LWT - Food Science and Technology 191 (2024) 115679

Contents lists available at ScienceDirect

LWT

journal homepage: www.elsevier.com/locate/lwt

Shedding light on the ageing of extra virgin olive oil: Probing the impact of temperature with fluorescence spectroscopy and machine learning techniques

Francesca Venturini [a,b,*], Silvan Fluri [a], Manas Mejari [c], Michael Baumgartner [a], Dario Piga [c], Umberto Michelucci [b,d]

[a] Institute of Applied Mathematics and Physics, Zurich University of Applied Sciences, Technikumstrasse 9, 8401, Winterthur, Switzerland
[b] TOELT LLC, Advanced AI Lab, Birchlenstrasse 25, 8600, Dübendorf, Switzerland
[c] IDSIA - Dalle Molle Institute for Artificial Intelligence, USI-SUPSI, Via la Santa 1, 6962, Lugano, Switzerland
[d] Lucerne University of Applied Sciences and Arts, Computer Science Department, Lucerne, Switzerland

ARTICLE INFO

Keywords:
Extra virgin olive oil
Fluorescence spectroscopy
UV spectroscopy
Machine learning
Optical sensor
Thermal stability
Quality control
Oxidation

ABSTRACT

This work systematically investigates the oxidation of extra virgin olive oil (EVOO) under accelerated storage conditions with UV absorption and total fluorescence spectroscopy. With the large amount of data collected, it proposes a method to monitor the oil's quality based on machine learning (ML) applied to highly-aggregated data.

EVOO is a high-quality vegetable oil that has earned worldwide reputation for its numerous health benefits and excellent taste. Despite its outstanding quality, EVOO degrades over time due to oxidation, which can affect both its health qualities and flavour. Therefore, it is highly relevant to quantify the effects of oxidation on EVOO and develop methods to assess it that can be easily implemented under field conditions, rather than in specialized analytical laboratories.

The ML approach indicates that the two excitation wavelengths (480 nm) and (300 nm) exhibit the maximum relative change in fluorescence intensity during the ageing for the majority of the oils, thus identifying the wavelengths which are more informative for quality prediction. Also, the paper proposes a method for the prediction of olive oil quality using highly-aggregated data. Such a method is of interest because it paves the way to the realization of a low-cost, portable device for in-field quality control.

The following study demonstrates that fluorescence spectroscopy has the capability to monitor the effect of oxidation and assess the quality of EVOO, even when the data are highly aggregated. It shows that complex laboratory equipment is not necessary to exploit fluorescence spectroscopy using the proposed method and that cost-effective solutions, which can be used in-field by non-scientists, could provide an easily-accessible assessment of the quality of EVOO.

1. Introduction

Extra virgin olive oil (EVOO) is a high-quality product that is widely consumed due to its health benefits and culinary properties (Frankel, 2011). Its historically predominant Mediterranean diffusion is spread to non-producing countries, and EVOO is becoming more popular globally. The assurance of the quality of EVOO, from the production throughout its shelf life, is therefore becoming of great importance. The quality of EVOO depends on multiple factors, such as the olive cultivar, the geographic location, the processing method, and storage conditions, just to name the most important ones.

During storage, EVOO undergoes chemical and physical changes that can result in the deterioration of its quality attributes, such as flavour, aroma, and colour, but also in the loss of nutritional properties. The main causes of this degradation are autoxidation (in darkness) and photo-oxidation (in presence of light) which depend on temperature, time, amount of available oxygen or water, and result in the degradation of fatty acids, tocopherols and chlorophylls and in formation of primary and secondary oxidation products (Gómez-Alonso, Salvador, and Fregapane (2004); Mancebo-Campos, Fregapane, and Desamparados

* Corresponding author. Institute of Applied Mathematics and Physics, Zurich University of Applied Sciences, Technikumstrasse 9, 8401, Winterthur, Switzerland.
E-mail address: vent@zhaw.ch (F. Venturini).

https://doi.org/10.1016/j.lwt.2023.115679
Received 30 August 2023; Received in revised form 12 December 2023; Accepted 22 December 2023
Available online 27 December 2023

Salvador (2008); Aparicio-Ruiz, Minguez-Mosquera, and Gandul-Rojas (2010); Esposto et al. (2017); Lam, Roy, and Chattopadhyay (2020)).

The quantitative assessment of the effects of ageing during storage and the monitoring of the quality of EVOO throughout its entire life cycle is crucial to ensure the high quality of the oil is guaranteed until the expiration time. Previous studies indicate that the concentration of several compounds change during storage and that the freshness is significantly influenced by the storage conditions and not only by the time elapsed from the harvest of the olives (Fadda et al. (2012); Aparicio Ruiz, Tena Pajuelo, Romero del Río, García-González and Morales Millán (2017); Conte et al. (2020). The compounds studied include chlorophylls, carotenoids, tocopherols and lipids (Aparicio-Ruiz and Gandul-Rojas (2012); Camerlingo, Portaccio, Delfino, Lepore et al. (2019); Botosoa, Chèné, and Karoui (2021); Lopes and Courrol (2023)). In a comprehensive study over several years using sensory and VOC analyses it was observed that only 47% of the oils samples collected worldwide conform to the declared EVOO category (Taiti, Marone, Fiorino, & Mancuso, 2022).

According to the European Regulation and its amendments (Union, 1991, 2013) the quality assessment of EVOO is performed through a series of chemical analysis and on sensory evaluation by certified laboratories and panelists respectively. These analyses are time-consuming, expensive, and are not easily accessible for a large number of producers. Alternatively to chemical analysis, this work investigates the use of fluorescence spectroscopy, which has the advantages of being a rapid, cost-efficient, and at the same time sensitive technique (El Orche, Bouatia, and Mbarki (2020); Guzmán, Baeten, Pierna, and García-Mesa (2015)). Fluorescence spectroscopy has been demonstrated to be a technique that can be used successfully to monitor the freshness of EVOO and the quality of olive oil in general because of the natural presence of fluorophores in olive oil, the strongest of which is chlorophyll (Sikorska et al. (2008); Karoui and Blecker (2011), Lobo-Prieto, Tena, Aparicio-Ruiz, García-González, and Sikorska (2020); Al Riza, Kondo, Rotich, Perone, and Giametta (2021)). Machine learning (ML) has emerged as a promising tool for the extraction from spectral analysis of relevant features and for the objective and efficient evaluation of the quality of EVOO (Ordukaya and Karlik (2017); Vega-Márquez, Nepomuceno-Chamorro, Jurado-Campos, and Rubio-Escudero (2020); Venturini et al. (2021); Zaroual, Chénè, El Hadrami, and Karoui (2022)). This is because ML algorithms can learn from data to identify patterns and classify samples based on their quality attributes.

This study aims to contribute to a general understanding of EVOO ageing by systematically studying the effects of thermal degradation intended as accelerating ageing condition. This is done by a comprehensive spectroscopic analysis of the absorption and fluorescence properties. To observe general trends in ageing, a high sample variability was chosen in terms of geographical origin and composition properties. The quality assessment of EVOO is performed by measuring the UV-absorption parameters K_{268} and K_{232} defined by the European Regulations (Union, 1991, 2013). It should be noted that although the EU defined quality assessment includes other parameters not measured by this study, if an oil exceeds the limits in terms of UV parameters, it can definitely not be sold as EVOO any more.

The use of ML methods is being explored to determine whether simple fluorescence sensors can be used to qualitatively control oil quality in the field and at production and storage sites. Therefore, the large amount of data is analyzed in an aggregated form. An analysis of the spectral information is expected to provide additional information. However, the theses put forward in this work is that a sensor measuring the fluorescence intensity at one or maximum two excitation wavelengths can reliably and robustly predict if the EVOO is still in this quality class. The advantage of such a sensor would be that, requiring only one or two LEDs for the excitation and a photodiode for the detection, its design, dimensions and working principle would be extremely simple and cheap. Such a sensor, which could be the subject of future work, would open the way to a fast and easy qualitative assessment of the quality of olive oil.

The contributions of this work are four: 1) extensive UV-absorption spectroscopy measurement at the wavelength specified by the European Regulations over time for commercial EVOO; 2) extensive fluorescence emission analysis of the same oils through the acquisition of excitation-emission matrices (EMMs); 3) identification of the two wavelengths for which the overall relative change in the fluorescence is maximal; 4) proposal of a method for the prediction of olive oil quality based on aggregated data.

