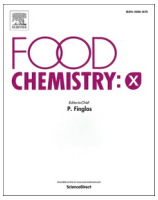
Food Chemistry: X 18 (2023) 100696

Contents lists available at ScienceDirect

Food Chemistry: X

journal homepage: www.sciencedirect.com/journal/food-chemistry-x



Rapid identification of adulteration in raw bovine milk with soymilk by electronic nose and headspace-gas chromatography ion-mobility spectrometry

Huaixiang Tian, Juanjuan Xiong, Shuang Chen, Haiyan Yu, Chen Chen, Juan Huang, Haibin Yuan, Xinman Lou *

School of Perfume and Aroma Technology, Shanghai Institute of Technology, Shanghai 201418, PR China

ARTICLE INFO

Keywords:
Raw bovine milk adulteration
Soymilk
Volatile compounds
Electronic nose
Headspace-gas chromatography ion-mobility spectrometry

ABSTRACT

The adulteration of soymilk (SM) into raw bovine milk (RM) to gain profit without declaration could cause a health risk. In this study, electronic nose (E-nose) and headspace-gas chromatography ion-mobility spectrometry (HS-GC-IMS) were applied to establish a rapid and effective method to identify adulteration in RM with SM. The obtained data from HS-GC-IMS and E-nose can distinguish the adulterated samples with SM by principal component analysis. Furthermore, a quantitative model of partial least squares was established. The detection limits of E-nose and HS-GC-IMS quantitative models were 1.53% and 1.43%, the root mean square errors of prediction were 0.7390 and 0.5621, the determination coefficients of prediction were 0.9940 and 0.9958, and the relative percentage difference were 10.02 and 13.27, respectively, indicating quantitative regression and good prediction performances of SM adulteration levels in RM were achieved. This research can provide scientific information on the rapid, non-destructive and effective adulteration detection for RM.

1. Introduction

Protein is the main nutritional component of raw bovine milk (RM), and its content determines the value of milk. However, unscrupulous businesses intentionally adulterate milk by adding non-dairy proteins in order to gain economic benefits which reduces RM quality and be hazardous to the interests of consumers. With the diversification of dietary structure and the increase of vegetarian groups, plant-based milk is rapidly emerging domestically and even globally. Plant-based protein has been regarded as a potential substitute for protein in dairy industry (FARAH et al., 2021; JARIYASOPIT et al., 2021). Among which, soymilk (SM) is a common plant-based milk which is much cheaper than RM, and importantly, not easy to be detected (FENG, ZHU, CHEN, BAO, & HE, 2019; UNCU & UNCU, 2020). A study showed that cheese made with 25% (v/v) SM (4.15%) in RM has a higher protein content than cheese made with whole RM (2.6%) (ALI, ABDULLA, MALIK, MAHMOUD & HAMADNALLA, 2020a). However, the protein of SM is statutorily defined as an allergen and its presence in unlabeled food has a potential risk, such as skin pruritus, rash, chest tightness dyspnea and even shock, which also triggers a trust crisis from consumers (JARIYASOPIT et al., 2021). According to the International Union of Immunological Societies, Gly m 1 to Gly m 8 has been verified as allergen proteins in soybean (KEREZSI, JACQUET & BLECKER, 2022). In Japan, it was found that infants and children were more susceptible to allergen proteins in soybean (Wang et al., 2020a).

Until now, various technologies have been used for detecting adulteration of protein from plants or animal milk, including gas chromatography-mass spectrometry (GC–MS), nuclear magnetic resonance (NMR) (LI et al., 2017), near-infrared spectroscopy (NIR) (DOS SANTOS PEREIRA, DE SOUSA FERNANDES, DE ARAUJO, DINIZ & MACIEL, 2020), fourier-transform infrared spectroscopy (FT-IR) (SEN et al., 2021), and liquid chromatography-high resolution mass spectrometry (LC-MS/MS) (JI et al., 2023; FAN et al., 2023) (Table S1). All these methods are sensitive, reliable, and useful for detecting specific exogenous proteins in dairy products, but most of them are analyzed based on metabolic profiling or different absorption of functional groups in protein/fat/carbohydrates, which are complicated and time-consuming for sample preparations, and the applied instrumentations are inconvenient and expensive. Thus, the development of rapid, sensitive and convenient methods to effectively identified adulteration in

* Corresponding author.
E-mail address: lxinman2@163.com (X. Lou).

https://doi.org/10.1016/j.fochx.2023.100696
Received 27 January 2023; Received in revised form 28 April 2023; Accepted 28 April 2023
Available online 29 April 2023

RM with SM is necessary. It was reported that protein, fat and solids-not-fat were susceptible to compositional differences in milk which can be easily detected by milk composition analyzer (ZHU, WEN & WANG, 2020). Our previous study (TIAN et al., 2022) revealed that milk composition analyzer could rapidly and accurately identify acid neutralizers adulterated in RM. Therefore, milk composition analyzer may have a potential ability to identify adulteration in RM with SM.

Volatile compounds (VOCs) are an important indicator of RM and SM which is also one of the factors that affect consumers to choose dairy products (UTZ et al., 2021). However, the beany odour of SM is a major manifestation caused by the enzymatic oxidation of lipids and other bioactive substances induced by lipoxygenase resulting in differences of VOCs compared with RM, especially hexanal hexanol and 2-pentylfuran (Wang et al., 2020a). Headspace gas chromatography-ion mobility spectrometry (HS-GC-IMS) is an emerging approach with no extensive complex and time-consuming of pretreatment, and its high sensitivity has been proved to be efficient in identifying the adulteration acts to control the quality in food industry (GU, CHEN, WANG, WANG & HUO, 2020). As previous studies reported (X. WANG, ROGERS, LI, YANG, CHEN & ZHOU, 2019), HS-GC-IMS combined with principal component analysis (PCA), partial least-squares discriminant analysis and soft independent modeling of class analogy can accurately identify adulteration of honey according to target the differences in VOCs. A similar study found that adulteration of sesame oil could be correctly classified and identified based on data provided by HS-GC-IMS with random forests classification model and recursive support vector machine method (ZHANG et al., 2016). Electronic nose (E-nose) is a bionic olfactory technology with an internal array sensor which can capture information about the VOCs to analyze differences in odour (ALI, HASHIM, ABD AZIZ & LASEKAN, 2020b). E-nose coupled with discriminant function analysis, the classification accuracy of identifying soybean oil adulteration levels in cow ghee could achieve more than 99% (ROY et al., 2022). Moreover, GU et al. (2020) investigated the performances of HS-GC-IMS and E-nose, respectively, combined with k-nearest neighbour for rapid detection of moldy rice grains from healthy ones and the correct classification rates reached 94.44% and 91.67%, respectively. A feasibility study reported the application of GC-IMS and E-nose for the detection of extra virgin olive oil adulterated with soft-refined olive oils (DAMIANI, CAVANNA, SERANI, DALL'ASTA & SUMAN, 2020). The available methodologies can distinguish kinds of adulterants in milk samples while most of them need to be improved in quantifying RM adulteration levels. Moreover, the research on identifying adulteration in RM with SM detected by HS-GC-IMS and E-nose with chemometric methods has been not reported, which needs to be further evaluated.

The objective of this work was to establish an efficient, accurate and sensitive method to rapidly identified the adulteration levels in RM with SM. First, the milk composition analyzer was used to preliminarily detect the contents of fat, protein and non-fat solids in RM and adulterated samples. Then, the VOCs were identified by HS-GC-IMS and E-nose, and the differences of VOCs among samples were showed through volatile fingerprints. PCA and partial least squares (PLS) regression were applied to classify adulterated samples and predict the concentration of SM in RM. The 3 times and 10 times the standard deviation of Y-residual/slope standard deviation were limit of detection (LOD) and limit of quantitation (LOQ), respectively, according to the previous method (Gu et al., 2020). Value of root means square error of calibration/root mean square error of prediction (RMSEC/RMSEP), R_c^2/R_p^2 (decision coefficient of calibration set/ decision coefficient of prediction set) and correlation coefficient (R^2) and ratio of prediction to deviation (RPD) were used to evaluate performances of models. This study is helpful to understand flavour variations in RM after adulterating with SM and provides a reference for the rapid identification of plant-based milk adulteration in RM.

