

# Unterstützung der automatisierten Auswertung chemisch-analytischer Messungen von Spirituosen durch validierte Algorithmen und eine intuitive Benutzeroberfläche

## Assisting the automated analysis of chemical-analytical measurements in spirits using validated algorithms and an intuitive user interface

Dr. rer. nat. Andreas T. Grasskamp<sup>1</sup>, M. Sc. Satnam Singh<sup>1</sup>, M. Sc. Helen Haug<sup>1</sup>, PD Dr. rer. nat. Tilman Sauerwald<sup>1,2</sup>

<sup>1</sup>Fraunhofer-Institut für Verfahrenstechnik IVV, Freising, DE <sup>2</sup>Universität des Saarlandes, Saarbrücken, DE

Korrespondenz: [andreas.grasskamp@ivv.fraunhofer.de](mailto:andreas.grasskamp@ivv.fraunhofer.de) oder [tilman.sauerwald@ivv.fraunhofer.de](mailto:tilman.sauerwald@ivv.fraunhofer.de)

### Kurzfassung

Die eingehende Analyse umfangreicher chemischer Messdaten bedarf eines hohen zeitlichen und personellen Einsatzes. Viele Aspekte dieser Auswertung können allerdings automatisch durchgeführt werden, wenn die ansonsten von Hand durchgeführten Schritte in objektive algorithmische Routinen übersetzt werden. Zu diesem Zweck haben wir eine anpassbare Auswertesoftware für gaschromatographische Daten entwickelt und die damit durchführbare Auswertung am Beispiel von Whisky-Proben validiert. Mittels einer anwendungsspezifisch kuratierten Datenbank aus 217 bekannten Inhaltsstoffen konnten wir unter optimiertem Arbeits- und Zeitaufwand gegenüber anderen Verfahren zeigen, dass charakteristische Whisky-Komponenten in jeder der 16 beispielhaft untersuchten Proben den Hauptanteil ausmachten. Damit gehen wir einen entscheidenden Schritt hin zu einer automatischen Bewertung des Geschmacks von Genussmitteln.

### Abstract

Exhaustive analysis of chemical measurements requires considerable expenditure of time and personnel. However, many aspects of this can be automated by translating manual work into objective algorithmic routines. To this end, we developed an adaptable software for gas chromatography data and validated analysis steps using whisky samples. Employing a specifically curated database of 217 known whisky chemicals, we managed to optimize the required manual and temporal expense by a good margin and showed that characteristic whisky components constituted the majority of all detected molecules in all 16 analyzed samples. With this approach, we present a decisive contribution towards the automated assessment of aroma profiles in food.

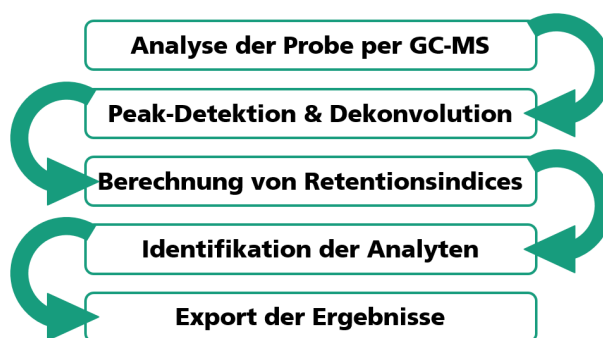
## 1 Hintergrund

Aromen bestimmen das Geruchs- und Geschmackserlebnis von Lebens- und Genussmitteln. In der weit verbreiteten Spirituose Whisky führt ein komplexes Gemisch aus vielen verschiedenen Aromen zu einem breiten Spektrum an Genussfaktoren (s. z.B. [1,2]). In der Regel stammen diese Aromen von flüchtigen organischen Verbindungen (auch *volatile organic compounds*, VOCs). Ein solches Gemisch aus nicht riechenden und riechenden Verbindungen mit variierenden Geruchsschwellen verlangt nach einer sehr selektiven, hochaufgelösten Messung mit hoher Messdynamik, um alle relevanten Geruchsstoffe zu erfassen. Als Quasi-Standard bietet instrumentelle chemische Analytik mittels Gaschromatographie-Massenspektrometrie (GC-MS) hier das derzeit leistungsfähigste Verfahren, um VOCs zu detektieren und zu identifizieren. Allerdings ist dieser Prozess aufgrund des hohen zeitlichen und technischen Aufwandes, besonders für die Auswertung der Daten, bisher mit hohen Kosten behaftet. Daher ist eine effiziente Methode zur verlässlichen Identifikation der entscheidenden Inhaltsstoffe wünschenswert. Eine solche Methode mit intuitiver Benutzerführung stellen wir hier vor

und validieren sie im Vergleich mit standardmäßig verwendeter Software. Eine zu diesem Zweck aus der Literatur gewonnene Liste von >200 Whisky-Komponenten (u.a. [1–5]) wurde genutzt, um die Qualität der Erkennung einzuschätzen.