2. Materials and methods

2.1. Extra virgin olive oil samples

This study involves 24 different commercially available extra virgin olive oils. The oils were chosen to cover a wide range of price classes and production regions (Italy, Spain, Portugal, Greece, and unspecified in Europe). The motivation for selecting samples as diverse as possible is rooted in the authors' desire to identify behavioral characteristics that are not specific to factors such as geographical regions or cultivars. Instead, they aim to discover traits that are common to all EVOO. The results thus obtained have greater applicability. The list of the EVOOs is reported in Table 1.

The oxidative stability of olive oils was investigated under accelerated storage conditions, following the modified Schaal oven test Celsius (Evans, List, Moser, & Cowan, 1973) by holding the samples at 60 °C in the dark. This approach enables to analyze the oxidation in shorter time because it reproduces oxidative changes similar to those observed under actual shelf life conditions (Morales and Przybylski (2013); Mancebo-Campos et al. (2008)).

All oils were measured unaged (from a just-opened bottle) and at nine different time intervals (ageing stages) up to a maximum of 53 days in the oven. The exact duration of the aging for each ageing step is reported in Table 2.

The samples were prepared from commercial bottles by filling to the top and hermetically closing 4 ml glass vials to minimize the amount of oxygen in the headspace. One vial per measurement was prepared: three samples for each of the 24 oils and for 10 ageing steps for a total of 720 samples. This procedure ensures that all the samples, at all stages, aged under identical conditions: temperature, darkness, and oxygen in the

Table 1
List of the olive oils samples analyzed in this study and their geographical origin. IT: Italy, ES: Spain, GR: Greece, PT: Portugal, EU: European not specified.

Label	Sample description	Origin
A	Coop Naturaplan Italienisches Olivenöl (BIO)	IT
B	Hacuinda Don Paolo	ES
C	Monocultivar Nocellara Bio	IT
D	Monini, Toscano IGP	IT, Tuscany
E	Monini, Classico	IT
F	Oliva, Favola	IT
G	Migros, M Classics	ES
H	Alexis, Manaki	GR
I	Migros, Bio Italienische Olivenöl	IT
J	Alnatura, Natives Olivenöl extra	ES
K	Migros, Bio Griechisches Olivenöl	GR
L	Demeter, Spanisches Olivenöl	ES
M	Filippo Berio, Il Classico	EU
N	Demeter, Bio Coop Naturaplan Portugisisches Olivenöl	PT
O	Castillo, Don Felpe	ES
P	Coop, Naturaplan Bio Griechisches Olivenöl	GR
Q	Demeter, Son Naava	ES, Mallorca
R	Iliada, Kalamata PDO	GR
S	Sapori d'Italia, Sicilia	IT, Sicily
T	San Giuliano, Sardegna DOP	IT, Sardinia
U	Coop, Naturaplan Bio Spanisches Olivenöl	ES
V	San Giuliano, Fruttato	IT
W	San Giuliano, L'Originale	IT
X	Coop, Qualité-Prix	IT, ES, GR

Table 2
Ageing duration in days for each ageing step.

Ageing step	Ageing Duration at 60 °C (days)
0	unaged
1	2
2	4
3	7
4	9
5	18
6	27
7	36
8	45
9	53

headspace. The importance of the latter needs not to be underestimated. In fact, it was shown that the presence of oxygen in the vial's headspace directly impacts the oxidation process, leading to an increase in the concentration of free fatty acids and the peroxide value (Iqdiam et al., 2020). At each ageing step, three identical samples of the 24 oils, for a total of 72 samples, were taken from the oven and measured. The redundancy was done to have multiple measurements of the same oil at any ageing conditions to check the reproducibility of the results. More details on EVOOs, sample preparation, and the ageing process are reported in the article (Venturini, Fluri, & Baumgartner, 2023).

2.2. UV-absorption and fluorescence spectroscopy measurements

The quality of EVOOs at the various phases of ageing can be monitored using UV absorption spectroscopy, which provides information on the degree of oxidation. As olive oil undergoes oxidation, primary and secondary oxidation products form which are characterized by absorption in the UV. The European Regulation and its amendments (Union, 1991, 2013) define precise wavelengths at which the UV absorbance should be measured and clear limits for the maximal absorbance. According to the Regulation the three parameters (called extinction coefficients) to quantify the absorbance in the UV are K_{232}, the absorbance at 232 nm, K_{268}, the absorbance at 268 nm, and ΔK determined from the absorbance at 264 nm, 268 nm and 272 nm according to the formula

$$\Delta K = \left| K_{268} - \left(\frac{K_{264} - K_{272}}{2} \right) \right| \tag{1}$$

The UV spectroscopy extinction coefficients were determined with an Agilent Cary 300 UV–Vis spectrophotometer on diluted samples. Both the sample preparation and the measurement method followed the procedures defined by the European Regulation and its amendments (Union, 1991, 2013). The three parameters K_{232}, K_{268}, and ΔK were measured for each type of oil and ageing step at a constant temperature of 22 °C.

The total fluorescence spectroscopy characteristics of the 720 samples were determined by acquiring the excitation emission matrix (EEM). The EEMs were measured with an Agilent Cary Eclipse Fluorescence Spectrometer by changing the wavelength of the illuminating excitation light from 300 to 650 nm in steps of 10 nm and measuring the intensity of the fluorescence emitted as a function of the wavelength between 300 and 800 nm in steps of 2 nm. All measurements were made at a constant temperature of 22 °C on undiluted samples. The detailed information on the acquisition parameters and procedures of both spectroscopic techniques are reported in (Venturini et al., 2023). The complete dataset including both UV-absorption and fluorescence spectroscopy raw measurements is publicly available (Venturini, Fluri, & Baumgartner, 2023).

2.3. Machine learning methods

This work aims to determine whether it is possible to predict if an oil aged, for example after a long time from processing or in the bottle, is still extra virgin using only aggregated information from fluorescence spectra. The goal is to explore whether a very simple sensor based on the measurement of the fluorescence intensity, thus with a non-destructive purely optical method on undiluted samples, could be used to perform a quality assessment. The motivation for the aggregation of the data is to explore the potentiality of a low-cost and portable device with a hardware as simple as possible: one or at most two LEDs for the excitation and a single photodiode for the fluorescence intensity detection. Clearly the aggregation of the information of such a device poses a serious challenge as compared to the analysis of the entire EMMs.

Therefore, the proposed approach first requires the identification of the excitation wavelengths containing the largest amount of information for this specific purpose (Section 2.3.1). Once the method determines the most relevant wavelengths, the machine learning algorithm classifies the oil as EVOO or non-EVOO according to the UV-spectroscopy criteria of the European regulation (Union, 1991, 2013) (Sections 2.3.2 and 2.3.3).

2.3.1. Information content maximization approach

To identify the fluorescence spectra that contain the most information about thermal degradation this work determines at which excitation wavelength λ_i $i = 1, \ldots, 35$ the resulting fluorescence spectrum shows the greatest change between different ageing steps. The change is quantified mathematically with the Relative Error (*RE*) for a sample *j*, an excitation wavelength λ_i and an ageing step *k* defined as

$$RE_{j,k}\left(\lambda_i\right) = \frac{\left\| I_{j,0}\left(\lambda_i\right) - I_{j,k}\left(\lambda_i\right) \right\|_2^2}{\left\| I_{j,0}\left(\lambda_i\right) \right\|_2^2} \tag{2}$$

where $I_{j,k}$ (λ_i) is a vector whose components represent the intensity at various emission wavelengths resulting from excitation at the wavelength λ_i, for sample j and aging step k. $\|\cdot\|_2^2$ indicates the Euclidean norm squared, or in other words, the squared sum of all the values of the vector (the symbol "·" is a placeholder for any possible vector to which the norm operation can be applied). 11 pixels of the measured spectra around the excitation frequency were set to zero (5 on the left, 5 on the right of the excitation frequency and the excitation frequency itself) to remove the intensity due to Rayleigh scattering.

Given the focus on long-term forecasting, to identify the two excitation wavelengths, $\lambda^{[1]}$ and $\lambda^{[2]}$, at which the fluorescence changes are most pronounced, the *RE* was calculated for $k = 9$. First the wavelength $\lambda_j^{[1]}$ for which the *RE* for $k = 9$ is largest was determined for each oil *j*. Then, the number of times that this wavelength appear in the list was counted. Finally, the two wavelengths that are most common are selected. Since the fluorophores in olive oil have broad absorption bands, wavelengths with variations of ± 10 nm excite the same fluorophores. Therefore, the wavelength 310 nm can be counted as "equivalent" to 300 nm. The analysis of the acquired data showed that taking the average of the three samples or taking one of the samples does not produce any changes in model performance. For the predictive models just one of the three samples was used.

