

Untersuche den Einfluss der Wassertemperatur auf den Kaffeegeschmack



Studentin/Studenten

*Thomas Mandelz
Alexander Schilling
Patrick Schürmann
Si Ben Tran
Alessia Vannini
Marvin von Rappard
Lukas Zemp
Jan Zwicky*

Betreuerinnen/Betreuer

*Barbara Dravec
Severin Gsponer
Andrea Kennel*

Auftraggeber

*Prof. Dr. Chahan Yeretizian, Zentrumsleitung Kaffee
Kompetenz Zentrum, Zürcher Hochschule für Ange-
wandte Wissenschaften (ZHAW)*

Projektnummer

Pro2D

Vereinbarung

vertraulich

GitHub

<https://github.com/marvinvr/fhnw-pro2d.git>

Fachhochschule Nordwestschweiz, Hochschule für Technik

Abstract

Im Jahr 2019 wurde das Regelwerk der World Barista Championship (WBC) geändert. Neu darf Kaffee mit einer individuell gewählten Wassertemperatur im Bereich von 90,5 bis 96 Grad zubereitet werden. Die vorliegende Arbeit untersucht die Auswirkung der Wasser- und Kaffeetemperatur auf das Aroma des Kaffees.

Es wurden zwei Experimente durchgeführt. Mit drei Wassertemperaturen (90, 93 und 96 Grad) und mit aktiver Kühlung des Spouts oder Eiswürfeln direkt im Kaffee. Bei beiden Stichproben wurde die Menge jedes Moleküls im Kaffee ermittelt. Dazu wurde die chemische Verfahren Gaschromatographie gekoppelt mit Massenspektrometrie und Headspace Gaschromatographie verwendet.

Die Untersuchung zeigte, dass flüchtige Moleküle bei tieferer Wasser- und Kaffeetemperatur häufiger vorkommen. Signifikante Unterschiede gab es bei der Wassertemperatur zwischen 90 und 93 Grad sowie 90 und 96 Grad. Bei der aktiven Kaffeekühlung sind Unterschiede zwischen keiner Kühlung und aktiver Kühlung signifikant. Kaum Signifikanzunterschiede gibt es zwischen den verschiedenen Kühlmengen und -methoden.

Das Projektteam konzentrierte sich darauf, Zusammenhänge zwischen der Menge der flüchtigen Moleküle im Kaffee basierend auf den vier Molekulareigenschaften (Molekülmasse, Löslichkeit, Polarisierbarkeit und Henry-Law-Konstante) in Abhängigkeit der Wasser- und Kaffeetemperatur, zu analysieren.

In der Tendenz nimmt bei höherer Molekülmasse, Löslichkeit, Polarisierbarkeit und Henry-Law-Konstante die Häufigkeit der flüchtigen Moleküle ab. Das höchste Bestimmtheitsmass lag bei 0.42. Die Zahlen gehen in die erwartete Richtung, zeigen aber keinen signifikanten Zusammenhang mit den analysierten Molekulareigenschaften.

Keywords:

Kaffee, flüchtige Moleküle, Normalisierung, Regression, Extraktion, Gaschromatographie, Massenspektrometrie

Vorwort/Dank

Der vorliegende Bericht entstand im Rahmen eines Zwisemester-Projekts an der Fachhochschule Nordwestschweiz, Hochschule für Technik, Studiengang Data Science.

An dieser Stelle möchten wir uns nochmals recht herzlich bei Chahan Yeretian und seinem Team für die Möglichkeit dieses Projekts und die Bereitstellung der Daten bedanken.

Ein besonderer Dank gilt auch dem Team der Fachexperten der FHNW für die Betreuung dieser Arbeit.

Für die kommende World Barista Championship in Melbourne (Australien) wünschen wir Chahan Yeretian und seinem Team von Herzen viel Glück und Erfolg.

Windisch, im Juni 2022

Die Studentin und Studenten

Inhaltsverzeichnis

Abbildungsverzeichnis	v
Tabellenverzeichnis	vii
1 Einleitung.....	1
2 Fragestellung.....	2
3 Daten	3
3.1 <i>Eingereichte Daten</i>	<i>3</i>
3.1.1 VOC-BlackEagle.....	3
3.1.2 Chilling Compound	3
3.2 <i>Daten zu Molekulareigenschaften.....</i>	<i>3</i>
4 Methoden und Resultate	5
4.1 <i>Normalisierung.....</i>	<i>5</i>
4.2 <i>Explorative Datenanalyse.....</i>	<i>6</i>
4.2.1 Molekulareigenschaften der flüchtigen Moleküle des VOC-BlackEagle Datensatzes	6
4.2.2 VOC-BlackEagle.....	8
4.2.3 Chilling Compound	12
4.3 <i>Lineare Regression.....</i>	<i>23</i>
4.3.1 Molekülmasse	24
4.3.2 Löslichkeit	26
4.3.3 Polarisierbarkeit.....	28
4.3.4 Henry-Law-Konstante	30
4.4 <i>Multiple Linear Regression.....</i>	<i>32</i>
5 Schlussbemerkungen	34
Anhang.....	35
A Ergänzende Visualisierungen zu Kapitel 4.2.2.3	36

Abbildungsverzeichnis

Abbildung 1: Visualisierung von der ZHAW bezüglich der Auswertung der sensorischen Experimente	2
Abbildung 2: Beispiel Normalisierung	5
Abbildung 3: Histogramme relevanter, kardinal skaliertter Molekulareigenschaften	7
Abbildung 4: Einteilung in funktionelle Gruppen	7
Abbildung 5: Häufigkeitsverteilung der funktionellen Gruppen	7
Abbildung 6: Anzahl cyclische Verbindungen	8
Abbildung 7: Histogramm normalisierte "Counts"	8
Abbildung 8: Dichteplots für drei zufällig ausgewählte Moleküle	9
Abbildung 9: Boxplots für drei zufällig ausgewählte Moleküle	9
Abbildung 10: Boxplots zur Analyse der Verteilung der Moleküle mit und ohne cyclischen Ring	10
Abbildung 11: Verteilung der "Counts" nach drei zufällig ausgewählten funktionellen Gruppen	10
Abbildung 12: Zusammenhang normalisierte "Counts" und Molekülmasse	11
Abbildung 13: Histogramme der normalisierten Werte für "Area"	12
Abbildung 14: Zusammenhang normalisierte "Areas" und Molekülmasse für Kaffeesorten Arab (oben) und Eugi (unten)	13
Abbildung 15: Zusammenhang normalisierte "Areas" und Löslichkeit für Kaffeesorten Arab (oben) und Eugi (unten)	15
Abbildung 16: Zusammenhang normalisierte "Areas" und Löslichkeit für Kaffeesorte Arab ohne Logarithmus-Transformation	16
Abbildung 17: Zusammenhang normalisierte "Areas" und Polarisierbarkeit für Kaffeesorten Arab (oben) und Eugi (unten)	17
Abbildung 18: Zusammenhang normalisierte "Areas" und Henry-Law-Konstante für Kaffeesorten Arab (oben) und Eugi (unten)	19
Abbildung 19: Zusammenhang normalisierte "Areas" und funktionelle Gruppen für Kaffeesorte Arab	20
Abbildung 20: Zusammenhang normalisierte "Areas" und funktionelle Gruppen für Kaffeesorte Eugi	21
Abbildung 21: Zusammenhang normalisierte "Areas" und Moleküle mit und ohne cyclische Verbindungen für Kaffeesorten Arab (oben) und Eugi (unten)	22
Abbildung 22: Lineare Regression für normalisierte "Areas" und Molekülmasse für Kaffeesorten Arab (oben) und Eugi (unten)	24
Abbildung 23: Lineare Regression für normalisierte "Areas" und Löslichkeit für Kaffeesorten Arab (oben) und Eugi (unten)	26

Abbildungsverzeichnis

Abbildung 24: Lineare Regression für normalisierte "Areas" und Polarisierbarkeit für Kaffeesorten Arab (oben) und Eugi (unten)	28
Abbildung 25: Lineare Regression für normalisierte "Areas" und Henry-Law-Konstante für Kaffeesorten Arab (oben) und Eugi (unten)	30
Abbildung 26: Ergebnisse multiple lineare Regression für Kaffeesorte Eugi und 5g mit der Eiswürfel-Methode aktiv gekühltem Kaffee	32
Abbildung 27: Zusammenhang normalisierte "Counts" und Löslichkeit	36
Abbildung 28: Zusammenhang normalisierte "Counts" und Polarisierbarkeit	37
Abbildung 29: Zusammenhang normalisierte "Counts" und Henry-Law-Konstante noch ohne Logarithmus-Transformation (siehe Kapitel Explorative Datenanalyse zum Datensatz Chilling Compound)	38

Tabellenverzeichnis

Tabelle 1: Relevante Molekulareigenschaften

4

1 Einleitung

Die Kaffeebohne ist neben Tee der weltweit am meisten gehandelte Rohstoff. Neben dem Handel mit dem Rohstoff spielt dessen Weiterverarbeitung eine entscheidende Rolle. Die Röstung, die Mahlung sowie das Aufbrühen von Kaffee wurde in den letzten Jahren immer professioneller. Chahan Yeretzian von der ZHAW Wädenswil ist Leiter der Fachgruppe "Kompetenzzentrum Kaffee und Analytische Technologie". Er untersucht Kaffee und seine chemischen Eigenschaften seit Jahren. Er befasst sich damit, was es genau braucht, damit Kaffee weltmeisterlich perfekt wird.

Die World Barista Championship (WBC) ist der bedeutendste internationale Kaffeewettbewerb. Im Jahr 2019 wurden die Regeln und Vorschriften aktualisiert, so dass die Teilnehmer und Teilnehmerinnen die Temperatur des Extraktionswassers in einem Bereich von 90,5 bis 96 °C variieren können. In den Jahren vor 2019 war die Temperatur festgelegt. Diese neue Flexibilität warf die Frage auf, welchen Einfluss die Temperatur auf den Extraktionsprozess und das sensorische Tassenprofil hat.

2 Fragestellung

Aufgrund der Änderung des Regelwerks hat die ZHAW sensorische Experimente mit einer Kaffeedegustatorin und zwei Kaffeedegustatoren durchgeführt. Jede dieser drei Personen testete 18 Espresso, sechs mit je 90, 93 und 96 Grad heissem Extraktionswasser. Die Espresso bewerteten sie nach acht verschiedenen Gütekriterien: Aroma, Geschmack, Nachgeschmack, Säuregehalt, Körper, Süsse, Bitterkeit und Crema. Die Bewertungen lagen sehr nah beieinander, zeigten jedoch ein Muster. Mit höherer Wassertemperatur stieg die Bitterkeit. Fast alle anderen Gütekriterien bewerteten die Degustatoren tiefer. Bei Crema erkannten sie keine Unterschiede. Signifikant bewerteten sie die Unterschiede zwischen 90 und 93 und 90 und 96 Grad.

Diese Visualisierung erstellte die ZHAW aufgrund des sensorischen Experiments. Weil die Degustatoren bei Crema keine Unterschiede feststellten, verzichteten sie darauf diese im Spinnendiagramm zu integrieren.

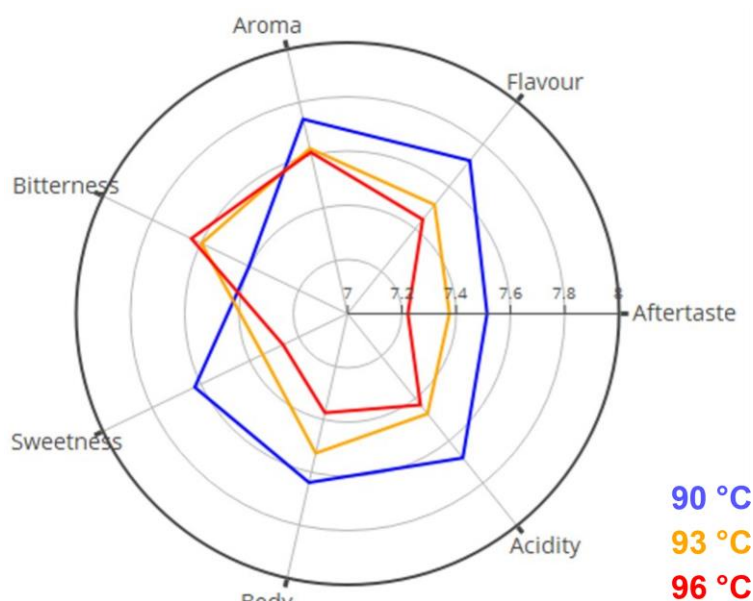


Abbildung 1: Visualisierung von der ZHAW bezüglich der Auswertung der sensorischen Experimente

Dieses sensorische Experiment führte zu der Vermutung, dass flüchtige Stoffe bei der Extraktion von Kaffee bei höheren Temperaturen eher verdampfen. Diese Vermutung sollte im Rahmen des Projektes untersucht und gestützt oder verworfen werden.

Zusätzlich sollten Zusammenhänge zwischen der Menge der flüchtigen Moleküle im Kaffee und chemischen Molekulareigenschaften in Abhängigkeit der Wasser- und Kaffeetemperatur analysiert und visualisiert werden. Dabei standen vier Molekulareigenschaften im Zentrum: Molekülmasse, Löslichkeit, Polarisierbarkeit und die Henry-Law-Konstante.

3 Daten

3.1 Eingereichte Daten

Um die Hypothese zu untersuchen, dass in heisserem Kaffee weniger flüchtige Stoffe vorhanden sind, führte das Team von Chahan Yeretizian zwei Experimente durch. Das Resultat beider Experimente war die Angabe der im Kaffee enthaltenen flüchtigen Moleküle mit ihrer qualitativen Menge.

3.1.1 VOC-BlackEagle

Beim ersten Experiment wurden mit dem Kaffeemaschinenmodell «Black Eagle» für drei Wassertemperaturen - 90, 93 und 96 Grad - je 8 Espresso extrahiert. VOC steht für «volatile organic compounds» (= flüchtige organische Verbindungen). Mit dem chemischen Verfahren Gaschromatographie gekoppelt mit Massenspektrometrie wurden dann die Mengen der einzelnen Moleküle ermittelt, genannt «Counts». Insgesamt wurden diese Werte für 49 flüchtige Moleküle eruiert.

3.1.2 Chilling Compound

Beim zweiten Experiment wurde der Kaffee aktiv gekühlt. Die eingestellte Wassertemperatur war dabei identisch. Für je 12 Espresso wurde das ganze Extrakt, 5g, 12g und 20g des Kaffees oder der Spout an sich, gekühlt. Alle 12 Experimente wurden für die beiden Kaffeesorten Arabica und Eugi durchgeführt. Für die Analyse der flüchtigen Moleküle und deren qualitative Menge, «Area» genannt, wurde das chemische Verfahren Headspace Gaschromatographie verwendet. Insgesamt wurden 46 flüchtige Moleküle festgestellt.

3.2 Daten zu Molekulareigenschaften

Grundlage für die Daten zu den chemischen Molekulareigenschaften bildeten die folgenden vier Datenbanken. Diese Quellen wurden uns von Chahan Yeretizian und seinem Team empfohlen.

