Microchemical Journal 194 (2023) 109222

Contents lists available at ScienceDirect

Microchemical Journal

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

Mixture design and multivariate image analysis to monitor the colour of strawberry yoghurt purée

Pier Lorenzo Rolando [a], Rosalba Calvini [a,b,c,*], Giorgia Foca [a,b,c], Alessandro Ulrici [a,b,c]

[a] Dipartimento di Scienze della Vita, Università di Modena e Reggio Emilia, Padiglione Besta, Via Amendola, 2, 42122 Reggio Emilia, Italy
[b] Centro Interdipartimentale BIOGEST-SITEIA, Università degli Studi di Modena e Reggio Emilia, Piazzale Europa, 1, 42122 Reggio Emilia, Italy
[c] Consorzio Interuniversitario Nazionale per la Scienza e Tecnologia dei Materiali (INSTM), Via Giusti, 9, 50121 Firenze, Italy

ARTICLE INFO

Keywords:
Strawberry yoghurt purée
Anthocyanins
Mixture design
Response surface
Multivariate image analysis
Colourgrams

ABSTRACT

Food colour is a commercial added value, since it represents the first appealing factor for consumers. In this context, this study was aimed at evaluating the effect of strawberry yoghurt purée (SYP) formulation on the corresponding colour and on its variation over time, which is mainly due to degradation and browning phenomena. To this aim, a combined approach was used that included mixture design and multivariate analysis of RGB images. Strawberry purée, sugar, lemon juice and two types of thickener were mixed in different proportions by I-optimal mixture design to obtain 44 SYP formulations. The samples were subjected to light and temperature stress conditions for five weeks; during this time the RGB images of the samples were acquired using a flatbed scanner, along with the images of the corresponding control samples. The dimensionality of the acquired images was reduced by two different approaches: i) the conversion of images into signals, namely colourgrams, which can be seen as the colour fingerprint of the imaged samples, and ii) the calculation of the median values of various colour-related parameters. The colourgrams dataset was then subjected to exploratory data analysis using Principal Component Analysis, while the median values of colour-related parameters were analysed using Response Surface Methodology and Partial Least Squares-Discriminant Analysis. The aim of data analysis was both to find the best colour parameters to describe colour variability over time, and to investigate the cause-effect relationship between mixture proportions and colour response. The results highlighted that, among the considered colour parameters, relative green (i.e., the ratio of green to lightness) and red could be used to monitor colour changes. Colour variation due to stress conditions was more pronounced for samples with a high percentage of strawberry purée, and the type of thickener also affected the colour degradation kinetics.

1. Introduction

Strawberry yoghurt purée (SYP) is a specific formulation – composed of strawberry, sucrose, lemon juice, thickener, natural strawberry flavour, and water – to be used as a semi-finished product in the manufacturing of flavoured yoghurts. In particular, SYP is added during yoghurt production in a percentage ranging from 15% to 20% w/w of the entire end product.

In order to satisfy consumers' expectations, a good SYP formulation has to transmit to the final product a balanced and recognisable strawberry flavour/taste and a pronounced and pleasant strawberry-like red colour. Focusing on composition variables only, the optimal taste strongly depends on the right balance between strawberry and sugar content in the formulation, as well as on the addition of strawberry food flavouring. On the other hand, colour dependencies are way more complex to foresee and understand: multiple interactions between all the components may come into play for the determination of an optimal colour. Furthermore, the colour is not always stable over time, but it is subjected to changes that depend on different factors. In this context, a topic of great concern for SYP manufacturers is the identification of the optimal formulation that ensures the desired colour appearance and its stability over time, since these two properties can greatly affect purchases of the final product.

Anthocyanins are the natural pigments which are mainly responsible for strawberry colour. Despite many studies showed a notable variability on anthocyanins profile based on strawberry cultivar and maturation

* Corresponding author at: Dipartimento di Scienze della Vita, Università di Modena e Reggio Emilia, Padiglione Besta, Via Amendola, 2-42122 Reggio Emilia, Italy.
E-mail address: rosalba.calvini@unimore.it (R. Calvini).

https://doi.org/10.1016/j.microc.2023.109222
Received 31 May 2023; Received in revised form 8 August 2023; Accepted 19 August 2023
Available online 21 August 2023

stage [1,2], scientific literature agrees on pelargonidin-3-glucoside as the most abundant compound, which ranges between 50% and 90% of total anthocyanins content [1–4]. Colour properties of these compounds originate from their resonating structures, which also confer them an intrinsic instability [5] that must be taken into consideration when trying to optimize and control the colour of food products containing anthocyanins. Many factors have a great influence on their colour and stability, including pH, temperature, light exposure, time, oxidase enzyme activity, water activity and total soluble content, among others.

While pH-dependent structural changes have been largely documented and studied [5], the impact of temperature as the second most crucial parameter on anthocyanin stability is not sufficiently understood from a chemical point of view [6]. However, it is well known that anthocyanins are thermolabile, leading to the so-called browning effects. These phenomena are reported in two different studies, in which pilot productions of strawberry jams [7] and juices [8] were put under observation to evaluate which process caused higher losses in anthocyanins content. In addition, Martinsen et al. (2020) [7] compared different levels of thermal treatment and observed a first order reaction kinetics. Browning processes also occur when strawberries are mashed into pulp, due to the effect of oxidizing enzymes such as polyphenol oxidase. In this case, a balanced blanching operation can inactivate these enzymes and prevent undesired colour variations in the final product [9].

Given the complex relationships between the various factors underlying SYP colour and its stability, in this work we have focused on the effect of compositional factors. A quaternary mixture design was elaborated, considering the I-optimality criterion [10] for the selection of the experimental conditions, i.e., of the recipes. This allowed to study the colour characteristics of SYP according to the recipe used for its production and how the colour changes over time under controlled storage conditions. Indeed, ingredients proportions in the recipe influence the properties of SYP, such as final pH of the solution, total soluble solids content, water activity ([11], where the authors also studied the effect of different thickeners), and citric acid content [12] among others. In turn these properties affect anthocyanins stability and therefore the final colour of the product.

In order to accelerate degradation and browning phenomena we considered extreme stress conditions, i.e., continuous exposure to intense light at a temperature of about 35 °C. Any other factor potentially affecting anthocyanins stability was kept constant.

To evaluate the colour of the SYP samples and its variation over time, we considered Red-Green-Blue (RGB) imaging due to its several advantages. Indeed, this tool allows to objectively assess colour related information about the investigated samples using affordable devices in a fast and non-destructive manner, resulting also a green approach.

Compared to traditional tristimulus colourimeters, RGB images can be acquired by commonly used devices such as digital cameras, smartphones or flatbed scanners, allowing also to evaluate colour variability within the sample surface [13]. RGB images are complex data arrays, and it is of utmost importance to apply proper image analysis strategies to extract the useful information from such data [14].

When multiple images have to be analysed altogether, data-dimensionality reduction is generally performed, which consists in converting each image into a feature vector acting like a fingerprint of the sample in terms of colour. Then, the feature vectors obtained from each acquired image can be collected into a data matrix and elaborated by multivariate data analysis to extract the useful information and relevant colour characteristics of the images. As a very straightforward method for data-dimensionality reduction, it is possible to calculate average or median values of R, G and B values and/or other colour parameters derived from RGB data, using then these descriptors to evaluate the colour properties of the considered samples [15 –17]. In order to preserve also the information related to colour variability within the images, some of us proposed an alternative approach for data dimensionality reduction, which consists in calculating the frequency distribution curves of a set of colour descriptors and merging them in sequence to obtain a signal named *colourgram* [18].

