

**Futtermittelauthentifizierung: Bewertung
metabolombasierter Ansätze zur geografischen Klassifizierung
am Beispiel von Körnermaisproben**

DISSERTATION

zur Erlangung des akademischen Grades

Doctor rerum naturalium

(Dr. rer. nat.)

der Fakultät für Mathematik, Informatik und Naturwissenschaften

Fachbereich Chemie

der Universität Hamburg

HAMBURG SCHOOL OF FOOD SCIENCE

Institut für Lebensmittelchemie

vorgelegt von

David Schütz

Hamburg 2025

1. Gutachter der Dissertation: Prof. Dr. Markus Fischer
2. Gutachter der Dissertation: Prof. Dr. Stephan Seifert

Mitglieder der Prüfungskommission:

Prof. Dr. Markus Fischer

Prof. Dr. Ralph Holl

Prof. Dr. José A. C. Broekaert

Datum der Disputation: 23.05.2025

Die vorliegende Arbeit wurde im Zeitraum von Dezember 2015 bis Juli 2019 an der Hamburg School of Food Science des Fachbereichs Chemie der Fakultät für Mathematik, Informatik und Naturwissenschaften der Universität Hamburg unter der Anleitung von Herrn Prof. Dr. Markus Fischer angefertigt.

Danksagung

Ich danke zuallererst meinem Doktorvater Prof. Dr. Markus Fischer, für die Bereitstellung des interessanten Themas, sowie seine stetige wissenschaftliche und organisatorische Unterstützung und das mir entgegengebrachte Vertrauen. Insbesondere möchte ich mich für die hervorragende, moderne Geräteausstattung, sowie die Unterstützung bei der Erstellung der Veröffentlichungen bedanken.

Ich danke Prof. Dr. Stephan Seifert für die Übernahme des Zweitgutachtens.

Außerdem danke ich den Projektpartnern beim BfR in Berlin. Namentlich meiner Mitstreiterin Elisabeth Achten, sowie Dr. Janet Riedl, Dr. Bettina Horn und Dr. Carsten Fahl-Hassek. Insbesondere für das Bereitstellen von vermahlenden Herkunftsproben. Frau Riedl sei besonders gedankt für ihre engagierte Unterstützung bei den beiden Veröffentlichungen.

Ich danke allen Probenspendern, die durch die Bereitstellung der Körnermaisproben diese Arbeit erst möglich gemacht haben. Insbesondere Norbert Erhardt von der Landwirtschaftskammer Nordrhein-Westfalen, Dr. Joachim Eder vom LfL Bayern, Institut für Pflanzenbau und Pflanzenzüchtung und Dr. Gert Barthelmes vom LELF Brandenburg für die Proben aus den Landessortenversuchen, sowie Silvie Kramer von Cargill Deutschland für die Stärkemaissproben.

Mein besonderer Dank gilt Dr. Marina Creydt, für ihren stetigen zusätzlichen Blick auf mein Projekt und die fachliche und methodische Unterstützung. Außerdem für die gute Zusammenarbeit und die gemeinsamen Mittagessen. Ein riesengroßes Dankeschön außerdem für die Durchsicht der Veröffentlichungen und dem Korrekturlesen dieser Arbeit.

Allen Mitarbeitern im AK Fischer und der HSFS sei für die Unterstützung und die angenehme Arbeitsatmosphäre gedankt, u. a. Stefanie Schelm, Maike Arndt, Alexandra Scharf, Joanna Fafińska, Dr. Erik Eschbach und Dr. Anna Schulz. Auch Clemens Bernhardt sei gedankt, für den Ausflug in die Welt des Vitamins B₁₂ und die gute Zusammenarbeit dabei. Ein ganz besonderes Dankeschön geht an Katharina Nagel.

Ein großes Dankeschön geht an die MS-Abteilung des FB Chemie, für die Unterstützung bei meinen vielen Messungen am Maxis. Allen voran Dr. Maria Riedner, Raphael Schuster, Dany Gellert, Gaby Graack und Christine Christ.

Ich danke weiterhin Dr. Christian Wittenburg, Conny Bretzke und Sinja Walter für eine schöne Zeit in den Nebenfachpraktika der AC, vor allem dem IFP.

Ich danke meinen Freunden aus der Chemie, Charlotte, Fenja, Julia, Sven, Raphael, Parnian, Maryam und vielen mehr, für eine schöne Unizeit und oft einen anderen Blickwinkel. Und nicht zuletzt meinen Eltern und Großeltern dafür, dass sie mich immer unterstützt haben.

Publikationsliste

Im Rahmen dieser Arbeit wurden zudem fünf Veröffentlichungen erstellt.

1) Non-Targeted LC-MS Metabolomics Approach towards an Authentication of the Geographical Origin of Grain Maize (*Zea mays* L.) Samples

David Schütz, Elisabeth Achten, Marina Creydt, Janet Riedl, Markus Fischer
Foods **2021**, 10, 2160.

2) Fourier-transform near-infrared spectroscopy as a fast screening tool for the verification of the geographical origin of grain maize (*Zea mays* L.)

David Schütz, Janet Riedl, Elisabeth Achten, Markus Fischer
Food Control **2022**, 136, 108892.

3) Food authentication goes green: Determination of the geographical origin of grain maize (*Zea mays* L.) using direct analysis in real time mass spectrometry (DART-MS)

Felix Schmauder, David Schütz, Marina Creydt, Janet Riedl, Markus Fischer
Food Control **2024**, 163, 110489.

4) Classification of Grain Maize (*Zea mays* L.) from Different Geographical Origins with FTIR Spectroscopy—a Suitable Analytical Tool for Feed Authentication?

Elisabeth Achten, David Schütz, Markus Fischer, Carsten Fahl-Hassek, Janet Riedl, Bettina Horn
Food Analytical Methods **2019**, 12, 2172–2184.

5) Cobalamin is produced by *Acetobacter pasteurianus* DSM 3509

Clemens Bernhardt, Xuan Zhu, David Schütz, Markus Fischer, Bernward Bisping
Applied Microbiology and Biotechnology **2019**, 103, 3875–3885.

Die Ergebnisse der ersten beiden Veröffentlichungen sind in **Kapitel 6.4** (1. Veröffentlichung) sowie **Kapitel 6.8** (2. Veröffentlichung) dieser Arbeit dargelegt.

Inhaltsverzeichnis

Danksagung	IV
Publikationsliste	V
Inhaltsverzeichnis	VI
Abkürzungsverzeichnis	XI
1 Zusammenfassung	1
2 Abstract	3
3 Einleitung	5
4 Theorie	8
4.1 Mais	8
4.1.1 Botanische Einordnung und Geschichte	8
4.1.2 Anbauformen und Sorten	8
4.1.3 Körnermais: Nacherntebehandlung und Nutzung als Futtermittel	10
4.1.4 Anbauländer und –mengen	11
4.2 <i>Metabolomics</i>	13
4.2.1 <i>Omics</i> – Technologien	13
4.2.2 Grundlagen <i>Metabolomics</i>	15
4.2.3 <i>Food / Feed Metabolomics</i>	16
4.2.4 Analytische Technologien und Strategien	17
4.3 Apparative Methoden	20
4.3.1 FT-NIR-Spektroskopie	20
4.3.2 Flüssigkeitschromatographie	22
4.3.3 Massenspektrometrie	26
4.4 Datenanalyse	32
4.4.1 Univariate Datenanalyse	32
4.4.2 Verfahren der multivariaten Datenanalyse	32
4.4.3 Datenvorverarbeitung	33
4.4.4 Hauptkomponentenanalyse (PCA)	35
4.4.5 <i>Partial Least Square Discriminant Analysis</i> (PLS-DA)	36
4.4.6 Erstellung und Bewertung von Klassifizierungs- und Vorhersagemodellen	37
4.4.7 <i>Support Vector Machines</i> (SVM)	38
4.4.8 <i>Random Forest</i> (RF)	40

4.4.9	Kreuzvalidierung zur Bewertung der Genauigkeit von Klassifikationsalgorithmen ..	40
4.4.10	Fusion von Analyseergebnissen verschiedener Methoden zur Klassifikation	42
5	Zielsetzung der Arbeit	44
6	Ergebnisse und Diskussion	46
6.1	Optimierung der Probenextraktion.....	46
6.1.1	Optimierung der Extraktionsmethode für polare Metabolite	46
6.1.2	Optimierung der Extraktionsmethode für unpolare Metabolite	48
6.2	Untersuchung einer möglichen Beeinflussung des Körnermaismetaboloms durch die Maissorte.....	51
6.2.1	Untersuchung einer Sortenabhängigkeit der Metabolite	51
6.2.2	Multivariate Analyse der Metabolite	52
6.2.3	Herkunftsbezogene Varianz vs. sortenbezogene Varianz der Metabolite.....	54
6.2.4	Einfluss von Sorte und geographischer Herkunft auf Klassifizierungsmodelle	56
6.2.5	Einfluss der Sorte und geographischer Herkunft je nach Analysemethode	58
6.3	Analyse der Herkunftspuren mittels LC-MS.....	59
6.3.1	Multivariate Analyse der Metabolite in den Herkunftspuren.....	60
6.3.2	Einfluss der Messmethode auf die Klassifikation	62
6.4	Bestimmung der geographischen Herkunft von Körnermaisproben anhand des Profils der unpolaren Metabolite im positiven Ionenmodus ¹	63
6.4.1	Einfluss der Lipidspezies auf die Eignung zur Klassifikation nach der geographischen Herkunft	65
6.4.2	Identifizierung von relevanten Markermetaboliten zur Klassifikation nach der geographischen Herkunft.....	66
6.4.3	Multivariate Datenanalyse und Klassifikation unter Nutzung aller 20 identifizierten Markermetabolite	68
6.4.4	Strukturaufklärung der optimalen Markermetabolite	71
6.5	Lagerstabilität der Lipidextrakte	72
6.6	Etablierung einer MRM-Methode für die Quantifizierung von Lipidspezies in Körnermais.	75
6.6.1	Ergebnisse der Klassifizierung durch eine MRM-Methode (MS ² -Level)	78
6.6.2	Vergleich mit der Tocopherol-basierten Methode	81
6.6.3	Herkunftsbestimmung mittels Lipidprofil vs. Tocopherole-basierter Methode	82
6.7	Untersuchungen mittels FT-NIR-Spektroskopie	84
6.7.1	Auswahl der optimalen Probenform für FT-NIR-Spektren.....	85
6.7.2	Multivariate Analyse der FT-NIR-Spektren des Sortenprobensets	85

6.7.3	Untersuchung einer möglichen Sortenabhängigkeit bei der Klassifikation von FT-NIR spektroskopischen Messungen	87
6.8	Analytik der Herkunftsproben mittels FT-NIR-Spektroskopie ²	88
6.8.1	Interpretation der FT-NIR-Spektren von Körnermais	88
6.8.2	Multivariate Datenanalyse der FT-NIR-Spektren von Körnermais	89
6.8.3	Unterschiede in den FT-NIR-Spektren je nach Herkunft	90
6.8.4	Auswirkungen des <i>Preprocessing</i> von Spektren auf die Herkunftsbestimmung	91
6.8.5	Einfluss der verwendeten Spektrenbereiche auf das SVM-Klassifikationsmodell für die Herkunftsbestimmung	92
6.8.6	Optimiertes SVM-Klassifikationsmodell für die Herkunftsbestimmung	94
6.8.7	Direktes vs. Zwei-Schritt SVM-Klassifikationsmodell für die Herkunftsbestimmung	95
6.8.8	Robustheit des Klassifikationsmodells bei Vorhersage weiterer Proben	97
6.9	Fusion von LC-MS- und FT-NIR-Messdaten	99
6.9.1	Datenvorbereitung für die <i>low-level</i> Fusion	99
6.9.2	Vergleich des Fusionsmodells mit den Einzelmodellen mittels multivariater Datenanalyse	100
6.9.3	Vergleich der Klassifikationsergebnisse des Fusionsmodells mit den Einzelmodellen .	100
7	Übergreifende Diskussion	102
7.1	Vergleich der Eignung der Herkunftsverifikation von Körnermais mittels LC-MS(/MS) und FT-NIR-Spektroskopie	102
7.2	Beurteilung einer möglichen Störung der Herkunftsverifikation durch einen Sorteneinfluss	103
7.3	Vergleich der Ergebnisse nach Datenfusion mit den Ergebnissen aus LC-MS-Daten und FT-NIRS-Daten	104
8	Ausblick	106
9	Methoden	108
9.1	Probenmaterial	108
9.1.1	QC-Kontrollprobe	109
9.2	Probenvorbereitung	110
9.3	Allgemeine Vorgehensweise zur Extraktion von Metaboliten aus Körnermaisproben ..	111
9.3.1	Extraktionsoptimierung für polare Metabolite	112
9.3.2	Optimierte Extraktionsmethode für polare Metabolite	113
9.3.3	Extraktionsoptimierung für unpolare Metabolite	113
9.3.4	Optimierte Extraktionsmethode für unpolare Metabolite	114
9.4	Überprüfung der Lagerungsstabilität von Extrakten unpolarer Metabolite	115

9.5	Tocopherol-spezifische Extraktionsmethode.....	115
9.6	LC-MS(/MS)-Messungen	116
9.6.1	LC-MS Methoden für polare Metabolite.....	116
9.6.2	LC-MS Methoden für unpolare Metabolite.....	117
9.6.3	Kalibration des Massenspektrometers	118
9.6.4	Akquirierung von MS/MS-Fragmentspektren	118
9.7	LC-MRM Methode für Lipide.....	119
9.7.1	Kalibration für quantitative Metabolit-Bestimmungen	120
9.7.2	Validierung der Lipid-MRM-Methode	120
9.8	LC-MRM Methode für Tocopherole	121
9.8.1	Kalibration für quantitative Tocopherol-Bestimmungen	122
9.9	Auswertung der LC-MS(/MS)-Messungen.....	122
9.9.1	<i>Feature-Finding</i>	122
9.9.2	Erstellung von vergleichenden Peaklisten.....	123
9.9.3	Bestimmung der Peakflächen von LC-MS(/MS)-Analysen und quantitative Auswertung	124
9.9.4	Rechnerische Bestimmung der Validierungskennzahlen	126
9.10	FT-NIR-Spektroskopie.....	127
9.10.1	Evaluation der Probenform für FT-NIR-Spektroskopie	127
9.10.2	Aufnahme von FT-NIR-Spektren.....	127
9.11	Probenmessungen.....	128
9.12	Univariate Datenanalyse	130
9.13	Datenvorverarbeitung.....	130
9.13.1	<i>Preprocessing</i> für LC-MS Daten	130
9.13.2	<i>Preprocessing</i> -Algorithmen für FT-NIR-Spektren	131
9.14	Multivariate Datenanalyse und Klassifikation.....	132
9.14.1	PCA sowie PLS-DA.....	132
9.14.2	Strategie zur Auswahl von Features zur Herkunftsbestimmung.....	132
9.14.3	Kreuzvalidierung der Klassifikationsmodelle	133
9.14.4	Klassifikationsalgorithmen	134
9.14.5	Vorhersage von weiteren Proben mittels existierender Klassifikationsmodelle	135
9.15	Datenfusion	136
9.15.1	<i>Preprocessing</i> der LC-MS bzw. FT-NIR-Messdaten vor Datenfusion.....	136
9.15.2	Durchführung der Datenfusion	136

10	Literaturverzeichnis	137
11	Abbildungsverzeichnis	161
12	Tabellenverzeichnis	163
13	Anhang.....	165
13.1	Material	165
13.2	Sicherheit und Entsorgung	172
13.3	Ergänzungen zum Ergebnisteil	174
	Eidesstattliche Versicherung	191

Abkürzungsverzeichnis

%	Prozent
°C	Grad Celsius
γ	Kernel Parameter
μ	Mikro
ANOVA	<i>analysis of variance</i>
ANP	wässrige Normalphase (engl. <i>aqueous normal phase</i>)
APCI	chemische Ionisation bei Atmosphärendruck (engl. <i>atmospheric pressure chemical ionization</i>)
API	Atmosphärendruckionisation (engl. <i>atmospheric pressure ionization</i>)
APPI	Atmosphärendruck-Photoionisation (engl. <i>atmospheric pressure photoionization</i>)
BB	Brandenburg
BfR	Bundesinstitut für Risikobewertung
BHT	Butylhydroxytoluol
BPC	<i>base peak chromatogram</i>
BY	Bayern
cm	Zentimeter
C	Cost Parameter (<i>soft margin</i> Parameter)
CC	korrekte Klassifizierungen (engl. <i>correct classifications</i>)
CCC	Gegenstromverteilungschromatographie (engl. <i>counter current chromatography</i>)
CE	Kapillarelektrophorese (engl. <i>capillary electrophoresis</i>)
CE-MS	Kapillarelektrophoresen-Massenspektrometrie (engl. <i>capillary electrophoresis mass spectrometry</i>)
Da	Dalton
DAD	Diodenarraydetektor
DART	<i>direct analysis in real time</i>
DDGS	Trockenschlempe (engl. <i>distiller's dried grains with solubles</i>)
DESI	Desorptions-Elektrospray-Ionisierung (engl. <i>desorption electrospray ionization</i>)
DG	Diglycerid
DNA	Desoxyribonukleinsäure (engl. <i>desoxyribonucleic acid</i>)
eV	Elektronenvolt
EI	Elektronenstoß-Ionisation (engl. <i>electron impact ionization</i>)
EIC	<i>extracted ion chromatogram</i>
ESI	Elektrospray-Ionisation (engl. <i>electrospray ionization</i>)
EU	Europäische Union
FAO	<i>Food and Agricultural Organisation of the United Nations</i>
FRA	Länderkürzel für "Frankreich"
FT	Fourier Transformation
FT-ICR-MS	Fouriertransformation-Ionenzyklotronresonanz-Massenspektrometer (engl. <i>fourier transform ion cyclotron resonance mass spectrometer</i>)
FT-IR	Fouriertransformierte-Infrarot(-Spektroskopie)
FT-NIR	Fouriertransformierte-Nahinfrarot(-Spektroskopie)
g	Gramm
g	Erdbeschleunigung
GC	Gaschromatographie
GC-EI-MS	Gaschromatographie gekoppelt mit Elektronenstoßionsations-Massenspektrometrie (engl. <i>gas chromatography electron impact ionization mass spectrometry</i>)

GC-MS	Gaschromatographie gekoppelt mit Massenspektrometrie (engl. <i>gas chromatography mass spectrometry</i>)
GHS	Global harmonisiertes System zur Einstufung und Kennzeichnung von Chemikalien (engl. <i>globally harmonized system</i>)
GM	Genetisch modifiziert (engl. <i>genetically modified</i>)
ha	Hektar
HILIC	Hydrophile Interaktionschromatographie (engl. <i>hydrophilic interaction liquid chromatography</i>)
HMDB	<i>Human Metabolome Database</i>
HPLC	Hochleistungsflüssigkeitschromatographie (engl. <i>high performance liquid chromatography</i>)
HSI	<i>Hyperspektrale Bildgebung</i> (engl. <i>hyperspectral imaging</i>)
Hz	Hertz
IR	Infrarot
IMS	Ionen-Mobilitäts-Spektrometrie (engl. <i>ion mobility spectrometry</i>)
IM-MS	Ionen-Mobilitätsspektroskopie gekoppelt mit Massenspektrometrie (engl. <i>ion mobility mass spectrometry</i>)
k	Kilo
KMR	krebserzeugend, mutagen oder reproduktionstoxisch
L	Liter
LC	Flüssigkeitschromatographie (engl. <i>liquid chromatography</i>)
LC-ESI-MS	Flüssigkeitschromatographie gekoppelt mit ESI-Massenspektrometrie (engl. <i>liquid chromatography electrospray ionization mass spectrometry</i>)
LC-MS	Flüssigkeitschromatographie gekoppelt mit Massenspektrometrie (engl. <i>liquid chromatography mass spectrometry</i>)
LDA	Lineare Diskriminanzanalyse (engl. <i>linear discriminant analysis</i>)
<i>m</i>	Masse
m	Milli
m	Meter
M	Molar
<i>m/z</i>	Verhältnis von Masse zu Ladung
MALDI	<i>Matrix-unterstützte Laser-Desorption/Ionisation</i> (engl. <i>matrix assisted laser desorption ionization</i>)
min	Minute
Mio.	Millionen
mol	Mol
Mrd.	Milliarden
mRNA	Boten-Ribonukleinsäure (engl. <i>messenger ribonucleic acid</i>)
MTBE	Methyl- <i>tert.</i> -butylether
MRM	<i>multiple reaction monitoring</i>
MS	Massenspektrometer
MS ¹	Ebene der Precursorionen (unfragmentierte Analyten)
MS ²	Ebene der Fragmentionen
MSC	<i>multiplicative scatter correction</i>
n	Nano
n	Anzahl Messungen
NIR	Nahinfrarot(-spektroskopie)
NMR	Kernresonanzspektroskopie (engl. <i>nuclear magnetic resonance</i>)
NP	Normalphase (engl. <i>normal-phase</i>)
NRW	Nordrhein-Westfalen
NWG	Nachweisgrenze

ppm	<i>parts per million</i>
PC	Phosphatidylcholin
PC	Hauptkomponente
PCA	Hauptkomponentenanalyse (engl. <i>principal component analysis</i>)
PE	Phosphatidylethanolamin
PER	Länderkürzel für "Peru"
pH	negativer dekadischer Logarithmus der H ⁺ -Konzentration
PLS-DA	<i>partial least squares discriminant analysis</i>
PLS-R	<i>partial least squares regression</i>
q	Quadrupol
QC	Qualitätskontrolle (engl. <i>quality control</i>)
q-ToF-MS	Quadrupol-Flugzeit(-Massenspektrometer) (engl. <i>quadrupol time-of-flight mass spectrometry</i>)
R ²	Bestimmtheitsmaß
RASFF	<i>Rapid Alert System Food and Feed</i>
RBF	Radial-Basis-Kernelfunktion
RF	<i>random forest</i>
RFID	Radiofrequenz-Identifikation (engl. <i>radio-frequency identification</i>)
RP	Umkehrphase (engl. <i>reversed phase</i>)
RSD	relative Standardabweichung
rpm	Umdrehungen pro Minute
s	Sekunde
SLO	Länderkürzel für "Slowakei"
SNV	<i>standard normal variate</i>
SPA	Länderkürzel für "Spanien"
SFC	überkritische Flüssigkeitschromatographie (engl. <i>supercritical fluid chromatography</i>)
SFC-MS	überkritische Flüssigkeitschromatographie gekoppelt mit Massenspektrometrie (engl. <i>supercritical fluid chromatography mass spectrometry</i>)
SVM	<i>support vector machines</i>
TG	Triglycerid
ToF	Flugzeit (engl. <i>time of flight</i>)
Tsd.	Tausend
UKR	Länderkürzel für "Ukraine"
UNG	Länderkürzel für "Ungarn"
UHPLC	Ultrahochleistungsflüssigkeitschromatographie (engl. <i>ultra high performance liquid chromatography</i>)
USA	Vereinigte Staaten von Amerika
UV	Ultraviolett
v/v	Volumenanteile
V	Volt
Vis	sichtbare Strahlung (engl. <i>visible</i>)
Vpp	<i>voltage peak-to-peak</i>
z	Ladung

1 Zusammenfassung

Die Rückverfolgbarkeit von Waren, insbesondere Lebensmitteln und Futtermitteln wird zunehmend relevanter, vor allem da der Handel und die Lieferketten im Zuge der Globalisierung immer komplexer werden. Körnermais als wichtiges Futtermittel ist besonders anfällig für betrügerische Umdeklarationen der Warenherkunft. Der Handelsmarkt für Körnermais ist international, gleichzeitig beeinflusst der Anbauort nicht nur die ernährungsphysiologische Qualität des Futtermittels, sondern ist auch ein Risikofaktor für eine Kontamination mit Mykotoxinen. Heutzutage erfolgt die Verifikation des Anbaulandes jedoch immer noch hauptsächlich über Lieferdokumente. Folglich ist eine objektive analytische Verifikation des Anbauortes von Körnermais erforderlich. Insbesondere die Analyse des Metaboloms (Gesamtheit aller Stoffwechselprodukte) erscheint naheliegend, da dieses dem Phänotyp am nächsten ist und somit empfänglich für Umwelteinflüsse ist.

Daher wurde die Möglichkeit der Herkunftsverifikation von Körnermaisproben in dieser Arbeit untersucht. Hierfür kamen zwei analytische Verfahren zum Einsatz: die flüssigkeitsgekoppelte Massenspektrometrie (LC-MS) sowie die FOURIER-transformierte Nahinfrarotspektroskopie (FT-NIR-Spektroskopie).

Um sicherzustellen, dass die beobachteten Unterschiede von der unterschiedlichen geographischen Probenherkunft herrührten und nicht aus Sortenunterschieden resultierten, wurden insgesamt 81 Proben aus Landessortenversuchen analysiert. Dabei konnte gezeigt werden, dass – im Einklang mit der Literatur – der Einfluss der Herkunft mit im Mittel 12.0-25.6 % viel bedeutsamer ist, als jener der Sorte mit lediglich 7.8-9.1 % beeinflussten Features.

Bei der Analytik mittels FT-NIR-Spektroskopie stellte sich eine Vermahlung der Körnermaisproben als vorteilhaft heraus. Insgesamt wurden 101 Proben aus fünf unterschiedlichen Ländern analysiert. Mittels statistischer Tests und multivariater Datenanalyse wurden herkunfts-spezifisch beeinflusste Bereiche der FT-NIR-Spektren identifiziert. Diese entsprachen den Absorptionsbanden von Proteinen, Stärke sowie Lipiden. Diese Bereiche des Spektrums wurden zur Klassifikation mittels des *Support-Vector-Machines*-Algorithmus mit einer linearen Kernelfunktion genutzt, wobei eine Erfolgsrate von 94.7 % mittels Kreuzvalidierung (10 *folds*, 100 Wiederholungen) erreicht wurde. Die Robustheit der Methode wurde zudem überprüft, indem weitere 72 Proben aus Frankreich und Ungarn analysiert wurden, die mittels eines anderen Verfahrens vermahlen und zudem für 24 h gefriergetrocknet wurden.

Im Falle der LC-MS-Analytik wurden insgesamt 151 Proben aus sieben unterschiedlichen Ländern untersucht. Um das Körnermais-Metabolom analytisch zu erfassen, wurden zunächst zwei Methoden für die Extraktion von polaren bzw. unpolaren Metaboliten optimiert. Dabei zeigte sich, dass polare Metabolite bevorzugt mittels einer Mischung aus Methanol und Wasser (2:1 v/v) extrahiert werden konnten, während unpolare Metabolite optimaler Weise mittels einer BLIGH & DYER-basierten Extraktionsmethode abgedeckt wurden.

Mittels statistischer Tests und multivariater Datenanalyse wurden herkunfts-spezifisch beeinflusste Features in den LC-MS-Daten der unpolaren Metabolite, welche im positiven Ionenmodus detektiert werden konnten, ermittelt. Dieses Subset des Metaboloms erwies sich

sowohl als sehr abundant, als auch gleichzeitig am robustesten gegenüber eventuellen Sorteneinflüssen.

Folglich wurde diese Methode zur analytischen Herkunftsverifikation herangezogen. Mittels statistischer Tests und multivariater Datenanalyse wurden 20 herkunfts-spezifisch beeinflusste Features in den LC-MS-Analysen identifiziert. Die Klassifizierung mittels *Random-Forest*-Algorithmus erzielte dabei eine Erfolgsrate von 90.5 % in der Kreuzvalidierung (10 *folds*, 100 Wiederholungen). Die Markersubstanzen wurden als Lipide identifiziert, die den Klassen der Phospholipide sowie den Di- und Triglyceriden angehörten. Die prinzipielle Eignung der zuvor mittels hochauflösender LC-MS-Analyse auf der Ebene der unfragmentierten Analyten (MS^1 -Ebene) halb-quantitativ detektierten Markermetabolite wurde für ein Subset von neun Markern ebenfalls quantitativ auf MS^2 -Ebene verifiziert. Dabei wurden kongruente Klassifizierungsergebnisse erhalten.

Die Robustheit der 20 identifizierten Markersubstanzen, wurde zudem ausführlich untersucht. Insbesondere wurden die Metabolite in Hinblick auf ihre Lagerungsstabilität der erstellten Extrakte verifiziert.

Schließlich wurde eine Datenfusion, der mittels der beiden Analysemethoden erhobenen Daten durchgeführt. Es konnte gezeigt werden, dass eine Fusion der Analysedaten auf *low-level* Ebene zu besseren Klassifizierungsergebnissen mittels des *Random-Forest*-Algorithmus führte (97.3 % für die Fusion der 20 LC-MS-Metabolite und der 216 spezifischen Wellenzahlen der FT-NIR-Spektren), als die jeweiligen Einzelmethode. Dieses Ergebnis unterstreicht die Möglichkeit der Optimierung eines Klassifikationsmodells durch die Nutzung zweier zueinander orthogonaler Analysemethoden.

2 Abstract

The traceability of goods, especially food and feed, is an emerging topic, as trade and supply chains are more and more complex in a globalized market. Especially grain maize is prone to fraud regarding the origin of growth: The market for maize is internationalized and the growth location could not only influence the quality of grain maize, but also influence the risk of mycotoxin contamination. Today, this verification of the origin is mainly done by checking shipping documents, which may be distorted. In conclusion there is a high need of analytical methods for verification e. g. to be able to objectively trace back the origin of feed samples. Analyzing the metabolome for origin verification is straightforward, as it is closest to the phenotype and therefore, very sensitive to environmental influences.

Therefore, the ability of analytical methods for a verification of the geographical origin of grain maize is accessed in this work. Two analytical methods, liquid chromatography online-coupled to mass spectrometry (LC-MS) as well as FOURIER-transformation near-infrared spectroscopy (FT-NIR spectroscopy) were used.

To assure that the observed variations were caused by differences of growth origin (and not by varietal differences) in total 81 samples of varietal trials were analyzed first. It could be shown, that in accordance with literature the influence of the geographical location of the growth origin had a much bigger influence (mean 12.0-25.6 vs. 7.8-9.1 % regulated features by origin resp. variety).

For FT-NIR spectroscopy, it was also found that the preferred sample form for the NIR beam is a ground sample. In total, 101 samples from five different countries were measured by FT-NIR spectroscopy for building an origin differentiation model. By using multivariate data analysis specific marker bands suitable for determining the origin of grain maize could be identified, which were located in the regions of absorbance of proteins, starches and lipids. Classification algorithms showed a successful rate of 94.7 % in a 100 times repeated 10-fold cross-validation support vector machines classification using a linear kernel function. The robustness of the method was further accessed by measuring another 72 samples from France and Hungary, which were prepared by a different mill and lyophilized afterwards. In case of LC-MS analysis, 151 samples from seven different countries were considered. To assure that the metabolome was optimal covered first two extraction methods were first optimized for maximum coverage. It turned out, polar metabolites of grain maize are preferently extracted using a mixture of methanol and water (2:1 v/v), while non-polar metabolites are best covered by using a BLIGH & DYER-based extraction method.

By using multivariate data analysis 20 promising specific marker features could be identified by means of LC-MS analysis from non-polar metabolites detected in the positive ion mode. Classification of statistical relevant markers using random forest algorithm showed a successful rate of 90.5 % in a 100 times repeated 10-fold cross-validation. The robustness of the model to different varieties was tested, as well as its performance when the number of marker features is downscaled to be more feasible to use. The selected markers were identified to belong to the class of triglycerides, diglycerides and phospholipids. The feasibility of these marker substances in terms of a quantitative analysis were exemplary conducted for a subset of nine marker substances, which resulted in comparable classification results.

The robustness of the 20 identified marker metabolites was carefully analyzed. In particular, the stability of the extracts obtained was investigated.

Finally, the data obtained by the two analytic methods were also combined to a low-level fusion classification model. Using this strategy correct classification results of 97.3 % (using fused data of 20 LC-MS marker substances and 216 wavenumbers of FT-NIR spectra) could be achieved using the random forest algorithm, which performed better than the respective analytical methods without combination. These results underline the possibility of optimization of a classification model by the combined usage of two different orthogonal analytical approaches.

3 Einleitung

Die Warenströme von Lebensmitteln, wie auch Futtermitteln – die Ausgangsstoffe für tierische Lebensmittel darstellen – sind zunehmend internationalisiert^{1,2}. Insbesondere Mais, aber auch andere Cerealien, wie Weizen oder Reis sind dabei einer starken Import- bzw. Exportbestrebung ausgesetzt^{1,3,4}. Die Nachfrage der Endverbraucher nach Lebensmitteln mit besonderen Eigenschaften, wie biologischer Anbau/Aufzucht, oder bestimmte Pflanzensorten/Tierrassen oder auch bestimmter geographische Herkunft („Regionalität“) hat stark zugenommen. Dies umfasst auch den Wunsch nach der regionalen Herkunft von Futtermitteln für Tiere⁵. Beide Faktoren erzeugen gemeinsam ein Bedürfnis nach einer möglichst genauen und lückenlosen Rückverfolgbarkeit der Erzeugnisse entlang der Lieferketten, sowohl für den Endverbraucher als auch die Lebensmittelindustrie^{2,6}. Insbesondere, da die zugesicherten Eigenschaften, manchmal nicht unmittelbar durch den Verbraucher selbst oder analytisch trivial zu verifizieren sind (z. B. „biologisch“, oder „fair trade“)⁶. Dies begünstigt einen weiteren Faktor, bewusstes Verfälschen von Lebens-/Futtermitteln („*Food Fraud*“) mit der Absicht einen finanziellen Gewinn zu erzielen^{7,8}.

Insbesondere die geographische Herkunft steht dabei in einem besonderen Fokus, da keine unmittelbar offensichtliche Eigenschaft versprochen wird, sodass der Verbraucher auf die korrekte Deklaration vertrauen muss. Daher wird der Verbraucher durch eine Reihe von Vorschriften des Gesetzgebers geschützt: Zunächst bietet das Lebensmittel- und Futtermittelgesetzbuch generell gesetzlichen Schutz vor einer Täuschung. Für Futter- und Lebensmittel gelten analoge Vorgaben, die in §19 bzw. §11 zu finden sind⁹. Die Rückverfolgung von Lebens- und Futtermitteln entlang der Lieferkette ist dabei ein zentrales Element, die in den Verordnungen der europäischen Union (EU) (EG) Nr. 178/2002 und insbes. für Futtermittel in (EG) Nr. 183/2005 genauer definiert wird¹⁰⁻¹².

Die Überprüfung der Lebensmittelherkunft erfolgt auch heutzutage immer noch vor allem über Lieferdokumente. Da die geographische Herkunft von Lebens- und Futtermitteln in bestimmten Fällen zu strengeren Kontrollen und zusätzlichen Analysen führen kann (siehe Durchführungsverordnung (EU) Nr. 884/2014¹³), besteht der grundsätzliche Anreiz, durch die Verschleierung des realen Ursprungslandes den Import zu vereinfachen.

Dies zeigt sich auch in den OPSON-Operationen von Interpol und Europol, die der Bekämpfung der kriminellen Lebensmittelverfälschung sowie des Verkaufs minderwertiger Lebensmittel dienen^{14,15}. Somit ist festzustellen, dass ein Bedarf besteht, die Rückverfolgbarkeit entlang der Lieferkette weiterzuentwickeln, da das System der Überprüfung per Lieferdokumenten fehleranfällig sein bzw. in betrügerischer Absicht getäuscht werden könnte^{6,7,16}.

Ein innovativer Ansatz besteht darin, Lebens- bzw. Futtermitteln die als Schüttgut gehandelt werden, *Tracer* beizumischen. Als *Tracer* dienen dabei bspw. RFID-Chips (engl. *radio-frequency identification*), aber auch Tabletten aus Methylcellulose, auf denen ein Barcode aufgedruckt wird¹⁷⁻¹⁹. Es wird erwartet, dass außerdem die Entwicklung von Blockchain-basierten Verfahren, die Rückverfolgbarkeit von Waren insbes. auch von Lebensmitteln und Futtermitteln, grundlegend verändern könnte^{6,16,20,21}.

Darüber hinaus sind analytische Methoden als objektiver Ansatz zur Verifikation der geographischen Herkunft ein wichtiges Betätigungsfeld der lebensmittelchemischen Forschung. In

der verwendeten Literatur sind daher bereits eine Reihe analytischer Ansätze zur Verifikation vorhanden^{6,22–27}.

Während die Rückverfolgbarkeit von Lebensmitteln aufgrund der direkten Auswirkungen auf den Verbraucher im Fokus der Öffentlichkeit steht, ergibt sich prinzipiell die Herausforderung ebenso für Futtermittel. Eine Übertragung negativer Eigenschaften von Futtermitteln auf Lebensmittel über lebensmittel-erzeugende Tiere ist möglich^{2,28,29}. Zu den herkunftsbezogenen Risikofaktoren gehören insbesondere Kontaminationen mit Schwermetallen, Mykotoxinen sowie Rückstände von Tierarzneimitteln und Pflanzenschutzmitteln^{28,30–33}. In der Verordnung (EU) Nr. 178/2002 wurde im Jahre 2002 der Grundsatz definiert, dass die Lebensmittelsicherheit vom Erzeuger zum Verbraucher sichergestellt werden muss¹⁰. Obwohl bereits dort Futtermittel als wichtiger Bestandteil der Lebensmittelkette erwähnt wurden, erfolgte die Definition der Rückverfolgbarkeit von Futtermitteln erst im Jahre 2009 durch die EU-Verordnung Nr. 767/2009³⁴.

Analog zu den Fällen von *Food Fraud* im Bereich der Lebensmittel, gab es ebenso Fälle der Verfälschung von Futtermitteln^{35–37}. Besonders zu nennen, sind hierbei in chronologischer Reihenfolge: Dioxin, Aflatoxine und polychlorierte Biphenyle. Diese Verunreinigungen hatten zum Teil auch Auswirkungen auf Lebensmittel, da die kontaminierten Futtermittel zur Fütterung von Nutztieren verwendet wurden. So führte die Verfütterung von Mais aus Serbien, welcher das Mykotoxin Aflatoxin B1 enthielt im Jahre 2013 zu einem Futtermittelskandal. Dieser wurde jedoch erst aufgedeckt, als bereits ca. 10000 Tonnen des belasteten Mais verkauft und z. T. bereits in die Lebensmittelkette gelangt waren³⁶.

Ein weiterer bedeutender Fall trat im Jahre 2000 in den USA auf, als die gentechnisch veränderte Maissorte StarLink – welche zu diesem Zeitpunkt lediglich eine Zulassung als Futtermittel besaß – in Lebensmitteln nachgewiesen werden konnte^{38,39}. Insgesamt wurden vermutlich bis zu 124 Mio. Bushel Mais (entspr. 3.15 Mio. Tonnen⁴⁰) kontaminiert. Die Rate an beanstandeten Ladungen betrug zeitweise 5-10 %⁴¹. Die Kontamination hatte auch Auswirkungen auf den US-Export, da Japan und Südkorea nur StarLink-freien Mais importierten. In der Folge brach der Import von US-Mais dort um bis zu 50 % ein⁴¹.

Die Notwendigkeit einer Herkunftsverfolgung ist vor allem für Futtermittel, die aus Cerealien bestehen, gegeben. Zum einen wird in Amerika bzw. Europa ein hoher Anteil an Cerealien in Futtermitteln für Schweine oder Geflügel von bis zu 60 % verfüttert²⁸. Der Handel mit Getreide, wie Weizen, oder Mais ist zudem stark internationalisiert^{3,4}, was die Rückverfolgung im Bedarfsfall erschwert. Eine herkunftsbezogene Risikoanalyse von importierten Lebens- und Futtermitteln erfolgt bereits heute in dem von der EU etablierten RASFF-System (*Rapid Alert System Food and Feed*)⁴².

Als besonders vielversprechend für die Sicherung der Rückverfolgung gelten dabei die analytischen Verfahren der flüssigkeits-gekoppelten Massenspektrometrie (LC-MS) sowie der Fourier-transformierten Nahinfrarot (FT-NIR)-Spektroskopie. Beide Technologien wurden bereits an der *Hamburg School of Food Science* erfolgreich zur Herkunftsunterscheidung von weißem Spargel (LC-MS und FT-NIR-Spektroskopie) bzw. Haselnüssen eingesetzt. Im Falle der LC-MS-Analyse erwies sich dabei das (unpolare) Metabolom als besonders geeignet^{43–46}, weshalb dieses ebenso in dieser Arbeit im Fokus stehen sollte. Eine Übertragung der dort gewonnenen Erkenntnisse auf

Körnermais erschien daher vielversprechend. Das Verfahren der FT-NIR-Spektroskopie war im Falle der Matrix Körnermais darüber hinaus von Interesse, da die Methode bereits in der Cerealienindustrie etabliert ist^{27,47,48}.

In dieser Arbeit sollten daher analytische Methoden zur Herkunftsverifikation von Körnermais etabliert werden. Das Ziel war, mittels chemischer bzw. physikalischer Analysen ein Profil der Körnermaisproben zu erhalten, welches mit jenen von authentischen Proben aus den Herkunftsländern verglichen werden kann. Die Erstellung der Profile der Proben sollte dabei zunächst mit *non-targeted* Strategien erfolgen, bei der ein Maximum an analytischen Informationen erhoben wird. In einem zweiten Schritt erfolgte eine Datenreduktion, mittels multivariater Datenanalyse verbunden mit der Fokussierung auf herkunftsspezifisch beeinflusste Messdaten. Anhand dieser herkunftsspezifischen Fingerprints der Körnermaisproben sollten dann die Herkunft der Proben analytisch ermittelt werden.

Die Analytik der Proben erfolgte dabei in Kooperation mit der Fachgruppe 83 („Produktidentität und Rückverfolgbarkeit“) des Bundesinstituts für Risikobewertung (BfR), welche dieselben Körnermaisproben ebenfalls mittels Kernresonanzspektroskopie (engl. *nuclear magnetic resonance spectroscopy*, NMR-Spektroskopie), Fourier-transformierter Infrarotspektroskopie (FT-IR) sowie Stabilisotopenanalyse analysierte.

4 Theorie

4.1 Mais

4.1.1 Botanische Einordnung und Geschichte

Die Bezeichnung „Mais“ wird sowohl für die Maispflanze (*Zea mays* L.), als auch für die an der Pflanze gebildeten Maiskörner verwendet. Die Maispflanze gehört botanisch zu der Familie der Süßgräser (*Poaceae*)⁴⁹.

Je nach Verwendungszweck wird dabei entweder die komplette Pflanze oder nur ein Teil von ihr, geerntet. Die komplette Pflanze wird bspw. zur Ethanolproduktion, zur Erzeugung von Silage als Futtermittel bzw. in Biogasanlagen verwendet. Beim „Corn-Cob-Mix“ wird nur ein Teil der Maispflanze geerntet und zwar die Maiskolben, genauer die Maiskörner mitsamt Maisspindel^{50,51}. Schließlich existiert noch der Anbau des Körnermais, bei dem nur die Körner geerntet werden. Während im Falle der zuvor genannten Anbauzwecke die Pflanze relativ früh geerntet wird, erfolgt bei der Nutzung als Körnermais die Ernte erst, nachdem der Reifeprozess der Körner abgeschlossen ist^{52–55}.

Die Maispflanze ist in Mittelamerika in den Ländern Mexiko, Honduras, Nicaragua heimisch und stammt evolutionär vermutlich von dem Wildgras Teosinte ab⁴⁹. Dieses bildet jedoch eine eher zierliche Ähre aus, die nur 5-12 Körner trägt. Im Gegensatz dazu, produzieren Maispflanzen, größere Kolben mit jeweils ca. 500 Körnern⁴⁹.

Die evolutionäre Wiege des Mais, liegt in Mittelamerika, von wo aus der Mais zunächst in Südamerika, später dann auch in Nordamerika angebaut wurde. Durch Christoph Kolumbus wurde er 1493 zudem nach Europa gebracht⁴⁹, wo er zunächst nur in Südeuropa angebaut wurde. In Deutschland wurden erst ab den 1960er Jahren vermehrt Mais kultiviert, nachdem die Einführung von Hybridsorten eine Anpassung an die klimatischen Bedingungen ermöglichte⁴⁹. In Deutschland verdreifachte sich der Ertrag von Körnermais im Zeitraum 1960-2000, der von Silomais verdoppelte sich im demselben Zeitraum⁵⁶.

4.1.2 Anbauformen und Sorten

Die Maispflanze ist eine einjährige sommerannuelle Pflanze. Als C₄-Pflanze ist sie an klimatisch wärmere Bedingungen der Tropen bzw. Subtropen adaptiert. Dies gestattet die optimale Nutzung von Kohlenstoffdioxid^{49, 57}. Die Wuchshöhe beträgt bis zu 2.5 Meter, es wird ein unverzweigter Halm mit einem Durchmesser bis 5 cm, ausgebildet. Die Laubblätter sind bis zu 1.5 Meter lang, 4-10 cm breit und befinden sich wechselseitig am Halm⁵⁷. Mais ist ein Fremdbefruchter, d. h. die Bestäubung erfolgt durch Pollen anderer Maispflanzen, welche durch den Wind herangetragen werden. Die männliche Blütenstände mit dem Pollen befinden sich oben am Halm; die weibliche Blüten – aus denen die Maiskolben entstehen – befinden sich hingegen in den Blattachseln⁴⁹. Nach der Bestäubung bilden sich pro Pflanze 1-3 Kolben aus, die von Hüllblättern (Lieschblättern) umgeben sind. Der Maiskolben besteht aus vielen kleinen Maiskörnern, welche auf der Spindel angeordnet sind. Die Aussaat erfolgt in Deutschland Ende April/Anfang Mai. Die Ernte kann je nach Verwendungszweck Ende September/Anfang Oktober für Silomais oder zwischen Oktober und

November bei Körnermais durchgeführt werden^{51,52,54,56}. Es werden ca. 6-10 Pflanzen pro Quadratmeter angepflanzt⁵⁸.

Ab den 1930er Jahren setzen sich in den USA zunehmend Hybridsorten gegenüber den bis dahin verwendeten offen bestäubten Sorten (*open pollinated variants*) durch. Hybridsorten wurden durch die Kreuzung von unterschiedlichen Maissorten erzeugt. Auf diese Weise konnte der Maisertrag in den USA pro Hektar von ca. 1380 Tonnen im Jahre 1930 auf 7970 Tonnen im Jahre 1990^{59,60} erhöht werden.

Körnermais umfasst die getrockneten Samen der Maispflanze ohne die Spindel. Das Maiskorn lässt sich botanisch in die Bestandteile Endosperm, Keimling, Perikarp und *Tip Cap* unterteilen⁶¹. Diese Bestandteile nehmen unterschiedliche Anteile des gesamten Maiskornes ein. Zudem sind die Gehälter von Stärke, Protein, Öl und Mineralien unterschiedlich auf die verschiedenen Kompartimente verteilt. Ein komplettes Maiskorn besteht zu rund 73 % aus Stärke, 9 % Protein, 4 % Öl sowie zu etwas über 1 % aus Mineralien⁶¹. Während das Endosperm mit ca. 87% den höchsten Anteil an Kohlenhydraten birgt, besteht der Keimling zu mehr als 1/3 aus Öl sowie zu je ca. 10 % aus Mineralien und Zucker⁶¹. Die Einlagerung von Stärke in das Endosperm erfolgt in Form von Stärkepartikeln eingebettet in einer Proteinmatrix⁶¹. Mais enthält insbesondere relativ viel Vitamin E. Vitamin E ist eine Sammelbezeichnung für Tocopherole und Tocotrienole, welchen eine antioxidative Aktivität gemein ist. Vornehmlich α -Tocopherol und γ -Tocopherol sind in größeren Mengen in Mais nachweisbar, in geringeren Mengen auch β -Tocopherol und δ -Tocopherol. Dabei ist charakteristisch, dass mehr γ -Tocopherol, als α -Tocopherol vorhanden ist⁶². Im Hinblick auf das Aminosäurespektrum ist hervorzuheben, dass Mais nur relativ wenig Lysin und Tryptophan enthält⁴⁹.

Der Ertrag von Körnermais und Silomais ist häufig abhängig von den angebauten Maissorten. Um die optimale Maissorte für den Anbauzweck und die klimatischen Bedingungen zu bestimmen, werden jährlich in den Bundesländern von den Landwirtschaftskammern Anbauversuche (Landessortenversuche) an bestimmten Standorten mit unterschiedlichen Maissorten unternommen. Die verschiedenen Maissorten werden sortenrein geerntet und in Hinblick auf ihren Ertrag evaluiert. Dabei erfolgt generell eine Unterteilung nach Verwendungszweck in Silomais und Körnermais. Es werden üblicherweise folgende Parameter für Körnermais erhoben: Wuchshöhe, Trockenstresssymptome, Anteil Stängelfäule, Trockenmassegehalt bei Ernte und der Relativertrag bei 86 % Trockenmasse⁶³. Die Sortenlandschaft im Maisanbau ist vielfältig. Für Körnermais werden bis zu ca. 50 verschiedene Sorten je Standort jährlich evaluiert^{63,64}. Dagegen ist die Fluktuation der Maissorten weniger ausgeprägt. Manche ertragreiche Sorten sind teilweise seit mehreren Jahren in der Evaluierung durch die Landessortenversuche. Durch die Nutzung von bei den Landessortenversuchen gut abschneidenden Maissorten, könnte es möglicherweise zur Bevorzugung bestimmter Sorten je nach Anbauregion kommen.

Die Körnermaissorte hat dabei jedoch lediglich für den Maisbauern bei der Anpflanzung eine praktische Bedeutung, bzw. beim Verkauf von Saatgut. Sie wird jedoch regelhaft nicht beim Verkauf der geernteten Körner ausgezeichnet (bspw. als Qualitätsmerkmal o.ä.). Zusätzlich ist von Belang, ob die Maissorte gentechnisch verändert wurde oder nicht; da die EU die Kennzeichnung von gentechnisch veränderten Lebens- bzw. Futtermitteln in der EU-Verordnung Nr. 1829/2003 vorsieht⁶⁵. In den USA existiert in ähnlicher Weise eine Unterscheidung, ob eine gentechnisch

veränderte Sorte (GM, engl. *genetically modified*), nur als Futtermittel, oder für den menschlichen Verzehr zugelassen ist^{38,39}. Da Körnermais beim Handel per Güterwaggon transportiert wird und die Zwischenlagerung in großen Silos erfolgt^{66–68}, ist eine Trennung der Ware nach exakter Körnermaissorte nicht möglich^{38,39,66,67,69–72}. Die fehlende Segregation kann jedoch auch zu Problemen führen, wie der Rückruf der StarLink-Körnermaissorte im Jahre 2000 eindrucksvoll zeigt^{38,39,67}. Die Einführung einer Segregation von GM und non-GM Mais, würde nach Lin *et al.*, 0.22 US-Dollar/*Bushel* zusätzlich kosten (entspr. 8.66 US-Dollar/Tonne)⁶⁷.

Die genaue Körnermaissorte wird bei geernteter Ware daher generell nicht ausgezeichnet, und ist – bis auf die Trennung zwischen GM vs. non-GM bzw. zwischen Lebens- und Futtermittel, für den Handel unbedeutend. Die exakte Sorte stellt dabei insofern auch kein Qualitätskriterium der Ware dar. Durch die Vermischung der Ernte von mehreren Bauern beim ersten Zwischenhändler, kann es zudem zu der Situation kommen, dass die gehandelten Körnermaischargen, nicht sortenrein mit unbekannter Sortenbezeichnung sind. Sondern dass vielmehr auch eine Mischung aus einer oder mehreren unbekannten Sorten in unbekannten Mischungsverhältnissen vorliegen kann.

4.1.3 Körnermais: Nacherntebehandlung und Nutzung als Futtermittel

Mais dient als Futtermittel für Rinder, Schweine und Geflügel, als Ausgangsstoff zur Ethanolproduktion, als Energieträger in Biogasanlagen sowie als Lebensmittel (Maisstärke, Maiskeimöl, Maismehl, Cornflakes, Popcorn, Zuckermais)^{73,74}. Die Nacherntebehandlung von geerntetem Mais ist, je nach dem Verwendungszweck unterschiedlich: Während Mais, welcher als Futtermittel für Rinder dienen soll, oftmals siliert wird, wird Mais, welcher zur Ethanolproduktion oder in Biogasanlagen verwendet wird, lediglich gehäckselt. Bei Körnermais der als Futtermittel v. a. für Geflügel oder auch als Lebensmittel in Form von Maisstärke, Maiskeimöl oder Cornflakes, dienen kann, werden die Körner von der Spindel und der restlichen Pflanze separiert, getrocknet und dann weiterverarbeitet (siehe **Abb. 4.1**).



Abb. 4.1: Verschiedene Teile der Maispflanze. **A:** Maispflanzen auf dem Feld. **B:** Maiskolben an der Pflanze. **C:** Maiskörner ohne Kolben.

Im Gegensatz zu Silomais, welcher bereits früher geerntet wird, erfolgt die Ernte von Körnermais erst später, wenn die Reifung der Körner weiter fortgeschritten ist und die Blätter der Pflanze bereits ihre grüne Farbe verloren haben. Trotz der fortgeschrittenen Reife ist der Wassergehalt der Maiskörner bei der Ernte mit ca. 35 % zu hoch für eine Lagerung⁷⁵. Ohne eine Trocknung der Maiskörner, würden diese innerhalb von zwei bis drei Tagen keimen.

Die Trocknung, erfolgt mittels angewärmter Luft, die mittels eines Gebläses durch die Maiskornschichtung geblasen wird⁷⁶. So ist es möglich, den Mais von 35 % Feuchte innerhalb von 6 h auf eine Restfeuchte von max. 15 % zu Trocknen⁷⁷. Der Energieaufwand, bzw. die Kosten für Brennstoff (Diesel bzw. Gas) pro Tonne Körnermais sind daher meistens der entscheidende Faktor, ob sich der Anbau von Körnermais, in Konkurrenz zu Weizen, finanziell lohnt. Ein Restfeuchtegehalt von 15 % ist nur zur kurzzeitigen Lagerung geeignet, für eine längerfristige Lagerung zwischen sechs bis 12 Monaten sind Werte von 13 % nötig. Eine Langzeitlagerung ist erst bei einer Restfeuchte von < 13 % sicher möglich⁷⁸. Zudem „atmen“ Maiskörner als lebende Samen auch weiterhin und geben bei Lagerung weiterhin entweder Feuchtigkeit ab, oder nehmen diese je nach Luftfeuchtigkeit wieder auf^{76,79}. Somit ist der Wassergehalt einer Körnermaischarge eher als unsicherer Parameter zu sehen, der ohne analytische Verifikation nicht als hoch verlässlich eingestuft werden kann.

Die Qualitätsparameter für die Beurteilung von gehandeltem Körnermais sind je nach Empfängerland leicht unterschiedlich. Generell wird die Qualität von Körnermais meist anhand folgender Parameter bewertet^{80–82}: Dichte, Feuchtigkeit, Anteil an Kornbruch, Anteil der Körner mit *stress cracks* / *heat damage* sowie Anteil an Fremdmaterialien. Die *Food and Agricultural Organisation* der Vereinten Nationen (FAO) betreibt eine Sammlung von weltweiten Richtlinien für Qualitätsparameter für Lebensmittel (genannt „*Codex Alimentarius*“)⁸¹. Für Mais umfassen diese diverse Anforderungen, u. a. bezüglich des Feuchtigkeitsgehalts, Anteil an Fremdmaterialien, Kornbruch, Insekten, Fremdsäaten sowie den Gehalten an Schwermetallen, Pestiziden und Mykotoxinen⁸¹.

Für Mais der in den USA gehandelt wird, werden die Handelsklassen „US No. 1-5“ festgesetzt⁸⁰. Zur Klassifikation „US No. 1“ ist ein Gewicht von 56 *pound* je Bushel (entspricht 721.6 kg/m³) notwendig. Der Anteil an Körnern, die hitzebedingte *stress cracks* aufweisen, darf 0.1 % und der von beschädigten Körnern insgesamt, darf 3 % nicht überschreiten. Fremdmaterialien dürfen nicht mehr als 2 % enthalten sein. Die Anforderung bezüglich des Feuchtigkeitsgehalts wurde 1985 aus den Anforderungen gestrichen⁴. Für Mais der in der EU gehandelt wird, gelten die Bestimmungen gemäß EU-Verordnung Nr. 642/210 bzw. 1272/2009^{82,83}. Demnach erfolgt die Analyse des Feuchtigkeitsgehaltes mittels Gravimetrie.

In Kanada die Bestimmung des Proteingehalts von Cerealien mittels FT-NIR-Spektroskopie bereits seit den 1970er Jahren üblich⁴⁷. Die FT-NIR-Spektroskopie ist insbesondere bei der Beurteilung des Proteingehaltes sowie der Restfeuchte von Relevanz⁴⁷. Im Falle von Mais wurde die Möglichkeit der Quantifizierung der Bestandteile Wasser, Protein, Stärke und Öl in der Literatur mittels NIR-Spektroskopie demonstriert^{27,84–87}. Ebenso kann eine Mykotoxin-Belastung mittels NIR-Spektroskopie nachgewiesen werden⁸⁸. Weitere Anwendungsfelder mittels hyperspektralen *Imagings* (HSI) sind das Auffinden von Maiskörnern, welche nicht keimfähig sind⁸⁹ oder Frostschäden⁹⁰ besitzen.

4.1.4 Anbauländer und –mengen

Die Produktion von Mais betrug im Jahre 2019/20 weltweit 1.12 Mrd. Tonnen⁹¹. In der EU ist Weizen (44 %, 131.8 Mio. Tonnen) das am meisten angebaute Getreide. Danach folgen Mais (23 %, 70.1 Mio. Tonnen) sowie Hafer (18.6 %, 55.6 Mio. Tonnen)⁹². Getreide wird in der EU

hauptsächlich als Tierfutter genutzt (zu ca. 2/3), ansonsten als Lebensmittel (1/3). Nur 3 % des angebauten Getreides wird für die Erzeugung von Biokraftstoffen verwendet⁹³.

Zu den Hauptanbauländern von Mais zählen die USA, China, Brasilien, Argentinien und Indien⁹⁴. In der EU sind die Hauptanbauländer nach erzeugter Menge die Ukraine, Rumänien, Russland, Frankreich und Ungarn^{92,95,96}. Nach Anbaufläche ergibt sich hingegen folgende Staffelung: Rumänien, Frankreich, Ungarn, Polen, Italien und Deutschland⁹⁵. Hauptimportländer der EU waren 2020/21 die Ukraine, Brasilien sowie Serbien⁹⁷.

In den USA findet der Maisanbau von allem im „*Corn Belt*“ statt, zu dem üblicherweise die Bundesstaaten Illinois, Indiana und Iowa gezählt werden. Aus dieser Region stammen über 40 % der angebauten Menge der USA⁹⁸. Diese lässt sich weiter nach dem beabsichtigten Verwendungszweck differenzieren: Futtermittel machen 39 % aus; für die Produktion von Bioethanol werden 34 % verwendet; knapp 7 % gehen in die Produktion von Stärke, Maiskeimöl und *Corn Sirup*. 17.5 % der Ernte werden exportiert⁹⁹. Die festen Reste des Maissubstrates aus der Ethanolgewinnung, können dabei abgetrennt und getrocknet werden, die sogenannte Trockenschlempe oder engl. *Distiller's Dried Grains with Solubles* (DDGS). Diese ist reich an Proteinen und kann – zu Pellets gepresst – als Futtermittel genutzt werden¹⁰⁰.

In den EU-Staaten wurden 2014/15 auf insgesamt 15.44 Mio. Hektar Maisanbau betrieben. Davon wurden 9.2 Mio ha für den Körnermais anbau genutzt. Unter Berücksichtigung der Ukraine und Russland ergibt sich eine Anbaufläche von insgesamt über 24 Mio. ha. In der Ukraine und Russland wird wesentlich mehr Körnermais als Silomais gemessen an der Ackerfläche, angebaut (2 Mio. ha Silomais gegenüber fast 6 Mio ha Körnermais). Körnermais wird in der EU zu 78 % verfüttert, 16 % werden industriell verwertet und 8 % werden zu Biokraftstoff verarbeitet⁹⁵. Während in Deutschland und Frankreich hauptsächlich Silomais angebaut wird (zusammen 58 % der Produktionsleistung), wird Körnermais hingegen überwiegend in Rumänien (29 %), Frankreich (17 %) und Ungarn (13 %) kultiviert. Weitere Anbauländer für Körnermais sind Deutschland, Italien und Polen⁹⁵. Zuckermais als Gemüse wird hingegen in Relation nur als Nischenprodukt auf 70 Tsd. Hektar, hauptsächlich in Ungarn und Frankreich angebaut⁹⁵.

In Deutschland wird aufgrund des kühleren Klimas wesentlich mehr Ackerfläche für den Anbau von Silomais genutzt (2.22 Mio. ha / 86.7 Mio Tonnen Ertrag), als für die Erzeugung von Körnermais (416 Tsd. ha / 3.66 Mio. Tonnen, incl. *Corn-Cob-Mix*)¹⁰¹. Der Anbau von Zuckermais als Gemüse spielt demgegenüber eine vernachlässigbare Rolle (314.8 ha / 2284.5 Tonnen)¹⁰².

Circa 15 % des weltweit erzeugten Körnermais im Erntejahr 2018/19 (insges. ca. 172 von 1122 Mio. Tonnen) wurden im- bzw. exportiert. Zu den größten Exportnationen gehörten dabei: die USA (50.5 Mio. Tonnen), Brasilien (39.5 Mio. Tonnen), Argentinien (32 Mio. Tonnen) sowie die Ukraine (29.5 Mio. Tonnen). Zu den größten Importationen gehörten hingegen die EU (24 Mio. Tonnen), Mexiko (16.7), Japan (15.8 Mio. Tonnen) sowie Südkorea (10.5 Mio. Tonnen)⁹⁶.

Aufgrund der hierzulande geringen Verbreitung des Körnermais anbaus und des hohen Bedarfs der Landwirtschaft, ist Körnermais eine Ware, die stark aus anderen Ländern importiert wird. Die 2017/18 nach Deutschland importierte Menge an Körnermais betrug 2.57 Mio. Tonnen. Der Selbstversorgungsgrad betrug lediglich 69 %⁹⁴. Die Länder, von denen Deutschland die größten Mengen Körnermais importierte, waren: die Ukraine (knapp 800 Tsd. Tonnen), Polen (ca. 600 Tsd.

Tonnen), Frankreich (knapp 375 Tsd. Tonnen), die Niederlande, Ungarn, Rumänien, Tschechien, die Slowakei und Österreich⁹⁴.

Die Preise für Körnermais (siehe **Abb. 4.2**) sind dabei ein Einflussfaktor für die weltweiten Im- und Exportströme. Auffällig ist dabei insbesondere, dass in China der Preis pro Tonne seit 2011 stark angestiegen ist, auf zeitweise über 400 US-Dollar/Tonne, während die Preise in den übrigen betrachteten Anbauländern unter der Marke von 200 US-Dollar/Tonne blieben.

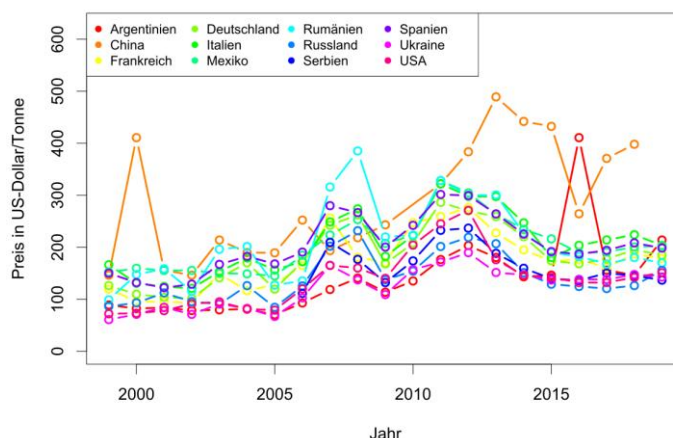


Abb. 4.2: Preise von Körnermais in verschiedenen Staaten von 1999 bis 2019. Nach Daten der FAO¹⁰³.

In den restlichen betrachteten Anbauländern ist ein weitestgehend ähnlicher Verlauf der Maispreise zu beobachten. Dennoch sind gewisse Preisunterschiede auch innerhalb Europas vorhanden. Mais aus den großen nicht-EU-Anbauländern Argentinien, USA, Russland, Ukraine oder Serbien war bspw. im Zeitraum 2015-2018 am günstigsten mit einem Preis von ca. 120-150 US-Dollar/Tonne, während Mais aus EU-Ländern etwas teurer war. Die Preisspanne zwischen bspw. Mais aus Italien mit ca. 205 US-Dollar/Tonne, und Mais aus Frankreich mit nur ca. 175 US-Dollar/Tonne, betrug im selben Zeitraum ca. 30 US-Dollar¹⁰³. Diese Preisunterschiede verdeutlichen, weshalb der Maisanbau sehr von den globalen Warenströmen abhängig ist. Dies erklärt auch die Motivation für *Feed Fraud*, sowie den daraus folgenden Bedarf an Analysemethoden für die objektive Verifikation des geographischen Ursprungs, abseits von Lieferdokumenten.

4.2 Metabolomics

4.2.1 Omics – Technologien

Im Zuge des technologischen Fortschritts hat sich unter anderem auch die biomedizinische Forschung weiterentwickelt, von der Analyse von Einzelparametern hin zu der Betrachtung ganzer Parametergruppen. Dabei werden insbesondere Veränderungen aufgrund von exogenen oder endogenen Einflüssen untersucht, welche sich mit den sogenannten *omics*-Technologien nachvollziehen lassen. Dazu gehören die Bereiche *Genomics*, *Transkriptomics*, *Proteomics* und *Metabolomics*, teilweise wird noch *Isotopolomics* hinzugerechnet^{6,25}.

Der Bereich *Genomics* befasst sich dabei mit der Sequenzierung von vielen bzw. möglichst allen Genen eines Organismus, also seines Genoms^{6,104}. Die Analyse des Genoms erfolgt über *Next-Generation-Sequencing-Verfahren*^{22,104}. Damit verwandt ist der Bereich *Transkriptomics*, welcher

sich mit dem Expressionsmuster der mRNA (Boten-Ribonukleinsäure, engl. *messenger ribonucleic acid*) beschäftigt, das wiederum von der DNA (Desoxyribonukleinsäure, engl. *deoxyribonucleic acid*) transkribiert wird^{105,106}. Diese Analyse erfolgte häufig mittels sogenannter Microarrays, auf denen die mRNA von bestimmten Genen einer Probe immobilisiert und anschließend quantifiziert wird, oder neuerdings ebenfalls mittels Next-Generation-Sequencing-Verfahren^{105,107}. Einen Schritt später setzt der Bereich „*Proteomics*“ an, unter diesem Begriff wird die Analyse der Proteine eines Organismus subsummiert²³. Die Analytik erfolgt überwiegend mittels massenpektrometrischer Verfahren^{108–110}, wobei je nach Analysenverfahren entweder intakte Proteine (*top-down* Ansatz), oder durch tryptischen Verdau erzeugte Peptide (*bottom-up* Ansatz) analysiert werden^{6,108}. Oftmals erfolgt vor der massenspektrometrischen Analyse eine Auftrennung der Proteine mittels Gelelektrophorese^{109,110}.

Das Forschungsfeld „*Metabolomics*“ umfasst laut Fiehn die Charakterisierung bzw. Quantifizierung aller Metabolite in einem Organismus¹¹¹. Mit dem Begriff *Metabolomics* verwandt ist der Begriff „*Metabonomics*“, der 1999 von Nicholson definiert wurde als „die quantitative Messung der Metabolite als Auswirkung auf Stimuli bzw. Genetik“¹¹⁰. Heutzutage werden beide Begriffe synonym verwendet¹¹².

Metabolite besitzen ein Molekulargewicht von $< 1500 \text{ Da}$ ^{6,113,114}. Beispiele für Metabolit-Stoffklassen sind z. B. Lipide, Aminosäuren, Nukleinsäuren, Kohlenhydrate, organische Säuren, aber auch sekundäre Pflanzenstoffe^{25,115–117}. Die Gesamtheit aller Metabolite in einem Organismus wird als „*Metabolom*“ bezeichnet^{118,119}. Das pflanzliche Metabolom umfasst laut Fiehn ca. 0.2 bis 1 Mio. Metabolite¹¹¹. Metabolite können außerdem durch Enzyme zu anderen/weiteren Metaboliten umgesetzt werden, dabei sind sie sowohl Edukte als auch Produkte enzymatischer Reaktionen. Sie besitzen eine biologische Halbwertszeit, weshalb die Zusammensetzung des Metaboloms in Gänze einer zeitlichen Variation unterworfen sein kann.

Die erläuterten *omics*-Forschungsfelder stehen untereinander in Verbindung, da von dem Genom die mRNA transkribiert wird und von der mRNA wiederum die Proteine translatiert werden, siehe **Abb. 4.3**^{25,120}. Die *omics*-Kaskade steht somit im Zusammenhang mit dem „zentralen Dogma der Molekularbiologie“ von Crick¹²⁰. Die synthetisierten Proteine agieren teilweise als Enzyme, die weitere Stoffwechselprodukte (Metabolite) ineinander umwandeln. Während die Genomsequenz nahezu unveränderlich ist, ist bereits deren Übersetzung in mRNA teilweise abhängig von Umweltfaktoren¹⁰⁶, welche sich auf das Proteom und das Metabolom noch stärker auswirken. Die zunehmende Relevanz der Umgebung, wird auch dadurch verdeutlicht, dass die *omics*-Kaskade die Wandlung vom Genotyp eines Organismus zum Phänotyp mit allen sichtbaren Merkmalen abbildet. Dabei steht insbesondere das Metabolom dem Phänotyp des Organismus am nächsten, da sich Veränderungen in der Kaskade am stärksten auf die Metabolite auswirken²⁵.

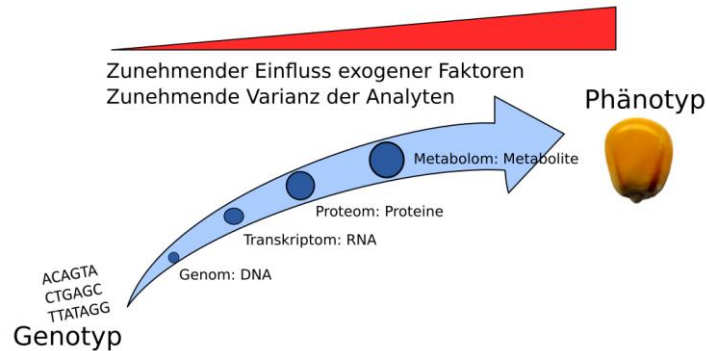


Abb. 4.3: Die *omics*-Kaskade. Abbildung modifiziert nach Dervilly-Pinel *et al.*¹²¹ und Dettmer *et al.*²⁵.

Neben den vier bedeutenden Teilen der *omics*-Kaskade wird teilweise auch noch der Bereich „Isotopomics“ hinzugezählt⁶. Dieses Forschungsgebiet bezeichnet die Analytik der in einem Organismus vorhandenen Elemente bzw. Isotope. So wird bspw. die Stabilisotopenanalyse als Referenzmethode für die Bestimmung der geographischen Authentizität von Wein¹²² eingesetzt.

4.2.2 Grundlagen Metabolomics

Einflussfaktoren auf das Metabolom sind neben dem Genotyp, also bspw. der Maissorte, vor allem auch exogene Faktoren, siehe **Abb. 4.4**. Tiere können sich, ein für sie vorteilhaftes Habitat aktiv suchen und so exogenen Stressfaktoren vermeiden. Im Falle von Pflanzen, die ortsfest sind, ist eine aktive Vermeidung von Stressoren durch Änderungen der Umgebung hingegen nicht möglich. Somit sind Pflanzen im besonderen Maße exogenen (Stress-)Faktoren ausgesetzt¹²². Zu diesen Einflussfaktoren gehören u. a. Temperatur, Wasserverfügbarkeit, Sonnenstunden, Bodenbeschaffenheit (Salzgehalt) sowie Pathogene^{118,122–124}. Daneben können für landwirtschaftlich genutzte Pflanzen weitere Faktoren bedeutend sein, wie Pestizide, Düngung, Bewässerung, Wachstumskontrolle, Anpflanzung in Gewächshäusern sowie der Erntezeitpunkt^{119,123}. Zudem ist eine gewisse Variabilität des Metaboloms zwischen verschiedenen Erntejahren anzunehmen¹²⁴.

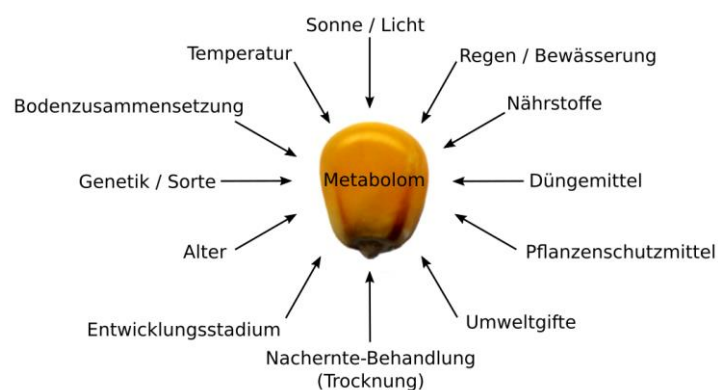


Abb. 4.4: Einflussfaktoren auf das Metabolom.

Neben Faktoren, die das Wachstum und das Gedeihen einer Pflanze beeinflussen, hat jedoch auch die Nacherntebehandlung das Potential, das Metabolom des Ernteguts zu verändern¹¹⁷. Bei Körnermais ist vor allem die Trocknung der Maiskörner hervorzuheben (siehe **Abschn. 4.1.3**). Der Metabolismus des Ernteguts wird durch die Ernte nicht gestoppt. Auch Arbeitsschritte nach dem Anbau haben somit einen Einfluss auf die Metabolitkonzentrationen in der Probe. Folglich können

die Schritte nach der Probennahme, aber vor der eigentlichen Messung, d. h. Lagerungsbedingungen, Homogenisation, Extraktion und ggfls. Anreicherung der Analyten ebenfalls einen starken Einfluss auf die erzielten Ergebnisse aufweisen^{25,119,121,123}.

Im Rahmen der biochemischen Forschung können unter Laborbedingungen generell Faktoren, die eine unerwünschte Variabilität der Analyseergebnisse darstellen würden, entlang der kompletten *omics*-Kaskade vermindert werden. Dies erfolgt in der Regel durch den gleichen genetischen Hintergrund der Versuchstiere/Pflanzen, sowie einheitliche Haltungsformen/Anbaubedingungen und eine einheitliche Ernährung/Düngung¹²¹. Hingegen wird die Variabilität der einzelnen Proben umso größer, je weiter sich von standardisierten Laborbedingungen entfernt wird. Insbesondere das Forschungsfeld *Food Authentication* steht bei der Probenbeschaffung vor der Herausforderung, zwischen einem möglichst definierten/homogenen Probenset und einem Probenset, welches möglichst gut die Realität der Proben auf dem Markt abbildet, was jedoch mit einer gesteigerten Inhomogenität der Proben einhergehen kann. Daher ist für den Bereich der *Food/Feed Metabolomics* eine größere Variabilität der Proben, welche aus marktrealen Bedingungen stammen, anzunehmen. Dies unterstreicht die Notwendigkeit der identischen Behandlung im Rahmen der Probenvorbereitungsschritte im Labor, um eine zusätzliche Variation der Analyten zu vermindern¹²³. Im Vergleich zu der notwendigen Studiengröße bei biochemischen Forschungsprojekten, ist daher für den *Food/Feed Metabolomics*-Bereich eine höhere Probenanzahl als notwendig anzusehen, als dies bei definierten Laborstudien der Fall ist.

4.2.3 Food / Feed Metabolomics

Die Analysemethoden der Forschungsfelder der *omics*-Kaskade wurden sowohl auf Lebensmittel pflanzlichen Ursprungs, als auch tierischen Ursprungs übertragen, um deren Authentizität analytisch zu bestimmen⁶. Vor allem *Metabolomics*-basierte Ansätze zur Authentifizierung sind weit verbreitet. Hierbei werden unterschiedlichste analytische Plattformen genutzt, wie die NMR-Spektroskopie^{24,119,125}, (N)IR-Spektroskopie^{26,27,126}, Gaschromatographie gekoppelt mit Massenspektrometrie (GC-MS)^{25,127} oder Flüssigkeitschromatographie gekoppelt mit Massenspektrometrie (LC-MS)^{6,25,119,128}. In der modernen Analytik werden darüber hinaus aber auch Ansätze wie die Kapillarelektrophorese (CE-MS) oder superkritische Flüssigkeitschromatographie (engl. *supercritical fluid chromatography*, SFC-MS) jeweils gekoppelt mit der Massenspektrometrie, eingesetzt^{6,119,127,129,130}. Ebenfalls ist die Art der untersuchten Lebensmittel sehr vielfältig. Prominente Beispiele sind Fleisch, Milch, Honig, Nüsse, Kakao sowie Tee und Gewürze^{46,131–136}. Da sich eine geringe Qualität von Futtermitteln auch auf die Qualität von Lebensmitteln auswirken kann, ist eine Überwachung der geographischen Authentizität von Futtermitteln ähnlich sinnvoll, wie im Lebensmittelbereich^{21,31,137–139}.

In Bezug auf die Authentizität von Futtermitteln allgemein, finden sich nur relativ wenige Veröffentlichungen: Fernández-Ibáñez *et al.* untersuchten die Möglichkeit Futtermittelbestandteile mittels FT-NIR-Mikroskopie zu unterscheiden, wobei eine Referenz-Bibliothek mit mehr als 48000 Spektren zur Klassifikation erstellt wurde¹³⁹. Tres *et al.* evaluierten die Authentifikation von Hühnerfutter in Bezug auf eine biologische/ökologische vs. konventionelle Herstellungsweise mittels GC-MS sowie mittels NIR-Spektroskopie^{140,141}. Weitere Studien beziehen sich auf die Analytik von Fleisch oder Milch, um entsprechende Rückschlüsse auf die Art der verwendeten Futtermittel zu ziehen^{142,143}.

Die Authentizitätsanalytik von Körnermais, bzw. der ähnlichen Matrix DDGS (einem Nebenprodukt aus der Bioethanolproduktion^{100,144}), wurde bisher vor allem mittels folgender Analysetechniken vorgenommen: GC-MS, LC-MS, Stabilisotopenanalytik, FT-IR und FT-NIR-Spektroskopie^{145–150}. Besonders hervorzuheben sind dabei die Arbeiten von Baniyadi *et al.* und von Tang *et al.* mittels GC-MS bzw. LC-MS zur Abhängigkeit der Regulierung des Metaboloms von Körnermais von der Maissorte bzw. geographischer Herkunft^{151,152}. Diese berichteten von der Tendenz, des Metaboloms eher nach geographischer Herkunft zu variieren, als nach Maissorte. Während 12–90 % der Metabolite von Maiskörnern je nach Herkunft variierten, zeigten weniger als 27 % eine Beeinflussung durch den Genotyp¹⁵². Weniger als 10 % aller betrachteten Metabolite waren dabei gleichzeitig von der geographischen Herkunft und dem Genotyp (Maissorte) abhängig¹⁵². Somit ist anzunehmen, dass bei der Analytik des Körnermais-Metaboloms mittels LC-MS-Analysen, die Metabolite nur relativ selten gleichzeitig einer Beeinflussung von Herkunft und Sorte ausgesetzt sind.

Die Analytik mittels NIR-Spektroskopie ist für Körnermais in Bezug auf die Verifikation der geographischen Herkunft noch nicht durchgeführt worden. Dafür ist diese Analysenmethode bei anderen Cerealien weit verbreitet: Untersuchte Matrices waren u. a. Weizen(mehl)^{153–155}, Hartweizen¹⁵⁶, Linsen¹⁵⁷ sowie Reis¹⁵⁸. So zeigten González-Martín *et al.* dass die NIR-Spektroskopie für die Bestimmung der geographischen Herkunft von Weizen aus verschiedenen Regionen von Chile geeignet ist¹⁵⁵. Zhao *et al.* untersuchten die Unterscheidbarkeit von Weizen aus verschiedenen chinesischen Provinzen. Dabei stellte die Sorte und das Erntejahr keine starken Störfaktoren für die Bestimmung der geographischen Herkunft dar^{153,154}.

Dagegen würde Körnermais bereits – abseits der Authentifikation der geographischen Herkunft - relativ intensiv mittels HSI untersucht: Dabei kommt eine Kamera zum Einsatz, welche neben dem Bereich des sichtbaren Lichts auch noch Teile des Nahinfrarotbereichs von 400 bis 1000 nm abdeckt⁹⁰. Dieser analytische Ansatz wurde genutzt, um die Vitalität von Maiskörnern zu beurteilen⁸⁹ sowie um Frostschäden an Maissaatgut zu detektieren⁹⁰. Daneben wurde von diversen Wissenschaftlern die Eignung der NIR-Spektroskopie zur Detektion von Mykotoxinen in Cerealien untersucht⁸⁸. In Bezug auf eine Authentifizierung existieren bisher lediglich Untersuchungen von Fernández Pierna *et al.*, bei denen mittels HSI die Zusammensetzung von u. a. maishaltigen Futtermitteln überwacht wurden^{27,159}.

Basierend auf den dargestellten Vorarbeiten erscheint eine analytische Authentifizierung der geographischen Herkunft von Mais, insbesondere mittels der LC-MS-Analytik sowie der FT-NIR-Spektroskopie als Methodik vielversprechend.

4.2.4 Analytische Technologien und Strategien

Die unter der Bezeichnung „Metaboliten“ subsumierten Substanzen weisen zum einen eine sehr unterschiedliche chemische Charakteristik auf (Lipide, Aminosäuren, Zucker, organische Säuren, Phenole etc.^{25,115–117}), welche die Erfassung sämtlicher Substanzen in einer einzigen Analyse unmöglich macht^{6,113,119}. Zum anderen liegen diese Substanzen über einen breiten Konzentrationsbereich verteilt vor, der mehrere Größenordnungen umfasst^{6,113,116}. Diese Konzentrationsunterschiede erschweren zusätzlich eine simultane Analyse des gesamten Metaboloms eines Organismus. Somit ist die Untersuchung von Metaboliten mittels einer einzelnen Probenvorbereitung und nachfolgenden einzelnen Analysentechnik auf lediglich eine

Subgruppe an Bestandteilen des Metaboloms beschränkt⁶. Häufig genutzte analytische Verfahren sind vor allem die NMR-Spektroskopie sowie die Massenspektrometrie, siehe **Tab. 4.1**^{25,119,128}.

Tab. 4.1: Vergleich von Massenspektrometrie und NMR-Spektroskopie als Metabolomics-Plattformen^{119,160–166}.

	Massenspektrometrie	NMR-Spektroskopie
Vorteile	• Geringe Nachweisgrenzen (pg/nmol-Bereich) • Hoher dynamischer Bereich • Hohe Anzahl detektierbarer Analyten in einer Analyse (mehrere Tausend) • Kopplung mit Chromatographie möglich	• <i>De novo</i> Strukturaufklärung einfach möglich • Non-destruktive Messung • Gute Reproduzierbarkeit • Quantifizierung leicht möglich
Nachteile	• Analyt muss ionisierbar sein • Destruktive Messung (aber geringe Probenmengen) • Verbindungen mit gleichem <i>m/z</i>-Verhältnis (isobar) nicht trivial zu unterscheiden • Vergleichsweise schlechte Reproduzierbarkeit	• Hohe Nachweisgrenzen ($\geq 1 \mu\text{M}$) • Geringe Anzahl detektierbarer Analyten in einer Analyse (< 200) • Lösungsmittel müssen in der Regel deuteriert sein • Selektiv für bestimmte Kerne (^1H, ^{13}C, ^{31}P, ^{19}F etc.)

Vorteile der NMR-Spektroskopie sind dabei, dass diese Methode einerseits nicht destruktiv ist, andererseits können mannigfaltige Informationen über die Struktur der Analyten direkt gewonnen werden. Nachteile dieser Methode sind hingegen, dass insbesondere die ^{13}C -NMR-Spektroskopie in Relation zu der LC-MS oder GC-MS relativ unempfindlich ist^{125,161}. Wasser stört die NMR-Analyse, weshalb der Wassergehalt von Lebensmittelproben meist zuvor per Gefriertrocknung reduziert werden, oder das Wassersignal unterdrückt werden muss¹²⁵. Außerdem zu nennen sind die hohen Anschaffungskosten und Unterhaltskosten eines NMR-Spektrometers¹⁶⁷.

Bei der Massenspektrometrie handelt es sich um eine destruktive Methode, jedoch ist die Empfindlichkeit weitaus größer als bei der NMR-Spektroskopie. Sie ist außerdem zur Quantifizierung von Analyten gut geeignet¹⁶⁷. Neben der direkten Verwendung als Analyseverfahren ist es üblich, die Massenspektrometrie als Detektor für eine vorgeschaltete chromatographische Trennmethode einzusetzen. Dies erhöht den Informationsgehalt der Analysen deutlich, jedoch auch den Aufwand bei der Datenauswertung. Übliche Kopplungen sind z. B. die Gaschromatographie gekoppelt mit der Elektronenstoßionisations-Massenspektrometrie (GC-EI-MS) oder der Flüssigkeitschromatographie gekoppelt mit der Elektrosprayionisations-Massenspektrometrie (LC-ESI-MS)^{25,119,127}. Daneben gibt es exotischere Varianten, wie SFC-MS¹³⁰ oder CE-MS^{25,168}. In Abhängigkeit der vorhergeschalteten Trennmethode ist die Untergruppe des analytisch zugänglichen Metaboloms unterschiedlich: Während die GC die Analyse von flüchtigen, temperaturstabilen Bestandteilen einer Probe ermöglicht, erlaubt die LC die Analytik von nicht-flüchtigen Bestandteilen einer Probe^{115,116,128}. Wohingegen die Methode der SFC, üblicherweise mit superkritischem Kohlendioxid als Laufmittel, die Detektion eines weiteren Unteranteils des Metaboloms ermöglicht^{130,169}. Die CE hingegen erlaubt prinzipbedingt nur die Auftrennung von geladenen Analyten^{25,168}.

Mit der Massenspektrometrie als Analysator bzw. Detektor können mehrere tausend Analyten zugleich während einer Messung nachgewiesen werden. Durch die Möglichkeit der

Fragmentierung von Ionen im MS ist zudem eine Strukturaufklärung der Analyten möglich. Nachteilig wirkt sich hingegen aus, dass für die GC-MS eine vergleichsweise aufwendige Probenvorbereitung mit Derivatisierung, notwendig sein kann und die Analysenzeiten bei Chromatographie-gekoppelten Methoden sehr lang sind. Zudem besteht die Gefahr von Ionensuppressionseffekten, was sich negativ auf die Quantifizierung der Analyten auswirken kann. Als Folge ist die Chromatographiemethode möglicherweise aufwendig zu optimieren¹⁷⁰.

Die Erfassung der Metaboliten kann dabei entweder semi-quantitativ, d. h. lediglich in Relation zueinander, ohne eine exakte Quantifizierung der einzelnen Markerkonzentrationen mittels Standardsubstanzen erfolgen. Alternativ kann die Messung quantitativ unter Einsatz von entsprechenden Standardsubstanzen durchgeführt werden. Dabei werden üblicherweise *non-targeted Metabolomics* Studien eher qualitativ bzw. semi-quantitativ auf Quadrupol-Flugzeit (engl. *quadrupole time-of-flight*, q-ToF) Massenspektrometern oder Orbitrap-Massenspektrometern durchgeführt. *Targeted* Studien dagegen zumeist quantitativ, überwiegend auf Triple-Quadrupol-Massenspektrometern (engl. *triple quadrupole*, QqQ)^{6,114,127}. Dabei ist die Taktik des stufenweisen Entwickelns einer Authentifizierungsmethode von besonderer Bedeutung: Zunächst wird mittels eines q-ToF-Massenspektrometers ein möglichst hochaufgelöstes Bild des Metaboloms der untersuchten Proben aufgenommen. Anschließend erfolgt mittels multivariater Datenanalyse die Selektion von ca. 20-30 Markermetaboliten, welche besonders zur Beantwortung der Fragestellung/Differenzierung der Probengruppen geeignet sind. Die ausgewählten Metaboliten werden daraufhin quantitativ auf MS/MS-Ebene z. B. mittels eines Triple-Quadrupol-Massenspektrometers analysiert¹¹⁴. Dabei erfolgt eine exakte, quantitative Erfassung der Konzentration in Relation zu Standardsubstanzen. Da diese Art von Geräten weitaus kostengünstiger ist, als hochauflösende q-ToF-Massenspektrometer, ermöglichen sie, Authentifizierungsmethoden als Routineanalytik in die Praxis zu überführen^{6,43,44}. Aufgrund der Messung auf MS/MS-Ebene (Ebene der Fragmentionen, auch MS²-Ebene) kann trotz der geringeren Massenauflösung eines Triple-Quadrupol-Massenspektrometers, eine ausreichende Sensitivität und Spezifität erreicht werden. Dieser Ansatz wurde beispielsweise von Klockmann *et al.* zur Bestimmung der geographischen Authentizität von Haselnüssen durchgeführt^{44,46}.

Aufgrund der großen Datenmengen, die insbesondere bei der Kopplung der Massenspektrometrie mit chromatographischen Trennmethode je Probe anfallen sowie der tendenziell hohen Probenanzahl bei Untersuchungen der Authentizität, erfolgt die Datenauswertung häufig mithilfe der multivariaten Datenanalyse (siehe **Abschn. 4.4.2**)^{6,127,171}.

Die Analytik im Anwendungsbereich der Authentizitätskontrolle erfolgt daneben häufig mittels der FT-NIR-Spektroskopie^{26,172}. Diese Technologie bietet zwar nicht die Möglichkeit, einzelne Biomoleküle in einem Lebensmittel/Futtermittel zu analysieren, dafür ist die Methode verglichen mit chromatographischen Trennmethode sehr schnell¹⁶³. Außerdem benötigt sie, neben der Homogenisation und ggfls. Trocknung von zu feuchten Proben, keine weitere (chemische) Probenvorbereitung. Der eigentliche Vorgang der Spektrenaufnahme ist zudem mit 60-90 s je Probe sehr schnell. Zudem sind die Unterhaltskosten eines FT-NIR-Spektrometers vergleichsweise gering¹⁷³.

4.3 Apparative Methoden

4.3.1 FT-NIR-Spektroskopie

Der Infrarotbereich des Lichtspektrums umfasst die Wellenzahlen von $12500\text{--}10\text{ cm}^{-1}$. Da dieser Bereich sehr groß ist, wird er üblicherweise weiter unterteilt in drei Regionen: die NIR-Region ($800\text{ bis }2500\text{ nm}$; $12500\text{--}4000\text{ cm}^{-1}$), die IR-Region ($2500\text{--}25000\text{ nm}$; $4000\text{--}400\text{ cm}^{-1}$), sowie die Region des fernen IR ($25\text{ }\mu\text{m} - 1\text{ mm}$; $400\text{--}10\text{ cm}^{-1}$). Die NIR-Spektroskopie beruht auf der Detektion der Absorption von Nahinfrarotstrahlung durch die Probe^{47,174,175}. An den Nahinfrarotbereich schließt sich in Richtung kürzere Wellenlängen der Bereich des sichtbaren Lichts an.