2. Material and methods

2.1. Sample and sample preparation

RM samples (0–4 ℃) were collected from Shanghai No. 4 Dairy Product Factory and kept in a refrigerator at 4 ± 2 ℃ and analyzed within 24 h. Soybeans were purchased from a local market (Shanghai, China), and SM was prepared according to a previous study (SHI et al., 2019). Soybeans were soaked in water for 12 h at a ratio of 1:3 (w/v) and then ground in water at a ratio of 1:7 (w/v) using a juicer equipped with an automatic centrifugal filter to separate the okara (SM-8000(GT), Hurom Co. Ltd., Jilin, China). Subsequently, the obtained SM was boiled to 100 ℃ for 10 min, then cooled to room temperature.

The RM samples were adulterated with SM at the concentrations of 0.0%, 0.5%, 1.0%, 5.0%, 10.0%, 15.0%, and 20.0% (w/w) to prepare the adulteration samples and all samples were mixed well before use. Three independent samples were prepared for each concentration, in a total of 21 test samples. No technical replicates per treatment were used.

2.2. Milk composition analyzer data acquisition

A Youchuang UL80BC milk composition analyzer (Hangzhou, CN) was used to obtain the content of fat, protein and solids-not-fat, which is susceptible to the addition of SM in RM (GAUTAM, SHARMA, LATA, RAJPUT & MANN, 2017). A 30 mL sample was added in a cup, which was placed below the injection port, then the instrument can start automatic detection after pressing the confirmation key. The data can be obtained after the end of test, waiting for 1–2 min. The calibration was needed before each sample detection.

2.3. HS-GC-IMS measurement

Volatile fingerprints of milk samples were determined by an IMS system (Flavorspec®, Gesellschaft für Analytische Sensorsysteme mbH, Dortmund, Germany), equipped with an Agilent 6890 N gas chromatograph (Palo Alto, USA) using a non-polar MXT-5 GC column (15 m × 0.53 mm × 0.1 μm, Restek, PA, USA).

Milk samples (2.0 g) were transferred into a 20 mL headspace sample vial and subsequently incubated at 60 °C for 10 min at 500 rpm. After incubation, a headspace volume of 100 μL was automatically injected utilizing a heated syringe (85 °C). The column was held at 60 °C, and the carrier gas and drift gas was nitrogen (99.999% purity). The carrier gas passed through the HS-GC-IMS injector transferring the sample into the GC column as follows: 2 mL/min for 2 min, then raised to 100 mL/min over 18 min and held at 100 mL/min until 30 min. The flow rate of drift gas is 150 mL/min. Each analysis was conducted in triplicate 3.

The GC-IMS data were analyzed by using LAV software (version 2.2.1, Gesellschaft fürAnalytische Sensorsysteme mbH, Dortmund, Germany). VOCs were qualitatively analyzed by comparing the retention indexes and drift times with those in the GC-IMS library.

2.4. E-nose measurement

For the fast gas chromatography-based E-nose analysis which was equipped with two different polarity columns working in a parallel mode, a 5.0 g of milk sample was accurately weighted and immediately placed in a 20 mL headspace sample vial. The vial was closed with a leak-proof cap, covered with a silicon/Teflon septum, then incubated at 50 °C for 20 min. The autosampler injected 5 mL of sample from the headspace into the GC at a rate of 125 μL/s and collected the analytes in a Tenax trap at 40 ℃ for 50 s. After rapid heating, the analytes were separated and transferred to two parallel short GC columns (Restek, Centre County, PA, U.S.): a non-polar column (MXT-5: 5% diphenyl, 95% methylpolysiloxane, 10 m × 0.180 mm × 0.4 μm film thickness) and a slightly polar column (MXT-1701: 14% cyanopropyl- phenyl, 86% methylpolysiloxane, 10 m × 0.180 mm × 0.4 μm film thickness).

Hydrogen (99.999% purity) was used as the carrier gas. The system was operated at a constant pressure of 80 kPa with a 10 mL/min split flow at the column head. The temperature conditions were as follows: the sampler temperature was 200 ℃; the temperature program included an isothermal step at the beginning of 80 ℃, a slope of 3 ℃/s to 250 ℃, and an isothermal step of 250 ℃ for 21 s; and flame-ionization detection (FID1/FID2) at 260 ℃. The samples were analyzed in triplicate to obtain a better parallel effect.

2.5. Data analysis

Unsupervised PCA and supervised PLS were employed to detect the latent information from the massive data, build the discriminant analysis model, and identify the tentative markers. PCA and PLS analyses with SIMCA-P software (Version 14.1, Umetrics AB, Umeå, Sweden), were used to cluster the RM sample and adulterated samples. One-way analysis of variance (ANOVA) was introduced to determine the significant difference in the concentrations of SM in RM with $p < 0.05$ using SPSS 22.0 software (SPSS Inc,Chicago, IL, USA).

3. Results and discussions

3.1. Identification of adulteration in raw bovine milk with soymilk by milk composition analyzer

The contents of fat, protein and non-fat solids in RM were determined by milk composition analyzer and the results are shown in Table 1. There were significant differences in the indicators of fat and protein content between adulterated and non-adulterated samples ($p < 0.05$). When the concentration of SM was up to 20% in RM, the protein content significantly decreased from 3.38 g/100 g to 3.02 g/100 g. Moreover, RM (8.84 g/100 g), SM (5.99 g/100 g) and adulterated RM samples with 15% − 20% SM (8.27–8.43 g/100 g) were significantly different in solids-not-fat ($p < 0.05$). As the concentration of SM increased in RM, fat, protein, and solids-not-fat content in adulterated samples showed a decreasing trend, indicating that milk composition analyzer could identify these differences between adulterated and non-adulterated samples. TIAN et al. (2022) reported that milk composition analyzer could judge whether RM was adulterated by neutralizing acid adulterants (such as sodium hydroxide, sodium thiocyanate and sodium carbonate) but the types cannot be identified. However, it can be seen in the present study that the content of fat, protein and solids-not-fat in all adulterated samples at six levels (0.5% − 20%) meet the requirements of GB 19301–2010 (Food Safety Standard for Raw Milk in China: the content of fat (g/100 g) ≥ 3.10; the content of protein (g/100 g) ≥ 2.80; the content of solids-not-fat (g/100 g) ≥ 8.10) which means milk composition analyzer failed to be used as a tool for the rapid detection of adulteration in RM with SM at the level of 0.5% − 20% based on the contents of fat, protein, and solids-not-fat in milk. Therefore, E-nose and HS-GC-IMS need to be further applied to identify the differences based on VOCs compounds between adulterated and non-adulterated samples.

Table 1
Results of raw bovine milk and the milk adulterated with soymilk samples using milk composition analysis.

