

# *Chemometric data treatment of spectroscopic measurements for characterization of monovarietal extra-virgin olive oil from Marche*

*Rosangela Leone*

*School of Science and Technology, University of Camerino*

## **ABSTRACT**

The ever increasing interest of consumers for safety, authenticity and quality of food commodities has driven the attention towards the analytical techniques used for analyzing these commodities. In recent years, rapid and reliable sensor, spectroscopic and chromatographic techniques have emerged that, together with multivariate and multiway chemometrics, have improved the whole control process by reducing the time of analysis and providing more informative results. Sensory quality is an important property of extra virgin olive oil influenced by different volatile and phenolic compounds. Their levels can be influenced by many factors, and one of the most important is the cultivar. Spectroscopic techniques such as IR spectroscopy and UV-Vis spectroscopy offer advantages as quick and accurate measurements that require minimal sample preparation and produce minimal amount of waste. For the characterization of monovarietal extra virgin olive oils from Marche, both IR spectra and UV-vis different chemometric techniques were applied: PCA, PLS-DA, and Class Modeling Techniques: UNEQ and SIMCA, obtaining excellent results in classifying the oils (olives) of the analyzed oils. In the case of processing IR data, the modeling method of the UNEQ allows to recognize all the 5 categories fairly well (average sensitivity = 88%). The method also shows a moderate specificity, especially the models of the Coroncina, Mogliano and Raggia have been able to refuse most of the samples belonging to the Mignola category. Processing with UV-vis is data produced a model that could classify the five varieties of extra virgin olive oil with excellent results. The prediction efficiency for SIMCA and UNEQ models is 90% and 85% respectively. The prediction of external sets has good sensitivity for all categories with both methods (SIMCA and UNEQ), while the specifics are unsatisfactory. Finally, the correlation between the spectroscopic technique (FTIR) and the volatiles, phenols and the sensory analysis of the extra-virgin olive oils of the Marche have also been determined. Calibration models with PLS were constructed to reveal any correlation between quality parameters and spectral data. The performance of these models is satisfactory for some phenolic compounds (P-Coumaric Acid, Pinoresinol), Volatiles (Hexanal, 2-Hexenal) and for some sensory parameters (Fruity and Bitter).

## **1. Introduction**

The survival of the national agricultural system is increasingly linked to the typical nature of the products. The farmers in our agro industrial system are losing the international competition with producers who have broader agricultural extensions and lower labor costs. However, the Italian production system is still rich in typical high quality products, which are also highly demanded on the most advanced markets. These products, with high added value, suffer commercial frauds including

counterfeiting. It is therefore important, for proper protection, of our specialties, the development of methods that, in an economic and easy way, allow to characterize the products. In this context, the aim of this study was to find a chemometric data treatment of spectroscopic measurements that permit the characterization of monovarietal extra-virgin olive oil from Marche. We dedicate our efforts to the recognition of five cultivars characteristic of Marche production: Ascolana, Coroncina, Mignola, Mogliano and Raggia. There are two analytical methods<sup>1</sup> for trying to characterize

an oil; one uses specific analytical methods to quantify each single substance, in this case the choice of analytical parameter depends on the purpose of the analysis<sup>II</sup>. A totally different method is one that uses fast analytical instrumental methods<sup>III</sup>, which provide an aspecific response related to many of the substances that define the characteristics of oil<sup>IV</sup>. The readings, in this type of analysis, give no direct information but treating them with appropriate chemometric methods, it is possible to disclose the information of interest. Chemometric methods are statistical techniques that need a sufficient number of samples to represent the overall product variability.

In this study the choice falls on non-specific instrumental techniques<sup>V</sup>, non-invasive, not destructive and easy to apply even on a large scale, capable of providing complete and global information about the sample being examined. An increasing number of works mention spectroscopic analysis, such as UV-vis<sup>VI</sup>, IR<sup>VII</sup>, to satisfy these requirements.

Our study aims to develop a method for characterization of monovarietal extra-virgin olive oil, produced by the cultivars Ascolana, Coroncina, Mignola, Mogliano and Raggia. To achieve this goal, non-specific measurements have been carried out, and particularly in my case: UV-VIS and IR spectra. The UV-VIS and IR spectra were chemometrically investigated to respond to our needs. As each spectrum has a large number of points, the first step was applying a method for reducing them. Several pre-treatment were investigated in order to find the one that combined with PCA and other methods enable us to reduce the dimensionality of the data and discriminate among categories (varieties of monovarietal extra-virgin olive oils). We applied several techniques of modelling and classification: PLS-DA, SIMCA, UNEQ, in order to find an analytical-chemometric model capable to classify each sample. It was subsequently proceeded to a selection of variables with SE-

LECT, and then it has been verified the accuracy of the reduced model. We have always adopted validation methods to estimate both the classification and prediction ability of the methods mentioned above.

