Analytica Chimica Acta 1280 (2023) 341884

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

Analysing olive ripening with digital image RGB histograms

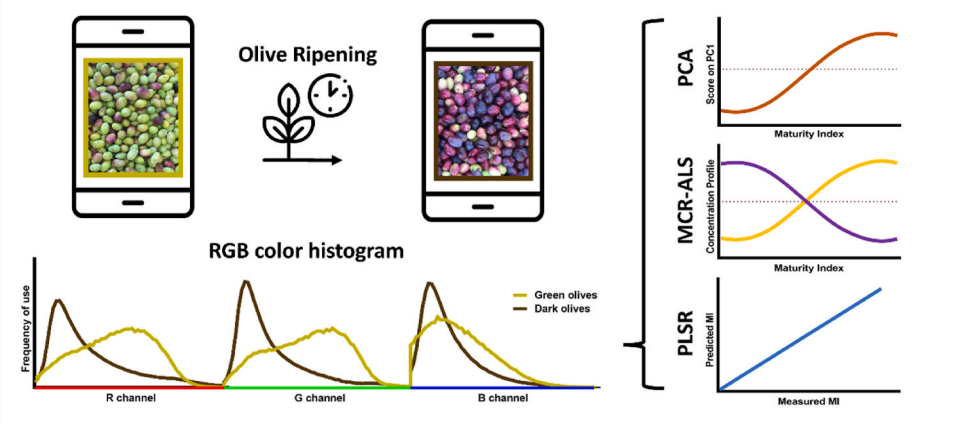
Jokin Ezenarro [a], Ángel García-Pizarro [a,b], Olga Busto [a], Anna de Juan [c], Ricard Boqué [a,*]

[a] Universitat Rovira i Virgili. Chemometrics and Sensorics for Analytical Solutions (CHEMOSENS) group, Department of Analytical Chemistry and Organic Chemistry, Campus Sescelades, Edifici N4, C/Marcel·lí Domingo 1, Tarragona, 43007, Spain
[b] Fruit Production Program, IRTA Mas Bové, Ctra. Reus-El Morell Km 3.8, Constantí, 43120, Spain
[c] Universitat de Barcelona. Chemometrics Group, Dept. of Chemical Engineering and Analytical Chemistry, Martí i Franqués 1, 08028, Barcelona, Spain

HIGHLIGHTS

- Digital images objectively analyse color changes for olive ripening.
- CACHAS offers reproducible MI assessment, replacing visual inspection.
- R, G, and B histograms determine MI of olives in CACHAS.
- A methodology potentially applicable to other fruits is proposed.
- Study confirms digital images' capability for accurate fruit ripeness prediction.

GRAPHICAL ABSTRACT



ARTICLE INFO

Handling Editor: Prof. L. Buydens

Keywords:
PCA
PLS
Multivariate curve resolution
MCR-ALS
Chemometrics Assisted Colour Histogram-based Analytical Systems (CACHAS)
Fruit
Maturity index
Cell phone

ABSTRACT

Digital images are commonly used to monitor processes that are based on colour changes due to their simplicity and easy capture. Colour information in these images can be analysed objectively and accurately using colour histograms. One such process is olive ripening, which is characterized by changes in chemical composition, sensory properties and can be followed by changes in physical appearance, mainly colour. The reference method to quantify the ripeness of olives is the Maturity Index (MI), which is determined by trained experts assigning individual olives into a colour scale through visual inspection. Instead, this study proposes a methodology based on Chemometrics Assisted Colour Histogram-based Analytical Systems (CACHAS) to automatically assess the MI of olives based on R, G, and B colour histograms derived from digital images. The methodology was shown to be easily transferable for routine analysis and capable of controlling the ripening of olives. The study also confirms the high potential of digital images to understand the ripening process of olives (and potentially other fruits) and to predict the MI with satisfactory accuracy, providing an objective and reproducible alternative to visual inspection of trained experts.

* Corresponding author.
E-mail address: ricard.boque@urv.cat (R. Boqué).

https://doi.org/10.1016/j.aca.2023.341884
Received 31 March 2023; Received in revised form 2 October 2023; Accepted 7 October 2023
Available online 9 October 2023

1. Introduction

Digital images have become a popular method to monitor processes since they are simple and easy to capture and can provide a measurement that reflects adequately phenomena that can be easily observed by the human eye. Besides, digital cameras are found in most commonly used electronic devices, which makes this technology available for any potential user. Digital image analysis has already been applied in food and agriculture for classification of samples of tea, honey or grain, and in other areas such as biomedicine and microbiology for the detection of cancer cells or different types of bacteria and yeast, even as a detector for biosensors, determining allergens in food, or other biomolecules using ELISA assays, which makes it a high-potential technology with multiple applications yet to be investigated [1,2].

A digital image is formed by a collection of small units of information called pixels. Each pixel is defined by two spatial coordinates (x and y) with associated colour information, represented in a three-dimensional colour space, typically Red, Green and Blue (RGB). This means that any colour can be represented by a linear combination of these three basic colours. Additionally, images can be stored in various file formats, such as JPEG or PNG, and can be easily shared or transferred digitally. Even if the images using the JPEG format are compressed in a lossy way, this is, some information about the pixels is lost in the process, and the ones using PNG are lossless; it does not affect the final results as in these images the noise is not significant and even with a lossy compression the information of interest is conserved [3].

