Innovative and thorough practice to certify reference materials for sensory defects of olive oil

Fidel Ortega-Gavilán [a,*], José Antonio García-Mesa [b], José Carlos Marzal-Fernández [c], Francisco J. Moreno-Ballesteros [d], Francisco P. Rodríguez-García [b], Antonio González-Casado [a], Luis Cuadros-Rodríguez [a]

[a] Department of Analytical Chemistry, Faculty of Science, University of Granada, C/ Fuentenueva s/n, E-18071 Granada, Spain
[b] Institute of Agricultural and Fishery Research & Training (IFAPA) "Venta del Llano", Crta. Nacional Bailén-Motril Km 18.5, E-23620 Mengíbar, Jaén, Sapin
[c] Foundation for the Promotion and Development of the Olive Grove and Olive Oil from Andalusia, GEOLIT Technological Park, C/ Sierra Morena 13a, E-23620 Mengíbar, Jaén, Spain
[d] Regulatory Council of the Protected Designation of Origin "Sierra de Segura", Industrial Park "Llanos de Armijo" s/n, E-23360, La Puerta de Segura, Jaén, Spain

ARTICLE INFO

ABSTRACT

An important problem in the olive sector is the occasional mismatch of results obtained by different tasting panels when the same olive oil sample is analysed. These discrepancies could be minimised by using reference materials (RM) for taster training. A comprehensive protocol based on the combined use of sensory and instrumental analysis for the certification of olive oil batches as RMs, developed within the framework of the project 'Operational Group INTERPANEL', is proposed. Similarity indices (R^2, $\cos\theta$ and NEAR) applied on GC–MS fingerprints, allow a successful homogeneity and stability assessment of produced batches. Furthermore, the use of robust statistics combined with a set of instructions developed to remove outliers were applied with excellent results on sensory data set provided by supra-panel composed by more than 100 qualified tasters. This work is the first to provide a comprehensive protocol for certification of real olive oil samples as RM for sensory analysis.

1. Introduction

Olive oil is one of the major characteristic foods of the Mediterranean diet, with a high added value for the consumer due to its high health benefits (Willett et al., 1995; Barjol, 2013). Indeed, the olive oil obtained by milling olive fruits is the only foodstuff that must be subjected to sensory analysis to verify the proper labelling of the two highest quality commercial categories: extra virgin olive oil (EVOO) and virgin olive oil (VOO) (Mariotti, 2014). The analytical methods employed for the physicochemical and organoleptic assessment of olive oils are detailed described in the updated version of Regulation (ECC) No 2568/91 (European Commission, 2016). The intensity of particular sensory attributes is determined by a tasting panel which is related to a group of tasters, who have been specifically selected and trained, and who gather to perform the sensory analysis of a product under controlled conditions (COI/T.20/Doc. No 4 / Rev.1, 2007; COI/T.20/Doc. No 14 / Rev. 7, 2021). The 'Panel Test' is a sensory method proposed by the

International Olive Council (IOC) to classify olive oil attending to their organoleptic characteristic measured by a panel of 8–10 tasters who works as a multisensor analytical instrument (COI/T.20/Doc. No 15 / Rev.10, 2018). This is a robust and statistically reliable method, but suffers from a certain lack of reproducibility in the results obtained by different panels due to the inherent subjectivity in the taster (Morales et al., 2013). One part of the industrial sector claims a solution to the differences among results, which have a negative effect, thus causing significant economic losses (Aparicio Ruiz, Morales, & Aparicio, 2019; Circi et al., 2017). In fact, this concern has been highlighted in a recent report requested by the European Commission (European Commission, 2020) in order to seek the origin of the problem and the solutions needed to improve the organoleptic evaluation of VOO and EVOO as conformity control systems. Despite the fact that some European Member States assume that the organoleptic assessment should be excluded from the analysis set carried out on VOO according current legislation, the above-mentioned report concludes that the organoleptic assessment should be

continued, albeit improved. The same report proposes the following solutions: (i) the development of reference samples for olive oil taste training, and (ii) the organization of regular tasting sessions with tasters from different countries. These coincide with the solutions proposed by several authors (COI/T.20/Doc. No 14 / Rev. 7, 2021; Aparicio Ruiz, Morales, & Aparicio, 2019; Aparicio, Conte & Fiebig, 2013) which consider the use of certified reference materials as the best way to harmonise tasting panels, as well as tasters from the same panel, which translates in improved results of the analysis.

A certified reference material (CRM) is defined as *a sufficiently homogeneous and stable material with respect to one or more specified properties, which has been established to be fit for intended use in a measurement process and is accompanied by a certificate that provides the value of the specified property, its associated uncertainty, and a statement of metrological traceability* (ISO Guide 30, 2015). According to ISO Guide 35 (ISO Guide 35, 2017) a reference material must be subjected to three experimental studies before being certified. The first study involves the homogeneity assessment of the produced RM batch. The packaging process is one of the most critical stages in RM production, so it is necessary to verify that the values of one or several targeted properties are similar enough regarding all packaged units of the batch. Furthermore, a stability study is necessary to ensure that the targeted property or properties remains unchanged over the time. RMs produced from biological matrices as perishable foodstuffs which may evolve over time, the stability study is carried out at the same time as the RM is being use. Finally, the characterisation study determines the value of one or several targeted properties or attributes that is used as the reference value. The results of these studies are gathered in the RM certificate, in which the certified value as well as the uncertainty value involved in the whole certification process are specified.

Regarding the experimental procedure, the characterisation study of organoleptic attributes should be performed by sensory analysis, but homogeneity and stability studies might be carried out using instrumental analysis (Rason, Laguet, Berge, Dufour & Lebecque, 2007; Valverde Som et al., 2018b). The chromatographic techniques coupled to different signal acquisition systems, *e.g.* gas chromatography coupled with flame ionization detector (GC-FID) or mass spectrometry (GC–MS), allow to study in an unspecific way the signal profile of the volatile organic compounds (VOCs) fraction responsible of the food flavour. Some recent studies use GC-FID to obtain VOC chromatographic fingerprints of olive oils in order to assess the similarity among these analytical signals by means of similarity indices (Valverde Som et al., 2018b; Valverde Som et al., 2018a; Ortega Gavilán, Valverde Som, Rodríguez García, Cuadros Rodríguez, & Bagur González, 2020). A high level of similarity can be assumed when no differences between successive fingerprints are found.

Nevertheless, alternative homogeneity studies use the intensity of an instrumental signal, from a specific chemical marker, to monitor the between-samples variability. This approach has been applied within the OLEUM project (www.oleumproject.eu/), a project funded by the EU's Horizon 2020 programme, for homogeneity assessment in samples intended for physical–chemical analysis (OLEUM Project, 2016), although this strategy could also be applied in sensory analysis if a suitable flavouring is used as a reference marker.

The following approaches may be considered in the production of reference materials for sensory analysis of virgin olive oil:

- Design of formulation types to prepare olive oils with a certain organoleptic attribute and with a precise intensity by adding flavourings –substances which cause a specific flavour– on an olive oil free of any perceptible negative attribute, called base. The previous knowledge of the corresponding odour threshold (or recognition threshold) of the substance to be used is always required. This is the strategy being addressed in the framework of OLEUM project (Barbieri et al., 2020). At least two artificial olfactory reference materials for sensory analysis of olive oil

having rancid and winey-vinegary defects have been produced within the project and their characteristics and production details discussed in Aparicio Ruiz et al. (2020).

