identified using the FID alone. To overcome these limitations, GC-qMS analysis was performed.

1.3.2 Compound Identification using GC-qMS

Due to the limited compound identification capability of the Intuvo 9000 GC system, GC-qMS analysis was used to enable compound identification in the samples. Figure 12 shows the chromatogram of the diluted test mixture, zoomed in over the 2–40 min range. Although the diluted test mixture concentration is relatively low (~16.7 ppm), the GC-qMS analysis still shows clear peak separation and well-resolved signals.

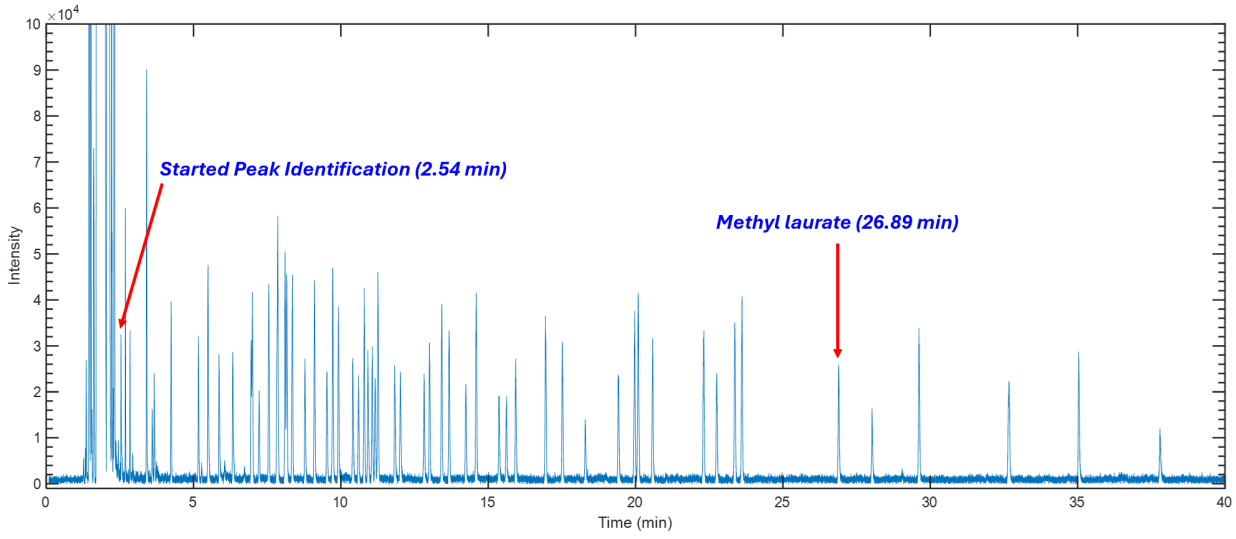


Figure 12. Baseline-corrected chromatogram of the diluted test mixture. From 0–40 min, zoomed in to visualize smaller peaks.

Mass spectra for MV calculation were usually extracted at the peak maximum. To calculate the MV, we used the same algorithm developed by Stein and Scott [9]. However, background noise in the extracted ion chromatogram (XIC) negatively affected the quality of the extracted mass spectra, resulting in MVs below 800. To illustrate this effect, the peak at 26.89 min, identified as methyl laurate, was selected as an example. Figure 13 shows the NIST library mass spectrum of methyl laurate, which was used as the reference spectrum for MV calculation. As shown in Figure 14, poor MVs were obtained when only the mass spectrum at the chromatographic peak apex was used, indicating that background noise significantly influenced the spectral quality.

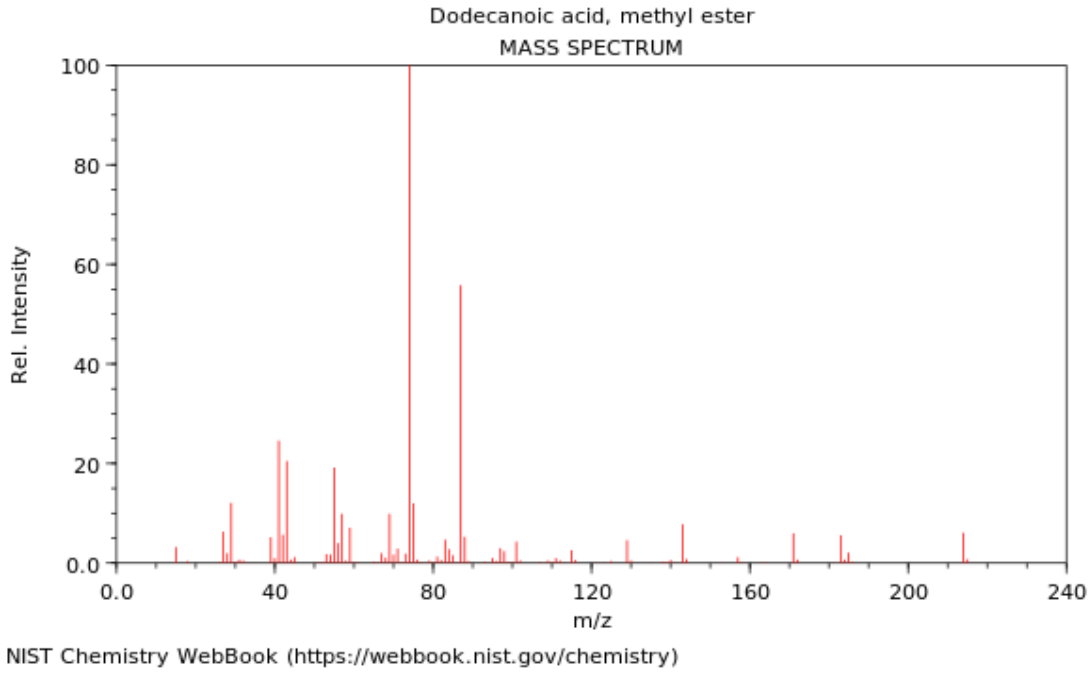


Figure 13. Reference mass spectrum of methyl laurate (methyl ester dodecanoic acid) from the NIST library.

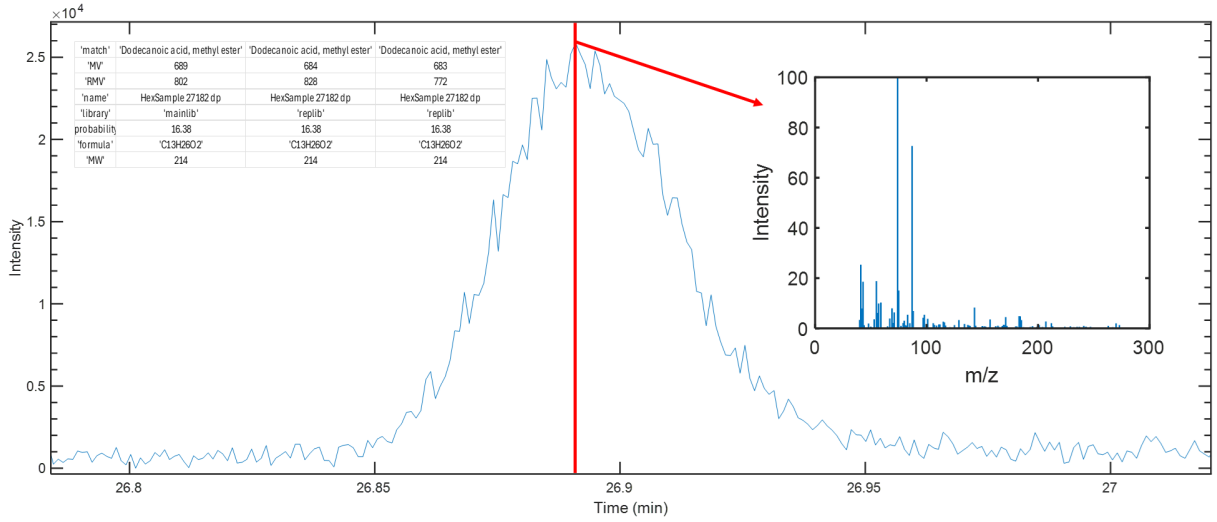


Figure 14. Mass spectrum of the compound eluting at 16.89 min at peak maximum, with NIST library match identification. Match value of 689 compared to reference mass spectrum of methyl laurate.

To solve this problem, different approaches were tested to obtain a cleaner mass spectrum. Figure 15 shows the procedure used to obtain the clean mass spectrum. Illustrates the approach used to improve mass spectral quality for compound identification. The TIC chromatographic peak at 26.89 min, corresponding to methyl laurate, is shown together with mass spectra acquired from three individual scans across the chromatographic peak (MS1–MS3). Although the spectra exhibit the same major fragment ions, differences in the relative abundances of low-intensity ions are observed due to scan-to-scan variability and background noise. The consistency of the spectra across the peak was evaluated to determine whether deskewing was necessary. As shown in Figure 16, spectra extracted from different scans show nearly identical fragmentation patterns, with high pairwise MVs 994 for MS1 vs MS2 and 998 for MS1 vs MS3, indicating insignificant spectral distortion across the chromatographic peak.

Despite this consistency, single-scan spectra often produced library MVs below 800, reflecting the influence of noise on low-abundance ions. To improve spectral quality, mass spectra from multiple scans across the chromatographic peak were summed and subsequently normalized to the base peak intensity. This summation increases the contribution of ions consistently associated with the analyte while minimizing the impact of random noise present in individual scans. The resulting averaged spectrum more closely matches the NIST reference spectrum of methyl laurate and has higher library MVs, enabling more reliable compound identification. Since no significant spectral skewing was observed, deskewing was not required, allowing direct summation of scans across the peak. After extracting the mass spectrum, signals below 1% or 2% of the base peak intensity were removed to reduce background noise. Figures 17 and 18 show the mass spectrum of methyl laurate peak at 26.89 min using different thresholds. These spectra were searched with the NIST library using MATLAB to get the compound identification and MV.

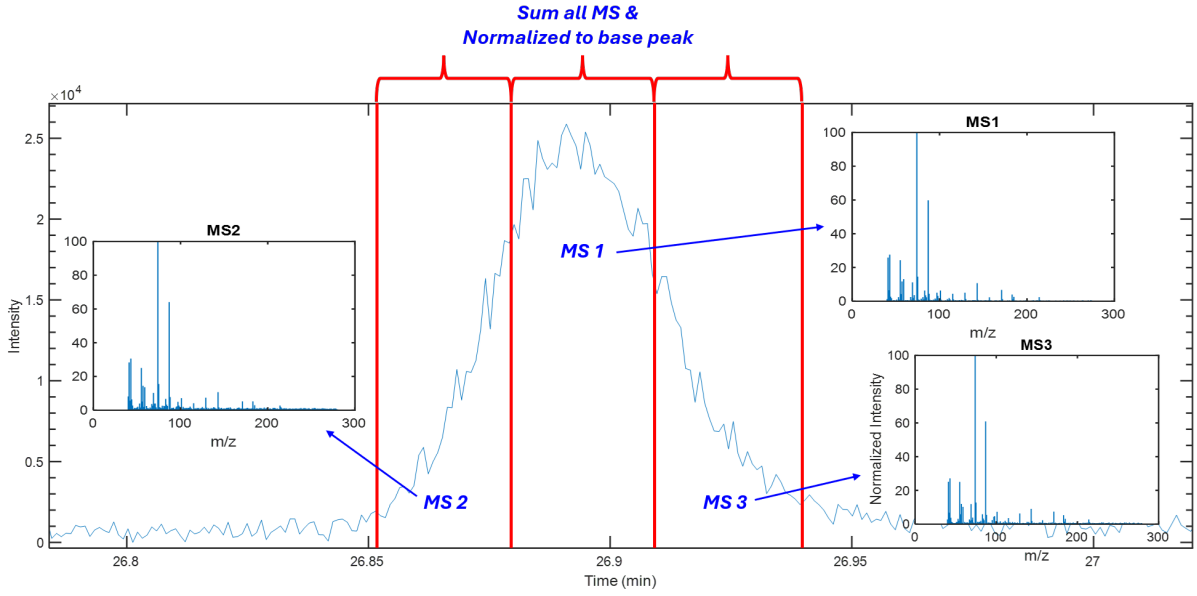


Figure 15. Illustration to obtain a clean mass spectrum. By summing MS scans across the peak region and normalizing it to the base peak.

