

Chemometric analysis of French lavender and lavandin essential oils by near infrared spectroscopy

Sofia Lafhal*, Pierre Vanloot, Isabelle Bombarda, Jacky Kister, Nathalie Dupuy

Aix Marseille Université, LISA, EA4672, Equipe METICA, 13397 Marseille cedex 20, France

A B S T R A C T

Chemometric treatments of near infrared (NIR) spectra were used firstly to understand data structure by principal component analysis (PCA), to discriminate, by partial least squares-discriminant analysis (PLS-DA) regression, French lavender and lavandin essential oils (EOs) samples ($n = 160$) and the seven varieties (Abrial, Fine, Grosso, Maillette, Matherone, Sumian and Super) and to quantify the main compounds such as linalyl acetate, linalool, eucalyptol and camphor by PLS regression models. The study was carried out over three crop years (2012–2014) to take seasonal variations into account. French lavender and lavandin EOs and their varieties were well classified (100% for lavender/lavandin EOs and between 96 and 100% for varieties) by PLS-DA regression models. The calibration models obtained by PLS regression for the determination of the main compound contents revealed good correlation (≥ 0.97) between the predicted and reference values. In the case of major compounds including linalyl acetate and linalool, the relative error of prediction (REP) is close to 2.5%. Partial least squares regression vectors allowed us to identify lavandulyl acetate, eucalyptol, linalool, camphor, *trans*- β -ocimene, β -caryophyllene and linalyl acetate as metabolomic indicators of Fine, Maillette, Matherone, Abrial, Grosso, Super and Sumian varieties respectively. The use of NIR spectra allowed for an improvement in French lavender and lavandin EOs characterization, quality control and traceability.

Keywords:
Essential oils
Lavender
Lavandin
Metabolomics
NIR
Chemometrics

1. Introduction

The *lavandula* family (Lamiaceae) is composed of thirty nine *lavandula* species which are mostly of Mediterranean origin, such as *Lavandula angustifolia* (lavender) and numerous intraspecific taxa and hybrids such as lavandin. The smell of the *lavandula* family and its color are typical of the Mediterranean landscape, particularly in France. Its essential oil is used for cosmetics, perfumes, and in medicine (Cawthorn, 1995; Knowlton and Pearce, 1993; Lis-Balchin, 2003; Lis-Balchin and Hart, 1999; Piccaglia et al., 1993; Raut and Karuppayil, 2014; Vakili et al., 2014). The Lamiaceae family is classified in several species which are subdivided into varieties, each with its physical and environmental characteristics. The best known is the population lavender including *L. angustifolia*, and among its varieties (Maillette, Matherone and Fine), the Fine variety is the most famous thanks to its essential oil yield and its high linalool content. Lavandin is a hybrid between *L. angustifolia* P. Mill. and *L. latifolia* (L.f.) Medikus, whose culture is more recent and more focused on “industrial” production. Among the

main varieties of lavandin EOs (Abrial, Grosso, Sumian and Super), the Grosso variety is the most famous for its essential oil yield (Lis-Balchin, 2004). Lavender and lavandin oils contain more than one hundred compounds including linalyl acetate, linalool, camphor, borneol, eucalyptol and β -caryophyllene, each contributing to the chemical and sensory properties of the oils. The major distinction between varieties of lavandin essential oils is in their relative contents of linalyl acetate, linalool, eucalyptol and camphor. The chemical composition of lavandin has a higher camphor content than true lavender. Chemical composition can be revealed using gas chromatography (Bicchi, 2000; Can Başer and Özek, 2012; Cserhádi et al., 2005; Daferera et al., 2002; Marriott et al., 2001; Nikolić et al., 2014; Pallado et al., 1997; Ter Heide et al., 1970) and gas chromatography-sniffing (Chin and Marriott, 2015; Cserhádi et al., 2005; Marriott et al., 2001). These methods are usually applied for quality control purposes (Do et al., 2015) and selection of high-quality plants, but they are very time-consuming, and attempts have been made to find alternative analytical analysis methods. Also, to our knowledge, the quality of an essential oil depends on three main factors: quality of the plant, harvesting, distillation. However, on the basis of these criteria, only specialists are able to differentiate one variety from another. In this context, vibrational spectroscopic methods such as NIR spectroscopy in combination

with chemometric treatments were successfully introduced for a non-destructive determination of metabolites occurring in essential oils, and to facilitate the implementation of this method on an industrial scale. Vibrational spectroscopy methods offer a global metabolic fingerprint of essential oils and allow the prediction of metabolite contents such as camphor (Allwood et al., 2007; Bombarda et al., 2008; Cozzolino, 2009; Dupuy et al., 2013, 2014; Maietti et al., 2013; Moco et al., 2007; Schulz et al., 2008; Tankeu et al., 2014).

The aim of this study was to show the advantages of NIR spectroscopy associated with chemometric treatments for direct and rapid test methods. These can be used for lavender and lavandin EOs characterization, quality control and traceability. This methodology can also be used to identify metabolomic indicators and establish a faster quality control for discrimination of lavender and lavandin essential oils based on their specificities.

To achieve these goals, in the first part of this study, data mining was carried out by principal component analysis (PCA) to evaluate differences and similarities between sample spectra; partial least squares (PLS) regression was used to develop lavender/lavandin EOs or varieties prediction models. Partial least squares regression was also used to quantify the main compounds such as linalyl acetate, linalool and camphor with gas chromatography as reference data. In the second part, the PLS regression vectors obtained for the prediction of the main compounds were compared with the PLS regression vectors obtained for the prediction of the EO varieties to identify the metabolomic indicators of each variety.

2. Materials and methods

2.1. Essential oil samples

One hundred sixty samples were analyzed including lavandin oil samples ($n=94$) and lavender oil samples ($n=66$) obtained from three French cooperatives ("Société Coopérative Parfums Provence Ventoux" in Sault, "Société Coopérative des Plantes à Parfums de Provence" in Simiane la Rotonde and "France Lavande" in Montguers), over three years of harvest (2012–2014), including various varieties and various French collect areas (Unknown department (00), Alpes-de-Haute-Provence (04), Ardèche (07), Drôme (26) and Vaucluse (84)). The lavender/lavandin essential oils studied were [*L. angustifolia* Miller] (lavender) and its hybrid [*L. angustifolia* Miller \times *L. latifolia* Linnaeus filis Medikus] (lavandin) type France; they were divided into varieties: Fine (FI, $n=21$), Maillette (MA, $n=29$) and Matherone (MT, $n=16$) for lavender samples and Abrial (AB, $n=16$), Grosso (GR, $n=37$), Sumian (SU, $n=19$) and Super (SP, $n=22$) for lavandin samples.

2.2. Pure standard samples

Pure standard substances linalyl acetate, linalool, camphor, eucalyptol, borneol, *trans*- β -ocimene, lavandulyl acetate and β -caryophyllene were purchased from Lavender France (Montguers, France), Fluka (Buchs, Switzerland), Alpha Aesar (Karlsruhe, Germany), Merck (Schuchardt, Germany), Alpha Aesar (Karlsruhe, Germany), Sigma-Aldrich (Steinheim, Germany), Adrian (Aix-les-Milles, France) and TCI Europe (Zwijndrecht, Belgium) respectively.