Data analysis provides many tools for the interpretation of the colour information in an objective and accurate manner. One of the most common ways to study the colour information of an image is the use of a colour histogram, made representing the frequency of the R, G, and B values in bin intervals that cover the full colour scale (0–255) for the ensemble of pixels of the Region of Interest (ROI) in the image. These histograms are really useful when the colour distribution of a heterogeneous image is studied, since not only the main or average colour in every RGB coordinate is measured but the variability of these color values across the full image. There are several analytical approaches to study the colour histograms, e.g., applying first order statistics to the RGB histograms, combining the RGB histograms with other colour-spaces (grayscale, HSV, CMYK …) where the result is called "colour-gram" [4], or using just the RGB histograms as a multivariate data input for further chemometric analysis [1].

Food chemistry is a field where changes of colour are often related to natural phenomena affecting the characteristics and the quality of food products. One of the processes where colour plays a major role is fruit ripening, where it can be the most important indicator of maturity and quality in many fruit species [5]. The colour of a fruit, particularly olives, is mainly influenced by the concentration and distribution of various anthocyanins, chlorophylls and carotenoids in the skin and flesh, which evolve throughout the ripening process [6]. Olive ripening is a complex and dynamic process that occurs during the final stages of the development of the fruit, and it is characterized not only by changes in colour but by a series of changes in the olive physical aspect, chemical composition and sensory properties. These changes are influenced by several environmental and genetic factors, including temperature, light exposure, watering and olive cultivar [7,8].

The ripening process starts when the olive fruit reaches full size and begins to change colour, typically from green to red or black. Over time, the skin and flesh of the fruit soften, and its flavour and aroma become more pronounced [9]. These changes are accompanied by a series of biochemical reactions, including the synthesis and breakdown of pigments, fatty acids and volatile compounds [10]. Ripe olives contain a higher percentage of oil and a lower percentage of water, which results in a higher yield of oil per batch of olives. Additionally, ripe olives also have a milder, fruitier flavour compared to unripe olives, which can be bitter. Therefore, assessing the ripeness of olives is crucial to ensure that only the highest quality olives are used to produce oil, resulting in a better tasting and higher quality oil [6].

The Maturity Index (MI) is a common parameter to quantify the maturity of olives and, thus, to determine their stage of ripeness [11]. This index is based on the perception of trained experts, which assign a colour value between 0 and 7 to every individual fruit of a set of randomly sampled olives. The MI is finally obtained by doing a mean of the colour values assigned to each individual olive of the set studied. Since MI is a parameter purely based on colour, the colour captured on digital images of olives is expected to relate to their MI; and a method to assess the MI based on this could be a better way to carry out the analysis since it is more objective and reproducible.

Methodologies to automatically assess the MI of olives based on infrared spectroscopy and digital images have been proposed [12]. Some of them are based on studying the pixels of individual olives one by one, assigning them a value in the colour scale to calculate the MI [13,14]. These approaches have the inconvenience of having to image olives separately or having to implement an automatic shape recognition algorithm or doing a manual separation of olives before acquiring the image of the total olive set. Other authors have analysed the percentage of black olives in a collective image applying a k-nearest-neighbours strategy on individual pixels [15].

Instead, this article proposes a methodology based on Chemometrics Assisted Colour Histogram-based Analytical Systems (CACHAS). R, G and B colour histograms can be derived from the single RGB image acquired on the set of randomly selected olives without the need of separating the individual fruits, in this way the whole population of olives in the sample is represented in the analytical signal and the heterogeneity of the sample is considered. In addition, the histograms thus obtained can be straightforwardly used as multivariate inputs to apply different chemometric techniques. As a first step, an exploratory approach applying principal component analysis (PCA) was used to study trends in the data. Then, multivariate curve resolution was applied as an unsupervised method oriented to model the evolution and the olive color change associated with the ripening process based on the use of color histograms obtained as a function of the ripening time. Finally, a partial least squares regression model was applied to assess the potential of quantitative estimation of the maturity index from the information in the color histograms of the analysed samples. The methodologies proposed allow assessing the ripening stage of olives (and potentially other fruits) using the images acquired by any user, making the approach easily transferable for real routine purposes, and providing a fast and *in situ* method to control the cultivation and processing of olives, ultimately benefiting both farmers and consumers.

2. Material and methods

2.1. Olives

Two sets of olives have been studied: data set A formed by olives of different varieties and data set B, formed by olives of a single variety monitored over time. Data set A is formed by seven varieties of olives grown in Catalonia, Spain, which are '*Arbequina*', '*Coratina*', '*Corbella*', '*Empeltre*', '*Koroneiki*', '*Morrut*' and '*Picual*' (see Fig. 1a). For each variety, several samples from the Camp de Tarragona production area (Spain) were analysed in three different ripening stages differenced by approximately three weeks. Samples from the year 2021 are included in the calibration set and samples from the year 2022 are used for validation purposes.

Since olives are a non-climacteric fruit, this is, they do not ripen once they are collected [16], they were harvested directly from the tree in each sampling point. A sample consists of a few hundreds of olives randomly collected from different positions in a single tree, placed in a box to be subsequently imaged.

Data set B was formed by a single variety of olive, '*Arbequina*', for which 120 samples (collected in the same way) were analysed over three months in different ripening stages (see Fig. 1b). The samples were

collected in 12 different olive mills from the Garrigues (Lleida, Spain) designation of origin, randomly at different moments in each mill to ensure that the full ripening process is covered and the inter-mill variability is considered.

2.2. Digital images

The digital images of the data set B, including the same olive variety from the 12 olive mills, were obtained once the sample arrived at the mill (indoors), by placing the sampled olives in a box and taking the image from above at around 30 cm. Neither position, distance nor lighting were controlled since the images were taken by mill workers for regular routine control purposes and not specifically for this study. This means the images obtained have different number of olives, different number of pixels that contain olives and different lighting conditions. The images acquired were taken with several Zebra ET56 tablet computers (Zebra Technologies Corporation, Illinois USA), all with the same model of tablet but different devices. Since the images were taken to have a visual register of a routine analysis of the olives, not only olives but also labels, leaves and the corners of the boxes and the table can appear in the images. Therefore, the ROI of the images has been selected by cropping them to contain mainly olives.

