Could the quality characteristics of virgin olive oil influence its geographical classification using untargeted infrared spectroscopy and chemometrics?

Davide De Angelis, Michela Pia Totaro, Francesco Caponio, Michele Faccia, Giacomo Squeo

PII: S2590-1575(25)00395-5

DOI: https://doi.org/10.1016/j.fochx.2025.102548

Reference: FOCHX 102548

To appear in: *Food Chemistry: X*

Received date: 20 March 2025

Revised date: 29 April 2025

Accepted date: 10 May 2025

Please cite this article as: D. De Angelis, M.P. Totaro, F. Caponio, et al., Could the quality characteristics of virgin olive oil influence its geographical classification using untargeted infrared spectroscopy and chemometrics?, *Food Chemistry: X* (2024), https://doi.org/10.1016/j.fochx.2025.102548

Could the quality characteristics of virgin olive oil influence its geographical classification using untargeted infrared spectroscopy and chemometrics?

Davide De Angelis, Michela Pia Totaro, Francesco Caponio, Michele Faccia, Giacomo Squeo

University of Bari “Aldo Moro”, Department of Soil, Plant and Food Science (DiSSPA), Via Amendola, 165/A, 70126 Bari, Italy

Michela.totaro@uniba.it

Francesco.caponio@uniba.it

Michele.faccia@uniba.it

Giacomo.squeo@uniba.it

*Corresponding author: davide.deangelis@uniba.it

Abstract

Ensuring the geographical authenticity of virgin olive oil (VOO) is essential for quality control, fraud prevention, and regional product protection. This study evaluates Near-Infrared (NIR), and Infrared (IR) spectroscopies combined with chemometrics to classify 97 VOO samples from Italy (Apulia, Tuscany) and foreign countries (Morocco, Jordan, Greece, Tunisia, Spain). Partial Least Squares Discriminant Analysis (PLS-DA) effectively classified VOOs according to geographical origin, with sensitivity and specificity values higher than 0.90 in prediction. However, we also examined whether chemical quality traits, such as peroxide values and fatty acid composition, could introduce biases in the classification models. The findings suggest that inconsistent quality grade of samples could affect classification. To mitigate this, a preliminary quality assessment is recommended before applying untargeted spectroscopic methods for authentication. This study highlights the importance of integrating quality control with untargeted approaches, giving suggestions for developing more reliable authentication techniques for VOO and other foods.

Keywords

Geographical authentication; Extra virgin olive oil; Infrared spectroscopy; Quality characteristics

1. Introduction

Olive oil production in the European Union is estimated to be 2079 kilo tons in 2024, with Spain as the major producer, followed by Greece and Italy (Agridata, EU, 2025). International regulations and standards like those issued by European Union (EU), International Olive Council (IOC), and Codex Alimentarius regulate the manufacturing process and set the quality and purity parameters for oils from olives (Conte et al. 2020). Although highly regulated, weaknesses in these frameworks remain, and they are mainly related to traceability and authentication along the production chain. In fact, while several analytical and official methods are designed to monitor product quality and eventually detect adulteration, identifying intentional misdescriptions of geographical origin is more challenging. There is no official method for testing geographical origin, and control is usually based on verifying documentation within the supply chain (Quintanilla-Casas et al. 2022). The motivation behind fraud related to geographical origin is the economic value of olive oil. Specifically for the geographical origin, the country of production is one of the main factors that guide consumer choice (Conte et al. 2020), therefore having a significant impact on product price. For example, the price of Italian extra virgin olive oil (EVOO) is generally higher than €9.3-9.5 per kg, whereas Spanish or Greek olive oils are typically 15-20% cheaper (IOC, 2024). Olive oils produced in other Mediterranean and North African countries are even lower priced (IOC, 2024), and increasing imports from these countries could stimulate fraudulent labeling.

Metabolomic approaches have been proposed as tools for assessing geographical origin of olive oil, as recently reviewed by Chaji et al. (2023). From this comprehensive review, wet-chemistry methods based on liquid- or gas-chromatography coupled with different kinds of detectors, have emerged as prevalent techniques. However, these analyses are time consuming and resource demanding, and require highly specialized operators, meaning that their implementation for ordinary controls in the olive oil sector remains challenging. To overcome these drawbacks, non-destructive methods of analysis have been developed, including those based on near- or mid-infrared radiation (Arroyo-Cerezo et al. 2024), fluorescence (Jiménez-Carvelo et al. 2019), nuclear magnetic resonance (Mannina and Sobolev, 2011), or combinations of these techniques with data fusion approaches (Lozano et al. 2025). NIR (Near-Infrared) and FT-IR (Fourier-Transform Infrared) spectroscopies are particularly promising because of the rapidity, the high cost-effectiveness ratio, and the possibility to scale them into portable or miniaturized sensors implementable within industrial processes (Gullifa et al. 2023). NIR and FT-IR spectroscopies have been previously used to detect adulterations of EVOO (Vanstone et al., 2018; Mousa et al., 2022; Vieira et al., 2021; Meng et al., 2023; Klinar et al., 2024). Whereas the assessment of the geographical origin has been earlier addressed for oils produced in different regions of Morocco (Laouni et al., 2023) and Argentina (Jiménez-Carvelo et al., 2019; Lozano et al., 2025), or in other countries (Hennessy et al., 2009; Bevilacqua et al., 2012; Bragolusi et al., 2021). The strength of untargeted approaches relies on the direct analysis of food samples, without any preparation or manipulation, that ideally leads to an unbiased estimation of the food characteristics. However, once produced, olive oils are very susceptible to chemical modification that is not related to geographical origin while depends mainly on storage conditions and time. In untargeted analytical approaches, classification models typically rely on the assumption that oils from different geographical regions exhibit distinct and stable metabolomic profiles. However, if the chemical quality of the oils is influenced by factors unrelated to geographical origin, this could introduce a bias into the classification process. Essentially, the model may distinguish oils based on their chemical quality rather than their true geographic origin. In fact, degradative modifications impact key quality parameters and thus might compromise the chemical stability required for reliable authentication using untargeted methods (Willenberg et al., 2019) whose successful application depends on the presence of stable markers (Lozano-Castellón et al., 2022). Despite the practical relevance that these issues have on the possible real application of untargeted vibrational spectroscopic methods for VOO origin classifications, only limited literature addressed such topic. Therefore, in this study, NIR and FT-IR spectroscopies were first tested for their performance in discriminating VOO based on their geographical origin. Samples were collected from Italy (Apulia and Tuscany) and abroad (Morocco, Jordan, Greece, Tunisia, and Spain). Then, the influence of oils quality on the performance of the classification models was investigated and discussed.

2. Material and Methods

2.1 Sampling

Ninety-six authentic samples of virgin olive oil were collected by trusted mills and resellers. Specifically, 70 oils were of Italian origin, and among them, 57 were produced in the Apulia region (southern Italy) and 13 were produced in Tuscany (central Italy). Additionally, 27 oils were collected from foreign countries recognized for their olive oil production, including samples from the Morocco (n = 21), Jordan (n = 3), Greece (n =1), Tunisia (n = 1) and Spain (n = 1). All the oils were produced in the 2022-2023 production years, and they were labelled as extra virgin by producers. Once the samples were collected, they were stored in dark glass bottles at 4 °C to prevent oxidation phenomena from light and heat. Before the NIR and IR spectra acquisition, the samples were left overnight at room temperature (25 °C).

2.2 Analytical determination

2.2.1 Routine analyses and fatty acids composition

The acidity, expressed as a percentage of free oleic acid, and the peroxide value, expressed in meq O_2/kg of oil, were analyzed according to standard methods of Commission regulation (EU) 2022/2105. Fatty acid composition was analyzed following sample transesterification with KOH 2N in methanol, according to EU standard methods (Commission regulation EU 2022/2105). The analysis was carried out using a gas chromatograph (Agilent 7890A, Agilent Technologies, Santa Clara, USA) equipped with a flame ionization detector (FID). The conditions of analysis were previously reported in Opaluwa et al. (2025). Specifically, the injector temperature was 210 °C and the column (SP2340 fused silica capillary column, 60 m × 0.25 mm i.d., 0.2 mm film thickness) (Supelco Park, Bellefonte, USA) was subjected to a temperature of 80 °C for 5 minutes, followed by an increase from 80 °C to 200 °C at a rate of 2 °C per minute. Then, the temperature was held at 200 °C for 5 minutes and finally the temperature was raised to 240 °C at a rate of 10 °C per minute. Helium served as carrier gas at a constat flow rate of 1 mL/min. The FID was set to 220 °C with an air flow of 400 mL/min and hydrogen flow of 40 mL/min. Fatty acids were identified by comparing their retention times with those of standard methyl esters (Merck KGaA, Darmstadt, Germany), and individual fatty acid quantities were expressed as a percentage of the total peak area.