Sample	Fat (g/100 g)	Protein (g/100 g)	Solids-not-fat (g/100 g)
Raw bovine milk	3.43 ± 0.15^{a}	3.33 ± 0.08^{a}	8.84 ± 0.20^{a}
0.5% soymilk	3.38 ± 0.04^{cd}	3.27 ± 0.03^{ab}	8.68 ± 0.08^{ab}
1% soymilk	3.34 ± 0.05^{b}	3.23 ± 0.03^{bc}	8.57 ± 0.07^{abc}
5% soymilk	3.33 ± 0.03^{b}	3.23 ± 0.03^{bc}	8.58 ± 0.08^{abc}
10% soymilk	3.29 ± 0.03^{bcd}	3.22 ± 0.02^{bc}	8.54 ± 0.04^{abc}
15% soymilk	3.27 ± 0.07^{d}	3.17 ± 0.02^{cd}	8.43 ± 0.05^{bc}
20% soymilk	3.16 ± 0.03^{c}	3.02 ± 0.01^{d}	8.27 ± 0.02^{c}
soymilk	1.51 ± 0.08^{e}	2.28 ± 0.09^{e}	5.99 ± 0.11^{d}
Standards in GB for bovine raw milk	≥ 3.10	≥ 2.80	≥ 8.10
Standards in GB for soymilk	≥ 0.80	≥ 2.00	≥ 4.00

GB represents Food Safety Standard for Raw Milk in China (GB 19301–2010); Any results with different superscripts in the same column are statistically significant ($p < 0.05$). The RM samples were adulterated with SM at the concentrations of 0.0%, 0.5%, 1.0%, 5.0%, 10.0%, 15.0%, and 20.0% (w/w), namely 0.5 %SM, 1.0 %SM, 5.0 %SM, 10.0 %SM, 15.0 %SM and 20.0 %SM.

3.2. Identification of adulteration in raw bovine milk with soymilk by E-nose

Gas chromatography-based E-nose is a new type of electronic analysis instrument which can accurately identify and analyze VOCs with the advantages of fast, simple, and good repeatability (M. ALI et al., 2020b). The response radar maps of two columns (MXT-5 and MXT-1701) towards RM and six adulterated samples were shown in the volatile fingerprint established based on E-nose. As shown in Fig. 1, response signals of polar and nonpolar columns both indicated that the overall spectra profile of non-adulterated and adulterated samples was similar, while the heights of multiple peaks were enhanced with the increase of SM concentration in RM. The similarity of the overall aroma profile indicated that there was little difference in VOCs between RM and adulterated samples. But the difference of peak heights revealed that the intensity of some VOCs in adulterated samples was higher than that of pure RM samples, especially the concentration of SM in RM was greater than 10 %. However, it was difficult to classify the concentrations of SM in RM sample only by observing the sensor signals of E-nose, thus the signals at the main peak position of 0–70 s were extracted as feature data for subsequent chemometrics analysis of PCA and PLS.

3.3. Analysis of volatile fingerprints of raw bovine milk and soymilk by HS-GC-IMS

To determine the difference of VOCs in E-nose fingerprints between RM sample and adulterated samples with different concentration of SM, the volatile fingerprints based on HS-GC-IMS was further investigated, and VOCs in different samples showed different peak position and intensity. Fig. 2A is the three-dimensional (3D) imaging consisting of ion drift time (X - axes), the retention time of gas chromatographic peaks (Y - axes), and the intensity of ion signals (Z - axes). Individual points in the 3D topographic plot indicated the VOCs detected from the sample, and it could be visualized that the VOCs in SM were more than those in RM. To observe conveniently, two-dimensional (2D) array full-size top view was used to compare the difference of VOCs in SM and RM (Fig. 2B), which was obtained by normalizing the ion drift time and reactive ion peak (RIP) position. Using the RM sample as a reference, the SM sample was subtracted by the reference and each dot represented a VOC whose signal of intensity was represented through the colour. If the concentrations of VOCs were consistent, the background would appear white after deduction, red indicated that the VOCs concentration in SM sample was higher than that in the reference, while blue meant lower. The results showed that most VOCs ions appeared at retention times of 100–500 s and normalized drift times of 1.0–1.7 ms.

Qualitative descriptions of the VOCs profiles in RM and SM are presented below by comparing their retention time and the ion drift time from their HS-GC-IMS spectra. A total of 46 VOCs was identified from the topographic maps via GC-IMS Library searches. However, many VOCs could generate several signals, which represented monomers (M), dimers (D), and trimer (T) in volatile fingerprints. The characteristic volatile fingerprints for two samples were established in Fig. 3A, where

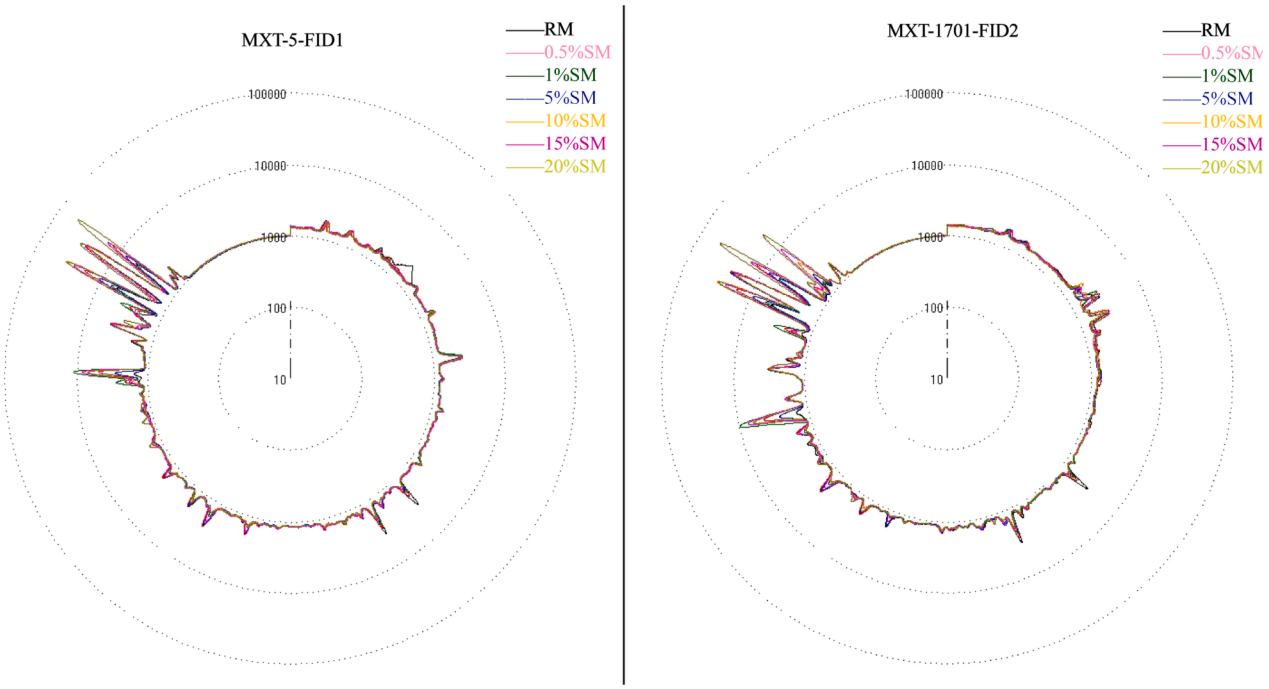


Fig. 1. E-nose fingerprints of raw bovine milk and the milk adulterated with soymilk samples at different concentrations. The RM samples were adulterated with SM at the concentrations of 0.0%, 0.5%, 1.0%, 5.0%, 10.0%, 15.0%, and 20.0% (w/w), namely 0.5 %SM, 1.0 %SM, 5.0 %SM, 10.0 %SM, 15.0 %SM and 20.0 %SM.

row represented the detected compound, and the column represented the content of the same VOCs in different samples. The brighter signal for each compound indicated a higher concentration, and unidentified substances were represented by numbers.

As shown in Fig. 3A and Table 2 9 ketones, 9 aldehydes, 4 alcohols, 3 esters, 3 terpenes, 1 furan, 1 pyrazine, were identified based on the retention time index and normalized drift time of standard substances in the database with a tolerable deviation < 5%. Eight substances were detected in both RM and SM samples, including acetophenone monomer, acetophenone dimer, benzaldehyde monomer, benzaldehyde dimer, furaneol monomer, furaneol dimer, heptanal monomer and heptanal dimer, corresponding to the region A of fingerprint (Fig. 3A). Diacetyl, 3-octanone, 2-hexanone monomer, 2-butanone monomer and 2-butanone dimer were only present in RM sample, meaning those substances were the characteristic VOCs of RM which were marked as area B in Fig. 3A. For SM sample, there were 33 substances detected, which were characteristic VOCs of SM, corresponding to region C of the fingerprint. Interestingly, trace amounts of acetophenone and 2-hexanone in present study were detected, which were not detected in a similar study by simultaneous distillation extraction -GC–MS method (SACCHI, MARRAZZO, MASUCCI, DI FRANCIA, SERRAPICA & GENOVESE, 2020). In addition, acetophenone, styrene, α-pinene, 2,3-pentanedione, furaneol and limonene trimers in SM were not detected in previous studies, which might be due to low content that simultaneous distillation extraction/ solid phase microextraction -GC–MS failed to detect (SACCHI et al., 2020). Therefore, it can be proposed that HS-GC-IMS was sensitive enough for the detection of VOCs, which could be a reliable technology to identify adulteration in RM with different concentration.

