The potential of near infrared spectroscopy to estimate the content of cannabinoids in *Cannabis sativa* L.: a comparative study

C. Sánchez-Carnerero Callado, N. Núñez-Sánchez, S. Casano, C. Ferreiro-Vera

www.elsevier.com/locate/talanta

PII: S0039-9140(18)30793-8
DOI: https://doi.org/10.1016/j.talanta.2018.07.085
Reference: TAL18914

To appear in: *Talanta*

Received date: 17 April 2018
Revised date: 24 July 2018
Accepted date: 26 July 2018

Cite this article as: C. Sánchez-Carnerero Callado, N. Núñez-Sánchez, S. Casano and C. Ferreiro-Vera, The potential of near infrared spectroscopy to estimate the content of cannabinoids in *Cannabis sativa* L.: a comparative study, *Talanta,* https://doi.org/10.1016/j.talanta.2018.07.085

The potential of near infrared spectroscopy to estimate the content of cannabinoids in *Cannabis sativa* L.: a comparative study.

C. Sánchez-Carnerero Callado[a,1], N. Núñez-Sánchez[b,1], S. Casano[a], C. Ferreiro-Vera[a,*]

[a]*Phytoplant Research S.L., The Science and Technology Park of Córdoba-Rabanales 21, Astronoma Cecilia Payne Street, Centauro Building, B-1, 14014. Córdoba, Spain.*

[b] *Department of Animal Production, Universidad de Córdoba, Campus Rabanales. Ctra Nacional IV-km 396. 14071 Córdoba, Spain.*

**Carlos Ferreiro Vera, Phytoplant Research S.L, The Science and Technology Park of Córdoba-Rabanales 21, Astronoma Cecilia Payne Street, Centauro Building, B-1, 14014. Córdoba, Spain. Tel/Fax: + 34 663 798 484 E-mail address: c.ferreiro@phytoplant.es*

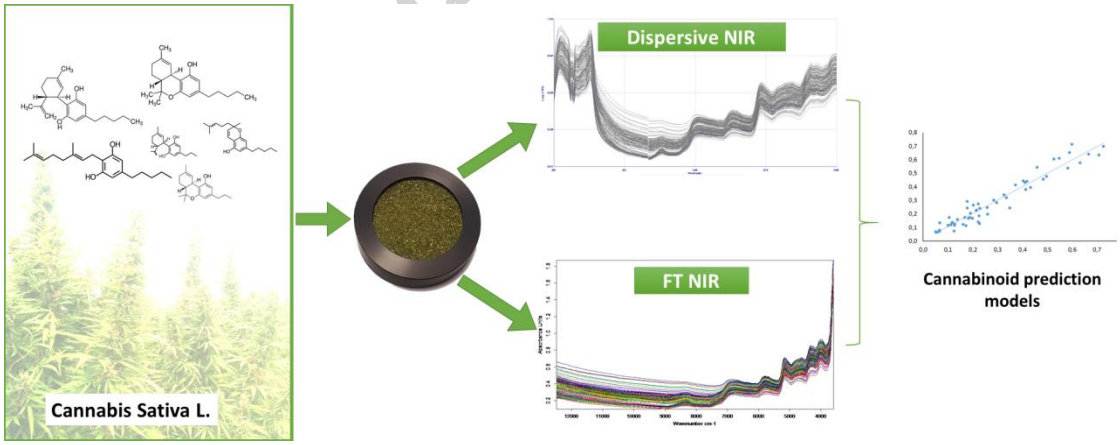
Abstract

Cannabis has been one of the oldest source of food, textile fiber and psychotropic substances. Cannabinoids are the main biologically active constituents of the *Cannabis* genus, with a demonstrated medicinal value. Its production is becoming legalized and regulated in many countries, thus increasing the need for a rapid analysis method to assess the content of cannabinoids. Gas chromatography (GC) is the preferred analytical method for the determination of these compounds, although is a slow and costly technique. Near infrared spectroscopy (NIR) has the potential for the quantitative prediction of quality parameters, and also of pharmacologically active compounds, but no references about cannabinoids prediction has been previously reported. The aim of the present research was to develop a fast, economical, robust and environmentally friendly method based on NIR technology that allow the quantification of the main

[1] *Both authors contributed equally to this work*

cannabinoids present in *Cannabis sativa* L. samples. A total of 189 grinded and dried samples from different genotypes and registered varieties were used. The content of the cannabinoids CBDV, Δ9-THCV, CBD, CBC, Δ8-THC, Δ9-THC, CBG and CBN were determined by gas chromatography. Spectra were collected in a dispersive NIR Systems 6500 instrument, and in a Fourier transform near Iinfrared (FT-NIR) equipment. The sample group was divided into calibration and validation sets, to develop modified partial lest squares (PLS) regression models with WINISI IV software with the dispersive data, and PLS models using OPUS 7.2 with the FT-NIR ones. Excellent coefficient of determination of cross validation (R^2_{CV} from 0.91 to 0.99), were obtained for the prediction of CBD, CBC, Δ8-THC, Δ9-THC, CBG and CBN, with standard error of prediction (SEP) values among 1.5 - 3 times the standard error of laboratory (SEL); and good for CBDV and Δ9-THCV cannabinoids (R^2 values of 0.89 and 0.83, respectively) with the dispersive instrument. Similar calibration and validation statistics have been obtained with the FT-NIR instrument with the same sample sets, using its specific OPUS software. In conclusion, a methodology of quantitative determination of cannabinoids in *Cannabis* raw materials has been developed for the first time using NIR and FT-NIR instruments, with similar good predictive results. This new analytical method would allow a simpler, more robust and precise estimation than the current standard GC.

Graphical Abstract:



Abbreviations:

Δ8-THC, Δ8- tetrahydrocannabinol; Δ9-THC, Δ9-tetrahydrocannabinol ; Δ9-THCV, Δ9-tetrahydrocannabivarin; CBD, cannabidiol ; CBC, cannabichromene; CBDV, cannabidivarin; CBG, cannabigerol; CBN, cannabinol; DT, detrending; MPLS, modified partial least squares ; MSC, multiplicative scatter correction ; N, number of samples; NIRS, near infrared spectroscopy; R^2, coefficient of determination of calibration; R^2_{CV}, coefficient of determination of cross validation ; RMSECV, root mean standard error of cross validation; SD, standard deviation; SEC, standard error of calibration; SECV, standard error of cross validation ; SEL, standard error of laboratory; SEP, standard error of prediction; RMSEC, root mean standard error of calibration ; SNV, standard normal variate; T, outlier samples; RSQ, coefficient of regression of reference data vs predicted data; RDP, residual predictive deviation, calculated as ratio SD/SEP

Keywords:

cannabinoids, near infrared spectroscopy, Cannabis sativa L., *quantification*

1. Introduction

Cannabis sativa L. is a species characterized for the versatility of the derived products, from food and textile fiber to psychotropic substances [1, 2, 3]. Even today it is still cultivated for industrial purposes such as food, paper and textile production [4, 5, 6] although it is mainly consumed for recreational or medicinal purposes [7], since *Cannabis* contains more than 500 different ingredients [8, 9, 10]. Chemical analysis of *Cannabis* in the 1940s and 1960s led to the discovery of a unique group of terpenophenolic secondary metabolites, known as cannabinoids, of which Δ9-tetrahydrocannabinol (Δ9-THC) was shown to be the primary psychoactive ingredient [11].

