Journal of Engineering Science and Technology
Vol. 20, No. 6 (2025) 2109 - 2125

CROSS-CORRELATION ANALYSIS FOR IDENTIFYING RELIABLE CHEMICAL MARKERS IN AGARWOOD OIL USING GC-MS AND GC-MS COUPLED WITH GC-FID TECHNIQUES

NOOR AIDA SYAKIRA AHMAD SABRI[1], MUHAMMAD IKHSAN ROSLAN[1], NUR ATHIRAH SYAFIQAH NORAMLI[1], NURLAILA ISMAIL[1,*], ZAKIAH MOHD YUSOFF[1], ALI ABD ALMISREB[2], SAIFUL NIZAM TAJUDDIN[3], MOHD NASIR TAIB[1]

[1]Advanced Signal Processing Research Interest Group, Faculty of Electrical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[2]Faculty of Computer Science and Engineering, International University of Sarajevo, Sarajevo, Hrasnička cesta 15, Ilidža 71210, Bosnia & Herzegovina
[3]Bioaromatic Research Centre of Excellence (BARCE), Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, 26300 Gambang Kuantan, Pahang, Malaysia
*Corresponding Author: nurlaila0583@uitm.edu.my

Abstract

This study employs cross-correlation analysis to evaluate the alignment of agarwood oil compounds detected using Gas Chromatography-Mass Spectrometry (GC-MS) and GC-MS coupled with Gas Chromatography-Flame Ionization Detector (GC-FID). A total of 96 GC-MS samples and 120 GC-MS/GC-FID samples were analysed, to identify significant chemical compounds for essential oil classification. The results identified γ-eudesmol and 10-epi-γ-eudesmol as the most consistently detected marker compounds, with γ-eudesmol and dihydro-β-agarofuran exhibiting the highest correlation (R=0.767). Additionally, 10-epi-γ-eudesmol and γ-eudesmol (R=0.763, lag = 0) showed strong alignment, confirming consistency between GC-MS and GC-MS/GC-FID detection. A correlation threshold of R>0.5 and a lag constraint of ±6, ensuring the selection of well-aligned compounds. Cross-correlation plots and heatmap visualizations validated alignment patterns and retention time stability. The findings demonstrate the effectiveness of cross-correlation analysis in identifying reliable marker compounds for agarwood oil classification. Future work could explore advanced retention time alignment techniques, machine learning-based classification models, and expanded datasets to improve the reliability of agarwood oil authentication and essential oil standardization.

Keywords: Agarwood oil, Cross-correlation, Essential oil classification, GC-MS, GC-MS/GC-FID.

1. Introduction

Agarwood oil, derived from the *Aquilaria* species, is one of the most valuable essential oils due to its distinctive fragrance, medicinal properties and high economic value [1-3]. Its therapeutic applications have long been recognized in traditional medicine systems and recent studies have highlighted its pharmacological potential, including anti-inflammatory and antimicrobial effects [2]. In addition to its medical use, agarwood oil is a prized component in luxury perfumes and incense products, and its global demand continues to grow, contributing significantly to the economies of producing countries [1, 4]. The quality of agarwood oil is primarily determined by its chemical composition, which varies depending on species, extraction techniques and environmental conditions [4]. An overview of agarwood production stages is shown in Fig. 1, and representative agarwood oil samples are presented in Fig. 2.

(a) *Aquilaria* trees cultivated in a plantation in Assam, India.

(b) Cross-sectional view of stems showing resin-impregnated zones.

(c) Chiseled and cleaned resinous wood fragments.

Fig. 1. Stages in the production of resinous agarwood from Aquilaria trees [1].

Fig. 2. Agarwood oil samples arranged by concentration from K1 (lowest) to K7 (highest).

To analyse this chemical profile, Gas Chromatography-Mass Spectrometry (GC-MS) and Gas Chromatography-Flame Ionization Detector (GC-FID) are widely used analytical techniques, with GC-MS enabling precise compound identification based on mass spectral data and GC-FID providing sensitive quantitative analysis of volatile compounds [5-11]. However, differences in detection sensitivity, peak intensities, and retention times between these techniques present challenges for establishing a standardized quality assessment system.

Cross-correlation analysis offers a statistical approach to evaluate the alignment of compounds detected by both techniques, enabling systematic comparison of detection consistency and retention time synchronization. By identifying highly correlated compounds, this study aims to establish reliable chemical markers that can support species classification, authentication, and quality grading of agarwood oil.

The research by Hoque et al. [12] has primarily focused on GC-MS profiling of agarwood oil, with limited studies examining compound consistency across multiple analytical techniques. The integration of GC-MS with GC-FID allows for a more comprehensive evaluation of compound stability and detection patterns, which is critical for improving standardized grading systems [11, 13-15]. According to Adhikari et al. [16], cross-correlation analysis has been widely utilized in various scientific fields, including metabolomics, environmental studies, medical imaging, and food science, to analyse relationships between variables over time or across different datasets [17].

In computational chemistry, Ali et al. [18] applied cross-correlation in their study on natural lead compounds against hemagglutinin-esterase surface glycoproteins in human coronaviruses, using principal component analysis and molecular dynamics simulations to investigate molecular interactions. This demonstrates the relevance of cross-correlation in studying molecular behaviour, similar to its application in compound profiling for essential oils.

In environmental science, Manco et al. [17] studied the cross-correlations of biogenic volatile organic compounds (BVOCs) emissions to differentiate phenological stages and environmental stress factors in Mediterranean sorghum plantations. Their findings underscore the role of cross-correlation in analysing volatile organic compound (VOC) emissions, which is directly relevant to agarwood oil research, where VOCs determine oil quality and classification. Similarly, Mi et al. [19] employed cross-correlation to evaluate the effects of salt concentration on microbial diversity and the quality of spontaneously fermented radish paocai, demonstrating the utility of cross-correlation in food chemistry. This application parallels the detection of compound variations in agarwood oil using cross-correlation techniques.

