Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

Chemical and ATR-FTIR characterisation of olive oils from small-scale producers in San Lorenzo Maggiore, Southern Italy

Laura Ricciotti

laura.ricciotti@unicampania.it

University of Campania "Luigi Vanvitelli"

Daniele Lucariello

University of Campania "Luigi Vanvitelli"

Valeria Perrotta

University of Campania "Luigi Vanvitelli"

Raffaella Aversa

University of Campania "Luigi Vanvitelli"

Research Article

Keywords: olive oil, territorial characterisation, ATR-FTIR, small-scale producers, chemical quality, chemometrics

Posted Date: October 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11131716/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Olive oil production in small rural areas is often dispersed among family-run farms, and many of these territories remain poorly represented scientifically. This study reports the first coordinated chemical and attenuated total reflectance Fourier-transform infrared (ATR-FTIR) characterisation of oils collected from small-scale producers across San Lorenzo Maggiore, Southern Italy. Forty-five samples were collected through direct contact with producers and municipal coordination. Free acidity, peroxide value and ultraviolet indices were measured alongside the ATR-FTIR profiles. Free acidity ranged from 0.08 to 1.01% oleic acid and peroxide value from 2.5 to 40.5 meq O_2 kg^{-1}. Forty-one samples were within both individual limits for these two parameters, while four exceeded one of them; a formal commercial classification was not assigned because the complete regulatory and sensory assessment was outside the scope of the campaign. Chemical PCA showed a continuous distribution, with samples exceeding an individual limit located toward the edge of the score space. The spectra displayed the characteristic lipid bands of olive oil and provided a complementary view of variation within the collection. An exploratory PLSR analysis of free acidity gave limited cross-validated performance (R^2CV = 0.032; RMSECV = 0.169%; RPD = 1.03), indicating that the present dataset is more suitable for descriptive spectral characterisation than for quantitative replacement of titration. Taken together, the results provide an analytical baseline for a previously undocumented local production area and show the value, as well as the practical constraints, of studying heterogeneous oils collected under real small-producer conditions.

1 Introduction

Olive oil is one of the foods most closely associated with the Mediterranean landscape and diet. Its value rests not only on its sensory identity, but also on a composition dominated by monounsaturated fatty acids and enriched by a diverse minor fraction. Phenolic compounds, tocopherols, phytosterols, pigments and squalene contribute to oxidative stability, flavour and nutritional interest, while their concentrations vary with cultivar, fruit maturity, agronomic conditions and extraction practice [1, 2]. The oil reaching the consumer is therefore the result of several linked stages, from fruit development and harvest to milling, storage and bottling. Studying that chain in real production settings is important because the variability found among farms is often wider than that reproduced in laboratory-prepared samples.

The conventional quality indices used for olive oil describe different chemical processes. Free acidity reflects the release of fatty acids from triacylglycerols and may increase when olives are damaged, stored for too long before milling or handled under unsuitable conditions. Peroxide value measures primary oxidation products, whereas ultraviolet absorbance indices provide information on conjugated systems associated with oxidative change. These measurements are complementary rather than interchangeable [3–5]. They are also only part of the assessment required for a formal commercial classification: the extra-virgin category depends on conformity with the full analytical framework and on sensory evaluation. In a territorial survey, it is therefore more informative to report the measured

parameters sample by sample and to discuss their distribution than to assign a collective grade to the entire production area.

Much of the published work on olive oil comes from established producing regions, authenticated cultivar collections or carefully controlled experimental series. Smaller rural districts are less often represented, even though they preserve local cultivars, mixed orchards and production practices that may differ considerably from industrial systems. These settings are scientifically relevant because they show how composition varies under the everyday conditions in which olives are grown and processed. They are also important for territorial valorisation: a reliable analytical baseline can help local producers understand the strengths and variability of their oils, identify aspects of processing that deserve attention and provide a reference for future monitoring campaigns.

San Lorenzo Maggiore is a small municipality in the Valle Telesina area of Benevento province, Campania, Southern Italy. Olive growing is embedded in a hilly agricultural landscape and is carried out mainly by small producers. Locally established cultivars include Racioppella, Ortice, Ortolana and Sprina, often represented in farm-specific blends rather than in rigorously authenticated monovarietal oils. Harvest date, fruit condition, milling interval and storage practice may therefore differ from one producer to another. This combination of cultivar diversity and small-scale processing makes the municipality a useful case in which to examine variation within a compact production area. It does not imply that the oils form a unique geographical class, but it does make their first systematic characterisation worthwhile.