On the other hand, the images of data set A, including the set of several olive varieties, were taken with an iPhone 6 plus (Apple Inc., California USA) in 2021 and with an iPhone 12 in 2022. In this case, since the images were taken after removing impurities from the olives (as it can be seen in Fig. S4), the images did not need to be cropped.

2.3. Colour histogram

To obtain the colour histogram of each image, the frequency of occurrence of each value between 0 and 255 in the R, G and B channels is represented. Other channels as HSV were tried obtaining suboptimal models, so the results are not shown. The bin of the histograms was augmented to four units to avoid the presence of null frequencies at some particular R, G and B values that may cause a distortion in the global form of the histogram (see Fig. S2). This action has a smoothing effect and removes zero values. In all cases, since the total number of pixels may differ between images due to the cropping step, the histograms were normalized by their Euclidean norm (2-norm). The normalization was carried out after concatenating the R, G and B histograms one after the other, that is, by following a low-level data fusion strategy, as shown in Fig. 2 [17].

In total, two datasets (**X**-blocks) were built using this data processing, one for each olive set described. Both datasets have samples in rows and values of R, G and B channels in columns. There are 189 variables in total, as there are 768 RGB values but each bin covers an interval of four values and the last bin of each colour channel was removed because it contains the white pixels that are caused by direct light reflection and do not contain any information of interest.

2.4. Maturity index

To assign a MI to each one of the collected samples for both data sets, a group of trained experts applied, the reference method for assessing olive ripening based on the method proposed by Uceda and Frías (1975) [11]. Thus, 50 olives were taken randomly from the box and each olive was assigned a value between 0 and 7 based on a colour scale of skin and flesh, where 0 is not mature at all and 7 is fully matured. Then, MI was calculated as:

$$\mathrm{MI} = \frac{\sum_{i=0}^{7} N_i * i}{N} \qquad \text{Eq. 1}$$

where i is the value assigned in the scale, N_i is the number of olives with value i assigned and N is the total number of olives analysed.

2.5. Chemometric algorithms

Based on the information contained in the colour histograms, the potential of several algorithms to explore and interpret the olive ripening process was tested. First, Principal Component analysis (PCA) was applied as it is one of the most general unsupervised data exploration methods and points out to general trends occurring in any data set (presence of outliers, detection of relevant color variables and process points). Then, Multivariate Curve Resolution (MCR) was used to model in a much more interpretable way the process evolution and the color change associated with the different ripening stages of olives taking advantage of the inclusion of some general characteristics known about the process evolution and the color histograms under the form of constraints. Finally, the feasibility to predict the MI using color information was tested using the Partial Least Squares Regression method, since it is

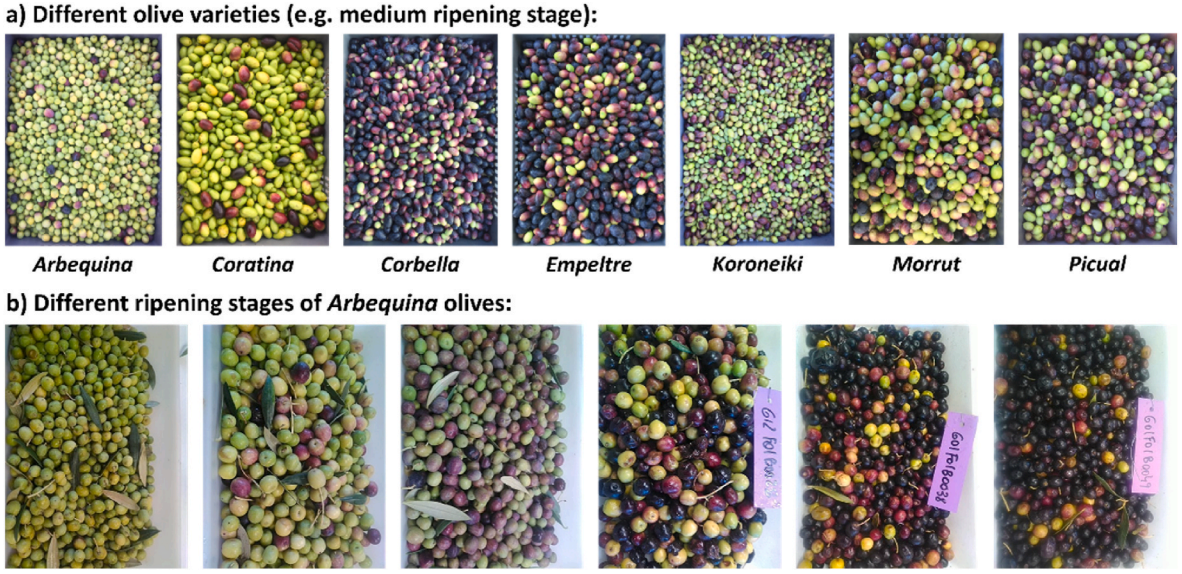


Fig. 1. Examples of digital images from the datasets used in the present work. a) Data set A: images of the seven olive varieties in a medium ripening stage. b) Data set B: images of *Arbequina* olives ripening during time.

