

Analytical Methods

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ARTICLE

Direct determination of tannin in *Acacia mearnsii* bark using near infrared spectroscopy

Cite this: DOI: 10.1039/x0xx00000x

Caren Machado Menezes^a, Adilson Ben da Costa^a, Ricardo Rocha Renner^a, Lucas Ferreira Bastos^b, Marco Flôres Ferrão^{c,d} and Valderi Luiz Dressler^e

Received 00th March 2014,
Accepted 00th March 2014

DOI: 10.1039/x0xx00000x

www.rsc.org/

This study aimed to investigate the application of the near infrared spectroscopy (NIRS) and multivariate calibration methods for direct determination of tannin content in the *Acacia mearnsii* bark, in order to improve control during the tannin content extraction process. For this purpose, 89 bark samples were collected in the industrial plant of an extractive tannin industry. The spectra in the near infrared were acquired using a FT-NIR spectrometer with an integrating sphere, an Indium-Gallium-Arsenic detector in the range of 7,500 to 4,000 cm⁻¹, resolution of 16 cm⁻¹ and 23 scans divided in two different ways: a) *in natura* samples (no sample processing); and, b) dried and milled samples. The partial least squares models (PLS) were developed and different strategies were investigated during the infrared spectra preprocessing. The results of the prediction were compared to the ones obtained through the reference methodology (NBR 11131), showing values for the root mean square error of prediction (RMSEP) between 2.11 and 2.42% for the dried and milled bark samples, and 2.31 and 2.54% for the *in natura* samples. These results show that NIRS combined with multivariate calibration methods may be used for direct determination of tannin content in *Acacia mearnsii* bark. Low need of sample preparation, short analysis time, no reagent consumption and, consequently, no waste generation are the main characteristics of the proposed method.

Introduction

Native from Australia, the *Acacia mearnsii* De Wild is a medium sized tree which comes from the *Fabaceae* (*Leguminosae*) family. Due to its high productivity and fast adaptation to different environmental conditions, the plant was introduced in other regions worldwide.¹ The major commercial plantations - aimed primarily at the production of wood, coal and tannins - are located in Eastern and Southern Africa, India and Southern Brazil.^{1,2} The bark from *Acacia mearnsii* is the most commonly used raw material in the extraction of tannin substances worldwide.³ They are used for manufacturing adhesives, flocculants and, also, for skin tanning.^{3,4} The tannin contents extracted from *Acacia mearnsii* are condensed tannins and they consist of high molecular weight oligomers or polymers, built by flavonoids (flavan-3-ol, catequin), which oxidative condensation takes place between carbons C-4 and C-6 or C-8. When in low degree of polymerization tannins are soluble in polar solvents, while in a high degree they are soluble in alkaline solutions.^{5,6} The concentration of tannins in the *Acacia mearnsii* bark may reach values above 45% (w/w), however this output may vary significantly according to the weather and pedological conditions of the growing

site, plant morphology and the cultivation techniques used.^{1,3} In Brazil, some studies show outputs of approximately 20 to 28%.^{7,8} Nowadays, the commercialization of *Acacia mearnsii* is based only in this material mass and not in the concentration of tannins. This is justified by the industrial sector, because of the slowness of the method used in the determination of the amount of tannin content accepted by the standard NBR 11131.⁹ This procedure may take up to 20 hours for each determination (with a maximum of 6 determinations in each shift), making it difficult to apply for quality control routine procedures, especially when a large amount of samples must be analyzed. In this regard, different studies have been conducted in order to develop alternative analytical methodologies. Lopes et al.¹⁰ applied a reversed phase-high-performance liquid chromatography (RP-HPLC) to determine the amount of flavan-3-ols present in semi-purified bark extract from three different species of *Stryphnodendron*. Prasad¹¹ applied near-infrared spectroscopy (NIRS) and a multiple linear regression in order to determine the total tannin and phenol content in dried and milled bark sample of *Leucaena mimosine* and *Acacia* sp. Using reflectance values measured in only four wavelengths (2152, 2178, 2183 e 2295 nm), the results obtained for the tannin content in *Acacia* sp. bark samples

showed a deviation of 1.12% and a correlation coefficient of 0.9632. Derkyi et al.¹² also used NIRS and multivariate calibration methods in the analysis of dried and milled *Pinus caribaea* bark samples. The authors obtained correlation coefficient of 0.95 and 0.98, and root mean square error of cross-validation values (RMSECV) of 0.72 and 0.23%, respectively, during the determination of extractable substances and polyphenols.

Nevertheless, neither of these studies resulted in a practical methodology to be used in the industrial production process, or even in the reforestation industry sector. Due to these circumstances, this study aimed to investigate NIRS applications and multivariate calibration methods for direct determination of tannin content in the *Acacia mearnsii* bark, in order to improve control during tannin extraction process.

The NIRS is a valuable tool for identification of organic and inorganic compounds^{13,14}, showing a great applicability in quantitative determinations of compounds presenting functional groups with hydrogen linked to carbon, nitrogen and oxygen. However, the great variety of chemical compounds presents in a sample result in the overlapping and disturbance of near-infrared spectrum absorption bands. Thereby, rarely there are spectrum bands clean enough to allow a simple correlation (univariate) with the concentration of analyte and, that is why, multivariate calibration methods are frequently used.^{14,15}

The spectroscopy has stood out as a powerful analytical tool, mainly, when associated to chemometrics. This tool is highly used in development of analytical methodologies for quantitative and qualitative evaluation of different types of material, like petroleum and its derivatives^{13,16,17}, vegetable oils and biofuels¹⁸⁻²², wood and cellulose^{12,23-26}.

