



Combining portable NIR spectroscopy and multivariate calibration for the determination of ethanol in fermented alcoholic beverages by a multi-product model

Ana Carolina da Costa Fulgêncio^a, Glaucimar Alex Passos Resende^a,
Marden Claret Fontoura Teixeira^a, Bruno Gonçalves Botelho^a, Marcelo Martins Sena^{a,b,*}

^a Chemistry Department, Universidade Federal de Minas Gerais, Belo Horizonte, MG 31270-901, Brazil

^b Instituto Nacional de Ciência e Tecnologia em Bioanálítica (INCT Bio), Campinas, SP 13083-970, Brazil

ARTICLE INFO

Keywords:

Alcoholic strength
Handheld NIR sensor
Portability
Chemometrics
Miniaturization
Direct analysis

ABSTRACT

In this study, a multivariate calibration multi-product model was built by combining partial least square regression (PLS) and portable near infrared (NIR) spectroscopy for the determination of ethanol content in fermented alcoholic beverages. Reference values were obtained by gas chromatography with flame ionization detection (GC-FID). Aiming at building a robust model, a great variety of beers, ciders, meads, and wines were incorporated into the model. NIR spectra were recorded between 908 and 1676 nm for 153 alcoholic beverage samples, corresponding to a range from 4.3 to 15.3% (v/v) of alcohol content. PLS model provided accurate results with root mean square errors of calibration (RMSEC) and prediction (RMSEP) of 0.8% and 0.9%, respectively. The developed method was validated through the estimate of proper figures of merit, such as linearity, trueness, precision, analytical sensitivity, bias, and residual prediction deviation (RPD). This method was simple, direct, rapid, of low-cost and environmentally friendly, not consuming reagents or solvents nor generating chemical waste. It could be incorporated in analytical platforms for quality inspection, contributing to provide better transparency in the food supply chain.

1. Introduction

The expanding global market of alcoholic beverages has increased the demand for their quality control and inspection. One of the main parameters for this control is the ethanol content, whose concentration characterizes the product and may represent its identity. This analytical demand is linked to the quality assurance in the food supply chain, taxation and declaration of ethanol content on the label [1]. From the forensic point of view, it is also important to verify whether a beverage was adulterated or counterfeit [2]. Some countries classify alcoholic beverages according to their ethanol content. The legal category of a beverage can impact the type of license required to manufacture, distribute, and sell the product; it may also influence on taxation [3,4]. Many countries have high excise duties for fermented alcoholic drinks, and the incorrect declaration of alcoholic strength is one of the fraud categories found in official regulations [5,6]. Fermented drinks usually have up to 16% alcohol by volume, while distilled beverages have much higher contents, between 30 and 60% [7]. Beer and wine are among the

most consumed beverages in the world, being the two most consumed alcoholic drinks in Europe [8]. In the present study, two other less consumed fermented beverages were analyzed, cider and mead, whose primary raw materials are apple juice and honey, respectively.

Ethanol content was one of the earliest parameters for which quantitative analytical methods were developed in the quality control of beverages [9]. The classical method for quantifying ethanol in alcoholic beverages is based on determining the density of the beverage after a distillation step [10]. This is a non-selective, laborious and time-consuming approach. Considering the need for quality control inspection in the food supply chain, several analytical methods have been developed for determining ethanol in different alcoholic beverages, including both fermented and distilled. Articles developing methods to quantify ethanol in more than one type of beverage have been based mainly on chromatographic techniques, such as gas chromatography coupled to mass spectrometry (GC-MS) [9] or flame ionization detection (GC-FID) [11], and high-performance liquid chromatography (HPLC) [12]. Besides being laborious and time-consuming, chromatographic

* Corresponding author.

E-mail address: marcsen@ufmg.br (M.M. Sena).

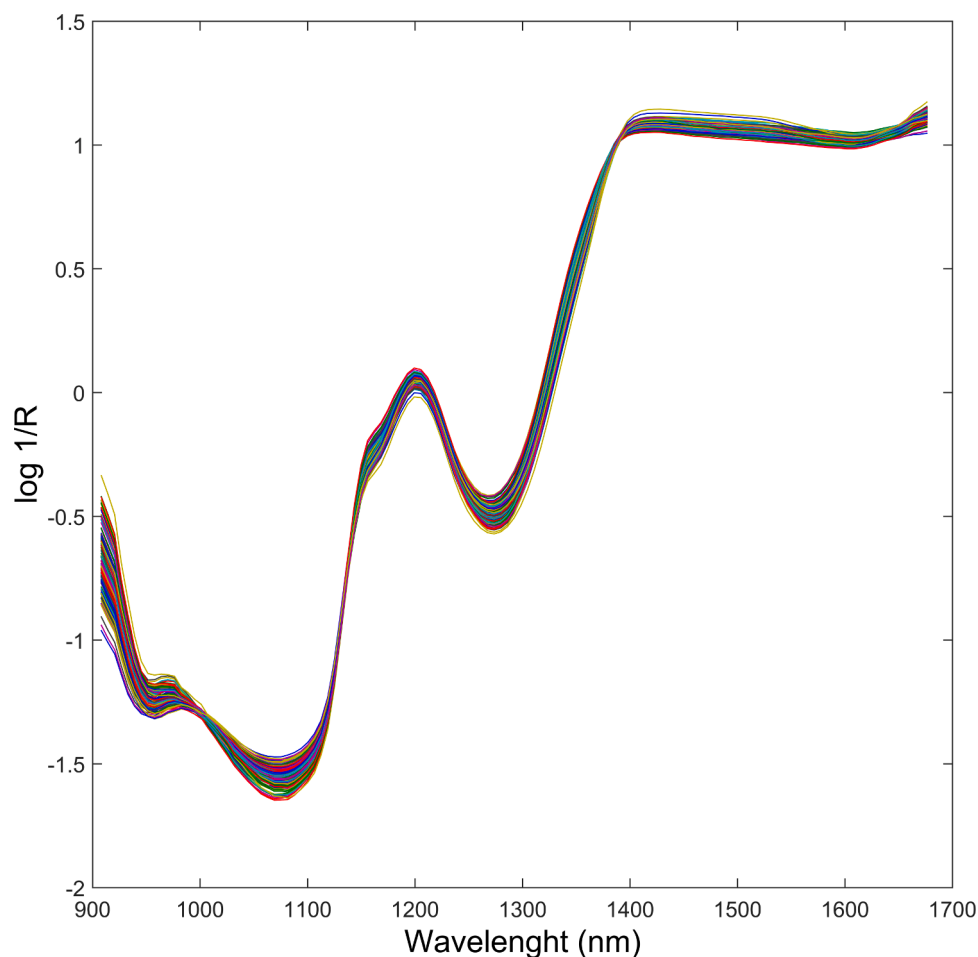


Fig. 1. NIR spectra preprocessed by Savitzky-Golay smoothing and standard normal variate (SNV) of all the 153 analyzed beverages samples.

Table 1

Results for the optimization of the developed PLS model through the detection of outliers.