Creating such a collection was not a routine sampling exercise. Unlike a survey carried out through a single mill or a commercial supplier, the present campaign required the research team and the Municipality of San Lorenzo Maggiore to contact producers individually, explain the purpose of the study, arrange collection across a dispersed territory, record the identity of each contribution and bring the oils into one analytical workflow. Participation was voluntary and depended on the availability and trust of family-run farms. These practical conditions explain both the value and the scale of the dataset: forty-five oils represent a substantial first coverage of the accessible local network, although they cannot be treated as an exhaustive census. To the best of our knowledge, no earlier peer-reviewed study has assembled and analysed a comparable collection specifically from San Lorenzo Maggiore.

The campaign was intended to preserve the variability encountered in local production rather than to select only samples with ideal analytical values. This distinction matters. A curated set can be useful for method development, but it may conceal the heterogeneity of real oils. By retaining the full collection, including samples outside an individual acidity or peroxide limit, the study records both favourable profiles and points at which production or storage practices may need closer attention. The resulting picture is more useful for a first territorial baseline than a dataset restricted in advance to uniformly compliant oils. Comparisons with published Mediterranean datasets can then place the local ranges in context without claiming superiority, authentication or geographical exclusivity.

ATR-FTIR spectroscopy offers a complementary way to examine this variability. It is rapid, requires little sample preparation and produces a molecular fingerprint dominated by the vibrational features of triacylglycerols. In olive oils, bands related to ester carbonyl groups, aliphatic chains and unsaturation have been widely used in studies of adulteration, cultivar discrimination and quality assessment [6–15]. Chemometric methods can summarise these complex spectra and explore their relationship with conventional measurements. Principal-component analysis (PCA) is especially useful for visualising the main directions of variation, provided that score patterns are not mistaken for validated classes.

Partial least-squares regression (PLSR) has also been used to relate spectra to quantitative reference data. Its performance, however, depends strongly on the sample design, reference range and independence of validation. Models developed from deliberately prepared mixtures or sharply contrasting oil categories address a different problem from the prediction of a relatively narrow natural property in genuine oils. Free acidity is particularly challenging because titration measures free fatty acids against a spectrum dominated by esterified lipids. In a modest observational dataset, PLSR is therefore best regarded as an exploratory test of the information carried by the spectra, not as the sole measure of the scientific value of the sampling campaign [16–20].

A further feature of the present archive is that the spectra were completed over three acquisition periods, with nine sample identifiers measured on two occasions. This was not designed as a formal inter-session reproducibility experiment, but it offers useful information on how the spectral profiles behaved across the practical course of the study. Examining acquisition history alongside chemical and spectral PCA helps separate variation associated with the oils from variation that may have arisen through storage, handling or measurement conditions. Such information is relevant to any future attempt to expand the local database or develop predictive applications.

The main aim of this work was to provide the first integrated chemical and ATR-FTIR characterisation of olive oils collected from small-scale producers in San Lorenzo Maggiore. Specifically, the study describes the distribution of free acidity, peroxide value and ultraviolet indices; examines their joint variation by PCA; compares the observed ranges with published olive-oil datasets; and documents the principal ATR-FTIR features of the collection. Spectral PCA and repeated-sample histories were used to explore variation across acquisition periods. Finally, PLSR was included as a secondary analysis to assess, cautiously, whether the available spectra could support an estimate of free acidity. In this way, the work remains centred on the territorial collection and its chemical and spectroscopic profile while defining a transparent starting point for larger, multi-year studies.

2 Materials and methods

2.1 Sampling campaign and study area

Forty-five olive-oil samples were obtained from small-scale producers in San Lorenzo Maggiore (Benevento, Campania, Italy) through a coordinated campaign conducted within a Master Agreement

between the Department of Architecture and Industrial Design of the University of Campania Luigi Vanvitelli and the Municipality of San Lorenzo Maggiore. Producers were contacted individually, collection had to be scheduled across a dispersed rural network, and each contribution was identified before the oils were brought into a common analytical workflow. The final number reflects the oils that could be assembled through voluntary participation and represents meaningful first coverage of the accessible local production network rather than a statistically exhaustive census. Before analysis, the oils were stored at 15–18°C and protected from light to limit oxidation. The geographical setting is shown in Online Resource 1, Fig. S1. The oils were supplied by the producers as extra virgin olive oil. Because sensory assessment and the complete set of conformity tests were not available, the study compares only the measured chemical criteria and does not assign a formal commercial category. The study involved neither human participants nor animals.