2.3. Gas chromatography (GC)

2.3.1. Gas chromatography coupled to a mass spectrometer (GC-MS) analysis

Gas chromatography coupled to a mass spectrometer analyses were performed on a 7890A GC system coupled to a 5975C VL mass spectrometer detector (Agilent Technologies) equipped with a HP-5MS capillary column (J&W Scientific, 30 m \times 0.25 mm,

0.25 μ m film thickness). Data acquisition and processing were performed using the MSD Chemstation E.01.01.335 (Agilent) software. One microliter of diluted essential oil (80 μ L in 1.5 mL of ethanol) was injected. The experimental conditions developed in the laboratory were: oven temperature program, 2 min at 80 °C, then 80 °C to 200 °C (5 °C/min), then 200 °C to 260 °C (20 °C/min), and held at final temperature for 5 min; temperatures injector (split ratio 60) and detector were set to 250 °C; carrier gas was helium at a flow rate of 1.2 mL/min; solvent delay 2 min; ionization voltage 70 eV; electron multiplier 1 kV.

2.3.2. Gas chromatography (GC) analysis

Gas chromatography analyses were performed on a 7890A GC (Agilent Technologies) system with a flame ionisation detector (FID) equipped with a HP5 capillary column (J&W Scientific, 30 m \times 0.25 mm, 0.25 μ m film thickness). The data acquisition and processing were performed using the Chemstation B.04.03-SP1 (87) (Agilent) software. The experimental conditions developed in the laboratory were the same as GC-MS except for carrier gas which was hydrogen. Linear retention indices were calculated with reference to *n*-alkanes (C8–C28).

2.4. Near infrared (NIR) spectroscopy

The spectra of each lavender or lavandin EOs and liquid standard compounds placed in a quartz cell (2 mm) were recorded from 4000 to 10,000 cm^{-1} using the software result integration 2.1 (Thermo Nicolet), with 4 cm^{-1} resolution and 64 scans on a Nicolet Antaris II spectrometer equipped with an InGaAs photodiode detector, an H₂ NIR source and a CaF₂-germanium beam splitter. An empty quartz cell was taken as reference for the background spectrum before analysis of each sample. The near infrared spectrometer was situated in an air-conditioned room (21 °C). Spectra of solid standard compounds (camphor and borneol) were recorded using an integrating sphere in diffuse reflectance mode from 4000 to 10,000 cm^{-1} , with 4 cm^{-1} resolution and 64 scans. A background spectrum was collected under the same conditions before measurement of each sample (interleaved mode).

2.5. Chemometric analysis

2.5.1. Principal component analysis (PCA)

Principal component analysis (PCA) (Cozzolino et al., 2011; Ebsensen et al., 2002; Kumar et al., 2014; Wold et al., 1987) is an unsupervised modeling method that allows for exploratory data analysis; it extracts information from data set and removes noise; it reduces the number of dimensions; and it allows for classification of samples by investigating similarities and differences between the samples. The principal component analysis projects into a smaller number of latent variables called principal components (PC). Each principal component explains part of the total information contained in the original data and the first PC is the one that contains the most information, followed in descending order in terms of information by PC2 and PC3 and so on. Plotting two PCs relatively to each other thus allows for interpretation of some groups, thanks to the similarities or differences between samples.

2.5.2. Partial least squares regression (PLS)

Partial least squares (PLS) regression (Daszykowski et al., 2007; Ebsensen et al., 2002; Kumar et al., 2014; Liang and Kvalheim, 1996; Martens, 1979; Sjöström et al., 1983) is a supervised method which is based on the relation between signal intensity (spectrum) and the characteristics of the sample (*Y* variable). Interference and overlapping information may be overcome by using a powerful multicomponent analysis such as PLS. The algorithm is based on the ability to mathematically correlate spectral data to a property

Table 1
Chemical composition of lavender EO varieties (GC%).

RI ^a	Compounds	FI (n=21)		MA (n=29)		MT (n=16)		Lavender (n=66)	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
923	α-Pinene	0.16	±0.06	0.07	±0.04	0.08	±0.05	0.10	±0.06
944	Camphene	0.10	±0.05	0.20	±0.07	0.07	±0.04	0.12	±0.08
973	Sabinene	0.35	±0.12	0.41	±0.14	0.32	±0.12	0.36	±0.14
980	β-Pinene	0.07	±0.08	0.02	±0.04	0.02	±0.03	0.04	±0.06
985	3-Octanone	0.78	±0.38	1.37	±0.36	0.36	±0.15	0.84	±0.52
989	β-Myrcene	0.39	±0.14	0.34	±0.19	0.42	±0.12	0.38	±0.16
1012	Hexyl acetate	0.26	±0.07	0.36	±0.13	0.09	±0.11	0.24	±0.15
1030	Limonene	0.32	±0.13	0.15	±0.06	0.11	±0.11	0.19	±0.14
1034	Eucalyptol and cis-β-Ocimene	4.13	±0.79	1.43	±0.63	5.95	±1.35	3.84	±2.07
1044	trans-β-Ocimene	2.95	±0.60	1.02	±0.64	7.42	±1.67	3.80	±2.72
1073	Linalool oxide	0.16	±0.04	0.38	±0.08	0.12	±0.03	0.22	±0.14
1087	α-Terpinolene	0.12	±0.10	0.22	±0.17	0.07	±0.08	0.14	±0.15
1098	Linalool	26.92	±3.47	39.09	±6.55	19.74	±3.55	28.58	±9.50
1106	Octen-1-ol acetate	0.97	±0.28	0.72	±0.30	1.02	±0.29	0.90	±0.32
1145	Hexyl isobutyrate	0.08	±0.04	0.06	±0.03	0.02	±0.03	0.05	±0.04
1150	Camphor	0.33	±0.17	0.56	±0.17	0.27	±0.08	0.39	±0.20
1168	Lavandulol	0.98	±0.38	0.18	±0.29	1.53	±1.11	0.90	±0.83
1171	Borneol	1.11	±0.21	1.72	±0.34	0.83	±0.22	1.22	±0.47
1183	Terpinen-4-ol	4.12	±1.54	0.38	±0.33	1.86	±0.40	2.12	±1.85
1193	Hexyl butyrate	0.34	±0.07	0.53	±0.11	0.11	±0.13	0.33	±0.20
1197	α-Terpineol	0.74	±0.32	0.72	±0.33	0.77	±0.43	0.74	±0.35
1260	Linalyl acetate	38.22	±2.72	40.01	±5.32	39.98	±2.26	39.40	±4.09
1293	Lavandulyl acetate	4.60	±1.66	0.71	±0.43	5.53	±1.36	3.61	±2.46
1386	Geranyl acetate	0.55	±0.18	0.54	±0.21	0.48	±0.22	0.52	±0.21
1429	β-Caryophyllene	5.09	±0.71	3.67	±0.89	6.12	±0.71	4.96	±1.28
1460	β-Farnesene	1.76	±0.47	1.85	±0.26	2.92	±0.48	2.18	±0.62
1492	Germacren D	0.83	±0.24	0.27	±0.11	0.69	±0.13	0.60	±0.30
1594	Caryophyllene oxide	0.53	±0.15	0.48	±0.14	0.50	±0.18	0.50	±0.16

^a RI: Retention indices on HP-5 capillary column and SD: Standard deviation.

