Industrial Crops & Products 187 (2022) 115349

Contents lists available at ScienceDirect

Industrial Crops & Products

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

Mid-infrared spectroscopy combined with multivariate analysis and machine-learning: A powerful tool to simultaneously assess geographical origin, growing conditions and bitter content in *Gentiana lutea* roots

Christian Coelho [a,*], Gilles Figueredo [b], Céline Lafarge [c], Elias Bou-Maroun [c], Stéphanie Flahaut [d]

[a] Université Clermont Auvergne, INRAE, VetAgro Sup campus agronomique de Lempdes, UMR F, F-15000 Aurillac, France
[b] LEXVA Analytique, 7 rue Henri Mondor, Biopôle Clermont Limagne, 63360 Saint Beauzire, France
[c] UMR PAM A 02.102, Université de Bourgogne Franche-Comté, L'institut Agro Dijon, 21000 Dijon, France
[d] CPPARM, ZA Les Quintrands, Route de Volx, 04100 Manosque, France

ARTICLE INFO

Keywords:
Gentiana lutea
Bitter composition
Mid-infrared spectra
Growing conditions
French mountains
Bitter content prediction
Multivariate analysis
Artificial neural networking

ABSTRACT

Mid-infrared spectroscopy was explored in order to evaluate its ability to classify geographical origin and bitter content of *Gentiana lutea* roots sourced from wild and cultivated growing conditions. Wild and cultivated *Gentiana lutea* roots from the four French mountains Massif Central, Jura, Alpes and Pyrénées were analyzed by Infrared spectroscopy and liquid chromatography. Unsupervised analyses assessed heterogeneity of *Gentiana lutea* roots in the different sampling sites due to their evolutive composition along plant growth. Predictive models using partial least squares discriminant analysis and probabilistic artificial neural network were discussed according to FTIR spectral regions and gave 100 % accuracy in authentication of geographical origin and growing conditions. Nevertheless, classification of gentian roots according to their bitter content, based on FTIR spectral signatures, was challenging due to their biological heterogeneity. We propose an unprecedented classification of *Gentiana lutea* roots according to their analyzed bitter content: Low, Medium and High, comprised in the range [6–8] %, [8–10] % and [10–12] % in dry weight, respectively. FTIR coupled with chemometrics applied directly on gentian roots allowed for a decent level of predictability by PLS-DA (Q2Cum = 0.41) and Artificial Neural Network (89.1 % accuracy), when using the (650–1800 cm^{-1}) infrared region.

1. Introduction

Gentiana lutea is a perennial plant species belonging to the *Gentianaceae* family of yellow gentians that grows naturally in the medium-altitude mountains (800–2500 m), and is present all around the world (González-López et al., 2014; Grabherr, 2009; Marković et al., 2019; Takahashi et al., 2014). *Gentiana lutea* contributes to the natural herbal biodiversity and landscape beauty, particularly during the flowering season (Grabherr, 2009). During its dormancy, *Gentiana lutea* also contributes to the sub-surface biodiversity thanks to its rhizomes that grow year after year, by accumulating numerous metabolites of interest, particularly secoiridoidal and iridoid glycosides, xanthones, monoterpene alkaloids; polyphenols, flavones and volatile compounds (Ando et al., 2007; Arberas et al., 1995; Ariño et al., 1997; Takahashi et al., 2014; Toriumi et al., 2003). Some of these biochemical compounds are currently of interest in the following sectors: herbal medicines, animal nutrition, agri-food, cosmetic applications, and more recently in agroecology (EFSA, 2005; EMA, 2009; Biehlmann et al., 2020; Franz Vienna et al., 2007; Mi et al., 2019; Park et al., 2018). In order to supply these human activities, *Gentiana lutea* is being cultivated (Marković et al., 2019; Mustafa et al., 2015).

Nevertheless, it has been shown that its composition, particularly its bitter compounds and volatile compounds varied depending on the environment in which they were grown: wild vs cultivated environments (Mustafa et al., 2016, 2015). Its wild harvesting by uprooting yellow gentians in wild environments should be accompanied by eco-friendly and sustainable management policies in these natural areas (Cambecedes et al., 2018). In that sense the French interprofessional association *Gentiana lutea* recently suggested harmonizing sustainable actions concerning its use, from the field to the final gentian-based

* Correspondence to: VetAgro Sup campus agronomique de Clermont, UMR F INRAE - Université Clermont Auvergne – VetAgro Sup, 89 avenue de l'Europe, BP 35, F-63370 Lempdes, France.
E-mail address: christian.coelho@vetagro-sup.fr (C. Coelho).

https://doi.org/10.1016/j.indcrop.2022.115349
Received 1 April 2022; Received in revised form 3 July 2022; Accepted 8 July 2022
Available online 19 July 2022

Table 1
Characteristics of *Gentiana lutea* roots samples collected from the four French mountains. GPS represents the Global Positioning Unit, Fresh weight represents the minimum and maximum of fresh root weighted on-field, N represents the number of uprooted gentians, WIL/CUL represents the wild vs cultivated growing practices.

Geographical origin	Sites	GPS coordinates	Altitude (m)	Fresh weight (g)	Number of samples (N)	Growing practices
Massif Central (MC)	A: Fraux	N: 45°02.971' E: 2°52.45'	1200	455–1155	8	WIL
	B: Liorangues	N: 45°.571' E: 3°38.723'	888	446–1120	5	CUL
	C: Malbo	N: 44°58.709' E: 2°45.921'	1220	438–1840	5	WIL
	D: Nasbinals	N: 44°40.736' E: 3°02.397'	1130	606–1947	5	WIL
	E: Gelles	N: 45°45.49' E: 2°44.970'	900	200–1000	5	CUL
	F: Pégrol	N: 45°36.586' E: 3°51.645'	1330	499–867	5	WIL
	G: Picherande	N: 45°28.512' E: 2°50.89'	1250	513–862	5	WIL
Jura (NMC)	H: La Chapelle des Bois	N: 46°37.481' E: 6°8.5′	1090	818–1038	5	WIL
Pyrénées (NMC)	I: Bagnères de Luchon	N: 42°44.430' E: 0°38.946'	1470	546–857	6	WIL
Alpes (NMC)	J: Samoens	N: 46°12.477' E: 6°28.920'	1340	488–737	6	WIL