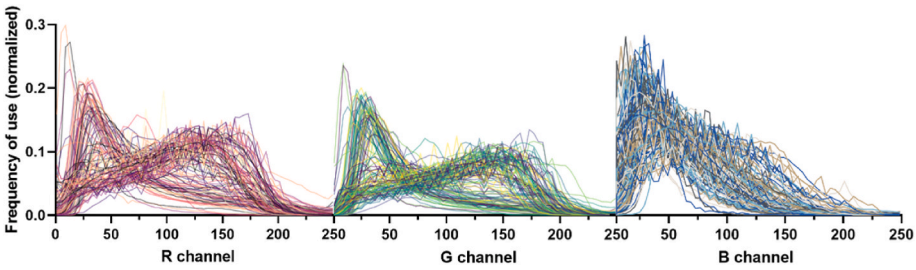


Fig. 2. RGB histograms (bin 4) of the images from data set A after low-level data fusion and normalization.

an algorithm that intends to maximize the covariance between the colourgram and MI, finding the model that best correlates both kinds of information.

Principal Component Analysis (PCA) is a statistical technique used to analyse the information contained in data tables formed by a set of samples described by the measurement of a set of variables. It is often used as a dimensionality reduction method, meaning that it reduces the complexity of the data by transforming it into a lower-dimensional space in order to help identifying patterns in the data. In PCA, the relevant information of the variables of the original data set are expressed by a set of new, uncorrelated variables called Principal Components (PCs). These principal components are found by looking for the directions of maximum variance of the original data. Afterwards, samples can be projected in this new dimensional space to be easily visualized [18,19]. A PCA model is mathematically described as:

$$\mathbf{X} = \mathbf{TP}^{\mathbf{T}} + \mathbf{E} \quad \text{Eq. 2}$$

where **X** is the matrix of original data (colour histograms of the different samples in our case), **T** are the scores matrix, which allows visualizing the samples in the principal component (PC) space and $\mathbf{P}^{\mathbf{T}}$ the loadings matrix, which allow representing the original variables in the PC space. **E** is the variance unexplained by the PCA model.

Multivariate Curve Resolution - Alternating Least Squares (MCR-ALS) is a statistical technique used to decompose a matrix of multi-component data into its individual components. The original data (**X**-block) is decomposed in concentration profiles (**C**) and spectral (response) profiles (**S**) in the following way:

$$\mathbf{X} = \mathbf{CS}^{\mathbf{T}} + \mathbf{E} \quad \text{Eq. 3}$$

where **E** is the residual matrix. This decomposition is carried out iteratively under suitable constraints, such as non-negativity, unimodality … that help to recover chemically meaningful concentration and response profiles, readily interpretable as the pure components of the initial data set. MCR-ALS is often used to characterize qualitative and quantitatively the individual components in complex mixtures, and to identify and understand the evolution of components in a process [20,21].

The Partial Least Squares (PLS) method seeks to calculate latent variables that maximize the covariance between the information gathered in an **X**-block, e.g. colour histograms, and the reference values of the dependent variable (**Y**-block), e,g. MI. These LVs can be used to build a Regression model (PLSR) to predict the value of the dependent variable **Y** from the measured information in the **X**-block. This method is particularly useful in situations where the **X**-block contains a large number of correlated variables, or when the nature of the samples studied and the relationship between **X** and **Y** is complex. PLSR can be used to predict the value of the dependent variable with great accuracy and to identify the most important variables to carry out the prediction task [22,23].

These three techniques are used to analyse and interpret complex data sets in order to understand the relationships between variables, identify patterns in the data, or make predictions. All software used ran under MATLAB environment. Routines in MATLAB 9.11 were used for data processing (image cropping, data arrangement and data fusion), the PLS_toolbox 9.0 (Eigenvector Research) for PCA and PLSR and the free downloadable MCR-ALS GUI 2.0 for MCR-ALS analysis [20].

3. Results and discussion

3.1. Principal component analysis

PCA was conducted on data set B, formed by the colour histograms of the images related to a single olive variety collected in 12 different mills as a function of time, after mean centring the data. The goal was understanding the main trends linked to the olive ripening process and only the first component was needed (which explains 52.70 % of the variance), since the following ones did not add any interpretable information about the ripening process. The results showed that the evolution of the scores of the first component of the RGB histogram are directly related to the ripening of the olives, as it can be seen when the scores are represented against the MI assigned to the different images (Fig. 3a). This happens because as olives ripen, their colour evolves from green to black, and this is the main source of variability in the image histograms, as it is reflected in the loadings of the first component (Fig. 3b). The loadings show positive peaks related to the dark RGB values (closer to zero), which can be related to the colour (red-black) obtained from an increasing concentration of anthocyanins during the ripening process, and clear negative peaks at higher values of Green and Red in the RGB scale, which can be related to the colour (yellow-green) obtained from a decreasing concentration of chlorophylls and carotenoids during the ripening process [6]. The blue channel does not show a very significant variation during ripening, seen through the lower intensity of their related loading values.

The olive ripening, as most biological processes, follows a sigmoid curve, where the properties of the green fruit stay constant for a period

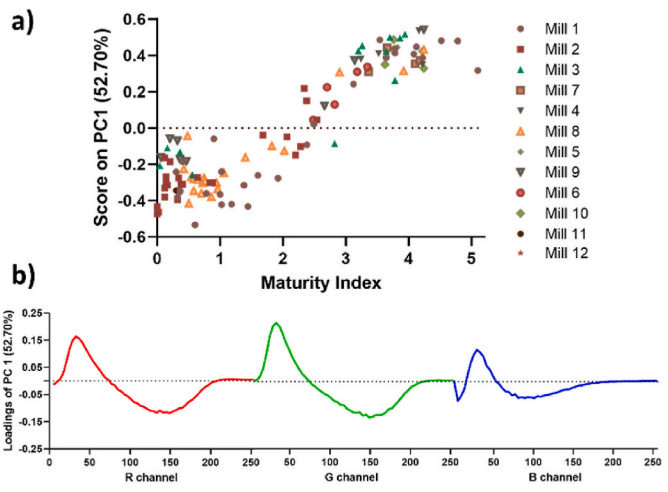


Fig. 3. First Principal Component (PC1) of the PCA performed on data set B. a) Scores of the RGB histograms vs. their maturity index. b) Loadings of each of the RGB values of the histograms.