- Use of olive oil types already produced with the required characteristics or production of these olive oil types from olive fruits or olive oils by applying controlled degradation processes. These olive oils, called concentrates, have a high intensity of a negative attribute so that the final products are obtained by mixing them with a compatible base in a certain percentage. The materials developed following this technique simulate the possible defects present in the olive oil much better than the previous approach. However, this strategy requires the search, development and certification of new materials in each campaign.

This paper describes a certification practice of olive oil reference materials for sensory analysis, which have been developed by applying the second approach previously described within the framework of the project 'Reference and harmonization system of virgin olive oil tasting panels from Andalusia – INTERPANEL' (www.interpanel.es; https://ec.europa.eu/eip/agriculture/en/find-connect/projects /sistema-de-referencia-y-armonización-de-paneles-de). The protocol described in this paper involves the combined application of instrumental techniques with sensory analysis to achieve the certification. The homogeneity and stability studies are based on the analysis of volatile organic compounds (VOCs) by gas chromatography coupled to mass spectrometry (GC–MS), while the characterisation study is achieved after the application of the 'Panel Test' method proposed by the IOC. The main novelty of this work is the development and application of an innovative approach which is applied on the characterisation study. It is based on several quality filters to detect and remove outlier values, combined with the use of robust statistics applied to the sensory data provided by a supra-panel consisting of more than a hundred tasters which belong to different laboratories.

2. Materials and methods

2.1. Olive oil materials

A total of eight olive oil batches, each with a total volume of 50 L, were produced in this work. The materials were designed to be representative of the main attributes and defects (fruity, fusty/muddy, musty and winey-vinegary) proposed by the IOC. The different olive oil batches were produced at the Institute of Agricultural and Fishery Research & Training (IFAPA) "Venta del Llano" using a "Il Molinetto" compact plant (Pieralisi, Jesi, Italy, http://www.pieralisi.com/Viewdoc?co_id = 373). The only exception was the wine/vinegary batch, which was provided by the Regulatory Council of the Protected Designation of Origin "Sierra de Segura". Only three of the five attributes or defects considered in this study were prepared at two levels of decreasing intensity symbolised by 2 (medium intensity) and 1 (low intensity), by blending an attribute-concentrated oil with a dilution base. The official agri-food quality control laboratories of Córdoba and Granada (Department of Agriculture, Livestock, Fisheries and Sustainable Development, Regional Govern of Andalusia, Spain) were involved in the setting and selection of the different mixing percentages. The main characteristics and composition are described in Table SM1 (in Supplementary Material).

Note that the material coded as PIC/1 was used as dilution base for materials with fusty/muddy and musty defect.

Each olive oil material was produced in a single batch and PET-amber bottled in units of 200 mL. Both the first and last five bottles from each material batch were then discarded. In order to evenly cover the entire batch, ten representative units were evenly sampled and recoded with numbers from 1 to 10 according to the order of packaging from each olive oil material batch, obtaining a set sample of 80 bottles which were reserved for the homogeneity study. In addition, twelve

consecutive bottles of each olive oil material batch were also reserved for a quarterly analytical control of the material stability once the batch has been declared homogeneous. Finally, three additional bottles of each material batch were set aside to further perform the characterisation of the materials. All batches were bottled in July 2019 and were kept in the dark and below 0 °C until use, in order to ensure the volatile profile stability. After the homogeneity assessment, all samples were stored in the dark at 7–8 °C, in order to keep the typical storage conditions in a real sensory analysis laboratory.

2.2. Standard solutions

A mixture of *n*-alkanes standard solution from C7 to C40 (1 mg mL^{-1} in hexane, Supelco Analytical, Bellefonte, PA, USA) was purchased. A working external standard mix solution of 100 μg mL^{-1} was prepared by dilution of the commercial mixture in *n*-hexane.

Furthermore, a mixture solution of fluorobenzene (99%; Sigma-Aldrich, St Louis, MO, USA) and 4-fluorobiphenyl (95%, Sigma-Aldrich, St Louis, MO, USA) of 0.1 and 0.5 mg g^{-1} respectively, prepared in recent refined olive oil, was used as internal standard.

2.3. Chromatography analysis

The chromatographic analysis of the samples was carried out using a Varian 3800 GC (Varian Inc., Palo Alto, CA, USA) equipped with a split/splitless injector coupled to a Varian 4000 ion trap mass spectrometer (Varian Inc., Palo Alto, CA, USA). The instrument includes an autosampler system (CombiPal; CTC Analytics, Zwingen, Switzerland). The analytical column was a capillary column covered with polyethylene glycol (Agilent HP-INNOWAX; 30 m × 0.25 mm internal diameter × 0.25 μm film thickness; maximum temperature 260 °C, Agilent Technologies, Santa Clara, CA, USA).

As previous step, the VOCs were extracted using a 50/30 μm divinylbenzene/carboxen/polydimethylsiloxane StableFlex 24 Ga fibre (Supelco Analytical, Bellefonte, PA, USA). 0.100 g of the IS mixture solution was accurate weighted in a 20 mL headspace vial; then 4.900 g of olive oil RM was carefully added until that the final mass was 5.000 g. The samples were incubated in a shaker oven at 40 °C for 2 min and the extraction was carried out for 30 min. Fibre thermal desorption into GC was conducted at 250 °C for 2 min in splitless injection mode, and then the fibre was cleaned into the GC with a 1:100 split ratio for 6 min. The column oven temperature of 80 °C was held for 2.5 min, then programmed at 40 °C min^{-1} to 240 °C, then this last temperature was kept for 7.5 min. Helium (99.999%; Air Liquid) was used as the carrier gas (flow rate 1.5 mL min^{-1}). The mass spectrometric conditions were as follow: the ion source temperature was held isothermally at 230 °C during the chromatographic separation and the transfer-line temperature at 250 °C was maintained through the analysis. The ionization electron energy was 70 eV and the emission current 10 μA. The chromatograms were recorded in full-scan mode and the average spectra were acquired in the *m/z* range of 10–400.

Chromatographic analysis was carried out for both homogeneity and stability studies. For homogeneity assessment, which was carried out in October 2019, the first unit from each homogeneity sample set was chosen as reference and analysed in triplicate, whereas the remaining nine units were analysed once. The stability study is carried out in parallel with the use of each reference material by the different laboratories. Once the homogeneity of the material has been verified and the certified property has been characterised, the stability study is carried out. Every three months, one unit of each material is taken from the set of samples reserved for this study. Each unit is analysed in triplicate to obtain the representative fingerprint of the month concerned. In order to control the performance of the instrument, 0.100 mg g^{-1} IS solution was analysed at the beginning and end of each chromatographic run using the same operational conditions.

2.4. Sensory analysis

Sensory analysis was carried out throughout November 2019, by applying the EU 'Panel Test' method for the characterisation study of RM candidates. A total of 22 laboratories, where each panel was composed of a group of 8–10 tasters trained in accordance with IOC guidelines (COI/T.20/Doc. No 14 / Rev. 7, 2021), participated in the characterisation study. This group consisted of 19 private laboratories authorised by the Regional Govern of Andalusia for quality control and 3 official ones, all of them located in Andalusia with the exception of an official laboratory located in Madrid.

The results obtained in each tasting session were provided through the panels' intercommunication platform called INTERPANEL, on which the statistical data treatment was also performed.

2.5. Data treatment

2.5.1. Chromatographic data treatment

MS Workstation (version 6.9) software data system (Varian Inc., Palo Alto, CA, USA) was used for chromatographic data acquisition and recording. The chromatograms obtained are firstly transformed in data vectors, each one representing the individual fingerprint of each selected material units for the involved study. In order to obtain the VOC chromatographic fingerprints, each chromatogram was exported to netCDF format using the OpenChrom® version 1.3 software (Lablicate & Scientific community, Hamburg, Germany), available free at www.lablicate.com/.

