



## Analytical Methods

## Application of a high-throughput process analytical technology metabolomics pipeline to Port wine forced ageing process

Cristiana C. Castro<sup>a</sup>, R.C. Martins<sup>b</sup>, José A. Teixeira<sup>a</sup>, António C. Silva Ferreira<sup>c,d,\*</sup><sup>a</sup>IBB – Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, Universidade do Minho, Campus de Gualtar, 4710-057 Braga, Portugal<sup>b</sup>Bioinformatics – Molecular and Environmental Biology Centre, Universidade do Minho, Campus de Gualtar, 4710-057 Braga, Portugal<sup>c</sup>Escola Superior de Biotecnologia – CBQF, Universidade Católica Portuguesa, Rua Dr. António Bernardino de Almeida, 4200-072 Porto, Portugal<sup>d</sup>Stellenbosch University, Private Bag X1, Matieland, 7602 Stellenbosch, South Africa

## ARTICLE INFO

## Article history:

Received 28 February 2013

Received in revised form 26 July 2013

Accepted 30 July 2013

Available online 8 August 2013

## Keywords:

Port wine

Forced ageing

GC–MS

Multivariate analysis

Co-expression

## ABSTRACT

Metabolomics aims at gathering the maximum amount of metabolic information for a total interpretation of biological systems. A process analytical technology pipeline, combining gas chromatography–mass spectrometry data preprocessing with multivariate analysis, was applied to a Port wine “forced ageing” process under different oxygen saturation regimes at 60 °C.

It was found that extreme “forced ageing” conditions promote the occurrence of undesirable chemical reactions by production of dioxane and dioxolane isomers, furfural and 5-hydroxymethylfurfural, which affect the quality of the final product through the degradation of the wine aromatic profile, colour and taste. Also, were found high kinetical correlations between these key metabolites with benzaldehyde, sotolon, and many other metabolites that contribute for the final aromatic profile of the Port wine. The use of the kinetical correlations in time-dependent processes as wine ageing can further contribute to biological or chemical systems monitoring, new biomarkers discovery and metabolic network investigations.

© 2013 Elsevier Ltd. All rights reserved.

## 1. Introduction

Metabolomics is a recent ‘omics’ field and is based on biology science devoted to fully identify phenotypes, the connectivity that relates their constituents and the dynamical response to perturbations, as well as the relationships with the rest of the cellular molecular machinery, namely at the regulatory levels such as genomics, transcriptomics and proteomics at the systems biology stage (Feist, Herrgård, Thiele, Reed, & Palsson, 2009).

The study of such complex matrix on a chemical composition perspective can be described as being *multi-scale*, i.e. several orders of magnitude (ppm–ppt) and *multivariate*, i.e. large diversity of chemical substances, and constitutes an extremely challenging problem being the main goal to obtain a comprehensive profile of the matrix. Ideally it would require a true ‘omic sensor’, i.e. with an unlimited linear zone of response and sensitive to all substances resulting in a global picture of the bioprocess (Jerez, 2008).

A high-throughput metabolomic chromatography systems consists in one part of hardware (analytical equipment) and the other part of bioinformatics (signal processing, data storage and multi-

variate analysis), which are becoming a trend in modern biotechnology. These methods become more appealing to systems biology and molecular biology sciences, when signal processing is performed which turns chromatography into a high-throughput system. This capacity of peak extraction by automated techniques makes possible the holistic assessment of the metabolism (Villas-Boas, Mas, Akesson, Smedsgaard, & Nielsen, 2005).

Gas chromatography–mass spectrometry (GC–MS) is specially suited for the study of yeast volatile metabolites. These are of great importance in fundamental biological functions, such as signaling and precursors of biochemical pathways, as well as in the formation of aroma from wines that result from fermentation and ageing, either in barrels or bottles. Several hundreds of compounds can be captured by GC–MS, and each compound produces a unique mass spectral fingerprint, which is afterwards used for metabolites recognition and quantification (Dunn et al., 2011). Classical mass spectroscopy is highly laborious and a significant amount of time is necessary for peak analysis, fingerprint recognition, identification and quantification by an analyst (Christensen, Tomasi, & Hansen, 2005). In this context nontargeted methodologies are extremely important, with an unbiased approach, toward understanding the biological system (Castro et al., 2012; Fiehn, 2002; Rodrigues, Barros, Carvalho, Brandão, Gil & Ferreira, 2011). When automatic chromatogram signal processing becomes a robust skill, mass-spectroscopy may be used as a high-throughput technology.

\* Corresponding author at: Escola Superior de Biotecnologia – CBQF, Universidade Católica Portuguesa, Rua Dr. António Bernardino de Almeida, 4200-072 Porto, Portugal. Tel./fax: +351 225 580 001.

E-mail address: [asferreira@porto.ucp.pt](mailto:asferreira@porto.ucp.pt) (A.C. Silva Ferreira).

Chromatography enables physico-chemical deconvolution of the complex matrix components during the analytical process originating a pattern, that is, a chromatogram. Although it allows samples classification based on the different patterns, it is important to relate that there are inherent variations in the chromatographic process, mainly related to the retention time (peak shifts) and baseline drifts. Therefore, in order to use GC–MS data on a high-throughput base, those issues must be focused on the preprocessing data handling prior to multivariate analysis (Christensen et al., 2005). In that regard, several authors have proposed some preprocessing methodologies, including noise filtering, baseline correction, peaks detection, alignment, identification and normalization (Castillo, Gopalacharyulu, Yetukuri, & Oresic, 2009).

In the present study, the aim is to develop a pipeline methodology based on Process Analytical Technology (PAT) multivariate analysis (MVA) and GC–MS data processing in order to provide a holistic view of the impact of the presence of oxygen and higher temperature during the “forced ageing” of a Port wine matrix. This ageing process, is an extremely complex chemical process that has been extensively studied by our research group (Silva Ferreira, Barbe, & Bertrand, 2002; Silva Ferreira, Barbe, & Bertrand, 2003a; Silva Ferreira, Monteiro, Oliveira, & Guedes de Pinho, 2008), where several chemical mechanisms take place and are responsible for differences in sensory perception affecting final product quality.

The metabolomics pipeline methodology here provided is applied to a “forced ageing” Port wine data set and includes both chromatographic signal preprocessing and MVA methodologies, together with the exploration of candidate metabolites expression and co-expression within the overall ageing process (see Fig. 1). Candidate metabolites correspond to MVA model variables, candidate to explain specific pathways of the metabolism that have chemical meaning and can be used for understanding the overall process. The co-expression of a candidate metabolite constitutes a powerful feature, enabling the study of potential interconnections between the candidates formation and consequently allowing further research for overlapping/connections between chemical and/or biochemical mechanisms. In this context, the global high-throughput pipeline methodology provides increased rate of metabolites identification involved in the chemical process contributing to further build the metabolic network of the overall process.