The main idea behind the present study is to combine mixture design and multivariate RGB image analysis to evaluate how colour properties of SYP are affected by compositional factors and stress conditions. More in detail, a mixture domain of the semi-finished strawberry product was defined considering four ingredients, i.e., strawberry purée, sugar, lemon juice and thickener. Within this domain, 44 different mixtures selected using the I-optimality criterion were prototyped and placed under observation over a five-week period, during which RGB images of the samples put under stress conditions and of the corresponding control samples were acquired every week using a flatbed scanner. The best parameters to describe colour variability were investigated by multivariate image analysis, also trying to establish a cause-effect relationship between mixture proportions and colour response.

Table 1
Lower and upper limits for mixture components expressed as weight fractions (w/w); for each mixture the sum of the four components is equal to 95% w/w, since a fixed quantity of water (5% w/w) was added.

Lower Limit		Component		Upper Limit
0.25	≤	A: Strawberry purée	≤	0.75
0.18	≤	B: Sugar	≤	0.696
0	≤	C: Lemon juice	≤	0.06
0.004	≤	D: Thickener	≤	0.01
		A + B + C + D	=	0.95

2. Materials and methods

2.1. Mixture design for strawberry yoghurt purée samples preparation

In order to model the variation of different colour-related properties, strawberry yoghurt purée samples preparation was carried out following the design of experiments (DoE) approach by using the Design Expert ver. 10 software (Stat-Ease Inc., USA). A double-constrained quaternary mixture design was applied, in which components variation ranges were derived from initial business recipe based on a number of considerations, including commercial interest for a strawberry formulation of up to 75% w/w, company know-how on the feasibility of some mixtures, legal/technical limits for some ingredients, such as the thickener, and finally the need to obtain a fairly pronounced colour variation for better modelling. The resulting ranges for each component are summarized in Table 1; the total sum of the four ingredients percentage was equal to 95% w/w, since a constant quantity of 5% w/w of water was added to each mixture. As it can be noticed, natural strawberry flavour was not added to samples, due to its small amount in the recipe (<0.1 % w/w) and its null effect on mixture colour; furthermore, no one of the above referred publications listed flavouring as a potential influencing factor.

Table 2 shows the design matrix consisting of 22 experiments, chosen based on the I-optimality criterion to fit the following quadratic polynomial model:

$$y = \sum_{i=1}^{m} b_i x_i + \sum_{i=1}^{m} \sum_{j>i}^{m} b_{ij} x_i x_j \quad (1)$$

where y is the colour parameter to be modelled, m is the number of mixture components, b_i is the one-factor term of the i-th component with proportion x_i, and b_{ij} is the interaction term for components i and j (with $i \neq j$).

All the mixtures resulting from DoE were prepared with two different thickeners, pectin (PEC) and locust bean gum (LBG), following the random order reported in Table 2 and obtaining a total of 44 SYP samples (=22 experiments × 2 thickener types).

Each mixture was prepared in a 1 kg batch using a Bimby TM21 robot (Vorwerk & Co. KG, Wuppertal, Germany). The complete cooking cycle for each mixture took in general between 20 and 25 min, considering ingredients loading, temperature rising till 93 °C and thermal treatment

Table 2
Design matrix for both PEC and LBG series of samples.

Run order	PEC samples name	LBG samples name	A: Strawberry Purée (%)	B: Sugar (%)	C: Lemon Juice (%)	D: Thickener (%)
1	PEC1	LBG1	70.60	18.00	6.00	0.40
2	PEC2	LBG2	54.12	37.76	2.72	0.40
3	PEC3	LBG3	25.00	69.60	0.00	0.40
4	PEC4	LBG4	72.50	18.00	3.50	1.00
5	PEC5	LBG5	25.00	63.60	6.00	0.40
6	PEC6	LBG6	75.00	19.40	0.00	0.60
7	PEC7	LBG7	39.53	53.59	1.33	0.55
8	PEC8	LBG8	54.12	37.76	2.72	0.40
9	PEC9	LBG9	55.20	33.00	6.00	0.80
10	PEC10	LBG10	75.00	19.40	0.00	0.60
11	PEC11	LBG11	25.00	67.20	2.00	0.80
12	PEC12	LBG12	58.33	35.67	0.00	1.00
13	PEC13	LBG13	25.00	63.00	6.00	1.00
14	PEC14	LBG14	72.50	18.00	3.50	1.00
15	PEC15	LBG15	54.06	37.58	2.66	0.70
16	PEC16	LBG16	50.00	44.00	0.00	1.00
17	PEC17	LBG17	40.00	48.00	6.00	1.00
18	PEC18	LBG18	47.65	40.65	6.00	0.70
19	PEC19	LBG19	64.53	28.59	1.33	0.55
20	PEC20	LBG20	54.06	37.58	2.66	0.70
21	PEC21	LBG21	25.00	67.20	2.00	0.80
22	PEC22	LBG22	54.06	37.58	2.66	0.70

at 93 °C ± 5 °C for 15 min. A unique batch of strawberry purée was picked from company's stock. It was obtained from strawberries of a single cultivar, which were subjected to a blanching process (78 °C for 3 min) and then mashed. This batch of strawberry purée was used for the preparation of all the samples in combination with other ingredients. After the cooking process, each mixture was divided into 6 aliquots of about 165 g and the aliquots were transferred into 105 ml jars, subsequently vacuum-sealed and labelled. The jars were filled until the edge with the mixtures still hot; in this manner no significant headspace was left for air oxygen, which is a factor that could affect the colour stability of the samples. Two jars of each mixture were used for the experimental part of this study, while the remainder jars served as backup.

2.2. Experimental set-up for stressing the samples

The 44 mixtures were randomly divided into 4 blocks of 11 samples each, paying attention to distribute as much uniformly as possible among the different blocks the various mixture compositions.

In order to accelerate oxidative phenomena that normally could take several months to occur, one jar for each one of the 44 mixtures (labelled as "S" series) was stored under stress conditions for approximately five weeks. More in detail, the jars were flipped over and stored in a closed polystyrene box containing a 50 W lamp (2 bulbs, full spectrum 380–720 nm). In this manner, the bottom of the jars was exposed to light for the entire period of five weeks (690 h of light exposure overall). In order to minimize the effect of heterogeneous lighting conditions inside the box, the internal area of the polystyrene box was divided in four areas, one for each block of samples, and once a week the jars were moved to the adjacent area (Figure S1). During this procedure each sample was inspected and shaken in order to prevent the possible formation of different density layers, even if this effect has never been observed over the five weeks of the experiment.

Light was not the only stress factor inside the box. In fact, due to the Joule effect, lamp bulbs warmed up the environment at a temperature of 34.5 °C ± 0.9 °C. Temperature was monitored with an Inkbird ITC-308 digital thermal regulator. The combination of light and temperature stress factors visibly accelerated oxidative reactions in samples, leading red anthocyanins towards more brownish compounds.

Simultaneously, the second jar of each one of the 44 mixtures was used as control sample (labelled as "C" series) and kept in the dark at a refrigerated temperature between 0 °C and 4 °C for the same period of five weeks.

