UNIVERSIDADE ESTADUAL DE CAMPINAS

FACULDADE DE ENGENHARIA DE ALIMENTOS

LUIS JAM PIER CRUZ TIRADO

HYPERSPECTRAL IMAGING FOR FOOD QUALITY CONTROL: COCOA BEANS HYBRIDS AND CHIA SEEDS SHELF-LIFE

IMAGENS HIPERESPECTRAIS PARA O CONTROLE DA QUALIDADE DE ALIMENTOS: HIBRIDOS DE GRÃOS DE CACAU E VIDA DE PRATELEIRA DE SEMENTES DE CHIA

CAMPINAS

2020

LUIS JAM PIER CRUZ TIRADO

HYPERSPECTRAL IMAGING FOR FOOD QUALITY CONTROL: COCOA BEANS HYBRIDS AND CHIA SEEDS SHELF-LIFE

IMAGENS HIPERESPECTRAIS PARA O CONTROLE DA QUALIDADE DE ALIMENTOS: HIBRIDOS DE GRÃOS DE CACAU E VIDA DE PRATELEIRA DE SEMENTES DE CHIA

Dissertation presented to the School of Food Engineering of the University of Campinas in partial fulfillment of the requirements for the degree of Master, in the area of Food Engineering.

Dissertação de mestrado apresentada à Faculdade de Engenharia de Alimentos da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em Engenharia de Alimentos.

Orientador: Prof. Dr. DOUGLAS FERNANDES BARBIN

ESTE EXEMPLAR CORRESPONDE À

VERSÃO FINAL DA DISSERTAÇÃO

DEFENDIDA PELO ALUNO LUIS JAM

PIER CRUZ TIRADO E ORIENTADA PELO

PROF. DR. DOUGLAS FERNANDES BARBIN.

CAMPINAS – SP

2020

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Faculdade de Engenharia de Alimentos
Claudia Aparecida Romano - CRB 8/5816

C889h

Cruz Tirado, Luis Jam Pier, 1992-
Hyperspectral imaging for food quality control: cocoa beans hybrids and chia seeds shelf-life / Luis Jam Pier Cruz Tirado. – Campinas, SP : [s.n.], 2020.

Orientador: Douglas Fernandes Barbin.
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Engenharia de Alimentos.

1. Híbrido de cacau. 2. Chia - Sementes. 3. Aprendizado de máquinas. 4. Vida de prateleira. 5. NIR. I. Barbin, Douglas Fernandes. II. Universidade Estadual de Campinas. Faculdade de Engenharia de Alimentos. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Imagens hiperespectrais para o controle da qualidade de alimentos: híbridos de graos de cacao de cacau e vida de prateleira de sementes de chia
Palavras-chave em inglês:
Cocoa hybrid
Chia - Seeds
Machine learning
Shelf life
NIR
Área de concentração: Engenharia de Alimentos
Titulação: Mestre em Engenharia de Alimentos
Banca examinadora:
Douglas Fernandes Barbin [Orientador]
Nuria Aleixos Borrás
Jose Blasco-Ivars
Data de defesa: 25-09-2020
Programa de Pós-Graduação: Engenharia de Alimentos

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-1963-4965
- Currículo Lattes do autor: http://lattes.cnpq.br/0860003173704816

BANCA EXAMINADORA

Prof. Dr. Douglas Fernandes Barbin – Orientador

Faculdade de Engenharia de Alimentos (FEA)

Universidade Estadual de Campinas (UNICAMP), Campinas - SP

Prof. Dr. José Blasco Ivars – Membro Titular

Instituto Valenciano de Investigaciones Agrarias - IVIA

Moncada, Valencia, Espanha

Dr. Nuria Aleixos Borras - Membro Titular

Departamento de Ingeniería Gráfica

Universitat Politècnica de València, Valencia - Espanha

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

This work is dedicated to my parents, Clara and Jorge, who motivate me every year to continue, even if they have to see me go.

To my sisters and brother, Marjory, Angie and Nick, for their unconditional affection.

To my grandmothers, Tarcila and Bartola, who were always mothers in my life.

To my grandfather, Francisco, from whom I couldn't say goodbye, sorry.

To my friends, from Peru, from Brazil and Belgium, they are and always will be my family.

AGRADECIMIENTOS

I thank God for taking care of me and for giving me strength every day to continue.

To my parents, Jorge and Clara, for teaching me to pursue my goals, not to lose myself, to be a correct person and to take advantage of each new day.

To my advisor, Prof. Dr. Douglas for his guidance, but above all for his friendship and his respect, for listening to me, supporting me, for being like a brother. Thank you for motivating me to always want more.

To my colleagues, Amanda and Marciano, for the conversations, the jokes, the partnership, the extensive audios of whatsapp (Amanda), the coexistence (that implies one or another beer (Marciano)), and for their respect, which is reciprocal.

To José Manuel Amigo, for his friendship and commitment. Thank you for always answering my emails and for your help in my professional development.

To Juan Fernández Pierna, Vincent Baeten and the workers of the CRA-W Research Center, for the support during my exchange, for the conversations and for the advice.

To my Belgian friends, Claudia Snipimoys Piccolo, David Beguin, Jérôme Dardenne, Vincent Exsteens and the guys of Gembloux Volleyball Team, for their friendship, the intense trainings, for the games, the hamburgers after training, for the countless beers, and for being that support during my stay in Belgium.

To my friends that I met during the master's degree, Noadia, Raquel, Ramon, Joaquim, Tatiane, Raffaela, Sabrina e Diego, Bárbara, Erick, Karina (gordinha), Monique, María Isabel, Sara Fraga, Bia, Fernanda Sievich, Cassia and María Paula, for the conversations, for the coffees, for the beers, for the parties, for the advices, for the camaraderie, for her friendship.

To my friends of Liga das Engenharias da Unicamp (LEU), George Paiva, Caio, Hugo, Pkizinho, Lucas (Megatron), Esperança, Gabriel (Contador), Vinicius, Leonardo, Victor, Mari Oliveira, Daniel (Bahía) and Ettore (Medina), for their friendship, for their love, for all training and games, for "after training" (Kadelao), for parties and for patience.

This work was supported by São Paulo Research Foundation (FAPESP) with project number 2019/04833-3 (BEPE fellowship), 2018/02500-4 (Master fellowship) and 2015/24351-2 (Young Researcher fellowship).

This work was carried out with the support of the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 (88882.329557/2019-01).

RESUMO

A imagem hiperespectral (HSI) permite a aquisição simultânea de informações espectrais e espaciais. Neste trabalho, HSI foi utilizado para o controle de qualidade de produtos agrícolas, que inclui a autenticação de híbridos de cacau e a estimativa do prazo de validade de sementes de chia. Para o trabalho com sementes de chia, as amostras foram armazenadas a 25, 35 e 45 ° C por 180 dias, para análises aceleradas do prazo de validade. Periodicamente, as amostras de chia eram removidas do armazenamento para obter imagens hiperespectrais (900 - 2500 nm), análise de acidez e perfil de ácidos graxos. O objetivo foi usar imagens hiperespectrais e análises multivariadas para desenvolver uma metodologia para estimar a vida de prateleira de sementes de chia, denominada Multivariate Accelerated Shelf Life Testing (MASLT). A Análise de Componentes Principais (PCA) foi usada para estudar a variabilidade durante o armazenamento e, em seguida, as pontuações do PC foram usadas para modelar a cinética e estimar os parâmetros da Equação de Arrhenius e, finalmente, para estimar a vida de prateleira. Além disso, pela primeira vez, uma nova estratégia foi proposta para validar essa metodologia, que chamamos de "Re-sampling", onde as amostras do conjunto de validação foram projetadas no conjunto de calibração com um número razoável de iterações. Os escores PC1 e gráficos cinéticos foram construídos ajustando os escores PC1 relacionados ao tempo versus o tempo por um modelo cinético fundido (R^2> 0,85). Os espectros de sementes de chia onde a acidez aumentou em 75% a partir do valor inicial foram usados para calcular o valor de corte (-0,9853). As estimativas de vida de prateleira foram 1300, 798 e 90 dias para sementes de chia armazenadas a 25, 35 e 45 ° C, respectivamente. Pela primeira vez, uma metodologia confiável é proposta para validar que todas as amostras foram previstas corretamente usando as pontuações PC1. No segundo estudo, cinco híbridos de cacau foram

cultivados e processados nas mesmas condições na CEPLAC (Medicilândia, Pará, Brasil). Os grãos de cacau foram então transportados para o Wallonie Research Center (Bélgica), onde foram obtidas imagens hiperespectrais na faixa de 1100 - 2500 nm. A análise parcial discriminante dos mínimos quadrados (PLS-DA) e a máquina de vetores de suporte (SVM) foram implementadas para classificar os híbridos de cacau, (1) duas classes de híbridos e (2) cinco classes de híbridos. Além disso, um novo conjunto de imagens foi usado para validação externa pixel a pixel. Os resultados mostraram que PLS-DA e SVM tiveramresultados comparáveis para modelos de duas classes (híbridos), mas o SVM (erro de previsão de 3,8 a 23,1%) foi superior ao PLS-DA (erro de previsão de 4,4 a 34,4%) quando todas as cinco classes de híbridos foram incluídas em um modelo. Os resultados de previsão pixel a pixel em um conjunto de imagens externas mostraram uma taxa de classificação correta de 50 a 100%. Os resultados para os modelos de duas classes e cinco foram comparáveis às técnicas de reação em cadeia da polimerase. Os resultados mostram o potencial do HSI para o controle de qualidade de produtos agrícolas, tanto para autenticação quanto para estimativa do prazo de validade.

PALAVRAS-CHAVE: híbrido de cacau; sementes de chia; aprendizado de máquina; re-sampling; imagem hiperespectral; NIR; vida de prateleira

ABSTRACT

Hyperspectral imaging (HSI) enables simultaneous acquisition of spectral and spatial information. In this work, HSI was used for quality control of agricultural products, which includes the authentication of cocoa bean hybrids and the estimation of shelf-life of chia seeds. Regarding the chia seeds study, samples were stored at 25, 35 and 45 ° C for 180 days, for accelerated shelf life analyzes. From time to time, chia samples were removed from storage to acquire hyperspectral images (900 - 2500 nm), acidity analysis, and fatty acid profile. The objective was to use hyperspectral images and multivariate analysis to develop a methodology for estimating the shelf-life of chia seeds, called Multivariate Accelerated Shelf Life Testing (MASLT). Principal Component Analysis (PCA) was used to study the variability during storage, and then, the PC scores were used to model the kinetics and estimate the parameters of the Arrhenius Equation, and finally to estimate the shelf life. Furthermore, for the first time a new strategy was proposed to validate this methodology, which we called "Re-sampling", where the samples from the validation set were projected onto the calibration set with a reasonable number of iterations. PC1 scores and kinetic charts were built fitting the time-related PC1 scores versus time by a fused kinetic model ($R^2 > 0.85$). The spectra of chia seeds where acidity increased at 75% from initial value were used to calculate the cut-off value (-0.9853). The shelf life estimations were 1300, 798 and 90 days for chia seeds stored at 25, 35 and 45 °C, respectively. For the first time, a reliable methodology is proposed to validate that all samples were correctly predicted using PC1 scores. In the second study, cocoa beans hybrids (five) were grown and processed under the same conditions in CEPLAC (Medicilândia, Para, Brazil). The cocoa beans were then transported to the Wallonie Research Center (Belgium), where hyperspectral images in the 1100 - 2500 nm range were acquired. Partial least square discriminant

analysis (PLS-DA) and Support vector machine (SVM) was implemented to classify cocoa bean hybrids, (1) two classes of hybrids and (2) five classes of hybrids. Additionally, a new set of images was used for external pixel-to-pixel validation. The results showed that PLS-DA and SVM demonstrate comparable results for two-class (hybrids) models, but SVM (3.8–23.1% prediction error) was superior to PLS-DA (4.4–34.4% prediction error) when all five classes (hybrids) were included in a model. Pixel-to-pixel prediction results on a set of external images showed a correct classification rate of 50 - 100%. The results for both the two-class models and the five-class model were comparable with polymerase chain reaction techniques. The results show the potential of HSI for quality control of agricultural products, both for authentication and estimation of shelf life.

KEYWORDS: cocoa bean hybrid; chia seeds; machine learning; re-sampling; hyperspectral imaging; NIR; shelf life

SUMMARY

CHAPTER 1:

GENERAL INTRODUCTION

1.1 General Introduction

The food industry is especially complex, both in its processes and in its supply chain. Initially, all supplies must be analyzed, to ensure they have the necessary quality, both in composition and authenticity. Next, during processing, the characteristics of the food must be evaluated at each stage, and in the product. Generally, here, at this stage, from a representative sample, the quality analysis results are expanded to entire batches of products. Finally, in the distribution chain, the products may be victims of economically motivated adulteration (EMA) or counterfeiting, potentially damaging prestigious brands and affecting consumer confidence.

Currently, science offers various technological approaches to analyze foods with great precision, such as mass spectrometry to quantify metabolites (Diomande et al., 2015) or nuclear magnetic resonance to identify the geographical origin of food (Caligiani, Palla, Acquotti, Marseglia, & Palla, 2014). However, the food industry is modernizing, within what is known as "Industry 4.0". Therefore, analysis and monitoring techniques that can be installed in the production lines are needed, preferably free of chemical reagents, that are precise, friendly and that allow analyzing a greater quantity of food in real time. To this end, imaging technology has become a powerful technique for food analysis. In this field, hyperspectral images (HSI) allow the simultaneous acquisition of spectral information, related to internal characteristics, and spatial information, associated with external physical characteristics (Hussain, Sun, & Pu, 2019). HSI has shown good performance for component quantification (Kamruzzaman, Makino, & Oshita, 2016), food classification (Velásquez, Cruz-Tirado, Siche, & Quevedo, 2017) and fraud detection (Orrillo et al., 2019).

In this research, the application of HSI for quality control in two products of high interest was proposed: chia seeds and cocoa beans. In the first case, Chia seeds were

stored for a long period, to evaluate their shelf life using accelerated conditions, acquiring hyperspectral images every certain period of time. With the established premise that (1) the samples had the same initial composition and (2) the changes in the samples are the effect of storage temperature, hyperspectral images were associated with the degradation of Chia seeds. Using the scores of the principal components (PCs) obtained from a Principal Component Analysis (PCA), it is possible to develop a methodology called Multivariate Accelerated Shelf Life Testing (MASLT). This method was previously established by Pedro & Ferreira (2006), therefore, the contribution in this research was to develop a validation strategy. For the first time, we used a "Re-sampling" strategy to validate that the samples were correctly predicted by MASLT. Therefore, it was possible to assume that the shelf-life was correctly estimated for each storage temperature. In the second case, HSI allowed to identify different classes of dry and fermented cocoa bean hybrids, which came from Pará (Brazil). The hybridization process was controlled, making it possible to select samples of hybrids of industrial interest. Furthermore, all the cocoa beans had the same drying process and fermentation time. The challenge in this study implied that the hybrids had similar ancestry, so their composition could be very similar. Thus, HSI should be able to discriminate hybrids only for small variations in composition. Accordingly, HSI showed a high performance to identify cocoa bean hybrids, with a correct classification of 40 - 100%.

The results encourage future studies to expand the technology to on-site applications, performance improvements using new artificial intelligence methods, and to include new samples and types of food.

CHAPTER 2:

LITERATURE REVIEW

2.1 Literature review

2.1.1 Hyperspectral imaging

Hyperspectral imaging (HSI) has been widely used in quality control of agricultural products (Jia et al., 2020). HSI allows to obtain spatial information (2 dimensions: X and Y) and spectral information (1 dimension: λ) simultaneously, obtaining a spectrum for each pixel in the image (Oliveira, Cruz-Tirado, & Barbin, 2019). Therefore, since each food has a particular composition, the spectra vary and can be used as a "*spectral fingerprint*" of that food. However, to analyze the large amount of information obtained from HSI, it is necessary to use multivariate analysis, in order to generate models that allow interpreting the variability in a data set, making correlations with characteristics or properties of foods, and obtaining and recognizing patterns to discriminate samples (Rehman, Mahmud, Chang, Jin, & Shin, 2019). Hyperspectral imaging has found several applications in quality control, safety, and in monitoring the processing of various agricultural products (Jia et al., 2020), including fruits and vegetables (Tsouvaltzis, Babellahi, Amodio, & Colelli, 2020), meat (Kamruzzaman et al., 2016; J. Ma & Sun, 2020), fish (Ivorra et al., 2013), milk and dairy products (Munir, Wilson, Yu, & Young, 2018), honey (Noviyanto & Abdulla, 2019), condiments (Oliveira et al., 2019), coffee (Calvini, Amigo, & Ulrici, 2017), cereals (Vermeulen, Suman, Fernández Pierna, & Baeten, 2018), among others.

2.1.2 Hyperspectral data processing

Hyperspectral images collect a large amount of information (thousands of data points), stored in pixels, with a high correlation between neighboring pixels (Vidal & Amigo, 2012). Therefore, the analysis of this information requires multivariate techniques, which allow analyzing hypercube and extracting useful information. However, before

performing any multivariate analysis, the images and spectra must be corrected and pre-processed, respectively. Some erroneous data values in the image such as *dead pixels* and *spike points* must be corrected beforehand and the background must be removed. Similarly, spectral data must be pre-processed to eliminate defects associated with the effect of particle size, light scattering, or morphological differences (surface roughness and detector artifacts) (Amigo, 2010).

2.1.2.1 Region of interest (ROI)

Generally speaking, a hyperspectral image is made up of the sample and the background. Depending on the geometry of the sample, background removal may or may not be an easy task. Commonly, a mask is created using wavelengths where there is a high contrast between the ROI and the background. Although it can also use methods such as "clustering", “PCA scores”, “histograms” or "PLS-DA" to remove the background (Oliveira et al., 2019; Vidal & Amigo, 2012).

2.1.2.2 Image correction

Image defects such as dead pixels and spike points are associated with anomalies in the detector, environmental conditions, or defects in the equipment components.

Spike points are defined as a sudden increase in spectrum, followed by a rapid decrease, which generally hides important imaging information (Zhang & Henson, 2007). For spike detection, manual monitoring is commonly used, which is difficult due to the high amount of information obtained for a hypercube (Nenadic & Burdick, 2005). Other techniques, such as comparing neighboring pixel information (Behrend, Tarnowski, & Morris, 2002) and wavelet transform (Ehrentreich & Summchen, 2001) can be used to remove or interpolate spikes, between other (Cannistraci, Montevecchi, & Alessio, 2009; Feuerstein, Parker, & Boutelle, 2009).

On the other hand, according to Firtha et al. (2008) a dead pixel (zero or missing values) number in a NIR hyperspectral imaging represents approximately 1% of total of pixels. They can be one pixel, a group of pixels or a full line of pixels (Vidal & Amigo, 2012). Dead pixels can be located using different criteria, which are well established in previous reviews and tutorials (Burger & Geladi, 2005; Mobaraki & Amigo, 2018). As mentioned above, due to the high correlation between neighboring pixels, it is possible to replace the dead pixels by interpolating the mean values of the neighboring pixels. It is important to mention that the presence of dead pixels and spike points can generally lead to errors in multivariate analysis. Some algorithms like PCA or MCR are highly influenced by a high number of dead pixels, distorting the result (Vidal & Amigo, 2012).

2.1.2.3 Spectral pre-processing

After extracting the spectral information from HIS, spectral pre-processing helps to minimize (or correct) the effect of unwavering phenomena that affect spectral measurement. These phenomena can be light scattering, the effect of particle size, differences in morphology, porosity, roughness and detector artifacts (Amigo, 2010). However, this step can be carried out with care, since an excess of the pre-processing (for example: an improper selection of window size in the smoothing) can cause loss of information (Jia et al., 2020). At present, there are spectral pre-processing techniques that can help solve these problems, such as:

- *Smoothing*: it allows the removal of part of the instrumental noise, with the Savitzky-Golay algorithm being the most widely used (Vidal & Amigo, 2012).

- *Lighter scattering correction*: light scattering is a problem associated with the acquisition of HSI spectra (especially in the NIR region), and it is common when

obtaining reflectance spectra of solid and semi-solid materials (Burger & Geladi, 2007). Among the algorithms most used to reduce the effect of light scattering we have: Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) (Mendez, Mendoza, Cruz-Tirado, Quevedo, & Siche, 2019).

- *Baseline correction*: Savitzky-Golay derivatives (1^{st} and 2^{nd} order) are the most popular algorithm to correct baseline and slopes. At the same time, the derivatives emphasize the characteristics of the spectrum. Therefore, it is of essential importance to be careful in the selection of the derivative parameters: derivative order, polynomial order and window size, to avoid noise being emphasized and to have erroneous measurements (Rinnan, Berg, & Engelsen, 2009).

2.1.3 Chemometrics

Multivariate analyses applied to predict chemical features are called chemometrics. There are several multivariate analysis methods for the analysis of hyperspectral images that are well discussed in previous reviews (Amigo, Babamoradi, & Elcoroaristizabal, 2015; Cortés, Blasco, Aleixos, Cubero, & Talens, 2019; Fernández Pierna et al., 2020; Jia et al., 2020), therefore, here we only expose slightly the multivariate methods most relevant for this research.

2.1.3.1 Principal component analysis (PCA)

PCA is probably the most widely used chemometric tool for removes multi-collinearity, as well as dimensionality reduction and feature extraction, using its functionalities such as solving for non-full-rank eigen problems, ellipse fitting, noise reduction and translation error attenuation (Wu, Chen, Ding, Hsu, & Huang, 2013). The component matrix transformation tries to find a new coordinate whose origin is the mean of the input data (spectra), reaching a maximum variance and generating a new uncorrelated

principal components (PCs) but preserved of maximum of information of original data (Hashim et al., 2012). Generally, the first PCs contain the greatest variability in the data (J. Li, Rao, & Ying, 2011). Therefore, the first main components (PC1 - 3) are generally used to show variations in samples as a consequence of storage conditions.

2.1.3.2 Regression analysis

- *Partial least squares regression (PLSR)*: it is perhaps the most popular multivariate linear model for monitoring the evolution (quantification) of components in a food throughout its shelf-life. PLSR linearly relates two matrices: the matrix X containing the input data (spectra) and the matrix Y containing the responses (i.e. moisture content), but in addition, it also models the structure of X and Y (Wold, Sjöström, & Eriksson, 2001). PLSR has the ability to analyze a large number of noisy and correlated variables, which is the case for spectral data. Furthermore, PLSR models can be improved by including more variables or a larger number of observations (samples) (Wold et al., 2001).

2.1.3.3 Discriminant analysis

- *Linear discriminant analysis (LDA)*: LDA is a probabilistic parametric classification technique whose objective is to find linear combinations of the X variables (discriminant functions) that discriminate between the classes (Sjöström, Wold, & Söderström, 1986). LDA maximizes the variance between classes and minimizes the variance within a class, obtaining orthogonal linear discriminant functions equal to the number of classes minus one (Meloun, Forina, & Militky, 1992).

- *Partial least squares discriminant analysis (PLS-DA)*: since the spectral data has a high correlation, a PLS version of LDA is required (Sjöström et al., 1986). PLS-DA is basically a PLS model, which has a variation on the dependent variables (matrix Y).

The dependent matrix Y is encoded with values of 0 and 1, that describes which class each sample belongs to (spectrum, matrix X) (Liu, He, & Wang, 2008). PLS-DA then rotates the latent variables to obtain a weight vector that promotes the best correlation with the response variable a and separating the classes (Lavine & Davidson, 2006). To delimit the classes, a cut-off value between 0 and 1 is established using probability density functions and Bayesian theory (Ferreira, 2015).

- *Support Vector Machine (SVM)*: SVM is supervised learning method, which reduced number of samples, called *support vectors*, where the input data is mapped into a high-dimensional vector space by a specific mapping function (Menesatti et al., 2009). Here, the nonlinear separable problem can be transformed into a linear separable problem (Grelet et al., 2020; X. Li, Zhu, Ji, & Liu, 2010). Therefore, by working non-linear and linear way, SVM can, in many cases, offer better performance in both classification and prediction.

2.1.4 Chia seed – Shelf life

Chia (*Salvia hispanica* L.) is oleaginous seed with high contents of essential fatty acids ω-3 and a favorable ratio ω-6/ω-3 for human consumption (de Falco, Amato, & Lanzotti, 2017). In addition, phenolic compounds from Chia are related to a protective effect against oxidative stress and obesity-related diseases (Marineli et al., 2014), reduced risk of cardiovascular disease and have hepatoprotective effect (Poudyal, Panchal, Waanders, Ward, & Brown, 2012). Industrially, chia oil has been commercialized throughout South America, since their species extends from Mexico to Argentina. This shows that these seeds are of great interest, especially in developed countries. The fatty acid and phenolic composition of chia seeds is dependent of origin, since climatic and cultivation conditions influence the development of these compounds. However, during storage, the acidity value, which is one of the most

important indicators for oil quality of chia seeds, can increase. High acidity values indicate that the product has been stored incorrectly or for long periods of time, making it difficult to process and causing loss of nutritional quality (Franklin et al., 2017; Mata et al., 2017). Therefore, it is important to estimate the shelf life of oilseeds such as chia, especially in tropical countries where storage conditions may not be ideal.

A more practical definition describes the shelf-life of a foodstuff as the *"duration of the consumer's sensory acceptability"*. Therefore, each food has a specific shelf-life, and this depends both on its specific quality characteristics and the storage conditions (Moschopoulou, Moatsou, Syrokou, Paramithiotis, & Drosinos, 2019). Currently, many of these shelf-life estimations are based on trial and error methods, with many risks of overestimating or underestimating the real shelf-life of foodstuffs (Wibowo, Buvé, Hendrickx, Van Loey, & Grauwet, 2018).