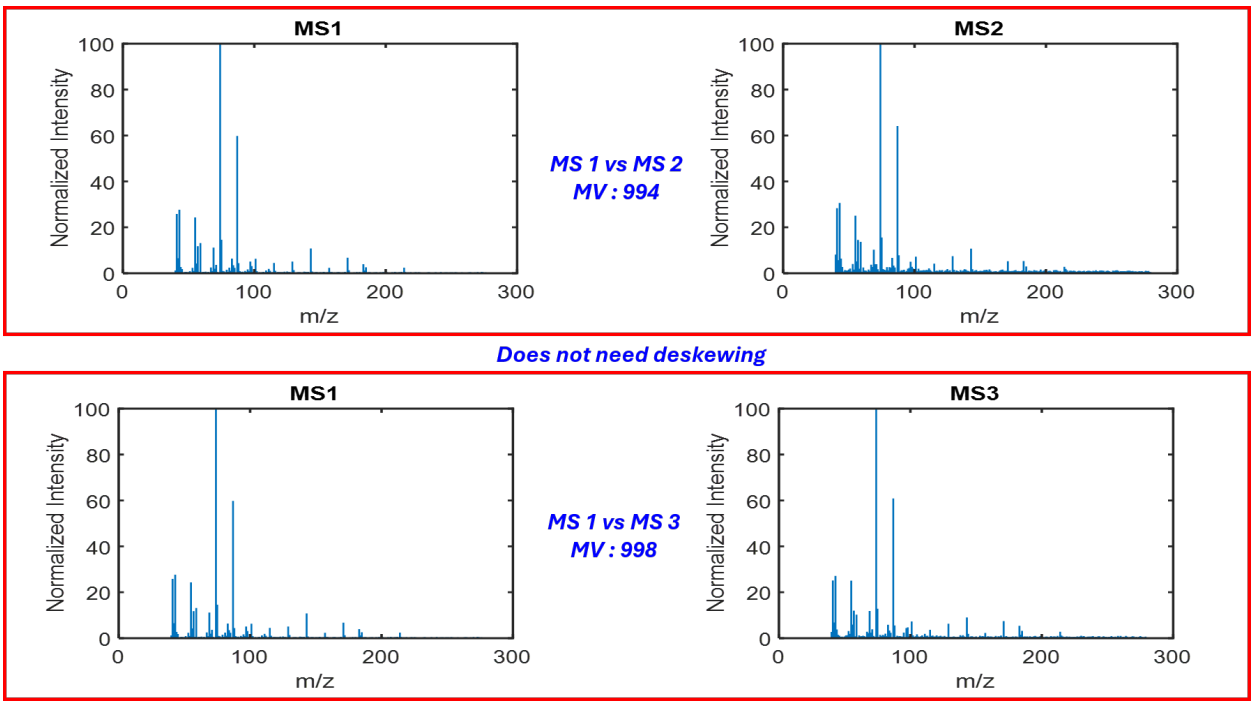


Figure 16. Illustration of mass spectral deskewing check. By comparing spectral similarity between MS1–MS2 and MS1–MS3.

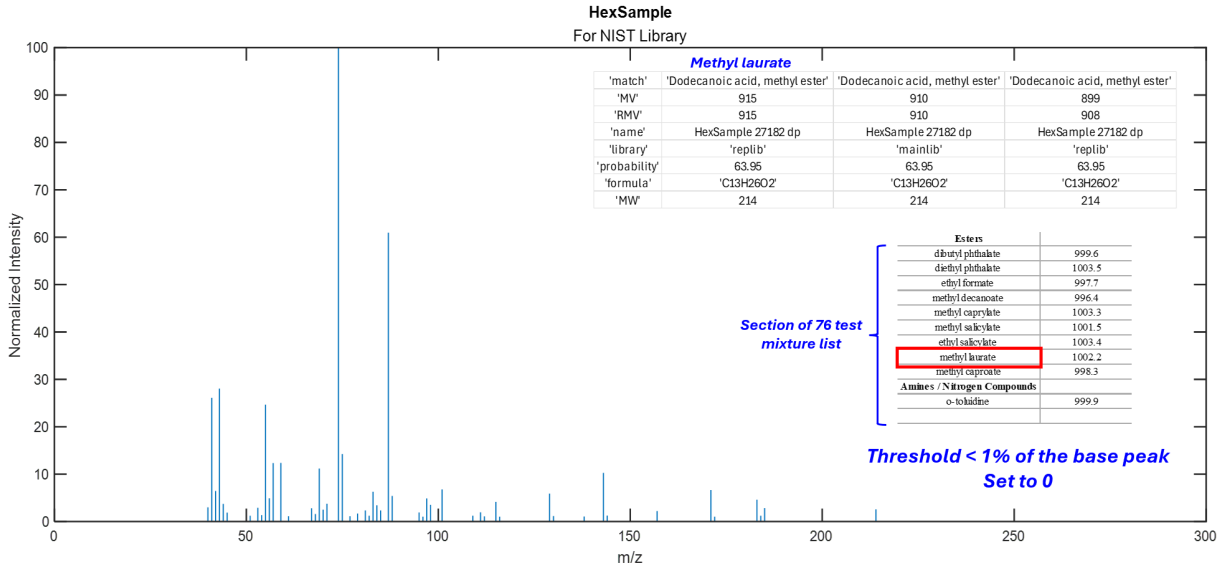


Figure 17. Mass spectrum of the methyl laurate peak obtained using a 1% base-peak intensity threshold. The NIST library MV is 915.

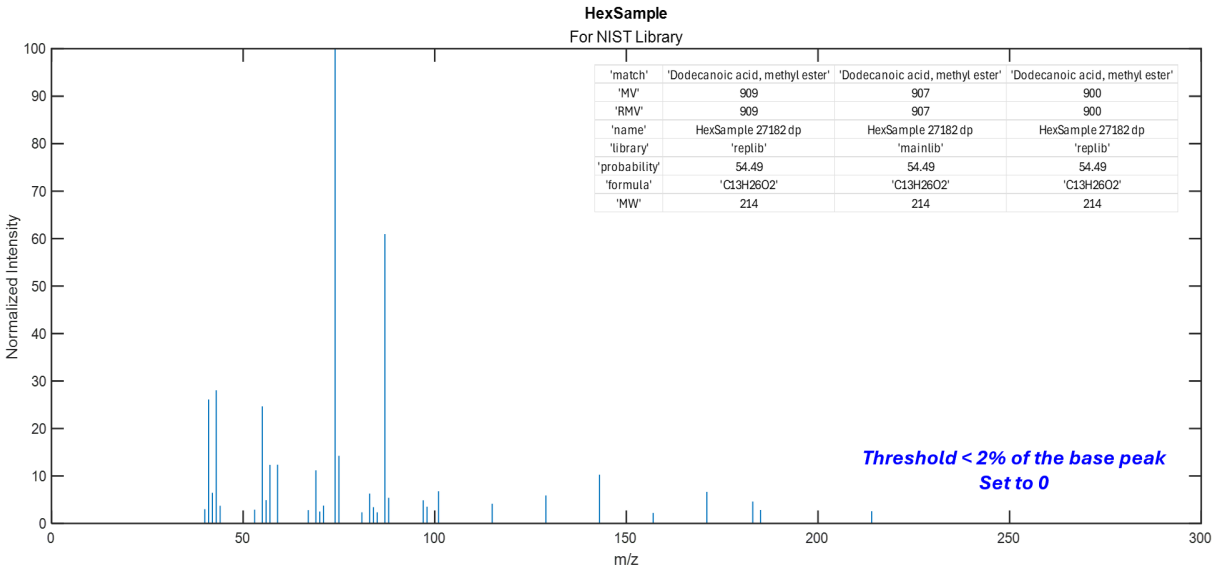


Figure 18. Mass spectrum of the methyl laurate peak obtained using a 2% base-peak intensity threshold. The NIST library MV is 909.

Using the 1% threshold produced a MV of 915 and identified the compound as methyl ester dodecanoic acid (methyl laurate). Methyl laurate was one of the compounds included in the complex test mixture. Using the 2% threshold gave a MV of 909 with the same compound identification. In contrast, the mass spectrum extracted only at the peak maximum produced a MV of 689 (Figure 14). These results show that averaging the full spectrum across the chromatographic peak and applying an intensity threshold improved the MV and spectral quality.

This process was applied across the entire chromatogram to identify as many of the compounds in the diluted test mixture as possible. Table 4 shows the peak identifications and MVs obtained using the 1% and 2% intensity thresholds. We could identify 59 compounds using 1% intensity threshold and 58 compounds using 2% intensity threshold.

Table 4. Peak Identification with Match Value Calculation.

Alkanes	Retention Time (min)	Match Value (TH < 1%)	Match Value (TH < 2%)
heptane	2.69	829	823
octane	4.24	810	812
decane	9.92	863	858
undecane	13.42	876	879
dodecane	16.94	903	883
tetradecane	23.61	899	880
hexadecane	29.62	868	877
pristane	32.67	862	852
octadecane	35.03	860	
Alkenes			
1-undecene	13.01	881	928
Alkynes			
1-heptyne	2.84	895	926
5-decyne	11.07	867	887
Cycloalkanes			
cyclooctane	7.00	879	904
butylcyclohexane	10.79	888	869
bicyclohexyl	19.97	904	909
Aromatics			
toluene	3.41	877	862
3-ethyltoluene	8.11	929	926
4-ethyltoluene	8.16	932	928
mesitylene	8.35	923	921
butylbenzene	11.26	942	935
tert-butyl benzene	9.10	929	916
1-propylbenzene	7.86	908	892
1-ethylnapthalene	22.32	915	909
cyclohexylbenzene	20.09	906	912
diphenylmethane	23.37	915	923
sec-butylbenzene	9.72	938	914
1,2,3,4-tetrahydronapthalene	14.60	932	926
Terpenes			
(-)-alpha-pinene	7.55	941	927
(-)-linalool	12.82	852	903

Alcohols			
2-pentanol	2.54	817	828
2-heptanol	6.33	850	862
1-octanol	11.83	844	879
1-nonanol	15.36	851	894
2-ethyl-1-hexanol	10.41	907	890
Ketones			
2-hexanone	3.66	903	905
2-heptanone	5.86	922	911
3-octanone	8.78	901	925
2-decanone	15.93	889	863
2-undecanone	19.41	889	853
2-dodecanone	22.76	879	841
acetophenone	10.92	939	957
butyrophenone	17.52	932	908
Esters			
dibutyl phthalate	37.80	873	909
diethyl phthalate	28.02	915	922
methyl decanoate	20.58	904	892
methyl caprylate	13.67	922	913
methyl salicylate	15.62	910	939
ethyl salicylate	18.29	894	907
methyl laurate	26.89	892	905
methyl caproate	6.95	941	945
Amines / Nitrogen Compounds			
o-toluidine	11.16	937	941
Sulfur Compounds			
2-(methylthio)-pyridine	12.01	848	881
1,4-Oxathiane	5.49	671	676
Halogenated Hydrocarbons			
1,5-dichloropentane	9.53	922	929
1-chlorohexane	5.17	922	933
1-bromohexane	7.23	910	934
1-bromoheptane	10.60	884	904
1-bromooctane	14.24	863	907

Chapter 2. Investigating the use of Activated Carbon Strips for Fragrance Residue Recovery, Storage, and Extraction Sample Prep for GC-MS Analysis in Forensics Fingerprinting Applications

2.1 Introduction

Accurate, reliable, and user-friendly forensic testing of chemically related substances is critically important in real-world investigations. Chemical evidence is often encountered in trace amounts and under conditions that are far from ideal, requiring analytical methods that can provide definitive information from potentially limited or degraded samples [10,11]. In many cases, such evidence must be collected at crime scenes, transported, stored, and analyzed at a later time, increasing the risk of chemical alteration or loss [12]. Consequently, forensic methodologies must not only deliver high analytical sensitivity and specificity but also be robust, practical, and adaptable to field conditions [13]. The development of workflows that preserve evidential integrity while remaining feasible for routine forensic use remains an ongoing challenge across many areas of forensic chemistry.

Fragrance residues (from e.g., perfumes, colognes, and similar products) deposited on fabric presents a significant forensic challenge because they consist largely of volatile and semi-volatile organic compounds that can easily evaporate, redistribute, or degrade after deposition [14]. Once transferred onto textile substrates, these compounds are influenced by fabric composition, airflow, temperature, environmental exposure, and time. As a result, the original chemical profile may change substantially between deposition and laboratory analysis. While gas chromatography coupled with mass spectrometry (GC–MS) is capable of detecting fragrance components, the reliability of the resulting chemical fingerprint depends strongly on how the sample is handled prior to analysis – which includes sample collection, transport, storage, and sample preparation right before analysis. Experimental studies have shown that fragrance

compounds can transfer between fabrics and remain detectable under controlled conditions, supporting their potential forensic value [15]. However, these studies also show that recovery efficiency depends on sampling configuration, sorbent properties, headspace volume, and exposure time [16,17]. Small procedural differences can alter compound-to-compound concentration ratios and reduce low-abundance compounds to levels in which their signal may no longer be detectible, particularly for highly volatile components. In addition, matrix interactions and adsorption/desorption variability may affect reproducibility, making it difficult to ensure that the recovered sample chemical profile accurately represents the original deposited fragrance [18]. Therefore, the primary challenge is not simply the analytical detection of fragrance compounds, but the preservation of a chemically representative and stable fingerprint from the moment of deposition through storage and following analysis [14]. Developing practical collection and storage strategies that minimize compositional change while remaining suitable for routine forensic use is essential for improving the evidential reliability of fragrance traces [18,19].

To address these challenges, there is a need for analytical workflows specifically designed to preserve fragrance chemical fingerprints from the point of deposition all the way through subsequent analysis. Such workflows must account for compound volatility, environmental exposure, adsorption/desorption processes, and storage conditions that may alter the sample chemical fingerprint over time [19,20]. An effective approach should enable consistent recovery of representative analytes while minimizing degradation and background interference, thereby supporting reliable comparisons between samples collected under different circumstances [20]. Developing strategies that integrate practical sample collection methods with

robust chemical analysis procedures is therefore essential for improving the forensic utility of fragrance evidence.

