UNIVERSITÀ DELLA CALABRIA

Dipartimento di Chimica

International Doctorate School of Science and Technique
"Bernardino Telesio"

*"ORGANIC MATERIALS OF PHARMACOLOGICAL INTEREST"*

**(OMPI, XXIV CYCLE; CHIM/01)**

# Traceability of Foodstuffs by High Tech methodologies of Mass Spectrometry

Ph.D. Thesis

**CANDIDATE**

DOTT. ATTILIO NACCARATO

---

A.Y 2008 - 2011

**UNIVERSITÀ DELLA CALABRIA**

**FACOLTÀ DI SCIENZE MATEMATICHE FISICHE E NATURALI**

DIPARTIMENTO DI CHIMICA

*SCUOLA DI DOTTORATO DI SCIENZA E TECNICA "BERNARDINO TELESIO"*

*"MATERIALI ORGANICI DI INTERESSE FARMACEUTICO"*

(OMPI, XXIV CYCLE; CHIM/01)

# *RINTRACCIABILITÀ HIGH TECH DI AGROALIMENTI MEDIANTE METODOLOGIE DI SPETTROMETRIA DI MASSA*

**DIRETTORE DELLA SCUOLA**

PROF. ROBERTO BARTOLINO

**COORDINATORE**

PROF. BARTOLO GABRIELE

**SUPERVISORE**

PROF. GIOVANNI SINDONA

**CANDIDATO**

DOTT. ATTILIO NACCARATO

A.Y. 2008 - 2011

**UNIVERSITÀ DELLA CALABRIA**

**FACOLTÀ DI SCIENZE MATEMATICHE FISICHE E NATURALI**

DIPARTIMENTO DI CHIMICA

DOCTORATE SCHOOL OF SCIENCE AND TECHNIQUE
"BERNARDINO TELESIO"

"ORGANIC MATERIALS OF PHARMACOLOGICAL INTEREST"
(OMPI, XXIV CYCLE; CHIM/01)

# TRACEABILITY OF FOODSTUFFS BY HIGH TECH METHODOLOGIES OF MASS SPECTROMETRY

**SCHOOL DIRECTOR**
PROF. ROBERTO BARTOLINO

**COORDINATOR**
PROF. BARTOLO GABRIELE

**SUPERVISOR**
PROF. GIOVANNI SINDONA

**CANDIDATE**
DOTT. ATTILIO NACCARATO

A.Y. 2008 - 2011

# Contents

**Preface** ..... V

## Chapter 1

*Food traceability: Legislation and Analytical methods*

1.1 Food traceability ..... 3

1.2 Analytical techniques and markers used in food traceability ..... 5

1.2.1 Markers identification by Gas Chromatography (GC) ..... 6

1.2.2 Markers identification by High Performance Liquid Chromatography (HPLC) ..... 9

1.2.3 Markers identification by Infrared Spectroscopy (IR) ..... 11

1.2.4 Markers identification by Nuclear Magnetic Resonance Spectroscopy (NMR) ..... 13

1.2.5 Multielement profile and isotopic ratio as traceability markers ..... 15

References ..... 22

## Chapter 2

*Chemometric treatment of data*

2.1 Introduction ..... 33

2.2 Preliminary data analysis ..... 34

2.2.1 Exploratory and unsupervised pattern recognition techniques ..... 36

2.2.2 Variable selection and reduction ..... 39

2.3 Supervised pattern recognition techniques ..... 40

2.3.1 Linear discriminant analysis (LDA) ..... 41

2.3.2 Partial least square discriminant analysis (PLS-DA) ..... 43

2.3.3 Soft independent modeling of class analogy (SIMCA) ..... 46

2.3.4 K-Nearest neighbor (KNN) ..... 47
2.3.5 Artificial neural networks (ANN) ..... 49

References ..... 53

## Chapter 3

*Results and discussion*

**3.1** Secondary metabolites of Olea europaea leaves as markers for the discrimination of cultivars and cultivation zones by multivariate analysis ..... 59

**3.2** Traceability of Tomatoes and Triple Concentrated Tomato Pastes ..... 68

**3.2.1** Investigating the Origin of Tomatoes and Triple Concentrated Tomato Pastes through Multielement Determination by Inductively Coupled Plasma Mass Spectrometry and Statistical Analysis ..... 69

**3.2.2** The volatile fraction profiling of fresh tomatoes and triple concentrate tomato pastes as parameter for the determination of geographical origin ..... 79

**3.3** Multielement Fingerprinting as a Tool in Origin Authentication of PGI Food Products: Tropea Red Onion ..... 92

**3.4** Authenticity of PGI Clementine of Calabria by multielement fingerprint ..... 97

References ..... 105

## Chapter 4

*Results and discussion*

**4.1** Secondary metabolites of Olea europaea leaves as markers for the discrimination of cultivars and cultivation zones by Multivariate analysis ..... 113

**4.1.1** Chemicals and Instrumentation ..... 113

**4.1.2** Sampling ..... 114

**4.1.3** Sample preparation ..... 115

**4.1.4** Statistical Analysis..... 115

**4.2** Investigating the Origin of Tomatoes and Triple Concentrated Tomato Pastes through Multielement Determination by Inductively Coupled Plasma Mass Spectrometry and Statistical Analysis .....116

**4.2.1** Chemicals and Instrumentation ..... 116

**4.2.2** Sampling..... 118

**4.2.3** Sample preparation ..... 118

**4.2.4** Statistical Analysis..... 119

**4.3** The volatile fraction profiling of fresh tomatoes and triple concentrate tomato pastes as parameter for the determination of geographical origin.....120

**4.3.1** Chemicals and Instrumentation ..... 120

**4.3.2** Sampling..... 120

**4.3.3** Sample preparation ..... 121

**4.3.4** Statistical Analysis..... 122

**4.4** Multielement Fingerprinting as a Tool in Origin Authentication of PGI Food Products: Tropea Red Onion.....123

**4.4.1** Chemicals and Instrumentation ..... 123

**4.4.2** Sampling..... 125

**4.4.3** Sample preparation ..... 125

**4.4.4** Statistical Analysis..... 126

**4.5** Authenticity of PGI Clementine of Calabria by multielement fingerprint .....127

**4.5.1** Chemicals and Instrumentation ..... 127

**4.5.2** Sampling..... 129

**4.5.3** Sample preparation ..... 130

**4.5.4** Statistical Analysis..... 132

References .....133

## Appendix

*Abstracts of published papers*

**A.1** Secondary metabolites of Olea europaea leaves as markers for the discrimination of cultivars and cultivation zones by multivariate analysis ..... 137

**A.2** Investigating the Origin of Tomatoes and Triple Concentrated Tomato Pastes through Multielement Determination by Inductively Coupled Plasma Mass Spectrometry and Statistical Analysis ..... 138

**A.3** The volatile fraction profiling of fresh tomatoes and triple concentrate tomato pastes as parameter for the determination of geographical origin ..... 139

**A.4** Multielement Fingerprinting as a Tool in Origin Authentication of PGI Food Products: Tropea Red Onion ..... 140

**A.5** Sarcosine as a marker in prostate cancer progression: a rapid and simple method for its quantification in human urine by solid-phase microextraction–gas chromatography–triple quadrupole mass spectrometry ..... 141

**A.6** Multistage mass spectrometry in quality, safety and origin of foods ..... 142

**A.7** Characterization of the micrites in the Late Miocene vermetid carbonate bioconstructions, Salento Peninsula, Italy: Record of a microbial/metazoan association fingerprint ..... 143

# Preface

The Mediterranean diet has its origins in the history of our country. It has come to us directly from the traditionally Greek eating habits. These habits has rooted and were maintained through the centuries. The traditional food in southern Italy produced, was originally based on starches (bread and pasta), vegetables, olive oil, small quantity of fish and meat, and include now other products that are established elements of this important nutritional style. Many of these products, unfortunately, do not come from the countries of the Mediterranean basin, but are often imported from other countries, with a negative economic impact, and in some cases, they affect consumer's health.

In this context, authentication of food products is of primary importance for consumers, farmers and producers. For consumers, geographical origin assures quality, organoleptic and nutritional characteristics whereas, from an economic point of view, product authentication is fundamental to prevent unfair competition that can eventually affect the regional and even national economy.

In the last decade the European Union moved important steps towards the protection of EU products and their consumers through the introduction of the premium brands (Council Regulation (EEC) No 2081/92) and the concept of traceability (Regulation (EC) No. 178/2002 of 28 January 2002). The implementation of traceability systems in agri-food companies realized in accordance to the standard UNI EN ISO 22005:2008 represents an indispensable tool not only to cope with law obligations, but also to add value to specific characteristics of a product, such as the origin/territoriality and the peculiarity of ingredients, and to satisfy customers' expectations.

This standard, however, does not offer a tool that can identify the origin of food products on reliable scientific basis. So the development of analytical method for

geographical origin identification of food products represents an important goal in order to assure the real provenance of products and has been the object of many scientific publications.

The research activity developed during the three years of Ph.D was devote to investigate markers capable of individuate the geographical origin of foodstuff using analytical techniques of mass spectrometry. In order to recognize the presence of patterns, the experimental data was subjected to chemometric data analysis. In these works unsupervised and supervised patterns recognition techniques such as principal component analysis (PCA), linear discriminant analysis (LDA), soft independent model class analogy (SIMCA), K-nearest neighbors (KNN), partial least squared discriminant analysis (PLS-DA) and artificial neural networks (ANN) were successful applied to cope discrimination issues.

Furthermore in the three years of Ph.D. I contributed to the realization of a review: "*Multistage mass spectrometry in quality, safety and origin of foods*" published on the European Journal Of Mass Spectrometry, and two research works in other research filelds. The first, concerns the development of an analytical method for the determination of the prostate cancer marker sarcosine in urine ("*Sarcosine as a marker in prostate cancer progression: a rapid and simple method for its quantification in human urine by solid-phase microextraction–gas chromatography–triple quadrupole mass spectrometry*", published on Analytical and Bioanalytical Chemistry). The latter concern the determination of biomarkers for the characterization of the micrite rock (“*Characterization of the micrites in the Late Miocene vermetid carbonate bioconstructions, Salento Peninsula, Italy: Record of a microbial/metazoan association”,* published on Sedimentary geology*)*.

In the appendix of the thesis are reported the abstract of all the articles published.

# Chapter 1

*Food traceability:*
*Legislation and Analytical methods*

## 1.1 Food traceability

The first definition of traceability was given in 1987 by an international standard EN ISO 8402.[1] Traceability was identified as "the ability to retrace history, use or location of an entity by the means of recorded identification". Within a firm, all the agents of the production and marketing chain must cooperate to make this traceability concept as efficient as possible. In 1999 the Codex Alimentarius Commission defined traceability as the ability to trace the history, application or location of an entity by means of recorded identifications.[2] Traceability is closely linked to product identity, but it can also relate to the origin of materials and parts, product processing history, and the distribution and location of the product after delivery.

In 2002 European Union (UE) approved the council regulation establishing the principles and requirements of food law in the European Union.[3] Primary objectives of this regulation are to ensure a high level of protection of consumer health and an effective functioning of the UE internal market. This regulation introduces the concept of traceability and complements the regulation (EEC) No 2081/92 [4] in which two separate classes of protected geographical name had been defined: protected designations of origin (POD) and protected geographical indications (PGI). PDO is a brand used for foodstuffs with a strong regional identity that are produced, processed and prepared in a specific geographical area using particularly techniques whereas PGI concerns agricultural products and foodstuffs closely linked to a geographical area in at least one of the stages of production, preparation or processing. In any case, the PGI product has to grow in the region whose name it bears, and it must have a reputation that can be attributed to its geographical origin. The principal aims of this legislation are to protect products against

fraud, imitation and to protect consumers by giving them information on the product for assuring quality, organoleptic and nutritional characteristics. The implementation of traceability systems in agri-food companies and supply chains in accordance to the standards UNI 10939:2001 and UNI 11020:2002 [5] represents an indispensable tool not only to cope with law obligations, but also to add value to specific characteristics of a product, such as the origin/territoriality and the peculiarity of ingredients, and to satisfy customers' expectations.

The standard UNI EN ISO 22005:2008 [6] has replaced the national standards for supply chain traceability (UNI 10939:2001) and for company traceability (UNI 11020:2002) becoming the international reference document for the certification of agri-food traceability systems. The traceability system, alone, cannot guarantee the safety of foodstuff, but it can surely give an important contribution to the achievement of this goal. In fact, whenever a hygienic-sanitary non conformity arises, it allows from one side to go up and retrieve the point of the supply chain where the problem has originated, and from the other side to proceed, if needed, to the "punctual" product recall.

These standards define the principles and specify the requirements for the implementation of a system of traceability in the agri-food companies, however, do not offer a tool that can identify on reliable scientific basis, the origin of food products. The development of analytical method for geographical origin identification of food products represents an important goal in order to assure organoleptic and nutritional characteristics to consumers and to prevent unfair competition that can eventually damage the whole agricultural sector. It has been received more attention with the increasing of mobility and product exchange due to the lowering of frontier and transport costs.

The unambiguous determination of origin has been the object of several scientific publications. A great number of different analytical techniques and parameters have been evaluated for the geographical origin authentication purpose. One of these is represented by the metabolomics approach, whereby the origin of the food is recognized from the distribution of the volatile compounds generated by metabolic pathway of major constituents of food or by the evaluation of markers of the secondary metabolism of plants. Another approach is represented by trace elements analysis carried out by inductively coupled plasma mass spectrometry (ICP-MS). Mono and multi-elemental techniques have been successfully employed in food authentication.[7] Stable isotope ratio and multi-element analysis are widely used in food authentication. The number of publications in which these analytical methods were used has grown significantly in recent years. Indeed is well known that the content of selected elements in foods is related with the soil type and the environmental growing conditions.

## 1.2 Analytical techniques and markers used in food traceability

There is an increasing interest by consumers for high quality food products with a clear geographical origin. These products are encouraged and suitable analytical techniques are needed for the quality control.[8] Various techniques have been studied based on organic constituents, mineral contents or composition, light- or heavy-element isotope ratios, or combinations thereof. These markers have been identified through the use of numerous analytical techniques, both spectroscopy and mass spectrometry. Mass spectrometry (MS) is a powerful and well know analytical technique widely used in several chemical applications. Its application in food traceability, concerns the analysis of markers that can

be both molecules or elements. With this purpose, it is usually coupled with separation techniques such as liquid and gas chromatography. Mass spectrometry has undoubtedly given a great contribution to increasing the applications in this field due to some of their features such as the ability to make a large number of samples, the low limit of detection that can be reached and the capability to identify univocally the analyzed markers. The principal analytical approaches used to traceability purpose can be subdivided into two macro groups: mass spectrometry techniques (GC-MS, HPLC-MS, ICP-MS) and spectroscopic techniques (AAS, ICP-AES, IR, NMR, ecc.)

### 1.2.1 Markers identification by gas chromatography (GC)

Gas chromatography is one of the most widely used techniques for qualitative and quantitative analysis. It is widely used in food analysis, mainly in volatile and semi-volatile composition studies, aromas, and pesticides.[9]

A mixture of compounds to be analyzed is initially injected into the GC where the mixture is vaporized and injected onto the head of a chromatographic column. Elution is brought about by the flow of an inert gaseous mobile phase. Commonly used gases include nitrogen, helium and hydrogen. Gas chromatography is based on partitioning of the analytes between a gaseous mobile phase and a liquid phase immobilized on the surface of an inert solid packing or on the walls of capillary tubing.[10]

There are different GC detectors with various types of selectivity and sensitivity: the flame ionization detector (FID), thermal conductivity detector (TCD), electron capture detector (ECD), photo-ionization detector (PID), flame photometric detector (FPD),

thermionic detector (TID), atomic emission detector (AED) and ozone- or fluorine-induced chemiluminescence detector (FCLD), mass spectrometry (MS).[11]

Some significant advances for GC have been the development of coupling GC to IRMS,[12] two-dimensional GC to analyze very complex samples,[13] and high speed GC.[14] These latter developments allow real-time process monitoring and on-site analysis (portable GC). In comparison with HPLC, the mobile phase in GC has a very limited role in the separation process. Moreover the analysis of polar and ionic molecules with GC and collecting components after GC separation for further analysis are difficult to do and are rarely done.[8]

GC-MS is an instrumental technique, comprising a gas chromatograph coupled to a mass spectrometer.[15] GC separates the components of a mixture and MS characterizes each of the components individually. In this way, one can both qualitatively and quantitatively analyze complex mixtures containing numerous compounds. In order for a compound to be analyzed by GC-MS it must be sufficiently volatile and thermally stable. In GC-MS, the ions required for mass analysis are generally formed by electron impact ionization. Gas molecules exiting the GC are bombarded by a high-energy electron beam (70 eV). As the MS detector is only designed to analyze clean materials careful sample preparation must be considered prior to injection in the gas chromatograph. GC-MS is one of the most widely used techniques and represents the method of choice for the analysis of food volatiles because of its high reproducibility.[16]

*Application*

By analyzing the GC profiles of various compounds (e.g., alkanes, aldehydes, alcohols, acids) present in wine it is possible to classify wines according to their geographical

origin.[17] Determination of the fatty acid composition and corresponding concentrations by GC allowed the geographical discrimination of milk samples[18] and olive oils.[19] Furthermore, determination of the geographical origin of cocoa masses [20] and orange juices [21] were accomplished via GC analysis. GC-MS has also been applied for the determination of the geographical origin of food products. This concerns mainly dairy products. For example, Emmental cheese samples from different countries and regions were easily differentiated by using GC-MS measured compounds.[16] Furthermore, GC-MS elucidated the relationship between the flavoring capabilities and geographical origin of natural whey cultures used for traditional water-buffalo Mozzarella cheese manufacture.[22] GC-MS has also contributed to the detection of specific markers for tracing the geographical origin of food products.[23] In this way the influence of pasture from a certain region on the volatile compounds in Ewes' dairy products (milk and cheese) was shown.[24] Specific markers were also found by GC-MS in honey.[25] These markers indicated if the honey was from Denmark, England, The Netherlands, Spain or Portugal. Recently other paper were published on honey. Alfieris et al. used the GC-MS fingerprinting of headspace volatile compounds to botanical discrimination and classification of honey samples according to their geographical origin.[26] Staminova et al. tracing the geographical origin of honeys based on volatile compounds determined by a head-space solid phase microextraction (SPME) combined with comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry.[27] GC continues to be used successful in traceability applications on other food products. Indeed, in works published during the last two years GC was employed in traceability markers detection of olive oil,[28] potatoes,[29] orujo brandy[30] and roasted hazelnuts.[31]

### 1.2.2 Markers identification by High Performance Liquid Chromatography (HPLC)

Liquid chromatography (LC), the generic name used to describe any chromatographic procedure in which a liquid mobile phase is used for analysis of complex mixtures of low volatile samples. Modern high performance LC systems (HPLC), are firmly established at the forefront of chromatographic techniques. HPLC is used for a wide range of applications. Volatility or thermal stability of the analytes is no longer a limit, as in the Gas Chromatographic (GC) applications, thus making HPLC the method of choice for polymers, polar, ionic and thermally unstable materials. Moreover, sample detection and quantitation can be achieved by means of continuous flow detectors; thus improving accuracy and precision of analysis. In HPLC, a liquid sample, is carried through a chromatographic column by a liquid mobile phase. Analytes separation is determined by the differences in their partitioning behavior between the mobile liquid phase and the stationary phase (column material). Dependent of the type of stationary phase, compounds can be separated based on their charge (weak/strong cation or anion exchange chromatography), molecular mass (size exclusion chromatography), hydrophobicity/polarity (reversed-phase HPLC, hydrophobic interaction chromatography), and specific characteristics (affinity chromatography). Although over the years a large number of LC detectors have been developed and described, the most common detectors in HPLC are an ultraviolet‑visible (UV‑vis) light absorbance detector, fluorescence detector, electrochemical detector and diffractometer detector. More recently, the coupling of HPLC to mass spectrometry supply with quadrupoles, magnetic sectors or time-of-flight (TOF) analyzer, has also had a great expansion into the field of food analysis. The most commonly employed ionization methods are the atmospheric

pressure ionization techniques such as the Electrospray Ionization (ESI) and atmospheric pressure chemical ionization (APCI). Those techniques are called "soft" because they allow to obtain only the molecular ion for each analyte thus avoiding the fragmentation.[32] Additional fragmentation can be achieved by performing in-source collision induced dissociation (CID)[33] in tandem or trap instruments. Tandem mass spectrometry in space and in time provide additional and unique information on the structure of analytes. ESI is useful for polar and ionic solutes ranging in molecular weight from 100 to $150x10^3$ dalton. APCI is applicable to non-polar and medium polarity molecules with a molecular weight from 100 to 2000 dalton.

One of the most important trends is the miniaturization of the HPLC analytical systems. The emergence of micro- and nano- columns have led to the development of Ultra Performance Liquid Chromatography systems (UPLC). The technology takes full advantage of chromatographic principles to run separations using columns packed with smaller particles and/or higher flow rates for increased speed, with superior resolution and sensitivity. Furthermore, two-dimensional LC makes HPLC an even more valuable and powerful analytical tool.[13]

*Application*

European wines from different geographical origin have been correctly classified on basis of the chromatography profiles obtained with HPLC in combination with a UV‑vis and/or fluorescence detector. These HPLC studies encompassed the analysis and quantification of either phenolic compounds,[34] amino acids and biogenic amines,[35] or contaminant ochratoxin A.[36] Apart from wines, HPLC has also been used to

geographically discriminate honey,[37] nuts,[38] olive oil[39] and cheese[40] on basis of the HPLC profiles of flavonoids, metal-binding proteins, triglycerides and peptides, respectively.

Recently HPLC continues to be used successful in traceability applications on other food products. Indeed, in works published during the current year HPLC was employed in traceability markers detection of extra-virgin olive oils,[41] cinnamon bark and cinnamon twig,[42] pigmented potatoes,[43] and main varieties of Argentinean wines.[44] Most studies described above were employed in conjunction with chemometric methods.

### 1.2.3 Markers identification by Infrared Spectroscopy (IR)

Infrared spectrophotometry is a powerful tool for identifying pure organic and inorganic compounds because, with the exception of a few homonuclear molecules, all molecular species absorb infrared radiation. IR spectroscopy is the measurement of the wavelength and intensity of the absorption of infrared light by a sample.[45] In fact IR measures the vibrations of molecules. Each functional group, or structural characteristic, of a molecule has a unique vibrational frequency that can be used to determine what functional groups are present in a sample. When the effects of all the different functional groups are taken together, the result is a unique molecular "fingerprint" that can be used to confirm the identity of a sample.

Infrared spectroscopy is a less satisfactory tool for quantitative analyses than its ultraviolet and visible counterparts because of lower sensitivity and frequent deviations from Beer's law.[10] The infrared portion of the electromagnetic spectrum is divided into three regions; the near-, mid- and far-infrared, named for their relation to the visible spectrum. The far-infrared, (approx. 400–10 $cm^{-1}$) lying adjacent to the microwave

region, has low energy and may be used for rotational spectroscopy. The mid-infrared (MIR) (approx. 4000–400 $cm^{-1}$) may be used to study the fundamental vibrations and associated rotational–vibrational structure, while the higher energy near-infrared (NIR) (14,000–4000 $cm^{-1}$) can excite overtone or harmonic vibrations. This means that NIR can provide more complex structural information than MIR.[8]

IR spectroscopy is a non-invasive and non-destructive technique.[46] The technique is rapid, relatively inexpensive and can be easily applied in fundamental research, in control laboratories, and on-line in the factory to analyze food products. The introduction of the Fourier transform technique in IR (FTIR) has increased the use of IR in food analysis.[47] A FTIR spectrometer obtains infrared spectra by first collecting an interferogram of a sample signal with an interferometer, which measures all of infrared frequencies simultaneously. In this way the technique allows a very rapid screening and quantification of components and therefore a high throughput of samples. IR spectroscopy cannot eliminate the need for more detailed laboratory analyses, but it may help to screen samples that require further examination. Special care is necessary with regard to possible interference between components possessing similar IR spectral regions, which occurs very frequently when food products are analyzed.

*Application*

With respect to MIR, various wines,[48] cheeses,[46] olive oils,[49] and honey[50] have been differentiated on the basis of geographical origin. With NIR the geographical classification of grapes,[51] wines,[52] rice,[53] soy sauce [54] and olive oils [55] have been accomplished. These results were obtained by combining the MIR and NIR data with chemometric methods. Recently application of IR fingerprint in the detection of food

adulteration and fraud has been review.[56] Zhang, Zhang, and Li in 2008 [57] established lamb origin tracing model by near-infrared spectroscopy combined with the cluster type of independent soft-mode method (SIMCA). The result confirmed that the near-infrared spectroscopy was a valid method for lamb origin tracing. Zhao et al.,[58] and Wu et al.,[59] utilized near infrared and high spectroscopy to study the tenderness of the beef. The model was established to describe the near infrared of beef tenderness and, because reached 84.21% of the correct prediction rate means that it can be used for beef tenderness prediction. IR was also used to detect genetically modified food. Bing, Luo, and Huang in 2005 detected genetically modified corn and its parents with near-infrared spectroscopy and they obtained very accurate results which showed that the method could quickly and accurately identify genetically modified corn.[60]

### 1.2.4 Markers identification by Nuclear Magnetic Resonance Spectroscopy (NMR)

NMR is based upon the measurement of absorption of radiofrequency radiation by atomic nuclei with non-zero spins in a strong magnetic field.[61] The absorption of the atomic nuclei is affected by the surrounding atoms, which cause small local modifications to the external magnetic field. In this way detailed information about the molecular structure of a food sample can be obtained. Among nuclei with non-zero spin, the isotopes of hydrogen-1 (spin = 1/2) and carbon-13 (spin = 1/2) are the most used in NMR, although other isotopes as nitrogen-15 (spin = 1/2), oxygen-17 (spin = 5/2), fluorine-19 (spin = 1/2), or phosphorous-31 (spin = 1/2) are also frequently employed.

In food analysis two types of NMR are applied, low resolution NMR (LR-NMR) and high resolution NMR (HR-NMR).[61] Nowadays, LR-NMR instruments (using frequencies of

10–40 MHz) are small, easy to use, and relatively inexpensive which make them suitable to perform rapid and reproducible measurements. However, LR-NMR requires reference methods to carry out quantitative analysis, and in many cases the precision of such reference method is a limiting factor. HR-NMR (using frequencies above 100 MHz) has been applied in many more food authenticity studies than LR-NMR.[62] The advantage of HR-NMR over LR-NMR is that it is possible to obtain much more detailed information regarding the molecular structure of a food sample. The major disadvantage is that it is one of the most expensive analytical techniques to employ, both in terms of initial capital outlay and running costs. Furthermore, extraction procedures may be necessary to enrich the studied compound as the sensitivity of HR-NMR is rather poor.[61] The combination of $^{1}$H NMR or $^{13}$C NMR fingerprinting with advanced chemometric methods provides an original approach to study the profile of a food product in relation to its geographical origin. Although the use of $^{13}$C NMR for this purpose is quite rare.

One of the major current applications of HR-NMR is site-specific natural isotope fractionation (SNIF)-NMR.[61] This technique is often used in food analysis and allows the determination of the geographical origin of foods based on the isotopic ratio of a given nucleus found in a constituent of the analyzed food. This can be explained by the fact that the specific proportions of the particular isotopes of hydrogen and oxygen present in molecules are dependent mainly on climatic and geographical conditions.[62] SNIF-NMR is often combined with IRMS and chemometric methods. The main drawback of SNIF-NMR is that it requires laborious sample preparation involving many purification and concentration steps.[61]

*Application*

$^{13}C$ NMR spectra of olive oils from different Italian regions were sufficiently different to permit their discrimination.[63] $^{1}H$ NMR has been applied more often to classify mediterranean olive oils according to their geographical origin.[64] For example, one approach concerned measuring 1H NMR spectra of phenolic extracts of olive oils for discrimination.[65] Besides olive oils, Italian wines could be differentiated from Slovenian wines[66] and the geographical origin of propolis samples could be determined by using $^{1}H$ NMR.[67] Propolis is a complex resinous substance used by bees to seal their hives and is marketed by health food stores for its claimed beneficial effect on human health. SNIF-NMR has been used for the geographical authentication of various wines.[68] By determining the natural abundance isotopic ratios of hydrogen, oxygen, and carbon, from water and ethanol extracted from the wine it is possible to distinguish between regions. Furthermore, SNIF-NMR has also been applied successfully to identify the geographical origin of natural mustard oils.[69] Recently NMR continues to be used successful in food traceability applications: milk,[70] Ligurian extra virgin olive oils.[71] In 2010 Consonni et al. reviewed the application of NMR to assess geographical origin and quality of traditional food products.[72] Researches published during the current year use NMR fingerprint in traceability of angelica gigas,[73] ginseng,[74] and durum wheat.[75]

### 1.2.5 Multielement profile and isotopic ratio as traceability markers

In the development of traceability methods often inorganic markers such as multielement content are used as well as organic molecules belonging to different classes. the normal range of organic compounds in foods varies with fertilization, climatic conditions in the

year of cultivation, history of fields and variety or species as well as geographical location and soil characteristics, so it is sometimes difficult to be definitive about the authenticity of a material from the determination of organic components. it is well known that the content of selected minerals and trace elements in foods clearly reflects the soil type and the environmental growing conditions. Because of that, evaluation of multielement profile has been proposed to assure the geographical origin of food samples.[76]

Atomic spectroscopic methods are used for the qualitative and quantitative determination of more than 70 elements. Typically, these methods can detect parts-per-million to parts-per-billion amounts, and, in some cases, even smaller concentrations. Atomic spectroscopic methods are, in addition, rapid, convenient, and usually of high selectivity. They can be divided into two groups: optical atomic spectrometry and atomic mass spectrometry.[10]

Spectroscopic determination of atomic species can only be performed on a gaseous medium. Consequently, the first step in all atomic spectroscopic procedures is atomization, a process in which a sample is volatilized and decomposed in such a way as to produce gas-phase atoms and ions. The efficiency and reproducibility of the atomization step can have a large influence on the sensitivity, precision, and accuracy of the method. In short, atomization is a critical step in atomic spectroscopy. Several methods are used to atomize samples for atomic spectroscopic studies. Inductively coupled plasmas, flames, and electrothermal atomizers are the most widely used atomization methods. Flames and electrothermal atomizers are widely used in atomic absorption spectrometry, while the inductively coupled plasma is employed in optical emission and in atomic mass spectrometry .

The selection of the atomic spectroscopic technique to be used for a particular application should be based on the desired result, since each technique involves different measurement approaches.

In atomic emission, thermal or electrical energy is used to bring the analyte species into an excited state, from which they return to their ground state through emission of radiation characteristic of all species that are present and that were sufficiently excited. Thus from the principle of atomic emission spectrometry it is a clearly multielement method. The number of elements that can be determined simultaneously is only limited by the availability of sufficiently sensitive interference-free spectral lines.

In atomic absorption spectrometry we need a primary source delivering monochromatic radiation of which the wavelength agrees with that of a resonance line of the element to be determined. The spectral width must be narrow with respect to the absorption profile of the analyte line. From this point of view atomic absorption is a single-elemental method, of which the dynamic range is usually much lower than in atomic emission spectrometry.

In atomic fluorescence, the excitation can be performed both with white as well as with monochromatic sources, which consequently affects the fluorescence intensities obtainable and the freedom from stray radiation limitations. The latter are particularly low with monochromatic primary sources and when using fluorescence lines with wavelengths differing from that of the exciting radiation. Generally, in atomic fluorescence the linear dynamic range is higher than in atomic absorption and spectral interference as well as background interferences are just as low.[77]

Inductively coupled plasma mass spectrometry (ICP-MS) is a combination of two established techniques, namely the inductively coupled plasma (ICP) and mass

spectrometry (MS). The ICP is an extremely suitable ion source for inorganic MS because the high temperature of the ICP ensures almost complete decomposition of the sample into its constituent atoms and the conditions within the ICP result in highly efficient ionization of most elements in the periodic table and, importantly, these ions are almost exclusively singly charged. ICP-MS is a well-known mature technique because of its high sensitivity and rapid multielement analysis capability for each run.[78] It typically provides limits of detection (LODs) in the range of ppb and ppt for many elements. For some elements (e.g., As, Cr, Mn, Ca, and Fe), not satisfactory LODs values are achieved by quadrupole based ICP-MS because of the polyatomic interferences which overlap the monitored isotopes. In order to eliminate these interferences, it is possible to use a instrument equipped with dynamic reaction cell (DRC) or a sector-field instrument that greatly improves the sensitivity and the selectivity. ICP-MS has continued to be a most powerful technique allowing multi-element detection over a wide linear dynamic range with very low detection limits, while also providing isotopic ratio measurement capability. These features explain because it is applied worldwide to the analysis of a wide variety of matrixes and its market continues to grow, unlike that of ICP optical emission spectrometry (ICP-OES), which has reached a steady state.[78a] Multi-element analysis involves obtaining of a large amount of data constitute by many variables monitored for each sample. Thus, information about geographical origin, need to be extracted by multivariate chemometric data analysis.