Afterwards, every chromatogram was pre-processed in order to: (i) remove the lower and upper range values of each fingerprint that did not contain any useful information; (ii) filter and smooth using a Savitzky-Golay least-squares polynomial filter with a window size of 5 points and a second-order polynomial for noise removing; (iii) correct the baseline drifting; (iv) align time shifting and warping with respect to the mix external standard solution; and (v) normalise the intensities by using a suitable internal standard as reference. Data pre-processing was performed using MATLAB, version 9.3 (Mathworks Inc., Natick, MA, USA) and an *ad-hoc* home-programmed MATLAB script, so-called 'GCnostic' (version 1), that runs a number of algorithms of MATLAB Bioinformatics Toolbox, and a specific algorithm for the alignment of chromatographic peaks based on Kovats indices (d'Acampora Zellner et al., 2008; Cuadros Rodríguez et al., 2021), instead of applying a self-alignment as usual. This function has a new alignment algorithm incorporated to correct the shifting and warping in retention time of chromatographic fingerprint observed both in the same or different sequences. For this purpose, the retention time values are transferred to a new empirical retention index obtained from the analysis of *n*-alkane mix used as external standard analysed at the beginning and the end of each chromatographic run.

After pre-processing, similarity indices were then calculated by comparing the fingerprint data vectors directly with MATLAB's capabilities. These indices were used to obtain reliable metrics for homogeneity and stability assessing (Valverde Som et al., 2018a; Ortega Gavilán, Valverde Som, Rodríguez García, Cuadros Rodríguez, & Bagur González, 2020).

In addition, an exploratory data analysis was performed in order to evaluate the natural grouping of the homogeneity samples analysed by GC–MS by the means of PLS_Toolbox software version 8.6.1 (Eigenvector Research Inc., Manson, WA, USA). Furthermore, a mean centring was applied on the fingerprint matrix prior to chemometric treatment.

2.5.2. Sensory analysis data treatment

The statistical data processing of the sensory analysis for the characterisation of materials was carried out through the INTERPANEL platform. Based on experience gained by the authors in processing data from sensory analysis, robust statistics were programmed and applied for the estimation of reference values. For each material, the arithmetic

mean and some robust estimations of the central value (median and Huber's average H15), and the corresponding standard deviations (SD) (interquartile range SD$_{IQR}$, equivalent median absolute deviation (MADe) and Huber's SD(H15)) were calculated in order to establish the characterisation value (Thompson and Lowthian, 2011). Furthermore, in order to identify and remove inconsistent biased values that may affect the quality of the results; the following quality statistical filters were designed and applied:

- First filter: rejecting the panels whose robust coefficient of variation (CV$_{IQR}$), obtained by dividing the interquartile range standard deviation (SD$_{IQR}$) by the median, exceeds 20% (COI/ T.20 / Doc. No 22, 2005).
- Second filter: rejecting inconsistent panels that have passed the first filter by means of the robust z*-score (z*$_p$), using the Equation (1).

$$z^*_p = \frac{m_p - m_{overall}}{MADe_{overall}} \quad (1)$$

where m$_p$ is the median obtained by the panel p for the attribute under consideration, m$_{overall}$ is the median of the panel medians and MADe$_{overall}$ is the equivalent median absolute deviation taking into account all panels, both for the same attribute. A tasting panel is considered inconsistent if the corresponding z*–score has an absolute value greater than 2; in this case, the panel is also removed of the study. Once this filter is applied, a single supra-panel is conformed, including all the tasters that belong to the non-eliminated panels.

- Third filter: rejecting inconsistent values of sensory data provided by each taster within the supra-panel. The statistical test is again based on a robust z$_t$*(H15)–score, specific to each taster t, which is calculated using the robust Huber's H15 estimators applying the following Equation (2).

$$z^*_t(H15) = \frac{x_t - Av_{sp}(H15)}{SD_{sp}(H15)} \quad (2)$$

where x$_t$ is the organoleptic intensity observed by the taster t, and Av$_{sp}$(H15) and SD$_{sp}$(H15) are the estimated average and standard deviation Huber's H15 parameters of the organoleptic intensities for the attribute under consideration from the supra-panel results. Again, a single sensory datum is considered inconsistent when the corresponding z*$_t$(H15) has an absolute value greater than 2; in this case, the datum is removed.

The results are considered statistically reliable if the number of panels removed by the first and second filters does not exceed 25% of the total number of participants. Regarding the third filter, the number of values removed should not exceed 20% of the total number of supra-panel tasters. If any of these limits are exceeded, the material is directly qualified as 'non-certifiable'.

After the application of the filters, the suitability of the candidate materials to be certified is checked depending on the value of SD$_{sp}$(H15) value. Three possible settings are covered:

- (i) SD$_{sp}$(H15) has a value lower than 1. The material under study is fitted for certification purpose.
- (ii) SD$_{sp}$(H15) has a value between 1 and 2.5. There is sufficient consensus among the tasters to establish a reference value, but not enough to be certified. The material is then labelled as 'consensus guiding material'.
- (iii) SD$_{sp}$(H15) has a value higher than 2.5. Due to the high data scattering, the candidate material is qualified as 'non-certifiable'.

The reference value assigned to the organoleptic intensity of the attribute under consideration for each material is defined by the Av$_{sp}$(H15) estimation from the results provides by the valid tasters that
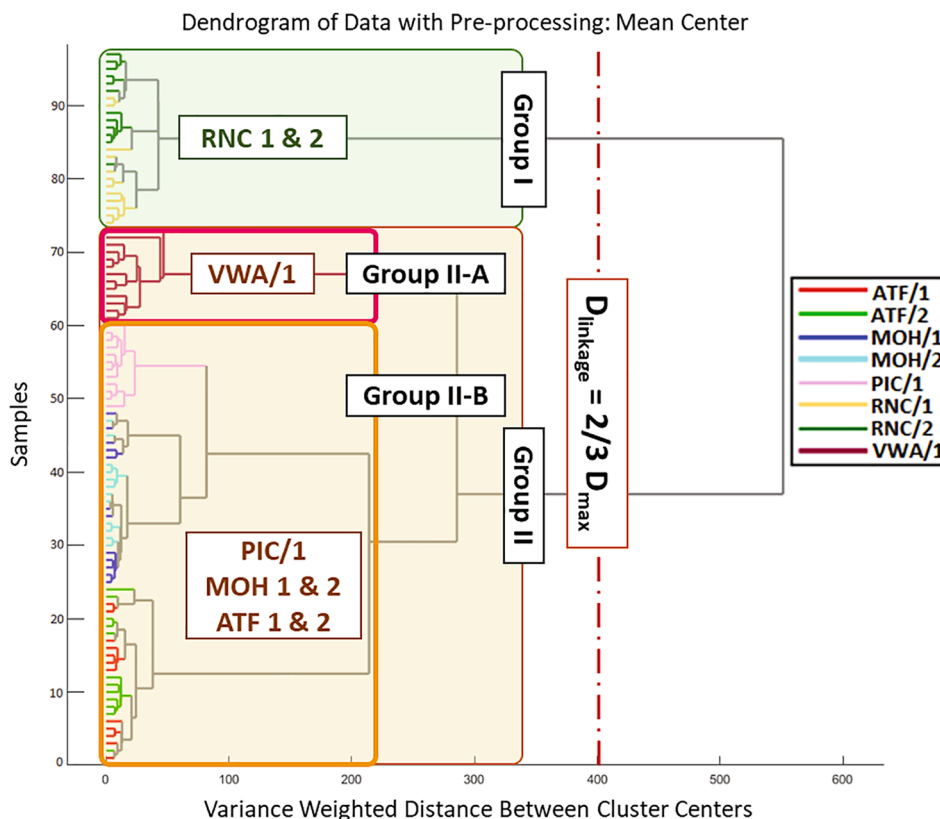

Fig. 1. Dendrogram of olive oils candidates to reference materials analysed according its VOCs fingerprints. HCA: Mahalanobis distance and Ward method as distance metric and linked criterion, respectively.

constitute the final supra-panel.