## 2 Methoden

Das prinzipielle Vorgehen in konventioneller und automatisierter Analyse, denen beiden die Analyse per GC-MS vorausgeht, ist in **Bild 1** zu sehen. Ein besonderer Fokus



**Bild 1** Ablauf der Auswertung in beiden Verfahren

wurde hierbei auf die Reduktion der durch den Nutzer durchzuführenden Einstellungen gelegt.

## 2.1 Datenerhebung: SBSE-TD GC/MS

Es wurden insgesamt 16 verschiedene, kommerziell erhältliche Whiskys analytisch untersucht. Unter den Proben befanden sich 9 Scotch und 7 Bourbon Whiskys mit Alkoholgehalt zwischen 40 und 50,5 % vol. Zur besseren Vergleichbarkeit der Proben wurden diese zunächst mit Trinkwasser auf einen einheitlichen Alkoholgehalt von ca. 40 % vol. eingestellt.

Die Aufarbeitung der Proben zur Extraktion der flüchtigen Komponenten erfolgte unter Einsatz der sog. *Stir Bar Sorptive Extraction* (SBSE [6]) mittels PDMS-beschichteter Gerstel Twister® (1 cm Länge, 0,5 mm Schichtdicke; Gerstel GmbH & Co. KG, Mühlheim a. d. Ruhr, DE). Ein Aliquot der Proben wurde jeweils entnommen, mit einer Standard-Lösung von 4-Chloro-2-Methoxyphenol und Undecan in Ethanol versetzt, mit deionisiertem Wasser verdünnt und anschließend mit Natriumchlorid versetzt. Infolge der Extraktion eines entsprechenden Aliquots mithilfe des Twisters® wurden die Analyten mittels Thermo-desorption und anschließender cryo-Fokussierung (Gerstel TDU und KAS, Gerstel GmbH & Co. KG, Mühlheim a. d. Ruhr, DE) in das GC/MS (Trace GC Ultra, DSQ II, jeweils Thermo Fisher Scientific GmbH, Dreieich, DE) überführt. Der Trägergasfluss (Helium) über die TDU wurde auf 50 mL/min gesetzt. Die Auftrennung erfolgte über eine DB-FFAP Säule (30 m x 0,25 mm, Schichtdicke 0,25 µm; J & W Scientific, Agilent Technologies, Waldbronn, DE), die massenspektrometrische Detektion im *Electron Ionization* (EI) Modus bei 70 eV, Ionen wurden dabei im Full-Scan ( $m/z$  35-399) erfasst. Parallel hierzu wurde zur Ermittlung der Retentionsindex-Werte (RI) ein Alkanstandard (C6-C26) unter denselben gaschromatographischen Bedingungen gemessen.

## 2.2 Auswertung in AMDIS

Die GC-MS Daten wurden mithilfe der Software AMDIS (Version 2.37, 2017 [7–9]) ausgewertet. Hierfür wurde anhand des Alkanlaufs zunächst eine entsprechende RI-Kalibrierdatei erstellt und auf korrekte Zuordnung der Alkane überprüft. Anschließend wurden die 16 Whisky-Dateien einzeln über die Analyse-Funktion „Use Retention Index Data“ anhand der passenden RI-Kalibrierdatei prozessiert. Hierbei kam eine hausinterne über AMDIS erstellte Analytdatenbank für RI und MS mit rund 700 Analyten zum Einsatz. Die durch AMDIS ermittelten RI-Werte sowie die MS der erkannten Peaks wurden entsprechend automatisch in AMDIS mit den Datenbankeinträgen abgeglichen. Das Ergebnis der AMDIS-Auswertung wurde als Text-Report abgespeichert und im Weiteren mit den Ergebnissen des hier präsentierten Ansatzes verglichen.