3.4. Identification of adulteration in raw bovine milk with soymilk by HS-GC-IMS

HS-GC-IMS was used to analyze the VOCs from RM adulterated with different concentrations of SM. As shown in Fig. 3C, the number of highlights in 2D spectral highlights were constantly increasing with the increase of SM concentration and the dots became brighter. A significant difference was observed in VOCs of adulterated samples when the SM concentration in RM reached 10% compared to pure RM sample. Moreover, in order to distinguish the VOCs between low concentration adulterated samples and raw milk, the differential plots were obtained by topographic plot deduction of RM plot where the difference even between the RM sample adulterated with 0.5% SM and RM could be reflected in Fig. 3D.

As Fig. 3B showed, the fingerprint presented the difference of VOCs between pure RM and RM adulterated samples with SM at different concentrations (0.5, 1.0, 5.0, 10.0, 15.0, 20.0%). Each row represented all the signal peaks detected in RM and adulterated samples, and each column represented the signal peaks of the same VOCs in different samples. Fig. 3B demonstrated that the VOCs in area A did not change significantly with the increase of SM concentration in RM. When the SM concentration in RM was 15%, the corresponding VOCs concentration in area B significantly decreased compared to the SM concentration<10%. The content of VOCs in region C was lower when SM concentration in RM was under 5%, while the peak of VOCs was significantly enhanced when SM concentration in RM was up to 15%. It was indicated that only when the concentration of SM in RM was more than 10%, the volatile fingerprints could be clearly distinguished among them (Fig. 3B). Therefore, data mining technology was needed to find more information and to obtain a more accurate identification of adulteration in RM with different concentrations of SM.

3.5. Discrimination of adulteration levels by principal component analysis

It was difficult to classify the differences in adulterated samples at low-concentration just by observing the E-nose signals and HS-GC-IMS spectra. PCA could provide an overview that revealed trends and groups of data, which was used to reduce the dimension of multidimensional data matrix, preserving most of the variation in the original set of variables (X. WANG et al., 2019). Therefore, PCA was applied to give more information about the differences of VOCs in RM with SM at different adulterated levels. Before the PCA analysis, the data from E-nose and HS-GC-IMS were cleaned, normalized, and scaled to minimize background interference. Normalization can eliminate errors from

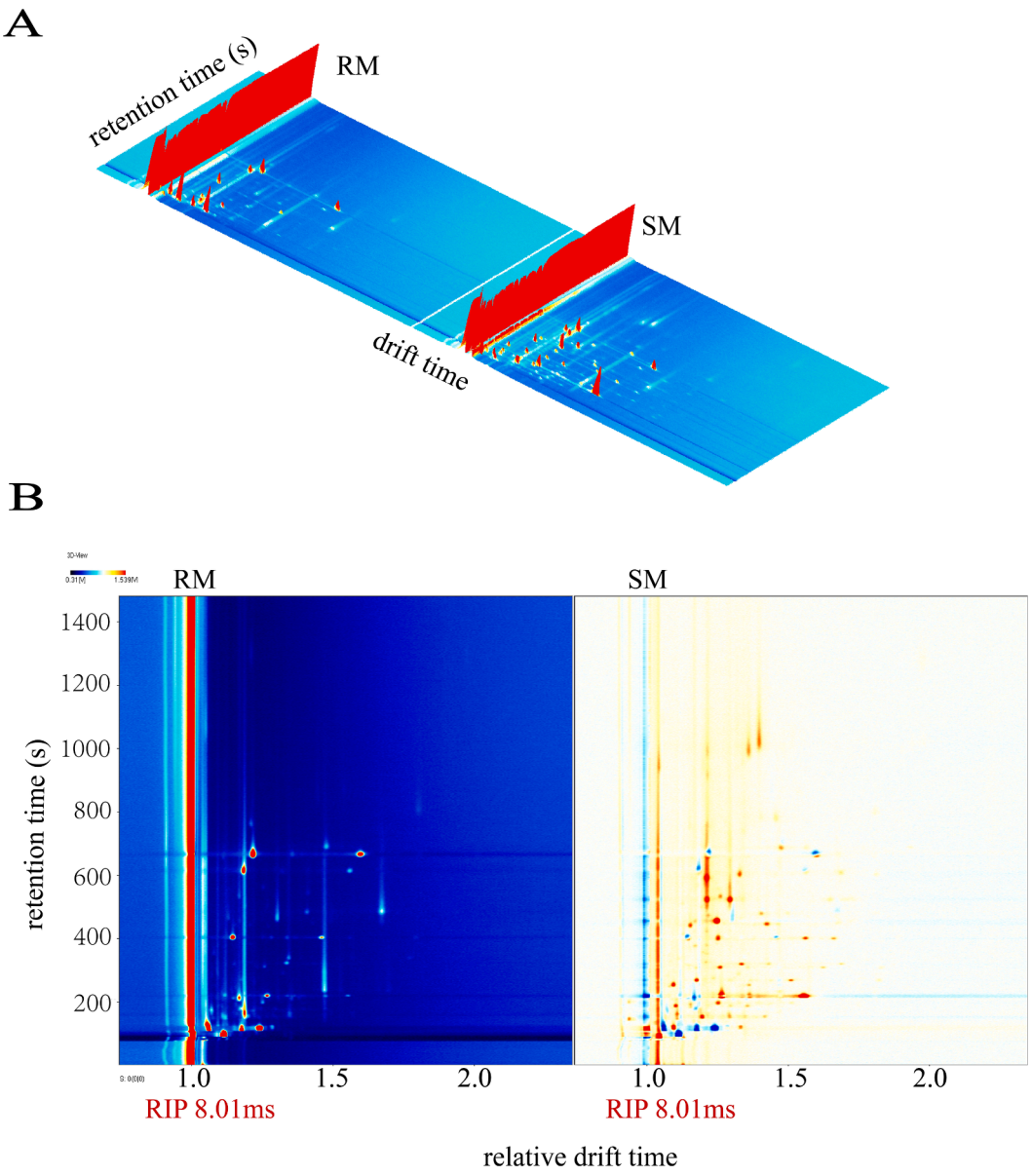


Fig. 2. 3D (A) and 2D (B) topographic plots of raw bovine milk and soymilk.

different sources, such as operation errors, instrument variations, and some uncontrollable biological variations. Scaling was used to convert concentration of different substances into a scaling coefficient. Three major scaling methods were compared: including unit variance (UV) scaling, mean centralization (Ctr) scaling and pareto (Par) scaling. The interpretation, predictive ability and separation of data treated by Par were superior to those of UV and Ctr scaling. Thus, the data were analyzed using the Par scaling before the cluster analysis. For the data of E-nose, five main components were extracted after PCA analysis from the 21 samples × 70 response values data matrix, with the value of R^2 = 96.1% and Q^2 = 0.8940 (cumulative prediction rate of the model). The contribution rates of the first principal component (PC1) and the second principal component (PC2) were 70.3% and 13.1%, respectively, and the total contribution rate of the first two PCs was 83.1%. It can be seen that the first two PCs contained most of the variable information, so the first two PCs were selected for data visualization. The analysis results of 2D scatter plots with PC1 as the × - axis, and PC2 as the y – axis in Fig. 4A. The score chart showed that the clear trend of separation trend between RM sample and adulterated samples with SM at different concentrations was obvious, even for samples with low adulteration concentrations such as 0.5% and 1%, whose score tended to right with the increase of adulteration concentration.