- 1) <https://pubchem.ncbi.nlm.nih.gov/>
- 2) <https://webbook.nist.gov/>
- 3) <http://www.chemspider.com/>
- 4) <http://www.thegoodscentscompany.com/>

Anhand des Molekülnamens konnten wir von der PubChem-Datenbank JSON-Dateien zu allen Molekülen beider Datensätze herunterladen. Mittels Python-Skript haben wir die JSON-Dateien ausgelesen. Nicht vollständige oder vorhandene Eigenschaften wurden händisch aus den anderen drei Quellen zusammengesucht. Relevante Molekulareigenschaften wurden pro Datensatz in je ein Dataframe gespeichert und anschliessend als CSV-Dateien exportiert.

3 Daten

Zu den relevanten Molekulareigenschaften zählen:

Tabelle 1: Relevante Molekulareigenschaften

CAS-Nummer	Eindeutige Identifikationsnummer für jedes Molekül.
Molekularformel	Formel einer chemischen Verbindung, die angibt, wieviel Atome der in dieser Verbindung enthaltenen Elemente das Molekül der Verbindung zusammensetzen.
Molekülmasse in g/mol	Gibt die Masse pro Stoffmenge an. Je grösser die Molmasse, desto eher ist es ein Feststoff.
Siedepunkt in °C	Temperatur, bei der ein Molekül vom Aggregatzustand flüssig in gasförmig wechselt.
Schmelzpunkt in °C	Temperatur, bei der ein Molekül vom Aggregatzustand fest in flüssig wechselt.
Löslichkeit	Umfang, in welchem ein Molekül durch ein Lösemittel (in unserem Fall Wasser) gelöst werden kann. Die Löslichkeit ist abhängig von der Temperatur.
Dichte in mg/ml	Anzahl der in einem Volumen befindlichen Teilchen dividiert durch das Volumen.
Sättigungsdampfdruck in mmHg	Druck, bei dem der gasförmige Aggregatzustand sich mit dem flüssigen oder festen Aggregatzustand im Gleichgewicht befindet.
Henry-Law-Konstante	Das Henry-Gesetz besagt, dass der Druck eines Gases über einer Flüssigkeit direkt proportional zur Konzentration des Gases in der Flüssigkeit ist. Die Proportionalität wird ausgedrückt durch die Henry-Konstante.
Polarisierbarkeit in cm ³	Mass, um einen Dipolmoment mit einem elektrischen Feld zu induzieren. Je höher die Polarisierbarkeit ist, desto leichter lässt sich ein Dipolmoment erzeugen.
Funktionelle Gruppen	Jedes Molekül besitzt eine bis mehrere funktionellen Gruppen.
Anzahl funktionelle Gruppen	Anzahl unterschiedliche funktionelle Gruppen, denen ein Molekül zugeordnet werden kann.
Anzahl cyclische Verbindungen	Organische Moleküle kann man weiter unterteilen in acyclische und cyclische Verbindungen. Eine cyclische Verbindung weist keine Verbindung zu einem Ring auf. Die Verbindungen können gerade oder verzweigt sein. Bei einer cyclischen Verbindungen, weisen die organischen Moleküle eine Verbindung zu einem Ring auf.

Anhand der weiteren Analysen wurden die eingereichten Daten mit den Molekulareigenschaften anhand der CAS-Nummer der Moleküle verknüpft.

4 Methoden und Resultate

4.1 Normalisierung

Normalisierung resp. Standardisierung bedeutet, Werte zu transformieren, sodass sie gewisse Eigenschaften besitzen. Zum Beispiel werden Werte in einer Liste mit Mittelwert 16.5 und Standardabweichung 7 so verändert, dass der Mittelwert neu 0 und die Standardabweichung 1 beträgt (Standardisierung). Innerhalb der Liste sind die Werte weiterhin mit gleicher relativer Differenz verteilt.

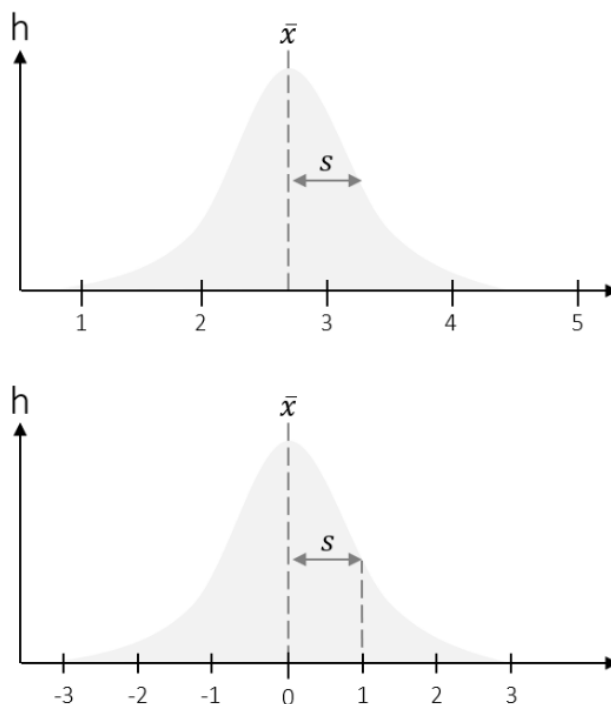


Abbildung 2: Beispiel Normalisierung

Für den VOC-BlackEagle Datensatz haben wir die «Counts», für den Chilling Compound Datensatz die «Areas» normalisiert.

Dafür haben wir folgenden Code verwendet:

VOC-BlackEagle Datensatz:

```
# Normalize all counts of the same molecule (based on the CAS number)
results["Norm_Counts"] = results.groupby(
    'CAS')['Counts'].transform(lambda x: (x - x.mean()) / x.std())
```

Chilling Compound Datensatz:

```
# Creating list of variabel names
absolute_name = ["Reference", "IR 5g", "IR 12g", "IR 20g", "Frozen Spout"]
relative_name = ["rel. Reference%", "rel. IR 5g%", "rel. IR 12g%", "rel. IR 20g%", "rel. Frozen Spout%"]

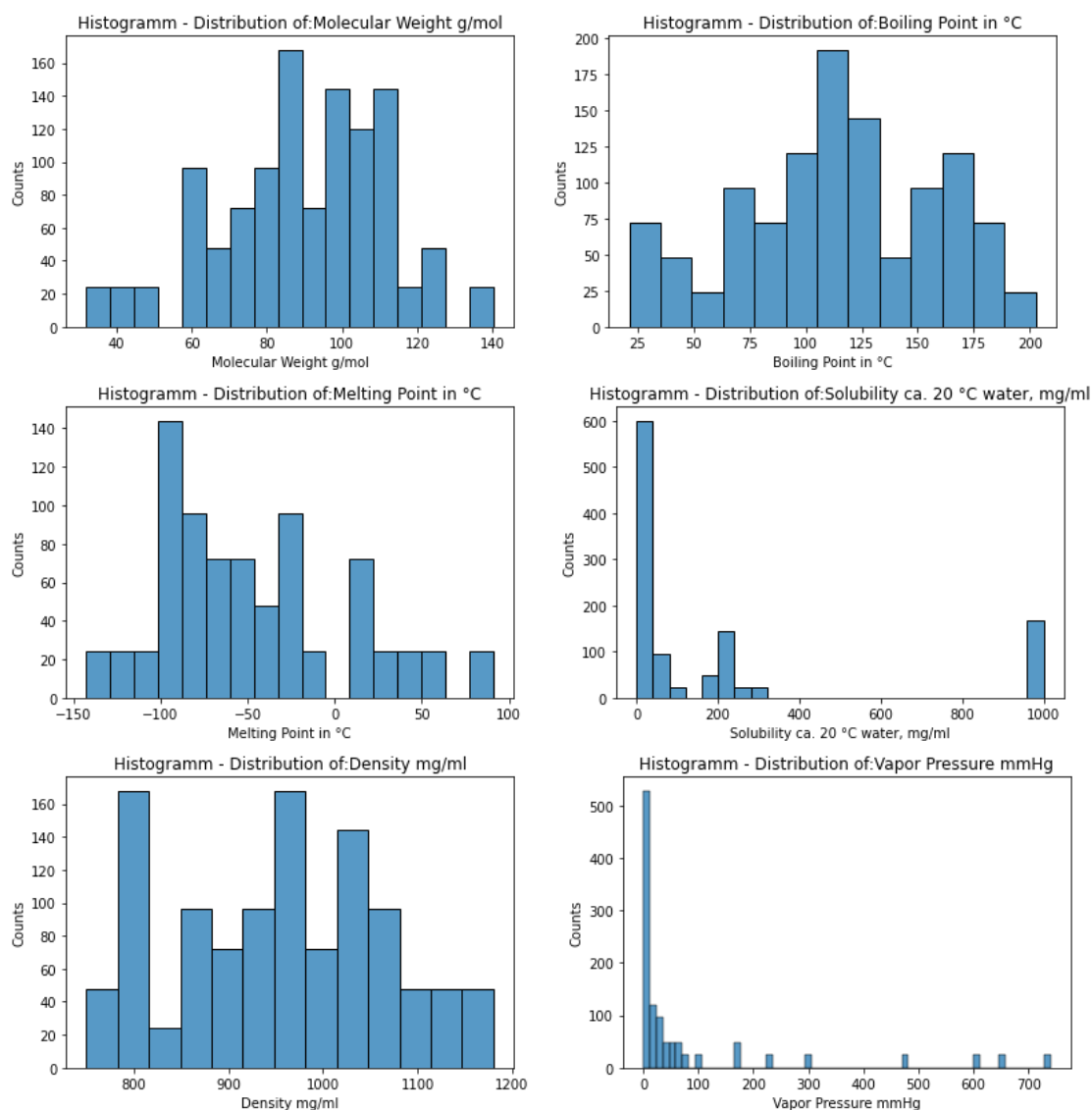
# Using for loops, to create for each molecule a relative value
for x, y in zip(absolute_name, relative_name):
    df_eugi_hsgc[y] = df_eugi_hsgc[x] / df_eugi_hsgc["Reference"] * 100
    df_arab_hsgc[y] = df_arab_hsgc[x] / df_arab_hsgc["Reference"] * 100
```

4.2 Explorative Datenanalyse

Die explorative Datenanalyse ist Ausgangspunkt jeder Analyse. Sie hilft die Qualität der Daten zu überprüfen, garantiert, dass die Daten zur Beantwortung der Fragestellung geeignet sind, erlaubt eine Plausibilisierung von Hypothesen und ermöglicht es, Wissen und Informationen aus Datenbeständen zu extrahieren.

4.2.1 Molekulareigenschaften der flüchtigen Moleküle des VOC-BlackEagle Datensatzes

Für das erstellte Molekulareigenschaften-Dataframe vom Datensatz VOC-BlackEagle erstellen wir Histogramme, um eine Gefühl für die Häufigkeitsverteilung der relevanten, kardinal skalierten Molekulareigenschaften (siehe Tabelle 1) zu erhalten.



4 Methoden und Resultate

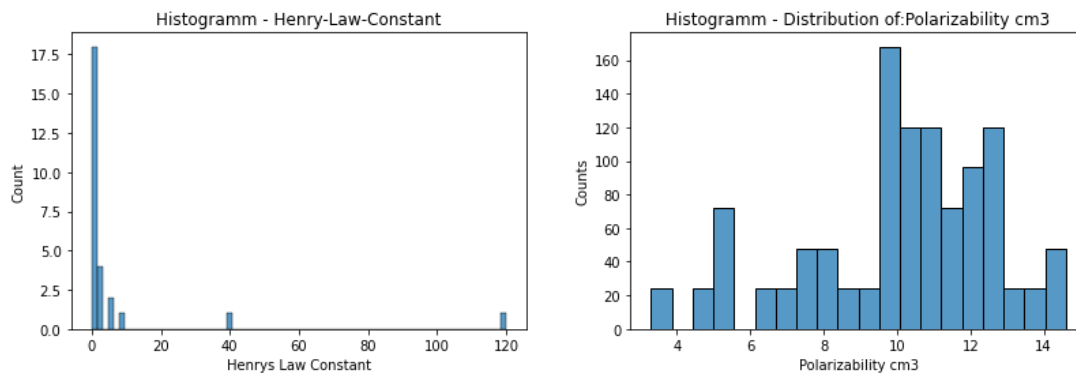


Abbildung 3: Histogramme relevanter, kardinal skaliertter Molekulareigenschaften

Als mögliche Klassifizierungen legten wir – neben den zentralen Molekulareigenschaften – folgende Kategorisierungen fest:

- Einteilung der Moleküle in funktionelle Gruppen
- Anzahl cyclische Verbindungen

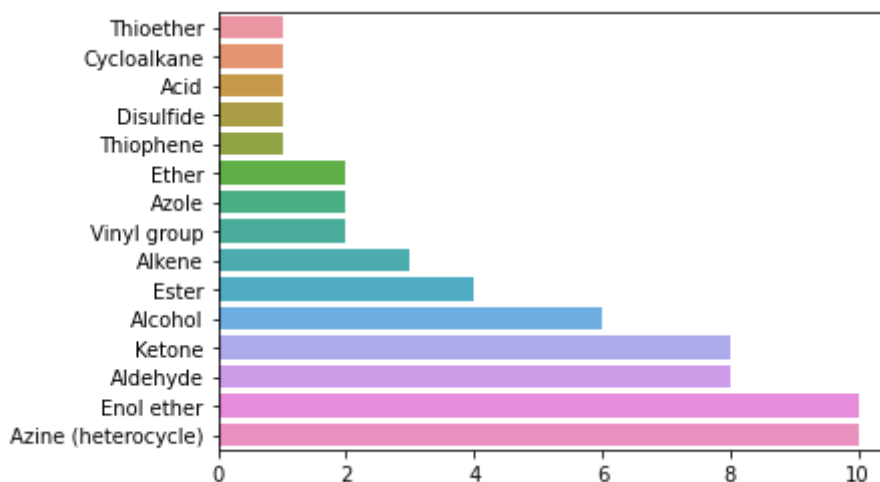


Abbildung 4: Einteilung in funktionelle Gruppen

Bei der Einteilung in funktionelle Gruppen fällt auf, dass über die Hälfte der Gruppen weniger als vier Moleküle beinhaltet. Eine Analyse der Anzahl funktioneller Gruppen pro Molekül ergibt, dass viele Moleküle zwei oder drei Gruppen zugeordnet werden können.

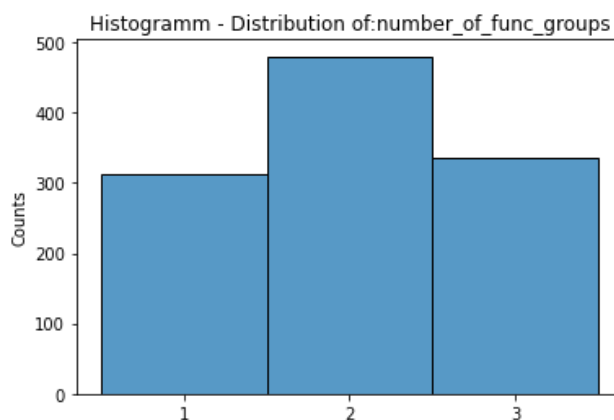


Abbildung 5: Häufigkeitsverteilung der funktionellen Gruppen

4 Methoden und Resultate

Das Histogramm von cyclischen Verbindungen zeigt, dass die Verteilung ausgeglichen ist.

