3. EXPERIMENTAL

3.1 Data sets

Three data sets are analyzed in this thesis; here their characteristics are described.

First data set regards the Lambrusco wine, it is a PDO wine produced in the Modena district in Italy, it exists in three different varieties Grasparossa, Salamino and Sorbara taking the name from the homonymous grape variety. The production protocol allows for the possibility of having grape blendings within each PDO, as long as some requirements are fulfilled. Thus, the "Lambrusco Sorbara" needs to consist of, at least, 60% Sorbara grapes, while the remaining 40% can be made of either Salamino grapes only, or from a mixture of Salamino grapes with other minor Lambrusco grapes (the latter must not exceed the 15% of the total composition). The "Lambrusco Salamino of Santa Croce" must consist of, at least, 85% Salamino grapes, while the remaining 15% can be from other minor Lambrusco grapes (Ancellotta and Fontana) and the "Lambrusco Grasparossa of Castelvetro" consists of, at least, 85% of Grasparossa grapes and 15% of other minor Lambrusco grapes (Malbo and Gentile). Therefore, with the only exception of Sorbara wine, for which the law provides a percentage of Sorbara grapes that can be as low as 60%, the other PDOs are composed of, at least, 85% of the respective pure grape varieties [1-4]. One hundred and ten bottles of Lambrusco wine, coming from different producers, were sampled and analyzed by High Performance Liquid Chromatography, coupled with UV detection (HPLC-DAD). The wine bottles were provided directly by the winegrowers and belong to three different varieties: 38 of Grasparossa, 38 of Salamino and 34 of Sorbara. Samples were randomized prior to analysis to avoid any experimental drifts, taking care of analyzing in a day samples belonging to different varieties. Some replicates are also provided of samples to verify reproducibility of experimental procedures.

Second data set regards Extra Virgin Olive Oil (EVOO) coming from different Mediterranean countries, this data set was provided by a previous work of doctoral thesis in our group of research from Dr. Durante [5]. The EVOO samples, object of this study, have been obtained from a Producer Consortium and are of assured geographical origin. They belong to different olive cultivars and come from different geographical areas namely, Liguria (northern of Italy), Apulia (southern of Italy), Greece, Tunisia and Spain. The main commercial interest is distinguishing Liguria EVOOs from the rest, being these the most estimated and high valued ones, since Taggiasca cultivar prevails, it is produced from the fruit of olive tree, Olea Europea, by mechanical press and without application of refining processes and more over is identified by the PDO trademark [6]. Samples of 72 oils have been analyzed by HS-SPME/GC-MS (Head Space – Solid Phase Micro Extraction/Gas Chromatography – Mass Spectrometry). The samples acquisition order was randomized, taking care to always acquire at least some samples belonging to each different region in the same day.

Third data set regard samples of red wine, it is a data set of literature, available for download at the web site http://www.models.kvl.dk/ the data set contain 42 samples of the 70 considered here, in which are included further samples collected in the same conditions by the same authors. Wines object of study were produced from the same grape (Cabernet Sauvignon) and belonging to different geographical areas and producers, they were collected from supermarkets and analyzed by mean of HS/GC-MS (Head-Space Gas Chromatography - Mass Spectrometry).
The countries of origin of samples are: Argentina and Chile, united in a unique class South America (30 samples), Australia (20 samples) and South Africa (20 samples). In this case the interest is that of recognize the country of provenience.

3.2 Chromatography

This section describes the modality of acquisition of each data set. Following the techniques used, are briefly illustrated for those who have not familiarity with chemical instrumentation.

Chromatography is one of the most widespread analytical techniques, which allows separation of the different components of a mixture thanks to their different affinity with two phases; these phases are a stationary one and a mobile one. The stationary phase can be solid or liquid supported by a solid and the mobile can be a liquid or a gas. The mixture is eluted with the mobile phase on a column containing (or being itself) the stationary phase and the different species of the mixture come out from the column at different time depending on their interaction with the phases. At the end of the column a detector convert in a signal the chemical information and for each compound a curve similar to a Gaussian is produced in the final plot. Since this signal is proportional to the concentration of the species in the analyzed mixture, also quantitative information can be obtained from chromatography [7]. Here we refer only to liquid chromatography (HPLC for the Lambrusco data set) and gas chromatography (GC for EVOO and WINE data set) the distinction is fundamentally due to the physical state of the mobile phase. This restricts the use of gas chromatography only for volatile analytes, namely small molecules, and makes it a perfect method for the analysis of aromas, using the instrumentation like an electronic nose, explained here. The volatile fraction is characteristics of the samples analyzed and represent a form of indirect indicators of traceability. On the contrary liquid chromatography analyzes components dissolved in solutions, and it is a more suitable methods for determining also molecules with higher weights. When the detector used, is complex and provides spectral information, whether a UV spectrum or a mass spectrum like here (but also fluorescence), it is said that the two techniques are coupled, from here the term hyphenated.

