

Adrian Antonsen

# Foundation for massive-scale analysis of fermented beverages

Master's thesis in Biotechnology

Supervisor: Eivind Almaas

Co-supervisor: Christian Schulz, Leesa J. Klau, Emil Karlsen

June 2021



Adrian Antonsen

# **Foundation for massive-scale analysis of fermented beverages**

Master's thesis in Biotechnology

Supervisor: Eivind Almaas

Co-supervisor: Christian Schulz, Leesa J. Klau, Emil Karlsen

June 2021

Norwegian University of Science and Technology

Faculty of Natural Sciences

Department of Biotechnology and Food Science



Norwegian University of  
Science and Technology



# Acknowledgements

As the completion of this thesis marks the end of an intensive and challenging two-year project, I would like to extend my gratitude to the people who have helped and guided me along the way. Firstly, I am grateful and appreciative of the opportunity to write my masters thesis in the Almaas Lab research group at NTNU Trondheim, and to be given the opportunity to work on a project I was allowed to shape myself, on a subject I am truly passionate about. I would like to express my gratitude to my supervisor Professor Eivind Almaas both for providing unwavering support and much needed guidance, and for believing in this project and allowing me to work on tenuous ideas others might have found silly.

Sincere gratitude goes to my co-supervisors for their support throughout this project: To Christian Schulz for introducing me to the strange world of beer research, and for convincing me to choose a subject based on real interest and not on some idea of what was expected of me, as well as for his generous support in disposing of excess samples from my research. To Emil Karlsen for academic and moral support, as well as for making sure I never took myself too seriously. And to Leesa J. Klau for shepherding me through the highly confusing field of NMR spectroscopy and multivariate statistical analysis with almost uncanny patience for my naive mistakes.

I also wish to extend my gratitude to the whole Almaas Lab research group for their support and guidance through this foray into the scientific community. I also wish to thank Professor Finn L. Aachmann, Gaston Courtade, and Trygve Andreassen for their contributions to the tools for NMR analysis used here, as well as Morten J. Dille and Trond Viggo Pettersen for their guidance in the field of brewing and their help with developing recipes for our studies.

Finally, I would like to thank my closest friends who have helped me keep both my sanity and my spirit through a challenging time of social restrictions, and my girlfriend Inger L. Lyngstad for her unending support, patience and work ethic that has kept me motivated in the face of adversity.

# Abstract

Brewing and consumption of beer is central to many cultures and is one of the oldest and most enduring forms of biotechnology. The commercial market for beer is large, and the available variety of styles and types is vast and constantly expanding. However, much of our understanding of beer and the categories thereof is based on tradition and artistic expression rather than a scientific foundation. The Systems Brewology Project in AlmaasLab at NTNU Trondheim aims to expand our scientific understanding of beer categorization, characterization, production process, and sensory experience. With the use of *data mining* and *machine learning* tools for analysis of a large database of chemical composition data on annotated beer samples, we aim to gain new insights into the ancient art of brewing.

The aim of this thesis has been to lay the foundation for creating a database capable of facilitating *massive-scale data analysis*. The efforts used in laying this foundation can be split into two categories: practical and theoretical. The practical efforts consisted of generating sample data to examine the viability of chemical analysis, data exploration by *multivariate statistical analysis*, and creating documentation for sample preparation and *one-dimensional proton nuclear magnetic resonance* (1D  $^1\text{H}$  NMR) acquisition. The theoretical efforts were focused on creating a storage structure for the coming database and evaluating options for database platforms, as well as the development of a platform for *citizen-science* by creating a website for contact with homebrewers.

The practical efforts aimed to explore the capabilities and limitations of *fingerprinting* by 1D  $^1\text{H}$  NMR in beer samples. For this, the following three case studies were completed: (1) study of traditional categorization of samples from a commercial brewery on the basis of chemical composition; (2) study of the effects of different *pitching rates* on taste and chemical spectra; (3) exploring the chemical changes caused by aging in beers flavoured with algae. Samples for these studies were produced in-house, as well as gathered through industry cooperation. These were analyzed using 1D  $^1\text{H}$  NMR spectroscopy. The resulting spectra were processed and analyzed using a variety of multivariate statistics, including *Principal Component Analysis* (PCA), *Partial Least-Squares Discriminant Analysis* (PLS-DA), *t-Distributed Stochastic Neighbor Embedding* (t-SNE), *hierarchical clustering*, *k-means clustering*, and *compound quantification*. The sample treatment process was standardized and compiled into a comprehensive manual for sampling, preparation and NMR acquisition.

The theoretical efforts aimed firstly at creating a storage structure for the database. To this end, a deeper understanding of the hierarchical structure used for categorization of beer types were required to sort and store a large number of beer samples. In order to document this structure and create a template for a database, an *ontology* of beer style categories was created. The ontology contains descriptions of most known beer styles, as well as records of the samples created in the practical efforts. The second aim of the theoretical efforts was to evaluate different database system solutions based on the characteristics of data

---

from the practical efforts.

The work done during this thesis has demonstrated that a streamlined method of 1D  $^1\text{H}$  NMR spectroscopy can be used to quickly analyze multiple samples of beer to produce informative results. It also demonstrates how statistical analysis tools for multivariate data analysis can be used to examine *high-dimensional* data produced by NMR spectroscopy of beer samples. Findings from case studies of beer samples did not detect statistically significant differences between conventional categories of beer, however, differences due to pitching rate and aging were detected. The approach using fingerprinting by multivariate analysis on 1D  $^1\text{H}$  NMR spectra showed promise for quick exploration of sample sets. Limitations in accuracy of 1D  $^1\text{H}$  NMR were encountered for complex solutions like beer, and suggestions for future improvements are detailed herein. Based on lessons from the data produced in the practical efforts, the database solution *openBIS* was chosen as the optimal database system framework for building a repository. The comprehensive ontology provided can be used to inform the storage structure of a database, as well as function as a controlled language for sample annotation.

Overall, this work presents an NMR-based *omics-level* analysis pipeline for beer. This includes an investigation of different central control parameters and a selection of statistical analyses for the resulting high-dimensional data. It also presents the groundwork for a comprehensive database, including an ontology and choice of an appropriate software for implementation of future compilation and analysis of data from beer, using both the above-mentioned and other new methods for data collection and analysis.

# Sammendrag

Brygging og konsum av øl er sentralt i svært mange kulturer, og er en av de eldste vedvarende formene for bioteknologi vi har. Det kommersielle markedet for øl er stort, med enormt mange ulike typer og stiler allerede i produksjon, og en hurtig utvikling på feltet. Likevel er mye av vår forståelse av produksjon og kategorisering av øl basert mer på tradisjon og håndverk enn på vitenskapelig grunnlag. Systems Brewology-prosjektet ved AlmaasLab, NTNU Trondheim har som mål å utvikle vår vitenskapelige kunnskap om kategorisering, vurdering, produksjon og sensoriske opplevelser av øl. Gjennom bruk av verktøy for *datautvinning* og *maskinlæring* for analyse av databaser med informasjon om kjemisk sammensetting av øl, håper vi å kunne utvinne ny kunnskap om bryggekunsten.

Målet for denne avhandlingen har vært å legge grunnlaget for å etablere en database som kan fasilitere storskala dataanalyse av øl. Arbeidet med å legge dette grunnlaget kan deles i to kategorier: praktiske og teoretiske. Det praktiske arbeidet har bestått av å produsere data for å undersøke mulighetene for bruk av kjemiske analysemetoder, evaluering av statistiske metoder for *multivariatanalyse*, og produksjon av protokoller for prøvebehandling og henting av *one-dimensional proton nuclear magnetic resonance* (1D  $^1\text{H}$  NMR). Det teoretiske arbeidet gikk ut på å etablere en struktur for den kommende databasen, og å evaluere potensielle databaseplattformer. En plattform for *citizen-science* ble også etablert ved å utvikle en nettside for kontakt mellom hjemmebryggere og Systems Brewology prosjektet.

Det praktiske arbeidet hadde som hensikt å utforske muligheter og begrensninger ved *fingerprinting* av prøver fra øl med 1D  $^1\text{H}$  NMR. For å oppnå dette ble de følgende tre studiene gjennomført: (1) Evaluering av tradisjonelle kategoriseringer av øl fra et kommersielt bryggeri på basis av kjemisk sammensetning; (2) undersøkelse av de sensoriske og kjemiske effektene ved ulike gjærmengder under fermentering; (3) studie av kjemiske endringer forårsaket av aldring i øl smaksatt med alger. Prøver for disse studiene ble produsert i våre laboratorier, eller innhentet gjennom samarbeid med industri. Alle prøvene ble analysert ved bruk av 1D  $^1\text{H}$  NMR-spektroskopi. Spektra fra prøvene ble analysert ved bruk av ulike metoder for multivariatanalyse, slik som *Principal Component Analysis* (PCA), *Partial Least-Squares Discriminant Analysis* (PLS-DA), *t-Distributed Stochastic Neighbor Embedding* (t-SNE), *hierarchical clustering*, *k-means clustering*, og *kvantifisering av forbindelser*. Prosessen for prøvebehandling ble standardisert og samlet til en detaljert manual for prøvetaking, behandling, og innhenting av spektra ved NMR.

Det teoretiske arbeidet siktet for det første på å etablere en lagringsstruktur for den kommende databasen. For å gjøre dette var det nødvendig å utvikle en dypere forståelse for den hierarkiske strukturen brukt for kategorisering av ulike øltyper, ettersom denne strukturen vil være sentral for hvordan prøver skal kategoriseres og lagres. En *ontologi* for kategorier og beskrivelser av øl ble derfor utviklet, noe som vil bidra til forståelse og dokumentering av strukturen, samt fungere som et grunnlag for en database. Det andre målet



---

for det teoretiske arbeidet var å undersøke ulike tilgjengelige databaseløsninger, og vurdere dem i forhold til nødvendige egenskaper basert på karakteristikken til dataen funnet under det praktiske arbeidet.

Arbeidet gjort i denne avhandlingen har vist at en strømlinjet metode for 1D  $^1\text{H}$  NMR-spektroskopi kan bli brukt til å raskt og effektivt analysere flere ulike øl og fremstille informative resultat. Det blir også vist hvordan ulike statistiske metoder kan bli brukt for å vurdere *høydimensjonell* data produsert fra NMR-spektroskopi av øl. Studiene gjort her kunne ikke avdekke statistisk signifikante forskjeller mellom konvensjonelle kategorier av øl, men forskjeller på basis av gjærmengde og aldring ble oppdaget. Bruk av fingerprinting ved multivariatanalyse av 1D  $^1\text{H}$  NMR-spektra gav lovende resultat for hurtig vurdering av prøver, men visse begrensinger ved bruk av 1D  $^1\text{H}$  NMR for komplekse medium, slik som øl, ble oppdaget. Forslag til mulige forbedringer beskrives i denne avhandlingen. Basert på kunnskap fra strukturen av dataen produsert under det praktiske arbeidet ble databasesystemet *openBIS* valgt som den optimale løsningen for etablering av en database for ølprøver. Ontologien som ble fremstilt vil kunne fungere som mal for databasestrukturen, og som et kontrollert vokabular ved annotering av prøver.

Totalt sett presenterer dette arbeidet en NMR-basert *omics-level* analysemetode for øl. Dette inkluderer undersøkelser av sentrale kontroll-parametere og valg av metoder for statistisk analyse av de resulterende høydimensjonelle dataene. Grunnarbeidet for en omfattende databaseløsning legges også frem, bestående av en ontologi samt valg av passende datasystemer for fremtidig ansamling og analyse av data fra øl, både ved nevnte og nye potensielle metoder for innhenting og analyse av data.

# Table of Contents

<b>List of Figures</b>	<b>xi</b>
<b>List of Tables</b>	<b>xv</b>
<b>Acronyms</b>	<b>xvi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Importance of beer . . . . .	1
1.2 Aim of the project . . . . .	2
1.2.1 Platform development . . . . .	3
1.2.2 Aim of the thesis . . . . .	3
<b>2 Background</b>	<b>5</b>
2.1 A brief history of beer . . . . .	5
2.1.1 Classification of beer . . . . .	6
2.1.2 Beer styles . . . . .	7
2.2 The brewing process . . . . .	7
2.2.1 Malt and hops . . . . .	8
2.2.2 Mashing . . . . .	9
2.2.3 Boiling . . . . .	9
2.2.4 Fermentation and yeast . . . . .	9
2.2.5 Storage and bottle fermentation . . . . .	10
2.3 Nuclear magnetic resonance spectroscopy . . . . .	11
2.3.1 Nuclear spin . . . . .	11
2.3.2 Resonance phenomenon . . . . .	11
2.4 Treatment and chemical analysis of beer by NMR spectroscopy . . . . .	12

---

2.4.1	Use of NMR spectroscopy for beer analysis . . . . .	12
2.4.2	Sample treatment . . . . .	13
2.4.3	Data treatment . . . . .	13
2.5	Statistical data analysis . . . . .	14
2.5.1	T-tests and ANOVA . . . . .	14
2.5.2	Principal Component Analysis . . . . .	14
2.5.3	Clustering and supervised analysis . . . . .	15
2.5.4	Fingerprinting . . . . .	16
2.5.5	Machine-guided analysis methods . . . . .	16
2.6	Ontology . . . . .	16
2.6.1	Defining ontologies . . . . .	17
2.6.2	Purpose of ontologies . . . . .	17
2.6.3	Ontology concepts . . . . .	18
2.7	Databases . . . . .	18
2.7.1	Fundamentals of databases . . . . .	18
2.7.2	Structured Query Language . . . . .	19
2.7.3	Application programming interface . . . . .	19
2.7.4	Challenges in beer studies . . . . .	19
<b>3</b>	<b>Materials and Methods</b>	<b>20</b>
3.1	Case studies . . . . .	20
3.1.1	Brewing procedure . . . . .	20
3.1.2	Categorization study of Austmann samples . . . . .	21
3.1.3	Study of the effects of differing pitching rates on beer composition . . . . .	21
3.1.4	Study of aging in algae-flavoured beer samples . . . . .	23
3.2	Sample preparation, NMR spectroscopy, and statistical analyses . . . . .	25
3.2.1	Pipeline overview . . . . .	25
3.2.2	Sample preparation . . . . .	25
3.2.3	NMR acquisition . . . . .	26
3.2.4	Spectrum processing . . . . .	27
3.3	Statistical analysis . . . . .	28
3.4	Choice of database solution . . . . .	29

---

3.5	Ontology creation . . . . .	29
3.6	Project website . . . . .	29
<b>4</b>	<b>Results and Analysis</b>	<b>30</b>
4.1	Case studies of beer samples . . . . .	30
4.1.1	Description and overview . . . . .	30
4.1.2	Data treatment and scaling . . . . .	31
4.1.3	Compound identification . . . . .	31
4.1.4	Categorization study of Austmann beers . . . . .	33
4.1.5	Pitch rate study . . . . .	42
4.1.6	Aging study of algae-flavoured beer . . . . .	48
4.2	Ontology . . . . .	54
4.2.1	General design . . . . .	54
4.2.2	Parts of the ontology . . . . .	55
4.2.3	Public access . . . . .	58
4.3	Database framework . . . . .	59
4.4	Outreach . . . . .	59
<b>5</b>	<b>Discussion</b>	<b>61</b>
5.1	Case studies of beer samples . . . . .	61
5.1.1	Categorization study of Austmann samples . . . . .	61
5.1.2	Pitch rate study . . . . .	62
5.1.3	Aging study in algae-flavoured beer . . . . .	63
5.1.4	Evaluation of the analysis pipeline as a whole in light of the three case studies explored . . . . .	64
5.2	Data treatment . . . . .	65
5.2.1	Challenges and decisions on scaling of chemical spectra . . . . .	65
5.3	Lessons from the case studies for data storage . . . . .	65
5.3.1	Metadata . . . . .	66
5.3.2	Analysis data . . . . .	66
5.3.3	Results data . . . . .	67
5.4	Outreach and contributions . . . . .	67
5.5	Database selection . . . . .	68

---

5.5.1	Alternatives . . . . .	68
5.6	Ontology . . . . .	71
5.6.1	Accuracy of classifications . . . . .	71
5.7	Summary of recommendations for future efforts . . . . .	72
<b>6</b>	<b>Outlook and summary</b>	<b>74</b>
	<b>Bibliography</b>	<b>77</b>
	<b>Appendix</b>	<b>81</b>
<b>A</b>	<b>Recipes for brewing experiments</b>	<b>82</b>
A.1	Pitch rate samples base recipe . . . . .	83
A.2	Alage-beer samples base recipes . . . . .	84
A	Brown Ale recipe . . . . .	84
B	IPA recipe . . . . .	85
C	Goose recipe . . . . .	86
D	Amber Lager recipe . . . . .	87
<b>B</b>	<b>Matlab scripts for alignment, normalization and binning</b>	<b>88</b>
B.1	Alignment steps . . . . .	88
B.2	Normalization script . . . . .	91
B.3	Binning script . . . . .	95
<b>C</b>	<b>R-scripts for statistical tests and plotting</b>	<b>98</b>
C.1	R-script for determining optimal clusters for K-means . . . . .	98
C.2	R-script for plotting and anova test of ethanol content . . . . .	98
<b>D</b>	<b>Python script for transforming Bruker files into Matlab files</b>	<b>100</b>
<b>E</b>	<b>Manual for sample treatment and NMR acquisition</b>	<b>104</b>
<b>F</b>	<b>Supplementary plots for statistical results</b>	<b>115</b>
F.1	Validation plots for PLS-DA analysis . . . . .	115
A	Austmann samples . . . . .	115
B	Pitch rate samples . . . . .	117

---

---

F.2	Detailed spectrum and loading plots . . . . .	118
A	Austmann samples loading plots . . . . .	118
B	Pitch rate samples loading plots . . . . .	123
C	Algae-beer samples loading plots . . . . .	128

# List of Figures

1.1	<b>Visual overview of the scope of the thesis and aims of the overarching project.</b> Purple nodes represent sources, blue nodes represent processes or general concepts, light blue nodes represent the direct products of this thesis. . . .	4
2.1	<b>Visual representation of a simplified brewing process.</b> . . . . .	8
3.1	<b>Experimental setup of pitch rate study.</b> The figure illustrates one of three parallels in the study. . . . .	22
3.2	<b>Algae beer experiment setup of one combination of wort and yeast.</b>	24
3.3	<b>Sample preparation procedure.</b> . . . . .	26
4.1	<b>Full spectrum view of a representative raw spectrum with all identified compounds annotated.</b> Identification supplied by Leesa J. Klau. . . .	32
4.2	<b>Scores plot of 1st and 2nd principal components from PCA of initial categorization analysis of samples from Austmann.</b> Each point represents a sample, and the categories are denoted by shape and color, as seen in the legend in the top right. Explained variances of each component displayed in brackets next to the principal component. . . . .	34
4.3	<b>Scores plot of 1st and 2nd components from PLS-DA of initial categorization analysis of samples from Austmann.</b> Each point represents a sample, and the categories are denoted by shape and color, as seen in the legend in the top right. Explained variances of each component displayed in brackets next to the principal component. . . . .	35
4.4	<b>Variable Importance in Projection (VIP) scores for PLS-DA analysis of initial categorization test of samples from Austmann.</b> X-axis represents VIP score, Y-axis displays the ppm of the bin, or feature. . . . .	36
4.5	<b>Clustering analysis of Austmann samples by t-SNE.</b> Data points are colored after conventional categorizations. Perplexity = 5. . . . .	37
4.6	<b>Clusters containing the different samples.</b> Original categorization is displayed in color code: Sour: light blue, dark: red, IPA: green, light: dark blue. .	38
4.7	<b>Clustering analysis of re-clustered Austmann samples by t-SNE.</b> Data points are colored according to the new clusters found by k-means. Perplexity = 4. . . . .	39

---

4.8	<b>Scores plot of 1st and 2nd components from PLS-DA of re-clustered categorization analysis of samples from Austmann.</b> Each point represents a sample, and the categories are denoted by shape and color, as seen in the legend in the top right. Explained variances of each component displayed in brackets next to the principal component. . . . .	39
4.9	<b>Important features for PLS-DA analysis of re-clustered categorization test of samples from Austmann.</b> X-axis represents importance score, Y-axis displays the ppm of the bin, or feature. . . . .	40
4.10	<b>Full spectra and loadings for reclustered Austmann samples.</b> Upper field shows a plot of all binned NMR spectra for the reclustered Austmann samples overlaid and colored by cluster (cluster 1: red, cluster 2: light blue, cluster 3: green, cluster 4: deep blue). Some identifiable compounds are annotated. Lower field shows PLS-DA loadings for the first PC, taller peaks indicate more weight in PLS-DA. X-axis: PPM, Y-axis: intensity of signal / weight in PLS-DA. . . . .	41
4.11	<b>Illustration of trends recorded for sensory analysis of pitching study.</b> Direction of trends are illustrated by arrows, indicating that the majority of participants described the intensity of the given characteristic for each category in the direction of the arrow. E.g: Overpitched samples were generally described as less sour than underpitched samples. . . . .	42
4.12	<b>PCA scores plot of 1st and 3rd principal components of differential analysis of samples from pitch rate study.</b> Each point represents a sample, and the categories are denoted by shape and color, as seen in the legend in the top right. Explained variances of each component displayed in brackets next to the principal component. . . . .	43
4.13	<b>Clustering analysis of pitch rate samples by t-SNE.</b> Data points are colored according pitching rate used. Perplexity = 4. . . . .	44
4.14	<b>Full spectra and loadings for pitch rate samples.</b> Upper field shows a plot of all binned NMR spectra for all pitch rate samples overlaid and colored by pitch rate (underpitched: blue, normal pitch: green, overpitched: red). Some identifiable compounds are annotated. Lower field shows PCA loadings for the first PC, more extreme peaks indicate more weight in the separation found in PCA. X-axis: PPM, Y-axis: intensity of signal / weight in PCA. . . . .	45
4.15	<b>Dendrogram showing clustering of samples from pitch rate study using Euclidean distance and Ward clustering algorithm.</b> . . . . .	46
4.16	<b>Boxplot of ethanol concentration in the three different pitching rates.</b> . . . . .	47
4.17	<b>Summary of ANOVA of differences in ethanol content between pitching rates.</b> . . . . .	47
4.18	<b>Paired t-test of difference in two age categories in the full algae-beer data set.</b> Unit variance scaling was performed prior to t-test. . . . .	48



---

4.19	<b>Overview of the first five principal components of the PCA and PLS-DA analyses of all algae-beer samples.</b> Each point represents a sample, red triangles represent non-aged samples, green crosses represent aged samples. Percentage of variance explained by each component is displayed below the component. . . . .	49
4.20	<b>Clustering analysis of entire sample set from aging study by t-SNE.</b> Data points are colored according to aging category. Perplexity = 8. . . . .	49
4.21	<b>PCA plot of 1st and 2nd component of samples from the Brown Ale group in the aging study.</b> . . . . .	50
4.22	<b>Clustering analysis of Brown Ale samples from aging study by t-SNE.</b> Data points are colored according to aging category. Perplexity = 6. . . . .	51
4.23	<b>T-test of difference in aged and non-aged groups for each feature in brown ale samples from the aging study.</b> Unit variance scaling performed on data set prior to t-test. . . . .	51
4.24	<b>Dendrogram showing clustering of samples from the brown ale group of the aging study.</b> Clustering using Euclidean distance and Ward clustering algorithm. . . . .	52
4.25	<b>Full spectra and loadings for algae flavoured samples.</b> Upper field shows a plot of all binned NMR spectra for all algae brown ale samples overlaid and colored by aging category (non-aged: blue, aged: red). Some identifiable compounds are annotated. Lower field shows PCA loadings for the first PC, more extreme peaks indicate more weight in the separation found in PCA. X-axis: PPM, Y-axis: intensity of signal / weight in PCA. . . . .	53
4.26	<b>Example of partial ontology structure for samples in the pitch rate study in network view for illustrative purposes.</b> Full lines represent parent/child relationships, dashed lines represent custom object properties. Figure created in Protege. . . . .	54
4.27	<b>Example of a few sub classes of the ale category in the ontology structure.</b> Figure created in Protege. . . . .	55
4.28	<b>Example of partial ontology class structure for the ale category.</b> Full lines represent parent/child relationships, dashed lines represent custom object properties. Figure created in Protege. . . . .	56
4.29	<b>Example of partial ontology structure for the ingredient category, including connections between an ingredient instance and a sample instance.</b> Full lines represent parent/child relationships, dashed lines represent custom object properties. Figure created in Protege. . . . .	57
4.30	<b>List of custom object properties used in the ontology.</b> Figure created in Protege. . . . .	58
4.31	<b>Logo for openBIS.</b> . . . . .	59
4.32	<b>Illustration of the Systems Brewology website.</b> . . . . .	60

---

F.1	Permutation test statistic for PLS-DA analysis of initial categorization test of samples from Austmann. . . . .	115
F.2	Permutation test statistic for PLS-DA analysis of re-clustered categorization test of samples from Austmann. . . . .	116
F.3	Permutation test statistic for PLS-DA analysis of pitch rate study samples.	117
F.4	Detailed spectra and loadings in 0.8-1.8 ppm region for Austmann samples.	118
F.5	Detailed spectra and loadings in 1.8-3.5 ppm region for Austmann samples.	119
F.6	Detailed spectra and loadings in 5.8-7.1 ppm region for Austmann samples.	120
F.7	Detailed spectra and loadings in 6.9-8.2 ppm region for Austmann samples.	121
F.8	Detailed spectra and loadings in 7.8-9 ppm region for Austmann samples. .	122
F.9	Detailed spectra and loadings in 0.8-1.8 ppm region for pitch rate samples. .	123
F.10	Detailed spectra and loadings in 1.8-3.5 ppm region for pitch rate samples. .	124
F.11	Detailed spectra and loadings in 5.8-7 ppm region for pitch rate samples. . .	125
F.12	Detailed spectra and loadings in 7-8 ppm region for pitch rate samples. . . .	126
F.13	Detailed spectra and loadings in 7.8-9 ppm region for pitch rate samples. . .	127
F.14	Detailed spectra and loadings in 0.8-1.8 ppm region for algae-beer samples.	128
F.15	Detailed spectra and loadings in 1.8-3.4 ppm region for algae-beer samples.	129
F.16	Detailed spectra and loadings in 5.8-7 ppm region for algae-beer samples. .	130
F.17	Detailed spectra and loadings in 7-8 ppm region for algae-beer samples. . .	131
F.18	Detailed spectra and loadings in 7.8-9 ppm region for algae-beer samples. .	132

# List of Tables

3.1	<b>List of sampled beers from Austmann brewery.</b> Including name, sample ID and categorization. . . . .	21
3.2	<b>Table of pitching rates for samples in pitch rate study</b> . . . . .	23
3.3	<b>List of samples produced for the study of aging in algae flavoured beer.</b> Including sample ID, beer type, yeast, algae type, brewing date, and sampling date. . . . .	24
4.1	<b>List of identified compounds and their location in the spectra.</b> Supplied by Leesa J. Klau. . . . .	32
4.2	<b>Table of original style categories of Austmann samples and number of samples in each category.</b> . . . . .	33
4.3	<b>Table of samples from Austmann.</b> Including new clusters assigned according to chemical spectra. . . . .	38

# Acronyms

- HDL** *High-Density Lipoprotein.* 2
- ROS** *Reactive Oxygen Species.* 10
- NMR** *Nuclear Magnetic Resonance.* 11
- FID** *Free Induction Decay.* 11
- MVDA** *Multivariate Data Analysis.* 12
- PCA** *Principal Component Analysis.* 12
- PLS-DA** *Partial Least Squares-Discriminatory Analysis.* 12, 13
- MS** *Mass Spectrometry.* 13
- TSP** *Trimethylsilylpropanoic Acid.* 13
- t-test** *Student t-test.* 14
- ANOVA** *Analysis of Variance.* 14
- t-SNE** *t-Distributed Stochastic Neighbor Embedding.* 15
- DBMS** *Database Management System.* 18
- SQL** *Structured Query Language.* 19
- API** *Application Programming Interface.* 19
- SL** *Saccharina latissima.* 24
- AE** *Alaria esculenta.* 24
- PQN** *Probabilistic Quotient Normalization.* 28
- ppm** *Parts per Million.* 31
- VIP** *Variable Importance in Projection.* 35
- EBC** *European Brewery Convention.* 58

---

**IBU** *International Bitterness Units.* 58

**HPLC** *High Performance Liquid Chromatography.* 64

**GC-MS** *Gas Chromatography-Mass Spectroscopy.* 64

# Chapter 1

## Introduction

This chapter will introduce the aims of this thesis and the larger project around it, as well as the motivation and reasoning behind the efforts made therein.

### 1.1 Importance of beer

The brewing of fermented beverages might be the oldest example of biotechnology we have. Simple beer-like drinks have been discovered as far back as 5 000 BC [1], and here in Norway beer has been an integral and spiritual part of our tradition ever since the Viking age. The consumption of beer is a significant part of the diet for many of us, according to statistics from *ourworldindata.org* [2] it represents a 45% share of our total alcohol consumption in Norway, with annual consumption per capita during 2014 at 52 L. The numbers for the US, UK and Germany were even higher with 76L (50%), 83L (37%) and 117L (54%) respectively. Although there the global brewing industry is quite large and there is a blooming interest in smaller craft beer focused breweries, there still is a pervading non-scientific approach when it comes to the brewing process and the categorization of beer. The brewing process is often viewed more as an art than a science, with a tendency towards *know-how* rather than *know-why*, possibly expressed best in two quotes accredited to Fritz Maytag, an important figure in the history of modern craft beer:

*"We brewers don't make beer, we just get all the ingredients together and the beer makes itself."*

*"Beer does not make itself properly by itself. It takes an element of mystery and of things no one can understand."*

— Fritz Maytag

The *Systems Brewology Project* aims to contribute to the unveiling of these proverbial mysteries and gain a deeper and more scientific understanding of the categorization of different beer types, the relations between them, and the processes involved in their creation.

With the aim of better understanding the minutiae in the difference between beer styles, brewing methods, treatments, and storage, we believe *massive-scale* chemical and compositional analysis will be a valuable tool. This means examining a large variety of beers with detailed methods of compositional analysis, creating a comprehensive library of samples representative of all the categories and types of beer available. To this end, a platform

---

is needed to facilitate the storage and analysis of detailed information on beer brewed in controlled laboratory experiments, samples gathered from breweries and home brewers, as well as samples from commercially available products. We therefore aim to construct a database for storing detailed information on every facet of a sample, including the recipe, brewing process, treatment, and other factors affecting a sample.

To be able to discover trends, patterns, and emergent properties from analysis of these samples, it is important to have the capability to store and export all relevant information about the samples and the chemical analysis performed on them. We therefore need to construct or tailor a database structure to our specific needs where we can easily enter and export all pertinent information for a given data analysis we wish to perform.

In order to set up such a data repository, we must understand the nature of the data to be stored there, the structure of sample categorization, as well as the input and export criteria for storing and extracting data. In addition to this, we also need to establish simple, universally viable sample preparation, analysis, and annotation protocols so that all relevant samples can be treated in a consistent manner for the lifetime of the project. There is also a need to establish a way of receiving samples from external sources and contributors, as generating the necessary sample volume for massive-scale analysis by in-house brewing is not feasible. There is a need for a solid foundation in order to implement a database that meets all the requirements, as well as having procedures and channels in place for generating and analyzing the data.

## 1.2 Aim of the project

There are multiple eventual goals for the Systems Brewology Project, a deeper understanding of beer and the brewing process is just a part of what we aim to accomplish. We believe that understanding beer making and composition can be a stepping stone to new discoveries in quite disparate fields like yeast and microbial fermentation, health, and consumer choice. Large-scale analysis of samples with many different strains of yeast and other microorganisms under different conditions and subjected to a vast array of different metabolites, coupled with detailed chemical analysis of the resulting product, could help us understand and discover new fermentation paths and products. Such discoveries could guide deeper research on pathway modeling and microbiological experiments on microorganisms that are used every day. This could lead to benefits not only for basic science, but also for society and industry.

Research into the production and composition of beer can also lead to a better understanding of the health effects of a product that is both used and abused worldwide on a daily basis. Beer is, as mentioned above, a large part of many people's diets, and this can have both beneficial and adverse effects on health. Multiple studies ([3], [4]) have shown beneficial effects of moderate beer consumption on cancer development, cardiovascular disease, immune system stimulation, and reduced risk of dementia. One example of such an effect is the impact on the risk of cardiovascular disease with elevated levels of *High-Density Lipoprotein* (HDL) and *apolipoprotein A1*, both of which are associated with a lowered risk of *atherosclerosis*. However, we also know that excessive consumption can have severe detrimental effects on aspects of our health, such as increased risk of obesity, allergy induction, cancer development and social misbehaviour, to name a few. We aim not only to understand beer production, but also to inform on it based on our research, and with a combination of information on the chemical components in beer, as well as

---

their effects on human health, we may well be able to inform on the production of beer and other fermented products with greater health benefits and less detrimental effects.

The industry around beer is substantial, with an estimated worth of US\$ 187.89 billion, and is predicted to rise steadily in the years to come [5]. A large part of the total beer market consists of the pale lager / pilsner category, with more specialized beer making up smaller portions. However, not everyone prefers pale lager beer, and the choice of what to produce may be based largely on guesswork or on fluctuating economic markets where beer preferences might not be fairly represented [6]. By combining data on the composition of different beers with sensory analysis data, as well as expert and consumer preference data, we believe it will be possible to discover certain customer group preferences for certain taste components or profiles, and thereby be able to create and tailor new products that might appeal to a certain category of customers. If such a method is achievable, it might be immensely valuable to the brewing industry, as well as to guide new discoveries in psychology and consumer preference.

### 1.2.1 Platform development

The study of beer is important, but there is another major benefit of the work in this project: The development of the research platform itself. The methodology we are developing in order to enable massive-scale analysis of beer can, with only minor changes, be adapted to enable the analysis of almost any other beverage or food. The process would hardly have to change if one were to use it for research on liquids like coffee or soda, and could be adopted to use on solid foods with small changes in the laboratory procedure for chemical analysis. There are of course specific attributes of the chemical analysis and customised data repositories that would have to change in order to apply the protocols to other materials, but the methodology and ideas would remain the same. This could increase the possible applications of our work many-fold, and pave the way for many new projects in food science and related disciplines.

### 1.2.2 Aim of the thesis

As explained above, there is a multitude of reasons both to study beer and for the development of the research platform. **The aim of this thesis** is therefore to lay the groundwork to facilitate massive-scale computer analysis of fermented beverages, as well as providing a structure for categorization and storage of the generated data. This will involve a combination of the following efforts:

- Documentation and standardization of sampling and analysis protocols in a user manual.
- Case studies analyzing different sets of samples as pilot studies for evaluation of the analysis pipeline.
- Creation of a custom ontology for categorization of beer samples.
- Evaluation of requirements for a database structure based on experience from case studies.
- Recommendations on specific database solutions based on the requirements found.





# Chapter 2

## Background

This section will present the theoretical background necessary to understand the work performed for this thesis. This will include brief explanations of central themes in brewing, nuclear magnetic resonance spectroscopy, and its use in beer research, multivariate analysis, and database structure.

### 2.1 A brief history of beer

This section is based on the chapter "A Brief Economic History of Beer", from the book "The Economics of Beer", a comprehensive review of the history and economic perspective of the brewing industry by E. Poelmans, 2012 [7]. As well as "A Comprehensive History of Beer Brewing", an exploration of brewing in ancient cultures by F. Meussdoerffer, 2009 [1].

Beer, or at least beer-like fermented beverages, has been discovered independently by almost all human societies. The production of such alcoholic beverages by fermentation of sugar-rich foods like fruits or grains represents some of the very first applications of what we may think of as biotechnology. Evidence of fermented drinks has been found in ancient China as early as 7 000 years BC, and traces of barley beer have been confirmed in early Mesopotamia and Egypt around 5 000 years BC [1]. Here in Norway beer has served an important role in Old Norse culture and ideology, with significant parts of the theogony and myths of Norse mythology being beer-related [8].

The history of fermented beverages and beer is as long as civilisation itself, but here we need only understand the more modern developments in Europe after the 14th century, as this is the source of the majority of modern beer making methods and traditions, and has thus largely shaped beer into what it is today.

Until the mid-to-late Middle Ages, brewing was mainly done in homes, farms, and monasteries. Monasteries were the only producers of beer on a large scale and brewed mainly for the consumption of the monks, pilgrims, and the poor who sought help there. Between 1300 and 1500 AD, beer evolved from being produced on a small scale, to becoming a large-scale export industry with dedicated breweries producing higher quality beer on a larger scale. The move from small-scale home production to larger-scale export was due to a myriad of factors, some of which were a rising demand for beer during the reexpansion after the *bubonic plague* that killed a large part of the European population, coupled with

---

a growing preference for beer over water, as the beer was boiled and thus significantly less likely to transmit disease compared to the polluted water that was the alternative. Another important factor in the growth of the brewing industry was travel. As more and more merchants traveled between towns, the demand for lodging and food grew, and so the number of inns and taverns increased. These became centers for social interaction, and naturally the sale of alcoholic beverages, primarily beer, became common in these places. One important factor in the change from monastic to commercial brewing were the taxation of beer. Prior to the use of *hops*, a mixture of herbs called *grut* were usually used for flavouring. The local rulers used to tax beer by forcing the brewers to buy *grut* from them instead of taxing beer directly, however, with the switch from *grut* to hops during the late Middle Ages, taxation became more difficult as the hops could be produced far more easily and the rulers thus lost their monopoly on flavouring additives. The solution was clearly to tax the beer itself, and since the Catholic monasteries were exempt from tax, the local rulers would favour commercial breweries who would provide taxes for the beer they produced [7], [1].

The use of hops started in northern Germany during the 13th century and significantly improved both the taste and the shelf life of beer. It was discovered that the hop plant (*Humulus lupulus*) had slight antibacterial effects that reduced spoilage and stabilized taste compounds in the brew. This allowed for longer storage times and thus the possibility of large-scale distribution and export. This, coupled with the use of larger brewing vessels and thus larger production volumes, turned brewing into a more capital-intensive industry, which also contributed to the centralization of the brewing process [9].

Unification of brewing into larger facilities and the increased commercial interest in beer lead to many important developments in brewing technology during the 18th and 19th centuries. One of the most noteworthy of these were an increased understanding of brewer's yeast and fermentation, which led to the development of the *lagering* process, and the production of *bottom-fermented* lager beers. The introduction of steam engines for production and transport, as well as new methods of refrigeration and new options for storage in glass bottles and later cans, significantly increased the ability to store and distribute beer. This increase in distribution and storage further pushed the brewing industry towards more centralization and large-scale production. [7], [1]

The modern day result of all this evolution in brewing technology is a substantial global market, comprising both large and small breweries that together produce a large variation of different beer types and styles [10].

### 2.1.1 Classification of beer

The definition of beer has changed widely since the first jars of fermented barley in ancient Mesopotamia, and many restrictions on what constitutes beer have been used since then. In Germany in 1516, William IV of Bavaria adopted a law called the *Reinheitsgebot*, which translates to *purity law*. He sought to specify and constrict the ingredients used in beer, as well as the price of the beer. Under this law, the only ingredients allowed were barley, hops and water. These restrictions may have had many reasons, such as preventing competition for grain with bakeries by restricting brewing to the use of barley, or to prevent the sale of beers from Northern Germany that contained additives not used in Bavaria. The observant reader may notice that yeast is not included in the purity law, which might seem like a grievous oversight. Brewers in this period saw yeast as a fixed part of the brewing process, the yeast culture in one beer batch was usually transferred to the subsequent

---

batch, thus specifying yeast in the law would be like specifying what kettle or stirring rod to use. Only later with the understanding of fermentation following the discoveries made by Louis Pasteur was yeast specified in the Reinheitsgebot. The law is of course no longer enforced, but many breweries still adhere to these strict specifications, and it still to this day represents our general understanding of what beer is. [7], [1]

### 2.1.2 Beer styles

Despite strict specifications of what goes into beer, a vast assortment of different beer styles has developed over the years by varying the ingredients and brewing procedure of the beer. There is no official consensus on what defines one beer style and separates it from another, but there is a general understanding of roughly what constitutes each style of beer. In most cases, the style is determined by the brewer who will classify the beer by the ingredients used and the style he or she intends the beer to be. Different styles of beer are usually separated either by the ingredients used, the taste profile, or by the area the style originates from. Most generally, we divide beer types in two main categories: Ales and lagers, defined by the use of top and bottom-fermenting yeasts, respectively. The main difference between these two types is their fermentation temperatures and fermentation time, as well as the flavour profile they produce. Ales usually ferment at temperatures between 17 and 22 °C, while lagers usually ferment in the range 8 to 12 °C. Traditionally, these yeast types have been *Saccharomyces uvarum* and *Saccharomyces cerevisiae* for lager and ale yeasts, respectively. Within both categories we find beer with varying characteristics, like dark or light color, high or low levels of bitterness, or even the inclusion of other additives like fruit or spices that contribute to the taste. Variations in the *malt bill*, hop type, yeast, and flavour additives, among other things, will all affect the style or type of a beer [11].

## 2.2 The brewing process

Background for the brewing process is based on the chapter "The Brewing Process", from the book "Brewing Yeast and Fermentation" by Boulton & Quain [12], a detailed work describing common industrial brewing practices and the theoretical background behind them. This is supplemented by commonly known brewing procedures that may not be referenced.

The details in the process for brewing beer vary a lot depending on the beer, but the general procedure is usually the same for most styles. In short, it consists of malting and milling grains, extracting the sugars from the grains into water, boiling the water with hops, and then fermenting the resulting sugar water for a certain amount of time using a specific yeast or other fermenting organism. The procedure can vary greatly for different beer types, as well as for the scale of production. Large scale commercial breweries will use a significantly more complex procedure for purification and stabilization of the product compared to what a small-scale home brewer might. The procedure described here pertains to the general method used in our laboratory and is illustrated in figure 2.1.

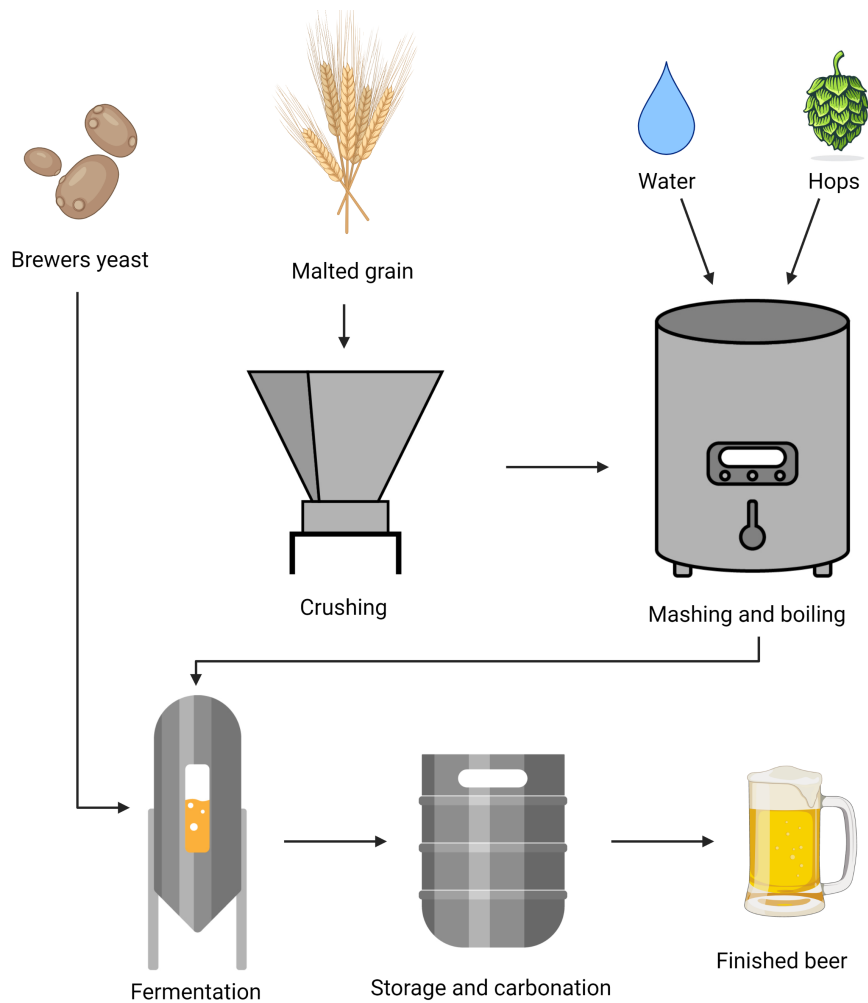


Figure 2.1: Visual representation of a simplified brewing process.