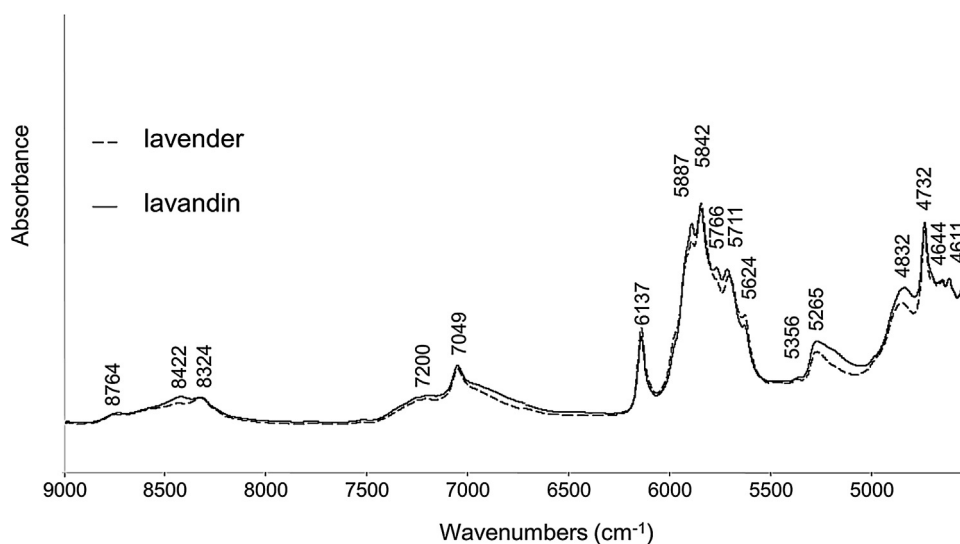


Fig. 1. NIR spectra of lavender (FI) and lavandin (AB) essential oils.

matrix of interest while simultaneously accounting for all other significant spectral factors that perturb the spectrum. It is thus a multivariate regression method that uses the full spectral region selected and is based on the use of factors. The evaluation of errors in calibration was carried out by computing the standard error of calibration (SEC) as follows:

$$SEC = \sqrt{\left(\frac{\sum_{i=1}^N (C_i - C'_i)^2}{N - 1 - p} \right)} \quad (1)$$

where C_i is the known value, C'_i is the value calculated by the calibration equation, N is the number of samples and p is the number

of independent variables in the regression optimized by cross-validation.

The standard error of prediction (SEP) gives an estimation of the prediction performance during the step of validation of the calibration and is calculated from the following equation:

$$SEP = \sqrt{\left(\frac{\sum_{i=1}^M (C_i - C_i)^2}{M} \right)} \quad (2)$$

where M is the number of samples in the prediction set.

Table 2
Chemical composition of lavandin EO varieties (GC%).

RI ^a	Compounds	AB (n = 16)		GR (n = 7)		SP (n = 19)		SU (n = 22)		Lavandin (n = 94)	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
923	α-Pinene	0.30	± 0.09	0.32	± 0.10	0.11	± 0.03	0.22	± 0.09	0.24	± 0.12
944	Camphene	0.27	± 0.06	0.22	± 0.06	0.14	± 0.04	0.20	± 0.06	0.21	± 0.07
973	Sabinene	0.59	± 0.15	0.36	± 0.10	0.13	± 0.05	0.27	± 0.08	0.34	± 0.18
980	β-Pinene	0.33	± 0.14	0.30	± 0.09	0.04	± 0.04	0.15	± 0.17	0.20	± 0.16
985	3-Octanone	0.22	± 0.07	0	± 0	0.68	± 0.18	0.86	± 0.36	0.44	± 0.41
989	β-Myrcene	0.42	± 0.06	0.48	± 0.10	0.51	± 0.14	0.43	± 0.09	0.46	± 0.11
1012	Hexyl acetate	0.16	± 0.05	0.13	± 0.04	0.41	± 0.17	0.04	± 0.05	0.18	± 0.16
1030	Limonene	0.67	± 0.09	0.62	± 0.10	0.70	± 0.40	0.87	± 0.21	0.71	± 0.24
1034	Eucalyptol and cis-β-Ocimene	8.41	± 0.98	5.23	± 0.97	3.55	± 0.46	7.52	± 1.96	6.18	± 2.10
1044	trans-β-Ocimene	2.89	± 0.55	0.29	± 0.06	1.49	± 0.27	1.72	± 0.63	1.60	± 1.01
1073	Linalool oxide	0.14	± 0.02	0.12	± 0.03	0.11	± 0.03	0.13	± 0.02	0.12	± 0.03
1087	α-Terpinolene	0.36	± 0.03	0.35	± 0.04	0.26	± 0.03	0.34	± 0.04	0.33	± 0.05
1098	Linalool	34.72	± 2.00	33.79	± 2.55	35.32	± 2.63	43.13	± 3.56	36.74	± 4.47
1106	Octen-1-ol acetate	0.48	± 0.11	0.31	± 0.07	0.28	± 0.07	0.19	± 0.05	0.31	± 0.12
1145	Hexyl isobutyrate	0.18	± 0.01	0.19	± 0.02	0.14	± 0.02	0.19	± 0.03	0.17	± 0.03
1150	Camphor	9.18	± 0.56	6.90	± 0.63	4.70	± 0.44	6.07	± 1.28	6.71	± 1.63
1168	Lavandulol	0.72	± 0.20	0.68	± 0.26	0.38	± 0.22	0.22	± 0.32	0.50	± 0.32
1171	Borneol	2.96	± 0.44	3.11	± 0.56	2.92	± 0.49	6.20	± 1.69	3.80	± 1.56
1183	Terpinen-4-ol	0.91	± 0.29	3.50	± 0.68	0.30	± 0.21	0.72	± 0.75	1.36	± 1.53
1193	Hexyl butyrate	0.39	± 0.06	0.40	± 0.06	0.69	± 0.12	0.52	± 0.11	0.50	± 0.15
1197	α-Terpineol	0.78	± 0.15	0.93	± 0.32	0.99	± 0.36	0.96	± 0.17	0.91	± 0.29
1260	Linalyl acetate	25.47	± 1.96	31.13	± 2.60	38.02	± 2.68	21.72	± 3.24	29.08	± 6.33
1293	Lavandulyl acetate	1.55	± 0.14	2.38	± 0.34	1.49	± 0.30	0.60	± 0.44	1.50	± 0.74
1386	Geranyl acetate	0.44	± 0.07	0.51	± 0.14	0.63	± 0.16	0.46	± 0.08	0.51	± 0.14
1429	β-Caryophyllene	2.62	± 0.27	1.76	± 0.17	1.36	± 0.14	1.43	± 0.38	1.79	± 0.49
1460	β-Farnesene	0.83	± 0.16	1.37	± 0.18	0.84	± 0.09	1.00	± 0.18	1.01	± 0.29
1492	Germacren D	0.71	± 0.09	0.75	± 0.11	0.57	± 0.08	0.53	± 0.12	0.64	± 0.14
1594	Caryophyllene oxide	0.12	± 0.03	0.08	± 0.02	0.09	± 0.03	0.08	± 0.05	0.09	± 0.04

^a RI: Retention indices on HP-5 capillary column and SD: Standard deviation.

