

Contents

1	1 Methods	2
2	1 Sampling of trees and heartwood	3
3	2 Extraction	3
4	2.1 Aim	3
5	2.2 Drying and avoiding extractive oxidation	4
6	2.3 Sample geometry	6
7	2.4 Extraction method and apparatus	6
8	3 Moisture determination	10
9	4 Gas Chromatography	11
10	4.1 Theory and Apparatus	11
11	4.2 Quantification	14
12	5 Vibrational Spectroscopy	15
13	5.1 Theory	15
14	5.2 Imaging	20
15	5.3 Point spectra	24
16	6 Analysis of Spectral Data	26
17	6.1 Data structure	27
18	6.2 Pre-processing and exploratory analysis	28
19	6.3 Multivariate Analysis Tools	33
20	References	41
21	A PAPER I	53
22	B PAPER II	54
23	C PAPER III	55
24	D PAPER IV	57
25	E PAPER V	58

²⁷	F Unpublished Extractives paper	59
---------------	--	-----------

²⁸ **Todo list**

29	■ check where all the abbreviations are introduced the first time - are there multiples?	1
30	■ unfortunately the pictures of the nitrogen shower were lost with a phone	5
31	■ find a source for Tg definitino // and Tg of 70 for hemicellulose??	9
32	■ in paperI (gc-ms) and paperIV(oxalate) the wrong water temperature was written!! - does this need to be corrected in the journal?	9
33	■ "and references therein" - enough or should they be mentioned explicitly? (ref214-219, noted in outcommented text)	19
34	■ @Lisbeth: could you please check if this is mentioned in the book you got with the Raman instrument? I cannot find this info in any of my sources...	31
35	■ cannot check, have no access to matlab	35
36	■ find out why this is not bold	37
37	■ in figure 4, one peak of the lignin BP has the wrong wavenumber - it should be 929, not 829	56

Contents

1

43
44

2

check where
all the abbreviations are
introduced
the first time
- are there
multiples?

⁴⁵ Chapter 1

⁴⁶ Methods

Tree	Total height	Age	Diameter at 1.3 m \year rings JW	Diameter at 6 m \year rings JW	Diameter at 11 m \year rings JW
spruce ist hier klein geschrieben, Larch groß					
spruce 1	28.9 m	48 y	32.3 cm \10	29.2 cm \10	26.7 cm \10
spruce 2	28.2 m	48 y	27.9 cm \10	26.7 cm \10	22.2 cm \10
Larch 1	29.5 m	65 y	38.1 cm \14-16	36.1 cm \14	30.5 cm \10
Larch 2	29.0 m	65 y	40.6 cm \20	34.8 cm \15	31.2 cm \7-15

Table 1.1: Summary of meta-data on the sampled spruce and larch trees and number of year rings considered to be juvenile wood (JW).

The following chapter describes all methodological considerations taken for this project, including the sampling of trees and sample preparation. It gives a theoretical background and practical aspects of all the methods used, which includes physical/chemical methods used in the laboratory, as well as data analytical methods. Since Raman planning

"This thesis aims at Raman imaging as its main method, therefore many aspects in its planning take this into consideration."

... sorry, mir fällt gerade nix wirklich viel besseres ein als daß ich meine, ich möchte diesen Satz umstellen

1 Sampling of trees and heartwood

Because large variations were expected at the level of Raman parameters potentially introducing variation were tentatively the trees felled for the following experiments were obtained from two near-by stands in the same forest, providing similar light, soil and climatic conditions. Furthermore, the pair of spruce and larch trees were genetically

tentatively.
[tentatively' bedeutet in erster Linie sowas wie 'zögerlich', ist das tatsächlich gemeint? Eher willt 'as best as possible'?

To follow the extractive distribution (Zeitform: 'dealt with'
ggt besser 'elaborated [in Chapter...]'? (dealed with in Chapter ??), discs of up to 200 mm thickness were sown out not only at 1.3 m (used for all publications), but also at 6 m and 11 m stem height. Age, height and diameter of all sample discs can be found in Table 1.1.

The heartwood of the individual slices were selected as described in Paper I (Appendix oben 'year rings'
hier 'annual rings') in Figure 1.1. Table 1.1 additionally shows the number of annual rings considered as juvenile heartwood for all trees and heights. Note that the heartwood of spruce 2 at 11 m could not be separated into juvenile and mature heartwood, due to the small diameter of the stem and resulting lack of material.

2 Extraction

2.1 Aim

The goal of the extraction of the heartwood extractives from Norway spruce and Kurile larch specimens was to generate samples lacking the hydrophobic or the hydrophilic part of the extractives, or all of them and that could be sectioned for Raman microscopy. To

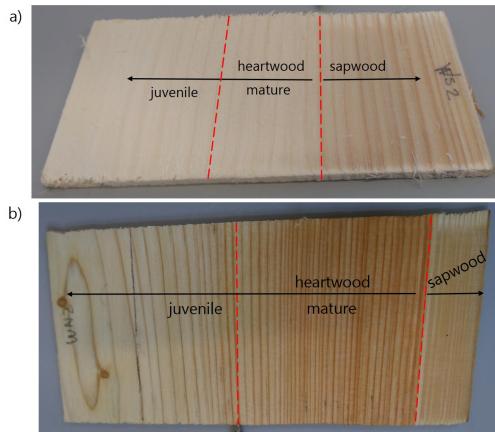


Figure 1.1: Mature heartwood is selected from **a)** Norway spruce and **b)** Kurile larch via counting the number of rings of juvenile wood, and looking at the color difference between sapwood and heartwood.

73 minimize a possible bias in the data caused by the extraction, the following requirements
74 to the extraction procedure were determined:

- 75 • **maximal extraction efficiency** – To remove all (or a big part) of extractives
76 from the heartwood, extraction efficiency should be maximized, which can be
77 tuned by the choice of the sample geometry, the drying strategy and the choice of
78 extraction conditions (solvent(s), temperature and time).
- 79 • **maximal selectivity** – To remove specific parts (hydrophobic/hydrophilic) of
80 the extractives, extraction should be selective, which is mainly achievable via the
81 choice of the extraction solvent(s).
- 82 • **Non- destructive**: To obtain wood samples with minimal changes of the struc-
83 tural polymers, the extraction procedure should be as mild as possible. Since
84 extractives were to be characterized as well, the extraction procedure and pro-
85 cessing also aimed at avoiding the oxidation of extractives.

86 The recommendations given in Holmbom (1999) and Willför et al. (2006) were largely
87 followed for much of the procedure development.

88 2.2 Drying and avoiding extractive oxidation

89 Because the drying of the wood samples was unavoidable for various reasons (e.g.,
90 determination of extraction yields and weight loss after fungal degradation), care was
91 taken to use methods that would minimize structural alterations of the wood due to
92 enzymatic activity and that would avoid the oxidation of extractives (Holmbom, 1999).

93 Firstly, the wooden disks sampled in the forest were packed into plastic bags right
94 after harvest and frozen within 10 hours, as often described in the literature (Willför
95 et al., 2003a, 2004; Zule et al., 2015). Unfortunately, the freezing process in a large
96 freezer at -20 °C was relatively slow (> 24h) due to the large amount of wood. During
97 the sawing of slices from the disks, and also cutting of sticks and chips from the slices
98 and milling of wood chips (Figure 1 in Paper I, Appendix A) samples were kept as cold
99 as possible using a walk-in freezer, cooling-boxes with ice during transport and thermal
100 packs as underlayment, respectively.

101 The work of Willför et al. (2006) made it clear that already the drying procedure
102 affected the types of extractives obtained from thermo-mechanical pulp (TMP) of spruce
103 after different drying procedures. Freeze-drying appeared to give the highest yields for
104 TMP (Willför et al., 2006). Samples were therefore frozen at -80 °C in preparation
105 for freeze-drying (lyophilization). Freeze-drying is a procedure where frozen water in
106 the sample is sublimated to its gaseous phase by a strong reduction of pressure, that
107 is the application of vacuum. It has the advantage that it may avoid the movement
108 of liquid water inside the tissue¹, but also avoiding exposure of the wood to high
109 temperatures and oxygen, which has been shown to reduce extraction yield in pulps
110 (Willför et al., 2006). A considerable disadvantage was found to be the movement of
111 resin, as the transversal planes of the sticks were often found to be sticky. Most likely
112 this resin originated from larger chambers (i.e.,resin channels and ray cells) and therefore
113 should not have affected the micro-distribution within the cell wall as found in Paper
114 II (Appendix B).

115 Because sample dryness is especially important when using hydrophobic solvents
116 (Holmbom, 1999), preliminary experiments on the duration of drying were conducted.
117 That is, the weight loss of sticks and milled wood samples was compared after at least
118 30 hours of freeze-drying and showed that milled wood was approx. 7% more dry than
119 sticks. This could be one of the factors that affected the sticks' lower yields (see Paper
120 I, Appendix A). As a side effect, this experiment showed that the moisture content of
121 spruce heartwood was 34-37 % and a bit lower for larch with 28-30 %. Wood samples
122 were stored under vacuum in a desiccator with fresh molecular sieves until usage, which
123 was not only relevant for weighing, but also for extraction efficiency with hydrophobic
124 solvents.

125 Since at least some extractive groups are prone to oxidation, the removal of extraction
126 solvents was done under vacuum and mild heating using a rotary evaporator. Water was
127 the only solvent where temperatures of at least 60 °C were needed, because the process
128 would have taken too long otherwise. Transfer of samples into the round bottom flasks
129 for solvent evaporation, as well as transfer from there into smaller vials was always
130 done under a stream of nitrogen gas, to minimize the contact with oxygen. In order to
131 remove residual solvent from the extracts, the latter were dried in a heating block under
132 a custom built nitrogen shower and/or freeze-dried.

¹This would be especially relevant for sapwood because it may avoid pit aspiration.

unfortunately
the pictures
of the nitro-
gen shower
were lost with
a phone

133 The wood samples extracted with hydrophobic solvents only were dried in the fume
134 hood over night. All other samples had water as the last solvent and were thus dried
135 by lyophilization.

136 2.3 Sample geometry

137 The sample geometry has a large influence on the extraction efficiency, since the proba-
138 bility of washing extractives out of the samples decreases with increasing size. Usually,
139 samples are milled to obtain good yields, which was also done here to obtain extraction
140 yields as a reference to the sticks' yields (largely discussed in Paper I, Appendix A).

141 The sample sticks were meant to be subjected to fungal degradation with a direc-
142 tional growth setup. Although fibers were oriented perpendicular to the growth direction
143 in the directional growth setup originally proposed by Schilling et al. (2013) (mimicking
144 a radial fungal attack and increasing chances of an even hyphal front), this could not
145 be implemented in our experiments due to the need for sectioning transverse sections
146 along the stick for microscopy later on.

147 The geometry should therefore be chosen such that the sample was:

- 148 • long enough to allow for the directional growth setup with the fungus
- 149 • small enough to ensure a reasonable extraction efficiency
- 150 • big enough, and fibers oriented parallel to the long axis of the stick to allow
151 sectioning for microscopy

152 We opted for 50 mm in longitudinal direction, which appeared sufficiently long to
153 allow sampling at several locations in the degradation gradient as seen in (Schilling
154 et al., 2013). Fibers were oriented parallel to the long axis, which would correspond to
155 a fungus growing upwards in the stem. The surface of $3 \times 4 \text{ mm}^2$ (radial \times tangential or
156 *vice versa*) allowed for easy handling during sectioning, as well as enough material to find
157 a good location for imaging and sampling from several year rings. Extraction efficiency
158 may have been better with tracheids oriented perpendicular to the long axis, because
159 many of them would have been open to the outside. Additionally, the inter-cellular
160 lignin network that was found to contain the highest concentrations of extractives, after
161 rays and resin channels, may have had a more exposed surface. Finally, I assume that
162 rays and resin channels, containing the highest amounts of oleoresin, would have had
163 similar chances of being washed out.

164 2.4 Extraction method and apparatus

165 Theoretically, for extractives to leave a solid block of wood, two things have to happen:

- 166 1. they have to dissociate from their matrix - success is determined by solvent and
167 temperature

- 168 2. leave the cell wall and the maze of wood cells by diffusion and/or convection -
169 presumably time and pressure are easily accessible variables for this step.

170 **Selection of solvents**

171 Since over 20 years wood extractives are frequently studied by extracting hydrophobic
172 and hydrophilic components separately. Most literature sources recommend a sequential
173 extraction with at least two solvents of very different polarity and used in the order of
174 increasing polarity (Fang et al., 2013; Willför et al., 2003a; Zule et al., 2015, 2017, a.o.).
175 The extraction of spruce and larch heartwood at different stem heights (Chapter ??)
176 was done using a slight modification of a much-used protocol (Nisula, 2018; Nuopponen
177 et al., 2004; Willför et al., 2003a, 2004, 2006, 2003c) which facilitated the direct com-
178 parison to published results. In the above mentioned studies, hexane was used for the
179 removal of hydrophobic components, and acetone:water (9:1) was used for the extrac-
180 tion of hydrophilic extractives. The extraction protocol was improved in terms of green
181 chemistry values, reducing environmental and/or health concerns where possible (apart
182 from energy demands and safety). In this case, hexane was substituted by heptane (H)
183 due to its lower volatility and because it is not neurotoxic to humans (Byrne et al.,
184 2016).

185 During their inquiry to substitute benzene for a less toxic solvent for the extraction
186 of two hardwood species, Sefara and Birkett (2006) tested various solvent systems and
187 noted that neither were the maximal yields obtained by the same solvent system in
188 either species, nor did different solvent systems deliver the same extract composition.
189 On the one hand, this is not surprising, since the yields depend on varying amounts of
190 different types of extractives, as well as their individual properties. On the other hand,
191 it does pose an analytical problem, because different solvent systems would have to be
192 used to obtain accurate results. Examples of this were seen for both, spruce lignans and
193 larch flavonoids, and are shortly discussed in Chapter ??.

194 As a consequence, the extraction solvents of wood specimen for fungal degradation
195 had to be carefully selected to ensure maximal completeness, but also selectivity. In-
196 spiration was taken from the work of Giwa (1973), who aimed to separate different
197 types of flavonoids of *Larix occidentalis* already while extracting the wood. The author
198 used a five-solvent system with increasing polarity. In the work presented herein, four
199 solvents were chosen for the preparation of wood specimens for fungal degradation, two
200 immiscible with water, one miscible with water and water itself. To replace petroleum
201 ether (Giwa, 1973) or hexane (Willför et al., 2003b), heptane was chosen again as the
202 most apolar solvent. Benzene (as used by Giwa (1973)) was replaced by dichloromethane
203 (DCM) based on the work of Sefara and Birkett (2006) and is a medium-good choice ac-
204 cording to green chemistry standards (Byrne et al., 2016). Ethanol (EtOH) was selected
205 instead of the benzene:ethanol mix used by (Giwa, 1973) to get a clear cut between the
206 hydrophobic and hydrophilic solvents. The selection of water as the most polar solvent
207 was primarily important for extraction of larch heartwood, due to the large amounts
208 of water-soluble arabinogalactan. It was also a good way to investigate whether any

other extractives (or other cell wall residues) in spruce or larch were soluble in water. Samples extracted with all four solvents were called TOT in Paper I and EXT in Paper IV. To check for completeness, the milled samples were extracted with acetone:water (9:1, AcO) as a final step, and indeed, more extractives were removed by this step (although at very low amounts, see Paper I in Appendix A). In an attempt to separate the effects of hydrophobic and hydrophilic extractives, partial extractions were done using only hydrophobic solvents (PHO samples) or hydrophilic solvents (PHI samples).

Selection of apparatus

Traditionally, the extraction of solids (including wood materials, e.g., by Adamopoulos and Voulgaridis (2012); Cheng et al. (2013); Giwa (1973)) was done with Soxhlet extraction, a form of continuous extraction, where the material was extracted in cycles, recycling the solvent by refluxing it within the glass apparatus. Despite its long tradition, the main disadvantages of Soxhlet extraction are long hours of extraction (typically for 24 h or longer (Adamopoulos and Voulgaridis, 2012; Cheng et al., 2013; Côté et al., 1966; Giwa, 1973; Holmbom, 1999, a.o.)) and the large amounts of solvent needed (about 250 ml for 10 g of wood in a large setup (Holmbom, 1999)). Apart from the amount of time needed to extract a larger amount of samples, the long duration of extraction certainly would affect the cell wall structure of extracted wood samples.

These limitations were overcome by Accelerated Solvent Extraction (ASE), which since has been used in numerous studies on wood extractives (Fang et al., 2013; Nisula, 2018; Wagner et al., 2020; Willför et al., 2003b; Zule et al., 2015, 2017, a.o.). It additionally yielded equal or higher amounts of extractives in comparative studies (Willför et al., 2006). It is a modern method for the continuous extraction of solids for which an extraction machine is used, in this case a Dionex 350 ASE from Thermo Fisher Scientific. With this method extractions can be done at high pressures allowing solvents to be used above their boiling points, which increases extraction efficiency. This is achieved by locking the sample in a pressure tight cell with tiny valves at each end, where the solvent is introduced and ejected. The extremely high pressure of 13.8 MPa is generated by the addition of nitrogen gas until reaching a target pressure. The sample cell is then heated in an integrated oven. The use of N₂ gas also has the advantage that oxidation of sample or extract is avoided due to the nitrogen atmosphere in the cell and the collection flasks. Due to the high temperatures used, much lower amounts of solvent are needed (e.g., 54 ml for 12 g of wood meal) and reduce the running costs.