Stable isotope ratio (SIR or IRMS) have been used to determine the authenticities of several food products.[7] The development of this analytical method was possible by the advent of high-resolution mass spectrometry techniques in elementary analysis. Indeed, mass spectrometers commonly used in multi and trace elements analysis, do not provide

the sensitivity or the precision required to detect the slight differences in natural isotopic abundances because of the instable nature of the ion source. Multiple-collector (MC) systems are commonly used to minimized the effect of ion beam fluctuations on the measurement. In these instrument each isotope ion beam is measured simultaneously in an appropriately spaced array of collectors. In this way, the error due to the temporal fluctuations of ion beams is removed during the determination of the isotope ratio. The two ion source commonly used in MC systems are inductively coupled plasma (MC-ICP-MS system) and thermal ionization (MC-TIMS system). ICP is a very common and well known ion source that needs no further description. TIMS is an analytical techniques first applied in the early of 1900. It operate under high vacuum and it is useful for elements with relatively low ionization energy. In TI only a small amount of analyte is absorbed onto a metal filament which is then heated causing the atom ionization due to Langmuir effect. For many decades, MC-TIMS was the isotope analytical techniques of choice, but due to instrumental developments (e.g. higher element sensitivities, faster isotope ratio measurements, comparable precision and accuracy) the use of MC-ICP-MS has gained more space in isotope ratio measurement.[79] Stable isotope ratios, of the light elements $^{2}H/^{1}H$, $^{11}B/^{10}B$, $^{13}C/^{12}C$, $^{15}N/^{14}N$, and $^{18}O/^{16}O$, and of the heavy elements $^{87}Sr/^{86}Sr$ and $^{207}Pb/^{206}Pb$, $^{208}Pb/^{206}Pb$, $^{208}Pb/^{207}Pb$, $^{206}Pb/^{204}Pb$ have been used to determine the authenticities of food products.[80] The changes in the ratio of the 'heavy' to 'light' isotope, called fractionation, is related to many physico-chemical processes. For example, the measurement of the stable isotope ratios of hydrogen and oxygen are applicable to the characterization of geographical origin because they are strongly latitude dependent and the fractionation that affect their isotope ratio is due to evaporation, condensation and precipitation mechanisms.[7a] Strontium isotope ratio is affected by the age and Rb/Sr ratio

of the rock materials in which the elements are present. Strontium (Sr) has four naturally occurring stable isotopes: $^{88}Sr$, $^{87}Sr$, $^{86}Sr$ and $^{84}Sr$. Isotopes $^{84}Sr$, $^{86}Sr$ and $^{88}Sr$ occur in constant relative proportions, while $^{87}Sr$ is increased over geologic time by the radioactive β-decay of $^{87}Rb$ (half-life of 48.8 billion years).[81] Thereby, $^{87}Sr/^{86}Sr$ isotope abundance ratio provides a fingerprint for different rock types and can therefore be helpful in authenticity determinations.

*Application*

Methods based on multi-element and isotopic ratio coupled with proper chemometric tools have been applied to a range of foodstuffs in order to permit their geographical origin determination. Published literature was reviewed by Kelly et al. in 2005,[7a] Armenta et al. in 2009,[7b] and Vanhaecke et al. in 2010.[80] The first two reviews closely concerning food forensic. They comment on published studies, organized by matrices and provide data on the analytical and chemometric approach. The work of Vanhaecke et al.[80] concerns "provenancing purposes" in archeometry, environmental, forensic, biological applications including agricultural products and is focused on the applications of isotope ratio of metals and metalloids. Van Ruth and Luykx in 2008 published an overview of analytical methods for determining the geographical origin of food products including the application of IRMS and ICP-MS fingerprint.[8] For each technique are highlighted strengths and weaknesses such as sensitivity, simplicity, time analysis and costs. Since the publication of the last review other articles were published further expanding the range of applications. Our research group contributed demonstrating how multielement distribution can be successful used for forensic food purpose.[82] The use of ICP-MS

equipped with a dynamic reaction cell (DRC) was used to determine 18 trace elements in 36 olive oil samples from different cultivation zone in Italy and LDA allowed to obtain good discrimination between the five groups investigated (Rossano, Andria, Lamezia, Spoleto and Pescara).[83] In the last year two papers concerning olive oil were published by Camin et al..[84] Multielement stable isotope ratio and mineral composition were applied to 539 authentic Italian PDO and PGI extra-virgin olive oils to establish a national databank for olive oils and 267 European extra-virgin olive oils. The analysis of oxygen isotopic ratio $^{18}O/^{16}O$ in bulk oil samples would seem preferable than the isotopic ratio $^{13}C/^{12}C$. With European samples by combining 3 isotopic ratios ($^{13}C/^{12}C$, $^{18}O/^{16}O$ and $^{2}H/^{1}H$) and 14 elements content (Mg, K, Ca, V, Mn, Zn, Rb, Sr, Cs, La, Ce, Sm, Eu, U), an improved separation between classes is obtained. The combination of Sr isotopic and multi-element (rare earth elements) fingerprints provides a suitable tool for the determination of geographical origin of Szegedi Füszerpaprika, an Hungarian paprika with PDO recognition.[85] Numerous works have been focused on discrimination by the origin of food samples using element composition and chemometric analysis. Seafood is a field in which this analytical approach has been used in less extension for geographical identification due to significant matrix-induced spectral and nonspectral interferences.[86] Bendicho et al.[87] demonstrate how the differentiation of mussels from Galicia (Northwest of Spain), product with PDO brand can be achieved using ICP-MS and different pattern recognition techniques (i.e. LDA, SIMCA and ANN). High field strength elements (HFSEs) and rare earth elements (REEs), were used with this aim and an hexapole collision cell using gas 8% $H_2$ and He and flow cell gas 6 ml $min^{-1}$ was used for polyatomic interference correction.

## References

1. EN ISO 8402.
2. Commission, C. A., Food and Agriculture Organization of the United Nations, Word Health Organization. *Codex Alimentarius* **1999**.
3. Simpkins, W. A.; Louie, H.; Wu, M.; Harrison, M.; Goldberg, D., Trace elements in Australian orange juice and other products. *Food Chemistry* **2000,** *71* (4), 423-433.
4. Pupin, A. M.; Dennis, M. J.; Parker, I.; Kelly, S.; Bigwood, T.; Toledo, M. C. F., Use of isotopic analyses to determine the authenticity of Brazilian orange juice (Citrus sinensis). *Journal of Agricultural and Food Chemistry* **1998,** *46* (4), 1369-1373.
5. (a) UNI 10939 (2001). Sistema di rintracciabilita` nelle filiere agroalimentari. Principi generali per la progettazione e l'attuazione. Ente Nazionale Italiano di Unificazione, Milano; (b) UNI 11020 (2002) Sistema di rintracciabilita` nelle aziende agroalimentari. Principi e requisiti per l'attuazione. Ente Nazionale Italiano di Unificazione, Milano.
6. ISO 22005:2008 "Traceability in the feed and food chain – General principles and basic requirements for system design and implementation".
7. (a) Kelly, S.; Heaton, K.; Hoogewerff, J., Tracing the geographical origin of food: The application of multi-element and multi-isotope analysis. *Trends Food Sci. Technol.* **2005,** *16* (12), 555-567; (b) Gonzalvez, A.; Armenta, S.; de la Guardia, M., Trace-element composition and stable-isotope ratio for discrimination of foods with Protected Designation of Origin. *TrAC Trends in Analytical Chemistry* **2009,** *28* (11), 1295-1311.
8. Luykx, D. M. A. M.; Van Ruth, S. M., An overview of analytical methods for determining the geographical origin of food products. *Food Chem.* **2008,** *107* (2), 897-911.
9. Chang, S. K. C.; Holm, E.; Schwarz, J.; Rayasduarte, P., Food. *Anal Chem* **1995,** *67* (12), R127-R153.
10. Skoog, D. A.; West, D. M.; Holler, F. J.; Crouch, S. R., *Fundamentals of analytical chemistry* 8th Edition ed.; Thomson: 2004.
11. Grob, R. L.; Barry, E. F., *Modern practice of gas chromatography*. John Wiley: 2004.
12. Wolfram, M.-A., Stable isotope analysis of fatty acids by gas chromatography–isotope ratio mass spectrometry. *Anal. Chim. Acta* **2002,** *465* (1-2), 63-79.
13. Tranchida, P. Q.; Dugo, P.; Dugo, G.; Mondello, L., Comprehensive two-dimensional chromatography in food analysis. *Journal of Chromatography A* **2004,** *1054* (1-2), 3-16.

14. Eiceman, G. A.; Hill, H. H.; Gardea-Torresdey, J., Gas chromatography. *Anal Chem* **1998,** *70* (12), 321R-339R.

15. Gohlke, R. S.; McLafferty, F. W., Early gas chromatography/mass spectrometry. *J. Am. Soc. Mass Spectrom.* **1993,** *4* (5), 367-371.

16. Pillonel, L.; Badertscher, R.; Froidevaux, P.; Haberhauer, G.; Hölzl, S.; Horn, P.; Jakob, A.; Pfammatter, E.; Piantini, U.; Rossmann, A.; Tabacchi, R.; Bosset, J. O., Stable isotope ratios, major, trace and radioactive elements in emmental cheeses of different origins. *LWT - Food Science and Technology* **2003,** *36* (6), 615-623.

17. (a) Etievant, P.; Schlich, P.; Cantagrel, R.; Bertrand, A.; Bouvier, J. C., Varietal and geographic classification of french red wines in terms of major acids. *J. Sci. Food Agric.* **1989,** *46* (4), 421-438; (b) Shimoda, M.; Shibamoto, T.; Noble, A. C., Evaluation of headspace volatiles of cabernet-sauvignon wines sampled by an on-column method. *J. Agric. Food Chem.* **1993,** *41* (10), 1664-1668.

18. Collomb, M.; Bütikofer, U.; Sieber, R.; Jeangros, B.; Bosset, J.-O., Composition of fatty acids in cow's milk fat produced in the lowlands, mountains and highlands of Switzerland using high-resolution gas chromatography. *International Dairy Journal* **2002,** *12* (8), 649-659.

19. Ollivier, D.; Artaud, J.; Pinatel, C.; Durbec, J. P.; Guerere, M., Triacylglycerol and fatty acid compositions of French virgin olive oils. Characterization by chemometrics. *J. Agric. Food Chem.* **2003,** *51* (19), 5723-5731.

20. Hernandez, C. V.; Rutledge, D. N., Multivariate statistical analysis of gas chromatograms to differentiate cocoa masses by geographical origin and roasting conditions. *Analyst* **1994,** *119* (6), 1171-1176.

21. Ruiz del Castillo, M. L.; Caja, M. M.; Blanch, G. P.; Herraiz, M., Enantiomeric distribution of chiral compounds in orange juices according to their geographical origins. *J. Food Prot.* **2003,** *66* (8), 1448-54.

22. Mauriello, G.; Moio, L.; Genovese, A.; Ercolini, D., Relationships between flavoring capabilities, bacterial composition, and geographical origin of natural whey cultures used for traditional water-buffalo Mozzarella cheese manufacture. *J. Dairy Sci.* **2003,** *86* (2), 486-497.

23. Fernandez, C.; Astier, C.; Rock, E.; Coulon, J. B.; Berdague, J. L., Characterization of milk by analysis of its terpene fractions. *Int. J. Food Sci. Technol.* **2003,** *38* (4), 445-451.

24. Povolo, M.; Contarini, G.; Mele, M.; Secchiari, P., Study on the influence of pasture on volatile fraction of ewes' dairy products by solid-phase microextraction and gas chromatography-mass spectrometry. *J. Dairy Sci.* **2007,** *90* (2), 556-569.

25. Radovic, B. S.; Careri, M.; Mangia, A.; Musci, M.; Gerboles, M.; Anklam, E., Contribution of dynamic headspace GC–MS analysis of aroma compounds to authenticity testing of honey. *Food Chem.* **2001,** *72* (4), 511-520.

26. Aliferis, K. A.; Tarantilis, P. A.; Harizanis, P. C.; Alissandrakis, E., Botanical discrimination and classification of honey samples applying gas chromatography/mass spectrometry fingerprinting of headspace volatile compounds. *Food Chem.* **2010,** *121* (3), 856-862.

27. Stanimirova, I.; Ustun, B.; Cajka, T.; Riddelova, K.; Hajslova, J.; Buydens, L. M. C.; Walczak, B., Tracing the geographical origin of honeys based on volatile compounds profiles assessment using pattern recognition techniques. *Food Chem.* **2010,** *118* (1), 171-176.

28. (a) Pouliarekou, E.; Badeka, A.; Tasioula-Margari, M.; Kontakos, S.; Longobardi, F.; Kontominas, M. G., Characterization and classification of Western Greek olive oils according to cultivar and geographical origin based on volatile compounds. *Journal of Chromatography A* **2011,** *1218* (42), 7534-7542; (b) Pizarro, C.; Rodriguez-Tecedor, S.; Perez-del-Notario, N.; Gonzalez-Saiz, J. M., Recognition of volatile compounds as markers in geographical discrimination of Spanish extra virgin olive oils by chemometric analysis of non-specific chromatography volatile profiles. *Journal of Chromatography A* **2011,** *1218* (3), 518-523.

29. Longobardi, F.; Casiello, G.; Sacco, D.; Tedone, L.; Sacco, A., Characterisation of the geographical origin of Italian potatoes, based on stable isotope and volatile compound analyses. *Food Chem.* **2011,** *124* (4), 1708-1713.

30. Garcia-Martin, S.; Herrero, C.; Pena, R. M.; Barciela, J., Solid-phase microextraction gas chromatography-mass spectrometry (HS-SPME-GC-MS) determination of volatile compounds in orujo spirits: Multivariate chemometric characterisation. *Food Chem.* **2010,** *118* (2), 456-461.

31. Cordero, C.; Liberto, E.; Bicchi, C.; Rubiolo, P.; Schieberle, P.; Reichenbach, S. E.; Tao, Q. P., Profiling food volatiles by comprehensive two-dimensional gas chromatography coupled with mass spectrometry: Advanced fingerprinting approaches for comparative analysis of the volatile fraction of roasted hazelnuts (Corylus avellana L.) from different origins. *Journal of Chromatography A* **2010,** *1217* (37), 5848-5858.

32. (a) Gaskell, S. J., Electrospray: Principles and practice. *J. Mass Spectrom.* **1997,** *32* (7), 677-688; (b) Kebarle, P., A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J. Mass Spectrom.* **2000,** *35* (7), 804-817.

33. Gabelica, V.; Pauw, E. D., Internal energy and fragmentation of ions produced in electrospray sources. *Mass Spectrom. Rev.* **2005,** *24* (4), 566-587.

34. Rodriguez-Delgado, M. A.; Gonzalez-Hernandez, G.; Conde-Gonzalez, J. E.; Perez-Trujillo, J. P., Principal component analysis of the polyphenol content in young red wines. *Food Chem.* **2002,** *78* (4), 523-532.

35. Heberger, K.; Csomos, E.; Simon-Sarkadi, L., Principal component and linear discriminant analyses of free amino acids and biogenic amines in Hungarian wines. *J. Agric. Food Chem.* **2003,** *51* (27), 8055-8060.

36. Bellí, N.; Marín, S.; Sanchis, V.; Ramos, A. J., Review: Ochratoxin A (OTA) in Wines, Musts and Grape Juices: Occurrence, Regulations and Methods of Analysis. *Food Science and Technology International* **2002,** *8* (6), 325-335.

37. Tomasbarberan, F. A.; Ferreres, F.; Garciaviguera, C.; Tomaslorente, F., Flavonoids in Honey of Different Geographical Origin. *Z Lebensm Unters For* **1993,** *196* (1), 38-44.

38. Gómez-Ariza, J. L.; Arias-Borrego, A.; García-Barrera, T., Multielemental fractionation in pine nuts (Pinus pinea) from different geographic origins by size-exclusion chromatography with UV and inductively coupled plasma mass spectrometry detection. *Journal of Chromatography A* **2006,** *1121* (2), 191-199.

39. Stefanoudaki, E.; Kotsifaki, F.; Koutsaftakis, A., The potential of HPLC triglyceride profiles for the classification of cretan olive oils. *Food Chem.* **1997,** *60* (3), 425-432.

40. Di Cagno, R.; Banks, J.; Sheehan, L.; Fox, P. F.; Brechany, E. Y.; Corsetti, A.; Gobbetti, M., Comparison of the microbiological, compositional, biochemical, volatile profile and sensory characteristics of three Italian PDO ewes' milk cheeses. *International Dairy Journal* **2003,** *13* (12), 961-972.

41. Ouni, Y.; Taamalli, A.; Gomez-Caravaca, A. M.; Segura-Carretero, A.; Fernandez-Gutierrez, A.; Zarrouk, M., Characterisation and quantification of phenolic compounds of extra-virgin olive oils according to their geographical origin by a rapid and resolutive LC-ESI-TOF MS method. *Food Chem.* **2011,** *127* (3), 1263-1267.

42. Ding, Y.; Wu, E. Q.; Liang, C.; Chen, J.; Tran, M. N.; Hong, C. H.; Jang, Y.; Park, K. L.; Bae, K.; Kim, Y. H.; Kang, J. S., Discrimination of cinnamon bark and cinnamon twig samples sourced from various countries using HPLC-based fingerprint analysis. *Food Chem.* **2011,** *127* (2), 755-760.

43. Ieri, F.; Innocenti, M.; Andrenelli, L.; Vecchio, V.; Mulinacci, N., Rapid HPLC/DAD/MS method to determine phenolic acids, glycoalkaloids and anthocyanins in pigmented potatoes (Solanum tuberosum L) and correlations with variety and geographical origin. *Food Chem.* **2011,** *125* (2), 750-759.

44. Di Paola-Naranjo, R. D.; Baroni, M. V.; Podio, N. S.; Rubinstein, H. R.; Fabani, M. P.; Badini, R. G.; Inga, M.; Ostera, H. A.; Cagnoni, M.; Gallegos, E.; Gautier, E.; Peral-Garcia, P.; Hoogewerff, J.; Wunderlin, D. A., Fingerprints for Main Varieties of Argentinean Wines: Terroir Differentiation by Inorganic, Organic, and Stable Isotopic Analyses Coupled to Chemometrics. *J. Agric. Food Chem.* **2011,** *59* (14), 7854-7865.

45. Putzig, C. L.; Leugers, M. A.; McKelvy, M. L.; Mitchell, G. E.; Nyquist, R. A.; Papenfuss, R. R.; Yurga, L., Infrared Spectroscopy. *Anal Chem* **1994,** *66* (12), 26R-66R.

46. Karoui, R.; Dufour, E.; Pillonel, L.; Picque, D.; Cattenoz, T.; Bosset, J. O., Determining the geographic origin of Emmental cheeses produced during winter and summer using a technique based on the concatenation of MIR and fluorescence spectroscopic data. *Eur. Food Res. Technol.* **2004,** *219* (2), 184-189.

47. McKelvy, M. L.; Britt, T. R.; Davis, B. L.; Gillie, J. K.; Graves, F. B.; Lentz, L. A., Infrared spectroscopy. *Anal Chem* **1998,** *70* (12), 119r-177r.

48. Picque, D.; Cattenoz, T.; Corrieu, G.; Berger, J. L., Discrimination of red wines according to their geographical origin and vintage year by the use of mid-infrared spectroscopy. *Sci Aliment* **2005,** *25* (3), 207-220.

49. Tapp, H. S.; Defernez, M.; Kemsley, E. K., FTIR Spectroscopy and Multivariate Analysis Can Distinguish the Geographic Origin of Extra Virgin Olive Oils. *J. Agric. Food Chem.* **2003,** *51* (21), 6110-6115.

50. Ruoff, K.; Luginbühl, W.; Künzli, R.; Iglesias, M. T.; Bogdanov, S.; Bosset, J. O.; von der Ohe, K.; von der Ohe, W.; Amadò, R., Authentication of the Botanical and Geographical Origin of Honey by Mid-Infrared Spectroscopy. *J. Agric. Food Chem.* **2006,** *54* (18), 6873-6880.

51. Arana, I.; Jaren, C.; Arazuri, S., Maturity, variety and origin determination in white grapes (Vitis Villifera L.) using near infrared reflectance technology. *J. Near Infrared Spectrosc.* **2005,** *13* (6), 349-357.

52. Liu, L.; Cozzolino, D.; Cynkar, W. U.; Gishen, M.; Colby, C. B., Geographic Classification of Spanish and Australian Tempranillo Red Wines by Visible and Near-Infrared Spectroscopy Combined with Multivariate Analysis. *J. Agric. Food Chem.* **2006,** *54* (18), 6754-6759.

53. Kim, S. S.; Rhyu, M. R.; Kim, J. M.; Lee, S. H., Authentication of rice using near-infrared reflectance spectroscopy. *Cereal Chem.* **2003,** *80* (3), 346-349.

54. Iizuka, K.; Aishima, T., Soy Sauce Classification by Geographic Region Based on NIR Spectra and Chemometrics Pattern Recognition. *J. Food Sci.* **1997,** *62* (1), 101-104.

55. Downey, G.; McIntyre, P.; Davies, A. N., Geographic classification of extra virgin olive oils from the eastern Mediterranean by chemometric analysis of visible and near-infrared spectroscopic data. *Appl. Spectrosc.* **2003,** *57* (2), 158-163.

56. Zhang, J.; Zhang, X.; Dediu, L.; Victor, C., Review of the current application of fingerprinting allowing detection of food adulteration and fraud in China. *Food Control* **2011,** *22* (8), 1126-1135.

57. Zhang, N.; Zhang, D.; Li, S., Preliminary study on origin traceability of mutton by near infrared reflectance spectroscopy coupled with SIMCA method. *Transactions of the Chinese Society of Agricultural Engineering* **2008,** *24* (12 ), 309-312.

58. Zhao, J. W.; Zhai, J. M.; Liu, M. H.; Cai, J. R., The determination of beef tenderness using near-infrared spectroscopy. *Spectroscopy and Spectral Analysis* **2006,** *26* (4), 640-642.

59. Wu, J.-H.; Peng, Y.-K.; Jiang, F., Hyperspectral scattering profiles for prediction of beef tenderness. *Transactions of the Chinese Society for Agricultural Machinery* **2009,** *40* (12), 135-138.

60. Rui, Y. K.; Luo, Y. B.; Huang, K. L.; Wang, W. M.; Zhang, L. D., Application of near-infrared diffuse reflectance spectroscopy to the detection and identification of transgenic corn. *Spectroscopy and Spectral Analysis* **2005,** *25* (10), 1581-1583.

61. Ibañez, E.; Cifuentes, A., New Analytical Techniques in Food Science. *Critical Reviews in Food Science and Nutrition* **2001,** *41* (6), 413-450.

62. Reid, L. M.; O'Donnell, C. P.; Downey, G., Recent technological advances for the determination of food authenticity. *Trends in Food Science & Technology* **2006,** *17* (7), 344-353.

63. Shaw, A. D.; di Camillo, A.; Vlahov, G.; Jones, A.; Bianchi, G.; Rowland, J.; Kell, D. B., Discrimination of the variety and region of origin of extra virgin olive oils using 13C NMR and multivariate calibration with variable reduction. *Anal. Chim. Acta* **1997,** *348* (1-3), 357-374.

64. (a) Mannina, L.; Patumi, M.; Proietti, N.; Bassi, D.; Segre, A. L., Geographical Characterization of Italian Extra Virgin Olive Oils Using High-Field 1H NMR Spectroscopy. *J. Agric. Food Chem.* **2001,** *49* (6), 2687-2696; (b) Rezzi, S.; Axelson, D. E.; Héberger, K.; Reniero, F.; Mariani, C.; Guillou, C., Classification of olive oils using high throughput flow 1H NMR fingerprinting with principal component analysis, linear discriminant analysis and probabilistic neural networks. *Anal. Chim. Acta* **2005,** *552* (1-2), 13-24.

65. Sacco, A.; Brescia, M.; Liuzzi, V.; Reniero, F.; Guillou, G.; Ghelli, S.; van der Meer, P., Characterization of italian olive oils based on analytical and nuclear magnetic resonance determinations. *J. Am. Oil Chem. Soc.* **2000,** *77* (6), 619-625.

66. Brescia, M. A.; Košir, I. J.; Caldarola, V.; Kidrič, J.; Sacco, A., Chemometric Classification of Apulian and Slovenian Wines Using 1H NMR and ICP-OES Together with HPICE Data. *J. Agric. Food Chem.* **2002,** *51* (1), 21-26.

67. Watson, D. G.; Peyfoon, E.; Zheng, L.; Lu, D.; Seidel, V.; Johnston, B.; Parkinson, J. A.; Fearnley, J., Application of principal components analysis to 1H-NMR data obtained from propolis samples of different geographical origin. *Phytochem. Anal.* **2006,** *17* (5), 323-331.

68. (a) Martin, G. J.; Mazure, M.; Jouitteau, C.; Martin, Y. L.; Aguile, L.; Allain, P., Characterization of the geographic origin of Bordeaux wines by a combined use of isotopic and trace element measurements. *Am J Enol Viticult* **1999,** *50* (4), 409-417; (b) Ogrinc, N.; Košir, I. J.; Kocjančič, M.; Kidrič, J., Determination of Authenticy,

Regional Origin, and Vintage of Slovenian Wines Using a Combination of IRMS and SNIF-NMR Analyses. *J. Agric. Food Chem.* **2001,** *49* (3), 1432-1440.

69. Remaud, G. S.; Martin, Y.-L.; Martin, G. G.; Naulet, N.; Martin, G. J., Authentication of Mustard Oils by Combined Stable Isotope Analysis (SNIF-NMR and IRMS). *J. Agric. Food Chem.* **1997,** *45* (5), 1844-1848.

70. Sacco, D.; Brescia, M. A.; Sgaramella, A.; Casiello, G.; Buccolieri, A.; Ogrinc, N.; Sacco, A., Discrimination between Southern Italy and foreign milk samples using spectroscopic and analytical data. *Food Chem.* **2009,** *114* (4), 1559-1563.

71. Mannina, L.; Marini, F.; Gobbino, M.; Sobolev, A. P.; Capitani, D., NMR and chemometrics in tracing European olive oils: The case study of Ligurian samples. *Talanta* **2010,** *80* (5), 2141-2148.

72. Consonni, R.; Cagliani, L. R., Nuclear magnetic resonance and chemometrics to assess geographical origin and quality of traditional food products. *Advances in food and nutrition research* **2010,** *59*, 87-165.

73. Kim, E. J.; Kwon, J.; Park, S. H.; Park, C.; Seo, Y. B.; Shin, H. K.; Kim, H. K.; Lee, K. S.; Choi, S. Y.; Ryu, D. H.; Hwang, G. S., Metabolite Profiling of Angelica gigas from Different Geographical Origins Using (1)H NMR and UPLC-MS Analyses. *J. Agric. Food Chem.* **2011,** *59* (16), 8806-8815.

74. Lee, A. R.; Gautam, M.; Kim, J.; Shin, W. J.; Choi, M. S.; Bong, Y. S.; Hwang, G. S.; Lee, K. S., A Multianalytical Approach for Determining the Geographical Origin of Ginseng Using Strontium Isotopes, Multielements, and (1)H NMR Analysis. *J. Agric. Food Chem.* **2011,** *59* (16), 8560-8567.

75. Lamanna, R.; Cattivelli, L.; Miglietta, M. L.; Troccoli, A., Geographical origin of durum wheat studied by (1)H-NMR profiling. *Magn. Reson. Chem.* **2011,** *49* (1), 1-5.

76. Gonzalvez, A.; Armenta, S.; de la Guardia, M., Trace-element composition and stable-isotope ratio for discrimination of foods with Protected Designation of Origin. *Trac-Trends in Analytical Chemistry* **2009,** *28* (11), 1295-1311.

77. Broekaert, J. A. C., *Analytical Atomic Spectrometry with Flames and Plasmas*. Wiley-VCH: 2002.

78. (a) Beauchemin, D., Inductively Coupled Plasma Mass Spectrometry. *Anal Chem* **2010,** *82* (12), 4786-4810; (b) Montaser, A., *Inductively Coupled Plasma Mass Spectrometry*. Wiley-VCH: New York, 1998; (c) S. Caroli, *The Determination of Chemical Elements in Food -Applications for Atomic and Mass Spectrometry*. John Wiley: 2007.

79. Becker, J. S., *Inorganic Mass Spectrometry: Principles and Applications*. John Wiley and Sons: Chichester, 2007.

80. L. Balcaen, L. M., F. Vanhaecke, Determination of isotope ratios of metals (and metalloids) by means of inductively coupled plasma-mass spectrometry for provenancing purposes — A review. *Spectrochimica Acta B* **2010,** *65*, 769-786.

81. Garciaruiz, S.; Moldovan, M.; Fortunato, G.; Wunderli, S.; Garciaalonso, J., Evaluation of strontium isotope abundance ratios in combination with multi-elemental analysis as a possible tool to study the geographical origin of ciders. *Anal. Chim. Acta* **2007,** *590* (1), 55-66.

82. Benincasa, C.; Lewis, J.; Sindona, G.; Tagarelli, A., The use of multi element profiling to differentiate between cow and buffalo milk. *Food Chemistry* **2008,** *110* (1), 257-262.

83. Benincasa, C.; Lewis, J.; Perri, E.; Sindona, G.; Tagarelli, A., Determination of trace element in Italian virgin olive oils and their characterization according to geographical origin by statistical analysis. *Anal. Chim. Acta* **2007,** *585* (2), 366-370.

84. (a) Camin, F.; Larcher, R.; Nicolini, G.; Bontempo, L.; Bertoldi, D.; Perini, M.; Schlicht, C.; Schellenberg, A.; Thomas, F.; Heinrich, K.; Voerkelius, S.; Horacek, M.; Ueckermann, H.; Froeschl, H.; Wimmer, B.; Heiss, G.; Baxter, M.; Rossmann, A.; Hoogewerff, J., Isotopic and Elemental Data for Tracing the Origin of European Olive Oils. *J. Agric. Food Chem.* **2010,** *58* (1), 570-577; (b) Camin, F.; Larcher, R.; Perini, M.; Bontempo, L.; Bertoldi, D.; Gagliano, G.; Nicolini, G.; Versini, G., Characterisation of authentic Italian extra-virgin olive oils by stable isotope ratios of C, O and H and mineral composition. *Food Chem.* **2010,** *118* (4), 901-909.

85. Brunner, M.; Katona, R.; Stefánka, Z.; Prohaska, T., Determination of the geographical origin of processed spice using multielement and isotopic pattern on the example of Szegedi paprika. *Eur. Food Res. Technol.* **2010,** *231* (4), 623-634.

86. Cubadda, F.; Raggi, A.; Coni, E., Element fingerprinting of marine organisms by dynamic reaction cell inductively coupled plasma mass spectrometry *Analytical and Bioanalytical Chemistry* **2006,** *384* (4), 887-896

87. Costas-Rodríguez, M.; Lavilla, I.; Bendicho, C., Classification of cultivated mussels from Galicia (Northwest Spain) with European Protected Designation of Origin using trace element fingerprint and chemometric analysis. *Anal. Chim. Acta* **2010,** *664* (2), 121-128.

# Chapter 2

*Chemometric treatment of data*

## 2.1 Introduction

Data analysis has become a fundamental task in analytical chemistry due to the great quantity of analytical information provided by modern analytical instruments. This leads to the availability of multivariate data matrices that require the use of mathematical and statistical procedures, in order to efficiently extract the maximum useful information from data. Several kinds of pattern recognition methods have been applied in food science, that can be divided into several categories depending on how they are compared. Considering the way they achieve the classification, can be divided in two groups: discrimination and modeling techniques. The first such as linear discriminant analysis (LDA), *k*-nearest neighbors (*k*NN), partial least squares discriminant analysis (PLS-DA) and artificial neural networks (ANN), focused on discrimination among classes. On the other hand, soft independent modeling of class analogy (SIMCA) and unequal dispersed classes (UNEQ), are oriented towards modeling classes.

Discriminating techniques are used to build models based on all the categories concerned in the discrimination, whereas disjoint class-modeling methods create a separate model for each category. One of the drawbacks of discriminating methods is that samples are always classified into one of the given categories, even if they do not belong to any of them. Class-modeling methods consider those objects that fit the model for a category as part of the model, and classify as non-members those that do not.

Pattern recognition techniques can also be grouped as unsupervised and supervised. In unsupervised techniques the different samples are classify without the a priori knowledge of their origin. This category includes techniques such as: cluster analysis (CA) and principal component analysis (PCA).