2. Materials and methods

2.1 Sample collection and processing

The samples of *Acacia mearnsii* bark were collected at the raw materials department of the Seta S/A Extrativa de Tanino de Acácia company (www.setaonline.com/) located in Estância Velha, RS, Brazil. The collections were performed in different days, during a period of 8 months, gathering 89 samples. These samples were identified from Am1 to Am89, keeping the order they were collected.

About 250 g of each sample were taken, in order to acquire the infrared spectra and determine the tannin content using the reference methodology.⁹

In order to acquire the NIR spectra, the samples were treated in two different ways: *in natura* samples and dried and milled samples. The first set of spectra was acquired in the *in natura* samples (original samples, no treatments applied). For this purpose, the samples were divided in 3 subsamples and identified as Am1-1, Am1-2 Am1-3 up to Am89-1, Am89-2, Am89-3 and NIR spectra were acquired direct in internal face to the bark, in triplicate, placed on a sapphire windows located on the top of the equipment. Thus, for the tests with *in natura* samples, it was used 267 subsamples of 15 cm² area each were used.

The second set was built from previously dried bark. The samples were oven dried at 105 °C, until they reached a constant weight. After then, they were milled using a mill (Quimis, model Q298A),

generating 89 samples (with no subsamples). The samples were milled in order to obtain particle size average lower than 50 µm. Particle size were determined by using a Mastersizer 2000 (Micro Particle Analyzer, Malvern Instruments). Therefore the infrared spectra acquisition, these samples were transferred to quartz cuvettes.

2.2 Reference methods for tannin determination

The standard method (NBR 11131⁹) was used as reference method to determine the tannin content in *Acacia mearnsii* bark samples. This method may be divided in three main steps. The first step was the extraction of the tannin and non-tannin contents from the bark samples. On the next step, the tannin and non-tannin contents were separated by filtration in skin powder and, finally, quantified by gravimetry. This procedure was performed in duplicate and the final result matched the arithmetic average.

2.3 Infrared spectra acquisition

The spectra were obtained using a FT-NIR spectrometer with an integrating sphere (PerkinElmer, model Spectrum 400) and an Indium-Gallium-Arsenic (InGaAs) detector set in the range of 7,500 to 4,000 cm⁻¹, resolution of 16 cm⁻¹ and 32 scans. The determinations were performed using a NIRA accessory.

2.4. Data modeling

The infrared spectra were registered in *Comma Separated Values* (CSV) and later moved to a spreadsheet created using Microsoft Excel®, corresponding to the data matrix X. A spreadsheet built on Microsoft Excel® was also used to register the average value of tannin contents, corresponding to the data vector Y. The data were modeled in the software Solo 6.5.3 (Eigenvector Research, Inc.), an environment in which all multivariate calibration models by Partial least squares (PLS), Interval partial least squares (iPLS) and Synergy interval partial least squares (siPLS) were developed.

For the PLS models, the entire range of the infrared spectra, between 7,500 to 4,000 cm⁻¹, was used. The iPLS and siPLS were built from the best preprocessing technique selected in the PLS models and dividing the infrared spectrum in 2, 3, 4, 8, 16 and 32 intervals. The siPLS models were developed in the software's auto mode, expressing the best combination of ranges (with maximum of 32 ranges) to create calibration models.²⁷

The algorithm Kennard-Stone²⁸ was used in the Matlab version 7.0 (Mathworks Inc.) to select the calibration and prediction sets. This step was conducted in such a way that 2/3 of the samples were bound to the calibration set and 1/3 of them to prediction set.

For the identification and exclusion of outliers in the calibration set, it was used a waste chart of Students versus Leverage, in which were considered as anomalous those samples showing, simultaneously, values for h of Leverage higher than h critical and waste of Students higher than 2.5%.²⁹

Aiming to eliminate the non relevant information about the spectra and make the data matrix well conditioned to the analysis, different preprocessing techniques were used) with the spectra (normalization, multiplicative scattering correction, mean center, Savitsky-Golay first derivative, Savitsky-Golay second derivative and autoscale) and evaluated according to the RMSECV results. In the same way, the

number of latent variables (LVs) used in the models were determined based on the least value obtained for RMSECV.³⁰⁻³³ The results obtained through the developed methodology in this research were compared to the results obtained by reference method using the Wilcoxon nonparametric test, using the Paleontological Statistics (PAST) software, version 1.91.³⁴

3. Results and discussions

Using the reference methodology⁹, the tannin content of the 89 samples of *Acacia mearnsii* bark ranged between 10.0±1.2% and 31.1±2.1% (w/w, based on humid weight), as seen on Table 1, showing a normal distribution ($P>0.05$) compared with the average (17.4%). The standard error³⁵ presented by the reference method was 1.2 %.

Table 1. Result (Average ± standard deviation for $n = 2$) of tannin concentration (% w/w) in 89 samples of *Acacia mearnsii* bark, obtained with the reference methodology.