product by creating a French label "Gentiane - Filière Développement Durable" (INPI, 2021). In that perspective, mapping the potential resource in relation to its composition and particularly its bitter composition, appeared crucial for facilitating adhesion, confidence, and organization of the gentian network. Analytical and chemical procedures could be implemented for that purpose, but they require prior treatment and separation of the bioactive compounds from the root samples at the laboratory scale (Aberham et al., 2011; Biehlmann et al., 2020; Carnat et al., 2005; Mustafa et al., 2015). The possibility to use spectroscopic measurements directly on gentian roots attracted our attention in this research study. UV–visible, near and mid-Infrared spectroscopies have already been used for herbal recognition (Bunaciu et al., 2011; Yang et al., 2014) and particularly with gentian species (Chuang et al., 2013; Mi et al., 2019; Shen et al., 2020). The potentiality of using infrared spectroscopy coupled with chemometric techniques has been recently proposed for quality control and geographical authentication of other botanical plants for medicinal and food purposes (Wang et al., 2021, 2018). Recently, infrared spectroscopy was combined with machine-learning analysis, which improved the potential of classification regarding biological samples with high variability (Abraham and Kellogg, 2021; Enders et al., 2021; Oliveira et al., 2021).

These chemometric-guided approaches have recently been reviewed in the area of botanical materials (Abraham and Kellogg, 2021) and have been implemented in this study dedicated to *Gentiana lutea* rhizomes. The objectives are to evaluate the impact of the geographical origin, to evaluate growing practices and to propose a classification of gentian roots based on their bitter content that could be directly applicable to the field.

2. Materials and methods

2.1. Gentian roots samples

A total of 55 *Gentiana lutea* roots were sampled entirely in July 2018 from ten different sites in four French mountains (Massif Central, Jura, Pyrénées and Alpes). In each site, a minimum of five gentian roots were uprooted with a "devil's pitchfork" in order to consider the biochemical variability of the sampling site. Two of these sites were two gentian cultivars (CUL) (N = 10) and the eight others were wild sites (WIL) (N = 45). Table 1 summarizes the number of sampled gentian roots used in this study. Due to the disproportionate number of gentian roots sampling sites between French mountains, it was decided to compare Massif Central (MC) (N = 38) vs Non-Massif Central (NMC) including Jura, Pyrénées and Alpes mountains (N = 17) for the comparison of geographical origin. The dirt was removed from all the roots, and they were manually sliced into 1–2 cm pieces, dried at 40 °C for two days and finally ground to a fine powder and stored at 4 °C until physico-chemical analysis. For bitter content analysis, gentian roots were extracted at (weight/volume) ratio of 10 g / 1 L of methanol. For Mid Infrared analysis by ATR-FTIR, dry powders were used in state.

2.2. Bitter content analysis

Bitter components were quantitatively determined by liquid chromatography using a LC-DAD (Acquity, Waters) device with a Raptor ARC C18 column (150 × 2.1 mm, particle size of 1.8 μm) and a mobile phase made of (ultrapure water/acetonitrile/formic acid) with (84.9/15/0.1),v/v/v) at constant flow rate of 0.2 mL.min^{-1}. Volume injection was set at 1 μL. Amarogentin, gentiopicroside, loganic acid, swertiamarine and sweroside, supplied by Extrasynthese (Genay, France) at the highest purity grade, were dissolved in methanol at 1 g.L^{-1} for stock solution preparation, stored at 4 C. They were respectively detected at 305, 280, 260, 260 and 260 nm. Retention times of pure standards were 6.0, 6.7, 7.1, 7.4 and 9.4 min respectively for loganic acid, swertiamarine, gentiopicroside, sweroside and amarogentin. UV spectra comparisons with pure standards were used for identification. Quantification was adapted from a previous published methodology (Biehlmann et al., 2020) and was both validated by an external calibration from 1 to 100 mg.L^{-1} and a standard addition of the mix of commercial standards from 1 to 100 mg.L^{-1} to gentian root methanolic extracts. Analyses were performed in technical triplicates. The content of individual bitter compounds was expressed in percent by weight (% wt) of dried gentian root. Gentian root bitter content was calculated by the sum of the mass percentages of the overall analyzed bitter compounds.

2.3. Mid Infrared analysis by ATR-FTIR

Fourier-Transform InfraRed (FTIR) spectra were acquired by attenuated total reflection (ATR) using a Thermo Scientific Nicolet iS5. All gentian root powders were gently placed on the surface of the diamond crystal and pressed with a constant force with a torque screwdriver. Each spectrum was recorded from 500 to 4000 cm^{-1} with a 4 cm^{-1} resolution resulting from the accumulation of 32 scans.

