A.D. 1308
unipg
DIPARTIMENTO
DI SCIENZE AGRARIE,
ALIMENTARI E AMBIENTALI

OLI E GRASSI ALIMENTARI: INNOVAZIONE E
SOSTENIBILITA' NELLA PRODUZIONE E NEL CONTROLLO

Perugia, 15 - 17/06/2022

SISSG
SOCIETÀ ITALIANA PER LO STUDIO DELLE SOSTANZE GRASSE

# Geographical characterization of virgin olive oils by headspace volatile compounds analysis (HS-SPME-GC-MS).

Selvaggini R., Urbani S., Taticchi A., Esposto S., Gianluca V., Sordini B., Daidone L., Servili M.

www.unipg.it

roberto.selvaggini@unipg.it

# EXTRA VIRGIN OLIVE OIL (EVOO)

Extra virgin olive oil (EVOO) is one of the products that best represents the tradition and culture of Mediterranean Countries and is one of the main products of the agri-food sector.

In addition to being one of the key foods of the Mediterranean diet, due to its nutritional and health characteristics, it also plays a major role in the commercial and economic aspect of the producing countries.

Due to the high importance it has on both the nutritional and commercial aspects, olive oil can be subjected to adulteration, for this the European Union pays particular attention to this problem.

# TRACEABILITY OF EXTRA VIRGIN OLIVE OIL (EVOO)

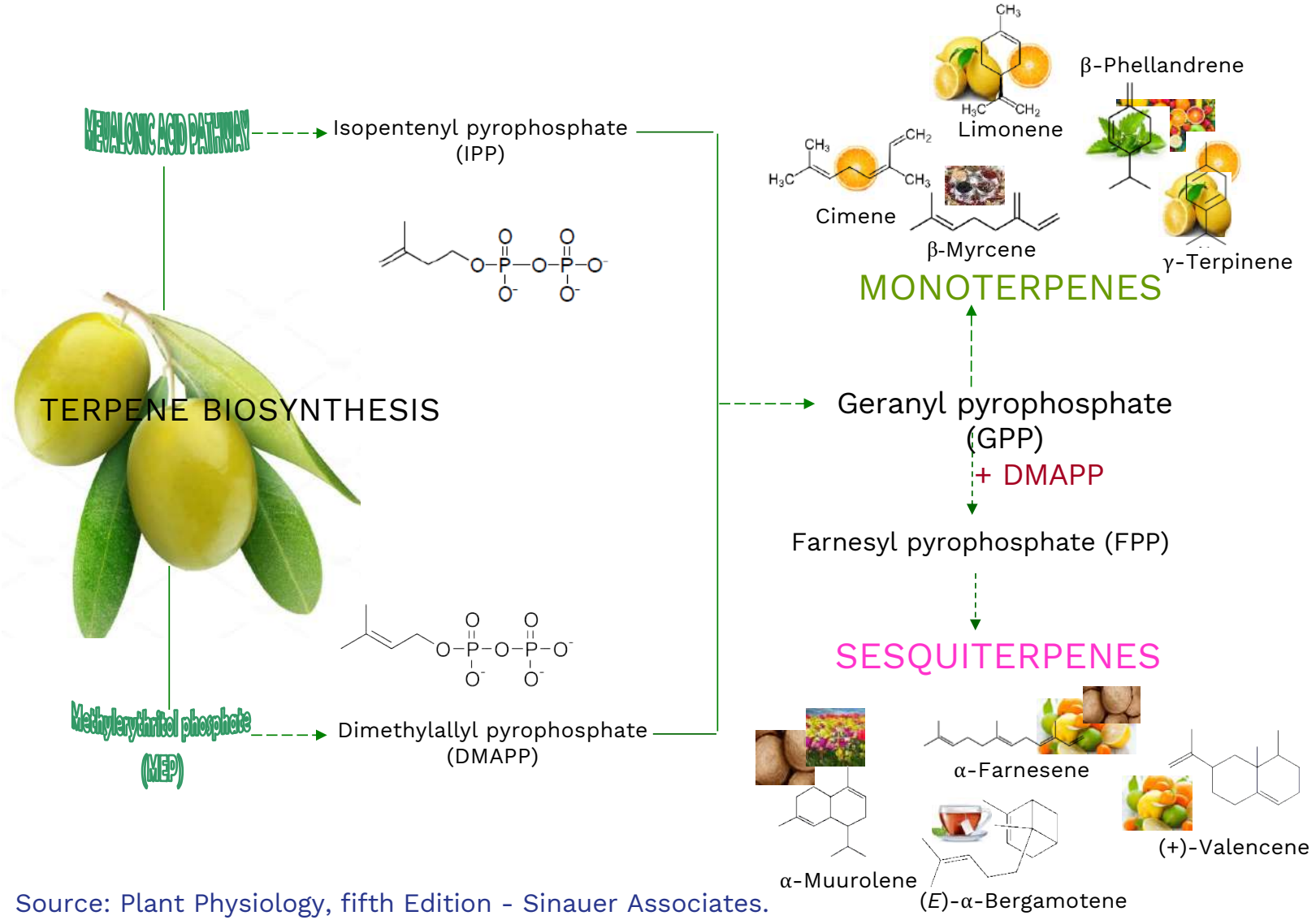
Great variability of EVOO composition in minor compounds due to genetic, agronomic, pedoclimatic, technological, and conservation factors.

Demand by consumers of high quality and authenticity extra virgin olive oils (EVOOs).

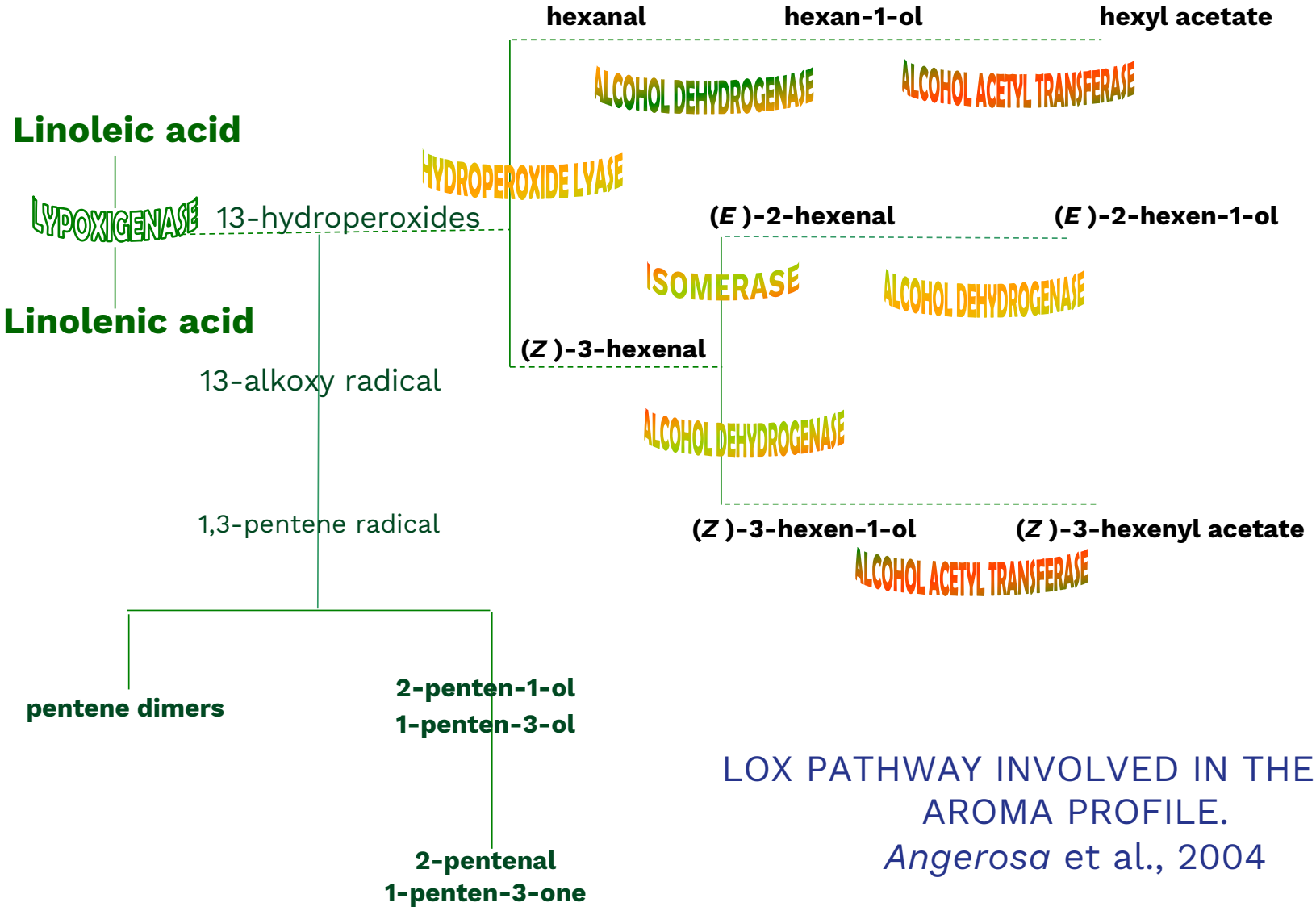
How to determine the geographical origin of EVOOs

- ✓ Analysis of stable isotopes
- ✓ Nuclear Magnetic Resonance (NMR)
- ✓ Near infrared spectroscopy (NIR)
- ✓ DNA analysis
- ✓ Headspace volatile compounds

# BIOSYNTHETIC PATHWAYS OF THE PRIMARY AROMAS IN THE EVOO




Source: Plant Physiology, fifth Edition - Sinauer Associates.