Sample	Average ± SD	Sample	Average ± SD	Sample	Average ± SD	Sample	Average ± SD
Am1	15.4±0.8	Am24	11.3±0.3	Am47	15.0±1.4	Am70	13.9±1.6
Am2	23.2±0.9	Am25	15.6±0.2	Am48	16.4±1.5	Am71	17.8±2.6
Am3	20.2±0.7	Am26	12.8±0.9	Am49	13.7±1.6	Am72	16.7±0.2
Am4	20.5±0.5	Am27	12.0±0.4	Am50	20.6±0.1	Am73	18.9±0.9
Am5	13.7±0.6	Am28	18.2±1.2	Am51	21.2±0.1	Am74	20.8±1.4
Am6	15.8±0.4	Am29	20.8±0.5	Am52	15.2±1.3	Am75	18.1±0.2
Am7	31.1±2.1	Am30	26.1±0.1	Am53	17.3±0.2	Am76	16.6±0.1
Am8	21.8±0.5	Am31	16.4±0.2	Am54	21.3±0.6	Am77	12.6±2.3
Am9	13.8±0.7	Am32	19.1±1.2	Am55	14.6±1.4	Am78	19.7±1.3
Am10	18.0±1.2	Am33	22.1±0.4	Am56	24.2±0.1	Am79	22.1±1.8
Am11	17.8±0.4	Am34	10.0±1.2	Am57	16.8±1.5	Am80	21.1±1.8
Am12	12.8±1.2	Am35	20.7±2.4	Am58	23.6±0.6	Am81	21.2±0.9
Am13	16.2±0.5	Am36	19.9±0.6	Am59	21.3±2.7	Am82	23.0±1.4
Am14	13.9±0.6	Am37	11.7±1.1	Am60	14.9±1.0	Am83	15.2±1.1
Am15	17.2±0.9	Am38	19.5±1.6	Am61	11.2±1.0	Am84	19.5±0.5
Am16	13.0±1.3	Am39	17.3±0.6	Am62	12.6±0.7	Am85	12.9±2.0
Am17	12.8±1.1	Am40	17.6±1.3	Am63	20.9±0.1	Am86	22.3±4.4
Am18	11.8±0.8	Am41	10.6±0.6	Am64	13.2±1.5	Am87	17.0±0.7
Am19	18.0±1.4	Am42	16.0±0.1	Am65	13.2±0.1	Am88	16.3±1.9
Am20	18.9±0.9	Am43	16.2±0.8	Am66	18.7±1.6	Am89	15.3±0.9
Am21	14.3±0.9	Am44	13.4±0.8	Am67	26.1±2.0		
Am22	19.2±0.2	Am45	14.8±0.2	Am68	15.0±2.9		
Am23	15.3±0.5	Am46	19.3±1.5	Am69	22.7±2.0		

Figure 1 presents the profile of the NIR spectra (between 7,500 and 4,000 cm^{-1}) obtained from the *in natura* and dried and milled bark samples.

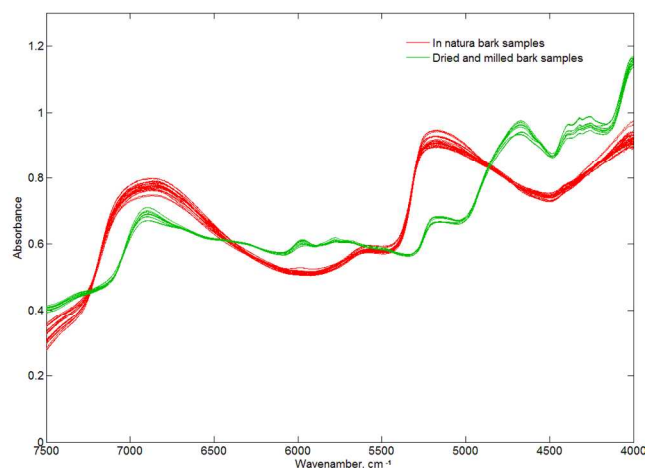


Figure 1. Profile of the NIR spectra obtained from the *Acacia mearnsii* bark samples.

The high intensity of the absorption of the O-H bonds of water is evident in the regions between 6,250 – 7,250 cm^{-1} and 5,400 – 4,750 cm^{-1} .²² However, according to Bakeev³⁶, these regions are important to characterize alcohols and organic acids, which contain the functional group O-H. The O-H group is the second main group (after the C-H) in the NIR spectrum. According to Holler et al.¹⁵, the band at 7,100 cm^{-1} is regularly used in the phenols qualification process. In the dried and milled bark samples, at 6,000 cm^{-1} , it is possible to see a discrete absorption band due to the first overtone of C-H stretching vibration of aromatic hydrocarbons.¹⁴

The bands between 4,000 – 5,000 cm^{-1} result from combinations of fundamental vibrations in the fingerprint region of the Mid infrared (MIR) with C-H stretching vibrations.¹⁴ For phenols, bands combination of C-H stretching vibration and deformation of C-H occur in 4,648, 4,550, 4,300 and 4,046 cm^{-1} .³⁷ Furthermore, the absorption bands of O-H vibration modes – normally less polar than water connections, as occurs in tannin molecules – may be present between 4,000-5,000 cm^{-1} .³⁸ According to Weyer³⁸, this region has an informative character and low noise level, and Derkyi et al.¹² and Prasad¹¹ used this region to determine the tannin content in dried samples.