Beyond chemistry and environmental science, cross-correlation has been applied in medical imaging to study blood flow patterns. A study conducted by Bakker et al. [20] used cascaded plane wave ultrasound to measure blood velocity vectors in the carotid artery, employing cross-correlation techniques for tracking motion in biological tissues. This highlights the versatility of cross-correlation in signal analysis, aligning with its use in compound detection in chromatography.

Additionally, Duma et al. [21] examined carbon-13 lineshapes in solid-state NMR and studied coherent CSA-dipolar cross-correlation, which is crucial for analysing structural properties of labelled compounds. This is directly related to

chromatographic techniques like GC-MS and GC-FID, where spectral data is analysed for compound identification. Similar analytical approaches, combining chromatographic and spectroscopic techniques, have been used for catalyst characterization and pharmaceutical analysis [22-24]. In addition to essential oil profiling, chromatographic and chemometric methods have been widely utilized in environmental monitoring and wastewater treatment studies, underscoring their versatility across diverse analytical domains [25-27].

While such studies highlight the broad applicability of these techniques, the specific role of cross-correlation in synchronizing GC-MS and GC-FID datasets remains underexplored. This study aims to evaluate the alignment of chemical compounds detected in agarwood oil using cross-correlation analysis applied to data obtained from GC-MS and GC-MS coupled with GC-FID techniques. It focuses on assessing the similarity in compound peak intensities and retention times between two datasets.

One dataset is based on four high-quality *Aquilaria malaccensis* samples, while the other includes four *Aquilaria* species. The objective is to identify consistently detected and strongly correlated compound pairs that can serve as reliable chemical markers. By establishing statistically aligned markers across platforms, this study provides a foundation for developing robust and standardized classification frameworks for agarwood oil. The results are expected to support industrial needs related to authentication, quality grading, and regulatory compliance based on chemical profiling.

2. Research Methodology

2.1. Data Collection and Compound Identification

This study utilizes two datasets obtained through GC-MS and GC-MS coupled with Gas GC-FID to analyse agarwood oil compounds. The first dataset consists of 96 samples from *Aquilaria malaccensis*, classified as high-quality essential oils, sourced from the Forest Research Institute Malaysia (FRIM) and Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA). These institutions were selected due to their established protocols and ability to provide well-characterized agarwood oil samples with consistent extraction and analysis conditions. Their contributions enabled a controlled environment for compound identification and cross-correlation analysis.

In this study, the samples underwent hydro-distillation after water soaking (5-21 days) and drying, and seven significant compounds were identified: β-agarofuran, α-agarofuran, 10-epi-γ-eudesmol, γ-eudesmol, eudesmol, longifolol and hexadecanolm [5], coded as C1, C2, C3, C4, C5, C6 and C7, respectively. These compounds are detailed in Table 1, which presents the chemical compound dataset and their respective peak areas (%) for different agarwood oil qualities. The second dataset contains 120 samples from four *Aquilaria* species (*A. beccariana*, *A. malaccensis*, *A. crassna* and *A. subintegra*), extracted at the BioAromatic

Research Centre of Excellence (BARCE), UMPSA, using hydro-distillation, followed by dilution in analytical-grade dichloromethane (DCM) for GC-MS and GC-FID analysis. The identification of compounds in this dataset was conducted using mass spectral libraries (NIST, Wiley, HPCH2205.L) for GC-MS and linear retention indices for GC-FID [28].

Among the compounds identified, six significant compounds consisting of β-selinene, dihydro-β-agarofuran, δ-guaiene, 10-epi-γ-eudesmol, γ-eudesmol and pentadecanoic acid were selected for further analysis, as they were consistently present across all four species of *Aquilaria* oil. These compounds, designated as Ca, Cb, Cc, Cd, Ce, and Cf, were chosen for their reliability as chemical markers for species classification. Their peak area (%) values are summarized in Table 2.

Notably, Cd and Ce exhibited relatively higher peak areas in *A. malaccensis*, suggesting their prominence in this species. In contrast, Cc and Cb showed a more balanced distribution across species, indicating broader market potential. The variation in compound intensity across species, despite overall presence, highlights their potential discriminative power. These datasets were specifically chosen to enable cross-correlation analysis, supporting the identification of well-aligned marker compounds for essential oil quality grading and species differentiation.

Table 1. Identified chemical compounds and peak area (%) from GC-MS analysis for different agarwood oil qualities

Code	Compounds	Peak Area (%)			
		RG	HG	MNS	HIGH
C1	β-agarofuran	3.96	3.47	2.21	2.37
C2	α-agarofuran	2.98	2.41	1.42	2.01
C3	10-epi-γ-eudesmol	21.01	9.58	5.04	10.34
C4	γ-eudesmol	1.38	11.11	12.36	9.92
C5	longifolol	0.00	0.00	0.00	0.00
C6	hexadecanol	0.00	0.00	0.00	0.00
C7	eudesmol	0.00	0.00	0.00	0.00

Table 2. Identified chemical compounds and peak area (%) from GC-MS coupled with GC-FID analysis from four *Aquilaria* oil species

Code	Compounds	Peak Area (%)			
		AB	AM	AC	AS
Ca	β-selinene	0.66	0.56	0.11	0.37
Cb	dihyro-β-agarofuran	1.25	0.55	0.48	0.44
Cc	δ-guaiene	0.74	2.02	0.21	0.35
Cd	10-epi-γ-eudesmol	0.34	6.73	2.54	2.16
Ce	γ-eudesmol	0.26	2.17	0.95	1.85
Cf	pentadecanoic acid	0.15	0.15	0.14	0.46

2.2. Data Pre-Processing

Figure 3 presents a structured workflow outlining the methodology used for cross-correlation analysis in agarwood oil compounds. The workflow consists of multiple stages, from data collection to compound selection and visualization, ensuring a systematic approach to identifying significant marker compounds for quality grading and species classification.

The process begins with data collection, where GC-MS data serves as the input, and GC-MS coupled with GC-FID data serves as the output. The GC-MS dataset consists of chemical compounds extracted from high-quality agarwood oil samples, while the GC-MS/GC-FID dataset contains compounds from four different

Aquilaria oil species. The integration of these datasets allows for a comparative analysis of chemical composition.