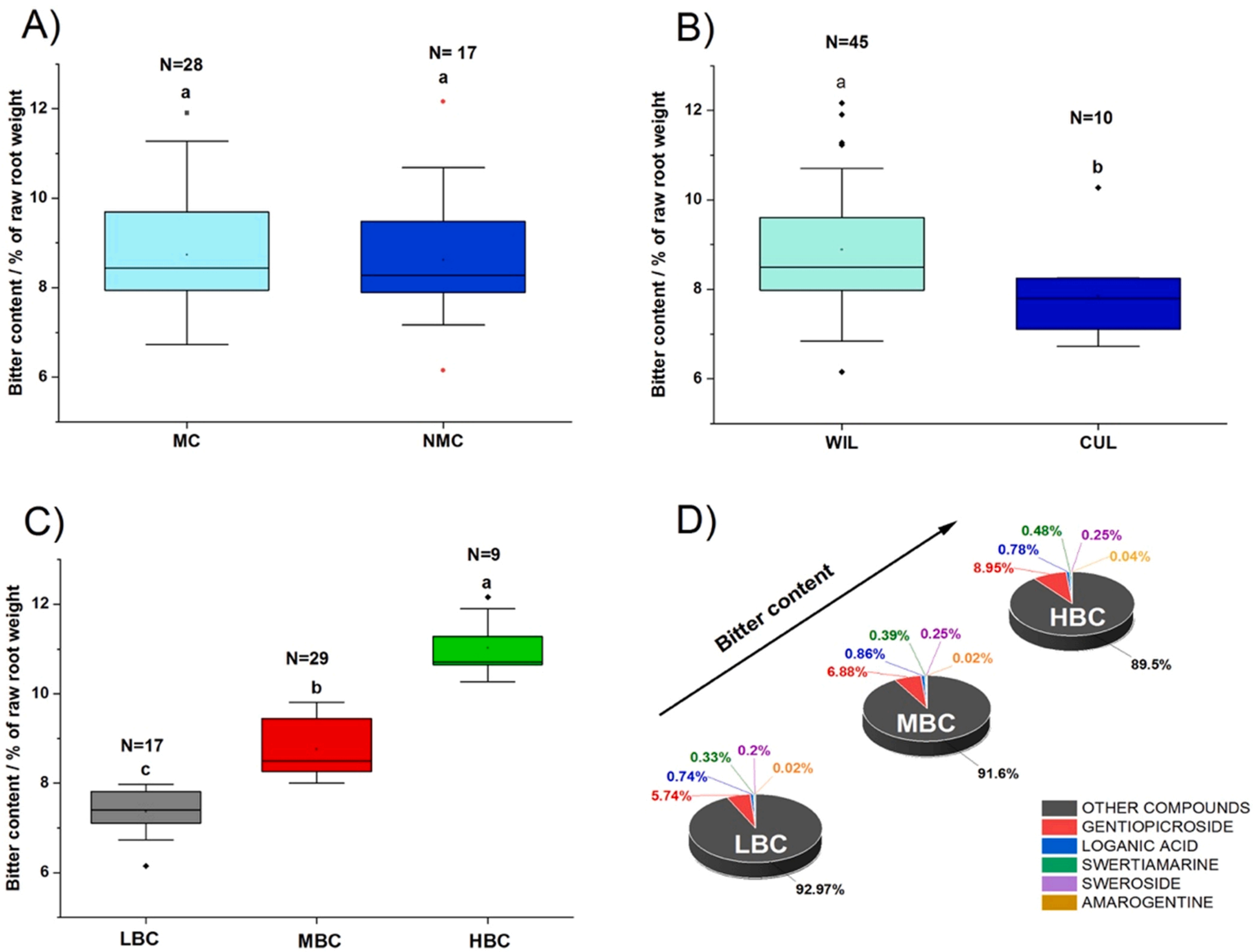

Fig. 1. Box plot of bitter content of gentian roots comparing geographical origin: Massif central (MC) vs Non-Massif Central (NMC) (A), growing conditions: Wild (WIL) vs Culture (CUL) (B), the different classes according to Low Bitter content (LBC), Medium Bitter content (MBC) and High Bitter content (HBC). Pie chart representation of the contribution of gentiopicroside, loganic acid, swertiamarine, sweroside and amarogentine to the overall raw gentian root weight, following the proposed classification according to their bitter content. On top of each boxplot, the letters represent a significant difference (p-value <0.05, ANOVA and Fisher's least significant difference post-hoc test) between gentian groups. The number N indicates the number of gentian roots used for the representation.

2.4. Data preprocessing

All the 55 raw FTIR spectra were preprocessed on OriginPro 9.0 software (Originlab, Massachussets, USA) using the 2nd derivative and the Savitzky-Golay smoothing using a quadratic polynomial order with 15 points in each sub model. This novel dataset was analyzed by classification modeling using gentian roots growing conditions (CUL/WIL), geographical origin (MC/NMC) and bitter content: Low Bitter Content (LBC) / Medium Bitter Content (MBC) / High Bitter Content (HBC) as the variables to predict with Partial Least Square-Discriminant Analysis (PLS-DA) and Artificial Neural Networking (ANN). These two models were chosen based on their predictive potential in terms of plant geographical authentication, culture practices and chemical composition (Abraham and Kellogg, 2021; Enders et al., 2021; Oliveira et al., 2021). Three ranges of FTIR wavenumbers were investigated in this study: 650–1800 & 2600–3700 cm^{-1}, 2600–3700 cm^{-1} and 650–1800 cm^{-1}, representing 6400, 4700 and 2400 datapoint values, respectively.

2.5. Classification modeling by multivariate analysis and machine-learning

SIMCA 17 software (Sartorius, Göttingen, Germany) was used for PLS-DA. A calibration model was evaluated by the coefficient of determination (R2X), the proportion of the variance of the response was explained by the generated model (R2Y) and the predictive values (Q2Y). Validation was evaluated by realizing permutation tests and by analyzing the values of Root Mean Squared Error of Prediction (RMSEP).

NeuralNetwork Toolbox v.4 from Matlab 2019b software (The MathWorks Natick, Massachussets, USA) was used to test several combinations of ANN models to evaluate our predictors and visualize the confusion matrix. All ANN models were built with 30 neurons with sigmoidal function in the hidden layer and a linear function in the output layer. The performance (ρ_i) of the model was estimated by the cross-entropy of the model and the lowest numbers of epoch enabling to stabilize quickly the generated ANN model (Murphy, 2012). In the following Eq. (1): d is the length of the test and $\widehat{y}_{(k)}$ is the estimated label found by the classes using the cross-validation procedure. The ρ_i is the mean of accuracies obtained in the cross-validation steps.

$$\rho_i = \frac{1}{d} \sum_{k=1}^{d} \delta(k), \text{with} \delta(k) \left\{ \begin{array}{ccc} 1 & \widehat{y}_{(k)} = & y_{(k)} \\ 0 & \widehat{y}_{(k)} \neq & y_{(k)} \end{array} \right\} \quad (1)$$

