Talanta 176 (2018) 526–536

Contents lists available at ScienceDirect

Talanta

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

Chemometric compositional analysis of phenolic compounds in fermenting samples and wines using different infrared spectroscopy techniques

Jose Luis Aleixandre-Tudo[a,b,*], Helene Nieuwoudt[c], Jose Luis Aleixandre[a], Wessel du Toit[b]

[a] Departamento de Tecnologia de Alimentos, Universidad Politécnica de Valencia, Camino de Vera s/n, 46022 Valencia, España
[b] Department of Viticulture and Oenology, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa
[c] Institute for Wine Biotechnology, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa

ARTICLE INFO

Keywords:
Infrared spectroscopy
Phenolic compounds
Chemometrics
Anthocyanins
Tannins

ABSTRACT

The wine industry requires reliable methods for the quantification of phenolic compounds during the winemaking process. Infrared spectroscopy appears as a suitable technique for process control and monitoring. The ability of Fourier transform near infrared (FT-NIR), attenuated total reflectance mid infrared (ATR-MIR) and Fourier transform infrared (FT-IR) spectroscopies to predict compositional phenolic levels during red wine fermentation and aging was investigated. Prediction models containing a large number of samples collected over two vintages from several industrial fermenting tanks as well as wine samples covering a varying number of vintages were validated. FT-NIR appeared as the most accurate technique to predict the phenolic content. Although slightly less accurate models were observed, ATR-MIR and FT-IR can also be used for the prediction of the majority of phenolic measurements. Additionally, the slope and intercept test indicated a systematic error for the three spectroscopies which seems to be slightly more pronounced for HPLC generated phenolics data than for the spectrophotometric parameters. However, the results also showed that the predictions made with the three instruments are statistically comparable. The robustness of the prediction models was also investigated and discussed.

1. Introduction

The highly competitive global wine market is currently demanding top quality products. Internationally wine producers are facing the challenge of an increasingly competitive international scenario [1,2]. The inclusion of state of the art analytical technologies to ensure high quality standards and process control is thus a priority [3,4]. Analytical technologies combine several analytical tools which include physical, chemical, mathematical, statistical and other analytical resources to provide a holistic insight into product properties. The information obtained can thus be beneficial for benchmarking, decision making, grading, process control, adulteration or geographical identification tasks, among others [5–7].

The use of spectroscopy with chemometrics combines several of these tools. Spectroscopy has been declared as suitable for process control and monitoring [4,8–10]. The use of infrared spectroscopy (IR) relies on the molecular overtones and vibrations of the atoms when infrared radiation is passed through a sample. The amount and frequency of the absorbed light as well as the amount of reflected or transmitted light provide information of the grape and wine biochemical components. In addition, IR has been defined as a non-destructive, fast and easy to perform analytical technique [10,11]. The fact that it can measure more than one parameter at a time makes it the analytical technology of preference in food-related and non-related industries [6,12]. In the past years an increased availability of IR instruments and applications, including quantification and discrimination tasks, have been reported [10], nevertheless its industrial implementation seems to be slow and only possible to medium and large size wineries [6].

Phenolic compounds in combination with other major wine constituents are mainly responsible for the mouth feel attributes of a red wine [13,14]. Moreover, the colour properties of a wine depend on the levels and chemical state of the phenolic compounds present at the time of evaluation [15,16]. Phenolic compounds are extracted during the fermentation mainly from the solid parts of the grape berry [17]. However, the level of these compounds is not the main factor contributing to their later presence in wine. The interactions and associations among phenolic substances, which occur as soon as the compounds coexist in the must, influence their further presence and consequently their contribution to the wine organoleptic properties [18,19].

* Corresponding author at: Departamento de Tecnologia de Alimentos, Universidad Politécnica de Valencia, Camino de Vera s/n, 46022 Valencia, España.
E-mail addresses: joaltu@upvnet.upv.es, joaltu@sun.ac.za (J.L. Aleixandre-Tudo).

http://dx.doi.org/10.1016/j.talanta.2017.08.065
Received 12 May 2017; Received in revised form 17 August 2017; Accepted 20 August 2017

A number of studies aiming to monitor phenolic compounds during the fermentation process using infrared spectroscopy have been reported in the literature. Visible-Near infrared spectroscopy was explored by Cozzolino [20]. Only calibration statistics for malvidin-3-glucoside, polymeric pigments and tannins models using HPLC analysis as reference method were reported in Cabernet Sauvignon and Shiraz fermentations collected over two vintages. Additionally, Fragoso [21] also reported quantification models for fermenting samples using Fourier Transform (FT) Mid-infrared with the spectrophotometric determination of total phenolics (TP), anthocyanins (TA) and methylcellulose precipitable (MCP) tannins as reference methods. Five different cultivars were included in the trail and microvinifications (4 kg) were performed at different ripening levels. Samples were collected during fermentation for 10 days. Validation residual predictive deviation (RPD) values higher than 3 for TP and TA and lower than 2.5 for MCP tannins were reported. Finally, di Egidio [8] investigated the use of Near- (NIR) and Mid-infrared (MIR) to monitor the levels of TP, TA and total flavonols (TF) in 15 micro-vinifications during the fermentation process of Nebbiolo grapes. Good calibration models were reported, however the lack of the standardization of the predictive accuracy makes it difficult to compare the results with those reported in other studies. The combination of ultraviolet-visible and near infrared (UV–VIS–NIR) spectroscopy was also investigate for some of the most representative phenolic compounds [22]. Accurate single cultivar models were observed for catechin and malvidin, however due to the limited number of samples no validation data was reported.

NIR and IR (WineScan™, Foss Electric) spectra in transmission mode as well as attenuated total reflectance (ATR) MIR spectra in reflexion mode was collected from a large set of fermenting samples and wines. Prediction models were built for 27 individual phenolic compounds quantified using an HPLC method as well as for the spectrophotometric determinations TP, TA, MCP tannins and colour density (CD). The aim of this study was thus to provide accurate externally validated prediction models for phenolic monitoring, quantification and profiling during the winemaking process. The goal of comparing three spectroscopic techniques relies on the identified need of providing an increasing number of applications to scientists and professionals. Despite the published studies, a direct comparison between three different spectroscopic techniques has not yet been investigated. An additional statistical treatment of the predictions obtained with the different instruments is reported in this study. The suitability of each technique has been evaluated based on the results obtained from the process of model calibration and validation.

2. Materials and methods

2.1. Reagents and standards

Phosphoric acid and caffeic acid were purchased from Fluka (Sigma-Aldrich Chemie, Steinheim, Germany). Acetonitrile was obtained from Merck (Darmstadt, Germany). Methyl cellulose, ammonium sulphate, hydrochloric acid (HCl), gallic acid, catechin, p-coumaric acid, quercetin-3-glucoside and quercetin were obtained from Sigma-Aldrich Chemie, (Steinheim, Germany) and malvidine-3-glucoside chloride was purchased from Extrasynthese (Lyon, France).

2.2. Samples

Samples during the fermentation process were collected from 13 commercial scale vinifications at the Welgevallen cellar (Stellenbosch, South Africa) over two consecutive vintages (2015–2016). Nine different fermentations were followed in 2015 and four fermentations were sampled in 2016. Four cultivars were represented including Cabernet Sauvignon, Shiraz, Pinotage and Grenache. Samples were collected daily the first 15 days of the fermentation and every 3 days for a maximum period of two months after fermentation. Samples were passed through a kitchen sieve and frozen immediately after collection. Varying phenolic and sugar ripening levels, cold maceration, the use of different yeast strains, extended maceration, tannin addition and malolactic fermentation in barrel were some of the winemaking variables included in the sample set. Fermentations took place in different fermenters, ranging from 3.000 to 10.000 L. A total of 391 samples were collected. The day of analysis, the samples were thawed at room temperature and centrifuged in a 7366 Hermle centrifuge (Wehingen, Germany) at 3248 g for 5 min before spectra collection or analysis were performed. Additionally, wine samples (178) spanning a range of vintages (from 2005 to 2016) and cultivars (12) as well as some blends were also collected and analysed. Before analysis the samples were also centrifuged at 3248 g for 5 min. A total number of 569 samples including fermenting samples and wines were used in the calibrations.

2.3. Spectrophotometric analysis of phenolic compounds

The method reported by Iland [23] was used for the quantification of total anthocyanin and total phenolic content. The samples were diluted 50 times with HCl 1 M and kept for 3 h before the absorbance at 280 nm and 520 nm was recorded using a Multiskan GO Microplate Spectrophotometer (Thermo Fisher Scientific, Inc., Waltham, MA, USA). The total phenolics index was calculated as the A280nm times the dilution factor (DF = 50). The anthocyanins content (mg/L malvidin-3-glucoside) was calculated using the molar extinction coefficient (ε) and the molecular weight (MW) of the most abundant anthocyanin found in wine malvidin-3-glucoside (ε = 28.000 L/cm*mol; MW = 529 g/mol) times the dilution factor. The total tannin content was quantified using the methyl cellulose (MCP) tannin assay developed by Sarneckis [24] and later adapted to a high throughput format by Mercurio [25]. Briefly the method consists of a polymer-tannin interaction that results in an insoluble complex that precipitates and can be measured by comparing a control sample and a treatment sample (where tannins are removed by centrifugation). In a 2 mL microfuge tube, 600 μL of MCP solution was mixed with 50 μL of wine and let to stand for 2–3 min. 400 μL of saturated ammonium sulphate solution and 950 μL of distilled water were added to a final volume of 2 mL (treatment sample). A control sample with distilled water (600 μL) instead of MCP solution was also prepared. Samples were left for 10 min prior to centrifugation in an Eppendorf 5415D centrifuge (Hamburg, Germany) at 9279 g for 5 min. The difference between the control A280 nm value and the treatment A280 nm value was converted into epicatechin equivalents (mg/L) using a calibration curve. A dilution factor of 40 was used to calculate the total tannin content. The methylcellulose solution (0.04% w/v; 1500 cP viscosity at 2%) was prepared according to the method's instructions [25]. Colour density was measured according to Glories [26]. Fifty μL of wine were directly pipetted into a UV–VIS Nunc F96 MicroWell plate (Nunc, Langenselbold, Germany) and the absorbance at 420 nm, 520 nm and 620 nm were recorded. The colour density was calculated as the sum of the three wavelengths. Absorbance values were always referenced to a standard 10 mm path length.