2.2.2 Phenolic compounds determination

The total phenolic content (TPC) was quantified following the method described by Squeo et al. (2019). Briefly, a liquid-liquid extraction was performed using 1 g of oil, 1 mL of hexane, and 5 mL of methanol-water (70:30 v/v). After vortexing for 10 minutes and centrifuging at 4032 × g for 10 minutes at 4 °C, the hydroalcoholic phase was collected, centrifuged again at 9072 × g for 5 minutes at 4 °C, and filtered through 0.45 µm nylon filters (Merck KGaA, Darmstadt, Germany). Subsequently, 100 µL of the extract was mixed with 100 µL of Folin-Ciocalteu reagent, followed by the addition of 800 µL of a 5 % (w/v) sodium carbonate solution after 4 minutes. The mixture was then incubated in a

water bath at 40 °C for 20 minutes, and the TPC was measured at 750 nm using an Agilent Cary 60 spectrophotometer (Agilent Technologies, Santa Clara, USA). The quantification was based on a calibration curve prepared with standard concentrations of pure gallic acid at 0, 5, 10, 20, 25, 50, 75, and 100 mg/L. The resulting linear equation is y = 0.0092x +0.0021 (R^2 = 0.9997). TPC was expressed as gallic acid equivalent (GAE mg/kg).

2.2.3 NIR and IR spectra acquisition

NIR spectra were acquired in transflectance mode using the Viscous Liquid Sampler (VLS) on a Nicolet iS50 spectrometer supported by the OMNIC software (Thermo Fisher Scientific Inc., Waltham, USA) and equipped with an integrating sphere modulus working in the wavelength range of 4000-12500 cm^{-1}, with a spectral resolution of 4 cm^{-1} and 16 scans. The trasflection pathlength was 1 mm. A CaF_2 beamsplitter was used. A dark correction was applied to exclude the contribution of reflected light from the sampling window. Three spectra replicates per sample were collected, and then a new background spectrum was taken prior the next measurement (Totaro et al. 2023). The VLS was thoroughly cleaned with pure ethanol and a cotton wipe before each measurement.
IR spectra were acquired in Attenuated Total Reflection (ATR) mode using the same Nicolet iS50 infrared spectrophotometer, under the following conditions: spectral range of 651-4000 cm^{-1}, spectral resolution of 4 cm^{-1}, and 32 scans for both sample and background. A KBr beamsplitter was used. Three spectra replicates per sample were collected, and then a new background was taken. The ATR crystal was thoroughly cleaned with pure ethanol and a cotton wipe before each measurement. All the spectrophotometric analyses were carried out at room temperature (approximately 25 °C).

2.4 Chemometric models and data analysis

The technical replicates measured per sample were averaged; the mean spectra saved as .xls file and imported in Solo v9.3.1 (Eigenvector Research Inc. Manson, USA) for data analysis. The datasets (sample × spectral variable) were 97 × 2205 and 97 × 1737 for NIR and IR, respectively. The NIR and IR datasets were elaborated separately, following the same flowchart. A Principal Component Analysis (PCA) was performed to explore the structure of the data in an unsupervised manner. Different preprocessing were used to minimize the influence of unwanted signal variations, including Standard Normal Variate (SNV), first derivative (1D; Savitzky-Golay, polynomial order 2, 25 smoothing points), second derivative (2D; Savitzky-Golay, polynomial order 2, 25 smoothing points) and Generalized Least Squares Weighting (GLSW; clutter source: x-block classes, α = 0.02). All the preprocessing included a final mean centering (MC) of the data. Before constructing the supervised classification models, the dataset was divided into calibration (70%) and validation (30%) sets using the Kennard-Stone algorithm. Therefore, n = 68 samples were used for calibration and cross-validation, i.e., to build and optimize the models, whereas the remaining 29 samples were used for model validation, i.e., to predict the results on an external dataset not used in the model construction.

In this study, Partial Least Square Discriminant Analysis (PLS-DA) was used as a supervised classification model to discriminate the samples according to the region of origin, considering three classes, namely Apulian, Tuscan, and foreign oils. Additional information about the PLS-DA algorithm can be found elsewhere (El Maouardi et al. 2024, Ballabio and Consonni, 2013). The venetian blind was chosen as a cross-validation scheme (10 data splits, i.e., number of sub-validations; per each split, on average, 90% of calibration samples were used for model building and the remaining 10% for internal validation). The optimal number of latent variables (LVs) was chosen according to the classification error in cross-validation. Sensitivity, specificity and classification error were used as figure of merits to evaluate the model performances in calibration, cross-validation (on the calibration set) and prediction on the external validation set (Ballabio et al. 2018; Allegretta et al. 2023). Sensitivity is defined as the true positive rate of the model, i.e. ability to correctly classify the samples of the class of interest. Specificity is the true negative rate of the model, indicating the ability to correctly reject samples of other classes from the class of interest. The classification error for a single class was calculated as: [1 – (average of sensitivity and specificity)] of that class. The models were finally validated using the test sets. The Non-Error Rate (NER), calculated as the average of sensitivity values per class, was used a global index of models performance in prediction.

Chemical data were subjected to One-way ANOVA followed by Games-Howell post-hoc test using Minitab Statystical Software (Minitab Inc., State College, PA, USA). The level of significant differences was set at $p<0.05$.

3. Results and Discussion

3.1 NIR Spectra and Exploratory analysis of data

The NIR spectra of the oil samples are depicted in Figure 1, and in general terms, they are similar to those reported in other works (Hourant et al. 2000; Bragolusi et al. 2021; Vieira et al. 2021; Du et al. 2021). The spectra were affected by a small scatter effect, observable by the differences in the signal intensity, which can be easily corrected by e.g., the SNV preprocessing.

Figure 1

The majority of the spectral information is observable at wavenumber below 10000 cm^{-1}. For this reason, the spectral region above this value was excluded from further data elaboration (data not shown). Main bands are observable in the regions between 4237 and 4464 cm^{-1}, 4545 and 4762 cm^{-1}, 5618 and 6061 cm^{-1}, 6897 and 7353 cm^{-1}, 7937 and 8696 cm^{-1}, already identified by Hourant et al. (2000), and that are representative of the bonds of fatty acids and triacylglycerols structure (Hourant et al. 2000; Du et al. 2021; Vieira et al. 2021). The combination of C-H and CH=CH stretching vibrations explain the observed bands in the first two regions highlighted, whereas in the last three regions, bands are due to the first and second overtones of CH_2, CH_3, and CH=CH. PCA was used to explore the dataset, testing different preprocessing . Figure 2 depicts the score plots and the loadings plots of the PCA carried out using i) SNV and ii) SNV + GLSW filter. Principal

Component (PC) 1 in SNV preprocessed dataset (Figure 2a) explains almost all the variance (96.96%), but did not show useful information linked to the origin of oil. However, it seems that PC2, only explaining 1.27% of variance, grabs valuable information, with Apulian oils mostly located in the positive score of PC2. By observing the loadings plot (Figure 2b), it seems that the major contribution to PC2 is given by the spectral regions near 4000 and 6000 cm^{-1}, suggesting an effect related to the amount and position of unsaturation on the fatty acid chains (Zielińska et al., 2020). The analysis of other PCs did not evidence other relevant information useful to our scope (data not shown).

Figure 2

The application of 1D and 2D preprocessing (data not shown) did not provide better information regarding the class distribution. The results of the PCA carried out on GLSW pre-processed spectra is reported in Figure 2c and 2d. The GLSW filter minimizes the within-class variance (i.e., the variability within the same category) while preserving the between-class variance (i.e., differences between distinct classes) (Serranti et al. 2013; Allegretta et al. 2023). In this case, the variance explained by PC1 is 26.55% (Figure 2c) and the score plot highlights a clearer separation of samples, promising for possible supervised classifications. Along PC1 two clusters emerged, Apulia (positive scores on PC1) and Tuscany + foreign (negative scores on PC1). The latter clusters were separated along PC2. The loadings plot shows that such distribution of the sample was related to the absorption around 4500, 6000 and 8300 cm^{-1}, the latter derived from the second overtones of the C-H stretching vibrations, overall suggesting possible influence of the oleic and linoleic fatty acids (Du et al. 2021, Vieira et al. 2021). Although the loadings were noisier compared to those observed on SNV pre-processed spectra (Allegretta et al. 2023), similarities can still be observed. In this case as well, the loadings highlight the maximum contribution of the spectral regions near 4500 and 6000 cm^{-1} (Figure 2b and 2d).

3.2 IR spectra and exploratory analysis

The IR spectra are shown in Figure 3. The region above 3200 cm^{-1} was excluded due to a low signal-to-noise ratio (data not shown), whereas the range 2290 and 2420 cm^{-1} was excluded to remove the influence of carbon dioxide absorption. The spectra display several bands representing functional groups typical of olive oils (Hennessy et al., 2009; Laouni et al. 2023). IR signals exhibited a slight scattering effect, which was minimized using the SNV function (Figures 3a and 3b).

Figure 3

The exploration by PCA, shown in Figure 4, led to the same conclusion already drawn for the NIR dataset. In particular, the application of the GLSW preprocessing allows to observe a distinct distribution of the samples, with both Italian classes positioned on the positive side of PC2 but separated along PC1. However, the loadings plot, which illustrates the contributions of the variables, was less interpretable, as observed in a previous study (Allegretta et al., 2023). Nevertheless, the regions with the highest contributions to PC1 and PC2 were similar to those observed with SNV preprocessing alone.

