Comparison of the performances of handheld and benchtop near infrared spectrometers: Application on the quantification of chemical components in maritime pine (Pinus Pinaster) resin

Morandise Rubini, Lisa Feuillerat, Thomas Cabaret, Léo Leroyer, Luc Leneveu, Bertrand Charrier

Title: Comparison of the performances of handheld and benchtop near infrared spectrometers: application on the quantification of chemical components in maritime pine (*Pinus Pinaster*) resin

Morandise Rubini [a*]; Lisa Feuillerat [a]; Thomas Cabaret [a]; Léo Leroyer [a]; Luc Leneveu [b]; Bertrand Charrier [a]

[a] CNRS/Université de Pau des Pays de l'Adour, Institut des sciences analytiques et de physico-chimie pour l'environnement et les matériaux, Xylomat, UMR5254, 40004, Mont de Marsan, France

[b] Biogemme - Holiste, 40600, Biscarrosse, France

* Corresponding authors

Corresponding authors e-mail address: morandise.rubini@univ-pau.fr

1 Introduction

Maritime pine (*Pinus pinaster*) tapping (resin collection) was a prominent activity in France from the 19th century until the 1970's [1]. Maritime pine covers in the Southwest of France more than one million hectares [2]. After the 1970s, tapping drastically declines because its lost its profitability, mainly because to the growth of the Chinese market [3,4]. Recently, tapping reemerged [5] with new, mechanized tapping techniques, in a context with concerns about harvesting in sustainable and environmentally friendly conditions [6].

Closed cup tapping is one of these techniques. The first works are related to (C. Courau, 1996) [7]. Closed cup tapping became increasingly popular in France and is currently being improved [8]. Briefly described, a superficial incision using a circular saw removes a section of the bark and of the secondary phloem, where the resin is formed [9]. Then, a stimulant paste composed of natural acids is applied to activate resin exudation. Finally, a bottleneck guides the drained resin into an airtight sealed plastic bag, which protects the resin from impurities such as insects, sand, plant debris or rainwater, and limits evaporation.

Resin exudates abundantly [10] as a complex mixture of a volatile fraction (turpentine) and a nonvolatile fraction (rosin) [11,12]. Turpentine consists of a few unsaturated hydrocarbon monoterpenes, with smaller amounts of other monoterpenes [13,14]. Rosin is mainly composed of diterpenic monocarboxylic acids, plus neutral components [15–17]. The various components can be analyzed quantitatively using gas chromatography [18–21] and or liquid chromatography [22,23]. These methods are used for quality control and to select higher quality resin, but they are time-consuming and hard to use on-site. Considering this, the present study looks for alternative analytical methods, including Near infrared (NIR) spectroscopy in combination with chemometrics [24,25].

NIR spectroscopy is a non-destructive, reliable, and rapid tool for quantification, and can be applied on portable devices [26]. Indeed, low cost handheld near-infrared spectrometer, such

as *SCiO* (*Consumer Physics*, Israel), has been reported in the literature as a reliable and inexpensive technology, which facilitates its use on an industrial scale [27–32]. The objective of this study is to assess the potential of *SCiO* as a direct and rapid analysis tool for the quantification of the main chemical components of *Pinus pinaster* resin with chemometric processing.

Partial Least Squares (PLS) regression was used to develop quantitative predictive models. Those models were optimized various different spectral preprocessing methods. Then, the performance of *SCiO* was compared to the baseline of a benchtop spectrometer.

2 Materials and methods

2.1 Sample collection and conservation

One hundred and twenty-five samples of maritime pine (*Pinus pinaster*) resin were harvested in Biscarrosse (Landes, France) during the summer of 2018. They were collected by tapping one hundred and twenty-five fifty years old trees using a closed cup tapping technique patented by (L. Leneveu, 2002) [8]. Briefly described, the method consisted in having the resin flow into a sealed plastic bag made of an ethylene vinyl alcohol (EVOH) inner film (70 µm thick) and a low-density polyethylene (LDPE) outer film (45 µm thick).

Once, the samples were stored in a box kept away from light and heat in order avoid chemical alteration.

2.2 Reference methods

2.2.1 Proportion of turpentine fraction in Pinus pinaster resin (% turpentine) using a ventilated oven

For convenience, the relative proportion of turpentine and rosin fractions is designated here collectively as the "Proportion of turpentine fraction".

The proportion of turpentine fraction in maritime pine (*Pinus pinaster*) resin (*% turpentine*) was quantified using a ventilated oven.

The *% turpentine* was an important data for this study, because, the chemical composition was determined with two different analytical instruments, and the chemical composition pertains respectively to each fraction. Therefore, *% turpentine* is an indicator of the actual amount of the chemical components in each of the fractions.

Briefly, 2 g of resin were poured in a aluminium cup. The weight difference before and after drying was used to determine *% turpentine*. Each sample was placed in a ventilated oven at a temperature close to the boiling point of monoterpenes, for a duration sufficient to provoke its evaporation. Specifically, the samples were placed at 150 °C for 30 min in the ventilated oven.

2.2.2 Chemical composition of the turpentine as determined via Gas Chromatography coupled to a Flame Ionization Detector (GC–FID) analysis and sample preparation

The chemical composition of the turpentine fraction from maritime pine (*Pinus pinaster*) resin was quantified using a Perkim-Elmer Clarus 500 gas chromatograph (GC) coupled to a Flame Ionization Detector (FID), and a Perkim-Elmer Elite-5MS capillary column (30 m x 0,25 mm x 0,5 μm film thickness).

To prepare the samples, the plastic bag containing the resin was homogenized manually for several minutes until the turpentine and rosin were completely mixed. Then, 10 mL of the resin were poured in a cylindrical glass vial (h = 50 mm, outer diameter = 22 mm). After three days, 0.5 mL of supernatant containing turpentine was diluted with 1.5 ml of hexane, which was, in turn, stirred in an ultrasonic bath for 1 min, then filtered through a 0.2 μm nylon filter before injecting. 1 μL of this solution.

The experimental conditions developed in the laboratory were as follow: both injector and detector temperatures were set at 290 °C; oven temperature program, 50 to 150 °C (2 °C/min), then 150 to 320 °C (10 °C/min); hydrogen was used as carrier gas. The first ramp was used to elute the volatile components of the turpentine, the second ramp was used to clean the column.

Between each sample injection, two flash cleaning programs with acetone and hexane were applied to remove any residue (50 to 250 °C (15 °C/min)).

For the identification of the components of the turpentine fraction, Kovats's linear retention index was calculated from the injection of a homologous series of hydrocarbons (C8 – C20) and compared with literature data [33], in accordance with ASTM D6730 [34]. For the quantitative analysis, the respective rates were expressed as percentages.