2.4. HPLC analysis of phenolic compounds

The phenolic composition was analysed following the method initially reported by Peng [27] with some modifications. An Agilent Technologies 1260 Infinity series (Agilent, Waldbronn, Germany) HPLC system with a PLRP-S polymeric reversed phased column (3 μm particle size, 100 Å pore size, 150 mm × 4.6 mm), at 35 °C was used for the quantification of phenolic compounds. The solvents used were 100% acetonitrile (A) and phosphoric acid in water at 1.5% (B) for a gradient elution flow rate of 1 mL/min: 0 min (5% solvent B), 73 (25% solvent B), 78 (50% solvent B), 86 (50% solvent B),

90 (5% solvent B), with a re-equilibration time of 15 min. The injection volume was 20 μL and the phenolic compounds were detected by a photodiode array detector at 280 (gallic acid, monomeric and polymeric flavan-3-ols), 320 (phenolic acids), 360 (flavonols) and 520 nm (anthocyanins and polymeric pigments). External calibrations were obtained for gallic acid, catechin, caffeic acid, p-cumaric acid, quercetin-3-glucoside, quercetin and malvidin-3-glucoside. When the standards were not available the phenolic compounds were quantified using the available calibration curve of the compounds that correspond to the same phenolic family e.g. all the anthocyanins were quantified using malvidin-3-glucoside as standard.

2.5. Infrared scanning

Prior scanning samples were centrifuged as indicated in the previous section. Spectra were collected using a Multi-Purpose analyser (MPA) FT-NIR instrument (Bruker Optics, Ettlingen, Germany) in transmission mode. Samples were placed in a 1 mm cuvette and the spectra were averaged from 64 scans and collected in absorbance mode at 2 cm^{-1} resolution, at 10 kHz scanner velocity in the wavenumber range of 12,500–4000 cm^{-1}. An air background spectrum was measured at regular intervals during spectra acquisition and automatically subtracted from each sample spectrum. ATR FT-MIR spectra were also acquired. Each sample was scanned on an Alpha-P ATR FT-MIR spectrometer (Bruker Optics, Ettlingen, Germany), fitted with a 2 mm^2 single bounce diamond crystal sample plate. Spectra collection was done under the following scanning parameters: 4000–600 cm^{-1} wavenumber range, 4 cm^{-1} resolution, 7.5 kHz scanner velocity, and 64 sample scans at constant temperature of 30 °C. The instrumental control was carried out using OPUS software (OPUS v. 7.0 for Microsoft, Bruker Optics, Ettlingen, Germany). Finally, FT-IR spectral measurements were also performed using a WineScan™ FT120 instrument (Foss Electric, Denmark). Samples were scanned from 5011 to 929 cm^1 (which includes a small section of the near-IR region) at 4 cm^{-1} resolutions, and 20 scan per sample at 40 °C. FT-IR spectra were recorded in transmittance mode, and converted to a linearized absorbance spectrum [28]. Background absorbance in the wine sample is corrected through the use of a Foss Zero Liquid S-6060 (WineScan™ manual), which is scanned prior to the wine samples.

2.6. Development and validation of the prediction models

Data analysis and model performance was evaluated using OPUS software (OPUS v. 7.2 for Microsoft, Bruker Optics, Ettlingen, Germany). Initially, cross validated models were optimized using the general B option included in the software package. The advantage of this option is that the spectral range is divided into 10 sub regions. The software builds models using single or combinations of regions until the best model performance (lower error) is provided. Twenty samples were left out and used as internal validation in the cross validated optimized models. A large number of different spectral pre-processing options, including smoothing, standardization, transformation and normalization, were investigated. The number of latent variables (rank) was defined based on the predicted residual error sum of squares (PRESS) that is not significantly larger than the minimum. After the identification of the minimum PRESS a quotient of the lower rank PRESS values and the minimum is calculated. From this value a probability is calculated and the first rank having a probability smaller than 0.75 is selected as the optimum rank. Once the CV model was optimized the sample set was divided into calibration and validation sets at 50/50 and 68/32 ratios respectively. Models were validated using the best spectral pre-processing method as indicated in the model optimization step.

The accuracy of the prediction models was investigated using the following statistics. The coefficient of correlations in calibration (R^2cal) and validation (R^2val) which estimate the percentage of variation explained by the model in the training and validation set, respectively. A R^2 close to 1 is a necessary condition for a good model but not the only requirement. Poor reproducibility or the lack of correlation between the spectral information and the chemistry lead to low R^2 values. The errors in calibration (root mean square error in cross-validations (RMSECV)) and validation (root mean square error of prediction (RMSEP)) indicate the fit of the observations to the model in both calibration and validation steps. RMSECV measures the average difference between the values determined by the reference methods and those predicted by the model. The cross validation (CV) approach leaves out a specified number of samples that are later used for the internal calibration of the model. This procedure is repeated until all the samples have been left out at least once. It provides the potential error of future samples not included in the calibration set. The RMSEP measures the average difference between the values predicted by the model and values obtained from the reference analytical methods during validation of new samples. It gives the real average prediction error. Both the RMSECV and RMSEP are expressed in the same units as the reference values and are influenced by the range of phenolic levels. With the aim to standardize the prediction accuracy and avoid the effect of range in the cross validated and prediction errors, the residual predictive deviation is also reported. RPDcal and RPDval are defined as the ratios of the standard deviation of the calibration/validation set and the root-mean-square error of cross validation/prediction (RPD = SD/RMSE). High standard deviations in combination with small prediction errors provide relatively high RPD. The higher the RPD values the greater the probability of the model to accurately predict the levels of phenolic compounds in new samples.

Additionally, a regression test to compare data obtained from different sources was also performed with the aim to investigate firstly, if there was any systematic error between the values observed and those predicted by the models and secondly, to compare the similarity of the predictions performed by the different instruments. For this comparison the Deming approach was used as reported earlier by Linnet [29]. The test consists on a joint analysis on slope and intercept (SI test) and it measures if the differences between the predictions and the observed values are only due to random noise. The null hypothesis (Ho) is accepted if the slope is not different from 1 and the intercept is not different from 0 between the two methods at a 95% confidence level. The regression test is a non-parametric approach suitable for measurements affected by non-negligible experimental error. The test does not impose that one of the methods (reference or infrared) is chosen as reference. The test was conducted for the test (validation) set levels of the phenolic compounds investigated in the study. To further investigate how reliable the predictions were, inter-class correlation coefficients (ICC) with confidence intervals at 95% and the standard "typical" error of measurements (SEM) were investigated. ICC provides information on the consistency of the predicted values as it is a relative measurements of reliability [30]. The ICC takes values from 0 (no reliability) to 1 (perfect reliability). The closer the ICC is to 1 the higher the reliability and the lower the error. The ICC confidence intervals were also reported in order to assess whether there were significant differences between the predictions obtained for the same samples that were analysed with the different instruments. On the other hand, the SEM provides an absolute measurement of reliability as it quantifies the precision of individual measurements. The SEM has the same units as the reported values for the reference methods and it indicates the amount of error that can be considered as measurement error. The calibration results were reported following the indications provided by Olivieri [31] and William et al. [32].

2.7. Reference values

The number of samples, average, standard deviation, minimum, maximum and coefficient of variation (CV) of the HPLC quantified phenolic compounds as well as the four spectrophotometric parameters

are included in Supplementary information 1 (S1). The variation in the concentration ranges seems to be large enough to ensure sufficient variability in the data set. On the other hand, the standard error of the reference methods was also reported [32]. For that, the same sample was analysed, the same day, eight times by the same operator. A CV of 6.3% was observed for the MCP tannin levels. Moreover, the total anthocyanins and total tannins determination showed a CV < 2%. The determination of the colour density presented CV of 5.2%. Moreover, CVs < 2% were observed for gallic acid, catechin, caftaric acid and malvidine-3-glucoside while quercetin-3-glucoside showed a CV 2.6%. Chromatographic techniques contain some limitations that might compromise the accuracy of the predictions. Compounds are sometimes quantified based on the most available compounds (e.g. malvidine-3-glucoside is used to quantify the individual anthocyanins) which may have an influence on the response factor. Moreover, co-elution, shifting or baseline drift are also sometimes observed. Finally, there is always a chance to quantify a different compound with similar spectral features.