Because the shelf-life-defining degradation reactions are many and very complex, new multivariate analysis approaches are necessary for correct prediction. For this purpose, there are technologies such as gas chromatography-mass spectrometry (GC-MS) (Nzekoue et al., 2019), Liquid chromatography – mass spectrometry (LC-MS) (Coelho et al., 2020), Nuclear magnetic resonance (NMR) (Bosmans, Lagrain, Ooms, Fierens, & Delcour, 2014), electronic nose sensor (Giovenzana, Beghi, Buratti, Civelli, & Guidetti, 2014; Song et al., 2019), computational vision system (CVS) (Taheri-Garavand, Fatahi, Omid, & Makino, 2019), Near-infrared spectroscopy (NIRS) (Di Egidio et al., 2009; Pérez-Marín et al., 2019) and Hyperspectral imaging (HSI) (Chaudhry et al., 2018; Siripatrawan & Makino, 2018), which, in tandem with multivariate analysis, allow to model the kinetics of degradation (or transformation), estimate the shelf-life and predict the evolution of compounds of interest in a food during storage.

Using hyperspectral imaging, in a controlled shelf-life experiment where degradation reactions are caused by temperature ($T_1$and T_2) and storage time (t_0, t_1, ..., t_n), the spectra for the same temperatures are different for each day of storage, and vice versa. For example, as shown in Figure 1, $\lambda(T_1, t_0) \neq \lambda(T_1, t_1)$ or $\lambda(T_1, t_n) \neq \lambda(T_2, t_n)$. Other factors may also influence spectral variation during shelf-life, such as packaging (Taghizadeh, Gowen, Ward, & O'Donnell, 2010), coatings (Yousuf, Qadri, & Srivastava, 2018), or special storage conditions (for example: modified atmosphere) (L. Ma, Zhang, Bhandari, & Gao, 2017; Tsironi, Ntzimani, & Taoukis, 2019).

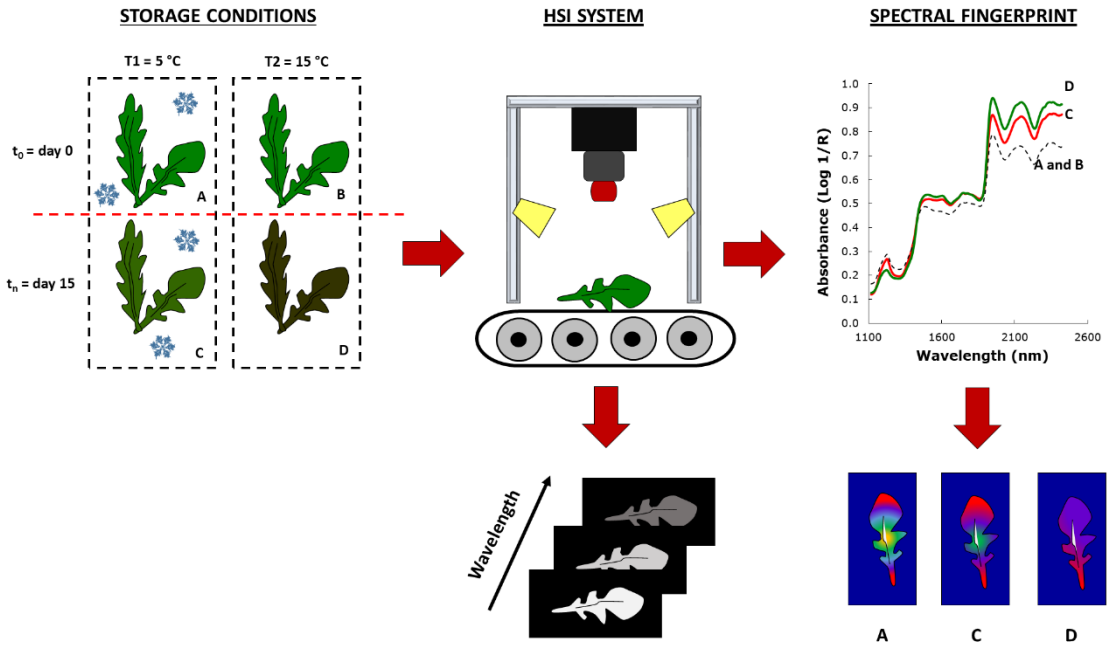


Figure 1. Spectral variations in a food as influence of temperature and storage time for spinach leaves

2.1.5 Cocoa bean hybrids – Authentication

Cocoa (*Theobroma cacao*) is one of the highly demanded crops, which are produced in tropical and sub-tropical regions. The largest producers are: (1) Cote D'Ivoire, (2) Ghana, (3) Indonesia, (4) Brazil, (5) Nigeria, (6) Cameroon, and (7) Ecuador (Teye,

Anyidoho, Agbemafle, Sam-Amoah, & Elliott, 2020). After being cultivated, the cocoa beans go through a fermentation process (between 5 - 7 days) and a drying process, which allows their particular flavor to develop. Cocoa beans presented a higher content and quality in proteins, vitamins, polyphenols, fat and carbohydrates compared to tea or wine (Lee, Kim, Lee, & Lee, 2003). Furthermore, cocoa bean is the raw material for chocolate and confectionery products, becoming more attractive and more consumed (Aprotosoaie, Luca, & Miron, 2016). However, the chocolate quality depends on the quality of the cocoa bean, which is influenced by geographic origin, soil and environmental conditions, growing and harvesting conditions, and post-harvest processing conditions (fermented and dried) (Efraim et al., 2013). Another factor that influences cocoa bean quality is the genetic variety. Hybridization is the technique that allows to create cocoa bean hybrids in order to improve some of its characteristics, such as resistance to diseases, greater efficiency, earliness and butterfat flavor expressed after optimal fermentation (Ji et al., 2013). Therefore, in a first approach, cocoa bean hybrids can have different prices in the market, making it possible for unscrupulous people to adulterate product batches. Second, the processing of cocoa beans is still rudimentary, thus, it is possible that different types of hybrids are mixed in the same batch, reducing the purity of the product.

Various analytical techniques such as multi-element analysis (Diomande et al., 2015), NIR spectroscopy (Barbin et al., 2018), microsatellites (Herrmann et al., 2015), mass spectrometry (Scollo, Neville, Oruna-Concha, Trotin, & Cramer, 2020), Raman spectroscopy (Vargas Jentzsch et al., 2016), computer vision (Mite-Baidal et al., 2019) and Polymerase chain reaction (PCR) (Motilal & Butler, 2003) have been developed to identify different cocoa hybrids. Although the results are encouraging, in some cases the focus was to identify just one variety of cocoa beans, while in others the sample had to

be ground and conditioned for analysis. Therefore, there is still a need to develop new methodologies to identify various varieties of cocoa from the same batch, without the need to destroy the sample.

Objective

In this work, it is proposed the application of hyperspectral imaging for the prediction of shelf-life of chia seeds, and to identify cocoa beans from different hibrids.

CHAPTER 3:

Shelf life estimation and kinetic degradation modeling of chia seeds (*Salvia hispanica*) using principal component analysis based on NIR-hyperspectral imaging

The results of this chapter is under peer review in *Food Control*

Shelf life estimation and kinetic degradation modeling of chia seeds (*Salvia hispanica*) using principal component analysis based on NIR-hyperspectral imaging

J.P. Cruz-Tirado[a]; Marciano Oliveira[a]; Milton de Jesus Filho[b]; Helena Teixeira Godoy[b]; José Manuel Amigo[c,d]; Douglas Fernandes Barbin[a]

[a] *Department of Food Engineering, University of Campinas, Rua Monteiro Lobato, 80, Cidade Universitária, Campinas, SP 13083-862, Brazil.*

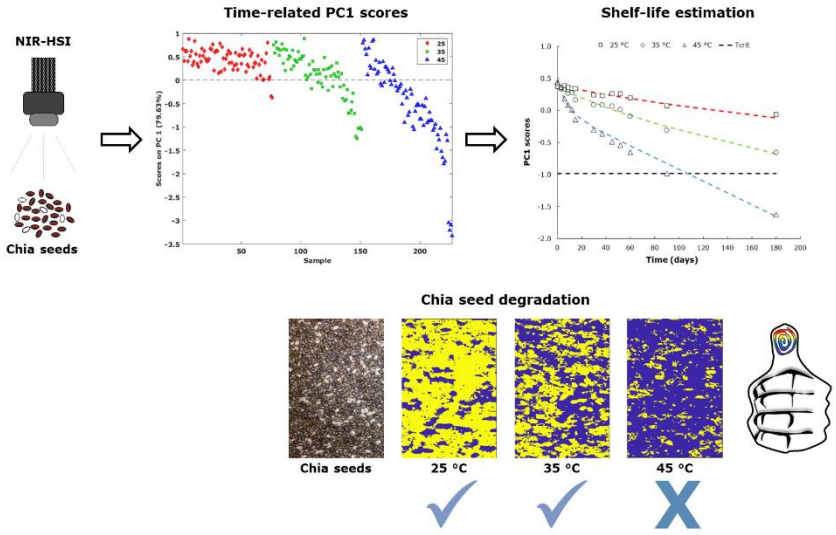
[b] *Department of Food Science, University of Campinas, Rua Monteiro Lobato, 80, Cidade Universitária, Campinas, SP 13083-862, Brazil.*

[c] *Ikerbasque, Basque Foundation for Sciences. María Díaz de Haro, 3. Bilbao 48013. Spain.*

[d] *Department of Analytical Chemistry, University of the Basque Country. Barrio Sarriena S/N. Leioa 48940. Spain.*

Corresponding author: dfbarbin@unicamp.br

Graphical abstract



Abstract

A new methodology based on Near Infrared-hyperspectral imaging and Principal Components Analysis (PCA) was developed and accurately validated to model the degradations kinetics and to estimate the multivariate accelerated shelf life (MASLT) of chia seeds (*Salvia hispanica*). Chia seeds were stored during 180 days at 25, 35 and 45 °C, observing fatty acid degradation and an increasing in the acidity. PC1 scores and kinetic charts were built fitting the time-related PC1 scores versus time by a fused kinetic model ($R^2 > 0.85$). The spectra of chia seeds where acidity increased at 75% from initial value were used to calculate the cut-off value (-0.9853). The shelf life estimations were 1300, 798 and 90 days for chia seeds stored at 25, 35 and 45 °C, respectively. For the first time, a reliable methodology is proposed to validate that all samples were correctly predicted using PC1 scores.

Keywords: Chia seeds; shelf-life; NIR Hyperspectral Imaging; PCA; kinetic degradation.

1. Introduction

Chia (*Salvia hispanica* L.) is an annual herbaceous plant belonging to the Salvia category of the *Labiatae* family. Chia seeds are the most consumed part of this plant, and they are commonly consumed as whole seeds, seed oil, seed flour and seed mucilage. Chia seeds nutritional value is centered in its higher concentration of polyunsaturated fatty acids, mainly α-linolenic acid (ω-3) (59.9–63.2 g/100 g) (Oliveira-Alves et al., 2017). In addition, chia seeds present a higher protein content (19.0–26.5%), phenolic compounds (highlighting rosmarinic, caffeic, and gallic acids), dietary fiber (47.1 to 59.8%) (de Falco et al., 2017; Grancieri et al., 2019), mucilage (Muñoz et al., 2012), vitamins, tocopherols and minerals. All together, they make chia a highly

nutritional value seed, so quality control regarding the nutritional composition of chia seeds is industrially important.

Depending on the market value, chia seeds can be stored for long periods, often without adequate control of storage conditions, producing the aging of the seed. Aging is a complex process that could lead to a diverse changes in chia seeds, including modifications in taste, flavor, fatty acid composition, protein, phenolic compounds and starch (Caruso et al., 2018). Storage conditions such as relative humidity and temperature are the most important factors that can cause important variations in seed quality during storage time, especially in tropical and subtropical regions (e.g. Brazil) (Delouche et al., 2016).

Accelerated shelf life tests (ASLT) is a method commonly used to determine the shelf life of any food. ASLT consists of collecting data related to quality parameters for various storage conditions (market and severe conditions) at different times and building kinetic models for shelf-life. By using ASLT, it is possible to evaluate the reaction velocity profile and to determine the reaction order, in order to convert the data obtained from the accelerated tests into the normal market conditions (Pedro & Ferreira, 2006). Some studies reported ASLT approach to predict shelf-life in seeds such as peanuts, linseeds (Cämmerer and Kroh, 2009) and chia seeds (Caruso et al., 2018).

For any food, the shelf life is defined by the variation of several quality parameters, and each parameter has its own cut-off values. Although many cut-off values of the parameters evaluated may be in the literature, legislation or within the quality parameters of each industry, when analyzing these parameters together there may be complications (Pedro & Ferreira, 2006). In this regard, the multivariate analysis techniques (like Principal Component Analysis (PCA)) can be useful for shelf life testing in several foods. PCA use linear combinations of the original variables to create

a new set of axes, called Principal Components (PCs). PC scores collect the global quality information of the sample, therefore, when incorporated into ASLT, a Multivariate Accelerating Shelf Life Testing (MASLT) method is established (Pedro & Ferreira, 2006).

Since all samples have the same initial composition and that storage conditions have been controlled correctly, MASLT assumes that the main cause of variation in PC scores (especially the first few PCs) are degradation reactions and, hence, scores values are related to time (Pedro & Ferreira, 2006). Firstly, MASLT use PC scores resulting of PCA model to describe the kinetic of degradation reactions. Subsequently, PC scores are used to model the Arrhenius equation allowing to explain the temperature dependence of degradation rate constants. Finally, the information obtained allows to calculate the shelf-life time (Chaudhry et al., 2018). Despite MASLT has been successfully tested in foods such as tomato paste (Pedro & Ferreira, 2006), sunflower oil (Upadhyay and Mishra, 2015), and fresh-cut lettuce (Derossi et al., 2016), its application in chia seeds has not been explored yet. It should be considered that MASLT is commonly applied using PC scores of a PCA model created with information from various specific parameters, which in the case of seeds may take several months for collecting data. Also, grinding seeds accelerate degradation reactions by exposing a larger area to temperature effect. Therefore, seed industries have focused their horizon towards the application of new non-destructive technologies, which are quick and efficient, that require a minimum of sample preparation and that can be applied in various parts of the supply chain.

Near-infrared hyperspectral imaging (NIR-HSI) is a technology that meets all those requirements. NIR-HSI allows a unique spectral fingerprint for each food, that is related with its chemical composition. NIR-HSI allows obtaining spatial information, which

often helps to overcome the problem of food heterogeneity. Therefore, it is possible to acquire images every certain period of time to observe the spectral variations as a consequence of the degradation reactions and use the spectral data as a quality parameter. This approach was previously employed to determine shelf- life in rocket leaves using vis-NIR HSI with MASLT (Chaudhry et al., 2018).

For chia seeds, NIR-HSI, can offer greater advantages, since degradation reactions can not necessarily generate color changes. While internal changes in the seed are result of lipid oxidation reactions, changes in carbohydrates and phenolic compounds may be expressed in the intensity of the spectral absorbance. In this context, the hypothesis of this work is that spectral variations in NIR-HSI may reflect chemical changes in chia seeds caused by temperature during storage. In addition, for the first time an attempt is made to design a "re-sampling" algorithm to validate the MASLT method to predict shelf-life.

2. Material and methods

2.1 Reagents

Chloroform (Synth, Brazil), methanol (J.T. Baker, EUA), sodium sulfate (Exôdo, Brazil), hexane (Fisher Scientific, EUA), sodium chloride (Synth, Brazil), sodium hydroxide (Exôdo, Brazil), boron trifluoride solution (BF_3) at 12% in methanol (Sigma Aldrich, EUA). Standards of methyl esters C4 - C24 and internal standar C23:0 were purchased from Sigma Aldrich (EUA).

2.2 Samples preparation and storage

Three batches of chia (*Salvia hispanica* L.) seeds were donated by R&S BLUMOS Industrial e Comercial Ltda (São Paulo, Brazil). Samples were transported to Laboratory of Food Inovation (LINA) (Campinas, Brazil), where they were inspected to

eliminate possible impurities or foreign materials. According to information obtained from sample supplier, 100 gr chia seeds contained: < 1g carbohydrates, 21g proteins, 31g lipids (3.4g saturated lipids, 19.72g alfa-linolenic acid and 5.44g linolenic acid) and 38g dietary fiber. Also, microbiological quality of the chia seeds was in accordance with the regulations of the Brazilian standards and there was no presence of insects (or part of them) in the samples.

Representative samples of 250 g of chia seeds were packed in glass containers and sealed using parafilm (Sinergia Científica, Brazil). Then, sample was stored at 25 °C (reference temperature), 35 °C and 45 °C (accelerated temperatures) in three different BOD TECNAL TE-371 climatic chambers, without light incidence on the samples. Maximum temperature variation in the chambers was ± 0.3 ° C. Twelve replicates were prepared for each storage temperature. Samples (approximately 10g) were taken for image acquisition at 0 (initial), 3, 6, 9, 12, 15, 30, 37, 45, 52, 60, 90, 180 days of storage.

2.3 Fatty acid composition and free fatty acid

Chia seeds were ground using a mill (model A 11 B S32, IKA, Germany). Later, chia oil was extracted by Bligh-Dyer method (Hartman and Lago, 1973). This method allows extracting lipids from chia seeds without applying heat, so it can be used to assess oil deterioration as a result of storage conditions. For fatty acid composition measurement, the lipids obtained were esterified as reported by Joseph and Ackman (1992) and the chromatographic conditions were based on Ballus et al. (2014), with modifications. Separation of methyl esters was performed on a 7890A gas chromatograph (GC-Agilent, Germany) equipped with a flame ionization detector (FID). The methyl esters were separated using a DB 23 capillary column (60 m, 0.25 mm d.i., 0.25 μm film thick, Agilent, USA). Methyl esters were identified by comparing

their retention times with those obtained with the standards (FAME mix C4-C24) under the same chromatographic conditions. Quantification was performed by internal standardization, using C23:0 as internal standard. Correction factors and fatty acid concentration (mg/g oil) were calculated according to Joseph and Ackman (1992). The methodology is reported extensively in Supporting Information.

Free fatty acid in stored chia seeds were determined according to Ca 5a-40 (AOCS, 1998). All samples were analyzed in triplicate.

2.4 NIR-HSI acquisition and processing

Hyperspectral images were acquired using a SisuCHEMA SWIR hyperspectral camera (Specim Spectral Imaging Ltd, FIN-90571 Oulu, Finland), in the NIR range 928 – 2524 nm with a spatial resolution of 320 pixels per line scan and spectral resolution of 10 nm. Hyperspectral images were acquired at 6.23 nm intervals in 256 wavelength channels. The spectra were acquired with an exposure time of 2.1 ms using a 50 mm lens. The instrumental calibration was performed using two-dimensional reference images: the dark (0% reflectance) and the white (~99% reflectance). Then, the Evince software (UmBio AB, Sweden) automatically subtracted the white and dark references from subsequently acquired images.

Chia seed samples were dispersed on a Teflon plate (5 mm thick) to acquired images. Twelve samples were taken for initial day (day 0) and eighteen samples were taken for subsequent days (6 samples per temperature). Hyperspectral images of samples were segmented using Evince software (UmBio AB, Sweden) and self-developed code in Matlab R2016a software (The Mathworks Inc., Natick, MA, USA) was used to extract the mean spectra and corrected base line effects using Standard Normal Variate (SNV), producing one mean spectrum per replicate.

2.5 Multivariate accelerated shelf life testing (MALST) modelling

For MALST algorithm design, traditional convention in linear algebra is followed: boldface upper case represents matrices, boldface lower case represents vectors, italic lower case represents scalar quantities, and italic subscripts denote case letters and sequences.

(1) First step: a matrix **X** (MxN), representing variability in quality parameters in seeds at 25, 35 and 45 °C, was constructed. In this matrix, m is the number of points collected during the time of storage (0, 3, 6, 9, 12, 15, 30, 37, 45, 52, 60, 90 and 180 days) where image acquisition was performed for each storage temperature; n represents the number of variables or wavelengths (256) in NIR range (928 – 2524 nm) included in the study. Since the spectral profiles serve as an attribute of property, mean centering was used for the normalization of the data (Chaudhry et al., 2018).

(2) Second step: PCA model based on matrix **X** (after data mean centering) was performed using PLSToolbox (Eigenvector Research, Seattle, USA) in Matlab R2016a (The Mathworks Inc., Natick, MA, USA). Loading matrix corresponding to PC1-2 was plotted vs storage time in order to delete regions in NIR range that did not provide any information to PCA model. A self-developed Matlab R2016a code was used to create scores matrix (**S**) of time-related PC scores in each storage temperature (25, 35 and 45 °C). Then, matrix **S** was plotted *vs* storage time to describe PC scores variation during storage time, well-known as kinetic plots or shelf-life charts. Kinetic plots were used to model reaction order (Eq. 1) and to estimate the multivariate kinetic parameters.

Quality degradation kinetics can be represented by equation (1):

$$\frac{dP}{dt} = kP^n \qquad (1)$$

where *P* denoted any quality attribute, *t* is the storage time, *n* is the reaction order and *k* is the degradation rate (*k* is negative if *P* decreases with time).

Also, acceleration factor (α_m) was calculated for 35 °C and 45 °C storage temperatures, according to Eq. 2:

$$\alpha_{T+\delta T,T} = \frac{k_{T+\delta T}}{k_T} \qquad (2)$$

where $\alpha_{T+\delta T,T}$ is the acceleration factor, *T* is the actual market temperature (25 °C), and *k* is the rate degradation for market and accelerated test conditions.

Further, because the degradation rates show a direct dependence on the storage temperature, Arrhenius equation (Eq. 3) (Labuza, 1982) was used to estimate dependence temperature of each kinetic model.

$$k = C * \exp\left(-\frac{E_a}{RT}\right) \qquad (3)$$

in which *C* is the pre-exponential or frequency term, E_a is the activation energy, *R* (8.314 J/mol) is the universal gas constant with a constant, T_{ref} is the reference temperature (25 °C).

All kinetic parameters were estimated using the Levemberg-Marquardt algorithm for non-linear fitting included in Curve Fitting toolbox in Matlab R2016a.

(3) Third step: The cut-off criteria (t_{crit}) for the property under study is the most important and significant aspect of the MASLT methodology. In this work, we used the spectra of an unacceptable sample and the loadings to simultaneously calculate the cut-off criteria x, according to Eq. 4:

$$\boldsymbol{t_{crit}^{T} = xa^{T} * L} \qquad (4)$$

where t_{crit}^{T} is the vector of critical scores, xa^{T} is a row vector and L is the loading matrix of the time-related PCs obtained in step 2. Finally, for calculating the actual shelf-life of chia seeds using $\alpha_{T+\delta T,T}$ and t_{crit}^{T} (Pedro & Ferreira, 2009).

The visualization of degradation/transformation process of Chia seeds during storage (0, 9, 30, 45, 60, 90 and 180 days) at different temperatures (25, 35 and 45 °C) were processed using PLS_Toolbox in Matlab R2016a.

2.6 Acidity value as cut-off criteria

As previously mentioned, the main advantage of chia consumption is its oil with a high content of polyunsaturated fatty acids. The acidity is defined as the amount of fatty acids no longer linked to their parent triglyceride molecules (Free fatty acids, FFA), and is one of the most important indicators for oil quality. High amounts of FFA indicate that the product has been stored for long periods of time and/or under inadequate storage conditions. In addition to nutritional and energy losses, the high FFA content increases the possibility of hydrolysis in the presence of moisture, during storage and industrial processing (Mata et al., 2017) and could affect the consumer acceptability (Franklin et al., 2017). Chia oil does not have a unique regulation for oil quality control, compared to olive oil, where the European Commission Regulation 2568/91 and subsequent amendments impose a maximum acidity of 0.8 g oleic acid/100 g oil (EC, 1991).In this sense, we arbitrarily defined a reference limit for the increase of acidity value as 75% of the initial content.

2.7 Validation

For the validation of the methodology proposed in this work, we developed a re-sampling-based algorithm that allows the projection of the validation samples on the calibration samples with a reasonable number of random iterations. For this purpose,

chia seeds samples were divided into calibration (70%) and validation (30%) sets. A PCA model was built using the calibration samples, and the PCs were projected onto the validation samples. The procedure was repeated for a number of iterations equal to 100, to ensure that all samples have been included at least once in the calibration and validation sets. Then, the frequency and standard deviation of the samples were calculated. The validation was performed using all the spectral information of the hyperspectral images.

2.8 Statistical analysis

Analyses were performed in triplicate for each sample for all tests, and data is presented as means ± standard deviation (SD). Analysis of variance (ANOVA) and Tukey's average comparison test ($p < 0.05$) were performed to compare fatty acid composition and free fatty acid values during storage using Statistica software version 7.0 (Statsoft, Oklahoma, USA).

3. Results and discussion

3.1 Fatty acid composition and free fatty acid

Table 1 shows the evolution on fatty acid profile of chia seeds stored at 25, 35 and 45 °C during 180 days. Palmitic (C16:0), stearic (C18:0) and Arachidic (C20:0) fatty acids increase (but not significantly, $p > 0.05$) during storage at 25 ° C and 35 ° C, reaching an increase of 5 and 8%, 13 and 22% and 15 and 25%, respectively. However, at 45 ° C these fatty acids are significantly degraded, reaching, at end of storage (180 days), reduction of 44% for palmitic acid, 60% for stearic acid and 93% for Arachidic acid. Imran et al. (2016) reported that, for raw chia oil stored during 60 days, an increase in palmitic acid of 28.8 – 30.8% and for stearic acid of 51.9 – 61.7%, when chia oil was stored at 25 °C.