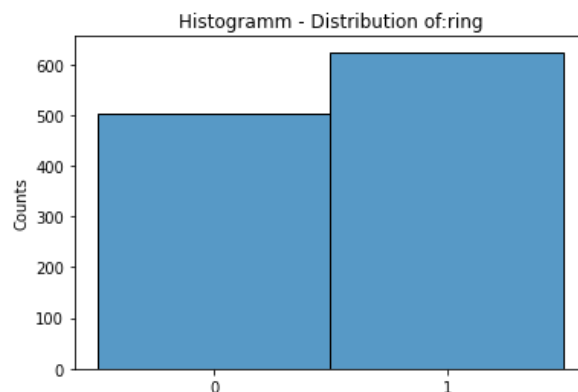


Abbildung 6: Anzahl cyclische Verbindungen

4.2.2 VOC-BlackEagle

4.2.2.1 Verteilung der «Counts»

Nach der Normalisierung der Daten sah das Histogramm der normalisierten «Counts» wie folgt aus:

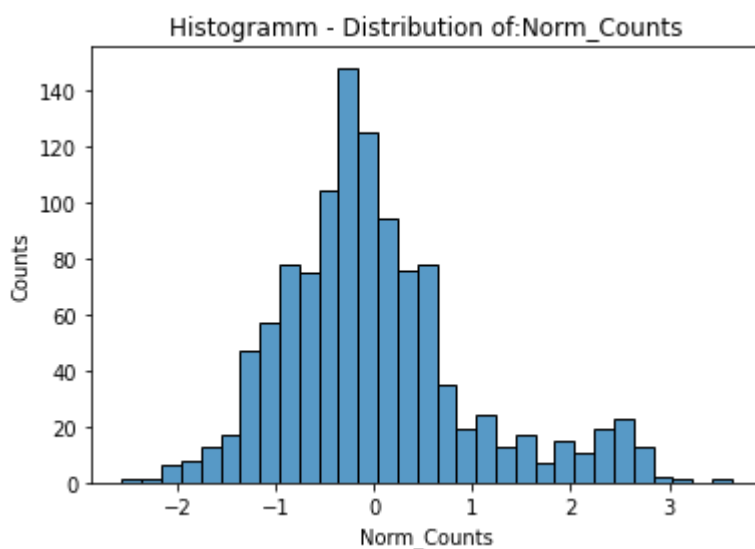


Abbildung 7: Histogramm normalisierte "Counts"

Als erstes wollten wir die Vermutung bestätigen, dass flüchtige Stoffe bei der Extraktion von Kaffee bei höheren Temperaturen eher verdampfen.

Dazu verglichen wir die Verteilung der normalisierten «Counts» bezüglich den drei unterschiedlichen Wassertemperaturen 90, 93 und 96 Grad. Als Referenz, also als 0 Wert, wurden die «Counts» von 93 Grad festgelegt.

Die nachfolgende Visualisierung zeigt drei zufällige Moleküle. Für alle anderen Moleküle sieht die Visualisierung vergleichbar aus.

4 Methoden und Resultate

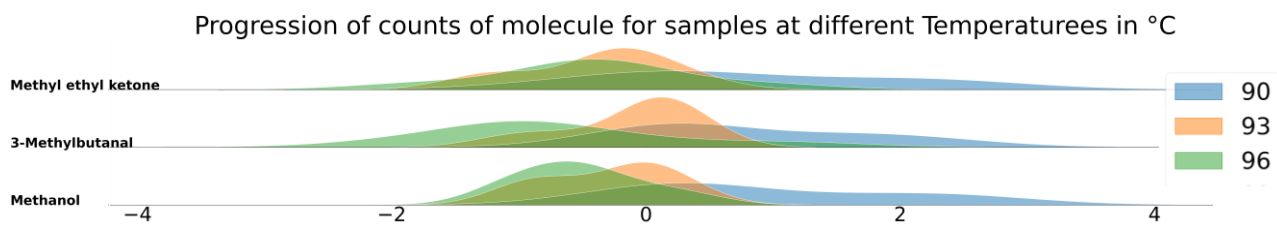


Abbildung 8: Dichteplots für drei zufällig ausgewählte Moleküle

Auffallend ist, dass die Kurven von 93 und 96 Grad deutlich steiler sind als diejenige von 90 Grad. Das bedeutet, dass die Mengen der Moleküle der acht Experimente näher beieinander liegen. Bei 90 Grad ist die Kurve flach und breit. Diese acht Experimente streuen mehr bezüglich vorhandener Menge im Kaffee als diejenigen bei wärmeren Temperaturen.

Zusätzlich ist erkennbar, dass die Verteilung von 90 Grad deutlich weiter rechts ausgerichtet ist als die Verteilung von 96 Grad. Weiter rechts besagt, es sind mehr flüchtige Moleküle vorhanden. Damit kann die Vermutung bereits unterstützt werden.

Eine zweite Darstellung zeigt dieselbe Analyse, für wiederum drei zufällige Moleküle, mit Boxplots.

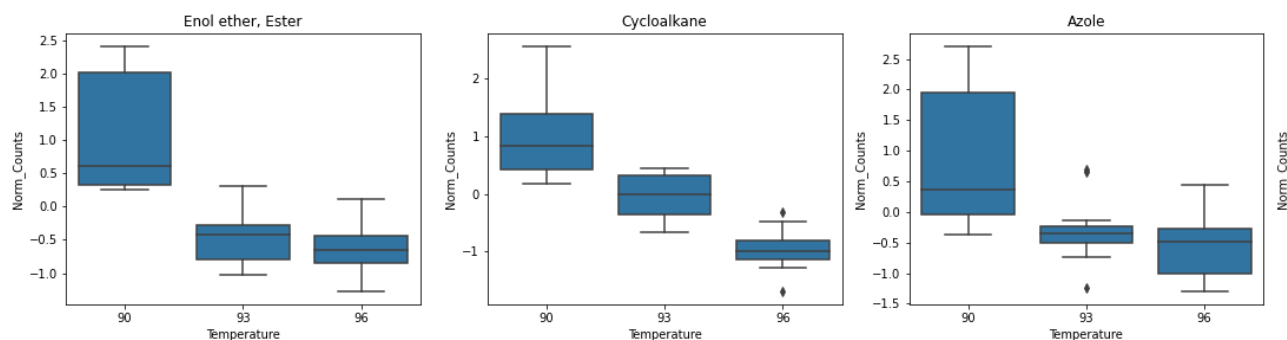


Abbildung 9: Boxplots für drei zufällig ausgewählte Moleküle

Mit einem Boxplot können die wichtigsten robusten Lage- und Streuungsmasse (Minimum, Maximum, unteres und oberes Quartil und Median) dargestellt werden. Punkte ausserhalb der «Whisker» stellen Ausreisser in den Daten, dar. Auch in dieser Darstellung ist das oben beschriebenen Verhalten beobachtbar.

4.2.2.2 Verteilung der «Counts» nach cyclischen Verbindungen und Gruppen

In einer weiteren Analyse wurde die Verteilung nach cyclischen Verbindungen untersucht. Es gibt kein Molekül, das mehr als eine cyclische Verbindung enthält.

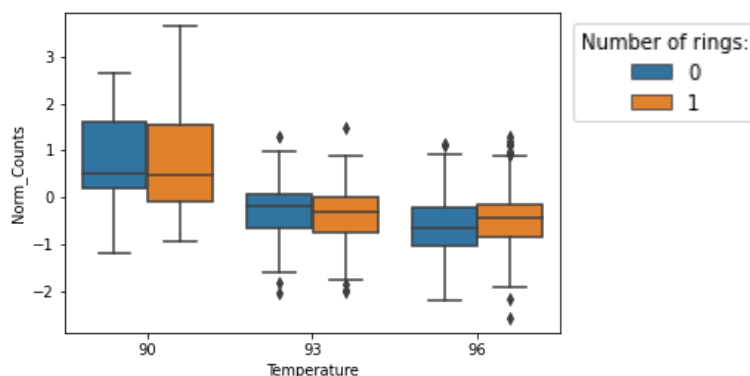


Abbildung 10: Boxplots zur Analyse der Verteilung der Moleküle mit und ohne cyclischen Ring

Bezüglich Verteilung gibt es kaum Unterschiede. Nur bei den Molekülen mit Ringen gibt es einige Ausreisser bei 96 Grad mehr.

Auch die Untersuchung der Verteilung nach funktionellen Gruppen brachte kaum weitere Erkenntnis.

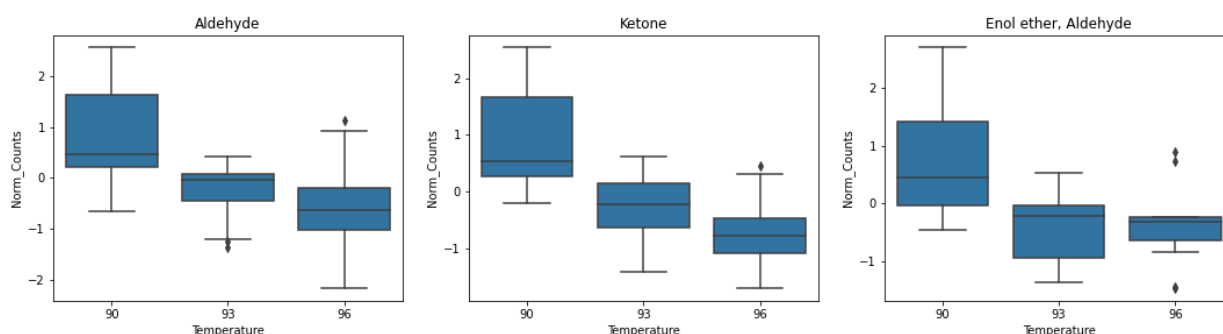


Abbildung 11: Verteilung der "Counts" nach drei zufällig ausgewählten funktionellen Gruppen

Es konnte nur erneut bestätigt werden, dass bei Kaffee, der mit 90 Grad extrahiert wurde, mehr flüchtige Moleküle enthalten sind.

4.2.2.3 Zusammenhang «Counts» und Molekulareigenschaften in Abhängigkeit der Wassertemperatur

Um die zweite Fragestellung des Projekts nach Zusammenhängen zwischen der Menge der flüchtigen Moleküle Molekulareigenschaften und Wassertemperatur zu untersuchen, wurden für alle relevanten Molekulareigenschaften (siehe Tabelle 1) und den normalisierten «Counts», den Mittelwerten sowie den Medianwerten der normalisierten «Counts» untersucht.

Nachfolgende Visualisierung zeigt die Gegenüberstellung des Medians der normalisierten «Counts» und der Molekülmasse in Abhängigkeit der Wassertemperatur. Dieselben Visualisierungen, aber für die drei Molekulareigenschaften Löslichkeit, Polarisierbarkeit und Henry-Law-Konstante befinden sich im Anhang dieses Dokumentes. Die Visualisierungen sind vergleichbar und die Interpretation davon ist identisch mit derjenigen der Molekülmasse.

4 Methoden und Resultate

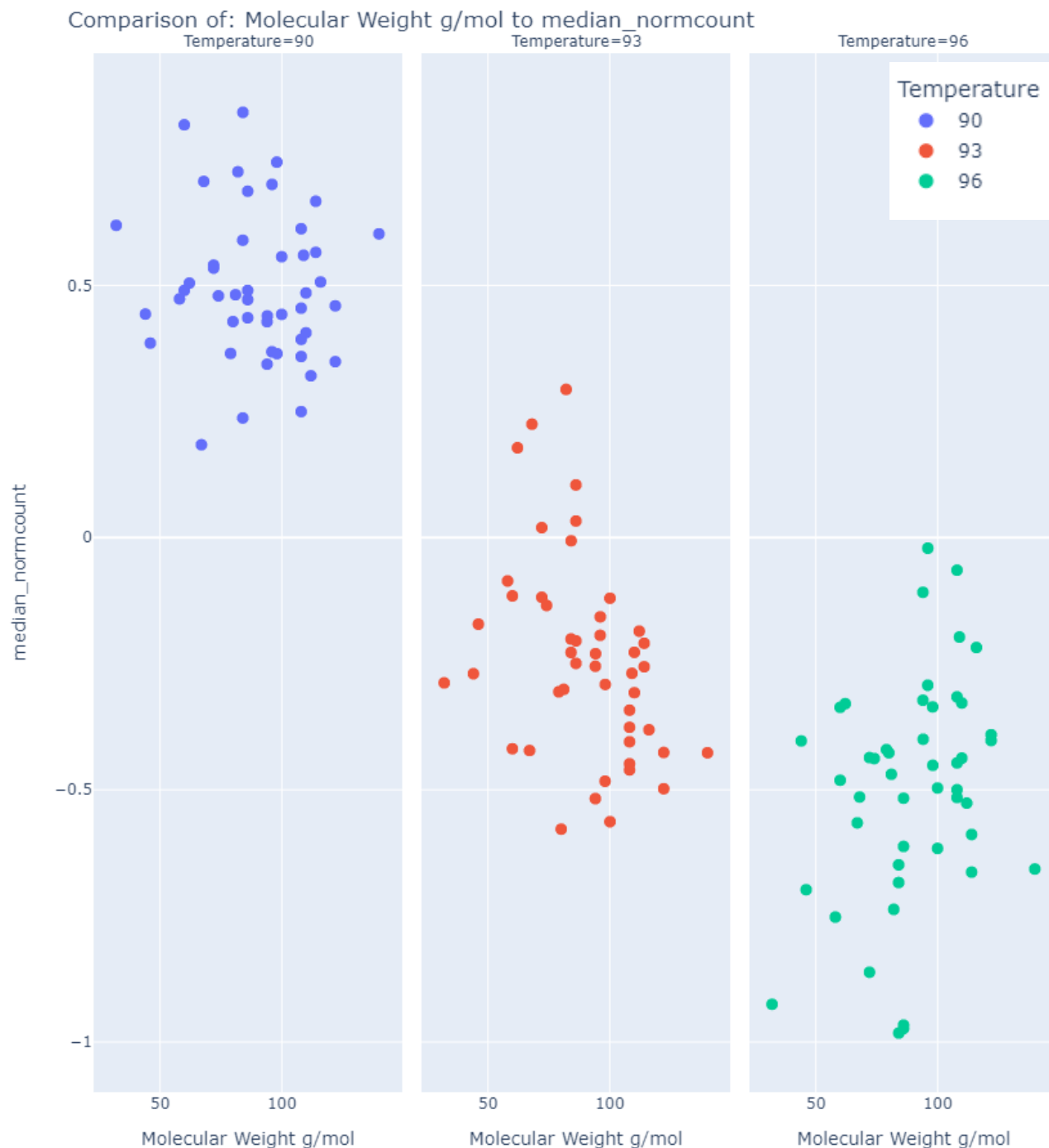


Abbildung 12: Zusammenhang normalisierte "Counts" und Molekülmasse

Die Punktwolke visualisiert wiederum das erhöhte Vorkommen flüchtiger Moleküle bei kühlerer Wassertemperatur. Die Streuung der Punkte lässt jedoch keinen Schluss auf einen Zusammenhang zwischen der Molekülmasse und der Menge zu.

Die Visualisierungen mit allen relevanten Molekulareigenschaften (siehe Tabelle 1) können mit dem Python Script «Explorative_Datenanalyse.py» unter dem Pfad Github\fhmw-pro2d\Scripts\Initial_Dataset generiert werden. Diese werden dann als HTML-Files im Unterordner «plots» abgelegt. Werden diese im Browser geöffnet sind sie interaktiv. Wir haben dafür die Programmbibliothek «plotly» (<https://plotly.com/python/line-and-scatter/>) verwendet.