Supervised pattern recognition techniques use the information about the class membership of the samples (class or category) in order to classify new unknown samples in one of the known classes on the basis of its pattern of measurements. Fall in this category techniques such as: linear discriminant analysis (LDA), *k*-nearest neighbors (*k*NN), partial least squares discriminant analysis (PLS-DA), artificial neural networks (ANN), Soft independent modeling of class analogy (SIMCA) and unequal dispersed classes (UNEQ). These techniques, generally, use a common strategy whichever the algorithm applied. Initially the dataset is split in 2 subsets: train, and test set consist of objects of known class membership for which variables are measured. The train set samples are used to build the model. It is derived between a certain number of variables measured on the samples that constitute the training set and their known categories. The last step is the validation of the model using an independent test set of samples, in order to evaluate the reliability of the classification achieved. In some techniques a further calibration set is used for the optimization of parameters characteristic of each multivariate technique. Often a variables selection step is needed in order to retain those variables that contain information for the aimed classification, whereas those variables encoding the noise and/or with no discriminating power are eliminated.

## 2.2 Preliminary data analysis

Generally analytical data are assumed to be obtained by validated analytical methods in terms of precision, accuracy, sensitivity, specificity, uncertainty and robustness. Moreover, the data used for training should be sufficiently large to cover the possible known variation in the problem domain, so the generalization of the models to the not-yet measured data is possible. Supervised pattern recognition requires a training set with

objects of known categories to derive a model for the identification of unknown samples. Therefore, it is mandatory to first establish whether chemical measurements are actually good enough to fit into the predetermined classes, because pattern recognition techniques cannot compensate for poorly designed experiments or inadequate experimental data.[1] This task can be arduous due to the fact that modern analytical techniques are able to generate so much data that the essential information may not be readily evident.

Exploratory data analysis (EDA) (and unsupervised pattern recognition) is commonly used to simplify and gain better knowledge of data sets. The challenge is to remove the redundancy and noise while retaining the meaningful information.[2] Data pre-treatment is essential to avoid wrong (or trivial) conclusions. The first step in EDA comprises a univariate data analysis using basic and descriptive statistics (e.g., calculation of mean, standard deviation, variance, skewness, kurtosis, correlation matrix, *t*-test, *F*-test, ANOVA, box and whisker plots and checking the normality). Then, the presence of outliers, i.e. observations that appear to break the pattern or grouping shown by the vast majority of the samples, should be evaluated since most conventional multivariate methods are sensitive to them.[3] Thus, outliers have to be identified, and then a decision should be made related to the acceptance or rejection of the outliers in the modeling process.

The values of the features measured can differ by orders of magnitude and/or can be measured in different units and/or by different analytical methods or instruments. So some variables could weigh more than others on the results. In order to modify the relative influences of the variables on a model, a data pre-treatment known as *weighting* and/or *scaling* can be performed. Weighting consists of giving each variable a new weight, i.e. multiplying the original values by a constant which differs between the

variables. Thus, the individual contributions are re-adjusted to the outcome on equal basis. The most common scaling methods are the following: *mean-centring* (the average is subtracted from each variable), *standardization* or *autoscaling* (each variable is first centred, and then divided by its standard deviation), *normalization* (variables are divided by the square root of the sum of the variable squares), *constant row sum* (each variable is divided by the sum of all variables for each sample), *normalization variable* (variables are normalized with respect to a single variable), *range transformation* (the minimum value for a variable is set to 0, the maximum value to 1, and all intermediate values lie along a linear range between 0 and 1).[1,4] In general, data pre-treatment is needed prior to the application of multivariate data analysis techniques. Sometimes, it is part of the chemometric technique itself, e.g. data pre-treatment is often performed in principal component analysis (PCA). In discriminating techniques, scaling can only be done over the entire data set. However, class-modeling techniques have the additional possibility of scaling each category separately. So, an additional decision must be made before a classification rule can be deduced since classification results depend on the way scaling is done, as demonstrated for SIMCA.[5]

### 2.2.1 Exploratory and unsupervised pattern recognition techniques

The main EDA technique is PCA, which is often the first step of the data analysis in order to detect patterns in the measured data. PCA is a projection and dimension reduction method for transforming the original measurement variables into new, uncorrelated variables called principal components, which retain as much as possible of the information present in the original data.[6] Each principal component is a linear

combination of the original measurement variables. A set of orthogonal axes, that represent the direction of greatest variance in the data is found.

The linear coefficients of the inverse relation of linear combinations are called the component *loadings*, i.e. the correlation coefficients between the original variables and the principal components. The values that represent the samples in the space defined by the principal components are the component *scores*. The scores were then used as input to the multivariate analyses.[7] A subset comprising a few of the transformed variables only may then also be used in subsequent calculations of relatively reduced complexity. It has been frequently demonstrated that variable reduction can be important for adequate generalization of the derived models.[8]

Other unsupervised pattern recognition techniques can be used for preliminary evaluation of the information contents in the data matrices, such as cluster analysis (CA). Cluster analysis is a method for dividing a group of objects into classes so that similar objects are in the same class. As in PCA, the groups are not known prior to the mathematical analysis and no assumptions are made about the distribution of the variables. Cluster analysis searches for objects which are close together in the variable space. The distance, $d$, between two points in $n$-dimensional space with coordinates $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$ is usually taken as the Euclidean distance defined by

$$d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}$$

As in PCA, a decision has to be made as to whether or not the data are standardized. Standardizing the data will mean that all the variables are measured on a common scale so that one variable does not dominate the others. There are a number of methods for searching for clusters. One method starts by considering each object as forming a ‘cluster’

of size one, and compares the distances between these clusters. The two points which are closest together are joined to form a new cluster. The distances between the clusters are again compared and the two nearest clusters combined. This procedure is repeated and, if continued indefinitely, will group all the points together. There are a variety of ways of computing the distance between two clusters which contain more than one member. The simplest conceptually is to take the distance between two clusters as the distance between nearest neighbors. This is called the single linkage method. It is illustrated in Figure 2.1.[9]

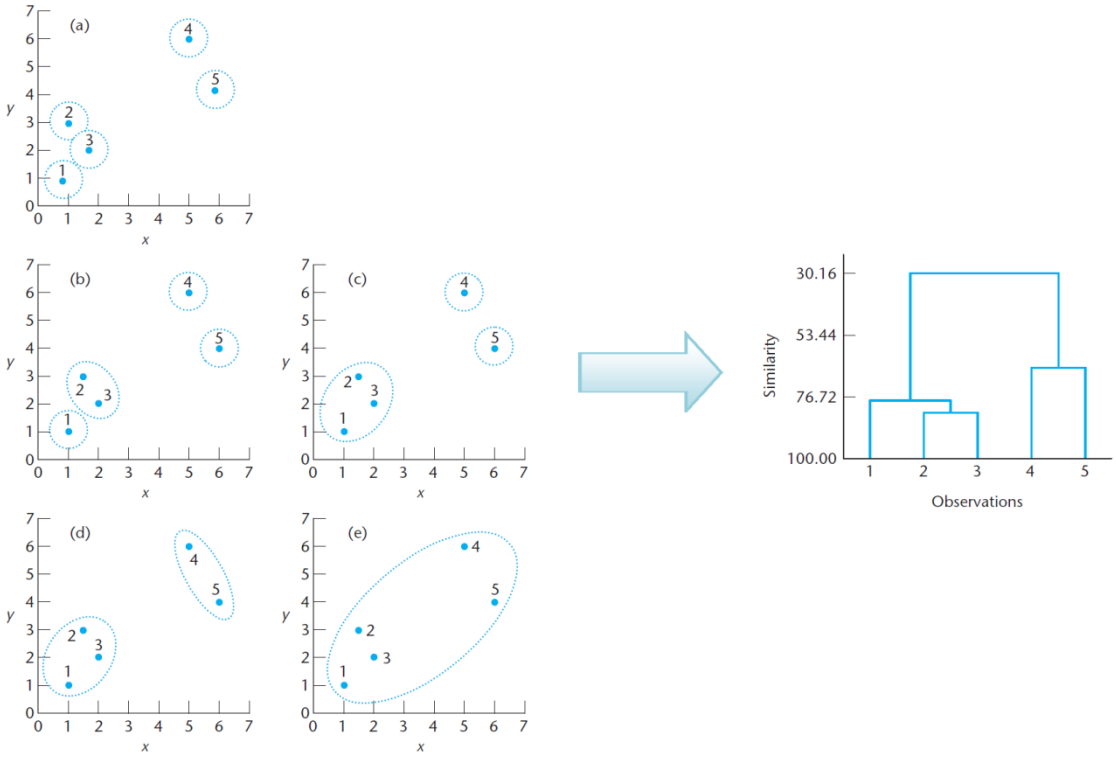


**Figure 2.1** Stages in clustering: the dotted lines enclose clusters (left) and the resulting dendrogram illustrating the stages of clustering (right)

The successive stages of grouping can be shown on a dendrogram. The vertical axis can show either the distance, $d_{ij}$, between two points $i$ and $j$ when they are joined, or alternatively the similarity, $s_{ij}$, defined by $s_{ij} = 100(1 - d_{ij}/d_{max})$.

### 2.2.2 Variable selection and reduction

Variable selection is a preliminary step used in multivariate data analysis, particularly if the number of objects is relatively small, the number of variables is large and many of these variables contain redundant or noisy information. In these cases, a variable or feature selection procedure is required in order not to fall into the overfitting problem. Overfitting takes place if the model learns the idiosyncrasy of the data; then, the noise is modeled as well, and the model loses its generalization ability. Variable selection for discriminant analysis selects a subset of variables that are the most discriminating. The preferred variable selection method is stepwise selection, which is based on a greedy search that sequentially adds or deletes variables from the pool of candidate variables. The addition or deletion of a single variable is performed regarding the largest improvement in the classification, and the process goes on until the search gets trapped in the first local optimum. Several stepwise strategies exist, such as forward stepwise, backward stepwise, forward entry and backward removal.[10] In the forward options, variables are moved into the model, whereas the backward options start with a model with all variables in the model, which are then removed successively. In the forward and backward stepwise strategies, the addition or removal of a variable is considered simultaneously based on probability or Fisher criteria ($p$ or $F$-values). So, forward stepwise variable selection evaluates the addition of new variable and removal of previously entered one. Backward stepwise selection examines the removal of a new variable and the addition of a previously deleted one. The forward entry and backward removal options only enter or remove variables, respectively.

In variable reduction, the number of variables is reduced by combining the original variables attaining a smaller number of principal component or latent variables (which are

variables derived rather than measured) such as respectively principal PCA or PLS components.[4a] This approach is useful for ill-posed data, i.e. data where the number of variables exceeds the number of objects. A certain number of principal components is extracted, deleting the higher order ones and thereby reducing the noise to some extent, and then, a supervised pattern recognition is applied. Elimination of principal components has to be done carefully so that important information for the discrimination should not be lost. The simplest approach to determine the number of significant components is by measuring the prediction error, and then evaluating the prediction residual error sum of squares (PRESS) or root mean square error (RMSE), which can also be expressed as a percentage of variance. This error can be used to decide how many components should be included in the model. The decision can be taken by using a standard cut-off percentage error; ignoring those components from which the error has declined to the noise level; considering the components up to which the error reaches a plateau; selecting PCs whose eigenvalues are equal to 1 or greater; or by looking for a break in a scree plot (eigenvalue versus component number).[1]

## 2.3 Supervised pattern recognition techniques

Supervised pattern recognition techniques have been applied to a wide variety of chemical data (chromatographic, spectrometric, spectrophotometric, spectroscopic, sensorial, etc.) with diverse purposes such as profiling, fingerprinting, authentication, detection of adulteration, food quality assessment, data interpretation, etc..[4a, 11]

### 2.3.1 Linear discriminant analysis (LDA)

Linear discriminant analysis (LDA) is certainly one of the most widely used classification techniques. The data in the training set is used to defines k-1 delimiters (where k is the number of categories) so as the multivariate space of the objects is divided in as many subspaces as the number of classes. Discriminant functions (canonical roots) are obtained as a linear combination of descriptor that maximize the ratio of variance between categories to variance within categories. Being k classes, k−1 canonical roots can be determined if the number of variables is larger than k.[12] For each case the relative canonical scores were computed. Their scatterplot for pairs of canonical roots can be very useful for determining how each discriminant function contributes to the discrimination between categories. The starting point of linear discriminant analysis (LDA) is to find a linear discriminant function (LDF), Y, which is a linear combination of the original variables $X_1$, $X_2$, etc

$$Y = a_1X_1 + a_2X_2 + \cdots a_nX_n$$

The original n measurements for each object are combined into a single value of *Y*, so the data have been reduced from n dimensions to one dimension. The coefficients of the terms are chosen in such a way that *Y* reflects the difference between groups as much as possible: objects in the same group will have similar values of *Y* and objects in different groups will have very different values of *Y*. Thus the discriminant function provides a means of discriminating between the two groups. The simplest situation is that in which there are two classes and two variables, $X_1$ and $X_2$, as illustrated in Figure 2.2(a). This diagram also shows the distribution of the individual variables for each group in the form of dot-plots. For both the variables, there is a considerable overlap in the distributions for the two groups.

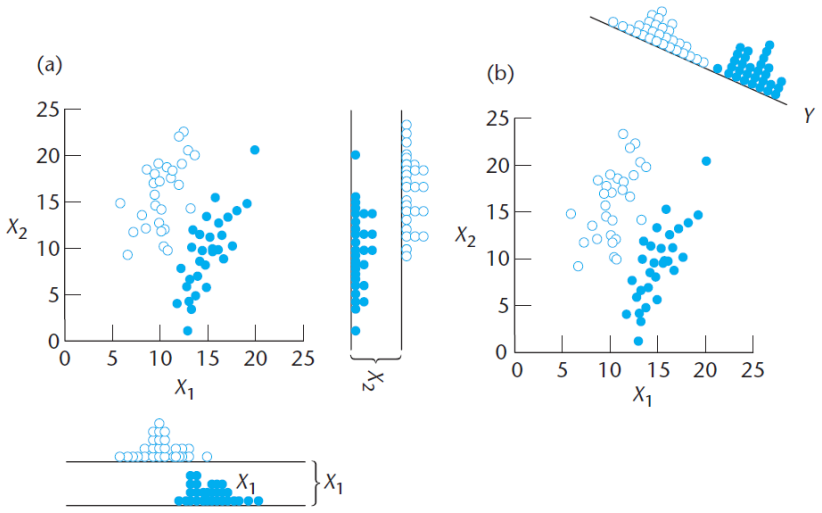


**Figure 2.2** (a) Two groups and the distributions of each variable for each group. (b) The distribution of the linear discriminant function for each group.

The discriminant function is shown by the line labelled Y in Figure 2.2(b), and the value which the function takes for a given point is given by the projection of the point on to this line. Figure 2.2(b) shows the dot-plots of the LDF, *Y*, for each group. It can be seen that there is no overlap between the distribution of *Y* for the two groups. This means that *Y* is better at discriminating between the groups than the original variables.

An unknown object will be classified according to its *Y* value. An initial common sense approach would be to compare *Y* with and , the *Y* values for the means of the two groups. If *Y* is closer to than to then the object belongs to group 1, otherwise it belongs to group 2.[9] In LDA, classes are supposed to follow a multivariate normal distribution and be linearly separated. LDA can be considered, as PCA, as a feature reduction method in the sense that both, LDA and PCA, determine a smaller dimension hyperplane on which the points will be projected from the higher dimension. However, whereas PCA selects a direction that retains maximal structure among the data in a lower dimension, LDA selects a direction that achieves maximum separation among the given classes. LDA is

one variant of discriminant analysis, in which the discrimination boundaries are linear. LDA requires that the variance–covariance matrices of the classes established can be pooled. This is only possible when these matrices can be considered to be equal, which means that 95% confidence ellipsoids have an equal volume (variance) and orientation in the space (covariance).[12a]

Several other functions can be used for discrimination, such as *quadratic discriminant function* (QDA) and *Bayesian classification function*, which are also sub-cases of *regularized discriminant function* (RDA). QDA, which establishes parabolic boundaries, is less subjected to constraints in the distribution of objects in space than LDA, but similarly requires that the number of samples is higher than that of variables. RDA presents the advantage compared to LDA and QDA that is less subjected to constraints without requiring more objects. The Bayesian approach is based on the principle that membership of each class has a prior probability and the measurements are primarily used to refine this.[12a, 13] Canonical variate analysis (CVA) is other discriminant technique, also known as canonical correlation analysis (CCA).CVA differs from LDA, e.g. in the fact that LDA uses a vector containing the membership information, whereas CVA can use a matrix.

### 2.3.2 Partial least square discriminant analysis (PLS-DA)

Partial least squares modeling is a multivariate projection method for modeling a relationship between dependent variables (***Y***) and independent variables (***X***). The principle of PLS is to find the components in the input matrix (***X***) that describe as much as possible the relevant variations in the input variables and at the same time have maximal correlation with the target value in ***Y***, given less weight to the variations that are

irrelevant or noisy. So, PLS models both ***X*** and ***Y*** simultaneously to find the latent variables in ***X*** that will predict the latent variables in ***Y***.

PLS-DA algorithm allows to establish a regression model between the data matrix **X** where each sample is described by a number of variables and a dummy matrix **Y** constructed with zeros and ones represented containing the class membership information. The matrix consisted of as many columns as there are classes and an observation had the value 1 for the class it belongs to and 0 for the rest. The ***X*** matrix consists of the original (preprocessed) data. Matrices ***X*** and ***Y*** are decomposed in a product of other two matrices of scores and loadings. Sample is assigned to one class when the value is above a specific prediction threshold. Indeed, as the predicted value is hardly always exactly 1 or 0, values $\geq 0.5$ are interpreted as indicating membership of the class considered whereas results $\leq 0.5$ indicating nonmembership nonmembership.[14]

Conversely to principal component analysis, PLS-DA model is able to classified unknown samples. During PLS-DA the principal components are rotated to generate latent variables (LVs), which represent those directions that maximize the variance between different classes rather than the total variance as in principal component analysis. An important feature of PLS is that it takes into account errors in both matrices, ***X*** and ***Y***, and assumes that they are equally distributed. Moreover, PLS is suitable for data sets with fewer objects than variables and a high degree of inter-correlation between the independent variables. Several algorithms exist for PLS, each one with certain advantages depending on the case studied. Among them, non-linear iterative partial least squares (NIPALS) algorithm allows the calculation of the principal components one by one.[4a]

PLS compute a number of latent variables with decreasing explained variance. After a number of latent variables, the variation explained by the others LVs can be mostly

attributed to noise. Therefore, in order to avoid overfitting, is extremely important to choose the correct number of latent variables in the calibration step of model development. The optimum number of latent variables can be selected by evaluating error function such as the Root Mean Square Error of Prediction (RMSEP) as function of the number of LVs.[15] Often an external test set is also used to evaluate the true prediction performance of the optimized regression model even though, in the case of PLS it does not seem strictly necessary. The regression coefficients of the PLS model was used to identify important variables. Their size give an indication of which experimental variables have an important impact on the response variables. Variables that have regression coefficients with high values play an important role in regression model.[16] Another parameter that shows the importance of variables is the loading weight. Each variable has a loading weight along each model component. They show how much each variable contributes to explaining the response variation along each model component.

When several dependent data are available for calibration, two approaches can be used in PLS regression: either properties are calibrated for one at a time (PLS1), or properties are calibrated at once (PLS2). In PLS1 model, the *Y* response consists of a single variable. When there is more than one *Y* response a separated model must be constructed for each *Y* response. In PLS2 model, responses are multivariate. PLS1 and PLS2 models provide different prediction set and PLS2 regression give better results than PLS1 regression only if *Y* variables are strongly correlated. In the other case, PLS1 models are generally more robust. For PLS1-DA, one regression for each class has been build. For PLS2-DA, all the classes are included in one regression.[17]

### 2.3.3 Soft independent modeling of class analogy (SIMCA)

SIMCA is the most used of the class-modeling techniques.[4a] It is a class modelling technique that builds a class model based on the significant principal components (PCs) of the category.

The range of the scores of the N principal components used to build the model for each categories are the edges of a hyper volume, the "normal" SIMCA model. Then, in this technique the models (one for each classes) can overlap and/or leave some regions of the multivariate space unassigned. An important consequence of this feature is that SIMCA is able to detect the number of false positive/negative for each class.[18]

The number of principal components for each class in the training set is determined by cross-validation. This way, a sufficient number of principal components are retained to account for most of the variation within each class, while ensuring high signal-to-noise ratio by not including the so-called secondary or noise-laden principal components in the class model.[13] SIMCA determines the class distance and the modeling and discriminatory powers.[1] The class distance can be calculated as the geometric distance from the principal component models. Another approach considers that each class is bounded by a region of space, which represents a percentage of confidence level (usually 95%) that a particular object belongs to a class. The discriminatory power measures how well a variable discriminates between two classes. This differs from the modeling power in the sense that a variable being able to model one class well, it does not necessarily imply being able to discriminate two groups effectively. SIMCA results can be graphically visualized. Thus, a plot of the loadings and the scores of the PCA performed on the training set provide information about outliers, sub-groupings and within-class structure. Moreover, a useful tool for the interpretation of SIMCA results is the so-called Coomans plot, which shows

the discrimination of two classes (Figure 2.3). The distance from the model for class 1 is plotted against that from model 2. The critical distances (usually at 95% of confidence level) are indicated on both axes.

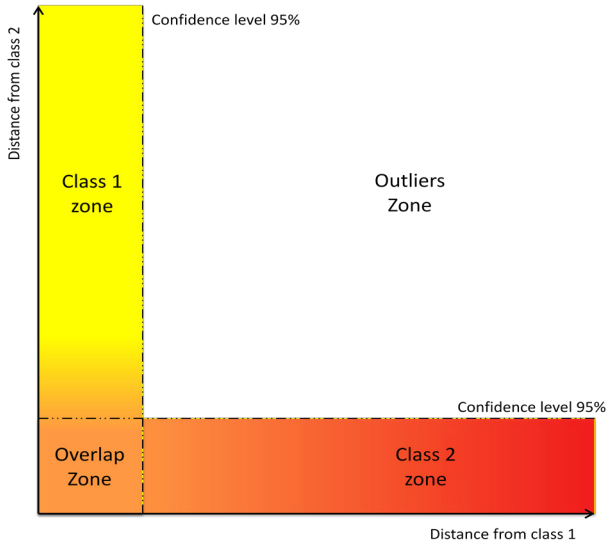


**Figure 2.3** Example of Coomans plot

So, four zones are defined on the plot: class 1, class 2, overlap of classes 1 and 2, and outlier zone (far from both classes). By plotting objects in this plot it is easy to visualize how certain a classification is.[12a]

### 2.3.4 K-Nearest neighbor (KNN)

Nearest neighbor methods are based on the determination of the distances between an unknown object and each of the objects of the training set. Usually, the Euclidean distance is used, but for strongly correlated variables, correlation-based measures are preferred. Then, the lowest distance is selected for the assignment of the class membership. In *k*NN, the *k*-nearest objects to the unknown sample are selected and a

majority rule is applied: the unknown is classified in the group to which the majority of the *k* objects belong. The choice of *k* is optimized by calculating the prediction ability with different *k* values. Small *k* values (3 or 5) are to be preferred frequently. A pre-processing of the data is required in order to avoid the effect of different scales of the variables. The technique can also be applied to the scores of the samples related to the principal components.

The method present several advantages: (i) its mathematical simplicity, which does not prevent it from achieving classification results as good as (or even better than) other more complex pattern recognition techniques; (ii) it is free from statistical assumptions, such as the normal distribution of the variables; and (iii) its effectiveness does not depend on the space distribution of the classes. On the other hand, this technique has similar limitations to LDA. Indeed *k*NN cannot work well if large differences are present in the number of samples in each class. An alternative criterion should be used then, instead of a simple majority criterion. For instance, another choice of criterion in *k*NN consists of weighing the importance as a neighbor of a known object to an unknown sample (inverse distance or inverse square distance). So, the nearest neighbors influence more the classification than the farthest ones. *k*NN provides poor information about the structure of the classes and of the relative importance of each variable in the classification. Furthermore, it does not allow a graphical representation of the results, and in the case of large number of samples, the computation can become excessively slow.[4a]

### 2.3.5 Artificial neural networks

Artificial neural networks (ANNs) are very sophisticated chemometric techniques capable of modeling extremely complex non-linear functions for classification and regression purposes. Although ANNs are suitable for dealing with issue related to food control and authentication, the diffusion of neural network for these aims is still at a relatively earlier stage of development.[19] Compared to other multivariate techniques, ANNs operate using a large number of parallel connected simple arithmetic units called neurons. Each neuron is a non-linear parameterized bounded function and the pattern of interconnection among them constitute the network architecture.[19] Construction of an artificial neural network occurs through the training process using a portion of the dataset and by means of opportunely designed training algorithms. The most frequently used ANNs are Kohonen-, counter propagation-, radial basis function- and probabilistic neural networks (PNN), but without doubt, the most popular, and widely used type of networks is the feed forward multilayer perceprons (MLP) trained with back-propagation (BP) algorithm in which the neurons operating on the same input variable are organized in layers. There are three kinds of layers in ANN: input layer, one or more hidden layers and output layer (Figure 2.4). The neurons are interconnected in a feed-forward way i.e. the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes.

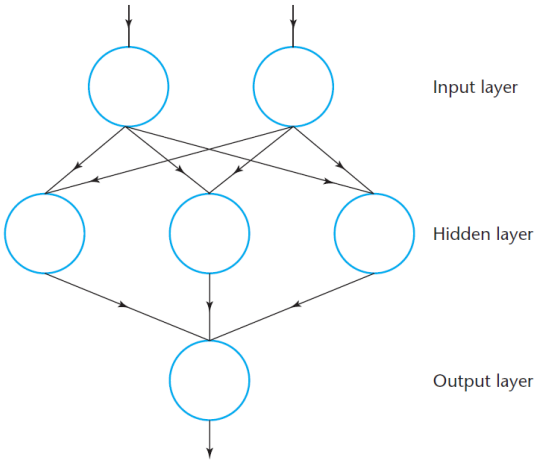


**Figure 2.4** Example of a neural network structure.

The number of neurons and hidden layers are some of the parameters that determine the predictive efficiency and reliability of network. The neurons are sorted in an input layer, containing one neuron for each independent variable (***X***); one or more hidden layers, where the data are processed; and an output layer, with one neuron for each dependent variable (***Y***). So, the data from the input layer are propagated through the network via synapses, which are associated with coefficients of connectivity called *weights* ($w$).[20] The net input ($a_i$) is computed as the inner product of the input variables ($\boldsymbol{x}_i$) impinging on the neuron and their weights ($w_i$). Once the node calculates this product, the result ($a_i$) is passed to a differentiable non-linear transfer (activation) function ($f(a_i)$), which transforms the weighted sum of all variables impinging onto the neuron in order to get the output value ($\boldsymbol{y}$). Several transfer functions have been used, e.g. variant logistic functions, hyperbolic tangent or sigmoid function.[21] The learning process identifies the weights that produce the best fit of the predicted outputs over the entire training data set. The hidden layers are particularly important to cope with non-linear classification problems.[20, 22]

Initialization of a network involves assigning random initial values to the weights (and thresholds) of all connections between neurons. The correction of the weights happens

iteratively.[23] In the construction of a neural network, particular attention has to be paid to the selection of the architecture, the selection of the learning parameters and the network validation. During the training step, the values for all the parameters involved in the learning process, the error between the net predicted output and the correct output are calculated. The size and number of the hidden layers and the number of epochs (training cycles) are evaluated by testing different values (trial and error) and checking the accuracy of the resulting prediction. The optimal number of epochs for an ANN is achieved when the error on the test set reaches a minimum. The number of hidden nodes is critical for the design of the network, because if too many hidden nodes are used, the network will overfit or memorize the training set data (noise). Conversely, if few hidden nodes are used, the network will fail to generalize and become unstable. Generally, one hidden layer is sufficient to approximate continuous functions, whereas two hidden layers may be necessary for learning functions with discontinuities.[22] An approach to determine the best number of hidden nodes is to start with the simplest architecture, i.e. one hidden layer, and to add nodes one at a time, until the network has learned the training set.[23] But with increasing number of hidden nodes, training becomes excessively time-consuming, so several rules of thumb are available in the literature, which relate hidden layer size to the number of nodes in input and output layers.[22] The quality of the ANN architecture and the best values of parameters involved in the learning process are evaluated using the root mean square error between the expected and the actual value of the output.[22,23]

Scaling of the data is essential to prevent larger numbers to override smaller ones and to prevent premature saturation of hidden nodes, which impedes the learning process. Balancing of data for preventing the net from being biased to the over-represented classes is also important. There are different approaches to stop the training of the network such

as the training error, gradient of error or cross-validation. Training proceeds until the error function reduces to a desired minimum. In classification, where the output is discrete values containing class membership information, the convergence criterion should be based on the hit (or miss) rate representing the percentage of examples classified correctly (or incorrectly).

A great advantage of ANNs is that causal knowledge of the relationship between the input and the output variables is not required. Instead, they learn these relationships through successive trainings. Moreover, ANNs present remarkable and attractive information processing characteristics: (i) non-linearity, allowing better fit to the data; (ii) noise insensitivity, providing accurate prediction in the presence of uncertain data and measurement errors; (iii) high parallelism, which implies fast processing and hardware failure tolerance; (iv) learning and adaptability, allowing the system to update (modify) its internal structure in response to changing environment; and (v) generalization, enabling application of the model to unlearned data.[22,24]

## References


1. Brereton, R. G., *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*. Wiley: Chichester, 2003.

2. Siebert, K. J., Chemometrics in brewing - A review. *J. Am. Soc. Brew. Chem.* **2001,** *59* (4), 147-156.

3. Moller, S. F.; von Frese, J.; Bro, R., Robust methods for multivariate data analysis. *J. Chemom.* **2005,** *19* (10), 549-563.

4. (a) Berrueta, L.; Alonsosalces, R.; Heberger, K., Supervised pattern recognition in food analysis. *Journal of Chromatography A* **2007,** *1158* (1-2), 196-214; (b) Esbensen, K. H., *The Introducer Package: Multivariate Data Analysis in Practice*. Camo: Oslo 2006; (c) Johnson, G. W.; Ehrlich, R., State of the Art Report on Multivariate Chemometric Methods in Environmental Forensics. *Environmental Forensics* **2002,** *3* (1), 59-79.

5. (a) Derde, M. P.; Coomans, D.; Massart, D. L., Effect of scaling on class modeling with the simca method. *Anal. Chim. Acta* **1982,** *141*; (b) Derde, M. P.; Massart, D. L., UNEQ: a disjoint modelling technique for pattern recognition based on normal distribution *Anal. Chim. Acta* **1986,** *184*.

6. (a) Wold, S.; Esbensen, K.; Geladi, P., Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987,** *2*; (b) Vandeginste, B. G. M.; Massart, D. L.; Buydens, L. M. C.; Jong, S. D. E.; Lewi, P. J.; Smeyers-Verbeke, J., Handbook of Chemometrics and Qualimetrics: Part B. Elsevier: Amsterdam, 1998; pp 88-149.

7. Kemsley, E. K., Discriminant analysis of high-dimensional data: a comparison of principal components analysis and partial least squares data reduction methods. *Chemom. Intell. Lab. Syst.* **1996,** *33* (1), 47-61.

8. (a) Seasholtz, M. B.; Kowalski, B., The parsimony principle applied to multivariate calibration. *Anal. Chim. Acta* **1993,** *277* (2), 165-177; (b) Onno E, d. N., The influence of data preprocessing on the robustness and parsimony of multivariate calibration models. *Chemom. Intell. Lab. Syst.* **1994,** *23* (1), 65-70; (c) Livingstone, D. J.; Manallack, D. T., STATISTICS USING NEURAL NETWORKS - CHANCE EFFECTS. *J. Med. Chem.* **1993,** *36* (9), 1295-1297.

9. Miller, J. C.; Miller, J. N., *Statistics and Chemometrics for Analytical Chemistry*. 5th Ed ed.; Pearson: 2005.

10. (a) Steyerberg, E. W.; Eijkemans, M. J. C.; van Houwelingen, J. C.; Lee, K. L.; Habbema, J. D. F., Prognostic models based on literature and individual patient data in logistic regression analysis. *Statistics in Medicine* **2000,** *19* (2), 141-160; (b)

Harrell, F. E.; Lee, K. L.; Mark, D. B., Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine* **1996,** *15* (4), 361-387.

11. (a) Lavine, B. K., Pattern recognition. *Crit. Rev. Anal. Chem.* **2006,** *36* (3-4), 153-161; (b) Kelly, S.; Heaton, K.; Hoogewerff, J., Tracing the geographical origin of food: The application of multi-element and multi-isotope analysis. *Trends Food Sci. Technol.* **2005,** *16* (12), 555-567; (c) Luykx, D. M. A. M.; Van Ruth, S. M., An overview of analytical methods for determining the geographical origin of food products. *Food Chem.* **2008,** *107* (2), 897-911.