LOX PATHWAY INVOLVED IN THE EVOO AROMA PROFILE.
*Angerosa* et al., 2004

# MATERIALS AND METHODS

## List of EVOO samples

| Region/ Country (2018-2019) | No. |
|---|---|
| Apulia | 31 |
| Calabria | 21 |
| Tuscany | 15 |
| Sicily | 14 |
| Lombardy | 13 |
| Liguria | 6 |
| Sardinia | 2 |
| Greece | 9 |
| **Total number of samples** | **111** |

| Region/ Country (2019-2020) | No. |
|---|---|
| Apulia | 33 |
| Calabria | 36 |
| Sicily | 22 |
| Basilicata | 4 |
| Sardinia | 2 |
| Italy | 22 |
| Greece | 31 |
| Spain | 27 |
| Tunisia | 12 |
| Portugal | 5 |
| **Total number of samples** | **194** |

## List of variables analysed by targeted HS-SPME-GC-MS

**Aldehydes**

Pentanal
(E)-2-Pentenal
Hexanal
(E)-2-Hexenal
(E,E)-2,4-Hexadienal
Benzaldehyde

**Alcohols**

1-Pentanol
1-Penten-3-ol
(E)-2-Penten-1-ol
(Z)-2-Penten-1-ol
1-Hexanol
(E)-2-Hexen-1-ol
(E)-3-Hexen-1-ol
(Z)-3-Hexen-1-ol
Phenylethyl alcohol
Benzyl alcohol

**Esters**

Hexyl acetate
(Z)-3-Hexenyl acetate
Methyl benzoate
Methyl salicylate

**Ketones**

3-Pentanone
1-Penten-3-one
6-Methyl-5-hepten-2-one

**Hydrocarbons**

3,4-Diethyl-1,5-hexadiene
3,4-Diethyl-1,5-hexadiene (i)
3-Ethyl-1,5-octadiene
3-Ethyl-1,5-octadiene (i)
3,7-Decadiene (i1)
3,7-Decadiene (i2)
3,7-Decadiene (i3)

**Monoterpenes**

D-Limonene
alpha-Pinene
Sabinene
beta-Myrcene
3-Carene
Camphene
alpha-Phellandrene
beta-Phellandrene
beta-Ocimene
p-Cymene

**Sesquiterpenes**

alpha-Muurolene
alpha-Farnesene
Cyclosativene
alpha-Copaene

# MATERIALS AND METHODS

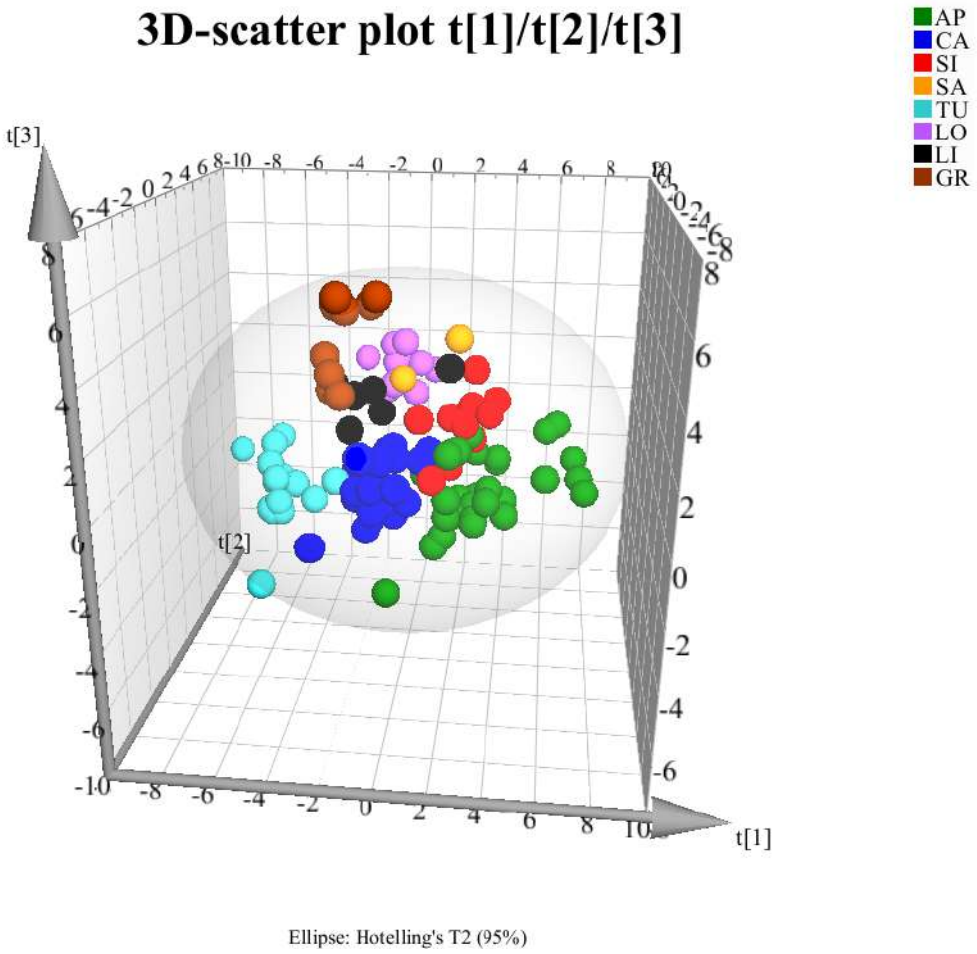
• The volatile compounds analysis was carried by HS-SPME-GC-MS (Taticchi *et al.,* 2021)*.

•The Multivariate Data Analysis (MVDA) was carried out using the partial least squares discriminant analysis (PLS-DA), the orthogonal partial least squares discriminant analysis (OPLS-DA), and the class modelling by means of Soft Independent Modelling by Class Analogy (SIMCA) methods. OPLS separates the systematic variation in X into two parts, one that is linearly related (and, therefore, predictive) to Y and one that is unrelated (orthogonal) to Y. OPLS-DA further enhances group separation.

The data were previously transformed logarithmically using the logarithm function in base 10 to obtain a distribution that was as close as possible to the Gaussian one. For the MVDA data were first normalized, for this reason variables were mean-centred and scaled to unit variance (autoscaling) in order to give all the variables the same importance.
Cross-validation was used to establish the number of significant components in the definition of the model.
MVDA was carried out using the chemometric package SIMCA v. 13.0.3.0 (Umetrics AB, Umeå, Sweden).

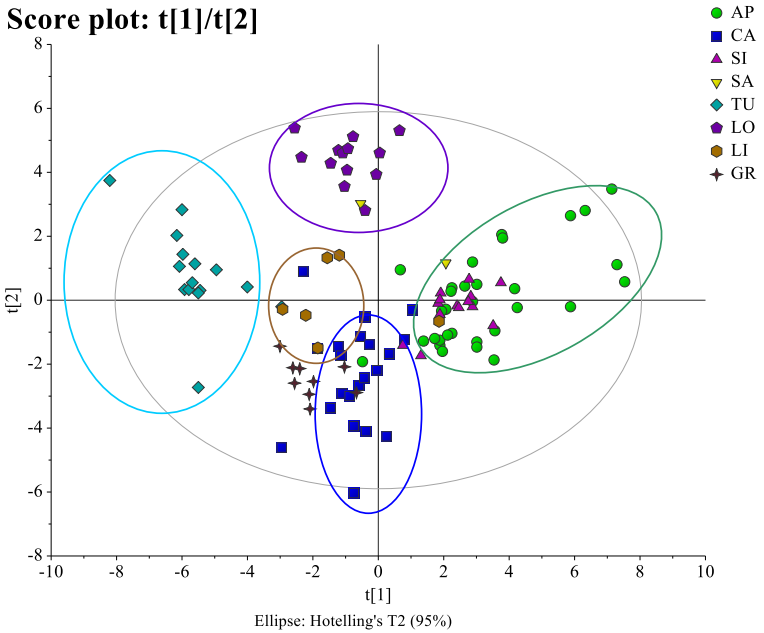
* Taticchi, A., Esposto, S., Veneziani, G., Minnocci, A., Urbani, S., Selvaggini, R., Sordini, B., Daidone, L., Sebastiani, L., Servili, M. (2021) High vacuum-assisted extraction affects virgin olive oil quality: Impact on phenolic and volatile compounds. Food Chemistry, 342, 128369. DOI: 10.1016/j.foodchem.2020.128369.

# PLS-DA model of EVOOs of the olive oil campaign 2018-2019



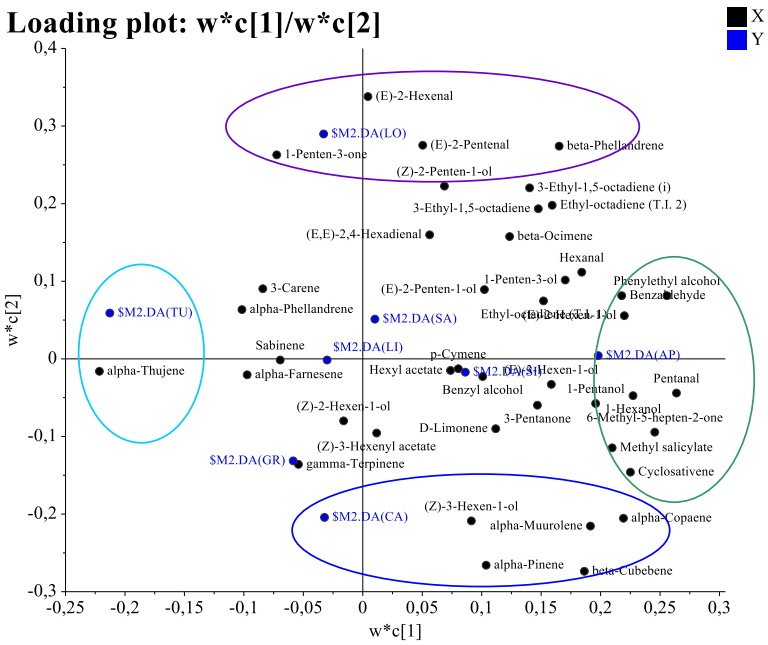

3D-scatter plot ($t_1/t_2/t_3$) of the first three latent variables of the **PLS-DA** model of EVOO samples of the olive oil campaign 2018-2019 of some Italian regions and Greece built with 111 samples (AP = Apulia; CA = Calabria; SI = Sicily; SA = Sardinia; TU = Tuscany; LO = Lombardy; LI = Liguria; GR = Greece) and **all the variables**. The model explains 73% of the total variance of X with eight significant latent variables (each explaining 25%, 14%, 10%, 6%, 6%, 5%, 5%, and 2%, respectively). In the 3D-scatter plot there is a more than good discrimination of the samples with separate clusters relating to Tuscan, Lombard, Apulian, Calabrian and Greek oils, showing the Ligurian ones in the central area of the plot.