### 2.2.1 Malt and hops

Malt is grain that has gone through a *malting process*. This process is performed on grain that is to be used in brewing, and can be performed on many types of grain, but the procedure described here is for barley, the grain most commonly used in brewing. The barley undergoes controlled germination by steeping in water, this activates enzymes that break down starch into energy for the developing seed. This process is then prematurely terminated by applying heat in a process known as *kilning*. The enzyme systems and sugar reserves are now dormant, and will be reactivated again to release sugars during the *mashing process*.

Hop is a taste and conservation additive used in brewing that imparts bitterness and other floral and herbal taste components. It is made from the seed-bearing cones of the female *Humulus lupulus* plant, that are usually dried and pressed into pellets. Hops can be used in different stages of the brewing process, either during wort boiling and/or during fermentation.

Milling, or crushing, is the process by which the malted grain is crushed to expose the contents of the malt to the mashing water. This step is usually done shortly before the

---

brewing starts [12].

### 2.2.2 Mashing

Mashing is the process by which the crushed malt is exposed to water and the enzymatic reactions started during malting are reactivated. These reactions, catalysed by  $\alpha$  and  $\beta$ -amylase enzymes, change the starch-based energy reserves of the malt endosperm into fermentable sugars that can be utilized by the yeast during fermentation. When the extraction is finished, the spent malt is removed from the water. The steps of adding and removing the mash is known as *mash-in* and *mash-out* respectively. The resulting sugar rich water is known as *wort*. The mashing process is usually done at temperatures in the range of 55-68 °C, with different temperatures favoring the two main amylase enzymes differently, which effects the final sugar composition of the wort [12].

### 2.2.3 Boiling

After the mash is removed and the sugar-rich wort is formed, the mixture is boiled. Boiling is done for at least 60 minutes, and the process serves a host of different purposes. Firstly, it sterilizes the wort and inactivates any amylase enzymes left after mashing. The sterilization is noteworthy, as it allows all previous steps of the process to be performed under less rigorous conditions, significantly simplifying the process. The boiling also gets rid of many undesirable volatile compounds, many of which are extracted from the malt during the mashing process.

Adding hops, known as *hopping*, also happens during boil. The boiling process helps extract and isomerize the alpha acids from the hops, a process that is necessary for the desired flavor development in many hop variants. Those that impart mostly bitterness are added early in the boil and are heavily isomerized, but most of their aromatic compounds are lost, while others are added later and have a more balanced isomerization to aromatic ratio. This all depends on the hop used and the intended taste profile of the beer. The hops are removed prior to fermentation.

Clarification of the wort is another benefit of the boiling process, where the elevated temperature causes proteins to coagulate and polyphenols to form an insoluble precipitate with other unwanted compounds. Most of this precipitate, called *trub*, gathers at the bottom of the boiler when the wort is cooled, and is therefore not transferred to the fermentor. This makes the final product clearer and gives a cleaner and more predictable taste profile [12].

### 2.2.4 Fermentation and yeast

When the sugars have been extracted during the mashing process, the mixture is sterile, and unwanted compounds have been removed during boiling, the wort is ready for fermentation. Firstly, the wort is cooled to an appropriate fermentation temperature, this depends on the yeast type but is generally in the range of 12-30 °C, and transferred into a fermentation vessel. The style of fermentor used can vary from large stainless steel fermentors in commercial breweries to small plastic buckets used for home brewing, however the general process is the same. At this point, it is important that all instruments that come in contact with the wort, including the fermentation vessel itself, are sterile. The sugar-rich wort is highly susceptible to contamination by unwanted microorganisms that

---

will affect the fermentation process and result in an imperfect product. When the wort has been transferred to a clean fermentor, a specific yeast strain is added to the wort, whereupon the liquid is mixed thoroughly to distribute the yeast evenly and introduce oxygen to the mixture to allow the yeast to multiply under aerobic conditions before the main fermentation starts.

The fermentation itself occurs within the fermentor under anaerobic conditions, during which the yeast metabolizes carbohydrates into ethanol and CO<sub>2</sub>. There are multiple complex reactions taking place during yeast metabolism, and the fermentation process produces a myriad of different taste compounds in addition to ethanol from different fermentable compounds in the wort [12].

The resulting taste compounds produced during the fermentation process depend on multiple factors, like yeast strain, temperature, wort composition, amount of yeast used for inoculation, known as *pitch rate*, as well as the use of taste additives like fruit, spices and hops.

Different yeast strains can produce vastly different taste profiles from the same wort, as different strains have slight variations in metabolic pathways, and thus vary in what taste compounds they produce and what substrates they can metabolize [13].

Pitching rate can also affect the taste profile of the fermentation product, as the yeast population is experiencing varying levels of stress during the fermentation process depending on the number of yeast cells used for inoculation. Stressing the yeast will change its gene expression pattern, and thus also its metabolic pathways, resulting in different taste profiles. The effect of pitching rate varies between yeast strains and depends heavily on their resilience to stress [14].

Hops can also be added during the fermentation process. This is called *dry hopping*, whereby hops are added to the fermentation vessel and allowed to steep. This will release different flavours from the hops added here compared to the hops added during boil, and will often impart more complex floral notes to the beer. The flavour imparted by the hops during dry hopping is heavily dependent on the hop variant, the amount used, and the length of time it spends in the wort [9].

### 2.2.5 Storage and bottle fermentation

When the fermentation is complete, the finished beer is transferred into either bottles or kegs for storage. Storing the beer will impact the taste of the beer, as reactions occur within the beer over time. In contrast to some wines, the aging of beer is usually not considered a benefit for flavor quality, however, it may result in a rounder and milder taste in cases where spices are heavily used for flavoring. There are many reactions responsible for the change in flavor, both oxidative reactions caused by the presence of oxygen and *Reactive Oxygen Species* (ROS), and non-oxidative reactions like *esterifications*, breakdown of products from *Maillard reactions*, and *glycoside* to name a few. The degradation of the beer during storage is highly dependent on the environment the beer is stored in as well as the initial ingredients and the brewing process employed [15]. Different beer types react quite differently to storage, and studies have shown that specialty beer like dark and hoppy ales are less stable in taste profile compared to light lagers [16].

Another aspect of storage is adding CO<sub>2</sub> to the beer, often referred to as *carbonation*. This can be done either by *forced carbonation*, where the beer filled in a keg and pressurized

---

with CO<sub>2</sub>, or by a process called *bottle fermentation*. Bottle fermentation is done by transferring the finished beer into a bottle and adding a small amount of sugar, this allows the remaining yeast in the beer to continue fermentation for a short time, and will thus produce CO<sub>2</sub> that carbonize the beer [17].

## 2.3 Nuclear magnetic resonance spectroscopy

This section describing the theory of NMR is based on the book "Spin Dynamics: Basics of Nuclear Magnetic Resonance" [18], a comprehensive and modern introduction to the principles of NMR by M. Levitt, 2008.

*Nuclear Magnetic Resonance* (NMR) spectroscopy is a technique that can be used to elucidate the chemical structure of molecules within solution. The basic principle of NMR spectroscopy is measurement of the phenomenon of magnetic resonance of nuclei. If a sample in a magnetic field is subjected to radiation (energy) at a specific frequency, the sample can absorb the energy and reach a new energy state, called *excitation*. The specific radiofrequency necessary to achieve excitation is dependent on a few things: The type of nucleus to be excited (e.g., <sup>1</sup>H or <sup>13</sup>C), the chemical environment of the nucleus (e.g., protons in a methyl, hydroxyl or amino group), and the spatial location within the magnetic field if the field is not uniform. [18]

### 2.3.1 Nuclear spin

Nuclei have magnetic fields, these fields are created by the combination of the positive charge of the nucleus and *nuclear spin*, where the nucleus acts as though it is spinning around a single axis. If the nucleus is subjected to a magnetic field, the spin axis will orient itself in a specific direction determined by the magnetic field. However, the axis of spin can be in one of two states, one being a stable state (low energy state) and the other being an unstable state at exactly 180° in the opposite direction (high energy state). This two-state situation is not the case for all nuclei, but for many of those with biological interest (i.e., <sup>1</sup>H, <sup>13</sup>C, <sup>15</sup>N, <sup>19</sup>F and <sup>31</sup>P) this is the case. These are said to have a *nuclear spin number* I = 1/2. [18]

### 2.3.2 Resonance phenomenon

If a nucleus in a magnetic field is subjected to irradiation in the radio frequency range, and the energy is exactly the same as the energy difference between the orientation states in the nucleus, the nucleus will "flip" and enter the other energy state. When the irradiation is removed, the nuclei will revert to their original state. Both the excitation and relaxation processes can be detected by finely tuned coils of wire that measure the voltage induced by the excitation process. This signal is known as *Free Induction Decay* (FID).

The energy required to excite the nuclei ( $\Delta E$ ) is dependent on the strength of the magnetic field ( $B_0$ ), the *gyromagnetic ratio* for the type of nucleus ( $\gamma$ ), the *Larmor resonance frequency* ( $\omega_0 = 2\pi\nu_0$ ), as well as the Planck's constant ( $h$ ). The equation for the excitation energy is expressed as:

$$\Delta E = \gamma h B_0 / 2\pi \tag{2.1}$$



---

Using the Bohr condition ( $\Delta E = h\nu$ ), stating that the frequency of the transition is equal to the change in energy between two states, we can express the frequency of the excitation energy as:

$$\nu_0 = \gamma B_0 / 2\pi \quad (2.2)$$

By measuring the frequency required to excite a nucleus, it is possible to determine its chemical shift. A chemical shift is caused by the magnetic field created by the electron cloud in the vicinity of the nucleus, having a shielding effect on the magnetic field being applied, and thereby changing the resonance frequency compared to an identical nucleus in a different chemical environment. Measurements of the excitation frequency of all proton nuclei in a sample can therefore yield accurate information on the environment around the different nuclei, which can be interpreted to reveal the molecular structure.

The explanation above gives a brief introduction to how NMR works. The most basic form of NMR is 1D proton spectroscopy, meaning that the magnetic field is tuned to affect hydrogen nuclei. This is quite useful as most compounds of biological interest contain at least one hydrogen atom. The measurements described can also elucidate further information, like the chemical neighbourhood of a nucleus based on *coupling constants* between other similar nuclei. NMR can also be used quantitatively to determine the concentrations of compounds, and while it has traditionally been used to analyze pure compounds, developments in techniques have allowed for use in complex mixtures like that of beer [18].

## 2.4 Treatment and chemical analysis of beer by NMR spectroscopy

This thesis aims to enable quantitative analysis of components in different beers using NMR spectroscopy. To gain any useful insight from the comparison of NMR spectra from different beer samples, both the sampling and data treatment methods must be as uniform as possible without introducing any bias towards certain sample types.

### 2.4.1 Use of NMR spectroscopy for beer analysis

NMR spectroscopy has been used in multiple studies of beer since the 1980s, and the methods have developed over the years since. NMR has been used both in qualitative and quantitative studies on beer, ranging from differentiation of beers from different brands [19], to quantification of specific compounds [20].

Most compositional analysis by NMR has been done on either hop compounds [20], or carbohydrate composition [21]. Hop compounds are of great interest because of their importance in the overall taste profile, as well as effects on the bitterness and stability of the finished product. Analyses of hops revolve mostly around the formation, stability, and taste of  $\alpha$ - and  $\beta$ -acids, as well as their isomerization process [22]. Carbohydrates are of interest due to their contribution to the mouthfeel and body of the beer, as well as their effects on viscosity, filtering requirements and haze formation, all important aspects of both taste and quality of beer products [23].

In recent studies, the combination of NMR spectroscopy with *Multivariate Data Analysis* (MVDA), like *Principal Component Analysis* (PCA) and *Partial Least Squares-Discriminatory*

---

*Analysis* (PLS-DA), has shown promising results in both differentiation of beers based on type (ale or lager) [24], [19], and in rapid quantification of certain important compounds in a beer sample [20], [25], [26]. A thorough review of the state of NMR beer research in 2011 by J. E. A. Rodrigues [22], recognized that to create models with predictive or real statistical power large scale data would be needed. They also noted that the combination of NMR spectroscopy and *Mass Spectrometry* (MS), with MVDA would be a powerful tool in future studies towards a better understanding of the effects of raw materials on the final product, tracking of compositional changes during different processes, and the development of new methods for correlating chemical analysis data to quality control and sensory analysis. [22].

#### 2.4.2 Sample treatment

Some of the important considerations in sample preparation is firstly that the samples must have a similar pH. Differences in pH cause significant differences in the resulting NMR spectra, as pH effects the protonation state of functional groups, and in turn the chemical shift. Control of pH can be achieved using an inorganic buffer to adjust the pH of different samples to be similar without introducing anomalies that could be mistaken for yeast metabolites [27]. Secondly the sample must be homogeneous, as gas or particles can cause disturbances in the magnetic field during NMR acquisition and can affect the quality of the spectra. Filtering and *ultrasonication* will remove particles and gas bubbles that could cause issues in the sample [28], as well as removing any yeast cells present, thereby preventing any further fermentation within the sample. Thirdly, the sample needs to contain a reference compound, with a known chemical shift to calibrate the shift of the spectrum. This can be solved by including a known quantity of *Trimethylsilylpropanoic Acid* (TSP) with the sample for reference as the  $^1\text{H}$  signal in the TSP will not overlap with any of the signals occurring in the beer [28].

#### 2.4.3 Data treatment

In order to evaluate differences in beers with NMR spectroscopy, it is crucial that the treatment of the data allows for proper comparison between spectra. An important step in this process is the alignment of the different spectra. In order to be able to statistically compare different spectra, they must first be calibrated using an internal standard, as well as being perfectly aligned with each other. Further misalignment can be caused by small differences in pH that remain after buffering, or by slight temperature differences. Alignment can be completed using the *icoshift* software in *Matlab*, which allows for quick and flexible alignment of 1D NMR spectra for metabolomics. *Icoshift* allows for processing of full resolution spectra, so that any dimensionality reduction we would wish to do can happen after alignment and therefore do not affect the accuracy. [29]

Normalization and scaling for statistical analysis are also important considerations in pretreatment that may affect the results of any analysis to be performed. Normalization is a process by which the relative intensities of the spectra are scaled to be the same. This is crucial to account for differences in dilution between samples. In this thesis, all normalization is done using probabilistic quotient normalization, which scales the spectra using a calculated median of all spectra in the study as a reference [30]. Scaling is the final pretreatment method that is relevant for our data. The aim is to adjust for the difference in variation between the compounds, by converting the values into differences in relative

---

abundance compared to a scaling factor. *Unit variance scaling*, also known as auto scaling, is a common approach that uses standard deviation as a scaling factor. It leaves all compounds with a standard deviation of one, and thus allows for analysis on the basis of correlation instead of covariance [31]. Scaling can however cause problems with baseline noise in spectroscopy, scaling is therefore not used for nonparametric statistical analysis in this thesis. For parametric tests like *Student t-test* (t-test) however, unit variance scaling was applied.

## 2.5 Statistical data analysis

For the actual statistical analysis of the spectra, one observation is important: There is no single outcome variable to evaluate. When comparing and analyzing differences in these NMR spectra, we are looking at what is essentially many measurements made on each experimental unit (sample) with multiple outcome variables to consider. This is a multivariate problem, and we must therefore use analysis methods that handle multivariate analysis, such as PCA, PLS-DA, and clustering methods [32]. There is a multitude of possible analysis tools for multivariate analysis, one of which is the web-based service *MetaboAnalyst*, a free service that enables rapid analysis using a variety of different analysis methods, including specialized tools for metabolomics data from NMR-analysis, which is the closest common analogy to our type of data. *MetaboAnalyst* also provides code in the R-syntax for all analysis performed to enable ease of reputability and transparency [33].

### 2.5.1 T-tests and ANOVA

The statistical tools t-test and *Analysis of Variance* (ANOVA) are both methods for evaluating if there is a true difference in the mean between groups. The t-test is used in cases of only two groups, while ANOVA is used in cases of more than two groups. Based on the means, variance, and sample size of the groups, a t-test will calculate the probability of random chance resulting in a difference in the means equal to or greater than the one observed. In the case of an ANOVA, the principle is the same, only expanded to evaluate differences in means between more than two groups [34].

### 2.5.2 Principal Component Analysis

PCA is a fundamental analysis method in multivariate analysis and is based on reducing the dimensionality of the data into the dimensions that explain the most variance. The algorithm creates consecutive *principal components* by finding the regression line that maximises the variance along the line, then finding the next such line that is orthogonal to the first one, and so on until no variance remains. In mathematical terms, this amounts to calculating the *eigenvectors* of the *covariance matrix* and ranking them by their corresponding *eigenvalues* in decreasing order. Thus, the principal components become the new axis of the data, constructed so that they explain the maximum variance possible. Each principal component explains as much of the remaining variation as possible, therefore the first components explain the majority of the variance. The *loadings* of the PCA describe the weight of each variable in the construction of the principal component. In practical terms, this means that variables with large absolute values in their loadings have more variation between samples, and is therefore more important for the placement of a sample

---

along a given principle component. To find variation in peaks between samples in our spectra, we can therefore look for data points with high loading values. [35].

### 2.5.3 Clustering and supervised analysis

Unsupervised methods for multivariate analysis, such as PCA, can be used to detect patterns, trends, and outliers in a data set. However, unsupervised methods are incapable of discriminating between variations within a sample group and between different groups. Supervised methods, like PLS-DA, take group identification into account when evaluating variation. These methods can therefore be used to highlight the actual differences in compounds that separate the different sample groups or treatments, and provide loading plots to identify the areas of the spectra that are responsible for the observed variation. Much like in PCA, the loadings from PLS-DA identify peaks that vary between samples, however, since PLS-DA tries to separate based on categories, these loadings show peaks with categorical differences. Note that supervised methods are prone to overfitting and must therefore be validated. This can be done using a permutation test, which runs the model multiple times, each time assigning incorrect class labels to some of the samples, and comparing the separation of the incorrect classes to the results from the original classifications. This way, the permutation test measures of the classification given are significantly better than randomly assigned classes, and provide a p-value for the probability of the separation observed to be due to random chance [36].

Other unsupervised methods for clustering can help group samples according to similarities, and may therefore be valuable in evaluating the categorization of the data or for validating analysis methods. The most commonly used unsupervised clustering method is *k-means clustering*, which creates k-number of clusters by minimizing the Euclidean distance from each sample to one of k-number of cluster centers, using the first two principal components [37]. *Agglomerative hierarchical clustering*, represented by a *dendrogram* is another valuable exploratory clustering algorithm that does not require a specified number of clusters, as opposed to the k-means algorithm. Dendrogram clustering creates binary pairs by grouping each sample with the most similar other sample in the set, creating nodes for each pair, and continuing in an iterative process to create new pairs of the nodes until a root node is reached. The resulting clusters can be easily visualized in a dendrogram [38]. Another powerful and flexible tool for examining clustering is an algorithm called *t-Distributed Stochastic Neighbor Embedding* (t-SNE). This algorithm is commonly used in the field of machine learning and is intended as a tool for faithfully mapping *high-dimensional* data onto a *low-dimensional* space [39], and allows for visualization of any apparent structure in the data that might not be evident by PCA or PLS-DA analysis. This is done by calculating *similarity scores* for all points in the high-dimensional space, these scores describe the distance between each set of two points. The scores are then scaled using a normal distribution, and then projected onto a lower-dimensional space. The algorithm recalculates the similarity scores for the lower-dimensional space, scales them using a *t-distribution*, and thereafter iteratively rearranges the points until the pattern of the *similarity score matrices* for the high and low dimensional data are identical. The t-distribution is used for scaling of the low-dimensional similarity scores because it creates more relative space between clusters in comparison to other distributions, making visualization much simpler. This way, the algorithm is able to map the data onto a lower-dimensional space while the clustering in the high-dimensional space is preserved [40]. The *perplexity* parameter in t-SNE allows us to tune the balance between local and global aspects of the data by providing a number of expected close neighbours for each data point [39].

---

### 2.5.4 Fingerprinting

When attempting to analyse highly complex media such as beer by NMR spectroscopy, it quickly becomes evident that identifying all compounds and peaks will be quite difficult if not impossible, as we have large areas with heavily overlapping signals. A traditional approach of compound identification and quantification will not be viable for regular 1D proton NMR, however the use of *metabolite fingerprinting* is a promising alternative. The approach aims to highlight compositional similarities and differences in samples and groups of samples, evaluating the spectra as whole systems instead of trying to quantify sample differences in each compound [41]. The fingerprinting approach will be used for most of this thesis, combined with efforts of identification and quantification of important compounds when possible.

### 2.5.5 Machine-guided analysis methods

An eventual goal for the database effort is to have access to a large data set for analysis. Such analysis can hardly be done manually, therefore we must consider possible machine-guided analysis methods for analyzing large quantities of data. The relevant methods for our efforts fall roughly into two categories: Knowledge discovery and knowledge prediction.

Knowledge discovery methods aim to explore data sets and uncover unknown structures within the data, such as patterns, connections, and similarities that might not be immediately apparent to human analysts. One major method for knowledge discovery is the use of *data mining algorithms*, these algorithms sift through the data searching for all discernible patterns therein. Normally such algorithms do not discriminate, however, and false discoveries will be highly prevalent. It is therefore important to develop accurate algorithms and to employ the knowledge of domain and data experts when evaluating the structures found [42].

Knowledge prediction pertains to constructing models for extrapolating on known data to predict unknown information. This can be predicting the results of trends in the data or creating connections between elements based on similar relationships in other elements. *Machine learning* is one of the most powerful modern tools of knowledge prediction, these complex algorithms are able to teach themselves to recognize and predict patterns based on knowledge gathered through data mining, and form new prediction algorithms that are not explicitly programmed beforehand [43].

In our case, such tools will likely be essential for studying the classification structure of beer, evaluating similarities and differences between sample treatments, and searching for patterns in customer preference. As well as predicting new patterns, preferences, and categories that might not be fully realised yet.

## 2.6 Ontology

This section is based on a manual for ontology development by Natalya F. Noy and Deborah L. McGuinness at Stanford University, titled "Ontology Development 101" [44]. The work is a comprehensive guide to establishing an ontology, as well as the theoretical background needed to do so.

---

### 2.6.1 Defining ontologies

Defined as “explicit formal specifications of terms in the domain and relations among them” by Thomas Gruber in 1993 [45]. Meaning it is essentially a vocabulary of terms within a given domain or field, where all relevant concepts are defined and named. This definition is quite cursory and there is a lot more to ontologies than just glorified vocabularies. Ontologies let us create accurate understandings of the concepts within our domain, the relationships between those concepts, and the rules that govern both the definitions and relations. Modern ontologies are both machine and human readable and they can help us structure, store, and infer information from our knowledge base [44].

### 2.6.2 Purpose of ontologies

The purpose of ontologies is to facilitate communication between users of a domain, as well as between the users and computer software. For any communication to be accurate, we need clear definitions of concepts and instances within our domain. This allows us to avoid misunderstandings or errors caused by a vague understanding of a concept, or miscommunication caused by the use of different synonyms or naming conventions among domain users. These definitions, as well as the relations between them may be recorded in an ontology [44], [46].

Another major benefit of using ontologies is the ability to link our knowledge structure to software. By constructing classifications and relations with rules that are machine readable, we can produce intuitive digital storage repositories based on the ontology structure. We may also use the rules for structures and relations in the ontology to guide software to infer information about the instances we include. The machine readability of ontology languages are among their most important and defining features, as the interaction between ontologies and software is what elevates them from mere dictionaries to truly powerful tools of annotation and reasoning. When digital knowledge bases with overlapping coverage can use the same exact naming conventions and definitions, the sharing of knowledge and data between repositories becomes significantly simpler. The use of universal ontological terms for describing domains also simplifies the extension of domain knowledge and facilitates the sharing of methods for describing complex concepts between domains. One example used by Noy & McGuinness(2001) [44] is the ability to share ontological representations of time, like time intervals, points in time, etc. Making such definitions universal would greatly improve our ability to share knowledge and data in exact ways.

The ability to make domain assumptions explicit and editable is also a benefit of digital ontologies. If the assumptions and definitions of a domain are stated in an ontology, they can easily be edited and changed over time as our understanding of the domain concepts evolve and expand. Having these definitions stated also simplifies the transition into a domain for new users, as having clearly stated and structured definitions of concepts will make learning them significantly easier [44].

The creation of an ontology for the classification of beer will be central to the aim of this thesis. Since there is no official consensus on the definitions of beer types and classes, we must define our own classifications in order to have our analysis be informative and useful. This classification structure will also be made publicly available, so that it may be used by others and thereby contribute to a standardization of beer categorization. This classification system will also serve as the template for the database structure that is to follow from this thesis, and serve as a temporary method of recording and annotating

---

samples generated or gathered for the project.

### 2.6.3 Ontology concepts

There are a few terms that will be necessary to understand in order to comprehend the descriptions of the ontology, these are as follows:

- *Instances*: The smallest units included in a class structure of an ontology, usually the objects or concepts the ontology is seeking to categorize.
- *Object properties*: Properties of a class or an instance that describe the object in relation to other objects. In this ontology these are properties that describe concepts like one object either being or having an ingredient.
- *Data properties*: Properties describing the nature of a data field, this can be concepts like date, age, or name.
- *Property assertions*: The connection of a data or object property between two objects. These function as links between objects, and describe certain concepts like a certain object being an ingredient in another.
- *Class axioms*: Specific rules for a class, describing a defining trait of the class, often in relation to other classes.

Note that property assertions and class axioms serve a similar purpose, the difference being property assertions acting on instances of a class, and class axioms acting on the class itself [44, 47].

## 2.7 Databases

### 2.7.1 Fundamentals of databases

A database is an organized repository of stored information. Storage is generally digital, and there is usually some implementation of software to enable data access through queries and data filtering. There are endless variations on how databases are structured, but a general design principle is based on there being multiple parts of a database system, with a central component called a *Database Management System* (DBMS) handling communication between the actual information repository and the users or applications that are fetching or delivering the information [48]. There are multiple types of databases, a widespread type is *relational databases*, which is what we will focus on here.

A relational database consists of multiple tables of columns and rows that represent relations. Generally, a table represents a data category where rows are referred to as *records* and columns as *attributes*. Each record is usually an instance of a type, such as a car, and the attributes represent some value of the instance, like the registration number for the car. Each record, or row, will have a unique key. This key can appear in different tables giving the instance multiple attributes within different data categories. Tables can also be linked by relationships, which function as logical links based on the interaction between the tables [48].

---

### 2.7.2 Structured Query Language

Querying, filtering, and retrieving the data stored in a repository is usually handled by a querying language. *Structured Query Language* (SQL) is the most commonly used language to handle relational databases. It consists of multiple sublanguages that handle querying, access control, data definition, and classification, as well as data manipulation like deletion, insertion, or updating. While there are multiple versions of SQL, their general functionalities are similar [49].

### 2.7.3 Application programming interface

An Application Programming Interface (API) is a part of many database systems. These interfaces facilitate controlled use of external programs to interact with the database system. Using an API, we can design custom applications to handle any sort of interaction with the database, such as importing, exporting, or reading information in ways required for specialized data structures. The flexibility of APIs allows for a far broader functionality of most database systems, and API support will be required for the database solution in this project [50].

### 2.7.4 Challenges in beer studies

Storage of data for beer research poses a challenge in sample categorization and metadata recording. Firstly, there is no established way to store scientific data on beer along with chemical spectra or data from other analysis methods. Beer is a complex medium made up of many different ingredients, as well as a variety of production processes that must all be captured in a database. Information on the beer production and recipe will be stored as metadata: supplementary data unique to each sample that describes its properties. Because the structure of the metadata can vary significantly between different beers, the database used must be highly flexible. The second challenge comes in the form of chemical spectra. These are large spreadsheets, whom in their native formats require proprietary software to interact with. This causes limitations in the querying and filtering processes applied to the spectra themselves.



# Chapter 3

## Materials and Methods

This section will describe the methodology and equipment used in the various parts of this project. The experimental efforts will be described in detail, while the theoretical efforts that make up a significant part of this thesis, including decision making and design choices, will be explained here briefly and be discussed in detail in chapter 5.

### 3.1 Case studies

To generate examples of the data to be included in the database, three different data sets were generated. The first is comprised of 19 commercial samples from *Austmann Brewery*. The second includes nine samples from an experiment manipulating the yeast pitch rate used in the same wort. The third data set is comprised of 30 samples from an experiment designed to evaluate the use of algae as a flavouring additive in beer, these samples were run twice at two different time points to examine the effects of aging, thus the data set contained 60 samples. The total number of spectra analyzed in this thesis was 88, however no analysis were done on the collection as a whole.

#### 3.1.1 Brewing procedure

For the samples produced in-house the following general brewing procedure was used and changed when necessary for the given recipe. Explanations of brewing-specific terms can be found in the background chapter, section 2.2.

The wort is generally brewed in a 55 L *Braumeister Speidel* using malt specified in the recipe of the beer. The malt is ground prior to brewing, and mash-in is performed at 58 °C. Salts are added to the water prior to mash-in, according to the water profile of the beer style. Mashing time and temperature is chosen according to the given recipe, and mash-out is done at 75 °C for 15 min. During mashing, stirring is performed every 15-20 minutes in order to break up channels forming in the mash and increase the sugar extraction rate. After mash-out, the mash is rinsed with water at 70 °C, which is later returned to the wort. The wort is boiled according to the recipe, usually for 60 min, and hops are added during the boiling according to the recipe. Flocculation material (*Irish moss* and *Whirlflock*) is usually added to the wort 15 min before the end of boil, this has no effect on taste and is intended to help larger particles flocculate before transferring the wort to fermentation tanks.

---

### 3.1.2 Categorization study of Austmann samples

For the study examining beer style categories, nineteen beers were sampled from the Austmann Brewery in Trondheim. The beers sampled represented the majority of the brewery’s assortment of beer. Each beer was sampled and treated as described in sections 3.2. They were divided into four basic categories for analysis based on their profile. The included beers and their categorization are shown in table 3.1.

Table 3.1: **List of sampled beers from Austmann brewery.** Including name, sample ID and categorization.

Beer name	Sample ID	Style category
Lynchburg Latte	AU10	Dark
Northie	AU11	Dark
Stille natt	AU18	Dark
Coffee Porter	AU3	Dark
Liquid Lumberjack	AU9	Dark
Old School IPA	AU12	IPA
Onkel I Amerika	AU13	IPA
Pine Shoot	AU16	IPA
Psycho Tropic	AU17	IPA
Captain Cold One	AU2	IPA
Hawaiian IPA	AU4	IPA
Hopbird	AU5	IPA
Pilsner	AU15	Light
Utpå	AU19	Light
Hoppy Blonde	AU6	Light
Lager 11	AU8	Light
Blood And Thunder	AU1	Sour
Peach Piñata	AU14	Sour
Hula Hula	AU7	Sour

### 3.1.3 Study of the effects of differing pitching rates on beer composition

For studying the effects of *pitching rate*, three different rates (underpitched, normal pitch, and overpitched) were performed in three parallels, each with the same wort using the same yeast, as illustrated in figure 3.1. One batch of wort was made and separated into nine fermentation vessels. Different amounts of yeast were added to each fermentor depending on the pitching rate, and three fermentors were pitched with each of the three rates. All fermentors contained 1 L wort and allowed to ferment for 14 days at 20 °C.

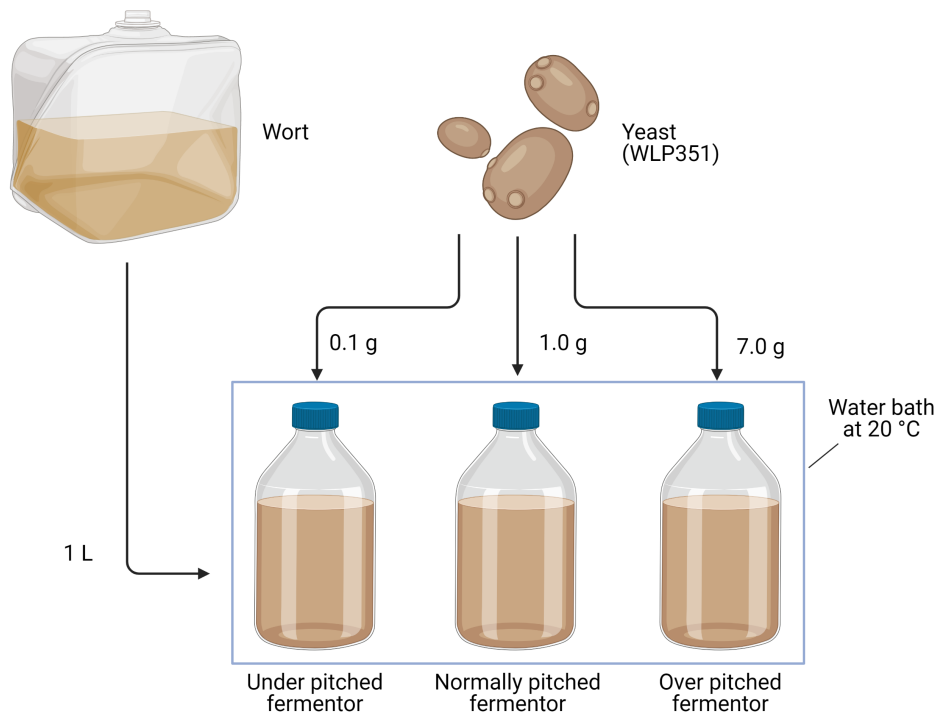


Figure 3.1: **Experimental setup of pitch rate study.** The figure illustrates one of three parallels in the study.

The recipe used for the beer was a *Belgian-style witbier*, a beer style with a light color and a significant portion of wheat in the malt bill. This recipe was chosen based on its yeast-dependent flavour profile and short fermentation time. The yeast used was a Bavarian *Weizen ale yeast*, produced by Whitelabs, designation WLP351. The yeast was chosen for its distinct flavour profile and high variability with different pitch rates.

The yeast was prepared by growing 40 mL (1 packet) of WLP351 in 2 L of malt extract in order to create enough yeast to inoculate the different samples. The total yeast mass produced by the starter was 30 g centrifuged yeast mass. Cell counts were done by centrifuging the yeast slurry in 50 mL Eppendorf centrifuge tubes for 2 minutes at 4500 rpm and 27 °C, discarding the supernatant and weighing out a small amount of centrifuged yeast mass. The yeast sample was then diluted in water and counted by microscopy in a Burker *hemocytometer* with 0.04 mm<sup>3</sup> counting chambers. Ten counting chambers were counted at varying locations on the hemocytometer, and the values were averaged. The cells / gram count was then calculated from the average hemocytometer count using the weight of the counted yeast mass and the dilution.

Pitch rates were chosen based on standard pitch recommendations of  $1 \times 10^6$  yeast cells / mL wort / °Plato [14]. The assumed gravity of the wort before fermentation was 15.3 °P, for 1 L wort the normal pitch rate was thus calculated at  $1.53 \times 10^{10}$  cells / fermentor. This results in 1.27 g of centrifuged yeast mass / fermentor based on the assumption that 1 mL centrifuged yeast mass weighs approximately 1 g. For this study, the standard pitch rate was set to 1 g centrifuged yeast slurry / fermentor. A "low" pitch rate was chosen at 0.1 g yeast / fermentor, and a "high" pitch rate was chosen at 7 g yeast / fermentor (due to limitations in yeast amount). Before inoculation, the yeast was centrifuged to an even mass and the supernatant removed. The amounts of yeast corresponding to each pitch

rate were weighed and subsequently used for inoculation of the corresponding fermentor. The amount of yeast used in the fermentation of each sample is shown in table 3.2.

The finished wort was transferred to 10 fermentation tanks (200 mL Schott Duran *protect* bottles with yeast locks) in 1 L batches. Each pitch rate were performed in three parallels, with one additional fermentor that was not inoculated as a control. The fermentors were then sealed with a yeast lock to provide an anaerobic environment, and placed in three different water baths containing one fermentor of each pitch rate to counteract potential temperature differences caused by the different pitch rates. Experimental setup for one parallel is shown in figure 3.1. The wort was then allowed to ferment for 14 days at 20 °C, before being transferred to a cooler to cold-crash for 4 days. After fermentation, a sample was taken from each fermentor, and the remaining beer was transferred to bottles for secondary bottle fermentation. The samples were then treated and analyzed as described in section 3.2.

Table 3.2: **Table of pitching rates for samples in pitch rate study**

Sample name	Pitch rate group	Pitch rate (cells / ml / °P)	Pitch rate (% of recommended)	Yeast used per 1 L (grams)	Sample ID
Under 1	Underpitched	$8.5 \times 10^4$	8.5	0.10	WU1
Under 2	Underpitched	$8.5 \times 10^4$	8.5	0.10	WU2
Under 3	Underpitched	$9.4 \times 10^4$	9.4	0.11	WU3
Normal 1	Normally pitched	$8.7 \times 10^5$	87.2	1.02	WN1
Normal 2	Normally pitched	$8.9 \times 10^5$	88.9	1.04	WN2
Normal 3	Normally pitched	$9.0 \times 10^5$	89.7	1.05	WN3
Over 1	Overpitched	$6.0 \times 10^6$	602.6	7.05	WO1
Over 2	Overpitched	$6.0 \times 10^6$	603.4	7.06	WO2
Over 3	Overpitched	$6.0 \times 10^6$	602.6	7.05	WO3

## Sensory analysis

For the pitch rate study, a quick qualitative sensory analysis was performed to evaluate if a taste difference in the samples could be detected, and if any trends were identifiable. The analyses were performed by 16 untrained participants in six independent groups. The participants had no prior knowledge of the experimental setup or variables of the samples, they were given one sample from each pitching rate and asked to identify any differences or trends in the taste profiles and sort the samples accordingly. Notes on the responses from the participants were compiled into a list of general taste descriptors and their intensities for each pitch rate.

### 3.1.4 Study of aging in algae-flavoured beer samples

The samples of beers flavoured with algae were produced in-house, and the worts for the four beers were brewed according to the procedure described in section 3.1.1, following the recipes in appendix A.2. After brewing, each wort were separated into either nine (Algae amber lager & Algae brown ale) or six (Algae IPA & Algae goose) small fermentors holding 2 L each. Yeast was added to the fermentors according to table 3.3, creating 10 different combinations of wort and yeast. The samples were fermented at 20 °C, however exceptions were made for fermentors with *Dahls yeast* that were set at 12 °C, and fermentors with *Oslo kveik* that were set at 30 °C. The temperature was regulated by placing the fermentors in

water baths holding the appropriate temperature. After half of the fermentation time had passed, 6 g of freeze dried and blanched algae of types *Saccharina latissima* (SL) and *Alaria esculenta* (AE) were added to each of the appropriate fermentors according to table 3.3. This setup resulted in 30 unique combinations of wort, yeast, and algae, including samples not containing algae. An example of the experimental setup for one beer type and one yeast is illustrated in figure 3.2. After fermentation, the samples were prepared, treated, and analyzed according to the procedure described in section 3.2.

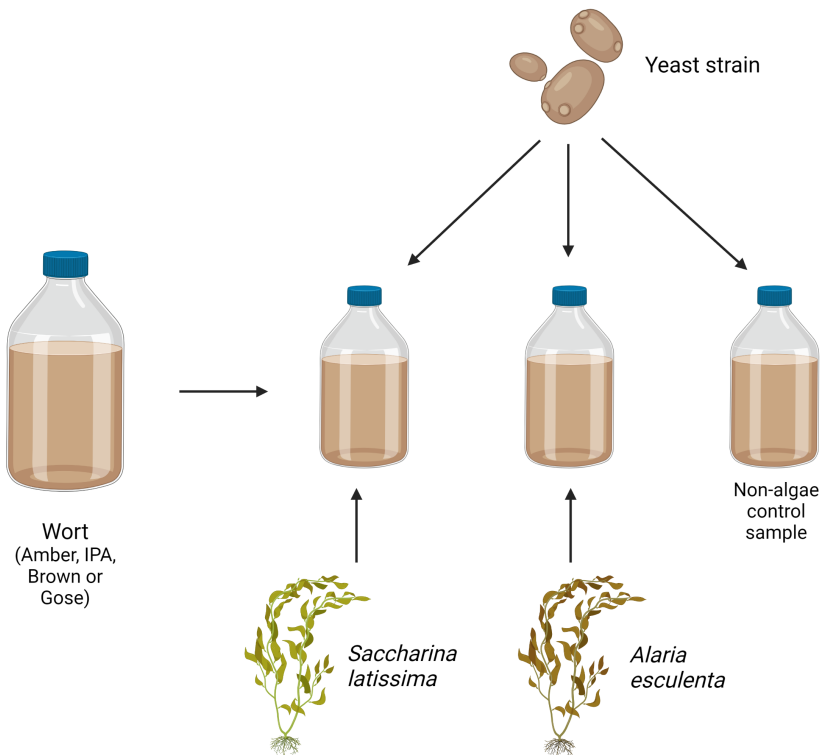


Figure 3.2: Algae beer experiment setup of one combination of wort and yeast.

The algae-flavoured beer samples were then stored in glass bottles in a dark cooler holding 4 °C for between 155-161 days (depending on the brewing date), a storage time of about 5 months or 22 weeks. After the storage time, samples were analyzed again according to the procedure described in section 3.2, and a comparative analysis were performed between the aged and non-aged samples.

Table 3.3: List of samples produced for the study of aging in algae flavoured beer. Including sample ID, beer type, yeast, algae type, brewing date, and sampling date.

Sample nr.	Beer type	Yeast	Algae	Brewdate	Sampling date
A1	Amber Lager	WLP001	SL	14-Oct	28-Oct
A2	Amber Lager	WLP565	AE	14-Oct	28-Oct
A3	Amber Lager	WLP001	No Algae	14-Oct	28-Oct
A4	Amber Lager	WLP001	AE	14-Oct	29-Oct
A5	Amber Lager	WLP565	SL	14-Oct	29-Oct
A6	Amber Lager	Dahls	AE	14-Oct	19-Nov
A7	Amber Lager	Dahls	SL	14-Oct	19-Nov
A8	Amber Lager	WLP565	No Algae	14-Oct	30-Oct
A9	Amber Lager	Dahls	No Algae	14-Oct	19-Nov

---

B1	Brown Ale	WLP013	AE	15-Oct	29-Oct
B2	Brown Ale	WLP001	SL	15-Oct	29-Oct
B3	Brown Ale	WLP002	SL	15-Oct	29-Oct
B4	Brown Ale	WLP002	AE	15-Oct	29-Oct
B5	Brown Ale	WLP001	AE	15-Oct	30-Oct
B6	Brown Ale	WLP013	SL	15-Oct	30-Oct
B7	Brown Ale	WLP001	No Algae	15-Oct	30-Oct
B8	Brown Ale	WLP013	No Algae	15-Oct	29-Oct
B9	Brown Ale	WLP002	No Algae	15-Oct	30-Oct
C1	IPA	WLP565	AE	16-Oct	30-Oct
C2	IPA	WLP565	SL	16-Oct	30-Oct
C3	IPA	WLP001	AE	16-Oct	30-Oct
C4	IPA	WLP001	SL	16-Oct	30-Oct
C5	IPA	WLP001	No Algae	16-Oct	30-Oct
C6	IPA	WLP565	No Algae	16-Oct	30-Oct
D1	Goose	Oslo Kveik	SL	20-Oct	2-Nov
D2	Goose	Oslo Kveik	AE	20-Oct	2-Nov
D3	Goose	WLP001	SL	20-Oct	2-Nov
D4	Goose	WLP001	AE	20-Oct	2-Nov
D5	Goose	Oslo Kveik	No Algae	20-Oct	2-Nov
D6	Goose	WLP001	No Algae	20-Oct	2-Nov

## 3.2 Sample preparation, NMR spectroscopy, and statistical analyses

### 3.2.1 Pipeline overview

The procedure to take a sample from the bottle through preparation, NMR acquisition, and analysis is based on the work done for the project by Kristoffer Zotchev [51], and is further standardized here. The procedure consists of four main steps:

1. Sample preparation
2. NMR acquisition
3. Spectral processing
4. Statistical analyses

Protocols for the first three of these procedures were standardized and a comprehensive instruction document was produced as a part of this thesis. This can be found in appendix E.

### 3.2.2 Sample preparation

The following is the protocol for preparing a sample acquired during the brewing process or from a finished beer for NMR analysis.

Samples of at least 2 mL were sterile filtered through a 25 mm diameter, *luer-lock*, 0.2  $\mu\text{m}$  pore size filter into 5 mL *Eppendorf tubes*. If at the time of sampling sterile filtering was unable to be carried out, the sample was frozen and then thawed before filtering. The tube was then placed in an *ultrasonic bath* for 10 minutes in order to properly degas the sample. Degassing was also performed on samples that have not been carbonated to ensure equal treatment of all samples. After degassing the samples were buffered by combining 720  $\mu\text{L}$  of sample with 80  $\mu\text{L}$  of buffer (500 mM sodium phosphate, pH 7.4, prepared in *deuterated* water ( $\text{D}_2\text{O}$  99.9 %) with TSP (1 %)) in a 1.5 mL *Sarstedt SafeSeal tube*. This resulted in a final buffering solution of 50 mM sodium phosphate, 10%  $\text{D}_2\text{O}$ , 0.1% TSP. For NMR acquisition, 600  $\mu\text{L}$  of the sample was transferred into a 5 mm *LabScape Stream NMR tube* (Bruker LabScape). A visual representation for sample preparation is shown in figure 3.3.