With regards to HS-GC-IMS fingerprint spectrum, 21 samples × 46 selected compounds data matrix was used for PCA analysis, and five main components were extracted with R^2 of 95.0% and Q^2 of 0.816. As shown in Fig. 4B, the first two principal components explained 70.9% and 12.0%, respectively, and the total contribution rate of the first two PCs were up to 82.9%. The score plot showed a general separation trend between RM sample and adulterated samples, accompanied with slight overlapped, particularly for the distance between the RM adulterated with 5% and 1%. There was an obvious distance between the RM adulterated sample with 0.5% SM and the pure RM sample, and the higher adulteration concentration was, the more inclined the score to the left side. Thus, RM sample adulterated with different concentrations could be well discriminated based on the E-nose and HS-GC-IMS data using the PCA model. The results also showed that PCA was a powerful and ideal general dimension reduction method when dealing with more than two different data. H. LI et al. (2022) applied HS-GC-IMS and PCA model for assessing milk quality through screening potential VOCs (such as 2-butanone and diacetyl) which were associated with deterioration.

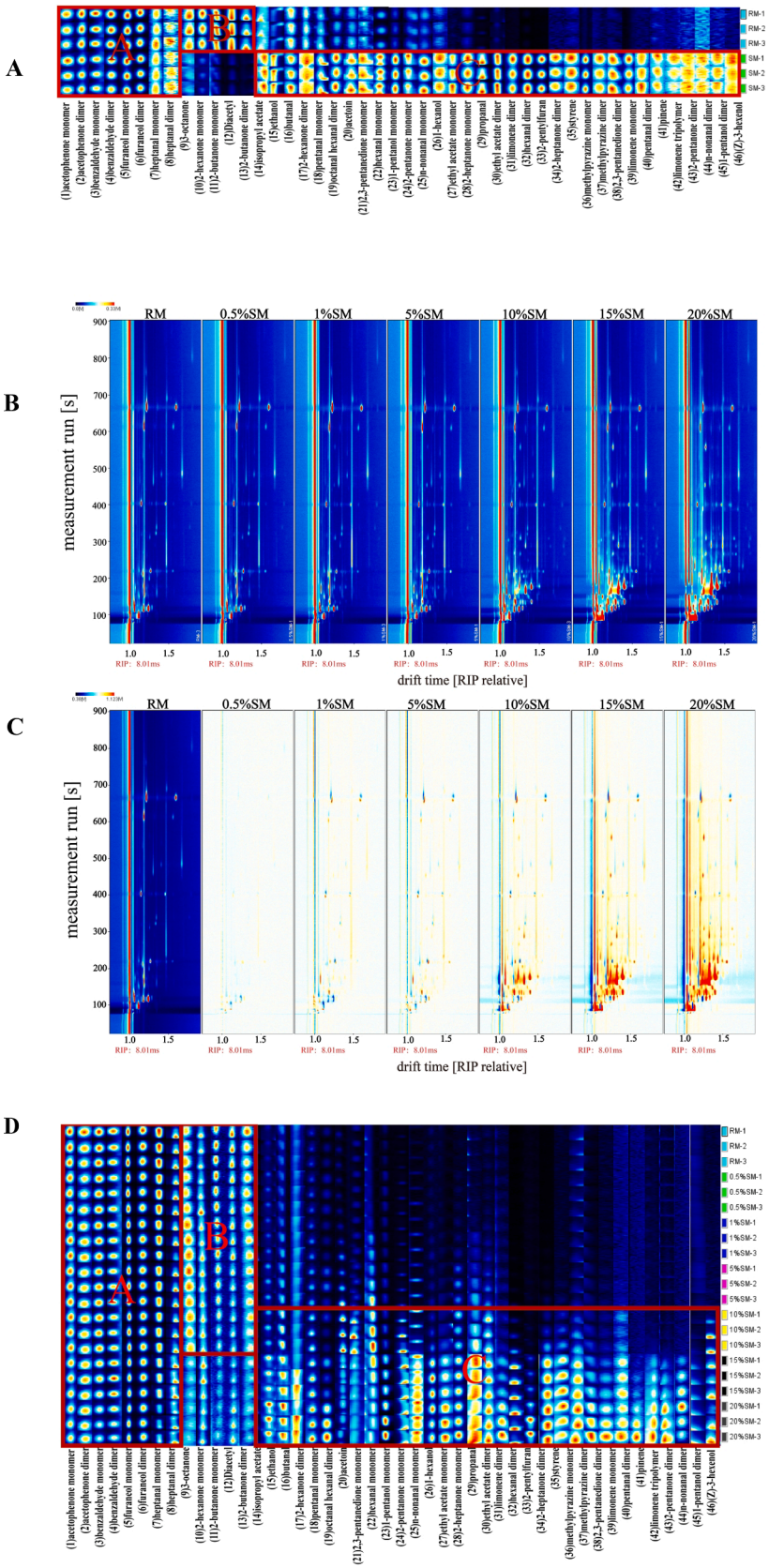


Fig. 3. The fingerprint of volatile profiles in raw bovine milk and soymilk (A) and in raw bovine milk and the milk adulterated with soymilk samples of different concentrations (B), 2D-topographic plots (C) and subtraction topographic plots (D) of volatile substances in raw bovine milk and the milk adulterated with soymilk samples of different concentrations (0.5%, 1.0%, 5.0%, 10.0%, 15.0%, and 20.0%, w/w).

Table 2
Qualitative detection of volatile compounds in raw bovine milk and soymilk using GC–IMS.