2.2 Reference chemical measurements

Free acidity was determined by acid–base titration and expressed as percentage oleic acid; peroxide value was determined by iodometric titration and expressed as meq O_2 kg^{-1}. Ultraviolet absorbance was recorded from 800 to 200 nm with a Cary 60 UV–Vis spectrophotometer (Agilent Technologies) in dual-beam mode, using a 1% (w/v) oil solution in isooctane and a 1 cm path-length cuvette. K232, K272 and ΔK were extracted from the spectra without baseline correction, following the laboratory workflow used for the original dataset. Each chemical determination was performed in duplicate. The interpretation considered Commission Delegated Regulation (EU) 2022/2104, Commission Implementing Regulation (EU) 2022/2105 and the International Olive Council trade standard [3–5]. The dataset contains K272; it is reported descriptively and is not presented as identical to the K268/K270 criteria used in current regulatory tables. Because this retrospective territorial dataset was not designed as a metrological validation study, formal between-day uncertainty and trueness estimates were not generated; small between-sample differences should therefore be interpreted cautiously.

2.3 ATR-FTIR measurements

Spectra were recorded with a Cary 630 ATR-FTIR spectrometer (Agilent Technologies) from 4000 to 650 cm^{-1} at 2 cm^{-1} nominal resolution using 128 co-added scans. The crystal was cleaned between measurements and a new background was acquired according to the laboratory workflow. The archive contained 54 files corresponding to 45 unique sample identifiers. One spectrum per identifier was used in the principal analyses: 11 spectra from the 2025 period, nine spectra acquired in July 2026 for previously missing identifiers and 25 spectra acquired in May 2026. Nine identifiers had spectra from two periods; these were examined separately as repeated histories and were not counted as additional samples.

2.4 Multivariate analysis

Chemical PCA used autoscaled free acidity, peroxide value, K232, K272 and ΔK for all 45 oils. Spectral PCA used standard-normal-variate (SNV) transformed absorbance values in the 1800–650 cm^{-1}

fingerprint region [21]. PCA was used to summarise covariance, not to build a cultivar or geographical classifier. Acquisition-period labels were added after fitting to help interpret the score distribution.

PLSR was included as an exploratory secondary analysis with free acidity as the response. The prespecified model used SNV spectra in the 1800–650 cm^{-1} region and one latent variable, evaluated by five-fold cross-validation repeated 20 times. A sensitivity screen considered raw, SNV and second-derivative Savitzky–Golay preprocessing, alternative spectral intervals and one to eight latent variables [22–24]. To avoid treating model selection as independent validation, the selected candidates were also examined by nested repeated cross-validation and by leaving each acquisition period out in turn. R^2, RMSE, mean absolute error, bias and RPD were calculated. The complete validation hierarchy is reported in Online Resource 1, Table S1.

2.5 Repeated-sample comparison and software

For the nine identifiers measured in two periods, Pearson correlation and root-mean-square distance were calculated after SNV transformation. Each earlier spectrum was also compared with the 25 May 2026 spectra, and the rank of its correct counterpart was recorded. All analyses used the complete spectral archive and the 45-row chemical table. The chemometric reanalysis was performed in Wolfram Language, while independent data checks, chemical PCA, figures and document assembly used Python. No spectra or reference values were imputed.

3 Results and discussion

3.1 Chemical profile of the local collection

The 45 oils covered a broad but interpretable chemical range (Table 1). Free acidity exceeded 0.8% in two samples (0.84 and 1.01%), while peroxide value exceeded 20 meq O_2 kg^{-1} in two different samples (29.4 and 40.5 meq O_2 kg^{-1}). Thus, 41 samples were within both individual limits. This predominance is encouraging for the local production context, but the four exceedances also show why the collection should not be described as uniformly compliant. Moreover, even the samples within these two limits cannot be assigned to the extra-virgin category here without the remaining required analyses and a sensory assessment.

Table 1
Descriptive statistics for the chemical measurements (n = 45)

Variable	Unit	Mean	SD	Minimum	Maximum
Free acidity	% oleic acid	0.303	0.174	0.080	1.010
Peroxide value	meq O_2 kg^{-1}	7.500	6.921	2.500	40.500
K232	—	2.099	0.560	1.047	3.438
K272	—	0.135	0.065	0.039	0.295
ΔK	—	0.002	0.002	-0.002	0.009

K272 is reported as measured and is not used as a substitute for the current regulatory K268/K270 criterion.

The first two components of the chemical PCA explained 45.88% and 22.18% of the autoscaled variance (68.06% in total). Samples exceeding an acidity or peroxide limit occupied peripheral positions, especially on the positive side of PC1, whereas most of the collection formed a continuous cloud (Fig. 1A). The loading plot shows that free acidity and peroxide value contribute in the same general PC1 direction, while ΔK contributes more strongly to PC2 (Fig. 1B). These patterns describe the main chemical contrasts within the collection; they do not establish discrete cultivar or geographical groups.

3.2 Comparison with published olive-oil datasets

Placing the San Lorenzo Maggiore oils beside published datasets helps to interpret the scale of the observed variability (Table 2). The ranges compiled from recent Mediterranean studies are not formal reference intervals, because the cited works differ in cultivar, harvest, extraction, storage and analytical design [25–31]. Nevertheless, they show that most values in the present collection fall within the broad compositional space reported for olive oils, while the upper ends of free acidity, peroxide value and K232 are wider than those in the selected comparison studies.