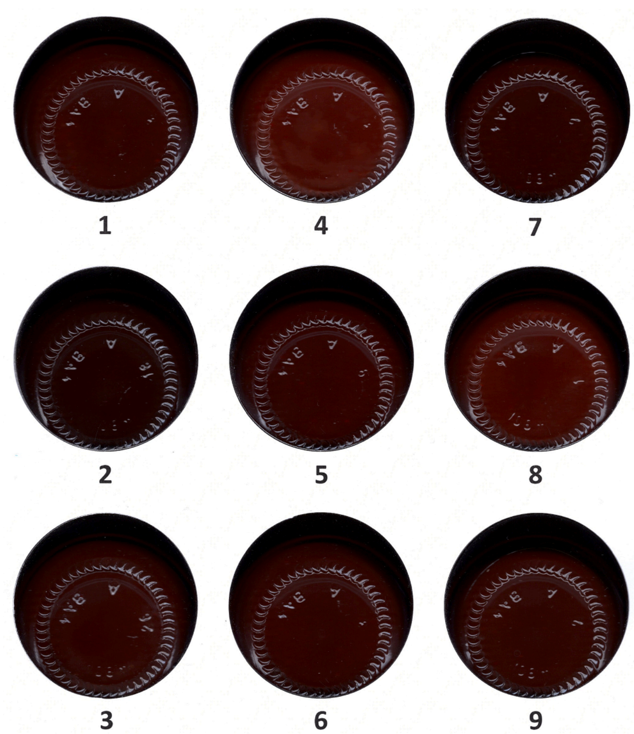

Fig. 1. Example of an image acquired with the flatbed scanner on a set of 9 samples. Following the numerical order (from 1 to 9) the samples are: PEC8, PEC7, LBG2, LBG1, LBG22, PEC18, LBG16, PEC6 and LBG12. All the samples belong to the stressed (S) series acquired at T04.

2.3. Image acquisition system

The evolution of samples colour during time in both stressing and controlled conditions was monitored weekly using an image acquisition system. The system was composed by a flatbed scanner (Epson Perfection V39), a white cardboard with 9 slots for the jars containing the mixtures, a carton box to cover the flatbed scanner during image acquisition and a computer to save and store images for further elaboration. Fig. 1 shows an example of an image acquired with this system.

Once a week for five weeks, both the S and the C samples series were taken from their respective storage places for images acquisition. Since each block was composed of 11 samples but only 9 slots were available in the white cardboard mask, two scans for each block were necessary. For this reason, the 11 samples of each block were divided into two sub-blocks, one with 5 samples and the other one with the remaining 6 samples. The two scans for each block were acquired as follows:

- the first scan contained the 5 samples of the first sub-block and 4 samples randomly selected from the second sub-block;
- the second scan contained the 6 samples of the second sub-block and 3 samples randomly selected from the first sub-block.

The position of the samples in the image scene was randomized for each image acquisition, as well as the subdivision of the samples of each block into the two sub-blocks.

Images of SYP mixtures were acquired at 6 different acquisition times from T00 to T05, where T00 corresponds to the day in which the mixtures were prepared and T01-T05 correspond to the weekly acquisitions performed from the first until the fifth storage weeks. Therefore, the final images dataset was composed of 96 images (=2 scans × 4 blocks × 2 series × 6 acquisition times).

For all the images, the following acquisition parameters were applied directly from Epson scanner software (EPSON Scan Ver. 3.9.4.7IT) and kept constant for all acquisition times: lightness + 20, contrast + 20, saturation + 60.

All the images were saved in .JPG format, with 24-bit depth and a spatial resolution of 9359 row pixels × 6800 column pixels. Considering a file size approximately equal to 191 MB per image, the overall dataset size was about 18 GB.

2.4. Data analysis

2.4.1. Data dimensionality reduction of RGB images dataset

Before further analysis, additional image pre-processing steps were necessary in order to crop the image of each single sample. These operations were automatically carried out by means of an image cropping algorithm written in MATLAB language (ver. 9.3, The Mathworks Inc., USA), obtaining on the whole 864 (=96 × 9) images of single samples.

This large dataset of images needed to be managed through proper data dimensionality reduction methods in order to retrieve useful information. In this study two different data dimensionality reduction approaches were considered. Initially, the colourgrams approach was used as a first strategy to gain a preliminary overview of the entire dataset of images and evaluate differences over time between stressed and control samples. Colourgrams, proposed the first time by Antonelli et al. (2004) [18] and later used in other successful applications [19–27], allow to codify the colour information contained in the three-dimensional array of each RGB image by reducing it to a signal (vector). Basically, the frequency distribution curves of the R, G and B channels of the original image and of other parameters derived from the values of R, G and B are obtained; then, the colourgram is built by merging these frequency distribution curves in sequence. In addition to R, G and B, the colour parameters considered for colourgrams computation include lightness (L = R + G + B), relative red (RR = R / L), relative green (RG = G / L), relative blue (RB = B / L), hue (H), saturation (S), intensity (I) and the three score vectors of each one of three PCA models calculated on raw, mean centered and autoscaled RGB data, respectively. For further details about colourgrams computation and the considered colour parameters, the reader is referred to Antonelli et al., 2004 [18] and Calvini et al., 2020 [28].

Such a transformation simplifies the analysis of a dataset of images that can be converted into a matrix of signals, where each row corresponds to a signal codifying the colour properties of a specific image in the dataset. Furthermore, the inclusion in the colourgram of more colour parameters in addition to the RGB channels allows to simultaneously evaluate different colour-related features of the imaged samples and to identify which of them are the more relevant for the considered application.

An easy-to-use graphical user interface developed by some of the authors [28] for the creation of colourgrams is freely downloadable from the web [29]. In this case, before calculating the colourgrams, background removal was performed by selecting a threshold limit: only the pixels with values of the blue channel < 200 have been included in colourgrams elaboration.

Subsequently, to simplify the identification of the colour parameters mainly influenced by mixture composition and stress conditions, an alternative approach of data dimensionality reduction was considered. At first, an additional cropping procedure was necessary to restrict sample images to the center of the jars in order to focus on mixtures colour and eliminate image noise deriving from shadows and reflections (Figure S2). From these re-cropped images, median values of red, green, blue, lightness, relative red, relative green, relative blue, hue, saturation and intensity were obtained, which correspond to the colour-related parameters considered in the calculation of colourgrams. Median values were calculated considering only the pixels with blue values lower than 50; this threshold was set to exclude remaining reflections due to signs of the jar glass. In this manner, each re-cropped image was converted into a feature vector of 10 elements, corresponding to the median values of the considered colour-related parameters.

2.4.2. Exploratory analysis of colourgrams by PCA

The mean centered matrix of colourgrams was analysed by Principal Component Analysis (PCA), by using the PLS Toolbox Version 8.8.1 (Eigenvector Inc.) running in MATLAB environment (ver. 9.3, The Mathworks Inc., USA). PCA performed on the colourgrams matrix allowed to visualize the overall structure of the dataset composed of 864 RGB images, to highlight clusters of similar images, and to identify the presence of outlier images.