According to ISO Guide 30 and 35 (ISO Guide 30, 2015; ISO Guide 35, 2017), the reference value of materials suitable for certification should be accompanied by an uncertainty value associated with the certification process. A section in ISO Guide 35 describes how the uncertainty value of a certified reference material (u_{CRM}) is calculated by the Equation (3).

$$u_{CRM} = \sqrt{u_{char}^2 + u_{homog}^2 + u_{stab}^2} \qquad (3)$$

where u_{char} is the uncertainty calculated in the characterisation study, and u_{homog} and u_{stab} are respectively the uncertainty components derived from the possible lack of homogeneity and stability under storage conditions. The same section of ISO Guide 35 specifies that, depending on the nature of the material to be certified, the uncertainty sources are different, thus the above equation could be modified by adding or removing terms. In the case of olive oil, the main uncertainty source comes from the sensory characterisation study and consequently the above Equation (3) is simplified in Equation (4).

$$u_{CRM} = u_{char} = \frac{SD_{sp}(H15)}{\sqrt{n}} \qquad (4)$$

where n refers to the number of data from which the standard deviation has been obtained; in this case, the number of tasters not removed in any of the filters described above that constitute the final supra-panel. For each material and attribute under consideration, the certification results are then summarised by the Equation (5).

$$Certified\,value\,(CRM) = X_{CRM} \pm U_{CRM} = X_{CRM} \pm (k \cdot u_{CRM}) \qquad (5)$$

where U_{CRM} is expanded uncertainty obtained by multiplying the u_{CRM} value by the coverage factor k (k = 2). Finally, the results obtained during the sensory and instrumental analysis are included in the certificate of the reference material according to ISO Guide 31 (ISO Guide 31, 2015).

3. Result and discussion

3.1. Chemometric exploratory analysis

Unsupervised pattern recognition methods such as hierarchical cluster analysis (HCA) and principal component analysis (PCA) were applied. A fingerprint matrix (96 × 2000) was built for the study (note that the 96 value were obtained from considering 80 samples and two additional replicates of the first 8 units selected as reference, and 2000 is

the number of numerical values acquired by the measuring device and which conform each chromatogram).

3.1.1. Hierarchical cluster analysis (HCA)

HCA were performed by selecting Manhattan distance and Ward method as a distance metric and linkage criterion respectively. Fig. 1 shows the dendrogram obtained, where each nesting group corresponds to the olive oil materials which have a specific defect independently of the intensity levels. This fact suggests that the mixing percentages used between intensity levels did not create large differences in the chromatograms.

Considering $D_{linkage} = 2/3\ D_{max}$ as a criterion to define the natural grouping clustering, two groups can be defined. Group I gathers the materials coded as RNC/1 and RNC/2, which were characterised by rancid defect. As detailed in Table SM1, these olive oil materials were obtained from fruits of 'arbequina' botanical variety. Group II is defined by the rest of materials, which were obtained from olives of the 'picual' botanical variety. In addition, this group is divided into two subgroups ($D_{linkage} = 2/5\ D_{max}$) according to the fruit ripeness degree. Group II-A is composed by the material VWA/1, which was obtained from olives with a medium degree of ripeness, while group II-B collects the remaining materials that was milled from ripe fruits. It should also be noted that the PIC/1 material was used as a base for diluting other materials, which justifies the grouping with the ATF and MOH materials.

3.1.2. Principal component analysis (PCA)

Three principal components (PCs) explaining 91.9% of the total variance were selected. Fig. 2 shows the 3D scores plot on PC1-PC2-PC3 (58.5%, 19.1% y 14.3% of variance, respectively).

The 3D scores plot (Fig. 2) clearly shows four groups. Three of them are consistent with the materials having the same defect while the fourth group is composed by miscellaneous materials. The grouping is described as follows: (I) ATF/1 and ATF/2 materials having fusty/muddy defect at two intensity levels, are characterised by values close to zero in PC1, negative values of PC2 and positive values of PC3; (II) MOH/1 and MOH/2 materials having musty defect at two intensity levels, as well as PIC/1 material, the undefective base, are characterised by values close to zero in PC1 and negative values for PC2 and PC3; (III) VWA/1 material, distinctive of wine/vinegary defect, shows positive values of PC1 and PC2 and near or below zero in PC3 (note the presence of possible outlier in this material); and (IV) RNC/1 and RNC/2 materials, which are qualified as rancid defectives at two intensity levels, have negative values in PC1 and PC3 and positive values in PC2.

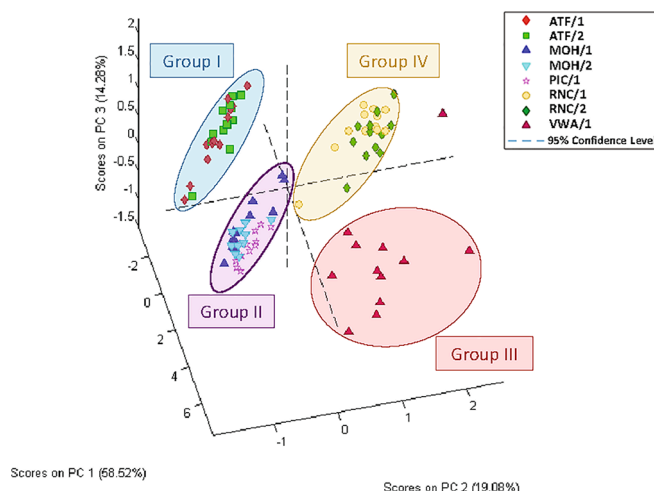Therefore, PC1 separates each group according to the type of olive oil