Table 2
Full-dataset ranges for the San Lorenzo Maggiore oils compared descriptively with values reported in published studies

Parameter	San Lorenzo Maggiore	Published values	References
Free acidity (% oleic acid)	0.08–1.01	0.08–0.35	[25–28]
Peroxide value (meq O_2 kg^{-1})	2.5–40.5	3.5–8.0	[25, 26, 31]
K232	1.047–3.438	1.60–2.25	[25, 29, 30]
K272	0.039–0.295	0.09–0.20	[25, 29, 30]
ΔK	−0.002–0.009	≤ 0.005	[25, 29]

Published values are a descriptive compilation from the cited studies, not regulatory thresholds or a meta-analytical reference range.

The broader local ranges were driven mainly by a small number of samples rather than by a uniform shift of the whole collection. This distinction is important for territorial interpretation. It suggests that the campaign captured oils with generally favourable values together with a few cases that merit attention to fruit condition, milling delay or storage. The comparison therefore supports a balanced reading: San Lorenzo Maggiore is not presented as analytically superior to other regions, but the first coordinated dataset provides a useful local benchmark and identifies the degree of variability that future producer support and multi-year monitoring should address.

3.3 ATR-FTIR profile of the oils

All spectra displayed the characteristic lipid profile of olive oil. The principal features included the = C–H stretching band near 3006 cm^{-1}, aliphatic C–H stretching near 2922 and 2853 cm^{-1}, the intense ester carbonyl band near 1744 cm^{-1}, deformation bands near 1460 and 1377 cm^{-1} and C–O stretching through the 1235–1030 cm^{-1} region [6, 13, 14]. Their positions were consistent across the collection, while differences in intensity and baseline reflected the combined effects of composition, handling and acquisition period. ATR-FTIR thus provided a rapid fingerprint of the full set and complemented the conventional chemical measurements.

Table 3
Principal ATR-FTIR bands observed in the olive-oil spectra

Approximate band (cm^{-1})	Assignment	Interpretive relevance
3006	=C–H stretching	Unsaturation-related lipid signal
2922, 2853	CH_2 asymmetric and symmetric stretching	Bulk hydrocarbon-chain signal
1744	Ester C = O stretching	Triacylglycerol ester functionality
1460, 1377	CH_2/CH_3 deformation	Aliphatic-chain signal
1235–1030	C–O stretching	Ester/glycerol fingerprint region
~ 720	CH_2 rocking	Long-chain aliphatic structure

Assignments are qualitative and are used to describe the spectral profile of the collection.

3.4 Spectral variation across acquisition periods

PCA of the SNV-treated fingerprint region concentrated 81.45% of the variance in PC1 and 13.49% in PC2. The scores were arranged partly by acquisition period, particularly along PC1 (Fig. 3A). This pattern

cannot be attributed to a single cause because acquisition period was not experimentally separated from storage duration and handling history. It does, however, show that the practical history of the samples contributed appreciably to the measured spectral variance and should be considered when this database is extended.

The repeated sample identifiers led to the same cautious interpretation. Their cross-period spectral correlations were high (median 0.965; range 0.936–0.988), yet the relative-distance analysis was less reassuring: the correct later counterpart had a median rank of 15 among 25 May 2026 spectra, and only one of nine appeared in the top three (Fig. 3B). The oils retained the same overall lipid fingerprint, but smaller sample-specific differences were not always preserved across periods. Future campaigns should therefore include replicate acquisitions and stable quality-control oils from the outset.

3.5 Exploratory relationship between ATR-FTIR spectra and free acidity

PLSR was examined as a secondary question rather than as the main purpose of the study. The prespecified one-latent-variable model gave $R^2CV = 0.032$, RMSECV = 0.169% oleic acid, MAE = 0.110% and RPD = 1.03. These values do not support quantitative prediction or replacement of the reference titration. A wider search of preprocessing choices and latent variables produced a modest apparent increase ($R^2CV = 0.189$; RPD = 1.12), but the gain was not retained under nested or acquisition-period-blocked validation. Full results are provided in Online Resource 1, Table S1 and Fig. S2.

This outcome is understandable in chemical terms. Free acidity measures a comparatively small free-fatty-acid fraction, whereas the infrared spectrum is dominated by esterified triacylglycerols. In a heterogeneous collection, variation in bulk composition and sample history can readily be larger than the spectral contribution associated with acidity. The result does not detract from the value of ATR-FTIR as a rapid fingerprinting tool for the collection; it simply defines the present boundary between descriptive characterisation and quantitative calibration. A predictive model would require a larger, prospectively balanced dataset, replicate measurements and genuinely independent validation.