Figure 4

3.3 Untargeted classification of samples based on the geographical origin using PLS-DA

The NIR and IR dataset presented have been used to develop PLS-DA models whose summaries are presented in Table 1 and Table 2, respectively. For the models based on NIR signals, four to seven latent variables were selected to minimize the cross-validation error. In the models constructed with IR spectra, the number of latent variables ranged from 2 to 7. This level of model complexity is consistent with previous studies on oils (Hennessy et al., 2009; Vieira et al., 2021; El Maouardi et al., 2024).

Generally, proper preprocessing of spectral data results in high sensitivity and specificity in both cross-validation and prediction. Sensitivity values greater than 0.90 could be considered satisfactory for classification purposes (Voccio et al., 2024). The values of specificity (always higher than 0.80) and NER confirms the adequate performance of the models. According to the results, it seems that the combination of SNV, 1D, and GLSW gave the best prediction results in terms of sensitivity, specificity and classification error for NIR signals, whereas for IR spectra, the best results were achieved using SNV+GLSW alone. Consequently, the PLS-DA scores and Variable Importance in Projection (VIP) scores for these preprocessings are shown in Figure 5 (NIR) and Figure 6 (IR). Compared to other results available in the literature regarding geographical authentication, the performance of the models was similar or even better to what was achieved using a GC-MS fingerprint (Quintanilla-Casas et al. 2022), visible and Raman spectroscopy (Kontzedaki et al. 2020), IR (Laouni et al. 2023), or NIR and IR (Jolayemi et al. 2017). By examining the Variable Importance in Projection (VIP) (Westad et al., 2013) for the NIR dataset, the first region of the spectrum, between 4500 and 6000 cm^{-1}, related to the type and the number of unsaturations of the fatty acids, is the most important (Figure 5d). This region, which also made the greatest contribution to the PCA, as discussed in Section 3.1, was similarly highlighted by Bevilacqua et al. (2012). For future implementation and optimization studies, including potential transfer to handheld instruments, this spectral region may be of particular interest.

Conversely, the VIP scores for all models based on IR spectra were notably low in the region between 1800 and 2700 cm^{-1} (Figure 6d).

Table 1. Figures of merit of the PLS-DA models computed on NIR spectra (n = 97) in Calibration, cross validation, and prediction using different preprocessing techniques.

			Calibration			Cross Validation			Prediction			
	Class	LVs	Sens	Spec	CE	Sens	Spec	CE	Sens	Spec	CE	NER
SNV + MC	APU	4	0.93	0.96	0.06	0.88	0.86	0.13	0.94	1.00	0.03	0.87
	FOR		0.90	0.83	0.13	0.80	0.79	0.20	0.86	0.82	0.16	
	TUS		1.00	0.98	0.01	0.88	0.93	0.10	0.80	0.92	0.14	
SNV + 1D + MC	APU	5	0.95	0.96	0.04	0.90	0.89	0.10	0.94	1.00	0.03	0.89
	FOR		0.85	0.92	0.12	0.80	0.85	0.17	0.71	0.91	0.19	
	TUS		1.00	0.93	0.03	0.88	0.92	0.10	1.00	0.88	0.06	
SNV + 2D + MC	APU	4	0.90	0.89	0.10	0.88	0.89	0.12	0.94	1.00	0.03	0.87

	FOR		0.85	0.85	0.15	0.80	0.85	0.17	0.86	0.91	0.12	
	TUS		1.00	0.95	0.03	0.88	0.93	0.10	0.80	0.88	0.16	
SNV + GLSW + MC	APU	7	0.98	1.00	0.01	0.90	0.86	0.12	0.94	1.00	0.03	0.93
	FOR		1.00	1.00	0.00	0.80	0.90	0.15	0.86	1.00	0.07	
	TUS		1.00	1.00	0.00	1.00	0.90	0.05	1.00	0.92	0.04	
SNV + 1D + GLSW + MC	APU	7	0.95	1.00	0.03	0.90	0.86	0.12	0.94	1.00	0.03	0.93
	FOR		0.95	1.00	0.03	0.75	0.90	0.18	0.86	1.00	0.07	
	TUS		1.00	0.97	0.02	0.88	0.95	0.09	1.00	0.96	0.02	
SNV + 2D + GLSW + MC	APU	6	0.95	0.96	0.04	0.93	0.93	0.07	0.88	1.00	0.06	0.87
	FOR		0.95	0.98	0.04	0.85	0.85	0.15	0.71	0.96	0.17	
	TUS		1.00	0.95	0.03	0.88	0.93	0.10	1.00	0.92	0.04	

APU: Apulia, FOR: foreign, TUS: Tuscany; LVs: latent variables; Sens: sensitivity; Spec: specificity; CE: classification error; NER: non-error rate; SNV: standard normal variate; 1D and 2D: first and second derivatives; GLSW: Generalized Least Squares Weighting; MC: mean centering.

Figure 5

While NIR and FT-IR spectroscopy have shown comparable performance in distinguishing different types of vegetable oils (Yuan et al., 2023) or detecting adulteration in oils (El Maouardi et al., 2024), some differences may emerge when assessing the geographical origin of oils. Specifically, studies by Jolayemi et al. (2017) and Bevilacqua et al. (2012) found that NIR tended to yield slightly better predictive accuracy than IR. Conversely, our results suggest that models built with both signals could achieve comparable and highly accurate predictions using proper preprocessing techniques.

Table 2. Figures of merit of the PLS-DA models computed on IR spectra (n = 97) in Calibration, cross validation, and prediction using different preprocessing techniques.

			Calibration			Cross Validation			Prediction			
	Class	LVs	Sens	Spec	CE	Sens	Spec	CE	Sens	Spec	CE	NER
SNV + MC	APU	7	0.92	0.93	0.07	0.79	0.87	0.17	0.84	1.00	0.08	0.95
	FOR		0.96	0.96	0.04	0.86	0.83	0.16	1.00	0.96	0.02	
	TUS		0.88	0.97	0.08	0.88	0.90	0.11	1.00	1.00	0.00	
SNV + 1D + MC	APU	3	0.55	0.93	0.26	0.53	0.83	0.32	0.47	0.90	0.31	0.69
	FOR		0.86	0.80	0.17	0.77	0.78	0.22	0.80	0.88	0.16	
	TUS		1.00	0.97	0.02	0.88	0.95	0.09	0.80	1.00	0.10	
SNV + 2D + MC	APU	3	0.50	0.93	0.28	0.50	0.80	0.35	0.47	0.90	0.31	0.82
	FOR		0.82	0.78	0.20	0.73	0.78	0.25	1.00	0.88	0.06	
	TUS		1.00	0.97	0.02	0.88	0.95	0.09	1.00	1.00	0.00	
SNV + GLSW + MC	APU	3	1.00	1.00	0.00	0.92	0.93	0.07	1.00	0.80	0.10	1.00
	FOR		1.00	1.00	0.00	0.86	0.87	0.13	1.00	0.96	0.02	
	TUS		1.00	1.00	0.00	0.88	0.98	0.07	1.00	1.00	0.00	
SNV + 1D + GLSW + MC	APU	2	0.97	0.97	0.03	0.92	0.80	0.14	0.90	0.90	0.10	0.97
	FOR		0.91	0.98	0.06	0.77	0.78	0.22	1.00	0.92	0.04	
	TUS		1.00	1.00	0.00	0.88	0.98	0.07	1.00	1.00	0.00	
SNV + 2D + GLSW + MC	APU	4	1.00	1.00	0.00	0.82	0.63	0.28	0.84	0.80	0.18	0.88

FOR	1.00	1.00	0.00	0.73	0.80	0.23	0.80	0.88	0.16
TUS	1.00	1.00	0.00	0.75	0.93	0.16	1.00	1.00	0.00

APU: Apulia, FOR: foreign, TUS: Tuscany; LVs: latent variables; Sens: sensitivity; Spec: specificity; CE: classification error; NER: non-error rate; SNV: standard normal variate; 1D and 2D: first and second derivatives; GLSW: Generalized Least Squares Weighting; MC: mean centering.

Figure 6

3.4 The possible influence of the quality parameters on the assessment of the geographical origin classification

Based on the results of the untargeted evaluation, we may conclude that the assessment of the geographical origin was successfully achieved using NIR and FT-IR spectroscopy. Indeed, as demonstrated by other studies, authentication problems can be effectively addressed with an untargeted approach (De Angelis et al. 2024; Mialon et al., 2023), which does not require prior information about the chemical composition of the products. While this approach is advantageous for real-case applications, e.g., a rapid quality control of the products, one might question whether the models could be biased to some extent by the chemical quality of the oils. Consequently, in this part of the study, we aim to conduct a more in-depth investigation of the problem, focusing on the chemical parameters that define the quality of olive oils collected in this study, and discuss their possible influence on the classification models.