Die Analytik von Probenmaterial erfolgt entweder anhand des transmittierten Anteils der (N)IR-Strahlung oder des reflektierten Anteils. Der von der Probe absorbierte Anteil führt aufgrund der periodischen Änderung des Dipolmoments bei Verbindungen mit einem permanenten Dipolmoment zu Schwingungen. Dabei ist der Schwingungszustand nicht der Grundzustand des Moleküls, d. h. die Quantenzahl ν ist größer als 0. Beträgt die Quantenzahl $\nu = 1$ so handelt es sich um die sogenannte „Grundschiwingung“, ist sie hingegen größer als 1 so handelt es sich um „Oberschwingungen“^{175,176}. Die Erreichung eines Schwingungsniveaus ist dabei abhängig von der Energie des Lichtquants $E = h \cdot \nu$. Je größer die Energie umso höher kann dabei die erreichte Quantenzahl sein, siehe **Gl. 4.1**.

$$E_{\nu} = \left(\nu + \frac{1}{2} \right) \cdot h \cdot \nu \quad (4.1)$$

Während bei der IR-Spektroskopie Grundschiwingungen ($\nu = 1$) zu beobachten sind, finden sich im Messbereich der NIR-Spektroskopie hauptsächlich die $1/2$. Oberschwingungen ($\nu = 1$ oder 2) sowie Kombinationsschwingungen. Die Schwingungen höherer Obertöne nehmen dabei an Intensität ab^{174,175,177}. Häufig beobachtete Schwingungen im NIR-Spektrum sind Streckschwingungen entlang der Bindungen von C-H, O-H und N-H, außerdem C=O sowie C $\equiv$ N-Schwingungen¹⁷⁴. NIR-Banden sind breiter als die Banden in einem IR-Spektrum. Daher kommt es häufig zum Überlappen von Banden einzelner Schwingungen, sodass eine direkte und eindeutige Zuordnung einer Bande zu einer bestimmten Schwingung einer funktionellen Gruppe, wie bei der IR-Spektroskopie, nicht möglich ist. Deshalb erfolgt die Auswertung von NIR-Spektren häufig mittels Algorithmen der multivariaten Datenanalyse oder (multivariater) Regressionsanalysen^{26,47,84,85}.

Es existieren verschiedene Bauweisen für (FT-)NIR-Spektrometer. Bei Transmissionsmessungen, liegen NIR-Strahlungsquelle und Detektor auf einer Linie und die Probe befindet sich dazwischen. Diese Bauart wird jedoch nur noch selten verwendet, seitdem Geräte welche den diffus reflektierten Strahlungsanteil messen, verfügbar sind. Deren Strahlungsquelle und Detektor befinden sich auf derselben Seite der Probe in einem Gerät. Diese Bauart wird auch für NIR-Sonden in Getreidemöhlen zur *online* Bestimmung der Restfeuchte und des Proteingehaltes vor bzw. nach der Vermahlung eingesetzt, um in nahezu Echtzeit die Produktströme analytisch zu überwachen^{47,178–180}.

Bei der Analyse von Proben mittels NIR-Spektroskopie, gilt prinzipiell das LAMBERT-BEER'SCHE Gesetz. Da NIR-Banden einen vergleichsweise geringen Extinktionskoeffizienten aufweisen, sind

Transmissionsmessungen mit ca. 1-2 cm Schichtdicke von Fleisch, Käse oder Getreidekörnern möglich⁴⁷. Bei pulverförmigen Proben ist jedoch zu beachten, dass aufgrund der Partikel in der Probe, der NIR-Strahl mit jedem einzelnen Partikel interagiert, also entweder absorbiert, reflektiert oder transmittiert wird. Der Strahlengang ist somit nicht mehr gradlinig und die Pfadlänge des Strahls ist unbekannt (im Gegensatz zu der UV/Vis-Photometrie, bei der die Schichtdicke der Küvette bekannt ist). Das LAMBERT-BEER'SCHE Gesetz ist somit bei pulverförmigen Proben nicht mehr erfüllt. Dies hat zudem zur Folge, dass bei diesen Proben ab einer Schichtdicke > 1 cm fast keine NIR-Strahlung mehr transmittiert wird. In der Summe wird die NIR-Strahlung, aufgrund der Anzahl an Interaktionen mit der Probe, entweder absorbiert oder diffus reflektiert^{47,175}. Daher ist die Messung des diffus reflektierten Anteils der durch eine Probe modulierten NIR-Strahlung weit verbreitet¹⁸¹.

Die Eindringtiefe der NIR-Strahlung in die Probe ist dabei wellenlängenabhängig und reicht bspw. von 4 mm (im Bereich 700-900 nm) bis zu 2-3 mm (900-1900 nm) in Äpfeln. In Cellulose hingegen reicht die Eindringtiefe laut Clarke *et al.* von 27 μm (bei 2380 nm), über 39-61 μm (bei 1675 nm), bis hin zu 180 μm (bei 1100 nm)^{181,182}.

Der Aufbau eines FT-NIR-Spektrometers für die Messung des diffus reflektierten Anteils der NIR-Strahlung ist wie folgt (siehe **Abb. 4.5**): Die NIR-Strahlung wird z. B. durch eine Quarz-Halogenlampe erzeugt, die auch „Wolfram-Halogen“-Lampe genannt wird. Die durch diesen Lampentyp abgegebene Strahlung fungiert als schwarzer Strahler und deckt daher einen breiten Bereich des Spektrums, über das sichtbare Licht, den MIR-Bereich und den NIR-Bereich mit einem Maximum bei ca. 2500 nm, ab¹⁷⁵. Andere mögliche Strahlungsquellen sind bspw. Leuchtdioden oder Nd-YAG-Laser^{175,183}.

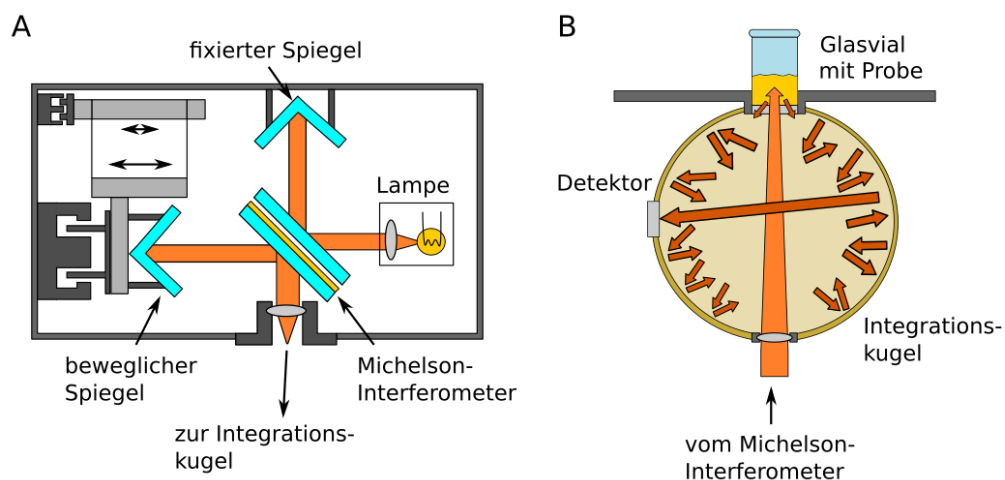


Abb. 4.5: Schematischer Aufbau eines FT-NIR-Spektrometers. **A:** Strahlengang im Spektrometer. **B:** Aufbau der Integrationskugel. Abbildung modifiziert nach Watari¹⁸⁴, Hansen *et al.*¹⁸⁵ und Kok¹⁸⁶.

Der Strahlengang der NIR-Strahlung ist abhängig von der Bauweise des Spektrometers. Es gibt grundsätzlich unterschiedliche Arten der Zerlegung der NIR-Strahlung in seine Wellenlängen: u.a. mittels Interferenzfilter, über Interferometer, mithilfe eines optischen Gitters oder Prismas. Bei FOURIER-Transformation-NIR-Spektrometern wird ein Interferometer verwendet¹⁸³. Ein MICHELSON-Interferometer, besitzt als zentrales Bauteil einen halbdurchlässigen Spiegel, welcher den Strahl aufteilt und auf verschiedenen Wegen über zwei weitere Spiegel leitet. Schließlich kommt es zur

Rekombination der beiden Strahlteile, wobei entsprechend der Weglänge und der Frequenz der Strahlung Interferenz auftritt. Daher kann über die Variation der Weglänge zwischen dem halbdurchlässigem Spiegel und einem der weiteren Spiegel die Wellenlänge variiert werden^{175,183}. Neben dem MICHELSON-Interferometer, existieren noch weitere Arten von Interferometern¹⁷⁵. Vor allem die Nutzung von mehreren Spiegeln in der Geometrie eines halben Würfels, anstelle von planen Spiegeln, bietet den Vorteil, das Interferometer besonders unempfindlich gegenüber Vibrationen zu machen^{186,187}.

Der NIR-Strahl trifft dann auf die zu untersuchende Probe und wechselwirkt mit dieser. Der diffus zurück reflektierte Anteil der Strahlung wird in einer mit Gold beschichteten Integrationskugel gesammelt und seitlich an der Kugel durch einen Detektor nachgewiesen^{186,188}.

Als Detektoren in NIR-Spektrometern dienen vor allem Halbleitermaterialien. Diese besitzen jeweils bestimmte Bereiche, in denen ihre Empfindlichkeit am größten ist. Detektoren aus Silizium können im Wellenlängenbereich von 400-1100 nm eingesetzt werden. Detektoren aus Bleisulfid hingegen im Bereich von 700-2500 nm. Detektoren aus InGaAs decken einen Bereich von 1100 bis 1700 bzw. max. 2300 nm ab¹⁸³.

Das FT-NIR-Spektrometer detektiert dabei die Intensität I des Strahls in Abhängigkeit von der Spiegelposition d (Interferogramm) und errechnet aus dem Detektorsignal ($I(d)$, zeitabhängig) mittels FOURIER-Transformation das NIR-Spektrum ($I(\nu)$ ist wellenzahlabhängig, X-Achse: Wellenzahl, Y-Achse: Absorption), siehe **Gl. 4.2**¹⁷⁵.

$$I(d) = \int I(\nu) \cdot (1 - \cos 2\pi\nu d) d\nu \quad (4.2)$$

Die Ermittlung des NIR-Spektrums über die FOURIER-Transformation hat dabei mehrere Vorteile: Zum einen ist kein Spalt im Spektrometer zur Selektion von einzelnen Wellenlängen notwendig. Alle Wellenlängen werden zudem zeitgleich gemessen, dies kann zu einem besseren Signal-zu-Rausch-Verhältnis führen. Schließlich ist die erzielte Auflösung im Vergleich zu anderen Spektrometer-Arten vergleichsweise hoch¹⁷⁵.

4.3.2 Flüssigkeitschromatographie

Die Flüssigkeitschromatographie (engl. *liquid chromatography*, LC) nutzt als Trennprinzip die unterschiedlich starken Wechselwirkungen zwischen den Analyten und dem Säulenmaterial als stationäre Phase einerseits und den Analyten und der mobilen Phase, dem Eluenten, andererseits. Die unterschiedlichen Wechselwirkungen führen zu einem sich immer wieder neu einstellendem Gleichgewicht der Analytenverteilung zwischen der stationären und der mobilen Phase. Als Resultat kommt es zu einer räumlichen Auftrennung von Stoffgemischen auf der Säule entlang der Flussrichtung. Dabei sind sowohl Eigenschaften der stationären Phase, als auch der mobilen Phase für die Trennwirkung verantwortlich. Die Trennwirkung beruht hauptsächlich auf adsorptiven Effekten^{189,190}.

Die Bewertung der Trennungsleistung kann mittels der VAN-DEEMTER-Gleichung^{189–191} (siehe **Gl. 4.3**) vorgenommen werden. Die Trennleistung eines Chromatographiesystems wird dabei in der „Trennstufenzahl“ H ausgedrückt. Diese gibt an, wie oft sich ein Gleichgewicht zwischen der stationären Phase und der mobilen Phase bezogen auf die Länge der Chromatographiesäule

ausbilden kann. Die Trennleistung ist dabei u. a. abhängig von der Flussrate u_x . Je mehr Trennstufen ein Chromatographiesystem aufweist, umso besser ist es dazu in der Lage, ein Trennproblem zu lösen^{189,190}. Die VAN-DEEMTER-Gleichung setzt sich dabei aus drei Termen zusammen, A die EDDY-Diffusion, B/u_x die (longitudinale) Diffusion und $C \cdot u_x$ der Massetransfer.

$$H = A + \frac{B}{u_x} + C \cdot u_x \quad (4.3)$$

Die EDDY-Diffusion (A) beschreibt den Umstand, dass die Analytmoleküle beim Umströmen der stationären Phase zufällig leicht unterschiedliche Wege durch die Chromatographiesäule nehmen können. Somit kommt es zu einer Verbreiterung der Analytenfront auf der Säule. Zusätzlich führt die Diffusion entlang der Konzentrationsgefälle (B) innerhalb der Säule zu einer Peakverbreiterung, da sich die Analyten als Peak mit vergleichsweise hoher Konzentration durch die Säule bewegen. Diese hohe Analytkonzentration hat beständig das Bedürfnis ein Gleichgewicht mit dem restlichen Laufmittel in der Säule herzustellen, welches eine niedrigere Analytkonzentration aufweist. Dabei ist insbesondere die Diffusion in Flussrichtung relevant, welche abhängig von der Flussrate ist. Der Massetransfer stellt den eigentlichen Chromatographievorgang dar. Aufgrund des geringen Porendurchmessers des Säulenmaterials erfolgt der Massetransfer innerhalb der porösen Partikel mittels Diffusion. Somit kommt es zu einer Verbreiterung der Peaks, da ein Teil der Analytmoleküle gerade innerhalb eines Chromatographiepartikels diffundieren können, während die restlichen Analytmoleküle außerhalb des Partikels sich in Relation schneller mit dem Laufmittelfluss mitbewegen. Dieses Phänomen verringert ebenfalls, wenn auch nur geringfügig, die Auflösung auf Chromatographieebene^{189,190}.

Die VAN-DEEMTER-Gleichung kann dazu genutzt werden, um die optimale Fließgeschwindigkeit eines Chromatographiesystems zu bestimmen. Die optimale Fließgeschwindigkeit ist bei der LC üblicherweise bei eher niedrigen Flussraten, da der Einfluss des Terms der Diffusion in der mobilen Phase (B/u_x) sehr gering ist¹⁸⁹.

Zu Beginn der 1970er Jahre wurde die LC durch die Hochdruckflüssigkeitschromatographie (engl. *high pressure liquid chromatography*, HPLC) abgelöst, die mittels Kurzhub-Kolbenpumpen dazu in der Lage ist, bis zu 400 bar Druck aufzubauen. Dies ermöglichte den Einsatz von geringeren Partikeldurchmessern im Bereich von $< 10 \mu\text{m}$, gegenüber der klassischen Säulenchromatographie^{189,190,192}. Während einfache Pumpen lediglich die Elution mittels eines fest, vorab gemischten Laufmittels leisten konnten (isokratische Elution), sind moderne Pumpen dazu in der Lage die Analyten mittels einer Gradientenelution aufzutrennen. Dabei können zwei oder mehr Eluenten mit unterschiedlichen Geschwindigkeiten gefördert werden. Die verschiedenen Eluenten werden in einer, der Pumpe nachgeschalteten Mischungskammer vermischt, sodass die Elution mittels eines frei wählbaren Laufmittelgemisches erfolgt. Dabei wird zu Beginn des Chromatographielaufs das Substanzgemisch bei einer Laufmittelzusammensetzung aufgebracht, bei der nahezu alle in der Probe enthaltenen Analyten auf der Säule retardiert werden. Anschließend wird die Zusammensetzung des Eluenten entlang eines vorher bestimmten Gradienten verändert. Im Laufe des Chromatographielaufs verändert sich daher die Laufmittelzusammensetzung, sodass alle Substanzen sukzessive wieder von der Säule eluiert werden¹⁹². Um möglichst sämtliche Analyten wieder von der Säule zu eluieren, folgt auf die Gradientenelution oftmals eine Plateauphase, bevor die Säule wieder auf die Ausgangsbedingungen zu Beginn des Gradienten gebracht wird (Äquilibration)¹⁹².

Der Partikeldurchmesser der Füllmaterialien von Chromatographiesäulen ist im Laufe der Zeit immer geringer geworden, da eine geringere Partikelgröße mit besseren Trennleistungen des Chromatographiesystems einhergeht. Dies folgt aus der VAN-DEEMTER-Gleichung¹⁸⁹. Andererseits steigt mit abnehmendem Partikeldurchmesser auch der auftretende Rückdruck, weshalb immer leistungsfähigere Pumpen entwickelt wurden. Dieser Trend mündete schließlich in der Entwicklung der UHPLC (engl. *ultra high pressure liquid chromatography*; Ultrahochleistungsflüssigkeitschromatographie), welche einen Druck von bis zu 1300 bar aufbauen können^{189,193}. Ein anderer Ansatz, um die Problematik eines hohen Rückdrucks zu umgehen, war die Entwicklung von teilporösen Partikeln. Während konventionelle Silicapartikel vollporös sind, besitzen teilporöse Partikel einen soliden Kern, welcher lediglich mit einer porösen Schicht aus Silica ummantelt sind. Dies führt zur Verringerung des Rückdruckes, aber auch des Diffusionsweges^{189,190}.

Der Aufbau einer HPLC-Anlage ist dabei wie folgt (siehe **Abb. 4.6**): Das Lösungsmittel wird zunächst meistens durch einen *Degasser* geführt. Dieser entfernt physikalisch gelöste Gase aus dem Laufmittel^{190,195}. Danach erreicht das entgaste Laufmittel die HPLC-Pumpe, die meistens zwei Kolben, einen je Laufmittel, enthält. Diese können unabhängig voneinander angesteuert werden, sodass eine flexible Laufmittelzusammensetzung erzielt werden kann. Nach den Pumpenkolben folgt eine Mischkammer, die der Vermischung der beiden Laufmittelsorten dient. Das Eluentengemisch wird danach entweder durch einen Autosampler oder ein manuelles Injektionsventil gefördert^{190,192}. Der Autosampler dient als Roboter, der die automatisierte Zuführung mehrerer Proben nacheinander ermöglicht. Dieser gewährleistet außerdem ein frei wählbares Injektionsvolumens, üblicherweise im Bereich von 1-100 μL ¹⁹⁵. Der Aufbau ist jedoch ansonsten identisch zu einem manuellen Injektionsventil, welches meistens als 6-Wege-Ventil ausgelegt ist. Dadurch wird die Probe möglichst ohne längere, starke Druckschwankungen auf die Trennsäule aufgebracht^{190,195}.

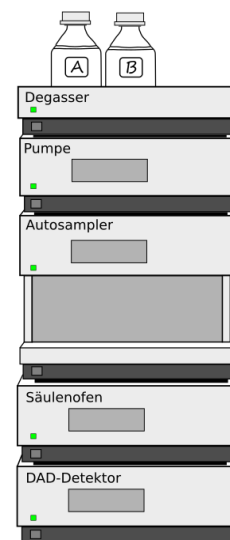


Abb. 4.6: Schematischer Aufbau einer HPLC-Anlage. Abbildung modifiziert¹⁹⁴.

Vor der eigentlichen Trennsäule (Hauptsäule) befindet sich oftmals noch eine Vorsäule, beide befinden sich im temperierten Säulenofen. Die Vorsäule besteht dabei aus demselben oder einem ähnlichen Chromatographiematerial, wie die Hauptsäule. Sie ist jedoch nur einige mm lang¹⁹⁵ und dient dem Schutz der Hauptsäule vor Verschmutzung¹⁹⁰. Auf der Hauptsäule erfolgt die eigentliche chromatographische Auftrennung der Probe. Sie hat üblicherweise eine Länge von 3-25 cm. Ihr Durchmesser kann 1-10 mm betragen, gebräuchliche Innendurchmesser sind insbesondere 2.1 mm, 3.0 mm oder 4.6 mm¹⁹⁵. Der Partikeldurchmesser liegt für analytische HPLC-Säulen üblicherweise bei 10-2 μm . Für UHPLC-Anlagen wurden Säulenmaterialien mit Partikelgrößen < 2 μm entwickelt^{189,190,192}.

Nach dem Verlassen der Hauptsäule werden die eluierenden Substanzen mittels eines Detektors erfasst. Es gibt viele verschiedene Bauarten von Detektoren, üblich sind vor allem UV-Vis/Dioden-

Array-Detektoren, Brechungsindex-Detektoren, Fluoreszenz-Detektoren sowie massenspektrometrische Detektoren (siehe **Abschn. 4.3.3.4**)^{190,192}.

4.3.2.1 Umkehrphasen Säulenmaterialien

Reversed-Phase (RP) Säulenmaterialien werden besonders häufig eingesetzt, da sie es erlauben, mithilfe von wässrigen Eluenten unpolare Analyten voneinander zu trennen. Die Durchführung der chromatographischen Trennung mittels wässriger Laufmittel ebnete den Weg zur Kopplung der LC-Analytik mit API- (Atmosphärendruck-Ionisierung) bzw. ESI-Massenspektrometern^{192,196,197}.

Die stationäre Phase besteht aus Silicagel, deren Oberfläche mit organischen Molekülen modifiziert wurde. Besonders häufig kommen C₁₈- oder C₈-Alkylketten zum Einsatz^{190,198}. Da die langen Alkylketten sterisch anspruchsvoll sind, können die Si-OH-Gruppen auf der Partikeloberfläche nicht zu 100 % mit Alkylketten versehen werden. Verbliebene Si-OH-Gruppen werden daher mittels des sterisch anspruchsloseren Trimethylchlorsilan behandelt („*end-capping*“)^{190,198}. So modifizierte Silicagele können als stationäre Phase für die Trennung von unpolaren bis mittelpolaren Gemischen genutzt werden. Als Elutionsmittel dienen Mischungen aus Wasser mit polaren organischen Lösungsmitteln (Acetonitril, Methanol oder Isopropanol). Der pH-Wert der mobilen Phase kann dabei durch Laufmittelzusätze, wie Säuren oder Puffersalzen eingestellt werden. RP-Säulenmaterial ist jedoch nur in einem pH-Bereich von zwei bis sieben oder acht stabil^{189,192,199}. Bestimmte RP-C₁₈ stationäre Phasen sind jedoch auch noch bei einem höheren pH-Wert von 9-10 nutzbar¹⁹⁹.

Der vorherrschende Trennmechanismus ist dabei das Prinzip der Verteilungschromatographie, dem Verteilungsgleichgewicht der Analyten zwischen mobiler und stationärer Phase. Daneben spielen Adsorptionseffekte der Analyten an der stationären Phase eine Rolle^{190,200}. Die Elution kann entweder isokratisch oder mittels eines Gradienten erfolgen. Die Gradientenelution ermöglicht dabei kürzere Trennläufe, bei gleichzeitig schmalere Peaks. Bei der Gradientenelution wird im Falle der RP-Elution das zu trennende Gemisch bei einem hohen Anteil an wässrigem Laufmittel und niedrigem organischen Anteil auf die Säule aufgegeben. Die Elution erfolgt mit im zeitlichen Verlauf ansteigendem organischen Anteil im Laufmittelgemisch. Daher werden sukzessive die Analyten von der Säule eluiert und dabei aufgetrennt¹⁹².

4.3.2.2 Silica Hybrid Phase Säulenmaterialien

Ein weiterer Trennmechanismus ist die wässrige Normalphasen-Chromatographie (ANP-Chromatographie, engl. *aqueous normal phase*). Diese Trennmethode wurde 1989 von Pesek entwickelt^{201,202}. Im Gegensatz zu der RP-Chromatographie, bei der mittelpolare bis unpolare Substanzgemische aufgetrennt werden, können mittels einer *Silica Hybrid* Phase polare Substanzgemische getrennt werden. Dabei bleibt eine Retention von unpolaren Analyten zum Teil erhalten. Sie stellt somit eine Alternative zur Normalphasenchromatographie (NP-Chromatographie) bzw. der hydrophilen Interaktionschromatographie (HILIC) dar und vereint Teile beider Prinzipien^{202–204}.

Bei der NP-Chromatographie wird Kieselgel mit Si-OH Gruppen auf der Partikeloberfläche als stationäre Phase verwendet¹⁹⁰. Als mobile Phase werden Mischungen aus unpolaren, organischen Lösungsmitteln wie Hexan, Pentan u. a., verwendet. Diese Eluenten sind jedoch nicht mit einer ESI-

Ionenquelle direkt kompatibel¹⁹⁶, sodass eine *online* Kopplung mit einem ESI-Massenspektrometer nicht ohne weitere Anpassungen möglich ist.

Hingegen stützt sich die ANP-Chromatographie auf eine stationäre Phase, die aus oberflächenmodifiziertem Kieselgel besteht. Dabei wurden die Silanolendgruppen modifiziert, sodass sich hauptsächlich Siliciumhydridgruppen (Si-H) auf der Oberfläche der Partikel befinden. Neben den Hydridgruppen wurden zudem zusätzlich ca. 2 % der Si-H-Gruppen mit langkettigen Alkylresten (C₈, C₁₈) modifiziert, ähnlich zu der RP-Chromatographie^{204–206}. Aufgrund der so veränderten Eigenschaften der Säulenpartikel, ist es möglich sowohl polare Substanzen aufgrund der Wechselwirkungen mit den Siliciumhydridgruppen, als auch mittelpolare Substanzen über Wechselwirkungen mit den Alkylgruppen, aufzutrennen. Dabei ist der Einsatz von wässrigen Elutionsmitteln analog zur RP-Chromatographie möglich, was den Einsatz eines ESI-Massenspektrometers als Detektor ermöglicht^{204,207–209}.

4.3.3 Massenspektrometrie

In einem Massenspektrometer wird der Analyt zunächst in der Ionenquelle ionisiert und dabei in die Gasphase überführt. Es gibt verschiedene Ionenquellen, die je nach Art des Analyten zum Einsatz kommen: Bei der GC-MS wird üblicherweise eine EI-Quelle zur Ionisation genutzt²¹⁰. Bei der Kopplung der LC mit einem MS kommt hingegen oft die Elektrospray-Ionisation (ESI) zum Einsatz. Bei der ESI-Ionisation werden die Analyten durch eine elektrisch geladene Kapillare zerstäubt, siehe **Abschn. 4.3.3.1**. Ein großer Vorteil dieser Methode ist, dass auch sehr große Analyten, wie bspw. Proteine schonend ionisiert werden können^{197,211}. Neben den genannten Ionisationsmethoden existieren noch weitere Ionisationsmethoden, wie MALDI (Matrix-unterstützte Laser-Desorption-Ionisierung), DESI (Desorptions-Elektrospray-Ionisierung), DART (engl. *direct analysis in real time*) oder APCI (engl. *atmospheric pressure chemical ionization*), bzw. APPI (engl. *atmospheric pressure photoionization*)^{197,212–215}.

Die geladenen Ionen werden über eine Ionenoptik fokussiert und weitergeleitet. Neben Linsen kommen dabei oftmals Multipole, welche analog zu Quadrupolen funktionieren, zum Einsatz bspw. Hexapole mit sechs Stäben oder Oktopole mit acht Stäben. Die Nutzung von vielen Stäben erweist sich als vorteilhaft bei Ionenoptiken, da dadurch die Fokussierung des Ionenstrahls besser gelingt¹⁹⁷.

Falls das Massenspektrometer technisch zur Fragmentierung der Ionen in der Lage ist, so erfolgt diese in einer Kollisionszelle. Diese besteht üblicherweise aus einem Quadrupol, der mit einem Kollisionsgas (meist Stickstoff, Helium oder Argon) unter niedrigem Druck gefüllt ist. Aufgrund des geringen Druckes des Kollisionsgases hat jedes Ion beim Durchlaufen eine gewisse Wahrscheinlichkeit mit einem Molekül des Kollisionsgases zu kollidieren und dabei in mehrere kleinere Fragmente zu zerfallen¹⁹⁷.

Schließlich ist die Art des Massenanalysators in einem Massenspektrometer von wichtiger Bedeutung, v. a. aufgrund der unterschiedlichen Genauigkeit der Bestimmung des *m/z*-Verhältnisses, siehe **Tab. 4.2**. Es gibt viele unterschiedliche Massenanalysatoren: Sektorfeld, Ionenfalle, Quadrupol, ToF, Orbitrap sowie das FT-ICR-MS (engl. *fourier transform ion cyclotron resonance mass spectrometer*)^{164,197}.

Tab. 4.2: Vergleich verschiedener Masseanalysatoren^{164,197,216}.

Analysator	Auflösung [FWHM]	Massengenauigkeit [ppm]	Massenbereich	Scanrate
Quadrupol	1000-5000	5-500	bis 3000 Da	bis 5000 Da/s
Ionenfalle	5000-10000	50-200	bis 4000 Da	bis 33000 Da/s
(q-)ToF	10000-50000	2-5	bis 40000 <i>m/z</i>	bis 100 Hz
Orbitrap	50000-200000	0.5-1	bis 3000 Da	bis 12 Hz
FT-ICR-MS	1-2 Mio.	0.1-1	bis 10000 <i>m/z</i>	bis 1 Hz

4.3.3.1 Elektrospray-Ionisation (ESI)

Die Elektrosprayionisation (ESI, engl. *electrospray ionisation*) ist eine der wichtigsten Ionisationstechniken, welche bei Atmosphärendruck stattfinden können. Sie wurde 1968 von Dole entwickelt und 1984 von Fenn eingeführt^{217,218}, der dafür 2002 den Nobelpreis erhielt^{197,212}. Das ESI-Prinzip beruht auf der Zerstäubung von Analyten und Lösungsmittel mittels einer elektrisch geladenen Kapillare unter Atmosphärendruck. Sie ist von besonderer Relevanz in der (bio)chemischen Analytik, da sie insbesondere auch die Ionisation von sehr großen Analyten, wie bspw. intakten Proteinen, ermöglicht.

Die Ionenentstehung läuft dabei wie folgt ab (siehe **Abb. 4.7**)^{197,219}: Zunächst wird die Mischung aus Eluent und Analyten durch eine Kapillare aus Stahl, die ESI-Nadel, gedrückt. An der Kapillare liegt eine Spannung von 3-6 kV an. Die entsprechende Gegenelektrode befindet sich in einigen Zentimetern Abstand als zylindrische Ringelektrode. Aufgrund der anliegenden Spannung an der Nadel werden Ladungsträger auf die Flüssigkeit übertragen, die unter Atmosphärendruck zerstäubt wird. Die gebildeten Tröpfchen aus Eluent und Analyten werden in Richtung der Gegenelektrode beschleunigt. Je nach gewähltem Ionenmodus dient die ESI-Nadel als Anode (negativer Ionenmodus) oder Kathode (positiver Ionenmodus). Beim Versprühen der Eluenten mitsamt der Analyten bildet sich ein sogenannter TAYLOR-Konus aus, da sich die Form der aus der Stahlkapillare austretenden Flüssigkeit aufgrund des elektrischen Feldes konisch verjüngt. Anschließend wird die Lösung in viele feine Tröpfchen mit einem Durchmesser von ca. 1.5 µm, zerstäubt. Die Erzeugung von freien Ionen aus dem Aerosol wird durch die Nutzung des *Dry-Gas* (Stickstoff, Temperatur bis zu 600 °C) sowie des *Nebulizer-Gas* (Stickstoff) unterstützt, wobei das Lösungsmittel/der Eluent verdampft und sich die Tröpfchengröße immer weiter verringert. Unterschreitet die Tröpfchengröße einen bestimmten Radius (RAYLEIGH-Limit), so kommt es aufgrund der zunehmenden Ladungsdichte in den Flüssigkeitströpfchen zu einer Abstoßung der Ionen darin und damit zu einer weiteren Zerstäubung der Tröpfchen, die COULOMB-Explosion genannt wird^{197,219}.

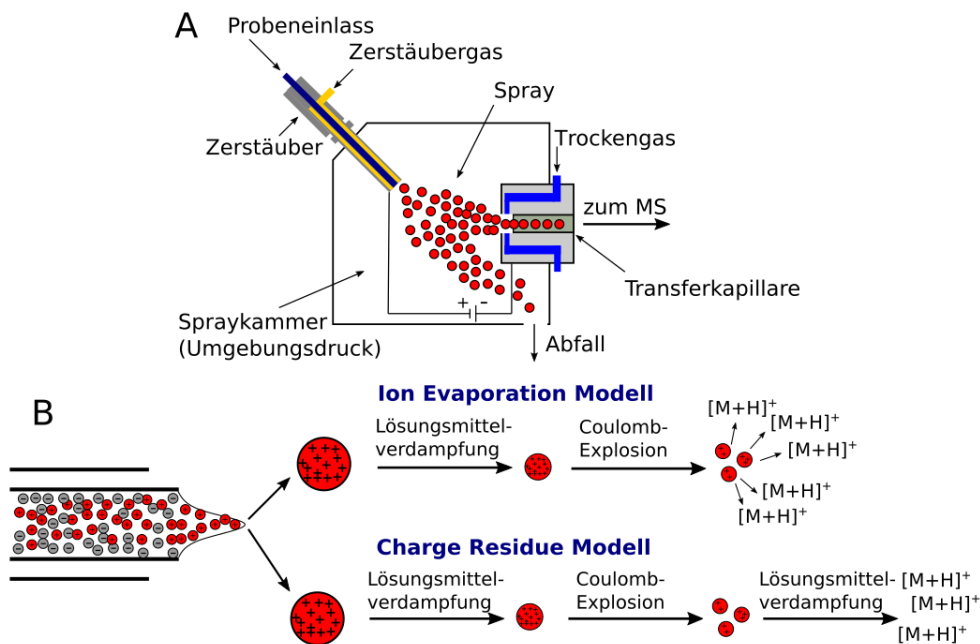


Abb. 4.7: Schematische Darstellung einer ESI-Ionenquelle und des Ionisationsprozesses. **A:** Aufbau einer ESI-Ionenquelle. **B:** Darstellung des Ionisationsprozesses. Abbildung nach Dass²²⁰, Wang²²¹ bzw. Gross¹⁹⁷ modifiziert.

Der Übergang der Analytionen aus einem elektrisch geladenen Flüssigkeitstropfen in die Gasphase als freie Ionen ist noch nicht vollständig geklärt. Bisher existieren zwei unterschiedliche Theorien, wobei es sich möglicherweise um konkurrierende Prozesse handelt: Zum einen existiert das *Ion Evaporation* Modell, welches die These aufstellt, dass Ionen direkt aus einem Flüssigkeitstropfen, also aus dem solvatisierten Zustand in die Gasphase (dem desolvatisierten Zustand) übergehen können^{197,219,222}. Dahingegen erklärt das *Charge-Residue*-Modell den Übergang der Ionen in die Gasphase als Resultat eines Prozesses der wiederholten Schrumpfung der Tröpfchen durch COULOMB-Explosion bzw. Verdunstung des Lösungsmittels, an dessen Ende nach vollständiger Verdampfung des Lösungsmittels, die desolvatisierten Ionen zurückbleiben^{197,219}.

Der stattfindende Ionisationsmechanismus ist vorwiegend die Protonierung $[M+H]^+$, dabei sind Mehrfachladungen $[M+nH]^{n+}$ mit $n = 2, 3, \dots$ etc. üblich. Die gebildeten Ionen werden auch „Pseudo-Molekülionen“ genannt. Oft können neben dem $[M+H]^+$ -Peak zusätzliche Adduktionen wie $[M+Na]^+$ oder $[M+K]^+$ beobachtet werden. Daneben können auch Addukte mit Laufmittelmolekülen wie $[M+Acetonitril+H]^+$ oder mit Laufmitteladditiven ($[M+Acetat]^-$, $[M+Formiat]^-$) möglich. Zusätzlich kann es zum Verlust von Neutralteilchen wie H_2O , CO_2 , oder NH_3 kommen^{197,212,223–225}.

4.3.3.2 Quadrupol

Nach der Bildung von Ionen in der Ionenquelle und dem Übergang in die Gasphase und Transport durch die Transferkapillare/Ionenoptik ins Hochvakuum, passieren die Ionen oftmals Quadrupole (q , Massenselektoren) bzw. Multipole. Während Quadrupole vier Stabelektroden besitzen und als Massenselektoren fungieren, dienen Hexapole mit $n = 6$ und Oktapole ($n = 8$) als Ionenoptik oder Kollisionszellen¹⁹⁷.

Der Aufbau eines Quadrupol ist dabei wie folgt: Er besteht aus vier parallel angeordneten Stabelektroden, die von oben betrachtet ein Quadrat bilden. Die jeweils gegenüberliegenden Elektroden sind an denselben Gleichspannungspol angeschlossen. Zusätzlich zu der Gleichspannung wird an die Elektroden eine Wechselspannung angelegt¹⁹⁷. Dieses Zusammenspiel ergibt ausschließlich für Ionen mit einem bestimmten Masse-zu-Ladungs-Verhältnis (m/z -Verhältnis) eine stabile Trajektorie durch den Quadrupol. Diese wird durch die Wechselspannung und dem Verhältnis aus Gleichspannung zu Wechselspannung bestimmt. Indem der Gleichspannungsanteil relativ zum Wechselstromanteil reduziert wird, kann ein zunehmend bereiteres m/z -Fenster den Quadrupol passieren. Indem die absolute Größe von Gleich- und Wechselstrom bei identischem Verhältnis zueinander variiert wird, können die m/z -Verhältnisse der Ionen, welche den Quadrupol passieren können, abgescannt werden¹⁹⁷.

Da auf die Ionen mit zunehmender Entfernung von der Mittelachse des Quadrupols ein ansteigendes Potential wirkt, werden Ionen mit bestimmten m/z -Verhältnissen wieder zurück auf die Mittelachse fokussiert. Ionen mit anderen m/z -Verhältnissen erreichen hingegen keine stabile Trajektorie und kollidieren mit einem der Stäbe des Quadrupols und werden dadurch entladen.

Ein Quadrupol oder Multipol kann auch als Kollisionszelle dienen, dafür wird er bei einem geringen Druck von ca. 1 Pascal mit einem inerten Gas gefüllt, meist Stickstoff, Argon oder Helium^{192,197}. Auf dem Weg durch die Kollisionszelle haben die Ionen eine bestimmte Wahrscheinlichkeit, mit einem der Teilchen des Kollisionsgases zu kollidieren. Bei der Kollision kommt es teilweise zur Fragmentierung der Ionen, die dann in einem weiteren Quadrupol selektiert werden können, falls es sich um geladene Teilchen handelt. Entstehen ungeladene Fragmente der *Precursoren* können diese in ihrer Flugbahn nicht beeinflusst werden, somit unterbleibt eine Fokussierung auf die Mittelachse des Quadrupols. Ungeladene Teilchen können zudem im weiteren Verlauf auch nicht detektiert werden¹⁹⁷.

4.3.3.3 Time-of-Flight-Massenanalysator

Der *Time-of-Flight*-Massenanalysator (ToF-MS) wurde bereits 1946 von Stephens erfunden¹⁹⁷. Er beruht auf dem Prinzip der Messung der Flugzeit der Ionen in einem feldfreien Flugrohr. Je nach m/z -Verhältnis brauchen die Ionen unterschiedlich lange für die Bewältigung einer feldfreien Strecke. Es gilt dabei der in **Gl. 4.4** gezeigte Zusammenhang. Die Beschleunigung aus der Ruheposition erfolgt nur durch die Beschleunigungsspannung U . Die Flugzeit t der Ionen ist somit proportional zu der Quadratwurzel aus dem Masse-zu-Ladungs-Verhältnis m/z ¹⁹⁷.

$$E_{kin} = \frac{1}{2} \cdot m \cdot v^2 = U \cdot e \cdot z \quad (4.4)$$

$$t = \frac{s}{v} = \frac{s}{\sqrt{2e \cdot U}} \sqrt{\frac{m}{z}}$$

Aufgrund dieses Prinzips kommen leichtere Ionen früher am Detektor an, als schwerere Ionen. Voraussetzung dabei ist, dass die Ionen alle zur selben Zeit beschleunigt werden. Dies wird mittels eines *Ion Pulsers* sichergestellt, welcher den kontinuierlichen Ionenstrom in einzelne, diskontinuierliche Ionenpakete aufteilt. Diese werden dann meistens orthogonal zur ursprünglichen Flugrichtung in das ToF-Rohr hinein beschleunigt^{192,197,226}. Somit erfolgt die Detektion von Ionen mit unterschiedlichen m/z -Verhältnissen zu verschiedenen Zeitpunkten. Da

die Flugzeit der Ionen allgemein im Bereich von nur einigen Nanosekunden liegt, können mit einem ToF-Massenanalysator prinzipiell auch sehr große Ionen detektiert werden^{197,216}.

Um die Messgenauigkeit der Flugzeit und somit die Auflösung zu erhöhen, besitzen ToF-Analysatoren oftmals am anderen Ende des Flugrohres einen Reflektor, der die Ionen zurückreflektiert. Dieser wurde 1994 von Mamyurin entwickelt^{197,227}. Er besteht aus mehreren Ringelektroden, deren Potential auf das ca. 1.05 bis 1.1-fache der Beschleunigungsspannung U eingestellt wird. Die Ionen dringen je nach ihrer kinetischen Energie unterschiedlich tief in das Reflektoren ein. Somit verweilen gleich große Ionen mit weniger kinetischer Energie dort kürzer als solche mit mehr kinetischer Energie. Als Resultat kommt es zu einer Refokussierung der Ionen, welche eine Erhöhung der Auflösung auf 1-2 ppm zur Folge hat^{192,197,216}.

Die Vorteile der ToF-Technologie sind die hohe Auflösung sowie der theoretisch nach oben hin offene Messbereich. Ein Nachteil ist die vergleichsweise häufig notwendige Kalibration des Massenspektrometers, aufgrund dessen Sensitivität gegenüber Temperaturschwankungen. Genauer als ToF-Analysatoren arbeiten FT-ICR-Massenspektrometer oder Orbitraps^{164,216,226}. Das FT-ICR-MS ist jedoch aufgrund der besonderen Detektorart, auf MS-Ebene langsamer. Deren Scanrate erreicht nicht die nötige zeitliche Auflösung zur Kopplung mit einer LC²²⁸.

4.3.3.4 LC-ESI-q-ToF-Massenspektrometrie

Bei dem q-ToF-Massenspektrometer handelt es sich um ein Hybridsystem bestehend aus einem Quadrupol, einer Kollisionszelle und einem dahinterliegenden ToF-Massenanalysator^{192,197,216}. Diese Bauweise bietet die Möglichkeit Ionen mittels des Quadrupols und einer Kollisionszelle zu fragmentieren. Die so erzeugten Fragmentionen können dann mit dem ToF-Massenanalysator ebenfalls mit hoher Massengenauigkeit detektiert werden, wie die unfragmentierten Precursorionen. Die Verwendung einer ESI-Quelle ermöglicht zudem die sanfte Ionisation einer Vielzahl von Analyten. Weitere Vorteile sind zudem die Eignung zur *online*-Kopplung mit chromatographischen Trennverfahren, wie der LC. Diese Eigenschaften machen LC-q-ToF-Massenspektrometer zu den bevorzugten Geräten, mit denen *non-targeted Metabolomics*-Studien durchgeführt werden^{6,127,192,216}. Sie besitzen gleichzeitig eine hohe Auflösung und sind in der Lage eine hohe Rate an Spektren pro Zeiteinheit aufzuzeichnen. Somit eignen sie sich hervorragend zur Kopplung mit einem Chromatographieverfahren^{192,216}.

Häufig kann bereits über die hohe Auflösung auf Ebene der unfragmentierten Analyten (MS¹-Ebene) mittels der exakten, monoisotopischen Masse und/oder unter Zuhilfenahme des Isotopenmusters ein Strukturvorschlag erstellt werden. Durch die zusätzliche Aufnahme von Fragmentspektren können weitere, hochaufgelöste Informationen über bestimmte Analyten erhalten werden.

Die Kopplung eines ESI-q-ToF-MS mit einer (U)HPLC-Anlage erfolgt, indem nach Passieren der Trennsäule das Eluat mittels einer Kapillare der ESI-Quelle zugeführt und dort ionisiert wird. Die geladenen Ionen werden dann durch das Vakuum im Massenspektrometer durch die beheizte Transferkapillare geleitet (siehe **Abb. 4.8**). Ungeladene Teilchen werden durch verschiedene Maßnahmen am Eintritt in das Massenspektrometer gehindert. Auf diese Weise wird eine Verschmutzung des Massenspektrometers bestmöglich verhindert. Im Inneren des MS durchlaufen die Ionen dann nacheinander zwei *Ion Funnels* und einen Multipol als Ionenoptik, den

Quadrupol, eine Kollisionszelle, einen *Ion Cooler* und schließlich das ToF-Rohr und den Detektor^{192,197,226,229}.

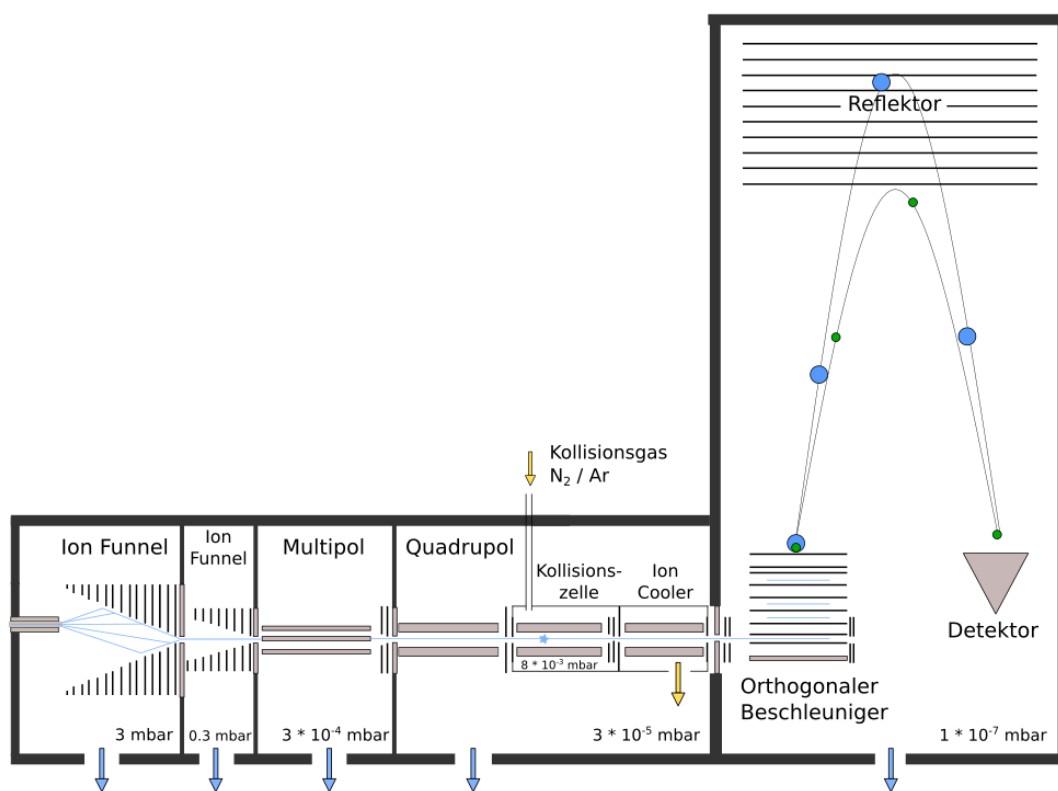


Abb. 4.8: Schematische Darstellung eines q-ToF-Massenspektrometers mit einer ESI-Ionenquelle. Abbildung modifiziert²²⁹.

Im MS^1 -Modus wird der Quadrupol lediglich als Ionenoptik genutzt, in der Kollisionszelle findet keine Fragmentierung statt. Alle Ionen erreichen somit das ToF-Rohr und in der Folge den Detektor^{192,197,226,229}. Im MS/MS -Modus (auch MS^2 -Modus genannt) wird hingegen der Quadrupol dazu genutzt, lediglich ein bestimmtes m/z -Verhältnis hindurchzulassen. Alle Ionen mit einem anderen m/z -Verhältnis können den Quadrupol nicht passieren. Die hindurchgelassenen Ionen werden dann in der Kollisionszelle fragmentiert und die Fragmentionen werden mittels des ToF-Massenanalysators, hochaufgelöst detektiert¹⁹².

Die hohe zeitliche Auflösung des ToF-Massenanalysators ermöglicht dabei eine Akquise von MS/MS -Spektren von dynamisch ausgewählten Precursorionen. Zusätzlich können sämtliche MS^2 -Fragmentationen hochaufgelöst detektiert werden¹⁹². Dies stellt ein wichtiges Werkzeug zur Methodenentwicklung für die Quantifizierung von Markersubstanzen dar, da es die Bestimmung sinnvoller Masseübergänge für den Transfer einer Methode auf ein Triple-Quadrupol-Massenspektrometer erleichtert. Die Beschränkung der pro Zeiteinheit erfassbaren Precursorionen ist jedoch auch bei der Verwendung von q-ToF-Hybrid-Massenspektrometern gegeben^{197,216,229}.

Zur Quantifizierung werden üblicherweise zwei Fragmentionen je Massenübergang detektiert: ein möglicherweise unspezifisches und intensives Fragmention (*Quantifier*) zur Quantifizierung. Sowie ein spezifisches, dafür aber möglicherweise weniger intensives Fragmention (*Qualifier*). Das Prinzip Selektion eines bestimmten *Precursors* in dem ersten Quadrupol, die Fragmentierung dieses Ions

und die Detektion eines bestimmten Fragmentions wird häufig für die Quantifizierung von Analyten eingesetzt. Es firmiert unter der Bezeichnung *Selected Reaction Monitoring* (SRM). Oftmals werden mehrere dieser *Precursor-Fragment*-Übergänge eines einzelnen Precursorions in einer Analysemethode parallel detektiert. Dies wird als *Multiple Reaction Monitoring* (MRM) bezeichnet. Somit wird insgesamt eine hohe Selektivität erreicht^{212,230}.

4.4 Datenanalyse

4.4.1 Univariate Datenanalyse

Die univariate Datenanalyse beschreibt die Untersuchung von Analysedaten mittels einfacher, statistischer Mittel bzw. Tests, wie bspw. dem t-Test. Dabei ist der sogenannte *p*-Wert zentral. Er gibt die Wahrscheinlichkeit dafür an, dass die Nullhypothese in Bezug auf eine Fragestellung, welche mittels der Messergebnisse beantwortet werden soll, zurückgewiesen wird. Die Nullhypothese lautet im Falle eines t-Tests: Die Mittelwerte der Messwerte aus den beiden Gruppen sind identisch. Das Signifikanzniveau gibt an, mit welcher Unsicherheit diese Aussage behaftet ist²³¹. Es beträgt in der Regel 95 % ($p = 0.05$) oder 99 % ($p = 0.01$).

Um die Signifikanz von Unterschieden zwischen mehr als zwei Probengruppen zu bestimmen, kann die einfaktorielle Varianzanalyse (engl. *one-way analysis of variance*; *One-Way ANOVA*-Test) angewandt werden. Dieser Test gibt die Wahrscheinlichkeit dafür an, dass die Null-Hypothese „*alle Mittelwerte der Probengruppen sind identisch*“ zurückgewiesen wird²³². Der ANOVA-Test lässt jedoch lediglich eine allgemeine Aussage darüber zu, ob alle Mittelwerte identisch sind. Um die Ursache für die Unterschiede zu ergründen, sind weitere Untersuchungen nötig. Mittels des Post-Hoc TUKEY-Tests können nach einem ANOVA-Test diejenigen Kombinationen aus zwei Probenklassen identifiziert werden, welche signifikant unterschiedlich sind²³².

4.4.2 Verfahren der multivariaten Datenanalyse

Die multivariate Datenanalyse dient der Reduktion der Komplexität von multidimensionalen Datensätzen, wobei möglichst für die Fragestellung nicht-relevante Informationen entfernt werden. Beispiele für eine multivariate Datenanalyse sind bspw. die multivariate Regressionsanalyse, Klassifikationsalgorithmen oder die Hauptkomponentenanalyse^{166,233–237}.

Mittels multivariater Datenanalyse wird die Informationsdichte in den Daten durch die Extraktion von relevanten Subdaten erzielt. Dabei werden die Daten oft von einem höherdimensionalen Raum in einen niederdimensionalen Raum transformiert²³⁸. Die Anzahl der Dimensionen entspricht dabei bspw. der Anzahl an analysierten Metabolite, Wellenlängen, oder anderen Messdaten.

Generell sind multivariate Klassifikationsverfahren danach zu unterscheiden, ob es sich um „*unsupervised*“ Verfahren, oder „*supervised*“ Verfahren²³⁹. Bei „*unsupervised*“ Verfahren ist die Probenstruktur im Sinne von Klassen, wie „Herkunft“, „Sorte“, „Anbauform“ oder „Alter“, nicht bekannt. D. h. es erfolgt neben der Verwendung der Messdaten keine Berücksichtigung von weiteren (Meta-)Daten. Bei „*supervised*“ Techniken sind diese zusätzlichen Metadaten hingegen bekannt und fließen bei der Datenbehandlung mit ein. Da dies ein *Bias* für eine optimistische Vorhersage darstellen kann, wurden für den Bereich der algorithmenbasierten Klassifikation, Validierungsstrategien entwickelt. Diese ermöglichen die Überprüfung der Modelle mittels nicht

zuvor zur Modellerstellung genutzten Messdaten, somit wird ein *Overfitting* möglichst vermieden^{240,241}.

Vor der Visualisierung bzw. Klassifikation der Messdaten findet oft eine Vorverarbeitung der Rohdaten statt²⁴². Dies ist insbesondere bei LC-MS-Analysen notwendig, da der Wertebereich im Bereich von 0 bis 10^6 schwanken kann. Ohne vorherige Anpassung würden dabei Messwerte mit höheren absoluten Werten dominant gegenüber Messwerten mit absolut geringeren Zahlenwerten auftreten²⁴². Dies könnte zur Folge haben, dass zahlenmäßig kleinere, dafür aber aussagekräftigere Messwerte zugunsten von zahlenmäßig großen, aber im Rahmen der Fragestellung nicht aussagekräftigen Messwerten weniger Einfluss in einem Modell bekommen. Somit würde die Aussagekraft schwinden. Um dies zu vermeiden, wird der Wertebereich oftmals angeglichen, wobei viele unterschiedliche Strategien/Verfahren existieren.

4.4.3 Datenvorverarbeitung

Der Prozess der Datenvorverarbeitung bei *Metabolomics*-Fragestellungen, insbesondere im *Food/Feed* Bereich ist elementar, da aufgrund der Art der Fragestellungen, die Ergebnisse auf z. T. nur geringen Konzentrationsänderungen zwischen unterschiedlichen Probengruppen beruhen. Gleichzeitig liegen oftmals nur vergleichsweise wenige Proben vor.

Die Datenvorverarbeitung unterteilt sich dabei in zwei unterschiedliche Bereiche: a) Vorkehrungen die direkt an den Messgeräten vorgenommen werden, um eine hohe Datenqualität zu sichern. Dies kann bspw. über eine Kalibration des Massenspektrometers oder die Hintergrundkorrektur des FT-NIR-Spektrometers erfolgen sowie b) Vorkehrungen bzw. Verfahren, die angewandt werden, um aus den Rohdaten möglichst Informationen von hoher Güte zu extrahieren. Aufgrund der Natur der Fragestellung bei *non-targeted Food/Feed Metabolomics*-Experimenten, die eine nicht-zielgerichtete Analyse des Metaboloms erfordern, ist eine manuelle Detektion aller Features in allen Proben jedoch nicht durchführbar.

Ein Feature im Bereich LC-MS ist dabei definiert als ein Signal bei einer bestimmten Retentionszeit, mit einem bestimmten m/z -Verhältnis und einer bestimmten Intensität/Peakfläche. Für eine systematische Detektion eines Features in vielen Proben, ist daher die automatische Zuordnung/Erkennung des Features in den Messungen entscheidend. Die computergestützte Peak-Erkennung würde jedoch bei einer zu stringenten Zuordnungsweise zu fehlerhaften Ergebnissen führen. Daher wird bei der Auswertung der LC-MS-Analysen zunächst neben einer Kalibration der m/z -Achse ein *retention time alignment* durchgeführt, um auch Abweichungen auf der Chromatographieebene zu begrenzen^{226,243,244}. Bei dem sich anschließenden systematischen *Feature Finding* werden schließlich in mehreren Proben vorkommende Signale detektiert. Dabei wird aufgrund von analytischen Beschränkungen eine gewisse Toleranz in Bezug auf die Retentionszeit (z. B. $\Delta RT < 0.5$ min) und das m/z -Verhältnis ($\Delta m/z < 20$ ppm) gewährt. Das Auffinden von Peaks im RT- bzw. m/z -Koordinatensystem ist daher eine zweidimensionale Herausforderung, wobei die Toleranz in x-Richtung (RT-Achse) weitaus größer ist, als jene in y-Richtung (m/z -Achse).

Darüber hinaus wird im Falle der LC-MS-Analytik bei einem Peak weiterhin unterscheiden, ob es sich um den monoisotopischen Peak $[M]^+$, oder um einen Isotopenpeak, bei $[M+1]^+$, $[M+2]^+$, oder $[M+3]^+$ etc., handelt^{212,245,246}. Daneben können durch ein und dieselbe Substanz mehrere Signale

im Massenspektrum hervorgerufen werden. Diese rühren von unterschiedlichen Addukten, wie $[M+H]^+$, $[M+Na]^+$, $[M+K]^+$, oder $[M+NH_4]^+$, her. Weiterhin können Neutralverluste auftreten, z. B. $[M-H_2O+H]^+$ oder $[M-CO_2+H]^+$ ^{212,224}. Zusammenfassend kann eine Verbindung mehrere Signale zur Folge haben. Da die Intensitäten dieser Peaks alle miteinander korrelieren, wird zum systematischen Vergleich ein monoisotopischer Peak eines einzelnen, intensiven, aber gut reproduzierbaren Adduktes als ein „Feature“ definiert, anhand dessen die Konzentration des Analyten semiquantitativ in allen Proben bestimmt wird²⁴⁷. Übliche Addukte dafür sind z. B. $[M+H]^+$ oder $[M+NH_4]^+$, im Falle der Addition von Ammoniumsalzen als Puffer. Ein Feature ist also genaugenommen ein Triple von RT, m/z -Verhältnis und Intensität bzw. Peakfläche. Während RT und m/z -Verhältnis in gewissen engen Toleranzen in allen Analysen konstant bleiben, variiert die Peakhöhe bzw. Peakfläche in Abhängigkeit von der Metabolitkonzentration²⁴⁴. Der monoisotopische Peak kann aufgrund der meist höheren Intensität in Relation zu den Isotopenpeaks identifiziert werden; Addukte aufgrund deren zeitgleichen Elution zu dem Hauptpeak^{245–247}.

Im Anwendungsfeld der FT-NIR-Spektroskopie resultieren hingegen andere Effekte, deren Abmilderung daher auch durch andere Verfahren erfolgt: Ein typisches Merkmal von pulverförmigen Proben ist, dass die Spektren einen Drift bzw. eine erhöhte Basislinie zu kleineren Wellenzahlen hin, aufweisen⁴⁷. Insbesondere bei pulverförmigen Proben kommt es aufgrund der ähnlichen Größenordnung der Partikel und der Wellenlänge der NIR-Strahlung vorwiegend zu LORENTZ-MIE-Streueffekten²⁴⁸. Daher ist bei der Etablierung von FT-NIR-spektroskopischen Methoden auf eine einheitliche Vermahlung der Proben zu achten, da die Partikelgrößenverteilung einen Einfluss auf die aufgezeichneten FT-NIR-Spektren ausüben könnte^{177,248,249}.

Die Korrektur dieser Streueffekte kann rechnerisch erfolgen, insbesondere mittels der *Standard Normal Variate* (SNV) und der *Multiplicative Scatter Correction* (MSC). Das *Preprocessing* mittels SNV wurde 1989 von Barnes eingeführt²⁵⁰. Es verändert das gemessene Spektrum derart, dass es zwar seine äußere Form behält, aber verschiebt es so im Raum, dass ein Offset entfernt wird und der Mittelwert aller Messpunkte 0 beträgt und skaliert es zudem so, dass die Standardabweichung aller Messpunkte gegen 1 tendiert^{248,251}. Das Verfahren der MSC erreicht mit ähnlichen Berechnungen ein analoges Ergebnis, jedoch wird ein Referenzspektrum benötigt^{248,252}. Dazu wird oftmals ein gemitteltes Spektrum über alle Analysen herangezogen^{253–255}. Dies erschwert allerdings insbesondere eine nachträgliche Erweiterung von Klassifikationsmodellen. Der SNV-Algorithmus benötigt hingegen kein Referenzspektrum^{248,250}. Da sich zudem die Resultate des SNV- und der MSC-Algorithmus laut Dhanoa *et al.* mittels eines Faktors ineinander überführen lassen²⁵⁶, ist die *Preprocessing*-Variante mittels SNV für die Erstellung eines Klassifikationsmodells im Vorteil.

Im Allgemeinen werden Messdaten, insbesondere im Falle der LC-MS-Analytik, vor der multivariaten Datenanalyse zunächst einer Datenvorverarbeitung unterzogen, um Metabolitsignale mit geringer Intensität und Metabolitsignale mit höherer Intensität anzugleichen. Dadurch werden jene Metabolitsignale mit geringer Intensität nicht gegenüber denen mit höherer Intensität diskriminiert. Es gibt dabei verschiedene Arten des *Preprocessing*, Beispiele sind *Mean-Centering*, *Normalisation*, *Auto-Scaling*, *Range-Scaling*, *Pareto-Scaling*, *Vast-Scaling* sowie das *Level-Scaling*²⁴².

Beim *Mean-Centering* wird der Mittelwert der Variablen $\bar{x}_i$ aus allen analysierten Proben von dem jeweils ermittelten Messwert x_{ij} abgezogen. Damit wird der Fokus auf die Unterschiede in dem Datensatz gesetzt, der Mittelwert der preprozessierten Daten beträgt 0²⁴². Diese Variante der Skalierung ist auch Teil von anderen *Preprocessing*-Varianten. Oftmals werden die auf den Mittelwert zentrierten Daten durch einen Faktor geteilt. Dieser Faktor ist bspw. beim *Auto-Scaling* die Standardabweichung aller analysierten Proben s_i , d. h. nach dem *Preprocessing* beträgt die Standardabweichung 1. Somit werden alle Metabolite in Bezug auf ihren Einfluss egalisiert, keiner der analysierten Metabolite besitzt ein größeres Gewicht²⁴². Beim *Pareto-Scaling* wird hingegen als Skalierungsfaktor $\sqrt{s_i}$, anstelle von s_i genutzt. Somit wird lediglich der Einfluss von sehr intensiven Metaboliten mit großen Messwerten reduziert. Dafür bleibt die relative Datenstruktur näher an derjenigen der Rohdaten²⁴². Das *Level-Scaling* benutzt anstelle der Standardabweichung, den Mittelwert $\bar{x}_i$ als Normierungsfaktor²⁴². Das *Vast-Scaling* entspricht weitestgehend dem *Auto-Scaling*, mit dem Unterschied das zusätzlich mit dem Faktor $\bar{x}_i/s_i$ multipliziert wird²⁴². Das *Range-Scaling* hingegen normiert die Messwerte, nach Zentrierung auf den Mittelwert, auf die abgedeckte Spanne der Messwerte ($x_{\max} - x_{\min}$). Zudem ist dieses Verfahren vulnerabel für Ausreißer, da die Minima/Maxima der Messwerte direkt in die Berechnung mit einfließen²⁴².

4.4.4 Hauptkomponentenanalyse (PCA)

Die Hauptkomponentenanalyse (engl. *principal component analysis*, PCA) ist ein häufig genutztes Verfahren der multivariaten Datenanalyse^{127,233,234,238,257}. Sie dient der Reduktion der Dimensionalität der Messdaten^{233,238,258}. Die multidimensionalen Messdaten werden in sogenannte Hauptkomponenten (*principal components*, PC) überführt, die eine Linearkombination der Analysedaten darstellen: $PC1 = W_1 * A_1 + W_2 * A_2 + \dots W_n * A_n$. Die resultierenden Hauptkomponenten stehen dabei orthogonal zueinander^{238,258,259}.

Die Projektionsebenen sind dabei so gewählt, dass diese möglichst große Anteile an der Varianz in dem Datensatz abbilden^{235,238}. Die erste Hauptkomponente erklärt den größten Anteil an der Varianz des Datensatzes, die Zweite den zweitgrößten Anteil, usw. Somit wird die Informationsdichte in dem Datensatz auf die ersten Hauptkomponenten konzentriert; die höheren Hauptkomponenten haben immer weniger Bedeutung für die Visualisierung von Unterschieden zwischen den Proben^{235,238,258}. Folglich ist es oftmals möglich, für eine Veranschaulichung der in einem multidimensionalen Datensatz vorhandenen Varianz, lediglich zwei oder drei der Hauptkomponenten darzustellen. Diese Art der Darstellung wird auch Scoresplot genannt^{234,258}.

In diesem Scatterplot ist pro Probe des Datensatzes ein Punkt gezeigt. Der Koordinatenursprung fällt dabei mit dem Schwerpunkt der Daten zusammen. Proben mit gleichen Eigenschaften, wie der geographischen Herkunft, Art oder Sorte etc., werden im Scoresplot oftmals mit gleicher Farbe dargestellt. Dabei handelt es sich jedoch nicht um eine Eigenschaft/Folge der Hauptkomponentenanalyse, denn diese ist ein *unsupervised* Verfahren. Die Einfärbung wird davon unabhängig, nachträglich zur besseren Veranschaulichung vorgenommen^{234,238,260}. Ähnliche Proben bilden im Scoresplot Cluster, deren Streuung von der Variabilität der Messdaten abhängt. Je geringer die Varianz zwischen bestimmten Proben ist, umso dichter sind die Proben im Scoresplot beieinander lokalisiert. Somit bilden Proben derselben Gruppe, welche sich ähnlich verhalten, in der PCA einen Cluster^{234,260}. Die PCA ist ein gängiges Verfahren, um Unterschiede zwischen

Probengruppen, von denen multidimensionale Messwerte mit sehr vielen Variablen existieren, zu visualisieren^{127,233,234,260}.

Die einzelnen Hauptkomponenten werden wie folgt berechnet^{234,258}: Die Datenmatrix X enthält n Proben, eine je Zeile und m Metabolite, eine je Spalte. Sie wird in die Matrizen T , P und E zerlegt, siehe **Gl. 4.5**.

$$X = T \cdot P^T + E \quad (4.5)$$

Die Matrix T ist die sogenannte Scoresmatrix, die zeilenweise die Proben und spaltenweise k Hauptkomponenten enthält. Die Matrix P wird Faktorenmatrix genannt, sie beinhaltet zeilenweise die m Faktoren der Linearkombination (Loadingsvektoren). In den Spalten der Matrix P sind die k Hauptkomponenten enthalten, und sie wird transponierte Matrix genannt. Neben diesen Matrizen entsteht zusätzlich die sogenannte Residuenmatrix E , welche die nicht von den Hauptkomponenten abgedeckte Restvarianz des Datensatzes enthält²⁵⁸.

Der Loadingsplot hingegen stellt eine Visualisierung der sogenannten Loadingsvektoren, also die Faktoren der Linearkombination aus der Matrix P von allen Metaboliten/Messwerten dar²⁵⁸. Je betrachteter Messgröße, also z. B. je Peak in einer LC-MS Analyse oder Wellenzahl bei FT-NIR-Spektroskopie, existiert somit einen entsprechenden Loadingsvektor. Je höher dabei die Varianz einer Variablen ist, desto weiter liegt der Loadingsvektor vom Koordinatenursprung entfernt²⁶⁰.

Dabei gilt außerdem, je größer der Einfluss eines Loadingsvektors auf die Positionierung der Proben im Scoresplot ist, umso weiter ist dieser vom Koordinatenursprung entfernt. Ist deren Einfluss hingegen gering, so liegt der Loadingsvektor nahe dem Koordinatenursprung; zugleich ist der Einfluss dieser Variablen auf die Positionierung der Proben im Scoresplot lediglich gering. Somit ermöglicht der Loadingsplot diejenigen Variablen zu identifizieren, welche hauptsächlich für die Anordnung der Proben im Scoresplot ursächlich sind^{234,258,260}. Ist im Scoresplot eine Auftrennung der Proben in Cluster entsprechend der Probenklassen zu beobachten, so können anhand des entsprechenden Loadingsplots mögliche potente Variablen zur Unterscheidung der Probenklassen identifiziert werden, da diese dort einen großen Abstand zum Koordinatenursprung aufweisen^{258,260}.

4.4.5 Partial Least Square Discriminant Analysis (PLS-DA)

Bei der *Partial Least Square Discriminant Analysis* (PLS-DA) handelt es sich um eine Modifikation der PCA. Während bei der PCA die Auftragung der Messdaten so erfolgt, dass die Varianz im transformierten Datensatz maximiert wird, so wird bei der PLS-DA die Co-Varianz zwischen den Messdaten (X) und den Probenklassen (Y) maximiert^{235,261}. Anders als bei der PCA, ist mittels der PLS-DA auch eine Klassifikation der Proben, durch die Abtrennung der Probenklassen mittels linearen Diskriminationsvektoren, möglich. Dabei werden die Proben im zweidimensionalen Raum mittels Geraden bzw. im dreidimensionalen Raum mithilfe von Ebenen in bestimmte Klassen unterteilt. Daneben gibt es zudem das Verfahren der *Partial Least Square Regression*, welches aufgrund von mehreren Ausgangsvariablen $X_1, X_2, \dots, X_n$ eine kontinuierliche Zielvariable Y vorhersagen kann^{236,262}.

Bei der PLS-DA wird die Klasseninformation zusätzlich mit in die Modellerstellung eingebracht, siehe **Gl. 4.6**. Dabei gilt rechnerisch neben dem Zusammenhang $X = T \cdot P^T + E$, zusätzlich der

Zusammenhang $Y = U \cdot Q^T + F$, mit der die Klasseninformation Y in die Transformation der Proben mit einfließt. Somit handelt es sich um eine *supervised* Methode, im Gegensatz zur Hauptkomponentenanalyse^{235,261,263}.

$$\begin{aligned} X &= T \cdot P^T + E \\ Y &= U \cdot Q^T + F \end{aligned} \quad (4.6)$$

Da die Bewertung des Einflusses einzelner Variablen anhand deren Unterschiedlichkeit zwischen den Probenklassen erfolgt, reichen bereits minimale Varianzen einer Variablen zwischen den Probenklassen, um deren Einfluss zu erhöhen und in Folge eine scheinbar hohe Separierbarkeit zwischen den Probengruppen vorzutäuschen. Das Verfahren der PLS-DA ist somit sehr anfällig für ein *Overfitting*. Dies geht so weit, dass es möglich ist, Messwerte, denen jeweils zufällig eine von zwei Klassen zugeordnet wurde, mittels PLS-DA voneinander zu trennen²⁶⁴. Weitere Proben, die nicht zur Modellerstellung genutzt wurden, werden jedoch völlig zufällig einer der beiden Klassen zugeordnet, weshalb das Modell in diesem Fall nicht aussagekräftig ist. Um dies zu erkennen bzw. zu bewerten, kann der Q^2 -Wert berechnet werden. Dieser gibt die Klassifikationsgenauigkeit des Modells für die Validierungsproben an und wird meistens durch eine *Leave-One-Out*-Klassifizierung, ermittelt^{265,266}. Analog gibt der R^2 -Wert, die Vorhersagegenauigkeit der Trainingsproben an²⁶⁵. Beide Werte können im Bereich zwischen 0 und 1 liegen. Während jedoch der R^2 -Wert oftmals sehr gute Werte erzielt, ist der Q^2 -Wert hingegen tendenziell kleiner. Niedrige Q^2 -Werte sprechen dabei für ein *Overfitting* der PLS-DA^{264,265}.

4.4.6 Erstellung und Bewertung von Klassifizierungs- und Vorhersagemodellen

Klassifizierungs- bzw. Vorhersagemodelle ermöglichen die Zuordnung von Proben zu bestimmten Eigenschaften (den sogenannten Klassen) anhand von Messdaten. Oftmals in Verbindung mit einer Reduktion, der Variablenanzahl durch eine multivariate Datenanalyse, wie z. B. PCA oder PLS. Die eigentliche Klassifikation erfolgt dann auf Grundlage der abgeleiteten (Haupt-) Komponenten^{21,261,267}.

Die Klassifikation kann auf Grundlage verschiedener Herangehensweisen erfolgen. Ihnen gemein ist jedoch in der Regel, dass eine Unterteilung der Proben aufgrund von gemessenen Variablen, bzw. daraus abgeleiteten Parametern erfolgt. Die einfachste Art der Unterteilung im zweidimensionalen Raum mittels einer Geraden ist in **Abb. 4.9A** gezeigt. Dieses Modell wird auch *Linear Discriminant Analysis* (LDA) genannt und wird häufig in Kombination mit einer PCA (PCA-LDA) eingesetzt^{233,267,268}.

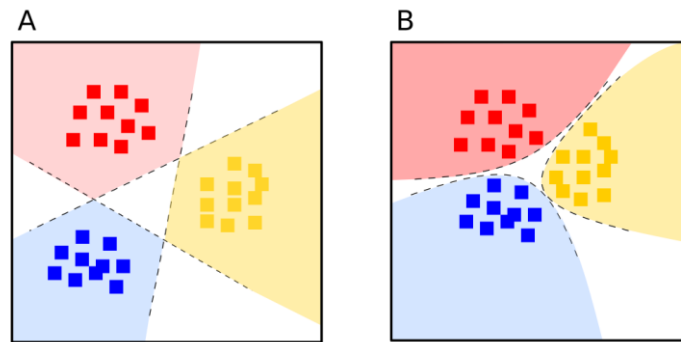


Abb. 4.9: Beispiele der Klassengrenzen bei verschiedenen Klassifikationsalgorithmen. **A:** LDA, **B:** *Support Vector Machines* (SVM).

Im gezeigten Beispiel sind die Proben, welche zur Modellerstellung genutzt wurden, alle korrekt klassifiziert worden (*Accuracy* = 100 %). Jedoch verlaufen die gezogenen Klassengrenzen je nach Algorithmus unterschiedlich. Während der LDA-Algorithmus lineare Klassengrenzen gezogen hat, wurden durch den SVM-Algorithmus mit quadratischer Kernel-Funktion nicht-lineare Klassengrenzen festgelegt (siehe **Abb. 4.9B**). Werden nun unbekannte Proben durch die Modelle vorhergesagt, kommt es zu Unterschieden in den Vorhersagen. Da nur die Trennebenen die Klassenzugehörigkeit definieren, die wiederum abhängig von der Lage der Cluster der Proben sind, welche zur Modellerstellung genutzt wurden, ist eine repräsentative Auswahl der Proben zur Modellerstellung elementar für die Leistungsfähigkeit des Modells. Eine übliche Methode stellt dafür die Kreuzvalidierung dar. Dabei wird das gesamte Probenset wiederholt in Trainings- und Validierungsproben unterteilt²⁴⁰.

Daneben ist bspw. bei der Klassifikation der Proben mittels LDA die Abgrenzung der Probencluster vergleichsweise ungenau, da die Trennebene nur unzureichend der Punktwolke der Trainingsproben folgt. Eine Klassifikation der Proben mittels eines anderen Algorithmus kann hingegen die Genauigkeit der Klassifikation erhöhen. So erlaubt beispielsweise die Klassifikation mittels SVM auch die Unterscheidung der Probengruppen mittels nicht-linearer Trennebenen²⁶⁹.

4.4.7 *Support Vector Machines* (SVM)

Ein weiterer Klassifikationsalgorithmus ist der *Support Vector Machines* (SVM) Algorithmus^{270,271}. Dieser ist insbesondere zu einer nicht-linearen Trennung von Probenklassen in der Lage^{263,269,272}. Dies führt aufgrund der nicht-linearen Verläufe der Cluster Grenzen in der Realität (siehe **Abb. 4.9B**) zu einer weitaus besseren Anpassung an die Gegebenheiten und in der Folge zu einer verbesserten Klassifikationsgenauigkeit.

Dabei erfolgt die Trennung der Probenklassen durch Übertragung der Daten in einen höherdimensionalen Raum, in dem eine Trennung der Probenklassen mittels einer linearen Hyperebene möglich ist. Die Trennung in dem höherdimensionalen Raum wird somit ähnlich zu der Trennung mittels des LDA-Algorithmus durchgeführt. Als Folge der Rücktransformation in einen niederdimensionalen Raum wird die Hyperebene verzerrt, sodass eine nicht-lineare Trennung resultiert^{166,269}. Der Verlauf der Hyperebene wird dabei durch namensgebende Stützvektoren (*support vectors*) bestimmt. Die Hyperebene wird so gewählt, dass deren Abstand zu den Probengruppen auf beiden Seiten der Ebene möglichst maximal wird^{270,273}.

Der *Cost*-Parameter C bestimmt dabei, wie viele der Variablen im Trainingsset als inkorrekt klassifiziert toleriert werden können. Damit einhergehend wird die „Auflösung“ der Trennebene reguliert. Kleine *Cost*-Werte führen zu einer eher grob definierten Trennebene aus relativ wenigen Stützvektoren, was zu einer gewissen Anzahl an inkorrekt klassifizierten Trainingsproben führt. Dagegen bietet eine solche Trennebene eine gewisse Flexibilität gegenüber dem Modell unbekannten Validierungsproben. Große *Cost*-Werte führen hingegen zu einer eher kleinteilig definierten Trennebene mit vielen Stützvektoren, welche zu überwiegend korrekt klassifizierten Trainingsproben führt. Eine solche Trennebene bietet jedoch keine besondere Flexibilität gegenüber dem Modell unbekannten Validierungsproben. Die Auswahl des *Cost*-Parameters stellt also ein Optimierungsproblem zwischen einer möglichst akkuraten Klassifikation von Trainingsproben und der möglichst guten Performance bei dem Modell unbekannten Validierungsproben dar. Somit muss der *Cost*-Parameter optimiert werden, um eine möglichst hohe *Accuracy* bei gleichzeitig minimalen *Cost*-Werten zu erzielen^{273–275}.

Stützvektoren definieren dabei die Hyperebene zur Klassentrennung, ausgehend von jenen Proben in einem Probencluster, welche eher am Rand eines Clusters liegen. Proben, welche weiter in der Mitte eines Clusters liegen, beeinflussen hingegen weniger den Klassifikationsprozess, da diese einen geringeren Einfluss auf die Lage der Hyperebene ausüben. Die Berechnung dieser Stützvektoren erfolgt grundsätzlich durch die Berechnung des Skalarproduktes. Durch geschickte Auswahl der Kernelfunktion, kann jedoch die Berechnung stark vereinfacht werden („Kernel-Trick“)^{166,263,269,273}. Über die Wahl bestimmter Kernelfunktionen, welche sowohl im niederdimensionalen Raum als auch im höherdimensionalen Raum ein gewisses, vorhersagbares Verhalten aufweisen, kann die Hyperebene bestimmt werden, ohne die Transformation in einen höherdimensionalen Raum tatsächlich zu berechnen. Dies verringert die notwendige Rechenleistung enorm. Zu den Kernelfunktionen welche diese Kriterien erfüllen, zählen u.a. die lineare Kernelfunktion, die polynomiale Kernelfunktion, die Radial-Basis Funktion (RBF) und die sigmoidale Kernelfunktion^{166,270,272}.

Das SVM-Modell muss vor allem aufgrund des *Cost*-Parameters, aber auch wegen der anderen Parameter der Kernelfunktion im Falle einer Änderung/Erweiterung des Modells erneut optimiert werden. Diese Parameteroptimierung wird auch *Grid-Search* genannt (siehe **Abschn. 4.4.7.1**)^{263,270,276}.

Der SVM-Algorithmus wurde eigentlich für eine Anwendung bei Zwei-Klassen-Problemen entworfen^{271,277}, jedoch erfolgte bald eine Erweiterung des Algorithmus auf Multiklassenprobleme^{263,277–279}.

4.4.7.1 *Grid-Search zur Parameteroptimierung*

Für optimale Klassifizierungsergebnisse unter Nutzung des SVM-Algorithmus sind teilweise mehrere Parameter zugleich zu optimieren²⁷⁶. Neben dem *Cost*-Parameter, welcher die Anzahl an genutzten Stützvektoren an den Klassengrenzen definiert, existieren zum Teil je nach genutzter Kernel-Funktion weitere Parameter, die zugleich zu optimieren sind. Beispiele dafür sind der Grad des Polynoms bei einer polynomialen Kernelfunktion ($n = 2, 3, \dots, k$), oder der *scale*-Faktor. Im Falle des RBF-Kernels ist neben dem *Cost*-Faktor C auch der Faktor γ (Gamma) zu validieren^{270,272,276,280}. Diese Parameter sind dabei oft über einen Bereich von mehreren Größenordnungen zu variieren, zudem stellen diese meist reelle Zahlen dar, sodass es theoretisch unendlich viele Möglichkeiten

der Variation gibt. Zur Bewertung werden die Parameter daher oft gleichmäßig im Zahlenraum des dekadischen Logarithmus abgerastert ($k = -2, -1, 0, +1, +2, \dots$ mit Parameterwert $= 10^k$). Bei zwei Parametern wird so ein 2-dimensionales Raster an Punkten gebildet, weshalb dieses Verfahren auch „*Grid Search*“ genannt wird^{270,276}.

Die Optimierung des Klassifikationsalgorithmus erfolgt dabei insofern, als das für alle Parameter in den oben definierten Abständen, jeweils eine Klassifikation mittels SVM durchgeführt und die *Accuracy* berechnet wird, die möglichst maximiert werden soll.

4.4.8 Random Forest (RF)

Der *Random-Forest*-Algorithmus (RF-Algorithmus) ist ein weiterer Algorithmus, der Proben aufgrund ihrer Ähnlichkeit zu Referenzproben bestimmten Probenklassen zuordnen kann. Er wurde 2001 von Leo Breiman entwickelt²⁸¹. Die Funktionsweise des Algorithmus basiert auf einer baumartigen Struktur, welche aus den Messdaten generiert wird. Pro Klassifikation werden mehrere dieser Bäume erstellt. Dabei stellen die zu klassifizierenden Proben je Baum eine zufällig aus der Gesamtmenge aller Proben ausgewählter Untermenge aller Proben dar („*bagging*“). Dadurch ist es möglich die Performance des Klassifikationsbaums über die übrigen Proben („*out-of-bag*“ samples) zu verifizieren^{282,283}.

Die Anzahl an insgesamt generierten Bäumen wird dabei durch den Parameter *ntree* festgelegt. Übliche Werte für *ntree* sind 500, 1000, seltener auch 2000 oder mehr^{284,285}. Das Klassifikationsergebnis des RF-Algorithmus ist dabei die häufigste gefällte Klassifikationsentscheidung aller Bäume je Probe²⁸².

Die Klassifikation anhand eines einzelnen RF-Baums wird folgendermaßen durchgeführt: Die Proben werden an jeder Verästelung (Knoten) des Baumes in zwei Untergruppen unterteilt. Diese Verteilung beruht auf den jeweiligen Messwerten eines Markers, welcher aus einer Subgruppe von wiederum zufällig ausgewählten Markern dazu am besten geeignet ist. Die Unterscheidung der Proben an jedem Knoten erfolgt dabei binär je nachdem, ob der genutzte Marker in den Proben größer oder kleiner als ein bestimmter Wert ist. Die Anzahl an Kandidaten-Markern, aus dessen Pool die Marker zur Unterscheidung der Proben gezogen wird, wird dabei durch den Parameter *mtry* bestimmt^{282,284}. Üblicherweise wird als Parameter *mtry* die Wurzel aus der Variablenanzahl genutzt^{282–285}. Die Bewertung der Marker bzw. die Ermittlung des besten Markers zur Unterscheidung der Proben an den Knoten des RF-Baums, erfolgt meistens anhand des GINI-Indexes²⁸².

Somit wird bei einem Random-Forest-Klassifikationsmodell nicht nur eine einzige Klassifikation der Proben durchgeführt. Vielmehr wird jede Probe von mehreren modellierten Bäumen vorhergesagt. Somit ist der RF-Algorithmus recht robust, gegenüber einem *Overfitting*^{286–288}.

4.4.9 Kreuzvalidierung zur Bewertung der Genauigkeit von Klassifikationsalgorithmen

Die Genauigkeit von Klassifikationsmodellen kann nicht anhand derselben Proben, welche bereits zur Erstellung des Modells genutzt wurden, geprüft werden, da dabei die Genauigkeit aufgrund der identischen Probenbasis in Relation zu anderen Proben tendenziell überschätzt wird. Daher ist zur objektiven Bestimmung der Klassifikationsgenauigkeit eine Validierung des erstellten Modells mit Proben, welche nicht bereits zur Modellerstellung genutzt wurden, nötig²⁸⁹. Diese

Vorgehensweise bietet eine objektive Überprüfung der Genauigkeit des Modells an. Zudem ist es möglich aufgrund der natürlichen, biologischen Variabilität der Probenzusammensetzung eine Abschätzung des auftretenden Fehlers durchzuführen.

Dabei werden die interne und die externe Validierung voneinander unterschieden. Bei der internen Validierung stammen die Testproben aus demselben Probensatz, während bei der externen Validierung ein komplett unabhängig analysiertes Probenset zwecks Validierung vorgehalten wird. Dabei werden in beiden Fällen, Trainingsproben, welche zur Modellerstellung herangezogen wurden, nie als Testproben zur Validierung genutzt und umgekehrt^{241,290}.

Im Falle der internen Validierung gibt es unterschiedliche Herangehensweisen hinsichtlich der Aufteilung der Proben auf das Trainingsprobenset und das Testprobenset. Neben der rein zufälligen Verteilung, der *Venetian-Blind* Verteilung und der *Leave-One-Out* Kreuzvalidierung, ist vor allem die *k-fold* Kreuzvalidierung (engl. *cross validation*) zu nennen^{240,291}. Bei der *k-fold* Kreuzvalidierung wird der Probensatz zunächst in *k* Teile unterteilt. Die Unterteilung erfolgt dabei zufällig, jedoch wird in der Regel eine gleichmäßige Verteilung der Probenklassen auf den Trainings-/Validierungs-Teildatensatz angestrebt. Diese Art der Probenaufteilung wird Stratifikation genannt²⁴⁰. Die insgesamt *k* Teildatensätze werden unterteilt in *k*-1 Teile, die zu einem Trainingsdatensatz zusammengefasst werden und einem Teil, der zur Validierung zu Verfügung steht. Diese Aufteilung wird insgesamt *k*-fach wiederholt, wobei derjenige Datenteil, welcher zur Validierung genutzt wird, jeweils wechselt, bis alle Teile jeweils einmal als Validierungsteilprobensatz gedient haben, siehe **Abb. 4.10**.

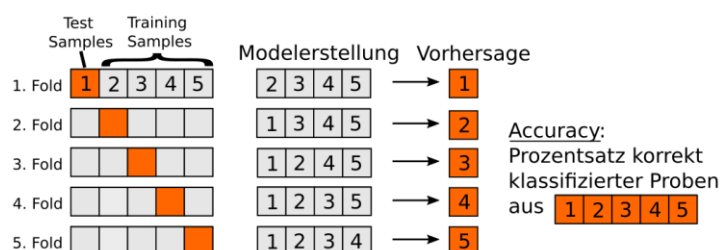


Abb. 4.10: Schema der *k-fold* Kreuzvalidierung mit $k = 5$, nach Raschka²⁴⁰.

Die Klassifikationsgenauigkeit wird bei der Kreuzvalidierung als Prozentsatz aller korrekt vorhergesagten Proben des Validierungsteilprobensatz angegeben. Da sämtliche *k* Teildatensätze jeweils einmal vorhergesagt wurden, liegen schlussendlich Vorhersagen für alle Proben des Datensatzes vor. Übliche Werte für *k* sind vor allem $k = 10$, was einer Aufteilung in 90 % der Proben als Trainingsset und 10 % als Validierungsset entspricht. Daneben sind jedoch auch $k = 5$ (80 %/20 %) oder 7 (86 %/14 %) in Gebrauch^{240,261}.

Je nach der genauen Aufteilung der Proben auf das Trainings/Validierungsset kommt es zu geringfügigen Unterschieden bei der ermittelten Klassifikationsgenauigkeit. Daher ist es durchaus üblich, die Kreuzvalidierung mehrfach durchzuführen („*repeated cross validation*“)^{292–294}. Die Anzahl der durchgeführten Wiederholungen beträgt dabei üblicherweise bis zu 100^{233,292,295,296}. Die in der Regel auftretende Schwankung der Klassifikationsgenauigkeit kann dabei als Standardabweichung angegeben werden. Allerdings ist es in der Literatur umstritten, ob die dabei errechnete Standardabweichung aussagekräftig ist, oder ob sie einem *Bias* unterliegt, aufgrund

der Tatsache, dass die Wiederholungen der Kreuzvalidierung nicht voneinander unabhängig sind^{293,294}.

4.4.10 Fusion von Analysenergebnissen verschiedener Methoden zur Klassifikation

Die Fusion von Analysedaten von unterschiedlichen Analysemethoden kann einen Vorteil bei der Algorithmen-basierten Klassifikation bieten. Können beispielsweise bestimmte Probenklassen anhand einer Analysemethode in Bezug auf ein untersuchtes Merkmal, wie bspw. die geographische Herkunft, botanische Sorte o. ä., nicht unterschieden werden, so kann sich die Situation im Falle einer komplementären Analysemethode anders darstellen. Somit kann die Kombination unterschiedlicher Analysemethoden bei der Klassifikation einen Vorteil liefern. Bspw. zeigten Wang *et al.*, dass die Klassifikation des Pilzes *Macrohyporia cocos* nach geographischer Herkunft mittels Analysedaten der FT-IR-Spektroskopie und UV-Chromatogrammen bei 210 nm, bei *low-level* Datenfusion eine Klassifizierungsgenauigkeit von 100 % erzielte, während die Einzelmethode lediglich Werte von 92.3 bzw. 76.9 % erreichten²⁹⁷.

Generell wird zwischen verschiedenen Strategien zur Datenfusion unterschieden, je nachdem, zu welchem Zeitpunkt der Datenverarbeitung die Zusammenführung der Daten der Messmethoden erfolgen: *low-level*, *mid-level* oder *high-level* Fusion^{298,299}. Die unterschiedlichen methodischen Ansätze sind in **Abb. 4.11** schematisch nach Pei *et al.*²⁹⁹ dargestellt.

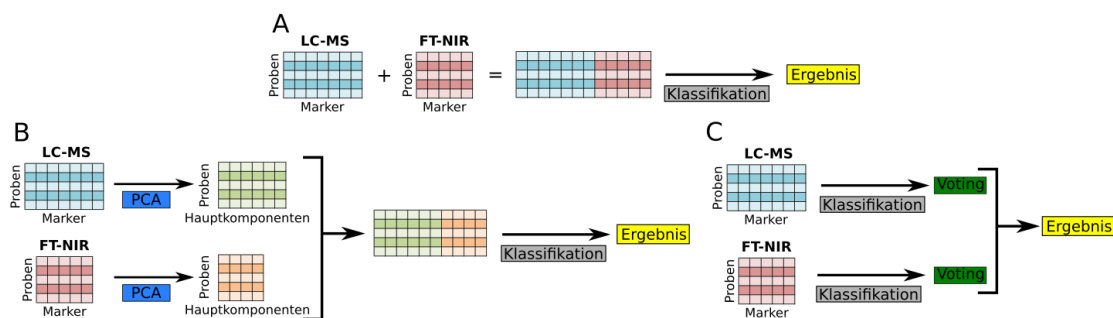


Abb. 4.11: Varianten der Datenfusion. **A:** *low-level* Fusion. **B:** *mid-level* Fusion. **C:** *high-level* Fusion. Schema adaptiert nach Pei *et al.*²⁹⁹.