Fig. 2. 3D scores plot of PC1-PC2-PC3 obtained from PCA of fingerprint matrix (96 × 2000).
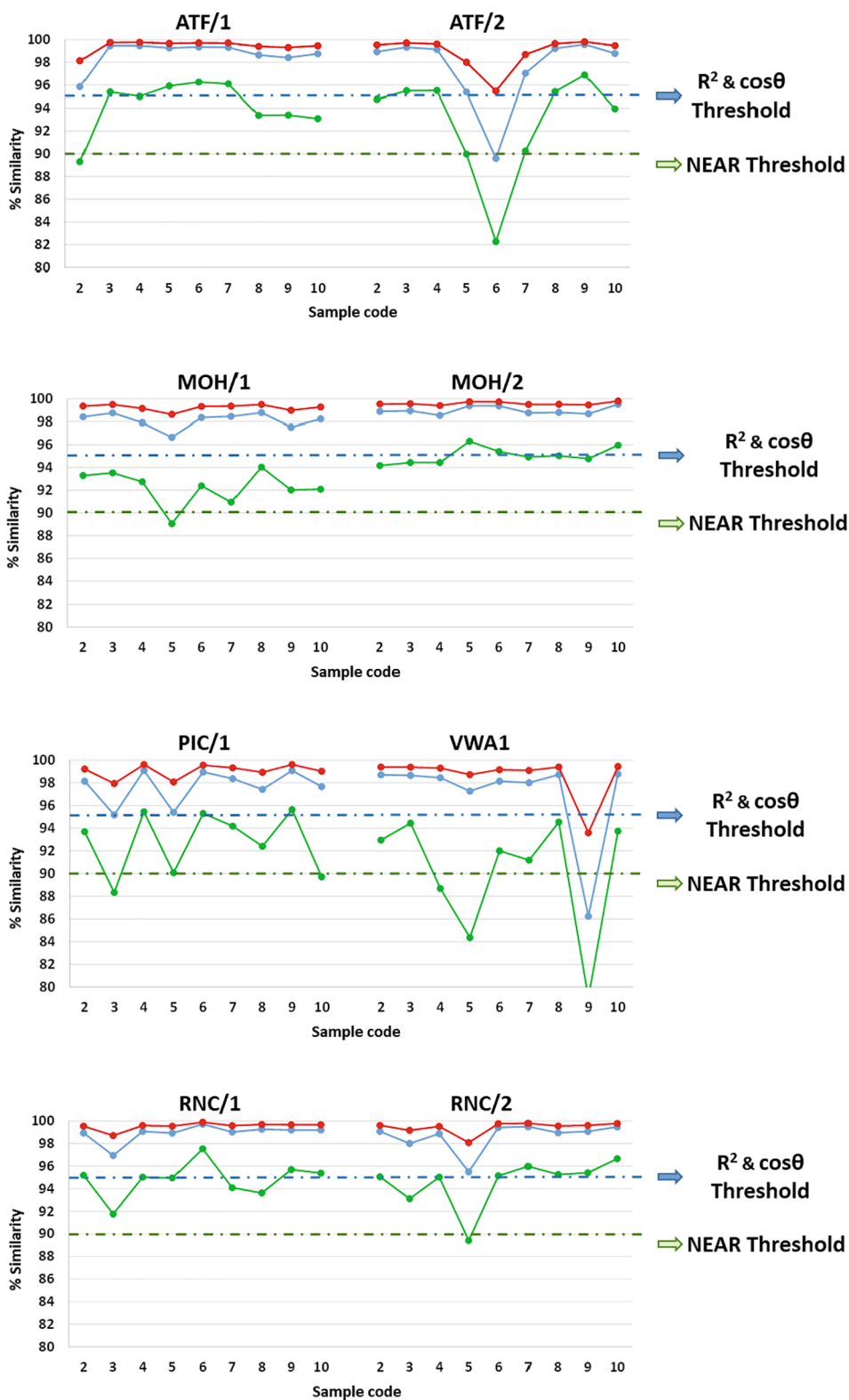
Fig. 3. Similarity-based control charts (percentage of similarity *vs* numerical identifying code of samples) obtained for the eight RMs candidates analysed for homogeneity assessment. The values of the similarity indices, R^2, $\cos\theta$ and NEAR, are plotted in blue, red and green respectively.

used as a dilution base in the production of the different RM. Groups I and II, whose dilution base was obtained from ripe fruits of the picual variety, have similar values in PC1 but very different from the values of Group III, consisting of olive oil obtained from fruits of a different picual

variety to the previous groups, and Group IV, whose base was obtained from olives of the arbequina variety. PC2 separates the materials according to the ripeness of the fruit. Groups I and II, RM from ripe fruit, present negative values in PC2 clearly opposed to those presented by
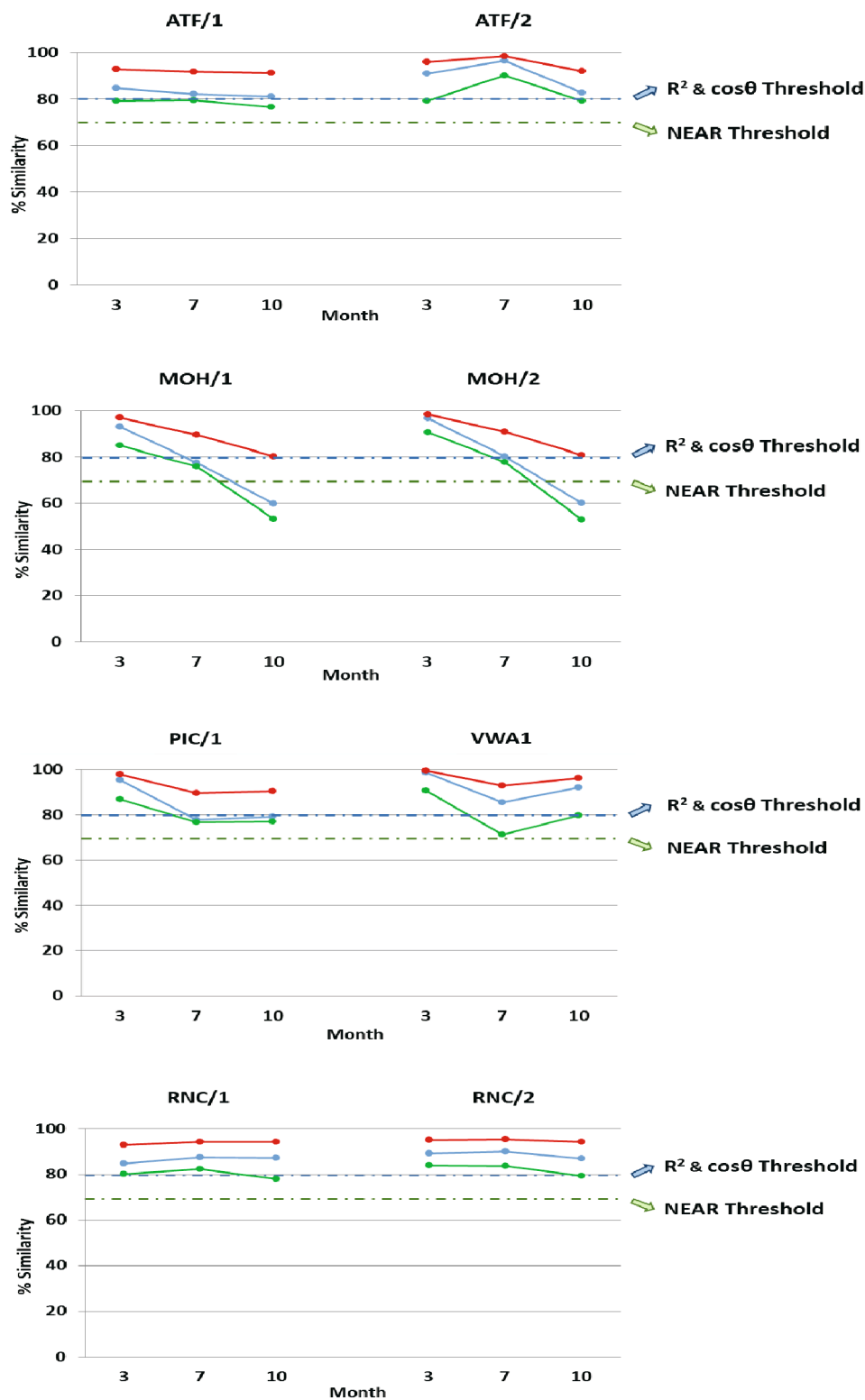
Fig. 4. Similarity-based control charts (percentage of similarity *vs* month analysis) obtained during the stability study for the eight RMs candidates. The values of the similarity indices used in this work, R^2, $\cos\theta$ and NEAR, are plotted in blue, red and green respectively. Note: the second stability analysis, corresponding with month 6, had to be postponed due to the COVID-19 pandemic.