In this study, we explore an experimental approach designed to evaluate whether fragrance residues left on fabric can be collected, then sufficiently preserved, and then retrieved for chemical analysis at a later time. For the analysis of fragrance residues left on fabric, GC–MS should provide sufficient separation capability; however, the complexity of fragrance mixtures and the potential loss of low-abundance components during storage might complicate data interpretation [16]. For this exploratory study, we examined the use of activated carbon (i.e., charcoal) strips to extract a suitable chemical fingerprint of fragrance compounds whereby the fragrance was initially adsorbed to cotton cloth serving as the initial sample. We selected the activated carbon strips due to their promising use when implemented for long term chemical fingerprinting of arson-related samples [12]. Figure 19 illustrates the overall experimental workflow for our study, including sample preparation, activated carbon strip capture of sufficiently volatile fragrance compounds, and storage conditions. This design provides a framework for systematically investigating the possibility of preserving fragrance chemical fingerprints on fabric for later forensic examination without presuming analytical outcomes. Herein, we define three sample forms: neat samples, initial carbon strip extracts (I-CSE), and stored carbon strip extracts (S-CSE). The neat samples are simply the neat fragrance, either perfume or cologne, prior to being adsorbed on cotton cloth. The I-CSE are obtained by first adsorbing a quantitative volume of fragrance on cotton cloth, then placing the cotton cloth in a paint can, where vapor phase extraction occurs onto a carbon strip section, followed by removal of the carbon strip section after 60 min (defined to be the *initial* time point) and extraction with methanol to form the I-CSE for GC-MS analysis. For the S-CSE sample prep, the same steps are

performed as with the I-CSE sample prep just described, except the vapor phase extraction step occurs over a much longer time period of 2.5 months using freezer storage for each paint can containing the cotton cloth sample with adsorbed fragrance and suspended carbon strip(s). We anticipate that short term storage under ambient conditions versus long term storage under freezer conditions should provide insight into potential compositional changes of the observed GC-MS chemical fingerprint for the various fragrances. Overall, our chemical fingerprint capture method using activated carbon strips facilitates an exploratory insight into several adsorption/desorption processes that can all occur: (1) adsorption of fragrance compounds from the neat fragrance solution to a suitable fabric substrate (cotton cloth), (2) vapor phase transfer of each compound from the fabric to a suspended activated carbon strip during storage in a paint can (either short term of long term under freezer conditions), and (3) extraction of the chemical fingerprint residing on the carbon strip using methanol prior to GC-MS data collection.

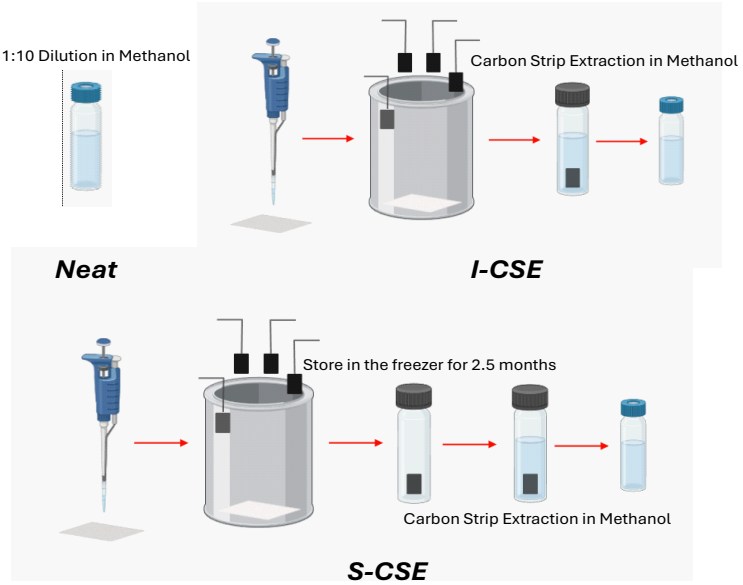


Figure 19. Illustration of sample preparation for neat, I-CSE, and S-CSE samples. Neat samples were diluted 1:10 in methanol prior to analysis. For I-CSE samples, fragrance was applied to cotton fabric and placed inside a sealed can, and four carbon strips were hanging inside the can for several hours to collect volatile organic compounds, followed by methanol extraction. S-CSE

samples were prepared using the same procedure, except that carbon strips were stored in sealed vials for 2.5 months in the freezer before the extraction.

By integrating a sampling method centered upon using activated carbon strip extracts, and GC–MS analysis with chemometric modeling and target-based peak evaluation, this work examines the feasibility of preserving fragrance chemical fingerprints on fabric for subsequent forensic analysis. In addition to GC-MS separation and detection, chemometric methods such as principal component analysis (PCA) will be applied to enhance classification and pattern recognition for the complex chemical dataset we will be study [21]. PCA enables dimensionality reduction while preserving meaningful variance in high-dimensional data, making it particularly suitable for GC-MS datasets composed of chromatographic and mass spectral variables [22]. However, when fragrance residues are subjected to environmental exposure or storage, analyte-specific signal reduction, background noise, and retention-time variability may influence the reliability of a TIC-based PCA classification approach. Beyond chromatographic profiling via PCA alone, we will explore a workflow based upon mass spectrum matching as an alternative approach to emphasize compound-specific spectrum features rather than signal intensity-based pattern matching approaches. In this context, peak alignment, spectral similarity assessment, feature filtering and signal threshold setting become critical steps to ensure meaningful comparisons across the samples collected and stored under different conditions. The use of structured feature selection and preprocessing strategies, including intensity thresholding and mean centering, is particularly important for high-dimensional and sparse datasets generated from concatenated mass spectral blocks. Match value (MV) calculations related to reference spectra provide a quantitative measure of spectral similarity and can be used to assess whether stored extracts retain chemically representative MS information [23]. In the present study, we

investigate whether fragrance residues left on fabric can be preserved and analytically retrieved using an activated carbon-based chemical fingerprint collection and storage workflow. Specifically, we anticipate using the S-CSE samples as potential reference targets for mass spectrum-based library comparisons using MVs, since the S-CSE chemical fingerprints should provide the *lowest common denominator* of analyte compounds that persist in the carbon strips for the 2.5-month freezer storage period.

2.2 Experimental

2.2.1 Samples Analyzed and their Preparation

A list of the 13 fragrance samples analyzed is provided in Table 5. An overview of the preparation of the three types of samples analyzed is presented in Figure 19. Initially, the neat fragrances were analyzed by GC-MS following 1:10 dilution in methanol (MeOH). For the preparation of the I-CSE and S-CSE samples, paint cans sized at ¼ pint (Amazon asin B0D7RLSJR4) were rinsed with methanol (MeOH) and dried in an oven. Sections of 100% cotton knit cloth (~0.5 g) were cut out from a package of T-shirts (Amazon asin B07GFPRG6D) and 100 µL of a given fragrance from local merchants was added to each fabric section. A parafilm barrier was used between the cotton and laboratory bench surface to prevent loss of sample volume, and the samples were left to air dry for ~ 5 min before being placed in their respective ¼ pint unlined and airtight paint cans. Each paint can was equipped with up to four carbon strips (product number A-1509) from Arrowhead Forensics (Lenexa, KS 66215) suspended in the headspace of a paint can via separate wires to ensure the carbon strips did not come into contact with one another. The carbon strip suspension wires were wedged between the paint can and its lid during the sealing process, after which the samples were placed in an oven at

75°C for 30 min. Each suspended activated carbon strip was transferred to an individual tube for storage. For the S-CSE samples, the tubes were stored in the freezer for 2.5 months. For the I-CSE samples, activated carbon strips in individual tubes were submerged in 500 µL of MeOH and 500 µL ethyl ether prior to analysis. After reaching room temperature, the tubes were vortexed for 30 s before 1 mL of the sample was transferred into its respective GC vials. The same preparation was applied to the S-CSE samples.

Table 5. 13 fragrance samples used with sample number and sample name.

Sample Number	Sample Name
113	Nest Midnight
116	Salvatore Ferragamo Tuscan
117	Kate Spade (KS) Truly Daring
118	KS Truly Dazzling
119	Burberry
121	Clubman Pinaud
122	Old Spice Classic
123	Aqua Velva Musk
124	Axe Phoenix
125	Vince Camuto Fiori
126	Ascent Playful Citrus
127	The Good Scent Mama
130	Calvin Klein (CK) Contradiction

2.2.2 GC-MS Parameters

Samples were analyzed using a GC-MS instrument comprised of an Agilent (model 6890) gas chromatograph with a model 5973-mass selective detector and a CTS Analytics CombiPAL autoinjector. All samples were introduced using an injection volume of 1 µL. The GC instrument was fit with a Restek Rtx-5MS column (catalog #12626) that is 60 m in length, an i.d. of 250 µm and film thickness of 0.25 µm. The GC oven temperature program began at 40°C

(held for 5 min), followed by a linear ramp of 10°C/min to a final temperature of 300°C over a 31-min period. The scan rate was 20 Hz over a *m/z* range of 50 to 650.

2.2.3 Data Processing

All GC-MS data were processed using MATLAB R2024a (Mathworks, Inc., Natick, MA, USA). After running the 13 fragrance samples 3 replicates of neat samples, 4 replicates of I-CSE, and 2 replicates of S-CSE were analyzed, comprising a total of 117 GC-MS sample runs. We initially focused on the total ion current (TIC) chromatograms. The TIC chromatograms were cropped to the time range to 14–28 min to remove late-eluting background noise and to focus the analysis on chromatographic regions containing meaningful analyte peaks. The data were then baseline-corrected using a rolling-ball minimum algorithm and normalized to the total TIC area [24]. The preprocessed TIC data were then used as X-block variables for PCA using PLS Toolbox 9.3 (Eigenvector Research, Manson, WA, USA).

For the MS–based workflow, a comprehensive peak table was first made using the GC-MS data from the neat samples. Peaks were selected from the TIC chromatograms using an intensity threshold of 30,000 to exclude low-intensity features that could adversely impact the quality of the final results, especially for the S-CSE samples that were found to undergo the largest loss of signal for many of the analytes in their chemical fingerprint relative to the neat samples, *vide infra*. The retention times of the peaks in the comprehensive peak table from the neat samples were applied to the GC-MS data for the I-CSE and S-CSE samples to ensure consistent analyte peak alignment across all sampling conditions, thereby minimizing errors arising from retention-time shifts and sampling variability. For each aligned chromatographic peak, mass spectra were extracted over the *m/z* range 50–300, since no *m/z* above *m/z* 300 were observed. Match values were calculated using the corresponding neat-sample spectra as

references, providing a measure of spectrum similarity across all three sampling conditions [25]. Only analyte peaks with a MV > 800 were retained as high-confidence fingerprinting features.

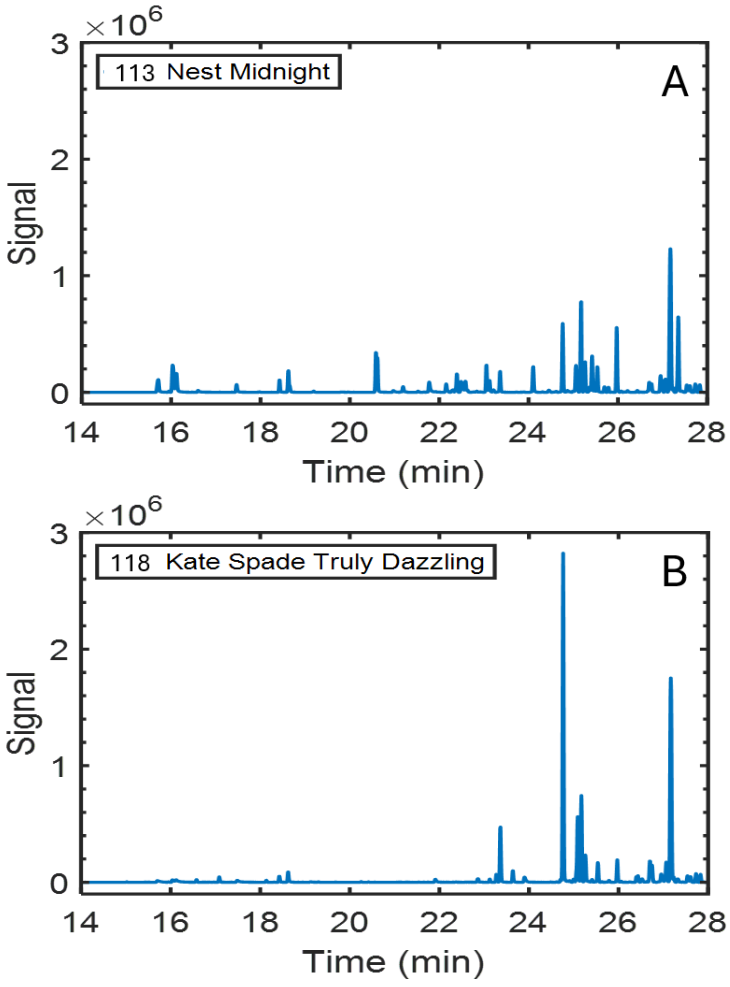
In terms of the specifics for the mass spectrum processing, it was determined that the random noise contribution at all *m/z* necessitated setting an intensity threshold of 600. Thus, this threshold was applied to the extracted mass spectra, with any *m/z* intensity was set to 0 if its intensity < 600. Mass spectrum signal thresholding substantially improved the spectrum quality for MV calculations by removing low-intensity noise while preserving meaningful *m/z* contributions. Each mass spectrum was then normalized to its base-peak intensity by scaling to 100 to facilitate comparing the mass spectrum for a given analyte peak across several sample runs even though the analyte concentration generally varied.

After MV-filtering, a final set of *lowest common denominator* analyte peaks was retained to serve as a target for subsequent analysis based on the lowest number of peaks observed among the S-CSE samples. The mass spectra corresponding to the selected peaks were concatenated into a single feature vector for each sample for multivariate analysis. Then the preprocessed MS data were then used as a X-block variables for PCA following mean-center preprocessing. To visualize fragrance discrimination, the number of peaks in each sample with a MV > 800 threshold relative to the target S-CSE was determined and averaged across replicates. Peaks that did not satisfy the MV threshold were classified as missing peaks within the corresponding samples. Variability between replicates was represented using sample standard deviation error bars for both matching and missing peaks.