Im Falle der *low-level* Fusion wird die Fusion der Datenmatrices nach dem *Preprocessing* durchgeführt. Die Klassifikation der Proben anhand der Analysedaten erfolgt somit gemeinsam für alle Analysemethoden in einer einzigen, gemeinsamen Datenmatrix^{297,299}. Die *mid-level* Fusion ist eine Methode, die in dem speziellen Fall der Kombination von multivariater Datenanalyse mit darauf folgender Klassifikation (z. B. PCA) angewendet werden kann. Bei der *mid-level* Fusion, werden die Scores-Matrizen der PCA je Analysemethode getrennt berechnet und vor der eigentlichen Klassifikation zu einer einzigen Matrix zusammengeführt^{297,299}. Die Klassifikation erfolgt somit mit Daten aus beiden Methoden. Bei der *low-level* sowie der *mid-level* Datenfusion ist das *Preprocessing* der Rohdaten von besonderer Bedeutung, da ohne entsprechendes *Preprocessing* die Gefahr bestünde, dass eine der beiden Analysenmethode im weiteren Verlauf dominiert und dadurch die Datenfusion keinen Vorteil bietet. Bei der *high-level* Fusion erfolgt die Fusion der beiden Methoden erst nach der Klassifikation der Proben^{297,298}. Dabei werden die Klassifikationsentscheidungen des Algorithmus jeweils ausschließlich anhand jeder Messmethode für sich betrachtet, getroffen. Die finale Klassifikation wird mittels eines *Voting*-Algorithmus der

die Klassifikationsentscheidungen der einzelnen Algorithmen bewertet und daraus ein Gesamtvotum errechnet, durchgeführt^{299,300}.

Diese verschiedenen Arten der Datenfusion gehen dabei einher mit einer unterschiedlichen Tendenz in Bezug auf die Klassifikationsgenauigkeit. Dabei ist im Allgemeinen die Reihenfolge der Klassifikationsgenauigkeit: *mid-level* > *high-level* > *low-level* Datenfusion²⁹⁸.

5 Zielsetzung der Arbeit

Das Ziel dieser Arbeit war es, die Möglichkeit einer Bestimmung/Verifikation des geographischen Ursprungslandes von Körnermais, mittels analytischer Methoden zu evaluieren. Dabei sollten die Analysen mittels LC-MS bzw. FT-NIR-Spektroskopie durchgeführt werden. Diese beiden Analyseverfahren wurden bereits erfolgreich zuvor für die Herkunftsbestimmung von Spargel bzw. Haselnüssen in der Arbeitsgruppe angewandt^{43,45,46}. FT-NIR-Analysen gehören zudem zu den üblichen analytischen Verfahren zur Bestimmung des Proteingehaltes sowie des Restfeuchtegehaltes bei der Verarbeitung von Cerealien in der Industrie^{27,47,48,178,301–303}. Daher erschien eine Erweiterung der Erkenntnisse auf der Grundlage der möglicherweise bereits für andere Zwecke erfolgten FT-NIR-Analysen besonders vielversprechend.

Optimierung der Probenvorbereitung

Um die Körnermaisproben optimal analysieren zu können, sollte zunächst die optimale Probenform für die FT-NIR-Analyse evaluiert werden. Dazu wurde dieselbe Probe einmal als ganze, intakte Körner und als vermahlene, pulverförmige Probe untersucht. Für die Analyse des Metaboloms sollte aufgrund der hohen Anzahl an Metaboliten in Pflanzen und der Schwierigkeit der Erfassung durch eine einzelne analytische Methode^{304–306}, ein mehrstufiges Vorgehen erfolgen: Zum einen sollte der Fokus einmal auf eher polaren Metaboliten (z. B. Zucker, Aminosäuren, organischen Säuren) liegen. Zum anderen sollten unpolare Metabolite (Lipide sowie weitere lipophile Substanzen) untersucht werden. Um den Abdeckungsgrad des Metaboloms möglichst zu maximieren, wurden jeweils aus zwölf literaturbekannten, üblichen Extraktionsmethoden für polare sowie unpolare Metabolite, die geeignetsten für die Analyse der (un)polaren Metabolite aus Körnermaisproben ermittelt.

Einflussfaktor Maissorte

Der möglicherweise störende Einfluss der Körnermaissorte sollte daraufhin systematisch anhand von Proben aus Landessortenversuchen bestimmt werden. Die Analyse von Proben aus Sortenversuchen bietet dabei den großen Vorteil, dass die im Handel üblicherweise nicht deklarierte Sorte bekannt ist. Gleichzeitig bieten die Landessortenversuche ein optimales Ausgangsmaterial, da der Anbau sehr definiert erfolgt. Daher konnte der Einfluss der Körnermaissorte auf die detektierbaren Metaboliten je Methode bestimmt werden, sodass eine optimale Methode zur Herkunftsbestimmung gewählt werden konnte.

Bestimmung von Markerbanden/Markersubstanzen für die Herkunftsbestimmung

Um die geographische Herkunft von Körnermaisproben mittels der LC-MS-Analytik zu bestimmen, sollte das Probenset mittels beider optimierter Extraktionsmethoden extrahiert werden. Anschließend erfolgte die Analyse mit den entsprechenden LC-MS-Methoden sowohl im positiven als auch im negativen Ionenmodus. Die Proben wurden zudem mittels eines FT-NIR-Spektrometers analysiert.

Für beide Analysemethoden erfolgte dann getrennt eine Selektion der herkunftsspezifisch beeinflussten Analysedaten mittels statistischer Methoden und/oder multivariater Datenanalyse. Die Zuordnung des Herkunftslandes anhand der erhobenen Daten sollte dabei mittels eines

Klassifikationsalgorithmus erfolgen. Dieser sollte die Proben anhand der Ähnlichkeit der Analysedaten der selektierten Features entsprechend zu ihren Herkunftsländern zuordnen. Zudem sollte die Anzahl der notwendigen Markerfeatures zur Klassifikation verringert werden, wobei deren Auswahl so erfolgte, dass die Klassifikationsraten sich nicht verschlechtern. So konnte ein einfacheres Modell generiert werden, welches in die Routineanalytik übertragen werden kann. Die dabei genutzten Metabolite bei der massenspektrometrischen Analytik wurden zudem mittels MS/MS-Spektren identifiziert, um die Etablierung einer quantitativen, zielgerichteten Methode zur Herkunftsbestimmung von Körnermais zu ermöglichen.

Evaluation der Eignung der Markermetabolite der LC-MS-Analyse zur Herkunftsbestimmung in Routinemethoden

Die Robustheit eines repräsentativen Teils der selektierten Markermetabolite der LC-MS-Methode zur Herkunftsbestimmung wurde zudem gezeigt, indem diese in eine MRM-Methode implementiert wurden. Anhand von Standardsubstanzen wurde außerdem eine Quantifizierung vorgenommen. Ebenso wurde die Stabilität der Markermetabolite in Hinblick auf die Lagerungsbedingungen überprüft. Zudem wurde ein Vergleich zu einer, lediglich auf dem Tocopherol-Profil beruhenden Unterscheidung der Herkunftsländer gezogen, da Tocopherole in der Vergangenheit bereits als Marker zur Herkunftsbestimmung genutzt wurden^{307–310}.

Analyse der Klassifikationsergebnisse bei Datenfusion der Markerbanden der FT-NIR-Analyse und der Markermetabolite der LC-MS-Analyse

Um die Leistungsfähigkeit eines kombinatorischen Ansatzes der analytischen Herkunftsbestimmung mittels der LC-MS-Analyse und der FT-NIR-Spektroskopie zu evaluieren, erfolgte außerdem eine Datenfusion auf *low-level* Ebene^{299,311–314}. Die Leistungsfähigkeit der kombinierten Nutzung der Analysedaten beider Methoden zur Herkunftsclassifizierung sollte insbesondere im Hinblick auf den Vergleich mit den jeweiligen Einzelmethoden untersucht werden.

6 Ergebnisse und Diskussion

6.1 Optimierung der Probenextraktion

Die Untersuchung von Maisproben auf ihre geographische Herkunft unter Ausnutzung der Unterschiede ihres Metaboloms machte es notwendig, die Proben entlang des analytischen Prozesses identisch zu behandeln. Die Minimierung von Schwankungen durch die Probenvorbereitung gewährleistete, dass gemessene Unterschiede auf die geographische Herkunft der Proben zurückzuführen sind.

Um eine artifizielle Veränderung der Körnermaisproben zu vermeiden, wurden die Proben bei niedrigen Temperaturen gelagert und verarbeitet. Nach der Vermahlung der Proben wurden diese für 24 h gefriergetrocknet. Dies diente der Verringerung der enzymatischen Aktivität durch Reduktion der in den Proben vorhandenen Restfeuchte von ca. 13 %⁷⁸.

Die Methode der Herkunftsbestimmung von Lebensmitteln mittels der Analyse des Metaboloms mittels LC-MS wurde im Arbeitskreis bereits zuvor von Klockmann *et al.* mit der Matrix Haselnüsse und Creydt *et al.* mit der Matrix Spargel durchgeführt^{43,46}.

Das komplette Metabolom einer einzelnen Pflanze umfasst laut Trethewey bis zu 25000 Metabolite^{304,305}, welche nicht alle mittels einer einzigen Analysenmethode erfasst werden können. Für die Analytik des Metaboloms von Körnermais wurden daher zwei unterschiedliche Extraktionsmethoden angewandt. Eine für eher unpolare Metabolite, v. a. Lipide, und eine für polare Metabolite, wie Aminosäuren, Polyphenole, organische Säuren u. a. Um störende Proteine aus der Probe möglichst zu präzipitieren, wurde jeweils das organische Lösungsmittel zuerst zugegeben und erst in einem zweiten Schritt ggfls. Wasser hinzugefügt. Da die verwendete Extraktionsmethode einen elementaren Einfluss auf die Ergebnisse der Metabolomanalytik hat, wurden die Extraktionsmethoden auf die kohlenhydratreiche Matrix Mais optimiert.

Hingegen wurden die UHPLC-Methoden und MS-Methoden weitestgehend von Creydt *et al.* übernommen⁴³. Für die Chromatographie von unpolaren Metaboliten wurde jedoch eine andere Chromatographiesäule als von Creydt *et al.* genutzt (Thermo Accucore RP-MS mit 2.6 µm Partikelgröße anstelle von Phenomenex Kinetex RP-C18 mit 1.7 µm Partikelgröße). Aufgrund des geringeren Rückdrucks wurde die Säulenofentemperatur auf 30 °C anstelle von 50 °C festgelegt.

Sämtliche Versuche, welche der Optimierung von Versuchen dienten, wurden mit der Qualitätskontrollprobe (QC-Probe) ausgeführt um die Vergleichbarkeit der Ergebnisse in Hinblick auf das Probenmaterial zu gewährleisten. Die QC-Probe stammte aus Deutschland und wurde in einem größeren Maßstab vermahlen, homogenisiert und dann in beiden Forschungsstellen als Kontrollprobe für eine gleichbleibende hohe analytische Qualität im Rahmen der Analysen wiederholt untersucht.

6.1.1 Optimierung der Extraktionsmethode für polare Metabolite

Polare Metabolite sind eine heterogene Mischung aus vielen unterschiedlichen Substanzklassen¹¹⁵. Deshalb ist es schwierig, eine generell anwendbare Extraktionsmethode zu benennen. Oft wird mit Mischungen aus Alkoholen und Wasser gearbeitet¹¹⁵, was den polaren Charakter des gewünschten

Metabolitspektrums widerspiegelt. Neben Ein-Phasen-Extraktionsmethoden, sind auch die polaren, wässrigen Phasen von Zwei-Phasen-Extraktionsmethoden interessant. Diese wurden bereits zuvor für die Untersuchung von polaren Metaboliten eingesetzt, insbesondere als Methode für die gleichzeitige Extraktion von polaren und unpolaren Metaboliten in einem Arbeitsgang^{315–318}. Sie bieten den Vorteil, dass die störenden Lipide der Probe in der unpolaren, organischen Phase angereichert werden und so von den polaren Metaboliten in der wässrigen Phase separiert werden können. Es wurden folgende Extraktionsmittel bzw. -methoden getestet: die Alkohole Methanol und Ethanol, sowie deren Mischungen mit Wasser. Außerdem die wässrigen Phasen, der eigentlich für Lipidextraktionen optimierten Extraktionsmethoden nach FOLCH³¹⁹, BLIGH & DYER³²⁰, LÖFGREN³²¹ und ARMIROTTI³²². Die getesteten Extraktionsmethoden für polare Metabolite sind in **Tab. 6.1** zusammen mit Literaturreferenzen aufgeführt.

Tab. 6.1: Evaluierte Extraktionsmethoden für die Detektion polarer Metabolite.

Nr.	Extraktionsgemisch	Literaturstelle
Ex01	Methanol/Wasser 1:1 (v/v)	115,306,323
Ex02	Methanol/Wasser 1:2 (v/v)	
Ex03	Methanol/Wasser 2:1 (v/v)	
Ex04	Ethanol/Wasser 1:1 (v/v)	306,324,325
Ex05	Ethanol/Wasser 1:2 (v/v)	
Ex06	Ethanol/Wasser 2:1 (v/v)	
Ex07	Methanol	115,306,323,324
Ex08	Ethanol	306,325
Ex09	Chloroform/Methanol/Wasser 4.5:2:1 (v/v/v), wässrige Phase	FOLCH ³¹⁹
Ex10	Chloroform/Methanol/Wasser 1:1:1 (v/v/v), wässrige Phase	BLIGH & DYER ^{306,320}
Ex11	1-Butanol/Methanol/ <i>n</i> -Heptan/Ethylacetat/Wasser 3:1:1.5:0.5:4 (v/v/v/v), wässrige Phase	LÖFGREN ³²¹
Ex12	Chloroform/Methanol/ <i>n</i> -Heptan/Wasser 3:4:1:2 (v/v/v/v), wässrige Phase	ARMIROTTI ³²²

Zur besseren Vergleichbarkeit wurden die Einphasen-Extraktionsmethoden auf ein Volumen von 1.0 mL und die Zweiphasen-Extraktionsmethoden auf ein Volumen von 1.5 mL normiert (siehe Creydt *et al.*)³⁰⁶, die eingewogene Probenmenge betrug in allen Fällen 50 mg.

Die Kompatibilität von Extrakt und Laufmittel wurde vorab durch Mischung der jeweiligen Extrakte im Verhältnis 1:1 mit Laufmittel B (Acetonitril/Methanol ohne Additiv, siehe **Abschn. 9.6.1**) geprüft. Da Vorversuche zeigten, dass die Konzentrationen an extrahierten polaren Metaboliten aus Körnermais relativ gering ausfielen, wurden je 7 µL injiziert des unverdünnten Extraktes injiziert. Polare Metabolite des Körnermais lagen außerdem nicht in einem Ausmaß vor, wie die unpolaren Metabolite (siehe **Abschn. 6.1.2**).

Die Ergebnisse der LC-MS-Messungen der polaren Extrakte sind in **Abb. 6.1** dargestellt. Die Summe der ermittelten Peakflächen der Features in beiden Ionenmodi ist gegen die eingesetzte Extraktionsmethode aufgetragen. Die Tendenz für die jeweils ermittelte Anzahl der detektierten Features war dabei ähnlich (siehe **Anhang Abb. 13.1**).

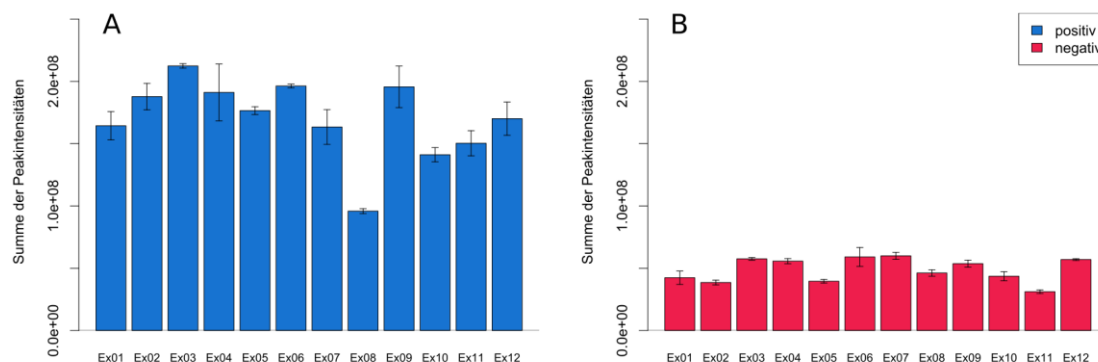


Abb. 6.1: Optimierung der Extraktionsmethode für polare Metabolite. Genutzt wurde jeweils die Summe der Peakintensitäten der Features in der QC-Probe. **A:** positiver Ionenmodus; **B:** negativer Ionenmodus. Evaluierte Extraktionsmethoden, siehe **Tab. 6.1** (n = 3).

Im negativen Ionenmodus fiel die Summe der Peakintensitäten um ca. 2/3 geringer aus als im positiven Ionenmodus. Außerdem variierten die summierten Peakintensitäten zwischen den Extraktionsmethoden stärker im positiven als im negativen Ionenmodus. Es zeigte sich zudem kein grundlegender Unterschied zwischen den Einphasen-Extraktionsmethoden (Ex01-Ex08) und den zwei Phasen ausbildenden Extraktionsmethoden (Ex09-Ex12). Lediglich die relative Standardabweichung (RSD) war bei den Zweiphasenmethoden um ca. 2 % erhöht.

Die folgenden drei Methoden schnitten im positiven Ionenmodus bei Summation der Peakintensitäten am besten ab: Methanol/Wasser 2:1 v/v, Ethanol/Wasser 2:1 v/v sowie FOLCH (wässrige Phase). Im negativen Ionenmodus waren die besten drei Extraktionsmethoden dagegen: reines Methanol, Ethanol/Wasser 2:1 v/v und Methanol/Wasser 2:1 v/v. Die RSDs der Methoden in beiden Ionenmodi stieg in der Reihenfolge von Methanol/Wasser 2:1 v/v $\approx$ Ethanol/Wasser 2:1 v/v < reines Methanol $\approx$ FOLCH von ca. 1 % auf ca. 5-8 % an. Das optimale Extraktionsgemisch war daher entweder Methanol/Wasser oder Ethanol/Wasser im Verhältnis 2:1 v/v.

Die Extraktionseffizienz für polare Metabolite sollte sich mit zunehmender Kettenlänge der Alkohole verschlechtern, da die Polarität des verwendeten Alkohols bei zunehmender Kettenlänge des Alkohols abnimmt. Dieser Trend konnte auch im Vergleich von Methanol und Ethanol als Extraktionsmittel nachvollzogen werden, wobei 40 % mehr extrahierte Metabolite mittels Methanol erhalten wurden. Daher wurde Methanol als organische Komponente im Extraktionsmittel gewählt. Die Extraktion mittels einer Mischung aus Methanol/Wasser im Verhältnis 1:2 v/v wurde zudem bereits zuvor von Creydt *et al.*, als optimale Extraktionsmethode für polare Metabolite des weißen Spargels ermittelt³⁰⁶. Die Verwendung von mehr Wasser im Extraktionsgemisch führte zwar zu ähnlich guten Ergebnissen bei Körnermais, jedoch zeigte sich bei dieser Matrix, dass dieses Mischungsverhältnis ein trübes Extrakt lieferte. Daher wurde die Methode mit einem Mischungsverhältnis von Methanol zu Wasser von 2:1 v/v als optimale polare Extraktionsmethode gewählt.

6.1.2 Optimierung der Extraktionsmethode für unpolare Metabolite

Anders als polare Metabolite sind unpolare Metabolite vor allem einer Stoffklasse unterzuordnen, den Lipiden. Extraktionsmethoden von Lipiden aus humanen bzw. tierischen Gewebe sind seit

vielen Jahren etabliert, seitdem die Zweiphasen-Extraktionsmethoden von FOLCH 1957 und BLIGH & DYER 1959 veröffentlicht wurden^{319,320}. Es wurden folgende Extraktionsmethoden für unpolare Metabolite evaluiert: Zweiphasen-Extraktionsmethoden für Lipide aus tierischen Gewebe – nach FOLCH³¹⁹, BLIGH & DYER³²⁰, LÖFGREN³²¹, ARMIROTTI³²², HARA³²⁶ sowie MATYASH³²⁷.

Aufgrund des Chloroformgehaltes der Extraktionsmittel der Methoden nach FOLCH bzw. BLIGH & DYER hat es zudem immer wieder Versuche gegeben, alternative Methoden zu etablieren, um das möglicherweise krebserregende Chloroform zu ersetzen. Vor allem die Extraktionsmethode, welche von Matyash *et al.* entwickelt wurde, ist hier zu nennen. Der Chloroformanteil im Extraktionsmittel wurde dabei durch das weniger bedenkliche Methyl-*tert.*-butylether (MTBE) ersetzt³²⁷. Zudem wurden die etablierten Methoden nach FOLCH³¹⁹, sowie BLIGH & DYER³²⁰, auch in einer modifizierten Form erprobt, indem bei der BLIGH & DYER-Methode das Chloroform durch Dichlormethan ersetzt wurde, bzw. in beiden Methoden das toxische Methanol durch das unbedenklichere Isopropanol substituiert wurde.

Zusätzlich wurden drei Einphasen-Extraktionsmethoden untersucht: eine Mischung aus Chloroform und Methanol 2:1 v/v³²⁸, eine Mischung aus Chloroform und Isopropanol 2:1 v/v⁴⁶ und eine Mischung aus den drei Lösungsmitteln MTBE, Methanol und Chloroform (4:3:3 v/v/v), die sich bei der Extraktion von Blutplasma als vorteilhaft erwies³²⁹. Die getesteten Extraktionsmethoden für unpolare Metabolite sind in **Tab. 6.2** zusammen mit Literaturreferenzen aufgeführt.

Tab. 6.2: Evaluierte Extraktionsmethoden für die Detektion unpolarer Metabolite.

Nr.	Extraktionsgemisch	Literaturstelle
Ex01	Chloroform/Methanol/Wasser 4.5:2:1 (v/v/v)	FOLCH ³¹⁹
Ex02	Chloroform/Methanol/Wasser 1:1:1 (v/v/v)	BLIGH & DYER ^{306,320}
Ex03	1-Butanol/Methanol/ <i>n</i> -Heptan/Ethylacetat/Wasser 3:1:1.5:0.5:4 (v/v/v/v)	LÖFGREN ³²¹
Ex04	Chloroform/Methanol/ <i>n</i> -Heptan/Wasser 3:4:1:2 (v/v/v/v)	ARMIROTTI ³²²
Ex05	<i>tert.</i> -MTBE/Methanol/Wasser 5:1.5:1 (v/v/v)	MATYASH ³²⁷
Ex06	<i>n</i> -Hexan/Isopropanol/Wasser 5:4.5:3 (v/v/v)	HARA ³²⁶
Ex07	BLIGH & DYER (Chloroform ersetzt durch Dichlormethan)	BLIGH & DYER*
Ex08	FOLCH (Methanol ersetzt durch Isopropanol)	FOLCH*
Ex09	BLIGH & DYER (Methanol ersetzt durch Isopropanol)	BLIGH & DYER*
Ex10	<i>tert.</i> -MTBE/Methanol/Chloroform 4:3:3 (v/v/v)	PELLEGRINO ³²⁹
Ex11	Chloroform/Isopropanol 1:2 (v/v)	⁴⁶
Ex12	Chloroform/Methanol 1:2 (v/v)	³²⁸

* Extraktionsgemisch modifiziert, siehe Spalte „Extraktionsgemisch“.

Zur besseren Vergleichbarkeit wurden die Volumina, wie bereits bei der Optimierung der Extraktion der polaren Metaboliten auf 1.0 mL bzw. 1.5 mL normiert, die Probenmenge betrug auch hier 50 mg. Bei den Zweiphasen-Extraktionsmethoden wurde jeweils die unpolare, organische Phase analysiert. Die Extrakte wurden mit Laufmittel B (Isopropanol/Acetonitril ohne Additiv, siehe **Abschn. 9.6.2**) um den Faktor 1:5 verdünnt und je 8 µL Probe injiziert. Vorversuche zeigten, dass die Signalintensitäten der extrahierten Metabolite, vor allem im Elutionsbereich der Triglyceride (TG) sehr hoch ausfiel. Wohingegen die Intensitäten im Bereich der Elution der Phospholipide eher gering waren. Ziel der Extraktionsoptimierung war es daher u. a. eine möglichst ausgewogene Extraktionsbilanz zwischen den Phospholipiden und den TGs in den Körnermaisproben zu erhalten.

Die Ergebnisse der LC-MS-Messungen der unpolaren Extrakte sind in **Abb. 6.2** dargestellt. Die Evaluation der Extraktionsmethoden erfolgte nach denselben Kriterien, die bereits für polare Metabolite angewandt wurden. Für die jeweilige Anzahl der detektierten Features konnte eine ähnliche Tendenz, wie für die Summe der Peakintensitäten festgestellt werden. Für die entsprechende Abbildung siehe **Anhang Abb. 13.2**.

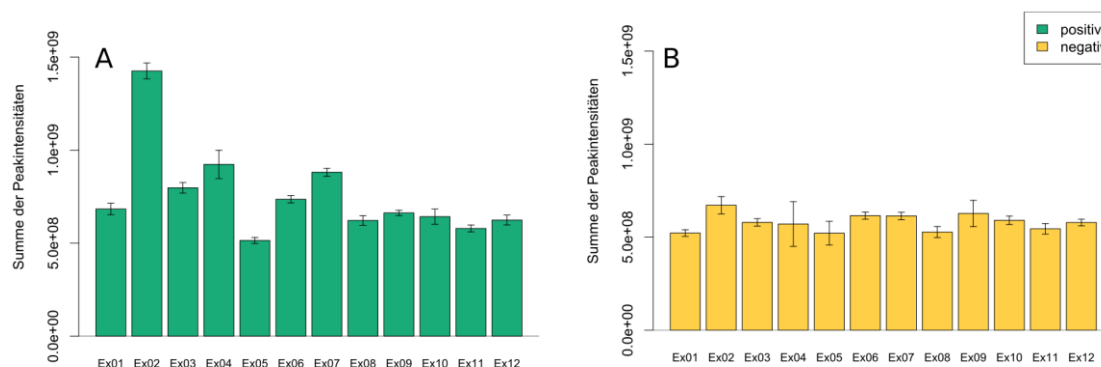


Abb. 6.2: Optimierung der Extraktionsmethode für unpolare Metabolite. Genutzt wurde die jeweilige Summe der Peakintensitäten der Features. **A:** positiver Ionenmodus; **B:** negativer Ionenmodus. Evaluierte Extraktionsmethoden, siehe **Tab. 6.2** (n = 3).

Die Leistungsfähigkeit der Extraktionsmethoden war – wie auch bereits bei der Analytik der polaren Metabolite im positiven Ionenmodus leicht besser als im negativen Ionenmodus. Mit um ca. 1/5 im Mittel größeren Intensitäten waren die Unterschiede jedoch nicht so gravierend, wie zuvor bei der Analytik der polaren Metabolite.

Hingegen zeigten sich weitaus grundlegendere Unterschiede zwischen den Zweiphasen- und den Einphasen-Extraktionsmethoden: Die Zwei-Phasen-Extraktionsmethoden wiesen für unpolare Metabolite im Allgemeinen eine z. T. deutlich größere Summe der Peakintensitäten auf, als die Einphasen-Extraktionsmethoden. Die besten drei Extraktionsmethoden nach absoluten Zahlen im positiven Ionenmodus waren: BLIGH & DYER, ARMIROTTI und BLIGH & DYER (Dichlormethan); erst an achter Stelle folgte die beste Einphasen-Extraktionsmethode nach PELLEGRINO. Im negativen Ionenmodus zeigte sich eine Staffellung nach: BLIGH & DYER, BLIGH & DYER (Isopropanol) sowie HARA. Diese Ergebnisse waren zu erwarten, da die getesteten Zweiphasen-Extraktionsmethoden für die Extraktion von Lipiden aus humanen oder tierischen Gewebe entwickelt wurden und z. T. seit über 60 Jahren im Einsatz sind^{319,320}.

Die Standardabweichung der Dreifachbestimmungen zeigte, dass die Extraktionsmethoden eine RSD von 3-12 % aufwiesen. Die RSD-Werte waren dabei im positiven Ionenmodus weitaus geringer (3-7 %), als im negativen (3-12 %). Eine Ausnahme stellte dabei nur die Extraktionsmethode nach ARMIROTTI dar, welche eine höhere RSD von 8 bzw. 22 % aufwies. Somit ergab sich, dass die Lipid-Extraktionsmethode nach BLIGH & DYER, aufgrund der herausragenden Ergebnisse in beiden Ionenmodi, die optimale Extraktionsmethode für unpolare Metabolite von Körnermais darstellt. Diese Extraktionsmethode wurde auch bereits von Creydt *et al.* als optimale Extraktionsmethode für unpolare Metabolite von weißem Spargel ermittelt³⁰⁶.

Bei der LC-MS-Analyse des verdünnten BLIGH & DYER-Extraktes wiesen lediglich vier der im positiven Ionenmodus detektierten Features eine Detektorsättigung auf, hauptsächlich verursacht durch TGs. Besonders hervorzuheben ist dabei das Triglycerid Trilineolin ($C_{57}H_{98}O_6$, m/z -Wert des $[M+NH_4]^+$ -Addukts 896.77). Dieses kommt literaturbekannt mit 25 % in Maiskeimöl am häufigsten von allen TGs vor³³⁰. Aufgrund des Ungleichgewichts zwischen TGs und Phospholipiden im Chromatogramm wurde jedoch eine geringe Anzahl von Features mit Detektorsättigung in Kauf genommen. Somit wurde das optimale Injektionsvolumen auf 8 μ L des mit Isopropanol/Acetonitril 3:1 v/v um den Faktor 1:5 verdünnten Extraktes bestimmt. Bei diesem Injektionsvolumen kam es zur Detektorsättigung bei insgesamt fünf TGs, sowie zwei Diglyceriden (DG).

6.2 Untersuchung einer möglichen Beeinflussung des Körnermaismetaboloms durch die Maissorte

6.2.1 Untersuchung einer Sortenabhängigkeit der Metabolite

Neben einem potentiell negativen Einfluss durch unterschiedliche Restfeuchtegehalte auf die Analytik, könnte es auch zu einer Beeinträchtigung der Herkunftsbestimmung aufgrund verschiedener Körnermaissorten kommen. Aufgrund unterschiedlicher klimatischer Verhältnisse in den Anbauländern und der Vielzahl an verfügbaren Maissorten könnte es zu einer Bevorzugung von bestimmten Körnermaissorten je nach geographischer Anbauregion kommen³³¹. Dadurch könnten bei der Untersuchung der geografischen Herkunft irrtümlich sortenbedingte Unterschiede im Körnermais-Metabolom für herkunftsbedingte Unterschiede gehalten werden.

Frank *et al.* untersuchten mittels GC-MS Unterschiede im Metabolom von GM bzw. nicht-GM Maispflanzen aus Deutschland und Südafrika. Dabei stellte sich heraus, dass der Genotyp einen weitaus geringeren Einfluss auf das Metabolom ausübt, als die geographische Herkunft³³². Insbesondere in Bezug auf die wechselseitige Beeinflussung von Genotyp und Phänotyp in Abhängigkeit von Einflüssen aus unterschiedlichen geographischen Anbauorten, wurden Studien von Baniyadi *et al.* und Tang *et al.* durchgeführt. Dabei zeigte sich, dass das Metabolom von Mais eher durch herkunftsabhängige (Umwelt-)Faktoren als durch den genetischen Sortenhintergrund beeinflusst wird^{151,152}. Baniyadi *et al.* analysierten 50 Proben der Silage und Maiskörner aus vier US- und zwei kanadischen Bundesstaaten, die jeweils fünf verschiedene Genotypen umfassten, mittels LC-MS¹⁵¹. Es wurden dabei tendenziell mehr Metabolite in der Maispflanze als in Maiskörnern detektiert. Die Metabolite waren generell eher signifikant in Abhängigkeit von der geographischen Herkunft, als vom Genotyp¹⁵¹. Diese Ergebnisse konnten später von Tang *et al.* in einer umfangreicheren Studie anhand von 480 Proben größtenteils bestätigt werden¹⁵². Dabei zeigte sich ebenso, dass mehr Metabolite bei der Analyse der Maispflanze (1341 Stück) als bei der Analyse der Maiskörner (635 Stück) detektiert werden konnten. 12-54 % der Metabolite in Maiskörnern waren abhängig von der geographischen Herkunft beeinflusst, im Gegensatz zu lediglich 3-10 % sortenspezifisch beeinflussten Metaboliten¹⁵².

Aus diesem Grund wurden insgesamt 81 Proben aus drei Landessortenversuchen des Jahres 2017 analysiert, um systematisch den Einfluss der Körnermaissorte und der geografischen Herkunft auf das Metabolitprofil zu untersuchen. Insbesondere in Bezug auf eine mögliche Störung der analytischen Bestimmung der geografischen Herkunft durch unterschiedliche Körnermaissorten.

Diese umfassten neun Sorten, die an jeweils drei Standorten in Nordrhein-Westfalen, Brandenburg und Bayern angebaut wurden.

Die Proben wurden jeweils mit den optimierten Extraktionsmethoden extrahiert. Für unpolare Metabolite wurde die adaptierte BLIGH & DYER-Methode genutzt, polare Metabolite wurden mittels eines Gemischs aus Methanol/Wasser 2:1 v/v extrahiert. Die LC-MS-Analytik wurde jeweils im positiven und negativen Ionenmodus durchgeführt.

6.2.2 Multivariate Analyse der Metabolite

In den LC-MS-Daten der 81 Körnermaisproben der Landessortenversuche konnten im positiven Ionenmodus 1347 polare bzw. 702 unpolare Features detektiert werden. Im negativen Ionenmodus war die Anzahl an detektierten Features dagegen mit 703 bzw. 646 niedriger. Somit zeigte sich insgesamt die Tendenz, dass die Anzahl an Features in der Reihenfolge polar-positiv > unpolar-positiv $\approx$ unpolar-negativ > polar-negativ in den Datensätzen abnahm.

Um zudem die Varianz der Metabolitsignale zu beurteilen, wurde eine PCA berechnet (siehe **Abschn. 9.14.1**). Anhand der PCA-Scoresplots kann geschlossen werden, dass eine Separierung des Sortenprobensets nach geographischer Herkunft erfolgte (siehe **Abb. 6.3**). Dies wurde durch die Anordnung der Proben in den PCA-Scoresplots deutlich, aus dem eine Clusterbildung der Proben entsprechend dem Herkunftsland hervorging. Die Tendenz der Proben sich nach geographischer Herkunft zu separieren, wurde dabei insbesondere für die polaren Metabolite deutlich. Dieselbe Trennung der Proben bei den unpolaren Metaboliten war im positiven Ionenmodus andeutungsweise zu erkennen. Hingegen war dies für die Messung im negativen Ionenmodus nicht mehr der Fall.

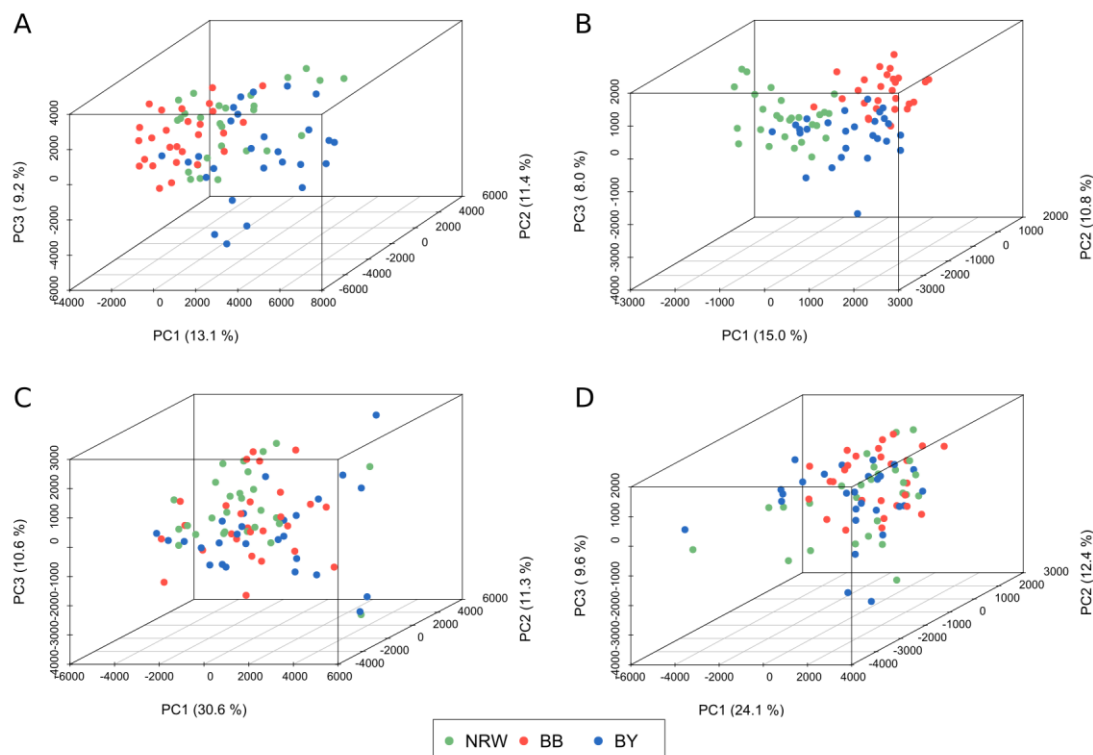


Abb. 6.3: PCA-Scoresplots (nach *Pareto-Scaling*) der Körnermaisproben. Polare Metabolite, positiver (A) bzw. negativer (B) Ionenmodus. Unpolare Metabolite, positiver (C) bzw. negativer (D) Ionenmodus. Die geographische Herkunft ist farblich gekennzeichnet, siehe Legende.

Um die Separation des Sortenprobensets auch anhand der vorhandenen gruppenspezifischen Unterschiede zu untersuchen, wurde ebenso PLS-DA-Analysen durchgeführt. Diese erfolgten jeweils nach Probenherkunft sowie Körnermaissorte. Dabei konnte ein Trend der Gruppierung der Proben nach ihrer Herkunft in Abhängigkeit von den betrachteten Metaboliten und dem Ionenmodus herausgearbeitet werden: Die Q^2 -Werte der multivariaten Datenanalyse für die geographische Herkunft, befanden sich im Bereich von 0.89 bis 0.55 und nahmen in der Reihenfolge polar-negativ $\approx$ polar-positiv > unpolar-positiv > unpolar-negativ ab. Die Unterscheidung der Sortenproben nach Körnermaissorte konnte nur über eine PLS-DA erzielt werden, während diese mittels PCA nicht möglich war. Deren Q^2 -Werte waren für polare Metabolite im positiven Ionenmodus mit 0.80 am höchsten, am niedrigsten hingegen für unpolare Metabolite im positiven Ionenmodus mit 0.39 (siehe **Anhang Abb. 13.3**). Im negativen Ionenmodus wurden Q^2 -Werte im Bereich von 0.6 bis 0.75 erzielt.

Zusammenfassend bietet die multivariate Datenanalyse mittels PCA und PLS-DA einen guten Ansatz, um die Varianz der detektierten Metabolite zu veranschaulichen. Jedoch erfolgt die Positionierung der Proben entweder entsprechend der maximalen Varianz der Proben mittels PCA^{259,333,334} bzw. der maximalen Kovarianz zwischen den Probenklassen über eine PLS-DA^{166,235,261} und nicht anhand der statistischen Signifikanz. Daher erscheint es vielversprechend, die Trennung der Proben in dem Scoresplot durch eine vorherige Auswahl an Markerfeatures zu verbessern.

6.2.3 Herkunftsbezogene Varianz vs. sortenbezogene Varianz der Metabolite

Da die 81 Proben sowohl verschiedene Herkunftsorte als auch mehrere Körnermaissorten umfassten, kann die statistische Signifikanz der Metabolite in Bezug auf die Herkunft oder die Sorte bestimmt werden. Dazu wurden alle Features einem *One-Way* ANOVA-Test (p -Wert ≤ 0.01) unterzogen um die Signifikanz der Unterschiede in den Peakintensitäten zwischen den jeweiligen Probenklassen (geographische Herkunft bzw. Maissorte) zu bestimmen. Um die zugrundeliegenden Unterschiede auf bestimmte Kombinationen aus Anbauorten/Sorten einzugrenzen, wurde zudem ein *post-Hoc* TUKEY-Test durchgeführt. Dieser erlaubt die Aufschlüsselung der ANOVA-signifikanten Marker auf konkrete Kombinationen der Probengruppen und wurde bereits in anderen Veröffentlichungen zur Identifikation von geeigneten Markersubstanzen zur Unterscheidung von Probenkategorien angewandt^{335,336}.

Da in dem Sortenprobenset neun Sorten aber nur drei Herkunftsorte enthalten sind, besteht ein Missverhältnis zwischen drei Zweier-Kombinationen der Herkunft und 36 Zweier-Kombinationen der Sorte. Da der ANOVA-Test lediglich Rückschlüsse hinsichtlich der generellen Signifikanz ermöglicht – ein signifikanter Unterschied zwischen zwei Probengruppen reicht bereits aus – ist es statistisch zwölfmal wahrscheinlicher, dass eine Signifikanz bezüglich der Sorte auftritt, als bezüglich der Herkunft. Daher wurde in **Tab. 6.3** der mittlere Prozentsatz der Metabolite, die im TUKEY-Test pro Zweier-Kombination signifikant waren, berechnet.

Tab. 6.3: Mittlere prozentuale Anteile der signifikant beeinflussten Features des Sortenprobensets. Bestimmt durch *One-Way* ANOVA p -Wert ≤ 0.01 und *post-Hoc* TUKEY-Test. Angegeben ist der mittlere Prozentsatz der Metabolite, die im TUKEY-Test pro Zweier-Kombination signifikant waren.

	Polare Metabolite		Unpolare Metabolite	
	<i>positiv</i>	<i>negativ</i>	<i>positiv</i>	<i>negativ</i>
Herkunftsabhängig	24.2 %	25.6 %	13.5 %	12.0 %
Sortenabhängig	8.5 %	7.9 %	7.8 %	9.2 %

Wie aus **Tab. 6.3** ersichtlich ist, waren die Metabolite zu einem gewissen Teil herkunftsabhängig beeinflusst; der Anteil betrug für die polaren Metabolite ca. ein Viertel und für die unpolaren Metabolite ca. ein Achtel der insgesamt detektierten Metabolite. Dagegen war der Anteil an sortenabhängigen Metaboliten geringer, er betrug lediglich ca. 8 % und war nicht abhängig von der gewählten Analysenmethode. Diese Ergebnisse sind im Einklang mit den Ergebnissen von Baniasadi *et al.*¹⁵¹ und Tang *et al.*¹⁵². Die Resultate des *post-Hoc* TUKEY-Tests der ANOVA-Ergebnisse sind in **Abb. 6.4** zudem als Heatmap aufgeschlüsselt entsprechend der geographischen Herkunft bzw. der Sorte.

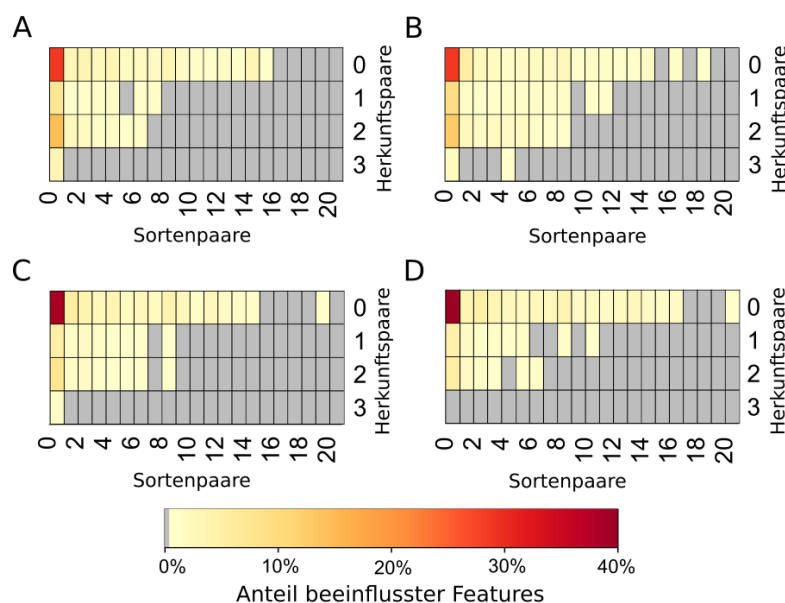


Abb. 6.4: Heatmaps der Verteilung der Signifikanzen der Features in dem Sortenprobenset. Die prozentuale Verteilung der Features in dem Probenset ist aufgeschlüsselt nach der Anzahl an signifikanten Herkunftsbzw. Sortenpaarungen in einem *post-Hoc* Tukey-Test (p -Wert ≤ 0.01). Polare Metabolite, positiver (A) bzw. negativer (B) Ionenmodus. Unpolare Metabolite, positiver (C) bzw. negativer (D) Ionenmodus. Die Anzahl ist farbcodiert, siehe Legende.

Abb. 6.4 ist zu entnehmen, dass insbesondere Metabolite, welche ein hohes Potenzial ($>$ neun Zweierpaare) für die Unterscheidung von mehreren Maissorten zugleich aufweisen, eher selten eine Signifikanz für die Unterscheidung der geographischen Herkunft aufwiesen. Metabolite, welche hingegen ein hohes Potenzial für die Unterscheidung der geographischen Herkunft besaßen, waren nur sehr selten zugleich signifikant für die Unterscheidung von vielen Maissorten. Der Anteil an Metaboliten welche gleichzeitig nach Herkunft und stark ($>$ neun Zweierpaare) nach Sorte beeinflusst waren, war dabei, mit lediglich 2.4 % und 3.0 % (polar) bzw. 0.7 % und 2.0 % (unpolar) sehr gering.

Dieses Verhalten ist dabei nicht für alle vier Messmodi gleichermaßen ausgeprägt: Es zeigte sich, dass insbesondere die polaren Metabolite, welche bei der multivariaten Datenanalyse ein gutes Clustering bezüglich der Herkunft zeigten, eher eine kombinierte Abhängigkeit nach Herkunft und Sorte aufwiesen, als die unpolaren Metabolite. Die unpolaren Metabolite zeigten hingegen sowohl ein gering ausgeprägtes Potenzial als Sortenindikator, als auch eine lediglich geringe Tendenz gleichzeitig für Herkunft und Sorte ein Unterscheidungsmerkmal zu sein.

Somit ist davon auszugehen, dass es bei der Etablierung einer analytischen Methode zur Herkunftsbestimmung, zu keinen Störeffekten durch die Körnermaissorte kommen sollte. Um dennoch einen möglicherweise gleichzeitigen Sorten- und Herkunftseinfluss auf einzelne Metabolite zu untersuchen, werden in **Abschn. 6.2.5** die Einflussfaktoren während der LC-MS-Analyse genauer analysiert. Der Sorteneinfluss wird nach Auswahl der Marker für ein Herkunftsbestimmungsmodell erneut evaluiert werden. Dazu sind Proben aus einem Sortenversuch in Spanien im Herkunftsprobenset enthalten. Über deren Klassifikationsrate bei der Verifikation der geographischen Herkunft, lassen sich Rückschlüsse auf das Ausmaß eines Sorteneinflusses ableiten.

6.2.4 Einfluss von Sorte und geographischer Herkunft auf Klassifizierungsmodelle

Um den wechselseitigen Einfluss von geographischer Herkunft und Körnermaissorte auf der Ebene der Klassifizierungsmodelle zu beleuchten und – im Gegensatz zu der multivariaten Analyse – die Ergebnisse in konkreten Zahlen auszudrücken, wurden unterschiedliche Klassifizierungsmodelle erstellt und getestet. Als Klassifikationsalgorithmus wurde hier einheitlich der RF-Algorithmus angewandt, welcher bereits von Creydt *et al.* erfolgreich zur Herkunftsbestimmung von weißem Spargel eingesetzt wurde⁴³.

Wie aus **Tab. 6.3** ersichtlich wird, waren in allen Messmodi jeweils durchschnittlich ca. 8 % der Features nach Maissorte beeinflusst und 1/4 bzw. 1/8 nach geographischer Herkunft. Dabei war ein kleinerer Prozentsatz von ≤ 3 % aller Features sowohl nach Herkunft beeinflusst als auch gleichzeitig stark nach Körnermaissorte. Die Klassifizierungsmodelle sollten die Sortenproben mittels verschiedener Subgruppen der Metabolite entsprechend ihrer geographischen Herkunft klassifizieren. Da das Sortenprobenset aus neun unterschiedlichen Maissorten besteht, die gleichmäßig über die Anbauländer verteilt sind, würde sich ein möglicher Sorteneinfluss negativ auf die Zuverlässigkeit der Bestimmung der geographischen Herkunft auswirken.

Die Metabolite des Datensatzes wurden zunächst mittels der p -Werte aus dem ANOVA-Test bzw. TUKEY-Test, bezüglich ihrer Signifikanz (p -Wert ≤ 0.01) zur Herkunfts- bzw. Sortenunterscheidung unterteilt. Diese Informationen wurden genutzt um die Auswirkungen auf die Klassifizierung bei der Nutzung folgender Marker-Subgruppen zu untersuchen: Zur Klassifikation wurden entweder alle Marker ausgewählt, welche signifikant bezüglich der Herkunft bzw. Sorte waren oder jeweils ein Subset von $n = 25$ Markern (25-fach wiederholt) zufällig selektiert. Die Klassifikation mittels RF-Algorithmus erfolgte dann entweder als 100-fach wiederholte 10-fache Kreuzvalidierung (alle Marker) oder als 10-fach wiederholte 10-fache Kreuzvalidierung (25 ausgewählte Marker). Die RF-Klassifikationsergebnisse bezüglich der Herkunft und Sorte je nach verwendeter Markersubgruppe sind in **Tab. 6.4** dargestellt.

Es ist festzustellen, dass der Prozentsatz der korrekt klassifizierten Proben erwartungsgemäß für die Bestimmung der Herkunft unabhängig von der konkret angewandten LC-MS-Methode abnahm, je weniger stringent nach Signifikanz der Herkunft (bzw. nicht-Signifikanz der Sorte) gefiltert wurde. Für die Bestimmung der Sorte gilt erwartungsgemäß ein inverser Zusammenhang. Zudem zeigte sich eine leicht schlechtere Klassifikationsgenauigkeit, falls anstatt sämtlicher detektierter Marker, lediglich ein Subset von $n = 25$ Markern betrachtet wurde. Dieses *Setting* kommt jedoch dem realen Anwendungsfall der Verifikation der geographischen Herkunft näher.

Tab. 6.4: RF-Klassifizierungsergebnisse des Sortenprobensets nach Herkunft bzw. Sorte je nach ausgewählter Subgruppe der detektierten Metabolite des Sortenprobensets (Details siehe Text).

	ANOVA-signifikant nach Herkunft (p -Wert ≤ 0.01) & <u>nicht</u> signifikant nach Sorte (p -Wert > 0.01)		ANOVA-signifikant nach Herkunft (p -Wert ≤ 0.01)		ANOVA-signifikant nach Sorte (p -Wert ≤ 0.01)		Hoch signifikant nach Sorte (Anzahl signifikanter TUKEY-Paarungen > 9)	
Alle Marker	<i>Herkunft</i>	<i>Sorte</i>	<i>Herkunft</i>	<i>Sorte</i>	<i>Herkunft</i>	<i>Sorte</i>	<i>Herkunft</i>	<i>Sorte</i>
polar-positiv	100 %	75.2 %	100 %	98.4 %	99.8 %	100 %	94.3 %	100 %
polar-negativ	100 %	80.8 %	100 %	98.1 %	99.5 %	100 %	91.5 %	99.7 %
unpolar-positiv	98.5 %	43.1 %	98.4 %	87.4 %	96.7 %	96.2 %	68.2 %	98.9 %
unpolar-negativ	95.8 %	49.0 %	97.7 %	89.5 %	95.9 %	93.7 %	71.9 %	92.4 %
n = 25 Marker	<i>Herkunft</i>	<i>Sorte</i>	<i>Herkunft</i>	<i>Sorte</i>	<i>Herkunft</i>	<i>Sorte</i>	<i>Herkunft</i>	<i>Sorte</i>
polar-positiv	98.8 %	48.5 %	97.8 %	71.6 %	85.8 %	86.6 %	79.3 %	96.4 %
polar-negativ	97.8 %	46.9 %	96.3 %	75.3 %	87.9 %	88.6 %	80.7 %	98.9 %
unpolar-positiv	94.4 %	33.1 %	92.7 %	63.0 %	78.0 %	81.0 %	61.1 %	90.7 %
unpolar-negativ	94.3 %	41.9 %	91.9 %	73.5 %	77.4 %	81.6 %	59.8 %	88.1 %

Optimale Ergebnisse wurden für die Herkunftsbestimmung mit dem Subset der polaren Metabolite mit > 95 % richtigen Zuordnungen erreicht, falls die Markermetabolite mittels ANOVA-Signifikanztest selektiert wurden. Dieses Ergebnis steht zunächst im Einklang mit den Ergebnissen aus dem vorherigen Abschnitt der multivariaten Datenanalyse, in denen die Proben ebenso für diese Messmodi am deutlichsten Cluster nach ihrer geographischen Herkunft bildeten. Die polaren Metabolite erzielten bei der Klassifikation nach Sorte jedoch gleichzeitig hohe Klassifikationsraten von 71-98 %, je nach gewähltem Subset. Dies ist insofern problematisch, als das bei der Verifikation der geographischen Herkunft von kommerziellen Maisproben, die Maissorte in der Regel nicht deklariert sein wird und somit unbekannt ist.

Eine bessere Eignung zur Herkunftsbestimmung ohne Beeinflussung durch die Maissorte, wies hingegen das Subset der unpolaren Metabolite auf. Hier konnte für die Bestimmung der geographischen Herkunft ebenfalls hohe Klassifikationsgenauigkeiten erzielt werden (> 90 %). Gleichzeitig wurden insbesondere für den positiven Ionenmodus sehr geringe Raten an korrekten Klassifikationen gemäß der Maissorte ermittelt (< 65 %). Dies galt vor allem auch, falls das Markersubset lediglich anhand der Signifikanz für die geographische Herkunft selektiert wurde. Diese Ergebnisse sind folglich besonders vielversprechend, da eine Markerselektion in Hinblick auf das Kriterium der Sorte im realen Anwendungsfall nur schlecht möglich wäre. Daher ist das Subset der unpolaren Metabolite, welche im positiven Ionenmodus gemessen werden können, als besonders geeignet für eine Verifikation der geographischen Herkunft von Körnermaisproben anzusehen. Es bietet dabei sowohl hohe Klassifikationsraten für die Herkunft als auch niedrige für die Sorten. Gleichzeitig blieben die Klassifikationsraten auch dann hoch, falls Sorteninformationen nicht in die Markerselektion mit eingebunden wurden.

6.2.5 Einfluss der Sorte und geographischer Herkunft je nach Analysemethode

Um eine Auswahl der am wenigsten von der Sorte beeinflussten LC-MS-Messmethode zu ermöglichen, wurden im Folgenden für die vier Messmethoden, diejenigen Metabolite, die signifikant für die Sorte oder die geographische Herkunft waren in zweidimensionalen *Survey-Views* abgebildet, siehe **Abb. 6.5**. Um den wechselseitigen Einfluss von Sorte und geographischer Herkunft auf die einzelnen Metabolite zu verdeutlichen, wurden darin die Metabolite je nach Anzahl der signifikanten Sorten- und Herkunfts-Zweierpaare im *post-Hoc* Tukey-Test ($p \leq 0.01$) farblich unterschiedlich markiert.

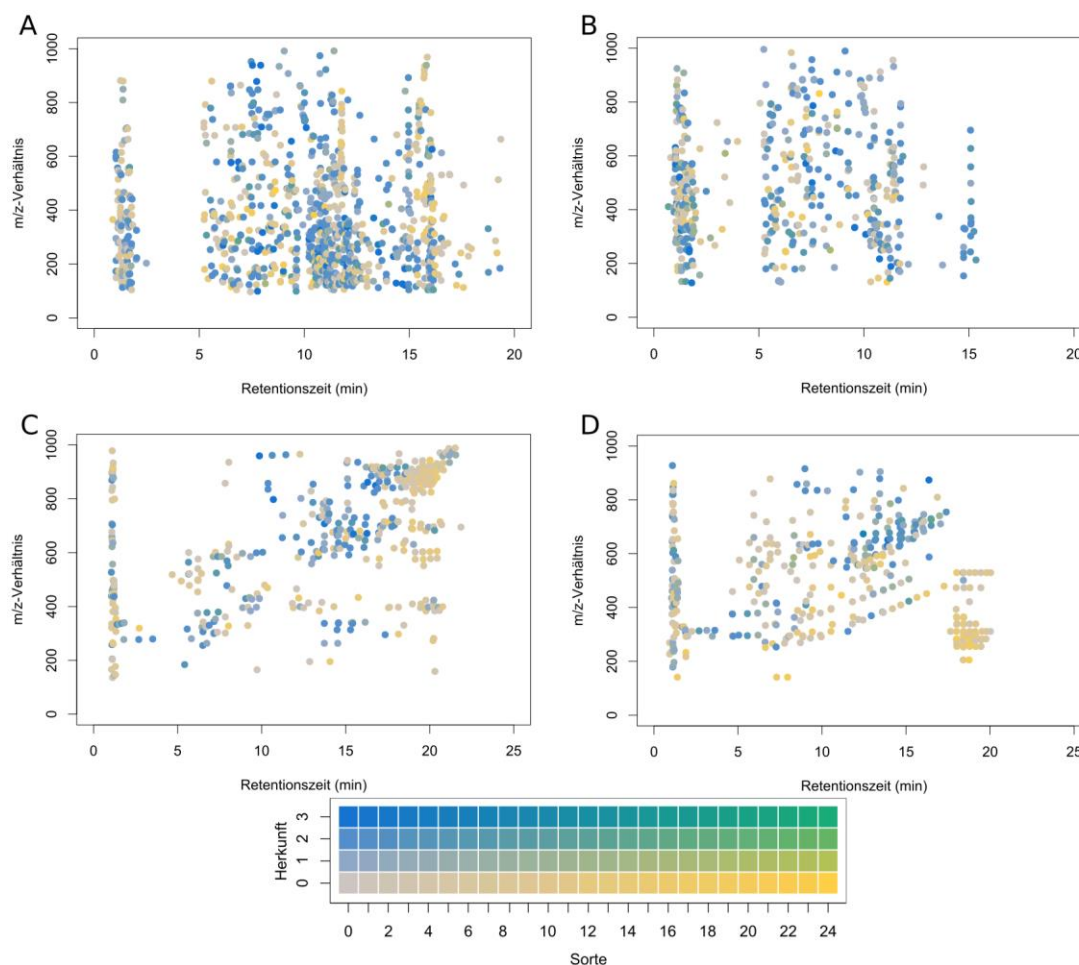


Abb. 6.5: Abstrahierte *Survey-Views* der herkunfts/sorten-bezogen beeinflussten Features je nach verwendeter LC-MS-Methode. Polare Metabolite, positiver (A) bzw. negativer (B) Ionenmodus. Unpolare Metabolite, positiver (C) bzw. negativer (D) Ionenmodus. Die Farbe zeigt an, inwieweit die Features entsprechend der Herkunft bzw. der Sorte signifikant (*post-Hoc* Tukey-Test p -Wert ≤ 0.01) unterschiedlich waren. Skala von grau bis blau: Anzahl unterscheidbarer Kombinationen von zwei Herkunftsländern (0-3); Skala von grau bis gelb: Anzahl unterscheidbarer Kombinationen von zwei Maissorten (0-36).

Wie aus **Abb. 6.5** ersichtlich ist, war das Ausmaß der Überlappung von sorten- bzw. herkunftsabhängig beeinflussten Peaks abhängig von der Polarität der analysierten Metaboliten und dem verwendeten Ionenmodus. Somit lässt sich ein gewisser störender Einfluss der Maissorte zunächst durch die Wahl der Analysenmethode bedeutend verringern. Während im Falle der polaren Methode herkunfts- und sortenabhängig beeinflusste Metabolite unabhängig von der

Retentionszeit oder des m/z -Verhältnisses auftraten (siehe **Abb. 6.5AB**), war bei der unpolaren Analysenmethode eine Korrelation mit diesen Eigenschaften zu beobachten (siehe **Abb. 6.5CD**). Die herkunftsabhängig beeinflussten Metabolite waren dabei eher bei einem m/z -Verhältnis > 500 Da zu finden, die sortenabhängig beeinflussten Features hingegen eher bei kleineren m/z -Verhältnissen < 500 Da. Sortenabhängige Metabolite eluierten eher bei Retentionszeiten < 2 min sowie > 18 min, während herkunftsabhängige Features meist eine Retentionszeit < 18 min aufwiesen. Somit zeigte sich, dass die Problematik einer Störung der Herkunftsbestimmung durch die Auswahl der geeignetsten Subfraktion aller analysierten Metaboliten, abgemildert werden kann.

Die unpolaren Metabolite, welche im positiven Ionenmodus detektiert werden konnten, sind hauptsächlich der Substanzklasse der Lipide zuzuordnen. Da freie Fettsäuren nicht im positiven Ionenmodus, sondern bevorzugt im negativen Ionenmodus ionisieren³¹⁸, beginnt der relevante m/z -Bereich erst bei > 450 Da, ab dem Lyso-Phospholipide auftreten^{337,338}. Somit würde eine Fokussierung auf Lipide, den Vorteil bieten, dass eine Beeinflussung der Herkunftsbestimmung durch die Körnermaissorte als unwahrscheinlicher als bei der Nutzung anderer Metabolitarten zu bewerten ist. Zudem würde die große Diversität der Lipidspezies, bei gleichzeitiger relativer Ähnlichkeit der Substanzgruppen einen großen Pool an potenziellen Markersubstanzen darstellen. Dieser könnte vergleichsweise einfach analysiert und auch ausgewertet werden. Unter den Lipiden konnten zudem keine hoch-sortenspezifischen Lipidmarker nachgewiesen werden, welche gleichzeitig für die Unterscheidung der geographischen Herkunft der Proben geeignet waren.

Jedoch könnte es in einem ungünstigen Fall dazu kommen, dass eine Sorte die vorher für eine bestimmte Region typisch war, dort nicht mehr angebaut wird und stattdessen in einer anderen Region. Dies könnte die Herkunftsbestimmung auf die Probe stellen, da die Körnermaissorte im Handel nicht ausgewiesen wird^{38,39,66,67,69–72}. Da der Sorteneinfluss in Relation zu dem herkunftsbezogenen Einfluss jedoch eher gering ist^{151,152}, ist der Umfang einer möglichen Störung der analytischen Herkunftsbestimmung per se eher unwahrscheinlich. Zudem können (negative) Auswirkungen auf das Herkunftsbestimmungsmodell dadurch abgemildert werden, indem tendenziell weniger sortenspezifisch anfällige Metabolite genutzt werden.

6.3 Analyse der Herkunftsproben mittels LC-MS

Um sowohl leistungsstarke als auch verlässliche Marker mit guter Differenzierungsmöglichkeit der Anbauländer zu extrahieren, wurden vielversprechende Markerfeatures zunächst per PCA-Loadingsplot identifiziert. Geeignete Peaks wurden integriert und auf Grundlage der Peakflächen eine *One-Way* ANOVA-Signifikanzanalyse sowie ein post-Hoc Tukey-Test durchgeführt. Zudem wurde die RSD der entsprechenden Marker in jeweils zehn QC-Proben bestimmt.

Die Klassifikation der Proben erfolgte dann anhand der selektierten Markermetabolite mittels RF-Algorithmus. Um eine unvoreingenommene Bewertung der Leistungsfähigkeit der Modelle vornehmen zu können, wurde die Klassifikation als 10-fache Kreuzvalidierung durchgeführt^{292,339}. Die Klassifikation wurde jeweils 100-fach wiederholt („*repeated cross-validation*“) um einen repräsentativen Durchschnitt der Klassifizierungsgenauigkeit abzubilden und die Standardabweichung zu bestimmen.

Außerdem wurde das entworfene Klassifizierungsmodell auf weitere Proben aus anderen Ländern, die nicht zu den Herkunftsorten, die für die Erstellung des Modells zählten, angewandt, um die Performance des Modells dahingehend zu studieren.

6.3.1 Multivariate Analyse der Metabolite in den Herkunftsproben

Die LC-MS-basierte Analyse der Körnermais-Herkunftsproben erfolgte mittels der vier bereits in **Abschn. 6.1** erläuterten Methoden für polare bzw. unpolare Metabolite. Anschließend wurden die Rohdaten kalibriert und ein *Feature-Finding* durchgeführt, wobei der Injektionspeak hier nicht exkludiert wurde (Retentionszeitbereich 0-20 min für polare und 0-23 min für unpolare Metabolite). Aus allen Features der Proben des Herkunft-Probensets wurde jeweils eine *Consensus*-Featureliste erstellt. Auf diese Weise wurden insgesamt vier Listen erhalten (siehe **Tab. 6.5**). Es wurden dabei nur jene Features berücksichtigt, welche in mindestens der Hälfte aller Proben detektiert werden konnten. Die Anzahl an nachgewiesenen Features je Probe sind in **Tab. 6.5** für die vier unterschiedlichen Analysevarianten zusammengestellt. Zudem ist vermerkt, wie viele der Metabolite signifikant bezüglich der Probenherkunft beeinflusst sind (ANOVA-Test p -Wert ≤ 0.01).

Tab. 6.5: Anzahl detektierter Features je nach genutzter LC-MS-Methode. Aufgeschlüsselt nach der Gesamtzahl an Features des Herkunftsprobensets, sowie Anzahl der Features, welche signifikant bezüglich der Probenherkunft (ANOVA-Test p -Wert ≤ 0.01) beeinflusst sind.

	LC-MS-Methode			
	Polar-positiv	Polar-negativ	Unpolar-positiv	Unpolar-negativ
Anzahl detektierter Features	775	708	1175	995
Davon ANOVA-signifikant für Herkunft	482 (62.2 %)	560 (79.1 %)	994 (84.6 %)	848 (85.2 %)

Wie aus **Tab. 6.5** hervorgeht, bestätigten sich auch hier die Ergebnisse, wie sie bereits bei der Extraktionsmethodenoptimierung und der Untersuchung der Sorteneinflüsse erhalten wurden. Es konnten mehr Metabolite im positiven Ionenmodus als im negativen Ionenmodus detektiert werden. Der prozentuale Anteil an Features, welche signifikant bezüglich der Probenherkunft beeinflusst sind, nahm in der Reihenfolge der Analysenmethoden unpolar-negativ $\approx$ unpolar-positiv > polar-negativ > polar-positiv ab. Die Varianz der Analysedaten je nach Messmethode ist in **Abb. 6.6** mittels PCA verdeutlicht.

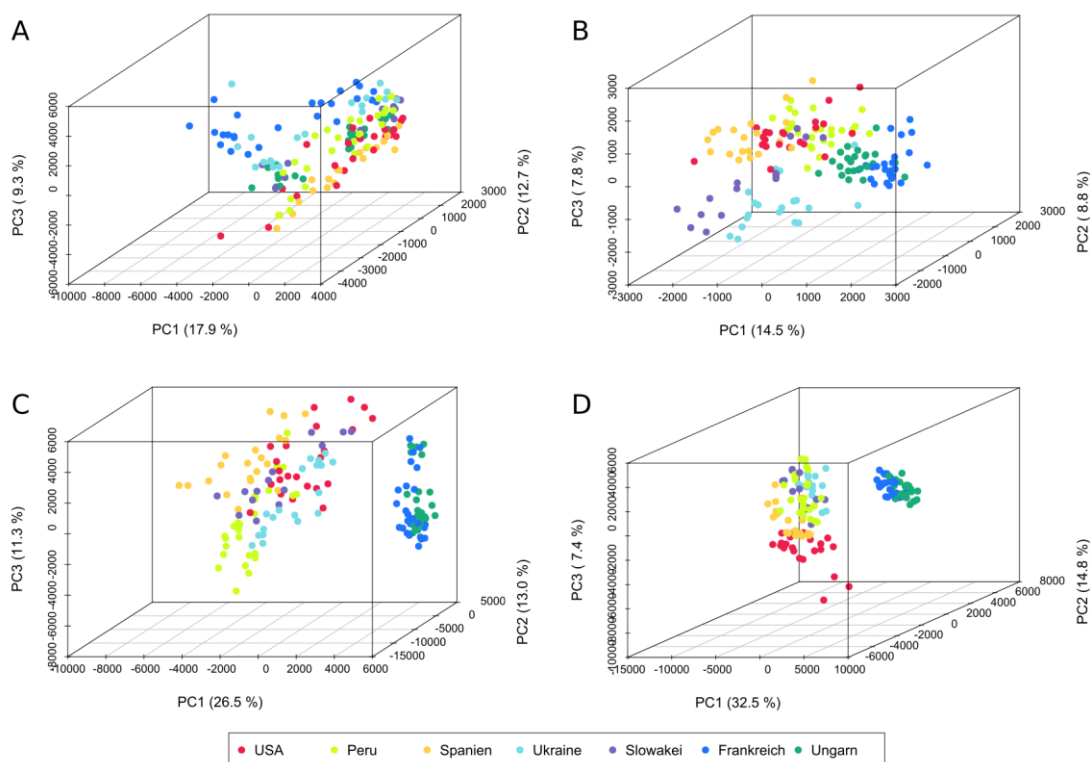


Abb. 6.6: PCA-Scoresplot der Metabolite je nach LC-MS-Methode. Polare Metabolite, positiver (A) bzw. negativer (B) Ionenmodus. Unpolare Metabolite, positiver (C) bzw. negativer (D) Ionenmodus. Die Probenherkunft ist entsprechend der Legende farbig indiziert.

Wie aus **Abb. 6.6** ersichtlich, war die Varianz in dem Probenstet hauptsächlich auf die unterschiedliche geographische Herkunft der Körnermaisproben zurückzuführen. Die Proben separierten gut in einzelne Cluster, die den unterschiedlichen Herkunftsländern entsprachen, wie durch die farbige Markierung verdeutlicht wird. Bei näherer Betrachtung zeigte sich eine Abstufung der Güte der Trennung nach der Probenherkunft für die Proben aus Frankreich/Ungarn, welche insbesondere im Falle der unpolaren Metabolite als ein Cluster abseits von den restlichen Proben, auftraten.

Wie aus den Loadingsplots (**Anhang Abb. 13.4**) hervorgeht, zeigte sich, dass die gute Separierung bei der Analytik der polaren Metabolite lediglich durch einige, wenige Features verursacht wurde. Diese wiesen zudem eine sehr geringe Retentionszeit auf, hauptsächlich eluierten die Metabolite während des Injektionspeaks der LC-MS-Analyse. Aus den Histogrammen der Featureintensitäten kann zudem entnommen werden, dass deren Intensitätenverteilung bei der Analytik des polaren Metaboloms überwiegend Features niedriger Intensität beinhaltete. Dagegen wiesen die detektierten unpolaren Metabolite eine ausgewogenere Verteilung der Features auf der Intensitätsskala auf, siehe **Anhang Abb. 13.5**.

Ein Abgleich der m/z -Werte und der Fragmentspektren der bedeutendsten Marker der Analyse der polaren Metabolite mit Vergleichsdaten aus den *Human Metabolome Database* (HMDB)³³⁷ und der LipidMaps-Datenbank³³⁸ ergab zudem, dass es sich bei den detektierten Metaboliten mit geringer Retentionszeit um unpolare Metabolite, genauer Lipide handelte. Beispiele sind im **Anhang Tab. 13.10** dargestellt. Diese unpolaren Metabolite können jedoch besser über Lipid-

Extraktionsmethoden und folgender Auftrennung auf RP-Chromatographiesäulen analysiert werden³¹⁸. Daher wurde der Fokus weiterer Analysen auf die unpolaren Extrakte und LC-MS-Methoden gelegt.

6.3.2 Einfluss der Messmethode auf die Klassifikation

Neben der Bestimmung für eine Herkunftsbestimmung relevanter Features, stellt sich außerdem die Frage nach der am besten geeigneten Klassifizierungsmethode. Die Klassifikation der Proben anhand der LC-MS-Messdaten erfolgte in dieser Arbeit aufgrund der überzeugenden Ergebnisse von Creydt *et al.* mit dem RF-Algorithmus. Die optimale Anzahl genutzter Bäume (*ntree*) wurde in Übereinstimmung mit Creydt *et al.*⁴³ auf 1000 festgelegt.

Die Klassifizierungsgenauigkeiten sind in **Tab. 6.6** in Abhängigkeit von der Analyse-methode gegenübergestellt. Da diese mit von der Anzahl der genutzten Markerfeatures abhängen, sind die Klassifizierungsergebnisse als Resultat aller ANOVA-signifikanten Features der Messmethode aufgeführt. Zudem wurden ebenso 25-fach, je $n = 25$ Marker aus den ANOVA-signifikanten Features ausgewählt und jeweils einer 10-fach wiederholten 10-fach Kreuzvalidierung unterzogen. Mittels dieses Ansatzes kann ein eventueller Einfluss der unterschiedlichen Anzahl der signifikanten Marker je Messmethode, vermieden werden.

Tab. 6.6: Erreichte RF-Klassifikationsgenauigkeiten mit den Analysedaten der vier LC-MS-Methoden des Körnermais-Herkunftsprobensets. Als Marker dienten entweder alle für die Herkunft signifikanten Metabolite (ANOVA p -Wert ≤ 0.01), oder $n = 25$ zufällig daraus ausgewählte Metabolite. Die Klassifikation erfolgte mittels einer 100- bzw. 10-fach wiederholten, 10-fachen Kreuzvalidierung (Standardabweichung in Klammern).

Messmethode	Alle ANOVA-signifikanten Markerfeatures	$n = 25$ zufällig ausgewählte ANOVA-signifikante Markerfeatures
Polar-positiv	97.7 % (0.6 %)	85.1 % (4.8 %)
Polar-negativ	99.1 % (0.3 %)	88.8 % (3.3 %)
Unpolar-positiv	98.5 % (0.3 %)	92.1 % (2.4 %)
Unpolar-negativ	99.5 % (0.7 %)	93.3 % (2.6 %)

In Bezug auf die genutzte LC-MS-Methode bzw. Metabolitfraktion konnten bei der Nutzung aller jeweils signifikanten Markerfeatures keine bedeutsamen Unterschiede zwischen den vier Messmethoden festgestellt werden. Die Klassifikationsraten befanden sich im Bereich von 97.7 bis 99.5 %, die Klassifikation anhand des Subsets der polar-positiven Metabolite erzielte die geringste Genauigkeit bei in Relation hoher Varianz der Ergebnisse (97.7 ± 0.6 %). Für die Klassifikation mittels 25 zufällig ausgewählten signifikanten Markermetabolite, konnte hingegen ein Abfall der Klassifikationsergebnisse in der Reihenfolge unpolar-negativ > unpolar-positiv > polar-negativ > polar-positiv ermittelt werden.

Im Hinblick auf die Unterschiede zwischen den beiden Ionenmodi, wiesen unpolare Metabolite geringere Unterschiede als die polare Metabolite auf, siehe **Tab. 6.5**. Im Hinblick auf den Ionenmodus ist zusätzlich festzustellen, dass im positiven Ionenmodus absolut jeweils mehr Features, als im negativen Ionenmodus detektiert werden konnten.

Daher wurde unter Berücksichtigung aller hier dargelegter Faktoren, die Klassifikation von Körnermaisproben entsprechend ihrer geographischen Herkunft, mit den unpolaren Metaboliten, welche im positiven Ionenmodus detektiert werden konnten, angegangen. Diese Methode versprach gleichzeitig hohe Klassifikationsgenauigkeiten und eine vorteilhafte Intensitätenverteilung der Features. Zudem zeigte sich in **Abschn. 6.2.4**, dass das unpolare Metabolom insbesondere im positiven Ionenmodus und bei der Auswahl von herkunftsbezogenen signifikanten Features, besonders unempfindlich gegenüber Sorteneinflüssen ist. Diese Metabolitfraktion war außerdem bereits zuvor von Klockmann *et al.* zur Herkunftsbestimmung von Haselnüssen genutzt worden⁴⁴. Auch Creydt *et al.* fanden u. a. geeignete unpolare Metabolite für die Bestimmung der geographischen Herkunft von weißem Spargel⁴³.

6.4 Bestimmung der geographischen Herkunft von Körnermaisproben anhand des Profils der unpolaren Metabolite im positiven Ionenmodus¹

Die analytische Bestimmung der geographischen Herkunft von Körnermais sollte entsprechend der Erkenntnisse aus den vorherigen Abschnitten (siehe **Abschn. 6.1** und **6.2**) über die Analyse des unpolaren Metaboloms, analysiert mittels LC-MS im positiven Ionenmodus erfolgen. Ein repräsentativer *Survey-View* der QC-Probe ist im **Anhang** in **Abb. 13.6A** gezeigt, zusammen mit Chromatogrammen der QC-Probe (**Abb. 13.6B**) sowie einer repräsentativen Herkunftssprobe je Herkunftsland (**Abb. 13.6C-I**).

Die von der unpolaren BLIGH & DYER-Extraktionsmethode erfassten Metabolite beinhalteten überwiegend die Stoffgruppe der Lipide. Im Hinblick auf die einzelnen Lipidklassen wurden hauptsächlich Phospholipide (Phosphatidylcholine (PC) und Phosphatidylethanolamine (PE)), sowie Acylglyceride (DGs und TGs) in dem unpolaren Extrakt mittels LC-MS-Analytik detektiert. DGs und Phospholipide eluierten beide in dem Retentionszeitbereich von 10.5 - 18.5 min. DGs wiesen ein m/z -Verhältnis von m/z 580 - 750 auf, während Phospholipide eher bei höheren m/z -Werten von 700 - 850 beobachtet werden konnten. TGs eluierten separat bei einer Retentionszeit > 17.5 min. DGs und TGs konnten als Ammoniumaddukt detektiert werden ($[M+NH_4]^+$), aber auch als Natriumaddukte ($[M+Na]^+$), gemäß Han³⁴⁰. Die Phospholipide waren weniger intensiv und konnten hauptsächlich als $[M+H]^+$ Addukte detektiert werden. PCs traten daneben auch als Natriumaddukte auf³⁴¹.

Die Anzahl an detektierten Features war unabhängig von der Probenherkunft. Folglich kann vornehmlich von quantitativen Unterschieden, als von qualitativen Unterschieden ausgegangen werden. Indem alle Intensitäten der Features ermittelt wurden, welche in mindestens 50 % aller Proben detektiert werden konnten und die ab einer Retentionszeit von 2 min eluierten, konnte eine *Consensus*-Tabelle erstellt werden. Diese enthielt insgesamt 1022 Features.

¹ Die in den folgenden Abschnitten (6.4 – 6.4.4; Seite 63 bis 72) dargelegten Ergebnisse wurden bereits zuvor weitestgehend publiziert in der Veröffentlichung:

D. Schütz, E. Achten, M. Creydt, Riedl, M. Fischer

Non-Targeted LC-MS Metabolomics Approach towards an Authentication of the Geographical Origin of Grain Maize (*Zea mays* L.) Samples

Foods **2021**, *10*, 2160.

Um Metabolite, welche zur Unterscheidung der Proben geeignet sind, zu identifizieren, wurde der Datensatz nach einem *Pareto-Scaling* einer PCA-Analyse unterzogen. Der PCA-Scoresplot sowie die RF-Klassifikationsmatrix sind in **Abb. 6.7** gezeigt.

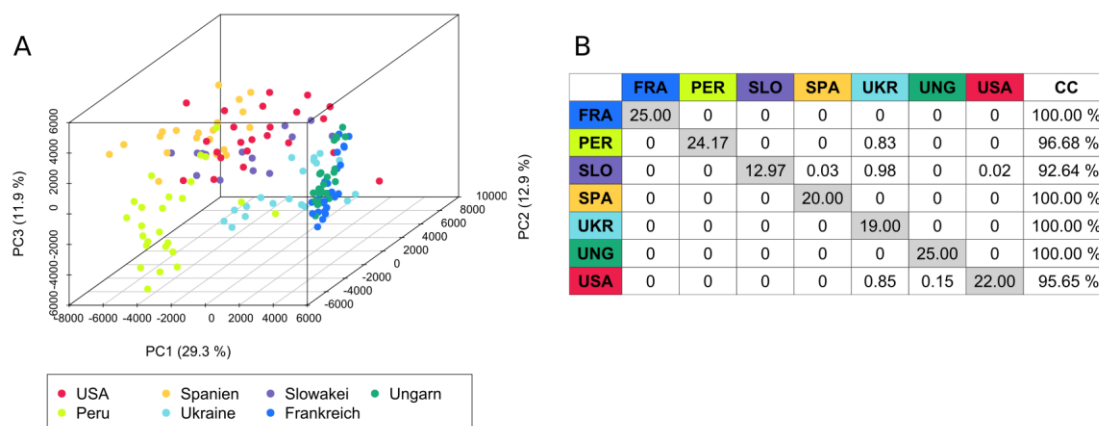


Abb. 6.7: PCA-Scoresplot und RF-Klassifikationsmatrix erstellt mittels der *Consensus-Feature* Tabelle der im positiven Ionenmodus detektierten Metabolite (nach *Pareto-Scaling*). **A:** Scoresplot; **B:** Klassifikationsmatrix. Die RF-Klassifikation ist 100-fach wiederholt und jeweils 10-fach kreuzvalidiert. Abbildung modifiziert nach Schütz *et al.*³⁴².

Wie in **Abb. 6.7A** ersichtlich, konnte eine zufriedenstellende Separation der Proben nach deren Herkunftsorten mittels der im positiven Ionenmodus analysierten unpolaren Metabolite erreicht werden. Dies untermauert die Hypothese, dass die in den LC-MS-Analysedaten der Herkunftproben enthaltene Varianz auf die unterschiedliche geographische Herkunft der Proben zurückzuführen ist. Die Summe der erklärten Varianz war moderat. Es waren sechs Hauptkomponenten notwendig, um mehr als 95 % der in dem Datensatz enthaltenen Varianz abzudecken. Dennoch ordneten sich die Proben in Clustern entsprechend ihrer Herkunft an. Die Proben aus den USA, Slowakei und Spanien liegen relativ dicht im Scoresplot zusammen. Die Nähe der Proben aus den USA zu einem Teil der europäischen Proben ist überraschend. Da die Proben aus den USA jedoch aus verschiedenen US-Bundesstaaten stammten, die in der Nähe des Golfs von Mexiko liegen, ist dieses Verhalten aufgrund des teilweise ähnlichen Klimas nachvollziehbar. Ein zweites Probencluster im Scoresplot bilden Proben aus Frankreich, Ungarn sowie der Ukraine, was in Übereinstimmung mit deren Herkunft aus Zentraleuropa ist. Proben aus Peru, bilden ein separates Cluster, was im Hinblick auf die große geographische Entfernung zu den Ursprungsorten der anderen Proben schlüssig erscheint.

Der Loadingsplot offenbarte allerdings, dass die Features unter den Top 20 % der absolut größten Loadings entweder während des Injektionspeaks eluierten, oder es sich um *in-source* Fragmente $[\text{Acyl}+74]^+$ von DGs oder TGs³⁴³ handelte.

Der Abgleich der m/z -Werte der weiteren einflussreichen Features mit den Werten aus der HMDB Datenbank ($\Delta m/z \leq 5$ ppm) ergab, dass es sich hauptsächlich um Lipide handelte. Dies ist im Einklang mit der Verwendung der BLIGH & DYER Extraktionsmethode, welche zur Extraktion von Lipiden entwickelt wurde³²⁰. Die Verteilung der Features auf die einzelnen Lipidunterklassen, ergab als Reihenfolge: TG >> DG > PC > PE. Neben diesen Lipidklassen, konnten einige wenige Monoglyceride, sowie freie Fettsäuren und niedermolekulare Substanzen detektiert werden.

Markersubstanzen, die zur analytischen Verifikation der geographischen Herkunft dienen sollen, müssen zugleich verlässlich und reproduzierbar sein und zudem statistisch signifikante Unterschiede zwischen den Herkunftsländern aufweisen. Es stellte sich heraus, dass Features, welche eine auffällige Positionierung im PCA-Loadingsplot abseits des Koordinatenursprungs aufwiesen, teilweise *in-source* Fragmente waren. Um diese nicht irrtümlich als potente Markerfeatures für die Herkunftsbestimmung auszuwählen, erschien es sinnvoll als zusätzliches Kriterium die Signifikanz in einem ANOVA-Test nach Herkunft (p -Wert ≤ 0.01) zu beachten. Dies sollte sowohl zu einer Verbesserung der Trennung im PCA-Scoresplot führen als auch sich positiv auf die Ergebnisse des Klassifizierungsmodells auswirken.

6.4.1 Einfluss der Lipidspezies auf die Eignung zur Klassifikation nach der geographischen Herkunft

Um vorab einen zuverlässigeren Überblick über den Zusammenhang zwischen Lipidspezies und dem Herkunftsort zu erhalten, wurden die Rohdaten zudem erneut mittels einer anderen Herangehensweise ausgewertet. Aufgrund der hohen Abundanz von Lipiden aus der Klasse der Phospholipide, sowie der Acylglyceride, wurde der Datensatz von LC-MS-Analysen nach Peaks mit zu Lipiden passenden m/z -Verhältnissen ($\Delta m/z \leq 5$ ppm) untersucht.

Übereinstimmende Peaks wurden integriert und die jeweils 20 den Lipidklassen der Phospholipide (genauer: PCs und PEs), DGs und TGs zuzuordnenden Peaks bestimmt, die sowohl die größten Peakflächen in dem Datensatz als auch einen ANOVA p -Wert ≤ 0.01 für die geographische Herkunft aufwiesen. So konnten insgesamt 60 Features ermittelt werden, die sich zu je 20 Stück auf die Lipidklassen der Phospholipide (genauer: PCs oder PEs), DGs und TGs verteilen.

Die allgemeine Reproduzierbarkeit dieser Markerfeatures ist dabei als gut anzusehen: 48 der insgesamt betrachteten 60 Marker wiesen eine RSD von ≤ 15 % in der QC-Probe auf. Die RSDs sind zusammen mit weiteren statistischen Kennzahlen im **Anhang Tab. 13.11** aufgeführt. Ein Abgleich der Ergebnisse der statistischen Signifikanz mit jenen der multivariaten Datenanalyse ergab weiterhin, dass sieben Phospholipide, acht DGs sowie 17 der 20 TGs nicht nur statistisch signifikant waren, sondern auch in den Top 20 % Features, mit dem größten Einfluss auf die ersten drei Komponenten der PCA, vorkamen.

Im Rahmen einer 100-fach wiederholten 10-fach kreuzvalidierten Klassifikation mittels des RF-Algorithmus, konnte eine Abnahme der Klassifizierungsgenauigkeit von 84.1 % nach 76.4 % in der Reihenfolge: Phospholipide > DGs > TGs beobachtet werden (für die detaillierten Klassifikationsmatrizen, siehe **Anhang Abb. 13.7**). Dabei zeigten sich insbesondere folgende Auffälligkeiten: Mittels Phospholipiden waren die Proben aus Frankreich und Ungarn nur schwer zu trennen, was jedoch mittels der TGs leicht gelang. Die Differenzierung zwischen Proben aus den USA und der Slowakei war für jede der betrachteten drei Lipidgruppen eine Herausforderung: Im Falle der Phospholipide konnte zwar keine systematische Fehlklassifikation beobachtet werden, jedoch war die Fehlerrate allgemein sehr hoch. Die Rate an korrekten Klassifikationen (CC, engl. *correct classifications*) betrug nur 54 % für slowakische Proben und 78 % für Proben aus den USA. Für DGs war der Trend ähnlich, jedoch verringerten sich die CC insbesondere für slowakische Proben auf nur 32 %. Das Modell ordnete zudem 5 von 14 slowakischen Proben fälschlich als aus den USA zu. Hier trat somit eine Schwäche des Modells zu Tage, dass die Trennung von Proben aus der Slowakei und den USA nicht mehr zuverlässig erfolgte. Im Falle der TGs, zeigten sich weitere

Schwächen: Zusätzlich war hier die Unterscheidung zwischen Proben aus den USA und Spanien eine Herausforderung. Weiterhin wurden ukrainische Proben bei Nutzung der TGs ebenfalls häufig fehlerhaft anderen Klassen zugeordnet (CC 75 %).

Gründe für die unterschiedliche Eignung der Lipidspezies zur Herkunftsbestimmung von Körnermais könnten auf deren unterschiedliche Rolle im Stoffwechsel der Pflanzen zurückzuführen sein. TGs fungieren als Speicherlipide^{344,345}, während DGs Zwischenprodukte in der Lipidbiosynthese darstellen³⁴⁶. Im Gegensatz dazu sind Phospholipide Bestandteil der Zellmembran^{346,347}, deren Regulation mit Trockenstress assoziiert ist oder allgemeiner gefasst mit der Anpassung an klimatische Umweltbedingungen^{345,346,348}. Daher erscheint eine höhere Variabilität der Phospholipide als Reaktion auf Umwelteinflüsse, in Relation zu den beiden anderen Lipidklassen denkbar.

6.4.2 Identifizierung von relevanten Markermetaboliten zur Klassifikation nach der geographischen Herkunft

Um eine weitere Nutzung des entwickelten Klassifikationsmodells, auch in einem produktionsorientierten *Setting* abseits der universitären Forschung zu ermöglichen, ist eine Reduktion der genutzten Markersubstanzen notwendig. Die *Downstream*-Analytik erfolgt in einem industriellen *Setting* oftmals mittels Triple-Quadrupol-Massenspektrometer, die im MRM-Modus betrieben werden^{6,43,44,123}. Zusätzlich sind kurze Analysendauern anzustreben, wobei gleichzeitig die Anzahl an parallel analysierten Substanzen wiederum nicht zu hoch sein sollte, um Ionensuppressionseffekten vorzubeugen^{119,349}. Um das Modell Routine-tauglicher zu machen, sollte die Anzahl an verwendeten Markersubstanzen möglichst auf ≤ 25 reduziert werden.

Eine Vorabanalyse des Einflusses der Marker aus einem Pool der Top 20 abundantesten Lipid-Marker der Phospholipide, DGs und TGs zeigte, dass auf Grundlage von 20 Markern noch eine hohe Klassifikationsgenauigkeit von $> 85\%$ erreicht werden konnte. Die erzielte Klassifikationsgenauigkeit in Abhängigkeit von der Anzahl an genutzten Markern ist im **Anhang in Abb. 13.8** dargestellt. Die entspr. Anzahl an Markern wurde dazu 25-mal jeweils zufällig aus dem Pool der Top 20 abundantesten Lipid-Marker ausgewählt und mittels 10-mal wiederholter 10-facher Kreuzvalidierung klassifiziert.

Die Klassifikationsgenauigkeit zeigte bei steigender Anzahl an genutzten Markern einen asymptotischen Verlauf gegen 91% (der erzielten Genauigkeit bei Nutzung aller 60 betrachteten Marker). Während eine Reduktion auf nur noch 15 Marker keinen nennenswerten Verlust an Klassifikationsgenauigkeit zur Folge hatte, fiel diese bei einer weiteren Reduktion der Markeranzahl jedoch relativ stark ab. Bei der wiederholten Selektion von 20 Markern betrug die Genauigkeit im Mittel immer noch 88.4% , somit konnten ähnliche Genauigkeiten wie bei Nutzung aller 60 Markersubstanzen ($91.1 \pm 1.0\%$) erreicht werden. Eine Reduktion der Markeranzahl auf eine Anzahl, welche in einer Routinemethode noch gut nutzbar ist, ist somit unter Bewahrung einer ähnlich hohen Klassifikationsgenauigkeit möglich.