The ASE has an auto-sampler and can be programmed, so that every cell can be treated with up to three solvents sequentially at the desired temperature, number of cycles and total solvent volume (i.e., rinse volume, corresponding to initial solvent volume plus volume added on each extraction cycle). The collection vial can also be selected, so that the user may chose whether to pool extracts or not.

Selection of extraction conditions After the choice of the solvents and the apparatus, the extraction conditions were fine tuned by the choice of temperature, number of extraction cycles and the rinse volume.

As mentioned at the beginning of the section, dissociation of extractives from the cell wall are driven not only by the solvent, but also the temperature. Since temperature may affect the cell wall polymers' motility, that is, they may become mobile when reaching the glass transition temperature, they may in turn affect the cell wall organization. To avoid this, first extractions were performed using only 65 °C. Unfortunately, the yields for the sticks were extremely low compared to the milled wood samples (data not shown), even with other parameters optimized (detailed below). According to Fengel and Wegener (1989, ch. 12) the glass transition temperature for wood is higher than it is for the individual polymers, and no changes in composition were recorded below 100 °C. Thus, the solvent temperatures were used as suggested in the literature (Willför et al., 2003a, a.o.): 90 °C for heptane and DCM and 100 °C for ethanol and water.

find a source
for Tg
definitino //
and Tg of 70
for hemicellulose??

The other factors largely affecting yields concerns diffusion/convection, which are driven by time and pressure. The rinse volume percentage corresponds to the amount of fresh solvent introduced over the course of the extraction cycles and is expressed as a percentage of initial solvent volume, for instance 34 ml in case of cells used in this work. Preliminary extractions showed that an increase of rinse volume from 10% (default) to 150 % (maximum) largely improved the sticks' yields for all solvents already at 65 °C, but also the milled wood yields to some degree (data not shown). This corresponds to about 50 ml of fresh solvent introduced over the course of the extraction and this solvent flow through the cell allowed dissociated analytes to be "pressed out" of the sample. Nevertheless larger volumes of solvent were needed this way.

in paperI
(gc-ms)
and pa-
perIV(oxalate)
the wrong
water
temperature
was written!!
- does this
need to be
corrected in
the journal?

Lastly, the number of extraction cycles was adapted. Initial experiments used only 4 extraction cycles, but with a 15 min duration, allowing extractives to diffuse on a larger time scale. At 150 % rinse volume, this results in a solvent flow of about 12.5 ml per cycle. Theoretically, the initial solvent volume was likely not saturated, but in the sticks - inside the lumina - there may have been local saturation. Because the dissociation phase from the matrix is assumed to be on a much shorter time scale, and passive diffusion on a longer one, a 15 min extraction cycle was deemed to long. Therefore, replacing longer diffusion times with more frequent solvent flow was deemed a more reasonable approach, that is, using 9 extraction cycles and a dwell time of 5 min each, translating into approx. 5.5 ml of fresh solvent per extraction cycle. The improvement of yields is exemplified based on the extraction yields of larch heartwood in table 1.2. The effect was especially significant for the water extract, which is rich in arabinogalactan occupying the cell lumina.²

Despite the efforts of developing a procedure that would efficiently extract the wooden sticks, while avoiding cell wall modifications, the findings presented in Paper I (Appendix A) show that only 20-80 % of the extractives could be removed compared to the milled sticks. It cannot be completely ruled out that the extraction procedure may have affected the micro-distribution of extractives within the tissue due to migration of

²It could or should be tested, whether removing the arabinogalactan first could improve the extraction yields of the other solvents.

solvent / T [°C]	yields mg/g dry wood			
	milled		sticks	
	4 x 15 min	9 x 5 min	4 x 15 min	9 x 5 min
heptane / 90	3,8	8,4	1,7	5,8
DCM / 90	2,2	2,1	1,2	1,6
EtOH / 100	41,2	43,4	7,2	11,7
water / 100	60,3	72,3	7,2	30,8

Table 1.2: Effect of increasing the number of extraction cycles and decreasing the dwell time for larch heartwood extracted with hydrophobic and hydrophilic solvents sequentially.

extractives, as was demonstrated by Nzokou and Kamdem (2004). Nevertheless, the results indicate that, if present, these effects may have been marginal, since the extractives - and especially phenolics - distributions largely agreed with findings from other studies, including *Pinus sylvestris* (Felhofer et al., 2018) and *Pseudotsuga menziesii* (Donaldson et al., 2019).

3 Moisture determination

In Paper I (Appendix A), the sorption isotherm was determined for Kurile larch samples in the native state and after the various extraction procedures to investigate potential effects of extractives on the moisture content in larch. This is relevant for wood in service not only because the moisture content of wood strongly influences its mechanical behavior and dimensional stability, but also because the moisture content of the wood drives the ability of wood degrading fungi and their spores to germinate and grow (Boddy, 1983; Brischke et al., 2017; Zeller, 1920; ?).

The mechanical behavior is mainly affected by the humidity range up to $\approx 98\%$ relative humidity (RH) where wood is hygroscopic and water inside the cell wall is bound by the cell wall polymers. While the cell walls gradually saturate with increasing relative humidity, liquid water may already accumulate in micro-voids such as pit chambers and tapered ends of tracheids. That is, capillary forces begin to dominate at 98 % RH, crossing over into the over-hygroscopic quality domain, ultimately filling up cell lumina (and remaining unsaturated cell wall) with water until full saturation is reached (?).³

When the same ambient temperature and moisture levels are used, but starting from a fully saturated sample, higher values are obtained for the moisture content of the wood. That is, wood shows sorption hysteresis, which can be visualized by the sorption isotherm - the equilibrium moisture content of the sample plotted against various relative

³Scientists would wish for a sharp transition between these two regimes and mark it by the fiber saturation point, but reality is not black and white, and so its definition and location are still under debate (?)

313 humidities for the desorption and absorption modes. This difference between absorption
314 and desorption in the over-hygroscopic range is caused by the "ink-bottle" effect, where
315 the RH at which cell lumina are emptied is dependent on the size of the pits.

316 The size of voids can be calculated, as capillary condensation will happen at different
317 RHs (expressed as water potential for the over-hygroscopic range). Therefore, knowledge
318 about the water sorption behavior in the over-hygroscopic range can be used to estimate
319 the pore size distribution (Defo et al., 1999; Fredriksson, 2019). This in turn can be used
320 to determine bulking effects of wood modification or extractives (Choong and Achmadi,
321 1991; Nzokou and Kamdem, 2004; Zeller, 1920), which was also done in this study. Since
322 the desorption isotherm shows the water content starting with filled cells, any bulking
323 effects should be visible from there, since bulking would reduce the overall volume that
324 water can occupy. Similarly, if pits are clogged by extractives (in addition to aspiration),
325 this should be visible in sorption isotherm.

326 To obtain the equilibrium moisture content of wood (and other porous materials)
327 separate methods are applied for the hygroscopic and over-hygroscopic range. For the
328 hygroscopic range, various saturated salt solutions are commonly used to equilibrate
329 the relative humidity to individual values when placed in an air-tight chamber (Defo
330 et al., 1999; Fitzpatrick et al., 2013; Nzokou and Kamdem, 2004), but this works well
331 only up to 95-97 % (Fredriksson, 2019). Samples are placed on a support platform
332 over the salt solution and allowed to equilibrate for a period of two month. A detailed
333 description can be found in the methods section of Paper I (Appendix A). Beyond that,
334 specialized techniques are needed for the over-hygroscopic moisture range and include
335 pressure membrane, tension plate, centrifuge and the pressure plate technique (Defo
336 et al., 1999; Fredriksson, 2019). The pressure plate technique was used as detailed in
337 ?, who adapted it for equilibration of samples in desorption and absorption modes in
338 a parallel setup. In short, for each humidity level, a pair of pressure-tight chambers
339 is set-up with the same pressure regulator, applying pressures of 0-5 bar (99.6-100 %
340 RH). Each cell contains a porous ceramic plate which is allowed to equilibrate with
341 water supplied from below at the applied pressure. The supply of water is especially
342 important for absorption experiments, for which a water levelling system was installed
343 (see ?). Again, samples were allowed to equilibrate for a period of 2 months.

344 For further details concerning sample preparation please refer to the detailed method
345 section of Paper I (Appendix A).

346 4 Gas Chromatography

347 4.1 Theory and Apparatus

348 Gas chromatography (GC) coupled to a Flame Ionization Detector (FID) and a Mass
349 Spectrometer (MS) was used to characterize the constituents of the heartwood extracts
350 of N. spruce and Kurile Larch. The chromatographic column allows separation of chem-

ical compounds with subsection detection for identification (MS) and quantification (FID) (Bartle and Myers, 2002) (Mermet et al., 2004, Ch. 21). The method also has a long tradition in the study of extractives, especially of conifers (Ekman, 1976; Nisula, 2018; Willför et al., 2003b), but was also applied to extracts of other species (Ekeberg et al., 2006; Nisula, 2018; Ostroukhova et al., 2012). The working principle, as well as the description of the apparatus are shortly outlined below and focus on the setup used for the present study, as detailed in the methods of Paper I (Appendix A).

As for all chromatographic techniques, GC relies on the partition coefficient of molecules between two immiscible media: a stationary phase, i.e., a capillary column, and a mobile phase, in this case an inert carrier gas such as helium. Most commonly, the stationary phase for GC setups consists of a highly viscous liquid. Thus, in the strict sense the method should be called gas-liquid chromatography. Put simply, the analyte molecules are transferred to the gas phase and interact with the stationary phase to varying degrees and through different inter-molecular forces. This causes a spatial separation of different analytes within the column, that is, each type of analyte will spend a different amount of time in the column.

The aforementioned partition coefficient describes the relative concentration of an analyte in the liquid stationary phase and the gas phase, which in turn is governed by their activity coefficients, i.e., the degree of interaction with either phase. The specific retention time of any analyte under given experimental conditions can then be determined by the fraction of analyte in the mobile phase, the velocity of the mobile phase and the length of the column. Since molecules are in the gas phase in GC, the main influencing factors of the retention time (or rather, retention volume) of an analyte are its individual vapor pressure and the column temperature, as well as the choice of the stationary phase (Mermet et al., 2004).

The sample is usually dissolved in a solvent and introduced to the column in the gas phase by means of a split/splitless injector. A small amount of liquid sample (e.g., 1 μ l) is transferred from the sample vial into the hot injector liner by an auto-sampler. (Mermet et al., 2004) The injector liner is a small chamber where the sample is vaporized and mixed with the carrier gas. During sample preparation, analytes with very low vapor pressure can be derivatized to increase their volatility (Li et al., 2001). For extractives, many of which are phenolic in nature, derivatization is done by the attachment of trimethyl-silyl groups to alcoholic residues (Ekman and Holmbom, 1989; Li et al., 2001; Nisula, 2018; Örså and Holmbom, 1994). When the injector is operated in split mode, the split valve is open during injection directing the majority of the sample (e.g., 90 %) to waste. (Mermet et al., 2004) The advantage of this mode is that highly concentrated samples can be diluted directly by the carrier gas and thereby column overload avoided (Chasteen, 2023). The requirements to the carrier gas itself are that it should not interact with the sample or the stationary phase and it should not produce a signal in the detector. Furthermore, the gas should be relatively small in size to avoid a large number of collisions with the analytes, which would cause peak broadening.

The column is located in an oven, which is operated via a temperature program that is set to yield sufficient separation of analyte peaks, but also to elute all analytes by the end of the run. The required temperatures, heating rates and holding times need to be set. As mentioned above, the properties of the stationary phase largely influence the retention times, i.e., they determine which kind of compounds are efficiently separated. For most applications, the inner tube lining, very often fused silica, is coated with a non-polar, viscous liquid such as dimethylpolysiloxane as in the case of the HP-1 column (Agilent J&W) used in this work. With such a setup, polar compounds will elute from the column first. The length of the capillary is important for band resolution. The longer the column, the higher is the resolution, but also the measurement time (Mermet et al., 2004). Additionally, some compounds may not elute at all if the column is very long. A good example for this can be seen in Zule et al. (2015), where hydrophobic extractives of *Larix decidua Mill.* were run on a 6 m and a 25 m HP-1 column. On the shorter column all four internal standards eluted, as well as sterol esters and triglycerides from the sample. On the long column, only the first two standards eluted together with well resolved fatty acids, resin acids, diterpenoids and sterols.

As mentioned above, detection of the extractive molecules was done over FID for quantification and MS for identification. The FID is one of the most universal and wide spread detectors for chromatography, as it has large linear range and high sensitivity. It is therefore very suited for quantification, provided the compounds are known. The signal is based on the change in conductivity of a hydrogen flame placed in an electric field when organic compounds eluting from the column are burned (Mermet et al., 2004). When unknown samples are run, coupling GC to mass spectrometry (MS) is the method of choice for organic compounds and greatly complements the FID measurements. Not only can the elemental composition of an analyte be determined, but unresolved peaks can be distinguished via their MS signal, which is demonstrated in chapter ??.

The MS operates under vacuum conditions, which is easily coupled to a GC setup with low flow-rates as used in this work (e.g., 0.9 ml/min) (Mermet et al., 2004, Ch. 25). The analytes leaving the column are first ionized, then pass through a mass analyzer and are finally detected with an electron multiplier. In the setup used, analytes were ionized using electron impact ionization, where accelerated electrons may knock an electron out of an analyte molecule upon collision, resulting in a cation with a single positive charge. The excess of internal energy leads to further fragmentation of the molecule. At an acceleration voltage of 70 eV, reproducible fragmentation patterns of the analytes are obtained, and accordingly, large databases were built to facilitate identification of compounds. Such a library was collected for extractives by our collaborators from the laboratory of Wood and Paper Chemistry at Åbo University in Finland. Each ion thereby produced has a specific mass-to-charge ratio (m/z) and travels through the mass analyzer, in the present case a quadrupole. The quadrupole consists of four stainless steel rods arranged radially and parallel to each other. A positive and a negative direct-current (DC) voltage is applied to the diagonal pairs, respectively. This field is then

433 disturbed by an alternating-current (AC) at radiofrequency. Only ions of a given m/z
 434 ratio will have stable, oscillating trajectories along the rods and will reach the detector.
 435 By altering the DC and AC voltages, keeping their ratios equal, a variety of m/z ratios
 436 can be scanned (e.g.,m/z 1-500).

437 4.2 Quantification

438 The quantification of each sample was done using a fatty acid (heneicosanoic acid, HIA)
 439 and a sterol (betulinol, BET) as internal standards (ISTDs). As described in Paper I
 440 (Appendix A), peak tables of the FID of each sample were obtained from MSD Enhanced
 441 ChemStation[®] (Agilent Technologies Inc., CA 95051, USA, www.agilent.com). These
 442 were imported into MATLAB (MATLAB[®] R2017a, The MathWorks, Inc.) and peak
 443 areas used for quantification of each identified extractive with a self-written MATLAB
 444 script based on the calculation shown below (1.1-1.7). The script was adapted to the
 445 needs that emerged for two and four extraction solvents, respectively. As stated in
 446 Paper I (Appendix A), fatty acids, resin acids and flavonoids were quantified using
 447 HIA, whereas sterols and lignans were quantified using BET.