2.8. Transmission Fourier transform near-infrared spectroscopy (FT- NIR) prediction models

The NIR region provides vibrational information in the form of overtones and combination bands. Despite the low intensity, the band shape is often characteristic of single compounds or group of compounds, fact that makes this technique suitable for quantification purposes [10]. Spectral properties in the near infrared region (12,500–4000 cm^{-1}) were collected in transmission mode and are reported in Supplementary information S2(a). The main dominant features observed in the raw spectra are absorption bands that correspond to the major components of grape and also to the ethanol generated during the fermentation process [21]. The main dominant features occur at 6900 cm^{-1} and 5100 cm^{-1}, which are related to water and ethanol. Changes during fermentation were observed in the 4500–4300 cm^{-1} region and around 6900 cm^{-1} which indicates the production of ethanol. Moreover, changes were also noticed in the region around 5600 cm^{-1}, which have been associated to the changes in glucose, fructose, phenolics and tannins that occur during the fermentation [20,33].

Prediction models were investigated for the quantified HPLC phenolic compounds as well as for the four spectrophotometric phenolic measurements. The summary statistics for the NIR calibrated and validated models are presented in Table 1. The number of samples included, the range of concentrations in the validation set, the spectral pre-processing method, the calibration/validation ratio and the statistical parameters that define the accuracy of the validated prediction models are shown. As can be observed in Table 1, only the flavan-3-ol dimer B1 and the grape reaction product (GRP) models sowed RPD_{val} values lower than 2.5. Moreover, 5 phenolic compounds, the MCP tannins and the TPI determination showed $RPD_{val} > 3$ which ensures increased model's robustness. In previous results reported by Cozzolino [20], RPD > 2.5 for the prediction of the HPLC polymeric phenols peak (reported as tannins by Cozzolino [20]), was observed for calibrated models (no validation data was reported) containing only Cabernet Sauvignon samples. Lower accuracy was obtained when the samples set, composed by two cultivars, two vintages and two different yeasts, were included in the calibration models (RPD < 2.5). The prediction models for polymeric phenols in our study showed RPD values of 2.89 in validation. Robust prediction models were also observed, and reported here for the first time, for the spectrophotometric determination of MCP tannins, anthocyanins, total phenolics and colour density. The calibration/validation ratio was almost always (except two models) set at 50/50, a fact that also adds robustness to the prediction models.

The results from the SI test showed differences at 95% confidence interval for slope and intercept (Ho rejected) between the reference and predicted values for most of the NIR calibrations. Interestingly, the spectrophotometric methods for phenolic analysis seem to provide increased accurate models as the null hypothesis was confirmed (expect for the anthocyanins). However, the results observed indicate a systematic error between the combination NIR instrument/PLS models, but does not provide information about the magnitude of the error, which means that there can be large differences between the reference and the predicted values despite the two criteria still being met. To investigate the reliability of the predictions ICC and SEM were evaluated. Higher ICC values (> 0.9) were observed almost throughout all the calibrations which indicated a high proportion (> 93% on average) of the differences attributable to the variance of the true vs. predicted values. SEM values were always lower and comparable in magnitude to the real error of validation given by the RMSEP which also indicate increased accuracy.

A large number of spectral pre-processing methods were selected during the model optimization step. Multiplicative scattering correction (MSC), named as one of the most common normalization technique for NIR spectroscopy [34] and straight line subtraction were preferred pre-processing methods. Min-max normalization and the absence of spectral pre-processing were also frequently used. On the other hand, the most commonly used spectral regions corresponded to the 6500–5700 cm^{-1} region, attributed to phenolic compounds and tannins as well as to the 8800–7250 cm^{-1} and 10,500–8800 cm^{-1} regions that are also selected in an important number of prediction models (Supporting information S3(a)).

2.9. Reflexion attenuated total reflectance mid-infrared (ATR-MIR) prediction models

The MIR range is also sensitive to some groups not detectable in the NIR range and is thus expected that it might offer increase sensitivity for the prediction of some wine compounds and properties [5]. MIR is also the region in the electromagnetic spectrum that provides fundamental vibrational frequencies for functional groups in molecules [10]. Table 2 shows calibration and validation statistics obtained for ATR-MIR prediction models. The spectral features obtained from the reflected spectra are dominated by absorption bands that belong mainly to organic acids, carbohydrates and ethanol produced during fermentation. The bands provided by these major components masks the characteristic IR vibrations of phenolic compounds [35]. The main absorption bands that correspond to the infrared absorption of water and ethanol can be observed at 3305 and 1640 cm^{-1} and 2985 and 1050 cm^{-1}, respectively (Supporting information S2(b)). The later absorption bands have also been associated to organic compounds such as fructose and glucose. Moreover, the CO_2 release by yeast during fermentation can be observed at 2341 cm^{-1}. The spectral regions associated with the absorptions of phenolic compounds have been identified in the 1500–1100 cm^{-1} region [36,37] also known as the fingerprint region (1500–900 cm^{-1}).

MIR prediction models showed $RPD_{val} > 2.5$ in 75% of the compounds or parameters under study. It thus seemed that the ability of MIR to predict phenolic compounds is slightly lower that what was observed by NIR spectroscopy. ATR-MIR seems to have difficulties to accurately predict GRP and quercetin, compounds that showed RPD_{val} values < 2. Moreover, the models built to predict the levels of six phenolic compounds including gallic acid, catechin, B1, *p*-coumaric acid, kaempherol and cyaniding-3-glucoside showed $RPD_{val} > 2$ but lower than 2.5. On the other hand, this technique seems to predict the anthocyanin composition with $RPD_{val} > 2.5$ for all the anthocyanin compounds (except for cyaniding-3-glucoside) as well as the levels of polymeric pigments. Contrarily to what was reported by other authors [21] $RPD_{val} > 2.5$ was observed for the MCP tannin assay prediction model. Moreover, a model with $RPD_{val} = 2.59$ is reported here for the first time for the colour density parameter. In this case 7 out of 31 models were built with a calibration validation ratio of 68/32.

As observed for NIR spectroscopy, the combined slope and inter-

Table 1
Summary statistics for the NIR calibration and validation models built for the studied phenolic measurements.

	N	Range (mg/L)	Pre-processing	Rank	R2cal	RMSECV	RPDcal	R2val	RMSEP	RPDval	Cal/Val	SI test	ICC	SEM
Gallic acid	541	0.31–71.4	No pre-processing	19	0.95	3.01	4.6	0.86	4.79	**2.67**	50/50	Ho rejected	0.93	3.39
Catechin	526	5.15–81.4	MSC	12	0.88	5.85	2.89	0.83	6.39	**2.5**	50/50	Ho rejected	0.91	4.42
B1	530	2.74–53.04	MSC	10	0.76	4.94	2.03	0.76	5.68	**2.09**	68/32	Ho rejected	0.87	3.97
Polymeric phenols	568	67.89–2005	Straight line subtraction	12	0.9	135	3.23	0.88	147	**2.89**	50/50	Ho accepted	0.94	104.57
GRP	512	0.36–17.39	No pre-processing	16	0.92	0.76	3.55	0.84	0.93	**2.48**	50/50	Ho rejected	0.91	0.66
Caftaric acid	554	0.31–139.5	Straight line subtraction	17	0.92	8.8	3.63	0.86	11.2	**2.72**	50/50	Ho rejected	0.93	7.9
Caffeic acid	522	0.36–10.36	MSC	13	0.92	0.82	3.59	0.87	1.05	**2.76**	50/50	Ho rejected	0.93	0.74
Coutaric acid	568	0.45–38.91	No pre-processing	19	0.92	2.63	3.59	0.84	3.6	**2.54**	50/50	Ho rejected	0.92	2.53
***p*-coumaric acid**	517	0.26–9.476	SNV	14	0.88	0.61	2.9	0.87	0.63	**2.78**	50/50	Ho rejected	0.93	0.44
Quercetin-3-glucoside	556	0.27–154.3	Straight line subtraction	11	0.91	10.3	3.3	0.88	12	**2.86**	50/50	Ho accepted	0.94	8.47
Quercetin	510	0.46–20.75	Straight line subtraction	13	0.88	1.65	2.84	0.84	1.71	**2.53**	50/50	Ho accepted	0.92	1.21
Kaempherol	504	0.27–5.022	SNV	19	0.98	0.15	6.53	0.85	0.32	**2.59**	50/50	Ho accepted	0.92	0.23
Delphinidin-3-glucoside	541	0.36–50.44	MSC	13	0.95	2.32	4.62	0.92	3.11	**3.46**	50/50	Ho rejected	0.96	2.2
Cyanidin-3-glucoside	242	0.34–1.126	MSC	14	0.87	0.05	2.74	0.86	0.06	**2.68**	68/32	Ho rejected	0.92	0.04
Petunidin-3-glucoside	546	0.47–45.91	Straight line subtraction	16	0.97	2.16	5.35	0.9	3.52	**3.2**	50/50	Ho rejected	0.95	2.48
Peonidin-3-glucoside	546	0.37–22.23	No pre-processing	13	0.88	1.73	2.93	0.85	1.74	**2.59**	50/50	Ho rejected	0.92	1.24
Malvidin-3-glucoside	563	0.94–287	Straight line subtraction	16	0.96	16.5	4.79	0.87	26.8	**2.8**	50/50	Ho rejected	0.93	18.98
Delphinidin-3-acetylglucoside	492	0.37–16.82	No pre-processing	20	0.98	0.65	6.43	0.88	1.38	**2.88**	50/50	Ho accepted	0.94	0.98
Cyanidin-3-acetylglucoside	324	0.50–7.308	No pre-processing	14	0.94	0.34	4.01	0.91	0.39	**3.47**	50/50	Ho rejected	0.95	0.27
Petunidin-3-acetylglucoside	521	0.42–15.17	No pre-processing	16	0.95	0.89	4.42	0.92	1.04	**3.62**	50/50	Ho accepted	0.96	0.74
Peonidin-3-acetylglucoside	544	0.59–12.55	MSC	18	0.97	0.65	5.89	0.91	1.08	**3.39**	50/50	Ho rejected	0.95	0.76
Malvidin-3-acetylglucoside	560	0.63–108.3	No pre-processing	17	0.94	7.15	4.04	0.85	10.5	**2.6**	50/50	Ho rejected	0.92	7.4
Delphinidin-3-cumarylglucoside	427	0.34–5.887	No pre-processing	21	0.98	0.19	7.61	0.86	0.50	**2.73**	50/50	Ho rejected	0.93	0.35
Petunidin-3-cumarylglucoside	494	0.45–9.29	MSC	13	0.95	0.57	4.38	0.85	0.90	**2.56**	50/50	Ho rejected	0.92	0.63
Peonidin-3-cumarylglucoside	544	0.31–15.26	MSC	15	0.95	0.84	4.58	0.86	1.38	**2.73**	50/50	Ho rejected	0.93	0.95
Malvidin-3-cumarylglucoside	562	0.55–50.33	Constant offset elimination	15	0.88	4.27	2.88	0.84	4.87	**2.5**	50/50	Ho rejected	0.91	3.39
Polymeric pigments	551	1.906–85.3	Min-max normalization	16	0.91	5.71	3.36	0.86	6.63	**2.62**	50/50	Ho accepted	0.93	4.73
MCP tannins	569	75.22–3228	Min-max normalization	15	0.94	204	4.09	0.92	231	**3.49**	50/50	Ho accepted	0.96	163.41
Anthocyanins	569	28.3–883	MSC	12	0.93	53.1	3.71	0.87	71.1	**2.75**	50/50	Ho rejected	0.93	49.85
TPI	569	5.748–89.35	Min-max normalization	17	0.97	2.88	5.86	0.95	3.48	**4.49**	50/50	Ho accepted	0.98	2.47
CD	560	2.733–39.55	Min-max normalization	17	0.93	1.91	3.82	0.85	2.7	**2.62**	50/50	Ho accepted	0.92	1.91