2.3.2. Classification model

To classify the sample in EVOO or non-EVOO the following algorithms were tested in this work.

- AdaBoost (with a decision tree as base model initialized with maximum depth of 1, with 5 estimators and a different random seed each training run. The algorithm was trained with a learning rate of 1 and the SAMME. R boosting algorithm (Hastie, Rosset, Zhu, & Zou, 2009)).
- Random forest (with 100 estimators and maximum depth of 2).
- Logistic regression.
- Naïve Bayes.

The results reported in this paper are for the AdaBoost algorithm, as this provided the best results for both parameters K_{268} and K_{232}. AdaBoost is known to outperform other algorithms by sequentially combining weak learners and assigning higher weights to misclassified instances, focusing on the most challenging data inputs and thus creating a robust ensemble model. The results obtained using the other algorithms are reported in full detail in the additional material. The AdaBoost algorithm was tested for various values of the number of learners, from 5 to 50, since this is known to be the parameter with the largest influence on the results (James, Witten, Hastie, & Tibshirani, 2013). The learning rate was not changed during the tests. The random forest classifiers were tested for multiple depths, ranging from 50 to 100.

The algorithms classify the oils following the criteria for the UV-spectroscopy limits for either the parameter K_{268} or the parameter K_{232} for EVOO defined by the European regulation.

- class 1 if $K_{268} < 0.22$ or class 0 if $K_{268} > 0.22$.
- class 1 if $K_{232} < 2.5$ or class 0 if $K_{232} > 2.5$.

The parameter ΔK was not used for the classification since the error on the measurements is comparable to the overall change observed (see Section 3.1). The dataset is unbalanced since after 53 days of ageing only seven oils are bad using the K_{232} criterion and only four oils are still good using the K_{268} criterion (see Section 3.1). Therefore, a random oversampling without replacement was used for the model training (Hastie et al., 2009). For more information on the chosen models, the reader is referred to (James et al., 2013).

In order to evaluate models' performance accuracy, sensitivity, and specificity were used. Accuracy is simply the ratio of correctly predicted instances to total instances. To define sensitivity and specificity, it is necessary first to defined true positives (TP), true negatives (TN), false negatives (FN) and false positives (FP). Positives and negatives indicate two classes, which in the case discussed in this paper are good (still EVOO) or bad quality olive oil (not EVOO).

TP are instances that are correctly predicted as positive by a classification model. FP are instances that are incorrectly predicted as positive when they are actually negative. TN are instances that are correctly predicted as negative by a classification model. FN are instances that are incorrectly predicted as negative when they are actually positive. Accuracy is defined as

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{3}$$

Sensitivity, also known as True Positive Rate or Recall, measures the proportion of actual positive instances correctly predicted and is given by

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{4}$$

Specificity represents the True Negative Rate and is defined as

$$\text{specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{5}$$

indicating the ability to correctly identify negative instances.

2.3.3. Validation

To validate the model, two different strategies were followed. These are summarized in Table 3. The approaches reflect the classical validation methods used in time series (Fan and Yao (2003); McQuarrie and Tsai (1998)), since the measurements were made at subsequent time points.

For both methods, an increasing number of ageing steps was used for the training to determine the amount if information needed for a reliable classification. In Method 1 the validation (prediction of the class) is always done on the 9th ageing step, and in Method 2 the class at the immediate next ageing step is predicted. For example, in Method 1 the *REs* calculated from the spectra at ageing steps 4 and 5 was used for training, and the model was validated using the *REs* corresponding to the ageing step 9, while in Method 2 the model was validated using the *REs* at ageing step 6.

Table 3
An overview of the different training methods used. For both methods an increasing number of ageing steps for training was used. In Method 1 the validation is done on step 9; in Method 2 the validation is done on the immediate next ageing step.

Input ageing steps used for training	Validated on ageing step (method 1)	Validated on ageing step (method 2)
4, 5	9	6
4, 5, 6	9	7
4, 5, 6, 7	9	8
4, 5, 6, 7, 8	9	9

3. Results and discussion

3.1. UV-absorption results

The evolution of the UV extinction coefficients is shown in Figs. 1–3. The first observation is that, generally, the rates of increase of the three parameters vary significantly between the samples. This different ageing patterns may depend, among other factors, on the cultivar of the olive oils used, on the processing method, and the geographical origin of the fruits. It is to be noticed that the EVOO labeled C was characterized by a strong absorption in the UV already before the start of the ageing in this study, that is in unaged conditions. This indicates that the specific bottle had already suffered significant ageing and was no longer an EVOO at the beginning of this study. This can be easily seen in Fig. 2.

Looking at the single extinction coefficients, Fig. 1 shows the values of K_{268} for all samples and the average (as a thick solid black line in the right panel). Although the average value is not chemically significant per se, it helps determine if there is a common behaviour despite the heterogeneous nature of the EVOOs. The results indicate that no significant changes are visible in the first 10 days, whereas a clear increase is visible after 10 days within the measurement errors. Only 4 oils have a K_{268} parameter that is below the threshold after 53 days (indicating still a high quality). The average rate of change is 0.004/day, the total change of the average over 53 days is 143 %. Similarly, Fig. 2 shows that the values of K_{232} also tend to increase over time, although the change is less pronounced. At the end of the study, only seven oils have exceeded the regulatory limit for EVOO. The average rate of change is 0.008/day and the change of the average over 53 days is 26%.

The behaviour of ΔK (Fig. 3) mimics the one of K_{268}, showing an appreciable increase only after 10–15 days. The uncertainty on the value of ΔK was estimated by repeating the measurement with the same oil at the same ageing step and was found to be approximately 0.02 (2σ). Since the ΔK values are comparable to the measurement error, it was decided not to use them as labels for the machine learning model.

The increase in K_{268} and K_{232} over time is consistent with the observations previously reported (Mancebo-Campos et al. (2008); Conte et al. (2020); Escudero, Ramos, La Rubia, and Pacheco (2016)). It is important to note that, differently from previous studies, the trends presented here arise from a variety of samples and are therefore not specific of a cultivar or of a geographical origin. Additionally, unlike previous studies, the ageing in this work was performed using closed vials with the minimum amount of oxygen possible to mimic the storage conditions of commercial bottles. This precaution avoids the

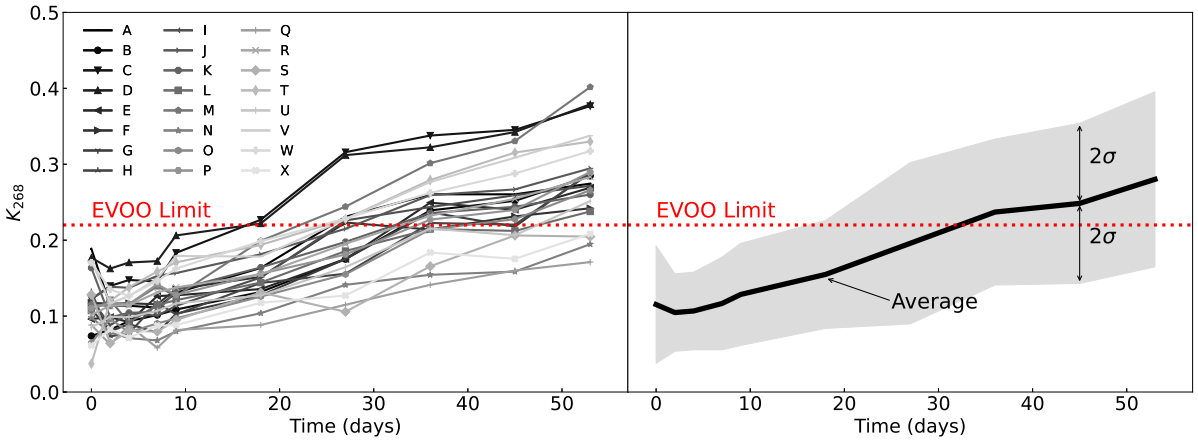

Fig. 1. Left panel: UV-Spectroscopy K_{268} parameter over the time during the ageing at 60 °C. Right panel: average of the values of all the oils and variation during the ageing. σ is the standard deviation at each ageing step. The red line is the limit of K_{268} according to the European Regulation.