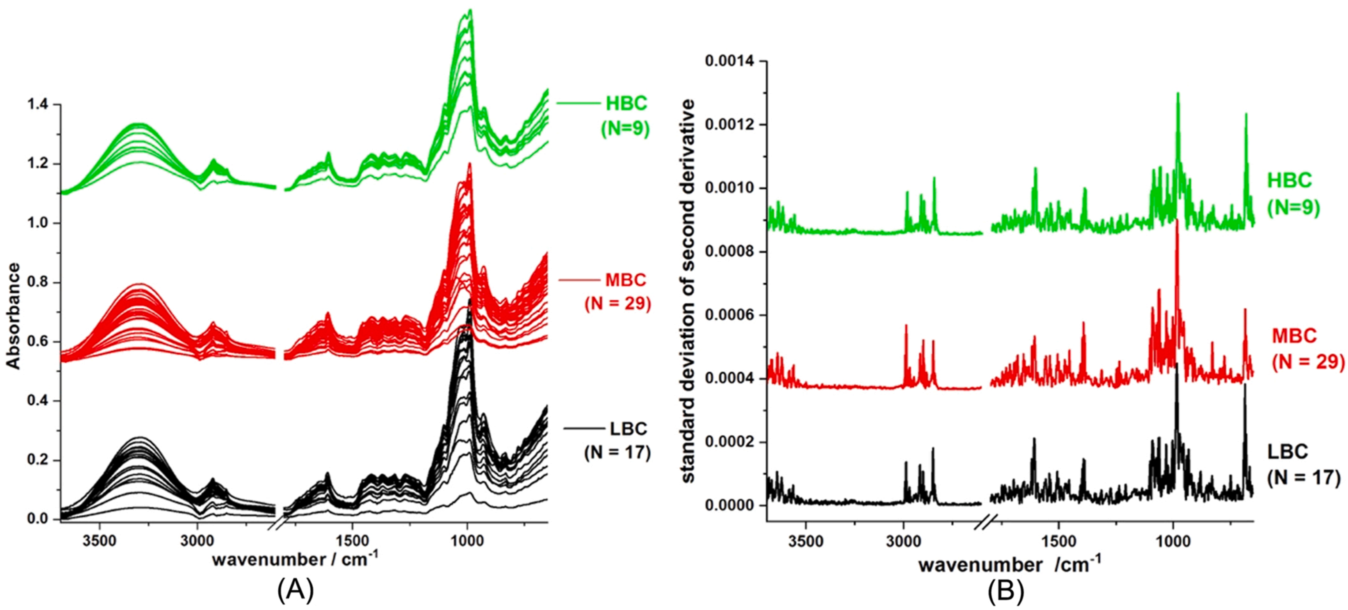

Fig. 2. (A) ATR-FTIR spectra and (B) the standard deviation of second derivative of the mean ATR-FTIR spectrum of gentian roots classified upon their bitter content: Low (LBC), Medium (MBC) and High (HBC), in black, red and green color, respectively.

3. Results and discussion

3.1. Bitter content in Gentiana lutea roots

The bitter content, representing the sum of the mass percentages of gentiopicroside, loganic acid, swertiamarine, sweroside and amarogentine are presented in Fig. 1. Gentian bitter content presents a similar mean value of (8.73 ± 1.32 %) and (8.61 ± 1.44 %) for MC and NMC gentian roots (Fig. 1A). Nevertheless, as illustrated by the elevated standard deviations, the heterogeneity among the sampled gentian roots was particularly high. As presented in Fig. S.I.1, the main reason of that observation could be due to the elevated variability of the biological roots sampled in a same geographical site. Such results could also be attributed to different physiological parameters of gentian roots growth (particularly their age), genetic variations among *Gentiana lutea* species, and the gentian secondary metabolism pathways leading to the production of seco-iridoids compounds (Franz et al., 1996; González-López et al., 2014; Rossi et al., 2016; Takahashi et al., 2014). In that sense, the impact of growing practices revealed significative differences in the bitter content of gentian roots (Fig. 1B). Gentian roots present a higher mean value of bitter content under wild growing conditions, (8.89 ± 1.35 %), compared to cultivated areas, (7.84 ± 1.00 %). Such observation is in accordance with previous published data (Mustafa et al., 2016, 2015). Gentiopicroside is the main contributor to such differentiation observed between wild and cultivated growing conditions (Table S.I.1). Taking into consideration this chemical heterogeneity, based on gentian geographical origin and growing conditions, it was decided to voluntarily classify them according to their bitter content. For that, three new classes of gentian roots were created as follows: Low Bitter Content (LBC), Medium Bitter Content (MBC) and High Bitter Content (HBC), based on the measured mass percentages of bitter content ranging in [6–8] %, [8–10] % and [10–12] % in dry weight, respectively (Fig. 2C). Fig. 2D highlights the mass percentages of individual bitter compounds among LBC, MBC and HBC. Gentiopicroside, representing the main bitter compound for the three typologies of gentian roots, increased with the overall bitter content of the classified LBC, MBC and HBC gentian roots from 5.74 % to 6.88 % and 8.95 % in dry weight, respectively. Classifying the gentian roots by means of their total bitter content increased significantly (p-value <0.05) the individual bitter compounds (gentiopicroside, swertiamarine, and amarogentine), except loganic acid and sweroside presenting not statistically distinguishable differences among pre-defined gentian bitter classes (Table S.I.2).

