A Metabolomic Approach using PARAFAC2 for the Identification of Volatile Markers for Vineyard Age, Ripening Stage and Storage Temperature in White Wines

IFAPA Instituto de Investigación y Formación Agraria y Pesquera

José M. Muñoz-Redondo[1], M. Victoria Jiménez-Povedano[2], Emma Cantos-Villar[2], Belén Puertas[2], Manuel J. Valcárcel-Muñoz[3], Beltrán Peña-Parra[3], José M. Moreno-Rojas[1]

[1]Department of Food Science and Health - Andalusian Institute of Agricultural and Fisheries Research and Training (IFAPA), Avenida Menéndez-Pidal, SN, 14004 Córdoba, Spain
[2]Department of Food Science and Health - Andalusian Institute of Agricultural and Fisheries Research and Training (IFAPA), Ctra. Cañada de la Loba PK 3.1 (CA-3102), 11471 Jerez de la Frontera, Spain
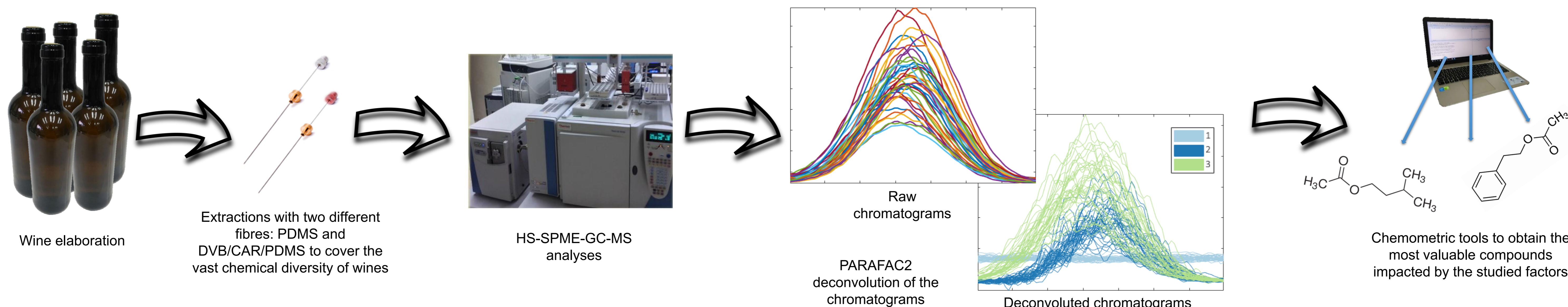[3]Bodegas Fundador, S.L.U., C/ San Ildefonso, 3, 11403 Jerez de la Frontera, Spain

INTRODUCTION

The recent revolution of the called "omic" platforms (transcriptomics, proteomics, metabolomics) allows for the simultaneous measurement of hundreds to thousands of molecular quantities. Thus, the focus on univariate classical statistic methodologies, where variables are studied one at a time, has already been shifted towards multivariate methods. Thus, a holistic point of view of the target samples can be obtained and the interaction between variables can be highlighted. In chromatographic techniques coupled to mass spectrometry detectors the aim is to identify the compounds presents in samples and compare their concentrations. Due to the often complexity of the chromatograms (such as coelution peaks, baseline contributions or low signal-to-noise peaks) and the large number of variables, processing of the data becomes a challenging and a time consuming task[1]. A wide range of softwares are available to automatically process the chromatograms as result of gas chromatography-mass spectrometry performances. PARADISE[1] is a recently developed software that implement PARAllel FACtor analysis2 (PARAFAC2) for deconvolution of the chromatograms, that has demonstrated its usefulness handling coeluted and retention time shifts[2]. The aim of this work was to study the effect of the vineyard age, ripening stage of grapes and storage temperature on the volatile composition of Palomino Fino white wines elaboration. The goal was to extract the maximum information from the samples by means of PARAFAC2 deconvolution treatment of the chromatographic data and chemometric tools.

