

CITAC
Cooperation on International
Traceability in Analytical Chemistry

Eurachem / CITAC Guide

Messunsicherheit durch Probenahme

Ein Leitfaden zu Methoden und Ansätzen

2. Auflage auf englisch – 2019

deutsch – 2023

**Gemeinsam erstellt mit Eurolab, Nordtest und
dem RSC Analytical Methods Committee**

Messunsicherheit durch Probenahme: Ein Leitfaden zu Methoden und Ansätzen

2. Auflage (2019) auf deutsch (2023)

Herausgeber

Michael H Ramsey (University of Sussex, UK)
Stephen L R Ellison (LGC, UK)
Peter Rostron (University of Sussex, UK)

Zusammensetzung der Arbeitsgruppe *

Eurachem

Michael H Ramsey (chair)	<i>University of Sussex, UK</i>
Stephen L R Ellison (secretary)	<i>LGC, Teddington, UK</i>
Paolo de Zorzi	<i>ISPRA (Italian National Institute for Environmental Protection and Research), Italy</i>
Pentti Minkkinen	<i>LUT University, Finland</i>
Silke Richter	<i>BAM, Germany</i>
Eskil Sahlin	<i>RISE (Research Institute of Sweden), Sweden</i>
Alex Williams	<i>UK</i>

Eurolab

Irache Visiers	<i>Applus</i>
Rüdiger Kaus	<i>Eurolab DE, Germany</i>

CITAC

Ilya Kuselman	<i>Independent Consultant on Metrology, Israel</i>
Jorge Eduardo S Sarkis	<i>Institute of Nuclear and Energy research, Brazil</i>

Nordtest

Bertil Magnusson	<i>Trollboken AB, Sweden</i>
------------------	------------------------------

RSC AMC

Peter Rostron	<i>RSC, UK</i>
Roger Wood	<i>RSC, UK</i>

**Zur Zeit der Genehmigung des Dokuments*

Zitierweise

Diese Veröffentlichung sollte wie folgt zitiert werden*:

M H Ramsey, S L R Ellison and P Rostron (eds.)
Eurachem/EUROLAB/ CITAC/Nordtest/AMC Guide:
Measurement uncertainty arising from sampling: a
guide to methods and approaches. Second Edition,
Eurachem (2019). ISBN (978-0-948926-35-8).
Available from <http://www.eurachem.org>

**Je nach Anforderungen der entsprechenden Zeitschrift*

Übersetzung ins Deutsche

Die Übersetzung erfolgte durch Dr. M. Koch, Universität Stuttgart, J. Sunderkötter, BAM Berlin und unter Mitwirkung weiterer Kollegen von Eurolab-D.

Im Falle von Differenzen zwischen dieser Übersetzung und dem englischen Original ist die englische Fassung maßgeblich.

Danksagung

Dieses Dokument wurde in erster Linie von einer gemeinsamen Arbeitsgruppe von Eurachem / EUROLAB / CITAC / Nordtest / AMC in der rechts abgebildeten Zusammensetzung im Zeitraum 2017–2019 erarbeitet. Die Herausgeber sind all diesen Personen und Organisationen und allen anderen, die Kommentare, Ratschläge und Unterstützung beigesteuert haben, dankbar.

Die Erstellung dieses Handbuchs wurde teilweise vom Analytical Methods Trust unterstützt.

[Hier eingeben]

Vorwort

Die Messunsicherheit ist der allerwichtigste Parameter, der die Qualität der Messungen beschreibt. Denn die Unsicherheit beeinflusst grundlegend die Entscheidungen, die auf dem Messergebnis basieren. Bei der Entwicklung von Verfahren zur Schätzung der Unsicherheit, die ihren Ursprung im analytischen Teil der Messung hat, wurden erhebliche Fortschritte erzielt, und es liegen Leitlinien für diese Verfahren vor [1]. Eine Messung beinhaltet jedoch fast immer den Prozess der Probenahme. Denn in der Regel ist es nicht möglich, die gesamte Masse des zu charakterisierenden Materials zu analysieren (das Probenahmezielobjekt). Wenn das Ziel der Messung darin besteht, den Wert der Analytkonzentration in einem Probenahmezielobjekt zu schätzen, dann muss die mit dem Probenahmeprozess verbundene Unsicherheit zwangsläufig zu der mit dem gemeldeten Ergebnis verbundenen Unsicherheit beitragen. Es hat sich immer deutlicher gezeigt, dass die Probenahme oft der wichtigere Beitrag zur Unsicherheit ist und daher ein ebenso sorgfältiges Management und Kontrolle erfordert. Daher sollte die Unsicherheit, die sich aus dem Probenahmeverfahren ergibt, bewertet werden. Während die bestehenden Leitlinien die Probenahme als möglichen Beitrag zur Unsicherheit in einem Ergebnis identifizieren, sind die Verfahren zur Schätzung der resultierenden Unsicherheit nicht gut entwickelt und es sind weitere, spezifische Leitlinien erforderlich.

Historisch gesehen haben sich die Messwissenschaftler in erster Linie mit Messungen innerhalb von Laboren beschäftigt, und der Prozess der Probenahme wurde von einer anderen Gruppe von Personen durchgeführt, die oft in getrennten Organisationen arbeiten. Das Wissen des Messwissenschaftlers über den Probenahmeprozess ist dann sehr begrenzt. Umgekehrt ermöglicht das Einführen von In-situ-Analyseverfahren dem Messwissenschaftler manchmal Messungen an der Probenahmestelle und im Kontakt mit dem zu untersuchenden Material. Beispiele für diese Situation sind Prozessanalysen in der industriellen Produktion und In-situ-Messungen auf kontaminierten Flächen. Die Platzierung des Analysesensors in diesen Situationen stellt dann die Entnahme einer Probe dar, und der Messwissenschaftler wird nicht nur über alle Phasen des Messvorgangs, einschließlich der Probenahme, informiert, sondern ist auch für diese verantwortlich. Ein solches Bewusstsein für den gesamten Prozess ist wichtig, unabhängig von der Arbeitsteilung. Da Analyse- und Probenahmeverfahren zur Unsicherheit im Ergebnis beitragen, kann die Unsicherheit nur geschätzt werden, wenn der gesamte Prozess verstanden wird. Darüber hinaus ist eine optimale Verteilung des Arbeitsaufwandes zwischen Probenahme und Analyse nur möglich, wenn sowohl Probenahme- als auch Analyseprozesse verstanden werden.

Wenn die verschiedenen Phasen in der Verantwortung verschiedener Personen liegen, muss eine gute Kommunikation zwischen allen Beteiligten stattfinden. Probenahmeplaner und Analytiker müssen das gesamte Messverfahren optimieren und eine Strategie zur Schätzung der Unsicherheit entwickeln. Beide müssen die Ziele der Messungen mit dem Kunden besprechen. Alle drei Parteien benötigen von der zuständigen Aufsichtsbehörde Anleitungen, wie diese Schätzungen der Unsicherheit zu handhaben sind, um die Zuverlässigkeit der auf den Messungen basierenden Entscheidungen sicherzustellen. Um diese Entscheidungen zu untermauern, benötigen alle Parteien zuverlässige Schätzungen der Unsicherheit, einschließlich derjenigen, die sich aus der Probenahme ergeben. Obwohl keine allgemeine

Anleitung den Expertenrat in komplexen oder kritischen Fällen ersetzen kann, werden in diesem Handbuch die Methoden beschrieben, die erforderlich sind, um die für die meisten analytischen Messsysteme erforderlichen zuverlässigen Schätzungen der Unsicherheit bei der Probenahme zu erfüllen.

Seit der Veröffentlichung der ersten Ausgabe dieses Leitfadens im Jahr 2007 wurden einige wichtige Entwicklungen in diese zweite Ausgabe aufgenommen. Diese schließen ein

- den Ausdruck der Messunsicherheit als Unsicherheitsfaktor (FU), wenn die die Probenahmeunsicherheit beschreibende Häufigkeitsverteilung nicht normal, sondern logarithmisch normal ist, und die Verwendung eines „Unsicherheitsfaktors“ in einem Unsicherheitsbudget.
- die Verwendung eines nicht-ausgewogenen Designs, um die Unsicherheit kostengünstiger abzuschätzen, als dies mit dem ausgewogenen Design nach der „Duplikatmethode“ möglich wäre;
- Aktualisierungen von Definitionen und Verweisen, um aktuelle internationale Dokumente und Literatur widerzuspiegeln, einschließlich der Anwendung dieser Methoden auf Messungen vor Ort und in-situ, sowohl im Makro- als auch im Mikromaßstab.

Zusammenfassung

In diesem Leitfaden werden verschiedene Methoden beschrieben, mit denen die Messunsicherheit abgeschätzt werden kann, insbesondere jene, die sich aus den Probenahmeprozessen und der physikalischen Probenvorbereitung ergibt. Es ist eine ganzheitliche Sichtweise des Messprozesses erforderlich, der alle diese Schritte sowie den Analyseprozess einschließt, wenn die Messgröße als Wert der Analytkonzentration im Probenahmezielobjekt definiert wird und nicht nur in der Probe, wie sie ins Labor geliefert wurde. Der Leitfaden zeigt zunächst, wie wichtig es ist, die Gesamtunsicherheit einer Messung zu kennen, um eine zuverlässige Interpretation der Messungen vorzunehmen und ihre Eignung für den Zweck zu beurteilen. Dabei wird der gesamte Messprozess betrachtet, Einzelschritte definiert und deren Auswirkungen und Fehler hinsichtlich der Gesamtunsicherheit der Messergebnisse beschrieben.

Es werden zwei grundlegende Ansätze zur Schätzung der Unsicherheit aus der Probenahme beschrieben. Der empirische Ansatz verwendet wiederholte Probenahmen und Analysen unter verschiedenen Bedingungen, um die Auswirkungen zu quantifizieren, die durch Faktoren wie die Heterogenität des Analyten im Probenahmezielobjekt und durch Schwankungen bei der Anwendung einer oder mehrerer Probenahmeverfahren verursacht werden, um die Unsicherheit (und normalerweise einige ihrer Komponenten) zu quantifizieren. Der Modellansatz verwendet ein vordefiniertes Modell, das jede Komponente der Unsicherheit identifiziert, Schätzungen für jede Komponente vornimmt und diese summiert, um eine Gesamtschätzung vorzunehmen. In diesem Ansatz können manchmal Modelle aus der Probenahme-Theorie verwendet werden, um einige der Komponenten aus der Kenntnis der Eigenschaften der einzelnen Teilchen abzuschätzen.

Für jeden dieser Ansätze werden Beispiele aus verschiedenen Anwendungsbereichen angeführt. Dazu gehören Untersuchungen zur Umwelt (Boden und Wasser), zu Lebensmitteln (Anbau und Verarbeitung) und zu Futtermitteln. Die Schätzungen der erweiterten Messunsicherheit reichen von wenigen Prozent bis zu mehr als 80 %, bezogen auf die Messgröße. Der Beitrag der Probenahme ist gelegentlich gering, aber häufig dominant (kann 90 % der Messunsicherheit, ausgedrückt als Varianz, überschreiten). Dies deutet darauf hin, dass ein größerer Aufwand für die Probenahme und nicht für die chemische Analyse betrieben werden muss, wenn die Gesamtunsicherheit verringert werden muss, um die Zwecktauglichkeit zu erreichen.

Zu den Managementfragen, die behandelt werden, gehört die Verantwortung für die Qualität des gesamten Messprozesses, der auch das Probenahmeverfahren umfassen muss. Es wird eine Anleitung zur Auswahl des für jede Anwendung am besten geeigneten Ansatzes gegeben und ob eine einzelne Erstvalidierung des Systems ausreicht oder ob eine fortlaufende Überwachung der Unsicherheit durch die Probenahme mithilfe der Qualitätskontrolle der Probenahme erforderlich ist. Die zusätzlichen Kosten für die Schätzung der Unsicherheit werden auch in Bezug auf die Kosteneinsparungen berücksichtigt, die durch eine zuverlässigere Kenntnis der Messunsicherheit erzielt werden können.

Solch ein Leitfaden kann niemals vollständig sein, und obwohl es Anhänge mit Einzelheiten zu einigen der angewandten statistischen Techniken und Quellen für detailliertere Ratschläge

gibt, wird es in komplexeren Situationen häufig einen Bedarf an Expertenrat geben. Dieser Leitfaden soll eine nützliche Einführung in dieses Thema sein, aber wir hoffen, dass er auch die weitere Erforschung verbesserter Methoden zur Unsicherheitsschätzung anregt.

Inhaltsverzeichnis

Vorwort	i
Zusammenfassung.....	iii
Inhaltsverzeichnis.....	v
Liste der Akronyme und Abkürzungen.....	viii

TEIL 1 - Einführung und Anwendungsbereich

1 Einleitung.....	1
1.1 Begründung für den Leitfaden	1
1.2 Ziel des Leitfadens	1
1.3 Anwendung zur Beurteilung der Zwecktauglichkeit	1
1.4 Vorgesehene Zielgruppe für den Leitfaden.....	2
1.5 Beziehung dieses Leitfadens zu anderen Dokumenten	2
1.6 Verwendung des Leitfadens	3
2 Geltungsbereich und Einsatzgebiet.....	4
3 Terminologie.....	5

TEIL 2 – Grundlegende Konzepte

4 Probenahme im Messprozess	6
5 Messunsicherheit	8
5.1 Definition der Messunsicherheit	8
5.2 Spezifizierung der Messgröße	10
5.3 Fehler, Präzision und Unsicherheit.....	11
5.4 Probenahme und physikalische Vorbereitung als Quellen der Messunsicherheit	11
5.5 Unsicherheitsquellen.....	12
5.6 Heterogenität als Unsicherheitsquelle	13
5.7 Variation der Unsicherheit mit dem Konzentrationsbereich	14

TEIL 3 – Schätzung der Messunsicherheit unter Einbezug der Probenahme

6 Ansätze zur Unsicherheitsschätzung.....	17
7 Die Messgröße.....	18
8 Allgemeine Studiendurchführung zur Probenahmeunsicherheit	18

9 Empirischer Ansatz	19
9.1 Überblick.....	19
9.2 Unsicherheitsquellen.....	19
9.3 Statistisches Modell für die empirische Schätzung der Unsicherheit.....	20
9.4 Empirische Schätzung der Unsicherheit	21
9.5 Berechnung der Unsicherheit und ihrer Komponenten.....	24
9.6 Alternative empirische Methoden zur Unsicherheitsschätzung.....	26
10 Der Modellierungs-Ansatz.....	27
10.1 Ursache-Wirkungs-Modellierung	27
10.2 Probenahmetheorie für die Schätzung der Unsicherheit	27
 Teil 4 – Managementfragen	
11 Verantwortung für die Qualität der Probenahme	32
12 Auswahl des Ansatzes zur Unsicherheitsschätzung.....	32
13 Qualitätskontrolle der Probenahme	34
13.1 Beziehung zwischen Validierung und Qualitätskontrolle.....	34
13.2 Methoden der internen Qualitätskontrolle der Probenahme	35
14 Berichterstattung und Interpretation von Unsicherheit.....	37
14.1 Einleitung.....	37
14.2 Erweiterte Unsicherheit	37
14.3 Aussagen zur relativen Unsicherheit.....	38
14.4 Unsicherheitsfaktor	38
14.5 Beiträge zur Unsicherheit	38
14.6 Anwendbarkeit der Schätzungen	39
14.7 Interpretation von Unsicherheitsangaben gegen Grenzwerte	39
15 Kosten für die Schätzung der Unsicherheit der Probenahme.....	40
16 Beurteilung der Zwecktauglichkeit von Messungen anhand der Unsicherheit.....	41
17 Auswirkungen auf die Planung von Probenahme- und Messstrategien.....	43
17.1 Fachwissen und Beratung.....	43
17.2 Probenahme-Bias vermeiden	43
17.3 Planung für Unsicherheitsschätzungen	44

17.4 Kriterien für die Zwecktauglichkeit	44
17.5 Verwendung von Daten aus früheren Validierungen.....	44
17.6 Angemessenheit der Probenahmeunsicherheit.....	44
17.7 Unsicherheitsschätzung für Vor-Ort- und In-situ-Messungen	44
Anhang A: Beispiele.....	46
Beispiel A1: Nitrat in Gewächshaus-Kopfsalat	47
Beispiel A2: Blei in kontaminiertem Oberboden.....	53
Beispiel A3: Im Grundwasser gelöstes Eisen	63
Beispiel A4: Vitamin A in Babybrei mit Obst und gemahlenem Getreide.....	71
Beispiel A5: Enzyme in Hühnerfutter	88
Beispiel A6: Cadmium und Phosphor in landwirtschaftlichem Oberboden mit dem Modellansatz	91
Anhang B: Terminologie	102
Anhang C: Nützliche statistische Verfahren.....	109
C.1 Schätzung des Bias zwischen zwei Probenahmeverfahren unter Verwendung gepaarter Proben.....	109
C.2 Weitere Beschreibung des Probenahmefehlers durch die Probenahmethorie.....	111
C3. Quellen für Berechnungssoftware	114
Anhang D: Alternative experimentelle Designs für empirische Unsicherheitsschätzungen	115
1. Mehrstufige Designs zur Schätzung sonstiger Komponenteneffekte	115
2. Vereinfachtes und nicht-ausgewogenes Design zur Reduzierung der Kosten.....	116
Anhang E: Änderung der Probenahmeunsicherheit durch Vorhersagen aus der Probenahmethorie	119
Literatur	120

Liste der Akronyme und Abkürzungen

ANOVA	analysis of variance - Varianzanalyse
AQC	analytical quality control – Analytische Qualitätskontrolle
BIPM	Bureau International des Poids et Mesures – Internationales Büro für Maße und Gewichte
CEN	European Committee for Standardization - Europäisches Komitee für Normung
CH	constitution heterogeneity*
CRM	certified reference material – Zertifiziertes Referenzmaterial
CTS	collaborative trial in sampling - Probenahmevergleichsuntersuchung
df	degrees of freedom - Freiheitsgrade
DH	distribution heterogeneity*
FAPAS	Markenname einer Organisation, die internationale Ringversuche organisiert
FSE	fundamental sampling error*
GEE	global estimation error*
GFAAS	graphite furnace atomic absorption spectrometry – Graphitrohr-Atomabsorptionsspektrometrie
GSE	grouping and segregation error*
GUM	Guide to the Expression of Uncertainty in Measurement - Leitfaden zur Angabe der Unsicherheit beim Messen
HPLC	high performance liquid chromatography – Hochleistungsflüssigkeitschromatographie
IDE	increment delimitation error*
IEC	International Electrotechnical Commission - Internationale Elektrotechnische Kommission
IFCC	International Federation of Clinical Chemistry and Laboratory Medicine
IPE	increment and sample preparation error*
ISO	International Organization for Standardization - Internationale Organisation für Normung
IUPAC	International Union of Pure and Applied Chemistry
IUPAP	International Union of Pure and Applied Physics
IXE	increment extraction error*
LOD	limit of detection – Nachweisgrenze
LOQ	limit of quantification – Bestimmungsgrenze
MU	measurement uncertainty - Messunsicherheit
NIFES	National Institute of Nutrition and Seafood Research
NIST	National Institute of Standards and Technology
OIML	International Organization of Legal Metrology - Internationale Organisation für das gesetzliche Messwesen
PSE	point selection error*
PME	point mineralisation error*
PT	proficiency test – Eignungstest(ringversuch)
QA	quality assurance – Qualitätssicherung
QC	quality control - Qualitätskontrolle
RANOVA	robust analysis of variance – robuste Varianzanalyse
RSD	relative standard deviation – relative Standardabweichung
RST	reference sampling target – Referenzprobenahmezielobjekt
SD	standard deviation – Standardabweichung
SPT	sampling proficiency test – Probenahme-Ringversuch
SS	sum of squares – Summe der Quadrate
SWE	weighting error – Gewichtungsfehler
TAE	total analytical error – gesamter analytischer Fehler
TSE	total sampling error – gesamter Probenahmefehler

* Anm. d. Übers.: Hier und an anderen Stellen wurde der englische Fachbegriff beibehalten, da es im Deutschen keinen allgemein üblichen Begriff dafür gibt.

TEIL 1- Einführung und Anwendungsbereich

1 Einleitung

1.1 Begründung für den Leitfaden

Der Hauptzweck einer Messung besteht darin, Entscheidungen zu ermöglichen. Die Zuverlässigkeit dieser Entscheidungen hängt von der Kenntnis der **Unsicherheit** der Messergebnisse ab. Wenn die Unsicherheit von Messungen unterschätzt wird, zum Beispiel, weil die **Probenahme** nicht berücksichtigt wurde, dann könnte das zu fehlerhaften Entscheidungen mit schweren finanziellen Konsequenzen führen. Die **Zwecktauglichkeit** von Messergebnissen kann nur bewertet werden, wenn zuverlässige Schätzungen ihrer Unsicherheit vorliegen. Aus diesem Grund ist es wichtig, dass wirksame Verfahren für die Schätzung der Unsicherheiten, die sich aus allen Teilen des Messprozesses ergeben, zur Verfügung stehen. Diese muss Unsicherheiten einschließen, die sich aus allen relevanten Probenahmen und physikalischen Vorbereitungen ergeben. Urteile darüber, ob der analytische Beitrag zur Unsicherheit akzeptabel ist, können nur getroffen werden mit Kenntnis der Unsicherheit, die durch den Rest des Messverfahrens entsteht.

1.2 Ziel des Leitfadens

1.2.1 Ziel dieses Leitfadens ist es, das Grundprinzip und die praktische Anwendung der verfügbaren Methoden zur Schätzung der Unsicherheit unter Berücksichtigung des Beitrags durch die Probenahme zu erläutern. Der Leitfaden zielt nicht darauf ab, einzelne Probenahmeverfahren zu empfehlen, die oft in anderen Dokumenten oder Vorschriften vorgeschrieben sind, sondern vielmehr die Messunsicherheit zu berücksichtigen, die durch die jeweilige Vorschrift erzeugt wird.

1.2.2 Der Leitfaden zielt weiterhin darauf ab, die Bedeutung der Probenahme für das gesamte Messunsicherheitsbudget und somit auch für die Verlässlichkeit der aus den Messungen resultierenden Entscheidungen zu erläutern. Der Leitfaden erklärt nicht nur, wie die Unsicherheit geschätzt werden kann, sondern liefert auch die Begründung für die Einbeziehung der Probenahme in das gesamte Management des Messprozesses.

1.2.3 Im Gegensatz zu der Annahme, die oft für Schätzungen der Unsicherheit einer Analyseverfahren gemacht wird, sollte eine Schätzung für eine Probenahmeverfahren für eine Materialcharge nicht als automatisch auf jede nachfolgende Materialcharge anwendbar angenommen werden. Zum Beispiel kann sich abhängig vom Probenahmezielobjekt der Grad der **Heterogenität** (d.h. Inhomogenität) wesentlich verändert haben. Es wird daher eine routinemäßige Überwachung der wichtigsten Parameter der Probenahmequalität zur Untersuchung und Aktualisierung von Unsicherheitsschätzungen für nachfolgende Chargen notwendig sein.

1.3 Anwendung zur Beurteilung der Zwecktauglichkeit

Einer der Hauptvorteile der Kenntnis der Unsicherheit einer Messung besteht darin, es einem Interessenvertreter zu ermöglichen, ihre Eignung für einen bestimmten Zweck zu beurteilen. Ein richtiges Verständnis der **Unsicherheit aus der Probenahme** sollte daher in eine umfassendere Sichtweise der Zwecktauglichkeit einbezogen werden. Dies ist aus zwei

Gründen wichtig. Erstens wird sichergestellt, dass die Schätzung der Unsicherheit jeder Messung realistisch ist, wenn man sie mit dem optimalen Wert der für zuverlässige Entscheidungen erforderlichen Unsicherheit vergleicht. Angesichts dieses für die Zwecktauglichkeit erforderlichen Unsicherheitsniveaus ist es zweitens notwendig, den Aufwand (oder die Ausgaben) auf die Probenahme und die analytischen Aspekte des Messprozesses aufzuteilen, um die erforderliche Unsicherheit am wirtschaftlichsten zu erhalten. Diese Ideen werden weiterentwickelt, und ein quantitativer Ansatz zur Beurteilung der Zwecktauglichkeit durch Abwägung von Unsicherheit und Kosten wird in Abschnitt 16 eingeführt.

1.4 Vorgesehene Zielgruppe für den Leitfaden

Dieser Leitfaden richtet sich in erster Linie an Spezialisten wie Probenahmeplaner und analytische Chemiker, die die mit ihren Messergebnissen verbundene Unsicherheit abschätzen müssen. Andere Interessenvertreter sollten für bestimmte Anwendungen fachlichen Rat einholen.

1.5 Beziehung dieses Leitfadens zu anderen Dokumenten

1.5.1 Die derzeitige Vorgehensweise bei der Schätzung der Unsicherheit für ein breites Spektrum von Messungen folgt dem 'Guide to the Expression of Uncertainty in Measurement' ('GUM') [2], veröffentlicht 2008 von ISO in Zusammenarbeit mit BIPM, IEC, IFCC, IUPAC, IUPAP und OIML. Der GUM legt die erforderlichen Konzepte fest, stellt die allgemeinen Grundsätze auf und stellt ein Verfahren zur Verfügung, das auf Fälle anwendbar ist, in denen ein geeignetes Modell des Messprozesses verfügbar ist. Die Anwendung dieses Ansatzes auf die chemische Analyse wurde 1995 in einem Eurachem-Leitfaden zur 'Quantifizierung der Unsicherheit bei analytischen Messungen' [3] beschrieben und 2012 in einer dritten Auflage um die Verwendung von Validierungs- und Methodenleistungsdaten erweitert [1]. Weitere nützliche Beiträge zur praktischen Schätzung der Unsicherheit von analytischen Messungen unter Verwendung von Daten aus Gemeinschaftsstudien wurden vom Analytical Methods Committee der Royal Society of Chemistry 1995 [4] und durch ISO TC/69 im Jahre 2017 geleistet [5]. Dieser Leitfaden zur Probenahme steht im Einklang mit den im GUM festgelegten allgemeinen Grundsätzen.

1.5.2 Die Probenahmetheorie hat sich weitgehend unabhängig von der analytischen Chemie und der chemischen Metrologie entwickelt. Die Qualität der Probenahme wurde in der Probenahmetheorie im Allgemeinen mit der Auswahl einer "richtigen" Probenahmevervorschrift, geeigneter Validierung und Schulung des Probenahme-Personals (d.h. **Probenehmer**) verknüpft, um sicherzustellen, dass diese Vorschrift korrekt angewendet wird [6]. Es wird dann angenommen, dass die Stichproben repräsentativ und frei von systematischen Fehlern sind und die Varianz der durch das Modell vorhergesagten entspricht.

1.5.3 Ein alternativer Ansatz ist die Schätzung der Unsicherheit der Probenahme für typische Materialien oder für **Probenahmezielobjekte** während der Validierung der Probenahmevervorschrift, und die laufende Bestätigung der Leistungsdaten (in der Praxis) durch geeignete Qualitätskontrollmaßnahmen. Diese Vorgehensweise entspricht eher den Verfahren, die für den Rest des Messprozesses bereits vorhanden sind. Interessanterweise ist die Qualität der Probenahme nur durch die Messungen, die an den resultierenden Proben vorgenommen werden, quantifizierbar.

1.5.4 Es wurden Probenahmenvorschriften verfasst, um das anzuwendende Verfahren für die Probenahme von unzähligen Materialarten und für viele verschiedene chemische Komponenten zu beschreiben. Diese Vorschriften sind manchmal in Regelwerken oder in internationalen Vereinbarungen festgelegt^a [7]. Diese Verfahren identifizieren dabei selten die jeweiligen Beiträge von Probenahme und chemischer Analyse zur kombinierten Unsicherheit^b.

1.5.5 Dementsprechend gibt es eine Reihe früherer Literatur über die Theorie und Praxis der Probenahme. Wie in Abschnitt 1.2.1 erläutert, strebt dieser Leitfaden daher nicht an, weitere Probenahmenvorschriften vorzuschlagen, sondern vielmehr Methoden zur Quantifizierung der auftretenden Unsicherheit bereitzustellen, wenn eine bestimmte Vorschrift genutzt wird.

1.5.6 Ein Handbuch, das Verfahren für die Schätzung der Unsicherheit der Probenahme, abgeleitet aus diesem Leitfaden, aber mit weiteren Fallstudien unterlegt beschreibt, wurde von der Nordtest Gruppe erstellt [8].

1.6 Verwendung des Leitfadens

1.6.1 Dieses Dokument fasst die Konzepte zusammen, die zum Verständnis von Unsicherheit bei Probenahme notwendig sind und stellt Verfahren zur Verfügung, die deren praktische Umsetzung ermöglichen. Der Leitfaden behandelt zusätzlich Fragen im Zusammenhang mit Management, Qualitätssicherung und Berichterstattung von Ergebnissen und deren Unsicherheit. Der Geltungsbereich und das beabsichtigte Einsatzgebiet sind in Abschnitt 2 dargelegt, der auch die behandelten Ansätze zusammenfasst. Die Terminologie wird in Abschnitt 3 erörtert und Schlüsselbegriffe werden in Anhang B definiert.

1.6.2 Grundlegende Konzepte werden in den Abschnitten 4 und 5 behandelt. Ein Überblick über den Messprozess wird in Abschnitt 4 beschrieben. Dazu gehört eine Erläuterung der für Probenahme verwendeten Terminologie und die Angabe, welche Prozessschritte in diesem Leitfaden detailliert betrachtet werden. Die Messunsicherheit und ihre Quellen werden in Abschnitt 5 weiter erörtert.

1.6.3 In den Abschnitten 6 bis 10 werden Methoden zur Schätzung der Unsicherheit inklusive Diskussion der Vorzüge der verschiedenen Optionen beschrieben. Die beiden verfügbaren grundlegenden Ansätze sind in Abschnitt 6 zusammengefasst und in den Abschnitten 9 und 10 ausführlich behandelt. Die Absicht ist, eine Reihe von Optionen zu bieten, die angewandt werden können, und nicht einen bestimmten Ansatz.

1.6.4 Management- und Qualitätsfragen werden in den Abschnitten 11 bis 13 behandelt. Diese beinhalten eine sehr kurze Erörterung der Zuständigkeiten für die Qualität bei der Probenahme (Abschnitt 11), bevor in Abschnitt 12 die Auswahl des Ansatzes zur Schätzung der Unsicherheit diskutiert wird. Der Einsatz Qualitätskontrolle der der Probenahme zur Überwachung der Leistungsfähigkeit der Probenahme wird in Abschnitt 13 behandelt. Berichterstattung und Verwendung von Unsicherheit und ihre Auswirkungen auf die

^a Die 'Annahmestichprobenprüfungen' werden bei der Probenahme einer breiten Palette von Materialien angewandt [7].

^b Einige Konzepte der Probenahmetheorie können für die Schätzung der Messunsicherheit sinnvoll angepasst werden (Abschnitt 10.2).

Verlässlichkeit von Entscheidungen werden in Abschnitt 14 erörtert. Kosten sind ein wichtiger Faktor und die Auswahl der kostengünstigsten und geeignetsten Methode zur Schätzung wird in Abschnitt 15 erläutert. Die Kenntnis des Wertes der Unsicherheit hilft bei der Beurteilung der Zwecktauglichkeit der Messung als Ganzes und ihrer Bestandteile. Dies wird in Abschnitt 16 diskutiert.

1.6.5 Eine Reihe von Beispielen, ein detailliertes Glossar der in diesem Leitfaden verwendeten Begriffe und Definitionen, einige wichtige statistische Verfahren und Versuchspläne sowie eine Diskussion zur Verbesserung der Probenahmeunsicherheit unter Verwendung von Vorhersagen aus der Probenahmetheorie werden als Anhänge bereitgestellt.

2 Geltungsbereich und Einsatzgebiet

2.1. Die Prinzipien dieses Leitfadens sind auf die Schätzung der Unsicherheit für das gesamte Spektrum an Materialien anwendbar, die einer analytischen Messung unterzogen werden (z.B. gasförmig, flüssig und fest). Dazu gehören Umweltmaterialien (z.B. Gestein, Boden, Wasser, Luft, Abfall und Biota), Lebensmittel, industrielle Materialien (z.B. Rohstoffe, Zwischenprodukte und Produkte), forensische Materialien und Pharmazeutika. Dieser Ansatz ist auf die Probenahme nach einer beliebigen Vorschrift anwendbar, unabhängig davon, ob es sich um Einzel- oder Mischproben, Einzel- oder Mehrfachbestimmungen handelt.

2.2. Der Leitfaden beschreibt die Schätzung der Unsicherheit durch i) wiederholte Messungen und Probenahme (der 'empirische Ansatz') sowie durch ii) Modellierung auf der Grundlage identifizierter Einflussgrößen und theoretischer Überlegungen (der "Modellierungs-Ansatz").

2.3. Die Verwendung von Unsicherheitsschätzungen bei der Beurteilung der Zwecktauglichkeit und bei der Optimierung des Aufwands zwischen den einzelnen Schritten des Messprozesses wird behandelt. Es werden Methoden der Beurteilung der Zwecktauglichkeit beschrieben, die auf einem Prozentsatz der Gesamtvarianz basieren, und solche, die auf der Grundlage einer Kosten-Nutzen-Analyse beruhen.

2.4. Obwohl die allgemeinen Grundsätze dieses Leitfadens gelten, wird darin nicht speziell auf mikrobiologische Probenahme eingegangen. Auch die Schätzung der Unsicherheit aus räumlichen oder zeitlichen Informationen, wie Lage oder Größe von Gebieten mit hoher Analytkonzentration, wird nicht erörtert.

3 Terminologie

3.1. Die genauen Definitionen vieler der in diesem Leitfaden verwendeten Begriffe variieren je nach Anwendungsgebiet. Eine vollständige Auflistung der Begriffe und ihrer verschiedenen Definitionen findet sich in Anhang B. In diesem Leitfaden wurden normative Definitionen der einzelnen Begriffe ausgewählt, die so allgemeingültig wie möglich für alle Anwendungsbereiche sind. Die verwendeten Begriffe sind in Anhang B aufgeführt und bei der ersten Verwendung im Text fett gedruckt .