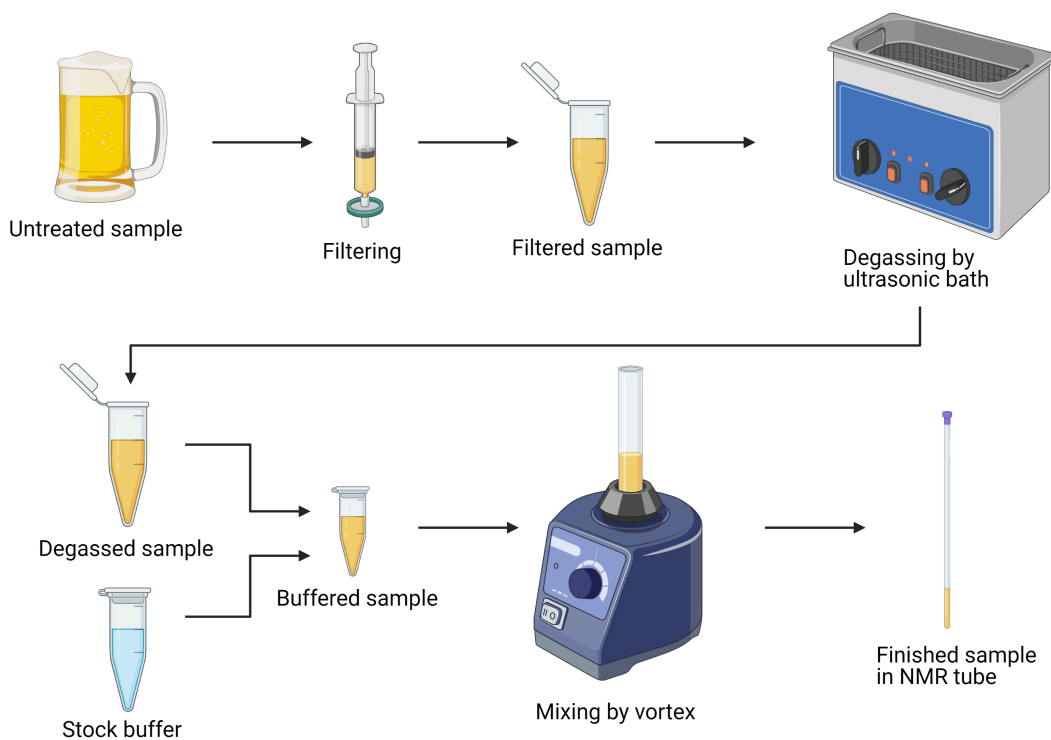


Figure 3.3: **Sample preparation procedure.**

### 3.2.3 NMR acquisition

NMR experiments were acquired on a *Bruker Avance III HD 800 MHz spectrometer*, using a 5 mm *Z-gradient CP-TCI (H/C/N) cryoprobe* and *SampleJet autosampler* at the NMR Laboratory of Natural Sciences at NTNU.

The NMR experiments used here are the same as those used previously in this project by Kristoffer Zotchev [51], which were adapted from Monakhova et al. 2011 [52]. Three sequential experiments were performed for each sample. Firstly, a standard water pre-saturation pulse program was used for water suppression, and the resulting spectrum was used to determine the frequencies of the seven peaks of the ethanol signal in the specific sample. These values were then used to guide the optimization of the shape pulse for multiple solvent suppression in the later experiments. This process was automated using a purpose-made ICON-NMR script written by Gaston Courtade and Trygve Andreassen.

---

The second experiment was a 1D  $^1\text{H}$  *noesygpps1d* spectrum where only water suppression was performed. The purpose of this experiment was to allow for recording of the ethanol peaks, and thereby enable measurements of ethanol content and attenuation by the yeast.

The third experiment was a 1D  $^1\text{H}$  *noesygpps1d* spectrum with both water and ethanol suppression, using the values recorded during the first experiment to generate the shape pulse. Continuous wave iterations were also applied to the center of the ethanol quartet during the relaxation delay to suppress the effects of  $^{13}\text{C}$  satellites.

Both *noesygpps1d* experiments were acquired with the following parameters:

- Time domain size (TD): 64k complex points
- Number of scans (ns): 8
- Dummy scans (ds): 4
- Relaxation delay (d1): 5
- Acquisition time (AQ): 2.5 seconds
- Spectral width (SW): 21.6966 ppm
- Receiver gain (rg): 10

### 3.2.4 Spectrum processing

Processing of the raw spectra from NMR acquisition is necessary for any meaningful analysis. This procedure consists of two main parts: processing and alignment.

#### TopSpin processing

Spectrum processing was done in the *TopSpin* software (Bruker, version 4.1.1), and consisted of a 0<sup>th</sup> order automatic phase correction and a *1D Fourier transformation* with exponential window multiplication. All samples were examined manually for anomalies or suboptimal signal suppression and deemed satisfactory. After processing, the spectra from each experiment were exported as an intensity matrix and ppm-scale by an in-house *Python* script (*bruker-to-mat*, appendix D) for import into Matlab.

#### Matlab processing

Further data processing was completed in Matlab (Mathworks, version R2020b), and involved alignment of spectra, normalization, and *binning*. Alignment was done using the *Icoshift* software for Matlab [53]. Firstly, the spectra within each data set were aligned to each other using the "average2" mode with a multiplier of 3.0 in the *Icoshift* function applied to the whole spectrum. Thereafter, the spectra were aligned by the TSP-peak, again using the "average2" mode with a multiplier of 3. A corrected ppm scale was calculated by calibrating the TSP peak to 0 ppm. After TSP-correction, the spectra were plotted and inspected for alignment errors. If further alignment was required, an additional interval-based alignment step was included. This was done by defining regions as a set of intervals



---

to be aligned independently using the "average2" mode with a multiplier of 3. The Matlab script used for alignment can be found in appendix B.1.

Normalization was done using an in-house Matlab-script written for normalising data using *Probabilistic Quotient Normalization* (PQN) based on [30], shown in appendix B.2. During normalisation, the spectra are also truncated to the region 0.8 and 9 ppm, while excluding the residual ethanol and water peaks. Negative values are also excluded. Finally, binning was performed on the spectra using an in-house Matlab script that sums the values for each spectrum within a given bin size, available in appendix B.3. This reduces the dimensionality of the data set, allowing for faster processing times of statistical analyses. Most data sets were analyzed both with and without binning with very similar outcomes.

### 3.3 Statistical analysis

After spectrum processing in Matlab, the spectra were exported to matrices in spreadsheets in *.csv* format. Files were uploaded to the *metaboanalyst.ca* website and statistical analysis was performed using the generic statistical analysis tool set. Files were uploaded as *spectral bins*, no filtering was applied, and only mean centering were used for normalization within Metaboanalyst. Thereafter, the data sets were analyzed with a variety of multivariate analysis tools depending on the study. Figures for the multivariate analyses presented here were exported from Metaboanalyst. In this thesis, no scaling was performed on the spectra due to issues with traditional scaling methods having adverse effects on noise in the baseline and crowded regions of the NMR spectra. In the case of parametric tests like t-tests, the data was scaled using unit variance scaling.

Within Metaboanalyst, both unsupervised and supervised methods of clustering and dimensionality reduction were used, as well as t-tests. PCA was performed on all data sets and decisions on what combination of principal components to evaluate were made based on *scree plots* of the variance explained by each component, as well as the apparent separation of the expected categories in the PCA. PLS-DA was performed in cases where we observed separation of categories in PCA. All PLS-DA plots were validated using permutation tests. T-tests were used to evaluate the difference between groups in the case of the aging study, for these tests the data was scaled using unit variance scaling.

Determination of the optimal number of clusters (k) for k-means clustering was done using *NbClust* package in R-studio. Optimal k was evaluated by three different methods: *Elbow method*, *silhouette methods*, and *gap statistic method*. Results from the gap statistic method were used in case of differing results. The script used can be found in appendix C.1.

Clustering analysis by t-SNE was performed in Matlab using the same data sets that were studied in Metaboanalyst. The analysis were performed using the *tsne function* with *Euclidean distance* as the distance measure. Perplexity was optimized for each data set by iteratively performing the analysis with increasing perplexity and choosing the value that provided the optimal separation of clusters.

Statistics on ethanol content in samples from the pitch rate study were achieved by comparing area-under-the-curve measures of the ethanol peak and the internal standard TSP-peak in TopSpin. Based on information about the amount of TSP, the amount of ethanol could be calculated. The differences in mean between groups were evaluated using ANOVA. Plots and statistical tests for these results were done using an in-house script in R-studio, and the script is available in appendix C.2.

---

### 3.4 Choice of database solution

The evaluation of different database solutions as a recommendation for future work in this project was done based on the necessities discovered during the studies performed for this thesis, as well as the general requirements for database solutions agreed upon by the research group. A shortlist of solutions for evaluation were provided by field experts, and examined by studying the documentation and use cases of the relevant solutions. A detailed description of the selection process and criteria used can be found in section 5.5.

### 3.5 Ontology creation

The beer ontology was created using the *Protege* ontology creation software. The classifications were based on two style guidelines published by the *Brewers Association* [54] and *CraftBeer.com* [55]. Final choices on preliminary classifications were made by consensus within the research group, however the classifications are at this point only guidelines, and are expected to change as new information is gathered. A detailed discussion on the design choices and procedure behind the ontology is presented in section 5.6.

### 3.6 Project website

The website for the project was created to enable *citizen science* by facilitating contact and sharing of knowledge and samples between the Systems Brewology Project and experts on beer, like professional brewers and experienced homebrewers. The website was created using *WordPress*, and the graphics and logo for the project was created in *Adobe Illustrator 2020* to increase the visual appeal of the website.

# Chapter 4

## Results and Analysis

The following chapter will present and analyze the results found during the different efforts of this thesis. This will include findings from the case studies of samples, as well as a brief presentation of decisions made regarding ontology design and evaluation of database solutions. This chapter will aim to evaluate and discuss the results on their own, the broader implications of these results will be discussed in chapter 5.

### 4.1 Case studies of beer samples

This section details the findings in the three different case studies performed on three sets of beer samples. These include an analysis of categories in commercial beer samples from the Austmann brewery, the study of the effect of varying the pitching rate during the brewing process, and the study of the results of aging in a complex sample set of algae-flavoured beer.

#### 4.1.1 Description and overview

To generate the data to be included in the database, three different brewing case studies were conducted. The resulting data were analyzed and examined in order to attain a clearer picture of the possibilities and limitations of the analytical and statistical methods used. Each data set were examined with one aspect of brewing or beer treatment in mind. The first data set is comprised of commercial samples from the Austmann brewery and were used to evaluate methods of categorization of beer by type. The second data set is from a study examining the effects of pitching rate in a heavily yeast-dependent beer brewed in-house. This data set was examined and cross-referenced with qualitative sensory data to evaluate the detection accuracy in the analysis. The third data set was from a study evaluating the viability of including algae as a flavor additive during beer brewing. This data set was sampled at two different time points, and the analysis performed here examined the effects of aging on the samples in this data set.

These analyses are intended as proof-of-concept and are mostly performed on small data sets from studies not specifically designed to examine the hypothesis in question. These results should therefore be considered not as conclusive evidence for or against the hypothesis, but rather as indications that may lead to future discoveries in large-scale targeted studies.

---

### 4.1.2 Data treatment and scaling

The general procedure for data treatment described in section 3.2.4 provided quick and easily analyzable raw data. The sample preparation procedure, NMR-acquisition, and spectrum processing done in Matlab were efficient and reproducible. A detailed instruction manual was created in order to document and standardize the process of taking a sample from a beer bottle through preparation, treatment, and acquisition by NMR. The manual was illustrated for better visual guidance in navigating complex software for NMR-acquisition. The manual is intended as a tool and reference material for later experiments, ensuring that the sampling procedure is consistent for every sample in the database. The full manual can be found in appendix E.

A problem with scaling procedures were discovered early during the process of standardizing data treatment, where common scaling methods such as unit variance scaling, *Pareto scaling*, and *range scaling* caused scaling of random noise in the baseline of the spectra to be overestimated compared to the actual differences in relevant peaks. This is an inherent problem in the full-spectrum approach because of the vast differences in the amplitude of peaks in different areas of the spectrum, the noise in one area might be larger than the relevant peaks in another, making scaling and filtering by normal methods an unviable option. To achieve a method of data treatment that would allow us to study the differences in peaks without baseline noise effecting the results, we opted to not perform scaling in these studies. This would allow the statistics to detect categorical differences in peaks in the nonparametric tests, although the relative statistical importance of different peaks were overestimated in cases of high amplitude and underestimated in areas of low amplitude. In cases of parametric testing, such as t-test, unit variance scaling was used, as these results would have been completely unreliable had scaling not been performed.

Evaluations of differences between binned and unbinned spectra were also done, and no perceptible differences were detected. For the following studies, all data was therefore binned, as it allowed for significantly shorter computing times and reduced errors due to the large size and high dimensionality of the unbinned data. Dimensionality were reduced by a factor of 80 without detecting noticeable effects on the statistical results.

### 4.1.3 Compound identification

A list of 22 compounds were identified by Leesa J. Klau, along with their location along the *Parts per Million* (ppm)-axis of the spectra. These are listed in table 4.1, and a visual representation of the location of the compounds in a representative NMR spectrum is given in figure 4.1. Later identification of taste compounds and annotation of spectra are based on this.

Table 4.1: **List of identified compounds and their location in the spectra.** Supplied by Leesa J. Klau.

<b>Molecule</b>	<b>ppm in spectra</b>
Histidine	8.00
Phenylalanine	7.43
Phenylalanine	7.39
Phenylalanine	7.33
Tyrosine	7.19
Histidine	7.06
Tyrosine	6.89
Acetylchline	3.22
Mallic acid	2.68
Succinic acid	2.41
Pyruvic acid	2.37
Acetylcholine	2.08
Proline	2.00
Acetic acid	1.92
Isobutanol	1.75
Isopentanol	1.66
Alanine	1.48
Isopentanol	1.44
Lactic Acid	1.33
Valine	1.04
Higher alcohols	0.89
TSP-d4	0.00

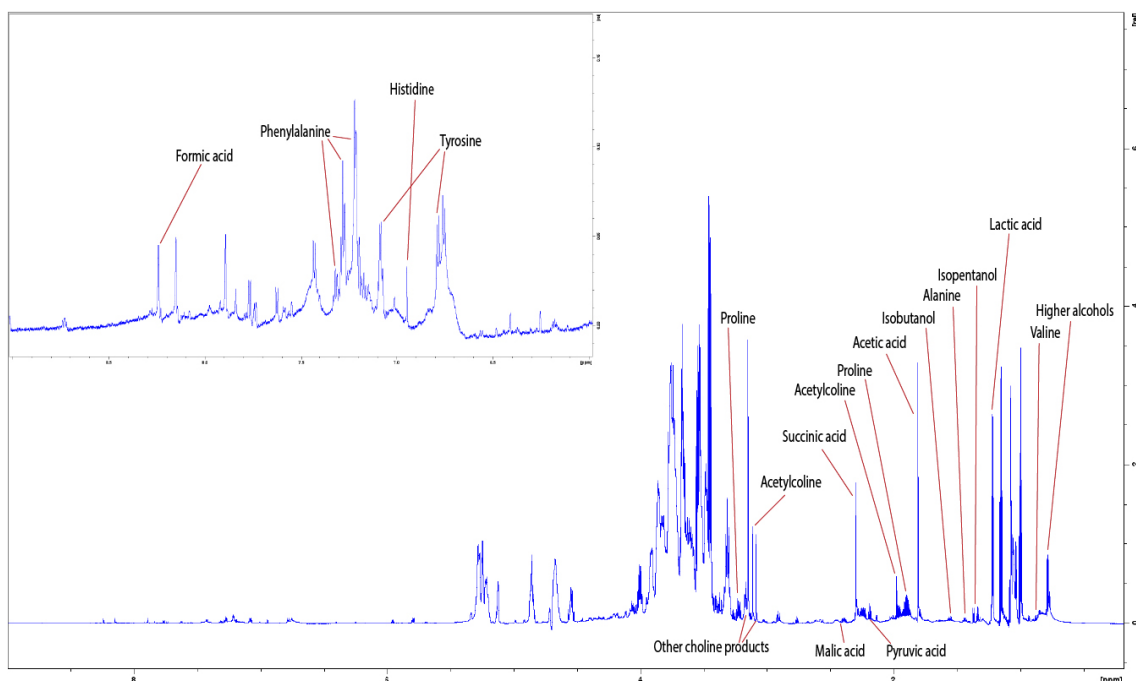


Figure 4.1: **Full spectrum view of a representative raw spectrum with all identified compounds annotated.** Identification supplied by Leesa J. Klau.

---

#### 4.1.4 Categorization study of Austmann beers

The purpose of this study was to determine if the conventional categorizations used for commercial beers would be supported by the clustering and categorization methods used in the statistical analysis of the  $^1\text{H}$  NMR spectra. The following hypothesis was formulated and tested: "Beer samples in the same categories would have more similar NMR profiles than those in different categories, and would therefore cluster together when analyzed." The beers and categorization used can be found in table 3.1, and an overview of the categories and number of samples in each can be found in table 4.2.

Table 4.2: **Table of original style categories of Austmann samples and number of samples in each category.**

Category	Number of samples in category
Dark	5
IPA	7
Light	4
Sour	3

#### Initial analysis of traditional categories

Principal component analysis score plot of the NMR data, figure 4.2, shows a clear separation of the sour beers, while the rest form a line of three loose clusters with members across three of the categories. We do not observe a clear differentiation of the purported categories here.

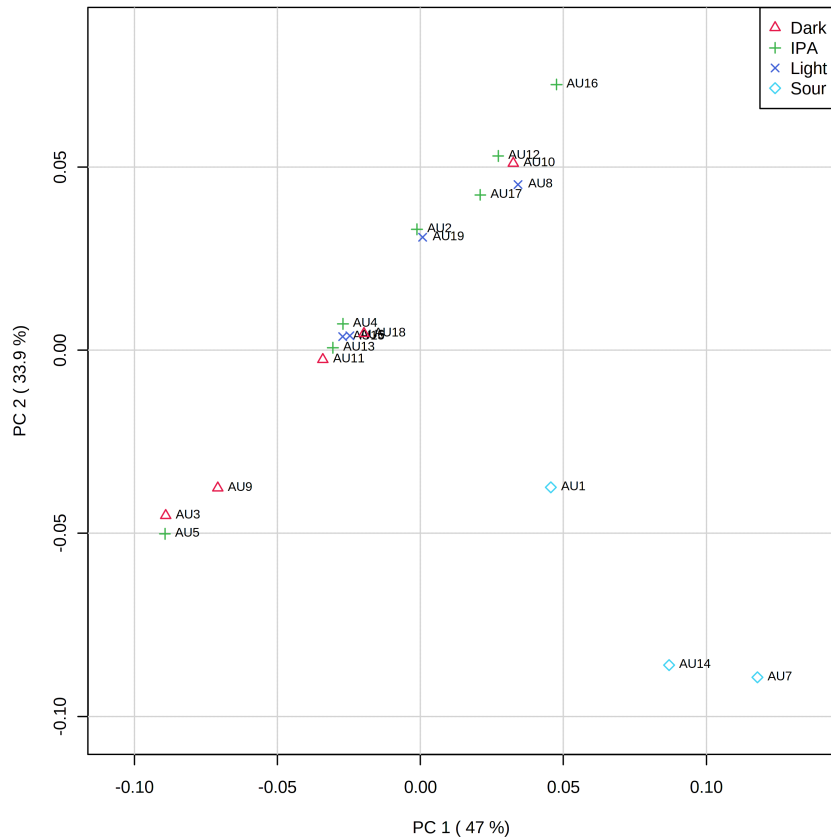


Figure 4.2: **Scores plot of 1st and 2nd principal components from PCA of initial categorization analysis of samples from Austmann.** Each point represents a sample, and the categories are denoted by shape and color, as seen in the legend in the top right. Explained variances of each component displayed in brackets next to the principal component.

The partial least squares discriminatory analysis (PLS-DA) score plot, figure 4.3, again confirms the findings from the PCA, with clear separation of the sour beers. Here we also find a slight separation of the darker beers, while the light and IPA categories are still indistinguishable from each other.

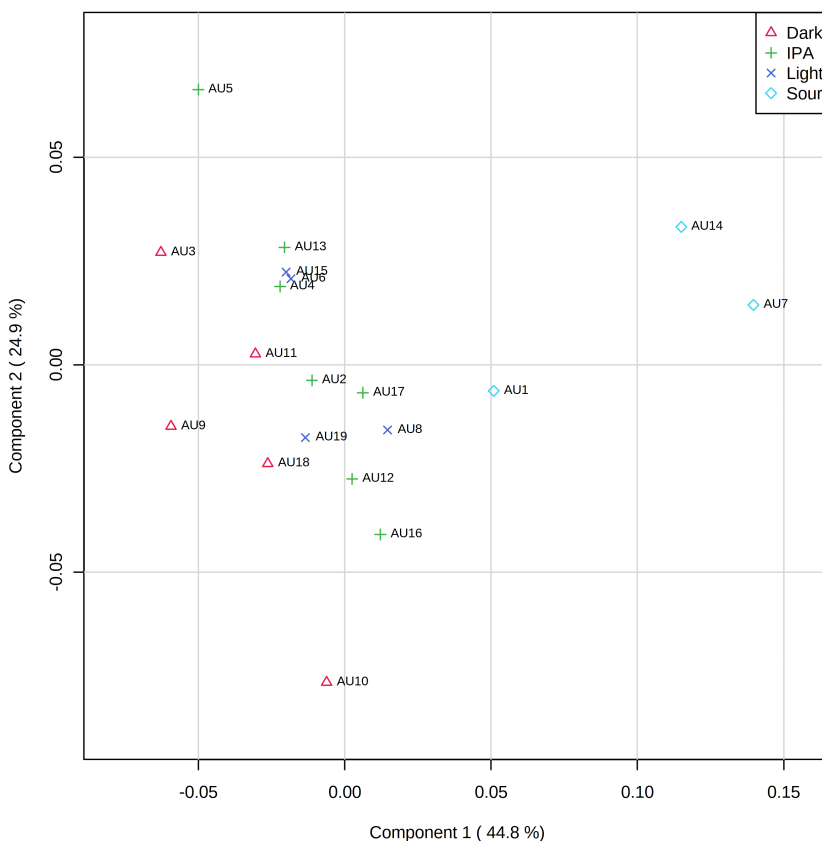


Figure 4.3: **Scores plot of 1st and 2nd components from PLS-DA of initial categorization analysis of samples from Austmann.** Each point represents a sample, and the categories are denoted by shape and color, as seen in the legend in the top right. Explained variances of each component displayed in brackets next to the principal component.

To ensure that these findings are valid and not due to overfitting of this supervised model, cross-validation was performed using a permutation test. The results, displayed in appendix figure F.1, show that the separation distances were indeed significant with an observed p-statistic of  $p < 0.001$ , indicating that the results are not due to overfitting of the model.

In order to see which peaks drive this separation of categories, we can examine the *Variable Importance in Projection* (VIP) scores in the PLS-DA analysis. Figure 4.4 displays the 15 ppm bins with the highest VIP scores, meaning that they explain the most variation responsible for categorical separation. We note that the two bins with the highest VIP scores are at 1.325 and 1.335 ppm. These bins contain the peak associated with lactic acid and are most likely responsible for the separation of the sour beers from the other categories. The rest of the bins have lower VIP scores, and we find those in the region of 3.5 to 4.5 ppm, which is a highly crowded region of the NMR spectrum, making compound identification difficult. Distinction between categories originating in this region might not be reliable, as multiple compounds create overlapping signals. However, as the model is validated using a permutation test, we can be reasonably certain that there is some validity behind these results.



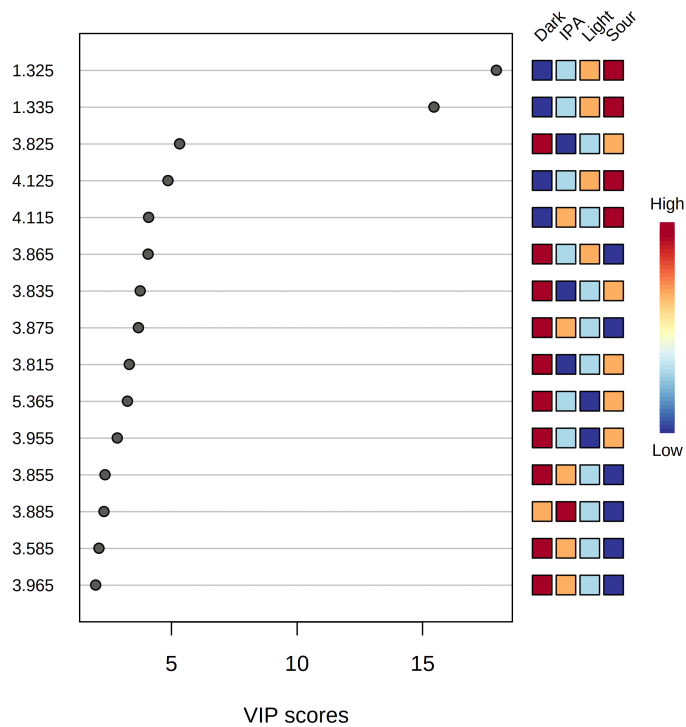


Figure 4.4: **Variable Importance in Projection (VIP) scores for PLS-DA analysis of initial categorization test of samples from Austmann.** X-axis represents VIP score, Y-axis displays the ppm of the bin, or feature.

To thoroughly investigate the data set for any possible clustering, we also performed t-SNE analysis of the data. The t-SNE plot shown in figure 4.5 supports the findings of PCA and PLS-DA, showing four groupings where only the sour beers display unequivocal separation from the rest of the data set.

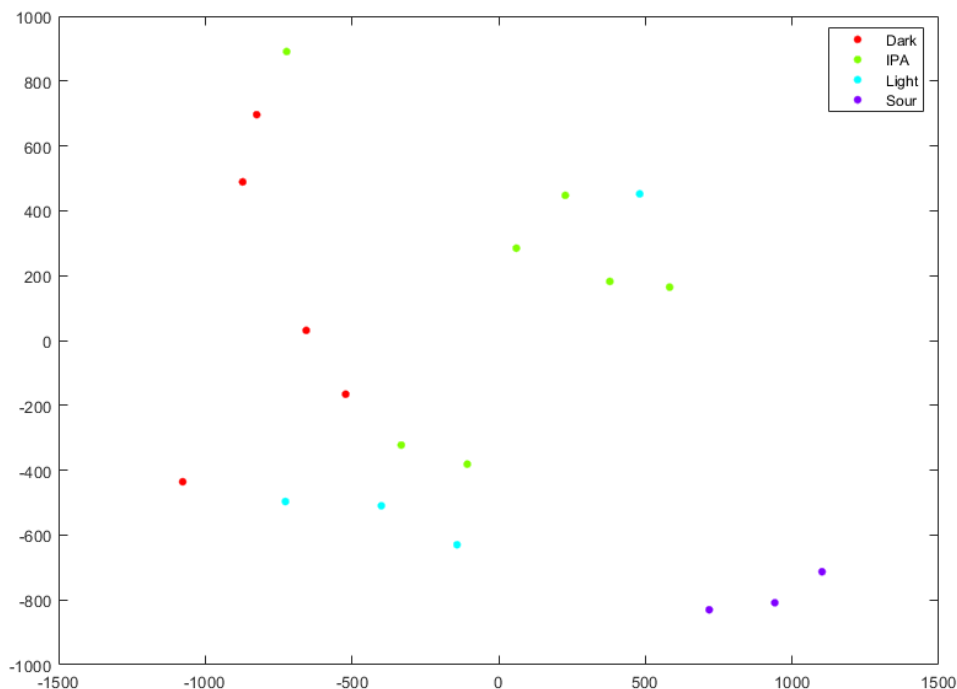


Figure 4.5: **Clustering analysis of Austmann samples by t-SNE.** Data points are colored after conventional categorizations. Perplexity = 5.

The initial analysis based on conventional categories did not show the clustering pattern expected, and thus did not support the assumptions of the hypothesis. Separation were found for the group of sour beers, however, the light and IPA categories did not separate from each other, although this is not surprising as these categories are rather similar in taste profile. It is important to note here that the lack of differences detected during this analysis does not reflect a weakness in the traditional categorization scheme based on the taste of the beer, only that an objective analysis of the NMR spectrum weighs the relative importance of taste components differently than the human pallet. As we did not observe the separation of clusters we expected, we chose to reject the hypothesis and attempt to create more objective categorizations by reclustering the samples using unsupervised methods.

### Unsupervised clustering of Austmann samples

With the aim of exploring a more objective approach to categorization, we applied an unsupervised clustering analysis to create new categories of the 19 samples. The clustering technique used was k-means clustering, and the optimal number of clusters was determined to be four, using the stop-gap method. The four clusters are shown in figure 4.6, and listed along with beer names, sample ID, and traditional categorization in table 4.3.

Cluster	Members
Cluster 1	AU1 , AU7 , AU14
Cluster 2	AU3 , AU9 , AU5
Cluster 3	AU11 , AU18 , AU2 , AU4 , AU13 , AU6 , AU15 , AU19
Cluster 4	AU10 , AU12 , AU16 , AU17 , AU8

Figure 4.6: **Clusters containing the different samples.** Original categorization is displayed in color code: Sour: light blue, dark: red, IPA: green, light: dark blue.

Table 4.3: **Table of samples from Austmann.** Including new clusters assigned according to chemical spectra.

Beer name	Category	Cluster based on chemical composition	Sample ID
Blood And Thunder	Sour		AU1
Peach Piñata	Sour	1	AU14
Hula Hula	Sour		AU7
Coffee Porter	Dark		AU3
Hopbird	IPA	2	AU5
Liquid Lumberjack	Dark		AU9
Northie	Dark		AU11
Onkel I Amerika	IPA		AU13
Pilsner	Light		AU15
Stille natt	Dark	3	AU18
Utpå	Light		AU19
Captain Cold One	IPA		AU2
Hawaiian IPA	IPA		AU4
Hoppy Blonde	Light		AU6
Lynchburg Latte	Dark		AU10
Old School IPA	IPA		AU12
Pine Shoot	IPA	4	AU16
Psycho Tropic	IPA		AU17
Lager 11	Light		AU8

We see that the sour beers appear in their own cluster, however the rest of the clusters contain samples from multiple categories.

The original data set was then relabeled with new clusters instead of the conventional categories, and reanalyzed to determine what aspects of the data were responsible for the new categorizations. PLS-DA and t-SNE were repeated using the clusters established with k-means. The score plot for PLS-DA, figure 4.8, and the t-SNE visualization, figure 4.7 now shows four clear clusters corresponding almost perfectly to the four clusters identified by k-means. Permutation test (appendix figure F.2) validates the fit of the PLS-DA model ( $p = 0.016$ ).

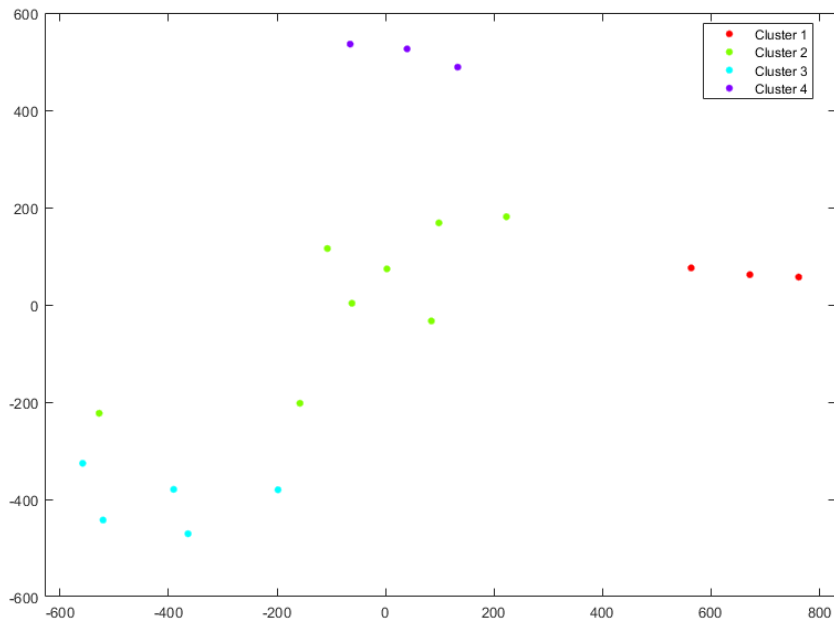


Figure 4.7: **Clustering analysis of re-clustered Austmann samples by t-SNE.** Data points are colored according to the new clusters found by k-means. Perplexity = 4.

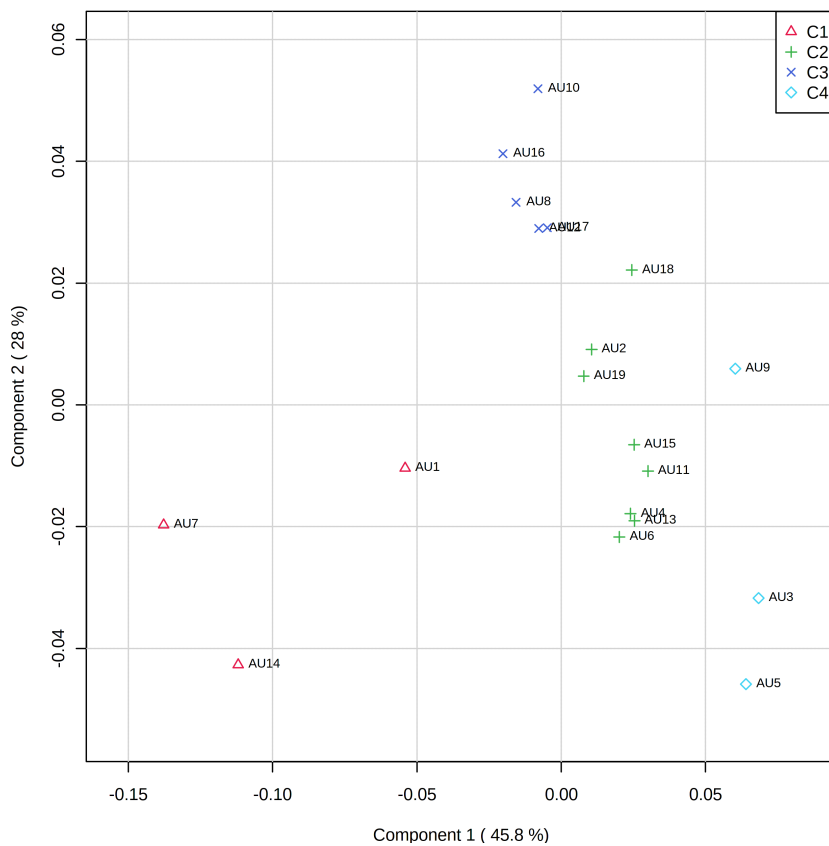


Figure 4.8: **Scores plot of 1st and 2nd components from PLS-DA of re-clustered categorization analysis of samples from Austmann.** Each point represents a sample, and the categories are denoted by shape and color, as seen in the legend in the top right. Explained variances of each component displayed in brackets next to the principal component.

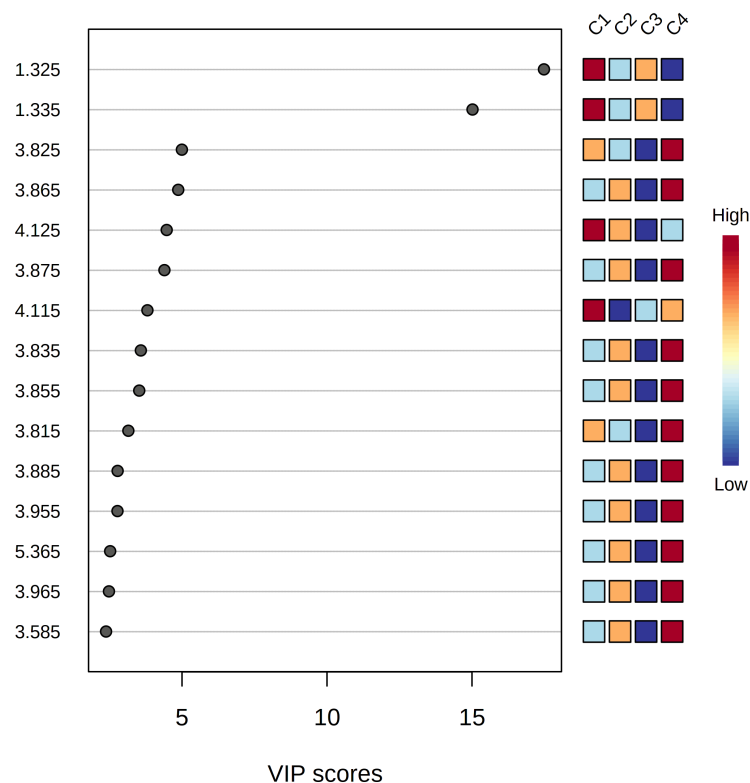


Figure 4.9: **Important features for PLS-DA analysis of re-clustered categorization test of samples from Austmann.** X-axis represents importance score, Y-axis displays the ppm of the bin, or feature.

We can now evaluate the VIP scores to determine which areas of the spectra drive the separation of the new clusters. The top 15 bins with the highest VIP scores are shown in figure 4.9. These bins are mostly the same as the ones found in the initial categorization.

Visualizing the loadings of the PLS-DA along with the collection of spectra analyzed, figure 4.10, we see that the most important peak for the separation in the first PC represents lactic acid, dominated by the cluster of sour beers (red lines) in the spectrum plot. This was expected from the PLS-DA score plot where the same cluster of sour beers (red lines) stand out from the rest. Most of the weight in the loadings plot is in the heavily crowded 3.5-4.5 region, where compound identifications are difficult. More detailed plots of the spectra and loadings are shown in appendix A, illustrating that the different clusters show slightly elevated levels in different peaks, but that most stand-out areas of difference are dominated by red lines, representing the category of sour beers.

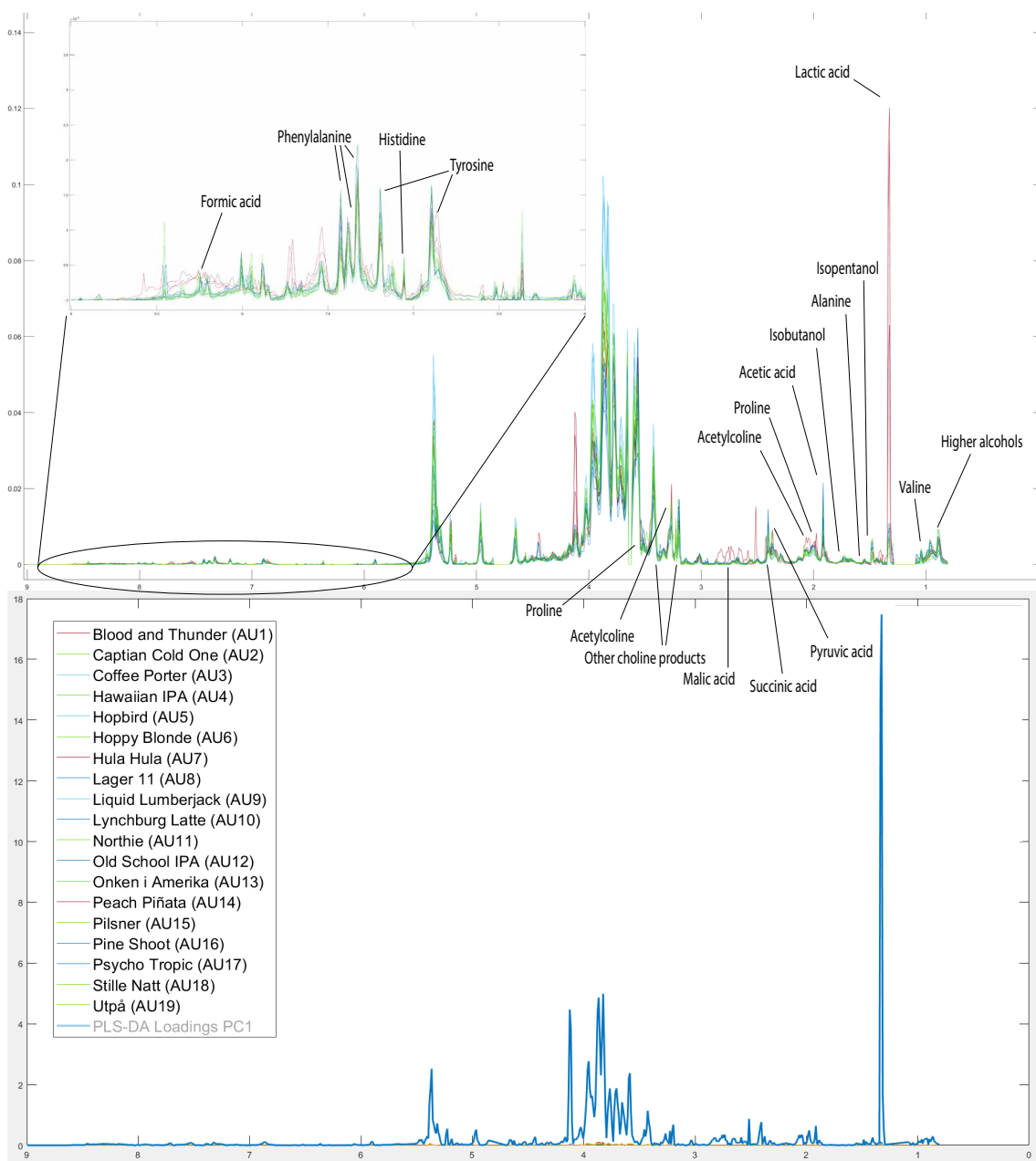


Figure 4.10: **Full spectra and loadings for reclustered Austmann samples.** Upper field shows a plot of all binned NMR spectra for the reclustered Austmann samples overlaid and colored by cluster (cluster 1: red, cluster 2: light blue, cluster 3: green, cluster 4: deep blue). Some identifiable compounds are annotated. Lower field shows PLS-DA loadings for the first PC, taller peaks indicate more weight in PLS-DA. X-axis: PPM, Y-axis: intensity of signal / weight in PLS-DA.

### Evaluation of categorization study

As we can see from the results above, an objective approach to classification does not support the conventional categories of these beers. The conventional categories are based on the taste and ingredients used in the brewing of these samples, and the difference between the categories suggested by the analysis of NMR spectra and the categories originally used may be due to differences in how some compounds are perceived while consuming the beer compared to their prevalence in the sample detected by NMR. Some compounds may be

---

essential for taste, but be present in only small amounts in the beer, and therefore may not be highlighted when analyzing the NMR spectra.

#### 4.1.5 Pitch rate study

The pitch rate study was a simple proof-of-concept study designed specifically for testing whether our analysis method could discern a difference between identically brewed beers with varying pitch rates (amount of yeast used for inoculation), the hypothesis being: "Differences in the pitching rate of yeast in samples create differences detectable by both NMR, and sensory analysis. These differences will cause samples with the same pitch rate to be grouped together under various analysis." Pitching rates and categories can be found in table 3.2. Three different approaches were used to examine differences in the three pitching rates; sensory analysis, multivariate statistical analysis of NMR spectra, and ethanol percentage calculation.

##### Sensory analysis

The sensory analysis was done only to determine if a difference could be found between samples, and if a trend could be identified. The tests were intended to collect qualitative sensory data to help inform database decisions, and were conducted in an informal setting. Sixteen individuals in 6 independent groups were asked to describe the taste of a sample from each pitch rate without prior knowledge of the difference or experiment setup. The participants were also asked to identify any trends in the samples and sort them accordingly.

Invariably, each independent group sorted the samples in the same order and attributed similar taste descriptors to the samples. A graphical representation of the descriptors and trends are shown in figure 4.11.

The qualitative sensory data, while limited, appear to indicate that there is a perceptible difference. However, to be able to investigate these relationships further, a rigorous sensory experiment is required. Based on these findings, we expect that differences should also be detectable using NMR spectroscopy.