Partial least square (PLS) calibration was used, with samples of year 2015, to study the correlation between the spectroscopic measurement (FTIR) and volatiles, phenols and sensorial analytical parameters.

## 2. Experimental

### 2.1. Olive oil samples

We gathered samples of monovarietal extra-virgin olive oils produced in Marche with the aim of identifying a specific fingerprint of the typical product. Several cause affect the characteristics of extra-virgin olive oils, between them we can highlight the season, the pedology weather condition, processing method and storage condition. In order to take into account of these causes we sampled over 4 years (2009, 2010, 2015, 2016 ); the samples come from wide area in Marche including the typical areal of each variety. Sampling was carried out at some producers, selected with the help of the staff “Osservatorio regionale Suoli-Servizio Agricoltura Regione Marche” and of” Studio Agronomico Demetrio Ruffini di Colmurano”. The producers guarantee the quality of the samples and the variety of the olive used. We used a total of 153 samples of oil, related to 5 categories: Ascolana, Coroncina, Mignola, Mogliano and Raggia.

### 2.2. Instruments and procedures

The samples were filtered on a 5 µm PVC filter before every spectroscopic measurement for removing any suspended material which may give rise to absorption and light diffusion phenomena during the spectrophotometric experiment. The two spectroscopic techniques that have been applied to olive oil samples are: FT-IR and UV-vis spectroscopy.

### 2.2.1. FT-IR spectroscopy

IR spectra of each sample were obtained, on previously filtered oil, by a spectrophotometer Perkin Elmer spectrum 100 FT-IR equipped with ATR system, it was operated in transmission mode, resolution  $0.5\text{ cm}^{-1}$  averaging on 11 scans to obtain the spectrum. Each spectrum was acquired in the spectral range between  $4000$  and  $650\text{ cm}^{-1}$  with subtraction of the background that was measured in air before every sample measurement. Two spectra for each sample were acquired by repeating the measurement procedure. We cleaned the sample holder by alcohol and lab paper before delivering  $40\text{ }\mu\text{L}$  of oil on the measuring crystal.

### 2.2.2. UV-vis spectroscopy

We prepared two solutions for each sample of filtered oil, at 10% and 1% respectively, carefully weighing suitable aliquots then taking to volume (volumetric dilution, 25 mL flask) with Isooctane(2,2,4-Trimethylpenta-ne) spectrophotometrically pure. Two instruments were used for the spectroscopic measurements: in years 2009-2010 the UV-vis spectra were acquired from 200 to 700 nm every 0.5 nm by a Varian Cary 50 Scan spectrophotometer, spectra were subtracted of the pure isooctane spectrum; every measurement was repeated twice; in years 2015-2016 the spectra were acquired from 190 to 1100 nm every 1 nm by a H.-Packard 8453 UV-visible Spectroscopy System, spectra were subtracted of the pure isooctane spectrum. Four measurements on each solution were performed. Code A in the file name stand for 1% solution while B indicates 10% solution. All that was done for increasing the number of objects in the dataset to consider the variability due to experimental noise.

## 2.3 Data analysis

The experimental data were processed with the chemometric software: Unscrambler<sup>VIII</sup> and V Parvus<sup>IX</sup>.

### 2.3.1. Processing IR and UV-vis data for classification

1)IR matrix: segments saturated or without significant absorptions were removed: 200–660 nm, 1755-2650 nm and 3560-4000 nm. Thus, the final IR data matrix contains 72 rows (samples) and 3524 variables, IRset1.

2)In UV-vis matrix we have considered two spectra related to different solutions 10% and 1% this for taking in account the correct values of absorbance, in fact the the two spectra were combined in order to respect the Lambert-Beer's law. The spectra has the absorbance recorded for solutions at 10% at wavelengths between 690 and 285 nm and the values of absorbance of samples at 1% from 284 to 241 nm. We have decided to not consider the spectral readings between 240 and 190 nm because they are out of scale, even in diluted solutions. Several pre-treatments and transforms of the data were investigated in pairs with the different pattern recognition methods. The following pre-treatments were applied on this dataset (IRset1): SNV followed by first derivatives computed by means of third-order polynomials through eleven points Savitzky-Golay method; the result was then autoscaled by column. The use of SNV, and first derivative was effective in eliminating unwanted variations, like global intensity effects and baseline shifts. Whereas autoscaling for eliminating the scatter effects and scaling on the data.

In the case of UV-vis spectra, by Standard Normal Variate (SNV) transform followed by column autoscaling.

The following chemometric methods were applied on the data matrices:

a) Principal component analysis (PCA) as a display method, in order to visualize the data structure.

b) PLS-DA (only for the IR matrix ) in order to investigate the possibility to predict the classification of the different categories of extra-virgin olive oil.

