Journal of Breath Research

PAPER • **OPEN ACCESS**

Wavelet-enhanced TD–GC–MS analysis of molecular pattern alterations in gas samples induced by breath sampling devices

To cite this article: Nicoletta Ardito *et al* 2026 *J. Breath Res.* **20** 016013

Journal of Breath Research

PAPER

OPEN ACCESS

RECEIVED
22 September 2025

REVISED
14 January 2026

ACCEPTED FOR PUBLICATION
26 January 2026

PUBLISHED
12 February 2026

Wavelet-enhanced TD–GC–MS analysis of molecular pattern alterations in gas samples induced by breath sampling devices

Nicoletta Ardito[1], Arianna Elefante[1,2,5], Marilena Giglio[1], Andrea Zifarelli[1,*], Laura Facchini[3], Pietro Patimisco[1], Vincenzo Spagnolo[1], Nicola Amoroso[4,5] and Angelo Sampaolo[1,5]

[1] PolySense Lab-Dipartimento Interateneo di Fisica, University and Polytechnic of Bari, Via Amendola 173, Bari, Italy
[2] Institute for Photonics and Nanotechnologies (IFN), National Research Council, Via Amendola 173, Bari, Italy
[3] Predict S.p.A., Viale Adriatico, c/o Fiera del Levante Pad. 105, 70132 Bari, Italy
[4] Dipartimento di Farmacia-Scienze del Farmaco, University of Bari Aldo Moro, Bari, Italy
[5] Centro di Tecnologie Innovative per la Rilevazione e l'Elaborazione del Segnale (T.I.R.E.S.), Università degli studi di Bari Aldo Moro, Bari, Italy
[*] Author to whom any correspondence should be addressed.

E-mail: andrea.zifarelli@uniba.it

Keywords: breath sampler, wavelet analysis, VOCs

Abstract

This study presents a computational method to identify volatile organic compound (VOC) artifacts introduced by breath sampling hardware. To exclude endogenous biological variability, ambient air was collected using two sampling devices working in the same experimental conditions: the Mistral end-tidal breath sampler and the ACTI-VOC PLUS pump, a low-emission reference system. VOCs were pre-concentrated on sorbent-packed thermal desorption (TD) tubes and analyzed by TD–gas chromatography–mass spectrometry (TD–GC–MS). Differential chromatograms obtained by subtracting ACTI-VOC signals from Mistral traces were processed using stationary wavelet transform (SWT) to selectively enhance high-frequency features indicative of artifactual emissions. Four new compounds not previously associated with Mistral sampling hardware were consistently detected in Mistral samples and were absent in ACTI-VOC pump controls: 1,3,5-trioxane, 1,3,5,7-tetroxane, (Acetyloxy)acetic acid, and N,N-dimethylformamide. These molecules are indicative of polymer degradation, acetal resin breakdown, and material off-gassing specific to the breath sampler.

1. Introduction

Analysis of exhaled breath has increasingly been recognized as a non-invasive approach for evaluating systemic metabolic, inflammatory, and pathological conditions. This diagnostic capability derives from the presence of volatile organic compounds (VOCs), carbon-based molecules produced by endogenous biochemical activity or host-microbiota interactions. Once formed, VOCs enter the bloodstream and subsequently diffuse across the alveolar-capillary membrane into exhaled breath, facilitating their analytical detection [1, 2].

The terminal phase of exhalation, commonly known as the end-tidal fraction (typically the last 150-200 ml of expired air), is considered the most representative of alveolar gas composition. This fraction minimizes dilution from anatomical dead space and upper airway contaminants, thereby enhancing the biological specificity and reliability of collected samples [3, 4].

Among the available analytical techniques, thermal desorption (TD)–gas chromatography–mass spectrometry (TD–GC–MS) is widely regarded as the gold standard for breath VOC analysis, offering ultra-trace sensitivity and high-resolution separation of complex VOC mixtures [5, 6]. Nonetheless, the accuracy and interpretability of breath VOC profiles are frequently compromised by artifactual signals introduced during sample collection, handling and storage. Traditionally, exhaled breath samples were collected using polymeric sampling bags, such as Tedlar (polyvinyl fluoride), Nalophan (polyethylene terephthalate), and Teflon (PTFE). Although cost-effective, these materials are known to introduce artifactual VOCs, such as aromatic hydrocarbons, ketones, phenols, cyclic ethers, and aliphatic hydrocarbons [7, 8]. Additionally, polymeric bags

are susceptible to permeation losses and polymer degradation, compromising sample integrity and reproducibility [9, 10]. To mitigate these limitations, modern breathomics workflows rely on direct pre-concentration of VOCs onto sorbent-packed TD tubes, thus minimizing exposure to reactive or emissive surfaces. Specialized breath sampling devices, such as the commercially available ReCIVA system (Owlstone Medical, UK) and the Mistral system (Predict s.r.l., Italy), have been developed for this purpose [11, 12]. The ReCIVA employs a silicone face mask with integrated carbon dioxide and pressure sensors, enabling fractionated sampling onto multiple sorbent tubes. Conversely, the Mistral system is tailored for volume-controlled, thermoregulated end-tidal collection while simultaneously sampling ambient air for background correction. Both systems employ low-emission materials to reduce artifactual contributions and controlled flow dynamics to improve sampling reproducibility. Despite recent advancements in sampling technologies, residual emissions, including 1,4-pentadiene, 2-hexanone, and cyclic siloxanes, persistently appear in device blanks, likely originating from PVC thermal degradation, polyethylene, and solubilizers used in manufacturing [13]. Although typically observed at low intensities, these background signals can co-elute with biological ones, thus representing critical confounders in biomarker discovery efforts.

Conventional chromatographic preprocessing techniques, such as baseline correction, peak deconvolution, and retention time (RT) alignment, are essential for improving data quality and enhancing the resolution of complex chromatographic profiles, but often prove to be inadequate for fully resolving overlapping artifactual and biological peaks [14–16]. As a result, there is growing interest in advanced signal processing tools capable of distinguishing biological signals from device-related noise. Among these, Wavelet transform-based methods have proven to be highly effective in performing multiscale time-frequency decomposition [17]. These approaches have been successfully applied in contexts like biosignal analysis and analytical spectrometry [18–21]. In particular, the stationary wavelet transform (SWT) offers key advantages over traditional discrete wavelet transform, most notably shift invariance and the preservation of full temporal resolution, features essential for chromatographic analysis where RT accuracy is critical [22–24].

In this work, a computational framework based on SWT is proposed to evaluate its potential for isolating high-frequency signal components associated with artifactual emissions of the sampling device for breath analysis. The methodology is applied to ambient air samples acquired using a commercial end-tidal sampling system (Mistral) and a low-emission reference device (ACTI-VOC PLUS), with the aim of identifying hardware-derived VOCs and assessing their potential impact on untargeted VOC profiling.

2. Materials and methods

To characterize VOCs potentially emitted by breath sampling device, a structured analytical workflow was developed. Ambient air was selected as a reference matrix due to its reduced contributions of human-derived VOCs, thereby enabling selective detection of artifactual emissions originating from device materials. During sampling, a single operator remained at a distance from the devices, while controlled laboratory conditions, including closed doors, active air recirculation, and filtered exhaust, were maintained to stabilize and reduce temporal variability in ambient air composition. The experimental workflow consisted of five sequential stages: (i) ambient air collection onto TD tubes using both the test device (Mistral breath sampler) and a low-emission reference system (ACTI-VOC PLUS pump) (ii) analysis by TD–GC–MS; (iii) chromatographic processing (iv) SWT decomposition of the chromatograms for artifact detection and (v) compound identification with mass spectral deconvolution.

