Multivariate data analysis strategy to monitor Trentingrana cheese real-scale production through volatile organic compounds profiling

Michele Ricci, Flavia Gasperi, Emanuela Betta, Leonardo Menghi, Isabella Endrizzi,
Danny Cliceri, Pietro Franceschi, Eugenio Aprea

PII: S0023-6438(22)01299-3

DOI: https://doi.org/10.1016/j.lwt.2022.114364

Reference: YFSTL 114364

To appear in: *LWT - Food Science and Technology*

Received Date: 6 September 2022

Revised Date: 16 December 2022

Accepted Date: 20 December 2022

Please cite this article as: Ricci, M., Gasperi, F., Betta, E., Menghi, L., Endrizzi, I., Cliceri, D., Franceschi, P., Aprea, E., Multivariate data analysis strategy to monitor Trentingrana cheese real-scale production through volatile organic compounds profiling, *LWT - Food Science and Technology* (2023), doi: https://doi.org/10.1016/j.lwt.2022.114364.

Author statement

Conceptualization, F.G., and E.A.; methodology, F.G., E.A., M.R. and P.F.; validation, P.F., M.R.; formal analysis, M.R.; investigation, M.R., F.G., I.E., L.M., D.C. and E.A.; data curation, M.R., F.G., I.E., L.M., E.B., D.C., and E.A.; writing—original draft preparation, M.R.; writing—review and editing, M.R., F.G., I.E., L.M., D.C., P.F., E.B. and E.A.; visualization, M.R.; supervision, F.G., and E.A.; project administration, F.G.; funding acquisition, F.G.

Multivariate data analysis strategy to monitor Trentingrana cheese real-scale production through volatile organic compounds profiling

Michele Ricci[a,b], *Flavia Gasperi*[a,b], *Emanuela Betta*[b], *Leonardo Menghi* [a,b,c], *Isabella Endrizzi*[b],
Danny Cliceri[a], *Pietro Franceschi*[b], *Eugenio Aprea*[a,b]

a University of Trento - Center Agriculture Food Environment; Via E. Mach, 1 – San Michele All'Adige (TN) – 38010 – Italy;

b Fondazione Edmund Mach - Research and Innovation Centre; Via E. Mach, 1 – San Michele All'Adige (TN) – 38010 – Italy;

c University of Southern Denmark – Department of Technology and Innovation; Campusvej 55 – Odense – 5230 – Denmark.

Abstract

Volatile organic compounds (VOCs) in cheese, as result of the chemical, physical and microbiological properties of the raw milk, are related to its sensory properties and consumer's acceptability. Measurement of VOCs can be related to the quality of the production process, highlighting changes in the raw materials or the process conditions. In the present study, we tested the suitability of ANOVA-Simultaneous Component Analysis (ASCA) to extract useful information from volatile organic compound data measured over two years of production of Trentingrana cheese in a real production context where several confounding factors are present.

A total of 317 cheese wheels were collected from the 15 cooperative dairy factories every two months. The ASCA analysis indicates that the milk collection affects the VOC profiles. To deeper investigate this factor, an Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) model was developed to estimate the associations between VOCs and process characteristics of the dairy factory. Results showed that the milk collection procedure affects the content of organic acids, esters, and ketones of the cheeses.

Keywords

Grana Cheese; Volatile Organic Compounds; Anova-simultaneous Component Analysis; Orthogonal Partial Least Squares Regression

Abbreviations

VOCs = Volatile Organic Compounds

PDO = Product Designation of Origin

SPME/GC-MS = Solid Phase Micro Extraction Gas Chromatography-Mass Spectrometry

ANOVA = Analysis of variance

ASCA = ANOVA Simultaneous Component Analysis

O-PLS-DA = Orthogonal Partial Least Squares Discriminant Analysis

DMC = Double milk collection

SMC = Single milk collection

MMC = Mixed milk collection

1. Introduction

Volatile organic compounds (VOCs) are molecules characterized by high vapor pressure at room temperature and low water solubility. Several VOCs in food contribute to odors and flavors that play a key role in sensory quality perception and liking responses (Liaw, Miracle, Jervis, Listiyani, & Drake, 2011, Khattab, Guirguis, Tawfik, & Farag, 2019). In cheese, VOCs are produced to a great extent during ripening by the catabolic activity of microorganisms on carbohydrates, lipids, and proteins naturally present in milk and rennet (Kilcawley, Faulkner, Clarke, O'Sullivan, & Kerry, 2018; Marilley & Casey, 2004; McSweeney & Sousa, 2000; McSweeney, Ottogalli, & Fox, 2004). The metabolic pathways responsible for the synthesis of VOCs are affected by the properties of raw milk and the conditions of the production process. For this reason, VOCs are considered reliable markers of process quality and traceability of cheese products (Pisano, Scano, Murgia, Cosentino, & Caboni, 2019; Suh, 2022).

Trentingrana cheese, produced under the European Protected Designation of Origin (PDO) of Grana Padano (EC Commission Regulation No. 1107, 1996), is a semi-fat, hard, cooked cheese that undergoes a slow ripening period of up to 2 years, even longer for some wheels. The production process of Trentingrana has distinctive aspects (MiPAF, 2006): the use of raw cow milk only from livestock on mountain terrains in a delimited area (Autonomous Province of Trento, Northeast Italyy), the application of restricted cattle feeding, and the removal of lysozyme and silage from the cow’s feeding (D.P.R. n. 1269, 30 October 1955).

The Trentingrana Consortium includes 15 dairy factories, producing first-quality cheese according to official guidelines. The official disciplinary allows producers to collect milk from different farms that differ according to cow’s breed, altitude, and use of unifeed mixer wagons or traditional feeding procedures (Bittante et al., 2011). Additionally, the disciplinary allows applying slight changes in rennet, whey starters, and heating/storing machinery used in the dairy factory. Altogether, this can reasonably affect the peculiar physical and sensory properties of the final product (Ricci et al., 2022).

Trentingrana dairy factories also differ in the adoption of a double or a single milk collection procedure (Endrizzi et al., 2012), which determines differences in storage time, temperature conditions of the raw milk before transformation, and intensity of the milk skimming process. This latter process decreases bacterial and somatic cell counts by natural gravity separation of fat, thus standardizing the properties of fat and casein/fat ratio (McSweeney et al., 2004). The effects of the milk collection procedure, the skimming process, and storage temperature on the sensory and chemical properties of Trentingrana cheese have already been studied. Endrizzi et al. (2012) found differences in physical properties and sensory quality attributes in cheese wheels produced in pilot plants with different milk collection procedures and in different seasons, showing significant differences in commercial quality, colorimetric properties, and the VOC profile. Franciosi, De Sabbata, Gardini, Cavazza, & Poznanski (2011) reported the effect of the milk collection procedure on chemical and VOCs composition in Trentingrana cheese showing a higher content of free fatty acids and related esters in cheese wheels produced using double collection without refrigeration. In a similar study, Fabris et al. (2010) trained a random forest classifier to recognize the milk collection procedure from the VOC content in Trentingrana cheese and to highlight which molecules are determinants for discriminating the cheese wheels produced in different seasons.

Monitoring the VOC profile of Trentingrana in its real-scale production process is functional to understanding how to operatively improve its quality: associating the presence of chemical compounds to a production process condition or a feature of the final product allows to develop a faster quality control procedure and to estimate how the issue studied is related to the chemical properties of food (Ellis & Mayhew, 2014).

Previous studies were done using a restricted batch of samples from pilot plants with a balanced experimental design to study the factors of interest, and thus excluding at multivariate level the effect of the other factors that may influence the final quality of the product in a real production context. Overall, previous results highlighted that there exists a need to develop a large-scale monitoring of the chemical properties of Trentingrana cheese, to estimate the significance and the importance of the factors investigated in previous experiments in the real context of the production process. Because of the presence of many factors and the multivariate structure of VOC data, there is also the need to develop a functional and reliable statistical procedure to infer the effect of the process-related factors at multivariate level removing the nuisance due to confounding factors.

The objective of this work is to test the suitability of ANOVA-Simultaneous Component Analysis (ASCA) to extract useful information from volatile organic compound data measured in Trentingrana cheese in a real production context where several confounding factors are present, and no experimental design is designed focusing on predefined *a priori* factors. To this end, within the collaboration with Trentingrana – Consorzio dei Caseifici Sociali Trentini (Italy), the VOC profile of Trentingrana cheese over two years of production was analyzed by SPME/GC-MS, sampling a representative selection of cheese wheels from its real-scale production process. We estimated how VOC profiles are related to distinct raw materials, different processing of the different cheese factories, the enzymatic activity of the rennet adopted, and different parts of the year when the milk is produced.

A two-step analytical process is presented: first, for the estimation of the significance and the effect size of two process-related factors (Dairy Factory and Time of the year when the wheels are produced), an ANOVA-Simultaneous Component Analysis was adopted. Analyzing the results, an overall tendency in the VOC content among dairy factors was detected. The effect of the milk collection procedure adopted by the dairy factory is proposed as an interpretation for this tendency.

To test this hypothesis and to estimate its effect, an O-PLS-DA predictive model was trained and validated (Smilde, Timmerman, Hendriks, Jansen, & Hoefsloot 2012; Trygg & Wold 2002).

