Prediction of yerba mate caffeine content using near infrared spectroscopy

*Gabriel Goetten de Lima [1,2], Henrique Zavattieri Ruiz [3], Mailson Matos [2], Cristiane Vieira Helm [4], Marcus Vinicius de Liz [3], Washington Luiz Esteves Magalhães [4]**

[1] Materials Research Institute, Athlone Institute of Technology, Athlone, Ireland.

[2] Integrated Engineering and Materials Sciences Program, Federal University of Paraná, Curitiba, Paraná, Brazil

[3] Academic Department of Chemistry and Biology, Federal University of Technology – Paraná (UTFPR), Curitiba, Brazil

[4] Embrapa Forestry, Colombo, Brazil.

* **Email address**: washington.magalhaes@embrapa.br

Abstract

There is a commercial and beneficial interest of producing yerba mate leaves into different grades of caffeine. This work uses a handheld and bench near infrared spectroscopy to compare and predict, using partial least squares regression, the amount of caffeine in yerba mate leaves. Standards of pure caffeine were compared, using high-performance liquid chromatography, with extracts of yerba mate. The bench spectroscopy gave a strong confidence model of caffeine prediction, whereas the handheld related to a fair model. For first detection and initial separation of yerba mate in the field, the modelling proposed can be used to predict caffeine intensity.

Keywords: yerba-mate, caffeine, Near infrared spectroscopy, Liquid chromatography,

1. Introduction

Ilex paraguariensis St Hilaire (Aquifoliaceae) leaves are used as tea-like beverage, commonly named as yerba mate tea, and is really important culturally and economically in South America, more specifically in South Brazil, Argentine, Paraguay, and Uruguay [1,2],

which are also the regions where it is cultivated. Nonetheless, yerba mate extracts are exported to USA, Asia, Europe, Syria and Lebanon [2,3] which are now becoming popular [4]. Yerba mate contains many antioxidant [5] and anti-inflammatory [6] ingredients and are found in large quantity [7], this proved to be effective against oxidative damage to DNA [8] and have potential to scavenge free radicals [9]. It has also been suggested that ingestion of yerba mate can be an effective and economic way to provide prevention against cardiovascular diseases [10]. In addition, mate also shows a high cytotoxicity for cancer cells [11].

This hot beverage is mainly consumed because of its stimulating factors, methylxanthine compound, which contains mainly caffeine and small amounts of theobromine [12]. The yerba mate tea cup content of caffeine is comparable to one of coffe [1]; however, as pointed out by Mazzafera *et al.* (1997) the way this tea is consumed which involves repeatedly pouring additional hot water over in the 'mate' can yield intakes greater than 260 mg of caffeine per serving, attributed to percent stem or woody content and extraction rate [13].

Depending on the harvest season, the quantity of these compounds in the leaves of the yerba-mate varies and the need to quantify the amount of these compounds with an easy and fast approach is needed [14]. Furthermore, it is commercially interesting to selective quantify the amount of caffeine [14] for each leaf to obtain different grades of yerba mate teas before it is processed. For these reasons, the purposed work performed in this manuscript presents an easy method to separate the amount of caffeine in different grades, or intensities; the idea is based on using a portable and bench near infrared spectroscopy device with the help of a model to quantify the amount of caffeine content for each tree. The model was achieved using partial least squares to find a correlation between NIR with the quantified compounds obtained via HPLC.

The advantages of using near infrared spectroscopy are simplicity in the pre-treatment of the sample and speed of obtaining the data; also, the cost of quantification when compared to chromatographic methods [15].

To predict quantity of compounds, one of the most used methods is through the partial least squares regression, PLS, which allows to evaluate the covariance between the near infrared spectrum and a variable of interest. If the evaluation presents a good correlation coefficient, it is possible to create a model that allows estimating the quantification of the desired variable only with the NIR spectrum [16]. Although the quantification of methylxanthines and phenolic compounds have already been evaluated on different substrates using the partial least squares regression method using near-infrared spectroscopy [15,17], the analysis on yerba mate leaves is particularly attractive given the potential interest in this field.

2. Methodology

2.1 Materials and initial data

For this work, 55 yerba mate trees with average of 1.6 m in height were employed during the harvest season, which occurs during May to September, in the month of July with prevalence of mature leaves. In the field, NIR spectra were collected from plucked leaves, five for each tree, where they were immediately placed over a flat surface to assist in the readings of the portable spectrometer Phazir Handheld NIR System (Polychromix). The spectra were collected between wavelengths of 1600 nm to 2350 nm, with resolution of 10.5 nm at diffuse reflection mode. Data collected was in absorbance, i.e. log (1/R) form, and were treated using the software Unscrambler® X (CAMO PROCESS AS, Norway). Five spectra of distinct leaves from the same tree were averaged into one spectrum. Thus, were obtained 55 spectra corresponding to the 55 yerba mate trees. The leaves used in this work belongs to Embrapa Forestry breeding program where cultivars were found with very high and very low content of caffeine .