Table 2
Summary statistics for the MIR calibration and validation models built for the studied phenolic measurements.

	N	Range (mg/L)	Pre-processing	Rank	R2cal	RMSECV	RPDcal	R2val	RMSEP	RPDval	Cal/Val	SI test	ICC	SEM
Gallic acid	511	0.31–48.55	SNV	10	0.90	3.42	3.16	0.83	4.93	**2.40**	68/32	Ho rejected	0.9	3.49
Catechin	511	5.92–81.4	No pre-processing	8	0.75	7.26	2.32	0.78	8.36	**2.23**	68/32	Ho rejected	0.88	5.77
B1	530	0.31–50.22	SNV	7	0.83	4.99	2.38	0.80	5.12	**2.25**	50/50	Ho rejected	0.86	3.82
Polymeric phenols	547	67.89–2005	SNV	11	0.91	128.0	3.33	0.85	163.0	**2.55**	50/50	Ho accepted	0.92	114.61
GRP	519	0.36–15.22	MSC	11	0.88	0.86	2.93	0.70	1.18	**1.82**	50/50	Ho rejected	0.82	0.839
Caftaric acid	552	0.31–139.5	Min max normalization	11	0.90	9.76	3.23	0.85	11.80	**2.61**	50/50	Ho rejected	0.92	8.314
Caffeic acid	534	0.36–10.36	MSC	8	0.87	1.07	2.77	0.86	1.09	**2.67**	50/50	Ho rejected	0.92	0.769
Coutaric acid	547	0.45–38.7	First derivative	11	0.86	3.14	2.70	0.85	3.95	**2.59**	68/32	Ho rejected	0.92	2.8
p-coumaric acid	515	0.26–8.04	Straight line subtraction	9	0.86	0.63	2.64	0.81	0.70	**2.29**	50/50	Ho rejected	0.89	0.5
Quercetin-3-glucoside	553	0.27–154.3	No pre-processing	10	0.86	13.00	2.67	0.85	13.10	**2.61**	50/50	Ho rejected	0.92	9.3
Quercetin	497	0,45–20,93	SNV	8	0.69	2.56	1.78	0.69	2.55	**1.79**	68/32	Ho rejected	0.82	1.8
Kaempherol	519	0.27–4.978	No pre-processing	11	0.81	0.34	2.32	0.82	0.44	**2.44**	68/32	Ho rejected	0.89	0.
Delphinidin-3-glucoside	541	0.36–50.44	SNV	11	0.92	2.96	3.61	0.88	3.77	**2.85**	50/50	Ho accepted	0.94	2.67
Cyanidin-3-glucoside	246	0.34–1.008	No pre-processing	7	0.77	0.06	2.06	0.76	0.08	**2.06**	68/32	Ho rejected	0.86	0.05
Petunidin-3-glucoside	546	0.47–45.91	Straight line subtraction	12	0.95	2.57	4.45	0.86	4.24	**2.65**	50/50	Ho rejected	0.92	3
Peonidin-3-glucoside	538	0.37–22.54	No pre-processing	11	0.92	1.47	3.47	0.84	1.82	**2.54**	50/50	Ho rejected	0.92	1.29
Malvidin-3-glucoside	563	0.94–287	Straight line subtraction	11	0.93	20.70	3.78	0.85	29.10	**2.58**	50/50	Ho rejected	0.92	20.57
Delphinidin-3-acetylglucoside	486	0.37–16.82	No pre-processing	9	0.88	1.44	2.87	0.86	1.48	**2.71**	50/50	Ho rejected	0.93	1.04
Cyanidin-3-acetylglucoside	337	0.50–7.308	SNV	9	0.88	0.48	2.82	0.85	0.51	**2.57**	50/50	Ho accepted	0.92	0.36
Petunidin-3-acetylglucoside	521	0.42–15.17	SNV	10	0.91	1.19	3.27	0.88	1.32	**2.86**	50/50	Ho rejected	0.94	0.93
Peonidin-3-acetylglucoside	544	0.59–12.55	SNV	9	0.92	1.03	3.62	0.89	1.24	**2.97**	50/50	Ho rejected	0.94	0.87
Malvidin-3-acetylglucoside	545	0.63–108.3	SNV	11	0.95	6.49	4.30	0.89	9.18	**2.96**	50/50	Ho rejected	0.94	6.47
Delphinidin-3-cumarylglucoside	406	0,30–5,28	No pre-processing	10	0.90	0.41	3.12	0.85	0.49	**2.57**	50/50	Ho rejected	0.92	0.35
Petunidin-3-cumarylglucoside	478	0,47–7881	No pre-processing	9	0.87	0.79	2.96	0.85	0.86	**2.67**	50/50	Ho rejected	0.92	0.59
Peonidin-3-cumarylglucoside	531	0.31–13.12	Straight line subtraction	11	0.94	0.91	3.99	0.85	4.20	**2.62**	50/50	Ho rejected	0.92	0.94
Malvidin-3-cumarylglucoside	542	0.55–39.32	No pre-processing	11	0.88	3.98	2.90	0.85	4.20	**2.62**	50/50	Ho rejected	0.92	2.62
Polymeric pigments	539	1.91–76.74	No pre-processing	12	0.91	5.50	3.25	0.85	6.56	**2.61**	50/50	Ho accepted	0.92	4.64
MCP tannins	569	75.22–3228	SNV	11	0.9	261	3.17	0.89	322	**2.51**	50/50	Ho accepted	0.92	227.61
Anthocyanins	569	28.3–883	Constant set elimination	14	0.94	47.20	4.20	0.86	71.70	**2.71**	50/50	Ho rejected	0.93	50.63
TPI	569	5.75–89.35	SNV	10	0.93	4.34	3.84	0.91	4.82	**3.25**	50/50	Ho accepted	0.95	3.41
CD	550	2.73–39.55	SNV	11	0.88	2.24	2.93	0.85	3.07	**2.59**	68/32	Ho accepted	0.92	2.16

cept test was rejected in most of the cases. Again the spectrophotometric methods (excluding anthocyanins) were shown to provide accurate estimations (Ho accepted) indicating the suitability of infrared techniques to predict these parameters. With regards to ICC slightly lower average values were observed (0.91), but again always higher than 0.9 for the majority of the models. Finally, SEM values lower and thus comparable to the prediction errors were always reported.

A large number of different pre-processing methods were used to optimize the prediction models, including standard normal variate (SNV), commonly used in IR [8] and no spectral pre-processing. Despite the importance of pre-processing methods to improve spectroscopy calibrations [38] the absence of spectral pre-processing was found as one of the preferred options. Straight line subtraction and MSC were also used to a lesser extent. With regards to the regions selected by the software when building the prediction models, the band corresponding to the fingerprint regions (900–1500 cm^{-1}) which also contains information regarding phenolic compounds was always included as can be observed in Supporting information S3(b). Two other regions (2250–3250 and 3500–3750 cm^{-1}) were also often included, which corresponds to the spectral regions around the main absorption bands (that account to water).