## 2. Materials and methods

### 2.1. Wine material and treatments

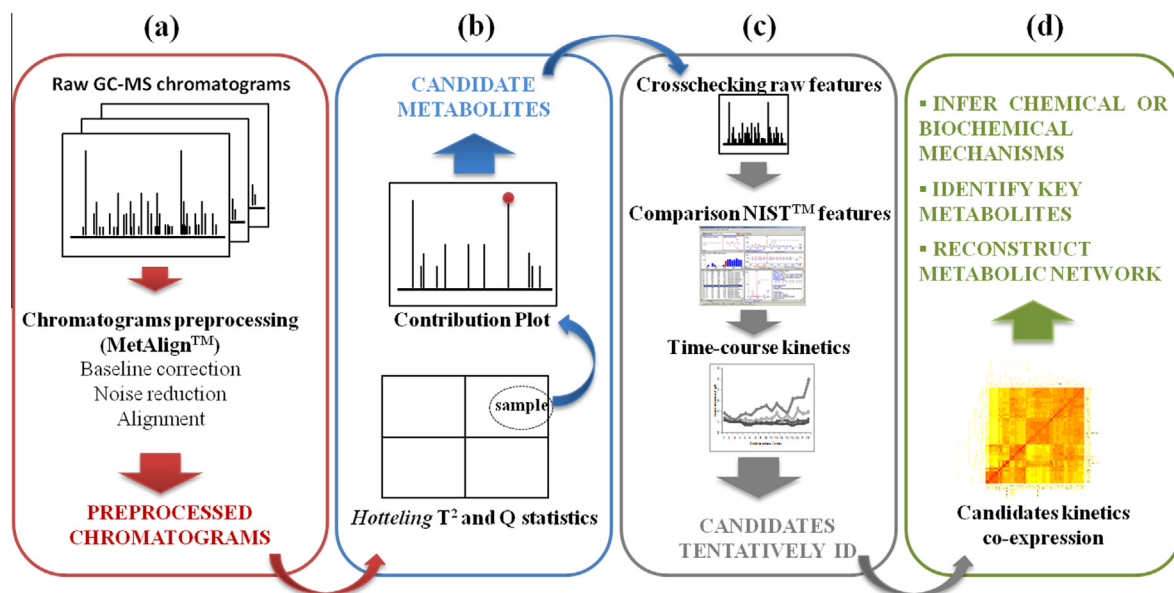
Young Port wine characterized by a pH = 3.4, 2.5 mg/L of dissolved oxygen, 17 mg/L of free SO<sub>2</sub>, 150 g/L of reducing sugars and 20.5% (v/v) alcohol made on the year of the experiment (without any oak contact) was used in this experiment.

Different oxygen treatments were introduced: 0 (P#\_NoO<sub>2</sub>), 1 (P#\_1inj), 2 (P#\_2inj) and 5 (P#\_5inj) saturations in glass vessels filled with 500 mL of Port wine. Oxygen saturation was obtained by stirring each sample vigorously for about 1 h until an oxygen concentration of about 8–9 mg/L was reached. This was performed in a laminar flow chamber under UV light to prevent microbial contamination. Oxygen injections were measured with a WTW 340 Oxygen Probe (Silva Ferreira, Hogg, & Guedes de Pinho, 2003b), during 18 weeks (‘P1’ to ‘P18’) of storage at 60 °C in a temperature controlled incubator, and discrete samples were obtained heuristically for each oxygen regime and further analyzed by GC–MS (see Section 2.4 GC–MS analysis, for specifications).

The forced ageing experiment was implemented to simulate the typical oxidation aroma of Port wine by promoting chemical changes on wine composition. Samples were supplemented with different oxygen regimes and kept at high temperatures (60 °C). Although those extreme conditions are not representative of real ageing process, they were selected in order to be able to reproduce in the laboratory on a reasonable time the ageing process, in spite of the risk of promoting other chemical reactions which would not occur in the normal process. The length of the duration for the forced aged protocol was sensory-driven. In fact, at each sampling point, samples were submitted to sensory analysis in order to validate that the product was still perceived as Port wine and it was observed that after 18 weeks it not accepted as such.

### 2.2. Chemicals

All chemicals employed were of analytical grade: anhydrous sodium sulphate (HPLC grade) (Merck, Darmstadt, Germany), dichloromethane (Lab Scan, Sowinskięo, Gliwice), 3-octanol (97%)



**Fig. 1.** High-throughput pipeline methodology for GC–MS data processing: (a) GC–MS data preprocessing; (b) Multivariate Analysis; (c) Candidate metabolites identification; (d) Temporal relationships of candidate metabolites for process contextualization.

(Sigma–Aldrich, USA), *cis*- and *trans*-5-hydroxy-2-methyl-1,3-dioxane (>99%, Sigma–Aldrich, USA), *cis*- and *trans*-4-hydroxymethyl-2-methyl-1,3-dioxolane (>99%, Sigma–Aldrich, USA), benzaldehyde (>98%, Sigma–Aldrich, USA), furfural (>97%, Sigma–Aldrich, USA), 5-hydroxymethylfurfural (5-HMF) (>98%, Sigma–Aldrich, USA); and 3-hydroxy-4,5-dimethyl-2(5H)-furanone (>99%, Sigma–Aldrich, USA).

### 2.3. Volatiles extraction

A liquid–liquid extraction was performed to extract the volatile fraction from each sample. The procedure used was as follows: 5 g of anhydrous sodium sulphate and 50  $\mu$ L of internal standard – 3-octanol – were added to 50 mL of sample and were extracted twice with 5 mL of dichloromethane using a magnetic stir bar for 5 min per extraction, 2 mL of the resulting organic phase were concentrated under a nitrogen stream 4 times (Silva Ferreira, Hogg, & Guedes de Pinho, 2003b).

### 2.4. GC–MS analysis

The forced ageing experimental protocol was performed in duplicate for practical reason and some samples were analysed by GC–MS on the replicate trial. These GC–MS targeted analysis were used as a crosscheck procedure on the evaluation and validation of both metabolites features preservation after the preprocessing methodology and the kinetics of each metabolite.