2.2 Yerba-Mate Leaves Sample Preparation and spectra collection with Bench Near-infrared Spectroscopy

After collection of the spectra using a portable device, around 10 leaves were plucked from each tree, at random without a specific preference from age of the leaves. The samples were dried in a microwave oven (Electrolux MEF41, power 1150W, frequency 2450 MHz) for approximately 3 minutes, the leaves being turned every minute for a more efficient drying until the leaves were crisp. The leaves were ground in a blender (Britânia Diamante Black Filter) and then sieved through a 25-mesh and 710 μm aperture before collecting the NIR spectra using bench NIR 900 (FEMTO, Brazil). The spectra were collected between wavelengths of 1100 nm to 2500 nm, with resolution of 2 nm using diffuse reflection mode. The same samples were used to extract caffeine and theobromine, which were subsequently characterized through high-performance liquid chromatography technique (HPLC).

Both portable and bench NIR used in this work can be operated in reflection and transmission modes with a spectral range of 900-1600 nm and 1100-2500 nm and maximum resolution of 10.5 nm and 2 nm respectively. The light source of the equipment is a tungsten-halogen lamp.

2.3 High-performance liquid chromatography Measurements

For the HPLC analyses, the Agilent (1260 Infinity Quaternary LC system) high efficiency liquid chromatograph equipped with automated injector and UV-DAD detector was used. Compounds were separated on C18 column (DIONEX, Acclaim 120 Å, 3 μm, 2.1 x 150 mm) with pre-column C18 (DIONEX, Acclaim 120 Å, 5 μm, 2 x 10 mm). The flow rate was 0.300 ml.min-1. Detection of caffeine and theobromine, derivatives of methylxanthines, was performed at fixed wavelength at 280 nm by the injection of 10 μL of aqueous extract. The mobile phase consisted of 0.5% acetic acid HPLC (Phase A) and

acetonitrile solution HPLC (Phase B), both solutions were filtered on 0.45 μm membranes and homogenized.

The caffeine standard with purity greater than 99.0% (Sigma Aldrich) was used in the preparation of stock solutions in ultrapure water. The calibration profiles were performed by diluting the stock solution in the mobile phase to furnish solutions with final concentrations of 1.5, 3.0, 15.0, 45.0, 75.0, 105.0, 135.0, 150.0 mg/L.

Three analytical curves were made for caffeine from the areas of the chromatographic peaks, corresponding to the concentration of each compound, and the area averages, for each concentration evaluated, were used to quantify these compounds in the extracts of yerba-mate.

To prepare the extracts for the HPLC, about 0.1 grams of previously thawed, crushed and sieved yerba-mate samples were weighed directly into a graduated 50 mL falcon tube. In each falcon, 25 mL of ultrapure water, previously heated to its boiling temperature (95.0 ± 2.0 ºC), were followed by extraction of its main components in Ultrasound (Ultracleaner 1400A) for 15 minutes. After cooling at room temperature, the extracts were filtered through a 0.45 μm nylon membrane using a syringe and holder.

The extracts were prepared with boiling water to simulate the preparation of the yerba-mate tea, commonly consumed in the South America.

About 10 μl of the filtered yerba-mate extract was injected into the HPLC and from the areas of the peaks, a correlation of caffeine concentration with the weight of extracted yerba-mate was performed to obtain the amount of caffeine in the extract.

2.4 Multivariate Partial Least Squares Regression

Unscrambler® X software was used to verify the correlation between the data by partial least squares (PLS) analysis. This software allowed to evaluate the region of the near

infrared spectrum that presented greater covariance with the realized quantification. The estimation models were evaluated according to these regions of the spectrum.

Both set of spectra – handheld and bench NIR – were treated using Unscrambler software and PLS was performed (projection to latent structure) through partial least square algorithm in order to construct a model to predict the caffeine and theobromine content. The model was based on the data obtained by near infrared spectroscopy and the values obtained from HPLC. Although smoothing was performed with both first and second derivatives of the signal, as well with standard normal variate (SNV), multiple scatter correction (MSC) and combination of both first and second derivative with SNV and MSC; the raw spectra exhibited a better profile (Supplemental file - Table.A.1). Forty-three samples were used to construct the multivariate model and the 12 remaining samples were used to validate the model. Therefore, these 12 samples were considered as an external validation point to the robustness of the procedure.

Regression coefficients; coefficients of the calibration equation were also obtained by the Unscrambler® X software and plotted. In the graph obtained, the y-axis corresponds to B_i, and has the value of B_0 on the x-axis. So, the equation would be as described in equation (1).

$$\text{Caffeine content} = B_0 + B_1 \times A_{L1} + B_2 \times A_{L2} + \cdots \quad (1)$$

where A_{Li} corresponds to the absorbance at the specific wavelength Li.