		Number of calibration samples	Number of validation samples	Number of latent variables (LV)	RMSEC (%)	RMSEP (%)	RMSECV (%)	RPD calibration	RPD validation
Before	Beer	26	13	8	1.4	1.8	1.7	2.1	2.0
	Cider	27	13						
	Mead	24	12						
	Wine	25	13						
After	Beer	25	13	8	0.8	0.9	1.0	3.5	4.1
	Cider	25	11						
	Mead	21	10						
	Wine	21	10						

methods go against the principles of green chemistry, which search to avoid the use of toxic solvents and to minimize the generation of chemical waste. Thus, many alternative methods have been developed, mainly in the last years, seeking to overcome these limitations. These methods have been based on amperometric detection [13], microbial biosensing employing a eukaryote double-mediator system [14], direct analysis in real time mass spectrometry (DART-MS) [2], a colorimetric reaction of ethanol with dichromate in a 3D printing system [15], UV-visible spectra acquired in a flow system with in-line gas-diffusion [1], visible chemical waves registered with a smartphone during a Belousov-Zhabotinsky (BZ) reaction [16], a chemo-chronometric assay utilizing the BZ oscillating reaction [17] and RGB digital images obtained with a desktop scanner [18]. Most of these methods consume reagents [1,15-18], in some cases toxic reagents [15]. While beer and wine are among the most analyzed beverages in all previously mentioned articles, studies involving cider and mead are scarce in the

literature. Only one article has determined ethanol specifically in cider by employing ^1H NMR spectroscopy [19].

In the last years, vibrational spectroscopic techniques have increasingly represented alternatives for developing simpler, direct, faster, and greener analytical methods. Since these techniques provide overlapped analytical signals, their combination with chemometric tools is almost mandatory. Recently, the advantages of these techniques have been improved due to the miniaturization of the spectrophotometers. In fact, portability is an important trend in analytical chemistry in general, not only in spectroscopy, providing advantages such as time saving, low costs of operation, small size and light weight of the equipments, lower energy consumption and environmentally friendly analysis [20-24]. A minimum or no sample pretreatment is required, small sample volumes are necessary, there is no need to consume reagents or solvents, and no chemical waste is generated. There are evident advantages in eliminating the need of sending samples to the laboratory, as it can be moved

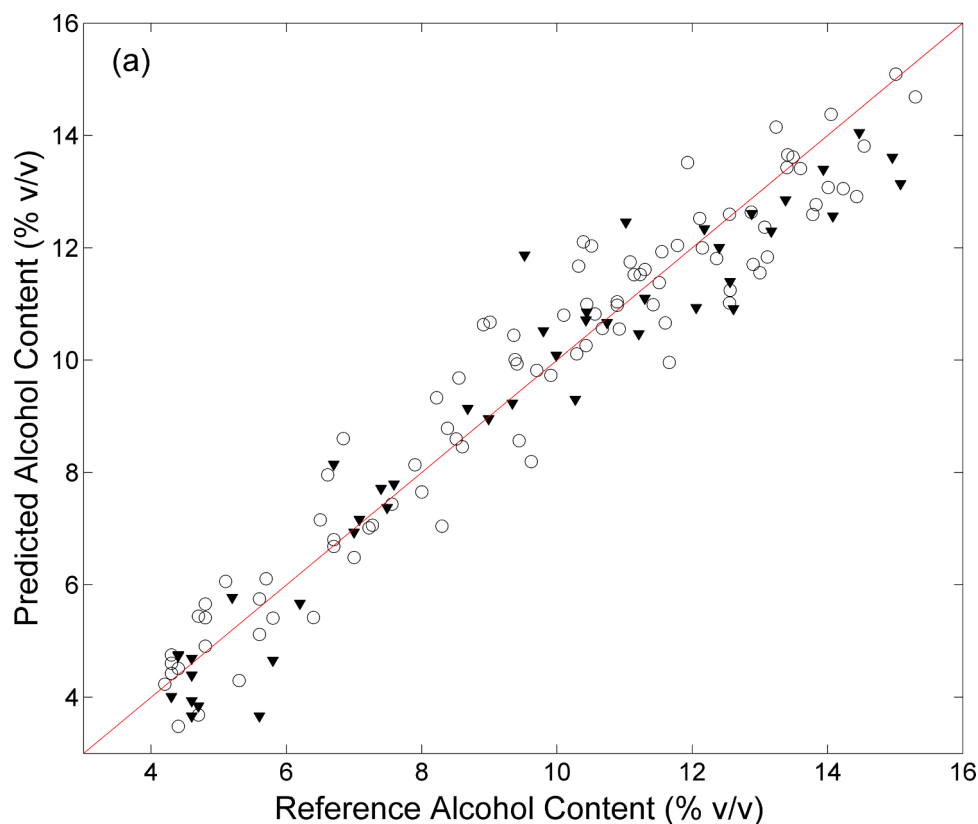


Fig. 2. Plot of reference alcohol content *versus* predicted values for the developed PLS model. (a) Representation as a function of calibration and validation samples. Empty circles indicate calibration samples and full down triangles indicate validation samples. (b) Representation as a function of the type of beverage. gray crosses represent beer samples, green asterisks cider samples, blue squares mead samples, and red down triangles wine samples.

to the field [20]. Particularly, portable devices can provide fast screening of beverage authenticity directly at the point of need [22]. Among vibrational techniques, handheld near infrared (NIR) devices present the advantages of being mostly cheaper, lighter and of smaller dimensions than Raman and mid infrared portable equipments [23,24]. Due to these characteristics, portable NIR spectroscopy is becoming one of the most commercially developed and employed analytical platforms for determining food authenticity, allowing integration into digital traceability systems [23]. In a previous and very recent article, our research group has employed portable NIR spectroscopy and multivariate calibration to determine the alcoholic strength of only one beverage matrix, beer [25]. In the present study, going beyond this previous reference, we developed a multi-product quantitative model by analyzing four different beverage matrices, which implies several advantages discussed in the next paragraph.

Multi-product models were obtained from spectral data for different types of samples or matrices, which are combined into a single multivariate calibration model [26,27]. They are also called global or robust calibration models (though these two terms have also been used for other situations) and can be very convenient and cost-effective. The maintenance of several one-product calibration models is laborious and time-consuming. Going from one-product to multi-product calibration may result in lower accuracy, though limited to an acceptable level, when a large number of samples for each matrix is available. However, when the products are relatively similar and especially if the number of samples of each product is small, making local modeling difficult, multi-product models tend to provide a better performance and lower prediction errors [26]. If the analytical range of the dependent variable is not too large, the traditional linear multivariate calibration model partial least squares (PLS) tends to provide good predictions. The use of portable devices also facilitates the development of this type of model

since the same spectrophotometer can be carried to different analysis sites. Multi-product multivariate calibration models have been developed in combination with NIR data for several types of matrices, such as different foods [26], several kinds of plant feed materials [27], various root and tuber powders [28], biomass from different product origins [29], industrialized fruit nectars and soy juices obtained from different fruits [30], wine grapes (*Vitis vinifera* L.) during on-vine ripening and at harvest from 25 white and red grape varieties [31], and five different vaccine products [32].

The objective of this paper was to develop and validate a multivariate calibration multi-product method based on the combination of portable NIR spectroscopy and PLS for directly determining ethanol content in four distinct fermented alcoholic beverages. Reference values were obtained with a method based on GC-FID. Concerning the robustness of the method, a relative great variety of beers, ciders, meads, and wines were incorporated into the model, corresponding to an analytical range from 4.3 to 15.3% (v/v). The proposed method was also submitted to a multivariate analytical validation according to the Brazilian and international guidelines [33,34].