Cannabinoids are the main biologically active constituents of the *Cannabis* genus and they are produced by the glandular trichomes that occur on most aerial surfaces of the plant [12]. In 1967, Mechoulam and Gaoni [13] defined cannabinoids as the group of C21 compounds typical of the species *Cannabis sativa* L., including their carboxylic acids, analogues and transformation products. Nowadays, a new definition includes related structures or any other compound that affects cannabinoid receptor, thus an expansion of the definition considering pharmacognostic aspects is obtained.

The most important cannabinoids are tetrahydrocannabinol (Δ9-THC), cannabidiol (CBD), cannabigerol (CBG), cannabichromene (CBC) and cannabinol

(CBN), the latter as result of a degradation process. These neutral cannabinoids do not occur at significant concentrations in the plants. *Cannabis* biosynthesizes primarily the carboxylic acid forms of Δ9-THC, CBD, CBG and CBC, namely Δ9-tetrahydrocannabinolic acid (THCA), cannabidiolic acid (CBDA), cannabigerolic acid (CBGA) and cannabichromenic acid (CBCA), respectively.

In 1970s, three chemotypes of *Cannabis* plants have been recognized based on Δ9-THC/CBD ratio: a Δ9-THC/CBD ratio >> 1.0 is characteristic of "drug-type" plants (chemotype I), a Δ9-THC/CBD ratio close to 1.0 is for "intermediate-type" plants (chemotype II), and a Δ9-THC/CBD ratio << 1.0 is characteristic of "fiber-type" plants (chemotype III) [14, 15, 16].

At present day almost 150 alkyl (C-5, C-3, C-1) phytocannabinoids have been reported [17] and several other chemotypes of *Cannabis* have been selected and described [18]. Clinical trials on *Cannabis* based medicines, pure cannabinoids and synthetic analogues have demonstrated some effectiveness as analgesics for chronic neuropathic pain, glaucoma, nausea, asthma, depression, insomnia, neuralgia [19,20], appetite stimulants for cancer or AIDS patients [21], multiple sclerosis [22] and palliative agents for several other disease states and symptoms [23, 24].

As a result of this increased clinical knowledge, *Cannabis* derived products are increasingly being recognized as substances with a demonstrated medicinal value, and this is confirmed by the fact that its production and use is becoming legalized and regulated in many countries. Breeders of medicinal *Cannabis* varieties carry on a constant process of selective breeding in order to develop varieties aiming to improve the yield of certain cannabinoids and other secondary metabolites of interest for the pharmaceutical industry, or even to guarantee the absence or reduction of certain unwanted molecules, as a practical example can be mentioned the reduction of Δ9-THC in varieties with a CBD chemotype and intended to be used for controlling convulsive seizures in epileptic children. The high intrinsic variability present in *Cannabis* plants (different brands, varieties, chemotypes and gender) leads to a great difficulty in obtaining a classification standard [25]. This has led to the need for a convenient and rapid analytical method to provide quantitative data on the cannabinoids content of different sources of *Cannabis* genetic resources and then to speed up and reduce costs of the ongoing breeding program.

Gas chromatography (GC) is the preferred analytical method for the determination of these compounds since acidic cannabinoids are thermally unstable and can be

decarboxylated when exposed to heat, becoming their respective neutral cannabinoids [26]. However, GC is a slow and costly technique, which requires a tedious sample preparation stage involving at least the extraction of the active ingredients by using organic solvents, whose subsequent residues must be managed with a considerable increase of cost, time and document management.

Near Infrared Spectroscopy (NIRS) has the potential for quantitative and qualitative prediction of the main parameters such as protein, fat, moisture, fiber, ash, starch or sugar content of raw materials related with the quality of the agricultural products [27]. Cozzolino [28] reviewed applications of NIRS to the qualitative and quantitative analysis of plant natural products.

Regarding *Cannabis* plant, very few applications to qualitative and quantitative analysis of cannabinoids have been developed and reported with NIRS. Qualitative NIRS has been applied to discriminate between "drug-type" (chemotype I) and "fiber-type" (chemotype III) [29], for spectral discrimination of *Cannabis sativa* L. leaves and canopies from other plant species [30], and more recently for the prediction of the growth stage of *Cannabis* plants in the early stages of an indoor cultivation [25]. With respect to quantitative applications, only the prediction of the chemical composition of fiber and core fraction of hemp (chemotype III) has been developed [31], and no scientific papers have been found in the literature dealing with the ability of NIRS to predict the content of cannabinoids in *Cannabis* samples.

The aim of the present research was to develop a fast, economical, robust and environmentally friendly method based on NIR technology that allowed the quantification of the main cannabinoids present in *Cannabis sativa* L. samples.

2. Materials and methods

2.1. Plant material and sample preparation

A total of 189 samples of *Cannabis sativa* L. were used in this study. The samples of leaves and inflorescences were selected from plants grown under different conditions and locations (i.e. indoor with diverse PAR light exposition, and in open field, high tunnels, or high technology greenhouse), irrigation doses, plant densities, sowing times and plant varieties. Cannabis samples used in this article have been supplied by the company Phytoplant Research S.L., which is authorized by the Spanish Agency of

Medicines and Medical Devices (AEMPS) for R&D on *Cannabis* breeding, cultivation and obtainment of derived products Some samples were taken from genetic resources in use in the internal breeding program (under the research authorization of the). The sample set was completed with medicinal varieties in process of registration at the Community Plant Variety Office (CPVO), which were identified by their denomination proposals and application numbers: Aida, 2016/0167; Beatriz, 2017/0146; Enza (Divina), 2017/0149; Juani, 2016/0117; Magda, 2017/0145; Mati, 2017/0147; Marisa (Moniek), 2016/0114; Octavia, 2017/0148; Pilar, 2016/0115; Sara, 2015/0098 and Theresa, 2016/0116.Industrial hemp varieties Carma and Ermes were also included in the sample set. Besides, the samples were collected from different zones of the plant (apical, median, basal), in order to obtain samples of the same chemotype with a certain variability of the cannabinoids of interest.

Samples of the selected plant material were processed as follows: in a standing crop of a given variety or genotype, a 30 cm part of each plant containing leaves and at least one female inflorescence was collected to get around 30 g of fresh vegetal material. Plant samples were oven dried at 40 ºC for 72 hours, and the stems and seeds of more than 2 mm in size were removed. The resulting samples were ground to obtain a semi-fine powder, passing through a 1 mm sieve (test sample), and were used for further analyses.

2.2. Reference analysis of cannabinoids

Cannabis samples were analyzed in the laboratory by gas chromatography flame ionization detector (GC-FID) for the determination of the content of cannabinoids.

2.2.1. Standards

Cannabinoids standard such as, cannabidivarin (CBDV), Δ9-tetrahydrocannabivarin (Δ9-THCV), cannabichromene (CBC), delta-8 tetrahydrocannabinol (Δ8-THC), CBD, Δ9-THC, CBG and CBN were purchased from THC Pharm (Frankfurt, Germany) with the import authorization of the AEMPS. n-Hexane HPLC grade (Panreac, Barcelona, Spain) was used as extractant and squalene from Sigma-Aldrich, (St. Louis, MO, USA) was used as internal standard.