Therefore, the absorbance spectrum of an unknown yerba-mate sample obtained via NIR was modelled using the most statistically relevant values of absorbance points multiplied by their respective coefficients, which were represented in a graph plot, and added with the constant B_0 making it able to predict the caffeine content.

The same procedure was attempted for the prediction of theobromine; however, due to low theobromine concentration, the lack of qualified calibration and the resultant poor prediction model, these results were omitted.

3. Results and discussion

The quantification by HPLC reported maximum and minimum values of caffeine of 2.39 g / 100 g and 0.01 g/100 g respectively; all extracts presented caffeine concentration within the limits of the prepared analytical curve and the average of caffeine among all samples evaluated was (0.83 ± 0.63) g/100g. The standard deviation obtained by HPLC demonstrate the great variation in the content of caffeine that can occur between leaves obtained from the same harvesting site.

Although the levels of caffeine are dependent on the extraction method due to their solubility in different solvents and pHs [18], caffeine values obtained herein were similar to other works using the same hot-water extraction [19]; however, these levels might vary depending on the region where the leaves were collected, such as 0.73 g/100 g [18] in the province of Corrientes, Argentina. Nonetheless, there is already a report of caffeine content from yerba-mate leaves obtained from the same harvesting site (Ivaí, Paraná) and the same hot-water extraction used in this work, the caffeine values ranged from (0.01 to 1.01 g)/100 g which was similar to what this work has reported.

The multivariate model based from the bench NIR spectra of dried leaves from yerba mate trees correlates with caffeine content measured through HPLC method (Figs. 1.i and 2.i). The standard error of calibration was 0.3 mg/kg with a coefficient of determination as high as 0.8.

The regression coefficients (shown in Fig. 2) that are statistically significant coincides with the highest absorption peak of the pure caffeine using bench equipment (Fig. 3).

The importance of the multivariate PLS can be observed in Figure 3, where the highest absorption peak of pure caffeine was not identified by the normal scan of NIR spectra of mate leaves using the bench equipment, even in the spectrum of the leaf with the highest content of caffeine; however, the multivariate PLS was able to detect and predict the caffeine content.

To determine the accuracy of this model, the content of caffeine predicated by the multivariate model was compared to HPLC results for samples that did not participate in the construction of model (external validation) - see Figures 1.iii and 1.iv. The coefficient of determination of this external test was as high as 0.81 presenting a standard error of prediction equal to 0.26 mg/kg, exhibiting a robust prediction for caffeine content.

The PLS multivariate model of the handheld NIR to predict caffeine exhibits a prediction error (Figure 1.ii) around 0.30 mg with a coefficient of determination around 0.78 (Table 1). These values are similar to that obtained for the NIR bench spectra.

Moreover, the predictive caffeine region of the raw handheld NIR spectra – 2120 to 2255 nm - is close to the region used for bench NIR (Fig. 2.i-ii). Although, the regions used to construct the prediction models for both equipment have overlapping frequencies, a noncoincident profile is perceived. This is due to the reduced number of absorbance points obtained and worse optical resolution of handheld NIR.

The external validation for handheld NIR model (Fig. 1.iv) exhibits a standard error value for caffeine prediction content equal to 0.38 mg/kg (Table 1). This error is higher than the external validation, by 27%, performed using spectra from NIR bench although the calibration model initially appears to have similar performance.

The ratio of performance to standard deviation (RPD = SD/SEP) is regularly used by researchers [20–22] to infer the usefulness of multivariate prediction models. Most researchers

agree that the model is excellent for RPD >2; fair models, with 1.4 < RPD < 2; and, non-reliable models, with RPD <1.4.

Strong water absorption bands (1430–1470 nm and 1920–1960 nm) [21] obtained in the NIR spectra can alter the quantitation of active ingredients through NIR. Consequently, removal of water in the sample on experimental pre-treatment and exclusion of water absorption bands in the estimated NIR spectra for PLS regression becomes the general procedure for quantitative analysis.

Therefore, the results could have potentially been weakened in the PLS model of yerba-mate when measuring the total caffeine content of leaves. However, the water interference on the developed PLS model did not influence so significantly because the loading of water absorption band being included in the models are small and the region used to develop the caffeine PLS model was outside the region of the main water absorption bands.

Nonetheless, studies on the effect of water in leaves for the effectiveness of PLS model's quantification of natural products are scarce in literature and one of the few who contributed to these studies, Chan et al., (2007) [22] reported that removing the water band to produce the PLS alkaloid content model decreases the root mean of square error of prediction set from 8.06% to 5.92%, a difference of 2.14%. In addition, the root mean of square error of cross-validation between the dried sample and pure sample leaves gave a difference value of 0.2, in which they attributed to physical interferences like humidity, temperature and packing density which affects the NIR spectra.