Another useful parameter is the relative error of prediction (REP), which shows the predictive ability of the model, calculated from the equation:

$$REP = \frac{SEP}{\bar{y}} \times 100 \quad (3)$$

where \bar{y} is the mean of the observed values of compound content.

Partial least squares regression has not been developed for pattern recognition problems such as classification. However, this technique can be adapted for classification, giving rise to the partial least square-discriminant analysis (PLS-DA) regression method. Partial least square-discriminant analysis regression is carried out using an exclusive binary coding scheme with one bit per class (lavender/lavandin or variety). For the codification of samples, the two EOs corresponding to lavender and lavandin were arbitrarily classified in that order. For instance, a lavender sample was codified by the vector {1; 0}. The sample was then assigned to the class showing the highest membership value. Because of the difficulty of calibrating and predicting samples with binary variables it is necessary to grade results between values 0 and 1. Sample values lower than 0.5 and higher than 1.5 were identified as outside the defined class and samples with values between 0.5 and 1.5 were identified as belonging to the defined class. The same protocol was used to predict varieties. For the codification of samples, the seven varieties corresponding to AB, GR, SU, SP, FI, MA and MT were arbitrarily classified in that order. For instance, an AB sample was codified by the vector {1; 0; 0; 0; 0; 0; 0}.

As Haaland and Thomas (1988) explained, the first coefficient of regression (noted B) is a good approximation of the pure compound spectrum in the case of PLS regression. If we extend this property of the first regression coefficient to the class prediction by PLS-DA, the first coefficient of regression is an approximation of the original feature of the complex chemical system representing a class. In the context of our study, those regression coefficients B were very useful to identify metabolomic indicators which are the orig-

inal features of the lavender and lavandin EOs. The first regression coefficients B obtained for each EO varieties prediction were compared with those obtained for the main compounds prediction. The metabolomic indicator of a variety was the major compound with the first regression coefficient the most similar to that obtained for the prediction of the variety. The evaluation of the similarity between regression coefficients was carried out by computing the Pearson coefficient as follows:

$$r = \frac{COV(X, Y)}{\sqrt{varX \times varY}} \quad (4)$$

where cov (X,Y) is the covariance and var is the variance.

To build all PLS and PLS-DA models, 107 samples were used: (FI, n = 14), (MA, n = 19), (MT, n = 11), (AB, n = 11), (GR, n = 25), (SU, n = 13) and (SP, n = 14) and the validation step was performed by full cross-validation. To test the performance of the models in prediction, 53 samples were used: (FI, n = 7), (MA, n = 10), (MT, n = 5), (AB, n = 5), (GR, n = 12), (SU, n = 6) and (SP, n = 8). For best prediction performance, in the case of eucalyptol, trans-β-ocimene, borneol, linalyl acetate, lavandulyl acetate, the PLS regression models were built from derived NIR spectra, and for linalool, camphor, and β-caryophyllene from baseline offset pretreated spectra. To classify samples, in the case of the FI variety, the PLS-DA regression model was built from derived NIR spectra and for other varieties from baseline offset pretreated spectra.

2.5.3. Software

Chemometric analysis was performed using The UNSCRAMBLER X V.10.3 (CAMO/Software, Oslo, Norway).

3. Results and discussion

3.1. Gas chromatography

The identification of the compounds was achieved by comparing their mass spectra with those of Wiley275 and NIST05a libraries as

Chemometric analysis

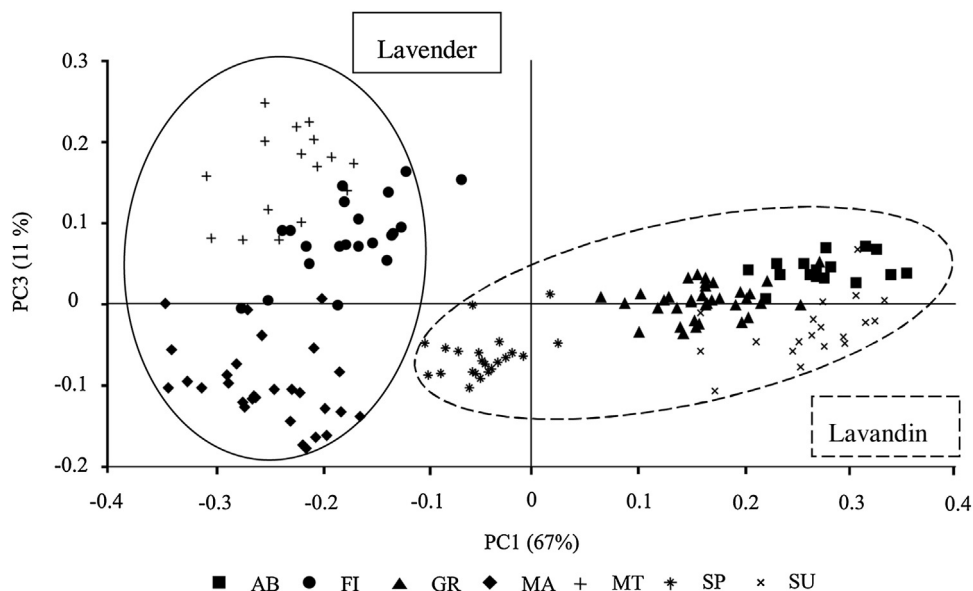


Fig. 2. Score plot (PC1/PC3) of the PCA of NIR data ($n = 160$). Lavender EOs: Fine (FI), Maillette (MA) and Matherone (MT) and lavandin EOs: Abrial (AB), Grosso (GR), Super (SP) and Sumian (SU).