4.2.3 Chilling Compound

Für das erstellte Molekulareigenschaften-Dataframe vom Datensatz Chilling Compound haben wir auf eine explorative Datenanalyse verzichtet. Die flüchtigen Moleküle unterscheiden sich nur in ein paar weniger Fällen von denjenigen des VOC-BlackEagle Datensatzes.

Wir starteten mit der Normalisierung der Werte der Mengen der flüchtigen Moleküle im Kaffee, nachfolgend «Area» genannt. Als Referenz-«Area» für ungekühlten Kaffee wurde der Wert 100 festgelegt.

Nach der Normalisierung der Daten sahen die Histogramme der normalisierten «Areas» für die Kaffeesorten Arab und Eugi wie folgt aus:

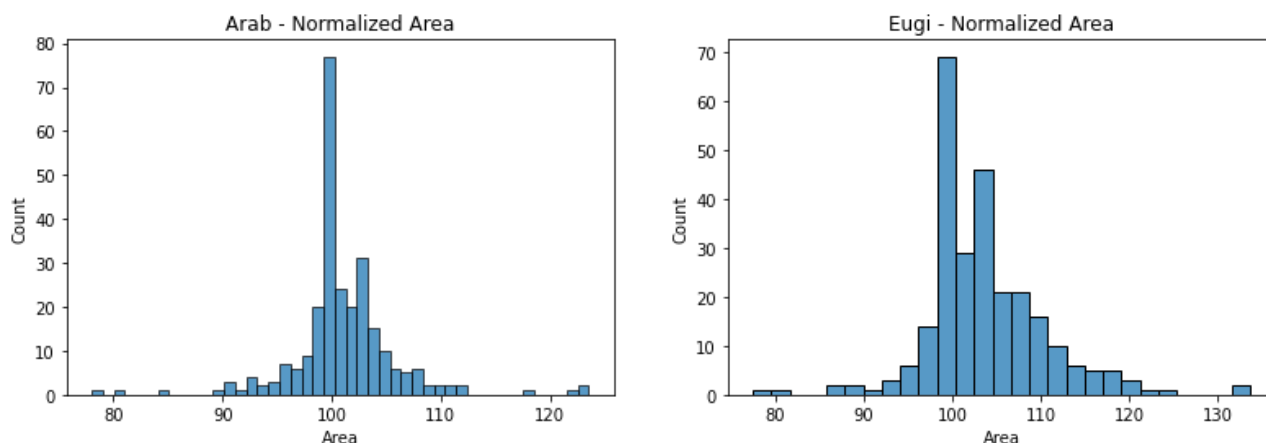


Abbildung 13: Histogramme der normalisierten Werte für "Area"

Wir visualisierten die Menge der flüchtigen Moleküle für beide Kaffeesorten gegenüber allen relevanten Molekulareigenschaften (siehe Tabelle 1) mit Streudiagrammen. Die unterschiedlichen Punktfarben stellen den Referenzkaffee (blau) und die unterschiedlichen Kühlungen dar (5g, 12g, 20g, Spout).

Während die aktive Kühlung mit einem Eiswürfel eine kontinuierliche Steigerung darstellt (5g → 12g → 20g) kann die Spout-Kühlung nicht hierarchisch eingeordnet werden.

In den nachfolgenden vier Kapiteln wird auf die Zusammenhänge zwischen Menge von flüchtigen Molekülen in Abhängigkeit der Kühlmethode basierend auf den vier zentralen Molekulareigenschaften Molekülmasse, Löslichkeit, Polarisierbarkeit und Henry-Law-Konstante näher eingegangen.

Die Visualisierungen mit allen relevanten Molekulareigenschaften (siehe Tabelle 1) können mit dem Python Script «EDA_Chilling_Comp_HSGC.py» unter dem Pfad Github\fhmw-pro2d\Scripts\Chilling_Compound\Explorative_Datenanalyse generiert werden. Diese werden direkt in diesem Ordner abgelegt.

4 Methoden und Resultate

4.2.3.1 Molekülmasse

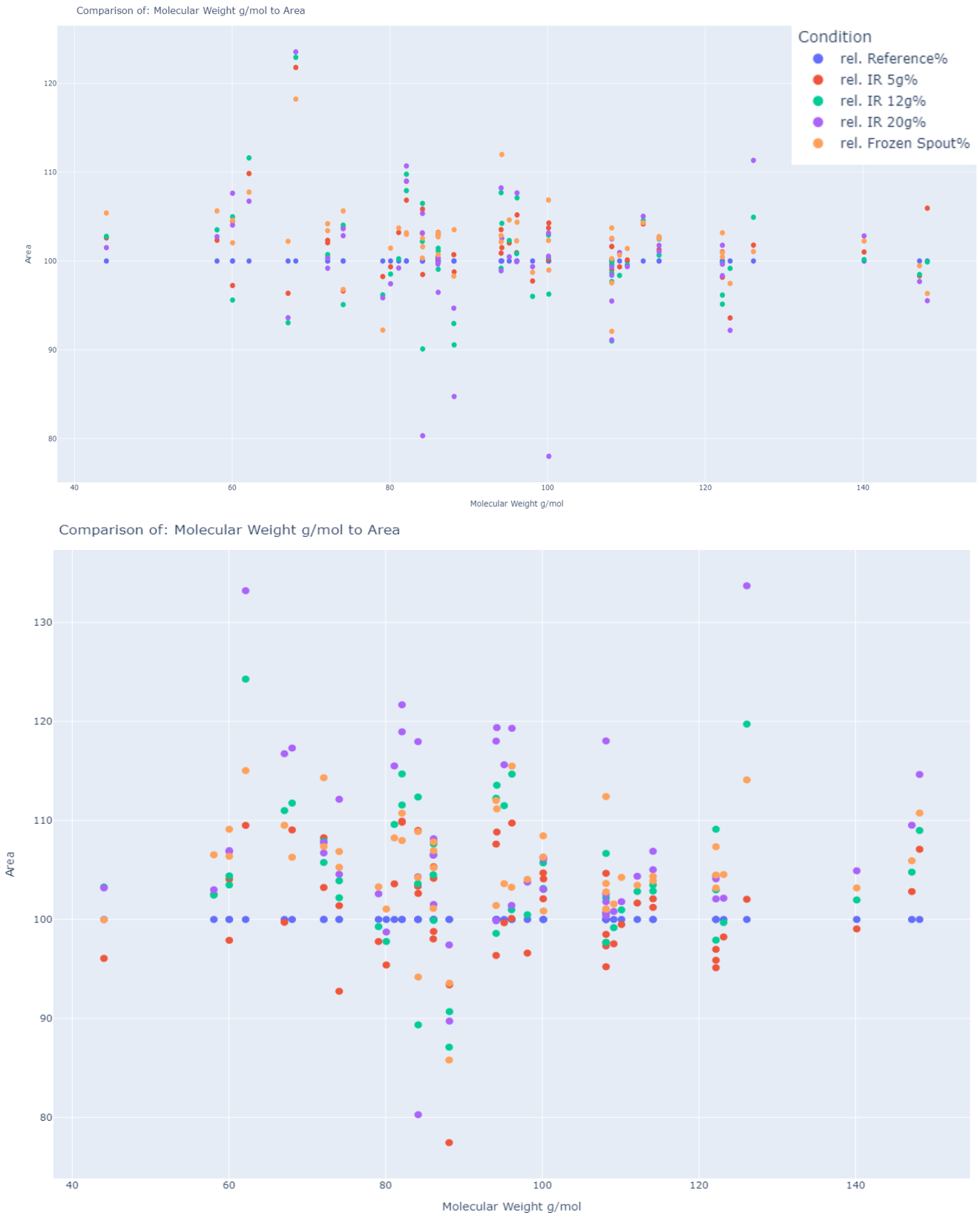


Abbildung 14: Zusammenhang normalisierte "Areas" und Molekülmasse für Kaffeesorten Arab (oben) und Eugi (unten)

4 Methoden und Resultate

Beide Punktwolken visualisieren in der Tendenz ein erhöhtes Vorkommen flüchtiger Moleküle bei aktiver Kühlung des Kaffees. Bei gewissen Molekülen weicht diese Tendenz ab und sie kommen sogar weniger oft vor. Eine Vermutung war, dass mehr aktiv gekühlter Kaffee dazu führt, dass die Menge dieser Moleküle sich der Menge im Referenzkaffee wieder angleichen. Diese Vermutung konnte weder gestützt noch verworfen werden, da sie nur teilweise zutraf. Vereinzelt nahm die Menge sogar mit mehr aktiver Kühlung ab.

Bei der Kaffeesorte Arab (obere Abbildung) ist die Tendenz von mehr flüchtigen Molekülen vor allem bei tieferer Molekülmasse (< 100 g/mol) sichtbar, gleicht sich aber mit höherer Molekülmasse (> 100 g/mol) der Menge im ungekühlten Referenzkaffee an. Es ist ein leichter Abwärtstrend erkennbar.

Bei der Kaffeesorte Eugi (untere Abbildung) ist der Abwärtstrend weniger stark. Die Grenze um 100 g/mol ist aber auch hier sichtbar. Die Menge von Molekülen mit Massen über 100 g/mol verhalten sich bei Eugi aber stabiler als bei Arab. Der Unterschied zu ungekühltem Referenzkaffee ist weniger deutlich als bei Molekülen mit Massen unter 100 g/mol. Das Vorkommen nimmt mit höherer Molekülmasse aber nicht mehr weiter ab, sondern bleibt relativ konstant.

4 Methoden und Resultate

4.2.3.2 Löslichkeit

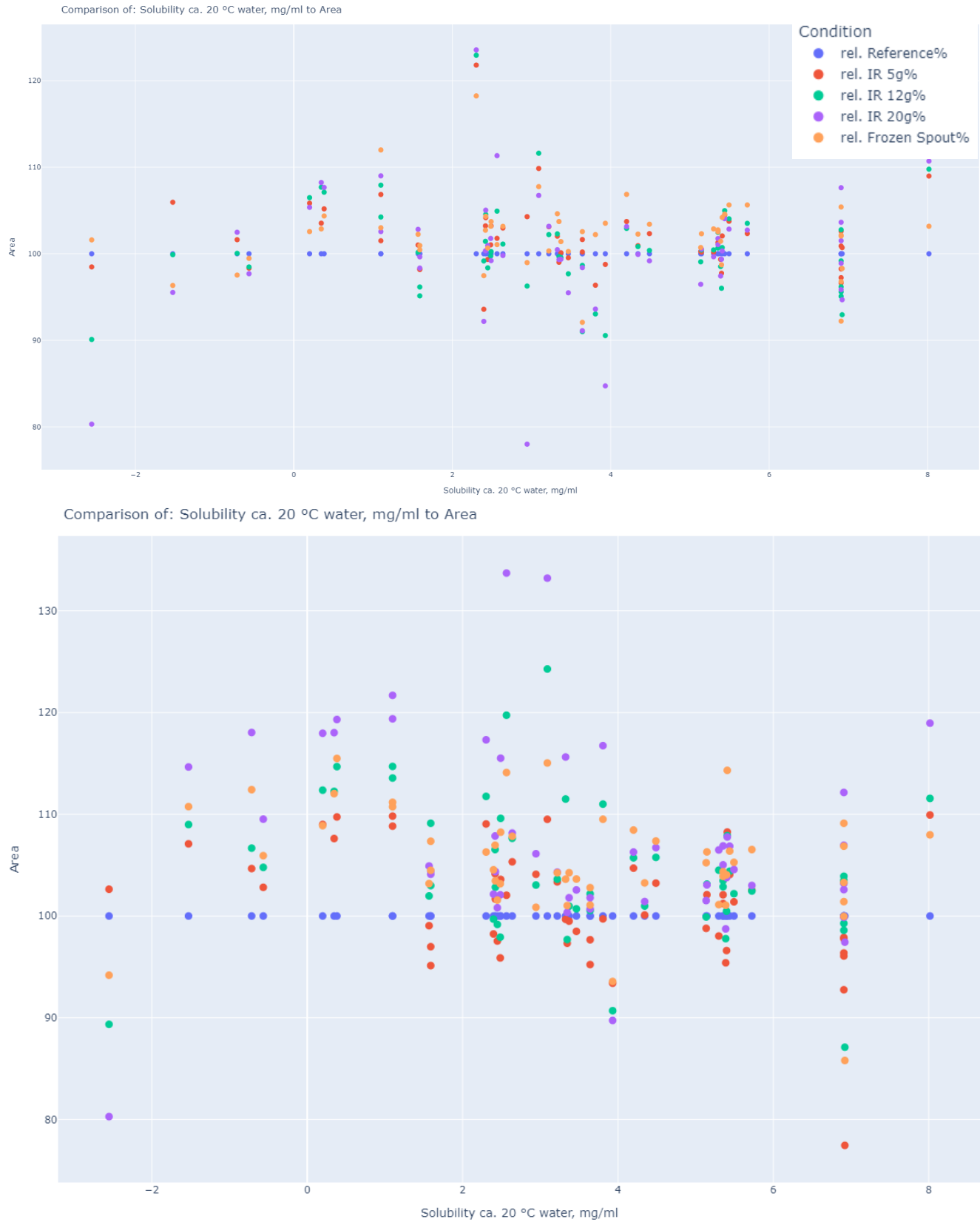


Abbildung 15: Zusammenhang normalisierte "Areas" und Löslichkeit für Kaffeesorten Arab (oben) und Eugi (unten)

4 Methoden und Resultate

Beide Punktwolken visualisieren in der Tendenz ein erhöhtes Vorkommen flüchtiger Moleküle bei aktiver Kühlung des Kaffees. Diese Tendenz ist bei der Kaffeesorte Arab (obere Abbildung) weniger ausgeprägt als bei Eugi (untere Abbildung). Bei beiden Kaffeesorten ist im unteren 2/3-Bereich eine, nach unten geöffnete, parabelförmiger Kurve erkennbar. Im oberen Drittel ist die Tendenz dann wieder eher steigend, wobei es in diesem Bereich wenig Werte gibt und diese stark streuen.

Die Werte auf der X-Achse der oberen beiden Abbildung zeigen logarithmierte Werte der Löslichkeit. Die Logarithmus-Transformation wurde mit folgendem Code durchgeführt:

```
df_chemical_data['Solubility ca. 20 °C water, mg/ml'] =  
np.log(df_chemical_data['Solubility ca. 20 °C water, mg/ml'].astype(float))
```

Logarithmus-Transformationen sind hilfreich, wenn der Wertebereich der dargestellten Daten gross ist. Mittels logarithmierter Darstellung werden Zusammenhänge im Bereich der kleinen Werte besser erkennbar. Die Logarithmus-Transformation wurde neben der Löslichkeit auch für die Henry-Law-Konstante und den Sättigungsdampfdruck durchgeführt.