2.2.3 Chemical composition of the rosin fraction determined via High-Performance Liquid Chromatography coupled to Diode Array Detectors (HPLC-DAD) and sample preparation

The chemical composition of the rosin fraction from maritime pine (*Pinus pinaster*) resin was quantified using High-Performance Liquid Chromatography coupled to a Diode Array Detector (HPLC-DAD). A methodology similar to the ones already reported was used [35,36]. Briefly, the plastic bag containing the resin was homogenized manually for several minutes until the turpentine and rosin were completely mixed. Then, 1 % w/v of the sample was diluted in Methanol. 10 µL of this solution was injected in a Thermo Scientific Ultimate 3000 chromatographic unit equipped with a Thermo Scientific LC Acclaim PolarAdvantage II C18 column (5 µm, 150 x 4.6 mm), and a Diode Array Detector. The separation was carried out at 20 °C with a binary gradient mixture of solvent A (methanol + 1.0 % formic acid) and solvent B (water + 1.0 % formic acid). A flow rate of 300 µl/min was used with a gradient

program as follow: 0-2 min (60 % of A); 2-17 min. (60-80 % of A); 17-34 min. (80 % of A); 34-48 min. (80-100 % of A); 48-67 min. (100 % of A); equilibrated (60 % of A).

The methodologies described by (Lee et al., 1997; Kersten et al., 2006) [22,23] were used for identification and quantification of the chemical components of the rosin fraction.

2.3 Near infrared (NIR) acquisition

Near infrared (NIR) acquisition was performed using a benchtop spectrometer and a handheld spectrometer, respectively, *MultiPurpose Analyzer I* (*Bruker*, USA) and *SCiO* (*Consumer Physics*, Israel).

MultiPurpose Analyzer I is a Fourier transform near infrared (FT-NIR) spectrometer implemented with an integrating sphere. The spectral range of acquisition was between 12,500 to 4,166 cm^{-1} (780 to 2400 *nm*) with a nominal resolution of 8 cm^{-1}. For the spectral acquisition, resin was poured in a cylindrical glass vial (h = 50 mm, outer diameter = 22 mm), and placed over the sphere window, with an average of 16 scans for each spectrum.

The SCiO handheld spectrometer [37] operated in a spectral range between 13,514 to 9,346 cm^{-1} (740 to 1070 *nm*) with values every 1 nm and an optical resolution of 30 nm [38]. The resin samples were scanned directly through the sealed plastic bag using an accessory provided by *Consumer Physics* to avoid ambient light interference, and also to ensure a constant distance of 10 mm between the light source and the sample. For each sealed plastic bag, the spectrum was calculated as the median of six measurements.

As the *SCiO* is a cloud-based device, *Consumer Physics* proposes a commercial online toolkit, which is required for spectra acquisition and download them. This online toolkit is also useful to create online mathematical models, which can be implemented in a mobile application using the *Software Development Kit* provided by *Consumer Physics*. Yet, this application was not used in the present study.

2.4 Multivariate Data Analysis

Fig. 1 details the process used during the Multivariate Data Analysis. The same methodology was applied on both spectrometers, in order to compare their quantitative predictive ability.

2.4.1 Preprocessing

During spectral data acquisition, scattering of photons leads to many physical phenomena (lengthening of the optical path, noise) that need to be corrected with spectra preprocessing algorithms [39] to limit the influence of physical phenomena on the relevant properties [40]. Such effects include baseline drift, nonlinearity, curvilinearity, additive and multiplicative effects, and irrelevant variations in the spectra [40]. Thus, several preprocessing techniques were applied in order to increase the accuracy of the predictive models: Standard Normal Variate (SNV), Detrend (DT2), combination of SNV and DT2 [41], Multiplicative Scatter Correction (MSC) [42], and Savitzky-Golay's first and second derivatives [43].

2.4.2 Subset Selection

Subset selection algorithms were used to extract calibration and independent validation subsets from a dataset [44]. The DUPLEX algorithm [45] promotes sample representativeness, ensuring that the selected spectra are uniformly distributed in the multidimensional data space and that the dependent variable has the same statistics characteristics (mean, variance, range, etc.).

In the first step, the two spectra most distant from each other are selected and placed in the calibration subset. On the remaining spectra, the following pair of spectra most distant from each other are selected for the validation subset. By alternating subsets, the procedure continues to select a single spectrum for each subset of data. The selected spectrum is the one farthest from the spectra already selected within the subset. Once the required number of

144 samples is reached in the validation subsets, the remaining samples are placed in the
145 calibration subset.
146 With this procedure, the dataset was split as follow: 70 % in the calibration subset, and 30%
147 in the validation subset.

148 *2.4.3 Modelling Phase*

149 Partial Least Squares (PLS) regression with Nonlinear Iterative Partial Least Squares
150 (NIPALS) algorithm is the favorite for regression in NIR applications, because of its ability to
151 analyze high-dimensional and multi-collinear data [46,47]. Its purpose is to use independent
152 variables (X) to predict a dependent variable (y): the spectra (X) are used to predict the
153 analytical reference values (y). To achieve this goal, it will attempt to meet three objectives:
154 (i) explain X; (ii) explain y; (iii) find relationships between X and y. Nonlinear Iterative
155 Partial Least Squares (NIPALS) has been used as algorithm to perform PLS.
156 In order to evaluate the quantitative predictive ability of Partial Least Squares (PLS) models,
157 several merit scores were computed. The coefficient of determination gave a measure of
158 linearity of the prediction based either on the calibration subset (R^2_{Cal}; R^2_{CV}), or on the
159 validation subset (R^2_{Val}). Root Mean Square Error of Calibration, Cross-Validation and
160 Prediction (respectively, RMSEC, RMSECV and RMSEP) were used to measure the accuracy
161 of the prediction calculated on both the calibration and validation subsets.
162 Another figure of merit, the Ratio of standard error of Performance to standard Deviation
163 (RPD), was computed to classify the applicability of the models. According to (Williams,
164 2014) [48], higher RPD values were associated with better quantitative prediction.
165 General rules emerge to classify the applicability of the models: with $2.5 \leq RPD < 3.0$, the
166 quantitative model makes approximate quantitative predictions (screening); with $3.0 \leq RPD <$
167 3.5, good quantitative predictions (quality control); for $3.5 \leq RPD < 4.1$, very good

quantitative predictions (process control); and RPD ≥ 4.1 indicates an excellent quantitative predictive model.

2.4.4 Software

Chemometric models construction was carried out using Matlab 2019a (Mathworks Inc., USA) with the SAISIR toolbox [49], and all the scripts were designed in-house.

3 Results and discussion

3.1 Chemical composition and % turpentine of maritime pine (Pinus pinaster) resin

Table 1 list the values of the studied parameters of maritime pine (*Pinus pinaster*) samples: main chemical components of the turpentine and rosin fractions, and proportion of the turpentine fraction.

Concerning the turpentine fraction, the results obtained via GC-FID showed that the analyzed turpentine of *Pinus pinaster* is mainly composed of unsaturated hydrocarbon monoterpenes, above all α-pinene and β-pinene: the α-pinene content was 66.2 ± 9.0 % and the β-pinene content was 18.5 ± 8.5 %. (Ghanmi et al., 2005) [13], comparing the turpentine content of maritime pine (*Pinus pinaster*) and of Aleppo pine (*Pinus halepensis*) from Morocco, report similar composition values for α-pinene and β-pinene as main components of turpentine: 78.7 % for α-pinene, and 7.3 % for β-pinene.