3.2.1 HPLC-DAD

Lambrusco. The wine samples were purified using solid phase extraction, as suggested by literature [8]. For this purpose, Supelco DSC-18 cartridges with 6mL tubes were selected and the extraction procedure was carried out using a manifold system connected to a vacuum pump. This kind of extraction is the most used in phenolic compound purification. The cartridges for the extraction were conditioned rinsing them with 5 mL of methanol and 5 mL of water; after that, 1 mL of the wine sample was loaded. The retained fraction in the cartridge was first washed with 5mL of water and then extracted with 6 mL of diethyl ether. The collected ether fraction was evaporated to dryness under nitrogen flow and then dissolved in a water-methanol (80-20 %) solution. Although ether is supposed to elute most of the phenolic compounds, the cartridges after elution still presented a dark purple color, which means that some compounds were retained. The obtained extract was then filtered and injected in a system for chromatography.

Wine extracts were analyzed by reversed phase liquid chromatography by a Beckman System Gold (USA) for HPLC, with a Model 126 pump built in binary high pressure gradient, coupled with a Model 168 diode array detector of the same producer. The column used was a reversed-phase Atlantis dC18 (250 mm x 4.6 mm, 5 µm packing) Waters-Milford-MA. The injector consisted of a Model 7015 Reodyne valve equipped with a 100 µL loop.

The mobile phase was formed by two solvents: solvent A was water (0.1% TFA) and solvent B was 80 % acetonitrile and 30% water (0.1% TFA). An elution linear gradient was used following the scheme: 0.00min: 0% B; 1.00min: 20% B; 19.00min: 40% B; 29.00min: 100% B; 39.00min:0% B. The flow rate was set at 0.6 mL/min. The system was thermostatted at 40°C. The wavelength range in the diode array detector was from 220 to 430 nm with a resolution of $\Delta\lambda$=2 nm. For the preparation of the mobile phase and the sample treatment, acetonitrile and methanol for HPLC-Gold-Ultragradient by Carlo Erba and water obtained from a Milli-Q purification system (Millipore) were used. Diethyl ether (a.r.) by Riedel-de Haën was used for the solid phase extraction. Gallic acid RPE provided by Carlo Erba Analyticals, (+)-

catechin, syringic acid, caffeic acid, vanillin, p-coumaric acid, myrecetin, quercetin provided by Sigma Aldrich were used to prepare the standards of phenolic compounds [9].
In arranging the dataset the first and last part of the chromatogram, are not considered since there are not chromatographic peak, so the final dimension of the dataset are 110x1211x106.

3.2.2 GC-MS

Liguria EVOO. For sampling the volatile compounds a PDMS/CAR/DVB 50/30 µm fiber was chosen. According to the literature [10-11] this fiber is absorbed compounds in the range of C_3-C_{20}.
After mild mixing, an aliquot of 5.00 mL of each sample was transferred by a calibrated micro-syringe in a 10 mL flask equipped with a screw-top quick fit adaptor and a Teflon-coated silicone septum. The flask was held in a warm water bath at 40.0 °C for 30 min to allow equilibrium between the sample and headspace, prior to SPME sampling. The fiber was inserted into the sample container trough the septum and exposed to the headspace for 30 min. Finally, it was desorbed in the GC injector for 4 min by splitless injection mode. GC-MS was performed by an Agilent 6890N Gas Chromatograph, equipped with an Agilent 5973 mass spectrometer detector and provided with a non polar capillary column of fused silica (CP-SIL5; length 60m; internal diameter 0.25 mm, film thickness 1 µm). Helium was used as carrier gas (helium column flow was kept constant during the chromatographic run at 0.8 mL/min). The injector temperature was 270 °C and the split valve was closed for 5 min. after the sample injection and then opened during all the chromatographic analysis. The injector was equipped with a special deactivated SPME glass insert (0.75 mm internal diameter): the GC oven starting temperature was 50 °C and it was increased at 2 °C/min to 120 °C. Then it was raised at a ramp rate of 6 °C/min up to 180 °C and afterwards at 12 C/min to 250 °C and held for 5 min. Finally it was increased at 15 °C /min to 270 °C and held for 5 min.
The chromatograms were acquired at constant sampling time, ν=20 Hz, for a total time of 67 minutes and 15 seconds, giving a 13458 data points signal.

The mass-spectrometer interface temperature was set at 250 °C. The temperature of the ion source was 230 °C, electron energy 70 eV and quadruple temperature 150 °C. The chromatograms were acquired for a total time of about 67 minutes, but the first 3 minutes and the last 10 minutes in the retention mode were cut because there were no peaks at all. The acquired mass range was 35-250 amu. Array dimensions are: 72 (samples) x 1514 (Retention times) x 216 (m/z). After elimination of variables with almost null variance, both for the chromatographic and mass spectrum directions, the final three-way array is of dimensionality: 72x1039x161. Duplex algorithm [12] has been used on TIC signals to split the samples into training and test sets.
In order to use two-way classification and class modeling methods the data array has been rearranged by row-wise unfolding obtaining respective dimensionality for training set and test set of 42x49644 and 30x49644, respectively.