2.10. Transmission Fourier transform infrared (FT-IR) prediction models

IR spectroscopy shows a strong absorption of water and other solvents such as ethanol compromising it applicability to hydro alcoholic solutions. Overlap between these major components and the analyte of interest limits its applicability for quantification purposes [10]. Despite the possible loss of information, spectral pre-processing and wavenumber selection techniques might be in this case required [10]. Spectral properties contained in the infrared region are presented in Supporting information S2(c). The presence of absorption bands corresponding to water (1717–1543 cm^{-1} and 3624–2971 cm^{-1}) dominate the spectra and mask the characteristic IR vibrations of the phenolic compounds. The relevant information on phenolic compounds is expected to be found in the 1600–1000 and 3000–1700 cm^{-1} as the region between 5000 and 3500 cm^{-1} is thought to contain very little information [39]. Table 3 shows summary statistics as well as range of predicted levels, pre-processing methods in the prediction models for the phenolic compounds and measurements investigated in the study. As can be observed only five HPLC quantified compounds showed $RPD_{val} < 2.5$ and within these, p-coumaric acid, quercetin and delphinidine-3-glucoside have RPD values of 2.45, 2.48 and 2.49, respectively. On the other hand, the polymeric phenols and peonidin-3-acetyl glucoside presented $RPD_{val} > 3$. The prediction of the anthocyanin content in wine samples using FT-IR spectroscopy with PLS regression in young, crianza and reserva wines was reported by Romera-Fernandez [40]. Prediction models for 13 different anthocyanins including the glucosides as well as the acetyl and cumaryl glucosides were reported. The authors found accurate models only for young wines as the prediction errors showed for aged wines (crianza and reserva) were unacceptable. In the current study the inclusion of wine samples from different vintages (2005–2016) indicates that robust models were also obtained for samples that have been aged in barrel or bottle for a certain period of time. Additionally, the four spectrophotometric parameters investigated also showed $RPD_{cal} > 3$ except for the total anthocyanin content and colour density with RPD_{val} = 2.97 and 2.77, respectively. The models were usually predicted with a 50/50 ratio calibration/validation with only five models using cal/val ratio of 68/32. As far as the authors know (only calibration data in finished wine samples has been reported [39]), this is the first validation data available on the use of a FT-IR for the prediction of phenolic compounds in fermenting samples and wines during aging.

IS tests showed the highest number of null hypothesis accepted of the three spectroscopies, but still the majority of the models showed significant differences between the true vs. predicted values. Moreover, the spectrophotometric measurements again confirmed the Ho, but in this case the exception was the MCP tannin model (Ho rejected). ICC (0.93 on average) and SEM values were comparable to those obtained with FT-NIR spectroscopy, which showed slightly higher accuracy than those observed for the ATR-MIR instrument.

With regards to spectral pre-processing, first derivative appears in this case as the most frequently used pre-processing method, followed by SNV, second derivative and the combination of first derivative and MSC. Interestingly, the absence of spectral pre-processing was not a favourite choice in this case, which can be attributed to the high noise caused to the calibrations by some of the interfering compounds [39]. The most frequently used spectral regions corresponded to the 930–1335 cm^{-1} and 1740–2950 cm^{-1}, regions that have been identified as phenolic compounds containing information regions. Interestingly the 3750–5000 cm^{-1} region, with a priori less information on phenolics, has also been included in a number of prediction models (Supporting information S3(c)).

2.11. Comparison between spectroscopies in terms of robustness and accuracy

The suitability of a specific spectroscopy technique to quantify the levels of phenolic compounds during fermentation and aging can be assessed by investigating the information extracted from the process of building the prediction models. The number of samples included in the validated models defines the range of concentrations that will be predicted by the model with a certain RPD accuracy. In the optimization step the outliers were identified using the Mahalanobis distance. From the Mahalanobis distances of each calibration spectrum the threshold of the Mahalanobis distance is calculated which defines the threshold under which the spectra of unknown samples can be reliably analysed (it measures the similarity between the analysed spectra and the calibration spectra). If the models showed robust accuracy the outliers were not excluded with the aim of not excluding important variability contained in the samples set, being thus able to predict the highest possible range of concentrations. The presence of outliers can be due to abnormal spectral properties, errors during the analytical reference methods performance or also due to non-linearity issues at higher levels of phenolic compounds [41].

The average number of outliers was thus investigated and showed ATR-MIR spectroscopy as the technique with the highest number of outliers on average (3.6%). The other two studied spectroscopies (FT-NIR and FT-IR) had a lower number of outliers excluded in the model optimization step (2.6% and 2.7%, respectively), which on a general basis, provided models able to predict a larger range of phenolic levels. As commonly observed the range of concentrations accurately predicted is smaller than those quantified from the HPLC or spectrophotometric analysis. Non-linearity behaviours were sometimes observed in the model optimization step with a reduced number of samples (generally corresponding to wine samples) in the higher range of concentrations, which were often identified as outliers. After the exclusion of outliers, the model performance was increased despite the smaller range of concentrations predicted. Secondly a smaller range of concentrations were predicted by ATR-MIR spectroscopy models for a certain number of compounds (e.g. gallic acid, GRP, p-coumaric acid, malvidin-3-cumaryl glucoside or polymeric pigments). From the range of phenolic levels quantified in the data set, ATR-MIR spectroscopy models were able to predict on average the 78% of the concentration range measured in the fermenting samples and wines included in the study (i.e. if the range contained in the data set for p-coumaric acid corresponds to 0.42–43.42 mg/L and the model was able to predict 0.452–38.7 mg/L, the percentage of the range accurately predicted by the model thus represent 89.36% of the total data set range). The average percentage corresponded to 82% for both FT-NIR and FT-IR spectroscopies, respectively. This indicates a bigger range of levels

Table 3
Summary statistics for the IR (FOSS Winescan) calibration and validation models built for the studied phenolic measurements.

	N	Range (mg/L)	Preprocessing	Rank	R2cal	RMSECV	RPDcal	R2val	RMSEP	RPDval	Cal/Val	SI test	ICC	SEM
Gallic acid	552	0.314–86.96	SNV	22	0.91	4.57	3.34	0.85	5.67	2.56	50/50	Ho accepted	0.93	3.39
Catechin	512	5.926–81.4	Constant offset elimination	23	0.9	5.39	3.1	0.85	6.12	2.55	50/50	Ho rejected	0.92	4.33
B1	504	2.737–53.35	SNV	22	0.86	3.91	2.7	0.84	4.69	2.52	68/32	Ho rejected	0.87	3.97
Polymeric phenols	567	67.89–2005	SNV	18	0.91	132	3.34	0.91	126	3.37	50/50	Ho accepted	0.96	83.54
GRP	537	0.251–17.39	MSC	22	0.74	1.29	1.97	0.78	1.55	2.17	68/32	Ho rejected	0.91	0.66
Caftaric acid	567	0.3123–139.5	MSC	23	0.91	9.87	3.31	0.87	11.2	2.76	50/50	Ho rejected	0.95	7.08
Caffeic acid	535	0.359–11.11	First derivative	18	0.88	1.02	2.92	0.86	1.1	2.64	50/50	Ho rejected	0.93	0.74
Coutaric acid	547	0.452–38.91	First derivative + straight line subtraction	23	0.91	2.8	3.41	0.84	3.63	2.53	50/50	Ho accepted	0.93	2.46
p-coumaric acid	506	0.264–9.476	Constant offset elimination	16	0.85	0.625	2.6	0.83	0.63	2.45	50/50	Ho rejected	0.93	0.44
Quercetin-3-glucoside	551	0.273–153.4	First derivative	24	0.86	13	2.65	0.82	14.5	2.33	50/50	Ho rejected	0.91	10.82
Quercetin	510	0.462–20.75	SNV	24	0.9	1.59	3.12	0.84	1.78	2.48	50/50	Ho accepted	0.92	1.21
Kaempherol	516	0.269–4.892	Constant offset elimination	24	0.83	0.328	2.45	0.87	0.361	2.75	68/32	Ho rejected	0.93	0.26
Delphinidin-3-glucoside	540	0.357–50.44	First derivative	21	0.86	4.15	2.63	0.84	4.32	2.49	50/50	Ho rejected	0.91	3.03
Cyanidin-3-glucoside	253	0.319–1.064	Straight line subtraction	18	0.82	0.0645	2.33	0.82	0.0621	2.37	68/32	Ho rejected	0.9	0.04
Petunidin-3-glucoside	545	0.468–45.91	MSC	22	0.88	4.1	2.85	0.88	3.85	2.91	50/50	Ho rejected	0.94	2.66
Peonidin-3-glucoside	532	0.374–22.36	First derivative + SNV	24	0.79	2.1	2.18	0.85	2.04	2.58	68/32	Ho rejected	0.92	1.44
Malvidin-3-glucoside	562	0.938–287	First derivative + MSC	21	0.91	24.2	3.29	0.86	27.7	2.71	50/50	Ho rejected	0.94	18.68
Delphinidin-3-acetylglucoside	485	0.374–16.82	First derivative+Straight line substraction	23	0.91	1.28	3.26	0.84	1.58	2.53	50/50	Ho rejected	0.94	0.98
Cyanidin-3-acetylglucoside	336	0.499–7.308	First derivative	23	0.87	0.513	2.78	0.86	0.48	2.71	50/50	Ho accepted	0.95	0.27
Petunidin-3-acetylglucoside	520	0.416–15–17	First derivative	24	0.9	1.24	3.21	0.88	1.29	2.92	50/50	Ho accepted	0.94	0.92
Peonidin-3-acetylglucoside	543	0.585–12.55	Min-max normalization	22	0.92	1.12	3.43	0.9	1.17	3.16	50/50	Ho rejected	0.95	0.81
Malvidin-3-acetylglucoside	559	0.625–108.3	First derivative	22	0.91	8.85	3.3	0.84	10.9	2.5	50/50	Ho rejected	0.92	7.59
Delphinidin-3-cumarylglucoside	426	0.339–5.887	First derivative	24	0.9	0.463	3.12	0.88	0.468	2.9	50/50	Ho rejected	0.94	0.33
Petunidin-3-cumarylglucoside	507	0.450–9.29	MSC	23	0.9	0.831	3.08	0.85	0.914	2.61	50/50	Ho rejected	092	0.63
Peonidin-3-cumarylglucoside	536	0.313–14.05	Min-max normalization	24	0.93	1.01	3.76	0.87	1.31	2.77	50/50	Ho rejected	0.93	0.95
Malvidin-3-cumarylglucoside	553	0.554–41.98	First derivative + MSC	14	0.85	4.7	2.56	0.85	4.36	2.58	50/50	Ho rejected	0.91	3.2
Polymeric pigments	562	1.906–85.3	SNV	22	0.97	7.49	2.8	0.82	7.02	2.52	50/50	Ho rejected	0.92	4.44
MCP tannins	568	75.22–3228	Straight line subtraction	23	0.93	224	3.77	0.92	223	3.63	50/50	Ho rejected	0.96	158.82
Anthocyanins	568	28.3–883	No spectral pre-processing	24	0.92	56.5	3.58	0.89	65.3	2.97	50/50	Ho accepted	0.95	44.46
TPI	568	5.748–89.35	First derivative + MSC	24	0.93	4.48	3.81	0.91	4.7	3.33	50/50	Ho accepted	0.96	3.27
CD	557	2.733–39.55	First derivative + MSC	23	0.88	2.58	2.86	0.87	2.54	2.77	50/50	Ho accepted	0.94	1.76