Concerning the rosin fraction, the results from HPLC-DAD show that the analyzed rosin of *Pinus pinaster* is mainly composed of diterpenic monocarboxylic acids, such as levopimaric acid, with a content of 62.1 ± 5.9 %. (Arrabal et al., 2005) [50] monitored the oleoresin chemical composition of 150 Spanish *Pinus pinaster* trees aged of 33 years old, from thirty sites. Authors reported similar composition values with levopimaric acid as the main constituent ranging from 41.7 to 45.3 %.

The turpentine content (*% turpentine*) was 34.3 ± 6.6 %, also in agreement with the literature: (Arrabal et al., 2005) [50] found *% turpentine* values ranging from 27.7 to 30.3 %.

3.2 NIR spectra

As reminded by (Y. Ozaki, 2012) [51], near infrared can be divided in three regions: region I (800 – 1200 *nm*; 12500 – 8500 cm^{-1}), region II (1200 – 1800 *nm*; 8500 – 5500 cm^{-1}) and region III (1800 – 2500 *nm*; 5500 – 4000 cm^{-1}). Those regions have specific spectral features, covering mainly electronic transitions, overtones and combination bands [52].

The benchtop *MPA I* spectrometer covers the 12,500 - 4,166 cm^{-1} range (780 to 2400 *nm*), which roughly corresponds to all these three regions. Along the spectra acquired with the *MPA I*, the CH first overtone, third overtone, deformation band and combination band; OH second overtone and $C = 0$ first and fourth overtone are observed.

The handheld *SCiO* spectrometer covers a narrow range from 13,514 to 9346 cm^{-1} (740 to 1070 *nm*), which covers the first region (Region I), also called "*near infrared*", where CH third overtone and OH second overtone are observed. In fact, the *SCiO's* capacity in the visible domain and the NIR domain are limited [32]. From this information, it is assumed that *MPA I* and *SCiO* spectrometers can consistently detect the same vibrational bands at different overtones.

Table 2 reports the vibrational bands of *Pinus pinaster* resin. The table was drafted based on the chemical structure of its major components, the spectra shapes, and the literature [53,54].

3.3 Modelling Phase

Statistics of Partial Least Squares (PLS) regression are presented in Table 3.

PLS regression was applied on calibration subsets to develop quantitative predictive models. In PLS modeling, Latent Variables (LVs) were calculated and considered as new eigenvectors

to reduce the dimensionality and compress the original spectral data. The number of LVs was determined as the one that provided the lowest RMSECV using a leave-one-out cross-validation. Once developed, the model was applied to validation subsets, which had not been used during the calibration procedures. Then, the predictive ability of each model was assessed by external validation on the validation subsets.

As a rule, a good model should have higher R^2_{Cal}, R^2_{CV}, R^2_{Val} and RPD values, and lower RMSEC, RMSECV and RMSEP. Here, the quantitative predictive ability of the model was selected based on highest RPD values.

3.3.1 Benchtop spectrometer: MultiPurpose Analyzer I (Bruker, USA)

Figure 2 (a) shows the raw spectra acquired with the *MultiPurpose Analyzer I*. Figure 2 (b to g) shows the preprocessed spectra. The raw and preprocessed spectra highlight many vibrational bands, which contain all the information needed for the modelling phase.

As seen in Table 3, all the parameters values of the coefficients of determination calculated between reference and predicted data on either calibration or validation subsets are excellent (R^2_{Cal} > 0.8; R^2_{CV} >0.8; R^2_{Val} > 0.8). In this case, PLS regression models predictions are well correlated to the references ones. RMSEC, RMSECV and RMSEP values are low in accordance with the reference data. In addition, they are close, which shows that the PLS regression models are robust. RPD statistics give poorest performance with PLS regression models computed without preprocessing (raw data). This means that spectral preprocessing is mandatory to improve the predictive ability of all models. Concerning the RPD values, all parameters are classified in two groups. In the first group, RPD ≥ 4.1, indicating excellent quantitative predictive models for α-pinene, β-pinene, and abietic and neoabietic acids. In the second group, the RPD ranges from 3.5 to 4.1, indicating very good quantitative predictions, for levopimaric acid and *% turpentine.*

Concerning the preprocessing, Detrend (DT2) gives better predictive ability for a mixture of abietic and neoabietic acids (RPD = 4.0); Savitzky–Golay first derivatives (SG1) give better predictive ability for α-pinene (RPD = 4.7) and levopimaric acid (RPD = 4.0); Savitzky–Golay second derivatives (SG2) give better predictive ability for β-pinene (RPD = 5.0) and *% turpentine* (RPD = 3.5).

3.3.2 Handheld spectrometer: SCiO (Consumer Physics, Israel)

Figure 3 (a) shows the raw spectra acquired with the *SCiO*. The raw spectra seem flat and did not exhibit any characteristic vibrational bands except for OH at 957 *nm* (10,449 cm^{-1}). This is seemingly due to physical phenomena, such as scattering effects and overlapping signals [55,56].

In practice, *SCiO* lacks repeatability and reproducibility. For example, the same sample area can be scanned (without touching or moving the device) several times, causing intensity shift. (Subedi *et al*., 2020) [38] is cited as a first attempt to test the repeatability of *SCiO*, and two other spectrometers: *MicroNIR* (*Viavi*, USA) and *F750* (*Felix Instruments*, USA). There the authors expressed the repeatability as the standard deviation of absorbance of the highest signal across the wavelength range monitored. Standard deviation was computed on 20 repeated measures of a PTFE reference material. Standard deviation was 0.04, 0.22 and 0.40 m Absorbance units for the *F750*, *MicroNIR* and *SCiO* spectrometers respectively. The values obtained with the *SCiO* were much higher than with the others two spectrometers. It is thus necessary, for each sample, to acquire several spectra, and to use a statistical method (*mean*, *median*) in order to have a representative spectrum, and also, to apply preprocessing algorithms. Once the spectra are preprocessed, vibrational bands are highlighted (Figure 3: b - g).

Similarly, with the PLS regression models developed for the benchtop spectrometer for all parameters, the values of coefficients of determination (R^2_{Cal}; R^2_{CV}; R^2_{Val}) show good

agreement between prediction and reference values. RMSEC, RMSECV and RMSEP values were higher than those obtained with the benchtop device, and show some discrepancy, suggesting that PLS regression models cannot be as robust as with the benchtop device. Looking at RPD statistics, spectral preprocessing is also required to improve the predictive ability of all models. As shown in Table 3, the five parameters can be classified into two groups. In the first group, the RPD ranges from 3.0 to 3.5, which indicate some good quantitative predictive ability for quality control. This group includes all major components of the resin, such as α-pinene, and levopimaric, abietic and neoabietic acids. In the second group, the RPD ranged between 2.5 and 3.0, i.e. β-pinene and *% turpentine* were less correctly predicted.