2. Materials and methods

2.1 Cheese samples

A total of 317 cheese wheels were bi-monthly sampled from the Trentingrana Consortium repository (located in the Autonomous Province of Trento, Northeast Italy) during the years 2017 and 2018. The wheels were produced from November 2015 to October 2017 by 15 different dairy factories located in the Province of Trento (Italy), ripened for 18±2 months, and labeled progressively from C-1 to C-15 (production traits of each dairy are summarized in Table 1). For each dairy factory, the number of samples varied in proportion to the volume of cheese wheels delivered from a minimum of 1 cheese wheel and a maximum of 3. For more details on sampling criteria, see Ricci et al. (2022). The samples for the analysis of VOCs were prepared by taking 24 parallelepipeds of cheese (3x1.5x1.5 cm) from various positions of the freshly cut half part of the wheel then finely grounded and well mixed. Approximately 3 grams of grounded cheese were weighted inside a 20 mL GC-MS vial (Supelco, Bellefonte, CA, USA), capped with PTFE/silicone septa (Supelco, Bellefonte, CA, USA), and stored at -80°C. For each cheese wheel, three vials from the same mixing were prepared. Before the analysis, samples were thawed for 1 hour at room temperature, then each vial was spiked with the internal standard, just before the beginning of the analysis. Each sample was classified according to the dairy factory where it was produced, the part of the year when the milk was collected, and the milk collection procedure adopted for its production. At the beginning of the sampling procedure, from one of the first cheese wheels, 100 vials were prepared with the same grounded cheese mix and stored at -80°C than used as quality control (QC sample) during GC-MS analysis over time.

In dairy factories where double milk collection (DMC) procedure is adopted, the full-fat milk of the evening milking is delivered to the cheese factory and undergoes a gravity separation process overnight in large vats (Ma & Barbano, 2000). After that, the milk of the morning milking is added to the semi-skimmed milk and used to produce cheese according to the standard cheese-making procedure of the Trentingrana. The single milk collection (SMC), instead, consists in storing the milk of the morning milking at the dairy farm in controlled conditions and then adding the evening milk, before moving the raw milk to the cheese factory, where the skimming procedure takes place overnight. Samples produced in the dairy factory labeled C-2 were classified in a third-class called mixed milk collection (MMC) because in that specific dairy factory both collection procedures are used according to the farm where the milk is collected.

2.2 SPME/GC-MS analysis

The procedure for Headspace solid-phase microextraction coupled with gas chromatography-mass spectrometry (SPME/GC-MS) was performed according to Endrizzi et al. (2012) with a few amendments. The samples were equilibrated at 40 °C for 30 min, and then in the headspace environment, a fused silica fiber coated with 2 cm of 50/30 µm divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS, Supelco, Bellefonte, PA, USA) was inserted and exposed for 30 min without changing the temperature. The desorption of the volatile compounds from the SPME fiber was performed at 250 °C for 5 minutes in the injector port of a GC-MS operating in electron ionization mode (EI, internal ionization source; 70 eV). The control of the procedure phases was managed using an auto-sampling system (CTC combiPAL, CTC Analysis AG, Zwingen, Switzerland) equipped with a cooling system that kept the vials at 4 °C before the start of the analysis. Separation was conducted on an HP-Innowax fused silica capillary column (30 m, 0.32 mm ID, 0.5 µm film thickness; Agilent Technologies, Palo Alto, CA, USA). Separation conditions were as follows: carrier gas was helium at a constant flow rate of 2 mL/min; oven temperature programming was 40 °C for 3 min, an increase from 40 to 180 °C at 4

°C/min, stationary at 180 °C for 6 min, then another increase from 180 to 220 °C at 5 °C/min and finally, 220 °C for 3 minutes. The mass spectrometer operated a mass scan range from 33 to 300 m/z (GC Clarus 500, PerkinElmer, Norwalk CT, USA).

Compound identification was based on mass spectra matching with those present in the standard NIST14 (NIST/EPA/NIH, 2014) library and linear retention times calculated injecting C7-C30 n-hydrocarbon series under the same chromatographic conditions. Compounds were semi-quantified spiking samples with 4-methyl-2-pentanone (Sigma-Aldrich) as I.S. 0.05 g/L in aqueous solution. Amount of VOCs in the samples were expressed as µg/kg equivalent of the I.S.

The analytical measurements were performed over a period of two months and required four different batches of SPME fibers to overcome the decline of the performances due to the deterioration of the fiber itself. The repeatability of the method was assessed for each batch of SPME fiber, analyzing twelve replicates of a reference cheese on the same day. The observed average variation, estimated for the classes of acids, esters, ketones, and aldehydes agreed with the literature for SPME analysis with this type of matrix (Barbieri et al., 1994; Bellesia et al., 2003, results in table A.1 in Appendix). Furthermore, a QC samples was measured every ten cheese samples over all the period of measurements.

2.3 Statistical analysis

VOCs data were analyzed using ANOVA simultaneous component analysis (ASCA, Smilde et al., 2012) to identify multivariate patterns significantly associated with the different study factors: Dairy Factories, Time of the year, and their interaction, and the effect of the batch of SPME fiber. The ASCA model decomposes the signal of each volatile x in the following form:

$$x_{jkni} = \mu + \alpha_j + \beta_k + \gamma_n + (\alpha\beta)_{jk} + \epsilon_i$$

Where μ represent the overall grand mean of the volatile compound, α_j the expected value for the jth Dairy Factory, β_k the expected value for the kth Time of the year, γ_n the expected value for the nth batch of SPME fibers, $(\alpha\beta)_{jk}$ the interaction between the Dairy Factory and Time and ϵ_i the residual error for the ith cheese wheel representing the natural variability of each cheese wheel. The effect of the batch of SPME fibers is considered a known and controllable nuisance source of variability and it is integrated into the model as a blocking factor (Montgomery, 2013).

A permutation test (n = 1000) was applied to assess univariate statistical significance of each factor for each volatile compound estimating empirical null distributions for the univariate sum of squares (α = 0.05).

The expected values matrix for each factor was estimated by calculating for each compound, after unit variance scaling, the effects for each level from ANOVA decomposition for that factor. Those matrices were mean centered for each compound and transformed using singular values decomposition (SVD), to analyze the multivariate structure of the effects of each level.

ASCA frameworks removes the effect of all the known confounding variables considering the internal correlation structure due to the common metabolic pathways for many volatile organic compounds.

An Orthogonal Partial Least Squares – Discriminant Analysis (O-PLS-DA, Trygg & Wold, 2002) classifier was developed to deeper investigate the Milk Collection procedure adopted to produce each cheese wheel from its VOC profile.

The model was built using a restricted dataset of the 14 dairy factories that use single or double milk collection (n=306, prevalence of DMC= 70.9%), to analyze only the two most prominent modalities of the milk collection process. Data were partitioned into train and test sets (train/test ratio = 0.8) and a cross-validation procedure was performed in the train set to estimate the optimal number of orthogonal projections. To assess the predictive capacity of the model, the results from repeated partitions were compared with a null distribution obtained by a permutation test (n = 1000) in terms of sensitivity, specificity, and overall accuracy (using Cohen's Kappa index, Ferri, Hernández-Orallo, & Modroiu, 2009).

To identify which VOCs were related to the variation of the Milk Collection Procedure, a bootstrap procedure (n = 1000) was employed to estimate the confidence intervals and the significance of the regression coefficient of every VOC included in the model (Lazraq, Cléroux, & Gauchi, 2003). The validation of the coefficients of the model identified the VOCs affected by the different milk collection procedure. This procedure of testing for significance considers the dimensionality and the structure of the data as it is modeled by O-PLS-DA and does not require standard statistical assumptions.

All the statistical analyses have been performed using R version 4.1.0 (R Core Team, 2021), the ggpubr package version 0.4.0 (Kassambara, 2020), the factoextra package version 1.0.7 (Kassambara & Mundt, 2020), the caret package version 6.0-90 (Kuhn, 2021), and the ropls package version 1.24.0 (Thevenot, Roux, Xu, Ezan, & Junot, 2015).

3. Results and discussion

3.1 Qualitative VOCs assessment

A total of 75 volatile organic compounds have been identified by SPME/GC-MS analysis. These compounds belong to the following chemical classes: esters (n = 17), alcohols (n = 13), ketones (n = 11), acids (n = 9), aldehydes (n = 8), sulfurs (n = 5), hydrocarbons (n = 4), phenols (n = 3), lactones (n = 2), terpenes (n = 2) and pyrazines (n = 1). Overall results are summarized in table A.2 in Appendix.

Identified compounds agreed with the literature on VOCs in grana cheese (Qian & Reineccius, 2002). The most prominent compound type by overall relative concentration is organic acids, followed by ketones and alcohols. Those classes contain several molecules that are directly related to the natural content of raw milk, such as medium-chain fatty acids, and they are naturally occurring in many milk-based products due to lipid catabolism by endogenous enzymes and microbial activity (Collins, McSweeney, & Wilkinson, 2004).

Esters were characterized by a high overall mean but also high variability. These compounds are synthetized from the lipidic fraction by the microbial activity in milk during ripening and they are often associated with positive sensory descriptors in hard seasoned cheese (Liu, Holland, & Crow, 2004; Qian & Reineccius, 2002). High level of variance in ester compounds were already detected in previous works on Trentingrana cheese (Fabris et al., 2010). Conversely, both terpenes and hydrocarbons were present at low levels with high variability because most of them were not detected in all the samples. According to literature, these compounds are related to the cows' diet and the seasonal effect and are not naturally occurring in ripened cheeses (Kilcawley et al., 2018). Lastly, phenol, 3-methyl phenol and 4-methyl phenol are mostly related to amino-acid metabolism, however their presence may also be related to the diet and the external environment (Curtin & McSweeney, 2004; Panseri, Luca, Zecconi, Soncini, & Noni, 2014).