2.4.3. Modelling of median red parameter

In order to study the effect of mixtures composition and of time on the SYP colour of the S series of samples, i.e., the series for which the time effect is expected to be much more relevant, the median values of different colour parameters were considered for possible modelling by DoE. Since median red was identified as the parameter showing better correlation with the percentage of strawberry content, for each thickener type the corresponding response surfaces were calculated considering the images acquired at the beginning (T00) and at the end of the stress time interval (T05). It should be noticed that, for the samples that were imaged twice in the same image acquisition session, the mean value of the two corresponding median red values was considered for model calculation.

Analysis of variance (ANOVA) was used to verify the statistical significance of the model and of the lack of fit; the comparison of the variation sources was based on the Fisher distribution ($P < 0.05$). The values of the coefficient of determination (R^2), of the adjusted R^2 (R^2 Adj) and of the predicted R^2 (R^2 Pred, estimated by leave-one-out crossvalidation) were considered to express the models performance, while response surfaces were obtained to represent the variation of median red with mixture composition at the different image acquisition times.

2.4.4. Evaluation of colour variation by PLS-DA

In order to highlight colour variations caused by stress conditions over time and to identify the colour parameters more suitable to monitor this aspect, the dataset of median values of the different colour parameters described in section 2.4.1 was analysed using a multivariate classification algorithm, namely Partial Least Squares-Discriminant Analysis (PLS-DA). In particular, a PLS-DA classification model was calculated to discriminate between stressed and control samples. In this case, only the samples belonging to the acquisition times from T01 to T05 were

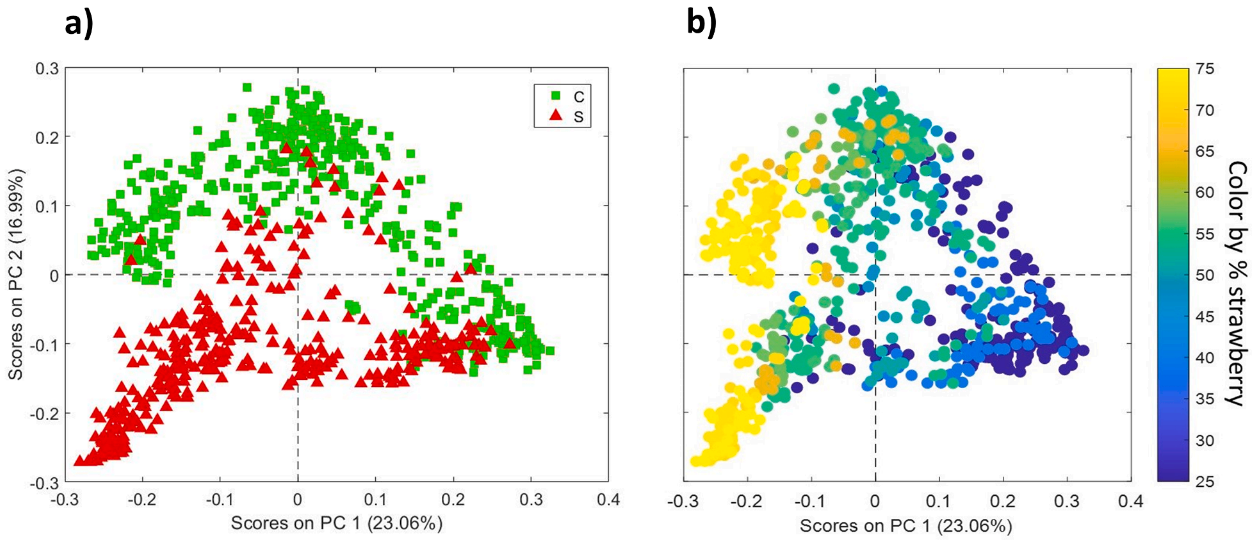

Fig. 2. PC1-PC2 score plot of the colourgrams matrix; the objects are coloured according to stressed and control samples (a) and according to strawberry percentage of the corresponding mixtures (b).

considered, while the T00 samples were excluded.

The dataset of the median values was preprocessed using autoscaling, and the optimal number of latent variables was selected in cross-validation considering 4 deletion groups, corresponding to the 4 randomized sample blocks described in Section 2.2.

The classification performances were evaluated considering sensitivity (SENS), specificity (SPEC) and classification efficiency (EFF), calculated in calibration and cross-validation. SENS is the percentage of samples of the modelled class correctly accepted by the class model, SPEC is the percentage of objects of the other classes correctly rejected by the class model, and EFF is the geometric mean of SENS and SPEC [30].

It has to be highlighted that in this study the PLS-DA model was not calculated for prediction purposes, but only to identify one or more colour descriptors that are more related to colour variations due to stress conditions. To this aim, the Variable Importance in Projection (VIP) scores were used to point out colour parameters with higher relevance in the discrimination between control and stressed samples. Indeed, VIP scores provide a measure of the relevance of each variable in the definition of the PLS-DA model; as a general rule, variables with VIP score greater than 1 can be considered as significant [31].

3. Results and discussion

3.1. PCA of colourgrams

An initial exploratory analysis was performed on the entire colourgrams dataset, considering all acquisition times (from T00 to T05) and both control and stressed samples. The analysis of the PC1 and PC2 score plot revealed interesting trends, as reported in Fig. 2, where the same samples are coloured according to stressed and control samples (Fig. 2a), and according to the percentage of strawberry of the corresponding mixtures (Fig. 2b).

PC1 describes the colour differences of the SYP samples according to the percentage of strawberry in the mixture (Fig. 2b), where the samples prepared with the highest amount of strawberry are located at negative PC1 score values, while the samples with the lowest strawberry percentage are located at positive PC1 score values. Indeed, considering the corresponding RGB images, it is possible to observe that the mixtures prepared with a higher amount of strawberry have a lighter and reddish

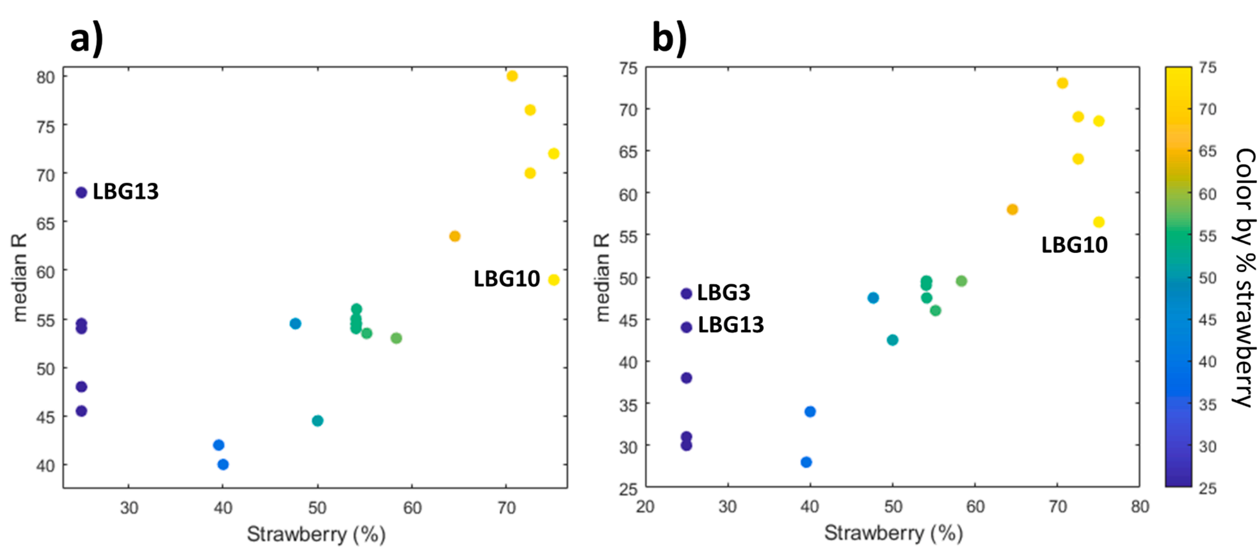

Fig. 3. Variation of median R with increasing strawberry percentage in the formulation containing LBG as thickener at T00 (a) and T05 (b).

colour, while the mixtures with a lower amount of strawberry have a darker colour.