Concerning the preprocessing, Detrend (DT2) gives better predictive ability for the mixture of abietic and neoabietic acids (RPD = 3.01). Combination of Standard Normal Variate and Detrend (SNV + DT2) give better predictive ability for *% turpentine* (RPD = 3.03). Multiplicative Scatter Correction (MSC) gives better predictive ability for levopimaric acid (RPD = 3.09). Savitzky–Golay second derivatives (SG2) give better predictive ability for α-pinene (RPD = 3.16) and β-pinene (RPD = 2.82).

4 Conclusion

The *SCiO* handheld NIR spectrometer was studied as a potential tool to quantify the chemical components of maritime pine (*Pinus pinaster*) resin, in view of its use as a quality control tool for the tapping industry.

This study looked at the performance of the handheld *SCiO* spectrometer in comparison with a benchtop *MPA I* spectrometer, and evaluated the performance based on statistical parameters.

Although, *SCiO* and *MPA I* work in different NIR regions, PLS regression models can be used to quantify major resin components, such as, α-pinene, β-pinene, levopimaric, abietic

and neoabietic acids, and % *turpentine*. Results shown that spectral preprocessing is mandatory to increase the predictive ability of the PLS regression models.

Considering the selected PLS models: the coefficients of determination (R^2_{Cal}; R^2_{CV}; R^2_{Val}) show good agreement between prediction and reference values for both, *SCiO* and *MPA I* spectrometers with values higher than 0.8; RMSEC, RMSECV and RMSEP values were higher for the *SCiO* compared to the *MPA I*, and they also show some discrepancy, suggesting that PLS regression models for *SCiO* cannot be as robust as *MPA I* models; RPD statistics values were lower for the *SCiO* compared to the *MPA I* – the RPD values for *MPA I* showed an excellent quantitative predictive ability, whereas those for *SCiO* have a good quantitative predictive ability for quality control purposes.

Considering the overall instrument parameters (wavelength, resolution, size, cost), the *SCiO* spectrometer showed very good results. Additionally, the *SCiO* is a handheld spectrometer, and can be easily used on-site. Thus, the *SCiO* could be a useful tool to predict main chemical components of maritime pine (*Pinus pinaster*) for quality control in the tapping industry.

Acknowledgements

The authors gratefully acknowledge the financial support from the Nouvelle Aquitaine regional council, the Landes departmental council, the Agence Nationale de la Recherche (National Agency for Research) and Xyloforest (ANR-10-EQPX-16). The authors warmly thank Luc Leneveu from the Holiste company, who provided the raw material, and Dr. Camille Lepoittevin for hosting Lisa Feuillerat during the data acquisition work on the *MultiPurpose Analyzer I* spectrometer (INRA, Pierroton, France).

Declaration of competing interest

The authors declare no actual or potential conflicts of interests.

References

[1] C. Courau, Le gemmage en forêt de Gascogne, 2014.

[2] B. Lemoine, N. Decourt, Tables de production pour le pin maritime dans le sud-ouest de la France, Rev. For. Française. (1969) 5. doi:10.4267/2042/20235.

[3] J.-C. Bussy, La gemme et les produits résineux en France, Rev. For. Française. 284 (1971) 377. doi:10.4267/2042/20503.

[4] A. Rodríguez-García, J.A. Martín, R. López, A. Sanz, L. Gil, Effect of four tapping methods on anatomical traits and resin yield in Maritime pine (Pinus pinaster Ait.), Ind. Crops Prod. 86 (2016) 143–154. doi:10.1016/j.indcrop.2016.03.033.

[5] M. Castillo Martos, Construir la tecnología: el caso de la resina de pino en Francia, siglos XVIII y XIX, Llull Rev. La Soc. Española Hist. Las Ciencias y Las Técnicas. 26 (2003) 1061–1066.

[6] C. Courau, La relance du gemmage en forêt de Gascogne, 2009.

[7] C. Courau, Procedé de traitement de la résine, dispositif pour la mise en œuvre de ce procedé et colophane obtenue, 1996.

[8] L. Leneveu, Procédé pour favoriser l'exsudation de l'oleorésine et composition pour mettre en œuvre ce procédé, 2012.

[9] M. Stoffel, M. Klinkmüller, 3D analysis of anatomical reactions in conifers after mechanical wounding: First qualitative insights from X-ray computed tomography, Trees - Struct. Funct. 27 (2013) 1805–1811. doi:10.1007/s00468-013-0900-2.

[10] F.A. Neis, F. de Costa, M.R. de Almeida, L.C. Colling, C.F. de Oliveira Junkes, J.P. Fett, A.G. Fett-Neto, Resin exudation profile, chemical composition, and secretory canal characterization in contrasting yield phenotypes of Pinus elliottii Engelm, Ind. Crops Prod. 132 (2019) 76–83. doi:10.1016/j.indcrop.2019.02.013.

[11] A.J.D. Silvestre, A. Gandini, Chapter 2 - Terpenes: Major Sources, Properties and

337 Applications, in: Elsevier, 2008: pp. 17–38. doi:https://doi.org/10.1016/B978-0-08-
338 045316-3.00002-8.
339 [12] A.J.D. Silvestre, A. Gandini, Chapter 4 - Rosin: Major Sources, Properties and
340 Applications, in: Elsevier, 2008: pp. 67–88. doi:https://doi.org/10.1016/B978-0-08-
341 045316-3.00004-1.
342 [13] M. Ghanmi, A. El Abid, A. Chaouch, A. Aafi, M. Aberchane, A. El Alami, A. Farah,
343 Étude du rendement et de la composition de l'essence de térébenthine du Maroc: Cas
344 du Pin maritime (Pinus pinaster) et du Pin d'Alep (Pinus halepensis), Acta Bot. Gall.
345 152 (2005) 3–10. doi:10.1080/12538078.2005.10515450.
346 [14] T.L. Eberhardt, P.M. Sheridan, J.M. Mahfouz, Monoterpene persistence in the
347 sapwood and heartwood of longleaf pine stumps: Assessment of differences in
348 composition and stability under field conditions, Can. J. For. Res. 39 (2009) 1357–
349 1365. doi:10.1139/X09-063.
350 [15] H. Wang, B. Liu, X. Liu, J. Zhang, M. Xian, Synthesis of biobased epoxy and curing
351 agents using rosin and the study of cure reactions, Green Chem. 10 (2008) 1190–1196.
352 doi:10.1039/b803295e.
353 [16] Fao, Gum naval stores: turpentine and rosin from pine resin, 1995.
354 [17] M. Ghanmi, B. Satrani, A. Aafi, R. Ismail, A. Farah, A. Chaouch, R. Ismaili,
355 Évaluation de la qualité de la colophane du pin maritime (Pinus pinaster) et du pin
356 d'Alep (Pinus halepensis) du Maroc Évaluation de la qualité de la colophane du pin
357 maritime (Pinus pinas-ter) et du pin d'Alep (Pinus halepensis) du Maroc, Acta Bot.
358 Gall. 156 (2009) 427–435. doi:10.1080/12538078.2009.10516168.
359 [18] M.J. Lombardero, J. Pereira-Espinel, M.P. Ayres, Foliar terpene chemistry of Pinus
360 pinaster and P. radiata responds differently to Methyl Jasmonate and feeding by larvae
361 of the pine processionary moth, For. Ecol. Manage. 310 (2013) 935–943.