3.2. Die analytische Wissenschaft befasst sich im Allgemeinen mit der Zusammensetzung (von Proben), daher ist die Messgröße in der Regel auf die Menge oder den Anteil eines Analyten in einer Mischung bezogen. Dies kann auf verschiedene Weise ausgedrückt werden, einschließlich (zum Beispiel) Massenanteil, Volumenanteil, Stoffmengen-(Mol-)anteil sowie Masse oder Stoffmenge pro Volumeneinheit. Leider gibt es keinen weithin allgemein anerkannten Begriff, der alle diese Bedeutungen abdeckt. Daher sollte in diesem Leitfaden der Begriff "Konzentration", wenn er nicht eingeschränkt ist, als auf jedes dieser verschiedenen Maße von Anteilen oder Mengen anwendbar verstanden werden. Wenn der Text eine eingeschränkte Interpretation verlangt, wird "Konzentration" eingeschränkt (zum Beispiel als "Stoffmengenkonzentration") oder durch einen spezifischeren Begriff ersetzt (z.B. "Massenanteil").

TEIL 2 – Grundlegende Konzepte

4 Probenahme im Messprozess

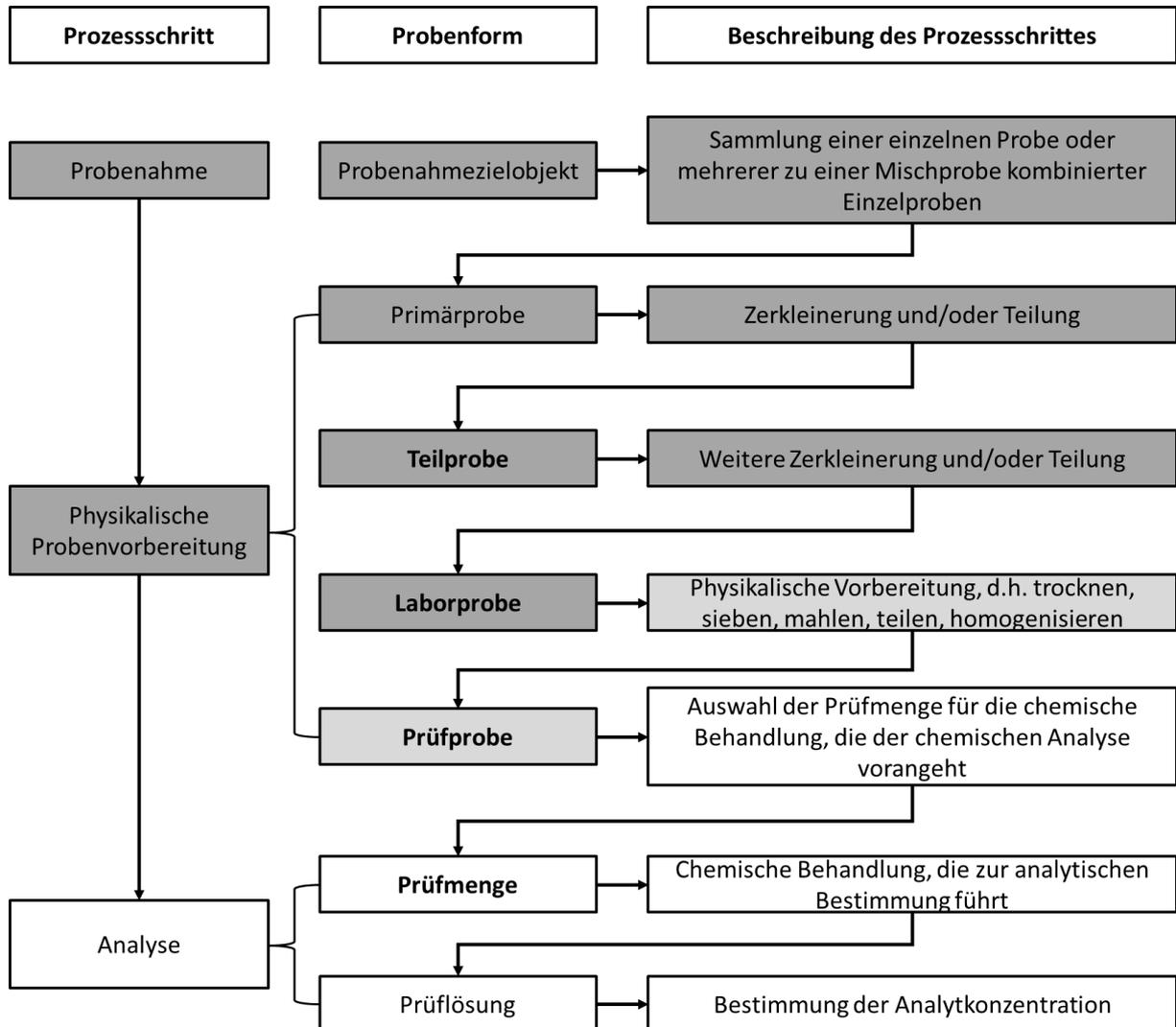


Abb. 1: Schematische Darstellung eines typischen Messprozesses

Die Abbildung zeigt einen vollständigen Messprozess, beginnend mit der Primärprobenahme und endend mit der analytischen Bestimmung. Es gibt viele Zwischenschritte, wie den Transport und die Konservierung von Proben, die nicht alle immer vorhanden sind. Jeder Schritt führt zu einem Beitrag zur Messunsicherheit. Dieser Leitfaden konzentriert sich auf die Prozessschritte der Probenahme und der physikalischen Probenvorbereitung (schattierte Kästen), da der letzte Schritt in früheren Leitlinien ausreichend behandelt wird [1]. Es ist zu beachten, dass zwei der Probenahmeschritte innerhalb des Labors stattfinden (hellgrau) und häufig als Teil des Analyseprozesses betrachtet werden. Für Definitionen der Begriffe siehe Anhang B.

Ein Probenahmezielobjekt ist der Teil des Materials zu einem bestimmten Zeitpunkt, den die Probe (und damit das Messergebnis) repräsentieren soll. Das Probenahmezielobjekt muss vor der Entwicklung des Probenahmeplans definiert werden. Es kann durch eine Verordnung definiert werden, wie z.B. die Gesamtheit einer Charge, eines Loses oder einer Lieferung. Wenn die Eigenschaften und Merkmale (z.B. Analytkonzentration) des Materials in einem

bestimmten Gebiet oder Zeitabschnitt von Interesse sind und ermittelt werden müssen, dann kann dies als Probenahmezielobjekt betrachtet werden. Wenn die Zusammensetzung einer ganzen Charge erforderlich ist (z.B. eines Lebensmittelrohstoffs), dann stellt die gesamte Charge das Zielobjekt dar. Wenn die räumliche (oder zeitliche) Variation der Konzentration benötigt wird (z.B. beim Auffinden von 'Hot Spots' innerhalb eines kontaminierten Materials), dann ist jeder Ort, an dem die Konzentration benötigt wird, ein separates Probenahmezielobjekt. Jedes einzelne Probenahmezielobjekt erzeugt letztendlich ein berichtetes Messergebnis und eine Unsicherheit.

Primärproben bestehen oft aus mehreren **Einzelproben**, die zu einer **Mischprobe** zusammengefasst werden, bevor eine Messung durchgeführt wird. Es wird die Unsicherheit dieses einzelnen Messwerts, der an dieser Mischprobe vorgenommen wurde, die durch alle erforderlichen vorbereitenden Schritte erzeugt wurde, benötigt. Der Wert dieser Unsicherheit wird oft durch die Anzahl der genommenen Einzelproben beeinflusst. Dies steht im Gegensatz zu der Situation, in der mehrere individuelle Primärproben (n) aus verschiedenen Teilen des Probenahmezielobjekts entnommen und getrennt gemessen werden. Wenn die Zusammensetzung des Probenahmezielobjekts unter Berücksichtigung des Mittelwertes dieser Einzelmessungen berechnet wird, dann wird die Unsicherheit des Mittelwertes mit Hilfe der Standardabweichung des Mittelwertes (s/\sqrt{n}) berechnet. Dies ist nicht dasselbe wie die Unsicherheit der einzelnen Messung, deren Schätzung das Ziel dieses Leitfadens ist.

4.1. Der gesamte Messprozess (Abbildung 1) beginnt typischerweise mit der Entnahme der Primärprobe aus einem Probenahmezielobjekt. Die resultierende Probe durchläuft vor der analytischen Bestimmung einen oder mehrere Schritte. Alle Schritte tragen zur Unsicherheit des Endergebnisses bei, wenn der geforderte Analytwert (d.h. **Messgröße** oder wahrer Wert) als Analytkonzentration im Probenahmezielobjekt ausgedrückt wird. Es existieren bereits Anleitungen zur Schätzung der analytischen Schritte des Messprozesses [1]. Diese schließen sicherlich die Auswahl der Prüfmenge, die der Messung vorausgehende chemische Behandlung und die analytische Bestimmung ein, können aber auch die physikalische Vorbereitung der Laborprobe durch Trocknen, Sieben, Mahlen, Teilen und Homogenisierung beinhalten.

4.2. Es ist gängige Praxis, all die verschiedenen Probenformen in der zweiten Spalte von Abbildung 1 oft einfach als "Probe" zu bezeichnen. Es ist zweifellos wichtig, sie in der Diskussion sorgfältig zu unterscheiden, insbesondere diejenigen, die in diesem Leitfaden besonders berücksichtigt werden (in schattierten Kästchen in Abbildung 1). Darauf wird in Abschnitt 5.2 näher eingegangen.

4.3. Die im Leitfaden beschriebenen Methoden helfen bei der Identifizierung der Hauptquelle der Unsicherheit, wie z.B. die Probenahme eher als die chemische Analyse, aber sie erklären nicht unbedingt die Ursache. Allerdings ist die Heterogenität innerhalb des Probenahmezielobjekts, sei sie räumlich oder zeitlich, als signifikanter Beitrag der Unsicherheit bekannt. Gesonderte Studien wären erforderlich, um die Variabilität, die zur Unsicherheit beiträgt, zu beschreiben. Für die Zwecke dieses Leitfadens wird die Heterogenität innerhalb des Probenahmezielobjekts als nur eine Ursache für die Unsicherheit bei der abschließenden Messung behandelt. Dies ist der Fall, unabhängig davon, welche

Maßnahmen ergriffen werden, um die Auswirkungen der Heterogenität durch die Anwendung einer bestimmten Probenahmenvorschrift zu minimieren.

5 Messunsicherheit

5.1 Definition der Messunsicherheit

5.1.1 Die Unsicherheit einer Messung oder Messunsicherheit (MU) ist in der metrologischen Terminologie [2] definiert als:

ein Parameter, assoziiert mit dem Ergebnis einer Messung, der die Streuung der Werte charakterisiert, die sinnvollerweise der Messgröße zugeordnet werden können.

Die Definition umfasst mehrere wichtige Merkmale, die in den folgenden Absätzen erörtert werden.

5.1.2 Der "Parameter" kann z.B. ein Bereich, eine Standardabweichung, ein Intervall (wie ein Konfidenzintervall) oder ein Halbinsintervall ($\pm u$ ist die Angabe eines Halbinsintervalls) oder ein anderes Maß für Streuung, wie z.B. eine relative Standardabweichung sein. Es ist zu beachten, dass, wenn MU als Standardabweichung ausgedrückt wird, der Parameter als "Standardunsicherheit" bezeichnet wird, üblicherweise mit dem Symbol u . Andere Ausdrucksformen werden in Abschnitt 14 betrachtet.

5.1.3 Mit jedem Messergebnis ist Unsicherheit "assoziiert". Ein vollständiges Messergebnis enthält typischerweise eine Angabe der Unsicherheit in der Form $x \pm U$, wobei x das Messergebnis und U eine Angabe der Unsicherheit ist (man wird feststellen, dass das Symbol U in diesem Leitfaden eine besondere Bedeutung hat; es zeigt eine "erweiterte Unsicherheit" an, die in Abschnitt 14 weiter erörtert wird). Diese Form, ein Ergebnis auszudrücken, ist ein Hinweis für den Endnutzer des Ergebnisses, dass das Ergebnis bedeutet, der Messwert liegt mit hinreichender Verlässlichkeit innerhalb dieses Intervalls.

5.1.4 Die Messgröße ist einfach eine Größe, wie z.B. eine Länge, eine Masse oder eine Konzentration eines Stoffes, die gemessen wird. Der Begriff 'Wert der Messgröße' ist eng verbunden mit dem traditionellen Konzept des "wahren Wertes" in der klassischen statistischen Terminologie. Unter diesem alternativen Blickwinkel ist "Unsicherheit" historisch definiert^c worden als

Eine einem Prüfergebnis beigefügte Schätzung, die den Wertebereich charakterisiert, innerhalb dessen der wahre Wert liegen soll.

Diese Definition (die im Folgenden als statistische Definition bezeichnet wird) hat den Vorteil, dass sie Entscheidungsträgern leichter zu erklären ist, da diese oft den Begriff "wahrer Wert" als den für ihre Entscheidung interessierenden Wert anerkennen. Sie hat den Nachteil, dass der wahre Wert selbst nie bekannt sein kann und dies im Allgemeinen einer weiteren Erklärung bedarf.

^c ISO 3534-1: 1993 Statistics – Vocabulary and Symbols, International Organization for Standardization, Genf

5.1.5 Die metrologische Definition besagt, dass die Unsicherheit "die Streuung der Werte, die der Messgröße sinnvollerweise zugeordnet werden können", ausdrückt. Dies ist eine besonders wichtige Formulierung. Sie zeigt an, dass, obwohl die Unsicherheit mit einem Messergebnis verbunden ist, der angegebene Bereich sich auf den möglichen Wertebereich der Messgröße beziehen muss. Zum Beispiel könnte die Messgröße die Gesamtmasse von Gold in einer geologischen Lagerstätte sein. Ein solcher Bereich ist deutlich verschieden von einer **Präzisions**angabe, die die Spannweite der Ergebnisse beschreiben würde, die bei einer Wiederholung der Messung beobachtet werden könnten. Wenn Informationen darüber verlangt werden, „wo der Messwert liegen könnte“, verlangt diese Definition der Unsicherheit implizit, dass der Messwissenschaftler *alle* Effekte berücksichtigt, die das Messergebnis beeinflussen könnten. Diese Effekte schließen offensichtlich die Ursachen der zufälligen Variation von einer Messung zur nächsten über den Zeitraum der Messung ein. Es ist aber auch wichtig, Quellen von **Bias** während des Experiments zu berücksichtigen, und sehr oft erzeugen diese größere Effekte, als durch wiederholte Messungen allein beobachtet werden können. Das heißt, die Messunsicherheit erfordert automatisch einen Bereich, der *sowohl* zufällige *als auch* systematische Effekte berücksichtigt.

5.1.6 Um ein einfaches analytisches Beispiel zu betrachten, beinhaltet eine einfache Messung der Konzentration in einem Feststoff typischerweise die Extraktion von Material, das Wiegen, volumetrische Tätigkeiten und vielleicht Spektrometrie oder Chromatographie. Wiederholte Messungen zeigen eine Streuung der Werte aufgrund zufälliger Schwankungen bei diesen Tätigkeiten. Aber alle Analytiker wissen, dass eine Extraktion selten vollständig ist, und dass bei einem bestimmten Material eine unzureichende Extraktion zu einem gleichbleibend niedrigen Ergebnis führt. Während eine gute analytische Praxis immer versucht, solche Effekte auf ein unbedeutendes Maß zu reduzieren, wird ein gewisser Bias bestehen bleiben. Bei der Darstellung der Unsicherheit für den Wert der Messgröße muss der Analytiker also die begründete Möglichkeit eines systematischen Fehlers durch solche Ursachen in Betracht ziehen. (Gewöhnlich geschieht dies, indem Informationen wie die Wiederfindungsrate des Analyten, die an Referenzmaterialien oder bei Aufstockungsexperimenten beobachtet wurde, berücksichtigt werden).

5.1.7 Die gleichen Überlegungen gelten im Falle einer Probenahme. Es ist bekannt, dass unterschiedliche Proben, die einem Grundmaterial entnommen werden, oft reale Wertschwankungen zeigen, was wiederholte Messungen klar belegen. Es ist auch bekannt, dass die Probenahme fehlerhaft sein kann, zum Beispiel durch unterschiedliche Materialentnahme, ungeeigneten Zeitpunkt der Probenahme, wenn zeitliche Schwankungen auftreten, oder durch Zugangsbeschränkungen. Diese Effekte beeinflussen die Beziehung zwischen dem Wert der Messgröße und dem beobachteten Ergebnis. Während eine gute Praxis bei der Probenahme darauf abzielt, diese Auswirkungen auf ein unbedeutendes Maß zu reduzieren, berücksichtigt eine sorgfältige Bewertung der Unsicherheit immer die Möglichkeit verbleibender systematischer Effekte.

5.1.8 Der aktuelle Leitfaden zur Messunsicherheit [2] macht deutlich, dass die Messunsicherheit (Abschnitt 2.2 der Referenz [2]) nicht dazu gedacht ist, "grobe Fehler" oder „Irrtümer“ zu berücksichtigen (Abschnitt 3.4.7 der Referenz [2]). Dies würde zum Beispiel Fehler durch Übertragungsfehler oder massiven unsachgemäßen Gebrauch der Messvorschrift ausschließen. Die Probenahme kann jedoch hohe Unsicherheiten (z.B. 80 %

des Konzentrationswertes) einfach durch die routinemäßige Anwendung einer anerkannten Messvorschrift auf ein sehr heterogenes Material erzeugen. Selbst wenn die Verfahren nominell korrekt sind, wird es aufgrund von Mehrdeutigkeit in den Messvorschriften und der geringfügigen Anpassungen, die in realen Probenahmesituationen an den Vorschriften vorgenommen werden, auch leichte Abweichungen in den tatsächlichen Verfahren geben. Ob dieses hohe Maß an Unsicherheit zu einem inakzeptablen Maß an Verlässlichkeit bei den darauf basierenden Entscheidungen führt, hängt von einer gründlichen Bewertung der Zwecktauglichkeit ab (siehe Abschnitt 16).

5.2 Spezifizierung der Messgröße

5.2.1 Wenn einem Endnutzer ein Konzentrationsergebnis für eine Gesamtprobe in der Form ' $x \pm U$ ' vorgelegt wird, wird er dieses Intervall ganz selbstverständlich so interpretieren, dass es den Bereich der Werte, die der Konzentration im Probenahmezielobjekt (z.B. einer Materialcharge) zuzuordnen sind, enthält. Diese Vorstellung impliziert, dass die Messgröße "die Konzentration (des Analyten) in der Materialcharge" ist, und dass die Unsicherheit jede notwendige Berücksichtigung von Heterogenität in der Gesamtprobe beinhaltet. Der Analytiker hingegen könnte sich auf "die Konzentration in der analysierten Laborprobe" beziehen, womit implizit die Abweichung zwischen Laborproben ausgeschlossen wird. Es ist klar, dass die eine Sichtweise die Auswirkungen der Probenahme einschließt, während die andere dies nicht tut. Die Auswirkung auf die Unsicherheit kann natürlich sehr beträchtlich sein. Aus metrologischer Sicht ergibt sich diese Unterscheidung, weil die beiden Ansichten unterschiedliche Messgrößen berücksichtigen. Die eine betrachtet „die Konzentration im Probenahmezielobjekt“, die andere "die Konzentration in der Laborprobe". Ein weiteres Beispiel wäre die "Schadstoffkonzentration in einem Fabrikabfluss zum Zeitpunkt der Probenahme", verglichen mit „der durchschnittlichen Schadstoffkonzentration über ein Jahr“.

5.2.2 Diese Mehrdeutigkeiten bei der Interpretation können nur durch eine sorgfältige Spezifizierung der Messgröße vermieden werden. Die Angabe der Maßeinheit (Masse, Länge, Konzentration usw.) ist zweifelsohne notwendig. Ebenso wichtig ist es, sich über den Anwendungsbereich der Messung klar zu werden, indem Informationen über Faktoren wie die Zeit, den Ort oder die Grundgesamtheit, auf die voraussichtlich das Messergebnis angewendet wird, einbezogen werden. Einige besondere Beispiele für die Spezifikation von Messgrößen und ihre Auswirkungen auf die Schätzung der Unsicherheit werden im Folgenden erörtert.

Es ist nie möglich, alle Unklarheiten bei der Verwendung der Formulierungen in der Probenahmevervorschrift zu vermeiden.

5.2.3 Wenn eine Mischprobe durch die Vereinigung mehrerer über ein Probenahmezielobjekt verteilter Einzelproben entnommen und als eine einzelne Primärprobe analysiert wird, liefert diese einzelne Bestimmung der Analytkonzentration eine Schätzung des Wertes der Messgröße (d.h. die durchschnittliche Zusammensetzung des Zielobjekts), wie in Abschnitt 4 kurz besprochen. Die Unsicherheit bei diesem Einzelwert spiegelt die Unsicherheit bei der Schätzung des Messgrößenwertes wider. Werden im Gegensatz dazu mehrere unabhängige Primärproben dem Zielobjekt entnommen, jede einmal analysiert, und der Mittelwert berechnet, so ist dieser Mittelwert ebenfalls ein Schätzwert für den Wert der Messgröße. Die Unsicherheit wird jedoch nicht die der Messung sein (ausgedrückt als

Standardabweichung, s), sondern der Standardfehler des Mittelwertes (ausgedrückt als s/\sqrt{n}). Diese letztere Unsicherheit des Mittelwertes kann durch die Entnahme von mehr Primärproben^d verringert werden, die Unsicherheit der Messung dagegen nicht.

5.3 Fehler, Präzision und Unsicherheit

5.3.1 Unsicherheit hängt mit anderen Begriffen zusammen, wie z.B. Genauigkeit, Fehler, Richtigkeit, Bias und Präzision. In anderen Leitfäden werden die Zusammenhänge im Detail erörtert [1,2]. Es lohnt sich jedoch, einige der wichtigen Unterschiede zu wiederholen:

- Unsicherheit ist ein Bereich von Werten auf der Grundlage des Messergebnisses und anderer bekannter Effekte, wohingegen Fehler als Differenz zwischen einem einzelnen Ergebnis und einem „wahren (oder Referenz-)Wert“ definiert ist. Bias kann als durchschnittlicher Fehler angesehen werden.
- Unsicherheit bezieht die Berücksichtigung aller Effekte, die ein Ergebnis beeinflussen können (d.h. sowohl zufällige und systematische Fehler) ein; Präzision umfasst nur die Effekte, die während der Beobachtungen variieren (d.h. nur einige zufällige Fehler).
- Die Unsicherheit ist für die korrekte Anwendung von Mess- und Probenahmeverfahren gültig, aber, wie in Abschnitt 5.1.8 erwähnt, ist sie nicht dazu bestimmt, grobe Bedienungsfehler zu berücksichtigen.

5.4 Probenahme und physikalische Vorbereitung als Quellen der Messunsicherheit

5.4.1 Die Entnahme einer Probe führt zu Unsicherheiten beim berichteten Messergebnis, wo immer das Ziel der Messung in Bezug auf die Analytkonzentration im Probenahmezielobjekt und nicht einfach in der Laborprobe definiert ist.

5.4.2 Probenahmeverfahren sind insofern nie perfekt, als sie niemals die erforderlichen Maßnahmen beschreiben können, die der Probenehmer für jede mögliche Eventualität benötigt, die in der realen Welt, in der die Probenahme stattfindet, auftreten kann. Der räumliche (oder zeitliche) Punkt für die Entnahme einer Probe ist selten genau spezifiziert (z.B. auf den Millimeter oder die Sekunde genau). Der Probenehmer muss solche Entscheidungen (idealerweise nach objektiven Kriterien) treffen, aber da Heterogenität (in Raum oder Zeit) unvermeidlich ist, werden diese Entscheidungen die geschätzte Konzentration beeinflussen. Ein Verständnis dieser Unsicherheitsquellen ist bei der Entwicklung und Umsetzung von Methoden zur Schätzung von Unsicherheit wichtig. Wenn z.B. **Doppelproben** genommen werden, kann es sein, dass die Probenahme an genau der *gleichen* Stelle und zur *gleichen* Zeit möglicherweise nicht die Unsicherheit der Messung widerspiegelt, die tatsächlich existiert. Dies wird in der Beschreibung der Schätzmethoden (Abschnitte 6 bis 10) und in den verschiedenen Arbeitsbeispielen (Anhang A) weiter ausgeführt.

^d Unter der Annahme, dass die Proben zufällig und unabhängig sind, und unter der Annahme eines Bias von Null.

5.4.3 Heterogenität gibt immer Anlass zu Unsicherheit. Wenn das Probenahmezielobjekt perfekt homogen wäre, dann wäre dieser Beitrag gleich Null, aber fast alle Materialien sind in gewissem Maße und auf einer bestimmten Größenskala heterogen. Beträgt die Prüfmenge einige Mikrogramm, dann ist fast jedes Material heterogen, und die Probenahme trägt zur Unsicherheit bei der Messung einer Analytkonzentration bei. Die Heterogenität kann in einem separaten Experiment quantifiziert werden, aber wenn das Ziel darin besteht, die Analytkonzentration in einem größeren Probenahmezielobjekt zu schätzen, dann ist diese Heterogenität nur eine Ursache für die Messunsicherheit (wie in den Abschnitten 4.3 und 5.6 ausgeführt).

5.4.4 Ähnliche Argumente können für die Unsicherheit im Rahmen der Prozesse der physikalischen Aufbereitung (z.B. Transport, Konservierung, Zerkleinerung, Teilung, Trocknung, Siebung, Homogenisierung) angeführt werden, die nach der Probenahme und vor jeder chemischen Behandlung der Prüfprobe stattfinden (Abbildung 1). Jeder Schritt kann Fehler aus einer Reihe von Mechanismen einbringen, wie z.B. Verlust des Analyten, Verlust von Feinpartikeln oder Kontamination durch Geräte oder vorhergehende Proben. Die angewandten Methoden und die erfolgte Ausbildung sollten darauf abzielen, diese Fehler auf ein Minimum zu reduzieren. Darüber hinaus sind jedoch Verfahren erforderlich, um die Unsicherheit zu schätzen, die alle diese Schritte, wenn sie in der Praxis angewendet werden, im endgültigen Messwert erzeugen.

5.5 Unsicherheitsquellen

5.5.1 Unsicherheit entsteht aus einer Vielzahl von Quellen, die auf unterschiedliche Weise kategorisiert wurden. Zum Beispiel identifiziert der Eurachem-Leitfaden zur Unsicherheit acht wichtige Kategorien von Effekten, die für die Schätzung der Unsicherheit wichtig sind [1], von denen die ersten beiden Probenahme und Probenvorbereitung sind. Spezifische Effekte, die in diesen beiden Kategorien identifiziert werden können, sind in Tabelle 1 dargestellt. Ein Modellierungsansatz könnte diese Effekte als Grundlage für ein mathematisches Modell nutzen. Alternativ identifiziert die Probenahmetheorie acht verschiedene Fehlerquellen bei der Probenahme (Tabelle 2); jede von diesen kann auch auf eine Vielzahl von Kausalfaktoren zurückgeführt werden, die wiederum in verschiedenen Modellierungsansätzen verwendet werden können. Ein weiterer alternativer Ansatz besteht darin, alle Schritte im Messprozess (Abbildung 1) als Quellen der Unsicherheit zu betrachten, die einen gewissen Beitrag zur Unsicherheit der endgültigen Messung leisten. In diesem Leitfaden behandeln die einfachsten Studiendesigns die Unsicherheit als aus vier Wirkungsklassen resultierend (Tabelle 3), und diese Klassen werden in einem einfachen statistischen Modell als Unsicherheitsquellen behandelt; dies ist konsistent mit der in Referenz [1] ausdrücklich vorgeschlagenen Kategorisierung von Unsicherheitsquellen. In ihrer einfachsten Form kann diese Kategorisierung auf zwei Kategorien reduziert werden: „**Probenahmeunsicherheit**“ und „analytische Unsicherheit“.

5.5.2 Das wichtige Merkmal jeder dieser verschiedenen Klassifizierungen ist, dass jede sicherstellen soll, dass, wie auch immer sie kategorisiert und bewertet werden, alle praktisch wichtigen Effekte bei der Schätzung der Unsicherheit berücksichtigt werden. Solange diese Voraussetzung erfüllt ist, kann jedes Kategorisierungsschema zur Schätzung der Unsicherheit

angewendet werden. Die in Tabelle 2 und Tabelle 3 aufgeführten Kategorisierungsschemata decken alle praktisch wichtigen Effekte ab.

5.5.3 Jede unterschiedliche Kategorisierung der Quellen führt im Allgemeinen zu einem anderen Studiendesign und sehr oft zu grundlegend unterschiedlichen Methoden der Bewertung von Unsicherheitsbeiträgen. Dies führt zu im Wesentlichen unabhängigen Schätzungen der Unsicherheit über verschiedene Ansätze. Wie an anderer Stelle [5] festgestellt wurde, deuten extrem unterschiedliche Schätzungen der Unsicherheit für das gleiche System darauf hin, dass mindestens eine Studienmethodik fehlerhaft ist. Dies bildet die Grundlage für eine Überprüfung der Gültigkeit eines Ansatzes. Daher wird, wo praktikabel, ein Vergleich der Unsicherheitsschätzungen, die sich aus unabhängigen Bewertungsansätzen ergeben, als ein Mittel zur Validierung bestimmter Schätzungen und zur Beurteilung der Gültigkeit verschiedener Ansätze empfohlen.

5.6 Heterogenität als Unsicherheitsquelle

5.6.1 Die IUPAC definiert derzeit sowohl Homogenität als auch Heterogenität als "Der Grad bis zu dem eine Eigenschaft oder ein Bestandteil gleichmäßig über eine Materialmenge verteilt ist." ([9] Definitionen siehe Anhang B). So definiert, gehört die Heterogenität zu den wichtigsten Faktoren, die zu der mit der Probenahme verbundenen Unsicherheit beitragen. Einzelproben von verschiedenen Orten im Probenahmezielobjekt haben unterschiedliche Konzentrationen des Analyten in einem heterogenen Material und es gibt eine Probe-zu-Probe-Variation der Analytkonzentration - in der Regel sichtbar als ein Beitrag zur beobachteten Variation der Ergebnisse. Im Allgemeinen ist die genaue Abhängigkeit der Konzentration vom Ort unbekannt, so dass keine Korrektur vorgenommen werden kann. Dies führt zu Unsicherheiten bei jedem beliebigen Ergebnis oder allgemein bei jedem Durchschnitt solcher Ergebnisse.

5.6.2 Die IUPAC stellt als Zusatz zu der obigen Definition [9] fest, dass "der Grad der Heterogenität (das Gegenteil von Homogenität) der entscheidende Faktor des Probenahmefehlers ist". Die Anmerkung ist ein guter Hinweis auf die Bedeutung der Heterogenität bei der Probenahme. Es gibt andere Fehler- und Unsicherheitsquellen bei der allgemeinen Durchführung von Probenahmen; zum Beispiel Kreuzkontamination und unvollkommene Stabilisierung der Proben, die beide zu (unbekanntem) Bias oder zusätzlicher Variabilität führen können. Doch Heterogenität und ihre Auswirkungen - wie z.B. Zufallsvariabilität und Stichproben-Bias - bleiben das größte Problem bei ordnungsgemäß durchgeführter Probenahme und werden generell die bedeutendste Quelle der Unsicherheit bilden.

5.6.3 Eine alternative Definition von Homogenität wird manchmal für partikelförmiges Material verwendet, das, wenn es aus Partikeln unterschiedlicher Materialien besteht, niemals "homogen" im Sinne der IUPAC-Definition sein kann. In diesem Zusammenhang kann eine Mischung, bei der die Wahrscheinlichkeit der Auswahl von verschiedenen Arten von Partikeln im gesamten Probenahmezielobjekt konstant ist, als „homogen“ bezeichnet werden, um anzuzeigen, dass die erwartete Konzentration in Proben, die an einem beliebigen Punkt im Materials entnommen wurden, die gleiche sein wird. Doch auch hier muss anerkannt werden, dass die partikuläre Beschaffenheit des Materials zu Probe-zu-Probe-Variationen

aufgrund leicht unterschiedlicher Zusammensetzung der tatsächlich entnommenen Einzelproben führt; die Heterogenität, wie sie von der IUPAC definiert wird, hat unter diesen Umständen immer noch eine Auswirkung und trägt folglich weiterhin zur Unsicherheit bei.

5.7 Variation der Unsicherheit mit dem Konzentrationsbereich

Vor der Berechnung der Unsicherheit der Probenahme, wie in Abschnitt 6 beschrieben, sollte die Variation der Unsicherheit mit dem Konzentrationsbereich in Betracht gezogen werden. Wenn es einen großen Konzentrationsbereich gibt, kann die Unsicherheit (ausgedrückt in Konzentrationseinheiten) mit der Konzentration zunehmen, und die Berechnungen werden vorzugsweise an log-transformierten Daten durchgeführt. Die Unsicherheit in relativen Einheiten auszudrücken ist nicht geeignet, wenn die Analytkonzentration innerhalb eines Faktors zehn der Nachweisgrenze liegt. Bei einer großen Anzahl von Doppelmessungen (z.B. > 50) kann es möglich sein, ein Modell der Probenahmeunsicherheit in Abhängigkeit von der Konzentration anzupassen [10, 11, 12]. In den Beispielen A1 und A4 ist der Konzentrationsbereich eng, und die Berechnungen werden mit den Daten in Konzentrationseinheiten durchgeführt, d.h. es wird keine Transformation der Daten vorgenommen. Im Beispiel A2 wird die Unsicherheit unter Verwendung von Daten sowohl in Konzentrationseinheiten als auch nach Log-Transformation berechnet und die Ergebnisse verglichen. Im Beispiel A3 werden die Berechnungen in absoluten Einheiten durchgeführt, und im Beispiel A5 in relativen Einheiten, d.h. in Prozent. In der Realität ist es oft schwierig zu wissen, ob und wie die Unsicherheit mit der Konzentration variiert und welche Methode der Schätzung daher am besten geeignet ist.