The results of the analytical determinations on olive oils are reported in Table 3. The dataset collected for this study presents a highly variable chemical profile, especially in terms of fatty acid profile, acidity and peroxide values. The statistical analysis of the data reveals significant differences in peroxide values across the three classes, with mean values of 9.42, 17.94 and 33.57 meq O_2/kg for Apulian, foreign and Tuscan oils, respectively.

Table 3. Results of statistical analysis (mean ± standard deviation) of chemical analysis (acidity, peroxide value, total phenol content) and fatty acids composition of virgin olive oils from different country.

	Total Phenol Content (mg/kg)	Acidity g/100g	Peroxide Value meq O_2/kg	Plamitic acid ($C_{16:0}$)	Palmitoleic acid ($C_{16:1}$)	Stearic acid ($C_{18:0}$)	Oleic acid ($C_{18:1}$)	Linoleic acid ($C_{18:2}$)
Apulia (n = 57)	277±98 a	0.38±0.13 ab	9.42±5.39 c	13.55±2.40 a	1.12±0.68 a	1.88±0.64 b	73.13±5.24 a	8.92±3.27 a
Foreign country (n = 27)	239±94 a	0.48±0.22 a	17.94±5.97 b	14.71±2.87 a	0.83±0.39 b	2.46±0.43 a	69.43±4.04 b	10.30±3.02 a
Tuscany (n = 13)	250±92 a	0.34±0.11 b	33.57±8.16 a	13.66±2.36 a	1.12±0.11 a	1.01±0.17 b	74.53±2.27 a	6.72±0.86 b
	p=0.237	p=0.044	p<0.001	p=0.207	p=0.003	p<0.001	p<0.001	p<0.001

Different letters for the same parameters indicate significant differences according to the one-way ANOVA followed by Games-Howell post-hoc test.

These differences raise a potential concern because the classification results could be influenced by peroxide values, which are clearly not related to the geographical origin but characterize differently the classes. To address this point, firstly a PLS-DA was computed using the chemical data, and the validation results are presented in Table 4. The classification performance was comparable to that achieved using spectral data, showing high sensitivity and specificity and as expected, the VIP scores (Figure 7) indicate a strong contribution of the peroxide value, along with a minor and little influence of some fatty acids. Hence, the concern about the possible effect of the quality traits seemed confirmed. However, the extent to which the peroxide value is affecting the classification based on the spectral analyses is hard to assess precisely. Not to mention that other features not measured but correlated to the peroxide value might have played a role in the classification results. This highlights the importance of performing a quality assessment of the food products, olive oils in this specific case, before moving toward the untargeted assessment of geographical origin.

Table 4. Figures of merit of the PLS-DA models computed on chemical data (n = 97), in cross validation, and prediction using autoscaling as preprocessing.

		CAL			CV			PRED			
Class	LVs	Sens	Spec	CE	Sens	Spec	CE	Sens	Spec	CE	NER
APU	2	0.84	0.87	0.15	0.84	0.80	0.18	0.95	0.90	0.08	0.92
FOR		0.86	0.80	0.17	0.82	0.76	0.21	0.80	0.83	0.18	
TUS		1.00	0.95	0.03	0.88	0.93	0.10	1.00	1.00	0.00	

APU: Apulia, FOR: foreign, TUS: Tuscany; LVs: latent variables; Sens: sensitivity; Spec: specificity; CE: classification error; NER: non-error rate.

Figure 7

Therefore, a subset selection according to the chemical determinations was done. A total of 24 samples can be classified as non-extra-virgin (peroxide value > 20 meqO_2/kg), including two Apulian, 9 foreign and the 13 Tuscan oils. These oils were removed from the original dataset, reducing it to 73 samples, of which 55 from Apulia and 18 from foreign countries. PLS-DA models were calculated considering both the chemical data and the spectral data. The performance of the model built using the chemical data were inferior compared to those of the NIR and IR spectra (Table 5). The specificity, sensitivity and NER of the models based on spectroscopic data were very promising.

As a final consideration, we should acknowledge that this latter dataset presents an imbalance within classes (55 vs 18 samples), which could lead to concerns regarding possible biases introduced into the model (Ballabio et al. 2018). To mitigate this point, considering NER as performance metric is important as it is less affected by class imbalance compared to other metrics commonly used to evaluate classification models (Ballabio et al. 2018). Moreover, two additional classifications were carried out, by randomly selecting two sub-sets of Italian oils to make a balanced class with the Foreign oils. In both cases we achieved prediction performances similar to the models built using

the whole dataset (Supplementary Table S1). Consequently, given the magnitude of the imbalance, the metrics used to discuss the models, and the results on reduced datasets, we could conclude that the presence of possible biases is minimized, and the models built on EVOO samples are valid on the investigated dataset.

Table 5. Figures of merit of the PLS-DA models computed on a reduced dataset of EVOO samples (n = 73) using NIR spectra, IR spectra and chemical data, in cross validation, and prediction.

			Cross Validation			Prediction			
Data and Preprocessing	Class	LVs	Sens	Spec	CE	Sens	Spec	CE	NER
NIR SNV + MC	APU	3	0.97	1.00	0.01	0.94	1.00	0.03	0.97
	FOR		1.00	0.97	0.01	1.00	0.94	0.03	
IR SNV + GLSW + MC	APU	2	0.75	0.89	0.18	1.00	0.94	0.03	0.97
	FOR		0.89	0.75	0.18	0.94	1.00	0.03	
Chemical data Autoscaling	APU	1	0.84	0.92	0.12	0.94	0.83	0.11	0.89
	FOR		0.92	0.84	0.12	0.83	0.94	0.11	

APU: Apulia, FOR: foreign, TUS: Tuscany; LVs: latent variables; Sens: sensitivity; Spec: specificity; CE: classification error; NER: non-error rate; SNV: standard normal variate; GLSW: Generalized Least Squares Weighting; MC: mean centering.

4. Conclusion

The geographical origin of food is one of the main economic leverages and drivers for products valorization. However, assessing food origin is a hard task, and methods to assess it are welcome. In this framework, untargeted, non-destructive, high-throughput and green analytical methods are gaining more and more attention. NIR and IR have been already tested for the discrimination of olive oils geographical origin obtaining promising results. The results of the present study confirm this evidence. Nonetheless, less attention has been given to the possible impact of chemical features on the performance of discrimination. The results presented showed that when the quality grade of the samples is inconsistent, special attention should be paid. In particular, the different oxidation level can drive the discrimination of the sample in spite of the origin. Hence, in view of a possible untargeted analytical pipeline, a preliminary assessment of foods quality grade seems fundamental to prevent erroneous classification. Fortunately, NIR/IR methods for qualitative assessment of oils already exist or can be developed in-house. With this regard, two scenarios can be foreseen: i) a NIR/IR method is already in place for chemical characterization of the oils and ii) the opposite situation, in which there is no untargeted NIR/IR method for oil analysis. In the first case, a subsequential approach can be envisioned in which the NIR/IR signal could be firstly used to chemically characterize the samples and, based on the results, chemical coherent classes are defined which are subsequently subjected to the origin discrimination. In the second case, the approach is the same but at least some key parameters should be evaluated using other analytical methods to ensure class coherence.

This study presents the limitations that the number of samples per the different classes is limited and unevenly distributed, which may affect the robustness of the classification. Second, a net overlap between the “chemical” class high- peroxide value and the “geographical” class Tuscany was also present. As a result, we recognize that full validation of the proposed models on a larger scale would require more numerous and more balanced sample sets across all classes. However, this was beyond the scope of the present work. The main goal of our study was to explore the feasibility of using untargeted spectroscopic approaches for discriminating samples based on geographical origin and to highlight the importance of integrating a preliminary quality assessment. In the author’s opinion this is fundamental to develop sound untargeted approaches for origin discrimination of oils and food products.

Acknowledgments

We thank Carmine Summo, our beloved friend and esteemed colleague, for his contribution to the conceptualization of this manuscript. Though he is no longer with us, his dedication continues to inspire us.

This research was carried out within the Agritech National Research Center and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR)–MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4—D.D. 1032 17/06/2022, CN00000022). This manuscript reflects only the authors’ views and opinions, neither the European Union nor the European Commission can be considered responsible for them. The authors would like to thank Gianni Martellini and Francesco Tricarico for providing the foreign oils, and Claudio Rossi and Gabriella Tamasi from the University of Siena (Italy) for providing the Tuscan oils.

CRediT Author contribution

Davide De Angelis: Conceptualization, Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing; **Michela Pia Totaro**: Data curation, Formal analysis, Writing – original draft; **Francesco Caponio**: Project administration, Supervision, Writing – review & editing; **Michele Faccia**: Project administration, Supervision, Writing – review & editing; **Giacomo Squeo**: Conceptualization, Data curation, Formal analysis, Project administration, Writing – original draft, Writing – review & editing.

Conflict of Interest

None.

