Short note

Multivariate analysis of biophenols content in extra-virgin olive oils of Italian and European origin

Antonio Perna✉
Pierangela Rovellini

Innovhub Stazioni Sperimentali per l'Industria
Via Giuseppe Colombo, 79
Milan, Italy

✉CORRESPONDING AUTHOR
Antonio Perna
e-mail: antonio.perna@mi.camcom.it
Tel: +39 02 8515 3582

Received: January 28, 2025
Accepted: May 22, 2025

The aim of this work was to apply multivariate methods to biophenol content in extra-virgin olive oils of Italian and European origin in order to screen datasets and, if possible, give some consideration to their origin and correlation between variables.
Over the years, Innovhub has analysed many extra-virgin oil samples, recording the biophenol content and chemical profile, which was useful for the construction of a wide database.
We have investigated a possible distinction between European and Italian extra-virgin oils by considering their biophenol content. For this purpose, we used Principal Component Analysis (PCA), a multivariate method useful for an explorative analysis of the data.
In the results, PCA made an effective distinction between Italian and European extra-virgin olive oils, allowing us to observe interesting aspects of the distribution of the oils along PC1 and PC2 and their relationship to some biophenol variables.

Keywords: Biophenols, Multivariate Analysis, PCA, EVO

1. INTRODUCTION

Over the years, Innovhub has recorded a great deal of data from the biophenol profile analysis. Phenolic compounds are important for the quality of virgin olive oil because of their contribution to the oil's flavour, aroma and nutritional properties. The most important phenolic compounds in extra-virgin olive oil are phenolic acids, alcohols, secoiridoids, derivatives lignans and flavonoids. Those data are stored in a database that is updated every year. We grouped data from 2013 to 2023, dividing them in two main groups: oils from Italy and those from other European different countries. Samples arrived at Innovhub already labelled, so we already knew whether the oil was Italian or European. Our aim was to seek out a possible a distinction between the two groups by considering the biophenol profile and content analysed. We investigated the data using a multivariate method analysis called Principal Component Analysis. PCA is an unsupervised analysis and a dimensionality reduction method. It is a powerful technique because it allows new information to be obtained about the system analysed simply by projecting the actual variables (which describe the overall experiment) in a new cartesian plane where, on the x axis, we find Principal Component 1 (PC1) and, on the y axis, Principal component 2 (PC2). The number of Principal components (PC) is usually equal to or less than the number of variables.

2. MATERIAL AND METHODS

2.1. DETERMINATION OF PHENOLIC COMPOUNDS BY HPLC

The method was based on direct extraction of the biophenolic minor polar compounds from extra virgin olive oils by means of a methanol solution and subsequent quantification by HPLC with the aid of a UV detector at 280 nm [2]. Syringic acid was used as the internal standard.
The content of the natural and oxidised oleuropein and ligstroside derivatives, lignans, flavonoids and phenolic acids are expressed in mg of tyrosol/kg. A solution of internal standard (1 ml of 0.015 mg $ml^{-1}$of syringic acid in water/MeOH 20:80 v/v) was added to a sample of virgin olive oil (2 g). The mixture was shaken (30 s) and 5 mL of the extraction solution containing water and MeOH (20:80 v/v) was added. The mixture obtained was shaken for 1 min, extracted for 15 min in an ultrasonic bath and then centrifuged at 5000 rpm (2500 g) for 25 min at T=20°C. The upper phase was filtered using a 0.45 µm PVDF syringic filter, and then 20 µL of the filtered solution was analysed by HPLC with a UV detector at 280 nm. The HPLC system consisted of a C18 Spherisorb ODS-2 reverse column (5 µm, 250 mm x 4.6 mm). Elution was performed at flow rate of 1 ml min-1 following a gradient composed of a mixture of water and orthophosphoric acid (99.8:0.2 v/v) (Solvent A), MeOH (Solvent B), and acetonitrile (Solvent C): from 96% (A)-2% (B)-2% (C) to 0% (A)-50% (B)-50% (C) in 60 min. The last gradient was kept for 10 min. The successive gradient was: from 0% (A)-50% (B)-50% (C) to 96% (A)-2% (B)-2% (C) at 2 min and then kept for 10 min. [3,4]

