Io voglio un mondo
all’altezza dei sogni che ho.

(L. L.)

PREFACE

These few pages are directed to those who have not familiarity with chemometrics, since it wants to be a summary of what people told me during these three years of PhD and what I have learned in my personal experience, so a chemometrician could be bored by it. But please, be patient, in the same way you were, when you wait for ‘the response of a Matlab calculation’ and let other people to be introduced gradually in our world.

The continuous improvement of technologies has recently changed the modern approach to life and the advent of the computer’s era has modified the way of working so much that it seems today impossible thinking about the place of work without the constant use of a PC station. In every field of science the benefits brought are consistent and some tools have become indispensable and of everyday use. Analytical chemistry has not been spared from this *revolution*: modern engineering develops instruments, which are able to provide a large number of information, so that understand correctly the results obtained is not so easy. Nowadays the main difficulty is to have a correct interpretation of these data and several questions make their way: ‘Am I really able to understand what I measured?’, ‘Is all the information obtained useful for my aims?’ and moreover ‘How can I be sure that it is the real interpretation?’.

Chemometrics is the science developed to give answer to these questions; this highly interfacial discipline employs different data-analytic instruments such as multivariate statistic, applied mathematics and computer science, in order to address in chemistry, physics, biochemistry, medicine, biology and chemical engineering. In particular one of the applications of multivariate analysis concerns the possibility of deal with many variables simultaneously, using the entire information provided by one or more experiment, to define the state of our system and, at the same time, differentiate the useful information from the noise.

Fathers of chemometrics use a lot of example in an attempt to explain what multivariate analysis is, but without doubts my favorite one is the one concerning the perception of the system around us. Everyday we make decisions taking into account at the same time different factors. We can say that our brain works instinctively in a multivariate way and elaborate simultaneously a large number of information to give an answer to our questions and, as well as instinctively, it selects what type of information is inherent to a context and which is not. So, if the approach to everyday life is multivariate, why can't it be the same in science? Chemometrics aim is to apply this natural way of thinking to science, when an event must be evaluated, more than a variable needed to get a correct knowledge of the analyzed system.
In this thesis I will deal with only few of the chemometrics tools nowadays available for data treatment but it is really interesting to focus on the possibilities these approaches can give to research and how can reduce the quantity of work and allow a better optimization of the time.
My aim in writing this thesis is that of giving an 'easy to read' manual for some useful tools of data analysis in my field of study, namely food authentication.

Hope you will enjoy the lecture.

ABSTRACT

The progress in design of analytical instrumentations, especially hyphenated techniques, has rapidly increased the need to analyze data, which have a multi-way structure, i.e. each sample is characterized by a 2D landscape (e.g. chromatography/mass spectrometry, excitation/emission fluorescence). Hence the need of data analysis tools capable of handling this complex data structure. This issue is topical in food analysis, dealing with quality and authentication assessment, nowadays mandatory both from producers and consumers points of view. From here the necessity of improving analytical methods, in terms of efficiency, robustness and rapidity, and coupling them with multivariate data analysis to effectively extract information. My thesis project explored two chemometrics tools handling multi-way data: Multivariate Curve Resolution (MCR) and Multi-way classification methods. As well as two application fields are investigated: resolution and quantification of food constituents, by the so called 'mathematical chromatography approach', where peaks are resolved 'a posteriori' (using MCR), which allows handling strong co-elution issues; and building authenticity models to recognize typical products label by protected denomination of origin (PDO). To this aim multi-way (NPLS-DA) and multiset (discriminant-MCR) discriminant methods as well as multi-way class modeling (N-SIMCA) were comparatively applied and discussed with respect to the current state of art methodology. Case studied derived from hyphenated techniques, here used as fingerprinting methods, in particular PDO Lambrusco wine of three different denomination were for the first time characterized (HPLC-DAD) with respect to the phenolic fraction. Additional data sets, from previous work of our group or from literature regards PDO Ligurian Extra Virgin Olive Oil and wines of different geographical for the identification of provenience characterized by (HS-SPME/GC-MS).

In the MCR application salient aspects dealt within the research work have been the use of local rank and selectivity constraints, that had been crucial

for the correct resolution and quantification of target compound.

Regarding classification, the study focuses on the n-dimensional version of the SIMCA (Soft Independent Modeling of Class Analogies) methodology, recently developed, NSIMCA. Here focus was on demonstrating advantages/disadvantages of class modeling with respect to discriminant approaches and of multi-way classification with respect to classical 2-way classification performed on unfolded data or on scores from decomposition methods, such us PARAFAC or Tucker3.

This part demonstrates that multi-way data analysis can represent a powerful strategy in authentication and classification tasks giving high values of sensitivity and specificity for all the model calculated; class modeling techniques respect to discriminant one results suboptimal in case of strong intra-class heterogeneity.

Furthermore, a new classification approach based on extension of MCR based regression to discrimination task is also proposed. Basically, it works including class information as an additional selectivity constraint. Main advantages is that, the whole resulting MCR concentration profiles can be interpreted as classical samples scores, the recovered component profiles (spectral matrix) provides a very interesting interpretation of the characteristic of the modeled classes in terms of spectral signature which are class specific.