Replicate measurements were performed across two independent sessions to assess repeatability and to account for background variability associated with the Mistral sampler itself. In the first session, both devices were run simultaneously in the same controlled environment to ensure direct comparability and minimize environmental fluctuations. In the second session, measurements were conducted in two separate environments to evaluate whether operation of the Mistral sampler could locally modify ambient VOC composition. Sampling procedures and instrumental conditions were standardized to ensure cross-device and inter-day consistency.

2.1. Data collection and TD-GC-MS acquisition

A schematic overview of the analytical configuration for sample collection and GC-MS detection is provided in figure 1.

VOC sampling was performed using two offline devices: the Mistral breath sampler and the ACTI-VOC PLUS pump (Markes International, UK). For both sampling systems, VOCs were collected onto stainless-steel TD tubes (Markes International, Part No. C2-AXXX-5149) packed with a dual-bed sorbent of Tenax® TA and Carbograph™ 5TD, suitable for capturing a broad range of volatile and semi-volatile compounds. Prior to sampling, all tubes were thermally preconditioned with nitrogen at high temperatures to remove residual contaminants.

The Mistral is a breath sampling system designed for end-tidal fractions collection. It features a dual-inlet configuration enabling sequential sampling of exhaled and background ambient air into the TD

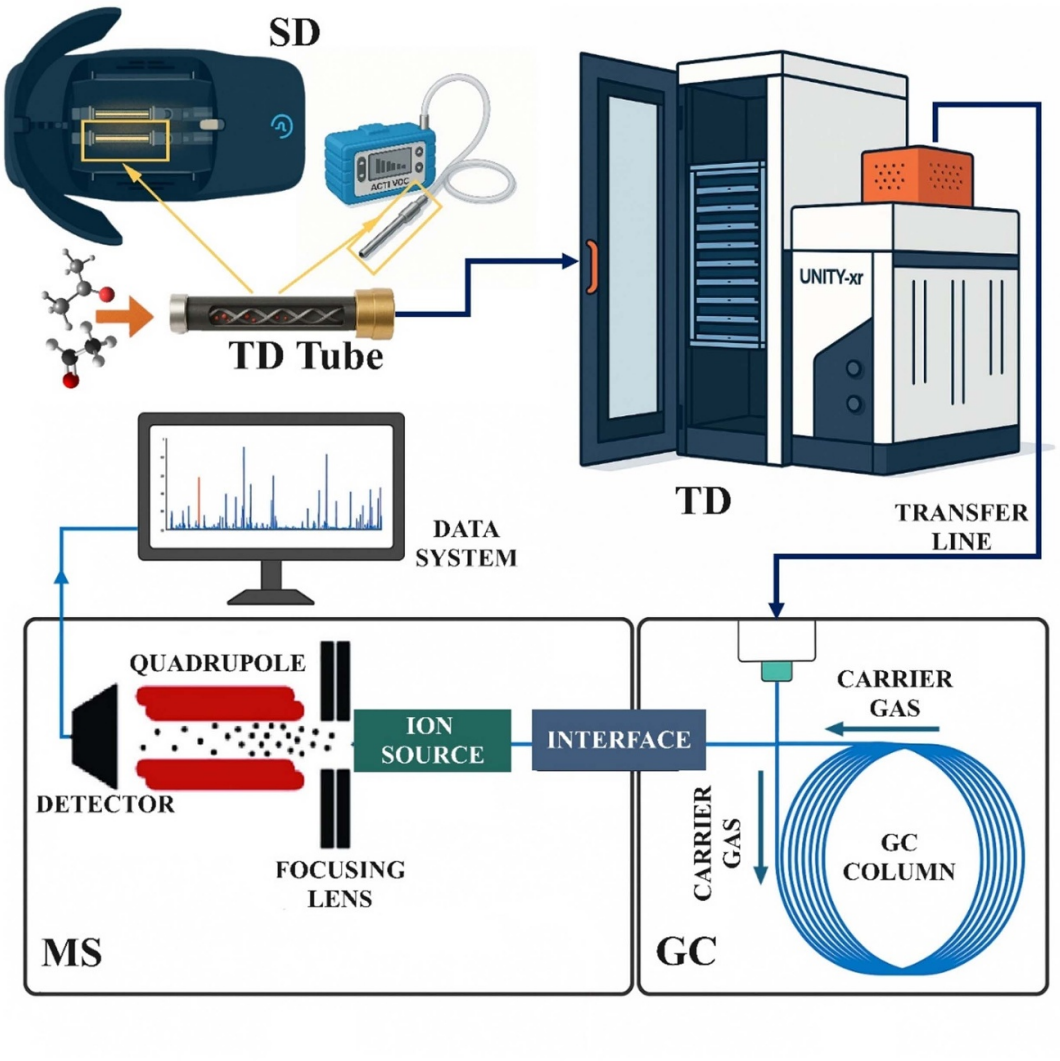

Figure 1. Schematic of the employed TD-GC-MS setup; SD, sampling devices: on the left Mistral sampler, on the right ACTI-VOC pump; TD Tube, thermal desorption tube; TD, thermal desorber; GC, gas chromatograph; MS, mass spectrometer.

tubes which are connected downstream of the dual-inlet manifold. A thermoregulated stainless-steel sampling line (36 °C-37 °C) prevents condensation and an integrated purge cycle cleans internal pathways before each sampling event. The device operates at a fixed flow rate of 200 ml min^{-1} and is programmed to collect samples of 800 ml over a 4 min period. For this study, the Mistral sampler was employed in ambient-collection mode, using only the air inlet.

The ACTI-VOC PLUS is a portable, battery-powered sampling pump optimized for ambient air collection onto TD tubes. It features adjustable flow control and is characterized by low intrinsic VOC emissions. To ensure comparability with the Mistral device, the pump was operated under identical sampling conditions, setting the flow rate to 200 ml min^{-1} and the sampling time to 4 min.

TD was performed using a Markes ULTRA-UNITY xR system. Prior to each experimental session, system performance was verified by loading TD tubes with multiple concentrations of a certified liquid standard using the Calibration Solution Loading Rig (CSLR, Markes International), followed by analysis to assess linearity and overall instrument response. Additionally, trap blanks were analyzed at the beginning and end of each session to confirm the absence of carryover. Following system validation, VOCs were desorbed using the ULTRA-UNITY in a two-stage desorption mode. A 3.0 min pre-purge was first applied at 30 ml min^{-1} under helium flow to remove water and light volatiles. Subsequently, during the primary desorption stage, VOCs were desorbed at 300 °C for 10.0 min under helium flow at 50 ml min^{-1}. To minimize back diffusion, a helium trap purge was applied at 70 ml min^{-1} for 1.0 min. The analytes were then cryofocused on a cold trap (Part No. U-T14WMT-2S, Markes International), maintained at 5 °C. In the secondary stage, the trap was rapidly heated to 300 °C at a rate of 25 °C s^{-1} and held for 3.0 min. The analytes were transferred to the GC inlet using

a deactivated fused-silica transfer line insert (Part No. SERUTE-5099, Markes International), heated to prevent condensation. Gas chromatographic separation was carried out using an Agilent 8860 GC equipped with a VOCOL capillary column (length: 30 m, internal diameter: 0.25 mm, 1.0 μm film thickness; Supelco, Merck). The GC oven was operated under a multi-stage temperature program with the following steps: an initial isothermal hold at 37 °C for 5.0 min, a ramp of 6 °C min^{-1} up to 180 °C, a ramp of 3 °C min^{-1} up to 185 °C to improve resolution of closely eluting mid-range analytes, and a final ramp of 70 °C min^{-1} up to 220 °C to ensure complete elution of late-retained compounds and thermal conditioning of the column. High-purity helium was used as the carrier gas at a constant flow rate of 1.6 ml min^{-1}.