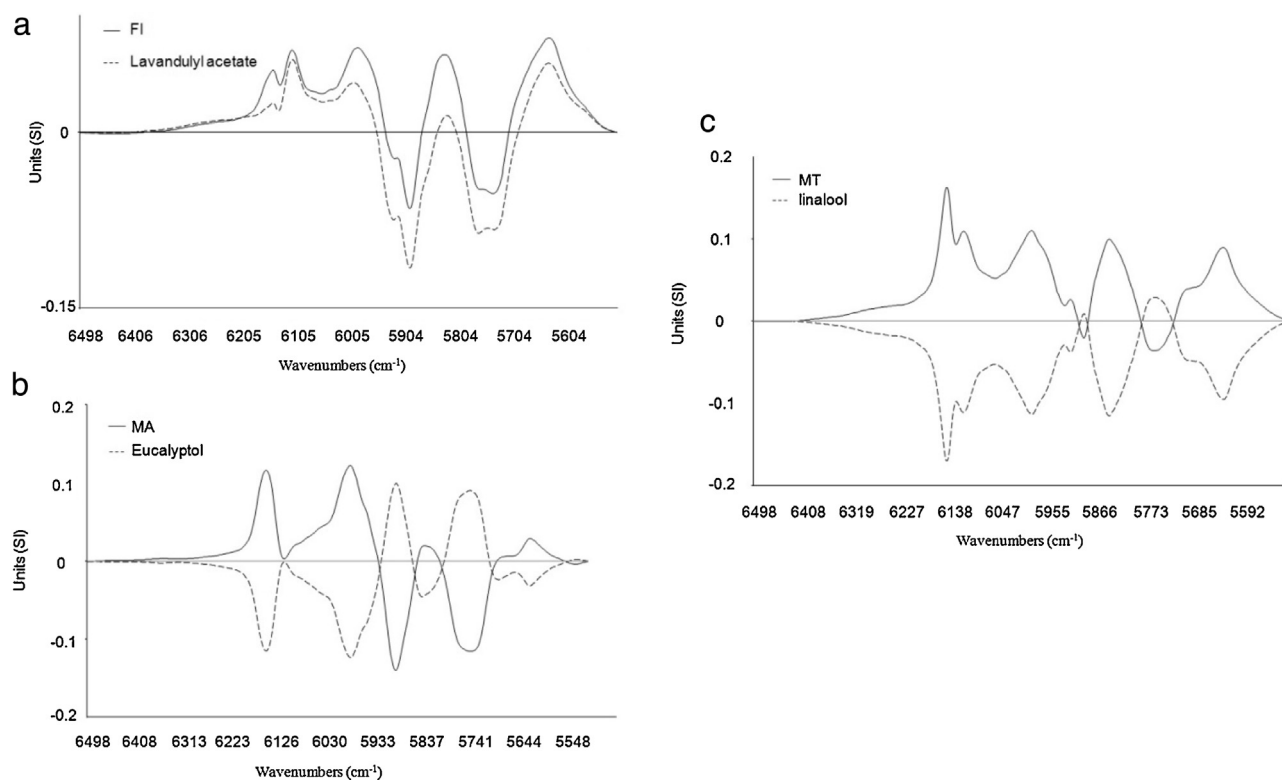


Fig. 3. Superposition of the first regression vectors obtained for FI variety and lavandulyl acetate (a) Superposition of the first regression vectors obtained for MA variety and eucalyptol (b) Superposition of the first regression vectors obtained for MT variety and linalool (c).

well as by comparing their retention indices with those of authentic samples. [Tables 1 and 2](#) list the relative composition (mean and range) determined from peak areas for the 28 major compounds, each accounting for more than 0.1% in the EOs. Concerning lavender EOs, the major compounds identified are linalyl acetate (39.40%), linalool (28.58%), β -caryophyllene (4.96%), eucalyptol coeluted with *cis*- β -ocimene (3.84%), *trans*- β -ocimene (3.80%) and lavandu-

lyl acetate (3.61%). Concerning lavandin EOs, the major compounds identified are linalool (36.74%), linalyl acetate (29.08%), camphor (6.71%), eucalyptol coeluted with *cis*- β -ocimene (6.18%) and borneol (3.80%). For most samples, lavender EOs were characterized by a higher content of linalyl acetate than linalool, while lavandin EOs presented higher amounts of linalool than linalyl acetate, except for the SP variety. As shown in [Tables 1 and 2](#), there are other signif-

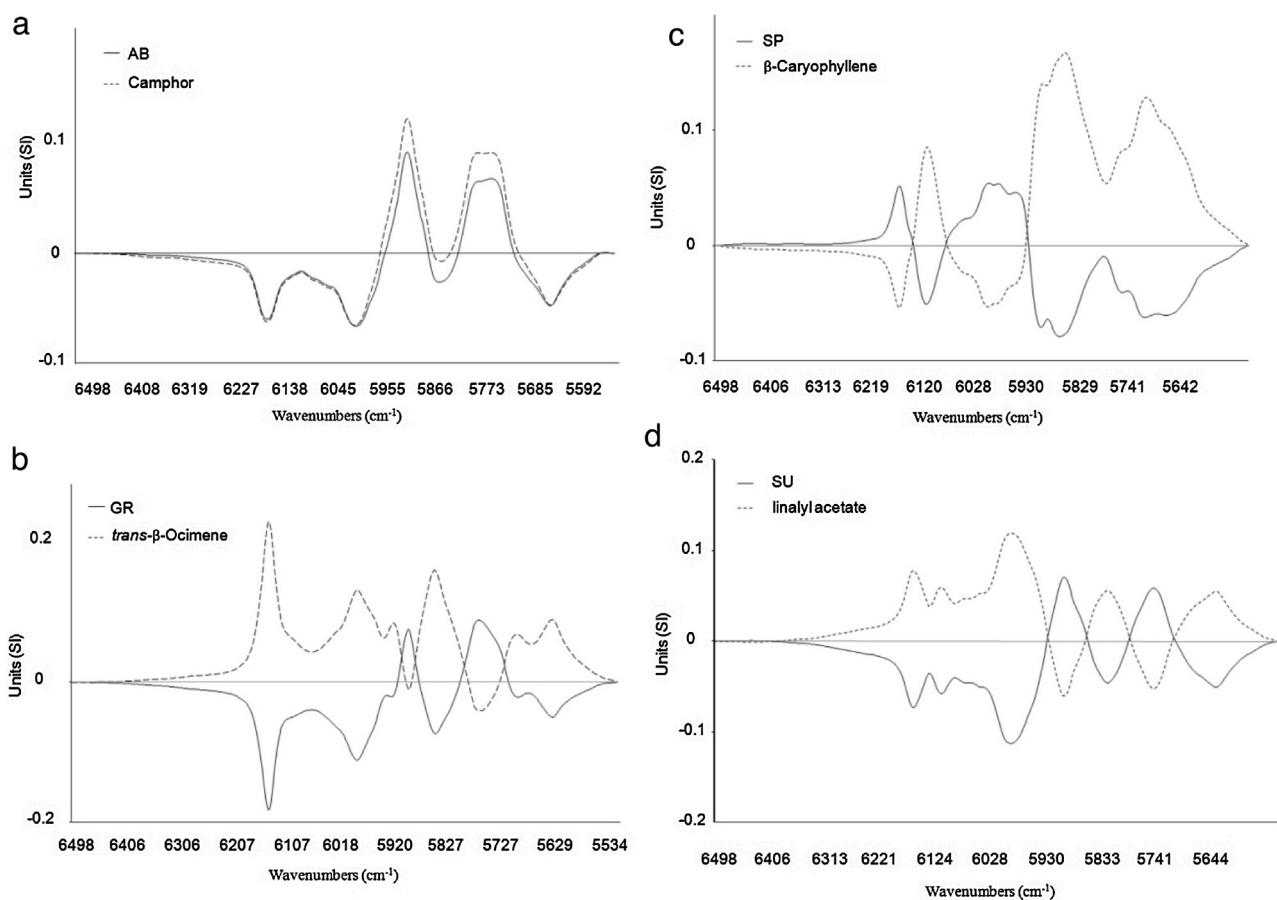


Fig. 4. Superposition of the first regression vectors obtained for AB variety and eucalyptol (a). Superposition of the first regression vectors obtained for GR variety and *trans*- β -ocimene (b). Superposition of the first regression vectors obtained for SP variety and β -caryophyllene (c). Superposition of the first regression vectors obtained for SU variety and linalyl acetate (d).