2.2.2. Instruments

An Agilent GC 7820 series (Agilent Technologies Inc, Santa Clara, CA, USA) equipped with a 7396A autosampler and a flame ionization detector (FID) was used. The instrument was equipped with a ZB-5MS capillary column (30 m length, 0.25 mm internal diameter, film thickness 0.25 µm; Phenomenex, Torrance, CA, USA), The injector temperature was 300 ºC, with an injection volume of 1 µL in splitless mode and a carrier gas (He) flow rate of 1.3 mL/min. The temperature gradient started at 160 ºC and linearly increased after 2 minutes at a rate of 25 ºC/min, until the temperature of 242 ºC, a linearly increased at a rate of 2 ºC/min until 250 ºC and a new linearly increased at a rate of 25 ºC/min until the final temperature of 300 ºC which was held for 5 min. The FID detector temperature was set to 300 ºC. The device was controlled by Agilent GC OpenLAB Workstation software.

2.2.3. GC-FID analysis

According with the law, an official method (EC) No 1177/2000 of 31 May 2000 Annex C, named "Community method for the quantitative determination of THC content of hemp varieties" was used in the wet way. This method is based on the quantitative determination of Δ9-THC by GC although slight modifications were made in the temperature ramp of the chromatographic oven for the simultaneous determination of cannabinoids (CBDV, Δ9-THCV, CBD, CBC, Δ8-THC, Δ9-THC, CBG and CBN).

Briefly, the wet analysis steps were: weigh 100 mg of the powdered test sample, place in a centrifuge tube and add 5 mL of n-hexane containing squalene, as an internal standard. Sonicate in an ultrasonic bath for 20 minutes. Centrifuge for five minutes at 3000 r.p.m. and then remove the supernatant cannabinoid solution. Inject the solution, after necessary dilution, into the GC-FID to carry out a quantitative analysis. All determinations in the laboratory were performed in duplicate, and the standard error of laboratory (SEL) was calculated for each of the reference parameters:

$$SEL = \sqrt{\frac{\sum_{i=1}^{n} (y_{i1} - y_{i2})^2}{2n}}$$

Where n is the number of samples and y_{i1} and y_{i2} are values obtained for the replicates 1 and 2, respectively, of sample i.

2.3. Spectra collection

All the 189 dried and grounded *Cannabis* samples were scanned consecutively in two NIR instruments with different optical characteristics. Circular quartz cuvettes (48 mm diameter) were used as sample holders for reflectance analysis in both instruments.

2.3.1. Dispersive NIR spectrometer

A dispersive NIR spectrometer, NIR Systems 6500 scanning monocromator System II (FOSS NIR Systems, Silver Spring, Maryland, USA) equipped with a transport module, was used to measure reflectance spectra from 400 to 2498 nm, every 2 nm (Si detector for 400–1100 nm region and PbS for 1100 - 2500 nm). Circular quartz cuvettes were placed in special holders for circular capsules. Spectral absorbance values were recorded in reflectance mode as log 1/R, where R is the sample reflectance. During the sample analysis, an internal white ceramic reference was scanned 16 times and afterwards the sample was scanned 32 times, obtaining a final spectrum for each cuvette and sample. WINISI II software version 1.50 (Infrasoft International LLC, Port Matilda, PA, USA) was used for spectral data collection.

2.3.2. FT-NIR spectrometer

Fourier transform near infrared (FT-NIR) spectra were recorded on a multipurpose analyzer (MPA) FT-NIR spectrometer (Bruker Optics GmbH, Ettlingen, Germany) equipped with an integrated Michelson interferometer and a highly sensitive PbS detector. Spectra were collected in diffuse reflectance mode in the wavenumber range 12500–4000 cm^{-1} (800-2500 nm), with a resolution of 8 cm^{-1}, using the macro sample integrating sphere measurement channel. Each spectrum was recorded as the average of 32 subsequent scans. The background was recorded automatically with a gold coated reference every hour, approximately. OPUS 7.2 software (Bruker Optics GmbH, Ettlingen, Germany) was used for spectra acquisition.

2.4. Population structuring and spectral outlier detection

Prior to NIRS calibration development, principal components analysis (PCA), available in WINISI IV software, was performed on the full set of samples (N=189) in order to reduce the dimensionality of the data matrix and to retain the maximal amount of variability in the spectral data. This PCA was performed by using standard normal

variate and detrending (SNV and DT) algorithms [32], as pre-treatments to remove the interferences of scatter usually found in absorbance data. This pretreatment was used together with a Norris–Williams spectral derivative "1,5,5,1", where the first digit is the order of the derivative, the second is the gap over which the derivative is calculated, the third is the number of data points in a running average or smoothing, and the fourth the second smoothing [33].

After the PCA analysis, the center of the spectral population was determined, and the GH value (Mahalanobis global distance to the center of the population) of each sample calculated, in order to detect outliers. The resulting population was sorted according to the GH value, considering as outliers those samples with GH value 3.0 or above, which were eliminated [34]. This procedure was repeated until no outliers were found. Before calibration development, the sample set without spectral outliers was divided into calibration and validation sets. Hence, the samples in the validation set were not used along the calibration process and were saved to develop external validation. The procedure for the sample selection to integrate each group was based on the spectral information as follows: one of each 3 sorted samples from the resulting population in the previous PCA step were selected to build up the validation set, and the rest remained as the calibration set. In this way, the samples of the validation set are consistently distributed in the population [34, 35].

2.5 Data processing and calibration development

Multivariate analysis was used to develop prediction equations, using the resulting calibration and validation sample sets from the PCA without outliers. Separate regression models were performed for each parameter in the calibration sets in both instruments using their respective chemometric software. Cannabis spectra obtained in the dispersive NIR Systems instrument and in the MPA FT-NIR instrument were fitted to the content of cannabinoids determined by GC to develop the correspondent NIR prediction equations. In this work, concentrations of zero cannabinoid found in the GC laboratory analysis were considered as values, both in calibration and validation procedures.

Partial least squares (PLS) regression method was used to develop predictive models in the FT-NIR instrument. The spectral and concentration data are first encoded in matrix form and then reduced to few factors. Hence, the resulting spectral vectors are

directly related to the constituents of interest. PLS1 algorithm provided by the OPUS 7.2 and Bruker Quant 2 software packs was used in this study with the FT-NIR spectra. A modification of this methodology, available in WINISI IV software, was used for the dispersive reflectance spectra obtained in the NIR Systems 6500 instrument, owing to MPLS is often more stable and accurate than the standard PLS algorithm. In modified partial least squares (MPLS) method of regression, the NIR residuals at each wavelength, obtained after each factor is calculated, are standardized (divided by the standard deviations of the residuals at a wavelength) before calculating the next factor. Cross validation was used to select the optimal number of factors and to avoid overfitting [36]. The number of PLS factors required to model the data represents the optimal number of factors or latent variables to retain for generating a predictive model. If this value is too small, the information of interest in the spectra is not fully utilized, whereas if it is too big, extra noise may be included in the model. Cross validation was performed by splitting the calibration set into four groups, according to Shenk and Westerhaus [37] in the NIR dispersive instrument, using WINISI IV. Each group was predicted using a calibration developed on the other samples, until all the samples have been used for validation. Finally, validation errors were combined to obtain a standard error of cross validation (SECV). The optimum number of factors was determined as the one that provided the lowest SECV. For the FT-NIR calibration data set, internal validation was performed using a leave-one-out cross validation using the OPUS software, and the optimum number of factors was determined here as the one that provided the lowest root mean square error of cross-validation (RMSECV), up to a maximum of 12, [38].