Regardless of storage temperature, oleic (ω-9, C18:1n9c), linoleic (ω-6, C18:3n6), γ-linoleic (C18:3n6), α-linolenic (ω-3, C18:3n6) and eicosenoic (C20:1n9) fatty acids are degraded during storage. However, the degradation of fatty acids increased with the increase of temperature; and some, such as α-linolenic, started to degrade early (30 days) when stored at 45 ° C (Tukey's test, $p < 0.05$). After 180 days of storage, oleic acid reduction was 8, 16 and 80%; linoleic acid reduction was 4, 8 and 68%, γ-linoleic reduction was 31, 50 and 88%; α-linolenic reduction was 1, 7 and 69%; and eicosenoic acid reduction was 7, 29 and 93%, from initial concentration (Day 0), at 25, 35 and 45 °C, respectively. Imran et al. (2016) reported a decrease for linoleic acid (ω-6) of 28.9 – 32.2% and for α-linolenic acid (ω-3) of 8.1 – 13.7%, for raw chia oil stored at 25 °C for 60 days. Other author also reported decrease in polyunsaturated fatty acid composition of oils as effect of temperature (Choe and Min, 2006; Imran et al., 2015).

Variations (mainly degradation) in fatty acids during storage are associated with oxidation or hydrolyzation reactions and lipolytic activity, which leads to the formation of carbonyl compounds, glycerol and fatty acids (Imran et al., 2015; Wang et al., 2012). Free fatty acids act as pro-oxidants in edible oils, thereby accelerating oil degradation (Miyashita and Takagi, 1986). These degradation reactions are accentuated with the increase in temperature (Choe and Min, 2006). Besides, oils with a high content of polyunsaturated fatty acids (like chia oil) are more susceptible to degradation processes (Timilsena et al., 2017). Therefore, processing, storage and transport conditions of chia seeds demand attention to avoid sample quality degradation.

Table 1. Evolution on fatty acid composition of chia seeds during storage

Fatty acid	**Tem**	**Time (days)**				
(mg/g oil)	**(°C)**	**0**	**30**	**60**	**90**	**180**

C16:0	25	110±4^{Aa}	110 ±4^{Aa}	111±4^{Aa}	113±5^{Aa}	115±5^{Aa}
	35		111 ±5^{Aa}	113±5^{Aa}	118±2^{Aa}	120±7^{Aa}
	45		106 ±4^{Aa}	104±6^{Aa}	92±2^{Bb}	62±24^{Bb}
C18:0	25	42±8^{Aa}	42±8^{Aa}	43±8^{Aa}	45±8^{Aa}	48±8^{Aa}
	35		44±7^{Aa}	45±8^{Aa}	50±8^{Aa}	52±10^{Aa}
	45		38±7^{Aa}	34±11^{Aab}	28±12^{Aab}	17±6^{Bb}
C18:1n9c	25	102±2^{Aa}	101±1^{Aa}	100±1^{Aa}	98±2^{Aab}	94±2^{Ab}
	35		99±2^{Aa}	97±3^{Aa}	90±0^{Bb}	86±4^{Bb}
	45		86±4^{Bb}	84±6^{Bb}	60±3^{Cc}	20±3^{Cd}
C18:2n6c	25	304±3^{Aa}	302±2^{Aa}	302±2^{Aa}	298±2^{Aa}	293±2^{Ab}
	35		299±3^{Aa}	296±4^{Aa}	289±5^{Ba}	279±1^{Bb}
	45		286±4^{Bb}	273±8^{Bb}	123±6^{Cc}	97±3^{Cd}
C18:3n6	25	5.06±1.40^{Aa}	4.80±1.56^{Aa}	4.69±1.52^{Aa}	3.92±0.93^{Aa}	3.51±1.03^{Aa}
	35		4.56±1.71^{Aa}	4.29±1.31^{Aa}	3.33±0.58^{Aa}	2.51±1.32^{ABa}
	45		2.80±0.69^{Aa}	2.68±0.98^{Aa}	0.71±0.34^{Bb}	0.60±0.61^{Bb}
C18:3n3	25	911±9^{Aa}	909±9^{Aa}	907±9^{Aa}	904±7^{Aa}	899±6^{Aa}
	35		904±8^{Aab}	898±8^{Aab}	890±10^{Aab}	846±36^{Ab}
	45		878±2^{Bb}	834±21^{Bc}	551±23^{Bd}	284±24^{Be}
C20:0	25	3.15±0.34^{Aa}	3.17±0.33^{Aa}	3.20±0.32A^{Aa}	3.31±0.32^{Aa}	3.63±0.24^{Aa}
	35		3.33±0.31^{Aa}	3.73±0.46^{Bab}	3.89±0.01^{Bb}	3.93±0.06^{Ab}

	45		2.93 ± 0.15^{Aa}	1.92 ± 0.90^{Aab}	0.96 ± 0.99^{Cbc}	0.23 ± 0.15^{Bc}
C20:1n9	25	2.59 ± 0.4^{Aa}	2.53 ± 0.4^{Aa}	2.81 ± 0.56^{Aa}	2.61 ± 0.60^{Aa}	2.40 ± 0.57^{Aa}
	35		2.3 ± 0.44^{Aa}	2.56 ± 0.40^{Aa}	2.05 ± 0.47^{Aa}	1.83 ± 0.47^{Aa}
	45		2.0 ± 0.26^{Aa}	1.20 ± 0.26^{Bb}	0.77 ± 0.47^{Bbc}	0.18 ± 0.07^{Bc}

[a-e] Mean with different lower case letter in the same column indicates significant difference between days of storage for each measured fatty acid according to Tukey's test, $p < 0.05$.

[A, B] Mean with different upper case letter in the same column indicates significant difference between storage temperature (25, 35 and 45 °C) for each measured fatty acid according to Tukey's test, $p < 0.05$.

Acidity is related to its oxidative stability and, therefore, allows to infer the storage conditions of the product. Figure 1 shows the acidity (%) behavior of chia seeds as a function of storage temperature (25, 35 and 45 ° C) for 180 days. For any storage temperature, acidity values increase during storage time, although no statistically significant difference (Tukey's test, $p > 0.05$) was observed until after 30 days of storage, both for the effect of storage temperature and storage time. The increase in the content of free fatty acids indicates a hydrolytic degradation and autoxidative process of fatty acids (Choe and Min, 2006; Ixtaina et al., 2012), which is in agreement with fatty acid composition (Table 1). The acidity and oxidation level (amount of hydroperoxides) of chia seeds are important quality control parameters, as they have an influence on palatability, nutritional quality, and toxicity, as well as influences the process of extraction and production of edible oil (Choe and Min, 2006; Franklin et al., 2017; Mata et al., 2017). As previously mentioned, FFA act as pro-oxidants, so an increase in the amount of FFA would accelerate the degradation processes of fatty acids (Miyashita and Takagi, 1986), especially oleic, linoleic and α-linolenic polyunsaturated fatty acids, which are more susceptible to oxidation and are predominant in chia seeds (Timilsena et al., 2017). Caruso et al. (2018) reported that the acidity of chia seeds increased 83%

after 5 months of storage at 25 ° C, Imran et al. (2016) found an increase in acidity in chia oil of 107% after 60 days of storage at 25 °C and Ixtaina et al. (2012) found an increase in acidity in chia oil of 8% when stored at 20 ° C for 225 days.

The temperature had a significant effect on the acidity of chia seeds, as expected, since high temperatures degrade the antioxidant compounds naturally present in chia (e.g. tocopherols), accelerating the processes of autoxidation and degradation reactions.

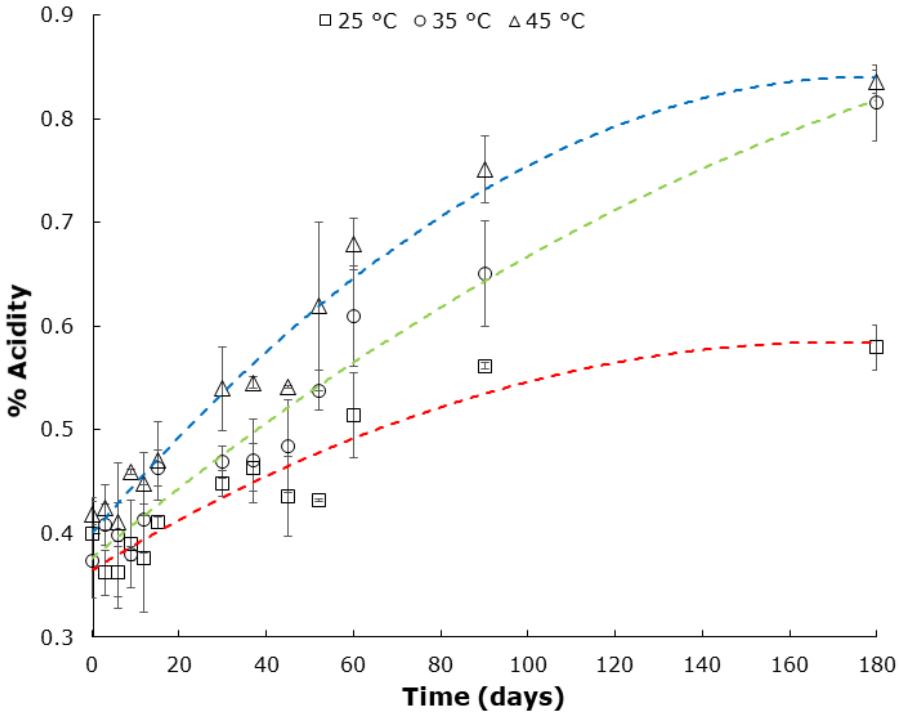


Fig. 1. Acidity (%) of Chia seeds stored at 25 °C, 35 °C and 45 °C over time.

3.2 Spectra profile

Fig. 2A shows the mean spectra based on storage temperature at day 180 and 2B demonstrates the changes in the mean spectra based on the days of storage at 45 °C. The absorption bands at 1200 - 1250 nm are related to C-H stretch second overtone and a great absorption band at 1400 – 1450 nm related to O-H stretch first overtone. The spectral region between 1600-1650 nm is related to the first overtone of C-H vibration,

influenced by the presence of CH_3 bonds, which may be associated with the presence of phenolic compounds (such as tocopherol, polyphenols, myricetin, quercetin, kaempferol, chlorogenic acid, etc.) and mucilage content (da Silva et al., 2017). The peaks at 1720 and 1760 nm are too associate to C-H stretch first overtone, and can be associated with the several lipid species found in chia oil. According to Hourant et al. (2000), the peak at 1720 nm is characteristic for oils that are rich in polyunsaturated fatty acids, which was expected for chia seeds. In this context, variations in absorbance at 1720 nm may be associated with various classes of oxygenated compounds, indicating oxidation and hydrolytic degradation of lipids in chia seeds as result of storage conditions (Murray, 1986). The region between 2300 – 2350 nm are associated with C-H combinations and deformation tones; and C=C and C–H stretch combination tones of *cis* unsaturated fatty acids were observed at 2144 nm (Mureșan et al., 2016). The absorption band at 2340 nm is related to CH combinations, and in the case of chia, could be associated to polysaccharides such as fiber and mucilage (Muñoz et al., 2012). The region at 2250 – 2260 nm is associated to protein content, which represent approximately 20% chia composition (Grancieri et al., 2019). Chia seed spectra show variations in absorbance values with temperature (Figure 2A) and with storage days (Figure 2B). Besides, changes with storage time were more noticeable at 45 ° C, which was expected since the seed offers protection to the seed from environmental effects.

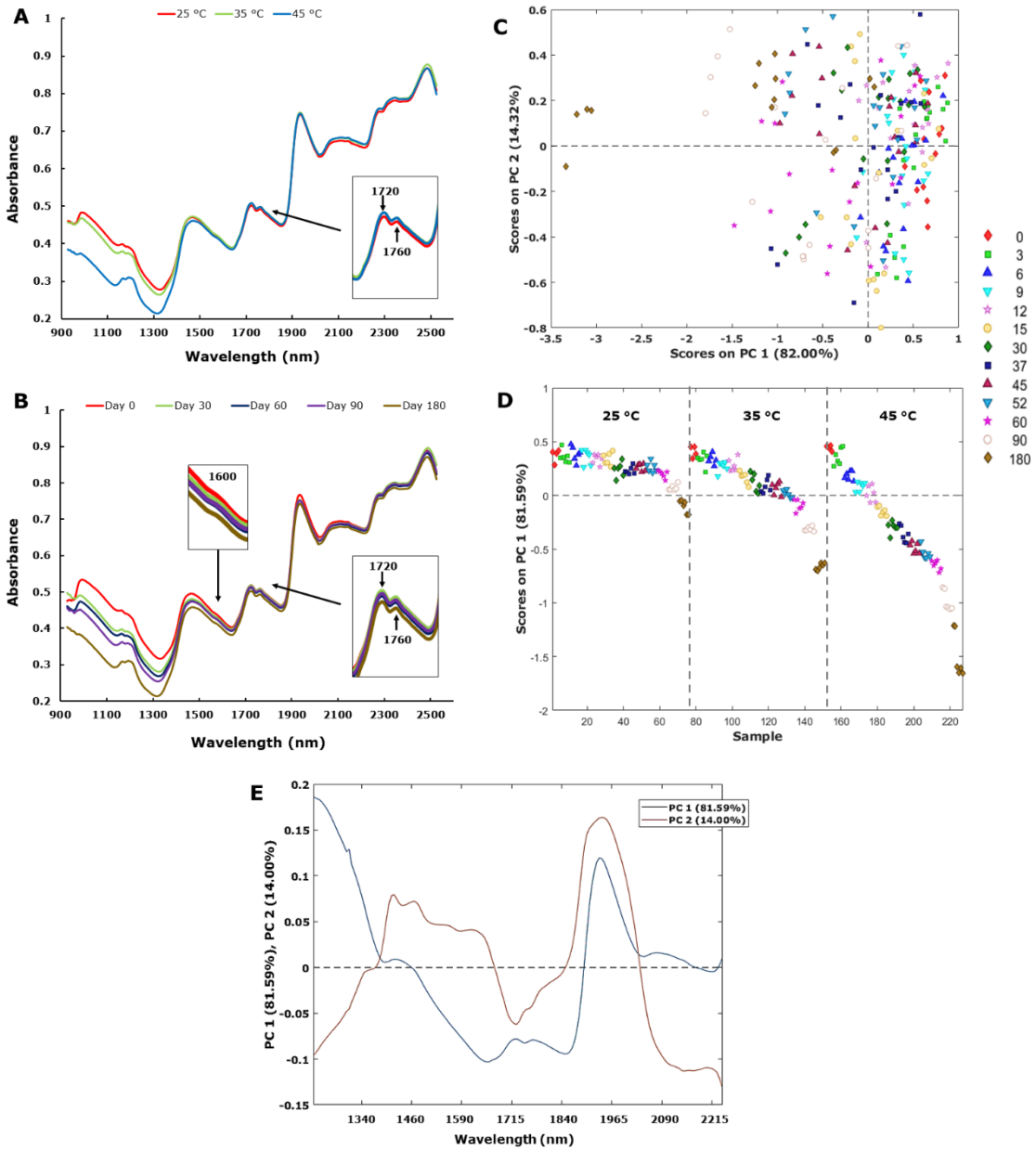


Fig. 2. A) Mean spectra based on temperature of storage 25 °C, 35 °C and 45 °C at day 180, B) Mean spectra based on days of storage 0, 30, 60, 90 and 180 at 45 °C, C) PC1/PC2 plot for days of storage in the wavelength range of 928–2524 nm; D) time-related PC1 scores at 25, 35 and 45 °C in the wavelength range of 1228–2238 nm; E) Loadings PC1 (red) PC2 (blue) in the wavelength range of 1228–2238 nm.

3.3 Time-related PC scores analysis

PCA of the pre-processed spectral data (SNV + mean center) in the wavelength range of 928–2524 nm was plotted in Figure 2C. PC1 (82%) and PC2 (14.32%) covered the maximum variance in the spectral data of chia seeds. The spectral region in 928 – 1211 nm and in 2244 – 2524 nm were removed because they did not contribute to the model from the preliminary PCA (Bro and Smilde, 2014). In the spectral range of 1218 – 2238 nm, the two first PCs explained a total variability of 95.59%. Also, in this reduced range (164 variables) PC1 explained a variability of 81.59% (Fig. 2D), almost invariable compared to the 82% explained using the full range (256 variables) (Fig. 2C).

It is notorious that the variability on PC1 scores (decreasing) were associated to storage time of chia seeds (Fig. 2C) and that this variability is greater for storage temperature of 45 °C compared to 35 °C or 25 °C (Fig. 2D). This would indicate that, despite the inherent protection of the seed structure, the temperature has a significant impact on the quality of chia seeds. This is important, since in many countries with tropical climates, poor storage management can lead to increases in the degradation rates of chia seeds (Delouche et al., 2016).

At 25 ° C, the variation in seed composition during storage time is minimal, especially in the fatty acid profile (Caruso et al., 2018; Imran et al., 2016). At 35 and 45 ° C, degradation processes are accelerated, and may be associated with changes in moisture and lipid degradation of chia seeds. In the first case, the decrease in moisture content is related to the integral entropy (degree of order/disorder) of chia seeds (Escalona-García et al., 2016). Temperature influences the seeds moisture loss, destabilizing the matrix and breaking the adsorbate (water) and adsorbent (food) bonds, which leads to a greater availability of free water for degradation reactions (Pérez-Alonso et al., 2006). PC loadings (Fig. 2E) clearly show the effect on the regions between 1400 - 1500 nm and 1950 – 1970 nm, related to O-H stretch first overtone, on the variability of the samples.

In the second case, the high content of polyunsaturated fatty acids in chia makes it much more susceptible to oxidation processes compared to other vegetable oils (Souza et al., 2017). The wavelengths at 1720 and 1760 nm shown in the loadings (Fig. 2E) are related, as previously mentioned, with various classes of oxygenated compounds, indicating oxidation and hydrolytic degradation of lipids in chia seeds (Murray, 1986). Imran et al. (2016), Guiotto et al. (2014) and Ixtaina et al. (2012) reported an increase in the oxidative stability parameters (peroxide value, free fatty acid and *p*-Anisidine value) during storage of chia oil and chia/sunflower blend oil stored at 4 - 25 °C by 60 - 360 days. Also, it should be considered that the variability of PC1 scores may be influenced by the degradation of phenolic compounds during storage (Mannucci et al., 2019), as can be seen in Fig. 2B and in the loadings graph (Fig. 2E) in the surrounding region from 1600-1650 nm. This is relevant, since a lower presence of antioxidant compounds (e.g. tocopherol) facilitate lipid degradation processes (Ixtaina et al., 2012).

Figure 3 shows the variability of PC1 scores in chia seeds stored at 25, 35 and 45 ° C in spectral range of 1228–2238 nm. In Figure 3A, the raw scores of PC1 applied on the hyperspectral images of chia seeds are observed, while in Fig. 3B a cut-off value (threshold) equal to mean scores was applied to hyperspectral images. Both figures show the color changes associated with the loss of quality or degradation of chia seeds, although these changes were more pronounced at 35 and 45 ° C. This is in accordance with Fig. 2D. In Figure 3A, it is possible to observe that the first 30 days of storage do not show variation in the scores for the samples stored at 25 ° C, which could be translated as an induction period, that is, from this period, degradation processes begin at this temperature (reduction of yellow and red color). This is expected, since the seed offers protection to the components of the seed. This induction period is shorter in samples stored at 35 ° C (~ 9 days) and was not observed at 45 ° C (also visible in Fig.

3B). Besides, it is possible to observe that at 45 ° C the greatest variability in the scores is observed, especially at the end of storage (180 days), where the degradation reaches its maximum point (Fig. 3B) and it is even possible to observe the morphology of the chia seeds (Fig. 3A). In effect, Fig. 3 shows how the degradation reactions that occur in chia seeds are dependent on time and temperature of storage. During storage of seeds, there are reports of changes in proteins, which are decomposed into amine macromolecules (Xu et al., 2018). Also, loss in phenolic compounds such as caffeic acid (chia seeds stored at 25 °C/10 months) (Caruso et al., 2018), tocopherols, carotenoids and chlorophyll (Mannucci et al., 2019), which are consumed to counteracting the oxidative reactions that can be generated over 6 months of storage. On the other hand, the oxidation and hydrolyzation processes of lipids in seeds or grains contribute to the formation of carbonyl compounds, glycerol and free fatty acids (Wang et al., 2012), which decreases the product quality, as observed in chia seeds (Caruso et al., 2018) and penaut and linseeds (Cämmerer and Kroh, 2009). Moreover, during the storage time there are variations in the composition of polysaccharides (Imran et al., 2015; Xu et al., 2018), which in the case of chia could be related to the degradation of the mucilage. Finally, chia seeds lose moisture during storage, which is caused by transpiration process and by the disposal of water for biological processes (degradation/formation of compounds). Water activity affects the mobility and reactivity of chemical species, indicating that greater availability of free water allows greater dissolution, mobility and reaction of pro-oxidant (Escalona-García et al., 2016). In chia seed (% humidity <10), the greater amount of water is linked to other structural components. Therefore, when the temperature increases, the bonds break and a greater amount of free water is arranged for the transformation reactions in the seed (Pérez-

Alonso et al., 2006). This can be seen in Figure 3, where seeds stored at 35 and 45 ° C show greater variability in PC1 scores.

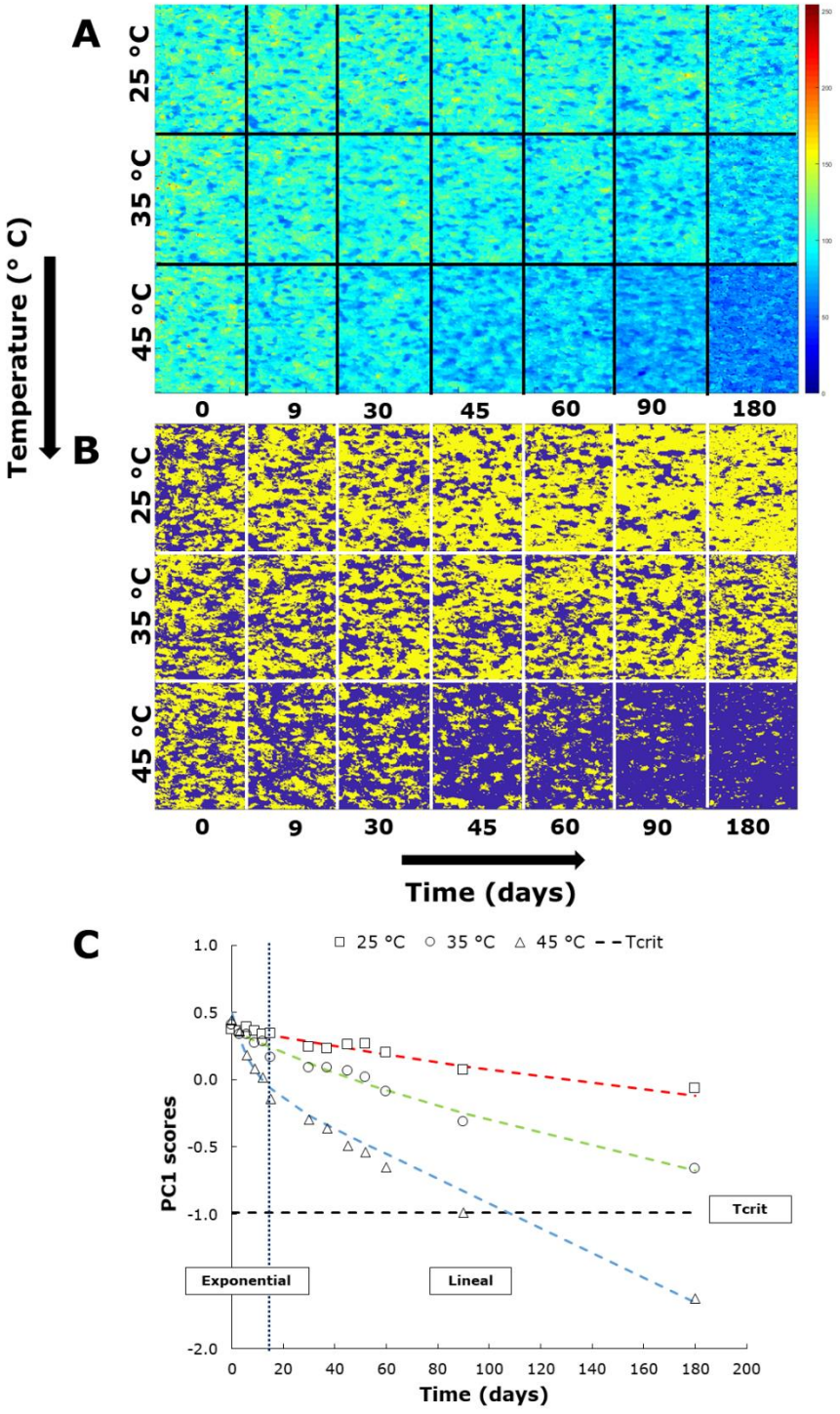


Fig. 3. (A) Variability in time-related PC1 scores at 25 °C, 35 °C and 45 °C, B) Variability in time-related PC1 scores at 25 °C, 35 °C and 45 °C using a cut-off value =

mean scores. Both figures were constructed in a wavelength range of 1228–2238 nm, C) Kinetic charts of the PC1 scores of Chia seeds stored at 25, 35, 45 °C. Black dotted-line represent the shelf –life cut-off value (-0.9853).