2. Materials and methods

2.1. Samples

Sample beverages utilized for the construction of the multi-product model were beer (thirty-nine samples), cider (forty samples), mead (thirty-six samples), and wine (thirty-eight samples). A total of 153 samples of alcoholic beverages were purchased in different local marketplaces (100% of both beer and wine samples, 4 samples of cider and 4 samples of mead) or artisanally prepared (the remaining samples of both cider and mead) for this study. Since only a small variety of cider and

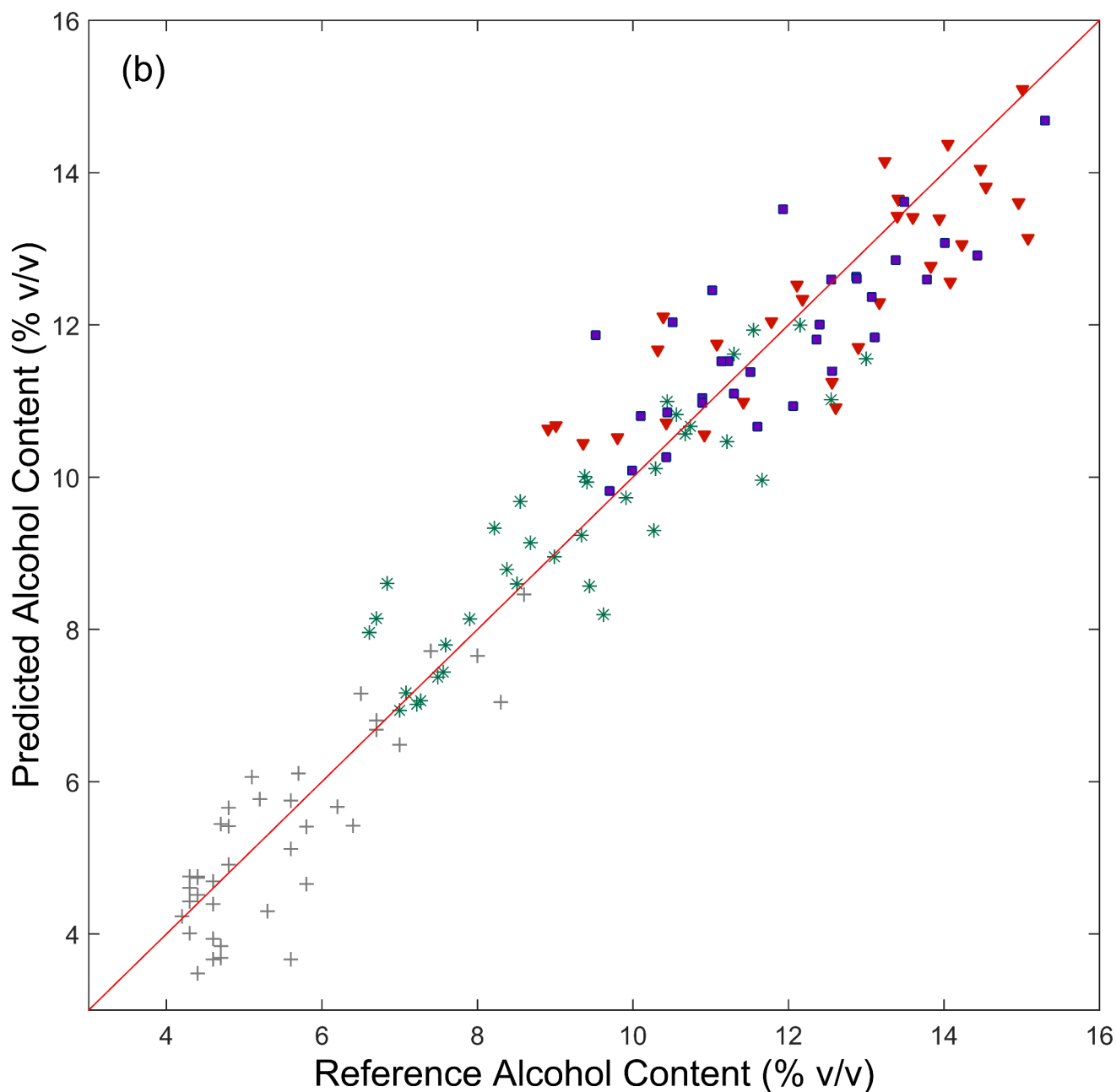


Fig. 2. (continued).

mead brands are available in the local market, about 90% of these beverage samples were artisanally produced in our laboratory with a wide variety of raw materials and alcohol content, aiming at obtaining a robust multi-product model. The whole set of samples was obtained or prepared to provide a wide variance with their ethanol content ranging from 4 to 16% (v/v).

2.2. Instruments and software

NIR spectra were acquired using a portable MicroNIR® 1700 spectrophotometer from Viavi Solutions (Milpitas, CA, USA), with an InGaAs photodiode array detector and linear variable filter (LVF) technology as dispersing element. This is one of the smallest NIR devices, weighing only 64 g, with one of the best performances in the market [23]. Each sample was scanned in the diffuse reflectance mode (and converted to $\log(1/R)$, which is equivalent to absorbance), with twenty scans at the wavelength range of 908–1676 nm, with resolution of 6.25 nm. Data

were processed using MATLAB® software, version 7.13 (MathWorks, Natick, USA), coupled with the PLS_Toolbox, version 6.5 (Eigenvector Technologies, Manson, USA). A Shimadzu GC-17A gas chromatography system (Kyoto, Japan) equipped with a flame ionization detector (FID) and a Poraplot Q column (10 m × 0.32 mm × 10 μm) was used for obtaining reference values.

2.3. Procedure

Approximately 5 mL of each beer and cider sample were degassed using an ultrasonic bath (Equilab ULTRASONIK 28H, Madrid, Spain) for 5 min. Mead and wine samples did not need to be degassed. After degassing, approximately 1.0 mL of each beverage sample was used for the GC-FID reference method and the remaining 4.0 mL for the proposed multi-product NIR method. In the sequence, 3.0 mL of each sample was transferred to a polystyrene Petri dish (1.7 cm radius × 0.9 cm height). The plate was placed above the MicroNIR® spectrophotometer for

Table 2
Other estimated quantitative FOM for the developed PLS model.

Figures of merit	Parameter		Value		
Precision	RSD repeatability (%)	Beer	0.3		
			0.2		
			0.5		
		Cider	1.7		
			0.3		
			0.4		
		Mead	0.4		
			0.1		
			0.2		
		Wine	0.4		
			0.3		
			0.3		
		RSD intermediate precision (%)	Beer	1.6	
				1.2	
				1.8	
				Cider	1.9
					1.1
Mead	1.8				
	0.8				
Wine	0.5				
	1.0				
	1.1				
Linearity	Durbin-Watson test parameter			Slope	1.78
		0.97			
		±0.03			
		0.85			
		±0.22			
Intercept	Correlation coefficient	0.9312			
		0.9			
Inverse of analytical sensitivity (%)			0.9		
Bias (%)			0.3 ± 0.9		
Working Range (%)			4.3–15.3		

spectra recording. During all these measurements, the temperature of the laboratory was set at 25 ± 1 °C. Precautions to maintain the scattering coefficient constant during the experiments were taken, such as the control of the sample volume and the use of a reflective artifact.

2.4. GC-FID analysis (reference method)

A GC-FID method was developed for obtaining reference values for the samples, which were used as dependent variables for the building of the PLS model. For this chromatographic method, 100 µL of each sample were mixed with 900 µL of the internal standard (IS) solution (1-butanol 1.0% v/v) in a 2 mL microcentrifuge tube. The carrier gas was hydrogen with a flow of $2.2 \text{ cm}^3 \text{ min}^{-1}$. The injection port and the detector were both held at 250 °C, and the oven temperature was maintained at 200 °C for 4 min. An aliquot of 0.2 µL of the diluted sample was injected with split ratio of 1:2.

2.5. Multivariate calibration model and analytical validation

A multi-product multivariate calibration model was built using PLS. Samples were split into two-thirds and one-third for the calibration and validation sets, respectively, by applying the Kennard-Stone algorithm [35]. This algorithm was applied separately to the samples of each matrix/product, guaranteeing the representativeness of each type of beverage in the calibration and validation sets. Kennard-Stone algorithm was chosen instead of random sampling because it ensures the selection of representative and homogeneously distributed samples in the whole analytical range of the multivariate space, in a reproducible manner based on a systematic criterion [36].