Multivariate regression equations for each cannabinoid were obtained in WINISI IV after combining several spectra pretreatments in different wavelength regions. SNV and DT [32], and multiplicative scatter correction (MSC) [39] pretreatments were used to remove additive and multiplicative effects in the spectra, together with Norris–Williams spectral derivatives [33], to emphasize steep edges of a peak. Four of the most usual derivative pretreatments (1, 5, 5, 1; 1, 10, 10, 1; 2, 5, 5, 1 and 2, 10, 10, 1), have been used in this study, as described by Núñez-Sánchez [40], following Shenk and Westerhaus [36,37] recommendations for agricultural products. Three spectral regions were considered to develop the equations: 400-2500 nm (complete VIS+NIR region of the NIR Systems 6500 instrument), 1100-2500 nm, as the NIR region recorded by the

PbS detector in that instrument, and 800-2500 nm, which corresponds to the scanning range of the FT-NIR instrument (9000-4000 cm^{-1}). Hence, twenty-four regression equations per parameter were developed by combining four spectral derivative math treatments and two scatter correction methods in three spectral regions. Two passes of elimination of T outlier samples (samples with abnormally high residuals of predicted versus reference values), were applied. A critical T value of 2.5 was set for T outliers.

In OPUS 7.2 software, several spectral pre-processing methods available in the Quant 2 algorithm were applied: first and second derivatives, first derivative combined with vector normalization (similar to the SNV treatment), and first derivative combined with MSC. Besides, a set of five frequency regions typical for NIR applications were used to optimize the models: 9400-7500, 7500-6100, 6100-5450, 5450-4600 and 4600-4250 cm^{-1}. These five frequency regions are tested on their own and in all possible combinations [38].

The performance of the calibration models was evaluated using the standard error of calibration (SEC), the coefficient of determination for calibration (R2), the standard error of cross validation (SECV) and the coefficient of determination for cross validation (R^2_{CV}) [36,41] provided by the WINISI IV software. Similarly, the performance of the PLS models developed with OPUS Quant 2 algorithm was evaluated in terms of the root mean square error of calibration (RMSEC) and cross validation (RMSECV), together with the R^2 and R^2_{CV} provided by this software.

The predictive ability of the equations was assessed by external validation, by means of a simple regression between NIRS predicted values and reference data. Hence, the models developed with the strategies described above were applied to the validation sample sets, which had not been used in the calibration procedures. Hence, the comparison of lab data *versus* predicted data provided the SEP value (standard error of prediction) or RMSEP (root mean square error of prediction) in WINISI IV and OPUS, respectively, together with the slope and regression coefficients (RSQ). Furthermore, paired t-test was performed to evaluate the significance of the differences between laboratory and predicted results of models in both instruments, in the validation sets.

Additionally, the RPD (residual predictive deviation) statistic, or ratio of the standard deviation (SD) of the original data to the SEP (SD/SEP) in WINISI models, or to the RMSEP (SD/RMSEP) in OPUS models, according to Williams [41], was used to

evaluate the predictive ability of the models. This statistic is provided by the OPUS software, whereas it has to be calculated for the WINISI IV equations.

For each cannabinoid, the best model was selected according to the higher R^2_{CV} and RPD values and lower SECV or RMSECV and SEP or RMSEP, in both instruments.

3. Results and discussion

3.1. Cannabinoids data

The strategy followed to select the plant material to generate the calibration set consisted on choosing *Cannabis* samples of different chemotypes from 11 medicinal varieties in process of registration at CPVO and 27 other genetic resources, which were characterized by different chemotypes, cultivated in various locations and collected from different areas of the plant. This strategyallowed to get a high variability of the content of the cannabinoids CBDV, Δ9-THCV, CBC, CBD, Δ8-THC, Δ9-THC, CBG and CBN (Table 1). In fact, there are samples with concentrations up to 12.93 % CBD and 22.14 % Δ9-THC.

The uniform distribution on the chemical measurements was described as the ideal to get robust calibrations by Williams [41], instead of a Gaussian distribution, as all the values are equally represented in the calibration set. Hence, the effect by which the results of future analyses show to regress towards the mean will be avoided. The histograms of frequencies showed in general a continuous distribution of the reference values across the concentration intervals (Figure 1). Although the distribution of samples was not uniform, the continuous presence of samples between minimum and maximum values, and not having a Gaussian distribution was considered a positive feature of the calibration set, as the models will not show the results of future analyses to regress towards the mean [41].

In addition, the procedure followed to split the sample set into calibration and validation sets, which consisted of sorting one every 3rd sample from the collective ordered upon the GH value for the validation set, allowed to select samples from all the cultivars and crop fields in a similar proportion (30%), thus providing a validation set

with similar chemical characteristics in both the calibration and validation sets (Table 1).

The standard error of laboratory (SEL) was calculated for each parameter to measure the prediction of the method (Table 1). The comparison of these values with the NIR calibration and validation errors was used to evaluate the performance of the models for the cannabinoid prediction in each instrument.

3.2. Spectral features and population structure

The spectra of the sample set of the C*annabis* plants are shown in Figure 2. Figure 2A corresponds to spectra collected in the FNS scanning spectrometer provided with two detectors (400-1098 nm and 1100-2500 nm). Therefore, visible region is also recorded with instrument. The slight discrepancies observed in the visible region (400–750 nm) among spectra can be attributed to the colour, resembling different intensities of green and ochre of the samples. With respect to the spectra collected in the FT-NIR instrument (Figure 2B), absorbance values were recorded in the wavenumber frequencies between 12500 and 4000 cm^{-1} (780-2500 nm), so the visible region is not included in the spectra. Besides, the region between 12500 and 9400 cm^{-1} does not contain information of interest in the FT-NIR spectra, so it was not used for calibration development.

Absorption bands typical of plant material appear in all the spectra and in both instruments in the NIR region from 1100-2500 nm, being the absorbance differences among spectra due to changes in their chemical composition. Bands at 1450 and 1930 nm, corresponding to the first overtone and combination bands of the -OH group, mostly present in water, can be found in the spectra. Although samples had been oven-dried, probably, residual water from samples handling was responsible for them. Bands associated with lipids were found around 1210 nm (second overtone of stretching vibrations), 1730 and 1760 nm (first overtone of stretching vibrations), and around 2310 and 2350 nm (combination of stretching and bending vibrations) of $-CH_2$ and $-CH_3$ functional groups. Bands associated at 2058 and 2166 nm have been associated with absorption of protein, and around 2078–2110 nm and 2268 nm are related to fibre content, as reported by Williams [41].

Besides, the absorbance band found at 1666 nm corresponds to aromatic hydrocarbons of the *Cannabis* terpenes (first overtone of CH-stretching). This band is more evident in samples having contents of CBD above 4%.

As for the population structure, the PCA applied to determine the structure the spectral variability of the population [35] allowed the detection of two samples as outliers (GH value > 3), which were removed. From the remaining 187 samples, every third sample was selected for the validation set (56 samples), and the rest remained as the calibration set (131 samples). Figure 3 shows the score plot of the three first principal components (PCs) of the calibration set (crosses) and the overplot of the validation samples spectra (squares). The structured selection of samples for validation, upon the spectral information, allowed obtaining a validation set homogeneously distributed in the population, as well as with similar chemical composition (Table 1).

3.3. Calibration development

In this work, the same calibration and validation sets (131 and 56 samples, respectively), resulting from the population structuring process, were used for the development and validation of the predictive models in both instruments using their respective chemometric software.