Number	Compound	CAS#	Formula	MW	RI	Rt [s]	Dt [RIPrel]
1	Diacetyl	431–03-8	C4H6O2	86.1	603.0	120.849	1.1739
2	2-Butanone(D)	78–93-3	C4H8O	72.1	603.2	120.924	1.2382
3	2-Butanone(M)	78–93-3	C4H8O	72.1	599.6	119.823	1.0592
4	3-Octanone	106–68-3	C8H16O	128.2	996.7	469.600	1.3016
5	Acetophenone(M)	98–86-2	C8H8O	120.2	1064.7	611.373	1.5634
6	Acetophenone(D)	98–86-2	C8H8O	120.2	1066.5	615.383	1.1847
7	Furaneol(M)	3658–77-3	C6H8O3	128.1	1090.1	669.401	1.2212
8	Furaneol(D)	3658–77-3	C6H8O3	128.1	1088.4	665.512	1.5934
9	2-Hexanone(M)	591–78-6	C6H12O	100.2	781.4	212.717	1.1732
10	2-Hexanone(D)	591–78-6	C6H12O	100.2	790.1	219.742	1.4963
11	Acetoin	513–86-0	C4H8O2	88.1	708.8	162.716	1.3490
12	Propanal	123–38-6	C3H6O	58.1	509.4	92.572	1.0486
13	Butanal	123–72-8	C4H8O	72.1	590.8	117.169	1.2924
14	Benzaldehyde(M)	100–52-7	C7H6O	106.1	957.6	404.092	1.1472
15	Benzaldehyde(D)	100–52-7	C7H6O	106.1	955.7	401.095	1.4663
16	Octanal	124–13-0	C8H16O	128.2	1005.1	485.428	1.4079
17	Heptanal	111–71-7	C7H14O	114.2	909.9	338.855	1.3423
18	Heptanal	111–71-7	C7H14O	114.2	895.6	321.750	1.6870
19	Ethanol	C64175	C2H6O	46.1	452.9	75.520	1.0563
20	1-Hexanol	111–27-3	C6H14O	102.2	866.2	289.637	1.3240
21	Ethyl acetate(M)	141–78-6	C4H8O2	88.1	614.5	124.414	1.0944
22	Ethyl acetate(D)	141–78-6	C4H8O2	88.1	615	124.566	1.3394
23	2-Methylpyrazine(M)	109–08-0	C5H6N2	94.1	813.4	239.459	1.1180
24	2-Methylpyrazine(D)	109–08-0	C5H6N2	94.1	803.9	231.223	1.3917
25	1-Pentanol Monomer	71–41-0	C5H12O	88.1	762.6	198.100	1.2488
26	1-Pentanol Dimer	71–41-0	C5H12O	88.1	760.4	196.499	1.5114
27	2-Pentanone (M)	107–87-9	C5H10O	86.1	683.5	149.898	1.3662
28	2-Pentanone(D)	107–87-9	C5H10O	86.1	679.8	148.200	1.1230
29	2-Heptanone(M)	110–43-0	C7H14O	114.2	886.6	311.615	1.2642
30	2-Heptanone(D)	110–43-0	C7H14O	114.2	886.1	311.002	1.6237
31	Pentanal(M)	110–62-3	C5H10O	86.1	724.1	171.700	1.1889
32	Pentanal(D)	110–62-3	C5H10O	86.1	694.2	155.014	1.4199
33	*n*-Nonanal	124–19-6	C9H18O	142.2	1099.4	691.015	1.4740
34	*n*-Nonanal	124–19-6	C9H18O	142.2	1099.0	690.080	1.9359
35	Hexanal(M))	66–25-1	C6H12O	100.2	790.9	220.349	1.269
36	Hexanal(D)	66–25-1	C6H12O	100.2	790.4	219.959	1.5541
37	Limonene (M)	138–86-3	C10H16	136.2	1025.1	525.229	1.2134
38	Limonene (D)	138–86-3	C10H16	136.2	1024.3	523.622	1.2900
39	Limonene (T)	138–86-3	C10H16	136.2	1025.4	525.812	1.6418
40	Styrene	100–42-5	C8H8	104.2	869.4	292.998	1.4051
41	Pinene	80–56-8	C10H16	136.2	949.0	391.279	1.6751
42	2–3-Pentanedione	600–14-6	C5H8O2	100.1	714.1	165.750	1.2320
43	2,3-Pentanedione	600–14-6	C5H8O2	100.1	716.9	167.328	1.3220
44	(Z)-3-Hexenol	928–96-1	C6H12O	100.2	852.3	275.555	1.2344
45	isoPropyl acetate	108–21-4	C5H10O2	102.1	654.1	137.756	1.1622
46	2-Pentylfuran	3777–69-3	C9H14O	138.2	989.3	456.240	1.2481

Mw represents molecular mass, RI represents relative retention index on MXT-5, Rt represents retention time of gas chromatography, Dt represents relative drift time of ion mobility spectrometry, M represents monomer, D represents dimer, T represents trimer.

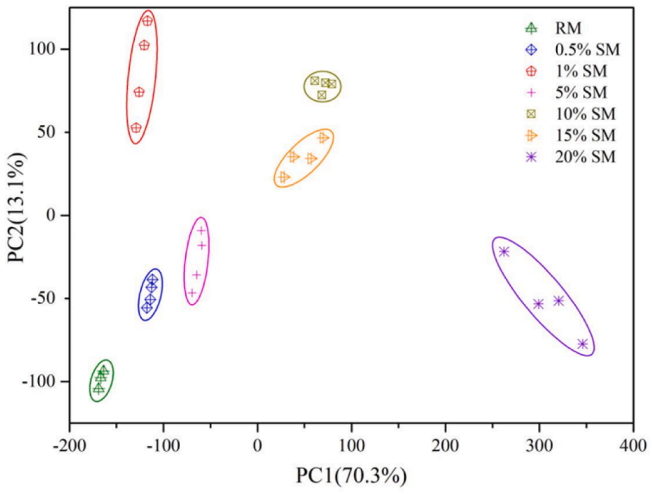
Another study revealed that GC–MS in combination with PCA could identify SM adulteration with RM as low as (1–2)% by determining fatty acids (JARIYASOPIT et al., 2021). However, PAC failed to predict the concentration of SM adulteration in RM samples and the establishment of quantitative model needs further help of more powerful technologies.

3.6. Quantification of adulteration in raw bovine milk with soymilk by HS-GC-IMS vs E-nose data

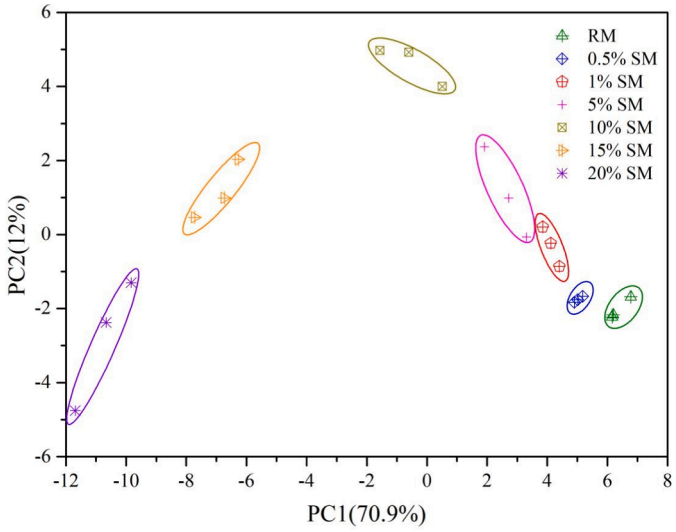
In this step, the PLS model was applied to predict the concentration of SM adulteration in RM samples based on the data of E-nose and HS-GC-IMS. Parameters for PLS models were shown in Table 3. Small value of RMSEC/RMSEP and higher value of R_c^2/R_p^2 indicated a better fit for the PLS model to the experimental data. In E-nose models, the R_C^2 and RMSEC were 0.9952 and 0.5498 in calibration set, the R_P^2 and RMSEP were 0.9931 and 0.7390 in prediction set. The ratio of RPD, LOD and LOQ were 10.02, 1.53% and 5.06%, respectively. The results suggested that the model based on E-nose data was also suitable for quantitative detection of SM adulteration in RM.

For HS-GC-IMS models, the performance of SM concentration prediction was slightly superior to that of E-nose models. The R_C^2 of the calibration set was 0.9969, and the R_P^2 of the prediction set was 0.9958, showing the prediction values of SM concentration were in excellent agreement with the true value. The RMSEC and RMSEP were 0.3642 and 0.5621, indicating the high robustness of the calibration models. The obtained RPD was up to 13.27, when PLS models with RPD $>$ 2 were recommended for screening purposes, and RPD greater than 3 was good for prediction (VESTIA, BARROSO, FERREIRA, GASPAR & RATO, 2019). LOD and LOQ were 1.43% and 4.72%. The obtained results indicated that HS-GC-IMS combined with PLS model was suitable for quantitative detection of SM adulteration in RM. In other research on adulteration detection of cow milk in buffalo and goat milk, the detection level was higher than 5% and R^2 values were in the range of 0.96–0.98 by FTIR combined with PLS-discriminant analysis (SEN, DUNDAR, UNCU & OZEN, 2021). The dispersions of calibration set composed of 21 samples based on HS-GC-IMS and E-nose data intuitionally (Fig. 4C) showed that most sample points were evenly distributed on the line, which confirmed that both HS-GC-IMS and E-nose technologies had the potential for rapid detection of possible adulteration in RM with SM. However, HAN et al. (2022) reported poor performance in quantitative analysis of adulteration level of safflower seed oil by E-nose and GC-IMS combined with PLS, indicating that the same