3.6 Strengths and limitations

The main strength of the work is the coordinated assembly of 45 traceable oils from a production system distributed among numerous small farms. This created the first shared chemical and spectral baseline for San Lorenzo Maggiore and retained the heterogeneity encountered in practice. The study also combines conventional measurements, chemical PCA, ATR-FTIR fingerprinting and transparent exploratory modelling. Its limitations are equally clear. The collection comes from one municipality and one campaign, participation was voluntary, cultivar identity was not established through an authenticated external panel, acquisition period was partly confounded with storage and handling, and no sensory assessment was available. These constraints limit geographical, cultivar and commercial-classification claims, but they do not remove the value of documenting the local production for the first time.

4 Conclusions

This study provides the first coordinated chemical and ATR-FTIR characterisation of olive oils collected directly from small-scale producers in San Lorenzo Maggiore. Organising the 45-sample campaign required individual engagement with a dispersed network of family-run farms, and the resulting dataset offers a meaningful first view of a territory that had not previously been described in this way. Most samples were within the individual free-acidity and peroxide-value limits considered here, while four exceedances revealed the variability that can arise in real production and highlighted where closer attention to processing or storage may be useful. Chemical PCA and comparison with published datasets placed that variability in context, and ATR-FTIR documented a coherent olive-oil fingerprint across the collection.

The exploratory PLSR did not provide a reliable quantitative estimate of free acidity, so titration remains the appropriate reference measurement. This is a secondary limitation rather than the central message of the study. The principal contribution is the territorial baseline created by combining direct producer sampling, conventional chemistry and rapid spectroscopy. Repeating the campaign across harvest years, recording production variables more systematically and adding replicate spectra, sensory assessment and authenticated comparison oils would allow this baseline to develop into a stronger resource for local quality monitoring and future traceability research.

Declarations

Funding: No external funding was reported for this study.

Competing interests: The authors declare no competing interests.

Author contributions: Conceptualization, L.R. and R.A.; investigation, L.R., D.L. and V.P.; data curation, L.R. and D.L.; methodology and validation, L.R. and R.A.; writing—original draft, L.R.; writing—review and editing, all authors; supervision, R.A. All authors approved the submitted version.

Data availability: The chemical dataset is included in Online Resource 1 (Table S2) and is also supplied in machine-readable form as Online Resource 2. The spectral-file inventory and raw ATR-FTIR spectra are available from the corresponding authors on reasonable request.

Ethics approval: Not applicable.

Clinical trial number: Not applicable.

Acknowledgements: The authors are deeply grateful to the olive-oil producers of San Lorenzo Maggiore, whose trust and participation made this coordinated territorial campaign possible. The authors also thank the Municipality of San Lorenzo Maggiore for facilitating contact across the dispersed local production network and the trainees who supported the laboratory analyses. The municipal collaboration was institutional; the Municipality had no role in sample selection, data analysis or interpretation.

References

1. Aparicio R, Harwood J (eds) (2013) Handbook of Olive Oil: Analysis and Properties, 2nd edn. Springer, New York. https://doi.org/10.1007/978-1-4614-7777-8
2. Boskou D (ed) (2015) Olive and Olive Oil Bioactive Constituents. AOCS, Urbana
3. European Commission (2022) Commission Delegated Regulation (EU) 2022/2104 of 29 July 2022 supplementing Regulation (EU) 1308/2013 as regards marketing standards for olive oil. Off J Eur Union L284:1–22
4. European Commission (2022) Commission Implementing Regulation (EU) 2022/2105 of 29 July 2022 laying down rules on conformity checks of marketing standards for olive oil and methods of analysis. Off J Eur Union L284:23–48
5. International Olive Council (2024) Trade standard applying to olive oils and olive-pomace oils, COI/T.15/NC No 3/Rev. 20. Madrid
6. Rohman A, Man YBC (2014) The use of FTIR spectroscopy and chemometrics for rapid authentication of extra virgin olive oil. J Am Oil Chem Soc 91:207–213. https://doi.org/10.1007/s11746-013-2370-5
7. Vasconcelos M, Coelho L, Barros A, de Almeida JMMM (2015) Study of adulteration of extra virgin olive oil with peanut oil using FTIR spectroscopy and chemometrics. Cogent Food Agric 1:1018695. https://doi.org/10.1080/23311932.2015.1018695
8. Sun X, Lin W, Li X, Shen Q, Luo H (2015) Detection and quantification of extra virgin olive oil adulteration with edible oils by FT-IR spectroscopy and chemometrics. Anal Methods 7:3939–3945. https://doi.org/10.1039/C5AY00472A
9. Jiménez-Carvelo AM, Osorio MT, Koidis A, González-Casado A, Cuadros-Rodríguez L (2017) Chemometric classification and quantification of olive oil in blends with edible vegetable oils using FTIR-ATR and Raman spectroscopy. LWT 86:174–184. https://doi.org/10.1016/j.lwt.2017.07.050
10. Rohman A, Windarsih A, Erwanto Y, Zakaria Z (2017) FTIR spectroscopy coupled with chemometrics for authentication of extra virgin olive oil. Int J Food Prop 20:S1173–S1181. https://doi.org/10.1080/10942912.2017.1336718
11. Lamas S, Rodrigues N, Peres AM et al (2021) Fourier transform infrared spectroscopy–chemometric approach as a non-destructive olive cultivar tool for discriminating Portuguese monovarietal olive oils. Eur Food Res Technol 247:2473–2484. https://doi.org/10.1007/s00217-021-03809-8
12. Malavi D et al (2023) Hyperspectral imaging and chemometrics for authentication of extra virgin olive oil: a comparative approach with FTIR, UV–VIS, Raman, and GC–MS. Foods 12:429. https://doi.org/10.3390/foods12030429
13. Guillén MD, Cabo N (1997) Characterization of edible oils and lard by Fourier transform infrared spectroscopy. J Am Oil Chem Soc 74:1281–1286. https://doi.org/10.1007/s11746-997-0058-4
14. Lerma-García MJ, Ramis-Ramos G, Herrero-Martínez JM, Simó-Alfonso EF (2010) Authentication of extra virgin olive oils by Fourier-transform infrared spectroscopy. Food Chem 118:78–83.