3.2. FT-IR spectra of gentian roots

In order to evaluate the possibility to use FTIR to differentiate bitter content in gentian roots regardless of their geographical origin and growing conditions, we classified the measured FTIR spectra of *Gentiana lutea* roots according to their bitter content (LBC, MBC and HBC). Fig. 2A shows the mid-infrared spectra of the gentian roots analyzed according to their bitter content. No obvious difference was observed at first glance between the three typologies of gentian roots. All gentian roots presented numerous absorption bands between 3700 and 650 cm^{-1}, particularly those corresponding to the O-H band centered at 3300 cm^{-1}. This can surely be attributed to the water content of the gentian roots still present after the drying process and to the presence of numerous alcohol functions of sugar units and seco-iridoids in gentian roots. The numerous peaks located between 2920 and 2850 cm^{-1} correspond to the C-H aliphatic functions. The absorption peaks centered at 985, 1000 and 1060 cm^{-1} correspond to the C-O stretching vibration of glycosidic linkages present among the cell walls polysaccharides and the abundant iridoids, particularly gentiopicroside (Mi et al., 2019; Yang et al., 2014). The absorption band centered at 1600 cm^{-1} could be attributed to the C=C stretching aromatic moieties present in gentian roots. Fig. 2B shows the standard deviation for each wavenumber for the three classes of gentian roots according to their bitter content. Fig. 2B offers a more precise visualization of the predominant peaks presenting a high standard deviation, largely distributed in the spectral range [1800–650 cm^{-1}]. Main absorption peaks at 682–689 cm^{-1}, 971, 980–990, 1004, 1061, 1066, 1091, 1606 cm^{-1} are revealed by this spectral treatment but with no clear differences among bitter classes at first glance. In order to elucidate the relations existing between spectral data and bitter content of gentian roots, multivariate and machine-learning were further carried out.

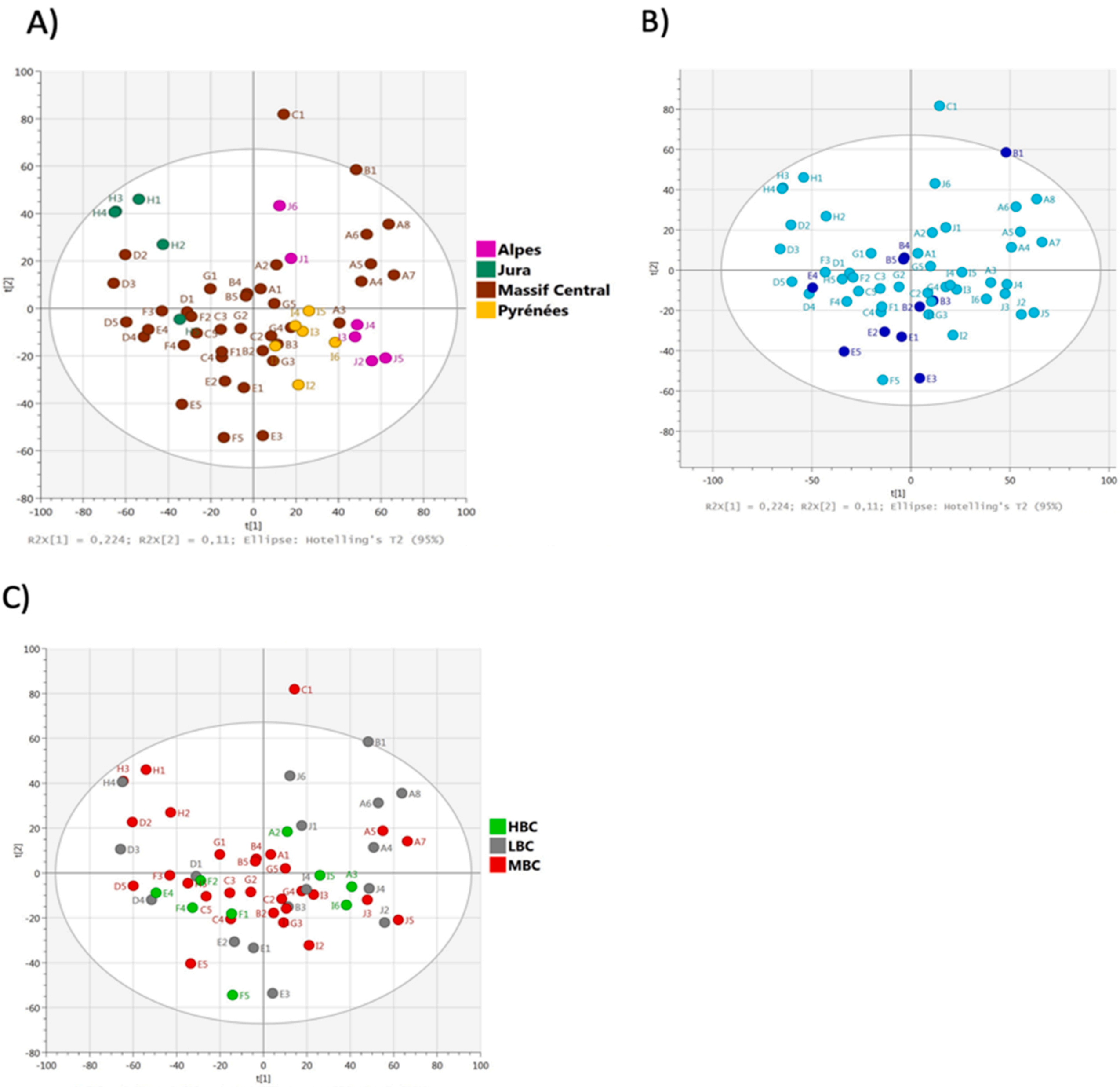

Fig. 3. Scatterplot obtained after principal component analysis on the second derivative and Savitzky-Golay smoothed FTIR spectrum of the N = 55 *Gentiana lutea* roots according to their geographical origin (A), growing conditions (B) and bitter content (C).