362 doi:https://doi.org/10.1016/j.foreco.2013.09.048.
363 [19] İ. Tümen, E.K. Akkol, H. Taştan, I. Süntar, M. Kurtca, Research on the antioxidant,
364 wound healing, and anti-inflammatory activities and the phytochemical composition of
365 maritime pine (Pinus pinaster Ait), J. Ethnopharmacol. 211 (2018) 235–246.
366 doi:https://doi.org/10.1016/j.jep.2017.09.009.
367 [20] E. Gonçalves, A.C. Figueiredo, J.G. Barroso, J. Henriques, E. Sousa, L. Bonifácio,
368 Effect of Monochamus galloprovincialis feeding on Pinus pinaster and Pinus pinea,
369 oleoresin and insect volatiles, Phytochemistry. 169 (2020) 112159.
370 doi:https://doi.org/10.1016/j.phytochem.2019.112159.
371 [21] M. Lai, L. Zhang, L. Lei, S. Liu, T. Jia, M. Yi, Inheritance of resin yield and main resin
372 components in Pinus elliottii Engelm. at three locations in southern China, Ind. Crops
373 Prod. 144 (2020) 112065. doi:https://doi.org/10.1016/j.indcrop.2019.112065.
374 [22] B.L. Lee, D. Koh, H.Y. Ong, C.N. Ong, High-performance liquid chromatographic
375 determination of dehydroabietic and abietic acids in traditional Chinese medications,
376 in: J. Chromatogr. A, 1997: pp. 221–226. doi:10.1016/S0021-9673(96)00901-6.
377 [23] P.J. Kersten, B.J. Kopper, K.F. Raffa, B.L. Illman, Rapid analysis of abietanes in
378 conifers, J. Chem. Ecol. 32 (2006) 2679–2685. doi:10.1007/s10886-006-9191-z.
379 [24] B.K. Via, C. Zhou, G. Acquah, W. Jiang, L. Eckhardt, Near infrared spectroscopy
380 calibration for wood chemistry: Which chemometric technique is best for prediction
381 and interpretation?, Sensors (Switzerland). 14 (2014) 13532–13547.
382 doi:10.3390/s140813532.
383 [25] J. Sandak, A. Sandak, R. Meder, Assessing trees, wood and derived products with near
384 infrared spectroscopy: Hints and tips, in: J. Near Infrared Spectrosc., IM Publications
385 LLP, 2016: pp. 485–505. doi:10.1255/jnirs.1255.
386 [26] R.A. Crocombe, Portable Spectroscopy., Appl. Spectrosc. 72 (2018) 1701–1751.

387 doi:10.1177/0003702818809719.
388 [27] A.V. Morillas, J. Gooch, N. Frascione, Feasibility of a handheld near infrared device
389 for the qualitative analysis of bloodstains, Talanta. 184 (2018) 1–6.
390 doi:10.1016/j.talanta.2018.02.110.
391 [28] V. Wiedemair, D. Langore, R. Garsleitner, K. Dillinger, C. Huck, V. Wiedemair, D.
392 Langore, R. Garsleitner, K. Dillinger, C. Huck, Investigations into the Performance of
393 a Novel Pocket-Sized Near-Infrared Spectrometer for Cheese Analysis, Molecules. 24
394 (2019) 428. doi:10.3390/molecules24030428.
395 [29] B.K. Wilson, H. Kaur, E.L. Allan, A. Lozama, D. Bell, A New Handheld Device for
396 the Detection of Falsified Medicines: Demonstration on Falsified Artemisinin-Based
397 Therapies from the Field, Am. J. Trop. Med. Hyg. 96 (2017) 1117–1123.
398 doi:10.4269/ajtmh.16-0904.
399 [30] F. Kosmowski, T. Worku, Evaluation of a miniaturized NIR spectrometer for cultivar
400 identification: The case of barley, chickpea and sorghum in Ethiopia, PLoS One. 13
401 (2018) e0193620. doi:10.1371/journal.pone.0193620.
402 [31] V. Wiedemair, C.W. Huck, Evaluation of the performance of three hand-held near-
403 infrared spectrometer through investigation of total antioxidant capacity in gluten-free
404 grains, Talanta. 189 (2018) 233–240. doi:10.1016/j.talanta.2018.06.056.
405 [32] M. Li, Z. Qian, B. Shi, J. Medlicott, A. East, Evaluating the performance of a consumer
406 scale SCiO^{TM} molecular sensor to predict quality of horticultural products, Postharvest
407 Biol. Technol. 145 (2018) 183–192. doi:10.1016/j.postharvbio.2018.07.009.
408 [33] G. Drive, A. Suite, C. Stream, C.M. Spectrometry, R.P. Adams, Identification of
409 Essential Oil Components by Gas Review : Identification of Essential Oil Components
410 by Gas Chromatography / Mass Spectrometry, (2005).
411 [34] A. D6730-19, Standard Test Method for Determination of Individual Components in

Spark Ignition Engine Fuels by 100-Metre Capillary (with Precolumn) High-Resolution Gas Chromatography, West Conshohocken, 2019. https://www.astm.org/Standards/D6730.htm (accessed December 27, 2019).

[35] T. Cabaret, Y. Gardere, M. Frances, L. Leroyer, B. Charrier, Measuring interactions between rosin and turpentine during the drying process for a better understanding of exudation in maritime pine wood used as outdoor siding, Ind. Crops Prod. 130 (2019) 325–331. doi:10.1016/j.indcrop.2018.12.080.

[36] T. Cabaret, F. Mariet, K. Li, L. Leroyer, B. Charrier, High temperature drying effect against resin exudation for maritime pine wood used as outdoor siding, Eur. J. Wood Wood Prod. 77 (2019) 673–680. doi:10.1007/s00107-019-01425-8.

[37] S. Rosen, D. Goldring, D. Sharon, U. Kinrot, I. Bakish, N. Ittai, Spectrometry system applications, 2017.

[38] P.P. Subedi, K.B. Walsh, Assessment of avocado fruit dry matter content using portable near infrared spectroscopy: Method and instrumentation optimisation, Postharvest Biol. Technol. 161 (2020) 111078. doi:https://doi.org/10.1016/j.postharvbio.2019.111078.

[39] M. Zeaiter, D. Rutledge, Preprocessing Methods, in: Compr. Chemom., Elsevier, 2009: pp. 121–231. doi:10.1016/B978-044452701-1.00074-0.

[40] Å. Rinnan, F. van den Berg, S.B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, TrAC Trends Anal. Chem. 28 (2009) 1201–1222. doi:10.1016/j.trac.2009.07.007.

[41] 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. doi:10.1366/0003702894202201.

[42] I.S. Helland, T. Næs, T. Isaksson, Related versions of the multiplicative scatter

correction method for preprocessing spectroscopic data, Chemom. Intell. Lab. Syst. 29 (1995) 233–241. doi:10.1016/0169-7439(95)80098-T.