Vor der Logarithmus-Transformation hatte die Visualisierung für die Kaffeesorte Arab wie folgt ausgesehen. Für die Kaffeesorte Eugi ist die Visualisierung vergleichbar:

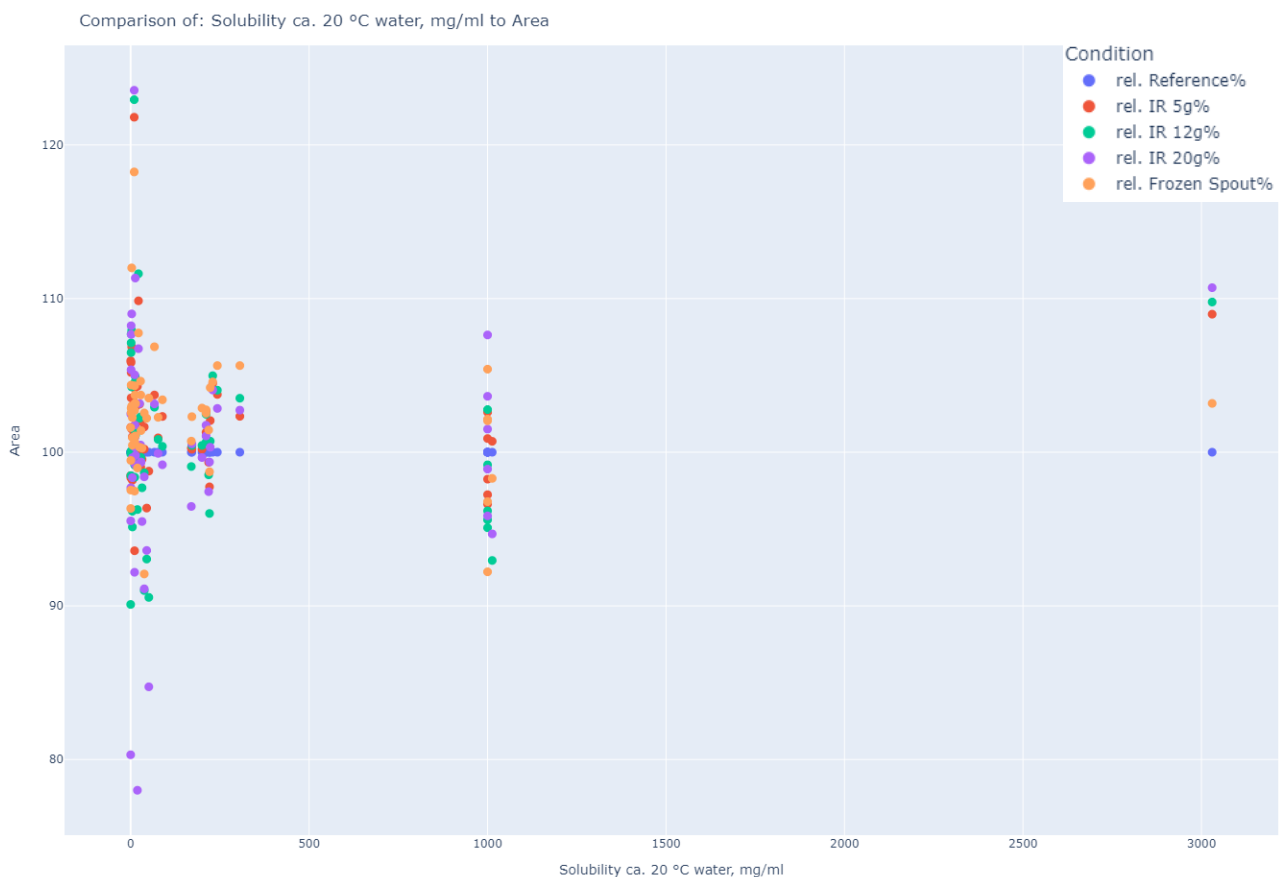


Abbildung 16: Zusammenhang normalisierte "Areas" und Löslichkeit für Kaffeesorte Arab ohne Logarithmus-Transformation

4 Methoden und Resultate

4.2.3.3 Polarisierbarkeit



Abbildung 17: Zusammenhang normalisierte "Areas" und Polarisierbarkeit für Kaffeesorten Arab (oben) und Eugi (unten)

4 Methoden und Resultate

Die Punktwolke der Kaffeesorte Arab (obere Abbildung) zeigt ein ähnliches Vorkommen flüchtiger Moleküle bei aktiver Kühlung des Kaffees wie bei der Menge im Referenzkaffee. Die Streuung ist oberhalb und unterhalb der Referenzwerts 100 in etwa gleich. Mit zunehmender Polarisierbarkeit ist ein minimaler Abwärtstrend im Vorkommen erkennbar.

Die Kaffeesorte Eugi (untere Abbildung) zeigt wiederum deutlich ein erhöhtes Vorkommen flüchtiger Moleküle. Die Kurve gleicht einer flachen, nach unten geöffneten, Parabel im Wertebereich.

4 Methoden und Resultate

4.2.3.4 Henry-Law-Konstante

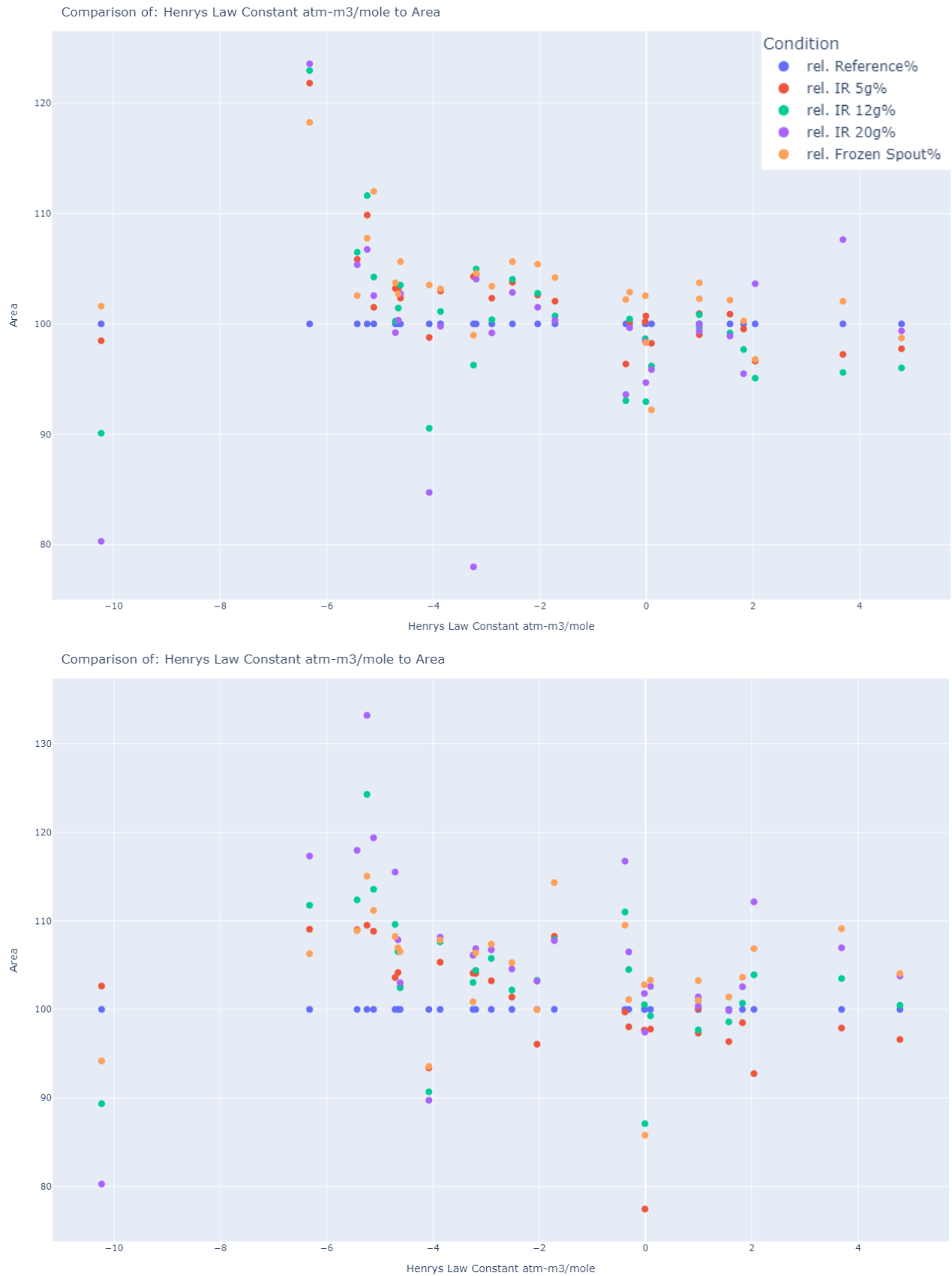


Abbildung 18: Zusammenhang normalisierte "Areas" und Henry-Law-Konstante für Kaffeesorten Arab (oben) und Eugi (unten)

4 Methoden und Resultate

Beide Punktwolken visualisieren im mittleren Bereich ein erhöhtes Vorkommen flüchtiger Moleküle bei aktiver Kühlung des Kaffees. Gegen Ende des Wertebereichs gleicht sich die Menge derjenigen im Referenzkaffee an. Die Ausprägungen sind bei beiden Kaffeesorten vergleichbar. Ebenfalls ist bei beiden Kaffeesorten ein deutlicher Abwärtstrend zu erkennen.

Die Werte auf der X-Achse der oberen beiden Abbildung zeigen logarithmierte Werte der Henry-Law-Konstante.

4.2.3.5 Funktionelle Gruppen

In Kapitel 4.2.1 werden zwei von uns festgelegte Kategorisierungen beschrieben, unter anderem nach funktionellen Gruppen.

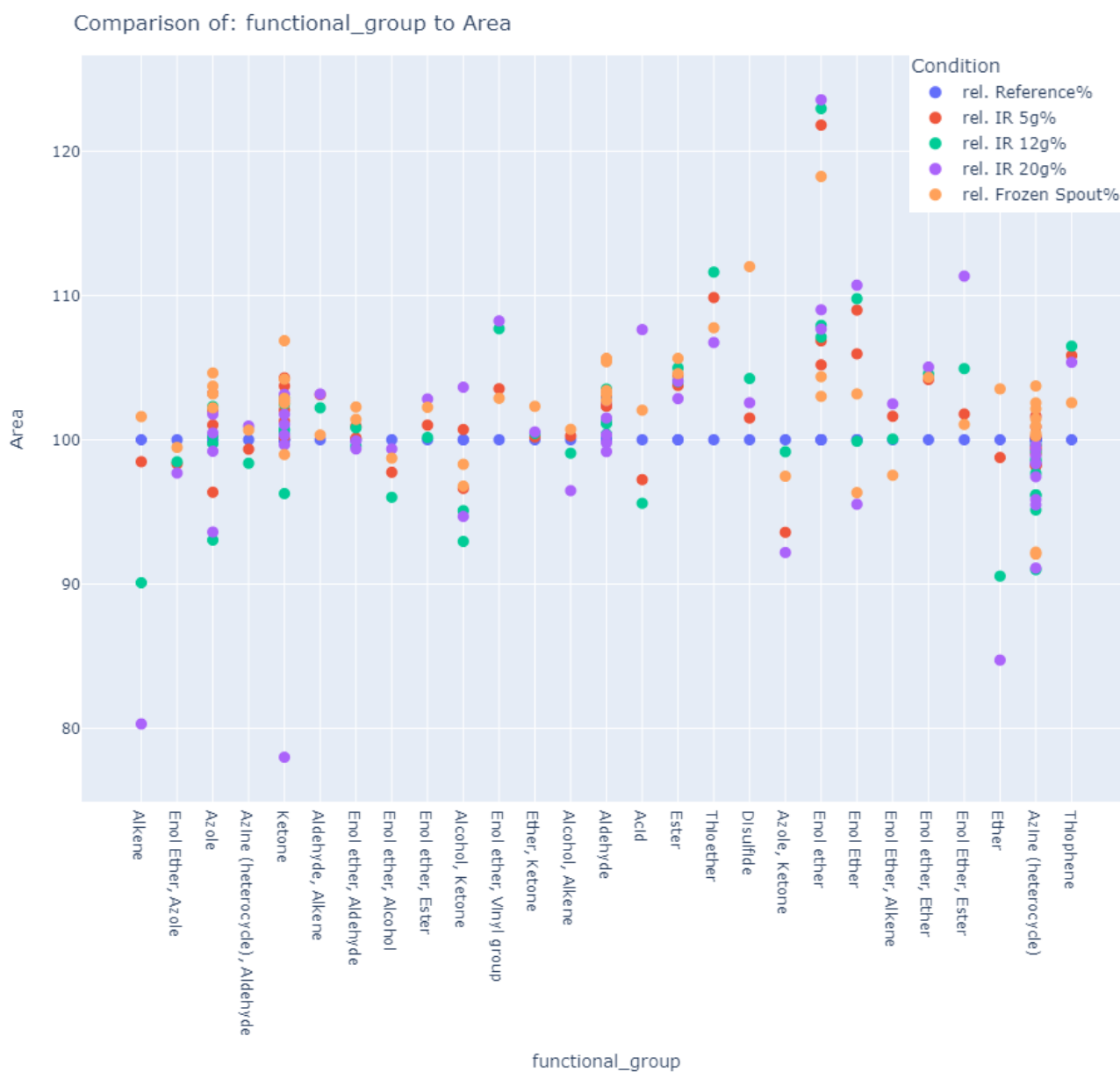


Abbildung 19: Zusammenhang normalisierte "Areas" und funktionelle Gruppen für Kaffeesorte Arab

4 Methoden und Resultate

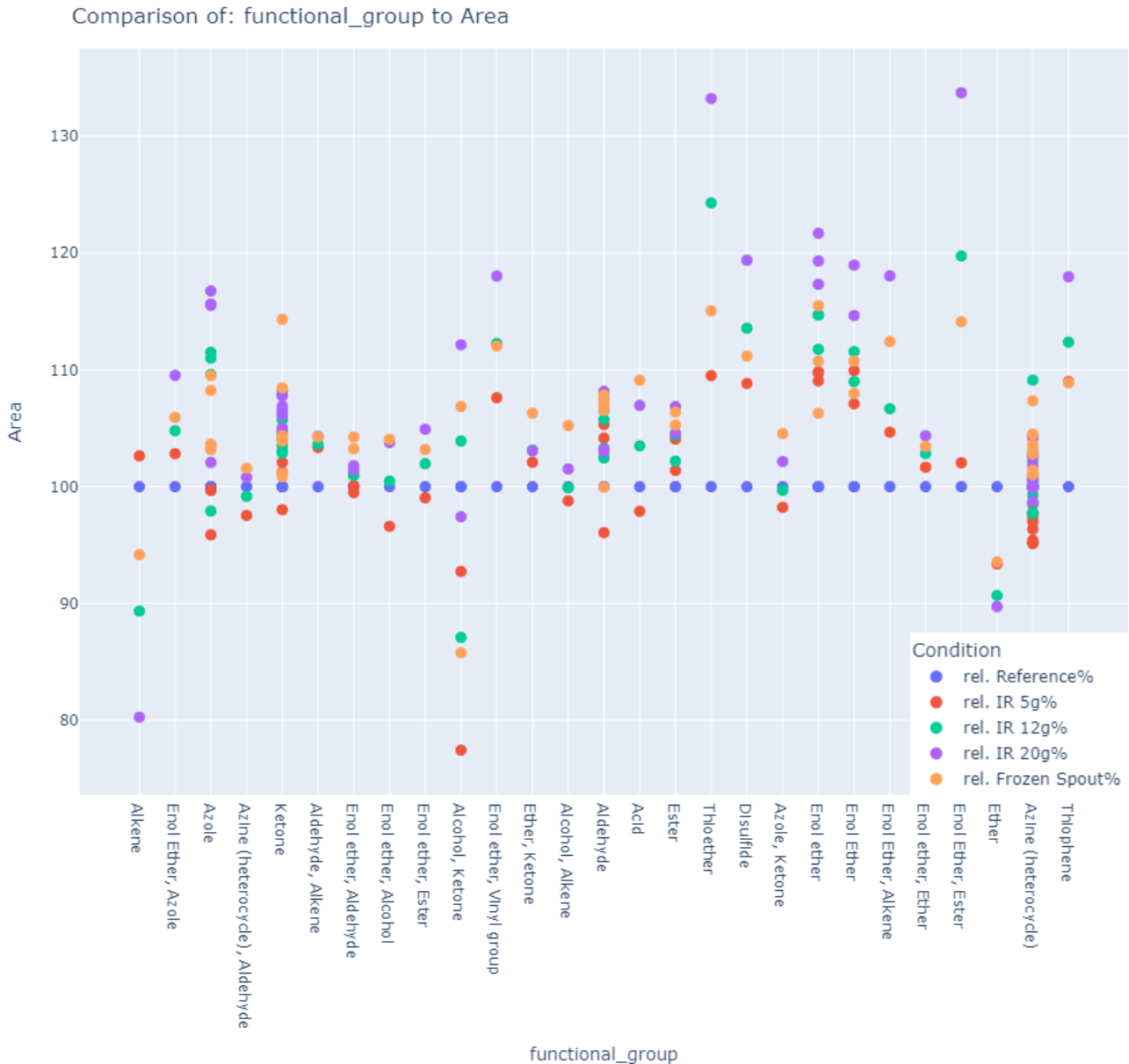


Abbildung 20: Zusammenhang normalisierte "Areas" und funktionelle Gruppen für Kaffeesorte Eugi

Die Menge an flüchtigen Molekülen in aktiv gekühltem Kaffee verhält sich je nach funktioneller Gruppe unterschiedlich. Auch sind Unterschiede zwischen den beiden Kaffeesorten erkennbar.