## 2.3 Vorverarbeitung und automatische Erkennung von Alkan-peaks

Die Identifizierung der zugrundeliegenden Substanzen und ihrer entsprechenden Peaks in den in Abschnitt 2.1 erzeugten GC-MS-Daten erfordert die Verarbeitung und Identifizierung von Alkan-Standardreferenzen. Diese Referenzen wurden (analog zu [10]) zur Berechnung der RI verwendet, die der Eingrenzung der anschließenden Datenbanksuche dienten. Um GC-MS-Daten zu verarbeiten, haben wir eine Vorverarbeitungspipeline entwickelt. Der erste Schritt der Pipeline umfasst die Extraktion der Intensitätsmatrix der Form Anzahl der Scans x Bereich der  $m/z$ . Darauf folgt die Anwendung der Biller-Biemann-Entfaltung [11]. Die so entfalten Peaks bestehen aus einer großen Menge Rauschen, das entfernt werden muss. Angesichts der Struktur der Alkan-Referenz besteht die Aufgabe darin, große Peaks zu erkennen, die in bekannten Abständen und vergleichbaren Intensitäten auftreten. Um Peaks der Alkane zu unterscheiden, wurden daher die Peaks des Signals mit der geringsten Ausprägung als Mittelwert des Signals bestimmt. Um die verrauschten Übereinstimmungen von Alkanen weiter zu reduzieren, nutzten wir die lineare Variation der Retentionszeit-Abstände zwischen den Alkanen. Zu diesem Zweck haben wir ein lineares Modell trainiert, mit dem abgeschätzt werden kann, ob der als Alkan ausgewählte Peak nahe genug am Modell liegt, indem wir einen oberen und einen unteren Schwellenwert festlegen. Benachbarte Abstände, die kleiner als der untere Schwellenwert sind, werden als redundant betrachtet, während größere Abstände einen fehlenden Nachbarn andeuten. Der fehlende Peak wird durch Interpolation geschätzt. Für den Abgleich der Alkan-Peaks wurde C17 aufgrund der erwarteten Position und der geringeren Konzentration in der verwendeten Standardlösung als Unterscheidungsspeak verwendet. Die Identifizierung eines Peaks ermöglicht die Kennzeichnung aller anderen Peaks zusammen mit ihren Retentionszeiten.

## 2.4 Peak-Detektion und Dekonvolution

Die unbekannten Substanzen durchliefen dieselbe Vorverarbeitung, einschließlich der Extraktion der Intensitätsmatrix und der Dekonvolution nach Biller Biemann [11]. Um die Anzahl der falsch erkannten Peaks zu reduzieren, wurde ein Entrauschungsansatz verfolgt, der die Schätzung des MAA-Wertes (Median der absoluten Abweichungen) mit Hilfe eines gleitenden Fensters beinhaltet. Für jedes Fenster in der Matrix wird ein MAA-Wert berechnet und mit einem Schwellenwert verglichen, um Ausreißer zu identifizieren und zwischen Rauschen und dem gewünschten Signal zu unterscheiden. Die Retentionszeiten dieser Peaks werden dann auf der Festplatte gespeichert und zusammen mit den in Abschnitt 2.3 ermittelten Retentionszeiten der Alkane zur Berechnung der Retentionsindizes der unbekannten Verbindungen verwendet. Schließlich wurde eine Liste der Retentionszeiten, der berechneten Retentionsindizes und alle Massenspektren auf der Festplatte gespeichert, um weitere Vergleiche anzustellen.

## 2.5 Abgleich von Massenspektren und Datenbankeinträgen

Die Identifikation von Molekülen aus einzelnen detektierten und entfalteten („dekonvolvierten“) Massenspektren wurde mittels Berechnung der Kosinusähnlichkeit [8] wie folgt durchgeführt:

$$(1) \quad score_{MS} = \frac{\sum_{i=1}^n (MS_{lib,i} * MS_{temp})}{\sqrt{\sum_{i=1}^n (MS_{lib,i})^2} * \sqrt{\sum_{i=1}^n (MS_{temp})^2}}$$

wobei  $MS_{lib}$  eines von  $n$  in der Bibliothek vorhandenen Massenspektren und  $MS_{temp}$  ein unbekanntes Massenspektrum darstellt. Der höchste Intensitätswert jedes MS (in Datenbank-Einträgen und unbekannten Spektren) wurde jeweils durch Normalisierung auf 1000 gesetzt. Alle m/z-Werte wurden auf ganzzahlige Werte gerundet.