VOCs detection was performed using an Agilent 5977C mass spectrometer operated in electron ionization (EI) mode with a single quadrupole as detector. Data were acquired in full-scan mode over the m/z range 35-250 at 6.2 spectra s^{-1} (mass resolution increment: 0.1 u). The ion source and quadrupole were thermally regulated at 230 °C and 150 °C, respectively, to maintain ionization efficiency and mass stability. Prior to analysis, instrument performance was verified using Agilent's autotune protocol, and the electron multiplier voltage was automatically optimized via eTune calibration to ensure consistent sensitivity and signal stability throughout the acquisition.

To distinguish VOCs potentially released by the Mistral sampler, the experimental design was conceived to isolate sampler-derived artifacts from ambient VOC background. Sampling was therefore conducted across two independent analytical sessions, with 20 consecutive replicates collected in each session for each sampling mode. In Session 1, both devices were operated simultaneously within a closed, access-controlled indoor environment (Env1) to minimize environmental fluctuations. Air conditions were maintained by a continuously operating filter and dehumidifier, providing quasi-stable conditions throughout the sampling period. Consecutive sampling for each device thus ensured quasi-stationary ambient conditions over a limited temporal window of a few hours, enhancing sampler-associated features while constraining short-term variability in ambient air composition. VOCs detected by both devices were attributed to the ambient air and were therefore excluded from classification as Mistral-derived artifacts. However, compounds potentially released by the Mistral sampler in Env1 could also be collected by the ACTI-VOC when both devices shared the same environment. This cross-collection effect reduces the intensity of the corresponding signals in the differential chromatograms and may hinder their detection during the fourth step of the experimental workflow, namely the SWT-based decomposition for artifact identification. In Session 2, the Mistral was again deployed in Env1, while the ACTI-VOC pump was operated in a separate indoor location (Env2). Although this configuration does not imply equivalence in terms of VOCs composition between the two environments, it prevents cross-contamination ensuring that VOCs potentially released by the Mistral sampler are not collected by the reference pump. This cross-session design allowed more effective and reproducible identification of device-specific artifacts, allowing them to be distinguished from ambient-derived VOCs.

2.2. Chromatograms processing

The GC-MS Agilent software returns the total ion current (TIC) chromatograms, which show the sum of the intensities of the ions detected across all masses as a function of time. The TIC traces were exported and processed using a custom Python-based computational pipeline. For both sampling devices, replicate TIC traces were normalized using min-max scaling and interpolated onto a common temporal grid. This normalization step was applied to facilitate retention-time alignment under inter-run variability in absolute signal intensity. To correct for RT deviations between replicates, typically arising from minor thermal or flow instabilities, a peak-based alignment process was employed. Peak detection was performed using the find_peaks function from the scipy.signal module [25]. Detection parameters, including peak prominence (0.01) and inter-peak distance (5 points), were chosen based on the instrument's acquisition rate and chromatographic resolution (see section 2.1). In particular, the minimum distance between peaks was constrained to ensure consistency with the chromatographic peaks' width (FWHM 2–3 Hz, see section 3.2, figure 3). Peak detection was therefore applied to normalized TIC traces to reliably identify reproducible peak centroids for alignment.

For each sampling device, the first acquired chromatogram was designated as the reference trace to define a common retention-time axis. Peak centroids from subsequent replicates were matched to the reference within a $\pm$0.2 min RT window, consistent with the instrument's acquisition rate and the observed RT stability across replicates. Time-domain corrections were computed via linear interpolation to construct a custom stretch function that realigned the chromatograms while preserving local peak shape and intensity. As this stretch function acts exclusively on the time axis, it was subsequently applied to the original, non-normalized TIC traces, thereby preserving the original signal intensities and relative peak contributions. This adaptive alignment enabled a multi-peak adjustment that accounted for non-uniform RT variation across the chromatograms. The aligned TIC traces were then averaged to obtain a session-specific

representative chromatogram for each device, thereby suppressing stochastic noise. Differential artifact profiling was performed by subtracting the session-matched ACTI-VOC pump average profile from the corresponding Mistral trace, for each session. The resulting differential signals constituted the input for SWT-based artifact localization.

2.3. Artifact detection with SWT

To further improve the detection of low-abundance, device-derived features, the SWT was applied to differential TICs obtained as outlined in section 2.2. Unlike traditional smoothing or baseline correction techniques, SWT enables multiscale decomposition while preserving the original RT axis. This shift-invariant property is crucial for TD–GC–MS data, where the temporal domain encodes chromatographic separation, as in TIC traces the RT acts as the critical reference for accessing compound-specific mass spectra. Therefore, preserving its integrity is essential for downstream molecule identification.

A subset of orthogonal Daubechies wavelets was selected based on their compact support, vanishing moments, and established efficacy in detecting localized, transient signal components in analytical data [26]. The maximum decomposition level was determined using the swt_max_level function, of the pywt pyhon library, and validated by comparing the wavelet's effective support L_j with the FWHM of the chromatographic peaks. To avoid loss of resolution, decomposition was constrained such that $L_j \ll \mathrm{FWHM}_{\mathrm{peak}}$ thus preventing oversmoothing of sharp features [27, 28]. Each wavelet was evaluated by reconstructing the signal via inverse SWT and calculating the mean squared error (MSE) with respect to the original differential TIC. The wavelet minimizing the MSE was selected for further processing. To extract reproducible artifacts while controlling for stochastic environmental variability, only features consistently present in the detail coefficients across both experimental sessions were retained. This filtering step enhanced the specificity of artifact identification by leveraging inter-session redundancy as a statistical constraint.

2.4. Artifact identification and validation

Candidate artifactual signals highlighted by SWT were subjected to tentative compound identification using two complementary Agilent software: Unknowns Analysis (v. 10.0) for automated spectral deconvolution and Qualitative Analysis (v. 12.0) for manual inspection [29, 30]. This dual-platform approach ensured internal consistency between automated and manual spectral evaluation, combining algorithmic resolution of overlapping signals with manual spectral validation, based on EI mass spectral similarity to the NIST library. Unknowns Analysis enabled automated deconvolution of co-eluting peaks and spectral matching against the National Institute of Standards and Technology (NIST) 2017 Mass Spectral Library (NIST17). Only compounds exhibiting match factors $\geqslant$80% and well-resolved chromatographic peaks were retained for further evaluation. Each candidate was then independently reviewed in Qualitative Analysis. The base peak identified during deconvolution was used to generate an extracted ion chromatogram (EIC), a chromatographic trace of a specific m/z ion over time, enabling assessment of RT reproducibility, and chromatographic isolation. From this EIC, the associated mass spectrum was re-extracted and subjected to a second library search (NIST17). This secondary match confirmed the internal consistency between the automated deconvoluted spectrum and the manually validated ion signal, thereby reinforcing confidence in the tentative compound assignment.