https://doi.org/10.1016/j.foodchem.2009.04.092

15. Rohman A, Ghazali MAB, Windarsih A et al (2020) Comprehensive review on application of FTIR spectroscopy coupled with chemometrics for authentication analysis of fats and oils in food products. Molecules 25:5485. https://doi.org/10.3390/molecules25225485
16. Varma S, Simon R (2006) Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics 7:91. https://doi.org/10.1186/1471-2105-7-91
17. Cawley GC, Talbot NLC (2010) On over-fitting in model selection and subsequent selection bias in performance evaluation. J Mach Learn Res 11:2079–2107
18. Westad F, Marini F (2015) Validation of chemometric models—a tutorial. Anal Chim Acta 893:14–24. https://doi.org/10.1016/j.aca.2015.06.056
19. Roberts DR, Bahn V, Ciuti S et al (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40:913–929. https://doi.org/10.1111/ecog.02881
20. Williams PC (2014) The RPD statistic: a tutorial note. NIR News 25:22–26. https://doi.org/10.1255/nirn.1419
21. Barnes RJ, Dhanoa MS, Lister SJ (1989) Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Appl Spectrosc 43:772–777. https://doi.org/10.1366/0003702894202201
22. Savitzky A, Golay MJE (1964) Smoothing and differentiation of data by simplified least squares procedures. Anal Chem 36:1627–1639. https://doi.org/10.1021/ac60214a047
23. Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. Chemom Intell Lab Syst 58:109–130. https://doi.org/10.1016/S0169-7439(01)00155-1
24. van den Rinnan Å, Engelsen SB (2009) Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends Anal Chem 28:1201–1222. https://doi.org/10.1016/j.trac.2009.07.007
25. Bechar S, Khamar M, Cherkaoui E et al (2024) Extraction methods and geographical variability influence on phenolic content, fatty acid composition and physicochemical quality of Moroccan Picholine olive oils. OCL 31:20. https://doi.org/10.1051/ocl/2024017
26. Ghorbel A, Guidara W, Rekik M et al (2025) Chemical and sensory characteristics of olive oils extracted from Tunisian olive varieties. ACS Omega 10:16579–16588. https://doi.org/10.1021/acsomega.4c11690
27. Loizzo MR, Sicari V, Spizzirri UG et al (2022) Evaluation of selected quality parameters of Agristigna monovarietal extra virgin olive oil during storage. Foods 11:1113. https://doi.org/10.3390/foods11081113
28. Manai-Djebali H, Trabelsi N, Medfai W et al (2023) Chemical composition of monovarietal extra virgin olive oils obtained from Tunisian mills. Pol J Food Nutr Sci 73:354–366. https://doi.org/10.31883/pjfns/174488

29. Benguennouna N, Benabdelmoumene D, Dahmouni S et al (2025) Impact of diverse irrigation water sources on olive oil quality and physicochemical properties. Sci Rep 15:15049. https://doi.org/10.1038/s41598-025-99425-7
30. Rey-Giménez R, Vázquez-Ayala S, Laya-Reig D, Sánchez-Gimeno AC (2023) Chemometric and physico-chemical characterization of oils from autochthonous cultivars grown in Aragon. Foods 12:803. https://doi.org/10.3390/foods12040803
31. Chiavaro E, Vittadini E, Rodriguez-Estrada MT et al (2007) Monovarietal extra virgin olive oils: correlation between thermal properties and chemical composition. J Agric Food Chem 55:10779–10786. https://doi.org/10.1021/jf0709909