MATERIALS AND METHODS

Samples of Palomino Fino white wines were elaborated in a factorial design. Three factors were studied to evaluate their impact on the volatile composition of the wines: vineyard age (4, 13 and 21 years), maturity of grapes from which the wines were elaborated (8.5, 10.5 and 12.5 °Baumé) and storage temperature of the wines once bottled (refrigerated and room temperature). The storage temperature was only assessed in the wines produced from the 12.5 °Baumé grape maturity wines. Analyses were performed using head space solid phase micro-extraction method (HS-SPME) followed by gas chromatography – mass spectrometry (GC–MS) in TIC mode, recording in the range of 50-450 m/z. Deconvolution of the chromatograms was accomplished using PARAFAC2 by means of PARADISE software. To explore the data Principal Component Analyses (PCA) were built. Afterwards, Partial Least Squares Discriminant Analysis (PLS-DA) was used to construct classification models. The variable importance in projection (VIP) values obtained from the PLS-DA models were used in an iterative way as previously described[3] to perform selection of the most informative variables (compounds). All the discriminant models were optimized and validated in a double cross-validation scheme based on the number of misclassification (NMC) and the area under the receiver operator curve (AUROC) as statistic diagnostics. For each classification task, 30 submodels were obtained as previously recommended[4]. Permutation test was carried out to validate de models.

Wine elaboration — Extractions with two different fibres: PDMS and DVB/CAR/PDMS to cover the vast chemical diversity of wines — HS-SPME-GC-MS analyses — Raw chromatograms — PARAFAC2 deconvolution of the chromatograms — Deconvoluted chromatograms — Chemometric tools to obtain the most valuable compounds impacted by the studied factors
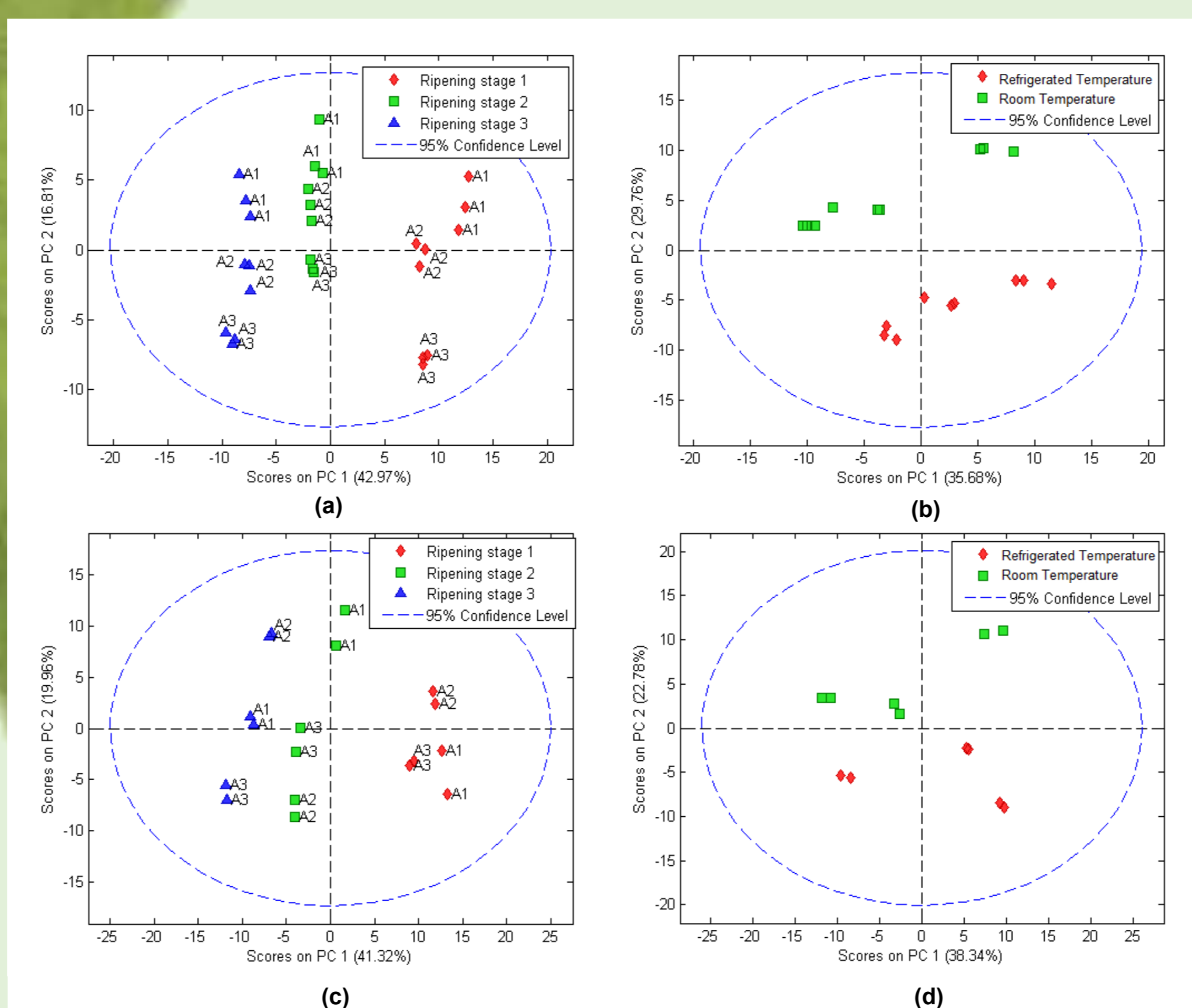
RESULTS