2.2. MULTIVARIATE STATISTICAL METHODS: PRINCIPAL COMPONENT ANALYSIS

PCA was performed using a correlation matrix. The number of samples analysed for biophenol profile content was 179, divided in 124 IEO (Italian Extra virgin Oil) and 55 EEO (European Extra virgin Oil) [Table I: https://drive.google.com/drive/folders/1jseDbqRfZ_t9-sZCwKbI1xsuHLxzaAFx?usp=sharing]. These oils were analysed with regard to all the parameters present in the Reg. UE n. 2022/2104 and were found to be in compliance. The variables used for the PCA analysis were 11: tyrosol, hydroxytyrosol, oleocanthal, total lignans, phenolic acids, luteolin, apigenin, elenolic acid, decarboxy methyl elenolic acid, oleuropein derivatives and ligstroside derivatives.
As regards the last two variables, we removed the hydroxytyrosol and tyrosol contribution since they were already present as variables.
PCA was performed using JMP 17(SAS Institute)

3. RESULTS AND DISCUSSION

3.1. PCA RESULT

PCA performed on the Biophenols dataset led to some important consideration.
Figure 1 shows the summary and screen plot for PCA. As we can see, the first 4 principal components explain around 80% of the variance, which is sufficient to make a classification. Figure 2 shows the score plot for PC1 and PC2. In this picture, blue points represent EEO while red points represent IEO. IEO are clearly distinguished on PC1 from EEO, where we find an overlap of some IEO. Usually, IEO have a higher biophenol content than EEO. The latter may contain oil from Italian olives, which could explain why some EEO oil also overlaps to the IEO. This overlap is also explained by the content of three variables (Ligstroside Derivatives, Oleocanthal, Oleuropein Derivatives), which can be low in some IEO similar to EEO. In fact, as we can see in Figure 3, the most important variables that impact on the discrimination between EEO and IEO on the first PC1 are Ligstroside Derivatives, Oleocanthal, Oleuropein Derivatives. Furthermore, a strong correlation is noted between those three variables, which is also seen in the correlation matrix and correlation plot (Figure 4 and 5). It is well known that Oleocanthal is a ligstroside derivatives, so this high correlation value (0.96) is explained by their biochemical pathway. It could also be explained by the fact that those oils probably came from cultivars of Southern Italy. Unfortunately, no further information about the region of origin is known to us and we only know they are Italian due to the labelling, which did not provide any additional information. Let's focus now on PC2.

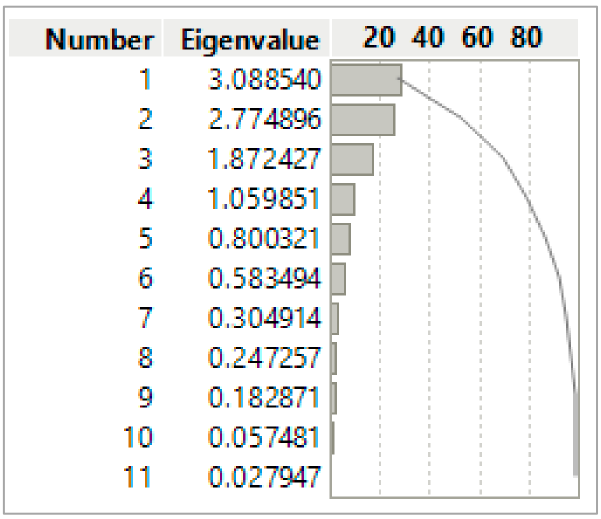

Figure 1 - Summary Plot of PCA. This diagram shows what is the contribution of the single variables to the overall variance of the system

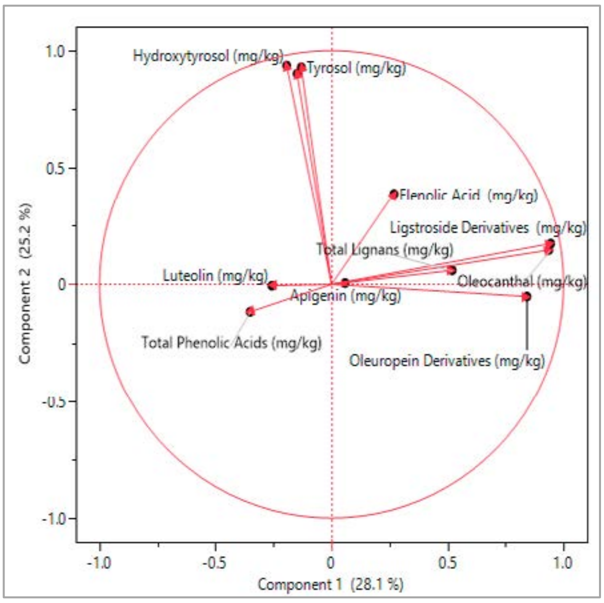

Figure 2 - Score Plot of PC1 vs PC2. Blue points represent EEO while Red Points represent IEO