The performed analytical validation searched to adapt the requirements of the Brazilian and international guidelines [37,38] to the state of the art of estimate FOM in multivariate calibration [33,34,39].

Appropriate figures of merit (FOM) were estimated for the proposed multi-product method in accordance with regulatory requirements, such as trueness, precision, linearity, working range, analytical sensitivity, bias, and residual prediction deviation (RPD). Samples with extreme leverages, large residuals in the X block (spectral data) or large residuals in the y block (studentized residuals of the predicted concentration values) at 95% confidence level were detected as outliers [33,34]. The uncertainty of the predicted values was determined to estimate confidence intervals by calculating the sample-specific standard prediction errors (SPE). SPE were calculated based on the error-in-variables equation (EIV) [40]. This equation considers the uncertainty of the reference method and should only be used in the absence of bias. Thus, the absence of bias in the model should be previously verified by a t-test with the validation samples (the worst case in relation to the calibration samples) at 95% confidence level [39]. In addition, an elliptical joint confidence region (EJCR) was calculated to verify the linearity and the absence of systematic errors in the model [41,42].

3. Results and discussion

Specific guidelines recommend the utilization of a minimum of 6 (nLV+1) and 4(nLV) samples in the calibration and validation sets, respectively, for building infrared multivariate quantitative models, where nLV means the number of latent variables [39]. Taking into account that the number of samples available for building one-product models was too small, it was decided to build only one multi-product multivariate calibration model.

As already mentioned in Section 2.5, the Kennard-Stone algorithm was employed to select samples for calibration and validation datasets, which were composed of 102 and 51 samples, respectively. The proportionality among each type of beverage was maintained in these two datasets. NIR data were preprocessed to improve signal-to-noise ratio, and to avoid drifts caused by multiplicative light scattering, which are typical of diffuse reflectance spectra [43]. Spectra were sequentially preprocessed by Savitzky-Golay smoothing (5 points filter width and second order polynomial fit), standard normal variate scaling (SNV), and mean centering. Fig. 1 shows all preprocessed spectra used in the model development (before mean centering). By observing these spectra, the most prominent signals are the absorption bands centered at about 1200 nm and 1400 nm. These two spectral bands can be assigned to the second overtone of C–H stretching of ethanol, and to the first overtone of O–H stretching vibrations of ethanol and water, respectively [44].

Random subsets (9 splits and 20 iterations) cross-validation was applied to select the best number of latent variables (LV). The LV number presenting the smallest root mean square error of cross-validation (RMSECV) was chosen. The best model was selected with 8 LV after optimization by outlier detection. Outlier detection was performed based on the parameters mentioned in Section 2.5, all at 95% confidence level. This procedure can improve model quality since outliers are defined as observations showing some type of departure from most of the data. The number of outliers must not exceed the limit of 22.2% of the training and test sets for each beverage [38,45]. From the original data, 17 out of 153 samples were removed as outliers (11.1%), ten from the calibration set (9.8%), and seven from the validation set (13.7%). Table 1 specifies how many samples of each beverage were removed from the model. After deleting outliers, the model was reconstructed and accounted for 99.82% of the variance in X block and 93.09% in Y block. The results for the optimization of the model, including FOM such as root mean square errors of calibration (RMSEC) and prediction (RMSEP), and RPD for calibration and validation sets, are showed in Table 1.

RMSEC and RMSEP are FOM related to the trueness of the multivariate calibration method. An improvement of these parameters was obtained after the outliers' detection, resulting in an RMSEC of 0.8% and an RMSEP of 0.9% v/v. An appropriate FOM to assess the predictive

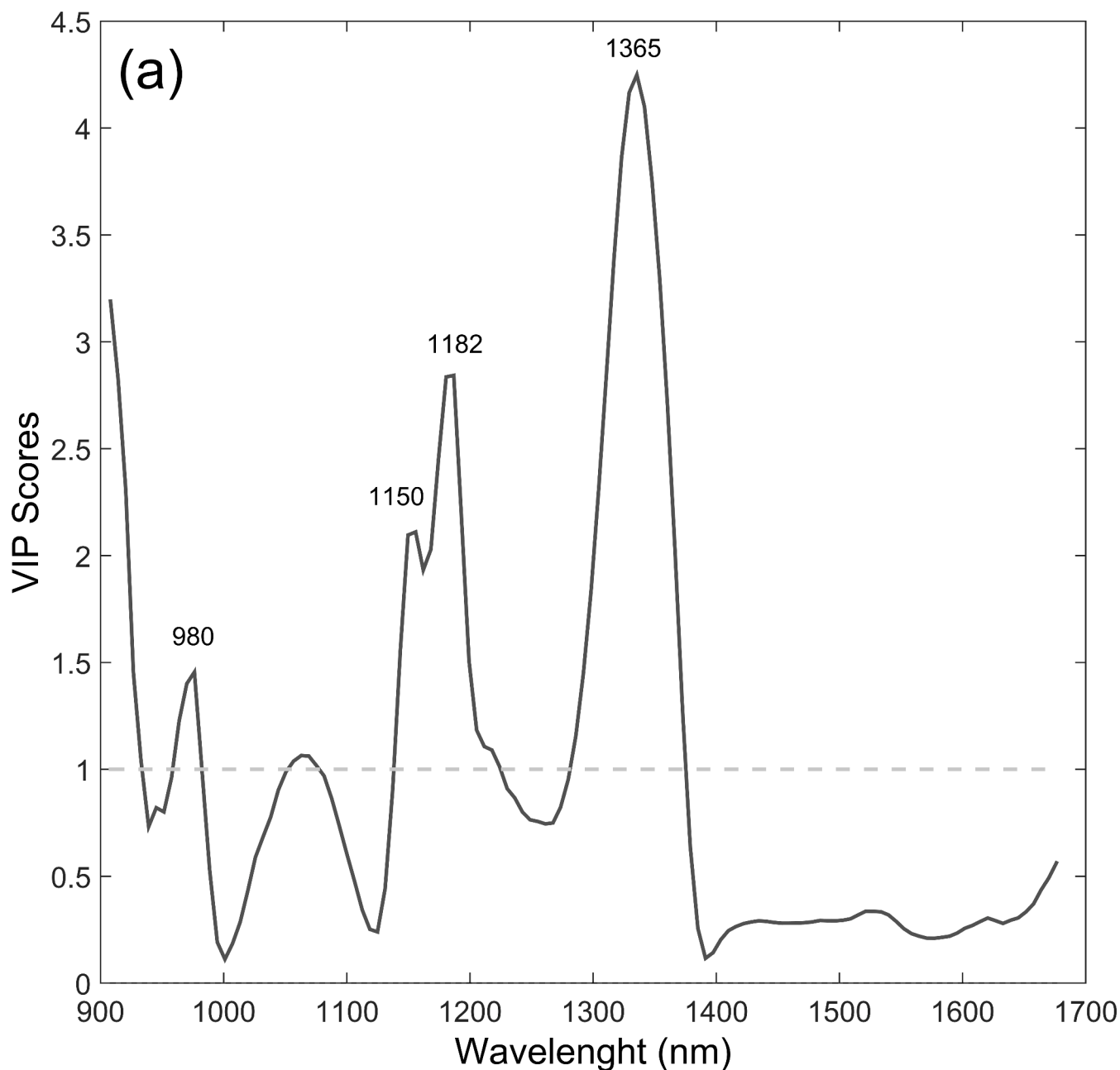


Fig. 3. Informative vectors for the PLS model. (a) Variable importance in projection (VIP) scores. Dashed horizontal line indicates the threshold of 1.0, above which variables are considered to contribute significantly to the PLS model. (b) Regression vector.

ability of NIR models in absolute terms is RPD, which is also a parameter for evaluating trueness. The concept of RPD was first introduced in 1993 and corresponds to the ratio of natural variation in samples to the size of probable prediction errors [46]. The developed PLS model presented RPD of 3.5 and 4.1 for the calibration and validation sets, respectively. These results are satisfactory, as PLS models with RPD values greater than 2.4 are considered to have a good predictive capacity [33,47]. The adjust can be evaluated through the regression plots of reference *versus* predicted values, which are shown in Fig. 2. The linearity will be evaluated by proper statistical tests presented in Table 2 and discussed later in this section, aiming to prove that residuals are randomly distributed. This figure represents the same fit as function of calibration and validation samples (Fig. 2a) and as a function of the type of beverage (Fig. 2b). It is possible to observe a difference in alcoholic strength amid the studied matrices. Wine and mead have the highest alcohol content,

ranging between 8% and 16%. Beer has the lowest content, ranging between 4% and 9%. Finally, cider has an intermediate alcohol content, between 7% to 14%.

