Chemometrics and Intelligent Laboratory Systems 154 (2016) 176–184

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Modelling the ageing process: A novel strategy to analyze the wine evolution towards the expected features☆

Ana C. Pereira [a,b], Maria J. Carvalho [b,c], Andreia Miranda [b,c], João M. Leça [b,c], Vanda Pereira [b,c], Francisco Albuquerque [d], José C. Marques [b,c], Marco S. Reis [a,*]

[a] Chemical Process Engineering and Forest Products Research Centre (CIEPQPF), Department of Chemical Engineering, University of Coimbra, Rua Sílvio Lima, 3030-790 Coimbra, Portugal
[b] Centre of Exact Sciences and Engineering, University of Madeira, Campus da Penteada, 9000-390 Funchal, Portugal
[c] Institute of Nanostructures, Nanomodelling and Nanofabrication (I3N), University of Aveiro, 3810-193 Aveiro, Portugal
[d] Madeira Wine Company, Rua Dos Ferreiros 82, 9000 Funchal, Portugal

ARTICLE INFO

Article history:
Received 30 November 2015
Received in revised form 24 March 2016
Accepted 27 March 2016
Available online 2 April 2016

Keywords:
Wine ageing
Chemical characterization
Feature extraction
Classification models
Process monitoring and evaluation

ABSTRACT

In this work we present a new strategy to monitor the wine evolution during the ageing process. More specifically, we validate a procedure for analyzing how wine evolves during the ageing process in relation to the desired and expected quality features and we apply the proposed methodology to the case of a Portuguese fortified wine, the Madeira wine, where we compare the wine evolution under two different ageing processes.

The approach developed consists on modeling samples labeled as aged reference wines (5 year old Madeira wines), produced from four different grape varieties, and then analyze how and in which extent young wines (up to 3 years old) come closer to the reference data set. The analysis is based on a comprehensive set of chemical data, including: polyphenolic composition, organic acids, reducing sugars, color and oenological parameters, commonly used as routine quality control information. The study considers several feature extraction methods, such as: Principal Components of Analysis (PCA), Independent Component of Analysis (ICA) and Partial Least Squares (PLS). The classification methodologies tested were: Linear Discriminant Analysis (LDA), nearest neighbor (*k*-NN) and Soft Independent Modelling by Class Analogy (SIMCA). The different options of preprocessing/feature extraction/classification were evaluated and compared using a Monte Carlo approach.

From our analysis, the best combination of feature extraction/classification methodologies was PLS/LDA, which presented a classification performance of approximately 90% for three out of the four classes modeled, and of about 78% for the remaining one. Regarding the wines monitored during the first 3 years, our analysis revealed that they indeed mature in relation to the five year old reference wines. Furthermore, for some wines, it is possible to detect differences between the two ageing processes analyzed.

This study is of particular importance for this type of wines, where the ageing process plays a central role for attaining the expected quality levels, implying significant risks and costs for local and industrial producers. Notwithstanding the specific case study presented, the strategy outlined can be extrapolated to other products with similar characteristics in terms of their monitoring and process control.

1. Introduction

In recent decades, chemometrics has yielded important tools for wine's safeguard and protection, in order to deal with possible adulteration practices and to verify product authenticity [1–5]. Pattern recognition techniques have been successfully employed in the definition of strategies to protect the wine designation of origin [6–10], to confirm the grape variety authenticity and the wine style typicity [11–14]. These studies are aligned with the endeavor of regulatory authorities to prevent and detect adulteration practices and also with the increasing demand of information by consumers, who are increasingly interested in knowing product details, such as where, when and how wines were produced [15]. Of particular interest from the producers' standpoint is the development of new winemaking process control tools. However, in this scope, patterns recognition techniques were explored to a lesser extent. In fact, one of the main concerns of wine producers it to guarantee that their wines will get the expected and desired features of their target market, ensuring the success of their investment. In this study, we developed a monitoring framework that assists them in this complex but critical task. Supervised pattern recognition techniques were explored with the purpose of creating a classification

☆ Selected paper from the 9th Edition of the In Vino Analytica Scientia Symposium in Trentino Alto Adige (Italy), 14-17 July 2015.

* Corresponding author.
E-mail address: marco@eq.uc.pt (M.S. Reis).

http://dx.doi.org/10.1016/j.chemolab.2016.03.030
0169-7439/

model for a specific set of aged wines, selected as reference wines, and then used to implement a strategy to verify if younger wines are evolving in accordance to the expected and desired characteristics.

Supervised pattern recognition techniques are commonly distinguished into two groups: one focused on discriminating among classes, like linear discriminant analysis (LDA), *k*-nearest neighbor ($k-$NN), decision trees, support vector machines (SVM) or artificial neural networks (ANN), and another group oriented towards modelling classes in an isolated way, the so-called one-class classification methods, for which soft independent modelling by class analogy SIMCA is a well-known example. The first group of discriminant techniques set up classification rules for a number of pre-specified classes that cover the entire space of analysis. These rules are then used for assigning class labels to new samples [25]. On the other hand, one-class modelling methods generate a separate model for each category, and define conditions under which such category should be assigned to future samples. Each new sample is then analysed in order to verify whether or not it is compatible with the characteristics of each class-specific model [26].