3.2 ANOVA simultaneous component analysis

The percentage of total variance explained by each factor and interaction was estimated according to Bertinetto, Engel, & Jansen (2020) by calculating the percentages for each factor of the sum of squares (supplementary figure A).

The high percentage of explained variance related to the effect of the SPME fibers highlighted that there is an important systematic error related to the 4 different batches of fibers used. Even if the effect of a measurement-related bias is important, the ASCA framework allows to analyze the effect of Dairy Factory and Time removing the effect of a potential confounding factor such as the variation of the SPME fiber.

Results of ASCA permutation test showed no significant effects at a univariate level for the interaction of Dairy Factory and Time, while 3 molecules were significantly responding to the factor Time. The Dairy Factory factor was significant for 46 molecules out of 75 and the SPME fiber factor was significant for 55 out of 75 compounds (Supplementary figure B).

Considering that none of the compounds was significantly responding to the interaction factor (Time: Dairy Factory), the multivariate decomposition of this term was not considered.

The Time factor has been included in the model, but it has been analyzed also at a univariate level.

The permutation test demonstrated that the ASCA decomposition with the model proposed is representative of the overall data structure. Results are discussed in sections 3.1.1 and 3.1.2 for the factor Dairy Factory and Time of the year, respectively.

To verify the presence of significant factors not included in the model of the multivariate decomposition, the PCA biplot of the residuals was analyzed to estimate the presence of effects not represented in the model (supplementary figure C).

3.2.1 Effect of the Dairy Factory

In ASCA decomposition, the Dairy Factory factor was the factor related to the production process that described the largest percentage of explained variance (14.5%, supplementary image A), excluding the blocking factor SPME fiber. The variations in the production process adopted in the dairy factory significantly affect the VOCs profile of Trentingrana cheese more than the other factors included in the model.

The results of the multivariate decomposition of the Dairy Factory term of the ASCA decomposition are shown in figure 1. The biplot indicates that the first two components account for 51.3% of the overall variability. The first principal component separated from left to right all the dairies for the content of organic acids, their esterified form, and 1- and 2- butanol. Along with this component, the samples were separated from the right to left for the content of two ketones (propanone and pentanone), and from left to right for the content of free fatty acids and their esterified forms.

On the first component, which explained 28.8% of the overall variance, the dairy factories were distributed in different groups, with three dairies (C-7, C-4, and C-15) placed on the far left, C-6 and C-13 placed on the left side, C-1, C-2, C-5, C-9, C-12, and C-14 placed in the central position of the axis, and C-3, C-8, C-10 and C-11 in the right side of the plot.

The second principal component explained 22.5% of the overall variance and was related to the variation among dairies due to propionic acid, limonene, alpha-thujene, 2,6-dimethyl pyrazine, and alcohols such as 2-propanol, 2-butanol, and 2-heptanol. Along with this component, the dairy factory C-11 was separated from all the others in the lower part of the graph. This was due mostly to the higher content of limonene, and to the lower content of 2,6-dimethyl pyrazine. Higher levels of limonene have previously been related to the cow's diet (Kilcawley et al., 2018) or to process related contaminants from industrial detergents. Conversely, the formation of 2,6-dimethyl pyrazine is related to the Maillard reaction occurring during milk cooking during cheese production (Divine, Sommer, Lopez-Hernandez, & Rankin, 2012), and to the higher content of propionic acid, which is related to the activity of contaminant microbes, which are associated to the handling of the raw milk and the condition of the production process (Giraffa, 2021). The presence of propionic bacteria in Trentingrana was already reported by Rossi, Gatto, Sabattini, & Torriani (2012), who also found significant differences in the microbial activity during ripening between dairies and between different parts of the year.

The formation of free fatty acids and their esterified form is related to the catabolism of triglycerides during ripening (Collins et al., 2004). The distinct levels of these molecules between dairies along the first principal component suggest that the concentration is related to the process of the milk collection procedure.

The effect of the dairy factory on the textural and colorimetric properties of Trentingrana cheese was reported in a previous work (Ricci et al., 2022). Comparing the results, the dairy factories that were similar for the overall physical properties of their cheese were not similar for the overall VOCs profile. This could be due to the fact that the factors that affect the physical properties and the factors that affect the formation of VOCs in Grana cheese are different: color and texture of cheese are mostly affected by the properties of the raw milk and the treatments of the curd, while VOCs formation is affected also in a large scale by the microbial activity during the ripening process (Kilcawley et al., 2018; Divine et al., 2012; Fox, Guinee, Cogan, & McSweeney, 2017). Noteworthily, the dairy factories differ in the farm producing the raw milk, the heating and storing machinery used, the properties of the whey starter, and the milk collection procedures, which can be double or single. Comparing the results to the information about the dairies available in table 1, it should be noted that the factor that affects the overall variation of VOCs between dairy factories is the milk collection procedure. Conversely, the average volume of production for year, the number of farms delivering the milk, and the adoption of unifeed alimentation system in the farms are not directly related to the formation of volatile organic compounds in hard seasoned cheese. These results highlight the importance of the milk collection procedure on the overall chemical profile of Trentingrana cheese, coherently with the results reported by Endrizzi et al. (2012).

Overall, the ASCA model highlighted the importance of the milk collection procedure on the VOC profile of Trentingrana cheese and separated the dairy C-11 from the others in the second principal component, detecting that the production process differs in that specific plant, modifying the concentration of other volatile organic compounds. These results indicated that ASCA is a valuable data analysis procedure to recognize significant differences at process levels in large-scale sampling procedures, such as routinary sampling procedures. The relations between volatile compounds and the milk collection procedure is further discussed in section 3.3.

3.2.2 Effect of the Time of the year

As shown in supplementary figure A, only 2.6 % of the multivariate variance was explained by the factor Time. Hence, considering the small number of significantly different molecules for this factor according to the permutation test (supplementary figure B), the multivariate ASCA decomposition could be misleading, and it was reported only in the supplementary figure D.

The permutation test reports that the three molecules that vary significatively by the production time of the year included 3-methyl phenol (*m*-cresol), *p*-xylene, and ethylbenzene.

Ethylbenzene is classified as a pollutant (Panseri et al., 2014), and hence was not included in further analysis.

The effect of the time of the year on these single compounds is reported in the boxplot in figure 2. The formation of p-xylene and 3-methyl-phenol is associated with the metabolism of aromatic amino acids (Curtin & McSweeney, 2004), and a seasonal effect in their concentrations was reported in raw milk used for cheesemaking by Faustini et al. (2019). Post-hoc analysis highlighted that the cheese wheels produced from September to December have a lower content of 3-methyl-phenol than those produced from January to April, and higher levels of ethyl benzene and p-xylene than those produced from March to June.

The estimated effect of time of the year when the wheels were produced on the content of VOCs was low at the multivariate level. Interestingly, Fabris et al. (2010), detected 8 molecules that changed significantly according to different part of the year when the milk is produced, analyzing with PTR-MS a small sample of cheese wheels from a pilot plant. The molecules were tentatively identified as medium-chain organic acids and ketones. The different results could be interpreted considering the smaller analytical power of the previous research due to the limited number of samples analyzed.

This could also be explained considering the variance added by the natural variability of the product: the seasonal effect may influence the volatile compounds in raw milk, but there are no indications that it could also modify the conditions during the production process and the ripening phase, thus a transformed product could be less affected by seasonal conditions.

3.3 O-PLS-DA predictive model

3.3.1 Model performances

The O-PLS-DA algorithm allows modeling separately the variations of the predictors correlated and orthogonal to the response. This model improves the explication of the effect of the predictors and their systematic variation compared to standard PLS (Pinto, Trygg, & Gottfries 2012). To estimate the effective presence of the effect of the milk collection procedure on the VOC content, the predictive O-PLS-DA model was validated using a permutation test. The significance was tested by comparing the performance indices of the models trained in the permutation test, considered as the null distribution, with the indices estimated from multiple partitions of the real data set. The comparisons between the replicates and the null distributions are shown in the violin plots in figure 3.

Results of the comparison between multiple partitions and null distribution demonstrated that the final O-PLS-DA model had significant predictive capacities, reported by the significant difference from the null distribution of the kappa index estimated from repeated partitions. Moreover, it efficiently separated the two groups, as reported by the high values of sensitivity and specificity and by their significant difference from null distribution. This model demonstrated that at multivariate level the content of VOCs in a single cheese wheel could be associated with a different milk collection procedure adopted by the dairy factory. The test adopted demonstrated that the model was representative of the underlying data structure and that its performances were reliable independently from the single data partition.

3.3.2 Importance of variables

The confidence interval of each regression coefficient of the model was estimated using a bootstrap approach, and the molecules whose confidence interval included 0 value were labeled as nonsignificant. The validated coefficient absolute values of the significant molecules are reported in the barplot of figure 4.

The bootstrap test of the O-PLS-DA model determined that the concentration of 44 volatile compounds is related to a different milk collection procedure. Results showed that the content of volatile compounds produced by the catabolism of fat in cheese, such as medium-chained free fatty acids, their esterified forms, and secondary alcohols, is related to the different milk delivery procedures.

The formation of 2-pentanone was related to the different pasture techniques of the cows by different studies (Kilcawley et al., 2018; Villeneuve et al., 2013). However, as the disciplinary of production regulates the food intake of the cows, the most reasonable production pathway is the oxidative pathway of fatty acids by microbial activity (Collins et al., 2004). Instead, 3-methyl butanal and 2-methyl-1-butanol are transitory compounds of branched amino acids' catabolism during ripening (Bovolenta et al., 2014).