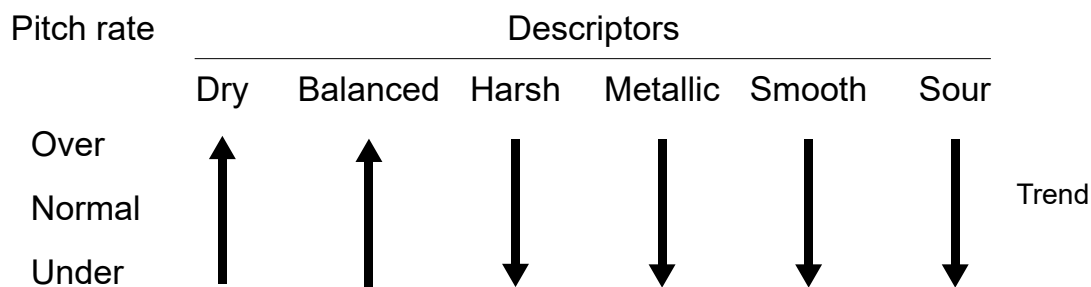


Figure 4.11: **Illustration of trends recorded for sensory analysis of pitching study.** Direction of trends are illustrated by arrows, indicating that the majority of participants described the intensity of the given characteristic for each category in the direction of the arrow. E.g: Overpitched samples were generally described as less sour than underpitched samples.

---

## Multivariate analysis

Examining the multivariate analysis of the pitch rate study samples, there is little to no clear separation that would reflect the findings in the initial sensory test. Only PCA gives some hints at a clear separation, and the results are ambiguous.

Examining the PCA score plot of the 1st and 3rd principal components (as these gave the best groupings), as well as the t-SNE plot of the data set, figures 4.12 and 4.13, we can visualize some separation in the three pitching rates. The underpitched sample in the second parallel (WU2) appears as an outlier in both plots, however the other samples fall somewhat into three clusters. Based on this, we would expect a supervised model like PLS-DA to be able to identify the areas driving the separation. However, the supervised model could not be used in this scenario as the PLS-DA permutation test heavily indicated overfitting, as seen in appendix figure F.3 ( $p = 0.168$ ). Therefore, we can only interpret the PCA results in this case.

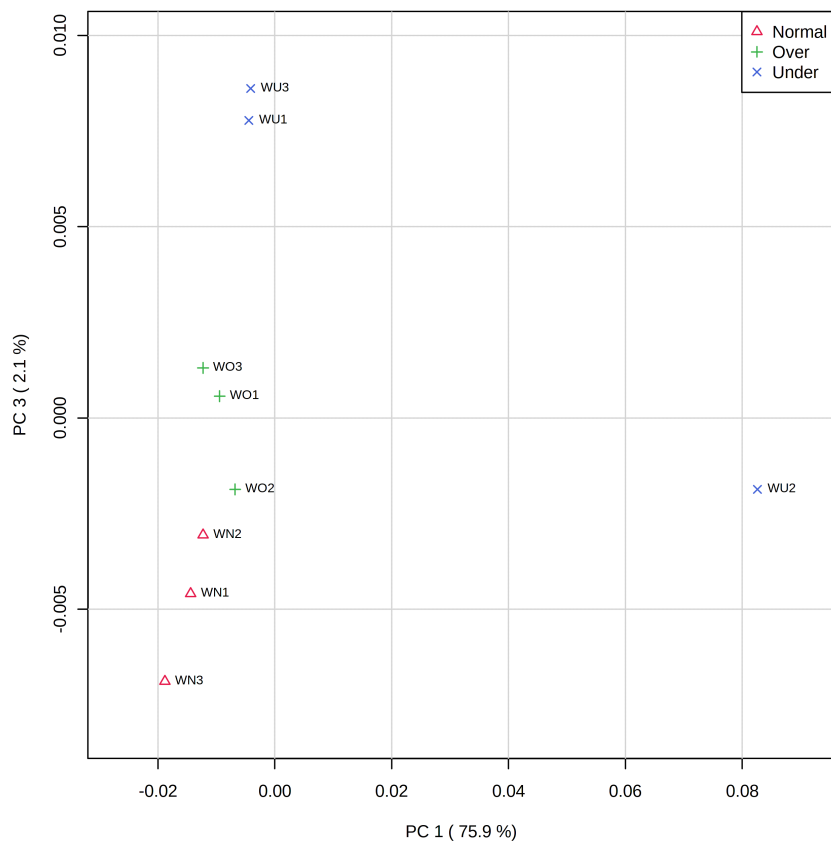


Figure 4.12: **PCA scores plot of 1st and 3rd principal components of differential analysis of samples from pitch rate study.** Each point represents a sample, and the categories are denoted by shape and color, as seen in the legend in the top right. Explained variances of each component displayed in brackets next to the principal component.



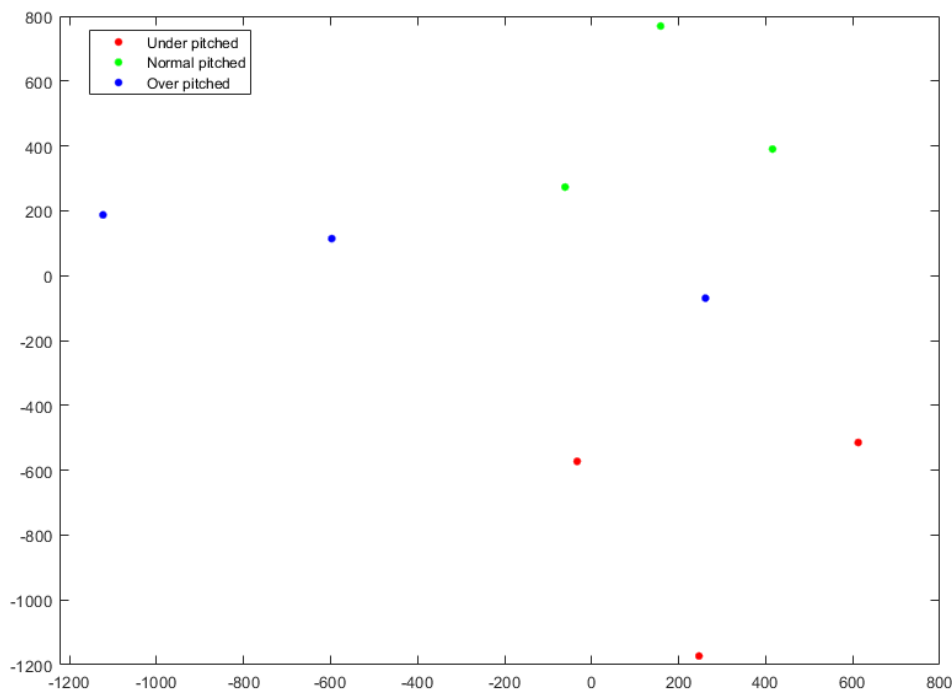


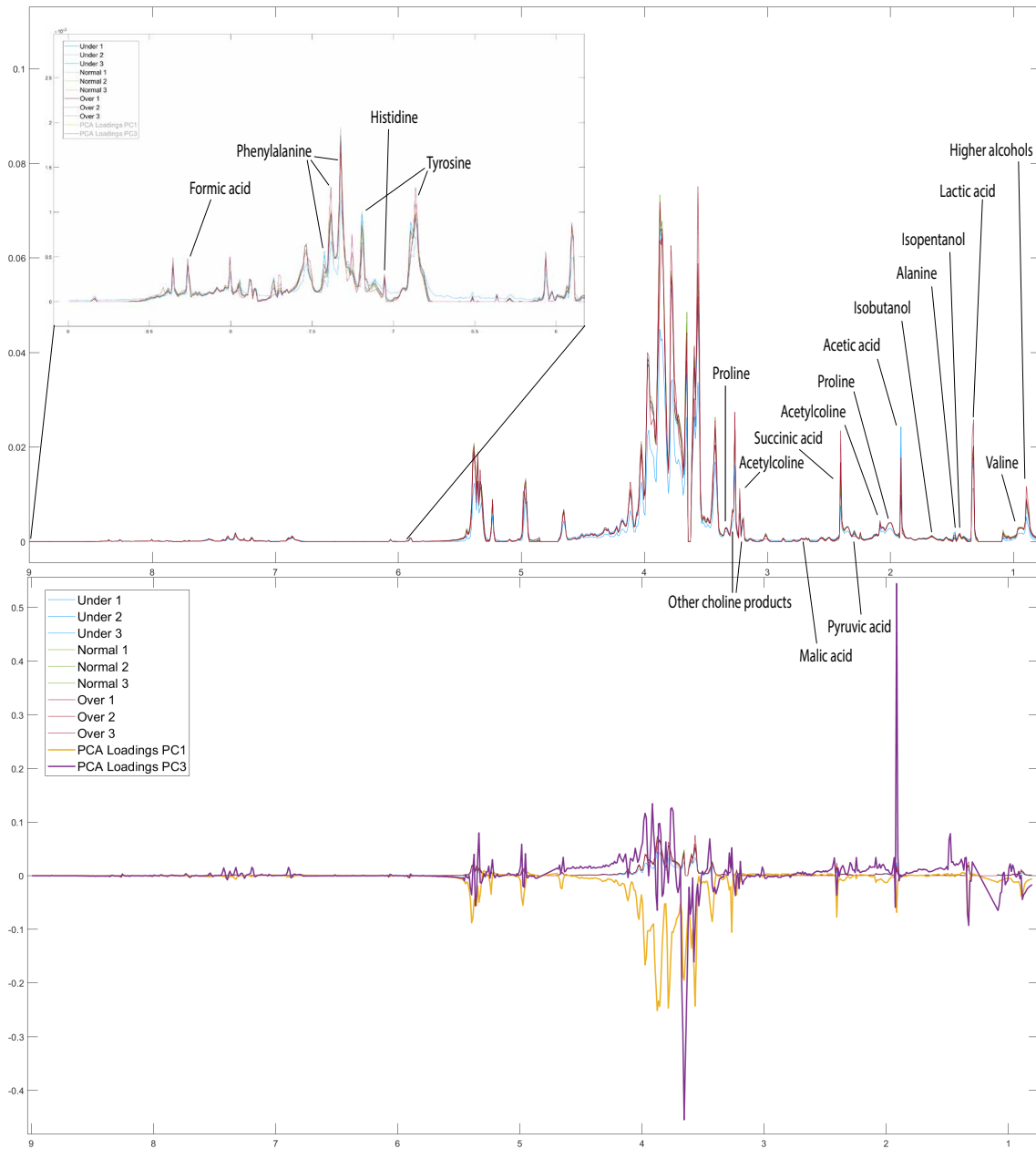
Figure 4.13: **Clustering analysis of pitch rate samples by t-SNE.** Data points are colored according pitching rate used. Perplexity = 4.

Evaluating the overlay of the spectra and the PCA loadings for the first and the third principal component, figure 4.14, we can identify a few notable compounds that are heavily weighted in the PCA loadings. More detailed spectra and loadings plots can be found in appendix B. Looking at the loadings for PC1 and PC3 (yellow and purple lines in the lower field of figure 4.14), we see that most of the highly contributing areas are in the crowded region between 3.5 and 4.5 ppm. However, we have a few identifiable peaks that display a difference between the pitch rates. Examples of these are:

- Lactic acid (ca. 1.32 ppm), which imparts a sour taste to beer [56], this peak is highest in the overpitched samples and lowest in the underpitched, indicating a positive association between pitching rate and lactic acid production in this study.
- Acetic acid (ca. 1.91 ppm) imparts a sour, vinegar-like taste that is generally considered undesirable in beer [57]. This compound seems to be present in higher levels in under-pitched samples, and lower in over pitched ones, indicating a negative correlation between pitching rate and acetic acid production.
- Succinic acid (ca. 2.41 ppm), which is not associated with a specific taste, but is considered an undesirable compound that may reduce the *drinkability* of a beer due to its role as a potential stimulant of gastric acid secretion [58]. In this study, succinic acid is clearly higher in overpitched samples and lower in underpitched ones, indicating a positive correlation between pitch rate and succinic acid prevalence.

Tying these compounds back to the trends in taste found during the initial sensory evaluation, there are no obvious parallels that stand out between the identified compounds and

the sensory trends, but we can note that underpitched beers are described as more sour tasting than the others although lactic acid is less prevalent in the underpitched samples, indicating that other taste compounds like acetic acid might be the cause of the sour taste in these beers. These preliminary results may be of interest for more detailed studies.



**Figure 4.14: Full spectra and loadings for pitch rate samples.** Upper field shows a plot of all binned NMR spectra for all pitch rate samples overlaid and colored by pitch rate (underpitched: blue, normal pitch: green, overpitched: red). Some identifiable compounds are annotated. Lower field shows PCA loadings for the first PC, more extreme peaks indicate more weight in the separation found in PCA. X-axis: PPM, Y-axis: intensity of signal / weight in PCA.

Using a dendrogram to visualize the hierarchical clustering to examine the similarity of the samples, shown in figure 4.15, we find that there is no clear trend here either. This somewhat weakens the assumption made in the hypothesis that we may find significant differences between the treatment categories using multivariate analysis.

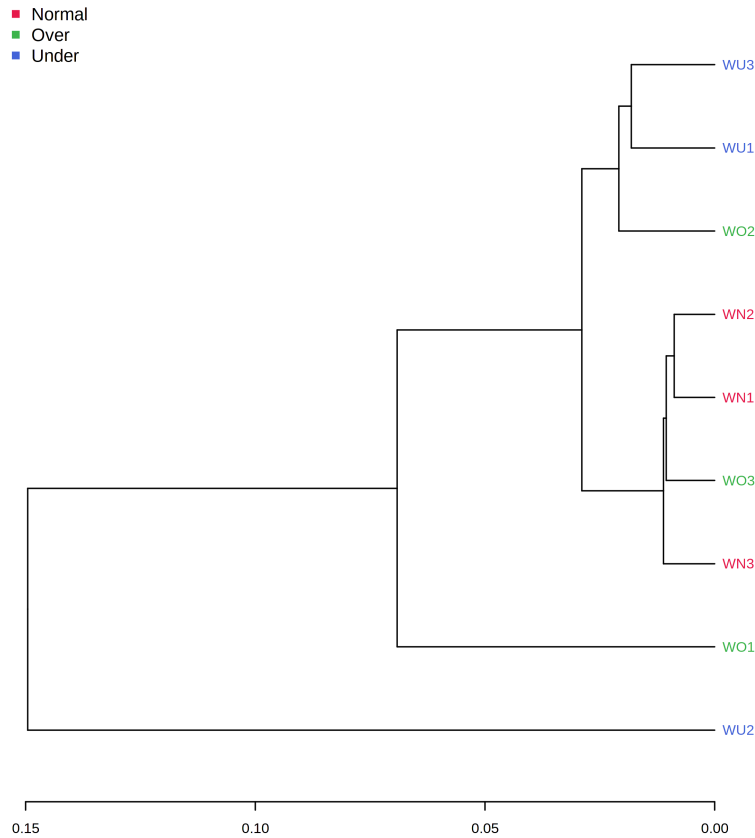


Figure 4.15: **Dendrogram showing clustering of samples from pitch rate study using Euclidean distance and Ward clustering algorithm.**

We receive somewhat unclear results from the multivariate analysis, however, we can still identify some trends and separation between the different pitching rates, or treatments. Therefore, we cannot reject the hypothesis based on this, although the differences are not as noticeable as we had expected. We move on to evaluate ethanol content, as this is one of the more prevalent components present but is not included in the spectra used for multivariate analysis.

### Ethanol content

Another interesting aspect of the different pitching rates is the ethanol content, as this will significantly affect the taste profile of the beer. Ethanol content is not explored in the normal analysis of the spectra, as that signal is suppressed in the spectra used for the main analysis in order to evaluate the presence of less prevalent taste compounds. Spectra without ethanol suppression were used to estimate the alcohol concentration for the three different pitching rates.

As we can see from figure 4.16, there is a trend in the alcohol content produced in the three different groups, indicating that the alcohol percentage of the brew decreases as the pitch rate increases. This trend might explain the findings from the preliminary sensory experiment, as some of the descriptors assigned to the underpitched samples could be associated with a higher alcohol content. Note that the difference is not statistically

significant, as the large variation within the groups and the small number of samples lead to great uncertainty in the results. Summary of ANOVA shown in figure 4.17.

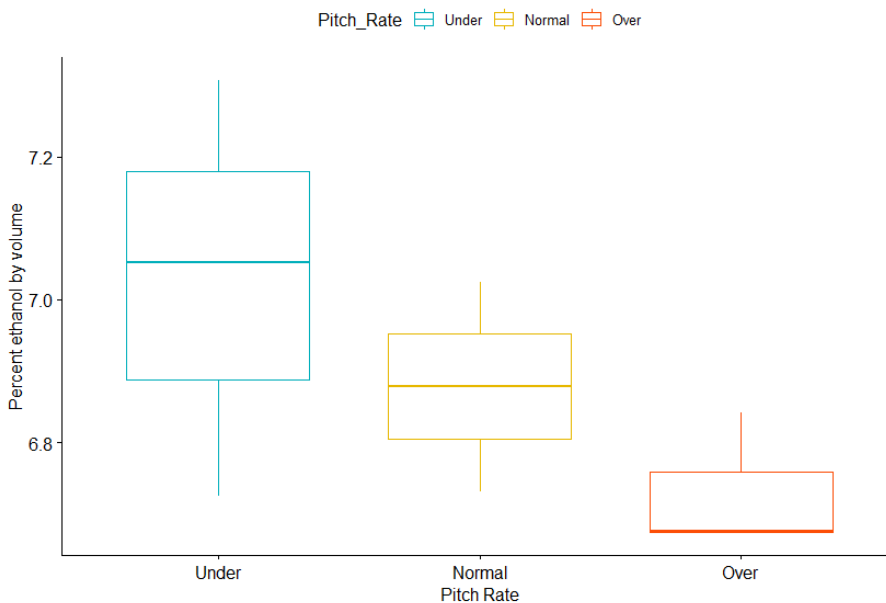


Figure 4.16: **Boxplot of ethanol concentration in the three different pitching rates.**

```

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.8784    0.1134  60.666 1.35e-09 ***
Pitch_RateOver -0.1479    0.1603  -0.923   0.392
Pitch_RateUnder 0.1492    0.1603   0.931   0.388
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1964 on 6 degrees of freedom
Multiple R-squared:  0.364,    Adjusted R-squared:  0.152
F-statistic: 1.717 on 2 and 6 DF,  p-value: 0.2572

```

Figure 4.17: **Summary of ANOVA of differences in ethanol content between pitching rates.**

### Evaluation of the hypothesis for the pitch rate study

Based on the findings above, the hypothesis still stands, as we did detect some differences. However, the size of the data set is insufficient for appropriate statistical analyses. One interesting aspect of this study is the clear results found by the sensory analysis in relation to the less clear results from the analysis of NMR spectra. This indicates that the compounds detected by the pallet and olfactory senses during tasting are weighted quite differently than the prevalence of these compounds detected by NMR. These discoveries should be investigated further, as more comprehensive studies are required for truly informative results.

---

#### 4.1.6 Aging study of algae-flavoured beer

Analysis of samples from the study of aging in algae-flavoured beer focused on exploring the differences in the same set of 30 beers sampled at two different time points. The aim was to determine if the proposed analysis method could determine any categorical differences caused by aging in the 30 samples, based on the hypothesis: "Aging will cause changes to the chemical composition of the samples, therefore samples in the same aging category will have more similar NMR spectra and will cluster together during analysis."

##### Full data set analysis

Firstly, the full data set were analyzed in order to explore global trends. Examining t-tests between the different groups for each feature, figure 4.18, there are many peaks where we see a significant difference in the two groups, indicating clear differences might be a difference in the different aging times.

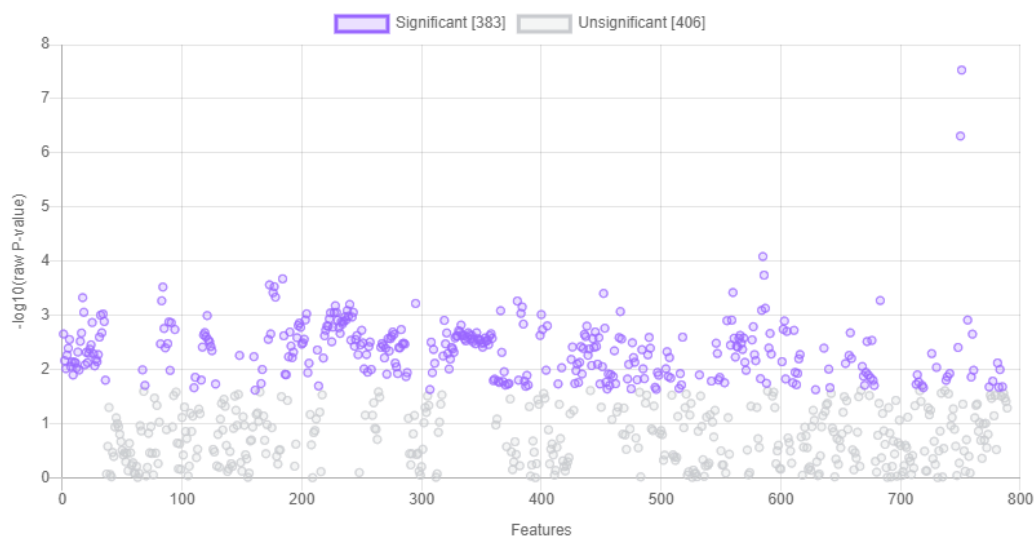
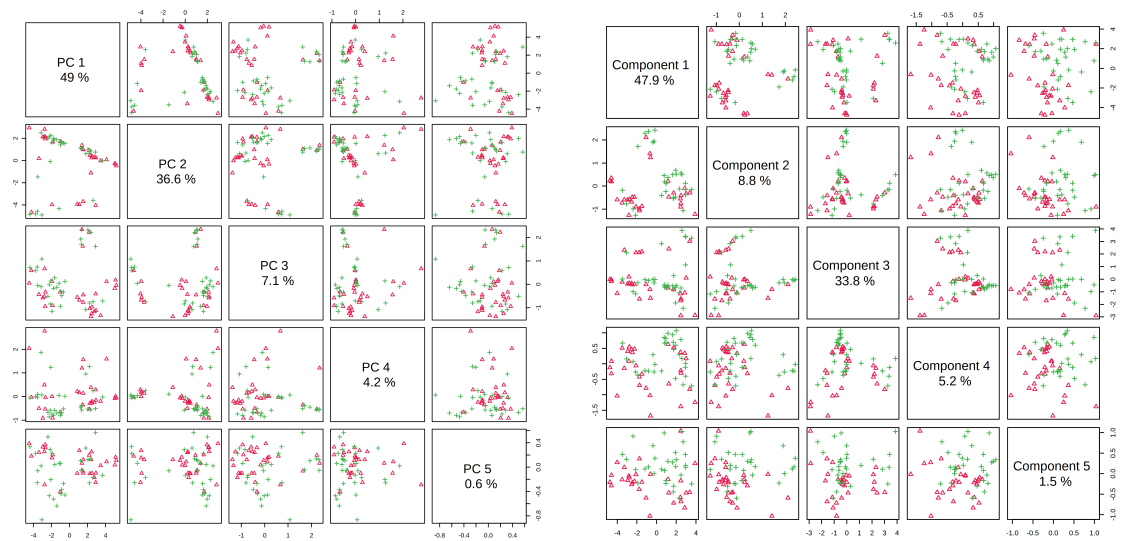


Figure 4.18: **Paired t-test of difference in two age categories in the full algae-beer data set.** Unit variance scaling was performed prior to t-test.

Examining the data using PCA, figure 4.19a, PLS-DA, figure 4.19b, and t-SNE, figure 4.20 we cannot discern a clear trend that indicates a difference between the aged and non-aged sample groups. We therefore separated the data set into the four different types of beer in the study and examined each type on its own.



(a) Overview of the first five principal components of the PCA analysis of all algae-beer samples. (b) Overview of the first five principal components of the PLS-DA analysis of all algae-beer samples.

Figure 4.19: **Overview of the first five principal components of the PCA and PLS-DA analyses of all algae-beer samples.** Each point represents a sample, red triangles represent non-aged samples, green crosses represent aged samples. Percentage of variance explained by each component is displayed below the component.

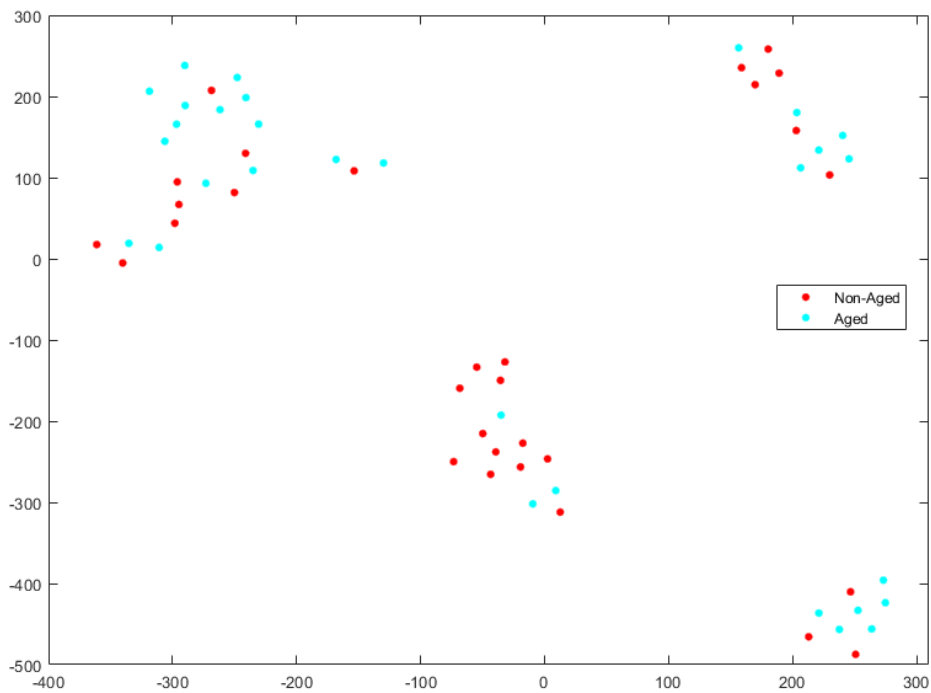


Figure 4.20: **Clustering analysis of entire sample set from aging study by t-SNE.** Data points are colored according to aging category. Perplexity = 8.

---

## Separate analysis

When evaluating each beer separately, only the brown ale category showed any signs of clear separation between the two groups. Examining the PCA of the brown ale samples, figure 4.21, we find that the 1st principal component divides the data in two main clusters with a clear distinction between sample groups. The same pattern is observed in the t-SNE plot, figure 4.22. This might be an indication of an actual difference between aging groups. This is also reflected in the t-test performed on the two groups, figure 4.23 resulting in 532 features with a significant difference between the aged and non-aged samples. The same trend is visible in the dendrogram, shown in figure 4.24, where the two differently aged groups separate quite well into two main clusters.

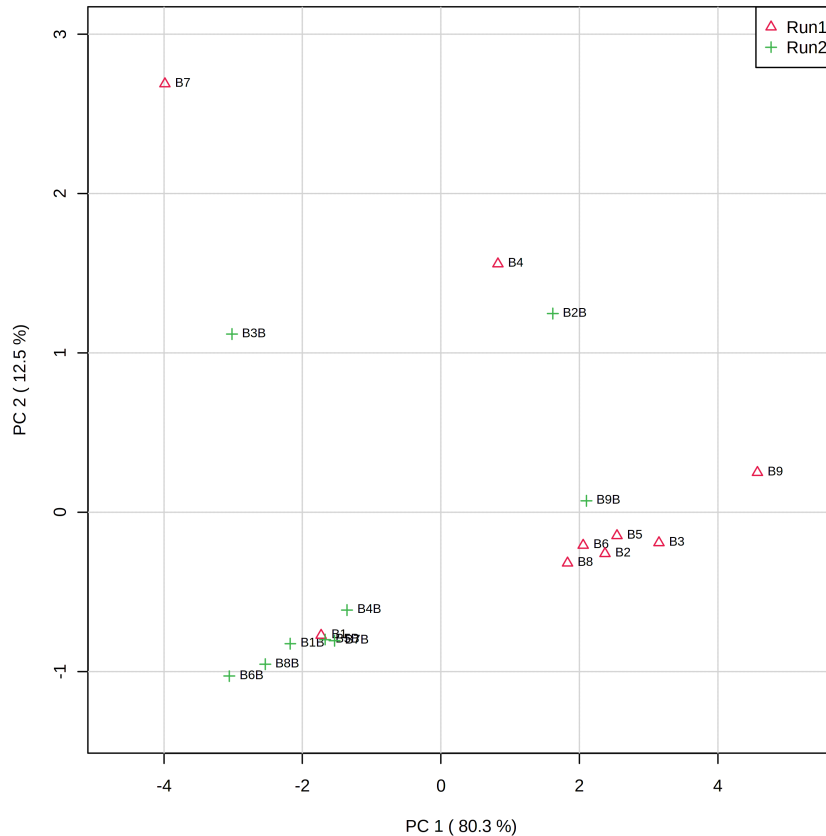


Figure 4.21: PCA plot of 1st and 2nd component of samples from the Brown Ale group in the aging study.

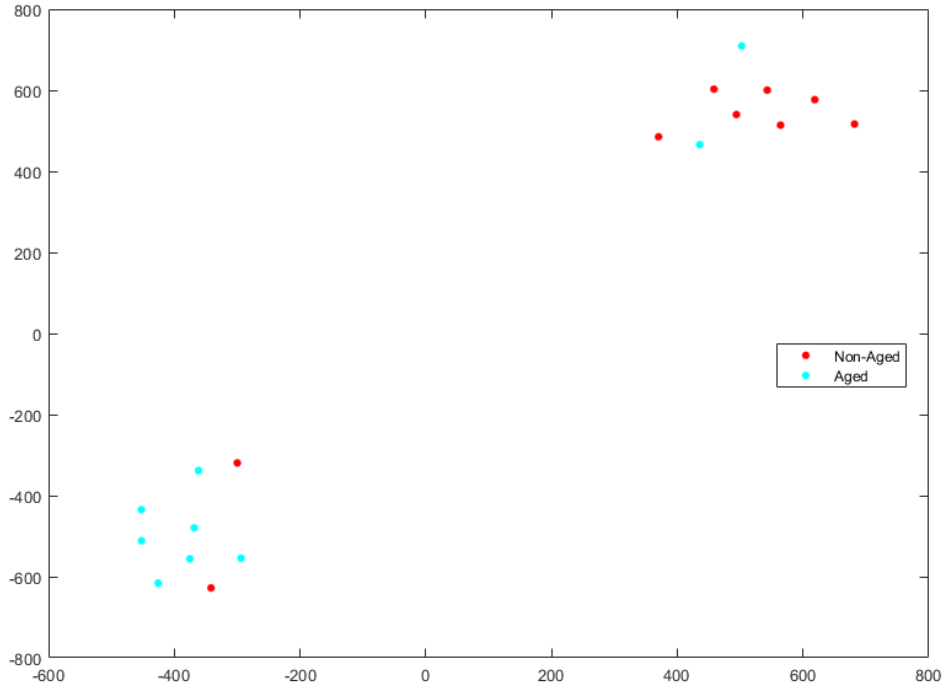


Figure 4.22: **Clustering analysis of Brown Ale samples from aging study by t-SNE.** Data points are colored according to aging category. Perplexity = 6.

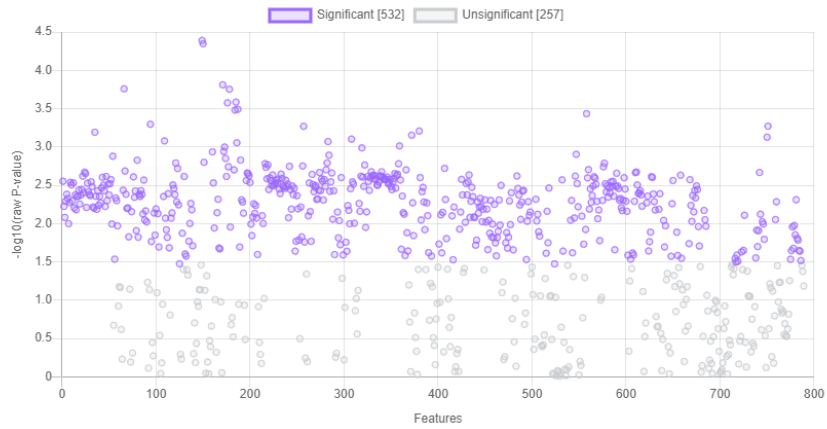


Figure 4.23: **T-test of difference in aged and non-aged groups for each feature in brown ale samples from the aging study.** Unit variance scaling performed on data set prior to t-test.



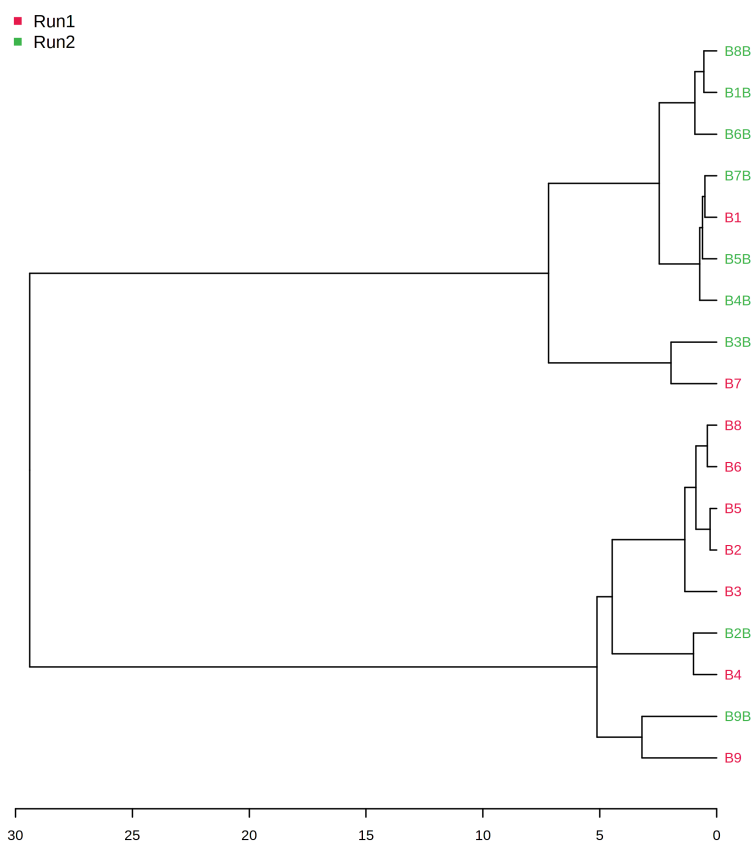


Figure 4.24: **Dendrogram showing clustering of samples from the brown ale group of the aging study.** Clustering using Euclidean distance and Ward clustering algorithm.

Taking a more detailed look at the NMR-spectra of the brown ales along with the loadings from the PCA in figure 4.25, we see that most of the points of separation are in the region from 3.5 to 4-5 ppm, where compound identification is nearly impossible. However, we notice four identifiable peaks that stand out in the loadings in the 0.8 to 2.5 ppm region of the spectrum (more detailed plots can be found in appendix C). These are the higher alcohols, lactic acid, acetic acid, and succinic acid:

- Higher alcohols (ca. 0.86 - 0.91 ppm) contribute to the flavour profile by intensifying alcohol perception and resulting in a warm mouthfeel [59]. In this case, the aged samples look to have higher concentrations of these compounds in general.
- Lactic acid (ca. 1.32 ppm), which imparts a sour taste to beer [56], this peak is highest in the non-aged samples, indicating a lessening of the lactic acid content over time.
- Acetic acid (ca. 1.91 ppm) imparts a sour, vinegar-like taste that is generally considered undesirable in beer [57]. This compound seems to be present in higher levels in non-aged samples, however both categories have samples both low and high in this compound, making categorical differences speculative at best.
- Succinic acid (ca. 2.41 ppm), which is not associated with a specific taste, but is considered an undesirable compound that may reduce the *drinkability* of a beer

due to its role as a potential stimulant of gastric acid secretion [58]. In this study, succinic acid content seems to be similar between the two categories, even though it is weighted in the loadings. This looks to be caused by some of the non-aged samples seemingly lacking a peak for the compound (same for pyruvic acid (ca. 2.37 ppm)), which may be caused by alignment errors in the data processing step, or contamination of the samples.

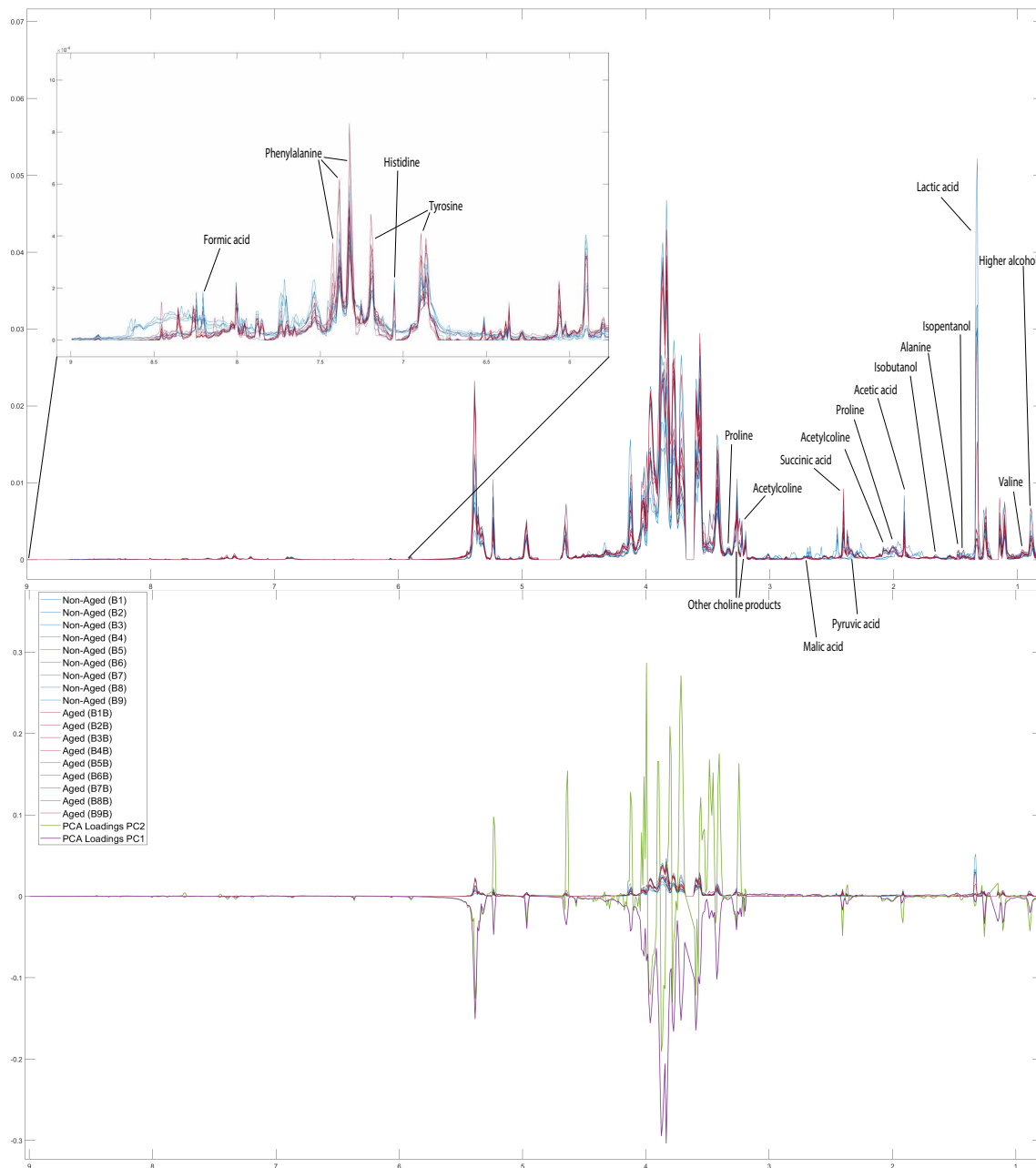


Figure 4.25: **Full spectra and loadings for algae flavoured samples.** Upper field shows a plot of all binned NMR spectra for all algae brown ale samples overlaid and colored by aging category (non-aged: blue, aged: red). Some identifiable compounds are annotated. Lower field shows PCA loadings for the first PC, more extreme peaks indicate more weight in the separation found in PCA. X-axis: PPM, Y-axis: intensity of signal / weight in PCA.

Analysis of the other beer types yielded no noteworthy results. This may be caused by differences in the composition of the beers reacting differently to aging, but as the analysis

was done on samples with a large number of variable factors, we cannot draw any definite conclusions. The results found above show that we did indeed detect differences between aging categories for one of the beer types, and that these differences caused most of the samples in the two categories to form two distinct clusters. Based on this, the hypothesis still stands, however only one of the four beer types showed a clear separation between age groups, therefore the results cannot be generalized for other types of beer.

## 4.2 Ontology

The beer ontology was created with the aim of allowing classification of all relevant samples the project might encounter, as well as the ingredients and other aspects of the samples. The aim was to allow as much of the information as possible to be represented as objects and relations, instead of just metadata in text fields.

### 4.2.1 General design

The ontology was created to categorize samples with a reasonable level of detail. To this end, the decision was made that the instances of the ontology were to be individual samples, sorted by the style of the beer the sample was based on. The reasoning behind this was that using object property assertions to create structures that would contain detailed recipe information would be more flexible than using class axioms in a similar way.

The final design is a structure where each sample is recorded as an instance, with each instance connected to both a style category describing the ingredients and the area of origin, as well as being connected to a project detailing the experimental or analytical strategy for the sample. Each sample is also connected to the different ingredients used in the brewing process, as well as to the producer of the sample. These connections are annotated as object properties, describing the nature of the link between the sample and the ingredient. In this way, a detailed network is created, containing the pertinent information on the sample in question. An example illustrating part of the structure within the pitch rate study can be seen in figure 4.26.

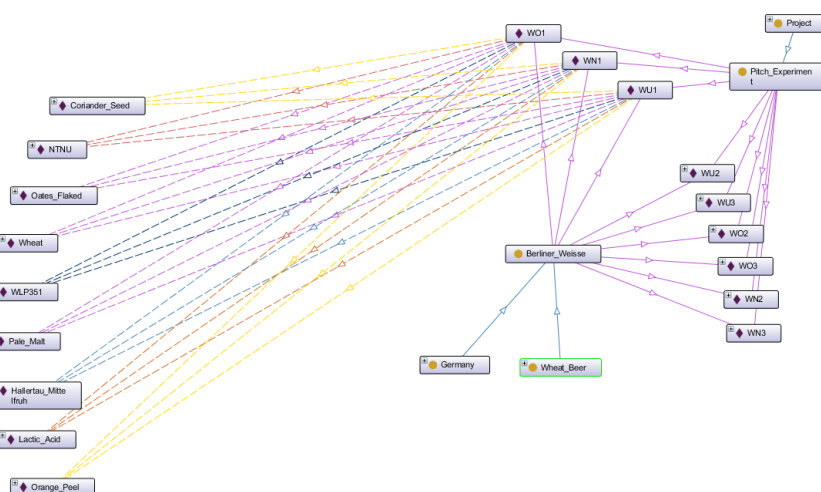


Figure 4.26: **Example of partial ontology structure for samples in the pitch rate study in network view for illustrative purposes.** Full lines represent parent/child relationships, dashed lines represent custom object properties. Figure created in Protege.

---

## 4.2.2 Parts of the ontology

### Beer types

The classification of beer types were approached from multiple different angles: by ingredient and taste profile, region of origin, producer, and project. Firstly, the different beer types represented in the ontology (132 at the time of writing) were classified in two main categories: ale and lager, depending on the yeast type (typically) used in the beer. A separate category for specialty beer styles that may use either lager or ale yeast were also added. Thereafter, each subcategory were divided into increasingly specific classes depending on the ingredients, brewing process or taste profile of the style, an example of some sub classes of ales are shown in figures 4.27 and 4.28. In some special cases, the class was shared between both ale and lager, as the style can be brewed with both yeast types. Such ambiguous definitions were avoided as often as possible.

Since beer is a deeply cultural product that traditionally were heavily influenced by the area it was brewed in, the decision was made to add a layer of categorization by region of origin to the list of beer styles. This means that taste-style were cross-referenced with a list of countries (or sometimes specific areas within countries) to create a more complete ontology that provides information on both forms of classification. Each sample was also categorized by what study and project it belonged to, as well as who produced it. This model of multiple angles of categorization for each sample allows for a more detailed and comprehensive ontology, as well as offering multiple approaches to categorization, making the ontology intuitive for different users with varying backgrounds and conventions for beer categorization. It also allows for a variety of category analyses down the line when chemical spectra are available for representative samples of different categories.