448 The calculation of original analyte concentration per gram of wood was done as
 449 follows:

$$c_{ISTD,1} = m_{ISTD}/V_1 \quad (1.1)$$

450 with $c_{ISTD,1}$ [mg/ml] the concentration of internal standards (ISTD) in the stock
 451 solution, m_{ISTD} [mg] the mass of the respective ISTDs, and V_1 [mg/ml] the volume
 452 of stock solution. Note that this is calculated for each internal standard separately ,
 453 i.e.,HIA and BET.

$$c_{ISTD,2} = c_{ISTD,1} * V_2/V_3 \quad (1.2)$$

454 with $c_{ISTD,2}$ [mg/ml] the concentrations of internal standards in the sample, V_2 [ml]
 455 the volume of internal standard used for the mixture, V_3 [ml] the final volume of GC-MS
 456 samples after derivatization.

$$c_{a_i,1} = (A_{a_i}/A_{HIA}) * c_{HIA,2} \quad (1.3)$$

457 with $c_{a_i,1}$ the i th analyte (a_i) concentration in the GC-vial of all chemical groups
 458 except sterols and lignans (see below), A_{a_i} and A_{HIA} the peak areas (A) of analyte i
 459 and HIA peak, respectively.

$$c_{a_{i,ST},1} = (A_{a_{i,ST}}/A_{BET}) * c_{BET,2} \quad (1.4)$$

460 where $a_{i,ST}$ designates analytes that are sterols, their concentration being calculated
 461 relative to the BET standard.

$$c_{a_i,LI,1} = (1.2 * A_{a_i,LI}/A_{BET}) * c_{BET,2} \quad (1.5)$$

462 where $a_{i,LI}$ are analytes that belong to the chemical class of lignans. Their concentration
 463 is calculated relative to the BET standard, with an additional correction factor
 464 of 1.2, as recommended in Willför et al. (2003b).

$$c_{a_i,extract} = (c_{a_i} * V_3)/V_{extract} \quad (1.6)$$

465 with $c_{a_i, extract}$ [mg/ml] is the analyte (incl. ST and LI) concentration in the extract
 466 and $V_{extract}$ [ml] is the volume aliquot of 0.5 mg calculated for each sample.

$$c_{a_i,wood} = (c_{a_i,extract} * V_4)/m_{wood} \quad (1.7)$$

467 with $c_{a_i,wood}$ [mg/g] the analyte concentration in the original wood sample, V_4 [ml]
 468 the total volume of extract and m_{wood} [g] the dry mass of each wood sample used for
 469 extraction

470 5 Vibrational Spectroscopy

471 5.1 Theory

472 Vibrational Spectroscopy is the umbrella term for analytical techniques that apply elec-
 473 tromagnetic radiation to induce vibrations of chemical bonds (i.e., electrons), allowing
 474 the identification and/or quantification of molecules. Because vibrations are studied,
 475 these techniques are usually non-invasive and suitable for a large variety of materials
 476 and states of aggregate. The main method used in this project was Confocal Raman
 477 Spectroscopy. It was used to obtain reference spectra of heartwood extracts and purified
 478 lignans, as well as to image heartwood cells in order to locate extractives in spruce and
 479 larch (Paper II, Appendix B) and to study brown rot degradation in spruce (Papers III
 480 and IV, Appendices C, D). Infra Red (IR) spectroscopy was used in two different, very
 481 specialized setups to study the heartwood formation process in Kurile larch (Paper V,
 482 Appendix E): Imaging was done in combination with a powerful source in the form of
 483 a synchrotron-derived IR beam (synchrotron-FT-IR) and spectra of individual cell wall
 484 layers were obtained from a combination with an atomic force microscope (AFM). Note
 485 that I was only involved in the synchrotron-FT-IR measurements. In the area of wood
 486 science, both IR and Raman spectroscopy have been extensively used to characterize
 487 the composition and organization of the cell wall (Agarwal, 1999, 2019; Gierlinger et al.,
 488 2002; Naumann et al., 2007).

489 Vibrations of chemical are of characteristic energy, depending on the type of move-
 490 ment, the atoms involved, as well as their environment. Out of all the possible move-
 491 ments (translation, vibration, rotation), for a given molecule with N atoms $3N-6$ vi-
 492 brational modes can be expected, except for linear molecules, where $3N-5$ modes exist.

These involve symmetric and asymmetric stretching and bending vibrations, where the nuclei of atoms change their positions relative to each other, which influences the electron distribution of the bonds. More energy is required to increase/decrease the distance between two bonded atoms (stretches), than for a change in bonding angle between neighboring atoms. Additionally the strength of the bond (i.e., the electronegativities of the atoms) and the weight of the atoms influence the amount of energy necessary for a given vibration, that is, the frequency at which a bond can vibrate. This relationship is described by Hooke's law (Eq. 1.8) for a diatomic molecule) and shows that strong bonds and light atoms increase the vibrational frequency (Larkin, 2011; Smith and Dent, 2005). For a better visualization of the spectra, wavenumbers are usually plotted instead of the frequency (or wavelength).

$$\bar{v} = \frac{1}{2\pi c} \sqrt{K \left(\frac{1}{m_1} + \frac{1}{m_2} \right)} \quad (1.8)$$

, where \bar{v} is the wavenumber in cm^{-1} , c the speed of light in cm/s , K the force constant between the atoms and m_1 and m_2 their respective masses (Larkin, 2011).

Depending on the excitation wavelength used, there are two different processes that lead to the transfer the photon energy to a molecular bond causing it to vibrate: inelastic scattering (Raman spectroscopy ⁴) or absorption (infra-red (IR) spectroscopy) (Larkin, 2011; Smith and Dent, 2005).

At higher energy in the ultra violet to near-IR (NIR) range and using a monochromatic light source, photon scattering can be observed. Thereby a photon is absorbed and the molecule transferred into a short lived "virtual" excited state where the electron cloud is distorted by the electromagnetic field of the light source (induced dipole moment). Upon relaxation, a new photon is emitted. In rare cases, that is, in $10^{-6}\text{-}10^{-8}$ of the original photons, the photon is inelastically scattered. A characteristic amount of energy is transferred to or from the photon by bond vibrations in the molecule, if the vibration causes a change in the bond polarizability (i.e., the ease of distorting the electron cloud). This is the basic selection rule for Raman activity. The detected photons have a slight shift in frequency relative to the original one, and therefore, the wavenumbers shown in the spectrum are not absolute, but relative to the original photon frequency. Whether the photon gains or loses energy, depends on the initial vibrational energy level of the molecules, which in turn is temperature dependent (Smith and Dent, 2005). Since experiments are usually conducted at room temperature, standard instrumentation for Raman spectroscopy records the portion of photons losing energy to the vibration (Stokes shift). In that case, a photon is excited from the ground vibrational state to the virtual state and relaxes to the first vibrational state of the molecule

⁴The foundation of Raman spectroscopy has its 100 years anniversary in the year these lines are written. The theory of the existence of inelastic light scattering was published in 1923 by the Austrian physicist Adolf Smekal. Only five years later it was first shown experimentally by led to the name Raman spectroscopy after one of its discoverers (Smith and Dent, 2005).

(schematically shown in Figure 1.2). The intensity of a Raman band (I_R is strongly dependent on the frequency of the incident radiation (ν , Eq.1.9), which implies that Raman bands produced by excitation with longer wavelengths are weaker. Other factors influencing the size of Raman bands is the intensity of the incident radiation (I_0 , the number of scattering molecules (N), and the square of the change in polarizability (α) relative to the vibrational amplitude (Q) (Larkin, 2011).

$$I_R \propto \nu^4 I_0 N \left(\frac{\delta\alpha}{\delta Q} \right)^2 \quad (1.9)$$

If no such energy transfer occurs, elastic Rayleigh scattering is seen. The emitted photon will have the same frequency as the original one and return to the vibrational ground state (Figure 1.2) This process is much more likely to occur (10^{-3} photons of the original beam) and thus appropriate filters have to be installed to detect the much weaker Raman signal (Larkin, 2011; Smith and Dent, 2005).

At lower energies in the NIR to IR energy range and a broad band source, non-destructive absorption processes may only occur if the photon energy corresponds to the energy gap between vibrational levels (resonance effect). That is, if the vibration causes a change in dipole moment of the molecule at the same frequency as the incident photon, the latter can be absorbed. This is the selection rule for IR activity. The intensity of the absorption (I_{IR}) is proportional to the square of the change in dipole moment (μ) relative to the vibrational amplitude (Eq. 1.10). The spectrum obtained shows how much of the radiation was absorbed (or transmitted) at the wavenumbers covered.

$$I_{IR} \propto \left(\frac{\delta\mu}{\delta Q} \right)^2 \quad (1.10)$$

, where $\mu = e_i r_i$, with e_i the atomic charges of a molecule and r_i their positions (Larkin, 2011).

Although vibrations of molecules are involved both in Raman scattering and IR absorption, the information obtained by the two processes is complementary and governed by the above mentioned selection rules (Agarwal and Atalla, 2010; Larkin, 2011; Smith and Dent, 2005). A change in polarizability, necessary for Raman scattering, is caused by symmetric vibrations, also in the case where the molecule does not have an inherent dipole moment (non-polar groups). On the other hand, absorption of IR radiation is caused by changes in dipole moment that can be observed primarily in polar groups and asymmetric vibrations. Some vibrations will be visible exclusively in Raman or IR spectroscopy, others may be seen in both.

A big advantage of Raman over IR spectroscopy is that the scattering of water and CO_2 are relatively weak, while they are strong absorbers of IR radiation (Smith and Dent, 2005). Thus, aqueous solutions and hydrated biological samples can be easily analyzed using Raman spectroscopy, and no corrections for atmospheric water and CO_2

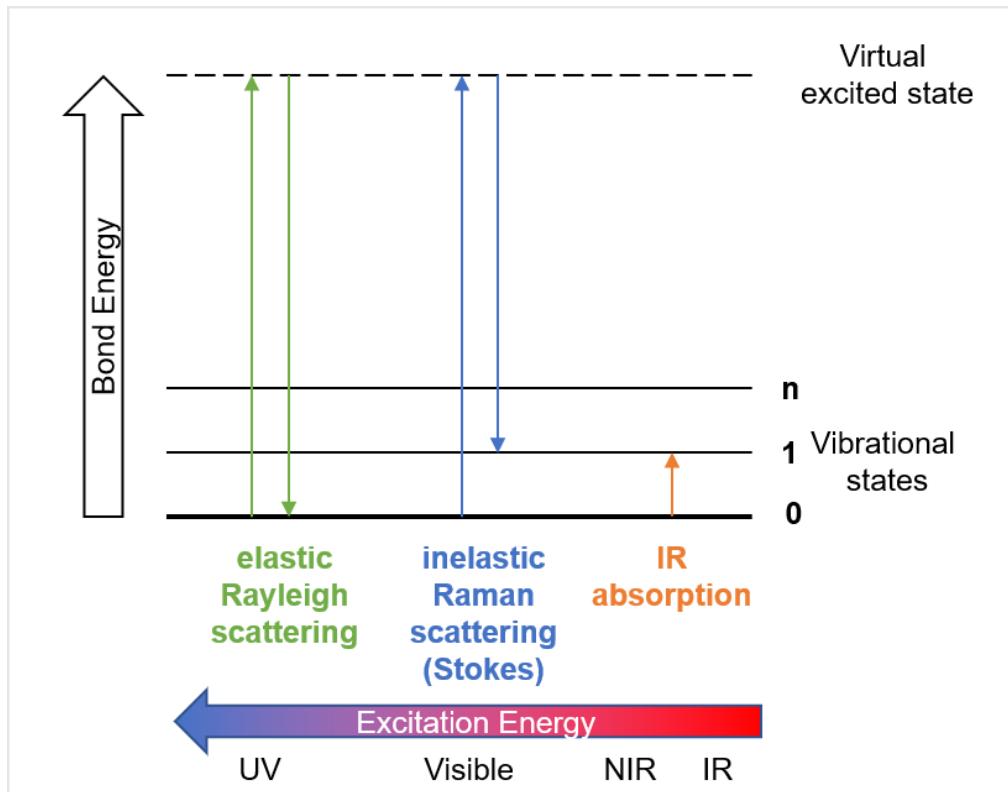


Figure 1.2: Schematic representation of the energy transitions of a molecular bond upon excitation with light of different wavelengths. In the higher energy regime, Rayleigh and Raman scattering can be observed (UV-NIR), while absorption of photons to higher vibrational levels of the molecule occurs in the NIR-IR range. Note that for Raman scattering the energy transitions for the Stokes shift are shown.

562 are necessary, as is the case for IR spectroscopy. On the other hand, the use of visible
563 light as excitation wavelength in Raman spectroscopy may result in absorption processes
564 that cause burning of the sample or electronic transitions leading to fluorescence. Laser-
565 induced fluorescence is a big issue with biological samples, especially when conjugated
566 aromatic molecules such as lignin are present (Agarwal, 1999; Lähdetie et al., 2013;
567 Smith and Dent, 2005). The main problem with fluorescence emission is that it may
568 cause such a large background in the spectrum that the Raman signals are concealed. As
569 both, lignin and phenolic extractives, especially flavonoids in larch heartwood, increased
570 the background drastically, several counter measures were used for acquisition of Raman
571 spectra:

- 572 • use of a confocal microscope setup (covered below).
- 573 • appropriate sample preparation, i.e., quenching the fluorescence by immersing
574 the sample in water (Agarwal, 1999), i.e., trapping the sample section with an
575 objective-slide + cover glass with a drop of water. This measure also dissipates
576 heat created by the laser from the sample.
- 577 • choice of a different excitation wavelength for images of larch heartwood, i.e., using
578 the 785 nm laser (NIR, faintly red), instead of 532 nm (visible green) for Raman
579 images.

580 Furthermore, for some samples it may also be helpful to quench the fluorescence by
581 bleaching the sample, i.e., irradiating it with the laser before acquisition of a spectrum.
582 This also works in the case of wood, but causes alterations of the lignin structures
583 (Prats-Mateu et al., 2018), for which reason bleaching was avoided during this work.

584 Fluorescence is not the only effect that may occur when the excitation wavelength
585 (nearly) resonates with an electronic transition: Because the polarizability of a molecule
586 increases near/at an electronic transition (Bock et al., 2020; Long, 2002), Raman bands
587 can be drastically enhanced (resonance enhancement). This is especially true when
588 the vibrating bonds are part of a π system (i.e., aromatic rings and ethylenic bonds),
589 since these also have an increased polarizability (Bock et al., 2020; Davies, 1952). Ad-
590 ditionally, certain aromatic C=C stretching vibrations at 1600 cm^{-1} are enhanced by
591 conjugation (Bock et al., 2020, and references therein), which is the reason why espe-
592 cially conconiferylaldehyde and -alcohol show strong Raman bands (Bock et al., 2020).
593 These enhancement effects strongly affect Raman spectra of lignin(s), where there is a
594 large amount of aromatic rings and some degree of conjugation. These effects explain
595 why the lignin C=C stretch is still strongly present in spectra of the S2 cell wall, al-
596 though the cellulose content is about twice as high than the lignin content (Holtzapple,
597 2003).

598 // A short overview of applications of Raman and IR spectroscopy, as well as the
599 necessary instrumentation used in this work follows.

"and refer-
ences therein"
- enough or
should they
be mentioned
explicitly?
(ref214-219,
noted in out-
commented
text)

600 5.2 Imaging

601 Since the aim of this study was to investigate the spatial resolution of heartwood ex-
 602 tractives, as well as to observe the initial stage of brown-rot degradation on a cellular
 603 level, obtaining chemical information on a submicrometer scale was essential. This was
 604 achieved with Confocal Raman Microscopy (spruce and larch, Papers B-D) and com-
 605 plemented with synchrotron- IR microscopy (for larch, Paper V, Appendix E).

606 In order to perform imaging, a microscope is used to direct the excitation beam at
 607 the sample. By means of a scanning stage, spectra are recorded at every pre-defined
 608 point. The maximal lateral resolution ($r_{\delta x, \delta y}$, Eq. 1.11) that can be achieved is limited
 609 by diffraction and is dependent on the wavelength (λ) and the objective aperture (NA)
 610 used (Atkins and Paula, 2006; Dieing et al., 2011)⁵. It follows from equation 1.11 that
 611 Raman images will generally have a higher resolution, compared to IR images due to
 612 the shorter wavelength. Similarly, Raman images acquired with a 785 nm laser will have
 613 a lower resolution than with a 532 nm laser. Additionally, the intensity of the signal
 614 is much weaker (1.9) with the 785 nm laser, which requires longer integration times
 615 per spectrum at maximal laser power to obtain reasonable signal-to-noise ratio, which
 616 drastically augments the measurement times for images.

$$r_{\delta x, \delta y} = 0.61\lambda/NA \quad (1.11)$$

617 Confocal Raman Microscopy

618 When using a confocal microscope, clear measurements can also be done on thick sam-
 619 ples, because the setup hinders light that is not in the focal plane to reach the detector.
 620 Thus a focus point below the surface of the sample can be chosen.

621 The key elements are a focused point source of light (i.e., a laser) in combination
 622 with a pinhole in front of the detector. A coherent, collimated laser beam causes a very
 623 small illumination volume, which limits the amount of light subsequently emitted from
 624 the sample. The confocally aligned pinhole physically blocks all radiation that is not
 625 part of the image plane (Conn, 2010; Dieing et al., 2011). Consequently, fluorescence is
 626 also reduced (Agarwal, 1999).

627 The maximal depth resolution ($r_{\delta z}$, Eq. 1.12) is dependent on the refractive index
 628 (n) of the sample, and has a higher dependence on the NA than the lateral resolution
 629 (Conn, 2010; Dieing et al., 2011). The NA determines the maximal angle of light that
 630 can be collected by the objective lens (Conn, 2010).

$$r_{\delta z} = 1.26n\lambda/NA^2 \quad (1.12)$$

⁵Depending on how the Rayleigh criterion is interpreted, a correction factor of 0.32 instead of 0.61 should be used (Conn, 2010).

The Raman instrument used for this thesis was the WITec alpha 300R (WITec GmbH), schematically shown in Figure 1.3 (adapted from (GmbH, 2021)). The system was equipped with a 532 nm and a 785 nm diode lasers (Fig. 1.301) which are linearly polarized, that is, the electromagnetic field of the light waves is uniformly oriented. This allows the determination of polarization in highly oriented samples (or areas thereof), such as the S1 layer of tracheid cells (Gierlinger et al., 2010). Optical fibers guide the laser beam to the microscope (Fig. 1.302), where various objectives can be selected for measurement (Fig. 1.303). Most of the work for this thesis was done using a 100x oil objective (Zeiss, "N-Achroplan", NA = 1.2) due to the method of sample preparation (wet sample enclosed in objective slide + cover glass). The high magnification together with the large aperture allowed for a very high lateral resolution of 270 nm and 400 nm for the green and the red laser, respectively. In practice, 333 nm were used for both. Before reaching the detector, the back-scattered light is filtered to remove a large part of the Rayleigh peak (Fig. 1.305). The photonic fiber leading to the detector simultaneously serves as pinhole (Fig. 1.3 06). For detection, a separate ultra-high throughput spectrometer with back-illuminated CCD cameras is installed for each laser (UHTS300 and UHTS400, respectively, Fig. 1.307, 08).