3.3. Unsupervised multivariate analysis carried out on gentian FTIR spectral features confirms gentian specificities associated to geographical origin

Unsupervised analyses on the second derivative and Savitzky-Golay smoothed FTIR spectra were undertaken to investigate the diversity of chemical composition of the gentian roots (N = 55). Fig. 3 shows the results of the scatterplot of the first two principal components of the Principal Component Analysis (PCA) carried out in regards of their growing conditions (CUL/WIL), geographical origin (MC/NMC) and bitter content (LBC/MBC/HBC). The two first components explain only 33.4 % of the total variability of the entire dataset. Fig. 3A indicates the possibility to authenticate gentian roots from NMC sites by differentiating the five biological replicates from Jura, Alpes and Pyrénées. Massif Central (MC) sites present a more heterogeneous distribution in the scatterplot, meaning a higher diversity in its chemical composition. Such diversity could be attributed to a higher diversity in its bitter composition, as illustrated in Fig. S.I.1. When looking at Fig. 3B and Fig. 3C, the PCA model with two principal components failed to correctly differentiate gentian roots based on growing conditions and bitter content classes, respectively. Such results confirm the high variability in the chemical composition of the biological samples of gentian roots and motivate the use of supervised approaches that integrate regression algorithms presenting linear and/or non-linear latent structures.

3.4. Possibility to predict gentian bitter contents by FTIR spectral features by PLS-DA and Artificial Neural network

Partial Least Square Discriminant Analysis (PLS-DA) and Artificial Neural Network (ANN) were carried out by attempting to associate FTIR measurements on *Gentiana lutea* rhizomes with their geographical origin, growing conditions and their bitter content. The results of the different linear and non-linear models using three different infrared

Table 2
Results of the classification modeling by means of Partial Least Square Discrimination Analysis (PLS-DA) and artificial neural networking (ANN).

Attempt of gentian root classification based on:	FTIR region / cm^{-1}	PLS-DA (models with 2 components)			ANN (models with 30 neurons)		
		R2X	Q2Cum	RMSEP	performance	gradient	% accuracy
Geographical origin	650–1800 & 2600–3700	0.5560	0.075	0.327	1.3E-6	5,5E-7	100 %
	2600–3700	0.401	0.162	0.376	0.17	3.0E-3	98,2 %
	650–1800	0.540	0.080	0.333	0.33	2.8E-2	96,4 %
Growing conditions	650–1800 & 2600–3700	0.519	0.426	0.263	5.6E-3	4.5E-4	100 %
	2600–3700	0.151	0.030	0.349	3.1E-2	6.9E-5	100 %
	650–1800	0.517	0.431	0.263	8.7E-7	6.6E-7	98.2 %
Bitter content	650–1800 & 2600–3700	0.530	0.361	0.702	0.24	8.6E-3	85.5 %
	2600–3700	0.252	0.056	1.007	0.35	0.17	58.2 %
	650–1800	0.556	0.413	0.632	0.26	0.11	89.1 %

regions are presented in Table 2.

First of all, as expected by the previous non-supervised analysis (Fig. 3A), the geographical origin of gentian roots could be modelized via PLS-DA and ANN, as visualized by the elevated values of the coefficient of determination (R2X) and the performance values, respectively. Nevertheless, the predictive ability (Q2Cum comprised between 0.075 and 0.162 for the three investigated FTIR regions) is very low, showing that gentian geographical origin modeling is not linear and is improved when using an ANN model, leading to a 100 % accuracy when analyzing the entire region of the FTIR spectra. The two others investigated FTIR regions lead to satisfactory accuracies above 95 % with the non-linear ANN models. Undoubtedly, ANN model, considering the overall FTIR region (650–1800 & 2600–3700 cm^{-1}), predicts better gentian geographical origin than the PLS-DA model.

Concerning the classification of gentians based on growing conditions, the PLS-DA linear model leads to very good predictive classifications, particularly on the entire infrared spectra (650–1800 & 2600–3700 cm^{-1}) and on the 650–1800 cm^{-1} region, as illustrated by the Q2Cum values: 0.426 and 0.431, respectively. Interestingly, the region 2600–3700 cm^{-1} is not linearly correlated to the growing conditions (Q2Cum = 0.03) whereas its non-linear modeling via ANN model is totally satisfactory (100 % accuracy). The non-linear ANN models with the two other regions could also nicely predict cultivated gentian roots growing conditions.

Finally, the linear and the non-linear modeling of gentian bitter content is totally appropriate in the 650–1800 cm^{-1} FTIR region, leading respectively to elevated predictive ability (Q2Cum = 0.413) and 89.1 % of global accuracy, respectively. Concerning the two others investigated FTIR spectral regions, the PLS-DA models lead to poorer results with lower predictive ability and higher RMSEP. The ANN models were not satisfactory in the region 2600–3700 cm^{-1} but the entire FTIR region (650–1800 & 2600–3700 cm^{-1}) presents a lower but good accuracy (85.5 %), with a similar performance of 0.24 and even better gradient of 8.6E-3, compared to the 650–1800 cm^{-1} FTIR region. When looking closely at the generated ANN models, as shown in the confusion matrices (Fig. S.I.2), it appears that HBC gentian roots (N = 9) were properly classified with 100 % accuracy while the group LBC/MBC presented 90 % accuracy in its classification. Such results highlighted that our pre-defined classification based on the analysis of gentian bitter content could lead to artificial biases on the predictive algorithms used by machine-learning (Mehrabi et al., 2021). The developed predictive models on gentian root bitterness based on a three-class categorization (LBC, MBC and HBC) could be satisfactorily implemented with both PLS-DA and ANN using the gentian root (650–1800) cm^{-1} FTIR region.

To conclude, optimized linear modeling of gentian FTIR spectra could be developed in the future to evaluate gentian growing conditions and gentian bitter content. Non-linear modeling is more suited for gentian geographical authentication purposes. Such results indicate that FTIR spectra coupled with chemometrics are reliable tools dedicated to qualitative and quantitative diagnostics of *Gentiana lutea* resources that consider their complex chemical composition and their biological variability. Such results could be extended to other botanical plants and could be used in future for helping in the sustainable management of endangered natural resources, based on their geographical origin, growing conditions and bioactive compounds content.