Bei Vorhandensein eines zum unbekannten MS gehörigen RI-Wertes wurde  $score_{MS}$  mit diesem zu einem neuen  $score_{MS+RI}$  verrechnet:

$$(2) \quad score_{MS+RI} = 1 - \frac{(1 - score_{MS}) + abs(\frac{RI_{curr} - RI_{ref}}{RI_{ref}})}{2}$$

Hierbei wurden nur Datenbankeinträge als mögliche Treffer berücksichtigt, für die galt:  $|RI_{curr} - RI_{ref}| \leq 30$ . Dafür wurde eine rechteckige Signum-Funktion verwendet und mit dem jeweiligen  $score$  multipliziert:

$$(3) \quad f(|RI_{curr} - RI_{ref}|) = \begin{cases} 0, falls & |RI_{curr} - RI_{ref}| > 30 \\ 1, falls & |RI_{curr} - RI_{ref}| \leq 30 \end{cases}$$

Die Vorgänge in **Gleichung (1) bis (3)** wurden für alle unbekannten Spektren in einem Chromatogramm wiederholt. Werte für den  $score$  lagen zwischen 0 (keine Übereinstimmung) und 1 (identisch). Derjenige Datenbankeintrag mit dem jeweils höchstenscore wurde als identifiziertes Molekül anerkannt, wenn der Wert des  $score \geq 0.6$  (bzw.  $\geq 0.8 / 0.9$ ) betrug.

## 2.6 Abgleich der Ergebnisse zwischen AMDIS und MATLAB

Um die Erkennung der Moleküle in allen 16 erhobenen Proben zwischen den beiden Ansätzen (AMDIS-Auswertung & Benutzeroberfläche) zu vergleichen, sind wir wie folgt vorgegangen. Für jede Probe wurde eine Liste aller einzigartigen Moleküle erstellt (d.h. Mehrfachdetektionen wurden entfernt). Dabei wurde jeweils dasjenige Vorkommen eines Moleküls behalten, das mit der höchsten Erkennungs-Sicherheit in Verbindung stand („Net score“ in AMDIS,  $score$  in MATLAB). Die Berechnung des „Net score“ in AMDIS konnte bisher nicht zweifelsfrei nachvollzogen werden, aber Metriken wie die Unterschiede zwischen Massenspektren und RI-Werten fließen auch hier ein [9]. Dann wurde für jedes in AMDIS detektierte Molekül ge-

prüft, ob dieses in der automatisierten Routine ebenso erkannt wurde. Aus der Anzahl der in beiden Ansätzen gefundenen Moleküle wurde dann die Übereinstimmung berechnet, indem durch die Gesamtzahl der in AMDIS detektierten Einzelbestimmungen geteilt wurde.

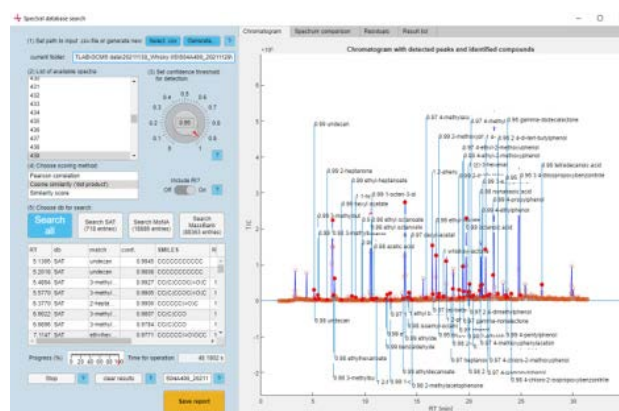
Die Einteilung der gefundenen Moleküle in die zwei Klassen (bekannter/nicht bekannter charakteristischer Whisky-Bestandteil) fand durch Abgleich mit der vorhandenen Datenbank und der darin enthaltenen binären Information über die Zugehörigkeit statt.