icant differences between lavender and lavandin EOs, particularly for camphor and β -caryophyllene contents. The low camphor content in lavender (0.39%) compared with lavandin (6.71%) justifies its use in the perfume industry and the use of lavandin in the toiletries industry. The high β -caryophyllene content in lavender (4.96%) compared with lavandin (1.79%) also characterizes lavenders. Other minor differences in compound contents were observed, showing a potential differentiation between lavender and lavandin EOs. Regarding lavender varieties, the low linalool content (19.74%) and high *trans*- β -ocimene content (7.42%) in MT are characteristics of this variety. The low eucalyptol content (1.43%) and high 3-octanone content (1.37%) in MA are characteristics of this variety. Fine variety is characterized by high content in terpinen-4-ol (4.12%). Regarding lavandin varieties, the low linalyl acetate content in SU (21.72%) and the high eucalyptol content in AB (8.41%) are characteristic of these varieties. The low *trans*- β -ocimene content (0.29%) is characteristic of GR variety. Super variety is characterized by high linalyl acetate content (38.02%) among lavandin EOs and is also characterized by low β -caryophyllene content (1.36%).

3.2. NIR spectroscopy

Fig. 1 shows the NIR spectra of lavender and lavandin EO samples over the spectral range 4500–9000 cm^{-1} . Even though the compositions of lavender and lavandin EOs differ greatly, their spectra are only slightly different. The major differences between the spectra were: the intensity in the 4950–5400 cm^{-1} region which was stronger for lavandin EOs (second overtone of C=O stretching vibrations); the intensity for the bands at 8422 cm^{-1} , 5887 cm^{-1} and

5766 cm^{-1} which were stronger for lavandin EOs; and also a relevant difference of intensity at 5624 cm^{-1} which was stronger for lavender EOs. The spectra of the eight major pure terpenoids (>2%) in lavender and lavandin EOs were recorded and the assignment of major bands was carried out using literature data (Cozzolino et al., 2005; Dupuy et al., 2010; Guo et al., 2006; Workman and Weyer, 2007). The bands between 4445 and 4895 cm^{-1} were characteristic of the combination of CH stretching vibrations of CH_3 and CH_2 with other vibrations (C=O and C=C stretching); the bands between 4950 and 5400 cm^{-1} were attributed to the second overtone of carbonyl compounds; the bands between 5690 and 6190 cm^{-1} corresponded to the first overtone of the CH stretching vibration of CH_3 , CH_2 , and $\text{CH}=\text{CH}$; the bands between 6600 and 7600 cm^{-1} corresponded to the first overtone combinations of the CH stretching vibration of CH_3 , CH_2 , and $\text{CH}=\text{CH}$ and the bands between 8100 and 8950 cm^{-1} corresponded to the second of the CH stretching vibrations of CH_3 , CH_2 , and $\text{CH}=\text{CH}$.

3.3. Chemometric analysis

The first and third components (PC1/PC3) of the PCA carried out on NIR data represent 78% of the total spectral variance (Fig. 2). Principal components one and three were presented instead of PC1 and PC2 because PC1 and PC2 allowed for discrimination between lavender and lavandin EOs but not for discrimination between varieties. The lavender and lavandin groups are separated on the first component (PC1), lavender is negatively projected while lavandin is positively projected, except for the SP variety which is negatively projected but near the center. Principal component one also allows

Table 3

Statistics of the PLS regression models for the 8 main compounds in the lavender and lavandin EOs established from NIR data.

Compounds	Mean (GC%)	R ²	SEC	LV	Q ²	SEP	REP (%)
Eucalyptol ^a	5.01	0.987	0.379	8	0.983	0.443	8.84
trans- β -Ocimene ^a	2.70	0.977	0.435	5	0.988	0.360	13.33
Linalool ^b	32.66	0.991	0.978	5	0.993	0.847	2.59
Camphor ^b	3.55	0.998	0.185	6	0.998	0.226	6.36
Borneol ^a	2.51	0.994	0.174	6	0.994	0.200	7.97
Linalyl acetate ^a	34.24	0.994	0.782	3	0.993	0.821	2.39
Lavandulyl acetate ^a	2.55	0.987	0.306	8	0.971	0.424	16.62
β -Caryophyllene ^b	3.37	0.989	0.253	7	0.986	0.284	8.42

^a Prediction regression model (PLS regression) established from derived NIR data.^b Prediction regression model (PLS regression) established from baseline offset NIR data. R²: coefficients correlation in calibration, Q²: coefficients correlation in prediction, SEC: standard error of calibration, SEP: standard error or prediction, LV: latent variables and REP: relative error of prediction.

for the separation of lavandin varieties. Principal component one can be described with linalyl acetate and camphor contents. Lavender EOs have low camphor content (0.39%) and high linalyl acetate (39.40%) compared with lavandin (6.71% and 29.08% respectively). As observed previously, among lavandin EOs the SP variety was characterized by high linalyl acetate and low camphor contents. That is why this variety is very close to lavender EOs in the PCA. Principal component three allows for separation of lavender varieties and can be described by linalool and eucalyptol contents. The varieties MA, FI and MT are separated on PC3 according to their linalool content (39.09%, 26.92% and 19.74% respectively) and eucalyptol content (1.43%, 4.13% and 5.95% respectively).

The determination of the main compound contents in the EOs by chemometric analysis from NIR spectra in the range 5500–6500 cm⁻¹ (a part of first overtone information) was achieved using PLS regression algorithms using gas chromatographic data as reference. Table 3 gives the statistics of the PLS regression models for the eight quantified compounds (accounting for more than 2%). For best results, in the case of eucalyptol, trans- β -ocimene, borneol, linalyl acetate and lavandulyl acetate, the PLS regression models were built from derived NIR spectra and for other compounds from baseline offset pretreated NIR spectra. As shown in Table 3, the eight main compounds can be classified in three groups. For the first group, comprising the two major compounds, very good results were obtained with a REP at 2.39% and 2.59% for the prediction of linalyl acetate and linalool respectively. In the second group, four compounds were correctly predicted since their REP were between 5 and 10% (eucalyptol, camphor, borneol and β -caryophyllene). In the third group, two compounds were not well-predicted with a REP at 13.33% and 16.62% for the prediction of trans- β -ocimene and lavandulyl acetate respectively. The performance of the PLS prediction models was directly related to the compound contents in the EOs.

Table 4 gives the prediction results for the 53 lavender and lavandin EOs recognition. The models were based on the same calibration and prediction set used for quantitative analysis. In the first step the discrimination between lavender and lavandin was studied. The results were excellent, with 100% of correct classification.

Table 4

Classification matrix obtained in prediction (PLS-DA regression) of lavender and lavandin EO varieties established from NIR data.