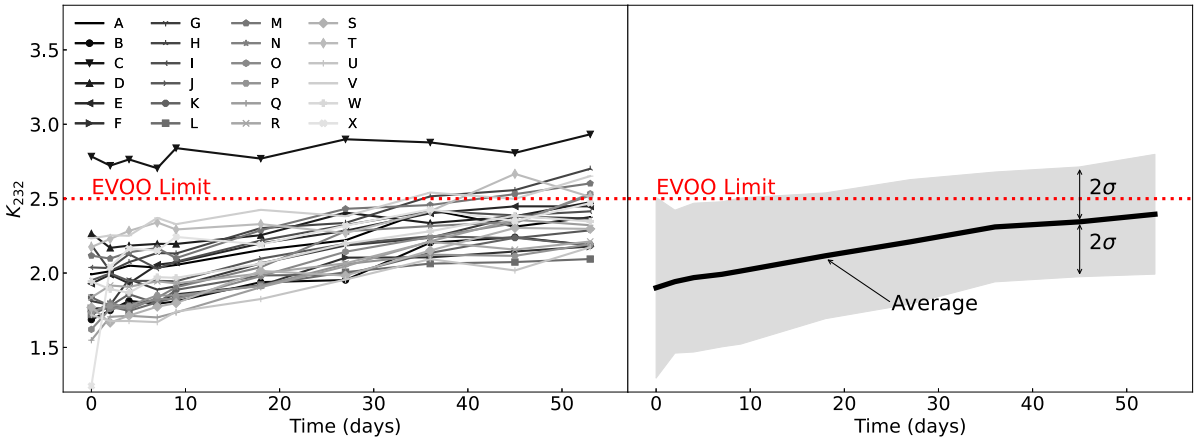

Fig. 2. Left panel: UV-Spectroscopy K_{232} parameter over the time during the ageing at 60 °C. Right panel: average of the values of all the oils and variation during the ageing. σ is the standard deviation at each ageing step. The red line is the limit of K_{232} according to the European Regulation.

demonstrated acceleration of oxidation and reduction of shelf life associated with oxygen in the headspace (Iqdiam et al., 2020).

The results of this work indicate that the absorbance at 268 nm (K_{268}) is the most sensitive parameter between the three to detect ageing due to temperature. This is consistent with the fact that the K_{268} extinction coefficient is an indicator of the presence of secondary oxidation products (Escudero et al., 2016).

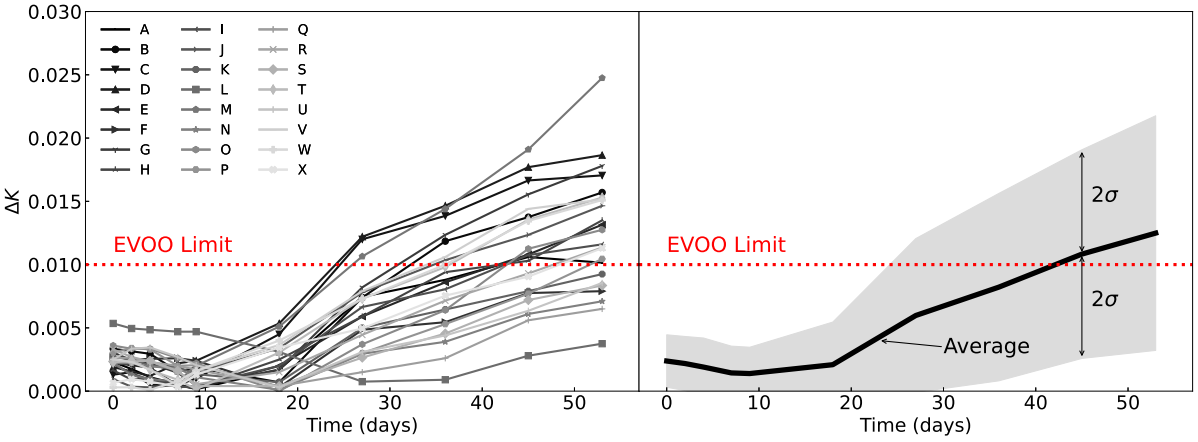

Fig. 3. Left panel: UV-Spectroscopy ΔK parameter over the time during the ageing at 60 °C. Right panel: average of the values of all the oils and variation during the ageing. σ is the standard deviation at each ageing step. The red line is the limit of ΔK according to the European Regulation.

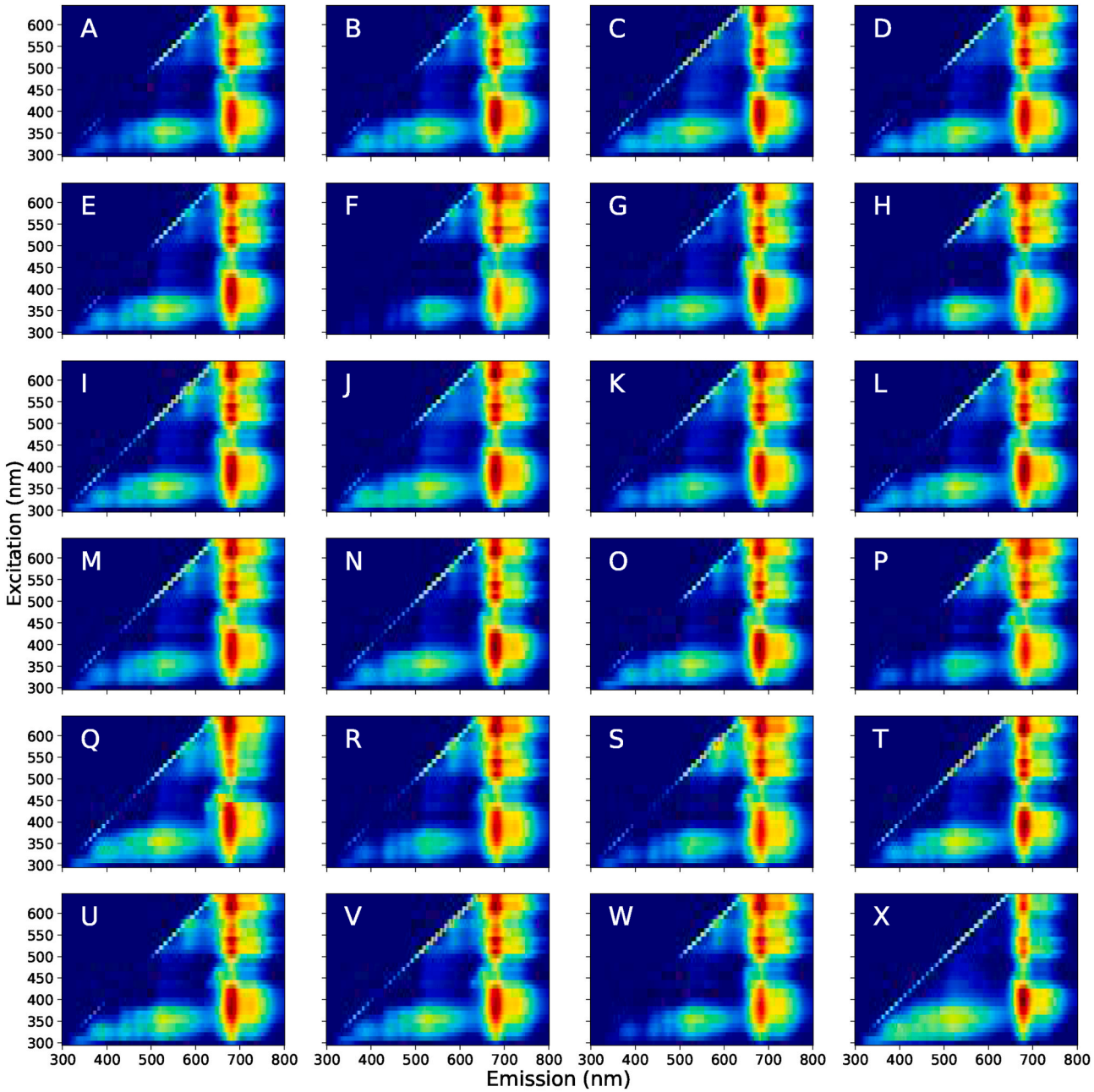

Fig. 4. Excitation-emission matrices (EEMs) of the 24 EVOO at ageing step 0 (unaged oils). The data are shown in logarithmic scale to make differences more visible. Figure from (Venturini et al., 2023).