## 3 Ergebnisse & Diskussion

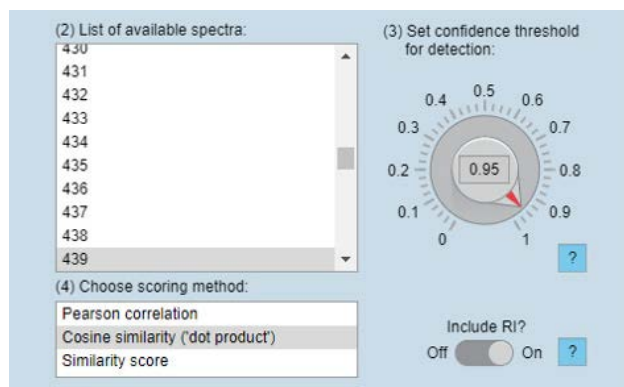
Es existieren viele proprietäre Lösungen für die Analyse von instrumentell-analytischen Datensätzen, bei denen allerdings oft eine Schnittstelle zum Export der Daten in andere Programme fehlt oder nicht ohne Umwege möglich ist. Wir haben hiermit ein erweiterbares und intuitives Werkzeug entwickelt, das auf viele Import-Dateiformate anpassbar ist, den Export in mehreren Formaten erlaubt und mit maßgeschneiderten Datenbanken arbeiten kann. Gegenüber einer großen Notwendigkeit für Benutzerinteraktion im standardmäßig verwendeten und weit verbreiteten Programm AMDIS [7] durchlief die automatische Auswertung per Benutzeroberfläche den Datensatz mit weniger manuellen Schritten, mit Ausnahme der Definition von Parametern und Arbeitsverzeichnis zu Beginn.

### 3.1 Aufbau und Validierung der Benutzeroberfläche

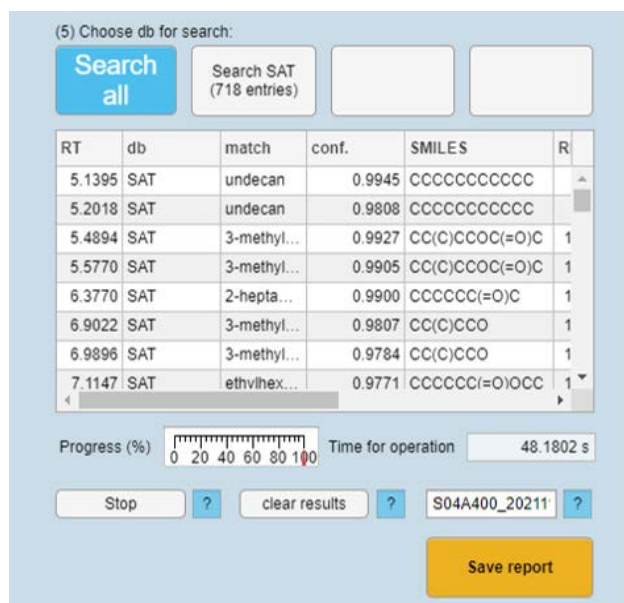
Zur benutzerfreundlichen Anwendung der beschriebenen Verfahren wurde eine intuitiv zu bedienende, modular erweiterbare graphische Benutzeroberfläche (GUI) entwickelt (**Bilder 2-5**). Mit dem hier beschriebenen Ansatz konnten wir die aufwändige Auswertung eines Datensatzes von 16 verschiedenen Whiskys zu einem entscheidenden Teil automatisieren und den manuellen Aufwand vermindern.



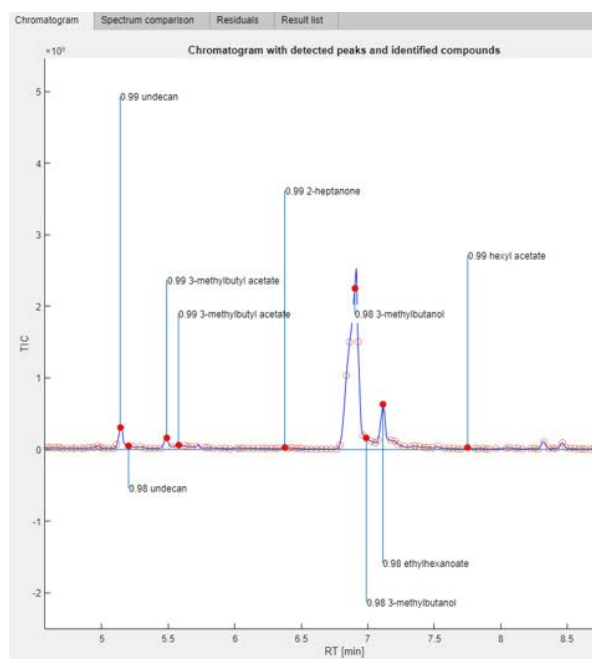
**Bild 2** Gesamtansicht der GUI mit ausgewertetem Chromatogramm