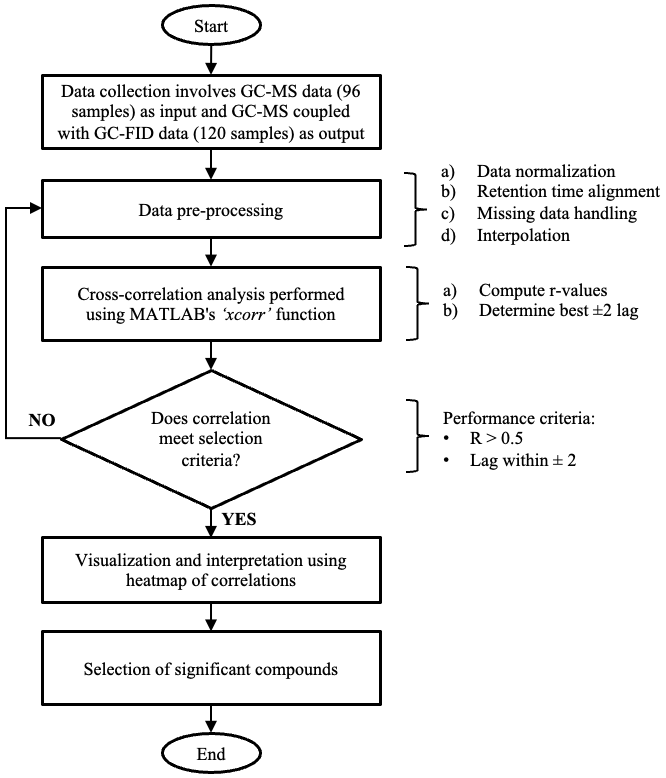


Fig. 3. Workflow of cross-correlation analysis for agarwood oil compounds.

Once the data is collected, it undergoes preprocessing to enhance consistency and accuracy. Several essential steps are involved in this phase:

- Normalization: Standardizes peak area values to minimize variations in instrument sensitivity.
- Retention Time Alignment: Ensures that compounds detected in both datasets are correctly matched by adjusting for slight variations in retention times.
- Handling of Missing Data: Removes compound pairs that lack corresponding matches in both datasets.
- Interpolation: Adjusts sample sizes to maintain uniformity, enabling meaningful comparisons between GC-MS and GC-FID results.

After preprocessing, the cross-correlation analysis is conducted using MATLAB's *'xcorr'* function [29]. This step quantifies the similarity between compounds by computing correlation coefficients (R-values) and evaluating the lag shift between peaks [18, 30]. The cross-correlation results provide insights into how well the compound intensities align across the two datasets.

Then, a heatmap representation of the correlation coefficients is generated to provide a comprehensive overview of the relationships between GC-MS and GC-

MS/GC-FID compounds. The heatmap enables the visual identification of` strongly correlated compound pairs, highlighting potential marker compounds for essential oil quality grading and species differentiation [31].

Following the heatmap analysis, a decision-making step is introduced to assess the significance of each compound pair based on the following criteria:

- R-value greater than 0.5, indicating a moderate to strong correlation.
- Lag within ± 6, ensuring minimal misalignment between GC-MS and GC-MS/GC-FID peaks.

Compound pairs that satisfy both conditions are prioritized for further classification analysis. The top 10 compound pairs with the highest correlation values ($R > 0.6R$) and minimal lag shifts are selected for detailed evaluation, where their alignment patterns, detection consistency, and peak shifts are examined in greater depth. For compound pairs that do not meet the selection criteria ($R \leq 0.5R$ or Lag $> \pm 6$), instead of being discarded, they undergo further assessment. Additional preprocessing adjustments, such as realigning retention times or refining normalization techniques, may be applied to improve correlation accuracy. If variations remain, these compounds are documented as potentially significant but requiring further validation for classification.

2.3. Performance Criteria for Cross-Correlation Selection

The selection of significant compound correlations between GC-MS and GC-MS coupled with GC-FID datasets is based on the discrete-time cross-correlation function, which measures the similarity between compound peak intensities across both techniques. The correlation coefficient, denoted as $R_{xy}[k]$, is computed using the following formula in Eq. (1) [32]:

$$R_{xy}[k] = \sum_n x[n] \cdot y[n+k] \tag{1}$$

where $x[n]$ and y$[n]$represent the peak intensities of a compound detected in the GC-MS and GC-MS/GC-FID datasets, respectively, while k denotes the lag shift between the two signals. The cross-correlation function helps determine how well a compound detected in GC-MS corresponds to its presence in GC-MS/GC-FID, considering potential shifts in detection time.

To ensure reliable selection of compounds, two performance criteria are applied. The first criterion is the correlation strength, where only compound pairs with a correlation coefficient greater than 0.5 ($R_{xy}[k] > 0.5$) are considered significant. A correlation coefficient closer to 1 indicates a strong relationship between the compound intensities in both datasets, whereas values below 0.5 suggest weak or inconsistent relationships, which may not be suitable for classification purposes.

The second criterion is the lag constraint, which ensures that the compound detection is well-aligned across both techniques. A compound pair is only considered valid if the lag shift k is within ± 2, meaning that its detection in GC-MS and GC-MS/GC-FID does not deviate by more than two time points. This tolerance accounts for minor variations in retention time due to instrument conditions or experimental factors. If the lag shift exceeds this range ($k > \pm 2$), the

compound pair is considered misaligned and excluded from further analysis to maintain data consistency.

These threshold values were selected based on standard practices in signal-based compound comparison and observed retention time variation in GC platforms. They serve to balance sensitivity and reliability by capturing compound pairs with consistent detection while minimizing noise and misalignment due to instrument variability.

3. Results and Discussion

This study analysed the correlation between seven compounds detected in the GC-MS dataset, coded as C1 to C7, and six significant compounds from the GC-MS/GC-FID dataset, labelled as Ca to Cf. The GC-MS compounds correspond to β-agarofuran (C1), α-agarofuran (C2), 10-epi-γ-eudesmol (C3), γ-eudesmol (C4), eudesmol (C5), longifolol (C6) and hexadecanol (C7), while the GC-MS/GC-FID compounds include β-selinene (Ca), dihydro-β-agarofuran (Cb), δ-guaiene (Cc), 10-epi-γ-eudesmol (Cd), γ-eudesmol (Ce) and pentadecanoic acid (Cf).