3.3.1. Dispersive NIR spectrometer

The WinISI IV software package was used for the chemometric management of data. Separate MPLS calibrations were performed for each parameter, using the calibration set (131 samples). The combination of four spectral derivative math treatments (1, 5, 5, 1; 1, 10, 10, 1; 2, 5, 5, 1 and 2, 10, 10, 1), and two scatter correction algorithms (SNV and DT, and MSC) in three spectral regions (400–2500; 1100–2500 and 800–2500 nm) provided 24 equations for the prediction of each cannabinoid. Several statistics (SEC, R^2, SECV and R^2_{CV}), describing the performance of the calibration, were provided by the software. The predictive ability of the equations was assessed by applying these equations to the external validation set (56 samples). SEP values, together with the slope and RSQ resulting from the comparison of lab *versus* predicted data, were obtained. The selection of the best calibrations was based on the higher R^2_{CV} and RPD values and lower SECV and SEP (Table 2).

The visible spectral range did not provide relevant information to improve the results, as none of the selected equations used the complete spectral range from 400-2500 nm. In general, better results were obtained using a second derivative treatment, and only Δ8-THC, Δ9-THC and CBG equations use the first derivative pre-treatment of the data. With respect to the scatter correction, the combination of SNV and DT pre-treatments showed to provide improved results in all the cannabinoids except Δ9-THCV and Δ9-THC. For these parameters, MSC scatter correction presented enhanced results.

The number of PLS terms of the selected models varied between 7 for CBN and 12 for CBDV and Δ9-THCV parameters.

Different criteria have been proposed for the evaluation of the best models selected for each parameter in each instrument and for the comparison of results. Williams [42] indicated that calibration models having a value for R^2 above 0.91 are considered to be excellent, and good prediction if R^2 is between 0.82 and 0.90. A value for R^2 between 0.66 and 0.81 indicates approximate quantitative predictions, whereas, values of R^2 between 0.50 and 0.65 indicates that more than 50% of the variance in Y is accounted for variance X, so that discrimination between high and low concentrations can be made The R^2_{CV} values of the selected models obtained with the dispersive instrument NIR Systems 6500 scanning monocromator (Table 2) were higher than 0.91 (from 0.91 to 0.99), thus showing to be excellent, for the prediction of CBD, CBC, Δ8-THC, Δ9-THC, CBG and CBN, and good for CBDV and Δ9-THCV cannabinoids (R^2_{CV} values of 0.89 and 0.83, respectively), according to Williams [42].

With respect to the comparison of SECV with SEL values, Shenk and Westerhaus [37] proposed the following criteria: SEP values 1 to 1.5 times SEL show excellent precision; good precision if SEP is 2 to 3 times SEL, moderate precision if SEP is 4 to 5 times SEL, whereas models with SEP values above 5 times SEL of the reference method are considered to have poor precision. According to this rule, the comparison of SEL and SEP in this work show that CBG model has an excellent precision, Δ8-THC, Δ9-THC, CBD and CBC have good precision, and Δ9-THCV moderate precision. Only SEP values of CBDV are over 5 times SEL (Tables 1 and 2).

RPD is a non-dimensional statistic for the quick assessment of a NIR spectroscopy calibration model, as it enables the evaluation of a SECV or SEP in terms of the standard deviation (SD) of the reference data. Williams [43] graded NIR

applications in forages, feeds and soils according to their RPD values, indicating that values from 2.5 to 2.9 are considered fair, and appropriate for screening purposes, from 3 to 3.4 can be considered as good for quality control, above 3.5 suitable for process control, and excellent those above 4.1. Calibrations with RPD values below 2 are considered to be poor, and not recommended. According to this statistic, the models to predict CBG and Δ9-THCV are not yet suitable to be used (RPD 1.25 and 1.84, respectively), whereas CBN, Δ8-THC, and CBDV models can be considered fair, and appropriate for screening purposes. The models for Δ9-THC, CBC and CBD can be evaluated as good, very good and excellent (RPD values 3.07, 3.79 and 6.03, respectively), as they have enough accuracy to be used in routine analysis (Table 2).

As for the validation results (Table 2; Figure 4A), after the application of the selected models to the validation set, in general, SEP values are similar to SECV and slope and RSQ are close to or above 0.9 in all parameters except CBG and CBN, which results are less accurate. Additionally, paired t-test showed no significant differences between reference and NIR predicted values in the validation set ($p > 0.05$), thus confirming the reliability of the NIR technology for the measurement of the cannabinoids in the dispersive instrument.

3.3.2. FT-NIR spectrometer

The PLS-1 algorithm included in the Quant2 module of the OPUS 7.2 software package was used for the development of predictive calibration, using the same 131 samples as with the FNS 6500 instrument.

For the model development, first and second derivative, first derivative combined with vector normalization, and first derivative combined with MSC were used as spectral pre-treatment of the data. Although second derivative showed better results in most of the cannabinoids, the selected equation to predict Δ9-THC used first derivative, as also did with the FOSS NIR Systems 6500. Unexpectedly, the use of algorithms to correct the scatter effect provided better results only in the equations for the prediction of CBN (MSC pre-treatment) and CBC (SNV pre-treatment).

Despite the instrument collects data between 12500 – 4000 cm^{-1} (800–2500 nm), the region from 12500 to 9400 cm^{-1} provides no relevant information (Figure 2B), so it was not used for the calibration development. Hence, five frequency regions typical for NIR applications were used to optimize the models: 9400 – 7500, 7500 –6100, 6100 –

5450, 5450 – 4600 and 4600 – 4250 cm^{-1}. None of the selected equations showed to use the complete wavelength region, and there was not a specific spectral region used for the prediction of all the cannabinoids (Figure 5). However, 6100 – 5450 cm^{-1} (1639 – 1835 nm) spectral region, which includes the band linked to terpenes absorbance, has been included in all the models except CBG, probably due to the fact that the molecular structure of this particular cannabinoid differs from the rest.

According to Williams [42], the selected models obtained with the FT-NIR instrument showed to be excellent for the prediction of all cannabinoids (R^2_{CV} values from 0.93 to 0.99), except CBDV and Δ9-THCV (R^2_{CV} values of 0.81 and 0.77, respectively), which can be considered of good precision (Table 3). The comparison of SEL and SEP offer similar results as the ones obtained with the dispersive instrument, and CBG has an excellent precision, Δ8-THC, Δ9-THC, CBD and CBC have good precision, Δ9-THCV moderate precision, and SEP values of CBDV are over 5 times SEL (Tables 1 and 3). According to RPD statistic, the CBN model, together with the ones to predict CBG and Δ9-THCV, are not yet suitable to be used, whereas, Δ8-THC, and CBDV models can be considered appropriate for screening purposes. The models for Δ9-THC, CBC and CBD can be evaluated as good, very good and excellent as their RPD values exceed 3, and reach a value of 6, for the latter one (Table 3).

External validation shows similar results to those obtained with the dispersive instrument (Table 3; Figure 4B), after the application of the selected models to the validation set. Likewise, paired t-test showed no significant differences ($p > 0.05$) between cannabinoids reference and NIR predicted values after applying the models developed in the FT-NIR instrument to the validation set.

The calibration and validation statistics obtained in both instruments, using the same calibration and validation sets but their specific software, are comparable (Tables 2 and 3). Similarly, the plots of measured values (X axis) versus predicted (Y axis) of cannabinoids of the validation data set (Figure 4A,B) show analogous results, and the same samples with higher residual values of each parameter can be found in both instruments. However, it is noticeable that the models obtained with the FT-NIR instrument tend to eliminate a higher number of samples along the calibration process, and not to use scatter correction pre-treatment to get the better performance results.