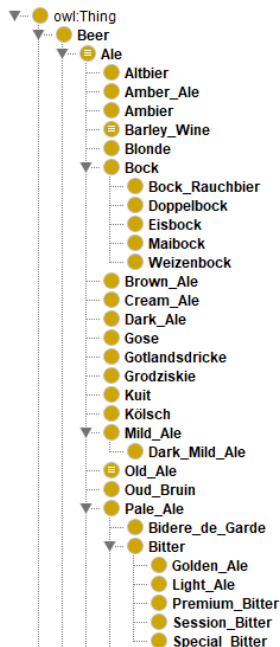


Figure 4.27: **Example of a few sub classes of the ale category in the ontology structure.** Figure created in Protege.

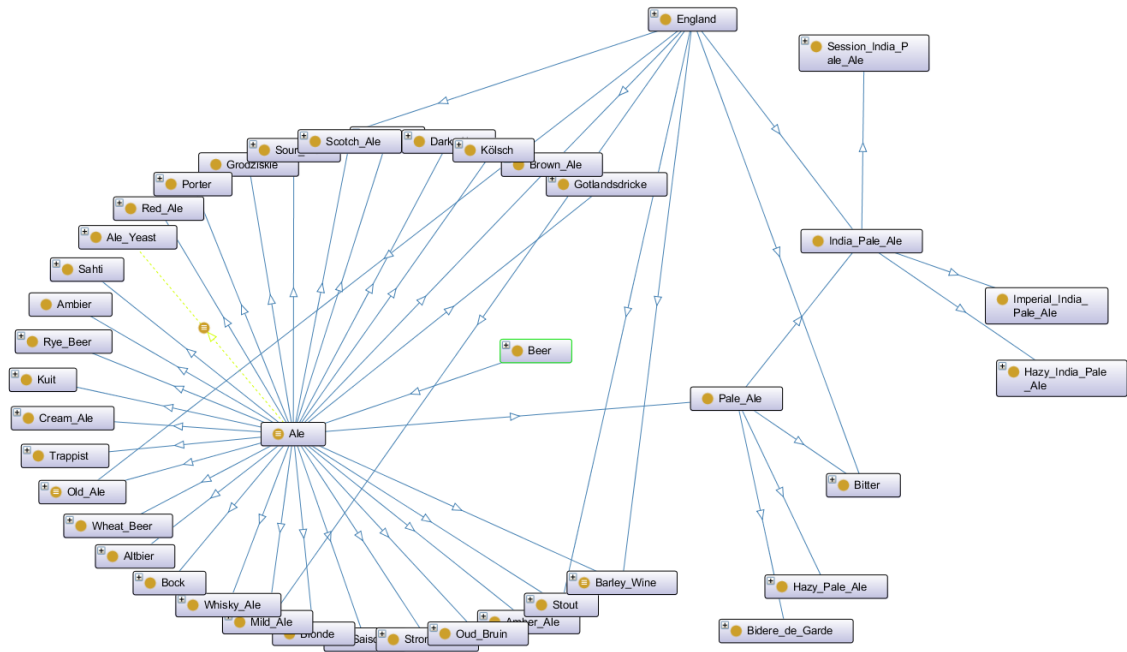


Figure 4.28: **Example of partial ontology class structure for the ale category.** Full lines represent parent/child relationships, dashed lines represent custom object properties. Figure created in Protege.

## Ingredients

Each ingredient that is used in the brewing process is also categorized in its own hierarchical structure, which is separated into seven general categories: *fermenting organism* (yeast and other fermenting organisms); *fermentable* (malt and other grains); *flavour additive* (spices, fruits, etc); *hops*; *salts*; *secondary additives* (flocculation agents, yeast nutrients, etc); and *water*.

In contrast to the beer styles, the instances of ingredients are types of product, not individuals of that product. An example of the "ingredient" category is shown in figure 4.29 This decision was made as we do not expect there to be enough differences in individual products of the same type from the same producer to warrant recording. If cases should arise where we expect specifics like this to be relevant, those can be recorded as metadata.

Ingredients were, as with beer styles, also categorized by area of origin, as well as by producer.

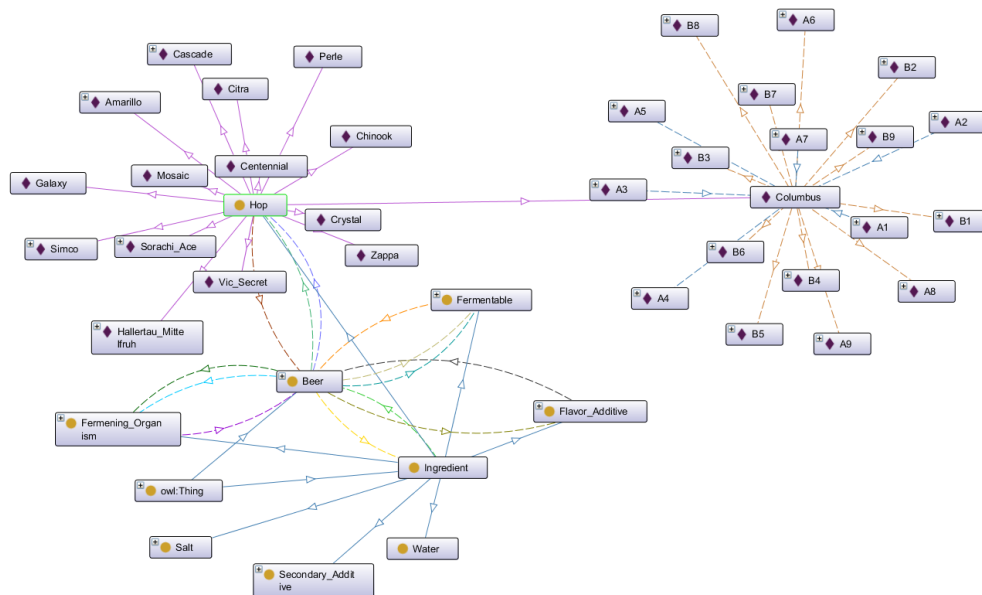


Figure 4.29: **Example of partial ontology structure for the ingredient category, including connections between an ingredient instance and a sample instance.** Full lines represent parent/child relationships, dashed lines represent custom object properties. Figure created in Protege.

## Object properties

Recipes for the different samples are represented as links between an instance of a beer type (sample) and an instance of an ingredient. This was achieved using custom object properties as links between the instances, like the property *has\_hop* being a link between a sample and a hop type used in the sample. The reasoning for representing the recipes in this manner is that this allows for a much more transparent and detailed network of information, as opposed to only describing the recipe in text format as metadata for each sample. This structure enables multiple approaches for investigating the use of an ingredient or other components in a certain beer. The current list of object properties used in the ontology is shown in figure 4.30.

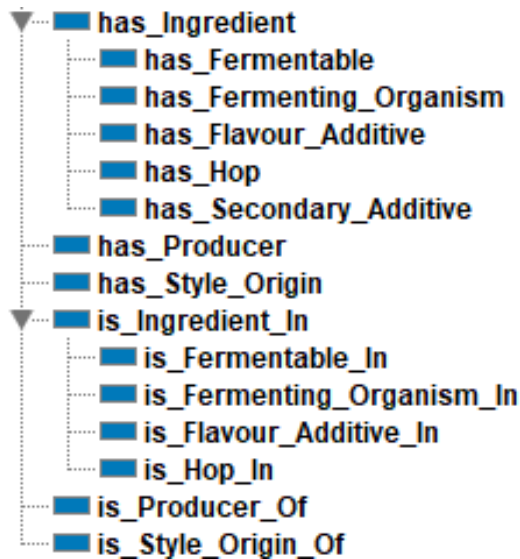


Figure 4.30: List of custom object properties used in the ontology. Figure created in Protege.

### Data types and data properties

Custom data types were created to represent sample ID, and custom data properties were created to record the beer name; brewing and sampling dates; and gravity measurements. Measurements of *European Brewery Convention* (EBC) standard for beer color and *International Bitterness Units* (IBU) scale were also recorded as custom data properties. In certain cases, custom properties were created to describe aspects relevant only to a few samples, like a property for recording the pitch rate.

#### 4.2.3 Public access

The OWL-code for the ontology is published on a GitHub page dedicated to the Systems Brewology Project, and is available at: <https://github.com/AlmaasLab/SystemsBrewology>. This allows for public access to the current ontology, as well as any future updates to be made throughout the lifetime of the project. This way, the ontology can be a valuable resource for any venture needing an organized way to structure and annotate beer samples.

---

## 4.3 Database framework

The database framework chosen was openBIS, as this was the most flexible and fully featured of the systems under evaluation. openBIS was the only system that would allow for a fully customizable structure that would be ideal for our use, allowing for less compromise in the database design compared to other solutions. A detailed explanation of the decisions made for this decision is presented in section 5.5.



Figure 4.31: **Logo for openBIS.**  
Procured from <https://openbis.ch/>

## 4.4 Outreach

Sample production for this thesis were done primarily in-house, but the aim is to set up a much larger collection system where both private and commercial brewers can send in samples to include in the database. As a start in this direction, one of the studies in this thesis was done using samples from commercial beers made by the Austmann Brewery, who were interested in sharing samples and recipes with the project. Similar conversations have been started with other commercial breweries in the region.

To spread information on the project and enable communication between us and potential contributors, a project website was created. The website contains brief information on the project and how others may contribute to the effort, as well as contact information for further communication. The website can be found at URL: <https://almaalab.nt.ntnu.no/index.php/research/systems-brewology/>, and an example is shown in figure 4.32.

The design and content of the website is intended to draw attention and interest from the general brewing enthusiast, rather than appealing to the scientific community. The aesthetic is simple and clean with a clear theme throughout, and does not include detailed technical information that would not be of interest to non-academic visitors. We do, however, intend to add additional pages where more details on the project, work, and aims can be found. The current version is only intended to supply cursory information and enable contact between us and potential contributors.



# Systems Brewology - Analysis of beer

Analyzing beer samples from home brewers, industry, and laboratory production  
to increase knowledge for brewers and scientific understanding.

## The beer project:

- Science for a better brew -



### What we do

We take a scientific approach to the art of brewing. The craft has always been more about know-how, than know-why. We delve into the scientific understanding of what gives different beers their different properties, and aim to understand the connections between choices in the brewing process, the resulting flavor compounds, and how this finally affects the experience of drinking the beer.



### Why we do it

We believe that brewing is a craft. Through scientific research, we believe that we can expand the understanding of how the components used in the brewing process affects the final beer and quantify the minute details in their combinations that makes the difference between a fine pint and a truly exceptional beer.



### How we do it

To quantify the components of a beer in detail, we use nuclear magnetic resonance (NMR) spectroscopy - similar to an MRI machine at the hospital - to give us a unique fingerprint for every sample. With many samples along with accurate recipes, we build a database of different beers, allowing us to look for patterns and trends, to see what compounds stem from which ingredients, and to discover new connections between process and product we have yet to understand.



### Why we need you

To build a database collection of different beers with detailed recipes, we are in need of a large number of samples where we have knowledge of each ingredient and the brewing procedure used. We need far more samples than we can brew ourselves, and we wish to cooperate with both professional breweries and homebrewers in order to produce all the samples we need.



### What's in it for you

If you are willing to share your beer and your knowledge with us, you can contribute to this amazing project and hopefully help increase our collective know-why of the brewing process. With enough people we can move mountains, and we plan on doing exactly that! Besides: In how many scientific projects have you already participated?



### How to help

If you have a spare deciliter of a beer, maybe even a full single bottle, and a recipe to go along with it, you have all we need! If you wish to contribute to our research, send it to us here at NTNU in Trondheim for analysis. Contact us for more detailed information.

## Please contact us:

[brewlab@biotech.ntnu.no](mailto:brewlab@biotech.ntnu.no)

Brewology - Sem Sælandsvei 6/8, Kjemblokk 3, 163 - Department of Biotechnology and Food Science, NTNU - 7491 Trondheim



Figure 4.32: Illustration of the Systems Brewology website.

# Chapter 5

## Discussion

This chapter will discuss in detail the findings presented in chapter 4. The implications of findings from the beer sample case studies will be discussed, as well as the limitations of these simple studies and suggested improvements for more reliable results. Experiences from the case studies in terms of implications for the proposed database in terms of requirements for importing, storing, and exporting the relevant data will be evaluated in detail, and the decisions made in the design and setup of the beer ontology, as well as the rationale behind the evaluation of different database solutions will also be laid out in this chapter.

### 5.1 Case studies of beer samples

#### 5.1.1 Categorization study of Austmann samples

The purpose of this analysis was to examine if our analysis method would support or refute a simple conventional categorization of the 19 samples from Austmann. As shown in section 4.1.4, only one of the assumed categories were supported by the analysis, while the three others showed strong overlap in chemical features and were placed in mixed clusters by the k-means clustering algorithm. Perhaps not surprisingly, the category of sour beers were the only one to be supported by this analysis. This may be due to the large difference in lactic acid content between sour beer and other types, as can be seen in the spectra in figure 4.10, as well as the presence of taste compounds from fruit and other additives used in these beers. Because the pH of these beers is quite low, it may have been outside of the buffering range for the buffer used in the sample preparation. Therefore, it is possible that the low pH of the resulting samples has affected the alignment algorithms and thereby caused false differences between the spectra.

For the other categories, most of the peaks driving the difference are located in the highly crowded regions of the spectra, making any compound identification difficult, and thereby making it hard to assign differences in compounds to the new sample clusters. Unfortunately, recipes for these commercial beers were not attainable at the time of writing, making any evaluation of the new clusters on the basis of ingredients purely speculative.

Ideally, we would like to be able to create more objective categorizations of these beers based on the chemical composition instead of taste, cultural origin, ingredients, or color as the conventional categories are. Such categorizations might be less informative to the

---

general consumer, as all compounds detected by NMR are not experienced the same while consuming beer. However, we believe understanding the chemical similarities and differences in beer categories, as well as exploring new approaches to categorization will lead to new insights that may be valuable to the field. As the full-spectrum approach cannot provide fully conclusive results in this study, further improvements are required. Some suggestions are as follows:

Larger sample size could potentially make trends and patterns more clear. As this study only consisted of nineteen samples, some of the original categories included only three samples. This provides low statistical power for many of the tests performed. A significantly larger collection of samples could potentially lead to the discovery of some common identifiers either supporting or refuting conventional categorization.

Another approach than full-spectrum analysis, that is used in many studies, is to analyze separate areas of the spectra individually to gain insight into the differences in a small collection of compounds, or sometimes a compound type occurring in a given ppm region [20, 25, 26]. This approach would avoid the problem of crowded regions affecting the analysis without providing information on what compounds actually differ. Another benefit to a partial spectrum approach would be the ability to apply scaling functions to a specific region of similar peak intensity, making the process of distinguishing between baseline and peaks easier for the scaling algorithm compared to the full spectrum. This approach would sacrifice some of the benefits of quick full-spectrum fingerprinting in order to gain greater accuracy. A combined approach could be a possibility, where a more detailed analysis could be guided by preliminary discoveries from a full-spectrum evaluation.

Finally, an even more detailed approach would be to quantify the concentrations of relevant compounds by peak integration, described in section 3.3. This would allow for even more accurate studies of the relative quantities of compounds between samples as well as between and within groups or categories. Such quantification processes are time consuming and labour intensive however, and would therefore not be suitable in a fingerprinting capacity, as we envision for the full-spectrum approach.

All in all, results of the categorization study did not support most of the conventional categories, but failed to construct more informative categorizations based on the chemical spectra of the samples. This failure seems mainly due to lacking sample metadata and resolution in the statistical analysis.

### 5.1.2 Pitch rate study

The study examining the effects of pitching rate was the only experiment designed specifically for this thesis with a clear hypothesis in mind. The goal was to explore the difference in chemical composition caused by different pitching rates, as described in section 3.1.3. The results of the analysis, section 4.1.5, show a few interesting trends but also reflect a general lack of clear separation between the three pitching rates reflected in the chemical spectra. The most noteworthy result here might be that multiple independent sensory analyses provided the same clear trends and similar descriptors for the different samples, while the statistical approach failed to do the same. This is most likely due to the human sensory system detecting and experiencing compounds very differently from the objective NMR approach.

Looking at the PCA along with the spectra and the PCA loadings, figures 4.12 and 4.14, we can see some categorical separation, however, the main source of separation lies in the

---

crowded region of the spectra, making it hard to identify what compounds are actually differing between the treatments. We are able to identify a few compounds with actual differences, however there is clearly much we are missing. The separation seen in PCA and t-SNE were not supported by the dendrogram in figure 4.15, where the three treatment groups form mixed clusters. This suggests that the separation found in PCA and t-SNE might not be accurate.

Interestingly, there seems to be a trend present in the ethanol content of the three samples. Illustrated in figure 4.16, we see a seemingly negative relationship between the amount of yeast used for inoculation and the ethanol content in the resulting beer. A detailed study of the effects of pitching rates in beer [14] demonstrates no difference in ethanol content between different pitching rates, putting our findings into question. The trend illustrated here is not statistically significant, figure 4.17, meaning we cannot draw definite conclusions from these results.

This pitch rate study was just a small-scale case study and would have to be improved and scaled up to yield truly significant results. Some suggestions for improvements are as follows:

The same improvements mentioned above for the categorization analysis of Austmann samples would apply here as well. Increased sample size, partial-spectrum analysis, and compound quantification would probably provide better insight into the effects of pitch rate.

The study design could also be improved by increasing the number of treatment categories to five, making trends easier to identify than with only three treatments. A more accurate system for growing and measuring yeast would also be beneficial if this was to become a large-scale experiment. Finally, a proper sensory analysis performed by a trained panel, or a large number of participants, would yield more accurate information on the sensory aspect of the samples, possibly revealing other trends than those found here.

### 5.1.3 Aging study in algae-flavoured beer

The analysis of aging in beer flavoured with algae was originally not intended for the study of any specific hypothesis, but rather designed to evaluate the suitability of algae as a flavouring additive in different beer types using different fermenter organisms. The nature of this data set with a large number of variables made it quite suitable as a case study for our analysis platform, as most of the samples we would expect to receive for an eventual database would be from vastly different samples not designed for scientific experiments. The purpose of this analysis was therefore to evaluate the possibility of reducing the number of variable factors to a level where we could study some specific aspects of the data. In this case, we had the opportunity to run two NMR analyses at different time points, allowing for a study of aging.

As seen in section 4.1.6 there was no discernible trend caused by aging across the whole data set of the four beer types, only in the brown ale category did we see a clear separation between the two aging groups. These results are in line with the findings of [16], where darker beers were shown to be more affected by aging than lighter beers.

Looking at the spectra along with the PCA loadings, figure 4.25, we see the usual trend of highly weighted compounds being located in the crowded region of the spectra, foiling our efforts of identifying the differences in compounds caused by aging. Still, we see a few

---

notable compounds change according to age group, as mentioned in section 4.1.6. This indicates that there might be a difference to be found, however, the analysis methods are limiting due to the large effect of the crowded region of NMR spectra, making our discoveries inconclusive.

The same improvements as mentioned above for the pitch rate study and the classification analysis would be applicable here. Larger sample size, partial-spectrum analysis, and quantification of compounds would most likely yield more information on the effects of aging. Additionally, a specialized study should be designed to examine the effects, with less variables like different yeasts and algae flavouring included.

#### 5.1.4 Evaluation of the analysis pipeline as a whole in light of the three case studies explored

As illustrated by the three different case studies performed in this thesis, the analysis pipeline has both benefits and some areas of potential improvement. As a method of quick evaluation of a sample set for indication of trends or effects for further study, the pipeline works fairly well. The required time between sampling a beer and having preliminary results from multivariate analysis can be less than 24 hours, depending on the complexity of the data set and the statistical analysis involved. This enables a high potential throughput of samples for evaluation, although with low accuracy for detection of certain compounds. More detailed studies on discovery rate would be beneficial in evaluating this method of analysis.

The shortcomings of the approach demonstrated here lie mostly in the lack of accuracy of a full 1D spectrum analysis approach. A common problem noted in all three case studies is the heavy weighting of the crowded 3.5-4.5 region of the spectra in the loadings of multivariate analysis. One approach that may lead to a better understanding of this region would be 2D NMR spectroscopy, which could potentially provide better resolution to the spectra by adding another dimension to the analysis. This would however increase the analysis time and complexity considerably, as well as complicating the statistical analysis. The 2D NMR approach to beer analysis has shown promise in previous studies [60], indicating it as a possible approach for later studies.

Another potential approach mentioned above would be to pick out identifiable peaks of interest and perform a more thorough quantitative analysis of these peaks. The problem with this approach alone is the lack of discoverability as unexpected peaks might be overlooked if only predetermined areas of the spectra are analyzed. A similar problem is present in other chemical analysis methods commonly used for studying certain compounds in beer, like *High Performance Liquid Chromatography* (HPLC) and *Gas Chromatography-Mass Spectroscopy* (GC-MS), where high resolution is achieved at the cost of a much narrower view where only a specific subset of compounds are detected. These methods are therefore not suited to obtain a full profile of a complex medium such as beer; however, as a supplementary analysis to a full-spectrum NMR approach, these detail-oriented analyses may be quite beneficial.

The statistical methods used here have varied between the different case studies, as different tests and measures were more suited to the different data sets and hypotheses. While we believe a flexible toolbox of statistical tests is important in order to adapt to different hypotheses, a standardized set of initial statistical analysis could be beneficial to evaluate trends in comparison to previous studies and to not only choose statistical tests based on

---

the results we expect, thereby reducing biases and increasing the chance of discovery of unexpected results.

Compound identification is another area that would benefit from more work. As of yet, we only have a few peaks of the spectra identified and linked to compounds, and only a few of these are possible to quantify with acceptable accuracy. Developing a comprehensive library of known compounds and their expected chemical shift in the spectra would significantly increase the insight gained from a simple 1D spectrum, and will be a prerequisite for more detailed quantitative analyses. This will require a combination of literature studies for known compounds of interest, as well as analysis by *heteronuclear single quantum coherence spectroscopy* (HSQC) to identify compounds present in small quantities.

In conclusion, the recommendation for future analysis is to create a robust and comprehensive analysis pipeline based on rapid 1D full-spectrum analysis for initial discovery, guiding more detailed quantitative methods like 2D NMR or GC-MS. Coupled with a versatile suite of multivariate analysis techniques and a comprehensive library of identifiable compounds, this approach to beer analysis could have great potential for the discovery of new and valuable insights into the production of beer.

## 5.2 Data treatment

### 5.2.1 Challenges and decisions on scaling of chemical spectra

As described in section 4.1.2, an issue with commonly used scaling methods was discovered during multivariate analysis, where the scaling process would scale the noise of the baseline or crowded regions of the spectra to the same level of significance as the peaks where the actual informative differences lie. The problem resulted in real peaks rarely driving the difference between groups, as random differences in the baseline might represent more relative differences in the scaled data.

To work around this problem, all nonparametric analyses were performed without scaling, as this allowed the differences in the peaks to be weighted more than those in the baseline. This is far from a perfect solution, and statistical results from the multivariate analyses must be viewed in this light.

We believe the unscaled approach to be a viable option for achieving quick results of indications of trends and structures in the data. However, we also suggest solutions for scaling and filtering of full spectra should be developed in order to gain a more accurate method of analysis.

## 5.3 Lessons from the case studies for data storage

Circling back to evaluate the original purpose of the three case studies of beer: To gain a better understanding of the strengths and weaknesses of our analysis pipeline, as well as to evaluate the structure and nature of the data to be stored in the proposed database. The first purpose is discussed at length above, as for the second purpose we distinguish between three main categories of data: Metadata, analysis data, and results data.

---

### 5.3.1 Metadata

Metadata in this sense encompasses all information about a sample that is not gained from the analysis, including recipe, style and type descriptions, producer, brewing and sampling dates, and any other pertinent information. After going through the case studies, it became clear that different data were important in different analyses, and that what information is needed might not be known prior to analysis. It is therefore recommended to record as much metadata as is practical for any given sample. Lack of metadata can severely limit analysis, as was the case in the Austmann case study where we were unable to obtain production recipes for the samples in time for this thesis, resulting in a somewhat stunted analysis effort.

Some metadata might seem relevant only for a given experiment, such as detailed information on the pitching rate for the second case study. However, having knowledge on pitching rates for all samples would enable a more comprehensive analysis, evaluating the effect across multiple variables.

The storage structures for metadata were not a concern for the small studies performed by hand for this thesis, however it might become important for larger analyses where machine-readability is a concern. A standardized way of storing metadata should therefore be developed in the proposed database.

### 5.3.2 Analysis data

All data recorded during some form of analysis, either chemical spectra from NMR or descriptors recorded by sensory analysis, will be considered *analysis data* in this section. Through the three case studies different forms, or *stages*, of analysis data were recorded.

Firstly, three raw NMR spectra were recorded for each sample, where two of these were relevant for further analysis. These spectra should be stored in a format that retains as much of the original information as possible, ideally in the native storage files of Bruker TopSpin, containing all the information of the spectra, as well as metadata on the NMR-acquisition, process, and parameters used. This data is stored in a folder structure readable in TopSpin, and might prove difficult to store and read within a database system. A solution could therefore be to store the raw spectra as spreadsheets which can be stored and read by a database system, while maintaining a separate backup of the spectra as Bruker files, possibly in a compressed format for storage purposes.

Secondly, *composite spectra* are created for each case study, where spectra from the relevant samples are merged and aligned into a matrix of multiple spectra, as described in section 3.2.4. This alignment must be done for each collection of spectra in order to avoid false discoveries in the statistics caused by misalignment of the spectra. The alignment procedure is time consuming, and it might have to be done slightly differently for each collection of spectra. For traceability and reproducibility of results, it is therefore important that these spectra collections are kept after a completed study. These collected spectra can be stored as large spreadsheets, which can be stored and read by the database system. The intermediate files and the procedure for creating these collections can also be stored in the form of Matlab workspace files. Although this is useful for reproducibility, it might be excessive in terms of storage space.

Thirdly, data from other forms of analysis must be considered. Other chemical analysis methods can most likely be stored in a similar way as the NMR spectra, as these would be

---

either the spectra themselves or tables of peak intensities. Storing more detailed results from sensory analysis could pose more of a problem, as the variables analysed by a tasting panel might not be identical for different studies. Standardization is important to machine-readability when considering large scale or global analysis of a whole library of samples using computer tools, and therefore the development of a standardized way of recording and storing sensory data is recommended.

### 5.3.3 Results data

In some cases, we might be interested in analyzing the outcomes of different studies, evaluating them in relation to each other, and performing internal meta-studies. To facilitate such efforts, it is important to store detailed results of the studies already performed, not only for the purposes of publication and reproducibility, but also for later evaluation of the findings. Making earlier discoveries easily available in an ordered fashion would also help in the exploration of possible findings from previous studies that were not followed up, making it more likely that indications of findings or trends are found and tested.

Storage of results would by its very nature be individual to each hypothesis tested, however, developing a unified storage solution for statistics from previous studies would be recommended for a long project of this type.

## 5.4 Outreach and contributions

The results of the effort to establish cooperation with the brewing community can be considered partially successful. Meetings with representatives from Austmann Brewery in Trondheim and Voss Brewery in Voss, have resulted in preliminary agreements to share recipes and samples from the brewing process. Smaller studies, like the study of algae as a flavouring agent in beer, was performed in cooperation with E.C. Dahls Brewery in Trondheim and SINTEF, with the aim of exploring potential new commercial products using algae. Such scientific and industry cooperation is highly beneficial, as it allows the same samples to serve multiple purposes.

Gathering of samples from breweries has been limited however, due to the restrictions caused by the SARS-CoV-2 epidemic starting in late 2019. These restrictions have caused delays in analysis development, as well as severe limitations in the brewing industry, making it necessary for some potential contributors to postpone involvement. We do, however, believe there is great potential for cooperation from the local brewing industry when the situation returns to normal. We now have a more streamlined analysis procedure that will enable quick results after sample gathering, as well as having access to previous results that may spark interest in potential contributors.

The homebrewing community could also be a resource for gathering a large number of samples. The website created is a first step in the direction of enabling contact with homebrewers, however, creating a base of contributors in the homebrewing community is beyond the scope of this thesis. For future efforts, we recommend establishing rapport with homebrewers through forums, social media, and conventions. Facilitating the sharing of results and knowledge between this project and the homebrewing community could be of great value for both parties, and a good opportunity for citizen science.



---

## 5.5 Database selection

A substantial part of the groundwork necessary for massive-scale analysis envisioned in this thesis was choosing a database structure based on our needs. Multiple alternative database models or services were evaluated based on the prerequisites we had. The following is a short overview of the requirements for our database solution:

- The system must be highly configurable and flexible in terms of structure and contents.
- The system must support a hierarchical structure of sufficient depth to handle detailed categorization.
- Objects in the database must be queryable and filterable by multiple identifiers.
- The system must allow for the creation of custom data fields or templates for recording of diverse metadata.
- The system must be able to store and read either NMR spectra in native format, or large spreadsheets.
- Queries must be exportable in formats that can be used in analysis.
- The system must include security options to restrict access to potential industry secrets.
- Custom input options or templates must be configurable so that non-experts can input information to the database.

It should also be mentioned that multiple projects within the same research group might be developing database solutions alongside the Systems Brewology Project. Suitability for other applications might therefore affect the final decision on what will end up being implemented. The conclusion in this thesis will be a recommendation based on the suitability for the solution for the problems posed in this thesis.

### 5.5.1 Alternatives

Based on the requirements above, multiple known database structures were evaluated. This section contains the considerations made for a short-list of the most promising alternatives, which includes the following:

- Fully custom-made database structure based in Python
- Metabolomics Workbench
- Fairdom Seek4Science
- ISA Tools
- openBIS

---

## Custom Structure

The first alternative considered for the database solution was to construct a fully custom database using Python and *MySQL*. This would allow for a fully customizable relational database structure where we would be able to construct whatever internal structures and external interfaces we would need. The versatility of this option makes it a compelling alternative.

The drawbacks of a fully custom solution is that it requires a great deal of skill and knowledge to construct, and constructing such a database on par with already available solutions would require time and resources that are far beyond the scope of this thesis, and possibly even the project as a whole.

## Metabolomics Workbench

Metabolomics Workbench is a browser-based data repository consisting of multiple metabolomics databases for biomedical research. Their service also allows for creating custom studies and uploading metabolomics data for the study, including metabolite analysis like MS and NMR spectra as well as accompanying metadata for the samples. [61]

Such a solution would fit well with many of our requirements, as the structure for storing NMR spectra and metadata is already in place, search and filtering systems are in place, and even simple analysis as well as data visualization is possible within the database. The service comes up short in terms of customizability, and the available templates for metadata and annotation will not fit our requirements for storing recipe information and brewing process used for beer samples. Another problem presents itself in the requirement for security, as Metabolomics Workbench stores all information in publicly available repositories, and we might need to store recipes that are not public knowledge.

## ISA-Tools

*ISA-Tools*, based on the *ISA framework*, is a software suite of multiple applications. It is based on the ISA data model, consisting of a hierarchical structure describing "Investigation" (project), "Study" (unit of research), and "Assay" (analytical measurement). This structure enables a comprehensive overview of the recorded study, with appropriate metadata recorded at each level in the ISA hierarchy. The software suite enables the creation of customizable repositories, using a host of different software packages for creating, annotating, and exporting data. External API support for *iPython* and *Jupyter notebooks* is also included, enabling fully custom interaction with the repository if needed. [62]

This solution could be a good fit for our needs, the customizability available with the software suite would most likely allow us to create a structure suitable for our needs. The limitations of the ISA structure is, however, a concern. Limiting the hierarchical structure to a scientific standard that may prove restrictive to the unconventional needs of studies on fermented beverages. These restrictions could possibly be circumvented by bending the rules of the hierarchical structure, but a more flexible solution would still be preferable. Security could be upheld with ISA-Tools using local hosting of the repository, and integrated connections to publications would ease efforts in knowledge sharing and dissemination.

---

## Seek

The *Fairdom Seek4Science platform* is another ISA-based platform. Seek expands on the regular ISA model, claiming it to be more configurable and usable outside traditional biology. Otherwise, seek has many of the same properties as ISA-Tools, although with a smaller suite of additional software. It also offers integrated use of ontologies with *RightField* enabled spreadsheets for metadata annotation. [63]

Seek seems to be a well-suited solution, as it lessens the restrictions imposed by the ISA format, while maintaining many of the attractive features offered by ISA-Tools. The integration of ontologies for metadata annotation could be a great benefit in this project, as the initial sample structure is based on a custom ontology. Seek also includes access control to keep contents in the database private and offers versions for private hosting. One notable drawback to Seek is the lack of support for data spreadsheets linked to a sample, this would require NMR spectra to be included as data files with their contents unavailable to be searched or interfaced with within the repository. Exporting options also seem to be limited, however this might be solved with external scripts through API. Otherwise, the sample annotation and filtering system would fit well with our needs.

## openBIS

openBIS is the most comprehensive option on our list, it includes solutions for inventory management, lab notebook, version tracking, and extensive systems for data management, analysis, import/export, as well as API and plugin support. Data storage capabilities include full data spreadsheets accessible from the interface, as well as configurable data fields for metadata. Import and export can be automated with the integration of API or plugins, and simple analysis can even be performed within the database. Data files of spectra can also be included and linked to the samples, however these cannot be manipulated within the database. Querying and filtering is possible in both annotated data tables and metadata fields. The whole setup can also be hosted locally with access control, ensuring secure storage if necessary. [64]

openBIS is a somewhat different solution from the others on this list, as it offers a more complete environment for up to multiple projects. This may or may not be a benefit in our case, but should be considered in case the need arises. The data storage solution offered by openBIS is as customizable as both ISA-based options, but with more support for interaction with raw data like spreadsheets. The extensive support for API and plugins also offer possibilities for simplified or even semiautomated input procedures of sample data, as well as customizable export options. The major drawback of openBIS is the complexity of the system and the time necessary to set up and configure the database.

## Final Decision

The final decision for a database solution will be openBIS, with Seek as a good alternative. Both of these solutions offer all the functionality required for implementation of the database, and stands out compared to other alternatives in terms of flexibility and support for external tools. From the initial research, openBIS appears to be a somewhat more flexible system compared to Seek, and offers a more complete solution that may have benefits for later efforts during the lifetime of the project. openBIS does, however, come with a more complicated installation and setup procedure that may require significant

---

time and resources to complete, which may not be the case with Seek as it is a far less complex system. Despite a more demanding initial process, we believe the benefits of a more comprehensive solution will outweigh the drawbacks.

## 5.6 Ontology

### 5.6.1 Accuracy of classifications

The ontology aims to introduce an attempt at standardization to a subject fraught with a wide array of differing opinions on definitions and standards. The purpose of constructing this new ontology, based on a collection of definitions offered by different experts and authors, is to capture a unified snapshot of the current state of understanding of beer style classification. Such a pieced-together set of definitions might have minute overlaps in the classifications used in our annotation, either due to a lack of intimate understanding of the beer by the annotator, or poor definitions offered by the author of the source. The creation and curation of such an ontology is by necessity an iterative process, and will have to be reevaluated and changed by those continuing my work on this project. With the aid of the analysis pipeline developed in this thesis, we can hopefully offer new perspectives on the classification of beer styles.

#### Sourcing of classifications

The beer style categorization was based on two main sources: The beer style guides of craftbeer.com [55], and the Brewers Association [54]. These are organizations based in the United States, promoting knowledge and education about craft beer and brewing procedures. They both publish style guides detailing the taste, characteristics, and important ingredients of a large number of beer styles. Together, these two guides cover most of the known beer styles in sufficient detail for the purposes of our ontology. However, a more detailed annotation process will have to be undertaken for the ontology to accurately identify and cover all possible styles of beer. In a few cases where the style guides were sparse on information, some descriptions were supplemented by information found on *Wikipedia.com*, we considered this an acceptable source in this endeavour as it represents a common understanding of the beer styles described. Which, by the very nature of beer styles, is what the ontology is designed to provide.

#### Class structure

The design of the class structure for the beer styles is somewhat unconventional as it categorizes a style both by a description of ingredients and characteristics, as well as by region of origin. This adds a layer of complexity to the categorization that might serve to make the ontology harder to incorporate into a database or other structure. However, we believe the added benefit of multiple avenues of categorization makes up for the additional complexity. The two different avenues of categorization are separate, and one can therefore be removed if needed.

Based partially on categorizations from the style guides and partially on our own knowledge on the ingredients and processes of the different styles, similar styles were grouped together under a common name, like the different versions of the German beer style *Bock* for

---

example. These groupings can definitely be improved with further research and annotation. As different sources disagree on these groupings, more detailed research are required to make sound and consistent design choices on how to define these in the ontology.

## Object properties and links

Choosing to construct the recipes of samples in a network form using object properties to describe the inclusion of ingredients, rather than just adding a recipe to the metadata, does also add to the complexity of this ontology. The reasoning behind this choice is that more information can be included in this way, as separate metadata can be annotated for a specific ingredient, giving us the ability to annotate the ingredients themselves with results from our studies or literature research. This results in a more flexible and interactive structure, again providing multiple avenues of inquiry into the ontology. Using an approach like this provides information in both directions for one annotation; not only can you easily figure out what hops are used in a certain beer, but also what beers contain a certain hop.

The current structure of object properties only covers the basic connections between a sample and instances of ingredients, producers, and styles. This annotation system still needs more development in order to enable the annotation of more complex information about the brewing procedure for a sample, like the time point an ingredient is added, or the amount used. This information could also be added as metadata for the sample, however recording it as part of a link would be beneficial in some instances.

All information on a sample cannot be included in the links for practical purposes, and the choice of what to include as metadata and what to include in custom connections is a complex one. This will require much practical testing and evaluation of the ontology by experienced brewers and other domain experts.

## 5.7 Summary of recommendations for future efforts

Following the reasoning in section 5.5, the openBIS database solution should be used for the construction of a the database. It meets all the requirements set in this thesis, as well as being a comprehensive solution suite that can be used for multiple projects and efforts in the future. Should openBIS not perform as expected, the recommendation is to use Seek, which also met all the criteria we set.

The final system will have to contain standardized templates for importing and annotating samples in order to keep the information structure similar and easily available for exporting algorithms. All available metadata should be recorded for a given sample, as seemingly unimportant features might be relevant for some studies. The import templates for the database should account for this, and should therefore be developed to account for every known type of ingredient or treatment, as well as have the flexibility to encompass new aspects of beer we have yet to predict.

The database should be designed to store large quantities of spectral data, multiple spectra may be needed to fully encompass each sample. The storage system must also be designed to handle not only 1D NMR spectra as first expected, but also other forms of chemical analysis like HSQC, HPLC, and GC-MS, as well as data from sensory analysis. There should be standardized ways of importing and storing all relevant forms of chemical analysis in order to make querying and filtering possible for all facets of a sample. The proposed

---

database must also have a solution for recording findings from previous studies, as well as any composite spectra or similar treated data used in the study. This will enable a more iterative approach to these studies, making illusive discoveries readily available by building upon previous discoveries or exploring weak indications in previous studies.

Looking at the results and discoveries from the case studies, sections 4.1 and 5.1, we find that the methods used have both limitations and potential. The limitations lie in the problems caused by highly crowded regions of the NMR spectra limiting compound identification, as well as in problems with scaling methods either amplifying noise in the spectra or underestimating differences in smaller peaks. These problems are inherent in the full-spectrum approach, but may be solved by customizing methods of filtering and scaling for the type of spectra. We see great potential in the approach, as we can very quickly treat, analyse, and evaluate a large number of samples. This gives us a solid platform for discovering potential features for more targeted approaches. Therefore, we believe the optimal analysis strategy for this project lies in a combination of different chemical analysis methods at different levels of specificity.

In terms of gathering samples, the time working on this thesis has shown that brewing our own samples is a valuable resource for targeted experiments, but it will be far too costly and impractical to produce samples in the quantities and variation we would need for a comprehensive database covering representative sample sets from all major types of beer. The need for external samples is clear, and strategies must be developed to facilitate this. During this thesis we had limited luck receiving samples and recipes from professional breweries, however, we believe this to be caused mainly by the Covid-19 pandemic severely impacting the brewing industry. In future efforts, we should focus on reestablishing contacts with potential contributors in the industry, as well as developing rapport with the homebrewing community to receive samples and input from them.

## Chapter 6

# Outlook and summary

This chapter will summarize the conclusions of this thesis, as well as present suggestions for the future of the Systems Brewology Project, what steps need to be taken to achieve the aims set for the project and what sort of findings might be reasonable to expect.

Summing up the findings and the work performed for this thesis, we see that we now have a solid foundation on which to build a database solution for later studies, a methodology for performing these studies, and structures for both procuring and annotating samples for the project.

Based on the results of the case studies of beer samples done for this thesis, we can conclude that 1D  $^1\text{H}$  NMR spectroscopy is a quick and efficient method for qualitative studies of beer. The statistical methods of PCA, PLS-DA, t-SNE, hierarchical clustering, k-means clustering, and compound quantification were valuable tools for analyzing the spectra produced by NMR. Some limitations in the accuracy of 1D NMR for use on complex media were discovered however, and for future efforts we recommend evaluating a combination of chemical analysis methods.

The case studies performed for this thesis resulted in interesting discoveries about traditional categorizations of beer styles, and new categories were explored based on clustering of the chemical spectra. Notable effects caused by varying pitching rate on the taste of beer were documented, along with preliminary results on differences in chemical spectra and in certain chemical compounds, as well as ethanol content in the finished beer. Differences in aging of certain beer types flavoured with algae were also studied, and clear effects of aging were discovered in the chemical compositions of samples of a darker beer style.

Based on evaluation of the data used in the case studies, critical recommendations were made regarding the implementation of a database for storing of the samples. The recommendation is to employ the openBIS database system as a framework for constructing a database. From observation of the requirements for statistical analysis, it is recommended that a highly standardized importing and exporting system is developed for the database, so that data can be both imported and exported in standardized formats. This will be critical for consistency in the data structure, both for the use of querying applications and for the use of exported data sets in data mining and machine learning algorithms.

The custom ontology created during this work will be valuable both for informing the structure of the database and for understanding the hierarchical structure of beer styles. This allows for a detailed annotation solution for the samples used in this project, as well as enables sharing of concepts between users and experts of the field using a standardized

---

knowledge base. This may also be of value in communication between researchers and brewers. The ontology is made publicly available, and it is highly recommended that future work in this project involves continued development of the ontology.

Procurement of future samples will be facilitated by the project website created as part of this thesis, while the sample treatment and NMR acquisition procedure is standardized and documented for future use in the manual found in appendix E. This contributes to usability and knowledge transfer to new users of the pipeline.

This project has great promise for future studies, the preliminary results from case studies presented in this thesis already demonstrate the ability to detect trends and characteristics in different sample sets. We believe that building on the work documented here, and with the right improvements to the analysis pipeline, data storage solutions, and sample procurement, this project will produce valuable results.

There are multiple paths forward for this project, the most important of which will be constructing a research platform based on the requirements and suggestions discussed in chapter 5. With this platform in place, the next step would be to start gathering samples in large quantities and running them through NMR analysis to incorporate them into the database. When the samples are in place, analysis can start by using mining algorithms to look for trends or patterns in the data, as well as to manually perform targeted analyses such as those used in this thesis to study specific hypotheses.

In platform development, the next step is to start constructing the actual database using the openBIS database structure, as presented in section 5.5. A separate work needs to be done to elucidate the final data structure to record the expected samples, although this effort may be guided by the ontology created in this thesis, as well as by the recommendations made here. However, there will most likely be unforeseen peculiarities of the openBIS system that will require extensive deliberation of the information structure to be implemented.

The required aspects of the database envisioned here include custom import and export procedures, sorting and querying systems, as well as security systems.

The importing and annotating system will have to be standardized to ensure that the metadata for all samples are structured in the same way. Our recommendation for achieving this is to construct custom templates in openBIS, or using scripts incorporated through APIs. Ideally, these templates would employ a controlled language based on the ontology created for this thesis, as well as specialized data fields for recording the most common items of metadata that would be present in all samples.

The exporting system will have to be designed along with the querying and filtering system. For exporting, it is critical that the files are both human and machine-readable, enabling both manual studies as well as the use of machine learning algorithms. For querying, it will be important that the importing system is standardized, and that the recorded spectra are readable by the querying algorithms. This may require some annotation during the importing step. We expect the exporting can be done mainly by the systems included in the openBIS suite, however there might also be a need for some use of external scripts through API should the exporting systems in openBIS be limiting. The querying system is already incorporated in the openBIS solution, and will be well suited to the tasks required for this project, provided the data is stored in a readable format.

In terms of gathering samples, this thesis has proven that brewing our own samples is a valuable resource for targeted experiments, but it will be far too costly and impractical



---

to produce samples in the quantities and variation we would need for a comprehensive database covering representative sample sets from all major styles of beer. The need for external samples is clear, these may come either from commercial breweries or the homebrewing community.