The heatmap in Fig. 4 provides an overview of the correlation strengths between 7 GC-MS compounds (C1-C7) and 6 GC-MS/GC-FID compounds (Ca-Cf). By visually analysing the colour gradient, distinct patterns emerge, indicating varying levels of compound alignment across the two analytical techniques. Warmer colours (yellow-green) indicate higher correlations (R>0.6). Moderate correlations (0.5<R<0.6) appear in green-blue tones, while cooler colours (blue) represent weaker correlations (R<0.5).

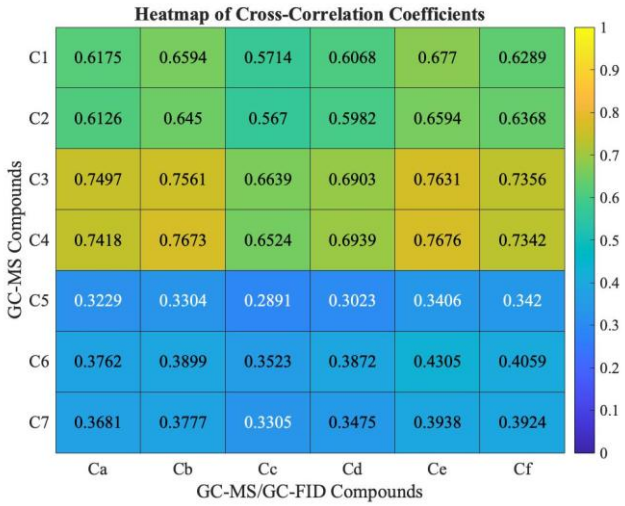


Fig. 4. Heatmap of cross-correlation coefficients between GC-MS and GC-MS/GC-FID compounds.

In general, C3 and C4 exhibit the strongest correlations with GC-MS/GC-FID compounds, as seen in the yellow-green regions of the heatmap (R>0.6). This suggests that these compounds are consistently detected across both GC-MS and GC-MS/GC-FID, making them strong candidates for essential oil classification and species differentiation. Their high correlation values indicate a stable presence in

agarwood oil, reinforcing their suitability as marker compounds. In contrast, C1 and C2 show moderate correlations (0.5<R<0.6), appearing in green-blue areas of the heatmap. This suggests that while these compounds are somewhat aligned across both techniques, their correlation is not as strong as C3 and C4, possibly due to variations in peak intensity, instrumental detection, or retention time shifts. These moderate correlations indicate that C1 and C2 may still be relevant for classification but require further validation.

On the other hand, C5, C6 and C7 exhibit the weakest correlations (R<0.5), as indicated by the predominantly blue regions in the heatmap. These low correlation values suggest poor alignment between GC-MS and GC-MS/GC-FID detections, which could be due to compound volatility, co-elution effects, or differing instrument sensitivities. While these compounds are still considered in the analysis, their inconsistent detection reduces their reliability as marker compounds for classification purposes.

Table 3 presents the compound pairs with the highest cross-correlation values with R > 0.7, ranked based on their correlation coefficient $R_{xy}[k]$. It quantifies the similarity between compound intensities in both datasets, while the best lag (k) represents the shift in detection time between GC-MS and GC-MS/GC-FID. A lag of 0 indicates simultaneous detection, whereas positive or negative lag values suggest that one technique detected the compound earlier than the other.

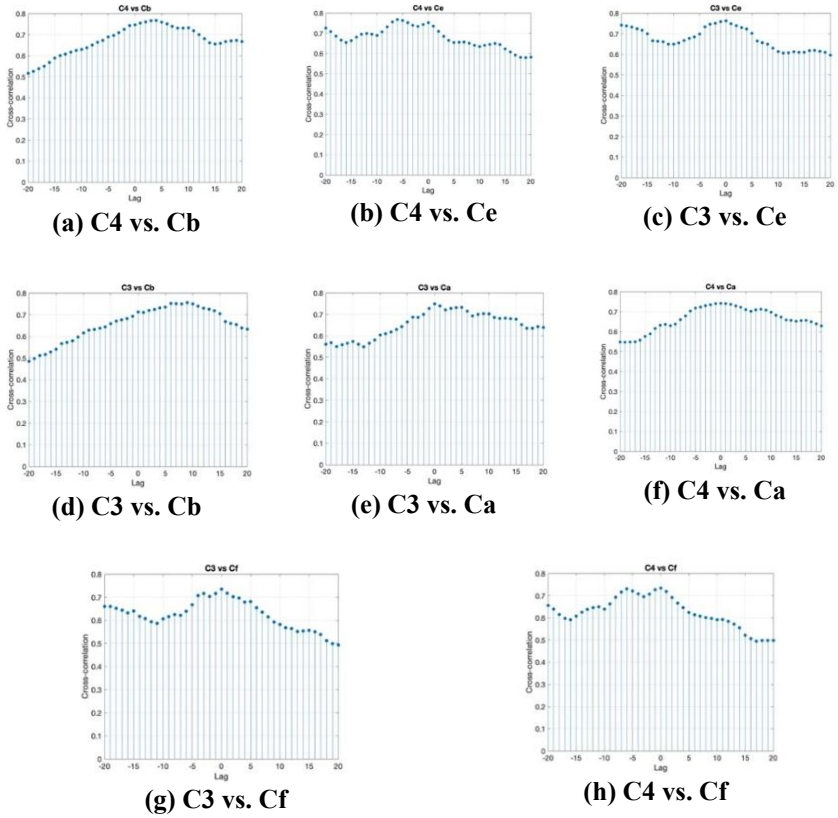
Table 3. Highest cross-correlation (R > 0.7) values and corresponding lag between GC-MS and GC-MS/GC-FID compounds

GC-MS compound	GC-MS/ GC-FID	Cross-Correlation	Lag	Interpretation
C4	Cb	0.767	4	Cb detects earlier by 4 steps
C4	Ce	0.767	-6	C4 detects earlier by 6 steps
C3	Ce	0.763	0	Detected at the same time
C3	Cb	0.756	9	Cb detects earlier by 9 steps
C3	Ca	0.749	0	Detected at the same time
C4	Ca	0.741	0	Detected at the same time
C3	Cf	0.735	0	Detected at the same time
C4	Cf	0.734	0	Detected at the same time

From Table 3, C4 exhibited the highest correlation with two GC-MS/GC-FID compounds, Cb and Ce, both with R = 0.767. However, their retention time alignment varied, with Cb detected 4 steps earlier (lag = 4) and Ce detected 6 steps later (lag = -6), indicating a strong correlation but with minor retention time shifts across techniques. Similarly, C3 vs. Ce demonstrated a strong correlation (R = 0.763) with a lag of 0, suggesting a perfect alignment between GC-MS and GC-MS/GC-FID detection.