Groups III and IV, consisting of RM whose fruit ripeness degree is lower than the previous ones. Furthermore, PC3 differentiates each group according to the defect which the materials were designed when the same type of base and maturity of the fruit is involved. In this way, PC3 allows a clear separation between Groups I and II, whose organoleptic defects were different from each other. It should be noted that the closeness of PIC/1 to MOH/1 and MOH/2 could be due to the low mixing percentages used in their production process, which implies small dissimilarity of the chromatographic signals.

Table 1
Results of the three central estimators calculated for the assignment of the certified value (*Note that the value of PIC/1 refers only to the intensity of the fruity attribute*).

Material code	Central value estimators			Deviation estimators		
	Mean	Median	Av(H15)	S_{IQR}	MADe	SD(H15)
ATF/2	4.2	4.2	4.2	0.9	0.7	0.8
ATF/1	2.5	2.6	2.5	2.6	2.0	1.9
MOH/2	4.2	4.2	4.3	1.2	1.0	1.0
MOH/1	2.8	2.8	2.8	1.1	0.9	0.9
PIC/1	3.0	3.0	3.1	0.8	0.7	0.6
RNC/2	2.5	2.4	2.5	0.9	0.9	0.8
RNC/1	1.2	1.3	1.2	1.8	1.8	1.2
VWA/1	2.5	2.4	2.5	0.8	0.6	0.6

The units of each estimation are expressed in cm, according to COI/T.20 Doc. N° 15, 2018.

3.2. Similarity analysis for verifying homogeneity and stability of the reference material

The similarity analysis for the homogeneity and stability studies were performed according to methodology described in previous papers (Valverde Som et al., 2018a; Valverde Som et al., 2018b; Ortega Gavilán, Valverde Som, Rodríguez García, Cuadros Rodríguez, & Bagur González, 2020). Briefly, the similarity values in the homogeneity study were obtained by comparing the fingerprint of each batch unit analysed with a reference fingerprint. For each material the reference fingerprint was the average of the results of the analysed unit in triplicate. Both the reference agnostized chromatographic fingerprint and each of the single unit agnostized chromatographic fingerprint for each RM candidate obtained within the homogeneity study are displayed in the supplementary material (figure SM1 - figure SM8). Chromatographic fingerprints were considered to be similar when the values of the similarity indices, namely R^2, cosθ and NEAR (Brereton, 2009; Andríc and Héberger, 2017; Pérez Robles et al., 2017) (expressed in percentage), were higher than the threshold imposed for this study, 95% for the first two indices and 90% for the third one. The thresholds were defined by the authors based on the experience gained from several studies. Similarity values below these decision thresholds suggest a potential heterogeneity among the bottled units of the packaged batch. Fig. 3 shows the similarity control charts of the eight olive oil materials developed. In this figure the relationship between the behaviour of the three similarity indices could be verified, as their values increase and decrease following the same trend.

All materials show values of the similarity indices above or very close to the decision threshold. Only two units (bottle 6 of the ATF/2 material and bottle 9 of the VWA/1 material) have values below the threshold for the three indices. Regarding the ATF/2 material, figures SM2 a and b reveal that the main differences in the similarity indices for the unit 6 correspond to a considerable decrease in the intensity values of the signals highlighted by the green arrows. The control chart of ATF/2 samples shows a significant deviation that is corrected throughout the analysis. This fact suggests a possible lack of homogeneity among the bottled units 4 and 8 that may have been caused by the bulk container used to package the different units having been refilled several times.

However, in order to establish the homogeneity of the material batch, the units within this interval were discarded. Regarding the control chart of the VWA/1 material, no clear trend is observed and sample 9 shows a value in all three indices that is evidently lower than the others. In order to confirm this behaviour, unit 9 was retested and the same results were obtained again. In fact, it can be seen in figures SM6 a and b how the signals highlighted by the green arrows are responsible for the lowest values of similarity indices. The lack of homogeneity may also be caused by the reasons mentioned in the case of ATF material. Consequently, all the bottled units from the original batch that were between the tested samples 8 and 10 of this material were rejected in order to declare the material batch as homogeneous. It should be noted that such low values of similarity for this unit were anticipated in the PCA, where the outlier corresponded indeed to unit 9 of this material. As for the remaining samples, the oscillations in the similarity values could be indicative of a lack of mixing throughout the packaging in the different units composing the batch. The zoomed regions of figures SM1 to SM8 bring forward the signals responsible of the fluctuations observed in the similarity indices, highlighting the differences among different samples for remaining RM candidates.

In contrast, the similarity indices of the stability study were obtained by comparing the representative material fingerprint corresponding to the month of study with the reference fingerprint at time zero. When the homogeneity of the different batches is ensured and all units are stored under the same conditions, it is assumed that the fingerprint obtained for each stability analysis is representative of the remaining units stored. The representative fingerprint is the average fingerprint of the three replicates from the analysed material, while the fingerprint at time zero is the average from the results of the homogeneity study of RM. The agnostized chromatographic fingerprints obtained in the stability study for each one of the different RM candidate are also included in the supplementary material (figures SM9 - SM16).

Furthermore, for stability study, the samples are considered under control when the values of R^2, cosθ and NEAR are higher than the threshold imposed for this study: 80% for the first two indices and 70% for the third one. Again, the thresholds were defined by the authors. It should be noted that the chromatographic analyses for the homogeneity study are carried out under repeatability conditions, whereas the

Table 2
Certified values of the olive oil reference materials obtained in according the characterisation study (*Note that the values of the certified reference materials are in bold; in addition, the 'consensus guiding values' are in italics and in parentheses*).

Material code	Attribute/Defect	Status	Certified value (cm)	$\pm U_{CRM}$ (k = 2) *
ATF/2	Fusty/muddy	Certified	**4.2**	**0.1**
ATF/1	Fusty/muddy	Consensus	*(2.5)*	–
MOH/2	Musty	Certified	**4.3**	**0.2**
MOH/1	Musty	Certified	**2.8**	**0.2**
PIC/1	Fruity	Non-Certified	–	–
RNC/2	Rancid	Certified	**2.5**	**0.2**
RNC/1	Rancid	Consensus	*(1.2)*	–
VWA/1	Wine/vinegary	Certified	**2.5**	**0.2**

*The U_{CRM} values have been obtained from equation (4) and (5).

stability study is carried out under quasi-reproducibility conditions. Therefore, the criterion to be applied in the latter should be less restrictive than in the former. A material is considered out of control when two consecutive stability analyses carried out in different months show values below the established threshold. Fig. 4 shows the stability control chart of the eight olive oil materials for different analysis months.

Currently, the stability study is in progress and only three units of each RM candidate reserved for it have been analysed. It can be observed in Fig. 4 how, throughout the different months of the study, the similarity indices of all the materials were higher than the established thresholds. This fact suggests that the remaining units that were cold stored during this period of time keep their organoleptic characteristics unchanged and therefore are considered to be under control. On the contrary, the samples coded as MOH/1 and MOH/2 were the only ones with similarity indices (R^2, $\cos\theta$ and NEAR) lower than the threshold for month 10. In order to discard any possible error in the instrumental measurement, these samples were analysed again, obtaining similar results. The decreasing trend observed in both materials is associated with changes to the VOC fraction as a consequence of some degradation process. In fact, in figures SM11 and SM12 (a and b) it can be seen how some of the signals have been disappearing since the first month of analysis while new signals have been appearing and increasing their intensity in the successive analyses. Therefore, the decreasing of similarity indices is caused by differences between the reference fingerprint (month 0) and the fingerprint considered as representative of all units of these materials at month 10. For the same reason, these compositional changes could suggest a shift in the certified intensity. If similar results are obtained for both materials in subsequent stability analyses, they should be subjected again to sensory analysis to confirm the change in the certified defect. However, if the tasters determine a change in intensity or the appearance of a new defect, the materials will be declared as out of stability, thus invalidating their certificate.