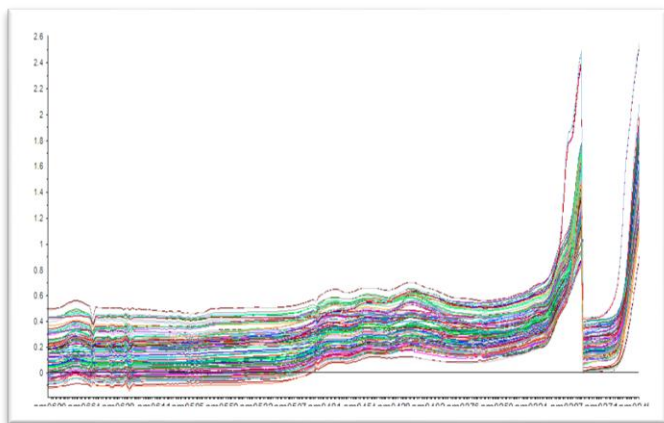


Fig. 1. Row UV-vis spectra

c) UNEQ and SIMCA as class-modelling techniques, in order to build models for Monovarietal extra-virgin olive oil from Marche.

UNEQ<sup>x</sup> is the name currently used in chemometrics for the class modelling technique developed by H. Hotelling and based on the  $T^2$  statistics. UNEQ is less sensitive than SIMCA to the presence of uninformative variables, but this method cannot be applied when the number of objects in each class is less than the number of variables. For this reason, a selection of discriminant wavelengths was performed by means of SELECT method as the first step.

SIMCA<sup>xi</sup> is the first class-modelling technique used chemometrics. An object is accepted by a model when its distance from the model is less than a critical distance, defined by means of Fisher statistics. Such critical distance is influenced by uninformative variables so that, when their number is large, many objects of non-pertinent classes may be accepted by a given model, whose specificity consequently decreases. For this reason, SIMCA is very sensitive to the presence of variables with no or small discriminant power.

In SIMCA, the principal components of auto-scaled data are computed and K components are used to build the model. These K components define the inner space, the space of the structure. The other dimensions are the outer space, the space of noise. The model of

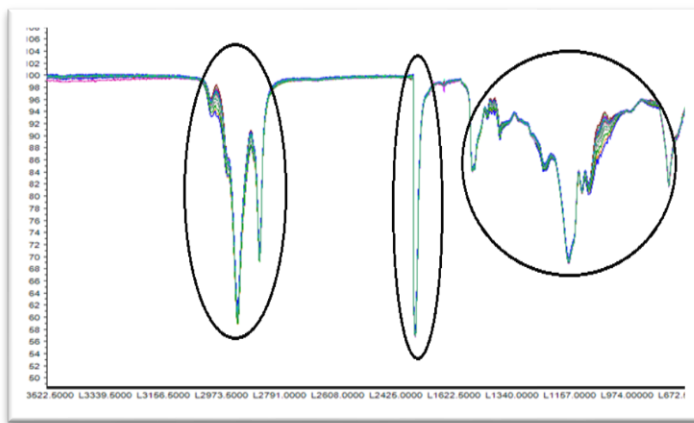


Fig. 2. Row IR spectra

SIMCA is a (hyper) parallelepiped in the space of the first components, delimited by the range of the scores. In this study, the complexity of the inner space was determined setting the minimum percentage of retained variance (95%).

Since the oil origin was guaranteed, the SIMCA and UNEQ models were built forcing the class boundary to retain all the samples of the respective class, thus the sensitivity is always 100%. The UNEQ, SIMCA and PLS models were validated using both a cross-validation procedure with 5 cancellation groups (5CV) and an external test set.

SELECT is a variable selection method for classification and regression and, in this study it was used to improve the performance of UNEQ, SIMCA and PLS.

SELECT<sup>xiii</sup> is a stepwise algorithm, based on the decorrelation of variables, that permits to select the most useful uncorrelated variables. It examines a variable at a time starting with that having the highest correlation. Linear correlation coefficient is the correlation index when the algorithm is applied with regression method while the Fisher classification weight is the correlation index when we want to optimize the classification. The variables are both decorrelated and selected on the base of the correlation index, upgraded at each step. The algorithm iterates until a fixed number of variables is chosen or no further selection

happened in the last cycle.

In Fig.1 and Fig.2 the row UV-vis and IR are shown, respectively.

### 2.3.2. Prediction of chemical parameters of olive oil with FT-IR

This study aims to determine some correlation between the spectroscopic measurements (FT-IR) and volatiles and phenols substances or sensorial features of extra-virgin olive oils. The samples are those of year 2015.

Partial least square (PLS) calibration models were computed in order to reveal possible correlation between quality parameters and spectral data. The evaluation of the performance of these models was based on the multiple coefficient of determination ( $R^2$ ), the root mean square error of calibration (RMSEC) and root mean square error of cross validation (RMSECV).