accurately predicted by the models, which is in accordance with the smaller number of outliers excluded.

In addition, the parameters derived from the calibration and validation steps were also investigated and compared. As expected lower calibration than validation errors were observed for the three spectroscopies. Following the cross validation procedure, as the samples are at some point used for the internal validation, and have therefore also been contained in the calibration data set, an increased model performance is expected. On the contrary in the validation step the samples have never been used to calibrate the model, which means that the information of the samples used for the external validation is completely unknown, being the model performance slightly decreased. Average RPD values in calibration of 4.14, 3.12 and 3.02 were observed for FT-NIR, ATR-MIR and FT-IR, respectively. Moreover, the different spectroscopies showed RPD values in validation of 2.86, 2.55 and 2.71 for FT-NIR, ATR-MIR and FT-IR, respectively. The results indicate that the FT-NIR technique is the most suitable for phenolics quantification of the three IR techniques assessed.

One of the first steps when performing PLS modelling concerns to the selection of the optimum number of latent variables or dimensionalities (also known as rank), which indicates the number of PLS components in the multivariate models. This is not a straightforward decision to make and in some cases may lead to overoptimistic results due to over-fitting of models. Over-fitting is the use of prediction models that use more complicated approaches than are necessary [42], which means that the model performance is challenged to a critical extent in the prediction stage. As the quantification of phenolic compounds using IR spectroscopy in this study handled complex data, with a large number of predictors and observations, models were optimized with a higher rank (greater than 15) for some compounds. Interestingly, the lower rank (10 on average) was observed for the phenolic substances quantified with the ATR-MIR instrument which gave on average models with lower prediction accuracy. The average rank for FT-NIR and for FT-IR (WineScan™) was observed at higher ranks of 15 and 22, respectively. It can be stated that a model is overfitted if the predictions are no better than in a simpler model [42] and this could be observed for the ATR-MIR models where higher ranks did not provide more accurate predictions than those observed using less complex models. The experimental design and some of the data treatment implemented during the model optimization steps were considered based on the possibility of minimizing the risk of overfitting (or in other works to ensure model robustness). As wine is a complex matrix with a changing environment, which modifies the relationships between the variables during the fermentation and aging processes, a large number of samples including high variability were used in the study. This decision allowed us to include in the external validation set a number of samples big enough to guarantee that the models have been intensively challenged to give accurate future predictions [43]. Additionally, the variable selection step avoided the inclusion of variables (spectral data points) that contain non or very little prediction information, thus minimizing the inclusion into the models of non-correlated information and noise [43]. The authors trust that by adopting these decisions the models showed in this study are robust enough for the prediction of external future samples.

Additionally, to the regression metrics reported here, the joint test on slope and intercept (SI test) was also investigated using the predicted data obtained from the three infrared instruments. A pairwise comparison between instruments was evaluated in terms of null hypothesis (slope equals to 1 and 0 intercept at 95% confidence interval). ICC and SEM were also investigated to report on the extent of the possible significant differences (Table 4). Interestingly and contrarily to what was observed for true vs predicted values, the null hypothesis is confirmed for the majority of the phenolic calibrations investigated in the study which points out that on a general basis the differences in the predictions provided by the different instrument are due to random noise. The ICC and SEM evaluation presented similar values for the three compared instruments, with ICC > 0.9 and SEM values within the error of prediction reported, indicating that overall there are not significant differences between the values predicted regardless of the instrument and spectroscopy technique used. The test also allowed to evaluate an instrument comparison by using the ICC test confidence intervals at 95%. In this case an overlapping between the CI at 95% means that there are not significant differences between the values predicted and on the other case if the CI do not show overlapping indicate that there are statistical differences between the values predicted with two different calibrations. As can be observed in Supporting information 4 (S4) only the GRP, quercetin, petunidine-3-acetylglucoside, MCP tannins and TPI models showed no overlapping between FT-NIR and ATR-MIR spectroscopies. As observed in Table 2, GRP and quercetin models performance showed lower accuracy while in the other three cases it seems that the highly accurate models reported for FT-NIR (Table 1) account for the differences observed. When comparing FT-NIR with FT-IR again GRP appears as one of the less accurate models in FT-IR while delphinide-3-glucoside and TPI models accuracy for FT-NIR made them being significantly different. Finally, the comparison ATR-MIR/FT-IR the differences observed were due to the poor quercetin model found for ATR-MIR and for the robust polymeric pigments and MCP tannins found for FT-IR.

3. Discussion

The wine industry is currently requiring easy to operate techniques for routine wine analysis. The possibility of on-line process control and monitoring at different stages of the wine production chain becomes a necessity [10]. Spectroscopy provides a suitable solution as it has been defined as a fast, reliable and simple technique that also allows for the quantification of several parameters simultaneously. The continuous improvements in software and hardware together with increased instrument availability places IR spectroscopy as a suitable option for industrial applications. The ability of infrared spectroscopy to provide phenolic compositional analysis during fermentation and aging has been proved [44]. Different regions in the electromagnetic spectrum as well as two measuring options, namely transmission and reflection, were tested and compared. This study showed the suitability of these techniques to estimate the levels of 27 individual phenolic compounds as well as four important spectrophotometric phenolic measurements.

The comparison of three different spectroscopies was considered with the aim of investigating different applications. The results showed that the three studied techniques provide an accurate quantification of the phenolic content during the fermentation and aging. The monitoring of the phenolic content during the winemaking provides a valuable application that will assist in the decision making during this process (pump overs frequency, pressing, addition of enological products, blending, microoxygenation…). Moreover, despite that the models were built statically, an in-line and on-line monitoring of the fermentation process would be highly possible with the current available technology (knowledge, software, probes, instruments…).

Several authors have stated that IR spectroscopy will soon become a routine method for the modern grape and wine industry [6,11]. However, a number of limitations are delaying this transition and preventing winemakers to benefit from the identified advantages. First of all, the lack of the understanding of the technology appears as the main issue which is often combined with the lack of dedicated personnel. Moreover, the required knowledge to interpret the values obtained is also in most cases absent. As an example, the total tannin content of a sample can be quantified using different principles [45], which provide results that are not in the same order of magnitude [46]. Moreover, other parameters can be expressed as indexes with no units or as concentration of a reference compound (e.g. total phenolic index). In order to obtain a better understanding of the results obtained, information on the reference analytical method used to develop the

Table 4
Slope and intercept regression test results for the three studies spectroscopy techniques.