Tabelle 1: Einige Unsicherheitsquellen bei Probenahme und Probenvorbereitung, übernommen aus Referenz [1]

Probenahme	Probenvorbereitung
<ul style="list-style-type: none"> - Heterogenität (oder Inhomogenität) - Auswirkungen der spezifischen Probenahme-Strategie (z. B. zufällig, geschichtet, proportional etc.) - Auswirkungen von der Bewegung des Massengutes (insbesondere Dichteselektion) - Aggregatzustand (fest, flüssig, gasförmig) - Auswirkungen von Temperatur und Druck - Auswirkungen des Probenahmeprozesses auf die Zusammensetzung (z. B. unterschiedliche Absorption im Probenahmesystem) - Transport und Konservierung der Probe 	<ul style="list-style-type: none"> - Homogenisierung und/oder Teilprobenahmeeffekte - Trocknung - Mahlung - Auflösung - Extraktion - Kontamination - Derivatisierung (chemische Effekte) - Verdünnungsfehler - (Vor-)Anreicherung - Speziationseffekte

Tabelle 2: Quellen der Probenahmeunsicherheit in der Probenahmethorie*

Quelle	Beschreibung
Fundamental sampling error (FSE)***	Ein Ergebnis der Heterogenität in der Zusammensetzung (die Teilchen sind chemisch oder physikalisch unterschiedlich)
Grouping and segregation error (GSE)	Ein Ergebnis der verteilungsbedingten Heterogenität
Long-range point selection error (PSE ₁)	Raum- oder zeitübergreifende Trends
Periodic point selection error (PSE ₂)	Periodische Änderung der Konzentration im Raum oder über die Zeit
Increment delimitation error (IDE)	Identifizierung der richtigen Probe für die Entnahme. Berücksichtigt die Volumengrenzen eines richtigen Probenahmegerätes
Increment extraction error (IXE)	Entnahme der vorgesehenen Probe. Berücksichtigt die Form der Schneidkanten des Probenahmegeräts
Increment and sample preparation error (IPE)	<i>Kontamination</i> (Fremdmaterial in der Probe): <i>Verluste</i> (Adsorption, Kondensation, Ausfällung usw.): <i>Veränderung der chemischen Zusammensetzung</i> (Konservierung): <i>Veränderung der physikalischen Zusammensetzung</i> (Agglomeration, Zerbrennen von Teilchen, Feuchtigkeit usw.): ** <i>Unbeabsichtigte Fehler</i> (Vertauschte Probennummern, Unkenntnis, Nachlässigkeit): ** <i>Absichtliche Fehler</i> (Vortäuschung falscher Tatsachen, absichtliche Fehler bei der Auswahl der Einzelproben, Fälschung usw.)
Weighting error (SWE) - Gewichtungsfehler	<i>Das Ergebnis von Fehlern bei der Zuordnung von Gewichtungen zu verschiedenen Teilen einer ungleich zusammengesetzten Probe</i>

* Diese Einteilung folgt der von Gy [13] und anderen (wird in Abschnitt 10 näher erläutert).

** Von Unsicherheitsschätzungen als grobe Fehler ausgeschlossen [2].

*** Anm. d. Übers.: Hier und an anderen Stellen wurde der englische Fachbegriff beibehalten, da es im Deutschen keinen allgemein üblichen Begriff dafür gibt.

Tabelle 3: Unsicherheitsbeiträge im empirischen Ansatz

Prozess	Einordnung *	
	<i>zufällig (Präzision)</i>	<i>systematisch (Bias)</i>
Analyse	Analytische Variabilität (kombinierter Beitrag zufälliger Effekte)	Analytischer Bias (kombinierter Effekt von Biasquellen)
Probenahme	Probenahmevariabilität (dominiert durch Heterogenität und Probenehmer-Einflüsse)	Probenahmebias (kombinierter Effekt von Auswahlbias, Probenehmer-Bias etc.)

* Die Unterscheidung von zufälligen und systematischen Auswirkungen kann vom Kontext abhängen. Eine systematische Auswirkung bei Messungen einer Organisation (z. B. analytischer Bias) kann auch als zufälliger Effekt betrachtet werden, wenn er im Kontext des Konsenswerts aus einer organisationsübergreifenden Eignungsprüfung betrachtet wird.

TEIL 3 – Schätzung der Messunsicherheit unter Einbezug der Probenahme

6 Ansätze zur Unsicherheitsschätzung

6.1 Es gibt zwei umfassende Ansätze für die Schätzung der Unsicherheit. Einer von ihnen, der als "empirisch", "experimentell", "retrospektiv" oder "top-down" beschrieben wird, verwendet ein gewisses Maß an Wiederholung des gesamten Messvorgangs, um eine direkte Schätzung der Unsicherheit für das Endergebnis der Messung zu erhalten. Dieser Ansatz wird in diesem Leitfaden als "empirischer" Ansatz bezeichnet. Der zweite Ansatz, der verschiedentlich als "Modellierung", "theoretisch", "prädiktiv" oder "bottom-up" bezeichnet wird, zielt darauf ab, alle Unsicherheitsquellen einzeln zu quantifizieren, und verwendet dann ein Modell, um sie zu kombinieren. Er wird daher als "modellierender" Ansatz bezeichnet. Diese Ansätze schließen sich nicht gegenseitig aus. Die empirische Methode kann angepasst werden, um die Beiträge zur Unsicherheit von einem oder mehreren Effekten oder Klassen von Effekten zu schätzen. Beide Ansätze können, falls erforderlich, sinnvollerweise zusammen verwendet werden, um dasselbe Messsystem zu untersuchen. Die Anwendbarkeit der beiden Ansätze schwankt zwischen den verschiedenen zu beprobenden Materialien.

6.2 Der in diesem Leitfaden verfolgte Ansatz besteht darin, den empirischen Ansatz detailliert zu beschreiben, der die breiteste Anwendbarkeit für eine Vielzahl von Messsystemen und Anwendungen (z.B. gasförmig, flüssig und fest) hat. Modellierende Ansätze werden für besondere Situationen, für die sie anwendbar sind (z. B. partikelförmige Feststoffe), beschrieben. Es werden auch Hinweise gegeben, wie eine Kombination dieser verschiedenen Ansätze verwendet werden kann, um zuverlässigere und kosteneffektivere Schätzungen der Unsicherheit in einer Reihe von Messsystemen zu erhalten. Dieser duale Ansatz soll es dem Benutzer des Leitfadens ermöglichen, die für seine spezielle Situation am besten geeignete Methode der Unsicherheitsschätzung auszuwählen. (Abschnitt 1 bietet eine Anleitung zur Auswahl des Ansatzes.)

6.3 Referenz [5] stellt fest, dass die Modellierungsansätze und die empirische Studie, die einen Ringversuch verwendet, Extreme eines Kontinuums sind:

Beachten Sie jedoch, dass die beobachtete Wiederholbarkeit oder eine andere Präzisionschätzung sehr oft als ein separater Beitrag zur Unsicherheit angesehen wird, sogar im GUM-Ansatz. In ähnlicher Weise werden einzelne Effekte in der Regel zumindest auf Signifikanz geprüft oder quantifiziert, bevor die Reproduzierbarkeit bewertet wird. Die praxisnahe Unsicherheitsbewertung verwendet daher oft einige Elemente aus beiden Extremen.

Wenn man sich auf eines der beiden Extreme bezieht, ist es daher wichtig, sich bewusst zu sein, dass dies Extreme sind und dass viele praxisnahe Schätzungen Elemente beider Ansätze beinhalten. Das übergeordnete Ziel eines jeden Ansatzes ist es, eine ausreichend zuverlässige Schätzung der gesamten Messunsicherheit zu erhalten. Dies erfordert nicht unbedingt, dass alle einzelnen Quellen der Messunsicherheit quantifiziert werden müssen, sondern nur der kombinierte Effekt bewertet werden muss. Wenn jedoch das Gesamtniveau der Unsicherheit als inakzeptabel angesehen wird (d. h., die Messungen sind nicht zwecktauglich), dann

müssen Maßnahmen ergriffen werden, um die Unsicherheit zu reduzieren. Im anderen Fall kann die Unsicherheit unnötig klein sein, wodurch es gerechtfertigt sein kann, die analytische Unsicherheit zu erhöhen und dadurch die Kosten der Analyse zu senken. Methoden zur Modifizierung der Unsicherheit werden in Anhang E diskutiert. In diesem Stadium ist es jedoch wichtig, Informationen darüber zu haben, welcher allgemeine Teil des Messverfahrens den dominanten Beitrag zur Gesamtunsicherheit verursacht, und es kann dann notwendig sein, die einzelnen Effekte zu bewerten. Der Vorteil einer detaillierten früheren Untersuchung ist, dass diese Informationen bereits verfügbar sind; der Nachteil ist, dass sie kostspielig zu beschaffen sind und sich als unnötig erweisen können, wenn die Unsicherheit akzeptabel ist. Die Planer sollten daher den erforderlichen Detaillierungsgrad einer Unsicherheitsschätzung abwägen, unter Berücksichtigung wahrscheinlich benötigter detaillierter Informationen für weitere Entwicklungen.

7 Die Messgröße

7.1 In der folgenden Diskussion wird davon ausgegangen, dass die Messgröße ein Durchschnittswert ist, der die Zusammensetzung des gesamten Probenahmezielobjekts repräsentiert, und dass die Messgröße durch einen Prozess der Probenahme und Analyse geschätzt werden soll. Dies bezieht sich auf die Spezifizierung der Messgröße (Abschnitt 5.2) und die Definition des Probenahmezielobjekts (Abschnitt 4.1).

8 Allgemeine Studiendurchführung zur Probenahmeunsicherheit

8.1 Die analytische Arbeit sollte im Rahmen eines geeigneten Qualitätssystems durchgeführt werden, einschließlich validierter Analysemethoden, Eignungstests, interner Qualitätskontrolle und externer Bewertung, wo dies zweckmäßig ist. Validierungsverfahren sollten alle Schritte umfassen, die normalerweise im Labor durchgeführt werden (einschließlich der Entnahme von Teilproben) und sollten Kontrollen des Bias unter Verwendung zertifizierter Referenzmaterialien oder anderer Methoden zur Schätzung des analytischen Bias einschließen [14]. Beachten Sie, dass die in diesem Leitfaden beschriebenen Methoden zur Unsicherheitsschätzung auch für die Schätzung der Unsicherheiten im Zusammenhang mit der Entnahme von Teilproben angewendet werden können.

8.2 Laboratorien, die die chemische Analyse durchführen, sollten die Konzentrationsschätzungen genau so berichten, wie sie gefunden wurden; insbesondere dürfen die Werte nicht zensiert, gekürzt oder in Form von "kleiner als" eines Schwellenwerts berichtet werden, egal ob unterhalb der Nachweisgrenze (LOD) oder kleiner Null. Werden negative oder Sub-LOD-Beobachtungen nicht gemeldet, führt dies zu einer Unterschätzung der Unsicherheit.

9 Empirischer Ansatz

9.1 Überblick

9.1.1 Mit dem empirischen Ansatz („Top-down“) soll eine zuverlässige Schätzung der Unsicherheit erzielt werden, ohne dass eine der Quellen einzeln bekannt sein muss. Es stützt sich auf allgemeine Präzisionsschätzungen aus internen oder organisationsübergreifenden Messversuchen. Es ist möglich, die allgemeine Art der Quelle, wie zufällige oder systematische Effekte, zu beschreiben und diese in solche zu unterteilen, die sich aus dem Probenahmeprozess oder dem Analysenprozess ergeben. Schätzungen der Größe jedes dieser Effekte können unabhängig von den Eigenschaften der Messverfahren vorgenommen werden, z. B. die Probenahmepräzision (für zufällige Effekte aus der Probenahme) oder der analytische Bias (für systematische Effekte aus der chemischen Analyse). Diese Schätzungen können kombiniert werden, um eine Schätzung der Unsicherheit des Messergebnisses zu erhalten. Dieser Ansatz wird detailliert in den Beispielen A1, A2, A3 und A4 veranschaulicht.

9.2 Unsicherheitsquellen

9.2.1 Es ist möglich, die Messunsicherheit zu berücksichtigen, die sich aus vier allgemeinen Fehlerquellen ergibt. Diese vier Quellen sind die zufälligen Fehler, die sich aus den Verfahren der Probenahme und der Analyse ergeben, sowie die systematischen Fehler, die sich aus diesen Verfahren ergeben. Diese Fehler wurden traditionell als Probenahmepräzision, analytische Präzision, Probenahmebias bzw. analytischer Bias quantifiziert (Tabelle 4). Wenn Fehler aus diesen vier Quellen einzeln oder kombiniert quantifiziert werden, ist es möglich, die Unsicherheit der von diesen Verfahren erzeugten Messungen abzuschätzen. Methoden zur Schätzung von drei der vier Fehler sind gut etabliert. Die Präzision der Probenahme und die analytische Wiederholpräzision können durch Duplizieren eines Anteils (z. B. 10 %) der Proben bzw. Analysen geschätzt werden. Andere Beiträge zur analytischen Unsicherheit, wie z. B. Abweichungen über verschiedenen Tagen und systematische Abweichungen, können hinzugefügt werden. Diese zusätzlichen Beiträge müssen separat geschätzt werden. Beispielsweise sollte die Abweichung über verschiedene Tage während der Validierung des Analysenverfahrens ermittelt werden. Der analytische Bias kann geschätzt werden, indem der Bias bei gut passenden zertifizierten Referenzmaterialien gemessen wird und angenommen wird, dass dieser Bias auch den für das Testmaterial gültigen darstellt, oder indem er direkt aus der Validierung des analytischen Verfahrens abgeleitet wird.

9.2.2 Die Verfahren zur Schätzung des Probenahmebias beinhalten die Verwendung eines Referenzzielobjekt für die Probenahme [15, 16] (das Probenahme-Äquivalent eines Referenzmaterials). Alternativ dazu verwenden sie Messungen aus organisationsübergreifenden Probenahmeversuchen, bei denen der von jedem Teilnehmer potenziell eingebrachte unvermutete Bias der Probenahme in der Schätzung der Unsicherheit auf der Grundlage der Gesamtvariabilität enthalten ist [17] (Abschnitt 9.5). Obwohl einige der Komponenten der Unsicherheit die mit systematischen Effekten verbunden sind schwer abzuschätzen sind, ist es vielleicht auch nicht nötig, dies zu tun, wenn es gute Hinweise dafür gibt, dass die systematischen Effekte klein und unter Kontrolle sind. Solche Nachweise können qualitativ sein, wie z.B. Vorkenntnisse über die chemische oder physikalische Natur des Probenahmezielobjekts, oder quantitativ, wie z.B. Informationen aus früheren Messungen von gesamten Chargen. (Siehe Beispiele A3 und A4, Anhang A.)

Tabelle 4: Schätzung der Unsicherheitsbeiträge im empirischen Ansatz

Prozess	Effektklasse	
	Zufällig	Systematisch (<i>bias</i>)
Analytik	Doppelbestimmungen ergeben Präzision unter Wiederholbedingungen	z.B. Validierungsdaten oder ZRM
Probenahme	Doppelte Proben	Referenzprobenahmezielobjekt, organisationsübergreifende Probenahmestudien

Vier Klassen von Effekten, die zur Unsicherheit von Messungen beitragen, und Methoden zu deren Schätzung.

9.3 Statistisches Modell für die empirische Schätzung der Unsicherheit

Um experimentelle Methoden zur Schätzung der Unsicherheit mit dem empirischen Ansatz zu entwerfen, ist ein statistisches Modell erforderlich, das die Beziehung zwischen den gemessenen und wahren Werten der Analytkonzentration beschreibt. Dieses Modell mit zufälligen Effekten betrachtet eine einzelne Messung der Analytenkonzentration (x), an einer einzelnen Probe (Misch- oder Einfachprobe), von einem einzelnen bestimmten Probenahmezielobjekt:

$$x = X_{true} + \varepsilon_{sampling} + \varepsilon_{analytical}$$

wobei X_{true} der wahre Wert der Analytkonzentration (wie auch immer ausgedrückt) im Probenahmezielobjekt (d. h. äquivalent zum Wert der Messgröße) ist. Der Gesamtfehler aufgrund der Probenahme ist $\varepsilon_{sampling}$ und der gesamte analytische Fehler ist $\varepsilon_{analytical}$.

Wenn bei einer Untersuchung eines einzelnen Probenahmezielobjekts die Variationsquellen unabhängig sind, ist die Varianz σ_{meas}^2 der Messung gegeben durch,

$$\sigma_{meas}^2 = \sigma_{sampling}^2 + \sigma_{analytical}^2$$

wobei $\sigma_{sampling}^2$ die Varianz zwischen den Proben eines Zielobjekts ist (größtenteils aufgrund der Heterogenität des Analyten), und $\sigma_{analytical}^2$ die Varianz zwischen den Analysen an einer Probe ist.

Wenn statistische Schätzungen der Varianz (s^2) verwendet werden, um diese Parameter zu approximieren, erhalten wir

$$s_{meas}^2 = s_{sampling}^2 + s_{analytical}^2$$

Die Standardunsicherheit (u) kann mit Hilfe von s_{meas} abgeschätzt werden, die somit gegeben ist durch

$$u = s_{meas} = \sqrt{s_{sampling}^2 + s_{analytical}^2} \quad \text{Gleichung 1}$$

Die durch die physikalische Probenvorbereitung verursachte Varianz kann in die Probenahmevarianz einbezogen, oder bei Bedarf als separater Term ausgedrückt werden.

Da $s_{analytical}$, so wie im empirischen Ansatz ermittelt, nur die Wiederholbarkeitskomponente der analytischen Unsicherheit ist, kann s_{meas} unterschätzt werden. Dies wird in Abschnitt 9.4.4 näher erläutert.

Bei einer Erhebung über mehrere Probenahmezielobjekte, die für die Schätzung der Stichprobenunsicherheit empfohlen wird (Abschnitt 9.4.2), muss das Modell erweitert werden, zu

$$x = X_{true} + \varepsilon_{target} + \varepsilon_{sampling} + \varepsilon_{analytical}$$

wobei der zusätzliche Term ε_{target} die Variation der Konzentration zwischen den Probenahmezielobjekten darstellt und die Varianz $\sigma_{between-target}^2$.

Eine geeignete ANOVA erzeugt Schätzungen der Varianzen $\sigma_{between-target}^2$, $\sigma_{sampling}^2$ und $\sigma_{analytical}^2$, und die Unsicherheit wird genau wie zuvor mit Gleichung 1 geschätzt.

Die Gesamtvarianz, gegeben durch

$$\sigma_{meas}^2 = \sigma_{between-target}^2 + \sigma_{sampling}^2 + \sigma_{analytical}^2$$

ist auch ein nützlicher Parameter bei der Beurteilung der Zwecktauglichkeit; dies wird in Abschnitt 16.2 näher erläutert. Für praktische Zwecke werden die Varianzen der Grundgesamtheit durch ihre Schätzungen s^2 ersetzt und ergeben

$$s_{total}^2 = s_{between-target}^2 + s_{sampling}^2 + s_{analytical}^2 \quad \text{Gleichung 2}$$

9.4 Empirische Schätzung der Unsicherheit

9.4.1 Vier Methodentypen sind für die Schätzung der Unsicherheit mit dem empirischen Ansatz anwendbar (Tabelle 5). Eine fünfte variographische Methode wird in Abschnitt 9.6 kurz beschrieben. Die Hauptmethode, die in diesem Leitfaden weiter beschrieben wird, ist die "Duplikatmethode" (Nr. 1). Wenn ein Probenehmer mehrere Probenahmeverfahren in Methode Nr. 2 verwendet, kann ein Bias zwischen den Vorschriften erkannt werden. Wenn mehrere Probenehmer alle dieselbe Vorschrift anwenden (Methode Nr. 3, was einer Vergleichsuntersuchung bei der Probenahme - CTS (collaborative trial in sampling) - oder einem Methodenleistungstest entspricht), dann kann ein Bias zwischen verschiedenen Probenehmern erkannt und in die Unsicherheitsschätzung einbezogen werden. Wenn mehrere Probenehmer verschiedene Vorschriften anwenden, die nach ihrer fachlichen Meinung für das angegebene Ziel am besten geeignet sind (Methode Nr. 4, die einem Eignungstest-Ringversuch für die Probenahme – SPT (sampling proficiency test) - entspricht), dann kann ein Bias der Probenahme, entweder zwischen den Probenahmeverfahren oder zwischen den Probenehmern, erkannt und in die Schätzung der Unsicherheit einbezogen werden. Ein ausgearbeitetes Beispiel für diesen Ansatz wurde beschrieben, allerdings nur mit einer einzigen Runde eines SPT [17].

Tabelle 5: Vier empirische Methoden zur Schätzung der kombinierten Unsicherheit einschließlich Probenahme

Methode Nr.	Methodenbeschreibung	Probennehmer (Personen)	Vorschriften	Geschätzte Komponente			
				Probenahme		Analytik	
				Präzision	Bias	Präzision	Bias
1	Duplikat	Eine(r)	Eine	Ja	Nein	Ja ³	Nein ¹
2	Vorschriften	Eine(r)	Mehrere	Zwischen Vorschriften		Ja ³	Nein ¹
3	CTS	Mehrere	Eine	Zwischen Probenehmern		Ja	Ja ²
4	SPT	Mehrere	Mehrere	Zwischen Vorschriften + Zwischen Probenehmern		Ja	Ja ²

¹Informationen zum analytischen Bias können durch Einbeziehung zertifizierter Referenzmaterialien in den Analyselauf erhalten werden (siehe Beispiel A2, Anhang A).

²Der analytische Bias ist teilweise oder vollständig in Vergleichsuntersuchungen enthalten, an denen mehrere Laboratorien beteiligt sind.

³Bei diesen Ansätzen wird die Präzision unter Wiederholbedingungen geschätzt.

9.4.2 Die Duplikatmethode ist die einfachste und wahrscheinlich kostengünstigste der vier in Tabelle 5 beschriebenen Methoden. Sie basiert darauf, dass ein einzelner Probennehmer einen kleinen Teil (d. h. 10 %, aber nicht weniger als acht Zielobjekte) der Primärproben dupliziert^e [18, 19]. Idealerweise werden die Duplikate von mindestens acht Probenahmezielobjekten entnommen, die nach dem Zufallsprinzip ausgewählt werden, um die typische Zusammensetzung solcher Zielobjekte zu repräsentieren. Wenn nur ein Zielobjekt vorhanden ist, können alle acht Duplikate davon genommen werden, aber die Unsicherheitsschätzung gilt dann nur für dieses eine Zielobjekt. Die Doppelproben werden durch Wiederholung derselben nominellen Probenahmenvorschrift entnommen, mit zulässigen Variationen, die die Mehrdeutigkeit der Probenahmenvorschrift und die Auswirkung der kleinräumigen Heterogenität des interessierenden Analyten auf die Umsetzung dieser Vorschrift widerspiegeln. Zum Beispiel wird bei einem "W"-Design zum Sammeln einer Mischprobe aus einem Salatfeld der anfängliche Startpunkt und die Ausrichtung des "W" für die Doppelprobe geändert; bei einem Gitterdesign werden wiederum Startpunkt und Ausrichtung geändert (Beispiel A1, Anhang A). Die Doppelproben werden mit einer einzigen Probenahmenvorschrift und von einer einzigen Person (Probennehmer) genommen. Die beiden

^eEs kann ein höherer Replikationsgrad verwendet werden, aber die Verdoppelung ist normalerweise die effektivste Form der Replikation in Probenahmestudien. Es ist zum Beispiel besser, Duplikate von 12 Probenahmezielobjekten zu nehmen, als Triplikate von acht Zielobjekten, da obwohl jede Schätzung der Unsicherheit der Probenahme (s_{sampling}) zwar einen geringeren Standardfehler hat, die Schätzung auf einem kleineren Anteil der Gesamtpopulation der Probenahmezielobjekte basiert und daher weniger repräsentativ ist. Die Mindestanzahl von acht Duplikaten ist erforderlich, um ausreichend zuverlässige Schätzungen der Unsicherheit zu liefern [19].

duplizierten Proben werden einer physikalischen Aufbereitung unterzogen, die zu zwei getrennten Prüfproben führt. Aus beiden Prüfproben werden doppelte Prüfmengen entnommen und in zweifacher Ausführung analysiert (d.h. doppelte chemische Analyse). Dieses System der doppelten Probenahme und der chemischen Analyse beider Proben wird als "ausgewogenes Design" bezeichnet (Abbildung 2). Beachten Sie, dass die Duplikatmethode keinen Beitrag eines Bias enthält, der entweder als vernachlässigbar angenommen oder separat geschätzt werden muss, z. B. mit Hilfe von mehreren Probenehmern, mehreren Vorschriften und/oder organisationsübergreifenden Probenahmeversuchen wie bei den anderen drei Methoden.

Anmerkung 1: Obwohl die "Duplikatmethode" im Allgemeinen in Bezug auf einen einzelnen Probenehmer und eine Vorschrift beschrieben wird, kann dasselbe Design mit verschiedenen Probenehmern verwendet werden, um den "Zwischen-Probenehmer"-Beitrag zur Unsicherheit zu berücksichtigen (entspricht Methode Nr. 3).

Anmerkung 2: Bei einem kostengünstigeren "nicht ausgewogenen" Design wird nur eine der beiden Proben doppelt analysiert, z. B. die Probe 1 in Abb. 2. Es hat die gleiche Anzahl von Probenduplikaten wie das ausgewogene Design, hat aber die Hälfte der Analysenduplikate. Dies reduziert die zusätzlichen Kosten für die Anwendung dieser Methode auf dieselbe Anzahl von Zielobjekten um 33 % ([20] Anhang D)

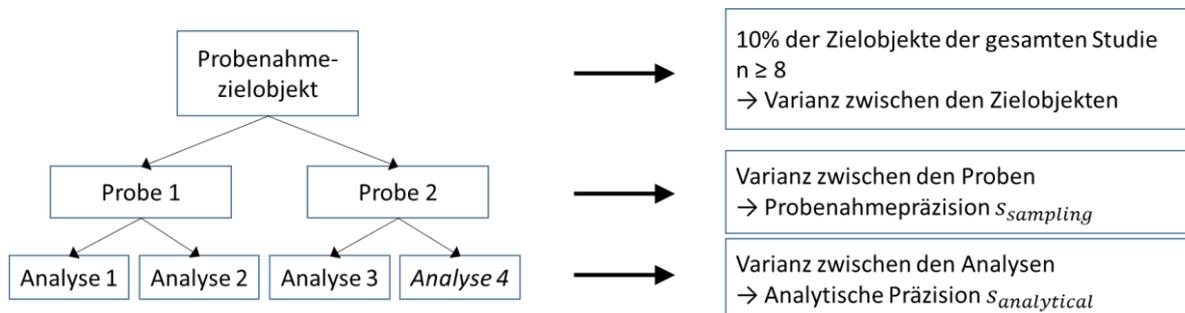


Abb. 2: Ein ausgewogenes Design

Ein ausgewogenes Design zur empirischen Schätzung der Unsicherheit (d. h. zweistufiges geschachteltes Design), unter Verwendung der "Duplikatmethode". Das Entfernen der Analyse 2 auf Probe 2 würde zu dem kostengünstigeren nicht ausgewogenen Design [Abb. D2(b)] führen, der in Anmerkung 2 oben besprochen wird.

9.4.3 Die Prüfmengen werden dann anonym mit einer geeigneten analytischen Methode unter Wiederholbedingungen chemisch analysiert (z. B. zufällig innerhalb einer analytischen Serie verteilt). Wenn Schätzungen des analytischen Anteils der Messunsicherheit vom Labor unabhängig vorgenommen wurden, ist dies für den Vergleich mit den Schätzungen dieser Methode oder ggf. für die separate Einbeziehung in das Unsicherheitsbudget nützlich [21]. Die durch die physikalische Probenvorbereitung verursachte Varianz kann in die Stichprobenvarianz einbezogen werden, indem jedes Probenduplikat unabhängig vorbereitet

wird. Alternativ kann diese Varianz separat geschätzt werden, indem eine zusätzliche Replikationsebene in das Design eingefügt wird (Anhang D).

9.4.4 Das hier vorgeschlagene ausgewogene Design liefert nur die Wiederholstandardabweichung der analytischen Messungen. Um den anderen Teil der analytischen Unsicherheit abzuschätzen, müssen andere Unsicherheitsquellen berücksichtigt werden, einschließlich der Schwankungen von Tag zu Tag und eines potenziellen analytischen Bias. Die Grenzen dieses Ansatzes und ein praktisches Beispiel finden Sie in Abschnitt 6 von Beispiel A2. Eine Alternative ist, das Messlabor nach der Wiederholbarkeit und der Messunsicherheit zu fragen und dann zu prüfen, ob die in dieser Studie ermittelte Wiederholbarkeit mit der vom Labor angegebenen vergleichbar ist. Wenn dies der Fall ist, können wir die vom Labor angegebene Messunsicherheit als $u(\text{analytisch})$ verwenden (normalerweise $U/2$). Wenn die beobachtete Wiederholbarkeit erheblich von der Schätzung der Messunsicherheit des Labors abweicht, ist es in der Regel ratsam, den größeren der beiden Werte zu verwenden. Eine zweite Alternative ist die Verwendung der Schätzung des analytischen Bias, die aus gut angepassten zertifizierten Referenzmaterialien in der analytischen Serie vorgenommen wird. Diese Bias-Schätzung kann dann mit der Wiederholbarkeit kombiniert werden, um die Messunsicherheit zu erhalten [1,22].

9.5 Berechnung der Unsicherheit und ihrer Komponenten

9.5.1 Die Zufallskomponente der Unsicherheit kann durch Anwendung der Varianzanalyse (ANOVA)^f oder der Spannweitenberechnung^g auf die Konzentrationsmessungen an den doppelten Proben geschätzt werden. Die Schätzung basiert auf dem in Abschnitt 9.3 beschriebenen Modell, angewandt auf das jeweils verwendete Messvorschrift (mit der festgelegten Anzahl von Probeninkrementen und Analysewiederholungen). Sowohl ANOVA als auch Spannweitenberechnungen können für Berechnungen in Konzentrationseinheiten verwendet werden. Wenn die Standardabweichung näherungsweise proportional zur Konzentration ist oder die relative Unsicherheit größer als 20 % ist, sollten die Berechnungen nach einer Log-Transformation der Daten durchgeführt werden. Siehe auch Abschnitt 5.7 zur Variation der Standardabweichung über die Konzentration.

9.5.2 Die Werte von s_{sampling} und $s_{\text{analytical}}$ aus der ANOVA sind Schätzungen der Probenahmepräzision bzw. der analytischen Präzision. Die Zufallskomponente der Messunsicherheit wird durch die Kombination dieser beiden Schätzungen berechnet (Gleichung 1). Für die erweiterte Unsicherheit, z. B. für ca. 95 % Konfidenz, muss dieser Wert mit einem Erweiterungsfaktor von 2 multipliziert werden. Die erweiterte Unsicherheit (U) wird dann berechnet mit

$$U = 2s_{\text{meas}} \quad \text{Gleichung 3}$$

^f In den Häufigkeitsverteilungen der analytischen, probeninternen und probenübergreifenden Variabilität gibt es oft einen kleinen Anteil (d. h. <10 %) an Ausreißerwerten. Dies erfordert die Verwendung einer Methode zur Herabgewichtung der Auswirkung der abweichenden Werte auf die klassische ANOVA, wie z. B. die Verwendung robuster statistischer Methoden. Dies ergibt eine zuverlässigere Schätzung der Varianzen der zugrundeliegenden Populationen. Eine ausführlichere Erläuterung dieser Methoden finden Sie im Arbeitsbeispiel in Anhang A1 (S. 47), A2 (S. 53).

^g Siehe Beispiel in Anhang A3.

U kann auch relativ zum berichteten Wert x und in Form eines Prozentsatzes ausgedrückt werden, als relative erweiterte Unsicherheit U' :

$$U' = 100 \frac{2s_{meas}}{x} \% \quad \text{Gleichung 4}$$

Die relative Unsicherheit ist breiter anwendbar als die Standardunsicherheit, da sie sich als Funktion der Konzentration bei Werten weit oberhalb der analytischen Nachweisgrenze (>10-fach) nicht nennenswert ändert. Andere Erweiterungsfaktoren können nach Bedarf gewählt werden. Die Verbesserung dieser Schätzung der Unsicherheit durch Einbeziehen des systematischen Fehlers aus der chemischen Analyse wird in Beispiel A2, Anhang A diskutiert.

Die relative erweiterte Unsicherheit für die Probenahme oder Analyse allein kann in ähnlicher Weise ausgedrückt werden als

$$U'_{sampling} = 100 \frac{2s_{sampling}}{x} \%$$

und

$$U'_{analytical} = 100 \frac{2s_{analytical}}{x} \%$$

Wenn die angegebene Standardunsicherheit des Labors $u_{analytical}$ größer ist als die durch ANOVA ermittelte $s_{analytical}$, ist das Ersetzen von $s_{analytical}$ durch $u_{analytical}$ eine bessere Schätzung der Unsicherheit.

9.5.3 Die klassische ANOVA geht davon aus, dass die Häufigkeitsverteilung des Effekts annähernd normal (d.h. gaußförmig) ist, aber in einigen Fällen hat sich gezeigt, dass sie annähernd log-normal ist. In solchen Fällen können die natürliche Logarithmen der Messwerte verwendet werden, und die ANOVA wird mit diesen log-transformierten Werten durchgeführt. Die Ergebnisse einer solchen ANOVA liegen nicht in den ursprünglichen Einheiten der Analytkonzentration vor, können aber zur Berechnung eines Standardunsicherheitsfaktors (${}^F u$) und des erweiterten Unsicherheitsfaktors (${}^F U$) verwendet werden. Ein Messwert kann dann mit ${}^F U$ multipliziert werden, um die obere Vertrauensgrenze zu berechnen, und durch ${}^F U$ geteilt werden, um die untere Vertrauensgrenze des Unsicherheitsintervalls zu berechnen.