**Bild 3** Detailansicht der GUI: Definition des Arbeitsverzeichnisses und Einstellung von Parametern



**Bild 4** Detailansicht der GUI: Ergebnistabelle und Exportoptionen



**Bild 5** Detailansicht der GUI: Identifizierte Analyten

### 3.2 Ein automatisierter Ansatz liefert qualitativ hochwertige Ergebnisse

Probe	einzigartige Moleküle (AMDIS)	Übereinstimmung (%)	Ø RT diff. (s)
S01	68	95.58	-1.44
S02	81	82.72	-1.10
S03	76	89.47	0.15
S04	79	93.67	0.35
S05	63	85.71	-2.28
S06	70	78.57	-0.59
S07	70	80.00	-1.25
S08	63	90.48	0.23
S09	73	80.82	-0.84
B01	70	80.00	1.00
B02	69	86.95	-0.44
B03	66	78.78	0.96
B04	71	87.32	-0.51
B05	70	87.14	-1.18
B06	66	92.42	-0.41
B07	66	92.42	0.46
Ø		<b>86.38</b>	<b>-0.43</b>

**Tabelle 1** Ergebnisse des Vergleichs zwischen der Auswertung in AMDIS und der Benutzeroberfläche.

Um die Qualität unseres neuen Ansatzes zu erheben, haben wir zuerst überprüft, welche der durch die konventionelle Auswertung detektierten Moleküle (mit mäßiger und hoher Sicherheit) auch in der automatisierten Auswertung aufzufinden waren. Wie **Tabelle 1** zu entnehmen, konnten wir durch die Benutzeroberfläche im Mittel ~86% der einzigartigen Moleküle, die durch Auswertung in AMDIS gefunden wurden, auch mit der entwickelten Benutzeroberfläche identifizieren. Hierbei lag die mittlere Abweichung in der Retentionszeit bei  $< \frac{1}{2}$  s.

Als nächstes wollten wir wissen, welcher Anteil der detektierten Moleküle in beiden Ansätzen auch in der Liste der 217 charakteristischen Whisky-Komponenten enthalten war. Außerdem haben wir betrachtet, ob eine bessere Erkennung mit dem Vorhandensein in der Liste von charakteristischen Whisky-Komponenten zusammenhängt. Dafür haben wir in jeder Probe die mittlere Qualität der Erkennung in zwei Klassen verglichen: die Moleküle, die in der Whisky-Liste vorhanden waren, und jene, für die das nicht der Fall war. Im Fall der automatisierten Routine war in allen Proben der Unterschied signifikant, während dies bei der AMDIS-Auswertung nur in 4 Proben der Fall war. Hierbei spielten jedoch mehrere Faktoren eine Rolle. Dies war zum einen die Tatsache, dass in AMDIS im Mittel 78.3% der gefundenen Stoffe als charakteristische Whisky-Moleküle erkannt wurden, wohingegen das bei der automatisierten Auswertung nur für durchschnittlich 45.8 % der erkannten Moleküle zutraf. Im automatisierten Ansatz wurden indes in allen Proben deutlich mehr Stoffe erkannt



als in der Auswertung durch AMDIS. Dies spricht für eine hohe Sensitivität des automatisierten Ansatzes für Spurenstoffe oder Hintergrund-Rauschen. Eine Änderung der Kriterien für die Auswertung (*net score* mindestens 80, entspricht hoher Sicherheit der Identifikation) und der automatisierten Auswertung (*score* mindestens 0.8) führte zu einer Erhöhung und Annäherung der Verhältnisse in beiden Klassen (AMDIS: 82.7 %, MATLAB: 59.9%) und einer leichten Verbesserung der Übereinstimmung auf 88.7 %. Eine Anhebung beider Kriterien auf 90 bzw. 0.9 führte zu einer Übereinstimmung von 88.2 % und einer weiteren Annäherung der Vorhersagequalität (85.2 % bzw. 75.1 %).