# PLS-DA model of EVOOs of the olive oil campaign 2018-2019

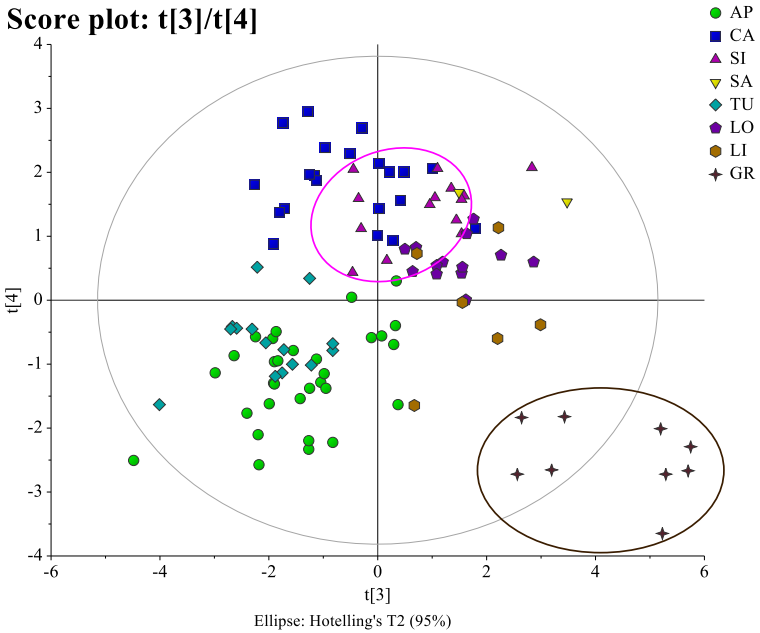



In the score plot is visible a more than good discrimination of the samples with separate clusters of the oils from Tuscany, Lombardy, Apulia, Calabria and Greece, showing the Liguria ones in the central area of the plot. The variables responsible of this differentiation are for first component on the left mainly α-thujene, and on the opposite side phenylethyl alcohol, 6-methyl-5-hepten-2-one, cyclosativene, (*E*)-2-hexen-1-ol, and benzaldehyde; in the second component in the top (*E*)-2-hexenal, 1-penten-3-one and (*E*)-2-pentenal, and β fellandrene and at the bottom some terpenes and (*Z*)-3-hexen-1-ol.

Score plot and loading plot of the first two latent variables of the **PLS-DA** model of EVOO samples of the olive oil campaign 2018-2019 of some Italian regions and Greece built with 111 samples (AP = Apulia; CA = Calabria; SI = Sicily; SA = Sardinia; TU = Tuscany; LO = Lombardy; LI = Liguria; GR = Greece) and **all the variables**.

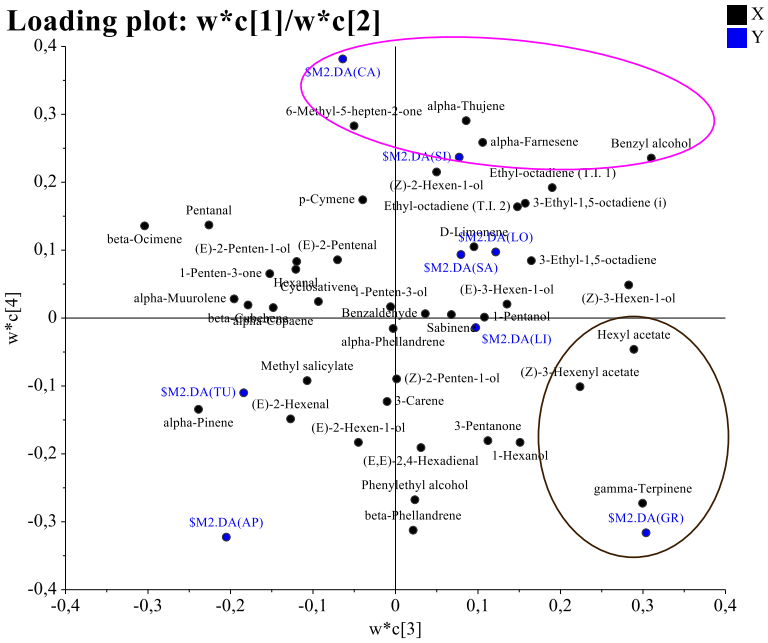


# PLS-DA model of EVOOs of the olive oil campaign 2018-2019




In the score plot $t_3/t_4$ a clear distinction is observed between the Greek samples, located in the lower right part, while in the upper right we find the Sicilian ones, discreetly separated from the others and in particular differentiated from the Apulian oils. The Greeks EVOOs have high contents of γ-terpinene (*Z*)-3-hexenyl acetate, and hexyl acetate, while the Sicilian ones have high concentrations of benzyl alcohol, 6-methyl-5-hepten-2-one, and α-thujene.

Score plot and loading plot of the third and the fourth latent variables of the **PLS-DA** model of EVOO samples of the olive oil campaign 2018-2019 of some Italian regions and Greece built with 111 samples (AP = Apulia; CA = Calabria; SI = Sicily; SA = Sardinia; TU = Tuscany; LO = Lombardy; LI = Liguria; GR = Greece) and **all the variables**.

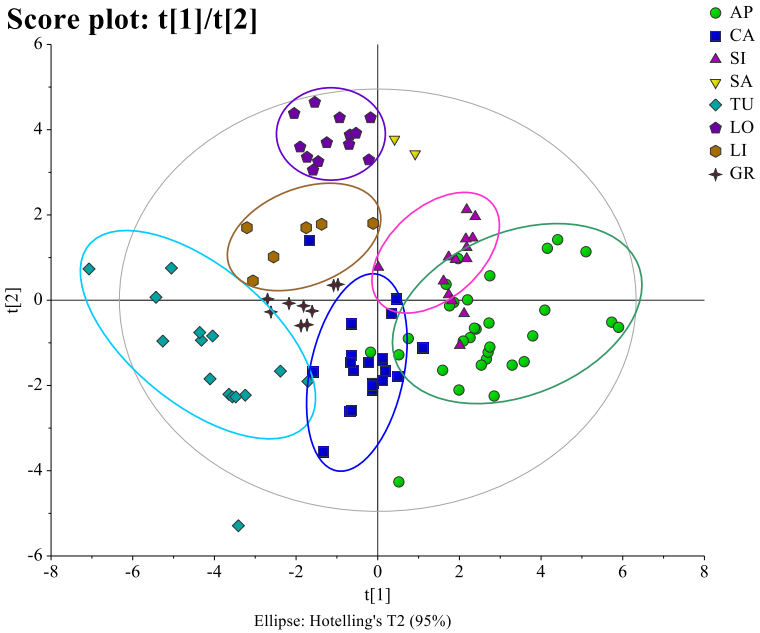


# Misclassification table of EVOOs of the olive oil campaign 2018-2019

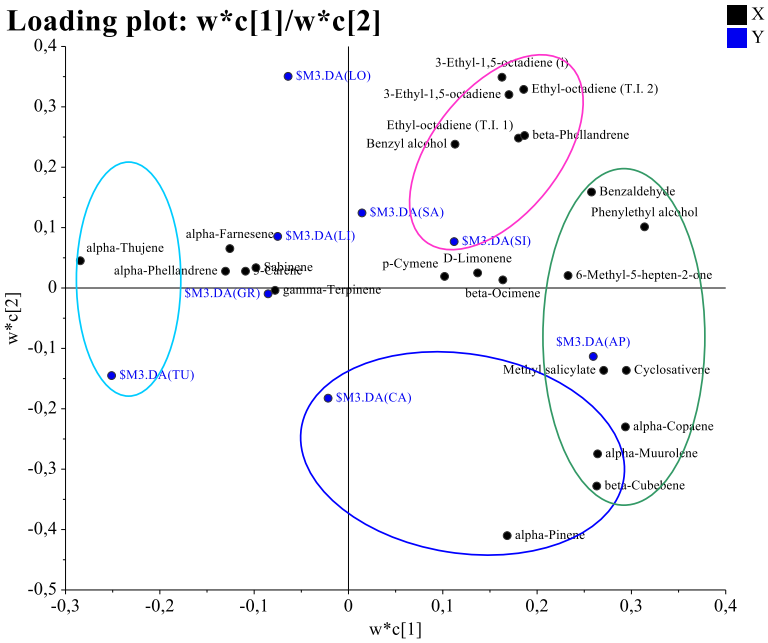
Misclassification table obtained from the **PLS-DA** model built with all the samples and **all the variables** of EVOO samples from different Italian regions and Greece.

| | Members | Correct | AP | CA | SI | SA | TU | LO | LI | GR | No class (YPred < 0) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **AP** | 31 | 100% | 31 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **CA** | 21 | 100% | 0 | 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **SI** | 14 | 100% | 0 | 0 | 14 | 0 | 0 | 0 | 0 | 0 | 0 |
| **SA** | 2 | 50% | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| **TU** | 15 | 100% | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 0 | 0 |
| **LO** | 13 | 100% | 0 | 0 | 0 | 0 | 0 | 13 | 0 | 0 | 0 |
| **LI** | 6 | 66,67% | 0 | 0 | 0 | 0 | 0 | 2 | 4 | 0 | 0 |
| **GR** | 9 | 100% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 9 | 0 |
| **No class** | 0 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total | 111 | 97,3% | 31 | 21 | 15 | 1 | 15 | 15 | 4 | 9 | 0 |

# PLS-DA model of EVOOs of the olive oil campaign 2018-2019

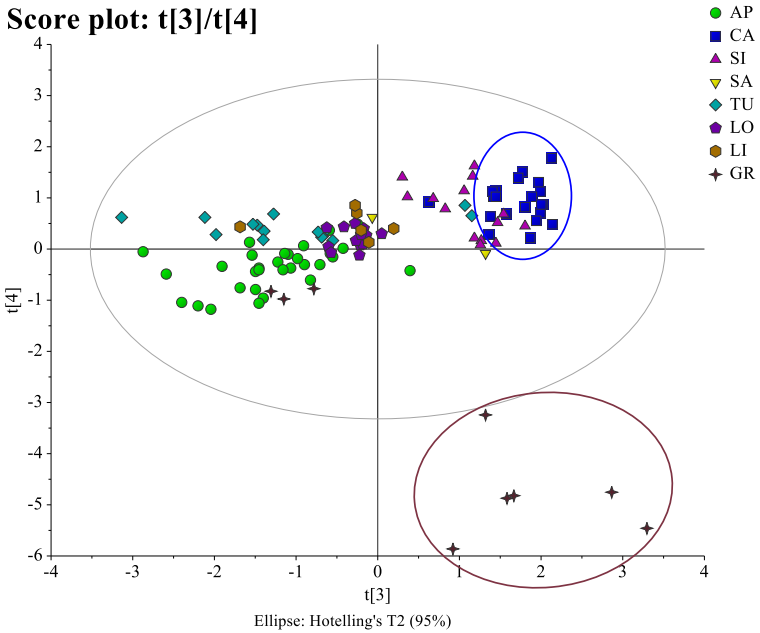



Score plot and loading plot of the first two latent variables of the **PLS-DA** model of EVOO samples of the olive oil campaign 2018-2019 of some Italian regions and Greece built with 111 samples (AP = Apulia; CA = Calabria; SI = Sicily; SA = Sardinia; TU = Tuscany; LO = Lombardy; LI = Liguria; GR = Greece) and **without the LOX variables**. The model explains 82% of the total variance of X with eight significant latent variables (each explaining 27%, 17%, 9%, 8%, 6%, 4%, 7%, and 4%, respectively).