PLS regression was used to relate 10 of the IR spectral data (IRselect) with the analytical measurements of important chemical parameters of olive oils, such as: Fruity, Bitter, P-coumaric, Hexanal etc. In order to increase the predictive ability of the PLS model area normalization was applied to the 10 of the spectral FTIR data. Models were constructed for each response separately, and to validate the developed PLS model, Leave-One-Out (LOO) cross-validation method was used<sup>xiii</sup>.

The Unscambler software was used for PLS calculus.

## 3. Result and discussion

### 3.1. PCA results

PCA was applied on the two matrices listed in the Section 2.3. We used principal component analysis to “see” if there is some kind of grouping hidden in the data. Principal components analysis using only the objects of the year 2015 is performed on the IRset1 variables set including 3524 wavenumbers selected as above explained. Pre-treatment used

for getting the next plot are: SNV, First derivatives and autoscaling, how we said before. Then we applied a Systematic (112233) Cross Validation method with 20 segments. The three-dimensional score plot of objects Fig.3 show that the objects group on the base of their category if projected on the PC2, PC4 and PC5. Ascolana is the best separated category but even Mogliano and Raggia are well apart. In the case of UV-vis spectra, the Fig.4 shows the scores plot of PCA obtained after application of SNV followed by column autoscaling. We noticed the improvement in grouping ability of the homogeneous data, particularly evident for Mogliano, Raggia and Ascolana.

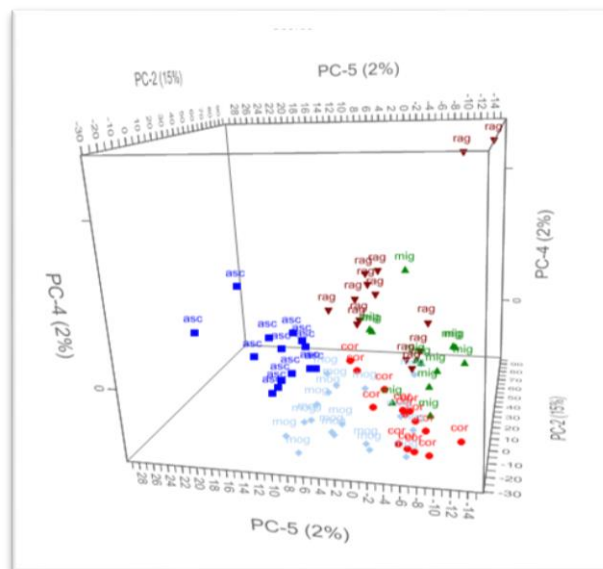


Fig.3 PCA on IRset1, score plot of PC2 (15%) PC4 (2%) and PC5 (2%)

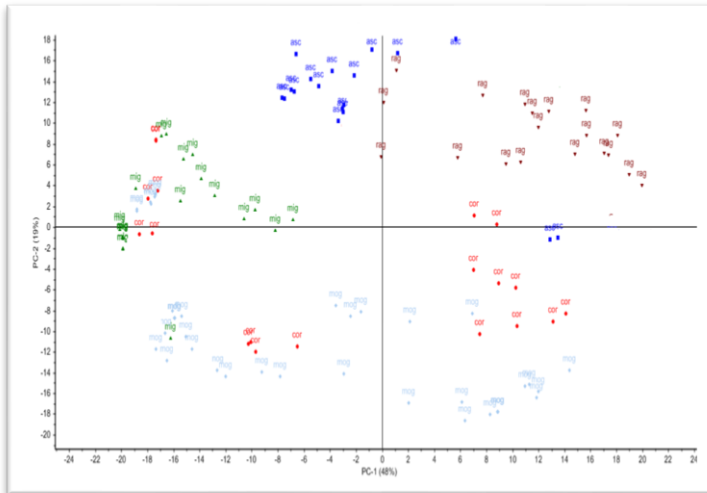


Fig.4 Score plot after SNV and column autoscaling with PC1 (48%) vs PC2 (19%)

### 3.2. PLS-DA results

The encouraging results obtained with PCA suggest that PLS-DA, a method based on a similar algorithm, could be profitably used to classify the oil varieties. Here variables in IR-set1 were analysed using all the objects of matrix. We evaluated the prediction ability of the method by 5 groups' cross-validation. Fig. 5 shows the score plot of Factor-2 (22%,4%), Factor-3 (4%, 12%) and Factor-4 (4%,12%) where is evident the separation between categories.