The most important informative vectors extracted from the developed PLS model were used for spectral interpretation. Regression coefficients and variable importance in projection (VIP) scores are shown in Fig. 3. By jointly evaluating these plots, it is possible to identify the variables or bands most contributing to model predictions. The VIP scores (Fig. 3a) represent a measure of the degree of importance of each variable in the projection used by a particular model in absolute values. Complementarily, regression vectors (Fig. 3b) discriminate the direction of each variable contribution. The most positive coefficients contribute to the prediction of the target analyte, while the most negative ones can be associated with the interferences. Variables with VIP scores above a threshold of 1.0 are considered significant for the model predictions. It is

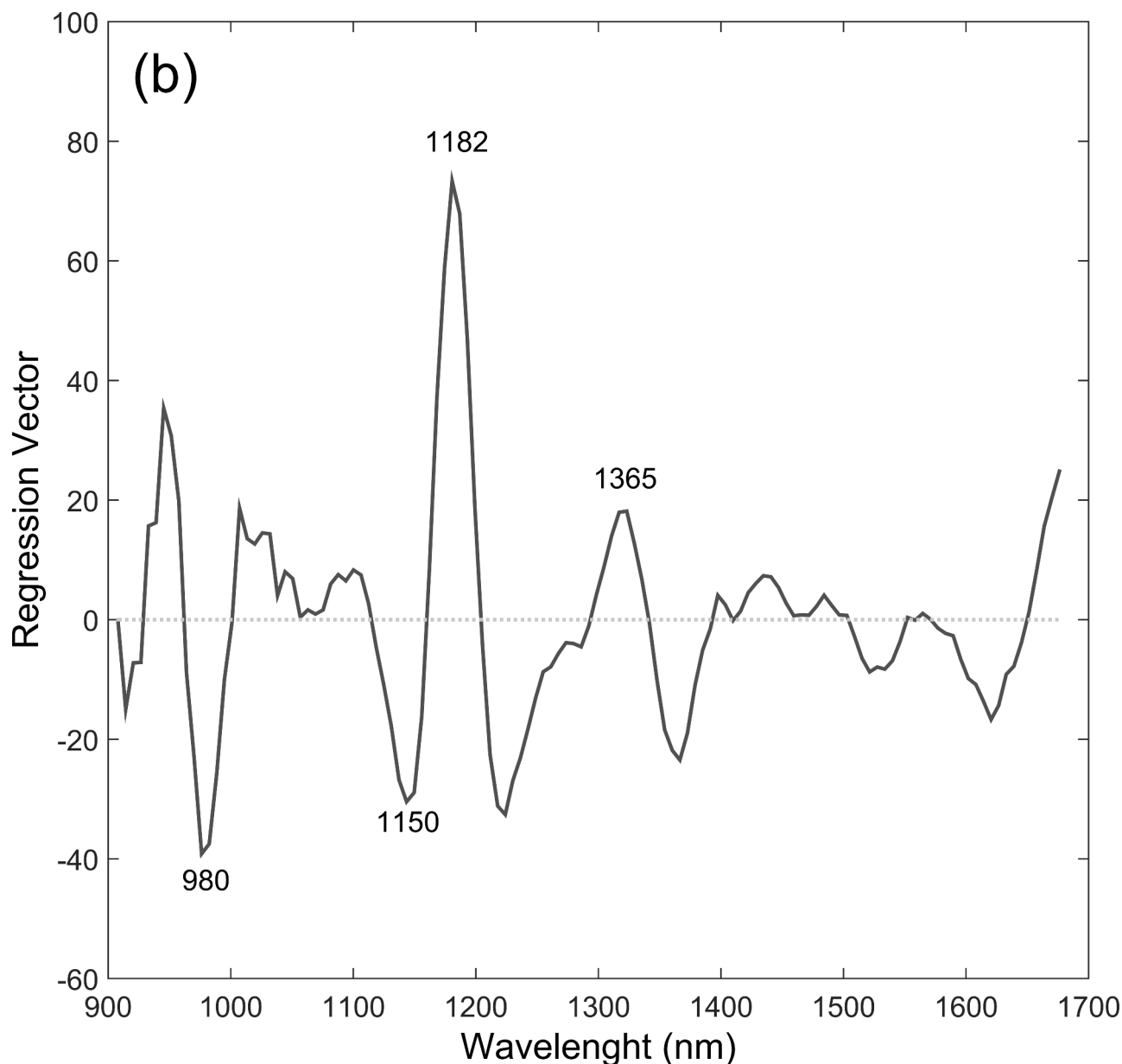


Fig. 3. (continued).

also important to note that NIR spectroscopy involves interaction between matter and light, and although the spectral bands may be assigned to vibrational modes of specific functional groups, they are not fully selective.

According to Fig. 3, there are four discriminant spectral bands to be highlighted, two positive and two negative regions of regression coefficients, which correspond to the four highest VIP scores peaks. The band centered around 980 nm can be assigned to the second overtone of the combination band of asymmetric and symmetric stretchings of the O—H bond of water, which is an interference in ethanol quantification. Spectral bands centered around 1150 nm and 1182 nm can be assigned to the second overtones of CH₃ and CH₂ stretchings of carbohydrates, which are common components of all analyzed beverages. Finally, the highest VIP score, at 1365 nm, is associated to the most predictive spectral band related to the target analyte. This band can be assigned to the first overtone of C—H stretching and the first overtone of the O—H stretching of ethanol [48]. It is interesting to note that the most intense NIR absorption band in the samples, above 1400 nm (Fig. 1) did not contribute significantly to the model, since the respective VIP scores

presented small values below the threshold of 1.0.