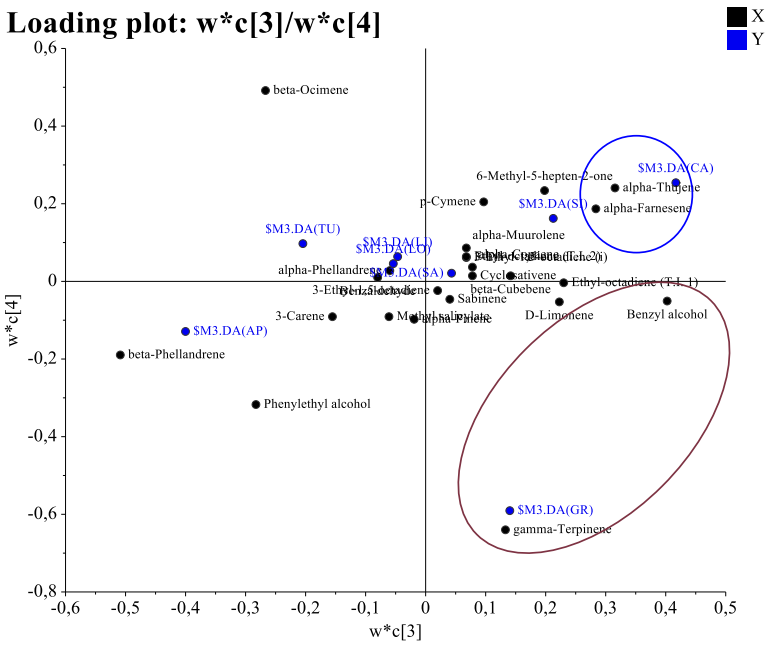

The score plot $t_1/t_2$ is very similar to that obtained with all the variables; however, the Sicilian samples differ more from those of Apulia that have higher values of phenylethyl alcohol, α-copaene, cyclosativene, α-muurolene, β-cubebene, and benzaldehyde; the Sicilian samples differ from the Apulian ones due to higher concentrations of benzyl alcohol and ethyl octadienes. Lombard and Sardinian oils have high contents of benzyl alcohol and ethyl-octadienes. Tuscan oils are characterized by high concentrations of α-thujene; Calabria EVOOs show high values of α-pinene, β-cubebene, and α-muurolene.

# PLS-DA model of EVOOs of the olive oil campaign 2018-2019




In the score plot $t_3/t_4$ there is a clear distinction of the Greek samples, positioned in the lower right side, and the Calabrian ones collocated in a fairly limited area in the upper right; the other classes are quite extensive and partially overlapping. In the relative loading plot it appears that Greek oils are characterized by γ-terpinene and benzyl alcohol, while the Calabrian EVOOs (top right) are characterized by high concentrations of benzyl alcohol, D-limonene, and α-thujene.

Score plot and loading plot of the third and the fourth latent variables of the **PLS-DA** model of EVOO samples of the olive oil campaign 2018-2019 of some Italian regions and Greece built with 111 samples (AP = Apulia; CA = Calabria; SI = Sicily; SA = Sardinia; TU = Tuscany; LO = Lombardy; LI = Liguria; GR = Greece) and **without the LOX variables**.



# Misclassification table of EVOOs of the olive oil campaign 2018-2019

Misclassification tables obtained from the **PLS-DA** and **SIMCA** models built with all the samples and **without the LOX variables** of EVOO samples from different Italian regions and Greece.

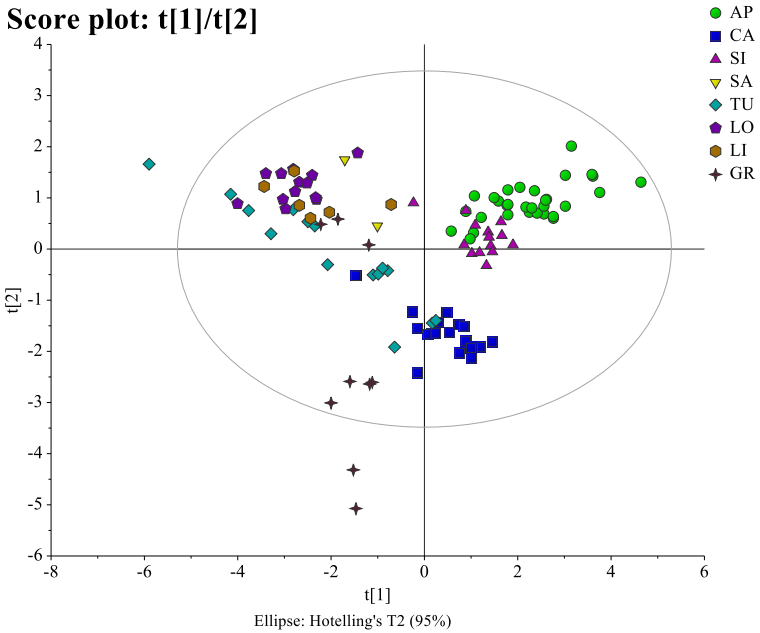
## PLS-DA

| | Members | Correct | AP | CA | SI | SA | TU | LO | LI | GR | No class (YPred < 0) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **AP** | 31 | 100% | 31 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **CA** | 21 | 100% | 0 | 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **SI** | 14 | 100% | 0 | 0 | 14 | 0 | 0 | 0 | 0 | 0 | 0 |
| **SA** | 2 | 0% | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| **TU** | 15 | 86,67% | 0 | 2 | 0 | 0 | 13 | 0 | 0 | 0 | 0 |
| **LO** | 13 | 100% | 0 | 0 | 0 | 0 | 0 | 13 | 0 | 0 | 0 |
| **LI** | 6 | 50% | 0 | 0 | 0 | 0 | 0 | 3 | 3 | 0 | 0 |
| **GR** | 9 | 66,67% | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 6 | 0 |
| **No class** | 0 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total | 111 | 90,99% | 31 | 23 | 15 | 0 | 16 | 17 | 3 | 6 | 0 |

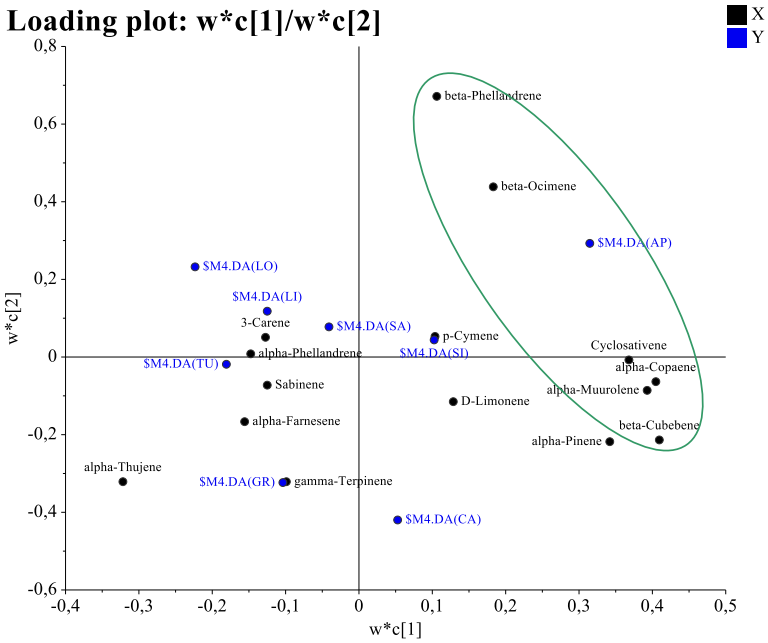
## SIMCA

| | Members | Correct | AP | CA | SI | SA | TU | LO | LI | GR | No class (PModX+ < 0) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **AP** | 31 | 100% | 31 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **CA** | 21 | 100% | 0 | 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **SI** | 14 | 100% | 0 | 0 | 14 | 0 | 0 | 0 | 0 | 0 | 0 |
| **SA** | 2 | 100% | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 |
| **TU** | 15 | 100% | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 0 | 0 |
| **LO** | 13 | 92,31% | 0 | 0 | 0 | 0 | 0 | 12 | 1 | 0 | 0 |
| **LI** | 6 | 100% | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 0 |
| **GR** | 9 | 100% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 9 | 0 |
| **No class** | 0 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total | 111 | 99,1% | 31 | 21 | 14 | 2 | 15 | 12 | 7 | 9 | 0 |

# PLS-DA model of EVOOs of the olive oil campaign 2018-2019




Score plot and loading plot of the first two latent variables of the **PLS-DA** model of EVOO samples of the olive oil campaign 2018-2019 of some Italian regions and Greece built with 111 samples (AP = Apulia; CA = Calabria; SI = Sicily; SA = Sardinia; TU = Tuscany; LO = Lombardy; LI = Liguria; GR = Greece) and **only the terpenes**. The model explains 89% of the total variance of X with eight significant latent variables (each explaining 31%, 14%, 10%, 7%, 10%, 9%, 4%, and 4%, respectively).