Bei der Sorte Arab weisen knapp 48 % der funktionellen Gruppen eine höhere Menge bei aktiv gekühltem Kaffee aus, etwa ein Drittel eine ähnliche Menge und ein Fünftel kommt weniger oft vor.

Bei der Sorte Eugi haben rund 60 % der Gruppen ein höheres Vorkommen, etwa ein Drittel ein ähnliches und nur gerade zwei Gruppen weisen deutlich niedrigere Werte aus.

4.2.3.6 Cyclische Verbindungen

Eine weitere in Kapitel 4.2.1 festgelegte Kategorisierung ist nach Moleküle mit und ohne cyclischen Verbindungen.

Bei der Sorte Arab ist eine leicht höhere Tendenz bei Molekülen ohne Ring mit der Spout-Kühlung erkennbar.

Die Sorte Eugi weist bei Molekülen mit einer cyclischen Verbindung bei 12g und 20g aktiv gekühltem Kaffee zwar eine viel grössere Streuung auf, im Mittel unterscheidet sich aber die Menge kaum von derjenigen, der Moleküle ohne cyclische Verbindung.

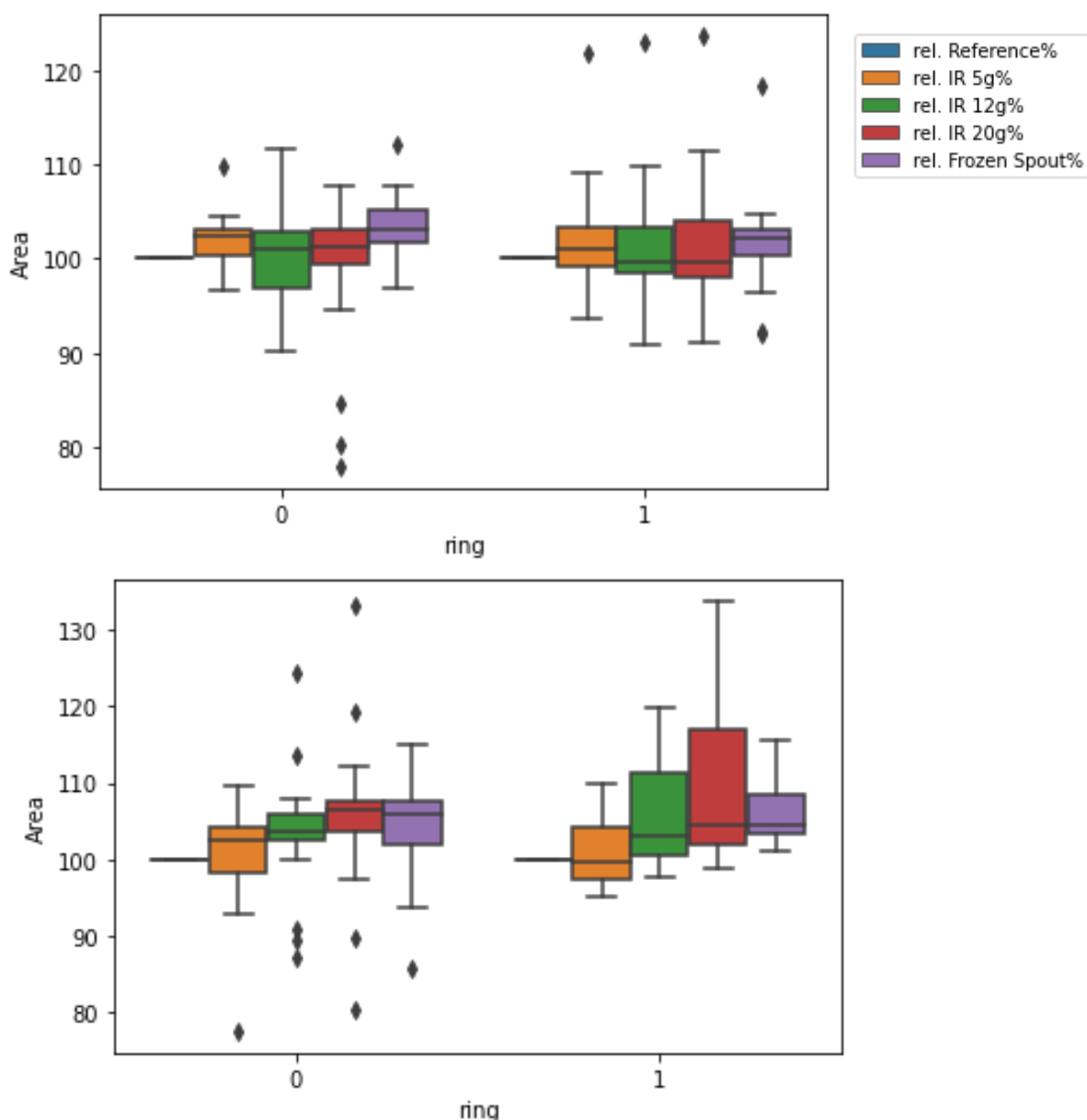


Abbildung 21: Zusammenhang normierte "Areas" und Moleküle mit und ohne cyclische Verbindungen für Kaffeesorten Arab (oben) und Eugi (unten)

4.3 Lineare Regression

Im Anschluss an die explorative Datenanalyse des Datensatzes Chilling Compound erstellten wir lineare Regressionen um die Zusammenhänge zwischen der Menge an flüchtigen Molekülen gegenüber den kardinal skalierten relevanten Molekulareigenschaften (siehe Tabelle 1) in Abhängigkeit der Kühlmethode visuell und numerisch sichtbar zu machen.

Die Visualisierungen und Berechnungen können mit dem Python Script «EDA_Chilling_Comp_HSGC.py» unter dem Pfad Github\fhnw-pro2d\Scripts\Chilling_Compound\Explorative_Datenanalyse generiert werden. Diese werden dann als HTML-Files im Unterordner «plots» abgelegt.

In diesem Kapitel werden nur die Visualisierungen und Ergebnisse der vier zentralen Molekulareigenschaften Molekülmasse, Löslichkeit, Polarisierbarkeit und Henry-Law-Konstante beschrieben.

Die Formel für die lineare Regression lautet:

$$Area_{[Kühlmethode]} = \alpha * [Molekulareigenschaft] + \beta$$

Für jede lineare Regression wurde zudem das Bestimmtheitsmass R^2 berechnet.

Das allerhöchste Bestimmtheitsmass über alle Molekulareigenschaften und Kühlmethoden konnte bei der Henry-Law-Konstante mit der Eiswürfel-Kühlmethode und 5g aktiv gekühltem Kaffee erzielt werden. R^2 liegt bei 0.305562, die Formel der linearen Regression lautet:

$$Area_{5g} = -1.08048 * Henry\ Law\ Constant\ atm\ m3/mole + 95.4392$$

4 Methoden und Resultate

4.3.1 Molekülmasse

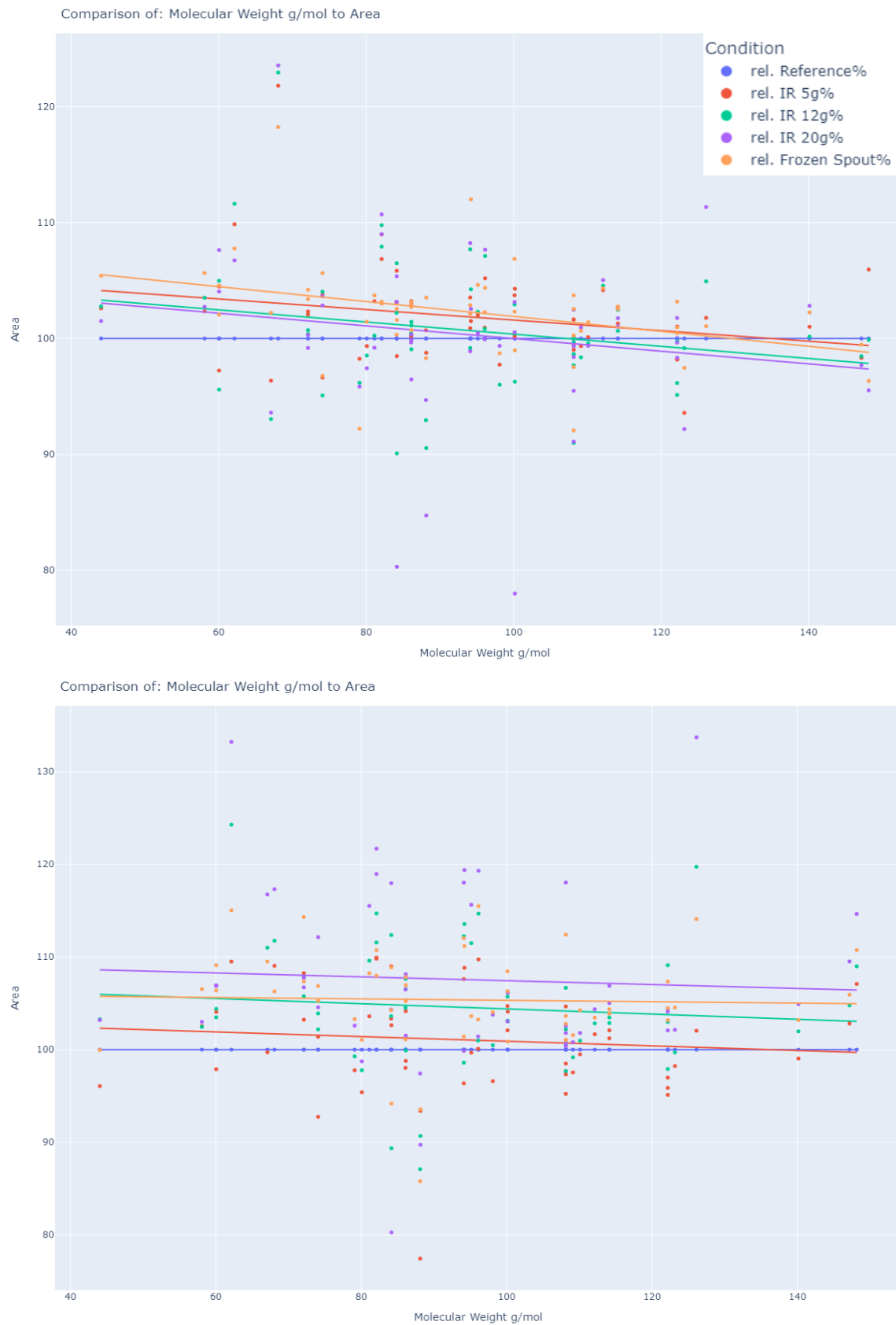


Abbildung 22: Lineare Regression für normalisierte "Areas" und Molekülmasse für Kaffeesorten Arab (oben) und Eugi (unten)

4 Methoden und Resultate

Für die Sorte Arab konnte das höchste Bestimmtheitsmass mit der Spout-Kühlmethode erzielt werden. R^2 liegt bei 0.121975, die Formel der linearen Regression lautet:

$$Area_{FrozenSpout} = -0.0642962 * Molecular Weight \text{ g/mol} + 108.332$$

Für die Sorte Eugi konnte das höchste Bestimmtheitsmass mit der Eiskwürfel-Kühlmethode und 5g aktiv gekühltem Kaffee erzielt werden. R^2 liegt bei 0.00962, die Formel der linearen Regression lautet:

$$Area_{5g} = -0.0249906 * Molecular Weight \text{ g/mol} + 103.409$$

4 Methoden und Resultate

4.3.2 Löslichkeit

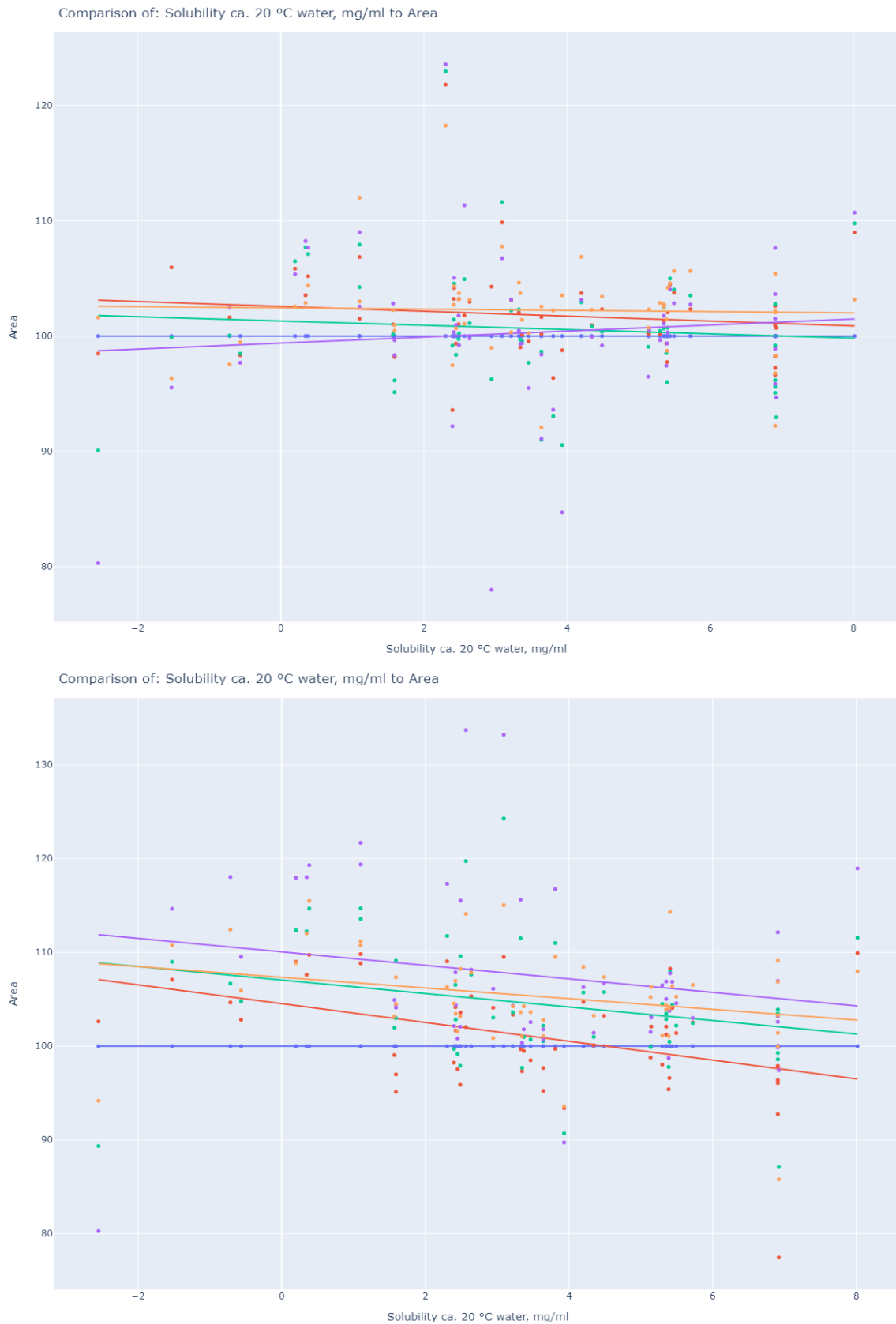


Abbildung 23: Lineare Regression für normalisierte "Areas" und Löslichkeit für Kaffeesorten Arab (oben) und Eugi (unten)