Medium-chain fatty acids, such as hexanoic, heptanoic, octanoic, nonanoic, and decanoic acids were significantly higher in cheese wheels produced by a double delivery procedure of milk. The different quantities could be explained by a higher lipolytic activity due to the storage and collecting procedure, as suggested by Franciosi et al. (2012).

The formation of ethyl hexanoate, ethyl octanoate, and ethyl decanoate is related to the esterification of an organic acid with ethanol due to microbial activity, and their presence is related to the availability of free fatty acids (Kilcawley et al., 2018). Interestingly, such compounds were significantly higher in cheese from double milk collection procedures, thus reasonably suggesting these molecules as other reliable markers of the process. Moreover, these results are coherent with those highlighted by the ASCA model for the factor Dairy Factory.

According to Collins et al. (2004), lipase activity in cheese is affected by process conditions and microbial and enzymatic activity, and it is a critical step for the synthesis of secondary products of lipid metabolism during ripening. Wang & Randolph (1978) reported a reduction of the lipase activity in skim milk after temperature inactivation of the lipase naturally present in milk in conditions similar to the milk collection procedures reported in the present work. Eugster, Fuchsmann, Schlichtherle-Cerny, Bütikofer, & Irmler (2019), instead, reported that Non-Starter-Lactic-Acid-Bacteria (NSLAB) can produce acetoin, 2-butanone, and 2-butanol at high levels from pyruvate in hard and semi-hard cheeses, affecting the catabolism of amino acids positively. Altogether, an explanation for the variations related to the milk collection process could be that the different duration of the skimming process may affect the activity of the endogenous lipase enzyme in milk and of the non-starter lactic bacteria (NSLAB) naturally present in the raw milk, which grows better in raw milk collected using the double delivery procedure (Giraffa, 2021; Franciosi et al., 2011). The routinary analysis of the total microbial population in raw milk sampled from vats done by Trentingrana Consortium reported a significant difference between dairy factories according to the different milk collection procedure (data not shown).

The results reported on the effect of milk collection procedure are coherent with previous research in a single dairy factory with controlled conditions of milk collection procedures (Endrizzi et al., 2012), which reported significant differences in the content of organic acids and esters between cheese wheels produced adopting a different milk collection procedure. This confirms the validity of the adopted multivariate strategy based on ANOVA-Simultaneous Component Analysis (ASCA) to extract useful information from volatile organic compound data in a real production context where several confounding factors are present.

4. Conclusions

The analysis of food products directly sampled from the production process allowed measuring the relation between the production process' variables and the properties of the food product at a real scale semi-industrial level.

ANOVA simultaneous component analysis showed the effect of the season and the production plant in the content of volatile compounds in cheese, highlighting the differences between dairy factories due to milk collection and the sub-products of Maillard reaction at low temperatures.

These data highlight how the Trentingrana cheese chemical profile is affected by the first steps of the production process: raw milk storing and skimming. For this reason, the quality control procedure to produce hard seasoned cheeses needs also to monitor and uniform the conditions of the process in those early stages to ensure the same properties.

In conclusion, the proposed analytical framework can be applied in other research related to large-scale food production processes to reliably and effectively highlight the factors responsible for the differences observed when the latter are masked by several confounding cofactors.

Further research is needed to estimate the underlying mechanism at the chemical and microbial level of technological variations in the production process and their effect on the quality of cheese at the technological and sensory levels. The association of microbial and VOCs data could help to understand the synthesis of VOCs in Trentingrana cheese. Furthermore, if associated with process conditions data, they could provide useful information for the optimization of the production process.

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.

Acknowledgments

The research was funded by Trentingrana - Consorzio dei Caseifici Sociali Trentini, Italy (as part of the contract with Fondazione Mach, signed on February 28, 2020) and by the Autonomous Province of Trento, Italy (as part of the ADP funding prot. n. 244380 dd 04/05/2020 and the TRENTINGRANA project - RDP 201-202, CUP C66D17000180008).

The authors thanks dr. Andrea Goss from Trentingrana Concast for helpful technical discussion.

5. Appendix

Author statement

Conceptualization, F.G., and E.A.; methodology, F.G., E.A., M.R. and P.F.; validation, P.F., M.R.; formal analysis, M.R.; investigation, M.R., F.G., I.E., L.M., D.C. and E.A.; data curation, M.R., F.G., I.E., L.M., E.B., D.C., and E.A.; writing—original draft preparation, M.R.; writing—review and editing, M.R., F.G., I.E., L.M., D.C., P.F., E.B. and E.A.; visualization, M.R.; supervision, F.G., and E.A.; project administration, F.G.; funding acquisition, F.G. All authors have read and agreed to the published version of the manuscript.

Bibliography

1. Barbieri, G., Bolzoni, L., Careri, M., Mangia, A., Parolari, G., Spagnoli, S., & Virgili, R. (1994). Study of the Volatile Fraction of Parmesan Cheese. *Journal of Agricultural and Food Chemistry*, 42(5), 1170–1176. https://doi.org/10.1021/jf00041a023
2. Bellesia, F., Pinetti, A., Pagnoni, U. M., Rinaldi, R., Zucchi, C., Caglioti, L., & Palyi, G. (2003). Volatile components of Grana Parmigiano-Reggiano type hard cheese. *Food Chemistry*, 83(1), 55–61. https://doi.org/10.1016/S0308-8146(03)00037-2

3. Bertinetto C., Engel J., & Jansen J. (2020). ANOVA simultaneous component analysis: A tutorial review, *Analytica Chimica Acta: X*, 6. https://doi.org/10.1016/j.acax.2020.100061
4. Bittante G., Cecchinato A., Cologna N., Penasa M., Tiezzi F., & De Marchi M. (2011). Factors affecting the incidence of first-quality wheels of Trentingrana cheese, *Journal of dairy science*, 94(7), 3700-3707. https://doi.org/10.3168/jds.2010-3746
5. Bovolenta S., Romanzin A., Corazzin M., Spanghero M., Aprea E., Gasperi F. & Piasentier E. (2014). Volatile compounds and sensory properties of Montasio cheese made from the milk of Simmental cows grazing on alpine pastures, *Journal of Dairy Science*, 97(12), 7373-7385, https://doi.org/10.3168/jds.2014-8396.
6. Collins Y.F., McSweeney P.L.H., & Wilkinson M.G. (2004). Lipolysis and Catabolism of Fatty Acids in Cheese. In: Patrick F. Fox, Paul L.H. McSweeney, Timothy M. Cogan, Timothy P. Guinee (eds.), *Cheese: Chemistry, Physics and Microbiology* (pp. 373-389), Academic Press, https://doi.org/10.1016/S1874-558X(04)80075-7
7. Curtin Á.C. & McSweeney P.L.H. (2004). Catabolism of Amino Acids in Cheese during Ripening. In: P.F. Fox, P.L.H. McSweeney, T.M. Cogan, T.P. Guinee (eds.), *Cheese: Chemistry, Physics and Microbiology* (pp. 435-454), Academic Press. https://doi.org/10.1016/S1874-558X(04)80077-0
8. D.P.R. 30 ottobre 1955, n. 1269. Riconoscimento delle denominazioni circa i metodi di lavorazione, caratteristiche merceologiche e zone di produzione dei formaggi. Gazz. Uff. N. 295 del 22 Dicembre 1955.
9. MiPaf, Decreto 20 Luglio 2006. Modifica del decreto 13 gennaio 2006 relativo alla protezione transitoria accordata a livello nazionale alla modifica del disciplinare di produzione della denominazione di origine protetta «Parmigiano Reggiano», registrata con regolamento (CE) 1107/96 della Commissione del 12 giugno 1996, Gazzetta Ufficiale, 185 (2006), pp. 28-34

10. Divine R.D., Sommer D., Lopez-Hernandez A., & Rankin S.A. (2012). Short communication: Evidence for methylglyoxal-mediated browning of Parmesan cheese during low-temperature storage, *Journal of Dairy Science*, 95(5), 2347-2354, https://doi.org/10.3168/jds.2011-4828
11. EC Commission Regulation No. 1107, 21/06/1996, On the registration of geographical indications and designations of origin under the procedure laid down in article 17 of council regulation (EEC) no 2081/92 , Official Journal of the European Union, L148 (1996), pp. 1-10
12. Ellis A.M. & Mayhew C.A. (2014). PTR-MS in the Food Sciences. In: A.M. Ellis and C.A. Mayhew (eds.) *Proton Transfer Reaction Mass Spectrometry* (pp. 221-265). https://doi.org/10.1002/9781118682883.ch6
13. Endrizzi I., Fabris A., Biasioli F., Aprea E., Franciosi E., Poznanski E., Cavazza A., & Gasperi F. (2012). The effect of milk collection and storage conditions on the final quality of Trentingrana cheese: Sensory and instrumental evaluation, *International Dairy Journal*, 23(2), 105-114, https://doi.org/10.1016/j.idairyj.2011.10.004
14. Eugster E., Fuchsmann P., Schlichtherle-Cerny H., Bütikofer U., & Irmler S. (2019). Formation of alanine, α-aminobutyrate, acetate, and 2-butanol during cheese ripening by Pediococcus acidilactici FAM18098, *International Dairy Journal*, 96, 21-28, https://doi.org/10.1016/j.idairyj.2019.04.001
15. Fabris A., Biasioli F., Granitto P.M., Aprea E., Cappellin L., Schuhfried E., Soukoulis C., Märk T.D., Gasperi F., & Endrizzi I. (2010). PTR-TOF-MS and data-mining methods for rapid characterization of agro-industrial samples: influence of milk storage conditions on the volatile compounds profile of Trentingrana cheese. *Journal of Mass Spectrometry*, 45, 1065-1074. https://doi.org/10.1002/jms.1797