3.2. *Fluorescence spectroscopy results*

The EEMs of the oils measured immediately after opening the bottles (Fig. 4) have strong similarities with a small variability in the fluorescence characteristics of specific spectral regions. The strongest signal is clearly the asymmetric band with emission between ca. 650 nm and 750 nm. This band is strongest for excitation wavelengths of 350–400 nm and 600–650 nm and is attributed to chlorophyll pigments, consistent with what has previously been reported in the literature (Al Riza et al. (2021); Sikorska, Khmelinskii, and Sikorski (2012); Zandomeneghi, Carbonaro, and Caffarata (2005); Baltazar, Hernández-Sánchez, Diezma, and Lleó (2020)). The most evident differences between the oils arise for excitations between 310 and 350 nm end emissions between 350 nm and 600 nm, such as pyridoxine, vitamins and flavins (Zandomeneghi et al., 2005). It is difficult to assign these bands because there are many endogenous molecules which fluoresce in this spectral range. Additionally, oxidation products that may be present in fresh olive oil also emit in a spectral range of 400–450 nm (Baltazar et al. (2020); Al Riza, Kondo, Catalano, and Giametta (2019)). It is to be noticed that the measurements were performed on undiluted samples. The reason is that in this work the authors want to identify a method which is as simple as possible, without any sample handling, to be useable in field conditions and without special laboratory equipment and knowledge. Therefore, even if the spectra may be subjected to the inner filter effect (Skoog et al., 2017), this is either not relevant or can be exploited to be more sensitive to thermal degradation.

The evolution of the fluorescence emission spectrum with thermal ageing is shown in Fig. 5 for two selected oils and at three excitation wavelengths. The excitation wavelengths 300 nm and 480 nm were chosen because they correspond to those selected by the information content maximization algorithm described in Section 2.3.1 (see the discussion in Section 3.3). 400 nm was chosen because it corresponds to

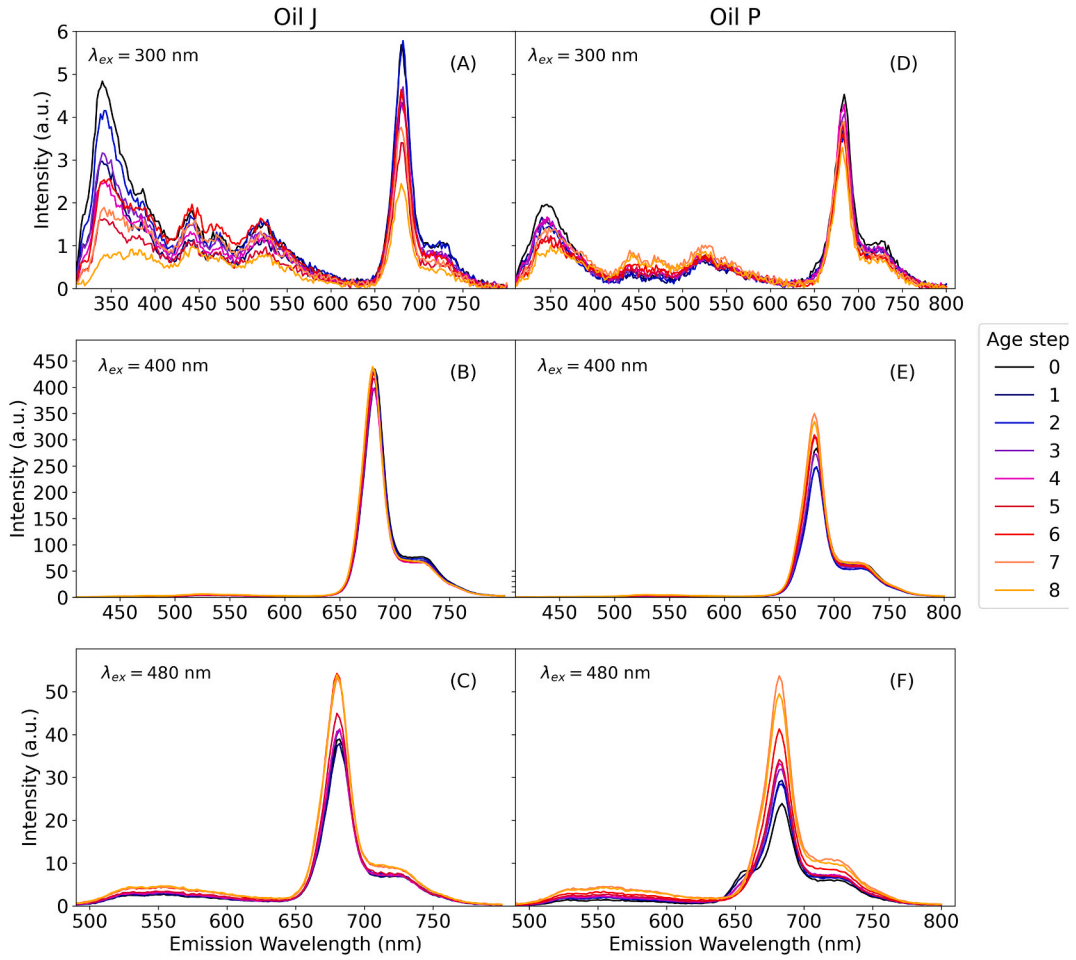

Fig. 5. Evolution of the fluorescence emission with the ageing for two selected oils at selected excitation wavelengths. Left panels: oil J; right panels: oil P. λ_{ex} is the excitation wavelength.

the absorption band of chlorophylls, which have the strongest fluorescence signal, as it can be seen from Fig. 4.

The analysis of the changes shows that there are some common changes and some oil-specific ones. For an excitation wavelength λ_{ex} = 300 nm, there is a decrease in the intensity of both peaks at ca. 350 and at ca. 680 nm for all oils. This can be interpreted as an oxidative reduction of tocopherols and chlorophylls, respectively. On the other hand, the changes in the region between 400 and 500–550 nm depend on the oil chemical properties, showing an increase (for example for oil F, O and P, the latter shown in Fig. 5 in panel (D)), no change (for example for oil A or H, not shown here), or an unclear evolution (for example for oil J shown in panel (A) in Fig. 5). This specificity can be interpreted as the result of the formation of primary and secondary oxidation products which are expected to be very different in the various samples due to the heterogeneity of the oil cultivar and geographical origin. For the excitation wavelength λ_{ex} = 400 nm there is a general tendency in the increase of the peak at ca. 680 nm (panel (E) (oil P in panel (E) in Fig. 5), but such a behaviour is less pronounced in some oils (for example, oil J, panel (B) in Fig. 5). Significant changes in the fluorescence spectra were observed for the excitation wavelength λ_{ex} = 480 nm (panels (C) and (F) in Fig. 5). Here there is a consistent increase in the intensity, which is observed for all the oils investigated both in the region between 500 nm 625 nm and of the peak at ca. 680 nm. In some of the oils the shape of the spectra also show a significant change (for example oil P shown in panel (F) in 5). These results can be interpreted as the result of the formation of oxidation products, the difference in the absorption and emission characteristics of chlorophyll *a and b* and pheophytin *a* and *b* (Galeano Díaz, Durán Merás, Correa, Roldán and Rodríguez Cáceres, 2003) which may be affected differently by the thermal degradation, in combination with the inner filter effect (Torreblanca-Zanca et al., 2019).

The spectral regions where the fluorescence undergoes the strongest change can be visualized by calculating the difference of the fluorescence intensity at a given ageing step and in unaged condition. This is shown in Fig. 6 for two selected oils at two ageing steps, namely after 9 and 53 days. In both oils there is an increase (red) of the fluorescence intensity for excitations above 400 nm and emissions between 650 and 680 nm, whereas the intensity of the shoulder between 680 nm and 700 nm tends to decrease, particularly for excitations below 400 nm. Other smaller changes are more specific to the oil and reflect the different chemical composition and heterogeneity of the oils investigated.