For industry cooperation, the recommendation for future efforts is to contact local breweries and brew-pubs to inquire about procuring small samples along with recipes and other metadata. For contact with homebrewers we recommend developing rapport with the community through a presence in forums and social media, as well as brewing conventions like *NorBrew Con*. The project website will also be a useful tool in contact with both commercial and amateur brewers.

When a final database is constructed and populated by samples, the analysis can begin. We suggest the use of data mining and machine learning algorithms for discovering trends in the samples. Early efforts would primarily focus on data mining where the aim of discerning new patterns in the data sets. This might, for example, lead to new ideas of beer style classification, or relations between ingredients that are believed unrelated. Machine learning focused on predicting new information based on known data could also be a useful tool to discover new possibilities that might not exist yet, or coupled with sensory analysis and preference data, might be able to predict new combinations of ingredients and procedures to brew new beer types targeted to certain customer groups.

Relating sensory experiences to chemical spectra is another important area of study with regard to machine learning. Using large data sets of chemical and sensory data on beer samples, we may be able to train our analysis algorithms to interpret the chemical spectra in similar ways to how a human experiences beer. This could lead to highly valuable discoveries when coupled with preference data and other metadata on how we experience beer. User data from applications like *Untapped* who collect information on peoples experience of beers may aid in such endeavours. The possibilities of the use of computational algorithms in beer studies are just speculation as of yet, however, we believe they will have an important role in the future of this project.

Manually targeted analysis will also be an important part for future studies using this database. Exploratory statistical efforts like those in this thesis will allow us to further study the trends we discover using computer algorithms, or statistically evaluate findings from targeted experiments using in-house samples.

In conclusion, I believe there is potential in this line of research and the project as a whole, although it may come with many challenges. It is my hope that the groundwork done in this thesis will be a solid foundation for the future development of a valuable and flexible research platform that will bring new understanding to the ancient art and science of beer brewing.

# Bibliography

1. Meussdoerffer, F. G. *A Comprehensive History of Beer Brewing* ISBN: 978-3-527-31674-8 (2009).
2. Ritchie Hannah, R. M. *Alcohol Consumption* 2018.
3. De Gaetano, G. *et al.* Effects of moderate beer consumption on health and disease: A consensus document. *Nutrition, Metabolism and Cardiovascular Diseases* **26**, 443–467. ISSN: 15903729 (June 2016).
4. Spaggiari, G., Cignarelli, A., Sansone, A., Baldi, M. & Santi, D. To beer or not to beer: A meta-analysis of the effects of beer consumption on cardiovascular health. *PLoS ONE* **15**, e0233619. ISSN: 19326203 (June 2020).
5. The Business research company. *Global Beer Market Data And Industry Growth Analysis*
6. Toro-González, D., McCluskey, J. J. & Mittelhammer, R. C. Beer Snobs Do Exist: Estimation of Beer Demand by Type. *Journal of Agricultural and Resource Economics* **39**, 174–187. ISSN: 10685502 (2014).
7. Poelmans, E. & Swinnen, J. F. in *The Economics of Beer* (Oxford University Press, Jan. 2012). ISBN: 9780191731884.
8. Lindow, J. *Norse Mythology: A Guide to Gods, Heroes, Rituals, and Beliefs* 47–326. ISBN: 0199839697 (Oxford University Press, 2002).
9. Moir, M. *Hops - A millennium review* 2000.
10. Gourvish, T. & Wilson, R. G. *The Dynamics of the Modern Brewing Industry* (ed Gourvish Terry, W. R. G.) ISBN: 1134756119 (Routledge, 1998).
11. Stewart, G. G. & Priest, F. G. in *Handbook of Brewing* 54–91 (CRC Press, Dec. 2006).
12. Boulton, C. & Quain, D. in *Brewing Yeast and Fermentation* 19–68 (Blackwell Science Ltd, Dec. 2007).
13. Steensels, J. *et al.* Improving industrial yeast strains: Exploiting natural and artificial diversity. *FEMS Microbiology Reviews* **38**, 947–995. ISSN: 15746976 (Sept. 2014).
14. Verbelen, P. J. *et al.* Impact of pitching rate on yeast fermentation performance and beer flavour. *Applied Microbiology and Biotechnology* **82**, 155–167. ISSN: 01757598 (Feb. 2009).
15. Vanderhaegen, B., Neven, H., Verachtert, H. & Derdelinckx, G. *The chemistry of beer aging - A critical review* Apr. 2006.
16. Vanderhaegen, B., Delvaux, F., Daenen, L., Verachtert, H. & Delvaux, F. R. Aging characteristics of different beer types. *Food Chemistry* **103**, 404–412. ISSN: 03088146 (Jan. 2007).

- 
17. Marconi, O., Rossi, S., Galgano, F., Sileoni, V. & Perretti, G. Influence of yeast strain, priming solution and temperature on beer bottle conditioning. *Journal of the Science of Food and Agriculture* **96**, 4106–4115. ISSN: 10970010 (Sept. 2016).
  18. Levitt, M. H. *Spin Dynamics: Basics of Nuclear Magnetic Resonance* 2nd. ISBN: 978-0-470-51117-6 (Wiley, 2008).
  19. Duarte, I. F., Barros, A., Almeida, C., Spraul, M. & Gil, A. M. Multivariate Analysis of NMR and FTIR Data as a Potential Tool for the Quality Control of Beer. *Journal of Agricultural and Food Chemistry* **52**, 1031–1038. ISSN: 00218561 (Mar. 2004).
  20. Nord, L. I., Vaag, P. & Duus, J. Quantification of organic and amino acids in beer by <sup>1</sup>H NMR spectroscopy. *Analytical Chemistry* **76**, 4790–4798. ISSN: 00032700 (Aug. 2004).
  21. Broberg, A., Thomsen, K. K. & Duus, J. O. Application of nano-probe NMR for structure determination of low nanomole amounts of arabinoxylan oligosaccharides fractionated by analytical HPAEC-PAD. *Carbohydrate Research* **328**, 375–382. ISSN: 00086215 (Sept. 2000).
  22. Rodrigues, J. E. & Gil, A. M. *NMR methods for beer characterization and quality control* Dec. 2011.
  23. Baxter, E. & Hughes, P. *Beer: Quality, safety and nutritional aspects* (2001).
  24. Duarte, I. *et al.* High-resolution nuclear magnetic resonance spectroscopy and multivariate analysis for the characterization of beer. *Journal of Agricultural and Food Chemistry* **50**, 2475–2481. ISSN: 00218561 (Apr. 2002).
  25. Almeida, C. *et al.* Composition of beer by <sup>1</sup>H NMR spectroscopy: Effects of brewing site and date of production. *Journal of Agricultural and Food Chemistry* **54**, 700–706. ISSN: 00218561 (Feb. 2006).
  26. Rodrigues, J. E. *et al.* Quantification of organic acids in beer by nuclear magnetic resonance (NMR)-based methods. *Analytica Chimica Acta* **674**, 166–175. ISSN: 00032670 (Aug. 2010).
  27. Wishart, D. S. Quantitative metabolomics using NMR. *TrAC - Trends in Analytical Chemistry* **27**, 228–237. ISSN: 01659936 (Mar. 2008).
  28. Lachenmeier, D. W. *et al.* Quality control of beer using high-resolution nuclear magnetic resonance spectroscopy and multivariate analysis. *European Food Research and Technology* **220**, 215–221. ISSN: 14382377 (Feb. 2005).
  29. Savorani, F., Tomasi, G. & Engelsen, S. B. icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. *Journal of Magnetic Resonance* **202**, 190–202. ISSN: 10907807 (Feb. 2010).
  30. Dieterle, F., Ross, A., Schlotterbeck, G. & Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in <sup>1</sup>H NMR metabolomics. *Analytical Chemistry* **78**, 4281–4290. ISSN: 00032700 (July 2006).
  31. Van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K. & van der Werf, M. J. Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics* **7**, 1–15. ISSN: 14712164 (June 2006).
  32. Olkin, I. & Sampson, A. in *International Encyclopedia of the Social & Behavioral Sciences* 10240–10247 (Elsevier, Jan. 2001).
  33. Xia, J., Psychogios, N., Young, N. & Wishart, D. S. MetaboAnalyst: A web server for metabolomic data analysis and interpretation. *Nucleic Acids Research* **37**, W652–W660. ISSN: 03051048 (July 2009).
-

- 
34. Mishra, P., Singh, U., Pandey, C., Mishra, P. & Pandey, G. Application of student's t-test, analysis of variance, and covariance. *Annals of Cardiac Anaesthesia* **22**, 407. ISSN: 0971-9784 (2019).
  35. Abdi, H. & Williams, L. J. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics* **2**, 433–459. ISSN: 19395108 (July 2010).
  36. Szymańska, E., Saccenti, E., Smilde, A. K. & Westerhuis, J. A. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* **8**, 3–16. ISSN: 1573-3882 (June 2012).
  37. Andreopoulos, B., An, A., Wang, X. & Schroeder, M. A roadmap of clustering algorithms: finding a match for a biomedical application. *Briefings in Bioinformatics* **10**, 297–314. ISSN: 1467-5463 (Dec. 2008).
  38. Nielsen, F. *Hierarchical Clustering* 195–211 (2016).
  39. Wattenberg, M., Viégas, F. & Johnson, I. How to Use t-SNE Effectively. *Distill* **1**, e2 (Jan. 2017).
  40. Van Der Maaten, L. & Hinton, G. *Visualizing Data using t-SNE* tech. rep. (2008), 2579–2605.
  41. Defernez, M. & Colquhoun, I. J. Factors affecting the robustness of metabolite fingerprinting using <sup>1</sup>H NMR spectra. *Phytochemistry* **62**, 1009–1017. ISSN: 00319422 (Mar. 2003).
  42. Hand, D. J. *Principles of data mining in Drug Safety* **30** (Springer, Nov. 2007), 621–622.
  43. Jordan, M. I. & Mitchell, T. M. *Machine learning: Trends, perspectives, and prospects* July 2015.
  44. Noy, N. F. & Mcguinness, D. L. *Ontology Development 101: A Guide to Creating Your First Ontology* tech. rep. (Stanford University, 2001).
  45. Gruber, T. R. A translation approach to portable ontology specifications. *Knowledge Acquisition* **5**, 199–220. ISSN: 10428143 (June 1993).
  46. Zhou, X. *et al.* Ontology development for unified traditional Chinese medical language system. *Artificial Intelligence in Medicine* **32**, 15–27. ISSN: 09333657 (Sept. 2004).
  47. Bao, J. *et al.* *OWL 2 Web Ontology Language Document Overview (Second Edition)* Dec. 2012.
  48. Yu, B. *Database Fundamentals* tech. rep. (University of the District of Columbia, 2008).
  49. Groff, J. R. & Weinberg, P. N. *SQL: The Complete Reference* (ed Poulin, H.) ISBN: 0072118458 (Brandon A Nordin, 1999).
  50. De Souza, C. R. B., Redmiles, D., Cheng, L.-T., Millen, D. & Patterson, J. *Sometimes You Need to See Through Walls-A Field Study of Application Programming Interfaces in Proceedings of the 2004 ACM conference on Computer supported cooperative work - CSCW '04* (ACM Press, New York, New York, USA, 2004). ISBN: 1581138105.
  51. Zotchev, K. *High-Resolution 1D 1 H Nuclear Magnetic Resonance, Multivariate Analysis and Heteronuclear Single Quantum Coherence Spectroscopy for Comparing Different Sample Preparation Methods and Styles of Beer* PhD thesis (Norwegian University of Science and Technology, 2020).
  52. Monakhova, Y. B. *et al.* Application of automated eightfold suppression of water and ethanol signals in <sup>1</sup>H NMR to provide sensitivity for analyzing alcoholic beverages. *Magnetic Resonance in Chemistry* **49**, 734–739. ISSN: 07491581 (Nov. 2011).
-

- 
53. Tomasi, G., Savorani, F. & Engelsens, S. B. Icoshift: An effective tool for the alignment of chromatographic data. *Journal of Chromatography A* **1218**, 7832–7840. ISSN: 00219673 (Oct. 2011).
  54. Swersey, C. *et al.* *Brewers Association 2021 Beer Style Guidelines* tech. rep. (Brewers Association, 2021).
  55. Burnham, T. *et al.* *2018 CraftBeer.com Beer Styles Guide* tech. rep. (Brewers Association, 2018).
  56. Suzuki, K. *125th anniversary review: Microbiological instability of beer caused by spoilage bacteria* 2011.
  57. Zhang, Y., Jia, S. & Zhang, W. Predicting acetic acid content in the final beer using neural networks and support vector machine. *Journal of the Institute of Brewing* **118**, 361–367. ISSN: 20500416 (Dec. 2012).
  58. Tyrell, T. & Fischer, F. Influencing the organic acid profile of beer by application of adsorbent materials. *Journal of the Institute of Brewing* **120**, 459–466. ISSN: 20500416 (Oct. 2014).
  59. He, Y. *et al.* *Wort composition and its impact on the flavour-active higher alcohol and ester formation of beer - A review* June 2014.
  60. Khatib, A. *et al.* Application of two-dimensional J-resolved nuclear magnetic resonance spectroscopy to differentiation of beer. *Analytica Chimica Acta* **559**, 264–270. ISSN: 00032670 (Feb. 2006).
  61. Sud, M. *et al.* Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Research* **44**, D463–D470. ISSN: 13624962 (Jan. 2016).
  62. Rocca-Serra, P. *et al.* Databases and ontologies ISA software suite: supporting standards-compliant experimental annotation and enabling curation at the community level. *BIOINFORMATICS APPLICATIONS NOTE* **26**, 2354–2356 (2010).
  63. Wolstencroft, K. *et al.* SEEK: A systems biology data and model management platform. *BMC Systems Biology* **9**, 33. ISSN: 17520509 (July 2015).
  64. Bauch, A. *et al.* OpenBIS: A flexible framework for managing and analyzing complex data in biology research. *BMC Bioinformatics* **12**, 1–19. ISSN: 14712105 (Dec. 2011).

# Appendix

## Appendix A

# Recipes for brewing experiments

## A.1 Pitch rate samples base recipe

Generated 2021-02-09T14:20:26+01:00 - Brewfather - <https://web.brewfather.app>

**Brewfather**  
www.brewfather.app

### Belgian Witbier - 6.6%

#### Witbier

Author: Hoppy Indian Brewery

Type: All Grain

IBU : 16 (Tinseth)  
Color : 7 EBC  
Carbonation : 2.6 CO2-vol

Pre-Boil Gravity : 1.057 / 14 °P  
Original Gravity : 1.062 / 15.3 °P  
Final Gravity : 1.012 / 3.1 °P

#### Fermentables (12.2 kg)

5.5 kg - Pale Malt, Ale 3.9 EBC (45.1%)  
5.5 kg - Wheat Malt 3 EBC (45.1%)  
1.2 kg - Oats, Flaked 2.8 EBC (9.8%)

#### Hops (100.5 g)

60 min - 50.5 g - Hallertauer Mittelfrueh - 3...  
30 min - 50 g - Hallertauer Mittelfrueh - 3.6...

#### Miscellaneous

Mash - 20 ml - Lactic Acid 80%  
5 min - Boil - 20 g - Coriander Seed  
5 min - Boil - 50 g - Orange Peel, Sweet

#### Yeast

1 pkg - White Labs Bavarian Weizen Yeast WLP351  
1 pkg - White Labs Belgian Ale WLP550

#### Starter

Step 1: 3.9 L (383 g DME / 468 g LME)  
100 billion cells overbuild: 0.61 L  
3.29 L pitch amount  
635 billion yeast cells  
12 million cells / ml

#### Speidel Braumeister 50L

Batch Size : 46 L  
Boil Size : 55 L  
Post-Boil Vol : 50 L

Mash Water : 55 L  
Sparge Water : 7.56 L

Boil Time : 60 min  
Total Water : 62.56 L

Brewhouse Efficiency: 74%  
Mash Efficiency: 77.2%

#### Mash Profile

More dextrinous  
69 °C - 60 min - Temperature  
75 °C - 10 min - Mashout

#### Fermentation Profile

Ale  
20 °C - 4 days - Primary  
26 °C - 12 days - Secondary  
4 °C - 4 days - Cold Crash

#### Measurements

Mash pH:

Boil Volume:

Pre-Boil Gravity:

Post-Boil Kettle Volume:

Original Gravity:

Fermenter Top-Up:

Fermenter Volume:

Final Gravity:

Bottling Volume:



7 EBC

#### Recipe Notes

I used fresh 50gms orange peel, bit next time I will use atleast 100gms. You can also use dry orange peel.



## A.2 Alage-beer samples base recipes

Generated: 2020-10-06T13:49:53+02:00 - Brewfather - https://web.brewfather.app

### A Brown Ale recipe

**Brewfather**  
www.brewfather.app

#### Algae Brown Ale - 5%

##### Engelsk Brown Ale

Forfatter: Viggo

Type: Kornbasert

IBU : 17 (Tinseth)  
Farge : 45 EBC  
Karbonering : 4.6 g/L

Gravity før kok : 1.045 / 11.2 °P  
Original Gravity : 1.050 / 12.4 °P  
Final Gravity : 1.012 / 3.1 °P

##### Gjærbare ingredienser (9.88 kg)

7.1 kg - Pale Malt, Maris Otter 5.9 EBC (71.9%)  
880 g - Brown Malt 175 EBC (8.9%)  
800 g - Chateau Biscuit 50 EBC (8.1%)  
800 g - Dark Crystal Malt 220 EBC (8.1%)  
300 g - Pale Chocolate 600 EBC (3%)

##### Humle (20 g)

60 min - 20 g - Columbus/Tomahawk/Zeus (CTZ)...

##### Diverse

Mesk - 5.22 g - Calcium Chloride (CaCl2)  
Mesk - 8.28 g - Epsom Salt (MgSO4)  
Mesk - 5.4 g - Gypsum (CaSO4)  
15 min - Kok - 1.8 stk - Servomyces  
15 min - Kok - 3.6 g - Yeast Nutrients  
10 min - Kok - 9 g - Irish Moss

##### Speidel Braumeister 50L

Størrelse på brygg: 45 L  
Kokestørrelse : 53.96 L  
Volum etter kok : 48.96 L

Meskevann : 55 L  
Skyllevann : 4.71 L

Koketid : 60 min  
Total vannmengde : 59.71 L

Brygghuseffektivitet: 74%  
Meskeeffektivitet: 77.3%

##### Meskeprofil

Medium fermentability  
66 °C - 90 min - Temperatur  
78 °C - 15 min - Mash out

##### Gjæringsprofil

Ale  
18 °C - 14 dager - Primær  
20 °C - 3 dager - Primær  
3 °C - 3 dager - Cold Crash  
3 °C - 1 dager - Karbonering  
3 °C - 14 dager - Modning

##### Vannprofil

Trondheim (Kilde: Øltid) (Style - N. English...  
Ca 45 Mg 9 Na 4 Cl 30 SO 61

SO/Cl-forhold: 2  
Mesk pH: 5.33

##### Målinger

Mesk pH:

Kokevolum:

Gravity før kok:

Volum i kjele etter kok:

Original Gravity:

Gjæringskar-påfyll:

Gjæringskar-volum:

Final Gravity:

Flaskevolum:



45 EBC

## B IPA recipe

Generated 2020-10-06T13:51:20+02:00 - Brewfather - <https://web.brewfather.app>

**Brewfather**  
www.brewfather.app

### Algae IPA - 5.6%

#### Amerikansk IPA

Forfatter: Trond Viggo Pettersen

Type: Kornbasert

IBU : 45 (Tinseth)  
Farge : 21 EBC  
Karbonering : 4.7 g/L

Gravity før kok : 1.054 / 13.3 °P  
Original Gravity : 1.059 / 14.5 °P  
Final Gravity : 1.016 / 4.1 °P

#### Gjærbare ingredienser (11.8 kg)

7 kg - Chateau Pilsen 2-Row 3.5 EBC (59.3%)  
2.8 kg - Carahell 25.5 EBC (23.7%)  
2 kg - Carared 47.5 EBC (17%)

#### Humle (302 g)

60 min - 37 g - Columbus/Tomahawk/Zeus (CTZ)...  
5 min - 25 g - Amarillo - 9.2% (3 IBU)  
5 min - 20 g - Simcoe - 12.8% (4 IBU)  
5 min - 20 g - Sorachi Ace - 13% (4 IBU)

#### Hop Stand

10 min hopstand @ 80 °C  
10 min - 35 g - Amarillo - 9.2% (1 IBU)  
10 min - 30 g - Cascade - 5.5% (1 IBU)  
10 min - 30 g - Centennial - 10% (1 IBU)  
10 min - 20 g - Simcoe - 12.8% (1 IBU)

#### Tørrhumle

5 dager - 25 g - Simcoe - 12.8%  
5 dager - 20 g - Amarillo - 9.2%  
5 dager - 20 g - Cascade - 5.5%  
5 dager - 20 g - Centennial - 10%

#### Speidel Braumeister 50L

Størrelse på brygg: 45 L  
Kokestørrelse : 53.96 L  
Volum etter kok : 48.96 L

Meskevann : 55 L  
Skyllevann : 6.24 L

Koketid : 60 min  
Total vannmengde : 61.24 L

Brygghuseffektivitet: 74%  
Meskeeffektivitet: 77.3%

#### Meskeprofil

4 trinn  
63 °C - 15 min - Temperatur  
67 °C - 15 min - Temperatur  
70 °C - 60 min - Temperatur  
78 °C - 15 min - Temperatur

#### Gjæringsprofil

Ale  
18 °C - 14 dager - Primær  
20 °C - 4 dager - Primær  
2 °C - 3 dager - Cold Crash  
2 °C - 1 dager - Karbonering  
10 °C - 14 dager - Modning

#### Målinger

Mesk pH:

Kokevolum:

Gravity før kok:

Volum i kjele etter kok:

Original Gravity:

Gjæringskar-påfyll:

Gjæringskar-volum:

Final Gravity:

Flaskevolum:



21 EBC

## C Goose recipe

Generated 2020-10-06T13:48:05+02:00 - Brewfather - <https://web.brewfather.app>

**Brewfather**  
www.brewfather.app

### Algae Tornwood Goose - 5.1%

#### Gose

Forfatter: Trond Viggo Pettersen

Type: Kornbasert

IBU : 8 (Tinseth)  
Farge : 7 EBC  
Karbonering : 5.8 g/L

Gravity før kok : 1.043 / 10.7 °P  
Original Gravity : 1.048 / 11.8 °P  
Final Gravity : 1.009 / 2.3 °P

#### Gjærbare ingredienser (9.1 kg)

4.5 kg - Chateau Wheat Blanc Nature 4.5 EBC (...  
4.4 kg - Pilsen Malt 3.5 EBC (48.4%)  
200 g - Acidulated 6 EBC (2.2%)

#### Humle (45.1 g)

60 min - 19.6 g - Hallertauer Mittelfrueh - 3...  
30 min - 9.8 g - Hallertauer Mittelfrueh - 3....  
15 min - 15.7 g - Hallertauer Mittelfrueh - 3...

#### Diverse

Mesk - 1 pkg - WildBrew Sour Pitch  
Mesk - 10 g - Calcium Chloride (CaCl2)  
Mesk - 1.3 g - Epsom Salt (MgSO4)  
Mesk - 1.4 g - Gypsum (CaSO4)  
Mesk - 10 ml - Lactic Acid 80%  
Mesk - 3 g - Magnesium Chloride (MgCl2)  
15 min - Kok - 2 stk - Servomyces  
15 min - Kok - 1 stk - Whirlfloc  
15 min - Kok - 4 g - Yeast Nutrients (WLN1000)  
Kokeslutt - 150 g - Algae - freeze dried  
Primær - 1500 g - Tindved - Thornwood

#### Gjær

1 pkg - White Labs California Ale WLP001

#### Starter

Steg 1: 2.2 L (216 g DME / 264 g LME)  
401 milliarder gjærceller  
9 millioner celler / ml

#### Speidel Braumeister 50L

Størrelse på brygg: 45 L  
Kokestørrelse : 53.96 L  
Volum etter kok : 48.96 L

Meskevann : 55 L  
Skyllevann : 4.08 L

Koketid : 60 min  
Total vannmengde : 59.08 L

Brygghuseffektivitet: 74%  
Meskeeffektivitet: 77.3%

#### Meskeprofil

Max fermentability with highest yield

63 °C - 30 min - Temperatur  
70 °C - 60 min - Temperatur  
78 °C - 15 min - Mash out

#### Gjæringsprofil

Lager  
18 °C - 14 dager - Primær  
20 °C - 3 dager - Primær  
3 °C - 3 dager - Cold Crash  
3 °C - 1 dager - Karbonering  
3 °C - 14 dager - Modning

#### Vannprofil

Trondheim (Kilde: Øltid) (Style - Fruit Lambic)  
Ca 40 Mg 9 Na 4 Cl 50 SO 25

SO/Cl-forhold: 0.5  
Mesk pH: 5.45

#### Målinger

Mesk pH:

Kokevolum:

Gravity før kok:

Volum i kjele etter kok:

Original Gravity:

Gjæringskar-påfyll:

Gjæringskar-volum:

Final Gravity:

Flaskevolum:



7 EBC

## D Amber Lager recipe

Generated 2020-10-06T13:49:15+02:00 - Brewfather - https://web.brewfather.app

**Brewfather**  
www.brewfather.app

### Amber Algae Lager - 4.9%

European-Style Dark Lager  
Forfatter: Trond Viggo Pettersen

Type: Kornbasert

IBU : 23 (Tinseth)  
Farge : 33 EBC  
Karbonering : 4.7 g/L

Gravity før kok : 1.047 / 11.7 °P  
Original Gravity : 1.051 / 12.7 °P  
Final Gravity : 1.014 / 3.6 °P

#### Gjærbare ingredienser (10.2 kg)

6.5 kg - Chateau Pilsen 2-Row 3.5 EBC (63.7%)  
2.5 kg - Carared 47.5 EBC (24.5%)  
1.2 kg - Triple Melanoidin 197 EBC (11.8%)

#### Humle (27 g)

60 min - 27 g - Columbus/Tomahawk/Zeus (CTZ)...

#### Diverse

15 min - Kok - 2 stk - Servomyces  
15 min - Kok - 1 stk - Whirlfloc  
15 min - Kok - 5 g - Yeast Nutrients (WLN1000)  
5 min - Kok - 50 g - Black pepper - crushed

#### Gjær

1000 ml - Dahls yeast

#### Speidel Braumeister 50L

Størrelse på brygg: 45 L  
Kokestørrelse : 53.96 L  
Volum etter kok : 48.96 L

Meskevann : 55 L  
Skyllevann : 4.96 L

Koketid : 60 min  
Total vannmengde : 59.96 L

Brygghuseffektivitet: 74%  
Meskeeffektivitet: 77.3%

#### Meskeprofil

4 trinn

63 °C - 15 min - Temperatur  
67 °C - 15 min - Temperatur  
70 °C - 60 min - Temperatur  
78 °C - 15 min - Temperatur

#### Gjæringsprofil

Lager

12 °C - 14 dager - Primær  
14 °C - 10 dager - Primær  
3 °C - 7 dager - Cold Crash  
3 °C - 1 dager - Karbonering  
3 °C - 14 dager - Modning

#### Målinger

Mesk pH:

Kokevolum:

Gravity før kok:

Volum i kjele etter kok:

Original Gravity:

Gjæringskar-påfyll:

Gjæringskar-volum:

Final Gravity:

Flaskevolum:



33 EBC

## Appendix B

# Matlab scripts for alignment, normalization and binning

### B.1 Alignment steps

Collection of steps for alignment compiled by Leesa J. Klau.

```
% Notes for processing beer 1H NMR spectra using icoshift
% buffered samples, with double suppression (H2O and EtOH) used as examples
%
%
% SUMMARY OF DATASETS
% -----
% filenames:
%   noesygpps1d_13Cdec_buff.mat
%   noesygpps1d_13Cdec_unbuff.mat
%   noesygpps1d_buff.mat
%   noesygpps1d_unbuff.mat
%
% experiment type
%   noesygpps1d_13Cdec:   double suppression (H2O and EtOH)
%   noesygpps1d:         single suppression (H2O)
% sample type
%   buff:                 buffered samples
%   unbuff:               unbuffered samples
%
% contents of each dataset:
%-----
% barcode:               barcode ID number for each sample
% buff_ints:             defined intervals used for alignment for the buffered
→ samples
% ppm:                   x-axis scale for [] in ppm
% ppm_corr:              ppm scale corrected using TSP reference peak
% TSPppm:                co-ordinates of TSP peak in [XcoshiftTSP] (used to correct
→ ppm scale)
```

---

```

% unbuff_ints:  defined intervals used for alignment for the buffered
→ samples
% X:           raw data, samples x ppm
% Xaligned:   aligned data, samples x ppm (use with ppm_corr)
% Xcoshift:   coshifted data, samples x ppm
% XcoshiftTSP: coshifted and referenced to TSP, samples x ppm (use with
→ ppm_corr)
%
%
% ANALIGNMENT STEPS
%-----
%
% 1. load the data file - add filename
% -----
% datafile should consist of:
% ppm : row vector of the ppm scale
% X : matrix of raw data, samples x ppm
%
load('noesygpps1d_13Cdec_buff.mat')
%
%
% to generate plot of raw data:
% -----
%
plot (ppm, X'); set(gca, 'XDir', 'reverse')
%
%
% 2. create a matrix of COshifted data
% -----
% this is an alignment based on the whole spectrum. the best mode for the
% buffered samples seems to be 'average2' with a multiplier of 3. for the
% unbuffered samples the best seems to be using 'median'. also using the
% 'b' setting: algorithm search for the best 'n' for each interval
%
% buffered samples
[Xcoshift] = icoshift ('average2',X,'whole','b',[2 1 0],ppm);
%
% unbuffered samples
[Xcoshift] = icoshift ('median',X,'whole','b',[2 1 0],ppm);
%
%
% 3. align coshifted data using TSP reference region
% -----
% define the TSP region from the coshifted matrix.
% Note:
%   in buffered samples, seems there are a couple that affect the TSP
%   peak? Therefore only used the main peak (on the right) for the TSP
%   alignment.
%   in unbuffered samples the TSP peak seems more widely affected by the
%   the individual sample?

```

---

---

```

%
%
% buffered samples
[XcoshiftTSP] = icoshift ('average2',Xcoshift,(52079:52730),'b',[2 1
→ 0],ppm);
%
% unbuffered samples
[XcoshiftTSP] = icoshift ('average2',Xcoshift,(52340:52443),'b',[2 1
→ 0],ppm);
%
%
%
% 4. compute a TSP corrected ppm axis
% -----
%
% plot the TSP corrected matrix data
plot (ppm, XcoshiftTSP'); set(gca, 'XDir', 'reverse')
%
% select the TSP peak using the "data tips" tool
% right-click, select "export curser data to workspace"
% name "TSPppm"
% compute corrected ppm scale:
ppm_corr = ppm - TSPppm.Position(1);
%
% plot with corrected scale
plot (ppm_corr, XcoshiftTSP'); set(gca, 'XDir', 'reverse')
%
%
% 5. define intervals and align coshifted data
% -----
% use the coshifted and TSP aligned matrix (XcoshiftTSP) generated from
% previous steps as a starting point. use corrected ppm scale (ppm_corr).
% best mode for alignment seems to be 'average2' with multiplier of 3.
% select missaligned regions and manually define intervals
% (buff_ints/unbuff_ints). this uses the column numbers from the
% ppm/ppm_corr vectors and therefore is the same for both scales. the
% buffered and unbuffered intervals differ slightly. They have the same
% number of intervals and same approximate positions but have been moved
% to accomodate approximately the same signals between both datasets
%XcoshiftTSP
%
% buffered samples
[Xaligned] = icoshift ('average2',XcoshiftTSP,buff_ints,'b',[2 1
→ 0],ppm_corr);
%
% unbuffered samples
[Xaligned] = icoshift ('average2',XcoshiftTSP,unbuff_ints,'b',[2 1
→ 0],ppm_corr);
%
%
%

```

---

---

## B.2 Normalization script

Script made by Christian Schulz.

```
function [normalised, ppmNew] = beernormal(inputdata, ppm, excludeAreas,
→ varargin)
% This function normalises data using pqn based on
% [1] Dieterle, F et al., Analytical Chemistry 78(13), 4281-4290
% Additionally, it excludes areas from the data, such as ethanol and water
% areas before the normalisation
% The function truncates the spectrum automatically: Typically only from
→ -0.5
% to 11 will be used.
% Furthermore, negative values will be eliminated.
%
% Scaling will be performed based on
% [2] R. A. van den Berg, H. C. Hoefsloot, J. A. Westerhuis, A. K. Smilde,
→ M. J. van der Werf, BMC Genomics. 7, 142 (2006).
% The options are:
% 'auto', 'pareto', 'vast', 'level' or 'range'
% which should be given as variable arguments.
%
%
% Example usage:
% excludeAreas = {{1.152,1.247},{3.608,3.681},{4.751,4.928}};
% Used (adrian): excludeAreas =
→ {{-0.0501,0.0501},{1.0902,1.2902},{3.6151,3.6500},{4.7001,4.8500}};
%
% [normalised, ppmNew] = beernormal(Xaligned, ppm_corr, excludeAreas,
→ 'scale', 'pareto');
% This removes the unwanted areas, truncates the data, normalises it and
% scales it according to pareto scaling.
%
% [normalised, ppmNew] = beernormal(Xaligned, ppm_corr);
% This normalises data, truncating it, without removing further areas nor
% scaling the data.
%
%
% HOWEVER: Scaling and removal of water/ethanol peaks is recommended!
%
%
% Christian 22.04.2020
%

% ppm truncate values:
ppmscaleEnd = 9;
ppmscaleStart = 0.8;
```



---

```

%% check inputs and add the truncates to possible excluded areas.

% Check ppm scale
if numel(ppm)~=size(inputdata,2)
    error('The ppm scale and input data do not match in length.')
end

% generating and checking the ppm truncats
startPPM = ppm(end);
endPPM = ppm(1);
if startPPM > endPPM
    error('Something wrong with the ppm scale (maybe inverted?).')
end
if ppmscaleEnd > endPPM
    error('The ppm scale is shorter than 11. Fix this script in the start
    ↪ for the truncation.')
end
if ppmscaleStart < startPPM
    error('The ppm scale is shorter than -0.5. Fix this script in the start
    ↪ for the truncation.')
end
addlisting = {[endPPM,ppmscaleEnd],[ppmscaleStart,startPPM]};

% Adding the ppm truncates to the list
if exist('excludeAreas', 'var')
    excludeAreas{end+1} = addlisting{1};
    excludeAreas{end+1} = addlisting{2};
else
    excludeAreas = addlisting;
end

% actually checking the excluded areas
for ii = 1:numel(excludeAreas)
    if numel(excludeAreas{ii}) ~= 2
        fprintf(2,"Error in excluded Areas: Area %i has not two ppm values
        ↪ defined.\n", ii);
        error('Returning.')
    end
end
inpLoc = excludeAreas;
for ii = 1:numel(excludeAreas)
    for jj = 1:numel(excludeAreas{ii})
        inpLoc{ii}{jj} = find(ismembertol(ppm,excludeAreas{ii}{jj},
        ↪ 0.00001));    % TO adjust for non exactly values in the ppm
        ↪ scale, since they have more digits
        if numel(inpLoc{ii}{jj}) ~=1
            fprintf("ppm couple:%i value: %i\n", ii, jj);
            error('Given ppm values not in ppm list. Maybe there is an
            ↪ error with the accuracy (NMR ppm are NOT EXACTLY)')
        end
    end
end

```

---

---

```

    end
end

% Checking input for rhe scaling
scaleType = [];
% Valid options for scaling
options = {'auto','pareto','vast','level','range','paretoNC'};

% Check input type
for j = 1:length(varargin)
    if isequal(varargin{j},'scale')
        for n = 1:length(options)
            if isequal(varargin{j+1},options{n})
                scaleType = n;
            end
        end
    end
end

end

if isempty(scaleType)
    scaling = false;
    fprintf(2,"No scaling will be performed!\n");
else
    scaling = true;
end

%% remove the exctration areas from the matrix

RemovedMatrix = inputdata;
ppmNew = ppm;
removeline = [];
for ii = 1:numel(excludeAreas)
    num = [inpLoc{ii}{1} inpLoc{ii}{2}];
    minnum = min(num);
    maxnum = max(num);
    numbers = [minnum:maxnum];
    removeline = [removeline,numbers];
end
RemovedMatrix(:,removeline) = [];
ppmNew(:,removeline) = [];

%% Adjusting for negative values

RemovedMatrix(RemovedMatrix<0) = realmin;
% This uses the smales value in matlab (2.225073858507201e-308) instead of
% a negative value.

%% normalistion using pqn

[row, col]=size(RemovedMatrix);

```

---

---

```

% Perform an integral normalization (typically a constant integral of 100
↳ is used).
integral=sum(RemovedMatrix,2);
integral_normed = 100*RemovedMatrix./repmat(integral,1,col); %now each
↳ spectra sum to 100
%choose a reference or template spectrum, am selecting the median
ref=median(integral_normed);

% Addition here
% Apparently, a 0 in the ref is a problem...
% Therefore, we will manually check it
if sum(ismember(ref, 0))>0
    posnill = find(ismember(ref, 0));
    for ii = 1:numel(posnill)
        med = median(integral_normed(posnill(ii)));
        if med ~=0
            ref(posnill(ii)) = med;
        else
            ref(posnill(ii)) =
↳ median([ref(posnill(ii)-1),ref(posnill(ii)+1)]);
        end
    end
end
end

% Calculate the quotients of all variables of interest of the test spectrum
↳ with those of the reference spectrum.
quotients = integral_normed./repmat(ref,row,1);
% Calculate the median of these quotients.
median_q = median(quotients,2);
% Divide all variables of the test spectrum by this median.
data = integral_normed./repmat(median_q,1,col);

%% scale data

% Handle scale type
if scaling
    if scaleType == 1 % auto scaling
        scaledData = zscore(data);
    elseif scaleType == 2 % pareto scaling
        for j = 1:size(data,2)
            scaledData(:,j) =
↳ (data(:,j)-mean(data(:,j)))./sqrt(std(data(:,j)));
        end
    elseif scaleType == 3 % level scaling
        for j = 1:size(data,2)
            scaledData(:,j) = (data(:,j)-mean(data(:,j)))./mean(data(:,j));
        end
    elseif scaleType == 4 % vast scaling
        for j = 1:size(data,2)

```

---

---

```

        scaledData(:,j) =
            ↪ ((data(:,j)-mean(data(:,j)))./std(data(:,j))).*(mean(data(:,j))./std(data(:,
    end
elseif scaleType == 5 % range scaling
    for j = 1:size(data,2)
        scaledData(:,j) =
            ↪ (data(:,j)-mean(data(:,j)))./(max(data(:,j))-min(data(:,j)));
    end
elseif scaleType == 6 % paretoNC scaling
    for j = 1:size(data,2)
        scaledData(:,j) = (data(:,j))./sqrt(std(data(:,j)));
    end
end
normalised = scaledData;
else
    normalised = data;
end

%% plott the new graph

plot (ppmNew, normalised); set(gca, 'XDir', 'reverse')

```

### B.3 Binning script

Script made by Leesa J. Klau.

```

%%
%% BINNING NMR SPECTRA %%%
%
%
% will bin the ppm scale to predefined width
% takes the mean of the signal within the bin
% outputs an adjusted ppm scale, and matrix of mean or sum binned spectra
%
%
% required inputs
% -----
% first define [Bins] by:
% Bins = (minppm:width:maxppm)
% e.g. Bins = (-0.5:0.01:11); this will return bins with a width of 0.01
% ppm from the range -0.5 to 11 ppm
%
% inputspectra is the matrix of aligned spectra (samples x ppm) - this
% matrix will be transposed during binning
%
% the ppm scale that matches the inputspectra
%

```

---

```

% define n as the number of samples
%
% 'mode' either as 'mean' or 'sum'
%
% outputs
% -----
%
% ppmBin is the new ppm scale of the binned data
%
% spectraBin is the binned data with mean of each bin
%
%
%%
function [ppmBin,spectraBin]=binning(inputspectra, ppm, n, Bins, varargin)

spectra=inputspectra';
[~,~,loc]=histcounts(ppm,Bins);

Mode = [];
% Valid options for scaling
options = {'mean','sum'};

% Check input type
for j = 1:length(varargin)
    if isequal(varargin{j},'mode')
        for m = 1:length(options)
            if isequal(varargin{j+1},options{m})
                Mode = m;
            end
        end
    end
end

if isempty(Mode)
    binningstuff = false;
    fprintf(2,"No binning will be performed!\n");
else
    binningstuff = true;
end

if binningstuff

for i = 1:n
    if Mode == 1 % mean
        spectraBin(:,i)=accumarray(loc(:),spectra(:,i),[],@mean);
    elseif Mode == 2 % sum
        spectraBin(:,i)=accumarray(loc(:),spectra(:,i),[],@sum);
    end
    spectraBin = spectraBin;
end

```

---

```
end
ppmBin = 0.5*(Bins(1:end-1)+Bins(2:end));

ppmBin = ppmBin;

plot(ppm,inputspectra), hold on; plot(ppmBin,spectraBin,'r'); set(gca,
→ 'XDir', 'reverse')
```

## Appendix C

# R-scripts for statistical tests and plotting

### C.1 R-script for determining optimal clusters for K-means

```
library(factoextra)
library(NbClust)

df <- scale(example_filename)
head(df)

fviz_nbclust(df, kmeans, method = "wss") +
  geom_vline(xintercept = 4, linetype = 2)+
  labs(subtitle = "Elbow method")

fviz_nbclust(df, kmeans, method = "silhouette")+
  labs(subtitle = "Silhouette method")

set.seed(123)
fviz_nbclust(df, kmeans, nstart = 25, method = "gap_stat", nboot = 50)+
  labs(subtitle = "Gap statistic method")
```

### C.2 R-script for plotting and anova test of ethanol content

```
library(dplyr)
library(ggpubr)
library(readr)

# Importing data set
ethValues <- read.csv("Pitch Exp. Ethanol integration values.csv")

# Cleaning up data set for plotting
dplyr::sample_n(ethValues, 9)
```

---

```

levels(ethValues$PitchNr)

group_by(ethValues, Pitch_Rate) %>%
  summarise(
    count = n(),
    mean = mean(Percentage, na.rm = TRUE),
    sd = sd(Percentage, na.rm = TRUE)
  )

# Box plots
# ++++++
# Plot weight by group and color by group

ggboxplot(ethValues, x = "Pitch_Rate", y = "Percentage",
  color = "Pitch_Rate", palette = c("#00AFBB", "#E7B800",
  ↪ "#FC4E07"),
  order = c("Under", "Normal", "Over"),
  ylab = "Percent ethanol by volume", xlab = "Pitch Rate")

# Mean plot
# ++++++
# Plot weight by group
# Add error bars: mean_se
ggline(ethValues, x = "Pitch_Rate", y = "Percentage",
  add = c("mean_se", "jitter"),
  order = c("Under", "Normal", "Over"),
  ylab = "Percent ethanol by volume", xlab = "Pitch Rate")

# Formulating a linear model for anova
model1 <- lm(Percentage ~ Pitch_Rate, data = ethValues)

# Residual plots for model verification
plot(model1)

# Summary of anova
summary(model1)

```





---

```

    " -3.3112754 ]\n",
    "[12.71151739 12.7112729 12.71102841 ... -3.31078641 -3.31103091\n",
    " -3.3112754 ]\n"
  ]
}
],
"source": [
  "# Read and plot processed 1H bruker spectra\n",
  "# Also save them as .mat files\n",
  "# Gaston Courtade - 06.03.20\n",
  "\n",
  "import nmrglue as ng\n",
  "import numpy as np\n",
  "import matplotlib.pyplot as plt\n",
  "import seaborn as sns\n",
  "import scipy.io as sio\n",
  "\n",
  "\n",
  "# 3 spectra per sample, zgpr, noesygps1d H2O supp, noesygps1d H2O +
  ↪ EtOH supp\n",
  "# sample order: IPA, Lager, AustmannStout, SourStout, Tripp, Weiss\n",
  "# alternate between vortex/ultrasonication, reps 1,2,3\n",
  "# buffered samples: 74-181\n",
  "# unbuffered samples: 185-292\n",
  "# skip samples: 95-97 (LAG_B_S1), 203-205 (LAG_U_V1), 182-184
  ↪ (testshims)\n",
  "\n",
  "# change these three lines maybe\n",
  "start = 3\n",
  "end = 33\n",
  "skip = []\n",
  "filename = 'Pitch_Experiment.mat'\n",
  "# ints = [52333,52555] #makes an interval file\n",
  "\n",
  "# no need to edit below this, really\n",
  "spectra_list = []\n",
  "i = start\n",
  "while i < end+1:\n",
  "    if i not in skip:\n",
  "        spectra_list.append(i)\n",
  "    i+=3\n",
  "\n",
  "\n",
  "data_matrix = []\n",
  "ppm_list = []\n",
  "SF = 800.13\n",
  "for expno in spectra_list:\n",
  "\n",
  "    # nmrglue magic\n",
  "    dic, data = ng.fileio.bruker.read_pdata(f'{expno}/pdata/1')\n",
  "    udic = ng.bruker.guess_udic(dic, data, strip_fake=True)

```

---

---

```

"    udic[0]['car'] = 3760.61 # 4.7 ppm\n",
"    uc_1h = ng.fileiobase.uc_from_udic(udic, dim=0)\n",
"    ppm_1h = uc_1h.ppm_scale()\n",
"    print(ppm_1h)\n",
"    #add this data to the matrix\n",
"    data_matrix.append(data)\n",
"\n",
"    \n",
"#save to matlab \n",
"sio.savemat(filename, {'ppm':ppm_1h, 'X':data_matrix})\n"
]
},
{
"cell_type": "markdown",
"metadata": {},
"source": [
"*repeat for both datasets (noesygppsid H2O supp, and H2O+EtOH  

↪ supp)\n",
"*make sure that SR = 0Hz\n",
"\n",
"\n",
"1. make sure script location is in same folder as experiment  

↪ folders\n",
"2. run the script above with Shift + Enter\n",
"3. load the .mat file into MatLab and see beer_alignment.m for  

↪ processing details"
]
},
{
"cell_type": "code",
"execution_count": null,
"metadata": {},
"outputs": [],
"source": []
}
],
"metadata": {
"kernel_spec": {
"display_name": "Python 3",
"language": "python",
"name": "python3"
},
"language_info": {
"codemirror_mode": {
"name": "ipython",
"version": 3
},
"file_extension": ".py",
"mimetype": "text/x-python",
"name": "python",
"nbconvert_exporter": "python",

```

---

```
    "pygments_lexer": "ipython3",  
    "version": "3.8.5"  
  }  
},  
"nbformat": 4,  
"nbformat_minor": 4  
}
```

## Appendix E

# Manual for sample treatment and NMR acquisition

# Beer project NMR Manual

A. Antonsen, L. Klau, C. Schulz

February 2021

## **1 Introduction**

This manual provides instructions for taking beer samples through all the steps from sample collection to preparation and NMR-acquisition. Please note that this document is not a replacement for proper training, and that none of these procedures should be performed without the appropriate HSE training.