The calibration data indicates that the PLS model will have difficulties in recognizing samples with low concentrations (below 0.2 g / 100 g), as they estimate very different values from those determined by high performance liquid chromatography. Nonetheless, it can be useful for quantities higher than 0.2 g/100g; likewise, the values obtained by HPLC of

caffeine from all the leaves were out of the error field from this region, meaning that yerba-mate leaves will mostly have higher caffeine quantity than the model error range.

Nonetheless, it is important to state that the time period which the leaves were collected corresponds to the harvesting season of yerba-mate [23]; therefore, the leaves should contain the highest total levels of compounds, including caffeine.

One important aspect of PLS is that it takes into account errors both in the concentration estimation and the spectra [24]. Although at first it seems that other methylxanthine groups might affect the results of the model [24]; the intensities of other groups from yerba-mate leaves are low – theobromine ~20x lower than caffeine and other groups are even lower than this [25]. Moreover, NIR equipment cannot measure such low values and, we believe, this quantity will not impact the model.

According to the calculated RPD (Table 1), for bench NIR it can be deduced, with the standard consensus on RPD, that the model is excellent for caffeine prediction and is a fair model for handheld NIR. However, for the first screening purpose to select yerba mate plants with high or low caffeine content, handheld equipment will be useful. Mainly for field applications if one keeps in mind the fast reading of the measurement, the absence of any pretreatment of the leaves, no need for highly trained personnel, and low cost of the whole procedure.

4. Conclusion

This work attempts to produce a model using PLS to predict caffeine content of yerba mate leaves that can be used in the field with fast readings using a fairly simple equipment - a handheld NIR - which were also modelled and compared to a bench NIR equipment. The results herein demonstrated that it is possible to use this model as a first selection of caffeine grade tool in raw bulk yerba mate leaves, with, as suggested, later confirmation and improvement in selection using bench near infrared equipment.

Conflicts of Interest

The authors declare that there is no conflicts of interest regarding the publication of this manuscript.

Acknowledgment

This work was supported by research grant Capes.

References

1. Heck, C.I.; De Mejia, E.G. Yerba mate tea (Ilex paraguariensis): A comprehensive review on chemistry, health implications, and technological considerations. Journal of Food Science 2007, 72(9) 10.1111/j.1750-3841.2007.00535.x.
2. Kujawska, M. Yerba Mate (Ilex paraguariensis) Beverage: Nutraceutical Ingredient or Conveyor for the Intake of Medicinal Plants? Evidence from Paraguayan Folk Medicine. Evidence-based Complementary and Alternative Medicine 2018, 2018 10.1155/2018/6849317.
3. Butiuk, A.P.; Maidana, S.A.; Martos, M.A.; Akakabe, Y.; Adachi, O.; Hours, R.A. Characterization and application of fungal chlorogenate hydrolase to enzymatic breaking down of chlorogenate from yerba mate. Biocatalysis and Agricultural Biotechnology Elsevier Ltd: 2018, 14(January), 395–401 10.1016/j.bcab.2018.04.005.
4. Frizon, C.; Perussello, C.; Sturion, J.; Hoffmann-Ribani, R. Novel Beverages of Yerba-Mate and Soy: Bioactive Compounds and Functional Properties. Beverages 2018, 4(1), 21 10.3390/beverages4010021.
5. Berté, K.A.S.; Beux, M.R.; Spada, P.K.W.D.S.; Salvador, M.; Hoffmann-Ribani, R. Chemical composition and antioxidant activity of yerba-mate (Ilex paraguariensis A.St.-Hil., Aquifoliaceae) extract as obtained by spray drying. Journal of Agricultural and Food Chemistry 2011, 59(10), 5523–5527 10.1021/jf2008343.
6. Gómez-Juaristi, M.; Martínez-López, S.; Sarria, B.; Bravo, L.; Mateos, R. Absorption

and metabolism of yerba mate phenolic compounds in humans. Food Chemistry Elsevier: 2018, 240(August 2017), 1028–1038 10.1016/j.foodchem.2017.08.003.