On the other hand, PC2 highlights the differences between stressed and control samples (Fig. 2a), which are more evident for mixtures with high strawberry percentage (Fig. 2b).

Even if these simple observations could have been done by visually comparing the RGB images of some representative samples, it has to be considered that it is not feasible to simultaneously evaluate all the 864 images of the dataset with the naked eye. Conversely, the colourgrams approach coupled with PCA allowed to gain a general overview of the colour properties of the considered samples and to highlight colour variation trends common to all the images of the dataset.

In addition, PCA results showed that the stress conditions considered in this study resulted to be appropriate to promote colour variations in the samples compared to the corresponding control samples, as well as that mixtures composition is responsible for a higher colour variability than the one due to stressing factors. It must be also highlighted that stress conditions and sample composition affected sample colour in different directions in the score plot, suggesting that different colour parameters should be used to describe composition-dependent and stress-dependent colour variation.

In particular, this PCA model allowed to highlight two relevant points: the colour of mixtures is mainly influenced by strawberry percentage, and the stress conditions led to a colour modification which is more pronounced for samples with a high strawberry percentage. However, the analysis of the corresponding PC1 and PC2 loading vectors did not allow to easily identify the specific colour parameters mainly influenced by mixture composition and/or stress conditions. For this reason, the subsequent modelling steps were carried out considering a simpler approach based on the use of median values of the colour parameters.

3.2. Mixture design models on median red parameter

The analysis of the mixture design models calculated for different colour parameters revealed that the median red value was the most relevant parameter to describe colour variation of S samples according to mixture composition. A preliminary visual inspection of the scatter plot of strawberry percentage vs. median red at T00 and T05 (Fig. 3a and Fig. 3b, respectively) has however revealed an unexpected behaviour for some LBG samples (namely: LBG10 and LBG13 samples at T00; LBG3, LBG10 and LBG13 samples at T05), suggesting a possible anomaly in their colour. In particular, LBG13 was characterised by a higher median red value at both T00 and T05 compared to other samples prepared with the same amount of strawberry, which corresponds to the lowest limit, i.e., 25%. At T05 LBG3 showed the same anomalous behaviour of LBG13. On the other hand, mixture LBG10, containing the highest amount of strawberry (75%), showed a median red value much lower than the one measured for the other samples with very similar strawberry percentage.

To further confirm this trend, the images corresponding to the anomalous samples were visually compared with those of the most similar samples in terms of composition, in particular with regard to strawberry percentage; the reader can view these images and compare them with the naked eye by accessing the Supplementary Material. Fig. 3S reports in subplot a) the images of the mixtures with a low percentage of strawberry (25%) and in subplot b) the images of the mixtures with a high percentage of strawberry (70.6–75%). In the figure, the anomalous samples are boxed in blue and the replicate mixtures have the name label in the same colour. As it can be seen, in Fig. 3Sa anomalous samples appear lighter than the similar mixtures, while in Fig. 3Sb they appear darker than the similar mixtures. LBG13 (Fig. 3Sa) and LBG10 (Fig. 3Sb) are anomalous from the beginning, i.e., at both T00 and T05. Moreover, it can be noticed that LBG10 is markedly different from its replicate mixture LBG6. The most likely cause for this behaviour is ascribable to problems that arose during the cooking process. LBG3 instead changed its colour in an anomalous way over time and its final colour appears also less homogeneous than that of the other samples.

Table 3
Statistics of the mixture models for the median red parameter obtained for the PEC and LBG thickeners at T00 and T05.

	PEC series		LBG series	
Stressing time	**T00**	**T05**	**T00**	**T05**
Model	Significant	Significant	Significant	Significant
Lack of fit	Not significant	Not significant	Not significant	Not significant
R^2	0.84	0.85	0.96	0.99
R^2 Adj	0.81	0.82	0.93	0.98
R^2 Pred	0.73	0.73	0.64	0.90
Model terms	Linear mixture: A, B, C, D	Linear mixture: A, B, C, D	Linear mixture: A, B, C, D Quadratic terms: AB, AC, AD, BC, BD, CD	Linear mixture: A, B, C, D Quadratic terms: AB, AC, AD, BC, BD, CD
Excluded samples			#10, #13	#3, #10, #13

These considerations led us to eliminate the anomalous samples from the dataset before calculating the mixture design models. In particular, the model calculated on T00 samples was obtained after excluding LBG10 and LBG13, while in the model calculated on T05 samples also LBG3 was excluded in addition to LBG10 and LBG13. Indeed, these samples resulted to be anomalous since they did not show the same relationship between sample colour, expressed in terms of median red value, and sample composition as the other mixtures and their inclusion in the mixture design models would have inevitably made them less reliable. However, for the sake of completeness, the performances of the models calculated including also the anomalous samples are reported in the supplementary material (Table S1).

Table 3 reports the significant terms of the models calculated for both PEC and LBG thickener types at T00 and T05, and the corresponding performances. It should be noted that before accepting each model it was verified that the errors had a normal distribution and that the error variance was constant for any value of the independent variable.

Overall, the models were satisfactory, with acceptable R^2 Pred values. The median red parameter was adequately fitted using a linear model when the mixture was prepared using PEC as a thickener, while a quadratic model was found to be more adequate to fit median red of the mixtures containing LBG.

Fig. 4 shows the response surfaces of the models, in the triangular domain defined by the strawberry purée, sugar and lemon juice components. The lower (0.40 %) and higher (1.00 %) levels of the fourth component, i.e., the thickener, are shown below and above, respectively. It is interesting to observe that all surfaces show a similar trend as regards the variation of the median red parameter with the composition of the mixture and with stressing time. In general, the median red values increase as the amount of strawberry puree increases, as expected, and they tend to decrease with stressing time. Regarding the PEC response surfaces, the decrease in median red observed with time is almost independent of the amount of thickener in the mixture; a slight positive effect of lemon juice on median red values can be observed both at T00 and at T05. The LBG response surfaces show that, considering the same proportions of the other ingredients, at T00 the median red value is generally lower with a higher amount of thickener, and in this case it seems that the addition of lemon juice does not lead to a significant effect. At T05 the decrease in median red values is more pronounced for mixtures containing a low amount of strawberry, lemon juice and thickener. However, the comparison between the response surfaces at T00 and T05 with LBG = 0.4% also suggests that the maximum amount