Die Selektion von potenziell vielversprechenden Markern erfolgte über den Loadingsplot der PCA. Qualitativ hochwertige Markersubstanzen zur Herkunftsbestimmung von Körnermaisproben sollten zudem gleichzeitig möglichst reproduzierbar und signifikant für die Probenherkunft sein. Deshalb mussten akzeptable Marker folgende Kriterien erfüllen:

- 1) Bei einem ANOVA-Test musste der p -Wert ≤ 0.01 sein.
- 2) Eine möglichst hohe Anzahl (> 7 von 21 insgesamt möglichen Zweipaarungen) im *post-Hoc* TUKEY-Test, die signifikant sind.
- 3) Für eine gute Reproduzierbarkeit sollte eine RSD von $\leq 15\%$ in $n = 10$ QC-Proben erreicht werden.

Um ein möglichst ausgeglichenes Modell für die Unterscheidung aller sieben in dem Herkunftsprobensatz vorhandenen Länder zu erhalten, wurde zusätzlich darauf geachtet, dass alle Marker jedes Herkunftsland von mindestens einem weiteren Land signifikant unterscheiden konnte.

Zusätzlich wurde im Hinblick einer besseren späteren Übertragbarkeit des Modells in die Routineanalytik auf folgende Aspekte geachtet: Die gewählten Markersubstanzen sollen relativ abundant in den Proben vorliegen, um eine quantitative Bestimmung zu erleichtern. Daher gehörten alle selektierten 20 Markermetabolite auch zu den Top 20 intensivsten Markern der jeweiligen Lipidklasse. Zudem sollten die Markersubstanzen relativ divers über unterschiedliche Substanzgruppen verteilt sein, um dem Modell auch bei der routinemäßigen Anwendung eine gewisse Robustheit mitzugeben. Gleichzeitig sollten die Substanzen nicht zu nahe beieinander eluieren, um die Erfassung in einem einzigen Analyselauf zu ermöglichen.

Anhand dieser Merkmale konnten insgesamt 20 potente Marker für die Herkunftsbestimmung von Körnermaisproben selektiert werden. Die Peakflächen dieser Markersubstanzen sind im **Anhang in Abb. 13.9** entsprechend ihrer geographischen Herkunft als Boxplots dargestellt. Diese waren dabei offensichtlich je nach Herkunftsland unterschiedlich abundant. Es konnten keine systematischen Unterschiede aller Marker bei einem einzigen Herkunftsland beobachtet werden. Manche Lipidklassen wiesen in Körnermais aus bestimmten Ländern Auffälligkeiten auf: Bspw. traten geringere Peakflächen für die Gruppe der Phospholipide in den französischen und ungarischen Proben, sowie den meisten Markern aus der Gruppe der DGs auf. Diese Körnermaisproben waren nicht für die Verwendung als Futtermittel produziert worden, sondern sollten der Herstellung von Maisstärke dienen. Zudem wurden diese Proben in gefrorenem Zustand anhand eines leicht abweichenden Protokolls mithilfe einer Messermühle unter Zusatz von Trockeneis im Verhältnis 1:1 homogenisiert. Da eine Vermahlung mit Zusatz von Trockeneis als besonders Metabolit-schonend literaturbekannt ist^{123,255}, ist zu vermuten, dass die geringe DG-Konzentration mit dem Verwendungszweck der Stärkeproduktion assoziiert ist. Es wurden hohe Konzentrationen von Phospholipiden in Proben aus Spanien und Ungarn detektiert; außerdem von DGs in spanischen Proben, sowie TGs in peruanischen Proben. Mögliche Gründe für die unterschiedliche Regulation der Lipidklassen könnten sein, dass TGs Speicherlipide sind, während DGs eine Rolle als Zwischenprodukte im Lipidstoffwechsel darstellen³⁴⁴. Im Gegensatz dazu, sind Phospholipide Membranlipide, deren Regulation üblicherweise mit Umwelteinflüssen, wie z. B. Temperatur, Bodenzusammensetzung, Trockenheit oder Pathogene^{344,348} in Verbindung gebracht wird. Diese Umwelteinflüsse können v. a. aufgrund des geographischen Anbauortes Schwankungen ausgesetzt sein, weshalb eine höhere Variation der Phospholipide im Herkunftsprobensatz im Vergleich zu den anderen Lipidklassen vorstellbar ist.

Variationen in dem Fettsäurespektrum in Abhängigkeit von der geographischen Herkunft wurden ebenso von Tres *et al.* mittels GC in DDGS aus Mais bzw. Weizen zu deren Unterscheidung

genutzt¹⁴⁵. Im Hinblick auf die Authentizität des geographischen Ursprungs anderer pflanzlicher Produkte anhand des mittels LC-MS analysierten Lipidprofils, untersuchten Lim *et al.* sowie Rubert *et al.* Reis³⁵⁰ sowie die Tigernuss³⁵¹. Dabei erwiesen sich vor allem Vertreter der Klasse der Phospholipide als gut geeignete Markermetabolite. Phospholipide, DGs und TGs wurden auch von Klockmann *et al.* zur Unterscheidung des geographischen Ursprungs von Haselnüssen genutzt^{44,46}. Zudem erwiesen sich in einer Studie zur geographischen Authentizität von chinesischem Schweinefleisch hauptsächlich Acylglyceride und Phospholipide als potentielle Marker-substanzen¹³⁵.

6.4.3 Multivariate Datenanalyse und Klassifikation unter Nutzung aller 20 identifizierten Markermetabolite

Um optimale Klassifikationsergebnisse bei einer gleichzeitig möglichst niedrigen Anzahl von Features zu erzielen, ist eine Auswahl möglichst hoch potenter Features notwendig. Um eine höhere Robustheit des Modells beizubehalten, ist es zudem wünschenswert, dass die Marker unterschiedlichen Substanzklassen angehören. Daher wurden für das optimierte Modell Features, welche allen vier Lipidklassen (PCs, PEs, DGs sowie TGs) zugehörig waren, genutzt. Diese wiesen sowohl große Peakflächen, als auch eine hohe Signifikanz im ANOVA-Test und im *post-Hoc* TUKEY-Test auf. Um eine stabile Klassifikationsrate oberhalb von 90 % zu erreichen, war die Verwendung von mindestens 20 Markersubstanzen aus den jeweiligen Top 20 intensivsten Markern je Lipidklasse notwendig, siehe **Anhang Abb. 13.8**.

Bei der Verwendung von 20 ausgewählten Markersubstanzen konnte eine Gesamtklassifikationsrate von 90.5 % erreicht werden. Die selektierten Markersubstanzen wurden im Folgenden identifiziert (siehe **Abschn. 6.4.4**) und sind im **Anhang Tab. 13.12** aufgelistet. Eine Reduktion auf 20 Markersubstanzen führte – aufgrund der stringenten Auswahl der Features aus allen vier Lipidklassen – lediglich zu einem leicht verminderten Genauigkeit der Klassifikation im Vergleich zu allen 1022 Markern der Gesamtfeatureliste (98.1 %).

Ein *post-Hoc* TUKEY-Test ergab, dass die Fähigkeit zwischen Zwei-Länder-Kombinationen zu unterscheiden, für die gewählten Markermetabolite variierte: Da insgesamt sieben Ursprungsländer in dem Herkunftsprobenset enthalten waren, konnte idealerweise jedes Land von sechs anderen Ländern abgetrennt werden. Die maximal erreichte Anzahl an signifikanten Unterschieden zwischen zwei Ländern waren 16 von 21 theoretisch möglichen Kombinationen; dies wurde von vier Markern erreicht. Die durchschnittliche Anzahl an unterscheidbaren Zwei-Länder-Kombinationen betrug zehn.

Auf Länderebene variierte die Anzahl der Features, die eine bestimmte Probenherkunft von allen anderen statistisch signifikant (TUKEY-Test p -Wert ≤ 0.01) abtrennen konnte von 1 bis 20, je nach betrachtetem Herkunftsland. Während französische, peruanische und ungarische Proben relativ viele signifikant unterschiedliche Metabolite für ihre Unterscheidung aufwiesen; waren slowakische Körnermaisproben nur durch relativ wenige Metabolite von den anderen Proben zu differenzieren. Die Tendenz der Ergebnisse der statistischen Signifikanz spiegelt sich ebenfalls in den Ergebnissen der RF-Klassifikation wider.

Die Separation der Proben im PCA-Scoresplot anhand der 20 selektierten Metabolite änderte sich nicht maßgeblich (siehe **Abb. 6.8**) im Vergleich zu den Ergebnissen unter Einbeziehung aller

1022 Features (vgl. **Abb. 6.7**). Die Varianz, welche durch die ersten drei Hauptkomponenten erklärt wurde, betrug 96 %.

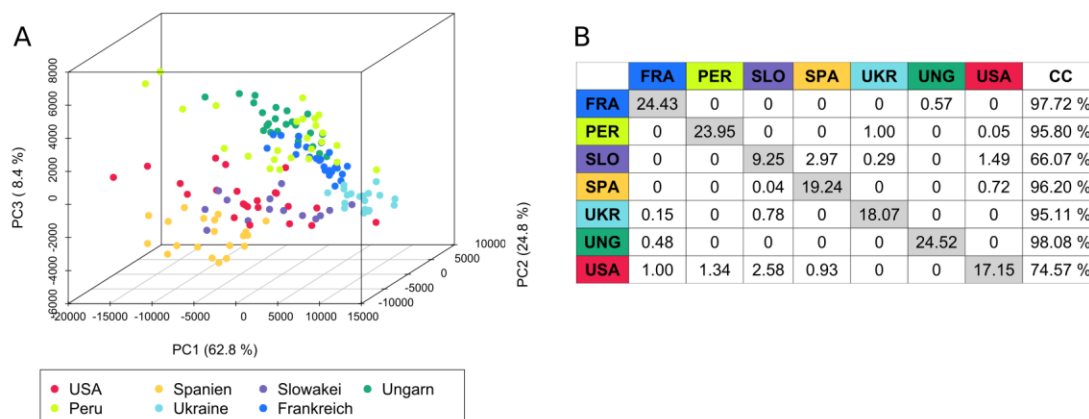


Abb. 6.8: PCA-Scoresplot und RF-Klassifikationsmatrix unter Nutzung der 20 ausgewählten Markersubstanzen (nach *Pareto-Scaling*). **A:** Scoresplot; **B:** Klassifikationsmatrix. Die RF-Klassifikation ist 100-fach wiederholt und jeweils 10-fach kreuzvalidiert. Abbildung modifiziert nach Schütz *et al.*³⁴².

Auf Länderebene verbesserte sich die Trennung der zuvor dicht beieinander gelegenen Proben aus Frankreich und Ungarn. Die Größe der Punktwolken der Proben aus der Ukraine sowie Spanien konnten ebenfalls verringert werden. Die Proben aus Peru veränderten ihre Position im Scoresplot und liegen nun näher bei den Proben aus Frankreich/Ungarn. Die Trennung von Proben aus den USA und der Slowakei blieb eine Herausforderung, was sich auch in den Klassifikationsergebnissen widerspiegelte.

Die Klassifikationsergebnisse auf Länderebene spiegelten im Wesentlichen die Erkenntnisse aus dem PCA-Scoresplot wieder: Proben aus der Slowakei und den USA sind ebenso für das RF-Modell eher schwierig zu unterscheiden. Weiterhin war auch die Separation von spanischen Proben herausfordernd, da einige der slowakischen Proben irrtümlich als spanische Proben klassifiziert wurden. Für die USA-Proben konnte ebenso eine größere Diversität bei der Klassifikation beobachtet werden. Diese größere Varianz konnte durch die Verteilung der Proben auf sechs US-Bundesstaaten, welche hauptsächlich an den Golf von Mexiko grenzen, erklärt werden. Insgesamt decken die USA-Proben eine Ost-West-Distanz von ca. 1400 km ab. Verglichen mit der Fläche auf denen die anderen Proben, insbesondere aus Spanien, der Slowakei aber auch Peru gesammelt wurden, waren die Proben aus den USA geographisch weiter verteilt. Das Auftreten von Schwankungen in den Metabolitkonzentrationen erfolgt jedoch nicht aufgrund von politischen Staatsgrenzen, sondern aufgrund von Veränderungen von klimatischen Bedingungen, sowie weiteren umweltbedingten Faktoren¹¹⁹. Somit ist eine Unterscheidung von zwei Probenklassen bei größerer geographischer Entfernung als wahrscheinlicher anzunehmen, als bei nur geringer Distanz der Anbauorte.

Die Varianz der Peakflächen der Markerlipide in Proben aus Sortenversuchen von demselben Standort in Lleida (Spanien), war nicht auffällig erhöht. Im PCA-Loadingsplot (siehe **Abb. 6.8A**) zeigte sich, dass die Varianz der spanischen Proben in Relation zu bspw. den Proben aus den USA bzw. Peru gering ausfiel. Außerdem war die Erfolgsrate der Klassifikation der spanischen Proben

mit 19 von 20 korrekt zugeordneten Proben sehr gut. Dies zeigte, dass der heterogene genetische Hintergrund der Proben, deren Klassifikation nicht negativ beeinflusste.

Generell reagieren Pflanzen auf veränderte klimatische Bedingungen, über die Modulation des Ausmaßes an ungesättigten Fettsäuren^{345,348,352}. In einem spezielleren Fall, wurde gezeigt, dass Blätter der Maispflanze auf Temperaturstress durch die Regulation von Lipiden reagieren³⁴⁶. Tang *et al.* verdeutlichten ebenfalls mittels LC-MS-basierter Analytik, dass umwelt-assoziierte Faktoren des geographischen Anbauorts einen größeren Effekt auf Futter- und Körnermais Metaboliten ausüben, als der Genotyp der Pflanze¹⁵².

Die Eignung des Markersets wurde weitergehend analysiert, indem binäre *One-vs.-Rest* Modelle berechnet wurden. Anhand dieser Klassifikationsmodelle kann jeweils ein Herkunftsland von allen anderen Ländern in dem Körnermaisherkunftsprobenset abgegrenzt werden. Dabei muss das Modell jedoch nicht wie in den vorherigen Modellen zwischen sieben Herkunftsorten entscheiden, sondern nur noch zwischen zwei Probenklassen: Entweder stammt die Probe aus einem bestimmten Land, bspw. Spanien, oder sie stammt aus einem der anderen Länder Peru, USA, Ukraine, Slowakei, Frankreich oder Ungarn; jedoch nicht weiter differenziert. Mittels dieser Modelle ließen sich Klassifikationsgenauigkeiten von > 89 % erreichen (siehe **Tab. 6.7**).

Tab. 6.7: Klassifikationsergebnisse von binären RF Modellen (ein Land vs. die restlichen Länder, 100-wiederholt, 10-fach kreuzvalidiert). Es wurden lediglich die 20 ausgewählten Markersubstanzen nach *Pareto-Scaling* genutzt. Tabelle nach Schütz *et al.*³⁴².

Modell	Klassifikationsgenauigkeit
Frankreich vs. Rest	98.9 ± 0.6 %
Peru vs. Rest	97.0 ± 0.4 %
Slowakei vs. Rest	91.5 ± 0.6 %
Spanien vs. Rest	96.1 ± 0.6 %
Ukraine vs. Rest	98.5 ± 0.3 %
Ungarn vs. Rest	99.2 ± 0.3 %
USA vs. Rest	89.3 ± 1.0 %

Die Separierung der ungarischen Proben von allen anderen Proben aus dem Herkunftsprobenset war am wenigsten fehleranfällig mit über 99 % Genauigkeit. Die Abtrennung der USA-Proben schnitt hingegen am schlechtesten mit lediglich 89 % Genauigkeit ab. Die erzielte Klassifikationsgenauigkeit war auch hier geringer für Proben aus der Slowakei und den USA. Insofern zeigt sich eine Übereinstimmung mit dem Klassifizierungsmodell auf Länderebene (siehe **Abb. 6.8B**). Die Ähnlichkeit der Proben aus den USA und der Slowakei konnte außerdem bereits anhand des PCA-Scoresplots vermutet werden (siehe **Abb. 6.8A**).

Anhand der Rangfolge der Variablenwichtigkeit konnte festgestellt werden, dass je nach zu separierendem Herkunftsland, jeweils unterschiedliche Marker besonders potent waren (siehe **Anhang Tab. 13.13**). Jeder einzelne Marker war dabei mindestens einmal unter den Top 5 Markersubstanzen gelistet, mit Ausnahme von TG (58:2), dessen bestes Ranking ein 7. Platz bei der Abtrennung der spanischen Proben darstellte. Hingegen waren die Substanzen PE (34:1) und PC (36:5) am besten geeignet (Median: 6.ter Platz). Die Substanzen TG (58:2) und TG (60:3)

erwiesen sich als am schlechtesten (Median: 17.ter Platz). Die Markersubstanz PC (34:1), trat im Durchschnitt am höchsten in der Rangfolge auf.

Bei Betrachtung der Lipidklasse konnten ebenso gewisse Charakteristiken in der Rangfolge der Wichtigkeit der Marker beobachtet werden: Phospholipide waren besonders zuverlässig, um Proben aus der Ukraine oder den USA von dem Rest des Datensatzes zu separieren. DGs waren potente Marker zur Abtrennung der ungarischen Proben. Im Falle von Proben aus Frankreich oder Spanien waren es ebenso DGs, die hauptsächlich zur Unterscheidung beitrugen. TGs waren insbesondere geeignet Proben aus Peru zu unterscheiden. Proben aus der Slowakei konnten am besten durch eine Mischung aus Phospholipiden und TGs, jedoch auch einigen DGs von anderen Probenherkünften identifiziert werden.

Diese Ergebnisse verdeutlichen, dass insbesondere die Nutzung einer Kombination von Markersubstanzen aus unterschiedlichen Lipidklassen, einen Vorteil für eine exakte Klassifikation darstellt.

6.4.4 Strukturaufklärung der optimalen Markermetabolite

Eine weitergehende Strukturaufklärung der Markermetabolite erfolgte anhand von MS/MS-Spektren. Aus den enthaltenen Fragmentierungsmustern wurde die Struktur der Metabolite hergeleitet. Die Fragmentionen wurden entweder mit denen aus Referenzspektren in einschlägigen Datenbanken (HMDB oder LipidMaps) verglichen oder strukturelle Informationen wurden direkt aus dem Fragmentierungsverhalten abgeleitet. Das Fragmentierungsmuster erlaubt generelle Rückschlüsse auf die jeweilige Struktur der Lipide. Einerseits können Informationen über die Kopfgruppe erhalten werden. Aber auch Informationen über enthaltenen Fettsäuren, wie die Länge der Acylkette sowie die Anzahl der Doppelbindungen. PCs wurden anhand eines sehr abundanten Peaks im MS²-Spektrum mit einem m/z -Verhältnis von 184 identifiziert, was auf die Bildung eines Phosphocholin-Ions $[C_5H_{15}NO_4P]^+$ aus der polaren Kopfgruppe zurückzuführen ist³⁴¹. Auf dieselbe Art und Weise wurden PEs anhand eines Fragmentpeaks $[M+H-141]^{+341}$, äquivalent zu dem Verlust der Phosphatidylethanolamin-Kopfgruppe ($C_2H_8NO_4P$) identifiziert. Die Anwesenheit von DGs oder TGs wurde durch die intensiven Fragmentionen, die durch den Verlust von einer Fettsäurekette und den Neutralverlust von Ammoniak $[M+NH_4-(RCOOH+NH_3)]^+$ bzw. den Neutralverlust von lediglich Ammoniak $[M+NH_4-17]^+$ entstehen, nachgewiesen³⁴⁰.

Die Substanzen, welche in dem optimierten Modell genutzt wurden, sind im **Anhang Tab. 13.12** zusammen mit spezifischen MS²-Fragmenten, welche zur Identifikation führten, aufgeführt. Von den 20 selektierten Metaboliten gehören 14 zu der Klasse der Acylglyceride, neun TGs und fünf DGs. Daneben wurden sechs Phospholipide identifiziert, jeweils drei PCs bzw. PEs.

Im **Anhang Abb. 13.9** sind die Peakflächen der selektierten Markerfeatures je Herkunftsland in Boxplots dargestellt. Die Peakflächen unterschieden sich dabei zwischen den einzelnen Herkunftsländern. Es konnte jedoch keine systematische Regulation aller Markersubstanzen in einem Herkunftsland beobachtet werden. Bemerkenswert waren jedoch folgende Effekte: In den ukrainischen Proben waren alle betrachteten Phospholipide in nur geringer Menge vorhanden; ebenso wie die betrachteten DGs in französischen und ungarischen Proben. Dagegen wurden relativ hohe Mengen an Phospholipiden in Proben aus Spanien sowie Ungarn detektiert, sowie von DGs in spanischen und TGs in peruanischen Proben. Proben aus den USA, sowie der Slowakei

wiesen eher ein „gemittelttes Metabolitprofil“ auf, während die Proben aus den anderen Ländern in Relation eher zu höheren oder niedrigeren Metabolitkonzentrationen neigten. Dies könnte auch einen Grund darstellen, weshalb Proben aus diesen beiden Ländern teilweise falsch klassifiziert worden sind.

Eine weitere Analyse der in dieser Arbeit verwendeten Proben fand nachgelagert durch Schmauder *et al.* mittels DART-MS statt³⁵³. Dabei wurden Körnermaisproben sowohl von dem Herkunftsprobenset als auch dem Sortenprobenset (geographische Herkunft: Deutschland) nach dem Verfahren von BLIGH & DYER extrahiert und anschließend analysiert. Die RF-Klassifikation der Proben anhand der Analysedaten nach geographischer Herkunft ergab eine Rate an CC von 84.4 %, was vergleichbar war mit den in dieser Arbeit erzielten Ergebnissen. Dabei konnte insbes. für die Sortenproben aus Deutschland eine CC von 92.7 % erzielt werden, was die gute Eignung des unpolaren Metaboloms zur Herkunftsbestimmung unterstreicht. Besonders potente Marker zur Herkunftsbestimmung gehörten der Klasse der Mono-, Di- und Triglyceride an. Auch in dieser Arbeit wurden Diglyceride und Triglyceride als potente Markersubstanzen identifiziert.

6.5 Lagerstabilität der Lipidextrakte

Um die Validität und Stabilität der ausgewählten Metabolite zu überprüfen, wird in diesem Abschnitt die Lagerfähigkeit der Probenextrakte unter verschiedenen Lagerungsbedingungen untersucht: die Extrakte wurden bei definierten Temperaturen (-21 °C, +4 °C und Raumtemperatur) für unterschiedlich lange Zeiten (24 h, 3 Tage, 7 Tage) gelagert und daraufhin mit der LC-MS-Methode für unpolare Metabolite analysiert.

Bei visuellem Vergleich der Messergebnisse zeigten sich keine gravierenden Veränderungen, falls die Lagertemperatur ≤ 4 °C betrug. War die Temperatur höher, so konnte bei fortschreitender Lagerungsdauer beobachtet werden, dass generell die Peakintensitäten im BPC anstiegen. Dies ist vermutlich auf die langsame Verdunstung des Extraktionsmittels bei einer längeren Lagerungsdauer von mehreren Tagen bei Raumtemperatur zurückzuführen, trotz dichtschießender Lagerungsgefäße.

Die Veränderungen der Peakflächen der 20 Markermetabolite (welche in **Abschn. 6.4.2** als herkunfts-spezifisch identifiziert wurden), sind in der Heatmap in **Abb. 6.9** für unterschiedliche Lagerungstemperaturen und -dauern dargestellt. Die Peakflächen der Analyse der direkt vor der Messung erstellten frischen Extrakte wurden dabei als Referenz definiert. Es zeigte sich, dass sowohl die Lagerungstemperatur als auch die Lagerungsdauer einen lediglich geringen Einfluss auf das Lipidprofil der Körnermais-Extrakte ausübten.

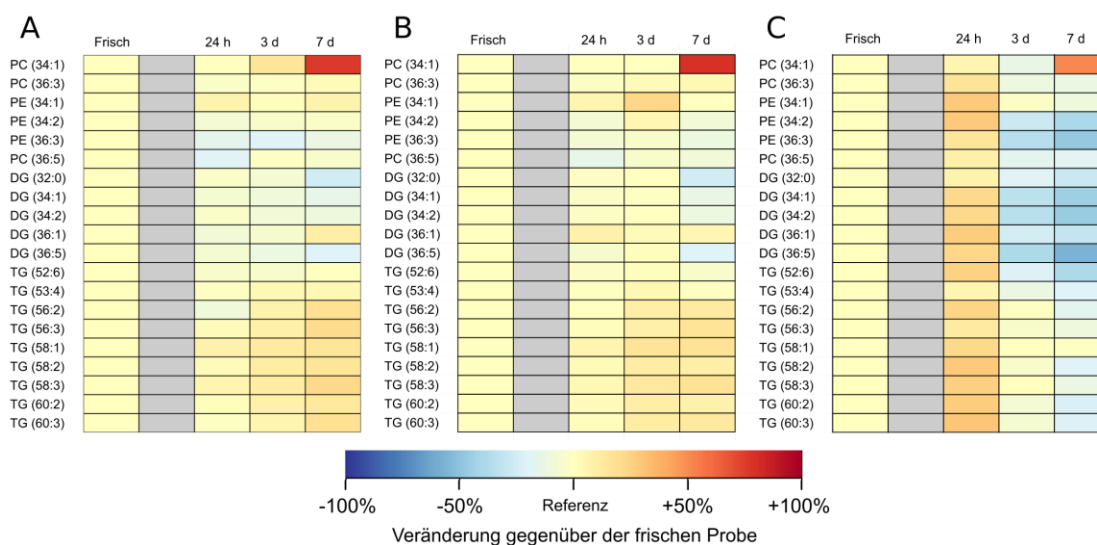


Abb. 6.9: Heatmaps der Peakflächen je nach Lagerungsbedingungen. Dargestellt sind die mittleren Peakflächen der 20 Markermetabolite aus BLIGH & DYER-Extrakten der QC-Probe in Abhängigkeit von der Lagerungsdauer (Spalten) bzw. der Lagerungstemperatur (**A**: -21 °C; **B**: +4 °C; **C**: Raumtemperatur, n = 3-5).

Im Vergleich hatte die Lagerungstemperatur der Extrakte, einen größeren Einfluss auf die Messergebnisse als die Lagerungsdauer. Für die einzelnen Metabolite zeigten sich im Verlauf der Lagerung schwankende Signalintensitäten innerhalb des Beobachtungszeitraumes von 7 Tagen, insbesondere bei einer Lagerung bei Raumtemperatur. Bei identischer Lagerungsdauer wiederum hatte die Lagerungstemperatur einen starken Einfluss auf die Peakflächen der Metabolite: Selbst bei einer Lagerungsdauer von 7 Tagen waren die betrachteten Peakflächen bei -21 °C sowie bei +4 °C nur geringfügig verändert, im Mittel um -25 % bzw. +23 %, ohne PC (34:1), siehe unten. Hingegen konnten während bei einer Lagerung bei Raumtemperatur nach bereits 3 Tagen starke Veränderungen in Relation zu frisch extrahierten Proben festgestellt werden (siehe **Abb. 6.9C**).

Bei den gekühlten Lagerungsweisen (-21 °C oder +4 °C) zeigte sich jedoch, dass eine Lagerung bis zu 3 Tage lediglich mit einer tolerablen Schwankung der Markerpeakflächen einherging. Die größte Peakflächenabweichung nach unten betrug -21 % in Relation zu dem frischen Extrakt, die größte Abweichung nach oben +24 %.

Bei Raumtemperatur nahmen die betrachteten Peakflächen nach 24 h zunächst etwas zu (durchschnittlich +21 %), dann jedoch nach 3 bzw. 7 Tagen im Mittel um -16 % bzw. -23 % ab. Während die initiale Zunahme vermutlich auf die Verdunstung des leicht flüchtigen Lösungsmittels zurückzuführen ist, ist die darauf folgende Abnahme wahrscheinlich auf Umwandlungs- bzw. Zersetzungsprozesse der Metabolite zurückzuführen.

Im Vergleich der unterschiedlichen Metabolitklassen zeigte sich, dass insbesondere die Phospholipide/DGs am empfindlichsten auf eine Lagerung reagierten. Bei diesen Lipiden betrug die Abweichung von dem Ursprungszustand des frischen Extraktes nach 7 Tagen bei +4 °C im Mittel -23 %. Unter denselben Bedingungen zeigten die TGs dagegen nur eine Abweichung von durchschnittlich -6 %.

Insbesondere hervorstechend war dabei die Markersubstanz PC (34:1) welche entgegen des Trends nach 7 Tagen stark erhöhte Werte aufwies. Danach zeigte sich jedoch, dass auch dieses

Lipid bei längerer Lagerung geringere Konzentrationen in Übereinstimmung mit dem generellen Trend aufwies. Somit ist anzunehmen, dass vermutlich ein *Carry-Over* von PC (34:1) von vorherigen Messungen stattgefunden hat.

Dagegen waren bei einer Lagerungsdauer von 24 h bei +4 °C lediglich Schwankungen im Bereich von -14.8 % bis +8.7 % beobachtbar. Daher ist eine gekühlte Lagerung bei max. +4 °C und nicht länger als 24 h anzustreben, da es ansonsten zu Abbauprozessen kommt – insbesondere der Phospholipide, was eine Verschlechterung der Vorhersage der geographischen Herkunft mittels eines Klassifikationsmodells zur Folge haben könnte.

Eine mittels PCA durchgeführte Beurteilung der Varianz anhand der 20 Markerfeatures ergab eine Zusammengruppierung der gelagerten QC-Proben sowohl entsprechend ihrer Lagerungsdauer als auch ihrer Lagerungstemperatur, siehe **Abb. 6.10**. Um die Relevanz dieser Änderungen besser einschätzen zu können, sind in **Abb. 6.10** zusätzlich die Herkunftspuben der Länder Ukraine und Spanien als frische Referenz berücksichtigt. Dabei ist ersichtlich, dass die Lagerungsproben sich entsprechend ihrer Lagerungsbedingungen auftrennten. Diese Unterschiede sind jedoch auf einer Relevanzskala für die Herkunftsbestimmung nicht bedeutsam. Vielmehr bleiben, die auf eine – teilweise extreme Lagerungsarten – zurückzuführenden Differenzen der Proben auf einen Bereich beschränkt, welcher bereits durch die biologische Variation der Herkunftspuben übertroffen wurde.

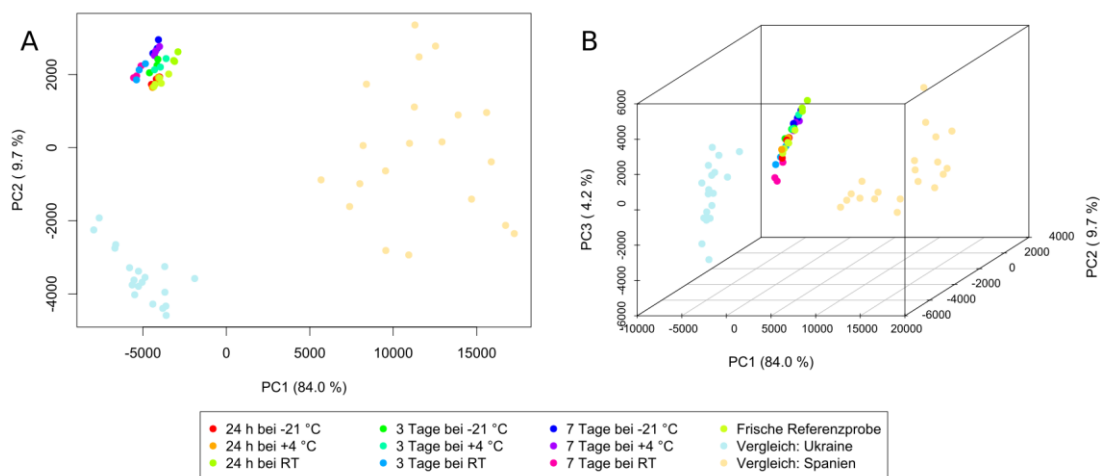


Abb. 6.10: PCA-Scoresplot nach *Pareto-Scaling* der Lagerungsproben. Dargestellt sind die Lagerungsproben (farbig), sowie als Vergleich Herkunftspuben aus den Ländern Ukraine bzw. Spanien (hellblau bzw. hellgelb), welche in einem anderen Batch, direkt nach der Extraktion analysiert wurden.

Zusammenfassend sind aus den Lagerungsversuchen folgende Erkenntnisse für eine verlässliche Analyse des Körnermaismetaboloms abzuleiten: Eine Lagerung bei ≤ 4 °C und eine möglichst kurze Lagerungsdauer erwiesen sich dabei als am geeignetsten, da es ansonsten zu einer Veränderung des Metabolitmusters kommen könnte. Dies ist vermutlich auf die teilweise Verflüchtigung des organischen Lösungsmittels zurückzuführen und auf Abbau- und Umwandlungsprozesse im Extrakt. Zudem zeigte sich bei länger dauernder Lagerung ab 3 Tagen, eine negative Beeinflussung der in der Probe enthaltenen Phospholipide, welche stark an Signalintensitäten verloren.

6.6 Etablierung einer MRM-Methode für die Quantifizierung von Lipidspezies in Körnermais

Um die Funktionsfähigkeit der Herkunftsbestimmung mittels der gewählten Marker bei einem möglichen Transfer der Methode auf ein Triple-Quadrupol-Massenspektrometer zu validieren, wurde ebenfalls die Zuverlässigkeit der Marker auf Ebene der MS²Fragmentationen untersucht.

Die MRM-Übergänge für alle 20 Metabolite sind im **Anhang** in **Tab. 13.14** tabelliert. Allgemein zeigte sich, dass bei der Nutzung von MRM-Fragmenten die absolute Höhe der Signale zurückging, dafür verbesserte sich das Signal-zu-Rausch-Verhältnis deutlich. Somit konnte trotz geringerer Peakflächen eine niedrigere Nachweisgrenze (NWG) erreicht werden, als bei Verwendung von Peakflächen auf MS¹-Ebene. Zudem wurde die Spezifität der Analysemethode stark erhöht, indem eng beieinander eluierende Lipidspezies mit gleichem m/z -Verhältnis (z. B. PC (33:4) und PE (36:4)) zuverlässiger voneinander getrennt wurden, als bei alleiniger Betrachtung des m/z -Verhältnis auf MS¹-Ebene. Die Eignung der Detektion der Metabolite auf MS²-Ebene zur Herkunftsbestimmung von Körnermais wird im Folgenden nun anhand einer repräsentativen Auswahl von neun Markersubstanzen gezeigt. Die weitere Beschränkung auf wenige, ausgewählte Metabolite, welche hingegen in einer kürzeren Analysendauer quantitativ bestimmt werden, demonstriert zudem die Eignung eines solchen methodischen Ansatzes für die Anwendung in der Routineanalytik.

Zunächst wurde die HPLC-Methode bezüglich ihrer Laufzeit optimiert. Dazu wurde gegenüber der *non-targeted*-LC-MS-Methode, die Säulenofentemperatur von 30 °C auf 40 °C erhöht. Die geringfügig schlechtere Trennleistung bei höherer Temperatur, wurde zugunsten kleinerer Retentionszeiten in Kauf genommen. Somit konnte die Analysendauer von 27 min auf 16 min reduziert werden.

Aufgrund des Mangels an kommerziell erhältlichen Standardsubstanzen, kam im Falle der Lipidmarker das Problem der korrekten Quantifizierung auf. Da jede Analysensubstanz je nach Konzentration eine leicht unterschiedliche Detektorempfindlichkeit aufweist, ist es zur exakten Bestimmung der Konzentration eines Analyten unerlässlich, eine möglichst ähnliche Standardsubstanz zu verwenden. Im Idealfall kommen isopenmarkierte Standardsubstanzen zum Einsatz³¹⁸, welche dem Analyten entsprechen, jedoch teilweise andere Isotope enthalten. Dies verändert deren m/z -Verhältnis, während das Retentionsverhalten bei der Chromatographie ähnlich bleibt. Somit ist eine interne Kalibrierung möglich. Alternativ können auch die Analysensubstanzen selbst, als externer Standard verwendet werden. Um die notwendigen Standardsubstanzen für die ausgewählten Markermetabolite zu erhalten, wäre vorher jedoch eine aufwendige Isolierung der Substanzen mittels präparativer HPLC³⁵⁴ oder anderer Verfahren, wie der *Counter Current* Chromatographie (CCC)³⁵⁵ notwendig gewesen.

Daher erfolgte die Quantifizierung mittels einer externen Kalibriergeraden anhand von vier repräsentativen Lipiden aller relevanten Substanzklassen (PC, PE, DG und TG). Die Herangehensweise der Quantifizierung von Lipiden bei *Lipidomics*-Experimenten mittels jeweils eines repräsentativen Standards je Lipidklasse wurde bereits durch Fauland *et al.* praktiziert³⁵⁶.

Die Validierung der LC-MRM-Methode zur quantitativen Bestimmung von Lipiden in Körnermais für den Herkunftsnachweis erfolgte durch Bestimmung verschiedener Kennzahlen der Methode. Es wurde sowohl eine Kalibrierung mit Standardsubstanzen in Reinform (Grundkalibrierung) als auch unter Matrixzusatz (Matrixkalibrierung) durchgeführt. Die weitere Teilvalidierung der MRM-Methode für neun Lipide der Lipidklassen PC, PE, DG und TG umfasste dabei folgende Parameter: Feststellung des linearen Bereichs, Bestimmung der NWG, der Präzision sowie des *Carry-Over*. Die ermittelten Validierungskennzahlen sind in **Tab. 6.8** zusammengefasst, für die Kalibrationsgeraden der Standardsubstanzen, siehe **Anhang Abb. 13.10**.

Bei der Grundkalibration wurden folgende Grenzen des linearen Arbeitsbereiches in Bezug auf die Substanzgruppen festgestellt: Der lineare Bereich für PCs wurde zu 0.02 bis 10 μM , der für PEs wurde zu 0.02 bis 20 μM bestimmt. Der lineare Bereich für DGs betrug ebenso 0.02 bis 20 μM und der für TGs wurde zu 0.02 bis 5 μM bestimmt. Der lineare Arbeitsbereich deckte für alle vier Substanzklassen mehrere Dekaden ab. Außerdem wurde in zwei von vier Fällen eine sehr gute Linearität mit einem $R^2 > 0.99$ der linearen Regression der Kalibriergeraden erhalten. Lediglich die Kalibrationsgeraden der PC- und DG-Standardsubstanzen erreichten nur einen R^2 Wert > 0.97 .

Für die Kalibriergerade der Standardsubstanzen unter Zusatz von Matrix (BLIGH & DYER-Extrakt der QC-Probe) ergaben sich folgende in **Tab. 6.8** gezeigte Ergebnisse. Alle Matrix-Kalibrationsreihen wiesen einen R^2 Wert > 0.99 auf. Im Vergleich zu der Grundkalibrierung zeigte sich bei der Matrixkalibration, dass bei den TGs der lineare Bereich höhere Konzentrationen teilweise nicht abdeckte (siehe **Tab. 6.8**).

Der Grund für diese Abweichung, ist vermutlich darin zu suchen, dass der Matrixzusatz sich vor allem bei der Detektion der TGs bemerkbar machte. Die Gruppe der TGs in Körnermais war gegenüber den anderen betrachteten Substanzklassen sehr abundant vorhanden. Daher schöpften einige nicht quantifizierte Triglyceridspezies in der Matrix bei der gewählten Verdünnung der Extrakte bereits die Detektionsgrenze nach oben aus.

Insbesondere die Metabolite, die der Klasse der PE angehören, wiesen eine vergleichsweise geringe Empfindlichkeit auf. Um den oberen, nicht-linearen Bereich der Kalibrierung zu erreichen, musste ein Standard mit einer Konzentration von 50 μM gewählt werden. Dagegen wurde der lineare Bereich der PC-Standardsubstanz bereits bei 20 μM überschritten.

Die NWG des PC-Gruppenstandards betrug 0.9 nM und die des PE-Standards 0.02 nM. Die NWG des DG-Standards belief sich auf 40.6 nM und die des TG-Standards betrug 4.9 nM. Die Nachweisgrenzen für die einzelnen Markersubstanzen wurden ebenfalls bestimmt. Dazu wurde eine Verdünnungsreihe einer Mischung von Extrakten der Herkunftsproben (ein Extrakt je Herkunftsland und die QC-Probe) im Bereich vom Faktor 1:5 bis zu 1:300 analysiert. Anhand der daraus ermittelten Kalibriergeraden, konnten sowohl die Nachweisgrenzen als auch die Arbeitsbereiche der 20 Markersubstanzen bestimmt werden. Diese sind in **Tab. 6.8** für die Markerlipide sowie die vier Standardsubstanzen dargestellt.

Da für die Markersubstanzen keine isotopenmarkierten Standardsubstanzen beschafft werden konnten, wurde die Quantifizierung anhand von Verbindungen aus den gleichen chemischen Analytgruppen durchgeführt. Um dennoch die individuellen Unterschiede der Metabolite zu berücksichtigen, wurden die Konzentrationen der Markermetabolite zunächst in Relation zu einer

Pool-Probe, bestehend aus Probenextrakten aller sieben Herkunftsländer und der QC-Probe, bestimmt. Die Berechnung der absoluten Markerkonzentrationen erfolgte dann über die Poolprobe. Die Ergebnisse des Herkunftsprobensets sind in **Abschn. 6.6.1 (Abb. 6.11)** dargestellt.

Tab. 6.8: Ermittelte Validierungskennzahlen für die quantitative Bestimmung von Markerlipiden zur Herkunftsbestimmung mittels Grund- und Matrixkalibrierung. Die Kalibration für die quantifizierten Metabolite erfolgte anhand einer Verdünnungsreihe eines gepoolten Extraktes repräsentativer Herkunftspuren, für die Diagramme der Kalibriergeraden siehe **Anhang Abb. 13.11**.

Substanz	Geradengleichung Kalibration	R ²	NWG	Lin. Bereich	Carry-Over	Konzentration in Extrakt QC-Probe
<i>Standards - Grundkalibration</i>						
PC (36:2)	$y = 2.18 \cdot 10^7 x + 1.59 \cdot 10^6$	0.97	0.896 nM	0.02 - 10 µM	0.08 %	5.1 µM
PE (36:0)	$y = 48142x - 9223$	> 0.99	15.1 nM	0.02 - 20 µM	< 0.01 %	1.9 µM
DG (32:0)	$y = 1.45 \cdot 10^5 x + 61734$	0.97	40.5 nM	0.02 - 20 µM	0.17 %	2.2 µM
TG (48:0)	$y = 6.39 \cdot 10^6 x + 50373$	> 0.99	4.89 nM	0.02 - 20 µM	0.44 %	0.5 µM
<i>Standards – unter Matrixbelastung</i>						
PC (36:2)	$y = 3.47 \cdot 10^7 x + 1.26 \cdot 10^7$	> 0.99	-/-	0.02 - 5 µM	-/-	-/-
PE (36:0)	$y = 2363x + 1148$	> 0.99	-/-	0.02 - 20 µM	-/-	-/-
DG (32:0)	$y = 1.70 \cdot 10^6 x - 4.91 \cdot 10^5$	> 0.99	-/-	0.02 - 20 µM	-/-	-/-
TG (48:0)	$y = 7.37 \cdot 10^6 x - 3.59 \cdot 10^5$	> 0.99	-/-	0.02 - 5 µM	-/-	-/-
<i>Lipidmarker</i>						
						Konzentration in Extrakt Pool-Probe
PC (36:3)	$y = 6.19 \cdot 10^8 x - 1.82 \cdot 10^6$	> 0.99	1:177423 (0.16 nM)	1:300 – 1:5	0.03 %	28.3 µM
PE (36:3)	$y = 1.21 \cdot 10^6 x - 3318$	0.98	1:200 * (140 nM)	1:200 – 1:10	< 0.01 %	28.0 µM
DG (34:2)	$y = 3.42 \cdot 10^7 x - 2.91 \cdot 10^5$	0.97	1:5139 (33.7 nM)	1:225 – 1:5	0.13 %	173.3 µM
TG (56:3)	$y = 7.08 \cdot 10^7 x + 1.85 \cdot 10^5$	0.98	1:51550 (0.22 nM)	1:300 – 1:5	0.22 %	11.2 µM
TG (56:2)	$y = 6.84 \cdot 10^7 x - 2.45 \cdot 10^5$	> 0.99	1:24716 (0.36 nM)	1:300 – 1:5	0.16 %	8.8 µM
TG (60:3)	$y = 7.70 \cdot 10^6 x - 25468$	> 0.99	1:14202 (0.07 nM)	1:300 – 1:5	0.18 %	1.0 µM
TG (58:2)	$y = 1.87 \cdot 10^7 x - 33976$	> 0.99	1:6948 (0.39 nM)	1:300 – 1:5	0.15 %	2.7 µM
TG (60:2)	$y = 1.57 \cdot 10^7 x - 52890$	> 0.99	1:14521 (0.15 nM)	1:300 – 1:5	0.12 %	2.2 µM
TG (52:6)	$y = 1.62 \cdot 10^7 x + 6.02 \cdot 10^5$	0.97	1:574 (6.1 nM)	1:200 – 1:5	10.0 %	3.5 µM

* abgeschätzt aus Kalibriergerade.

Aus **Tab. 6.8** ist weiterhin ersichtlich, dass die Nachweisgrenzen der Markerlipide je nach deren Lipidklasse generell in der Reihenfolge PC < TG < DG < PE zunahm. Die NWG lag für die einzelnen Markerlipide dabei im Bereich von 1:177423 bis 1:125, somit könnte also beispielsweise der Marker PC (36:3) theoretisch noch in einem 1:177423 verdünnten gepoolten Extrakt bestimmt

werden. Unter der Annahme, dass sich der Marker PC (36:3) analytisch genauso wie der Gruppenstandard PC (36:2) verhält, entspricht dies einer NWG von 0.16 nM. Dabei ist zu beachten, dass die NWG lediglich rechnerisch bestimmt wurde und nicht in der Praxis über die Analyse eines entsprechend verdünnten Messstandard/Probe als Peak verifiziert wurde. Die tatsächliche NWG liegt vermutlich bei höheren Konzentrationen.

Die NWG der einzelnen Marker liegen in derselben Größenordnung wie die NWG der zugehörigen Standardsubstanzen. Lediglich der Metabolit TG (52:6) zeigte ein auffällig anderes Verhalten als der repräsentative Gruppenstandard TG (48:0) und wies vergleichsweise schlechte NWG auf (siehe **Tab. 6.8**). Dies ist vermutlich darauf zurückzuführen, dass TGs auf dem C₁₈-Säulenmaterial vergleichsweise spät und nicht immer vollständig eluieren, sodass die Blindwerte erhöht waren.

Die Quantifizierung der Menge der Standardsubstanzen in dem undotierten QC-Probenextrakt mittels der Matrixkalibration ergab Konzentrationen im Bereich von 0.5 bis 5.1 µM (siehe **Tab. 6.8**). Insbesondere die Konzentration des TG-Standards war dabei relativ gering, wogegen die Konzentrationen der übrigen Standards eher im zu erwartenden Bereich der anderen Markerlipide lagen.

Die Zuverlässigkeit der LC-MRM-Methode wurde weiter überprüft, indem ein Mischstandard von äquivalenten Mengen der vier Standardsubstanzen in einen BLIGH & DYER-Extrakt der QC-Probe dotiert wurde (Konzentration je 5 µM) und die Präzision bestimmt wurde. Dabei wurde eine RSD von 1.4-20.0 % erreicht. Die geringste RSD wurde bei dem TG-Standard erreicht, dann folgten die Standards DG und PC, die schlechteste RSD stammte von dem PE-Standard, welcher eine Schwankungsbreite von 20.0 % aufwies. Ein *Carry-Over*-Effekt der zudotierten Standardsubstanzen konnte nicht beobachtet werden, der höchste Wert wurde mit lediglich 0.4 % für den TG-Gruppenstandard detektiert.

Für alle quantifizierten Metabolite wurde ebenfalls eine niedrige RSD von ≤ 15 % in der QC-Probe (2.1 bis 10.4 %) erzielt. Ein *Carry-Over* nach der Messung einer höher konzentrierten Probe (Verdünnungsfaktor 1:5) konnte nicht beobachtet werden. Die Verschleppung der Markersubstanzen in den folgenden Blank überstieg in keinem Fall 1.0 %, mit Abweichung des Triglycerids TG (52:6). Dieses stellte insofern eine Ausnahme dar, da es zudem einen hohen Blindwert aufwies, da 15.5 % des Signals der Poolprobe (1:10 verd.) bereits im Blindwert erreicht wurden.

6.6.1 Ergebnisse der Klassifizierung durch eine MRM-Methode (MS²-Level)

Die Ergebnisse der beispielhaften Quantifizierung von neun ausgewählten Metaboliten in 151 Proben des Herkunftsprobensets sind in **Abb. 6.11** zusammengestellt. Da für die meisten der quantifizierten Metabolite keine Reinsubstanzen kommerziell zugänglich waren, wurde jeweils ein Lipid der Klassen PC, PE, DG sowie TG als Gruppenstandard genutzt. Über diese Gruppenstandards wurden die Lipidkonzentrationen in der Poolprobe absolut bestimmt. Die Bestimmung der Konzentrationen der Marker in den Extrakten der Proben des Herkunftsprobensets erfolgte indirekt, indem die Markerkonzentrationen zunächst in Relation zu der entsprechenden Konzentration in der Poolprobe bestimmt wurden.

Es zeigte sich hier ebenfalls eine herkunftsabhängige Schwankung der Markerkonzentrationen (siehe **Abb. 6.11**). Die Verteilungsmuster der Lipidklassen in Bezug auf die Herkunftsorte waren dabei ähnlich. Im Vergleich zu den ermittelten Peakflächen auf LC-MS-Ebene (siehe **Anhang Abb. 13.9**), zeigten sich bei den Markern vergleichbare Auffälligkeiten je Herkunftsland: Während für die Phospholipide PC(36:3) und PE (36:3) die Konzentrationen für Mais aus der Ukraine besonders gering war, war das Diglycerid DG (34:2) vor allem in den Stärkemaisproben aus Frankreich und Ungarn vergleichsweise niedrig konzentriert. Die TGs waren allgemein in den Proben aus Peru in Relation zu den anderen Probengruppen erhöht.

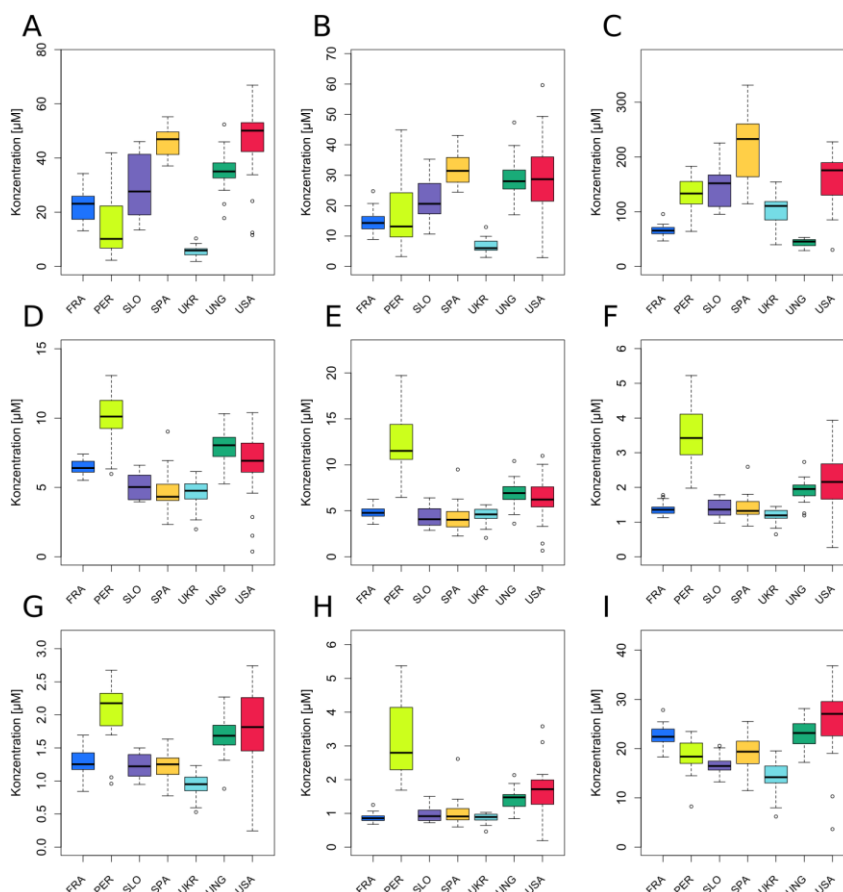


Abb. 6.11: Darstellung der herkunftsabhängigen Schwankungen der quantifizierten Marker in den Extrakten des Herkunftssensets. Ermittelt in Relation zu den Konzentrationen in einer Poolprobe, die anhand der Standardreihe der Gruppenstandards berechnet wurde. **A:** PC (36:3), **B:** PE (36:3), **C:** DG (34:2), **D:** TG (56:3), **E:** TG (56:2), **F:** TG (60:3), **G:** TG (58:2), **H:** TG (60:2), **I:** TG (52:6).

Insgesamt kann festgestellt werden, dass die Bestimmung von ausgewählten Lipiden mittels einer LC-MRM-Methode für alle untersuchten Metabolite kongruente Ergebnisse zu der Bestimmung auf LC-MS-Ebene ergab. Somit ist davon auszugehen, dass die identifizierten Marker signifikant unterschiedlich je nach geographischer Herkunft beeinflusst waren.

In **Abb. 6.12A** sind die Varianzen der Proben anhand der Peakflächen auf MS^2 -Ebene mittels einer PCA dargestellt. Die Punktwolken im PCA-Scoresplot, wiesen eine ähnliche Verteilung und Gruppierung im Koordinatenraum auf, wie unter Nutzung der Peakflächen der LC-MS-Analyse (siehe **Abb. 6.8A**). Entsprechend ihrer geographischen Herkunft gruppieren sich die Herkunftssproben ebenfalls in Cluster, auch wenn sich ihre absolute Position im PCA-Scoresplot

unterschied. Die ersten drei Hauptkomponenten deckten dabei mit 98.7 % einen hohen Anteil der Varianz ab, zudem waren lediglich zwei Hauptkomponenten zur Erklärung von mehr als 95 % der Varianz notwendig.

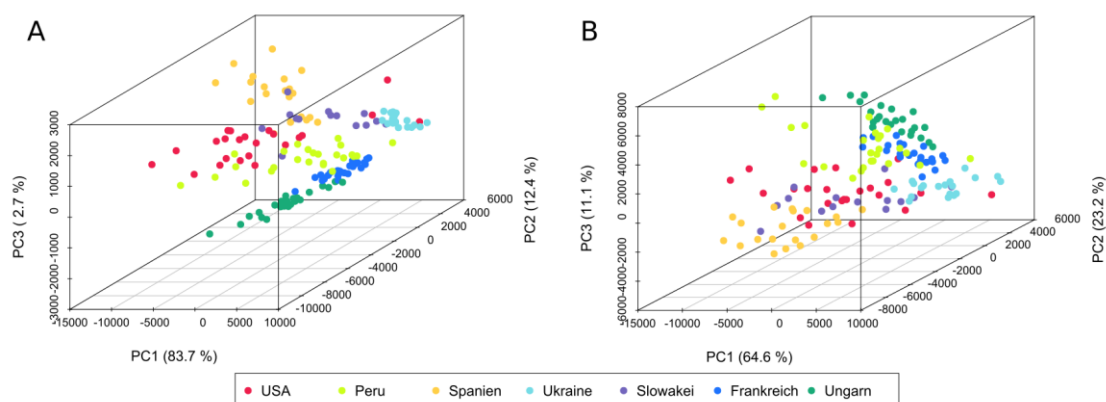


Abb. 6.12: PCA-Scoresplots der Herkunftspuren (nach *Pareto-Scaling*) der MS^1 - und MS^2 -Peakflächen derselben neun Markermetabolite im Vergleich. **A:** Auf Grundlage der ermittelten MS^1 -Peakflächen. **B:** Anhand der MS^2 -Peakflächen, die mittels der LC-MRM-Methode bestimmt wurden.

Im Vergleich zu den Klassifikationsraten, welche in **Abschn. 6.4.3** für das Modell basierend auf den Peakflächen aller 20 Marker erzielt werden konnten, wurden für die MRM-basierte Quantifizierung einiger ausgewählter Metabolite im Vergleich durchaus ähnlich gute Werte erreicht: Auf globaler Ebene konnten anhand der neun Markersubstanzen 87.1 ± 0.9 % der Herkunftspuren korrekt vorhergesagt werden (100-fach wiederholte, 10-fache Kreuzvalidierung). Im direkten Vergleich, ermöglichten die neun Marker sowohl unter Nutzung der MS^1 -Peakflächen, als auch der MS^2 -Peakflächen mit 87.2 % bzw. 87.1 % korrekten Zuordnungen bei der RF-Klassifikation eine vergleichbare, robuste Herkunftsindikation (siehe **Abb. 6.13** für die Klassifikationsmatrizen).

A		FRA	PER	SLO	SPA	UKR	UNG	USA	CC
	FRA	24.87	0	0	0	0	0.13	0	99.48 %
	PER	0	23.95	0	0	1.02	0	0.03	95.80 %
	SLO	0	0	7.45	2.99	0.93	0	2.63	53.21 %
	SPA	0.01	0	0.89	18.09	0	0	1.01	90.45 %
	UKR	0	0	1.00	0	18.00	0	0	94.74 %
	UNG	0.10	0	0	0	0	24.90	0	99.60 %
	USA	0.01	2.47	4.44	0.65	1.07	0	14.36	62.43 %

B		FRA	PER	SLO	SPA	UKR	UNG	USA	CC
	FRA	23.86	0	0	0	0	1.14	0	95.44 %
	PER	0	23.09	0.11	0	1.00	0	0.80	92.36 %
	SLO	0	0	8.94	3.00	0.01	0	2.05	63.86 %
	SPA	0	0.03	1.33	17.08	0	0	1.56	85.40 %
	UKR	0	0	0.85	0	18.15	0	0	95.53 %
	UNG	1.59	0.03	0	0	0	23.38	0	93.52 %
	USA	0	1.01	2.48	1.00	1.50	0	17.01	73.96 %

Abb. 6.13: Vergleich der RF-Klassifikationsmatrizen der Proben des Herkunftspurensets. **A:** Anhand der Peakflächen der neun quantifizierten Metabolite auf MS^1 -Ebene. **B:** Anhand der Peakflächen der neun quantifizierten Metabolite auf MS^2 -Ebene.

Auf Länderebene war die korrekte Zuordnung, insbesondere der Länder Spanien (CC 85.4 %), Slowakei (CC 63.9 %) sowie den USA (CC 74.0 %) auf MS^2 -Ebene schwierig (siehe **Abb. 6.13B**). Im Vergleich zu dem Klassifikationsmodell mit 20 Markersubstanzen auf MS^1 -Ebene zeigten sich ähnliche Schwächen in Bezug auf die Proben aus der Slowakei und den USA (siehe **Abschn. 6.4.3**). Für die spanischen Proben waren die CC-Werte jedoch bei Nutzung von 20 Markern mit 96.2 %

weitaus besser. Da es sich bei diesen Proben um 20 unterschiedliche Körnermaissorten handelte, ist anzumerken, dass eine Klassifikation mittels vieler Marker vermutlich eine höhere Robustheit in Bezug auf (negative) Einflüsse der Sorte auf die Herkunftsbestimmung einhergehen könnte, als wenn weniger Marker verwendet werden.

Die Klassifikationsgenauigkeit insgesamt war jedoch im Vergleich zu dem Modell mit 20 Markersubstanzen nur leicht erniedrigt (90.5 % vs. 87.1 %, d. h. -3.4 %). Da eine Reduktion der Anzahl an verwendeten Markern von 20 auf neun erfolgte, ist zu vermuten, dass die Genauigkeit der Klassifikation durch die erhöhte Spezifität aufgrund der Nutzung von MRM-Analysedaten erhöht wird. Dieser generelle Trend konnte auch von Klockmann *et al.* bei der geographischen Herkunftsbestimmung von Haselnussproben festgestellt werden⁴⁴. Die Ursache ist vermutlich darin zu suchen, dass durch die Beobachtung eines jeweils spezifischen MS²-Fragmentes die Genauigkeit der Bestimmung der Metabolitkonzentration in der MRM-Methode höher ist – und somit die Realität besser wiedergibt – als die bloße Betrachtung von Peakflächen auf MS¹-Ebene.

Die MRM-Methode wurde zusätzlich anhand der Analyse von insgesamt 20 Validierungsproben aus Frankreich bzw. Ungarn (je n = 10) getestet. Diese Proben wurden anschließend durch das RF-Klassifikationsmodell vorhergesagt. Dabei wurde eine Vorhersagegenauigkeit von 71.7 % erreicht (14 von 20 Proben korrekt zugeordnet). Dies unterstreicht die generelle Eignung der Verifikation der geographischen Herkunft von Körnermais mittels einer LC-MRM-Methode, welche lediglich auf einer geringen Anzahl von Markermetaboliten beruht.

6.6.2 Vergleich mit der Tocopherol-basierten Methode

Die Bestimmung der Authentizität vieler – meist fettreicher – Nahrungsmittel, wie Paranüsse, Kaffee, Olivenöl oder Mandelöl konnte bereits in den 1970er-Jahren aufgrund der Analyse ihres Tocopherol-Musters erfolgen^{307–310}. Da Körnermais nach Goffman & Böhme hauptsächlich γ -Tocopherol und α -Tocopherol enthält³⁵⁷, erschien die Möglichkeit einer Herkunftsverifikation mittels der Gehalte dieser beiden Tocopherole zumindest vielversprechend. Um die Eignung einer Tocopherol-basierten Klassifizierungsmethode mit der in **Abschn. 6.6** entwickelten Lipidmarker-basierten LC-MRM-Methode zu vergleichen, wurde zunächst eine LC-MRM-Methode für die parallele Analytik von α - und γ -Tocopherol in Körnermais entwickelt. Die Extraktion der Tocopherole aus den Körnermaisproben erfolgte dabei mittels Isopropanol^{358–360}. Dem Extraktionsmittel wurde Butylhydroxytoluol (BHT) als Antioxidans beigemischt^{44,306,361}.

6.6.2.1 Bestimmung von α -Tocopherol & γ -Tocopherol mittels einer MRM-Methode

Die Bestimmung von α - und γ -Tocopherol erfolgte parallel in einem einzigen LC-Lauf mittels massenspektrometrischer Detektion. Die HPLC-Methode wurde dabei von Klockmann *et al.*⁴⁴ adaptiert. Da α -Tocopherol in dieser Methode bereits bei einer Retentionszeit von 2.6 min eluierte, wurde der Gradient entsprechend angepasst. Der Laufmittelanteil zu Gradientenbeginn wurde von 85 % Laufmittel B auf 75 % B reduziert, entsprechend wurde die Länge des Gradienten von 3 min auf 4 min verlängert, die Gesamtlänge der Methode blieb mit 13.5 min unverändert. Unter diesen Bedingungen gelang es, die beiden Tocopherole chromatographisch als basisliniengetrennte Peaks zu detektieren: γ -Tocopherol eluierte zuerst bei 4.8 min, dann folgte α -Tocopherol bei 5.2 min. Beide Tocopherole waren über deren $[M+H]^+$ -Addukt mit einem m/z -

Verhältnis von 431.3889 für α -Tocopherol und 417.3733 für γ -Tocopherol massenspektrometrisch detektierbar.

Um eine bessere Quantifizierung zu ermöglichen, wurde zur Detektion eine MRM-Methode entwickelt. Als Kollisionsenergie wurden dabei 40 eV gewählt. Die für die Auswertung verwendeten MRM-Übergänge waren für α -Tocopherol m/z 431.39 $\rightarrow$ 165.1 (*Quantifier*) bzw. 137.1 (*Qualifier*). Für γ -Tocopherol wurden als Übergänge m/z 417.37 $\rightarrow$ 151.1 (*Quantifier*) bzw. 123.1 (*Qualifier*) genutzt.

In Körnermaisproben ist nach Grams *et al.* α -Tocopherol im Bereich von 14.7-25.4 mg/kg bzw. γ -Tocopherol im Bereich von 38.6-70.0 mg/kg zu erwarten^{357,362}. Da in Relation zu Lipiden die Empfindlichkeit des ESI-Massenspektrometers für die Tocopherole relativ gering ist^{44,363}, wurde für die Bestimmung von α - bzw. γ -Tocopherol in Körnermais ein geringerer Verdünnungsfaktor als bei der Analytik der Lipide eingesetzt.

6.6.3 Herkunftsbestimmung mittels Lipidprofil vs. Tocopherole-basierter Methode

Die entwickelte Tocopherol-Methode wurde daraufhin auf einen Teil der Herkunftsproben angewendet und deren Eignung für die Herkunftsverifikation mit jener der Lipid-basierten Methode verglichen. Dabei wurden Tocopherol-Konzentrationen im Bereich von 6.1-17.6 mg/kg α -Tocopherol sowie 6.6-34.7 mg/kg γ -Tocopherol in den Herkunftsproben ermittelt, siehe **Abb. 6.14**. Die nachgewiesenen absoluten Mengen lagen dabei unter den Werten, die von Grams *et al.* angaben, jedoch konnte der Trend, dass mehr γ -Tocopherol als α -Tocopherol in Mais vorhanden ist, bestätigt werden³⁶². Vermutlich wurden die Analyten mittels der durchgeführten Flüssigextraktion nicht vollständig extrahiert, da im Gegensatz zu Grams *et al.* keine Soxhlet-Extraktion durchgeführt wurde.

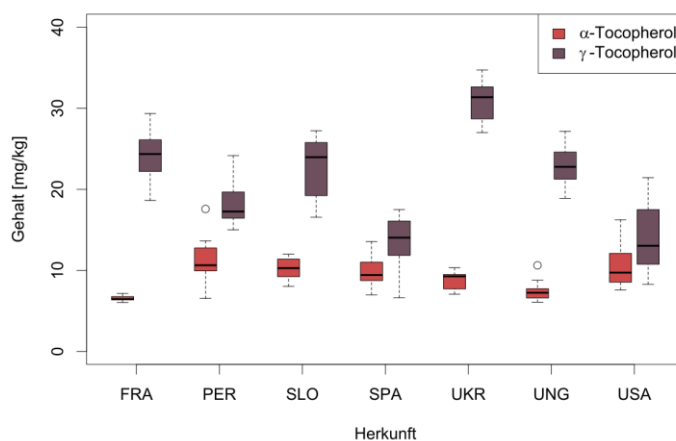


Abb. 6.14: Darstellung der Tocopherol-Konzentrationen in Abhängigkeit von der geographischen Herkunft der Maisproben.

Insgesamt waren beide Tocopherole mit einem ANOVA p -Wert von $5.85 \cdot 10^{-11}$ (α -Tocopherol) bzw. $4.57 \cdot 10^{-29}$ (γ -Tocopherol) stark signifikant nach der Probenherkunft beeinflusst. Der *post-Hoc* TUKEY-Test offenbarte jedoch, dass γ -Tocopherol mit 16 signifikant unterschiedlichen Länderpaarungen einen besseren Marker als α -Tocopherol mit lediglich acht signifikanten Länderpaarungen darstellte. Auf Länderebene bereitete insbesondere die Trennung von

Konstellationen Schwierigkeiten, an denen die Länder Ungarn, Frankreich, USA, Peru oder Spanien beteiligt waren.

Diese Ergebnisse wurden unterstützt durch die RF-Klassifikation, welche insbesondere bei der Zuordnung der Proben zu den Ländern Peru, USA und Spanien hohe Fehlerraten aufwies. Die erreichten Klassifikationsgenauigkeiten (RF-Algorithmus, 10-fach wiederholte 10-fache Kreuzvalidierung) sind in **Abb. 6.15** dargestellt. Die Effizienz der Klassifikation anhand der beiden Tocopherole wurde dazu verglichen mit je zwei Lipidmarkern. Diese wurden zufällig ausgewählt, entweder aus den neun Lipid-MRM Markern, aus den 20 optimalen Markern, oder aus den Top 20 der Phospholipid, DG oder TG-Marker. Diese Vorgehensweise wurde hier gewählt, um einen *Bias* aufgrund der unterschiedlichen Anzahl der zur Klassifikation verwendeten Markersubstanzen zu umgehen.

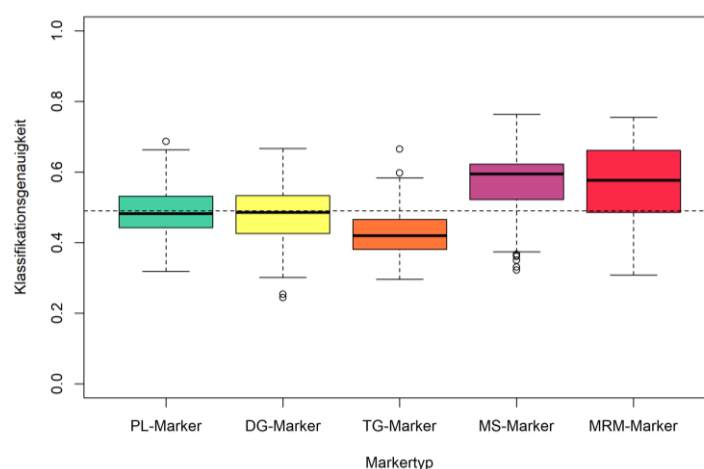


Abb. 6.15: Boxplot der erreichten Klassifikationsgenauigkeit (RF-Algorithmus, 10-mal wiederholte, 10-fache Kreuzvalidierung) bei Nutzung von je zwei Lipidmarkern. Die Art der Lipidmarker ist auf der x-Achse notiert. Die gestrichelte Linie gibt die Klassifikationsgenauigkeit mittels α -Tocopherol und γ -Tocopherol an.

Die Klassifikation der Proben anhand der beiden Tocopherole erzielte eine Genauigkeit von 49.0 ± 1.6 %. Im Vergleich dazu nahm die Fähigkeit, die Proben korrekt zu klassifizieren, in der Reihenfolge der Markersubsets optimale MS^1 -Marker > MRM-Marker > PC/PE $\approx$ DG > TG ab. Die besten Ergebnisse wurden im Mittel durch das Subset der 20 ausgewählten Lipide (MS^1 -Ebene) mit 57.0 ± 8.7 % erreicht. Dagegen schnitt die Klassifikation anhand von je zwei TGs mit 42.9 ± 6.6 % am schlechtesten ab. Die Lipid-MRM-Marker erwiesen sich dabei im Mittel um 7.2 % besser als die Tocopherole. In Relation zu einem Subset aus den 20 Markermetaboliten erzielten die Tocopherole ein um 8 % schlechteres Ergebnis.

Gründe für diese schlechtere Eignung der Tocopherol-basierten Methode sind zum einen in einer möglichen Sortenabhängigkeit (niedrige CC von 38.7 % bei den spanischen Proben) und auch in der Beschränkung auf lediglich zwei Marker zu suchen, was zudem im Hinblick auf die Fälschungssicherheit der Methode eher kritisch zu bewerten ist.

Der Ansatz einer Herkunftsbestimmung mittels des Tocopherolmusters wurde aus den aufgeführten Gründen nicht weiter verfolgt: Die Proben aus Spanien offenbarten eine gewisse Sortenabhängigkeit beider Tocopherole. Im Hinblick auf die Rolle von Vitamin E in der pflanzlichen

Zelle^{357,364} ist dies als schlüssig anzusehen. Zudem ist diese Variabilität im Einklang mit den Ergebnissen von Grams *et al.*, welche ebenso eine sortenabhängige Varianz der Konzentrationen von α - γ -Tocopherol in Mais feststellen konnten³⁶². Galliher *et al.* berichteten von einer starken Varianz für diese Tocopherole auch in Maiskeimlingen³⁶⁵. Die beiden Marker wiesen zudem im Vergleich eine schlechtere Klassifizierungsleistung, als die neun quantifizierten Lipidmarker auf MS²-Ebene, auf. Problematisch ist weiterhin, dass die Detektion der Tocopherole mittels Massenspektrometrie recht unempfindlich ist^{44,363}, weshalb eine gleichzeitige Quantifizierung der Tocopherole zusammen mit weiteren Lipid-basierten Markersubstanzen sich als Herausforderung darstellt.

6.7 Untersuchungen mittels FT-NIR-Spektroskopie

Die FT-NIR-Spektroskopie ist als schnelle spektroskopische Methode in der Cerealien-verarbeitenden Industrie weit verbreitet^{27,47,48,178,301–303,366}. Sie kommt vor allem als Qualitätssicherung zum Einsatz, um die Restfeuchte, den Proteingehalt und weitere Parameter zu überwachen. Im Falle von Mais wird dazu meist das unvermahlene, intakte Maiskorn analysiert^{27,88,89}, jedoch wird dies kontrovers gesehen, da intakte Maiskörner asymmetrisch geformt sind. Nansen *et al.* zeigten zudem, dass das Vermahlen von Maisproben einen starken Einfluss auf die Reproduzierbarkeit hat²⁴⁹. Bei zunehmendem Mahlgrad nahm die Varianz in den NIR-Spektren gegenüber intakten Maiskörnern ab.

Deshalb wurde zunächst die am besten für die FT-NIR-Spektroskopie geeignete Probenform evaluiert. Zur Methodenoptimierung wurde die QC-Probe II genutzt, welche sowohl als ganze Körner, als auch als vermahlene, pulverförmige Probe vorlag.

FT-NIR-Spektren weisen zwei dominante Wasserbanden, bei 6870 cm⁻¹ und 5180 cm⁻¹ auf^{303,367}. Diese könnten die Herkunftsbestimmung von Körnermais stören, falls die Proben schwankende Wassergehälte aufweisen. Daher war zu überprüfen, ob die Wasserbanden durch Gefriertrocknung der Probe vor der Analyse entfernt bzw. auf ein konstantes Level gebracht werden müssen, oder ob es ausreichend ist, die Wasserbanden für die Datenauswertung zu exkludieren.

Um eine möglichst schonende Homogenisierung der Proben zu gewährleisten, erfolgte die Probenvorverarbeitung analog zu jener der LC-MS-Analytik, indem die Maiskörner zunächst in Flüssigstickstoff schockgefroren und anschließend in einer Messermühle unter Zusatz von Trockeneis im Verhältnis 1:1 vermahlen wurden. Um zudem sicherzugehen, dass eine möglicherweise variable Restfeuchte der Proben nicht die Probenklassifikation beeinträchtigen könnte, wurden die Sortenproben zudem einer 24 h dauernden Gefriertrocknung unterzogen. Diese Vermahltechnik wurde ebenso von Arndt *et al.* für die FT-NIR-Analytik von Mandeln genutzt²⁵⁵.

Methodisch wurde zur Klassifikation der Proben mittels der FT-NIR-Spektroskopie analog zu **Abschn. 6.4** vorgegangen: Als Klassifikationsmethode diente jedoch in diesem Fall der SVM-Algorithmus. Diese Klassifikationsmethode wurden bereits von Richter *et al.* bei der Herkunftsanalyse von weißem Spargel nach dessen Ursprungsland genutzt⁴⁵ sowie von Arndt *et al.* zur Klassifikation von Mandeln²⁵⁵.

6.7.1 Auswahl der optimalen Probenform für FT-NIR-Spektren

In **Abb. 6.16** sind jeweils 100 FT-NIR-Spektren von intakten Maiskörnern der QC-Probe II in beiden möglichen Orientierungen dargestellt. Außerdem sind ebenso 100 Spektren derselben Probe in Pulverform gezeigt.

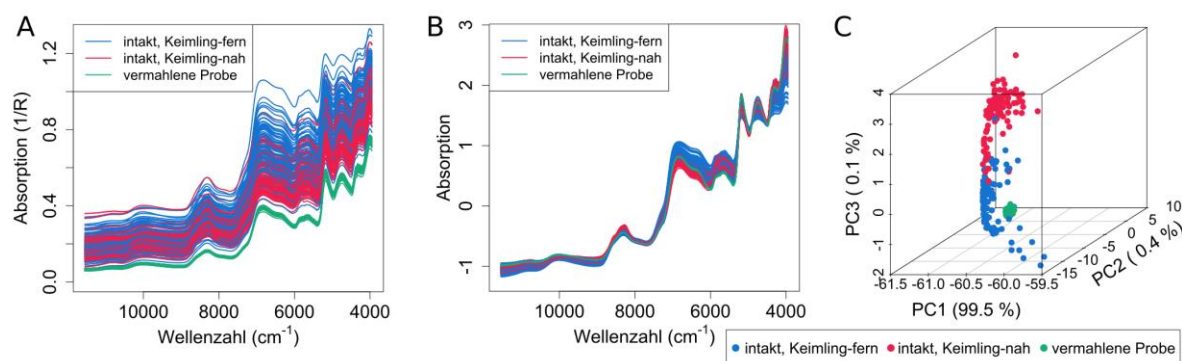


Abb. 6.16: FT-NIR-Spektren der QC-Probe II nach unterschiedlicher Probenvorbereitung. **Rot:** „intaktes Korn, Keimling-zugewandte Seite“. **Blau:** „intaktes Korn, Keimling-abgewandte Seite“. **Grün:** vermahlene Maiskörner (je $n = 100$ Spektren). **A:** Rohdaten der Spektren. **B:** Spektren nach *Preprocessing* mittels SNV. **C:** PCA der FT-NIR-Spektren nach SNV-*Preprocessing*.

Es konnte beobachtet werden, dass die Varianz der FT-NIR-Spektren derselben Probe für intakt analysierte Maiskörner, sehr viel größer ausfiel, als bei der homogenisierten Probe. Die beiden Orientierungen des Maiskorns konnten anhand der Clusterbildung im PCA-Scoresplot relativ exakt voneinander unterschieden werden. Verglichen mit der sehr dichten Clusterung der Spektren der vermahlenden Probe, hätte jedoch die Analytik von intakten Maiskörnern eine inakzeptable zusätzliche Variabilität in den Datensatz hineingetragen.

Ein *Preprocessing* der FT-NIR-Spektren mittels SNV (siehe **Abb. 6.16B**) führte zu einer offensichtlichen Reduktion, der Varianz. Da jedoch die Varianz der Spektren von intakten Maiskörnern in Relation zu vermahlenden Maiskörnern – trotz *Preprocessing* – inakzeptabel hoch erschien, wurden alle Proben für die FT-NIR-Spektroskopie in Form von vermahlenem Pulver analysiert.

6.7.2 Multivariate Analyse der FT-NIR-Spektren des Sortenprobensets

Um einen eventuellen Einfluss der Maissorte auch bei der FT-NIR-Analytik auszuschließen, wurde das Probenet aus den Landessortenversuchen untersucht. Die 81 FT-NIR-Spektren wurden dabei zunächst einem *Preprocessing* mittels SNV unterzogen, die Spektren sind in **Abb. 6.17** gezeigt. Diejenigen Wellenzahlen/Spektrenbereiche, welche einem signifikanten Herkunftseinfluss (ANOVA-Test p -Wert ≤ 0.01) unterlagen, sind in **Abb. 6.17A** im unteren Bereich mittels schwarzer Markierungen dargestellt. Die entsprechenden Banden, die hingegen einem signifikanten Sorteneinfluss (ANOVA-Test p -Wert ≤ 0.01) aufwiesen, sind in **Abb. 6.17B** in entsprechender Weise markiert.

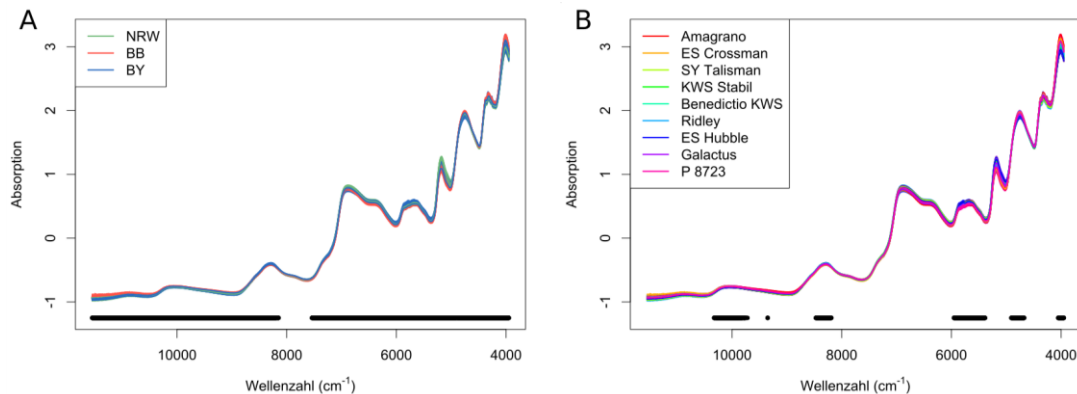


Abb. 6.17: Darstellung der FT-NIR-Spektren des Sortenprobensets (*Preprocessing* mittels SNV). **A:** Markierung der Proben entsprechend des Herkunftsbundeslandes, siehe Legende. **B:** Markierung der Proben entsprechend der Körnermaissorte, siehe Legende. In **(A)** sind jene Wellenzahlen, die einem signifikanten Herkunftseinfluss (ANOVA-Test p -Wert ≤ 0.01) unterliegen, im unteren Bereich schwarz markiert. In **(B)** sind jene Wellenzahlen, die einem signifikanten Sorteneinfluss (ANOVA-Test p -Wert ≤ 0.01) unterliegen, markiert.

Die enthaltene Varianz in den FT-NIR-Spektren nach *Preprocessing* mittels SNV ist im **Anhang** in **Abb. 13.12A** mittels einer PCA dargestellt. Dem Scoresplot war zu entnehmen, dass die Varianz der FT-NIR-Spektren innerhalb des Sortenproben-Datensatzes zu einer Gruppierung der Proben nach deren geographischen Herkunft führte. Anhand der Länge der Loadingsvektoren im Loadingsplot der PCA konnten insgesamt sechs besonders herkunftsrelevante Spektralbereiche identifiziert werden (8400, 6980-6395, 5800, 5180-5160, 4775-4810, 4010 cm⁻¹). Die Spektralbereiche waren dabei insbesondere in den Absorptionsbanden für ungesättigte Lipide (8400 cm⁻¹), Wasser/Stärke (6980-6395 cm⁻¹), Lipide (5800 und 4010 cm⁻¹), Wasser (5180-5160 cm⁻¹) und Proteine (4775-4810 cm⁻¹) zu finden^{303,367}. Eine Separation nach Maissorte konnte hingegen in der PCA nicht festgestellt werden.

Die Sortenabhängigkeit einiger Wellenzahlen/Spektralregionen konnte jedoch mittels PLS-DA gezeigt werden. Der entsprechende Scoresplot ist im **Anhang** in **Abb. 13.12B** abgebildet. Hier war nun eine eindeutige Separation der Sortenversuchsproben nach ihrer Maissorte – unabhängig von der Anbauregion – zu beobachten. Die dabei im Loadingsplot relevanten Spektralbereiche lagen v. a. bei 5160-5180 cm⁻¹, innerhalb einer Absorptionsbande von Wasser³⁶⁷, sowie um 4010 cm⁻¹. Da die Bande bei 4010 cm⁻¹ sehr nahe am Rand des Messbereichs lag, wurde diese nicht als störend für die Herkunftsbestimmung angesehen, weil ausgewählte Bereiche zur Herkunftsbestimmung nicht aus diesem Randbereich des Spektrums ausgewählt werden. Weitere, weniger stark hervorstechende Spektralbereiche lagen bei 6880 sowie 5800 cm⁻¹. Diese Bereiche befinden sich in den charakteristischen Absorptionsbereichen von Stärke (6880 cm⁻¹) sowie Lipiden (5800 und 4010 cm⁻¹)³⁰³. Der Q²-Wert für die PLS-DA Analyse nach Körnermaissorte betrug lediglich 0.11. Dies deutet daraufhin, dass die Maissorte mittels NIR nicht erfasst wird. Die Sortenunabhängigkeit der Zuordnung der Herkunftsproben wurde dennoch zusätzlich anhand von 20 Körnermaissorten von einem Standort (Lleida, Spanien) im Rahmen des Herkunftsbestimmungsmodells überprüft.

Es zeigte sich somit ein ähnliches Bild, wie bereits zuvor bei der LC-MS-basierten Analytik. Waren lediglich 22.7 % der erfassten Wellenzahlen signifikant sortenabhängig beeinflusst, so waren hingegen 91.6 % der erfassten Wellenzahlen signifikant herkunftsabhängig beeinflusst. Eine

nähere Aufschlüsselung der Signifikanz mittels *post-Hoc* TUKEY-Test zeigte zudem, dass unter den sortenabhängig beeinflussten Wellenzahlen Kombinationen, welche die Körnermaissorte P 8723 enthielten, überwiegend für die Unterschiede verantwortlich waren. Abseits von dieser Sorte waren lediglich 119 von 3725 Wellenzahlen signifikant bezüglich der Körnermaissorte. Dies entspricht 3.2 %.

Die von der geographischen Herkunft beeinflussten Banden der FT-NIR-Spektren befanden sich quasi im gesamten Bereich des aufgenommenen Spektrums verteilt, mit Ausnahme eines schmalen Fensters bei 8185-7540 cm^{-1} . Sortenbezogen beeinflusst waren lediglich sechs schmale Bereiche des Spektrums. Die bedeutendsten befanden sich bei 10325-9350, 5955-5385 und 4910-4675 cm^{-1} . Diese Bereiche sind folgenden Substanzgruppen ursächlich zuzuordnen: Wasser (3. Oberschwingung), Lipide (1. Oberschwingung der C-H Streckschwingung; $-\text{CH}_3$ bzw. $-\text{CH}_2-$) und Proteine³⁰³.

6.7.3 Untersuchung einer möglichen Sortenabhängigkeit bei der Klassifikation von FT-NIR spektroskopischen Messungen

Der eventuell störende Einfluss der Maissorte auf die Bestimmung der Herkunft mittels Klassifizierungsalgorithmen wurde mittels des SVM-Algorithmus abgeschätzt. Zur Klassifikation wurde jeweils das komplette Spektrum eingesetzt. Als Kernel-Funktion diente eine lineare Kernelfunktion. Der Cost-Parameter C wurde im Bereich von $1 \cdot 10^{-4}$ bis $1 \cdot 10^0$ für die Herkunftsbestimmung optimiert zu 0.0041 sowie für die Sortenbestimmung zu 0.0083. Die erhaltenen Klassifikationsergebnisse sind im **Anhang** in **Abb. 13.13** dargestellt.

Die Ergebnisse zeigten, dass die Genauigkeit der Klassifikation nach geographischer Herkunft mit 92.7 % sehr viel besser war, als jene der Sorten mit nur 39.0 %. Im Hinblick auf die Fehleranfälligkeit der Klassifikation nach Herkunft ist festzustellen, dass die Trennung der Proben für alle drei Probenklassen etwa gleich effizient war. Bei der Klassifikation nach Sorten zeigte sich hingegen, dass die Gesamtrate an korrekten Klassifikationen lediglich auf die korrekte Zuordnung von Proben der Sorte P 8723 zurückzuführen waren (CC von 77.9 %). Alle anderen Sorten waren hingegen mittels FT-NIR-Analytik (CC im Bereich von 6.3 bis 54 %; Mittelwert 34.2 %) nicht voneinander zu unterscheiden.

Diese Ergebnisse untermauern die Erkenntnisse aus der Signifikanzanalyse, die ebenso ergab, dass lediglich die Spektren der Körnermaissorte P 8723 ursächlich für signifikant unterschiedliche Spektrenbereiche war.

6.8 Analytik der Herkunftspuren mittels FT-NIR-Spektroskopie²

6.8.1 Interpretation der FT-NIR-Spektren von Körnermais

In **Abb. 6.18** sind die FT-NIR-Spektren des Herkunftsprobensets nach *SNV-Preprocessing* abgebildet (Wellenzahlenbereich von 10000 bis 3950 cm^{-1}), zusätzlich die die Medianspektren je Herkunftsländ in dunkleren Farben abgebildet. Alle unprozessierten Spektren können **Anhang Abb. 13.14** entnommen werden. Die Spektrenform war für alle Proben unabhängig von der Probenherkunft ähnlich. Da die Region $> 10000 \text{ cm}^{-1}$ nur vergleichsweise geringe Absorptionsbanden aufwies, wurden diese von der weiteren Auswertung ausgeschlossen.

Die erste intensive Absorptionsbande befand sich bei ca. 8350 cm^{-1} , welche der zweiten Oberschwingung der C=C-H Streckschwingung entspricht. Diese rührt von ungesättigten Lipiden her. Bei 6870 cm^{-1} und 5180 cm^{-1} , befanden sich zwei Absorptionsbanden von Wasser (1. Oberschwingung der O-H Streckschwingung bzw. Kombinationsschwingung aus einer O-H Streckschwingung und einer H-O-H Deformationsschwingung). Eine schwache Absorptionsbande war bei 6385 cm^{-1} lokalisiert, der ebenso die 1. Oberschwingung einer O-H Streckschwingung zugrunde liegt. Diese stammt von Stärkemolekülen. Die entsprechenden Schwingungen der aliphatischen Reste von Lipiden befanden sich im Bereich bei 5875 cm^{-1} (1. Oberschwingung der C-H Streckschwingung; $-\text{CH}_3$) und 5680 cm^{-1} (1. Oberschwingung der C-H Streckschwingung; $-\text{CH}_2-$). Eine weitere Lipid-bezogene Spektralregion lag bei ca. 4335 cm^{-1} und resultierte aus der 2. Oberschwingung der C-H Deformationsschwingung. Die Absorptionsbande in der Region von 4985 bis 4515 cm^{-1} wurde durch die Kombination verschiedener Schwingungen der Amidgruppen in Proteinen verursacht^{303,367}.

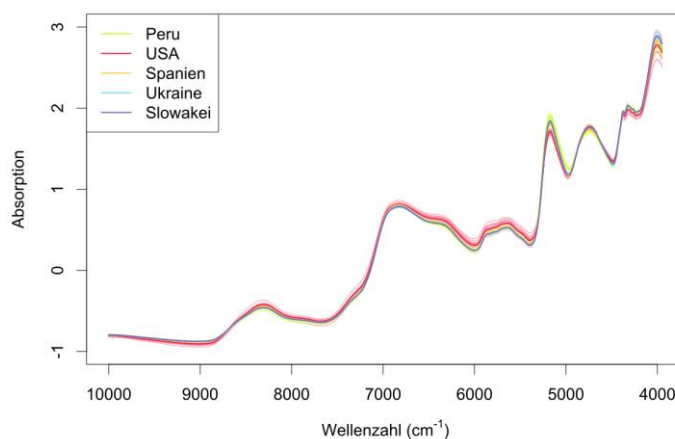


Abb. 6.18: FT-NIR-Spektren der Proben des Herkunftsprobensets nach *Preprocessing* mittels SNV. Das Medianspektrum je Herkunftsländ ist dunkler dargestellt. Abbildung modifiziert nach Schütz *et al.*³⁶⁸.

² Die in den folgenden Abschnitten (6.8 – 6.8.7; Seite 88 bis 97) dargelegten Ergebnisse wurden bereits zuvor weitestgehend publiziert in der Veröffentlichung:

D. Schütz, J. Riedl, E. Achten, M. Fischer

Fourier-transform near-infrared spectroscopy as a fast screening tool for the verification of the geographical origin of grain maize (*Zea mays* L.)

Food Control **2022**, 136, 108892.

6.8.2 Multivariate Datenanalyse der FT-NIR-Spektren von Körnermais

Um einen Überblick über das Potenzial der FT-NIR-Spektroskopie zur Unterscheidung der geographischen Herkunftsorte von Körnermaisproben zu erhalten, wurden die FT-NIR-Spektren einer PCA unterzogen. Der resultierende Scoresplot ist in **Abb. 6.19** abgebildet.