7. Kungel, P.T.A.N.; Correa, V.G.; Corrêa, R.C.G.; Peralta, R.A.; Soković, M.; Calhelha, R.C.; et al. Antioxidant and antimicrobial activities of a purified polysaccharide from yerba mate (Ilex paraguariensis). International Journal of Biological Macromolecules 2018, 114, 1161–1167 https://doi.org/10.1016/j.ijbiomac.2018.04.020.
8. Gonzalez De Mejia, E.; Young, S.S.; Ramirez-Mares, M.V.; Kobayashi, H. Effect of yerba mate (Ilex paraguariensis) tea on topoisomerase inhibition and oral carcinoma cell proliferation. Journal of Agricultural and Food Chemistry 2005, 53(6), 1966–1973 10.1021/jf048158g.
9. Riachi, L.G.; Simas, D.L.R.; Coelho, G.C.; Marcellini, P.S.; Ribeiro da Silva, A.J.; Bastos de Maria, C.A. Effect of light intensity and processing conditions on bioactive compounds in maté extracted from yerba mate (Ilex paraguariensis A. St.-Hil.). Food Chemistry 2018, 266, 317–322 https://doi.org/10.1016/j.foodchem.2018.06.028.
10. Gebara, K.S.; Gasparotto-Junior, A.; Santiago, P.G.; Cardoso, C.A.L.; De Souza, L.M.; Morand, C.; et al. Daily Intake of Chlorogenic Acids from Consumption of Maté (Ilex paraguariensis A.St.-Hil.) Traditional Beverages. Journal of Agricultural and Food Chemistry 2017, 65(46), 10093–10100 10.1021/acs.jafc.7b04093.
11. de Mejía, E.G.; Song, Y.S.; Heck, C.I.; Ramírez-Mares, M. Yerba mate tea (Ilex paraguariensis): Phenolics, antioxidant capacity and in vitro inhibition of colon cancer cell proliferation. Journal of Functional Foods 2010, 2(1), 23–34 https://doi.org/10.1016/j.jff.2009.12.003.
12. Oellig, C.; Schunck, J.; Schwack, W. Determination of caffeine, theobromine and theophylline in Mate beer and Mate soft drinks by high-performance thin-layer chromatography. Journal of Chromatography A 2018, 1533, 208–212

https://doi.org/10.1016/j.chroma.2017.12.019.

13. Mazzafera, P. Maté drinking: caffeine and phenolic acid intake. Food Chemistry 1997, 60(1), 67–71 https://doi.org/10.1016/S0308-8146(96)00311-1.

14. Montagnini, F.; Eibl, B.I.; Barth, S.R. Organic yerba mate: an environmentally, socially and financially suitable agroforestry system. Bois Et Forets Des Tropiques 2011, 308(308), 59–74.

15. Huck, C.W.; Guggenbichler, W.; Bonn, G.K. Analysis of caffeine, theobromine and theophylline in coffee by near infrared spectroscopy (NIRS) compared to high-performance liquid chromatography (HPLC) coupled to mass spectrometry. Analytica Chimica Acta 2005, 538(1), 195–203 https://doi.org/10.1016/j.aca.2005.01.064.

16. Roggo, Y.; Chalus, P.; Maurer, L.; Lema-Martinez, C.; Edmond, A.; Jent, N. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. Journal of Pharmaceutical and Biomedical Analysis 2007, 44(3), 683–700 https://doi.org/10.1016/j.jpba.2007.03.023.

17. Sena, M.M.; Poppi, R.J. N-way PLS applied to simultaneous spectrophotometric determination of acetylsalicylic acid, paracetamol and caffeine. Journal of pharmaceutical and biomedical analysis Elsevier: 2004, 34(1), 27–34.

18. Deladino, L.; Teixeira, A.S.; Reta, M.; García, A.D.M.; Navarro, A.S.; Martino, M.N. Major Phenolics in Yerba Mate Extracts (Ilex paraguariensis) and Their Contribution to the Total Antioxidant Capacity. Food and Nutrition Sciences Scientific Research Publishing: 2013, 04(08), 154–162 10.4236/fns.2013.48A019.

19. Anesini, C.; Turner, S.; Cogoi, L.; Filip, R. Study of the participation of caffeine and polyphenols on the overall antioxidant activity of mate (Ilex paraguariensis). LWT-Food Science and Technology Elsevier: 2012, 45(2), 299–304.

20. Chang, C.-W.; Laird, D.A.; Mausbach, M.J.; Hurburgh, C.R. Near-infrared reflectance

spectroscopy–principal components regression analyses of soil properties. Soil Science Society of America Journal Soil Science Society: 2001, 65(2), 480–490.

21. Forouzangohar, M.; Cozzolino, D.; Kookana, R.S.; Smernik, R.J.; Forrester, S.T.; Chittleborough, D.J. Direct comparison between visible near-and mid-infrared spectroscopy for describing diuron sorption in soils. Environmental science & technology ACS Publications: 2009, 43(11), 4049–4055.
22. Bellon-Maurel, V.; Fernandez-Ahumada, E.; Palagos, B.; Roger, J.-M.; McBratney, A. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. TrAC Trends in Analytical Chemistry Elsevier: 2010, 29(9), 1073–1081.
23. Camotti Bastos, M.; Cherobim, V.F.; Reissmann, C.B.; Fernandes Kaseker, J.; Gaiad, S. Yerba mate: Nutrient levels and quality of the beverage depending on the harvest season. Journal of Food Composition and Analysis: 2018, 69, 1–6
24. Brereton, R.G. Introduction to multivariate calibration in analytical chemistry. The Analyst: 2000, 125(11), 2125–2154.
25. Filip, R.; Lopez, P.; Coussio, J.; Ferraro, G. Mate substitutes or adulterants: study of xanthine content. Phytotherapy Research: 1998, 12(2), 129–131.