3.4 Multivariate modeling and shelf life estimation

The PC1 scores values over time were fitted using the non-linear regression by applying the kinetic model composed by two terms: exponential and lineal (Table 2). The fitting to this model showed better results, based on R^2 and standard error, compared to the first-order and second-order models (data not shown). Initially, changes in PC1 scores presented an exponential behavior (Fig. 3C), which may be associated with the loss of water and/or volatile compounds due to temperature. Then, a greater availability of free water (breakdown of water-food bonds) allows degradation reactions to begin (Pérez-Alonso et al., 2006). These degradation processes (such as the formation of hydrogenated compounds) are constant and explained by the linear term of the kinetic equation. The lowest value of R^2 (0.853) was for samples stored at 25 ° C, probably because at this temperature the changes are minimal, as seen in Fig. 3. A_0 values were different for samples stored at 25 (0.384), 35 (0.370) and 45 ° C (0.496). This could indicate that the degradation process starts earlier for temperatures of 35 and 45 ° C, being more noticeable at 45 ° C. Multivariate rate constant *k* associated to exponential term increases with increasing temperature, being 0.0058 (d^{-1}) at 25 °C, 0.0146 (d^{-1}) at 35 °C and 0.1294 (d^{-1}) at 45 °C. Instead, multivariate rate constant *c* associated to lineal term decreases with increasing temperature, being -0.0014 (d^{-1}) at 25 °C, -0.0039 (d^{-1}) at 35 °C and -0.0092 (d^{-1}) at 45 °C. The variation in the multivariate rate constant is greater for the exponential term than for the linear term. This could be caused by the rapid loss of water in the beginning of the storage process. Also, probably at the beginning of the storage process, phenolic compounds (e.g. tocopherols) may have been

rapidly consumed to avoid lipid oxidation processes. Together, these changes are expressly visible in Fig. 3, especially for 45 °C. Therefore, the acceleration factor (mean value of exponential and linear terms) was expected to be greater for 45 °C ($\alpha_{45,25} = 2.6$) compared to 35 °C ($\alpha_{35,25} = 14.4$). The estimated values for *k* and *c* by means of proposed kinetic model of PC1 scores are results of the input variables (spectra), which are related to the variation in the overall quality of chia seeds stored at different temperatures (Derossi et al., 2016). Both *k* and *c* are related to variations in moisture content, lipid self-oxidation and degradation of polysaccharides such as mucilage, reserve proteins and phenolic compounds and biological transformation process (germination) (Guiotto et al., 2014; Souza et al., 2017). Hence, this makes comparison with other oilseeds difficult. To our knowledge, there are no works that evaluate the shelf-life of chia seeds using the proposed kinetic model. However, various works reported association of kinetic parameters to the shelf-life of chia oil (Escalona-García et al., 2016; Guimarães-Inácio et al., 2018; Ixtaina et al., 2012).

Table 2. Kinetic parameters of PC1 scores as a function of time and estimated parameters of the Arrhenius model for chia seeds samples stored at 25, 35 and 45 °C.

Kinetic model: PC1 scores=A_0*exp(-k*t) + c*t

Temp (°C)	**A_0 (dimensionless)**	**Conf. Intervals**	**k (1/d)**	**Conf. Intervals**	**c (1/d)**	**Conf. Intervals**	**SSE**	**R^2**	**αT,25**
25	0.384	[0.360;0.408]	0.0058	[0.0018; 0.0098]	-0.0014	[-0.0020;-0.0007]	0.228	0.853	…
35	0.370	[0.341;0.	0.0146	[0.0102;	-0.0039	[-0.0042;-	0.28	0.955	2.6

		399]		0.0189]		0.0035]		
45	0.496	[0.397;0.595]	0.1294	[0.0854; 0.1734]	-0.0092	[-0.0096;-0.0088]	0.953	0.954	14.4

Arrhenius model, $k = C*exp(-E_a/RT)$

C	**Activation energy (E_a, kJ.mol^{-1})**	**R^2**
43.79	121.9	0.939

The beginning of the transformation processes within the seed begin with the breaking of the water-food bonds (Pérez-Alonso et al., 2006). Therefore, the values of *k* of the exponential term were used to calculate the activation energy of Arrhenius (R^2 = 0.939, Table 2). The activation energy for chia seeds (E_a = 128.9 kJ.mol^{-1}) was higher than for chia oil (E_a = 44.4 – 73.5 kJ.mol^{-1}) (Escalona-García et al., 2016; Guimarães-Inácio et al., 2018; Ixtaina et al., 2012) and pure α-linolenic (E_a = 60 - 70 kJ.mol^{-1}) (Litwinienko, 2001), which is the main fatty acid present in chia seeds. This is in agreement with the previous results, since it is likely that there are other components, such as phenolic compounds or mucilage (Caruso et al., 2018; Imran et al., 2015), that begin to degrade before the oil in chia seeds. That is because the activation energy was calculated based on the *k* obtained from the kinetic modeling of the PC1 scores, which were obtained in NIR spectral data. Therefore, E_a is influenced by the degradation of all components within the structure of chia seeds, and is specific to this type of product.

Figure 3C shows the PC1 scores as a function of time for each storage temperature, which were used to calculate the shelf life. For this study, we propose to use the acidity (%) of the chia seeds as a cut-off criterion to estimate the shelf-life. This limit was

established as the sample that presents a 75% increase in acidity compared to the initial value, which corresponded to the chia seeds stored at 45 ° C for 90 days. Then, the spectrum and the loadings of the time-related PC of this sample was taken to determine the shelf-life of chia seeds according to Eq. 4, and the cut-off criteria was calculated to be -0.9853 (black dashed line in Fig. 3).

Finally, to calculate the shelf life of chia seeds stored at 25 ° C and 35 ° C, we project the cut-off time of the chia seeds stored at 45 ° C (90 days) on the PC1 scores, using the acceleration factor previously calculated (Table 2) (Pedro & Ferreira, 2009).

$$Shelf - life_{25\,^{\circ}C} = \alpha_{25,45} * time_{crit,45} = 14.4 * 90 = 1300\ days$$

$$Shelf - life_{35\,^{\circ}C} = \alpha_{35,45} * time_{crit,45} = 8.9 * 90 = 798\ days$$

The time of shelf-life of chia seeds decreases with storage at higher temperatures, which is not an unexpected result. Although acidity was used as a cutting factor, due to its industrial importance (Mata et al., 2017) and its negative relationship with consumer acceptability (Franklin et al., 2017), the estimated useful life using the PC1 scores obtained from the NIR-hyperspectral images represents the overall degradation of chia seeds, such as previously discussed (see Fig. 3), yielding a more realistic estimation. Moreover, by encompassing the variability of all components of chia seeds, various cut-off criteria (such as proteins or specific phenolic compounds) can be established to estimate the shelf life based on the needs or uses of the chia seeds.

3.5 Validation

The re-sampling-based validation methodology was applied to validate the category of the chia samples (day *x* temperature). Thus, it is possible to assume that a new sample can be predicted reliably. The validation was performed using all the spectral

information of the NIR-HSI images. However, to facilitate the visualization, Figure 4 shows the validation obtained by using the mean spectrum of each sample.

Fig. 4A indicates the frequency with which the samples have been validated. The frequency values are between 20-40 for all samples, indicating a satisfactory result of the proposed number of iterations. The low values in the standard deviation (Fig. 4B) indicate that the validation has been satisfactory. That is, all the validated samples were correctly assigned in the range of scores corresponding to the day (0 - 180 days) and temperature (25, 35 and 45 ° C) of the chia seeds samples. More clearly, Fig. 4C shows how the validation samples (red circle) were correctly assigned within the corresponding calibration group. Fig. 4D shows the validation of the variability of the PC1 scores of chia seeds stored at 25, 35 and 45 ° C for 180 days. Pixel-to-pixel validation shows how chia seeds degrade as a function of time and temperature. The results are consistent with those shown in the calibration, where samples at 45 ° C have a high degradation rate (Fig. 3B). Therefore, the validation confirms that PC1 scores vary in relation to time and temperature, and that they reflect the chemical changes occurred by the degradation/transformation processes. Fig. 4E shows the standard deviation of the NIR-HSI. Here we can see that all the pixels within the validated image were corresponding to the pixels assigned for the PCA model calibration. Therefore, this would indicate that the PC1 scores are useful for estimating the shelf life, since all seeds had the same variability in the scores in both calibration and validation. Then, the validation of the methodology allows to make some statements: 1) PC1 scores are associated with chemical changes during storage time at different temperatures, 2) chia samples belonging to the same category (time *x* temperature) present the same variability in the PC1 scores, and 3) chia seeds within the same sample have the same

degradation rate. Thus, it is possible to assume that PCA scores can be used reliably to estimate the shelf life of chia seeds.

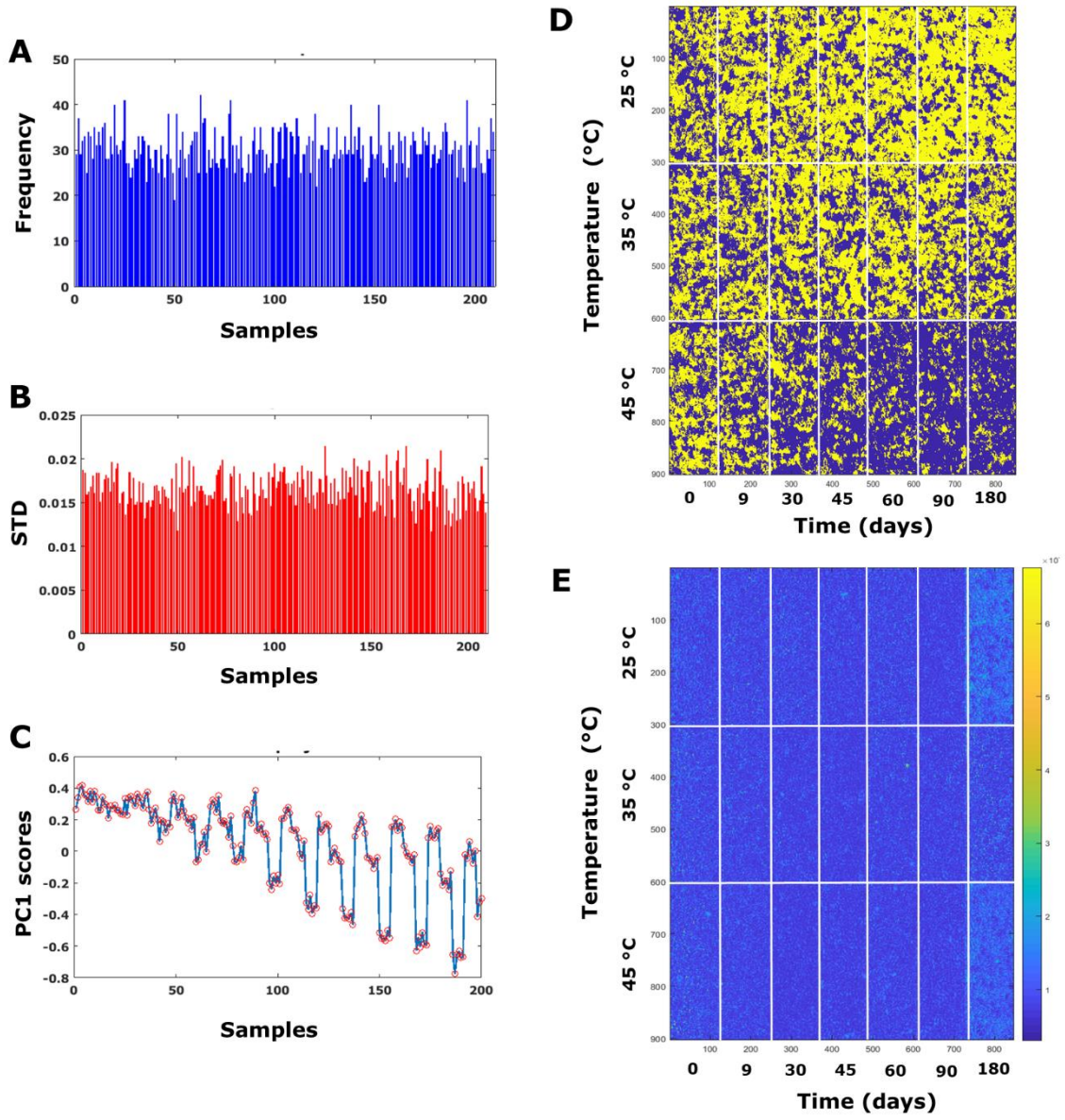


Fig. 4. Validation of time-related PC scores method for estimate shelf life of Chia seeds. A) Frequency, B) Standard deviation, C) Validation score projection, D) Variability in time-related PC1 validated scores at 25 °C, 35 °C and 45 °C using a cut-off value = mean scores, and E) Standard deviation in time-related PC1 validated scores at 25 °C, 35 °C and 45 °C.

4. Conclusions

NIR-HSI was used to explain the variability in the composition of chia seeds subjected to accelerated tests. The time-related PCs scores were used to model degradation kinetics and to estimate the multivariate accelerated shelf-life (MASLT) of chia seeds. Variability in PC1 is associated with the global degradation of chia seeds, which includes changes in proteins, lipids, carbohydrates and phenolic compounds, so their use to estimate the shelf life is appropriate. PC1 scores over time were adjusted to a fused kinetic model, with an $R^2 > 0.85$, which indicates a degradation of the chia seed components in two steps: exponential and linear. The increase in temperature accelerated the processes of seed degradation, as shown by the composition of fatty acids and acidity. Using the spectral information of chia seed samples with an increase of 75% of their initial value in their acidity, the shelf-life of chia seeds was estimated, being 1300, 798 and 90 days for chia seeds stored at 25, 35 and 45 °C, respectively. A new approach for the validation of the methodology of estimation of shelf-life using PCs scores was proposed. The results show that all the samples were correctly predicted within the same category (day *x* temperature). Therefore, for the first time, it is possible to state that the method developed can estimate the shelf-life of chia seeds reliably. Future work could endeavor to estimate the useful life of chia seeds or others, depending on other characteristics that are associated with the final use of the product, such as a specific protein or phenolic compound, because the spectral information collected encompasses the entire composition of the sample.

Funding

This work was supported by São Paulo Research Foundation (FAPESP), project n° 2015/24351/2, and Brazilian National Council for Scientific and Technological Development (CNPq) project n° 404852/2016-5). This study was financed in part by the

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

Acknowledgements

Marciano M. Oliveira acknowledges scholarship funding from CAPES. J. P. Cruz-Tirado acknowledges scholarship funding from FAPESP, grant n° 2018/02500/4. The authors would like to thank Marcia Ferreira for her explanation of the MASLT methodology.

Conflict of interest

The authors declare no competing financial interest.

Associated content

Supporting information

Supporting information for this study includes the expanded methodology of extraction, methylation and quantification of fatty acids from chia seeds.

References

AOCS, 1998. Method Cd 1-25: Iodine Value of Fats and Oils. Wijs Method.

Ballus, C.A., Meinhart, A.D., de Souza Campos Jr, F.A., da Silva, L.F. de O., de Oliveira, A.F., Godoy, H.T., 2014. A quantitative study on the phenolic compound, tocopherol and fatty acid contents of monovarietal virgin olive oils produced in the southeast region of Brazil. Food Res. Int. 62, 74–83.

Bro, R., Smilde, A.K., 2014. Principal component analysis. Anal. Methods 6, 2812–2831.

Cämmerer, B., Kroh, L.W., 2009. Shelf life of linseeds and peanuts in relation to

roasting. LWT - Food Sci. Technol. 42, 545–549.

Caruso, M.C., Favati, F., Di Cairano, M., Galgano, F., Labella, R., Scarpa, T., Condelli, N., 2018. Shelf-life evaluation and nutraceutical properties of chia seeds from a recent long-day flowering genotype cultivated in Mediterranean area. LWT 87, 400–405.

Chaudhry, M.M.A., Amodio, M.L., Babellahi, F., de Chiara, M.L. V, Amigo Rubio, J.M., Colelli, G., 2018. Hyperspectral imaging and multivariate accelerated shelf life testing (MASLT) approach for determining shelf life of rocket leaves. J. Food Eng. 238, 122–133.

Choe, E., Min, D.B., 2006. Mechanisms and Factors for Edible Oil Oxidation. Compr. Rev. Food Sci. Food Saf. 5, 169–186.

da Silva, B.P., Anunciação, P.C., Matyelka, J.C. da S., Della Lucia, C.M., Martino, H.S.D., Pinheiro-Sant'Ana, H.M., 2017. Chemical composition of Brazilian chia seeds grown in different places. Food Chem. 221, 1709–1716.

de Falco, B., Amato, M., Lanzotti, V., 2017. Chia seeds products: an overview. Phytochem. Rev. 16, 745–760.

Delouche, J.C., Matthes, R.K., Dougherty, G.M., Boyd, A.H., 2016. Storage of seed in sub-tropical and tropical regions.

Derossi, A., Mastrandrea, L., Amodio, M.L., de Chiara, M.L. V, Colelli, G., 2016. Application of multivariate accelerated test for the shelf life estimation of fresh-cut lettuce. J. Food Eng. 169, 122–130.

EC, 1991. Commission Regulation (EEC) 2568/91 of July 11th 1991 on characteristics of olive oil and on the relevant methods of analysis.

Escalona-García, L.A., Pedroza-Islas, R., Natividad, R., Rodríguez-Huezo, M.E., Carrillo-Navas, H., Pérez-Alonso, C., 2016. Oxidation kinetics and thermodynamic analysis of chia oil microencapsulated in a whey protein concentrate-polysaccharide matrix. J. Food Eng. 175, 93–103.

Franklin, L.M., Chapman, D.M., King, E.S., Mau, M., Huang, G., Mitchell, A.E., 2017. Chemical and Sensory Characterization of Oxidative Changes in Roasted Almonds Undergoing Accelerated Shelf Life. J. Agric. Food Chem. 65, 2549–2563.

Grancieri, M., Martino, H.S.D., Gonzalez de Mejia, E., 2019. Chia Seed (Salvia hispanica L.) as a Source of Proteins and Bioactive Peptides with Health Benefits: A Review. Compr. Rev. Food Sci. Food Saf. 18, 480–499.

Guimarães-Inácio, A., Francisco, C.R.L., Rojas, V.M., Leone, R. de S., Valderrama, P., Bona, E., Leimann, F.V., Tanamati, A.A.C., Gonçalves, O.H., 2018. Evaluation of the oxidative stability of chia oil-loaded microparticles by thermal, spectroscopic and chemometric methods. LWT 87, 498–506.

Guiotto, E.N., Ixtaina, V.Y., Nolasco, S.M., Tomás, M.C., 2014. Effect of Storage Conditions and Antioxidants on the Oxidative Stability of Sunflower–Chia Oil Blends. J. Am. Oil Chem. Soc. 91, 767–776. https://doi.org/10.1007/s11746-014-2410-9

Hartman, L., Lago, R.C., 1973. Rapid preparation of fatty acid methyl esters from lipids. Lab. Pract. 22, 475–476.

Hourant, P., Baeten, V., Morales, M.T., Meurens, M., Aparicio, R., 2000. Oil and Fat Classification by Selected Bands of Near-Infrared Spectroscopy. Appl. Spectrosc. 54, 1168–1174.

Imran, M., Anjum, F.M., Ahmad, N., Khan, M.K., Mushtaq, Z., Nadeem, M., Hussain, S., 2015. Impact of extrusion processing conditions on lipid peroxidation and storage stability of full-fat flaxseed meal. Lipids Health Dis. 14, 92.

Imran, M., Nadeem, M., Manzoor, M.F., Javed, A., Ali, Z., Akhtar, M.N., Ali, M., Hussain, Y., 2016. Fatty acids characterization, oxidative perspectives and consumer acceptability of oil extracted from pre-treated chia (Salvia hispanica L.) seeds. Lipids Health Dis. 15, 162.

Ixtaina, V.Y., Nolasco, S.M., Tomás, M.C., 2012. Oxidative Stability of Chia (Salvia hispanica L.) Seed Oil: Effect of Antioxidants and Storage Conditions. J. Am. Oil Chem. Soc. 89, 1077–1090.

Joseph, J.D., 1992. Capillary column gas chromatographic method for analysis of encapsulated fish oils and fish oil ethyl esters: collaborative study. J. AOAC Int. 75, 487–506.

Labuza, T.P., 1982. Shelf-life dating of foods. Food & Nutrition Press, Inc.

Litwinienko, G., 2001. Autooxidation of unsaturated fatty acids and their esters. J. Therm. Anal. Calorim. 65, 639–646.

Mannucci, A., Castagna, A., Santin, M., Serra, A., Mele, M., Ranieri, A., 2019. Quality of flaxseed oil cake under different storage conditions. LWT 104, 84–90.

Mata, T.M., Correia, D., Pinto, A., Andrade, S., Trovisco, I., Matos, E., Martins, A.A., Caetano, N.S., 2017. Fish oil acidity reduction by enzymatic esterification. Energy Procedia 136, 474–480.

Miyashita, K., Takagi, T., 1986. Study on the oxidative rate and prooxidant activity of free fatty acids. J. Am. Oil Chem. Soc. 63, 1380–1384.

Muñoz, L.A., Cobos, A., Diaz, O., Aguilera, J.M., 2012. Chia seeds: Microstructure, mucilage extraction and hydration. J. Food Eng. 108, 216–224.

Mureșan, V., Danthine, S., Mureșan, A.E., Racolța, E., Blecker, C., Muste, S., Socaciu, C., Baeten, V., 2016. In situ analysis of lipid oxidation in oilseed-based food products using near-infrared spectroscopy and chemometrics: The sunflower kernel paste (tahini) example. Talanta 155, 336–346.

Murray, I., 1986. The NIR spectra of homologous series of organic compounds, in: Proceedings of the International NIR/NIT Conference. Akademiai Kiado: Budapest, Hungary, pp. 13–28.

Oliveira-Alves, S.C., Vendramini-Costa, D.B., Betim Cazarin, C.B., Maróstica Júnior, M.R., Borges Ferreira, J.P., Silva, A.B., Prado, M.A., Bronze, M.R., 2017. Characterization of phenolic compounds in chia (Salvia hispanica L.) seeds, fiber flour and oil. Food Chem. 232, 295–305.

Pedro, A.M.K., Ferreira, M.M.C., 2009. The Use of Near-Infrared Spectroscopy and Chemometrics for Determining the Shelf-Life of Products. Appl. Spectrosc. 63, 1308–1314.

Pedro, A.M.K., Ferreira, M.M.C., 2006. Multivariate accelerated shelf-life testing: a novel approach for determining the shelf-life of foods. J. Chemom. 20, 76–83.

Pérez-Alonso, C., Beristain, C.I., Lobato-Calleros, C., Rodríguez-Huezo, M.E., Vernon-Carter, E.J., 2006. Thermodynamic analysis of the sorption isotherms of pure and blended carbohydrate polymers. J. Food Eng. 77, 753–760.

Souza, A.L., Martínez, F.P., Ferreira, S.B., Kaiser, C.R., 2017. A complete evaluation of thermal and oxidative stability of chia oil. J. Therm. Anal. Calorim. 130, 1307–

1315.

Timilsena, Y.P., Vongsvivut, J., Adhikari, R., Adhikari, B., 2017. Physicochemical and thermal characteristics of Australian chia seed oil. Food Chem. 228, 394–402.

Upadhyay, R., Mishra, H.N., 2015. Multivariate Analysis for Kinetic Modeling of Oxidative Stability and Shelf Life Estimation of Sunflower Oil Blended with Sage (Salvia officinalis) Extract Under Rancimat Conditions. Food Bioprocess Technol. 8, 801–810.

Wang, F., Wang, R., Jing, W., Zhang, W., 2012. Quantitative dissection of lipid degradation in rice seeds during accelerated aging. Plant Growth Regul. 66, 49–58.

Xu, M., He, D., Teng, H., Chen, L., Song, H., Huang, Q., 2018. Physiological and proteomic analyses of coix seed aging during storage. Food Chem. 260, 82–89.

Supporting Information

Shelf life estimation and kinetic degradation modeling of Chia seeds (Salvia hispanica) using Principal Component Analysis based on NIR-Hyperspectral imaging

1. Experimental

1.1 Fatty acid composition and free fatty acid

Chia seeds were ground using a mill model A 11 B S32 (IKA, Germany). Later, Chia oil was extracted by Bligh-Dyer method (Hartman & Lago, 1973). This method allows extracting lipids from Chia seeds without applying heat, so it can be used to assess oil deterioration as a result of storage conditions. Summarized, 3 g of ground sample was weighed into test tubes with chloroform (8 mL), methanol (16 mL) and water (6.4 mL). The tubes were agitated for 30 minutes and then an additional 8 mL chloroform and 8 mL anhydrous sodium sulfate (1.5%) were added, with phase separation occurring. The obtained lipid extract was dried at 40 ° C on a rotary evaporator model 801 (Fisatom, Brazil) until solvent drying. The oil was stored at -86 °C until the analysis.

For fatty acid composition measurement, the lipids obtained were esterified as reported by Joseph and Ackman (1992). Initially, 200 µL of C23:0 was added to tubes as an internal standard at a concentration of 5 mg/mL and then dried with N_2 flow. The extracted fat (25 mg) together with 4 mL of sodium hydroxide (0.5 M in methanol) and C23:0 were heated at 100 °C for 20 min. Then the extract was derivatized with 3 mL of boron trifluoride solution (12% in methanol) at 100 °C for 5 min, and then 4 mL of saturated sodium chloride solution and 4 mL of hexane was added. The tubes were rested until the phases separated. Three consecutive partition steps were performed with hexane (2 mL) under stirring. The collected phases were mixed and concentrated to dryness on a rotary evaporator and resuspended in 2 mL hexane. Extraction was performed in triplicate (n = 3) and the extract was stored in at -86 °C until the analysis.

The chromatographic conditions were based on Ballus et al. (2014), with modifications. Separation of methyl esters was performed on a 7890A gas chromatograph (GC-Agilent, Germany) equipped with a flame ionization detector (FID). The methyl esters were separated using a DB 23 capillary column (60 m, 0.25 mm d.i., 0.25 µm film thick, Agilent, USA). An aliquot of the extract (1 µL) was injected at a ratio of 1:50. Injector and detector temperatures were maintained at 250 ° C and 280 ° C, respectively. The oven temperature ramp was 50 °C (5 min, hold time) increasing to 175 °C to 25 °C/min, finally to 230 ° C to 4 °C/min and maintained for 25 min. The flow rate of carrier gas (N_2) was 1 $mL.min^{-1}$ and detector gas flow rate (N_2:H_2:synthetic air) were 30 $mL.min^{-1}$, 30 $mL.min^{-1}$ and 300 $mL.min^{-1}$, respectively. Methyl esters were identified by comparing their retention times with those obtained with the standards (FAME mix C4-C24) under the same chromatographic conditions. Quantification was performed by internal standardization, using C23:0 as internal standard. Correction factors and fatty acid concentration (mg/g oil) were calculated according to Joseph and Ackman (1992).