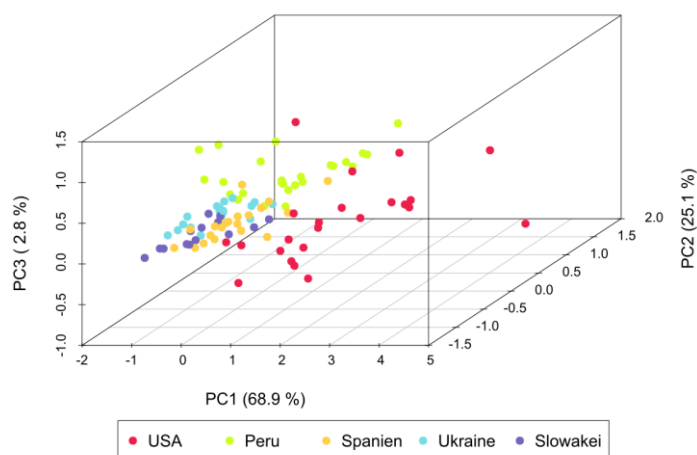


Abb. 6.19: PCA-Scoresplot von SNV preprozessierten FT-NIR-Spektren der Proben des Herkunftsprobensets. Nur die Wellenzahlen $< 10000 \text{ cm}^{-1}$ wurden genutzt ($n = 14\text{-}25$ Proben pro Herkunftsland; das Herkunftsland ist farblich kenntlich gemacht, siehe Legende). Abbildung modifiziert nach Schütz *et al.*³⁶⁸.

Wie **Abb. 6.19** zu entnehmen ist, bildeten die Proben in der PCA-Analyse entsprechend ihrer geographischen Herkunft separate Cluster. Proben aus Spanien, der Ukraine und der Slowakei lagen dabei im Scoresplot relativ dicht beieinander. Sie waren aber dennoch als eigene Cluster erkennbar. Wohingegen die Proben aus den USA und Peru jeweils weiter von den europäischen Proben entfernt positioniert waren. Diese Ergebnisse stimmen mit den beobachteten Unterschieden im Bereich der Wasserbande bei $5235\text{-}4920 \text{ cm}^{-1}$ überein. Weitere Unterschiede wurden für peruanische Proben in der für Proteine spezifischen Bande ($4845\text{-}4645 \text{ cm}^{-1}$) und für die USA-Proben in der Spektralregion bei $6745\text{-}5340 \text{ cm}^{-1}$ gefunden. Letztere Bande steht im Zusammenhang mit dem Stärke- bzw. Lipidgehalt der Proben³⁰³. Gründe für die Unterschiedlichkeit der Proben aus den USA bzw. Peru könnten darin liegen, dass die Proben aus den USA aus insgesamt sechs unterschiedlichen US-Bundesstaaten stammten, welche entlang des Golfes von Mexiko liegen. Die Ost-West-Ausdehnung reichte dabei von Texas bis Georgia (Luftlinie ca. 1400 km). Gleichzeitig wurden relativ viele unterschiedliche Klimazonen abgedeckt, da auch Proben aus Mississippi und Louisiana in dem Probenset enthalten waren. Dies erhöht die Wahrscheinlichkeit, dass die Variabilität innerhalb der Klasse der USA-Proben stärker ausfällt, als in den anderen Probenklassen. So wurden bspw. die Proben aus Spanien alle in Lleida gesammelt. Weiterhin stammten die Proben aus der Slowakei, ebenfalls aus einem relativ kleinen Gebiet. Dies war tendenziell auch aus dem PCA-Scoresplot ersichtlich.

Gründe, für die ebenso beobachteten Unterschiede der peruanischen Proben, könnten in der Nähe der Anbauregion zum Äquator liegen. Zudem stammten die Proben aus einer Küstenregion zwischen dem Südpazifik und den Anden. Diese besondere geographische Lage könnte klimatische Umstände zur Folge haben, welche das Metabolom der Körnermaisproben in Relation zu den Proben aus anderen Ländern relativ stark abweichend beeinflussten.

6.8.3 Unterschiede in den FT-NIR-Spektren je nach Herkunft

Varianzen zwischen den FT-NIR-Spektren von Körnermaisproben unterschiedlicher geographischer Herkunft konnten anhand der ersten drei Loadingsvektoren der PLS-DA ausfindig gemacht werden (siehe **Abb. 6.20**). Da die PLS-DA die Körnermaisproben nach der größten Varianz in den Spektren zwischen den Herkunftsorten trennt, haben Wellenzahlen mit großem Einfluss im Loadingsplot, eine hohe Diskriminierungsfähigkeit zwischen den Herkunftsländern.

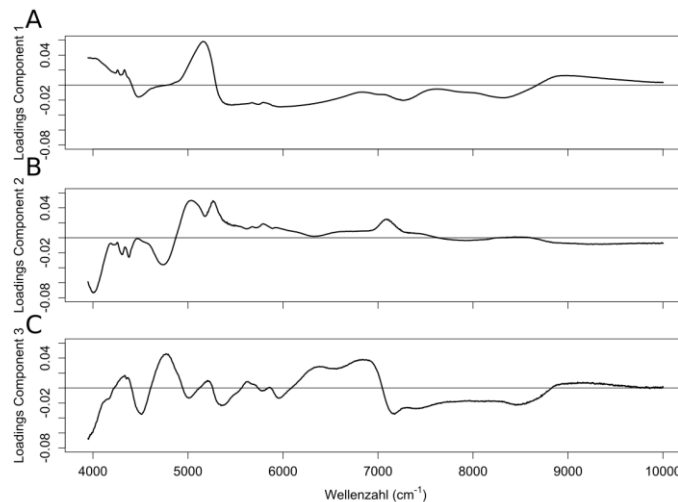


Abb. 6.20: PLS-DA Loadingsvektoren der Trennung nach der geographischen Herkunft. Dargestellt sind die ersten drei Komponenten (A-C). Alle Spektren wurden zuvor mittels SNV preprozessiert. Abbildung modifiziert nach Schütz *et al.*³⁶⁸.

Geeignete Spektrenbereiche für eine Herkunftsunterscheidung von Körnermaisproben wurden hauptsächlich im Bereich von niedrigeren Wellenzahlen ($< 6000 \text{ cm}^{-1}$) detektiert. Nur wenige Spektrenbereiche im Bereich von 6000 bis 8500 cm^{-1} waren besonders für die Herkunftsbestimmung geeignet. Wellenzahlen, die einen starken Einfluss auf die Trennung der Proben entsprechend ihrer Herkunft in der PLS-DA ausübten, waren assoziiert mit den Absorptionsbanden von Lipiden, Stärke und Proteinen sowie von der Feuchtigkeit^{303,367}.

Insbesondere die Spektralregionen $5860\text{--}5775$, $4575\text{--}4450$ und $4270\text{--}4240 \text{ cm}^{-1}$ waren konsistent mit den Wellenzahlen, welche von Zhou *et al.* für die Herkunftsbestimmung von DDGS angegeben wurden¹⁴⁹. Sie berichteten ebenso von der Wichtigkeit des Öl-, Stärke- sowie des Proteingehalts zur Herkunftsbestimmung von DDGS mittels FT-NIR-Spektroskopie¹⁴⁹. González-Martín *et al.* stellten fest, dass die Spektralregion $7142\text{--}5376 \text{ cm}^{-1}$ bei Weizenmehlproben aus unterschiedlichen geographischen Regionen in Chile unterschiedliche Absorptionen aufwies¹⁵⁵. Diese Ergebnisse stimmten mit den Ergebnissen dieser Arbeit überein, die ebenso diese Region des FT-NIR-Spektrums als variabel entsprechend dem Körnermais-Anbauort hervorheben (siehe **Abb. 6.20**).

Wie die Resultate aus **Abb. 6.20** zeigen, befanden sich geeignete Regionen zur Herkunftsunterscheidung von Körnermais eher im Bereich niedriger Wellenzahlen, insbes. bei den beiden Absorptionsbanden von Wasser³⁶⁷ ($7290\text{--}6530$ und $5400\text{--}4970 \text{ cm}^{-1}$). Üblicherweise wird Körnermais nach der Ernte durch maschinelle Trocknung auf einen niedrigeren Wassergehalt eingestellt, um eine längerfristige Lagerung zu ermöglichen. Die maximal tolerable Restfeuchte ist jedoch von der Lagerungsdauer abhängig und beträgt $< 15 \%$ für kurzzeitige Lagerung, bis $< 13 \%$

für längerfristige Lagerung (> 1 Jahr)⁷⁸. Zudem ist der Trocknungsprozess vergleichsweise teuer, da er sehr energieintensiv ist. Folglich kann der Wassergehalt von Körnermais nicht als konstant angenommen werden, wie es für eine Verwendung als Indikator für die geographische Herkunft notwendig wäre. Daher wurden die beiden Wasserbanden aus dem FT-NIR-Spektrum nicht für die Erstellung des SVM-Modells berücksichtigt.

6.8.4 Auswirkungen des *Preprocessing* von Spektren auf die Herkunftsbestimmung

FT-NIR-Spektren werden regulär einem sogenannten *Preprocessing* unterzogen, um multiplikative Streueffekte abzumildern, die insbesondere bei der Analyse von pulverförmigen Proben auftreten²⁴⁸. In dieser Arbeit wurde der SNV-Algorithmus als *Preprocessing*-Methode genutzt, da er einerseits sehr populär ist, die Spektrenform weitestgehend beibehält und im Gegensatz zu dem ebenfalls populären MSC-Algorithmus²⁵⁶, kein Referenzspektrum benötigt. Neben einem *Preprocessing* der Spektren, welches Streueffekte reduziert, ist die Nutzung der Ableitungen von Spektren üblich²⁴⁸.

Um den Einfluss von unterschiedlichen *Preprocessing*-Strategien auf die SVM-Klassifizierungsergebnisse vergleichen zu können, wurden die Spektren der Herkunftssprobensets neben SNV unterschiedlichen Vorverarbeitungsstrategien vor der Klassifikation unterzogen (1. bzw. 2. Ableitung, *Binning*). Die resultierenden Spektren sind im **Anhang in Abb. 13.15** gezeigt.

Die SVM-Klassifikation erlaubt die Abtrennung von Punktwolken voneinander durch verschiedene Kernel-Funktionen, die auch nicht-lineare Grenzverläufe abbilden können^{274,369–371}. Da die genutzte Kernelfunktion einen Einfluss auf die SVM-Klassifizierungsergebnisse ausübt³⁷², wurden unterschiedliche Kernel-Funktionen (linear, quadratisch, RBF) evaluiert.

Die Klassifizierung mittels SVM benötigt zudem mehrere Parameter, die zu optimieren sind: Neben C , dem *Cost*-Parameter, welcher die Feinheit der Klassifikationsgrenze definiert, gibt es zum Teil zusätzliche Kernel-Funktionsparameter³⁷². Die erhaltenen Klassifizierungsgenauigkeiten in Abhängigkeit von dem *Preprocessing* sowie der genutzten SVM Kernel-Funktion sind in **Tab. 6.9** zusammengefasst. Für zugehörige Details zu den genutzten Kernelparametern und weiteren Parametern für die SVM-Klassifikation, siehe **Anhang Tab. 13.15**.

Tab. 6.9: SVM Klassifizierungsgenauigkeiten in Abhängigkeit vom Spektrenpreprocessing bzw. der genutzten Kernel-Funktion. Die mittlere Klassifizierungsgenauigkeit sowie deren Standardabweichung (in Klammern) wurden durch ein *Grid-Search* (10-mal wiederholte 10-fache Kreuzvalidierung) ermittelt. Für die optimalen (Kernel-)Parameter siehe **Anhang Tab. 13.15**. Es wurden Wellenzahlen von $10000\text{--}4000\text{ cm}^{-1}$ zur Klassifikation genutzt. Tabelle nach Schütz *et al.*³⁶⁸.

<i>Preprocessing</i>	Lineare Kernel-Funktion	Quadratische Kernel-Funktion	RBF-Kernel
SNV	92.2 % (1.0 %)	92.2 % (0.9 %)	92.9 % (1.6 %)
SNV + <i>Binning</i>	92.2 % (0.9 %)	93.0 % (0.3 %)	92.8 % (0.9 %)
SNV + 1. Ableitung	95.0 % (1.0 %)	95.1 % (1.1 %)	95.0 % (0.7 %)
SNV + 1. Ableitung + <i>Binning</i>	95.3 % (0.9 %)	95.4 % (1.0 %)	95.1 % (0.9 %)
SNV + 2. Ableitung	94.9 % (0.9 %)	94.3 % (0.6 %)	95.0 % (0.8 %)
SNV + 2. Ableitung + <i>Binning</i>	95.0 % (1.5 %)	95.0 % (1.1 %)	94.7 % (1.3 %)

Wie aus dem *Grid-Search* hervorging (siehe **Tab. 6.9**), waren die erreichten Klassifizierungsgenauigkeiten eher abhängig von dem *Preprocessing* der Spektren, als von der verwendeten SVM-Kernelfunktion. Die erreichten Klassifikationsgenauigkeiten waren im Falle der 1. oder 2. Ableitung der Spektren vergleichbar, sanken jedoch minimal, falls die Originalspektren zur Klassifikation herangezogen wurden. Ein *Binning* der FT-NIR-Spektren auf 1/5 ihrer Rohdatenpunktzahl (Auflösung 10 cm^{-1} anstelle von 2 cm^{-1}) führte zu keiner Verschlechterung der Genauigkeit der Klassifikationsmodelle. Die Art der genutzten Kernelfunktion hatte nur einen marginalen Einfluss auf die erzielte Genauigkeit der Klassifikation. Die besten Resultate wurden mittels eines RBF-Kernels oder einer linearen Kernelfunktion erzielt. Aus Gründen der einfacheren Optimierung der Parameter, wurde eine lineare Kernelfunktion genutzt, um Modelle zur Herkunftsunterscheidung zu erstellen. Die Spektren wurden mittels SNV, 1. Ableitung und einem *Binning* preprozessiert, um eine optimale Vorbereitung für die SVM-Klassifikation zu erreichen.

6.8.5 Einfluss der verwendeten Spektrenbereiche auf das SVM-Klassifikationsmodell für die Herkunftsbestimmung

Es erschien vorteilhaft, dass ein Klassifikationsmodell für die Authentizität von Körnermaisproben einerseits eine geringfügig unterschiedliche Restfeuchte der Körnermaisproben toleriert. Andererseits trägt eine möglichst breite Verteilung der zur Klassifikation eingesetzten FT-NIR-Messdaten auf verschiedene chemische Substanzen, welche die NIR-Banden erzeugen, zu einem verlässlichen Modell bei.

Um das Potenzial bestimmter Bereiche des FT-NIR-Spektrums für die Bestimmung der geographischen Herkunft zu untersuchen, wurden verschiedene SVM-Modelle mit einer linearen Kernelfunktion verglichen. Dabei wurde ein Modell (M1), welches auf dem kompletten Spektrum basierte, mit einem Modell (M2), welches nicht die Wasserbanden enthielt sowie weiteren Modellen (M3-M6), die auf einzelnen Absorptionsbanden von Lipiden, Proteinen oder Stärke beruhten, verglichen. Darüber hinaus wurde ein kombiniertes Modell (M7), welches die Absorptionsbanden von Lipiden, Proteinen und Stärke gemeinsam nutzte, erstellt. Die resultierenden Klassifikationsgenauigkeiten sind in **Tab. 6.10** dargestellt. Für die Klassifikationsmatrizen siehe **Anhang Abb. 13.16**.

Tab. 6.10: Auflistung der SVM-Modelle, welche auf unterschiedlichen Spektrenregionen beruhen. Dargestellt sind die Spektrenregionen, Wellenzahlenbereiche sowie die mittlere Klassifikationsgenauigkeiten (100-fach wiederholt, 10-fach kreuzvalidiert) für die Zuordnung der Herkunftsländer auf globaler Ebene. Tabelle nach Schütz *et al.*³⁶⁸.

Modell	Spektrenregion	Wellenzahlen (Anzahl Datenpunkte)	CC [%] (Standardabweichung)
M1	Komplettes Spektrum	10000-4000 cm ⁻¹ (589)	95.4 % (0.8 %)
M2	Spektrum ohne Wasserbanden*	10000-4000 cm ⁻¹ * (471)	92.5 % (1.3 %)
M3	Protein-assoziierte Bande	4960-4490 cm ⁻¹ (47)	90.0 % (1.1 %)
M4	Stärke-assoziierte Bande	6520-6010 cm ⁻¹ (51)	87.2 % (1.2 %)
M5	Lipid-assoziierte Bande I	5990-5420 cm ⁻¹ (57)	91.8 % (0.9 %)
M6	Lipid-assoziierte Bande II	8600-7980 cm ⁻¹ (61)	89.3 % (1.8 %)
M7	Kombination der Banden assoziiert mit Proteinen, Stärke und Lipiden	8600-7980, 6520-6010, 5990-5420, 4960-4490 cm ⁻¹ (216)	94.7 % (1.2 %)

* ohne Wellenzahlenbereiche spezifisch für Wasser (7290-6530 und 5400-4970 cm⁻¹).

Die beste Klassifikationsgenauigkeit konnte mit dem Modell M1 mit 95.4 %, welches auf dem kompletten Spektrenbereich von 10000-4000 cm⁻¹ basierte, erzielt werden. Die erzielte Rate an korrekten Klassifikationen fiel geringer aus, falls die Wasserbanden nicht berücksichtigt wurden. Dieser Umstand war aufgrund deren starken Einfluss im PLS-DA-Loadingsplot (siehe **Abb. 6.20**) zu erwarten. Der Anteil von falsch zugeordneten Proben auf Länderebene blieb dabei für beide Modelle M1 oder M2 ähnlich (siehe **Anhang Abb. 13.16** und **Abb. 13.17**). Während die Klassifikation von Proben aus den USA bzw. Peru sehr genau erfolgte (CC von 100 % für beide Probenklassen), war im Falle der europäischen Proben die Zuordnung zu den entsprechenden Ländern mit CC-Werten zwischen 96-68 % weniger genau. Das Klassifikationsergebnis der slowakischen Proben verschlechterte sich aufgrund der Entfernung der Wasserbanden aus dem Modell von 85 % auf 68 %.

Die Klassifikationsmodelle, die auf bestimmten Spektrenbereichen beruhten, welche die Absorptionsbanden für Lipide, Proteine oder Stärke enthielten (M3-M6), erzielten Klassifikationsgenauigkeiten von 87-92 % und waren damit ebenfalls weniger genau als das Modell M1, welches auf dem gesamten FT-NIR-Spektrum basierte.

Diese Modelle legen den Einfluss der spezifischen Absorptionsbanden auf die korrekte Zuordnung der Proben zu ihrem Herkunftsland offen: Die Nutzung der mit Stärke assoziierten Absorptionsbande (M4) führte zu einer in Relation geringsten Genauigkeit bei der Vorhersage des Herkunftslandes (siehe **Tab. 6.10**). Jedoch wurde bei diesem Modell die höchste Genauigkeit für die Probenklassen USA und Peru erzielt, verglichen mit den Modellen M3, M5 und M6. Dafür ist die Spezifität für die Differenzierung der europäischen Proben bei dem Modell M4 am geringsten ausgeprägt. Geringfügig bessere Klassifikationsraten konnten dagegen mittels den Protein- oder Lipid-assoziierten Absorptionsbanden (M3, M5 und M6) erzielt werden.

Mithilfe der Lipid-Absorptionsbande bei ca. 8350 cm⁻¹ (M6) konnte die beste Trennung der Proben aus der Ukraine und der Slowakei vorgenommen werden. Allerdings wurden dabei häufiger spanische Proben falsch klassifiziert. Diese Ergebnisse suggerieren, dass der Protein- bzw. der Ölgehalt von Körnermais möglicherweise besser dazu geeignet sind, die geographische Herkunft zu

bestimmen, als der Stärkegehalt. Das Fettsäurespektrum wurde beispielsweise erfolgreich von Tres *et al.* zur geographischen Herkunftsbestimmung von DDGS auf Weizen- bzw. Maisbasis genutzt¹⁴⁵ sowie zur geographischen Herkunftsbestimmung von Pistazien³⁷³.

Mithilfe einer Kombination aus den Absorptionsbanden von Proteinen, Stärke und Lipiden (M7) konnte eine vergleichbare Klassifikationsgenauigkeit ($94.7 \pm 1.2 \%$), wie für das Modell, welches auf den kompletten Spektren basierte (M1) erzielt werden. Dabei trat ein ähnliches Muster in Bezug auf die CC per Herkunftsland auf, wie beim Modell M1. Während Proben aus den USA und Peru sich vergleichbar zu M1 verhielten, zeigte sich für die Slowakei-Proben eine Verschlechterung der CC (vermutlich aufgrund des Vernachlässigens der Informationen aus den Wasserbanden) und für die Ukraine-Proben eine Verbesserung der erzielten Genauigkeiten in Relation zum Modell M1.

Die Nutzung von lediglich relativ kleinen Bereichen eines FT-NIR-Spektrums zur algorithmenbasierten Klassifikation von Proben ist in der Literatur häufig anzutreffen: De Girolamo *et al.* evaluierten größere Spektrenabschnitte auf ihre Eignung zur Herkunftsbestimmung von Hartweizen¹⁵⁶. Zhang *et al.* verwendeten eine Kombination von vielen kleineren Spektrenausschnitten zur Herkunftsbestimmung von Stärke, welche aus *Radix puerariae* gewonnen wurde³⁷⁴. Ein FT-NIR basiertes Klassifikationsmodell, das auf Metabolit-assoziierten Spektrenregionen basiert, ist daher möglicherweise zur Verifikation der geographischen Authentizität von Körnermaisproben vielversprechend. Um eine hohe Klassifikationsgenauigkeit zu erreichen, sollte der Einfluss der Probenfeuchtigkeit sowie des Hintergrundrauschens in absorptionsfreien Bereichen des Spektrums reduziert werden. Daher wurde das Modell M7, welches eine Kombination aus den absorptionsreichen Bereichen des FT-NIR-Spektrums, die durch Proteine, Stärke sowie Lipide in den Körnermaisproben hervorgerufen werden, als optimal ausgewählt.

6.8.6 Optimiertes SVM-Klassifikationsmodell für die Herkunftsbestimmung

Das optimierte Klassifikationsmodell M7 basiert auf einer Kombination der NIR-Spektrenbereiche, denen die Absorptionsbanden von Proteinen, Stärke und Lipiden zuzuordnen sind. Dabei konnte eine hohe Klassifikationsgenauigkeit von $94.7 \pm 1.2 \%$ im Rahmen einer 100-fach wiederholten 10-fachen Kreuzvalidierung erzielt werden. Auf Länderebene betrachtet, wurden die Proben aus den USA und Peru je zu 100 % korrekt klassifiziert (siehe **Abb. 6.21**), obwohl diese Proben eine starke Streuung im PCA-Scoresplot aufwiesen und die USA-Proben aus insgesamt sechs unterschiedlichen US-Bundesstaaten stammten. Innerhalb der europäischen Proben wurden gute Ergebnisse für die Proben aus Spanien sowie der Ukraine erzielt. Die Proben aus der Slowakei konnten weniger gut vorhergesagt werden.

	PER	SLO	SPA	UKR	USA	CC
PER	25.00	0	0	0	0	100.00 %
SLO	0	9.28	1.00	3.72	0	66.29 %
SPA	0	0.21	19.77	0.02	0	98.85 %
UKR	0	0.37	0	18.63	0	98.05 %
USA	0	0	0	0	23.00	100.00 %

Abb. 6.21: SVM-Klassifikationsergebnisse bei kombinierter Nutzung von Spektrenbereichen, welche auf den Absorptionsbanden von Proteinen, Stärke und Lipiden basierten.

Die Herausforderung, Proben aus der Ukraine und der Slowakei auseinander zu halten, wurde als Hauptursache für das Auftreten von falsch zugeordneten Proben innerhalb des Subsets der europäischen Proben identifiziert. Während die CC für die ukrainischen Proben relativ hoch ist, wurden gleichzeitig relativ viele slowakische Proben irrtümlich als aus der Ukraine klassifiziert. Dies führte zu einer vergleichsweise hohen Fehlerrate für die slowakischen Proben.

Gründe für die Schwierigkeiten, die Proben aus der Ukraine und der Slowakei zu unterscheiden, könnten in der nur geringen geographischen Distanz zwischen den beiden Nachbarländern begründet sein. Unterschiede in der Probenzusammensetzung und als weitere Folge in den FT-NIR-Spektren, entstehen aufgrund von verschiedenen klimatischen Bedingungen oder aufgrund anderer umweltbedingter Einflüsse, wie bspw. der Umgebungstemperatur, der Anbauhöhe, sowie der Wasserverfügbarkeit oder der Bodenzusammensetzung¹¹⁹. Den politischen Staatsgrenzen kommt diesbezüglich keinerlei Bedeutung zu.

Zusammen legen die Ergebnisse den Schluss nahe, dass eine korrekte Zuordnung von Proben aus unterschiedlichen Kontinenten erfolgversprechender war, als von Proben von demselben Kontinent (Europa) auf Länderebene.

6.8.7 Direktes vs. Zwei-Schritt SVM-Klassifikationsmodell für die Herkunftsbestimmung

Sowohl die Ergebnisse der multivariaten Datenanalyse als auch der SVM-Klassifikation, ließen vermuten, dass die Körnermaisproben aus den USA sowie Peru leichter von den Proben, welche aus Europa stammten, abgetrennt werden könnten. Daher wurde neben der Möglichkeit ein einziges SVM-Modell mit allen fünf Ländern zu erstellen, auch ein schrittweises Modell evaluiert. Dieses sollte zunächst die Proben nach Kontinenten trennen und in einem zweiten Schritt die europäischen Proben entsprechend dem Land differenzieren. In **Abb. 6.22** sind PCA-Scoresplots dargestellt, welche die Varianz der Proben je Kontinent bzw. innerhalb von Europa visualisieren.

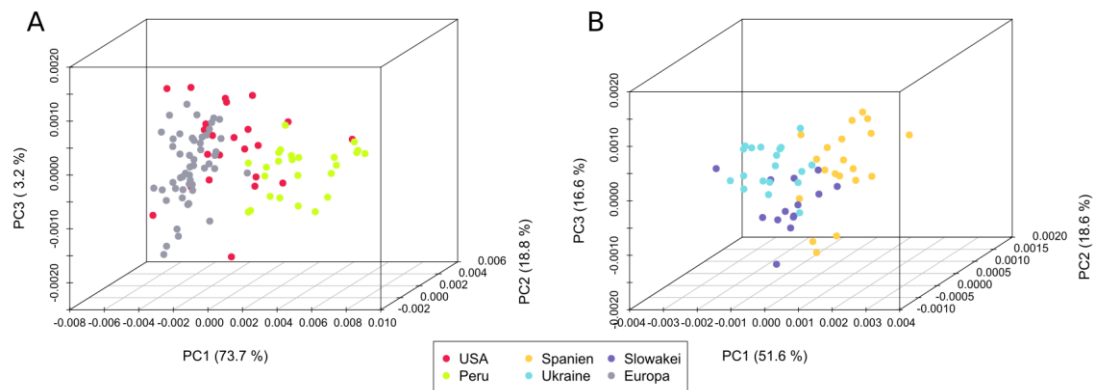


Abb. 6.22: PCA-Scoresplots der Körnermaisproben auf kontinentaler Ebene sowie auf Länderebene. **A:** Herkunftsorte USA, Peru und Europa. **B:** Europäische Herkunftsorte Spanien, Slowakei und Ukraine. Für die beiden Modelle wurden lediglich die Absorptionsbanden von Proteinen, Stärke und Lipiden genutzt. Abbildung modifiziert nach Schütz *et al.*³⁶⁸.

Auf Ebene der Kontinente separierten sich die Proben eindeutig im PCA-Scoresplot (siehe **Abb. 6.22A**). Die Klassifikationsergebnisse unterstützten die Annahme, dass die Körnermaisproben zuverlässig entsprechend ihren Herkunftskontinenten Europa, Nord- und Südamerika zugeordnet werden konnten, ohne dass Fehlklassifikationen auftraten. Daneben bildeten die Proben aus Europa ebenso Cluster im Scoresplot entsprechend ihrem Herkunftsland (siehe **Abb. 6.22B**). In letzterem Fall war jedoch ein gewisser Überlappungsanteil der Probenklassen erkennbar.

Die insgesamt erreichte Klassifikationsgenauigkeit des schrittweise evaluierten Modells betrug 93.8 %. Dieses Ergebnis ist vergleichbar mit der Genauigkeit des direkten Modells. Die Rate an korrekt klassifizierten Proben je Land war dabei ungleich verteilt: Während die Zuordnung der Proben zu den Kontinenten zu jeweils 100 % korrekt erfolgte, sank die Genauigkeit für die europäischen Proben auf lediglich 88.1 ± 1.9 % ab. Während in beiden Modellen, Proben aus den USA, Peru und Spanien hohe CC-Raten von > 97 % aufwiesen, war die Rate für ukrainische sowie slowakische Proben geringer. In dem direkten Modell wurden fast alle ukrainischen Proben – auf Kosten von Falschzuordnungen der slowakischen Proben – korrekt zugeordnet. Dagegen waren die Falschzuordnungen in dem europäischen Modell gleichmäßiger auf die Probenklassen verteilt (CC von 88 % und 75 % für ukrainische resp. slowakische Proben). Dabei schien eine Tendenz zu bestehen, dass ukrainische und slowakische Proben mittels der etablierten Klassifikationsmodelle nicht voneinander unterscheidbar waren. Es ist anzunehmen, dass sich diese Körnermaisproben analytisch sehr ähnlich sind, was vermutlich in der geringen geographischen Distanz der Staaten begründet ist. Die Überschneidung von FT-NIR-Spektren von Körnermaisproben dieser beiden Staaten kann daher vermutlich nicht bloß mithilfe der Optimierung der Klassifikationsmodelle gelöst werden. Voraussichtlich ist zum Herausstellen von Unterschieden, eine größere Datenbasis, als hier mit 14 bzw. 19 Proben je Land gegeben ist, notwendig.

Die Ergebnisse der schrittweisen Klassifikation der Körnermaisproben zeigten, dass die Zuordnung der Proben zu ihrem Ursprungskontinent verlässlicher erfolgte, als die Bestimmung des jeweiligen Herkunftslands. Die Genauigkeit der Klassifikation nahm für die europäischen Proben dabei nicht zu, falls anstelle des einstufigen Modells mit allen Herkunftsländern, das zweistufige Modell genutzt wurde. In der Literatur wurde eine Unterscheidung von Proben nach Kontinenten

beispielsweise für DDGS-Proben¹⁴⁷ und Körnermaisproben³⁷⁵ mittels FT-IR-Spektroskopie erzielt. Studien, in denen unterschiedliche räumliche Auflösungen anhand derselben Proben untersucht wurden, sind selten. De Girolamo *et al.* erzielten bessere Ergebnisse bei der Differenzierung von Hartweizen mittels FT-NIR aus Italien vs. non-Italien, als bei der Unterscheidung verschiedener Anbauregionen innerhalb Italiens¹⁴⁷.

Zusätzlich ist anzumerken, dass die Proben aus Spanien, die sich aus 20 unterschiedlichen Körnermaissorten zusammensetzten, die geringste Rate an fehlerhaften Klassifizierungszuordnungen aller Herkunftsländer innerhalb Europas aufwiesen. Diese Ergebnisse lassen vermuten, dass sortenspezifische Einflüsse die Verifikation der geographischen Herkunft von Körnermaisproben mittels FTR-NIR-Spektroskopie und SVM-Klassifikation nicht negativ beeinflussen.

Unterschiede in der Probenzusammensetzung können aufgrund von unterschiedlichen Umwelteinflüssen auftreten¹¹⁹. Folglich sind Unterschiede in den FT-NIR-Spektren als Folge wahrscheinlicher, falls eine größere geographische Distanz zwischen den Probengruppen liegt, als eine geringere. Diese Erklärung wird durch die Ergebnisse der Klassifizierungsmodelle gestützt: Eine Separation von Proben auf unterschiedlichen Kontinenten ist ohne bedeutende Fehlerrate möglich, im Gegensatz zu der Unterscheidung von Proben von demselben Kontinent (Europa). Somit kann geschlossen werden, dass die FT-NIR-Spektroskopie zur Herkunftsbestimmung von Körnermais verlässlicher auf größere geographische Entfernungen durchgeführt werden kann, als auf Länderebene.

6.8.8 Robustheit des Klassifikationsmodells bei Vorhersage weiterer Proben

Um die generelle Eignung der FT-NIR-Spektroskopie zur Herkunftsbestimmung von Körnermais zu zeigen, wurden zusätzlich weitere Proben aus Frankreich (Region Elsaß) bzw. Ungarn analysiert (je $n = 36$). Diese Proben wurden einer etwas anderen Probenvorbereitung unterzogen: Die Proben wurden in flüssigem Stickstoff schockgefroren und dann in einer Messermühle unter Zusatz der gleichen Menge an Trockeneis vermahlen und anschließend gefriergetrocknet.

Eine Vermahlung unter Zusatz von Trockeneis kann vorteilhaft sein, um das Metabolitprofil in seinem natürlichen Zustand zu bewahren, da die Probe während des Vermahlungsprozesses weniger Hitze ausgesetzt ist. Die Gefrier Trocknung von Proben reduziert eine möglicherweise enzymatisch oder mikrobiologisch hervorgerufene Veränderung der Probe, indem deren Wassergehalt herabgesetzt wird¹¹⁹. Somit können weniger stark veränderte, homogenisierte Proben erhalten werden, wobei jedoch ein erhöhter Zeitaufwand für die Probenvorbereitung in Kauf genommen werden muss²⁵⁵.

Andererseits ist eine gewisse Toleranz des Klassifikationsmodells im Hinblick auf den Wassergehalt der Proben erstrebenswert. Zum einen könnte dieser Schwankungen unterliegen, die nicht unbedingt mit der geographischen Herkunft assoziiert sind, zum anderen sind die Wasserbanden in FT-NIR-Spektren relativ intensiv³⁶⁷. So könnte eine zeitaufwendige Gefrier Trocknung vermieden werden.

Um daher das Verhalten der entwickelten Klassifikationsmodelle gegenüber den in den Proben enthaltenen Restfeuchte zu untersuchen, wurden die zusätzlichen Proben aus Frankreich/Ungarn

den Klassifikationsmodellen hinzugefügt. Der Cost-Parameter *C* wurde erneut optimiert, und die erzielten Klassifikationsgenauigkeiten wurden zusammen mit ihrer Standardabweichung in **Tab. 6.11** dargestellt.

Tab. 6.11: SVM-Klassifikationsgenauigkeiten (100-mal wiederholte 10-fache Kreuzvalidierung) incl. der Proben aus Frankreich bzw. Ungarn. Die Genauigkeit der Klassifikation ist sowohl über alle Probenklassen, als auch individuell für jede Probenherkunft einzeln aufgeschlüsselt.

	Mittlere Klassifikationsgenauigkeit (Standardabweichung)						
	PER	USA	SPA	UKR	SLO	FRA	UNG
Globales Modell							
Komplettes Spektrum (10000-4000 cm ⁻¹)	90.2 % (1.0 %)						
	100 %	100 %	92.9 %	84.5 %	72.4 %	88.9 %	87.1 %
Optimiertes Modell (Spektrenbereiche für Proteine, Stärke und Lipide)	87.2 % (0.6 %)						
	100 %	100 %	99.2 %	89.4 %	41.9 %	78.4 %	88.9 %
Europäisches Modell							
Komplettes Spektrum (10000-4000 cm ⁻¹)	87.1 % (0.9 %)						
	-/-	-/-	99.6 %	83.5 %	82.1 %	84.5 %	86.5 %
Optimiertes Modell (Spektrenbereiche für Proteine, Stärke und Lipide)	82.0 % (0.8 %)						
	-/-	-/-	99.7 %	83.5 %	44.9 %	79.1 %	88.6 %

Bei der Erweiterung des Klassifikationsmodells um zwei Probenklassen auf insgesamt sieben Probenklassen, veränderten sich die Klassifizierungsergebnisse der ursprünglichen Probenklassen nicht wesentlich, mit Ausnahme der CC von slowakischen Proben des optimierten Modells, die stark auf unter 45 % abfielen (siehe **Tab. 6.9**). Ursächlich waren dabei insbesondere Schwierigkeiten, diese Proben von ukrainischen Proben zu unterscheiden. Die Gesamtklassifikationsgenauigkeiten aller Probenklassen fielen um 3-5 % leicht ab. Jedoch waren die Proben aus Frankreich bzw. Ungarn mit ähnlich guten CC im Bereich > 78 % wie die übrigen Proben zuzuordnen. Interessanterweise zeigte sich, dass die beiden unterschiedlich vorbereiteten Probenvarianten in den SVM-Modellen immer separat für sich blieben. Proben aus Frankreich/Ungarn wurden somit ausnahmslos entweder als französische oder als ungarische Proben klassifiziert (siehe **Anhang, Abb. 13.18**). Daher ist ein Vergleich der zusätzlich klassifizierten Proben aus Frankreich/Ungarn mit den restlichen Proben nur bedingt bis gar nicht aussagekräftig. Gründe für das unterschiedliche Verhalten der Körnermaisproben aus Frankreich/Ungarn können entweder aus der anderen Art der Probenvorbereitung hervorgehen oder auf unterschiedliche chemische Profile zurückzuführen sein, da diese Proben zur Stärkegewinnung und nicht zur Nutzung als Futtermittel kultiviert wurden. Dies lässt auf eine geringere Variabilität der Proben aus Frankreich/Ungarn, welche schockgefroren und unter Zusatz von Trockeneis vermahlen wurden, schließen; was mit Ergebnissen von Arndt *et al.*²⁵⁵ übereinstimmte.

Die generelle Eignung der FT-NIR-Spektroskopie zeigte sich jedoch auch darin, dass die Separation der Proben aus Frankreich/Ungarn voneinander, hingegen mit einer Genauigkeit von 95.4 % ± 1.1 % (SVM-Klassifikation, 100-mal wiederholte, 10-fache Kreuzvalidierung, *Cost*

$C = 0.0062$) durchgeführt werden konnte, falls das komplette akquirierte FT-NIR-Spektrum von 10000 cm^{-1} bis 4000 cm^{-1} genutzt wurde. Es zeigte sich, dass die Egalisierung der enthaltenen Wasserbanden durch Gefriertrocknung der Proben, nicht zu einer Verschlechterung der Trennleistung im Vergleich zu den anderen Modellen führte. Somit unterstreichen diese Ergebnisse, dass die Verifikation der geographischen Herkunft von Körnermais unabhängig von dem Wassergehalt sowie der Vermahlungsart der Proben, mittels FT-NIR-Spektroskopie erfolgen kann.

6.9 Fusion von LC-MS- und FT-NIR-Messdaten

Im Folgenden wurde eine *low-level* Datenfusion der selektierten 20 Markermetabolite sowie die Kombination aus spezifischen FT-NIR-Banden der zuvor optimierten Einzelmethode durchgeführt. Um ein eventuelles *Bias* der Klassifikationsmethode aufgrund der unterschiedlichen Analysemethoden zu vermindern, wurden dabei zunächst die Wertebereiche der Analysedaten der beiden Methoden aneinander angeglichen (siehe **Abschn. 6.9.1**). Die anschließende Klassifikation der Proben nach Fusion der Markermatrizen erfolgte mittels des RF-Algorithmus^{281,282,286}.

6.9.1 Datenvorbereitung für die *low-level* Fusion

Die *low-level* Datenfusion ermöglichte es, die Performance der Klassifikation mittels LC-MS- bzw. FT-NIR-Daten in direkter Konkurrenz zu analysieren. Vor der Fusion der Daten wurden die mittels SNV, 1. Ableitung und *Binning* preprozessierten FT-NIR-Messdaten an den Messbereich der LC-MS-Messdaten angepasst: Da q-ToF-Massenspektrometer und FT-NIR-Spektrometer auf unterschiedlichen physikalischen Prinzipien beruhen, decken die gewonnenen Rohdaten unterschiedliche Wertebereiche ab: Während die aufgenommenen Massenspektren theoretisch Signale im Wertebereich von $[0 \dots 4.7 \times 10^6]$ counts aufweisen können, umfassten die FT-NIR-Spektren hingegen einen Bereich von $[0.05 \dots 1.34] \log(1/R)$. Nach *Preprocessing* der FT-NIR-Daten mittels SNV + 1. Ableitung und *Binning* änderte sich dieser Wertebereich zwar zu $[-0.95 \dots +2.96]$, jedoch ist die Inkompatibilität offensichtlich.

Um die Wertebereiche der beiden Datenblöcke anzugleichen, wurden sowohl die Rohdaten der massenspektrometrischen Analyse, als auch die FT-NIR-Daten einem (weiteren) *Preprocessing* mittels des *Range-Scalings* unterzogen. Das *Range-Scaling* skaliert die Analysedaten entsprechend den Schwankungsbreiten innerhalb der Proben²⁴². Daher ist dieses *Preprocessing*-Verfahren besonders geeignet um der Diskriminierung einer der beiden Analysemethoden bei einer Datenfusion entgegenzuwirken.

Da das *Range-Scaling* nach Analysemethoden getrennt erfolgt, ergeben sich nach dem *Preprocessing* ähnliche Wertebereiche für die vorher sehr unterschiedlichen Messwerte. Nach der Skalierung umfasste der Wertebereich der LC-MS-Daten $[-0.54 \dots +0.73]$, jener der FT-NIR-Daten $[-0.76 \dots +0.77]$. Die Spanne der LC-MS-Daten konnte somit an jene der FT-NIR-Daten angeglichen werden. Auf diese Weise war es möglich einen zuvor vorhandenen *Bias* aufgrund der verschiedenen Messbereiche der beiden Methoden im Zuge der Datenfusion deutlich zu reduzieren.

6.9.2 Vergleich des Fusionsmodells mit den Einzelmodellen mittels multivariater Datenanalyse

Um die in den Datensätzen enthaltene Varianz besser abschätzen zu können, wurden die preprozessierten Analysedaten fusioniert und gemeinsam einer PCA unterzogen. Der erhaltene Scoresplot ist in **Abb. 6.23C** gezeigt. Die entsprechenden PCA-Scoresplots der LC-MS-Daten bzw. der FT-NIR-Daten sind in **Abb. 6.23A** und **B** dargestellt.

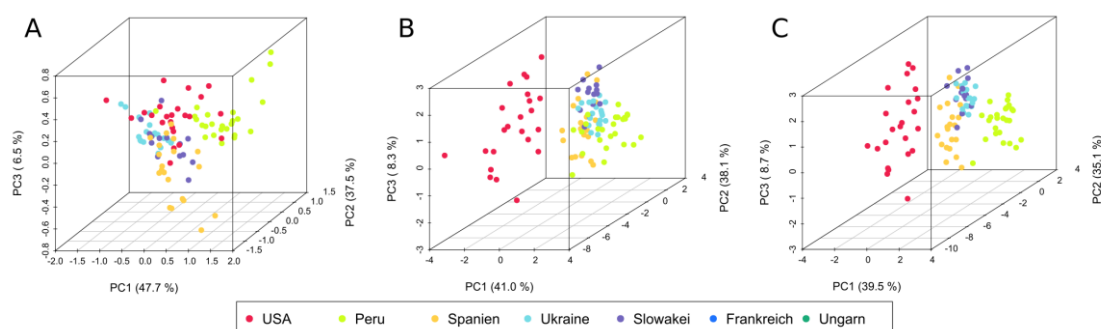


Abb. 6.23: PCA-Scoresplots der mittels *Range-Scaling* preprozessierten und angeglichenen Analysedaten. **A:** LC-MS-Daten (20 Markerfeatures), **B:** FT-NIR-Daten (Spektrenbereiche spezifisch für Proteine, Stärke und Lipide), **C:** fusionierter Datensatz aus beiden Analyseverfahren.

Die Trennung der Proben nach ihrem geographischen Ursprungsort verbesserte sich bei der Verwendung beider analytischen Methoden, in Relation zu den beiden Einzelmethode. Die Ursache dafür kann anhand des Loadingsplots der PCA nachvollzogen werden, der im **Anhang** in **Abb. 13.19** abgebildet ist. In Relation besaßen die Loadingsvektoren der Markerfeatures aus der LC-MS-Analyse relativ große absolute Werte, sie trugen also einen entscheidenden Beitrag zu der Varianz in dem fusionierten Datensatz bei.

6.9.3 Vergleich der Klassifikationsergebnisse des Fusionsmodells mit den Einzelmodellen

Mittels des fusionierten RF-Klassifikationsmodells konnte insgesamt eine Klassifikationsgenauigkeit von $97.3 \pm 0.7 \%$ der 101 Proben des Herkunftsprobensets im Rahmen einer 100-fach wiederholten 10-fachen Kreuzvalidierung erzielt werden. Verglichen mit der Genauigkeit der Einzelmethode, zeigte sich eine verbesserte Klassifikationsleistung: Die auf LC-MS-Daten basierende Methode erreichte lediglich $86.7 \pm 1.8 \%$ korrekte Klassifikationen; die auf FT-NIR-basierende Methode hingegen erreichte $92.3 \pm 1.1 \%$ korrekte Klassifikationen, bei Verwendung des RF-Algorithmus. Die zugehörigen detaillierten Klassifikationsmatrizen sind in **Abb. 6.24** abgebildet.

A							B							C						
	PER	SLO	SPA	UKR	USA	CC		PER	SLO	SPA	UKR	USA	CC		PER	SLO	SPA	UKR	USA	CC
PER	23.89	0	0	1.00	0.11	95.56 %	PER	25.00	0	0	0	0	100.00 %	PER	25.00	0	0	0	0	100.00 %
SLO	0	9.08	2.95	0.08	1.89	64.86 %	SLO	0	9.09	1.93	2.98	0	64.93 %	SLO	0	12.29	1.71	0	0	87.79 %
SPA	0.03	0.07	18.98	0	0.92	94.90 %	SPA	0	0.26	19.67	0.02	0.05	98.35 %	SPA	0	0.90	19.09	0	0.01	95.45 %
UKR	0	0.82	0	18.18	0	95.68 %	UKR	0	1.89	0.66	16.45	0	86.58 %	UKR	0	0.10	0	18.90	0	99.47 %
USA	1.34	2.35	0.86	1.00	17.45	75.87 %	USA	0	0	0	0	23.00	100.00 %	USA	0	0	0	0	23.00	100.00 %

Abb. 6.24: Detaillierte Klassifikationsmatrizen der RF-Klassifikation (100-fach wiederholte, 10-fache Kreuzvalidierung) nach *Range-Scaling*. **A:** nur LC-MS-Daten (20 Markerfeatures), **B:** nur FT-NIR-Daten (Spektrenbereiche spezifisch für Proteine, Stärke u. Lipide). **C:** Auf *low-level* Ebene fusionierter Datensatz der Analysedaten von (A) und (B).

Auf Ebene der Länderzuordnung zeigte sich dabei, dass eine Datenfusion insbesondere bei der Klassifikation von Körnermaisproben aus den USA zu Verbesserungen führte. Diese Proben wurden mittels der LC-MS-Einzelmethode relativ unzuverlässig klassifiziert (CC von 75.9 %), während die Zuordnung anhand der FT-NIR-Einzelmethode exakt (CC von 100 %) erfolgte. Umgekehrt konnte die Klassifikation von Proben aus der Ukraine mittels der LC-MS-Methode zuverlässiger erfolgen, als mittels FT-NIR-Spektroskopie. Bei der Klassifikation anhand der fusionierten Datensätze wurde somit die schlechte Trennleistung der einen Methode durch die bessere Trennleistung der anderen Methode ausgeglichen. Dies ist ein Beispiel dafür, wie die zueinander orthogonalen Methoden LC-MS und FT-NIR-Spektroskopie bei Anwendung der Datenfusion in kooperativer Art und Weise die Klassifikationsergebnisse verbessern können. Somit resultierte als Endergebnis eine zuverlässigere Klassifikationsleistung des fusionierten Modells.

Aus dem *Ranking* der Variablenwichtigkeit der fusionierten Daten (siehe **Anhang Tab. 13.16**) wurde deutlich, dass sich bei der *low-level* Fusion, die LC-MS-Methode gegenüber der FT-NIR-Methode, trotz vorheriger Angleichung des Wertebereichs der Messwerte durchsetzte. Von den fünf besten Markern bei der Klassifikation stammten vier von der LC-MS-Methode. Dies entspricht einer Bevorzugung der LC-MS-Markerfeatures, da weitaus mehr FT-NIR- als LC-MS-Marker in dem fusionierten Datensatz vorhanden waren. Diese Bevorzugung der LC-MS-Markerfeatures deutet dabei auf eine bessere Leistungsfähigkeit der LC-MS-basierten Analytik im Vergleich zur FT-NIR-basierten Analytik hin.

7 Übergreifende Diskussion

7.1 Vergleich der Eignung der Herkunftsverifikation von Körnermais mittels LC-MS(/MS) und FT-NIR-Spektroskopie

In dieser Arbeit wurden die LC-MS(/MS)-Analytik und die FT-NIR-Spektroskopie genutzt um die geographische Herkunft von Körnermais analytisch zu bestimmen. Dafür wurden Metabolite des Körnermais extrahiert und (halb-)quantitativ mittels LC-MS(/MS) analysiert. Die FT-NIR-spektroskopische Analyse erfolgte dagegen direkt an der lediglich vermahlenden Körnermaisprobe.

Die FT-NIR-Spektroskopie, ist eine zeit- und kostengünstige Analysemethode, die vor allem in der Cerealienindustrie eine routinemäßige Anwendung findet^{47,366,376}. Der Arbeitsaufwand für die Probenvorbereitung (Vermahlung) ist gering und die Messzeit der FT-NIR-Spektren ist im Vergleich zu anderen apparativen Analysemethoden oder nasschemischen Methoden sehr kurz, da eine Extraktion der Analyten, wie z. B. bei der LC-MS-Analytik, entfällt^{123,376,377}. Ein Einsatz für die Herkunftsbestimmung von Körnermais ist somit naheliegend. Im Gegensatz zur LC-MS-Analyse, die halb-quantitative Aussagen über den Gehalt von bestimmten Metaboliten erbringen kann, ist die Aussagekraft der FT-NIR-Spektroskopie jedoch verhältnismäßig begrenzt^{123,366}. Eine direkte Korrelation von spezifischen Metaboliten mit bestimmten Spektrenbereichen in der FT-NIR-Spektroskopie ist leider nicht möglich³⁶⁶.

Die LC-MS-Methodik ist im Vergleich langsamer als die FT-NIR-Spektroskopie und zudem teurer in der Anschaffung und dem Unterhalt. Dafür ermöglicht diese Analysetechnik ein viel detailliertes Bild auf die Zusammensetzung einer Körnermaisprobe zu erhalten, da die einzelnen Metabolite spezifisch nachgewiesen werden und zudem quantifiziert werden können¹²³. Die Probenvorbereitung (Vermahlung, Gefriertrocknung und Extraktion) der Proben ist ebenfalls zeitaufwendiger^{123,255}.

Im direkten Vergleich ist ersichtlich, dass beide methodischen Ansätze zur Verifikation des Herkunftsortes von Körnermaisproben insgesamt ähnlich gute Klassifizierungsergebnisse nahe 90 % erbringen konnten. Während die FT-NIR-Methode mit 94.7 % für die Futtermais-Proben das beste Ergebnis erzielte, konnte die LC-MS-Methode mit 86.8 % ein ähnlich gutes Resultat erbringen. Unter Einbeziehung der Proben zur Stärkeherstellung aus Frankreich und Ungarn konnte ein besseres Ergebnis von 90.5 % mittels LC-MS-Analysedaten als über FT-NIR-Analysedaten (87.2 %) erzielt werden.

Ein Teil der LC-MS-Markersubstanzen wurde zudem mittels externer Standards quantifiziert, wobei die genutzte LC-MRM-Methode erfolgreich einer Teilvalidierung unterzogen wurde. Die ausgewählten Metabolite waren dabei auch bei Betrachtung der MRM-*Quantifier* potente und valide Markersubstanzen für die Herkunftsbestimmung: Im direkten Vergleich ermöglichten die neun Marker sowohl unter Nutzung der MS¹-Peakflächen, als auch der MS²-Peakflächen mit 87.2 % bzw. 87.1 % korrekten Zuordnungen bei der RF-Klassifikation eine vergleichbare, robuste Herkunftsbestimmung. Dies zeigte, dass die zur Herkunftsbestimmung genutzten Markerlipide nicht nur einer halb-quantitativen Betrachtung der relativen Peakflächen auf MS¹-Ebene, sondern auch einer quantitativen Betrachtung der Markerkonzentrationen standhielten.

Bei genauerer Betrachtung der Klassifikationsergebnisse auf Länderebene wurde deutlich, dass gewisse Unterschiede in den CCs für bestimmte Herkunftsländer zwischen dem LC-MS-basierten Modell und dem auf FT-NIR-Spektroskopie basierenden Modell bestehen. So war anhand von LC-MS-Daten insbesondere die korrekte Zuordnung von Proben aus den USA und der Slowakei nur eingeschränkt möglich, da diese sich in den ausgewählten Metaboliten nur geringfügig unterschieden (CC von 75 % für USA-Proben). Die korrekte Zuordnung der USA-Proben ist jedoch mittels des auf FT-NIR-Daten basierenden Modells problemlos möglich (CC von 100 %). Auch hier bereitet die Zuordnung der Proben aus der Slowakei die meisten Schwierigkeiten, jedoch nicht in Relation zu den USA-Proben, sondern in Abgrenzung zu den übrigen europäischen Proben insgesamt. Daneben war die Zuordnung der Proben aus Frankreich und Ungarn, welche für die Stärkeproduktion genutzt wurden, als Herausforderung anzusehen, was jedoch auch in deren leicht unterschiedlichen Probenvorbereitung, sowie deren anderem Verwendungszweck begründet sein kann.

Insgesamt erscheint die FT-NIR-Spektroskopie eher für die korrekte Verifikation des geographischen Herkunftsortes über größere Distanzen auf Ebene der Kontinente geeignet, während die LC-MS-Analytik eher für die Zuordnung zu einzelnen Ländern geeignet ist. Das schlechte Abschneiden der LC-MS basierten Klassifikation bei der Vorhersage der Proben aus den USA, steht möglicherweise mit den über eine große Fläche verstreuten Orten der Probenahme und der geringen Probenzahl (vier pro Ort) im Zusammenhang. Hier könnte eine größere Probenanzahl pro Anbauort, sowie ggfls. eine feinere Unterteilung der USA in einzelne Bundesstaaten, möglicherweise Abhilfe schaffen.

7.2 Beurteilung einer möglichen Störung der Herkunftsverifikation durch einen Sorteneinfluss

Mithilfe von Proben aus den Landessortenversuchen der Bundesländer Nordrhein-Westfalen, Brandenburg und Bayern konnte gezeigt werden, dass der Anteil herkunftsbezogen-beeinflusster Features bei der LC-MS-Analytik mit 12.0-25.6 % in allen vier Messmethoden, den sortenbezogen-beeinflussten Features (7.8-9.1 %) weitaus überlegen war. Die Messmethoden waren jedoch unterschiedlich anfällig für einen Sorteneinfluss: Während das polare Metabolom sensitiver gegenüber sortenbezogenen Schwankungen war, erwies sich das unpolare Metabolom in dieser Hinsicht als vergleichsweise robust. Gleichzeitig zeigten sich mit ≤ 3 % aller Features nur relativ wenige Metabolite, die sowohl von der Herkunft als auch von der Sorte beeinflusst wurden.

Klassifikationsmodelle, die lediglich Metabolite nutzten, welche signifikant bezüglich der geographischen Herkunft sind, zeigten eine hohe Genauigkeit der Zuordnung nach Herkunft von > 90 % für alle vier Ionenmodi. Mit lediglich 25 Markern aus dieser Subgruppe trat eine Separierung nach Körnermaissorte kaum auf (CC von ≤ 75 %). Insbesondere galt dies für unpolare Metabolite, welche im positiven Ionenmodus detektiert wurden (CC von 63 %).

Für die FT-NIR-Analytik ergab sich ein ähnliches Bild für das Sortenprobenset. 91.6 % der Wellenzahlen waren herkunftsbezogen beeinflusst, sortenbezogen hingegen nur 22.7 %. Zudem zeigte sich, dass der sortenbezogene Einfluss auf lediglich eine der insgesamt neun untersuchten Maissorten (P 8723) zurückzuführen war. Die Klassifikationsergebnisse des Sortenprobensets ergaben zu 92.7 % eine korrekte Zuordnung entsprechend der geographischen Herkunft,

gegenüber nur 39.0 % für die Sorte. Daraus lässt sich schlussfolgern, dass die Maissorte auch bei der Analytik mittels FT-NIR-Spektroskopie unter bestimmten Umständen einen geringen Störfaktor darstellt.

Die Regulation des Metaboloms insbesondere von bestimmten Lipiden in Körnermais erschien somit mehr von klimatischen bzw. allgemeinen Umweltfaktoren abhängig zu sein als von der Genetik. Gleichzeitig konnte in dieser Arbeit keine negative Beeinflussung der Klassifikation durch unterschiedliche Sorten festgestellt werden. So wurden die Proben aus Spanien, die sich aus jeweils verschiedenen Sorten zusammensetzten, zu 96 % (LC-MS) und 99 % (FT-NIRS) korrekt klassifiziert.

Daher konnte der Einfluss der angebauten Körnermaissorte auf die Klassifizierung nach geographischer Herkunft durch zwei unterschiedliche Maßnahmen stark reduziert werden. Zum einen, indem bei der LC-MS-Analytik die Messmethode für unpolare Metabolite im positiven Ionenmodus gewählt wurde, die eine Metabolitfraktion erfasste, welche geringfügiger durch die Körnermaissorte beeinflusst wird. Zum anderen, indem die Gefahr des Sorteneinflusses zusätzlich vermindert wurde, indem generell nur hochsignifikante Marker bezüglich der Herkunft zur Herkunftsverifikation genutzt wurden.

Eine Herkunftsbestimmung von Körnermais mittels LC-MS und FT-NIR-Spektroskopie war somit möglich, auch ohne die in der Regel nicht deklarierte/bekannte genaue Sortenbezeichnung bzw. -zusammensetzung der Charge^{38,39,66,67,69–72} kennen zu müssen. Die prinzipielle Eignung der beiden angewandten apparativen Methoden zur Herkunftsbestimmung konnte anhand von sehr guten Ergebnissen bei der Klassifikation nach der geographischen Herkunft belegt werden.

7.3 Vergleich der Ergebnisse nach Datenfusion mit den Ergebnissen aus LC-MS-Daten und FT-NIRS-Daten

Die Fusion der Marker beider Analysemethoden auf *low-level* Ebene, führte zu einer Verbesserung der Klassifikationsergebnisse auf 97.3 %. Im direkten Vergleich der beiden Einzelmethoden bei der *low-level* Fusion zeigte sich im Rahmen einer multivariaten Datenanalyse mittels PCA, dass Markerfeatures in der LC-MS-Analyse eine stärkere Variabilität bezüglich der geographischen Herkunft von Körnermaisproben aufwiesen, als FT-NIR-Spektren. Somit ist zu vermuten, dass die Methodik der Herkunftsbestimmung mittels LC-MS tatsächlich leicht bessere Ergebnisse, als die FT-NIR-Spektroskopie liefert. Dies ist vor allem im Hinblick darauf zu betrachten, dass die Klassifikation beim LC-MS-Ansatz anhand von nur 20 Variablen erreicht wurde, wohingegen der FT-NIR-Ansatz 216 Variablen nutzte. Auf Ebene der CC je Herkunftsland ist festzustellen, dass beide Analysemethoden jeweils unterschiedliche Stärken, aber auch Schwächen aufwiesen. Durch die Kombination beider Methoden ist der RF-Klassifikationsalgorithmus jedoch dazu in der Lage, die Stärken beider Methoden zu vereinen, so verbesserte sich bspw. für slowakische Proben die CC von je 65 % auf 88 % im fusionierten Modell.

Im Hinblick auf eine Datenfusion ist somit als Fazit festzustellen, dass die Fusion von LC-MS-Daten und FT-NIR-Spektren in bestimmten Anwendungsbereichen vorteilhaft sein kann. Dies gilt insbesondere für Szenarien, wo bspw. die korrekte Herkunft vorgegeben wird, und lediglich eine qualitative Aussage zum Wahrheitsgehalt der Probenherkunft gegeben werden soll.

Eine initiale Analyse könnte beispielsweise mittels FT-NIR-Spektroskopie und *One-vs.-All*-Modellen erfolgen. Ergibt das Analysenergebnis, die Herkunft sei nicht korrekt deklariert, wird zusätzlich eine LC-MS-Analyse durchgeführt. Eine Fusion der FIT-NIR- und LC-MS-Messergebnisse könnte dann mit guter Aussagekraft das tatsächliche Herkunftsland der Probe bestimmen. Somit würde nur in denjenigen Fällen, in denen das *One-vs.-All*-Klassifikationsmodell eine Falschdeklaration anzeigt, eine LC-MS-Analyse erfolgen, die zusätzlichen personellen, analytischen und apparativen Aufwand erzeugt. Dieser zusätzliche Aufwand würde jedoch durch die dann erhaltene genauere Aussage über die geographische Herkunft gerechtfertigt.

8 Ausblick

Die in den vorangegangenen Kapiteln dargestellten und im zeitlichen Rahmen einer Doktorarbeit erzielbaren Ergebnisse legen dar, dass die Analyse des unpolaren Metaboloms mittels LC-MS bzw. die Analyse mittels FT-NIR-Spektroskopie eine Validierung des geographischen Anbauortes von Körnermais ermöglicht. Die selektierten 20 Markermetabolite aus der LC-MS-Analyse bzw. die vier Markerbanden des FT-NIR-Spektrums mit insges. 216 Wellenzahlen, zeigten in Klassifikationsmodellen gute Vorhersagequalitäten. Jedoch können die Konzentrationen der Lipide generell auch in Abhängigkeit von den klimatischen Bedingungen variabel sein^{43,46,346,348}. Somit ist ein wichtiger, noch zu erfolgender Arbeitsschritt die Validierung der Stabilität der Markermetabolite über mehrere Ernteperioden. Analoges gilt für die Konstanz der FT-NIR-Markerbanden⁴⁵, da ein FT-NIR-Spektrum einen Summenparameter für verschiedenste chemische Stoffe bzw. Stoffgruppen darstellt.

Die entwickelten Klassifikationsmodelle sollten in Zukunft auf weitere Herkunftsländer ausgeweitet werden, um noch besser die marktreale Situation abzubilden. Insbesondere ist hierbei China zu nennen, dass 2019/20 den größten Produzenten von Körnermais weltweit mit > 260 Mio. Tonnen darstellte⁹¹. Momentan gehört China zwar noch zu den Ländern, welche Mais importieren³⁷⁸. Sollte sich diese Lage jedoch in Zukunft ändern, ist es empfehlenswert das Klassifikationsmodell um dieses Land zu erweitern. Ähnliches gilt für Argentinien, wenn auch aus anderen Gründen: Aufgrund der unterschiedlichen Jahreszeiten auf der Südhalbkugel der Erde, erfolgt die Ernte dort von Mai bis Juli^{91,379}, weshalb ein zu den Körnermaisproduzenten der Nordhalbkugel antizyklisches Angebot besteht. In Bezug auf die Proben aus den USA konnte im Rahmen dieser Arbeit zudem festgestellt werden, dass ein größerer Probenumfang mit genauen Herkunftsangaben aus den USA wünschenswert gewesen wäre. Eine Erweiterung auf die US-Bundesstaaten des „Corn Belt“ wäre im selben Zuge sinnvoll.

Außerdem ist für die Routineanwendung der Herkunftsbestimmung die Etablierung einer Triple-Quadrupol-Methode zur Quantifizierung sämtlicher 20 identifizierter Markerlipide in einer einzigen LC-MS/MS-Analyse anzustreben. Die Problematik, dass die identifizierten Metabolite nicht als käufliche (isotopenmarkierte) Standardsubstanzen existieren, muss zudem dabei berücksichtigt werden. Mögliche Lösungsansätze wären hier, entweder eine chemische Synthese durchzuführen oder die Isolation der Substanzen im semi-präparativen Maßstab mittels HPLC oder anderen innovativen Verfahren, wie der CCC^{380,381}.

Um die Routinetauglichkeit der LC-MS-Methode zu zeigen, wurden bereits die Quantifizierung eines Teils der Markermetabolite durchgeführt und die Eignung zur Herkunftsbestimmung gezeigt. Zudem wurde zur Vorbereitung des Routineeinsatzes, die Stabilität der Extrakte unter verschiedenen Lagerungsbedingungen überprüft. Besondere Aufmerksamkeit ist darüber hinaus der Verlässlichkeit der selektierten Metabolite in Bezug auf Schwankungen zwischen den einzelnen Erntejahren zu widmen.

Weitere vielversprechende Forschungsansätze:

Bei den Matrices Spargel⁶ bzw. Haselnüsse⁴⁶ wurden ebenso u. a. Lipide als vielversprechende Marker zur Herkunftsbestimmung gefunden. Gleichzeitig erwies sich auch hier die Analyse der

polaren Metabolite als weniger vielversprechend. Daher wäre die These zu überprüfen, dass eine Allgemeingültigkeit der Eignung von Lipiden als Marker für die Herkunft bestehen könnte.

Zur weiteren Verfeinerung der Genauigkeit der Vorhersagemodelle könnte die Anwendung neuerer Analysetechniken einen Beitrag leisten. Besonders vielversprechend ist hierbei die Kopplung der Ionenmobilitätsspektroskopie (IMS) mit der Massenspektrometrie (IM-MS). Durch die Auftrennung von coeluierenden Isomeren vor der massenspektrometrischen Detektion in einer Ionenmobilitätszelle als Ionen durch ein *Driftgas*³⁸², könnte eine bedeutsame Verbesserung der Genauigkeit der Herkunftsbestimmung unter Nutzung der bereits in dieser Arbeit ausgewählten Lipidmarker eintreten. Die IMS kann einzelne Lipidspezies auftrennen, welche sich entweder lediglich in der *cis/trans*-Konfiguration der Doppelbindungen in den Fettsäuren³⁸³, oder in Bezug auf deren Position unterscheiden³⁸⁴. Die Anwendung der IMS im Lebensmittelbereich liegt zunehmend im Trend, so wurden bspw. von Lu *et al.* Phenolsäuren in Pflanzenwurzeln mittels HPLC-ESI-IMS untersucht. Dabei konnten diverse Phenolsäuren zwar in der HPLC-Ebene nicht aufgetrennt werden, deren Separation gelang jedoch leicht mittels IMS^{382,385}. Somit könnte die Verwendung eines IM-MS eine vielversprechende Verbesserung der Herkunftsbestimmung auch für Körnermais ermöglichen.

Auch die Fusion mit weiteren, mittels anderer analytischer Methoden gewonnen Daten, könnte für die Herkunftsbestimmung von Körnermais aussichtsreich sein. Insbesondere die Fusion mit Daten aus der Analyse mittels NMR-Spektroskopie sowie FT-IR-Spektroskopie³⁷⁵ könnte erstrebenswert sein. Vor allem, da diese Analytik bereits von den Projektpartnern an denselben Proben, die auch in dieser Arbeit verwendet wurden, vorgenommen wurden.

Im Zuge der genaueren Untersuchung der Unterschiede zwischen Körnermaisproben verschiedener Herkunftsländer, wäre es zudem sehr interessant auch Mischungen zu betrachten. In der Realität ist das Auftreten einer nur anteiligen Verfälschung von Chargen durchaus gegeben^{386–388}. Vermutlich wären in einem solchen Falle die zu beobachtenden Unterschiede noch geringer, als bei Reinproben. Daher erfordert die Erforschung dieses Themenkomplexes vorrausichtlich den Einsatz von leistungsfähigeren Analysemethoden bzw. -strategien, wie der IM-MS.

9 Methoden

9.1 Probenmaterial

Es wurden verschiedene Körnermais-Proben in dieser Arbeit verwendet. Diese stammten aus unterschiedlichen Herkunftsländern und beinhalteten zum Teil unterschiedliche Körnermaissorten, um möglichst sortenunabhängige Markermetabolite zur Herkunftsbestimmung aufzufinden.

Zur Methodenoptimierung wurde dabei durchgehend QC-Kontrollproben verwendet (siehe **Abschn. 9.1.1**), um die biologische Varianz der Proben zu eliminieren. Eine QC-Probe wurde ebenso bei der Messung größerer Probenzahlen in Batches alle zehn bzw. zwölf LC-MS-Analysen analysiert, um eventuelle gerätebezogenen Schwankungen in den Messungen im zeitlichen Verlauf aufzudecken.

Um einen entsprechenden Sorteneinfluss der – üblicherweise im Handel nicht deklarierten Körnermaissorte – auf die Herkunftsbestimmung zusätzlich systematisch zu untersuchen, wurden außerdem insgesamt 81 Proben aus drei Landessortenversuchen aus Deutschland (Nordrhein-Westfalen, Brandenburg, Bayern) des Jahres 2017 analysiert. Die Proben stammten aus den „Landessortenversuchen des LEIF Brandenburg 2017“ (Dr. Gert Barthelmes), den „Landessortenversuchen der LfL Bayern, dem Institut für Pflanzenbau und Pflanzenzüchtung, des Jahres 2017“ (Dr. Joachim Eder) sowie den „Landessortenversuchen der Landwirtschaftskammer Nordrhein-Westfalen des Jahres 2017“ (Norbert Erhardt). Das verwendete Probenset umfasste neun Sorten. Je Kombination aus Herkunftsbundesland und Körnermaissorte wurden drei Replikate analysiert, siehe **Tab. 9.1**.

Tab. 9.1: Probenumfang des Probensets aus Landessortenversuchen. Genutzt zur Evaluierung des Einflusses der Körnermaissorte auf das Metabolitprofil in Relation zu dem Herkunftseinfluss.

Probenbezeichnung	Probenanzahl
<i>nach Herkunft</i>	
Nordrhein-Westfalen	27
Brandenburg	27
Bayern	27
<i>nach Sorte</i>	
Amagrano	9
ES Crossman	9
SY Talisman	9
KWS Stabil	9
Benedictio KWS	9
Ridley	9
ES Hubble	9
Galactus	9
P 8723	9
Σ Gesamtprobenanzahl	81

Zur Etablierung bzw. Validierung der Klassifizierungsmodelle zur Herkunftsbestimmung kamen insgesamt 173 Proben aus sieben verschiedenen Herkunftsländern zum Einsatz. Die Probenanzahl

je Land betrug dabei in der Regel $n = 19$ bis 25 , lediglich die Proben aus der Slowakei waren mit $n = 14$ etwas unterrepräsentiert. Die Probenzahlen und deren Verteilung auf die einzelnen Herkunftsländer sind in **Tab. 9.2** aufgeschlüsselt.

Tab. 9.2: Probenumfang des Probensets zur Herkunftsbestimmung. Genutzt zur Etablierung von analytischen Methoden zur Herkunftsbestimmung von Körnermais.

Herkunftsland	Probenanzahl
Frankreich (Elsaß)*	36
Peru	25
Spanien†	20
Slowakei	14
Ukraine	19
Ungarn*	36
USA	23
- Alabama	4
- Arkansas	4
- Georgia	4
- Louisiana	4
- Mississippi	4
- Texas	3
Σ Gesamtprobenanzahl	173

* Diese Proben wurden mittels Variante B vermahlen. Für die LC-MS-Analytik wurde eine geringere Anzahl ($n = 25$) zur Modellerstellung verwendet, je $n = 10$ Proben dienten dort zusätzlich zu Validierungszwecken.

† Die Proben aus Spanien entstammten einem Sortenversuch und enthielten 20 unterschiedliche Körnermaissorten.

Für einen Teil der Proben waren genauere Ortsangaben vorhanden: 14 Proben kamen aus der West-Slowakei. Die 20 Proben aus Spanien stammten von einem Sortenversuch in Lleida. Daher konnten diese Proben dazu genutzt werden, um die Varianz der gefundenen Markermetabolite bzw. FT-NIR-Spektrenbereiche zur Herkunftsbestimmung in Bezug auf die Körnermaissorte zu evaluieren. Die Proben aus den USA, stammten jeweils zu $n = 3-4$, aus den Bundesstaaten Alabama, Arkansas, Georgia, Louisiana, Mississippi sowie Texas. Bei diesen Proben war zudem eine Ortangabe bis auf *County*-Ebene vorhanden. Die Proben aus Peru stammten aus zehn Orten in der Provinz Lambayeque. Insgesamt wurden je $n = 36$ Proben aus Frankreich (Region Elsass) bzw. Ungarn mittels FT-NIR-Spektroskopie analysiert. Mittels LC-MS wurde hingegen nur ein Teil dieser Proben (je 25 Stück) untersucht. Zur zusätzlichen Verifizierung der entwickelten LC-MRM-Methode wurden in einem weiteren Batch jeweils zehn Proben beider Herkunftsebenen analysiert.

9.1.1 QC-Kontrollprobe

Als QC-Kontrollprobe wurde eine kommerzielle Probe (Mühle Gladen, Dorsten/Lembeck) genutzt, die aus dem Rheintal (Deutschland) stammte. Sie wurde entsprechend der in **Abschn. 9.2** dargelegten Probenvorbereitung der Variante A homogenisiert und anschließend gefriergetrocknet.

Die QC-Kontrollprobe wurde zur Überwachung der Leistungsfähigkeit der UHPLC und des Massenspektrometers eingesetzt, indem alle zehn oder zwölf Herkunftsproben ein entsprechender Extrakt der QC-Probe analysiert wurde. Dies diente dem Zweck Unregelmäßigkeiten innerhalb der Messreihen zu erkennen. Des Weiteren kam die QC-Probe

immer dann zum Einsatz, wenn eine Methodenoptimierung durchgeführt wurde, um eventuelle Störeinflüsse durch die Körnermaisprobe auszuschließen.

Neben der QC-Kontrollprobe (Kurzform: „QC-Probe“) kam im Bereich der FT-NIR-Analytik eine weitere QC-Probe II zum Einsatz. Diese diente zur genaueren Evaluation des Einflusses der Probenvorbereitung auf die Analyseergebnisse der FT-NIR-Spektroskopie. Es handelte sich um eine kommerzielle Probe von 2.5 kg Körnermais zur Verwendung als Haustierfutter (Dehner Gartencenter, Hamburg). Diese Probe wurde sowohl einmal nach **Abschn. 9.2 Variante A** als auch *Variante B* prozessiert, als auch die intakten Maiskörner der Probe direkt, ohne weitere Probenvorbereitung analysiert. Dies ermöglichte einen direkten Vergleich des Einflusses der Probenvorbereitung auf die Messergebnisse ohne einen störenden Einfluss durch unterschiedliche sonstige Probeneigenschaften befürchten zu müssen.

9.2 Probenvorbereitung

Ein Teil der in dieser Arbeit verwendeten Proben (Herkunftsproben aus den Ländern Peru, Spanien, Slowakei, Ukraine und den USA) wurden von den Projektpartnern am BfR in Berlin akquiriert. Die Probenvorbereitung dieser Proben erfolgte ebenfalls dort durch die Projektpartner. Die Probenvorbereitung der Maisproben erfolgte je nach aufarbeitender Forschungsstelle nach leicht unterschiedlichen Verfahren:

Variante A – (Forschungsstelle BfR/Berlin):

Die Proben wurden kurzzeitig vakuum-verpackt bei Raumtemperatur bis zur Vermahlung gelagert. Die Maiskörner wurden in einer Ultra-Zentrifugalmühle ZM 200 mit einer Siebweite von 0.5 mm (Retsch GmbH, Haan) vermahlen. Die vermahlenen Proben wurden in einem Rhönradmischer RRM 100 (J. Engelsmann AG, Ludwigshafen) für 3 h bei 28 rpm homogenisiert³⁷⁵.

Für die FT-NIR-Analytik wurden die Proben wie von der Forschungsstelle BfR/Berlin zu Verfügung gestellt, verwendet. Teilproben wurden in der Forschungsstelle Hamburg vor der dortigen Analyse mittels LC-MS, gefriergetrocknet.

Für die Gefriertrocknung wurden jeweils ca. 3 g vermahlene Körnermaisprobe in ein 5 mL Glasvial ohne Deckel gefüllt, mit einem Stück fuselfreiem Zellstoff verschlossen und in flüssigem Stickstoff schockgefroren. Die Gefriertrocknung erfolgte für 24 h in einer Alpha 1-4 LSC Gefriertrocknungsanlage (Martin Christ Gefriertrocknungsanlagen GmbH, Osterode). Anschließend wurden die Proben in 15 mL-Probenröhrchen abgefüllt, erneut mittels flüssigen Stickstoffs eingefroren und bei -21 °C dauerhaft gelagert.

Variante B – Forschungsstelle Uni Hamburg

Die Proben wurden in der Regel umgehend nach Erhalt vermahlen, ggfls. wurden die Proben kurzzeitig bei Raumtemperatur bis zur Vermahlung gelagert. Die Maiskörner wurden in flüssigem Stickstoff schockgefroren und anschließend in einer GM 300 Messermühle (Retsch GmbH, Haan) mit derselben Menge Trockeneis zusammen vermahlen. Die Gesamtmenge an Mahlgut betrug maximal 500 g. Falls mehr Probe zu vermahlen war, wurden mehrere Mahlgänge durchgeführt und

die Chargen anschließend miteinander vermischt. Das verwendete Mahlprogramm ist in **Tab. 9.3** detailliert dargestellt.

Tab. 9.3: Verwendete Mahlsequenz zur Vermahlung von Körnermaisproben in der Grindomix GM 300 Messermühle.

Dauer [s]	Geschwindigkeit [rpm]
20	1000
30	4000
5	Pause
30	4000
5	Pause
30	4000

Nach der Vermahlung wurden die Proben unter Lichtausschluss in einer Beta 1-8 LDplus Gefriertrocknungsanlage (Martin Christ Gefriertrocknungsanlagen GmbH, Osterode) für 24 h gefriergetrocknet. Anschließend wurden die Proben zu 40 g in 50 mL-Probenröhrchen abgefüllt, erneut mittels flüssigen Stickstoffs schockgefroren und bei -21 °C dauerhaft gelagert.

9.3 Allgemeine Vorgehensweise zur Extraktion von Metaboliten aus Körnermaisproben

Die Extraktionen wurden mit jeweils 50 mg des Lyophilisats durchgeführt. Um die Varianz des Extraktes durch äußere Einflüsse zu minimieren, wurden die Extraktionsansätze jeweils zwischen den Arbeitsschritten auf Eis gelagert sowie die verwendeten Lösungsmittel zuvor auf Eis gekühlt.

Es wurde eine definierte Menge vermahlende Körnermaisprobe auf einer Analysenwaage genau in ein 2 mL-Reaktionsgefäß eingewogen und zusätzlich zwei Stahlkugeln mit einem Durchmesser von 1/8 Zoll = 3.2 mm hinzugefügt. Die Extraktion erfolgte in einer BeadRuptor24 Kugelmühle (Omni International, USA) bei einer Gesamtextraktionsdauer von 3 min und einer Intensität von 3.1 m/s. Im Falle von Einphasen-Extraktionsmethoden wurde insgesamt 1 mL Extraktionsmittel eingesetzt, im Falle von Zweiphasen-Extraktionsmethoden wurde insgesamt 1.5 mL Extraktionsmittel eingesetzt (siehe **Tab. 9.4** und **Tab. 9.5**).

Je nach angewandter Extraktionsmethode erfolgte die Extraktion ggfls. in mehreren Schritten, d. h. nach 1 min Extraktion wurden erneut Lösungsmittel hinzugegeben und danach weitere 2 min lang extrahiert. Diese Vorgehensweise wurde bei der Nutzung von Wasser als Teil des Extraktionsgemisches genutzt, indem zunächst für 1 min nur mit organischen Lösungsmitteln extrahiert wurde, um die bei der Chromatographie störende Mitextraktion von Proteinen zu verringern.

Anschließend wurden die Proben für 3 min bei 16000 g und 4 °C zentrifugiert. Im Falle von Zweiphasen-Extraktionsmethoden erfolgte hierbei eine Phasentrennung. Diejenige Phase, welche die Metabolite von Interesse enthielt, wurde in ein neues 2 mL-Reaktionsgefäß überführt und ggfls. verdünnt. Anschließend wurde erneut für 10 min bei 16000 g und 4 °C zentrifugiert und min. 600 µL des Überstandes in ein HPLC-Vial überführt. Die Lagerung der Vials bis zur Messung erfolgte (außer bei Lagerungsversuchen, siehe **Abschn. 9.4**) bei 4 °C.

9.3.1 Extraktionsoptimierung für polare Metabolite

Die Optimierung der Extraktionsmethode für polare Metabolite wurde jeweils in Dreifachbestimmung mit vier Zweiphasen-Extraktionsmethoden und acht Einphasen-Extraktionsmethoden entsprechend dem in **Abschn. 9.3** geschilderten Ablauf, durchgeführt. Als einheitliches Probenmaterial wurde die QC-Probe verwendet.

Dazu wurden jeweils 50 ± 0.5 mg der gefriergetrockneten QC-Probe in ein 2 mL Reaktionsgefäß eingewogen, und extrahiert. Die Dauer der einzelnen Extraktionsschritte variierte, je nachdem wie oft Lösungsmittel hinzugegeben werden mussten, die Gesamtextraktionsdauer betrug einheitlich 3 min. Das Volumen des Extraktionsmittels wurde für Einphasen-Extraktionen auf 1 mL normiert und für Zweiphasen-Extraktionen auf 1.5 mL.

Die Extraktionsmethoden sind in **Tab. 9.4** zusammenfassend aufgelistet. Dargestellt sind Extraktionsmittelzusammensetzung und die Reihenfolge der Zugabe der Lösungsmittel^{115,306,319–325}.

Tab. 9.4: Untersuchte Extraktionsmethoden bei der Optimierung der Analyse von polaren Metaboliten von Körnermais mittels LC-ESI-MS.

Methanol/Wasser 1:1 (v/v)	Methanol/Wasser 1:2 (v/v)	Methanol/Wasser 2:1 (v/v)	Ethanol/Wasser 1:1 (v/v)
+ 500 µL Methanol	+ 333 µL Methanol	+ 667 µL Methanol	+ 500 µL Ethanol
1 min Kugelmühle	1 min Kugelmühle	1 min Kugelmühle	1 min Kugelmühle
+ 500 µL Wasser	+ 667 µL Wasser	+ 333 µL Wasser	+ 500 µL Wasser
2 min Kugelmühle	2 min Kugelmühle	2 min Kugelmühle	2 min Kugelmühle
Ethanol/Wasser 1:2 (v/v)	Ethanol/Wasser 2:1 (v/v)	Methanol	Ethanol
+ 333 µL Ethanol	+ 667 µL Ethanol	+ 1000 µL Methanol	+ 1000 µL Ethanol
1 min Kugelmühle	1 min Kugelmühle	3 min Kugelmühle	3 min Kugelmühle
+ 667 µL Wasser	+ 333 µL Wasser		
2 min Kugelmühle	2 min Kugelmühle		
FOLCH	BLIGH & DYER	LÖFGREN	ARMIROTTI
+ 900 µL Chloroform + 400 µL Methanol	+ 250 µL Chloroform + 500 µL Methanol	+ 450 µL 1-Butanol + 150 µL Methanol	+ 150 µL Chloroform + 600 µL Methanol + 150 µL Heptan
1 min Kugelmühle	1 min Kugelmühle	1 min Kugelmühle	1 min Kugelmühle
+ 200 µL Wasser	+ 250 µL Chloroform + 500 µL Wasser	+ 225 µL Heptan + 75 µL Ethylacetat	+ 300 µL Chloroform + 300 µL Wasser
2 min Kugelmühle	2 min Kugelmühle	1 min Kugelmühle	2 min Kugelmühle
		+ 600 µL Wasser	
		1 min Kugelmühle	

Im Falle der Zweiphasenextraktionen erfolgte die Analytik der polaren Metabolite aus der polaren, wässrigen Phase. In der Regel befindet sich die wässrige Phase bei Zweiphasensystemen aus wässriger und organischer Lösung aufgrund der höheren Dichte von Wasser unten. Jedoch ist dies bei der Verwendung halogenierter Lösungsmittel in der Regel umgekehrt; aufgrund deren hoher Dichte³²⁷. Im Falle der Extraktionen nach FOLCH, BLIGH & DYER, sowie ARMIROTTI wurde daher die obere, wässrige Phase analysiert. Nur bei der Extraktion nach LÖFGREN stellte die untere Phase, die

wässrige Phase dar. Die Extrakte wurden mit Acetonitril/Methanol 9:1 v/v (organisches Laufmittel der Flüssigkeitschromatographie ohne Additiv) um den Faktor 1:2 verdünnt, um die Kompatibilität der Extrakte mit den Anfangsbedingungen der Chromatographie zu gewährleisten. Die anschließende Messung der Proben erfolgte mittels der LC-MS Methode für polare Metabolite (**Abschn. 9.6.1**), mit einem Injektionsvolumen von 5 µL. Die Kalibration/Auswertung erfolgte wie in **Abschn. 9.9.1** beschrieben.

9.3.2 Optimierte Extraktionsmethode für polare Metabolite

Anhand der Auswertung (siehe **Abschn. 6.1.1**) wurde die Extraktionsmethode, welche auf einer Mischung aus Methanol/Wasser 2:1 v/v basiert, als optimal zur Extraktion von polaren Metaboliten des Körnermais bestimmt. Die Versuchsvorschrift lautet wie folgt:

Von der jeweiligen Körnermaisprobe wurden 50 ± 0.25 mg in ein 2 mL-Reaktionsgefäß genau eingewogen. In dem Gefäß befanden sich bereits zwei Stahlkugeln. Um Veränderungen der Metabolitzusammensetzung und die Extraktion von Proteinen zu minimieren, wurden sowohl die Proben als auch die Extraktionsmittel bei der Extraktion auf Eis gekühlt.

Die Extraktion wurde in einer BeadRuptor 24 Kugelmühle (Omni International, USA) bei einer Frequenz von 3.1 m/s für insgesamt 3 min durchgeführt. Es wurden 667 µL Methanol zu der Probe gegeben und für 1 min in der Kugelmühle extrahiert. Dann folgte die Zugabe von 333 µL Wasser und eine erneute Extraktion mittels Kugelmühle für 2 min. Anschließend wurden die Proben für 3 min bei 16000 g und 4 °C zentrifugiert. 750 µL des Überstandes wurden in ein neues 2 mL-Reaktionsgefäß überführt. Dann wurde erneut für 10 min bei 16000 g und 4 °C zentrifugiert und 700 µL des Überstandes in ein HPLC-Vial überführt. Eine kurzzeitige Lagerung der Vials bis zur Messung erfolgte bei 4 °C.

9.3.3 Extraktionsoptimierung für unpolare Metabolite

Entsprechend der Optimierung der Extraktionsmethode für polare Metabolite wurde analog die optimale Extraktionsmethode für unpolare Metabolite bestimmt. Diese erfolgte ebenso in Dreifachbestimmung von acht Zweiphasen- und vier Einphasen-Extraktionsmethoden; entsprechend der in **Abschn. 9.3** geschilderten allgemeinen Vorgehensweise. Dabei kamen ebenso je 50 ± 0.5 mg der gefriergetrockneten QC-Probe als einheitliche Körnermaisprobe zum Einsatz. Die verwendeten Extraktionsmethoden sind in **Tab. 9.5** zusammenfassend aufgelistet^{46,306,319–322,326–329}.

Tab. 9.5: Untersuchte Extraktionsmethoden bei der Optimierung der Analyse von unpolaren Metaboliten von Körnermais mittels LC-ESI-MS.

FOLCH	BLIGH & DYER	LÖFGREN	ARMIROTTI
+ 900 µL Chloroform + 400 µL Methanol	+ 250 µL Chloroform + 500 µL Methanol	+ 450 µL 1-Butanol + 150 µL Methanol	+ 600 µL Methanol + 150 µL Chloroform + 150 µL Heptan
1 min Kugelmühle	1 min Kugelmühle	1 min Kugelmühle	1 min Kugelmühle
+ 200 µL Wasser	+ 250 µL Chloroform + 500 µL Wasser	+ 225 µL Heptan + 75 µL Ethylacetat	+ 300 µL Chloroform + 300 µL Wasser
2 min Kugelmühle	2 min Kugelmühle	1 min Kugelmühle	2 min Kugelmühle
		+ 600 µL Wasser	
		1 min Kugelmühle	
MATYASH	BLIGH & DYER*	FOLCH†	BLIGH & DYER†
+ 1000 µL tert.-MTBE + 300 µL Methanol	+ 250 µL Dichlormethan + 500 µL Methanol	+ 900 µL Chloroform + 400 µL Isopropanol	+ 250 µL Chloroform + 500 µL Isopropanol
1 min Kugelmühle	1 min Kugelmühle	1 min Kugelmühle	1 min Kugelmühle
+ 200 µL Wasser	+ 250 µL Dichlormethan + 500 µL Wasser	+ 200 µL Wasser	+ 250 µL Chloroform + 500 µL Wasser
2 min Kugelmühle	2 min Kugelmühle	2 min Kugelmühle	2 min Kugelmühle
HARA	PELLEGRINO	Chloroform/Isopropanol 1:2 (v/v)	Chloroform/Methanol 1:2 (v/v)
+ 600 µL Hexan + 400 µL Isopropanol	+ 300 µL Chloroform + 300 µL MTBE + 400 µL Methanol	+ 333 µL Chloroform + 667 µL Isopropanol	+ 333 µL Chloroform + 667 µL Methanol
1 min Kugelmühle	3 min Kugelmühle	3 min Kugelmühle	3 min Kugelmühle
+ 500 µL Wasser			
2 min Kugelmühle			

* Chloroform ersetzt durch Dichlormethan. † Methanol ersetzt durch Isopropanol.

Im Falle der Extraktionen nach FOLCH, BLIGH & DYER sowie ARMIROTTI wurde die untere, organische Phase analysiert. Bei der Extraktion nach LÖFGREN und MATYASH wurde die obere Phase analysiert, welche die organische Phase darstellte. Die Dauer der einzelnen Extraktionsschritte variierte, je nachdem wie oft Lösungsmittel hinzugegeben werden mussten, die Gesamtextraktionsdauer betrug auch hier 3 min. Das Gesamtvolumen des zugegebenen Extraktionsmittels betrug 1 mL für Einphasen-Extraktionen und 1.5 mL für Zweiphasen-Extraktionen, analog Creydt *et al.*³⁰⁶.

Die Extrakte wurden mit Isopropanol/Acetonitril 3:1 v/v (organisches Laufmittel der Flüssigkeitschromatographie ohne Additiv) um den Faktor 1:10 verdünnt, um hier ebenso die Kompatibilität der Extrakte mit den Chromatographiebedingungen sicherzustellen. Die Messung der Proben erfolgte mittels der LC-MS Methode (**Abschn. 9.6.2**) für unpolare Metabolite. Das Injektionsvolumen betrug 5 µL. Die Kalibration/Auswertung erfolgte wie in **Abschn. 9.9.1** beschrieben.