Aus der Standardabweichung der logarithmisch transformierten Werte ($s(\ln(x)) = s_G$) kann der Standardunsicherheitsfaktor (${}^F u$) berechnet werden [23], mit

$${}^F u = \exp(s_G)$$

Der nützlichere erweiterte Unsicherheitsfaktor (${}^F U$), der $k = 2$ verwendet, kann auf zwei Arten berechnet werden, entweder als

$${}^F U = \exp(2s_G) \quad \text{Gleichung 5}$$

Alternativ, weil die Multiplikation mit k im logarithmischen Raum dasselbe ist wie die Potenzierung von k im linearen Raum, kann ${}^F U$ auch einfacher berechnet werden als

$${}^F U = ({}^F u)^2$$

Die obere Vertrauensgrenze für das erweiterte Unsicherheitsintervall wird berechnet, indem die gemessene Konzentration (x) mit $^F U$ multipliziert wird, und die untere Vertrauensgrenze als x dividiert durch $^F U$. Dieses Vertrauensintervall ist also nicht symmetrisch um den Messwert im normalen linearen Konzentrationsraum.

Die relative Unsicherheit u' , ausgedrückt als Bruch, kann berechnet werden aus

$$u' = \sqrt{\exp(s_G^2) - 1}$$

was für kleine Werte von u' , (z.B. $< 0,2$) ungefähr gleich s_G ist.

Ein ausgearbeitetes Beispiel für die Berechnung von $^F U$ finden Sie in Beispiel A2 in Anhang A.

9.5.4 Da die Unsicherheit vieler Messsysteme von der Heterogenität innerhalb des Probenahmezielobjekts dominiert wird, ergibt die Verwendung der einfachsten "Duplikatmethode" oft eine einigermaßen zuverlässige Schätzung der Unsicherheit. Untersuchungen von Umweltsystemen haben gezeigt, dass die Effekte zwischen den Probenehmern und zwischen den Vorschriften oft viel kleiner sind als die durch Heterogenität verursachten Effekte [24]. Weitere Informationen zur Auswahl der effektivsten Methode zur Unsicherheitsschätzung finden Sie in Abschnitt 1. Beispiele für Anwendungen der Duplikatmethode finden Sie in den Beispielen A1 und A2, Anhang A.

9.5.5 Zusätzlich zu einer ersten einmaligen Schätzung der Unsicherheit für eine bestimmte Probenahmenvorschrift, die auf ein bestimmtes Probenahmezielobjekt angewendet wird, ist die routinemäßige Anwendung der "Duplikatmethode" auch als Möglichkeit zur Überwachung der laufenden Probenahmequalität nützlich (Abschnitt 13). Dadurch kann die Auswirkung von Änderungen der Heterogenität des Probenahmezielobjekts zwischen verschiedenen Anwendungen derselben Probenahmenvorschrift auf die Unsicherheit berücksichtigt werden. Auf diese Weise kann ein quantitativer Nachweis über die Qualität der Probenahme gewonnen werden, anstatt sich nur auf die Annahme zu verlassen, dass die Proben repräsentativ sind, wenn sie nach einer korrekten Vorschrift entnommen wurden.

9.6 Alternative empirische Methoden zur Unsicherheitsschätzung

9.6.1 Die Variographie wurde als weiteres empirisches Mittel zur Schätzung der Messunsicherheit aus den kombinierten Quellen der Probenahme und der Analyse vorgeschlagen [25] und demonstriert [26]. Sie ist besonders nützlich in Situationen, in denen es große räumliche und/oder zeitliche Variationen der Analytkonzentration gibt, die quantifiziert und modelliert werden können. Dies ist in einigen Fällen der Gesteins- und Bodengeochemie sowie bei der Emissionskontrolle (z. B. Abwasser) der Fall, wenn eine große Anzahl ($n > 100$) gleichmäßig verteilter Proben genommen wurde.

Eine weitere Anleitung zu den Prinzipien und der Anwendung der Variographie für diesen Zweck, mit einer Fallstudie, ist verfügbar [8].

10 Der Modellierungs-Ansatz

10.1 Ursache-Wirkungs-Modellierung

10.1.1 Der Modellierungsansatz, umgangssprachlich oft als "bottom-up" bezeichnet, wurde für Messverfahren im Allgemeinen beschrieben [2] und auf analytische Messungen angewendet [1]. Er identifiziert zunächst alle Quellen der Unsicherheit, quantifiziert die Beiträge aus jeder Quelle und kombiniert dann alle Beiträge als Budget, um eine Schätzung der kombinierten Standardunsicherheit zu erhalten. Dabei wird das Messverfahren in alle seine Einzelschritte zerlegt. Dies kann zweckmäßigerweise in Form eines Ursache-Wirkungs-Diagramms oder eines "Fischgräten-Diagramms" geschehen [1]. Die Messunsicherheit, die durch jeden dieser Schritte entsteht, wird unabhängig voneinander geschätzt, entweder empirisch oder durch andere Methoden. Die kombinierte Unsicherheit wird dann berechnet, indem die Unsicherheiten aus allen Schritten mit geeigneten Methoden kombiniert werden. Dieser Ansatz ist für analytische Methoden gut etabliert [1], wurde aber erst in jüngerer Zeit auf den Prozess der Probenahme angewendet [25, 27]. Für Partikelsysteme verwendet die Stichprobentheorie einen ähnlichen Ansatz, um sieben Arten von Probenahmefehlern zu identifizieren. Einer dieser Fehler (Fundamentalfehler) wird mithilfe einer Gleichung geschätzt, die auf einer detaillierten Kenntnis der einzelnen Partikel basiert, die als Probe genommen werden, wie im nächsten Abschnitt (und Beispiel A5, Anhang A) beschrieben.

10.2 Probenahmetheorie für die Schätzung der Unsicherheit

10.2.1 Die Probenahmetheorie wurde als geeignete Methode zur Schätzung der Unsicherheit von Stichproben vorgeschlagen [28]. Dieser Ansatz stützt sich auf die Verwendung eines theoretischen Modells, wie z. B. das von Gy. Pierre Gy hat eine vollständige Probenahmetheorie entwickelt, die in vielen Veröffentlichungen beschrieben ist [6, 13, 29, 30, 31, 32]. Abbildung 3 zeigt die Gy'sche Klassifizierung der Probenahmefehler. Die meisten Probenahmefehler, außer den Probenvorbereitungsfehlern, sind auf die Materialheterogenität zurückzuführen, die in zwei Klassen unterteilt werden kann: 1) Heterogenität in der Zusammensetzung (constitution heterogeneity - CH), und 2) Verteilungsheterogenität (distribution heterogeneity - DH). Beide Heterogenitäten können mathematisch definiert und experimentell abgeschätzt werden. Die Heterogenität in der Zusammensetzung bezieht sich auf die Tatsache, dass alle natürlichen Materialien heterogen sind, d. h. sie bestehen aus verschiedenen Arten von Teilchen (Moleküle, Ionen, Körner). In dieser Terminologie wird eine Verteilung als "heterogen" bezeichnet, wenn die Partikel im zu untersuchenden Probenahmezielobjekt (oder Los) nicht zufällig verteilt sind.

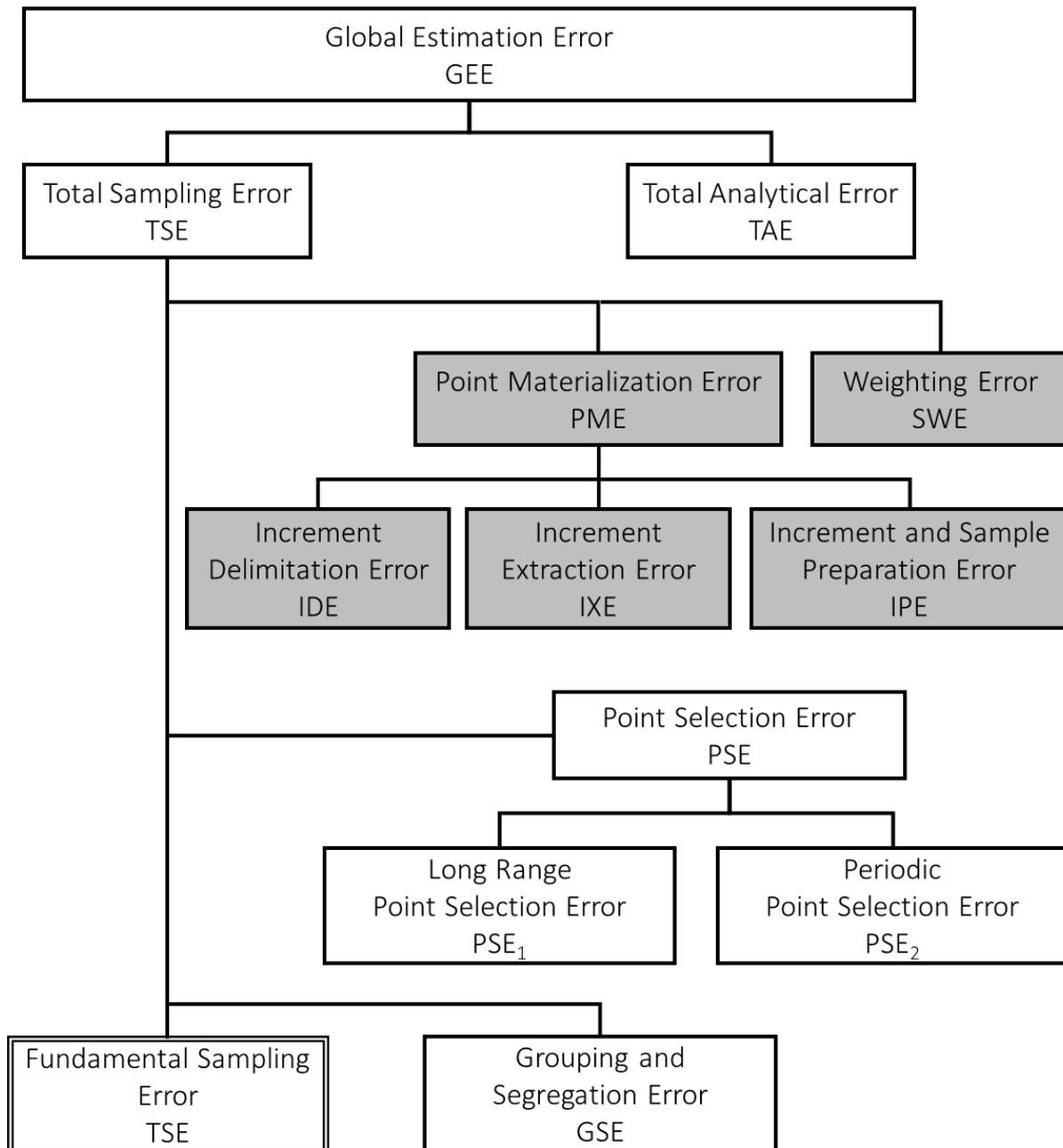


Abb. 3: Klassifizierung von Probenahmefehlern in der Stichprobentheorie von Gy

* Fehler aus falscher Probenahme sind durch schattierte Kästchen gekennzeichnet und werden bei den Schätzungen der Unsicherheit nicht berücksichtigt.

10.2.2 Die Klassifizierung von Probenahmefehlern bildet einen logischen und nützlichen Rahmen für die Gestaltung und Prüfung von Probenahmeverfahren. Diejenigen, die für die Schätzung der Unsicherheit von zentraler Bedeutung sind (z. B. FSE in Abbildung 3), werden im Folgenden diskutiert, andere (SWE, PSE und GSE) in Anhang C.

10.2.3 Der Gesamtbestimmungsfehler, den Gy den *global estimation error* (GEE) nennt, ist die Summe aus dem *total sampling error* (gesamter Probenahmefehler) (TSE) und dem *total analytical error* (gesamter analytischer Fehler) (TAE). Die Komponenten des TSE können in zwei Hauptgruppen unterteilt werden: 1) Fehler aus falscher Probenahme, 2) Fehler aus korrekter Probenahme. Einige Fehler aus falschen Probenahme entstehen durch das, was der GUM [2] als grobe Fehler bezeichnet und als solche würden sie von den Unsicherheitsschätzungen ausgeschlossen werden. Die Fehler aus korrekter Probenahme treten im Rahmen der guten Praxis auf und können bei der Schätzung der Unsicherheit nach dem GUM-Ansatz [2] berücksichtigt werden.

10.2.4 Fehler aus falscher Probenahme entstehen durch Probenahmegeräte und -verfahren, die nicht den in der Probenahmetheorie definierten Regeln der korrekten Probenahme folgen. In Abbildung 3 sind diese Fehler in schattierten Kästchen dargestellt. Der *increment delimitation error* (IDE) ist ein Fehler, der entsteht, wenn die Form der Probe nicht korrekt ist. Zum Beispiel ist die richtige Probe aus einem Prozessstrom eine komplette Scheibe gleicher Dicke, die durch den Prozessstrom geschnitten wird. Das Probenahmegerät sollte so ausgelegt sein, dass sie das vorgesehene Probenprofil extrahieren kann (d.h. alle Bestandteile haben eine gleiche Chance, in der Probe zu landen). Andernfalls wird ein *sample or increment extraction error* (IXE) entstehen. *Sample preparation errors* (Probenvorbereitungsfehler) (IPE) haben mehrere mögliche Ursachen, die in Tabelle 2 aufgelistet sind, von denen zwei aus der Berechnung ausgeschlossen werden, da sie nach der GUM-Definition [2] als grobe Fehler gelten.

10.2.5 Fehler aus falscher Probenahme haben folgende Eigenschaften gemeinsam: 1) sie erzeugen einen Probenahmebias und erhöhen die Gesamtvarianz auf unvorhersehbare Weise, 2) sie sind von den Umständen abhängig und daher ist jeder Versuch, sie experimentell abzuschätzen, normalerweise nicht sinnvoll, weil er teuer ist und die Ergebnisse nicht verallgemeinert werden können. Der richtige Weg ist, diese Fehler zu minimieren oder zu eliminieren, indem man die Geräte und Verfahren sorgfältig überprüft, strukturell falsche Geräte und Verfahren durch solche ersetzt, die den Regeln der korrekten Probenahme folgen, und indem man das Probenahmepersonal ausreichend schult. Nur wenn dieser technische Teil korrekt ausgeführt wird, hat der theoretische Teil der Unsicherheitsbewertung einen Vorhersage-Wert.

10.2.6 Fehler aus korrekter Probenahme sind im unteren Teil von Abbildung 3 dargestellt. Wenn die Fehler aus falscher Probenahme eliminiert werden, können diese Fehler modelliert und zur Schätzung der Unsicherheit der Probenahme verwendet werden. Der *fundamental sampling error* gehört zu den wichtigsten und wird hier weiter betrachtet; andere werden in Anhang C2 diskutiert.

10.2.7 Der **fundamental sampling error (FSE)** ist der minimale Fehler eines idealen Probenahme-Verfahrens. Letztlich hängt er von der Anzahl der kritischen Partikel in den Proben ab (ein "kritischer Partikel" ist eines, das den Analyten enthält). Bei homogenen Gasen und Flüssigkeiten ist er sehr klein, aber bei Feststoffen, Pulvern und partikelförmigen Materialien, insbesondere bei niedrigen Konzentrationen kritischer Partikel, kann der *fundamental sampling error* sehr groß sein. Wenn die zu beprobende Charge als eindimensionales Objekt behandelt werden kann, können Modelle für den *fundamental sampling error* verwendet werden, um die Unsicherheit der Probenahme abzuschätzen. Wenn die Charge nicht als eindimensionales Objekt behandelt werden kann, muss zumindest der *point selection error* berücksichtigt werden, wenn die Varianz der Primärproben geschätzt wird. Wenn die Probenvorbereitung und die Reduzierung der Probenmenge durch Aufteilung korrekt durchgeführt werden, können Modelle für den *fundamental sampling error* zur Schätzung der Varianzkomponenten verwendet werden, die durch diese Schritte erzeugt werden. Wenn der Erwartungswert für die Anzahl der kritischen Partikel in der Probe leicht als Funktion der Probenmenge geschätzt werden kann, kann die Poisson-Verteilung oder die Binomial Verteilung als Probenahmemodelle verwendet werden, um die Unsicherheit der Probenahme zu schätzen. In den meisten Fällen kann das Modell für den *fundamental sampling error* verwendet werden.

10.2.8 Wenn das zu beprobende Material aus Partikeln mit unterschiedlichen Formen und Größenverteilungen besteht, ist es schwierig, die Anzahl der kritischen Partikel in der Probe abzuschätzen. Mit einer Gleichung kann die relative Varianz des *fundamental sampling error* geschätzt werden:

$$\sigma_r^2 = C d^3 \left(\frac{1}{M_s} - \frac{1}{M_L} \right) \quad \text{Gleichung 6}$$

mit

$$\sigma_r = \frac{\sigma_a}{a_L} = \text{relative Standardabweichung des } \textit{fundamental sampling error}$$

σ_a = absolute Standardabweichung (in Konzentrationseinheiten)

a_L = mittlere Konzentration der Charge

d = charakteristische Partikelgröße = obere 95 % Grenze der Größenverteilung

M_s = Probenmenge

M_L = Chargenmenge

C ist eine Probenahmekonstante, die von den Eigenschaften des beprobten Materials abhängt; C ist das Produkt aus vier Parametern:

$$C = f g \beta c \quad \text{Gleichung 7}$$

f = Formfaktor (siehe Abb. 4)

g = Größenverteilungsfaktor ($g = 0,25$ für breite Größenverteilung und $g = 1$ für einheitliche Partikelgrößen)

β = Freisetzungsfaktor (siehe Abb. 4). Für Materialien, in denen die Partikel vollständig freigesetzt sind, ist $\beta = 1$. Für nicht freigesetzte Materialien wird ein empirischer Faktor, $\beta = \left(\frac{L}{d}\right)^x$, verwendet, wobei für x Werte im Bereich von 0,5 bis 1,5 vorgeschlagen wurden.

c = Strukturfaktor, der wie folgt abgeschätzt werden kann, wenn die notwendigen Materialeigenschaften verfügbar sind:

$$c = \frac{\left(1 - \frac{a_L}{\alpha}\right)^2}{\frac{a_L}{\alpha}} \rho_c + \left(1 - \frac{a_L}{\alpha}\right) \rho_m \quad \text{Gleichung 8}$$

Dabei ist a_L die mittlere Konzentration in der Charge, α die Konzentration des Analyten in den kritischen Partikeln, ρ_c die Dichte der kritischen Partikel und ρ_m die Dichte der Matrix oder der verdünnenden Partikel. Eine vollständigere Beschreibung dieser Terme findet sich in [29].

10.2.9 Stehen die Materialeigenschaften nicht zur Verfügung und sind sie schwer abzuschätzen, kann die Probenahmekonstante C immer experimentell abgeschätzt werden. Zertifizierte Referenzmaterialien sind z. B. eine spezielle Gruppe von Materialien, für die die Probenahmekonstante aus vorhandenen Daten geschätzt werden kann.

10.2.10 Ein Beispiel dafür, wie das Modell für den *fundamental sampling error* in der Praxis verwendet werden kann, finden Sie in Beispiel A5, Anhang A.

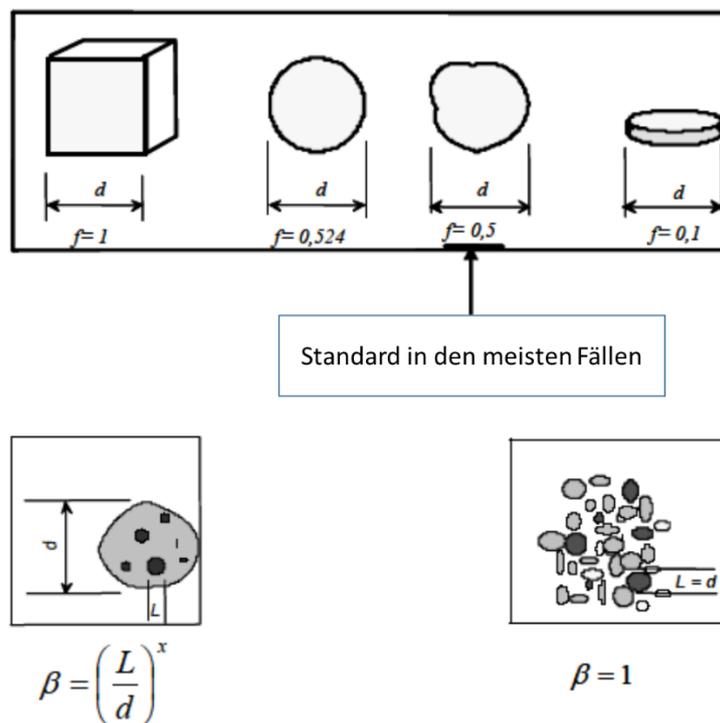


Abb. 4: Schätzung des Faktors für die Schätzung des *fundamental sampling error*

Der Partikelformfaktor f (oberer Teil) und der Freisetzungsfaktor β für nicht freigesetztes Material (unten links) und freigesetztes Material (unten rechts). L ist die Freisetzunggröße der kritischen Partikel.

Teil 4- Managementfragen

11 Verantwortung für die Qualität der Probenahme

11.1 Die Auswirkungen der Betrachtung von Probenahme als integraler Bestandteil des Messprozesses sind weitreichend und umfassen auch Managementfragen. Die Strenge, die für die Bewertung und Verbesserung der Qualität von Aktivitäten innerhalb des analytischen Labors angewandt wird, sollte auch auf die Probenahmeverfahren angewandt werden. Die Verantwortung für die Qualität des gesamten Messprozesses sollte letztlich bei einer Organisation liegen, und die Verantwortlichkeiten für verschiedene Teile des Prozesses müssen zusätzlich definiert werden. In ähnlicher Weise sollte eine Stelle die Verantwortung für die Schätzung der Messunsicherheit übernehmen, auf der Grundlage der Informationen von allen Teilnehmern. Diese Organisation kann dann alle Teilnehmer über die Beiträge informieren, die die sich aus den wichtigsten Schritten des Messverfahrens ergeben.

12 Auswahl des Ansatzes zur Unsicherheitsschätzung

12.1 Der empirische Ansatz (Top-down) und der Modellierungsansatz (Bottom-up) haben unter bestimmten Umständen jeweils ihre Vorteile. Diese sollten bei der Auswahl des Ansatzes für ein bestimmtes Probenahmeverfahren berücksichtigt werden.

- Der empirische Ansatz umfasst alle Quellen der Unsicherheit, ohne dass der Wissenschaftler ihre Identität im Voraus kennen muss. Er ist zum Beispiel automatisch auf die jeweiligen Verunreinigungen und die Mineralogie eines geologischen Standorts anwendbar. Die Berechnungen erfordern keine Vorkenntnisse über die Beschaffenheit des Materials (z. B. Korngröße, Analytenspeziation, Grad der Heterogenität). Er ist relativ schnell und einfach in der Praxis anwendbar (vor allem die "Duplikatmethode"). Es sind mindestens vier Optionen verfügbar, die schrittweise genauere (und teurere) Schätzungen der Unsicherheit ermöglichen, je nach Bedarf (Tabelle 5). Einige dieser Methoden können systematische Fehler (z. B. Probenahme-Bias) innerhalb der Schätzung der Unsicherheit berücksichtigen. Probenahme-Eignungstests und Referenz-Probenahmezielobjekte befinden sich noch im Anfangsstadium der Entwicklung, sind aber bereits sehr vielversprechend für diese Anwendung [17].
- Zu den Nachteilen des empirischen Ansatzes gehört, dass er nicht unbedingt die einzelnen Unsicherheitskomponenten quantifiziert (obwohl dieses Wissen mit begrenzter Auflösung hinzugefügt werden kann). Er basiert nicht auf einem theoretischen Modell der Partikelentnahme, aber dies kann ein Vorteil bei der Anwendung auf Materialien sein, die nicht in Partikelform vorliegen (z. B. gasförmige, flüssige Stoffe, Biota). Der empirische Ansatz liefert nur einen Näherungswert für die Unsicherheit, die über das Zielobjekt hinweg als konstant angenommen wird, aber das gilt auch für den Modellierungsansatz. Extremwerte bei den Wiederholungsmessungen können zu einer Überschätzung des Unsicherheitswerts führen, der für die meisten Messungen nicht repräsentativ ist. Dieser Effekt kann jedoch durch die Verwendung robuster Statistik minimiert werden ([33] und Abschnitt 6 der Beispiele A1 und A2).

- Der Hauptvorteil des Modellierungsansatzes besteht darin, dass er es ermöglicht, die größte Quelle der Ungewissheit leicht zu identifizieren, wenn sie im Modell enthalten ist. Er bietet eine transparente Methode, die zeigt, welche Unsicherheitskomponenten bei der Summierung der Unsicherheit berücksichtigt wurden. Schließlich können Modellierungsansätze, wenn vorherige Informationen verfügbar sind, weniger kostspielig sein als umfangreiche experimentelle Studien.
- Zu den Nachteilen des Modellierungsansatzes gehört, dass die theoretischen Vorhersagen der Unsicherheit detaillierte vorherige Messungen der Mineralogie, der Korngröße und der Analytspezies des zu beprobenden Materials (z. B. Boden) und der Art und Weise, wie diese im Zielobjekt variieren, erfordern können. Es müssen daher idealisierte Annahmen über die Zusammensetzung des Materials getroffen werden (z. B. Mineralogie, Korngröße und Analytspezies). Der Modellierungsansatz unter Verwendung der Probenahmetheorie erfordert Schätzungen oder Annahmen über acht Arten von Probenahmefehlern, und auch darüber, wie diese über das Zielobjekt variieren könnten. Sowohl theoretische als auch empirische Ansätze können relativ zeitaufwendig und daher teuer sein. Generische Schätzungen sind möglicherweise zu allgemein und spiegeln nicht die spezifischen Umstände eines bestimmten Probenahmezielobjekts wider. Außerdem werden möglicherweise nicht alle Quellen der Unsicherheit identifiziert werden, was zu einer Unterschätzung der Gesamtunsicherheit führt.

Alles in allem sind die empirischen Methoden daher tendenziell allgemeiner auf einen breiteren Bereich von Materialtypen anwendbar und hängen nicht so stark von den Vorkenntnissen über das System oder über alle Quellen der Unsicherheit ab. Dadurch sind sie weniger zeit- und damit kostenaufwendig in der Anwendung, was vor allem bei der einmaligen Prüfung von verschiedenen Probenahmezielobjekten von Vorteil ist. Die Modellierungsansätze hingegen führen zu einer detaillierteren Bewertung der einzelnen bekannten Unsicherheitsquellen und sind besser geeignet für die Entwicklung eines langfristigen Probenahmeplans für eine spezifische, gut charakterisierte Anwendung.

13 Qualitätskontrolle der Probenahme

13.1 Beziehung zwischen Validierung und Qualitätskontrolle

13.1.1 Wenn eine Unsicherheit ermittelt wurde, die die Messungen für ihren Zweck geeignet macht, kann eine Bewertung der vorgeschlagenen Probenahme- und Analyseverfahren vorgenommen werden, um diesen Zweck zu erfüllen. Zu diesem Zweck werden zwei Bewertungsinstrumente benötigt: Validierung und kontinuierliche Qualitätskontrolle.

13.1.2 Die Validierung umfasst eine einmalige Schätzung der Unsicherheitskomponenten, die unter Bedingungen ermittelt werden, die bei der routinemäßigen Anwendung der Verfahren zu erwarten sind. Die Validierung kann allgemein für das Probenahmeverfahren (Erstvalidierung) oder standortspezifisch für das Verfahren erfolgen, das "vor Ort" für das ausgewählte Zielobjekt verwendet wird (Vor-Ort-Validierung). Die Erstvalidierung wird verwendet, wenn die Probenahme als einmalige Kampagne durchgeführt wird (Einzelprobenahme, z. B. bei der Untersuchung kontaminierter Standorte), während die Vor-Ort-Validierung in regelmäßigen Abständen wiederholt wird (wiederholte Probenahme, z. B. zeit- oder durchflussproportionale Probenahme von Abwässern). Kurz gesagt, die Validierung zeigt, was erreicht werden kann, und wenn dies den Anforderungen an die Zwecktauglichkeit entspricht, gelten die Verfahren als für den Routineeinsatz geeignet.

13.1.3 Die Validierung allein kann jedoch nicht gewährleisten, dass die Routineergebnisse tatsächlich für den Zweck geeignet sind. Routine- oder standortspezifische Bedingungen können systematisch oder gelegentlich von denen der Validierung abweichen. Dies gilt insbesondere für die Probenahme, bei der der größte Teil der Unsicherheitskomponente häufig auf die Heterogenität des Zielobjekts zurückzuführen ist, d. h. der Grad der Heterogenität kann von einem Ziel zum anderen stark variieren. Dies gilt auch, wenn eine Probenahmemethode an verschiedenen Standorten angewendet wird. Diese Umstände unterstreichen die Notwendigkeit einer kontinuierlichen internen Qualitätskontrolle, die auch die Probenahme einschließt, um sicherzustellen, dass die bei der Validierung herrschenden Bedingungen (und damit die erwartete Unsicherheit der Ergebnisse) bei jeder Durchführung der Probenahme- und Analyseverfahren weiterhin gelten. Die kombinierte Anwendung von Validierung und Qualitätskontrolle ist in Tabelle 6 dargestellt.

Tabelle 6: Veranschaulichung des kombinierten Einsatzes von Validierung und Qualitätskontrolle der Probenahme

	Ein an vielen Standorten angewandtes Verfahren	Ein an einem Standort wiederholt angewandtes Verfahren
Validierung	Erstvalidierung mit generischen Leistungsdaten	Vor-Ort-Validierung zur Ermittlung der Leistungsdaten für das jeweilige Zielobjekt
Qualitätskontrolle	Umfassende Qualitätskontrolle mit standortspezifischer Überprüfung der generische Leistungsdaten	Stichprobenartige Qualitätskontrolle zur Überprüfung der Konsistenz der Leistungsdaten im Zeitverlauf

13.1.4 Die Notwendigkeit einer internen Qualitätskontrolle der Probenahme wird derzeit nicht allgemein gesehen, und die Methoden für ihre Durchführung sind, außer in einigen Spezialgebieten wie der geochemischen Prospektion [34], nicht gut etabliert. Konkrete Vorschläge für die Qualitätskontrolle von Probenahmen werden für einige Umweltprobenmatrizes in [35] gegeben. Dabei handelt es sich jedoch nicht um neue Grundsätze; mit geringfügigen Einschränkungen sind die Grundsätze der internen Qualitätskontrolle von Analysen auch auf Probenahmen anwendbar [16, 36, 37]. Außerdem sind die bei der Validierung verwendeten Methoden mit einigen Vereinfachungen auch für die interne Qualitätskontrolle anwendbar. Der Grund für die Vereinfachung ist, dass die Validierung eine gute Schätzung der Unsicherheit liefern muss, während die Qualitätskontrolle lediglich die Übereinstimmung mit der bei der Validierung ermittelten Unsicherheit über Raum und Zeit nachweisen muss.

13.2 Methoden der internen Qualitätskontrolle der Probenahme

13.2.1 Der Schwerpunkt des Interesses liegt fast ausschließlich auf dem Aspekt der Präzision. Bias ist bei der Validierung schwer und bei der internen Qualitätskontrolle fast unmöglich zu berücksichtigen. Das "Referenzzielobjekt", das konzeptionelle Äquivalent eines zertifizierten Referenzmaterials [15], ist selten verfügbar. Außerdem ist der Nutzen eingeschränkt: Wir müssen sehen, ob die Ergebnisse für individuelle Probenahmezielobjekte geeignet sind, und nicht, ob reproduzierbare Ergebnisse ohne Bias für ein möglicherweise nicht repräsentatives Referenzzielobjekt erzielt werden können.

13.2.2 Das wichtigste Instrument ist die Replikation. Sie wird minimal durchgeführt, indem von jedem Ziel zwei Proben durch eine vollständige (und angemessen randomisierte) Verdoppelung der Probenahme genommen werden. Jede Probe wird einmal analysiert und die Differenz zwischen den Ergebnissen $D = |x_1 - x_2|$ berechnet. Wenn die validierten Wiederholstandardabweichungen der Probenahme und der Analyse $s_{r,s}$ bzw. $s_{r,a}$ betragen, ist die kombinierte Wiederholstandardabweichung $s_{r,meas} = \sqrt{s_{r,s}^2 + s_{r,a}^2}$. Folglich kann eine einseitige Spannweitenregelkarte mit einer Warngrenze (95 %-Konfidenzintervall) von $2,83 \times s_{r,meas}$ und einer Eingriffsgrenze (99 %-Konfidenzintervall) von $3,69 \times s_{r,meas}$ erstellt werden [37] (Abbildung 5). Ein außer Kontrolle geratener Wert von d zeigt an, dass das Ergebnis als möglicherweise unbrauchbar betrachtet werden sollte. Ein solches Ergebnis ist nicht diagnostisch und kann auf eine Störung entweder bei der Probenahme oder bei der Analyse zurückzuführen sein; letzteres sollte durch Standardmethoden der analytischen Qualitätskontrolle festgestellt werden.