3.3. Feature analysis

Using the relative error *RE* calculated according to Eq. (2), the most relevant wavelengths resulted $\lambda^{[1]}$ = 480 nm and $\lambda^{[2]}$ = 300 nm, which appeared 15 and 6 times respectively. The remaining three oils showed maximal *RE* at 320 nm (twice) and 450 nm (once). It is to be noted that for oils in which *RE* did not result highest at 480 nm, for example $\lambda_C^{[1]}$ = 450 nm, the second most relevant wavelength in terms of *RE* was always 480 nm. This motivated the selection $\lambda^{[1]}$ = 480 nm and $\lambda^{[2]}$ = 300 nm for all the oils. The results can be interpreted with the help of Fig. 5: for both $\lambda^{[1]}$ = 480 nm and $\lambda^{[2]}$ = 300 the spectra undergoes significant changes. Since the *RE* is normalized by the norm squared, the relative changes have a more important role than the absolute intensity, which is

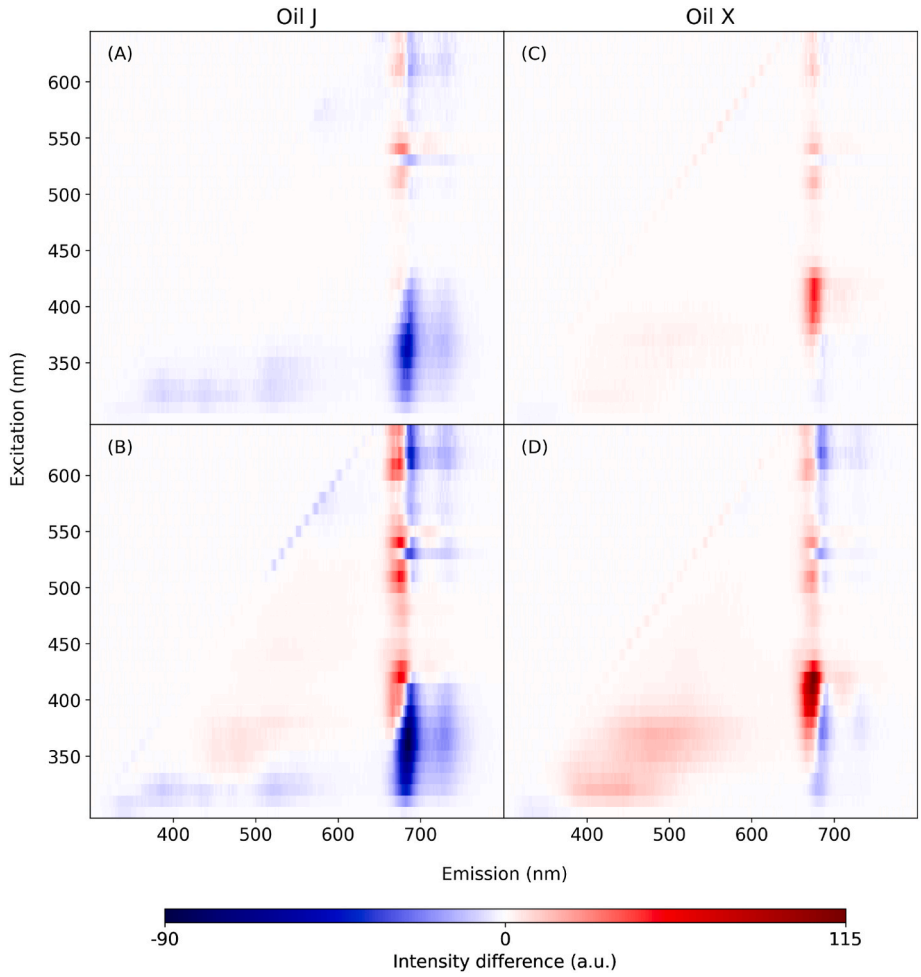

Fig. 6. Difference of the fluorescence intensity at selected ageing step and unaged for two selected oils. Red indicates positive values, blue negative ones. Left panels: oil J; right panels: oil X. Panels A, C: Difference of the intensity after 9 days and unaged; Panels B, D: Difference of the intensity after 53 days and unaged. The intensity is plotted in logarithmic scale to make differences more visible.

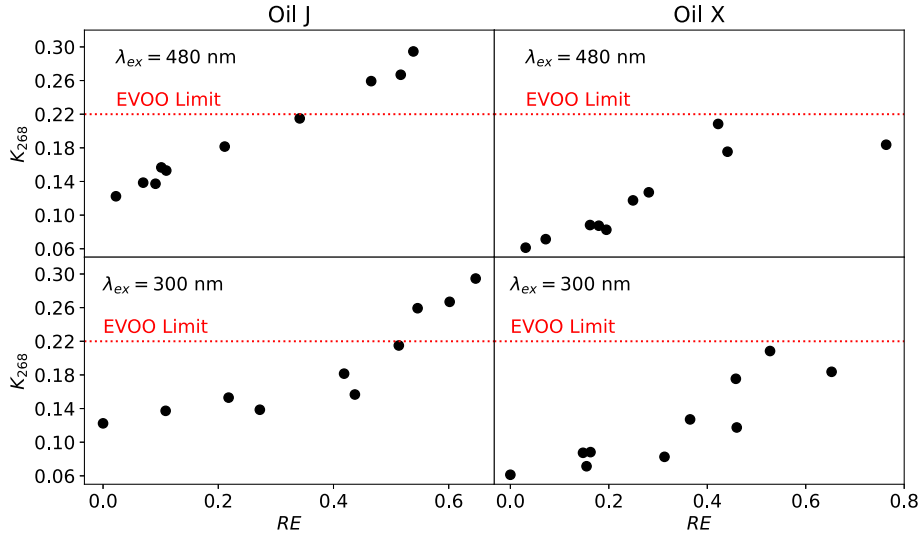

Fig. 7. Evolution of the UV parameter K_{268} versus the relative error with at the two wavelengths containing the most information. Left panels: oil X; right panels: oil j. Top panels: $\lambda^{[2]} = 300$ nm; $\lambda^{[1]} = 480$ nm. The limits allowed for EVOO are also shown in red.
(a) Results for the classification done with the class derived from the parameter K_{268}.
(b) Results for the classification done with the class derived from the parameter K_{232}.

much stronger for excitation of 400 nm (corresponding to the absorption of chlorophylls), as discussed in Section 3.2.

The *RE* allows to identify which excitation wavelength has the overall strongest impact upon ageing on the fluorescence spectrum. To be useable for the classification of oil quality, however, *RE* must correlate with changes in the UV parameters. Fig. 7 shows the coefficient K_{268} vs. relative errors RE (480) and RE (300) for two selected oils, and demonstrates that this is indeed the case. The evolution of the UV parameter K_{232} shows a similar behaviour (see Additional Material).

3.4. Predictive models

The performance of the model is quantified in terms of accuracy, sensitivity, and specificity (Hastie et al., 2009). The results are shown in Fig. 8, where the metrics are plotted for all the different training listed in Table 3 and for the classes obtained from the parameters K_{268} in panel (a) and K_{232} in panel (b), respectively. The results show that the model is able to detect oils that have passed the thresholds with a high accuracy, above 90%. As expected, the greater the size of the data set (or, in other words, the more aging steps employed during training), the better the performance becomes. When using the parameter K_{268} to label the class (see panel (a) in Fig. 8), already with three ageing steps (in particular 4,5 and 6) the performance is exceptional, with an accuracy of 90%. The results are consistent with the fact that the *RE* grows noticeably only after the first ageing steps, thus including step 6 give the model enough information on the evolution of the ageing process. This is not reflected when predicting the class from the parameters K_{232}. In this case the best results are obtained when using all available ageing steps (4,5,6,7 and 8) as can be seen from panel (b) in Fig. 8. This can be understood in term of the less sensitivity of this parameter to thermal oxidation as discussed previously.

The values of the metrics are summarized in Table 4 for clarity. The values in the table correspond to the training performed using the largest number of ageing steps. It should be noted that the metric values are the same in both cases. The reason is that the dataset has only 24 oils, thus the accuracy, for example, can only assume a discrete set of values (0, 1/24, 1/12, 3/24, etc.).

The results where the validation is done only on the next ageing step (*method 2* in Table 3) are shown in Fig. 9 for the class obtained from K_{268} and K_{232} in panels (a) and (b) respectively.