Discrete sample extracts were analyzed using a Varian CP-3800 gas chromatograph (USA) equipped with a Varian Saturn 2000 mass selective detector (USA) and a Saturn GC/MS workstation software version 5.51. The column used was STABILWAX-DA (60 m  $\times$  0.25 mm, 0.25  $\mu$ m) fused silica (Restek, USA). The injector port was heated to 220 °C. The split vent was opened after 30 s. The carrier gas was Helium C-60 (Gasin, Portugal) at 1 mL/min, constant flow. The oven temperature was 40 °C (for 1 min), then increased at 2 °C/min to 220 °C and held for 30 min. All mass spectra were acquired in the electron impact (EI) mode. The ion trap detector was set as follows: The transfer line, manifold and trap temperatures were respectively 230, 45 and 170 °C. The mass range was 33–350 *m/z*, with a scan rate of 6 scan/sec. The emission current was 50  $\mu$ A, and the electron multiplier was set in relative mode to auto-tune procedure. The maximum ionization time was 25,000  $\mu$ sec, with an ionization storage level of 35 *m/z*. The injection volume was 1  $\mu$ L and the analysis was performed in Full Scan mode.

### 2.5. GC–MS data preprocessing

GC–MS ion-trap-Trap raw chromatograms were converted to text format (\*.txt) using MASSTransit™ (Version 3.0, Palisade Corporation) and imported to MetAlign™ software. This preprocessing software was used for baseline correction, accurate mass calculation, data smoothing and noise reduction, followed by a spectral alignment of the GC–MS data (Lommen, 2009). Aligned chromatograms matrix was exported as “.csv” file and subjected to multivariate analysis using R-Project 2.15.0 and also to build the identification and quantification tables used to obtain the expression and co-expression of candidate molecules within the “forced ageing” Port wine process.

### 2.6. Multivariate analysis (MVA)

MVA involves the use of mathematical and statistical tools to extract the information from complex datasets (e.g. GC–MS datasets) (Cozzolino, Cynkar, Shah, Damberg, & Smith, 2009), considering the system as a multi-dimensional comparison of the

different chemical information present in the sample (Rodrigues et al., 2011). Relevant Principal Component Analysis (RPCA) (Martins, Lopes, Vicente, & Teixeira, 2008) is a non-supervised methodology that determines relevant orthogonal decompositions of the information able to discriminate between samples (scores), and variables (loadings), that provide relevant interpretation of significant variables (or metabolites) on each principal component (Martins et al., 2008).

The use of both scores, Hotelling  $T^2$  and *Q* statistics – namely diagnostic plot – and the contribution plot for the identification of variables responsible for a specific behavior/observation being outside the normal operation conditions is quite common (Conlin, Martin, & Morris, 2002; Hotelling, 1947; Kourti & MacGregor, 1996; Mason, Tracy, & Young, 1995; Mason & Young, 1999).

In this pipeline methodology, diagnostic and contribution plots were used together for the selection of samples standing out of the latent model according to their metabolic characteristics, in response to the external conditions imposed, and to analyze variables/candidate metabolites, to be responsible for these differences during the Port wine “forced ageing” process.

### 2.7. Metabolites identification and quantification

In order to identify variables, that is candidate metabolites, which are responsible for the deviations of samples in the diagnostic plot, an in-house application was developed in R language, in order to build both the identification and composition tables using the aligned matrix (\*.csv) derived from MetAlign™ preprocessing software (Lommen, 2009).

The identification table was built using the discrimination of the *m/z* channels and the corresponding normalized intensity in each scan. This was executed by imposing the minimum number of fragments, from the original feature, necessary for compounds identification (3 or more) and performing the Sinkhorn factorization for checking peak consistency among samples (Sinkhorn, 1964; Walczak & Wu, 2005). Only identifiable and coherent peaks which hold the same mass fingerprint, among all samples build up the identification table, are used for a correct identification of molecules by comparing their fingerprint with databases (Linstrom, Mallard, National Institute of, & Technology, 1997). The composition table is built using compounds concentrations obtained directly by linear relation to the internal standard for each sample (in this case, 3-octanol), taking into account the maximum intensity of one *m/z* channel.

### 2.8. Metabolites kinetics and co-expression in time-course

The kinetics or time-expression of each candidate molecule was directly obtained from the composition table, which was determined by the maximum intensity of one *m/z* channel. This fact can make the interpretation less sensitive, leading to variations of the preprocessed expression when compared to the raw data expression of each molecule. However, the main interest in the time-expression is to explore the tendency of the molecule concentration through samples.

The co-expression of compounds is obtained by Pearson's correlation coefficient (Rodgers & Nicewander, 1988) of the composition table. Correlations between the kinetics of candidates and a key metabolite, known to be involved in a specific chemical phenomenon in the process, are displayed in a heatmap (performed in R-Project 2.15.0), in a colour range from white-to-red, which correspond to correlations between  $0.8 \leq R^2 \leq 1$ . Thus, clusters from the heatmap can be used for interpretation of temporal relationships of candidate metabolites allowing process contextualization, identification of the involved metabolites and metabolic network reconstruction.

### 3. Results and discussion

#### 3.1. Robustness of GC–MS preprocessing using MetAlign™

Due to *multi-scale* nature of Port wine chromatograms matrix, operational parameters of MetAlign™ (Lommen, 2009) preprocessing software tool, need to be optimized in order to provide simultaneous extraction of high and low concentration compounds. The optimization was performed according to the analytical equipment characteristics, and taking into account the preservation of the original features of a sub-set of volatiles. Volatiles were selected to monitor the impact of the different algebraic treatments in each metabolite spectra fingerprint within data processing. The criteria for metabolites selection were: (i) volatiles recognized as products of oxidation, Maillard reaction products and combination of both; (ii) substances present at different scales (i.e. *ppm-ppt*) and simultaneously acquired on the same analytical run.

The selected volatiles were: the heterocyclic acetals of glycerol – dioxolane 1 and dioxolane 2 (isomers *cis*- and *trans*-4-hydroxymethyl-2-methyl-1,3-dioxolane) – the furanic aldehydes: furfural, 5-hydroxymethylfurfural (5-HMF) and phenylacetaldehyde, a Strecker aldehyde (Hofmann & Schieberle, 2000), dependent of oxygen and temperature during ageing.

The original features of each metabolite were compared with the preprocessed resulting features in order to ensure the validity of the preprocessing methodology. Table 1 presents the correlation coefficients between features of both raw data and MetAlign™ (Lommen, 2009). These features can be used as a validation index for the preprocessing software tool preserving the original chromatographic features. The high correlation coefficients ( $R^2$ ) are presented in Table 1, indicating the preprocessing parameters used to maintain the integrity of the overall matrix in the forced ageing of Port wine. As consequence, the features of the selected compounds with different magnitudes (*ppm-ppt*) were preserved, showing the strength of the preprocessed matrix which can be used for understanding chemical compound interaction in the Port wine “forced ageing” process.