Additionally, C3 vs. Cb showed a slightly lower but still strong correlation (R = 0.756) with a lag of 9, indicating that Cb was detected significantly earlier in GC-MS/GC-FID. Several other compound pairs, such as C3 vs. Ca (R = 0.749), C4 vs. Ca (R = 0.741), C3 vs. Cf (R = 0.735) and C4 vs. Cf (R = 0.734) also exhibited

high correlation values and demonstrated strong correlations with perfect alignment (lag = 0), suggesting their consistency across both techniques and indicating their potential role as marker compounds for agarwood oil classification. To further illustrate these findings, Fig. 5 presents the cross-correlation plots for the compound pairs with the highest correlation (R > 0.7).



(a) C4 vs. Cb **(b) C4 vs. Ce** **(c) C3 vs. Ce**

(d) C3 vs. Cb **(e) C3 vs. Ca** **(f) C4 vs. Ca**

(g) C3 vs. Cf **(h) C4 vs. Cf**

Fig. 5. Cross-correlation plots for compound pairs with the highest correlation (R > 0.7).

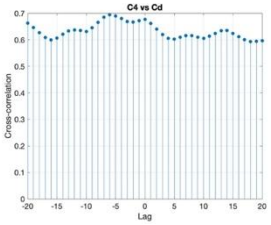
In addition to the highest correlation values (R > 0.7), 13 compound pairs exhibited moderately high correlations within the range of $0.6 \leq R \leq 0.7$, as shown in Table 4. These results indicate strong detection alignment between GC-MS and GC-MS/GC-FID, though with slightly greater variations in retention time synchronization compared to the highest correlation group.

Several compounds in this category demonstrate near-perfect alignment (lag = 0), suggesting that they are consistently detected by both techniques at the same time, reinforcing their reliability as potential marker compounds for agarwood oil classification. Examples include C3 vs. Cd (R = 0.690), C3 vs. Cc (R = 0.663), C2 vs. Ce (R = 0.659), C2 vs. Cf (R = 0.636) and C2 vs. Ca (R = 0.612). These pairs exhibit detection consistency, further validating the compatibility between the two analytical methods.

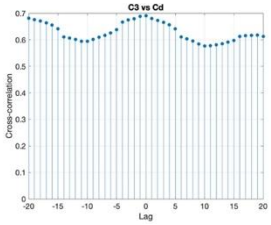
Table 4. High moderate cross-correlation (0.6 ≤ R ≤ 0.7) values and corresponding lag between GC-MS and GC-MS/GC-FID compounds.

GC-MS compound	GC-MS/ GC-FID	Cross-Correlation	Lag	Interpretation
C4	Cd	0.693	-6	C4 detects earlier by 6 steps
C3	Cd	0.690	0	Detected at the same time
C1	Ce	0.676	-3	C1 detects earlier by 3 steps
C3	Cc	0.663	0	Detected at the same time
C1	Cb	0.659	7	Cb detects earlier by 7 steps
C2	Ce	0.659	0	Detected at the same time
C4	Cc	0.652	-5	C4 detects earlier by 5 steps
C2	Cb	0.645	10	Cb detects earlier by 10 steps
C2	Cf	0.636	0	Detected at the same time
C1	Cf	0.628	-4	C1 detects earlier by 4 steps
C1	Ca	0.617	-2	C1 detects earlier by 2 steps
C2	Ca	0.612	0	Detected at the same time
C1	Cd	0.606	-3	C1 detects earlier by 3 steps

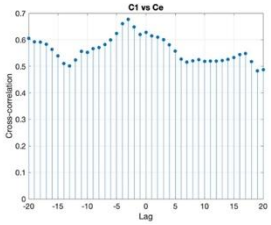
For compounds with positive or negative lag values, the detection shifts may be attributed to instrumentation sensitivity, chromatographic separation differences or sample preparation factors. Notable examples of positive lag values include C1 vs. Cb (R = 0.659, lag = 7) and C2 vs. Cb (R = 0.645, lag = 10), indicating that the GC-MS/GC-FID detection occurred earlier than GC-MS. Conversely, negative lag values were observed in C4 vs. Cd (R = 0.693, lag = -6), C1 vs. Ce (R = 0.676, lag = -3), C4 vs. Cc (R = 0.652, lag = -5), C1 vs. Cf (R = 0.628, lag = -4), C1 vs. Ca (R = 0.617, lag = -2) and C1 vs. Cd (R = 0.606, lag = -3), where GC-MS detection occurred earlier. These findings suggest moderate correlation with slight retention shifts, highlighting variations in how these compounds are detected across the two analytical methods. To complement the tabulated results, Fig. 6 displays the cross-correlation plots for selected compound pairs within the high moderate correlation range (0.6 ≤ R ≤ 0.7).