	Variety	Number of latent variables	AB (n=5)	GR (n=12)	SP (n=8)	SU (n=6)	FI (n=7)	MA (n=10)	MT (n=5)	False-negative samples	False-positive samples	
NIR models	Lavandin	AB ^a	12	5	0	0	1	0	0	0	0	1
		GR ^a	10	0	12	1	0	0	0	0	0	1
		SP ^a	10	0	0	8	0	0	0	0	0	0
		SU ^a	8	0	0	0	5	0	0	0	1	0
	Lavender	FI ^b	8	0	0	0	0	5	0	0	2	0
		MA ^a	11	0	0	0	0	0	10	0	0	0
		MT ^a	8	0	0	0	0	0	0	5	0	0

^a Prediction regression model (PLS regression) established from baseline offset NIR data.^b Prediction regression model (PLS regression) established from derived NIR data.

In the second step, the seven varieties were discriminated and well-predicted with a percentage of correct classification higher than 96% (Table 4). The varieties MA, MT and SP gave 100% of correct classification. For lavender varieties, the FI variety presented two false negative samples. Regarding the lavandin varieties, AB presented one false positive sample, GR presented one false positive sample too and SU presented one false negative sample. To predict varietal origin, the difficulty comes from the variability in the quality of samples, which depends on harvest year, collect area, storage conditions and extraction process, regardless of varietal origin.

As it was possible to classify samples as a function of lavender/lavandin and the origin of their varieties, it was interesting to understand how these classifications were established in order to identify metabolomic indicators. Figs. 3 and 4 present the superposition of the first regression coefficients B obtained for lavender and lavandin varieties respectively and their corresponding metabolomic indicators. For the lavender varieties, the FI first regression coefficient B was correlated with the regression coefficient obtained for quantitative analysis of lavandulyl acetate (Fig. 3a). The Pearson coefficient between the first two regression coefficients was 0.86. Fine variety was characterized by the higher content of lavandulyl acetate (mean of 4.60%). Lavandulyl acetate could be considered as a metabolomic indicator for the FI variety. Some Matherone samples had high concentrations in lavandulyl acetate and this could explain that three samples of the MT variety overlay the FI group on the PCA score plot. Maillette first regression coefficient B was anti-correlated with the regression coefficient obtained for quantitative analysis of eucalyptol (Fig. 3b). The Pearson coefficient between the two first regression coefficients was -0.97. Maillette variety was characterized by a low content (mean 1.43%) in eucalyptol. Eucalyptol could be considered as a metabolomic indicator for the MA variety. Matherone first regression coefficient B was anti-correlated with the regression coefficient obtained for quantitative analysis of linalool (Fig. 3c). The Pearson coefficient between the first two regression coefficients was -0.99. Matherone variety was characterized by a low content (19.74%) in linalool, which was a metabolomic indicator for this variety. Concerning the lavandin varieties, the AB and cam-

phor first regression coefficient B were correlated with a Pearson coefficient of 0.98 (Fig. 4a). Abrial variety presented a mean content of 9.18% in camphor. Abrial variety was characterized by a high content of camphor which was a metabolomic indicator for this variety. Grosso first regression coefficient B was anti-correlated with the one obtained for *trans*- β -ocimene quantitative analysis with a Pearson coefficient of -0.87 (Fig. 4b). Grosso variety presented a mean content of 0.29% for *trans*- β -ocimene. A low *trans*- β -ocimene content in EOs was a metabolomic indicator of the GR variety. Super first regression coefficient B was anti-correlated with that obtained for β -caryophyllene quantitative analysis with a Pearson coefficient of -0.95 (Fig. 4c). Super variety presented a mean content of 1.36% in β -caryophyllene. A low content in β -caryophyllene could be considered as a metabolomic indicator of the SP variety. Sumian first regression coefficient B was anti-correlated with that obtained for linalyl acetate quantitative analysis with a Pearson coefficient of -0.99 (Fig. 4d). Sumian variety presented a mean content of 21.72% in linalyl acetate. A low content in linalyl acetate could be considered as a metabolomic indicator of the SU variety. To sum up, first PLS regression vectors allowed us to identify lavandulyl acetate, eucalyptol, linalool, camphor, *trans*- β -ocimene, β -caryophyllene and linalyl acetate as metabolomic indicators of FI, MA, MT, AB, GR, SP and SU varieties respectively.

4. Conclusion

Near infrared spectroscopy combined with chemometric analysis was used to quantify the main compounds in lavender and lavandin EOs. Developed PLS models can be used to predict the linalyl acetate and linalool contents with an error rate lower than 3%; eucalyptol, camphor, borneol and β -caryophyllene 5 and 10%; *trans*- β -ocimene and lavandulyl acetate with an error between 10 and 17%. The results matched the prediction for compounds occurring in high concentration quite accurately. Camphor, which is an important compound to detect adulteration, has a lower error rate at 6.36%. Near infrared spectroscopy was also used as a metabolomic fingerprint combined with chemometric analysis to discriminate lavender from lavandin essential oils and their varieties (Abrial, Grosso, Sumian, Super, Fine, Maillette and Matherone). Lavender/lavandin and their varieties were very well-predicted with a percentage of correct classification at 100% for lavender/lavandin and between 96 and 100% for their varieties. The examination of the PLS first regression coefficient B allowed for identification of metabolomic indicators for all lavender/lavandin varieties. Near infrared spectroscopy associated with chemometric analysis was thus demonstrated to be a powerful tool for lavender authenticity. This approach could be applied to other essential oils commonly used in the perfume industry, medicine and other fields.

Acknowledgments

The authors are grateful to the “Société Coopérative Parfums Provence Ventoux”, “Société Coopérative des Plantes à Parfums de Provence” and “France Lavande”, for providing samples of lavender and lavandin essential oils.

References

Allwood, J.W., Ellis, D.I., Goodacre, R., 2007. Metabolomic technologies and their application to the study of plants and plant–host interactions. *Physiol. Plant.* 132 (2).

Bicchi, C., 2000. Essential oils | gas chromatography. In: Wilson, I.D. (Ed.), *Encyclopedia of Separation Science*. Academic press, Oxford, pp. 2744–2755.

Bombarda, I., Dupuy, N., Le Vanda, J.-P., Gaydou, E.M., 2008. Comparative chemometric analyses of geographic origins and compositions of lavandin var. Grosso essential oils by mid infrared spectroscopy and gas chromatography. *Anal. Chim. Acta* 613, 31–39.

Can Başer, K.H., Özek, T., 2012. Chapter 22—analysis of essential oils and fragrances by gas chromatography. In: Poole, C.F. (Ed.), *Gas Chromatography*. Elsevier, Amsterdam, pp. 519–527.

Cawthorn, A., 1995. A review of the literature surrounding the research into aromatherapy. *Complement. Ther. Nurs. Midwifery* 1, 118–120.

Chin, S.-T., Marriott, P.J., 2015. Review of the role and methodology of high resolution approaches in aroma analysis. *Anal. Chim. Acta* 854, 1–12.

Cozzolino, D., Murray, I., Chree, A., Scaife, J.R., 2005. Multivariate determination of free fatty acids and moisture in fish oils by partial least-squares regression and near-infrared spectroscopy. *LWT* 38, 821–828.

Cozzolino, D., 2009. Near infrared spectroscopy in natural products analysis. *Planta Med.* 75, 746–756.