2.3 Results and Discussion

To get a sense of the chemical composition variety of these representative fragrances and determine reproducibility among replicates, Fig. 20 shows the total ion chromatograms (TICs) of

the neat samples for the four arbitrarily selected fragrances: 113 – Nest Midnight, 118 – KS Truly Dazzling, 122 – Old Spice Classic, and 127 – The Good Scent Mama. These TIC chromatograms are presented following baseline correction and normalization to their total area. Two normalization approaches were evaluated: normalization to internal standards and normalization to total area. Normalization to total area provided better reproducibility within replicates for the neat samples and was therefore selected for further analysis [26]. In Fig. 20D, with the TIC chromatogram for Sample 127 – The Good Scent Mama, there red arrows point to analyte peaks of interest which we discuss next to bring attention to the main issues with the overall efficiency of collecting, storing, and then analyzing the chemical fingerprint of these fragrances that were initially loaded onto cotton fabric.



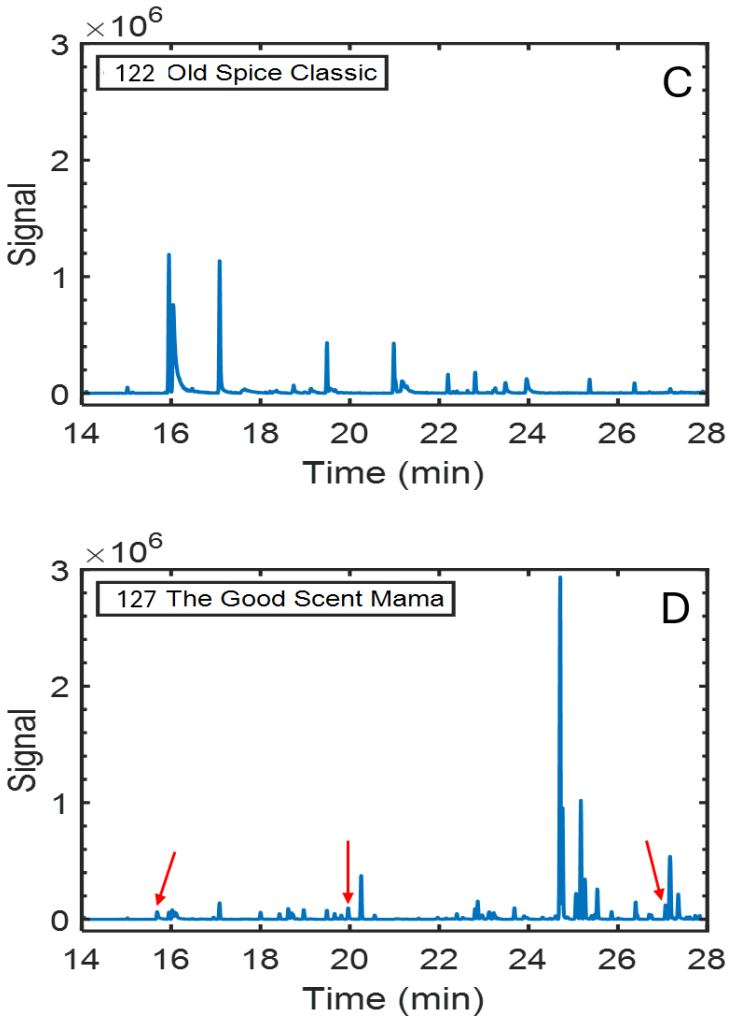


Figure 20. TIC chromatograms after preprocessing. (A) Sample 113 Nest Midnight neat replicate 1. (B) Sample 118 Kate Spade Truly Dazzling neat replicate 1. (C) Sample 122 Old Spice Classic neat replicate 1. (D) Sample 127 The Good Scent Mama neat replicate 1, with three red arrows indicating selected peaks.

Using overlays of zoomed-in regions of the TIC chromatograms for the three peaks in Sample 127 – The Good Scent Mama (Fig. 20D), the representative effects of activated carbon strip capture, storage, and solvent extraction prior to GC-MS analysis are readily visualized in Fig. 21. These overlays highlight differences in compound presence among the neat, I-CSE, and S-CSE samples and demonstrate storage effects with the activated carbon strips. Figure 21A

shows an analyte peak detected only in the neat sample, identified as 1,1′-oxybis-2-propanol with a NIST library MV of 923. Thus, 1,1′-oxybis-2-propanol represents compounds that were not sufficiently, if at all, captured by the activated carbon strip and/or were not eluted from the activated carbon strip via methanol extraction prior to GC-MS analysis. Figure 21B shows an analyte peak present in both the neat and I-CSE samples, corresponding to cis-4-(1-methylethyl)cyclohexanemethanol with a MV of 932. This result suggests that cis-4-(1-methylethyl)cyclohexanemethanol is sufficiently captured by the activated carbon strip and then extracted by methanol for the GC-MS analysis, however the absence of the cis-4-(1-methylethyl)cyclohexanemethanol peak in the S-CSE samples suggests that the compound was either not retained during storage or lost overtime [12]. In contrast, Fig. 21C shows an analyte peak observed in all three conditions neat, I-CSE, and S-CSE identified as 13-methyl-oxacyclotetradecane-2,11-dione with a MV of 889. The persistence of this peak indicates that it is relatively stable and can be effectively captured by the carbon strip even after storage [27]. Furthermore, this then means that 13-methyl-oxacyclotetradecane-2,11-dione and others with similar capture, storage, and methanol extraction behavior are excellent candidates to provide a persistent, long-term chemical fingerprint for forensics purposes.

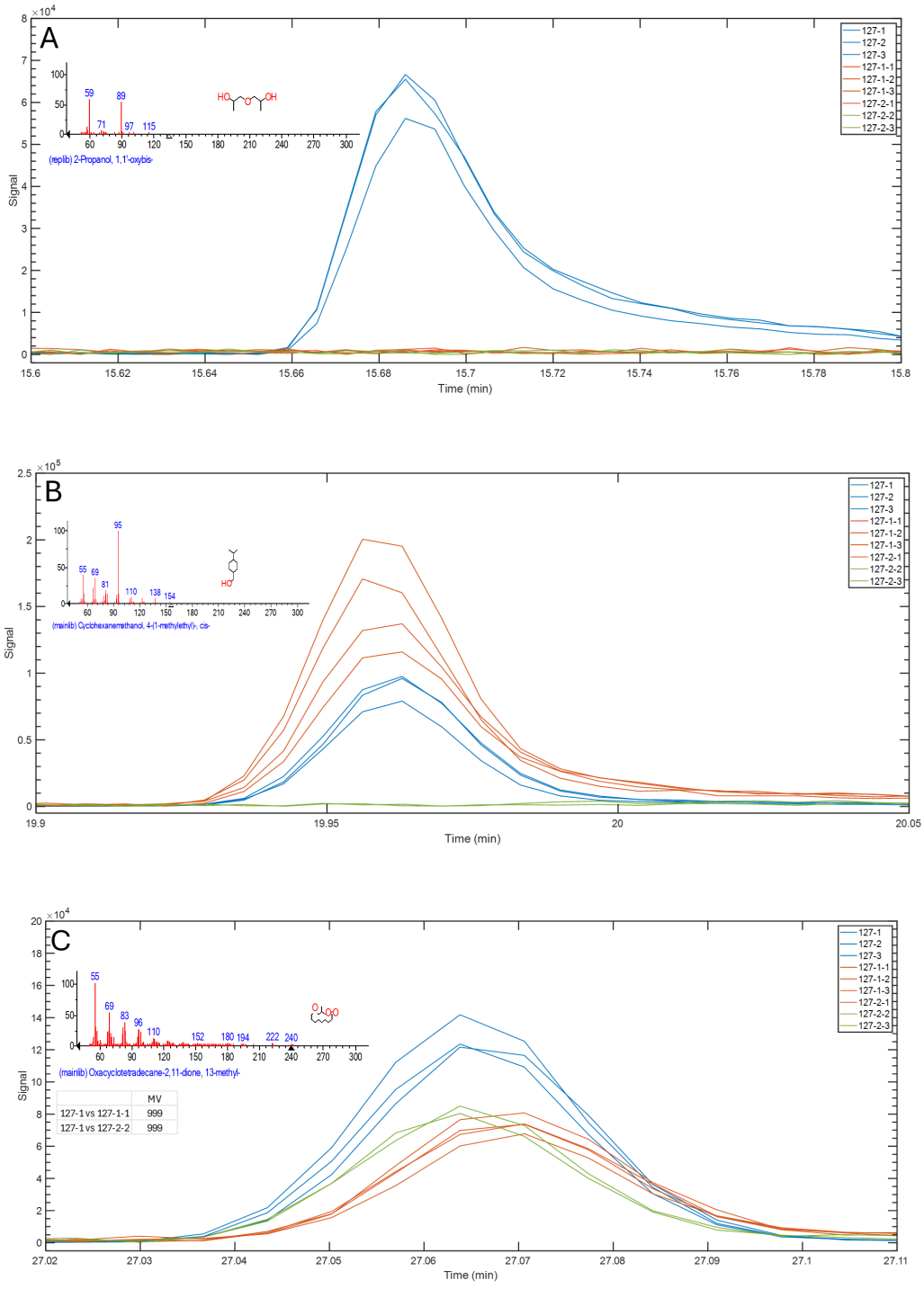


Figure 21. Overlay of neat, I-CSE and S-CSE samples zoomed-in chromatograms for three selected peaks in sample 127 "The Good Scent Mama". (A) Peak present only in the neat sample, with a library match value of 923 for 1,1′-oxybis-2-propanol. (B) Peak present in the neat and I-CSE samples, with a library match value of 932 for cis-4-(1-methylethyl)cyclohexanemethanol. (C) Peak presents in the neat, I-CSE, and S-CSE sample, with a library match value of 889 for 13-methyl-oxacyclotetradecane-2,11-dione.

To determine whether or not the TIC chromatograms alone could reliably classify the 13 fragrance samples in their three sample types (neat, I-CSE, and S-CSE), PCA scores plots were generated without preprocessing of the X-block. In Fig. 22A, the scores plot is presented for the 3 replicates of the 13 neat fragrances, with PC1 and PC2 accounting for 59.17% and 9.85% of the variance, respectively. Meanwhile, in Fig. 22B the scores plot for all samples is presented (3 neat replicates, 4 I-CSE replicates, and 2 S-CSE replicates), with a similar amount of variance captured on PC1 and PC2 of 56.22% and 10.09%, respectively. However, the sample class clustering per fragrance is substantially different in Figs. 22A versus 22B. The PCA scores plot for the neat-only samples (Fig. 22A) showed good reproducibility among replicates for a given fragrance sample (relatively tight clusters). When I-CSE and S-CSE samples were included in Fig. 22B, no clear clustering was observed. Most I-CSE and S-CSE samples were widely spread along PC1 and PC2, so the individual "clusters" for many of the fragrances overlapped with each other substantially. This behavior is attributed to signal reduction, compound loss, and increased background noise associated with the activated carbon strip capture process [13,27], consistent with the representative analytes discussed in Fig. 21. These results suggest that TIC chromatograms alone are insufficient for reliable fragrance classification and motivate the use of mass spectrum–based approach to compare and/or classify the fragrances. Our approach will take away the dependence on signal intensity to the extent that a given analyte peak must have *sufficient signal intensity* in the S-CSE samples to produce a high-quality mass spectrum for matching to any sample from the neat or I-CSE samples. As an example, going back to Fig. 21, only 13-methyl-oxacyclotetradecane-2,11-dione with a MV of 889 in Fig. 21C is a suitable candidate for a chemical fingerprint marker compound with potential forensic utility. Thus, the

next step in our study overall was to work our way through all 13 fragrances to find the subset of analyte peaks that behave like 13-methyl-oxacyclotetradecane-2,11-dione in Fig. 21C.