4 Methoden und Resultate

Für die Sorte Arab konnte das höchste Bestimmtheitsmass mit der Eiskwürfel-Kühlmethode und 20g aktiv gekühltem Kaffee erzielt werden. R^2 liegt bei 0.007463, die Formel der linearen Regression lautet:

$$Area_{20g} = -0.261159 * Solubility\ ca.\ 20\ ^\circ C\ water\ mg/mol + 99.3884$$

Für die Sorte Eugi konnte das höchste Bestimmtheitsmass mit der Eiskwürfel-Kühlmethode und 5g aktiv gekühltem Kaffee erzielt werden. R^2 liegt bei 0.170816, die Formel der linearen Regression lautet:

$$Area_{5g} = -1.00317 * Solubility\ ca.\ 20\ ^\circ C\ water\ mg/mol + 104.532$$

4 Methoden und Resultate

4.3.3 Polarisierbarkeit

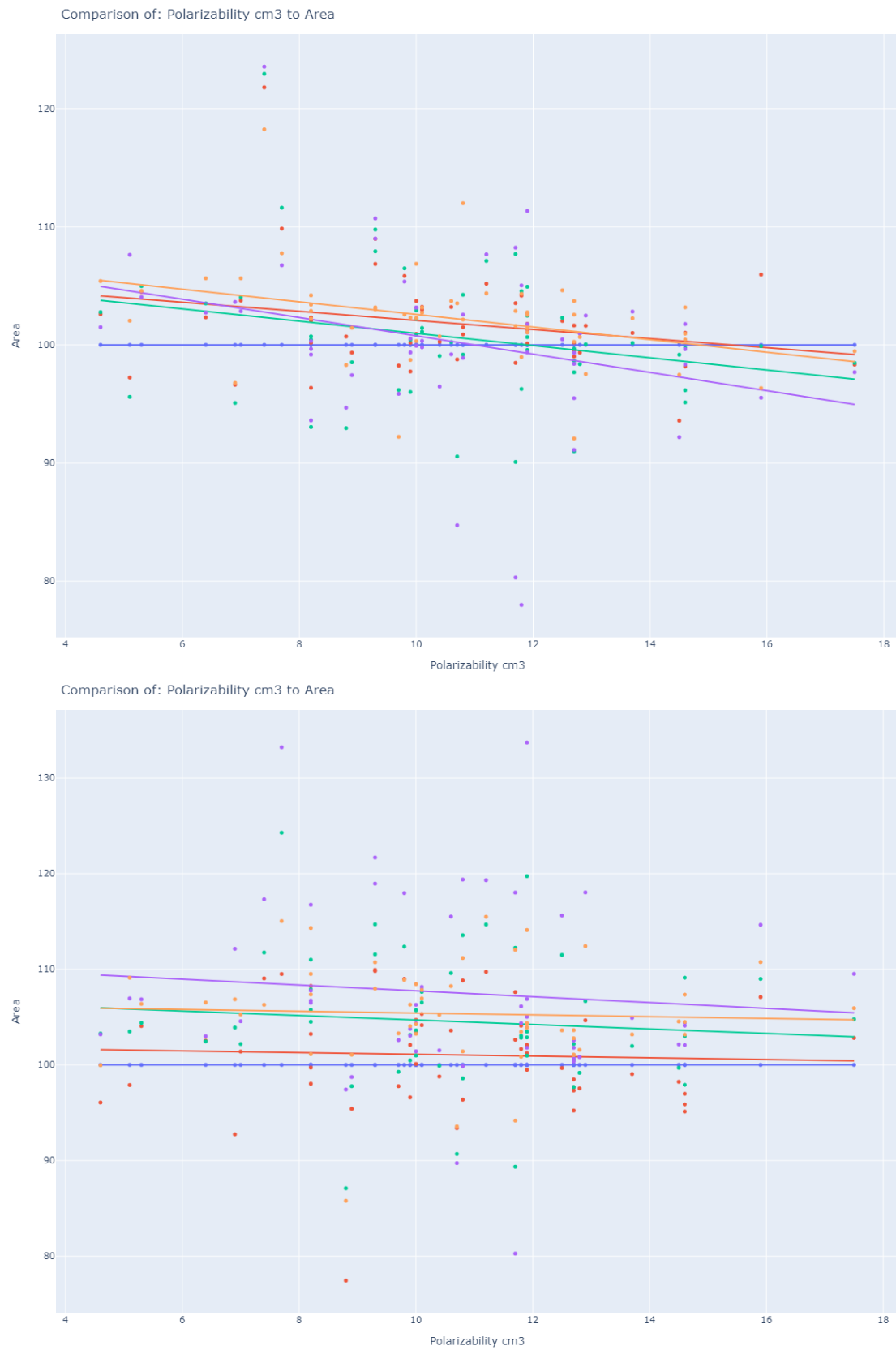


Abbildung 24: Lineare Regression für normalisierte "Areas" und Polarisierbarkeit für Kaffeesorten Arab (oben) und Eugi (unten)

4 Methoden und Resultate

Für die Sorte Arab konnte das höchste Bestimmtheitsmass mit der Eiskwürfel-Kühlmethode und 20g aktiv gekühltem Kaffee erzielt werden. R^2 liegt bei 0.129051, die Formel der linearen Regression lautet:

$$Area_{20g} = -0.533613 * Polarizability\ cm^3 + 107.918$$

Für die Sorte Eugi konnte das höchste Bestimmtheitsmass mit der Eiskwürfel-Kühlmethode und 5g aktiv gekühltem Kaffee erzielt werden. R^2 liegt bei 0.170816, die Formel der linearen Regression lautet:

$$Area_{12g} = -0.234489 * Polarizability\ cm^3 + 107.041$$

4 Methoden und Resultate

4.3.4 Henry-Law-Konstante

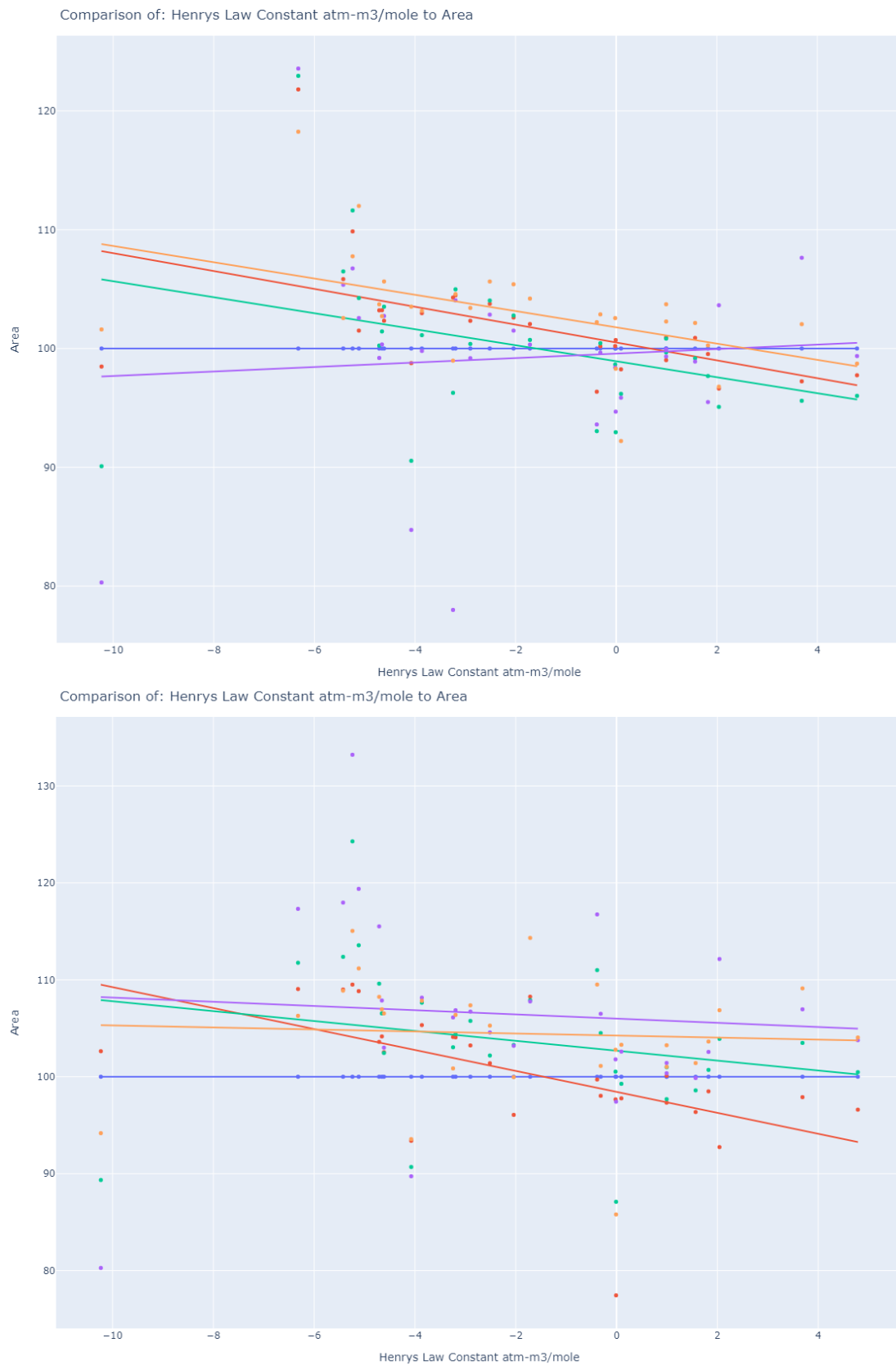


Abbildung 25: Lineare Regression für normalisierte "Areas" und Henry-Law-Konstante für Kaffeesorten Arab (oben) und Eugi (unten)

4 Methoden und Resultate

Für die Sorte Arab konnte das höchste Bestimmtheitsmass mit der Eiskwürfel-Kühlmethode und 5g aktiv gekühltem Kaffee erzielt werden. R^2 liegt bei 0.268076, die Formel der linearen Regression lautet:

$$Area_{5g} = -0.751667 * Henry\ Law\ Constant\ atm\ m^3/mole + 100.506$$

Für die Sorte Eugi konnte das höchste Bestimmtheitsmass mit der Eiskwürfel-Kühlmethode und 5g aktiv gekühltem Kaffee erzielt werden. R^2 liegt bei 0.305562, die Formel der linearen Regression lautet:

$$Area_{5g} = -1.08048 * Henry\ Law\ Constant\ atm\ m^3/mole + 95.4392$$

4.4 Multiple Lineare Regression

Um Zusammenhänge zwischen der Menge der flüchtigen Moleküle und mehreren unabhängigen Variablen in Abhängigkeit der Kühlmethode zu untersuchen, führten wir eine multiple lineare Regression durch.

Wir beschränkten uns auf die vier zentralen Molekulareigenschaften Molekülmasse, Löslichkeit, Polarisierbarkeit und Henry-Law-Konstante. Die Berechnungen wurde für beide Kaffeesorten und alle Kühlmethoden zweimal durchgeführt. Im ersten Durchlauf mit allen Molekülen und beim zweiten Durchlauf nur mit denjenigen, bei denen die Henry-Law-Konstante bekannt ist. Das sind 25 von 53.

Die Ergebnisse können mit dem Python Script «EDA_Chilling_Comp_HSGC.py» unter dem Pfad Github\fhmw-pro2d\Scripts\Chilling_Compound\Explorative_Datenanalyse generiert werden. Diese können einfach in einen Texteditor exportiert werden.

Für die Kaffeesorte Eugi bei der 5g Kaffee mit dem Eiswürfel gekühlt wurde, ohne Berücksichtigung der Moleküle, bei denen die Henry-Law-Konstante unbekannt ist, konnte das höchste Bestimmtheitsmass erzielt werden. Die Ergebnisse sehen wie folgt aus:

```
Eugi DataFrame rel. IR 5g%
OLS Regression Results
=====
Dep. Variable:          y      R-squared:          0.438
Model:                  OLS    Adj. R-squared:       0.340
Method:                 Least Squares      F-statistic:       4.476
Date:                  Wed, 22 Jun 2022    Prob (F-statistic): 0.00804
Time:                  12:17:14    Log-Likelihood:    -84.066
No. Observations:      28      AIC:              178.1
Df Residuals:          23      BIC:              184.8
Df Model:              4
Covariance Type:       nonrobust
=====
               coef      std err          t      P>|t|      [0.025      0.975]
-----
const          122.3124      11.337      10.789      0.000      98.860      145.764
x1             -0.1251       0.210      -0.597      0.556      -0.559      0.308
x2             -2.1049       0.920     -2.287      0.032      -4.009     -0.201
x3              0.0789       0.620       0.127      0.900      -1.204      1.362
x4             -0.2954       1.489     -0.198      0.844      -3.375      2.784
=====
Omnibus:           18.334    Durbin-Watson:       1.780
Prob(Omnibus):     0.000    Jarque-Bera (JB):    23.580
Skew:              -1.566    Prob(JB):            7.58e-06
Kurtosis:          6.225    Cond. No.            938.
=====

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
x1: Molecular Weight g/mol
x2: Solubility ca. 20 °C water, mg/ml
x3: Henrys Law Constant atm-m3/mole
x4: Polarizability cm3

coef intercept 122.3124443553308
model score r2 0.43768878398715594
```

Abbildung 26: Ergebnisse multiple lineare Regression für Kaffeesorte Eugi und 5g mit der Eiswürfel-Methode aktiv gekühltem Kaffee

4 Methoden und Resultate

Die Formel für eine multiple lineare Regression mit dem Chilling Compound Datensatz lautet:

$$\begin{aligned} \text{Area}[\text{Kühlmethode}] \\ = \beta + [\text{Molekulareigenschaft1}] * a_1 + [\text{Molekulareigenschaft2}] * a_2 \\ + [\text{Molekulareigenschaft3}] * a_3 + \dots + [\text{Molekulareigenschaftn}] * a_n \end{aligned}$$

Für das obige Ergebnis also:

$$\begin{aligned} \text{Area}_{5g} = & 122.3124443553308 + \text{Molecular Weight } g/mol * (-0.1251) \\ & + \text{Solubility ca. 20 } ^\circ C \text{ water } mg/mol * (-2.1049) + \text{Polarizability } cm^3 * (-0.2954) \\ & + \text{Henry Law Constant atm } m^3/mole * 0.0789 \end{aligned}$$

Mit dem p-Wert (in der Abbildung oben im zweiten Absatz die dritte Spalte von rechts) wird die Nullhypothese getestet, dass der Koeffizient gleich null ist. Dies würde bedeuten, dass die Variable (in unserem Fall der Wert der Molekulareigenschaft) keinen Effekt auf das Modell hat, also dass eine Konstante einen gleich guten Effekt hätte. Ein niedriger p-Wert (<0.05) bedeutet, dass der Wert der Variable mit dem Resultat (in unserem Fall dem Wert der «Area») zusammenhängt.