[43] H. Mark, J. Workman, Chapter 57 - Derivatives in Spectroscopy: Part 3—Computing the Derivative (the Savitzky-Golay Method)☆, in: H. Mark, J.B.T.-C. in S. (Second E. Workman (Eds.), Academic Press, 2018: pp. 371–381. doi:https://doi.org/10.1016/B978-0-12-805309-6.00057-X.

[44] J.-R. Bouveresse, Maalouly, Jaillais, Sélection d'échantillons représentatifs par des méthodes chimiométriques. Application à des modèles d'étalonnage, Spectra Anal. 33 (2004).

[45] R.D. Snee, Validation of Regression Models: Methods and Examples, Technometrics. (1977). doi:10.1080/00401706.1977.10489581.

[46] H. Wold, Estimation of Principal Components and Related Models by Iterative Least squares, Acad. Press. New York. (1966) 391–420. doi:[1] E. Devices, D.E. Mccumber, A.G. Chynoweth, A.G. Foyt, B. Elschner, M. Schlaak, References 1., 24 (1967) 10–12.

[47] S. Wold, H. Martens, H. Wold, The multivariate calibration problem in chemistry solved by the PLS method, in: Springer, Berlin, Heidelberg, 1983: pp. 286–293. doi:10.1007/BFb0062108.

[48] P. Williams, The *RPD* Statistic: A Tutorial Note, NIR News. 25 (2014) 22–26. doi:10.1255/nirn.1419.

[49] C. Cordella, D. Bertrand, SAISIR: A new general chemometric toolbox, TrAC - Trends Anal. Chem. 54 (2014) 75–82. doi:10.1016/j.trac.2013.10.009.

[50] C. Arrabal, M. Cortijo, B.F. de Simón, M.C. García Vallejo, E. Cadahía, Differentiation among five Spanish Pinus pinaster provenances based on its oleoresin terpenic composition, Biochem. Syst. Ecol. 33 (2005) 1007–1016. doi:https://doi.org/10.1016/j.bse.2005.03.003.

462 [51] Y. Ozaki, Near-infrared spectroscopy-its versatility in analytical chemistry, Anal. Sci.
463 28 (2012) 545–563. doi:10.2116/analsci.28.545.
464 [52] V. Wiedemair, C.W. Huck, Evaluation of the performance of three hand-held near-
465 infrared spectrometer through investigation of total antioxidant capacity in gluten-free
466 grains, Talanta. 189 (2018) 233–240. doi:10.1016/j.talanta.2018.06.056.
467 [53] M. Schwanninger, J.C. Rodrigues, K. Fackler, A review of band assignments in near
468 infrared spectra of wood and wood components, J. Near Infrared Spectrosc. 19 (2011)
469 287–308. doi:10.1255/jnirs.955.
470 [54] L. Workman, J. Weyer, Practical Guide and Spectral Atlas for Interpretive Near-
471 Infrared Spectroscopy, CRC Press, 2012. doi:10.1201/b11894.
472 [55] M. Blanco, I. Villarroya, NIR spectroscopy: A rapid-response analytical tool, TrAC -
473 Trends Anal. Chem. 21 (2002) 240–250. doi:10.1016/S0165-9936(02)00404-1.
474 [56] Y.B. Ma, K.S. Babu, J.K. Amamcharla, Prediction of total protein and intact casein in
475 cheddar cheese using a low-cost handheld short-wave near-infrared spectrometer, Lwt.
476 (2019). doi:10.1016/j.lwt.2019.04.039.
477

Figure 1 Detailed overview of the Multivariate Data Analysis.

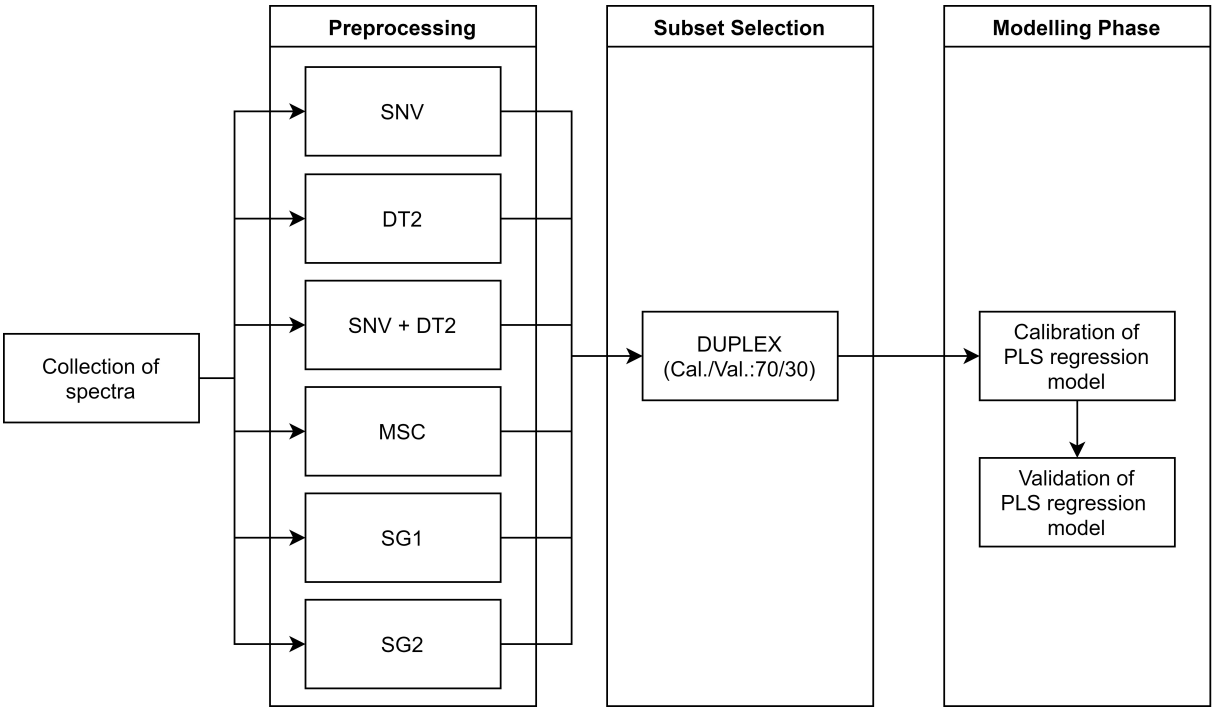

Figure 2 Different preprocessing algorithms were applied to the raw spectra acquired with *SCiO* (*Consumer Physics*, Israel). (a) Raw spectra; (b) SNV preprocessed spectra; (c) DT2 preprocessed spectra; (d) SNV + DT2 preprocessed spectra; (e) MSC preprocessed spectra; (f) SG1 preprocessed spectra; (g) SG2 preprocessed spectra.