NIR has been used in *Cannabis* for qualitative purposes (discrimination between Δ9-THC-rich and hemp [29], and discrimination of *Cannabis* leaves and canopies from other plant species [30]), but no previous reports on the development of predictive models with samples of chemotypes different than I, II, and III have been found in our bibliographic research, so it was not possible to contrast the results obtained here. With respect to other active pharmaceutical parameters, similar statistics have been reported on NIR prediction of alkaloids in *Papaver somniferum* [44] and in perennial ryegrass [45] or phenolic compounds in plants [46].

4. Conclusions

The methodology of quantitative determination of cannabinoids in *Cannabis* raw materials has been developed using near infrared spectroscopy (NIR) and Fourier-Transform Near-Infrared Spectroscopy (FT-NIR) for the first time. This new analytical method would allow a simpler, a more robust and precise estimation than the current standard GC. The results obtained allow to confirm that there is sufficient information in the NIRS spectral region for the development of cannabinoid prediction models in dried and ground *Cannabis* plant materials. Hence, the NIR technique has been proven as an alternative method to the conventional GC analysis for the assessment of *Cannabis sativa* L. varieties and genetic resources as regard to their cannabinoids contents. The major advantage of such spectroscopic technique is that the sample is measured with no or little pre-treatment (e.g. grinding, drying), is faster, simpler, and uses no pollutants for the environment compared with the reference analytical method. The comparison of the predictive ability of the models obtained with the dispersive NIR and the FT-NIR spectrometers show no significant differences between them, although pre-treatments for the scatter correction and derivatives, as well as the outlier detection during the calibration process with the chemometric software were diverse.

References

[1] R.E. Schultes, W.M. Klein, T. Plowman, T.E. Lockwood, Cannabis: an example of taxonomic neglect, Harvard Univ. Bot. Mus. Leafl. 23 (1974) 337−367.

[2] M.D. Merlin, Archaeological evidence for the tradition of psychoactive plant use in the old world, Econ. Bot. 57 (2003) 295−323.

[3] H.E. Jiang, X. Li, Y.X. Zhao, F. Hueber, S. Bera, Y.F. Wang, L.C. Zhao, C.J. Liu, C.S. Li, A new insight into Cannabis sativa (cannabaceae) utilization from 2500-year-old Yanghai Tombs, Xinjiang, China, J. Ethonopharmacol. 108 (2006) 414−422.

[4] P. Ranalli, M. Di Candilo, G. Mandolino, G. Grassi, A. Carboni, Hemp for sustainable agricultural systems, Agro Food Ind Hi Tech. 10 (1999) 33−38.

[5] J.C. Callaway, Hempseed as a nutritional resource: an overview, Euphytica. 140 (2004) 65−72.

[6] U. Kriese, E. Schumann, W.E. Weber, M. Beyer, L. Brühl, B. Matthäus, Oil content, tocopherol composition and fatty acid patterns of the seeds of 51 Cannabis sativa L. genotypes, Euphytica. 137 (2004) 339−351.

[7] M.A. ElSohly, Marijuana and the Cannabinoids (Forensic Science and Medicine), Humana Press Totowa, New Jersey, 2007.

[8] C.E. Turner, M.A. ElSohly, E.G. Boeren, Constituents of Cannabis sativa L. XVII. A review of the natural constituents, J. Nat. Prod. 43 (1980) 169–234.

[9] S.A. Ross, M.A. ElSohly, Constituents of Cannabis sativa L. XXVIII. A review of the natural constituents: 1980–1994, Zagazig J. Pharm. Sci. 4 (1995) 1–10.

[10] M.A. ElSholy, W. Gul, Handbook of Cannabis, Oxford University Press, Oxford, 2014.

[11] R.G. Pertwee, Cannabinoid pharmacology: the first 66 years. Br. J. Pharmacol. 147 (2006) 163−171.

[12] P. Dayanandan, P.B. Daufman, Trichomes of Cannabis sativa L. (Cannabaceae), Am. J. Bot. 63 (1976) 578–591.

[13] R. Mechoulam, Y. Gaoni, The absolute configuration of Δ1-tetra-hydrocannabinol, the major active constituent of hashish, Tetrahedron Lett. 12 (1967) 1109−1111.

[14] P.S. Fetterman, E.S. Keith, C.W. Waller, O. Guerrero, N.J. Doorenbos, M.W. Quimb, Mississippi-grown Cannabis sativa L. Preliminary observation on chemical definition of phenotype and variations in tetrahydrocannabinol content versus age, sex, and plant part, J. Pharm. Sci. 60 (1971) 1246–1249.

[15] E. Small, H.D. Beckstead, Cannabinoid phenotypes in Cannabis sativa, Nature. 245 (1973) 147–148.

[16] E. Small, H.D. Beckstead, Common cannabinoid phenotypes in 350 stocks of Cannabis, Lloydia. 36 (1973) 144–165.

[17] L.O. Hanus, S. Martin Meyer, E. Muñoz, O. Taglialatela Scafati, G. Appendino, Phytocannabinoids: a unified critical inventory, Nat. Prod. Rep. 33 (2016) 1357–1392.

[18] E.P.M. de Meijer, K.M. Hammond, The inheritance of chemical phenotype in *Cannabis sativa* L. (V): regulation of the propyl-/pentyl cannabinoid ratio, completion of a genetic model, Euphytica. 210 (2016) 291–307.

[19] S. Mechoulam, N. Lander, S. Dikstein, E.A. Carlini, M. Blumenthal, On the therapeutic possibilities of some cannabinoids, in: S. Cohen, R. Stillman (Eds.), The Therapeutic Potential of Marihuana, New York: Plenum Med. Book Co, 1976, pp. 35–45.

[20] J.A. Duke, K.K. Wain, Medicinal Plants of the world, Computer index with more than 85.000 entries, in: J.A. Duke (Ed.), Handbook of Medicinal Herbs, Boca Raton, FL: CRC Press, 1981, pp. 96.

[21] D.I. Abrams, C.A. Jay, S.B. Shade, H. Vizoso, H. Reda, S. Press, M.E. Kelly, M.C. Rowbotham, K.L. Petersen, Cannabis in painful HIV-associated sensory neurpathy: a randomized placebo-controlled trial, Neurology. 68 (2007) 515–521.

[22] G. Pryce, D. Baker, Emerging properties of cannabinoid medicines in management of multiple sclerosis, Trends. Neurosci. 28 (2005) 272–276.

[23] R. Mechoulam, E.A. Carlini, Toward drugs derived from cannabis, Naturwissenschaften. 65 (1978) 174–179.

[24] J.M. Cunha, E.A. Carlini, A.E. Pereira, O.L. Ramos, C. Pimentel, R. Gagliardi, W.L. Sanvito, N. Lander, R. Mechoulam, Chronic administration of cannabidiol to healthy volunteers and epileptic patients, Pharmacology. 21 (1980) 175–185.

[25] B.T. Borille, R.S. Ortiz, K.C. Mariotti, G. Vanini, L.V. Tose, P.R. Filgueiras, M.C.A. Marcelo, M.F. Ferrão, M.J. Anzanello, R.P. Limbergerac, W. Romão, Chemical profiling and classification of cannabis through electrospray ionization coupled to Fourier transform ion cyclotron resonance mass spectrometry and chemometrics, Anal. Methods. 9 (2017) 4070–4081.

[26] A.K. Hewavitharana, G. Golding, G. Tempany, G. King, N. Holling, Quantitative GC-MS analysis of Δ9-tetrahydrocannabinol in fiber hemp varieties. J. Anal. Toxicol. 29 (2005) 258–261.