References

1. Agridata EU (2025). Olive oil production accessible from https://agridata.ec.europa.eu/extensions/DashboardOliveOil/OliveOilProduction.html (Accessed on 13th March 2025).
2. Allegretta, I., Squeo, G., Gattullo, C. E., Porfido, C., Cicchetti, A., Caponio, F., ... & Terzano, R. (2023). TXRF spectral information enhanced by multivariate analysis: A new strategy for food fingerprint. *Food Chemistry*, *401*, 134124.
3. Arroyo-Cerezo, A., Yang, X., Jiménez-Carvelo, A. M., Pellegrino, M., Savino, A. F., & Berzaghi, P. (2024). Assessment of extra virgin olive oil quality by miniaturized near infrared instruments in a rapid and non-destructive procedure. *Food Chemistry*, *430*, 137043.
4. Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: linear models. PLS-DA. *Analytical methods*, *5*(16), 3790-3798.
5. Ballabio, D., Grisoni, F., & Todeschini, R. (2018). Multivariate comparison of classification performance measures. *Chemometrics and Intelligent Laboratory Systems*, *174*, 33-44.
6. Bevilacqua, M., Bucci, R., Magrì, A. D., Magrì, A. L., & Marini, F. (2012). Tracing the origin of extra virgin olive oils by infrared spectroscopy and chemometrics: A case study. *Analytica chimica acta*, *717*, 39-51.
7. Bragolusi, M., Massaro, A., Zacometti, C., Tata, A., & Piro, R. (2021). Geographical identification of Italian extra virgin olive oil by the combination of near infrared and Raman spectroscopy: A feasibility study. *Journal of Near Infrared Spectroscopy*, *29*(6), 359-365.
8. Chaji, S., Bajoub, A., Cravotto, C., Voss, M., Tabasso, S., Hanine, H., & Cravotto, G. (2023). Metabolomics in action: Towards producing authentic virgin olive oil rich in bioactive compounds and with distinctive organoleptic features. *LWT*, 115681.
9. Commission Regulation of European Union. "Commission implementing regulation (EU) 2022/2105 of 29 July 2022." *Official Journal L* 284 (2022): 23-48.
10. Conte, L., Bendini, A., Valli, E., Lucci, P., Moret, S., Maquet, A., ... & Toschi, T. G. (2020). Olive oil quality and authenticity: A review of current EU legislation, standards, relevant methods of analyses, their drawbacks and recommendations for the future. *Trends in Food Science & Technology*, *105*, 483-493.
11. De Angelis, D., Summo, C., Pasqualone, A., Faccia, M., & Squeo, G. (2024). Advancements in food authentication using soft independent modelling of class analogy (SIMCA): a review. *Food Quality and Safety*, *8*, fyae032.
12. Du, Q., Zhu, M., Shi, T., Luo, X., Gan, B., Tang, L., & Chen, Y. (2021). Adulteration detection of corn oil, rapeseed oil and sunflower oil in camellia oil by in situ diffuse reflectance near-infrared spectroscopy and chemometrics. *Food Control*, *121*, 107577.
13. El Maouardi, M., De Braekeleer, K., Bouklouze, A., & Vander Heyden, Y. (2024). Comparison of Near-Infrared and Mid-Infrared spectroscopy for the identification and quantification of argan oil adulteration through PCA, PLS-DA and PLS. *Food Control*, *165*, 110671.

14. Gullifa, G., Barone, L., Papa, E., Giuffrida, A., Materazzi, S., & Risoluti, R. (2023). Portable NIR spectroscopy: The route to green analytical chemistry. *Frontiers in Chemistry*, *11*, 1214825.
15. Hennessy, S., Downey, G., & O′ Donnell, C. P. (2009). Confirmation of food origin claims by Fourier transform infrared spectroscopy and chemometrics: Extra virgin olive oil from Liguria. *Journal of Agricultural and Food Chemistry*, *57*(5), 1735-1741.
16. Hourant, P., Baeten, V., Morales, M. T., Meurens, M., & Aparicio, R. (2000). Oil and fat classification by selected bands of near-infrared spectroscopy. *Applied spectroscopy*, *54*(8), 1168-1174.
17. International Olive Council (IOC). World market of olive oil and table olives – Data from April 2024, accessible from https://www.internationaloliveoil.org/world-market-of-olive-oil-and-table-olives-data-from-april-2024/?lang=it (Accessed on 13th March 2025).
18. Jiménez-Carvelo, A. M., Lozano, V. A., & Olivieri, A. C. (2019). Comparative chemometric analysis of fluorescence and near infrared spectroscopies for authenticity confirmation and geographical origin of Argentinean extra virgin olive oils. *Food Control*, *96*, 22-28.
19. Jolayemi, O. S., Tokatli, F., Buratti, S., & Alamprese, C. (2017). Discriminative capacities of infrared spectroscopy and e-nose on Turkish olive oils. *European Food Research and Technology*, *243*, 2035-2042.
20. Klinar, M., Benković, M., Jurina, T., Jurinjak Tušek, A., Valinger, D., Tarandek, S. M., ... & Gajdoš Kljusurić, J. (2024). Fast Monitoring of Quality and Adulteration of Blended Sunflower/Olive Oils Applying Near-Infrared Spectroscopy. *Chemosensors*, *12*(8), 150.
21. Kontzedaki, R., Orfanakis, E., Sofra-Karanti, G., Stamataki, K., Philippidis, A., Zoumi, A., & Velegrakis, M. (2020). Verifying the geographical origin and authenticity of greek olive oils by means of optical spectroscopy and multivariate analysis. *Molecules*, *25*(18), 4180.
22. Laouni, A., El Orche, A., Elhamdaoui, O., Karrouchi, K., El Karbane, M., & Bouatia, M. (2023). A preliminary study on the potential of FT-IR spectroscopy and chemometrics for tracing the geographical origin of Moroccan virgin olive oils. *Journal of AOAC International*, *106*(3), 804-812.
23. Lozano, V. A., Carvelo, A. M. J., Olivieri, A. C., Kucheryavskiy, S. V., Rodionova, O. Y., & Pomerantsev, A. L. (2025). Authentication of Argentinean extra-virgin olive oils using three-way fluorescence and two-way near-infrared data fused with multi-block DD-SIMCA. *Food Chemistry*, 463, 141127.
24. Lozano-Castellón, J., López-Yerena, A., Domínguez-López, I., Siscart-Serra, A., Fraga, N., Sámano, S., ... & Pérez, M. (2022). Extra virgin olive oil: A comprehensive review of efforts to ensure its authenticity, traceability, and safety. *Comprehensive Reviews in Food Science and Food Safety*, *21*(3), 2639-2664.
25. Mannina, L., & Sobolev, A. P. (2011). High resolution NMR characterization of olive oils in terms of quality, authenticity and geographical origin. *Magnetic Resonance in Chemistry*, *49*, S3-S11.

26. Meng, X., Yin, C., Yuan, L., Zhang, Y., Ju, Y., Xin, K., ... & Hu, L. (2023). Rapid detection of adulteration of olive oil with soybean oil combined with chemometrics by Fourier transform infrared, visible-near-infrared and excitation-emission matrix fluorescence spectroscopy: A comparative study. *Food Chemistry*, *405*, 134828.
27. Mialon, N., Roig, B., Capodanno, E., & Cadiere, A. (2023). Untargeted metabolomic approaches in food authenticity: A review that showcases biomarkers. *Food chemistry*, *398*, 133856.
28. Mousa, M. A., Wang, Y., Antora, S. A., Al-Qurashi, A. D., Ibrahim, O. H., He, H. J., ... & Kamruzzaman, M. (2022). An overview of recent advances and applications of FT-IR spectroscopy for quality, authenticity, and adulteration detection in edible oils. *Critical Reviews in Food Science and Nutrition*, *62*(29), 8009-8027.
29. Opaluwa, C., De Angelis, D., Summo, C., & Karbstein, H. P. (2025). Effect of different vegetable oils on extruded plant-based meat analogs: Evaluation of oxidative degradation, textural, rheological, tribological and sensory properties. *Food Hydrocolloids*, *163*, 111038.
30. Quintanilla-Casas, B., Torres-Cobos, B., Guardiola, F., Servili, M., Alonso-Salces, R. M., Valli, E., ... & Tres, A. (2022). Geographical authentication of virgin olive oil by GC–MS sesquiterpene hydrocarbon fingerprint: Verifying EU and single country label-declaration. *Food Chemistry*, *378*, 132104.
31. Serranti, S., Cesare, D., Marini, F., & Bonifazi, G. (2013). Classification of oat and groat kernels using NIR hyperspectral imaging. *Talanta*, *103*, 276-284.
32. Squeo, G., Silletti, R., Napoletano, G., Greco Miani, M., Difonzo, G., Pasqualone, A., & Caponio, F. (2022). Characterization and effect of refining on the oil extracted from durum wheat by-products. *Foods*, *11*(5), 683.
33. Totaro, M. P., Squeo, G., De Angelis, D., Pasqualone, A., Caponio, F., & Summo, C. (2023). Application of NIR spectroscopy coupled with DD-SIMCA class modelling for the authentication of pork meat. *Journal of Food Composition and Analysis*, *118*, 105211.
34. Vanstone, N., Moore, A., Martos, P., & Neethirajan, S. (2018). Detection of the adulteration of extra virgin olive oil by near-infrared spectroscopy and chemometric techniques. *Food Quality and Safety*, *2*(4), 189-198.
35. Vieira, L. S., Assis, C., de Queiroz, M. E. L. R., Neves, A. A., & de Oliveira, A. F. (2021). Building robust models for identification of adulteration in olive oil using FT-NIR, PLS-DA and variable selection. *Food Chemistry*, *345*, 128866.
36. Voccio, R., Malegori, C., Oliveri, P., Branduani, F., Arimondi, M., Bernardi, A., ... & Cettolin, M. (2024). Combining PLS-DA and SIMCA on NIR data for classifying raw materials for tyre industry: A hierarchical classification model. *Chemometrics and Intelligent Laboratory Systems*, *250*, 105150.
37. Westad, F., Bevilacqua, M., & Marini, F. (2013). Regression. In *Data handling in science and technology* (Vol. 28, pp. 127-170). Elsevier.