16. Faustini M., Quintavalle Pastorino G., Colombani C., Chiesa L. M., Panseri S., Vigo D., & Curone G. (2019). Volatilome in Milk for Grana Padano and Parmigiano Reggiano Cheeses: A First Survey. *Veterinary sciences*, 6(2), 41. https://doi.org/10.3390/vetsci6020041
17. Ferri C., Hernández-Orallo J., & Modroiu R. (2009). An experimental comparison of performance measures for classification, *Pattern Recognition Letters*, 30(1), 27-38, https://doi.org/10.1016/j.patrec.2008.08.010
18. Fox P.F., Guinee T.P., Cogan T.M., & McSweeney P.L.H. (2017). Cheese: Structure, Rheology and Texture. In: P.F. Fox, T.P. Guinee, T.M. Cogan, P.L.H. McSweeney (eds.) *Fundamentals of Cheese Science* (pp. 475-532). Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7681-9_14
19. Franciosi E., De Sabbata G., Gardini F., Cavazza A., & Poznanski E. (2011). Changes in psychrotrophic microbial populations during milk creaming to produce Grana Trentino cheese, *Food Microbiology*, 28(1), 43-51, https://doi.org/10.1016/j.fm.2010.08.003
20. Franciosi E., Gardini F., Monfredini L., Tabanelli G., Fabris A., Endrizzi I., Poznanski E., Gasperi F., & Cavazza A. (2012). Does milk treatment before cheesemaking affect microbial and chemical traits of ripened cheese? Grana Trentino as a case study, *Journal of Dairy Science*, 95(10), 5485-5494. https://doi.org/10.3168/jds.2011-4693
21. Giraffa G. 2021, The Microbiota of Grana Padano Cheese. A Review, *Foods*, 10(11), 2632. https://doi.org/10.3390/foods10112632
22. Kassambara A. (2020). ggpubr: 'ggplot2' Based Publication Ready Plots. R Package Version 0.4.0. Available online: https://CRAN.R-project.org/package=ggpubr (accessed on 10/04/2022).
23. Kassambara A. & Mundt, F. (2020). Factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R Package Version 1.0.7. https://CRAN.R-project.org/package=factoextra (accessed on 10/04/2022).

24. Khattab A. R., Guirguis H. A., Tawfik S. M., & Farag M. A. (2019). Cheese ripening: A review on modern technologies towards flavor enhancement, process acceleration and improved quality assessment, *Trends in Food Science & Technology*, 88, 343-360, https://doi.org/10.1016/j.tifs.2019.03.009
25. Kilcawley K. N., Faulkner H., Clarke H.J., O'Sullivan M.G., & Kerry J.P. (2018). Factors Influencing the Flavour of Bovine Milk and Cheese from Grass Based versus Non-Grass Based Milk Production Systems. *Foods*, 7(3), 37. https://doi.org/10.3390/foods7030037
26. Kuhn M. (2021). caret: Classification and Regression Training. R package version 6.0-90. https://CRAN.R-project.org/package=caret
27. Lazraq A., Cléroux R., & Gauchi J. P. (2003). Selecting both latent and explanatory variables in the PLS1 regression model, *Chemometrics and Intelligent Laboratory Systems*, 66(2),117-126. https://doi.org/10.1016/s0169-7439(03)00027-3
28. Liaw I.W., Miracle R.E., Jervis S.M., Listiyani M.A.D., & Drake M.A. (2011). Comparison of the flavor chemistry and flavor stability of Mozzarella and Cheddar wheys, *Journal of Food Science*, 76(8), 1188-1194, https://doi.org/10.1111/j.1750-3841.2011.02360.x
29. Liu, S.Q., Holland, R., & Crow, V.L. (2004). Esters and their biosynthesis in fermented dairy products: A review. *International Dairy Journal*, 14(11), 923-945. 10.1016/j.idairyj.2004.02.010.
30. Ma Y. & Barbano D. M. (2000). Gravity Separation of Raw Bovine Milk: Fat Globule Size Distribution and Fat Content of Milk Fractions, *Journal of Dairy Science*, 83(8), 1719-1727, https://doi.org/10.3168/jds.S0022-0302(00)75041-7
31. Marilley L. & Casey M.G. (2004). Flavours of cheese products: metabolic pathways, analytical tools and identification of producing strains, *International Journal of Food Microbiology*, 90(2) 139-159. https://doi.org/10.1016/S0168-1605(03)00304-0

32. McSweeney P.L.H. & Sousa M.J. (2000). Biochemical pathways for the production of flavour compounds in cheeses during ripening: A review, *Lait*, 80(3), 293-324. https://doi.org/10.1051/lait:2000127
33. McSweeney P.L.H., Ottogalli G., & Fox P.F. (2004). Diversity of cheese varieties: An overview. In: P.F. Fox, P.L.H. McSweeney, T.M. Cogan, T.P. Guinee (eds.), *Cheese: Chemistry, Physics and Microbiology* (pp. 1-23), Academic Press. https://doi.org/10.1016/S1874-558X(04)80037-X
34. Montgomery D. C. (2013). Randomized blocks, latin squares, and related designs. In D. C. Montgomery (Eds.), Design and Analysis of Experiments (pp. 139–182). John Wiley & Sons Inc
35. NIST/EPA/NIH (National Institute of Standards and Technology/ Environmental Protection Agency/ National Institutes of Health) (2014). NIST Mass Spectral Library. National Institute of Standards and Technology, US Secretary of Commerce, Washington, DC.
36. Panseri S., Luca M., Zecconi A., Soncini G., & Noni I. (2014). Determination of volatile organic compounds (VOCs) from wrapping films and wrapped PDO Italian cheeses by using HS-SPME and GC/MS. *Molecules*, 19(7), 8707-8724. https://doi.org/10.3390/molecules19078707
37. Pinto R. C., Trygg J., & Gottfries J. (2012). Advantages of orthogonal inspection in chemometrics. *Journal of Chemometrics*, 26, 231–235. https://doi.org/10.1002/cem.2441
38. Pisano M. B., Scano P., Murgia A., Cosentino S., & Caboni P. (2016). Metabolomics and microbiological profile of Italian mozzarella cheese produced with buffalo and cow milk, *Food Chemistry*, 192, 618-624. https://doi.org/10.1016/j.foodchem.2015.07.061.
39. Qian M. & Reineccius G. (2002). Identification of Aroma Compounds in Parmigiano-Reggiano Cheese by Gas Chromatography/Olfactometry, *Journal of Dairy Science*, 85(6), 1362-1369, https://doi.org/10.3168/jds.S0022-0302(02)74202-1

40. R Core Team. (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available online: https://www.R-project.org/ (accessed on 10/04/2022).

41. Ricci M., Gasperi F., Endrizzi I., Menghi L., Cliceri D., Franceschi P., & Aprea E. (2022) Effect of Dairy, Season, and Sampling Position on Physical Properties of Trentingrana Cheese: Application of an LMM-ASCA Model. *Foods*, 11(1), 127. https://doi.org/10.3390/foods11010127

42. Rossi F., Gatto V., Sabattini G., & Torriani S. (2012). An assessment of factors characterising the microbiology of Grana Trentino cheese, a Grana-type cheese. *International Journal of Dairy Technology*, 65(3), 401–409. https://doi.org/10.1111/j.1471-0307.2012.00844.x

43. Smilde A.K., Timmerman M.E., Hendriks M.M.W.B., Jansen J.J., & Hoefsloot H.C.J. (2012). Generic framework for high-dimensional fixed-effects ANOVA. *Briefings in Bioinformatics*, 13(5), 524–535. https://doi.org/10.1093/bib/bbr071

44. Suh J. H. (2022). Critical review: Metabolomics in dairy science – Evaluation of milk and milk product quality, *Food Research International,* 154, 110984. https://doi.org/10.1016/j.foodres.2022.110984.

45. Thevenot E.A., Roux A., Xu Y., Ezan E., & Junot C. (2015). Analysis of the human adult urinary metabolome variations with age, body mass index and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. *Journal of Proteome Research*. 14(8), 3322-3335. http://dx.doi.org/10.1021/acs.jproteome.5b00354

46. Trygg J. & Wold S. (2002). Orthogonal projections to latent structures (O-PLS). *Journal of Chemometrics*, 16, 119–128. https://doi.org/10.1002/cem.695

47. Villeneuve M.-P., Lebeuf Y., Gervais R., Tremblay G.F., Vuillemard J.C., Fortin J., & Chouinard P.Y. (2013). Milk volatile organic compounds and fatty acid profile in cows fed

timothy as hay, pasture, or silage. *Journal of Dairy Science*, 96(11), 7181-7194. https://doi.org/10.3168/jds.2013-6785.

48. Wang L. & Randolph H.E. (1978). Activation of Lipolysis. I. Distribution of Lipase Activity in Temperature Activated Milk, Journal of Dairy Science, 61(7), 874-880. https://doi.org/10.3168/jds.S0022-0302(78)83664-9

Table A1: repeatability index for different classes of molecules for each SPME fiber adopted in SPME/GC-MS analysis. Values are estimated from 12 repeated analysis on a reference cheese sample.