of time, evolve gradually into the ripe form, to be stabilized when maturity has been reached. This pattern is well described with the scores of PC1, which means that the variation of the colour histograms reflects adequately the ripening process. In addition, it can be observed that there is no difference among the evolution shown by images acquired on samples from different mills, since all of them follow the same sigmoid pattern. Therefore, the ripening curve observed in this PCA does not differ according to the olive mills. On one hand, this means that ripening happens in a similar manner among olives of the same designation of origin; on the other hand, it also implies that the scale variations of colour histograms due to differences in image acquisition (e.g., different user, number of pixels containing olives, illumination conditions, …) are properly corrected by the histogram normalization.

3.2. Multivariate curve resolution

When it comes to process modelling, MCR is a better tool since the concentration profiles and pure responses can be constrained using the information obtained from experience and literature about the system under study. When studying the ripening process monitored in data set B by MCR-ALS, the initial data are the table of colour histograms ordered according to their related MI. The MCR model provides concentration profiles (**C**) related to the evolution of the ripening process and response profiles ($\mathbf{S}^{\mathbf{T}}$), which will be the colour histograms representing the components linked to the different stages of the ripening process. By calculating the eigenvalues using a Singular Value Decomposition (SVD) on the initial matrix of colour histograms, it was concluded that the ripening process can be modelled using two components. Then, a (SIMPLe-to-use Interactive Self-modelling Mixture Analysis) SIMPLISMA-based purest variable detection method was applied to obtain the spectral profiles to be used as initial guesses in the iterative optimization of the MCR-ALS algorithm [24]. Since the concentration profiles of a process and the related pure colour histograms cannot be negative, a non-negativity constraint was applied both to the spectral and concentration profiles. Additionally, since the samples were sorted in increasing MI order in the input matrix **X**, the unimodality constraint was applied to the concentration profiles reflecting that olive ripening is a continuously progressive process, i.e, when the olives ripen, they cannot go back in the maturity stage. Therefore, the concentration profiles should have a single maximum. However, since ripening is a biological process and the data used correspond to different olives at the different maturity stages and involves images taken in a non-controlled environment, a mild deviation from a perfect unimodal shape in the concentration profiles was allowed.

The concentration profiles obtained from the MCR-ALS model are plotted against the related MI values (Fig. 4) to study the evolving curves of the two components during the ripening process. Component 1 (purple) is associated with unripe olives and component 2 (orange) to ripe olives, as it can be concluded from the evolution of their related concentration profiles. The sigmoid curve derived from the concentration profile of the ripe olives can be used to accurately model the ripening process of olives and to understand the time at which the olives will reach their peak ripeness. Additionally, by analysing the shape and slope of the curve, it may be potentially possible to identify factors that affect the rate of ripening, such as temperature and humidity, and develop strategies for optimizing the process.

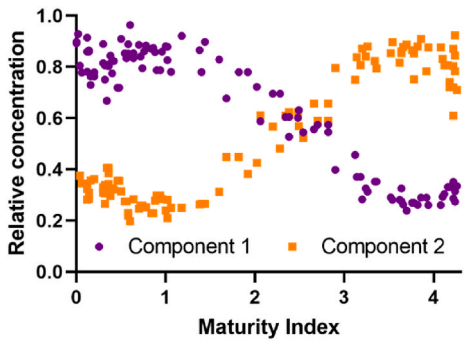


Fig. 4. Evolution of the relative concentrations of the two components derived from the MCR-ALS decomposition of the colour histograms of the olive ripening process monitored in data set B.

When olives ripen, the concentration of pigments such as anthocyanins increases since the oil accumulates in the fruit, while the concentration of other pigments, such as chlorophylls and carotenoids, progressively decreases due to the reduction of photosynthetic activity [6]. This shift in pigment composition results in a change in the colour of the olives from green to red or black. Since the concentration of the pigments that give olives their green colour decreases, the curve will gradually descend, eventually reaching a point where the concentration of the pigments that give olives their dark colour dominates. Such an evolution of the olive colour when ripening is clearly seen in the pure colour histograms associated with the components 1 and 2 (see Fig. 5). It is interesting to notice that only two components with evolving intensities (concentrations) during the ripening process are needed to describe the olive ripening process. This fact can help us to discard the presence of intermediate species that could have appeared if some intermediate stage involving the presence of completely different pigments had been involved in the ripening process.

As it can be seen in Fig. 5, the colour histogram of unripe olives (Component 1) shows high Red and Green values and low Blue values, providing a typical olive green colour when the maximum values of the R, G and B colour histograms are represented (as seen in the related square of Fig. 5). Ripe olives (Component 2), instead, show higher frequencies at low Red, Green and Blue values, corresponding to very dark colours, close to black.

The MCR resolved colour histograms can be additionally used to connect the colour information of the global image to the colour of individual pixels. In this way, it is possible to perform a fast approximate identification of pixels corresponding to unripe (green) olives and to ripe olives by comparing the RGB values of the individual pixels with those in the resolved MCR histograms. This comparison was carried out taking as a reference the maximum value of R, G and B in the histograms of the unripe and ripe olive components and setting an interval around these points that comprehends one quartile of the values. In this way, two sets of RGB values and their confidence intervals were defined, one for the ripe (dark) olives and another for the unripe (green) olives. Thus, any individual pixel in the image that presented the three R, G and B values inside the pre-set intervals for one of the resolved components, was classified as belonging to dark olives or to green olives. When this was not the case, the pixel was left as unassigned, as it can be seen in Fig. 6. Such an approach, although not being the central objective of this work, can be a simple approximate manner to connect the global colour of the image with the individual pixel information.