Figures

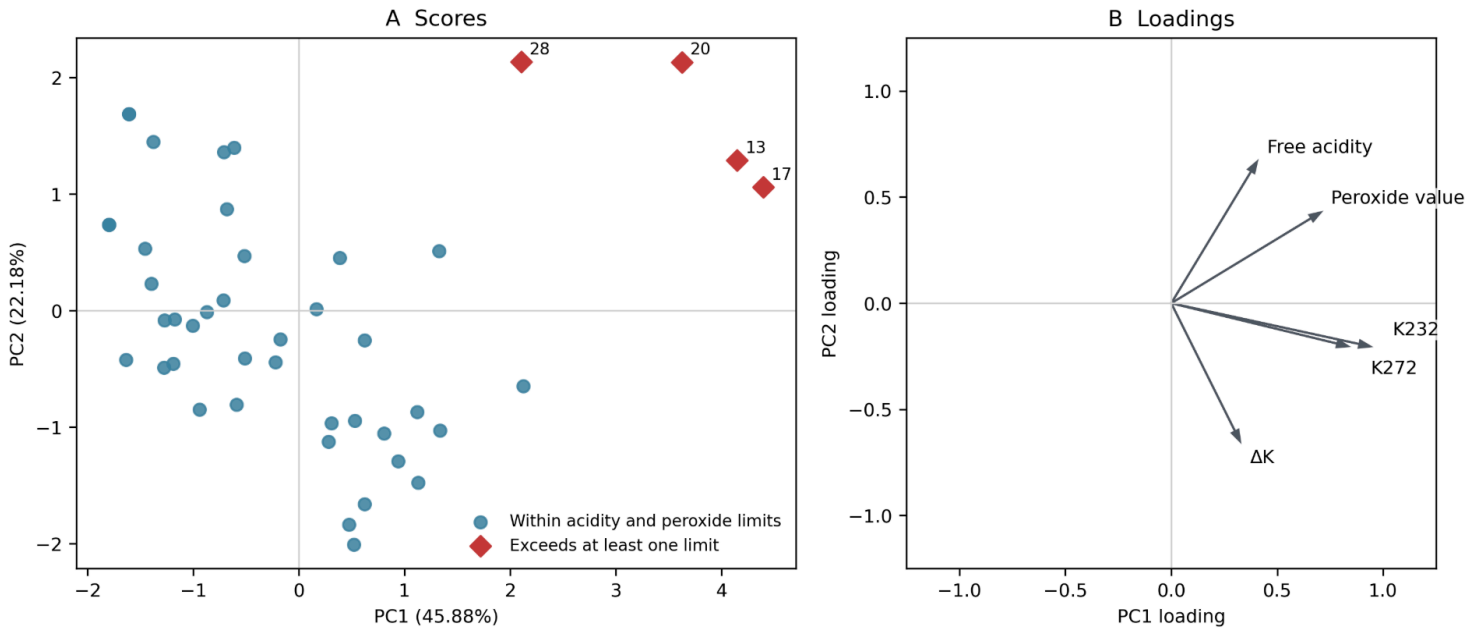

Figure 1

Chemical PCA based on autoscaled free acidity, peroxide value, K232, K272 and ΔK. A Scores; diamonds identify samples exceeding the individual free-acidity or peroxide-value limit. B Variable loadings. The analysis describes covariance within the collection and is not a validated classification