SPME fiber	Acids	Aldehydes	esters	ketones
1	14%	12%	6%	8%
2	27%	26%	15%	25%
3	37%	28%	10%	34%
4	22%	25%	6%	20%

Table A2: Overall values for SPME/GC-MS measurements for each volatile organic compound detected in Trentingrana Cheese expressed in µg/kg equivalent of i.s., each measurement has been carried out in triplicate. The values of the retention index NIST are obtained from NIST 14 database (NIST/EPA/NIH, 2014).

Compound Category	Compound Name	Minimum	Mean	Maximum	Retention Index Estimated	Retention Index NIST
Acids	acetic acid	30.21	222.45	633.28	1529	1449
	propionic acid	0.00	50.76	876.65	1594	1535
	butanoic acid	85.66	682.33	5040.80	1689	1625
	2-methyl butanoic acid	0.00	0.72	12.11	1759	1662
	hexanoic acid	78.00	870.81	9791.60	1914	1846
	heptanoic acid	0.00	5.44	85.49	2049	1950
	octanoic acid	25.84	232.10	3846.18	2127	2060
	nonanoic acid	0.00	7.74	51.94	2211	2171
	decanoic acid	5.87	48.29	678.66	2339	2276
	Total Acids mean	235.63				
	Total acids Standard deviation	472.30				
Alcohols	2-propanol	1.14	6.66	43.26	932	927
	ethanol	6.79	1030.07	4455.16	940	932
	2-butanol	0.00	6.75	230.39	1035	1025
	2-methyl, 1-butanol	6.19	35.68	159.91	1142	1119
	1-butanol	0.00	9.11	69.39	1164	1142
	3-methyl 1 butanol	0.00	13.35	72.21	1222	1209
	1-pentanol	0.00	1.36	6.85	1261	1250
	3-methyl, 3-buten 1-ol	2.11	9.14	24.43	1261	1248
	2-heptanol	0.00	10.08	70.88	1330	1320
	prenol	1.27	5.06	11.31	1332	1320
	hexanol	0.00	6.48	66.12	1364	1355
	2-ethyl hexanol	0.00	1.00	10.40	1501	1491
	2-nonanol	0.00	0.74	35.89	1531	1521
	Total alcohols mean	87.34				
	Total alcohols Standard deviation	522.54				
Aldehydes	2-methyl butanal	1.47	6.92	22.06	919	914
	3-methyl butanal	5.45	23.81	61.35	922	918
	hexanal	0.00	2.20	7.31	1100	1083
	3-methyl 2-butenal	0.00	0.29	2.06	1212	1215
	nonanal	0.00	1.71	19.78	1402	1391
	decanal	0.00	0.38	3.29	1510	1498
	benzaldehyde	0.00	3.96	36.74	1533	1520
	phenyl acetaldehyde	0.73	4.24	14.55	1654	1640
	Total Aldehydes mean	5.44				

	Total aldehydes standard deviation	8.61				
Esters	ethyl acetate	1.27	21.13	110.65	900	888
	ethyl propanoate	0.00	4.70	145.03	965	953
	isopropyl isobutanoate	0.00	32.12	188.59	970	959
	ethyl butanoate	15.96	299.17	1668.76	1049	1035
	2-methyl ethyl butanoate	0.00	0.01	0.50	1062	1051
	butyl acetate	0.00	0.87	12.00	1093	1074
	ethyl valerate	0.00	1.53	9.76	1147	1134
	butyl butanoate	0.00	0.95	20.41	1229	1220
	ethyl hexanoate	6.04	240.55	1777.48	1244	1233
	isoamyl butanoate	0.00	0.77	9.86	1273	1259
	butyl pentanoate	0.00	0.00	0.19	1324	1310
	propyl hexanoate	0.00	0.13	8.98	1329	1316
	isopentyl hexanoate	0.00	0.00	0.31	1423	1451
	butyl hexanoate	0.00	0.07	2.79	1423	1408
	ethyl octanoate	0.44	21.28	203.50	1444	1435
	2-hydroxy, 4-methyl, methyl pentanoate	0.00	0.36	9.16	1481	1513
	ethyl decanoate	0.00	4.17	32.61	1649	1638
	Total esters mean	26.16				
	Total esters standard deviation	107.10				
Hydrocarbons	(E) 3-octene	0.00	0.97	4.65	885	850
	ethyl benzene	0.00	0.94	18.64	1135	1129
	p-xylene	0.00	0.36	5.35	1143	1138
	m-xylene	0.00	0.88	14.04	1148	1143
	Total hydrocarbons mean	0.63				
	Total hydrocarbons standard deviation	1.39				
Ketones	2-propanone	3.30	29.61	80.99	882	819
	2-butanone	1.42	13.73	448.25	909	907
	2-pentanone	20.72	117.10	460.19	986	981
	2-hexanone	0.00	6.18	15.84	1099	1083
	3-heptanone	0.00	2.56	12.87	1165	1161
	2-heptanone	87.21	240.30	545.57	1194	1182
	2-octanone	0.00	0.54	6.68	1294	1287
	acetoin	0.00	5.19	60.71	1295	1284
	2-nonanone	7.64	27.49	69.51	1397	1390
	2-undecanone	1.47	5.37	11.76	1608	1598
	acetophenone	0.00	0.36	6.24	1661	1647
	Total ketones mean	37.37				

	Total ketones standard deviation	77.97				
Lactones	butanolactone	0.00	0.71	7.28	1637	1632
	delta decalactone	0.96	3.12	6.88	2162	2194
	Total lactones mean	1.92				
	Total lactones standard deviation	1.64				
Phenols	phenol	0.54	1.37	2.63	2021	2000
	4-methyl phenol	0.23	1.03	10.60	2095	2080
	3-methyl phenol	0.75	5.24	18.66	2103	2091
	Total phenols mean	2.55				
	Total phenols standard deviation	2.74				
Pyrazines	2, 6 dimethyl pyrazine	0.00	5.61	21.84	1336	1328
Sulfurate Compounds	methanthiol	0.00	1.24	6.97	866	692
	carbon disulfide	0.00	2.86	31.88	870	735
	dimethyl sulfide	0.53	3.77	12.90	872	754
	dimethyl disulfide	0.00	1.14	5.84	1089	1077
	dimethyl sulfone	0.00	1.75	6.43	1911	1903
	Total sulfurate mean	2.15				
	Total sulfurate standard deviation	2.20				
Terpenes	alpha thujene	0.00	1.39	23.56	1026	1028
	limonene	0.00	4.02	189.39	1201	1200
	Total terpenes mean	1.80				
	Total terpenes standard deviation	10.08				

Figure 1

Figure 1: Results of ASCA decomposition for the factor Dairy Factory.

In the plot A are reported the score values for each level are labeled from C-1 to C-15 for each dairy factory and they are represented by blue, orange, or black dots according to the milk collection procedure they adopt (respectively blue for DMC, yellow for SMC, and black for MMC). In plot B are reported loading values are represented using dark-gray text, only VOCs significantly different are represented**.**

Figure 2

Figure 2: Boxplot reporting values of the three significantly different volatiles compounds for the factor Time.

Values are centered for the mean value of each batch of SPME fiber adopted. Each dot corresponds to a single measurement, the bold black line represents the median value, upper and lower margins of the blocks indicate the limit of the second and third quartile respectively, and whiskers indicate upper confidence intervals at 95%. The dashed red line represents the overall median for each compound. Letters shows the groups estimated from post-hoc pairwise comparison tests.

Figure 3

Figure 3: Violin plot reporting the comparisons between the distributions of the model performance on repeated partitions and the null distribution of the permutation test.

Results of the Wilcoxon statistical test are reported in the box of text above each parameter, each dot represents a single measurement, null distribution and multiple partitions distribution are colored in blue and red respectively.