### 3.3. UNEQ and SIMCA results

As far as the two spectral matrices (UV-vis and IR ) are concerned, UNEQ and SIMCA were performed after feature selection, because a necessary considering the high number of variables. In this study, SELECT was performed reduction setting the in size maximum number of selections. The prediction and classification ability of a method, as well as by the confusion matrix, can best be judged if other indices such as sensitivity, specificity, and efficiency are also considered. We were able, using IR spectra, to classify the varieties Coroncina, Mignola, Mogliano, Asc olana and Raggia. Select algorithm enabled us to choose 10 wavelengths of the spectra with which to classify the oil varieties, really

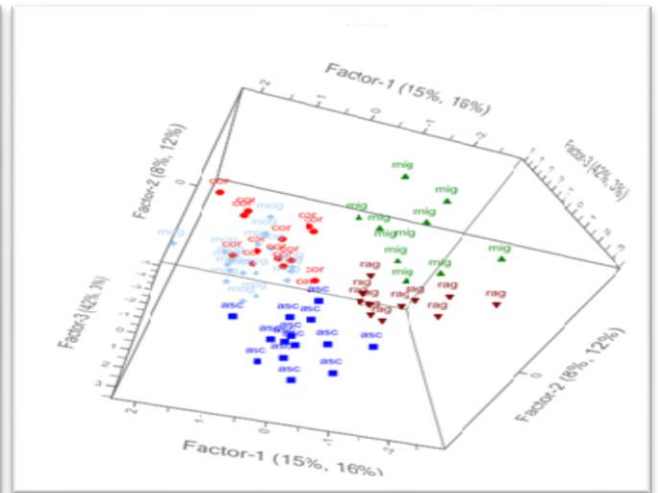


Fig.5 PLS-DA on IRselect, Score plot of Fact-1 (15%,16%),Fact-2 (8%, 12%),Fact-3 (42%,3%)

models of classes. These 10 variables can be used both with SIMCA and UNEQ modelling methods. The prediction ability of the methods has been investigated with a large external set including samples of several years. In spite of this large test UNEQ modelling method permits to recognize quite well all samples belong to the 5 categories (average Sensitivity=88%). The method show also moderate specificity, in particular the category models Coroncina, Mogliano and Raggia succeed to reject the majority of samples belong to the category Mignola and the samples of the other of these categories. Coroncina and Mogliano show low specificity among them, (see Tab.1) SIMCA modelling method performs worse in the same condition than UNEQ with 5 categories. We used as training set the most part of the samples of year 2015 were there are 5 classes, however two of these classes Raggia and Ascolana were not present in the data of the years preceding 2015. The external set have unbalanced number of objects that can negatively affect the results. To check this, we made a training set including only the samples of year 2015 belonging to the categories: Coroncina, Mignola and Mogliano. The prediction results showed an improvement of both Sensitivity and Specificity.

SENSITIVITIES EXT SET (Level = 95%)		
Ascolana	75%	(Accepted 15/ 20)
Coroncina	88%	(Accepted 56/ 64)
Mignola	98%	(Accepted 57/ 58)
Mogliano	86%	(Accepted 76/ 88)
Raggia	82%	(Accepted 23/28)
<b>Mean sensitivity</b>	<b>88%</b>	

EFFICIENCY EXT SET (Level = 95%)	
Ascolana	57%
Coroncina	58%
Mignola	17%
Mogliano	62%
Raggia	69%

SPECIFICITIES EXT SET (Level = 95%)			
Ascolana		Coroncina	
Coroncina	47% (Rejected 30/64)	Ascolana	5% (Rejected 1/20)
Mignola	43% (Rejected 25/58)	Mignola	84% (Rejected 49/58)
Mogliano	42% (Rejected 37/88)	Mogliano	14% (Rejected 12/88)
Raggia	36% (Rejected 10/28)	Raggia	39% (Rejected 11/28)
<b>Mean specificity for Ascolana</b>	<b>43%</b>	<b>Mean specificity for Coroncina</b>	<b>38%</b>
Mignola		Mogliano	
Ascolana	0% (Rejected 0/10)	Ascolana	10% (Rejected 2/20)
Coroncina	2% (Rejected 1/64)	Coroncina	33% (Rejected 8/24)
Mogliano	6% (Rejected 5/88)	Mignola	85% (Rejected 49/58)
Raggia	0% (Rejected 0/28)	Raggia	54% (Rejected 15/28)
<b>Mean specificity for Mignola</b>	<b>3%</b>	<b>Mean specificity for Mogliano</b>	<b>44%</b>
Raggia			
Ascolana	35% (Rejected 7/20)		
Coroncina	52% (Rejected 33/64)		
Mignola	69% (Rejected 40/58)		
Mogliano	60% (Rejected 53/88)		
<b>Mean specificity for Raggia</b>	<b>58%</b>		

Tab.1 Parameters: IR Modeling 5 cat. by UNEQ.

While the sensitivity is high both with SIMCA and UNEQ, the first method performs much better with respect to specificity. Even if specificity is high for all three categories it is noteworthy the case of Coroncina.