	FT-NIR/ATR-MIR	ICC	SEM	FT-NIR/FT-IR	ICC	SEM	ATR-MIR/FT-IR	ICC	SEM
Gallic acid	Ho rejected	0.88	3.91	Ho rejected	0.9	4	Ho rejected	0.89	3.92
Catechin	Ho rejected	0.92	4.65	Ho rejected	0.93	3.81	Ho accepted	0.91	4.63
B1	Ho rejected	0.93	2.71	Ho accepted	0.95	2.6	Ho accepted	0.92	2.99
Polymeric phenols	Ho accepted	0.93	108.3	Ho accepted	0.96	83.95	Ho accepted	0.92	114.6
GRP	Ho rejected	0.86	0.71	Ho rejected	0.87	0.95	Ho rejected	0.77	1.04
Caftaric acid	Ho accepted	0.93	7.6	Ho accepted	0.9	9.16	Ho accepted	0.92	8.46
Caffeic acid	Ho rejected	0.94	0.65	Ho accepted	0.94	0.65	Ho accepted	0.95	0.59
Coutaric acid	Ho rejected	0.94	2.56	Ho rejected	0.92	2.63	Ho rejected	0.92	3.04
p-coumaric acid	Ho accepted	0.91	0.42	Ho accepted	0.89	0.46	Ho accepted	0.9	0.44
Quercetin-3-glucoside	Ho accepted	0.93	8.71	Ho rejected	0.9	10.48	Ho rejected	0.89	10.63
Quercetin	Ho accepted	0.91	1.36	Ho accepted	0.93	1.15	Ho rejected	0.91	1.32
Kaempherol	Ho rejected	0.85	0.33	Ho accepted	0.93	0.23	Ho accepted	0.92	0.31
Delphinidin-3-glucoside	Ho accepted	0.95	2.37	Ho rejected	0.93	2.54	Ho rejected	0.91	3.03
Cyanidin-3-glucoside	Ho accepted	0.9	0.04	Ho accepted	0.86	0.05	Ho accepted	0.91	0.04
Petunidin-3-glucoside	Ho accepted	0.94	2.61	Ho accepted	0.93	2.92	Ho accepted	0.91	3.19
Peonidin-3-glucoside	Ho rejected	0.95	0.98	Ho accepted	0.93	1.49	Ho accepted	0.92	1.62
Malvidin-3-glucoside	Ho accepted	0.95	16.11	Ho accepted	0.94	17.9	Ho accepted	0.93	19.1
Delphinidin-3-acetylglucoside	Ho accepted	0.94	0.9	Ho accepted	0.92	1.13	Ho accepted	0.91	1.16
Cyanidin-3-acetylglucoside	Ho accepted	0.95	0.29	Ho accepted	0.95	0.29	Ho rejected	0.95	0.29
Petunidin-3-acetylglucoside	Ho accepted	0.95	0.8	Ho accepted	0.93	0.98	Ho accepted	0.94	0.91
Peonidin-3-acetylglucoside	Ho rejected	0.95	0.74	Ho rejected	0.95	0.78	Ho rejected	0.94	0.84
Malvidin-3-acetylglucoside	Ho accepted	0.95	5.73	Ho accepted	0.92	7.16	Ho accepted	0.93	6.89
Delphinidin-3-cumarylglucoside	Ho accepted	0.89	0.38	Ho accepted	0.94	0.31	Ho accepted	0.92	0.33
Petunidin-3-cumarylglucoside	Ho rejected	0.9	0.65	Ho accepted	0.94	0.56	Ho accepted	0.89	0.71
Peonidin-3-cumarylglucoside	Ho rejected	0.92	0.95	Ho accepted	0.93	0.91	Ho accepted	0.94	0.82
Malvidin-3-cumarylglucoside	Ho accepted	0.93	2.89	Ho rejected	0.91	3.19	Ho accepted	0.94	2.56
Polymeric pigments	Ho rejected	0.93	4.47	Ho rejected	0.93	4.32	Ho accepted	0.93	4.33
MCP tannins	Ho accepted	0.94	192.59	Ho accepted	0.97	134.6	Ho accepted	0.93	207.07
Anthocyanins	Ho accepted	0.94	45.43	Ho accepted	0.93	49.74	Ho accepted	0.91	55.71
TPI	Ho accepted	0.95	3.51	Ho accepted	0.96	2.95	Ho accepted	0.95	3.96
CD	Ho accepted	0.94	1.82	Ho accepted	0.95	1.63	Ho accepted	0.93	2.02

calibration models would thus need to be always provided. Additionally, is very often found that the model given by the supplier is not able to accurately predict the samples of a specific cellar or location. This can be due to different reasons including, the models not covering the range of concentrations present in the cellar (models built for red wines used to predict the levels of phenolics in rose wines) or models not containing the chemical variability present in the cellar samples (for instance models built to predict Pinot Noir phenolic content without including Pinot Noir samples in the calibration step). The implementation of the technology should be accompanied by a model optimization step where samples from the winery or analytical laboratory will be included in both the calibration and validations steps. In house model calibration using the same analytical and operational conditions would thus be the ideal scenario.

Finally, overoptimistic results are very often reported in the literature and the prediction ability of the models needs thus to be considered with caution. This includes reporting only calibration data (with no independent validation), the inclusion of a limited number of samples in the calibration models or the use of experimental wine-making conditions. In order to ensure robustness and accuracy enough sample variability needs to be included in the model calibrations. Moreover, the number of external samples used to validate the model should be kept as high as possible (50/50 ratio calibration/validation is desirable). Besides, fermenting samples were obtained from the three most representative South African cultivars (Cabernet Sauvignon, Shiraz and Pinotage) and also from a fermenting tank containing Grenache grapes. The inclusion of Grenache and Pinotage samples was done with the objective of challenging the model with the inclusion of samples containing different phenolic profiles (compared against dominant varieties). As observed in the results section the models were able to handle the additional variability. The use of models which include only chemical information from a single cultivar should be avoided if the aim is to predict the phenolic content of any other different cultivars. Moreover, the inclusion of different vintages (to cover varying seasonal effects) and the definition of experimental parameters at the same level of industrial conditions are also necessary requirements to ensure prediction accuracy. Finally, additional tests have been also evaluated with the aim of extending on the interpretation of the results obtained and also with the aim of comparing the prediction performance of the three instruments. The SI test showed a systematic error between the values predicted by any of the three instruments and those measured through the reference analytical methods. The rejection of the null hypothesis indicates that the predictions might provide larger error towards the extremes (slope different from 1) and/or possible systematic over- or under-estimation (intercept not placed at 0). With the aim of further investigating on the type of error, Bland and Altman plots were evaluated (data not shown). The plots consist on a representation of the difference between observed vs. predicted against the average observed/predicted for each sample. It is thus possible to visualize the structure of the differences between observed vs. predicted values and therefore the error type if any. Generally, the data points were scattered through the plot space, indicating an equal distribution of the error (constant error) within the 95% confidence interval for a specific comparison. However, the values were not statistical significant as was observed with the ICC and SEM values reported. On the other hand, it seems that the three spectroscopy techniques investigated provide predictions statistically comparable, which indicates that the phenolic composition of fermenting samples and wines can be obtained using any of the three investigated instruments. As far as the we know this study represents the first attempt to provide validated prediction models for the phenolic profiling of commercial fermenting and wine samples using three different infrared spectroscopy applications.

Acknowledgments

The authors gratefully acknowledge the Vali+d program of the Conselleria de Educacio Cultura i Esport APOSTD/2014/035

(Generalitat Valenciana), the National Research Foundation (NRF Free-standing postdoctoral fellowship (Grant no. 85242)) and Winetech South Africa WdT 13/01 and WdT 15/02 for financial support during the postdoctoral research collaboration between Stellenbosch University and Polytechnic University of Valencia.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.talanta.2017.08.065.