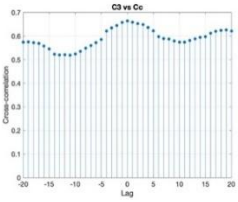
(a) C4 vs. Cd



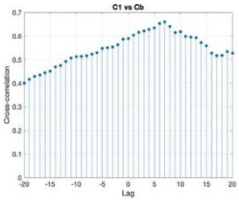
(b) C3 vs. Cd



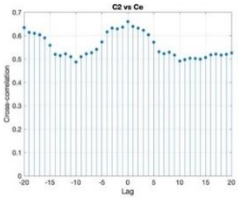
(c) C1 vs. Ce



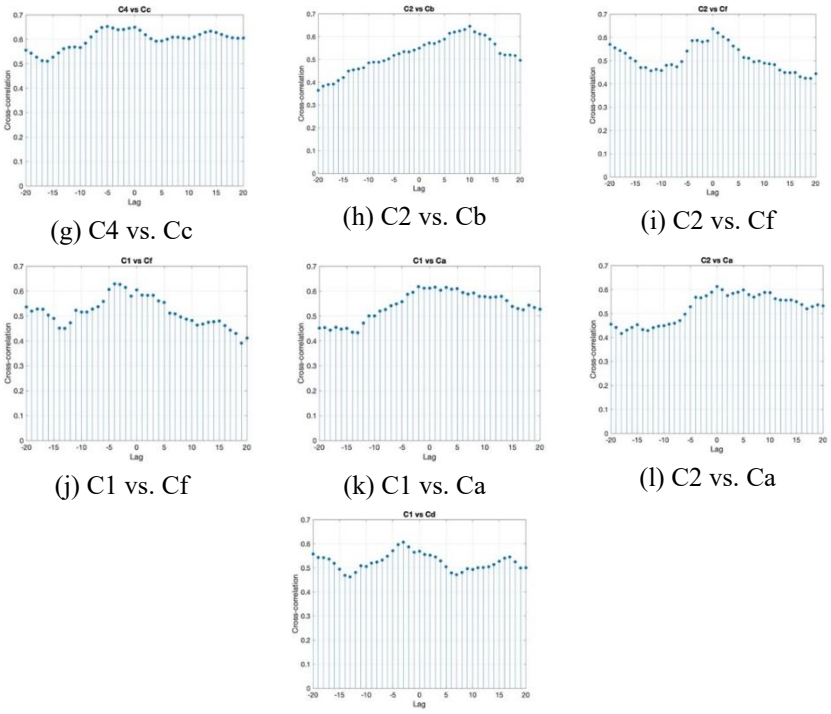
(d) C3 vs. Cc



(e) C1 vs. Cb



(f) C2 vs. Ce



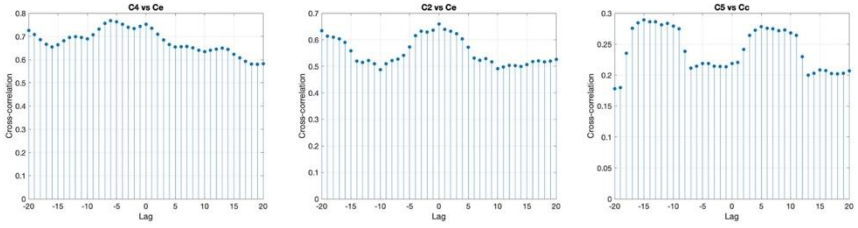
(g) C4 vs. Cc (h) C2 vs. Cb (i) C2 vs. Cf

(j) C1 vs. Cf (k) C1 vs. Ca (l) C2 vs. Ca

(m) C1 vs. Cd

Fig. 6. Cross-correlation plots for high | moderate compound pairs ($0.6 \leq R \leq 0.7$).

Figure 7 illustrates three representative cross-correlation plots, showcasing different levels of correlation strength which are high, moderate and low correlations between GC-MS and GC-MS coupled with GC-FID compounds. These plots provide insight into how compound intensities align across the two analytical techniques and how retention time shifts influence correlation values.



(a) (b) (c)

Fig. 7. Representative cross-correlation plots showing the relationship between GC-MS and GC-MS coupled with GC-FID Compounds. (a) High correlation: Strong alignment with minimal lag, (b) Moderate correlation: Partial alignment with slight variations, (c) Low correlation: Weak alignment with inconsistent detection patterns.

As shown in Fig. 7(a), the cross-correlation plot for C4 vs. Ce exhibits the highest correlation (R=0.767), with a lag of -6. This strong correlation suggests that

γ-eudesmol is consistently detected across both GC-MS and GC-MS/GC-FID, with only a minor retention time shift. The negative lag value indicates that C4 (γ-eudesmol in GC-MS) appears slightly earlier than Ce (γ-eudesmol in GC-MS/GC-FID), which may be attributed to differences in column polarity, instrumental settings, or slight variations in peak detection.

The cross-correlation plot displays a well-defined peak, confirming that the signal patterns are highly similar despite the time shift. This result suggests that γ-eudesmol is a reliable marker compound for agarwood oil characterization, as it maintains its presence across different chromatographic techniques. Figure 7(b) shows the cross-correlation plot for C2 vs. Ce (R=0.659, lag = 0) represents a moderate correlation, indicating that while these compounds show a reasonable level of similarity, they are not as strongly aligned as highly correlated pairs. The lag of 0 suggests that both compounds are detected at the same time across GC-MS and GC-FID, meaning no significant retention time shift is observed.

However, the peak alignment in the cross-correlation function is broader and less defined than in the high-correlation case, implying that variations in peak intensity and shape may have contributed to a slightly lower correlation coefficient. This result suggests that while α-agarofuran and γ-eudesmol exhibit some degree of similarity in their detection patterns, their classification potential may require further investigation. Additional factors such as compound volatility, co-elution effects or minor instrumental fluctuations may influence the correlation strength.

The cross-correlation plot for C5 vs. Cc (R= 0.289) illustrated in Fig. 7(c) shows a low correlation case, where the compound intensities exhibit significant inconsistencies between GC-MS and GC-MS/GC-FID. The low correlation coefficient suggests that these compounds are not strongly aligned, meaning their detection patterns differ across the two techniques. The plot exhibits a scattered and less pronounced peak, indicating weak similarity between the two signals. This could be due to multiple factors, such as variability in compound extraction efficiency, differences in detector response, or compound-specific interactions with the chromatographic column. The absence of a strong peak alignment suggests that C6 is not an ideal candidate for a marker compound, as its inconsistent detection and significant retention time shift reduce its reliability for classification. In contrast, Cc, while exhibiting a lower correlation, aligns better with other compounds and may still be reconsidered for further evaluation.