In this context, the absorption bands at 4,650 cm^{-1} , identified in the spectra of dried and milled bark samples and tannin samples, can be attributed to the functional group OH of the tannin molecule and/or the combination of C-H stretching vibrations and C-H deformation. These bands appear to be overlap by the absorption bands of the O-H bonds in samples presenting higher moisture levels.

Table 2 shows the results obtained in the models developed using the *in natura* bark samples of *Acacia mearnsii*. The calibration models using PLS indicated that the spectra preprocessing by normalization (inf-Norm, maximum = 1) followed by standard normal variate (SNV), 1st derivative (1D) and autoscale (Auto) show minor cross-validation errors (RMSECV = 2.39%), after the exclusion of 20 calibration and 24 prediction samples, which corresponded to 11 and 28% of the total calibration and prediction samples.

It is observed that the calibration model by iPLS, in which the spectrum is divided into three sub-regions (iPLS3) using 5 latent variables, selected the region between 6,334 – 5,170 cm^{-1} as the smallest cross-validation error (RMSECV = 2.59%), shows a correlation coefficient (r (CV)) of 0.635 (Figure 2-3). The best result of RMSECV pointed to 5 latent variables, which were used maintaining the balance with the root mean square error values of prediction (RMSEP).

Applying the best calibration for siPLS, the best model obtained was siPLS32auto, which employed a combination of regions between 7,392 – 7,286, 6,528 – 6,314, 6,096 – 5,990 and 4,476 – 4,370 cm^{-1} , resulting in a lower error rate (RMSECV = 2.31%) with correlation coefficient of 0.723 (Figures 4-5). The same happened for the iPLS model, the choice of the number of latent variables was based on the correlation between the prediction and validation errors and, therefore, 6 latent variables were used.

Table 2. Results of calibration models for tannin determination by PLS, iPLS and siPLS in *in natura* samples from *Acacia mearnsii* bark.

Model	Preprocessing	LV	RMSECV	RMSEP	r(CV)	r (P)
PLS	Normal, MSC, 1D, MC (x), Auto (y)	10	2.65	2.61	0.678	0.616
PLS	Normal, MSC, 2D, MC (x), Auto (y)	4	3.22	2.85	0.454	0.458
PLS	Normal, MSC, 1D, Auto (x), Auto (y)	5	2.58	2.86	0.657	0.592
PLS	Normal, MSC, 2D, Auto (x), Auto (y)	2	2.69	2.66	0.627	0.566
PLS	Normal, SNV, 1D, MC (x), Auto (y)	7	2.73	2.75	0.608	0.538
PLS	Normal, SNV, 2D, MC (x), Auto (y)	4	3.22	2.85	0.452	0.469
PLS	Normal, SNV, 1D, Auto (x), Auto (y)	5	2.39	2.53	0.698	0.632
PLS	Normal, SNV, 2D, Auto (x), Auto (y)	2	2.72	2.66	0.613	0.547
iPLS and siPLS models (Preprocessing: Normal, SNV, 1D, Auto (x), Auto (y))						
Model	IR selected regions, cm ⁻¹	LV	RMSECV	RMSEP	r(CV)	r (P)
iPLS2	5,750-4,002	5	2.70	2.730	0.590	0.557
iPLS3	6,334-5,170	5	2.59	2.54	0.635	0.640
iPLS4	6,626-5,754	4	2.83	2.84	0.547	0.510
iPLS8	6,192-5,758	5	2.79	2.80	0.564	0.538
iPLS16	5,974-5,758	4	2.75	3.14	0.574	0.360
iPLS32	6,096-5,990	6	2.79	2.98	0.554	0.447
siPLS8auto	7,500-7,066, 6,628-5,758, 5,320-4,450	5	2.54	2.59	0.655	0.616
siPLS16auto	7,500-7,284, 6,192-5,758, 5,538-5,104, 6,666-4,450	4	2.54	2.51	0.655	0.656
siPLS32auto	7,392-7,286, 6,420-6,314, 6,096-5,990, 4,476-4,370	6	2.31	2.23	0.723	0.608

Note: Normal: Normalization (Norm-inf, max = 1); MSC: Multiplicative scattering correction; MC: Mean center; 1D: 1st derivative (Savitsky-Golay); 2D: 2nd derivative (Savitsky-Golay); Auto: Autocalse; (x): x matrix (NIR); (y): y matrix (tannin concentration); LV: Number of latent variables; RMSECV: Root mean square error of cross validation; RMSEP: Root mean square error of prediction; r(CV): Cross validated correlation coefficient; r(P): Predicted correlation coefficient. The highlighted lines comprehend the best calibration model.

Analyzing the regions targeted in the models, it is possible to observe that between $6,334 - 5,170 \text{ cm}^{-1}$, selected in regression for iPLS can be from C-H stretching vibration that, according to Burns and Ciurczak¹⁴, refers to the aromatic hydrocarbons, which are present in the condensed tannin. Thus, the region selected by the regression model for siPLS, between $7,392 - 7,286 \text{ cm}^{-1}$, can be associated to the presence of phenols which, according to Holler et al.¹⁵ are identified at the band of $7,100 \text{ cm}^{-1}$ in quantitative analysis. While the region between $6,096 - 5,990 \text{ cm}^{-1}$ can be associated to C-H stretching vibrations regarding the aromatic hydrocarbons. Finally, the selection of the region between $4,476 - 4,370 \text{ cm}^{-1}$ in siPLS can be attributed to the OH functional group in the tannin molecule and/or the combination of C-H stretching vibrations and C-H deformation. In the original spectrum, these bands appear to be overlapped by the absorption bands from the O-H bonds in samples presenting higher moisture levels.