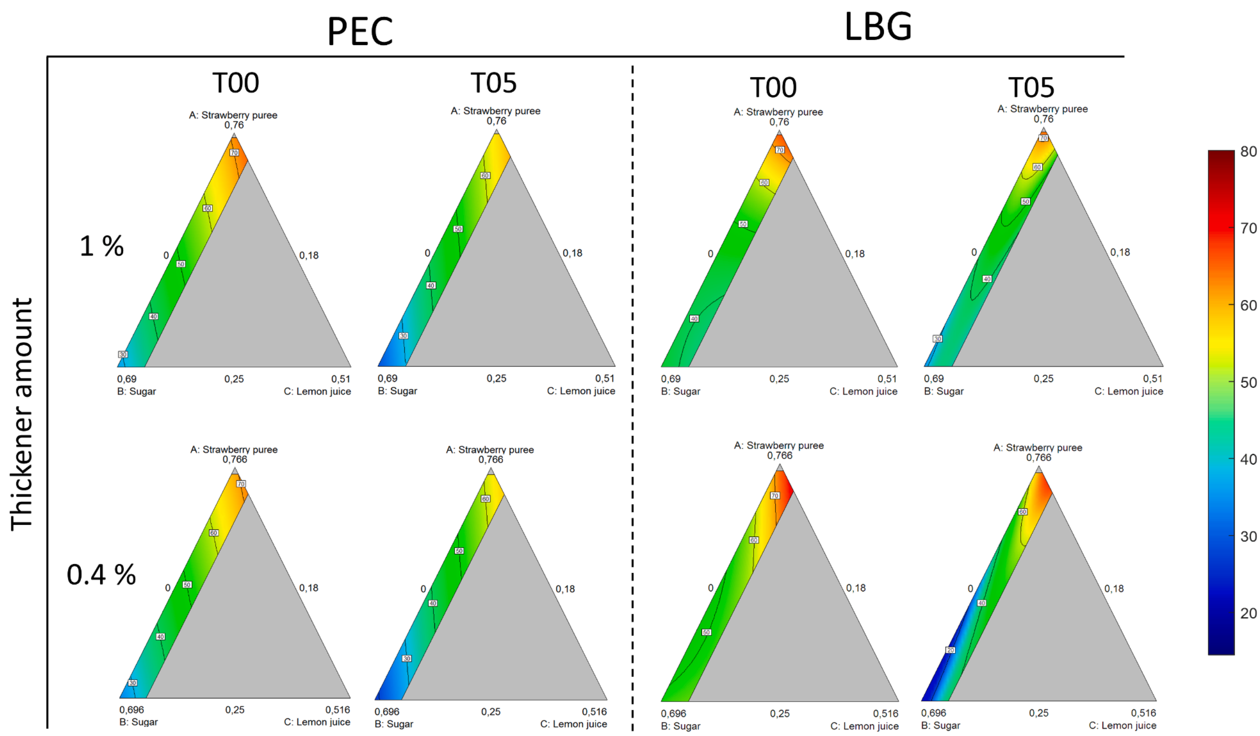

Fig. 4. Response surfaces for median R calculated from SYP images at initial (T00) and final (T05) acquisition times (samples containing PEC as a thickener on the left, samples containing LBG on the right; thickener amount decreases top-down).

Table 4
Results of the PLS-DA model calculated on the dataset of median colour values; the SENS and SPEC values are referred to the control (C) class.

	SENS	SPEC	EFF
Cal	93.2 %	86.2 %	89.6 %
CV	91.7 %	85.0 %	88.3 %

of lemon juice combined with the highest possible level of strawberry leads to the highest and most stable median red levels, which could therefore correspond to the optimal condition.

3.3. PLS-DA model on median values

Since LBG3, LBG10 and LBG13 samples resulted to be outlier at T00 and/or at T05, they were also excluded from the calculation of the PLS-DA model to discriminate between stressed and control samples considering all the acquisition times from T01 to T05.

The PLS-DA model was calculated with 4 latent variables selected according to cross-validation (*see* Section 4.4) and the corresponding results are reported in Table 4. The satisfactory results expressed in terms of SENS, SPEC and EFF values in calibration (Cal) and cross-validation (CV) confirmed a detectable colour difference between stressed and control samples. Fig. 5a shows the Y values in cross-validation (Y CV) for control class versus the acquisition order, where the samples are coloured according to control and stressed classes. In particular, the samples with a Y CV predicted value higher than the discriminant threshold (dashed red line) are classified as control samples, while the other samples are classified as stressed samples. As expected, the colour difference between control and stressed samples becomes progressively more pronounced over time; indeed, the number of misclassified samples is much lower at T05 than at T01.

Furthermore, Fig. 5b shows the Y CV predicted values for control class against strawberry percentage in the mixtures. In this case it is possible to observe that the samples with a strawberry percentage equal to 55% or higher are better classified compared to the samples with lower strawberry amount. This finding confirms that the higher the strawberry percentage, the more pronounced is colour variation due to stress conditions.

As previously stated in Section 2.4.4., this PLS-DA model has not been calculated for prediction purposes, i.e., to predict unknown samples assignation to stressed or control classes, but to obtain a latent variables space oriented toward the maximum separation of the samples according to storage conditions. In this manner it was possible to focus on colour variability related to degradation phenomena due to the considered stressing factors and observe how this variability is related with mixtures composition and storage time, as previously discussed in the description of Fig. 5.

In addition, by observing the VIP scores and the regression vector of stressed class of the PLS-DA model it is possible to identify the colour parameters that are more relevant to describe colour variations due to stress conditions and evaluate how they vary from control to stressed samples. More in detail, Fig. 6a shows the VIP scores of the PLS-DA model, were it is possible to observe that median values of relative red (RR), relative green (RG), saturation (S) and hue (H) have VIP score values higher than one (Fig. 6a), thus resulting to be the more relevant variables for the discrimination. In particular, the median value of RG parameter shows the highest VIP scores value, suggesting that this parameter is the most relevant to characterize colour differences due to stress conditions. Furthermore, considering the PLS-DA regression vector of stressed class (Fig. 6b), it is possible to observe that median RG has a positive value of the regression coefficient, therefore RG values of the SYP samples tend to increase for stressed samples over time.

4. Conclusions

In this work we investigated how the ingredients proportions influence colour and colour stability of a strawberry-based preparation widely used in industry, namely strawberry yoghurt purée (SYP). The natural pigments mainly responsible for the colour of this food product are anthocyanins, whose colour may vary due to a multitude of factors