One of the challenges when measuring wood tissue is the uneven surface, especially if rays are included. In order to minimize intensity variations due to different focal points, the WITec Control Suit 5 software provides a feature to "learn" the surface of a sample prior to imaging. This feature was used for all images published. Using the on-line detection mode (oscilloscope), the focus is adjusted by the user based on the maximal spectral intensity using a predefined number of points (25 in this case). A strategy was developed to avoid damaging the sample by too long exposure times, due to delays while looking for where to place the points (Fig. 1.4a): The points were first marked on the white light image (Fig. 1.4b), then the routine was started and the focus quickly adjusted manually, resulting in images without visible damage (Fig. 1.4c).

658 Synchrotron IR imaging

659 As mentioned earlier, the fluorescence observed in larch samples with the 532 nm laser
660 could only partially be compensated with the 785 nm laser, including long measurement
661 times. Thus recurrence to mid-IR radiation was necessary, but the resolution achievable
662 with a conventional IR microscope is limited compared to Raman imaging due to the
663 longer wavelength, and does not allow high resolution imaging of subcellular details.
664 Therefore, the investigation of the formation of heartwood extractives in Kurile larch
665 (Paper IV, Appendix E) was conducted at a synchrotron facility (ALBA, Cerdanyola del
666 Vallés, Spain, proposal 2018022761) using the IR beamline MIRAS (BL-01) for imaging.

667 In a synchrotron, electrons are accelerated in a large ring (270 m circumference) to
668 extremely high potentials (3 GeV at ALBA) (ALB, 2023), and upon application of strong
669 magnetic fields, electromagnetic radiation can be extracted. Using the appropriate
670 filters, the desired wavelength is extracted for each beamline. The extreme brightness

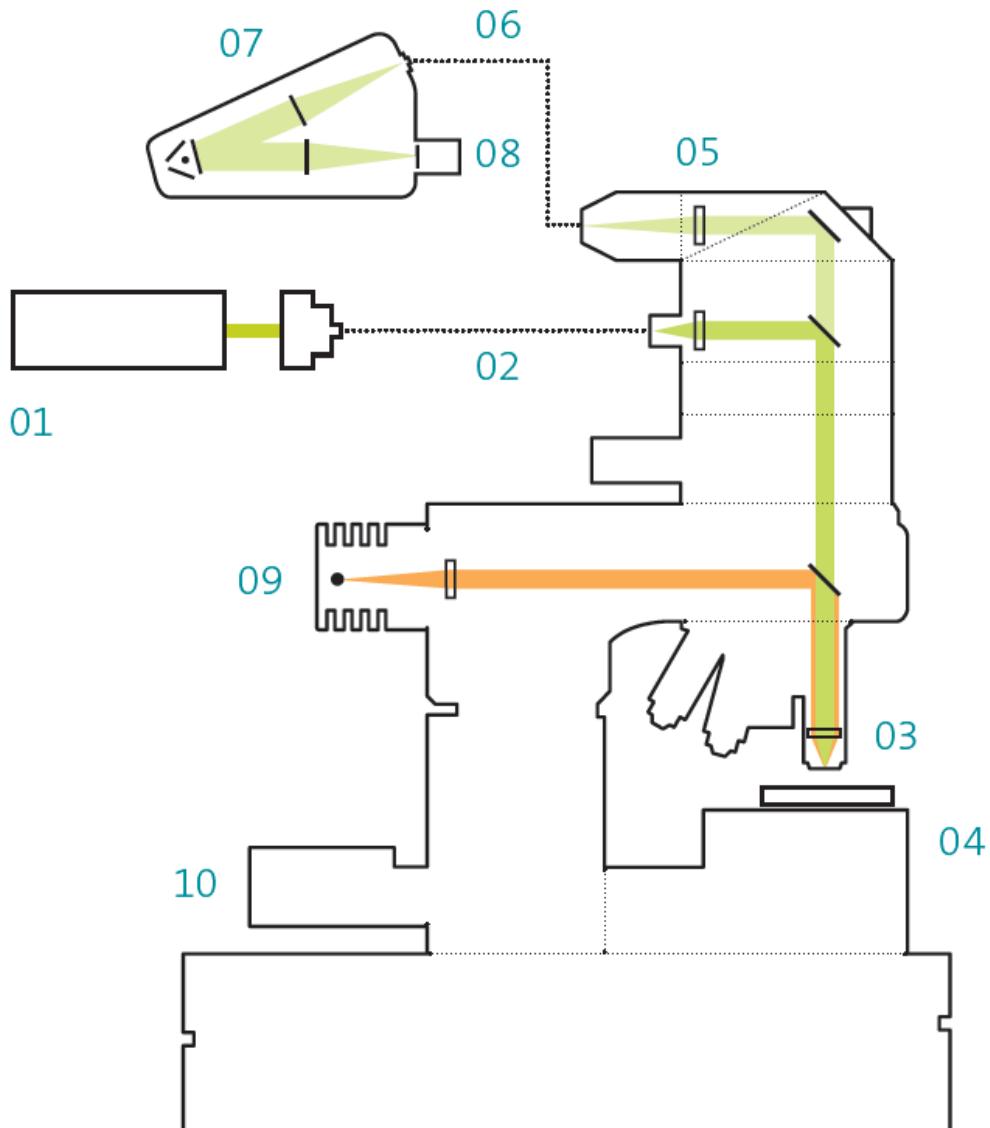


Figure 1.3: Schematic representation of the WITec alpha 300R Raman microscope (WITec GmbH) adapted from (GmbH, 2021). 01 - laser, 02 - optical fiber, 03 - objective, 04 - scanning stage, 05 - filter set, 06 - optical fiber (pinhole), 07 - lens-based UHTS Raman spectrometer, 08 - CCD detector, 09 - Koehler white-light illumination, 10 - motorized z-stage for focusing

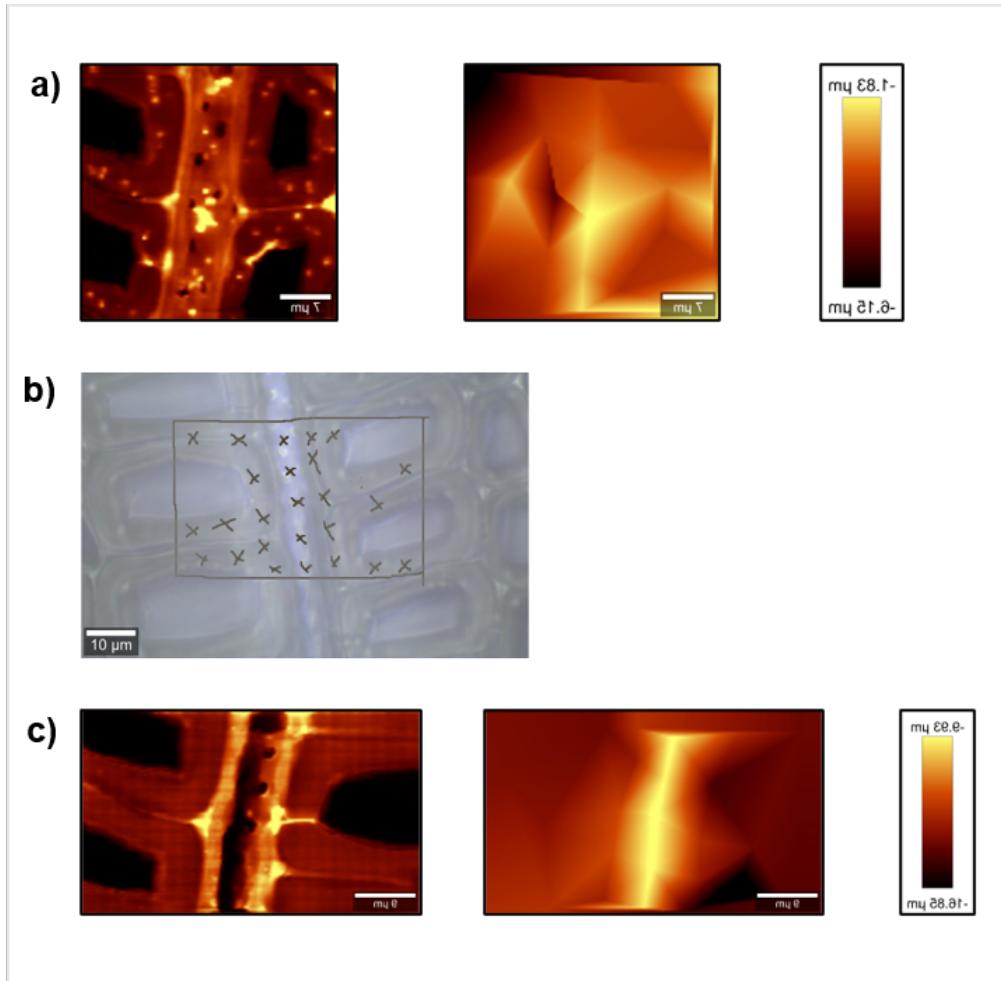


Figure 1.4: Images resulting from manually learning surfaces. **a)** First trials using the routine resulted in visibly burned spots on the sample. **b)** Marking of points to use for focus adjustment in white light image. **c)** Exemplary image resulting from marking the points first.

of the beam allows for a much better signal-to-noise ratio. At ALBA, a Bruker Hyperion 3000 microscope is installed and measurements were performed with a 36x objective lens in transmission mode, with ultra-thin samples placed on a zink selenide crystal. The near-perfect collimation of the beam allows for a superior resolution compared to a bench-top source. That is, with an aperture size of $3 \times 3 \mu\text{m}^2$, a nominal resolution of $1.1 \mu\text{m}$ is achieved at ALBA (ALB, 2023), while the same microscope operated with a regular source could achieve a spatial resolution of $2.5 \mu\text{m}$ (Lasch and Naumann, 2006). Thus, with a cell wall thickness of $1\text{-}5 \mu\text{m}$ for conifer tracheids (Fengel and Wegener, 1989; Schwarze et al., 2000), the chances of hitting distinct cellular areas, such as the middle lamellae, are more than twice as large with a synchrotron source. The spectra were acquired in the spectral region of $4000\text{ - }800 \text{ cm}^{-1}$, with 64 co-added scans using a Bruker Vertex 70 fourier-transform spectrometer together with a HgCdTe (MCT) detector cooled with N_2 .

5.3 Point spectra

In this work, point spectra were used as valuable add-on for the confirmation of results from imaging. Reference Raman spectra of extracts were acquired for the identification of MCR-ALS components of extractives in spruce (Paper II, Appendix B). To confirm data obtained from the synchrotron-IR measurements, a specialized IR technique was used, coupling IR spectroscopy with atomic force microscopy (AFM). A short description and considerations on both applications of point spectra follows.

Raman references

Reference spectra of extracts were taken as point spectra using an air objective (100x air) after solvent evaporation on a glass slide. The extracts remaining on the slide reflected the composition determined by GC-MS. As shown in the white light images (Figure 1.5), the surface structures of the dried extracts were somewhat similar between both species. The heptane extracts formed oil droplets (Fig. 1.5a,e), therefore focusing to get a good spectrum was easiest with the 100x objective. The DCM extracts dried inhomogeneously with areas that appeared crystalline (Fig. 1.5b,f), therefore spectra were recorded at various locations and averaged. Interestingly, the ethanol extracts (Fig. 1.5c,g) formed more homogeneous, but brittle films (seen well in the 10x image in Fig. 1.5f). The water extracts also dried with a rather inhomogeneous surface (Fig. 1.5d,h). The laser intensity had to be reduced in case of the larch extract as it was composed mainly of arabinogalactan, which apparently burns easily, as shown in Fig. 1.5f.

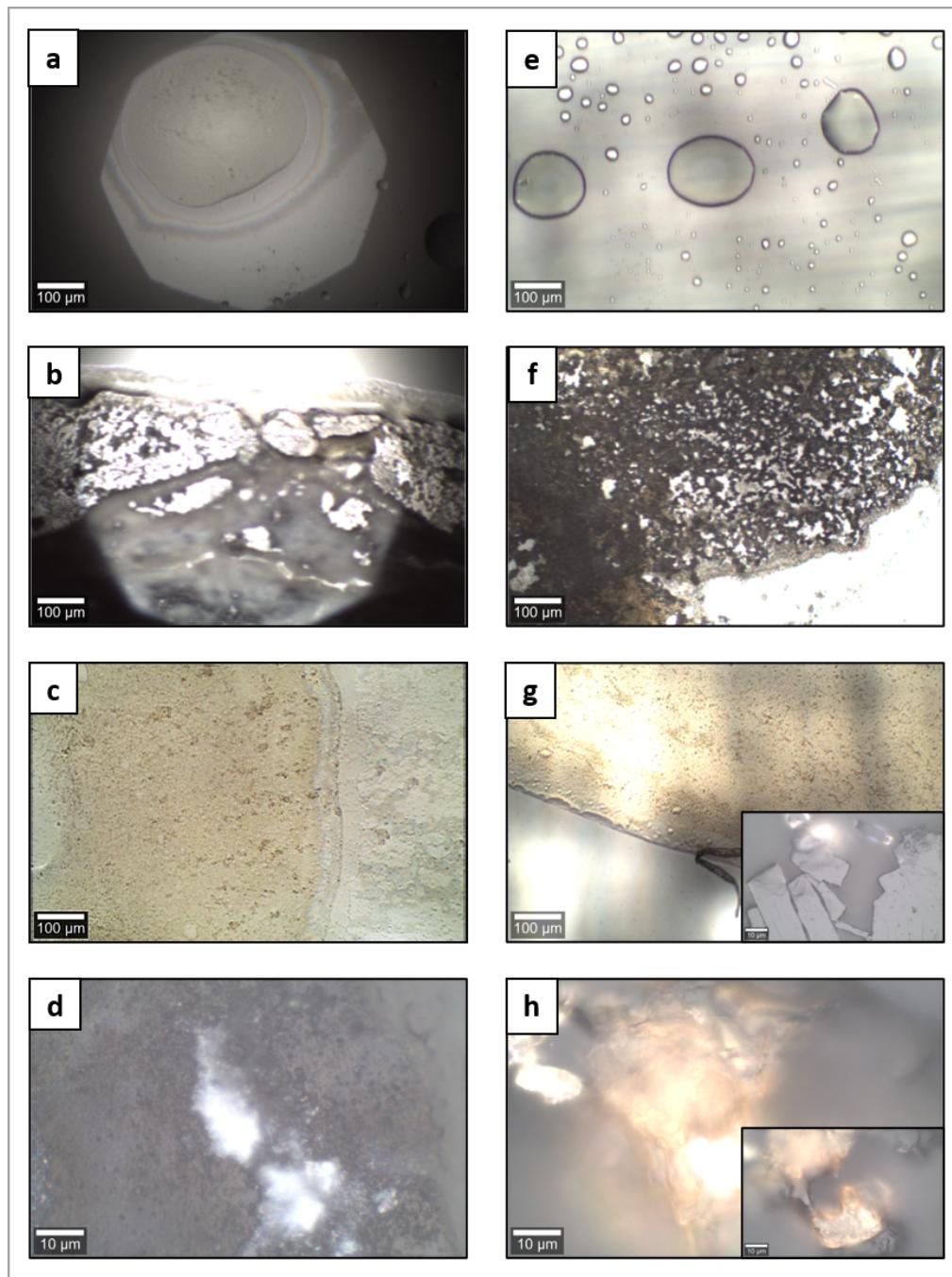


Figure 1.5: White light images of spruce and larch extracts. **a-d** spruce heptane, DCM, EtOH and water extract, resp.; **e-h** larch heptane, DCM, EtOH and water extract, resp.

705 AFM-IR

706 The combination of AFM and IR spectroscopy is a relatively new development (Dazzi
707 et al., 2005) and has already proven valuable for a large number of applications, including
708 microbiology (Dazzi and Prater, 2017). Nevertheless, this technique has only made its
709 debut in wood science (Martin et al., 2021; McDonald et al., 2020; Wang et al., 2016,
710 2018). AFM-IR allows the acquisition of point spectra well beyond the diffraction limit
711 set by the wavelength, because, rather than measuring the IR absorption directly, the
712 signal is obtained via the deflection of an AFM probe after the sample is excited by an
713 IR laser pulse.