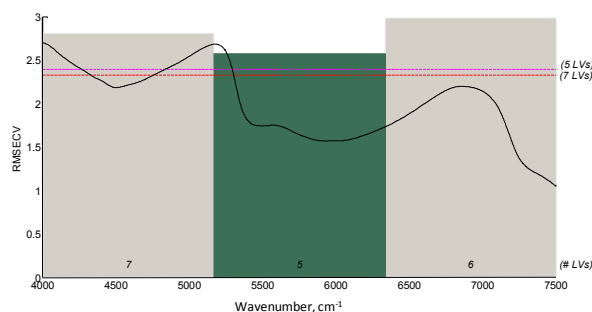


Figure 2. Result from the RMSECV for ranges (columns) and for the global model (dotted red lines), obtained by the iPLS3 model. The numbers inside the columns identify the number of latent variables (LVs) indicated in each sub-region.

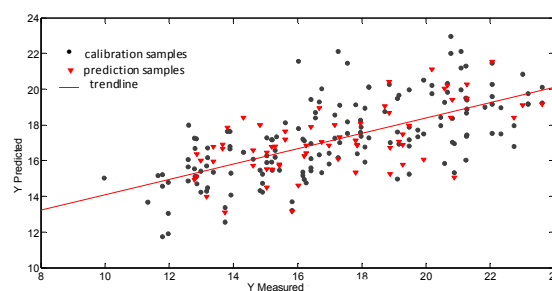


Figure 3. Correlation curve between the measured values (Y Measured) of tannins (% w/w) by the reference method and the predicted results (Y Predicted) by the iPLS3 model.

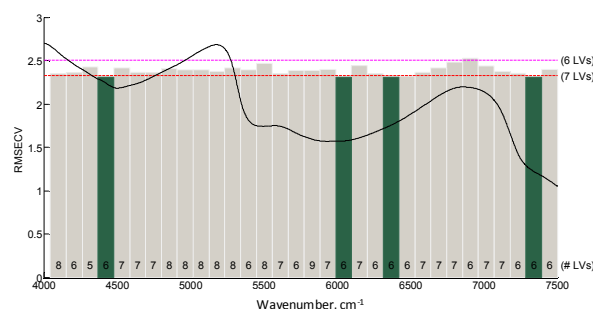


Figure 4. Result from the RMSECV for ranges (columns) and for the global model (dotted red lines), obtained by the siPLS32auto model. The numbers inside the columns refer to the number of latent variables (LVs) suggested in each sub-region.

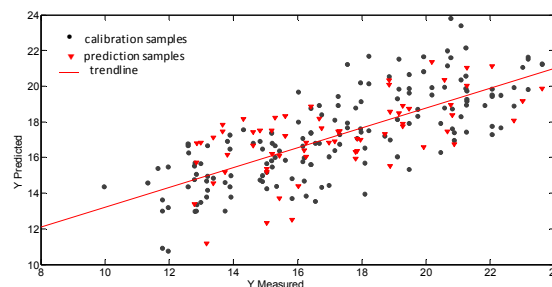


Figure 5. Correlation curve between the measured values (Y Measured) for tannins (% , w/w) by the reference methodology and the predicted results (Y Predicted) by the siPLS32auto model.

In Table 3 are shown the results obtained in the models developed for the determination of tannin content in dried and milled bark sample of *Acacia mearnsii*. The calibration models using PLS indicated that the preprocessing of the spectra through normalization (inf-Norm, maximum = 1) followed by MSC, first derivative (Savitsky-Golay) and autoscale, presented the smallest validation errors (RMSECV = 2.04%), after excluding 9 outliers samples identified in the calibration set (Am2, Am7, Am22, Am37, Am61, Am70, Am79, and Am72 Am89) and one in the prediction set (Am41). In this way, this preprocessing strategy was applied in the studies of iPLS and siPLS.

The calibration model for iPLS in which the spectra were divided into 4 sub-regions (iPLS4), selected the region between 6,626 – 5,754 cm^{-1} as the smallest cross-validation error (RMSECV =

2.04%), showing a correlation coefficient (r (CV)) of 0.848 (Figures 6-7). The calibration model by synergism of intervals selected the model siPLS16auto, using the combination between the regions 6,846 – 6,630, 5,974 – 5,758, 5,538 – 5,322 and 4,230 – 4,014 cm^{-1} , such as of smallest cross-validation error (RMSECV = 1.79%) showing a correlation coefficient (r (CV)) of 0.786 (Figures 8-9).

Table 3. Results of calibration models to tannin determination by PLS, iPLS and siPLS in dried and milled samples from *Acacia mearnsii* bark.