#### 3.2. Multivariate analysis (MVA)

Performing the RPCA to the MetAlign™ (Lommen, 2009) final matrix, four relevant orthogonal decompositions were found to explain a total of 63.2% of the chromatographic data variability (22.3% PC1, 21.3% PC2, 11.7% PC3 and 7.9% PC4). The most relevant components for capturing samples variability within the oxidation under each oxygen saturation regime are PC1 and PC2. According to both relevant decompositions, samples are distributed through the components as a function of the oxidation process occurring in the Port wine which induces chemical composition changes in samples (Figure A in supplementary material). Chemical differences occurring in the beginning of the “forced ageing” process are less representative than those taking place in the final weeks (samples closest to week 18 – “P18”) of the oxidation process as samples corresponding to these times are spread through the

**Table 1**  
Correlation coefficients between raw and MetAlign™ data features using quantifier ion of the selected metabolites.

Metabolite <sup>b</sup>	$m/z^a$	$R^2$
Dioxolane 1	103	0.9242
Dioxolane 2	103	0.9200
Furfural	95	0.9700
5-HMF	97	0.7120
Phenylacetaldehyde	91	0.9564

<sup>a</sup> Quantifier ion.

<sup>b</sup> Normalized value.

PC2 and PC1, respectively and the variability of the chemical information is higher in PC1. So, it can be stated that the complexity of the chemical profile of the resulting Port wine made under the different oxygen saturation regimes at high temperature increases during the maturation process leading to the production of compounds of during the storage time.

#### 3.3. Diagnostic and contribution plots – selection of candidate molecules

To understand the role of dissolved-oxygen regime and high temperatures during Port wine “forced ageing” process, the selection of candidate metabolites was performed by employing diagnostic ( $Q$  statistics and *Hotelling  $T^2$* ) and contribution plots, which are presented in Fig. 2 and Fig. 3, respectively (Hotelling, 1947) (Mason & Young, 1999; Mason et al., 1995).

The Diagnostic plot ( $Q$  statistics vs. *Hotelling  $T^2$* ) is performed because not all features in the metabolic matrix (resulting from MetAlign™ preprocessing) preserve the same quality after data decomposition into relevant principal components. In these cases, their reconstruction is statistically impossible, and their metabolic information is widely different from the average.  $Q$  statistics (square prediction error) is determined to assess the feature extraction quality, and the line crossing the  $Q$  axis correspond to the statistic confidence interval ( $Q_\alpha$ ) determined using the average and standard deviation of  $Q$  at a level of significance of  $\alpha = 0.05$ . Samples above  $Q_\alpha$  do not represent robust feature extractions (Conlin, Martin, & Morris, 2000). On the other hand, *Hotelling  $T^2$*  measures the distance to the center of the data. The line crossing the *Hotelling  $T^2$*  axis corresponds to the upper confidence interval of the distribution, determined at a level of significance of  $\alpha = 0.05$ . Samples above the  $T\alpha^2$  are considered to present significantly different features (Castro et al., 2009; Qin, 2003).

The determination of the  $Q$  statistic confidence interval ( $Q_\alpha$ ) limits the  $Q$  values above which features are not robust. According to the Diagnostic plot in Fig. 2, sample P18\_5inj is above the  $Q_\alpha$ , which means that the reconstruction of its metabolic information is statistically impossible.

In practice, P18\_5inj was exposed to extreme conditions (high temperature, 60 °C, under a saturation regime of five injections of  $O_2$ ) for 18 weeks, and its sensorial analysis highlights an unconformity of its chemical profile comparing to other samples recognized as Port wine. So, as the chemical information of the sample reflects the impact of the extreme conditions during storage, and it cannot be accepted as a normal Port wine, this was used as a control for monitoring the ageing process of these wines.

In contribution plot of Fig. 3, are visible the candidate variables/scans (1 to 14) representing metabolites that mostly explain chemical differences from the latent model. These candidates addresses the chemical quality of the final Port wine in the presence of the higher levels of oxygen after 18 weeks of “forced ageing” at high temperature (60 °C). The contextualization of these candidate molecules through the different chemical pathways involved in the “forced ageing” of Port wine entails their correct identification according to each mass spectra fingerprint, being essential for further comprehension of the temporal relationships with the entire matrix which constitutes the aromatic profile of the Port wine samples.

#### 3.4. Candidate metabolites identification and time-expression

In the present pipeline methodology, mass spectra features from the candidate metabolites selected in the contribution plot are validated by (i) crosschecking their presence in the raw data, (ii) comparing their features with those present in NIST 98 MS library (Linstrom et al., 1997) and (iii) time-expression of each metabolite throughout the maturation process under each condition.