38. Willenberg, I., Matthäus, B., & Gertz, C. (2019). A new statistical approach to describe the quality of extra virgin olive oils using near infrared spectroscopy (NIR) and traditional analytical parameters. *European Journal of Lipid Science and Technology*, *121*(2), 1800361.
39. Yuan, L., Meng, X., Xin, K., Ju, Y., Zhang, Y., Yin, C., & Hu, L. (2023). A comparative study on classification of edible vegetable oils by infrared, near infrared and fluorescence spectroscopy combined with chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *288*, 122120.
40. Zielińska, A., Wójcicki, K., Klensporf-Pawlik, D., Dias-Ferreira, J., Lucarini, M., Durazzo, A., ... & Nowak, I. (2020). Chemical and physical properties of meadowfoam seed oil and extra virgin olive oil: focus on vibrational spectroscopy. *Journal of Spectroscopy*, *2020*(1), 8870170.

Figure Captions:

Figure 1. NIR spectra of virgin olive oil samples.

Figure 2. Score and loadings plots generated by the principal component analysis of NIR spectra, computed on SNV+MC preprocessed data (a, b, respectively) and on SNV+GLSW+MC preprocessed data (c, d, respectively). SNV: Standard Normal Variate; GLSW: Generalized Least Squares Weighting; MC: mean centering.

Figure 3. IR spectra of virgin olive oil samples before (a) and after (b) Standard Normal Variate preprocessing.

Figure 4. Score and loadings plots generated by the principal component analysis of IR spectra, computed on SNV+MC preprocessed data (a, b, respectively) and on SNV+GLSW+MC preprocessed data (c, d, respectively).

Figure 5. PLS-DA score plots and VIP scores for NIR spectra processed using SNV, 1D, and GLSW. The score plot illustrates sample clustering based on geographical origin (a-c), while the VIP scores (d) highlight the most influential spectral regions in the classification model.

Figure 6. PLS-DA score plots and VIP scores for IR spectra processed using SNV, and GLSW. The score plot illustrates sample clustering based on geographical origin (a-c), while the VIP scores (d) highlight the most influential spectral regions in the classification model.

Figure 7. VIP scores of the PLS-DA models calculated using the chemical data preprocessed with autoscaling.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Bari, 20th March 2025

Dr. Davide De Angelis

On the behalf of all the authors

Highlights

- NIR and FT-IR spectroscopy effectively classify EVOO based on geographical origin
- Prediction performance in sensitivity and specificity were higher than 0.9
- Quality traits might influence origin classification models
- A prior quality assessment is recommended for reliable untargeted authentication

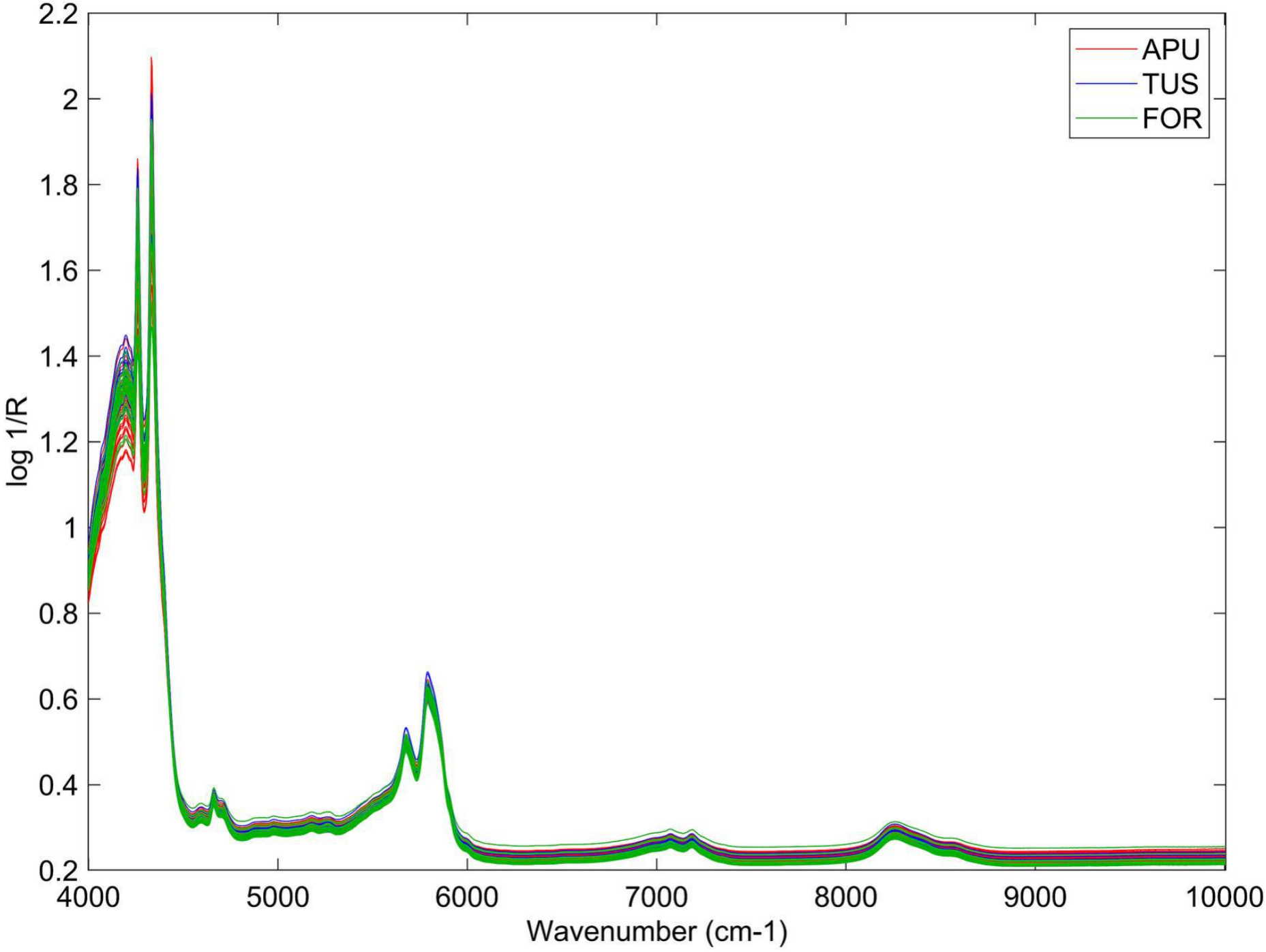
APU
TUS
FOR
log 1/R
Wavenumber (cm-1)
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
4000
5000
6000
7000
8000
9000
10000