Tables

Table 1. Coefficient of determination (R^2) and standard error of calibration (SEC) from models based on both near infrared equipment used in this study and standard error of prediction (SEP) based on external validation, with samples that do not participate in model construction, and ratio of performance to deviation (RPD).

Equipment	Factors	Calibration		External validation		RDP
		R^2	SEC (mg/kg)	R^2	SEP (mg/kg)	SD[a]/SEP
bench	5	0.80	0.32	0.81	0.26	2.60
handheld	5	0.78	0.30	0.68	0.38	1.80

[a]SD* is the standard deviation of the whole sample set (=0.6769 mg/kg)

Figure Captions

Figure 1. Partial least square multivariate model correlation of (i) bench and (ii) handheld near-infrared spectroscopy of leaves from yerba mate trees with caffeine content measured through high performance liquid chromatography method; also, (iii) bench and (iv) handheld partial least square multivariate model evaluation based on external validation, calculated using samples that do not participate in the construction of multivariate model. Reference are values obtained by the hot-water extraction of dried leaves from high performance liquid chromatography. Predicted can be related to bench and portable near infrared spectroscopy. The bench are values of leaves dried in a microwave oven, grounded and sieved with no extraction of its components by hot-water. The handheld are values of fresh leaves obtained in the field.

Figure 2. Plot of regression coefficients from the prediction model constructed using partial least square regression with near infrared spectroscopy spectra using (i) bench and (ii) handheld equipment. The region with the highest values of regression coefficients corresponds to the main absorption band of pure caffeine

Figure 3. Near infrared spectra of pure caffeine (a); dried leaf of yerba-mate from bench equipment (b) and (c) live leaf of yerba-mate from portable equipment, all raw data, on samples containing the highest value of caffeine, obtained from Ivaí region, - 2.39% of caffeine. The line in the figure corresponds to the region range of (i) handheld near infrared spectra that were used for the construction of the multivariate model; in the case of (ii) bench spectra, the region exhibits the most statistically significant region used for the construction of the multivariate model, which correlates with the highest absorption of caffeine (a).

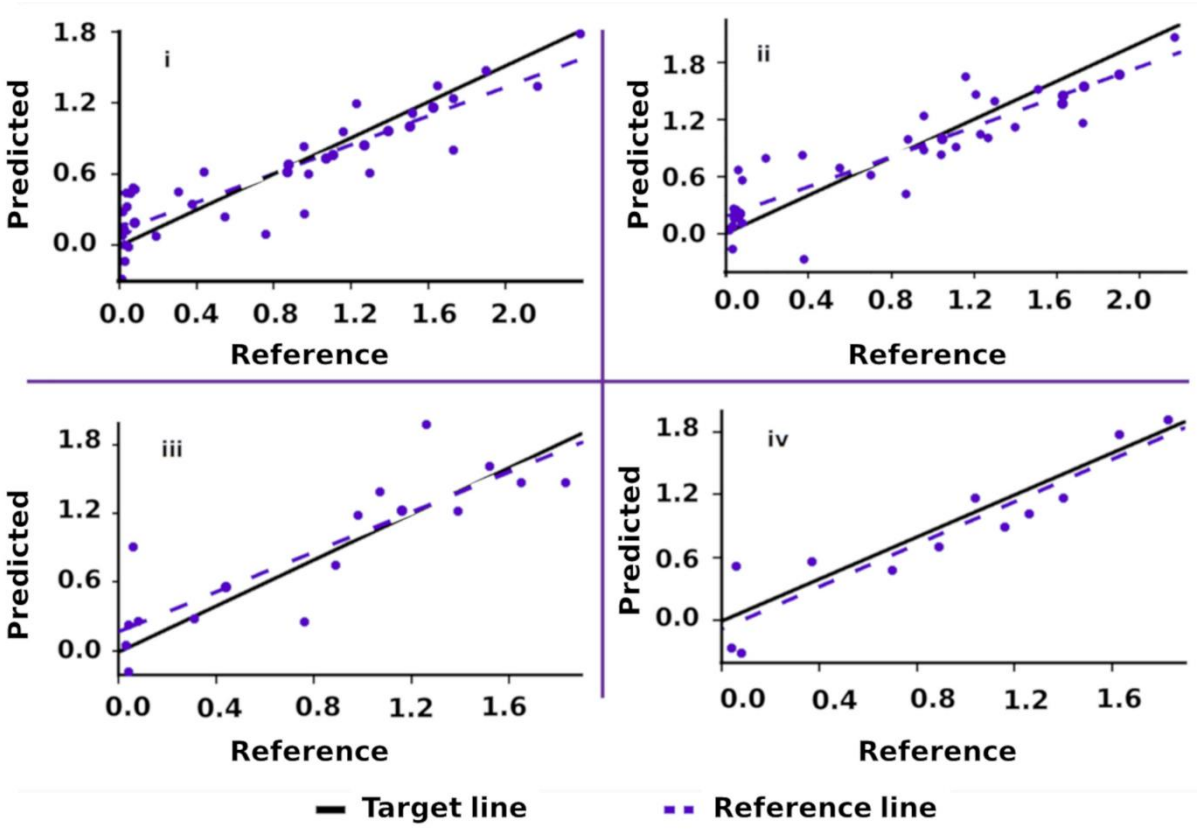
i
ii
iii
iv
Predicted
Reference
0.0
0.4
0.8
1.2
1.6
2.0
0.6
1.8
Target line
Reference line