This result reinforces the importance of excluding low-correlation compound pairs from further analysis, as they may introduce errors in quality grading and species differentiation. The findings highlight the necessity of using a correlation threshold (R>0.5) to ensure that only compounds with strong and meaningful relationships are considered in classification models.

4. Conclusion

This study demonstrates the successful application of cross-correlation analysis to evaluate the alignment of agarwood oil compounds between GC-MS and GC-MS coupled with GC-FID techniques. By analysing 96 GC-MS samples and 120 GC-MS/GC-FID samples, using peak area percentages (%) of seven GC-MS compounds (C1-C7) and six GC-MS/GC-FID compounds (Ca-Cf). The results indicate that C4 vs. Ce (γ-eudesmol vs. γ-eudesmol) and C4 vs. Cb (γ-eudesmol vs. dihyro-β-agarofuran) exhibited the highest correlation (R=0.767), confirming

strong consistency across both analytical methods. Additionally, C3 vs. Ce (10-epi-ϒ-eudesmol vs. γ-eudesmol) demonstrated a strong correlation (R=0.763, lag = 0), indicating perfect alignment between GC-MS and GC-MS/GC-FID detection.

The heatmap visualization revealed distinct correlation patterns, with C3 (10-epi-γ-eudesmol) and C4 (γ-eudesmol) consistently showing high correlations across multiple compound pairs, while C1 (β-agarofuran) and C2 (α-agarofuran) exhibited moderate correlations, and C5 (longifolol), C6 (hexadecanol) and C7 (eudesmol) displayed weaker alignment. The application of a correlation threshold of R>0.5 and lag within ±6, ensured the selection of well-aligned compounds reinforcing the robustness of the findings.

Cross-correlation plots further reinforced these findings. Highly correlated compounds with minimal lag shifts, such as C4 vs. Ce and C3 vs. Ce, were identified as the most reliable marker compounds for agarwood oil classification. In contrast, C6 vs. Cc exhibited poor alignment (R=0.289, lag=15), suggesting poor alignment and inconsistent detection across both techniques. However, Cc, despite having a lower correlation value, demonstrated stable alignment with other compounds, suggesting potential for further validation.

These findings underscore the effectiveness of cross-correlation analysis in identifying reliable chemical markers for agarwood oil quality classification. The identification of γ-eudesmol and 10-epi-γ-eudesmol as the most consistently detected compounds provides valuable insights for essential oil authentication. Future research should explore advanced retention time alignment techniques to further refine compound correlation accuracy across different analytical platforms. Incorporating machine learning models and expanding the dataset with more species and extraction methods can further enhance classification robustness. These strategies would improve the analytical reliability of agarwood oil assessment and support standardized grading for industrial and regulatory applications.

While this study prioritized a focused subset of compounds to ensure analytical consistency and interpretability, it is acknowledged that the limited number of chemical markers may constrain the generalizability of the findings. Given the inherent chemical complexity of agarwood oil, which varies across species, environmental factors and extraction techniques, future work will aim to incorporate a broader range of compounds and more diverse sample sources. This expansion will be critical for validating the proposed approach and enhancing its applicability to large-scale essential oil authentication and classification systems.

5. Acknowledgement

The authors gratefully acknowledge the financial support provided by the Ministry of Higher Education Malaysia through the Fundamental Research Grant Scheme (FRGS), under grant number FRGS/1/2023/ICT11/UITM/02/1. The insightful feedback and contributions from members of the Advanced Signal Processing Research Interest Group are deeply appreciated. The authors also wish to extend their gratitude to the Bio-Aromatic Research Centre of Excellence (BARCE) at Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA) for their invaluable assistance with data extraction. Special thanks are also due to the Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM) Shah Alam, for their generous funding support.

References

1. Naziz, P.S.; Das, R.; and Sen, S. (2019). The scent of stress: Evidence from the unique fragrance of agarwood. *Frontiers in Plant Science*, 10, 840.
2. Gogoi, R. et al. (2023). Agarwood (Aquilaria malaccensis L.) a quality fragrant and medicinally significant plant based essential oil with pharmacological potentials and genotoxicity. *Industrial Crops and Products*, 197, 116535.
3. Adhikari, S.R.; K. Pokhrel, K.; and Baral, S.D. (2021). Economic value of agarwood and its prospects of cultivation. *International Journal of Applied Sciences and Biotechnology*, 9(1), 23-31.
4. Van, T.Y.; Duong, T.T.; and Doan, T. (2024). An updated review of aquilaria species: distribution, chemical constituents and authentication methods. *Asian Journal of Plant Sciences*, 23(2), 146-167.
5. Ismail, N.; Azah, M.A.N.; Jamil, M.; Rahiman, M.H.F.; Tajuddin, S.N. and Taib, M.N. (2013). Analysis of high quality agarwood oil chemical compounds by means of SPME /GC-MS and Z-score technique. *Malaysian Journal of Analytical Scienc*es, 17(3), 403-413.
6. Sharma, S.S.S.M.; Atyam, V.; Avidi, M.; and Ayinala, D. (2024). A Review on Gas chromatography-Mass spectroscopy. *Journal of Pharma Insights and Research*, 2(1), 9-13.
7. Ismail, N. (2021). Analysis of GC-FID and GC-MS microwave-assisted hydrodistillation extraction (MAHD) of agarwood chips. *International Journal of Integrated Engineering*, 13(6), 180-189.
8. Fauziah, A. (2020). Effect of essential oils extracted from Chromolaena odorata (L.) RM KING and H. ROB. against Macrotermes carbonarius (HAGEN, 1858) and Globertermes sulphurous (HAVILANG, 1898). Universiti Kebangsaan Malaysia, Serangga, 25(1), 23-38.
9. Sawadogo, I. et al. (2022). Antifungal and antiaflatoxinogenic effects of cymbopogon citratus, cymbopogon nardus, and cymbopogon schoenanthus essential oils alone and in combination. *Journal of Fungi*, 8(2), 117.
10. Ghobakhloo, S.; Khoshakhlagh, A.H.; Morais, S.; and Mazaheri Tehrani, A. (2023). Exposure to Volatile Organic Compounds in Paint Production Plants: Levels and Potential Human Health Risks. *Toxics*, 11(2), 111.
11. Teonata, N.; Wijaya, V.A.; Vithaloka, V.S.; and Thariq, M. (2021). An introduction to different types of gas chromatography. *Jurnal Berkala Ilmiah Sains dan Terapan Kimia*, 15(1), 8-17.
12. Hoque, M.N.; Mondal, M.F.; Sagor, G.H.M.; Hasan, M.M.; and Hannan, A. (2021). Inclusive scenario of natural products obtained from agar plant (Aquilaria sp.) in Bangladesh- A Review. *Journal of Bio-Science*, 29(2), 151-173.
13. Kamatou, G.P.P.; and Viljoen, A.M. (2017). Comparison of fatty acid methyl esters of palm and palmist oils determined by GCxGC-ToF-MS and GC-MS/FID. *South African Journal of Botany*, 112, 483-488.
14. Abdellatif, F. et al. (2021). Minerals, essential oils, and biological properties of melissa officinalis L. *Plants*, 10(6), 1066.
15. Barp, L.; Franchina, F.A.; Purcaro, G.; Tranchida, P.Q.; and Mondello, L. (2017). In-pipette solid-phase extraction prior to flow-modulation