714 AFM allows the mapping of the topography of a sample at extremely high resolution
715 by moving a piezo-driven scanning probe over the sample surface. In contact-mode the
716 probe is in direct contact with the surface. The simplest form of an AFM probe consists
717 of an approx. 2 mm long cantilever with a tip at its free-swinging end. While the
718 properties of the cantilever (i.e., its spring constant) determine the type of sample that
719 can be analyzed, the tip's end determines the resolution (e.g., 2-30 nm). The material
720 for the tip determines the type of interactions with the sample, but commonly consists
721 of silicon and derivatives. For AFM-IR a silicon tip coated with gold is recommended
722 (Dazzi et al., 2005). The deflection of the cantilever is tracked by a visible laser directed
723 on the top face of the cantilever bearing a reflective window that will direct the laser
724 onto a four-quadrant detector (Dazzi and Prater, 2017; Dazzi et al., 2012).

725 When an IR laser pulse is directed at the sample, the sample will locally expand at
726 its absorbing wavelength and thereby give an impulse to the cantilever. The amplitude
727 of the thereby induced oscillation of the cantilever is proportional to the sample's IR
728 absorption.⁶ By scanning through various wavelengths at the same spot, an IR ab-
729 sorption spectrum is obtained as sensed through the "nanoscopic" contact area of the
730 tip (Dazzi et al., 2005; Dazzi and Prater, 2017; Dazzi et al., 2012).⁷ In practice and
731 as shown in Paper V (Appendix E), an AFM topography image is acquired and single
732 spectra obtained in the regions of interest.

733 6 Analysis of Spectral Data

734 The following section describes the multivariate data analysis methods used, including
735 how the data are structured and pre-processed. All of the following procedures were
736 performed in MATLAB R2019a (The MathWorks Inc.), some of them with the MATLAB
737 extention PLS toolbox (Eigenvector Research Inc.)

⁶The frequency of the oscillation will depend on the mechanical stiffness of the sample, leading to another variable that can be obtained during the same measurement.

⁷Alternatively, a scan can be performed at a single wavelength, thereby generating an image.

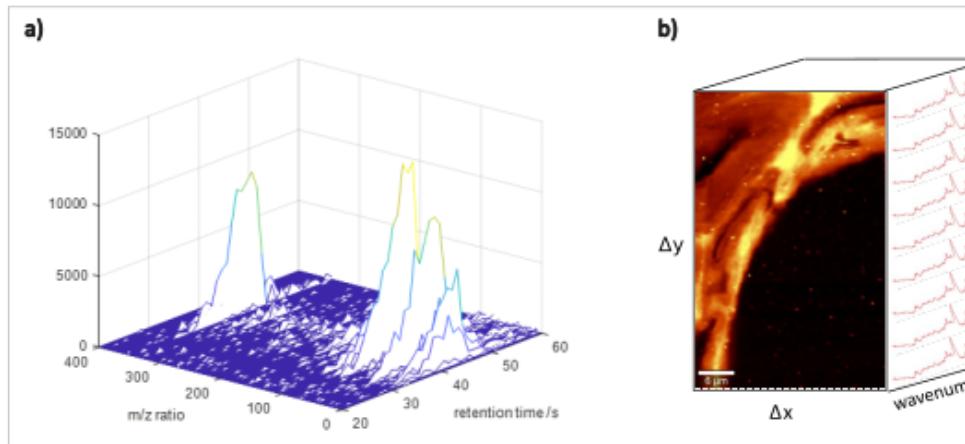


Figure 1.6: Natural data structure of a) MS data, with m/z ratios and retention times vs. intensity and b) Raman imaging data, with location (x and y coordinates) and wavenumbers vs. intensity

738 6.1 Data structure

739 Any spectrum consists of intensity values obtained for multiple variables such as wavenumbers
 740 or m/z ratios. A set of such spectra, that is, the probing at various time points, as
 741 in GC-MS, or locations, as in Raman imaging, constitute a multivariate dataset for one
 742 sample, i.e., chromatogram or image. In order to visualize the chromatogram based on
 743 the MS spectra, the Total Ion Current (TIC) is calculated as the sum of m/z ratio in-
 744 tensities at each time point. Similarly, for a Raman image, the intensity of a single pixel
 745 results from the sum of intensities of all spectral channels, called the global intensity.
 746 The concept is illustrated in Figure 1.6.

747 Before analyzing such a dataset, the data must be arranged in a matrix where the
 748 samples are arranged row-wise, while the variables are located in the columns (Fel-
 749 ten et al., 2015; Peré-Trepaut et al., 2005). Thus, for a GC-MS dataset, the variables
 750 (columns) are the m/z ratios, and the time points the rows. A Raman image has to be
 751 unfolded to fit into a matrix, as shown in Felten et al. (2015). In order to increase the
 752 robustness of the results or to compare various samples, the data of several chromato-
 753 graphic runs or Raman images can be combined into a multiset (Felten et al., 2015;
 754 Peré-Trepaut et al., 2005; Piqueras et al., 2015). In a multiset the matrix is augmented
 755 column-wise, that is, rows are added for each sample (Felten et al., 2015; Peré-Trepaut
 756 et al., 2005).

757 6.2 Pre-processing and exploratory analysis

758 Pre-processing is in many cases an essential step in the analysis of complex datasets and
 759 aims to remove any effects in the data that disturb the linear relationship of absorbance
 760 and concentration, as stated by Beer's law (Eq. 1.13).

$$A = \eta lc \quad (1.13)$$

761 , where A is the absorbance, η the molar absorptivity, l the path length and c the
 762 concentration (Åsmund Rinnan et al., 2009).

763 These effects certainly differ for different measurement techniques, but consist of
 764 random and/or systematic noise. After inspection of the raw data, a large pool of
 765 techniques offers possibilities to remove the latter. Furthermore, it may be relevant to
 766 additionally remove unnecessary variables (i.e., areas with no signal from the sample)
 767 or samples (i.e., outliers) to speed up processing times for big data sets and/or improve
 768 model quality. Some datasets may also require normalization or other forms of scaling
 769 in order to make data from different samples comparable. (Brereton, 2003; Åsmund
 770 Rinnan et al., 2009)

771 In order to decide on the right pre-processing method(s), a close look has to be
 772 taken at the results of a subsequent exploratory analysis, for example using Principal
 773 Component Analysis. Without these measures, potential outliers may not be identified
 774 and removed from the data (decreasing the model fit), or even worse, artifacts due to
 775 pre-processing may introduce and falsify the interpretation of results. The strategies
 776 used for the MS and Raman data in this work are outlined below. Inspiration for the
 777 treatment of Raman imaging data was drawn from the recommendations in Felten et al.
 778 (2015).

779 Pre-processing of MS data

780 A separate dataset was constructed for each species and solvent. The MS datasets were
 781 run as multisets, each containing a total of 9 submatrices: 6 samples (2 trees, 3 heights),
 782 2 internal standard and 1 blank. Inspired by Peré-Trepaut et al. (2005), the pure standard
 783 samples were added to allow relative quantification based on the peak area, provided
 784 separate components would be obtained for them⁸. The reason for adding the data from
 785 totally blank samples was to ensure that no m/z from column artefacts would make it
 786 into the pure spectra. Along with the building of the multiset as described in chapter
 787 6.2, descriptive variables for each row were created, indicating the sample (i.e. Tree 1,
 788 Tree 2, Blank, ISTD 1 (HIA) and ISTD 2 (BET) or the height (i.e. 1.3 m, 6 m, 11 m
 789 and NaN for blanks and ISTDs).

⁸Which was not the case, but had another interesting side effect on the deconvolution of the datasets - see Chapter ??

Because these datasets were too big to be analyzed with a bench-top computer, a strategy for the reduction of data without significant data loss had to be found. Existing routines described in the literature were tested, but could not be implemented directly, either because the data were changed in unpredictable ways (Tauler et al., 2015) or because a different data structure was needed (Bedia et al., 2016). These trials are not described in more detail, but the solution deduced for this thesis is detailed below.

The steps below were followed for each sample and matrix:

- Step 1: Import each sample matrix to the MATLAB workspace, with the retention times in rows and the mass spectral channels in columns. The time domain of each sample was cut at 5 and 32 min.
- Step 2: Row-wise concatenation of the sample-matrices. Blanks and the internal standards were included later (see Steps 5-7).
- Step 3: Removal of systematic noise (column artifacts) - Exploratory data analysis showed that the m/z ratios of 40 and 44 were unspecific column artifacts by applying a median filter to the data (Figure 1.7a). Similarly, an ion at 207 appeared at the end of the runs (Figure 1.7b). Note that in some datasets other ions were additionally found to be artifacts, which will be mentioned where appropriate.
- Step 4: Filtration of the time domain (reduction of rows) - Every time unit where the sum of m/z ratios was below a manually selected threshold were excluded from the dataset, i.e., areas without peaks were removed (Figure 1.7c).
- Step 5: Reduction of rows of the blank - the sample with least RTs left was used as a scaffold for the removal of the same rows in the blank.
- Step 6: Reduction of the rows of ISTDs - only the RTs where ISTD-peaks appeared were kept by manual selection.
- Step 7: Row-wise concatenation of pre-filtered blanks and ISTD matrices below the sample matrices.
- Step 8: Filtration of the m/z domain (columns, optional) - the mean of every m/z channel was calculated and a threshold set at (typically) 2% of the highest mean. All columns with a mean m/z below 2 % of the highest were excluded (Figure 1.7d).

820 Pre-processing of Raman data

821 The procedure used is described in Paper I (Appendix A) and more background information shall be given below. The same procedure was used for images of spruce
822 and larch heartwood, with slight differences for some parameters. Using MATLAB, all
823

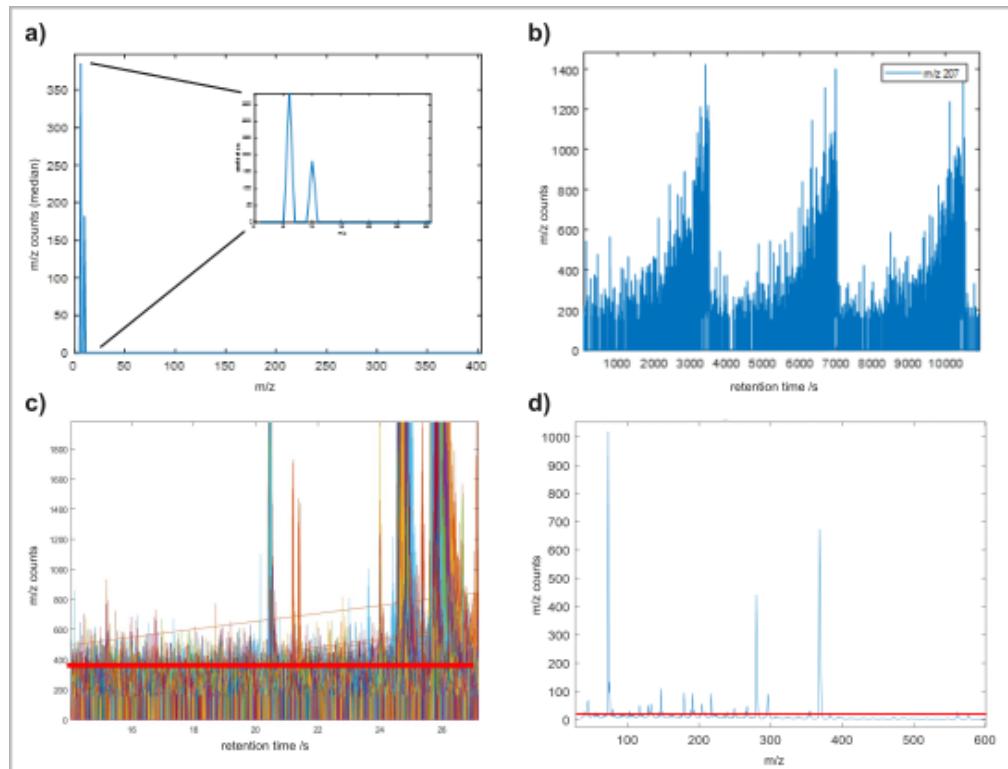


Figure 1.7: Illustration of several steps in the pre-processing of the MS datasets. a) Median of the dataset showing m/z 40 and 44 with much higher intensity than any other ions and were removed. b) m/z 207 appears with higher intensity at the end of each run and was also removed c) Filtration of the time domain by a manually selected threshold (exemplified by the red line) on the sum of m/z ratios present at each time point. Removes areas of the chromatogram without peaks. d) Filtration of the m/z domain by calculating the mean intensity of every m/z ratio and setting a threshold at 2 % of the highest mean (red line).

spectra of an image were first overlain to find obvious outliers based on the spectrum. The corresponding pixels were removed.

As radiation from outer space frequently hits the detector and causes large spikes in the spectra, cosmic ray removal is an important initial step in the pre-processing of Raman spectra. Many spectra of an image are usually affected this way and contribute to noise in the data, which in turn worsens the quality of any subsequent analysis. A relatively simple way to remove cosmic rays without distorting the spectrum significantly, is the use of a median filter of 3rd order along the spectral channels. Third order refers to the window size of three, which means that the intensity of a specific wavenumber is replaced by the median of the intensities of the two neighboring image pixels.

The next step is to remove the background radiation from the spectra, arising due to fluctuations of the optical focus (caused by sample unevenness) and varying degrees of auto-fluorescence in different areas of the image (due to e.g., lignin-rich middle lamellae or phenolic extractives). A very flexible approach for background correction is the use of asymmetric least squares algorithm to find the optimal parameters of the Whittaker smoother (Eilers, 2004). By selecting very small values for p and large values for λ , the algorithm will adapt to the valleys of the signal - the background. An example of the modeled baseline is shown in Figure 1.8, based on a spectrum obtained from a spruce resin channel. Nevertheless, at times the differences between the backgrounds in one image require the selection of different values for p and λ . Thus, the image pixels were first clustered using k-means cluster analysis (detailed in section 6.3) and the parameters adapted as necessary to each of the clusters.

The short integration time per spectrum has the drawback of reducing the signal-to-noise ratio. This is especially true in the case of measurements with the near-IR laser (785 nm) where the Raman effect is weaker and thus peaks intrinsically smaller. In order to improve the signal-to-noise ratio, Principle Component Analysis (detailed in section 6.3) was used for denoising. This was done by reconstructing the spectra from 10-15 principal components, as recommended in _____

Finally, the first derivatives (first order) of the spectra were calculated with a Savitzky-Golay filter which is commonly used in spectroscopy (Åsmund Rinnan et al., 2009). This step removes residual background contributions (smund Rinnan et al., 2009). Furthermore, the spectral resolution is somewhat increased, because the maxima of broad peaks are better distinguished, as they appear as inflection points (usually at the zero crossing) in the first derivative. The result is better models in the subsequent analysis. The Savitzky-Golay filter is a symmetric filter that calculates a user selected polynomial (quadratic in our case) for a number of points in a user selected window (11 in our case) to replace the central value in the window based on the coefficients obtained for the fit. The data are thereby smoothed and optionally derivatized (Åsmund Rinnan et al., 2009). Due to the symmetry of the window, variables are lost at each end. This did not pose a problem in our case, since there are no important signals at the edges of the Raman spectra recorded.

@Lisbeth:
could you
please check
if this is men-
tioned in the
book you got
with the Ra-
man instru-
ment? I can-
not find this
info in any of
my sources...

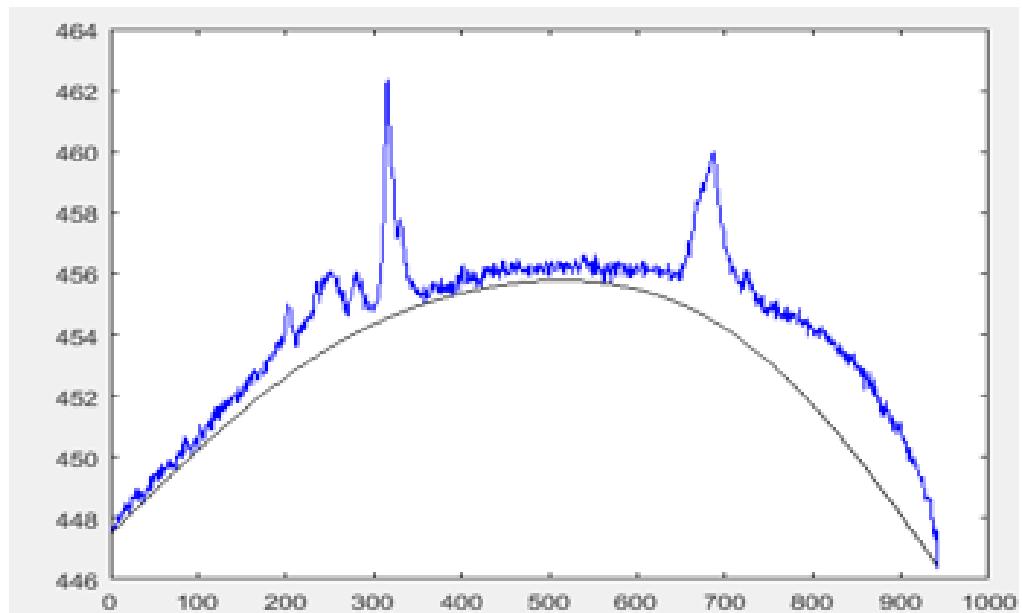


Figure 1.8: Example of a baseline calculated using the asymmetric least squares algorithm after Eilers (2004).