Model	Preprocessing	LV	RMSECV	RMSEP	r (CV)	r (P)
PLS	Normal, MSC, 1D, MC (x), Auto (y)	2	2.74	2.38	0.697	0.656
PLS	Normal, MSC, 2D, MC (x), Auto (y)	2	2.59	2.45	0.737	0.624
PLS	Normal, MSC, 1D, Auto (x), Auto (y)	2	2.04	2.26	0.849	0.686
PLS	Normal, MSC, 2D, Auto (x), Auto (y)	2	2.81	2.17	0.750	0.721
PLS	Normal, SNV, 1D, MC (x), Auto (y)	2	2.56	2.26	0.732	0.714
PLS	Normal, SNV, 2D, MC (x), Auto (y)	3	2.84	2.55	0.685	0.624
PLS	Normal, SNV, 1D, Auto (x), Auto (y)	2	2.45	2.36	0.772	0.656
PLS	Normal, SNV, 2D, Auto (x), Auto (y)	2	3.20	2.28	0.678	0.678
iPLS and siPLS models (Preprocessing: Normal, MSC, 1D, Auto (x), Auto (y))						
Model	IR selected regions, cm^{-1}	LV	RMSECV	RMSEP	r (CV)	r (P)
iPLS2	5,750-4,002	2	2.19	2.27	0.823	0.671
iPLS3	6,334-5,170	2	2.21	2.29	0.820	0.678
iPLS4	6,626-5,754	2	2.04	2.11	0.848	0.728
iPLS8	5,320-4,886	3	2.34	2.81	0.800	0.490
iPLS16	5,974-5,758	3	2.51	2.26	0.776	0.685
iPLS32	6,744-6,638	3	2.17	3.96	0.827	0.264
iPLS8	6,192-4,886	2	2.14	2.32	0.832	0.663
iPLS8auto	5,320-4,886	3	2.34	2.81	0.800	0.490
siPLS16auto	6,846-6,630, 5,974-5,758, 5,538-5,322, 4,230-4,014	4	1.79	2.42	0.786	0.624
siPLS32auto	6,846-6,630, 5,974-5,758, 5,538-5,322, 4,230-4,014	3	1.98	2.50	0.858	0.600

Note: Normal: Normalization (Norm-inf, max = 1); MSC: Multiplicative scattering correction; MC: Mean center; 1D: First derivative (Savitsky-Golay); 2D: Second derivative (Savitsky-Golay); Auto: Autoscale; (x): x matrix (NIR); (y): y matrix (tannin concentration); LV: Number of latent variables; RMSECV: Root mean square error of cross validation; RMSEP: Root mean square error of prediction; r (CV): Cross validated correlation coefficient; r (P): Predicted correlation coefficient. The highlighted lines comprehend the best calibration model.

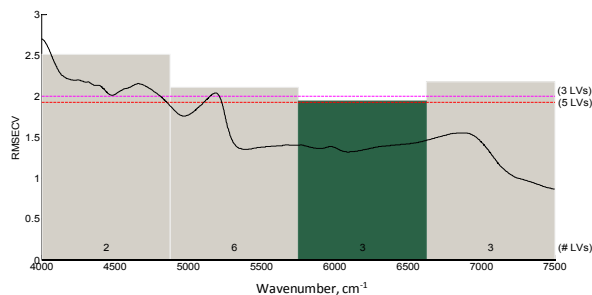


Figure 6. Results of RMSECV for the ranges (columns) and global model (dotted red lines), obtained by the iPLS4 model for the determination of tannin content in *Acacia mearnsii* dried and milled bark samples. The numbers inside the columns refer to the number of latent variables (LVs) suggested in each sub-region.

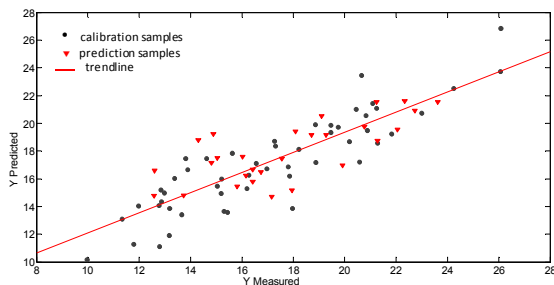


Figure 7. Correlation curve between the measured values (Y Measured) of tannin content (% w/w) by the reference method and the predicted results (Y Predicted) by the iPLS4 model.

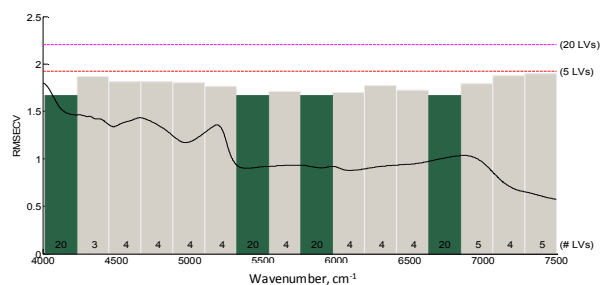


Figure 8. Results of RMSECV for the ranges (columns) and global model (dotted red lines), obtained by the siPLS16auto model for the determination of tannin content in *Acacia mearnsii* dried and milled bark samples. The numbers inside the columns refer to the number of latent variables (LVs) suggested in each sub-region.

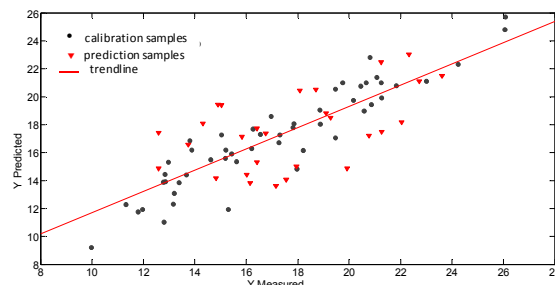


Figure 9. Correlation curve between the measured values (Y Measured) of tannin content (%) by the reference methodology and the predicted results (Y Predicted) by the siPLS16auto model, preprocessed with normalization, MSC, 1st Derivative and autoscaling.