From Fig. 9, and from the comparison with Fig. 8, it can be concluded

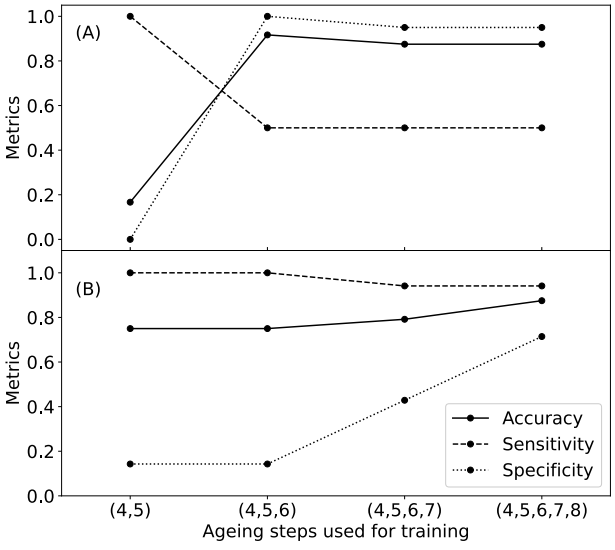

Fig. 8. Results for the prediction of the classes at ageing step 9 (Method 1). (A) Classification done with the class derived from the parameter K_{268}. (B) Classification done with the class derived from the parameter K_{232}.

Table 4
Metrics for the AdaBoost model described in the text. Note that the values are the same, since the number of oils that are over or under the threshold is very small and thus the metrics can assume a limited number of values.

Metric	Classes derived from K_{268}	Classes derived from K_{232}
Predictions on ageing step 9		
Accuracy	0.92	0.88
Sensitivity	0.5	0.94
Specificity	0.95	0.71
Predictions on next ageing step		
Accuracy	0.92	0.88
Sensitivity	0.5	0.94
Specificity	0.95	0.71

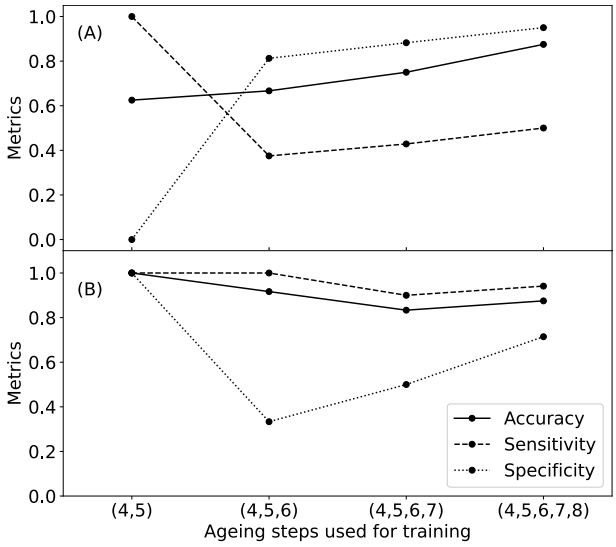

Fig. 9. Results for the prediction of the immediate next ageing step (Method 2). (A) Classification done with the class derived from the parameter K_{268}. (B) Classification done with the class derived from the parameter K_{232}.

that having enough ageing steps is a pre-requisite for a good prediction. The best results are obtained when using ageing steps 4,5,6,7 and 8 as input for the training.

The first point in Fig. 9 in panel (b) requires a discussion. All three metrics (accuracy, sensitivity, and specificity) are equal to 1. This would indicate a perfect prediction, but it is in reality to be attributed to a random effect due (most probably) to the specific data distribution. If this would be a real effect, this perfect prediction would also be visible when additional ageing steps are added to the training data. This is not clearly the case, as can be seen in panel (b) in Fig. 9. Thus, this point should not be considered as indicating that a model can predict the quality of oil perfectly.

4. Conclusions

In conclusion, this work makes several significant contributions to the study of the quality assessment of extra virgin olive oil (EVOO). Four key contributions can be identified.

Firstly, the work involves extensive UV-absorption spectroscopy measurements at the specific wavelengths required for the quality assessment specified by the European Regulations over time for a range of commercially available EVOOs. The measurements of three parameters (K_{232}, K_{268}, and ΔK) provide insights into the oxidation processes giving rise to absorption in the UV range, which is crucial for assessing EVOO's quality.

Secondly, the study conducts extensive total fluorescence spectroscopy analysis of the same oil samples through the acquisition of

excitation-emission matrices (EMMs). By measuring the fluorescence intensity at different excitation and emission wavelengths, the study explores the fluorescence characteristics of EVOOs during different ageing stages.

Thirdly, the research identifies the two excitation wavelengths (480 nm and 300 nm) that exhibit the maximum relative change in fluorescence for the majority of the oils. This finding highlights specific wavelengths that are highly informative in terms of assessing EVOO quality based on fluorescence properties.

Lastly, the work proposes a method for the prediction of olive oil quality based on aggregated data. By considering the fluorescence intensity at the two identified excitation wavelengths, the study employs machine learning algorithms (such as AdaBoost, Random Forest, Logistic Regression, and Naïve Bayes) to classify EVOOs as either extra virgin or non-extra virgin based on the UV-spectroscopy criteria defined by the European regulations with an accuracy above 90%. The advantage of the proposed method, is that it does not require a spectrometer or spectral analysis for the classification of EVOO: the realization of a portable device would only require two LEDs for the excitation at the two identified wavelengths, and one photodiode for the collection of the integrated fluorescence spectrum. Also, being the oils used in this study very heterogeneous in olive cultivar, geographical origin and advertised quality, the authors expect the results to be quite general in validity, thus with a broad range of applicability. The predictive models, particularly based on the fluorescence spectra described here, carry significant practical implications in the field. Potential applications include the early detection of oxidation or adulteration, enabling producers to take corrective actions promptly and maintain product quality. Additionally, these models may aid in classifying olive oils into different quality grades, facilitating sorting and pricing strategies. However, it's crucial to acknowledge certain limitations. The models heavily rely on the quality and representativeness of the training data, necessitating a diverse dataset that encompasses various olive oil types and conditions. Furthermore, the predictive power of the models may be constrained by factors not accounted for in the training data, such as variations in regional characteristics or production methods.

Future research directions of this study may be the expansion of the analysis to include a larger number of oils, additional analytical parameters used for quality control or the sensory analysis of the oils during the oxidation. Additionally, the realization of a compact, portable fluorescence-based instrument could be realized to test the proposed method in field.

Overall, the contributions of this work advance our understanding of EVOO quality assessment by combining UV-absorption spectroscopy, fluorescence analysis, and machine learning techniques. The findings provide valuable insights for the development of low-cost and portable methods for assessing the quality of EVOO, which can have significant implications for the olive oil industry and consumers.

Funding

This research was supported by the Hasler Foundation project "ARES: AI for fluoREscence Spectroscopy in oil".

CRediT authorship contribution statement

Francesca Venturini: Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization. **Silvan Fluri:** Investigation. **Manas Mejari:** Software, Methodology, Conceptualization. **Michael Baumgartner:** Investigation. **Dario Piga:** Writing – review & editing, Conceptualization. **Umberto Michelucci:** Writing – review & editing, Writing – original draft, Supervision, Software, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

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

Data availability

The data is available and linked in the paper.

References

Al Riza, D. F., Kondo, N., Catalano, P., & Giametta, F. (2019). A preliminary study on the potential of front face fluorescence spectroscopy for Italian mono-cultivar extra virgin olive oil discrimination. *Journal of Agricultural Engineering, 50*, 5–11.

Al Riza, D. F., Kondo, N., Rotich, V. K., Perone, C., & Giametta, F. (2021). Cultivar and geographical origin authentication of Italian extra virgin olive oil using front-face fluorescence spectroscopy and chemometrics. *Food Control, 121*, Article 107604.

Aparicio Ruiz, R., Tena Pajuelo, N., Romero del Río, I., García-González, D. L., & Morales Millán, M. T. (2017). Predicting extra virgin olive oil freshness during storage by fluorescence spectroscopy. *Grasas Y Aceites, e219*(68), 4.

Aparicio-Ruiz, R., & Gandul-Rojas, B. (2012). Thermal degradation kinetics of neoxanthin, violaxanthin, and antheraxanthin in virgin olive oils. *Journal of Agricultural and Food Chemistry, 60*, 5180–5191.

Aparicio-Ruiz, R., Minguez-Mosquera, M. I., & Gandul-Rojas, B. (2010). Thermal degradation kinetics of chlorophyll pigments in virgin olive oils. 1. compounds of series a. *Journal of Agricultural and Food Chemistry, 58*, 6200–6208.

Baltazar, P., Hernández-Sánchez, N., Diezma, B., & Lleó, L. (2020). Development of rapid extra virgin olive oil quality assessment procedures based on spectroscopic techniques. *Agronomy, 10*, 41.