## **2 Working with beer samples**

When working with beer samples it is important to keep in mind that the samples are in a more or less active state of fermentation, and are quite sensitive to both contamination and change over time. Make sure to always work with proper sterile technique, and to keep samples frozen when not used. Prepared NMR-samples should be run within 24h of preparation.

### 3 Sample filtering

- Fill sterile syringe with beer sample and attach sterile filter.
  - Filter: 0.2  $\mu\text{m}$  pore size, 25mm diameter, luer lock (female).
- Press sample through filter into appropriately sized tube, you should have at least 2 mL filtered sample. Change filter during the process if necessary.
- Note sample ID/name and date on tube.

### 4 Sample preparation for filtered samples

- Thaw sample (if frozen).
- Place sample in ultrasonic bath for 10 minutes (with the tube lid open) for degassing.
- Move 720  $\mu\text{L}$  of the sample into an appropriately sized tube.
- Add 80  $\mu\text{L}$  buffered stock solution to the same tube.
  - Buffered stock solution: 500 mM sodium phosphate pH 7.4 buffer containing  $\text{D}_2\text{O}$  (99%) and TSP (1%).
  - If there is precipitate in the buffered stock solution, warm slightly in water/hand and vortex until mixed.
- Vortex shortly to mix.
- Move 600  $\mu\text{L}$  mixed sample into 5mm NMR-tube.
- Note barcode on tube lid and record this along with sample ID/name.
- Gently press the tube lid onto the NMR-tube until a click is heard.
- If bubbles are present in the tube, shake firmly to settle the sample in the bottom of the NMR-tube.
- Wipe the glass tube with clean, lint free cleaning paper to remove grease and debris that can affect the acquisition.
- Place the tube in the sample holder. Remember the placement of the different samples, this will make it easier to place in the NMR auto sampler later.

## 5 Software-interface overview

When operating the NMR-magnets, there are two main software interfaces you will use, these are Topspin and IconNMR. Topspin is the main software that controls the acquisition process, as well as lets us view and analyze our spectra after acquisition. Here you will not use Topspin directly to run samples, instead using the automation software IconNMR that lets us queue multiple samples with pre-made parameter settings in order to maintain the same procedure for all samples. The following figures will show an example of the interfaces of both applications with important sections annotated.

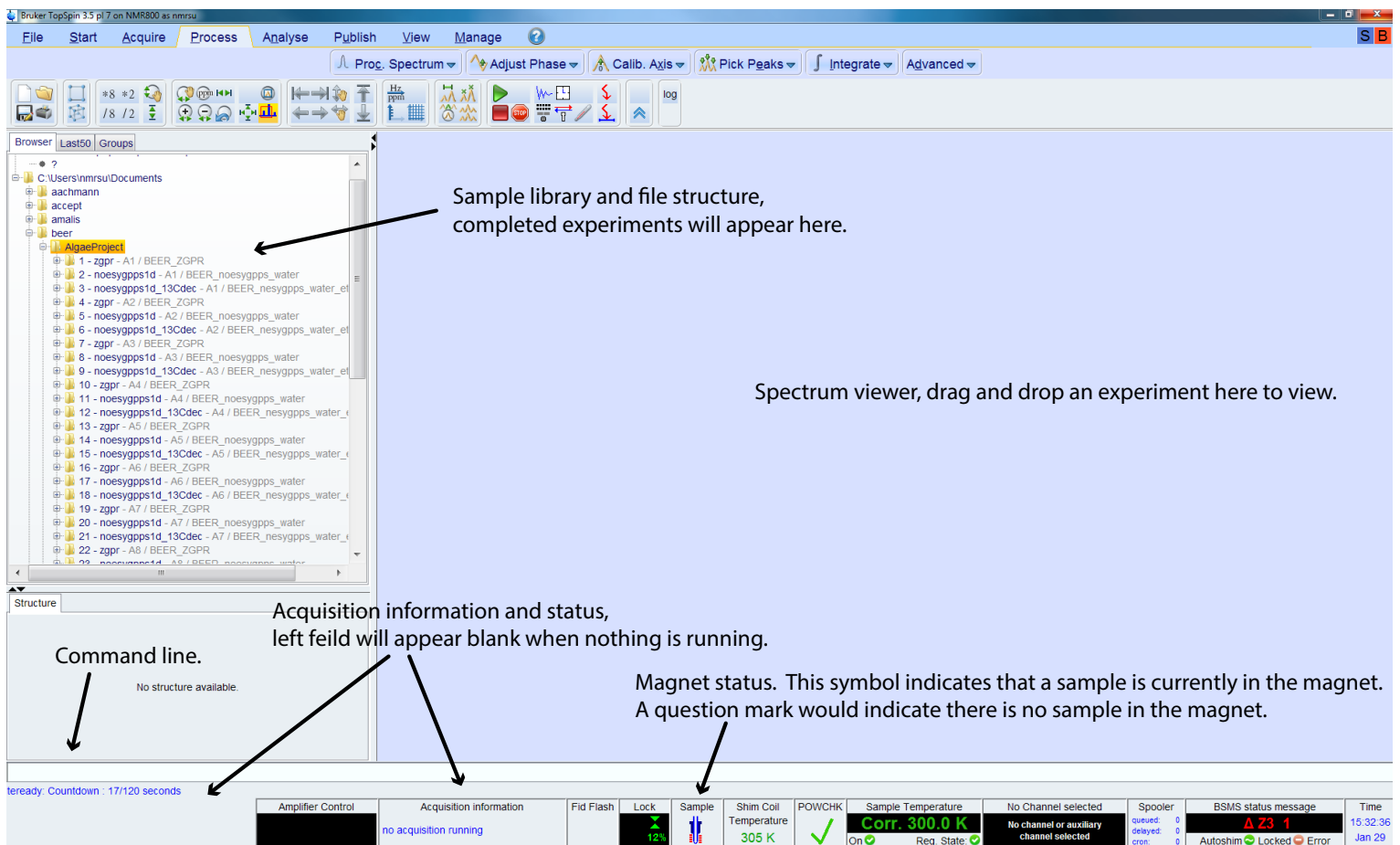


Figure 1: TopSpin interface overview.



Automation control panel.

Sample and experiment parameter settings.

Sample editing options.

Job status view.

User selector.

Holder	Type	Status	Name	No.	Solvent	Experiment	Pri	Par	Title/Orig	Time	User	Start Time
1 H11 - 195	U	Available										
1 H12 - 196	U	Available										
2 A1 - 201	U	Available										
2 A2 - 202	U	Available										
2 A3 - 203	U	Available										
2 A4 - 204	U	Available										
2 A5 - 205	U	Available										
2 A6 - 206	U	Available										
2 A7 - 207	U	Available										
2 A8 - 208	U	Available										
2 A9 - 209	U	Available										
2 A10 - 210	U	Available										
2 A11 - 211	U	Available										
2 A12 - 212	U	Available										
2 B1 - 213	U	Available										
2 B2 - 214	U	Available										
2 B3 - 215	U	Available										
2 B4 - 216	U	Available										
2 B5 - 217	U	Available										
2 B6 - 218	U	Available										
2 B7 - 219	U	Available										

#	Date	Tube ID	Holder	Name	No.	Experiment	Load	ATM	Lock	Shim	Acq	Proc	User	Disk	Title/Orig	Remarks
303	2021-01-27 19:11:52	5006090315	13	paenan	35	auto_COSY		✓	✓	✓	✓	✓	leesak	C:\Users\nmrsu\Documents\data\leesak\nmr	paenan variant dpepFJ heat-treated -10mg in 600uL 99.9% D2O 50030095697	log_err: This RG value has a
302	2021-01-27 18:39:38	5006090315	13	paenan	34	auto_1H_noesyPR	✓	✓	✓	✓	✓	✓	leesak	C:\Users\nmrsu\Documents\data\leesak\nmr	paenan variant dpepFJ heat-treated -10mg in 600uL 99.9% D2O 50030095697	
301	2021-01-15 20:22:47	5003004164	21	Jochen	10	auto_DQF_COSY		✓	✓	✓	✓	✓	aachmann	C:\Users\nmrsu\Documents\data\leesak\nmr	EPS #7 in D20 140121 DQF-COSY #5008105347 @ 21	Automation was stopped
300	2021-01-15 14:27:22	5003004164	21	Jochen	9	auto_C_HMBC		✓	✓	✓	✓	✓	aachmann	C:\Users\nmrsu\Documents\data\leesak\nmr	EPS #7 in D20 140121 13C HMBC #5008105347 @ 21	
299	2021-01-15 12:10:49	5003004164	21	Jochen	8	auto_H2BC		✓	✓	✓	✓	✓	aachmann	C:\Users\nmrsu\Documents\data\leesak\nmr	EPS #7 in D20 140121 13C H2BC #5008105347 @ 21	
298	2021-01-15 08:18:07	5003004164	21	Jochen	7	auto_C13C_HSQC_edHS QCs		✓	✓	✓	✓	✓	aachmann	C:\Users\nmrsu\Documents\data\leesak\nmr	EPS #7 in D20 140121 13C HSQC #5008105347 @ 21	
297	2021-01-15 07:55:52	5003004164	21	Jochen	6	auto_1H_ES	✓	✓	✓	✓	✓	✓	aachmann	C:\Users\nmrsu\Documents\data\leesak\nmr	EPS #7 D20 140121 1H NMR with excitation sculpting water suppression #5008105347 @ 21	
296	2021-01-15 02:44:06	5003004127	20	Jochen	5	auto_DQF_COSY		✓	✓	✓	✓	✓	aachmann	C:\Users\nmrsu\Documents\data\leesak\nmr	EPS #1 in D20 140121 DQF-COSY #50081015361 @ 20	

Figure 2: IconNMR interface overview.

## 6 NMR analysis procedure

- - Check that Topspin and IconNMR is running
  - Topspin can be started through the desktop icon for Topspin.
  - IconNMR can be opened by typing “iconnmr” in the topspin command field, and selecting the ”automation” option in the startup window, see figure 3.
- Check if acquisition is running in Topspin by reading the lower right-hand side of the Topspin interface, see figure 1.
- Check if automation is running in IconNMR by reading of the top left-hand side of the IconNMR interface. If the start button is active automation is off, if it’s inactive automation is running.
  - If acquisition and automation is running the experiments can be set up and queued as usual, with some exceptions in the procedure for placing samples in the autosampler, noted in section 7.
  - If acquisition is running but automation is not, the experiment can be set up, but not queued.
  - If no acquisition or automation is running, automation needs to be started before the samples can be run, see section 9.
  - If no acquisition is running, but automation is, you can proceed to the next steps.



Figure 3: IconNMR startup interface.

## 7 Sample setup in IconNMR

- Change to the correct user, in this case the “beer” user, if you did not do this when opening IconNMR.
- Select a set of available sample holders in the IconNMR interface, preferable in sequential order and in the same sample rack if possible.
- Make sure the holders are available by checking the physical rack.
  - If you are placing multiple samples, or replacing old ones, taking the rack out of the autosampler and into the computer room will make the next steps easier. Note that this can only be done when no acquisition or automation is running. For instructions on how to retrieve and replace the rack, see section 8 of this manual.
- Select the first sample holder in your sequence and click the add option in the lower panel. A line will appear containing four drop-down menus, as well as a few symbols, see figure 4.
  - \* Note: Be careful with scrolling while editing, as this will scroll through dropdown menus if hovering over one. Best practice is to always collapse a holder after editing and double check all experiments before submitting.
  - The first field indicates the directory, or folder, where the samples will be placed.
  - The second field indicates the experiment number within the directory. Each sample will create its own folder in the directory with the number specified in this field. Make sure these are unique and sequential. If you provide an experiment number that already exists in the directory the experiment will be skipped.
  - The third field indicates the solvent used in the sample. In our case you will use the “beer” preset.
  - The fourth field indicates what experiment to run. In our case you will use the “beer” presets here. There are four available, see table ?? for names and descriptions. If you are unsure of what to use, talk to your supervisor.
    - \* You may want to run multiple experiments on the same sample, to set this up simply press the “add” button on the lower panel while the current sample is highlighted and select a different experiment.
  - To the right of the six symbols there is a field where you can name your sample. Use sample ID or other identifier. It may also be useful to add the barcode of the NMR-tube in this field, this will appear in the “title” tab of the experiment, and can be done with the barcode scanner next to the computer.



Table 1: Name, description and usage of NMR experiments

Name in IconNMR from experiment drop down list (EXP)	Pulse program name (PUL-PROG)	Description	When to use
N auto_BEER_ZGPR	zgpr	1D proton spectrum with water suppression	This is a short experiment used to determine the position of the ethanol signals. After acquisition a processing script is automatically run that picks the ethanol peaks (a quartet at 3.6 ppm and triplet at 1.2 ppm) and generates a shape file (beer-shape)
N auto_BEER_NOESYGPPS_water	noesygpps1d	1D proton spectrum with water suppression	This experiment suppresses only the residual water signal, the ethanol signals remain.
N auto_BEER_NOESYGPPS_ethanol	noesygpps1d.13Cdec	1D proton spectrum with water suppression, ethanol suppression, and <sup>13</sup> C decoupling of ethanol CH <sub>2</sub> group at 3.6 ppm	This experiment uses the shape file generated in the first experiment and suppresses the ethanol signals. There is also <sup>13</sup> C decoupling but only for the CH <sub>2</sub> quartet (3.6 ppm)
N auto_BEER_13C-HSQC	hsqcedetgpsisp2.3	2D 1H- <sup>13</sup> C HSQC spectrum	This experiment is not required for all samples. It is used when confirmation of identified components is required.

**Routine:**

Routine acquire:

1. N auto\_BEER\_ZGPR
2. N auto\_BEER\_NOESYGPPS\_water
3. N auto\_BEER\_NOESYGPPS\_ethanol

No need to run the ethanol suppression experiment for samples that contain no ethanol (e.g. control samples). If 2D HSQC spectrum is required can acquire at the same time as the fourth experiment or can acquire later. If acquired later then use:

1. N auto\_BEER\_NOESYGPPS\_water
2. N auto\_BEER\_13C-HSQC

## 8 Autosampler

- On the 800-magnet you will find an autosampler, this mechanism will run our samples for us, we only need to place them in the right locations and give the right instructions.
- On the autosampler you will find a screen, check if the magnet is empty (indicated by two lines. A number on screen indicates a sample is running).
- Select wanted location by pressing “select rack”, then the rack number. The autosampler will now rotate the appropriate rack to the small door in the sampler. If a sample is running this rotation might take a while, make sure the door is closed while rotating.
- Lift the rack out carefully and place your samples in the desired location.
  - NOTE: Do not lift the rack out if automation is running. You may still place or remove samples while the rack is in the autosampler.
  - NOTE: If automation is running you will have a time limit to place/remove samples from the rack, this will be displayed on the autosampler screen, to extend this limit press the “more time” option on the screen.
  - Remove any old samples if necessary.
- Carefully replace the rack in the correct orientation, indicated by the slanting corner.
- Reset screen by pressing “system status”.

## 9 Submitting samples and starting automation

- When all samples are correctly entered in their sample holders they need to be submitted in IconNMR. To do this, highlight all the relevant sample holders by shift-clicking the first and last holder, and select the “submit” option in the lower panel.
- If automation is not running it must be started in IconNMR. Before you do this, make sure no acquisition is running in Topspin and that no sample is inside the magnet. Starting automation is done by selecting the “start” option in the top left-hand corner of IconNMR and selecting what sample to start on, this will be the first sample in your queue.
- Take note of run time and estimated time of completion in the bottom panel of IconNMR.

## **10 Removing and storing samples**

After your experiments are done, and you are have made sure all the acquisitions were successfully, you should remove your samples in order to make room for others using the machines. Remove physical samples from the auto sampler and delete them from IconNMR so the sample holder becomes available. Store the samples in a freezer or other appropriate location.

# Appendix F

## Supplementary plots for statistical results

### F.1 Validation plots for PLS-DA analysis

#### A Austmann samples

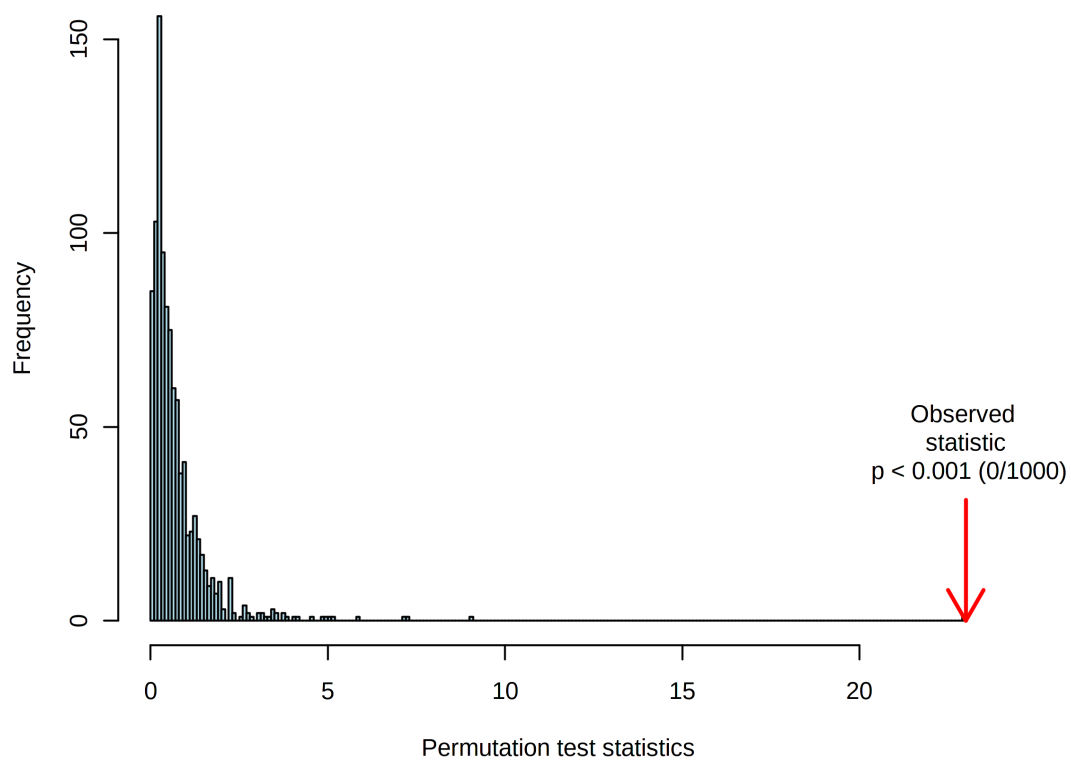


Figure F.1: Permutation test statistic for PLS-DA analysis of initial categorization test of samples from Austmann.



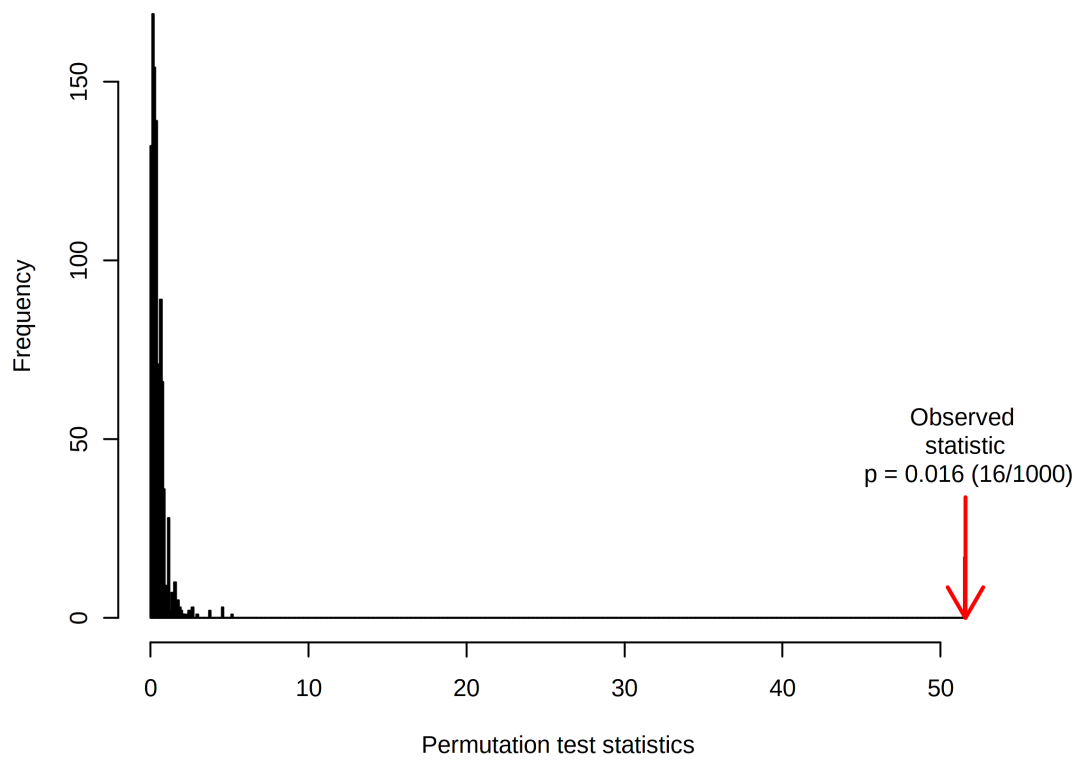


Figure F.2: Permutation test statistic for PLS-DA analysis of re-clustered categorization test of samples from Austmann.

---

## B Pitch rate samples

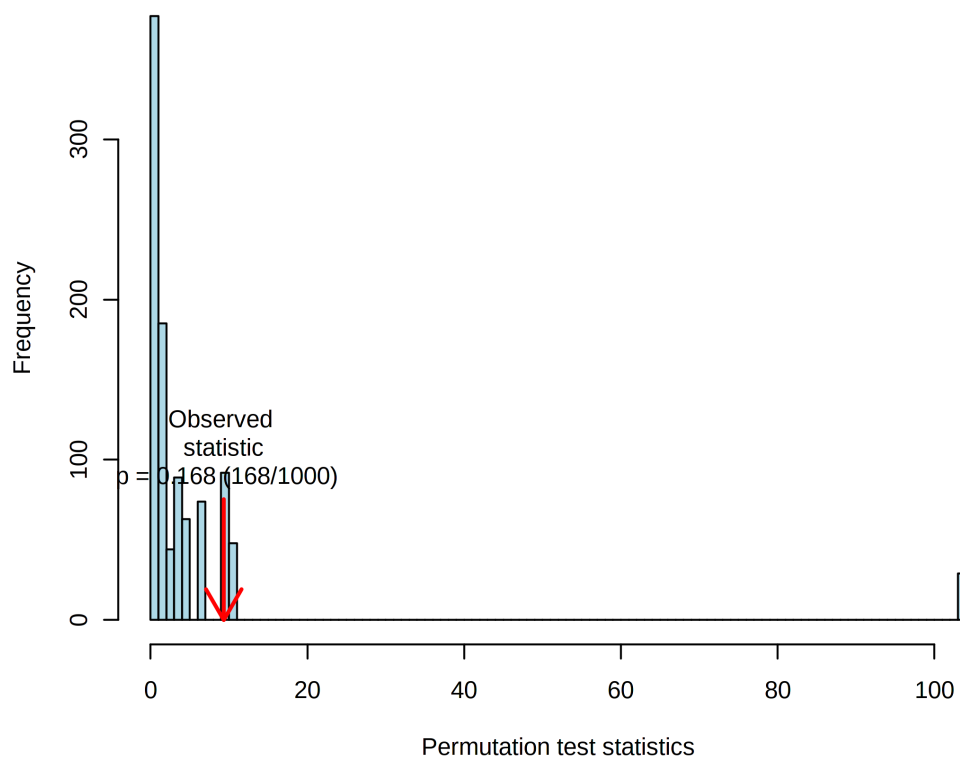


Figure F.3: Permutation test statistic for PLS-DA analysis of pitch rate study samples.

## F.2 Detailed spectrum and loading plots

### A Austmann samples loading plots

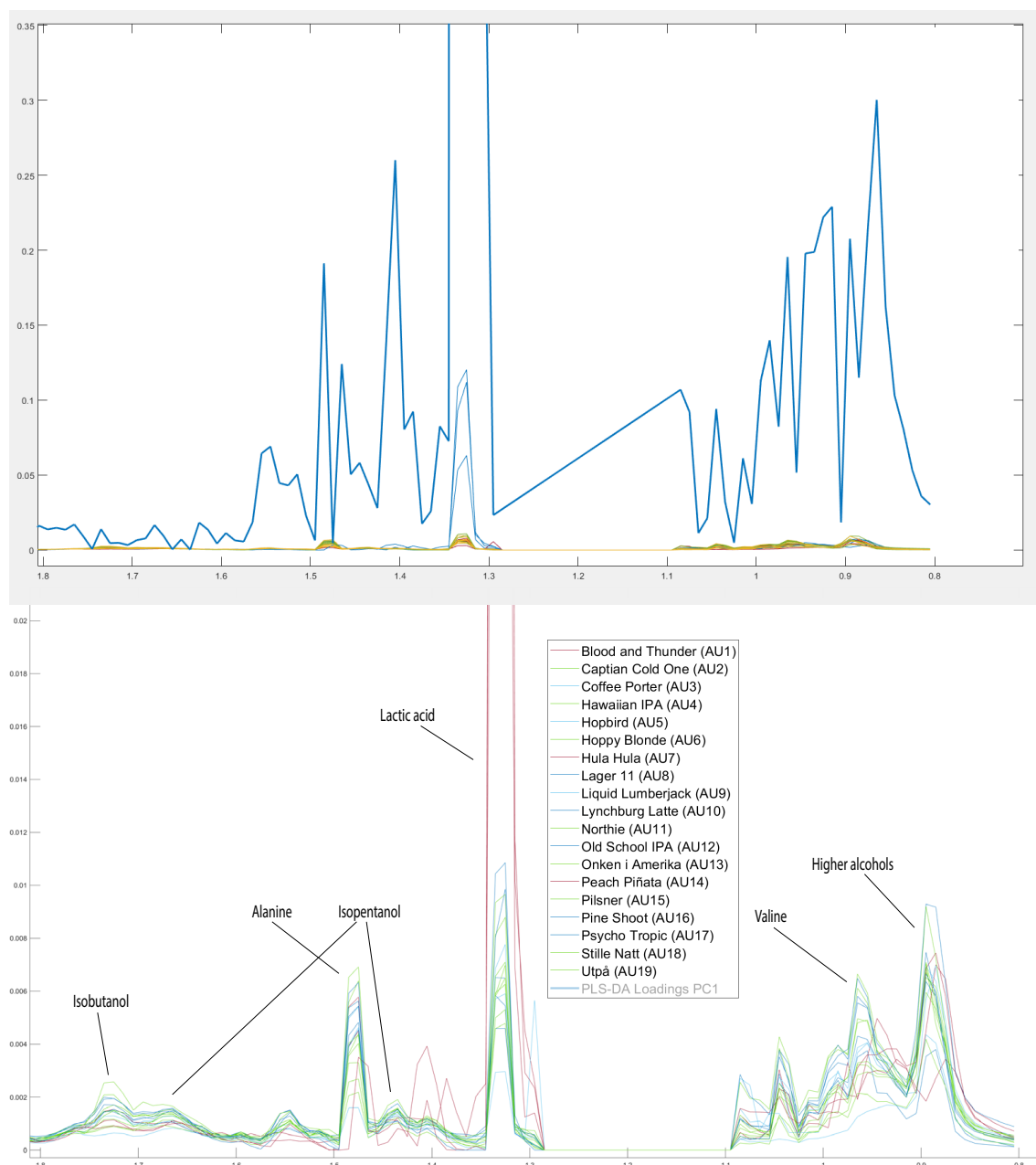


Figure F.4: Detailed spectra and loadings in 0.8-1.8 ppm region for Austmann samples.

Lower field shows a plot of 0.8-1.8 ppm region of all binned NMR spectra for the reclustered Austmann samples overlaid and colored by cluster (cluster 1: red, cluster 2: light blue, cluster 3: green, cluster 4: deep blue). Some identifiable compounds are annotated. Upper field shows PLS-DA loadings for the first PC, taller peaks indicate more weight in PLS-DA. X-axis: PPM, Y-axis: intensity of signal / weight in PLS-DA.

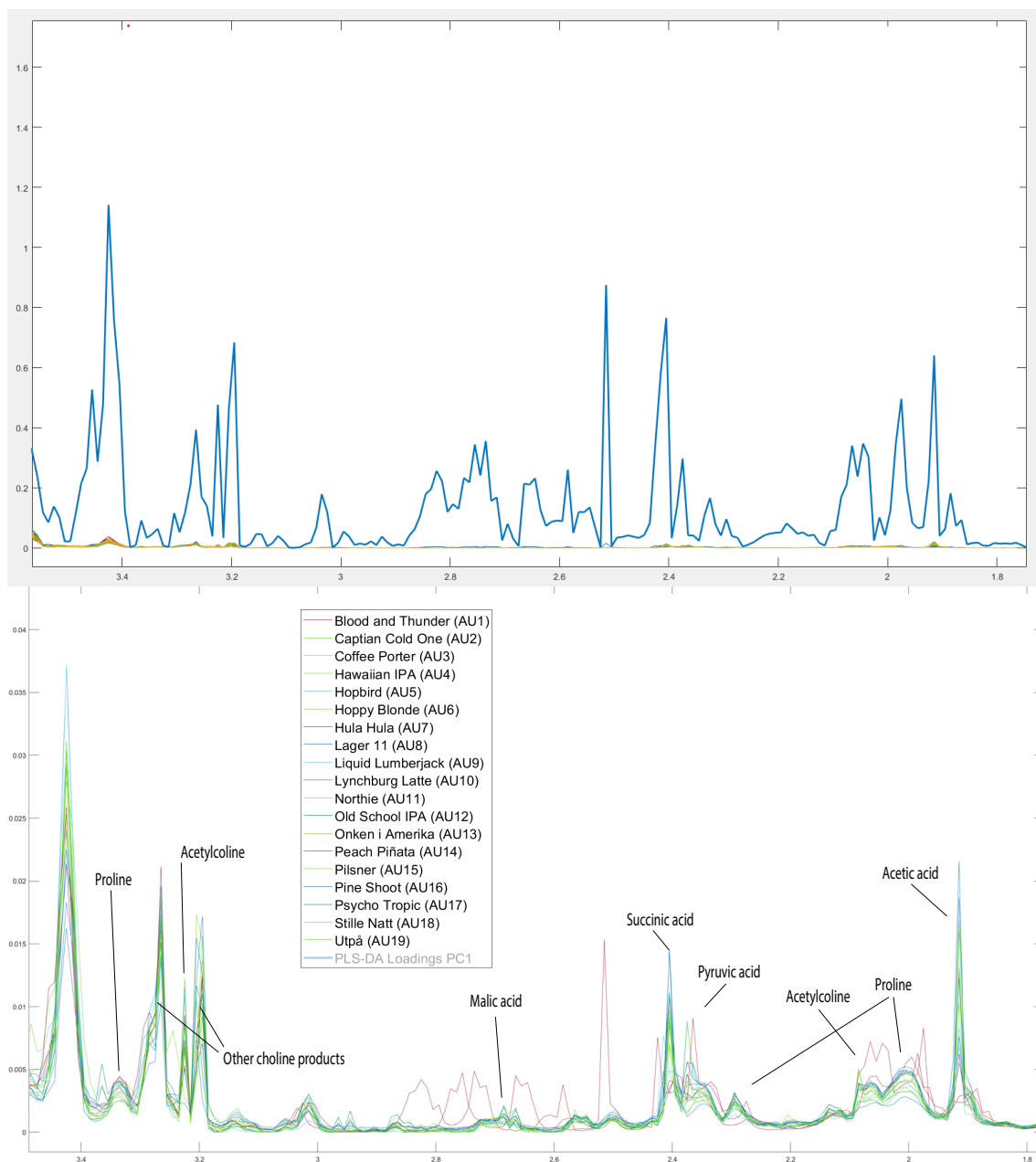


Figure F.5: Detailed spectra and loadings in 1.8-3.5 ppm region for Austmann samples. Lower field shows a plot of 1.8-3.5 ppm region of all binned NMR spectra for the reclustered Austmann samples overlaid and colored by cluster (cluster 1: red, cluster 2: light blue, cluster 3: green, cluster 4: deep blue). Some identifiable compounds are annotated. Upper field shows PLS-DA loadings for the first PC, taller peaks indicate more weight in PLS-DA. X-axis: PPM, Y-axis: intensity of signal / weight in PLS-DA.

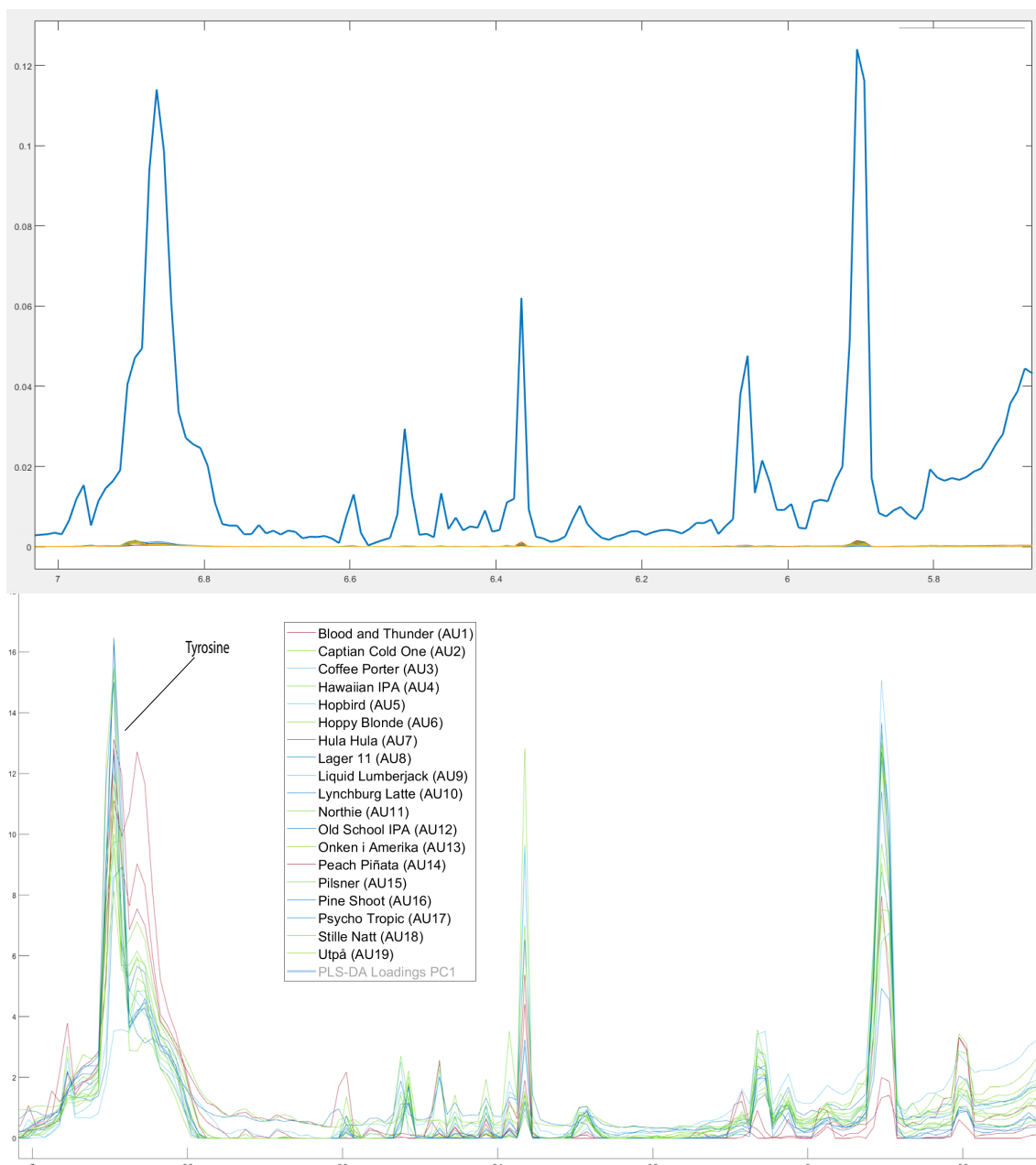


Figure F.6: Detailed spectra and loadings in 5.8-7.1 ppm region for Austmann samples. Lower field shows a plot of 5.8-7.1 ppm region of all binned NMR spectra for the reclustered Austmann samples overlaid and colored by cluster (cluster 1: red, cluster 2: light blue, cluster 3: green, cluster 4: deep blue). Some identifiable compounds are annotated. Upper field shows PLS-DA loadings for the first PC, taller peaks indicate more weight in PLS-DA. X-axis: PPM, Y-axis: intensity of signal / weight in PLS-DA.

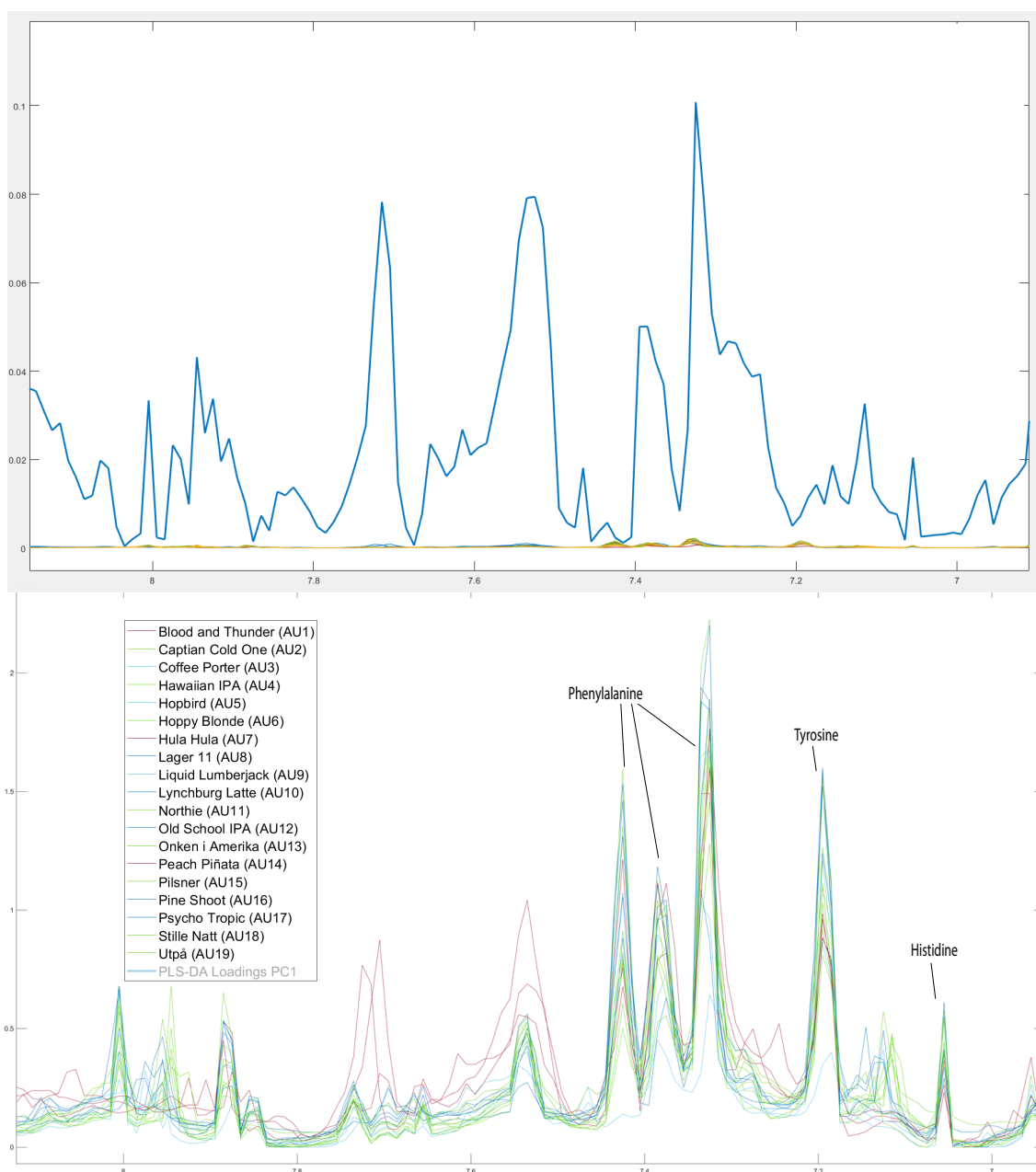


Figure F.7: Detailed spectra and loadings in 6.9-8.2 ppm region for Austmann samples. Lower field shows a plot of 6.9-8.2 ppm region of all binned NMR spectra for the reclustered Austmann samples overlaid and colored by cluster (cluster 1: red, cluster 2: light blue, cluster 3: green, cluster 4: deep blue). Some identifiable compounds are annotated. Upper field shows PLS-DA loadings for the first PC, taller peaks indicate more weight in PLS-DA. X-axis: PPM, Y-axis: intensity of signal / weight in PLS-DA.