The modelling of oil types was attempted even with UV-Vis spectra. The training set consists of samples of year 2015. The elaborations produced model able to classify with good results the five varieties of extra-virgin olive oil. The prediction efficiency, estimated by 5 deletion groups cross-validation, for SIMCA and UNEQ models is 90% and 85%, respectively. Prediction estimated by external set has good sensitivity for all categories with both methods (SIMCA and UNEQ) whereas the specificities are poor.

Thus in SIMCA only the category Mogliano succeeds to recognize well all samples and exclude most of those not belonging to it; whereas in UNEQ only the category Coroncina do, (see Tab.2).

QUALITY PARAMETERS		R <sup>2</sup> <sub>cal</sub>	R <sup>2</sup> <sub>cv</sub>	RMSEC	RMSECV
FRUITY	SENSORY	0.99	0.99	0.47	0.52
BITTER		0.97	0.97	0.67	0.62
P-COUMARIC AC.	PHENOLS	0.93	0.91	0.03	0.03
PINORESINOL		0.88	0.85	2.73	3.12
2-HEXENAL	VOLATILES	0.83	0.79	1.6+e^8	1.8+e^8
HEXANAL		0.77	0.73	7.9+e^6	8.7+e^6

Tab.3 the best predicted quality parameters.

SIMCA		UNEQ	
EFFICIENCY MODEL (Level = 95%)		EFFICIENCY MODEL (Level = 95%)	
Ascolana	90%	Ascolana	78%
Coroncina	90%	Coroncina	91%
Mignola	97%	Mignola	93%
Mogliano	85%	Mogliano	65%
Raggia	89%	Raggia	85%
EFFICIENCY EXT-SET2 (Level = 95%)		EFFICIENCY EXT-SET2 (Level = 95%)	
Ascolana	17%	Ascolana	32%
Coroncina	53%	Coroncina	65%
Mignola	42%	Mignola	32%
Mogliano	71%	Mogliano	16%
Raggia	28%	Raggia	45%

Tab.2 Efficiency UV-vis Modeling

### 3.4. PLS regression results

One of the multivariate statistical analysis tools, PLS regression was used to relate the FT-IR spectral data with the analytical results of several important chemical parameters of olive oils. Models were constructed for each response separately. The best correlation, among sensory parameters, was found for Fruity (see Tab.3) with R<sup>2</sup><sub>cal</sub> = 0.99 and R<sup>2</sup><sub>cv</sub> = 0.99 when using a model with six latent variables. The high value of R<sup>2</sup> and the lowest of RMSEC (0.47) and RMSECV (0.52) indicate the good performance and precision of PLS model. The Weighted Regression Coefficients plot showed that the Fruity note is mainly influenced by variables: 1273.5 cm<sup>-1</sup>, related to the alkylic chains due to the terminal (CH<sub>3</sub>) groups symmetric bending, and 1123.5 cm<sup>-1</sup>, related to the (-C-C-O-H) antisymmetric stretch C-O compared to the Stretch C-C.

The Bitter taste (see Tab.3), modelled with six latent variables, also correlated well to IR; the values R<sup>2</sup><sub>cal</sub> = 0.97 and R<sup>2</sup><sub>cv</sub> = 0.97 indicates good stability of the model. This is also supported by the values of tolerable differences between RMSEC (0.67) and RMSECV (0.62). The plot of the weights of variables, showed us that Bitter taste is highly connected with IR variables 1273.5 and 1132.5 cm<sup>-1</sup>. The variable 1132.5 cm<sup>-1</sup> is related to the (-C-C-O-H) antisymmetric stretch C-O compared to

the Stretch C-C. Interestingly, Bitter note is correlated to 2-Hexenal by the same IR variables (1273.5, 1123.5  $\text{cm}^{-1}$ ) and it also influenced by phenols; in fact it is correlated to the Pinoresinol compound by the variable (1123.5  $\text{cm}^{-1}$ ), see Fig. 6.

The Phenolic compound P-Coumaric correlate well enough to IR features,  $R^2_{\text{cal}} = 0.91$  and  $R^2_{\text{cv}} = 0.89$  indicates a good model This is also supported by the low RMSE (that are  $\text{RMSE}=\text{RMSECV}=0.03$ ).

Besides P-Coumaric, another phenolic compound, Pinoresinol showed high correlation with IR variables, with  $R^2_{\text{cal}}= 0.88$  and  $R^2_{\text{cv}} = 0.85$ .

This is also supported by the low difference between  $\text{RMSEC}(2.8)$  and  $\text{RMSECV}(3.1)$ . Comparing the Weighted Regression Coefficients plots we saw that the wavenumber 1451  $\text{cm}^{-1}$  is highly significant for both phenolic compound models. This wavelength is related to the scissoring vibration of the  $\text{CH}_2$  groups<sup>xiv</sup> of the alkylic chains. The other important IR variable (1732  $\text{cm}^{-1}$ ) for P-Coumaric acid corresponding to the carbonyl ( $\text{C}=\text{O}$ ) stretching vibration, in this case associated with the triglyceride ester bond, as well as the carboxylic group of free fatty acids.