A feature was classified as a confirmed artifact only if it met the following criteria: (i) reproducible detection in all Mistral replicates across both analytical sessions; (ii) complete absence in ACTI-VOC pump controls; and (iii) temporal co-localization using SWT in the differential TIC. Final tentative identifications were accepted exclusively when internal consistency between automated deconvolution and manual spectral evaluation was observed across both software platforms.

3. Results and discussion

3.1. Chromatograms processing

The effectiveness of the chromatographic alignment and differential artifact profiling is reported in figure 2.

Panel (a) shows the session-averaged TIC chromatograms obtained with the Mistral sampler and the ACTI-VOC pump after alignment processing. In the inset, the unaligned TIC acquired with the ACTI-VOC pump (red dashed line) is shown together with the aligned TICs, highlighting the temporal shift in RTs before the alignment procedure. Panel (b) presents the differential TIC profile, computed by subtracting the aligned, session-averaged ACTI-VOC trace from the Mistral counterpart. This subtraction effectively removed shared environmental contributions, isolating VOC features more likely attributable to emissions inherent to the Mistral sampling system.

3.2. Artifact localization and identification

A SWT was applied to the signal acquired at 6.2 Hz to isolate chromatographic transients from high-frequency noise and baseline drift. Among the Daubechies wavelet family, the Daubechies-3 (db3) wavelet yielded the lowest reconstruction error ($\mathrm{MSE} = 2.69 \times 10^{-20}$) and was therefore selected for subsequent analysis. The SWT was first applied to the chromatographic signal without padding. The maximum obtained decomposition level, corresponding to 1, enabled effective suppression of high-frequency

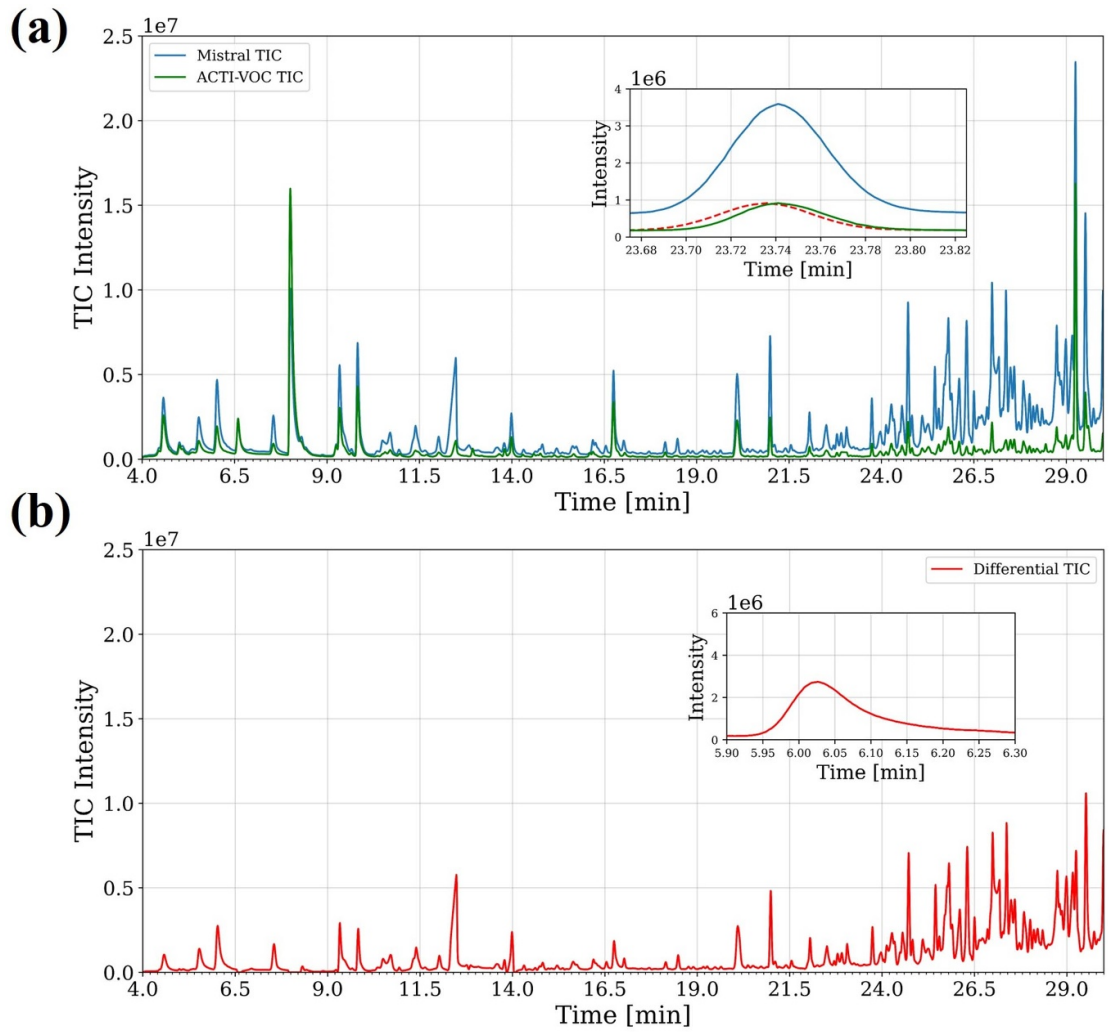

Figure 2. (a) Total ion chromatograms (TIC) obtained using the Mistral sampler (blue solid line) and the ACTI-VOC pump (green solid line) after alignment processing. In the inset, in red is shown the ACTI-VOC pump TIC prior to the alignment (red dashed line) (b) Differential TIC (red solid line).

noise and baseline drift while preserving the transient structures. To visualize the full multiscale structure of the signal, the SWT was subsequently applied to a zero-padded version of the chromatogram. Padding allowed the decomposition to reach the theoretical maximum level determined by the padded signal length, enabling the inspection of transient features across a broad range of time scales. The temporal resolution associated with each decomposition level was estimated based on the wavelet filter length, sampling interval, and scale factor. Only decomposition levels with a temporal support shorter than the total acquisition time were retained for further analysis.For each detected transient, the optimal decomposition level was determined by comparing its FWHM to the effective temporal resolution of each scale. The highest level at which the time resolution remained smaller than or comparable to the measured FWHM was assigned to each event [31]. This adaptive approach allowed narrower transients to be analyzed at the finest scales, preserving temporal resolution, while broader peaks were assigned to coarser scales to avoid oversmoothing. Figure 3 shows the results of the multiscale analysis along with the distribution of optimal decomposition levels assigned to each detected peak.