Weitere Evaluationen sind hier nötig, um diese Metriken zu optimieren. Angesichts des stark verminderten Aufwandes ist dieses Ergebnis jedoch vielversprechend. Zudem spricht eine fehlende Zuordnung von <25% der erkannten Moleküle zu charakteristischen Whisky-Komponenten nicht *per se* dafür, dass diese fälschlich identifiziert wurden. In nachfolgenden Versuchen werden wir die Anwendbarkeit der hier vorgestellten automatisierten Auswerterroutine auf die Analyse anderer VOC-Emittenten (z.B. andere Lebensmittel, Kunststoffe o.Ä.) demonstrieren. Des Weiteren möchten wir anhand von künstlich hergestellten Mischungen eine sichere Referenz für den Vergleich unseres Ansatzes mit der manuellen Expertenauswertung schaffen, da erst dann eine vollends verlässliche Validierung möglich ist. Die hier zugegebenen Standard-Substanzen (Undecan und 4-Chloro-2-Methoxyphenol, s. Methoden) wurden in allen Proben detektiert. Die noch einmal weitaus zeitaufwändigere manuelle Expertenauswertung unter Einsatz von AMDIS (gegenüber dem reinen Erstellen und Exportieren von Ergebnissen), bei der vorhandene Moleküle nach bestimmten Kriterien beurteilt und gewählt werden, bietet zudem noch weiteres Potenzial für die Verringerung der aufzuwendenden Kosten.

## 4 Danksagung

Wir bedanken uns beim Bayerischen Staatsministerium für Wirtschaft, Landesentwicklung und Energie für die Förderung im Rahmen des Projekts "Campus der Sinne".

## 5 Literatur

- [1] Poisson, L.; Schieberle, P.: Characterization of the most odor-active compounds in an American Bourbon whisky by application of the aroma extract dilution analysis. *Journal of Agricultural and Food Chemistry*, 56 (2008), S. 5813–5819(14).
- [2] Câmara, J. S.; Marques, J. C.; Perestrelo, R. M.;...; Caldeira, M.: Comparative study of the whisky aroma profile based on headspace solid phase microextraction using different fibre coatings. *Journal of Chromatography. A*, 1150 (2007), S. 198–207(1-2).
- [3] Daute, M.; Jack, F.; Baxter, I.;...; Walker, G.: Comparison of Three Approaches to Assess the Flavour Characteristics of Scotch Whisky Spirit. *Applied Sciences*, 11 (2021), S. 1410(4).
- [4] Demyttenaere, J.; Sanchezmartinez, J.; Verhe, R.; Sandra, P [P.]; Dekimpe, N.: Analysis of volatiles of malt whisky by solid-phase microextraction and stir bar sorptive extraction. *Journal of Chromatography. A*, 985 (2003), S. 221–232(1-2).
- [5] Jeleń, H. H.; Majcher, M.; Szwengiel, A.: Key odorants in peated malt whisky and its differentiation from other whisky types using profiling of flavor and volatile compounds. *LWT*, 107 (2019), S. 56–63.
- [6] Baltussen, E.; Sandra, P [Pat]; David, F.; Cramers, C.: Stir bar sorptive extraction (SBSE), a novel extraction technique for aqueous samples: Theory and principles. *Journal of Microcolumn Separations*, 11 (1999), S. 737–747(10).
- [7] Stein, S. E.: An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *Journal of the American Society for Mass Spectrometry*, 10 (1999), S. 770–781(8).
- [8] Stein, S. E.; Scott, D. R.: Optimization and testing of mass spectral library search algorithms for compound identification. *Journal of the American Society for Mass Spectrometry*, 5 (1994), S. 859–866(9).
- [9] D'Arcy, P.; Mallard, W. G.: AMDIS-user guide. <https://www.nist.gov/system/files/documents/srd/AMDISMan.pdf>
- [10] Haug, H.; Singh, S.; Grasskamp, A.; Sauerwald, T.: P5.1 - Automatisierte Auswertung chromatographischer Daten für die Bewertung von Whisky. Tagungsband Zum 15. Dresdner Sensor-Symposium 2021, 2021, pp. 245–250.
- [11] Biller, J. E.; Biemann, K.: Reconstructed Mass Spectra, A Novel Approach for the Utilization of Gas Chromatograph – Mass Spectrometer Data. *Analytical Letters*, 7 (1974), S. 515–528(7).