9.3.4 Optimierte Extraktionsmethode für unpolare Metabolite

Anhand der Auswertung (siehe **Abschn. 6.1.2**) wurde die Extraktionsmethode nach BLIGH & DYER³²⁰, welche eine Mischung aus Chloroform, Methanol und Wasser nutzt, als optimal zur Extraktion von unpolaren Metaboliten des Körnermais bestimmt. Die Versuchsvorschrift lautet wie folgt:

Von den vermahlenden Körnermaisproben wurden jeweils 50 ± 0.25 mg in ein 2 mL-Reaktionsgefäß eingewogen, in das zuvor zwei Stahlkugeln gegeben wurden. Sowohl die Proben, als auch die Extraktionsmittel wurden vor und zwischen den Extraktionsphasen auf Eis gekühlt. Die Extraktion erfolgte in einer BeadRuptor 24 Kugelmühle (Omni International, USA) für insgesamt 3 min und einer Intensität von 3.1 m/s. Es wurden zunächst 750 µL einer Mischung aus Chloroform und Methanol (1:2 v/v) zu der Probe gegeben und für 1 min in der Kugelmühle extrahiert. Dann folgte die Zugabe von 250 µL Chloroform und 500 µL Wasser und eine erneute Extraktion mittels Kugelmühle für 2 min. Anschließend wurden die Proben für 3 min bei 16000 g und 4 °C zentrifugiert, was zu einer Phasentrennung führte.

Im Falle der LC-MS-Messungen gemäß **Abschn. 9.6.2**, wurden 160 µL der unteren, organischen Phase in ein neues 2 mL-Reaktionsgefäß überführt, welches bereits 640 µL einer Mischung aus Isopropanol/Acetonitril (3:1 v/v) enthielt. Die Extraktionslösung wurde also um den Faktor 1:5 verdünnt. Im Falle von LC-MRM-Messungen gemäß **Abschn. 9.7** wurden 100 µL der unteren, organischen Phase in ein neues 2 mL-Reaktionsgefäß überführt, welches bereits 900 µL einer Mischung aus Isopropanol/Acetonitril (3:1 v/v) enthielt. Die Extraktionslösung wurde folglich um den Faktor 1:10 verdünnt. Dann wurde erneut für 10 min bei 16000 g und 4 °C zentrifugiert und 700 µL des Überstandes in ein HPLC-Vial überführt. Die kurzzeitige Lagerung der Vials bis zur Messung erfolgte bei 4 °C.

9.4 Überprüfung der Lagerungsstabilität von Extrakten unpolarer Metabolite

Zur Überprüfung der Lagerungsstabilität der Extrakte unpolarer Metabolite wurde die QC-Probe insgesamt 32 Mal auf 50 ± 0.25 mg genau eingewogen. Die Extraktion der Proben erfolgte nach der BLIGH & DYER-Methode (siehe **Abschn. 9.3.4**) jeweils in Dreifachbestimmung 7, 3 Tage oder 24 h vor Beginn der LC-MS-Analytik. In der Zeit bis zum Messbeginn wurden die Probenextrakte unter verschiedenen Bedingungen gelagert: bei -21 °C im Gefrierschrank, bei +4 °C im Kühlschrank oder bei Raumtemperatur. Die Verdünnung der Proben um den Faktor 1:5 mit Isopropanol/Acetonitril 3:1 v/v erfolgte dabei erst direkt vor der Messung, um explizit die Stabilität des Extraktes zu überprüfen. Als Vergleich und Referenz für eine sofortige Messung wurde außerdem die QC-Probe in Fünffachbestimmung frisch vor Messbeginn nach derselben Methode extrahiert, verdünnt und sofort analysiert. Die Auswertung der Messdaten erfolgte entsprechend wie in **Abschn. 9.9.3** dargelegt.

9.5 Tocopherol-spezifische Extraktionsmethode

Analog zu Moreau *et al.* und der §64-Methode zur Bestimmung von Vitamin E mittels HPLC wurde die Extraktion der Tocopherole mittels Isopropanol durchgeführt^{358,360}. Zudem wurde dem Extraktionsmittel 500 mg/L (0.05 %) BHT zugesetzt, um die Oxidation der Probenbestandteile möglichst abzumildern^{44,306,361}.

Zur Quantifizierung von α -/ γ -Tocopherol wurde wie folgt vorgegangen: Die Proben wurden zu 50 ± 0.5 mg auf einer Analysenwaage eingewogen. Die Extraktion erfolgte gemäß **Abschn. 9.3**, wobei je Probe 1000 µL Isopropanol verwendet wurden, welches 500 µg/mL BHT enthielt. 500 µL des Extraktes wurden in einem neuen 2 mL Reaktionsgefäß mit 500 µL Isopropanol mit 0.05 % BHT verdünnt (Verdünnungsfaktor 1:2) und nach erneuter Zentrifugation für 10 min bei 16000 g und

4 °C, wurden je 800 µL in ein HPLC-Vial überführt und bei 4 °C bis zur Analyse gelagert. Die LC-MS/MS-Messung erfolgte gemäß **Abschn. 9.8** mit der MRM-Methode für die Tocopherolbestimmung. Das Injektionsvolumen betrug 5 µL.

9.6 LC-MS(/MS)-Messungen

Die LC-MS(/MS)-Messungen erfolgten mittels einer Dionex 3000 UHPLC an die ein Maxis 4G ESI-qToF-Massenspektrometer von Bruker als Detektor gekoppelt war. Die Steuerung der UHPLC erfolgte mittels der Software Chromeleon 2.14. Das Massenspektrometer wurde mit oToFControl 3.4 gesteuert. Die Abstimmung der unterschiedlichen Programme erfolgte mittels HyStar 3.2.

9.6.1 LC-MS Methoden für polare Metabolite

Die LC-MS Analytik von polaren Metaboliten erfolgte entsprechend Creydt *et al.*⁴³. Als Trennsäule wurde eine Cogent DiamondHydrid 2.0 Chromatographiesäule (MicroSolv Technology, Leland, USA) mit entsprechender Vorsäule eingesetzt. Diese wies eine Länge von 150 mm auf, einen Durchmesser von 2.1 mm und eine Partikelgröße von 2.2 µm. Als Laufmittel A wurde Wasser und als Laufmittel B eine Mischung aus Acetonitril/Methanol 9:1 v/v eingesetzt. Als Additiv wurde je nach Ionisationsmodus 0.1 % Essigsäure (positiver Ionenmodus) bzw. 0.1 % Ameisensäure (negativer Ionenmodus) verwendet. Die Autosamplertemperatur betrug 4 °C, um eventuelle Abbauprozesse des Metaboloms zu verlangsamen. Das Injektionsvolumen betrug 5 µL (Extraktionsmethoden-Optimierung) bzw. 7 µL (Messungen der Sortenproben oder Herkunftsproben). Die Trennung erfolgte bei einer Säulenofentemperatur von 50 °C und einer Flussrate von 0.4 mL/min. Zur Trennung wurde ein Gradient verwendet, dessen Laufmittelzusammensetzung in **Tab. 9.6** aufgeführt ist.

Tab. 9.6: Zeitlicher Verlauf des Gradienten der LC-MS-Methode für die Analyse von polaren Metaboliten.

Retentionszeit [min]	Anteil Laufmittel B [%]	Anteil Laufmittel A [%]
0	100	0
2	100	0
14	50	50
17	50	50
18	100	0
22	100	0

Das Massenspektrometer wurde sowohl im positiven oder im negativen Ionenmodus betrieben. Die dabei verwendete UHPLC-Methode war jeweils identisch. Die entsprechenden Einstellungen des Massenspektrometers für beide Ionenmodi sind in **Tab. 9.7** zusammengestellt.

Tab. 9.7: Einstellungen der ESI-Ionenquelle sowie weiterer Parameter des q-ToF-Massenspektrometers für die LC-ESI-MS-basierte Analyse von polaren Metaboliten.

Allgemeine Parameter		QToF-MS-Parameter	
Messbereich	<i>m/z</i> 80-1000	Transfer Funnel 1 RF	300 Vpp
Spektratenrate	1 Hz	Transfer isCID Energy	0.0 eV
Methodendauer	22 min	Transfer Multipole RF	250 Vpp
ESI-Quellparameter		Quadrupol Low Mass	<i>m/z</i> 80
End Plate Offset	± 500 V	Quadrupol Ion Energy	5.0 eV
Kapillarspannung	± 4500 V	Collision Energy	10 eV
Nebulizer	4.0 bar	Collision RF	400 Vpp
Dry Gas	9.0 L/min	Pre Puls Storage	1.0 µs
Dry Temperature	200 °C	Transfer Time	75.0 µs

9.6.2 LC-MS Methoden für unpolare Metabolite

Die LC-MS Analytik von unpolaren Metaboliten erfolgte in Analogie zu Creydt *et al.*⁴³, mit Abweichungen bei der verwendeten Trennsäule (größerer Partikeldurchmesser) und in Folge einer geringeren Säulenofentemperatur. Alle anderen Parameter der Chromatographieanlage und des Massenspektrometers waren identisch wie von Creydt *et al.*²²⁴ optimiert.

Als Trennsäule wurde eine Accucore RP-MS Chromatographiesäule (Thermo Fischer Scientific, Braunschweig) mit entsprechender Vorsäule eingesetzt. Diese hatte eine Länge von 150 mm, einen Durchmesser von 2.1 mm und eine Partikelgröße von 2.6 µm. Als Laufmittel A wurde Wasser und als Laufmittel B eine Mischung aus Isopropanol/Acetonitril 3:1 v/v eingesetzt. Als Additiv wurde je nach Ionisationsmodus 10 mM Ammoniumformiat (positiver Ionenmodus) bzw. 0.02 % Essigsäure (negativer Ionenmodus) eingesetzt. Die Autosamplertemperatur betrug 4 °C. Das Injektionsvolumen betrug 5 µL für die Extraktionsmethoden-Optimierung bzw. 8 µL für die Messungen der Sortenproben oder Herkunftsproben. Die Trennung erfolgte bei einer Säulenofentemperatur von 30 °C und einer Flussrate von 0.3 mL/min. Zur Trennung wurde ein Gradient verwendet, dessen Laufmittelzusammensetzung **Tab. 9.8** entnommen werden kann.

Tab. 9.8: Zeitlicher Verlauf des Gradienten der LC-MS-Methode für die Analyse von unpolaren Metaboliten.

Retentionszeit [min]	Anteil Laufmittel B [%]	Anteil Laufmittel A [%]
0	55	45
2	55	45
4	75	25
18	100	0
22	100	0
23	55	45
27	55	45

Das Massenspektrometer wurde sowohl im positiven als auch im negativen Ionenmodus betrieben. Die dabei verwendete UHPLC-Methode war jeweils identisch. Die entsprechenden Einstellungen des Massenspektrometers für beide Ionenmodi waren ebenfalls identisch zu jenen

der polaren Methode (siehe **Tab. 9.7**), bis auf die Methodenlänge (22 min) sowie den Parameter „Collision RF“ welcher in der unpolaren Methode 500 Vpp, anstelle von 400 Vpp betrug.

9.6.3 Kalibration des Massenspektrometers

Die Kalibration des Massenspektrometers erfolgte mittels Natriumformiat- bzw. Natriumacetat-Clusterionen, je nach verwendetem Laufmitteladditiv. Die Zusammensetzung der Kalibrationslösung ist in **Tab. 9.9** aufgeschlüsselt. Die Kalibration erfolgte jeweils vor jedem gemessenen Batch und zudem durch Infusion der Kalibrationslösung mit 0.1 mL/h am Ende eines jeden LC-MS-Laufs über ein Sechs-Wege-Ventil mit einer 20 µL-Probenschleife, welche über eine Spritzenpumpe gespeist wurde.

Tab. 9.9: Pipettierschema für die Erstellung der Kalibrationslösung zur Kalibrierung der m/z -Achse des Massenspektrometers.

Bestandteil	Volumen
Natronlauge (1 M)	1 mL
Ameisensäure bzw. Essigsäure	100 µL
Isopropanol	50 mL
Wasser	50 mL

Dadurch, dass die Kalibrationslösung über die Spritzenpumpe permanent mit 0.1 mL/h während der eigentlichen LC-MS-Messung die Probenschleife durchströmte, erfolgte beim Umschalten des Sechs-Wege-Ventils für 30 s am Ende des LC-Laufes die Infusion von 20 µL Kalibrant in das Massenspektrometer. So wurde eine charakteristische Reihe von Clusterionen mit dem Schema $[\text{Na}(\text{NaCOOH})_n]^+ / [\text{HCOO}(\text{NaOOH})_n]^-$ bzw. $[\text{Na}(\text{NaC}_2\text{H}_3\text{O}_2)_n]^+ / [\text{C}_2\text{H}_3\text{O}_2(\text{NaC}_2\text{H}_3\text{O}_2)_n]^-$ erzeugt. Über die Clusterionen konnte entweder im Massenspektrometer direkt mittels *oTof-Control* oder später bei der Datenauswertung in *DataAnalysis 4.1*, die Kalibration mittels der Methode *High Precision Calibration* für die polaren Methoden oder über die Methode *Enhanced Quadratic* für die unpolaren Methoden erfolgen.

Neben der Kalibration mittels Clusterionen, wurde zusätzlich eine Lockmasse verwendet. Dazu dienten 40 µL einer 1 mg/mL Lösung von Hexakis(2,2-difluoro-ethoxy)phosphazene in Isopropanol. Diese erzeugte im positiven Ionenmodus ein Signal im Massenspektrum bei m/z 622.0295 $[\text{M}+\text{H}]^+$ bzw. im negativen Ionenmodus bei m/z 556.0014 $[\text{M}-\text{C}_2\text{H}_3\text{F}_2]^-$.

9.6.4 Akquirierung von MS/MS-Fragmentspektren

Die Messung von Fragmentspektren bestimmter m/z -Werte erfolgte mittels drei unterschiedlichen Strategien, je nach Abundanz der zu fragmentierenden Peaks:

- 1) Erstellung einer AutoMSMS-Methode und Eintrag der betreffenden m/z -Verhältnisse in die „Scheduled Precursor List“
- 2) Erstellung einer AutoMSMS-Methode und Eintrag der betreffenden m/z -Verhältnisse ± 0.5 Da in die „Precursor Ion List“ unter „Included“.
- 3) Erstellung einer MRM-Methode und Eintrag der betreffenden m/z -Verhältnisse als MRM-Precursor-Werte.

Die Fragmentierung erfolgte in allen Fällen mittels „Collision Induced Dissociation“. Die verwendete Fragmentierungsenergie betrug 20, 40 oder 60 eV. Alle sonstigen Einstellungen des Massenspektrometers verblieben entsprechend denen, die auch bei LC-MS-Messungen zur Anwendung kamen. Auch die Chromatographie-bezogenen Parameter blieben unverändert. Lediglich das Injektionsvolumen wurde teilweise, bei nur in niedrigen Konzentrationen vorhandenen Metaboliten nach oben (bis zu 16 µL eines 1:5 verdünnten Extraktes) angepasst.

Die Auswertung von MS/MS-Spektren zur Strukturaufklärung der Metabolite erfolgte mithilfe der Software DataAnalysis 4.1. Die Rohdaten der Massenspektren wurden zum Teil auch mit dem Softwarepaket ProteoWizard 3.0 inspiziert.

9.7 LC-MRM Methode für Lipide

Als Trennsäule wurde eine Accucore RP-MS Chromatographiesäule (150x2.1 mm; 2.6 µm Partikeldurchmesser) mit entsprechender Vorsäule verwendet. Als Laufmittel A wurde Wasser und als Laufmittel B eine Mischung aus Isopropanol/Acetonitril 3:1 v/v jeweils mit 10 mM Ammoniumformiat als Additiv eingesetzt. Die Autosamplertemperatur betrug 4 °C. Das Injektionsvolumen betrug 5 µL eines 1:10 verdünnten Lipidextraktes (siehe **Abschn. 9.3.4**). Die Trennung erfolgte bei einer Säulenofentemperatur von 40 °C und einer Flussrate von 0.4 mL/min, der verwendete Gradient ist in **Tab. 9.10** aufgeführt.

Tab. 9.10: Zeitlicher Verlauf des Gradienten der LC-MRM-Methode für die quantitative Bestimmung von unpolaren Metaboliten.

Retentionszeit [min]	Anteil Laufmittel B [%]	Anteil Laufmittel A [%]
0	80	20
2	80	20
8	100	0
12	100	0
13	80	20
16	80	20

Das Massenspektrometer wurde im positiven Ionenmodus betrieben. Die entsprechenden Einstellungen der ESI-Ionenquelle sowie weitere Einstellungen des Massenspektrometers sind identisch zu jenen der LC-MS-Methode für unpolare Metabolite (siehe **Abschn. 9.6.2**).

Zur Fragmentierung wurde eine Kollisionsenergie von 40 eV verwendet, das Selektionsfenster für das Precursorion betrug 4 Da. Es wurden insgesamt die MRM-Übergänge von neun Markern analysiert. Mittels des q-ToF-Massenspektrometers wurden hochaufgelöste *Full-Scan*-Massenspektren aufgenommen, weshalb geräteseitig lediglich das Precursorion definiert werden musste. Da nicht alle Ionenübergänge parallel erfasst werden konnten, erfolgte die Aufnahme der Fragmentionen nach deren Retentionszeiten gestaffelt. Die entsprechenden Fragmentionen, die als *Qualifier* bzw. *Quantifier* dienten, konnten nach der Analyse bei der Datenauswertung entsprechend ausgewählt werden. Die gewählten MRM-Übergänge sind in **Tab. 9.11** zusammengestellt, zusammen mit den Retentionszeiten der Marker sowie der Zeitfenster zur MRM-Datenaufnahme.

Tab. 9.11: Auflistung der quantifizierten Metabolite sowie die entsprechenden MRM-Übergänge.

Metabolit	Precursor <i>m/z</i>	Quantifier <i>m/z</i>	Qualifier <i>m/z</i>	RT (RT-Bereich)
PC (36:3)	784.586	184.07	601.52	5.3 min (0-6.0 min)
PE (36:3)	742.538	601.52	339.29	4.1 min (0-4.5 min)
DG (34:2)	610.514	313.27	337.27	5.5 min (4.5-7.0 min)
TG (52:6)	868.740	573.49	597.49	8.2 min (0-8.8 min)
TG (56:3)	930.852	631.57	633.58	9.4 min (8.8-9.5 min)
TG (56:2)	932.868	633.58	603.54	9.6 min (9.3-15 min)
TG (60:3)	986.914	689.65	687.63	9.8 min (9.5-15 min)
TG (58:2)	960.898	575.51	661.62	9.8 min (9.3-15 min)
TG (60:2)	988.929	689.65	603.54	10.0 min (9.8-15 min)
<i>Standardsubstanzen</i>				
STD PC (36:2)	786.602	184.073	522.356	6.0 min
STD PE (36:0)	748.586	607.567	267.267	6.3 min
STD DG (32:0)	586.541	313.273	239.236	6.0 min
STD TG (48:0)	824.771	551.505	313.272	9.2 min

Das Massenspektrometer wurde auf *m/z*-Ebene entsprechend nach dem in **Abschn. 9.6.3** dargestellten Verfahren vor Messbeginn mittels Natriumformiatcluster kalibriert. Zur Kalibration wurden eine Reihe Lipid-Standards gemessen sowie eine Verdünnungsreihe der Poolprobe als Referenz (siehe **Abschn. 9.7.1** und **9.7.2**).

9.7.1 Kalibration für quantitative Metabolit-Bestimmungen

Die quantitative Bestimmung der neun Markersubstanzen erfolgte über eine externe Kalibrationsgerade. Aufgrund eingeschränkter Verfügbarkeit von Standardsubstanzen wurde – ähnlich zu Fauland *et al.*³⁵⁶ – je Lipidklasse eine Substanz als Standard eingesetzt. Es wurde ein Mischstandard bestehend aus den vier Lipiden PC (36:2), PE (36:0), DG(32:0) sowie TG (48:0) angesetzt. Dazu wurden die Einzelsubstanzen als Stammlösungen zu je 1 mg/mL in Chloroform gelöst und daraus ein Mischstandard zu je 250 µM erstellt. Der Mischstandard wurde in Isopropanol entsprechend verdünnt, sodass folgende Konzentrationen erhalten wurden: 100, 50, 20, 10, 5, 2, 1, 0.5, 0.2, 0.1, 0.05 und 0.02 µM.

Die Kalibrierstandards wurden bis zur Analyse bei 4 °C im Kühlschrank gelagert. Die Analytik erfolgte entsprechend der in **Abschn. 9.7** dargestellten LC-MRM-Methode für die Quantifizierung von Lipiden. Es wurden jeweils 5 µL der Probenstandards injiziert.

9.7.2 Validierung der Lipid-MRM-Methode

Um das Detektionsvermögen der vier Standardsubstanzen mit denen der neun quantifizierten Markersubstanzen zu vergleichen, wurde eine Verdünnungsreihe des unverdünnten BLIGH & DYER-Extraktes (gemäß **Abschn. 9.3.4**) einer Poolprobe (je Herkunftsland eine Probe, sowie der QC-Probe) in Isopropanol/Acetonitril (3:1 v/v) mit 17 Verdünnungen im Bereich von 1:5 bis 1:300 erstellt. Darin waren alle quantifizierten Markermetabolite in den für die Quantifizierung relevanten Konzentrationsbereichen enthalten, da die Messung der eigentlichen Proben in einer 1:10 Verdünnung erfolgte.

Die Präzision der Bestimmung der vier Standardsubstanzen wurde anhand von insgesamt $n = 3$ QC-Proben (Einwaage 50 ± 0.25 mg), denen je $5 \mu\text{M}$ der vier Standardsubstanzen zudosiert wurden, bestimmt. Dies entspricht $20 \mu\text{L}$ Stammlösung zu $250 \mu\text{M}$ je Lipidstandard. Die Extraktion/Verdünnung der Proben um den Faktor 1:10 erfolgte entsprechend **Abschn. 9.3.4**. Das Injektionsvolumen betrug auch hier $5 \mu\text{L}$.

Die Bestimmung der NWG erfolgte, indem eine Extraktion nach BLIGH & DYER ohne Probe in fünffacher Bestimmung entsprechend **Abschn. 9.3.4** durchgeführt wurde. Die Blindextrakte wurden mittels der LC-MRM-Methode für Lipide (siehe **Abschn. 9.7**) analysiert. Das Injektionsvolumen betrug wie auch bei den anderen Messungen $5 \mu\text{L}$.

9.8 LC-MRM Methode für Tocopherole

Als Trennsäule wurde eine Poroshell 120 EC-C18 Chromatographiesäule (Agilent Technologies, Waldbronn) eingesetzt. Diese besaß eine Länge von 50 mm, einen Durchmesser von 4.6 mm und eine Partikelgröße von $2.7 \mu\text{m}$. Als Laufmittel A wurde Wasser und als Laufmittel B Isopropanol eingesetzt. Als Additiv wurde jeweils 5 mM Ammoniumformiat zugesetzt. Die Autosampler-temperatur betrug 4°C und das Injektionsvolumen $5 \mu\text{L}$. Die Trennung erfolgte bei einer Säulenofentemperatur von 30°C und einer Flussrate von 0.5 mL/min . Zur Trennung der Probenbestandteile wurde ein Gradient verwendet, dessen Laufmittelzusammensetzung in **Tab. 9.12** dargestellt ist.

Tab. 9.12: Zeitlicher Verlauf des Gradienten der LC-MRM-Methode für die quantitative Bestimmung von α -Tocopherol und γ -Tocopherol.

Retentionszeit [min]	Anteil Laufmittel B [%]	Anteil Laufmittel A [%]
0	75	25
2	75	25
6	100	0
9.5	100	0
10	75	25
13.5	75	25

Das Massenspektrometer wurde im positiven Ionenmodus betrieben. Die Parameter der ESI-Ionenquelle entsprachen jenen der LC-MRM-Methode für die Lipidspezies (siehe **Abschn. 9.7**). Für die Quantifizierung der Tocopherol-Konzentrationen in Körnermaisproben wurde eine LC-MRM-Methode erstellt, welche die Bestimmung der beiden Tocopherole anhand deren MRM-Übergänge ermöglichte. Zur Fragmentierung wurde eine Kollisionsenergie von 40 eV verwendet. Das Selektionsfenster für die Precursorionen bei m/z 431.39 (α -Tocopherol $[\text{M}+\text{H}]^+$) bzw. m/z 417.37 (γ -Tocopherol $[\text{M}+\text{H}]^+$) betrug jeweils 4 Da. Das q-ToF-Massenspektrometer erzeugte hochaufgelöste *Full-Scan*-Massenspektren der generierten Fragmentationen. Als *Quantifier* dienen die Ionen bei m/z 165.10 $[\text{C}_{10}\text{H}_{13}\text{O}_2+\text{H}]^+$ bzw. m/z 151.08 $[\text{C}_9\text{H}_{11}\text{O}_2+\text{H}]^+$ für α - bzw. γ -Tocopherol. Als *Qualifier* wurden die Ionen mit einem m/z 137.10 bzw. m/z 123.08 genutzt.

α -Tocopherol eluierte bei einer Retentionszeit von 5.2 min und γ -Tocopherol bei 4.8 min. Um das Massenspektrometer vor den teilweise mitextrahierten und in den Proben in hoher Konzentration

vorkommenden TGs zu schützen, wurde ab 7.0 min der Eluent nicht mehr ins Massenspektrometer infundiert, sondern über das Sechs-Wege-Ventil verworfen.

Die Kalibration des Massenspektrometers erfolgte vor jeder Messreihe mittels Natriumformiatcluster analog zu dem in **Abschn. 9.6.3** beschriebenen Verfahren. Außerdem wurde eine externe Kalibration mittels Standardsubstanzen wie in **Abschn. 9.7.1** beschrieben, durchgeführt.

9.8.1 Kalibration für quantitative Tocopherol-Bestimmungen

Zur quantitativen Bestimmung von α - und γ -Tocopherol wurde eine Kalibriergerade mit zehn Punkten im Konzentrationsbereich von 250 ng/mL bis 3.5 μ g/mL erstellt. Die einzelnen Verdünnungsstufen wurden aus einem Mischstandard der je 1 mg/mL α -Tocopherol und γ -Tocopherol enthielt, angesetzt. Der Mischstandard wurde in Isopropanol entsprechend verdünnt, sodass folgende Konzentrationen erhalten wurden: 3.5, 3, 2.5, 2, 1.5, 1.25, 1, 0.75, 0.5 und 0.25 μ g/mL.

Die Kalibrierstandards wurden bis zur Analyse bei 4 °C im Kühlschrank gelagert. Die Analytik erfolgte entsprechend der in **Abschn. 9.8** dargestellten LC-MS/MS-Methode für die Quantifizierung von α -/ γ -Tocopherol. Es wurden jeweils 5 μ L injiziert.

9.9 Auswertung der LC-MS(/MS)-Messungen

9.9.1 Feature-Finding

Sowohl die LC-MS als auch die LC-MS/MS bzw. MRM-Messungen wurden von dem Programm tof-Control als Rohdaten im d-Dateiformat abgespeichert. Dieses konnte in DataAnalysis 4.1 geöffnet werden. Dort erfolgte zunächst die Kalibration der Messdaten mittels des Kalibrationssegmentes am Ende des LC-Laufes. Im Falle, dass eine Kalibration der Messdaten nicht möglich war, wurde die Kalibration derjenigen zuvor gemessenen Probe – bei der eine Kalibration möglich war – übertragen.

Die weitere Auswertung erfolgte mittels der Funktion „*Find Molecular Features*“ im Programm DataAnalysis. Diese erlaubt das Zusammenfassen von Addukten derselben Substanz, die zur gleichen Retentionszeit auftreten, z. B. $[M+H]^+$, $[M+Na]^+$, $[M+K]^+$ oder $[M-H]^-$, $[M-H_2O-H]^-$, $[M-CO_2-H]^-$. Die dazu verwendeten Programmeinstellungen sind in **Tab. 9.13** zusammengefasst. Das *Feature-Finding* wurde im Retentionszeitbereich nach dem Injektionspeak (polar: ab 1.5 min; unpolar: ab 2.0 min) und vor dem Rückspülschritt (polar: bis 20 min; unpolar: bis 23 min) durchgeführt.

Tab. 9.13: Verwendete Einstellungen im Programm DataAnalysis für die Detektion von Features mittels der Funktion „Find Molecular Features“.

Parameter	Positiver Ionenmodus	Negativer Ionenmodus
S/N threshold	5	5
Correlation coefficient	0.7	0.7
Minimum compound length	8	8
Smoothing width	2	2
Additional smoothing	ausgewählt	ausgewählt
Addukte	M+H, M+NH ₄ , M+Na, M+K, M-H ₂ O+H, M-CO ₂ +H, M- NH ₃ +H	M-H, M-H ₂ O-H, M+Na-2H, M+Cl, M+HCOOH-H, M+CH ₃ COOH-H
Proteomics, CHNO	ausgewählt	ausgewählt
Background subtraction	Spectral	Spectral

Die detektierten Features konnten daraufhin entweder in DataAnalysis für jede individuelle Messung bezüglich m/z -Werte, Retentionszeiten, Peakhöhen sowie Peakflächen begutachtet oder mit ProfileAnalysis 2.1 weiter analysiert werden. Ersteres wurde für die Beurteilung der Eignung verschiedener Extraktionsmethoden zur Metabolitextraktion genutzt. Letzteres kam zur Evaluation des Sorteneinflusses sowie von herkunftsabhängig beeinflussten Markerfeatures zum Einsatz (für die exakte Verfahrensweise siehe **Abschn. 9.9.2** bzw. **9.9.3**).

Die Evaluation der optimalen Extraktionsmethoden erfolgte, indem die Features je Probe aus DataAnalysis exportiert wurden und dann mittels des Programms R 3.6.3 die Anzahl der Features, die Summe der Peakintensitäten und die Summe der Peakflächen (Areas) je Extraktionsmethode bestimmt wurden.

9.9.2 Erstellung von vergleichenden Peaklisten

Die vergleichende Untersuchung der Peakintensitäten in den einzelnen Proben eines Batches wurde mithilfe der Software ProfileAnalysis 2.1 durchgeführt. Diese ermöglichte die zusammenfassende Auswertung der Peakhöhen in mehreren Proben je Feature. Die dabei genutzten Parameter je nach LC-MS Methode, sind in **Tab. 9.14** zusammengestellt.

Die Zuordnung der detektierten Features innerhalb der verschiedenen Proben zueinander erfolgte in ProfileAnalysis über Toleranzbereiche der Retentionszeiten und der m/z -Verhältnisse (siehe auch **Tab. 9.14**). Wurden diese Kriterien für einen Peak in einer der Proben in Relation zu den übrigen Proben nicht erreicht, entsteht eine Lücke in der Peaktafel. In der Regel ist es jedoch nicht korrekt, dass eine einzelne Probe aus einer Reihe Proben derselben Matrix keinen Peak enthält. Daher wurde in ProfileAnalysis die Option fehlende Werte durch den kleinsten Wert zu ersetzen, ausgewählt.

Tab. 9.14: Einstellungen des Programms ProfileAnalysis zur Erstellung von vergleichenden Peaklisten.

Parameter	Auswahl	
Data selection and preprocessing	Find Molecular Features	
	Use time alignment	
Find molecular features	Load from previous processing without parameter checking	
Bucketing basis	Maximum Isotope	
Retention time range	Polar: 1.5-20 min	Unpolar: 2-23 min
Bucketing	Advanced bucketing	
	Time: 20 s	
	Mass: 20 mDa	
	Split bucket with multiple compounds	
Normalization	None	
Value count of Bucket	Polar: ≥ 33 %	Unpolar: ≥ 50 %
Missing value substitution	Minimum bucket value	
Bucket value transformation	None	

Die mittels ProfileAnalysis erzeugte Feature-Tabelle enthielt jeweils die Peakhöhen des jeweils größten Peaks eines jeden Features neben dessen m/z -Verhältnis und der Retentionszeit (RT in min) für die untersuchten Proben. Diese Tabelle wurde als csv-Datei exportiert und in Microsoft Excel 2010 umformatiert. Die weitere statistische Auswertung, insbesondere die multivariate Datenanalyse und die Erstellung von Klassifikationsmodellen basierte auf dieser Tabelle und erfolgte mittels R 3.6.3, wie in **Abschn. 9.14** dargestellt.

9.9.3 Bestimmung der Peakflächen von LC-MS(/MS)-Analysen und quantitative Auswertung

Zur Bestimmung der Peakflächen wurden die kalibrierten LC-MS- bzw. LC-MRM-Rohdaten in das Programm QuantAnalysis 2.1 geladen. Dieses erlaubte die halb-automatische Integration von Peakflächen der entsprechenden Ionenspuren (*extracted ion chromatograms*, EICs) der Peaks bzw. der betrachteten MRM-Übergänge der zuvor in DataAnalysis 4.1 kalibrierten Messdaten. Die Extraktionsbreite auf Ebene der m/z -Skala betrug dabei einheitlich ± 0.3 Da. Die Toleranz in Bezug auf die Retentionszeit wurde auf ± 30 s gesetzt. Bei MRM-Messungen wurden die Precursorionen auf MS^1 -Ebene geräteseitig vor der LC-MS-Analyse in der Messsoftware festgelegt. Da das q-ToF-Massenspektrometer die generierten Fragmentionen als hochaufgelöste *Full-Scan*-Massenspektren aufzeichnete, konnten die betrachteten Fragmentionen auf MS^2 -Ebene während der Auswertung im Nachhinein frei gewählt werden. Die Selektion der Fragmentionen fand daher in QuantAnalysis statt, entsprechend den in den LC-MS/MS-Methoden genannten MRM-Übergängen (siehe **Abschn. 9.7** und **9.8**).

Da die automatisierte Erkennung der Integrationsgrenzen fehlerbehaftet sein konnte, wurde jeder Peak in jeder Probe visuell auf korrekte Integrationsgrenzen überprüft und falls nötig diese manuell korrigiert. Die Peakflächen der *Qualifier* von MRM-Messungen wurden in QuantAnalysis zudem auf deren Vorhandensein zur selben Retentionszeit, wie der *Quantifier* sowie ein Signal-zu-Rausch-Verhältnis > 1 überprüft.

Die Peakflächen wurden anschließend als Tabelle im csv-Dateiformat exportiert und in Microsoft Excel 2010 umformatiert. Die weitere statistische Auswertung, insbesondere die multivariate Datenanalyse und die Erstellung von Klassifikationsmodellen basierte auf so erstellten Tabellen und erfolgte mittels R 3.6.3, wie in **Abschn. 9.14** dargestellt.

Im Falle der (relativen) Quantifizierung der Lipide sowie Tocopherole wurden aus den ermittelten Peakflächen der *Quantifier* die enthaltenen Konzentrationen wie folgt berechnet: In Microsoft Excel 2010 wurden die Peakflächen der Standardreihe zu den eingesetzten Konzentrationen in Relation gesetzt, sodass der Gehalt in der Probe bestimmt werden konnte, siehe **Gl. 9.1**. Somit konnten für jede Probe die absolute Konzentration der analysierten Substanzen angegeben werden.

$$y_{\text{Gehalt}} = F \cdot \frac{x - b}{m} \quad (9.1)$$

y_{Gehalt}	Berechneter Gehalt in der Probe (in mg/kg)
F	Einwiege-Faktor
x	Gemessene Peakfläche (in <i>counts</i>)
m	Steigung der Kalibriergeraden
b	Achsenabschnitt der Kalibriergeraden

Gl. 9.1: Gleichung zur Berechnung der Konzentrationen der Standardsubstanzen in den Proben aus den Peakflächen der *Quantifier*-Ionen auf MS²-Ebene.

Während dieses Vorgehen im Falle der Tocopherolbestimmung gewählt wurde, da jeweils die entsprechenden Analyten auch als Standardsubstanz vorlagen, war diese direkte analytische Bestimmung der selektierten Lipidmarker nicht trivial.

Einen einfacheren Weg zur absoluten Quantifizierung von Lipiden schlugen Fauland *et al.* vor, welche zur Bestimmung von Lipiden in Lipidtröpfchen aus Mausleber mittels LC-FT-ICR-Massenspektrometrie lediglich einen Gruppenstandard pro Lipidklasse (PC, PE, PS und TG) zur Validierung der Methode einsetzten³⁵⁶. Diese Quantifizierungsstrategie wurde in dieser Arbeit für die quantifizierten Lipidmarker angewandt, indem als Gruppenstandards die Lipide PC (36:2), PE (36:0), DG (32:0) und TG (48:0) zum Einsatz kamen. Ein Nachteil bei dieser Vorgehensweise ist jedoch, dass leicht unterschiedliche Ionisierung- bzw. Detektionseigenschaften jedes individuellen Lipids vernachlässigt werden, da für alle Vertreter einer Lipidklasse ein homogenes Verhalten angenommen wird. Dafür ist eine absolute Quantifizierung möglich, auch in dem Fall, dass für Analyten keine isotoopenmarkierte Standards verfügbar sind.

Eine andere Option stellt die relative Quantifizierung in Relation zu einer anderen Probe dar. Dabei werden die zu quantifizierenden Analyten ins Verhältnis zu einer Referenzprobe gesetzt, welche alle diese Analyten enthält. Auch diese Quantifizierungsstrategie wurde in dieser Arbeit angewandt. Dazu wurde eine Poolprobe aus je gleiche Mengen an Extrakt von jeweils einer Herkunftprobe je Land, genutzt. Von dieser Probe wurde eine Verdünnungsreihe erstellt, sodass für jede quantifizierte Markersubstanz in den Herkunftsproben ein „Verdünnungsfaktor“ in Relation zu der Poolprobe bestimmt werden konnte. So wurden die Nachteile der Quantifizierung mittels Gruppenstandard im Hinblick auf das individuelle Verhalten der einzelnen Markersubstanzen bei der LC-MS/MS-Analytik abgemildert. Jedoch war so keine absolute Quantifizierung der Marker in den Proben möglich.

Im Falle der Tocopherolbestimmung war dies nicht notwendig, da jeweils die entsprechenden Analyten auch als Standardsubstanz vorlagen und daher mittels externer Kalibration analysiert wurden.

9.9.4 Rechnerische Bestimmung der Validierungskennzahlen

Die Präzision bzw. Richtigkeit der Analysen wurde bestimmt, indem der Mittelwert sowie die RSD der analytisch ermittelten Konzentrationen in R 3.6.3 oder Microsoft Excel 2010 berechnet wurde^{389,390}. Die Berechnungen erfolgten mittels der Excel-Funktionen MITTELWERT() und STABW.S() bzw. der R-Funktionen *mean* und *sd*, siehe **Gl. 9.2** bzw. **Gl. 9.3**. Zudem wurde auch der Median berechnet, welcher unempfindlicher gegenüber Ausreißern ist. Dies wurde mittels der R-Funktion *median* aus dem R-Paket stats³⁹¹ durchgeführt.

$$\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n \quad (9.2)$$

$\bar{x}$	Mittelwert der Messwerte
x_n	Einzelne Messwerte
N	Anzahl der Messwerte

Gl. 9.2: Gleichung zur Berechnung von Mittelwerten.

$$s = \sqrt{\frac{1}{N-1} \sum_{n=1}^N (x_n - \bar{x})^2} \quad (9.3)$$

s	Standardabweichung der Messwerte
$\bar{x}$	Mittelwert der Messwerte
x_n	Einzelne Messwerte
N	Anzahl der Messwerte

Gl. 9.3: Gleichung zur Berechnung der Standardabweichung.

Die Präzision beschreibt inwieweit die gemessenen Konzentrationen eines Analyten um den tatsächlichen Gehalt streuen^{389,390}. Sie kann somit mittels der RSD anhand des Mittelwerts und der Standardabweichung rechnerisch bestimmt werden, siehe **Gl. 9.4**.

$$x_{RSD} = \frac{s}{\bar{x}} \cdot 100\% \quad (9.4)$$

x_{RSD}	RSD der Messwerte (in %)
s	Standardabweichung der Messwerte
$\bar{x}$	Mittelwert der Messwerte

Gl. 9.4: Gleichung zur Berechnung der RSD.

Die lineare Regression der Kalibrationsreihen erfolgte in Microsoft Excel 2010, indem einer Trendlinie im Diagramm erstellt wurde und Geradengleichung, sowie das Bestimmtheitsmaß R^2 angezeigt wurden.

Die Berechnung der NWG wurde nach der Kalibriergeradenmethode bestimmt^{389,392}. Die dazu notwendige Berechnung ist in **Gl. 9.5** dargestellt.

$$x_{NWG} = \frac{s \cdot t \cdot \sqrt{\frac{1}{p} + \frac{1}{n}}}{m} \quad (9.5)$$

x_{NWG}	NWG
s	Standardabweichung der Blindwerte
t	Konstante (aus Tabelle für einseitige STUDENT-t-Verteilung; $t = 2.132$ bei $\alpha = 0.05$; $P = 95\%$ und 4 Freiheitsgeraden) ^{389,392} .)
p	Anzahl der Parallelbestimmungen der Kalibration
n	Anzahl der Blindproben
m	Steigung der Kalibriergeraden

Gl. 9.5: Gleichung zur Berechnung der NWG.

9.10 FT-NIR-Spektroskopie

9.10.1 Evaluation der Probenform für FT-NIR-Spektroskopie

Da die Probenform (unvermahlen oder vermahlen) einen großen Unterschied in Hinblick auf die Schnelligkeit und damit Kosteneffizienz bedeuten würde, wurde der Einfluss der Probenform auf die erhaltenen FT-NIR-Spektren mittels der QC-Probe II untersucht. Dabei wurden folgende drei Fälle unterschieden:

- 1) Intaktes Maiskorn, „Keimling-ferne“ Seite.
- 2) Intaktes Maiskorn, „Keimling-nahe“ Seite.
- 3) Pulverförmige Probe, Probe im Vial zwischen Messungen jeweils dreifach invertiert.

Die Aufnahme von FT-NIR-Spektren erfolgte wie in **Abschn. 9.10.2** beschrieben. Jede der drei Messvarianten wurde in 50-facher Bestimmung durchgeführt. Im Fall 3 (pulverförmige Probe) wurde die nach *Variante A* (siehe **Abschn. 9.2**) ohne folgende Gefriertrocknung homogenisierte QC-Probe II genutzt.

9.10.2 Aufnahme von FT-NIR-Spektren

FT-NIR Spektren wurden mit einem Bruker TANGO FT-NIR-Spektrometer aufgenommen. Dabei kamen zwei verschiedene Messverfahren zum Einsatz: ganze, intakte Maiskörner wurden direkt auf den Strahlenausgang des Spektrometers gelegt und die Messung gestartet. Vermahlene Körnermaisproben wurden hingegen in Glasvials gefüllt und dann auf den Strahlenausgang des Spektrometers gestellt.

Die Steuerung des FT-NIR-Spektrometers erfolgte mittels des Messprogramms OPUS 7.5. Ein Hintergrundkorrektur wurde jeweils zur Beginn jeder Messreihe durchgeführt und innerhalb einer Messreihe jeweils nach Bedarf bei Aufforderung durch die Software. Die Aufnahme von FT-NIR-Spektren erfolgte dabei in allen Fällen mit identischen Parametern, die in **Tab. 9.15** zusammengestellt sind.

Tab. 9.15: Einstellungen für FT-NIR-Analysen in der Messsoftware OPUS.

Parameter	Wert
Anzahl Scans Probe	50
Anzahl Scans Hintergrund	50
Messbereich	11500-3950 cm ⁻¹
Auflösung	2 cm ⁻¹

Ein FT-NIR-Spektrum war somit das Resultat von 50 einzelnen Scans, die summiert wurden. Die Messung ergab 3725 Datenpunkte im Bereich von 11500 cm⁻¹ bis 3950 cm⁻¹. Die Messdaten wurden im JCAMP/DX-Format exportiert und in R 3.6.3 mittels des R-Paketes *hyperspec*³⁹³ eingelesen. Es wurde eine Datentabelle mit den Messwerten je Wellenzahl in den Spalten mit einer Probe je Zeile erstellt. Diese Tabelle diente als Grundlage für die *Preprozessierung* der FT-NIR-Spektren (siehe **Abschn. 9.13.2**) und die sich daran anschließende multivariate Datenanalyse sowie der Erstellung von Klassifikationsmodellen (siehe **Abschn. 9.14**).

9.11 Probenmessungen

Um die Herkunft von Körnermaisproben zu bestimmen, wurden insgesamt folgende in **Tab. 9.16** aufgeführten Analysen durchgeführt.

Tab. 9.16: Zusammenstellung der in dieser Arbeit durchgeführten Analysen.

Nr.	Probenart	Analysemethode	Analysezweck
1	QC-Probe	<i>non-targeted</i> LC-MS	Extraktionsoptimierung polare Metabolite
2	QC-Probe	<i>non-targeted</i> LC-MS	Extraktionsoptimierung unpolare Metabolite
3	Sortenprobenset	<i>non-targeted</i> LC-MS	Polares Metabolom
4	Sortenprobenset	<i>non-targeted</i> LC-MS	Unpolares Metabolom
5	Herkunftsprobenset	<i>non-targeted</i> LC-MS	Polares Metabolom
6	Herkunftsprobenset	<i>non-targeted</i> LC-MS	Unpolares Metabolom
7	QC-Probe	<i>non-targeted</i> LC-MS	Lagerungsstabilität von Lipidmarkern zur Herkunftsbestimmung
8	Herkunftsprobenset	LC-MRM-MS/MS	Bestimmung von Lipidmarkern zur Herkunftsbestimmung
9	Weitere Herkunftsproben (Frankreich bzw. Ungarn)	LC-MRM-MS/MS	Bestimmung von Lipidmarkern zur Herkunftsbestimmung
10	Teil des Herkunftsprobensets	LC-MRM-MS/MS	Bestimmung von α -Tocopherol und γ -Tocopherol
11	QC-Probe II	FT-NIR	Optimierung der Probenform
12	Sortenprobenset	FT-NIR	Sorteneinfluss vs. Herkunftseinfluss
13	Herkunftsprobenset	FT-NIR	Herkunftsbestimmung
14	Herkunftsprobenset	LC-MS + FT-NIR	Datenfusion zur Herkunftsbestimmung

Diese dienten unterschiedlichen Aspekten: Zunächst wurden Extraktionsmethoden zur Gewinnung von Extrakten des polaren sowie des unpolaren Metaboloms von Körnermais untersucht, indem die QC-Probe jeweils in Dreifachbestimmung mit insgesamt je zwölf Extraktionsmethoden getrennt für das polare sowie das unpolare Metabolom extrahiert und analysiert wurden (siehe **Abschn. 9.3.1** und **9.3.3**). Die dazugehörigen LC-MS-Analysen (siehe **Abschn. 9.6**) fanden dabei sowohl im positiven als auch im negativen Ionenmodus statt. Um einen eventuellen Einfluss der

Körnermaissorte auf die Herkunftsbestimmung zu analysieren, wurden zudem 81 Proben aus Landessortenversuchen mittels der LC-MS-Methoden untersucht.

Das Modell zur Herkunftsbestimmung von Körnermais mittels massenspektrometrischer Methoden wurde anhand der 151 Proben aus dem Herkunftsset erstellt. Dazu wurden die Proben sowohl mit der LC-MS-Methode für polare (Extraktion mit Methanol/Wasser 2:1 v/v, siehe **Abschn. 6.1.1**), als auch der Methode für die Detektion der unpolaren Metabolite (Extraktion nach BLIGH & DYER, siehe **Abschn. 6.1.2**) in beiden Ionenmodi analysiert (siehe **Abschn. 9.3.2** und **9.3.4**). Die Leistungsfähigkeit der Klassifikationsmodelle wurde anhand der jeweils insgesamt detektierten Metabolite untersucht. Die Entwicklung einer weitergehenden Klassifizierungsmethode zur Herkunftsbestimmung erfolgte aufgrund der Ergebnisse auf Grundlage der unpolaren Metabolite, welche im positiven Ionenmodus detektiert werden konnten. Hochpotente Markersubstanzen wurden zunächst selektiert (siehe **Abschn. 9.14.1** und **9.14.2**) und dann mittels LC-MS/MS-Fragmentierung identifiziert (siehe **Abschn. 9.6.4**). Ausgehend von 20 identifizierten Markersubstanzen wurde ein optimiertes RF-Klassifikationsmodell zur Herkunftsbestimmung erstellt (siehe **Abschn. 9.14.3** und **9.14.4**).

Die Eignung der in **Abschn. 6.4** selektierten und identifizierten 20 Marker zur Herkunftsbestimmung nicht nur in einem halb-quantitativen Maßstab, sondern auch bei einer absoluten Quantifizierung wurde anhand von neun ausgewählten Lipidmarkern mittels einer LC-MRM-Methode und einer externen Standardreihe quantifiziert (siehe **Abschn. 9.7**). Die Leistungsfähigkeit der quantitativen LC-MRM-Methode wurde außerdem durch die zusätzliche Messung von je zehn Proben aus Frankreich bzw. Ungarn überprüft, die mithilfe des RF-Herkunftsbestimmungsmodells vorhergesagt wurden (siehe **Abschn. 9.14.5**).

Die Routinetauglichkeit der LC-MRM-Methode wurde zudem untersucht, indem die Lagerungsstabilität der verwendeten Extrakte verifiziert wurde (siehe **Abschn. 9.4**).

Als Vergleichsmethode wurde in einem Teil der Herkunftsproben (104 von 151 Herkunftsproben) ebenfalls die Konzentrationen von α -Tocopherol sowie γ -Tocopherol massenspektrometrisch bestimmt (siehe **Abschn. 6.6.2** und **6.6.3**) und die Vorhersagequalität der Herkunftsbestimmung über das Tocopherolmuster mit den Ergebnissen der absoluten Quantifizierung von Lipidmarkern verglichen.

Im Bereich der FT-NIR-Spektroskopie wurden die optimale Probenform für die Aufnahme von FT-NIR-Spektren anhand der QC-Probe II gemäß **Abschn. 9.10.2** evaluiert. Anschließend erfolgte die Messung sämtlicher 81 Sortenproben sowie von insgesamt 101 Herkunftsproben als vermahlene Proben gemäß **Abschn. 9.10.2**. Die Verifizierung der Sortenunabhängigkeit der FT-NIR-Analytik erfolgte mittels der Proben aus den Landessortenversuchen (siehe **Abschn. 9.1**), während das Klassifikationsmodell zur Herkunftsbestimmung anhand der Herkunftsproben erstellt wurde. Die allgemeine Eignung des Modells und der Analysenmethode wurde zudem mittels je 36 Proben aus Frankreich bzw. Ungarn überprüft, welche mittels einer anderen Mühle vermahlen wurden und zudem gefriergetrocknet wurden (siehe **Abschn. 6.8.8**).

Um eine mögliche verbesserte Leistungsfähigkeit eines fusionierten Klassifikationsmodells zu untersuchen, wurden zudem beide analytische Methoden zusammen zur Klassifikation eingesetzt (siehe **Abschn. 9.15**). Dazu wurden die Analysedaten der 20 selektierten Markermetabolite der LC-

MS-Analyse und der vier FT-NIR-Markerbanden nach entsprechendem *Preprocessing* und einem *Range-Scaling* miteinander fusioniert und mittels dem RF-Algorithmus eine Klassifikation durchgeführt.

9.12 Univariate Datenanalyse

Die Überprüfung der Signifikanz von Unterschieden zwischen Probengruppen erfolgte auf der Ebene der unprozessierten Rohdaten mittels statistischer Tests.

Die Bestimmung der Signifikanz der Unterschiede zwischen Messwerten von zwei unterschiedlichen Probengruppen erfolgte über einen *t*-Test in R 3.6.3, indem die Funktion *t.test* aus dem R-Paket *stats* verwendet wurde³⁹¹. Die Grenze ab der ein Unterschied als signifikant angesehen wurde, war dabei ein *p*-Wert ≤ 0.01 . Dies entspricht einer Irrtumswahrscheinlichkeit von 1 %.

Die Bestimmung der Signifikanz der Unterschiede zwischen den Messwerten mehrerer ($n > 2$) verschiedener Probengruppen erfolgte über einen *One-Way*-ANOVA-Test in R 3.6.3, mittels der Funktion *aov.test* aus dem R-Paket *stats*³⁹¹. Das verwendete Signifikanzniveau betrug 99 % (*p*-Wert ≤ 0.01), d. h. die Irrtumswahrscheinlichkeit betrug 1 %.

Um die Ursache der Signifikanz bei mehreren Probengruppen weiter einzugrenzen, wurde zudem ein *post-Hoc* TUKEY-Test auf Grundlage der Ergebnisse des *One-Way*-ANOVA-Tests durchgeführt (*p*-Wert ≤ 0.01). Dies erfolgte durch die Funktion *TukeyHSD* in dem R-Paket *stats*³⁹¹.

9.13 Datenvorverarbeitung

9.13.1 *Preprocessing* für LC-MS Daten

Zur Gewährung der Ausgewogenheit des Einflusses der Veränderungen von niedrig konzentrierten Metaboliten gegenüber höher konzentrierten Metaboliten ist für die multivariate Datenauswertung sowie die Klassifikation der Proben ein *Preprocessing* der Rohdaten notwendig. Die *Preprozessierung* der LC-MS-Daten erfolgte in R 3.6.3³⁹¹. Die Peaktafel enthielt dabei spaltenweise die Features (Peakhöhen bzw. –flächen, erstellt entsprechend **Abschn. 9.9**) bzw. zeilenweise die Proben. Das *Pareto-Scaling* durchgeführt, indem zuerst jeweils der Mittelwert von den Einzelwerten abgezogen wurde (*mean centering*) und dann durch die Wurzel der Standardabweichung geteilt wurde²⁴². Die zugrundeliegende Gleichung ist in **Gl. 9.6** dargestellt.

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{\sqrt{s_i}} \quad (9.6)$$

$\tilde{x}_{ij}$	<i>Pareto</i> -skalierter Wert des <i>i</i> -ten Markers der <i>j</i> -ten Probe
x_{ij}	Unskalierter Wert des <i>i</i> -ten Markers der <i>j</i> -ten Probe
$\bar{x}_i$	Mittelwert der Messwerte des <i>i</i> -ten Markers in allen Proben
s_i	Standardabweichung der Messwerte des <i>i</i> -ten Markers in allen Proben

Gl. 9.6: Gleichung zur Durchführung des *Pareto-Scalings* von LC-MS-Analysedaten.

Es resultierte eine skalierte Matrix, welche wiederum entweder der multivariaten Datenanalyse (siehe **Abschn. 9.14.1**) bzw. dem RF-Klassifikationsalgorithmus (siehe **Abschn. 9.14.4**) als Ausgangsdaten diente. Die Skalierungsparameter je Markervariable, Mittelwert und

Standardabweichung, wurden gespeichert, um im Falle der Untersuchung von weiteren Proben zur Skalierung der vorherzusagenden Proben zum Einsatz zu kommen (siehe **Abschn. 9.14.5**).

Das *Preprocessing* der LC-MS-Daten für die Datenfusion erfolgte mittels des *Range-Scalings* (siehe **Abschn. 9.15.1**).

9.13.2 *Preprocessing*-Algorithmen für FT-NIR-Spektren

Die *Preprozessierung* von FT-NIR-Spektren erfolgte in R 3.6.3³⁹¹. Die *Consensus*-Datentabelle „*raw.data.matrix*“, mit den formatierten Rohdaten, wurde entsprechend den in **Tab. 9.17** gelisteten Algorithmen verarbeitet. Zunächst wurden diese mittels der Funktion *standardNormalVariate* aus dem R-Paket *prospectr*³⁹⁴ transformiert. Dann folgte die Ableitung der Messdaten mittels der Funktion *gapDer* und ein *Binning* mittels der Funktion *bin*, ebenso aus dem R-Paket *prospectr*. Es resultierte eine entsprechende skalierte Matrix, welche wiederum entweder der multivariaten Datenanalyse bzw. dem SVM-Klassifikationsalgorithmus (siehe **Abschn. 9.14**) als Ausgangsdaten diene.

Tab. 9.17: Evaluierte Algorithmen zur Auswahl eines optimalen *Preprocessing*-Verfahrens für FT-NIR-Spektren, sowie deren Funktionsaufrufe in R.

<i>Preprocessing</i> -Schritt	Funktionsaufruf	R-Paket
<i>Standard Normal Variate</i>	<code>snv.matrix = standardNormalVariate(raw.data.matrix)</code>	prospectr 0.2.1
1. Ableitung + <i>Binning</i>	<code>abl.matrix = gapDer(snv.matrix, m=1, w=11, s=11)</code>	
2. Ableitung + <i>Binning</i>	<code>abl.matrix = gapDer(snv.matrix, m=2, w=11, s=11)</code>	
<i>Binning</i>	<code>bin.matrix = bin(abl.matrix, bin.size=5)</code>	

Die Bewertung der verschiedenen *Preprocessing*-Varianten erfolgte anhand des Vergleichs der Klassifizierungsgenauigkeiten der prozessierten Messdaten von $n = 101$ Spektren der Herkunftsproben ohne Proben der Herkunft Frankreich und Ungarn. Als optimale *Preprocessing*-Methode für FT-NIR-Spektren wurde in **Abschn. 6.8.4** eine Kombination aus SNV gefolgt von der 1. Ableitung und einem *Binning* auf einen Abstand der Messpunkte von 10 cm^{-1} ermittelt. Der SNV-Algorithmus skaliert die einzelnen Spektren derart, dass sämtliche Messpunkte einen Mittelwert von 0 aufweisen, sowie eine Standardabweichung von 1^{248,250}. In **Gl. 9.7** ist der mathematische Zusammenhang des *Preprocessing* dargestellt.

$$\tilde{x}_i = \frac{x_i - \bar{x}}{s} \quad (9.7)$$

$\tilde{x}_i$	SNV-skaliertes Wert bei der i-ten Wellenzahl des Spektrums
x_i	Unskaliertes Wert bei der i-ten Wellenzahl des Spektrums
$\bar{x}$	Mittelwert der Messwerte im kompletten Spektrum
s	Standardabweichung der Messwerte im kompletten Spektrum

Gl. 9.7: Gleichung zur Durchführung des SNV-*Preprocessings* von FT-NIR-Spektren.

Das *Preprocessing* der FT-NIR-Spektren für die Datenfusion erfolgte mittels des *Range-Scalings* (siehe **Abschn. 9.15.1**).

9.14 Multivariate Datenanalyse und Klassifikation

9.14.1 PCA sowie PLS-DA

Um die Varianz eines Probensets allgemein zu beurteilen, wurden die preprozessierten Analysedaten einer PCA unterzogen. Um die Varianz der Analysedaten hinsichtlich der Unterschiede zwischen bestimmten Probenklassen zu visualisieren, wurden die Daten einer PLS-DA unterzogen.

Die praktische Durchführung beider Verfahren erfolgte jeweils in dem Statistikprogramm R 3.6.3³⁹¹. Die Matrix mit den preprozessierten Werten für die Features (in den Spalten) und die Proben (in den Zeilen) wurde für eine PCA der Funktion *prcomp* aus dem R-Paket *stats* übergeben³⁹¹. Die Funktion speicherte die ermittelten Scores in der Matrix *result\$x* und die Loading-Vektoren in der Matrix *result\$loadings*. Für eine PLS-DA wurde die Funktion *plsda* aus dem R-Paket *caret*³⁹⁵ genutzt. Im Falle der graphischen Darstellung beschränkte sich die Anzahl der verwendeten (Haupt-)Komponenten auf die ersten zwei bzw. drei zur zwei- bzw. dreidimensionalen Visualisierung des Scoresplots bzw. Loadingsplots.

Die Güte der Trennung der Proben mittels PLS-DA, wurde anhand des Q^2 -Wertes beurteilt. Der Q^2 -Wert entspricht dem Prozentsatz an korrekten Vorhersagen von Testproben im einem *Partial Least Squares Regression* Analysis (PLS-R) Modell. Die Berechnung wurde in R 3.6.3 durchgeführt, indem die Funktion *pls* aus dem R-Paket *pls* genutzt wurde³⁹⁶. Als Validierungsart wurde die 10-fache Kreuzvalidierung unter Nutzung der ersten drei Komponenten der PLS-DA gewählt.

9.14.2 Strategie zur Auswahl von Features zur Herkunftsbestimmung

Die Optimierung von Modellen zur Herkunftsbestimmung mittels Massenspektrometrie erfolgte als Folge der Erkenntnisse aus **Abschn. 6.1-6.3** lediglich für die unpolaren Metabolite, welche im positiven Ionenmodus detektiert werden konnten.

Die Vorauswahl von Features für die Herkunftsbestimmung erfolgte im Falle der LC-MS-Daten wie folgt: Die Analysedaten wurden in DataAnalysis 4.1 kalibriert und einem *Feature Finding* unterzogen (siehe **Abschn. 9.9.1**). Die so vorverarbeiteten Daten wurden in ProfileAnalysis 2.1 geladen und eine *Consensus*-Featureliste erstellt, welche sämtliche Features enthielt, die in mehr als der Hälfte der Proben vorkamen (siehe **Abschn. 9.9.2**). Die Analyse der Varianz erfolgte mittels PCA. Durch Abgleich der im PCA-Loadingsplot herausstechenden Markersubstanzen mit *m/z*-Werten aus den Metabolit-Datenbanken HMDB und LipidMaps, konnten diese als Lipide identifiziert werden.

Vielversprechende Features wurden mittels QuantAnalysis 2.1 integriert (siehe **Abschn. 9.9.3**) und einem *One-Way* ANOVA-Test nach Herkunft ($p \leq 0.01$), sowie einem *post-Hoc* TUKEY-Test unterzogen (**Abschn. 9.12**). Zudem wurden die entsprechenden Peaks in $n = 10$ QC-Proben integriert, um mittels der auftretenden RSD eine Aussage über die Analysengenauigkeit der jeweiligen Marker zu erhalten.

Der Markerpool wurde anschließend weiter auf hoch potente Marker eingegrenzt um die Anzahl der genutzten Marker auf ca. 20 Stück für ein Herkunftsbestimmungsmodell zu begrenzen. Diese Reduktion ist eine Voraussetzung, um eine Quantifizierung der Marker im Rahmen eines

Routineeinsatzes auf Triple-Quadrupol-Massenspektrometern zu erleichtern. Kriterien für die Annahme als hoch potentes Feature für die Bestimmung der Herkunft von Körnermais waren dabei:

- 1) Bei einem *One-Way* ANOVA-Test nach Herkunft einen p -Wert ≤ 0.01 ,
- 2) > 7 von insgesamt 21 möglichen Länderpaarungen mit p -Wert ≤ 0.01 im *post-Hoc* TUKEY-Test,
- 3) eine RSD ≤ 15 % in wiederholt analysierten Proben der QC-Probe ($n = 10$),

aufzuweisen. Um ein möglichst ausgeglichenes Modell zur Herkunftsbestimmung zu erhalten, wurde die Bedingung 2) zusätzlich erweitert, indem jedes der sieben Herkunftsländer die im Herkunftsprobenset vorhandenen Herkunftsproben, mindestens TUKEY-Test signifikant von einem weiteren Land unterscheidbar sein musste.

Es wurde außerdem auf eine Diversität der Markersubstanzen im Hinblick auf die Lipidklassen der PC, PE, DG und TG geachtet, sowie auf eine Verteilung der Substanzen auf der Retentionszeitskala, sodass nicht mehr als zwei Markersubstanzen zeitgleich eluierten. Marker mit höherer Intensität wurden im Zweifel bevorzugt, um die Analytik zu erleichtern. Daher waren alle selektierten Marker ebenso in den entsprechenden Gruppen der 20 intensivsten und signifikanten Marker vertreten.

Zusätzlich wurden Gruppen der 20 intensivsten Lipide je Lipidklasse (Phospholipide, DG bzw. TG) erstellt, welche einen signifikanten ANOVA-Test aufwiesen. Dieser Markerpool wurde ebenso auf seine Eignung als Marker in Klassifizierungsmodellen untersucht, um weitere markerbezogene Erkenntnisse zu gewinnen. Zudem dienten diese Lipidgruppen als Vergleichswerte zur Bewertung der Eignung von α - und γ -Tocopherol zur Herkunftsbestimmung sowie der Klassifikationsergebnisse mittels der quantifizierten Markersubstanzen.

Im Falle der FT-NIR-Spektren erfolgte keine Selektion von einzelnen Wellenzahlen. Hier wurde entweder das komplette Spektrum im Bereich von $10000\text{--}4000\text{ cm}^{-1}$ (*Grid Search*, Probenform, Analyse der Sortenproben) oder Spektren bei denen die beiden Wasserbanden (bei ca. 6870 und 5180 cm^{-1})³⁶⁷ entfernt wurden zur Klassifikation genutzt (Herkunftsbestimmung). Der Bereich von $11500\text{--}10000\text{ cm}^{-1}$ wurde nicht in die Auswertung mit einbezogen, da dieser Spektralbereich wenig Absorptionsbanden und ein starkes Rauschen aufwies.

Zur Analyse der Korrelation von NIR-Absorptionsbanden und Klassifizierungserfolg wurden zudem SVM-Klassifikationen anhand von relativ schmalen Spektrenbereichen vorgenommen. Diese deckten dabei bestimmte Absorptionsbanden von Proteinen ($4960\text{--}4490\text{ cm}^{-1}$), Stärke ($6520\text{--}6010\text{ cm}^{-1}$) oder Lipiden ($5990\text{--}5420\text{ cm}^{-1}$ bzw. $8600\text{--}7980\text{ cm}^{-1}$) ab³⁰³. Diese Spektrenbereiche wurden zuvor anhand des Loadingsplots einer PLS-DA Analyse nach Herkunft (siehe **Abschn. 9.14.1**) ermittelt. Als optimiertes Modell zur Herkunftsbestimmung wurde außerdem ein kombiniertes Modell aus den spezifischen Banden für Proteine, Stärke und Lipide erstellt.

9.14.3 Kreuzvalidierung der Klassifikationsmodelle

Zur Validierung der Klassifikationsergebnisse mit für das Modell unbekannten Proben und zur Optimierung von Parametern der Klassifikationsalgorithmen, ist es notwendig, das Probenset in einen Trainingsdatensatz und einen Testdatensatz zu unterteilen. Da der Trainingsdatensatz bereits zur Etablierung des Modells dient, ist dessen Rate an korrekten Zuordnungen der Proben

zu deren (Herkunfts-)Klassen nicht allgemein auf neue, dem Modell unbekannte Proben übertragbar^{240,265}. Die Allgemeingültigkeit der erstellten Modelle wurde daher mittels eines, dem Modell unbekannten Testdatensatzes überprüft, welches nicht zuvor an der Modellerstellung beteiligt war.

Die Aufteilung des gesamten Datensatzes in Trainings- bzw. Test-/Validierungsdatensätze erfolgte in dieser Arbeit durch die Methode der 10-fachen Kreuzvalidierung. Dieser Ansatz ist eine übliche Vorgehensweise zur Validierung von Klassifikationsmodellen bei Aufteilung in Trainings- und Test-/Validierungsdatensätze^{240,261,266,397}. Dabei werden sämtliche Proben eines Datensatzes zufällig in 10 ungefähr gleich große Gruppen unterteilt. Nacheinander werden nun 10 Klassifikationsmodelle erstellt, wobei jede der Subgruppen einmal das Test-Datenset darstellt, welches vorhergesagt wird. Das Modell wird jeweils anhand der übrigen neun Subgruppen erstellt. Somit sind die vorhergesagten Proben dem Modell jeweils unbekannt. Im Verlauf der Kreuzvalidierung fungiert jede individuelle Probe genau einmal als Testprobe, welche vorhergesagt wird. Dabei dienen je 90 % der Proben zur Modellerstellung und 10 % der Proben werden vorhergesagt²⁴⁰. Um die Vorhersagegüte der Klassifikationsmodelle weiter abzusichern, wurde der gesamte Vorgang der Kreuzvalidierung mehrfach wiederholt („*repeated cross validation*“^{240,261,266}). Diese Methode der Validierung wurde bereits in der Literatur mehrfach angewandt^{398–400}.

Die Teilung des Datensatzes wurde automatisch durch die Funktion *train* aus dem R-Paket *caret* in R 3.6.3 in Zuge der Durchführung der Klassifikation vorgenommen^{391,395}.

9.14.4 Klassifikationsalgorithmen

Die Klassifikation mittels verschiedener Klassifikationsalgorithmen erfolgte in R 3.6.3³⁹¹. Für die Klassifikation der Proben anhand von LC-MS-Analysedaten wurde der RF-Algorithmus genutzt; für FT-NIR-Analysedaten wurde der SVM-Algorithmus angewandt. Die Klassifikationsmodelle wurden mithilfe preprozessierter Messdaten (siehe **Abschn. 9.13**) erstellt. Die Klassifikationsmatrizen stellen die Zuordnungen der vorhergesagten Probenklassen zu den tatsächlichen Klassen der Proben dar.

Die Klassifikationen wurden mittels des R-Paketes *caret* sowohl 10-fach kreuzvalidiert, um die Leistungsfähigkeit bei dem Modell unbekannten Proben abzuschätzen, als auch 10 bzw. 100-fach wiederholt ausgeführt, um zufallsbedingte Effekte zu nivellieren (Funktion *train* mit den Parametern *number* = 10 (Anzahl Folds) und *repeats* = 10 oder 100)³⁹⁵. Teilweise wurden wiederholt ausgeführten Klassifikationen mit unterschiedlichen Markersubsets durchgeführt, dies ist im Ergebnisteil kenntlich gemacht.

Für Klassifikationen mittels RF nutzt die Funktion *train* dabei intern die Funktion *randomForest* aus dem gleichnamigen R-Paket²⁸⁶. Die Anzahl an Bäumen, die für eine RF-Klassifikation genutzt wurden, wurde auf *ntree* = 1000 festgesetzt, entsprechend wie von Creydt *et al.*⁴³ zur Klassifikation des Herkunftslandes von Spargel. Der Parameter *mtry* wurde auf die Quadratwurzel der Probenanzahl festgesetzt, was dem üblichen Wert^{282,285,401}, sowie den Standardeinstellungen der *randomForest*-Funktion entspricht.

Für SVM-Klassifikationen wurde intern von der *train*-Funktion auf die Funktion *ksvm* aus dem R-Paket *kernlab*²⁸⁰ zurückgegriffen. Für die Optimierung verschiedener Parameter der SVM-Modelle

wurden die Parameter systematisch variiert („*Grid Search*“). Es wurden folgende Kernelfunktionen untersucht: linear, quadratisch (entspr. polynomial 2.ten Grades) und die Radial Basis-Funktion (RBF). Während die lineare Kernelfunktion nur einen *Cost*-Parameter *C* zur Optimierung aufweist, sind im Falle der quadratischen Funktion bzw. der RBF zusätzlich zum *Cost*-Parameter, ein weiterer Parameter *scale* bzw. *sigma* zu optimieren. Der *Cost*-Parameter *C* wurde im Bereich von 10^{-5} bis 10^{+5} variiert (100 Abstufungen (linear) bzw. 20 Abstufungen (quadratisch/RBF)), *scale* im Bereich von 10^{-5} bis 10^{+1} (12 Abstufungen) und *sigma* im Bereich 10^{-6} bis 10^{-1} (10 Abstufungen). Die gewählte optimale Kernelfunktion für Klassifikationen von FT-NIR-Spektren von Körnermais war eine lineare Kernelfunktion. Für weitere Modelle wurde daher der *Cost*-Parameter *C* ebenso im Bereich von 10^{-4} bis 10^0 variiert, jedoch in 289 Abstufungen.

Die Variation erfolgte auch beim *Grid Search* im Rahmen einer 10-fachen Kreuzvalidierung, welche jeweils aus Gründen der langen Rechendauer in diesem Fall lediglich 10-fach wiederholt wurde. Die Selektion der optimalen *Cost*- sowie Kernelfunktionsparameter erfolgte daraufhin anhand der größten Klassifizierungsgenauigkeit (*Accuracy*). Die so ermittelten Parameter wurden schließlich in einer 100-fach wiederholten 10-fachen Kreuzvalidierung angewendet, deren Ergebnisse mitsamt Standardabweichung auf Ebene der Wiederholungen berichtet wurden (siehe auch **Abschn. 9.14.3**).

9.14.5 Vorhersage von weiteren Proben mittels existierender Klassifikationsmodelle

Die Klassifikation mittels verschiedener Klassifikationsalgorithmen erfolgte in R 3.6.3³⁹¹. Für die Vorhersage der je $n = 10$ weiteren Proben aus Frankreich bzw. Ungarn wurde das *Pareto-Scaling* dieser Proben anhand der bereits zuvor ermittelten Mittelwerte und Standardabweichungen der Proben mit denen das Vorhersagemodell erstellt wurde, vorgenommen. Somit konnten die neuen Proben exakt identisch zu jenen bereits im Modell vorhandenen behandelt werden.

Das R-Paket *caret* ermöglicht zwar die Vorhersage weiterer Proben mittels eines bereits erstellten Klassifikationsmodells mithilfe der Funktion *predict*. Jedoch beruht dabei die Vorhersage auf einem einzigen Modell aus allen Proben gemeinsam – die Ergebnisse der Funktion *predict* sind also nicht 10-fach kreuzvalidiert. Daher wurde ein leicht adaptiertes Verfahren zur Vorhersage von weiteren, dem Modell zuvor nicht bekannten Proben, angewandt:

Zunächst wurden die Unterteilungen des Probensets zur Erstellung des Modells in Trainings- und Validierungsset ausgelesen, welche bei der 100-fach wiederholten 10-fachen Kreuzvalidierung zum Einsatz kamen. Aus den entsprechenden Trainingssubsets der Proben im Rahmen der Kreuzvalidierungen, wurden jeweils die Klassifikationsmodelle erstellt²⁸⁶. Diese konnten dann genutzt werden, um einerseits die Validierungssubsets mittels der Funktion *predict* als Kontrolle vorherzusagen sowie die weiteren, unabhängigen Proben vorherzusagen. Die 10-fache Kreuzvalidierung wurde entsprechend dem Originalmodell ebenso 100-fach wiederholt. Dabei zeigte sich, dass die Vorhersagen der Validierungsproben mit den Ergebnissen, die über die Funktion *train* aus dem R-Paket *caret* erhalten wurden, übereinstimmten. Somit konnte die Methode der manuell kreuzvalidierten Modelle validiert werden und die neuen Proben vorhergesagt werden.

9.15 Datenfusion

9.15.1 *Preprocessing* der LC-MS bzw. FT-NIR-Messdaten vor Datenfusion

Für die Angleichung der unterschiedlichen Wertebereiche von LC-MS-Messdaten von $[0 \dots 4.6 \times 10^6]$ und der FT-NIR-Spektren nach SNV und 1. Ableitung und *Binning* von $[-0.95 \dots +2.96]$ vor einer *low-level* Datenfusion wurde in dieser Arbeit das *Range-Scaling* genutzt. Die *Preprozessierung* der Messdaten erfolgte in R 3.6.3³⁹¹. Die Tabellen enthielten spaltenweise die Features (Peakflächen der LC-MS-Analysen oder Absorptionen der FT-NIR-Analysen) und zeilenweise die Proben. Das *Range-Scaling* wurde durchgeführt, indem zuerst jeweils der Mittelwert von den Einzelwerten abgezogen wurde (*mean centering*) und dann durch die Differenz zwischen Maximum und Minimum der Messwerte geteilt wurde²⁴², siehe **Gl. 9.8**.

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_i}{(x_{i_{max}} - x_{i_{min}})} \quad (9.8)$$

$\tilde{x}_{ij}$	Range-skaliertes Wert des i-ten Markers der j-ten Probe
x_{ij}	Unskaliertes Wert des i-ten Markers der j-ten Probe
$\bar{x}_i$	Mittelwert der Messwerte des i-ten Markers in allen Proben
$x_{i_{max}}$	Maximum der Messwerte des i-ten Markers in allen Proben
$x_{i_{min}}$	Minimum der Messwerte des i-ten Markers in allen Proben

Gl. 9.8: Gleichung zur Durchführung des *Range-Scalings* von Analysedaten zur Datenfusion.

9.15.2 Durchführung der Datenfusion

Die Datenfusion wurde in dieser Arbeit auf *low-level* Ebene durchgeführt²⁹⁹. Als Klassifikationsalgorithmus wurde – aufgrund der guten Ergebnisse bei der Klassifikation von LC-MS-Messdaten – der *Random-Forest*-Algorithmus gewählt. Die Fusion der Daten erfolgte in R 3.6.3³⁹¹, indem die Matrizen der preprozessierten Analysedaten vor der Erstellung des Klassifikationsmodells zu einer Gesamtmatrix fusioniert wurden. Um den Einfluss des unterschiedlichen Wertebereiches der Analysedaten je nach apparativer Methode zu kompensieren, wurden die Wertebereiche der Daten zunächst entsprechend **Abschn. 9.15.1** mittels eines *Range-Scalings* angeglichen. Die Erstellung eines RF-Klassifikationsmodells erfolgte anschließend anhand der kombinierten Matrix mittels der *train*-Funktion aus dem R-Paket *caret*³⁹⁵. Das RF-Modell wurde gemäß **Abschn. 9.14.4** 10-fach kreuzvalidiert und die Klassifikationsgenauigkeit als Resultat von 100 Wiederholungen angegeben.

10 Literaturverzeichnis

- (1) Puma, M. J.; Bose, S.; Chon, S. Y.; Cook, B. I. Assessing the Evolving Fragility of the Global Food System. *Environ. Res. Lett.* **2015**, *10* (2), 024007.
- (2) Pinotti, L.; Dell'Orto, V. Feed Safety in the Feed Supply Chain. *Biotechnol Agron Soc Env.* **2011**, *15* (S1), 9–11.
- (3) Shiferaw, B.; Prasanna, B. M.; Hellin, J.; Bänziger, M. Crops That Feed the World 6. Past Successes and Future Challenges to the Role Played by Maize in Global Food Security. *Food Secur.* **2011**, *3* (3), 307–327.
- (4) United States Department of Agriculture (USDA); Federal Grain Inspection Service (FGIS). *Historical Compilation of Changes in the Grain Standards of the United States*; U.S. Standards; **1998**. https://www.gipsa.usda.gov/fgis/standards/history/standards_history.pdf, abgerufen am 15.08.2021.
- (5) Bundesministerium für Ernährung und Landwirtschaft (BMEL). *Deutschland, wie es isst. Der BMEL-Ernährungsreport 2020*; Berlin, **2020**; S. 32.
https://www.bmel.de/SharedDocs/Downloads/DE/Broschueren/ernaehrungsreport-2020.pdf?__blob=publicationFile&v=26, abgerufen am 06.10.2021.
- (6) Creydt, M.; Fischer, M. Omics Approaches for Food Authentication. *ELECTROPHORESIS* **2018**, *39* (13), 1569–1581.
- (7) Spink, J.; Moyer, D. C. Defining the Public Health Threat of Food Fraud. *J. Food Sci.* **2011**, *76* (9), R157–R163.
- (8) Manning, L.; Soon, J. M. Food Safety, Food Fraud, and Food Defense: A Fast Evolving Literature. *J. Food Sci.* **2016**, *81* (4), R823–R834.
- (9) Lebensmittel-, Bedarfsgegenstände- und Futtermittelgesetzbuch (Lebensmittel- und Futtermittelgesetzbuch - LFGB) § 11 Vorschriften zum Schutz vor Täuschung.
- (10) Verordnung (EG) Nr. 178/2002 des Europäischen Parlaments und des Rates vom 28. Januar 2002 zur Festlegung der allgemeinen Grundsätze und Anforderungen des Lebensmittelrechts, zur Errichtung der Europäischen Behörde für Lebensmittelsicherheit und zur Festlegung von Verfahren zur Lebensmittelsicherheit.
- (11) Verordnung (EG) Nr. 183/2005 des europäischen Parlaments und des Rates vom 12. Januar 2005 mit Vorschriften für die Futtermittelhygiene.
- (12) Bundesministerium für Ernährung und Landwirtschaft (BMEL); Bundesanstalt für Landwirtschaft und Ernährung (BLE). *Leitfaden zur Kontrolle der Umsetzung der Anforderungen an die Rückverfolgbarkeit von Futtermitteln in den Futtermittelunternehmen*; Futtermittelhygiene; Leitfaden Band 3; **2019**.
https://www.bmel.de/SharedDocs/Downloads/DE/_Tiere/Futtermittel/Betriebslisten/Leitfaden_Rueckverfolgbarkeit.pdf?__blob=publicationFile&v=2, abgerufen am 14.01.2022.
- (13) Durchführungsverordnung (EU) Nr. 884/2014 der Kommission vom 13. August 2014 zur Festlegung besonderer Bedingungen für die Einfuhr bestimmter Futtermittel und Lebensmittel aus bestimmten Drittländern wegen des Risikos einer Aflatoxin-Kontamination und zur Aufhebung der Verordnung (EG) Nr. 1152/2009.
- (14) Europol. *Operations: Opson*. <https://www.europol.europa.eu/operations/opson>, abgerufen am 19.09.2021.

- (15) Bundesamt für Verbraucherschutz und Lebensmittel. *OPSON-Operationen*. https://www.bvl.bund.de/DE/Arbeitsbereiche/01_Lebensmittel/03_Verbraucher/16_Food_Fraud/06_OPSON_Operationen/OPSON_Operationen_node.html, abgerufen am 14.08.2021.
- (16) Motta, G. A.; Tekinerdogan, B.; Athanasiadis, I. N. Blockchain Applications in the Agri-Food Domain: The First Wave. *Front. Blockchain* **2020**, *3*, 6.
- (17) Bibi, F.; Guillaume, C.; Gontard, N.; Sorli, B. A Review: RFID Technology Having Sensing Aptitudes for Food Industry and Their Contribution to Tracking and Monitoring of Food Products. *Trends Food Sci. Technol.* **2017**, *62*, 91–103.
- (18) Liang, K.; Chen, X.; He, R.; Li, J.; Okinda, C.; Han, D.; Shen, M. Development and Parameter Optimization of Automatic Separation and Identification Equipment for Grain Tracing Systems Based on Grain Tracers with QR Codes. *Comput. Electron. Agric.* **2019**, *162*, 709–718.
- (19) Qian, J.; Dai, B.; Wang, B.; Zha, Y.; Song, Q. Traceability in Food Processing: Problems, Methods, and Performance Evaluations—a Review. *Crit. Rev. Food Sci. Nutr.* **2020**, 1–14.
- (20) Galvez, J. F.; Mejuto, J. C.; Simal-Gandara, J. Future Challenges on the Use of Blockchain for Food Traceability Analysis. *TrAC Trends Anal. Chem.* **2018**, *107*, 222–232.
- (21) Esslinger, S.; Riedl, J.; Fauhl-Hassek, C. Potential and Limitations of Non-Targeted Fingerprinting for Authentication of Food in Official Control. *Food Res. Int.* **2014**, *60*, 189–204.
- (22) Teske, D.; Peters, A.; Möllers, A.; Fischer, M. Genomic Profiling: The Strengths and Limitations of Chloroplast Genome-Based Plant Variety Authentication. *J. Agric. Food Chem.* **2020**, *68* (49), 14323–14333.
- (23) Ortea, I.; O'Connor, G.; Maquet, A. Review on Proteomics for Food Authentication. *J. Proteomics* **2016**, *147*, 212–225.
- (24) Consonni, R.; Cagliani, L. R. The Potentiality of NMR-Based Metabolomics in Food Science and Food Authentication Assessment. *Magn. Reson. Chem.* **2019**, *57* (9), 558–578.
- (25) Dettmer, K.; Aronov, P. A.; Hammock, B. D. Mass Spectrometry-Based Metabolomics. *Mass Spectrom. Rev.* **2007**, *26* (1), 51–78.
- (26) Mendez, J.; Mendoza, L.; Cruz-Tirado, J. P.; Quevedo, R.; Siche, R. Trends in Application of NIR and Hyperspectral Imaging for Food Authentication. *Sci. Agropecu.* **2019**, *10* (1), 143–161.
- (27) Caporaso, N.; Whitworth, M. B.; Fisk, I. D. Near-Infrared Spectroscopy and Hyperspectral Imaging for Non-Destructive Quality Assessment of Cereal Grains. *Appl. Spectrosc. Rev.* **2018**, *53* (8), 667–687.
- (28) Guerre, P. Worldwide Mycotoxins Exposure in Pig and Poultry Feed Formulations. *Toxins* **2016**, *8* (12), 350.
- (29) Bundesinstitut für Risikobewertung (BfR). *Übergang von Aflatoxinen in Milch, Eier, Fleisch und Innereien*; Stellungnahme des BfR 009/2013; Berlin, 2013. <https://mobil.bfr.bund.de/cm/343/uebergang-von-aflatoxinen-in-milch-eier-fleisch-und-innereien.pdf>, abgerufen am 01.10.2021.
- (30) Carey, M.; Meharg, C.; Williams, P.; Marwa, E.; Jiujiu, X.; Farias, J. G.; De Silva, P. M. C. S.; Signes-Pastor, A.; Lu, Y.; Nicoloso, F. T.; Savage, L.; Campbell, K.; Elliott, C.; Adomako, E.; Green, A. J.; Moreno-Jiménez, E.; Carbonell-Barrachina, Á. A.; Triwardhani, E. A.; Pandiangan, F. I.; Haris, P. I.; Lawgali, Y. F.; Sommella, A.; Pigna, M.; Brabet, C.; Montet, D.; Njira, K.; Watts, M. J.; Meharg, A. A. Global Sourcing of Low-Inorganic Arsenic Rice Grain. *Expo. Health* **2020**, *12* (4), 711–719.
- (31) Gruber-Dorninger, C.; Jenkins, T.; Schatzmayr, G. Global Mycotoxin Occurrence in Feed: A Ten-Year Survey. *Toxins* **2019**, *11* (7), 375.

- (32) Van, T. T. H.; Yidana, Z.; Smooker, P. M.; Coloe, P. J. Antibiotic Use in Food Animals Worldwide, with a Focus on Africa: Pluses and Minuses. *J. Glob. Antimicrob. Resist.* **2020**, *20*, 170–177.
- (33) European Parliament. Directorate General for External Policies of the Union. *The Use of Pesticides in Developing Countries and Their Impact on Health and the Right to Food.*; **2021**.
- (34) Verordnung (EG) Nr. 767/2009 des Europäischen Parlaments und des Rates vom 13. Juli 2009 über das Inverkehrbringen und die Verwendung von Futtermitteln, zur Änderung der Verordnung (EG) Nr. 1831/2003 des Europäischen Parlaments und des Rates und zur Aufhebung der Richtlinien 79/373/EWG des Rates, 80/511/EWG der Kommission, 82/471/EWG des Rates, 83/228/EWG des Rates, 93/74/EWG des Rates, 93/113/EG des Rates und 96/25/EG des Rates und der Entscheidung 2004/217/EG der Kommission.
- (35) Bundesinstitut für Risikobewertung (BfR). *Bewertung von Dioxingehalten in Eiern auf Grund einer Warnung im EU-Schnellwarnsystem*; Stellungnahme des BfR 020/2010; Berlin, **2010**. https://mobil.bfr.bund.de/cm/343/bewertung_von_dioxingehalten_in_eiern_auf_grund_einer_warnung_im_eu_schnellwarnsystem.pdf, abgerufen am 01.10.2021.
- (36) Niedersächsisches Ministerium für Ernährung, Landwirtschaft und Verbraucherschutz. *Aflatoxin B1 in Futtermais - Aktueller Sachstand*; **2013**. https://www.ml.niedersachsen.de/startseite/service/archivierte_beitrage/amtsperiode_meyer/aflatoxin-b1-in-futtermais---chronologie-113258.html, abgerufen am 14.08.2021.
- (37) Niedersächsisches Ministerium für Ernährung, Landwirtschaft und Verbraucherschutz. *Mit ndl-PCB-belastetes Futtermittel: Aktueller Sachstand*; **2019**. <https://www.ml.niedersachsen.de/startseite/aktuelles/gefluegel-futter-belastet--aktuelle-fakten-171193.html>, abgerufen am 14.08.2021.
- (38) Bucchini, L.; Goldman, L. R. Starlink Corn: A Risk Analysis. *Environ. Health Perspect.* **2002**, *110* (1), 5–13.
- (39) Bratspies, R. M. Myths of Voluntary Compliance: Lessons from the StarLink Corn Fiasco. *William & Mary Environmental Law and Policy Review* **2003**, *27*, 593–647.
- (40) United States Department of Agriculture Economic Research Service. *Weights, Measures, and Conversion Factors for Agricultural Commodities and Their Products*; Agricultural Handbook; **1992**.
- (41) Lin, W.; Price, G. K.; Allen, E. W. StarLink: Impacts on the U.S. Corn Market and World Trade. *Agribusiness* **2003**, *19* (4), 473–488.
- (42) Bundesamt für Verbraucherschutz und Lebensmittel. *Das Europäische Schnellwarnsystem für Lebens- und Futtermittel (RASFF)*. https://www.bvl.bund.de/DE/Arbeitsbereiche/02_Futtermittel/01_Aufgaben/04_EU_Schnellwarnsystem/lm_schnellwarnsystem_aktuelle_rasff_meldungen_node.html, abgerufen am 10.10.2021.
- (43) Creydt, M.; Hudzik, D.; Rurik, M.; Kohlbacher, O.; Fischer, M. Food Authentication: Small-Molecule Profiling as a Tool for the Geographic Discrimination of German White Asparagus. *J. Agric. Food Chem.* **2018**, *66* (50), 13328–13339.
- (44) Klockmann, S.; Reiner, E.; Cain, N.; Fischer, M. Food Targeting: Geographical Origin Determination of Hazelnuts (*Corylus Avellana*) by LC-QqQ-MS/MS-Based Targeted Metabolomics Application. *J. Agric. Food Chem.* **2017**, *65* (7), 1456–1465.
- (45) Richter, B.; Rurik, M.; Gurk, S.; Kohlbacher, O.; Fischer, M. Food Monitoring: Screening of the Geographical Origin of White Asparagus Using FT-NIR and Machine Learning. *Food Control* **2019**, *104*, 318–325.

- (46) Klockmann, S.; Reiner, E.; Bachmann, R.; Hackl, T.; Fischer, M. Food Fingerprinting: Metabolomic Approaches for Geographical Origin Discrimination of Hazelnuts (*Corylus Avellana*) by UPLC-QTOF-MS. *J. Agric. Food Chem.* **2016**, *64* (48), 9253–9262.
- (47) Osborne, B. G. Near-Infrared Spectroscopy in Food Analysis. In *Encyclopedia of Analytical Chemistry*; Meyers, R. A., (Hrsg.); John Wiley & Sons: Chichester, UK, **2000**; S. a1018.
- (48) Cozzolino, D. An Overview of the Use of Infrared Spectroscopy and Chemometrics in Authenticity and Traceability of Cereals. *Food Res. Int.* **2014**, *60*, 262–265.
- (49) Miedaner, T. *Kulturpflanzen: Botanik - Geschichte - Perspektiven*; Kulturpflanzen; Springer Spektrum: Berlin, **2014**.
- (50) Bundesverband Deutscher Pflanzenzüchter e. V. *Mais*. <https://www.bdp-online.de/de/Pflanzenzuechtung/Kulturarten/Mais/>, abgerufen am 06.10.2021.
- (51) Peyker, W.; Degner, J.; Zorn, W.; Götz, R. *Leitlinie zur effizienten und umweltverträglichen Erzeugung von Körnermais*; Thüringer Landesanstalt für Landwirtschaft, (Hrsg.); 5. Aufl.; Jena, **2017**. http://www.tll.de/www/daten/publikationen/leitlinien/II_kmais.pdf, abgerufen am 16.08.2021.
- (52) proplanta Das Informationszentrum für die Landwirtschaft. *Nutzungsformen des Mais*. https://www.proplanta.de/Mais/Nutzungsformen_Pflanze1146727393.html, abgerufen am 24.09.2021.
- (53) Berkemeier, K. *Silomais: Den TS-Gehalt zur Abreife kontrollieren*. <https://www.elite-magazin.de/herdenmanagement/silomais-den-ts-gehalt-zur-abreife-kontrollieren-14885.html>, abgerufen am 24.09.2021.
- (54) Daccord, R.; Arrigo, Y.; Vogel, R. Nährwert von Maissilage. *Agrarforschung* **1995**, *2* (9), 397–400.
- (55) Müller, K. S. Futterwert aktueller Silomaisorten: In situ-Abbau der Trockenmasse der Restpflanze sowie in vivo-Verdaulichkeit der Rohnährstoffe der Ganzpflanze. Masterarbeit, Universität für Bodenkultur Wien, Wien, **2011**.
- (56) Bauböck, R.; Kappas, M. *Bioenergie im Landkreis Göttingen: GIS gestützte Biomassepotentialabschätzung anhand ausgewählter Kulturen, Triticale und Mais*; Ibidem Verlag: Berlin, **2012**.
- (57) Lieberei, R.; Reisdorff, C.; Franke, W. *Nutzpflanzenkunde*, 7. Aufl.; Thieme: Stuttgart, **2007**.
- (58) Landwirtschaftskammer Nordrhein-Westfalen. *Bestandesdichte und Reihenabstände*; **2015**. <https://www.landwirtschaftskammer.de/Landwirtschaft/ackerbau/mais/bestandsdichte-pdf.pdf>, abgerufen am 06.10.2021.
- (59) Kutka, F. Open-Pollinated vs. Hybrid Maize Cultivars. *Sustainability* **2011**, *3* (9), 1531–1554.
- (60) NASS - Agriculture Counts; United States Department of Agriculture (USDA). *Statistics by Subject: CROPS*. https://www.nass.usda.gov/Statistics_by_Subject/index.php?sector=CROPS, abgerufen am 02.10.2021.
- (61) Dei, H. K. Assessment of Maize (*Zea Mays*) as Feed Resource for Poultry. In *Poultry Science*; Manafi, M., (Hrsg.); InTech: London, UK, **2017**.
- (62) Shahidi, F.; de Camargo, A. Tocopherols and Tocotrienols in Common and Emerging Dietary Sources: Occurrence, Applications, and Health Benefits. *Int. J. Mol. Sci.* **2016**, *17* (10), 1745.
- (63) Landwirtschaftskammer Nordrhein-Westfalen; Erhardt, Norbert. *Ergebnisse der Landessortenversuche mit Körnermais 2020 in Nordrhein-Westfalen*; **2020**. <https://www.landwirtschaftskammer.de/Landwirtschaft/ackerbau/pdf/tabellen-koernermais-sv-2020.pdf>, abgerufen am 16.08.2021.

