XX Simposio CEA de Control Inteligente
25-27 de junio de 2025, Huelva

Imagen Hiperespectral en Campo para la Estimación del Contenido Graso sobre Seco en Aceitunas: Investigación en curso

Vaz, F.*, Argüello, D., Noguera, M., Aquino, A.

Centro de Investigación en Tecnología, Energía y Sostenibilidad, Carretera Huelva-Palos, 21819, La Rábida, Huelva, España.

To cite this article: Vaz, F.*, Argüello, D., Noguera, M., Aquino, A. 2025. In-Field Hyperspectral Imaging for Estimation of Olive Fruit Fat Content per Dry Matter: Ongoing Research. XX Simposio CEA de Control Inteligente

Resumen

La evaluación precisa y no destructiva del contenido graso de aceitunas en condiciones de campo es crucial para optimizar los programas de cosecha, mejorar la calidad del aceite, y avanzar en la introducción de la agricultura de precisión en los olivares. Este trabajo investiga el potencial de las imágenes hiperespectrales, adquiridas directamente en un olivar comercial en condiciones de iluminación natural, para la estimación del contenido graso de aceitunas en fase de maduración visibles en la superficie de los olivos. La investigación en curso se centra en el desarrollo de modelos de estimación capaces de, a partir de los perfiles espectrales de las aceitunas muestreadas, estimar con precisión su contenido graso sobre materia seca. Las primeras aproximaciones empleando Regresión de Mínimos Cuadrados Parciales (PLSR) arrojaron resultados prometedores considerando la limitada capacidad de este enfoque de modelado, lo que induce a la exploración del desempeño de modelos no lineales, más complejos y con mayor potencial descriptivo.

Palabras clave: agricultura de precisión, análisis de imágenes en agricultura, calidad de alimentos, sensores de software en agricultura, evaluación de la calidad.

In-Field Hyperspectral Imaging for Estimation of Olive Fruit Fat Content per Dry Matter: Ongoing Research

Abstract

Accurate and non-destructive assessment of olive fruit fat content in field conditions is crucial for optimising harvest programs, improving oil quality, and advancing the introduction of precision agriculture in olive groves. This work investigates the potential of hyperspectral imaging, acquired directly in a commercial olive grove under natural lighting conditions, for estimating the fat content of ripening olives visible on the surface of the trees. The ongoing research focuses on developing estimation models capable of accurately predicting the olives' fat content per dry matter from their sampled spectral profiles. Initial approaches using Partial Least Squares Regression (PLSR) yielded promising results considering the limited capability of this modelling approach, which leads to exploring the performance of more complex, nonlinear models with greater descriptive potential.

Keywords: precision farming, image analysis in agriculture, food quality, software sensors in agriculture, quality assessment.

1. Introducción

The olive tree (*Olea europaea L.*) plays a significant role in the Mediterranean economy, constituting a fundamental part of the global olive oil industry. This Mediterranean staple, valued for its gastronomic and nutraceutical properties, supports numerous industries. Precision agriculture is innovatively taking shape for optimal management in olive groves, and especially in high density orchards to make production profitable. This would involve modifying the way farming is done to account for the spatial and temporal variability that lurks inside agricultural fields, and thus it would require more advanced tools to monitor important parameters.

The accurate measurement of crucial agricultural indicators is increasingly being recognised as important in the field of precision agriculture, including fruit ripening. Ripening is defined as a complex series of physical and biochemical changes, among which oil accumulation is a key indicator of

*Autor para correspondencia: federico.vaz@pi.uhu.es

maturity, thus being indicative of the optimum time for harvesting (Hermoso *et al*., 1991).

Conventional chemical methods for determining fat content, such as Soxhlet extraction (AENC, 1961), are costly and impractical for large-scale fieldwork. Consequently, olive growers typically assess ripeness at a general level by observing the olives' external colour and subjectively comparing it against a defined grading scale. Recent investigations, however, have revealed a mismatch between fat content and external colour, thereby discrediting colour as an accurate and reliable indicator of ripeness. [Grilo *et al*., 2019].

This report describes the preliminary findings of ongoing experiments aimed at developing a methodology, based on hyper spectral imaging, for the objective, accurate, and non-destructive evaluation of olive fruit fat content per dry matter straight in the field.

2. Materials and methods

2.1 Experimental layout

The study was conducted in a commercial olive orchard (*Olea europaea L., cv. Picual*) managed by Nuestra Señora de la Oliva S.C.A. in Huelva, Andalusia, Spain (37°20'10.986" N, 7°1'49.46" W). Sampling occurred over 7 weeks in September and October, coinciding with the end of pit hardening and the harvest season.