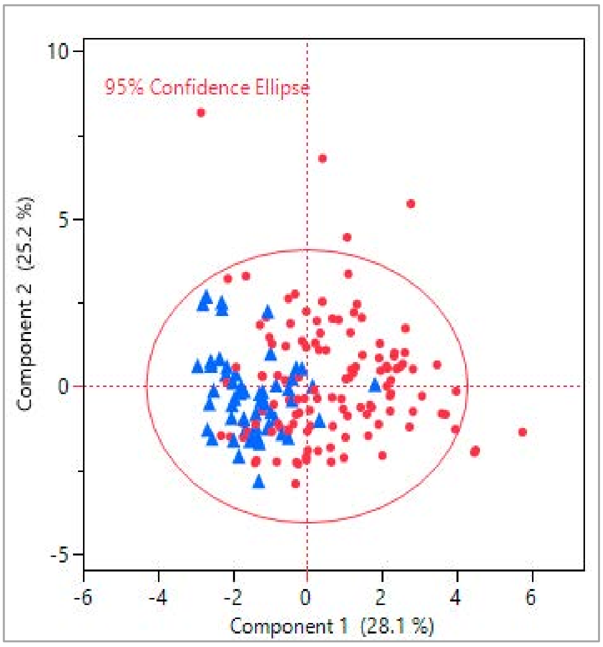

Figure 3 - Loading Plot of PC1 vs PC2

	Hydroxytyrosol (mg/kg)	Tyrosol (mg/kg)	Oleuropein Derivatives (mg/kg)	Ligstroside Derivatives (mg/kg)	Luteolin (mg/kg)	Apigenin (mg/kg)	Oleocanthal (mg/kg)	Decarboxymethylelenolic acid (mg/kg)	Elenolic Acid (mg/kg)
Hydroxytyrosol (mg/kg)	1.0000	0.8785	-0.1207	-0.0403	0.1055	0.0764	-0.0801	0.8099	0.3407
Tyrosol (mg/kg)	0.8785	1.0000	-0.2095	0.0642	-0.0650	-0.0287	0.0359	0.7971	0.1808
Oleuropein Derivatives (mg/kg)	-0.1207	-0.2095	1.0000	0.7695	-0.0028	0.2089	0.7148	-0.1902	0.3071
Ligstroside Derivatives (mg/kg)	-0.0403	0.0642	0.7695	1.0000	-0.2600	0.0392	0.9644	0.0461	0.2089
Luteolin (mg/kg)	0.1055	-0.0650	-0.0028	-0.2600	1.0000	0.6064	-0.2672	0.0018	0.1443
Apigenin (mg/kg)	0.0764	-0.0287	0.2089	0.0392	0.6064	1.0000	0.0209	-0.0731	0.1190
Oleocanthal (mg/kg)	-0.0801	0.0359	0.7148	0.9644	-0.2672	0.0209	1.0000	0.0346	0.1906
Decarboxymethylelenolic acid (mg/kg)	0.8099	0.7971	-0.1902	0.0461	0.0018	-0.0731	0.0346	1.0000	0.1794
Elenolic Acid (mg/kg)	0.3407	0.1808	0.3071	0.2089	0.1443	0.1190	0.1906	0.1794	1.0000

Figure 4 - Correlation Matrix of variables estimated by row-wise-method

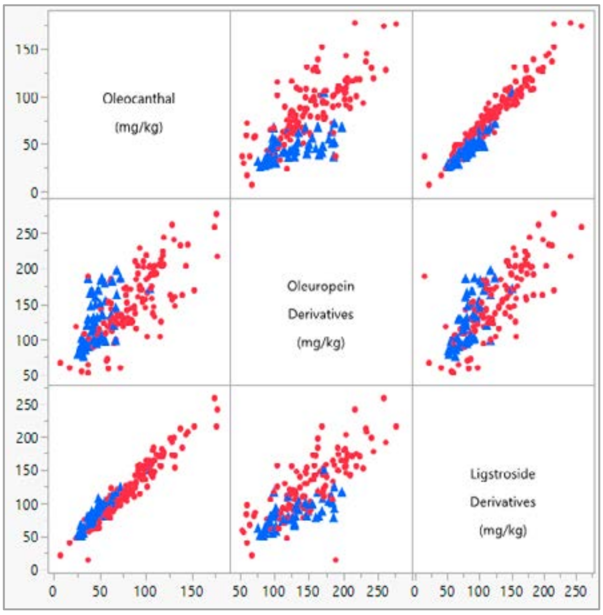

Figure 5 - Correlation Plot of Oleocanthal, Oleuropein Derivatives, Ligstroside Derivatives estimated by row-wise-method