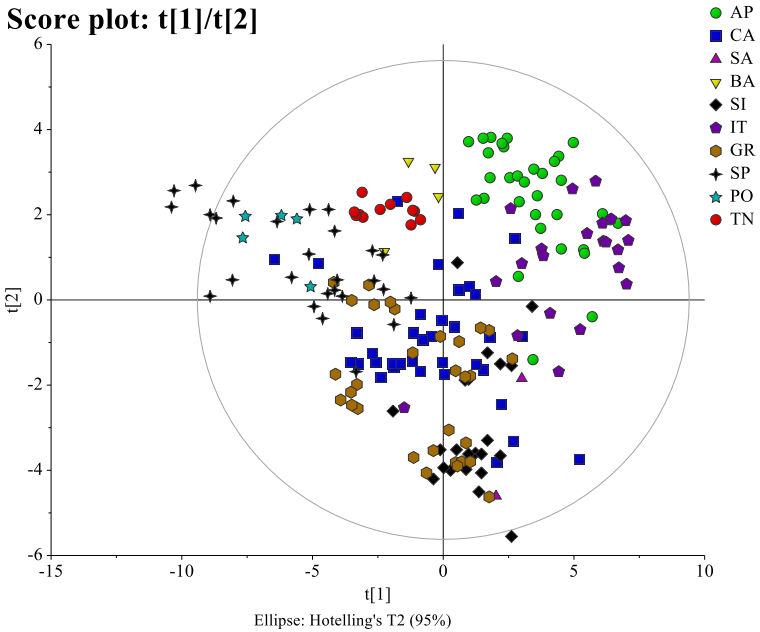

The score plot $t_1/t_2$ shows 5 well-spaced Greek champions; furthermore, it can be noted that Apulia EVOOs, located at the top right, are only partially overlapped with those of Sicily. From the respective loading plot it can be deduced that the variables responsible for this discrimination are mainly β-ocimene, cyclosativene, β fellandrene, α copaene, α muurolene, and β cubebene..

# Misclassification table of EVOOs of the olive oil campaign 2018-2019

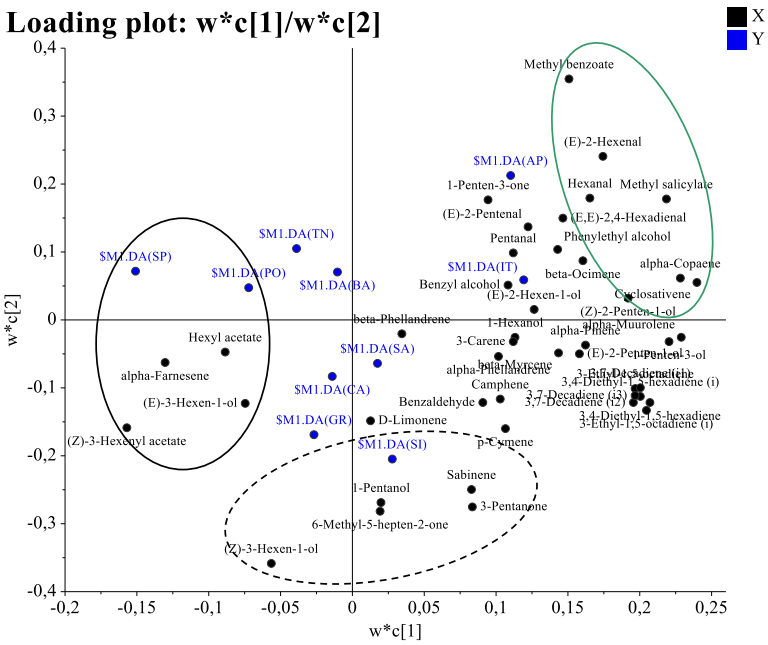
Misclassification table obtained from the **SIMCA** model built with all the samples and **with only the terpenes** of EVOO samples from different Italian regions and Greece.

| | Members | Correct | AP | CA | SI | SA | TU | LO | LI | GR | No class (PModX+ < 0) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **AP** | 31 | 100% | 31 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **CA** | 21 | 100% | 0 | 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **SI** | 14 | 57,14% | 6 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0 |
| **SA** | 2 | 100% | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 |
| **TU** | 15 | 93,33% | 0 | 1 | 0 | 0 | 14 | 0 | 0 | 0 | 0 |
| **LO** | 13 | 92,31% | 0 | 0 | 0 | 0 | 0 | 12 | 1 | 0 | 0 |
| **LI** | 6 | 83,33% | 0 | 0 | 0 | 0 | 0 | 1 | 5 | 0 | 0 |
| **GR** | 9 | 100% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 9 | 0 |
| **No class** | 0 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total | 111 | 91,89% | 37 | 22 | 8 | 2 | 14 | 13 | 6 | 9 | 0 |

# PLS-DA model of EVOOs of the olive oil campaign 2019-2020

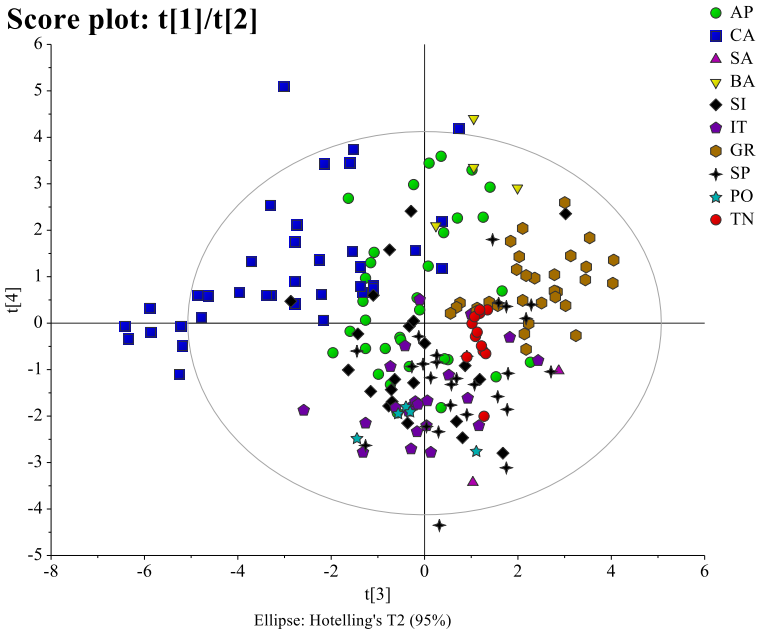



Score plot and loading plot of the first two latent variables of the **PLS-DA** model of EVOO samples of the olive oil campaign 2019-2020 of some Italian regions and some foreign Countries built with 194 samples (AP = Apulia; CA = Calabria; SA = Sardinia; BA = Basilicata; SI = Sicily; IT = Italian; GR = Greece; SP = Spain; PO = Portugal; TN = Tunisia) and **all the variables**. The model explains 85% of the total variance of X with twelve significant latent variables (each explaining 34%, 12%, 10%, 7%, 6%, 3%, 3%, 2%, 3%, 2%, 2%, and 1%, respectively).



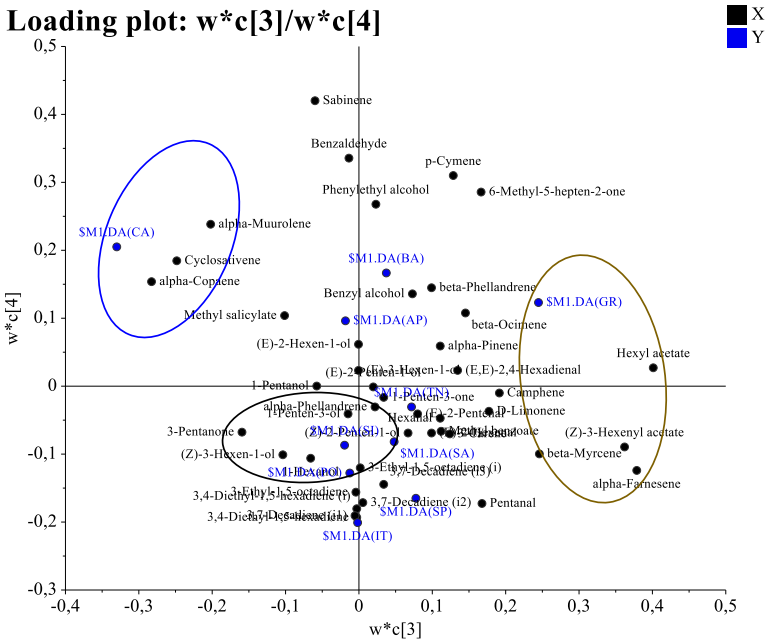

In the score plot the oils of Spain and Portugal are located on the left side of the first component being characterized by high concentrations of (*Z*)-3-hexenyl acetate, hexyl acetate, and α-farnesene, on the right and at the top in the second component are the Apulia and the Italian EVOOs characterized by methyl salicylate, methyl benzoate, (*E*)-2-hexenal, and hexanal and some terpenes. At the bottom of the second component are placed the Sicilian oils which are partially overlapped with the Greek ones; these are characterized by (*Z*)-3-hexen-1-ol, 6-methyl-5-hepten-2-one, 3-pentanone, 1-pentanol and sabinene.

# PLS-DA model of EVOOs of the olive oil campaign 2019-2020




In the score plot $t_3/t_4$ is observed along the third component a good discrimination of the Calabria EVOOs, on the left side, from Greece, on the right,. From the loading plot the Calabria oils are characterized by higher values of α-copaene, cyclosativene, and α-muurolene, while the Greek ones show higher concentrations of hexyl acetate, α-farnesene and (*Z*)-3-hexenyl acetate. The Sicilian samples differ from the Greeks in that the former are placed at the bottom of the fourth component and this is due to higher quantities of unsaturated $C_{10}$ hydrocarbons and alcohols such as 1-hexanol and (Z)-3-hexen-1-ol.

Score plot and loading plot of $t_3/t_4$ of the **PLS-DA** model of EVOO samples of the olive oil campaign 2019-2020 of some Italian regions and some foreign Countries built with 194 samples (AP = Apulia; CA = Calabria; SA = Sardinia; BA = Basilicata; SI = Sicily; IT = Italian; GR = Greece; SP = Spain; PO = Portugal; TN = Tunisia) and **all the variables**.