- (64) LfL Bayern Institut für Pflanzenbau und Pflanzenzüchtung. *Ergebnisse der Bayerischen Landessortenversuche Körnermais 2020*; **2020**.
https://www.lfl.bayern.de/mam/cms07/ipz/dateien/km20_ergebnisse_lsv.pdf, abgerufen am 24.09.2021.
- (65) Verordnung (EG) Nr. 1829/2003 des Europäischen Parlaments und des Rates vom 22. September 2003 über genetisch veränderte Lebensmittel und Futtermittel.
- (66) Bullock, D. S.; Desquilbet, M. The Economics of Non-Gmo Segregation and Identity Preservation. *Food Policy* **2002**, 27 (1), 81–99.
- (67) Lin, W.; Price, G. K.; Allen, E. StarLink: Impacts on the U.S. Corn Market and World Trade. In *Feed Situation and Outlook Yearbook*; **2001**; S. 46–54.
- (68) Albala, K. *The SAGE Encyclopedia of Food Issues*; SAGE Publications, Los Angeles, CA, USA, **2015**.
- (69) Bullock, D. S.; Desquilbet, M.; Nitsi, E. The Economics of Non-GMO Segregation and Identity Preservation. In *Food Policy*; Tampa, FL, USA, **2000**.
- (70) Flora, G. Realities of GMO Segregation. Institute for Agriculture and Trade Policy; **1999**.
https://www.iatp.org/sites/default/files/Realities_of_GMO_Segregation_051799.htm, abgerufen am 28.09.2021.
- (71) Agricultural Biotechnology Council of Australia. *StarLink Corn*; Issues Paper; 6; **2012**.
https://www.abca.com.au/wp-content/uploads/2012/09/ABCA_IssuesPaper_6_v2.pdf, abgerufen am 27.09.2021.
- (72) Federation of American Scientists. *StarLink Corn: A Cautionary Tale*.
<https://biosecurity.fas.org/education/dualuse-agriculture/2.-agricultural-biotechnology/starlink-corn.html>, abgerufen am 27.09.2021.
- (73) Deutsches Maiskomitee e.V. (DMK). *Tierische Veredlung*.
https://www.maiskomitee.de/Verwertung/Tierische_Veredlung, abgerufen am 18.08.2021.
- (74) Kumar, D.; Singh, V. Bioethanol Production From Corn. In *Corn*; Elsevier: Amsterdam, Niederlande **2019**; S. 615–631.
- (75) Hugger, H. Trocknung – der Kostenfaktor im Körnermaisbau. *Mais* **2005**, 3/2005 (32. Jg.), 76–79.
- (76) Strehler, A. *Trocknung von Getreide, Körnermais und Raps im landwirtschaftlichen Betrieb*, überarbeitete Aufl.; Deutsche Landwirtschafts-Gesellschaft: Frankfurt a. M., **1996**.
- (77) Böckelmann, M. Erfolgreiche Trocknung von Getreide Und Körnermais, **2016**.
- (78) Bern, C. J.; Quick, G.; Herum, F. L. Harvesting and Postharvest Management. In *Corn*; Elsevier: Amsterdam, Niederlande **2019**; S. 109–145.
- (79) Scherer, R.; Kutzbach, H. D.; Thaler, M.; Müller, H.-M. Atmungswärme und Atmungsverluste von Körnermais. *Grundl Landtech.* **1980**, 30 (4), 104–110.
- (80) United States Department of Agriculture (USDA); Federal Grain Inspection Service (FGIS). *Subpart D -- United States Standards for Corn*; **1996**.
<https://www.gipsa.usda.gov/fgis/standards/810corn.pdf>, abgerufen am 14.08.2021.
- (81) Codex Alimentarius. *Standard for Maize (Corn)*; CXS 153-1985; FAO, **1985**.
https://www.fao.org/input/download/standards/51/CXS_153e.pdf, abgerufen am 01.02.2025.
- (82) Verordnung (EU) Nr. 642/2010 der Kommission vom 20. Juli 2010 mit Durchführungsbestimmungen zur Verordnung (EG) Nr. 1234/2007 des Rates hinsichtlich der Einfuhrzölle im Getreidesektor.

- (83) Verordnung (EU) Nr. 1272/2009 der Kommission vom 11. Dezember 2009 mit gemeinsamen Durchführungsbestimmungen zur Verordnung (EG) Nr. 1234/2007 des Rates hinsichtlich des An- und Verkaufs von landwirtschaftlichen Erzeugnissen im Rahmen der öffentlichen Intervention.
- (84) Egesel, C. Ö.; Kahriman, F. Determination of Quality Parameters in Maize Grain by NIR Reflectance Spectroscopy. *Tarım Bilim. Derg.* **2012**, *18* (1), 31–42.
- (85) Jiang, H. Y.; Zhu, Y. J.; Wei, L. M.; Dai, J. R.; Song, T. M.; Yan, Y. L.; Chen, S. J. Analysis of Protein, Starch and Oil Content of Single Intact Kernels by near Infrared Reflectance Spectroscopy (NIRS) in Maize (*Zea Mays* L.). *Plant Breed.* **2007**, *126* (5), 492–497.
- (86) Paulsen, M. R.; Singh, M.; Singh, V. Measurement and Maintenance of Corn Quality. In *Corn*; Elsevier: Amsterdam, Niederlande, **2019**; S. 165–211.
- (87) Vernieres, C.; Hurburgh, C. R.; Rippke, G.; Hardy, C. L. *Near Infrared Spectroscopy (NIRS) Calibrations for Moisture in High Moisture Corn Co-Products*; Iowa State University: Ames, IA, USA, **2016**.
- (88) Levasseur-Garcia, C. Updated Overview of Infrared Spectroscopy Methods for Detecting Mycotoxins on Cereals (Corn, Wheat, and Barley). *Toxins* **2018**, *10* (1), 38.
- (89) Qiu, G.; Lü, E.; Lu, H.; Xu, S.; Zeng, F.; Shui, Q. Single-Kernel FT-NIR Spectroscopy for Detecting Supersweet Corn (*Zea Mays* L. *Saccharata* Sturt) Seed Viability with Multivariate Data Analysis. *Sensors* **2018**, *18* (4), 1010.
- (90) Zhang, J.; Dai, L.; Cheng, F. Classification of Frozen Corn Seeds Using Hyperspectral VIS/NIR Reflectance Imaging. *Molecules* **2019**, *24* (1), 149.
- (91) United States Department of Agriculture Foreign Agricultural Service. *Grain: World Markets and Trade*; **2021**. <https://downloads.usda.library.cornell.edu/usda-esmis/files/zs25x844t/6m312f84z/3r075m960/grain.pdf>, abgerufen am 27.05.2021.
- (92) eurostat. *Agricultural production - crops*. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Agricultural_production_-_crops#Cereals, abgerufen am 18.08.2021.
- (93) Europäische Kommission. *Agricultural production - crops*. https://ec.europa.eu/info/food-farming-fisheries/plants-and-plant-products/plant-products/cereals_de, abgerufen am 18.08.2021.
- (94) Bundesanstalt für Landwirtschaft und Ernährung (BLE). *Bericht zur Markt- und Versorgungslage Getreide 2019*; Bonn, **2019**. https://www.ble.de/SharedDocs/Downloads/DE/BZL/Daten-Berichte/Getreide_Getreideerzeugnisse/2019BerichtGetreide.pdf?__blob=publicationFile&v=3, abgerufen am 06.10.2021.
- (95) DMK / bwagrar.de. *Wie Körnermais verwendet wird*. <https://www.bwagrar.de/Aktuelles/Politik/Wie-Koernermais-verwendet-wird,QUIEPTQ5ODk5NTYmTUIEPTUyNjEz.html>, abgerufen am 15.08.2021.
- (96) United States Department of Agriculture Foreign Agricultural Service. *Grain: World Markets and Trade*; **2019**. <https://downloads.usda.library.cornell.edu/usda-esmis/files/zs25x844t/pr76fg43p/j098zq631/grain.pdf>, abgerufen am 01.10.2019.
- (97) Confédération Européenne de la Production de Mais (CEPM). *Newsletter CEPM*; Newsletter 28; Paris, **2021**. <https://www.maizeurop.com/wp-content/uploads/2021/07/Newsletter-CEPM-28-EN.pdf>, abgerufen am 17.08.2021.
- (98) Hunt, E. D.; Birge, H. E.; Laingen, C.; Licht, M. A.; McMechan, J.; Baule, W.; Connor, T. A Perspective on Changes across the U.S. Corn Belt. *Environ. Res. Lett.* **2020**, *15* (7), 071001.

- (99) National Corn Growers Association. *Corn Usage by Segment 2020*. <https://www.worldofcorn.com/#corn-usage-by-segment>, abgerufen am 17.08.2021.
- (100) U.S. Grains Council. *A Guide To Distiller's Dried Grains with Solubles (DDGS)*. <https://grains.org/wp-content/uploads/2018/01/Complete-2012-DDGS-Handbook.pdf>, abgerufen am 10.07.2021.
- (101) destatis Statistisches Bundesamt. *Anbauflächen, Hektarerträge und Erntemengen ausgewählter Anbaukulturen im Zeitvergleich*. <https://www.destatis.de/DE/Themen/Branchen-Unternehmen/Landwirtschaft-Forstwirtschaft-Fischerei/Feldfruechte-Gruenland/Tabellen/liste-feldfruechte-zeitreihe.html>, abgerufen am 18.08.2021.
- (102) destatis Statistisches Bundesamt. *Anbauflächen und Erntemengen von ökologisch angebautem Gemüse*. <https://www.destatis.de/DE/Themen/Branchen-Unternehmen/Landwirtschaft-Forstwirtschaft-Fischerei/Obst-Gemuese-Gartenbau/Tabellen/oekologisches-gemuese.html>, abgerufen am 18.08.2021.
- (103) Foods and Agriculture Organisation (FAO) of the United Nations. *FAOSTAT*. <http://www.fao.org/faostat/en/#data/PP>, abgerufen am 18.08.2021.
- (104) Goodwin, S.; McPherson, J. D.; McCombie, W. R. Coming of Age: Ten Years of next-Generation Sequencing Technologies. *Nat. Rev. Genet.* **2016**, *17* (6), 333–351.
- (105) Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: A Revolutionary Tool for Transcriptomics. *Nat. Rev. Genet.* **2009**, *10* (1), 57–63.
- (106) Gibson, G. The Environmental Contribution to Gene Expression Profiles. *Nat. Rev. Genet.* **2008**, *9* (8), 575–581.
- (107) Stefano, G. B. Comparing Bioinformatic Gene Expression Profiling Methods: Microarray and RNA-Seq. *Med. Sci. Monit. Basic Res.* **2014**, *20*, 138–142.
- (108) Zhou, H.; Ning, Z.; E. Starr, A.; Abu-Farha, M.; Figeys, D. Advancements in Top-Down Proteomics. *Anal. Chem.* **2012**, *84* (2), 720–734.
- (109) Rabilloud, T. Two-Dimensional Gel Electrophoresis in Proteomics: Old, Old Fashioned, but It Still Climbs up the Mountains. *Proteomics* **2002**, *2* (1), 3–10.
- (110) Nicholson, J. K.; Lindon, J. C.; Holmes, E. “Metabonomics”: Understanding the Metabolic Responses of Living Systems to Pathophysiological Stimuli via Multivariate Statistical Analysis of Biological NMR Spectroscopic Data. *Xenobiotica* **1999**, *29* (11), 1181–1189.
- (111) Fiehn, O. Combining Genomics, Metabolome Analysis, and Biochemical Modelling to Understand Metabolic Networks. *Comp. Funct. Genomics* **2001**, *2* (3), 155–168.
- (112) Nicholson, J. K.; Lindon, J. C. Metabonomics *Nature* **2008**, *455*, 1054–1056.
- (113) Aretz, I.; Meierhofer, D. Advantages and Pitfalls of Mass Spectrometry Based Metabolome Profiling in Systems Biology. *Int. J. Mol. Sci.* **2016**, *17* (5), 632.
- (114) Begou, O.; Gika, H. G.; Wilson, I. D.; Theodoridis, G. Hyphenated MS-Based Targeted Approaches in Metabolomics. *The Analyst* **2017**, *142* (17), 3079–3100.
- (115) Liu, Z.; Rochfort, S. Recent Progress in Polar Metabolite Quantification in Plants Using Liquid Chromatography-Mass Spectrometry. *J. Integr. Plant Biol.* **2014**, *56* (9), 816–825.
- (116) Courant, F.; Antignac, J.-P.; Dervilly-Pinel, G.; Le Bizec, B. Basics of Mass Spectrometry Based Metabolomics. *PROTEOMICS* **2014**, *14* (21–22), 2369–2388.
- (117) Benkeblia, N. Metabolomics and Postharvest Sciences: Challenges and Perspectives. *Acta Hort.* **2014**, No. 1047, 303–308.
- (118) Oliver, S. Systematic Functional Analysis of the Yeast Genome. *Trends Biotechnol.* **1998**, *16* (9), 373–378.

- (119) Ernst, M.; Silva, D. B.; Silva, R. R.; Vêncio, R. Z. N.; Lopes, N. P. Mass Spectrometry in Plant Metabolomics Strategies: From Analytical Platforms to Data Acquisition and Processing. *Nat. Prod. Rep.* **2014**, *31* (6), 784.
- (120) Cobb, M. 60 Years Ago, Francis Crick Changed the Logic of Biology. *PLOS Biol.* **2017**, *15* (9), e2003243.
- (121) Dervilly-Pinel, G.; Courant, F.; Chéreau, S.; Royer, A.-L.; Boyard-Kieken, F.; Antignac, J.-P.; Monteau, F.; Le Bizec, B. Metabolomics in Food Analysis: Application to the Control of Forbidden Substances. *Drug Test. Anal.* **2012**, *4*, 59–69.
- (122) Christoph, N.; Hermann, A.; Wachter, H. 25 Years Authentication of Wine with Stable Isotope Analysis in the European Union – Review and Outlook. *BIO Web Conf.* **2015**, *5*, 02020.
- (123) Creydt, M.; Fischer, M. Food Authentication in Real Life: How to Link Nontargeted Approaches with Routine Analytics? *ELECTROPHORESIS* **2020**, *41* (20), 1665–1679.
- (124) Girelli, C. R.; Del Coco, L.; Papadia, P.; De Pascali, S. A.; Fanizzi, F. P. Harvest Year Effects on Apulian EVOOs Evaluated by ¹H NMR Based Metabolomics. *PeerJ* **2016**, *4*, e2740.
- (125) Hatzakis, E. Nuclear Magnetic Resonance (NMR) Spectroscopy in Food Science: A Comprehensive Review. *Compr. Rev. Food Sci. Food Saf.* **2019**, *18* (1), 189–220.
- (126) Valand, R.; Tanna, S.; Lawson, G.; Bengtström, L. A Review of Fourier Transform Infrared (FTIR) Spectroscopy Used in Food Adulteration and Authenticity Investigations. *Food Addit. Contam. Part A* **2020**, *37* (1), 19–38.
- (127) Rubert, J.; Zachariasova, M.; Hajslova, J. Advances in High-Resolution Mass Spectrometry Based on Metabolomics Studies for Food – a Review. *Food Addit. Contam. Part A* **2015**, *32* (10), 1685–1708.
- (128) Danezis, G. P.; Tsagkaris, A. S.; Camin, F.; Brusic, V.; Georgiou, C. A. Food Authentication: Techniques, Trends & Emerging Approaches. *TrAC Trends Anal. Chem.* **2016**, *85*, 123–132.
- (129) Herrero, M.; García-Cañas, V.; Simo, C.; Cifuentes, A. Recent Advances in the Application of Capillary Electromigration Methods for Food Analysis and Foodomics. *ELECTROPHORESIS* **2010**, *31* (1), 205–228.
- (130) van de Velde, B.; Guillarme, D.; Kohler, I. Supercritical Fluid Chromatography – Mass Spectrometry in Metabolomics: Past, Present, and Future Perspectives. *J. Chromatogr. B* **2020**, *1161*, 122444.
- (131) Dai, W.; Xie, D.; Lu, M.; Li, P.; Lv, H.; Yang, C.; Peng, Q.; Zhu, Y.; Guo, L.; Zhang, Y.; Tan, J.; Lin, Z. Characterization of White Tea Metabolome: Comparison against Green and Black Tea by a Nontargeted Metabolomics Approach. *Food Res. Int.* **2017**, *96*, 40–45.
- (132) Sundekilde, U.; Larsen, L.; Bertram, H. NMR-Based Milk Metabolomics. *Metabolites* **2013**, *3* (2), 204–222.
- (133) Humston, E. M.; Zhang, Y.; Brabeck, G. F.; McShea, A.; Synovec, R. E. Development of a GC×GC-TOFMS Method Using SPME to Determine Volatile Compounds in Cacao Beans. *J. Sep. Sci.* **2009**, *32* (13), 2289–2295.
- (134) Drivelos, S. A.; Danezis, G. P.; Halagarda, M.; Popek, S.; Georgiou, C. A. Geographical Origin and Botanical Type Honey Authentication through Elemental Metabolomics via Chemometrics. *Food Chem.* **2021**, *338*, 127936.
- (135) Mi, S.; Shang, K.; Li, X.; Zhang, C.-H.; Liu, J.-Q.; Huang, D.-Q. Characterization and Discrimination of Selected China's Domestic Pork Using an LC-MS-Based Lipidomics Approach. *Food Control* **2019**, *100*, 305–314.

- (136) Rubert, J.; Lacina, O.; Zachariasova, M.; Hajslova, J. Saffron Authentication Based on Liquid Chromatography High Resolution Tandem Mass Spectrometry and Multivariate Data Analysis. *Food Chem.* **2016**, *204*, 201–209.
- (137) Pinotti, L.; Moretti, V. M.; Baldi, A.; Bellagamba, F.; Campagnoli, A.; Savoini, G.; Cantoni, C.; Dell’Orto, V. Feed Authentication as an Essential Component of Food Safety and Control. *Outlook Agric.* **2005**, *34* (4), 243–248.
- (138) Santos Pereira, C.; C. Cunha, S.; Fernandes, J. O. Prevalent Mycotoxins in Animal Feed: Occurrence and Analytical Methods. *Toxins* **2019**, *11* (5), 290.
- (139) Fernández-Ibáñez, V.; Fearn, T.; Soldado, A.; de la Roza-Delgado, B. Development and Validation of near Infrared Microscopy Spectral Libraries of Ingredients in Animal Feed as a First Step to Adopting Traceability and Authenticity as Guarantors of Food Safety. *Food Chem.* **2010**, *121* (3), 871–877.
- (140) Tres, A.; van Ruth, S. M. Verification of Organic Feed Identity by Fatty Acid Fingerprinting. *J. Agric. Food Chem.* **2011**, *59* (16), 8816–8821.
- (141) Tres, A.; van der Veer, G.; Perez-Marin, M. D.; van Ruth, S. M.; Garrido-Varo, A. Authentication of Organic Feed by Near-Infrared Spectroscopy Combined with Chemometrics: A Feasibility Study. *J. Agric. Food Chem.* **2012**, *60* (33), 8129–8133.
- (142) Schwendel, B. H.; Morel, P. C. H.; Wester, T. J.; Tavendale, M. H.; Deadman, C.; Fong, B.; Shadbolt, N. M.; Thatcher, A.; Otter, D. E. Fatty Acid Profile Differs between Organic and Conventionally Produced Cow Milk Independent of Season or Milking Time. *J. Dairy Sci.* **2015**, *98* (3), 1411–1425.
- (143) Ellis, K. A.; Innocent, G.; Grove-White, D.; Cripps, P.; McLean, W. G.; Howard, C. V.; Mihm, M. Comparing the Fatty Acid Composition of Organic and Conventional Milk. *J. Dairy Sci.* **2006**, *89* (6), 1938–1950.
- (144) Shurson, J.; Spiehs, M.; Whitney, M.; Knott, J. *Nutritional and Value Added Benefits of Feeding Maize DDGS and Other Dry-Mill Co-Products to Swine*.
<https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=abcb9b4f5ff28ca97f346eb0c9eda45e85abdbd1>, abgerufen am 09.07.2021.
- (145) Tres, A.; Heenan, S. P.; van Ruth, S. Authentication of Dried Distilled Grain with Solubles (DDGS) by Fatty Acid and Volatile Profiling. *LWT - Food Sci. Technol.* **2014**, *59* (1), 215–221.
- (146) Vermeulen, P.; Nietner, T.; Haughey, S. A.; Yang, Z.; Tena, N.; Chmelarova, H.; van Ruth, S.; Tomaniova, M.; Boix, A.; Han, L.; Elliott, C. T.; Baeten, V.; Fauhl-Hassek, C. Origin Authentication of Distillers’ Dried Grains and Solubles (DDGS)—Application and Comparison of Different Analytical Strategies. *Anal. Bioanal. Chem.* **2015**, *407* (21), 6447–6461.
- (147) Nietner, T.; Pfister, M.; Glomb, M. A.; Fauhl-Hassek, C. Authentication of the Botanical and Geographical Origin of Distillers Dried Grains and Solubles (DDGS) by FT-IR Spectroscopy. *J. Agric. Food Chem.* **2013**, *61* (30), 7225–7233.
- (148) Nietner, T.; Haughey, S. A.; Ogle, N.; Fauhl-Hassek, C.; Elliott, C. T. Determination of Geographical Origin of Distillers Dried Grains and Solubles Using Isotope Ratio Mass Spectrometry. *Food Res. Int.* **2014**, *60*, 146–153.
- (149) Zhou, X.; Yang, Z.; Haughey, S. A.; Galvin-King, P.; Han, L.; Elliott, C. T. Classification the Geographical Origin of Corn Distillers Dried Grains with Solubles by near Infrared Reflectance Spectroscopy Combined with Chemometrics: A Feasibility Study. *Food Chem.* **2015**, *189*, 13–18.
- (150) Tena, N.; Boix, A.; von Holst, C. Identification of Botanical and Geographical Origin of Distillers Dried Grains with Solubles by near Infrared Microscopy. *Food Control* **2015**, *54*, 103–110.

- (151) Baniasadi, H.; Vlahakis, C.; Hazebroek, J.; Zhong, C.; Asiago, V. Effect of Environment and Genotype on Commercial Maize Hybrids Using LC/MS-Based Metabolomics. *J. Agric. Food Chem.* **2014**, *62* (6), 1412–1422.
- (152) Tang, W.; Hazebroek, J.; Zhong, C.; Harp, T.; Vlahakis, C.; Baumhover, B.; Asiago, V. Effect of Genetics, Environment, and Phenotype on the Metabolome of Maize Hybrids Using GC/MS and LC/MS. *J. Agric. Food Chem.* **2017**, *65* (25), 5215–5225.
- (153) Zhao, H.; Guo, B.; Wei, Y.; Zhang, B. Near Infrared Reflectance Spectroscopy for Determination of the Geographical Origin of Wheat. *Food Chem.* **2013**, *138* (2–3), 1902–1907.
- (154) Zhao, H.; Guo, B.; Wei, Y.; Zhang, B. Effects of Grown Origin, Genotype, Harvest Year, and Their Interactions of Wheat Kernels on near Infrared Spectral Fingerprints for Geographical Traceability. *Food Chem.* **2014**, *152*, 316–322.
- (155) González-Martín, M. I.; Wells Moncada, G.; González-Pérez, C.; Zapata San Martín, N.; López-González, F.; Lobos Ortega, I.; Hernández-Hierro, J.-M. Chilean Flour and Wheat Grain: Tracing Their Origin Using near Infrared Spectroscopy and Chemometrics. *Food Chem.* **2014**, *145*, 802–806.
- (156) De Girolamo, A.; Cortese, M.; Cervellieri, S.; Lippolis, V.; Pascale, M.; Logrieco, A. F.; Suman, M. Tracing the Geographical Origin of Durum Wheat by FT-NIR Spectroscopy. *Foods* **2019**, *8* (10), 450.
- (157) Revilla, I.; Lastras, C.; González-Martín, M. I.; Vivar-Quintana, A. M.; Morales-Corts, R.; Gómez-Sánchez, M. A.; Pérez-Sánchez, R. Predicting the Physicochemical Properties and Geographical ORIGIN of Lentils Using near Infrared Spectroscopy. *J. Food Compos. Anal.* **2019**, *77*, 84–90.
- (158) Sun, J.; Lu, X.; Mao, H.; Jin, X.; Wu, X. A Method for Rapid Identification of Rice Origin by Hyperspectral Imaging Technology: Identification of Rice Origin by Hyperspectral Imaging. *J. Food Process Eng.* **2017**, *40* (1), e12297.
- (159) Fernández Pierna, J. A.; Baeten, V.; Dardenne, P. Screening of Compound Feeds Using NIR Hyperspectral Data. *Chemom. Intell. Lab. Syst.* **2006**, *84* (1–2), 114–118.
- (160) Pan, Z.; Raftery, D. Comparing and Combining NMR Spectroscopy and Mass Spectrometry in Metabolomics. *Anal. Bioanal. Chem.* **2007**, *387* (2), 525–527.
- (161) Emwas, A.-H.; Roy, R.; McKay, R. T.; Tenori, L.; Saccenti, E.; Gowda, G. A. N.; Raftery, D.; Alahmari, F.; Jaremko, L.; Jaremko, M.; Wishart, D. S. NMR Spectroscopy for Metabolomics Research. *Metabolites* **2019**, *9* (7), 123.
- (162) Marshall, D. D.; Powers, R. Beyond the Paradigm: Combining Mass Spectrometry and Nuclear Magnetic Resonance for Metabolomics. *Prog. Nucl. Magn. Reson. Spectrosc.* **2017**, *100*, 1–16.
- (163) Ellis, D. I.; Brewster, V. L.; Dunn, W. B.; Allwood, J. W.; Golovanov, A. P.; Goodacre, R. Fingerprinting Food: Current Technologies for the Detection of Food Adulteration and Contamination. *Chem. Soc. Rev.* **2012**, *41* (17), 5706.
- (164) Wishart, D. S. Advances in Metabolite Identification. *Bioanalysis* **2011**, *3* (15), 1769–1782.
- (165) Gathungu, R. M.; Kautz, R.; Kristal, B. S.; Bird, S. S.; Vouros, P. The Integration of LC-MS and NMR for the Analysis of Low Molecular Weight Trace Analytes in Complex Matrices. *Mass Spectrom. Rev.* **2020**, *39* (1–2), 35–54.
- (166) Gromski, P. S.; Muhamadali, H.; Ellis, D. I.; Xu, Y.; Correa, E.; Turner, M. L.; Goodacre, R. A Tutorial Review: Metabolomics and Partial Least Squares-Discriminant Analysis – a Marriage of Convenience or a Shotgun Wedding. *Anal. Chim. Acta* **2015**, *879*, 10–23.

- (167) Wishart, D. S. NMR Metabolomics: A Look Ahead. *J. Magn. Reson.* **2019**, *306*, 155–161.
- (168) Ramautar, R.; Somsen, G. W.; de Jong, G. J. CE-MS in Metabolomics. *ELECTROPHORESIS* **2009**, *30* (1), 276–291.
- (169) Pilařová, V.; Plachká, K.; Khalikova, M. A.; Svec, F.; Nováková, L. Recent Developments in Supercritical Fluid Chromatography – Mass Spectrometry: Is It a Viable Option for Analysis of Complex Samples? *TrAC Trends Anal. Chem.* **2019**, *112*, 212–225.
- (170) Scalbert, A.; Brennan, L.; Fiehn, O.; Hankemeier, T.; Kristal, B. S.; van Ommen, B.; Pujos-Guillot, E.; Verheij, E.; Wishart, D.; Wopereis, S. Mass-Spectrometry-Based Metabolomics: Limitations and Recommendations for Future Progress with Particular Focus on Nutrition Research. *Metabolomics* **2009**, *5* (4), 435–458.
- (171) Katerinopoulou, K.; Kontogeorgos, A.; Salmas, C. E.; Patakas, A.; Ladavos, A. Geographical Origin Authentication of Agri-Food Products: A Review. *Foods* **2020**, *9* (4), 489.
- (172) Roberts, J.; Power, A.; Chapman, J.; Chandra, S.; Cozzolino, D. A Short Update on the Advantages, Applications and Limitations of Hyperspectral and Chemical Imaging in Food Authentication. *Appl. Sci.* **2018**, *8* (4), 505.
- (173) Yakubu, H. G.; Kovacs, Z.; Toth, T.; Bazar, G. The Recent Advances of Near-Infrared Spectroscopy in Dairy Production—a Review. *Crit. Rev. Food Sci. Nutr.* **2020**, 1–22.
- (174) *Near-Infrared Spectroscopy: Theory, Spectral Analysis, Instrumentation, and Applications*; Ozaki, Y., Huck, C., Tsuchikawa, S., Engelsen, S. B., (Hrsg.); Springer Singapore: Singapur, **2021**.
- (175) *Near-Infrared Spectroscopy: Principles, Instruments, Applications*; Siesler, H. W., (Hrsg.); Wiley-VCH: Weinheim, **2002**.
- (176) Siebert, F.; Hildebrandt, P. *Vibrational Spectroscopy in Life Science*; Wiley-VCH: Weinheim, **2007**.
- (177) Pasquini, C. Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. *J. Braz. Chem. Soc.* **2003**, *14* (2), 198–219.
- (178) Fernández Pierna, J. A.; Abbas, O.; Lecler, B.; Hogrel, P.; Dardenne, P.; Baeten, V. NIR Fingerprint Screening for Early Control of Non-Conformity at Feed Mills. *Food Chem.* **2015**, *189*, 2–12.
- (179) Cozzolino, D.; Restaino, E.; La Manna, A.; Fernandez, E.; Fassio, A. Usefulness of near Infrared Reflectance (NIR) Spectroscopy and Chemometrics to Discriminate between Fishmeal, Meat Meal and Soya Meal Samples. *Cienc. E Investig. Agrar.* **2009**, *36* (2).
- (180) K. Maertens; P. Reyns; J. De Baerdemaeker. On-Line Measurement of Grain Quality with Nir Technology. In *International Conference on Crop Harvesting and Processing*; American Society of Agricultural and Biological Engineers, **2003**.
- (181) Nicolaï, B. M.; Beullens, K.; Bobelyn, E.; Peirs, A.; Saeys, W.; Theron, K. I.; Lammertyn, J. Nondestructive Measurement of Fruit and Vegetable Quality by Means of NIR Spectroscopy: A Review. *Postharvest Biol. Technol.* **2007**, *46* (2), 99–118.
- (182) Clarke, F. C.; Hammond, S. V.; Jee, R. D.; Moffat, A. C. Determination of the Information Depth and Sample Size for the Analysis of Pharmaceutical Materials Using Reflectance Near-Infrared Microscopy. *Appl Spectrosc* **2002**, *56* (11), 1475–1483.
- (183) Ciurczak, E. W.; Igne, Benoît; Workman, J.; Burns, D. A. *Handbook of Near-Infrared Analysis*; **2021**.
- (184) Watari, M. Applications of Near-Infrared Spectroscopy to Process Analysis Using Fourier Transform Spectrometer. *Opt. Rev.* **2010**, *17* (3), 317–322.

- (185) Nielsen, M. M.; Hansen, Å. Rapid High-Performance Liquid Chromatography Determination of Tocopherols and Tocotrienols in Cereals. *Cereal Chem. J.* **2008**, *85* (2), 248–251.
- (186) Kok, André. From Benchtop to Process NIR Control, **2013**.
<http://www.uco.es/investigacion/proyectos/nirsplatform/icnirs/images/NIR2013/KOK.pdf>,
abgerufen am 29.08.2021.
- (187) Bruker Optics. *The Bruker ROCKSOLID Interferometer*; Product Note T14-01/04; **2004**.
https://www.cigs.unimore.it/CigsDownloads/labs/FTIR2/manuali_lettore/rocksolid.pdf, abgerufen
am 30.08.2021.
- (188) Hanssen, L.; Snail, K. Integrating Spheres for Mid- and Near-Infrared Reflection Spectroscopy. In *Handbook of Vibrational Spectroscopy*; Chalmers, J. M., Griffiths, P. R., (Hrsg.); John Wiley & Sons: Chichester, UK, **2002**.
- (189) Fornstedt, T.; Forssén, P.; Westerlund, D. Basic HPLC Theory and Definitions: Retention, Thermodynamics, Selectivity, Zone Spreading, Kinetics, and Resolution. In *Analytical Separation Science*; Pino, V., Anderson, J. L., Berthod, A., Stalcup, A. M., (Hrsg.); Wiley-VCH: Weinheim, **2015**; S. 1–24.
- (190) Meyer, V. R. *Praxis der Hochleistungs-Flüssigchromatographie*, 9. Aufl.; Wiley-VCH: Weinheim, **2006**.
- (191) van Deemter, J. J.; Zuiderweg, F. J.; Klinkenberg, A. Longitudinal Diffusion and Resistance to Mass Transfer as Causes of Nonideality in Chromatography. *Chem. Eng. Sci.* **1995**, *50* (24), 3869–3882.
- (192) Allwood, J. W.; Goodacre, R. An Introduction to Liquid Chromatography–Mass Spectrometry Instrumentation Applied in Plant Metabolomic Analyses. *Phytochem. Anal.* **2010**, *21* (1), 33–47.
- (193) Sanders, K. L.; Edwards, James L. Nano-Liquid Chromatography-Mass Spectrometry and Recent Applications in Omics Investigations. *Anal Methods* **2020**, *12*, 4404–4417.
- (194) Thermo Fischer Scientific. *UltiMate™ 3000 binäres Standardsystem*.
<https://www.thermofisher.com/order/catalog/product/IQLAAAGABHFAPBMBEX#/IQLAAAGABHFA PBMBEX>,
abgerufen am 03.10.2021.
- (195) Moldoveanu, S. C.; David, V. Basic Information about HPLC. In *Essentials in Modern HPLC Separations*; Elsevier: Amsterdam, Niederlande, **2013**; S. 1–51.
- (196) Jiang, P.; Lucy, C. A. Coupling Normal Phase Liquid Chromatography with Electrospray Ionization Mass Spectrometry: Strategies and Applications. *Anal. Methods* **2016**, *8* (35), 6478–6488.
- (197) Gross, J. H. *Massenspektrometrie*; Springer: Berlin, Heidelberg, **2013**.
- (198) Vailaya, A. Fundamentals of Reversed Phase Chromatography: Thermodynamic and Exothermodynamic Treatment. *J. Liq. Chromatogr. Relat. Technol.* **2005**, *28* (7–8), 965–1054.
- (199) Kirkland, J. J.; van Straten, M. A.; Claessens, H. A. High PH Mobile Phase Effects on Silica-Based Reversed-Phase High-Performance Liquid Chromatographic Columns. *J. Chromatogr. A* **1995**, *691* (1–2), 3–19.
- (200) Dorsey, J. G.; Dill, K. A. The Molecular Mechanism of Retention in Reversed-Phase Liquid Chromatography. *Chem. Rev.* **1989**, *89* (2), 331–346.
- (201) Sandoval, J. E.; Pesek, J. J. Synthesis and Characterization of a Hydride-Modified Porous Silica Material as an Intermediate in the Preparation of Chemically Bonded Chromatographic Stationary Phases. *Anal. Chem.* **1989**, *61* (18), 2067–2075.

- (202) Pesek, J. J.; Matyska, M. T. Silica Hydride-Based Packing Materials: HPLC Stationary Phases for a Global Approach to Complex Sample Analysis. *Curr. Chromatogr.* **2018**, *5* (1), 33–42.
- (203) Pesek, Joseph; Matyska, Maria T. A Comparison of Two Separation Modes: HILIC and Aqueous Normal Phase Chromatography. *LCGC N. Am.* **2007**, *25* (2), 480–490.
- (204) Matyska, M.; Pesek, J. The Development of Silica Hydride Stationary Phases for High-Performance Liquid Chromatography from Conception to Commercialization. *Separations* **2019**, *6* (2), 27.
- (205) Pesek, J. J.; Matyska, M. T. Hydride-Based Silica Stationary Phases for HPLC: Fundamental Properties and Applications. *J. Sep. Sci.* **2005**, *28* (15), 1845–1854.
- (206) Pesek, J. J.; Matyska, M. T.; Boysen, R. I.; Yang, Y.; Hearn, M. T. W. Aqueous Normal-Phase Chromatography Using Silica-Hydride-Based Stationary Phases. *TrAC Trends Anal. Chem.* **2013**, *42*, 64–73.
- (207) Pesek, J. J.; Matyska, M. T.; Fischer, S. M.; Sana, T. R. Analysis of Hydrophilic Metabolites by High-Performance Liquid Chromatography–Mass Spectrometry Using a Silica Hydride-Based Stationary Phase. *J. Chromatogr. A* **2008**, *1204* (1), 48–55.
- (208) Callahan, D. L.; Souza, D. D.; Bacic, A.; Roessner, U. Profiling of Polar Metabolites in Biological Extracts Using Diamond Hydride-Based Aqueous Normal Phase Chromatography. *J. Sep. Sci.* **2009**, *32* (13), 2273–2280.
- (209) Soukup, J.; Janás, P.; Jandera, P. Gradient Elution in Aqueous Normal-Phase Liquid Chromatography on Hydrosilated Silica-Based Stationary Phases. *J. Chromatogr. A* **2013**, *1286*, 111–118.
- (210) Gross, J. H. *Massenspektrometrie: Spektroskopiekurs kompakt*; Springer Spektrum: Berlin, **2019**.
- (211) Smith, R. D.; Loo, J. A.; Edmonds, C. G.; Barinaga, C. J.; Udseth, H. R. New Developments in Biochemical Mass Spectrometry: Electrospray Ionization. *Anal. Chem.* **1990**, *62* (9), 882–899.
- (212) Pitt, J. J. Principles and Applications of Liquid Chromatography–Mass Spectrometry in Clinical Biochemistry. *Clin. Biochem. Rev.* **2009**, *30* (1), 19–34.
- (213) El-Aneel, A.; Cohen, A.; Banoub, J. Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. *Appl. Spectrosc. Rev.* **2009**, *44* (3), 210–230.
- (214) Awad, H.; Khamis, M. M.; El-Aneel, A. Mass Spectrometry, Review of the Basics: Ionization. *Appl. Spectrosc. Rev.* **2015**, *50* (2), 158–175.
- (215) Gross, J. H. Direct Analysis in Real Time—a Critical Review on DART-MS. *Anal. Bioanal. Chem.* **2014**, *406* (1), 63–80.
- (216) Junot, C.; Fenaille, F.; Colsch, B.; Bécher, F. High Resolution Mass Spectrometry Based Techniques at the Crossroads of Metabolic Pathways. *Mass Spectrom. Rev.* **2014**, *33* (6), 471–500.
- (217) Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B. Molecular Beams of Macroions. *J. Chem. Phys.* **1968**, *49* (5), 2240–2249.
- (218) Yamashita, M.; Fenn, J. B. Electrospray Ion Source. Another Variation on the Free-Jet Theme. *J. Phys. Chem.* **1984**, *88* (20), 4451–4459.
- (219) Kebarle, P.; Verkerk, U. H. Electrospray: From Ions in Solution to Ions in the Gas Phase, What We Know Now. *Mass Spectrom. Rev.* **2009**, *28* (6), 898–917.
- (220) Dass, C. *Fundamentals of Contemporary Mass Spectrometry*; Wiley-Interscience: Hoboken, N.J., **2007**.
- (221) *Bioinformatics of Human Proteomics*; Wang, X., (Hrsg.); Translational bioinformatics; Springer Dordrecht: Dordrecht, Niederlande, **2013**.

- (222) Iribarne, J. V. On the Evaporation of Small Ions from Charged Droplets. *J. Chem. Phys.* **1976**, *64* (6), 2287.
- (223) Godzien, J.; Ciborowski, M.; Martínez-Alcázar, M. P.; Samczuk, P.; Kretowski, A.; Barbas, C. Rapid and Reliable Identification of Phospholipids for Untargeted Metabolomics with LC–ESI–QTOF–MS/MS. *J. Proteome Res.* **2015**, *14* (8), 3204–3216.
- (224) Creydt, M.; Fischer, M. Plant Metabolomics: Maximizing Metabolome Coverage by Optimizing Mobile Phase Additives for Nontargeted Mass Spectrometry in Positive and Negative Electrospray Ionization Mode. *Anal. Chem.* **2017**, *89* (19), 10474–10486.
- (225) Huang, N.; Siegel, M. M.; Kruppa, G. H.; Laukien, F. H. Automation of a Fourier Transform Ion Cyclotron Resonance Mass Spectrometer for Acquisition, Analysis, and e-Mailing of High-Resolution Exact-Mass Electrospray Ionization Mass Spectral Data. *J. Am. Soc. Mass Spectrom.* **1999**, *10* (11), 1166–1173.
- (226) Romson, J.; Emmer, Å. Mass Calibration Options for Accurate Electrospray Ionization Mass Spectrometry. *Int. J. Mass Spectrom.* **2021**, *467*, 116619.
- (227) Mamyrin, B. A. Laser Assisted Reflectron Time-of-Flight Mass Spectrometry. *Int. J. Mass Spectrom. Ion Process.* **1994**, *131*, 1–19.
- (228) Schrader, W.; Klein, H.-W. Liquid Chromatography/Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (LC-FTICR MS): An Early Overview. *Anal. Bioanal. Chem.* **2004**, *379* (7–8).
- (229) Bruker Daltonics. MaXis User Manual. **2008**, Version 1.1.
- (230) Analytical Techniques in Forensic Science; Wolstenholme, R., Jickells, S., Forbes, S., (Hrsg.); John Wiley & Sons: Hoboken, NJ, USA, **2020**.
- (231) Kim, T. K.; Park, J. H. More about the Basic Assumptions of T-Test: Normality and Sample Size. *Korean J. Anesthesiol.* **2019**, *72* (4), 331–335.
- (232) Osborn, C. E. Statistical Applications for Health Information Management, 2. Aufl.; Jones and Bartlett Publishers: Sudbury, MA, USA, **2006**.
- (233) Gromski, P.; Xu, Y.; Kotze, H.; Correa, E.; Ellis, D.; Armitage, E.; Turner, M.; Goodacre, R. Influence of Missing Values Substitutes on Multivariate Analysis of Metabolomics Data. *Metabolites* **2014**, *4* (2), 433–452.
- (234) Bro, R.; Smilde, A. K. Principal Component Analysis. *Anal Methods* **2014**, *6* (9), 2812–2831.
- (235) Worley, B.; Powers, R. Multivariate Analysis in Metabolomics. *Curr. Metabolomics* **2012**, *1* (1), 92–107.
- (236) Wold, S.; Sjöström, M.; Eriksson, L. PLS-Regression: A Basic Tool of Chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58* (2), 109–130.
- (237) Fordellone, M.; Bellincontro, A.; Mencarelli, F. Partial Least Squares Discriminant Analysis: A Dimensionality Reduction Method to Classify Hyperspectral Data. *ArXiv180609347 Stat* **2018**.
- (238) Archana, T.; Sachin, D. Dimensionality Reduction and Classification through PCA and LDA. *Int. J. Comput. Appl.* **2015**, *122* (17), 4–8.
- (239) Baxter, M. J. A Review of Supervised and Unsupervised Pattern Recognition in Archaeometry. *Archaeometry* **2006**, *48* (4), 671–694.
- (240) Raschka, S. Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. *ArXiv181112808 Cs Stat* **2020**.
- (241) Kessler, W. Multivariate Datenanalyse für die Pharma-, Bio- und Prozessanalytik; **2007**.

- (242) 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* **2006**, 7 (1), 142.
- (243) Karaman, I. Preprocessing and Pretreatment of Metabolomics Data for Statistical Analysis. In *Metabolomics: From Fundamentals to Clinical Applications*; Sussulini, A., (Hrsg.); Springer International Publishing: Cham, Schweiz, **2017**; Vol. 965, S. 145–161.
- (244) Katajamaa, M.; Orešič, M. Data Processing for Mass Spectrometry-Based Metabolomics. *J. Chromatogr. A* **2007**, 1158 (1–2), 318–328.
- (245) Böcker, S.; Letzel, M. C.; Lipták, Z.; Pervukhin, A. SIRIUS: Decomposing Isotope Patterns for Metabolite Identification. *Bioinformatics* **2009**, 25 (2), 218–224.
- (246) Kuhl, C.; Tautenhahn, R.; Böttcher, C.; Larson, T. R.; Neumann, S. CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets. *Anal. Chem.* **2012**, 84 (1), 283–289.
- (247) Alonso, A.; Marsal, S.; Julià, A. Analytical Methods in Untargeted Metabolomics: State of the Art in 2015. *Front. Bioeng. Biotechnol.* **2015**, 3.
- (248) Rinnan, Å.; Berg, F. van den; Engelsen, S. B. Review of the Most Common Pre-Processing Techniques for near-Infrared Spectra. *TrAC Trends Anal. Chem.* **2009**, 28 (10), 1201–1222.
- (249) Nansen, C.; Kolomiets, M.; Gao, X. Considerations Regarding the Use of Hyperspectral Imaging Data in Classifications of Food Products, Exemplified by Analysis of Maize Kernels. *J. Agric. Food Chem.* **2008**, 56 (9), 2933–2938.
- (250) Barnes, R. J.; Dhanoa, M. S.; Lister, S. J. Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Appl. Spectrosc.* **1989**, 43 (5), 772–777.
- (251) Romero-Torres, S.; Pérez-Ramos, J. D.; Morris, K. R.; Grant, E. R. Raman Spectroscopic Measurement of Tablet-to-Tablet Coating Variability. *J. Pharm. Biomed. Anal.* **2005**, 38 (2), 270–274.
- (252) Liu, Y.; Liu, Y.; Chen, Y.; Zhang, Y.; Shi, T.; Wang, J.; Hong, Y.; Fei, T.; Zhang, Y. The Influence of Spectral Pretreatment on the Selection of Representative Calibration Samples for Soil Organic Matter Estimation Using Vis-NIR Reflectance Spectroscopy. *Remote Sens.* **2019**, 11 (4), 450.
- (253) Lu, H.; Xu, H.; Ying, Y.; Fu, X.; Yu, H.; Tian, H. Application Fourier Transform near Infrared Spectrometer in Rapid Estimation of Soluble Solids Content of Intact Citrus Fruits. *J. Zhejiang Univ. Sci. B* **2006**, 7 (10), 794–799.
- (254) Herrera, J.; Guesalaga, A.; Agosin, E. Shortwave near Infrared Spectroscopy for Non-Destructive Determination of Maturity of Wine Grapes. *Meas. Sci. Technol.* **2003**, 14 (5), 689–697.
- (255) Arndt, M.; Rurik, M.; Drees, A.; Bigdowski, K.; Kohlbacher, O.; Fischer, M. Comparison of Different Sample Preparation Techniques for NIR Screening and Their Influence on the Geographical Origin Determination of Almonds (*Prunus Dulcis* MILL.). *Food Control* **2020**, 115, 107302.
- (256) Dhanoa, M. S.; Lister, S. J.; Sanderson, R.; Barnes, R. J. The Link between Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) Transformations of NIR Spectra. *J. Infrared Spectrosc.* **1994**, 2 (1), 43–47.
- (257) Cevallos-Cevallos, J. M.; Reyes-De-Corcuera, J. I.; Etxeberria, E.; Danyluk, M. D.; Rodrick, G. E. Metabolomic Analysis in Food Science: A Review. *Trends Food Sci. Technol.* **2009**, 20 (11–12), 557–566.

- (258) Risvik, H. Principal Component Analysis (PCA) & NIPALS algorithm. http://neuroinformatics.be/wp-content/uploads/2016/01/pca_nipals.pdf, abgerufen am 21.08.2021.
- (259) Jolliffe, I. T.; Cadima, J. Principal Component Analysis: A Review and Recent Developments. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **2016**, 374 (2065), 20150202.
- (260) Liland, K. H. Multivariate Methods in Metabolomics – from Pre-Processing to Dimension Reduction and Statistical Analysis. *TrAC Trends Anal. Chem.* **2011**, 30 (6), 827–841.
- (261) Lee, L. C.; Liong, C.-Y.; Jemain, A. A. Partial Least Squares-Discriminant Analysis (PLS-DA) for Classification of High-Dimensional (HD) Data: A Review of Contemporary Practice Strategies and Knowledge Gaps. *The Analyst* **2018**, 143 (15), 3526–3539.
- (262) Geladi, P.; Kowalski, B. R. Partial Least-Squares Regression: A Tutorial. *Anal. Chim. Acta* **1986**, 185, 1–17.
- (263) Trainor, P.; DeFilippis, A.; Rai, S. Evaluation of Classifier Performance for Multiclass Phenotype Discrimination in Untargeted Metabolomics. *Metabolites* **2017**, 7 (2), 30.
- (264) Westerhuis, J. A.; Hoefsloot, H. C. J.; Smit, S.; Vis, D. J.; Smilde, A. K.; van Velzen, E. J. J.; van Duijnhoven, J. P. M.; van Dorsten, F. A. Assessment of PLS-DA Cross Validation. *Metabolomics* **2008**, 4 (1), 81–89.
- (265) Bevilacqua, M.; Bro, R. Can We Trust Score Plots? *Metabolites* **2020**, 10 (7), 278.
- (266) Rodríguez-Pérez, R.; Fernández, L.; Marco, S. Overoptimism in Cross-Validation When Using Partial Least Squares-Discriminant Analysis for Omics Data: A Systematic Study. *Anal. Bioanal. Chem.* **2018**, 410 (23), 5981–5992.
- (267) Arndt, M.; Drees, A.; Ahlers, C.; Fischer, M. Determination of the Geographical Origin of Walnuts (*Juglans Regia* L.) Using Near-Infrared Spectroscopy and Chemometrics. *Foods* **2020**, 9 (12), 1860.
- (268) Tharwat, A.; Gaber, T.; Ibrahim, A.; Hassanien, A. E. Linear Discriminant Analysis: A Detailed Tutorial. *AI Commun.* **2017**, 30 (2), 169–190.
- (269) Murty, M. N.; Raghava, R. Kernel-Based SVM. In *Support Vector Machines and Perceptrons*; Springer International Publishing: Cham, Schweiz, **2016**; S. 57–67.
- (270) Hsu, C.-W.; Chang, C.-C.; Lin, C.-J. *A Practical Guide to Support Vector Classification*; Department of Computer Science National Taiwan University: Taipei, Taiwan, **2003**. <https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>, abgerufen am 10.10.2021.
- (271) Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, 20 (3), 273–297.
- (272) Patle, A.; Chouhan, D. S. SVM Kernel Functions for Classification. In *2013 International Conference on Advances in Technology and Engineering (ICATE)*; IEEE: Mumbai, Indien, **2013**; S. 1–9.
- (273) Yu, H.; Kim, S. SVM Tutorial — Classification, Regression and Ranking. In *Handbook of Natural Computing*; Rozenberg, G., Bäck, T., Kok, J. N., (Hrsg.); Springer: Berlin, Heidelberg, **2012**; S. 479–506.
- (274) Xu, Y.; Zomer, S.; Brereton, R. G. Support Vector Machines: A Recent Method for Classification in Chemometrics. *Crit. Rev. Anal. Chem.* **2006**, 36 (3–4), 177–188.
- (275) Jakkula, V. *Tutorial on Support Vector Machine (SVM)*; School of EECS, Washington State University, **2006**; S. 13. <https://course.ccs.neu.edu/cs5100f11/resources/jakkula.pdf>, abgerufen am 23.08.2021.
- (276) Lameski, P.; Zdravevski, E.; Mingov, R.; Kulakov, A. SVM Parameter Tuning with Grid Search and Its Impact on Reduction of Model Over-Fitting. In *Rough Sets, Fuzzy Sets, Data Mining, and*

- Granular Computing*; Yao, Y., Hu, Q., Yu, H., Grzymala-Busse, J. W., (Hrsg.); Springer International Publishing: Cham, Schweiz, **2015**; Vol. 9437, S. 464–474.
- (277) Chih-Wei Hsu; Chih-Jen Lin. A Comparison of Methods for Multiclass Support Vector Machines. *IEEE Trans. Neural Netw.* **2002**, 13 (2), 415–425.
- (278) Duan, K.-B.; Keerthi, S. S. Which Is the Best Multiclass SVM Method? An Empirical Study. In *Multiple Classifier Systems*; Oza, N. C., Polikar, R., Kittler, J., Roli, F., (Hrsg.); Springer: Berlin, Heidelberg, **2005**; Vol. 3541, S. 278–285.
- (279) Weston, J.; Watkins, C. Support Vector Machines for Multi-Class Pattern Recognition. In *Proceedings of the 7th European symposium on artificial neural networks (ESANN-99)*; Brügge, Belgien, **1999**; S. 219–224.
- (280) Karatzoglou, A.; Smola, A.; Hornik, K.; Zeileis, A. Kernlab - An S4 Package for Kernel Methods in R. *J. Stat. Softw.* **2004**, 11 (9).
- (281) Breiman, L. Random Forests. *Mach. Learn.* **2001**, 45 (1), 5–32.
- (282) Cutler, A.; Cutler, D. R.; Stevens, J. R. Random Forests. In *Ensemble Machine Learning*; Zhang, C., Ma, Y., (Hrsg.); Springer US: Boston, MA, USA, **2012**; S. 157–175.
- (283) Cutler, A. Remembering Leo Breiman. *Ann. Appl. Stat.* **2010**, 4 (4).
- (284) Probst, P.; Wright, M. N.; Boulesteix, A. Hyperparameters and Tuning Strategies for Random Forest. *WIREs Data Min. Knowl. Discov.* **2019**, 9 (3).
- (285) Díaz-Uriarte, R.; Alvarez de Andrés, S. Gene Selection and Classification of Microarray Data Using Random Forest. *BMC Bioinformatics* **2006**, 7 (1), 3.
- (286) Liaw, A.; Wiener, M. Classification and Regression by RandomForest. *R News* **2002**, 2/3, 18–22.
- (287) Fawagreh, K.; Gaber, M. M.; Elyan, E. Random Forests: From Early Developments to Recent Advancements. *Syst. Sci. Control Eng.* **2014**, 2 (1), 602–609.
- (288) Ali, J.; Khan, R.; Ahmad, N.; Maqsood, I. Random Forests and Decision Trees. *Int. J. Comput. Sci. Issues* **2012**, 9 (5), 7.
- (289) Krstajic, D.; Buturovic, L. J.; Leahy, D. E.; Thomas, S. Cross-Validation Pitfalls When Selecting and Assessing Regression and Classification Models. *J. Cheminformatics* **2014**, 6 (1), 10.
- (290) Steyerberg, E. W.; Bleeker, S. E.; Moll, H. A.; Grobbee, D. E.; Moons, K. G. M. Internal and External Validation of Predictive Models: A Simulation Study of Bias and Precision in Small Samples. *J. Clin. Epidemiol.* **2003**, 56 (5), 441–447.
- (291) Ballabio, D.; Consonni, V. Classification Tools in Chemistry. Part 1: Linear Models. PLS-DA. *Anal. Methods* **2013**, 5 (16), 3790.
- (292) Yang, K.; Wang, H.; Dai, G.; Hu, S.; Zhang, Y.; Xu, J. Determining the Repeat Number of Cross-Validation. In *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)*; IEEE: Shanghai, China, **2011**; S. 1706–1710.
- (293) Wong, T.-T.; Yeh, P.-Y. Reliable Accuracy Estimates from k -Fold Cross Validation. *IEEE Trans. Knowl. Data Eng.* **2020**, 32 (8), 1586–1594.
- (294) Vanwinckelen, G.; Blockeel, H. On Estimating Model Accuracy with Repeated Cross-Validation. In *BeneLearn 2012: Proceedings of the 21st Belgian-Dutch Conference on Machine Learning*; Ghent, Belgien, **2012**; S. 39–44.
- (295) Khalid, T.; Aggio, R.; White, P.; De Lacy Costello, B.; Persad, R.; Al-Kateb, H.; Jones, P.; Probert, C. S.; Ratcliffe, N. Urinary Volatile Organic Compounds for the Detection of Prostate Cancer. *PLOS ONE* **2015**, 10 (11), e0143283.

- (296) Kasakin, M. F.; Rogachev, A. D.; Predtechenskaya, E. V.; Zaigraev, V. J.; Koval, V. V.; Pokrovsky, A. G. Targeted Metabolomics Approach for Identification of Relapsing–Remitting Multiple Sclerosis Markers and Evaluation of Diagnostic Models. *MedChemComm* **2019**, *10* (10), 1803–1809.
- (297) Wang, Q.-Q.; Huang, H.-Y.; Wang, Y.-Z. Geographical Authentication of Macropyria Cocos by a Data Fusion Method Combining Ultra-Fast Liquid Chromatography and Fourier Transform Infrared Spectroscopy. *Molecules* **2019**, *24* (7), 1320.
- (298) Callao, M. P.; Ruisánchez, I. An Overview of Multivariate Qualitative Methods for Food Fraud Detection. *Food Control* **2018**, *86*, 283–293.
- (299) Pei, Y.-F.; Zuo, Z.-T.; Zhang, Q.-Z.; Wang, Y.-Z. Data Fusion of Fourier Transform Mid-Infrared (MIR) and Near-Infrared (NIR) Spectroscopies to Identify Geographical Origin of Wild *Paris Polyphylla* Var. *Yunnanensis*. *Molecules* **2019**, *24* (14), 2559.
- (300) Márquez, C.; López, M. I.; Ruisánchez, I.; Callao, M. P. FT-Raman and NIR Spectroscopy Data Fusion Strategy for Multivariate Qualitative Analysis of Food Fraud. *Talanta* **2016**, *161*, 80–86.
- (301) Delwiche, S.; Miskelly, D. Analysis of Grain Quality at Receival. In *Cereal Grains*; Elsevier: Amsterdam, Niederlande, **2017**; S. 513–570.
- (302) Pandey, P.; Mishra, G.; Mishra, H. N. Development of a Non-Destructive Method for Wheat Physico-Chemical Analysis by Chemometric Comparison of Discrete Light Based near Infrared and Fourier Transform near Infrared Spectroscopy. *J. Food Meas. Charact.* **2018**, *12* (4), 2535–2544.
- (303) Shenk, J. S.; Workman, J. J.; Westerhaus, M. O. Application of NIR Spectroscopy to Agricultural Products. In *Handbook of Near-Infrared Analysis*; Burns, D. A., Ciurczak, E. W., (Hrsg.); CRC Press: Boca Raton, FL, USA, **2007**.
- (304) DellaPenna, D.; Last, R. L. Genome-Enabled Approaches Shed New Light on Plant Metabolism. *Science* **2008**, *320* (5875), 479–481.
- (305) Trethewey, R. N. Metabolite Profiling as an Aid to Metabolic Engineering in Plants. *Curr. Opin. Plant Biol.* **2004**, *7* (2), 196–201.
- (306) Creydt, M.; Arndt, M.; Hudzik, D.; Fischer, M. Plant Metabolomics: Evaluation of Different Extraction Parameters for Nontargeted UPLC-ESI-QTOF-Mass Spectrometry at the Example of White *Asparagus Officinalis*. *J. Agric. Food Chem.* **2018**, *66* (48), 12876–12887.
- (307) Funasaki, M.; Menezes, I. S.; Barroso, H. dos S.; Zanotto, S. P.; Carioca, C. R. F. Tocopherol Profile of Brazil Nut Oil from Different Geographic Areas of the Amazon Region. *Acta Amaz.* **2013**, *43* (4), 505–509.
- (308) González, A. G.; Pablos, F.; Martín, M. J.; León-Camacho, M.; Valdenegro, M. S. HPLC Analysis of Tocopherols and Triglycerides in Coffee and Their Use as Authentication Parameters. *Food Chem.* **2001**, *73* (1), 93–101.
- (309) Chen, H.; Angiuli, M.; Ferrari, C.; Tombari, E.; Salvetti, G.; Bramanti, E. Tocopherol Speciation as First Screening for the Assessment of Extra Virgin Olive Oil Quality by Reversed-Phase High-Performance Liquid Chromatography/Fluorescence Detector. *Food Chem.* **2011**, *125* (4), 1423–1429.
- (310) Gurfinger, T.; Letan, A. Detection of Adulteration of Almond Oil with Apricot Oil through Determination of Tocopherols. *J. Agric. Food Chem.* **1973**, *21* (6), 1120–1123.
- (311) Borràs, E.; Ferré, J.; Boqué, R.; Mestres, M.; Aceña, L.; Busto, O. Data Fusion Methodologies for Food and Beverage Authentication and Quality Assessment – A Review. *Anal. Chim. Acta* **2015**, *891*, 1–14.

- (312) Biancolillo, A.; Boqué, R.; Cocchi, M.; Marini, F. Data Fusion Strategies in Food Analysis. In *Data Handling in Science and Technology*; Elsevier: Amsterdam, Niederlande, **2019**; Vol. 31, S. 271–310.
- (313) Smolinska, A.; Engel, J.; Szymanska, E.; Buydens, L.; Blanchet, L. General Framing of Low-, Mid-, and High-Level Data Fusion With Examples in the Life Sciences. In *Data Handling in Science and Technology*; Elsevier: Amsterdam, Niederlande, **2019**; Vol. 31, S. 51–79.
- (314) Azcarate, S. M.; Ríos-Reina, R.; Amigo, J. M.; Goicoechea, H. C. Data Handling in Data Fusion: Methodologies and Applications. *TrAC Trends Anal. Chem.* **2021**, *143*, 116355.
- (315) Medina, J.; van der Velpen, V.; Teav, T.; Guitton, Y.; Gallart-Ayala, H.; Ivanisevic, J. Single-Step Extraction Coupled with Targeted HILIC-MS/MS Approach for Comprehensive Analysis of Human Plasma Lipidome and Polar Metabolome. *Metabolites* **2020**, *10* (12), 495.
- (316) Fei, F.; Bowdich, D. M. E.; McCarry, B. E. Comprehensive and Simultaneous Coverage of Lipid and Polar Metabolites for Endogenous Cellular Metabolomics Using HILIC-TOF-MS. *Anal. Bioanal. Chem.* **2014**, *406* (15), 3723–3733.
- (317) Villaret-Cazadamont, J.; Poupin, N.; Tournadre, A.; Batut, A.; Gales, L.; Zalko, D.; Cabaton, N. J.; Bellvert, F.; Bertrand-Michel, J. An Optimized Dual Extraction Method for the Simultaneous and Accurate Analysis of Polar Metabolites and Lipids Carried out on Single Biological Samples. *Metabolites* **2020**, *10* (9), 338.
- (318) Cajka, T.; Fiehn, O. Comprehensive Analysis of Lipids in Biological Systems by Liquid Chromatography-Mass Spectrometry. *TrAC Trends Anal. Chem.* **2014**, *61*, 192–206.
- (319) Folch, J.; Lees, M.; Sloane Stanley, G. H. A Simple Method for the Isolation and Purification of Total Lipides from Animal Tissues. *J. Biol. Chem.* **1957**, *226* (1), 497–509.
- (320) Bligh, E. G.; Dyer, W. J. A Rapid Method of Total Lipid Extraction and Purification. *Can. J. Biochem. Physiol.* **1959**, *37* (8), 911–917.
- (321) Löfgren, L.; Ståhlman, M.; Forsberg, G.-B.; Saarinen, S.; Nilsson, R.; Hansson, G. I. The BUMe Method: A Novel Automated Chloroform-Free 96-Well Total Lipid Extraction Method for Blood Plasma. *J. Lipid Res.* **2012**, *53* (8), 1690–1700.
- (322) Armirotti, A.; Basit, A.; Realini, N.; Caltagirone, C.; Bossù, P.; Spalletta, G.; Piomelli, D. Sample Preparation and Orthogonal Chromatography for Broad Polarity Range Plasma Metabolomics: Application to Human Subjects with Neurodegenerative Dementia. *Anal. Biochem.* **2014**, *455*, 48–54.
- (323) Dettmer, K.; Nürnberger, N.; Kaspar, H.; Gruber, M. A.; Almstetter, M. F.; Oefner, P. J. Metabolite Extraction from Adherently Growing Mammalian Cells for Metabolomics Studies: Optimization of Harvesting and Extraction Protocols. *Anal. Bioanal. Chem.* **2011**, *399* (3), 1127–1139.
- (324) Martin, A. C.; Pawlus, A. D.; Jewett, E. M.; Wyse, D. L.; Angerhofer, C. K.; Hegeman, A. D. Evaluating Solvent Extraction Systems Using Metabolomics Approaches. *RSC Adv* **2014**, *4* (50), 26325–26334.
- (325) Oh, J.; Yoon, D.-H.; Han, J.-G.; Choi, H.-K.; Sung, G.-H. ¹H NMR Based Metabolite Profiling for Optimizing the Ethanol Extraction of *Wolfiporia cocos*. *Saudi J. Biol. Sci.* **2018**, *25* (6), 1128–1134.
- (326) Hara, A.; Radin, N. S. Lipid Extraction of Tissues with a Low-Toxicity Solvent. *Anal. Biochem.* **1978**, *90* (1), 420–426.
- (327) Matyash, V.; Liebisch, G.; Kurzchalia, T. V.; Shevchenko, A.; Schwudke, D. Lipid Extraction by Methyl-Tert-Butyl Ether for High-Throughput Lipidomics. *J. Lipid Res.* **2008**, *49* (5), 1137–1146.

- (328) Cequier-Sánchez, E.; Rodríguez, C.; Ravelo, Á. G.; Zárate, R. Dichloromethane as a Solvent for Lipid Extraction and Assessment of Lipid Classes and Fatty Acids from Samples of Different Natures. *J. Agric. Food Chem.* **2008**, *56* (12), 4297–4303.
- (329) Gil, A.; Zhang, W.; Wolters, J. C.; Permentier, H.; Boer, T.; Horvatovich, P.; Heiner-Fokkema, M. R.; Reijngoud, D.-J.; Bischoff, R. One- vs Two-Phase Extraction: Re-Evaluation of Sample Preparation Procedures for Untargeted Lipidomics in Plasma Samples. *Anal. Bioanal. Chem.* **2018**, *410* (23), 5859–5870.
- (330) Ghazani, S. M.; Marangoni, A. G. Healthy Fats and Oils. In *Encyclopedia of Food Grains*; Elsevier: Amsterdam, Niederlande, **2016**; S. 257–267.
- (331) Hilbeck, A.; Lebrecht, T.; Vogel, R.; Heinemann, J. A.; Binimelis, R. Farmer's Choice of Seeds in Four EU Countries under Different Levels of GM Crop Adoption. *Environ. Sci. Eur.* **2013**, *25* (1), 12.
- (332) Frank, T.; Röhligh, R. M.; Davies, H. V.; Barros, E.; Engel, K.-H. Metabolite Profiling of Maize Kernels—Genetic Modification versus Environmental Influence. *J. Agric. Food Chem.* **2012**, *60* (12), 3005–3012.
- (333) Lever, J.; Krzywinski, M.; Altman, N. Principal Component Analysis. *Nat. Methods* **2017**, *14* (7), 641–642.
- (334) Cozzolino, D.; Power, A.; Chapman, J. Interpreting and Reporting Principal Component Analysis in Food Science Analysis and Beyond. *Food Anal. Methods* **2019**, *12* (11), 2469–2473.
- (335) Maraschin, M.; Somensi-Zeggio, A.; Oliveira, S. K.; Kuhnen, S.; Tomazzoli, M. M.; Zeri, A. C. M.; Carreira, R.; Rocha, M. A Machine Learning and Chemometrics Assisted Interpretation of Spectroscopic Data – A NMR-Based Metabolomics Platform for the Assessment of Brazilian Propolis. In *Pattern Recognition in Bioinformatics*; Shibuya, T., Kashima, H., Sese, J., Ahmad, S., (Hrsg.); Springer: Berlin, Heidelberg, **2012**; Vol. 7632, S. 129–140.
- (336) Jung, Y.; Lee, J.; Kwon, J.; Lee, K.-S.; Ryu, D. H.; Hwang, G.-S. Discrimination of the Geographical Origin of Beef by ¹H NMR-Based Metabolomics. *J. Agric. Food Chem.* **2010**, *58* (19), 10458–10466.
- (337) Wishart, D. S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M.-A.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeroncic, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, David. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; MacInnis, G. D.; Weljie, A. M.; Dowlatbadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Querengesser, L. HMDB: The Human Metabolome Database. *Nucleic Acids Res.* **2007**, *35* (Database), D521–D526.
- (338) Sud, M.; Fahy, E.; Cotter, D.; Brown, A.; Dennis, E. A.; Glass, C. K.; Merrill, A. H.; Murphy, R. C.; Raetz, C. R. H.; Russell, D. W.; Subramaniam, S. LMSD: LIPID MAPS Structure Database. *Nucleic Acids Res.* **2007**, *35* (Database), D527–D532.
- (339) Rodriguez, J. D.; Perez, A.; Lozano, J. A. Sensitivity Analysis of K-Fold Cross Validation in Prediction Error Estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32* (3), 569–575.
- (340) Han, X. Fragmentation Patterns of Glycerolipids. In *Lipidomics: comprehensive mass spectrometry of lipids*; John Wiley & Sons: Hoboken, NJ, USA, **2016**.
- (341) Pi, J.; Wu, X.; Feng, Y. Fragmentation Patterns of Five Types of Phospholipids by Ultra-High-Performance Liquid Chromatography Electrospray Ionization Quadrupole Time-of-Flight Tandem Mass Spectrometry. *Anal. Methods* **2016**, *8* (6), 1319–1332.

- (342) Schütz, D.; Achten, E.; Creydt, M.; Riedl, J.; Fischer, M. Non-Targeted LC-MS Metabolomics Approach towards an Authentication of the Geographical Origin of Grain Maize (*Zea Mays* L.) Samples. *Foods* **2021**, *10* (9), 2160.
- (343) Kalo, P. J.; Ollilainen, V.; Rocha, J. M.; Malcata, F. X. Identification of Molecular Species of Simple Lipids by Normal Phase Liquid Chromatography–Positive Electrospray Tandem Mass Spectrometry, and Application of Developed Methods in Comprehensive Analysis of Low Erucic Acid Rapeseed Oil Lipids. *Int. J. Mass Spectrom.* **2006**, *254* (1–2), 106–121.
- (344) Bates, P. D.; Stymne, S.; Ohlrogge, J. Biochemical Pathways in Seed Oil Synthesis. *Curr. Opin. Plant Biol.* **2013**, *16* (3), 358–364.
- (345) Murphy, D. J. Plant Storage Lipids. In *eLS*; John Wiley & Sons Ltd, (Hrsg.), **2016**; S. 1–7.
- (346) Gu, Y.; He, L.; Zhao, C.; Wang, F.; Yan, B.; Gao, Y.; Li, Z.; Yang, K.; Xu, J. Biochemical and Transcriptional Regulation of Membrane Lipid Metabolism in Maize Leaves under Low Temperature. *Front. Plant Sci.* **2017**, *8*, 2053.
- (347) Cowan, A. K. Phospholipids as Plant Growth Regulators. *Plant Growth Regul.* **2006**, *48* (2), 97–109.
- (348) Upchurch, R. G. Fatty Acid Unsaturation, Mobilization, and Regulation in the Response of Plants to Stress. *Biotechnol. Lett.* **2008**, *30* (6), 967–977.
- (349) Oldiges, M.; Lütz, S.; Pflug, S.; Schroer, K.; Stein, N.; Wiendahl, C. Metabolomics: Current State and Evolving Methodologies and Tools. *Appl. Microbiol. Biotechnol.* **2007**, *76* (3), 495–511.
- (350) Lim, D. K.; Mo, C.; Lee, J. H.; Long, N. P.; Dong, Z.; Li, J.; Lim, J.; Kwon, S. W. The Integration of Multi-Platform MS-Based Metabolomics and Multivariate Analysis for the Geographical Origin Discrimination of *Oryza Sativa* L. *J. Food Drug Anal.* **2018**, *26* (2), 769–777.
- (351) Rubert, J.; Hurkova, K.; Stranska, M.; Hajslova, J. Untargeted Metabolomics Reveals Links between Tiger Nut (*Cyperus Esculentus* L.) and Its Geographical Origin by Metabolome Changes Associated with Membrane Lipids. *Food Addit. Contam. Part A* **2018**, *35* (10), 1861–1869.
- (352) Murphy, R. C.; James, P. F.; McAnoy, A. M.; Krank, J.; Duchoslav, E.; Barkley, R. M. Detection of the Abundance of Diacylglycerol and Triacylglycerol Molecular Species in Cells Using Neutral Loss Mass Spectrometry. *Anal. Biochem.* **2007**, *366* (1), 59–70.
- (353) Schmauder, F.; Schütz, D.; Creydt, M.; Riedl, J.; Fischer, M. Food Authentication Goes Green: Determination of the Geographical Origin of Grain Maize (*Zea Mays* L.) Using Direct Analysis in Real Time Mass Spectrometry (DART-MS). *Food Control* **2024**, *163*, 110489.
- (354) Van der Meeren, P.; Vanderdeelen, J. Chapter 3 - Preparative High Performance Liquid Chromatography of Lipids. In *Advances in Lipid Methodology*; Christie, W. W., (Hrsg.); Woodhead Publishing: Cambridge, UK, **2013**; S. 83–118.
- (355) Matsuda, K.; Matsuda, S.; Saito, M.; Ito, Y. Separation of Phospholipids and Glycolipids Using Analytical Toroidal-Coil Countercurrent Chromatography. I. Separation of Human Brain Lipids. *J. Liq. Chromatogr. Relat. Technol.* **2002**, *25* (8), 1255–1269.
- (356) Fauland, A.; Köfeler, H.; Trötz Müller, M.; Knopf, A.; Hartler, J.; Eberl, A.; Chitraju, C.; Lankmayr, E.; Spener, F. A Comprehensive Method for Lipid Profiling by Liquid Chromatography-Ion Cyclotron Resonance Mass Spectrometry. *J. Lipid Res.* **2011**, *52* (12), 2314–2322.
- (357) Goffman, F. D.; Böhme, T. Relationship between Fatty Acid Profile and Vitamin E Content in Maize Hybrids (*Zea Mays* L.). *J. Agric. Food Chem.* **2001**, *49* (10), 4990–4994.
- (358) BVL L 00.00-62:2015-06 Untersuchung von Lebensmitteln - Bestimmung von Vitamin E (α -, β -, γ - und δ -Tocopherol) in Lebensmitteln mit Hochleistungs-Flüssigchromatographie (Übernahme der gleichnamigen Norm DIN EN 12822, Ausgabe August 2014).

- (359) Hu, W.; Wells, J. H.; Shin, T.-S.; Godber, J. S. Comparison of Isopropanol and Hexane for Extraction of Vitamin E and Oryzanols from Stabilized Rice Bran. *J. Am. Oil Chem. Soc.* **1996**, *73* (12), 1653–1656.
- (360) Moreau, R. A.; Powell, M. J.; Singh, V. Pressurized Liquid Extraction of Polar and Nonpolar Lipids in Corn and Oats with Hexane, Methylene Chloride, Isopropanol, and Ethanol. *J. Am. Oil Chem. Soc.* **2003**, *80* (11), 1063–1067.
- (361) Nagy, K.; Courtet-Compondu, M.-C.; Holst, B.; Kussmann, M. Comprehensive Analysis of Vitamin E Constituents in Human Plasma by Liquid Chromatography–Mass Spectrometry. *Anal. Chem.* **2007**, *79* (18), 7087–7096.
- (362) Grams, G. W.; Blessin, C. W.; Inglett, G. E. Distribution of Tocopherols within the Corn Kernel. *J. Am. Oil Chem. Soc.* **1970**, *47* (9), 337–339.
- (363) Lu, D.; Yang, Y.; Li, Y.; Sun, C. Analysis of Tocopherols and Tocotrienols in Pharmaceuticals and Foods: A Critical Review. *Curr. Pharm. Anal.* **2014**, *11* (1), 66–78.
- (364) Murkovic, M.; Hillebrand, A.; Winkler, J.; Pfannhauser, W. Variability of Vitamin E Content in Pumpkin Seeds (*Cucurbita Pepo* L). *Z. Für Lebensm.-Unters. -Forsch.* **1996**, *202* (4), 275–278.
- (365) Galliher, H. L.; Alexander, D. E.; Weber, E. J. Genetic Variability of Alpha-Tocopherol and Gamma-Tocopherol in Corn Embryos. *Crop Sci.* **1985**, *25* (3), 547–549.
- (366) Agelet, L. E.; Hurburgh, C. R. A Tutorial on Near Infrared Spectroscopy and Its Calibration. *Crit. Rev. Anal. Chem.* **2010**, *40* (4), 246–260.
- (367) Büning-Pfaue, H. Analysis of Water in Food by near Infrared Spectroscopy. *Food Chem.* **2003**, *82* (1), 107–115.
- (368) Schütz, D.; Riedl, J.; Achten, E.; Fischer, M. Fourier-Transform near-Infrared Spectroscopy as a Fast Screening Tool for the Verification of the Geographical Origin of Grain Maize (*Zea Mays* L.). *Food Control* **2022**, *136*, 108892.
- (369) Cai, Y.; Lin, S. L. Support Vector Machines for Predicting RRNA-, RNA-, and DNA-Binding Proteins from Amino Acid Sequence. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2003**, *1648* (1–2), 127–133.
- (370) Burges, C. J. C. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Min. Knowl. Discov.* **1998**, *2* (2), 121–167.
- (371) Bhavsar, H.; Panchal, M. H. A Review on Support Vector Machine for Data Classification. *Int. J. Adv. Res. Comput. Eng. Technol.* **2012**, *1* (10), 5.
- (372) Tharwat, A. Parameter Investigation of Support Vector Machine Classifier with Kernel Functions. *Knowl. Inf. Syst.* **2019**, *61* (3), 1269–1302.
- (373) Arena, E.; Campisi, S.; Fallico, B.; Maccarone, E. Distribution of Fatty Acids and Phytosterols as a Criterion to Discriminate Geographic Origin of Pistachio Seeds. *Food Chem.* **2007**, *104* (1), 403–408.
- (374) Zhang, H.; Jiang, H.; Liu, G.; Mei, C.; Huang, Y. Identification of Radix Puerariae Starch from Different Geographical Origins by FT-NIR Spectroscopy. *Int. J. Food Prop.* **2017**, 1–11.
- (375) Achten, E.; Schütz, D.; Fischer, M.; Faulh-Hassek, C.; Riedl, J.; Horn, B. Classification of Grain Maize (*Zea Mays* L.) from Different Geographical Origins with FTIR Spectroscopy—a Suitable Analytical Tool for Feed Authentication? *Food Anal. Methods* **2019**, *12* (10), 2172–2184.
- (376) Fernández Pierna, J. A.; Vermeulen, P.; Amand, O.; Tossens, A.; Dardenne, P.; Baeten, V. NIR Hyperspectral Imaging Spectroscopy and Chemometrics for the Detection of Undesirable Substances in Food and Feed. *Chemom. Intell. Lab. Syst.* **2012**, *117*, 233–239.

- (377) De Girolamo, A.; Lippolis, V.; Nordquist, E.; Visconti, A. Rapid and Non-Invasive Analysis of Deoxynivalenol in Durum and Common Wheat by Fourier-Transform Near Infrared (FT-NIR) Spectroscopy. *Food Addit. Contam.* **2009**, *26* (6), 907–917.
- (378) United States Department of Agriculture Foreign Agricultural Service. *World Agricultural Supply and Demand Estimates*; **2024**.
<https://www.usda.gov/oce/commodity/wasde/wasde0524.pdf>, abgerufen am 28.06.2024.
- (379) euronews. *Argentinien: Rekord-Maisernte, aber zu spät*.
<https://de.euronews.com/2017/07/04/argentinien-rekord-maisernte-aber-zu-spat>, abgerufen am 24.09.2021.
- (380) Geurts van Kessel, W. S. M.; Tieman, M.; Demel, R. A. Purification of Phospholipids by Preparative High Pressure Liquid Chromatography. *Lipids* **1981**, *16* (1), 58–63.
- (381) Vetter, W.; Hammann, S.; Müller, M.; Englert, M.; Huang, Y. The Use of Countercurrent Chromatography in the Separation of Nonpolar Lipid Compounds. *J. Chromatogr. A* **2017**, *1501*, 51–60.
- (382) Hernández-Mesa, M.; Ropartz, D.; García-Campaña, A. M.; Rogniaux, H.; Dervilly-Pinel, G.; Le Bizec, B. Ion Mobility Spectrometry in Food Analysis: Principles, Current Applications and Future Trends. *Molecules* **2019**, *24* (15), 2706.
- (383) Fouque, K. J. D.; Ramirez, C. E.; Lewis, R. L.; Koelmel, J. P.; Garrett, T. J.; Yost, R. A.; Fernandez-Lima, F. Effective Liquid Chromatography–Trapped Ion Mobility Spectrometry–Mass Spectrometry Separation of Isomeric Lipid Species. *Anal. Chem.* **2019**, *91* (8), 5021–5027.
- (384) Groessl, M.; Graf, S.; Knochenmuss, R. High Resolution Ion Mobility-Mass Spectrometry for Separation and Identification of Isomeric Lipids. *The Analyst* **2015**, *140* (20), 6904–6911.
- (385) Lu, Y.; Guo, J.; Yu, J.; Guo, J.; Jia, X.; Liu, W.; Tian, P. Two-Dimensional Analysis of Phenolic Acids in Seedling Roots by High Performance Liquid Chromatography-Electrospray Ionization-Ion Mobility Spectrometry. *Anal. Methods* **2019**, *11* (5), 610–617.
- (386) Bachmann, R.; Shakiba, N.; Fischer, M.; Hackl, T. Assessment of Mixtures by Spectral Superposition. An Approach in the Field of Metabolomics. *J. Proteome Res.* **2019**, *18* (6), 2458–2466.
- (387) Schmitt, C.; Bastek, T.; Stelzer, A.; Schneider, T.; Fischer, M.; Hackl, T. Detection of Peanut Adulteration in Food Samples by Nuclear Magnetic Resonance Spectroscopy. *J. Agric. Food Chem.* **2020**, *68* (49), 14364–14373.
- (388) Horn, B.; Esslinger, S.; Fauhl-Hassek, C.; Riedl, J. ¹H NMR Spectroscopy, One-Class Classification and Outlier Diagnosis: A Powerful Combination for Adulteration Detection in Paprika Powder. *Food Control* **2021**, *128*, 108205.
- (389) Wellnitz, J.; Gluschke, M. *Leitlinie zur Methodvalidierung*; Umweltbundesamt, (Hrsg.); Berlin, **2005**; Vol. 01/05.
- (390) McPolin, O. *The Validation of Analytical Methods for Pharmaceutical Analysis*; Mourn training services: Warrenpoint, UK, **2009**.
- (391) R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Wien, Österreich, **2020**. <https://www.R-project.org/>, abgerufen am 16.07.2020.
- (392) Telgheder, U. *Praktikum Wasserchemie/Wasseranalytik im Bachelor-Studiengang Water Science; Chemie - Leitfaden zur Auswertung analytischer Ergebnisse*; Praktikumsskript; Universität Duisburg-Essen: Duisburg/Essen, **2013**; S. 21. <https://www.uni->

- due.de/imperia/md/content/water-science/ws1314/1351_2661_ws1314_leitfaden_auswertung.pdf, abgerufen am 28.09.2021.
- (393) Beleites, C.; Sergo, V. *hyperSpec: a package to handle hyperspectral data sets in R*. <https://github.com/cbeleites/hyperSpec>, abgerufen am 25.07.2020.
- (394) Stevens, A.; Ramirez-Lopez, L. *An introduction to the prospectr package*. <https://cran.r-project.org/web/packages/prospectr/vignettes/prospectr.html>, abgerufen am 25.06.2021.
- (395) Kuhn, M. Building Predictive Models in R Using the Caret Package. *J. Stat. Softw.* **2008**, *28* (5).
- (396) Mevik, B.-H.; Wehrens, R.; Liland, K. H. *pls: Partial Least Squares and Principal Component Regression*. <https://CRAN.R-project.org/package=pls>, abgerufen am 30.06.2021.
- (397) Eliasson, M.; Rannar, S.; Trygg, J. From Data Processing to Multivariate Validation - Essential Steps in Extracting Interpretable Information from Metabolomics Data. *Curr. Pharm. Biotechnol.* **2011**, *12* (7), 996–1004.
- (398) Alewijn, M.; van der Voet, H.; van Ruth, S. Validation of Multivariate Classification Methods Using Analytical Fingerprints – Concept and Case Study on Organic Feed for Laying Hens. *J. Food Compos. Anal.* **2016**, *51*, 15–23.
- (399) Shao, X.; Li, H.; Wang, N.; Zhang, Q. Comparison of Different Classification Methods for Analyzing Electronic Nose Data to Characterize Sesame Oils and Blends. *Sensors* **2015**, *15* (10), 26726–26742.
- (400) Gromski, P. S.; Correa, E.; Vaughan, A. A.; Wedge, D. C.; Turner, M. L.; Goodacre, R. A Comparison of Different Chemometrics Approaches for the Robust Classification of Electronic Nose Data. *Anal. Bioanal. Chem.* **2014**, *406* (29), 7581–7590.
- (401) Scornet, E. Tuning Parameters in Random Forests. *ESAIM Proc. Surv.* **2017**, *60*, 144–162.
- (402) Sigma Aldrich. . <https://www.sigmaaldrich.com/DE/en/search>, abgerufen am 02.10.2021.
- (403) Carl Roth GmbH + Co. KG. https://www.carlroth.com/de/de/chemikalien/c/web_folder_260523, abgerufen am 02.10.2021.
- (404) Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA). Das Global Harmonisierte System (GHS) in der EU. **2019**.
- (405) Institut für Arbeitsschutz der Deutschen Gesetzlichen Unfallversicherung (IFA). Liste der krebserzeugenden, keimzellmutagenen und reproduktionstoxischen Stoffe (KMR-Stoffe). August **2021**. <https://publikationen.dguv.de/widgets/pdf/download/article/3517>, abgerufen am 22.09.2021.