3.3. Sensory analysis for characterising the reference materials

According to COI/T.20 Doc. No. 15, the assessment of sensory characteristics of an olive oil should be performed by sensory analysis. Therefore, once the eight material batches were declared as homogeneous, they were subjected to the EU panel test method in order to characterise the organoleptic intensity of the attribute/defect to be certified. Note that the characterisation study was carried out between the homogeneity and stability studies. For this study, three tasting sessions were organised, in which 22 recognised tasting panels participated. The sensory analysis was conducted as part of the routine tasting of the panels, so the tasters were not aware that the oil samples were candidates for reference material. In all cases, the results were uploaded to the INTERPANEL platform by the corresponding panel leaders, where the statistical data processing was performed.

The first preliminary session was aimed at determining the most intensely perceived defect by the tasters in each material. In five of the eight materials (coded as ATF/1 & ATF/2, MOH/1 & MOH/2 and VWA/1) the selected attribute corresponds to the original designed defects (see Table SM1). In case of the RNC/1 material (low intensity of rancid), at least 2/3 of the tasters did not observe any defect in the samples. Conversely, 25% of the tasters considered the RNC/2 material to remain without any appreciable defects, while the rest of tasters divided their scores equally between musty and rancid defect. The PIC/1 material is a particular case as it was developed from ripe fruit of 'picual' variety. Although no panel had difficulty in identifying the fruity attribute, some of them perceived the ripeness of the fruit as a difficult defect to classify. These results highlight the problem of olive oils obtained from ripe fruit of this variety, even though being recognised by governmental

institution as an EVOO, some of the panels are not able to evaluate it properly.

The second session was planned to carry out a sensory analysis targeted to the defective attribute selected in the first session, in order to obtain a representative value of the organoleptic intensity. The tasters' score was the sum of all the sensory perceptions detected in the tasting session. As a result, the obtained intensity was about twice the expected value with which the materials were designed and thus the result of the second sensory analysis was discarded.

Therefore, the third and definitive session was devoted to establish the organoleptic intensity to be certified. Only 20 of the 22 panels presented results through the INTERPANEL platform. The defects and intensities with which the materials were designed were correctly identified on this occasion. Thus, for the material coded as RNC/1, at least 2/3 of the tasters perceived the rancid defect, while the remaining third did not notice any defect. In contrast with first tasting session, more than 95% of the tasters identified the rancid defect correctly in the material coded as RNC/2, while the remaining 5% did not identify any defect. Although all tasters identified correctly the fusty defect in the ATF/1 material, the scores given by the taster were distributed between 0 and 4. Again the material coded as PIC/1 similar results were obtained to the first tasting session. All tasters identified correctly the fruity attributed but considered the ripeness of the fruit as a difficult defect to be classified.

After applying the statistical filters described in section 2.5.2, it was found that more than 70% of the tasters were considered suitable for establishing the certified value. The results of the three central value estimators as well as their corresponding deviation estimators are shown in Table 1.

It can be seen how, after detecting and removing inconsistent values for each material, the results of the different central tendency and deviation estimators were similar to each other. Only the materials with a lower consensus –ATF/1 and RNC/1– show poor similarity among the different estimated standard deviations. In both cases, the lowest standard deviation value corresponded to the statistics proposed by Huber (Thompson and Lowthian, 2011). With the use of an iterative process, Huber's H15 estimation minimises the influence of outliers on the data set, which considerably reduces the deviation from the final result. Therefore, in this study, Huber's estimators were selected to establish the certified and uncertainty value.

The final results of the certification study are shown in Table 2. According to the criteria established in previous sections, five of the eight materials produced (codified as ATF/2, MOH/1, MOH/2, RNC/2 and VWA/1) were certified, assigning each reference value the corresponding uncertainty value. The materials coded as ATF/1 and RNC/1 presented standard deviation values (SD(H15)) higher than the unit, so they did not satisfy the criteria to be declared as CRM. However, the final results showed sufficient consensus among the tasters to use these candidates to RM as consensus guiding materials. The materials with this designation were also suitable for the tasters' training, although their certificate did not provide any uncertainty value. Finally, the material coded as PIC/1 was discarded as a training reference. Although there was sufficient consensus for the fruity attribute, the discrepancies produced by considering the maturity of the fruit as a defect made the material unsuitable for certification.

4. Conclusion

The combined use of sensory and instrumental analysis, developed within the 'Reference and harmonization system of virgin olive oil tasting panels from Andalusia – INTERPANEL', is presented as a rigorous methodology for the certification of the organoleptic intensity of sensory attributes using eight real olive oils batches as reference materials. The

instrumental techniques allowed the homogeneity and stability assessment by evaluating the similarity among chromatographic fingerprints of VOC fraction obtained by SPME-GC–MS. In addition, unsupervised pattern recognition techniques (HCA & PCA) have been used to screen the natural grouping of each material produced as a preliminary step to the calculation of similarity indices. This analysis provides prior information from the data set about possible heterogeneity within a particular material batch.

The set of quality statistical filters developed for data processing of the different tasting sessions allowed to obtain statistically reliable results quickly and easily. This set of quality filters is not restricted to olive oil characterisation, since it may also be applied for panel harmonisation. Although the results of the three central value estimators applied on the supra-panel of tasters were similar, the standard deviations obtained when the Huber's H15 statistical estimator was applied were lower than the remaining ones. Considering these results, Huber's H15 method was selected for the assignment of the certified value and uncertainty assessment.

Complex olive oil materials such as those obtained from ripe fruit of the 'picual' botanical variety reflect the need to use valid references, such as CRM, for the training of tasters, whether obtained by the addition of flavourings or through mixtures of real samples. Even though the latter need to be replaced with each campaign, tasters recognise that the defect has been produced by natural processes. In addition, these materials require an exhaustive study of their sensory characteristics, which directly affects the time required to achieve certification. Through the development of the panel intercommunication system developed within the INTERPANEL project, it is intended to reduce the time required for certification of each olive oil RM. Therefore, both the objectives of the INTERPANEL project and those of this work are in tune with the solutions proposed in a recent report contracted by the EU (European Commission, 2020).

CRediT authorship contribution statement

Fidel Ortega-Gavilán: Methodology, Validation, Formal analysis, Investigation, Writing – original draft. **José Antonio García-Mesa:** Conceptualization, Methodology, Resources, Writing – review & editing. **José Carlos Marzal-Fernández:** Conceptualization, Resources, Project administration, Funding acquisition. **Francisco J. Moreno-Ballesteros:** Methodology, Resources. **Francisco P. Rodríguez-García:** Conceptualization, Resources, Supervision, Funding acquisition. **Antonio González-Casado:** Conceptualization, Supervision. **Luis Cuadros-Rodríguez:** Conceptualization, Methodology, Formal analysis, Writing – review & editing, Supervision, Project administration, Funding acquisition.

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2022.132195.