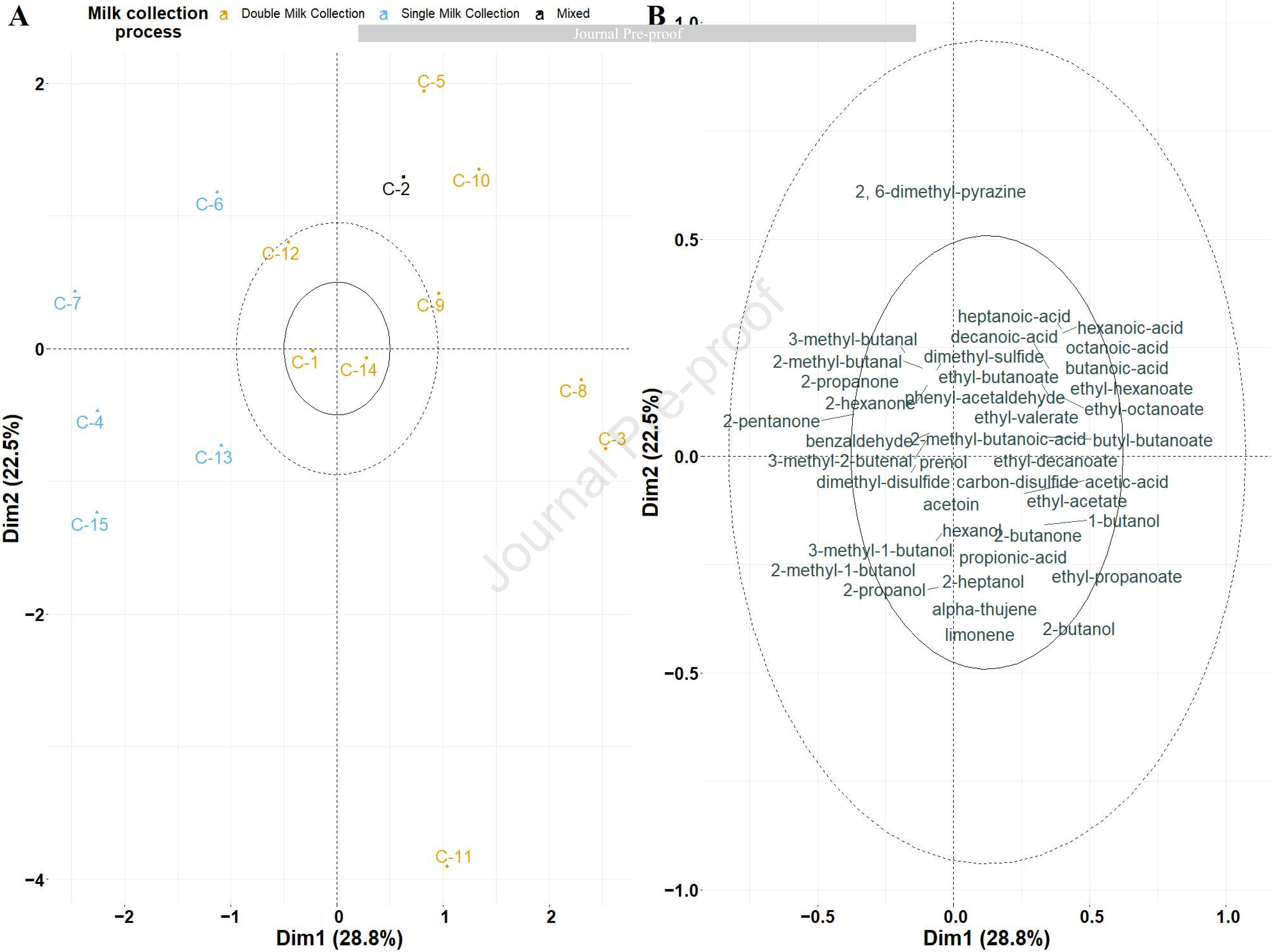
Figure 4

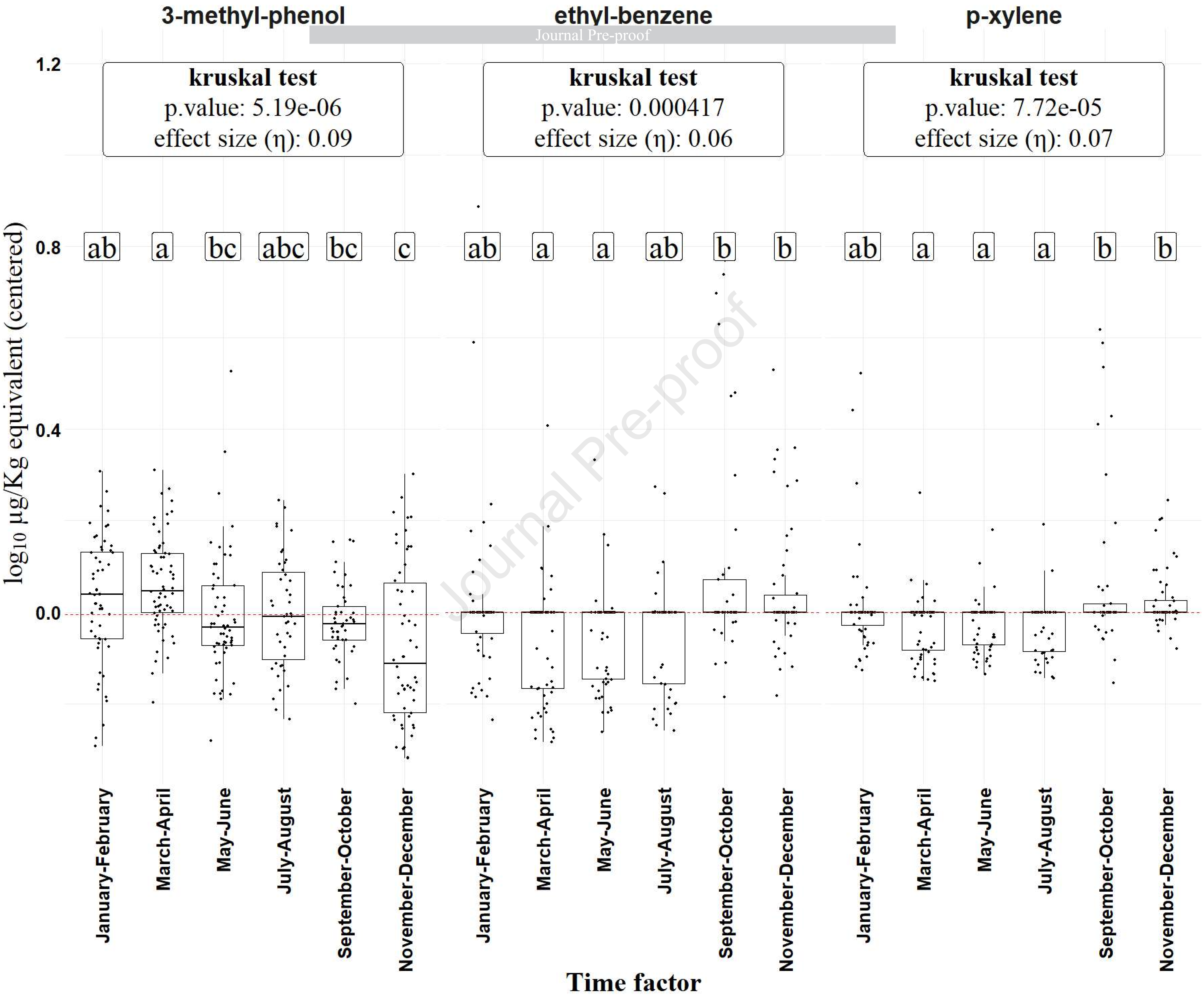
Figure 4: Barplot showing the absolute value of the significant regression coefficients of the O-PLS-DA model validated using bootstrap.

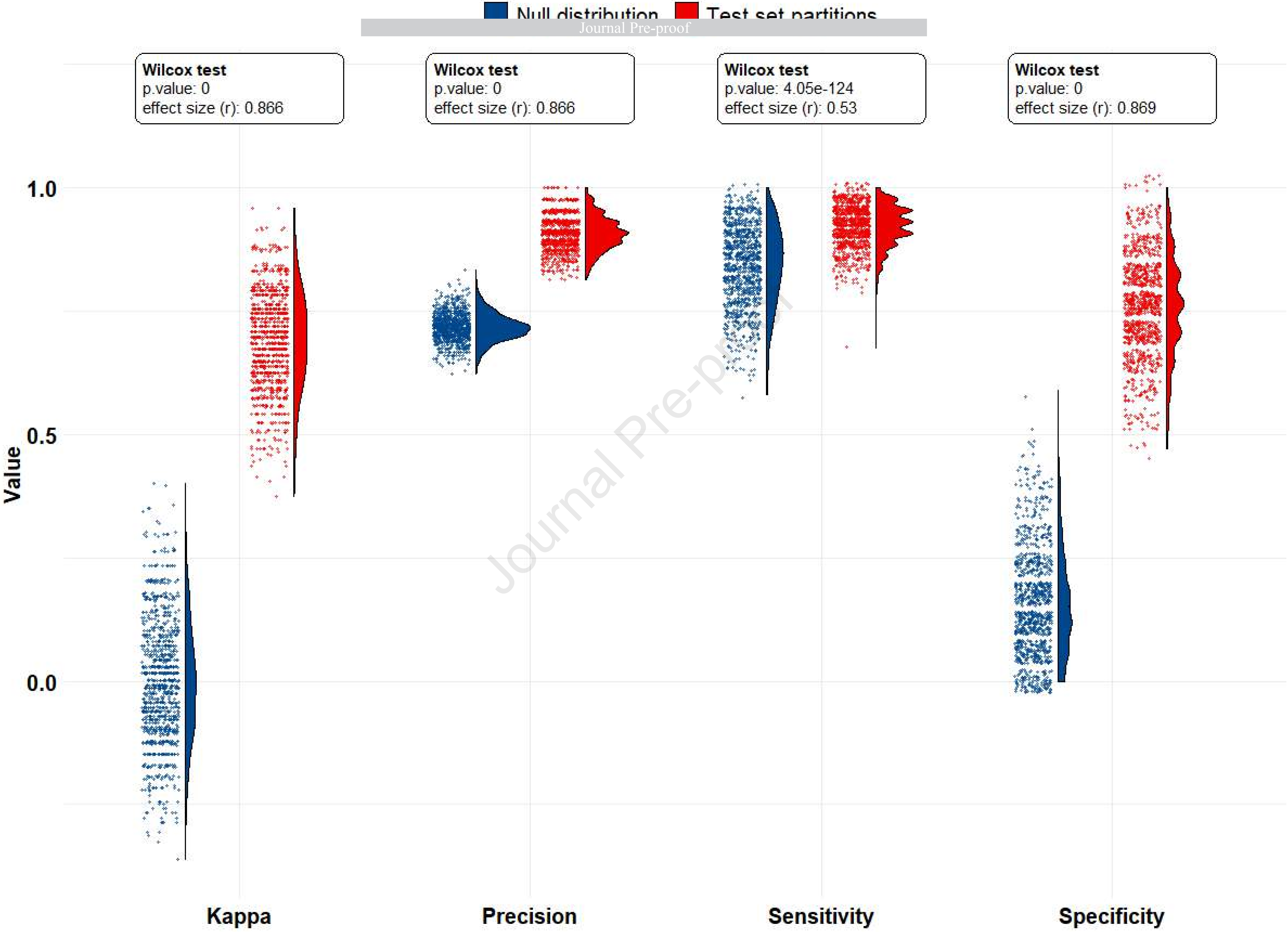
Values reported in light-blue bars are referred to the Single Milk Collection procedure (SMC) and values in orange bars are referred to the Double Milk Collection procedure (DMC). Bars show the absolute values of the coefficients of the model, dashed error bars report the confidence intervals estimated.