4. Conclusions

This study aims to evaluate the use of mid-infrared spectroscopy combined with multivariate and machine-learning analysis on *Gentiana lutea* roots sourced from cultivated and wild areas located in the French mountains (Massif Central, Jura, Pyrénées, Alpes). As expected, geographical authentication and plant-growing conditions are rather well differentiated among the 55 gentian samples and are discussed according to their chemical composition. Their bitter content was chemically analyzed and enabled us to propose an unprecedented bitter content classification that could be reliably predictable using FTIR spectroscopy and chemometrics, regardless of gentian geographical origin and growing conditions: Low (LBC), Medium (MBC) and High (HBC) gentians contained [6–8] %, [8–10] % and [10–12] % in dry weight, respectively.

This study paves the way for future development on implementable infrared tools that could be used by gentian actors to estimate bitter content on gentian roots and that could help in sustainable management on natural resource policies and uses.

CRediT authorship contribution statement

Christian Coelho: Writing, Data analysis, Software, Conceptualization, Funding acquisition. **Gilles Figueredo**: Methodology, Review & Editing. **Céline Lafarge**: Writing – Review & Editing, **Elias Bou-Maroun**: Writing – Review & Editing. **Stéphanie Flahaut**: Methodology, Funding acquisition, Review & Editing.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

Acknowledgments

We thank our financial supporters: Agence Nationale de la Cohésion des Territoires (ANCT) and the Fonds national d′aménagement et de développement du territoire (FNADT). We are very grateful to the French *Gentianalutea* association and to the farmers who allowed us to sample *Gentiana lutea* roots in their fields. We finally thank J. Billon for carefully proof-reading and editing the manuscript.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.indcrop.2022.115349.

References

Aberham, A., Pieri, V., Croom Jr., E.M., Ellmerer, E., Stuppner, H., 2011. Analysis of iridoids, secoiridoids and xanthones in *Centaurium erythraea*, *Frasera caroliniensis* and *Gentiana lutea* using LC-MS and RP-HPLC. J. Pharm. Biomed. Anal. 54, 517–525. https://doi.org/10.1016/j.jpba.2010.09.030.

Abraham, E.J., Kellogg, J.J., 2021. Chemometric-guided approaches for profiling and authenticating botanical materials. Front. Nutr. 8 https://doi.org/10.3389/fnut.2021.780228.

Ando, H., Hirai, Y., Fujii, M., Hori, Y., Fukumura, M., Niiho, Y., Nakajima, Y., Shibata, T., Toriizuka, K., Ida, Y., 2007. The chemical constituents of fresh Gentian Root. J. Nat. Med. 61, 269–279. https://doi.org/10.1007/s11418-007-0143-x.

Arberas, I., Leiton, M.J., Domínguez, J.B., Bueno, J.M., Ariño, A., de Diego, E., Renobales, G., de Renobales, M., 1995. The volatile flavor of fresh *Gentiana lutea L.* Roots. In: Charalambous, G. (Ed.), Developments in Food Science, Vol. 37. Elsevier, pp. 207–234. https://doi.org/10.1016/S0167-4501(06)80158-5.

Ariño, A., Arberas, I., Leiton, M.J., de Renobales, M., Dominguez, J.B., 1997. The extraction of yellow gentian root (*Gentiana lutea L.*). Z. Lebensm. -Forsch. A 205, 295–299. https://doi.org/10.1007/s002170050168.

Biehlmann, M., Nazaryan, S., Krauss, E., Ardeza, M.I., Flahaut, S., Figueredo, G., Ballester, J., Lafarge, C., Bou-Maroun, E., Coelho, C., 2020. How chemical and sensorial markers reflect gentian geographic origin in chardonnay wine macerated with *Gentiana lutea* roots? Foods 9. https://doi.org/10.3390/foods9081061.

Bunaciu, A.A., Aboul-Enein, H.Y., Fleschin, S., 2011. Recent applications of fourier transform infrared spectrophotometry in herbal medicine analysis. Appl. Spec. Rev. 46, 251–260. https://doi.org/10.1080/05704928.2011.565532.

Cambecedes, J., Garreta, G., Gire, L., Morisson, B., Garcia, J., Durand, B., 2018. Exploiter et préserver: vers un plan de gestion durable de la gentiane jaune dans les Pyrénées. CBNPMP, p. 110.

Carnat, A., Fraisse, D., Carnat, A.-P., Felgines, C., Chaud, D., Lamaison, J.-L., 2005. Influence of drying mode on iridoid bitter constituent levels in gentian root. J. Sci. Food Agric. 85, 598–602. https://doi.org/10.1002/jsfa.2011.

Chuang, Y.-K., Chen, S., Lo, Y.M., Yang, I.C., Cheng, Y.-F., Wang, C.-Y., Tsai, C.-Y., Hsieh, R.-M., Wang, K.-H., Lai, C.-C., Chen, W.-C., 2013. Quantification of bioactive gentiopicroside in the medicinal plant *Gentiana scabra* Bunge using near infrared spectroscopy. J. Food Drug Anal. 21, 317–324. https://doi.org/10.1016/j.jfda.2013.07.011.

Enders, A.A., North, N.M., Fensore, C.M., Velez-Alvarez, J., Allen, H.C., 2021. Functional group identification for FTIR spectra using image-based machine learning models. Anal. Chem. 93, 9711–9718. https://doi.org/10.1021/acs.analchem.1c00867.

European Food Safety Authority (EFSA), 2005. Study on the Assessment of plants/herbs, plant/herb Extracts and Their Naturally or Synthetically Produced Components As "additives" for Use in Animal Production, CFT/EFSA/FEEDAP/2005/01."

European Medicines Agency (EMA), 2009. Assessment Report on *Gentiana lutea L.*, Radix, Doc. Ref.EMA/HMPC/578322/2008."

Franz, C., Bezzi, A., Belliardo, F., 1996. Variability and breeding of *Gentiana lutea L.* with respect to bitter substances. L'uomo e l'ambiente.