Figure 3 a shows the scalogram obtained from the squared modulus of the SWT detail coefficients, providing a time-frequency representation of the chromatographic signal. The quantized levels on the frequency scale of figure 3 arise from the dyadic discretization of the SWT, producing a discrete pattern of frequency rather than a fully continuous spectrum. A dense and continuous distribution of transient activity was observed throughout the entire chromatographic run, indicating that transient events are not confined to specific time windows but occur along the full elution profile. Most of the signal power is concentrated at frequencies below 1 Hz, which is consistent with the characteristic timescales of typical

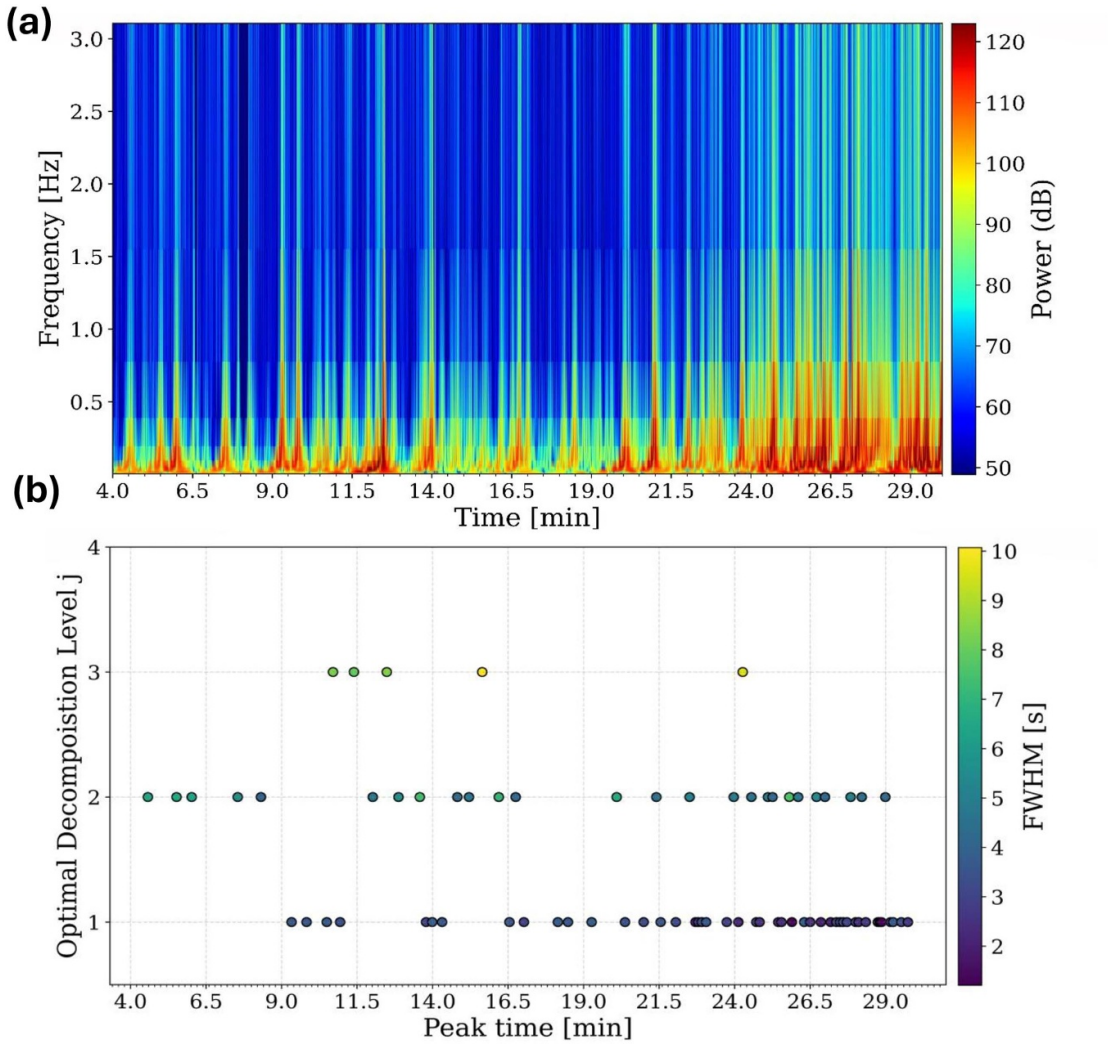

Figure 3. (a) Multiscale scalogram of the stationary wavelet transform (levels 1–11) using a Daubechies-3 wavelet. (b) Optimal decomposition level compared to the peaks FWHM.

chromatographic peaks. However, transient components extending up to 2-3 Hz were also detected, particularly in association with sharp, short-duration events, reflecting the presence of narrow peaks. The multiscale representation effectively enhances both short- and long-duration transient features, allowing for the visualization of minor transients that would otherwise be difficult to detect in the original signal due to noise and baseline fluctuations.

The adaptive selection of the optimal decomposition level for each detected transient is summarized in figure 3(b). The majority of peaks were assigned to level 1, corresponding to a temporal resolution of approximately 1.9 s. This indicates that most chromatographic transients are short-lived, typically with durations between 2 and 4 s, and are best captured at the finest scale of the decomposition. A smaller subset of peaks was assigned to level 2, associated with a temporal resolution of approximately 3.9 s, reflecting transients lasting up to 8 s. Only a few broader events required assignment to level 3, indicative of durations extending up to approximately 10-15 s.

To assess the amount of information preserved across different decomposition levels, the energy of the approximation coefficients was evaluated, as shown in figure 4.

The analysis indicates that more than 90% of the informative content is retained up to level 4, which is consistent with the FWHM-based assessment of peak widths. Therefore, levels above 4 are not expected to provide additional insights into chromatographic peak dynamics. Since the majority of detected transients were assigned to level 1, the corresponding detail coefficients were further analyzed to characterize the time-localized artifact features. The results are shown in figure 5.

Figure 5 displays the level-1 detail coefficients obtained from the SWT using the db3 wavelet [26]. Several well-localized high-frequency deviations are visible, corresponding to a sharp chromatographic event. Most features detected in the wavelet domain did not meet reproducibility or spectral validation criteria (see section 2.4) and were excluded from further analysis, as they likely stem from background

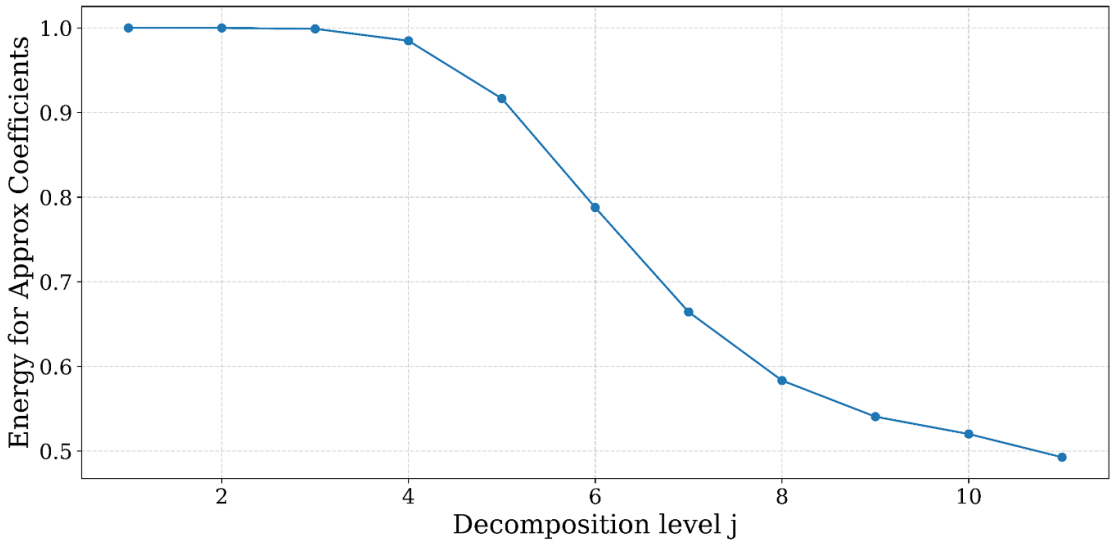

Figure 4. Energy of approximation coefficients as a function of the wavelet decomposition level (j).