Cozzolino, D., Cynkar, W.U., Shah, N., Smith, P., 2011. Multivariate data analysis applied to spectroscopy: potential application to juice and fruit quality. *Food Res. Int.* 44, 1888–1896.

Cserháti, T., Forgács, E., Deyl, Z., Miksik, I., 2005. Chromatography in authenticity and traceability tests of vegetable oils and dairy products: a review. *Biomed. Chromatogr.* 19, 183–190.

Daferera, D.J., Tarantilis, P.A., Polissiou, M.G., 2002. Characterization of essential oils from *Lamiaceae* Species by Fourier transform raman spectroscopy. *J. Agric. Food Chem.* 50, 5503–5507.

Daszykowski, M., Kaczmarek, K., Vander Heyden, Y., Walczak, B., 2007. Robust statistics in data analysis—a review: basic concepts. *Chemom. Intell. Lab. Syst. Syst.* 85, 203–219.

Do, T.K.T., Hadji-Minaglou, F., Antonioti, S., Fernandez, X., 2015. Authenticity of essential oils. *TrAC Trends Anal. Chem.* 66, 146–157.

Dupuy, N., Galtier, O., Le Dréau, Y., Pinatel, C., Kister, J., Artaud, J., 2010. Chemometric analysis of combined NIR and MIR spectra to characterize French olives. *Eur. J. Lipid Sci. Technol.* 112, 463–475.

Dupuy, N., Gaydou, V., Kister, J., 2014. Quantitative analysis of lavender (*Lavandula angustifolia*) essential oil using multiblock data from infrared spectroscopy. *Am. J. Anal. Chem.* 5, 633–645.

Dupuy, N., Molinet, J., Mehl, F., Nanlohy, F., Le Dréau, Y., Kister, J., 2013. Chemometric analysis of mid infrared and gas chromatography data of Indonesian nutmeg essential oils. *Ind. Crops Prod.* 43, 596–601.

Esbensen, K.H., Guyot, D., Westad, F., Houmoller, L.P., 2002. Multivariate data analysis: in practice: an introduction to multivariate data analysis and experimental design multivariate data analysis, fifth ed. pp. 19–75 and 115–171.

Guo, C., Shah, R.D., Dukor, R.K., Freedman, T.B., Cao, X., Nafie, L.A., 2006. Fourier transform vibrational circular dichroism from 800 to 10,000 cm^{-1} : near-IR-VCD spectral standards for terpenes and related molecules. *Vib. Spectrosc.* 42, 254–272.

Haaland, D.M., Thomas, E.V., 1988. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Anal. Chem.* 60, 1193–1202.

Knowlton, J., Pearce, S., 1993. Perfumery. In: Knowlton, J., Pearce, S. (Eds.), *Handbook of Cosmetic Science & Technology*. Elsevier, pp. 317–361.

Kumar, N., Bansal, A., Sarma, G.S., Rawal, R.K., 2014. Chemometrics tools used in analytical chemistry: an overview. *Talanta* 123, 186–199.

Liang, Y.-Z., Kvalheim, O.M., 1996. Robust methods for multivariate analysis—a tutorial review. *Chemom. Intell. Lab. Syst. Syst.* 32, 1–10.

Lis-Balchin, M., Hart, S., 1999. Studies on the mode of action of the essential oil of lavender (*Lavandula angustifolia* P. Miller). *Phytother. Res.* 13, 540–542.

Lis-Balchin, M., 2003. Lavender: the Genus *Lavandula*. CRC Press.

Lis-Balchin, M., 2004. 11—Lavender. In: Peter, K.V. (Ed.), *Handbook of Herbs and Spices*, Woodhead Publishing Series in Food science, Technology and Nutrition. Woodhead Publishing, pp. 179–195.

Maietti, S., Rossi, D., Guerrini, A., Useli, C., Romagnoli, C., Poli, F., Bruni, R., Sacchetti, G., 2013. A multivariate analysis approach to the study of chemical and functional properties of chemo-diverse plant derivatives: lavender essential oils: chemodiversity and multivariate analysis: the lavender case. *Flavour Fragr. J.* 28, 144–154.

Marriott, P.J., Shellie, R., Cornwell, C., 2001. Gas chromatographic technologies for the analysis of essential oils. *J. Chromatogr. A* 936, 1–22.

Martens, H., 1979. Factor analysis of chemical mixtures: Non-negative factor solutions for spectra of cereal amino acids 11. In: Presented at the International Conference on Computers and Optimization in Analytical Chemistry, Amsterdam, April 1978.

Moco, S., Vervoort, J., Moco, S., Bino, R.J., De Vos, R.C.H., Bino, R., 2007. Metabolomics technologies and metabolite identification. *TrAC Trends Anal. Chem.* 26, 855–866.

Nikolić, M., Jovanović, K.K., Marković, T., Marković, D., Gligorijević, N., Radulović, S., Soković, M., 2014. Chemical composition, antimicrobial, and cytotoxic properties of five *Lamiaceae* essential oils. *Ind. Crops Prod.* 61, 225–232.

Pallado, P., Tassinato, G., D’Alpaos, M., Traldi, P., 1997. Gas chromatography/mass spectrometry in aroma chemistry: a comparison of essential oils and flavours extracted by classical and supercritical techniques. *Rapid Commun. Mass Spectrom.* 11, 1335–1341.

Piccaglia, R., Marotti, M., Giovanelli, E., Deans, S.G., Eaglesham, E., 1993. Antibacterial and antioxidant properties of Mediterranean aromatic plants. *Ind. Crops Prod.* 2, 47–50.

Raut, J.S., Karuppaiyil, S.M., 2014. A status review on the medicinal properties of essential oils. *Ind. Crops Prod.* 62, 250–264.

Schulz, H., Vaidyan, V.K., Jayakumar, V.S., 2008. Potential of Raman and Infrared Spectroscopy for plant analysis. *AIP*, 8–12.

- Sjöström, M., Wold, S., Lindberg, W., Persson, J., Martens, H., 1983. [A multivariate calibration problem in analytical chemistry solved by partial least-squares models in latent variables.](#) *Anal. Chim. Acta* 150, 61–70.
- Tankeu, S.Y., Vermaak, I., Kamatou, G.P.P., Viljoen, A.M., 2014. [Vibrational spectroscopy and chemometric modeling: an economical and robust quality control method for lavender oil.](#) *Ind. Crops Prod.* 59, 234–240.
- Ter Heide, R., Timmer, R., Wobben, H.J., 1970. [Investigation into the composition of lavender and lavandin oil: Some new components.](#) *J. Chromatogr. A* 50, 127–131.
- Vakili, A., Sharifat, S., Akhavan, M.M., Bandegi, A.R., 2014. [Effect of lavender oil \(*Lavandula angustifolia*\) on cerebral edema and its possible mechanisms in an experimental model of stroke.](#) *Brain Res.* 1548, 56–62.
- Wold, S., Esbensen, K., Geladi, P., 1987. [Principal component analysis, Chemom. Intell. Lab. Syst., Proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists](#) 2, 37–52.
- Workman, J., Weyer, L., 2007. [Practical Guide to Interpretive Near-Infrared Spectroscopy.](#) CRC Press.