Fifteen trees were selected for the study. During each sampling session throughout the lifetime of the experiment, olive fruits visible within a 60 x 40 cm wooden frame placed over the canopy of the study trees (Figure 1) were photographed to acquire hyperspectral information. Immediately after photographing, the same sets of olives within the frame were carefully harvested, labelled, and transported to the laboratory for chemical determination of fat content per dry matter for each sample. This procedure resulted in 105 paired sampling points of hyperspectral and laboratory reference data.

Figure 1: Sampling point wood frame.

2.2 Materials

A Headwall Co-Aligned VNIR – SWIR push broom-type hyperspectral imaging system (Headwall Photonics Inc., Bolton, U.S.A.) was employed for in-field hyperspectral data acquisition. The VNIR sensor captures 273 spectral bands spanning the range of 399-1003 nm, with a spectral sampling interval ranging between 2 and 3 nm. The SWIR sensor captures 270 spectral bands spanning the range of 899-2509 nm, with a fixed spectral sampling interval of 6 nm. The device utilised in this study transforms each spatial image into a three-dimensional hyperspectral data cube. The dimensions of this cube are defined by m pixels in height, which varies depending on the displacement across the scanning direction, and 640 pixels in width, with a spectral dimension comprising 273 bands for VNIR, and 270 bands for SWIR.

A purpose-built acquisition device, incorporating a tripod with a precision rotation head and a stepper motor for controlled scanning speed, was developed for the present study. The system (Figure 2) was equipped with a height adjustment mechanism, enabling a range of 1.3 to 2 m, to ensure coplanar alignment between the camera and the wooden frame that delineated the sampling area. The stepper motor regulated a constant scanning speed, which was determined by the Headwall’s software based on illumination conditions and camera-to-target distance. This was essential for the push broom operation, as it guaranteed the acquisition of spatially accurate hyperspectral images of the targeted fruit clusters within the defined frame. A high-reflectivity fabric was placed within the scanning field of view for subsequent radiometric correction and reflectance normalisation.

Figure 2: Figure 2: Designed scanning device, consisting of a tripod with a rotating head to enable push-broom operation with the hyperspectral camera.

Headwall's Hyperspec III and SpectralView software packages were used to perform the initial transduction of data from the Headwall hyperspectral imager to a digital processing workstation. Radiometric correction and reflectance normalisation were implemented to reduce sensor-induced artefacts. MATLAB, Version R2024b, was subsequently employed for processing and analysis of the corrected hyperspectral information, and to explore the descriptive potential of the data in terms of fruit fat content estimation.

2.2 Experimental data processing and analysing

Based on the specialised literature, the SWIR spectrum was selected for this preliminary study, from which those bands associated to relevant water absorption peaks (1354-1458 nm, 1797-2090 nm) and low-information wavelengths (>2227 nm) were discarded (Saha & Jackson, 2018; Yang *et al.*, 2005).

First, a region of interest (ROI) consisting of olive fruit pixels within the wooden sampling frame shown in Figure 1 was defined from radiometrically corrected images. Then, for each of the 105 sampling points, the average spectral response of ROI pixels was calculated for the 147 selected bands, which resulted in a 105 × 147 feature matrix. An outlier analysis finally discarded 9 sample points, thus definitely configuring a 96 × 147 feature matrix.

2.5 Methodology for Olive Fruit Fat Content Prediction

The first derivative was calculated on the 147 features for the 96 samples in the feature matrix by applying Savitzky-Golay filtering. The transformed feature vector per sample describes and highlights subtle spectral reflectance variations, potentially due to biochemical variations in the fruits.

Next, Partial Least Squares Regression (PLSR) was employed to preliminarily explore the potential of the processed samples' spectral profiles to quantitatively estimate fat content. To establish statistically coherent training and validation datasets for modelling, the 96 feature vectors were first clustered using the K-Nearest Neighbours (KNN) algorithm, setting the reference fat content per dry matter as the clustering variable. The training set was built with 85% of samples in the dataset taken from the 10 generated clusters according to clusters' population ratios. The validation dataset was compiled with the remaining 15% of samples.

The optimal number of PLSR components for the predictive model was determined through an iterative process, in which estimation results for a specified wide range of components were monitored. The root-mean-square error (RMSE) and the coefficient of determination (R^2) were employed as indicators to assess model performance.