6.3 Multivariate Analysis Tools

Alternative to displaying the TIC or the global intensity map, the intensity of just one m/z ratio or Raman band can be used to show the temporal, respectively spatial distribution of a certain kind of molecule. Multivariate analysis tools such as those introduced below represent a more holistic approach to data analysis, as they exploit of the vast amount of variables available to decode the phenomena underlying the data. This is, instead of discarding most of the information obtained and analyzing just single bands, as is done in univariate analysis.

Both, Principal Component Analysis, as well as Multivariate Curve Resolution-Alternating Least Squares are data decomposition methods based on a bilinear model (Felten et al., 2015) of data governed by Beer's law. The model assumes, that any data matrix X can be approximated by the product of it's constituents S^T , i.e., individual spectral signatures, and their respective concentrations C (Equation 1.14) (Brereton, 2003).

$$X \approx C.S^T \quad (1.14)$$

Once the pre-processing is completed, PCA and MCR-ALS can be applied on data matrices composed of Raman spectra and TICs of GC-MS with very similar considerations.

Principal Component Analysis

Principal Component Analysis (PCA) is one of the most basic multivariate analysis tools. It is a linear, unsupervised data decomposition method with a wide range of applications. As mentioned in the section on pre-processing, it can be used to reduce the dimensions of a dataset before any subsequent analysis. Spectral noise, if not predominant, can be filtered out of the data by reconstructing the dataset based on a limited number of PCs, i.e., those where the loadings do not describe noise (Brereton, 2003; Dieing et al., 2011). Using PCA, an overview of the relationship of samples is quickly obtained, as well as it gives insight into the variables causing the sample relations. As such, it is extremely handy, if not indispensable, for exploratory data analysis. An idea of the data behavior (e.g., outliers) and of how many components are needed to describe a system is quickly obtained. In this thesis it was used prior to all MCR-ALS analyses and some examples are shown for the analysis of GC-MS data in Chapter ??.

PCA is also well suited as a stand-alone analysis to obtain insight on factors influencing the similarity of samples. In Chapter ??, an example is shown where concentration data were analyzed to study the influence of an extraction solvent. Although PCA is not qualified for classification in the strict sense, it can give insight on the proximity of known to unknown samples. In Paper II, III and IV (Appendices B-D), it was used to assign known reference samples (or previously identified MCR-ALS component spectra) to unidentified MCR-ALS component spectra from Raman image analysis. Likewise, it

can be used to obtain an overview of the results of various MCR-ALS analyses or cluster analyses, grouping a large amount of unassigned spectra. This was done for MCR-ALS components and cluster average spectra in Paper III (Appendix C).

skip Theory

The aim of a PCA is to project the samples of a given data matrix into a new coordinate system such that the maximal variance can be explained. The new coordinate system is spanned by a number of Principle Components (PCs), which are linear combinations of the original variables, and independent of each other. Any two PCs can be plotted against each other in a scores plot, where the sample scores determine the location of the samples within the new coordinate system, and may show groupings of the data. The PCs may be interpreted as overarching effects causing the variable trends, and can be deduced from the variable loadings. In the case of spectral data, these are usually plotted PC by PC in order find the peaks that cause the sample distribution. Note that some peaks may be negatively oriented and should therefore be interpreted as negatively correlated.

The original matrix (X , with dimensions $I \times J$), can be approximated by the product of the sample scores (T ($I \times N$) and the transposed loading matrix (P ($N \times J$), plus the residuals (E) to account for the variance not described by the scores and loadings matrices. In this case, N corresponds to the number of PCs, which is determined by the rank of the matrix.

$$X = T \times P^T + E = U \times S \times V^T + E \quad (1.15)$$

Note that there are several ways to calculate principal components, and Singular Value Decomposition (SVD) is the default method used in MATLAB. The loadings matrix is termed V and normalized to length 1. The scores matrix is obtained by the product of $U \times S$, where U corresponds to the normalized matrix T and S ($I \times J$) to the singular values of T , that is, the square root of the eigenvalues. (Brereton, 2003; Mermet et al., 2004, Ch.8) The eigenvalues inform about the size of the vectors, which translates into how important they are to explain the variance. The explanatory power of the PCs can then be expressed as the percentage of the sum over all eigenvalues. Since the SVD method calculates all PCs simultaneously, the calculation should be truncated at a desired number of PCs to avoid memory issues.

In order to determine whether a PCA model accommodates well all the data points, or whether any of the data points should be excluded as outliers, the Q and T^2 statistics were used during exploratory analysis. The Q residuals for each sample are calculated as the sum of squares of their residuals and indicates how much variation of a sample was not explained by the model. It is thus a kind of "lack-of-fit" for each sample. Hotelling's T^2 is a measure for the distance of a sample to the model center and is calculated as the sum of normalized squared scores Biancolillo and Marini (2018); Inc. (b). In MATLAB Q and T^2 are plotted against each other. If a sample has both, high Q and T^2 values, it is advisable to remove it from the data.

Cross-validation is usually done to assess the predictive power of a calibration model, but is also used to assess whether the correct number of components was chosen for a PCA (Biancolillo and Marini, 2018; Diana and Tommasi, 2002). For PCA the data are systematically split into L subsets (cancellation groups) and L models are built, always leaving out one group. The values of the left out group are then predicted (Biancolillo and Marini, 2018). For each of the models the root mean square error of cross-validation (RMSECV) is calculated according to equation 1.16. For comparative reasons, the root mean square error of calibration (RMSEC) is also calculated using the same formula, but based on the model that includes all the data (no splitting).⁹ Both of these errors are then plotted against the number of principal components calculated. The error should be smallest when the model describes the data, but not the noise. Therefore, it can be assumed that the right complexity of the model is obtained where the RMSECV is the smallest (Biancolillo and Marini, 2018; Brereton, 2003). The latter tends to increase when the model becomes too complex, while the RMSEC keeps decreasing. The two errors should be in the same range, otherwise there is a problem with the model.

$$RMSECV = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (1.16)$$

, where N is the number of samples in the dataset, y_i is the measured response and \hat{y}_i the predicted one.

The choice of the method for the formation of the cancellation groups depends among others on the size and sample types (e.g., replicates) and arrangement of the data matrix (Inc., a). The Venetian blinds method systematically forms cancellation groups with every (L to $L+x$)th sample, where x is the thickness of the blind. The number of splits and the blind thickness determine how many samples are left out at once, but also how frequently samples are left out (Inc., a). In MATLAB, 10 % of the data are left out by default, but it was deemed more representative to test the models based on 3/4 of the data, that is, leaving out 25 %. This is true for cross-validations performed in Papers III and IV (Appendices C and D), where 4 or 5 data splits were done. In Paper II (Appendix B), it appears that the default options were used, since 10 splits were done.

The blind thickness was adapted as needed to obtain 25% of left-out-data for the cancellation groups, and was 1 or 2 samples per blind depending on the dataset.

cannot check,
have no ac-
cess to mat-
lab

Multivariate Curve Resolution-Alternating Least Squares

Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) is a decomposition method that results in chemically meaningful component spectra and their distri-

⁹Technically, the RMSEC is used on a training set during building of predictive models, hence "calibration", and the prediction error is assessed based on a test set which is not involved in the model optimization process.

bution in the sample space. This is opposed to PCA, where the loading spectra are rather abstract (e.g., could be negative and positive).

MCR-ALS is a powerful method to untangle complex signals as obtained from Raman spectra of wood cell walls. During this project, MCR-ALS was used to decompose Raman image datasets in order to detect the extractive distribution in spruce and larch heartwood (Paper II (Appendix B)), track fungal degradation of the cell wall at various stages (Paper III (Appendix C)) and to detect fungal agents inside the cell wall (Paper IV (Appendix D)). Furthermore, the heartwood formation of larch was studied by MCR-ALS analysis of synchrotron IR images in Paper V (Appendix E). Finally, the use of MCR-ALS for the relative quantification of extractives analyzed by GC-MS chromatograms was explored (Chapter ??).

Theory

The spectral data matrix X is decomposed into its constituent spectra S^T and their respective concentration profiles C at each measurement point (image pixel or time point), as well as the residual error E (Eq. 1.17). Note that literature often refers to the component spectra as "pure" spectra, in the sense that they cannot be further differentiated (Felten et al., 2015). This vocabulary was not adopted in this work, because it is misleading, especially if the reader is not acquainted with the algorithm.

$$X = C.S^T + E \quad (1.17)$$

Before starting the analysis, the number of components that should be modeled has to be set, as well as spectra to initialize the ALS calculation have to be provided. The number of components could be chosen by *a priori* knowledge of the system (e.g., a tracheid cell imaged by Raman spectroscopy in the wet state will typically result in four components: lignin, cellulose S2 with microfibrils oriented perpendicular to the incident beam polarization, cellulose S1 with microfibrils oriented parallel to the incident beam polarization and water). Alternatively, this number is found by looking at the eigenvalues of a PCA and determining at which point the percentage of variation explained by adding another component becomes negligible, that is, the curve flattens out. A similar conclusion should be drawn from the RMSECV of the PCA. Compounds with low concentrations will probably only be found if a higher number of components is modeled. The final number of components is ultimately found by inspecting the solutions of the model (clarity of distribution maps and shape of component spectra), as well as the R^2 value and lack-of-fit (LOF). The R^2 describes how much of the variation was explained by the model and is based on how much the ratio of residual sum of squares and total sum of squares differs from 1 (i.e., if the residuals are close to zero, R^2 will be close to 1). The LOF in turn is the squared difference of the mean response to the predicted one and should be in the range of the experimental noise (Felten et al., 2015).

The spectra for initialization of both vibrational and mass data are typically chosen by applying SIMPLE-to-use Inter-active Self-modeling Mixture Analysis (SIMPLISMA)

(Windig et al., 1992; Windig and Guilment, 1991). This algorithm looks for the purest spectra within the given data matrix to be modeled. For example, when using Raman spectroscopy to find extractives within the cell walls of heartwood tracheids, it was important to include ray cells with resin drops, so that a pure spectrum could be extracted for initialization.

The application of constraints on the algorithm is what enables the finding of chemically meaningful solutions. The amount of models fitting the data equally well is narrowed down, which reduces the risk of finding ambiguous solutions (e.g., intensity or rotational ambiguities) (de Juan and Tauler, 2006). A case where constraints alone did not solve the issue of rotational ambiguities was encountered in the data of Paper III (Appendix C) and is shortly outlined farther below. Depending on the problem, various constraints can be applied. In the case of spectral data the most relevant constrain certainly is non-negativity of concentrations, as there are no negative concentrations in chemistry. This is also available for the spectral profiles, but is only applicable if the zero-order data are used (i.e., without derivatives in Raman data and for MS data). The preprocessing has to be done carefully as well, so that the baseline is ensured to be positive. Furthermore, through normalization of the component spectra by division by the Euclidean norm, it is ensured that they are comparable and that the intensity information is represented in the concentration profiles (Felten et al., 2015; Peré-Trepap et al., 2005). This step is especially important for quantification. A constrain that can be highly relevant in the case of multiset analysis is the "correspondence amongst species" (cas) constrain. It allows to add information about the presence or absence of certain components in each of the samples in the form of a binary matrix. This constrain is part of the refining step of the model.

Once the algorithm converges, the concentration profiles are split into individual samples and visualized by refolding into images in the case of Raman data, or overlaid to show chromatograms in the case of MS data. For a multiset, the matrices of the individual samples have to be separated first. When the images are to be compared after plotting in MATLAB, care has to be taken that the color scale is the same for all images. Due to the use of derivatives for the analysis of Raman imaging data, the spectral profiles were transformed back into zero-order spectra as described in Paper II (Appendix B).

1044 Rotational ambiguity case

1045 Rotational ambiguity describes a situation where a model results in distorted component spectra, although having a small lack of fit value, similar to the "true" solution (de Juan
1046 and Tauler, 2006). The case of rotational ambiguity appears in the Raman data in
1047 the degraded spruce at HF multiset. It is the reason for which the lumen was removed
1048 in several images of this and subsequently other datasets. Although the first model
1049 had acceptable model statistics (i.e., LOF 6.0665 %, $R^2 = 99.632$), the last component
1050 spectrum was very distorted after reconstruction from first derivatives (Fig. 1.9a, red

find out why
this is not
bold

rectangle). What appeared as (extractive) droplets in the lumina of the second and fourth image of the sixth component, was also faintly visible in the third component (marked in red). The spectra both indicated lipids by their strong C-H stretching bands, but the distribution maps did not appear clear. For example, was the concentration score of image one, component 3 much weaker than in the other images, even though a ray was imaged (dashed, red rectangle in Fig. 1.9a). After building several models with the correspondence amongst species constrain, it turned out that the droplets were fungal hyphae (treated in chapter (??)). The rotational ambiguity was caused by fungal hyphae rich in lipids being present in the same dataset where oleoresin was also present, but mixed with other substances, and but very similar to each other, compared to the other component spectra. The pixels with signals from fungal hyphae were thus removed by KMCA and cas was applied to one of the images. The final model (Fig. 1.9b) was published in Paper III (Figure 3 in Appendix C) and had almost the same model statistics as the original one (LOF 6.0795 % , $R^2 = 99.6304$), but clean distribution profiles and all-positive component spectra with only little distortion. Additionally, the lignin biphenyl substructure component could be isolated. This case also demonstrates how important it is not to focus merely on the model statistics.

1069 K-means Cluster Analysis

1070 K-means Cluster Analysis (KMCA) is an unsupervised clustering method, used to group
1071 multivariate data points into clusters based on the minimization of the within-cluster
1072 variance of all clusters, that is, the smallest distance of each data point to the cluster
1073 mean (centroid) of each of the clusters. First the number of clusters k has to be set by
1074 the user. An estimate can again be obtained by looking at the eigenvalues of a PCA, or
1075 by *a priori* knowledge.

1076 In MATLAB, the k-means++ algorithm is used, which improves the speed at which
1077 the initial cluster centroids are found (Arthur and Vassilvitskii, 2007; Inc., c), as com-
1078 pared to the classical Lloyd's algorithm (Lloyd, 1982). The algorithm uses a different
1079 approach to selecting the initial cluster centroids, leading to improved clustering so-
1080 lutions. Nevertheless, to increase the chances of finding a solution close to or at the
1081 global minimum, the algorithm should be run several times (e.g., 5 times) and the best
1082 solution chosen. The first centroid is selected at random and the distance to each other
1083 data point computed. The following centroids up to k are subsequently chosen from the
1084 remaining data points at random too, but using a probability function so that points
1085 with a bigger distance to the previously chosen centroids are selected with a higher
1086 probability. (Inc., c)

1087 After determining the initial centroids, each datapoint is assigned to the centroid to
1088 which it has the smallest distance. Based on all the datapoints assigned to the same
1089 cluster, a new centroid is computed. The procedure is repeated until the sum of squared
1090 distances of all clusters cannot be further minimized reassigning data points to other
1091 clusters (Inc., c).

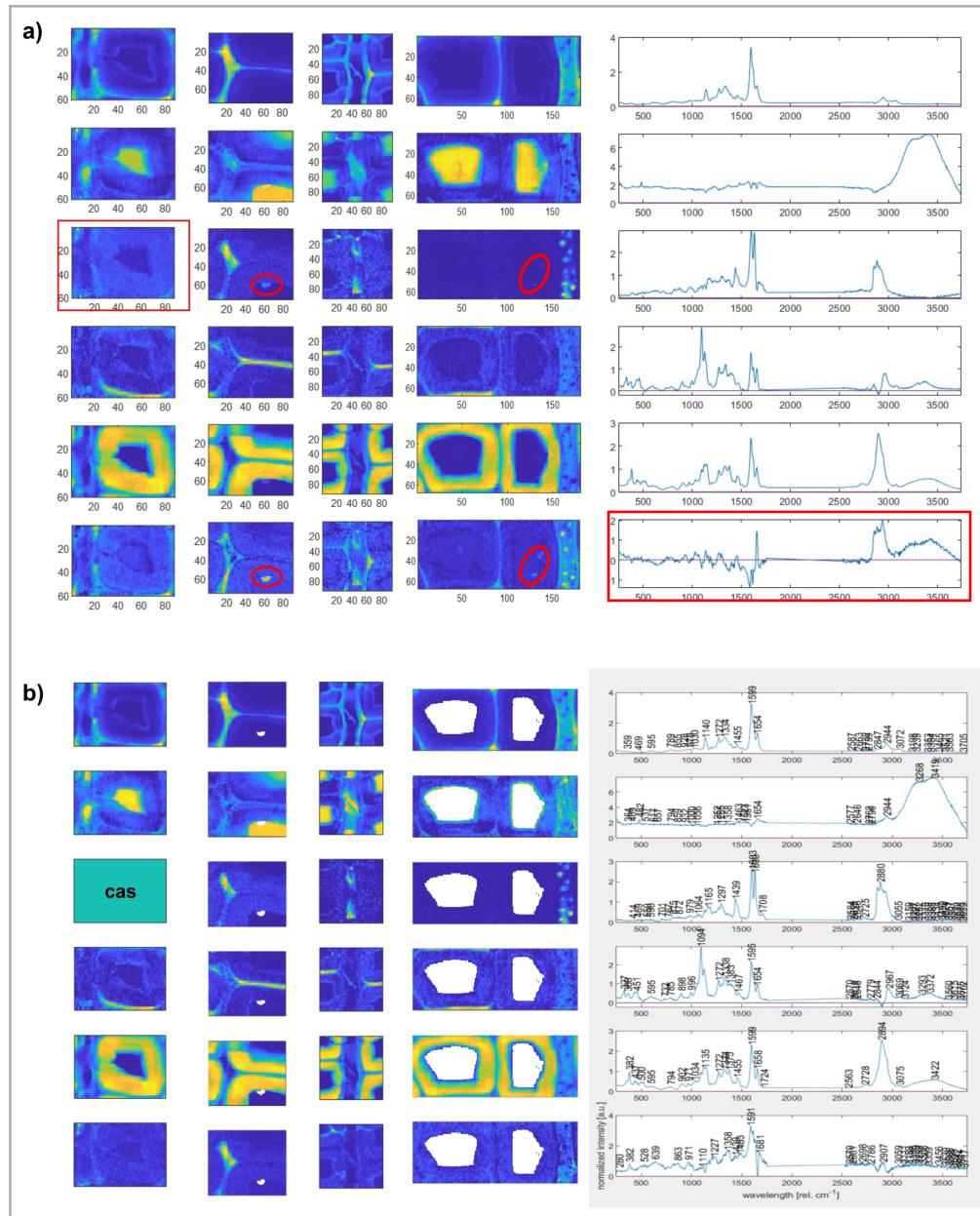


Figure 1.9: Rotational ambiguity issue of MCR-ALS encountered in a Raman multiset of native, degraded spruce tissue. **a)** original dataset (NAT at HF), with 6th spectral component distorted after reconstruction from first derivative (red rectangle), droplets visible in the lumina of tracheids in image 2 and 4 of components 3 and 6 (encircled in red) and at least one image with no extractives present in the ray (red dashed rectangle); LOF = 6.0665 %; $R^2 = 99.632$ **b)** final dataset, with pixels containing fungal hyphae removed; LOF 6.0795 %, R^2 99.6304.