In Table 4 are shown the most important intervals chosen by the models, represented by iPLS. It also relates the most relevant combinations, represented by siPLS, for the determination of tannin content.

Despite the cross-validation errors being higher than those obtained by other authors, as Derkyi et al.¹² and Prasad¹¹ in similar samples, no significant differences were identified ($P>0.05$) in the set of results produced by the reference methods and the ones showed by the methodology developed in this study (Table 5).

It should be noted that different from work developed by Derkyi et al.¹² and Prasad¹¹, in this study the samples were collected during a period of 8 months, directly from the industrial process. In this way, important variables such as the origin of the samples, storage time after cutting the trees (time between forest extraction and the beginning of the process), bark age, transportation and storage conditions are unknown. Despite these variables, the results reflect the application of NIRs for the determination of tannins content in the bark of *Acacia mearnsii* trees in real industrial process operation conditions.

The errors presented by PLS and siPLS models (Table 4) were upper to the standard error presented by reference method (1.2%), however, are secure enough for applied at line monitoring in industrial process and in the raw materials buying process, with the screening methods.

Table 4. Ranges selected by iPLS and siPLS and respective cross-validation errors (RMSECV, %) and prediction (RMSEP, %) for tannin determination.

Type of sample	iPLS			siPLS		
	Ranges	RMSECV	RMSEP	Range	RMSECV	RMSEP
<i>In natura</i>	6334-5170	2.59	2.54	7392-7286, 6420-6314, 6096-5990, 4476-4370	2.31	2.23
Dried and milled	6626-5754	2.04	2.11	6846-6630, 5974-5758, 5538-5322, 4230-4014	1.79	2.42

4. Final considerations

The results presented in this study prove that the near-infrared spectroscopy combined with multivariate calibration methods can be applied for the direct determination of tannin content in *Acacia mearnsii* bark.

The selected regions for the determination of tannin content in bark of *Acacia mearnsii* correspond, in part, between 4,000 – 5,000 cm⁻¹, which was also used by other authors^{11,12} to predict the concentration of this compound. However, regions of higher wavenumber also seem important for tannin quantification.

In addition, it is possible to notice that during the determination of tannin content, a minor validation error is obtained after drying and milling the samples. Possibly, this result is related to the high homogeneity of the sample and the minimization of moisture interference by overlapping the absorption bands of the O-H bonds, from water absorption at 4,650 cm⁻¹, which can be attributed to the OH functional group of tannin molecule and/or the combination of C-H stretching vibrations and C-H deformation.

At last, the methodology proposed in this work presents an analytical frequency of 10 determinations per hour (including sample preparation procedures), which is significantly higher than the reference methodology (20 hours for each determination). Furthermore, there is no need to use chemical reagents and, consequently, there is no waste generation. In this way, these advantages serve as a stimulus to the implementation of this methodology in tannin extractive industry quality control, especially to assist the manufacturing process on the segregation of raw materials in stock, in the valuation of raw materials in the buying process, and in the efficiency control of the extraction processes. Activities which are currently unviable due to time consuming of the reference method available.

Acknowledgements

The authors would like to thank SETA S/A Company, to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and to Fundo de Apoio a Pesquisa da UNISC (FAP-UNISC) for supporting this study.

Notes and references

^a Universidade de Santa Cruz do Sul, Programa de Pós-Graduação em Sistemas e Processos Industriais, 96815-900, Santa Cruz do Sul–RS, Brazil.

^b Universidade de Santa Cruz do Sul, Curso de Química Industrial, 96815-900, Santa Cruz do Sul–RS, Brazil.

^c Universidade Federal do Rio Grande do Sul, Departamento de Química Inorgânica 91501-970, Porto Alegre, RS, Brazil.

^d Instituto Nacional de Ciência e Tecnologia de Bioanalítica, 13083-970, Campinas, SP, Brazil.

^e Universidade Federal de Santa Maria, Departamento de Química, 97105-900, Santa Maria, RS, Brazil.