3. Results and discussion

Figure 3 shows the evolution of R^2 for the training and external validation sets as the number of latent variables included in the PLSR model increases. As it can be confirmed, model's regression performance on the validation set increases until reaching a maximum of 0.64 when 6 latent variables are considered. From this point, latent variables increasing involve accelerated modelling degradation, what might indicate the existence of complex relationships among the spectral responses when they are combined to model fat content of fruits. This, along with the respectable R^2 value achieved by this simple linear modelling approach, considering the complexity of the formulated regression task, encourages to explore more sophisticated and non-linear modelling frameworks, such as neural networks. Moreover, the addition of the second derivative for enriching the model with even more subtle transitions, might also have a positive impact in modelling performance. These conclusions are consistent to those derivable by analysing the model performance in terms of estimation capabilities, as RMSE values of 3.39% and 4% (both percentage of fat content per dry) were measured for the training and validation datasets, respectively, when the optimum PLSR configuration with 6 latent variables was employed. Table 1 summarises the best achieved results.

Table 1: Best PLSR modelling performance records (6 latent variables)

Training set		Test set	
R^2	RMSE (%)	R^2	RMSE (%)
0.76	3.39	0.64	4

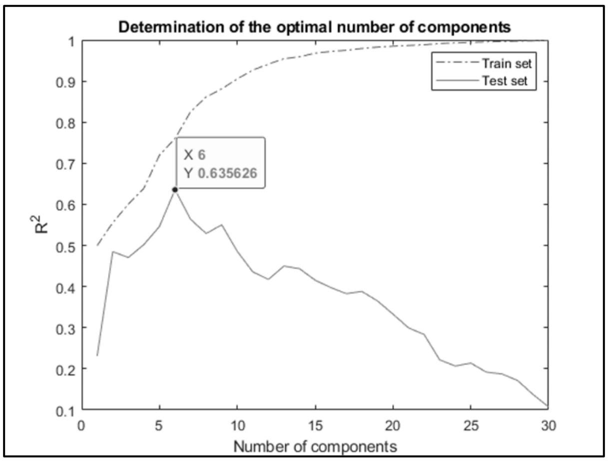

Figure 3: results of PLSR in function of the number of components selected

4. Conclusions

The intrinsic complexity of the relationship between the spectral signatures and the biochemical composition of the olive fruit — more especially the fat content—is probably the reason for the PLSR limited, although non negligible, estimation results. As a linear regression method, PLSR relies on the predictor and response variables having a high linear correlation. However, non-linearities in the spectral response linked to fat content may result from the complex interactions of different light absorbing and scattering components within the fruit matrix. In this context, the respectable R^2 and RMSE values registered with PLSR, encourages to continue this investigation considering more complex non-linear modelling approaches.

Acknowledgment

This work was supported by the Cooperation Program VI-A SPAINPORTUGAL (POCTEP) 2021–2027 and co-financed with ERDF [0067_OLIVAR_IA_5_E]; and the Research Consolidation Programme 2022 of the Ministry of Science and Innovation (Spain) [CNS2022–136137].

References

Asociación Espanola de Normalización y Certificación (AENC), 1961. UNE 55030:1961, determination of the content in total fat of olives, España. https://www.une. org/encuentra-tu-norma/busca-tu-norma/norma?c=N0005670 (accessed June 1, 2025).

Grilo, F., Caruso, T., Wang, S.C., 2019. Influence of fruit canopy position and maturity on yield determinants and chemical composition of virgin olive oil. Journal of the Science of Food and Agriculture 99, 4319–4330. DOI: https://doi.org/10.1002/JSFA.9665.

Hermoso, M., Uceda, M., García-Ortiz, A., Morales, J., Frías, L., Fernández, A., 1991. Elaboración de aceite de oliva de calidad. Colección Apunt 5 91.

Malavi, D., Nikkhah, A., Raes, K., Van Haute. S., 2023. Hyperspectral Imaging and Chemometrics for Authentication of Extra Virgin Olive Oil: A Comparative Approach with FTIR, UV-VIS, Raman, and GC-MS. Foods 12(3). DOI: 10.3390/foods12030429.

Saha, U., Jackson, D.,. 2018. Analysis of Moisture, Oil, and Fatty Acid Composition of Olives by near-Infrared Spectroscopy: Development and Validation Calibration Models. Journal of the Science of Food and Agriculture 98(5), 1821–31. DOI: 10.1002/JSFA.8658,.

Yang, H., Irudayaraj, J., Paradkar, M.M.., 2005. Discriminant Analysis of Edible Oils and Fats by FTIR, FT-NIR and FT-Raman Spectroscopy. Food Chemistry 93(1), 25–32. DOI: 10.1016/J.FOODCHEM.2004.08.039.