12. (a) D.L. Massart, B. G. M. V., L.M.C. Buydens, S. De Jong, P.J.; Lewi, J. S.-V., *Handbook of Chemometrics and Qualimetrics*. Elsevier: Amsterdam, 1997; Vol. Part A; (b) Heberger, K.; Csomos, E.; Simon-Sarkadi, L., Principal component and linear discriminant analyses of free amino acids and biogenic amines in Hungarian wines. *J. Agric. Food Chem.* **2003,** *51* (27), 8055-8060.

13. Lavine, B. K., Encyclopedia of Analytical Chemistry. Wiley: Chichester, 2000; p 9689.

14. Alexandrakis, D.; Downey, G.; Scannell, A. G. M., Detection and identification of bacteria in an isolated system with near-infrared spectroscopy and multivariate analysis. *J. Agric. Food Chem.* **2008,** *56* (10), 3431-3437.

15. Forina, M.; Lanteri, S.; Casale, A., Multivariate calibration. *Journal of Chromatography A* **2007,** *1158* (1-2), 61-93.

16. Liu, F.; He, Y.; Wang, L., Determination of effective wavelengths for discrimination of fruit vinegars using near infrared spectroscopy and multivariate analysis. *Anal. Chim. Acta* **2008,** *615* (1), 10-17.

17. Galtier, O.; Abbas, O.; Le Dréau, Y.; Rebufa, C.; Kister, J.; Artaud, J.; Dupuy, N., Comparison of PLS1-DA, PLS2-DA and SIMCA for classification by origin of crude petroleum oils by MIR and virgin olive oils by NIR for different spectral regions. *Vib. Spectrosc.* **2011,** *55* (1), 132-140.

18. Lo Feudo, G.; Naccarato, A.; Sindona, G.; Tagarelli, A., Investigating the Origin of Tomatoes and Triple Concentrated Tomato Pastes through Multielement Determination by Inductively Coupled Plasma Mass Spectrometry and Statistical Analysis. *J. Agric. Food Chem.* **2010,** *58* (6), 3801-3807.

19. Marini, F., Artificial neural networks in foodstuff analyses: Trends and perspectives A review. *Anal. Chim. Acta* **2009,** *635* (2), 121-131.

20. Gonçalves, E. C.; Minim, L. A.; Coimbra, J. S. R.; Minim, V. P. R., Modeling sterilization process of canned foods using artificial neural networks. *Chemical Engineering and Processing: Process Intensification* **2005,** *44* (12), 1269-1276.

21. Pazourek, J.; Gajdošová, D.; Spanilá, M.; Farková, M.; Novotná, K.; Havel, J., Analysis of polyphenols in wines: Correlation between total polyphenolic content and antioxidant potential from photometric measurements: Prediction of cultivars and vintage from capillary zone electrophoresis fingerprints using artificial neural network. *Journal of Chromatography A* **2005,** *1081* (1), 48-54.

22. Basheer, I. A.; Hajmeer, M., Artificial neural networks: fundamentals, computing, design, and application. *J. Microbiol. Methods* **2000,** *43* (1), 3-31.

23. Barile, D.; Coı¨sson, J. D.; Arlorio, M.; Rinaldi, M., Identification of production area of Ossolano Italian cheese with chemometric complex approach. *Food Control* **2006,** *17* (3), 197-206.

24. Jain, A.; Chaurasia, A.; Sahasrabuddhey, B.; Verma, K. K., Determination of bromide in complex matrices by pre-column derivatization linked to solid-phase extraction and high-performance liquid chromatography. *Journal of Chromatography A* **1996,** *746* (1), 31-41.

# Chapter 3

## *Results and discussion*

### *Olive Oil*

*Secondary metabolites of Olea europaea leaves as markers for the discrimination of cultivars and cultivation zones by multivariate analysis.*

### *Tomato*

*Investigating the Origin of Tomatoes and Triple Concentrated Tomato Pastes through Multielement Determination by Inductively Coupled Plasma Mass Spectrometry and Statistical Analysis*

*The volatile fraction profiling of fresh tomatoes and triple concentrate tomato pastes as parameter for the determination of geographical origin*

### *Tropea red Onion*

*Multielement Fingerprinting as a Tool in Origin Authentication of PGI Food Products: Tropea Red Onion*

### *Clementine di Calabria*

*Authenticity of PGI Clementine of Calabria by multielement fingerprint*

## 3.1 Secondary metabolites of Olea europaea leaves as markers for the discrimination of cultivars and cultivation zones by multivariate analysis

Olive (Olea europaea) is one of the most ancient cultivated fruit tree species in the Mediterranean basin. Olive trees show multiple phenotypic expressions, generally designated as varieties or cultivars. Some cultivars are specific of a particular country whereas others are present in more than one cultivation area.

The correct classification of the variety and cultivation zone of the olive oil, which also includes the tree and its different parts, is shown as a new problem in order to control the quality and the appellation of origin of the olive oils due to the fact that each combination of variety and cultivation zone has a different chemical composition.[1] Indeed distribution of macro- and microcomponents in olive oils varies, among others, with the cultivars.[2] Traditionally, morphological and phonological traits are used to identify olive cultivars.[3] This method, however, presents some limitations because of environmental influence. Accordingly, this works aimed at discriminating different varieties of olive trees cultivated in the same geographical area using as markers a set of biophenol contents in olive leaves determined by an HPLC-MS/MS approach. Moreover, the possibility of differentiation between samples of Carolea cultivated in different geographical zones was take into account.

Samples of olive leaves, utilized in this study, were collected from five different cultivars (Carolea, Cassanese, Coratina, Nocellara del Belice and Leccino), in the experimental field of the C.R.A. Istituto Sperimentale per l'Olivicoltura in Rende (Calabria, Italy). Other Carolea leaves were picked up from fields situated in Mirto (Calabria) and Spoleto (Umbria, Italy). Samples were treated according to procedure proposed by Sindona et al..[4]

Twelve phenolic compounds (Table 3.1) that originate from the secondary metabolism of the plant, through mevalonic acid, phenylpropanoid and catecol methoxylation pathways were selected as possible traceability markers. The data of each sample have been reported as concentrations (mg/kg) referred to salicin (internal standard) of the selected 12 phenolic compounds for the subsequent chemometric treatment.

| Ion $[M+NH_4]^+$ | Compound | Retention Time (min) |
|---|---|---|
| m/z 304 | Salicin (Internal Standard) | 8.51 |
| m/z 364a | 2-methoxyhydroxytyrosol glucoside | 5.57 |
| m/z 334a | Hydroxytyrosol glucoside | 5.76 |
| m/z 348a | Methoxytyrosol glucoside | 7.34 |
| m/z 408 | Oleoside | 12.22 |
| m/z 422 | Oleoside 11-methyl ester | 14.05 |
| m/z 642 | Verbascoside | 17.98 |
| m/z 720 | Angustifolioside A | 18.68 |
| m/z 720 | Angustifolioside B | 18.93 |
| m/z 562a | Saturated oleuropein | 19.00 |
| m/z 586a | Dimethyl oleuropein | 21.28 |
| m/z 558 | Oleuropein | 21.75 |
| m/z 542 | Ligstroside | 24.03 |

**Table 3.1** Selected 12 phenolic compounds ( [a] New phenolic compounds detected in the leaves of *Olea Europaea* :Di Donna, Mazzotti, Salerno, Tagarelli, Taverna, & Sindona, 2007)

In the first instance the data of olive leaves samples harvested in March and April in Rende were subjected to principal component analysis (PCA). The scores of samples and loadings of the variables on the two first principal components are plotted in Figure 3.1: the information retained is 52.65% of the total variance.

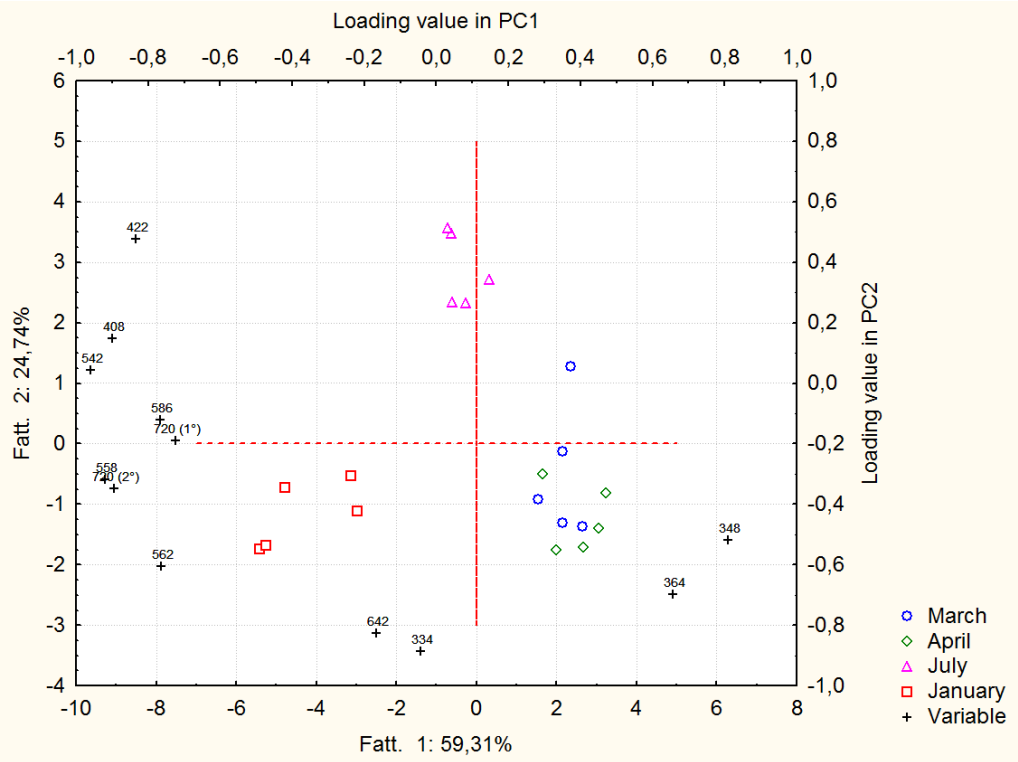


**Figure 3.1** Biplot of principal component scores and loadings for leaves of the 5 varieties of olive tree harvested in March and April (Variables are indicated as m/z values, see Table 1)

The plot shows differentiation between harvesting period regardless of cultivar. In samples corresponding to the Leccino, Coratina and, to a smaller extent, Cassanese it was observed a shifting along PC1 going from March to April which indicates the decrease of concentration of compounds angustifolioside A, angustifolioside B, oleuropein, satured oleuropein and dimethyl oleuropein in leaves harvested in April. On the contrary, tendency shown by Carolea variety was totally different from all the others. Actually, for this cultivar a shift along PC2 only is clear that means a decrease of concentration of oleoside, oleoside 11-methyl ester and ligstroside and an increase of hydroxytyrosol glycoside and methoxytyrosol glycoside going from March to April. For Nocellara samples it is not observed meaningful variations, even though it can be noted a slight shift to more positive values of PC1. The supervised pattern recognition techniques (LDA,

SIMCA and KNN) were used in order to obtain classification rules for distinguishing between five varieties of olive trees (Carolea, Cassanese, Coratina, Nocellara del Belice and Leccino) cultivated in the same geographical area (Rende, Calabria, Italy) and between samples of Carolea cultivated in different geographical zones. These techniques were applied to concentration values (mg/kg) referred to salicin of the selected twelve phenolic compounds, as in PCA analysis. The reliability of the classification rules needs to be validated and therefore a 10-fold cross-validation procedure was performed. The goodness of the classification models was evaluated in terms of prediction ability, which is equal to the percentage of the test set members correctly classified by the rules developed in the training step.

***Cultivars discrimination***

In the first instance linear discriminant analysis has been applied to concentration values of the selected 12 phenolic compounds of samples harvested in March and April, using five groups corresponding to the five cultivar type, as input a priori.

The differentiation between groups is significant since the low Wilks λ value (0.0011656) shows that the model is discriminating. Moreover, the information from data treatment is characterized by a high degree of reliability since the *p* level is extremely low (<0.00001).

To check the usefulness of the method for prediction purposes cross validation was performed. The proposed model predict 82% of correct variety and this suggests that the presented method may be a potential choice for checking cultivar type (Table 3.2).

| | Cassanese | Coratina | Carolea | Leccino | Nocellara |
|---|---|---|---|---|---|
| **Cassanese** | 9 | 0 | 0 | 0 | 1 |
| **Coratina** | 0 | 7 | 0 | 0 | 3 |
| **Carolea** | 1 | 0 | 9 | 0 | 0 |
| **Leccino** | 1 | 0 | 0 | 9 | 0 |
| **Nocellara** | 1 | 2 | 0 | 0 | 7 |

**Table 3.2** Prediction matrices for LDA model of olive trees varieties(rows represent the true class, columns report the assigned class)

The Analyses of the Discriminating Functions show that four new phenolic compounds [2-methoxyhydroxytyrosol glucoside (364), dimethyl oleuropein (586), hydroxytyrosol glucoside (334) and methoxytyrosol glucoside (348)] are the most important variables in the differentiation of cultivars. This result underline the important role of the HPLC-ESI-MS/MS method that allowed the identification of the minor phenolic compounds.

SIMCA was also applied to the same data matrix subjected to LDA and models obtained were based on two components for each category (explaining about 70% of class variance for olive trees varieties and about 80% of variance for cultivar zone), normal range and 5% as the significance level for critical distance. To study the predictive capability of SIMCA, the same cross-validation procedure was applied. The proposed models predict correctly 88% and 87% of the varieties and cultivation zones, respectively (Table 3.3).

| | Cassanese | Coratina | Carolea | Leccino | Nocellara | SENS(%) | SPEC(%) |
|---|---|---|---|---|---|---|---|
| **Cassanese** | 6 | 1 | 0 | 1 | 2 | 100 | 100 |
| **Coratina** | 0 | 10 | 0 | 0 | 0 | 100 | 100 |
| **Carolea** | 0 | 0 | 8 | 0 | 2 | 90 | 100 |
| **Leccino** | 0 | 0 | 0 | 10 | 0 | 100 | 100 |
| **Nocellara** | 0 | 0 | 0 | 0 | 10 | 100 | 95 |

**Table 3.3** Prediction matrices and SENS and SPEC values of SIMCA classification for olive trees varieties (rows represent the true class, columns report the assigned class)

In particular, the SIMCA model shows higher predictive capability for Coratina, Leccino and Nocellara varieties respect to classification obtained with Linear Discriminant Analysis. On the other hand, SIMCA shows a worse performance for Cassanese and Carolea. Also in this case, the discriminant powers of variables show that three minor phenolic compounds [2-methoxyhydroxytyrosol glucoside (364), hydroxytyrosol glucoside (334) and methoxytyrosol glucoside (348)] are among compounds that mostly contribute to discrimination of cultivars. Sensitivity and specificity values obtained were very satisfactory. Indeed SIMCA model presented a mean sensitivity of 98% and a mean specificity of 99%.

KNN was also applied to the same data sets subjected to LDA and SIMCA using the square inverse of the Euclidean distance and the decision criteria of two nearest neighbors and a majority vote. The number of neighbors was chosen after study of the success in classification with K values between 1 and 8. The best results were obtained with K = 1 (Table 3.4). The model is capable of achieving poor percentage for prediction (54.0%), especially for Nocellara (20%) and only for Leccino KNN afforded acceptable results (90%).

| | **Cassanese** | **Coratina** | **Carolea** | **Leccino** | **Nocellara** |
|---|---|---|---|---|---|
| **Cassanese** | 5 | 0 | 2 | 1 | 2 |
| **Coratina** | 1 | 6 | 2 | 1 | 0 |
| **Carolea** | 1 | 0 | 5 | 2 | 2 |
| **Leccino** | 0 | 0 | 1 | 9 | 0 |
| **Nocellara** | 3 | 1 | 2 | 2 | 2 |

**Table 3.4** Prediction matrices for KNN model for olive trees varieties (rows represent the true class, columns report the assigned class)

In conclusion, among the pattern recognition techniques for cultivar discrimination, LDA and SIMCA have shown successful results whereas KNN allowed to obtain unsuccessful results.

**Cultivation zones discrimination**

Linear discriminant analysis requires the data matrices for each category to have an high ratio between the number of training samples and the number of variables used in order to obtain stable chemometric model. Then attempt to distinguish between samples of Carolea cultivated in different geographical zones (5 objects for each zone, Rende, Mirto, Spoleto) could not be performed by considering all the variables. Accordingly, PCA have been applied to the concentration values of all phenolic compounds and coordinates of scores of the first four principal components (87.33% of the total available information) were submitted to LD analysis, using three groups corresponding to the three geographical zones, as input a priori. In this case, the high value (21.39) of the F (8.18) parameter indicates a significant difference among the means of the groups, whereas the information from data treatment is characterized by a high degree of reliability since the p level is extremely low (<0.00001). Moreover, the differentiation between groups is significant since the low Wilks λ value (0.009057) shows that the model is discriminating. When LDA was applied to the data set described above, model produced good percentages of correct prediction (93%) in cross-validation procedure (Table 3.5).

| | Rende | Mirto | Spoleto |
|---|---|---|---|
| **Rende** | 5 | 0 | 0 |
| **Mirto** | 0 | 5 | 0 |
| **Spoleto** | 1 | 0 | 4 |

**Table 3.5** Prediction matrices for LDA model of olive trees cultivation zones (rows represent the true class, columns report the assigned class)

As for cultivar discrimination, SIMCA was also applied to the same data matrix subjected to LDA without variables reduction step. The differentiation between samples of Carolea cultivated in different geographical zones can be achieved with good results for Mirto and Spoleto whereas percentage of prediction is only 60% for Rende. SIMCA model developed presented a SENS mean value of 93% and SPEC mean value of 100% (Table 3.6).

| | Rende | Mirto | Spoleto | SENS(%) | SPEC(%) |
|---|---|---|---|---|---|
| **Rende** | 3 | 2 | 0 | 100 | 100 |
| **Mirto** | 0 | 5 | 0 | 100 | 100 |
| **Spoleto** | 0 | 0 | 5 | 80 | 100 |

**Table 3.6** Prediction matrices and SENS and SPEC values of SIMCA classification for olive trees cultivation zones (rows represent the true class, columns report the assigned class)

KNN was also applied to the same data sets subjected to LDA and SIMCA using the square inverse of the Euclidean distance and the decision criteria of two nearest neighbors and a majority vote. As for cultivar, the number of neighbors was chosen after study of the success in classification with K values between 1 and 8. The best results were obtained with K = 1 (Table 3.7).

| | **Rende** | **Mirto** | **Spoleto** |
|---|---|---|---|
| **Rende** | 5 | 0 | 0 |
| **Mirto** | 0 | 4 | 1 |
| **Spoleto** | 0 | 2 | 3 |

**Table 3.7** Prediction matrices for KNN model for olive trees varieties (rows represent the true class, columns report the assigned class)

The model allowed to obtain 80% of mean prediction ability. The model be considered good only for categories Rende and Mirto. Also in this case, as for cultivar discrimination, better results were obtained using LDA and SIMCA approaches than using KNN. However this latter chemometric technique showed better results than those obtained with cultivar discrimination.

## 3.2 Traceability of Tomatoes and Triple Concentrated Tomato Pastes

Tomatoes (Lycopersicum esculentum) is a well known herbaceous plant belong to Solanaceae family and worldwide used in the diet. Their intake is referred to fresh fruits and also to semifinished products as sauces, purées and pastes (mono, double and triple concentrated). Tomato is certainly among the most important foods of the Mediterranean diet and has important healing effects associated with the presence of polyphenols, antioxidants such as carotenoids, and vitamins.

One of the principal adulteration of tomato products consists in the dilution of tomato pastes with water instead of using fresh fruits alone. Another aspect is represented by the importation of triple concentrated pastes from countries other than those of the producers. The globalization of food markets and, consequently, the easiness in the circulation of foodstuffs has caused the loss of identity of the origin of foods which may expose consumers to risks derived by the manufacturing processes. The imported products come often from countries where quality and safety rules are less stringent than those adopted in those countries where EMEA and FDA directives are followed.

In 2007, 160 million kilograms of tomato pastes, corresponding to about 25% of the Italian tomato production, have in fact been imported from abroad and the China is one of the principal exporter.[5] The introduction of foodstuff coming from foreign country may represent an unfair competition for producers that can affect the regional economy and even national economy. In order to protect both the consumers and the national production, a decree was issued by Italy in 2006 that established the obligation from 1 January 2008 to state clearly in the label the origin (Region or State) of fresh tomato used in the preparation of sauces.[6] The number of critical points in the production chain of tomato derivatives calls for the introduction of efficient analytical methods to reliably

check the origin of the raw materials. In recent years, there has been a growing interest about geographical characterization and authenticity of tomato but few traceability study are known for tomato juice and tomato paste.[7] An investigations based on a DNA extraction procedure has been published.[8] Two studies deal with the determination of geographical origin of cherry tomatoes[9] and triple concentrated tomato pastes[10] by 1H NMR spectroscopy.

### 3.2.1 Investigating the Origin of Tomatoes and Triple Concentrated Tomato Pastes through Multielement Determination by Inductively Coupled Plasma Mass Spectrometry and Statistical Analysis

Multielement analysis has been applied to a range of foodstuff to develop methods that allow the identification of their geographical origins.[11] Trace element concentration from tomatoes subjected to a proper chemometric approach was applied to distinguish between conventional and organic cultivation.[12]

The aim of this study was to develop a simple and rapid method, such as ICP-MS analysis, followed by a convenient statistical processing of multielement profile to discriminate between tomato samples cultivated in different areas and between tomato paste samples coming from different countries.

This research was the continuation and completion of my M.Sc. thesis work. In the first part of this work it has been shown that the multielement distributions can be a parameter for the characterization of the geographical origin of tomato and triple concentrated tomato pastes samples. In my PhD work this investigation was extended to samples belonging to the harvest season 2008 in order to confirm the goodness and robustness of the statistical model. In this way we also take into account possible climatic effects.

The tomatoes samples come from four Italian production regions, namely, Calabria (Crotone and Rocca di Neto), Basilicata (Matera) and Emilia Romagna (Collecchio) (Figure 3.2). The Italian tomato concentrate samples come from Collecchio and Rocca di Neto. Moreover, in comparison with the thesis work, samples coming from China, California and Greece were take into account in order to show the possibility to differentiate between tomato paste samples made in Italy with those produced in other countries.


**Figure 3.2** Italy map (Collecchio, Δ; Crotone, □; Matera, ○; Mesagne, ◊)

In the first part of the work the effectiveness of the digestion procedure was assessed in terms of accuracy for the 35 elements certified in reference material NCS ZC85006 tomato (Be, Na, Mg, Al, K, Ca, Sc, V, Mn, Fe, Ni, Co, Cu, Zn, As, Se, Rb, Sr, Y, Cd, Cs, Ba, La, Ce, Nd, Sm, Eu, Gd, Dy, Er, Yb, Lu, Pb, Th, U)

About 0.5 g of the certified reference material was subjected to digestion treatment using a mixture of 4.5mL of $HNO_3$/1mL of $H_2O_2$ and 0.5 mL of each of the following third components: HF, HCl, $H_3BO_3$ and $HClO_4$. The extracts were diluted to 50 mL with ultrapure water and subjected to ICP-MS analysis. The $HNO_3/H_2O_2$/HF mixture provided the best digestion results. Accuracy values were considered acceptable for 25 elements (values in the range 75-120%). Analytical parameters for these elements were shown in Table 3.8.

| | Isotope | Calibration range (µg/l) | $R^2$ | LOD (µg/kg)[a] | LOQ (µg/kg)[a] | Certified value (mg/kg)[b] | Found value (mg/kg)[c] | Accuracy (%) |
|---|---|---|---|---|---|---|---|---|
| **Al** | 27 | 0.1-2000 | 0.9998 | 159 | 346 | 0.295±0.043 | 0.279±0.008 | 94.5 |
| **As** | 75 | 0.1-2000 | 0.9999 | 0.42 | 0.55 | 1.05±0.13 | 1.03±0.02 | 98.6 |
| **Ba** | 138 | 0.1-2000 | 0.9999 | 0.68 | 0.98 | 55.2±5.2 | 52.9±0.8 | 95.9 |
| **Be** | 9 | 0.1-2000 | 0.9999 | 0.031 | 0.071 | (0.084) | 0.067±0.009 | 80.2 |
| **Ca** | 44 | 0.1-2000 | 0.9982 | 1441 | 1968 | 5.31 (%)±0.19 | 4.58±0.06 | 86.3 |
| **Cd** | 114 | 0.1-2000 | 0.9999 | 0.13 | 0.31 | 0.82±0.09 | 0.685±0.004 | 83.6 |
| **Ce** | 140 | 0.01-150 | 0.9999 | 0.063 | 0.084 | 3.08±0.22 | 2.94±0.26 | 95.5 |
| **Cu** | 63 | 0.1-2000 | 0.9997 | 0.76 | 0.95 | 21.1±2.5 | 23.5±0.5 | 111.5 |
| **Dy** | 164 | 0.01-150 | 0.9999 | 0.007 | 0.014 | (0.23) | 0.217±0.010 | 94.3 |
| **Fe**[d] | 56 | 0.1-2000 | 0.9999 | 142 | 179 | 0.138(%)±0.015 | 0.139±0.003 | 100.9 |
| **K** | 39 | 0.1-2000 | 0.9991 | 92.2 | 127.2 | 0.579(%)±0.052 | 0.564±0.010 | 97.4 |
| **La** | 139 | 0.01-150 | 0.9999 | 0.031 | 0.047 | 1.78±0.17 | 1.34±0.15 | 75.4 |
| **Lu** | 175 | 0.01-150 | 0.9999 | 0.001 | 0.002 | (0.019) | 0.019±0.002 | 102.2 |
| **Mg** | 24 | 0.1-2000 | 0.9998 | 50.7 | 105 | 0.736(%)±0.057 | 0.645±0.006 | 87.6 |
| **Mn** | 55 | 0.1-2000 | 0.9997 | 2.81 | 4.22 | 87.1±5.6 | 104.1±1.1 | 119.6 |
| **Na** | 23 | 0.1-2000 | 0.9997 | 304 | 722 | (0.13) | 0.126±0.003 | 96.8 |
| **Nd** | 142 | 0.01-150 | 0.9999 | 0.015 | 0.038 | (1.28) | 1.36±0.13 | 106.0 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Pb** | 208 | 0.1-2000 | 0.9999 | 0.24 | 0.33 | 4.97±0.54 | 5.49±0.03 | 110.5 |
| **Rb** | 85 | 0.1-2000 | 0.9999 | 1.18 | 2.89 | 6.66±0.47 | 6.98±0.01 | 104.8 |
| **Sm** | 152 | 0.01-150 | 0.9999 | 0.013 | 0.030 | 0.270±0.020 | 0.299±0.019 | 110.8 |
| **Sr** | 88 | 0.1-2000 | 0.9999 | 2.02 | 3.15 | 569±40 | 610±8 | 107.2 |
| **Th** | 232 | 0.01-150 | 0.9992 | 0.020 | 0.042 | (0.486) | 0.488±0.076 | 100.5 |
| **U** | 238 | 0.1-2000 | 0.9999 | 0.053 | 0.086 | (0.202) | 0.185±0.005 | 91.4 |
| **V** | 51 | 0.1-2000 | 0.9999 | 1.74 | 2.65 | 3.84±0.30 | 3.83±0.14 | 99.7 |
| **Zn** | 64 | 0.1-2000 | 0.9999 | 81.8 | 188 | 36.2±3.1 | 33.8±0.56 | 93.3 |

**Table 3.8** Summary of calibration parameters, limits of detection (LODs), limits of quantitation (LOQs) and mean accuracies (%) referred to certified reference material NCS ZC85006 Tomato. ([a] LOD and LOQ values are referred to analysis of tomato samples; [b] Certified values without standard deviation are reported in parenthesis; [c] ±SD, n=3; [d] Analyzed in DRC mode).

In order to develop the analytical method 46 elements were initially investigated: $^{7}Li$, $^{9}Be$, $^{23}Na$, $^{24}Mg$, $^{27}Al$, $^{39}K$, $^{44}Ca$, $^{45}Sc$, $^{51}V$, $^{52}Cr$, $^{53}Cr$, $^{55}Mn$, $^{57}Fe$, $^{58}Ni$, $^{60}Ni$, $^{59}Co$, $^{63}Cu$, $^{64}Zn$, $^{66}Zn$, $^{69}Ga$, $^{75}As$, $^{82}Se$, $^{85}Rb$, $^{88}Sr$, $^{89}Y$, $^{107}Ag$, $^{114}Cd$, $^{115}In$, $^{133}Cs$, $^{138}Ba$, $^{139}La$, $^{140}Ce$, $^{141}Pr$, $^{142}Nd$, $^{152}Sm$, $^{153}Eu$, $^{158}Gd$, $^{159}Tb$, $^{164}Dy$, $^{165}Ho$, $^{166}Er$, $^{169}Tm$, $^{174}Yb$, $^{175}Lu$, $^{205}Tl$, $^{208}Pb$, $^{209}Bi$, $^{232}Th$ and $^{238}U$. The determination of some elements by ICP-MS is known to suffer from polyatomic isobaric interferences. For example, regarding the elements of our interest, the scandium signal at m/z 45 is affected by $^{13}C^{16}O_2^+$ and $^{29}Si^{16}O^+$, the chromium signal at m/z 53 is affected by $^{40}Ar^{13}C^+$ and $^{37}Cl^{16}O^+$, the iron signal at m/z 56 is affected by $^{40}Ca^{16}O^+$, the nickel signal at m/z 58 is affected by $^{42}Ca^{16}O^+$, the zinc signal at m/z 64 is affected by $^{48}Ti^{16}O^+$ and $^{32}S_2^+$, the selenium signal at m/z 80 is affected by $^{40}Ar^{40}Ar^+$ and, finally, the europium signal at m/z 153 is affected by $^{137}Ba^{16}O^+$. The dynamic reaction cell (DRC) is proved to be an effective method for relieving such isobaric interferences. Then Sc, Cr, Fe, Ni, Zn, Se and Eu have been monitored in both modes (Standard and DRC modes). Methane (99.996% purity) was used as reaction gas. The optimization of the $CH_4$ flow rate and the RPq value were carried out using matrix blank solution prepared with HCl (2%), $HNO_3$ (1%), $CH_3OH$ (1%), Ca at 20 mg/L, Si and Ba at

50 μg/L and a matrix blank solution spiked with Sc, Cr, Fe, Ni, Zn and Se at 1 μg/L and Eu at 0.1 μg/L. The best background equivalent concentrations (BEC) were obtained at flow rate value (mL/min) of 0.7 for Cr, Fe, Se and Zn and at value of 1.15 for Sc, Ni and Eu. The best S/N ratio was obtained with a RPq value of 0.6 for all elements except for Eu (0.5), Cr (0.7) and Fe (0.7). ICP-MS analyses were performed following the operating program and parameters reported in Chapter 4 (Table 4.2). Copper and potassium values were not submitted to statistical analysis although accuracy values obtained using the certified matrix were satisfactory because of their use in agricultural practice as fertilizers. The chemometric treatments have been applied to the data set containing the concentration of those 23 elements that have shown the best accuracy values, and to the concentration of 9 elements (Ag, Cr, Ga, Ho, Li, Pr, Tb, Tl and Tm), for which results were above the limit of detection (LOD).

In the assay of $^{54}$Fe and $^{52}$Cr polyatomic isobaric interferences had a considerable impact on their signal. Thus, among the 32 elements monitored $^{54}$Fe and $^{52}$Cr were detected in DRC mode.

***Statistical analysis on tomato samples***

Pattern recognition analysis was carried out using three supervised chemometric techniques (linear discriminant analysis, soft independent modeling of class analogy, and K-nearest neighbors) to obtain classification rules for distinguishing between four Italian cultivation areas of tomatoes (Calabria, Basilicata, Puglia, Emilia Romagna) and between triple concentrated tomato paste samples coming from Italy and from foreign countries (California, China, Greece). The data matrix constituted of the concentration values (μg/kg) of the selected 32 elements were submitted to the three chemometric approach.

The reliability of the classification rules was validated through a cross-validation procedure.

For tomato, since the number of samples (110 objects) is more than three times the number of elements (32 variables) then standard linear discriminant analysis (LDA) was performed by considering all the elements without applying any method for reduction of the variables and using four groups corresponding to the four cultivation zones, as input a priori. In order to test the reliability of the obtained model in term of prediction ability, a 5-fold cross validation was performed. In this way, for each cross validation cycle, the 80% of the components of sample set was used for calculating models, and the remaining 20% of the components of the sample set was considered as an unknown, and classified. The proposed model has correctly predicted all samples for each zone and this suggests that the presented method may be a potential choice for checking origin of tomatoes. This discrimination is visually represented by the bi-dimensional plot of the first two roots which shows the great separation among the four clusters that represent areas of origin (Figure 3.3).