Table 2 presents complementary FOM assessed for the optimized PLS model. For evaluating linearity (Fig. 2), a correlation coefficient of 0.9312 ± 0.03 was estimated, jointly with a slope of 0.97 ± 0.03 and an intercept of 0.85 ± 0.22 . These results reinforced the accuracy results represented by RMSEC and RMSEP. In addition to these parameters, the linearity of the model should be verified by applying proper statistical tests to check the absence of systematic residuals. Thus, three tests were applied in the sequence: Ryan–Joiner for checking normality, Brown–Forsythe for checking homoscedasticity, and Durbin–Watson for confirming the absence of autocorrelation of the residuals, all at 95% confidence level [49]. The results of these tests assured the random behavior of the residuals, with an estimated value of 1.78 for Durbin–Watson test, within the acceptance range (1.50 - 2.50). The estimated small value of bias for the validation set, $0.3 \pm 0.9\%$, assured the absence of systematic errors in the model. A Student's *t*-test with 44 degrees of freedom corroborated the absence of a bias different from zero. Precision was evaluated as repeatability and intermediate precision at three different concentration levels for each of the four beverages, through estimating relative

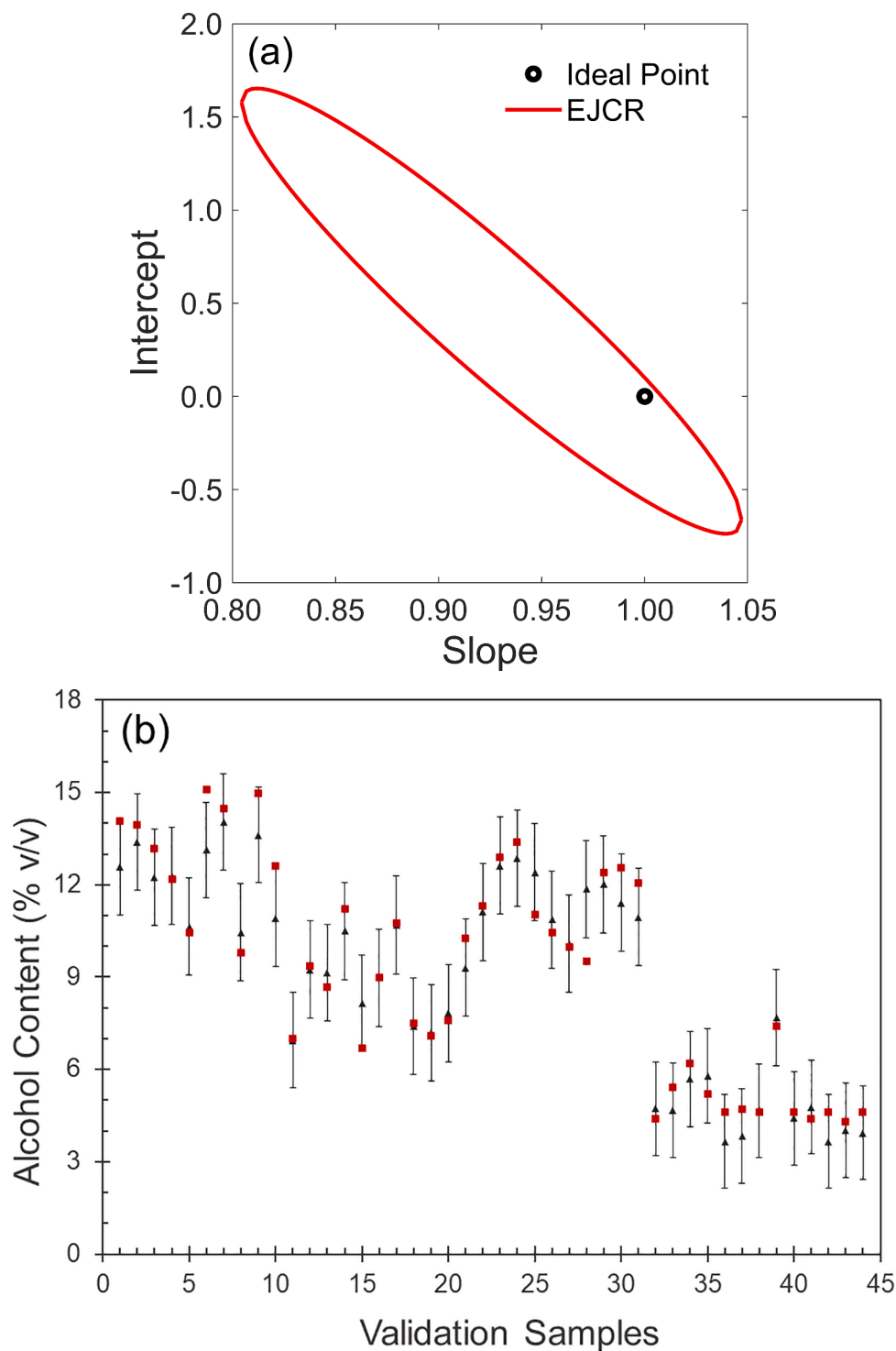


Fig. 4. (a) An elliptical joint confidence region (EJCR) estimated at 95% confidence level for the linearity (slope *versus* intercept) of the developed PLS model. (b) Confidence intervals estimated for the prediction of the validation samples. Sample-specific standard prediction errors (SPE) were calculated based on errors-in-variables (EIV) equation. gray up triangles mean predicted values and red squares mean the respective reference values.

standard deviation (RSD). RSD values varied between 0.1% and 1.7% at the level of repeatability and between 0.5% and 1.9% at the level of intermediate precision. These results were considered good, in accordance with the international guidelines that prescribe a maximum RSD of 3.7% for repeatability and 6.0% for intermediate precision, taking into account the level of the concentration of the analyte (above 1 g kg^{-1}) [10]. The analytical sensitivity was calculated from the estimate of the instrumental noise, which was obtained through the pooled standard

deviation of ten spectra replicates of an empty Petri dish [33]. Its inverse value, 0.9%, provides an estimate of the minimum concentration difference that the method can distinguish and established the use of only one decimal place to represent the results.

An elliptical joint confidence region (EJCR) was also estimated to confirm the linearity of the developed method (Fig. 4a). The limits of the estimated EJCR at 95% confidence level include the ideal point corresponding to an angular coefficient (SPE) of 1 and a linear coefficient

(intercept) of 0 for the plot of predicted versus reference values. Moreover, EJCRC can be considered an accuracy confirmation, showing no difference between the proposed method and the reference one. Therefore, the results of linearity, precision and trueness allowed to assure that the developed method can be considered accurate in the working range from 4.3% to 15.3%. Lastly, confidence intervals were estimated for the prediction of the validation samples. EIV equation was employed to estimated SPE varying between 1.5–1.6% (Fig. 4b). The reliability of the use of this equation was assured by the estimate of a non-significant bias, as previously mentioned in the last paragraph. Only two out of 44 validation samples were not predicted within the respective estimated confidence intervals (95%), representing an agreement of 95.5%. Therefore, the method was able to accurately quantify the alcohol content in beer, cider, mead, and wine with only one multivariate calibration model. The results shown in Fig. 4b also assured the trueness of the method jointly with other previously discussed FOM, such as RMSEC, RMSEP and RPD.

Conclusions

In this article, portability, vibrational spectroscopy and chemometrics were combined for the development of a direct, simple, rapid and environmentally friendly analytical method. Though the reference chromatographic method used solvents, once the model was established, the method will no longer consume reagents or solvents and can be considered green. A robust multi-product multivariate calibration model was built with NIR spectra for the quantification of alcohol content in four different fermented alcoholic beverages, demonstrating its applicability to real samples. The method was submitted to a full multivariate analytical validation through the estimate of proper figures of merit, thus ensuring its fit for purpose. The consistency of the model was also corroborated by a coherent spectral interpretation through analyzing its informative vectors. The developed method presented several advantages over traditional more laborious and time-consuming analytical alternatives based on densitometry or chromatography. These advantages allow to suggest the incorporation of the new method in portable analytical platforms for quality control of beverages, thus contributing to provide better transparency in the food supply chain. The synergy fostered by the combination of portability, NIR spectroscopy and a single multivariate calibration multi-product model can provide a green analytical tool for regulatory agencies performing quality inspection control of alcoholic beverages. This method may be expanded to incorporate other matrices, particularly distilled beverages, thus enlarging its analytical range.

Funding

This work was supported by Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), grant numbers APQ03457-16 and RED00042-16; and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), grant number 1383/2020.