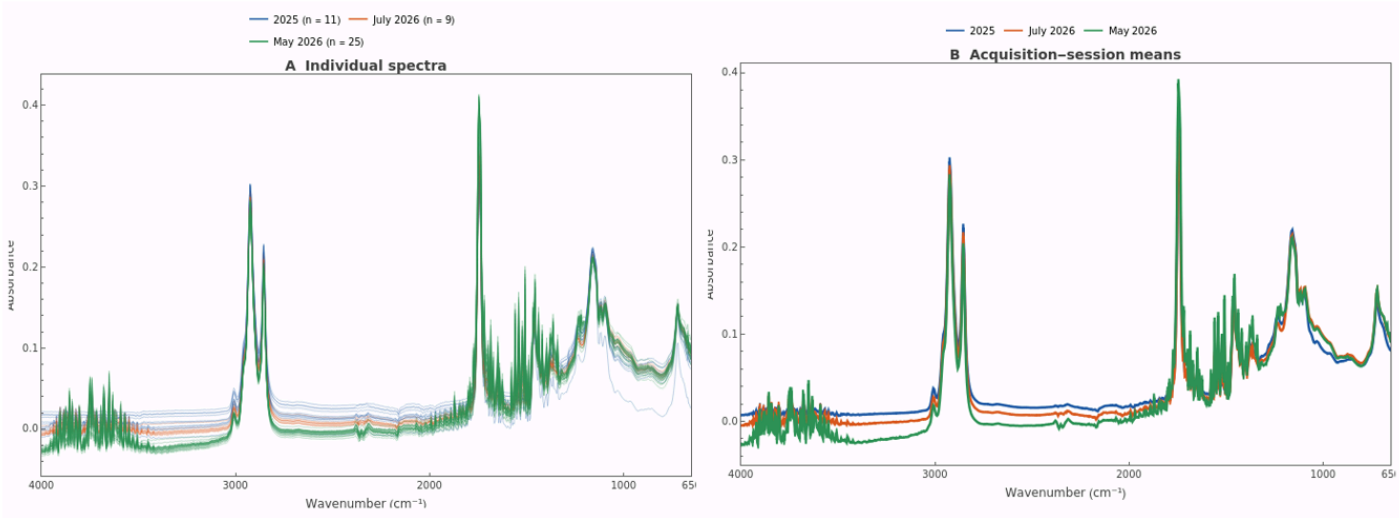

Figure 2

Raw ATR-FTIR spectra for the 45 unique sample identifiers. A Individual spectra coloured by acquisition period. B Mean spectrum for each acquisition period. The main band positions are shared across the collection, while baseline and amplitude vary between periods

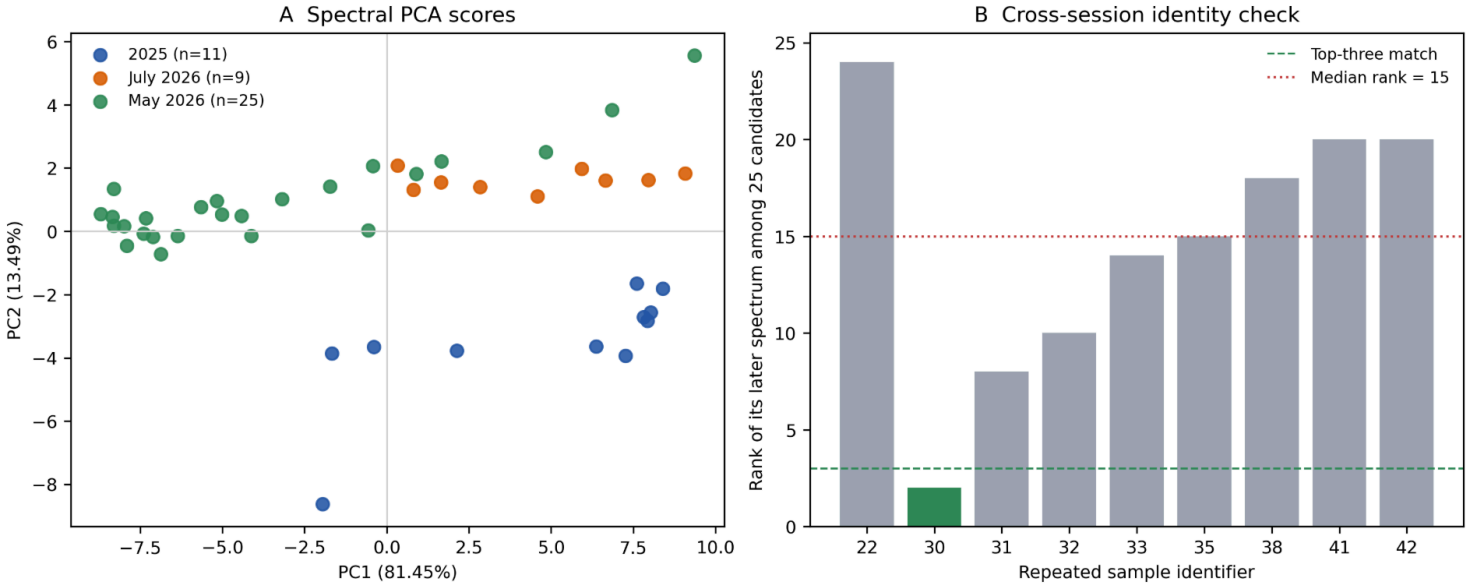

Figure 3

Spectral variation across acquisition periods. A PCA scores of SNV-treated fingerprint spectra, coloured by acquisition period. B Rank of the correct later counterpart for nine repeated sample identifiers; rank 1 denotes the closest of the 25 May 2026 spectra

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 03Supplementarymaterial.pdf