3.3. Partial least squares regression

Scores from PCA showed that colour histograms clearly reflect the evolution of the maturity index. In turn, the concentration profiles and the colour histograms obtained in the MCR model were clearly related to the colour variation undertaken by olives during the ripening process and, hence, to the evolution of the Maturity Index. Taking this into account, a PLSR model appeared as a potential tool to find the quantitative relationship between MI values and the information provided by the colour histograms.

To explore this possibility, a PLS calibration model between the colour histograms (**X**) and the maturity index, MI (**Y**) was built. As a calibration set, the dataset made up of olives of seven different varieties at three ripening stages in the same year (2021) was used. For this model, only mean centring was used for the **X**-block, this is, for the histograms, as well as for the **Y**-block. The model was cross-validated

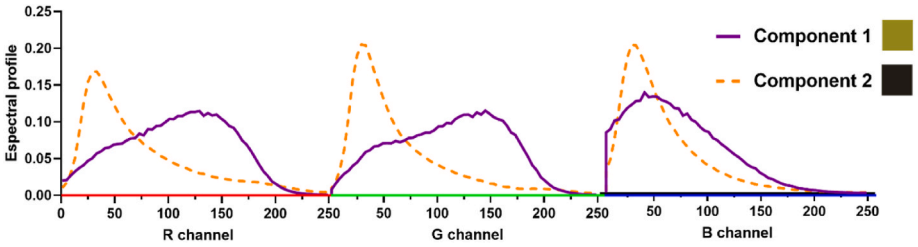


Fig. 5. MCR-ALS spectral profiles related to the two components needed to model the evolution of different pigments during the olive ripening process monitored in data set B. The colour associated with the maximum R, G and B values for each component is represented in the squares located at the right of the legend.

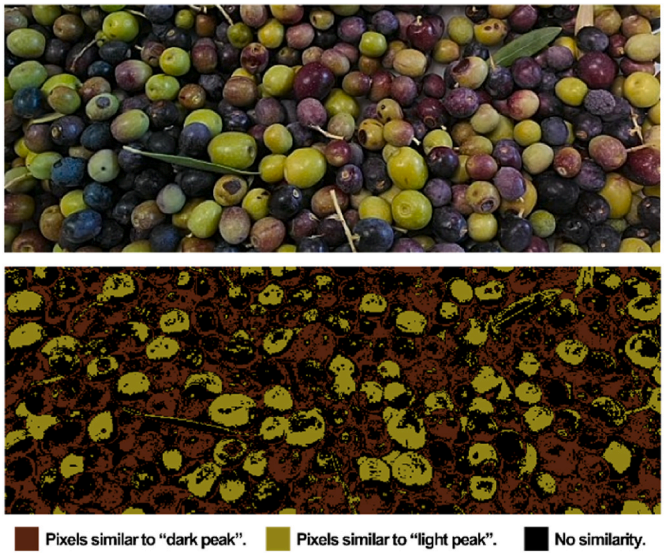


Fig. 6. Separation of the pixels of an olive image of data set B as belonging to green or dark olives, based on the information obtained in the resolved MCR colour histograms. Pixels in black were unassigned. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

(CV) using a 10-fold cross-validation, and the CV predictions can be seen in blue dots in Fig. 7a. Only the first LV was included, since it explained around 90 % of the **Y**-block variance and the CV error did not decrease significantly with more LVs. The external validation of the model was carried out using a validation set formed by images taken during the year 2022 on samples related to the same olive varieties and area as those in the calibration set.

As can be seen in red squares in Fig. 7a, the predictions for this external test set show that this model could be used for the on-site prediction of olive MI since the prediction RMSE (0.3) is low enough for routine analysis, and even more considering the subjectivity of the reference method. It has to be said that in 2022 the sampling started later in the year, so some of the olives had a MI too high to be predicted with this model and they were removed from the validation set so as not to extrapolate any result. When comparing the CV predictions and external validation predictions (Fig. 7a) it can be seen that the two campaigns are overlapped, this is, there is no major bias when using the model calibrated with the samples from 2021 for predicting the MI of samples from 2022.

In addition, to assess the heterogeneity that corresponds to the olives sampled, each image was divided in four quadrants, since there are around 50 olives per quadrant in this set of images (Fig. S4). Based on this division, a new PLS model was built for MI prediction taking into consideration as individual samples the colour histogram of every quadrant of the image, which was related to the same MI value. In this case, the cross-validation was carried out using a venetian blinds strategy, taking into account that all four quadrants of the same image were at the same time either used to perform the model or used for external prediction. The predictions performed with this new model can be seen in Fig. 7b.

Fig. 7b shows that there is a variability associated with the sampling of olives, but it is not significant when the image contains at least 50 olives. In this case the average relative standard deviation (RSD) between quadrants is below five percent (4.2 %).