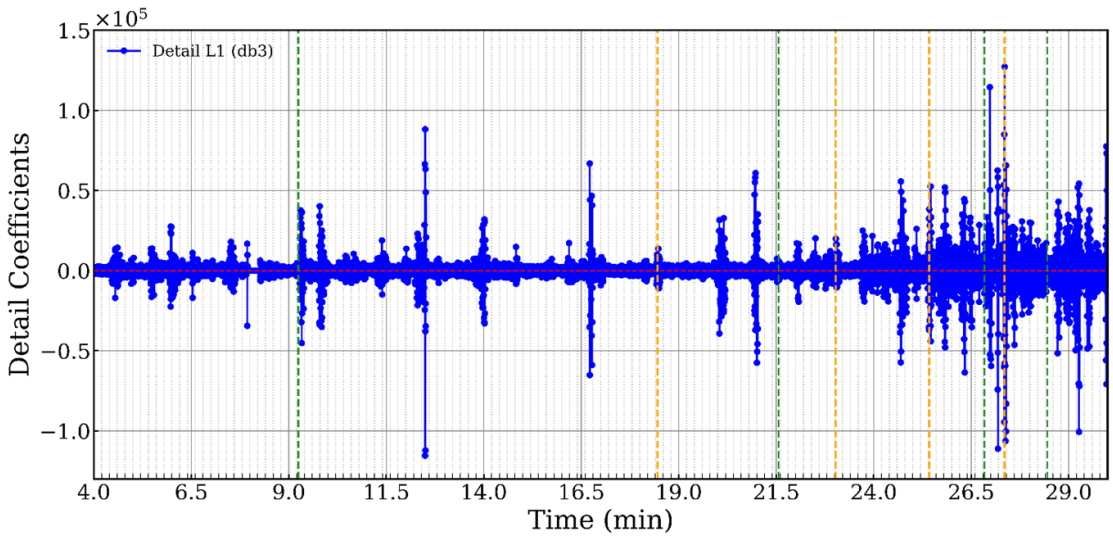

Figure 5. Detail coefficients of the stationary wavelet transform (Debauchian3, level <1).

variability in the sampling cartridges rather than systematic emissions from the device, for example, benzene and toluene at 16.75 min and 20.09 min, respectively (figure 5). Validated features, instead, are marked by vertical dashed lines indicating artifact-associated RTs: green lines refer to compounds previously reported in the literature as Mistral artifacts (1,4 pentadiene, 2-hexanone, 6-methyl-5-hepten-2-one, and 2,2,4,6,6-pentamethyl-heptane), while orange lines indicate newly identified features [13]. All annotated peaks consistently represent compounds detected across both sessions, with reproducibility and device-specific origin confirmed through cross-validation of wavelet coefficients. Notably, wavelet analysis enabled the identification of low-intensity artifactual features by isolating structured detail coefficients from noise.

Four new compounds, 1,3,5-trioxane, N,N-dimethylformamide, (acetyloxy)acetic acid, and 1,3,5,7-tetroxane, were tentative identified by matching their mass spectra against spectral libraries (see section 2.4). Specifically, N,N-dimethylformamide and (acetyloxy)acetic acid originate from polymeric components of the Mistral device, likely as degradation products or residual monomers [32]. Trioxane is associated with acetal resins, a class of thermoplastic polymers based on polyoxymethylene (POM), known to release cyclic oligomers such as trioxane under thermal or oxidative stress [33]. Although 1,3,5,7-tetroxane is not directly attributable to a

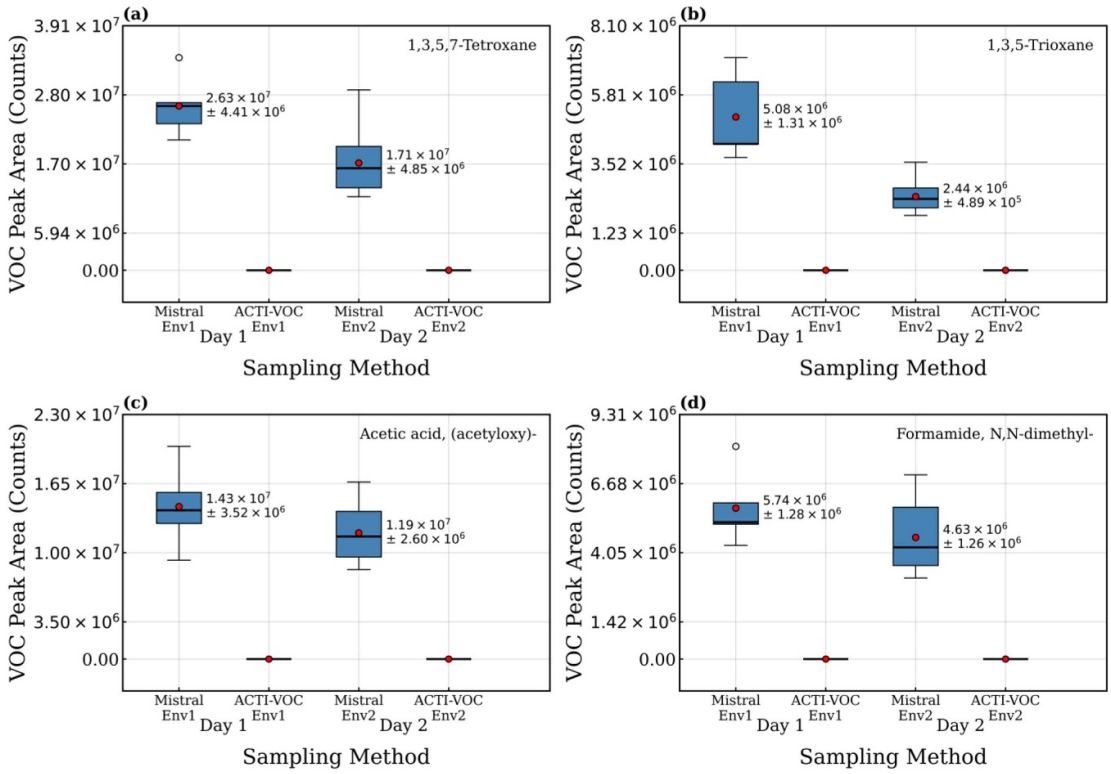

Figure 6. Boxplots showing artifactual VOC peak area in Mistral samples across two sessions. Compounds: (a) 1,3,5,7-Tetroxane, (b) 1,3,5-Trioxane, (c) (Acetyloxy)acetic acid, and (d) N,N-Dimethylformamide.

specific source, its frequent co-detection with trioxane suggests a possible indirect association, potentially arising from related degradation mechanisms or shared precursors [34]. Notably, none of these artifactual compounds were observed in ACTI-VOC pump controls, reinforcing their attribution to Mistral-specific hardware emissions.

3.3. Method reproducibility and temporal emission dynamics

The reproducibility and temporal patterns of the four newly identified artifactual VOCs, 1,3,5,7-tetroxane, 1,3,5-trioxane, (Acetyloxy)acetic acid, and N,N-dimethylformamide, were assessed across the two experimental sessions detailed in section 2.1. Figure 6 illustrates peak area distributions of each compound in Mistral and ACTI-VOC pump samplers across both experimental sessions.