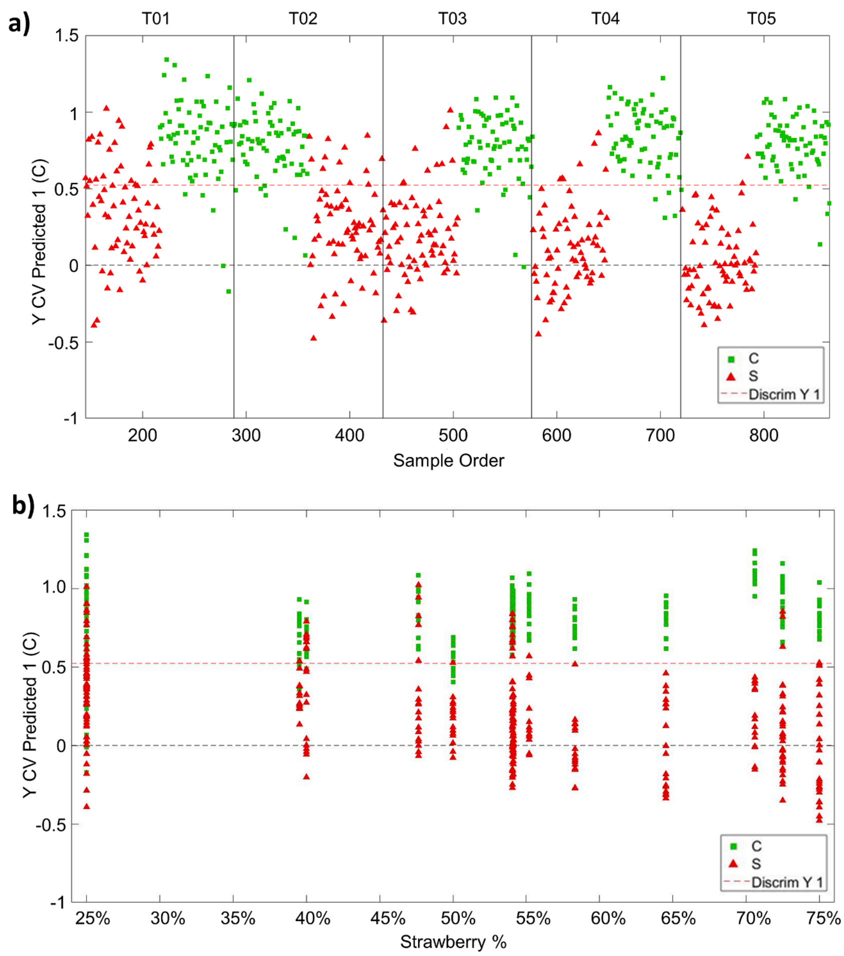

Fig. 5. Y CV predicted for C class vs. acquisition order (a) and vs. strawberry percentage of the corresponding mixtures (b) obtained from the PLS-DA model calculated on the dataset of median colour values.

related to food composition and storage conditions, including pH, temperature, light exposure, oxidase enzyme activity, water activity, and total soluble content among others.

To tackle these problems, we have chosen to investigate how colour depends on SYP composition, by analysing a series of SYP samples whose composition varied according to an I-optimal mixture design. Furthermore, we focused on the effect of exposure to intense light at a temperature of 35 °C, measuring by RGB imaging how the colour of the product changed during a period of 5 weeks, compared to the colour of control samples with the same composition stored for the same period in the dark at a temperature between 0 and 4 °C.

The RGB images of the samples were analysed using different strategies, including multivariate methods that allowed to consider various colour parameters deriving from the RGB ones in order to explore different colour features of the samples. The results highlighted that the optimal parameter to monitor SYP colour is the median value of the red channel (R) of the RGB images, which as expected depends strongly on the amount of strawberry purée, but also on other factors such as the amount of lemon juice and the type of thickener, although this latter component is present in small amounts.

Furthermore, the comparison between the samples stored under stress conditions with the control samples showed a progressive browning over time, which is more pronounced the more the amount of strawberry purée is high, since anthocyanins are the thermolabile components of the mixture. Results also showed that the type of thickener had an effect on the colour degradation kinetics of the product. Finally, it was observed that, among the colour parameters extracted from the images, relative green tends to increase for stressed samples over time. This colour parameter, together with red, could be used as an index of the degradation process of SYP colour.

These results are only a first step in the understanding of such complex phenomena. Future applications of these preliminary results will be focused on the optimization and validation of the components proportions, in order to find the best compromise between a bright red colour of the fresh product and colour stability over time. In turn, this goal will be achieved through the development of targeted image analysis strategies aimed at building efficient and robust predictive models. The same approach could then be used on other strawberry-based products for the rapid and non-destructive evaluation of colour related properties on an industrial scale in view of the implementation of green and eco-friendly strategies to monitor product quality.

CRediT authorship contribution statement

Pier Lorenzo Rolando: Methodology, Investigation, Formal analysis, Writing – original draft. **Rosalba Calvini:** Conceptualization, Methodology, Investigation, Formal analysis, Writing – review & editing. **Giorgia Foca:** Conceptualization, Methodology, Investigation, Formal analysis, Writing – review & editing. **Alessandro Ulrici:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

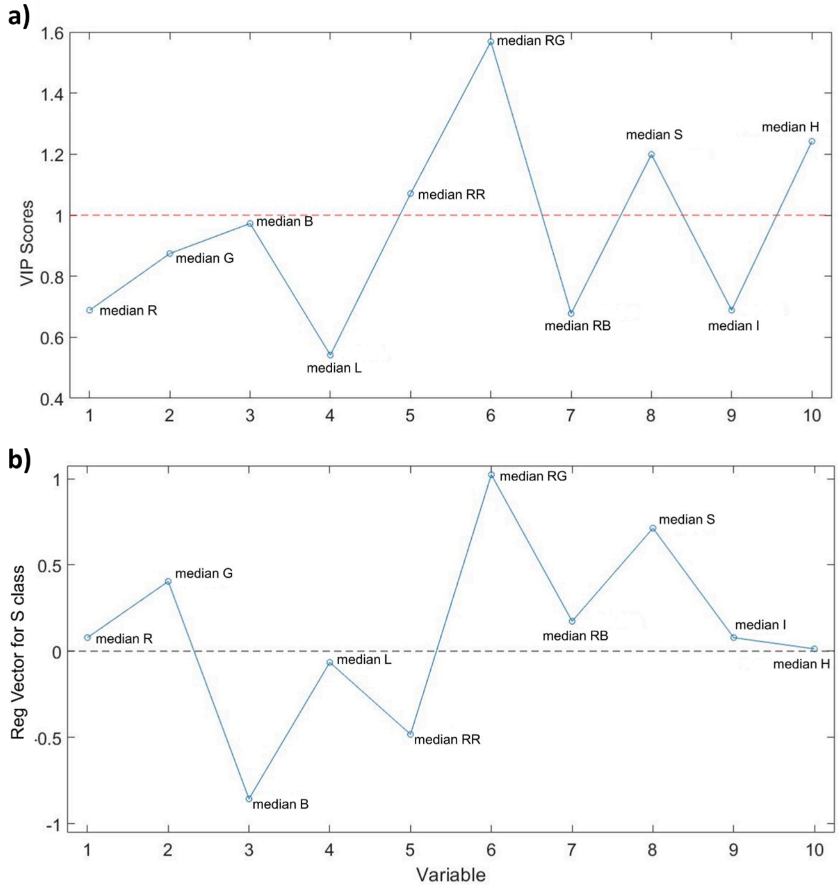

Fig. 6. PLS-DA model calculated on the dataset of median colour values: VIP scores (a) and regression vector for S class (b).

the work reported in this paper.

Acknowledgements

The authors gratefully acknowledge the staff of *DiSanaPianta Società Agricola Cooperativa* (Almese, Turin, Italy) for SYP ingredients supplying and assistance in the experimental phase. A special thanks to Riccardo Ceccato, production and R&D manager of the site.

Rosalba Calvini would like to thank the Italian funding programme Fondo Sociale Europeo REACT-EU - PON "Ricerca e Innovazione" 2014–2020 – Azione IV.6 Contratti di ricerca su tematiche Green (D.M. 1062 del 10/08/ 2021) for supporting her research (CUP: E95F21002330001; contract number 17-G-13884–4).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.microc.2023.109222.

References

[1] F.L. da Silva, M.T. Escribano-Bailón, J.J. Pérez Alonso, J.C. Rivas-Gonzalo, C. Santos-Buelga, Anthocyanin pigments in strawberry, LWT Food Sci. Technol. 40 (2) (2007) 374–382.