Figure 1.

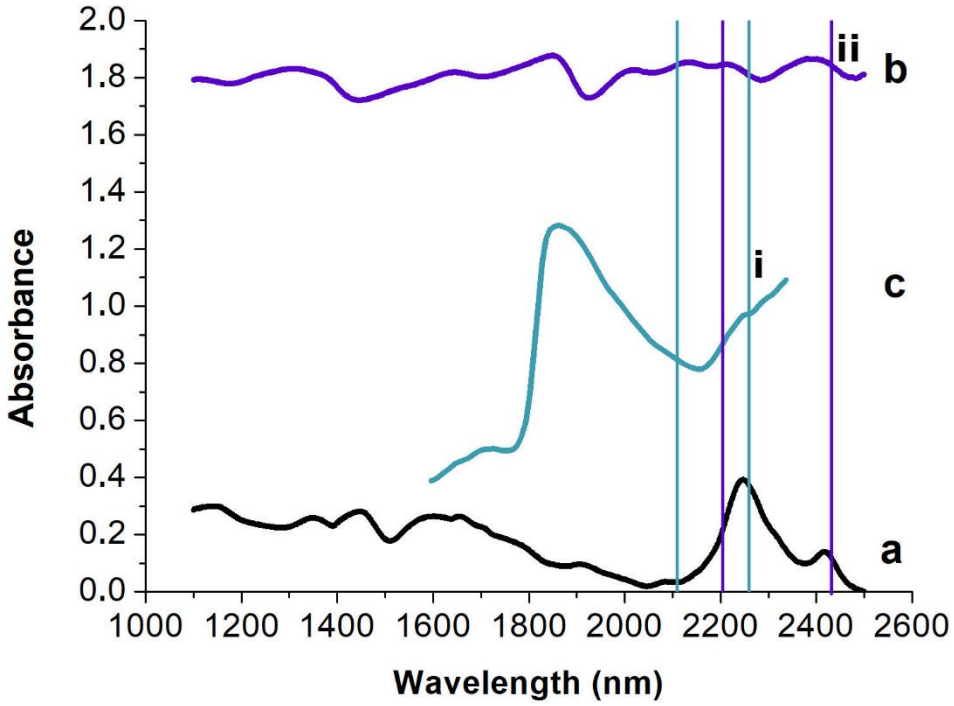
Absorbance
Wavelength (nm)
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
1000
1200
1400
1600
1800
2000
2200
2400
2600
ii
b
i
c
a

Figure 2.