11 Abbildungsverzeichnis

Abb. 4.1: Verschiedene Teile der Maispflanze.	10
Abb. 4.2: Preise von Körnermais in verschiedenen Staaten von 1999 bis 2019.....	13
Abb. 4.3: Die <i>omics</i> -Kaskade.....	15
Abb. 4.4: Einflussfaktoren auf das Metabolom.	15
Abb. 4.5: Schematischer Aufbau eines FT-NIR-Spektrometers.	24
Abb. 4.7: Schematische Darstellung einer ESI-Ionenquelle und des Ionisationsprozesses.....	28
Abb. 4.8: Schematische Darstellung eines q-ToF-Massenspektrometers mit einer ESI-Ionenquelle.	31
Abb. 4.9: Beispiele der Klassengrenzen bei verschiedenen Klassifikationsalgorithmen.	38
Abb. 4.10: Schema der <i>k-fold</i> Kreuzvalidierung mit $k = 5$, nach Raschka.	41
Abb. 4.11: Varianten der Datenfusion.	42
Abb. 6.1: Optimierung der Extraktionsmethode für polare Metabolite.	48
Abb. 6.2: Optimierung der Extraktionsmethode für unpolare Metabolite.	50
Abb. 6.3: PCA-Scoresplots (nach <i>Pareto-Scaling</i>) der Körnermaisproben.....	53
Abb. 6.4: Heatmaps der Verteilung der Signifikanzen der Features in dem Sortenprobenset.	55
Abb. 6.5: Abstrahierte <i>Survey-Views</i> der herkunfts/sorten-bezogen beeinflussten Features je nach verwendeter LC-MS-Methode.	58
Abb. 6.6: PCA-Scoresplot der Metabolite je nach LC-MS-Methode.	61
Abb. 6.7: PCA-Scoresplot und RF-Klassifikationsmatrix erstellt mittels der <i>Consensus-Feature</i> Tabelle der im positiven Ionenmodus detektierten Metabolite (nach <i>Pareto-Scaling</i>).	64
Abb. 6.8: PCA-Scoresplot und RF-Klassifikationsmatrix unter Nutzung der 20 ausgewählten Markersubstanzen (nach <i>Pareto-Scaling</i>).	69
Abb. 6.9: Heatmaps der Peakflächen je nach Lagerungsbedingungen.	73
Abb. 6.10: PCA-Scoresplot nach <i>Pareto-Scaling</i> der Lagerungsproben.....	74
Abb. 6.11: Darstellung der herkunftsabhängigen Schwankungen der quantifizierten Marker in den Extrakten des Herkunftsprobensets.....	79
Abb. 6.12: PCA-Scoresplots der Herkunftspuben (nach <i>Pareto-Scaling</i>) der MS^1 - und MS^2 - Peakflächen derselben neun Markermetabolite im Vergleich.	80
Abb. 6.13: Vergleich der RF-Klassifikationsmatrizen der Proben des Herkunftsprobensets.....	80
Abb. 6.14: Darstellung der Tocopherol-Konzentrationen in Abhängigkeit von der geographischen Herkunft der Maisproben.	82
Abb. 6.15: Boxplot der erreichten Klassifikationsgenauigkeit (RF-Algorithmus, 10-mal wiederholte, 10-fache Kreuzvalidierung) bei Nutzung von je zwei Lipidmarkern.....	83
Abb. 6.16: FT-NIR-Spektren der QC-Probe II nach unterschiedlicher Probenvorbereitung.	85
Abb. 6.17: Darstellung der FT-NIR-Spektren des Sortenprobensets (<i>Preprocessing</i> mittels SNV)...	86
Abb. 6.18: Median der FT-NIR-Spektren der Proben des Herkunftsprobensets nach <i>Preprocessing</i> mittels SNV.....	88
Abb. 6.19: PCA-Scoresplot von SNV preprozessierten FT-NIR-Spektren der Proben des Herkunftsprobensets.....	89
Abb. 6.20: PLS-DA Loadingsvektoren der Trennung nach der geographischen Herkunft.	90
Abb. 6.21: SVM-Klassifikationsergebnisse bei kombinierter Nutzung von Spektrenbereichen, welche auf den Absorptionsbanden von Proteinen, Stärke und Lipiden basierten.	95

Abb. 6.22: PCA-Scoresplots der Körnermaisproben auf kontinentaler Ebene sowie auf Länderebene.....	96
Abb. 6.23: PCA-Scoresplots der mittels <i>Range-Scaling</i> preprozessierten und angeglichenen Analysedaten.....	100
Abb. 6.24: Detaillierte Klassifikationsmatrizen der RF-Klassifikation (100-fach wiederholte, 10-fache Kreuzvalidierung) nach <i>Range-Scaling</i>	101
Abb. 13.1: Anzahl der Features in der QC-Probe je Extraktionsmethode für polare Metabolite...	174
Abb. 13.2: Anzahl der Features in der QC-Probe je Extraktionsmethode für unpolare Metabolite.	174
Abb. 13.3: Scoresplots der PLS-DA (nach <i>Pareto-Scaling</i>) des Sortenprobensets bei Trennung nach Körnermaissorte.	175
Abb. 13.4: Loadingsplots der PCA der Metabolite unterteilt nach genutzter LC-MS-Methode.	176
Abb. 13.5: Histogramme der Intensitätenverteilung der detektierten Features der Metabolite. .	177
Abb. 13.6: <i>Survey-View</i> sowie <i>Base-Peak</i> -Chromatogramm (BPC) eines repräsentativen BLIGH & DYER-Extraktes der QC-Probe.	178
Abb. 13.7: RF-Klassifikationsmatrizen (100-mal wiederholt und 10-fach kreuzvalidiert) unter Verwendung der Peakflächen der jeweils Top 20 signifikanten Phospholipide, DGs und TGs.....	180
Abb. 13.8: RF-Klassifikationsergebnisse unter Verwendung der Peakflächen der Top 20 intensivsten, signifikanten Phospholipide, DGs und TGs.	181
Abb. 13.9: Boxplots der mittleren Peakflächen der 20 Markerfeatures in Abhängigkeit von der geographischen Herkunft.	181
Abb. 13.10: Kalibrationsgeraden der Standardsubstanzen.	184
Abb. 13.11: Kalibrationsgeraden der Verdünnungsreihe der neun quantifizierten Markermetabolite.	185
Abb. 13.12: Multivariate Datenanalyse der FT-NIR-Spektren des Sortenprobensets.	186
Abb. 13.13: SVM-Klassifikationsmatrizen einer 100-fach wiederholten 10-fachen Kreuzvalidierung der Proben des Sortenprobensets.	186
Abb. 13.14: Rohdaten der FT-NIR-Spektren des Herkunftsprobensets.	187
Abb. 13.15: Einfluss von unterschiedlichen spektralen <i>Preprocessings</i> auf die FT-NIR-Spektren des Herkunftsprobensets.....	187
Abb. 13.16: SVM-Klassifikationsmatrizen (100-mal wiederholte 10-fache Kreuzvalidierung) unter Nutzung unterschiedlichen Spektralbereiche.	188
Abb. 13.17: SVM-Klassifikationsmatrizen (100-mal wiederholte 10-fache Kreuzvalidierung) der FT-NIR-Spektren ohne Wasserbanden.	189
Abb. 13.18: SVM-Klassifikationsmatrizen der erweiterten Modelle (Futtermais und Stärkernmais; 100-mal wiederholte 10-fache Kreuzvalidierung).....	189
Abb. 13.19: PCA-Loadingsplot der fusionierten LC-MS-Daten und der FT-NIR-Daten nach <i>Preprocessing</i> und Angleichung des Wertebereichs mittels <i>Range-Scaling</i>	189

12 Tabellenverzeichnis

Tab. 4.1: Vergleich von Massenspektrometrie und NMR-Spektroskopie als Metabolomics-Plattformen.	18
Tab. 4.2: Vergleich verschiedener Masseanalysatoren.	27
Tab. 6.1: Evaluierte Extraktionsmethoden für die Detektion polarer Metabolite.	47
Tab. 6.2: Evaluierte Extraktionsmethoden für die Detektion unpolarer Metabolite.	49
Tab. 6.3: Mittlere prozentuale Anteile der signifikant beeinflussten Features des Sortenprobensets.	54
Tab. 6.4: RF-Klassifizierungsergebnisse des Sortenprobensets nach Herkunft bzw. Sorte je nach ausgewählter Subgruppe der detektierten Metabolite des Sortenprobensets.	57
Tab. 6.5: Anzahl detektierter Features je nach genutzter LC-MS-Methode.	60
Tab. 6.6: Erreichte RF-Klassifikationsgenauigkeiten mit den Analysedaten der vier LC-MS-Methoden des Körnermais-Herkunftsprobensets.	62
Tab. 6.7: Klassifikationsergebnisse von binären RF Modellen (ein Land vs. die restlichen Länder, 100-wiederholt, 10-fach kreuzvalidiert).	70
Tab. 6.8: Ermittelte Validierungskennzahlen für die quantitative Bestimmung von Markerlipiden zur Herkunftsbestimmung mittels Grund- und Matrixkalibrierung.	77
Tab. 6.9: SVM Klassifikationsgenauigkeiten in Abhängigkeit vom Spektrenpreprocessing bzw. der genutzten Kernel-Funktion.	91
Tab. 6.10: Auflistung der SVM-Modelle, welche auf unterschiedlichen Spektrenregionen beruhen.	93
Tab. 6.11: SVM-Klassifikationsgenauigkeiten (100-mal wiederholte 10-fache Kreuzvalidierung) incl. der Proben aus Frankreich bzw. Ungarn.	98
Tab. 9.1: Probenumfang des Probensets aus Landessortenversuchen.	108
Tab. 9.2: Probenumfang des Probensets zur Herkunftsbestimmung.	109
Tab. 9.3: Verwendete Mahlsequenz zur Vermahlung von Körnermaisproben in der Grindomix GM 300 Messermühle.	111
Tab. 9.4: Untersuchte Extraktionsmethoden bei der Optimierung der Analyse von polaren Metabolite von Körnermais mittels LC-ESI-MS.	112
Tab. 9.5: Untersuchte Extraktionsmethoden bei der Optimierung der Analyse von unpolaren Metabolite von Körnermais mittels LC-ESI-MS.	114
Tab. 9.6: Zeitlicher Verlauf des Gradienten der LC-MS-Methode für die Analyse von polaren Metaboliten.	116
Tab. 9.7: Einstellungen der ESI-Ionenquelle sowie weiterer Parameter des q-ToF-Massenspektrometers für die LC-ESI-MS-basierte Analyse von polaren Metaboliten.	117
Tab. 9.8: Zeitlicher Verlauf des Gradienten der LC-MS-Methode für die Analyse von unpolaren Metaboliten.	117
Tab. 9.9: Pipettierschema für die Erstellung der Kalibrationslösung zur Kalibrierung der m/z -Achse des Massenspektrometers.	118
Tab. 9.10: Zeitlicher Verlauf des Gradienten der LC-MRM-Methode für die quantitative Bestimmung von unpolaren Metaboliten.	119
Tab. 9.11: Auflistung der quantifizierten Metabolite sowie die entsprechenden MRM-Übergänge.	120

Tab. 9.12: Zeitlicher Verlauf des Gradienten der LC-MRM-Methode für die quantitative Bestimmung von α -Tocopherol und γ -Tocopherol.	121
Tab. 9.13: Verwendete Einstellungen im Programm DataAnalysis für die Detektion von Features mittels der Funktion „ <i>Find Molecular Features</i> “.....	123
Tab. 9.14: Einstellungen des Programms ProfileAnalysis zur Erstellung von vergleichenden Peaklisten.	124
Tab. 9.15: Einstellungen für FT-NIR-Analysen in der Messsoftware OPUS.	128
Tab. 9.16: Zusammenstellung der in dieser Arbeit durchgeführten Analysen.	128
Tab. 9.17: Evaluierte Algorithmen zur Auswahl eines optimalen <i>Preprocessing</i> -Verfahrens für FT-NIR-Spektren, sowie deren Funktionsaufrufe in R.	131
Tab. 13.1: Metadaten der Körnermaisproben zur Erstellung von Klassifikationsmodellen bezüglich der geographischen Herkunft.....	165
Tab. 13.2: Sortenbezeichnungen der Körnermaisproben aus Spanien zur Bestimmung der Robustheit der Bestimmung der geographischen Herkunft gegenüber Sorteneinflüssen.	166
Tab. 13.3: Metadaten der Körnermaisproben aus Landesortenversuchen zur Bestimmung des wechselseitigen Einflusses der Körnermaissorte und der geographischen Herkunft.....	166
Tab. 13.4: In dieser Arbeit verwendete Geräte.....	167
Tab. 13.5: In dieser Arbeit verwendete Verbrauchsmaterialien.....	168
Tab. 13.6: In dieser Arbeit verwendete Chemikalien.....	169
Tab. 13.7: In dieser Arbeit verwendete Software.	171
Tab. 13.8: Gefahrensymbole nach GHS, H- und P-Sätze sowie Hinweise zur Entsorgung, der in dieser Arbeit verwendeten Chemikalien.....	172
Tab. 13.9: In dieser Arbeit verwendete KMR-Stoffe der Kategorien GHS 1A und 1B.	173
Tab. 13.10: Bedeutende Markerfeatures in polaren Extrakten für die Herkunftsbestimmung von Körnermais.	177
Tab. 13.11: Top 20 der intensivsten, signifikanten (ANOVA-Test p -Wert ≤ 0.01) Peaks je Lipidklasse.	179
Tab. 13.12: Informationen zu den Markerfeatures für die Unterscheidung der geographischen Herkunft von Körnermaisproben.	182
Tab. 13.13: <i>Ranking</i> der 20 Markersubstanzen nach ihrer Variableneinfluss in binären RF-Klassifikationsmodellen.....	183
Tab. 13.14: MRM-Übergänge der 20 selektierten Markermetabolite zur Herkunftsbestimmung von Körnermais.	184
Tab. 13.15: Optimierte Parameter für die Klassifikation mittels SVM.	188
Tab. 13.16: <i>Ranking</i> der Marker nach ihrem Variableneinfluss in dem auf <i>low-level</i> Ebene fusioniertem RF-Klassifikationsmodell.	190

13 Anhang

13.1 Material

Die in dieser Arbeit dargestellten Ergebnisse beruhen auf insgesamt 173 Körnermaisproben aus verschiedenen Herkunftsländern. Deren genaue Herkunftsorte sowie weitere Metadaten sind in **Tab. 13.1** aufgelistet, die Sortenbezeichnungen der spanischen Proben sind in **Tab. 13.2** aufgeführt. Die wechselseitige Beeinflussung von Sorte und Herkunftsort wurde zudem anhand von 81 Körnermaisproben aus drei Landessortenversuchen verschiedener deutscher Bundesländer analysiert. Diese Proben sind in **Tab. 13.3** mit ihren Metadaten aufgeführt.

Tab. 13.1: Metadaten der Körnermaisproben zur Erstellung von Klassifikationsmodellen bezüglich der geographischen Herkunft.

Land	Region	Sub-Region	Anzahl Proben
Frankreich	Elsass		36
Peru	Lambayeque		
		Pátapo	3
		Tumán	2
		Pomalca	3
		Ferrenafe	2
		Mochumí	3
		Túcume	2
		Pacora	3
		Jayanca	2
		Mórrope	3
		Lambayeque	2
Spanien	LLeida	El Poal	20
Slowakei			
	- Levice		10
	- Trnava		2
	- Senec		1
	- unbekannt		1
Ukraine	unbekannt		19
Ungarn	Ostungarn		36
USA			
	- Alabama	Blount County, Madison County, Cullman County, Marshall County	4
	- Arkansas	Greene County, Chicot County	4
	- Georgia	Bulloch County, Jefferson County, Washington County	4
	- Louisiana	Morehouse County, Avoyelles County, Pointe Couppe County, East Carroll County	4
	- Mississippi	Yazoo County, Jefferson Davis County, Perry County, Simpson County	4
	- Texas	San Patricio County, Ellis County	3

Tab. 13.2: Sortenbezeichnungen der Körnermaisproben aus Spanien zur Bestimmung der Robustheit der Bestimmung der geographischen Herkunft gegenüber Sorteneinflüssen. Tabelle nach Schütz *et al.*³⁴².

68.K	Charleston	DKC6728	DKC6650	DKC5830
Ixabel	Kefieros	Keridos	Kontigos	LG30.600
LG30.681	MAS 64.P	MAS 66.C	MAS 75.A	Miloxan
Nystar	P1524	P1570	P1574	P1921

Tab. 13.3: Metadaten der Körnermaisproben aus Landesortenversuchen zur Bestimmung des wechselseitigen Einflusses der Körnermaissorte und der geographischen Herkunft.

Landessortenversuch	Versuchsort	Anzahl Proben /Sorten
Brandenburg	14532 Güterfelde	81 (9 Sorten à 3 technische Replikate)
Nordrhein-Westfalen	48249 Dülmen	81 (9 Sorten à 3 technische Replikate)
Bayern	86666 Burgheim - Straß	81 (9 Sorten à 3 technische Replikate)

Die in dieser Arbeit verwendeten Geräte sind in **Tab. 13.4** aufgeführt. Die ebenfalls verwendeten Verbrauchsmaterialien sind zudem in **Tab. 13.5** aufgelistet. Die verwendeten Chemikalien sind in **Tab. 13.6** aufgelistet.

Tab. 13.4: In dieser Arbeit verwendete Geräte.

Gerät	Modellbezeichnung	Hersteller
Tiefkühltruhe	-/-	Liebherr International AG, Bulle, Schweiz
Kühlschrank	MediLine LKUv 1610	Liebherr International AG, Bulle, Schweiz
Analysenwaage I	Sartorius Talent M-power	Sartorius Weighing Technology GmbH, Göttingen, Hamburg, Deutschland
Analysenwaage II	XA105 DualRange	Mettler Toledo, Greifensee, Schweiz
Kolbenhubpipetten	Eppendorf Research variabel	Eppendorf AG, Hamburg, Deutschland
Messermühle	Grindomix GM 300	Retsch Technology GmbH, Haan, Deutschland
Gefriertrocknungsanlage I	Beta 1-8 LDplus	Martin Christ Gefriertrocknungsanlagen GmbH, Osterode, Deutschland
Gefriertrocknungsanlage II	Alpha 1-4 LSC	Martin Christ Gefriertrocknungsanlagen GmbH, Osterode, Deutschland
Eismaschine	Ice Queen 135 V17 Compacta	Whirlpool Corporation, Benton Harbor, USA
Wasseraufbereitungssystem	Direct-Q 3 UV-R	Merck Millipore, Burlington, Massachusetts, USA
Kugelmühle	BeadRuptor24	Omni International, Kennesaw, USA
Zentrifuge	SIGMA 1-14 Microfuge	Sartorius Stedim BiotechAG, Göttingen, Deutschland
UHPLC	Dionex Ultimate 3000	Dionex GmbH, Idstein, Deutschland
q-ToF-Massenspektrometer	Maxis 4G	Bruker Daltonik GmbH, Bremen, Deutschland
Spritzenpumpe	KDS 100	KD Scientific Inc., Holliston, USA
Hamilton-Spritze	Gasdichte Spritze #1010	Hamilton Company, Bonaduz, Schweiz
FT-NIR-Spektrometer	TANGO	Bruker Optik GmbH, Ettlingen, Deutschland

Tab. 13.5: In dieser Arbeit verwendete Verbrauchsmaterialien.

Material	Bezeichnung	Hersteller/Lieferant
Alufolie	-/-	handelsüblich
Filterpapier	Durchmesser 90 mm	VWR International GmbH, Darmstadt, Deutschland
Kunststoffgefäße	Schraubröhre, Spitzboden, steril, pyrogenfrei, endotoxinfrei, PP	Sarstedt AG & Co., Nümbrecht, Deutschland
Einmalspatel	Einmal-Kunststoff-Spatel, mikro, antistatisch, unsteril, PP	Carl Roth GmbH & Co. KG, Karlsruhe, Deutschland
Reaktionsgefäße	Eppendorf Safe-Lock-Tubes 2 mL	Eppendorf AG, Hamburg, Deutschland
Stahlkugeln	Fahrradlagerkugeln, Durchmesser 3/8 Zoll	Fahrradbedarfsgeschäft, Hamburg, Deutschland
Pipettenspitzen	epT.I.P.S.	Eppendorf AG, Hamburg, Deutschland
HPLC-Vials	1.5 mL, braun, N11, flach, Rollrand, 11,6 x 32 mm	Macherey-Nagel GmbH & Co. KG, Düren, Deutschland
Bördelkappen	Aluminium-Bördelkappen N11, mit Dichtscheibe, Mittelloch, Naturkautschuk/Butyl/TEF, silber	Macherey-Nagel GmbH & Co. KG, Düren, Deutschland
Chromatographiesäule I	Accucore RP-MS Chromatographiesäule (150x2.1 mm; 2.6 µm Partikeldurchmesser)	Thermo Fischer Scientific, Braunschweig, Deutschland
Chromatographiesäule II	Poroshell 120 EC-C18 Chromatographiesäule (50x4.6 mm; 2.7 µm Partikeldurchmesser)	Agilent Technologies, Waldbronn, Deutschland
Chromatographiesäule III	Cogent DiamondHydrid 2.0 Chromatographiesäule (150x2.1 mm, 4 µm Partikeldurchmesser)	MicroSolv Technology, Leland, USA
Zellstofftücher	-/-	Kimtech, Kimberley-Clark Professional, Koblenz-Rheinhafen, Deutschland
Glasvials	5 mL, klar, flach, Rollrand, 52x22x1.2 mm	Nipro Diagnostics Germany GmbH, Ratingen, Deutschland

Tab. 13.6: In dieser Arbeit verwendete Chemikalien.

Substanz	Formel	CAS-Nummer	Reinheitsgrad	Hersteller
Acetonitril	CH_3CN	75-05-8	LC-MS grade	Carl Roth GmbH & Co. KG, Karlsruhe, Deutschland
Ameisensäure	HCOOH	64-18-6	<i>p.a.</i>	Acros Organics, (Thermo Fisher Scientific), Geel, Belgien
Ammoniumacetat	$\text{NH}_4\text{CH}_3\text{COO}$	631-61-8	<i>p.a.</i>	Carl Roth GmbH & Co. KG, Karlsruhe, Deutschland
Ammoniumformiat	NH_4HCOO	540-69-2	$\geq 95 \%$	Carl Roth GmbH & Co. KG, Karlsruhe, Deutschland
1-Butanol	$\text{C}_4\text{H}_9\text{OH}$	71-36-3	<i>p.a.</i>	Honeywell Deutschland GmbH, Offenbach, Deutschland
Butylhydroxytoluol	$\text{C}_{15}\text{H}_{24}\text{O}$	128-37-0	$\geq 99 \%$	Carl Roth GmbH & Co. KG, Karlsruhe, Deutschland
Chloroform	CHCl_3	67-66-3	HPLC grade	Carl Roth GmbH & Co. KG, Karlsruhe, Deutschland
Dichlormethan	CH_2Cl_2	75-09-2	$> 99.5 \%$, zur Synthese	Carl Roth GmbH & Co. KG, Karlsruhe, Deutschland
Essigsäure	CH_3COOH	64-19-7	<i>p.a.</i>	Acros Organics, (Thermo Fisher Scientific), Geel, Belgien
Ethanol	$\text{C}_2\text{H}_5\text{OH}$	64-17-5	HPLC grade	VWR International GmbH, Darmstadt, Deutschland
Essigsäureethylester	$\text{C}_4\text{H}_8\text{O}_2$	141-78-6	<i>p.a.</i>	Merck KGaA, Darmstadt, Deutschland
<i>n</i> -Heptan	C_7H_{16}	142-82-5	HPLC grade	Merck KGaA, Darmstadt, Deutschland
<i>n</i> -Hexan	C_6H_{14}	110-54-3	<i>p.a.</i>	Carl Roth GmbH & Co. KG, Karlsruhe, Deutschland
Isopropanol	$\text{C}_3\text{H}_7\text{OH}$	67-63-0	LC-MS grade	Carl Roth GmbH & Co. KG, Karlsruhe, Deutschland
Methanol	CH_3OH	67-56-1	LC-MS grade	Carl Roth GmbH & Co. KG, Karlsruhe, Deutschland

Tab. 13.6: In dieser Arbeit verwendete Chemikalien (Fortsetzung).

Substanz	Formel	CAS-Nummer	Reinheitsgrad	Hersteller
Methyl- <i>tert.</i> -butylether	C ₅ H ₁₂ O	1634-04-4	HPLC grade	Carl Roth GmbH & Co. KG, Karlsruhe, Deutschland
Natriumhydroxid	NaOH	1310-73-2	ca. 99.5 %, reinst	Merck KGaA, Darmstadt, Deutschland
Hexakis(2,2-difluoroethoxy) phosphazene	C ₁₂ H ₁₈ F ₁₂ N ₃ O ₆ P ₃	186817-57-2	> 95 %	Santa Cruz Biotechnology, Dallas, Texas, USA
Stickstoff, flüssig	N ₂ (flüssig)	7727-37-9	-	Linde AG, Pullach, Deutschland
Trockeneis	CO ₂ (fest)	124-38-9	-	Linde AG, Pullach, Deutschland
<i>Analytische Standardsubstanzen</i>				
α-Tocopherol	C ₂₉ H ₅₀ O ₂	10191-41-0	≥ 95 %	Sigma Aldrich, München, Deutschland
γ-Tocopherol	C ₂₈ H ₄₈ O ₂	54-28-4	≥ 95 %	Sigma Aldrich, München, Deutschland
PC (36:2)	C ₄₄ H ₈₄ NO ₈ P	4235-95-4	> 99 %	Avanti Polar Lipids, Alabaster, Alabama, USA
PE (36:0)	C ₄₁ H ₈₂ NO ₈ P	1069-79-0	> 99 %	Avanti Polar Lipids, Alabaster, Alabama, USA
DG (32:0)	C ₃₅ H ₆₈ O ₅	30334-71-5	> 99 %	Avanti Polar Lipids, Alabaster, Alabama, USA
TG (48:0)	C ₅₁ H ₉₈ O ₆	122-32-7	≥ 99 %	Sigma Aldrich, München, Deutschland

Die in dieser Arbeit durchgeführten Analysen beruhen zudem stark auf der mittels Software durchgeführten Auswertung der Analysedaten. Die dabei verwendeten Programme sind in **Tab. 13.7** zusammengestellt.

Tab. 13.7: In dieser Arbeit verwendete Software.

Software	Hersteller	Zweck
Chromelion 2.14	Dionex, Idstein, Deutschland	Steuerung von UHPLC / Massenspektrometer
HyStar 3.2	Bruker Daltonik, Bremen, Deutschland	
oToFControl 3.4	Bruker Daltonik, Bremen, Deutschland	
DataAnalysis 4.1	Bruker Daltonik, Bremen, Deutschland	Auswertung Rohdaten Massenspektrometrie / „Feature Finding“
ProfileAnalysis 2.1	Bruker Daltonik, Bremen, Deutschland	Erstellung Consensus Peaklisten / Multivariate Datenanalyse
QuantAnalysis 2.1	Bruker Daltonik, Bremen, Deutschland	Auswertung Rohdaten Massenspektrometrie / Integration von Peaks
ProteoWizard 3.0	https://proteowizard.sourceforge.io , Vanderbilt University, Niederlande Matt Chambers <i>et al.</i>	Auswertung Rohdaten Massenspektrometrie
OPUS 7.5	Bruker Optik, Ettlingen, Deutschland	Steuerung von FT-NIR- Spektrometer
R 3.6.3	https://www.r-project.org/	Datenauswertung (Statistik, multivariate Datenanalyse, Klassifikation etc.) / Grafiken
Inkscape 0.92.4	https://inkscape.org/de/	Grafiken
LaTeX to SVG (MathJax)	https://viereck.ch/latex-to-svg/ , Thomas Lochmatter	Erstellung von Formeln
Microsoft Excel 2010	Microsoft, Redmond, USA	Datenauswertung quantitative Bestimmungen
Microsoft Word 2010	Microsoft, Redmond, USA	Textverarbeitung

13.2 Sicherheit und Entsorgung

Die in dieser Arbeit verwendeten Gefahrstoffe sind in **Tab. 13.8** zusammen mit ihrer GHS-Einstufung, den H- und P-Sätzen^{402–404} und Hinweisen zur Entsorgung aufgeführt. In **Tab. 13.9** sind zudem diejenigen Chemikalien aufgeführt, welche eine Einstufung als karzinogen, mutagen oder reproduktionstoxisch (KMR-Stoffe) aufweisen⁴⁰⁵.

Tab. 13.8: Gefahrensymbole nach GHS, H- und P-Sätze sowie Hinweise zur Entsorgung, der in dieser Arbeit verwendeten Chemikalien.

Substanz	GHS-Symbole	H-Sätze	P-Sätze	Entsorgung
Acetonitril	GHS02, GHS07	H225, H302+H312+H332, H319	P210, P280, P301, P312 + P330, P302 + P352 + P312, P304 + P340 + P312, P305 + P351 + P338	O
Ameisensäure	GHS02, GHS05, GHS06	H226, H302, H314, H331	P210, P280, P301 + P312, P303 + P361 + P353, P304 + P340 + P310, P305 + P351 + P338	O
Ammoniumacetat	Kein gefährlicher Stoff gemäß der Verordnung (EG) Nr. 1272/2008.			B
Ammoniumformiat	GHS07	H319	P264, P280, P305 + P351 + P338, P337 + P313	B
1-Butanol	GHS02, GHS05, GHS07	H226, H302, H315, H318, H335, H336	P210, P233, P280, P301 + P312, P303 + P361 + P353, P305 + P351 + P338	O
Butylhydroxytoluol	GHS09	H410	P273, P391, P501	O
Chloroform	GHS06, GHS08	H302, H315, H319, H331, H336, H351, H361d, H372, H412	P201, P273, P301 + P312 + P330, P302 + P352, P304 + P340 + P311, P308 + P313	H
Dichlormethan	GHS07, GHS08	H315, H319, H336, H351	P201, P202, P261, P302 + P352, P305 + P351 + P338, P308 + P313	H
DG (32:0)	Noch nicht vollständig untersuchter Stoff.			O
Essigsäure	GHS02, GHS05	H226, H314	P210, P233, P240, P280, P303 + P361 + P353, P305 + P351 + P338	O
Ethanol	GHS02, GHS07	H225, H319	P210, P233, P240, P241, P242, P305 + P351 + P338	O
Ethylacetat	GHS02, GHS07	H225, H319, H336	P210, P233, P240, P241, P242, P305 + P351 + P338	O
<i>n</i> -Heptan	GHS02, GHS07, GHS08, GHS09	H225, H304, H315, H336, H410	P210, P273, P301+P310, P304+P340, P331, P403+P235	O
<i>n</i> -Hexan	GHS02, GHS07, GHS08, GHS09	H225, H304, H315, H336, H361f, H373, H411	P202, P280, P303+P361+P353, P304+P340, P308+P313	O

Tab. 13.8: Gefahrensymbole nach GHS, H- und P-Sätze sowie Hinweise zur Entsorgung, der in dieser Arbeit verwendeten Chemikalien (Fortsetzung).

Substanz	GHS-Symbole	H-Sätze	P-Sätze	Entsorgung
Isopropanol	GHS02, GHS07	H225, H319, H336	P210, P233, P240, P241, P242, P305 + P351 + P338	O
Methanol	GHS02, GHS06, GHS08	H225, H301 + H311 + H331, H370	P210, P233, P280, P301 + P310, P303 + P361 + P353, P304 + P340 + P311	H
Methyl- <i>tert.</i> -butylether	GHS02, GHS07	H225, H315	P210, P233, P240, P241, P242, P303 + P361 + P353	O
Natriumhydroxid	GHS05	H290, H314	P234, P260, P280, P301 + P330 + P331, P303 + P361 + P353, P305 + P351 + P338	B
PC (36:2)	Noch nicht vollständig untersuchter Stoff.			O
PE (36:0)	Noch nicht vollständig untersuchter Stoff.			O
Hexakis(2,2-difluoroethoxy) phosphazene	GHS07	H315, H319, H335	P261, P271, P280	H
Stickstoff, flüssig	GHS04	H281	P282, P336 + P315, P403	A
α -Tocopherol	GHS07	H317	P261, P272, P280, P302 + P352, P333 + P313, P362 + P364	O
γ -Tocopherol	Kein gefährlicher Stoff gemäß der Verordnung (EG) Nr. 1272/2008.			O
Trockeneis	Nicht nach der Verordnung (EG) Nr. 1272/2008 eingestuft.			A
TG (48:0)	Kein gefährlicher Stoff gemäß der Verordnung (EG) Nr. 1272/2008.			O

A = Reste unter einem Laborabzug abdampfen lassen.

B = (Feststoffe ggfls. zuvor in Wasser auflösen), dann in Behälter für „andere Basen“ geben.

S = (Feststoffe ggfls. zuvor in Wasser auflösen), dann in Behälter für „andere Säuren, HCl/H₂SO₄-haltig“ geben.

O = (Feststoffe ggfls. zuvor in geeignetem Lösungsmittel auflösen), dann in Behälter für „organische Lösungsmittel, halogenfrei“ geben.

H = (Feststoffe ggfls. zuvor in geeignetem Lösungsmittel auflösen), dann in Behälter für „organische Lösungsmittel, halogenhaltig“ geben.

Tab. 13.9: In dieser Arbeit verwendete KMR-Stoffe der Kategorien GHS 1A und 1B.

IUPAC-Name	CAS-Nummer	GHS-Kategorie	Verwendete Menge
Trichlormethan	67-66-3	1B	insgesamt ca. 750 mL

13.3 Ergänzungen zum Ergebnisteil

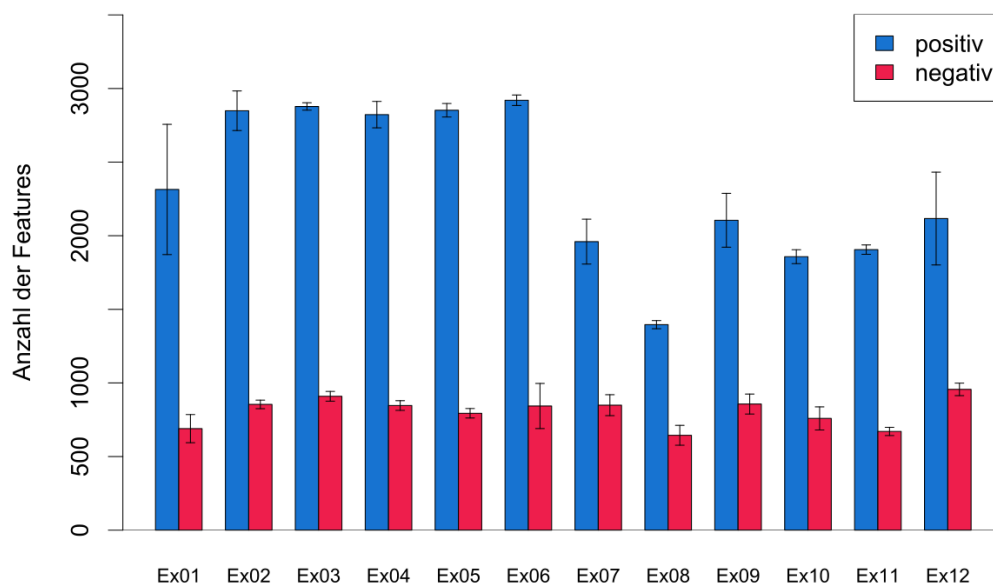


Abb. 13.1: Anzahl der Features in der QC-Probe je Extraktionsmethode für polare Metabolite. **Blau:** positiver Ionenmodus. **Rot:** negativer Ionenmodus. Evaluierte Extraktionsmethoden, siehe **Tab. 6.1** (n = 3).

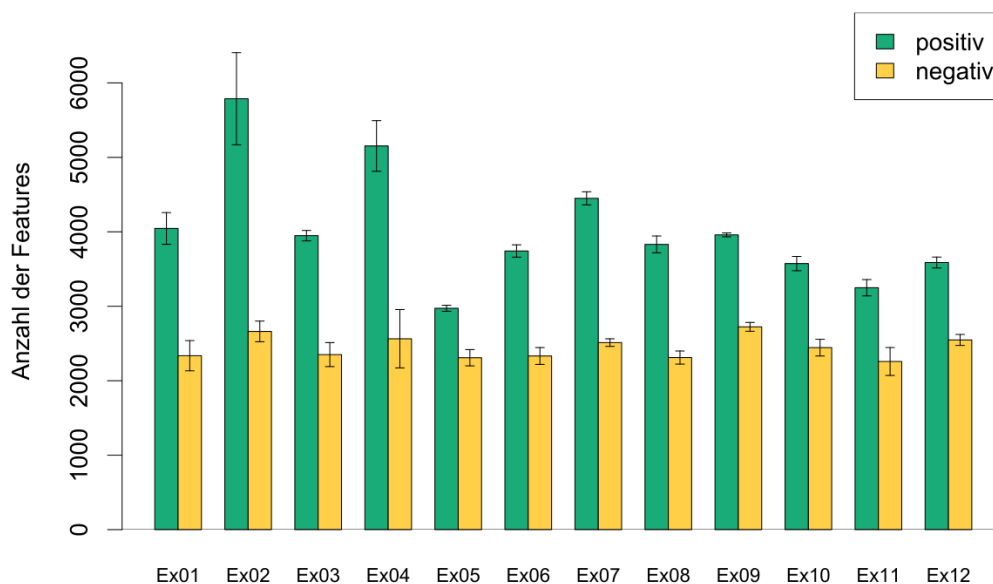


Abb. 13.2: Anzahl der Features in der QC-Probe je Extraktionsmethode für unpolare Metabolite. **Grün:** positiver Ionenmodus. **Gelb:** negativer Ionenmodus. Evaluierte Extraktionsmethoden, siehe **Tab. 6.2** (n = 3).

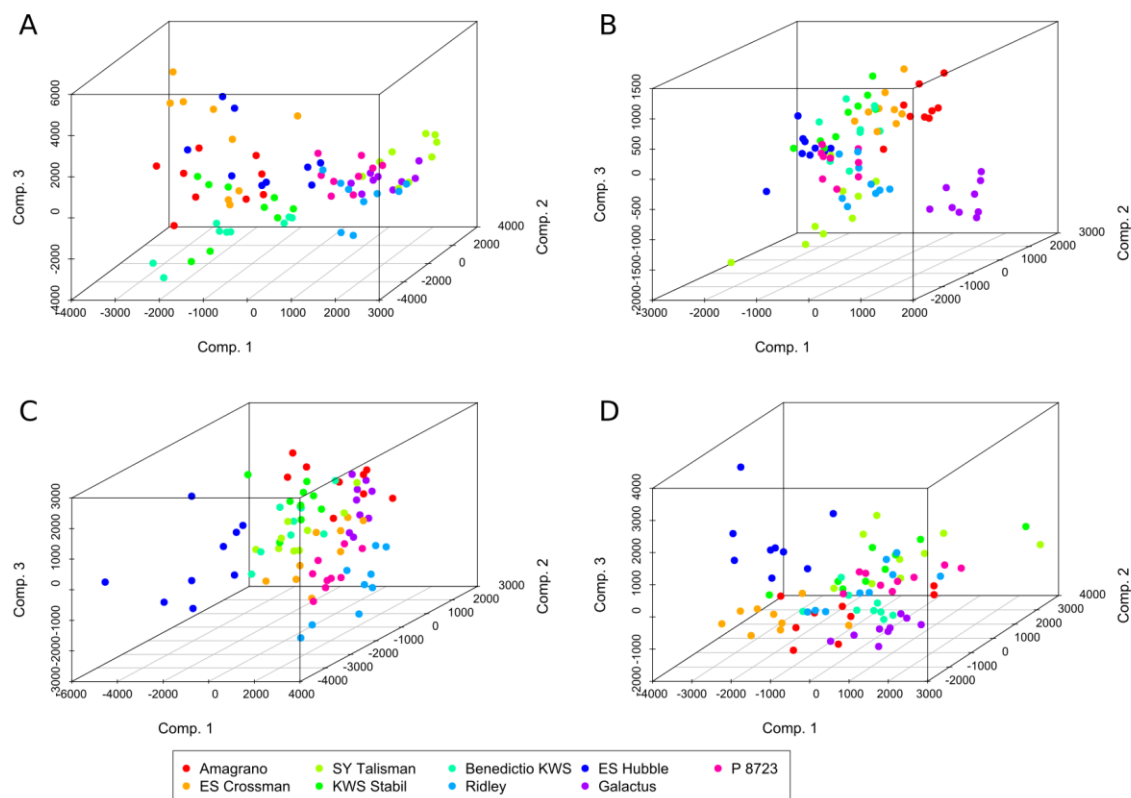


Abb. 13.3: Scoresplots der PLS-DA (nach *Pareto-Scaling*) des Sortenprobensets bei Trennung nach Körnermaissorte. Polare Metabolite detektiert im positiven Ionenmodus (A) bzw. im negativen Ionenmodus (B). Unpolare Metabolite, detektiert im positiven (C) bzw. im negativen (D) Ionenmodus. Die Körnermaissorte ist farblich gekennzeichnet, siehe Legende. Die Q^2 -Werte betrugen 0.80, 0.61, 0.39 und 0.74.

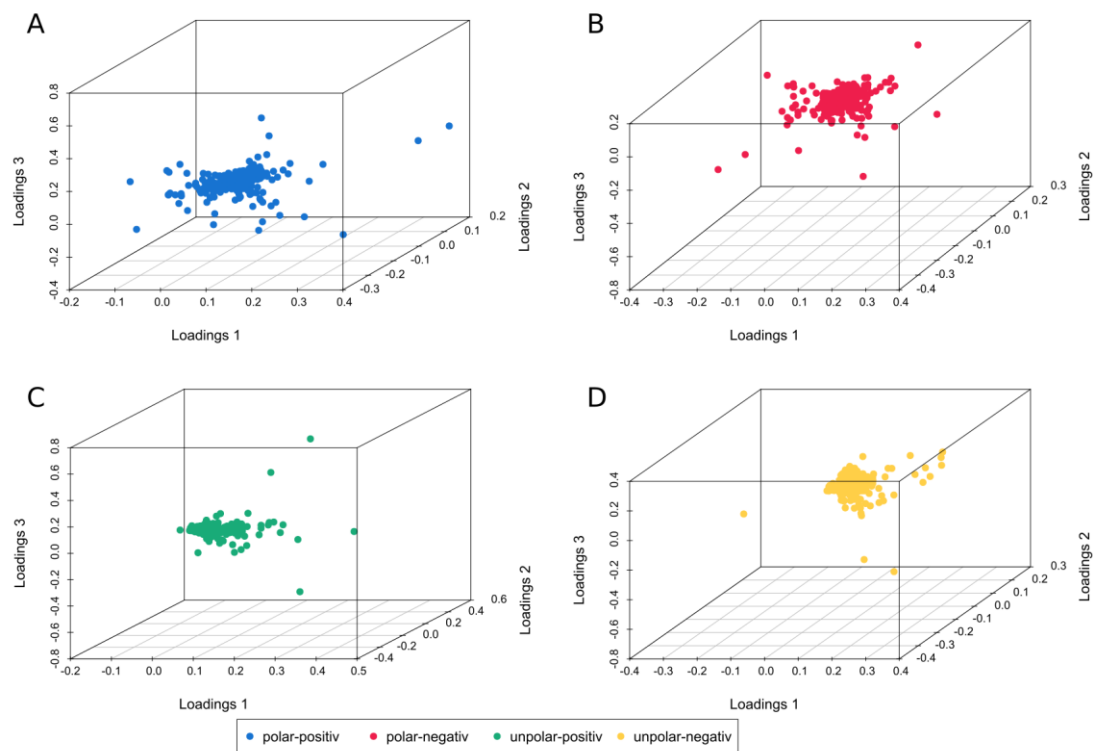


Abb. 13.4: Loadingsplots der PCA der Metabolite unterteilt nach genutzter LC-MS-Methode. Polare Metabolite detektiert im positiven Ionenmodus (A) bzw. im negativen Ionenmodus (B). Unpolare Metabolite, detektiert im positiven (C) bzw. im negativen (D) Ionenmodus.

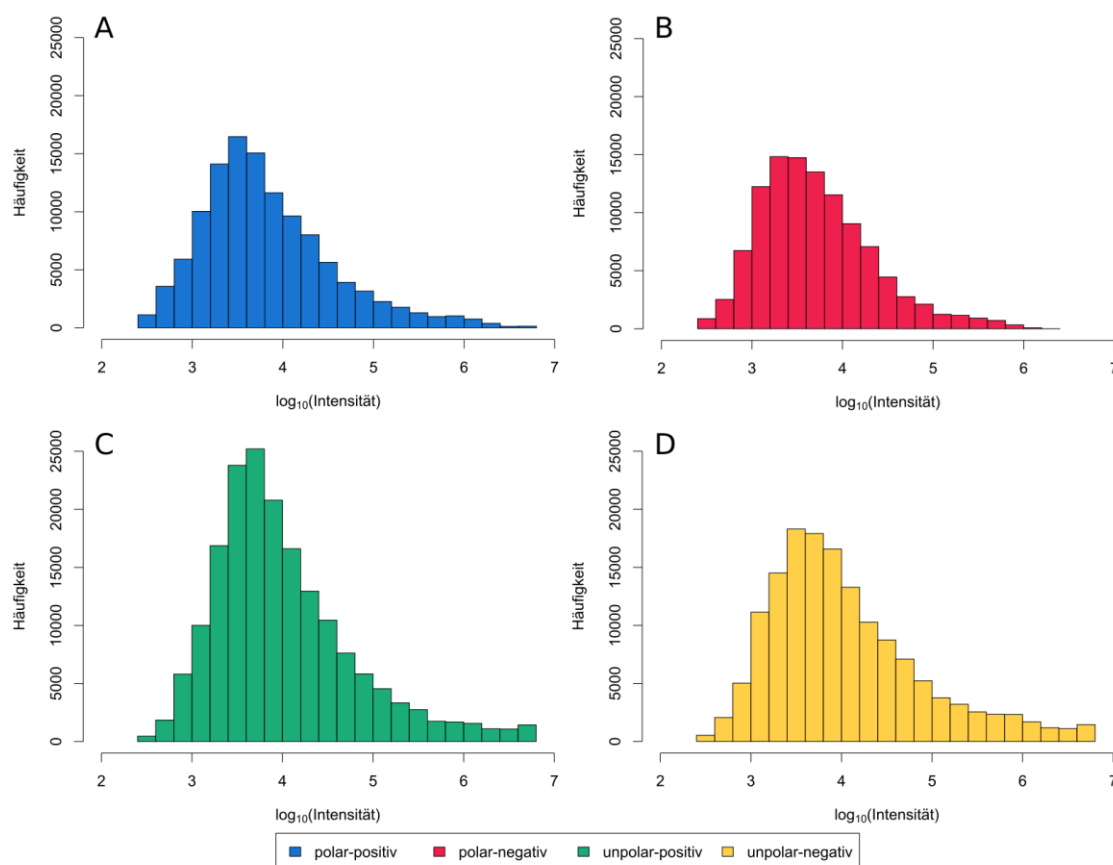


Abb. 13.5: Histogramme der Intensitätenverteilung der detektierten Features der Metabolite. Unterteilt nach genutzter LC-MS-Methode. Polare Metabolite detektiert im positiven Ionenmodus (A) bzw. im negativen Ionenmodus (B). Unpolare Metabolite, detektiert im positiven (C) bzw. im negativen (D) Ionenmodus.

Tab. 13.10: Bedeutende Markerfeatures in polaren Extrakten für die Herkunftsbestimmung von Körnermais. Ermittelt durch die Analyse der Metabolite mittels der LC-MS-Methode für polare Metabolite; nach Extraktion der Proben mittels Methanol/Wasser 2:1 v/v.

Ionenmodus	Retentionszeit [min]	m/z-Verhältnis	Addukt	Summenformel	Identifikation
positiv	9.52	756.554	$[M+H]^+$	$C_{42}H_{78}NO_8P$	PC (34:3)
positiv	9.54	758.570	$[M+H]^+$	$C_{42}H_{80}NO_8P$	PC (34:2)
positiv	9.45	782.570	$[M+H]^+$	$C_{44}H_{80}NO_8P$	PC (36:4)
negativ	8.07	738.507	$[M+Na-2H]^-$	$C_{39}H_{76}NO_8P$	PE (34:1)
negativ	9.86	802.560	$[M+HCOOH-H]^-$	$C_{42}H_{80}NO_8P$	PC (34:2)
negativ	9.78	826.561	$[M+HCOOH-H]^-$	$C_{44}H_{80}NO_8P$	PC (36:4)
negativ	9.77	828.575	$[M+HCOOH-H]^-$	$C_{44}H_{82}NO_8P$	PC (36:3)

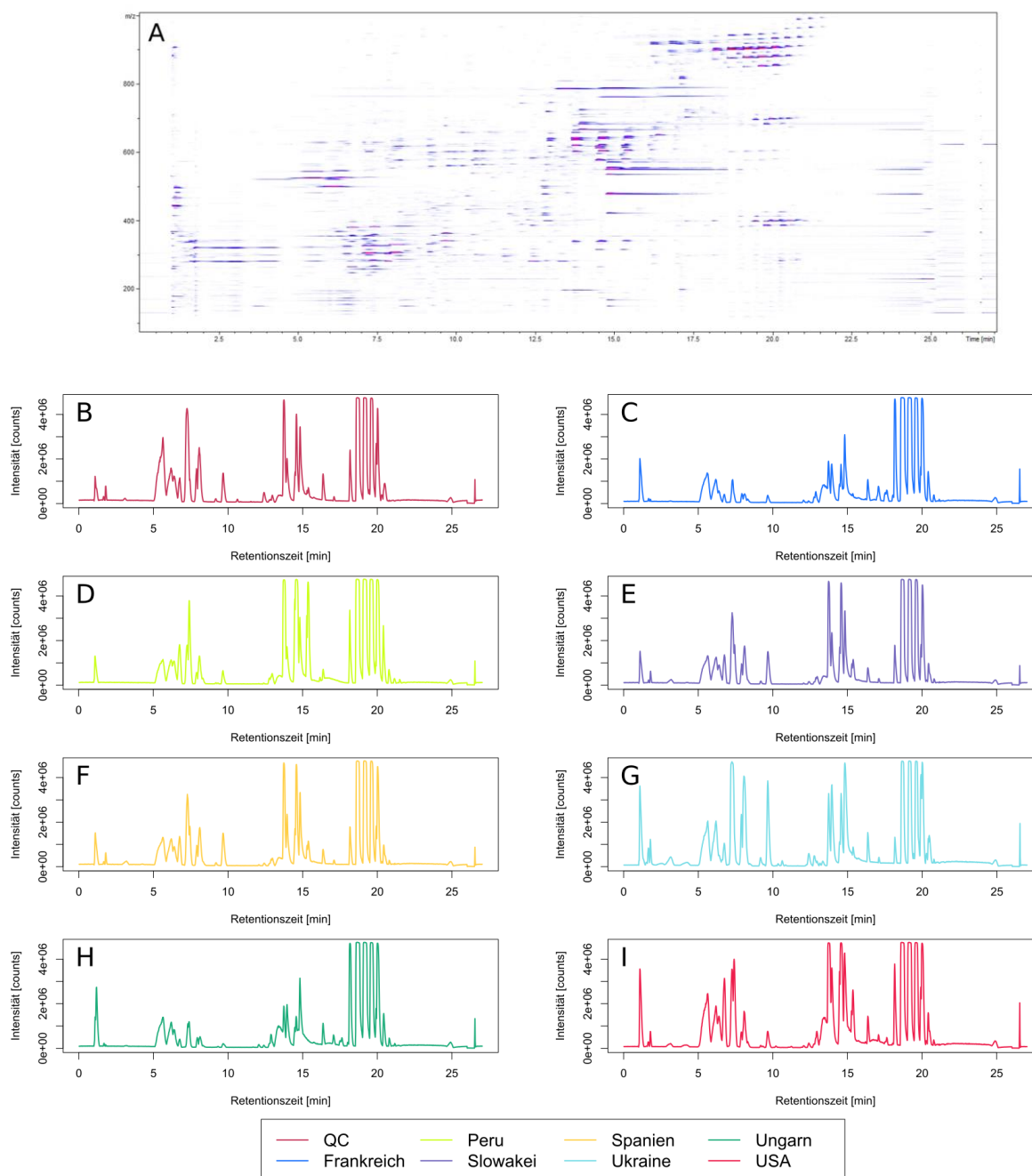


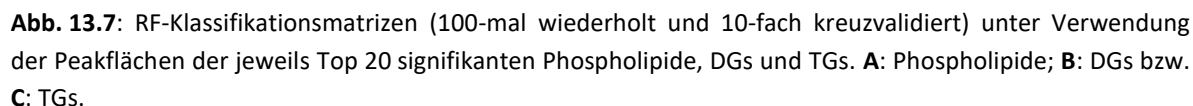
Abb. 13.6: *Survey-View* sowie *Base-Peak-Chromatogramm* (BPC) eines repräsentativen BLIGH & DYER-Extraktes der QC-Probe. **A:** *Survey-View* und **B:** BPC der QC-Probe. Außerdem weitere repräsentative BPCs von Herkunftsproben: **C:** Ukraine; **D:** Slowakei; **E:** Ungarn; **F:** Frankreich; **G:** Spanien; **H:** USA und **I:** Peru. Abbildung modifiziert nach Schütz *et al.*³⁴².

Tab. 13.11: Top 20 der intensivsten, signifikanten (ANOVA-Test p -Wert ≤ 0.01) Peaks je Lipidklasse. Die Peaks wurden als PCs, PEs, DGs resp. TGs entsprechend ihrem m/z -Verhältnis ($Error \leq 5$ ppm) identifiziert. Tabelle teilweise mit Daten aus Schütz *et al.*³⁴² erstellt.

RT [min]	m/z -Verhältnis [Da]	Substanz	Addukt	RSD in QC-Probe (n = 10)	Anzahl sign. Paare im TUKEY-Test
13.3	782.570	PC (36:4)	[M+H] ⁺	11.3 %	14
14.6	758.571	PC (34:2)	[M+H] ⁺	10.5 %	15
14.6	784.586	PC (36:3)*	[M+H] ⁺	11.7 %	14
16.5	760.586	PC (34:1)*	[M+H] ⁺	11.7 %	12
16.4	786.601	PC (36:2)	[M+H] ⁺	11.0 %	11
12.9	716.523	PE (34:2)*	[M+H] ⁺	10.8 %	15
13.0	742.538	PE (36:3)*	[M+H] ⁺	13.2 %	12
12.1	740.522	PE (36:4)	[M+H] ⁺	7.4 %	14
12.3	780.553	PC (36:5)*	[M+H] ⁺	5.7 %	16
13.3	756.553	PC (34:3)	[M+H] ⁺	8.8 %	13
13.9	718.537	PE (34:1)*	[M+H] ⁺	6.6 %	10
14.0	744.553	PE (36:2)	[M+H] ⁺	6.8 %	10
13.0	764.520	PE (38:6)	[M+H] ⁺	12.9 %	11
14.4	732.553	PC (32:1)	[M+H] ⁺	41.5 %	15
13.0	730.537	PC (32:2)	[M+H] ⁺	15.8 %	12
15.7	772.586	PE (38:2)	[M+H] ⁺	23.0 %	16
15.0	746.569	PE (36:1)	[M+H] ⁺	32.5 %	3
13.9	740.519	PC (33:4)	[M+H] ⁺	6.8 %	11
13.3	778.537	PC (36:6)	[M+H] ⁺	35.9 %	15
12.9	768.552	PE (38:4)	[M+H] ⁺	39.8 %	12
14.6	610.541	DG (34:2)*	[M+NH ₄] ⁺	12.6 %	15
15.4	612.556	DG (34:1)*	[M+NH ₄] ⁺	9.1 %	16
13.0	632.525	DG (36:5)*	[M+NH ₄] ⁺	9. %	16
16.2	640.588	DG (36:1)*	[M+NH ₄] ⁺	5.3 %	16
16.3	666.603	DG (38:2)	[M+NH ₄] ⁺	7.2 %	15
15.3	586.540	DG (32:0)*	[M+NH ₄] ⁺	8.0 %	15
13.8	608.524	DG (34:3)	[M+NH ₄] ⁺	6.8 %	14
16.1	614.572	DG (34:0)	[M+NH ₄] ⁺	13.7 %	14
17.8	722.663	DG (42:2)	[M+NH ₄] ⁺	16.0 %	14
18.4	724.684	DG (42:1)	[M+NH ₄] ⁺	7.2 %	14

* Als Marker im optimierten RF-Modell genutzt und mittels MS/MS-Spektren verifiziert.

* Als Marker im optimierten RF-Modell genutzt und mittels MS/MS-Spektren verifiziert.



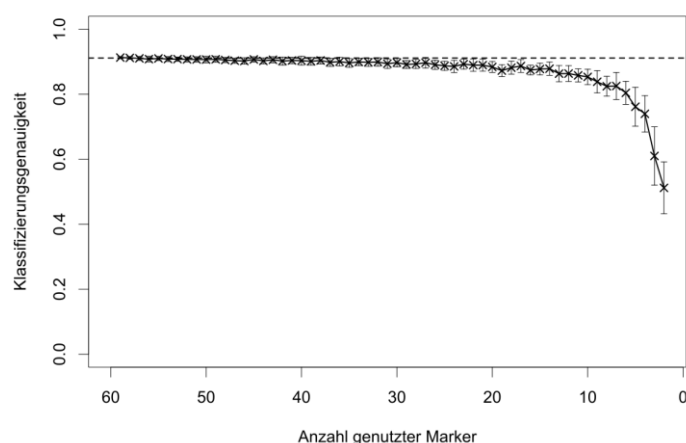


Abb. 13.8: RF-Klassifikationsergebnisse unter Verwendung der Peakflächen der Top 20 intensivsten, signifikanten Phospholipide, DGs und TGs. Die Anzahl an verwendeten Marker wurde dabei von 2-59 variiert, die Marker wurden jeweils 25-mal wiederholt, zufällig aus den 60 Markern ausgewählt. Die RF-Klassifikation wurde 10-fach kreuzvalidiert, sowie jeweils 10-mal pro Markersubset wiederholt.

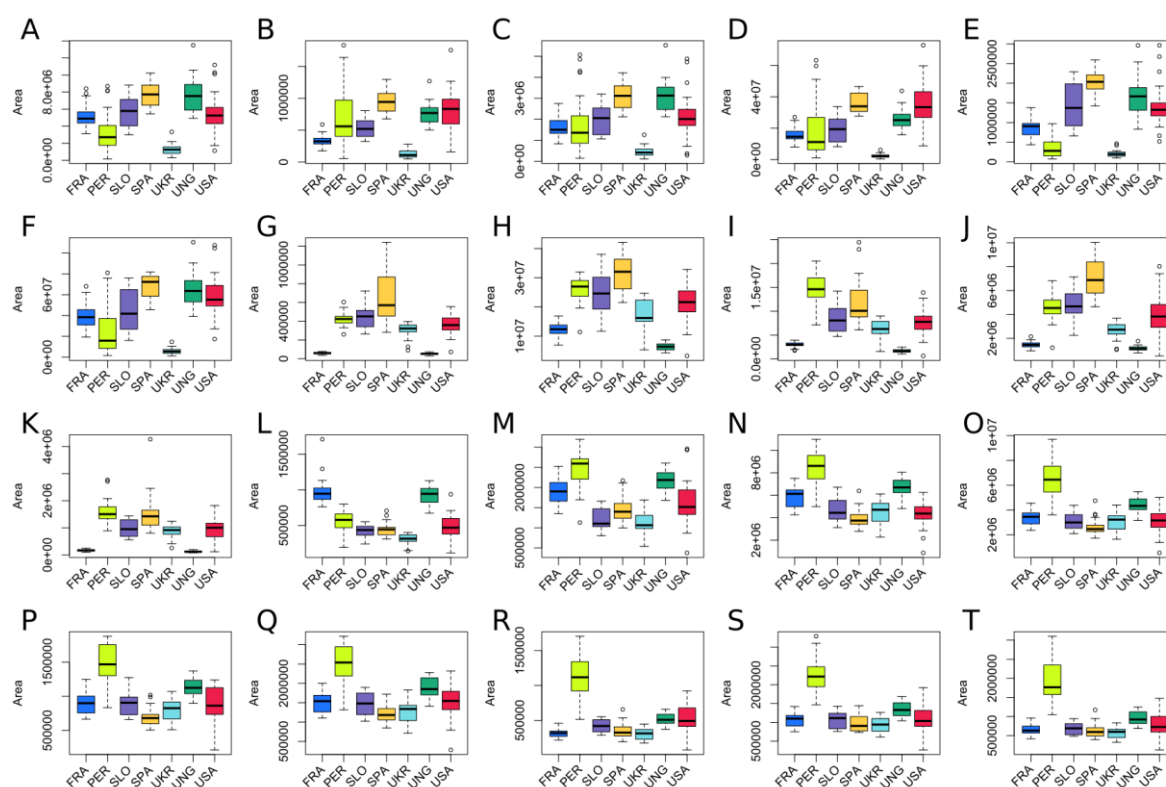


Abb. 13.9: Boxplots der mittleren Peakflächen der 20 Markerfeatures in Abhängigkeit von der geographischen Herkunft. **A:** PC (36:3), **B:** PC (34:1), **C:** PE (34:2), **D:** PE (36:3), **E:** PC (36:5), **F:** PE (34:1), **G:** DG (34:2), **H:** DG (34:1), **I:** DG (36:5), **J:** DG (36:1), **K:** DG (32:0), **L:** TG (56:3), **M:** TG (56:2), **N:** TG (58:3), **O:** TG (60:3), **P:** TG (58:2), **Q:** TG (53:4), **R:** TG (60:2), **S:** TG (52:6), **T:** TG (58:1). Abbildung modifiziert nach Schütz *et al.*³⁴².

Tab. 13.12: Informationen zu den Markerfeatures für die Unterscheidung der geographischen Herkunft von Körnermaisproben. Angegeben sind Substanzname, Retentionszeit, m/z -Verhältnis, Addukt sowie MS/MS-Fragmente, die zur Identifikation der Substanz führten. Tabelle teilweise nach Schütz *et al.*³⁴², modifiziert.

Marker	Summenformel	Addukt	m/z -Verhältnis berechnet [Da]	m/z -Verhältnis beobachtet [Da]	Error [ppm]	Relevante MS/MS-Fragmente	RT [min]
PC (36:3)	C ₄₄ H ₈₂ NO ₈ P	[M+H] ⁺	784.5851	784.5862	1	601.5 184.1	14.6
PC (34:1)	C ₄₂ H ₈₂ NO ₈ P	[M+H] ⁺	760.5851	760.5859	1	577.5 184.1	16.5
PE (34:2)	C ₃₉ H ₇₄ NO ₈ P	[M+H] ⁺	716.5225	716.5229	1	575.5 337.3	12.9
PE (36:3)	C ₄₁ H ₇₆ NO ₈ P	[M+H] ⁺	742.5381	742.5383	< 1	601.5 339.3	13.0
PC (36:5)	C ₄₄ H ₇₈ NO ₈ P	[M+H] ⁺	780.5538	780.5532	1	597.5 184.1	12.3
PE (34:1)	C ₃₉ H ₇₆ NO ₈ P	[M+H] ⁺	718.5381	718.5374	1	577.5 339.3	13.9
DG (34:2)	C ₃₇ H ₆₈ O ₅	[M+NH ₄] ⁺	610.5405	610.5408	< 1	337.3 313.3	14.6
DG (34:1)	C ₃₇ H ₇₀ O ₅	[M+NH ₄] ⁺	612.5561	612.5561	< 1	339.3 313.3	15.4
DG (36:5)	C ₃₉ H ₆₆ O ₅	[M+NH ₄] ⁺	632.5248	632.5245	1	337.3 335.3	13.0
DG (36:1)	C ₃₉ H ₇₄ O ₅	[M+NH ₄] ⁺	640.5874	640.5875	< 1	341.3 339.3	16.2
DG (32:0)	C ₃₅ H ₆₈ O ₅	[M+NH ₄] ⁺	586.5405	586.5396	2	313.3	15.3
TG (56:3)	C ₅₉ H ₁₀₈ O ₆	[M+NH ₄] ⁺	930.8484	930.8515	3	633.6 631.6 601.5	20.5
TG (56:2)	C ₅₉ H ₁₁₀ O ₆	[M+NH ₄] ⁺	932.8641	932.8677	4	633.6 603.5	20.9
TG (58:3)	C ₆₁ H ₁₁₂ O ₆	[M+NH ₄] ⁺	958.8797	958.8827	3	661.6 659.6 601.5	20.8
TG (60:3)	C ₆₃ H ₁₁₆ O ₆	[M+NH ₄] ⁺	986.9110	986.9139	3	689.6 687.6 601.5	21.2
TG (58:2)	C ₆₁ H ₁₁₄ O ₆	[M+NH ₄] ⁺	960.8954	960.8983	3	689.6 661.6 575.5	21.2
TG (53:4)	C ₅₆ H ₁₀₀ O ₆	[M+NH ₄] ⁺	886.7858	886.7881	3	599.5 589.5	19.4
TG (60:2)	C ₆₃ H ₁₁₈ O ₆	[M+NH ₄] ⁺	988.9267	988.9287	2	689.6 603.5	21.6
TG (52:6)	C ₅₅ H ₉₄ O ₆	[M+NH ₄] ⁺	868.7389	868.7401	1	597.5 573.5 571.5	18.3
TG (58:1)	C ₆₁ H ₁₁₆ O ₆	[M+NH ₄] ⁺	962.9116	962.9128	2	689.7 663.6 577.5	21.6

Tab. 13.13: *Ranking* der 20 Markersubstanzen nach ihrer Variableneinfluss in binären RF-Klassifikationsmodellen (ein Land vs. die restlichen Länder, 100-fach wiederholt und 10-fach kreuzvalidiert). Tabelle nach Schütz *et al.*³⁴².

Ranking	FRA vs. Rest	PER vs. Rest	SLO vs. Rest	SPA vs. Rest	UKR vs. Rest	UKR vs. Rest	USA vs. Rest
1	PE (34:1)	TG (60:3)	TG (53:4)	DG (36:5)	PC (34:1)	DG (34:2)	PC (34:1)
2	TG (52:6)	TG (60:2)	PE (34:1)	DG (32:0)	PE (34:1)	DG (34:1)	PE (36:3)
3	DG (34:1)	TG (58:1)	PC (34:1)	DG (34:2)	PC (36:3)	DG (36:1)	PE (34:2)
4	DG (36:1)	TG (56:2)	TG (52:6)	PC (36:5)	PE (36:3)	DG (36:5)	TG (56:3)
5	DG (32:0)	TG (58:3)	DG (36:5)	TG (58:3)	PE (34:2)	DG (32:0)	PC (36:5)
6	DG (34:2)	TG (56:3)	PC (36:5)	TG (56:3)	PC (36:5)	PE (34:1)	TG (58:1)
7	TG (58:1)	DG (34:1)	DG (32:0)	TG (58:2)	TG (52:6)	PE (36:3)	DG (36:5)
8	DG (36:5)	TG (58:2)	TG (58:3)	TG (56:2)	TG (60:2)	TG (58:1)	PC (36:3)
9	PE (36:3)	PC (36:5)	PE (36:3)	PC (34:1)	TG (58:1)	TG (52:6)	DG (34:1)
10	PC (34:1)	DG (36:1)	DG (36:1)	TG (58:1)	TG (53:4)	PC (34:1)	PE (34:1)
11	PC (36:5)	PE (34:2)	DG (34:2)	PE (34:1)	TG (60:3)	TG (60:2)	DG (36:1)
12	PC (36:3)	TG (53:4)	TG (56:3)	PE (36:3)	DG (36:1)	TG (56:2)	TG (58:3)
13	TG (60:2)	PC (36:3)	PE (34:2)	TG (60:2)	DG (34:1)	TG (58:3)	DG (32:0)
14	TG (53:4)	DG (36:5)	TG (60:2)	PE (34:2)	TG (58:3)	TG (53:4)	TG (56:2)
15	TG (58:3)	PC (34:1)	TG (58:1)	DG (34:1)	DG (36:5)	PE (34:2)	DG (34:2)
16	PE (34:2)	PE (36:3)	DG (34:1)	PC (36:3)	TG (56:3)	PC (36:3)	TG (58:2)
17	TG (58:2)	DG (34:2)	TG (58:2)	TG (60:3)	DG (32:0)	TG (60:3)	TG (60:2)
18	TG (56:2)	TG (52:6)	PC (36:3)	DG (36:1)	TG (56:2)	TG (58:2)	TG (60:3)
19	TG (56:3)	DG (32:0)	TG (56:2)	TG (52:6)	TG (58:2)	PC (36:5)	TG (52:6)
20	TG (60:3)	PE (34:1)	TG (60:3)	TG (53:4)	DG (34:2)	TG (56:3)	TG (53:4)

Tab. 13.14: MRM-Übergänge der 20 selektierten Markermetabolite zur Herkunftsbestimmung von Körnermais. Neben der Retentionszeit in der verkürzten LC-Methode, sind zudem die Änderungen bei Peakflächen beim Übergang von MS-Daten zu MS/MS-Daten angegeben.

Marker	Precursorion		Quantifier			Qualifier
	m/z-Verhältnis	Area QC-Probe	m/z-Verhältnis	Area QC-Probe	Verhältnis MS ² /MS ¹	
PC(36:3)*	784.586	10585986	184.073	10265230	97 %	601.523
PC(34:1)	760.586	4988586	184.073	10784273	216 %	577.518
PE(34:2)	716.523	272020	575.503	37245	14 %	337.273
PE(36:3)*	742.538	209180	601.518	17019	8 %	339.29
PC(36:5)	780.553	393350	184.073	19576380	4977 %	597.485
PE(34:1)	718.537	76914	577.515	12716	17 %	339.289
DG(34:2)*	610.541	2552942	313.272	185508	7 %	337.275
DG(34:1)	612.556	1114084	313.276	409684	37 %	339.292
DG(36:5)	632.525	918434	337.275	102904	11%	335.259
DG(36:1)	640.588	204562	341.307	79737	39 %	339.292
DG(32:0)	586.540	105878	313.275	89363	84 %	551.506
TG(56:3)*	930.852	5110529	631.574	408325	8 %	633.585
TG(56:2)*	932.868	2560492	633.589	229425	9 %	603.539
TG(60:3)*	986.914	1124771	689.648	27567	2 %	687.635
TG(58:2)*	960.898	1145538	575.510	87323	8 %	661.618
TG(53:4)	886.788	573853	589.524	184274	32 %	599.508
TG(60:2)*	988.929	579310	689.649	53031	9 %	603.538
TG(58:3)	958.883	350436	661.622	350759	100 %	659.605
TG(52:6)*	868.740	443198	573.493	416231	94 %	597.492
TG(58:1)	962.913	114043	663.636	108129	95 %	577.525

* Marker, die später beispielhaft in allen Proben quantifiziert und eingehender validiert wurden.

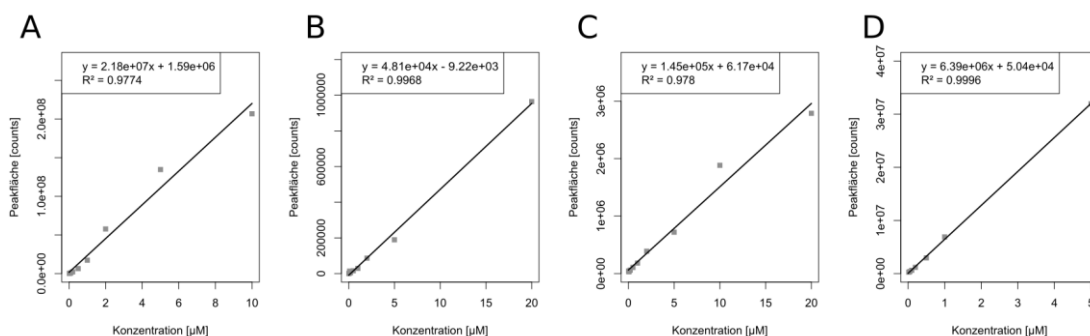


Abb. 13.10: Kalibrationsgeraden der Standardsubstanzen. **A:** PC (36:2), **B:** PE (36:0), **C:** DG (32:0) und **D:** TG (48:0).

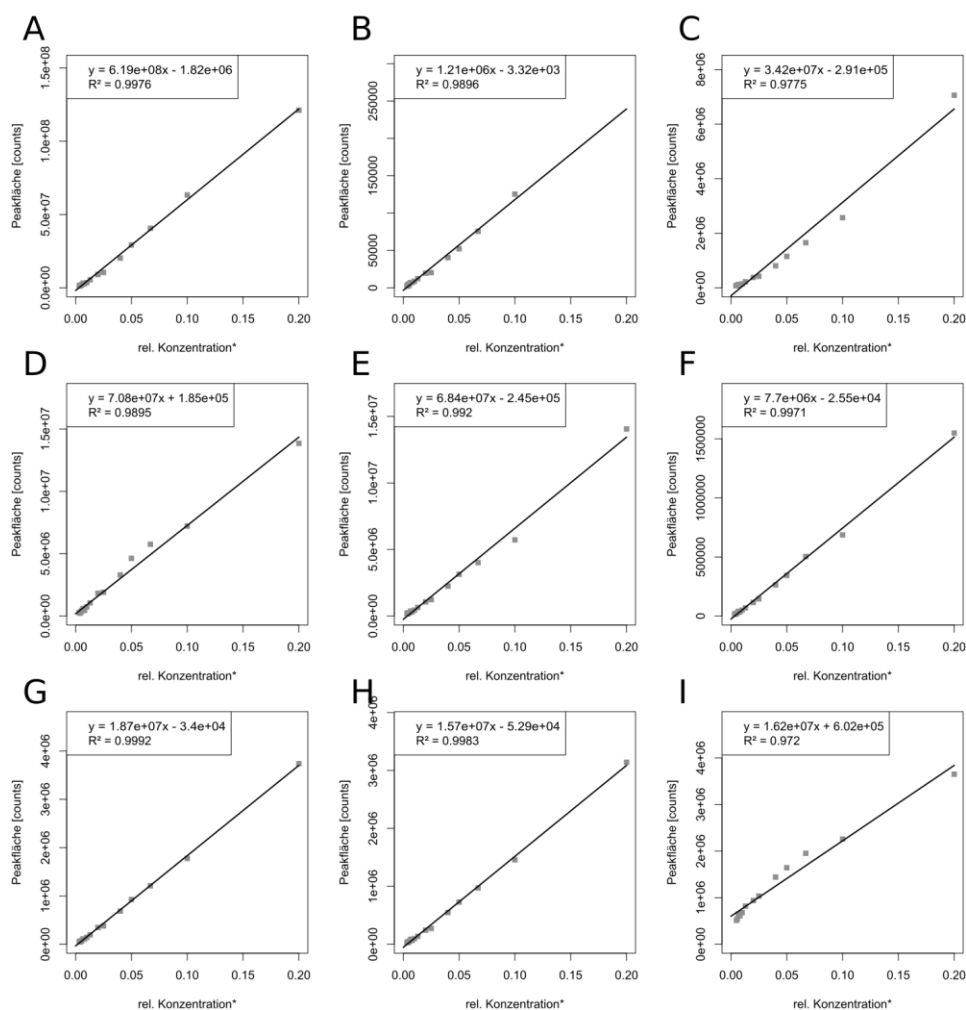


Abb. 13.11: Kalibrationsgeraden der Verdünnungsreihe der neun quantifizierten Markermetabolite (**A-I**).
A: PC (36:3), **B:** PE (36:3), **C:** DG (34:2), **D:** TG (56:3), **E:** TG (56:2), **F:** TG (60:3), **G:** TG (58:2), **H:** TG (60:2) und **I:** TG (52:6).

* Auf der x-Achse ist die Konzentration des jeweiligen Markers in Relation zu der Verdünnungsreihe der Poolprobe angegeben.

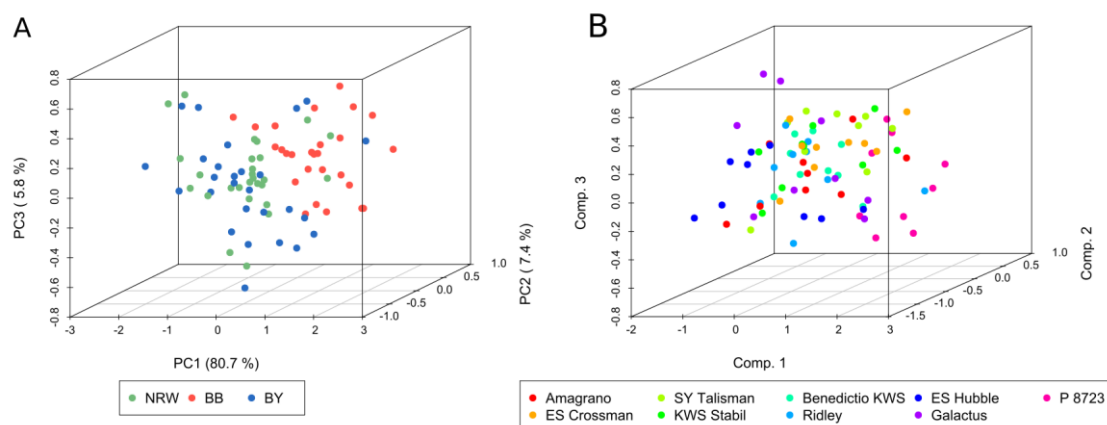


Abb. 13.12: Multivariate Datenanalyse der FT-NIR-Spektren des Sortenprobensets. **A:** PCA-Scoresplot (Proben nach geographischer Herkunft eingefärbt, siehe Legende). **B:** Scoresplot einer PLS-DA-Analyse nach Maissorte (Proben nach Maissorte eingefärbt, siehe Legende; Q^2 -Wert = 0.11).

A					B										
	BB	BY	NRW	CC		Ama	Ben	Cro	Gal	Hub	P8723	Rid	Sta	Tal	CC
BB	25.96	1.00	0.04	96.15 %	Ama	0.57	1.90	0.82	1.66	1.20	0	1.07	1.78	0	6.33 %
BY	1.52	24.50	0.98	90.74 %	Ben	1.42	4.69	0.27	0	0.76	0	0.02	1.84	0	52.11 %
NRW	0.99	1.35	24.66	91.33 %	Cro	0.14	1.12	4.00	0.99	0.42	0	0	2.33	0	44.44 %
					Gal	2.75	0	1.24	2.33	0.51	0.15	1.45	0.56	0.01	25.89 %
					Hub	1.31	0.68	0.80	0.32	4.86	0	0.92	0.11	0	54.00 %
					P8723	0	0	0	0	0	7.01	0.75	0	1.24	77.89 %
					Rid	2.82	0.71	0.05	0.64	0.02	1.01	2.36	1.19	0.20	26.22 %
					Sta	1.34	1.94	1.15	0.50	0.18	0	0.74	2.47	0.68	27.44 %
					Tal	0.07	0.08	0.02	0.06	0	1.15	2.31	1.98	3.33	37.00 %

Abb. 13.13: SVM-Klassifikationsmatrizen einer 100-fach wiederholten 10-fachen Kreuzvalidierung der Proben des Sortenprobensets. **A:** Klassifikation nach Herkunft. **B:** Klassifikation nach Maissorte.

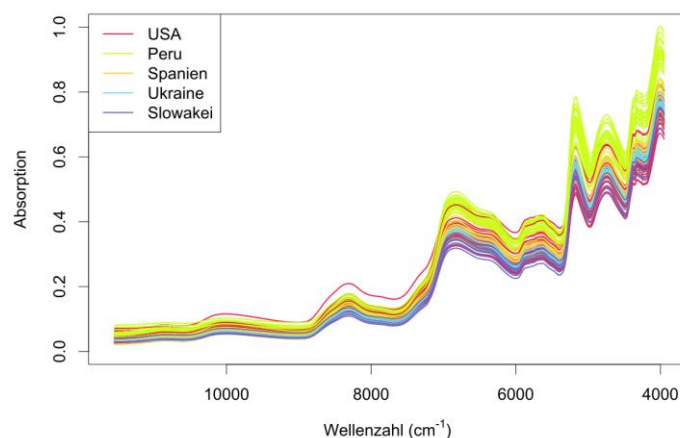


Abb. 13.14: Rohdaten der FT-NIR-Spektren des Herkunftssprobensets. Die geographische Herkunft ist farbkodiert, siehe Legende. Abbildung modifiziert nach Schütz *et al.*³⁶⁸.

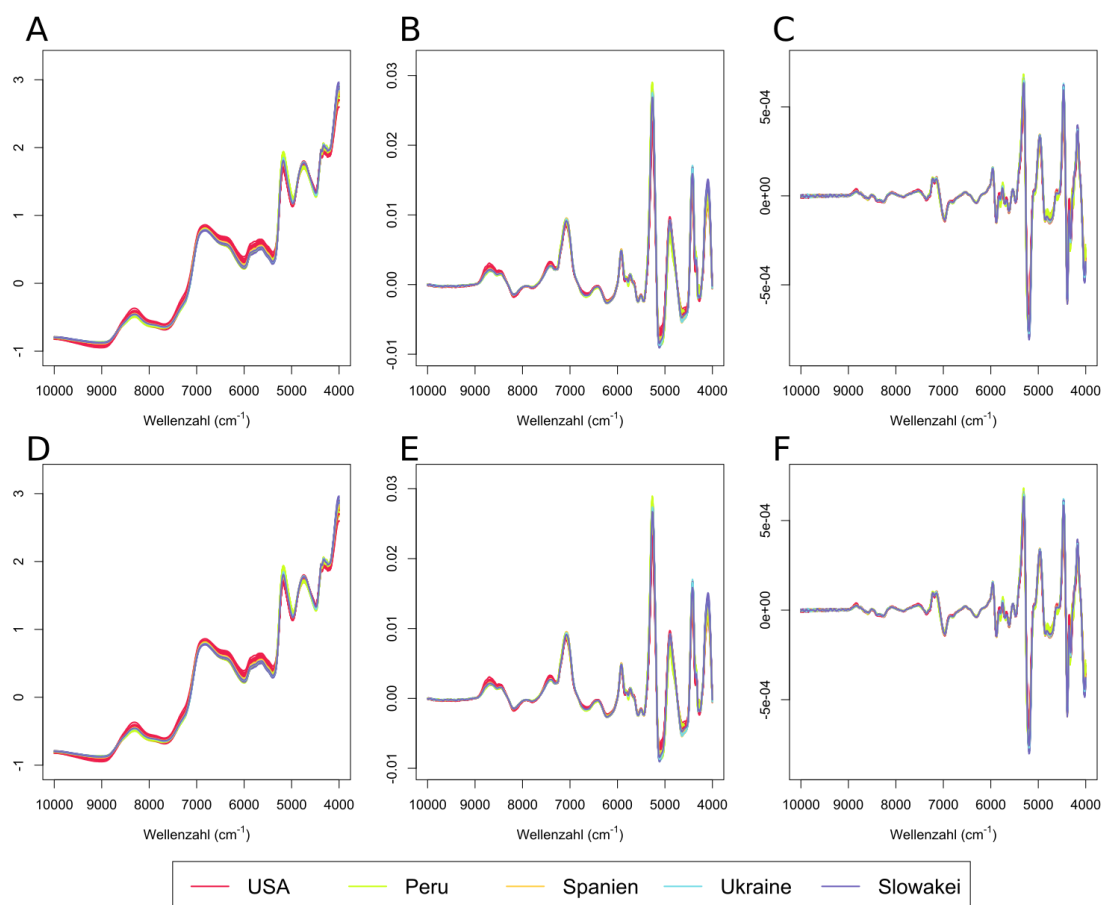


Abb. 13.15: Einfluss von unterschiedlichen spektralen *Preprocessings* auf die FT-NIR-Spektren des Herkunftssprobensets. **A:** SNV; **B:** SNV + 1. Ableitung; **C:** SNV + 2. Ableitung; **D:** SNV + *Binning*; **E:** SNV + 1. Ableitung + *Binning*; **F:** SNV + 2. Ableitung + *Binning*. Die Spektren in der oberen Reihe haben eine Auflösung von 2 cm⁻¹, während die Spektren in der unteren Reihe eine aufgrund des *Binnings* auf 10 cm⁻¹ reduzierte Auflösung aufweisen. Abbildung modifiziert nach Schütz *et al.*³⁶⁸.

Tab. 13.15: Optimierte Parameter für die Klassifikation mittels SVM. Die Parameter wurden mittels eines *Grid-Search* optimiert (10-mal wiederholte, 10-fache Kreuzvalidierung). Zudem wurden verschiedene *Preprocessing*-Varianten untersucht. Tabelle nach Schütz *et al.*³⁶⁸.

	Lineare Kernelfunktion	Quadratische Kernelfunktion		RBF-Kernel	
	<i>Cost</i>	<i>Cost</i>	<i>Scale</i>	<i>Cost</i>	<i>Gamma</i>
SNV	$1.6 \cdot 10^{-3}$	0.32	$1.0 \cdot 10^{-3}$	3.16	$3.2 \cdot 10^{-4}$
SNV + Binning	$6.3 \cdot 10^{-3}$	31.6	$1.0 \cdot 10^{-4}$	3.16	$3.2 \cdot 10^{-3}$
SNV + 1. Ableitung	$5.0 \cdot 10^{-4}$	31.6	$1.0 \cdot 10^{-5}$	3.16	$1.0 \cdot 10^{-4}$
SNV + 1. Ableitung + Binning	$2.5 \cdot 10^{-3}$	31.6	$3.2 \cdot 10^{-5}$	31.6	$3.2 \cdot 10^{-5}$
SNV + 2. Ableitung	$6.3 \cdot 10^{-4}$	31.6	$1.0 \cdot 10^{-5}$	31.6	$1.0 \cdot 10^{-5}$
SNV + 2. Ableitung + Binning	$3.2 \cdot 10^{-3}$	0.32	$3.2 \cdot 10^{-3}$	10.0	$1.0 \cdot 10^{-4}$

A

	PER	SLO	SPA	UKR	USA	CC
PER	25.00	0	0	0	0	100.00 %
SLO	0	11.87	0	2.13	0	84.79 %
SPA	0	0.76	19.24	0	0	96.20 %
UKR	0	1.75	0	17.25	0	90.79 %
USA	0	0	0	0	23.00	100.00 %

B

	SLO	SPA	UKR	CC
SLO	10.93	1.00	2.07	78.07 %
SPA	1.00	19.00	0	95.00 %
UKR	1.62	0	17.38	91.47 %

C

	PER	SLO	SPA	UKR	USA	CC
PER	25.00	0	0	0	0	100.00 %
SLO	0	9.65	0	4.35	0	68.93 %
SPA	0	1.13	18.79	0.05	0.03	93.95 %
UKR	0	2.31	0	16.69	0	87.84 %
USA	0	0.45	1.82	0	20.73	90.13 %

D

	SLO	SPA	UKR	CC
SLO	9.15	0	4.85	65.36 %
SPA	1.94	18.06	0	90.30 %
UKR	2.03	0.07	16.90	88.95 %

E

	PER	SLO	SPA	UKR	USA	CC
PER	25.00	0	0	0	0	100.00 %
SLO	0	8.02	0.98	5.00	0	57.29 %
SPA	0	2.89	16.86	0.25	0	84.30 %
UKR	0	2.73	1.09	15.18	0	79.89 %
USA	0	0	0	0	23.00	100.00 %

F

	SLO	SPA	UKR	CC
SLO	8.36	0.98	4.66	59.71 %
SPA	2.36	17.55	0.09	87.75 %
UKR	3.08	1.41	14.51	76.37 %

G

	PER	SLO	SPA	UKR	USA	CC
PER	25.00	0	0	0	0	100.00 %
SLO	0	9.80	0.16	4.04	0	70.00 %
SPA	0	0.42	18.90	0.68	0	94.50 %
UKR	0	0.12	2.02	16.86	0	88.74 %
USA	0	0.20	0.67	0	22.13	96.22 %

H

	SLO	SPA	UKR	CC
SLO	10.22	0.16	3.62	73.00 %
SPA	1.40	17.83	0.77	89.15 %
UKR	1.52	0.86	16.62	87.47 %

I

	PER	SLO	SPA	UKR	USA	CC
PER	24.57	0.31	0	0.12	0	98.28 %
SLO	0.73	10.45	1.00	1.82	0	74.64 %
SPA	0	3.08	15.43	0.59	0.90	77.15 %
UKR	0	1.60	0.02	17.38	0	91.47 %
USA	0	0	0.66	0	22.34	97.13 %

J

	SLO	SPA	UKR	CC
SLO	11.86	1.01	1.13	84.71 %
SPA	2.82	16.25	0.93	81.25 %
UKR	2.04	0	16.96	89.26 %

Abb. 13.16: SVM-Klassifikationsmatrizen (100-mal wiederholte 10-fache Kreuzvalidierung) unter Nutzung unterschiedlichen Spektrenbereiche. **Links:** Bei Klassifikation aller Proben ; **Rechts:** Bei Klassifikation lediglich der europäischen Proben. Es wurden der komplette FT-NIR-Spektralbereich von 10000 - 4000 cm^{-1} (**A, B**), bzw. lediglich Spektrenausschnitte (**C-J**) zur Klassifikation genutzt. Die Teilspektren umfassten Absorptionsbanden für Proteine (**CD**), für Stärke (**EF**) sowie Absorptionsbanden für Lipide (5990-5420 cm^{-1} , **GH** und 8600-7980 cm^{-1} , **IJ**). Das *Preprocessing* erfolgte mittels SNV + 1. Ableitung + *Binning*.

A							B				
	PER	SLO	SPA	UKR	USA	CC		SLO	SPA	UKR	CC
PER	25.00	0	0	0	0	100.00 %	SLO	10.81	1.00	2.19	77.21 %
SLO	0	9.48	0.31	4.21	0	67.71 %	SPA	1.00	19.00	0	95.00 %
SPA	0	0.94	18.99	0.07	0	94.95 %	UKR	1.47	0.07	17.46	91.89 %
UKR	0	1.08	1.00	16.92	0	89.05 %					
USA	0	0	0	0	23.00	100.00 %					

Abb. 13.17: SVM-Klassifikationsmatrizen (100-mal wiederholte 10-fache Kreuzvalidierung) der FT-NIR-Spektren ohne Wasserbanden. **A:** globales Modell. **B:** europäisches Modell. Es wurden FT-NIR-Spektren von 10000 - 4000 cm^{-1} unter Ausschluss der zwei Wasserbanden bei 7290-6530 und 5400-4970 cm^{-1} , zur Klassifikation genutzt. Das *Preprocessing* erfolgte mittels SNV + 1. Ableitung + *Binning*.

A									B						
	FRA	PER	SLO	SPA	UKR	UNG	USA	CC		FRA	SLO	SPA	UKR	UNG	CC
FRA	28.22	0	0	0	0	7.78	0	78.39 %	FRA	28.47	0	0	0	7.53	79.08 %
PER	0	25.00	0	0	0	0	0	100.00 %	SLO	0	6.28	1.00	6.72	0	44.86 %
SLO	0	0	5.86	1.00	7.14	0	0	41.86 %	SPA	0	0	19.94	0.06	0	99.70 %
SPA	0	0	0.11	19.83	0.06	0	0	99.15 %	UKR	0	1.94	1.20	15.86	0	83.47 %
UKR	0	0	1.01	1.00	16.99	0	0	89.42 %	UNG	4.11	0	0	0	31.89	88.58 %
UNG	3.98	0	0	0	0	32.02	0	88.94 %							
USA	0	0	0	0	0	0	23.00	100.00 %							

Abb. 13.18: SVM-Klassifikationsmatrizen der erweiterten Modelle (Futtermais und Stärkemais; 100-mal wiederholte 10-fache Kreuzvalidierung). **A:** alle Herkunftsländer. **B:** nur europäische Länder. Es wurden die kombinierten FT-NIR-Banden von Protein, Stärke und Lipiden zur Klassifikation genutzt. Das *Preprocessing* erfolgte mittels SNV + 1. Ableitung + *Binning*.

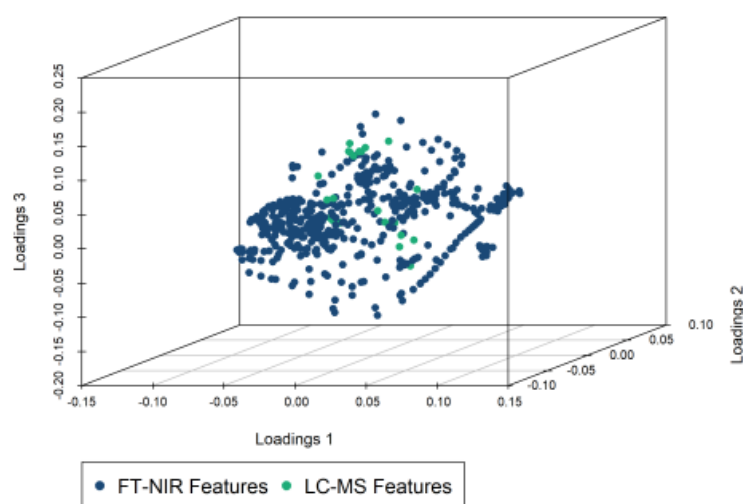


Abb. 13.19: PCA-Loadingsplot der fusionierten LC-MS-Daten und der FT-NIR-Daten nach *Preprocessing* und Angleichung des Wertebereichs mittels *Range-Scaling*. **Blau:** FT-NIR-Datenpunkte, **grün:** LC-MS-Markerfeatures.

Tab. 13.16: *Ranking* der Marker nach ihrem Variableneinfluss in dem auf *low-level* Ebene fusioniertem RF-Klassifikationsmodell. Dargestellt sind die Top 30 von 20 LC-MS-Markern und 216 FT-NIR-Wellenzahlen. Das Klassifikationsmodell zur Herkunftsbestimmung ist 100-fach wiederholt und 10-fach kreuzvalidiert.

Ranking	Variable	Ranking	Variable
1	PC (36:5)	16	6429 cm ⁻¹
2	PE (34:1)	17	6399 cm ⁻¹
3	4960 cm ⁻¹	18	8582 cm ⁻¹
4	PC (34:1)	19	8531 cm ⁻¹
5	PC (36:3)	20	6358 cm ⁻¹
6	5419 cm ⁻¹	21	6480 cm ⁻¹
7	4950 cm ⁻¹	22	4940 cm ⁻¹
8	PE (36:3)	23	8511 cm ⁻¹
9	6511 cm ⁻¹	24	6388 cm ⁻¹
10	6409 cm ⁻¹	25	6419 cm ⁻¹
11	6521 cm ⁻¹	26	4929 cm ⁻¹
12	6491 cm ⁻¹	27	DG (36:5)
13	PE (34:2)	28	6470 cm ⁻¹
14	5429 cm ⁻¹	29	8521 cm ⁻¹
15	8500 cm ⁻¹	30	6368 cm ⁻¹

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt zu haben. Sofern im Zuge der Erstellung der vorliegenden Dissertationsschrift generative Künstliche Intelligenz (gKI) basierte elektronische Hilfsmittel verwendet wurden, versichere ich, dass meine eigene Leistung im Vordergrund stand und dass eine vollständige Dokumentation aller verwendeten Hilfsmittel gemäß der Guten wissenschaftlichen Praxis vorliegt. Ich trage die Verantwortung für eventuell durch die gKI generierte fehlerhafte oder verzerrte Inhalte, fehlerhafte Referenzen, Verstöße gegen das Datenschutz- und Urheberrecht oder Plagiate.

Hamburg, den 02.02.2025