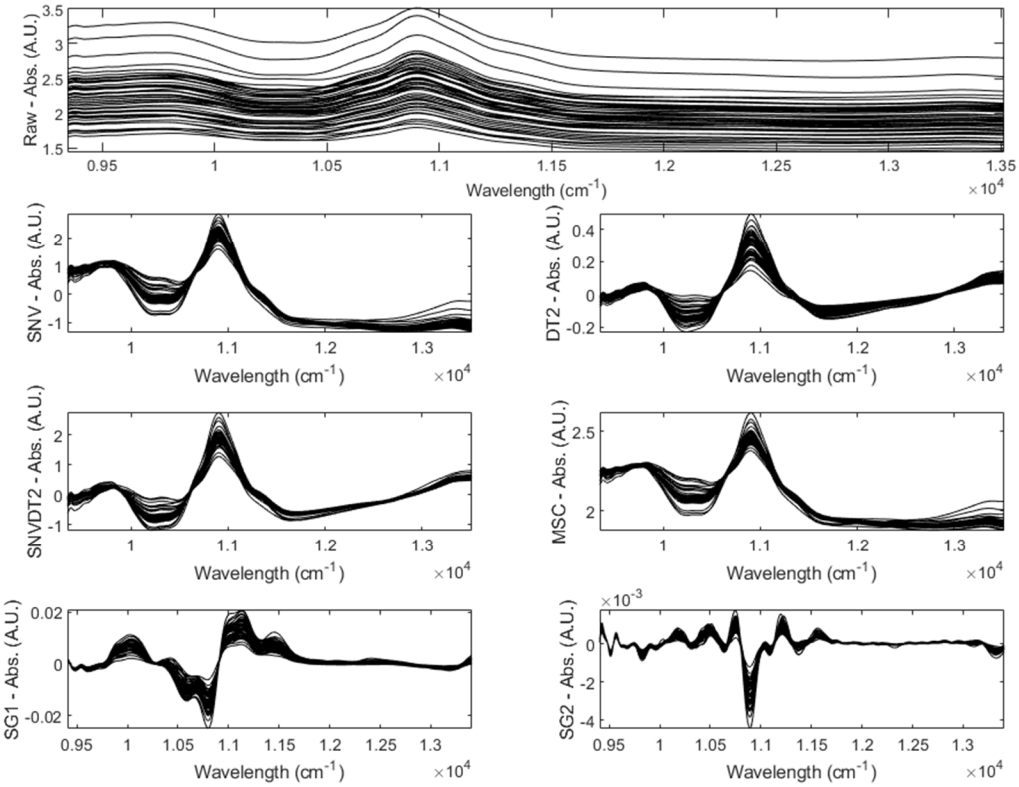

Figure 3 Different preprocessing algorithms were applied to the raw spectra acquired with *MultiPurpose Analyzer I* (*Bruker*, USA). (a) Raw spectra; (b) SNV preprocessed spectra; (c) DT2 preprocessed spectra; (d) SNV + DT2 preprocessed spectra; (e) MSC preprocessed spectra; (f) SG1 preprocessed spectra; (g) SG2 preprocessed spectra.

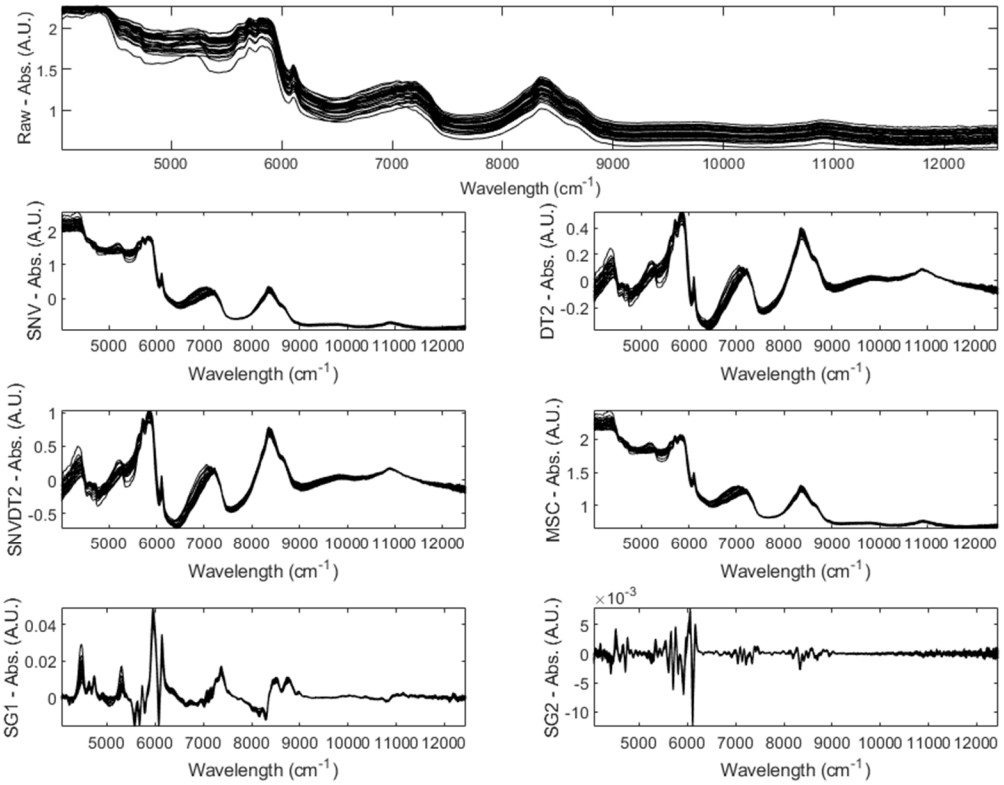

Table 1 Values of the parameters of the analyzed samples of maritime pine (*Pinus pinaster*): main chemical components of the turpentine and rosin fractions, and proportion of the turpentine fraction.

	Parameters	Number of samples	Minimum (%)	Maximum (%)	Mean (%)	Standard deviation (%)
Turpentine Chemical Composition	α-pinene	94	44.05	84.37	66.16	9.01
	β-pinene		3.24	37.47	18.47	8.45
	Myrcene or/and δ-3 carene		0.04	0.52	0.28	0.11
	Limonene		0.88	6.12	4.08	0.68
	ɣ-terpinene		0.64	2.19	1.42	0.30
Rosin Chemical Composition	Levopimaric acid	49	49.66	75.25	62.10	5.93
	Abietic and Neoabietic acids		24.75	50.34	37.90	5.93
% Turpentine	% Turpentine	112	15	52.8	34.25	6.58

Table 2 Band assignments for the chemical components of maritime pine (*Pinus Pinaster*) resin.

Device	Wavenumber		Functional group	Possible band assignments for:	
	(cm^{-1})	(nm)		Turpentine constituents	Rosin constituents
MultiPurpose Analyzer I	10891	918	CH methylene (CH_2)		
	9889	1011	OH from tertiary alcohols $(C-OH)$		
	8662	1154	$C=O$ from Carbonyl $(C=O)$		
	7211	1387	CH methyl associated with aromatic $(Ar-CH_3)$		
	6911	1447	CH aromatic (Ar)		
	6101	1639	CH from vinyl group as $(CH_2=CH-)$		
	5816	1719	CH methyl (CH_3)		
	5708	1752	CH methylene (CH_2)		
	5600	1786	CH methylene (CH_2)		
	5222	1915	$C=O$ (carbonyl) from acid $(C=OOH)$		
	4535	2205	CH aromatic (Aryl)		
	4420	2262	CH (CH Bending)		
SCiO	11074	903	CH methylene (CH_2)		
	10787	927	CH methylene (CH_2)		
	10449	957	OH Alkyl Alcohols $(R-C-OH)$		
	9843	1016	OH from tertiary alcohols as $(-C-OH)$		
	9443	1059	OH with hydrogen bonding $(R-C-OH)$		

Table 3 Statistics of Partial Least Squares (PLS) regression models.