# Misclassification table of EVOOs of the olive oil campaign 2019-2020

Misclassification table obtained from the **SIMCA** model built with all the samples and **all the variables** of EVOO samples of the olive oil campaign 2019-2020 of some Italian regions and some foreign Countries.

| | Members | Correct | AP | CA | SA | BA | SI | IT | GR | SP | PO | TN | No class (PModX+ < 0) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **AP** | 33 | 100% | 33 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **CA** | 36 | 100% | 0 | 36 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **SA** | 2 | 100% | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **BA** | 4 | 100% | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **SI** | 22 | 95,45% | 0 | 1 | 0 | 0 | 21 | 0 | 0 | 0 | 0 | 0 | 0 |
| **IT** | 22 | 90,91% | 2 | 0 | 0 | 0 | 0 | 20 | 0 | 0 | 0 | 0 | 0 |
| **GR** | 31 | 100% | 0 | 0 | 0 | 0 | 0 | 0 | 31 | 0 | 0 | 0 | 0 |
| **SP** | 27 | 100% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 27 | 0 | 0 | 0 |
| **PO** | 5 | 100% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 |
| **TN** | 12 | 100% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 0 |
| **No class** | 0 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total | 194 | 98,45% | 35 | 37 | 2 | 4 | 21 | 20 | 31 | 27 | 5 | 12 | 0 |

# Misclassification table of EVOOs of the olive oil campaign 2019-2020

Misclassification tables obtained from the **SIMCA** models built with all the samples and **without the LOX variables (on the left)** and **with only the terpenes (on the right)** of EVOO samples of the olive oil campaign 2019-2020 of some Italian regions and some foreign Countries.

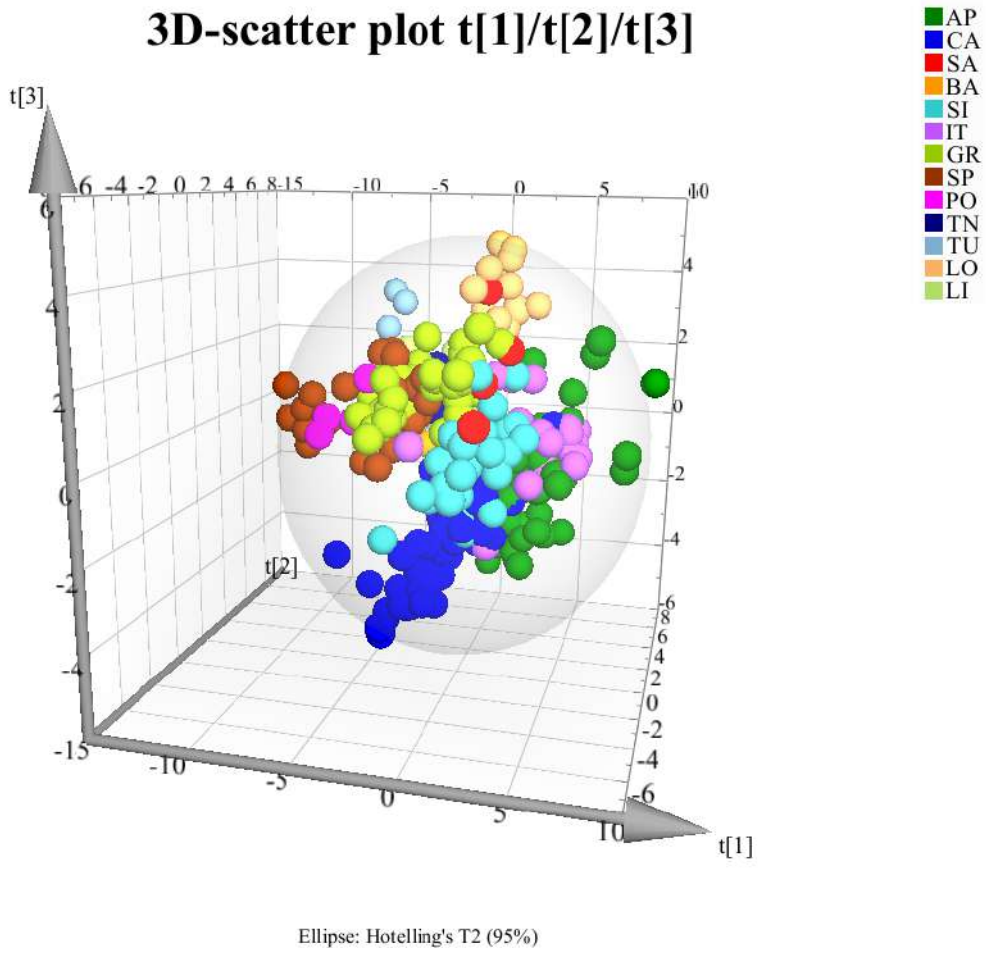
## Without the LOX variables

| | Members | Correct | AP | CA | SA | BA | SI | IT | GR | SP | PO | TN | No class (PModX+ < 0) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **AP** | 33 | 100% | 33 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **CA** | 36 | 100% | 0 | 36 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **SA** | 2 | 100% | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **BA** | 4 | 100% | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **SI** | 22 | 86,36% | 0 | 3 | 0 | 0 | 19 | 0 | 0 | 0 | 0 | 0 | 0 |
| **IT** | 22 | 86,36% | 3 | 0 | 0 | 0 | 0 | 19 | 0 | 0 | 0 | 0 | 0 |
| **GR** | 31 | 96,77% | 0 | 0 | 0 | 0 | 0 | 0 | 30 | 1 | 0 | 0 | 0 |
| **SP** | 27 | 100% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 27 | 0 | 0 | 0 |
| **PO** | 5 | 100% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 |
| **TN** | 12 | 83,33% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 10 | 0 |
| **No class** | 0 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total | 194 | 95,36% | 36 | 39 | 2 | 4 | 19 | 19 | 30 | 30 | 5 | 10 | 0 |

## With only the terpenes

| | Members | Correct | AP | CA | SA | BA | SI | IT | GR | SP | PO | TN | No class (PModX+ < 0) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **AP** | 33 | 90,91% | 30 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 0 |
| **CA** | 36 | 100% | 0 | 36 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **SA** | 2 | 100% | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **BA** | 4 | 100% | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **SI** | 22 | 90,91% | 1 | 0 | 0 | 0 | 20 | 1 | 0 | 0 | 0 | 0 | 0 |
| **IT** | 22 | 81,82% | 2 | 0 | 0 | 1 | 1 | 18 | 0 | 0 | 0 | 0 | 0 |
| **GR** | 31 | 83,87% | 0 | 0 | 0 | 2 | 1 | 0 | 26 | 2 | 0 | 0 | 0 |
| **SP** | 27 | 96,3% | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 26 | 0 | 0 | 0 |
| **PO** | 5 | 100% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 |
| **TN** | 12 | 100% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 0 |
| **No class** | 0 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total | 194 | 92,27% | 33 | 36 | 2 | 7 | 23 | 21 | 27 | 28 | 5 | 12 | 0 |

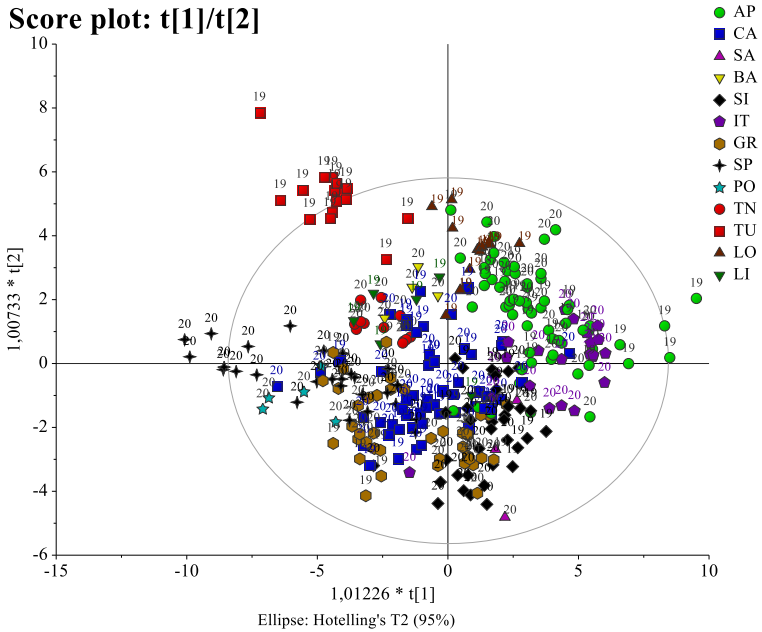
# OPLS-DA model of EVOOs of the olive oil campaigns 2018-2019 and 2019-2020



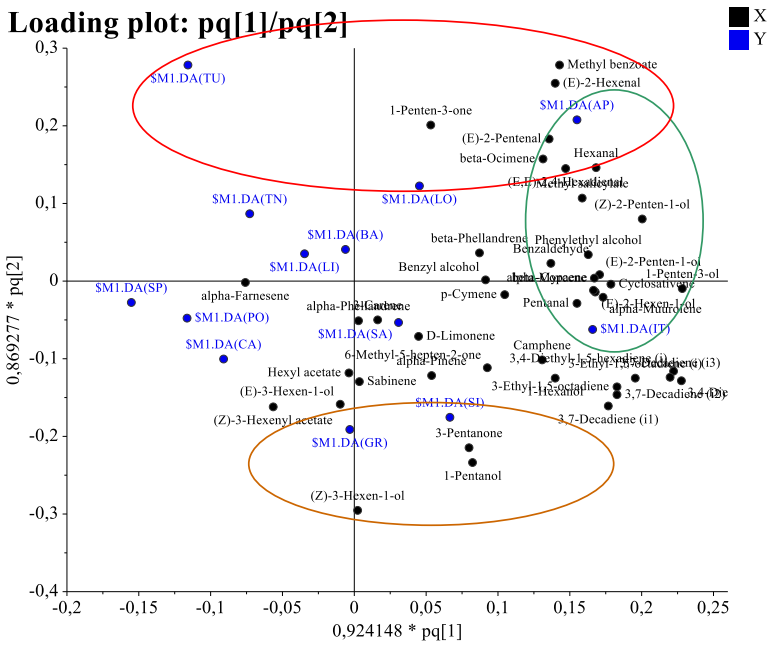

3D-scatter plot ($t_1/t_2/t_3$) of the first three latent variables of the **OPLS-DA** model of EVOO samples of the olive oil campaigns 2018-2019 and 2019-2020 of some Italian regions and some foreign Countries built with 305 samples (AP = Apulia; CA = Calabria; SA = Sardinia; BA = Basilicata; SI = Sicily; IT = Italian; GR = Greece; SP = Spain; PO = Portugal; TN = Tunisia; TU = Tuscany; LO = Lombardy; LI = Liguria) and **all the variables**. The model explains 80% of the total variance of X (R2X) and 42% of the total variation of data (R2) with ten plus one components (10 predictive and 1 orthogonal).