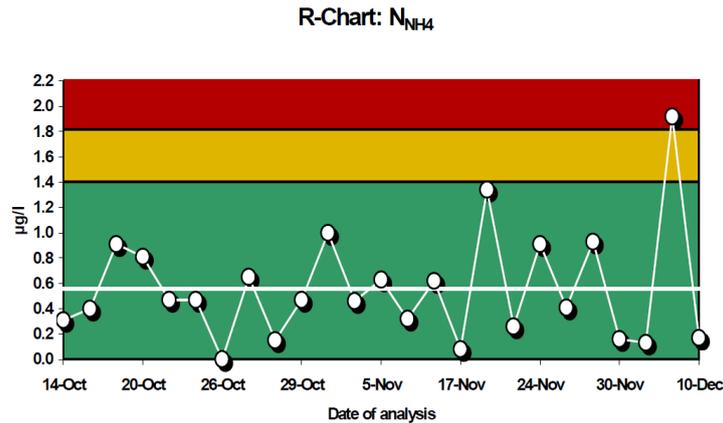


Abb. 4: Beispiel einer Spannweitenregelkarte für die Qualitätskontrolle der Probenahme

Für die Erstellung dieser Spannweiten-Regelkarte siehe Nordtest Guide [37]

13.2.3 Die Daten aus der Qualitätskontrolle können auch verwendet werden, um die Präzision des Probenahmeverfahrens zu aktualisieren, die bei der Methodvalidierung mit denselben Methoden, ANOVA oder relativen Differenzberechnungen ermittelt wurde.

13.2.4 In einigen Fällen können die zusätzlichen Kosten für die doppelte Probenahme durch die Anwendung der SAD-Methode (Split Absolute Difference) vermieden werden, bei der die normale Anzahl von Einzelproben, die als Probe zu kombinieren sind, nach dem Zufallsprinzip in zwei gleich große Teilmengen aufgeteilt wird, von denen jede separat verarbeitet und analysiert wird [38, 39]. Die Differenz zwischen solchen Ergebnissen hat eine Unsicherheit von $\sqrt{4u_s^2 + 2u_a^2}$, wenn die Bedingungen eingehalten werden, die bei der Validierung galten. Dies wiederum könnte zur Festlegung einer Eingriffsgrenze in einer einseitigen Regelkarte verwendet werden.

14 Berichterstattung und Interpretation von Unsicherheit

14.1 Einleitung

14.1.1 Es muss unbedingt sichergestellt werden, dass in den Berichten eindeutig angegeben wird, um welche Messgröße es sich handelt. Insbesondere muss klar sein, ob sich das Ergebnis und seine Unsicherheit auf eine einzelne Prüfmengen, eine Laborprobe, das gesamte Probenahmezielobjekt (z. B. ein Schüttgut) oder eine Reihe von Zielobjekten bezieht. Unter Anwendung der Grundsätze des GUM [2] und früherer Eurachem/CITAC-Leitfäden [z. B. 1] wird die Unsicherheit zunächst in Form einer Standardunsicherheit u geschätzt, die alle Effekte berücksichtigt, die das Ergebnis sinnvollerweise beeinflussen können. Die Unsicherheit kann ohne Änderung in dieser Form angegeben werden. Oft ist es jedoch zweckmäßig, in anderer Form, mit einem höheren Vertrauensniveau oder für eine erweiterte Anwendbarkeit, zu berichten. Es ist wichtig, alle Einschränkungen bei der Schätzung der Unsicherheit anzugeben, wie z. B. den Ausschluss von Probenahme-Bias oder anderen vernachlässigten Effekten. In den folgenden Abschnitten werden die wichtigsten Punkte beschrieben und einige Hinweise zu deren Interpretation gegeben.

14.2 Erweiterte Unsicherheit

14.2.1 Die Standardunsicherheit u , die auf ein Ergebnis in der Form $x \pm u$ angewendet wird und mit einer Normalverteilung verbunden ist, beschreibt ein Intervall, das nur etwa 68 % der Fläche der Verteilung umfasst. Dies bedeutet in der Regel, dass die Wahrscheinlichkeit, dass der Messwert außerhalb dieses Intervalls liegt, größer als 32 % ist. Dies wird für die meisten praktischen Anwendungen als unzureichende Sicherheit angesehen. Es ist daher üblich, einen geeigneten Multiplikator auf die Standardunsicherheit anzuwenden, so dass das angegebene Intervall einen größeren Teil der Streuung umfasst. Üblicherweise wird dieser Multiplikator, der normalerweise mit k bezeichnet wird, als *Erweiterungsfaktor* bezeichnet, und das Produkt $ku = U$ wird als *erweiterte Unsicherheit* bezeichnet.

14.2.2 Die Wahl von k wird in anderen Veröffentlichungen ausführlich erörtert [1, 2]. Die wichtigsten Grundsätze sind jedoch folgende:

- Das k spiegelt für die jeweilige Verteilung ein ungefähres Konfidenzintervall wider.
- Wenn bekannt ist, dass eine bestimmte Verteilung anwendbar ist, wird sie verwendet. Andernfalls wird eine Normalverteilung als vernünftige Annahme betrachtet, wenn die vorherrschenden Beiträge zur Unsicherheit alle normalverteilt sind oder es mehrere annähernd gleiche Beiträge aus beliebigen Verteilungen gibt. Bei dieser Annahme basiert k typischerweise auf dem Wert von Student's t für ein angemessenes (zweiseitiges) Vertrauensniveau und eine angemessene Anzahl von Freiheitsgraden.
- Beim Modellierungsansatz wird die Anzahl der Freiheitsgrade formal aus den Freiheitsgraden für die beitragenden Unsicherheiten nach einer veröffentlichten Formel [1, 2] abgeleitet oder aus der Anzahl der Freiheitsgrade für den dominierenden Beitrag [1] angenähert. In der Regel wird angenommen, dass die Anzahl der Freiheitsgrade ausreichend groß ist, um die Wahl von $k = 2$ für ein Vertrauen von etwa 95 % zu rechtfertigen.

Für die meisten praktischen Zwecke wird $k = 2$ als akzeptabel angesehen und ist manchmal sogar vorgeschrieben [40]. Es ist jedoch wichtig, den verwendeten Wert von k und das

ungefähre Vertrauensniveau, das k impliziert, bei der Angabe der erweiterten Unsicherheit anzugeben.

14.3 Aussagen zur relativen Unsicherheit

14.3.1 Häufig stellt man fest, dass die Standardunsicherheit bei Stichproben ungefähr proportional mit dem Wert des Ergebnisses zunimmt. Unter diesen Umständen ist es oft am praktischsten, die Unsicherheit in einer relativen Form anzugeben, z. B. als relative Standardabweichung u/\bar{x} oder als prozentuales Intervall unter Verwendung von Gleichung 4 (z. B. 10 %). Der angegebene relative Wert basiert in der Regel auf einer Schätzung der Unsicherheit für ein oder mehrere repräsentative Ergebnisse, ist aber über einen größeren Bereich von Konzentrationswerten anwendbar.

14.3.2 Es ist wichtig, eine einfache relative Standardabweichung nicht auf eine Konzentration von Null zu extrapolieren, da die Unsicherheit bei sehr niedrigen Konzentrationen normalerweise nicht vollständig verschwindet und die Proportionalitätsannahme nicht mehr gültig ist. Allgemeinere Ansätze für diese Situationen können entweder einen Konzentrationsbereich angeben, für den der relative Unsicherheitswert gilt [37], oder die Unsicherheit als eine Funktion der Konzentration ausdrücken [1,10]. Ausgearbeitete Beispiele, die die Berechnung und die Angabe der Unsicherheit zeigen, sind in Anhang A aufgeführt. In einigen Beispielen erfolgt die Berechnung der Unsicherheit in den ursprünglichen Konzentrationseinheiten (A1 und A2), in anderen wird die relative Unsicherheit verwendet (A3-A6), obwohl in allen Beispielen die Unsicherheit schließlich in relativer Form ausgedrückt wird.

14.4 Unsicherheitsfaktor

Wie in Abschnitt 9.5.3 [23] erläutert, ist es manchmal angemessener, die Messunsicherheit als erweiterten Unsicherheitsfaktor ($^F U$) auszudrücken. Dies ist der Fall, wenn die Häufigkeitsverteilung des Effekts annähernd lognormal ist, was bei hohen Werten der Stichprobenunsicherheit nicht ungewöhnlich ist.

14.5 Beiträge zur Unsicherheit

Die genauen Schritte, die in jedem Beitrag zur Messunsicherheit enthalten sind, müssen angegeben werden. Es ist möglich, einige spezifische Komponenten der Messunsicherheit zu quantifizieren, aber die erforderliche Methodik hängt von der angewandten Schätzmethode, den Details des Versuchsplans und der Person ab, für die die Informationen bestimmt sind. Der Versuchsplan in Abbildung 2 liefert beispielsweise getrennte Schätzungen für zwei Komponenten, die als "Probenahme" und "Analytik" bezeichnet werden. Wenn die Details dieser speziellen Umsetzung des Versuchsplans untersucht werden, wird deutlich, dass die Unsicherheit aus der physikalischen Probenvorbereitung unter dem allgemeinen Titel "Probenahme" erfasst wird, während die Unsicherheit aus der chemischen Vorbereitung unter "Analytik" erfasst wird. Falls erforderlich, ist es möglich, eine weitere Ebene der Verdoppelung der physikalischen Vorbereitung in den Versuchsplan einzufügen, um den separaten Beitrag zu schätzen, den dieser spezielle Schritt einbringt [22]. Wenn die Probenahmeunsicherheit als Unsicherheitsfaktor (im logarithmischen Bereich) und die analytische Unsicherheit als relative Unsicherheit (im linearen Bereich) ausgedrückt wird, ist es möglich, diese beiden Komponenten zu addieren [21]. Ein Ansatz besteht darin, sowohl die

Probenahme- als auch die analytische Unsicherheitskomponente zu berechnen und logarithmisch auszudrücken, wie in Beispiel A2 gezeigt. Ein zweiter Ansatz besteht darin, für die analytische Komponente anzunehmen, dass die relative Standardunsicherheit ($s'_{analytical}$) ungefähr gleich der Standardabweichung der natürlichen Logarithmen ($s_{G,analytical}$) ist. Dies ist ein akzeptabler Näherungswert, wenn $s'_{analytical} < 0,2$ ist, was normalerweise der Fall ist. Die beiden Komponenten können dann als Varianzen im logarithmischen Raum addiert werden, wie im ersten Ansatz. Die genauen Schritte, die in jedem Beitrag zur Messunsicherheit enthalten sind, müssen dokumentiert werden, wie auch immer sie geschätzt werden. Für weniger erfahrene Benutzer von analytischen Messungen kann es besser sein, einen Wert für die gesamte Messunsicherheit anzugeben und dabei zu vermerken, welche Quellen berücksichtigt wurden.

14.6 Anwendbarkeit der Schätzungen

Unter Hinweis auf die Erörterung der Spezifikation der Messgröße (Abschnitt 5.2) ist es von entscheidender Bedeutung, dass in den Berichten klar angegeben wird, um welche Messgröße es sich handelt. Wie in Abschnitt 14.1.1 ausgeführt, ist es besonders wichtig, klar anzugeben, ob sich das Ergebnis und seine Unsicherheit auf eine einzelne Prüfmenge, eine Laborprobe, das gesamte Probenahmezielobjekt oder auf eine Reihe von Zielobjekten beziehen. Im Gegensatz zu Unsicherheitsschätzungen für analytische Messungen ist es sehr wahrscheinlich, dass dieselbe Probenahmeverfahren zu Messungen mit unterschiedlich großer Unsicherheit führt, wenn sie auf ein neues Probenahmezielobjekt angewendet wird. Neue Schätzungen sind für wesentlich andere Zielobjekte erforderlich, insbesondere dann, wenn es Grund zu der Annahme gibt, dass sich der Grad der Heterogenität geändert hat.

14.7 Interpretation von Unsicherheitsangaben gegen Grenzwerte

Ergebnisse werden häufig mit Toleranzen oder Grenzwerten verglichen, um die Einhaltung einer Anforderung zu bewerten. Bei solchen Vergleichen ist es wichtig, die Unsicherheit zu berücksichtigen. Eine ausführliche Erörterung würde den Rahmen dieses Leitfadens sprengen; detailliertere Informationen sind in den Referenzen [1] und [41] zu finden. Die Grundprinzipien sind:

- Entscheiden Sie, ob die Entscheidung den Nachweis der Einhaltung der Vorschriften, den Nachweis der *Nichteinhaltung* der Vorschriften oder einen Ansatz mit "*geteiltem Risiko*" erfordert, und legen Sie ein angemessenes Vertrauensniveau fest.
- Zum Nachweis der Einhaltung müssen das Ergebnis und sein Unsicherheitsintervall vollständig innerhalb des zulässigen Bereichs liegen.
- Zum Nachweis der *Nichteinhaltung* müssen das Ergebnis und sein Unsicherheitsintervall vollständig *außerhalb* des zulässigen Bereichs liegen.
- Bei Ansätzen mit geteiltem Risiko ist ein Bereich für akzeptable Messergebnisse auf der Grundlage des zulässigen Intervalls festzulegen, der so angepasst wird, dass eine bestimmte Wahrscheinlichkeit falscher Akzeptanz und falscher Ablehnung gegeben ist. Neuere Leitlinien enthalten nützliche Details zum Verfahren [42].

Für regulatorische Zwecke ist es wichtig, die jeweils geltenden Vorschriften zu konsultieren, da derzeit keine allgemeine Anleitung alle Fälle abdecken kann. So gilt es beispielsweise im Allgemeinen als unsicher, Material "weiterzugeben", das nachweislich nicht konform ist, so

dass ein Nachweis der Einhaltung erforderlich ist. Die Strafverfolgung in den meisten Ländern erfordert jedoch eindeutige Beweise für die *Nichteinhaltung* der Vorschriften, und unter diesen Umständen (z. B. bei der Strafverfolgung im Zusammenhang mit Blutalkohol) ist es gängige Praxis, den Nachweis der Nichteinhaltung mit hoher Zuverlässigkeit zu verlangen.

15 Kosten für die Schätzung der Unsicherheit der Probenahme

15.1 Es erscheint logisch, das Gesamtbudget für die Validierung und Qualitätskontrolle von Probenahmen gegen die Kosten abzuwägen, die durch Fehlentscheidungen aufgrund unzureichender Unsicherheitsschätzungen entstehen. Es ist bekannt, dass die Durchführung der Unsicherheitsschätzung die Gesamtkosten der Messung erhöht. Die Anwendung der Duplikatmethode kann beispielsweise die Kosten für die Probenahme um bis zu 10 % und für die Analyse um 30 % erhöhen (d. h. es sind drei zusätzliche Analysen erforderlich, um das ausgewogene Design auf 10 % der Probenahmeziele anzuwenden). Diese höheren Kosten lassen sich jedoch durch die zusätzlich gewonnenen Informationen und die geringeren potenziellen Verluste aufgrund falscher Entscheidungen rechtfertigen, die ohne Kenntnis der Unsicherheit getroffen worden wären (Abschnitt 16).

Anmerkung: Ein kostengünstigerer nicht-ausgewogener Versuchsplan kann durchgeführt werden, bei dem nur eine der Doppelproben zweimal analysiert wird, wenn eine geeignete statistische Behandlung vorgenommen wird (Anhang D, [20]).

15.2 Es ist schwieriger, die allgemeinen Kosten für die anderen Methoden der Unsicherheitsschätzung zu ermitteln. Organisationsübergreifende Probenahmeversuche erfordern die Kosten von mindestens acht verschiedenen Teilnehmern (um eine akzeptable Zuverlässigkeit zu erreichen [17]) und sind daher wahrscheinlich wesentlich höher als die Kosten für die Duplikatmethode. Modellierungsmethoden erfordern detaillierte Informationen über das zu beprobende Material. Für einige Materialien, die über viele Chargen hinweg relativ konstant sind, können diese Werte allgemein anwendbar sein und machen diesen Ansatz daher kosteneffizienter als empirische Methoden, die eine größere Anzahl zusätzlicher Messungen bei jeder Charge erfordern. Bei dieser Diskussion muss daher berücksichtigt werden, inwieweit der Unsicherheitswert für eine bestimmte Vorschrift/Material-Kombination bei einer Erstvalidierung geschätzt wird und inwieweit der Wert kontinuierlich überwacht und/oder durch ein laufendes Qualitätskontrollprogramm für die Probenahme aktualisiert wird (Abschnitt 1). Es erscheint logisch, das Gesamtbudget für die Validierung und Qualitätskontrolle der Probenahme zusammen mit den Kosten zu betrachten, die durch Fehlentscheidungen auf der Grundlage unzureichender Unsicherheitsschätzungen entstehen.

16 Beurteilung der Zwecktauglichkeit von Messungen anhand der Unsicherheit

16.1 Ein angemessenes Verständnis der Probenahmeunsicherheit muss in die breitere Perspektive der Zwecktauglichkeit eingebettet sein. Es wurden drei Ansätze für die Festlegung von Zwecktauglichkeitskriterien vorgeschlagen. Der erste Ansatz besteht darin, einen Grenzwert für den maximalen Wert der Unsicherheit (d. h. die **Zielunsicherheit**) [43, 44] festzulegen, der als akzeptabel angesehen wird. Dieser Ansatz ist im analytischen Bereich weit verbreitet, wo eine relative Zielunsicherheit (z. B. 10 %) verwendet wird. Das Problem bei diesem Ansatz ist, dass er sich nicht unbedingt auf den beabsichtigten Zweck bezieht, für den der Benutzer die Messung benötigt.

16.2 Der zweite Ansatz besteht darin, die durch die Messung (Probenahme und Analyse) erzeugte Varianz mit der Varianz der Messungen zwischen den verschiedenen Probenahmezielobjekten zu vergleichen. Es gibt viele Situationen, in denen das Ziel der Messungen darin besteht, Konzentrationen zwischen verschiedenen Zielobjekten zu vergleichen, wie z. B. bei der Mineralexploration, wo das Ziel darin besteht, ein Zielobjekt mit einer signifikant höheren Konzentration eines Elements von Interesse (z. B. Gold) zu finden. Bei einer Anwendung dieses Ansatzes wird das Kriterium der Zwecktauglichkeit beispielsweise so festgelegt, dass die Messabweichung nicht mehr als 20 % zur Gesamtabweichung beiträgt (definiert in Gleichung 2) [45].

16.3 Der dritte und am häufigsten anwendbare Ansatz zur Beurteilung der Zwecktauglichkeit von Messungen besteht darin, die Auswirkungen der Messung auf ihren eigentlichen Zweck zu berücksichtigen. Alle analytischen Messungen werden durchgeführt, um eine Entscheidung zu unterstützen. Eine Entscheidung kann entweder richtig oder falsch sein. Eine falsche Entscheidung ist mit zusätzlichen Kosten verbunden, und eine falsche Entscheidung ist wahrscheinlicher, wenn die Unsicherheit größer ist. Nehmen wir zum Beispiel die Herstellung eines Materials, für das ein maximal zulässiger Gehalt an Verunreinigungen vorgegeben ist^h. Jede Materialcharge wird analysiert, um den Gehalt an Verunreinigungen zu bestimmen. Ein "falsch positives" Ergebnis hat zur Folge, dass die Materialcharge verworfen oder unnötig nachbearbeitet wird, um den scheinbar inakzeptablen Verunreinigungsgrad zu verringern. Ein "falsch negatives" Ergebnis bedeutet, dass eine fehlerhafte Charge an den Kunden ausgeliefert wird, was eine finanzielle Entschädigung erforderlich machen kann. Beide Situationen sind wahrscheinlicher, wenn die Unsicherheit höher ist. Dies scheint darauf hinzudeuten, dass die Messung so durchgeführt werden sollte, dass die Unsicherheit so gering wie möglich ist. Die Verringerung der Unsicherheit eines Messergebnisses verursacht jedoch schnell steigende Kosten. Eine nützliche Regel besagt, dass in Fällen, in denen zufällige Schwankungen die Unsicherheit dominieren, die Kosten einer Messung umgekehrt proportional zum Quadrat der Unsicherheit sind; eine Verringerung der Unsicherheit um einen Faktor 2 verursacht einen Anstieg der Ausgaben um den Faktor 4.

16.4 Die wahren Kosten einer Entscheidung ergeben sich aus der Summe der Messkosten und der Mehrkosten für Fehlentscheidungen. Aus den obigen Ausführungen geht hervor, dass diese Summe bei einem bestimmten Unsicherheitsgrad einen Minimalwert hat (Abbildung 6), und diese Unsicherheit ist die Definition der Zwecktauglichkeit.

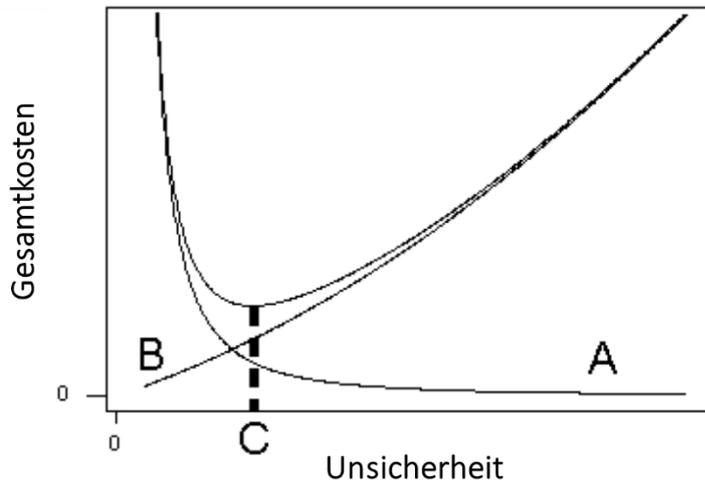


Abb. 6: Schematische Verlustfunktionen in Abhängigkeit von der Messunsicherheit

Die Linie A zeigt die Kosten der Messung. Linie B zeigt die Kosten von Fehlentscheidungen. Die Summe dieser beiden Linien (die Gesamtkosten, die durch die oberste Linie dargestellt werden) zeigt ein Kostenminimum bei Punkt C, der Unsicherheit, die als zwecktauglich angesehen werden kann.

16.5 Die optimale Aufteilung der Ressourcen zwischen Probenahme und Analyse ist auch eine Frage der Kosten. Schon eine elementare Betrachtung (ohne Berücksichtigung der Kosten) zeigt, dass die Unsicherheiten von Probenahme und Analyse ungefähr ausgeglichen sein sollten. Wenn beispielsweise die Unsicherheiten der Probenahme und der Analyse 10 bzw. 3 Einheiten betragen, ist die Gesamtunsicherheit der Messung $\sqrt{10^2 + 3^2} = 10,4$. Die Gesamtunsicherheit wird durch eine Verringerung der Unsicherheit der Analyse kaum beeinflusst: Wird sie auf (sagen wir) 1 Einheit reduziert, verringert sich die Gesamtunsicherheit auf $\sqrt{10^2 + 1^2} = 10,05$, eine unbedeutende Änderung. Ein differenzierterer Ansatz berücksichtigt die unterschiedlichen Kosten von Analyse und Probenahme. Wenn die Einheitskosten der Probenahme und der Analyse A und B für das gleiche spezifische Unsicherheitsniveau sind, ist das optimale Verhältnis der Probenahmeunsicherheit u_{samp} zur Analyseunsicherheit u_{anal} gegeben durch

$$\frac{u_{samp}}{u_{anal}} = \left(\frac{A}{B}\right)^{1/4}$$

Dieses Verhältnis liefert die Mindestausgaben für eine gegebene Gesamtunsicherheit von

$\sqrt{u_{samp}^2 + u_{anal}^2}$ oder, alternativ, die minimale Unsicherheit für einen gegebenen Aufwand [46].

Methoden zur Verringerung der Unsicherheit bei der Probenahme werden in Anhang E erörtert, auch wenn es nicht immer möglich oder notwendig ist, mit "minimalen Gesamtkosten" zu arbeiten.

^h Dieses Konzept ist auch auf Situationen anwendbar, in denen Materialien regulierte Mindestkonzentrationen von Analyten aufweisen; in diesem Fall sind die Begriffe "falsche Übereinstimmung" und "falsche Nichtübereinstimmung" anwendbar.

17 Auswirkungen auf die Planung von Probenahme- und Messstrategien

17.1 Fachwissen und Beratung

Wie Abschnitt 4 zeigt, umfassen die Probenahme- und Analyseverfahren eine Reihe von Tätigkeiten. Verschiedene Teile des Prozesses werden häufig verschiedenen Mitarbeitern zugewiesen, die möglicherweise sehr unterschiedliche Kenntnisse über die Ziele und - was noch wichtiger ist - unterschiedliche Kenntnisse über die Auswirkungen der verschiedenen Teile des Prozesses haben. Im Allgemeinen haben alle Beteiligten gute Kenntnisse über einen Teil des Prozesses, aber nur wenige sind in der Lage, über den gesamten Prozess Ratschläge zu geben. Daher ist es wichtig, dass die Probenplaner nach Möglichkeit analytische Chemiker und erfahrene Probenehmer in die Planung der Probenahme einbeziehen. In den meisten Fällen ist es auch ratsam, Statistikexperten einzubeziehen (siehe unten). Entscheidungsträger (d. h. Unternehmensmanager und diejenigen, die auf der Grundlage der Ergebnisse von Probenahmeaktivitäten handeln) sollten in die Planung neuer Anwendungen einbezogen werden, und auch die Aufsichtsbehörden sollten konsultiert werden, wenn eine Vorschriften zur Unterstützung von gesetzlichen Regularien dienen soll.

Obwohl die Grundsätze dieses Leitfadens weitgehend anwendbar sind, ist eine statistische Beratung durch Experten immer wertvoll und sollte unter bestimmten Umständen als unerlässlich angesehen werden, zum Beispiel:

- wenn die beobachteten oder erwarteten Häufigkeitsverteilungen nicht normal sind, z. B. wenn die Ergebnisse mehr als 10 % Ausreißer enthalten, oder wenn die Ergebnisse deutlich asymmetrische Verteilungen aufweisen;
- wenn große finanzielle oder soziale Auswirkungen von einer zuverlässigen Schätzung der Unsicherheit abhängen;
- wenn Konfidenzintervalle für die Unsicherheitsschätzungen oder - bei komplexeren Probenahmeplänen - für die Messergebnisse benötigt werden;
- wenn die Probenahmestrategie komplexer ist als eine einfache Zufallsstichprobe mit wiederholten Messungen, z. B. bei der Durchführung einer Probenahme geschichteter Stichproben.

17.2 Probenahme-Bias vermeiden

Die in diesem Leitfaden beschriebenen Methoden sind geeignet, um die Variabilität der Probenahme zu ermitteln, aber nur die komplexeren Methoden sind in der Lage, Unsicherheiten im Zusammenhang mit möglichem Bias bei der Probenahme zu bewerten. Aus diesem Grund sollte der Minimierung potenzieller Quellen von Bias besondere Aufmerksamkeit gewidmet werden. Dazu gehört möglicher Bias bei Probenahmen, bei denen Partikelgröße, Dichte oder Durchflussmenge einen Unterschied machen, ein Bias bei der Auswahl der Probenahmepunkte, Auswirkungen unterschiedlicher Probenahmegeräte usw. Es sollte spezielles Fachwissen zur Probenahmemethodik eingeholt werden, es sei denn, diese Faktoren können nachweislich angemessen kontrolliert werden oder sind durch eine festgelegte Probenahmevervorschrift vollständig spezifiziert.

17.3 Planung für Unsicherheitsschätzungen

Bei der Probenahme sollten immer zumindest einige Wiederholungsproben und -messungen vorgesehen werden, um die Unsicherheit der Ergebnisse beurteilen zu können.

17.4 Kriterien für die Zwecktauglichkeit

Die Planung sollte idealerweise mit der Festlegung klarer Zwecktauglichkeitskriterien beginnen, wobei die relativen Kosten und Unsicherheiten von Probenahme und Analyse zu berücksichtigen sind, sofern sie bekannt sind oder sinnvoll im Voraus bestimmt werden können. Abschnitt 16 enthält Hinweise darauf, wie der Analyse- und Probenahmeaufwand optimiert werden kann.

17.5 Verwendung von Daten aus früheren Validierungen

Die wichtigsten Unsicherheiten im Zusammenhang mit analytischen Messungen werden häufig während oder auf der Grundlage der Validierung der Analysemethode geschätzt, einem Prozess, der vor der Anwendung der Methode durchgeführt wird. Dementsprechend ist zu überlegen, ob die im Rahmen des Probenahmeexperiments festgestellte Variabilität die anhand früherer Informationen abgeschätzte analytische Messunsicherheit ersetzen, berücksichtigen oder einfach als Kontrolle dienen sollte. Bei der Betrachtung dieser Frage ist zu beachten, dass die während einer relativ kurzen Analysenserie beobachtete Variabilität selten als Schätzung der Unsicherheit ausreicht. Langfristige Studien sind im Allgemeinen zuverlässiger. Dementsprechend ist es sicherer, sich auf vorherige Validierungsdaten zu verlassen, es sei denn, die beobachtete Schwankung ist deutlich höher.

Die mit der Variabilität der Probenahme verbundenen Unsicherheiten können selbst im Voraus geschätzt werden, insbesondere wenn ein langfristiges Probenahmeprogramm geplant und durchgeführt werden soll. Unter diesen Umständen ist es in der Regel ratsam, eine erste Schätzung der Probenahmeunsicherheit vorzunehmen. Laufende Untersuchungen können dann zur Überprüfung der fortlaufenden Gültigkeit der Unsicherheitsschätzung dienen, z. B. durch Anwendung interner Qualitätskontrollprinzipien, wie in Abschnitt 13 beschrieben.

17.6 Angemessenheit der Probenahmeunsicherheit

Vor dem Berichten von Messergebnissen sollte bewertet werden, ob sie akzeptabel sind und mit den Qualitätszielen übereinstimmen, die für die gesamte Unsicherheit und ihre Probenahmekomponente festgelegt wurden, wahrscheinlich auf der Grundlage eines Eignungskriteriums vor der Durchführung der Messungen.

17.7 Unsicherheitsschätzung für Vor-Ort- und *In-situ*-Messungen

Es wurde nachgewiesen, dass die "Duplikatmethode" der empirischen Unsicherheitsschätzung auf Messungen angewandt werden kann, die sowohl vor Ort [47] als auch *in situ* [48] (d. h. ohne Entnahme einer Prüfmenge) durchgeführt werden. Sie ist auf *In-situ*-Messungen anwendbar, die sowohl auf der Makro- (cm) als auch auf der Mikroskala (μm) durchgeführt werden [49]. Es gibt jedoch noch einige offene Fragen für *In-situ*-Messungen. So ist beispielsweise bei Feldböden nicht klar, ob *In-situ*-Messungen und entsprechende Labormessungen vergleichbar sein sollten, wenn eine für die Labormessung verwendete Prüfmenge getrocknet, gesiebt oder anderweitig verändert wurde. Eine solche Behandlung kann sich eindeutig auf den offensichtlichen "Bias" zwischen laborgestützten und *In-situ*-

Messungen auswirken, was wiederum die Bewertung der Unsicherheit beeinflusst. Ein ähnliches Problem ergibt sich bei der spezifischen Anwendung auf *In-situ*-Messungen der Radioaktivität, bei denen Labormessungen in der Regel an einer weitaus kleineren Prüfmenge vorgenommen werden als bei der *In-situ*-Messung von z. B. Gammastrahlen [50].

Anhang A: Beispiele

Einleitung

Der effektivste Weg, die im Haupttext dieses Leitfadens beschriebenen Methoden zu erläutern, ist das Aufzeigen von Beispielen aus der Praxis. Diese Beispiele sollen nicht alle Umstände abdecken, sondern zeigen, wie die allgemeinen Grundsätze auf eine Vielzahl von Situationen in einer Reihe verschiedenen Sektoren angewendet werden können. Dazu gehören Lebensmittel (Produktion und Einzelhandel), Tierfutter und Umwelt (Boden und Wasser). Die Beispiele sind alle nach dem gleichen Schema aufgebaut, um Verständlichkeit und die Vergleichbarkeit zu erleichtern.

Anmerkung: Um eine Überprüfung der Berechnungen zu ermöglichen, werden in diesen Beispielen mehr Stellen angegeben, als normalerweise für die Messunsicherheit angegeben werden würden. In der Praxis ist es selten sinnvoll, bei der Angabe der Messunsicherheit mehr als zwei signifikante Ziffern anzugeben ([1], Abschnitt 9.5.1).

Beispiel A1: Nitrat in Gewächshaus-Kopfsalat

Messgröße				Unsicherheitsschätzung		
Analyt/ Technik	Einheit	Sektor/ Matrix	Probenahme- zielobjekt(e)	Zweck	Design	Statistik
Nitrat / Heißwasser- extraktion und Bestimmung mit HPLC	mg kg ⁻¹	Lebens- mittel / Kopf- salat	1 Partie Eisbergsalat, gezogen unter Glas	Unsicherheit – gesamte Messung, Probenahme und Analytik	Empirisch – Duplikat- methode	Robuste ANOVA mit Konzentrations- werten

1 Anwendungsbereich

Schätzung der Messunsicherheit und der Beiträge von Probenahme und Analyse für die routinemäßige Überwachung von im Gewächshaus angebautem Kopfsalat unter Verwendung einer Standard-Probenahmervorschrift.

2 Szenario und Probenahmezielobjekt

Nitrat ist für die Pflanzengesundheit unerlässlich; es gibt jedoch Bedenken hinsichtlich der menschlichen Gesundheit beim Verzehr erhöhter Nitratkonzentrationen. Die Nitratkonzentration in Salat wird gemäß den EG-Vorschriften regelmäßig überwacht. Für jede "Partie" mit bis zu 20.000 Salatköpfen werden Konzentrationsschätzungen vorgenommen, und die Ergebnisse für jede Partie werden einzeln zur Beurteilung der Übereinstimmung mit der einschlägigen Verordnung herangezogen. Jede Partie wird daher als ein Probenahmezielobjekt betrachtet und nicht einzelne Salatköpfe. Um einen zuverlässigen Vergleich der gemessenen Nitratkonzentrationen mit dem europäischen Grenzwert [51] (4500 mg kg⁻¹) zu ermöglichen, ist eine Schätzung der Messunsicherheit wünschenswert.