Free fatty acid in stored Chia seeds were determined according to Ca 5a-40 (AOCS, 1998). All samples were analyzed in triplicate.

CHAPTER 4:

Authentication of cocoa (*Theobroma cacao*) bean hybrids by NIR-hyperspectral imaging and chemometrics

The results of this chapter is published in *Food Control.* DOI:

10.1016/j.foodcont.2020.107445

Authentication of cocoa (*Theobroma cacao*) bean hybrids by NIR-hyperspectral imaging and chemometrics

J.P. Cruz-Tirado[a]; Juan Antonio Fernández Pierna[b]; Hervé Rogez[c]; Douglas Barbin[a,]; Vincent Baeten[b]*

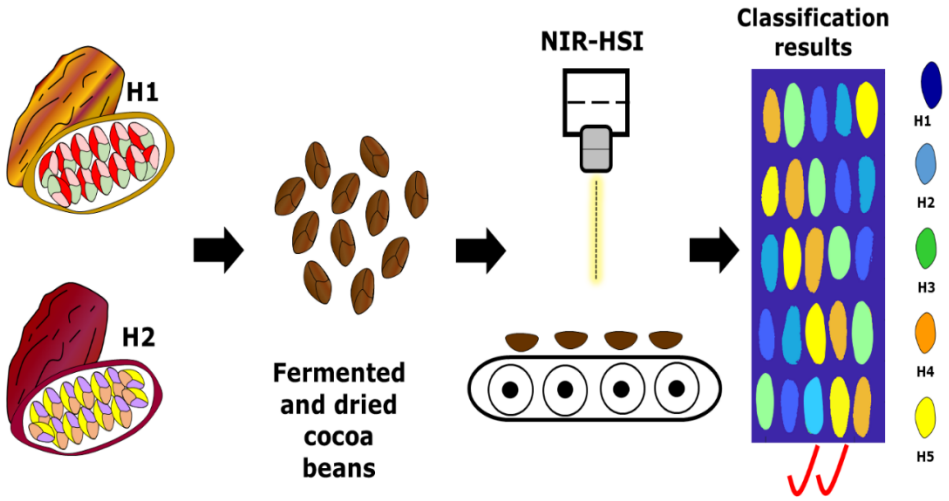
[a]*Department of Food Engineering, School of Food Engineering, University of Campinas, Campinas, SP, Brazil*

[b]*Walloon Agricultural Research Centre (CRA-W), Knowledge and valorization of agricultural products Department, Quality and authentication of products Unit, Gembloux, Belgium*

[c]*CVACBA (Center for Valorization of Amazonian Bioactive Compounds), UFPA (Federal University of Pará), Avenida Perimetral, 01, 66075-150, Guamá, Belém, Pará, Brazil*

**Corresponding author: dfbarbin@unicamp.br*

Graphical abstract



Abstract

The hybridization of cocoa allows generating new varieties with the aim of opening new horizons in terms of yielding, disease resistance and flavor. The objective of this work was the development and validation of classification models based on NIR hyperspectral imaging and chemometrics for the discrimination of five valuable cocoa bean hybrids. The chemometrics tools, PLS-DA and SVM, showed comparable results for 2-class (hybrids) models, but SVM (3.8 – 23.1 % prediction error) was superior to PLS-DA (4.4 – 34.4 % prediction error) when all five classes (hybrids) were included in a model. PLS-DA maps showed a simple and informative way to discriminate hybrids, allowing a correct classification between 50 – 100 %. Finally, it can be concluded that the models created in this work could be a good and reliably alternative to the actual visual method for the discrimination of cocoa bean hybrids.

Keywords: genotype; imaging; computer vision; near-infrared imaging

1. Introduction

According to International Trade Centre (2001), there is not a general rule to use both terms ‘cacao’ or ‘cocoa’ (*Theobroma cacao*) to refer to the bean. However, it is common to use the term ‘cacao’ to describe the scientific and horticultural aspects of the plant, reserving ‘cocoa’ for fermented and dried bean. The cocoa bean is one of the agricultural commodities highly demanded in the world for its high benefits in terms of nutrition and economics. Ivory Coast, Ghana, Indonesia, Nigeria, Cameroon and Brazil are the world’s largest cocoa producer (4.6 million tons of cocoa harvested in 2016) (The International Cocoa Organization, 2018).

Hybridization is a common technology that allowed to create cocoa bean hybrids with different features such as a greater disease resistance (e.g. “witches broom disease” caused by *Moniliophthora perniciosa*), however it also can affect pod and bean yield

parameters, precocity and butterfat flavor expressed after optimal fermentation. Traditionally, cocoa genotypes are grouped into Criollo, Forastero and Trinitario. Later, Motamayor et al. (2008) proposed a new sub-classification for Forastero group, including: Marañon (PA), Curaray (AGU), Iquitos (IMC), Nanay (NA), Contamana (SCA), Amelonado (BE), Purús (CAB), Nacional (MO) and Guiana (CJ). These genotypes were reported for South America, and from them, through hybridization, cocoa producers create hybrids with disease resistance, a greater yielding, and interesting flavor. Nevertheless, many times hybrids are planted together and the seeds are mixed, making it difficult to identify the purity of cocoa in relation to a variety of high economic value, in order to assure the quality of the desired final product, especially the chocolate.

Diverse analytical techniques such as multi-element and multi-compound isotope profiling (^{13}C, ^{15}N, % C, % N) (Diomande et al., 2015), proteomic and peptidomic fingerprinting by ultra-performance liquid chromatography tandem mass spectrometry method with electrospray ionization (UHPLC-ESI-MS/MS) (Kumari et al., 2018; Scollo, Neville, Oruna-Concha, Trotin, & Cramer, 2020), nanofluidic single nucleotide polymorphism (SNP) genotyping (Fang et al., 2014) or microsatellite markers (Dinarti et al., 2015; Herrmann et al., 2015) have been developed to identify cocoa bean hybrid. These techniques turn out to be accurate and reliable to identify cocoa bean hybrids, however, they consume a lot of time, use chemical reagents and destroy the samples. For the cocoa industry, it is essential to identify cocoa bean hybrids according to quality criteria without destroying the sample, quickly and reagent-free. Some non-destructive techniques such as Raman spectroscopy (Vargas Jentzsch et al., 2016) and computer vision (Jimenez et al., 2018; Mite-Baidal et al., 2019) showed a good performance to identify cocoa beans genotypes. However, in both cases the objective was to identify a

specific variety (CCN-51) and two-classes model was developed. New methods are still needed to quickly differentiate different types of hybrids, which can often be confused within the same batch.

Near infrared spectroscopy (NIRS) is the technology that can help to solve this problem. For the cocoa beans, NIRS has proven to be efficient in determining protein, fat, caffeine, theobromine, (−)-Epicatechin, carbohydrates and moisture content in cocoa flour (Álvarez et al., 2012; Barbin et al., 2018; Veselá et al., 2007), to detect adulterations by substitution (e.g. carob flour) in cocoa flour (Quelal-Vásconez et al., 2019; Quelal-Vásconez, Pérez-Esteve, Arnau-Bonachera, Barat, & Talens, 2018), geographical and varietal origin of the cocoa beans and shells of cocoa beans (Mandrile et al., 2019; Teye, Huang, Dai, & Chen, 2013; Trognitz et al., 2013). However, for NIR spectra acquisiton, cocoa bean samples were ground.

Near infrared hyperspectral imaging (NIR-HSI) is a NIR-based technology that allows obtaining spectral and spatial information simultaneously (Baeten, Pierna, Vermeulen, & Dardenne, 2010; Dale et al., 2013; Fernández Pierna, Baeten, & Dardenne, 2006). Some previous works showed the ability of NIR-HSI in tandem with multivariate analysis to identify hybrids of rice (X. Liu, Feng, Liu, & He, 2017), okra seeds (Nie, Zhang, Feng, Yu, & He, 2019), maize seeds (Guo, Zhu, Huang, Guo, & Qin, 2017), sweet potato (Su, Bakalis, & Sun, 2019) and soybean (Y. Liu, Wu, Yang, Tan, & Wang, 2019), among others. Regarding cocoa beans, Caporaso et al. (2018) showed that the spectral information obtained from NIR-HSI can be used to predict the fermentation index, total polyphenols and antioxidant activity in single peeled dried fermented cocoa beans.

According to Okiyama et al. (2017), one kg dried cocoa bean shell is composed approximately by 504 – 606 g fiber, 116 – 181 g protein, 47 – 101 g moisture, ~178 g

carbohydrates, 20 – 68 g fat and 18 – 58 g phenols. Previous work reported that the chemical information obtained from the cocoa bean shell can be used to identify genetic varieties of cocoa beans (Mandrile et al., 2019). In this work, NIR-HSI technology was chosen in order to extract, from images, the adequate spectral information from fermented and dried unpeeled cocoa beans, in order to allow the discrimination of cocoa beans hybrids, for quality control purposes (purity). Thus, the objective of this work was to identify and classify cocoa bean hybrids originating in the Brazilian Basin, using NIR-HSI in combination with chemometric tools as a non-destructive procedure.

2. Material and methods

2.1 Sample collection

Cocoa bean hybrids samples were collected by CEPLAC (Medicilândia, Para, Brazil) in August 2018. Samples were collected in the same geographical zone, to avoid the effects related to the soil composition and climatic factors. Five cocoa bean hybrids from *Forastero* genotype were selected considering their importance for Brazilian cacao industry, and they are summarized in Figure 1a.

All samples were fermented during 7 days, under same conditions of temperature and relative humidity. After fermentation, samples were dried during 4 days under sun shining. Then, samples were stored at 25 °C until analysis. Figure 1b shows the final internal appearance of cocoa bean hybrids, confirming the good compartmentation and typical brown colour of these beans. No additional processing such as grinding, shelling, among others of the cocoa bean hybrids was performed.

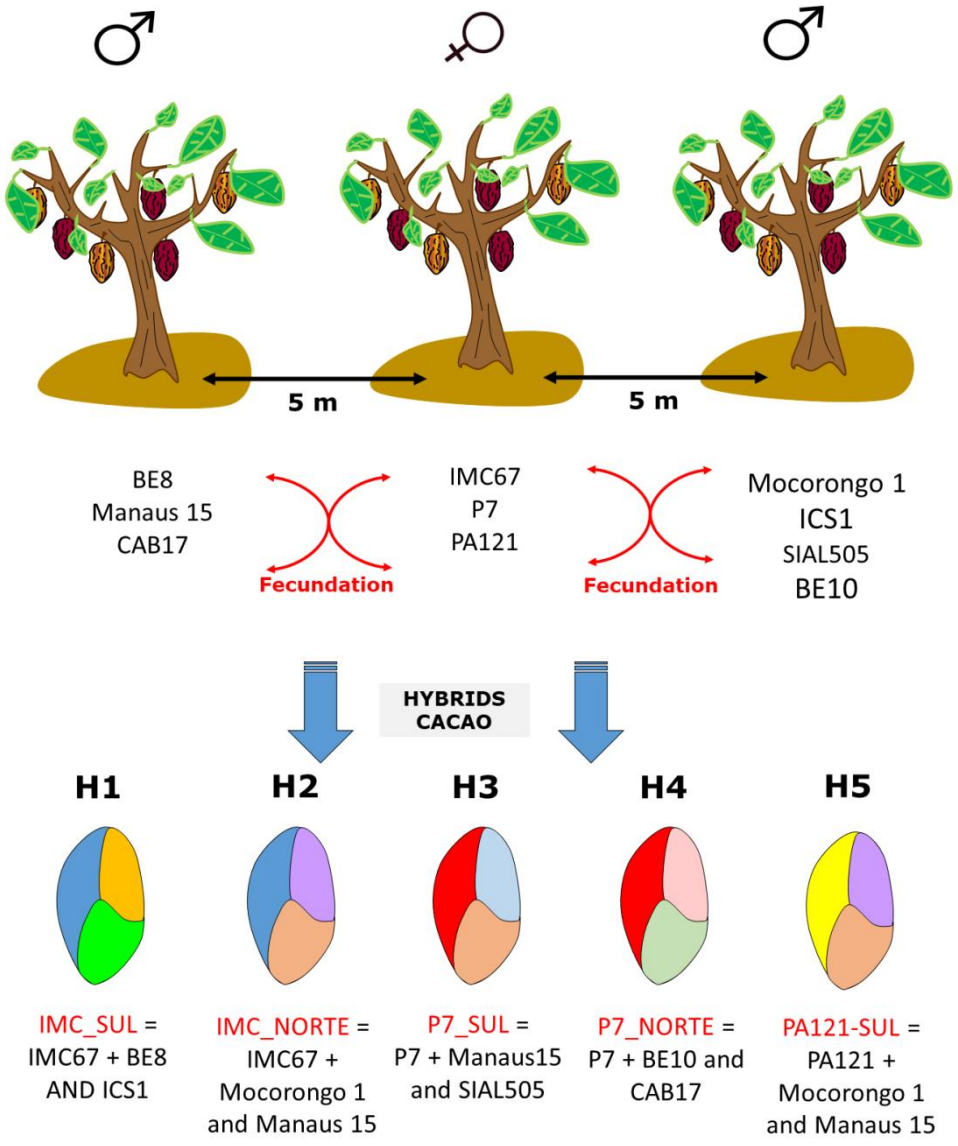


Fig. 1a. Scheme showing the process for production of cocoa bean hybrids in CEPLAC (Medicilândia, Para, Brazil).

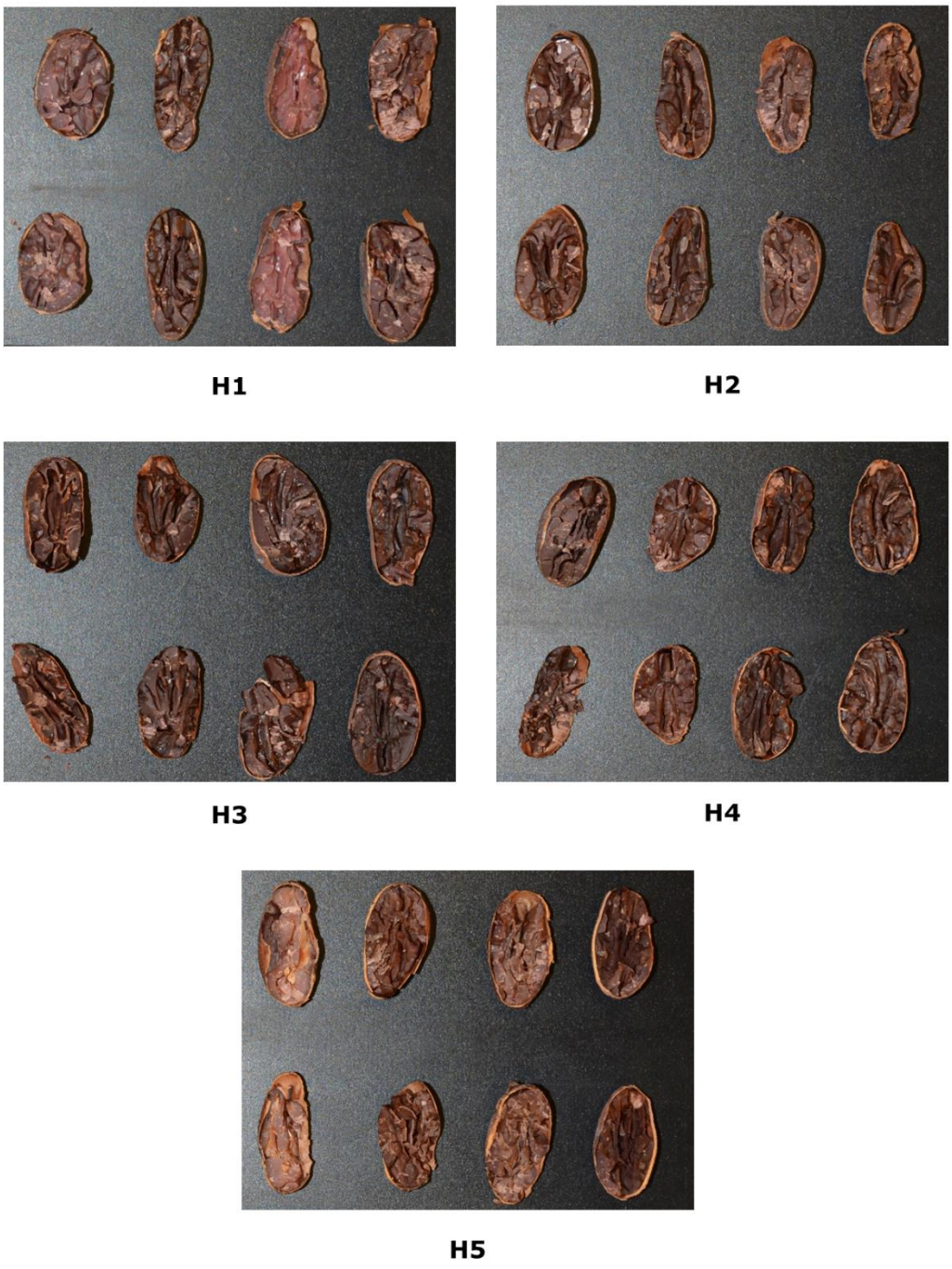


Fig. 1b. Photo of the visual aspect of the internal quality of cocoa bean hybrids.

2.2 Instrumentation

NIR hyperspectral images of cocoa bean hybrids were acquired using a line scan imaging system combined with a conveyor belt (BurgerMetrics SIA, Riga, Latvia) in a room temperature of 25 °C. This device consisted of a SWIR XEVA CL2.5 320 TE4

camera (Specim Ltd., Oulu, Finland) which has a resolution of 320-pixel lines and an ImSpector N25E spectrograph (Xenics nv, Leuven, Belgium) with a spectral range of 1100 – 2400 nm (209 wavelengths) and a spectral resolution of 6.3 nm. 32 scans per image have been averaged and each pixel provides an absorbance spectrum at each point of the image. Acquisition is performed using the HyperPro software (BurgerMetricsSIA, Riga, Latvia). More details of the used device and its components can be found in Eylenbosch et al. (Eylenbosch, Bodson, Baeten, & Fernández Pierna, 2018).

Two-dimensional references are needed for dark and white in order to perform instrumental calibration prior to further analysis. For this, a dark image is collected by blocking the access of light to the camera and a white image using a standard white reference board. The spectra are then corrected according to equation 1. Then, cocoa beans were placed on a conveyor belt (speed of 1.1 mm/s) in groups of 10 beans per image. Image acquisition was performed at room conditions.

$$I = \frac{(I_0 - B)}{(W - B)} * 100 \qquad \text{(Eq. 1)}$$

where I_0 is the original hyperspectral image; B is the dark image and W is the white image.

2.3 Spectral data collection

In total, 50 fermented and dried cocoa beans from each hybrid were chosen for the construction of the models. Additionally, a second set of 200 cocoa beans was used for pixel-to-pixel external validation. The number of cocoa bean hybrids chosen was sufficient to create robust models, since the experiment was controlled in (1) soil type (geographical origin), (2) same fermentation and drying process and (3) same harvest

date and (4) same pre-analysis storage time. The mean spectra (ROI) was extracted from whole cocoa beans for each hybrid (both sides). For this purpose, a mask was built to isolate the cocoa beans from the background using the difference in intensity. After ROI segmentation, spectral libraries for hybrids of cocoa beans were compiled by calculating the mean spectra on each of the 50 cocoa beans per hybrid, on both sides, finally resulting in 100 mean spectra for each variety (library with 500 spectra). In addition, a new set of cocoa bean hybrids containing 200 beans (200 images) was processed under the same conditions. This new dataset was acquired in order to demonstrate that the models created can predict on a completely independent set of dice, enhancing their robustness for industrial applications.

2.4 Data treatment

Chemometric data analysis was carried out using the PLS Toolbox from Eigenvector Research, Inc. (Manson, WA, USA) for Matlab R2017 (Mathworks, Natick, USA). First, the reflectance signal is converted to absorbance prior to spectral data treatments. The mean spectra of cocoa bean hybrids were pre-processed using standard normal variate (SNV) and the first derivative (Savitzky Golay, filter width 15 and a polynomial order of 2) to remove random shift of the baseline offset, light scattering interferences and noise (Vidal & Amigo, 2012).

Principal component analysis (PCA) was carried out to investigate systematic differences between samples and to find and remove outliers (Sendin, Manley, Baeten, Fernández Pierna, & Williams, 2019). For this purpose, spectral data were mean centered, then, PCA model were performed by singular value decomposition (SVD) algorithm with a confidence level of 0.95 for Q and T^2.

For discrimination purposes, Partial Least Squares discriminant analysis (PLS-DA) and Support Vector Machines (SVM) were applied on the data set. PLS-DA is a supervised lineal method widely used for food quality control and authentication. SVM is a supervised learning technique that works through the searching of hyperplanes (high dimensional feature space) by the use of kernel functions and penalization criteria, allowing both linear and non-linear classifications (Pierna, Baeten, Renier, Cogdill, & Dardenne, 2004).

Two approaches have been tested in this work. In the first approach, the aim was to construct PLS-DA and SVM models to discriminate between two cacao bean hybrids. The spectral data was constituted of 200 mean spectra for each model, of which 80 % (160 mean spectra) were randomly selected for the calibration and cross-validation of the models and the remaining 20 % (40 spectra) was used as test set to assess the discriminative capacity of the models. In the second approach, the aim was to construct one PLS-DA and SVM models to discriminate between the five cacao bean hybrids. Spectral data were randomly divided into two subsets: calibration and cross-validation set (80 % or 400 mean spectra) and test set (20 % or 100 mean spectra). In both cases, leave-one-out cross validation was applied when building the PLS-DA models, and the number of latent variables was chosen through the evaluation of sensitivity (Eq. 1), specificity (Eq. 2) and classification error (Eq. 3) of cross-validation and prediction (Nie et al., 2019). For SVM models, the penalty parameters of cost (c) and kernel function parameters gamma (g) were optimized using a grid search (Fernández Pierna et al., 2012; Pierna et al., 2011). The model performance was also statistically evaluated according to sensitivity, specificity and classification error.

$$Sensitivity\ (\%) = \frac{TP}{TP+FN} \quad \text{(Eq. 1)}$$

$$Specisficity\ (\%) = \frac{TN}{TP+FN} \quad \text{(Eq. 2)}$$

$$Error\ (\%) = \frac{FP+FN}{TP+FP+TN+FN} \quad \text{(Eq. 3)}$$

Where TP: true positive (positive samples correctly classify), TN: true negative (negative sample correctly classify), FP: false positive (positive samples incorrectly classify), FN: false negative (negative samples incorrectly classify).

For external validation of the models and to develop classification maps, a new sample test set (new cocoa bean hybrids not considered in the calibration or in the initial validation of the models) was used. The new test set consisted of 200 samples of the 5 cacao bean hybrids (20 % of each hybrid) and were divided as follows on two subsets: SET 1 contained hybrids of the same class and SET 2 contained the 5 hybrids (20 % of each hybrid) placed in known spatial positions in the image. The discriminative capacity of the PLS-DA models for two classes and for five classes were evaluated based on the correct classification rate (% CCR) for each hybrid in the new test set. Predictive ability was assessed on hyperspectral images (pixel-to-pixel) and measured as the % CCR using an algorithm with 4 approaches: (1) prediction using raw model, (2) prediction using majority vote, (3) filtered model by deleted samples no possible to classify (difference between 2 classes with more probability < 65 pixel), and 4) PLS-DA model applied to pixel with mean spectra value. All PLS-DA maps were constructed using their own program developed in Matlab R2017 (Mathworks, Natick, USA).

3. Results and discussion

3.1 Spectra profile

The raw spectra and pre-processed spectra of five cocoa bean hybrids are presented in Figures 2A and 2B, respectively. The mean spectra of all hybrids had similar pattern of

absorbance, but their relative absorbance was different in some spectral regions. Similarities in the shape of the spectrum are inherent to the specie (*Theobroma cacao*), while differences in absorbance are related to variations in the composition of the shell of cocoa beans (Quelal-Vásconez et al., 2019). Because all hybrids came from the same geographic area and the postharvest process (fermentation and drying conditions), it is possible to assume that variations in composition, and therefore spectral similarities and differences, are a consequence of hybrid genetics.

Absorption bands 1181 and 1426 nm correspond to the second overtone of O–H stretching and O–H deformation (Teye et al., 2013), which is associated with water and fiber content (Okiyama et al., 2017). Also, the band at 1426 nm can also be associated to the first overtone of the N–H stretching vibration. Therefore this band is associated to –CONH– structure (peptide) and related to protein content in shell of cacao beans (Mandrile et al., 2019; Quelal-Vásconez et al., 2019; Veselá et al., 2007). Band absorption 1263 nm is associated to C–H stretching second overtone ($–CH_3$ or $–CH_2$), due to the presence of fibers and carbohydrates (Okiyama et al., 2017; Osborne, Fearn, Hindle, & Osborne, 1993). The absorption bands of 1533, 1577 and 1650 - 1780 nm are associated to functional groups like (−)-epicatechin (flavanol), theobromine and caffeine (alkaloids), proteins, volatile and non-volatile acids (Álvarez et al., 2012; Teye et al., 2013; Veselá et al., 2007). Also, absorption band at 1715 nm is mainly related to fat content and fatty acid. The spectral region at 1900 - 1950 nm is mainly associated with O-H combinations, influenced by moisture content in cocoa bean. It has also been reported that the band at 1916 nm is related to the second overtone of C=O and it is associated with the content of (−)-epicatechin (Álvarez et al., 2012) and lignin content. While the band at 1990 nm is related to O-H combinations and with asymmetric stretching of N-H and amide II group, strongly influenced by the content of protein.