[27] P.C. Williams, K. Norris, Near-Infrared Technology in Agricultural and Food Industries, second ed., American Association of Cereal Chemists, Inc. St. Paul Minnesota, USA, 2001, pp. 295.

[28] D. Cozzolino, Near Infrared Spectroscopy in Natural Products Analysis, Planta Med. 75 (2009) 746–756.

[29] N. Wilson, M. Heinrich, The use of near infrared spectroscopy to discriminate between THC-rich and hemp forms of cannabis, Planta Med. 72 (2006) 1059–1059.

[30] C.S.T. Daughtry, C.L. Walthall, Spectral discrimination of Cannabis sativa L. leaves and canopies, Remote Sens. Environ. 64 (1998) 192–201.

[31] M.A.J. Toonen, C. Maliepaard, T.H. Reijmers, H. Van der Voet, H.D. Mastebroek, H.C. Van den Broeck, M.J.M. Ebskamp, W. Kessler, R.W. Kessler, Predicting the chemical composition of fibre and core fraction of hemp (Cannabis sativa L), Euphytica. 140 (2004) 39–45.

[32] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard Normal Variate transformation and De-trending of near infrared diffuse reflectance spectra. Appl. Spectrosc. 43 (1989) 772–777.

[33] K.H. Norris, P.C. Williams, Optimization of mathematical treatment of raw near-infrared signal in the measurement of protein in hard red spring wheat. I. Influence of particle size, Cereal Chem. 61 (1984) 158–165.

[34] J.S. Shenk, M.O. Westerhaus, Population definition, sample selection and calibration procedures for near infrared spectra and modified partial least squares regression, Crop Sci. 31 (1991) 469–474.

[35] J.S. Shenk, M.O. Westerhaus, Population structuring of Near Infrared spectra and modified partial least squares regression, Crop Sci. 31 (1991) 1548–1555.

[36] J.S. Shenk, M.O. Westerhaus, Analysis of agriculture and food products by near infrared reflectance spectroscopy. Monograph, Infra Int. Port Matilda, PA12101 Tech Road, Silver Spring, MD, USA, 1995.

[37] J.S. Shenk, M.O. Westerhaus, Calibration de ISI way, in: Near Infrared Spectroscopy: the future waves, A.M.C. Davies, P. Williams (Eds.), NIR Publications, Chichester, West Sussex, UK, 1996, pp. 198–202.

[38] OPUS, User manual. Ettlingen, Germany. 2007.

[39] P. Geladi, D. McDougall, H. Martens, Linearization and scatter corrections for near-infrared reflectance spectra of meat, Appl. Spectrosc. 39 (1985) 491–500.

[40] N. Núnez-Sánchez, D. Carrion, F. Peña, V. Domenech, A. Garzón, A.L. Martínez-Marín, Evaluation of botanical and chemical composition of sheep, Anim. Feed Sci. Technol., 222 (2016), 1–6.

[41] P.C. Williams, Implementation of Near-Infrared technology, in: P.C. Williams, K.H. Norris (Eds.), Near-Infrared Technology in the Agricultural and Food Industries, 2001, pp. 145–171.

[42] P. Williams, Near-infrared technology-Getting the best out of light, A short course in the practical implementation of near infrared spectroscopy for the user (ed 1.1), Nanaimo, Canada, 2003, pp. 109.

[43] P. Williams, The RPD Statistic: A Tutorial Note, NIR News, 25 (2014) 22–26.

[44] S. Gergely, E. Izsó, A. Salgó, Comparison of different types of near infrared (NIR) instruments in ability to measure alkaloids in capsule of poppy (P. somniferum), Period. Polytech. Chem. Eng. 59 (2015) 277–282.

[45] M.C. Soto-Barajas, I. Zabalgogeazcoa, I. González-Martin, B.R. Vázquez-de-Aldana, Qualitative and quantitative analysis of endophyte alkaloids in perennial ryegrass using near-infrared spectroscopy, J. Sci. Food Agric. 97 (2017) 5028–5036.

[46] Bittner, K., Lukas; A., Schonbichler, S, Bonn, K., Gunther; W., Huck, C, 2013. Near Infrared Spectroscopy (NIRS) as a tool to analyze phenolic compounds in plants Current Analytical Chemistry, 9 (3): 417–423.

Figure 1. ***Distribution of the cannabinoid values for the initial sample set.***

CBDV: cannabidivarin; Δ9-THCV: Δ9-tetrahydrocannabivarin; CBD: cannabidiol; CBC: cannabichromene; Δ8-THC: Δ8-tetrahydrocannabinol; Δ9-THC: Δ9-tetrahydrocannabinol; CBG: cannabigerol; CBN: cannabinol

Figure 2. ***Spectra of the complete set of samples obtained with the dispersive instrument NIRSystems 6500, in the wavelength region 400 – 2500 nm (A), and the FT-NIR instrument MPA, in the 12.000 – 4000 cm^{-1} region (B).***

Figure 3. ***Score plot of the three first PCA components of the calibration set (crosses), and the overplot of the validation set (squares).***

Figure 4A. ***Plot of measured values (x axis) versus predicted (y axis) of cannabinoids of the validation data set with the selected equations obtained in the dispersive NIR Systems 6500 instrument.***

CBDV: cannabidivarin; Δ9-THCV: Δ9-tetrahydrocannabivarin; CBD: cannabidiol; CBC: cannabichromene; Δ8-THC: Δ8-tetrahydrocannabinol; Δ9-THC: Δ9-tetrahydrocannabinol; CBG: cannabigerol; CBN: cannabinol

Figure 4B. ***Plot of measured values (x axis) versus predicted (y axis) of cannabinoids of the validation data set with the selected equations obtained in the FT-NIR MPA instrument.***

CBDV: cannabidivarin; Δ9-THCV: Δ9-tetrahydrocannabivarin; CBD: cannabidiol; CBC: cannabichromene; Δ8-THC: Δ8-tetrahydrocannabinol; Δ9-THC: Δ9-tetrahydrocannabinol; CBG: cannabigerol; CBN: cannabinol

Figure 5. ***Spectral regions used in the selected equations developed with the FT-NIR MPA, Bruker optics.***

CBDV: cannabidivarin; Δ9-THCV: Δ9-tetrahydrocannabivarin; CBD: cannabidiol; CBC: cannabichromene; Δ8-THC: Δ8-tetrahydrocannabinol; Δ9-THC: Δ9-tetrahydrocannabinol; CBG: cannabigerol; CBN: cannabinol

Table 1. Cannabinoid contents (%) quantified by gas chromatography of the calibration and validation sample sets. The standard error of laboratory of each parameter are also included.

Cannabinoid	Calibration set (N = 131)				Validation set (N = 56)				SEL
	Mean	SD	Minimum	Maximum	Mean	SD	Minimum	Maximum	
CBDV	0.3	0.4	0.0	2.0	0.3	0.6	0.0	2.1	0.02
Δ9-THCV	0.1	0.1	0.0	0.3	0.1	0.1	0.0	0.3	0.01
CBD	3.3	3.8	0.1	12.9	3.2	3.7	0.0	12.9	0.29
CBC	0.3	0.2	0.0	0.7	0.3	0.2	0.1	0.7	0.02
Δ8-THC	0.2	0.2	0.0	0.8	0.1	0.2	0.0	0.7	0.02
Δ9-THC	4.0	5.8	0.0	22.0	3.9	5.5	0.0	22.1	0.82
CBG	1.2	1.2	0.1	6.7	1.1	1.0	0.2	5.4	0.45
CBN	0.1	0.1	0.0	0.5	0.1	0.1	0.0	0.4	0.02

CBDV: cannabidivarin; Δ9-THCV: Δ9-tetrahydrocannabivarin; CBD: cannabidiol; CBC: cannabichromene; Δ8-THC: Δ8-tetrahydrocannabinol; Δ9-THC: Δ9-tetrahydrocannabinol; CBG:

cannabigerol; CBN: cannabinol; SD: standard deviation; N: number of samples; SEL: standard error of laboratory.