3 Probenahmervorschrift

Die anerkannte Vorschrift für diesen Zweck sieht vor, dass aus 10 Salatköpfen, die aus jeder Salatpartie geerntet werden, eine Mischprobe hergestellt wird [52]. Die Salate werden ausgewählt, indem man eine W-Form oder eine Fünf-Punkte-Würfelform durch die zu untersuchende Partie geht. Dieses Protokoll wird auf alle Partien angewandt, unabhängig von ihrer Größe. Die Proben wurden am Morgen entnommen und in eisgekühlten Kühlboxen zum beauftragten Analyselabor transportiert, wo sie innerhalb von 24 Stunden nach der Probenahme eintrafen.

4 Studiendesign - Duplikatmethode (Abschnitt 9.4.2)

Das Minimum von acht Zielobjekten wurde ausgewählt, die in die Vorschrift zur Unsicherheitsschätzung aufgenommen wurden. Für jede dieser Partien wurde zusätzlich zur Routineprobe (S1) eine zweite 10-Kopf-Probe genommen (S2). Diese Doppelprobe wurde so entnommen, dass sie die Variationen repräsentiert, die aufgrund von Unklarheiten in der

Probenahmeverfahren auftreten können, z. B. die Positionierung des Ursprungs des W-Designs und seine Ausrichtung.

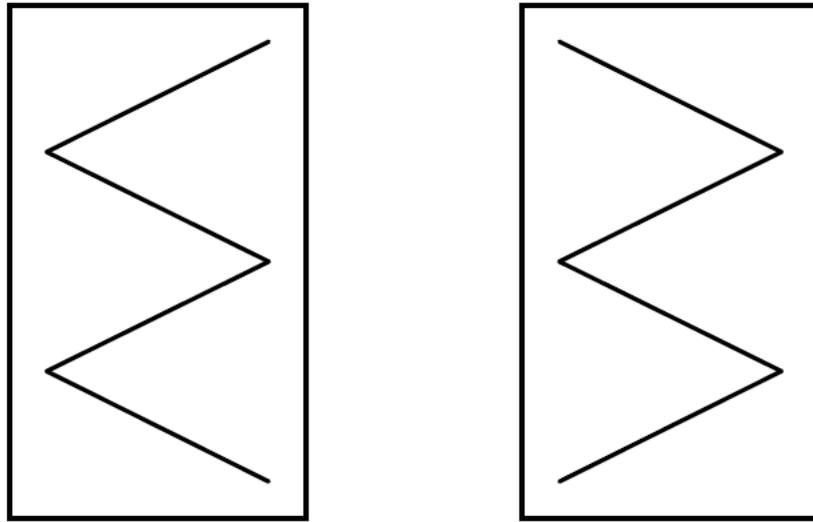


Abb. A1.1: Beispiel für die „Duplikatmethode“

Beispiel für die Anwendung der Duplikatmethode. Am Beispiel des W-Musters: Die Vorschrift schreibt das Muster vor, nicht aber die Position oder Ausrichtung. Das "W" kann gleichermaßen auf der linken oder rechten Seite beginnen. Entlang der Linie des W werden zehn Köpfe entnommen, um eine Mischprobe für ein Zielobjekt zu erstellen.

5 Probenahme und Analyse im Labor

Die Primärproben wurden bei ihrem Eingang im Labor eingefroren. Von jeder 10-Kopf-Probe wurde ein Kopfsalat (Einzelstück) in vier gleiche Viertel geschnitten und zwei Viertel aufbewahrt. Dies wurde für jede der 10 Einzelproben wiederholt. Die daraus resultierenden 20 Viertel wurden in einen Hobart Food Processor gegeben und mazeriert, um eine Mischprobe zu erhalten. Es wurden zwei analytische Prüfmengen (10 g) entnommen. Jede Prüfmenge wurde mit heißem Wasser extrahiert und die Nitratkonzentration mittels HPLC (Ultraviolett-detektor) bestimmt. Die Qualitätskontrollproben (Wiederfindung einer Aufstockung) wurden gleichzeitig mit den realen Proben analysiert. Es konnte kein signifikanter analytischer Bias festgestellt werden, so dass eine Biaskorrektur für die resultierenden Daten als unnötig erachtet wurde. Die ursprünglichen Messwerte, die für die Schätzung der Unsicherheit verwendet wurden, waren angemessen gerundet und es wurden keine Werte unterdrückt, die entweder unter Null oder unter der Nachweisgrenze lagen.

6 Ergebnisse

Die besten Schätzungen der Nitratkonzentration an jedem der acht Zielorte sind in Tabelle A1.1 aufgeführt.

Tabelle A1.1: Messungen der Nitratkonzentration (mg kg^{-1}) in acht doppelten Proben. Die Doppelproben sind mit S1 und S2 gekennzeichnet. Ebenso sind die Doppelanalysen A1 und A2 bezeichnet. DS1A2 (Wert 4754 mg kg^{-1}) ist somit die Analyse 2 von Probe 1 aus Probenahmezielobjekt D

Probenahme-Zielobjekt	S1A1	S1A2	S2A1	S2A2
A	3898	4139	4466	4693
B	3910	3993	4201	4126
C	5708	5903	4061	3782
D	5028	4754	5450	5416
E	4640	4401	4248	4191
F	5182	5023	4662	4839
G	3028	3224	3023	2901
H	3966	4283	4131	3788

Vor der Anwendung statistischer Methoden ist es sinnvoll, die Daten zu prüfen, um das allgemeine Niveau der Variabilität zu ermitteln. Die analytischen Duplikate (z. B. BS1A1 und BS1A2) liegen im Allgemeinen innerhalb von 300 mg kg^{-1} , was auf eine analytische Präzision von weniger als 10 % schließen lässt. Die Probenduplikate (z. B. DS1 und DS2) stimmen weniger gut überein, unterscheiden sich aber im Allgemeinen um weniger als 20 %. Ein Zielwert (C) weist jedoch eine größere Differenz auf, was auf einen Ausreißer-Wert schließen lässt.

Die Quantifizierung der Zufallskomponente der Messunsicherheit und von zwei ihrer Hauptkomponenten (Probenahme und Analyse) erfolgte mit Hilfe der robusten Varianzanalyse (RANOVA [53], Anhang C3), deren Ergebnisse in Abbildung A1.2 dargestellt sind. Die robuste ANOVA wurde hier verwendet, da abweichende Zielobjekte in dieser relativ gut kontrollierten Umgebung eher als Anomalien betrachtet werden und nicht die zugrundeliegende Statistik der Grundgesamtheit widerspiegeln, und als Vorsichtsmaßnahme gegen analytische Ausreißer.

Anmerkung: Robuste Methoden sollten nicht verwendet werden, wenn offensichtliche Ausreißer als Teil der typischen Population von Einzelproben oder beprobten Zielobjekten auftreten, es sei denn, die spezifische Implementierung berücksichtigt nicht-normale Verteilungen für einen Teil der angenommenen Fehlerstruktur ([33, 23]).

ERGEBNISSE DER KLASSISCHEN ANOVA				
Mittelwert = 4345,5625				
Standardabweichung (gesamt) = 774,5296				
Quadratsummen = 12577113 4471511 351320				
	zwischen Zielobjekten	Probenahme	Analytik	
Standardabweichung	556,2804	518,16089	148,18063	
prozentuale Varianz	51,583582	44,756204	3,7702174	
ERGEBNISSE DER ROBUSTEN ANOVA				
Mittelwert = 4408,3237				
Standardabweichung (gesamt) = 670,57617				
	zwischen Zielobjekten	Probenahme	Analytik	Messung
Standardabweichung	565,39868	319,04834	167,94308	360,5506
prozentuale Varianz	71,090791	22,636889	6,2723172	28,909209
relative Unsicherheit (% mit 95%iger Sicherheit)	-	14,474814	7,6193626	16,357719

Abb. A1.2: Ergebnisse der klassischen und robusten ANOVA für die Daten in Tabelle A1.1

Das Ergebnis der ANOVA für Daten aus einem ausgewogenen Versuchsplan ($n = 8$, Tabelle A1.1). Zum Vergleich werden sowohl robuste als auch klassische Schätzungen angegeben. Die Schätzungen der Standardabweichung werden für "Zwischen-Zielobjekten" ($s_{between-target}$), für "Innerhalb-Zielobjekt" (s_{samp}) und für „Innerhalb der chemischen Analyse“ (s_{anal}) berechnet. Die Ergebnisse werden in denselben Konzentrationseinheiten wie die Eingabedaten angegeben (d. h. in diesem Fall mg kg^{-1}).

Extraktion der robusten Schätzungen aus diesen Resultaten ergibt:

$$s_{samp} = 319,05 \text{ mg kg}^{-1}$$

$$s_{anal} = 167,94 \text{ mg kg}^{-1}$$

Gleichung 1 kann genutzt werden, um zu berechnen:

$$s_{meas} = \sqrt{(s_{samp}^2 + s_{anal}^2)} = 360,55 \text{ mg kg}^{-1}$$

Dies kann als Schätzung der Zufallskomponente der Standardunsicherheit (u) genutzt werden.

Die erweiterte relative Unsicherheit ist durch Gleichung 4 (Seite 25) gegeben als:

$$U_{meas}' = 2 * 100 * 360,55 / 4408 = 16,4 \% \text{ (des Konzentrationswertes)}$$

Für die Probenahme allein ist die erweiterte relative Unsicherheit (Zufallskomponente) entsprechend gegeben als:

$$U_{anal'} = 2 * 100 * 167,94 / 4408 = 7,6 \%$$

Dieser Wert liegt unter den normalen Grenzwerten der internen analytischen Qualitätskontrolle (z. B. 10 %).

Die analytischen Wiederfindungsschätzungen unterschieden sich statistisch nicht von der 100 %igen Wiederfindung (d.h. es wurde kein analytischer Bias festgestellt). Für dieses Beispiel wurde daher keine zusätzliche Unsicherheit im Zusammenhang mit analytischem Bias berücksichtigt.

7 Anmerkungen

Diese Unsicherheitsschätzung beinhaltet keine Schätzung des möglichen Probenahme-Bias.

8 Bewertung der Eignung dieser Messungen

Als Eignungskriterium wird zunächst der Prozentsatz der Gesamtvarianz (Abschnitt 16.2) genutzt. Bei der Verwendung der RANOVA berechnet das Programm, wie viel die Varianz Zwischen Zielobjekten, die Varianz innerhalb des Zielobjekts (oder der Probenahme) und die analytische Varianz (als Prozentsatz) zur Gesamtvarianz beitragen (Abbildung A1.2). Bei dieser Studie über Nitrat in Salat ist der maximale Beitrag zur Gesamtvarianz die Variabilität zwischen den Zielobjekten (71,1 %). Kombiniert man die Beiträge der Probenahme (22,6 %) und der Analyse (6,3 %) wird deutlich, dass der kombinierte Messprozess 28,9 % der Gesamtvarianz beiträgt. Dies ist geringfügig mehr als der Idealwert von 20 %. Von der Varianz der Messung ist die Probenahme der dominierende Faktor, der für 78,2 % der Varianz der Messung verantwortlich ist. Die Eignung kann auch mit der Methode der optimierten Unsicherheit (OU) bewertet werden. Bei dieser Methode wird die Eignung unter Berücksichtigung finanzieller Aspekte bewertet (Abschnitt 16.3) [54]. In diesem Fall kann gezeigt werden, dass eine Erhöhung von einer 10-Kopf- auf eine 40-Kopf-Mischprobe erforderlich ist, um Eignung zu erreichen (Anhang E, und [55]).

9 Berichterstattung und Interpretation

Für jede Salat-Partie (Probenahmezielobjekt) wird die Nitratkonzentration der 10-Kopf-Mischprobe mit dem Schwellenwert (4500 mg kg^{-1}) verglichen. Jede Nitratkonzentration sollte mit der Messunsicherheit (16,4 % des Messwerts) angegeben werden (Tabelle A1.2). Die Klärung, ob die jeweilige Charge einen Schwellenwert aufgrund der Messung und der damit verbundenen Unsicherheit überschreitet, hängt vom Wortlaut der entsprechenden Verordnung ab [42].

10 Zusammenfassung

Messunsicherheit		
Probenahme	Analyse	Gesamt
14,5 %	7,6 %	16, %

Tabelle A1.2 Ermittelte Nitrat-Konzentrationen von jeder Charge mit ihrer eigenen Messunsicherheit

Probenahme- zielobjekt	S1A1	Erweiterte Unsicherheit
A	3898	639,3
B	3910	641,2
C	5708	936,1
D	5028	824,6
E	4640	761,0
F	5182	849,8
G	3028	496,6
H	3966	650,4

Die Nitratkonzentrationen von S1A1 (Routineprobe) sind mit der zugehörigen Messunsicherheit (berechnet aus $U = 16,4\%$) dargestellt. Als Beispiel: Zielobjekt F hat einen Wert der Messgröße (oder wahren Wert) zwischen 4332 mg kg^{-1} und 6032 mg kg^{-1} .

Beispiel A2: Blei in kontaminiertem Oberboden

Messgröße				Unsicherheitsschätzung		
Analyt/ Technik	Einheit	Sektor/ Matrix	Probenahme- zielobjekt(e)	Zweck	Design	Statistik
Blei, gesamt / ICP-OES	mg kg ⁻¹ , bezogen auf Trocken- masse	Umwelt / Ober- boden	100 Zielbereiche – jeweils in einer Größe von 30 m x 30 m und Tiefe von 0 - 150 mm	Unsicherheit – gesamte Messung, Probenahme und Analytik Unsicher- heitsfaktor	Empirisch – Duplikat- methode	Robuste ANOVA mit Konzentrations- werten & klassische ANOVA nach Log- Transformation

1 Anwendungsbereich

Schätzung der Messunsicherheit und der Beiträge von Probenahme und Analyse an jedem von 100 verschiedenen Probenahmezielobjekten an einem Standort unter Verwendung einer gebräuchlichen Probenahmervorschrift.

2 Szenario und Probenahmezielobjekt

Es wurde eine Untersuchung eines 9 Hektar großen Geländes im Rahmen der Bewertung einer Fläche für potenzielle Wohnbebauung durchgeführt [56]. Als wichtigstes Analytelement für die Bewertung des Gesundheitsrisikos für Menschen wurde Blei ermittelt. Um die Bleikonzentration im Boden mit dem national geltenden Grenzwert (450 mg kg⁻¹) zu vergleichen, war eine Schätzung der Bleikonzentration und der Messunsicherheit für jedes der 100 Probenahmezielobjekte erforderlich.

3 Probenahmervorschrift

100 Proben des Oberbodens (nominale Tiefe 0 - 150 mm) wurden mit einem Handbohrer (Durchmesser 25 mm) an 100 Stellen entnommen. Diese Stellen wurden auf einem regelmäßigen Raster mit einem Stichprobenabstand von 30 m verteilt (Tabelle A2.1), so dass jede Stelle eine Fläche von 30 m x 30 m repräsentiert. Die Vermessung erfolgte mit Maßband und Kompass.

4 Studiendesign - Duplikatmethode (Abschnitt 9.4.2)

Zehn der Proben (d. h. 10 % der Gesamtzahl) wurden an zufällig ausgewählten Stellen zweifach entnommen unter Verwendung des ausgewogenen Designs (Abbildung 2). Die Doppelproben wurden in einem Abstand von 3 m zur ursprünglichen Probe in einer zufälligen Richtung entnommen. Damit sollen Unklarheiten der Probenahmervorschrift, die Unsicherheit bei der Lokalisierung des Probenahmezielobjekts (z. B. der Vermessungsfehler) und auch die Auswirkungen kleinräumiger Heterogenität auf die gemessene Konzentration innerhalb des festgelegten Zielobjektes berücksichtigt werden. Sechs zertifizierte Bodenreferenzmaterialien (ZRM) wurden für die Analyse ausgewählt, um den analytischen Bias über einen Konzentrationsbereich abzuschätzen.

5 Probenahme und Analyse im Labor

Die Primärproben wurden über Nacht bei 60 °C im Ofen getrocknet, zerteilt und gesiebt, um Partikel mit einer natürlichen Partikelgröße von mehr als 2 mm (gemäß der Definition von Boden) zu entfernen. Die gesiebten Proben (<2 mm) wurden gemahlen (95 % < 100µm) und gemischt. Prüfmengen von 0,25 g wurden zum Lösen mit Salpeter- und Perchlorsäure im Vorfeld der Bleibestimmung mittels ICP-OES entnommen [57]. Die Messungen wurden einer vollständigen analytischen Qualitätskontrolle (AQC) unterzogen und um die Konzentration der Blindprobe korrigiert, wenn sich diese Werte statistisch von Null unterschieden. Die ursprünglichen Messwerte, die für die Schätzung der Unsicherheit verwendet wurden, wiesen keine Rundung oder Unterdrückung von Werten kleiner Null oder der Nachweisgrenze auf.

6 Ergebnisse

6.1 Anfangsschätzung der relativen Unsicherheit

Die besten Schätzwerte für die Bleikonzentration an jeder der 100 Zielstellen sind in Form einer Karte dargestellt (Tabelle A2.1).

Tabelle A2.1: Gemessene Bleikonzentrationen an jedem Zielobjekt des Probenahmerasters (mg kg⁻¹), dargestellt anhand der tatsächlichen Koordinaten, die im regulären Probenahmeraster (Abstand 30 m) verwendet wurden [56]. Sie zeigen einen hohen Grad an Variabilität zwischen den Stellen von etwa einem Faktor 10. Die Variabilität innerhalb von 10 zufällig ausgewählten Stellen (d. h. A4, B7, C1, D9, E8, F7, G7, H5, I9 und J5) wurde für die Schätzung der Unsicherheit der Probenahme verwendet (Tabelle A2.2). Diese Variation innerhalb des Zielobjektes ist erheblich (z. B. ein Faktor von 2), aber wesentlich geringer als die Variabilität zwischen den Zielobjekten.

Reihe	A	B	C	D	E	F	G	H	I	J
1	474	287	250	338	212	458	713	125	77	168
2	378	3590	260	152	197	711	165	69	206	126
3	327	197	240	159	327	264	105	137	131	102
4	787	207	197	87	254	1840	78	102	71	107
5	395	165	188	344	314	302	284	89	87	83
6	453	371	155	462	258	245	237	173	152	83
7	72	470	194	82.5	162	441	199	326	290	164
8	71	101	108	521	218	327	540	132	258	246
9	72	188	104	463	482	228	135	285	181	146
10	89	366	495	779	60	206	56	135	137	149

Vier Messungen im ausgewogenen Design für jedes der 10 für die Duplizierung ausgewählten Probenahmezielobjekte wurden für die Schätzung der Unsicherheit verwendet (Tabelle A2.2). Die visuelle Betrachtung der Daten ermöglicht eine erste qualitative Bewertung der relativen Bedeutung der beiden Quellen der Messunsicherheit. Das geringe Maß an Übereinstimmung zwischen den Konzentrationswerten von einigen Duplikatproben deutet auf ein hohes Maß an Probenahmeunsicherheit hin (z. B. S1 im Vergleich zu S2 für das Zielobjekt "D9"). Die

Übereinstimmung zwischen den analytischen Duplikaten (A1 und A2) ist jedoch im Allgemeinen für die meisten Proben viel besser (< 10 % Unterschied) als die zwischen den Probenduplikaten.

Tabelle A2.2: Messungen der Konzentration (mg kg^{-1}) von Blei an 10 duplizierten Proben von insgesamt 100 Zielobjekten in einer Untersuchung von kontaminierten Flächen (Tabelle A2.1) [56]. Die Doppelproben sind mit S1 und S2 gekennzeichnet. Ebenso sind die Doppelanalysen mit den Bezeichnungen A1 und A2 gekennzeichnet. D9S1A2 (Wert 702 mg kg^{-1}) ist also die Analyse 2 von Probe 1 des Probenahmezielobjektes D9. Die angegebenen Werte sind der Übersichtlichkeit halber gerundet und werden für nachfolgende Berechnungen verwendet, aber im Allgemeinen sind ungerundete Werte für diese Berechnungen vorzuziehen.

Probenahme- zielort	S1A1	S1A2	S2A1	S2A2
A4	787	769	811	780
B7	338	327	651	563
C1	289	297	211	204
D9	662	702	238	246
E8	229	215	208	218
F7	346	374	525	520
G7	324	321	77	73
H5	56	61	116	120
I9	189	189	176	168
J5	61	61	91	119

Die Sichtprüfung eines Histogramms der gemessenen Werte der Bleikonzentration am gesamten Standort (Abb. A2.1a) zeigt, dass die Häufigkeitsverteilung rechtsschief ist. Diese schiefe Verteilung lässt sich durch die Bildung der natürlichen Logarithmen dieser Messwerte weitgehend beseitigen, was zu einer annähernd normalen Verteilung führt (Abb. A2.1b).

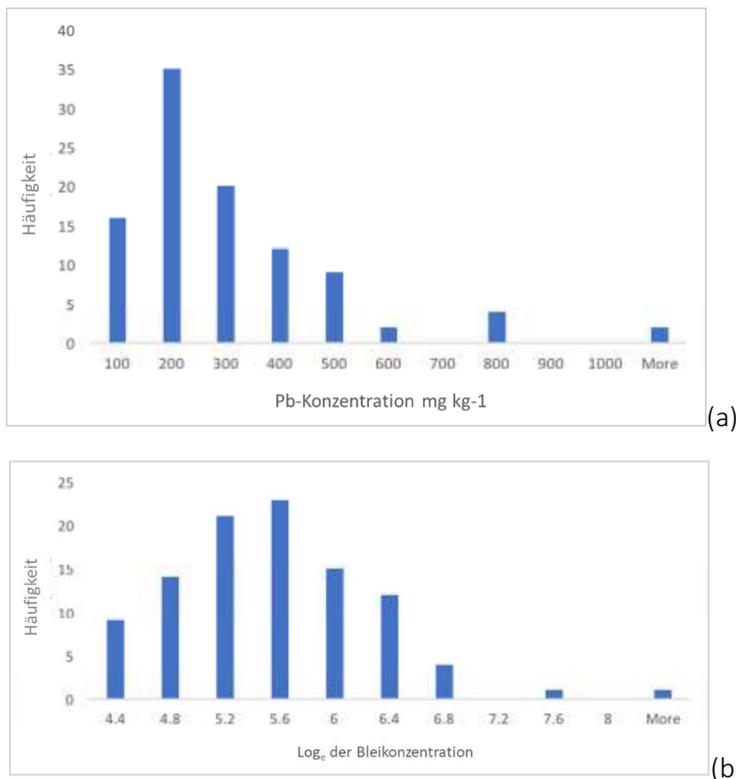


Abb. A2.1 Histogramme der in 100 Bodenproben gemessenen Pb-Konzentration (in mg kg⁻¹) dargestellt auf (a) der ursprünglichen linearen Skala (b) nach Bildung der natürlichen Logarithmen.

Unter Verwendung der ursprünglichen, nicht transformierten Messwerte (Tabelle A2.2) können die Zufallskomponente der Messunsicherheit und zwei ihrer Hauptkomponenten (Probenahme und Analyse) mittels robuster Varianzanalyse quantifiziert werden (z. B. mit dem Programm RANOVA2 [53]). Die robuste Statistik geht von einer zugrunde liegenden Normalverteilung aus, berücksichtigt aber die Ausreißerwerte, die in diesen Daten offensichtlich sind (z. B. Zielobjekt A4, Probenduplikat D9S1/S2, analytisches Duplikat B7S2A1/A2), und in den meisten ähnlichen Datensätzen [33] (siehe jedoch die Anmerkung in Beispiel A1, Abschnitt 6). Die Schätzungen der Unsicherheit sind über die 10 Zielobjekte gemittelt, unter der Annahme, dass die Unsicherheit in diesem Konzentrationsbereich nicht signifikant variiert. Die Unsicherheit wird in relativen Werten ausgedrückt, so dass sie über diesen Konzentrationsbereich anwendbar ist (Abschnitt 14.3).

Die Extraktion der robusten Schätzungen aus diesen Resultaten (Abb. A2.2) ergibt:

$$s_{\text{sampling}} = 123,8 \text{ mg kg}^{-1}$$

$$s_{\text{analytical}} = 11,1 \text{ mg kg}^{-1}$$

Gleichung 1 kann zur Berechnung verwendet werden:

$$s_{\text{meas}} = \sqrt{(s_{\text{sampling}})^2 + (s_{\text{analytical}})^2} = 124,3 \text{ mg kg}^{-1}$$

Dies kann als Schätzwert für die Zufallskomponente der Standardunsicherheit (u) verwendet werden.

Die erweiterte relative Unsicherheit ergibt sich aus Gleichung 4 (Seite 25) mit einem Erweiterungsfaktor von 2 wie folgt:

$$U_{meas}' = 2 * 100 * 124,3 / 297,3 = 83,63 \% \text{ (des Konzentrationswertes)}$$

Für die Probenahme allein ist die erweiterte relative Unsicherheit (Zufallskomponente) in ähnlicher Weise gegeben durch:

$$U_{sampling}' = 2 * 100 * 123,8 / 297,3 = 83,29 \%$$

Klassische ANOVA

Mittelwert	317,8	Anzahl der Zielobjekte 10		
Gesamt-Standardabweichung	240,19			
	<u>v. Zielobjekten</u>	<u>Probenahme</u>	<u>Analyse</u>	<u>Messung</u>
Standardabweichung	197,55	135,43	17,99	136,62
% der Gesamtvarianz	67,65	31,79	0,56	32,35
erweiterte relative Messunsicherheit (95 %)		85,23	11,32	85,98
Unsicherheitsfaktor (95 %)		2,6032	1,12	2,6207

Robuste ANOVA

Mittelwert	297,31			
Gesamt-Standardabweichung	218,49			
	<u>zw. Zielobjekten</u>	<u>Probenahme</u>	<u>Analyse</u>	<u>Messung</u>
Standardabweichung	179,67	123,81	11,144	124,31
% der Gesamtvarianz	67,63	32,11	0,26	32,37
erweiterte relative Messunsicherheit (95 %)		83,29	7,50	83,63

Abbildung A2.2: Das Ergebnis von RANOVA2 für Daten aus einem ausgewogenen Versuchsplan auf die ursprünglichen Messwerte angewandt ($n = 10$, Tabelle A2.2). Sie zeigt die klassischen und robusten Schätzungen des Mittelwerts und der Varianzen der Komponenten, wie in dem statistischen Modell (Gleichungen 1 & 2) gezeigt, ausgedrückt als Standardabweichungen. Das Resultat zeigt auch den erweiterten Unsicherheitsfaktor, der aus den logarithmisch transformierten Werten berechnet wurde (in der 8. Zeile) und der in Abschnitt 6.2 weiter unten behandelt wird.

Zum Vergleich werden sowohl robuste als auch klassische Schätzungen angegeben. Die Schätzungen der Standardabweichung werden berechnet für „zwischen Zielobjekten“ ($s_{btn-target}$), „innerhalb der Zielobjekte“ ($s_{sampling}$) und innerhalb der chemischen Analyse ($s_{analytical}$). Die Ergebnisse haben die gleichen Konzentrationseinheiten wie die Eingabedaten (d. h. in diesem Fall $mg\ kg^{-1}$), mit Ausnahme des prozentualen Anteils der Gesamtvarianz, der erweiterten relativen Unsicherheit und des erweiterten Unsicherheitsfaktors (FU).

Zum Vergleich: Die erweiterte Unsicherheit für den analytischen Beitrag (Zufallskomponente) ist gegeben durch

$$U_{analytical}' = 2 * 100 * 11,1 / 297,3 = 7,5 \%$$

Dieser Wert liegt unter den typischen Grenzwerten der internen analytischen Qualitätskontrolle (z. B. 10 %).

6.2 Log-transformierte Werte und der Unsicherheitsfaktor

Die zweite Möglichkeit zur Schätzung der Unsicherheit besteht in der Log-Transformation der Messwerte (Tabelle A2.2), in der Regel unter Verwendung des natürlichen Logarithmus (abgekürzt \ln oder \log_e), unter der Annahme, dass sie lognormal verteilt sind. Das Histogramm (Abb. A2.1) zeigt, dass die Variabilität zwischen den Zielobjekten annähernd lognormal verteilt ist, jedoch muss idealerweise die Häufigkeitsverteilung der Stichprobenvariabilität innerhalb jedes Zielobjekts bewertet werden. Bei nur zehn Probanden in diesem Beispiel ist es nicht möglich, diese Stichprobenverteilung vollständig zu charakterisieren. Das ungewöhnlich große Verhältnis der Durchschnittswerte der Probanden an zwei Zielobjekten (4,3 für G7 und 2,8 für D9) deutet jedoch auf eine Rechtsschiefe hin, die mit einer lognormalen Verteilung für die Stichprobenvariabilität verbunden ist, ähnlich wie bei der Variabilität zwischen den Zielobjekten (Abb. A2.1). Darüber hinaus werden beide Arten der Variabilität weitgehend durch die Heterogenität des Analyten gesteuert, die nachweislich oft eine ähnliche Form der Häufigkeitsverteilung über große räumliche Entfernungen zeigt [58].

Die Werte, die sich aus der Anwendung der klassischen ANOVA auf die logarithmisch transformierten Werte ergeben, sind in Tabelle A2.3 aufgeführt. Der Mittelwert im log-Bereich (5,478) ergibt das geometrische Mittel ($239,4 \text{ mg/kg} = e^{5,478}$). Die Messstandardabweichung der \log_e -transformierten Werte ($s_{G,meas} = 0,4817$) kann zur Berechnung des erweiterten Unsicherheitsfaktors ($2,6207 = e^{2*0,4817}$) unter Verwendung von Gleichung 5 verwendet werden.

$$F_U = \exp(2s_G) \dots \dots \dots \text{Gleichung 5 (wiederholt aus Abschnitt 9.5.3)}$$

Diese Berechnung kann auch automatisch von einem ANOVA-Programm durchgeführt werden, das die ursprünglichen Messwerte, die die Dateneingabe darstellen, log-transformiert (Resultat in Abb. A2.2, Zeile 8) [53].

Tabelle A2.3. Klassisches ANOVA-Resultat bei Anwendung auf die natürlichen Logarithmen jedes gemessenen Konzentrationswertes in Tabelle A2.2.

Durchschnitt	5,478			
Gesamt-Standardabweichung	0,82337			
	<u>zw. Zielobjekten</u>	<u>Probenahme</u>	<u>Analyse</u>	<u>Messung</u>
Standardabweichung	0,66775	0,4784	0,0567	0,4817
% der Gesamtvarianz	65,77	33,76	0,47	34,23

Für dieses Beispiel beträgt der Wert von $^F U_{\text{meas}}$ 2,62. Dieser Unsicherheitsfaktor wird offensichtlich vom Beitrag der Probenahme dominiert, $^F U_{\text{sampling}}$, der 2,60 beträgt. Diese Komponente wird direkt im ANOVA-Resultat (Abb. 2.2, Zeile 8) dargestellt. Sie kann auch direkt als 2,6034 ($= e^{2 \cdot 0,4784}$) berechnet werden, indem man den Wert von $s_{G, \text{sampling}}$ ($= 0,4784$, Tabelle A2.3) in Gleichung 5 einsetzt.

Tabelle A2.4: Konfidenzgrenzen (untere LCL und obere UCL) und Konfidenzintervall (CI), berechnet für einen nominalen Konzentrationsmesswert von 300 mg kg⁻¹, wenn die Messunsicherheit entweder als erweiterte relative Unsicherheit (U') oder als erweiterter Unsicherheitsfaktor ($^F U$) ausgedrückt wird. Die viel höhere UCL für den $^F U$ -Ansatz spiegelt die Rechtsschiefe der zugrundeliegenden Häufigkeitsverteilung besser wider (Abb. A2.1)

	Messwert	untere Konfidenzgrenze	Ermittlung der unteren Konfidenzgrenze	obere Konfidenzgrenze	Ermittlung der oberen Konfidenzgrenze	Konfidenzintervall
Einheit		mg kg ⁻¹		mg kg ⁻¹		mg kg ⁻¹
U'	83,6 %	49	300 - 83,6 %	551	300 + 83,6 %	± 251
$^F U$	2,62	115	300/2,62	784	300 x 2,62	-185, + 484

Eine sinnvolle Möglichkeit, diese beiden Ansätze zur Berechnung und Angabe der Unsicherheit zu vergleichen, ist der Vergleich der sich ergebenden 95 %-Konfidenzgrenzen (Tabelle A2.4). Für einen typischen gemessenen Bleikonzentrationswert von 300 mg kg⁻¹ ergibt die erweiterte Unsicherheit von 83,6 % eine untere Konfidenzgrenze (LCL) von 49 mg kg⁻¹ (d.h. 300 - 83,6 %) und eine obere Konfidenzgrenze (UCL) von 551 mg kg⁻¹ (d.h. 300 + 83,6 %). Dies ergibt ein symmetrisches Konfidenzintervall von 251 mg kg⁻¹. Bei Verwendung des Unsicherheitsfaktor-Ansatzes beträgt die LCL jedoch 115 mg kg⁻¹ (d. h. 300/2,62), und die UCL ist 784 mg kg⁻¹ (d.h. 300x2,62). Dieses Konfidenzintervall liegt deutlich asymmetrisch um den gemessenen Wert mit -185 mg kg⁻¹ bzw. + 484 mg kg⁻¹ vom gemessenen Wert. Diese Rechtsschiefe spiegelt die Annahmen über die Verteilung der ursprünglichen Messergebnisse wider (Abb. A2.1a). Der Ansatz des Unsicherheitsfaktors schließt die wenigen sehr hohen Werte viel stärker ein, wobei die UCL bis zu 784 und nicht nur bis zu 551 mg kg⁻¹ reicht, und ist daher repräsentativer für die Streuung der ursprünglichen Messungen. Außerdem deutet die Häufigkeitsverteilung der ursprünglichen Messwerte darauf hin, dass es mehr als 10 % abweichende Werte gibt, was die Grenze für eine zuverlässige Schätzung der Unsicherheit mit dieser speziellen robusten ANOVA darstellt [53]. Diese beiden Faktoren deuten darauf hin, dass der Unsicherheitsfaktor-Ansatz in diesem Fall zuverlässiger ist.