- J. A. Duke, Handbook of legumes of world economic importance. Plenum Press, New York, 1981, pp. 345.
- R. W. Dunlop, in *Acacia mearnsii* De Wild. ; ed, P. C. M. Jansen, and D. Cardon, . vol 3: Dyes and tannins/Colorant. [CD-Rom]. PROTA, Wageningen, Netherlands, 2005, pp. 216.
- A. G. Brown, T. H. Booth, D. J. Boland, P. N. Martensz, L. A. J. Thomson, N. Hall, ed. J. C. Doran and J. W. Turnbull. Australian trees and shrubs: species for land rehabilitation and farm planting in the tropics. Australian Centre for International Agricultural Research (ACIAR), 1997, pp. 384.
- S. P. Sherry, The black wattle (*Acacia mearnsii* de Wild.). University of Natal Press. Pietermaritzburg, 1971, pp. 402.
- L. Bravo, Nutrition reviews, 1998, vol. 56, n. 11, pp. 317-333.
- R. M. Rowell (Ed.), Handbook of wood chemistry and wood composites. CRC press, New York, 2nd edn, 2012, pp. 687.
- A. F. Santos, Embrapa Florestas. Circular técnica, 2001, vol. 44.
- M. R. Pansera, G. A. Iob, A. C. Atti-Santos, M. Rossato, L. Atti-Serafini, and E. Cassel, Brazilian Archives of Biology and Technology, 2004, vol. 47, n. 6, pp. 995-998.
- BRASIL. Associação Brasileira de Normas Técnicas (ABNT). NBR 11131 - Insumos - Tanantes - Determinação do teor de constituintes não-tanantes e tanantes, 2008.
- G. C. Lopes, A. C. C. Sanches, C. E. M. D. Toledo, A. C. Isler and J. C. D. Mello, Brazilian Journal of Pharmaceutical Sciences, 2009, vol. 45, n. 1, pp. 135-143.
- M. N. V. Prasad, Biomass and Bioenergy, 1995, vol. 8, n. 3, pp. 203-205.
- S. A. Derkyi, B. Adu-Amankwa, D. Sekyere and A. Darkwa, Journal of Emerging Trends in Engineering and Applied Sciences (JETEAS), 2011, vol. 2, n. 1, pp. 132-136.
- C. Pasquini and A. F. Bueno, Fuel, 2007, vol. 86, n. 12, pp. 1927-1934.
- D. A. Burns and E. W. Ciurczak. Handbook of near-infrared analysis. CRC press, New York, 3rd edn, 2007, vol 35, pp 808 .
- F. J. Holler and D. A. Skoog, S. R. Crouch. Princípios de análise instrumental. Bookman, 2009.
- G. Parisotto, M. F. Ferrão, A. L. Müller, E. I. Müller, M. F. Santos, R. C. Guimarães, J. C. M. Dias and É. M. Flores, Energy & Fuels, 2010, vol. 24, n. 10, pp. 5474-5478.
- R. M. Balabin, R. Z. Safieva and E. I. Lomakina, Microchemical Journal, 2011, vol. 98, n. 1, pp. 121-128.
- N. Sinelli, L. Cerretani, V. D. Egidio, A. Bendini and E. Casiraghi, Food research international, 2010, vol. 43, n. 1, pp. 369-375.
- M. H. Killner, J. J. Rohwedder and C. Pasquini, Fuel, 2011, vol. 90, n. 11, pp. 3268-3273.
- R. M. Balabin and R. Z. Safieva, Analytica chimica acta, 2011, vol. 689, n. 2, pp. 190-197.
- D. Mueller, M. F. Ferrão, L. Marder, A. B. Costa, and R. C. S. Schneider, Sensors, 2013, vol. 13, n. 4, pp. 4258-4271.
- C. F. C. Ruschel, C. T. Huang, D. Samios and M. F. Ferrão, Química Nova, 2014, vol. 37, n. 5, pp. 810-815.
- C. Pasquini, C. J. G. Carneiro and P. M. Bomfim, O Papel, 2007, vol. 68, pp. 71-83.
- T. M. Venas, A. Rinnan, Chemometrics and Intelligent Laboratory Systems, 2008, vol. 92, n. 2, pp. 125-130.
- P. R. G. Hein, J. T. Lima and C. Gilles, FLORESTA, 2010, vol. 40, n. 3.
- S. A. Derkyi, B. Adu-Amankwa, D. Sekyere and A. Darkwa, International Journal of Applied Sciences (IJAS), 2011, vol. 2, n. 1.
- Z. Xiaobo, Z. Jiewen, M. J. Povey, M. Holmes, and M. Hanpin, Analytica chimica acta, 2010, vol. 667, n. 1, pp. 14-32.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

28 R. W. Kennard and L. A. Stone, *Technometrics*, 1969, vol. 11, n. 1, pp. 137-148.

29 É. S. FERREIRA, Tese (Programa de Pós-Graduação em Engenharia Florestal) - Universidade Federal do Paraná, Curitiba, 2010.

30 G. D. Matos, E. R. P. Filho, R. J. Poppi and M. A. Z. Arruda, *Revista Analytica*, 2003, n. 6, pp. 38-50.

31 A. Rinnan, F. V. D. Berg and S. B. Engelsen, *TrAC Trends in Analytical Chemistry*, 2009, vol. 28, n. 10, pp. 1201-1222.

32 H. Martens, *Multivariate calibration*. John Wiley & Sons, Norway, 1991, pp 423.

33 P. Geladi and B. R. Kowalski, *Analytica Chimica Acta*, 1986, vol. 185, pp. 1-17.

34 Ø. Hammer, D. A. T. Harper and P. D. Ryan. PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica*, 2001, vol. 4, n.1, pp. 9.

35 R. E. Bruns, I. S. Scarminio, B. de Barros Neto, *Statistical Design: Chemometrics*. Elsevier, Amsterdam, 2006, pp. 412.

36 K. A. Bakeev (Ed.), *Process analytical technology: spectroscopic tools and implementation strategies for the chemical and pharmaceutical industries*. Wiley, 2nd edn, 2010, pp. 576.

37 S. I. Ishiuchi, M. Fujii, T.W. Robinson, B. J. Miller and H. G. Kjaergaard, *The Journal of Physical Chemistry A*, 2006, vol. 110, n. 23, pp. 7345-7354.

38 L. G. Weyer, *Applied Spectroscopy Reviews*, 1985, vol. 21, n. 1-2, pp. 1-43.

Elsevier Science Limited, 2006 - 412 páginas

Analytical Methods Accepted Manuscript