Figure 1

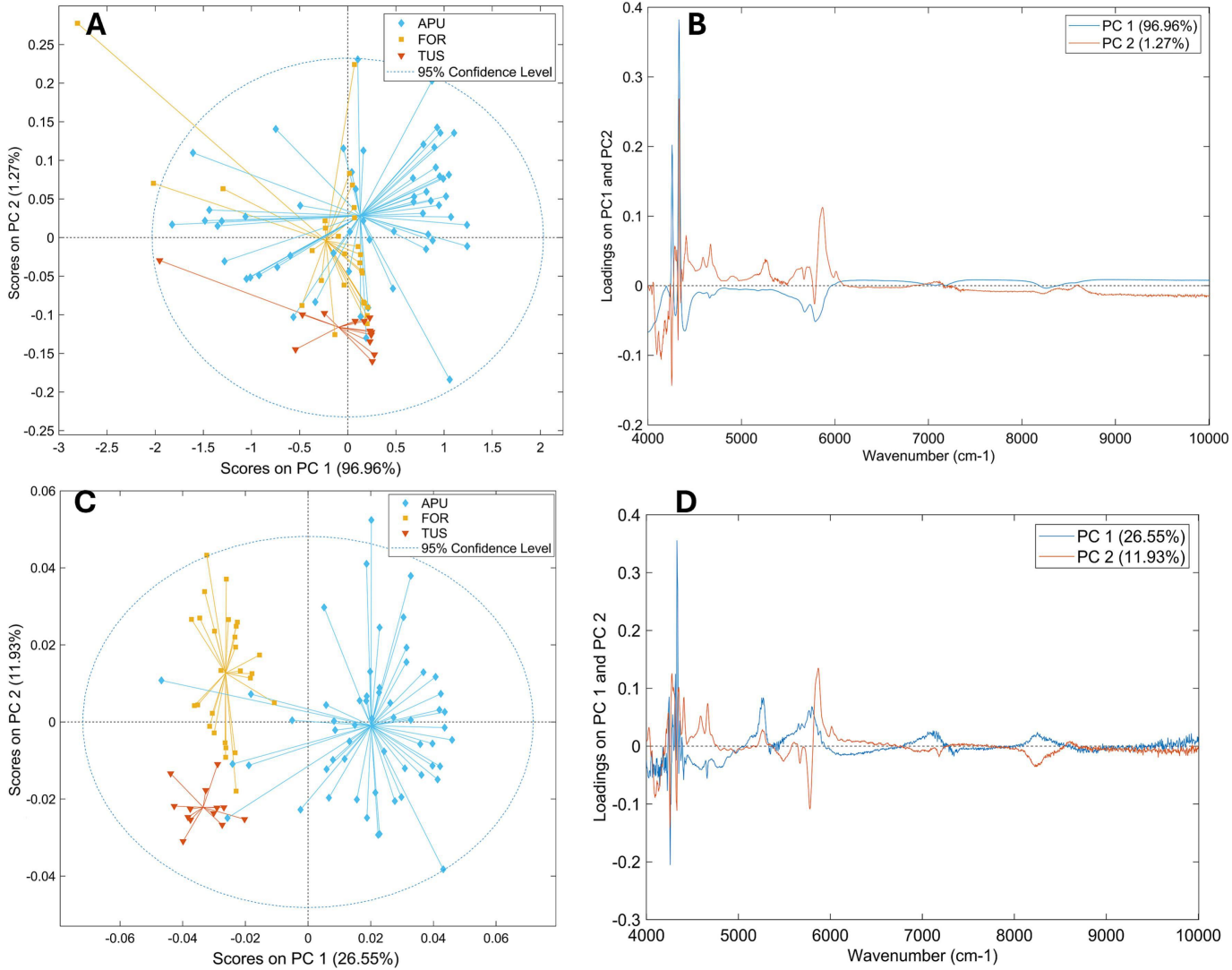
A
APU
FOR
TUS
95% Confidence Level
Scores on PC 2 (1.27%)
Scores on PC 1 (96.96%)
B
PC 1 (96.96%)
PC 2 (1.27%)
Loadings on PC1 and PC2
Wavenumber (cm-1)
C
APU
FOR
TUS
95% Confidence Level
Scores on PC 2 (11.93%)
Scores on PC 1 (26.55%)
D
PC 1 (26.55%)
PC 2 (11.93%)
Loadings on PC 1 and PC 2
Wavenumber (cm-1)

Figure 2

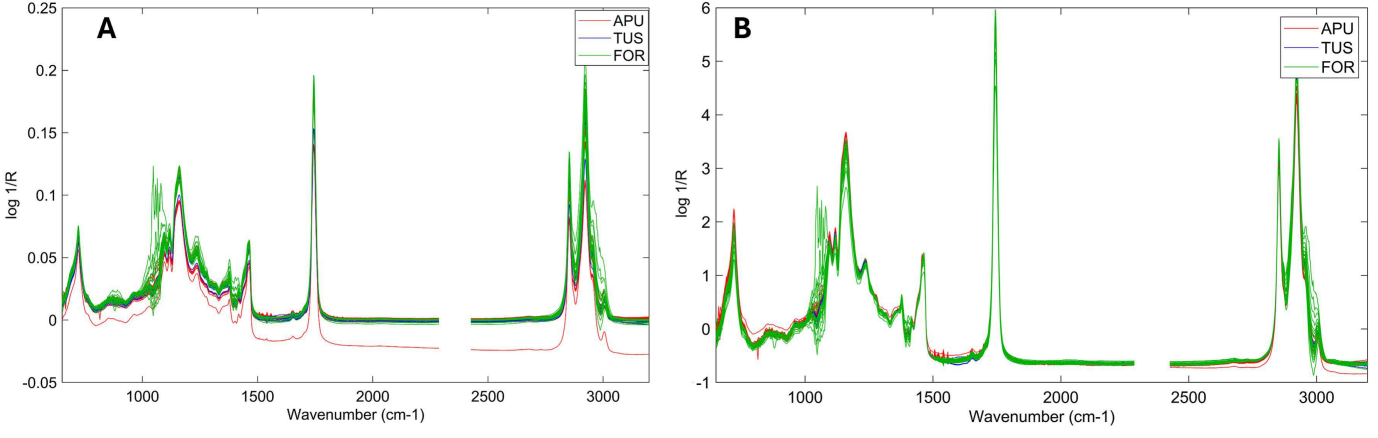
A
B
APU
TUS
FOR
log 1/R
Wavenumber (cm-1)
0.25
0.2
0.15
0.1
0.05
0
-0.05
6
5
4
3
2
1
-1
1000
1500
2000
2500
3000

Figure 3

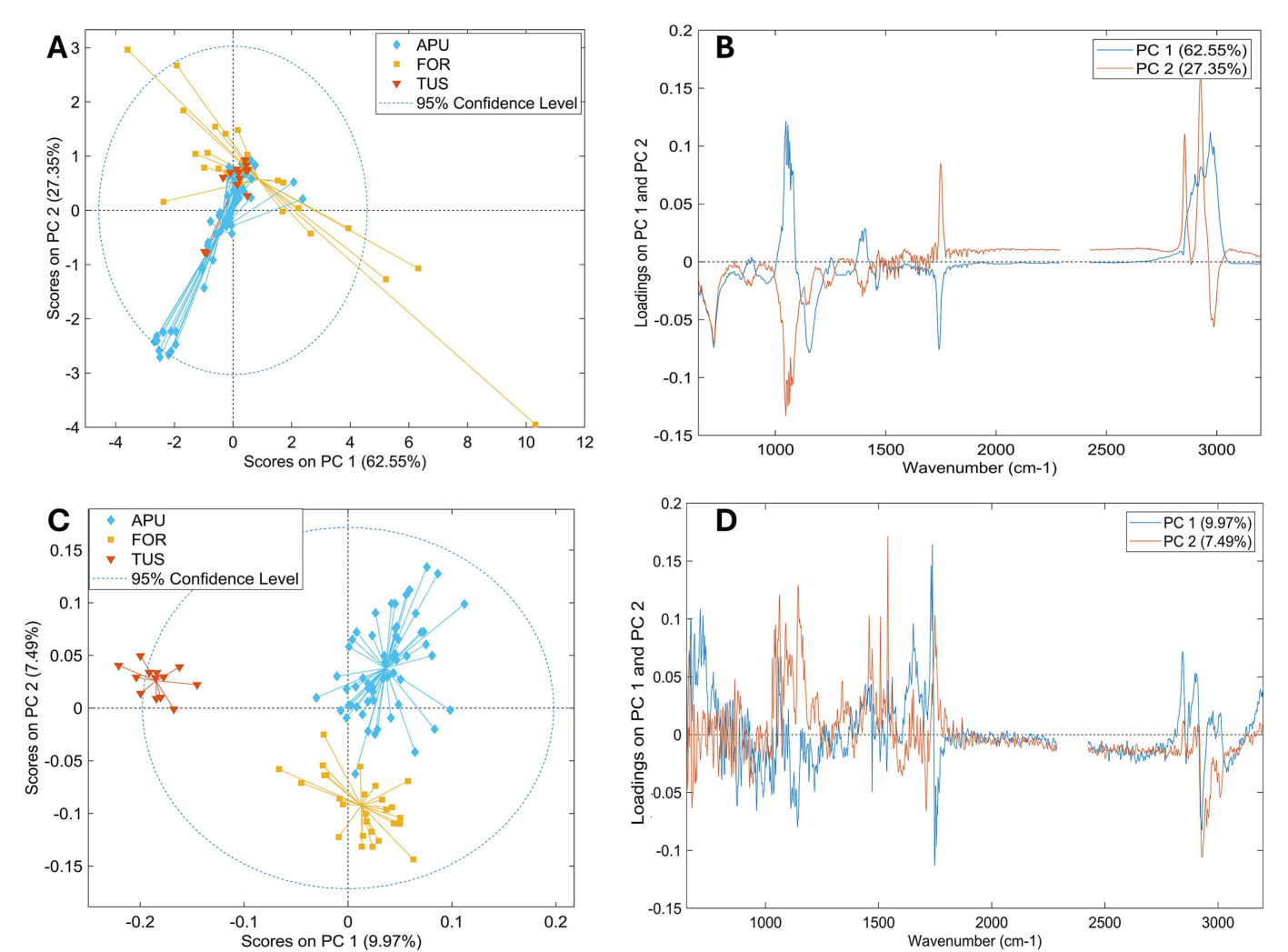
A
APU
FOR
TUS
95% Confidence Level
Scores on PC 2 (27.35%)
Scores on PC 1 (62.55%)
B
PC 1 (62.55%)
PC 2 (27.35%)
Loadings on PC 1 and PC 2
Wavenumber (cm-1)
C
APU
FOR
TUS
95% Confidence Level
Scores on PC 2 (7.49%)
Scores on PC 1 (9.97%)
D
PC 1 (9.97%)
PC 2 (7.49%)
Loadings on PC 1 and PC 2
Wavenumber (cm-1)