The methodology previously described allowed us to study in deep the modulation of the volatile composition of Palomino Fino white wines in a factorial design (vineyard age, grape ripening stage and storage temperature). In the Figure 1, PCA models were built to explore the data. Samples were clearly separated based on the different factors. Subsequently, PLS-DA models were performed to the observed differences and to carry out variable reductions. In all cases, the previously described validation of the PLS-DA models was satisfactory (Table 1). The variable reduction by means of an iterative VIP selection during the optimization and validation of the models allowed us to obtain candidate marker compounds for each factor.

Table 1. Performance of the discriminant models on the basis of 30 values obtained from 30 submodels for the two diagnostic statistics used for optimization and validation: number of missclasifications (NMC) and area under the receiver operating characteristic curve (AUROC) for PDMS and DVB/CAR/PDMS fibres.

	Model	NMC min	NMC max	NMC mean	AUROC min	AUROC max	AUROC mean
PDMS	Vineyard Age	0	0	0	1	1	1
PDMS	Ripening Stage	0	0	0	1	1	1
PDMS	Storage Temperature	0	0	0	1	1	1
DVB/CAR/PDMS	Vineyard Age	0.0	2.0	1.1	0.973	1.000	0.997
DVB/CAR/PDMS	Ripening Stage	0.0	1.0	0.1	0.993	1.000	0.999
DVB/CAR/PDMS	Storage Temperature	0	0	0	1	1	1

Figure 1. Principal Component Analyses (PCA) performed for the PDMS fibre (top) and DVB/CAR/PDMS (bottom) for vineyard age and ripening stage of the grapes (a) and (c) and for storage temperature factor (b) and (d). A1, A2, A3: vineyard age corresponding to 4, 13 and 21 years respectively. Ripening stage 1, 2 and 3: wines elaborated from grapes harvested at 8.5, 10.5 and 12.5 °Baumé respectively

CONCLUSIONS

The HS-SPME-GC-MS methodology was successfully applied alongside PARAFAC2 deconvolution and chemometric tools to study differences across white wine samples. We observed variations in the volatile profile of the wines and related to the three factors under study. The use of PARADISE for building the PARAFAC2 models allowed a fast and easy processing of the deconvoluted chromatograms. The variable reduction allowed us to highlight the most impacted compounds by each factor. This methodology proved to be suitable for the identification of subtle information of the complex datasets used in this study.

REFERENCES
1. Johnsen, L. G., Skou, P. B., Khakimov, B., & Bro, R. Journal of Chromatography A, 1503, 57-64 (2017).
2. Amigo, J. M., Skov, T., Bro, R., Coello, J., & Maspoch, S. TrAC Trends in Analytical Chemistry, 27(8), 714-725 (2008).
3. Khakimov, B., Poulsen, S. K., Savorani, F., Acar, E., Gürdeniz, G., Larsen, T. M., ... & Engelsen, S. B. Journal of proteome research, 15(6), 1939-1954 (2016).
4. Szymańska, E., Saccenti, E., Smilde, A. K., & Westerhuis, J. A. Metabolomics, 8(1), 3-16 (2012).

MINISTERIO DE ECONOMÍA, INDUSTRIA Y COMPETITIVIDAD — CDTI Centro para el Desarrollo Tecnológico Industrial — Instituto de Investigación y Formación Agraria y Pesquera, CONSEJERÍA DE AGRICULTURA, PESCA Y DESARROLLO RURAL — UNIÓN EUROPEA Fondo Europeo de Desarrollo Regional (FEDER) Una manera de hacer Europa — bestBrandy