Bands at 2092 and 2280 nm are also related to second overtone of CH=CH and CH_3 combination respectively, and they are characteristics of lignin, aromatics, polyphenols and fatty acids in cacao beans (Sunoj, Igathinathane, & Visvanathan, 2016; Teye et al., 2015). However, the peak at 2092 nm is associated to protein content (Caporaso et al., 2018; Veselá et al., 2007). The absorption bands at 2199 and 2375 nm could be attributed to the combination of C–H (Ma, Wang, Chen, Cheng, & Lai, 2017) and to the stretching and rocking vibrations of C-H and C-C of cellulose and hemicellulose (Okiyama et al., 2017; Wang et al., 2018), respectively. These bands were previously reported for cocoa bean shell (Mandrile et al., 2019; Quelal-Vásconez et al., 2019).

3.2 PCA

PCA models were performed using pre-processed data (Figure 2C) to identify possible clusters and to evaluate the effect of the pre-processing (SNV + 1st derivate).

PCA shows the possibility, although not easily, to observe groups of hybrids of the same class. H1 hybrid (purple symbol) was characterized by negative scores on PC1 and by negative scores on PC2. H2 (red symbol) and H4 (green symbol) hybrids were located on positive scores on PC1. While H2 hybrid presents greater variability on PC2, the H5 hybrid were characterized by positive scores on PC2. H3 (greenlight blue symbol) hybrid are characterized by negative scores on PC2, although they have great variability throughout PC1. Besides, H1 hybrids overlap on H2 hybrids, with whom they share a common ancestor (Figure 1). The results of the PCA show that for the same hybrid there is great variability between the samples, while there is an overlap between the hybrids of different classes. This could occur due to two situations: (1) the cocoa bean shell has fiber and protein as the majority components (Mandrile et al., 2019) and, (2) the five hybrids share genetic ancestry, so the concentration of fiber and protein can be very similar. Therefore, the overlap may be associated with the similarity in its major

components, while the clustering would be associated with the difference in minor components such as fatty acid tryptamides (FATs, more abundant in shell) (Quelal-Vásconez et al., 2018), phenolic compounds and aromatic compounds (Okiyama et al., 2017) developed during fermentation and drying for each class of hybrid.

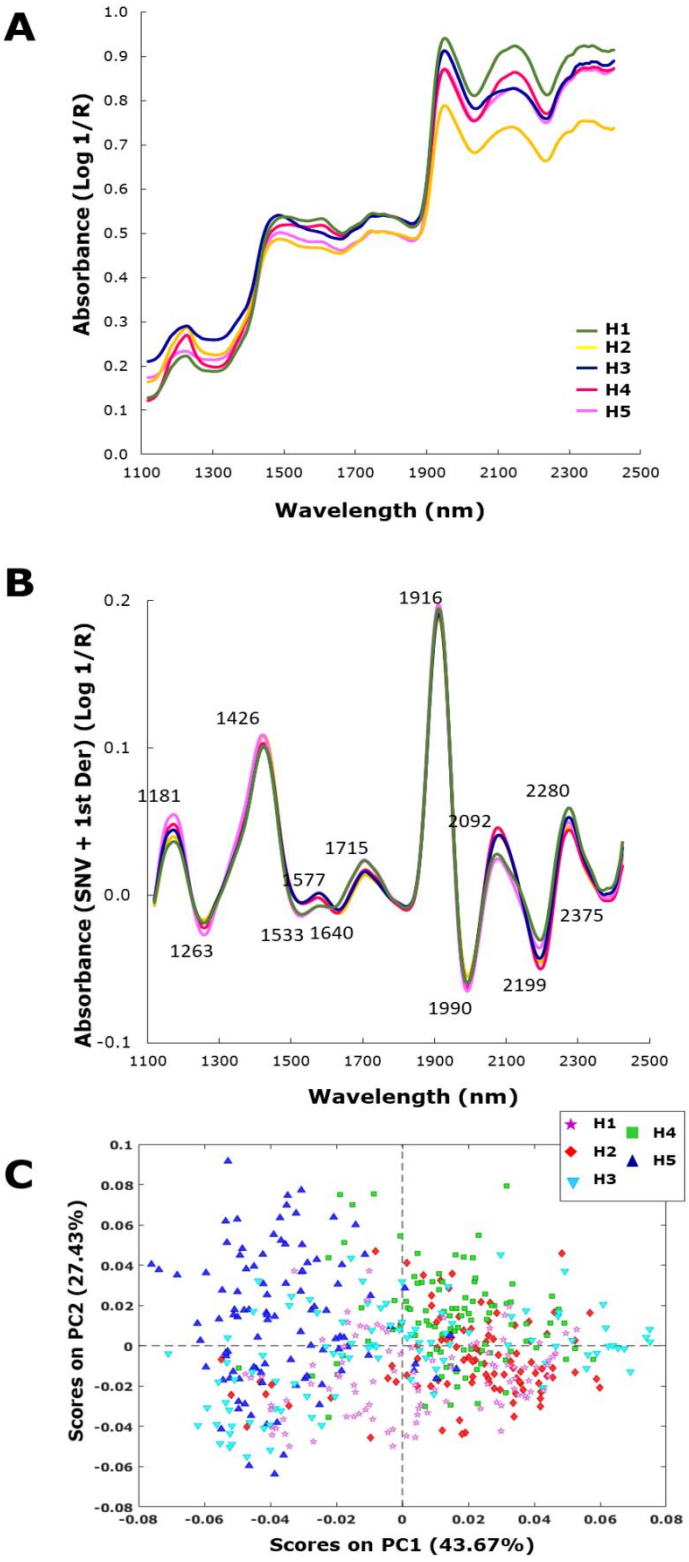


Figure 2. Spectral profile of raw spectra (A), spectra treated with SNV + first derivative (B) and PCA scores based on spectra treated with SNV + first derivate (C).

3.3 Discriminant analysis

PLS-DA models were created using the full spectral region (1100 - 2400 nm), however, the results (data no show) showed little sensitivity to discriminate between the 5 hybrids. Therefore, in this work, PLS-DA and SVM was carried out in two approaches using spectral data in the range 1369 – 2054 nm. This spectral region was more significant for discrimination (lower % classification error) than using the full spectrum. PLS-DA and SVM discrimination models were calibrated and validated using the mean spectrum of the cocoa bean hybrids. Discrimination models were created to identify a specific hybrid (genetic cross-linking) within a heterogeneous batch of cocoa beans, where the provenance of each bean was known. These models were created to help the industry establish the purity of a batch of cocoa beans, more necessarily to associate it with the origin of a specific tree, since cocoa trees are hermaphrodites, and can contain more than two hybrids to the same time.

3.3.1 IMC vs P7 classes

Here, we investigated the possibility of constructing PLS-DA and SVM classification models to discriminate between hybrids according to their "mother", being IMC 67 for H1 and H2, and P7 for H3 and H4. The PLS-DA model obtained a better result than the SVM model (11.3 % *vs* 17.5 % error, Table 1), although with a large number of latent variables (14). Recently, Scollo, Neville, Oruna-Concha, Trotin, & Cramer (2020) reported that the IMC 67 hybrid had a different protein profile from other types of hybrids (e.g. content of aminohydrolase), which allowed it to be clearly classified. Loading plot (Figure 2A, supplementary material) shows the importance of spectral region 1900 – 1990 nm for PLS-DA model, which are strongly associated with O-H combinations, mainly water, but it also is associated to functional groups of proteins, polyphenols, fatty acid and aromatics (Caporaso et al., 2018; Teye et al., 2013).

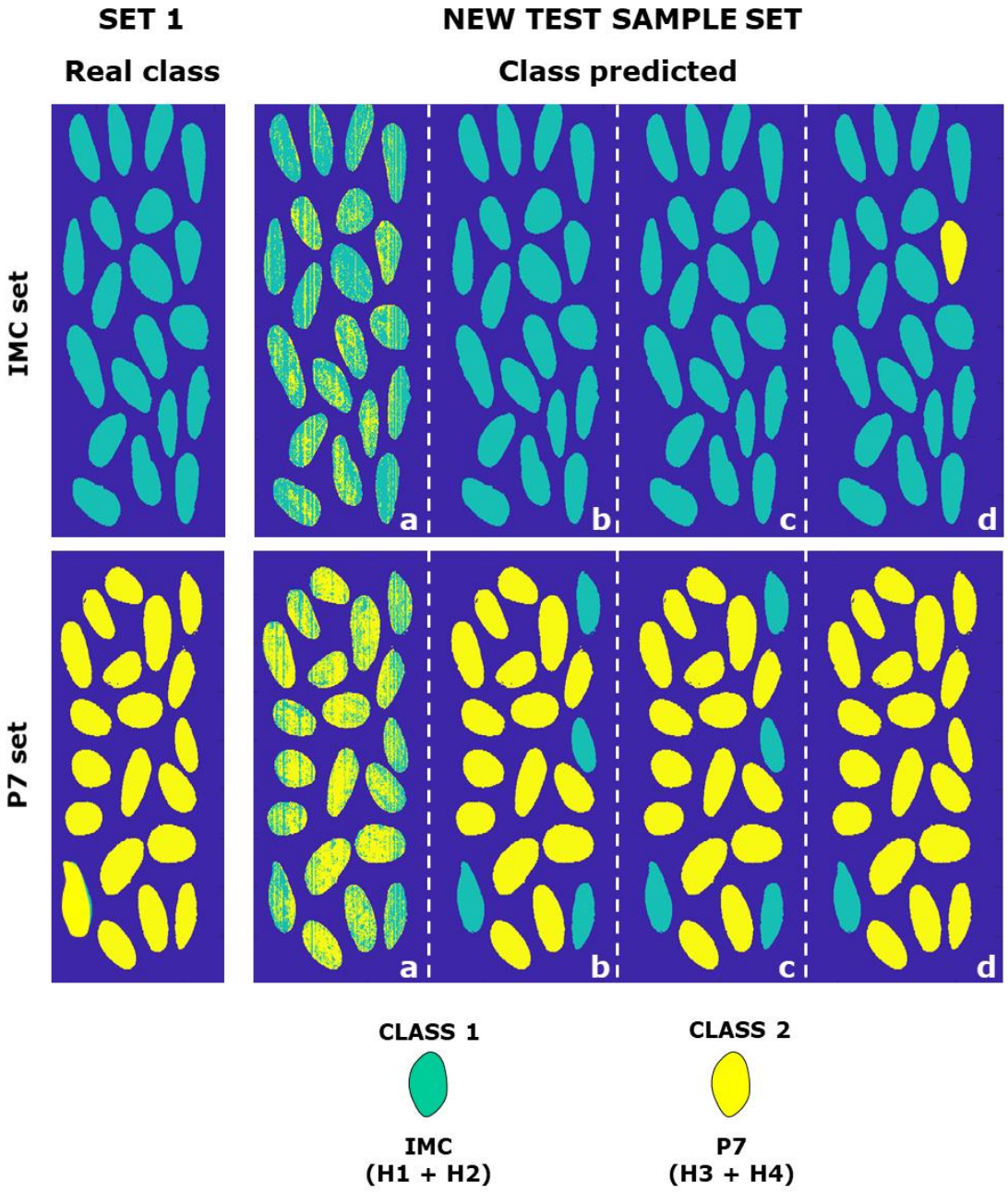


Figure 3. PLS-DA analysis for two classes of hybrids of cacao beans according to 'mother': IMC *vs* P7. PLS-DA maps for the most probable class assigned to cocoa bean hybrids were performed according approach: a) PLS-DA model; b) model applying majority vote; c) filtered model by deleted samples no possible to classify (difference between 2 classes with more probability < 65 pixel); d) PLS-DA model applied to pixel with mean spectra value.

Figure 3 shows the classification results (based on mean spectrum) and the pixel-to-pixel allocation of each hybrid class (IMC *vs.* P7). The classification results show as five beans of IMC hybrids and four beans of P7 hybrids are misclassified. For the pixel-to-pixel classification, using majority vote (Figure 3b), all grains are correctly assigned to the ICM class (100 %), while for class P7 only 80 % of the correct classification was reached (16/20). The approach c (Figure 3c) shows that cocoa bean hybrids from P7 misclassified as IMC always had a larger number of pixels similar to the IMC class (class 1 - class 2 < 65 pixels). This is also visible in Figure 3a. The higher discrimination capability may be associated with the ancestry of the "mothers" of the hybrids. IMC 67 is a variety from the genotype "Iquitos", while P7 probably comes from the "Namay" or "Marañon" groups (Motamayor et al., 2008). Satisfactorily, the P7 hybrids classification improved when the average spectrum value was applied to each pixel for classification (Figure 3d), reaching 95% correct classification (19/20). Because the shell is maternally derived from the integuments of the ovary, its composition is preferably dominated by "mother" genetic. In this sense, success in the discrimination of hybrids is probably associated with the particularities in composition of each "mother" species.

Table 1. Performance of 2-classes PLS-DA and SVM classification models for the hybrids of cocoa beans obtained by HSI in the spectral region 1369–2054 nm.

Hybrid	Model	Parameter*	Sensitivity		Specificity		Error	
			CV	Pred	CV	Pred	CV	Pred
IMC *vs* P7	PLS-DA	14	0.969	0.875	0.963	0.900	0.034	0.113
	SVM	(1;0.0316)	0.956	0.775	0.963	0.875	0.041	0.175
H1 *vs* H2	PLS-DA	9	0.975	1.000	0.975	1.000	0.025	0.000
	SVM	(10;0.0316)	0.988	1.000	0.963	0.750	0.025	0.125
H1 *vs* H3	PLS-DA	6	0.975	0.700	0.925	0.900	0.069	0.200
	SVM	(0.3;0.0316)	0.975	0.600	0.975	0.950	0.025	0.225

H2 *vs* H4	PLS-DA	6	0.963	0.650	0.975	0.900	0.031	0.225
	SVM	(1;0.0316)	1.000	0.650	0.988	0.950	0.006	0.200
H2 *vs* H5	PLS-DA	4	0.975	0.750	0.975	0.950	0.025	0.150
	SVM	(100;0.00031)	0.975	0.700	1.000	1.000	0.125	0.150
H1 *vs* H3	PLS-DA	4	0.963	1.000	0.938	0.850	0.050	0.075
	SVM	(31.62;0.001)	0.963	1.000	0.925	8.000	0.050	0.050
H1 *vs* H4	PLS-DA	8	1.000	1.000	0.975	9.000	0.013	0.050
	SVM	(1;0.1)	1.000	0.988	1.000	0.950	0.006	0.025
H1 *vs* H5	PLS-DA	5	0.988	1.000	1.000	0.950	0.006	0.025
	SVM	(100;0.00316)	1.000	1.000	0.988	0.950	0.006	0.025
H3 *vs* H4	PLS-DA	12	0.963	0.900	0.988	0.900	0.025	0.100
	SVM	(3.16;0.0316)	0.875	0.900	0.950	0.850	0.088	0.125
H3 *vs* H5	PLS-DA	8	0.938	0.800	0.950	0.900	0.056	0.150
	SVM	(3.16;0.001)	0.963	0.950	0.975	0.900	0.031	0.075
H4 *vs* H5	PLS-DA	8	0.988	0.900	0.975	1.000	0.019	0.050
	SVM	(10;0.01)	1.000	0.900	1.000	1.000	0.000	0.050

PLS-DA: partial least square discriminant analysis; SVM: support vector machine discriminant analysis; CV: cross-validation; Pred: prediction.

*Parameter for PLS-DA model's means the optimal number of LVs and for SVM model's mean different penalty parameters: cost (c) and kernel function parameters gamma (g).

3.3.1 2-class models

Here, 2-classes PLS-DA and SVM models were built to discriminate between two specific cocoa bean hybrids, and the performance of models is showed in Table 1. In general, the PLS-DA and SVM models showed a good performance with a % sensitivity of 60 – 100 % and a % error of 0 – 22.5 %. In most cases, the% error for SVM models was less than or equal to that for PLS-DA models, except for models H1 *vs* H2 and H3 *vs* H4. This is probably because the SVM algorithm is more complex and can work linearly and non-linearly way. However, generally speaking, a significant improvement was probably not observed for SVM models due to the complexity of the hybrids. The classification models were more sensitive for H1 in all cases (Table 1), which allows their clear distinction from any other hybrid.

For all PLS-DA models, the loadings (Figure 2B-K, supplementary material) showed the importance of the spectral region around 1650 nm is associated to functional groups like (−)-epicatechin, theobromine and caffeine content, and the band at 1715 is associated to fatty acid in cocoa beans, as previously reported by Alvarez et al. (2012). Also, the spectral region at 1900 – 1990 nm is associated with asymmetric stretching of N-H and amide II group present in the protein, which is second larger component in cocoa shell (116 – 181 g protein/kg dried cocoa shell) (Okiyama et al., 2017). Also, each hybrid has a particular protein profile (Scollo et al., 2020), allowing the formation of flavor precursors in cocoa beans, however, not all protein is degraded or transformed, since the proteolytic processes are different in each hybrid (Moreira, Vilela, Santos, Lima, & Schwan, 2018).

The H1 *vs* H2 model allowed, in the new external validation set (Figure 4A) to correctly classify all cocoa bean hybrids (100 %). Figure 4A, (approach *a*) shows the model sensitivity for the H1 class, where almost 100 % of the pixels corresponding to H1 hybrid were assigned to that class. That sensitivity was lower for H2, although not sufficient to incorrectly classify hybrids in any of the approaches (Figure 4A, approach *b-d*). For H1 hybrid, protein content and profile, which are quite specific and distinctive in their predecessors IMC 67 and ICS 1 (fine cocoa) (Scollo et al., 2020).

The model H3 *vs* H4 instead showed less sensitivity for prediction pixel-to-pixel (Figure 4B, approach *a*). Using majority vote (Figure 4B, approach *b*), the hybrids of class H4 were correctly classified at 100 %, while class H3 was correctly classified at 70 %. However, the best prediction was achieved using the third approach, where H3 and H4 hybrids were correctly classified at 100 % and 90 % (Figure 4B, d). Here, parents SIAL505 (H4) and BE10 (H3) (Figure 1) belong to the genotype "Amelonado”

(Motamayor et al., 2008). This would explain why some hybrids of the H3 class are confused with hybrids of the H4 class.

Figure 4C (approach *a*) shows that H5 hybrids show a definite assignment of the class, with almost all pixels correctly assigned. This facilitated their authentication using any of the three approaches (Figure 4C, b-d) achieved a 100 % correct classification in the new test set of H5 and 90 % in the new test set of H2 (data not shown). Probably, the success in the classification of the hybrids H2 and H5 is related to the difference in the genotype to which their mothers belonged, being IMC 67 (mother of H2) belonging to the genotype "Iquitos" and PA121 (mother of H5) belonging to the genotype “Marañon (Parinari I)” (Motamayor et al., 2008).

The differentiation between the H5 and H4 hybrids was more complicated (Figure 4D), reaching a 70 % correct classification for the H5 hybrid and 90 % for the H4 hybrid. A similar result with 80% correct classification for H5 and 100 % for H3 (data not shown) was found. None of the three approaches (Figure 4D, b-d) allowed an improvement in the prediction of cocoa bean hybrids. Hypothetically, the greater sensitivity of the model to identify the P7 samples suggests that these hybrids have a distinctive composition in residual compounds (e.g. protein fractions) that in quantity and type may be greater than in H5 hybrids.

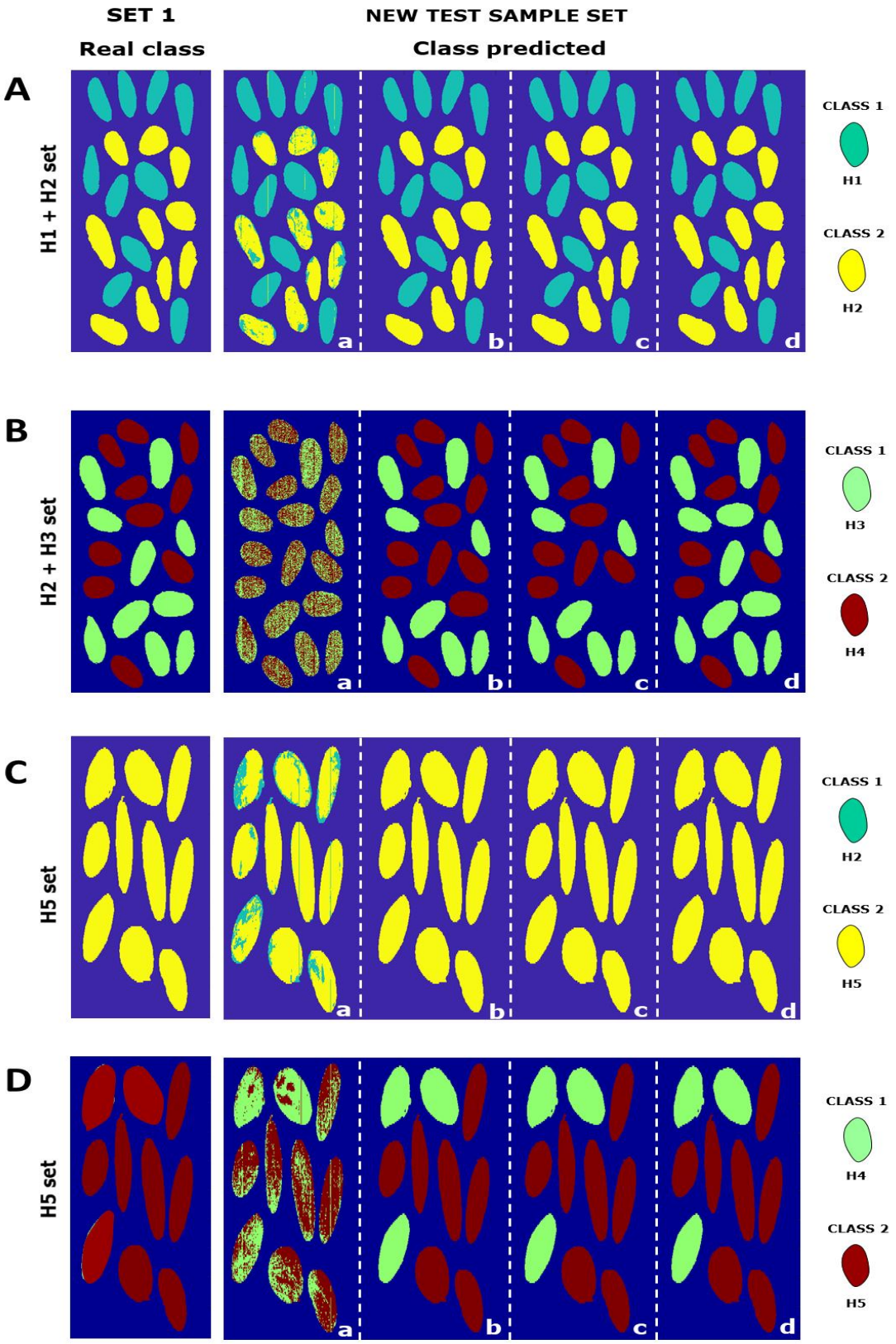


Figure 4. PLS-DA analysis for two classes of hybrids of cacao beans. A) Model H1 *vs* H2; B) Model H3 *vs* H4; C) Model H1 *vs* H5; D) Model H3 *vs* H5. PLS-DA maps for

the most probable class assigned to new test sample set were performed according approach: a) PLS-DA model; b) model applying majority vote; c) filtered model by deleted samples no possible to classify (difference between 2 classes with more probability < 65 pixel); d) PLS-DA model applied to pixel with mean spectra value.

From the results presented in this section, we can conclude the offspring hybrids of IMC 67 (H1 and H2) presumably have a particularly distinctive composition of *fine cocoa* than the other hybrids, especially the H1 hybrid. This may be associated with its ancestor ICS1, which is a hybridization of cocoa with a fine aroma (Criollo and Trinitario) (Castro-Alayo, Idrogo-Vásquez, Siche, & Cardenas-Toro, 2019; Kongor et al., 2016; Scollo et al., 2020). The performance of classification models was similar to reported to discriminate two-classes of Ecuadorian cocoa bean genotypes using Raman spectroscopy (accuracy = 91.8%) (Vargas Jentzsch et al., 2016) and computer vision (until 98%) (Jimenez et al., 2018), even reaching an error of 0 % for some pairs of hybrids that was not previously reported.

3.3.2 5-classes model

In this step, one PSL-DA and one SVM model for the 5 cocoa bean hybrids. Table 2 shows the performance evaluated as sensitivity, specificity and error for PLS-DA and SVM models. Both PLS-DA and SVM showed good performance to discriminate between the five types of cocoa hybrids (Figure 1). The lowest sensitivity and the highest error in cross-validation and prediction set was recorded for the hybrids H2 and H3 for both PLS-DA and SVM models. However, the prediction error for the test set was lower (18.1 % for H2 and 23.1 % for H3) for SVM models compared to PLS-DA models (23.1 % for H2 and 34.4 % for H3). In contrast, both models showed a good performance (lower error in test set) for the H1 hybrid, suggesting that this cocoa hybrid could have a rather particular composition compared to the other hybrids. Therefore, its

classification is more reliable, which is clearly shown in the low prediction error in test set (4.4 % for PLS-DA and 3.8 % for SVM).

Table 2. Performance of the 5-classes PLS-DA and SVM classification model for the hybrids of cacao beans obtained by HSI in the spectral region of 1369–2054 nm.