The other IR variable (1123.5  $\text{cm}^{-1}$ ) in the case Pinoresinol model (Fig.6d) is also one of the important wavelengths in the Bitter sensory model (Fig.6c). Volatile compounds such as: ethanol, 3-pentanone, 1-penten-3-one, 1-penten-3-ol, 1-hexanol, (Z)-3-hexen-1-yl acetate, (Z)-3-hexen-1-ol and (E)-3-hexen-1-ol are not correlated to IR wavelengths so well as the predict as the phenolic compounds. We observed, however, that 2-Hexenal and Hexanal correlate well enough to IR with 6 latent PLS model,(see Tab.3), for both calibration and validation  $R^2_{\text{cal}}$  ( 0.83, 0.77) and  $R^2_{\text{cv}}$  (0.79, 0.73) values provides an idea about the correlation with IR wavelength investigated. RMSEC and RMSECV values are related with the error between measured value and predicted value at each calibration and cross-validation step, respectively, are quite high, but the differences between them is small, 0.8 (in the case of Hexanal) and 0.15 ( in the case of 2-Hexenal).A comparison of the “Weighted Regression Coefficients” plots for 2-Hexenal (see fig.6b) with Bitter (see Fig.6c) permit to see that wavenumbers 1273.5, 1123.5  $\text{cm}^{-1}$  are important for both models.

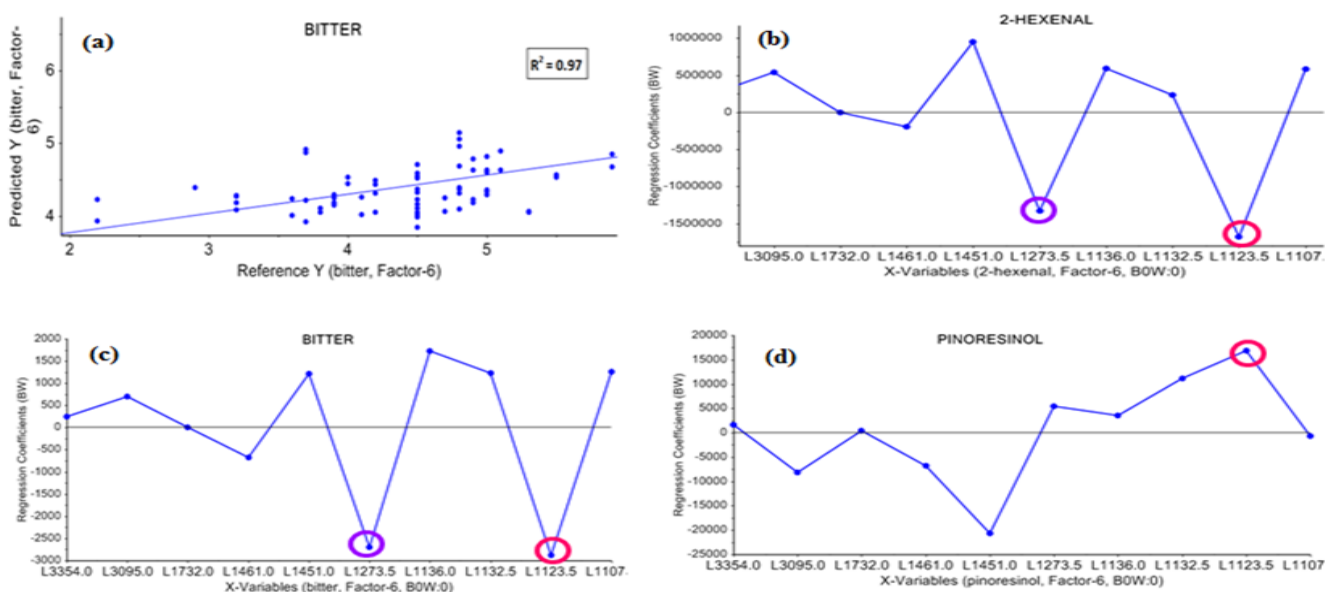


Fig.6 (a) Predicted vs Reference(Bitter); (b) ,(c) and (d) Weighted Regression Coefficient plots.



#### 4. Conclusion

The first aim of this study was to classify the monovarietal extra virgin olive oil from Marche and so we succeeded for Processing IR data with:

- Coroncina, Mogliano and Raggia in the UNEQ model with 5 Categories;
- Coroncina in the SIMCA model with 3 categories.

Then we have been successful for Processing UVvis Data with:

- Mogliano in SIMCa model with 5 categories;
- Coroncina in UNEQ model with 3 categories.

In order to build stable and strong models, this study could be extended with a higher number of samples, because it would be very helpful in safeguarding and promoting monovarietals olive oils in our territory.