Figure 4

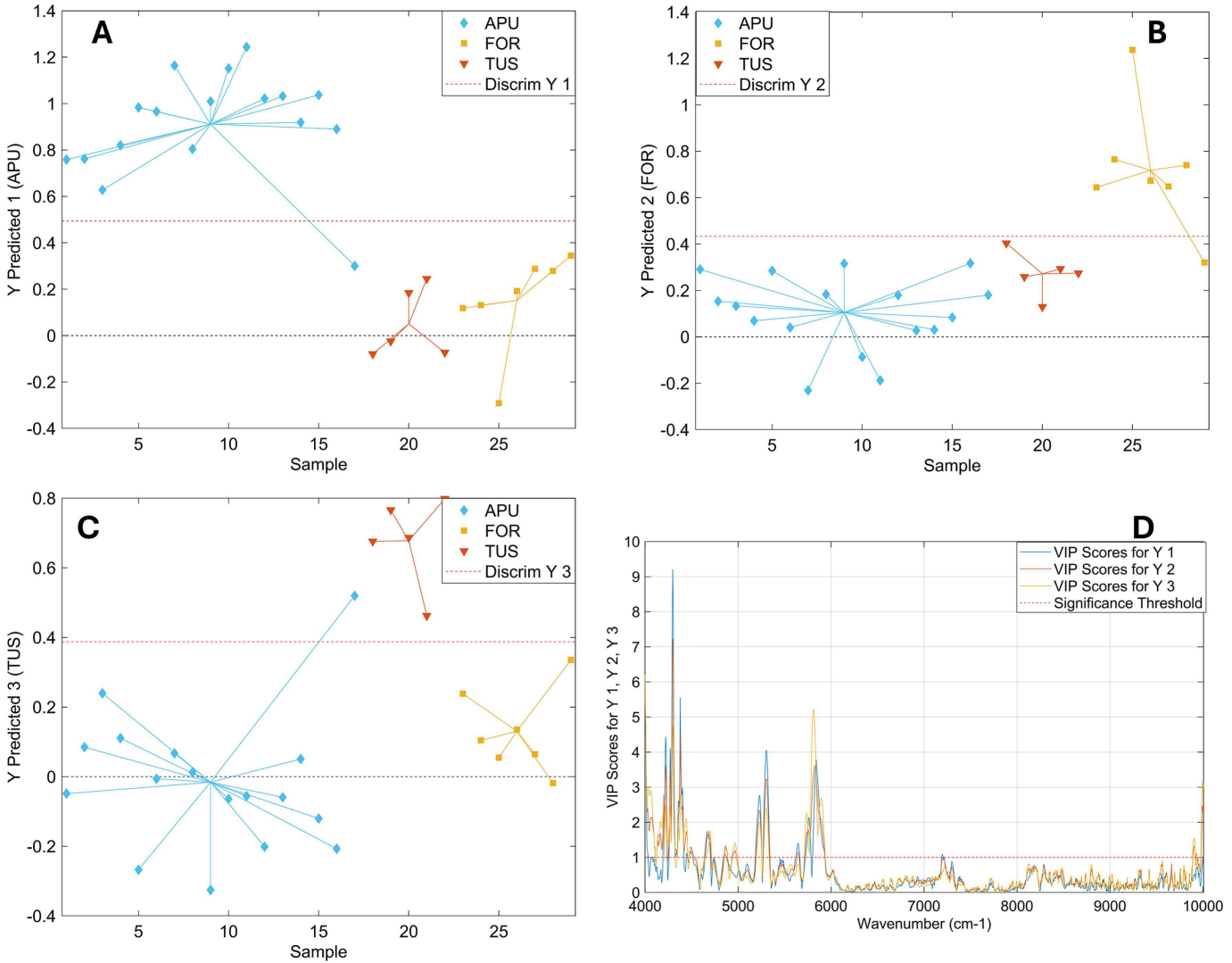
A
APU
FOR
TUS
Discrim Y 1
Y Predicted 1 (APU)
Sample
B
APU
FOR
TUS
Discrim Y 2
Y Predicted 2 (FOR)
Sample
C
APU
FOR
TUS
Discrim Y 3
Y Predicted 3 (TUS)
Sample
D
VIP Scores for Y 1
VIP Scores for Y 2
VIP Scores for Y 3
Significance Threshold
VIP Scores for Y 1, Y 2, Y 3
Wavenumber (cm-1)

Figure 5

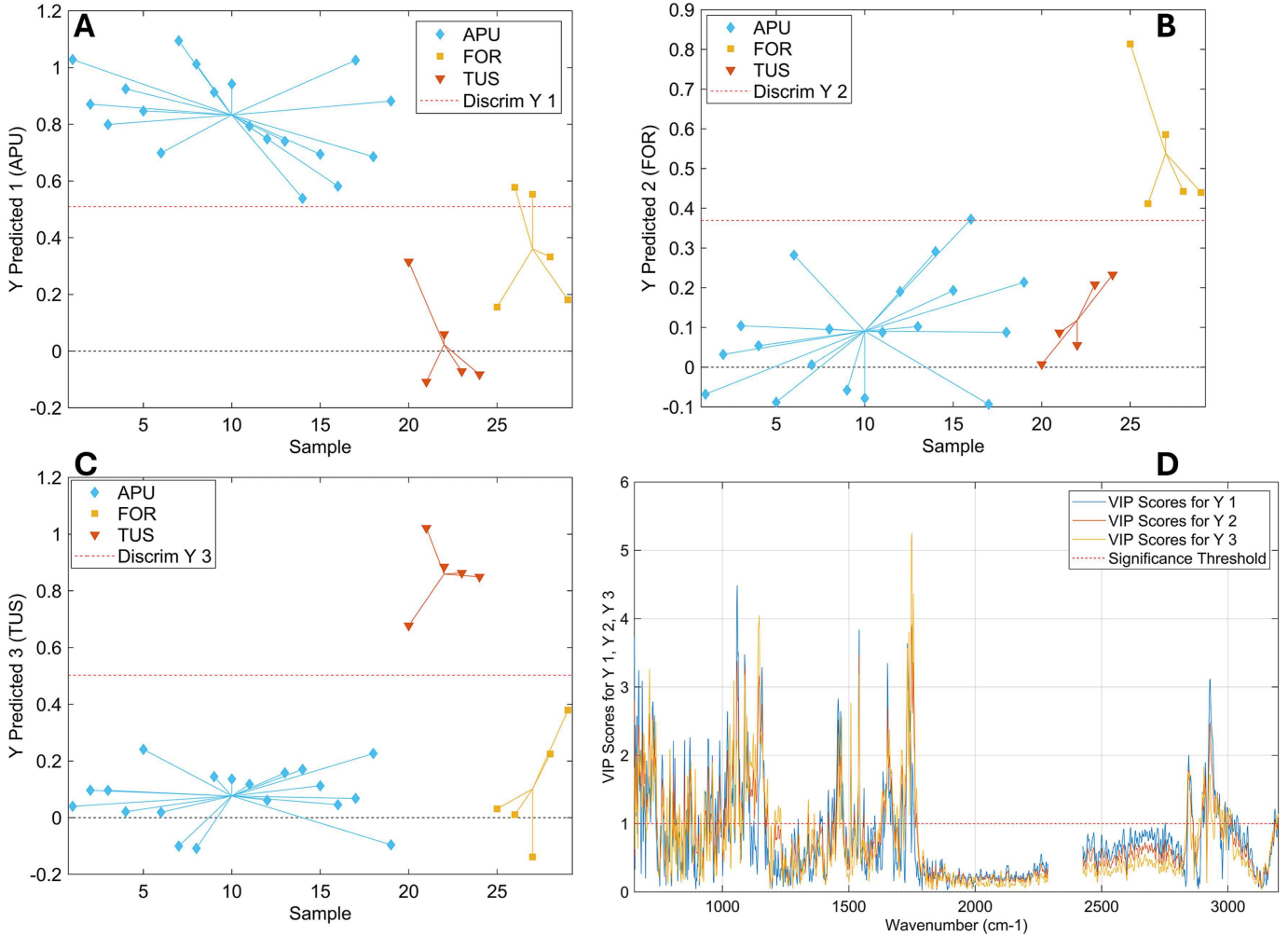
A
APU
FOR
TUS
Discrim Y 1
Y Predicted 1 (APU)
Sample
B
APU
FOR
TUS
Discrim Y 2
Y Predicted 2 (FOR)
Sample
C
APU
FOR
TUS
Discrim Y 3
Y Predicted 3 (TUS)
Sample
D
VIP Scores for Y 1
VIP Scores for Y 2
VIP Scores for Y 3
Significance Threshold
VIP Scores for Y 1, Y 2, Y 3
Wavenumber (cm-1)

Figure 6