A



B



C

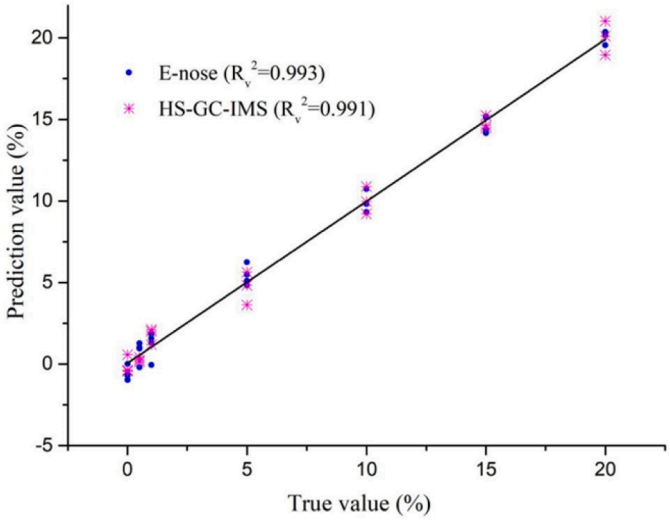


Fig. 4. PCA analysis of E-nose (A) and HS-GC-IMS (B), linear regression plots of true and prediction concentration of soymilk in milk validation sample set obtained by HS-GC-IMS and E-nose (C). The RM samples were adulterated with SM at the concentrations of 0.0%, 0.5%, 1.0%, 5.0%, 10.0%, 15.0%, and 20.0% (w/w), namely 0.5 %SM, 1.0 %SM, 5.0 %SM, 10.0 %SM, 15.0 %SM and 20.0 %SM.

Table 3
Statistical parameters of the PLS models for the quantification of soymilk in raw bovine milk using HS-GC-IMS and E-nose.

Method	Latent variables	Calibration			Prediction			RPD	LOD	LOQ
		n	R_c^2	RMSEC	n	R_P^2	RMSEP			
HS-GC-IMS	6	21	0.9969	0.3642	12	0.9958	0.5621	13.27	1.43%	4.72%
E-nose	4	21	0.9952	0.5498	12	0.9940	0.7390	10.02	1.53%	5.06%

R_c^2: decision coefficient of calibration set; R_P^2: decision coefficient of prediction set; RMSEC: root mean square error of calibration; RMSEP: root mean square error of prediction; LOD: limit of detection; LOQ: limit of quantification; RPD: the ratio of prediction to deviation; r: pearson correlation coefficient between reference and predicted value.

detection method was greatly affected by different mechanisms. Moreover, HS-GC-IMS is more usable to E-nose to some extent, which is in line with previous studies (Gu et al., 2020). This may be due to the fact that GC-IMS is extremely sensitive to low ppbv (parts per billion by volume) levels, thus components with trace content still could be detected among samples for significant differences (Wang et al., 2020b). Another predominant factor might be the influence of environmental conditions (temperature, humidity) on the responses of E-nose sensor, which will lead to some errors (DAMIANI et al., 2020). Additionally, a similar study on cow-buffalo milk using attenuated total reflectance Fourier transform infrared spectroscopy (ATR-FTIR) has also been reported, obtained inferior results in $R_C^2 = 0.99$ and $R_P^2 = 0.92$ (JAISWAL, JHA, BORAH, GAUTAM, GREWAL & JINDAL, 2015). Moreover, the LOD of present study (1.43% for HS-GC-IMS and 1.53% for E-nose) was lower than the previous results (1.75%) reported by GAUTAM et al. (2017). Therefore, as shown in Table S1, GC-IMS and E-nose combined with chemometric methods as rapid, non-destructive, and effective tools showed more convincing performances for the detection of adulteration in RM with SM which also provided a potential to predict the concentration of SM in RM.

4. Conclusions

In the present study, the volatile fingerprints of RM and adulterated sample with SM were established using fast Gas chromatography-based E-nose and HS-GC-IMS. Significant differences were observed between RM and SM in VOCs, while RM had fewer types compared to SM. The obtained data from HS-GC-IMS showed that diacetyl, 3-octane, 2-hexanone monomer, 2-butanone monomer and 2-butanone dimer were only identified in RM while 33 substances, e.g., butanal, isopropyl acetate was just detected in SM. E-nose data combined with PCA could afford a clearer distinction of adulterated and non-adulterated samples according to VOC profiles than that with HS-GC-IMS data, and those results were also better than the independent E-nose or HS-GC-IMS data on classification. Subsequently, the feasibility of SM adulterated levels in RM detected using HS-GC-IMS and E-nose fingerprints was assessed by PLS models. The RMSEP, R_P^2 of the prediction, and RPD of HS-GC-IMS and E-nose PLS model were 0.5621 and 0.7390, 0.9958 and 0.9940, 13.27 and 10.02, respectively. Meanwhile, the results indicated that the PLS models could rapidly identify SM levels adulterated in RM and the detection limits of HS-GC-IMS quantitative models was 1.53% which was superior to that of E-nose (1.43%). Both HS-GC-IMS and E-nose methods have the potential to recognize SM adulteration in RM with little sample preparation, and HS-GC-IMS fingerprinting could be used as an alternative tool for a truly fully automatable, labor-saving, cost-efficient, and in particular highly sensitive method. Nevertheless, HS-GC-IMS failed to detect alkane compounds, and there were still many signal peaks that had not been identified, so the HS-GC-IMS database needs further to be supplemented and refined. Follow-up studies are still needed to expand the scale of samples to improve the stability and reliability of models.

Declaration of Competing Interest

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

Data availability

The data that has been used is confidential.

Acknowledgement

This research was supported by “Science and Technology Innovation Action Plan” Agricultural Project of Shanghai Science and Technology Commission, grant number 19391902600.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fochx.2023.100696.

References

ALI, A., ABDULLA, M., MALIK, I., MAHMOUD, H. & HAMADNALLA, H. (2020a). Effect of Addition of Soymilk to Cheese Manufacturing and the Effect to the Quality of the Cheese during Storage Period. *Journal of Soil Science and Plant Physiology*, *2*(3), 126. 10.36266/JSSPP/126.

DOS SANTOS PEREIRA, E. V., DE SOUSA FERNANDES, D. D., DE ARAUJO, M. C. U., DINIZ, P. H. G. D. & MACIEL, M. I. S. (2020). Simultaneous determination of goat milk adulteration with cow milk and their fat and protein contents using NIR spectroscopy and PLS algorithms. *LWT-Food Science and Technology, 127*, 109427. 10.1016/j.lwt.2020.109427.

Ali, M. M., Hashim, N., Abd aziz, S., & Lasekan, O. (2020b). Principles and recent advances in electronic nose for quality inspection of agricultural and food products. *Trends in Food Science and Technology, 99*, 1–10. https://doi.org/10.1016/j.tifs.2020.02.028

Damiani, T., Cavanna, D., Serani, A., Dall'Asta, C., & Suman, M. (2020). GC-IMS and FGC-Enose fingerprint as screening tools for revealing extra virgin olive oil blending with soft-refined olive oils: A feasibility study. *Microchemical Journal, 159*, Article 105374. https://doi.org/10.1016/j.microc.2020.105374

FAN, R., XIE, S. B., WANG, S. F., Yu, Z. N., SUN, X. H., DU, Q. J., YANG, Y. X. & HANA, R. (2023). Identification markers of goat milk adulterated with bovine milk based on proteomics and metabolomics. *Food Chemistry: X*, *17*, 100601. 10.1016/j.fochx.2023.100601.

FARAH J. S., CAVALCANTI, R. N., GUIMARAES, J. T., BALTHAZAR, C. F., COIMBRA, P. T., PIMENTEL, T. C., ESMERINO, E. A., DUARTE, M. C. K. H., FREITAS, M. Q., GRANATO, D., NETO, R. P. C., TAVARES, M. I. B., CALADO, V., SILVA, M. C. & CRUZ, A. G. (2021). Differential scanning calorimetry coupled with machine learning technique: An effective approach to determine the milk authenticity. *Food Control, 121*, 107585. 10.1016/j.foodcont.2020.107585.