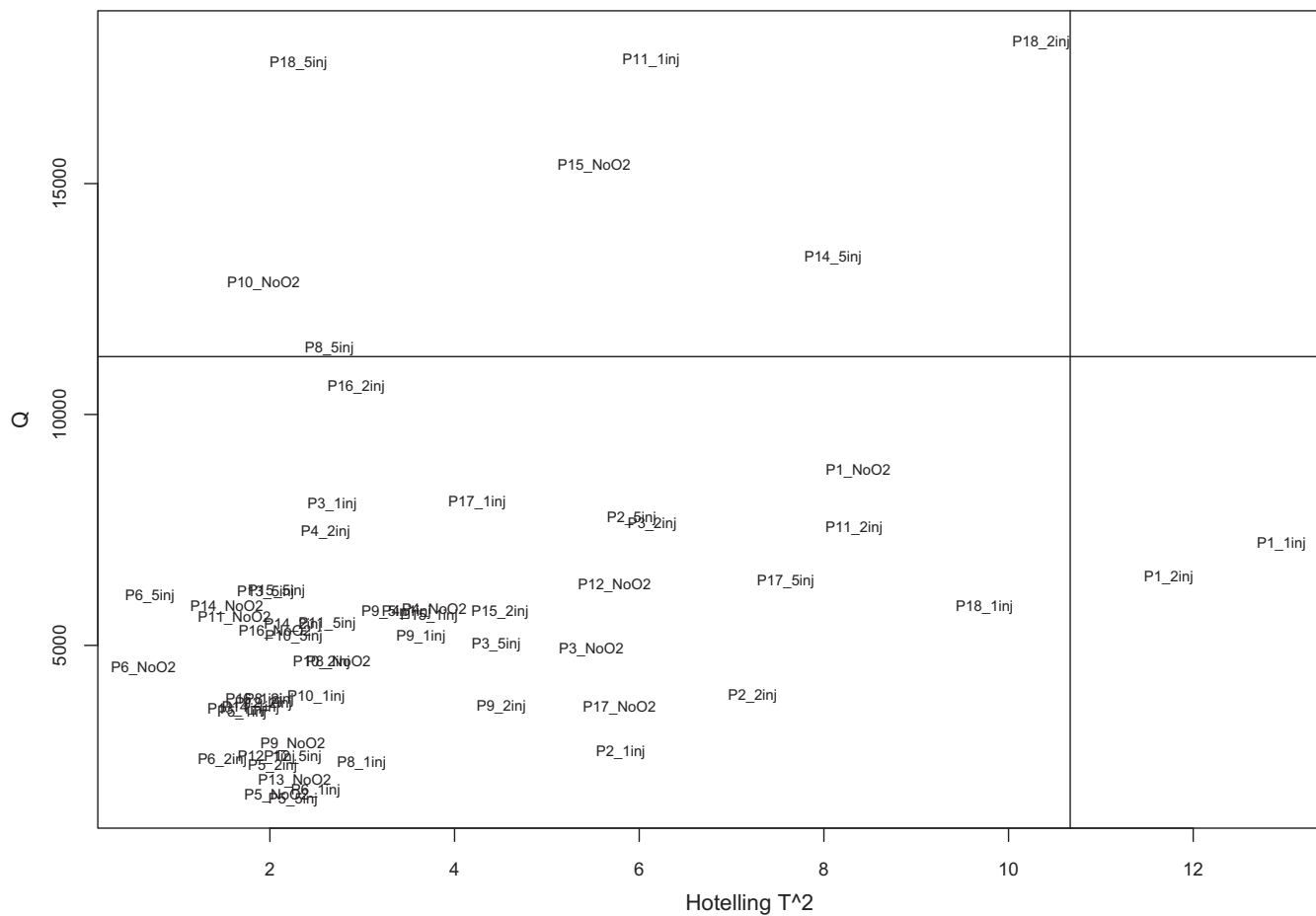


Fig. 2. Diagnostic plot (Hotelling  $T^2$ - $Q$  statistics) of multivariate analysis model.

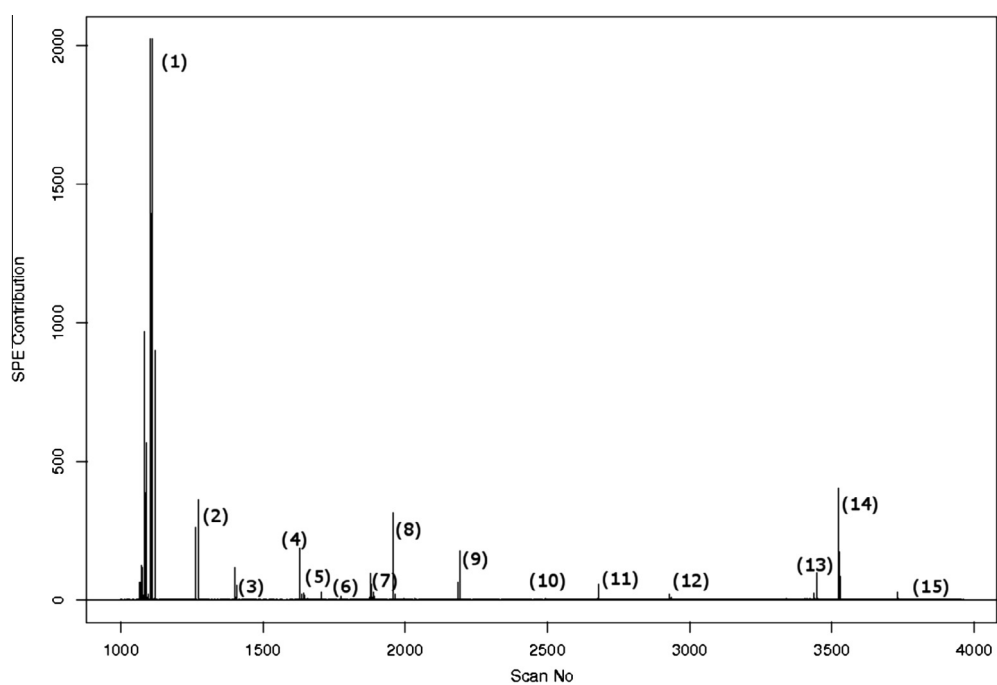


Fig. 3. Contribution plot of sample "P18\_5inj" with relevant candidate scans highlighted.

The mass spectra (MS) fingerprints of the highlighted metabolites of Fig. 3 were built using the identification table from the data matrix resulting from MetAlign™ preprocessing and were used for tentatively identifying the key metabolites [ $m/z$  (%)] that mostly influence the Port wine ageing process. MS features of metabolites tentatively identified as unknown are presented: 1 – 3-methyl-1-butanol; 2 – unknown: unkn\_1 [(100); 43 (67); 55(15); 88 (7); 59 (7)] – Scan 1274; 3 – unknown: unkn\_2 [(100); 43 (21); 91 (13); 119 (12); 75 (5); 59 (5)] – Scan 1403; 4 – unknown: unkn\_3 [43 (100); 115 (46); 55 (41); 45 (30); 59 (29); 73 (21); 39 (21); 67 (19)] – Scan 1629; 5 – furfural; 6 – dioxane isomer; 7 – dioxolane isomer; 8 – unknown: unkn\_4 [73 (100); 45 (33); 91 (15); 43 (14); 55 (7); 74 (6); 57 (4)] – Scan 1959; 9 – unknown: unkn\_5 [(100); 93 (91); 59 (90); 45 (83); 121 (80); 136 (70)] – Scan 2216; 10 – 2-phenylethanol; 11 – diethyl malate; 12 – diethyl tartrate; 13 – monoethyl succinate; 14 – 5-hydroxymethylfurfural (5-HMF).

The presence of furfural, dioxane and dioxolane isomers within the “forced ageing” of Port wine is indicative of the occurrence of wine oxidation during the process, as it was reported by Silva Ferreira et al. (2002) and (Silva Ferreira, Oliveira, Hogg, & Guedes de Pinho, 2003c). According to time-course kinetics of furfural and dioxane isomer, presented in Fig. 4(a) and Fig. 4(b), respectively, concentrations of both metabolites increase during “forced ageing”. Furfural and 5-hydroxymethylfurfural are derived from carbohydrate dehydration followed by cyclation in Maillard-type systems and are generally correlated to wine browning during ageing (Maillard, 1971). It is also evident that higher concentrations of

oxygen during the maturation process induce higher production of this heterocyclic acetal (Fig. 4(b)) and so, the formation of these off-flavors is due to the oxidative degradation of the wine (Silva-Ferreira, Guedes-de-Pinho, Rodrigues, & Hogg, 2002). The presence of 2-phenylethanol and 3-methyl-1-butanol at high concentrations is mainly due to the occurrence of the ethanol fermentation and widely contribute to the aromatic profile of Port wine (Ehrlich, 1907; Hazelwood et al., 2008). Another noteworthy result is the high impact of diethyl esters (diethyl tartrate and diethyl malate) as well as the monoethyl succinate in the Port wine ‘forced’ ageing, which concerns to the extended esterification reactions occurring through the process (Pereira, Reis, Saraiva, & Marques, 2010).