References

Andríc, F., & Héberger, K. (2017). How to compare separation selectivity of high-performance liquid chromatographic columns properly? *Journal of Chromatography A, 1488,* 45–56. https://doi.org/10.1016/j.chroma.2017.01.066

Aparicio, R., Conte, L. S., & Fiebig, H. J. (2013). Olive oil authentication. In R. Aparicio, & J. Harwood (Eds.), *Handbook of Olive Oil: Analysis and Properties* ((2nd ed.), pp. 590–653). New York: Springer. https://doi.org/10.1007/978-1-4614-7777-8_16

Aparicio Ruiz, R., Barbieri, S., Gallina Toschi, T., & García González, D. L. (2020). Formulations of rancid and winey-vinegary artificial olfactory reference materials (AORMs) for virgin olive oil sensory evaluation. *Foods, 9,* 1870. https://doi.org/10.3390/foods9121870

Aparicio Ruiz, R., Morales, M. T., & Aparicio, R. (2019). Does authenticity of virgin olive oil sensory quality require input from chemistry? *European Journal of Lipid Science and Technology, 121*(12), 1900202. https://doi.org/10.1002/ejlt.v121.1210.1002/ejlt.201900202

Barbieri, S., Brkić Bubola, K., Bendini, A., Bučar-Miklavčič, M., Lacoste, F., Tibet, U., … Gallina Toschi, T. (2020). Alignment and proficiency of virgin olive oil sensory panels: The OLEUM approach. *Foods, 9*(3), 355. https://doi.org/10.3390/foods9030355

Barjol, J. L. (2013). Introduction. In R. Aparicio, & J. Harwood (Eds.), *Handbook of Olive Oil: Analysis and Properties* ((2nd ed.), pp. 1–15). New York: Springer. https://doi.org/10.1007/978-1-4614-7777-8_1

Brereton, R. G. (2009). Exploratory data analysis. In R. G. Brereton (Ed.), *Chemometrics for Pattern Recognition* (pp. 47–105). Chichester: John Wiley & Sons.

Circi, S., Capitani, D., Randazzo, A., Ingallina, C., Mannina, L., & Sobolev, A. P. (2017). Panel test and chemical analyses of commercial olive oils: A comparative study. *Chemical and Biological Technologies in Agriculture, 4,* 1–10. https://doi.org/10.1186/s40538-017-0101-0

COI/T.20/Doc. No 4 / Rev.1 (2007). Sensory analysis of alive ail. General basic vocabulary. Madrid: International Olive Council (IOC).

COI/T.20/Doc. No 14 / Rev. 7 (2021). Sensory analysis of olive oil. Guide for the selection, training and quality control of virgin olive oil taster-qualifications of taster, panel leaders and trainers. Madrid: International Olive Council (IOC).

COI/T.20/Doc. No 15 / Rev.10 (2018). Sensory analysis of olive oil. Method for the organoleptic assessment of virgin olive oil. Madrid: International Olive Council (IOC).

COI/T.20 / Doc. No 22 (2005). Organoleptic assessment of extra virgin olive oil applying to use a designation of origin. Madrid: International Olive Council (IOC).

Cuadros Rodríguez, L., Ortega Gavilán, F., Martín Torres, S., Medina Rodríguez, S., Jimenez Carvelo, A. M., González Casado, A., & Bagur González, M. G. (2021). Standardization of chromatographic signals – Part I: Towards obtaining instrument-agnostic fingerprints in gas chromatography. *Journal of Chromatography A, 1641,* 461983. https://doi.org/10.1016/j.chroma.2021.461983

d'Acampora Zellner, B., Bicchi, C., Dugo, P., Rubiolo, P., Dugo, G., & Mondello, L. (2008). Linear retention indices in gas chromatographic analysis. *Flavour and Fragrance Journal, 23,* 297–314. https://doi:10.1002/ffj.1887.

Commission, E. (2016). Regulation (EEC) No. 2568/91 on the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis (consolidated version 01991R2568). *Official Journal of the European Union, L326,* 1–128.

European Commission (2020). *Study on the implementation of conformity checks in the olive oil sector throughout the European Union. Publication office of European Union.* Retrieved from: https://op.europa.eu/en/publication-detail/-/publication/606555af-46ff-11ea-b81b-01aa75ed71a1#. Accessed April 27, 2020.

ISO Guide 30. (2015). *Reference Materials – Selected Terms and Definitions.* Geneva: International Organization for Standardization.

ISO Guide 31. (2015). *Reference materials — Contents of certificates, labels and accompanying documentation.* Geneva: International Organization for Standardization.

ISO Guide 35. (2017). *Reference Materials – Guidance for characterisation and assessment of homogeneity and stability.* Geneva: International Organization for Standardization.

Mariotti, M. (2014). Virgin olive oil: Definition and standards. In C. Peri (Ed.), *The Extra-Virgin Olive Oil Handbook* (pp. 11–20). Chichester: Wiley.

Morales, M. T., Aparicio Ruiz, R., & Aparicio, R. (2013). Chromatographic methodologies: Compounds for olive oil odor issues. In R. Aparicio, & J. Harwood (Eds.), *Handbook of Olive Oil: Analysis and Properties* ((2nd ed.), pp. 261–309). New York: Springer. https://doi.org/10.1007/978-1-4614-7777-8_8.

OLEUM Project (2016), *Protocol for the standardization of a procedure in terms of homogeneity and representativeness of the samples to be analysed in T4.2 and T4.3.* Retrieved from: http://www.oleumproject.eu/images/uploads/files/D4.1_def.pdf. Accessed April 27, 2020.

Ortega Gavilán, F., Valverde Som, L., Rodríguez García, F. P., Cuadros Rodríguez, L., & Bagur González, M. G. (2020). Homogeneity assessment of reference materials for sensory analysis of liquid foodstuffs. The virgin olive oil as case study. *Food chemistry, 322,* 126743. https://doi.org/10.1016/j.foodchem.2020.126743

Pérez Robles, R., Navas, N., Medina Rodríguez, S., & Cuadros Rodríguez, L. (2017). Method for the comparison of complex matrix assisted laser desorption ionization-time of flight mass spectra. Stability of therapeutical monoclonal antibodies.

Chemometrics and Intelligent Laboratory Systems, 170, 58–67. https://doi.org/10.1016/j.chemolab.2017.09.008

Rason, Jonathan, Laguet, Arlette, Berge, Phillipe, Dufour, Eric, & Lebecque, Annick (2007). Investigation of the physicochemical and sensory homogeneity of traditional French dry sausages. *Meat Science, 75*(3), 359–370. https://doi.org/10.1016/j.meatsci.2006.06.004

Thompson, M., & Lowthian, P. J. (2011). *Notes on Statistics and Data Quality for Analytical Chemists* (pp. 135–139). London: Imperial College Press.

Valverde Som, L., Ruiz Samblás, C., Rodríguez García, F. P., & Cuadros Rodríguez, L. (2018a). Multivariate approaches for stability control of the olive oil reference materials for sensory analysis – Part I: Framework and fundamentals. *Journal of the Science of Food and Agriculture, 98*, 4237–4244. https://doi.org/10.1002/jsfa.8948

Valverde Som, L., Ruiz Samblás, C., Rodríguez García, F. P., & Cuadros Rodríguez, L. (2018b). Multivariate approaches for stability control of the olive oil reference materials for sensory analysis – Part II: Applications. *Journal of the Science of Food and Agriculture, 98*, 4245–4252. https://doi.org/10.1002/jsfa.8946

Willett, W. C., Scacks, F., Trichopoulou, A., Drescher, G., Ferro Luzzi, A., Helsing, E., & Trichopoulos, D. (1995). Mediterranean diet pyramid: A cultural modern for healthy eating. *The American Journal of Clinical Nutrition, 61*, 1402S–1406S. https://doi.org/10.1093/ajcn.61.6.1402S