References

[1] R.R. Cusmano Lucia, Morrison Andrea, Catching- up trajectories in the wine sector, World Dev. 38 (2010) 1588–1602.
[2] S. Ponte, J. Ewert, Which way is "Up" in upgrading? Trajectories of change in the value chain for South African Wine, World Dev. 37 (2009) 1637–1650. http://dx.doi.org/10.1016/j.worlddev.2009.03.008.
[3] D. Cozzolino, M. Parker, R. Dambergs, Chemometrics and visible-Near infrared spectroscopic monitoring of red wine fermentation in a pilot scale, Biotechnology (2006) ⟨http://onlinelibrary.wiley.com/doi/10.1002/bit.21067/full⟩ (accessed 27 January 2017).
[4] N.D. Lourenço, J.A. Lopes, C.F. Almeida, M.C. Sarraguça, H.M. Pinheiro, Bioreactor monitoring with spectroscopy and chemometrics: a review, Anal. Bioanal. Chem. 404 (2012) 1211–1237. http://dx.doi.org/10.1007/s00216-012-6073-9.
[5] M. Gishen, R.G. Dambergs, D. Cozzolino, Grape and wine analysis - enhancing the power of spectroscopy with chemometrics, Aust. J. Grape Wine Res. 11 (2005) 296–305. http://dx.doi.org/10.1111/j.1755-0238.2005.tb00029.x.
[6] R. Dambergs, M. Gishen, D. Cozzolino, A review of the state of the art, limitations, and perspectives of infrared spectroscopy for the analysis of wine grapes, must, and grapevine tissue, Appl. Spectrosc. Rev. 50 (2015) 261–278. http://dx.doi.org/10.1080/05704928.2014.966380.
[7] S. Lohumi, S. Lee, H. Lee, B.-K. Cho, A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration, Trends Food Sci. Technol. 46 (2015) 85–98. http://dx.doi.org/10.1016/j.tifs.2015.08.003.
[8] V. Di Egidio, N. Sinelli, G. Giovanelli, A. Moles, E. Casiraghi, NIR and MIR spectroscopy as rapid methods to monitor red wine fermentation, Eur. Food Res. Technol. 230 (2010) 947–955. http://dx.doi.org/10.1007/s00217-010-1227-5.
[9] D. Cozzolino, W. Cynkar, N. Shah, P. Smith, Quantitative analysis of minerals and electric conductivity of red grape homogenates by near infrared reflectance spectroscopy, Comput. Electron. Agric. 77 (2011) 81–85. http://dx.doi.org/10.1016/j.compag.2011.03.011.
[10] A. Ricci, G. Parpiniello, L. Laghi, M. Lambri, A. Versari, Application of infrared spectroscopy to grape and wine analysis, in: D. Cozzolino (Ed.), Infrared Spectroscopy: Theory, Developments and Applications 2013, Nova Science Publishers, Inc., Hauppauge, NY, 2013, pp. 17–41 (ISBN: 978-1-62948-521-8).
[11] C. Daniel, The role of visible and infrared spectroscopy combined with chemometrics to measure phenolic compounds in grape and wine samples, Molecules 20 (2015) 726–737. http://dx.doi.org/10.3390/molecules20010726.
[12] B.M. Nicolaï, K. Beullens, E. Bobelyn, A. Peirs, W. Saeys, K.I. Theron, J. Lammertyn, Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review, Postharvest Biol. Technol. 46 (2007) 99–118. http://dx.doi.org/10.1016/j.postharvbio.2007.06.024.
[13] W. Ma, A. Guo, Y. Zhang, H. Wang, Y. Liu, H. Li, A review on astringency and bitterness perception of tannins in wine, Trends Food Sci. Technol. 40 (2014) 6–19. http://dx.doi.org/10.1016/j.tifs.2014.08.001.
[14] J.M. McRae, J.A. Kennedy, Wine and grape tannin interactions with salivary proteins and their impact on astringency: a review of current research, Molecules 16 (2011) 2348–2364. http://dx.doi.org/10.3390/molecules16032348.
[15] J.A. Kennedy, J. Ferrier, J.F. Harbertson, C. Peyrot Des Gachons, Analysis of tannins in red wine using multiple methods: correlation with perceived astringency, Am. J. Enol. Vitic. 57 (2006) 481–485.
[16] J. Garrido, F. Borges, Wine and grape polyphenols - a chemical perspective, Food Res. Int. 54 (2013) 1844–1858. http://dx.doi.org/10.1016/j.foodres.2013.08.002.
[17] A. Teixeira, J. Eiras-Dias, S.D. Castellarin, H. Gerós, Berry phenolics of grapevine under challenging environments, Int. J. Mol. Sci. 14 (2013) 18711–18739. http://dx.doi.org/10.3390/ijms140918711.
[18] V. Cheynier, M. Dueñas-Paton, E. Salas, C. Maury, J.M. Souquet, P. Sarni-Manchado, H. Fulcrand, Structure and properties of wine pigments and tannins, Am. J. Enol. Vitic. 57 (2006) 298–305.
[19] H. Fulcrand, M. Dueñas, E. Salas, V. Cheynier, Phenolic reactions during wine-making and aging, Am. J. Enol. Vitic. 57 (2006) 289–297.
[20] D. Cozzolino, M.J. Kwiatkowski, M. Parker, W.U. Cynkar, R.G. Dambergs, M. Gishen, M.J. Herderich, Prediction of phenolic compounds in red wine fermentations by visible and near infrared spectroscopy, Anal. Chim. Acta 513 (2004) 73–80. http://dx.doi.org/10.1016/j.aca.2003.08.066.
[21] S. Fragoso, L. Aceña, J. Guasch, M. Mestres, O. Busto, Quantification of phenolic compounds during red winemaking using FT-MIR spectroscopy and PLS-regression, J. Agric. Food Chem. 59 (2011) 10795–10802. http://dx.doi.org/10.1021/jf201973e.
[22] M.J. Martelo-Vidal, M. Vázquez, Determination of polyphenolic compounds of red wines by UV–VIS–NIR spectroscopy and chemometrics tools, Food Chem. 158 (2014) 28–34. http://dx.doi.org/10.1016/j.foodchem.2014.02.080.
[23] N. Iland, P. Ewart, A. Sitters, J. Markides, A. Bruer, Techniques for Chemical Analysis and Quality Monitoring during Winemaking, Patrick Iland Wine Promotions, Campbelltown, South Australia, 2000.
[24] C.J. Sarneckis, R.G. Dambergs, P. Jones, M. Mercurio, M.J. Herderich, P.A. Smith, Quantification of condensed tannins by precipitation with methyl cellulose: development and validation of an optimised tool for grape and wine analysis, Aust. J. Grape Wine Res. 12 (2006) 39–49. http://dx.doi.org/10.1111/j.1755-0238.2006.tb00042.x.
[25] M.D. Mercurio, R.G. Dambergs, M.J. Herderich, P.A. Smith, High throughput analysis of red wine and grape phenolics - adaptation and validation of methyl cellulose precipitable tannin assay and modified somers color assay to a rapid 96 well plate format, J. Agric. Food Chem. 55 (2007) 4651–4657. http://dx.doi.org/10.1021/jf063674n.
[26] Y. Glories, La couleur des vins rouges, 2eme partie, Connaiss, La Vigne Du Vin 18 (1984) 253–271.
[27] Z. Peng, P.G. Iland, A. Oberholster, M.A. Sefton, E.J. Waters, Analysis of pigmented polymers in red wine by reverse phase HPLC, Aust. J. Grape Wine Res. 8 (2002) 70–75. http://dx.doi.org/10.1111/j.1755-0238.2002.tb00213.x.
[28] J. Pink, M. Naczk, D. Pink, Evaluation of the quality of frozen minced red hake: use of Fourier transform near-infrared spectroscopy, J. Agric. Food Chem. 47 (1999) 4280–4284. http://dx.doi.org/10.1021/jf990170z.
[29] K. Linnet, Evaluation of regression procedures for methods comparison studies, Clin. Chem. 39 (1993) 424–432.
[30] J.P. Weir, Quatifying test-retest reliability using the intraclass correlation coefficient and the SEM, J. Strength Cond. Res. 19 (2005) 231–240. http://dx.doi.org/10.1519/15184.1.
[31] A.C. Olivieri, Practical guidelines for reporting results in single- and multi-component analytical calibration: A tutorial, Analytica Chimica Acta 868 (2015) 10–22.
[32] P. Williams, P. Dardenne, P. Flinn, Tutorial: items to be included in a report on a near infrared spectroscopy project, J. Near Infrared Spectrosc 25 (2017) 85–90. http://dx.doi.org/10.1177/0967033517702395.
[33] A.V. Goodchild, F.J. El Haramein, A. Abd El Moneim, H.P.S. Makkar, P.C. Williams, Prediction of phenolics and tannins in forage legumes by near infrared reflectance, J. Near Infrared Spectrosc 6 (1998) 175–181. http://dx.doi.org/10.1255/jnirs.134.
[34] B. Lavine, A user-friendly guide to multivariate calibration and classification, Tomas Naes, Tomas Isakson, Tom Fearn and Tony Davies, NIR Publications, Chichester, 2002, J. Chemom. (2002) http://onlinelibrary.wiley.com/doi/10.1002/cem.815/full.
[35] J.L. Moreira, L. Santos, Spectroscopic interferences in Fourier transform infrared wine analysis, Anal. Chim. Acta 513 (2004) 263–268. http://dx.doi.org/10.1016/j.aca.2003.09.029.
[36] J.S. Jensen, M. Egebo, A.S. Meyer, Identification of spectral regions for the quantification of red wine tannins with Fourier transform mid-infrared spectroscopy, J. Agric. Food Chem. (2008) 3493–3499. http://dx.doi.org/10.1021/jf703573f.
[37] S. Fragoso, L. Aceña, J. Guasch, O. Busto, M. Mestres, Application of FT-MIR spectroscopy for fast control of red grape phenolic ripening, J. Agric. Food Chem. 59 (2011) 2175–2183. http://dx.doi.org/10.1021/jf104039g.
[38] L. Laghi, A. Versari, G.P. Parpinello, D.Y. Nakaji, R.B. Boulton, FTIR spectroscopy and direct orthogonal signal correction preprocessing applied to selected phenolic compounds in red wines, Food Anal. Methods 4 (2011) 619–625. http://dx.doi.org/10.1007/s12161-011-9240-2.
[39] A. Versari, R.B. Boulton, G.P. Parpinello, Effect of spectral pre-processing methods on the evaluation of the color components of red wines using fourier-transform infrared spectrometry, Ital. J. Food Sci. 18 (2006) 423–431.
[40] M. Romera-Fernández, L.A. Berrueta, S. Garmón-Lobato, B. Gallo, F. Vicente, J.M. Moreda, Feasibility study of FT-MIR spectroscopy and PLS-R for the fast determination of anthocyanins in wine, Talanta 88 (2012) 303–310. http://dx.doi.org/10.1016/j.talanta.2011.10.045.
[41] R.G. Dambergs, D. Cozzolino, W.U. Cynkar, L. Janik, M. Gishen, The determination of red grape quality parameters using the LOCAL algorithm, J. Near Infrared Spectrosc. 14 (2006) 71–79. http://dx.doi.org/10.1255/jnirs.593.
[42] D.M. Hawkins, The problem of overfitting, J. Chem. Inf. Model. 44 (2004) 1–12.
[43] W. You, S. May, N. Be, W. You, N. Introduction, R. Models, What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models, Psychosom. Med. 66 (2004) 411–421.
[44] A. Versari, G.P. Parpinello, L. Laghi, Application of infrared spectroscopy for the prediction of color components of red wines, Spectroscopy 27 (2012) 36–47.
[45] D. De Beer, J.F. Harbertson, P.A. Kilmartin, V. Roginsky, T. Barsukova, D.O. Adams, A.L. Waterhouse, Phenolics: a comparison of diverse analytical methods, Am. J. Enol. Vitic. 55 (2004) 389–400. http://dx.doi.org/10.1016/S0308-8146(02)00590-3.
[46] A. Cáceres-Mella, Á. Peña-Neira, J. Narváez-Bastias, C. Jara-Campos, R. López-Solís, J.M. Canals, Comparison of analytical methods for measuring proanthocyanidins in wines and their relationship with perceived astringency, Int. J. Food Sci. Technol. 48 (2013) 2588–2594. http://dx.doi.org/10.1111/ijfs.12253.