comprehensive two-dimensional gas chromatography with dual detection for the determination of minor components in vegetable oils. *Talanta*, 165, 598-603.

16. Adhikari, B.; Dahal, S.; Sapkota, N.; Baruwal, P.; Bhattarai, B.; Khanal, K.; and Chapagain, N.P. (2018). Field-aligned current and polar cap potential and geomagnetic disturbances: A review of cross-correlation analysis. *Earth and Space Science*, 5(9), 440-455.
17. Manco, A. et al. (2021). Cross-correlations of biogenic volatile organic compounds (BVOC) emissions typify different phenological stages and stressful events in a mediterranean sorghum plantation. *Agricultural and Forest Meteorology*, 303, 108380.
18. Ali, I. et al. (2023). Identification of natural lead compounds against hemagglutinin-esterase surface glycoprotein in human coronaviruses investigated via md simulation, principal component analysis, cross-correlation, H-bond plot and MMGBSA. *Biomedicines*, 11(3), 793.
19. Mi, T.; Wang, D.; Yao, S.; Yang, H.; Che, Y.; and Wu, C. (2022). Effects of salt concentration on the quality and microbial diversity of spontaneously fermented radish paocai. *Food Research Internati*onal, 160, 111622.
20. De Bakker, J.M.; De Korte, C.L.; and Saris, A.E. 2024. Cascaded plane wave ultrasound for blood velocity vector imaging in the carotid artery. IEEE *Transactions on Ultrasonics*, *Ferroelectrics*, *and Frequency Control*, 71(3), 366-379.
21. Duma, L.; Hediger, S.; Lesage, A.; Sakellariou, D.; and Emsley, L. (2003). Carbon-13 lineshapes in solid-state NMR of labeled compounds. Effects of coherent CSA-dipolar cross-correlation. *Journal of Magnetic Resonance*, 162(1), 90-101.
22. Mohammed, M.M.; Ali, N.S.M.; Alalwan, H.A.; Alminshid, A.H.; and Aljaafari, H.A. (2021). Synthesis of ZnO-CoO/Al_2O_3 nanoparticles and its application as a catalyst in ethanol conversion to acetone. *Results in Chemistry*, 3, 100249.
23. Alalwan, H.A.; Alminshid, A.H.; Mohammed, M.M.; and Mohammed, M.F. (2023). Methane activation on metal oxide nanoparticles: spectroscopic identification of reaction mechanism. *Particulate Science and Technology*, 41(5), 653-660.
24. Alminshid, A.H.; Alalwan, H.A.; Abdulghani, H.A.; and Mohammed, M.M. (2022). Spectrophotometric study of ephedrine hydrochloride in drug using molecular absorption UV-Visible. *Spectrochimica Acta Part A Molecular and Biomolecular Spectroscopy*, 270, 120828.
25. Alalwan, H.A.; Ali, N.S.M.; Mohammed, M.M.; Mohammed, M.F.; and Alminshid, A.H. (2023). A comparison study of methyl green removal by peroxi-coagulation and peroxi-electrocoagulation processes. *Cleaner Engineering and Technology*, 13, 100623.
26. Habl, A.; Amooey, A.; Mustafa, M.; and Alalwan, H.A. (2024). Electro oxidation process for wastewater treatment in petroleum refineries. *Pollution*, 10(2), 819-832.
27. Al‐Amshawee, S.K.A.; Yunus, M.Y.B.M.; Alalwan, H.A.; Lee, W.H.; and Dai, F. (2022). Experimental investigation of biofilm carriers of varying

shapes, sizes, and materials for wastewater treatment in fixed bed biofilm reactor: a qualitative study of biocarrier performance. *Journal of Chemical Technology and Biotechnology*, 97(9), 2592-2606.

28. Yusoff, Z.M.; and Ismail, N. (2024). Datasets of chemical compounds in three different species of aquilaria using GC-MS coupled with GC-FID analysis. *Data in Brief*, 53, 110209.
29. The MathWorks, Inc. (2024). *MATLAB* (Version R2024a), data accessed: 2021-04-29.
30. Jaradat, A.; Lutfiyya, H.; and Haque, A. (2020). Demand response for residential uses: A data analytics approach. P*roceedings of the IEEE* 6th *world forum on Internet of Things*, *WF-IoT*, New Orleans, LA, USA.
31. Zhang, L.; Wang, H.; Tang, X.; Lu, S.; Tan, Y.; Li, Q.; and Wu, J. (2022). Characterization of aroma volatiles in xilin fire ginger oils by HS-SPME-GC-MS. *International Journal of Food Properties*, 25(1), 53-64.
32. Attia, I.; and Sadot, D. (2024). *Interlocking of mode-locked lasers utilizing optical communication equipment*. Signal Processing in Photonic Communications (SpTh1H-4). Optica Publishing Group.