In this work, KMCA was used on Raman data only, but for two distinct purposes: as part of the pre-processing flow and as an actual analytical technique (Paper III, Appendix C). During preprocessing of the Raman spectra (Papers II, III and IV, Appendices B-D), KMCA was used to find spectra with distinct backgrounds in order to individually adapt the parameters used for the asymmetric least squares algorithm to each cluster. For most, but not all datasets, the backgrounds were so similar that the same parameters could be used for all clusters.

In Paper III (Appendix C), KMCA was applied to the concentration profiles resulting from the MCR-ALS analysis of native and extracted Spruce images with and without fungal degradation. Due to the significant reduction of dimensions from > 700 spectral variables to a set of 5-7 variables identified as chemical compounds, the resulting cluster distribution maps showed a very distinct and sharp separation of the anatomical units of the cells. This was further improved by color coding the clusters according to their L/C ratios, as detailed in Paper III (Appendix C).

1106 References

- 1107 Alba synchrotron, miras beamline, 2023. URL <https://www.albasynchrotron.es/en/beamlines/bl01-miras>.
- 1109 Stergios Adamopoulos and Elias Voulgaridis. Effect of hot-water extractives on
1110 water sorption and dimensional changes of black locust wood. *Wood Research*,
1111 57:69–78, 2012. ISSN 13364561. URL <http://users.teilar.gr/~mantanis/guest-work-Adamopoulos-Voulgaridis.pdf>. does not use the real sorption
1112 isotherm, but the scanning isotherm (fredriksson2018)

no differences
1113 found between EMC after treatments

their arguments why it did not
1114 work
- inefficient extraction (.
- 1116 Umesh P Agarwal. An overview of raman spectroscopy as applied to lignocellulosic
1117 materials. *Advances in Lignocellulosics Characterization*, pages 201–225, 1999.
- 1118 Umesh P. Agarwal. Analysis of cellulose and lignocellulose materials by raman spec-
1119 troscopy: A review of the current status. *Molecules*, 24:1659, 4 2019. ISSN 1420-
1120 3049. doi: 10.3390/molecules24091659. URL <https://www.mdpi.com/1420-3049/24/9/1659>.
- 1122 U.P. Agarwal and R.H. Atalla. Chapter 4: Vibrational spectroscopy, 2010.
- 1123 David Arthur and Sergei Vassilvitskii. k-means++: the advantages of careful seeding.
1124 2007. doi: 10.5555/1283383.1283494.
- 1125 Peter Atkins and Julio De Paula. *PHYSICAL CHEMISTRY*. W.H. Freeman and
1126 Company, 8 edition, 2006. ISBN 0-7167-8759-8.
- 1127 Keith D. Bartle and Peter Myers. History of gas chromatography. *TrAC - Trends in An-
1128 alytical Chemistry*, 21:547–557, 9 2002. ISSN 01659936. doi: 10.1016/S0165-9936(02)
1129 00806-3.
- 1130 Carmen Bedia, Romà Tauler, and Joaquim Jaumot. Compression strategies for the
1131 chemometric analysis of mass spectrometry imaging data. *Journal of Chemometrics*,
1132 30:575–588, 2016. ISSN 1099128X. doi: 10.1002/cem.2821.
- 1133 Alessandra Biancolillo and Federico Marini. *Chemometrics Applied to Plant Spectral
1134 Analysis*. Elsevier B.V., 1 edition, 2018. doi: 10.1016/bs.coac.2018.03.003. URL
1135 <http://dx.doi.org/10.1016/bs.coac.2018.03.003>.
- 1136 Peter Bock, Paula Nousiainen, Thomas Elder, Markus Blaukopf, Hassan Amer, Ronald
1137 Zirbs, Antje Potthast, and Notburga Gierlinger. Infrared and raman spectra of lignin
1138 substructures: Dibenzodioxocin. *Journal of Raman Spectroscopy*, 51:422–431, 3 2020.
1139 ISSN 10974555. doi: 10.1002/jrs.5808. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jrs.5808>.

- 1141 Lynne Boddy. Effect of temperature and water potential on growth rate of wood-rotting
1142 basidiomycetes. *Transactions of the British Mycological Society*, 80:141–149, 2 1983.
1143 ISSN 00071536. doi: 10.1016/S0007-1536(83)80175-2. URL <https://linkinghub.elsevier.com/retrieve/pii/S0007153683801752>.
- 1145 Richard G. Brereton. *Chemometrics: Data Analysis for the Laboratory and Chemical
1146 Plant*. John Wiley Sons Ltd, 2003. ISBN 0-471-48977-8 (HB); 0-471-48978-6 (PB).
- 1147 Christian Brischke, Arved Soetbeer, and Linda Meyer-Veltrup. The minimum moisture
1148 threshold for wood decay by basidiomycetes revisited. a review and modified pile
1149 experiments with norway spruce and european beech decayed by coniophora puteana
1150 and trametes versicolor. *Holzforschung*, 71:893–903, 2017. ISSN 1437434X. doi:
1151 10.1515/hf-2017-0051.
- 1152 Fergal P. Byrne, Saimeng Jin, Giulia Paggiola, Tabitha H. M. Petchey, James H. Clark,
1153 Thomas J. Farmer, Andrew J. Hunt, C. Robert McElroy, and James Sherwood. Tools
1154 and techniques for solvent selection: green solvent selection guides. *Sustainable Chem-
1155 ical Processes*, 4, 12 2016. doi: 10.1186/s40508-016-0051-z.
- 1156 Thomas G. Chasteen. Split/splitless and on-column gas chromatographic injectors,
1157 2023. URL https://www.shsu.edu/~chm_tgc/GC/GCinject.html.
- 1158 Kun Cheng, Hagit Sorek, Herbert Zimmermann, David E. Wemmer, and Markus Pauly.
1159 Solution-state 2d nmr spectroscopy of plant cell walls enabled by a dimethylsulfoxide-
1160 d 6 /1-ethyl-3-methylimidazolium acetate solvent. *Analytical Chemistry*, 85:3213–
1161 3221, 2013. ISSN 0003-2700. doi: 10.1021/ac303529v. URL <http://pubs.acs.org/doi/10.1021/ac303529v>.
- 1163 E. Choong and S. Achmadi. Effects of extractives on moisture sorption and shrinkage
1164 in tropical woods. *Wood and fiber science*, 23:185–196, 1991. ISSN 0735-6161.
- 1165 P. Michael Conn, editor. *Techniques in Confocal Microscopy: Reliable Lab Solutions*.
1166 Elsevier Inc., 2010. ISBN 978-0-12-384658-7. URL <http://site.ebrary.com/id/10399297>.
- 1168 W. A. Côté, A. C. Day, B. W. Simson, and T. E. Timell. Studies on larch
1169 arabinogalactan. i. the distribution of arabinogalactan in larch wood. *Holz-
1170 forschung*, 20:178–192, 1 1966. ISSN 0018-3830. doi: 10.1515/hfsg.1966.20.
1171 6.178. URL <https://www.degruyter.com/view/j/hfsg.1966.20.issue-6/hfsg.1966.20.6.178.hfsg.1966.20.6.178.xml>. printed - can't find the print??
- 1173 P. L. Davies. Polarizabilities of long chain conjugated molecules. *Transactions of the
1174 Faraday Society*, 48:789, 1952. ISSN 0014-7672. doi: 10.1039/tf9524800789.

- 1175 A. Dazzi, R. Prazeres, F. Glotin, and J. M. Ortega. Local infrared microspectroscopy with subwavelength spatial resolution with an atomic force microscope tip
1176 used as a photothermal sensor. *Optics Letters*, 30:2388, 9 2005. ISSN 0146-9592.
1177 doi: 10.1364/OL.30.002388. URL <https://opg.optica.org/abstract.cfm?URI=ol-30-18-2388>.
- 1180 Alexandre Dazzi and Craig B. Prater. Afm-ir: Technology and applications in nanoscale
1181 infrared spectroscopy and chemical imaging. *Chemical Reviews*, 117:5146–5173, 4
1182 2017. ISSN 0009-2665. doi: 10.1021/acs.chemrev.6b00448. URL <https://pubs.acs.org/doi/10.1021/acs.chemrev.6b00448>.
- 1184 Alexandre Dazzi, Craig B. Prater, Qichi Hu, D. Bruce Chase, John F. Rabolt, and Curtis
1185 Marcott. Afm-ir: Combining atomic force microscopy and infrared spectroscopy for
1186 nanoscale chemical characterization. *Applied Spectroscopy*, 66:1365–1384, 12 2012.
1187 ISSN 0003-7028. doi: 10.1366/12-06804. URL <http://journals.sagepub.com/doi/10.1366/12-06804>.
- 1189 Anna de Juan and Romà Tauler. Multivariate curve resolution (mcr) from 2000:
1190 Progress in concepts and applications, 7 2006. ISSN 10408347.
- 1191 Maurice Defo, Yves Fortin, and Alain Cloutier. Moisture content-water potential relationship
1192 of sugar maple and white spruce wood from green to dry conditions. *Wood and Fiber Science*,
1193 31:62–70, 1999. ISSN 07356161.
- 1194 Giancarlo Diana and Chiara Tommasi. Cross-validation methods in principal component analysis:
1195 A comparison. *Statistical Methods and Applications*, 11:71–82, 2 2002. ISSN 1613981X. doi:
1196 10.1007/BF02511446. URL http://venus.unive.it/romanaz/ada2/cross-validation_pca.pdfhttp://link.springer.com/article/10.1007/BF02511446%5Cnhttp://link.springer.com/10.1007/BF02511446http://link.springer.com/10.1007/BF02511446.
- 1200 Thomas Dieing, Olaf Hollricher, and Jan Toporski, editors. *Confocal Raman microscopy*,
1201 volume 158. Springer Series in Optical Sciences, 2011. ISBN 978-3642125218.
- 1202 Lloyd A. Donaldson, Adya Singh, Laura Raymond, Stefan Hill, and Uwe Schmitt. Extractive
1203 distribution in pseudotsuga menziesii: effects on cell wall porosity in sapwood and heartwood.
1204 *IAWA Journal*, 40:721–740, 11 2019. ISSN 0928-1541. doi:
1205 10.1163/22941932-40190248. URL https://brill.com/view/journals/iawa/40/4/article-p721_6.xml.
- 1207 Paul H.C. Eilers. Parametric time warping. *Analytical Chemistry*, 76:404–411, 2004.
1208 ISSN 00032700. doi: 10.1021/ac034800e.
- 1209 Dag Ekeberg, Per-Otto Flæte, Morten Eikenes, Monica Fongen, and Carl Fredrik Naess-
1210 Andresen. Qualitative and quantitative determination of extractives in heartwood of

- 1211 scots pine (*pinus sylvestris* l.) by gas chromatography. *Journal of Chromatography A*, 1109:267–272, 3 2006. ISSN 00219673. doi: 10.1016/j.chroma.2006.01.027. URL
1212 http://ac.els-cdn.com/S0021967306001397/1-s2.0-S0021967306001397-main.pdf?_tid=4dfa625e-7787-11e7-8109-0000aacb35f&acdnat=1501681049_a6fc537a91e7abccfa84c5ee6ee887a4<http://www.sciencedirect.com/science/article/B6TG8-4J616H4-6/2/28915b49b06c401265b458f1f>. extraction efficacy for
1213 differently ground materials vs whole wood.
- 1218 Rainer Ekman. Analysis of lignans in norway spruce by combined gas chromatog-
1219 raphy — mass spectrometry. *Holzforschung*, 30:79–85, 1 1976. ISSN 0018-3830.
1220 doi: 10.1515/hfsg.1976.30.3.79. URL <https://www.degruyter.com/downloadpdf/j/hfsg.1976.30.issue-3/hfsg.1976.30.3.79.pdf>
1221 <https://www.degruyter.com/view/j/hfsg.1976.30.issue-3/hfsg.1976.30.3.79/hfsg.1976.30.3.79.xml>
1222 <https://www.degruyter.com/downloadpdf/j/hfsg.1976.30.3.79.pdf>. not specific for HW or SW –gt; all together.
- 1225 Rainer Ekman and Bjarne R Holmbom. Analysis by gas chromatography of the wood
1226 extractives in pulp and water samples from mechanical pulping of spruce. *Nordic
1227 Pulp and Paper Research Journal*, pages 16–24, 1989.
- 1228 Wenwen Fang, Jarl Hemming, Markku Reunanen, Patrik Eklund, Enma Conde Pineiro,
1229 Ida Poljansek, Primoz Oven, and Stefan Willför. Evaluation of selective extraction
1230 methods for recovery of polyphenols from pine. *Holzforschung*, 67:843–851, 2013.
1231 ISSN 00183830. doi: 10.1515/hf-2013-0002. printed

ask them
-
1232 was hemicellulose observed in the hpsec at all temperatures?
- why did you
1233 ultimately chose the hex+ac/w system for the rest of the extractions?
- 1234 Martin Felhofer, Batirtze Prats-Mateu, Peter Bock, and Notburga Gierlinger. Anti-
1235 fungal stilbene impregnation: Transport and distribution on the micron-level. *Tree
1236 Physiology*, 38:1526–1537, 2018. ISSN 17584469. doi: 10.1093/treephys/tpy073.
1237 stilbenes associated to lipids in CC and CML

impregnation of spruce
1238 w/ ethanol extract of pine –gt; pinosylvin+lipids also end up in the CC and
1239 CML

deposits found in lumina with higher lipid amounts in sapwood,
1240 higher aromatic content in heartwood.
- 1241 Judith Felten, Hardy Hall, Joaquim Jaumot, Romà Tauler, Anna De Juan, and
1242 András Gorzsás. Vibrational spectroscopic image analysis of biological mate-
1243 rial using multivariate curve resolution-alternating least squares (mcr-als). *Nature
1244 Protocols*, 10:217–240, 2015. ISSN 17502799. doi: 10.1038/nprot.2015.008.
1245 URL <http://dx.doi.org/10.1038/nprot.2015.008%5Cnhttp://www.nature.com/doifinder/10.1038/nprot.2015.008>. printed.
- 1247 Dietrich Fengel and Gerd Wegener. *Wood, Chemistry, Ultrastucture, Reactions*. Berlin
1248 ; New York : de Gruyter,, 1989. ISBN 3110120593.