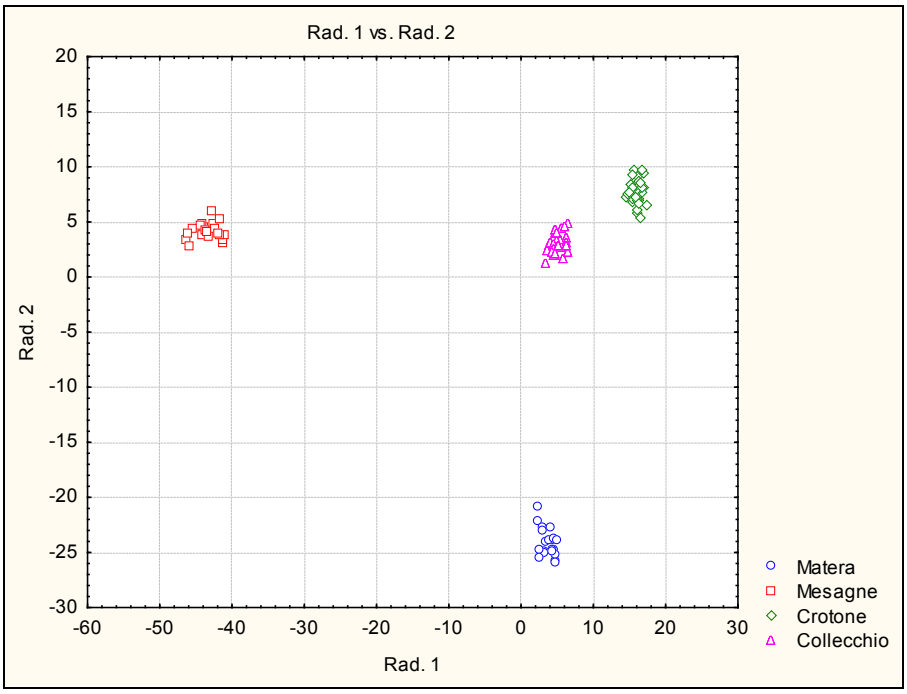


**Figure 3.3** LDA plot for tomato samples

This difference is statistically significant due to the very low Wilks λ value (<0.0000011) that shows how the model is highly discriminatory. Moreover the high value (325.1) of the F(96.22) ratio is indicative of a significant difference between the averages of groups. The low p-level (<0.00001) underline the very high probability of a correct classification. The analyses of the LDA discriminating model shows that Cd, alkaline metals and alkaline earth metals are the most important variables to determining the tomato geographical origin. The important role of alkaline metals and alkaline earth metals in differentiation of cultivation zones is in agreement with previous studies about the identification of the origin of virgin olive oil[13] carried out by the research group in which I have been worked toward Ph.D. This results are also in agreement with other two studies that I carried out during Ph.D on traceability of Tropea red Onion[14] and

Clementine di Calabria.[15] The results obtained suggests that the presented method may be a potential choice for checking origin of tomatoes.

Soft independent modeling of class analogy (SIMCA) was applied to the same data matrix used in standard LDA, and models obtained were based on normal range, 5% as the significance level for critical distance and a number of components for each category so that explains about 90% of class variance. Validation of the model was carried out by a 5-fold cross-validation procedure. The SIMCA model has provided very good results in terms of mean prediction ability classifying correctly all the samples submitted (Table 3.9). Moreover, the SIMCA model is satisfactory in terms of mean sensitivity (84.5%) and mean specificity (100%).

| | **Matera** | **Mesagne** | **Crotone** | **Collecchio** | **SENS(%)** | **SPEC(%)** |
|---|---|---|---|---|---|---|
| **Matera** | 20 | 0 | 0 | 0 | 80 | 100 |
| **Mesagne** | 0 | 20 | 0 | 0 | 80 | 100 |
| **Crotone** | 0 | 0 | 40 | 0 | 85 | 100 |
| **Collecchio** | 0 | 0 | 0 | 30 | 90 | 100 |

**Table 3.9** Prediction matrix and SENS and SPEC values of SIMCA classification (rows represent the true class, columns report the assigned class)

As for LDA model, the SIMCA discriminant powers of variables show that alkaline metals and alkaline earth metals are among the compounds that mostly contribute to discrimination of cultivation zones.

K-nearest neighbors (KNN) was applied to the same data set subjected to LDA and SIMCA using the square inverse of the Euclidean distance and the decision criterion of majority vote. The number of neighbors was chosen after evaluation of the success in classification with K values between 1 and 12. The best results (Table 3.10) were obtained with K = 3. The differentiation between geographical zones can be considered

acceptable for all classes. The mean prediction ability obtained is 88.2%. These results are worse than those obtained by LDA and SIMCA techniques.

| | Matera | Mesagne | Crotone | Collecchio |
|---|---|---|---|---|
| **Matera** | 17 | 0 | 3 | 0 |
| **Mesagne** | 0 | 20 | 0 | 0 |
| **Crotone** | 1 | 0 | 35 | 4 |
| **Collecchio** | 0 | 0 | 5 | 25 |

**Table 3.10** Prediction matrix for KNN model (rows represent the true class, columns report the assigned class)

***Statistical analysis on triple concentrated tomato paste samples***

Linear discriminant analysis, soft independent modeling of class analogy, and K-nearest neighbors have been applied to the concentration of the same 32 elements considered in the tomato samples. Also in this case, since the total number of objects (100 samples) is more than three times the number of variables standard LDA has been carried out using two categories associated to samples coming from Italy (Italy group) and samples from foreign countries (non-Italy group), as input a priori. As for tomato samples the statistical results obtained were very satisfactory: Wilks $\lambda = 0.00975$, $F(32.67) = 212.6$, p-level < 0.00001 and prediction ability of 98%for each class (Table 3.11). Lithium and rubidium (alkaline metals) are the most important variables in the distinction of geographical origin. Contrary to tomato samples the contribution of alkaline earth metals seems not important for differentiation of triple concentrated paste tomato samples.

| LDA | | | | |
|---|---|---|---|---|
| | **Italy** | **non-Italy** | | |
| **Italy** | 49 | 1 | | |
| **non-Italy** | 1 | 49 | | |
| **SIMCA** | | | | |
| | **Italy** | **non-Italy** | **SENS(%)** | **SPEC(%)** |
| **Italy** | 48 | 2 | 90 | 100 |
| **non-Italy** | 0 | 50 | 80 | 100 |
| **KNN** | | | | |
| | **Italy** | **non-Italy** | | |
| **Italy** | 50 | 0 | | |
| **non-Italy** | 0 | 50 | | |

**Table 3.11** Prediction matrices for supervised pattern recognition techniques (rows represent the true class, columns report the assigned class)

SIMCA was applied to the same data matrix used in standard LDA, and models obtained were based on five PCs for category "Italy" and four components for category "non-Italy that explain about 90% of variance for both classes. This chemometric model has provided very good results in terms of mean prediction ability (96%for Italy, 100%for non-Italy) (Table 3.11). Moreover, the SIMCA model is satisfactory in terms of mean sensitivity (85%) and mean specificity (100%).

KNN was applied using the same criteria chosen for tomato samples. The number of neighbors was evaluated, and the best results were obtained with K = 5 (Table 3.11). In this case, KNN has provided better results than those obtained by the LDA and SIMCA methods. Indeed the KNN method is capable to classify correctly all samples for both categories.

### 3.2.2 The volatile fraction profiling of fresh tomatoes and triple concentrate tomato pastes as parameter for the determination of geographical origin

Beside the multielement profile, another possible approach to establish the geographical origin of fresh tomatoes and triple concentrate tomato pastes could be the use of the profiling for the volatile products of secondary metabolisms pathway which are in part responsible for the aroma of tomato and triple concentrate tomato pastes.

Among the many factors affecting the profiling of the volatile components, a peculiar role is, in fact, played by the cultivar, the atmospheric, pedologic, and fostering conditions, the ripening degree and the storing procedure. The characterization of volatile fraction of tomato is a topic that has attracted the interest of many scientists. Many studies have concerned development and optimization of analytical methods for the characterization of volatile components using various analytical techniques.[16] Other studies show the change in the distribution of volatile compounds are related with the physical-chemical changes suffered by tomatoes after some industrial processes of production and/or the common preservation conditions used in retail outlets and households.[17] Further studies were performed in order to use the aroma as marker to differentiate between hybrid cultivars and traditional ones. Furthermore it was used as a marker in tomatoes breeding programs for selecting varieties with better quality characteristics.[18]

The main precursors of volatile compounds in tomato are free amino acids, fatty acids and carotenoids. Over 400 volatile components have been identified in fresh tomatoes and their formation is the result of different biosynthetic pathways.[16d] For instance saturated and unsaturated C6 and C9 alcohols and aldehydes, that are impact compounds of fresh tomato, are originated through lipoxygenase activity[19] while carotenoid and terpene

derivatives are generated by glycosidase activities.[17a] The different expression of the enzyme pool could affect the biosynthetic pathways involved in the formation of volatile compounds. The availability of these enzymes could be dependent on the pedoclimatic conditions and, therefore, the components of volatile fraction could be used as markers for the identification of geographical origin.

Accordingly, we evaluated the capability of volatile fraction profiling detected by SPME-GC-MS to act as marker for the discrimination between tomatoes harvested in different Italian region. The same aim was pursued for triple concentrate tomato pastes produced in two different Italian zones. Two supervised pattern recognition procedures such as Linear Discriminant Analysis (LDA) and Soft Independent Modeling of Class Analogy (SIMCA) were applied to tomato samples and triple concentrate tomato paste samples and the discrimination and classification ability of these multivariate techniques was evaluated.

The analytical method used was chosen referring to the many sources available in literature concerning the analysis of volatile components of fresh tomato and its derivatives using SPME-GC-MS. Ten frozen tomatoes were randomly chosen for each region and homogenized at 25 °C using an electric mixer. Then 2 g of homogenized tomatoes was put in a 10 mL vial and 2 mL of saturated $CaCl_2$ solution was added [16a] before thawing in order to avoid loss of volatile components[16d] easily degradable and promoting the salting out effect. Beltran et al. add at the ground tomato a 5% (w/w) of the $CaCl_2$ solution whereas Buttery et al. used a volume (mL) equal to the sample weight (g). The methods reported were tested and the best performance in term of S/N ratio was obtained using the procedure developed by Buttery et al.[16a]. Then 80 μL of ethyl isobutanoate solution at 5 mg/L in methanol[19a] was added as internal standard to obtain a

concentration of 200 µg/kg in each sample. The same procedure was followed for the concentrate paste samples, but 3 g was weighted and 120 µL of internal standard solution was added. The choose of the SPME fiber that gives best analytical performance was in accordance to the study conducted by Beltran et al.[16d] that compared the extraction efficiency of volatile compounds in head space of samples of fresh tomato. This study shows that Carboxen/PDMS fiber is the more efficient then PDMS/DVB and CW/DVB fibers. In order to test the possibility of using the autosampler, several tests were performed to evaluate peak areas of analytes of interest after 8, 10, 12 hours after the initial analysis. Since there was not a significant loss of analytes in this range of time, we decided to perform the analysis using the autosampler. The sample was preheating to 35 °C for 30 min and the extraction of volatile compounds took place for 60 min at the same temperature. The analytes were desorbed for 10 min from the fiber into the GC injector set at 300 °C in splitless mode. Integration was conducted on chromatographic peaks presenting a signal to noise ratio equal to or greater than 100 in at least one sample in order to consider only the most abundant detected compounds. Accordingly, 38 compounds in fresh tomato and 32 compounds in triple concentrate samples were considered (Table 3.12, Table 3.13). Quantification was carried out through comparison to ethyl isobutanoate area and expressed as µg/kg. These volatile molecules were identified by matching their recorded mass spectra with those present in the NIST02 library and by comparison of their retention indices (RI) relative to (C6-C20) n-alkanes with those of the literature. The non-isothermal Kovats retention indices, using definition of Van den Dool and Kratz (Van den dool & Kratz, 1963), were determined by using n-alkanes at the same chromatographic conditions and calculated as follows:

$$RI_{x=}100\,\mathrm{n} + 100(\mathrm{t_x} - \mathrm{t_n})/(\mathrm{t_{n+1}} - \mathrm{t_n})$$

where $t_x$ is the retention time of each unknown compound (x), $t_n$+1 and $t_n$ are retention times of n-alkanes eluting directly after and before the compound (x).

| Compound | Retention Time | Emilia Romagna | Basilicata | Calabria | Puglia | RI[a] | Methods of identification[b] |
|---|---|---|---|---|---|---|---|
| **2,3dihydrofuran** | 2.14 | 41.3 | 26.5 | 6.87 | 31.2 | 601 | MS |
| **2-methylpropanal** | 2.36 | 120 | 55.1 | 33.4 | 63.5 | 619 | MS |
| **2-methylfuran** | 2.53 | 247 | 128 | 103 | 102 | 632 | MS |
| **3-methylbutanal** | 3.13 | 27.3 | 4.54 | 5.81 | 12.0 | 680 | MS, RI |
| **2-methylbutanal** | 3.65 | 84.9 | 24.9 | 31.8 | 60.2 | 706 | MS, RI |
| **methyl butanoate** | 4.61 | 26.5 | 4.08 | 5.57 | 7.22 | 725 | MS, RI |
| **1-pentanol** | 5.17 | 29.7 | 9.92 | 3.99 | 20.8 | 737 | MS, RI |
| **methyl 3-methylbutanoate** | 6.81 | 46.9 | 21.6 | 15.6 | 40.0 | 770 | MS, RI |
| **Hexanal** | 8.20 | 258 | 217 | 192 | 130 | 798 | MS, RI |
| **trans-2-hexenal** | 12.11 | 10.8 | 5.27 | 1.10 | 1.32 | 851 | MS, RI |
| **Hexanol** | 13.58 | 1.90 | 5.05 | 1.79 | 1.80 | 870 | MS, RI |
| **Heptanal** | 16.19 | 2.28 | 5.87 | 5.89 | 1.16 | 905 | MS, RI |
| **1-nitropentane** | 16.31 | 20.8 | 11.4 | 10.8 | 32.6 | 907 | MS, RI |
| **Benzaldehyde** | 20.53 | 11.2 | 9.36 | 8.39 | 9.74 | 966 | MS, RI |
| **β-pinene** | 21.34 | 9.86 | 18.2 | 45.6 | 8.07 | 977 | MS, RI |
| **5-decene** | 21.91 | 12.1 | 15.1 | 12.0 | 13.0 | 985 | MS, RI |
| **6-methyl-5-hepten-2-one** | 22.58 | 1078 | 763 | 1132 | 665 | 992 | MS, RI |
| **6-methyl-5-hepten-2-ol** | 22.91 | 97.2 | 105 | 129 | 88.2 | 999 | MS, RI |
| **Octanal** | 23.52 | 1.04 | 22.8 | 48.8 | 2.03 | 1010 | MS, RI |
| **p-cymene** | 24.64 | 1.95 | 22.1 | 89.5 | 3.30 | 1028 | MS, RI |
| **D-limonene** | 24.93 | 41.1 | 597 | 1677 | 96.4 | 1033 | MS, RI |
| **2-isobutylthiazole** | 25.25 | 55.4 | 140 | 65.0 | 68.3 | 1039 | MS, RI |
| **γ-terpinene** | 26.29 | 1.23 | 0.798 | 6.76 | 0.509 | 1056 | MS, RI |
| **Terpinolene** | 26.89 | 14.8 | 8.65 | 14.3 | 7.07 | 1067 | MS, RI |
| **Linalool** | 29.33 | 17.0 | 42.9 | 403 | 12.6 | 1109 | MS, RI |
| **6-methyl-3,5-heptadien-2-one** | 29.48 | 52.4 | 21.5 | 34.4 | 26.7 | 1112 | MS, RI |
| **Unknown** | 29.54 | 71.3 | 28.2 | 30.0 | 27.2 | 1113 | - |
| **5-methyl-5-nonanol** | 30.37 | 16.6 | 124 | 279 | 127 | 1129 | MS |
| **3,7-dimethyl-3-octanol** | 31.40 | 6.74 | 15.2 | 52.3 | 15.3 | 1148 | MS, RI |
| **2,6,6-trimethyl-1-cyclohexene-1-carboxaldehyde** | 35.39 | 12.8 | 6.73 | 10.1 | 5.87 | 1227 | MS, RI |
| **linalyl acetate** | 36.89 | 3.54 | 2.15 | 58.3 | 1.46 | 1258 | MS, RI |
| **Genanial** | 37.79 | 16.4 | 18.4 | 25.9 | 6.39 | 1277 | MS, RI |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Unknown** | 40.00 | 10.7 | 47.4 | 69.0 | 27.4 | 1326 | - |
| **Unknown** | 41.09 | 2.82 | 14.9 | 22.7 | 9.05 | 1351 | - |
| **geranyl acetate** | 41.88 | n.d.c | n.d.c | 2.18 | n.d.c | 1369 | MS, RI |
| **2-buten-1-one, 1-(2,6,6-trimethyl-1,3-cyclohexadien-1-yl)** | 42.73 | 16.5 | 0.197 | 4.49 | 4.28 | 1388 | MS, RI |
| **geranyl acetone** | 45.66 | 4.06 | 27.5 | 29.6 | 6.34 | 1458 | MS, RI |
| **ß-ionone** | 46.95 | 2.31 | 2.58 | 3.29 | 1.36 | 1489 | MS, RI |

**Table 3.12** Mean concentration values (µg/kg, referred to ethyl isobutanoate) of volatile compounds detected in tomato samples, retention indices and methods of identification ([a] Retention index on ZB-5ms column; [b] MS: mass spectrum; RI: retention index when the calculated RI corresponds to the RI in the literature; [c] n.d.= not detected)

| Compound | Retention Time | Emilia Romagna | Calabria | RI[a] | Methods of identification[b] |
|---|---|---|---|---|---|
| **ethanol** | 1.67 | 105 | 1.22 | - | MS |
| **dimethyl sulfide** | 1.87 | 1452 | 564 | - | MS |
| **unknown** | 2.36 | 1096 | 359 | - | - |
| **2-methylfuran** | 2.53 | 853 | 131 | 632 | MS |
| **3-methylbutanal** | 3.19 | 44.8 | 131 | 680 | MS, RI |
| **2-pentanone** | 3.67 | 47.6 | 10.1 | 706 | MS, RI |
| **2-ethylfuran** | 3.90 | 67.3 | 35.5 | 711 | MS, RI |
| **methyl butanoate** | 4.59 | 10.5 | 7.36 | 725 | MS, RI |
| **unknown** | 5.14 | 10.2 | 4.59 | 736 | - |
| **dimethyl disulfide** | 5.27 | 7.83 | 13.6 | 739 | MS, RI |
| **Tetrahydro-2-(3-pentynyloxy)-2H-pyran** | 7.31 | 10.4 | 7.41 | 780 | MS |
| **hexanal** | 8.15 | 11.1 | 5.99 | 798 | MS, RI |
| **3-furaldehyde** | 10.4 | 1.49 | 53.7 | 824 | MS, RI |
| **5-methyl-5-hexen-2-ol** | 13.05 | 22.1 | 30.5 | 861 | MS, RI |
| **heptanal** | 16.23 | 0.705 | 0.467 | 905 | MS, RI |
| **1-(2-furanyl) ethanone** | 16.67 | 5.65 | 13.9 | 912 | MS, RI |
| **methyl 4,4-dimethyl-2-pentenoate** | 17.06 | 4.60 | 2.00 | 917 | MS |
| **α-pinene** | 18.12 | 4.76 | 1.72 | 932 | MS, RI |
| **2,7-dimethyloxepine** | 18.48 | 7.52 | 1.55 | 937 | MS |

| | | | | | |
|---|---|---|---|---|---|
| **benzaldehyde** | 20.51 | 6.61 | 12.0 | 966 | MS, RI |
| **β-pinene** | 21.31 | 9.14 | 1.72 | 977 | MS, RI |
| **6-methyl-5-heptene-2-one** | 22.38 | 453 | 204 | 992 | MS, RI |
| **p-cymene** | 24.62 | 18.6 | 3.69 | 1028 | MS, RI |
| **D-limonene** | 24.99 | 439 | 39.7 | 1033 | MS, RI |
| **cis-linalyl oxide** | 27.44 | 6.72 | 5.92 | 1076 | MS, RI |
| **trans-linalyl oxide** | 28.45 | 7.60 | 2.04 | 1093 | MS, RI |
| **linalool** | 29.27 | 22.1 | 10.4 | 1109 | MS, RI |
| **nonanal** | 29.40 | 20.1 | 9.28 | 1110 | MS, RI |
| **unknown** | 35.81 | 14.0 | 9.75 | 1236 | - |
| **unknown** | 36.73 | 7.34 | 5.07 | 1255 | - |
| **geranial** | 37.83 | 12.9 | 0.562 | 1278 | MS, RI |
| **2-buten-1-one, 1-(2,6,6-trimethyl-1,3-cyclohexadien-1-yl)** | 42.75 | 13.4 | 3.16 | 1388 | MS, RI |

**Table 3.13** Mean concentration values (µg/kg, referred to ethyl isobutanoate) of volatile compounds detected in triple concentrate tomato paste samples, retention indices and methods of identification.

The compounds detected were in agreement with literature data on volatile compounds detected in tomato and tomato paste using SPME and purge-and-trap methods.[16d, 16f, 17a, 17d] The most abundant volatile compound belong to terpenes. Hexanal, trans-2-hexenal and hexanol derive from lipoxygenase (LOX) pathway. the C6 compounds were generated by the biochemical pathway through conversion of linoleic and α-linolenic acids.[20]

### *Statistical analysis on tomato samples*

Initially, the obtained data were subjected to Principal Component Analysis (PCA) in order to have an overview of data, underline the most important variables and their

possible correlations. Figure 3.4 shows the biplot PC1 vs PC2 of scores and loadings for tomato samples which retain 49.32% of the total variance.

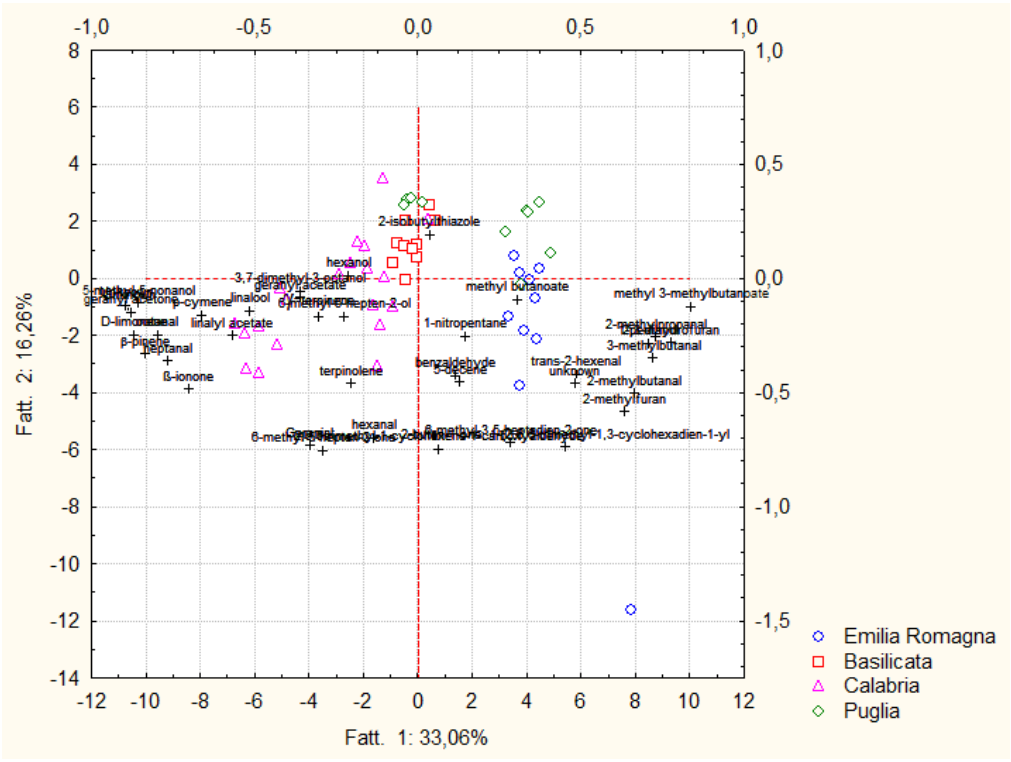


**Figure 3.4** Biplot of principal component scores and loadings for tomato samples.

The plot shows a separation of the production areas only on the basis of the first principal component. The compounds methyl 3-methylbutanoate, 2,3-dihydrofuran, 2-methylpropanal, 3-methylbutanal, 1-pentanol, 2-methylbutanal and 2-methylfuran (compounds with the highest positive loading values on the PC1) have higher concentrations in the samples coming from Emilia Romagna whereas those on the left of plot (geranyl acetone, D-limonene, 5-methyl-5-nonanol, β-pinene, octanal and heptanal) are present at higher concentration in Calabrian tomato samples. Variables with loading values near to zero have similar concentrations in all tomato samples regardless the

production zone. Samples coming from Puglia appeared to have an heterogeneous position in the bidimensional plot whereas Basilicata have shown values between Calabria and Emilia Romagna.

Pattern recognition analysis was carried out using linear discriminant analysis (LDA) and Soft Independent Modeling of Class Analogy (SIMCA) in order to obtain classification rules for distinguishing four Italian cultivation areas of tomatoes (Calabria, Basilicata, Puglia, Emilia Romagna) and between triple concentrated tomato paste samples coming from the producing areas of Collecchio (Emilia Romagna) and Crotone (Calabria).

In order to obtain models that have good stability, the number of samples must be at least 3 times the number of variables so, since the samples available were 50 and the volatile compounds detected were 38, the application of a variables reduction techniques is needed. Stepwise linear discriminant analysis (S-LDA) permits the variables with a major discriminant capacity to be selected discarding redundant information.

The forward stepwise analysis (F to enter=2.00 and F to remove=1.00) has retained 16 variables showed in Table 3.14

| | Wilks' λ | Parziale Wilks' λ | F-remove | p-level | Tollerance |
|---|---|---|---|---|---|
| **2,3-dihydrofuran** | 0,001843 | 0,344904 | 19,62667 | 0,000000 | 0,079846 |
| **p-cymene** | 0,001336 | 0,475681 | 11,38990 | 0,000034 | 0,375713 |
| **3-methylbutanal** | 0,001248 | 0,509330 | 9,95477 | 0,000094 | 0,441855 |
| **1-nitropentane** | 0,001206 | 0,527194 | 9,26731 | 0,000158 | 0,086426 |
| **2-methylfuran** | 0,001131 | 0,561879 | 8,05734 | 0,000411 | 0,060373 |
| **2-isobutylthiazole** | 0,001093 | 0,581265 | 7,44397 | 0,000682 | 0,321804 |
| **2-methylbutanal** | 0,001045 | 0,608415 | 6,65070 | 0,001345 | 0,124285 |
| **heptanal** | 0,000952 | 0,667882 | 5,13845 | 0,005319 | 0,113189 |
| **2-buten-1-one, 1-(2,6,6-trimethyl-1,3-cyclohexadien-1-yl)** | 0,000933 | 0,681399 | 4,83156 | 0,007131 | 0,053847 |

| | | | | | |
|---|---|---|---|---|---|
| **methyl 3-methylbutanoate** | 0,000832 | 0,764061 | 3,19090 | 0,037199 | 0,456252 |
| **octanal** | 0,000824 | 0,771292 | 3,06410 | 0,042518 | 0,163921 |
| **linalyl acetate** | 0,000823 | 0,772181 | 3,04867 | 0,043217 | 0,569403 |
| **1-pentanol** | 0,000804 | 0,790192 | 2,74366 | 0,059817 | 0,235784 |
| **2,6,6-trimethyl-1-cyclohexene-1-carboxaldehyde** | 0,000797 | 0,797138 | 2,62972 | 0,067618 | 0,469023 |
| **5-methyl-5-nonanol** | 0,000795 | 0,799760 | 2,58721 | 0,070793 | 0,115752 |
| **hexanal** | 0,000718 | 0,884790 | 1,34552 | 0,277630 | 0,188022 |

**Table 3.14** Summing up of the forward stepwise LDA for tomato samples.

The compounds showing a great discriminant power arise from different biosynthetic pathways. Between these, 3-methylbutanal, 2-isobutylthiazole, and hexanal are among the most important compounds in characterizing the tomato flavor.[21] Those results are in accordance with results recently published by Walczak et al. in which p-cymene, heptanal, octanal and hexanal mostly contribute to the LDA discriminant model to trace the geographical origin of honey.[22] The presence of hexanal among themost important variables in the differentiation of cultivation zones is also in accordance with our work concerning geographical origin determination of olive oil.[19a] The differentiation between groups is very significant since the very low Wilks λ value (0.00064) shows that the model is highly discriminating, whereas the very low p-level value (<0.00001) indicates that the classification occurs with an high degree of reliability. The scatterplot of canonical scores on the first three discriminant functions (Figure 3.5) shows a good separation between the four clusters corresponding to the four cultivation zones

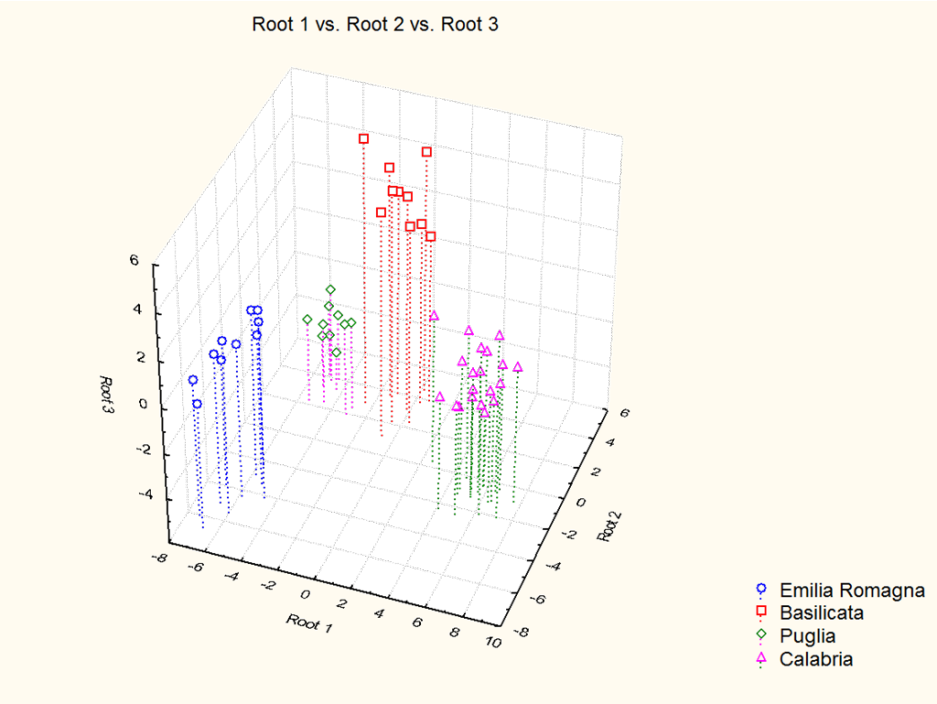


**Figure 3.5** LDA scatterplot of canonical scores on the first three discriminant functions for tomato samples.

A 10-fold cross-validation was performed in order to verify the goodness of the method in terms of prediction ability (Table 3.15). The proposed model has correctly predicted 96% of cultivation zones and this suggests that the presented method may be successfully used for checking origin of tomatoes

| | **Emilia Romagna** | **Basilicata** | **Calabria** | **Puglia** |
|---|---|---|---|---|
| **Emilia Romagna** | 10 | 0 | 0 | 0 |
| **Basilicata** | 0 | 10 | 0 | 0 |
| **Calabria** | 0 | 2 | 18 | 0 |
| **Puglia** | 0 | 0 | 0 | 10 |

**Table 3.15** Prediction matrices for LDA model of fresh tomato samples (rows represent the true class, columns report the assigned class)

SIMCA was applied to the same data matrix subjected to LDA without applying any method for reduction of the variables and validation was carried out by 10-fold cross-validation. The models obtained were based on normal range, 5% as the significance level for critical distance and a number of components for each category so that explains about 90% of class variance. The SIMCA model developed showed good results in terms of prediction ability. It correctly predicted 94% of geographical origin. SENS and SPEC values obtained were very satisfactory: mean sensitivity of 82% and a mean specificity of 100% (Table 3.16)

| | **Emilia Romagna** | **Basilicata** | **Calabria** | **Puglia** | **SENS(%)** | **SPEC(%)** |
|---|---|---|---|---|---|---|
| **Emilia Romagna** | 10 | 0 | 0 | 0 | 70 | 100 |
| **Basilicata** | 0 | 8 | 2 | 0 | 90 | 100 |
| **Calabria** | 0 | 0 | 20 | 0 | 85 | 100 |
| **Puglia** | 1 | 0 | 0 | 9 | 80 | 100 |

**Table 3.16** Prediction matrices for SIMCA model of fresh tomato samples (rows represent the true class, columns report the assigned class)

The same chemometric approaches applied on tomato samples were applied to triple concentrate tomato paste samples. Since the sample:variable ratio is less than three as for tomato samples, stepwise linear discriminant analysis was applied to the concentration values of the 32 volatile compounds detected in concentrated tomato samples in order to build a classifier capable to distinguish between samples coming from Emilia Romagna and Calabria.The choice of F to enter=3 and F to remove=2 has led to the selection of 3 variables (Table 3.17) that allows the construction of a model with very satisfactory statistical parameters. (Wilks $\lambda$=0.02880, F(3.26)=292.27 and p-level <0.00001).

| | Wilks' λ | Parziale Wilks' λ | F-remove | p-level | Tollerance |
|---|---|---|---|---|---|
| **3-methylbutanal** | 0.083367 | 0.345445 | 49.26516 | 0.000000 | 0.432682 |
| **ethanol** | 0.072101 | 0.399423 | 39.09398 | 0.000001 | 0.729928 |
| **tetrahydro-2-(3-pentynyloxy)-2H-pyran** | 0.034273 | 0.840269 | 4.94248 | 0.035107 | 0.469605 |

**Table 3.17** Summing up of the forward stepwise LDA for concentrate samples

The cross-validation procedure, performed as for tomato samples, has shown a prediction ability of 100% for each class (Table 3.18).

| | Emilia Romagna | Calabria |
|---|---|---|
| **Emilia Romagna** | 10 | 0 |
| **Calabria** | **0** | 20 |

**Table 3.18** Prediction matrices for LDA model of triple concentrate tomato paste samples (rows represent the true class, columns report the assigned class)

SIMCA was also applied to the triple concentrate tomato paste data matrix. Models obtained were based on a number of components for each class so that explains about 85% of the variance of category. The model developed was capable to correctly classify 97% of the samples submitted during cross-validation procedure and shows satisfactory values of mean sensitivity (90%) and mean specificity (100%) (Table 3.19).

| | Emilia Romagna | Calabria | SENS(%) | SPEC(%) |
|---|---|---|---|---|
| **Emilia Romagna** | 10 | 0 | 90 | 100 |
| **Calabria** | 1 | 19 | 90 | 100 |

**Table 3.19** Prediction matrices for SIMCA model of triple concentrate tomato paste samples (rows represent the true class, columns report the assigned class)

In conclusion, in this study the volatile fraction determined by a rapid and easy SPME-GC-MS analytical method was used as discriminating parameter for identifying the origin area of fresh tomatoes and triple concentrate tomato pastes. For tomatoes samples the two considered chemometric techniques (SLDA and SIMCA) showed comparable results and produced models capable to achieve very good percentage of prediction: 96% for stepwise LDA and 94% for SIMCA. Moreover, a sensitivity (SENS) mean value of 82% and specificity (SPEC) mean value of 100%were achieved by SIMCA model. The same chemometric approaches applied to triple concentrated tomato paste samples coming from two different Italian regions, showed excellent results. Stepwise LDA has correctly classified all samples for both category (Emilia Romagna and Calabria) whereas SIMCA showed very good results in terms of predictive ability (97%), sensitivity (90%) and specificity (100%).