In the 3D-scatter plot there is a more than good separation of Calabrian, Apulian, Italian, Sicilian, Lombard, Greek and Spanish oils, despite the annual variability.

# OPLS-DA model of EVOOs of the olive oil campaigns 2018-2019 and 2019-2020

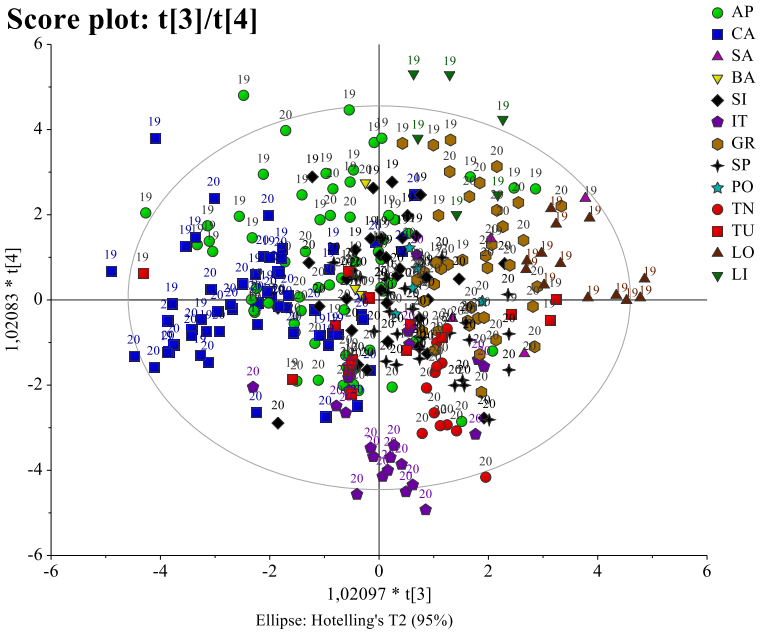



Score plot and loading plot of the first two components of the **OPLS-DA** model of EVOO samples of the olive oil campaigns 2018-2019 and 2019-2020 of some Italian regions and some foreign Countries built with 305 samples (AP = Apulia; CA = Calabria; SA = Sardinia; BA = Basilicata; SI = Sicily; IT = Italian; GR = Greece; SP = Spain; PO = Portugal; TN = Tunisia; TU = Tuscany; LO = Lombardy; LI = Liguria) and **all the variables**.



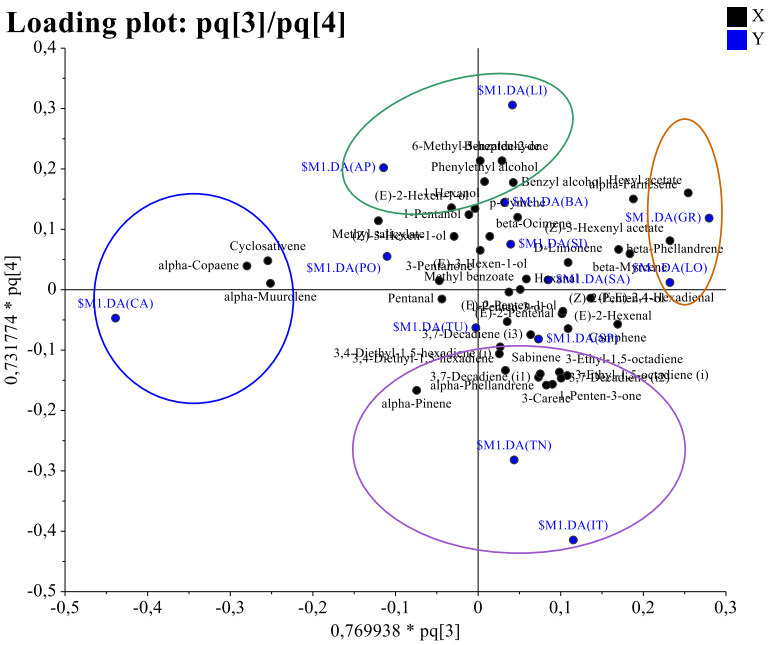

In the score plot $t_1/t_2$ there is a clear discrimination of the Tuscan EVOOs (to the left and to the top), the Spain EVOOs is partly overlapped to the Portuguese ones (to the left) the Apulians and Italians on the opposite side along the first; the Sicilians oils, partially overlapped to the Greek samples, at the bottom. In the loading plot on the right of the first component the variables with the highest weights are unsaturated hydrocarbons, some $C_5$ unsaturated alcohols, and some terpenes (cyclosativene, α-muurolene, and β-myrcene); on the opposite side, there are α-farnesene and (*Z*)-3-hexenyl acetate. Along the second component, at the top the most important variables are methyl benzoate, (*E*)-2-hexenal, 1-penten-3-one, (*E*)-2-pentenal, β-ocimene, hexanal, (*E, E*)-2,4-hexadienal, while at the bottom we find (*Z*)-3-hexen-1-ol, 1-pentanol, and 3-pentanone.

# OPLS-DA model of EVOOs of the olive oil campaigns 2018-2019 and 2019-2020




In the score plot $t_3/t_4$ there is a discrimination of the Calabrian samples placed on the left that are characterized by higher concentrations of α-copaene, cyclosativene, and α-muurolene, opposed to the Greek ones which instead show higher values of hexyl acetate and (Z)-3-hexenyl acetate; along the fourth below there are the Italian EVOs opposed to the Apulian oils: the first characterized by higher absolute values of the loadings of α-pinene, 3-carene, and 1-penten-3-one, the second by 6-methyl-5-hepten-2-one, benzaldehyde, benzyl alcohol, and phenylethyl alcohol.

Score plot and loading plot of the third and the fourth components of the **OPLS-DA** model of EVOO samples of the olive oil campaigns 2018-2019 and 2019-2020 of some Italian regions and some foreign Countries built with 305 samples (AP = Apulia; CA = Calabria; SA = Sardinia; BA = Basilicata; SI = Sicily; IT = Italian; GR = Greece; SP = Spain; PO = Portugal; TN = Tunisia; TU = Tuscany; LO = Lombardy; LI = Liguria) and **all the variables**.



# Misclassification table of EVOOs of the olive oil campaigns 2018-2019 and 2019-2020

Misclassification table obtained from the **SIMCA** model built with all the samples and **all the variables** of EVOO samples of the olive oil campaigns 2018-2019 and 2019-2020 of some Italian regions and some foreign Countries.

| | Members | Correct | AP | CA | SA | BA | SI | IT | GR | SP | PO | TN | TU | LO | LI | No class (PModX+ < 0) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **AP** | 64 | 92,19% | 59 | 1 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| **CA** | 57 | 100% | 0 | 57 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **SA** | 4 | 100% | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **BA** | 4 | 75% | 0 | 1 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **SI** | 36 | 97,22% | 0 | 1 | 0 | 0 | 35 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **IT** | 22 | 59,09% | 5 | 4 | 0 | 0 | 0 | 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **GR** | 40 | 95% | 0 | 0 | 1 | 0 | 0 | 0 | 38 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| **SP** | 27 | 96,3% | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 26 | 0 | 0 | 0 | 0 | 0 | 0 |
| **PO** | 5 | 100% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 0 |
| **TN** | 12 | 100% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 0 | 0 | 0 | 0 |
| **TU** | 15 | 100% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 0 |
| **LO** | 13 | 92,31% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 1 | 0 |
| **LI** | 6 | 100% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 0 |
| **No class** | 0 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total | 305 | 93,44% | 64 | 64 | 5 | 3 | 37 | 14 | 39 | 27 | 5 | 12 | 15 | 12 | 8 | 0 |

# Misclassification table of EVOOs of the olive oil campaigns 2018-2019 and 2019-2020

Misclassification tables obtained from the **SIMCA** models built with all the samples and **without the LOX variables (on the left)** and **with only the terpenes (on the right)** of EVOO samples of the olive oil campaign 2019-2020 of some Italian regions and some foreign Countries.