The second aim of this study was going to determine some correlation between the spectroscopic measurements (FT-IR) and chemical parameters of olive oil, and we found that, for:

- Fruity and Bitter (sensory parameters)
- P-Coumaric and Pinoresinol (phenols)
- 2-Hexenal and Hexanal (volatiles)

The success of this and other similar studies<sup>xv</sup> in the literature indicates that FTIR spectroscopy in combination with chemometric techniques have potential of predicting other quality parameters of olive oil in a shorter analysis time compared to chromatographic techniques<sup>xvi</sup>. This could be interesting for the industry, in fact rapid analyses of these chemical components would provide better control of quality during processing and storage and also help in determining the authenticity of the product.

#### References

- 
- <sup>I</sup> Aparicio R., "Characterisation: mathematical procedures for chemical analysis". In Handbook of Olive Oil: Analysis and Properties; Harwood J., Aparicio, R., Eds.; Aspen Publications: Gaithersburg, MD, 2000.
- <sup>II</sup> Angerosa F., Basti C., Vito R., "Virgin olive oil volatile compounds from lipoxygenase pathway and characterization of some Italian cultivars" *J. Agric. Food Chem.* 47 (1999) 836-839.
- <sup>III</sup> C. Oliveros, R. Boggia, R. Casale, M. Casale, C. Armanino, M. Forina "Optimisation of a new headspace mass spectrometry instrument. Discrimination of different geographical origin olive oils" *Journal of Chromatography A* 1076(1-2) (2005) 7-15.
- <sup>IV</sup> M. Casale, C. Casolino, P. Oliveri, M. Forina, "The potential of coupling information using three analytical techniques for identifying the geographical origin of Liguria extra virgin olive oil" *Food Chemistry* 118 (2010) 163-170.
- <sup>V</sup> María E. Escuderos, Sebastián Sánchez, Antonio Jiménez; "Quartz Crystal Microbalance (QCM) sensor arrays selection for olive oil sensory evaluation" *Food Chemistry* 124 (2011) 857-862.
- <sup>VI</sup> Forina, M., Boggia, R., & Casale, M. (2007). "The information content of visible spectra of extra virgin olive oil in the characterization of its origin" *Annali di Chimica*, 97(8), 615-693.
- <sup>VII</sup> Tapp, Henri S. Defernez, Marianne. Kemsley, E Katherine. "FTIR Spectroscopy and Multivariate Analysis Can Distinguish the Geographic Origin of Extra Virgin Olive Oils" *J. Agric. Food Chem.* 51(2003) 6110-6115.
- <sup>VIII</sup> The Unscrambler 9.8, Camo Software AS, 1986-2008.
- <sup>IX</sup> Forina, M., Lanteri, S., Armanino, C., Casolino & Casale, M. 2007. V- PARVUS

---

"An extendable package of programs for explorative data analysis, classification and regression analysis" Dip. Chimica e Tecnologie Farmaceutiche ed Alimentari, University of Genova.

<sup>x</sup> D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.L. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics: Part B, Elsevier, Amsterdam, 1998, p. 213.

<sup>xi</sup> S. Wold, M. Sjostrom, in: B.R. Kowalski (Ed.), Chemometrics, Theory and Application ACS Symposium Series 52, American Chemical Society, Washington, DC, 1977, pp. 243-255.

<sup>xii</sup> M. Forina, S. Lanteri, M. Casale, M.C. Cerato Oliveros, Chemom. Intell. Lab. Syst. 87 (2007) 252–261.

<sup>xiii</sup> Barros, A., Freire, I., Gonçalves, B., Baccelar, F., Gomes, S., Lopes, J., et al. (2012). Evaluation of chemical and phenotypic changes in Blanqueta, Cobrançosa and Galega during ripening. CYTA – Journal of Food. <http://dx.doi.org/10.1080/1947.6337.2012.705329>.

<sup>xiv</sup> INE (2012). Inquérito à produção de azeite. International Olive Oil, key-figures on the world market for olive oils, Data adopted at the 100th session of the IOOC (Madrid, Spain), 19–23 November 2012).

<sup>xv</sup> Oguz Uncu, Banu Ozen, Prediction of various chemical parameters of olive oils with Fourier transform infrared spectroscopy, Izmir Institute of Technology, Department of Food Engineering, Urla-Izmir, Turkey.

<sup>xvi</sup> Irene Gouvinhas, José M.M.M. de Almeida, Teresa Carvalho, Nelson Machado, Ana I.R.N.A. Barros Discrimination and characterization of extra virgin olive oils from three cultivars in different maturation stages using Fourier transform infrared spectroscopy in tandem with chemometrics. CITAB University of Trás-os-Montes and Alto Douro, 5001-801 Vila Real, Portugal. Department of

---

Physics, University of Trás-os-Montes and Alto Douro, 5000-801 Vila Real, Portugal.