CRediT authorship contribution statement

Ana Carolina da Costa Fulgêncio: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft. **Glaucimar Alex Passos Resende:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation. **Marden Claret Fontoura Teixeira:** Investigation, Data curation. **Bruno Gonçalves Botelho:** Conceptualization, Data curation, Investigation, Methodology, Resources, Supervision, Project administration, Funding acquisition. **Marcelo Martins Sena:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors would like to thank the Brazilian government agencies CNPq, FAPEMIG, and CAPES for financial support. ACCF acknowledges CAPES and GAPR acknowledges CNPq for Ph.D. scholarships. MCFT acknowledges CNPq for a scientific initiation scholarship. FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais) is especially acknowledged for the acquisition of the portable NIR spectrophotometer (Viavi MicroNIR® 1700) through the project APQ03457-16.

References

- [1] N. Ratanawimarnwong, M. Sinpun, P. Chankaw, N. Choengchana, D. Nacapricha, Simple flow system with in-line gas-diffusion unit for determination of ethanol employing hypsochromic shift of visible absorbance band of methyl orange, *Talanta* 206 (2020), 120234, <https://doi.org/10.1016/j.talanta.2019.120234>.
- [2] E. Sisco, E.L. Robinson, Determination of ethanol concentration in alcoholic beverages by direct analysis in real time mass spectrometry (DART-MS), *Forensic Chem.* 18 (2020), 100219, <https://doi.org/10.1016/j.forc.2020.100219>.
- [3] European Commission, Taxation and customs union - excise duty on alcohol. Brussels, 2020. https://ec.europa.eu/taxation_customs/taxation-1/excise-duties/excise-duty-alcohol_en (accessed in August 2022).
- [4] Department of Taxation and Finance of New York State, Alcoholic beverages tax. Albany, USA, 2022. <https://www.tax.ny.gov/bus/bev/abt.htm#beer> (accessed in August 2022).
- [5] Official Journal of the European Union: Regulation (EU) No 1168/2011 of the European Parliament and of the Council of 25 October 2011, Luxembourg, 2011.
- [6] European Commission, Labelling of alcoholic beverages in the EU: some facts. Brussels, 2022. https://ec.europa.eu/food/safety/labelling-and-nutrition/food-information-consumers-legislation/alcohol-labelling_en, (accessed in August 2022).
- [7] European Commission, Food information to consumers - legislation. Brussels, 2022. https://ec.europa.eu/food/safety/labelling-and-nutrition/food-information-consumers-legislation_en, (accessed in August 2022).
- [8] WHO – World Health Organization, Global status report on alcohol and health 2018. Geneva, Switzerland, 2018. <https://www.who.int/publications/i/item/9789241565639>, (accessed in August 2022).
- [9] M. Stupak, V. Kocourek, I. Kolouchova, J. Hajslova, Rapid approach for the determination of alcoholic strength and overall quality check of various spirit drinks and wines using GC-MS, *Food Control* 80 (2017) 307–313, <https://doi.org/10.1016/j.foodcont.2017.05.008>.
- [10] G.W. Latimer, *Official Methods of Analysis of AOAC International*, 19th ed., AOAC International, Gaithersburg, USA, 2012.
- [11] O. Wacheiko, P. Szpot, M. Zawadzki, The application of headspace gas chromatographic method for the determination of ethyl alcohol in craft beers, wines and soft drinks, *Food Chem.* 346 (2021), 128924, <https://doi.org/10.1016/j.foodchem.2020.128924>.
- [12] T. Yarita, R. Nakajima, S. Otsuka, T. Ihara, A. Takatsu, M. Shibukawa, Determination of ethanol in alcoholic beverages by high-performance liquid chromatography-flame ionization detection using pure water as mobile phase, *J. Chromatogr. A* 976 (2002) 387–391, [https://doi.org/10.1016/S0021-9673\(02\)00942-1](https://doi.org/10.1016/S0021-9673(02)00942-1).
- [13] T.R.L.C. Paixão, D. Corbo, M. Bertotti, Amperometric determination of ethanol in beverages at copper electrodes in alkaline medium, *Anal. Chim. Acta* 472 (2002) 123–131, [https://doi.org/10.1016/S0003-2670\(02\)00942-X](https://doi.org/10.1016/S0003-2670(02)00942-X).
- [14] H. Nakamura, R. Tanaka, K. Suzuki, M. Yataka, Y. Mogi, A direct determination method for ethanol concentrations in alcoholic beverages employing a eukaryote double-mediator system, *Food Chem* 117 (2009) 509–513, <https://doi.org/10.1016/j.foodchem.2009.04.026>.
- [15] A.C.N. Pinheiro, V.S. Ferreira, B.G. Lucca, Stamping method based on 3D printing and disposable napkin: cheap production of paper analytical devices for alcohol determination in beverages aiming forensics and food control, *Microchem. J.* 180 (2022), 107604, <https://doi.org/10.1016/j.microc.2022.107604>.
- [16] T. Somboona, S. Sansuk, An instrument-free method based on visible chemical waves for quantifying the ethanol content in alcoholic beverages, *Food Chem.* 253 (2018) 300–304, <https://doi.org/10.1016/j.foodchem.2018.01.168>.
- [17] S. Sansuk, P. Juntarakod, W. Tongphoothorn, A. Sirimungkala, T. Somboon, Visual chemo-chronometric assay for quantifying ethanol in alcoholic drinks by the