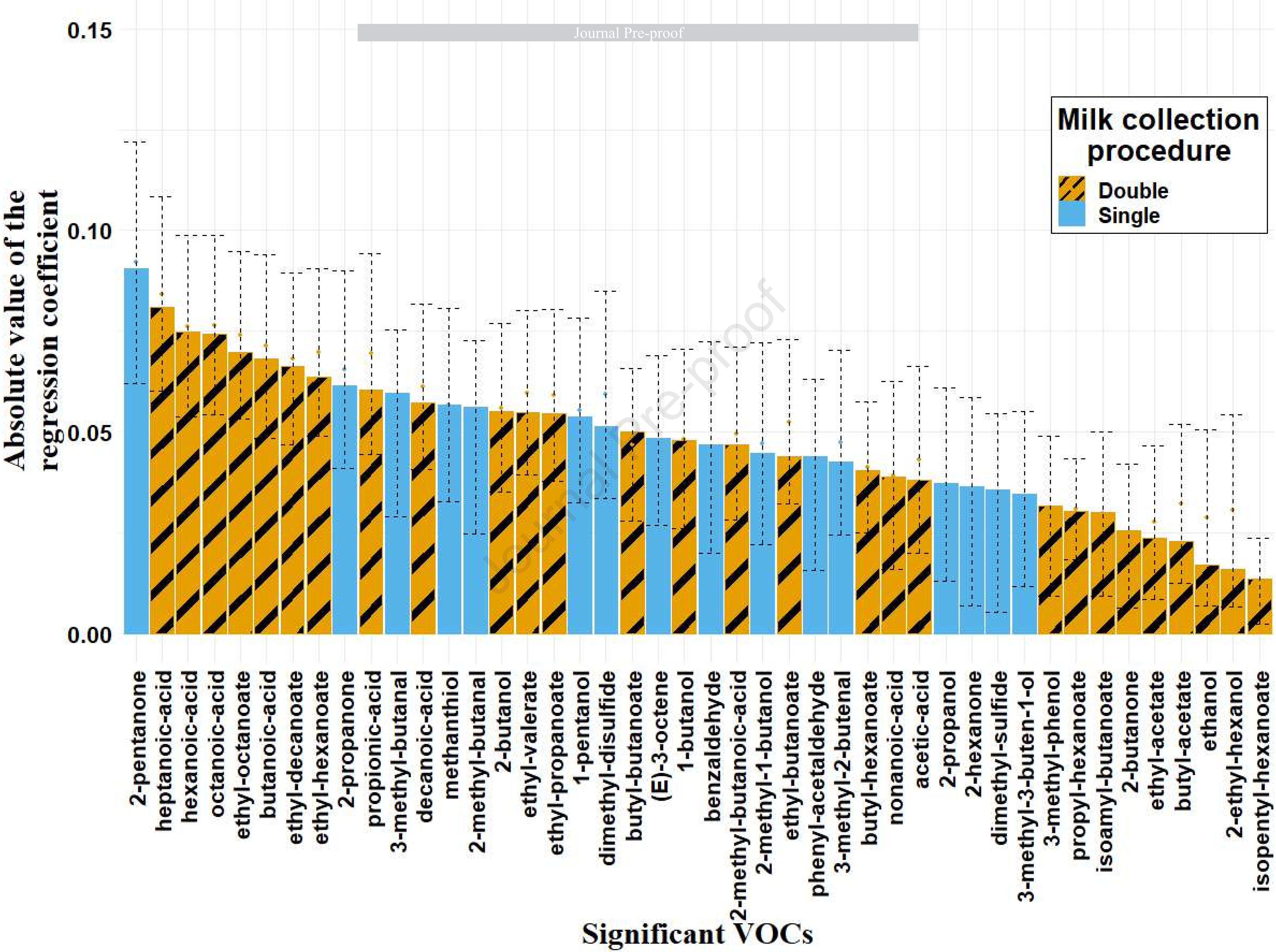
Table 1: Production traits across dairy factories of the Trentingrana consortium

Dairy Factory	Number of farms associated to the dairy	Percentage of farms using unifeed alimentation procedure	Number of cheese wheels produced during production year 2015/16	Number of cheese wheels produced during production year 2016/17	Milk collection procedure adopted
C-1	17	58.8	14476	14929	Double
C-2	18	16.7	1207	1534	Mixed
C-3	51	3.9	9882	10415	Double
C-4	38	26.3	11881	10537	Singular
C-5	10	30.0	6145	6296	Double
C-6	59	6.8	3010	2387	Singular
C-7	38	10.5	6695	7390	Singular
C-8	47	4.3	5286	5900	Double
C-9	46	0.0	5119	5407	Double
C-10	12	8.3	1985	3721	Double
C-11	78	1.3	7072	7795	Double
C-12	95	1.1	7961	8609	Double
C-13	9	0.0	5137	5028	Singular
C-14	72	2.8	8287	9183	Double
C-15	27	0.0	2635	1801	Singular


A
Milk collection process
a Double Milk Collection
a Single Milk Collection
a Mixed
C-5
C-2
C-10
C-6
C-12
C-7
C-9
C-1
C-14
C-8
C-4
C-3
C-13
C-15
C-11
2
0
−2
−4
−2
−1
0
1
2
Dim1 (28.8%)
Dim2 (22.5%)
B
1.0
0.5
0.0
−0.5
−1.0
−0.5
0.0
0.5
1.0
Dim1 (28.8%)
Dim2 (22.5%)
2, 6-dimethyl-pyrazine
heptanoic-acid
hexanoic-acid
3-methyl-butanal
decanoic-acid
octanoic-acid
2-methyl-butanal
dimethyl-sulfide
butanoic-acid
2-propanone
ethyl-butanoate
ethyl-hexanoate
2-hexanone
phenyl-acetaldehyde
ethyl-octanoate
2-pentanone
ethyl-valerate
benzaldehyde
2-methyl-butanoic-acid
butyl-butanoate
3-methyl-2-butenal
prenol
ethyl-decanoate
dimethyl-disulfide
carbon-disulfide
acetic-acid
acetoin
ethyl-acetate
1-butanol
hexanol
2-butanone
3-methyl-1-butanol
propionic-acid
2-methyl-1-butanol
ethyl-propanoate
2-propanol
2-heptanol
alpha-thujene
limonene
2-butanol


3-methyl-phenol
ethyl-benzene
p-xylene
kruskal test
p.value: 5.19e-06
effect size (η): 0.09
kruskal test
p.value: 0.000417
effect size (η): 0.06
kruskal test
p.value: 7.72e-05
effect size (η): 0.07
ab
a
bc
abc
bc
c
ab
a
a
ab
b
b
ab
a
a
a
b
b
1.2
0.8
0.4
0.0
$\log_{10}$ μg/Kg equivalent (centered)
January-February
March-April
May-June
July-August
September-October
November-December
January-February
March-April
May-June
July-August
September-October
November-December
January-February
March-April
May-June
July-August
September-October
November-December
Time factor


Null distribution
Test set partitions
Wilcox test
p.value: 0
effect size (r): 0.866
Wilcox test
p.value: 0
effect size (r): 0.866
Wilcox test
p.value: 4.05e-124
effect size (r): 0.53
Wilcox test
p.value: 0
effect size (r): 0.869
1.0
0.5
0.0
Value
Kappa
Precision
Sensitivity
Specificity


Absolute value of the regression coefficient
0.15
0.10
0.05
0.00
Milk collection procedure
▨ Double
■ Single
2-pentanone
heptanoic-acid
hexanoic-acid
octanoic-acid
ethyl-octanoate
butanoic-acid
ethyl-decanoate
ethyl-hexanoate
2-propanone
propionic-acid
3-methyl-butanal
decanoic-acid
methanthiol
2-methyl-butanal
2-butanol
ethyl-valerate
ethyl-propanoate
1-pentanol
dimethyl-disulfide
butyl-butanoate
(E)-3-octene
1-butanol
benzaldehyde
2-methyl-butanoic-acid
2-methyl-1-butanol
ethyl-butanoate
phenyl-acetaldehyde
3-methyl-2-butenal
butyl-hexanoate
nonanoic-acid
acetic-acid
2-propanol
2-hexanone
dimethyl-sulfide
3-methyl-3-buten-1-ol
3-methyl-phenol
propyl-hexanoate
isoamyl-butanoate
2-butanone
ethyl-acetate
butyl-acetate
ethanol
2-ethyl-hexanol
isopentyl-hexanoate
Significant VOCs

Highlights

- ASCA allows removing the effect of a potential confounding factors
- Specific groups of cheese VOCs are associated with sources of variation in dairies
- OPLSDA estimated the effect of milk collection procedure on VOCs removing data noise
- Milk collection procedure affects esters formation in ripened cheese

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.