[2] T. Dzhanfezova, G. Barba-Espín, R. Müller, B. Joernsgaard, J.N. Hegelund, B. Madsen, D.H. Larsen, M. Martínez Vega, T.B. Toldam-Andersen, Anthocyanin prole, antioxidant activity and total phenolic content of a strawberry (Fragaria × ananassa Duch) genetic resource collection, Food Biosci. 36 (2020), 100620.

[3] D.B. Kovacevic, P. Putnik, V. Dragovic-Uzelac, N. Vahcic, M.S. Babojelic, B. Levaj, Influences of organically and conventionally grown strawberry cultivars on anthocyanins content and colour in purées and low-sugar jams, Food Chem. 181 (2015) 94–100.

[4] K. Ertan, M. Türkylmaz, M. Özkan, Colour and stability of anthocyanins in strawberry nectars containing various co-pigment sources and sweeteners, Food Chem. 310 (2020), 125856.

[5] F. Delgado-Vargas, A.R. Jimenez, O. Paredes-Lopez, F.J. Francis, Natural pigments: Carotenoids, anthocyanins, and betalains - characteristics, biosynthesis, processing, and stability, Crit. Rev. Food Sci. Nutr. 40 (2000) 173–289.

[6] E. Sadilova, R. Carle, F.C. Stintzing, Thermal degradation of anthocyanins and its impact on colour and in vitro antioxidant capacity, Mol. Nutr. Food Res. 51 (2007) 1461–1471.

[7] B.K. Martinsen, K. Aaby, G. Skrede, Effect of temperature on stability of anthocyanins, ascorbic acid and colour in strawberry and raspberry jams, Food Chem. 316 (2020) 1–9.

[8] A. Hartmann, C.-D. Patz, W. Andlauer, H. Dietrich, M. Ludwig, Influence of processing on quality parameters of strawberries, J. Agric. Food Chem. 56 (20) (2008) 9484–9489.

[9] A. Sulaiman, F.V.M. Silva, High pressure processing, thermal processing and freezing of 'Camarosa' strawberry for the inactivation of polyphenoloxidase and control of browning, Food Control 33 (2) (2013) 424–428.

[10] P. Goos, J. Bradley, U. Syafitri, I-Optimal Design of Mixture Experiments, J. Am. Stat. Assoc. 111 (514) (2016) 899–911.

[11] M. Holzwarth, S. Korhummel, T. Siekmann, R. Carle, D.R. Kammerer, Influence of different pectins process and storage conditions on anthocyanin and colour retention in strawberry jams and spreads, LWT Food Sci. Technol. 52 (2) (2013) 131–138.

[12] M. Holzwarth, J. Wittig, R. Carle, D.R. Kammerer, Influence of putative polyphenoloxidase (PPO) inhibitors on strawberry (*Fragaria × ananassa* Duch.) PPO, anthocyanin and colour stability of stored purées, LWT Food Sci. Technol. 52 (2013) 116–122.

[13] D. Wu, D.W. Sun, Colour measurements by computer vision for food quality control–A review, Trends Food Sci. Technol. 29 (1) (2013) 5–20.

[14] J.M. Prats-Montalbán, A. de Juan, A. Ferrer, Multivariate image analysis: A review with applications, Chemom. Intel. Lab. Syst. 107 (1) (2011) 1–23.

[15] P.M. Santos, P.D. Wentzell, E.R. Pereira-Filho, Scanner digital images combined with colour parameters: a case study to detect adulterations in liquid cow's milk, Food Anal. Methods 5 (1) (2012) 89–95.

[16] A. Solana-Altabella, M.H. Sánchez-Iranzo, J.I. Bueso-Bordils, L. Lahuerta-Zamora, A.M. Mellado-Romero, Computer vision-based analytical chemistry applied to determining iron in commercial pharmaceutical formulations, Talanta 188 (2018) 349–355.

[17] L. Pagnin, R. Calvini, R. Wiesinger, J. Weber, M. Schreiner, Photodegradation kinetics of alkyd paints: the influence of varying amounts of inorganic pigments on the stability of the synthetic binder, Front. Mater. 7 (2020), 600887.
[18] A. Antonelli, M. Cocchi, P. Fava, G. Foca, G.C. Franchini, D. Manzini, A. Ulrici, Automated evaluation of food colour by means of multivariate image analysis coupled to a wavelet-based classification algorithm, Anal. Chim. Acta 515 (2004) 3–13.
[19] A. Borin, M.F. Ferrão, C. Mello, L. Cordi, L. Pataca, N. Durán, R.J. Poppi, Quantification of Lactobacillus in fermented milk by multivariate image analysis with least-squares support-vector machines, Anal. Bioanal. Chem. 387 (3) (2007) 1105–1112.
[20] F. Masino, G. Foca, A. Ulrici, L. Arru, A. Antonelli, A chemometric study of pesto sauce appearance and of its relation to pigment concentration, J. Sci. Food Agric. 88 (8) (2008) 1335–1343.
[21] G. Foca, F. Masino, A. Antonelli, A. Ulrici, Prediction of compositional and sensory characteristics using RGB digital images and multivariate calibration techniques, Anal. Chim. Acta 706 (2) (2011) 238–245.
[22] A. Ulrici, G. Foca, M.C. Ielo, L.A. Volpelli, D.P. Lo Fiego, Automated identification and visualization of food defects using RGB imaging: Application to the detection of red skin defect of raw hams, Innov. Food Sci. Emerg. Technol. 16 (2012) 417–426.
[23] G. Orlandi, R. Calvini, G. Foca, A. Ulrici, Automated quantification of defective maize kernels by means of Multivariate Image Analysis, Food Control 85 (2018) 259–268.
[24] 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.
[25] A. Giraudo, R. Calvini, G. Orlandi, A. Ulrici, F. Geobaldo, F. Savorani, Development of an automated method for the identification of defective hazelnuts based on RGB image analysis and colourgrams, Food Control 94 (2018) 233–240.
[26] E.T. dos Santos Caramês, M.R. Baqueta, D.A. Conceição, J.A.L. Pallone, Near infrared spectroscopy and smartphone-based imaging as fast alternatives for the evaluation of the bioactive potential of freeze-dried acai, Food Res. Int. 140 (2021), 109792.
[27] Z. Liu, S. Yang, Y. Wang, J. Zhang, Discrimination of the fruits of Amomum tsao-ko according to geographical origin by 2DCOS image with RGB and Resnet image analysis techniques, Microchem. J. 169 (2021) 106545.
[28] R. Calvini, G. Orlandi, G. Foca, A. Ulrici, Colourgrams guiGUI: A graphical user-friendly interface for the analysis of large datasets of RGB images, Chemom. Intel. Lab. Syst. 196 (2020) 1–11.
[29] Colourgrams GUI for Matlab: http://www.chimslab.unimore.it/downloads/.
[30] D. Ballabio, F. Grisoni, R. Todeschini, Multivariate comparison of classification performance measures, Chemom. Intel. Lab. Syst. 174 (2018) 33–44.
[31] R. Gosselin, D. Rodrigue, C. Duchesne, A Bootstrap-VIP approach for selecting wavelength intervals in spectral imaging applications, Chemom. Intel. Lab. Syst. 100 (1) (2010) 12–21.