So, according to the chemical background of sample “P18\_5inj”, translated with the variables or candidate metabolites identified in the contribution plot, it is possible to analyze that the presence of such extreme conditions (18 weeks of storage under the highest oxygen saturation regime studied) induces the occurrence of redox mechanisms which are closely associated to the “oxidative spoilage” of the wine during the maturation storage (Silva Ferreira, Oliveira, Hogg, & Guedes de Pinho, 2003c). This mechanism, and thus the presence of the candidate molecules identified, is in most cases associated with the loss of the original Port wine aromas, the development of unpleasant aromas as well as with changes in the Port wine colour or taste (Silva Ferreira et al., 2002). For these reasons, it is possible to consider that “P18\_5inj” sample is not representative of a typical aged Port wine by the conventional ageing process, since the chemical composition evidences the occurrence

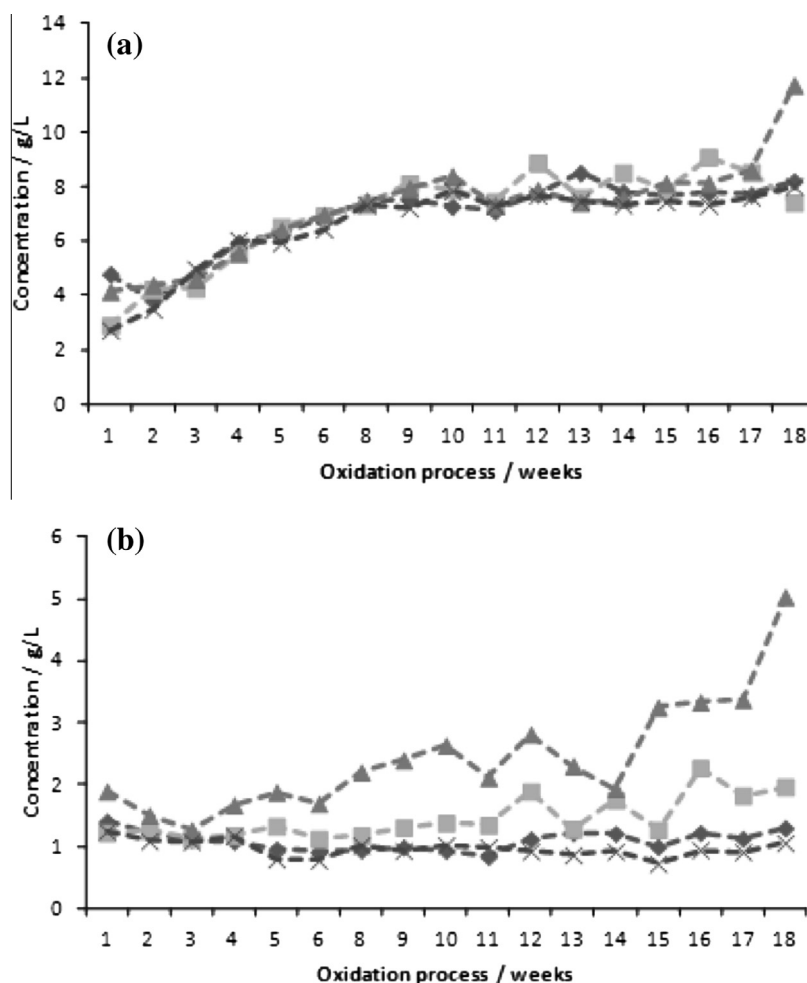
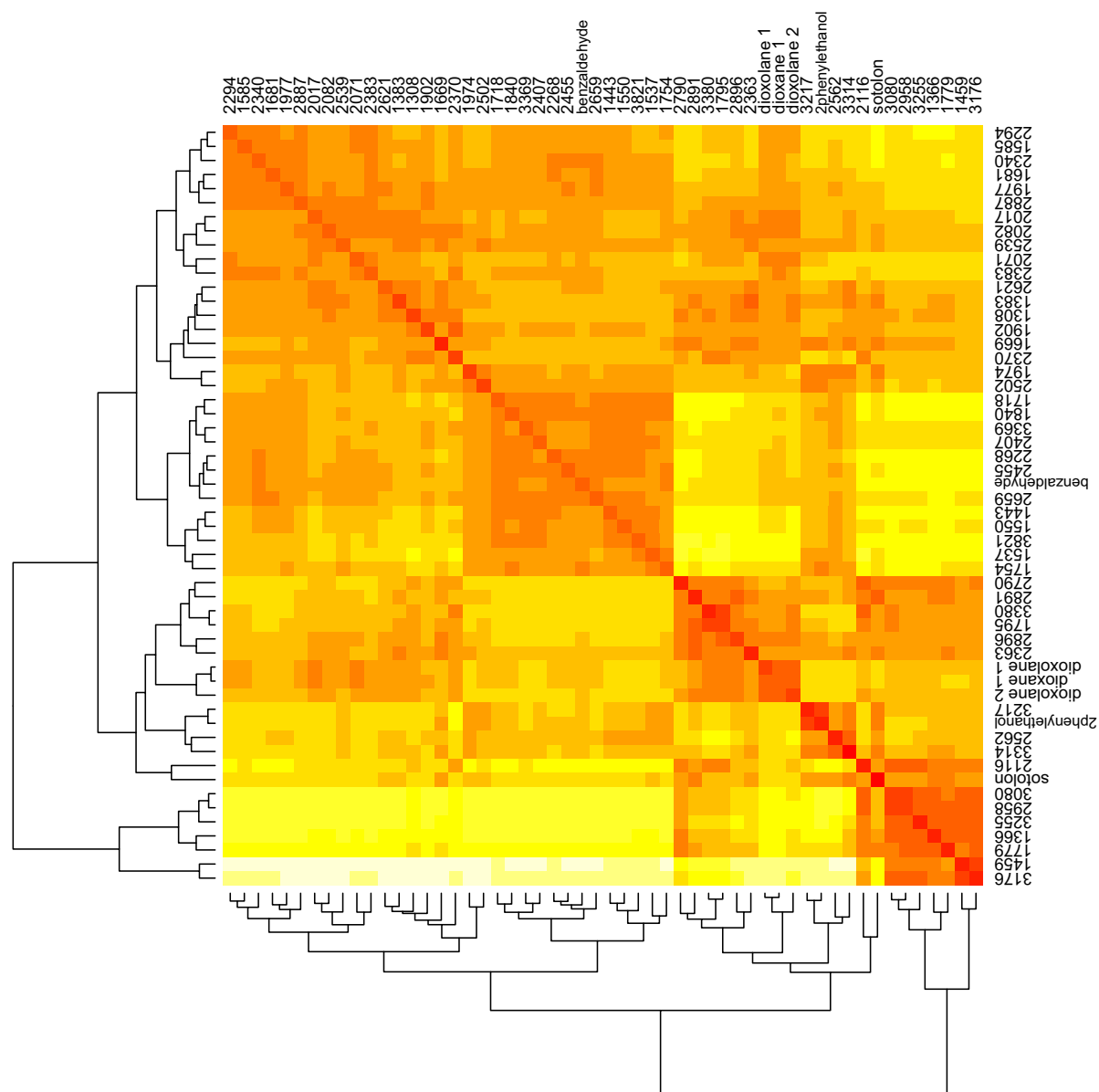