Moreover, the conclusions that can be obtained from the Regression Vector coefficients of this PLSR model, as shown in Fig. 8, are similar to those obtained from observing the spectral profiles of the two components obtained from the MCR-ALS algorithm. The histograms of all three channels (RGB) can be divided in two parts: a peak in the lower values, which is associated with dark pixels (as previously discussed, related to anthocyanins) and is positively correlated to the MI, and another peak in the higher values (as previously discussed, related to chlorophylls and carotenoids), negatively correlated to the MI. It can also be appreciated that the coefficients linked to the Blue part of the histogram are clearly closer to zero than those related to the Green and Red part, which were

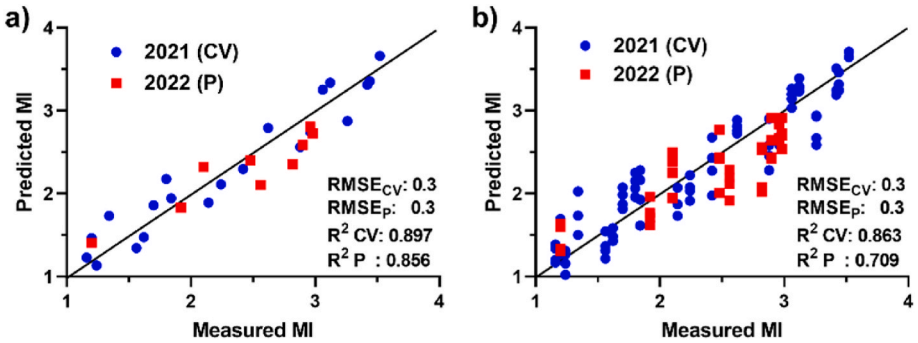


Fig. 7. MI predictions of images of olives of different varieties using a PLSR model against their measured MI. In blue circles, the 10-fold cross-validated predictions of the 2021 campaign images. In red squares, the validation predictions of the 2022 campaign images. a) using the whole images b) using the images divided in four quadrants to simulate sampling replicates.

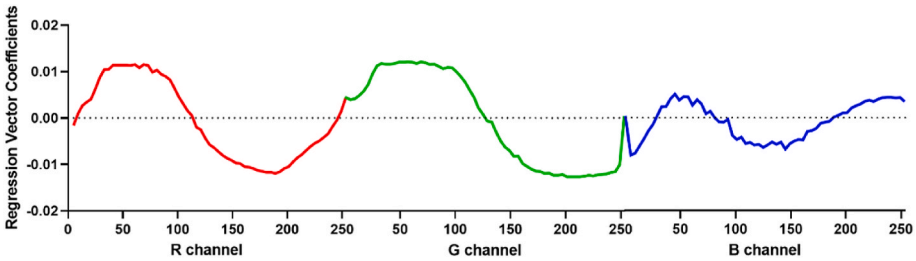


Fig. 8. Regression Vector coefficients of the PLSR model used to predict the MI of olive images.

seen to be the colours that varied the most during the olive ripening process. This qualitative information shows that the model is considerably robust, since the regression vector is very smooth and consistent with the physicochemical properties associated with the olive ripening process.

4. Conclusions

This study has confirmed the high potential of digital images to understand the olive ripening process and to predict the Maturity Index with a satisfactory accuracy. The tests carried out have proven that this methodology is easily transferable for routine analysis since the results obtained are not significantly affected by variations induced in the digital images due to differences in environmental conditions and user expertise in the image acquisition process. These promising results suggest the possibility to propose prototypes combining image and incorporated models for MI predictions on-site, providing more objective and reproducible results than those obtained through visual inspection of trained experts [11].

The combination of digital images and multivariate curve resolution (MCR-ALS) was very useful to model the ripening process of olives, providing valuable insights to understand the underlying changes in composition that occur along ripening time. In particular, through the interpretation of colour information, MCR was able to model accurately the changes in colour components related to the reduction of chlorophylls and carotenoids and to the increase of anthocyanins [6]. Further research is needed to fully understand the underlying mechanisms of olive ripening and to explore the potential applications of these techniques in a practical setting, but it can be envisioned that this kind of modelling can help to interpret differences in ripening among olives from different varieties and subject to different climatic conditions.

Additionally, the use of colour histograms as seeding information to predict the maturity index was also shown to be effective. By analysing the colour of the olives at different stages of ripening, it was possible to accurately predict their Maturity Index and thus their readiness for harvest.

In conclusion, the use of chemometrics based on the information provided by colour histograms can provide valuable insights into the ripening process of olives and can be used to predict maturity index. The methodologies proposed have the potential to improve the understanding of the olive ripening process and to support the development of more efficient and sustainable olive production practices.

Funding

Grant PID2019-104269RR-C33 funded by MCI/AEI/10.13039/501100011033. Grant URV Martí i Franqués –Banco Santander (2021PMF-BS-12). Grant URV-IRTA Martí i Franqués (2020PMF-PIPF-6). A.J. acknowledges funding from grant PID2019-1071586B-IOO.

CRediT authorship contribution statement

Jokin Ezenarro: Methodology, Software, Formal analysis, Investigation, Writing – original draft. **Ángel García-Pizarro:** Validation, Resources, Writing – original draft. **Olga Busto:** Conceptualization, Writing – review & editing. **Anna de Juan:** Investigation, Conceptualization, Writing – review & editing. **Ricard Boqué:** Investigation, Conceptualization, Writing – review & editing, 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.

Data availability

Data will be made available on request.

Acknowledgements

The authors would like to thank the olive mills from *Les Garrigues* production area that allowed the collection of data for this study. And all the members of the Oliviculture group of IRTA.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aca.2023.341884.