## 3.3 Multielement Fingerprinting as a Tool in Origin Authentication of PGI Food Products: Tropea Red Onion.

Onion (Allium cepa L.) is one of the oldest cultivated plants, and it is now used both as a food and for medical purposes. In fact, onion is a rich source of a number of phytonutrients, as flavonoids, fructo-oligosaccharides, thiosulfinates and other sulphur compounds which make it an important food of the Mediterranean diet.

The red onion variety Tropea is a typical Italian variety, cultivated in specific areas of Calabria. This cultivar, characterized by both white and purple flowers, is known for its distinctive red and sweet bulb (lengthened or oval).[23] Due to its characteristics, it was awarded with PGI certification by the European Union as "Cipolla Rossa di Tropea Calabria".[15] Analysis of the methanolic extract from the bulbs of Tropea onions revealed the presence of high concentrations of flavonoids which continue to attract attention as potentially useful agents with implications for inflammation, cardiovascular diseases, and cancer.[23-24] Respect to other onion cultivars, the variety Tropea seems to have peculiar nutritional properties such as a relatively high content of delphinidin derivatives (about 30% of the total anthocyanin content), the presence of petunidin derivatives and the highest amount of free quercetin (557.8 mg/kg in fresh bulbs).[25] Due to its special characteristics, the Tropea red onion is a product known throughout the world, important for local and national economy. On the other hand, the Tropea red onion is subject to food fraud. In 2008 it was estimated that, against a production of Tropea red onion in Calabria of about twenty thousand tons, the red onions labeled with the PGI brand were over one hundred thousand tons. This discrepancy can be evidently explained by an importation of onion fraudulently labeled with the PGI brand from abroad.[16f, 17a] Identification of geographical origin and authenticity of food products represents an important goal in

order to assure organoleptic and nutritional characteristics to consumers and to prevent unfair competition that can eventually damage the whole agricultural sector.

Onions have been the subject of numerous works some of which concerned the discrimination of their production area. Ariyama et al. published some important papers in food forensic field concerning onions and welsh onions. The elemental concentration profile evaluated by flame atomic absorption spectroscopy (FAAS), inductively coupled plasma emission spectroscopy (ICP-AES) and inductively coupled plasma mass spectrometry (ICP-MS) was applied to welsh onions samples coming from Japan and China for the determination of their geographical origin.[26] Subsequently, a similar method was successfully applied by the same research group to discriminate the Japanese onions from those coming from abroad.[27] These studies demonstrated also that the differences in elements content by fertilization were smaller than those between production zones when judged from an overview of numerous elements.[28] In this work, as carried out for tomato and triple concentrate tomato paste samples, the capability of multielement profiling to act as marker for the classification of Tropea red onion was evaluated. Samples were provided by the Consortium for the Protection of Tropea Red Onion and were harvested during the crop season 2009 from different fields belonging to the municipalities of Capo Vaticano, Amantea, Nocera Terinese, and Briatico (Figure 3.6) and immediately stored at -20 °C. Non-Tropea onion samples came from three Italian regions (Campania, Sicilia, and Piemonte) and The Netherlands. A total of 120 samples grown according to the production regulations of the Consorzio di tutela della Cipolla Rossa di Tropea (Consortium for the protection of Tropea Red Onion) and 80 onion samples from fields not belonging to the cultivation areas specified in the production regulations were used for the development of the research work.

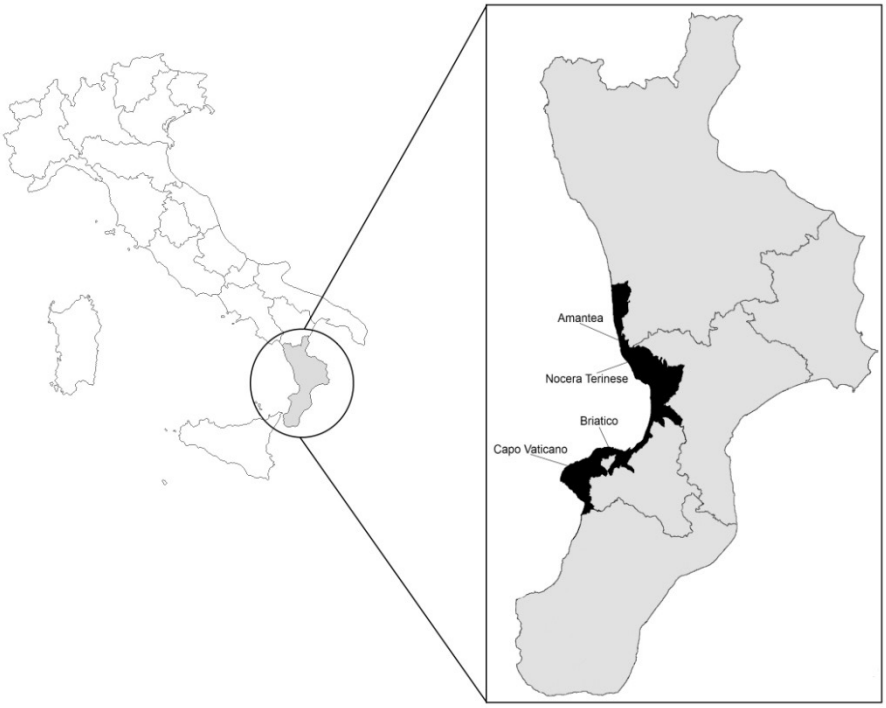


**Figure 3.6** Calabria map. In black is spotlighted the cultivation areas of Tropea red Onion.

The first part of the research activity provided for the development of the mineralization and ICP-MS analytical method. The proper amount of onion sample and acid mineralization mixture were investigated in order to obtain good limit of quantification without stress the mineralization vessel. In order to assess the accuracy of the method, certified reference material NCS ZC85006 Tomato (Be, Na, Mg, Al, K, Ca, Sc, V, Mn, Fe, Ni, Co, Cu, Zn, As, Se, Rb, Sr, Y, Cd, Cs, Ba, La, Ce, Nd, Sm, Eu, Gd, Dy, Er, Yb, Lu, Pb, Th, U; China National Analysis Center for Iron & Steel 2000) already used for the traceability study on tomato, was submitted to the whole analytical process and the accuracy of the elements evaluated. Initially 46 elements were investigated $^{7}$Li, $^{9}$Be, $^{23}$Na, $^{24}$Mg, $^{27}$Al, $^{39}$K, $^{44}$Ca, $^{45}$Sc, $^{51}$V, $^{52}$Cr, $^{53}$Cr, $^{55}$Mn, $^{54}$Fe, $^{56}$Fe,$^{58}$Ni, $^{60}$Ni, $^{59}$Co, $^{63}$Cu, $^{64}$Zn, $^{66}$Zn, $^{69}$Ga, $^{75}$As, $^{82}$Se, $^{85}$Rb, $^{88}$Sr, $^{89}$Y, $^{107}$Ag, $^{114}$Cd, $^{115}$In, $^{133}$Cs, $^{138}$Ba, $^{139}$La, $^{140}$Ce, $^{141}$Pr,

$^{142}$Nd, $^{152}$Sm, $^{153}$Eu, $^{158}$Gd, $^{159}$Tb, $^{164}$Dy, $^{165}$Ho, $^{166}$Er, $^{169}$Tm, $^{174}$Yb, $^{175}$Lu, $^{205}$Tl, $^{208}$Pb, $^{209}$Bi, $^{232}$Th and $^{238}$U. According to the polyatomic isobaric interferences for these elements, the optimization of the $CH_4$ flow rate and the RPq value were carried out. Finally 25 elements were monitored: 22 elements detected in standard mode ($^{27}$Al, $^{138}$Ba, $^{44}$Ca, $^{114}$Cd, $^{140}$Ce, $^{164}$Dy, $^{153}$Eu, $^{69}$Ga, $^{158}$Gd, $^{165}$Ho, $^{139}$La, $^{24}$Mg, $^{55}$Mn, $^{23}$Na, $^{142}$Nd, $^{141}$Pr, $^{85}$Rb, $^{152}$Sm, $^{88}$Sr, $^{205}$Tl, $^{89}$Y, $^{66}$Zn) and 3 elements in DRC mode ($^{54}$Fe, $^{52}$Cr, $^{58}$Ni). ICP-MS analyses were performed following the operating program and parameters reported in Chapter 4 (Table 4.5). One hundred twenty Tropea onion samples and eighty non Tropea onion samples were processed. Pattern recognition analysis was carried out using four supervised chemometric techniques (LDA, Stepwise-LDA, SIMCA and BP-ANN) in order to obtain classification rules for distinguishing between Tropea samples grown according the production regulations and non-Tropea samples. The reliability of the classification rules was validated through a cross validation procedure. All these techniques allowed us to obtain satisfactory results because all of the models obtained showed prediction ability >90%. Standard LDA shows a mean prediction ability of 94%. The elimination of redundant information in stepwise-LDA by means of the selection of only those variables that actually contributed to the discrimination between classes allowed us to obtain only a slight increase of the prediction ability (94.5%). The analyses of both LDA discriminating models show that lanthanides, alkaline metals and alkaline earth metals are the most important variables in the distinction of geographical origin. The importance of lanthanides and in particular of dysprosium shows that rare earth metals, not considered in the previously study concerning onions authentication,[27] should be take into account. The important role of alkaline metals and alkaline earth metals in differentiation of cultivation zones is in agreement with previous studies about the

identification of the origin of virgin olive oil, and the research work previously discussed on tomato and tomato paste.[13, 29] SIMCA model has provided very good results in terms of mean prediction ability (95.5%) with 99.2% of correct prediction for the class Tropea. Moreover, the SIMCA model is satisfactory in terms of mean sensitivity (88.5%) and mean specificity (85.5%). Before the training of the neural network, principal component analysis (PCA) was performed in order to reduce the dimensionality of a data set retaining the most part of the information present in the original data. The scores of different number of principal component were considered as input in the ANN training. The best results were achieved with a multilayer percepron MLP 10-7-1 with the first 10 PCs as input variables (explaining 92.8% of total variance) and 100 epoch of back-propagating algorithms. The optimum learning rate and momentum value were $\eta = 0.15$ and $\alpha = 0.3$, respectively. The mean validation performance was 91.5% and the root mean square error was 0.312 for the training set, 0.484 for the test set and 0.363 for the validation set.

## 3.4 Authenticity of PGI Clementine of Calabria by multielement fingerprint.

Clementine (Citrus clementina Hort. ex Tan.) is one of the most important cultivated variety of citrus mandarins in the Mediterranean basin. In Italy, cultivations are located in the southern part of the country where the best weather conditions for their growth exist. Calabria is a region where the cultivation of clementines is widespread. The peculiar pedoclimatic conditions of Calabrian cultivation areas of clementines have contributed to develop a product which, due to its special qualities, was awarded with PGI certification by the European Union as "Clementine di Calabria".[30]

Several papers concern the beneficial health effects of citrus fruits and citrus-derived products. Some of these properties have been found to include anticancer, antiviral, and anti-inflammatory activities which are related to the presence of antioxidants including vitamin C, carotenoids and phenolic compounds.[31] Clementine of Calabria is a fruit known throughout the world with a great impact on regional and national Gross Internal Product. Due to its special characteristics it has been the subject of food fraud in which clementines cultivated abroad (e.g. Spain, Tunisia, Algeria,) were passed off as Clementine of Calabria and labeled with PGI brand.[5]

In this experimental work, similar to what was reported with tomato, triple concentrate tomato paste and Tropea red onion, the capability of multielement profiling to act as marker for the classification of the PGI Clementine of Calabria was evaluated.

Three pattern recognition chemometric models were applied to discriminate between clementine with PGI brand and non-PGI samples developing in this way a reliable analytical tool for traceability purpose. Multielement fingerprints of both peel and juice samples were considered as discriminative markers.

Clementine samples with PGI brand came from four different Calabrian cultivation zones located in the municipalities of Corigliano Calabro, Lamezia Terme, Pizzo Calabro and Rosarno whereas non-PGI clementine samples came from Spain, Tunisia and Algeria.

The first part of the research activity was devoted to the development of the mineralization and ICP-MS analytical methods for the accurate determination of the markers. The proper amount of peel and juice sample, together with acid mineralization mixture were investigated in order to obtain good limit of quantification without stress the mineralization vessel. Accordingly to the optimization results of the polyatomic isobaric interferences and the limit of quantification obtained, concentrations of $^{23}$Na, $^{24}$Mg, $^{27}$Al, $^{44}$Ca, $^{52}$Cr, $^{55}$Mn, $^{54}$Fe, $^{58}$Ni, $^{63}$Cu, $^{64}$Zn, $^{69}$Ga, $^{85}$Rb, $^{88}$Sr, $^{89}$Y, $^{138}$Ba, $^{139}$La, $^{140}$Ce, $^{141}$Pr, $^{142}$Nd, $^{152}$Sm, $^{153}$Eu, $^{158}$Gd, $^{166}$Er, $^{208}$Pb in peel samples and the concentrations of $^{7}$Li, $^{23}$Na, $^{24}$Mg, $^{27}$Al, $^{44}$Ca, $^{52}$Cr, $^{55}$Mn, $^{54}$Fe, $^{58}$Ni, $^{59}$Co, $^{63}$Cu, $^{64}$Zn, $^{69}$Ga, $^{82}$Se, $^{85}$Rb, $^{88}$Sr, $^{89}$Y, $^{138}$Ba, $^{139}$La, $^{140}$Ce, $^{141}$Pr, $^{142}$Nd, $^{152}$Sm, $^{153}$Eu, $^{158}$Gd, $^{166}$Er, in juice samples were submitted to statistical analysis. Among the elements monitored in standard and DRC modes, significant differences have been observed in the assay of $^{56}$Fe and $^{52}$Cr. Therefore, the statistical analysis has been carried out by means of the data acquired in standard mode except for the determination of chromium in juice samples and iron in both peel and juice samples, in which DRC values were used

In first instance principal component analysis (PCA) was carried out on the data matrices containing the concentration values (µg/kg) of the selected elements for peel and juice samples in order to perform an exploratory analysis for obtaining an overview of data and finding patterns in complex experimental data. The scores and loadings values of the first two PCs for juice and peel samples are plotted in Figure 3.7A and Figure 3.7B, respectively.

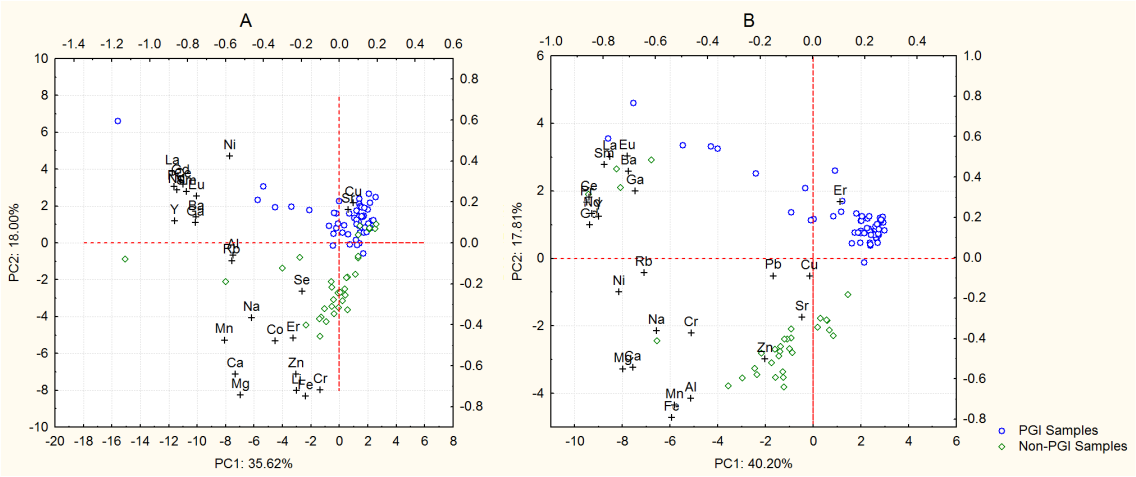


**Figure 3.7** Biplot PC1 vs PC2 scores and loadings for clementine juice samples (A) and for clementine peel samples (B)

Juice samples of the two production areas have similar score values on the first principal component whereas a separation of objects on the second PC is clear. The elements at top of the plot (i.e. Eu, Gd, Ce, Pr, Nd, Sm, La, Ni) are present at higher concentration in PGI samples whereas clementine samples coming from abroad have higher concentrations in elements with the highest negative loading values on the PC2 (i.e. Mg, Ca, Zn, Li, Fe, Cr, Na). Strontium and copper, which have loading values near to zero for both considered PCs, have similar concentrations in all clementine samples regardless the production zone. The four samples coming from the Algerian province of Blida are characterized by higher score values on the PC1 and slightly negative on the PC2.

For peel samples, the biplot of the first two principal components, which explain respectively 40.20% and 17.81% of the total variance, shows the presence of two different clusters corresponding to clementine samples with PGI brand and clementine samples coming from abroad (Figure 3.7 B). Also in this case, separation of the groups corresponding to the two production areas occurs principally of the second principal component and the position of variables in the plot of loadings is similar to those

observed for juice samples. Moreover, the cluster corresponding to the samples coming from the Algerian province of Blida is again characterized by higher negative score values on the PC1 and slightly positive on the PC2.

Pattern recognition analysis was carried out applying linear discriminant analysis (LDA), soft independent model of class analogy (SIMCA) and partial least square discriminant analysis (PLS-DA). In LDA, since to obtain models that have good stability, the number of samples must be at least 3 times the number of variables, the application of a variables reduction technique is needed. Forward stepwise LDA was used to classify clementines according to the categories PGI (samples grown in accordance with the production regulations) and non-PGI (samples cultivated in zones different from those cited in the production regulations) used as input a priori. The forward stepwise analysis performed on juice samples (F to enter=2.00 and F to remove=1.00) has retained 11 elements (Table 3.19).

| **Juice samples** | | | | | **Peel samples** | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Wilks' λ** | **Parziale Wilks' λ** | **F-remove** | **p-level** | | **Wilks' λ** | **Parziale Wilks' λ** | **F-remove** | **p-level** |
| **Cr** | 0.190443 | 0.757903 | 24.2766 | 0.000005 | **Fe** | 0.060440 | 0.722102 | 25.7846 | 0.000003 |
| **Ni** | 0.188653 | 0.765096 | 23.3339 | 0.000007 | **Ga** | 0.054743 | 0.797263 | 17.0375 | 0.000104 |
| **Y** | 0.183784 | 0.785367 | 20.7700 | 0.000019 | **Cr** | 0.052886 | 0.825243 | 14.1882 | 0.000351 |
| **Mg** | 0.180554 | 0.799415 | 19.0694 | 0.000039 | **Ba** | 0.052332 | 0.833981 | 13.3375 | 0.000511 |
| **Gd** | 0.176614 | 0.817249 | 16.9949 | 0.000095 | **Ca** | 0.050504 | 0.864177 | 10.5304 | 0.001834 |
| **Sr** | 0.172065 | 0.838855 | 14.5996 | 0.000270 | **Sm** | 0.046042 | 0.947913 | 3.68158 | 0.059278 |
| **Cu** | 0.167593 | 0.861239 | 12.2449 | 0.000784 | **Ce** | 0.046018 | 0.948424 | 3.64352 | 0.060572 |
| **Na** | 0.166547 | 0.866648 | 11.6942 | 0.001011 | **Zn** | 0.045950 | 0.949817 | 3.53991 | 0.064254 |
| **Co** | 0.153686 | 0.939174 | 4.92218 | 0.029499 | **Mn** | 0.045762 | 0.953727 | 3.25069 | 0.075890 |
| **Al** | 0.151891 | 0.950269 | 3.97734 | 0.049703 | **Cu** | 0.045580 | 0.957539 | 2.97102 | 0.089380 |
| **Zn** | 0.148973 | 0.968887 | 2.44056 | 0.122389 | **La** | 0.045011 | 0.969642 | 2.09770 | 0.152185 |

**Table 3.20** Summing up of the forward stepwise LDA: selected elements for juice samples and the 11 most important elements for peel samples

Among the most discriminating elements, the presence of alkali metals and alkaline earth metals (Mg, Na and Sr) is in agreement with other works carried out by our research group on the geographical traceability of oil,[13] tomatoes[29] and Tropea red onions.[14] On the other hand, these results are not in agreement with those obtained in the differentiation between frozen orange juice coming from Brazil and Florida[32] or Spain and Morocco.[33] To verify the goodness of method in terms of prediction ability, cross validation with cancellation group 10 was performed. The proposed model showed a total prediction ability of 96.6% and misclassification involved one sample belonging to PGI category and two samples belonging to non-PGI category.

For peel samples, such as for juice samples, forward S-LDA has been performed (F to enter=0.4 and F to remove=0.00) and four variables have been eliminated (Al, Eu, Ni and Pb, Table 3.20). The cross-validation procedure showed a prediction ability of 100% for each category.

SIMCA was applied to the same data matrices used for LDA and validation was carried out by 10-fold cross-validation procedure for both juice and peel samples. As regards juice samples, the model obtained was based on 9 PCs for first class (PGI samples) and 6 PCs for second class (Non-PGI samples) that explains 89.2% and 91.2% of total variance, respectively. SIMCA model has provided very good prediction ability classifying correctly all samples submitted (Table 3.21). Moreover, the SIMCA model is satisfactory in terms of mean sensitivity (81.8%) and mean specificity (96.6%). Better results are achieved by the SIMCA model built using peel samples data matrix. For these data, the two SIMCA classes are modeled considering 8 PCs for the first class (explained variance 89.7%) and 7 PCs for the second class (explained variance 90.2%). As for juice samples, SIMCA model is capable to correctly classify all samples submitted during cross-

validation procedure and, in this case, shows better values of mean sensitivity (88.6%) and mean specificity (100%) (Table 3.21).

| | Juice Samples | | | | | Peel Samples | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **PGI** | **non-PGI** | **SENS(%)** | **SPEC(%)** | | **PGI** | **non-PGI** | **SENS(%)** | **SPEC(%)** |
| **PGI** | 54 | 0 | 83.3 | 91.2 | **PGI** | 54 | 0 | 85.2 | 100 |
| **non-PGI** | 0 | 34 | 79.4 | 100 | **non-PGI** | 0 | 34 | 94.1 | 100 |

**Table 3.21** Prediction matrices for SIMCA of the cross validation procedure for juice and peel clementine samples (rows represent the true class, columns report the assigned class)

Finally partial least square discriminant analysis (PLS-DA) was applied to the same data matrix submitted to SIMCA approach. The optimum number of latent variables was selected by evaluating the parameter Root Mean Square Error of Prediction (RMSEP) as function of the number of LVs. Accuracy of the PLS model developed was evaluated as prediction ability on the basis of the correct classification of test set samples which were submitted as unknowns to the regression model. All variables were column centered and standardized by 1/standard deviation. The weighted regression coefficients of the PLS model were used to identify the most important variables. Their value gives an indication on experimental variables which have an significant impact on the response variables. The use of weighted coefficient is preferred because it allow to identify the real importance of variables as their sizes do not depend on the range of variation.[34] Variables that have weighted regression coefficients with high values play an important role in regression model and, in particular, positive values indicate a great deal in the relationship with the response for PGI category whereas negative mean a great deal to Non-PGI category.[35]

For juice samples 2 latent variables were chosen by which the RMSEP function has reached the first minimum. The 2D plot of the scores of the first three LVs is shown in Figure 3.8A. The model explained 73.8% of total variance and RMSEP and RMSEC values were 0.28 and 0.25, respectively. The closeness between these values can be interpreted as a lack of overfitting and good ability of the model to describe other data well.[36]

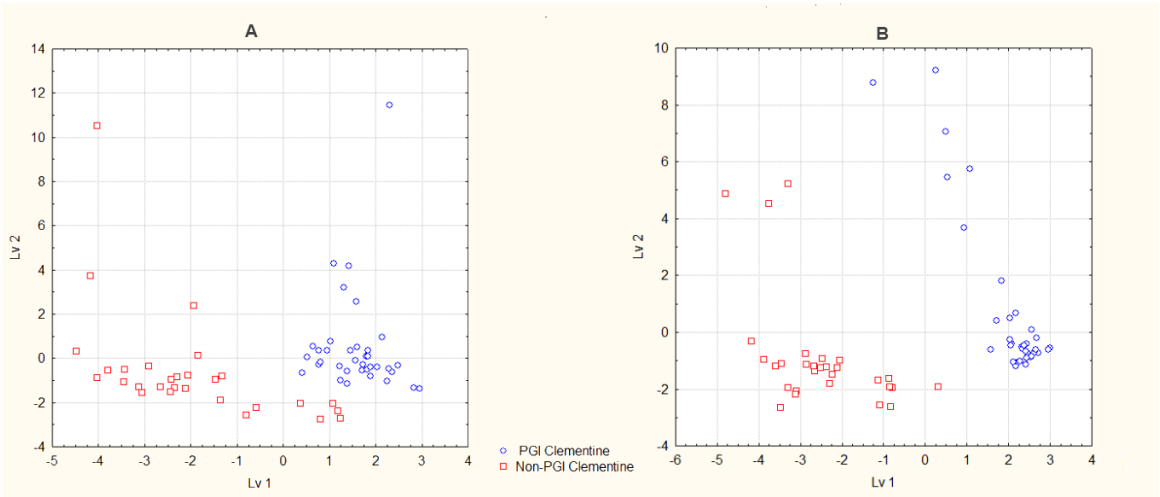


**Figure 3.8** PLS-DA Plot LV1 vs LV2 for clementine juice samples (A) and for clementine peel samples (B).

All samples of the test set were submitted to the model and only two of them (one PGI sample and one non-PGI sample) were wrongly classified in the class they belong to. The analysis of the PLS regression coefficients (Figure 3.9A) shows that copper, nickel and lanthanides are the most representative elements for the PGI category whereas magnesium and calcium are representative for Non-PGI category.

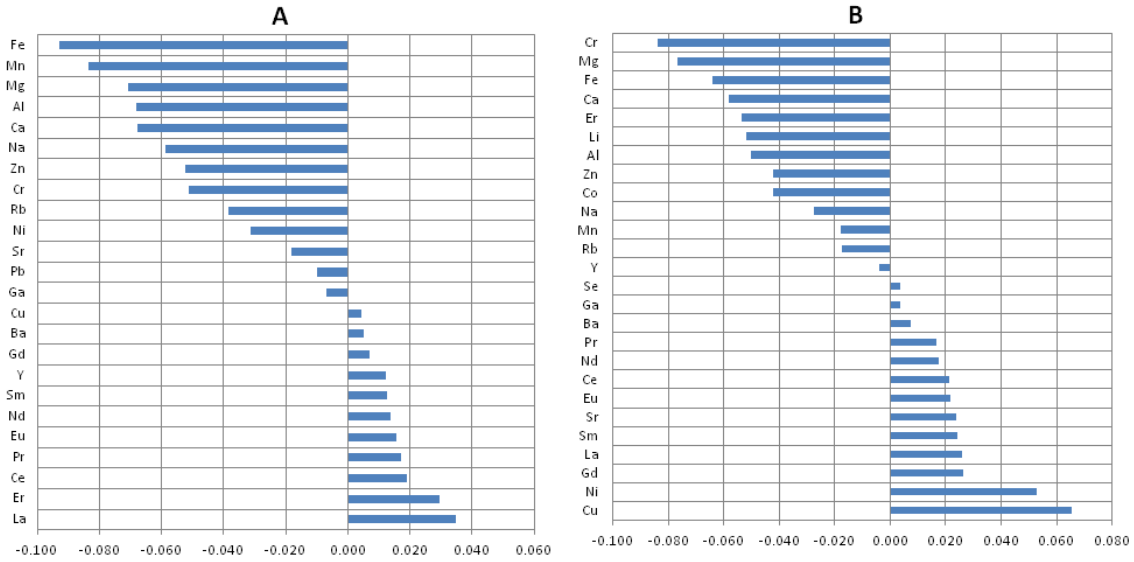


**Figure 3.9** Weighted regression coefficients for the variables in the PLS-DA models: clementine juice samples (A) and for clementine peel samples (B).

For peel samples 2 latent variables were chosen by which the RMSEP function has reached the first minimum. The 2D plot of the scores of the first two LVs is shown in Figure 3.8B. The model explained 92% of total variance and RMSEP and RMSEC values were 0.16 and 0.14, respectively. Also in this case, the PLS model shows good prediction ability since only one sample belonging to Non-PGI category was erroneously predicted. By looking at the regression coefficients (Figure 3.9B) it is possible to observe that most of the elements (8 out of 11) characterizing PGI peel samples belong to lanthanides. This result confirms the important role that these elements can play in applications of food authentication.[37] On the other hand, iron and manganese are the most representative elements for the non-PGI category. Moreover, such as observed for juice data, also for peel samples magnesium, calcium and sodium have an important role in the distinction of geographical origin.

## References


1. (a) Sharma, R.; Kaushal, B. B. L.; Sharma, P. C.; Singh, R. P., Evaluation of olive (Olea europaea L.) cultivars grown in different regions of Himachal Pradesh for oil extraction. *Acta Hortic.* **2005,** *696*, 559-562; (b) Cerretani, L.; Bendini, A.; Del Caro, A.; Piga, A.; Vacca, V.; Caboni, M. F.; Toschi, T. G., Preliminary characterisation of virgin olive oils obtained from different cultivars in Sardinia. *Eur. Food Res. Technol.* **2006,** *222* (3-4), 354-361.