- 1287 Eigenvector Research Inc. Using cross-validation, a. URL https://www.wiki.eigenvector.com/index.php?title=Using_Cross-Validation.
- 1288
- 1289 Eigenvector Research Inc. T-squared q residuals and contributions, b. URL https://www.wiki.eigenvector.com/index.php?title=T-Squared_Q_residuals_and_Contributions.
- 1290
- 1291
- 1292 Eigenvector Research Inc. kmeans, c. URL <https://de.mathworks.com/help/stats/kmeans.html>.
- 1293
- 1294 Peter Larkin. *IR and Raman Spectroscopy: Principles and Spectral Interpretation*. El-
1295 sevier Inc., 2011. ISBN 9780123869845. doi: 10.1016/B978-0-12-386984-5.10001-1.
1296 URL <http://medcontent.metapress.com/index/A65RM03P4874243N.pdf> <http://linkinghub.elsevier.com/retrieve/pii/B9780123869845100011>.
- 1297
- 1298 Peter Lasch and Dieter Naumann. Spatial resolution in infrared microspectroscopic
1299 imaging of tissues. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1758:
1300 814–829, 7 2006. ISSN 00052736. doi: 10.1016/j.bbamem.2006.06.008. URL <https://linkinghub.elsevier.com/retrieve/pii/S0005273606002197>.
- 1301
- 1302 D. Li, J. Park, and J. R. Oh. Silyl derivatization of alkylphenols, chlorophenols, and
1303 bisphenol a for simultaneous gc/ms determination. *Analytical Chemistry*, 73:3089–
1304 3095, 2001. ISSN 00032700. doi: 10.1021/ac001494l. silylation: use acetone as
1305 solvent to increase reaction rate! complete reaction in less than 15s at RT!! mixtures
1306 above 60heavily increase the reaction rate.

also: deactivation of the ex-
1307 cess reagent with a bit of water, then drying over Na₂SO₄ –gt; long term storage.
- 1308
- 1309 S. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*,
28:129–137, 3 1982. ISSN 0018-9448. doi: 10.1109/TIT.1982.1056489.
- 1310 Derek A. Long. *The Raman Effect*. Wiley, 4 2002. ISBN 9780471490289. doi: 10.1002/
1311 0470845767.
- 1312 Anni Lähdetie, Paula Nousiainen, Jussi Sipilä, Tarja Tamminen, and Anna-Stiina
1313 Jääskeläinen. Laser-induced fluorescence (lif) of lignin and lignin model com-
1314 pounds in raman spectroscopy. *Holzforschung*, 67:531–538, 7 2013. ISSN
1315 1437-434X. doi: 10.1515/hf-2012-0177. URL <https://www.degruyter.com/view/j/hfsg.2013.67.issue-5/hf-2012-0177/hf-2012-0177.xml> <https://www.degruyter.com/doi/10.1515/hf-2012-0177>. for thesis?
- 1316
- 1317
- 1318 Lucy S. Martin, Stanislav Jelavić, Simon M. Cragg, and Lisbeth G. Thygesen. Furfury-
1319 lation protects timber from degradation by marine wood boring crustaceans. *Green
1320 Chemistry*, 23:8003–8015, 2021. ISSN 1463-9262. doi: 10.1039/D1GC01524A. URL
1321 <http://xlink.rsc.org/?DOI=D1GC01524A>.

- 1322 H. McDonald, S. Morsch, and S. M. Rowland. Chemical analysis of tree growth in epoxy
1323 resin using afm-ir spectroscopy. *IEEE Transactions on Dielectrics and Electrical*
1324 *Insulation*, 27:773–781, 6 2020. ISSN 15584135. doi: 10.1109/TDEI.2019.008573.
- 1325 Jean-Michel Mermet, Matthias Otto, Miguel Valcárcel, Robert A. Kellner, and
1326 H. Michael Widmer, editors. *Analytical Chemistry: a modern approach to analytical*
1327 *science*. WILEY-VCH Verlag GmbH Co, KGaA Weinheim, 2 edition, 2004. ISBN
1328 3527305904. URL https://archive.org/details/analyticalchemis0000unse_u6v0/. Ch 8 - chemometrics
Ch 21 - chromatography
ch25 - MS (+GC-
1329 MS).
- 1331 Annette Naumann, Ursula Kües, and Andrea Polle. Fourier transform infrared
1332 microscopy in wood analysis, 2007. URL <https://www.researchgate.net/publication/262179733>.
- 1334 Linda Nisula. Wood extractives in conifers a study of stemwood and knots of industrially
1335 important species, 2018.
- 1336 M. Nuopponen, S. Willför, A. S. Jääskeläinen, A. Sundberg, and T. Vuorinen.
1337 A uv resonance raman (uvrr) spectroscopic study on the extractable com-
1338 pounds of scots pine (*pinus sylvestris*) wood: Part i: Lipophilic compounds.
1339 *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 60:
1340 2953–2961, 11 2004. ISSN 13861425. doi: 10.1016/j.saa.2004.02.008. URL
1341 http://ac.els-cdn.com/S1386142504000575/1-s2.0-S1386142504000575-main.pdf?_tid=d4228aca-7780-11e7-aa6b-00000aacb360&acdnat=1501678269_a9c16104d1a1a0e43a8cdea7bf02c9c5http://linkinghub.elsevier.com/retrieve/pii/S1386142504000575
- 1345 Pascal Nzokou and D. Pascal Kamdem. Influence of wood extractives on moisture sorp-
1346 tion and wettability of red oak (*quercus rubra*), black cherry (*prunus serotina*), and
1347 red pine (*pinus resinosa*). *Wood and Fiber Science*, 36:483–492, 2004. ISSN 07356161.
1348 in conclusions they say, that the difference in EMC between samples that were toluene-
1349 ethanol and water extracted was smaller as compared to the only ethanol and water
1350 extracted samples. They hypothesize that migration of hydrophobic substances in the
1351 subsequent water extraction might cause this.

heartwood; HOW MANY
1352 SAMPLES?
- 1353 L. A. Ostroukhova, V. A. Raldugin, V. A. Babkin, N. A. Onuchina,
1354 and A. A. Levchuk. Investigation of the chemical composition of larch
1355 wood resin. *Russian Journal of Bioorganic Chemistry*, 38:775–779, 12
1356 2012. ISSN 1068-1620. doi: 10.1134/S1068162012070151. URL <https://link.springer.com/content/pdf/10.1134%2FS1068162012070151.pdfhttp://link.springer.com/10.1134/S1068162012070151>. printed

this

- 1359 publication is really not well researched, and there are assumptions without
1360 references.
- 1361 Emma Peré-Trepaut, Silvia Lacorte, and Romà Tauler. Solving liquid chromatography
1362 mass spectrometry coelution problems in the analysis of environmental samples by
1363 multivariate curve resolution. *Journal of Chromatography A*, 1096:111–122, 2005.
1364 ISSN 00219673. doi: 10.1016/j.chroma.2005.04.089.
- 1365 S. Piqueras, C. Krafft, C. Beleites, K. Egodage, F. von Eggeling, O. Guntinas-
1366 Lichius, J. Popp, R. Tauler, and A. de Juan. Combining multiset resolution
1367 and segmentation for hyperspectral image analysis of biological tissues. *Analyt-
1368 ica Chimica Acta*, 881:24–36, 6 2015. ISSN 00032670. doi: 10.1016/j.aca.2015.04.
1369 053. URL <http://dx.doi.org/10.1016/j.aca.2015.04.053><https://linkinghub.elsevier.com/retrieve/pii/S0003267015005802>.
- 1370 Batirtze Prats-Mateu, Peter Bock, Martina Schroppenegger, José Luis Toca-Herrera,
1371 and Notburga Gierlinger. Following laser induced changes of plant phenyl-
1372 propanoids by raman microscopy. *Scientific Reports*, 8:11804, 12 2018. ISSN 2045-
1373 2322. doi: 10.1038/s41598-018-30096-3. URL <http://www.nature.com/articles/s41598-018-30096-3>. read, notes are in the booklet with the wooden cover.
- 1375 Jonathan S. Schilling, Shona M. Duncan, Gerald N. Presley, Timothy R. Filley, Joel A.
1376 Jurgens, and Robert A. Blanchette. Colocalizing incipient reactions in wood de-
1377 graded by the brown rot fungus postia placenta. *International Biodeterioration and
1378 Biodegradation*, 83:56–62, 2013. ISSN 09648305. doi: 10.1016/j.ibiod.2013.04.006.
1379 printed.
- 1380 Francis W. M. R. Schwarze, Julia Engels, and Claus Mattheck. *Fungal Strategies of
1381 Wood Decay in Trees*. Springer Berlin Heidelberg, 2000. ISBN 978-3-642-63133-
1382 7. doi: 10.1007/978-3-642-57302-6. URL <http://link.springer.com/10.1007/978-3-642-57302-6>.
- 1383 Nelson L. Sefara and Mike Birkett. Development of an alternative solvent to replace ben-
1384 zene in the determination of organic soluble extractives in wood. *TAPPSA Journal*,
1385 2006. URL http://tappa.co.za/archive2/APPW_2004>Title2004/Development_of_an_alternative/_development_of_an_alternative_.html. toluene and
1386 dichloromethane (DCM) alone removed the lowest amount of extractives. DCM was
1387 much more effective in.
- 1388 Ewen Smith and Geoffrey Dent. *Modern Raman Spectroscopy - A Practical Approach*.
1389 John Wiley Sons, Ltd, 12 2005. ISBN 9780470011836. doi: 10.1002/0470011831.
1390 URL <http://doi.wiley.com/10.1002/0470011831>.

- LC-MS metabolomic data processing using chemometric tools</i> - Gorrochategui, Eva; Jaumot, Joaquim; Tauler, Roma Romà; Gorrochategui, Eva; Jaumot, Joaquim; Tauler, Roma Romà)

From Duplicate 1 (<i>A protocol for LC-MS metabolomic data processing using chemometric tools</i> - Gorrochategui, Eva; Jaumot, Joaquim; Tauler, Roma Romà)

downloaded the supplementary - contains the functions

From Duplicate 2 (<i>A protocol for LC-MS metabolomic data processing using chemometric tools</i> - Gorrochategui, Eva; Jaumot, Joaquim; Tauler, Roma Romà)

From Duplicate 2 (<i>A protocol for LC-MS metabolomic data processing using chemometric tools</i> - Gorrochategui, Eva; Jaumot, Joaquim; Tauler, Roma Romà)

downloaded the supplementary - contains the functions.
- Kerstin Wagner, Maurizio Musso, Stefan Kain, Stefan Willför, Alexander Petutschnigg, and Thomas Schnabel. Larch wood residues valorization through extraction and utilization of high value-added products. *Polymers*, 12:359, 2 2020. ISSN 2073-4360. doi: 10.3390/polym12020359. URL <https://www.mdpi.com/2073-4360/12/2/359>.
- Xinzhou Wang, Yuhe Deng, Yanjun Li, Kevin Kjoller, Anirban Roy, and Siqun Wang. In situ identification of the molecular-scale interactions of phenol-formaldehyde resin and wood cell walls using infrared nanospectroscopy. *RSC Advances*, 6:76318–76324, 8 2016. ISSN 20462069. doi: 10.1039/c6ra13159j. URL <https://pubs.rsc.org/en/content/articlehtml/2016/ra/c6ra13159jhttps://pubs.rsc.org/en/content/articlelanding/2016/ra/c6ra13159j>.
- Xinzhou Wang, Linguo Zhao, Yuhe Deng, Yanjun Li, and Siqun Wang. Effect of the penetration of isocyanates (pmdi) on the nanomechanics of wood cell wall evaluated by afm-ir and nanoindentation (ni). *Holzforschung*, 72:301–309, 3 2018. ISSN 1437434X. doi: 10.1515/hf-2017-0123. URL <https://www.degruyter.com/document/doi/10.1515/hf-2017-0123/html>.
- Stefan Willför, Jarl Hemming, Markku Reunanen, Christer Eckerman, and Bjarne Holmbom. Lignans and lipophilic extractives in norway spruce knots and stemwood. *Holzforschung*, 57:27–36, 1 2003a. ISSN 0018-3830. doi: 10.1515/HF.2003.005. URL <https://www.degruyter.com/downloadpdf/j/hfsg.2003.57.issue-1/hf.2003.005.pdfhttps://www.degruyter.com/view/j/hfsg.2003.57.issue-1/hf.2003.005.hf.2003.005.xml>. finnish spruce

detection by GC-MS

knotwood contained more hydrophilic than lipophilic extr., stemwood same range as prev. lit.
- Stefan Willför, Jarl Hemming, Markku Reunanen, and Bjarne Holmbom. Phenolic and lipophilic extractives in scots pine knots and stemwood. *Holzforschung*, 57:359–372, 1 2003b. ISSN 0018-3830. doi: 10.1515/HF.2003.054.

- 1473 URL <https://www.degruyter.com/downloadpdf/j/hfsg.2003.57.issue-4/hf.2003.054.pdf>
1474 <https://www.degruyter.com/view/j/hfsg.2003.57.issue-4/hf.2003.054.xml>.
- 1475
- 1476 Stefan Willför, Linda Nisula, Jarl Hemming, Markku Reunanan, and Bjarne Holm-
1477 bom. Bioactive phenolic substances in industrially important tree species.
1478 part 1: Knots and stemwood of different spruce species. *Holzforschung*,
1479 58:335–344, 7 2004. ISSN 0018-3830. doi: 10.1515/HF.2004.052. URL
1480 <https://www.degruyter.com/downloadpdf/j/hfsg.2004.58.issue-4/hf.2004.052.pdf>
1481 <https://www.degruyter.com/view/j/hfsg.2004.58.issue-4/hf.2004.052.xml>
1482 <https://www.degruyter.com/view/j/hfsg.2004.58.issue-4/hf.2004.052.xml>
- 1483
- 1484 Stefan Willför, Jarl Hemming, and Ann-Sofie Leppanen. Analysis of extractives in
1485 different pulps – method development, evaluation, and recommendations. *Analytical
1486 tools with applications for wood and pulping chemistry*, pages 1–17, 2006. URL www.cost.eu/domains_actions/fps/Actions/E41. printed.
- 1487
- 1488 Stefan M Willför, Markku O Ahotupa, Jarl E Hemming, Markku H T Reuna-
1489 nen, Patrik C Eklund, Rainer E. Sjöholm, Christer S E Eckerman, Suvi P
1490 Pohjamo, and Bjarne R Holmbom. Antioxidant activity of knotwood extrac-
1491 tives and phenolic compounds of selected tree species. *Journal of Agricultural
1492 and Food Chemistry*, 51:7600–7606, 12 2003c. ISSN 0021-8561. doi: 10.1021/
1493 jf030445h. URL <http://pubs.acs.org/doi/pdf/10.1021/jf030445h> http://pubs.acs.org/doi/abs/10.1021/jf030445h. antioxidant props - table 2
- picea
1494 abies extract stronger than isolated oligolignans and this is stronger than larix decidua
1495 extract at same dosage –gt; 2quantitites for the radical scavenging essays.
- 1496
- 1497 W. Windig, C. E. Heckler, F. A. Agblevor, and R. J. Evans. Self-modeling mixture anal-
1498 ysis of categorized pyrolysis mass spectral data with the simplisma approach. *Chem-
1499 metrics and Intelligent Laboratory Systems*, 14:195–207, 4 1992. ISSN 01697439. doi:
1500 10.1016/0169-7439(92)80104-C.
- 1501 Willem Windig and Jean Guilment. Interactive self-modeling mixture analysis. *Analyt-
1502 ical Chemistry*, 63:1425–1432, 1991. ISSN 15206882. doi: 10.1021/ac00014a016.
- 1503 Sanford M. Zeller. Humidity in relation to moisture imbibition by wood and to spore
1504 germination on wood author. *Annals of the Missouri Botanical Garden*, 7:51–74, 1920.
- 1505 Janja Zule, Katarina Čufar, and Vesna Tišler. Lipophilic extractives in heartwood of eu-
1506 ropean larch (larix decidua mill.). *Drvna industrija*, 66:305–313, 2015. ISSN 00126772.
1507 doi: 10.5552/drind.2015.1442. URL <http://hrcak.srce.hr/file/221885>.

- 1508 Janja Zule, Katarina Čufar, and Vesna Tišler. Hydrophilic extractives in heart-
1509 wood of european larch (*larix decidua* mill.). *Drvna industrija*, 67:363–370, 2017.
1510 ISSN 00126772. doi: 10.5552/drind.2016.1618. URL <http://hrcak.srce.hr/file/221885><http://hrcak.srce.hr/file/254614><https://hrcak.srce.hr/172565>.
- 1512 Åsmund Rinnan, Lars Nørgaard, Frans van den Berg, Jonas Thygesen, Rasmus Bro,
1513 and Søren Balling Engelsen. Data pre-processing, 2009.
- 1514 Folke Örså and Bjarne R Holmbom. A convenient method for the determination of
1515 wood extractives in papermaking process waters and effluents. *Journal of Pulp and*
1516 *Paper Science*, 20:361–366, 1994.
- 1517 smund Rinnan, Frans van den Berg, and Søren Balling Engelsen. Review of the
1518 most common pre-processing techniques for near-infrared spectra. *TrAC Trends*
1519 *in Analytical Chemistry*, 28:1201–1222, 11 2009. ISSN 01659936. doi: 10.
1520 1016/j.trac.2009.07.007. URL <https://linkinghub.elsevier.com/retrieve/pii/S0165993609001629>.

₁₅₂₂ **Appendix A**

₁₅₂₃ **PAPER I**

₁₅₂₄ Appendix B

₁₅₂₅ PAPER II

₁₅₂₆ Appendix C

₁₅₂₇ PAPER III

in figure 4, 1528
one peak of
the lignin
BP has
the wrong
wavenumber
- it should be
929, not 829

₁₅₂₉ **Appendix D**

₁₅₃₀ **PAPER IV**

₁₅₃₁ Appendix E

₁₅₃₂ PAPER V

¹⁵³³ Appendix F

¹⁵³⁴ Unpublished Extractives paper