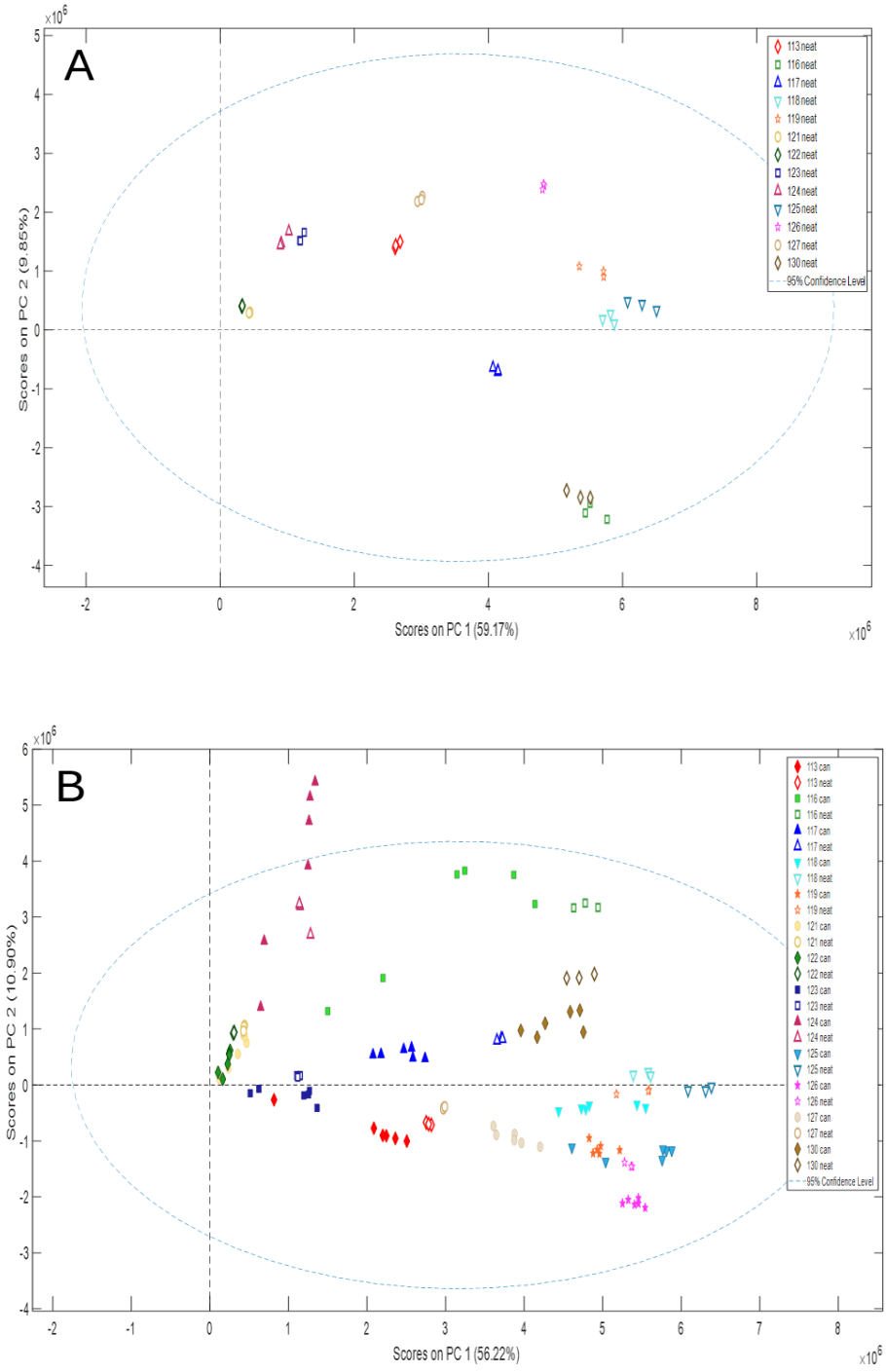
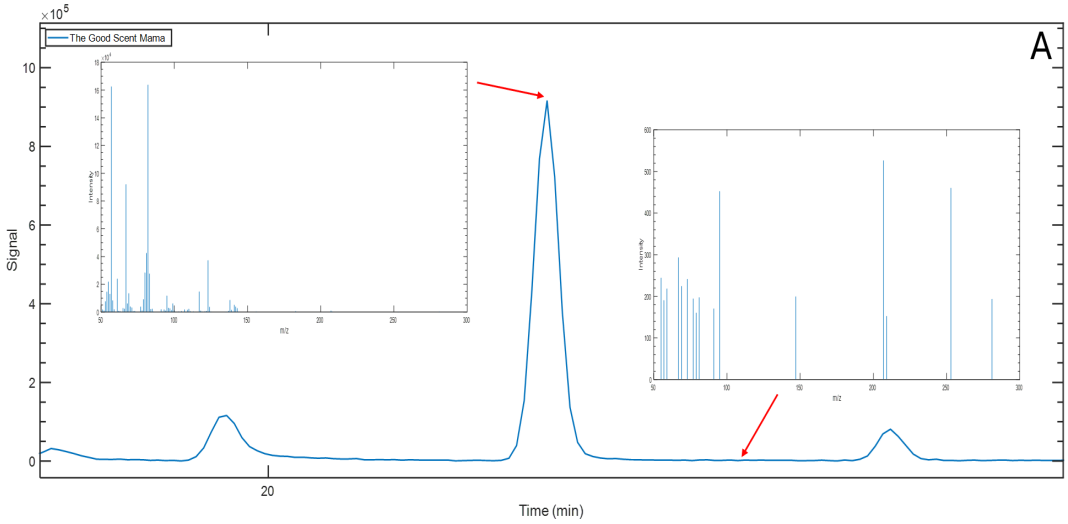
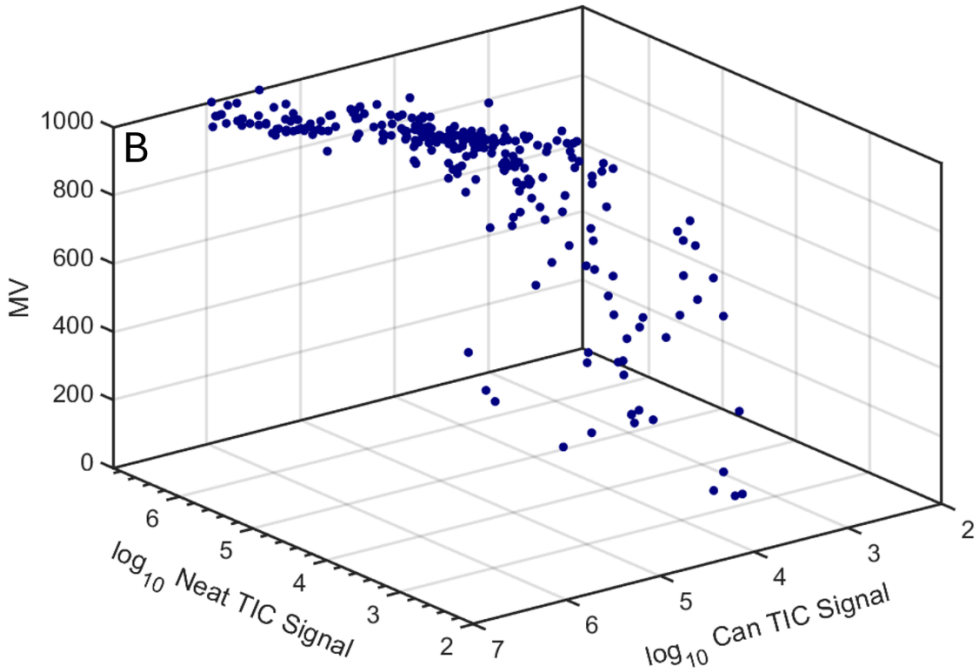


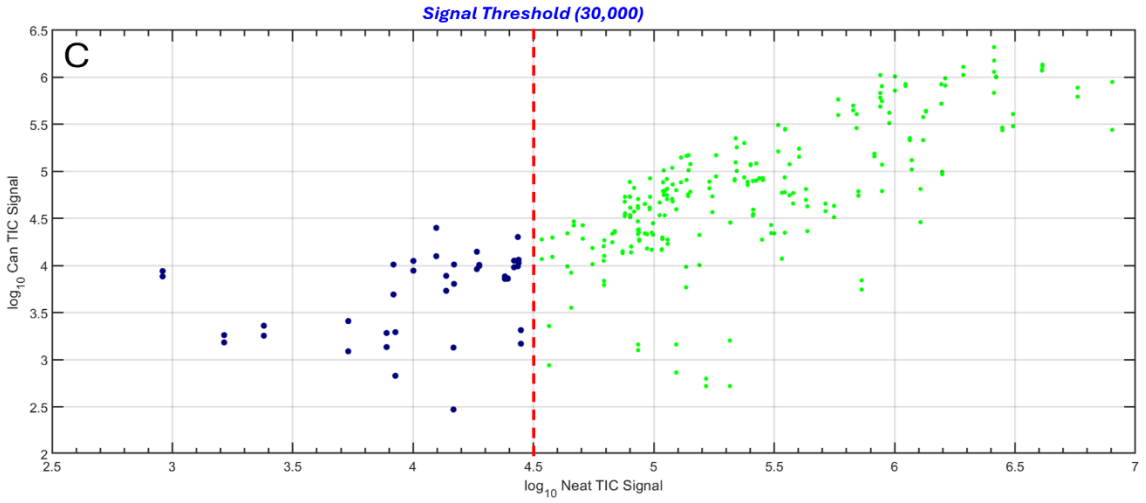
Figure 22. PCA scores plots based on TIC data. (A) PCA scores plot of neat samples. (B) PCA scores plot of neat, I-CSE, and S-CSE sample.

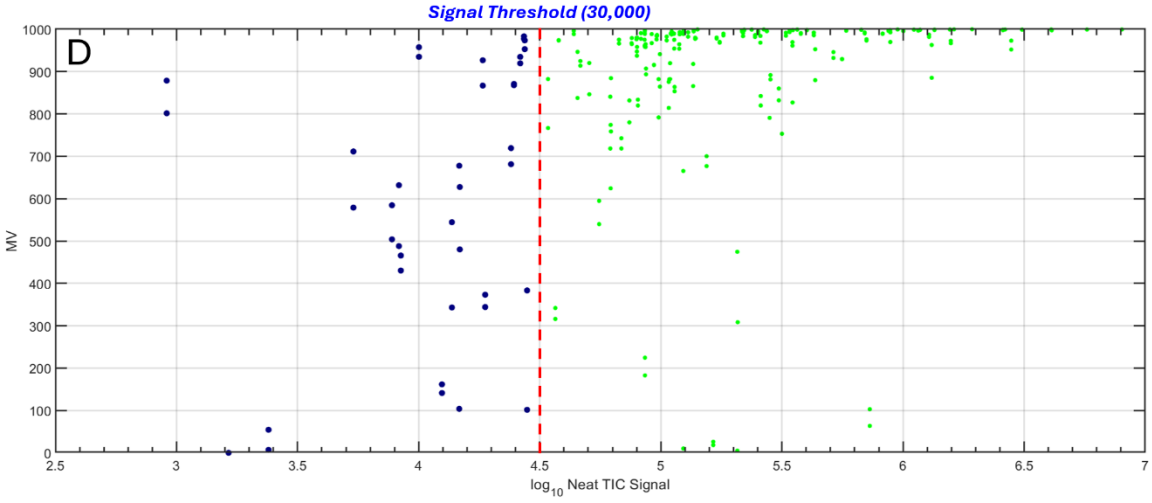
In order to improve data quality and to ensure meaningful feature selection, the intensity thresholds for data filtering the mass spectra and TIC data were evaluated in the Fig. 23. Figure 23A compares a typical analyte peak with a surrounding baseline "noise" region. Specifically, a representative mass spectrum for baseline noise (right inset spectrum) shows the presence of noise "spikes" that were observed with the GC-MS used. Several similar baseline spectra were also evaluated, and we arrived at selecting a signal intensity threshold of 600 to effectively separate meaningful analyte signal in the mass spectra from background noise. Application of this signal threshold is as follows. Once a spectrum of an analyte peak is obtained from the GC-MS data, all signals in its mass spectrum that were below the threshold of 600 were set to zero. For example, for the analyte peak spectrum in Fig. 23A (left inset spectrum), application of this signal threshold removed 24 ions that were initially in the in-set spectrum. While the analyte peak in Fig. 23A is of high *S/N*, for lower *S/N* analytes, application of this threshold of 600 substantially improved the mass spectrum quality and thus the MV obtained when comparing across samples. Additionally, Fig. 23B–E shows the relationship between MV and TIC peak maximum signal intensity for all 54 analyte peaks shown in Table 7 across four fragrances (sample 113 Nest Midnight, sample 118 Kate Spade Truly Dazzling, sample 122 Old Spice Classic, and sample 127 The Good Scent Mama) for the S-CSE samples and the corresponding signals of the neat samples. For the MV calculations in Fig. 23B, D, and E, the target spectrum for each analyte peak was the average spectrum across three neat replicates. Since the mass spectral quality begins to degrade as the *S/N* approaches the *LOD* in the TIC, Fig. 23B shows the overall trend between neat TIC signal, S-CSE TIC signal, and MV as a three dimensional plot. Peaks with higher signal in both neat and S-CSE samples generally exhibit higher MV due to higher-quality mass spectra. To see a clear trend in a two-dimensional picture, Fig. 23C shows a

signal threshold of 30,000 ($\log_{10}(30{,}000) \approx 4.5$) was applied to the neat TIC peak intensities. Peaks below this threshold were kept in dark blue same as Fig. 23B, while peaks above the threshold were colored green to visualize the distribution relative to signal intensity. Fig. 23D shows the relationship between MV of the S-CSE samples and the corresponding neat TIC signal. Most analyte peaks follow the expected trend; however, 14 peaks out of 214 total peaks exceeded the MV threshold of 800 but did not meet the TIC signal threshold of 30,000. 93.5% of the peak made it through the threshold criteria. Furthermore, in Fig. 23E is summarized the MV behavior across all analyte peaks using MV > 800 as the selection criterion for reliable spectral matching in Table 7. These results indicate that mass spectrum quality is strongly dependent on chromatographic signal intensity. As TIC peak height decreases toward the *LOD*, the resulting spectra becomes increasingly noisy, which reduces MV performance even when analyte is present. Fig. 23B–E demonstrates that a neat TIC peak height of approximately 30,000 corresponds to an MV threshold of ~800, providing a practical link between chromatographic signal intensity and spectral matching quality. This relationship helps define a consistent criterion for selecting analyte peaks with sufficiently high spectral reliability. This ensures that retained peaks correspond to reliable compound identifications while minimizing low-quality matches and background signals. Collectively, these criteria improve data quality by reducing noise contributions and retaining analytically meaningful features for subsequent analysis.