Foreign WINE. In both the wine samples preparation and GC run, the samples were randomized. Wine samples of 10 mL each, without sample pretreatment, were added directly into the 100 mL purge flask and 2 mL 4-methyl-1-penthanol in water (50 mg/L) was added as an internal standard. The samples were equilibrated to 30 ± 1°C in a circulating water bath and then purged with nitrogen (75 mL/min) for 20 min. The volatile compounds were collected on a Tenax-TA trap. The trapped volatiles were desorbed using an automatic thermal desorption unit (ATD 400, Perkin Elmer, Norwalk, USA) and the volatiles transferred to a gas chromatograph - mass spectrometer through a heated (225 °C) transfer line.
Separation of aroma compounds was carried out on a gas chromatography system (HP 6890 GC with an autosampler for liquid samples) with a 30 m long DB-Wax capillary column with an internal diameter of 0.25 mm, and with a 0.25 mm film thickness. The column flow rate was 1.0 mL/min using helium as a carrier gas. The GC was equipped with a mass spectrometric detector (Agilent 5973 Mass Selective Detector) operating in the electron

ionization (EI) mode at 70 eV. Mass-to-charge ratios between 15 and 300 were scanned twice for each GC scan. GC inlet was held at 250 °C and MS transfer line maintained at a temperature of 280 °C [13].

The data array has the following dimension 70 wine samples, 2700 scans of chromatogram in the second mode and 200 points for the spectra in the third mode (70x2700x200) reduced to 70x684x99.

The data set is split in two subsets, one set is used to build the model (training set, 46 samples partitioned in 20 South America, 13 Australia, 13 South America) and the other is projected in order to test predictive ability of the models (test set, 24 samples partitioned in 10 South America, 7 Australia, 7 South America). The sets have been selected by using the Duplex algorithm [12], it is applied class-wise with a 2:1 training/test splitting ratio on the data matrix containing the TIC chromatograms.

For the two way methods, the data are unfolded in matrices, variables having zero or almost zero variance are deleted, leading to training and test matrices of final size 46x21350 and 24x21350, respectively.

References

[1] Decree n. 187 Jul. 2009; published on Italian G.U. Aug. 2009 (Production Regulations of Lambrusco Sorbara).

[2] Decree n. 188 Jul. 2009; published on Italian G.U. Aug. 2009 (Production Regulations of Lambrusco Salamino).

[3] Decree n. 184 Jul. 2009; published on Italian G.U. Aug. 2009 (Production Regulations of Lambrusco of Modena).

[4] (Production Regulations of Lambrusco, available on line only in Italian). http://www.tutelalambrusco.it/disciplinari/sorbara.htm

[5] C. Durante; *Quality and Safety assurance of foods: a multi-way approach*. Doctoral Thesis (2005).

[6] EC Regulation n° 123/1997, (Production Regulations of Extra Virgin Olive Oil DOP Riviera Ligure, available on line in English). http://www.oliorivieraligure.it/controllo/disciplinare-piano-di-controllo?lang=en

[7] D. A. koog, J. J. Leary; *Principles of Instrumental Analysis*. 4th edition. Journal of Chemical Education (1992) *69* (8), p A224.

[8] S. Pérez-Magariño, M. Ortega-Heras, E. Cano-Mozo; *Optimization of a solid phase extraction method using copolymer sorbents for isolation of phenolic compounds in red wines and quantification by HPLC*. Jounrnal of Agricultural and Food Chemistry (2008) 56 11560-11570.

[9] E. Salvatore, M. Cocchi, A. Marchetti, F. Marini, A. de Juan; *Determination of phenolic compounds and authentication of PDO Lambrusco wines by HPLC-DAD and chemometric techniques*. Analytica Chimica Acta (2013) 761 34-45.

[10] S. Vichi, L. Pizzale, L. S. Conte, S. Buxaderas, E. López-Tamames; *Solid-Phase Microextraction in the Analysis of Virgin Olive Oil Volatile Fraction: Characterization of Virgin Olive Oils from Two Distinct Geographical Areas of Northern Italy*. Journal of Agricultural and Food Chemistry (2003) 51 6572-6577.

[11] S. Vichi, A. I. Castellote, L. Pizzale, L. S Conte, S. Buxaderas, E. López-Tamames; *Analysis of virgin olive oil volatile compounds by headspace solid-phase microextraction coupled to gas chromatography with mass spectrometric and flame ionization detection*. Journal of Chromatography A (2003) 983 19-33.

[12] R. D. Snee; *Validation of Regression Models: Methods and Examples*. Technometrics (1977) 19 415-428.

[13] T. Skov, D. Balabio, R. Bro; *Multiblock Variance Partitioning. A new approach for comparing variation in multiple data blocks*. Analytica Chimica Acta (2008) 615 18-29.