References

[1] P.H. Gonçalves Dias Diniz, Chemometrics-assisted color histogram-based analytical systems, J. Chemom. 34 (2020), e3242, https://doi.org/10.1002/CEM.3242.
[2] A. Roda, E. Michelini, M. Zangheri, M. Di Fusco, D. Calabria, P. Simoni, Smartphone-based biosensors: a critical review and perspectives, TrAC, Trends Anal. Chem. 79 (2016) 317–325, https://doi.org/10.1016/J.TRAC.2015.10.019.
[3] N. Ponomarenko, S. Krivenko, V. Lukin, K. Egiazarian, J.T. Astola, Lossy compression of noisy images based on visual quality: a comprehensive study, EURASIP J. Appl. Signal Process. 2010 (2010) 13, https://doi.org/10.1155/2010/976436.
[4] G. Orlandi, R. Calvini, L. Pigani, G. Foca, G. Vasile Simone, A. Antonelli, A. Ulrici, Electronic eye for the prediction of parameters related to grape ripening, Talanta 186 (2018) 381–388, https://doi.org/10.1016/j.talanta.2018.04.076.
[5] V. Usenik, F. Štampar, R. Veberič, Anthocyanins and fruit colour in plums (Prunus domestica L.) during ripening, Food Chem.. 114 (n.d.) 529–534. https://doi.org/10.1016/j.foodchem.2008.09.083.
[6] A. Dag, Z. Kerem, N. Yogev, I. Zipori, S. Lavee, E. Ben-David, Influence of time of harvest and maturity index on olive oil yield and quality, Sci. Hortic. 127 (2011) 358–366, https://doi.org/10.1016/J.SCIENTA.2010.11.008.
[7] M.A. Mele, M.Z. Islam, H.M. Kang, A.M. Giuffrè, Pre-and post-harvest factors and their impact on oil composition and quality of olive fruit, Emir. J. Food Agric. 30 (2018) 592–603, https://doi.org/10.9755/EJFA.2018.V30.I7.1742.
[8] R. Mafrica, A. Piscopo, A. de Bruno, M. Poiana, Effects of climate on fruit growth and development on olive oil quality in cultivar carolea, Agriculture 11 (2021) 147, https://doi.org/10.3390/AGRICULTURE11020147, 11 (2021) 147.
[9] C. Nergiz, Y. Engez, Compositional variation of olive fruit during ripening, Food Chem. 69 (2000) 55–59, https://doi.org/10.1016/S0308-8146(99)00238-1.
[10] C. Conde, S. Delrot, H. Gerós, Physiological, biochemical and molecular changes occurring during olive development and ripening, J. Plant Physiol. 165 (2008) 1545–1562, https://doi.org/10.1016/J.JPLPH.2008.04.018.
[11] M. Uceda, L. Frías, Épocas de recolección, Evolución del contenido graso del fruto y de la composición y calidad del aceite, in: Proceedings of II Seminario Oleícola International, Córdoba, Spain, 1975, pp. 25–46.

[12] C. Alamprese, S. Grassi, A. Tugnolo, E. Casiraghi, Prediction of olive ripening degree combining image analysis and FT-NIR spectroscopy for virgin olive oil optimisation, Food Control 123 (2021), 107755, https://doi.org/10.1016/j.foodcont.2020.107755.
[13] E. Guzmán, V. Baeten, J.A.F. Pierna, J.A. García-Mesa, Determination of the olive maturity index of intact fruits using image analysis, J. Food Sci. Technol. 52 (2015) 1462–1470, https://doi.org/10.1007/S13197-013-1123-7/TABLES/4.
[14] A. El, R. Abd, E.-R. Ahmed, H.E. Hassan, A.A. Abd El-Rahman, M.M. Attia, Color properties of olive fruits during its maturity stages using image analysis, in: Color Properties of Olive Fruits during its Maturity Stages Using Image Analysis, 2011, https://doi.org/10.1063/1.3631817.
[15] H.E. Hassan, A.A.A. El-Rahman, M.M. Attia, Color properties of olive fruits during its maturity stages using image analysis, AIP Conf. Proc. 1380 (2011) 101, https://doi.org/10.1063/1.3631817.
[16] D. Gamrasni, L. Li, A. Lichter, D. Chalupowicz, T. Goldberg, O. Nerya, R. Ben-Arie, R. Porat, Effects of the ethylene-action inhibitor 1-methylcyclopropene on postharvest quality of non-climacteric fruit crops, in: Postharvest Biology and Technology, 2016, https://doi.org/10.1016/j.postharvbio.2015.09.031. Article in.
[17] E. Borràs, J. Ferré, R. Boqué, M. Mestres, L. Aceña, O. Busto, Data fusion methodologies for food and beverage authentication and quality assessment – a review, Anal. Chim. Acta 891 (2015) 1–14, https://doi.org/10.1016/J.ACA.2015.04.042.
[18] H. Abdi, L.J. Williams, Principal component analysis, Wiley Interdiscipl. Rev. Comput. Stat. 2 (2010) 433–459, https://doi.org/10.1002/WICS.101.
[19] I.T. Jolliffe, Principal Component Analysis, Springer-Verlag, New York, 1986, https://doi.org/10.1007/B98835.
[20] J. Jaumot, A. de Juan, R. Tauler, MCR-ALS GUI 2.0: new features and applications, Chemometr. Intell. Lab. Syst. 140 (2015) 1–12, https://doi.org/10.1016/J.CHEMOLAB.2014.10.003.
[21] A. de Juan, R. Tauler, Multivariate Curve Resolution: 50 years addressing the mixture analysis problem – a review, Anal. Chim. Acta 1145 (2021) 59–78, https://doi.org/10.1016/J.ACA.2020.10.051.
[22] P. Geladi, B.R. Kowalski, Partial least-squares regression: a tutorial, Anal. Chim. Acta 185 (1986) 1–17.
[23] H. Wold, Model Construction and Evaluation when Theoretical Knowledge Is Scarce: Theory and Application of Partial Least Squares, Evaluation of Econometric Models, 1980, pp. 47–74, https://doi.org/10.1016/B978-0-12-416550-2.50007-8.
[24] W. Windig, J. Guilment, Interactive self-modeling mixture analysis, Anal. Chem. 63 (1991) 1425–1432, https://doi.org/10.1021/AC00014A016/ASSET/AC00014A016.FP.PNG_V03.