Table 2. Calibration and validation statistics of the selected equations obtained to predict the cannabinoid contents of samples analyzed in the dispersive NIR Systems 6500 scanning monocromator.

Cannabinoid	CALIBRATION											VALIDATION			
	Region (nm)	Scatter correction	Derivative	T	N	Mean	SD	SEC	R^2	SECV	R^2_{CV}	SEP	SLOPE	RSQ	RPD
CBDV	800-2500	SNV and DT	2,5,5,1	12	127	0.27	0.43	0.10	0.95	0.15	0.89	0.16	1.06	0.92	2.68
Δ9-THCV	1100-2500	MSC	2,5,5,1	12	127	0.05	0.06	0.02	0.92	0.02	0.83	0.03	1.15	0.87	1.84
CBD	1100-2500	SNV and DT	2,10,10,1	10	123	2.95	3.48	0.35	0.99	0.42	0.99	0.58	0.97	0.98	6.03
CBC	800-2500	SNV and DT	2,5,5,1	9	124	0.30	0.19	0.03	0.97	0.04	0.96	0.05	0.95	0.93	3.79
Δ8-THC	1100-2500	SNV and DT	1,5,5,1	9	118	0.13	0.16	0.03	0.97	0.03	0.96	0.07	0.89	0.85	2.40
Δ9-THC	1100-2500	MSC	1,10,10,1	11	122	3.30	5.28	0.58	0.99	0.77	0.98	1.72	1.00	0.90	3.07
CBG	800-2500	SNV and DT	1,5,5,1	8	120	1.04	0.98	0.25	0.94	0.28	0.92	0.79	0.67	0.54	1.25
CBN	800-2500	SNV and DT	2,5,5,1	7	122	0.12	0.10	0.02	0.95	0.03	0.91	0.05	0.94	0.76	2.09

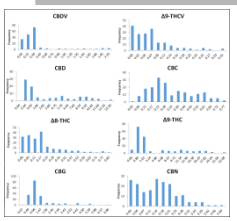
CBDV: cannabidivarin; Δ9-THCV: Δ9-tetrahydrocannabivarin; CBD: cannabidiol; CBC: cannabichromene; Δ8-THC: Δ8-tetrahydrocannabinol; Δ9-THC: Δ9-tetrahydrocannabinol; CBG: cannabigerol; CBN: cannabinol; SNV: standard normal variate; DT: detrending; MSC: multiplicative scatter correction; Derivative pre-treatment: the first digit is the number of the derivative, the second is the gap over which the derivative is calculated, the third is the number of data points in a running average or smoothing and the fourth is the second smoothing; T: number of PLS terms; Mean: mean of the calibration set; SD: standard deviation of the calibration set; SEC:

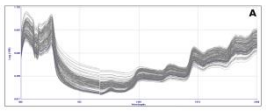
standard error of calibration; R^2: coefficient of determination of calibration; SECV: standard error of cross validation; r^2: coefficient of determination of cross validation; SEP: standard error of prediction; RSQ: coefficient of regression of reference data vs predicted data; RPD: ratio SD/SEP.

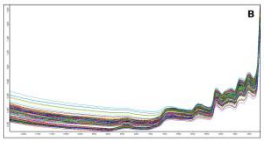
Table 3. Calibration and validation statistics of the selected equations obtained to predict the cannabinoid contents of samples analyzed in the FT-NIR instrument MPA.

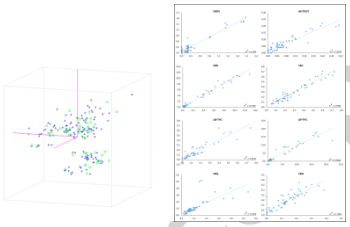
Cannabinoid	CALIBRATION										VALIDATION			
	REGION	SCATTER	Derivative	RANK	N	Mean	SD	RMSEC	R^2	RMSECV	R^2_{CV}	SEP	Slope	RSQ	RPD
CBDV	9403.7-8447.2; 6102-4242.9	none	2	11	119	0.24	0.38	0.13	0.89	0.17	0.81	0.21	0.86	0.93	2.71
Δ9-THCV	9403.7-5446.3;	none	2	10	123	0.06	0.06	0.02	0.89	0.03	0.77	0.04	0.73	0.86	1.95
CBD	7506-5446.3; 4428-4242.9	none	2	12	119	2.91	3.59	0.29	0.99	0.38	0.99	0.62	0.97	0.99	6.00
CBC	9403.7-4597.7	SNV	1	10	124	0.31	0.18	0.04	0.96	0.05	0.93	0.05	0.97	0.96	3.52
Δ8-THC	9403.7-7498.3; 6102-4242.9	none	2	10	117	0.13	0.15	0.02	0.98	0.03	0.96	0.07	0.93	0.91	2.38
Δ9-THC	9403.7-5446.3	none	1	12	119	3.23	5.29	0.49	0.99	0.62	0.99	1.79	0.88	0.95	3.09
CBG	7506-6796.3; 4428-4242.9	none	2	12	114	0.97	0.89	0.18	0.96	0.22	0.94	0.68	0.76	0.78	1.52
CBN	6102-5446.3; 4605.4-4242.9	MSC	1	10	112	0.12	0.09	0.02	0.96	0.02	0.95	0.06	0.83	0.83	1.71

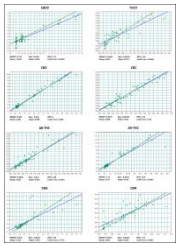
CBDV: cannabidivarin; Δ9-THCV: Δ9-tetrahydrocannabivarin; CBD: cannabidiol; CBC: cannabichromene; Δ8-THC: Δ8-tetrahydrocannabinol; Δ9-THC: Δ9-tetrahydrocannabinol; CBG: cannabigerol; CBN: cannabinol; SNV: standard normal variate; DT: detrending; MSC: multiplicative scatter correction; Derivative pre-treatment: the first digit is the number of the derivative, the second is the gap over which the derivative is calculated, the third is the number of data points in a running average or smoothing and the fourth is the second smoothing; RANK: number of PLS terms; Mean: mean of the calibration set; SD: standard deviation of the calibration set; RMSEC: root mean standard error of calibration; R^2: coefficient of determination of calibration; RMSECV: root mean standard error of cross validation; r^2: coefficient of determination of cross validation; SEP: standard error of prediction; RSQ: coefficient of regression of reference data vs predicted data; RPD: ratio SD/SEP.

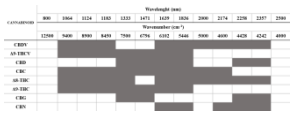












Highlights

- NIRS allowed predicting the cannabinoid content of *Cannabis sativa* L. in dry material.
- Comparison of FT-NIR and dispersive instruments showed similar results.
- Fast monitoring of dry plant material for practical field applications was possible.
- Cheaper characterization of *Cannabis sativa* L. chemotypes as compared to GC.
- Exponential future possibilities of applying NIRS techniques and protocols at industry level.