A
The Good Scent Mama
×10^5
Signal
Time (min)
20


B
MV
1000
800
600
400
200
0
log10 Neat TIC Signal
log10 Can TIC Signal


Signal Threshold (30,000)
C
log10 Can TIC Signal
log10 Neat TIC Signal


Signal Threshold (30,000)
D
MV
log10 Neat TIC Signal

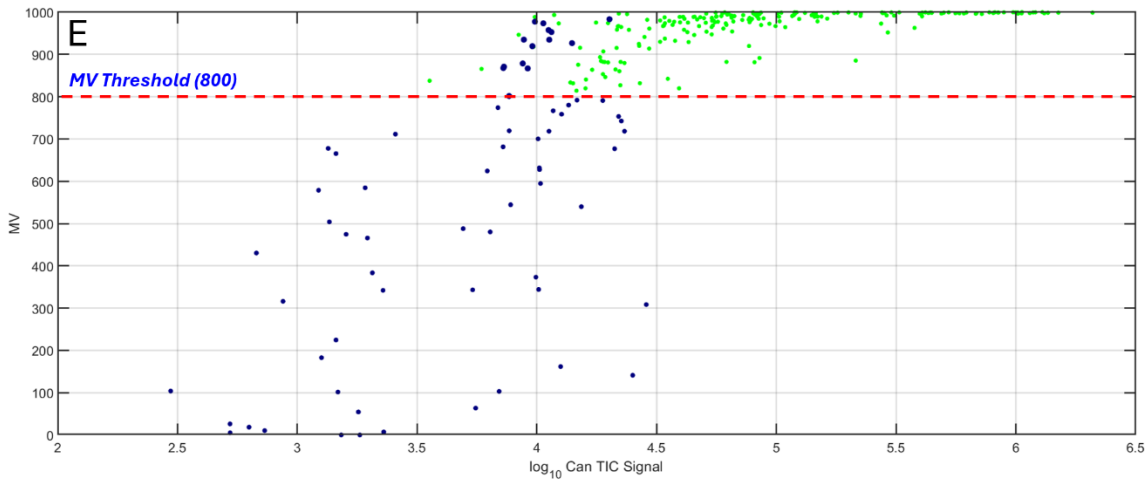
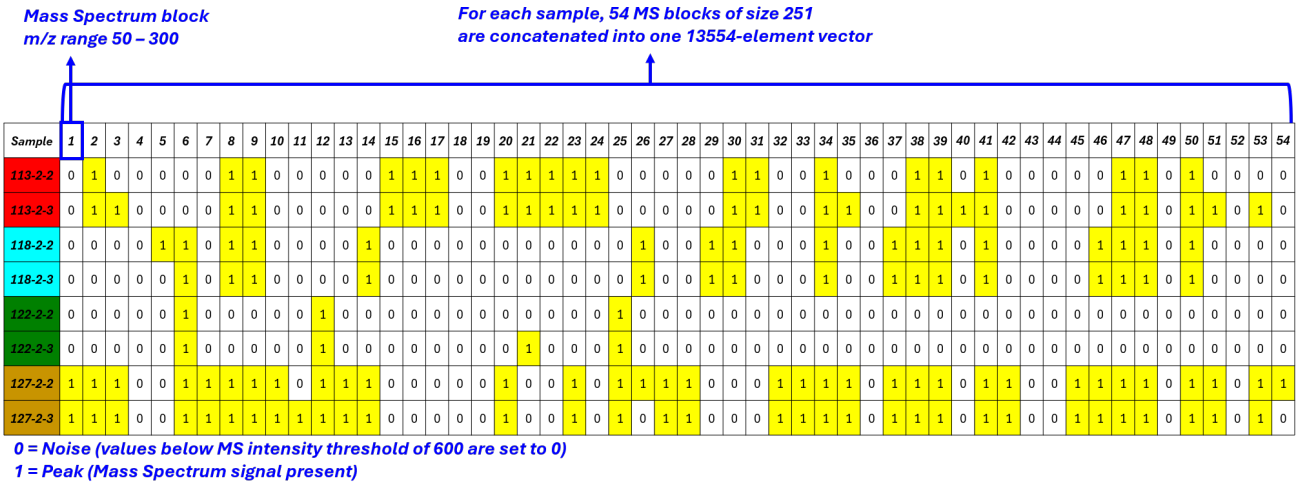

E
MV Threshold (800)
MV
log10 Can TIC Signal

Figure 23. Evaluation of intensity thresholds used for data filtering of MS and TIC data. (A) Comparison of an actual peak (left) and a noise region (right) in MS data, showing the selection of an intensity threshold of 600. (B) Three-dimensional relationship between TIC peak maximum intensity (neat and S-CSE samples) and mass spectrum MV for 54 analyte peaks across four fragrance samples (113 Nest Midnight, 118 Kate Spade Truly Dazzling, 122 Old Spice Classic, and 127 The Good Scent Mama). MV calculations used averaged spectra from three neat replicates as the target. (C) Two-dimensional projection of MV versus neat TIC peak intensity, with a TIC threshold of 30,000 applied; peaks below this threshold are shown in dark blue and those above in green. (D) Relationship between MV (S-CSE samples) and corresponding neat TIC peak intensity, highlighting the dependence of spectral quality on chromatographic signal intensity; 14 of 214 peaks exceeded MV > 800 without meeting the TIC threshold of 30,000. (E) Distribution of MV values for all analyte peaks using MV > 800 as the selection criterion, summarizing spectral matching reliability across the dataset. Collectively, these results demonstrate a strong dependence of mass spectrum quality on TIC signal intensity, with a practical correspondence between a neat TIC peak height of ~30,000 and MV, providing a unified criterion for reliable feature selection.

A TIC threshold of 30,000 was applied to ensure the neat samples made it possible to generate a comprehensive peak list of all 13 neat samples for subsequent analysis, resulting in 103 peaks in Table 6. This peak list was then applied to the I-CSE and S-CSE samples, retaining only analyte peaks with a MV > 800 compared to the neat samples. Among the S-CSE samples, the lowest observed peak count was 54 analyte peaks across all 13 fragrance samples, with the results summarized in Table 7, which was used as the final number of peaks for further analysis.

To emphasize the capability of the mass spectra provided across sample chemical fingerprinting, while reducing the effect of signal intensity variation, the Fig. 24 shows the MS-based PCA was performed using concatenated mass spectral data. Each MS block had an *m/z* range of 50–300. Signals with intensities above 600 were considered to represent qualifying mass spectra and were denoted as "1" (spectrum present), whereas signals below the threshold were denoted as "0" (no qualifying spectrum or noise). This binary notation was only used to simplify the visualization and discussion of spectrum presence or absence; the underlying data remained the original mass spectra. The resulting 54 MS blocks were concatenated into a single vector representing each sample. Each spectrum contained 251 *m/z* values (*m/z* 50–300),

resulting in a 13,554-element vector per sample. This approach allowed consistent comparisons across samples and provided a structured input for PCA based on mass spectrum patterns rather than absolute intensities.



Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
113-2-2	0	1	0	0	0	0	0	1	1	0	0	0	0	0	1	1	1	0	0	1	1	1	1	1	0	0	0	0	0	1	1	0	0	1	0	0	0	1	1	0	1	0	0	0	0	0	1	1	0	1	0	0	0	0
113-2-3	0	1	1	0	0	0	0	1	1	0	0	0	0	0	1	1	1	0	0	1	1	1	1	1	0	0	0	0	0	1	1	0	0	1	1	0	0	1	1	1	1	0	0	0	0	0	1	1	0	1	1	0	1	0
118-2-2	0	0	0	0	1	1	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	1	0	0	1	1	1	0	1	0	0	0	0	1	1	1	0	1	0	0	0	0
118-2-3	0	0	0	0	0	1	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	1	0	0	1	1	1	0	1	0	0	0	0	1	1	1	0	1	0	0	0	0
122-2-2	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
122-2-3	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
127-2-2	1	1	1	0	0	1	1	1	1	1	0	1	1	1	0	0	0	0	0	1	0	0	1	0	1	1	1	1	0	0	0	1	1	1	1	0	1	1	1	0	1	1	0	0	1	1	1	1	0	1	1	0	1	1
127-2-3	1	1	1	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	0	0	1	0	1	0	1	1	0	0	0	1	1	1	1	0	1	1	1	0	1	1	0	0	1	1	1	1	0	1	1	0	1	0

Figure 24. Schematic illustration of the workflow used to concatenate MS data. Each MS block has the *m/z* range of 50 – 300. Signals above the intensity threshold of 600 were assigned to a value of 1 (peak present), while signals below the threshold were assigned a value of 0 (noise). The 54 MS blocks were then concatenated into a single vector.

Table 6. Comprehensive peak list for neat samples (103 peaks total).

Peak Numbers	t_R (min)	1	2	3	4	5	6	7	8	9	10	11	12	13
1	14.0432												x	
2	14.0975		x											
3	15.0208		x	x		x	x					x	x	
4	15.6861	x	x											
5	15.8626	x	x			x						x	x	
6	15.9508	x	x			x	x			x		x	x	x
7	16.0323	x	x			x	x							x
8	16.1137	x												
9	16.4735						x							
10	16.5754		x	x	x	x				x				x
11	16.6025	x												
12	16.9487												x	
13	17.0845	x	x	x	x	x	x	x	x	x	x		x	x
14	17.4647	x												
15	18.0010												x	
16	18.2114		x	x			x							
17	18.4219	x	x	x	x				x		x	x	x	x
18	18.6188	x	x	x	x						x	x	x	x
19	18.6934			x						x	x		x	
20	18.7274		x				x	x				x	x	
21	19.1007			x			x				x	x		x
22	19.4877			x	x	x	x	x	x	x			x	
23	19.5216			x										
24	19.6574		x	x									x	
25	19.8068		x										x	
26	19.9629			x									x	
27	20.0036			x										
28	20.2548			x	x						x	x	x	
29	20.5807	x		x		x			x					
30	20.6146	x				x								
31	21.1916	x		x										
32	21.2595		x	x					x			x		x
33	21.2799								x					
34	21.5243	x												
35	21.7755	x												
36	21.9045			x										x
37	22.1556	x									x	x	x	x
38	22.1964		x	x				x		x				
39	22.3050	x		x					x	x				
40	22.3932	x							x		x	x	x	x
41	22.4883	x												
42	22.5426		x	x						x	x	x	x	x
43	22.5901													x
44	22.6308	x							x	x		x		
45	22.8006							x	x		x		x	x
46	22.8616		x	x	x							x	x	
47	22.9363												x	x
48	22.9635		x							x			x	x
49	23.0585	x										x		x
50	23.1196	x	x	x							x		x	x
51	23.2147	x								x	x	x	x	
52	23.2486								x					x
53	23.2690		x		x									
54	23.3029											x	x	x
55	23.3640	x	x	x	x							x	x	x
56	23.4726											x		
57	23.6423	x	x	x	x									
58	23.6831										x		x	x
59	23.8935									x	x			
60	23.9343		x				x							
61	24.1040	x	x			x								x
62	24.4502	x												
63	24.5724	x	x										x	
64	24.6063								x		x	x	x	
65	24.7014												x	
66	24.7625	x	x	x	x	x					x	x	x	x
67	24.8643	x				x					x	x	x	x
68	24.9118					x						x		
69	24.9933				x									
70	25.0612	x				x				x	x		x	
71	25.0883		x	x	x									x
72	25.1698		x		x	x				x	x	x	x	x
73	25.2580		x	x	x	x				x	x	x	x	x
74	25.3667	x	x					x	x	x				
75	25.4210	x			x	x						x	x	x
76	25.4821	x		x					x	x			x	
77	25.5364	x	x		x	x				x	x	x	x	x
78	25.6993					x					x			
79	25.7876			x	x									
80	25.8554									x			x	
81	25.9165										x			
82	25.9708	x	x	x	x									x
83	26.0795									x				
84	26.2152	x												
85	26.3714								x			x		
86	26.3985					x			x	x			x	
87	26.4325			x	x							x	x	
88	26.5207		x	x										x
89	26.6972	x	x	x	x	x			x		x	x	x	x
90	26.7447	x	x	x	x	x			x		x	x	x	x
91	26.9009											x		
92	26.9552	x		x	x									
93	26.9891	x		x	x									
94	27.0638	x		x	x								x	
95	27.1656		x	x	x	x	x		x		x	x	x	x
96	27.2335		x		x	x					x	x	x	x
97	27.3354					x					x	x	x	
98	27.4372	x												
99	27.5390	x	x		x	x					x	x	x	x
100	27.6069	x			x	x					x	x	x	x
101	27.7291	x	x		x	x					x	x	x	x
102	27.8309	x			x	x					x	x	x	x
103	27.8852									x	x			
Total Number of Peaks		47	41	39	31	29	11	6	19	22	33	39	52	40

Table 7. Final peak list from the two S-CSE samples with the lowest peak counts (54 peaks total).