- colorimetric Belousov-Zhabotinsky oscillator, *Food Control* 110 (2020), 107042, <https://doi.org/10.1016/j.foodcont.2019.107042>.
- [18] M.F. Filgueiras, B.O. Lima, E.M. Borges, A high-throughput, cheap, and green method for determination of ethanol in cachaça and vodka using 96-well-plate images, *Talanta* 241 (2022), 123229, <https://doi.org/10.1016/j.talanta.2022.123229>.
- [19] A. Zuriarrain, J. Zuriarrain, M. Villar, I. Berregi, Quantitative determination of ethanol in cider by ¹H NMR spectrometry, *Food Control* 50 (2015) 758–762, <https://doi.org/10.1016/j.foodcont.2014.10.024>.
- [20] A. Gatuszka, Z.M. Migaszewski, J. Namieśnik, Moving your laboratories to the field—Advantages and limitations of the use of field portable instruments in environmental sample analysis, *Environ. Res.* 140 (2015) 593–603, <https://doi.org/10.1016/j.envres.2015.05.017>.
- [21] R.A. Crocombe, Portable spectroscopy, *Appl. Spectrosc.* 72 (2018) 1701–1751, <https://doi.org/10.1177/0003702818809719>.
- [22] K.C.A. Rezende, L.M. Duarte, K.M.P. Pinheiro, T.M.G. Cardoso, S.A. Nogueira, W.K. T. Coltro, Portable analytical platforms associated with chemometrics for rapid screening of whisky adulteration, *Food Anal. Methods* 15 (2022) 2451–2461, <https://doi.org/10.1007/s12161-022-02303-5>.
- [23] C. McVey, C.T. Elliott, A. Cannavan, S.D. Kelly, A. Petchkongkaew, S.A. Haughey, Portable spectroscopy for high throughput food authenticity screening: advancements in technology and integration into digital traceability systems, *Trends Food Sci. Technol.* 118 (2021) 777–790, <https://doi.org/10.1016/j.tifs.2021.11.003>.
- [24] C. Zhu, X. Fu, J. Zhang, K. Qin, C. Wu, Review of portable near infrared spectrometers: current status and new techniques, *J. Near Infrared Spectrosc.* 30 (2022) 51–66, <https://doi.org/10.1177/09670335211030617>.
- [25] A.C.C. Fulgêncio, G.A.P. Resende, M.C.F. Teixeira, B.G. Botelho, M.M. Sena, Determination of alcohol content in beers of different styles based on portable near-infrared spectroscopy and multivariate calibration, *Food Anal. Methods* 15 (2022) 307–316, <https://doi.org/10.1007/s12161-021-02126-w>.
- [26] E. Micklander, K. Kjeldahl, M. Egebo, L. Norgaard, Multi-product calibration models of near infrared spectra of foods, *J. Near Infrared Spectrosc.* 14 (2006) 395–402, <https://doi.org/10.1255/jnirs.659>.
- [27] X. Fan, S. Tang, G. Li, X. Zhou, Non-invasive detection of protein content in several types of plant feed materials using a hybrid near infrared spectroscopy model, *PLoS ONE* 11 (2016), e0163145, <https://doi.org/10.1371/journal.pone.0163145>.
- [28] R.E. Masithoh, S. Lohumi, W.-S. Yoon, H.Z. Amanah, B.-K. Cho, Development of multi-product calibration models of various root and tuber powders by fourier transform near infra-red (FT-NIR) spectroscopy for the quantification of polysaccharide contents, *Heliyon* 6 (2020) e05099, <https://doi.org/10.1016/j.heliyon.2020.e05099>.
- [29] M.K.D. Rambo, M.M.C. Ferreira, E.P. Amorim, Multi-product calibration models using NIR spectroscopy, *Chemom. Intell. Lab. Syst.* 151 (2016) 108–114, <https://doi.org/10.1016/j.chemolab.2015.12.013>.
- [30] D.A. Santos, K.P. Lima, V. Cavalcante, A. Coqueiro, M.F.B. Consolin, N. Consolin Filho, P.H. Março, P. Valderrama, Multiproduct, multicomponent and multivariate calibration: a case study by using vis-NIR spectroscopy, *Food Anal. Methods* 11 (2018) 1915–1919, <https://doi.org/10.1007/s12161-017-1099-4>.
- [31] V. González-Caballero, D. Pérez-Marín, M.-I. López, M.-T. Sánchez, Optimization of NIR spectral data management for quality control of grape bunches during on-vine ripening, *Sensors* 11 (2011) 6109–6124, <https://doi.org/10.3390/s110606109>.
- [32] Y. Zheng, X. Lai, S.W. Bruun, H. Ipsen, J.N. Larsen, H. Löwenstein, I. Søndergaard, S. Jacobsen, Determination of moisture content of lyophilized allergen vaccines by NIR spectroscopy, *J. Pharm. Biomed. Anal.* 46 (2008) 592–596, <https://doi.org/10.1016/j.jpba.2007.11.011>.
- [33] B.G. Botelho, B.A.P. Mendes, M.M. Sena, Development and analytical validation of robust near-infrared multivariate calibration models for the quality inspection control of mozzarella cheese, *Food Anal. Methods* 6 (2013) 881–891, <https://doi.org/10.1007/s12161-012-9498-z>.
- [34] B.G. Botelho, L.P. Assis, M.M. Sena, Development and analytical validation of a simple multivariate calibration method using digital scanner images for sunset yellow determination in soft beverages, *Food Chem.* 159 (2014) 175–180, <https://doi.org/10.1016/j.foodchem.2014.03.048>.
- [35] R.W. Kennard, L.A. Stone, Computer aided design of experiments, *Technometrics* 11 (1969) 137–148, <https://doi.org/10.2307/1266770>.
- [36] R.D.A. Ferreira, G. Teixeira, L.A. Peternelli, Kennard-Stone method outperforms the Random Sampling in the selection of calibration samples in SNPs and NIR data, *Cienc. Rural* 52 (2021), e20201072, <https://doi.org/10.1590/0103-8478cr20201072>.
- [37] Brasil., *Dispõe Sobre a Padronização, a Classificação, o Registro, a Inspeção, a Produção e a Fiscalização de Bebidas; Regulamenta a Lei no 8918, De 14 De Julho De 1994 (Decreto no 6.871, De 04 De Junho De 2009), Diário Oficial da República Federativa do Brasil, Brasília, 2009.*
- [38] M. Thompson, S.L.R. Ellison, R. Wood, Harmonized guidelines for single-laboratory validation of methods of analysis (IUPAC Technical Report), *Pure Appl. Chem.* 74 (2002) 835–855, <https://doi.org/10.1351/pac200274050835>.
- [39] ASTM, *Standard Practices for Infrared Multivariate Quantitative Analysis - E1655-05*, ASTM International, West Conshohocken, USA, 2012.
- [40] N.M. Faber, X.H. Song, P.K. Hopke, Sample-specific standard error of prediction for partial least squares regression, *TrAC - Trends Anal. Chem.* 22 (2003) 330–334, [https://doi.org/10.1016/S0165-9936\(03\)00503-X](https://doi.org/10.1016/S0165-9936(03)00503-X).
- [41] A.S. Luna, F.B. Gonzaga, W.F. Rocha, I.C. Lima, A comparison of different strategies in multivariate regression models for the direct determination of Mn, Cr, and Ni in steel samples using laser-induced breakdown spectroscopy, *Spectrochim. Acta B* 139 (2018) 20–26, <https://doi.org/10.1016/j.sab.2017.10.016>.
- [42] M. Ito, T. Suzuki, S. Yada, H. Nakagami, H. Teramoto, E. Yonemochi, K. Terada, Development of a method for nondestructive NIR transmittance spectroscopic analysis of acetaminophen and caffeine anhydrate in intact bilayer tablets, *J. Pharm. Biomed. Anal.* 53 (2010) 396–402, <https://doi.org/10.1016/j.jpba.2010.04.029>.
- [43] A. Rinnan, F. van den Berg, S.B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, *TrAC - Trends in Anal. Chem.* 28 (2009) 1201–1222, <https://doi.org/10.1016/j.trac.2009.07.007>.
- [44] A.C.C. Fulgêncio, G.A.P. Resende, M.C.F. Teixeira, B.G. Botelho, M.M. Sena, Screening method for the rapid detection of diethylene glycol in beer based on chemometrics and portable near-infrared spectroscopy, *Food Chem* 391 (2022), 133258, <https://doi.org/10.1016/j.foodchem.2022.133258>.
- [45] MAPA, Ministério da Agricultura, Pecuária e Abastecimento, *Manual de Garantia da Qualidade Analítica, Áreas de Identidade e Qualidade de Alimentos e de Insumos*, Brasília, 2014.
- [46] P.C. Williams, D.C. Sobering, Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds, *J. Near Infrared Spectrosc.* 1 (1993) 25–32, <https://doi.org/10.1255/jnirs.3>.
- [47] P.C. Williams, Implementation of near-infrared technology, in: P.C. Williams, K. Norris (Eds.), *Near-infrared Technology in the Agricultural and Food Industries*, American Association of Cereal Chemists, St. Paul, 2001, pp. 145–169.
- [48] B.H. Stuart, *Infrared Spectroscopy: Fundamentals and Applications*, John Wiley & Sons, New York, 2004.
- [49] S.V.C. Souza, R.G. Junqueira, A procedure to assess linearity by ordinary least squares method, *Anal. Chim. Acta* 552 (2005) 25–35, <https://doi.org/10.1016/j.aca.2005.07.043>.



Marcelo Martins Sena is an Associate Professor at the Federal University of Minas Gerais, Belo Horizonte, Brazil, and Research Fellow from the Brazilian National Council for Scientific and Technological Development (CNPq). He holds a Ph. D. in Chemistry, and his current work focuses on chemometrics in analytical chemistry, including the application of molecular spectroscopic techniques (NIR, MIR, UV–vis, spectrofluorimetry, MS) to the analysis of food, pharmaceuticals, agricultural products and forensic samples. He has also published around 75 articles in international peer-reviewed journals.