All compounds were consistently detected across all Mistral replicates and were absent from ACTI-VOC pump controls. A smaller decrease in peak area from Session 1 (Day 1) to Session 2 (Day 2) was observed for all analytes, suggesting a depletion effect likely driven by volatilization or desorption from internal components of the device. GC–MS performance stability was assessed prior to each experimental session by triplicate injections of a liquid calibration standard. Consistent peak areas and RTs were observed across the two sessions, confirming instrument stability and excluding instrumental drift as a contributing factor to the observed changes in analytes peak area. Temporal trends observed in Session 2 are further illustrated in figure 7.

Each compound exhibited a smooth, monotonic decrease in signal intensity across subsequent replicates, without abrupt fluctuations, consistent with a passive, concentration-driven release rather than analytical noise. The reproducibility of these decay trends supports a physicochemical origin, plausibly driven by slow volatilization from polymeric substrates, or thermally labile materials intrinsic to the device. Despite the decreasing trend, intra-session variability remained within acceptable limits, with coefficients of variation (CVs) ranging from 16.8% to 25.8% in Session 1 and from 20.0% to 28.4% in Session 2 for the four molecules. Although these values exceed the 20% repeatability threshold specified by EPA Method TO-17 for standard gas mixtures, they are aligned with typical variance observed in TD–GC–MS applications involving real-world matrices [35].

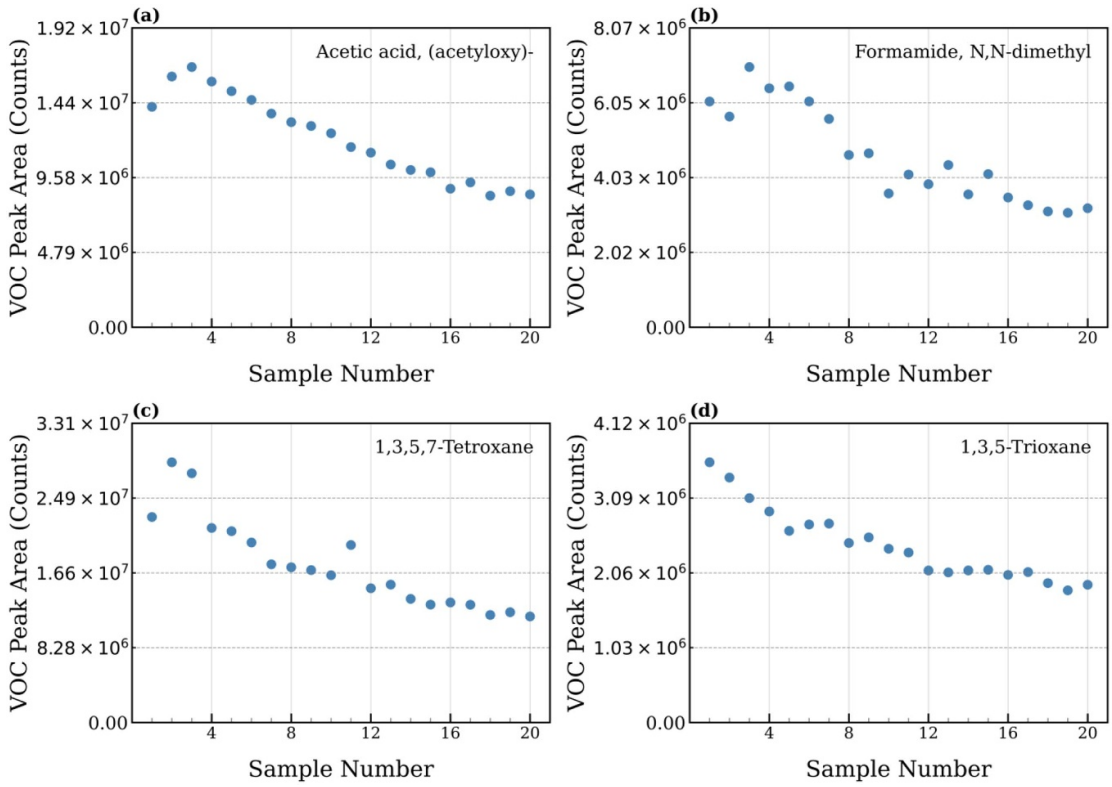

Figure 7. Time-resolved emission profiles for the four artifactual VOCs across 20 Mistral replicates in Session 2. Compounds: (a) 1,3,5,7-Tetroxane, (b) 1,3,5-Trioxane, (c) (Acetyloxy)acetic acid, and (d) N,N-Dimethylformamide.

4. Conclusions

In this work, a computational pipeline integrating differential chromatographic analysis and SWT was developed to detect and suppress artifactual VOCs emitted by breath sampling hardware. The Mistral sampler, operated exclusively in ambient air mode, was benchmarked against a low-emission reference device (ACTI-VOC PLUS pump) across two independent experimental sessions conducted on different days and in distinct indoor environments. This cross-validated design enabled the robust tentative identification of reproducible, device-specific emissions while accounting for background variability. SWT enabled the selective enhancement of low-abundance, high-frequency features while preserving chromatographic RT resolution. Daubechies 3 (db3) provided the lowest reconstruction error, emerging as the most effective for artifact detection.

Four compounds, 1,3,5-trioxane, 1,3,5,7-tetroxane, (Acetyloxy)acetic acid, and N,N-dimethylformamide, were consistently detected in all Mistral replicates and were entirely absent in ACTI-VOC PLUS controls. These compounds are chemically consistent with off-gassing from acetal polymers and plasticizers commonly used in medical sampling equipment. Specifically, 1,3,5-trioxane is a well-documented degradation product of acetal resins, while (Acetyloxy)acetic acid and N,N-dimethylformamide are both known to arise from the degradation of polymeric materials. Although no direct material association has been established for 1,3,5,7-tetroxane, its structural similarity to 1,3,5-trioxane, along with consistent co-detection, suggests a potential link to acetal resin degradation, potentially driven by related degradation pathways or common precursors. Time-resolved analyses revealed a gradual, monotonic decrease in artifact intensity across consecutive replicates, consistent with passive depletion mechanisms, such as progressive outgassing or desorption from internal surfaces, rather than analytical noise or environmental variability.

Beyond identifying device-derived artifacts, the proposed methodology offers broader applicability in real-world breathomics by providing an automated approach for subtracting ambient air background and hardware-related noise in clinical sampling workflows, thereby enhancing the reliability and interpretability of untargeted VOC profiling. This work contributes to the standardization of pre-analytical quality control procedures and offers a platform-independent strategy for artifact mitigation in exhaled breath analysis across both clinical and environmental contexts.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgment

The authors from Dipartimento Interuniversitario di Fisica di Bari acknowledge financial support from the Project D3 4 Health—Digital Driven Diagnostics, prognostics and therapeutics for sustainable Health care (No. B53C22006120001); and from the project BRIEF—Biorobotics Research and Innovation Engineering Facilities (No. J13C22000400007).

Conflict of interest

Dr Laura Facchini works as a Researcher and Breath Analyst at Predict S.p.A. (Viale Adriatico, c/o Fiera del Levante Pad. 105–70132 Bari, Italy) that develops, produces and sells the Mistral Breath Sampler device that was utilised in this study. All other authors declare no conflict of interest.