Fig. 4. Time-course kinetics of the highlighted candidate metabolites: (a) furfural; (b) *cis*- and *trans*-5-hydroxy-2-methyl-1,3-dioxane; (x) No O<sub>2</sub>; (◆) 1 Injection O<sub>2</sub>; (■) 2 Injection O<sub>2</sub>; (▲) 5 Injection O<sub>2</sub>.



**Fig. 5.** Heatmap matrix correlations ( $0.8 \leq R^2 \leq 1$  corresponding to white to red range colours) with dioxane isomer. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of deterioration mechanisms inherent to the ageing process occurring under drastic conditions.

Furthermore, regarding the correlation between candidate molecules kinetics over time is an innovative approach that allows increasing the knowledge of the overall process, providing the co-expression, enhancing time-course relationships of the highlighted metabolites and other relevant molecules. The co-expression between compounds allows a kinetical comparison enabling correlations discovery between them, facilitating “mechanistic contextualization”.

### 3.5. Candidate metabolites co-expression

Dioxane isomer, was reported to be one of the indicators of the Port wine ageing (Silva Ferreira et al., 2002), it was used to understand the co-expression of other compounds present in the overall metabolites matrix.

The exploration of dioxane kinetical correlations allows understanding its temporal relationships with other metabolites for further contextualize them within a chemical phenomenon that can

be being activated through the process, which is essential for further metabolic network reconstruction.

Scans corresponding to metabolites with correlation coefficients similar or higher than 0.8 ( $R^2 \geq 0.8$ ) according to dioxane isomer are present in the heatmap of Fig. 5.

Features from scans included in the heatmap of Fig. 5 were also tentatively identified, by supervising with the raw-data chromatogram features and also by comparing the mass feature with NIST 98 MS library (Linstrom et al., 1997), and using the time-expression of scans discriminated in the heatmap.

So, according to Fig. 5, it was found that the production of dioxane isomer, is highly correlated with the production of dioxolane isomers – dioxolane 1 ( $R^2 = 0.98$ ), dioxolane 2 ( $R^2 = 0.99$ ), benzaldehyde ( $R^2 = 0.89$ ) and sotolon ( $R^2 = 0.81$ ). The high correlations between these compounds is in agreement with the reported bibliography, as all these metabolites are known to be related with the Port wine ageing process (Silva Ferreira et al., 2002; ; Silva Ferreira et al., 2003a; ; Silva Ferreira et al., 2008). Dioxane isomer production is a result of the condensation reaction between glycerol and acetaldehyde under the low pH during wine ageing (Müller, Kepner, & Webb, 1978).

So, considering the consistency of the high correlations found, in consonance to the present pipeline methodology, it is of great interest the exploration of the entire heatmap correlations, improving compounds identification, contextualizing metabolites itemized and providing a different dimension in the understanding of the Port wine “forced ageing” process.

#### 4. Conclusions

In this research, was presented a metabolic pipeline methodology which allowing great progress in identifying metabolites involved in dynamic processes in a high-throughput mode. This pipeline will enable the network reconstruction, starting with a set of known compounds clustered together according to their kinetical correlations and representing a specific chemical pathway.

Applying the present pipeline to a Port wine “forced ageing” process, furfural and dioxane isomer were found to be key molecules involved in the process, which production is relevant for the quality of the Port wine. Moreover, dioxane isomer production is highly correlated with dioxolane isomers, dioxane isomer, benzaldehyde and sotolon production, which have been already reported in other studies, as being result of physical and chemical changes occurring over the wine ageing process.

Based on the temporal relationships between a known compound and the entire matrix, it was possible to understand potential interconnections between the candidates metabolites formation. Future researches of connections between chemical and/or biochemical mechanisms would provide chemical responses according to the interest on a specific process.

This study revealed several advantages leading to the use of the metabolomics pipeline methodology, encouraging also its application in different research fields, as it provides a greater, faster and reliable understanding of biological or chemical processes occurring through time-course.

#### Acknowledgements

The author Castro C.C. (SFRH/BD/46737/2008) gratefully acknowledges her Doctoral grant to the Fundação para a Ciência e Tecnologia (FCT). This research was funded by the projects PTDC/BIO/69310/2006, PTDC/AGR-ALI/121062/2010, FCOMP-01-0124-008775 through FCT, and partially supported by CBMA, IBB and ESB/UCP plurianual funds through the POS-Conhecimento Program that includes FEDER funds through the program COMPETE (Programa Operacional Factores de Competitividade).

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.foodchem.2013.07.138>.