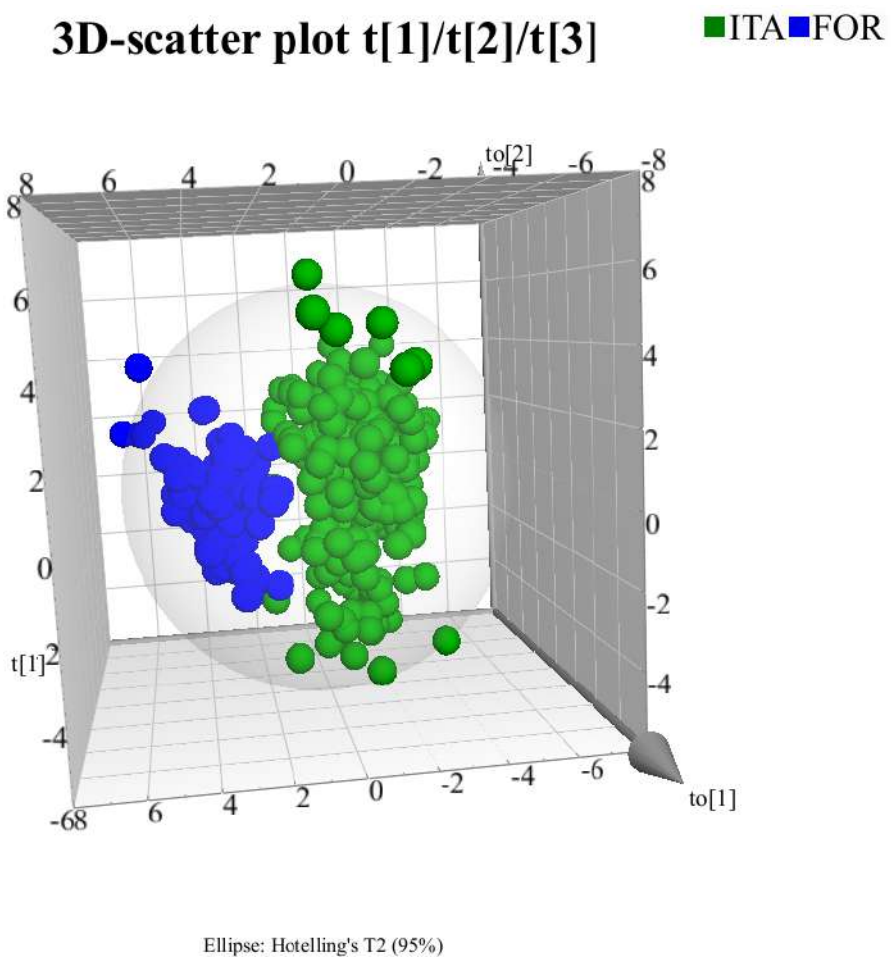
**Without the LOX variables**

| | Members | Correct | AP | CA | SA | BA | SI | IT | GR | SP | PO | TN | TU | LO | LI | No class (PModX+ < 0) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AP | 64 | 93,75% | 60 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 |
| CA | 57 | 98,25% | 0 | 56 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| SA | 4 | 100% | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| BA | 4 | 100% | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| SI | 36 | 86,11% | 0 | 3 | 1 | 0 | 31 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| IT | 22 | 36,36% | 11 | 2 | 0 | 0 | 0 | 8 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GR | 40 | 95% | 0 | 0 | 1 | 0 | 0 | 0 | 38 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| SP | 27 | 96,3% | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 26 | 0 | 0 | 0 | 0 | 0 | 0 |
| PO | 5 | 100% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 0 |
| TN | 12 | 83,33% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 10 | 0 | 0 | 0 | 0 |
| TU | 15 | 100% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 0 |
| LO | 13 | 92,31% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 1 | 0 |
| LI | 6 | 100% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 0 |
| No class | 0 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total | 305 | 90,16% | 71 | 61 | 7 | 4 | 31 | 9 | 40 | 28 | 5 | 10 | 15 | 12 | 12 | 0 |

**With only the terpenes**

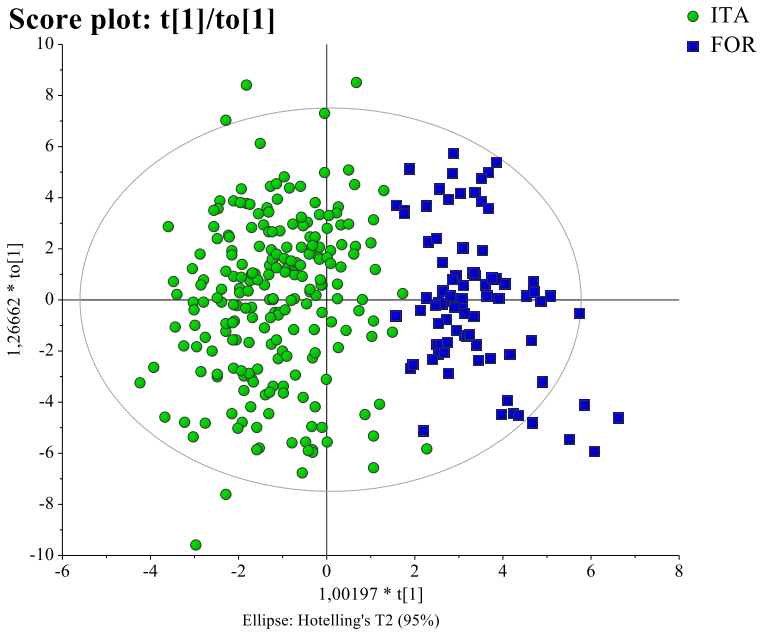
| | Members | Correct | AP | CA | SA | BA | SI | IT | GR | SP | PO | TN | TU | LO | LI | No class (PModX+ < 0) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AP | 64 | 82,81% | 53 | 4 | 1 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 1 | 0 |
| CA | 57 | 91,23% | 2 | 52 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| SA | 4 | 50% | 0 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| BA | 4 | 100% | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| SI | 36 | 58,33% | 8 | 4 | 0 | 0 | 21 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| IT | 22 | 63,64% | 3 | 2 | 0 | 1 | 1 | 14 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| GR | 40 | 82,5% | 1 | 0 | 3 | 0 | 1 | 1 | 33 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| SP | 27 | 88,89% | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 24 | 0 | 0 | 1 | 0 | 0 | 0 |
| PO | 5 | 100% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 0 |
| TN | 12 | 66,67% | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 8 | 3 | 0 | 0 | 0 |
| TU | 15 | 93,33% | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 0 | 0 | 0 |
| LO | 13 | 92,31% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 1 | 0 |
| LI | 6 | 83,33% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 5 | 0 |
| No class | 0 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total | 305 | 80,98% | 67 | 63 | 7 | 5 | 25 | 16 | 38 | 24 | 5 | 8 | 24 | 13 | 10 | 0 |

## OPLS-DA model of EVOOs of the olive oil campaigns 2018-2019 and 2019-2020 Italian and Foreign

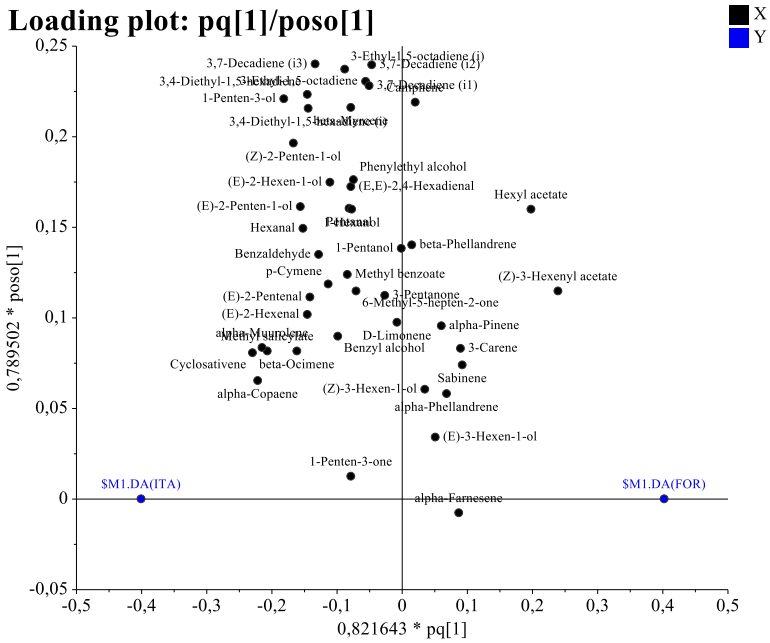



3D-scatter plot ($t_1/t_2/t_3$) of the first three latent variables of the **OPLS-DA** model of EVOO samples of the olive oil campaigns 2018-2019 and 2019-2020 grouped as Italian and Foreign (305 samples) (ITA = Italy; FOR = Foreign) and **all the variables**. The model explains 53% of the total variance of X (R2X) and 75% of the total variation of data (R2) with one plus three components (1 predictive and 3 orthogonal).
In the 3D-scatter plot there is very good discrimination of the oils coming from Italy with respect to the Foreign ones, despite the annual variability.

# OPLS-DA model of EVOOs of the olive oil campaigns 2018-2019 and 2019-2020 Italian and Foreign




Score plot and loading plot of the first two components of the **OPLS-DA** model of EVOO samples of the olive oil campaigns 2018-2019 and 2019-2020 grouped as Italian and Foreign (305 samples) (ITA = Italy; FOR = Foreign) and **all the variables**.




In the score plot $t_1/t_2$ there is an excellent discrimination of the EVOOs coming from Italy with respect to the Foreign ones, despite the annual variability. From the loading plot it is observed that the variables that most differentiate Italian oils along the first component are cyclosativene, α-copaene, α-muurolene, methyl salicylate, 1-penten-3-ol, (*Z*)-2-penten-1-olo, ß-ocimene, (*E*)-2-penten-1-ol, hexanal, and (*E*)-2-hexenal; while the Foreign samples are characterized by higher concentrations of esters such as (*Z*)-3-hexenyl acetate and hexyl acetate.

## Misclassification table of EVOOs of the olive oil campaigns 2018-2019 and 2019-2020 Italian and Foreign

Misclassification table obtained from the models built with all the samples of EVOO (305) of the olive oil campaigns 2018-2019 and 2019-2020 grouped as Italian and Foreign (ITA = Italy; FOR = Foreign).

### With all the variables (OPLS-DA)

| | Members | Correct | ITA | FOR | No class (YPred < 0) |
|---|---|---|---|---|---|
| **ITA** | 221 | 98,64% | 218 | 3 | 0 |
| **FOR** | 84 | 100% | 0 | 84 | 0 |
| **No class** | 0 | | 0 | 0 | 0 |
| Total | 305 | 99,02% | 218 | 87 | 0 |
| Fishers prob. | 0 | | | | |

### Without the LOX variables (SIMCA)

| | Members | Correct | ITA | FOR | No class (PModX+ < 0) |
|---|---|---|---|---|---|
| **ITA** | 221 | 97,74% | 216 | 5 | 0 |
| **FOR** | 84 | 94,05% | 5 | 79 | 0 |
| **No class** | 0 | | 0 | 0 | 0 |
| Total | 305 | 96,72% | 221 | 84 | 0 |
| Fishers prob. | 0 | | | | |

### With only the terpenes (OPLS-DA)

| | Members | Correct | ITA | FOR | No class (YPred < 0) |
|---|---|---|---|---|---|
| **ITA** | 221 | 95,93% | 212 | 9 | 0 |
| **FOR** | 84 | 90,48% | 8 | 76 | 0 |
| **No class** | 0 | | 0 | 0 | 0 |
| Total | 305 | 94,43% | 220 | 85 | 0 |
| Fishers prob. | 0 | | | | |

# CONCLUSIONS

- By employing the headspace volatile compounds analysis in the discrimination of the various samples of extra virgin olive oil, the targeted method gave excellent results, in fact a correct reclassification of almost all the samples was obtained.

- For optimal discrimination and reclassification, it is good to combine the terpene compounds with volatile compounds originating from the lipoxygenase pathway, hydrocarbons and some other compounds important in the differentiation of olive oils.

- It is also possible to discriminate EVOO samples using only terpene compounds, even if the reclassification gives less correct results.

# THANKS FOR YOUR ATTENTION!