6.3 Berechnung der relativen Unsicherheit aus dem Unsicherheitsfaktor

Obwohl der Unsicherheitsfaktor manchmal der zuverlässigste Weg ist, um die Messunsicherheit auszudrücken, kann es auch nützlich sein, sie in Form der bekannteren relativen Unsicherheit auszudrücken. Die relative Unsicherheit (u') wird normalerweise prozentual (Gleichung 4, Seite 20) angegeben, kann aber auch als Bruchteil der Einheit ausgedrückt werden. Wie in Abschnitt 9.5.3 beschrieben, kann der Bruchteil von u' aus der Standardabweichung der natürlichen Logarithmen der Messwerte ($s(\ln(x)) = s_G$) unter Verwendung der Beziehung

$$u' = \sqrt{\exp(s_G^2) - 1}$$

berechnet werden:

In diesem Beispiel, in dem $s_{G,meas}$ 0,4817 beträgt, ergibt diese Gleichung u' als 0,5111 (d. h. 51 %). Da u' größer ist als der festgelegte Richtwert von 20 % (d. h. $u' > 0,2$), kann er nicht mit 2 multipliziert werden, um eine erweiterte Unsicherheit zu erhalten.

6.4 Einbeziehung des analytischen Bias

Der analytische Bias wurde auf -3,41 % ($\pm 1,34$ %) geschätzt, wobei eine lineare funktionale Beziehung [59] zwischen den gemessenen Werten und den zertifizierten Werten der sechs ZRM hergestellt wurde (Tabelle A2.5).

Derzeit gibt es keinen Konsens darüber, wie zufällige und systematische Effekte am besten in einer Schätzung der Unsicherheit kombiniert werden können, obwohl vier Optionen identifiziert wurden [22]. Eine Option [37] besteht darin, den geschätzten analytischen Bias (z. B. -3,41 %) als typischen Wert für Teilnehmer an einer organisationsübergreifenden Studie zu betrachten. Wenn dieser Bias und seine eigene Unsicherheit (1,34 %) dann zur Zufallskomponente der Unsicherheit (unter Verwendung der Summe der Quadrate) addiert werden, erhöht sich die Varianz auf den Wert, der in einem solchen Versuch gefunden werden würde. Die Logik dieses Ansatzes besteht darin, dass die zusätzliche Unsicherheit, die in der Regel bei Versuchen zwischen Organisationen festgestellt wird, auf den unerwarteten Bias innerhalb jeder Organisation zurückzuführen ist. Wenn eine Schätzung der zusätzlichen Varianz, die durch diesen Bias zwischen verschiedenen Laboratorien verursacht wird, vorgenommen werden kann, kann diese zu der Zufallskomponente innerhalb einer Organisation addiert werden. In diesem Fall wird die relative Standard-Analyseunsicherheit auf 5,24 % erhöht [$= (3,75^2 + 3,41^2 + 1,34^2)^{0,5}$]. Die erweiterte analytische Unsicherheit (10,48 %) ist dann größer als der analytische Zielwert von 10 %, kann aber auch sinnvoll mit einer unabhängigen Schätzung der analytischen Messunsicherheit innerhalb des Labors verglichen werden. Die erweiterte Unsicherheit für die gesamte Messung erhöht sich dadurch auf 83,95 % [$= (83,29^2 + 10,48^2)^{0,5}$], was praktisch identisch mit der rein zufälligen Komponente von 83,63 % ist. Bei Verwendung des Unsicherheitsfaktoransatzes wäre der Beitrag des analytischen Bias in diesem Beispiel ebenfalls vernachlässigbar. In Fällen, in denen er nicht vernachlässigbar ist, könnte er mit Hilfe veröffentlichter Methoden [21] zu FU addiert werden.

Tabelle A2.5: Gemessene und zertifizierte Bleikonzentrationswerte von ZRMs für die Schätzung des Bias der Analysemethode [56]

ZRM Bezeichnung (n=4)	Mittelwert (mg kg ⁻¹)	Standardabweichung (mg kg ⁻¹)	Zertifizierter Wert (mg kg ⁻¹)	U des zertifizierten Werts (95 % Vertrauen)
NIST2709	19,7	3,2	18,9	0,5
NIST2710	5352,0	138,0	5532,0	80,0
NIST2711	1121,4	14,7	1162,0	31,0
BCR141	34,4	3,9	29,4	2,6
BCR142	36,2	4,6	37,8	1,9
BCR143	1297,5	33,0	1333,0	39,0

7 Anmerkungen

Diese Schätzung der Unsicherheit berücksichtigt keinen unentdeckten Bias bei der Probenahme (Abschnitt 9.4.2). Da die Unsicherheit jedoch häufig von der Heterogenität des Probenahmezielobjekts dominiert wird, kann die zusätzliche Unsicherheit, die durch einen Bias bei der Probenahme entsteht, im Vergleich dazu oft als unbedeutend angesehen werden (wie für den analytischen Bias gezeigt). Wenn höchste Qualität der Unsicherheitschätzung erforderlich ist, vielleicht wegen potenziell großer finanzieller Folgen einer Unterschätzung der Unsicherheit, kann es sinnvoll sein, eine der aufwändigeren Methoden mit mehreren Probenehmern und/oder Vorschriften anzuwenden (Tabelle 5).

Wäre die Messgröße (oder der wahre Wert) als die mittlere Bleikonzentration am gesamten Standort definiert worden, hätte die Unsicherheit den Beitrag des Standardfehlers zum berechneten Mittelwert enthalten müssen, ausgedrückt als s_{gesamt}/\sqrt{n} . In diesem Beispiel beträgt s_{gesamt} 403 mg kg⁻¹, $n = 100$ (Tabelle A2.1), und die Unsicherheit des Mittelwerts (291,9 mg kg⁻¹) beträgt daher 27,6 % des Werts bei 95 % Vertrauen. Dieser Wert kann berechnet werden, ohne dass der individuelle Beitrag der Unsicherheit durch die Probenahme oder die Analyse bekannt ist, und wird oft durch $s_{between-sample}$ dominiert. Wenn man alternativ eine lognormale Verteilung aller 100 Messwerte annimmt (Tabelle A2.1), beträgt das geometrische Mittel 208 mg kg⁻¹ und $s_{G,gesamt}$ für diese Werte 0,733. Der Unsicherheitsfaktor FU für das geometrische Mittel beträgt 1,158 (d. h. $\exp[2s_G/\sqrt{n}]$ unter Verwendung von Gleichung 5), was ein kleineres, aber asymmetrisches Konfidenzintervall von -13,7 % bis +15,8 % des geometrischen Mittels ergibt.

8 Bewertung der Zwecktauglichkeit dieser Messungen

Unter Verwendung der Methode "Prozentsatz der Gesamtvarianz" (Abschnitt 16.2) gibt das Ergebnis in Abbildung A2.1 den Prozentsatz der Gesamtvarianz ($[\text{Standardabweichung (gesamt)}]^2$) an, der auf "zwischen Zielobjekten", die Probenahme (innerhalb der Zielobjekte) und die Analyse (innerhalb der Stichprobe) entfällt. In diesem speziellen Beispiel dominiert eindeutig die "zwischen-Zielobjekten"-Varianz (67,6 % der Gesamtvarianz), obwohl dies unter dem idealen Schwellenwert von 80 % liegt (Abschnitt 16.2). Darüber hinaus dominiert die Probenahme (32,11 % der Gesamtvarianz) gegenüber der chemischen Analyse (0,26 % der

Gesamtvarianz) als Beitrag zur Messvarianz. Die Stichprobenvarianz (d. h. innerhalb des Zielobjekts) wird in diesem Fall als Hauptfaktor (99,2 %) für die Unsicherheit des Messverfahrens identifiziert (d. h. $100 * 32,11 / [32,11 + 0,26]$).

Die Bewertung der Zwecktauglichkeit von Messungen bei der Untersuchung kontaminierter Böden unter Verwendung der Methode der optimierten Unsicherheit (Abschnitt 16.3) wird an anderer Stelle beschrieben [60, 61].

9 Berichterstattung und Interpretation

Einzelne Messungen der Bleikonzentration, die für diese Zielobjekte berichtet werden, sollten mit einem $^F U$ -Wert von 2,62 oder einem U' -Wert, der 83,9 % des Konzentrationswertes entspricht angegeben werden. Dies gilt für alle die Messwerte (Tabelle A2.1), die mindestens 10-mal höher sind als die analytische Nachweisgrenze (die in diesem Fall auf 2 mg kg^{-1} geschätzt wird). In Anwendungen, bei denen dies nicht der Fall ist, ist es notwendig, die Unsicherheit als Funktion der Konzentration auszudrücken [10]. Darüber hinaus ergibt sich für die Unsicherheit der Mittelwertmessungen an den 10 Zielobjekten, an denen Doppelproben genommen wurden (z. B. die in Tabelle A2.2 aufgeführten), einen reduzierter $^F U$ von 1,98 (d. h. $\exp(2 * s_G / \sqrt{2})$ unter Verwendung von Gleichung 5) oder eine U' -Schätzung von 59,3 % ($= 83,9 / \sqrt{2}$).

Wenn man den Wert der Unsicherheit kennt, ist es auch möglich, eine probabilistische Interpretation des Ausmaßes der Bleikontamination am Standort vorzunehmen [24].

10 Zusammenfassung

	Messunsicherheit*		
	Probenahme	Analytisch	Gesamt
$^F U$	2,60	1,12	2,62
U'	83,3 %	10,5 %	83,9 %

* mit einem Erweiterungsfaktor von 2 (d.h. 95 % Vertrauen)

Beispiel A3: Im Grundwasser gelöstes Eisen

Messgröße				Unsicherheitsschätzung		
Analyt/ Technik	Einheit	Sektor/ Matrix	Probenahme- zielobjekt(e)	Zweck	Design	Statistik
Gelöstes Eisen / ICP- OES	mg l ⁻¹	Umwelt / Grund- wasser	Das Grund- wasser in der Nähe einer ausgewählten Messstelle in einem Grundwasser- körper	Gesamt- unsicherheit	Empirische Duplikate für die Validierung und Qualitäts- kontrolle	Spannweiten- methode in absoluten Einheiten

1 Anwendungsbereich

Ziel ist die Bestimmung der Gesamtunsicherheit bei der Messung von gelöstem Eisen in einer Probenahmevalidierungsstudie und die anschließende Kontrolle der Probenahmeunsicherheit bei der Überwachung.

2 Szenario und Probenahmezielobjekt

Für einen Grundwasserkörper, der eine wichtige Trinkwasserressource für die Stadt Århus, die zweitgrößte Stadt Dänemarks, darstellt, wurde durch Kontrollüberwachung festgestellt, dass eine Gefährdung hinsichtlich der Verschlechterung der Qualität aufgrund intensiver Trinkwasserentnahme besteht. Es wurde ein betriebliches Überwachungsprogramm erstellt, um den Trend bei der Entwicklung der Wasserqualität zu kontrollieren.

Der Grundwasserkörper befindet sich in glazialen Auswaschungssand mit Sanden und Tonen aus dem Miozän darunter und glazialen Moränenmaterial darüber. Die Geologie am Standort ist kompliziert mit mehreren lokalen Grundwasserleitern (unterirdische Schicht aus wasserführendem, durchlässigem Gestein oder durchlässigen Gemischen aus losem Gestein) und Aquitarden (geologische Formation aus Schichten, die entweder aus Ton oder aus nicht porösem Gestein bestehen und den Wasserfluss von einem Grundwasserleiter zum anderen einschränken). Der ermittelte Grundwasserkörper ist 2 km x 2 km x 10 m groß und beginnt 20 - 30 m unter der Oberfläche. Die natürliche Qualität des Grundwassers ist anaerob ohne Nitrat, mit Sulfat und reduziertem Eisen, aber ohne Schwefelwasserstoff und Methan. Eine der Bedrohungen für den Grundwasserkörper ist das Eindringen von Sauerstoff in den Grundwasserleiter infolge der Wasserentnahme und der damit verbundenen Absenkung des Grundwasserspiegels.

Im Grundwasserkörper wurden im Rahmen der Kontrollüberwachung neun Messstellen für chemische Analysen beprobt, und sechs Messstellen stehen jetzt für Probenahmen zur Verfügung. Im Plan für die operative Überwachung wurde beschlossen, dass eine Messstelle zweimal pro Jahr überwacht werden soll. Als Ziel für die operative Überwachung wurde festgelegt, dass mit einer Wahrscheinlichkeit von 95 % eine Qualitätsverschlechterung von 20 % erkannt werden kann. Es wurde beschlossen, gelöstes Eisen als Zielparameter zu

verwenden, das ein empfindlicher Indikator für die Oxidation eines Grundwasserleiters ist (abnehmende Eisenkonzentration bei zunehmender Oxidation), wobei das Redoxpotential als unterstützender Nachweis dient. Sauerstoff, pH-Wert, elektrische Leitfähigkeit und Redoxpotential wurden als Online-Indikatoren für die Stabilität der Probenahme verwendet, Natrium, Calcium und Chlorid als allgemeine Parameter für die Grundwasserqualität. Hier werden nur die beiden Schlüsselparameter, gelöstes Eisen und Redoxpotential, behandelt.

Zur Erreichung des Überwachungsziels ist eine Messunsicherheit von höchstens 10 % erforderlich, die sowohl die Probenahme als auch die Analyse umfasst (Vergleich von zwei Mittelwerten für jeweils zwei Proben, 95 % Konfidenzintervall, zweiseitiger Test), was einer erweiterten Messunsicherheit von 20 % entspricht. Um die Übereinstimmung des Überwachungsprogramms mit dem genannten Ziel zu gewährleisten, wurde zunächst eine Studie zur Validierung der Probenahme durchgeführt, die alle verfügbaren Messstellen umfasste, und auf der Grundlage der Ergebnisse dieser Studie wurde ein Programm zur routinemäßigen Qualitätskontrolle der Probenahme erstellt, das zusammen mit dem Überwachungsprogramm für die ausgewählte Messstelle umgesetzt werden sollte.

Die Eigenschaften des Grundwasserkörpers wurden auf der Grundlage früherer Überwachungstätigkeiten (Kontrollüberwachung) zusammengefasst. Tabelle A3.1 zeigt eine Zusammenfassung für die beiden Schlüsselparameter, einschließlich der zeitlichen und räumlichen Variabilität sowie der Messunsicherheit (Probenahme und Analytik).

Tabelle A3.1: Chemische Schlüsselparameter für neun Messstellen des Grundwasserkörpers, Kontrollüberwachung

	Redoxpotential mV	Gelöstes Eisen mg l ⁻¹
Durchschnitt	-123	1,11
Relative Standardabweichung	27 %	56 %
Hauptursache für Unsicherheit	Auswirkung von Sauerstoff während der Probenahme und der On-Line-Messung	Filterung

Die chemischen Daten deuten darauf hin, dass die Zusammensetzung des Grundwassers im Hinblick auf die Hauptbestandteile zeitlich und räumlich recht einheitlich ist (Daten nicht gezeigt, relative Standardabweichung 1,9 – 16 %), während die Variabilität bei den Redoxparametern (Sauerstoff, Redoxpotential und gelöstes Eisen) hoch ist. Die erwarteten Hauptursachen für die Unsicherheit sind in der Tabelle für die beiden Schlüsselparameter angegeben, und die Ursachen wurden während der Probenahme kontrolliert.

3 Probenahmeverfahren

Die Probenahme erfolgte gemäß der Grundwasserüberwachungsvorschrift der Provinz Århus mit permanenten, speziellen Pumpen (Grundfos MP1), die in der Mitte des verfilterten Bereichs jeder Messstelle aufgestellt wurden. Die Pumpraten betragen 1-2 m³ h⁻¹ (Messstellenspülung) mit einer 10 %igen Reduzierung kurz vor der Probenahme. Bei zwei der sechs Messstelle handelte es sich um Entnahmekübel mit großem Durchmesser, die mit

leistungsstarken Pumpen ausgestattet waren. Diese wurden mit $40\text{-}60\text{ m}^3\text{ h}^{-1}$ für die Messstellenspülung gepumpt, gefolgt von einer Reduzierung der Pumprate kurz vor der Probenahme. Während der Messstellenspülung wurde die Entwicklung der Wasserqualität durch Online-Messungen von Sauerstoff, pH-Wert, elektrischer Leitfähigkeit und Redoxpotential verfolgt, bis die Werte stabil waren, und dann wurden Proben genommen. Während der Probenahme wurde ein Protokoll erstellt, in dem die Pumpleistung und die Pumpzeiten sowie die Messungen des Grundwasserspiegels festgehalten wurden.

4 Studiendesign - empirisch

Der empirische Ansatz wurde für das Studiendesign gewählt, um Schätzungen der Heterogenität im Grundwasserkörper (Variation zwischen den Zielobjekten von Messstelle zu Messstelle und über die Zeit) und der Messunsicherheit zu erhalten und um die Unsicherheit bei der Probenahme und die analytische Unsicherheit getrennt aufzuzeigen.

4.1 Validierung

Ziel des Validierungsprogramms war es, sicherzustellen, dass eine Messunsicherheit erreicht werden kann, die dem festgelegten Qualitätsziel entspricht, und die Komponenten der Unsicherheit zu beschreiben, um gegebenenfalls Verbesserungsmöglichkeiten zu ermitteln. Das Validierungsprogramm umfasste die Beprobung von sechs Messstellen, zwei unabhängige Probenahmen pro Messstelle und zwei Teilproben pro analysierter Probe (siehe Abbildung A3.1).

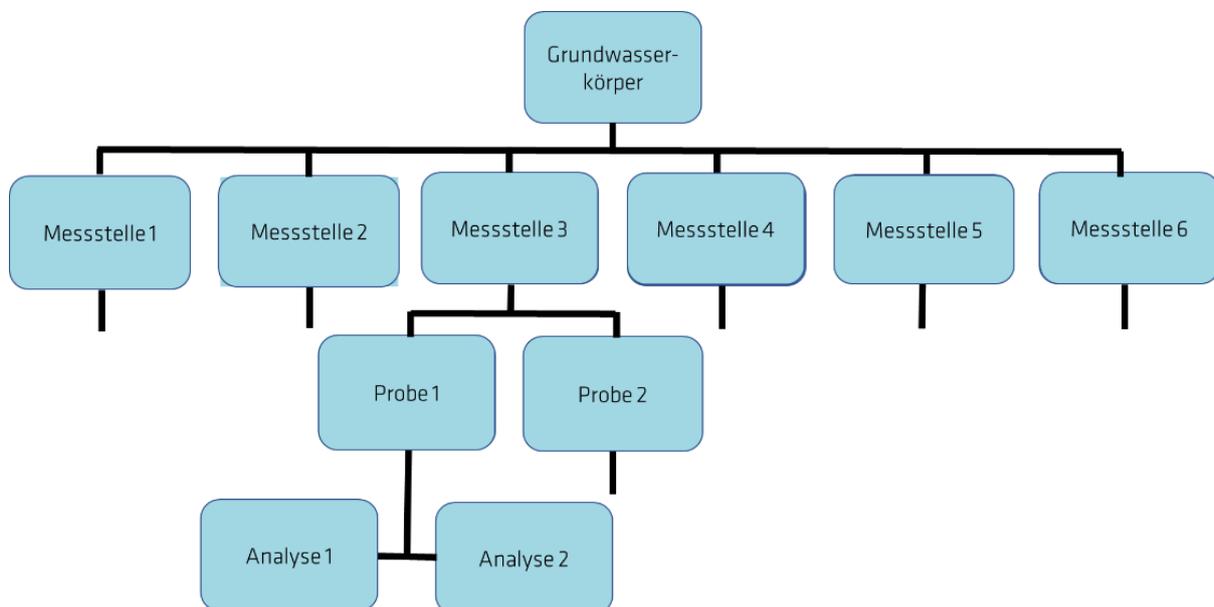


Abbildung A3.1 Entwurfsskizze für die Validierung

Insgesamt wurden in einer Beprobungsrunde 12 Proben entnommen und 24 Teilproben als Validierungsstudie zur Analyse gegeben.

4.2 Qualitätskontrolle

Ziel des Qualitätskontrollprogramms für die operative Überwachung war es, sicherzustellen, dass die Messunsicherheit während der Überwachung nicht zunimmt. Das

Qualitätskontrollprogramm wurde nach sorgfältiger Auswertung der Ergebnisse der Validierungsstudie aufgestellt. Die Qualitätskontrolle wurde so konzipiert, dass bei einer der beiden jährlichen Probenahmen Im Rahmen des Überwachungsprogramms jeweils eine Doppelbeprobung mit Doppelanalyse durchgeführt wurde (siehe Abbildung A3.2). Insgesamt wurden in der ersten Phase des Qualitätskontrollprogramms sechs Probenahmen mit 12 Proben und 24 analysierten Teilproben durchgeführt.

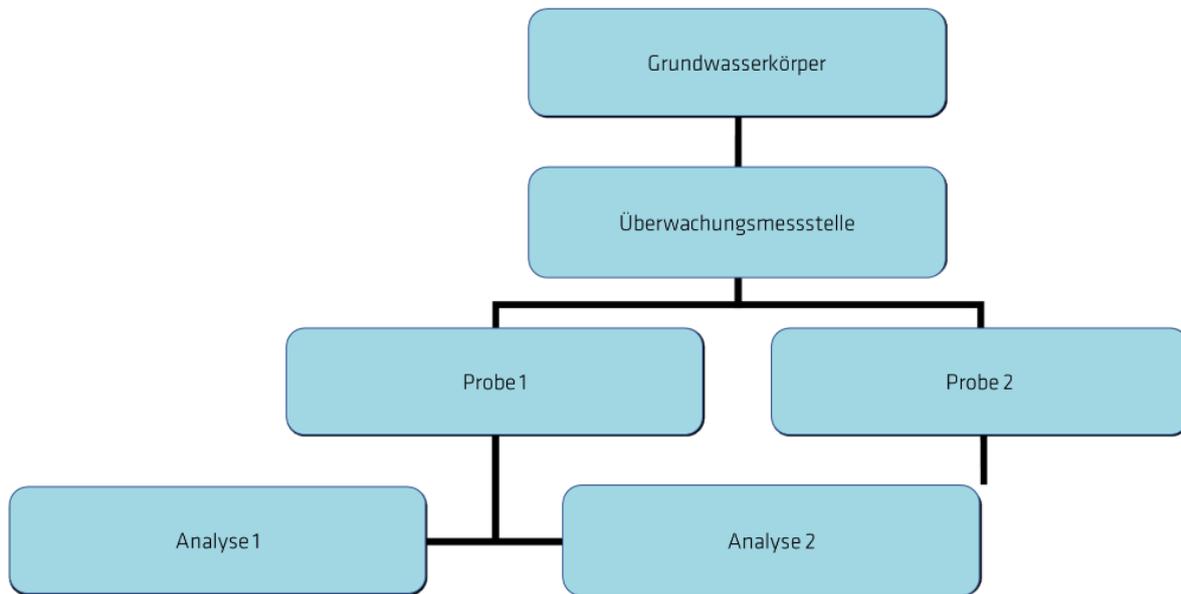


Abbildung A3.2 Schema für die Qualitätskontrolle, dargestellt für eine Probenahme

5 Teilprobenahme und Analyse

Die Probenvorbereitung und Analyseschema für die beiden Schlüsselparameter (Redoxpotential und gelöstes Eisen) sind in Tabelle A3.2 dargestellt.

Tabelle A3.2: Vorbereitungs- und Analyseprogramm

	Redoxpotential	Gelöstes Eisen
Vorbereitung	On-line-analysiert	On-line-gefiltert, mit Salpetersäure stabilisiert, im Labor analysiert

5.1 Teilprobenahme und Vorbereitung der Proben

Doppelte On-Line-Messungen/Teilprobenahmen für die Laboranalyse wurden durchgeführt, indem getrennte Probenströme entnommen und jeder Strom unabhängig behandelt wurde. Das bedeutet, dass die "analytische Unsicherheit", die mit dem Duplikatdesign ermittelt wurde, auch die Teilprobenahme, die Vorbereitung, wie z. B. Filtrierung, und den Transport umfasst. Eine Schätzung der analytischen Unsicherheit allein könnte aus den Qualitätskontrolldaten des Labors gewonnen werden (siehe Abschnitt 5.3).

Die Proben wurden on-line unter Ausschluss von Sauerstoff durch 0,45-µm-Zelluloseacetat-Membranfilter filtriert, und die Teilproben wurden vor Ort durch Ansäuern mit Salpetersäure

für die Metallanalyse stabilisiert. Die Teilproben wurden während des Transports zum Analyselabor in Polyethylenbehältern im Dunkeln bei weniger als 10 °C gelagert.

5.2 Laboranalyse

Die Analysen wurden in einem unabhängigen, akkreditierten (ISO/IEC 17025) Labor mit akkreditierten Methoden durchgeführt, die der erforderlichen Qualitätssicherung und analytischen Qualitätskontrolle unterliegen. Die Methoden und Kenndaten der Qualitätskontrolle sind in Tabelle A3.3 aufgeführt.

Tabelle A3.3: Methoden und Kenndaten der analytischen Qualitätskontrolle für Laboranalysen mit einem Gehalt von > 0,2 mg l⁻¹.

	Methode	Reproduzierbarkeit, s_r	Reproduzierbarkeit innerhalb des Labors, s_{RW}	Erweiterte Unsicherheit	Nachweisgrenze
Eisen	ICP-OES	0,95 %	4,3 %	8,6 %	0,01 mg l ⁻¹

Das zertifizierte Referenzmaterial (ZRM) VKI Metall LL2, nominal 0,200 mg l⁻¹ Fe, wurde zur Qualitätskontrolle mit einer scheinbaren Wiederfindung von 101,9 % (Mittelwert von 92 Kontrollergebnissen) verwendet.

5.3 Berechnungsmethoden

Die replizierten Daten wurden mit der Spannweitenmethode behandelt. Zum Vergleich wurden die Unsicherheits-schätzungen durch Varianzanalyse (ANOVA) und robuste ANOVA (RANOVA) mit ROBAN Version 1.0.1 (Anhang C3) berechnet.

Die angewandten Berechnungsmethoden werden in Abschnitt 7 (unten) erläutert. Die Spannweitenberechnungen lassen sich leicht mit einem Standard-Rechenprogramm durchführen.

Das Auftreten von systematischen Probenahme- und/oder Analysefehlern wurde nicht quantitativ bewertet, sondern die Konsistenz der erzielten Ergebnisse wurde als qualitative Kontrolle systematischer Fehler verwendet. Wurde beispielsweise in der gleichen Probe, in der der Sauerstoffgehalt höher 0,1 mg l⁻¹ bestimmt wurde, auch ein Gehalt an gelöstem Eisen von über 0,1 mg l⁻¹ festgestellt, so würde dies auf einen systematischen Probenahme- und/oder Vorbehandlungsfehler hindeuten. In ähnlicher Weise wurde überprüft, ob Redoxpotential und Sauerstoffgehalt übereinstimmen, um systematische Fehler zu vermeiden.

6 Ergebnisse

Die Daten der Validierungsstudie (sechs verschiedene Messstellen) unter Verwendung von Spannweitenberechnungen sind in Tabelle A3.4 und die detaillierten Berechnungen in Tabelle A3.7 dargestellt.

Tabelle A3.4: Relative erweiterte Unsicherheit (% , Erweiterungsfaktor 2) für die Analyse, die Probenahme und zwischen den Zielobjekten (zwischen den Messstellen), wie bei der Validierung unter Verwendung von Spannweitenberechnungen ermittelt

Spannweitenberechnung	Analysen	Probenahme	Zwischen den Zielobjekten
Gelöstes Eisen	1,8 %	10,5 %	70 %

Zum Vergleich sind in Tabelle A3.5 die statistischen Schätzungen aufgeführt, die mittels ANOVA und RANOVA ermittelt wurden.

Tabelle A3.5: Relative erweiterte Unsicherheit (% , Erweiterungsfaktor 2) für die Analyse, Probenahme und zwischen den Zielobjekten (zwischen den Messstellen), wie sie für gelöstes Eisen während der Validierung mit ANOVA- und RANOVA-Berechnungen ermittelt wurde

Gelöstes Eisen	Analysen	Probenahme	Zwischen den Zielobjekten
ANOVA	1,6 %	9,6 %	70 %
RANOVA	1,8 %	9,9 %	72 %

Die statistischen Schätzungen, die mit der Spannweitenstatistik während der Qualitätskontrolle (sechs Probenahmen) ermittelt wurden, sind in Tabelle A3.6 dargestellt.

Tabelle A3.6: Relative erweiterte Unsicherheit (% , Erweiterungsfaktor 2) für die Analyse, Probenahme und zwischen den Zielobjekten (zwischen den Messstellen), wie bei der Qualitätskontrolle unter Verwendung von Spannweitenberechnungen ermittelt

	Analysen	Probenahme	Zwischen den Zielobjekten ¹
Gelöstes Eisen	2,5 %	3,6 %	9,9 %

¹Die Variabilität zwischen den Zieleobjekten war bei der Qualitätskontrolle die Variabilität zwischen den Probenahmezeitpunkten - die zeitliche Variabilität für eine Messstelle

Tabelle A3.7: Ergebnisse und Spannweitenberechnungen für die Validierungsstudie, gelöstes Eisen, Basisdaten fett gedruckt, Symbole nur zur Beschreibung der Berechnungen verwendet (T: Zielobjekt, S: Probe, A: Analyse, R: Betrag der Differenzen)

Messstelle	S1A1	S1A2	S2A1	S2A1	R1	R2	R _{S+A}	Durchschnitt
	mg l ⁻¹	mg l ⁻¹	mg l ⁻¹	mg l ⁻¹	mg l ⁻¹	mg l ⁻¹	mg l ⁻¹	mg l ⁻¹
99,474	0,815	0,834	0,912	0,893	0,019	0,019	0,078	0,86
99,468	1,8	1,83	1,94	1,93	0,030	0,010	0,12	1,88
99,469	1,69	1,68	1,79	1,77	0,010	0,020	0,095	1,73
99,919	2,62	2,61	2,83	2,84	0,010	0,010	0,22	2,73
99,327	1,66	1,63	1,58	1,59	0,030	0,010	0,06	1,62
99,371	1,52	1,53	1,47	1,50	0,010	0,030	0,04	1,51
				Durchschnitt	0,018	0,017	0,102	1,719
							Standardabweich.	0,604
Analyse	$R_A = (\overline{R_1} + \overline{R_2}) / 2$		$R_A =$	0,017	$S_A = R_A / 1,128$		$s_A =$	0,015
							$CV_A =$	0,89 %
Probenahme	$S_{S+A} = \overline{R_{S+A}} / 1,128$		$S_{S+A} =$	0,091	$S_S = \sqrt{S_{S+A}^2 - \left(\frac{S_A}{\sqrt{2}}\right)^2}$		$s_S =$	0,090
							$CV_S =$	5,23 %
Zwischen Zielobjekten			$S_{T+S+A} =$	0,604	$S_T = \sqrt{S_{T+S+A}^2 - \left(\frac{s + A}{\sqrt{2}}\right)^2}$		$s_T =$	0,601
							$CV_T =$	35 %

In keiner der Grundwasserproben wurde ein gelöster Sauerstoffgehalt von über 0,1 mg l⁻¹ gemessen, und das gemessene niedrige Redoxpotential (-110 bis -200 mV) steht im Einklang mit dem Fehlen von Sauerstoff (<0,1 mg l⁻¹) und den hohen Konzentrationen an gelöstem Eisen (0,92 bis 2,8 mg l⁻¹).

7 Anmerkungen

Für das während der Validierung gemessene gelöste Eisen lieferte die Verwendung von ANOVA- und RANOVA-Berechnungen statistische Schätzungen, die sich nur geringfügig von denen unterschieden, die mit den einfachen Spannweitenberechnungen ermittelt wurden.

Die aus der Validierung und der anschließenden Qualitätskontrolle geschätzten Werte für die analytische Unsicherheit (d. h. 1,9 % und 2,5 %) stimmen weitgehend überein. Die Unsicherheit bei der Probenahme war geringer, wenn bei der Qualitätskontrolle nur eine Messstelle bei verschiedenen Gelegenheiten beprobt wurde (3,6-3,8 %), als wenn bei der Validierung verschiedene Messstellen zur gleichen Zeit beprobt wurden (10-15 %). Auch die

Unsicherheit zwischen den Zielobjekten zu verschiedenen Zeitpunkten war geringer als bei der Beprobung verschiedener Messstellen.

Wäre eine kontinuierliche Kontrolle der Probenahmeunsicherheit erforderlich gewesen, hätten die Kontrolldaten in Kontrollkarten dargestellt werden können, um eine frühzeitige Warnung vor übermäßiger Unsicherheit (Zufallsfehler) bei jeder Probenahme zu erhalten.

Die Anzahl der Wiederholungen (sechs) in dieser Studie war geringer als in den meisten anderen Fällen, und das Risiko eines geringeren Vertrauens in die Unsicherheitsschätzungen sollte bei der Bewertung der Ergebnisse berücksichtigt werden.

Der Unsicherheitsbeitrag, der sich aus dem Bias der Probenahme ergibt, wurde nur durch die Bewertung der Konsistenz der Messungen verschiedener, miteinander in Wechselbeziehung stehender chemischer Parameter (Sauerstoff, gelöstes Eisen, Redox) berücksichtigt, und die Bewertung bestätigte die Schlussfolgerung, dass es durch die Probenahme und die Vorbehandlung der Proben gelungen war, Bias durch Sauerstoffeinfluss und Filterverstopfung zu vermeiden.

8 Zusammenfassung

Die Messunsicherheit (%-Unsicherheit mit Erweiterungsfaktor 2) ist im Folgenden für gelöstes Eisen zusammengefasst.