2. (a) Angerosa, F.; Mostallino, R.; Basti, C.; Vito, R.; Serraiocco, A., Virgin olive oil differentiation in relation to extraction methodologies. *J. Sci. Food Agric.* **2000,** *80* (15), 2190-2195; (b) Angerosa, F.; Basti, C.; Vito, R., Virgin olive oil volatile compounds from lipoxygenase pathway and characterization of some Italian cultivars. *J. Agric. Food Chem.* **1999,** *47* (3), 836-839; (c) Benincasa, C.; De Nino, A.; Lombardo, N.; Perri, E.; Sindona, G.; Tagarelli, A., Assay of aroma active components of virgin olive oils from southern Italian regions by SPME-GC/ion trap mass spectrometry. *J. Agric. Food Chem.* **2003,** *51* (3), 733-741.

3. Fabbri, A.; Hormaza, J. I.; Polito, V. S., Random amplified polymorphic dna analysis of olive (olea-europaea l) cultivars. *Journal of the American Society for Horticultural Science* **1995,** *120* (3), 538-542.

4. Di Donna, L.; Mazzotti, F.; Salerno, R.; Tagarelli, A.; Taverna, D.; Sindona, G., Characterization of new phenolic compounds from leaves of Olea europaea L. by high-resolution tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2007,** *21* (22), 3653-3657.

5. http://www.coldiretti.it.

6. Marini, F., Artificial neural networks in foodstuff analyses: Trends and perspectives A review. *Anal. Chim. Acta* **2009,** *635* (2), 121-131.

7. (a) Sobolev, A. P.; Segre, A.; Lamanna, R., Proton high field NMRstudy of tomato juice. *Magn. Reson. Chem.* **2003,** *41*, 237-245; (b) Tiziani, S.; Schwartz, S. J.; Vodovotz, Y., Profiling of carotenoids in tomato juice by one- and two-dimensional NMR. *J. Agric. Food Chem.* **2006,** *54*, 6094-6100; (c) Capanoglu, E.; Beekwilder, J.; Boyacioglu, D.; Hall, R.; De Vos, R., Changes in antioxidant and metabolite profiles during production of tomato paste. *J. Agric. Food Chem.* **2008,** *56*, 964-973; (d) Arvanitoyannis, I. S.; Vaitsi, O. B., A review on tomato authenticity: Quality control methods in conjunction with multivariate analysis (Chemometrics). *Crit. Rev. Food Sci. Nutr.* **2007,** *47*, 675-699.

8. Turci, M.; Savo Sardaro, M. L.; Visioli, G. M., E.; Marmiroli, M.; Marmiroli, N., Evaluation of DNA extraction procedures for traceability of various tomato products. *Food Control* **2010,** *21*, 143-149.

9. Sequi, P.; Dell'Abate, M. T.; Valentini, M., Identification of cherry tomatoes growth origin by means of magnetic resonance imaging *J. Sci. Food Agric.* **2007,** *87*, 127-132.

10. Consonni, R.; Cagliani, L. R.; Stocchero, M.; Porretta, S., Triple Concentrated Tomato Paste: Discrimination between Italian and Chinese Products. *J. Agric. Food Chem.* **2009,** *57*, 4506-4513.

11. (a) Kelly, S.; Heaton, K.; Hoogewerff, J., Tracing the geographical origin of food: The application of multi-element and multi-isotope analysis. *Trends Food Sci. Technol.* **2005,** *16* (12), 555-567; (b) Gonzalvez, A.; Armenta, S.; de la Guardia, M., Trace-element composition and stable-isotope ratio for discrimination of foods with Protected Designation of Origin. *Trac-Trends in Analytical Chemistry* **2009,** *28* (11), 1295-1311; (c) Arvanitoyannis, I. S.; Tzouros, N. E., Agricultural produces: Synopsis of employed quality control methods for the authentication of foods and application of chemometrics for the classification of foods according to their variety or geographical origin. *Crit. Rev. Food Sci. Nutr.* **2001,** *41*, 287-319.

12. Kelly, S. D.; Bateman, A. S., Comparison of mineral concentrations in commercially grown organic and conventional crops - Tomatoes (Lycopersicon esculentum) and lettuces (Lactuca sativa). *Food Chem.* **2010,** *119*, 738-745.

13. Benincasa, C.; Lewis, J.; Perri, E.; Sindona, G.; Tagarelli, A., Determination of trace element in Italian virgin olive oils and their characterization according to geographical origin by statistical analysis. *Anal. Chim. Acta* **2007,** *585* (2), 366-370.

14. Furia, E.; Naccarato, A.; Sindona, G.; Stabile, G.; Tagarelli, A., Multielement fingerprint as tool in origin authentication of PGI food products: Tropea red onion. *J. Agric. Food Chem.* **2011,** *59*, 8450-8457.

15. H. Benabdelkamel; L. Di Donna; F. Mazzotti; A. Naccarato; G. Sindona; A. Tagarelli; Taverna., D., Multielement fingerprint as tool in origin authentication of PGI food products: Clementine di Calabria. *Submitted* **2011**.

16. (a) Buttery, R. G.; Teranishi, R.; Ling, L. C., Fresh tomato aroma volatiles - a quantitative study. *J. Agric. Food Chem.* **1987,** *35* (4), 540-544; (b) Buttery, R. G.; Teranishi, R.; Ling, L. C.; Flath, R. A.; Stern, D. J., Quantitative studies on origins of fresh tomato aroma volatiles. *J. Agric. Food. Chem.* **1988,** *36* (6), 1247-1250; (c) Bezman, Y.; Mayer, F.; Takeoka, G. R.; Buttery, R. G.; Ben-Oliel, G.; Rabinowitch, H. D.; Naim, M., Differential effects of tomato (Lycopersicon esculentum Mill) matrix on the volatility of important aroma compounds. *J. Agric. Food. Chem.* **2003,** *51* (3), 722-726; (d) Beltran, J.; Serrano, E.; Lopez, F. J.; Peruga, A.; Valcarcel, M.; Rosello, S., Comparison of two quantitative GC-MS methods for analysis of tomato aroma based on purge-and-trap and on solid-phase microextraction. *Anal. Bioanal. Chem.* **2006,** *385* (7), 1255-1264; (e) Song, J.; Fan, L. H.; Beaudry, R. M., Application of solid phase microextraction and gas chromatography time-of-flight mass spectrometry for rapid analysis of flavor volatiles in tomato and strawberry fruits. *J. Agric. Food. Chem.* **1998,** *46* (9), 3721-3726; (f) Serrano, E.; Beltran, J.; Hernandez, F., Application of multiple headspace-solid-phase microextraction

followed by gas chromatography-mass spectrometry to quantitative analysis of tomato aroma components. *Journal of Chromatography A* **2009,** *1216* (1), 127-133; (g) Aubert, C.; Baumann, S.; Arguel, H., Optimization of the analysis of flavor volatile compounds by liquid-liquid microextraction (LLME). Application to the aroma analysis of melons, peaches, grapes, strawberries, and tomatoes. *J. Agric. Food. Chem.* **2005,** *53* (23), 8881-8895; (h) Berna, A. Z.; Lammertyn, J.; Saevels, S.; Di Natale, C.; Nicolai, B. M., Electronic nose systems to study shelf life and cultivar effect on tomato aroma profile. *Sensors and Actuators B-Chemical* **2004,** *97* (2-3), 324-333.

17. (a) Servili, M.; Selvaggini, R.; Taticchi, A.; Begliomini, A. L.; Montedoro, G., Relationships between the volatile compounds evaluated by solid phase microextraction and the thermal treatment of tomato juice: optimization of the blanching parameters. *Food Chem.* **2000,** *71* (3), 407-415; (b) Zhang, Z. M.; Zeng, D. D.; Li, G. K., Study of the volatile composition of tomato during storage by a combination sampling method coupled with gas chromatography/mass spectrometry. *J. Sci. Food Agric.* **2008,** *88* (1), 116-124; (c) Maul, F.; Sargent, S. A.; Sims, C. A.; Baldwin, E. A.; Balaban, M. O.; Huber, D. J., Tomato flavor and aroma quality as affected by storage temperature. *J. Food Sci.* **2000,** *65* (7), 1228-1237; (d) Markovic, K.; Vacic, N.; Ganic, K. K.; Banovic, M., Aroma volatiles of tomatoes and tomato products evaluated by solid-phase microextraction. *Flavour and Fragrance Journal* **2007,** *22* (5), 395-400; (e) Krumbein, A.; Peters, P.; Brückner, B., Flavour compounds and a quantitative descriptive analysis of tomatoes (Lycopersicon esculentum Mill.) of different cultivars in short-term storage. *Postharvest Biology and Technology* **2004,** *32*, 15−28.

18. (a) Alonso, A.; Garcia-Aliaga, R.; Garcia-Martinez, S.; Ruiz, J. J.; Carbonell-Barrachina, A. A., Characterization of Spanish Tomatoes using Aroma Composition and Discriminant Analysis. *Food Science and Technology International* **2009,** *15* (1), 47-55; (b) Carbonell-Barrachina, A. A.; Agusti, A.; Ruiz, J. J., Analysis of flavor volatile compounds by dynamic headspace in traditional and hybrid cultivars of Spanish tomatoes. *Eur. Food Res. Technol.* **2006,** *222* (5-6), 536-542; (c) Mayer, F.; Takeoka, G. R.; Buttery, R. G.; Whitehand, L. C.; Naim, M.; Rabinowitch, H. D., Studies on the aroma of five fresh tomato cultivars and the precursors of cis- and trans-4,5-epoxy-(E)-2-decenals and methional. *J. Agric. Food Chem.* **2008,** *56* (10), 3749-3757; (d) Ruiz, J. J.; Alonso, A.; Garcia-Martinez, S.; Valero, M.; Blasco, P.; Ruiz-Bevia, F., Quantitative analysis of flavour volatiles detects differences among closely related traditional cultivars of tomato. *J. Sci. Food Agric.* **2005,** *85* (1), 54-60.

19. (a) Cavaliere, B.; De Nino, A.; Hayet, F.; Lazez, A.; Macchione, B.; Moncef, C.; Perri, E.; Sindona, G.; Tagarelli, A., A metabolomic approach to the evaluation of the origin of extra virgin olive oil: A convenient statistical treatment of mass spectrometric analytical data. *J. Agric. Food Chem.* **2007,** *55* (4), 1454-1462; (b) Russo, A.; Caputo, S.; Pantusa, M.; Perri, E.; Sindona, G.; Sportelli, L., Amino acids as modulators of lipoxygenase oxidation mechanism. The identification and structural characterization of spin adducts intermediates by electron spin resonance and tandem mass spectrometry. *Food Chem.* **2010,** *119* (2), 533-538.

20. Gray, D. A.; Prestage, S.; Linforth, R. S. T.; Taylor, A. J., Fresh tomato specific fluctuations in the composition of lipoxygenase-generated C6 aldehydes. *Food Chem.* **1999,** *64* (2), 149-155.

21. Buttery, R. G.; Ling, L. C., Volatile components of tomato fruit and plant parts:Relationship and biogenesis. In *Bioactive Volatile Compounds from Plants: ACS Symposium Series*, American Chemical Society: Washington, DC, 1993; Vol. 525, p 23−34.

22. Stanimirova, I.; Ustun, B.; Cajka, T.; Riddelova, K.; Hajslova, J.; Buydens, L. M. C.; Walczak, B., Tracing the geographical origin of honeys based on volatile compounds profiles assessment using pattern recognition techniques. *Food Chem.* **2010,** *118* (1), 171-176.

23. Corea, G.; Fattorusso, E.; Lanzotti, V.; Capasso, R.; Izzo, A. A., Antispasmodic saponins from bulbs of red onion, Allium cepa L. var. Tropea. *J. Agric. Food Chem.* **2005,** *53* (4), 935-940.

24. (a) Okamoto, T., Safety of quercetin for clinical application (Review). *Int J Mol Med* **2005,** *16* (2), 275-278; (b) Middleton, E.; Kandaswami, C.; Theoharides, T. C., The effects of plant flavonoids on mammalian cells: Implications for inflammation, heart disease, and cancer. *Pharmacol Rev* **2000,** *52* (4), 673-751.

25. Slimestad, R.; Fossen, T.; Vagen, I. M., Onions: A source of unique dietary flavonoids. *J. Agric. Food Chem.* **2007,** *55* (25), 10067-10080.

26. (a) Ariyama, K.; Horita, H.; Yasui, A., Application of inorganic element ratios to chemometrics for determination of the geographic origin of Welsh onions. *J. Agric. Food Chem.* **2004,** *52* (19), 5803-5809; (b) Ariyama, K.; Horita, H.; Yasui, A., Chemometric techniques on inorganic elements composition for the determination of the geographic origin of Welsh onions. *Anal. Sci.* **2004,** *20* (5), 871-877.

27. Ariyama, K.; Aoyama, Y.; Mochizuki, A.; Homura, Y.; Kadokura, M.; Yasui, A., Determination of the geographic origin of onions between three main production areas in Japan and other countries by mineral composition. *J. Agric. Food Chem.* **2007,** *55* (2), 347-354.

28. Ariyama, K.; Nishida, T.; Noda, T.; Kadokura, M.; Yasui, A., Effects of fertilization, crop year, variety, and provenance factors on mineral concentrations in onions. *J. Agric. Food Chem.* **2006,** *54* (9), 3341-3350.

29. Lo Feudo, G.; Naccarato, A.; Sindona, G.; Tagarelli, A., Investigating the Origin of Tomatoes and Triple Concentrated Tomato Pastes through Multielement Determination by Inductively Coupled Plasma Mass Spectrometry and Statistical Analysis. *J. Agric. Food Chem.* **2010,** *58* (6), 3801-3807.

30. Council Regulation (EEC) No 2081/92 (1992). On the protection of geographical indications and designations of origin for agricultural products and foodstuffs. Official Journal, L 208 (Corrigenda:[Official Journal L 27, 30.01.1997 and L 53,

24.02.1998]. Amended by Council Regulation (EC) No 535/97of 17 March 1997 (Official Journal L 83, 25.03.1997).

31. (a) Benavente-Garcia, O.; Castillo, J., Update on uses and properties of Citrus flavonolds: New findings in anticancer, cardiovascular, and anti-inflammatory activity. *J. Agric. Food Chem.* **2008,** *56* (15), 6185-6205; (b) Benavente-Garcia, O.; Castillo, J.; Marin, F. R.; Ortuno, A.; Del Rio, J. A., Uses and properties of Citrus flavonoids. *J. Agric. Food Chem.* **1997,** *45* (12), 4505-4515; (c) Gardner, P. T.; White, T. A. C.; McPhail, D. B.; Duthie, G. G., The relative contributions of vitamin C, carotenoids and phenolics to the antioxidant potential of fruit juices. *Food Chem.* **2000,** *68* (4), 471-474; (d) Sanchez-Moreno, C.; Plaza, L.; de Ancos, B.; Cano, M. P., Quantitative bioactive compounds assessment and their relative contribution to the antioxidant capacity of commercial orange juices. *J. Sci. Food Agric.* **2003,** *83* (5), 430-439; (e) Dhuique-Mayer, C.; Caris-Veyrat, C.; Ollitrault, P.; Curk, F.; Amiot, M. J., Varietal and interspecific influence on micronutrient contents in citrus from the Mediterranean area. *J. Agric. Food Chem.* **2005,** *53* (6), 2140-2145.

32. (a) McHard, J. A.; Bayer, S.; Winefordner, J. D., Determination of the geographical origins of frozen concentrated orange juice via pattern recognition. *J. Agric. Food Chem.* **1980,** *28* (6), 1306-7; (b) McHard, J. A.; Foulk, S. J.; Winefordner, J. D., A comparison of trace element contents of Florida and Brazil orange juice. *J. Agric. Food Chem.* **1979,** *27* (6), 1326-8.

33. Martin, G. J.; Fournier, J. B.; Allain, P.; Mauras, Y.; Aguile, L., Optimization of analytical methods for origin assessment of orange juices .2. ICP-MS determination of trace and ultra-trace elements. *Analusis* **1997,** *25* (1), 7-13.

34. (a) Alonso-Salces, R. M.; Moreno-Rojas, J. M.; Holland, M. V.; Reniero, F.; Guillou, C.; Héberger, K. r., Virgin Olive Oil Authentication by Multivariate Analyses of 1H NMR Fingerprints and δ13C and δ2H Data. *J. Agric. Food Chem.* **2010,** *58* (9), 5586-5596; (b) Alonso-Salces, R. M.; Héberger, K.; Holland, M. V.; Moreno-Rojas, J. M.; Mariani, C.; Bellan, G.; Reniero, F.; Guillou, C., Multivariate analysis of NMR fingerprint of the unsaponifiable fraction of virgin olive oils for authentication purposes. *Food Chem.* **2010,** *118* (4), 956-965.

35. Liu, F.; He, Y.; Wang, L., Determination of effective wavelengths for discrimination of fruit vinegars using near infrared spectroscopy and multivariate analysis. *Anal. Chim. Acta* **2008,** *615* (1), 10-17.

36. (a) Forina, M.; Lanteri, S.; Casale, A., Multivariate calibration. *Journal of Chromatography A* **2007,** *1158* (1-2), 61-93; (b) Marengo, E.; Robotti, E.; Bobba, M.; Milli, A.; Campostrini, N.; Righetti, S. C.; Cecconi, D.; Righetti, P. G., Application of partial least squares discriminant analysis and variable selection procedures: a 2D-PAGE proteomic study. *Anal. Bioanal. Chem.* **2008,** *390* (5), 1327-1342.

37. Oddone, M.; Aceto, M.; Baldizzone, M.; Musso, D.; Osella, D., Authentication and Traceability Study of Hazelnuts from Piedmont, Italy. *Journal of Agricultural and Food Chemistry* **2009,** *57* (9), 3404-3408.

# Chapter 4

## *Experimental*

### *Olive Oil*

*Secondary metabolites of Olea europaea leaves as markers for the discrimination of cultivars and cultivation zones by multivariate analysis.*

### *Tomato*

*Investigating the Origin of Tomatoes and Triple Concentrated Tomato Pastes through Multielement Determination by Inductively Coupled Plasma Mass Spectrometry and Statistical Analysis*

*The volatile fraction profiling of fresh tomatoes and triple concentrate tomato pastes as parameter for the determination of geographical origin*

### *Tropea red Onion*

*Multielement Fingerprinting as a Tool in Origin Authentication of PGI Food Products: Tropea Red Onion*

### *Clementine di Calabria*

*Authenticity of PGI Clementine of Calabria by multielement fingerprint*

## 4.1 Secondary metabolites of *Olea europaea* leaves as markers for the discrimination of cultivars and cultivation zones by multivariate analysis

### 4.1.1 Chemicals and Instrumentation

HPLC grade solvents were purchased from Carlo Erba (Rodano, Italy). Salicin was obtained from Sigma–Aldrich (Saint Louis, MO).

The collected fractions containing each compound were obtained using a fractionlynx semi-preparative HPLC system (Waters Corporation, Milford, MA, USA) composed of a autosampler/collector Waters 2767 sample manager, a 600E pump working in analytical mode, a 486 UV detector and a ZMD mass spectrometer equipped with an ESI source. The separation was performed using a 250×4.6mm 5mm reversed-phase C18 Luna-Phenomenex column at a flow rate of 1 mL/min. The run time was 70 min and the gradient was built using 5mM $NH_4^+CH_3COO^-$ in $H_2O$ (solvent A) and acetonitrile (ACN) (solvent B) as eluting phase. The solvent run was composed by the following steps: isocratic 90%A for 1min; linear gradient from 90% A to 50% A in 14min; isocratic 50% A for 8min; linear gradient from 50% A to 0%A in 10 min; isocratic 0% A for 5min; linear gradient 0% A to 90% A in 5 min; equilibration of the column for 10 min. The MS conditions were as follows: Capillary voltage 3.15 kV, cone voltage 7V, extractor 2V, RF lens 0.34 V, source block and desolvation temperature 120, 2508C, respectively, ion energy 0.5V, LM resolution 14.5, HM resolution 15.0 and multiplier 650 V. The nebulizer gas was set to 650 L/h. The samples were collected at the exhaust of the UV detector at the same time as the appearance of the mass signal.

The high-resolution ESI experiments were carried out in a hybrid Q-Star Pulsar-i (MDS Sciex Applied Biosystems, Toronto, Canada) mass spectrometer equipped with an ion

spray ionization source. Samples were introduced by direct infusion (5mL/min) of the solution coming from the HPLC separation at the optimum ion spray voltage of 4800 V. The nitrogen gas flow was set at 30 psi and the declustering and the focusing potentials were kept at 70 and 140V relative to ground, respectively. $MS^2$ experiments were performed in the collision cell *q* on the isotopically pure ($^{12}C$) peak of the selected precursor ions by keeping the first quadrupole analyzer at unit resolution, and scanning the time-of-flight (TOF) analyzer. The collision energy was set to 15 eV, for each compound, while the gas pressure of the collision chamber was regulated at the instrumental parameters CAD 5, which corresponds to a pressure of the chamber of $6.86\times10^{-3}$ Torr and a gas thickness of $9.55\times10^{-15}$ molecules/$cm^2$. All the acquisitions were averaged over 60 scans at a TOF resolving power of 8000. The molecular formula was evaluated by means of Analyst$^{TM}$ QS software (MDS-Sciex).

### 4.1.2 Sampling

Samples of olive leaves, utilized in this study, were collected from five different cultivars (Carolea, Cassanese, Coratina, Nocellara del Belice and Leccino), in the same experimental field of the C.R.A. Istituto Sperimentale per l'Olivicoltura in Rende (Calabria, Italy) in March and April 2006. Five samples of Carolea leaves were also collected in July 2006 and January 2007. Finally, five samples of Carolea leaves were picked from different cultivation areas, Rende and Mirto (Calabria) and Spoleto (Umbria, Italy), in July 2006. All the collected samples are summed in Table 4.1.

| Number of samples | Origin | Region | Cultivar | Harvest period |
|---|---|---|---|---|
| 5 | Rende | Calabria | Coratina | March 2006 |
| 5 | Rende | Calabria | Coratina | April 2006 |
| 5 | Rende | Calabria | Leccino | March 2006 |
| 5 | Rende | Calabria | Leccino | April 2006 |
| 5 | Rende | Calabria | Cassanese | March 2006 |
| 5 | Rende | Calabria | Cassanese | April 2006 |
| 5 | Rende | Calabria | Nocellara | March 2006 |
| 5 | Rende | Calabria | Nocellara | April 2006 |
| 5 | Rende | Calabria | Carolea | March 2006 |
| 5 | Rende | Calabria | Carolea | April 2006 |
| 5 | Rende | Calabria | Carolea | July 2006 |
| 5 | Mirto | Calabria | Carolea | July 2006 |
| 5 | Spoleto | Umbria | Carolea | July 2006 |

**Table 4.1** Olive leaf samples

### 4.1.3 Sample preparation

A mixture of methanol and water (20 mL; 1:1 v/v) was added to 2 g of olive leaf powder. Five cultivars (Carolea, Cassanese, Coratina, Nocellara del Belice and Leccino) were used for the assays. The mixture was homogenized by vortex for 3 min and subsequently sonicated for 20 min. The residual solution was filtered into a Buchner funnel, and diluted to 25 mL; 2mL of the solution were filtered on a 0.45 micron filter and submitted to HPLC analysis.

### 4.1.4 Statistical analysis

Principal component analysis (PCA) was performed by Statistica 8.0 (StatSoft 2007 Edition) and linear discriminant analysis (LDA), soft independent modelling of class analogy (SIMCA) and K-nearest neighbours (KNN) were executed by V-Parvus 84 2004.[1]

## 4.2 Investigating the Origin of Tomatoes and Triple Concentrated Tomato Pastes through Multielement Determination by Inductively Coupled Plasma Mass Spectrometry and Statistical Analysis

### 4.2.1 Chemicals and Instrumentation

The reagents used for mineralization ($HNO_3$ (65%), $H_2O_2$ (30%), HCl (30%), HF (40%), $HClO_4$ (70%), $H_3BO_3$) were Suprapur (Merck, Darmstadt, Germany).All other reagents used for analysis were of analytical reagent grade (Merck, Darmstadt, Germany). Two multielement solutions of Ag, Al, As, Ba, Be, Bi, Ca, Cd, Co, Cr, Cs, Cu, Fe, Ga, In, K, Li, Mg, Mn, Na, Ni, Pb, Rb, Se, Sr, Tl, V,U and Zn (100 mg/L, Merck) and Ce, Dy, Er, Eu, Gd, Ho, La, Lu, Nd, Pr, Sm, Sc, Tb, Th, Tm, Y and Yb (10 mg/L, PerkinElmer) were used to prepare the calibration standards. Aqueous solutions were prepared using ultrapure water, with a resistivity of 18.2 MΩcm, obtained from a Milli-Q plus system (Millipore, Bedford, MA). All glassware was decontaminated with nitric acid (2%, v/v) overnight, rinsed with ultrapure water and dried. The accuracy of the method was evaluated by analyzing the certified reference material NCS ZC85006 Tomato (China National Analysis Center for Iron & Steel 2000).

The sample preparation was carried out using the following system for microwave digestion: Anton Paar Multiwave 3000 with programmable power control (maximum power 1400 W) and rotor XF100 (operating pressure up to 120 bar maximum; operating temperature 260°C maximum; construction material PTFE-TFM for the vessel). The determination of the elements of interest was carried out utilizing an Elan DRC-e ICP-MS instrument (Perkin-Elmer SCIEX, Canada). Samples were introduced by means of a cross flow quartz nebulizer with a Scott-type spray chamber. The ICP torch was a standard torch (Fassel type torch) with platinum injector. A solution containing Rh, Mg, Pb, Ba

and Ce (10 μg/L) was used to optimize the instrument in terms of sensitivity, resolution and mass calibration. The $^{140}Ce^{16}O^{+}$ /$^{140}Ce^{+}$ ratio was used to check the level of oxide ions in the plasma that could interfere in the determination of some elements; also, instrumental parameters such as RF power and carrier gas flow were optimized and the level of doubly charged ion monitored by means of the signal $^{137}Ba^{2+}/^{137}B^{+}$. The $^{140}Ce^{16}O^{+}$ /$^{140}Ce^{+}$ and $^{137}Ba^{2+}/^{137}B^{+}$ ratios obtained after optimization were 2.8 and 3.2%, respectively. ICP-MS analyses were performed following the operating program and parameters shown in Table 4.2.

| | |
|---|---|
| **rf power (W)** | 1130 |
| **Nebulizer (carrier gas) flow rate (L min$^{-1}$)** | 0.95 |
| **Lens voltage (V)** | 6.25 |
| **Analog stage voltage (V)** | -1900 |
| **Pulse stage voltage (V)** | 1050 |
| **Discriminator threshold (V)** | 70 |
| **Quadrupole rod offset (V)** | 0 |
| **Resolution (amu)** | 0.70 |
| **Detector** | Dual |
| **Speed of peristaltic pump (rpm)** | 24 |
| **Sweeps/reading** | 20 |
| **Replicates** | 3 |
| **Dwell time** | 50 ms |
| **Scan mode** | Peak hopping |
| **DRC Parameters** | |
| **CH4 reaction gas flow (ml/min)** | 0.70 for Cr, Fe Se and Zn<br>1.15 for Sc, Ni and Eu |
| **Rejection parameter a (RPa)** | 0 |
| **Rejection parameter q (RPq)** | 0.5 for Eu<br>0.6 for Sc, Ni, Se and Zn<br>0.7 for Cr and Fe |

**Table 4.2** Instrumental Parameters and Operating Conditions for the ICP-MS Instrument

### 4.2.2 Sampling

Italian and foreign samples of certain origin were provided by Istituto Nazionale delle Conserve Alimentari (INCA). A total of about 130 tomato fruits of cv. Perfectpeel were hand harvested in August 2007 and 2008 from four different Italian regions (Calabria, Basilicata, Puglia and Emilia Romagna) and immediately stored at -20°C. One hundred samples (500 g for each samples) of triple concentrated tomato paste, 50 Italian, 10 Chinese, 20 Californian and 20 Greek, were screened. Italian tomato paste samples were obtained from the same producers of the tomato samples (Table 4.3).

| Zone | Region | Tomato | | Paste | |
|---|---|---|---|---|---|
| | | 2007[a] | 2008[a] | 2007[b] | 2008[b] |
| Mesagne | Puglia (Italy) | 10 | 10 | - | - |
| Collecchio | Emilia Romagna (Italy) | 10 | 20 | 10 | 10 |
| Matera | Basilicata (Italy) | 10 | 10 | - | - |
| Crotone | Calabria (Italy) | 20 | 20 | 20 | 10 |
| Unknown | (China) | - | - | - | 10 |
| Unknown | (California) | - | - | - | 20 |
| Unknown | (Greece) | - | - | - | 20 |

**Table 4.3** Number of tomato and triple concentrated tomato paste samples ([a] harvest year; [b]production year)

### 4.2.3 Sample preparation

An aliquot of tomato sample (10 g) or triple concentrated tomato paste sample (1 g) was directly weighted into the vessel of the microwave system. The digestion was performed by adding 4.5 mL of $HNO_3$, 1 mL of $H_2O_2$ and 0.5 mL of HF to each sample. The operating conditions used for the microwave digestion system is shown were 1000Wover ten minutes and hold at this power for eight minutes. After digestion the extracts were quantitatively transferred to a graduated polypropylene test-tube and made up to volume

(50 mL) with ultrapure water. The analytical batch consisted of a set of calibration standards, that were analyzed at the beginning of the run, samples, a minimum of three procedural blanks, one procedural blank spiked with a solution containing the elements of interest and the certified reference material. A mid-range calibration standard was measured at the end of each analytical run, in order to assess instrumental drift throughout the run. A eight point calibration curves covering the range 0.1-2000 μg/L were used for quantitative analysis. Standard solutions were prepared by diluting the multielement solutions cited in section 4.2.1 Chemicals and Instrumentation.

### 4.2.4 Statistical analysis

LDA was performed by Statistica 8.0 (StatSoft 2007 Edition); KNN and SIMCA were executed by V-Parvus 2004.[1]

## 4.3 The volatile fraction profiling of fresh tomatoes and triple concentrate tomato pastes as parameter for the determination of geographical origin.

### 4.3.1 Chemicals and Instrumentation

Calcium chloride 96% was purchased from Carlo Erba (Rodano, Italy). Methanol, ethyl isobutanoate, n-hexane and C7-C30 saturated alkanes were obtained from Sigma-Aldrich (Saint Louis, MO). The 85 μm carboxen/polydimethylsiloxane fiber for SPME analysis was supplied by Supelco (Bellefonte, PA). Sample analyses were performed using a Varian (Walnut Creek, CA) Saturn 2000 GC-MS ion trap (ITD) system in EI modes, coupled to a Varian 3800 gas chromatograph equipped with a Varian 8200 autoinjector. The ion trap temperature was set at 210 °C with an ionization time of 2 ms, a reaction time at 50 ms, and a scan rate at 1000 ms. The transfer line temperature was set at 230 °C. The column was a 30 m Zebron ZB-5ms low bleed (0.25 mm i.d., 0.25 μm film thickness). The gas chromatography oven temperature was initially held at 35 °C for 8 min, then increased at 1.5 °C/min to 45 °C, increased at 3 °C/min to 150 °C, increased again at 2.5 °C/min to 250 °C and held for 6 min. The carrier gas was helium at 1 mL/min. For SPME analyses, a narrow bore Supelco 0.8 mm i.d. GC inlet liner was used.

### 4.3.2 Sampling

Tomato and concentrate paste samples were provided by Istituto Nazionale delle Conserve Alimentari (I.N.C.A). Tomatoes of cultivar Perfectpeel were hand harvested in August 2007 from four different Italian regions (Calabria, Basilicata, Puglia and Emilia Romagna) and immediately stored at -20°C.

Harvest occurred when tomatoes providing a Color Stage 6 according with the United States Department of Agriculture (USDA) tomato ripeness color chart. Samples are presented in Table 4.4.

| Zone | Region | Tomato | Paste |
|---|---|---|---|
| | | 2007[a] | 2008[b] |
| Mesagne | Puglia | 10 | - |
| Collecchio | Emilia Romagna | 10 | 10 |
| Matera | Basilicata | 10 | - |
| Crotone | Calabria | 20 | 20 |

**Table 4.4** Number of tomato and triple concentrated tomato paste samples ([a] harvest year; [b]production year)

### 4.3.3 Sample preparation

Ten tomatoes were randomly chosen for each region and homogenized at 25°C using an electric mixer. Then two grams of homogenized tomatoes were put in a 10 mL vial and 2 mL of saturated $CaCl_2$ solution were added.[2] Then 80 µL of ethyl isobutanoate solution at 5 mg/L in methanol [3] were added as internal standard to obtain a concentration of 200 µg/kg in each sample. The mixture was homogenized and the vials was sealed. The same procedure was followed for the concentrate paste samples, but three grams were weighted and 120 µL of internal standard solution were added. All these steps were conducted before complete defrosting to avoid loss of volatile compounds.[4] Solid phase microextraction (SPME) technique in headspace mode was used for the extraction of volatile compounds. The sampling was performed in automated mode using the autoinjector equipped with the 85 µm carboxen/polydimethylsiloxane fiber.[4] The sample was preheating to 35°C for 30 min and the extraction of volatile compounds took place

for 60 minutes at the same temperature. The analytes were desorbed for 10 min from the fiber into the GC injector set at 300°C in splitless mode.