Franz Vienna, C., Bauer Graz, R., Carle Hohenheim, R., Tedesco Milano, D., Tubaro Trieste, A., Zitterl-Eglseer Wien, K., 2007. "Study on the assessment of plants/herbs, plant/herb extracts and their naturally or synthetically produced components as 'additives' for use in animal production. European Food Safety Authority (EFSA)."

González-López, O., Polanco, C., György, Z., Pedryc, A., Casquero, P.A., 2014. Genetic variation of the endangered *Gentiana lutea L.* var. aurantiaca (Gentianaceae) in populations from the Northwest Iberian Peninsula. Int. J. Mol. Sci. 15, 10052–10066. https://doi.org/10.3390/ijms150610052.

Grabherr, G., 2009. Biodiversity in the high ranges of the Alps: ethnobotanical and climate change perspectives. Glob. Environ. Change 19, 167–172. https://doi.org/10.1016/j.gloenvcha.2009.01.007.

Institut National de la propriété Industrielle (INPI), 2021. ⟨https://data.inpi.fr/marques/FR4789308?q=Gentiane%20#⟩FR4789308. Marque collective " Gentiane – filière développement durable ".

Marković, T., Radanović, D., Nastasijević, B., Antić-Mladenović, S., Vasić, V., Matković, A., 2019. Yield, quality and safety of yellow gentian roots produced under dry-farming conditions in various single basal fertilization and planting density models. Ind. Crops Prod. 132, 236–244. https://doi.org/10.1016/j.indcrop.2019.02.027.

Mehrabi, N., Morstatter, F., Saxena, N.A., Lerman, K., Galstyan, A.G., 2021. A Survey on Bias and Fairness in Machine Learning. ACM Comput. Surv. 54, 1–35. https://doi.org/10.1145/3457607.

Mi, L.-J., Zhang, J., Zhao, Y.-L., Zuo, Z.-T., Wang, Y.-Z., Li, F.-S., 2019. Discrimination and evaluation Gentiana rigescens–Camellia sinensis with different planting year using Fourier transform infrared spectroscopy. Agrofor. Syst. 93, 1157–1166. https://doi.org/10.1007/s10457-018-0216-8.

Murphy, K.P. Machine Learning: A Probabilistic Perspective (Adaptive Computation and Machine Learning series). Ed. The MIT Press. 2012.

Mustafa, A.M., Caprioli, G., Maggi, F., Vittori, S., Sagratini, G., 2016. Comparative Analysis of the Volatile Profiles from Wild, Cultivated, and Commercial Roots of *Gentiana lutea L.* by Headspace Solid Phase Microextraction (HS–SPME) Coupled to Gas Chromatography Mass Spectrometry (GC–MS). Food Anal. Method 9, 311–321. https://doi.org/10.1007/s12161-015-0196-5.

Mustafa, A.M., Caprioli, G., Ricciutelli, M., Maggi, F., Marín, R., Vittori, S., Sagratini, G., 2015. Comparative HPLC/ESI-MS and HPLC/DAD study of different populations of cultivated, wild and commercial *Gentiana lutea L.* Food Chem. 174, 426–433. https://doi.org/10.1016/j.foodchem.2014.11.089.

Oliveira, I.C., Franca, T., Nicolodelli, G., Morais, C.P., Marangoni, B., Bacchetta, G., Milori, D.M.B.P., Alves, C.Z., Cena, C., 2021. Fast and accurate discrimination of *Brachiaria brizantha* (A.Rich.) stapf seeds by molecular spectroscopy and machine learning. ACS Agric. Sci. Technol. 1, 443–448. https://doi.org/10.1021/acsagscitech.1c00067.

Park, E., Kim, J., Yeo, S., Kim, G., Ko, E.H., Lee, S.W., Li, W.Y., Choi, C.W., Jeong, S.Y., 2018. Antiadipogenic effects of loganic acid in 3T3-L1 preadipocytes and ovariectomized mice. Molecules 23. https://doi.org/10.3390/molecules23071663.

Rossi, M., Fisogni, A., Galloni, M., 2016. Biosystematic studies on the mountain plant *Gentiana lutea L.* reveal variability in reproductive traits among subspecies. Plant Eco. Div. 9, 97–104. https://doi.org/10.1080/17550874.2015.1074625.

Shen, T., Yu, H., Wang, Y.-Z., 2020. Discrimination of *Gentiana* and its related species using IR spectroscopy combined with feature selection and stacked generalization. Molecules 25, 1442. https://doi.org/10.3390/molecules25061442.

Takahashi, H., Imamura, T., Konno, N., Takeda, T., Fujita, K., Konishi, T., Nishihara, M., Uchimiya, H., 2014. The gentio-oligosaccharide gentiobiose functions in the modulation of bud dormancy in the herbaceous perennial *Gentiana*. Plant Cell 26, 3949–3963. https://doi.org/10.1105/tpc.114.131631.

Toriumi, Y., Kakuda, R., Kikuchi, M., Yaoita, Y., Kikuchi, M., 2003. New triterpenoids from *Gentiana lutea*. Chem. Pharm. Bull. 51, 89–91. https://doi.org/10.1248/cpb.51.89.

Wang, H., Song, W., Tao, W., Zhang, J., Zhang, X., Zhao, J., Yong, J., Gao, X., Guo, L., 2021. Identification wild and cultivated licorice by multidimensional analysis. Food Chem. 339, 128111 https://doi.org/10.1016/j.foodchem.2020.128111.

Wang, Y., Huang, H.-Y., Zuo, Z.-T., Wang, Y.-Z., 2018. Comprehensive quality assessment of *Dendrubium officinale* using ATR-FTIR spectroscopy combined with random forest and support vector machine regression. Spectrochim. Acta A 205, 637–648. https://doi.org/10.1016/j.saa.2018.07.086.

Yang, H.X., Ma, F., Du, Y.Z., Sun, S.Q., Wei, L.X., 2014. Study on the Tibetan medicine *Swertia mussotii* Franch and its extracts by Fourier transform infrared spectroscopy. Guang Pu Xue Yu Guang Pu Fen. Xi 34, 2973–2977.