Author contributions

Nicoletta Ardito 0009-0005-9275-4129
Data curation (equal), Formal analysis (equal), Investigation (equal), Software (equal), Writing – original draft (lead)

Arianna Elefante
Formal analysis (equal), Methodology (equal), Supervision (equal)

Marilena Giglio
Conceptualization (equal), Data curation (equal), Writing – review & editing (equal)

Andrea Zifarelli 0000-0003-1875-4671
Data curation (equal), Formal analysis (equal), Writing – review & editing (equal)

Laura Facchini
Data curation (equal), Methodology (equal), Validation (equal)

Pietro Patimisco
Formal analysis (equal), Supervision (equal)

Vincenzo Spagnolo 0000-0002-4867-8166
Funding acquisition (equal), Project administration (equal), Supervision (equal)

Nicola Amoroso
Data curation (equal), Formal analysis (equal), Methodology (equal), Writing – review & editing (equal)

Angelo Sampaolo
Conceptualization (equal), Methodology (equal), Supervision (equal), Writing – review & editing (equal)

References

[1] De Lacy Costello B, Amann A, Al-Kateb H, Flynn C, Filipiak W, Khalid T, Osborne D and Ratcliffe N M 2014 A review of the volatiles from the healthy human body *J. Breath Res.* **8** 014001
[2] Chen T, Liu T, Li T, Zhao H and Chen Q 2021 Exhaled breath analysis in disease detection *Clin. Chim. Acta* **515** 61–72
[3] Cope K A, Watson M T, Foster W M, Sehnert S S and Risby T H 2004 Effects of ventilation on the collection of exhaled breath in humans *J. Appl. Physiol.* **96** 1371–9
[4] Lawal O, Ahmed W M, Nijsen T M E, Goodacre R and Fowler S J 2017 Exhaled breath analysis: a review of "breath-taking" methods for off-line analysis *Metabolomics* **13** 110
[5] Westphal K, Dudzik D, Waszczuk-Jankowska M, Graff B, Narkiewicz K and Markuszewski M J 2023 Common strategies and factors affecting off-line breath sampling and volatile organic compounds analysis using thermal desorption-gas chromatography-mass spectrometry (TD-GC-MS) *Metabolites* **13** 8
[6] Bajo-Fernández M, Souza-Silva É A, Barbas C, Rey-Stolle M F and García A 2023 GC-MS-based metabolomics of volatile organic compounds in exhaled breath: applications in health and disease. A review *Front. Mol. Biosci.* **10** 1295955
[7] Palmisani J, Di Gilio A, Cisternino E, Tutino M and de Gennaro G 2020 Volatile organic compound (VOC) emissions from a personal care polymer-based item: simulation of the inhalation exposure scenario indoors under actual conditions of use *Sustainability* **12** 2577122577
[8] Ghimenti S, Lomonaco T, Bellagambi F G, Tabucchi S, Onor M, Trivella M G, Ceccarini A, Fuoco R and Di Francesco F 2015 Comparison of sampling bags for the analysis of volatile organic compounds in breath *J. Breath Res.* **9** 047110
[9] Miekisch W, Kischkel S, Sawacki A, Liebau T, Mieth M and Schubert J K 2008 Impact of sampling procedures on the results of breath analysis *J. Breath Res.* **2** 026007
[10] Beauchamp J, Herbig J, Gutmann R and Hansel A 2008 On the use of Tedlar® bags for breath-gas sampling and analysis *J. Breath Res.* **2** 046001
[11] Owlstone Medical Ltd ReCIVA® breath sampler to collect a breath sample
[12] PREDICT Life Care S.p.A. Mistral inside the breath.
[13] Di Gilio A *et al* 2020 Breath analysis: comparison among methodological approaches for breath sampling *Molecules* **25** 5823
[14] Li M and Wang X R 2019 Peak alignment of gas chromatography–mass spectrometry data with deep learning *J. Chromatogr. A* **1604** 460476
[15] Veltri P 2008 Algorithms and tools for analysis and management of mass spectrometry data *Brief Bioinform.* **9** 144–55
[16] Xu Z, Sun X and de Boves Harrington P 2011 Baseline correction method using an orthogonal basis for gas chromatography/mass spectrometry data *Anal. Chem.* **83** 7464–71
[17] Wahab M F and O'Haver T C 2020 Wavelet transforms in separation science for denoising and peak overlap detection *J. Sep. Sci.* **43** 1998–2010
[18] Barroso-García V, Gutiérrez-Tobal G C, Gozal D, Vaquerizo-Villar F, Álvarez D, Del Campo F, Kheirandish-Gozal L and Hornero R 2021 Wavelet analysis of overnight airflow to detect obstructive sleep apnea in children *Sensors* **21** 1491211491
[19] Nakajima Y *et al* 2021 Complex-valued wavelet spectrum analysis of respiratory conditions and its feasibility

in the detection of low-functional respiration *Healthcare* **9** 9819981
[20] Guo T, Zhang T, Lim E, Lopez-Benitez M, Ma F and Yu L 2022 A review of wavelet analysis and its applications: challenges and opportunities *IEEE Access* **10** 58869–903
[21] Rafiee J, Reif J, Prause N and Schoen M P 2011 Wavelet basis functions in biomedical signal processing *Expert Syst. Appl.* **38** 6190–201
[22] Nason G P and Silverman B W 1995 The stationary wavelet transform and some statistical applications *Wavelets and Statistics* and (*Lecture Notes in Statistics*) vol 103 (Springer) pp 281–99
[23] Merah M, Abdelmalik T A and Larbi B H 2015 R-peaks detection based on stationary wavelet transform *Comput. Methods Programs Biomed.* **121** 149–60
[24] Kumar A, Tomar H, Mehla V K, Komaragiri R and Kumar M 2021 Stationary wavelet transform based ECG signal denoising method *ISA Trans.* **114** 251–62
[25] The SciPy community Signal processing (scipy.signal)—SciPy v1.15.3 Manual
[26] Daubechies I 1992 *Ten Lectures on Wavelets* (*CBMS-NSF Regional Conference Series in Applied Mathematic*) vol 61 (Society for Industrial and Applied Mathematics (SIAM)) (https://doi.org/10.1137/1.9781611970104)
[27] Percival D B and Walden A T 2000 Wavelet methods for time series analysis wavelet methods for time seriesanalysis (https://doi.org/10.1017/CBO9780511841040)
[28] Sharma J K, Gopal Y, Birla D and Lalwani M 2018 An algorithm for selecting compatible wavelet function in electrical signals to detect and localize disturbances
[29] Agilent Technologies Agilent MassHunter workstation-unknowns analysis familiarization guide
[30] Agilent Technologies Agilent technologies agilent MassHunter workstation software qualitative analysis familiarization guide for GC/MS
[31] Percival D B and Mofjeld H O 1997 Analysis of subtidal coastal sea level fluctuations using wavelets *J. Am. Stat. Assoc.* **92** 868–80
[32] Kim K-H, Pandey S K, Kim Y-H, Sohn J R and Oh J-M 2015 Emissions of amides (N,N-dimethylformamide and formamide) and other obnoxious volatile organic compounds from different mattress textile products *Ecotoxicol. Environ. Saf.* **114** 350–6
[33] Lüftl S, Archodoulaki V-M and Seidler S 2006 Thermal-oxidative induced degradation behaviour of polyoxymethylene (POM) copolymer detected by TGA/MS *Polym. Degrad. Stab.* **91** 464–71
[34] Haubs M, Kreuznach B, Mann E and Lingnau J 2009 (54) process for the preparation of oxymethylene polymers and apparatus suitable for this purpose (76) inventors
[35] Epa U 1999 Compendium of methods for the determination of toxic organic compounds in ambient air second edition compendium method TO-17 determination of volatile organic compounds in ambient air using active sampling onto sorbent tubes