Hybrid	Model	Parameter*	Sensitivity		Specificity		Error	
			CV	Pred	CV	Pred	CV	Pred
H1	PLS-DA	12	0.963	1.000	0.944	0.912	0.047	0.044
H2			0.938	0.600	0.897	0.938	0.083	0.231
H3			0.787	0.500	0.806	0.813	0.203	0.344
H4			0.925	0.850	0.881	0.813	0.097	0.169
H5			0.938	0.950	0.956	0.925	0.048	0.053
H1	SVM	(10;0,01)	0.963	1.000	0.994	0.925	0.022	0.038
H2			0.950	0.650	0.984	0.988	0.033	0.181
H3			0.875	0.600	0.975	0.938	0.075	0.231
H4			0.975	0.900	0.978	0.938	0.023	0.081
H5			0.925	0.900	0.991	0.975	0.042	0.063

PLS-DA: partial least square discriminant analysis; SVM: support vector machine discriminant analysis; CV: cross-validation; Pred: prediction.

*Parameter for PLS-DA model's means the optimal number of LVs and for SVM model's mean different penalty parameters: cost (c) and kernel function parameters gamma (g).

In Figure 5 it can be seen how many of the hybrids presented pixels of all classes. More clearly, PLS-DA maps (approach *a*) show how for the same hybrid, pixels of all classes can exist. This is mainly associated with shared genetic information, such as sharing the same mother (1/2 genetic information, especially for H3 and H4 hybrids), especially "mother" genetic that is responsible for shell composition. Meanwhile, it is likely that not all compounds present in the shell have been transformed after the fermentation and

drying process and, therefore, similar pixels in each cocoa bean are associated with the fiber, alkaloids (e.g. theobromine (Biehl & Ziegleder, 2003)) or some proteins (e.g. albumins (Dodo, Fritz, & Furtek, 1992)) that do not degrade or transformed completely during fermentation.

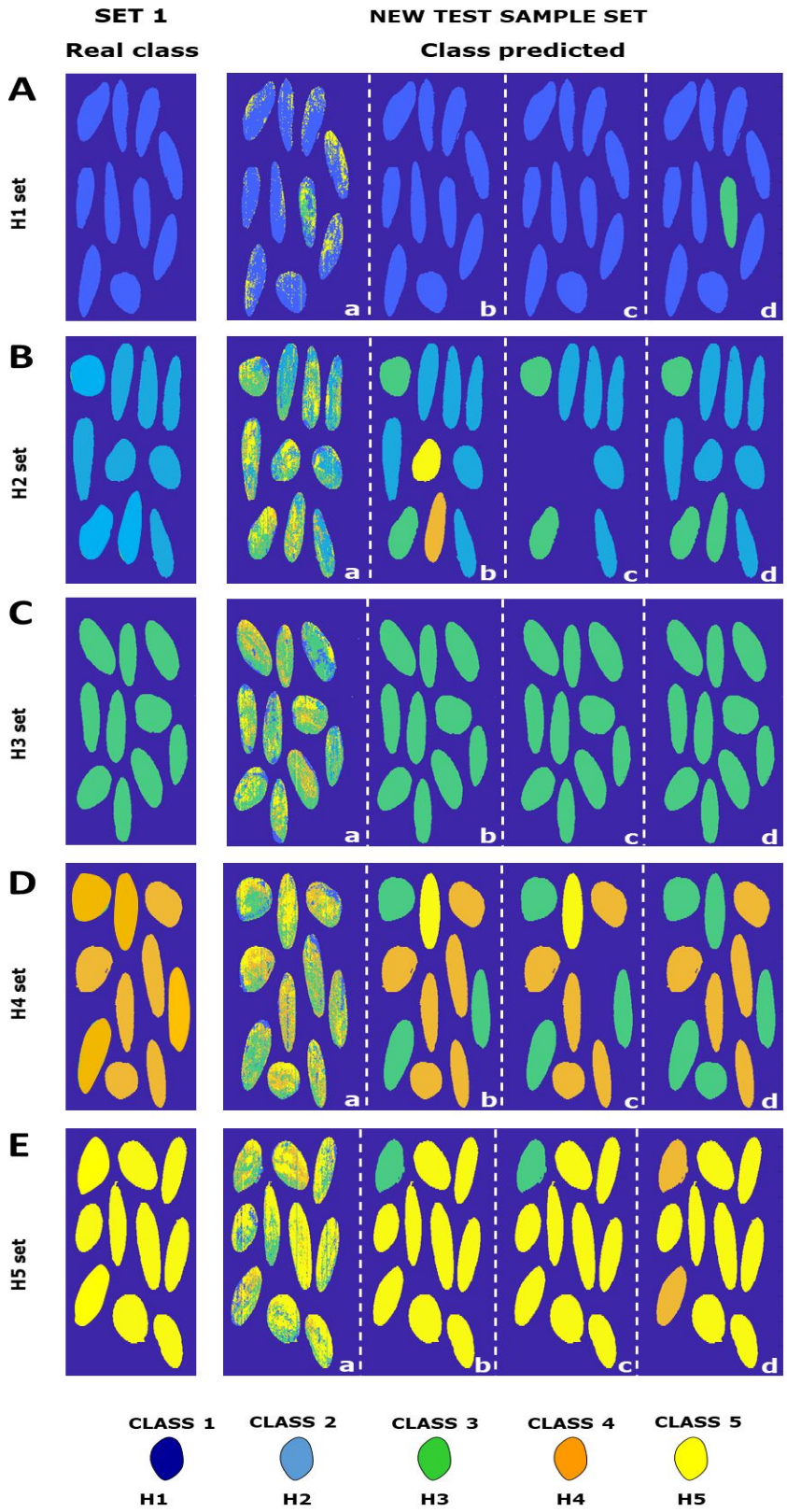


Figure 5. PLS-DA analysis for five classes of hybrids of cacao beans. A) H1; B) H2; C) H3; D) H4; E) H5. PLS-DA maps for the most probable class assigned to new test

sample set were performed according approach: a) PLS-DA model; b) model applying majority vote; c) filtered model by deleted samples no possible to classify (difference between 2 classes with more probability < 65 pixel); d) PLS-DA model applied to pixel with mean spectra value.

Figure 6 shows the 5-classes PLS-DA map created using a new data set (SET 2), which had samples with known spatial position and the same number of samples for each hybrid (10). For new data set (SET 2), Figure 6 (approach *a*) reflects the genetic complexity of hybrids, showing how some cocoa beans can be assigned pixels of 2 or more classes. However, as in 2-class models, the H1 (Class 1) hybrid clearly differs from the other hybrids. Using approaches *b* and *c*, H1 correct classification reached 90 %, but it reached 100 % when the average spectrum value is assigned to each pixel (Figure 6, d). This further reinforces the theory that hybrids descended from IMC 67 are compositionally particular. While for the H3 samples, the correct classification was 80 % in all cases. Also, Figure 6c shows that two cocoa beans (1 from H3 and 1 H4) were delighted because there was no significant number of pixels assigned to a single class (difference between number of pixels in class H3 and class H4 was <65 pixels).

For its part, H4 achieved a 100 % correct classification using any of the proposed approaches. The greatest difficulty for the 5-class model was to identify the H5 and H3 hybrids, which were confused with the H4 hybrid (correct classification 40 – 60 %), using any approach. This behavior was also previously observed in the 2-class models. This could indicate that these hybrids are likely to be compositionally very similar, so they could have developed the same flavor compounds in cocoa.

Finally, using 5-classes models we reached a correct prediction was 60 – 100 % for the set of cocoa beans that contained a single hybrid and between 40 – 100 % for cocoa beans set containing all hybrids. Previously, it was reported that using SSRs markets,

the classification error of cocoa germplasm was 15 – 44 % (Motilal & Butler, 2003). So we can say that our 5-classes model, in addition to not being destructive, chemical-free and requiring a minimum sample preparation, is reliable for the discrimination of cocoa hybrids.

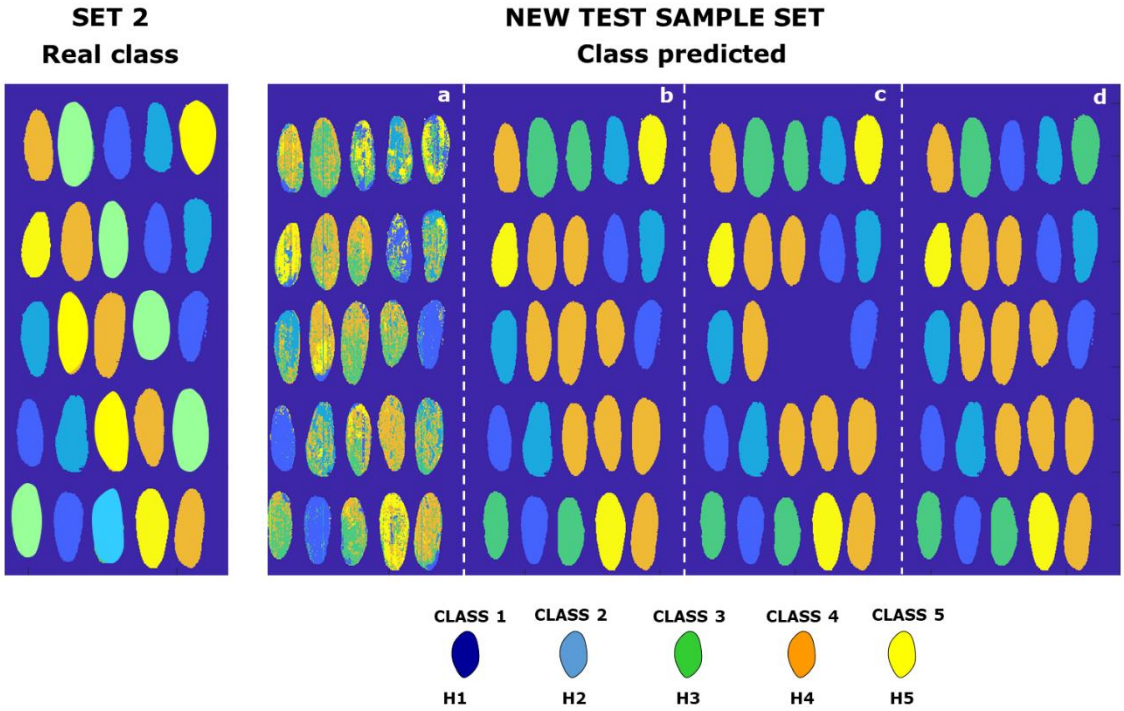


Figure 6. PLS-DA maps for the most probable class assigned to new test sample set. (Set 2) with hybrids with known spatial position. PLS-DA maps for the most probable class assigned to new test sample set were performed according approach: a) PLS-DA model; b) model applying majority vote; c) filtered model by deleted samples no possible to classify (difference between 2 classes with more probability < 65 pixel); d) PLS-DA model applied to pixel with mean spectra value.

4. Conclusion

In this paper, the development and robust validation of a method based on hyperspectral images to identify and classify cocoa bean hybrids has been proposed. The results indicated that comparable results are obtained for both PLS-DA and SVM for the 2-

class models, however, for the model that included the 5 classes, the SVM models showed a significant improvement reducing the prediction error. The H1 hybrid was the most distinguishable in all cases. A second validation using a new data set (SET 1 and SET 2) was performed for pixel-to-pixel prediction of hybrid classes. The maps show the reliability of the 2-class models to classify all the hybrids correctly (70 – 100 % CCR). While the prediction in the image using the 5-class model allows a 100 % correct classification of the H1, H2 and H4 hybrids. Future works should investigate the composition of these hybrids as well as that of their ancestors, thus allowing less factual conclusions to be established. In addition, the feasibility of the method in other hybrids and with a larger number of samples should be investigated.

Acknowledgements

This study was financed in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001; and São Paulo Research Foundation (FAPESP) (project number 2019/04833-3, 2018/02500-4, 2015/24351-2). This study was also financed by the BELCOBRA 'Optimization and implementation of analytical methods for the traceability and authenticity of cocoa and chocolate' (Brazil-Belgium cooperation CAPES-WBI/2017-2019/361394). Authors are grateful to CEPLAC (Medicilandia, PA, Brazil) for the mature fruits, to Teixeira L.E.O. and Freitas A.L. (UFPA) for their technical assistance for fermentation and drying process and to N. Kayoka (CRA-W) for hi technical assistance in NIR-HSI measurements. J. P. Cruz-Tirado acknowledges scholarship funding from FAPESP/BEPE, grant n° 2019/04833- 3.

References

Álvarez, C., Pérez, E., Cros, E., Lares, M., Assemat, S., Boulanger, R., & Davrieux, F. (2012). The Use of near Infrared Spectroscopy to Determine the Fat, Caffeine, Theobromine and (−)-Epicatechin Contents in Unfermented and Sun-Dried Beans of Criollo Cocoa. *Journal of Near Infrared Spectroscopy*, *20*(2), 307–315.

Baeten, V., Pierna, J. A. F., Vermeulen, P., & Dardenne, P. (2010). NIR Hyperspectral Imaging Methods for Quality and Safety Control of Food and Feed Products: Contributions to Four European Projects. *NIR News*, *21*(6), 10–13.

Barbin, D. F., Maciel, L. F., Bazoni, C. H. V., Ribeiro, M. da S., Carvalho, R. D. S., Bispo, E. da S., … Hirooka, E. Y. (2018). Classification and compositional characterization of different varieties of cocoa beans by near infrared spectroscopy and multivariate statistical analyses. *Journal of Food Science and Technology*, *55*(7), 2457–2466.

Biehl, B., & Ziegleder, G. (2003). *COCOA | Chemistry of Processing* (B. B. T.-E. of F. S. and N. (Second E. Caballero, Ed.).

Caporaso, N., Whitworth, M. B., Fowler, M. S., & Fisk, I. D. (2018). Hyperspectral imaging for non-destructive prediction of fermentation index, polyphenol content and antioxidant activity in single cocoa beans. *Food Chemistry*, *258*, 343–351.

Castro-Alayo, E. M., Idrogo-Vásquez, G., Siche, R., & Cardenas-Toro, F. P. (2019). Formation of aromatic compounds precursors during fermentation of Criollo and Forastero cocoa. *Heliyon*, *5*(1), e01157.

Dale, L. M., Thewis, A., Boudry, C., Rotar, I., Dardenne, P., Baeten, V., & Pierna, J. A. F. (2013). Hyperspectral Imaging Applications in Agriculture and Agro-Food Product Quality and Safety Control: A Review. *Applied Spectroscopy Reviews*, *48*(2), 142–159.

Dinarti, D., Susilo, A. W., Meinhardt, L. W., Ji, K., Motilal, L. A., Mischke, S., & Zhang, D. (2015). Genetic diversity and parentage in farmer selections of cacao from Southern Sulawesi, Indonesia revealed by microsatellite markers. *Breeding Science*, *65*(5), 438–446.

Diomande, D., Antheaume, I., Leroux, M., Lalande, J., Balayssac, S., Remaud, G. S., & Tea, I. (2015). Multi-element, multi-compound isotope profiling as a means to distinguish the geographical and varietal origin of fermented cocoa (Theobroma cacao L.) beans. *Food Chemistry*, *188*, 576–582.

Dodo, H. W., Fritz, P. J., & Furtek, D. B. (1992). A cocoa 21 kilodalton seed protein has trypsin inhibitory activity. *Cafe Cacao The (France)*.

Eylenbosch, D., Bodson, B., Baeten, V., & Fernández Pierna, J. A. (2018). NIR hyperspectral imaging spectroscopy and chemometrics for the discrimination of roots and crop residues extracted from soil samples. *Journal of Chemometrics*, *32*(1), e2982.

Fang, W., Meinhardt, L. W., Mischke, S., Bellato, C. M., Motilal, L., & Zhang, D. (2014). Accurate Determination of Genetic Identity for a Single Cacao Bean, Using Molecular Markers with a Nanofluidic System, Ensures Cocoa Authentication. *Journal of Agricultural and Food Chemistry*, *62*(2), 481–487.

Fernández Pierna, J. A., Baeten, V., & Dardenne, P. (2006). Screening of compound feeds using NIR hyperspectral data. *Chemometrics and Intelligent Laboratory Systems*, *84*(1), 114–118.

Fernández Pierna, J. A., Vermeulen, P., Amand, O., Tossens, A., Dardenne, P., & Baeten, V. (2012). NIR hyperspectral imaging spectroscopy and chemometrics for the detection of undesirable substances in food and feed. *Chemometrics and*

Intelligent Laboratory Systems, *117*, 233–239.

Guo, D., Zhu, Q., Huang, M., Guo, Y., & Qin, J. (2017). Model updating for the classification of different varieties of maize seeds from different years by hyperspectral imaging coupled with a pre-labeling method. *Computers and Electronics in Agriculture*, *142*, 1–8.

Herrmann, L., Felbinger, C., Haase, I., Rudolph, B., Biermann, B., & Fischer, M. (2015). Food Fingerprinting: Characterization of the Ecuadorean Type CCN-51 of Theobroma cacao L. Using Microsatellite Markers. *Journal of Agricultural and Food Chemistry*, *63*(18), 4539–4544.

Jimenez, J. C., Amores, F. M., Solórzano, E. G., Rodríguez, G. A., La Mantia, A., Blasi, P., & Loor, R. G. (2018). Differentiation of Ecuadorian National and CCN-51 cocoa beans and their mixtures by computer vision. *Journal of the Science of Food and Agriculture*, *98*(7), 2824–2829.

Kongor, J. E., Hinneh, M., de Walle, D. Van, Afoakwa, E. O., Boeckx, P., & Dewettinck, K. (2016). Factors influencing quality variation in cocoa (Theobroma cacao) bean flavour profile — A review. *Food Research International*, *82*, 44–52.

Kumari, N., Grimbs, A., D'Souza, R. N., Verma, S. K., Corno, M., Kuhnert, N., & Ullrich, M. S. (2018). Origin and varietal based proteomic and peptidomic fingerprinting of Theobroma cacao in non-fermented and fermented cocoa beans. *Food Research International*, *111*, 137–147.

Liu, X., Feng, X., Liu, F., & He, Y. (2017). Identification of hybrid rice strain based on near-infrared hyperspectral imaging technology. *Transactions of the Chinese Society of Agricultural Engineering*, *33*(22), 189–194.

Liu, Y., Wu, T., Yang, J., Tan, K., & Wang, S. (2019). Hyperspectral band selection for soybean classification based on information measure in FRS theory. *Biosystems Engineering*, *178*, 219–232.

Ma, H., Wang, J., Chen, Y., Cheng, J., & Lai, Z. (2017). Rapid authentication of starch adulterations in ultrafine granular powder of Shanyao by near-infrared spectroscopy coupled with chemometric methods. *Food Chemistry*, *215*, 108–115.

Mandrile, L., Barbosa-Pereira, L., Sorensen, K. M., Giovannozzi, A. M., Zeppa, G., Engelsen, S. B., & Rossi, A. M. (2019). Authentication of cocoa bean shells by near- and mid-infrared spectroscopy and inductively coupled plasma-optical emission spectroscopy. *Food Chemistry*, *292*, 47–57.

Mite-Baidal, K., Solís-Avilés, E., Martínez-Carriel, T., Marcillo-Plaza, A., Cruz-Ibarra, E., & Baque-Bustamante, W. (2019). *Analysis of Computer Vision Algorithms to Determine the Quality of Fermented Cocoa (Theobroma Cacao): Systematic Literature Review BT - ICT for Agriculture and Environment* (R. Valencia-García, G. Alcaraz-Mármol, J. del Cioppo-Morstadt, N. Vera-Lucio, & M. Bucaram-Leverone, Eds.). Cham: Springer International Publishing.

Moreira, I. M. da V., Vilela, L. de F., Santos, C., Lima, N., & Schwan, R. F. (2018). Volatile compounds and protein profiles analyses of fermented cocoa beans and chocolates from different hybrids cultivated in Brazil. *Food Research International*, *109*, 196–203.

Motamayor, J. C., Lachenaud, P., da Silva e Mota, J. W., Loor, R., Kuhn, D. N., Brown, J. S., & Schnell, R. J. (2008). Geographic and Genetic Population Differentiation of the Amazonian Chocolate Tree (Theobroma cacao L). *PLOS ONE*, *3*(10), e3311.

Motilal, L., & Butler, D. (2003). Verification of identities in global cacao germplasm collections. *Genetic Resources and Crop Evolution*, *50*(8), 799–807.

Nie, P., Zhang, J., Feng, X., Yu, C., & He, Y. (2019). Classification of hybrid seeds using near-infrared hyperspectral imaging technology combined with deep learning. *Sensors and Actuators B: Chemical*, *296*, 126630.

Okiyama, D. C. G., Navarro, S. L. B., & Rodrigues, C. E. C. (2017). Cocoa shell and its compounds: Applications in the food industry. *Trends in Food Science & Technology*, *63*, 103–112.

Osborne, B. G., Fearn, T., Hindle, P. H., & Osborne, B. G. (1993). *Practical NIR spectroscopy with applications in food and beverage analysis* (Vol. 2). Longman Scientific & Technical Harlow.

Pierna, J. A. F., Baeten, V., Renier, A. M., Cogdill, R. P., & Dardenne, P. (2004). Combination of support vector machines (SVM) and near-infrared (NIR) imaging spectroscopy for the detection of meat and bone meal (MBM) in compound feeds. *Journal of Chemometrics*, *18*(7-8), 341–349.

Pierna, J. A. F., Lecler, B., Conzen, J. P., Niemoeller, A., Baeten, V., & Dardenne, P. (2011). Comparison of various chemometric approaches for large near infrared spectroscopic data of feed and feed products. *Analytica Chimica Acta*, *705*(1), 30–34.

Quelal-Vásconez, M. A., Lerma-García, M. J., Pérez-Esteve, É., Arnau-Bonachera, A., Barat, J. M., & Talens, P. (2019). Fast detection of cocoa shell in cocoa powders by near infrared spectroscopy and multivariate analysis. *Food Control*, *99*, 68–72.

Quelal-Vásconez, M. A., Pérez-Esteve, É., Arnau-Bonachera, A., Barat, J. M., &

Talens, P. (2018). Rapid fraud detection of cocoa powder with carob flour using near infrared spectroscopy. *Food Control*, *92*, 183–189.

Scollo, E., Neville, D. C. A., Oruna-Concha, M. J., Trotin, M., & Cramer, R. (2020). UHPLC–MS/MS analysis of cocoa bean proteomes from four different genotypes. *Food Chemistry*, *303*, 125244.

Sendin, K., Manley, M., Baeten, V., Fernández Pierna, J. A., & Williams, P. J. (2019). Near Infrared Hyperspectral Imaging for White Maize Classification According to Grading Regulations. *Food Analytical Methods*, *12*(7), 1612–1624.

Su, W.-H., Bakalis, S., & Sun, D.-W. (2019). Chemometrics in tandem with near infrared (NIR) hyperspectral imaging and Fourier transform mid infrared (FT-MIR) microspectroscopy for variety identification and cooking loss determination of sweet potato. *Biosystems Engineering*, *180*, 70–86.

Sunoj, S., Igathinathane, C., & Visvanathan, R. (2016). Nondestructive determination of cocoa bean quality using FT-NIR spectroscopy. *Computers and Electronics in Agriculture*, *124*, 234–242.

Teye, E., Huang, X., Dai, H., & Chen, Q. (2013). Rapid differentiation of Ghana cocoa beans by FT-NIR spectroscopy coupled with multivariate classification. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *114*, 183–189.

Teye, E., Huang, X., Sam-Amoah, L. K., Takrama, J., Boison, D., Botchway, F., & Kumi, F. (2015). Estimating cocoa bean parameters by FT-NIRS and chemometrics analysis. *Food Chemistry*, *176*, 403–410.

The International Cocoa Organization. (2018). *Cocoa producing and cocoa consuming*

countries (p. Retrieved February 1, 2018, from https://www.icco.). p. Retrieved February 1, 2018, from https://www.icco.

Trognitz, B., Cros, E., Assemat, S., Davrieux, F., Forestier-Chiron, N., Ayestas, E., … Hermann, M. (2013). Diversity of Cacao Trees in Waslala, Nicaragua: Associations between Genotype Spectra, Product Quality and Yield Potential. *PLOS ONE*, *8*(1), e54079.

UNCTAD/WTO., I. T. C. (2001). *Cocoa: a guide to trade practices*. United Nations Publications.

Vargas Jentzsch, P., Ciobotă, V., Salinas, W., Kampe, B., Aponte, P. M., Rösch, P., … Ramos, L. A. (2016). Distinction of Ecuadorian varieties of fermented cocoa beans using Raman spectroscopy. *Food Chemistry*, *211*, 274–280.

Veselá, A., Barros, A. S., Synytsya, A., Delgadillo, I., Čopíková, J., & Coimbra, M. A. (2007). Infrared spectroscopy and outer product analysis for quantification of fat, nitrogen, and moisture of cocoa powder. *Analytica Chimica Acta*, *601*(1), 77–86.

Vidal, M., & Amigo, J. M. (2012). Pre-processing of hyperspectral images. Essential steps before image analysis. *Chemometrics and Intelligent Laboratory Systems*, *117*, 138–148.

Wang, J., Zhang, X., Sun, S., Sun, X., Li, Q., & Zhang, Z. (2018). Online determination of quality parameters of dried soybean protein–lipid films (Fuzhu) by NIR spectroscopy combined with chemometrics. *Journal of Food Measurement and Characterization*, *12*(3), 1473–1484.