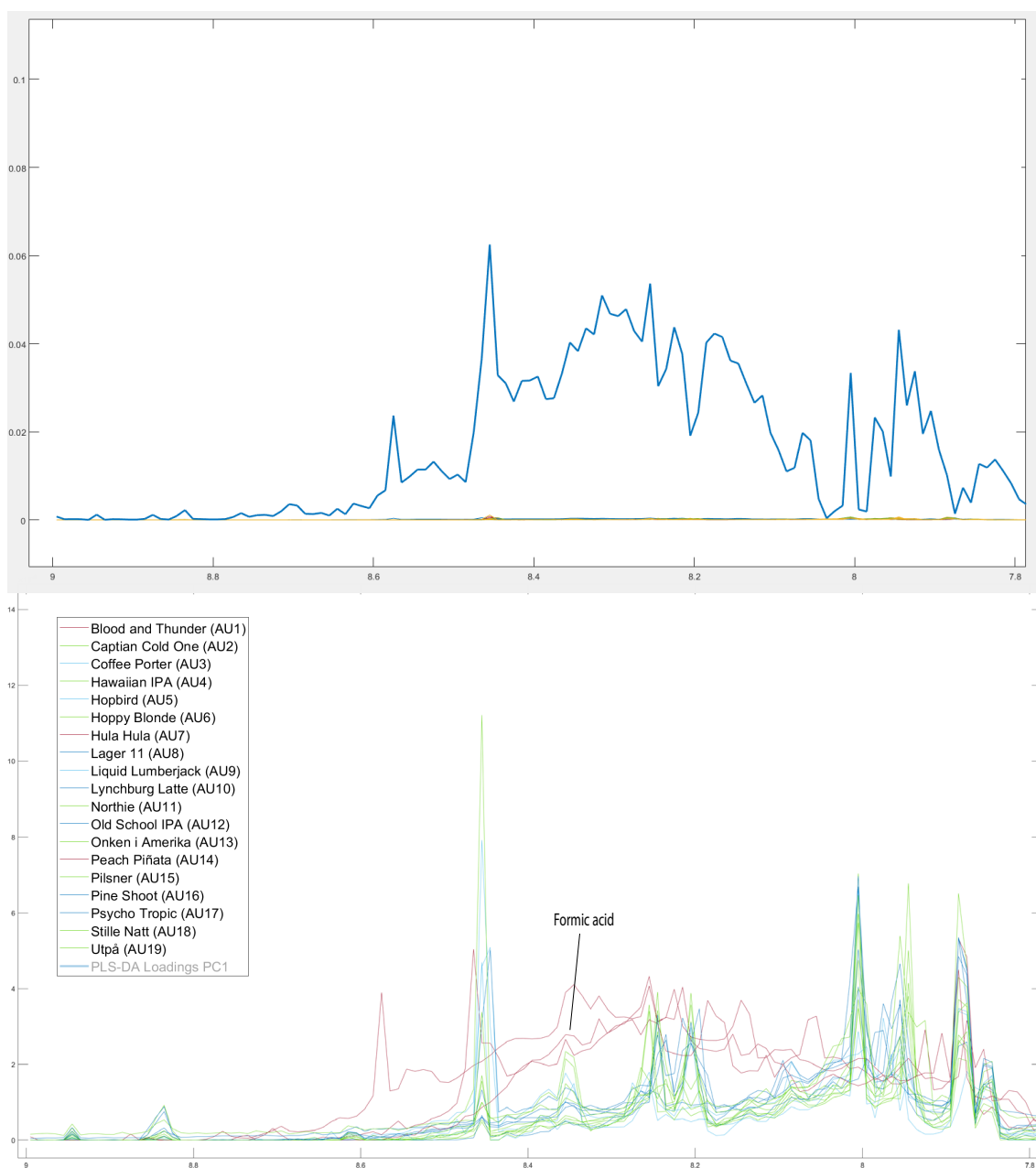


Figure F.8: Detailed spectra and loadings in 7.8-9 ppm region for Austmann samples.

Lower field shows a plot of 7.8-9 ppm region of all binned NMR spectra for the reclustered Austmann samples overlaid and colored by cluster (cluster 1: red, cluster 2: light blue, cluster 3: green, cluster 4: deep blue). Some identifiable compounds are annotated. Upper field shows PLS-DA loadings for the first PC, taller peaks indicate more weight in PLS-DA. X-axis: PPM, Y-axis: intensity of signal / weight in PLS-DA.

## B Pitch rate samples loading plots

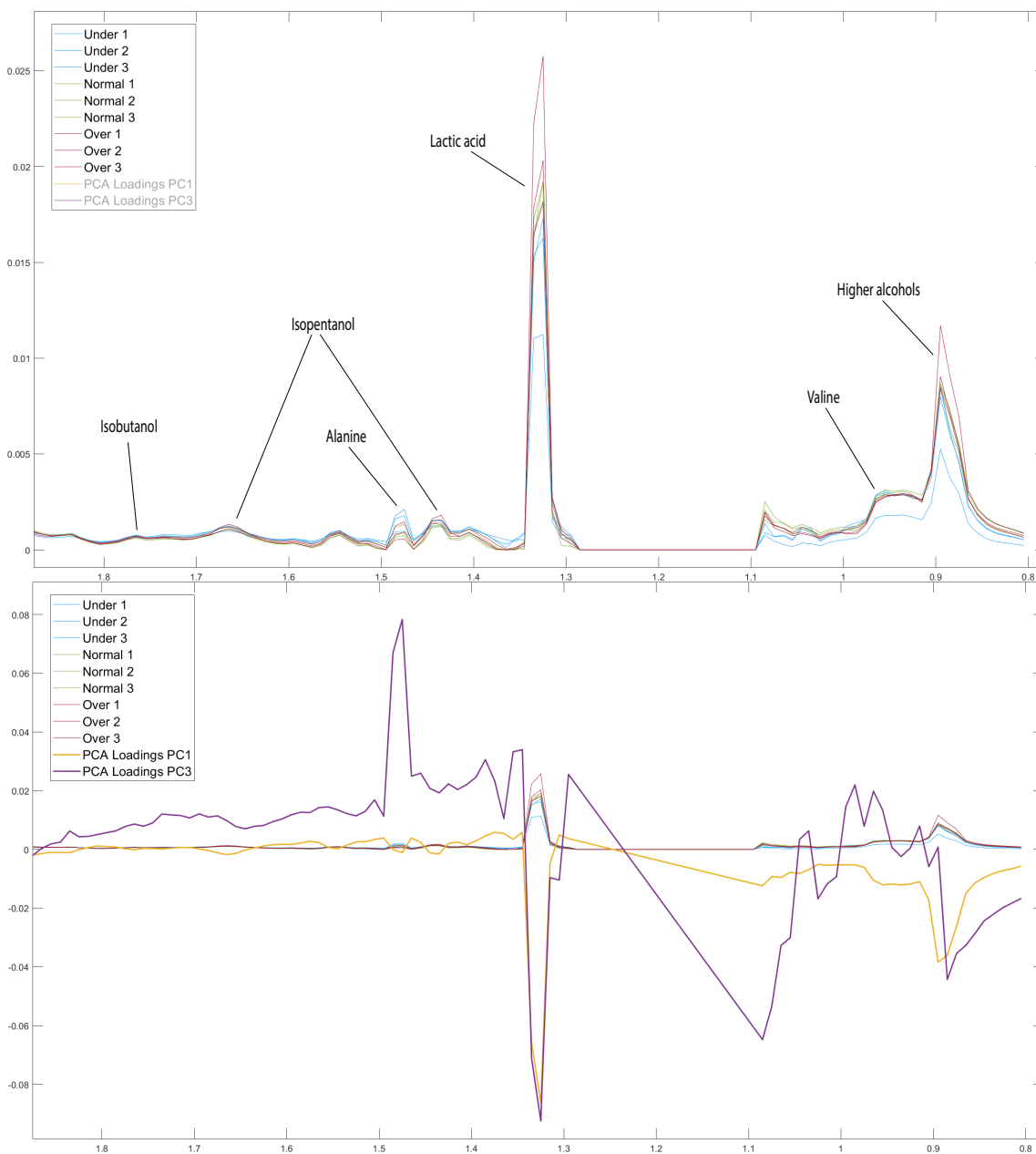


Figure F.9: Detailed spectra and loadings in 0.8-1.8 ppm region for pitch rate samples. Upper field shows a plot of 0.8-1.8 ppm region of all binned NMR spectra for all pitch experiment samples overlaid and colored by pitch rate (underpitched: blue, normal pitch: green, overpitched: red). Some identifiable compounds are annotated. Lower field shows PCA loadings for the first PC, more extreme peaks indicate more weight in the separation found in PCA. X-axis: PPM, Y-axis: intensity of signal / weight in PCA.



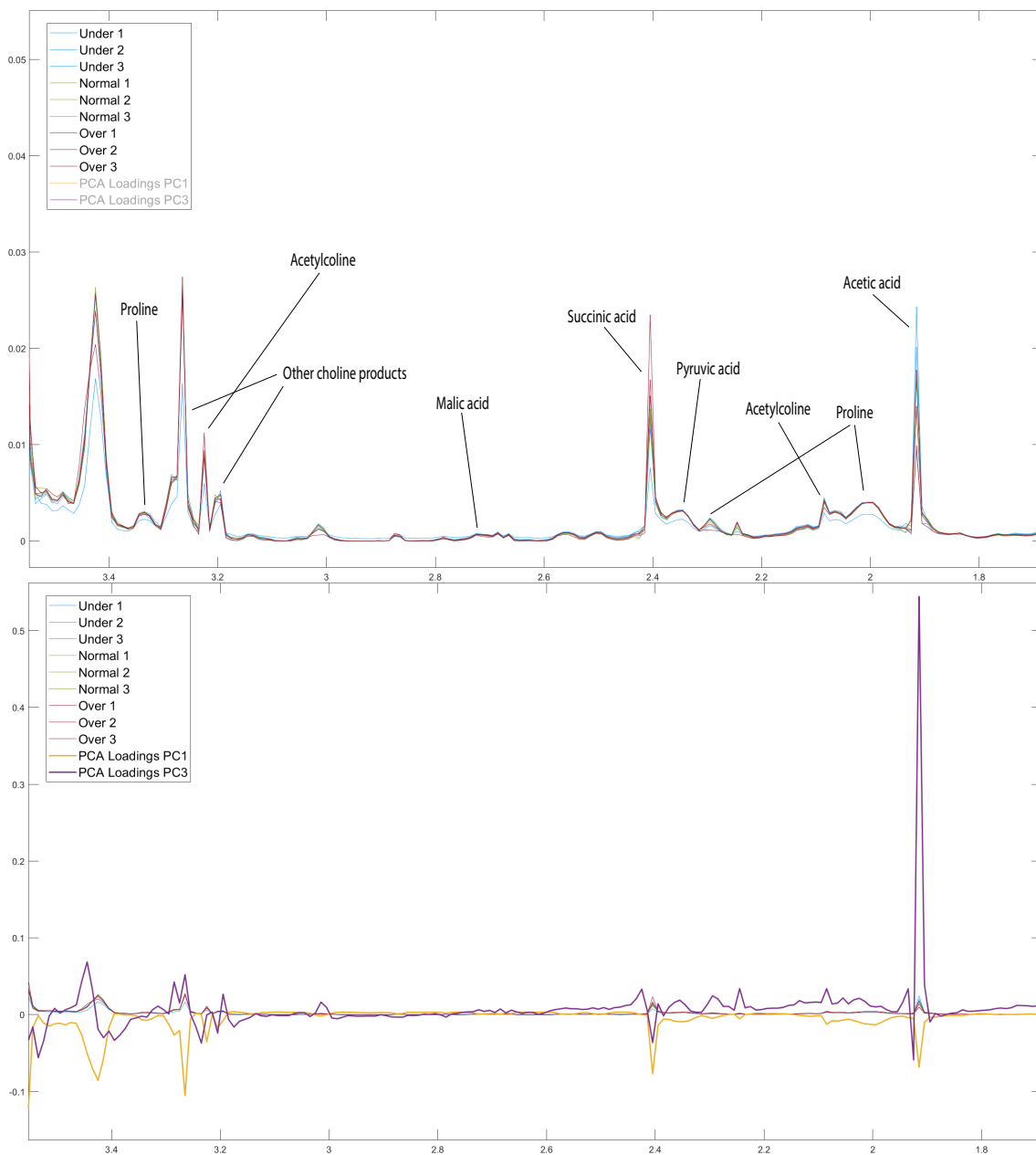


Figure F.10: Detailed spectra and loadings in 1.8-3.5 ppm region for pitch rate samples. Upper field shows a plot of 1.8-3.5 ppm region of all NMR spectra for all pitch experiment samples overlaid and colored by pitch rate (underpitched: blue, normal pitch: green, overpitched: red). Some identifiable compounds are annotated. Lower field shows PCA loadings for the first PC, more extreme peaks indicate more weight in the separation found in PCA. X-axis: PPM, Y-axis: intensity of signal / weight in PCA.

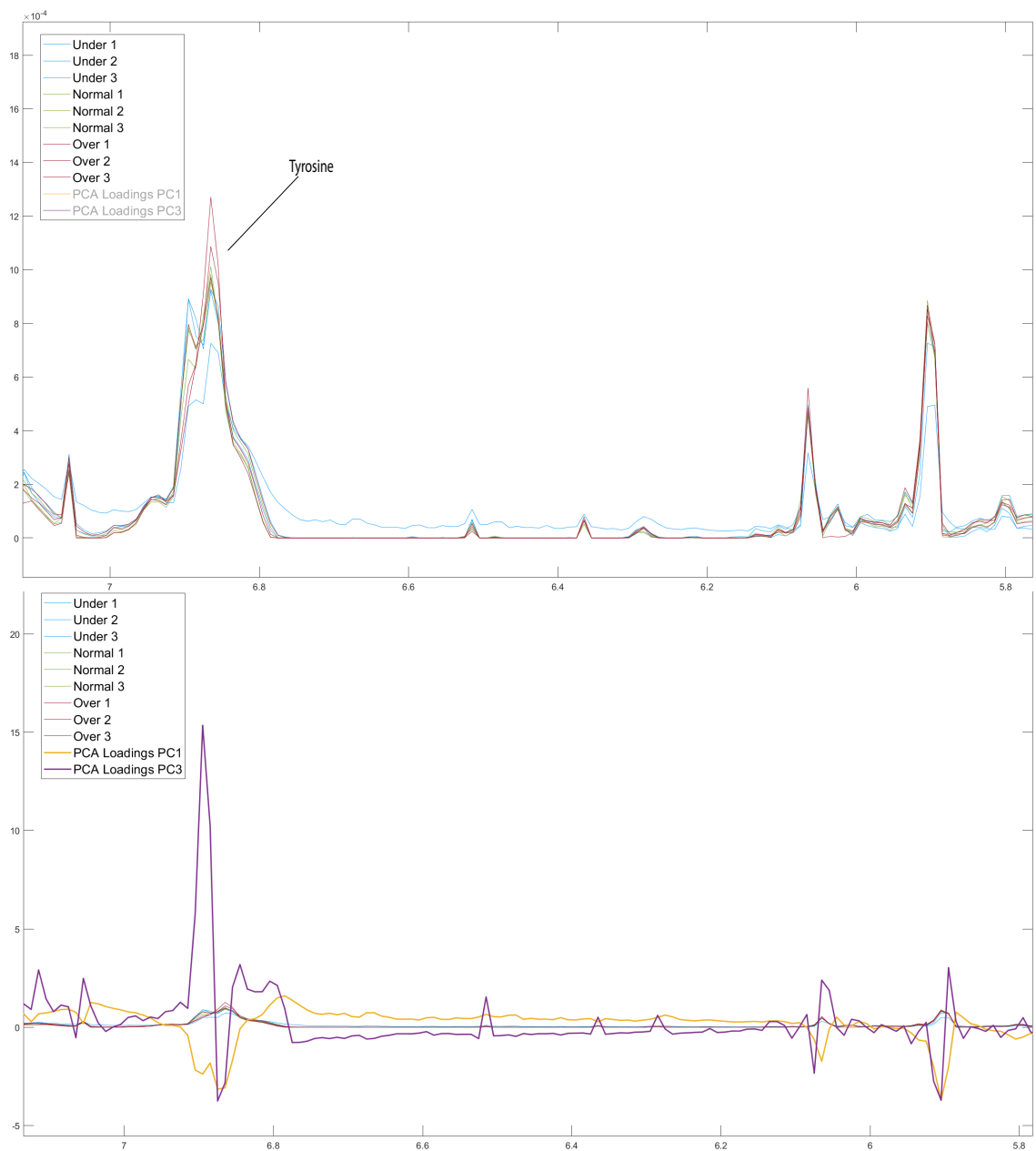


Figure F.11: Detailed spectra and loadings in 5.8-7 ppm region for pitch rate samples. Upper field shows a plot of 5.8-7 ppm region of all binned NMR spectra for all pitch experiment samples overlaid and colored by pitch rate (underpitched: blue, normal pitch: green, overpitched: red). Some identifiable compounds are annotated. Lower field shows PCA loadings for the first PC, more extreme peaks indicate more weight in the separation found in PCA. X-axis: PPM, Y-axis: intensity of signal / weight in PCA.

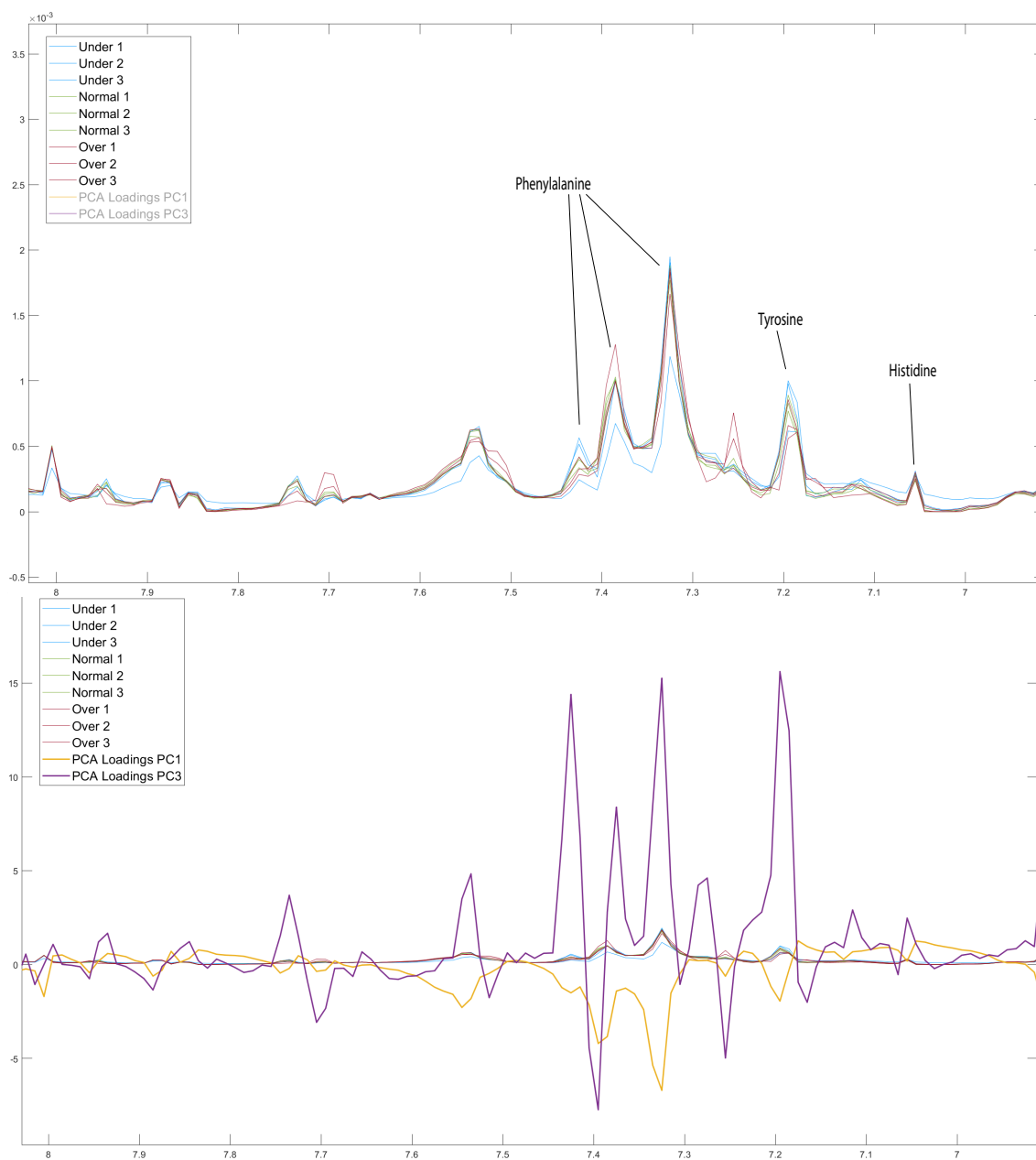


Figure F.12: Detailed spectra and loadings in 7-8 ppm region for pitch rate samples. Upper field shows a plot of 7-8 ppm region of all binned NMR spectra for all pitch experiment samples overlaid and colored by pitch rate (underpitched: blue, normal pitch: green, overpitched: red). Some identifiable compounds are annotated. Lower field shows PCA loadings for the first PC, more extreme peaks indicate more weight in the separation found in PCA. X-axis: PPM, Y-axis: intensity of signal / weight in PCA.

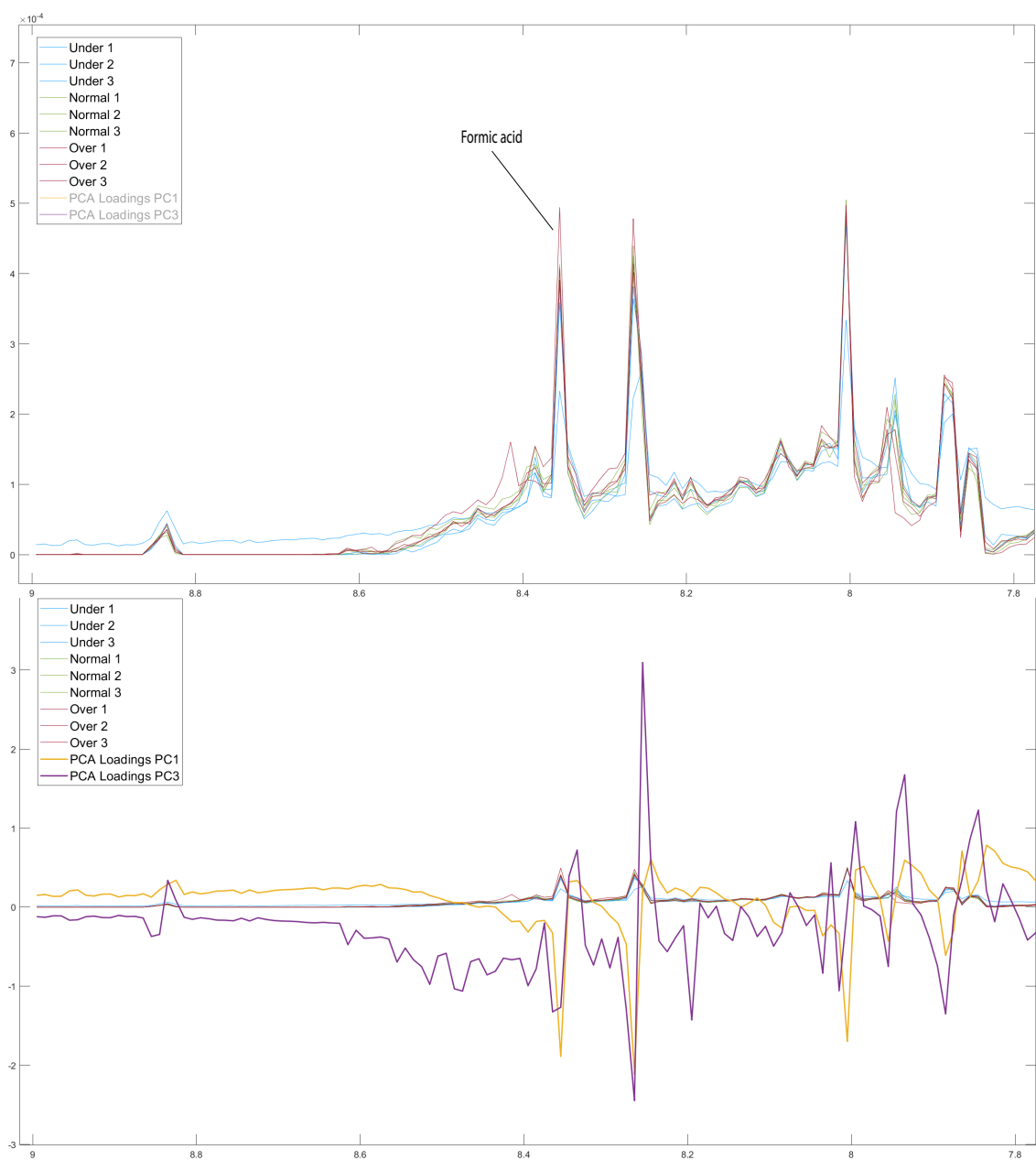
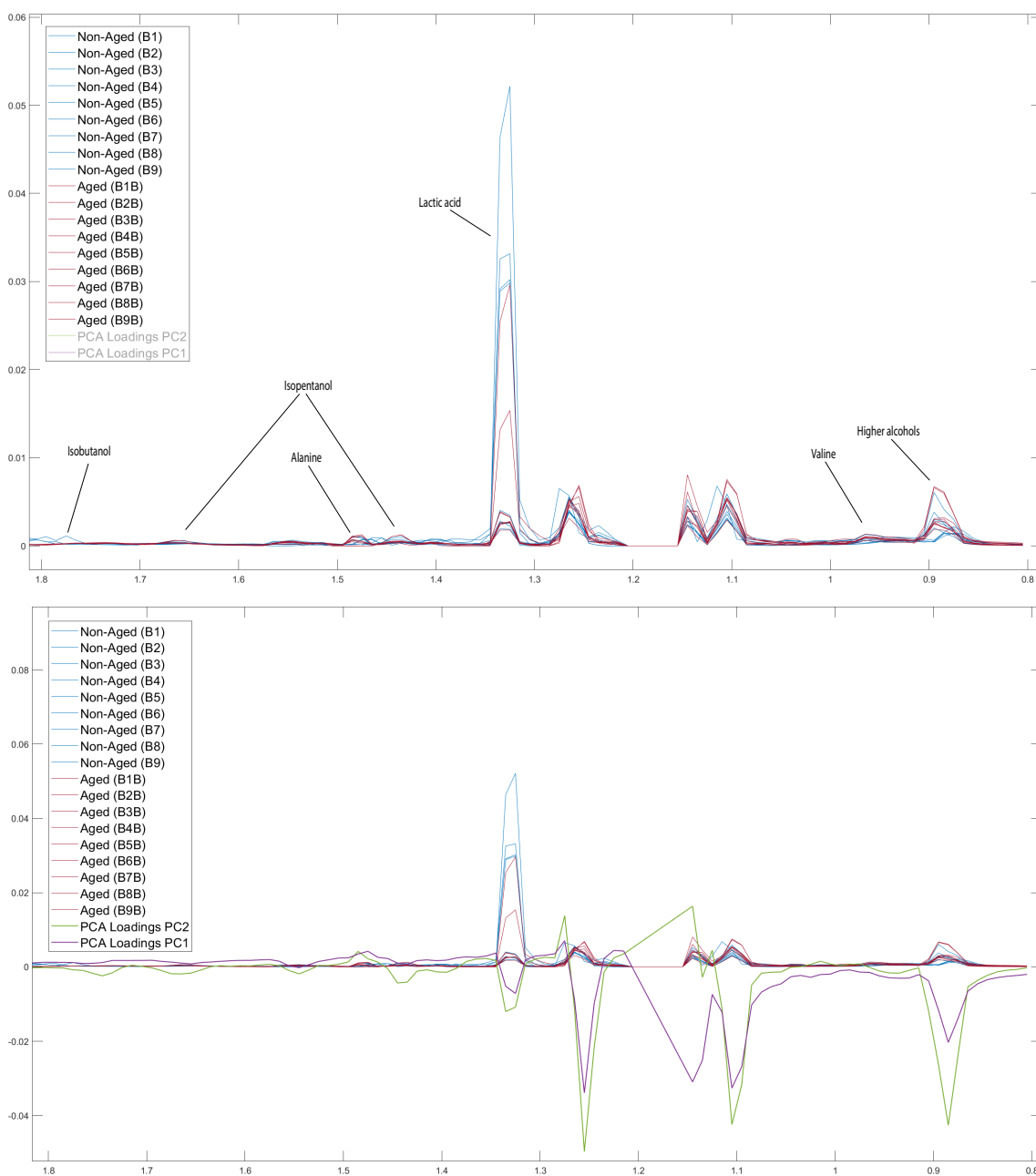


Figure F.13: Detailed spectra and loadings in 7.8-9 ppm region for pitch rate samples. Upper field shows a plot of 7.8-9 ppm region of all binned NMR spectra for all pitch experiment samples overlaid and colored by pitch rate. Some identifiable compounds are annotated. Lower field shows PCA loadings for the first PC, more extreme peaks indicate more weight in the separation found in PCA. X-axis: PPM, Y-axis: intensity of signal / weight in PCA.

## C Algae-beer samples loading plots



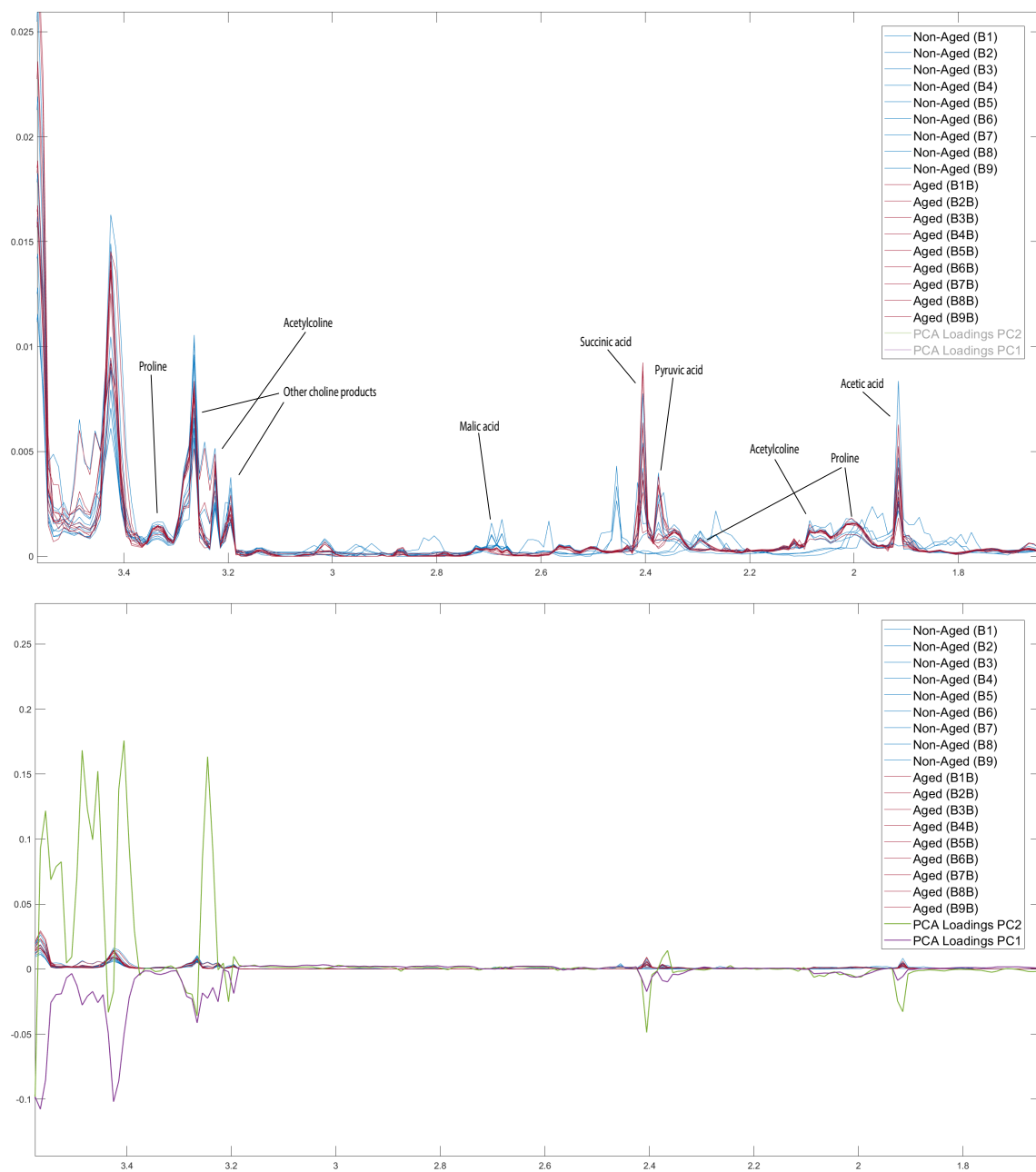


Figure F.15: Detailed spectra and loadings in 1.8-3.4 ppm region for algae-beer samples. Upper field shows a plot of 1.8-3.4 ppm region of all binned NMR spectra for all algae brown ale samples overlaid and colored by aging category (non-aged: blue, aged: red). Some identifiable compounds are annotated. Lower field shows PCA loadings for the first PC, more extreme peaks indicate more weight in the separation found in PCA. X-axis: PPM, Y-axis: intensity of signal / weight in PCA.

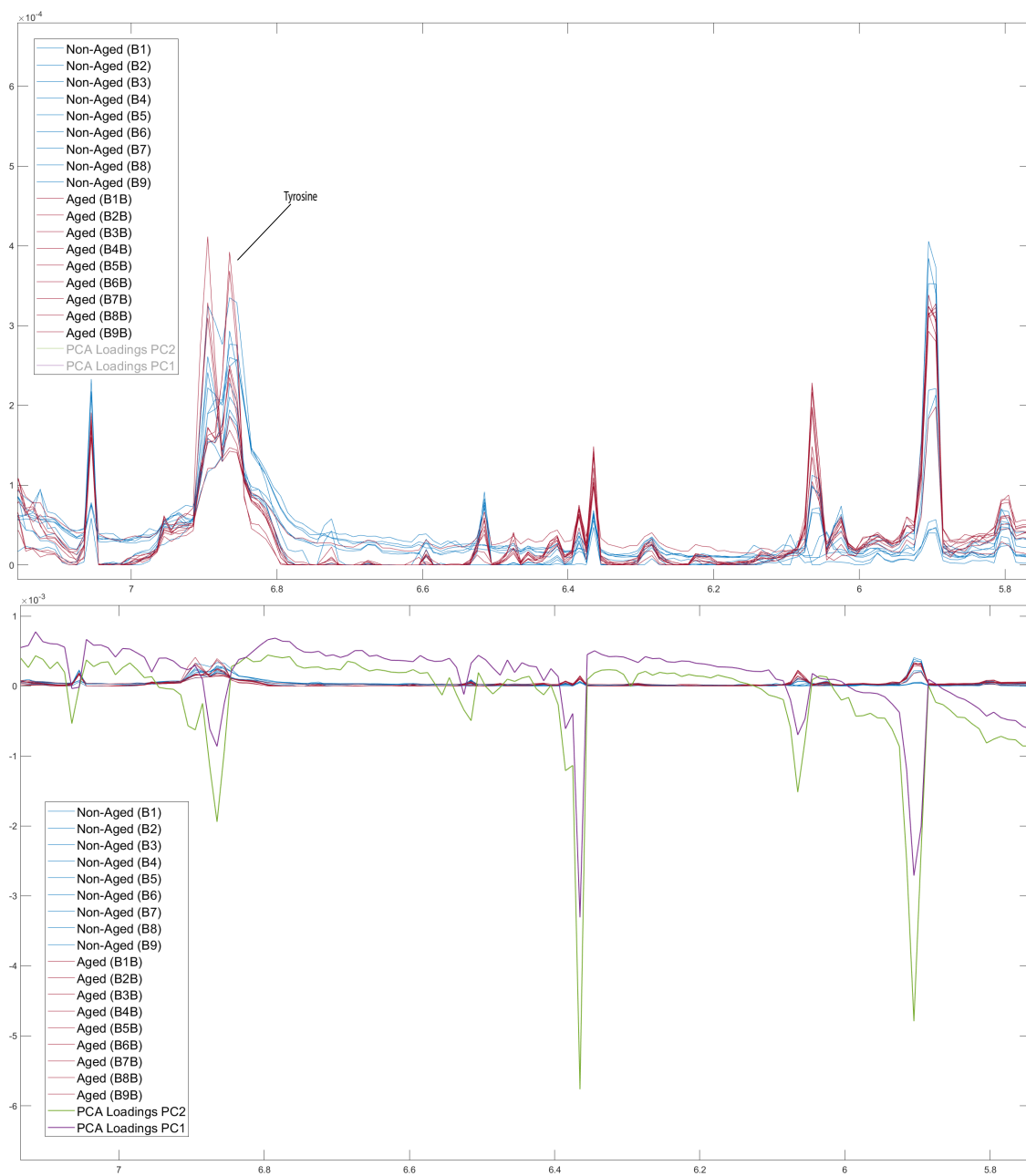


Figure F.16: Detailed spectra and loadings in 5.8-7 ppm region for algae-beer samples. Upper field shows a plot of 5.8-7 ppm region of all binned NMR spectra for all algae brown ale samples overlaid and colored by aging category (non-aged: blue, aged: red). Some identifiable compounds are annotated. Lower field shows PCA loadings for the first PC, more extreme peaks indicate more weight in the separation found in PCA. X-axis: PPM, Y-axis: intensity of signal / weight in PCA.

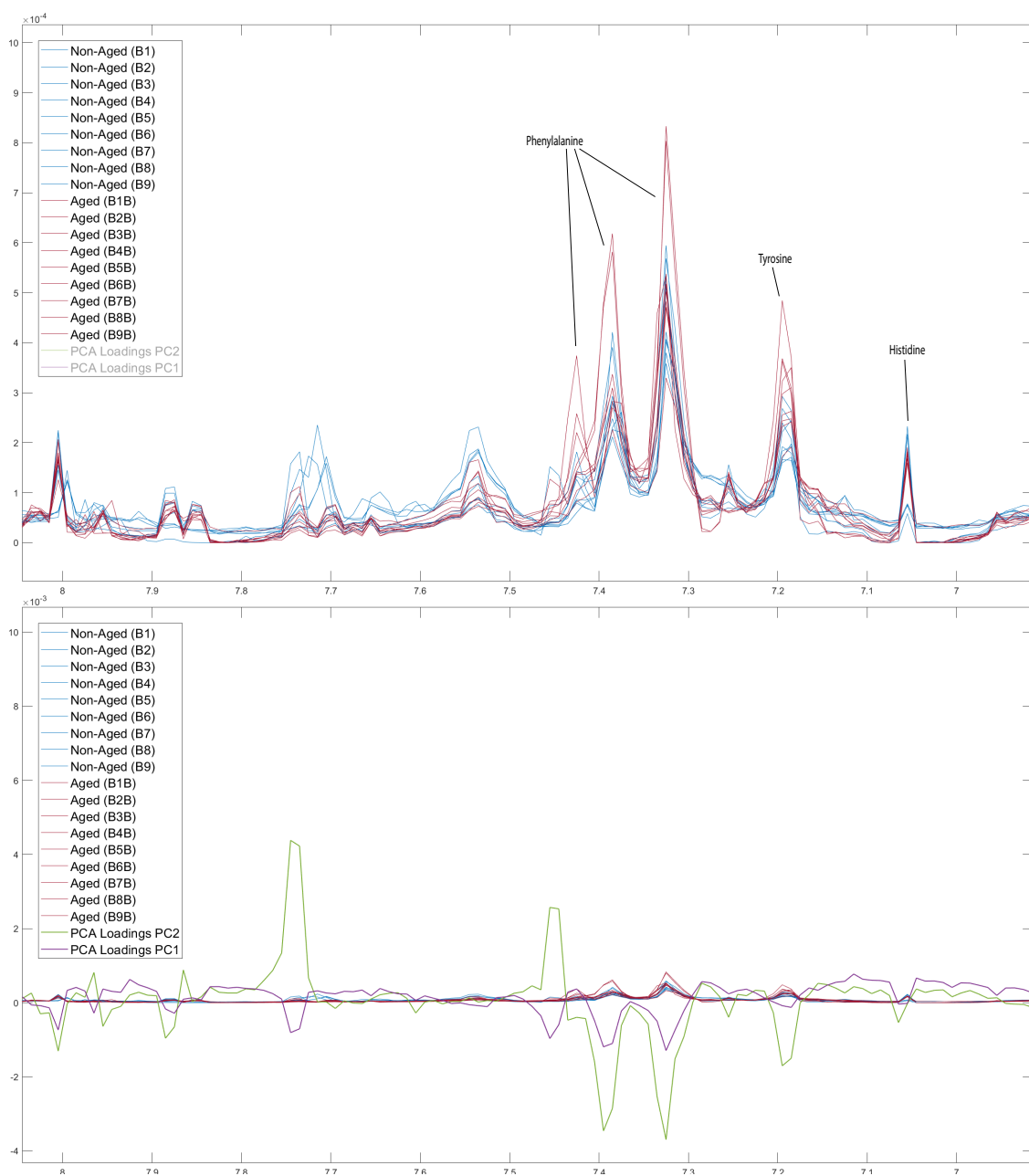


Figure F.17: Detailed spectra and loadings in 7-8 ppm region for algae-beer samples. Upper field shows a plot of 7-8 ppm region of all binned NMR spectra for all algae brown ale samples overlaid and colored by aging category (non-aged: blue, aged: red). Some identifiable compounds are annotated. Lower field shows PCA loadings for the first PC, more extreme peaks indicate more weight in the separation found in PCA. X-axis: PPM, Y-axis: intensity of signal / weight in PCA.



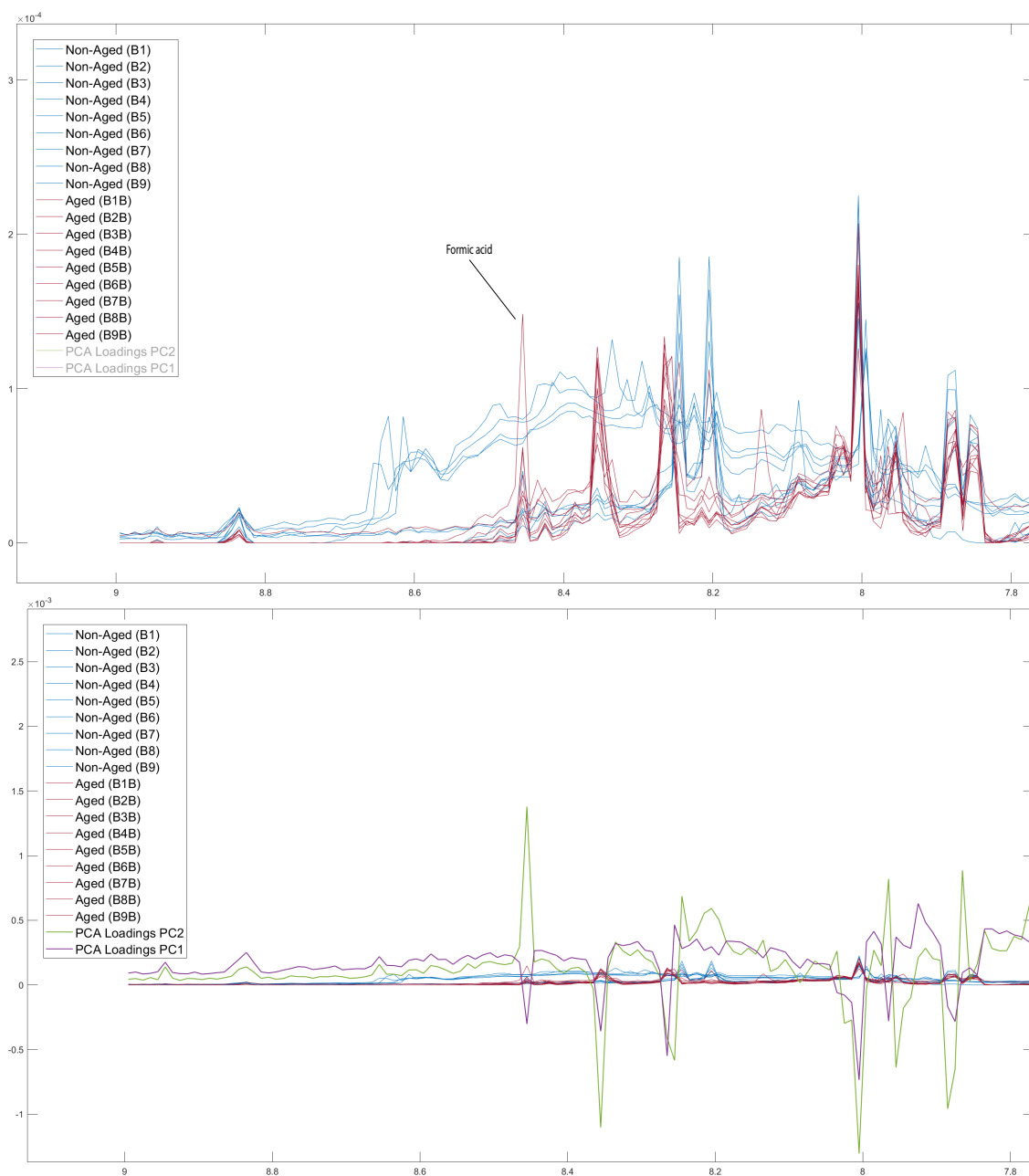


Figure F.18: Detailed spectra and loadings in 7.8-9 ppm region for algae-beer samples. Upper field shows a plot of 7.8-9 ppm region of all binned NMR spectra for all algae brown ale samples overlaid and colored by aging category (non-aged: blue, aged: red). Some identifiable compounds are annotated. Lower field shows PCA loadings for the first PC, more extreme peaks indicate more weight in the separation found in PCA. X-axis: PPM, Y-axis: intensity of signal / weight in PCA.