In contrast to supervised pattern recognition techniques, multivariate prediction methodologies have found broader applications in winemaking process control [16]. Usually, in this context, spectrometers are the analytical devices chosen to collect the data, which are then processed and used to follow target parameters during the winemaking process. These multivariate approaches are able to provide fast, in-line and reliable measurements in terms of a wide range of wine quality control specifications, such as alcohol content, total and volatile acidity, relative density, reducing sugars, tartaric, malic and lactic acid concentration [17–19]. The untargeted information from spectrometer devices has also been used to follow the wine ageing process [20]. In this particular case, the ageing time prediction estimates were compared with the real wine ageing time in casks and used to evaluate if wines showed the expected characteristics. Similar strategies for wine ageing control have also been proposed, but using other type of chemical information, namely the volatile and phenolic wine composition [21,22]. In spite of the lack of consensus regarding the type of chemical information to be used in studies of this nature, several works have demonstrated that the volatile profile is a good choice when it comes to distinguish between different grape varieties and the profile of trace elements is effective when the problem is to discriminate amongst possible geographical origin [23]. Concerning other classification purposes, namely wine classification by quality, brands or winemaking technology involved, the chemical data set used is quite diverse, including also the untargeted fingerprints [24]. In the present study, the chemical data set used comprises information from phenolic and organic acid composition, reducing sugar concentrations, two of the main furanic compounds, color data and other routine quality control parameters.

In this work we focus on the Madeira wine (MW) as a case study, which is an important Portuguese fortified wine, produced exclusively in the Madeira island (Designation of Origin, DO). The alcohol content of this fortified wine varies from 17 to 22% and its main characteristics are a result of the grape varieties from which they are produced and from the particular ageing process implemented. Four different styles of Madeira wines are considered in this study, namely sweet, medium sweet, medium dry, and dry wines, which are produced from *Malvasia*, *Boal*, *Verdelho* and *Sercial* grape varieties, respectively. These Madeira wines represent about 53% of all the wines sold with indication of age. In spite of this group of wines represent a small percentage in Madeira wines sales (7% of the 4.0 million of liters produced on average per year) they are recognized as the finest and highest quality Madeira wines.

Madeira wines can follow two typical ageing processes: the typical ageing process in casks, called the *Canteiro* ageing process, in which wines are stored in casks placed in warmed lofts heated by the sun (the temperature can range from 20 to 30 °C), and the *Estufagem* ageing process, in which the wine is heated and kept in stainless steel vats at a temperature of about 45 °C at least during three months. The aim of this second process is to accelerate wine ageing. The *Estufagem* process is not a common practice for the high quality Madeira wines that are under study in the present work. However, some attempts are being made in order to evaluate the impact of this ageing process for this selected group of wines. Specifically, the goal is to verify if the *Estufagem* process can indeed bring any advantage to these wines and evaluate the best way to conduct it for each style of wines. This study addresses this issue, by first developing a procedure to follow the ageing process and to assist in decision-making about wine compliance on the basis of a large variety of intrinsic wine characteristics.

This paper is organized as follows. In Sections 2.1 and 2.2 the data collection process is presented and the data sets are described, as well as wine chemical characterization carried out. In Section 2.3, we present the data analysis methodologies used, including those employed with the purpose for preprocessing, feature extraction and classification. The evaluation of classification models is presented in Section 3, together with the proposed approach to follow the evolution of the ageing process. This approach is applied to the analysis of younger wines that have been monitored during their ageing process. Finally, a brief discussion of the issue addressed in this study is presented in Section 4 and the main conclusions summarized in Section 5.

2. Materials and methods

2.1. Sample collection and data sets

Data collection for this study contemplated two different sets of samples. The first one comprise wines from the 2007 and 2008 harvests, collected and analysed after 5 years of ageing in casks. These wines were selected by an oenologist as reference wines of a five year old Madeira wine. The selection was made from wines of two harvests in order to incorporate in our analysis the inter-harvest variability. It comprises wines from the four white grape varieties recommended for Madeira wine production: *Malvasia*, *Boal*, *Verdelho* and *Sercial*. *Verdelho* and *Sercial* are produced in smaller quantities (together they represent about 35% of sales), and therefore they are also represented by a smaller number of samples than the remaining wines. More specifically, the following numbers of samples were collected and analyzed: 13 *Malvasia* wines (7 from the 2007 harvest and 6 from 2008), 9 *Boal* wines (5 from the 2007 harvest and 4 from 2008), 4 *Verdelho* and 4 *Sercial* wines (2 from 2007 and the other two from 2008, in both cases). The typical features of these wines can be summarized as follows. *Malvasia* grape variety is commonly used to produce sweet wines (residual sugars can range between 99 and 140 mg/L). These wines are frequently described as having rich honey notes, raisin and caramel aromas and present a dark brown color. *Boal* wines are medium sweet wines (residual sugars can vary between 80–100 mg/L). These wines typically have nutty and vanilla characteristics, flavors of smoke and dried fruit aromas, and are quite dark (mahogany) in color. *Verdelho* is a medium sweet wine, with sugar level ranging from 60 up to 81 mg/L. These are the more aromatic Madeira wines produced from white grape varieties, showing hints of caramel flavors and moderate acidity. *Sercial* is known to produce excellent dry wines (residual sugar varies from 25 to 61 mg/L), standing out for its mineral tones, subtle fruit and nut flavors, and high acidity [27].

The second group of samples is formed by wines from the 2011 harvest, each one produced exclusively from one of the four grape varieties (*Malvasia*, *Boal*, *Verdelho* and *Sercial*). These wines follow the common vinification practices [27] and they were separated in two sets. For the first set, wines were stored in casks and aged according to the *Canteiro* process, while for the second set wines were submitted to the *Estufagem* ageing process. Both sets of wines were monitored during the first three years of ageing. Samples were collect at the end of the *Estufagem* process, in the fourth month and then at the end of the first, second and third years of ageing.