ANEXO 1

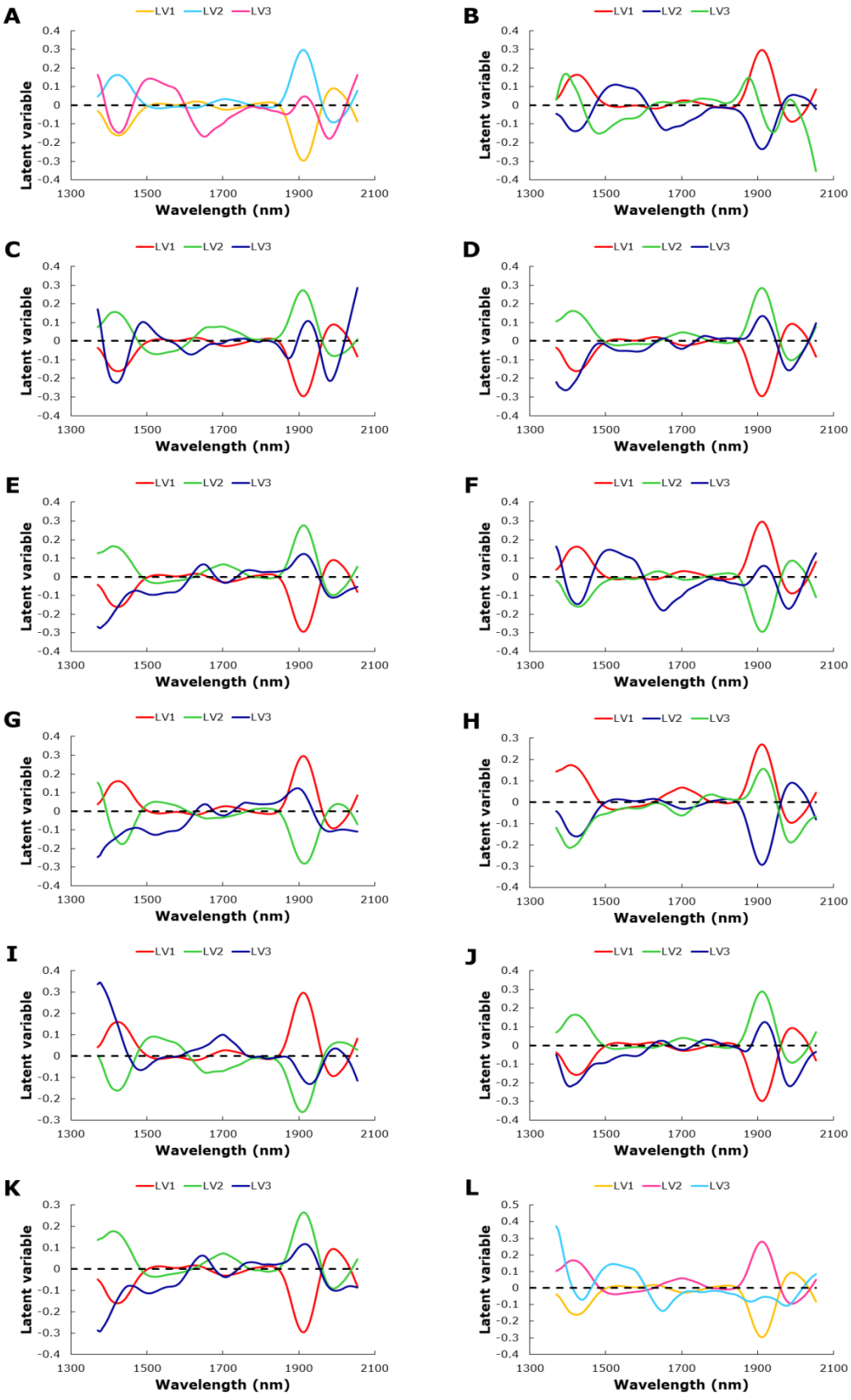

A
B
C
D
E
F
G
H
I
J
K
L
LV1
LV2
LV3
Latent variable
Wavelength (nm)
1300
1500
1700
1900
2100

Figure 1. Loading plot of PLS-DA models. A) IMC *vs* P7; B) H1 *vs* H2; C) H2 *vs* H3; D) H2 *vs* H4; E) H2 *vs* H5; F) H1 *vs* H3; G) H1 *vs* H4; H) H1 *vs* H5; I) H3 *vs* H4; J) H3 *vs* H5; K) H4 *vs* H5; L) 5-classes model

CHAPTER 5:

GENERAL DISCUSSION

5.1 General discussion

A hyperspectral imaging system (HSI) was implemented to ensure quality control of agricultural products in two different applications: (1) estimate the shelf life of chia seeds and (2) identify the variety of cocoa bean hybrid.

In the first case, HSI and PCA were used to develop a methodology to estimate the shelf life of chia seeds. The PC scores allowed modeling the kinetics of chia seed degradation, which obeyed two-stage kinetics: exponential and linear (R^2> 0.85). In addition, the increase in temperature accelerated the degradation of chia seeds, with an increase in acidity and the degradation of fatty acids. Using the cut-off value of the PC1 score when the seeds showed an increase in acidity of 75%, it was possible to estimate the useful life of chia seeds in approximately 1300 days at 25 ° C. Furthermore, it was possible to create a "Re-sampling" strategy that allowed validating the methodology, by projecting the validation samples on the calibration set with an acceptable number of iterations. Thus, it was possible to prove that the pixels within each image were being correctly predicted.

In the second case, HSI showed a high performance to discriminate between 2-classes of hybrids and 5-classes of hybrids, both for PLS-DA and SVM models, with results comparable to those obtained by polymerase chain reaction. The pixel-to-pixel prediction showed the high predictability of the PLS-DA models to identify cocoa bean hybrids, in an external set. Therefore, the methodology developed here could be plausibly implemented in cocoa bean production centers.

CHAPTER 6:

CONCLUSION AND FUTURE TRENDS

6.1 General conclusion

Generally speaking, hyperspectral images in combination with multivariate analysis allow, in a reliable way, an aid in the quality control of agricultural products. In this specific case, we can conclude that HSI can be implemented to estimate the shelf life of chia seeds at commercial temperatures, and in the identification of cocoa bean hybrids in Brazil.

6.1. Future trends

- The MASLT method should be expanded to other types of products, both fresh and processed products. In addition, the cut-off value to estimate shelf-life must be selected based on the end use of the product.

- Re-sampling method can be used to validate analysis methods using PCA or the information from the image itself (spectral pixel value and/or texture features).

- New varieties of cocoa bean hybrids produced in different locations in Brazil must be incorporated into the model to ensure the traceability of the products.

- Other machine learning methodologies such as Random Forest or Neural networks should be tested to reduce the error in the classification of the different cocoa bean hybrids.

- Variable selection methodology should be evaluated to reduce the error in the classification of cocoa bean hybrids, to reduce the calculation time of the models and to facilitate the construction of lower cost equipment.

CHAPTER 6:

REFERENCES

Amigo, J. M. (2010). Practical issues of hyperspectral imaging analysis of solid dosage forms. Analytical and Bioanalytical Chemistry, 398(1), 93–109.

Amigo, J. M., Babamoradi, H., & Elcoroaristizabal, S. (2015). Hyperspectral image analysis. A tutorial. Analytica Chimica Acta, 896, 34–51.

Aprotosoaie, A. C., Luca, S. V., & Miron, A. (2016). Flavor Chemistry of Cocoa and Cocoa Products—An Overview. Comprehensive Reviews in Food Science and Food Safety, 15(1), 73–91.

Barbin, D. F., Maciel, L. F., Bazoni, C. H. V., Ribeiro, M. da S., Carvalho, R. D. S., Bispo, E. da S., … Hirooka, E. Y. (2018). Classification and compositional characterization of different varieties of cocoa beans by near infrared spectroscopy and multivariate statistical analyses. Journal of Food Science and Technology, 55(7), 2457–2466.

Behrend, C. J., Tarnowski, C. P., & Morris, M. D. (2002). Identification of Outliers in Hyperspectral Raman Image Data by Nearest Neighbor Comparison. Applied Spectroscopy, 56(11), 1458–1461.

Bosmans, G. M., Lagrain, B., Ooms, N., Fierens, E., & Delcour, J. A. (2014). Storage of parbaked bread affects shelf life of fully baked end product: A 1H NMR study. Food Chemistry, 165, 149–156.

Burger, J., & Geladi, P. (2005). Hyperspectral NIR image regression part I: calibration and correction. Journal of Chemometrics, 19(5-7), 355–363.

Burger, J., & Geladi, P. (2007). Spectral Pre-Treatments of Hyperspectral near Infrared Images: Analysis of Diffuse Reflectance Scattering. Journal of Near Infrared Spectroscopy, 15(1), 29–37.

Caligiani, A., Palla, L., Acquotti, D., Marseglia, A., & Palla, G. (2014). Application of 1H NMR for the characterisation of cocoa beans of different geographical origins and fermentation levels. Food Chemistry, 157, 94–99.

Calvini, R., Amigo, J. M., & Ulrici, A. (2017). Transferring results from NIR-hyperspectral to NIR-multispectral imaging systems: A filter-based simulation applied to the classification of Arabica and Robusta green coffee. Analytica Chimica Acta, 967, 33–41.

Cannistraci, C. V, Montevecchi, F. M., & Alessio, M. (2009). Median-modified Wiener filter provides efficient denoising, preserving spot edge and morphology in 2-DE image processing. PROTEOMICS, 9(21), 4908–4919.

Chaudhry, M. M. A., Amodio, M. L., Babellahi, F., de Chiara, M. L. V, Amigo Rubio, J. M., & Colelli, G. (2018). Hyperspectral imaging and multivariate accelerated shelf life testing (MASLT) approach for determining shelf life of rocket leaves. Journal of Food Engineering, 238, 122–133.

Coelho, S. R. M., Alves Filho, E. G., Silva, L. M. A., Bischoff, T. Z., Ribeiro, P. R. V, Zocolo, G. J., … de Brito, E. S. (2020). NMR and LC-MS assessment of compound variability of common bean (Phaseolus vulgaris) stored under controlled atmosphere. LWT, 117, 108673.

Cortés, V., Blasco, J., Aleixos, N., Cubero, S., & Talens, P. (2019). Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review. Trends in Food Science & Technology, 85, 138–148.

de Falco, B., Amato, M., & Lanzotti, V. (2017). Chia seeds products: an overview. Phytochemistry Reviews, 16(4), 745–760.

Di Egidio, V., Sinelli, N., Limbo, S., Torri, L., Franzetti, L., & Casiraghi, E. (2009). Evaluation of shelf-life of fresh-cut pineapple using FT-NIR and FT-IR spectroscopy. Postharvest Biology and Technology, 54(2), 87–92.

Diomande, D., Antheaume, I., Leroux, M., Lalande, J., Balayssac, S., Remaud, G. S., & Tea, I. (2015). Multi-element, multi-compound isotope profiling as a means to distinguish the geographical and varietal origin of fermented cocoa (Theobroma cacao L.) beans. Food Chemistry, 188, 576–582.

Efraim, P., Pires, J. L., Garcia, A. de O., Grimaldi, R., Luccas, V., & Pezoa-Garcia, N. H. (2013). Characteristics of cocoa butter and chocolates obtained from cocoa varieties grown in Bahia, Brazil. European Food Research and Technology, 237(3), 419–428.

Ehrentreich, F., & Summchen, L. (2001). Spike removal and denoising of Raman spectra by wavelet transform methods. Analytical Chemistry, 73(17), 4364–4373.

Fernández Pierna, J. A., Vermeulen, P., Eylenbosch, D., Burger, J., Bodson, B., Dardenne, P., & Baeten Molecular Sciences and Chemical Engineering, V. B. T.-R. M. in C. (2020). Chemometrics in NIR Hyperspectral Imaging: Theory and Applications in the Agricultural Crops and Products Sector☆.

Ferreira, M. M. C. (2015). Quimiometria: conceitos, métodos e aplicações. Editora da Unicamp.

Feuerstein, D., Parker, K. H., & Boutelle, M. G. (2009). Practical Methods for Noise Removal: Applications to Spikes, Nonstationary Quasi-Periodic Noise, and Baseline Drift. Analytical Chemistry, 81(12), 4987–4994.

Firtha, F., Fekete, A., Kaszab, T., Gillay, B., Nogula-Nagy, M., Kovács, Z., & Kantor, D. B. (2008). Methods for improving image quality and reducing data load of NIR hyperspectral images. Sensors, 8(5), 3287–3298.

Franklin, L. M., Chapman, D. M., King, E. S., Mau, M., Huang, G., & Mitchell, A. E. (2017). Chemical and Sensory Characterization of Oxidative Changes in Roasted Almonds Undergoing Accelerated Shelf Life. Journal of Agricultural and Food Chemistry, 65(12), 2549–2563.

Giovenzana, V., Beghi, R., Buratti, S., Civelli, R., & Guidetti, R. (2014). Monitoring of fresh-cut Valerianella locusta Laterr. shelf life by electronic nose and VIS–NIR spectroscopy. Talanta, 120, 368–375.

Grelet, C., Froidmont, E., Foldager, L., Salavati, M., Hostens, M., Ferris, C. P., … Dehareng, F. (2020). Potential of milk mid-infrared spectra to predict nitrogen use efficiency of individual dairy cows in early lactation. Journal of Dairy Science, 103(5), 4435–4445.

Hashim, N., Janius, R. Bin, Baranyai, L., Rahman, R. A., Osman, A., & Zude, M. (2012). Kinetic Model for Colour Changes in Bananas During the Appearance of Chilling Injury Symptoms. Food and Bioprocess Technology, 5(8), 2952–2963.

Herrmann, L., Felbinger, C., Haase, I., Rudolph, B., Biermann, B., & Fischer, M. (2015). Food Fingerprinting: Characterization of the Ecuadorean Type CCN-51 of Theobroma cacao L. Using Microsatellite Markers. Journal of Agricultural and Food Chemistry, 63(18), 4539–4544.

Hussain, N., Sun, D.-W., & Pu, H. (2019). Classical and emerging non-destructive technologies for safety and quality evaluation of cereals: A review of recent applications. Trends in Food Science & Technology, 91, 598–608.

Ivorra, E., Girón, J., Sánchez, A. J., Verdú, S., Barat, J. M., & Grau, R. (2013). Detection of expired vacuum-packed smoked salmon based on PLS-DA method using hyperspectral images. Journal of Food Engineering, 117(3), 342–349.

Ji, K., Zhang, D., Motilal, L. A., Boccara, M., Lachenaud, P., & Meinhardt, L. W. (2013). Genetic diversity and parentage in farmer varieties of cacao (Theobroma cacao L.) from Honduras and Nicaragua as revealed by single nucleotide polymorphism (SNP) markers. Genetic Resources and Crop Evolution, 60(2), 441–453.

Jia, B., Wang, W., Ni, X., Lawrence, K. C., Zhuang, H., Yoon, S.-C., & Gao, Z. (2020). Essential processing methods of hyperspectral images of agricultural and food products. Chemometrics and Intelligent Laboratory Systems, 198, 103936.

Kamruzzaman, M., Makino, Y., & Oshita, S. (2016). Parsimonious model development for real-time monitoring of moisture in red meat using hyperspectral imaging. Food Chemistry, 196, 1084–1091.

Lavine, B. K., & Davidson, C. E. (2006). Classification and pattern recognition. Practical Guide to Chemometrics, 339–377.

Lee, K. W., Kim, Y. J., Lee, H. J., & Lee, C. Y. (2003). Cocoa has more phenolic phytochemicals and a higher antioxidant capacity than teas and red wine. Journal of Agricultural and Food Chemistry, 51(25), 7292–7295.

Li, J., Rao, X., & Ying, Y. (2011). Detection of common defects on oranges using hyperspectral reflectance imaging. Computers and Electronics in Agriculture, 78(1), 38–48.

Li, X., Zhu, W., Ji, B., & Liu, B. (2010). Weed identification based on features optimization and LS-SVM in the cotton field. Nongye Jixie Xuebao= Transactions of the Chinese Society for Agricultural Machinery, 41(11), 168–172.

Liu, F., He, Y., & Wang, L. (2008). Comparison of calibrations for the determination of soluble solids content and pH of rice vinegars using visible and short-wave near infrared spectroscopy. Analytica Chimica Acta, 610(2), 196–204.

Ma, J., & Sun, D.-W. (2020). Prediction of monounsaturated and polyunsaturated fatty acids of various processed pork meats using improved hyperspectral imaging technique. Food Chemistry, 321, 126695.

Ma, L., Zhang, M., Bhandari, B., & Gao, Z. (2017). Recent developments in novel shelf life extension technologies of fresh-cut fruits and vegetables. Trends in Food Science & Technology, 64, 23–38.

Marineli, R. da S., Moraes, É. A., Lenquiste, S. A., Godoy, A. T., Eberlin, M. N., & Maróstica Jr, M. R. (2014). Chemical characterization and antioxidant potential of Chilean chia seeds and oil (Salvia hispanica L.). LWT - Food Science and Technology, 59(2, Part 2), 1304–1310.

Mata, T. M., Correia, D., Pinto, A., Andrade, S., Trovisco, I., Matos, E., … Caetano, N. S. (2017). Fish oil acidity reduction by enzymatic esterification. Energy Procedia, 136, 474–480.

Meloun, M., Forina, M., & Militky, J. (1992). Chememetrics for Analyhtical Chemistry: PC-Aided Statistical Data Analysis. Prentice Hall Professional Technical Reference.

Mendez, J., Mendoza, L., Cruz-Tirado, J. P., Quevedo, R., & Siche, R. (2019). Trends in application of NIR and hyperspectral imaging for food authentication . Scientia Agropecuaria , Vol. 10, pp. 143–161. scielo .

Menesatti, P., Zanella, A., D'Andrea, S., Costa, C., Paglia, G., & Pallottino, F. (2009). Supervised Multivariate Analysis of Hyper-spectral NIR Images to Evaluate the Starch Index of Apples. Food and Bioprocess Technology, 2(3), 308–314.

Mite-Baidal, K., Solís-Avilés, E., Martínez-Carriel, T., Marcillo-Plaza, A., Cruz-Ibarra, E., & Baque-Bustamante, W. (2019). Analysis of Computer Vision Algorithms to Determine the Quality of Fermented Cocoa (Theobroma Cacao): Systematic Literature Review BT - ICT for Agriculture and Environment (R. Valencia-García, G. Alcaraz-Mármol, J. del Cioppo-Morstadt, N. Vera-Lucio, & M. Bucaram-Leverone, Eds.). Cham: Springer International Publishing.

Mobaraki, N., & Amigo, J. M. (2018). HYPER-Tools. A graphical user-friendly interface for hyperspectral image analysis. Chemometrics and Intelligent Laboratory Systems, 172, 174–187.

Moschopoulou, E., Moatsou, G., Syrokou, M. K., Paramithiotis, S., & Drosinos, E. H. (2019). 1 - Food quality changes during shelf life (C. M. B. T.-F. Q. and S. L. Galanakis, Ed.).

Motilal, L., & Butler, D. (2003). Verification of identities in global cacao germplasm collections. Genetic Resources and Crop Evolution, 50(8), 799–807.

Munir, M. T., Wilson, D. I., Yu, W., & Young, B. R. (2018). An evaluation of hyperspectral imaging for characterising milk powders. Journal of Food Engineering, 221, 1–10.

Nenadic, Z., & Burdick, J. W. (2005). Spike detection using the continuous wavelet transform. IEEE Transactions on Biomedical Engineering, 52(1), 74–87.

Noviyanto, A., & Abdulla, W. H. (2019). Segmentation and calibration of hyperspectral imaging for honey analysis. Computers and Electronics in Agriculture, 159, 129–139.

Nzekoue, F. K., Caprioli, G., Fiorini, D., Torregiani, E., Vittori, S., & Sagratini, G. (2019). HS-SPME-GC-MS technique for FFA and hexanal analysis in different cheese packaging in the course of long term storage. Food Research International, 121, 730–737.

Oliveira, M. M., Cruz-Tirado, J. P., & Barbin, D. F. (2019). Nontargeted Analytical Methods as a Powerful Tool for the Authentication of Spices and Herbs: A Review. Comprehensive Reviews in Food Science and Food Safety, 18(3), 670–689.

Orrillo, I., Cruz-Tirado, J. P., Cardenas, A., Oruna, M., Carnero, A., Barbin, D. F., & Siche, R. (2019). Hyperspectral imaging as a powerful tool for identification of papaya seeds in black pepper. Food Control, 101, 45–52.

Pedro, A. M. K., & Ferreira, M. M. C. (2006). Multivariate accelerated shelf-life testing: a novel approach for determining the shelf-life of foods. Journal of Chemometrics, 20(1-2), 76–83.

Pérez-Marín, D., Calero, L., Fearn, T., Torres, I., Garrido-Varo, A., & Sánchez, M.-T. (2019). A system using in situ NIRS sensors for the detection of product failing to meet quality standards and the prediction of optimal postharvest shelf-life in the case of oranges kept in cold storage. Postharvest Biology and Technology, 147, 48–53.

Poudyal, H., Panchal, S. K., Waanders, J., Ward, L., & Brown, L. (2012). Lipid redistribution by α-linolenic acid-rich chia seed inhibits stearoyl-CoA desaturase-1 and

induces cardiac and hepatic protection in diet-induced obese rats. The Journal of Nutritional Biochemistry, 23(2), 153–162.

Rehman, T. U., Mahmud, M. S., Chang, Y. K., Jin, J., & Shin, J. (2019). Current and future applications of statistical machine learning algorithms for agricultural machine vision systems. Computers and Electronics in Agriculture, 156, 585–605.

Rinnan, Å., Berg, F. van den, & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends in Analytical Chemistry, 28(10), 1201–1222.

Scollo, E., Neville, D. C. A., Oruna-Concha, M. J., Trotin, M., & Cramer, R. (2020). UHPLC–MS/MS analysis of cocoa bean proteomes from four different genotypes. Food Chemistry, 303, 125244.

Siripatrawan, U., & Makino, Y. (2018). Simultaneous assessment of various quality attributes and shelf life of packaged bratwurst using hyperspectral imaging. Meat Science, 146, 26–33.

Sjöström, M., Wold, S., & Söderström, B. (1986). PLS DISCRIMINANT PLOTS (E. S. GELSEMA & L. N. B. T.-P. R. in P. KANAL, Eds.).

Song, Y., Hu, Q., Wu, Y., Pei, F., Kimatu, B. M., Su, A., & Yang, W. (2019). Storage time assessment and shelf-life prediction models for postharvest Agaricus bisporus. LWT, 101, 360–365.

Taghizadeh, M., Gowen, A., Ward, P., & O'Donnell, C. P. (2010). Use of hyperspectral imaging for evaluation of the shelf-life of fresh white button mushrooms (Agaricus bisporus) stored in different packaging films. Innovative Food Science & Emerging Technologies, 11(3), 423–431.

Taheri-Garavand, A., Fatahi, S., Omid, M., & Makino, Y. (2019). Meat quality evaluation based on computer vision technique: A review. Meat Science, 156, 183–195.

Teye, E., Anyidoho, E., Agbemafle, R., Sam-Amoah, L. K., & Elliott, C. (2020). Cocoa bean and cocoa bean products quality evaluation by NIR spectroscopy and chemometrics: A review. Infrared Physics & Technology, 104, 103127.

Tsironi, T. N., Ntzimani, A. G., & Taoukis, P. S. B. T.-R. M. in F. S. (2019). Modified Atmosphere Packaging and the Shelf Life of Meat.

Tsouvaltzis, P., Babellahi, F., Amodio, M. L., & Colelli, G. (2020). Early detection of eggplant fruit stored at chilling temperature using different non-destructive optical techniques and supervised classification algorithms. Postharvest Biology and Technology, 159, 111001.

Vargas Jentzsch, P., Ciobotă, V., Salinas, W., Kampe, B., Aponte, P. M., Rösch, P., … Ramos, L. A. (2016). Distinction of Ecuadorian varieties of fermented cocoa beans using Raman spectroscopy. Food Chemistry, 211, 274–280.

Velásquez, L., Cruz-Tirado, J. P., Siche, R., & Quevedo, R. (2017). An application based on the decision tree to classify the marbling of beef by hyperspectral imaging. Meat Science, 133.

Vermeulen, P., Suman, M., Fernández Pierna, J. A., & Baeten, V. (2018). Discrimination between durum and common wheat kernels using near infrared hyperspectral imaging. Journal of Cereal Science, 84, 74–82.

Vidal, M., & Amigo, J. M. (2012). Pre-processing of hyperspectral images. Essential steps before image analysis. Chemometrics and Intelligent Laboratory Systems, 117, 138–148.

Wibowo, S., Buvé, C., Hendrickx, M., Van Loey, A., & Grauwet, T. (2018). Integrated science-based approach to study quality changes of shelf-stable food products during storage: A proof of concept on orange and mango juices. Trends in Food Science & Technology, 73, 76–86.

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems, 58(2), 109–130.

Wu, P., Chen, C., Ding, J., Hsu, C., & Huang, Y. (2013). Salient Region Detection Improved by Principle Component Analysis and Boundary Information. IEEE Transactions on Image Processing, 22(9), 3614–3624.

Yousuf, B., Qadri, O. S., & Srivastava, A. K. (2018). Recent developments in shelf-life extension of fresh-cut fruits and vegetables by application of different edible coatings: A review. LWT, 89, 198–209.

Zhang, L., & Henson, M. J. (2007). A Practical Algorithm to Remove Cosmic Spikes in Raman Imaging Data for Pharmaceutical Applications. Applied Spectroscopy, 61(9), 1015–1020.

ANEXO

Copyright permission to reproduce the article

Copyright

Personal use

Authors can use their articles, in full or in part, for a wide range of scholarly, non-commercial purposes as outlined below:

- Use by an author in the author's classroom teaching (including distribution of copies, paper or electronic)
- Distribution of copies (including through e-mail) to known research colleagues for their personal use (but not for Commercial Use)
- **Inclusion in a thesis or dissertation (provided that this is not to be published commercially)**
- Use in a subsequent compilation of the author's works
- Extending the Article to book-length form
- Preparation of other derivative works (but not for Commercial Use)
- Otherwise using or re-using portions or excerpts in other works

These rights apply for all Elsevier authors who publish their article as either a subscription article or an open access article. In all cases we require that all Elsevier authors always include a full acknowledgement and, if appropriate, a link to the final published version hosted on Science Direct.

Authentication of cocoa (Theobroma cacao) bean hybrids by NIR-hyperspectral imaging and chemometrics

Author:
J.P. Cruz-Tirado,Juan Antonio Fernández Pierna,Hervé Rogez,Douglas Fernandes Barbin,Vincent Baeten

Publication: Food Control

Publisher: Elsevier

Date: December 2020