Feng, L., Zhu, S. S., Chen, S. S., Bao, Y. D., & He, Y. (2019). Combining Fourier Transform Mid-Infrared Spectroscopy with Chemometric Methods to Detect Adulterations in Milk Powder. *Sensors (Basel), 19*(13), 2934–2948. https://doi.org/10.3390/s19132934

Gautam, P. B., Sharma, R., Lata, K., Rajput, Y. S., & Mann, B. (2017). Construction of a lateral flow strip for detection of soymilk in milk. *J Food Sci Technol, 54*(13), 4213–4219. https://doi.org/10.1007/s13197-017-2890-3

Gu, S., Chen, W., Wangr, Z. H., Wangr, J., & Huo, Y. J. (2020). Rapid detection of Aspergillus spp. infection levels on milled rice by headspace-gas chromatography ion-mobility spectrometry (HS-GC-IMS) and E-nose. *LWT-Food Science and Technology, 132*, Article 109758. https://doi.org/10.1016/j.lwt.2020.109758

Jaiswal, P., Jha, S. N., Borah, A., Gautam, A., Grewal, M. K., & Jindal, G. (2015). Detection and quantification of soymilk in cow-buffalo milk using Attenuated Total Reflectance Fourier Transform Infrared spectroscopy (ATR-FTIR). *Food Chemistry, 168*, 41–47. https://doi.org/10.1016/j.jfca.2020.103785

JARIYASOPIT, N., KHAMSAENG, S., PANYA, A., VINAISURATERN, P., METEM, P., ASAWALERTPANICH, W., VISESSANGUAN, W., SIRIVATANAUKSORN, VORAPAN. & KHOOMRUNG, S. (2021). Quantitative analysis of nutrient metabolite

compositions of retail cow's milk and milk alternatives in Thailand using GC-MS. *Journal of Food Composition and Analysis, 97*, 103785. 10.1016/j.jfca.2020.103785.

Kerezsi, A. D., Jacquet, N., & Blecker, C. (2022). Advances on physical treatments for soy allergens reduction-A review. *Trends in Food Science & Technology., 122*, 24–39. https://doi.org/10.1016/j.tifs.2022.02.007

Han, L., Chen, M., Li, Y. T., Wu, S. S., Zhang, L., Tu, K., … Song, L. J. (2022). Discrimination of different oil types and adulterated safflower seed oil based on electronic nose combined with gas chromatography-ion mobility spectrometry. *Journal of Food Composition and Analysis, 114*, Article 104804. https://doi.org/10.1016/j.jfca.2022.104804

Ji, Z. Y., Zhuang, J. Y., Deng, C. X., Hu, Z. S., Du, Q. J., Guo, T. J., … Yang, Y. X. (2023). Identification of mare milk adulteration with cow milk by liquid chromatography-high resolution mass spectrometry based on proteomics and metabolomics approaches. *Food Chemistry, 405*, Article 134901. https://doi.org/10.1016/j.foodchem.2022.134901

Li, H. Q., Xi, B., Yang, X. L., Wang, H. B., He, X. N., Li, W. H., & Gao, Y. Q. (2022). Evaluation of change in quality indices and volatile flavor components in raw milk during refrigerated storage. *LWT-Food Science and Technology, 165*, Article 113674. https://doi.org/10.1016/j.lwt.2022.113674

Li, Q. Q., Yu, Z. B., Zhu, D., Meng, X. H., Pang, X. M., Liu, Y., … Chen, G. (2017). The application of NMR-based milk metabolite analysis in milk authenticity identification. *Journal of the Science of Food and Agriculture, 97*(9), 2875–2882. https://doi.org/10.1002/jsfa.8118

Roy, M., Doddappa, M., Yadav, B. K., Jaganmohan, R., Sinija, V. R., Manickam, L., & Sarvanan, S. (2022). Detection of soybean oil adulteration in cow ghee (clarified milk fat): An ultrafast study using flash gas chromatography electronic nose coupled with multivariate chemometrics. *Journal of the Science of Food and Agriculture, 102*, 4097–4108. https://doi.org/10.1002/jsfa.11759

Sacchi, R., Marrazzo, A., Masucci, F., di Francia, A., Serrapica, F., & Genovese, A. (2020). Effects of Inclusion of Fresh Forage in the Diet for Lactating Buffaloes on Volatile Organic Compounds of Milk and Mozzarella Cheese. *Molecules, 25*(6), 1332. https://doi.org/10.3390/molecules25061332

Sen, S., Dundar, Z., Uncu, O., & Ozen, B. (2021). Potential of Fourier-transform infrared spectroscopy in adulteration detection and quality assessment in buffalo and goat milks. *Microchemical Journal, 166*, Article 106207. https://doi.org/10.1016/j.microc.2021.106207

Shi, Y. G., Yang, Y., Piekoszewski, W., Zeng, J. H., Guan, H. N., Wang, B., … Zhang, N. (2019). Influence of four different coagulants on the physicochemical properties, textural characteristics and flavour of tofu. *International Journal of Food Science and Technology, 55*(3), 1218–1229. https://doi.org/10.1111/ijfs.14357

Tian, H. X., Chen, B., Yu, H. B., Lou, X. M., Li, Y., Yu, H. Y., & Chen, C. (2022). Rapid detection of neutralising acid adulterants in raw milk using a milk component analyser and chemometrics. *Food Additives & Contaminants: Part A, 39*(9), 1501–1511. https://doi.org/10.1080/19440049.2022.2093985

Uncu, A. O., & Uncu, A. T. (2020). A barcode-DNA analysis method for the identification of plant oil adulteration in milk and dairy products. *Food Chemistry, 326*, Article 126986. https://doi.org/10.1016/j.foodchem.2020.126986

Utz, F., Kreissl, J., Stark, T. D., Schmid, C., Tanger, C., Kulozik, U., … Dawid, C. (2021). Sensomics-Assisted Flavor Decoding of Dairy Model Systems and Flavor Reconstitution Experiments. *Journal of Agricultural and Food Chemistry, 69*(23), 6588–6600. https://doi.org/10.1021/acs.jafc.1c02165

Vestia, J., Barroso, J. M., Ferreira, H., Gaspar, L., & Rato, A. E. (2019). Predicting calcium in grape must and base wine by FT-NIR spectroscopy. *Food Chemistry, 276*, 71–76. https://doi.org/10.1016/j.foodchem.2018.09.116

Wang, J., Kuang, H. Q., Zhang, Z. H., Yang, Y. Q., Yan, L., Zhang, M., … Guan, Y. F. (2020a). Generation of seed lipoxygenase-free soybean using CRISPR-Cas9. *The Crop Journal, 8*(3), 432–439. https://doi.org/10.1016/j.cj.2019.08.008

Wang, S. Q., Chen, H. T., & Sun, B. G. (2020b). Recent progress in food flavor analysis using gas chromatography-ion mobility spectrometry (GC-IMS). *Food Chemistry, 315*, Article 126158. https://doi.org/10.1016/j.foodchem.2019.126158

Wang, X. R., Rogers, K. M., Li, Y., Yang, S. P., Chen, L. Z., & Zhou, J. H. (2019). Untargeted and Targeted Discrimination of Honey Collected by Apis cerana and Apis mellifera Based on Volatiles Using HS-GC-IMS and HS-SPME-GC-MS. *Journal of Agricultural and Food Chemistry, 67*(43), 12144–12152. https://doi.org/10.1021/acs.jafc.9b04438

Zhang, L., Shuai, Q., Li, P., Zhang, Q., Ma, F., Zhang, W., & Ding, X. (2016). Ion mobility spectrometry fingerprints: A rapid detection technology for adulteration of sesame oil. *Food Chemistry, 192*, 60–66. https://doi.org/10.1016/j.foodchem.2015.06.096

Zhu, X. Y., Wen, J., & Wang, J. G. (2020). Effect of environmental temperature and humidity on milk production and milk composition of Guanzhong dairy goats. *Veterinary and Animal Science, 9*, Article 100121. https://doi.org/10.1016/j.vas.2020.100121