Peak Numbers	t_R (min)	1	2	3	4	5	6	7	8	9	10	11	12	13
1	14.04												x	
5	15.86	x	x			x						x	x	x
6	15.95		x	x		x						x	x	x
10	16.58		x							x				x
11	16.60		x	x						x				x
13	17.08		x	x	x	x	x	x		x	x		x	
15	18.00												x	
17	18.42	x		x	x						x	x	x	x
18	18.62	x		x	x						x	x	x	x
19	18.69			x									x	
20	18.73												x	
22	19.49		x	x		x	x	x	x	x			x	
24	19.66												x	
28	20.25				x						x	x	x	
29	20.58	x				x								
30	20.61	x				x								
31	21.19	x												
32	21.26		x	x			x					x		x
33	21.28								x					
37	22.16	x							x		x	x	x	x
38	22.20	x												
39	22.30	x										x		
40	22.39	x										x	x	
44	22.63	x												
45	22.80							x	x		x		x	x
46	22.86		x		x									
51	23.21										x		x	
52	23.25								x				x	x
53	23.27				x									
55	23.36	x	x	x	x									x
61	24.10	x	x			x								x
64	24.61											x	x	
65	24.70												x	
66	24.76	x	x	x	x	x					x	x	x	x
67	24.86					x						x	x	
68	24.91											x		
71	25.09		x		x						x		x	x
72	25.17	x	x		x	x				x	x	x	x	x
73	25.26	x	x		x	x				x	x	x	x	x
74	25.37								x					
77	25.54	x	x		x	x				x	x	x	x	x
80	25.86												x	
82	26.08									x				
85	26.37											x		
86	26.40					x				x			x	
87	26.43				x							x	x	
89	26.70	x		x	x	x			x		x	x	x	
90	26.74	x		x	x	x					x	x	x	
91	26.90											x		
94	27.06	x		x	x								x	
95	27.17	x		x	x	x			x		x	x	x	x
99	27.54											x	x	
100	27.61											x		
101	27.73					x						x	x	
102	27.83											x		
Total Peak Numbers		21	15	14	17	17	3	3	8	9	15	26	34	18

Preparation of MS-based PCA scores plots after mean centering of the X-block allowed assessment of whether or not preprocessing (primarily the thresholding of the signals in the mass

spectra) improves discrimination amongst fragrance and sample preparation types. Figure 25 shows applying the preprocessing improves clustering of the I-CSE and S-CSE samples. In Fig. 25A, prior to applying the signal threshold of 600 in the mass spectra per Fig. 23, although the I-CSE and S-CSE samples show improved clustering within each fragrance compared with the TIC-chromatogram based PCA in Fig. 22B, substantial overlap and dispersion among fragrances remain, making classification difficult. In contrast, Fig. 25B, after applying the signal threshold of 600 in the mass spectra per Fig. 23, improved separation and clustering is observed, with more distinct clustering within individual fragrance samples obtained, indicating enhanced discriminatory capability. Mean centering was applied to the X-block because the data are high-dimensional and sparse, which benefits from preprocessing to improve interpretability. This preprocessing step centers the data around the mean of each variable, allowing the PCA model to focus on relative differences in spectral features rather than absolute intensity values [28]. Consequently, the variance captured by the PCs reflects meaningful structural differences between samples, enabling more effective discrimination during analysis. Overall, Fig. 25 A versus B serves to validate the application of the signal threshold of 600 to improve the quality of the mass spectra.

Considering the interpretability of sparse, thresholded data, the relatively low percentage of variance explained by the PCs were considered (25% sum of PC1 and PC2 in Fig. 25A and 29% in Fig. 25B) [29,30]. Noise removal produces a sparse data matrix in which variance is distributed across many variables rather than dominated by a few high-intensity signals [31]. Despite the lower variance explained, the thresholded mass spectra provides meaningful structure for classification by emphasizing the presence or absence of characteristic mass spectrum information for each fragrance. For sparse data sets, a lower percentage of variance

captured by the PCs is expected and does not indicate poor model performance [32]. Instead, the improved separation observed in the scores plots supports that thresholding enhances discriminatory information while reducing noise contributions.

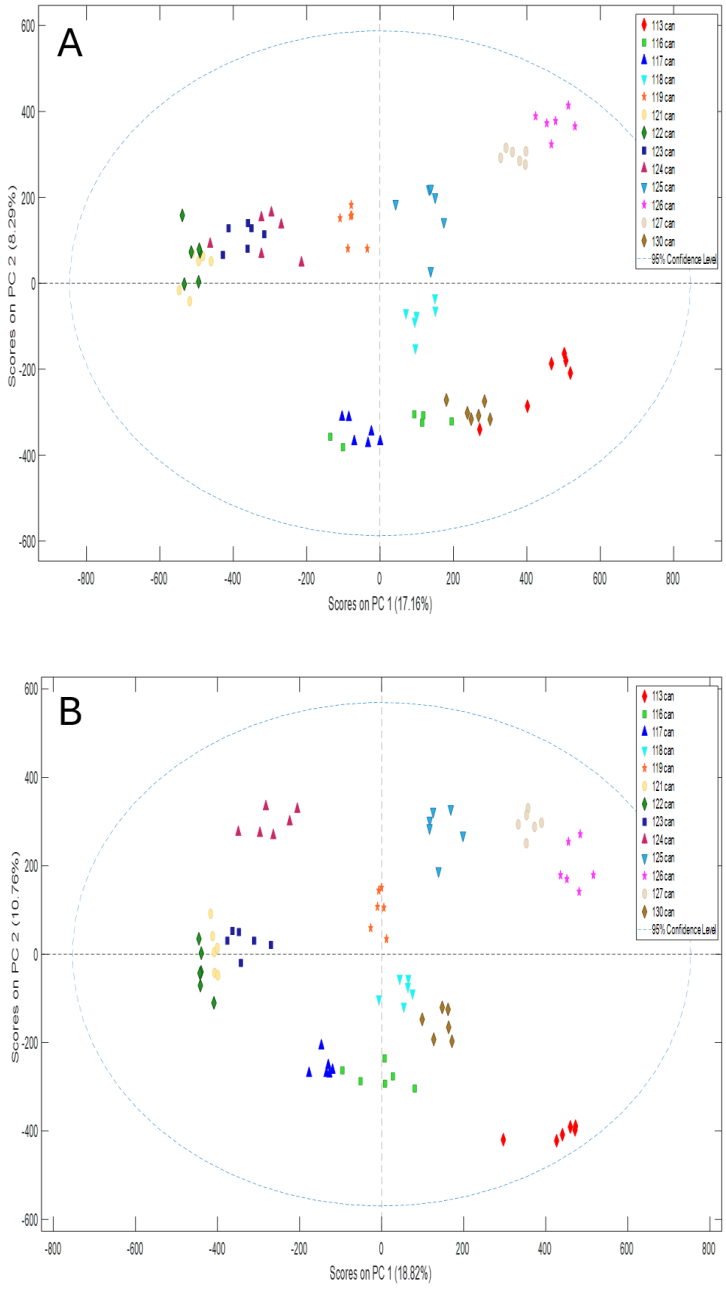


Figure 25. PCA scores plots based on MS data. (A) PCA scores plot of I-CSE and S-CSE samples without a threshold. (B) PCA scores plot of I-CSE and S-CSE samples with a threshold of 600.

To quantify the similarity (or lack thereof) of individual fragrance samples to a chosen reference fragrance, a peak-matching procedure was performed in Fig. 26. Match values were first calculated within each sample using the final peaks in Table 7, with peaks absent in each sample assigned a value of zero when it has the MV below 800. This allowed direct comparison of detected analyte components both among replicates and across different fragrance samples. Peaks with a MV > 800 indicate strong similarity to the reference, while zeros correspond to missing and not matching features. For each sample, the total number of analyte peaks matching the selected S-CSE was then counted to provide a quantitative measure of similarity. Lastly, error bars were calculated using the sample standard deviation for both matching and missing peaks, where variability between replicates was assessed based on the distribution of peak counts within each sample group (I-CSE and S-CSE).

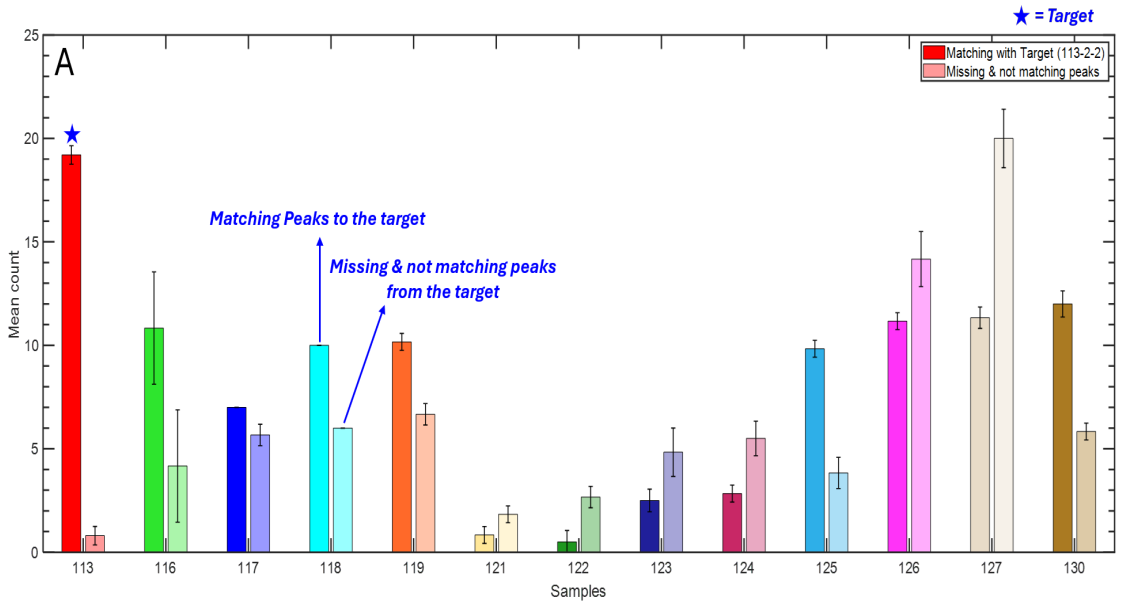
Using four different reference samples as examples (each of the 13 fragrances could in principle be used as a reference), Fig. 26 facilitates understanding whether or not target-based peak counting reliably distinguishes between fragrances. In each case, the reference sample shows the highest number of matching peaks and the fewest missing peaks (as it should if the method is reproducible), reflecting good internal consistency. Other fragrance samples, serving as “unknowns” relative to a given reference fragrance, display substantially fewer matching peaks and a greater number of missing peaks, highlighting differences in chemical composition, suggesting this approach has potential for matching reference fragrances to field testing situations. This trend is consistent across all four reference selections of S-CSE samples 113-2-2, 118-2-3, 122-2-2, and 127-2-3, indicating that the method reliably distinguishes between fragrances regardless of the chosen reference. The distinct separation between reference and

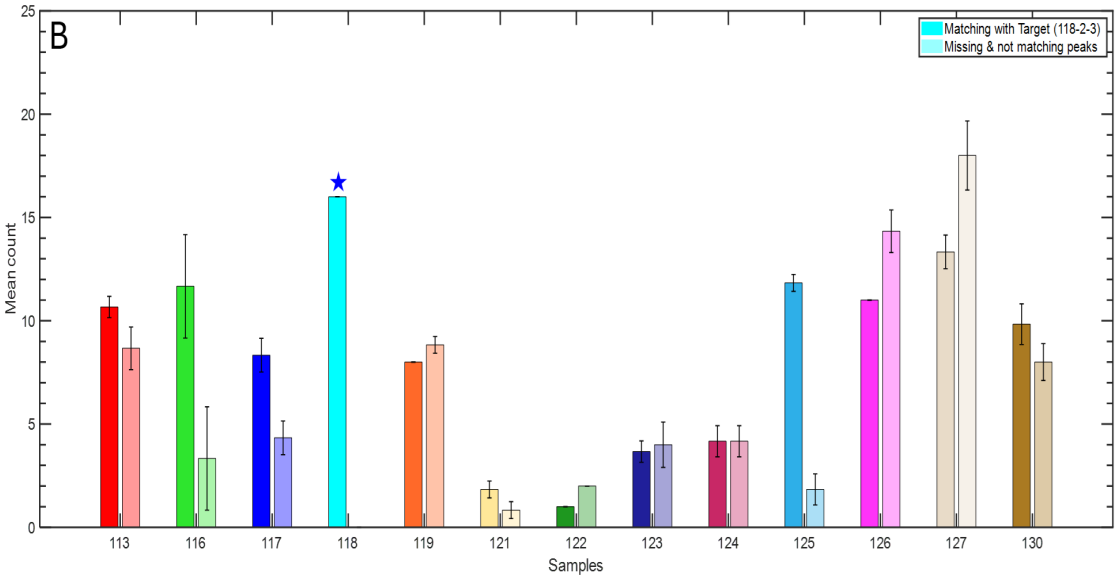
“unknown” samples demonstrates that target-based peak counting provides a robust measure for fragrance classification.

For example, when sample 113-2-2 (Nest Midnight) was used as the reference target (Fig. 26A), several fragrance samples, including samples 116, 118, 119, 125, 126, and 127, showed a relatively high number of matching peaks (exceeding half of the target set). However, these samples simultaneously exhibited a substantial number of missing and non-matching peaks, which enables discrimination between fragrances despite partial overlap. This behavior is consistent with the separation observed in the PCA scores plot (Fig. 25B), where these samples do not cluster together.

In contrast, when sample 122-2-2 (Old Spice Classic) was used as the reference target (Fig. 26C), a generally lower number of matching peaks was observed across many fragrance samples, indicating weaker similarity to the target profile. However, samples 121 and 123 showed comparatively similar matching and missing peak patterns to sample 122, consistent with their classification as cologne-type formulations. This suggests that structurally or formulation-wise similar products can exhibit overlapping peak-count profiles.

An inverse relationship between matching peaks and missing peaks further supports this interpretation. Samples with fewer matching analyte peaks consistently show higher numbers of missing and not matching peaks, indicating the absence of key analyte components present in the reference fragrance. This pattern shows that the peak-counting metric effectively captures both shared and unique components of each fragrance and complements multivariate analysis supporting the identification of distinctive fragrance profiles.