	Preprocessing	*SCiO* (*Consumer Physics*, Israel)								*MultiPurpose Analyzer I* (*Bruker*, USA)							
		LVs	R^2_{Cal}	R^2_{CV}	R^2_{Val}	*RMSEC*	*RMSECV*	*RMSEP*	*RPD*	LVs	R^2_{Cal}	R^2_{CV}	R^2_{Val}	*RMSEC*	*RMSECV*	*RMSEP*	*RPD*
α-pinene	RAW	9	0,87	0,83	0,46	3,33	3,71	5,57	1,23	5	0,86	0,87	0,81	3,28	3,25	3,81	2,33
	SNV	10	0,90	0,86	0,64	2,71	3,33	6,57	1,53	4	0,90	0,89	0,89	2,88	2,92	3,10	2,87
	DT2	10	0,93	0,92	0,85	2,36	2,57	4,13	2,28	6	0,94	0,92	0,94	1,99	2,54	2,85	3,63
	SNV + DT2	12	0,91	0,92	0,89	2,40	2,57	3,49	3,09	6	0,94	0,94	0,92	2,10	2,13	2,79	3,42
	MSC	9	0,86	0,86	0,81	3,31	3,38	3,96	2,35	4	0,91	0,90	0,91	2,79	2,87	3,03	2,82
	SG1	8	0,95	0,92	0,76	1,95	2,47	4,56	1,93	**6**	**0,97**	**0,97**	**0,95**	**1,53**	**1,66**	**2,00**	**4,70**
	SG2	**5**	**0,93**	**0,92**	**0,91**	**2,26**	**2,47**	**3,24**	**3,16**	4	0,96	0,95	0,94	1,72	1,91	2,44	4,16
β-pinene	RAW	10	0,92	0,87	0,47	2,35	3,07	5,49	1,35	10	0,96	0,96	0,94	1,61	1,78	2,19	3,69
	SNV	10	0,90	0,88	0,52	2,70	2,95	7,63	1,11	9	0,97	0,96	0,94	1,47	1,59	2,01	4,12
	DT2	8	0,89	0,90	0,85	2,71	2,70	3,64	2,52	5	0,88	0,88	0,91	2,71	2,90	3,13	3,11
	SNV + DT2	9	0,90	0,89	0,86	2,39	2,75	3,99	2,54	7	0,96	0,94	0,93	1,69	2,08	2,38	3,83
	MSC	9	0,89	0,87	0,84	2,58	2,99	3,88	2,42	8	0,96	0,95	0,95	1,62	1,91	1,97	4,45
	SG1	10	0,96	0,95	0,85	1,62	1,91	3,29	2,66	8	0,99	0,98	0,89	0,78	1,13	2,99	3,09
	SG2	**4**	**0,93**	**0,92**	**0,87**	**2,16**	**2,45**	**3,26**	**2,82**	**4**	**0,96**	**0,96**	**0,96**	**1,66**	**1,79**	**1,91**	**4,96**
Levopimaric acid	RAW	6	0,84	0,86	0,81	3,77	4,27	3,45	1,98	9	0,92	0,92	0,92	2,48	2,67	2,60	3,42
	SNV	6	0,80	0,84	0,85	3,86	4,22	3,99	2,51	7	0,93	0,93	0,94	2,38	2,76	2,29	3,87
	DT2	5	0,82	0,84	0,87	3,71	4,25	3,70	2,55	9	0,95	0,94	0,94	1,95	2,15	2,72	3,81
	SNV + DT2	4	0,76	0,78	0,88	4,01	4,09	3,77	2,85	5	0,89	0,90	0,83	2,90	2,88	4,26	2,24
	MSC	**6**	**0,78**	**0,81**	**0,89**	**4,09**	**4,66**	**3,01**	**3,09**	6	0,90	0,90	0,95	2,89	2,80	2,27	3,75
	SG1	6	0,89	0,93	0,79	2,89	3,04	4,14	2,13	**6**	**0,93**	**0,93**	**0,94**	**2,43**	**2,65**	**2,38**	**3,95**
	SG2	5	0,85	0,86	0,86	3,26	3,69	3,93	2,60	4	0,90	0,91	0,94	2,69	2,91	2,72	3,72
Abietic and neoabietic acids	RAW	7	0,86	0,91	0,86	3,50	3,61	3,00	2,28	8	0,93	0,93	0,94	2,29	2,59	2,34	3,80
	SNV	6	0,72	0,73	0,59	4,76	5,39	6,63	1,51	**9**	**0,94**	**0,94**	**0,96**	**2,17**	**2,53**	**2,06**	**4,30**
	DT2	**8**	**0,86**	**0,88**	**0,90**	**3,23**	**3,53**	**3,13**	**3,01**	7	0,95	0,95	0,95	1,87	1,98	2,59	3,99
	SNV + DT2	3	0,28	0,29	0,65	7,08	7,44	7,34	1,47	7	0,95	0,96	0,94	2,06	2,18	2,24	4,25
	MSC	7	0,61	0,61	0,82	6,56	6,89	3,98	2,34	8	0,94	0,95	0,95	2,26	2,57	2,17	3,92
	SG1	6	0,91	0,91	0,84	2,67	2,62	3,52	2,51	4	0,90	0,91	0,90	2,88	3,04	2,94	3,19
	SG2	3	0,85	0,86	0,88	3,22	3,78	3,56	2,87	3	0,90	0,90	0,94	2,66	2,96	2,64	3,83
% Turpentine	RAW	7	0,89	0,91	0,79	3,09	3,25	3,35	2,04	5	0,91	0,92	0,90	2,63	2,58	2,82	3,16
	SNV	6	0,85	0,85	0,68	3,37	3,92	6,87	1,46	6	0,89	0,90	0,87	2,98	3,11	3,24	2,74
	DT2	8	0,88	0,90	0,90	3,07	3,34	3,17	2,98	7	0,92	0,92	0,90	2,42	2,72	3,42	3,02
	SNV + DT2	**5**	**0,78**	**0,82**	**0,89**	**3,86**	**4,06**	**3,56**	**3,03**	6	0,90	0,91	0,88	2,78	3,04	3,30	2,89
	MSC	7	0,73	0,79	0,78	4,98	5,25	4,67	1,99	6	0,89	0,91	0,92	3,09	3,42	2,96	2,88
	SG1	4	0,90	0,92	0,83	2,87	3,29	3,56	2,48	5	0,89	0,91	0,91	3,03	3,20	2,73	3,44
	SG2	5	0,87	0,88	0,89	3,05	3,28	3,42	2,99	**4**	**0,89**	**0,91**	**0,92**	**2,86**	**3,07**	**2,94**	**3,45**

Graphical Abstract