Im obigen Ergebnis ist die für das Ergebnis signifikanteste Molekulareigenschaft, die Löslichkeit. Der p-Wert liegt 0.032. Die anderen drei p-Werte sind dafür eher hoch.

Bei den verschiedenen multiplen linearen Modellen wurden vereinzelt sehr gute p-Werte erzielt. Das obige Ergebnis erzielte jedoch das höchste Bestimmtheitsmass.

5 Schlussbemerkungen

Die ursprüngliche Vermutung, dass bei kühlerer Wasser- und Kaffeetemperatur mehr flüchtige Moleküle im Kaffee enthalten sind, konnte in der Tendenz bestätigt werden.

Die Analysen der explorativen Datenanalyse wie auch der einfachen und multiplen linearen Regression haben aber gezeigt, dass die Vermutung zu kurz gedacht ist. Weitere Einflussfaktoren wie zum Beispiel die Kaffeesorte beeinflussen die enthaltenen Moleküle, deren Molekulareigenschaften und Bestätigung oder das Verwerfen der Vermutung.

Zwischen einzelnen Molekulareigenschaften und der Menge an flüchtigen Molekülen konnten in Abhängigkeit einer aktiven (Eiswürfel/Spout) oder passiven Kühlung (Veränderung Wassertemperatur) schwache lineare Zusammenhänge festgestellt werden. Zu beachten ist, dass R^2 nur lineare Zusammenhänge darstellt.

In einem nächsten Schritt sollten nicht lineare Modelle zur Analyse von Zusammenhängen in Betracht gezogen werden.

Weiter sehen wir Potential in der chemischen Forschung. Dazu gehört zum Beispiel eine vollständige Molekulardatenbank. Gerade bei einer wichtigen Kerngrösse (Henry-Law-Konstante) ist es suboptimal, wenn bei mehr als der Hälfte der Moleküle Werte fehlen. Ergänzend könnten sensorische Werte als Eigenschaften erforscht werden, um so das Kaffeeprofil noch genauer analysieren zu können.

Anhang

A Ergänzende Visualisierungen zu Kapitel 4.2.2.3

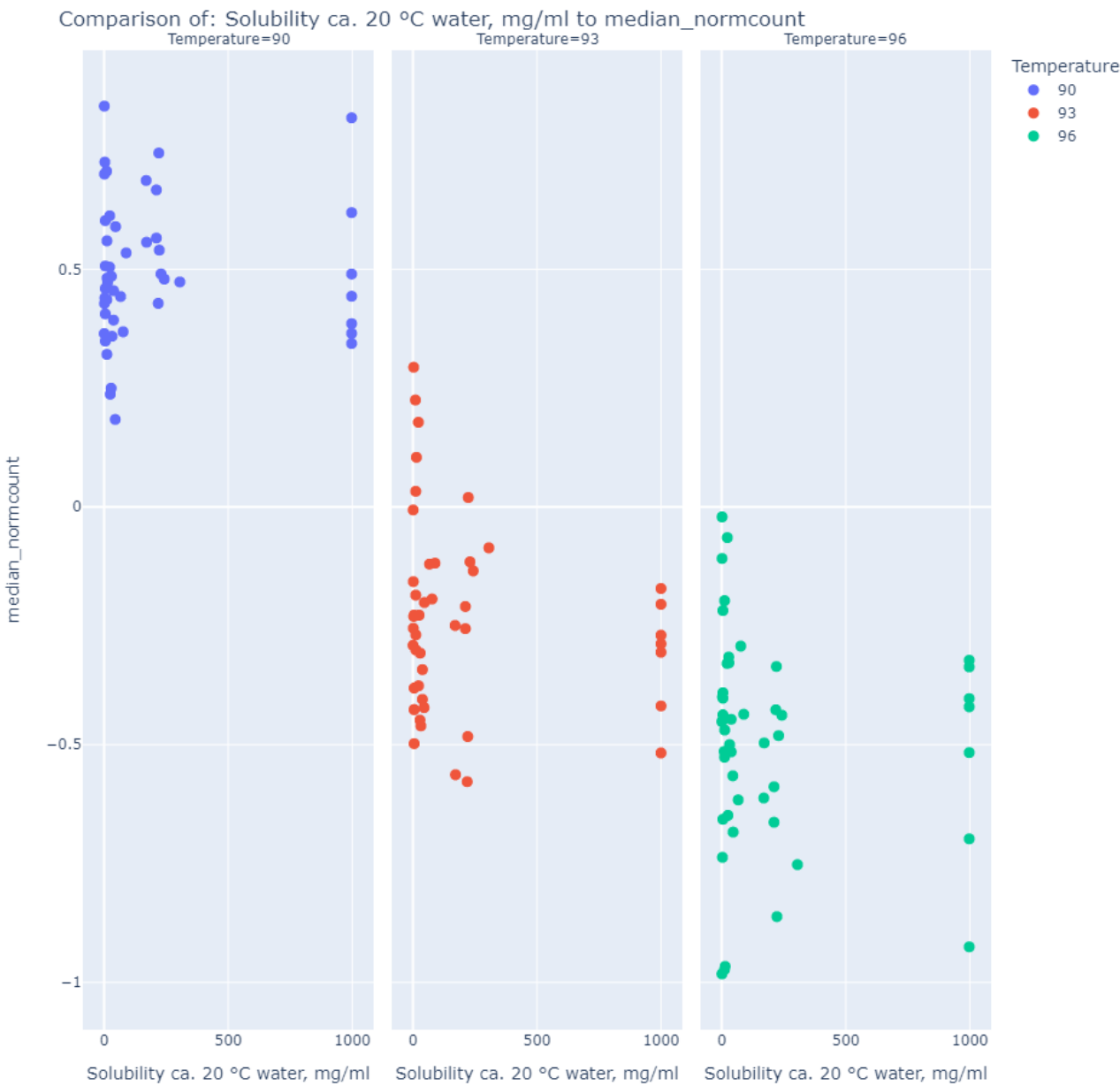


Abbildung 27: Zusammenhang normalisierte "Counts" und Löslichkeit

A Ergänzende Visualisierungen zu Kapitel 4.2.2.3

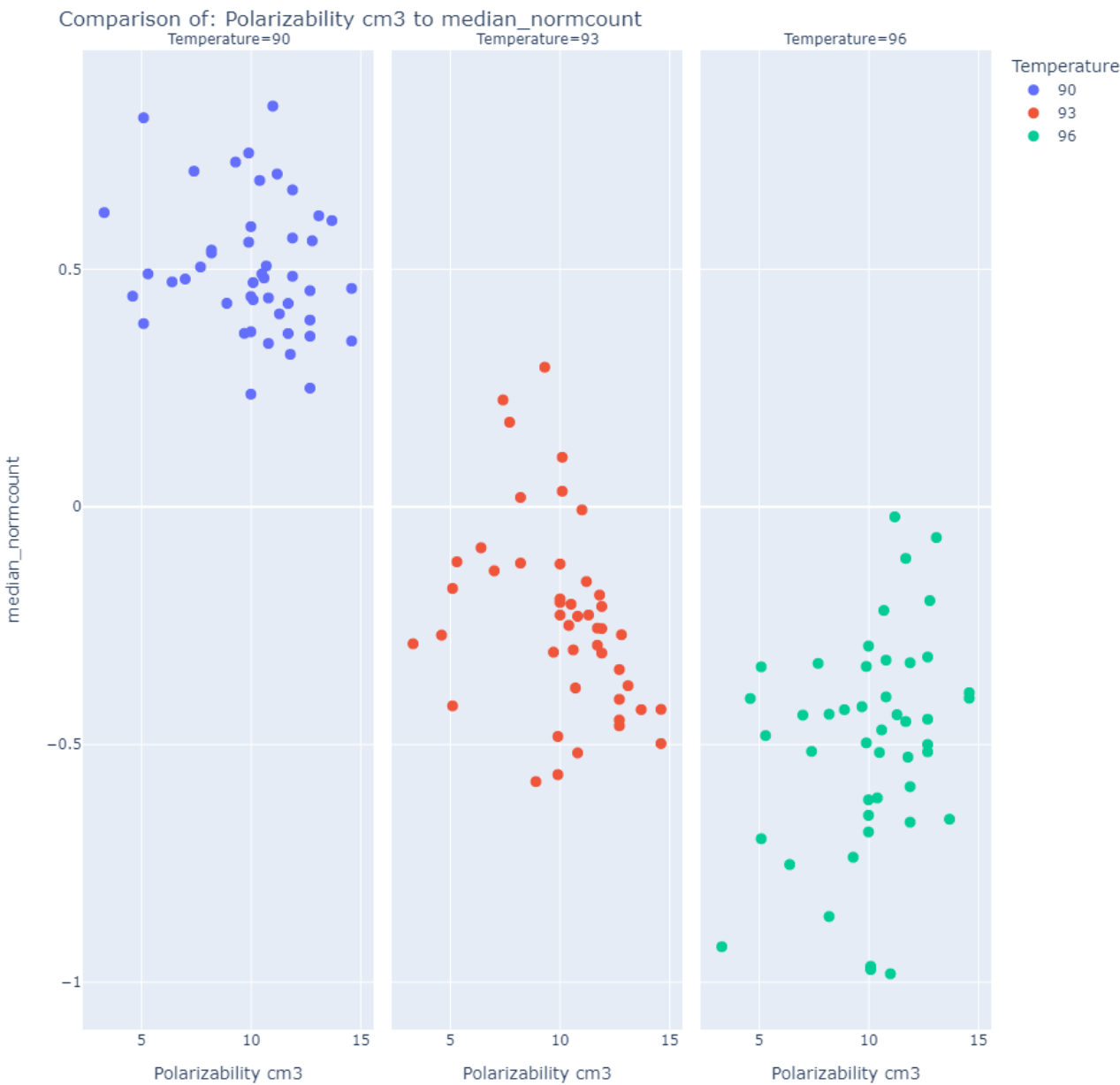


Abbildung 28: Zusammenhang normalisierte "Counts" und Polarisierbarkeit

A Ergänzende Visualisierungen zu Kapitel 4.2.2.3

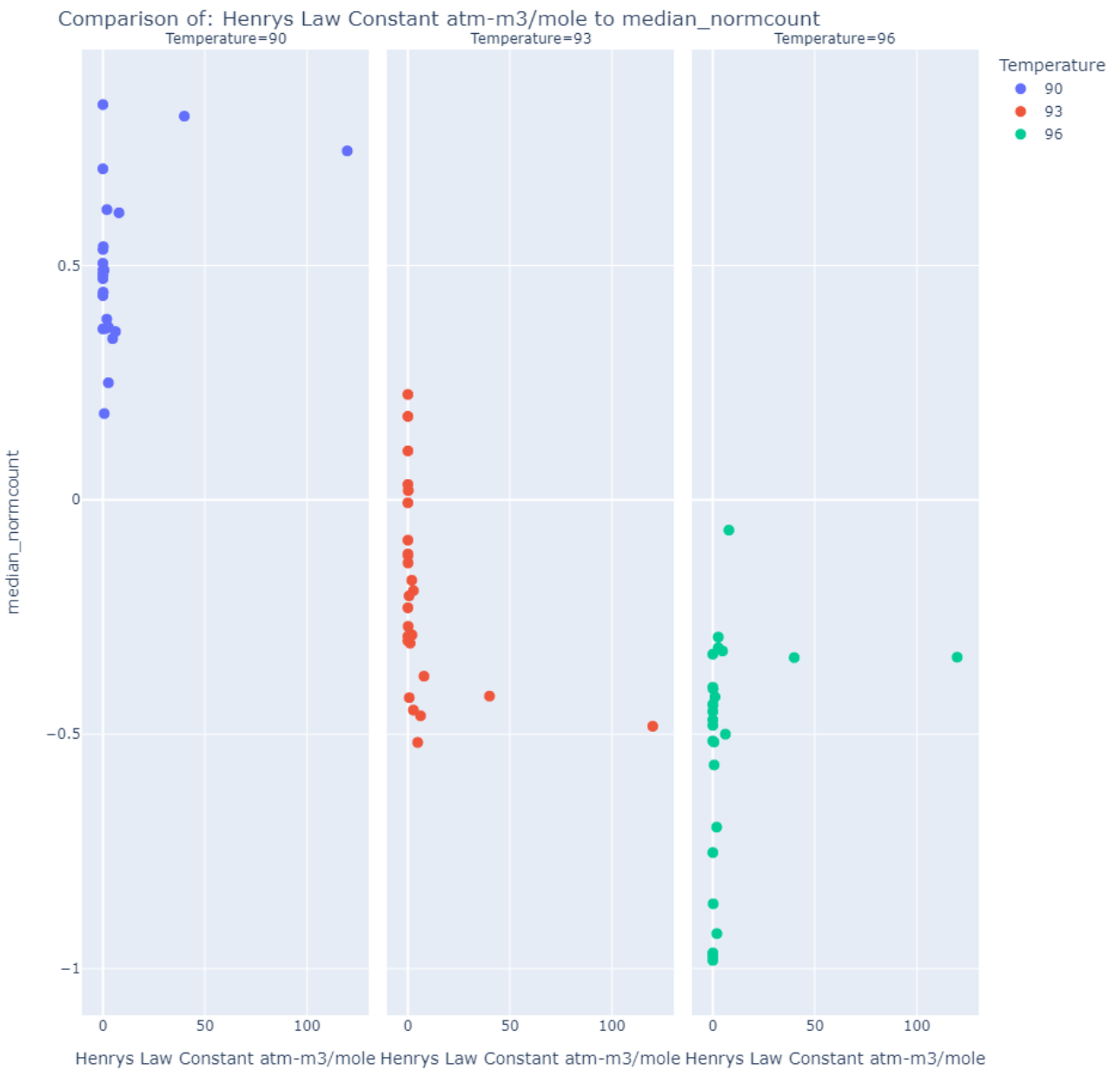


Abbildung 29: Zusammenhang normalisierte "Counts" und Henry-Law-Konstante noch ohne Logarithmus-Transformation (siehe Kapitel Explorative Datenanalyse zum Datensatz Chilling Compound)