#### References

- Castillo, S., Gopalacharyulu, P., Yetukuri, L., & Oresic, M. (2009). Algorithms and tools for the preprocessing of LC–MS metabolomics data. *Chemometrics and Intelligent Laboratory Systems*, 108(1), 23–32.
- Castro, C. C., Silva, J. S., Lopes, V. V., & Martins, R. C. (2009). Yeast metabolic state identification by fiber optics spectroscopy. In: Applications of optics and photonics, Portugal, (pp. 8001–8202).
- Castro, C. C., Gunning, C., Oliveira, C. M., Couto, J. A., Teixeira, J. A., Martins, R. C., et al. (2012). *Saccharomyces cerevisiae* oxidative response evaluation by cyclic voltammetry and gas chromatography–mass spectrometry. *Journal of Agricultural and Food Chemistry*, 60, 7252–7261.
- Christensen, J. H., Tomasi, G., & Hansen, A. B. (2005). Chemical fingerprinting of petroleum biomarkers using time warping and PCA. *Environmental Science and Technology*, 39(1), 255–260.
- Conlin, A., Martin, E., & Morris, A. (2000). Confidence limits for contribution plots. *Journal of Chemometrics*, 14, 725–736.
- Conlin, A. K., Martin, E. B., & Morris, A. J. (2002). Confidence limits for contribution plots. *Chemometrics and Intelligent Laboratory Systems*, 14, 725–736.
- Cozzolino, D., Cynkar, W. U., Shah, N., Damberg, R. G., & Smith, P. A. (2009). A brief introduction to multivariate methods in grape and wine analysis. *International Journal of Wine Research*, 1, 123–130.
- Dunn, W. B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., et al. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols*, 6(7), 1060–1083.
- Ehrlich, F. (1907). Über die Bedingungen der Fuselölbildung und über ihren Zusammenhang mit dem Eiweissaufbau der Hefe. *Berichte der Deutschen Chemischen Gesellschaft*, 40(1), 1027–1047.
- Feist, A. M., Herrgård, M. J., Thiele, I., Reed, J. L., & Palsson, B. Ø. (2009). Reconstruction of biochemical networks in microorganisms. *Nature Reviews Microbiology*, 7(2), 129–143.
- Fiehn, O. (2002). Metabolomics – The link between genotypes and phenotypes. *Plant Molecular Biology*, 48(1), 155–171.
- Hazelwood, L. A., Daran, J. M., van Maris, A. J., Pronk, J. T., & Dickinson, J. R. (2008). The Ehrlich pathway for fusel alcohol production: A century of research on *Saccharomyces cerevisiae* metabolism. *Applied and Environmental Microbiology*, 74(8), 2259–2266.
- Hofmann, T., & Schieberle, P. (2000). Formation of aroma-active Strecker-aldehydes by a direct oxidative degradation of Amadori compounds. *Journal of Agricultural and Food Chemistry*, 48(4301–4305).
- Hotteling, H. (1947). *Multivariate Quality Control Illustrated by the Air Testing of Sample Bombsites*. (p. 111): New York: McGraw-Hill.
- Jerez, C. A. (2008). The use of genomics, proteomics and other OMICS technologies for the global understanding of biomining microorganisms. *Hydrometallurgy*, 94, 162–169.
- Kourti, T., & MacGregor, J. F. (1996). Multivariate SPC methods for process and product monitoring. *Journal of Quality Technology*, 28, 409–427.
- Linstrom, P. J., Mallard, W. G., National Institute of, S., & Technology (1997). *NIST Chemistry WebBook*. National Institute of Standards and Technology Gaithersburg, MD.
- Lommen, A. (2009). MetAlign: Interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. *Analytical Chemistry*, 81(8), 3079–3086.
- Maillard, B. (1971). Additions radicalaires de diols et de leurs dérivés: diesters et acétals cycliques. vol. Ph.D Thesis (p. 325). Bordeaux: University Bordeaux 1 - France.
- Martins, R. C., Lopes, V. V., Vicente, A. A., & Teixeira, J. A. (2008). Computational shelf-life dating: Complex systems approaches to food quality and safety. *Food and Bioprocess Technology*, 1(3), 207–222.
- Mason, R. L., Tracy, N. D., & Young, J. C. (1995). Decomposition of T2 for multivariate control chart interpretation. *Journal of Quality Technology*, 27(2), 99–1108.
- Mason, R. L., & Young, J. C. (1999). Improving the sensitivity of the T2 statistic in multivariate process control. *Journal of Quality Technology*, 31(2), 155–165.
- Müller, C., Kepner, R., & Webb, A. (1978). 1,3-dioxanes and 1,3-dioxolanes as constituents of the acetal fraction of Spanish fino sherry. *American Journal of Enology and Viticulture*, 29, 207–212.
- Pereira, A. C., Reis, M. S., Saraiva, P. M., & Marques, J. C. (2010). Aroma trends in GC/MS profiles of liqueur wines. *Analytica Chimica Acta*, 659, 93–101.
- Qin, S. (2003). Statistical process monitoring: basics and beyond. *Journal of Chemometrics*, 17, 480–502.
- Rodgers, J. L., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *American Statistical Association*, 42(1), 59–66.
- Rodrigues, J. A., Barros, A. S., Carvalho, B., Brandão, T., Gil, A. M., & Ferreira, A. C. S. (2011). Evaluation of beer deterioration by gas chromatography–mass spectrometry/multivariate analysis: A rapid tool for assessing beer composition. *Journal of Chromatography A*, 1218(7), 990–996.
- Silva Ferreira, A. C., Barbe, J. C., & Bertrand, A. (2002). Heterocyclic acetals from glycerol and acetaldehyde in Port wines: Evolution with ageing. *Journal of Agricultural and Food Chemistry*, 50(9), 2560–2564.
- Silva Ferreira, A. C., Barbe, J. C., & Bertrand, A. (2003a). 3-Hydroxy-4,5-dimethyl-2(5H)-furanone: A key odorant of the typical aroma of oxidative aged Port wine. *Journal of Agricultural and Food Chemistry*, 51(15), 4356–4363.
- Silva Ferreira, A. C., Hogg, T., & Guedes de Pinho, P. (2003b). Identification of key odorants related to the typical aroma of oxidation-spoiled white wines. *Journal of Agricultural and Food Chemistry*, 51(5), 1377–1381.
- Silva Ferreira, A. C., Monteiro, J., Oliveira, C., & Guedes de Pinho, P. (2008). Study of major aromatic compounds in port wines from carotenoid degradation. *Food Chemistry*, 110(1), 83–87.
- Silva Ferreira, A. C., Oliveira, C., Hogg, T., & Guedes de Pinho, P. (2003c). Relationship between potentiometric measurements, sensorial analysis, and some substances responsible for aroma degradation of white wines. *Journal of Agricultural and Food Chemistry*, 51(16), 4668–4672.
- Silva-Ferreira, A., Guedes-de-Pinho, P., Rodrigues, P., & Hogg, T. (2002). Kinetics of oxidative degradation of white wines and how they are affected by selected technological parameters. *Journal of Agricultural and Food Chemistry*, 50(21), 5919–5924.
- Sinkhorn, R. (1964). A relationship between arbitrary positive matrices and doubly stochastic matrices. *The Annals of Mathematical Statistics*, 35(2), 876–879.
- Villas-Boas, S. G., Mas, S., Akesson, M., Smedsgaard, J., & Nielsen, J. (2005). Mass spectrometry in metabolome analysis. *Mass Spectrometry Reviews*, 24(5), 613–646.
- Walczak, B., & Wu, W. (2005). Fuzzy warping of chromatograms. *Chemometrics and Intelligent Laboratory Systems*, 77(1–2), 173–180.