### 4.3.4 Statistical analysis

Classification was carried out by two multivariate chemometric techniques: Linear Discriminant Analysis (LDA) and Soft Independent Modeling of Class Analogy (SIMCA). LDA was performed by Statistica 7.1; SIMCA was executed by V-Parvus 84 2004.[1] Principal component analysis (PCA) was performed by Statistica 7.1 statistical package.

## 4.4 Multielement Fingerprinting as a Tool in Origin Authentication of PGI Food Products: Tropea Red Onion.

### 4.4.1 Chemicals and Instrumentation

The reagents used for mineralization, $HNO_3$ (65%) and $H_2O_2$ (30%), were Suprapur (Merck, Darmstadt, Germany). All other reagents used for analysis were of analytical reagent grade (Merck, Darmstadt, Germany). Two multielement solutions of Ag, Al, As, Ba, Be, Bi, Ca, Cd, Co, Cr, Cs, Cu, Fe, Ga, In, K, Li, Mg, Mn, Na, Ni, Pb, Rb, Se, Sr, Tl, V, U and Zn (100 mg/l, Merck) and Ce, Dy, Er, Eu, Gd, Ho, La, Lu, Nd, Pr, Sm, Sc, Tb, Th, Tm, Y and Yb (10 mg/l, PerkinElmer) were used to prepare the calibration standards. Aqueous solutions were prepared using ultrapure water, with a resistivity of 18.2 MΩ cm, obtained from a Milli-Q plus system (Millipore, Bedford, MA, USA). All glassware, polyethylene flasks and tubes involved in sample preparation and measurement process were cleaned with nitric acid (2%, v/v) overnight, rinsed with ultrapure water and dried.

An Anton Paar Multiwave 3000 with programmable power control (maximum power 1400 W) and rotor XF100 (operating pressure up to 120 bar maximum; operating temperature 260°C maximum; construction material PTFE-TFM for the liner) was used for the microwave digestion of the samples.

The determination of the elements of interest was carried out utilizing an Elan DRC-e ICP-MS instrument (Perkin-Elmer SCIEX, Canada) equipped with dynamic reaction cell for suppressing or reducing polyatomic interferences operating with $CH_4$ (99.996% purity) as reaction gas.

The sample delivery system consisted of a PerkinElmer auto sampler model AS-93 Plus with peristaltic pump and a cross-flow nebulizer with a Scott type spray chamber. Samples were introduced by means of a quartz nebulizer. The ICP torch was a standard

torch (Fassel type torch) with platinum injector. A solution containing Rh, Mg, Pb, Ba and Ce (10 μg/L) was used to optimize the instrument in terms of sensitivity, resolution and mass calibration. The $^{140}Ce^{16}O^{+}/^{140}Ce^{+}$ and $^{137}Ba^{2+}/^{137}Ba^{+}$ ratios were used to check respectively the level of oxide ions in the plasma and the level of doubly charged ion that could interfere in the determination of some elements. Moreover, instrumental parameters such as RF power and carrier gas flow were optimized. In order to assess the accuracy of the method, certified reference material NCS ZC85006 Tomato (China National Analysis Center for Iron & Steel 2000) was submitted to the whole analytical process. ICP-MS analyses were performed following the operating program and parameters shown in Table 4.5.

| | |
|---|---|
| **rf power (W)** | 1100 |
| **Nebulizer (carrier gas) flow rate (L min$^{-1}$)** | 0.90 |
| **Lens voltage (V)** | 8.5 |
| **Analog stage voltage (V)** | -1800 |
| **Pulse stage voltage (V)** | 1120 |
| **Discriminator threshold (V)** | 70 |
| **Quadrupole rod offset (V)** | 0 |
| **Resolution (amu)** | 0.70 |
| **Detector** | Dual |
| **Speed of peristaltic pump (rpm)** | 24 |
| **Sweeps/reading** | 20 |
| **Replicates** | 3 |
| **Dwell time** | 50 ms |
| **Scan mode** | Peak hopping |
| **DRC Parameters** | |
| **CH4 reaction gas flow (ml/min)** | 0.70 for Cr, Fe Se and Zn<br>1.15 for Sc, Ni and Eu |
| **Rejection parameter a (RPa)** | 0 |
| **Rejection parameter q (RPq)** | 0.5 for Eu<br>0.6 for Sc, Ni, Se and Zn<br>0.7 for Cr and Fe |

**Table 4.5** Instrumental Parameters and Operating Conditions for the ICP-MS Instrument

### 4.4.2 Sampling

Samples were provided by the Consortium for the Protection of Tropea Red Onion and were harvested during the crop season 2009 from different fields belonging to the municipalities of Capo Vaticano, Amantea, Nocera Terinese, Briatico and immediately stored at -20°C. Non-Tropea onion samples came from three Italian region (Campania, Sicilia, Piemonte) and Netherlands (Table 4.6).

| Cultivation zone | Region/State | Harvest time | Number of samples |
|---|---|---|---|
| Capo Vaticano | Calabria | May 2009 | 20 |
| Amantea | Calabria | May 2009 | 20 |
| Nocera Terinese | Calabria | June 2009 | 40 |
| Briatico | Calabria | July 2009 | 40 |
| Agrigento | Sicily | July 2009 | 20 |
| Salerno | Campania | June 2009 | 20 |
| - | Piedmont | June 2009 | 20 |
| - | Holland | June 2009 | 20 |

**Table 4.6** Number of Tropea and non-Tropea onion samples

### 4.4.3 Sample preparation

Ten onion were randomly chosen for each harvest lot. For each onion, outer tunic, leaves, and basal plate with roots were discarded. Only the bulb was homogenized at 25°C using an electric mixer. An aliquot of onion sample (2 g) was directly weighted into the liner of the microwave system. The digestion was performed by adding 3 ml of $HNO_3$ and 1 ml of $H_2O_2$ to each sample. The operating conditions used for the microwave digestion system are shown in Table 4.7. After digestion the extracts were quantitatively transferred to a graduated polypropylene test-tube and diluted with ultrapure water up to 50 ml. Ten

grams of homogenized sample were dried at 80°C to constant weight in order to obtained the moisture content for each harvested onion batch.

| Step | Power (W) | Hold (min) |
|---|---|---|
| 1 | 100 | 5 |
| 2 | 600 | 5 |
| 3 | 1000 | 10 |
| 4 | 0 | 15 |

**Table 4.7** Operating conditions used for the microwave digestion system

The analytical batch consisted of a set of calibration standards, that were analyzed at the beginning of the run, samples, a minimum of three procedural blanks and one procedural blank spiked with a solution containing the elements of interest. A mid-range calibration standard was measured at the end of each analytical run, in order to assess instrumental drift throughout the run. An eight point calibration curves covering the range 0.1-2000 µg/l were used for quantitative analysis. Standard solutions were prepared by diluting the multielement solutions cited in 4.4.1 Chemicals and instrumentation section.

### 4.4.4 Statistical analysis

Classification was carried out by three multivariate chemometric techniques: Artificial Neural Network (ANN), Linear Discriminant Analysis (LDA) and Soft Independent Modeling of Class Analogy (SIMCA). ANN and LDA were performed by Statistica 7.1 (StatSoft 2005 Edition); SIMCA was executed by V-Parvus 84 2004.[1] Before ANN construction, principal component analysis (PCA) was performed by Statistica 7.1 statistical package.

## 4.5 Authenticity of PGI Clementine of Calabria by multielement fingerprint.

### 4.5.1 Chemicals and Instrumentation

The mineralization was carried out using acids with Suprapur grade ($HNO_3$ (65%), $H_2O_2$ (30%), HCl (30%), HF (40%), $HClO_4$ (70%), $H_3BO_3$) (Merck, Darmstadt, Germany). All other reagents used for analysis were of analytical reagent grade (Merck, Darmstadt, Germany). Two multielement solutions of Ag, Al, As, Ba, Be, Bi, Ca, Cd, Co, Cr, Cs, Cu, Fe, Ga, In, K, Li, Mg, Mn, Na, Ni, Pb, Rb, Se, Sr, Tl, V, U and Zn (100 mg/L, Merck) and Ce, Dy, Er, Eu, Gd, Ho, La, Lu, Nd, Pr, Sm, Sc, Tb, Th, Tm, Y and Yb (10 mg/L, PerkinElmer) were used for the preparation of aqueous calibration standard solutions after appropriate dilution. All glassware, polyethylene flasks, squeezer and tubes involved in sample preparation and measurement process was cleaned with nitric acid (2%, v/v) by soaking overnight and rinsed with ultrapure water prior to use. Aqueous solutions were prepared using ultrapure water, with a resistivity of 18.2MΩ cm, obt ained from a Milli-Q plus system (Millipore, Bedford, MA, USA).

The sample preparation was carried out using the following system for the microwave digestion: Anton Paar Multiwave 3000 with programmable power control (maximum power 1400 W) and rotor XF100 (operating pressure up to 120 bar maximum; operating temperature 260°C maximum; construction material PTFE-TFM for the liner and the seal). The elements determination was carried out utilizing an Elan DRC-e ICP-MS instrument (Perkin-Elmer SCIEX, Canada). The sample delivery system consisted of a PerkinElmer auto sampler model AS-93 Plus with peristaltic pump and a cross-flow nebulizer with a Scott type spray chamber. The ICP torch was a standard torch (Fassel type torch) with platinum injector.

A solution containing Rh, Mg, Pb, Ba and Ce (10 µg/L, Merck, Darmstadt, Germany) was used to optimize the instrument in terms of sensitivity, resolution and mass calibration.

A full optimization procedure was carried out before the analysis of calibration solutions and samples. The $^{140}Ce^{16}O^{+}/^{140}Ce^{+}$ ratio was used to check the level of oxide ions in the plasma that could interfere in the determination of some elements; also, instrumental parameters such as RF power and carrier gas flow were optimized and the level of doubly charged ion monitored by means of the signal $^{137}Ba^{2+}/^{137}Ba^{+}$. In order to obtain a better accuracy and extension of linear dynamic range a dual detector cross calibration was performed using a reference solution 200 µg/L prepared by diluting stock solutions XXI (Merck, Darmstadt, Germany) containing 10 mg/L of some elements considered in the analytical method with ultrapure water. The operating conditions and parameters of ICP-MS analyses are shown in Table 4.7.

| | |
|---|---|
| **rf power (W)** | 1100 |
| **Nebulizer (carrier gas) flow rate (L min$^{-1}$)** | 0.90 |
| **Lens voltage (V)** | 8.5 |
| **Analog stage voltage (V)** | -1800 |
| **Pulse stage voltage (V)** | 1120 |
| **Discriminator threshold (V)** | 70 |
| **Quadrupole rod offset (V)** | 0 |
| **Resolution (amu)** | 0.70 |
| **Detector** | Dual |
| **Speed of peristaltic pump (rpm)** | 24 |
| **Sweeps/reading** | 20 |
| **Replicates** | 3 |
| **Dwell time** | 50 ms |
| **Scan mode** | Peak hopping |
| **DRC Parameters** | |
| **CH4 reaction gas flow (ml/min)** | 0.70 for Cr, Fe Se and Zn<br>1.15 for Sc, Ni and Eu |
| **Rejection parameter a (RPa)** | 0 |
| **Rejection parameter q (RPq)** | 0.5 for Eu<br>0.6 for Sc, Ni, Se and Zn<br>0.7 for Cr and Fe |

**Table 4.8** Instrumental parameters and operating conditions for the ICP/MS instrument

### 4.5.2 Sampling

Clementine samples with PGI brand came from four different Calabrian cultivation zones located in the municipalities of Corigliano Calabro, Lamezia Terme, Pizzo Calabro and Rosarno. These samples were provided by six farms and hand harvested in October, November and December 2007. For each harvesting month three significant samples were randomly chosen for each farm so a total of 54 samples were collected. Non-PGI clementine samples came from Spain, Tunisia and Algeria (Table 4.9). All samples were immediately stored at -20°C.

| Cultivation zone | Region/State | Juice samples | Peel samples |
|---|---|---|---|
| Corigliano Calabro | Calabria | 18 | 18 |
| Lamezia Terme | Calabria | 9 | 9 |
| Pizzo Calabro | Calabria | 9 | 9 |
| Rosarno | Calabria | 18 | 18 |
| Algiers | Algeria | 10 | 10 |
| Blida | Algeria | 4 | 4 |
| Valencia | Spain | 8 | 8 |
| - | Tunisia | 12 | 12 |

**Table 4.9** Number of PGI and non-PGI clementine samples

### 4.5.3 Sample preparation

The analytical batch consisted of a set of calibration standards, samples, a procedural blank for each mineralization batch and one procedural blank spiked with a solution containing the elements of interest. A mid-range calibration standard was measured at the end of each analytical run in order to assess instrumental drift throughout the run. The quantitative determination of elements was carried out with external standards. Ten point calibration curves covering the range 0.1–2000 µg/L were used. Standard solutions were prepared by diluting the multielement solutions cited in Section 2.1. The concentration range for the elements Ce, Dy, Er, Eu, Gd, Ho, La, Lu, Nd, Pr, Sc, Sm, Tb, Th, Tm, U, Y, Yb was 0.1–150 µg/L whereas the concentration range for the elements Ag, Al, As, Ba, Be, Bi, Ca, Cd, Co, Cr, Cs, Cu, Fe, Ga, In, K, Li, Mg, Mn, Na, Ni, Pb, Rb, Se, Sr, Tl, V, Zn was 0.1–2000 µg/L.

*Juice samples*

Clementines were thoroughly washed with tap water and rinsed with ultrapure water. Clementine juice was obtained by hand squeezing using a plastic squeezer to prevent

metal contamination. Each clementine has constituted a sample and was squeezed separately paying attentions to obtain the juice from only the edible part of the fruit without including the albedo. For each cultivation zone and harvesting month three replicates were used in the quantitative analysis. An aliquot of clementine juice (5 g) at their natural Brix values was weighted directly into the PTFE-TFM digestion tube of the microwave system. Digestion was performed by adding 2.5 mL of $HNO_3$ to each sample. The operating conditions used for the microwave digestion system are shown in Table 4.9.

| | | Juice samples | Peel samples |
|---|---|---|---|
| **Step** | **Power (W)** | **Time (min)** | |
| **1** | 800 | 10:00 | 15:00 |
| **2** | 0 | 10:00 | 10:00 |
| **3** | 900 | 10:00 | 15:00 |
| **4** | 0 | 30:00 | 35:00 |

**Table 4.10** Mineralization power programs used for the microwave digestion of clementine juices and peels samples.

After mineralization process, extracts were quantitatively transferred to a graduated polypropylene test-tube and diluted with ultrapure water up to 50 mL. Blank samples were prepared by subjecting 5 mL of ultrapure water to the same digestion procedure used for juice samples.

*Peel samples*

Peel samples were prepared from the same clementine used for the preparations of the juice samples. Each clementine peel (albedo and flavedo) was grated on a plastic kitchen

grater to shred the peel without metal contamination. For each zone and harvesting month, three clementines were used in quantitative analysis as for juice samples. An aliquot of shredded Clementine peel (300 mg) was weighted directly into the PTFE-TFM digestion tube of the microwave system. Digestion was performed by adding 2 mL of $HNO_3$ and 4 mL of ultrapure water to each sample. The digestion was carried out using the microwave power program shown in Table 4.9. Digested samples were quantitatively transferred to a graduated polypropylene test-tube and the volume adjusted to 50 mL with ultrapure water. Blank samples were prepared in a similar way as juice samples using the peel microwave conditions of peel digestion.

### 4.5.4 Statistical analysis

Principal component analysis (PCA) was performed by Statistica 7.1 statistical package. Classification was carried out by three multivariate chemometric techniques: Linear Discriminant Analysis (LDA), Soft Independent Modeling of Class Analogy (SIMCA) and Partial Least Squares-Discriminant Analysis (PLS-DA). LDA were performed by Statistica 7.1 (StatSoft 2005 Edition), SIMCA was executed by V-Parvus 2009[1] whereas the PLS-DA algorithms was supported by the software packages "The Unscrambler 9.1" (Camo Process As., Oslo, Norway).

## References

1. Forina, M.; Lanteri, S.; Armanino, C.; Cerrato Oliveros, C.; Casolino, C. V-Parvus (2004) Department of Chimica e Tecnologie Farmaceutiche e Alimentari, University of Genova, Genova, Italy. (Free download at www.parvus.unige.it).

2. Buttery, R. G.; Teranishi, R.; Ling, L. C., Fresh tomato aroma volatiles - a quantitative study. *J. Agric. Food Chem.* **1987,** *35* (4), 540-544.

3. Cavaliere, B.; De Nino, A.; Hayet, F.; Lazez, A.; Macchione, B.; Moncef, C.; Perri, E.; Sindona, G.; Tagarelli, A., A metabolomic approach to the evaluation of the origin of extra virgin olive oil: A convenient statistical treatment of mass spectrometric analytical data. *J. Agric. Food Chem.* **2007,** *55* (4), 1454-1462.

4. Beltran, J.; Serrano, E.; Lopez, F. J.; Peruga, A.; Valcarcel, M.; Rosello, S., Comparison of two quantitative GC-MS methods for analysis of tomato aroma based on purge-and-trap and on solid-phase microextraction. *Anal. Bioanal. Chem.* **2006,** *385* (7), 1255-1264.

# Appendix

*Abstracts of published papers*

# A.1

Food Chemistry 121 (2010) 492–496

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

## Secondary metabolites of *Olea europaea* leaves as markers for the discrimination of cultivars and cultivation zones by multivariate analysis

Leonardo Di Donna, Fabio Mazzotti, Attilio Naccarato, Raffaele Salerno, Antonio Tagarelli *, Domenico Taverna, Giovanni Sindona

*Dipartimento di Chimica, Università della Calabria, Via P. Bucci, cubo 12/C, I-87030 Arcavacata di Rende (CS), Italy*

ARTICLE INFO

*Article history:*
Received 29 July 2009
Received in revised form 18 December 2009
Accepted 22 December 2009

*Keywords:*
Olive phenols
Tandem mass spectrometry
Olive tree cultivars
Chemometrics
Classification

ABSTRACT

New phenolic compounds from *Olea europaea*, identified by high performance liquid chromatography/electrospray ionization tandem mass spectrometry, are suitable markers for differentiation between different varieties of olive trees cultivated in the same geographical area (Rende, Italy). Five cultivars (Carolea, Cassanese, Coratina, Nocellara and Leccino) were considered for the discrimination. Samples of Carolea, cultivated in three different geographical zones (Rende, Mirto and Spoleto, Italy), were as well checked to evaluate possible differences. Three supervised pattern recognition procedures, linear discriminant analysis (LDA), soft independent modelling of class analogy (SIMCA) and *K*-nearest neighbours (*K*NN) were used to classify samples in five groups corresponding to the five cultivars and in three groups corresponding to the three areas of production. The results show that *K*NN provides a model unable to predict a proper assignment of the cultivar, at least for those olive trees considered in this work, whereas LDA and SIMCA allow the achievement of good percentage of prediction for the cultivars as well as cultivation zones.

# A.2

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY ARTICLE

*J. Agric. Food Chem.* **2010,** *58,* 3801–3807 **3801**
DOI:10.1021/jf903868j

## Investigating the Origin of Tomatoes and Triple Concentrated Tomato Pastes through Multielement Determination by Inductively Coupled Plasma Mass Spectrometry and Statistical Analysis

Gabriella Lo Feudo,[†] Attilio Naccarato,[‡] Giovanni Sindona,*[,‡] and Antonio Tagarelli[‡]

[†]INCA Istituto Nazionale delle Conserve Alimentari—Sede di Cosenza, Via N. Parisio 13, I-87100 Cosenza, Italy, and [‡]Dipartimento di Chimica, Università della Calabria, Via P. Bucci Cubo 12/C, I-87030 Arcavacata di Rende (CS), Italy

The concentration of 32 elements (Al, As, Ba, Be, Ca, Cd, Ce, Cu, Dy, Fe, K, La, Lu, Mg, Mn, Na, Nd, Pb, Rb, Sm, Sr, Th, U, V, Zn) was determined in tomatoes harvested in different four Italian regions and in triple concentrated tomato paste samples coming from Italy, China, Greece and California. The resulting multielement profiles were processed using three chemometric techniques to evaluate the possibility of discrimination between different cultivation areas. The closed-vessel microwave digested samples were diluted and analyzed by DRC-ICP-MS with $CH_4$ as reaction gas. The accuracy of the proposed method was considered acceptable (values in the range 75−120%) for 25 out of the 35 elements of the reference material NCS ZC85006 Tomato. The origin of tomato fruits and the areas of production as “Italy” and “non-Italy” of the triple concentrated pastes were evaluated by three supervised pattern recognition procedures, linear discriminant analysis (LDA), soft independent modeling of class analogy (SIMCA) and *K*-nearest neighbors (KNN).

# A.3

Food Research International 44 (2011) 781–788

Contents lists available at ScienceDirect

Food Research International

journal homepage: www.elsevier.com/locate/foodres

## The volatile fraction profiling of fresh tomatoes and triple concentrate tomato pastes as parameter for the determination of geographical origin

Gabriella Lo Feudo [a], Barbara Macchione [b], Attilio Naccarato [b], Giovanni Sindona [b], Antonio Tagarelli [b,*]

[a] *I.N.C.A. Istituto Nazionale delle Conserve Alimentari–Sede di Cosenza, Via N. Parisio 13, I-87100 Cosenza, Italy*
[b] *Dipartimento di Chimica, Università della Calabria, Via P. Bucci Cubo 12/C, I-87030 Arcavacata di Rende (CS), Italy*

ARTICLE INFO

*Article history:*
Received 18 October 2010
Accepted 8 January 2011

*Keywords:*
Tomato
Volatile compounds
SPME
GC-MS
Authenticity
Statistical analysis
Classification

ABSTRACT

Two unsupervised pattern recognition techniques such as stepwise linear discriminant analysis (SLDA) and Soft Independent Modeling of Class Analogy (SIMCA) were used to classify tomato samples in categories corresponding to the cultivation areas. The same approach was used for triple concentrated pastes for discrimination between two different Italian production areas. Accordingly, HS-SPME-GC-MS with 85 μm carboxen/polydimethylsiloxane fiber was used for the determination of the volatile fraction in tomatoes and triple concentrate tomato pastes samples. Ethyl isobututanoate was used as internal standard for semiquantitative analysis and the concentration data (μg/kg) of 38 compounds for tomatoes and of 32 compounds for triple concentrates were used in following chemometric analysis. Sixteen and three variables were selected by forward stepwise LDA for tomatoes and pastes, respectively. SLDA and SIMCA models showed respectively 96% and 94% in term of prediction ability for tomatoes. The two supervised techniques provided 100% and 97% in prediction of the production areas of tomato pastes, respectively.

# A.4

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

ARTICLE

pubs.acs.org/JAFC

## Multielement Fingerprinting as a Tool in Origin Authentication of PGI Food Products: Tropea Red Onion

Emilia Furia, Attilio Naccarato, Giovanni Sindona, Gaetano Stabile, and Antonio Tagarelli*

Dipartimento di Chimica, Università della Calabria, Via P. Bucci Cubo 12/C, I-87030 Arcavacata di Rende (CS), Italy

Supporting Information

**ABSTRACT:** Tropea red onion (*Allium cepa* L. var. Tropea) is among the most highly appreciated Italian products. It is cultivated in specific areas of Calabria and, due to its characteristics, was recently awarded with the protected geographical indications (PGI) certification from the European Union. A reliable classification of onion samples in groups corresponding to "Tropea" and "non-Tropea" categories is now available to the producers. This important goal has been achieved through the evaluation of three supervised chemometric approaches. Onion samples with PGI brand (120) and onion samples not cultivated following the production regulations (80) were digested by a closed-vessel microwave oven system. ICP-MS equipped with a dynamic reaction cell was used to determine the concentrations of 25 elements (Al, Ba, Ca, Cd, Ce, Cr, Dy, Eu, Fe, Ga, Gd, Ho, La, Mg, Mn, Na, Nd, Ni, Pr, Rb, Sm, Sr, Tl, Y, and Zn). The multielement fingerprint was processed using linear discriminant analysis (LDA) (standard and stepwise), soft independent modeling of class analogy (SIMCA), and back-propagation artificial neural network (BP-ANN). The cross-validation procedure has shown good results in terms of the prediction ability for all of the chemometric models: standard LDA, 94.0%; stepwise LDA, 94.5%; SIMCA, 95.5%; and BP-ANN, 91.5%.

**KEYWORDS:** Tropea red onion, trace elements, ICP-MS, authenticity, chemometric analysis, protected geographical indications

# A.5

Anal Bioanal Chem (2011) 400:2903–2912
DOI 10.1007/s00216-011-4960-0

ORIGINAL PAPER

# Sarcosine as a marker in prostate cancer progression: a rapid and simple method for its quantification in human urine by solid-phase microextraction–gas chromatography–triple quadrupole mass spectrometry

**Brunella Cavaliere · Barbara Macchione · Marcello Monteleone · Attilio Naccarato · Giovanni Sindona · Antonio Tagarelli**

Received: 11 January 2011 / Revised: 18 March 2011 / Accepted: 26 March 2011 / Published online: 15 April 2011


**Abstract** Sarcosine is an amino acid derivative of *N*-methylglycine and is involved in the amino acid metabolism and methylation processes that are enriched during prostate cancer progression. It could also serve as a new target to be measured during therapeutic interventions and help in the identification of aggressive tumors for radical treatment. In this study, we present a new urine test that can help early diagnosis of prostate cancer. The method for the quantification of sarcosine in urine consists of a solid-phase microextraction (SPME) step followed by gas chromatography–triple quadrupole mass spectrometry analysis. We used a preliminary derivatization step with ethyl chloroformate/ethanol and the corresponding ester was then extracted by SPME in immersion mode. Several fibers were evaluated and the optimization of the parameters affecting the SPME process was carried out using an experimental design. The optimal values were 20 min extraction time, 10% NaCl, and 270°C using a divinylbenzene/Carboxen/polydimethylsiloxane fiber. The triple quadrupole analyzer acquired data in selected reaction monitoring mode, allowing us to obtain reconstructed chromatograms with well-defined chromatographic peaks. The accuracy and precision of this method were evaluated at concentrations of 70, 250, and 800 ng/ml and were found to be acceptable. Very satisfactory values (0.10 and 0.16 ng/ml, respectively) were also achieved for the limit of detection and the limit of quantification. The proposed protocol represents a rapid, simple, selective, and sensitive tool to quantify sarcosine in urine samples for prostate cancer diagnosis and for a screening test.

**Keywords** Sarcosine · Prostate cancer · Alkyl chloroformate · Gas chromatography · Solid-phase microextraction · Tandem mass spectrometry

B. Cavaliere · B. Macchione · M. Monteleone · A. Naccarato · G. Sindona (✉) · A. Tagarelli
Dipartimento di Chimica, Università della Calabria,
Via P. Bucci Cubo 12/C,
87030 Arcavacata di Rende, CS, Italy
e-mail: sindona@unical.it

## Introduction

Markers for cancer detection are molecules that may be present in higher than normal concentrations in many body fluids, such as serum and urine, of patients with a tumor. The widespread use of marker tests has led to early diagnosis of diseases such as cancer. In particular, the detection and treatment of prostate cancer was significantly affected by the detection of prostate-specific antigen (PSA) in the blood test. PSA has become an indispensable marker both for diagnosis and for follow-up of patients, particularly after radical prostatectomy. Despite its outstanding performance, the PSA test has significant limitations, such as its lack of specificity, failure to detect a significant number of PSA-negative tumors, and its high levels in both cancerous and healthy tissues, resulting in significant numbers of false-positive cases. Hence, more specific markers are needed for early and reliable diagnosis.

Sreekumar et al. [1] recently applied the metabolic approach for the detection of prognostic biomarkers that enhanced the clinical management of prostate cancer and this was supported by Jamaspishvili et al. [2]. In the original work, the authors profiled 1,126 metabolites across 262 clinical samples (42 tissues, 110 urine samples, and 110 plasma samples). The results showed that the concen-

**A.6**

EJMS EUROPEAN JOURNAL OF MASS SPECTROMETRY

IYC 2011 International Year of CHEMISTRY

OPEN ACCESS

Review

# Multistage mass spectrometry in quality, safety and origin of foods

Donatella Aiello,[a] Damiano De Luca,[b] Emanuela Gionfriddo,[a] Attilio Naccarato,[a] Anna Napoli,[a] Elvira Romano,[a,c] Anna Russo,[c] Giovanni Sindona[a,*] and Antonio Tagarelli[a]

[a]Dipartimento di Chimica, Università della Calabria, via P. Bucci, Cubo 12/C, I-87030 Arcavacata di Rende (CS), Italy. E-mail: sindona@unical.it

[b]CALAB, Laboratorio Chimico Merceologico Della Calabria, 87046 Montalto Uffugo (CS), Italy

[c]CRA, Istituto Sperimentale per l'Olivicoltura, c.da Li Rocchi, I-87036 Arcavacata di Rende (CS), Italy

Quality and safety control and the validation of origin are hot issues in the production of food and its distribution, and are of primary concern to food and agriculture organization. Modern mass spectrometry (MS) provides unique, reliable and affordable methodologies to approach with a high degree of scientificity any problem which may be posed in this field. In this review the contribution of mass spectrometry to food analysis is presented aiming at providing clues on the fundamental role of the basic principles of gas-phase ion chemistry in applied research fields. Applications in proteomics, allergonomics, glycomics, metabolomics, lipidomics, food safety and traceability have been surveyed. The high level of specificity and sensitivity of the MS approach allows the characterization of food components and contaminants present at ultra-trace levels, providing a distinctive and safe validation of the products.

# A.7

Sedimentary Geology xxx (2011) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Sedimentary Geology

journal homepage: www.elsevier.com/locate/sedgeo

## Characterization of the micrites in the Late Miocene vermetid carbonate bioconstructions, Salento Peninsula, Italy: Record of a microbial/metazoan association

A. Guido [a,*], A. Vescogni [b], A. Mastandrea [a], F. Demasi [a], F. Tosti [a], A. Naccarato [c], A. Tagarelli [c], F. Russo [a]

[a] *Dipartimento di Scienze della Terra, Museo di Paleontologia, Università della Calabria,Via Bucci Cubo 15b, 87036 Rende (CS), Italy*
[b] *Dipartimento di Scienze della Terra, Università di Modena e Reggio Emilia, Largo S. Eufemia 19, 41100 Modena, Italy*
[c] *Dipartimento di Chimica, Università della Calabria, Via Bucci Cubo 12c, 87036 Rende (CS), Italy*

ARTICLE INFO

*Article history:*
Received 26 January 2011
Received in revised form 29 September 2011
Accepted 10 October 2011
Available online xxxx

*Keywords:*
Vermetid bioconstruction
Automicrite
Organic matter
Late Miocene
Salento Peninsula
Italy

ABSTRACT

Small carbonate bioconstructions, composed of micrite and vermetids, occur in the Salento Peninsula (southern Italy) at the base of the reefal early Messinian Novaglie Formation. These lens-shaped bioconstructions are tens of meters in length and up to 2.5 m in thickness, and crop out patchily along the South-Eastern Salento coast line.

Micrite is the most abundant component. Its origin and role in this association have been investigated by means of microfacies, epifluorescence, and organic matter analyses. Three different types of micrite were recognized: (I) not/weakly fluorescent detrital micrite with a few fine bioclasts; (II) fluorescent micrite rich in fine bioclasts; (III) fluorescent clotted peloidal micrite (automicrite). The first type of micrite, mainly present in the basal layer of the carbonate bioconstructions surrounds vermetids in life position. The type (II) and type (III) micrites occur in the overlying deposits, characterized by sub-horizontally isooriented vermetid shells.

The high fluorescence of the type (II) and (III) micrite can be related to organic matter derived respectively from decaying metazoan and microbial organisms. Twofold organic matter origin was supported by FT-IR and GC-MS analyses, carried out on the extracted organic matter. Micrite (I) shows very low intensity of organic matter functional groups, confirming its abiotic origin. The spectra of the automicrite (III) are characterized by the presence of stretching C=C vibrations attributable to alkene and/or unsaturated carboxylic acids, that may be synthesized by microbes. GC-MS investigations indicate the presence of extended hopane series, straight chain saturated ($C_{14}$, $C_{16}$), monounsaturated ($C_{16}$, $C_{18}$), and diunsaturated $C_{18}$-acids, diagnostic of microbial activity. Microbial communities appear to have played a prominent role in the deposition and stabilization of Salento micrite–vermetid carbonate bioconstructions. The type (III) micrite, classifiable as microbialite or automicrite, can be regarded a sort of "primary framework" of these small "bioconstructions".