★ = Target
A
Matching with Target (113-2-2)
Missing & not matching peaks
Matching Peaks to the target
Missing & not matching peaks
from the target
Mean count
Samples
0
5
10
15
20
25
113
116
117
118
119
121
122
123
124
125
126
127
130


B
Matching with Target (118-2-3)
Missing & not matching peaks
Mean count
Samples
0
5
10
15
20
25
113
116
117
118
119
121
122
123
124
125
126
127
130

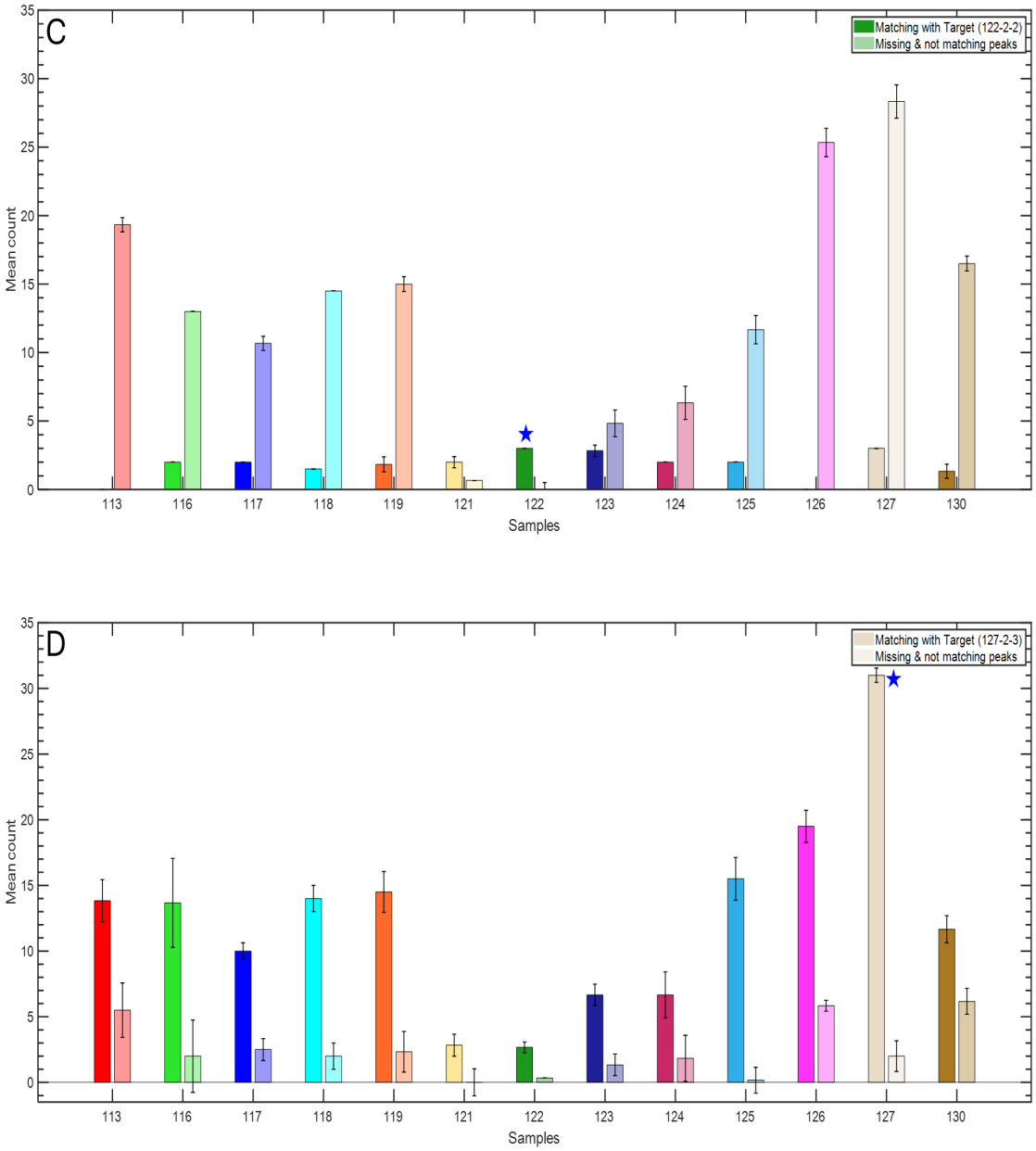


Figure 26. Average number of peaks matching the target, and number of missing and not matching peaks relative to the same fragrance. The solid bars represent the average number of peaks matching the reference target, while the lighter bars represent the number of missing and non-matching peaks relative to the same reference target fragrance. (A) Target-based peak counting using Sample 113-2-2 (Nest Midnight) as the reference. (B) Sample 118-2-3 (Kate Spade Truly Dazzling) as the reference. (C) Sample 122-2-2 (Old Spice Classic) as the reference. (D) Sample 127-2-3 (The Good Scent Mama) as the reference. In each case, the reference sample exhibits the highest number of matching peaks and the lowest number of missing peaks, while other fragrance samples show reduced matching and increased non-matching peaks. This demonstrates that target-based peak counting provides a reproducible and discriminating metric for fragrance differentiation.

References

[1] H.J. Hübschmann, Handbook of GC-MS: fundamentals and applications. John Wiley & Sons. (2026). https://doi.org/10.1002/9783527674305.

[2] T.M. Lovestead, Gas Chromatography, ASM Handbook. 10 (2019). https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=926381

[3] H.M. McNAIR, J.M. Miller, Basic Gas Chromatography, John Wiley & Sons, Inc. (1997)

[4] D.M. Ottenstein, The chromatographic support in gas chromatography, J. Chromatographic. Sci. 11 (1973) 136-144. https://doi.org/10.35629/4494-10031833.

[5] H.H. Ali, Gas Chromatography: Principles, Medicine and Pharmaceutical, Energy, Fuel Applications and Environmental Analysis, SAR J Med Biochem, 6(4) (2025) 94-105. https://doi.org/10.36346/sarjmb.2025.v06i04.005.

[6] B.A. Eckenrode, Environmental and Forensic applications of field-portable GC-MS: An overview, J. Am. Soc. Mass Spectrom. 12(6) (2001) 683-693. https://doi.org/10.1016/S1044-0305(01)00251-3.

[7] R.C. Halvorsen, T.J. Trinklein, C.G. Warren, R.D. Rogan, R.E. Synovec, Optimizing column-to-column retention time alignment in high-speed gas chromatography by combining retention time locking and correlation optimized warping, Talanta. 254 (2023). https://doi.org/10.1016/j.talanta.2022.124173.

[8] H.-G. Schmarr, J. Bernhardt, Profiling analysis of volatile compounds from fruits using comprehensive two-dimensional gas chromatography and image processing techniques, J. Chromatogr. A 1217 (2010) 565–574. https://doi.org/10.1016/j.chroma.2009.11.063.

[9] S.E. Stein, Chemical substructure identification by mass spectral library searching, Journal of the American Society for Mass Spectrometry. 6(8) (1995) 644-655. https://doi.org/10.1016/S1044-0305(05)80054-6.

[10] V. Cuzuel, G. Cognon, I. Rivals, C. Sauleau, F. Heulard, D. Thièbaut, J. Vial, Origin, Analytical Characterization, and Use of Human Odor in Forensics, Journal of Forensic Sciences. 62(2) (2017) 330–350. https://doi.org/10.1111/1556-4029.13394.

[11] N. D. Giovanni, D. Marchetti, A Systematic Review of Solid-Phase Microextraction Applications in the Forensic Context, Journal of Analytical Toxicology. 44(3) (2020) 268–297. https://doi:10.1093/jat/bkz077.

[12] S. C. Tran, M. R. Williams, M. E. Sigman, Hydrocarbon retention on activated carbon: Preservation of fire debris evidence, Forensic Chemistry. 39 (2024). https://doi.org/10.1016/j.forc.2024.100579.

[13] C. Martín-Alberca, F. E. Ortega-Ojeda, C. Gracía-Ruiz, Analytical tools for the analysis of fire debris. A review: 2008-2015, Analytica Chimica Acta. 928 (2016) 1-19. https://doi.org/10.1016/j.aca.2016.04.056.

[14] S. Gherghel, R. M. Morgan, J. F. Arrebola-Liebanas, G. S. Blackman, A. Garrido-Frenich, I. P. Parkin, Persistence of transferred fragrance on fabrics for forensic reconstruction applications. Science & Justice 2020, 60(1), 53-62. https://doi.org/10.1016/j.scijus.2019.09.002.

[15] S. Gherghel, R. M. Morgan, J. F. Arrebola-Liébanas, C. S. Blackman, I. P. Parkin, Fragrance transfer between fabrics for forensic reconstruction applications, Science and Justice. 59 (2019) 256-267. https://doi.org/10.1016/j.scijus.2019.02.002.

[16] S. Gherghel, R. M. Morgan, J. F. Arrebola-Liébanas, R. Romero-González, C. S. Blackman, A. Garrido-Frenich, I. P. Parkin, Development of a HS-SPME/GC-MS method for the analysis of volatile organic compounds from fabrics for forensic reconstruction applications, Forensic Sci Int. 290 (2018) 207-218. https://doi.org/10.1016/j.forsciint.2018.07.015.

[17] V. Totten, J. Willis, S. Eddins, B. Brooks, Qualitative identification of volatile compounds in foods and flowers using passive headspace extraction with activated charcoal fabric, Food Res Int. 163 (2023). https://doi.org/10.1016/j.foodres.2022.112130.

[18] E. Audette, P-L. Dumont, F. Crispino, PowerSorb® for forensic investigation of VOC traces: Application on perfume traces, Forensic Sci Int. 376 (2025). https://doi.org/10.1016/j.forsciint.2025.112605.

[19] L. Büchler, D. Werner, O. Delémont, Detection of gasoline on suspects' hands: Study of different sampling alternatives, Forensic Sci Int. 318 (2021). https://dx.doi.org/10.1016/j.forsciint.2020.110590.

[20] N. Boegelsack, J. Walker, C. D. Sandau, D. W. McMartin, J. M. Withey, G. O'Sullivan, Cross-Contamination of Ignitable Liquid Residues on Wildfire Debris—Effects of Packaging and Storage on Detection and Characterization, Separations. 11(2) (2024) 58. https://doi.org/10.3390/separations11020058.

[21] V. Cuzuel, R. Leconte, G. Cognon, D. Thiébaut, J. Vial, C. Sauleau, I. Rivals, Human odor and forensics: Towards Bayesian suspect identification using GC × GC-MS characterization of hand odor, Journal of Chromatography B. 1092 (2018) 379-385. https://doi.org/10.1016/j.chromb.2018.06.018.

[22] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemometrics and Intelligent Laboratory Systems. 2 (1987) 37-52. https://doi.org/10.1016/0169-7439(87)80084-9.

[23] I. Koo, S. Kim, X. Zhang, Comparative analysis of mass spectral matching-based compound identification in gas chromatography-mass spectrometry, Journal of Chromatography A. 1298 (2013) 132-138. http://dx.doi.org/10.1016/j.chroma.2013.05.021.

[24] C.N. Cain, S. Schöneich, R.E. Synovec, Development of an enhanced total ion current chromatogram algorithm to improve untargeted peak detection, Anal. Chem. 16 (2020) 11365-11373. https://doi.org/10.1021/acs.analchem.0c02136.

[25] S.E. Stein, D.R. Scott, Optimization and Testing of Mass Spectral Library Search Algorithms for Compound Identification, J. Am. Soc. Spectrom. 5(9) (1994) 859-866. https://doi.org/10.1016/1044-0305(94)87009-8.

[26] D. Weindl, A. Wegner, C. Jäger, K. Hiller, Isotopologue ratio normalization for non-targeted metabolomics, Journal of Chromatography A. 1389 (2015) 112-119. https://doi.org/10.1016/j.chroma.2015.02.025.

[27] M. M. Fabritius, A. Broillet, S. König, W. Weinmann, Analysis of volatiles in fire debris by combination of activated charcoal strips (ACS) and automated thermal desorption-gas chromatography-mass spectrometry (ATD/GC-MS). Forensic Sci Int. 289 (2018) 232-237. https://doi.org/10.1016/j.forsciint.2018.05.048.

[28] N.B Gallager, D. O'Sullivan, M. Palacios, The Effect of Data Centering on PCA Models, White paper, 2020

[29] R. Guerra-Urzola, K. V. Deun, J. C. Vera, K. Sijtsma, A Guide for Sparse PCA: Model Comparison and Applications, Psychometrika. 86 (2021) 893-919. https://doi:10.1007/s11336-021-09773-2.

[30] J. Machkour, A. Breloy, M. Muma, D. P. Palomar, F. Pascal, Sparse PCA with False Discovery Rate Controlled Variable Selection, ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2024) 9716-9720. https://doi.org/ 10.1109/ICASSP48485.2024.10448237.

[31] Y. Xie, T. Wang, J. Kim, K. Lee, M. K. Jeong, Lease angle sparse principal component analysis for ultrahigh dimensional data, Ann Oper Res. (2024). https://doi.org/10.1007/s10479-024-06428-0.

[32] D. Shen, H. Shen, J. S. Marron, Consistency of sparse PCA in High Dimension, Low Sample Size contexts, Journal of Multivariate Analysis. 115 (2013) 317-333. https://doi.org/10.1016/j.jmva.2012.10.007.