Botosoa, E. P., Chèné, C., & Karoui, R. (2021). Front face fluorescence spectroscopy combined with PLS-DA allows to monitor chemical changes of edible vegetable oils during storage at 60° C. *European Journal of Lipid Science and Technology, 123*, Article 2000088.

Camerlingo, C., Portaccio, M., Delfino, I., Lepore, M., et al. (2019). Surface-enhanced Raman spectroscopy for monitoring extravirgin olive oil bioactive components. *Journal of Chemistry*, 1–10, 2019.

Conte, L., Milani, A., Calligaris, S., Rovellini, P., Lucci, P., & Nicoli, M. C. (2020). Temperature dependence of oxidation kinetics of extra virgin olive oil (evoo) and shelf-life prediction. *Foods, 9*, 295.

El Orche, A., Bouatia, M., & Mbarki, M. (2020). Rapid analytical method to characterize the freshness of olive oils using fluorescence spectroscopy and chemometric algorithms, 2020 *Journal of Analytical Methods in Chemistry*. https://doi.org/10.1155/2020/8860161. article ID 8860161, 9 pages.

Escudero, A., Ramos, N., La Rubia, M. D., & Pacheco, R. (2016). Influence of extreme storage conditions on extra virgin olive oil parameters: Traceability study, 2016 *Journal of Analytical Methods in Chemistry*. https://doi.org/10.1155/2016/7506807. article ID 7506807, 10 pages.

Esposto, S., Taticchi, A., Urbani, S., Selvaggini, R., Veneziani, G., Di Maio, I., et al. (2017). Effect of light exposure on the quality of extra virgin olive oils according to their chemical composition. *Food Chemistry, 229*, 726–733.

Evans, C., List, G., Moser, H. A., & Cowan, J. (1973). Long term storage of soybean and cottonseed salad oils. *Journal of the American Oil Chemists' Society, 50*, 218–222.

Fadda, C., Del Caro, A., Sanguinetti, A. M., Urgeghe, P. P., Vacca, V., Arca, P., et al. (2012). Changes during storage of quality parameters and in vitro antioxidant activity of extra virgin monovarietal oils obtained with two extraction technologies. *Food Chemistry, 134*, 1542–1548.

Fan, J., & Yao, Q. (2003). *Nonlinear time series: Nonparametric and parametric methods* (Vol. 20). Springer.

Frankel, E. N. (2011). Nutritional and biological properties of extra virgin olive oil. *Journal of Agricultural and Food Chemistry, 59*, 785–792.

Galeano Díaz, T., Durán Merás, I., Correa, C. A., Roldán, B., & Rodríguez Cáceres, M. I. (2003). Simultaneous fluorometric determination of chlorophylls a and b and pheophytins a and b in olive oil by partial least-squares calibration. *Journal of Agricultural and Food Chemistry, 51*, 6934–6940.

Gómez-Alonso, S., Salvador, M., & Fregapane, G. (2004). Evolution of the oxidation process in olive oil triacylglycerol under accelerated storage conditions (40–60 c). *Journal of the American Oil Chemists Society, 81*, 177–184.

Guzmán, E., Baeten, V., Pierna, J. A. F., & García-Mesa, J. A. (2015). Evaluation of the overall quality of olive oil using fluorescence spectroscopy. *Food Chemistry, 173*, 927–934.

Hastie, T., Rosset, S., Zhu, J., & Zou, H. (2009). Multi-class adaboost. *Statistics and Its Interface, 2*, 349–360.

Iqdiam, B. M., Welt, B. A., Goodrich-Schneider, R., Sims, C. A., Baker IV, G. L., & Marshall, M. R. (2020). Influence of headspace oxygen on quality and shelf life of extra virgin olive oil during storage. *Food Packaging and Shelf Life, 23*, Article 100433.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Karoui, R., & Blecker, C. (2011). Fluorescence spectroscopy measurement for quality assessment of food systems—a review. *Food and Bioprocess Technology, 4*, 364–386.

Lam, H. Y., Roy, P. K., & Chattopadhyay, S. (2020). Thermal degradation in edible oils by surface enhanced Raman spectroscopy calibrated with iodine values. *Vibrational Spectroscopy, 106*, Article 103018.

Lobo-Prieto, A., Tena, N., Aparicio-Ruiz, R., García-González, D. L., & Sikorska, E. (2020). Monitoring virgin olive oil shelf-life by fluorescence spectroscopy and sensory characteristics: A multidimensional study carried out under simulated market conditions. *Foods, 9*, 1846.

Lopes, C. R. B., & Courrol, L. C. (2023). Evaluation of steady-state and time-resolved fluorescence spectroscopy as a method for assessing the impact of photo-oxidation on refined soybean oils. *Foods, 12*, 1862.

Mancebo-Campos, V., Fregapane, G., & Desamparados Salvador, M. (2008). Kinetic study for the development of an accelerated oxidative stability test to estimate virgin olive oil potential shelf life. *European Journal of Lipid Science and Technology, 110*, 969–976.

McQuarrie, A. D., & Tsai, C. L. (1998). *Regression and time series model selection*. World Scientific.

Morales, M. T., & Przybylski, R. (2013). Olive oil oxidation. In R. Aparicio, & J. Harwood (Eds.), *Handbook of olive oil* (pp. 479–522). Boston, MA: Springer US. URL: http://link.springer.com/10.1007/978-1-4614-7777-8_13, 10.1007/978-1-4614-7777-8_13.

Ordukaya, E., & Karlik, B. (2017). Quality control of olive oils using machine learning and electronic nose. *Journal of Food Quality*, 1–7, 2017.

Sikorska, E., Khmelinskii, I., & Sikorski, M. (2012). Analysis of olive oils by fluorescence spectroscopy: Methods and applications. *Olive oil-constituents, quality, health properties and bioconversions*, 63–88.

Sikorska, E., Khmelinskii, I. V., Sikorski, M., Caponio, F., Bilancia, M. T., Pasqualone, A., et al. (2008). Fluorescence spectroscopy in monitoring of extra virgin olive oil during storage. *International Journal of Food Science and Technology, 43*, 52–61.

Skoog, D. A., Holler, F. J., & Crouch, S. R. (2017). *Principles of instrumental analysis*. Cengage learning.

Taiti, C., Marone, E., Fiorino, P., & Mancuso, S. (2022). The olive oil dilemma: To be or not to be evoo? Chemometric analysis to grade virgin olive oils using 792 fingerprints from ptr-tof-ms. *Food Control, 135*, Article 108817.

Torreblanca-Zanca, A., Aroca-Santos, R., Lastra-Mejias, M., Izquierdo, M., Cancilla, J. C., & Torrecilla, J. S. (2019). Laser diode induced excitation of pdo extra virgin olive oils for cognitive authentication and fraud detection. *Sensors and Actuators B: Chemical, 280*, 1–9.

Union, E. (1991). Commission regulation (eec) no. 2568/91 of 11 july 1991 on the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis official journal l 248, 5 september 1991. *Offic. JL, 248*, 1–83.

Union, E. (2013). Commission implementing regulation no 1348/2013 of december 17 2013. *Official Journal of the European Union, 338*, 31–67.

Vega-Márquez, B., Nepomuceno-Chamorro, I., Jurado-Campos, N., & Rubio-Escudero, C. (2020). Deep learning techniques to improve the performance of olive oil classification. *Frontiers of Chemistry, 7*, 929.

Venturini, F., Fluri, S., & Baumgartner, M. (2023). Dataset of fluorescence eem and uv spectroscopy data of olive oils during ageing. *Mendeley Data*. https://doi.org/10.17632/g6y69g8gwm.1

Venturini, F., Fluri, S., & Baumgartner, M. (2023). Dataset of fluorescence eem and uv spectroscopy data of olive oils during ageing. *Data, 8*, 1–6.

Venturini, F., Sperti, M., Michelucci, U., Herzig, I., Baumgartner, M., Caballero, J. P., et al. (2021). Exploration of Spanish olive oil quality with a miniaturized low-cost fluorescence sensor and machine learning techniques. *Foods, 10*, 1010.

Zandomeneghi, M., Carbonaro, L., & Caffarata, C. (2005). Fluorescence of vegetable oils: Olive oils. *Journal of Agricultural and Food Chemistry, 53*, 759–766.

Zaroual, H., Chénè, C., El Hadrami, E. M., & Karoui, R. (2022). Application of new emerging techniques in combination with classical methods for the determination of the quality and authenticity of olive oil: A review. *Critical Reviews in Food Science and Nutrition, 62*, 4526–4549.