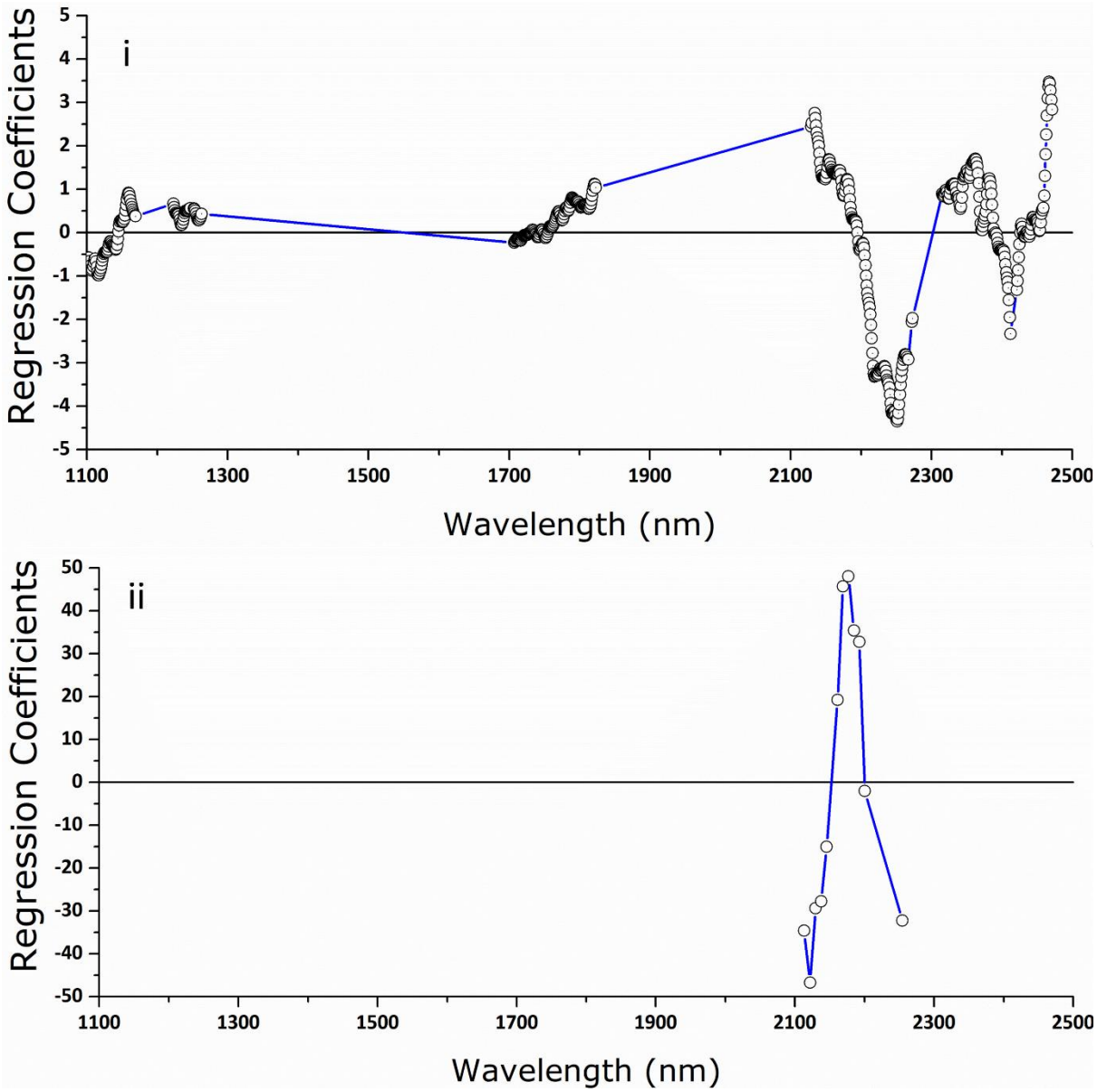
i
Regression Coefficients
5
4
3
2
1
0
-1
-2
-3
-4
-5
1100
1300
1500
1700
1900
2100
2300
2500
Wavelength (nm)
ii
Regression Coefficients
50
40
30
20
10
0
-10
-20
-30
-40
-50
1100
1300
1500
1700
1900
2100
2300
2500
Wavelength (nm)

Figure 3.

Table.A.1 - Coefficient of determination (R^2) and standard error of calibration (SEC) from models based on both near infrared equipment used in this study and standard error of prediction (SEP) based on external validation, with samples that do not participate in model construction, and ratio of performance to deviation (RPD). Pre-processing methods first and second derivative, multiple scatter correction (MSC), standard normal variate (SNV) and combination of both first and second derivative with MSC and SNV is shown. The NA means that the values were lower than acceptable values for PLS model.

Handheld NIR

Data Processing	Factors	Calibration		External validation		RDP
		R^2	SEC (mg/kg)	R^2	SEP (mg/kg)	
Original	5	0.78	0.30	0.68	0.38	1.8
1st derivative	5	0.70	0.36	NA	0.86	0.8
2nd derivative	1	0.21	0.59	NA	0.92	0.7
1st derivative : MSC	5	0.79	0.30	NA	NA	3.0E-4
2nd derivative : MSC	1	0.33	0.54	NA	0.73	0.9
1st derivative : SNV	5	0.79	0.30	NA	1.45	0.5
2nd derivative : SNV	1	0.33	0.54	NA	1.15	0.6

Bench NIR

Data Processing	Factors	Calibration		External validation		RDP
		R^2	SEC (mg/kg)	R^2	SEP (mg/kg)	
Original	5	0.8	0.32	0.81	0.26	2.6
1st derivative	1	0.43	0.53	NA	0.63	1.07
2nd derivative	1	0.26	0.61	NA	0.63	1.07
1st derivative : MSC	1	0.92	0.20	NA	NA	4.4E-4
2nd derivative : MSC	1	0.10	0.67	NA	0.62	1.09
1st derivative : SNV	3	0.92	0.20	NA	1.28	0.53
2nd derivative : SNV	1	0.26	0.61	NA	0.63	1.07

Graphical Abstract