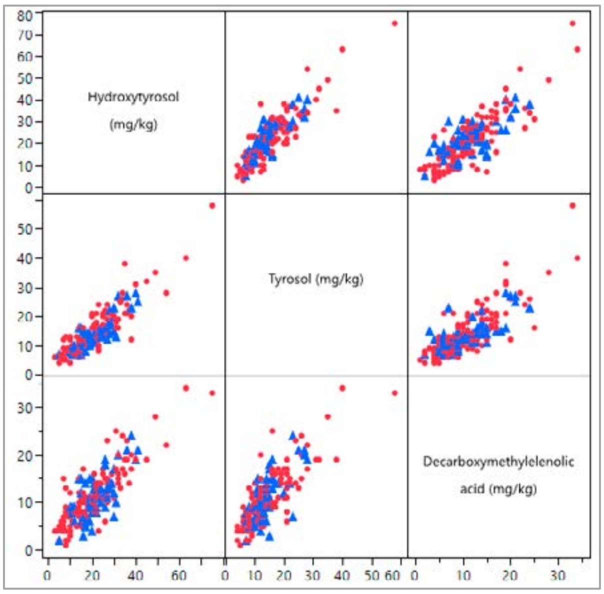

Figure 6 - Correlation Plot of Hydroxytyrosol, Tyrosol, Decarboxymethylelenolic acid estimated by row-wise-method

No clear distinction is seen in IEO and EEO on PC2 compared to PC1. Furthermore, we observe that three variables (tyrosol, hydroxytyrosol, decarboxy methyl elenolic acid) are strongly correlated, even if they don't make a distinction on PC2 between the two classes of oils. This is confirmed in Figure 4 and 6, where there is another indication regarding the correlation of these three variables. The possible explanation for the high correlation could be linked to the high hydrolysis ratio for some oils in the upper part of the biplot. In fact, when an oil is highly hydrolysed, the content of tyrosol, hydroxytyrosol and methyl elenolic acid is proportionally increased. Furthermore, we note a big intragroup variance within the Italian group, due to the different origin and blend of cultivars mixed in those samples.

4. CONCLUSION

In conclusion, we can state that there is a clear separation between EEO and IEO, even if some Italian oils overlap with European, since the former could have a biophenol content (due to their cultivar and origin) similar to European oils. This separation is explained by the contribution of Ligstroside Derivatives, Oleocanthal, Oleuropein Derivatives, highly inter-correlated. Another interesting aspect is the correlation of tyrosol, hydroxytyrosol and methyl elenolic acid, related to those oils which could have a high hydrolysis ratio. We observed a large intragroup variance within the Italian group. In fact, many of those oils come from different parts of Italy and have different biophenol content depending on their cultivar, origin of the olives, any blending of multiple cultivars. Further studies must focus on gathering the biophenol content of mono cultivar extra-virgin oils in order to build a database for each cultivar, and then analyse the possible distribution of each cultivar among the other Italian mono cultivars. Ultimately, it could be interesting to analyse the mono cultivar database of Italian oils versus the European version. This will be our next project.

REFERENCES

[1] Esalami, S.M, Dimic E.B, Rabrenovic, B.B. Phytochemical profile and antioxidant capacity of a virgin olive oil obtained from the olive cultivar 'Roghiani' from different regions of northern Libya. Grasas y Aceites 69: 2-10, (2018).

[2] IOC, 2022 Determination of Biophenols in olive oils by HPLC. International Olive Council. COI/T.20/Doc.No29/Rev.2 June 2022

[3] P. Rovellini, Indice di qualità dell'olio extra vergine di oliva, antiossidanti naturali e stato di ossidazione. Riv. Ital. Sostanze Grasse, Vol 81(4), 335-341, (2004).

[4] P. Rovellini, P.Fusari, S.Venturini- Indice di qualità salutistico nutrizionale dell'olio extra vergine di oliva novello- Controllo dello stato di ossidazione lipidica e valorizzazione del contenuto in micronutrienti ad alta attività antiossidante nella definizione globale della qualità e nella previsione di stabilità. Riv. Ital. Sostanze Grasse 89(4), 221-230, (2012)

[5] James N.Miller, Jane C.Miller and Robert D. Miller, Statistics and Chemometrics for Analytical Chemistry, Pearson, Seventh Edition (2018)