Die Daten zeigen, dass die Forderung nach weniger als 20 % erweiterter Messunsicherheit für gelöstes Eisen erfüllt werden konnte (Probenahmevalidierung) und dass die geforderte Messunsicherheit in der Realität während der Routineüberwachung erreicht wurde (Qualitätskontrolle der Probenahme). Darüber hinaus zeigen die Daten, dass, wenn eine Verbesserung der Überwachungssicherheit erforderlich wäre, der offensichtliche Ansatzpunkt für eine Verbesserung eine erhöhte Überwachungsichte für gelöstes Eisen wäre (die Unsicherheit zwischen den Zielobjekten dominiert).

Gelöstes Eisen im Grundwasser	Erweiterte Unsicherheit, Erweiterungsfaktor 2			Variabilität zwischen den Zielobjekten
	Probenahme	Analysen	Messung	
Validierung	11 %	1,9 %	11 %	70 % ¹
Qualitätskontrolle	3,6 %	2,5 %	4,4 %	9,9 % ²
¹ In der Validierungsstudie war die Variabilität zwischen den Zielobjekten die zwischen den Messstellen ² Bei der Qualitätskontrolle war die Variabilität zwischen den Zielobjekten die zwischen den einzelnen Probenahmen				

9 Danksagung

Die hier vorgestellten Arbeiten wurden vom Nordic Innovation Centre, dem Ausschuss für Boden- und Grundwasserkontamination des dänischen Rates für technische Wissenschaften und dem Kreis Århus, Dänemark, unterstützt. Die Untersuchungen vor Ort wurden von Mogens Wium, GEO, fachkundig durchgeführt.

Beispiel A4: Vitamin A in Babybrei mit Obst und gemahlenem Getreide

Messgröße				Unsicherheitsschätzung		
Analyt/ Technik	Ein- heit	Sektor/ Matrix	Probenahme- zielobjekt	Zweck	Design	Statistik
Vitamin A (als Retinol)/ HPLC	μg 100 g^{-1} in Pulver- form	Lebens- mittel/ Babybrei mit Obst	Produzierte Charge	Gesamtmes- unsicherheit	Empirische Duplikat- methode	Einfache ANOVA der Konzentrations- werte

1 Anwendungsbereich

Ziel ist die Schätzung der Messunsicherheit und die Beiträge von Probenahme und Analyse. Die Schätzungen basieren auf Proben einer Sorte von Babybrei - entnommen aus 10 verschiedenen Chargen - unter Verwendung einer Probenahmervorschrift, die Doppelproben aus jeder Charge sammelt.

2 Szenario und Probenahmezielobjekt

Bei der Herstellung von Babybrei wird das Vitamin A (Retinol) als Vormischung (zusammen mit Vitamin D und Vitamin C) zugesetzt. Die Vormischung ist eine geringfügige Zutat. Alle Zutaten werden vor der Verteilung auf die Verpackungen gründlich gemischt. Frühere Analysen haben gezeigt, dass die Schwankungen der Analyseergebnisse zwischen den einzelnen Packungen größer sind als erwartet. Eine Messunsicherheit von 20 - 30 % wird als akzeptabel angesehen. Es wurde die Frage aufgeworfen, ob die Schwankungen hauptsächlich auf die analytische Unsicherheit oder auf die Unsicherheit der Probenahme zurückzuführen sind. Eine der Theorien besagt, dass das Vitamin lokal ungleichmäßig in der Packung verteilt ist und daher zu einer größeren analytischen Unsicherheit führt, wenn die Prüfmenge zu klein ist (z. B. 3-5 g).¹ Eine mögliche Erklärung für die Heterogenität ist, dass die Vitamin-Vormischung aufgrund elektrostatischer Wechselwirkungen mit den Fruchtpartikeln im Breipulver in kleinen Hotspots aggregiert. Die Hersteller empfehlen für die Analyse von Vitamin A, D und C in Babybreipulver eine Prüfmenge von 40 - 50 g.

Um die gemessene Vitamin-A-Konzentration mit den angegebenen Werten und den europäischen Grenzwerten vergleichen zu können, ist eine Schätzung der Messunsicherheit wünschenswert. Zur Bestimmung der Zufallskomponente der Messunsicherheit wurde ein empirischer Ansatz unter Verwendung der Duplikatmethode (siehe Abschnitt 9.4.2) gewählt. Zur Schätzung der systematischen Komponente wurde ein Vergleich mit einem Referenzwert durchgeführt.

¹ EN-12823-1 ,Lebensmittel - Bestimmung von Vitamin A mittels HPLC' gibt eine Probe von etwa 2 - 10 g an.

3 Probenahmenvorschrift

Normalerweise wird ein Stichprobenverfahren angewandt, bei dem eine Probe (eine Verpackung) einer Charge als Screening-Probe verwendet wird, indem ihr Gehalt mit den angegebenen Werten und gesetzlichen Grenzwerten verglichen wird.

Validierung - In dieser Studie werden aus 10 verschiedenen Chargen *einer Sorte* von Babybrei-Pulver jeweils zwei Proben entnommen (d. h. 10 Probenahmezielobjekte). Jede Probe entspricht einer Packung von etwa 400 g Pulver.

Qualitätskontrolle - Die Qualitätskontrolle (QC) der Probenahme von verschiedenen Sorten von Babybrei erfolgt durch die Entnahme von zwei Proben aus jeder von acht Chargen *verschiedener* Sorten von Babybrei (d. h. acht Probenahmezielobjekte). Alle Breisorten enthalten neben gemahlenem Getreide auch Obst.

Um die Qualität in jeder Verpackung des Produkts zum Zeitpunkt des Mindesthaltbarkeitsdatums des Breipulvers zu gewährleisten, verpackt der Hersteller das Produkt in einen luftdichten und lichtgeschützten Beutel. Es wird daher davon ausgegangen, dass der Abbau von Vitamin A während der normalen Lagerzeit vernachlässigbar ist. Die Probenahme für die Validierung erfolgte am Ort der Herstellung. Für die Qualitätskontrolle wurden die Proben zum Teil am Herstellungsort und zum Teil bei den Einzelhändlern gekauft. Bei der Entnahme der Proben bei den Einzelhändlern wurde darauf geachtet, dass die beiden Proben (von jedem Produkt) von verschiedenen Einzelhändlern entnommen wurden, aber auch darauf, dass die Proben die gleiche Chargenkennzeichnung hatten. Dies ist wichtig, um zu vermeiden, dass Schwankungen zwischen den Chargen die offensichtliche Unsicherheit der Probenahme erhöhen, da die Probenahmenvorschrift in diesem Fall die Probenahme aus einer bestimmten Charge vorschreibt.

4 Studiendesign – empirischer Ansatz

Es wurde ein empirischer ("Top-down"-) Ansatz – Duplikatmethode gewählt, um die Zufallskomponente der Probenahmeunsicherheit zu schätzen. Die Validierung wird an einer Sorte Babybrei mit Obst und gemahlenem Getreide durchgeführt. Bei der Probenahme für die Qualitätskontrolle werden verschiedene Babybreiprodukte (die alle Obst und gemahlene Getreide enthalten) getestet, um festzustellen, ob die Schätzung der Messunsicherheit aus der Validierungsstudie für verschiedene Sorten von Babybreien, die Obst und gemahlene Getreide enthalten, zutreffend ist.

4.1 Validierung

Die Proben werden on-line (unmittelbar nach dem Abfüllen der Packungen) zu zufälligen Zeitpunkten entnommen. Von jeder der 10 Produktionseinheiten (Chargen) einer Sorte Babybrei-Pulver werden zwei Proben (zwei Packungen mit jeweils ca. 400 g) entnommen.

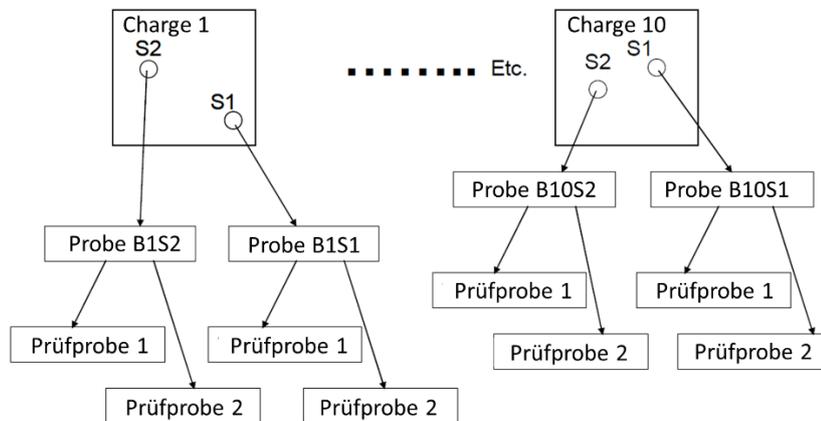


Abbildung A4.1: Probenahme für die Validierung. Von 10 Produktionseinheiten/Chargen der gleichen Sorte Babybrei werden jeweils zwei Proben entnommen

4.2 Qualitätskontrolle

Für die Qualitätskontrolle (QC) werden zwei Proben aus je einer Charge von acht verschiedenen Sorten von Babybrei, die Obst und gemahlenes Getreide enthalten, entnommen. Bei den Breien handelt es sich um Produkte von drei verschiedenen Herstellern. Die Proben (mit Ausnahme von zwei Sorten Brei) wurden von zwei Herstellern zur Verfügung gestellt. Der Rest wurde beim Einzelhändler gekauft.

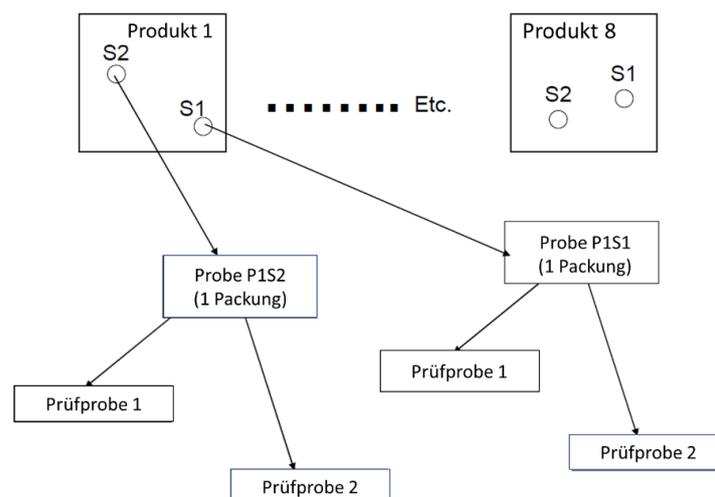


Abbildung A4.2: Probenahme für die Qualitätskontrolle. Zwei Proben werden von je einer Charge von acht verschiedenen Sorten Babybrei entnommen

5 Probenahme und Analysen im Labor

Die analytischen Arbeiten wurden vom "National Institute of Nutrition and Seafood Research" (NIFES) durchgeführt. Das Labor ist nach der Norm EN ISO/IEC 17025 akkreditiert.

Das Labor nimmt an Laboreignungstests (FAPAS und Bipea) mit guten Ergebnissen teil (im Zeitraum 2000-2005, $|Z\text{-score}| < 1$). Die Methode wurde mit einem zertifizierten Referenzmaterial (CRM) validiert. Die Kennzahlen des Labors sind in Tabelle A4.1 aufgeführt.

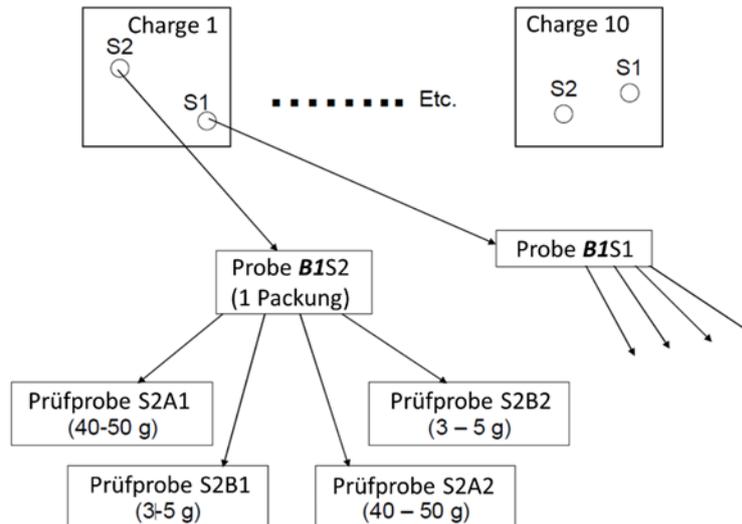
Tabelle A4.1: Methoden und Kennzahlen aus der Qualitätskontrolle - Laboranalysen

Parameter	Vitamin A – bestimmt als Retinol
Methode	HPLC – Normalphasensäule – UV-Detektion
Wiederholbarkeit	2RSD(%) = 6
Vergleichbarkeit innerhalb des Labors	2RSD(%) = 8
Messunsicherheit	14 % (95 % Konfidenzintervall)
Wiederfindung	<ul style="list-style-type: none"> ➤ Standardaddition, im Labor: 90 – 110 % ➤ Auf der Grundlage von Laboreignungstests (im Zeitraum 1999-2005), verschiedene Matrices: 88-113 %, mittlere Wiederfindung 100,5 %
Bestimmungsgrenze (LOQ)	0,14 mg kg ⁻¹
Genutztes ZRM	NIST 2383 – Babynahrung (gemischte Nahrung)
➤ ZRM – zertifizierter Wert	0,80 ± 0,15 mg kg ⁻¹ (95 % Konfidenzintervall)
➤ ZRM – analysierter Wert	0,77 ± 0,14 mg kg ⁻¹ (n = 28, 95 % Konfidenzintervall)

5.1 Sekundäre Probenahme

Zur Teilung der Proben wird ein mechanischer Probenteiler (Retsch) verwendet. Von jeder der Primärproben werden vier Prüfproben entnommen: zwei Portionen von etwa 3 – 5 g und zwei Portionen von etwa 40-50 g.

Abbildung A4.3: Teilung der Primärprobe zur Herstellung von vier Prüfproben



5.2 Analysen

Die Analysenmethode basiert auf EN-12823-1 (Lebensmittel - Bestimmung von Vitamin A mit Hochleistungs-Flüssigchromatographie - Teil 1: Bestimmung von all-E-Retinol und 13-Z-Retinol). Retinol wird mit ethanolischem Kaliumhydroxid, das Antioxidantien enthält, verseift. Vitamin A wird mit Hexan extrahiert. Die Analyse erfolgt mittels Hochleistungs-Flüssigchromatographie (HPLC) mit einem UV-Detektor.

Bei der Validierung werden für jede der Primärproben zwei Analysen an Proben von 40 - 50 g und zwei Analysen an Proben von 3 - 5 g durchgeführt. Bei der Qualitätskontrolle werden zwei Analysen an Proben von 40 - 50 g durchgeführt. Für jede Probe wird nur eine analytische Bestimmung durchgeführt (keine analytische Doppelbestimmung).

6 Informationen des Herstellers

Die Daten zur Schätzung des "zertifizierten Werts" von Vitamin A in Babybrei werden vom Hersteller (Nestlé) des für die Validierung ausgewählten Produkts zur Verfügung gestellt (siehe Tabelle A4.2).

Tabelle A4.2: Produktdaten des Herstellers

Produkt	Haferflockenbrei mit Bananen und Aprikosen (Nestlé)
Gewicht der Charge, einschließlich Vormischung (1 Charge = 2 Mischbehälter)	1092 kg
Gewicht des zugesetzten Vitamin-Vorgemisches in der Charge	9016 IU g ⁻¹ = 2705 µg g ⁻¹ (Retinol)
Vitamin A in der Vormischung (Angaben aus dem Analysenzertifikat)	304 µg 100 g ⁻¹ (Retinol)
Vitamin A in den Zutaten gemäß der Produktspezifikation	45 µg 100 g ⁻¹ (Retinol)
Geschätzter "wahrer Wert" von Vitamin A	349 µg 100 g ⁻¹ (Retinol)
Vitamin A deklariert als	Retinol - (Summe aus Trans- und Cis-Retinol)

7. Ergebnisse

Prüfprobe à 40 g – Babybrei

Tabelle A4.3: Validierungsdaten - von demselben Produkt, Ergebnisse in µg 100 g⁻¹ Pulver

Charge	S1A1	S1A2	S2A1	S2A2
B1	402	325	361	351
B2	382	319	349	362
B3	332	291	397	348
B4	280	278	358	321
B5	370	409	378	460
B6	344	318	381	392
B7	297	333	341	315
B8	336	320	292	306
B9	372	353	332	337
B10	407	361	322	382

S1 und S2: Primärproben von Probenahmestelle 1 und 2 einer Produktionscharge

A1 und A2: Analysen von Doppelproben einer Primärprobe S

Analysierter Mittelwert (Probe 40 g): 348 µg 100 g⁻¹.

Prüfprobe à 4 g – Babybrei

Tabelle A4.4: Validierungsdaten - von demselben Produkt, Ergebnisse in $\mu\text{g } 100 \text{ g}^{-1}$ Pulver

Charge	S1B1	S1B2	S2B1	S2B2
B1	400	491	323	355
B2	413	159	392	434
B3	315	391	252	454
B4	223	220	357	469
B5	462	343	262	293
B6	353	265	305	456
B7	298	234	152	323
B8	425	263	417	353
B9	622	189	291	272
B10	292	397	142	568

S1 und S2: Primärproben von Probenahmestelle 1 und 2 einer Produktionscharge

B1 und B2: Analysen von Doppelproben einer Primärprobe S

Analysierter Mittelwert (Probe 4 g): $341 \mu\text{g } 100 \text{ g}^{-1}$

7.1 Berechnungen

Die ANOVA-Berechnung kann mit den in Excel, Minitab, SPSS usw. verfügbaren Tools durchgeführt werden. In dieser Studie werden die Berechnungen in einem Excel-Tabellenblatt durchgeführt, und die Einzelheiten sind in Abschnitt 11 - ANOVA-Berechnungen - aufgeführt.

Berechnung der Unsicherheit der Analysen, einseitige ANOVA, Prüfprobe à 40 g

Tabelle A4.5: Ergebnisse der ANOVA-Berechnungen - Unsicherheit der Analysen - Summe der Quadrate der Differenzen, innerhalb der Gruppen (SS_{Error} SS_E). Für Details siehe Abschnitt A4.11

$SS_{E-\text{Anal}}$ ($\mu\text{g } 100 \text{ g}^{-1}$) ²	Freiheits- grad (df)	Varianz = $SS_{E-\text{Anal}}/\text{df}$ ($\mu\text{g } 100 \text{ g}^{-1}$) ²	Standardabweichung SD_{anal} = $\sqrt{(SS_{E-\text{Anal}})/\text{df}}$ ($\mu\text{g } 100 \text{ g}^{-1}$)	Relative Standardabweichung $RSD_{\text{anal}}(\%)$ = $(SD / \bar{X}_a) * 100 \%$
16595	20	829,75	28,805	8,28

Berechnung der Unsicherheit der Probenahme, einfache ANOVA, Prüfprobe à 4 g

Tabelle A4.6: Ergebnisse der ANOVA-Berechnungen - Unsicherheit der Probenahme - Summe der Quadrate der Differenzen. Für Einzelheiten siehe Abschnitt A4.11

SS_S ($\mu\text{g } 100 \text{ g}^{-1}$) ²	Freiheits- grad (df)	Varianz $V_{\text{Samp}} =$ $(SS_S/df_S - SSE_{\text{Anal}}/df_A)2$ ($\mu\text{g } 100 \text{ g}^{-1}$) ²	Standardabweichung SD_{samp} $= \sqrt{V_{\text{samp}}}$ ($\mu\text{g } 100 \text{ g}^{-1}$)	Relative Standard- abweichung $RSD_{\text{samp}}(\%)$ $= (SD / \bar{X}_s) * 100 \%$
14231	10	296,7	17,22	4,95

Berechnung der Messunsicherheit – Prüfprobe à 40 g

Der RSD-Wert (%) aus der ANOVA-Berechnung kann als Schätzwert für die Standardunsicherheit u (%) verwendet werden. Das Analyselabor hat die analytische Standardunsicherheit auf 7 % geschätzt, was niedriger ist als die zufällige analytische Komponente für diesen Probentyp von 8,28 %. Der höhere der beiden Werte wird für die Berechnungen verwendet. Kombiniert man die RSD-Werte aus den Tabellen A4.5 und A4.6 mit Gleichung 1, so können die Ergebnisse wie in Tabelle A4.7 dargestellt werden.

$$u_{\text{meas}} = \sqrt{(u_{\text{samp}})^2 + (u_{\text{anal}})^2} \quad (\text{Gleichung A1})$$

Tabelle A4.7: Messunsicherheit - 40 g Prüfprobe

Messunsicherheit, ANOVA Berechnungen – 40 g Prüfprobe			
	Probenahme	Analyse	Gesamt
Unsicherheit u (%)	4,95	8,28	9,7
Erweiterte Unsicherheit U (%) = $2 * u$ mit einem Erweiterungsfaktor von 2 (d. h. 95 % Konfidenz)	9,9	17	20

Berechnung der Unsicherheit der Analysen, einfache ANOVA, Prüfprobe à 4 g

Es werden dieselben Berechnungen angestellt wie bei der Probengröße von 40 g (siehe Tabelle A4.14 in Abschnitt 11 dieses Beispiels).

Tabelle A4.8: Ergebnisse der ANOVA-Berechnungen - Unsicherheit der Analysen, Prüfprobe à 4 g - Summe der Quadrate der Differenzen, innerhalb der Gruppen (SS-Error SS_E)

SS_E ($\mu\text{g } 100 \text{ g}^{-1}$) ²	Freiheits- grad (df) $(N * 2 - N) = 20$	Varianz $= SS_E / df$ ($\mu\text{g } 100 \text{ g}^{-1}$) ²	Standardabweichung SD_{anal} $= \sqrt{SS_E / df}$ ($\mu\text{g } 100 \text{ g}^{-1}$)	Relative Standardabweichung $RSD_{\text{anal}}(\%)$ $= (SD / X_a) * 100 \%$
--	---	--	---	--

312206,5	20	15610,325	124,9413	36,6800
----------	----	-----------	----------	---------

Berechnung der Unsicherheit der Probenahme, einfache ANOVA, Prüfprobe à 4 g

Tabelle A4.9: Ergebnisse der ANOVA-Berechnungen - Unsicherheit der Probenahme, Prüfprobe à 4 g - Summe der Quadrate der Differenzen

SS_S ($\mu\text{g } 100 \text{ g}^{-1}$) ²	Freiheits- grad (df)	Varianz $V_{\text{Samp}} =$ $(SS_S/df_S - SSE_{\text{Anal}}/df_A)2$ ($\mu\text{g } 100 \text{ g}^{-1}$) ²	Standardabweichung SD_{samp} $= \sqrt{V_{\text{Samp}}}$ ($\mu\text{g } 100 \text{ g}^{-1}$)	Relative Standardabweichung $RSD_{\text{samp}}(\%)$ $= (SD / \bar{X}_s) * 100 \%$
102860,25	10	-2662,15	$\sqrt{-2662,15}$ auf null gesetzt	Per Konvention auf null gesetzt

Es werden die gleichen Berechnungen angestellt wie bei der Probengröße von 40 g (Tabelle A4.15, Abschnitt 11 dieses Beispiels).

Der negative Wert von V_{Samp} in Tabelle A4.9 zeigt an, dass SD_{samp} im Vergleich zum berechneten Wert von SD_{anal} klein ist. In diesem Fall bestätigten die Schätzungen von SD_{anal} und SD_{samp} unter Verwendung der robusten ANOVA die kleinere Probenahme-Standardabweichung; die robusten ANOVA-Schätzungen waren $u_{\text{samp}}(\%) = 6,9 \%$ und $u_{\text{Anal}}(\%) = 30 \%$.

Da die Probenahme für die Experimente mit 40 g und 4 g Prüfproben identisch ist (und die Unsicherheit der Probenahme daher gleich sein sollte), wird eine $RSD_{\text{samp}}(\%) = 5 \%$ ($\approx 4,95$ siehe Tabelle A4.7) als Schätzwert verwendet.

Berechnung der Messunsicherheit - 4 g Prüfprobe

Unter Verwendung des berechneten RSD-Wertes (%) in den Tabellen A4.5 und A4.6 als Schätzung der Messunsicherheit und in Kombination mit Gleichung A1 können die Ergebnisse wie folgt dargestellt werden (Tabelle A4.10).

Tabelle A4.10: Messunsicherheit - Prüfprobe à 4 g

Messunsicherheit, ANOVA Berechnungen – Prüfproben à 4 g			
	*Probenahme	Analyse	Messung
Unsicherheit u (%)	5	36,7	37
Erweiterte Unsicherheit U (%) = $2 * u$ mit einem Erweiterungsfaktor von 2 (d. h. 95 % Konfidenz)	10	73,4	74

*Der Wert u (%) ergibt sich aus Berechnungen mit Prüfprobe à 40 g

7.2 Auswirkung der Größe der Prüfprobe auf die Messunsicherheit

Das Babybrei-Pulver sieht homogen aus, so dass eine geringe Messunsicherheit (u) erwartet wird. Die Analysen des Pulvers ergaben jedoch eine überraschend große u bei einer Prüfprobengröße von 4 g (die CEN-Norm EN-12823-1 gibt eine Prüfprobengröße von etwa 2 - 10 g an). Die Hersteller empfehlen eine Prüfprobengröße von 40 - 50 g.

Die Validierungstests ergaben folgende Ergebnisse

Tabelle A4.11: Vergleich der Messunsicherheit - Prüfprobe von 40 g und 4 g

Größe der Prüfprobe	Messunsicherheit (U_{meas})	Erweiterte Messunsicherheit U_{meas}
Prüfprobe à 40 g	9,7 %	20 %
Prüfprobe à 4 g	37 %	74 %

Es kann festgestellt werden, dass $u_{40g} \ll u_{4g}$. Eine U_{meas} von etwa 20 % ist akzeptabel, wenn man das Herstellerkriterium von 20 - 30 % zugrunde legt, während ein U_{meas} von 74 % als zu hoch angesehen wird, wenn man die Matrix und die Produktionsbedingungen dieser Produktsorte berücksichtigt.

Daraus lässt sich schließen, dass ein Prüfprobengewicht von 4 g für die Analyse von Vitamin A (Retinol) in Babybrei-Pulver, das gemahlene Getreide und Obst enthält, den Zweck nicht erfüllt. Es wird eine Probengröße von 40 - 50 g empfohlen. Dies unterstützt auch die Theorie, dass das Vitamin ungleichmäßig im Produkt verteilt ist, möglicherweise als lokale "Hot Spots" aufgrund elektrostatischer Wechselwirkungen.

7.3 Qualitätskontrolle

Gemäß Abschnitt 13.2.2 dieses Leitfadens ist das wichtigste Instrument der Qualitätskontrolle die Wiederholung. Diese wird mit geringfügigem Aufwand durchgeführt, indem zwei Proben von jedem Zielobjekt durch eine vollständige (und angemessen randomisierte) doppelte Ausführung der Probenahmenvorschrift genommen werden. Die Probe muss nur einmal analysiert werden, und die Differenz zwischen den Ergebnissen $D = |x_1 - x_2|$ wird berechnet. In dieser Studie wurde jede Probe zweimal analysiert, aber die Vergleiche wurden zwischen einer Analyse jeder Probe (doppelter Satz) durchgeführt.

In der Studie zur Qualitätskontrolle wurden Prüfmengen von 40 g verwendet. Den Angaben zufolge enthalten die Produkte unterschiedliche Mengen an Vitamin A.

Tabelle A4.12: Qualitätskontrolldaten für eine Prüfmenge von 40 g verschiedener Produkte, $\mu\text{g } 100 \text{ g}^{-1}$ Pulver

Produkt	Hersteller	Zutaten für Breipulver	S1A1	S1A2	S2A1	S2A2
P1	1	Hafer, Reis und Birne	322	319	350	375
P2	1	Hafer, Roggen, Reis und Birne	332	317	358	393
P3	1	Weizen, Banane und Apfel	443	430	461	388
P4	1	Weizen und Apfel	318	383	390	334
P5	2	Hafer, Reis und Banane	252	219	265	227
P6	2	Weizen und Apfel	274	239	233	217
P7	2	Hafer, Reis und Apfel	206	225	198	195
P8	3	Weizen, Dinkel, Hafer und Apfel (Bio-Produkt)	392	335	375	416

S1 und S2: Primärproben (Laborproben) von den Probenahmestellen 1 und 2 einer Charge jeden Produkts.

A1 und A2: Analysen von zwei Prüfproben aus jeder Laborprobe.

Qualitätskontrolle - Berechnung und Regelkarte

Die validierten Unsicherheiten der Probenahme und der Analyse sind u_{samp} bzw. u_{anal} . Die Erstellung einer Regelkarte wird in Abschnitt 13.2 beschrieben. Im Fall von Babybrei (40 g Probe) können die folgenden Berechnungen durchgeführt werden:

$$\text{Warngrenze: } WL = 2,38 * \sqrt{u_{\text{anal}}^2 + u_{\text{samp}}^2} = 2,83 * \sqrt{(4,95^2 + 8,28^2)}\% = 27 \%$$

$$\text{Eingriffsgrenze: } AL = 3,69 * \sqrt{(4,95^2 + 8,28^2)}\% = 36 \%$$

$$\text{Mittellinie: } CL = 1,128 * \sqrt{(4,95^2 + 8,28^2)}\% = 11 \%$$

Die berechnete D (%) in Tabelle A4.13 kann direkt mit der Eingriffsgrenze verglichen werden, oder die Ergebnisse können in einer Regelkarte dargestellt werden, siehe Abbildung A4.4.

Tabelle A4.13: Qualitätskontrolle: Berechnung der Differenzen D und D (%) - zwischen Proben einer Charge ($\mu\text{g } 100 \text{ g}^{-1}$ Pulver)

Produkt	Analysen	Probe S1 x_{S1}	Probe S2 x_{S2}	$D = x_{S1} - x_{S2} $	\bar{x}	$D(\%) = (D/\bar{x}) * 100$
P1	A1	322	350	28	336	8
P2		332	358	26	345	8
P3		443	461	18	452	4
P4		318	390	72	354	20
P5		252	265	13	259	5
P6		274	233	41	254	16
P7		206	198	8	202	4
P8		392	375	17	384	4
P1	A2	319	375	56	347	16
P2		317	393	76	355	21
P3		430	388	42	409	10
P4		383	334	49	359	14
P5		219	227	8	223	4
P6		239	217	22	228	10
P7		225	195	30	210	14
P8		335	416	81	376	22

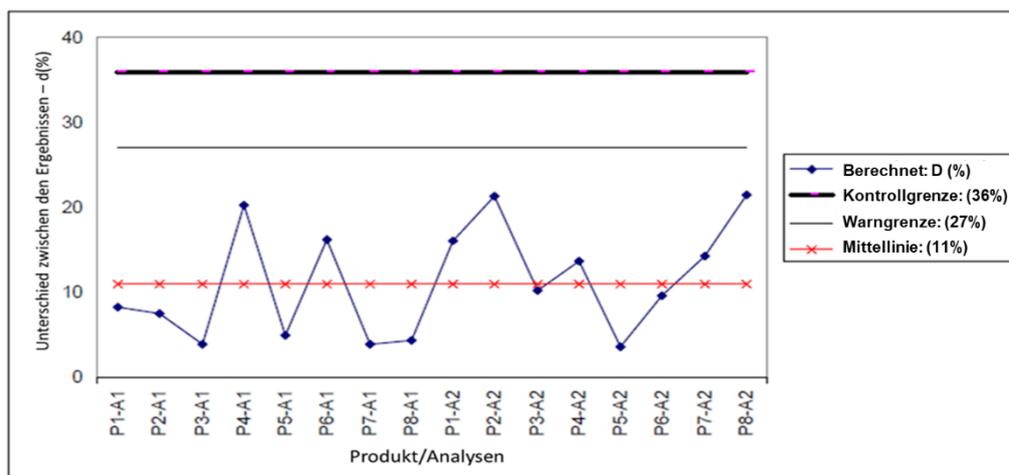


Abbildung A4.4: Regelkarte: Qualitätskontrollanalysen von Vitamin A in getreide- und obsthaltigem Babybrei

Die Regelkarte in Abbildung A4.4 zeigt, dass bei der Entnahme von Doppelproben aus derselben Charge die Differenz zwischen den Analyseergebnissen D (%) kleiner ist als die Eingriffsgrenze. Alle berechneten Differenzen sind tatsächlich kleiner als die Warngrenze von 27 %.

Die im Validierungsschritt ermittelte Messunsicherheit wird daher als geeignet für die Qualitätskontrolle der Probenahme von Babybrei mit gemahlenem Getreide und Obst angesehen.

Wenn das normale Verfahren darin besteht, eine Probe aus jeder Charge zu analysieren, wird empfohlen, mindestens in einer von zehn der beprobten Chargen Doppelproben aus derselben Charge zu entnehmen.

Messunsicherheit

Probenahmeunsicherheit

Berechnungen aus der Validierungsstudie ergaben eine erweiterte Probenahmeunsicherheit U_{samp} (%) = 10 % (40 g Prüfprobe - siehe Tabelle A4.7). Die berechnete Unsicherheit enthält keine Beiträge zur Unsicherheit aufgrund von Differenzen „zwischen den Vorschriften“ und „zwischen den Probennehmern“.

Analytische Unsicherheit

Die Berechnungen aus der Validierungsstudie ergaben eine erweiterte Messunsicherheit der Analysen (U_{anal}) von 17 % - 40 g Prüfprobe. Das Labor gibt seine eigene Schätzung der analytischen Unsicherheit an (siehe Tabelle A4.1): $2 \cdot RSD_{inlab}(\%) = 14$ %. $2 \cdot RSD_{inlab}(\%)$ wird als Schätzwert für U_{anal} im Labor verwendet. Das in der Validierungsstudie gefundene U_{anal} lag auf demselben Niveau, aber immer noch ein wenig höher als die vom Labor angegebene U_{anal} .

Das verwendete zertifizierte Referenzmaterial (ZRM) ist 2383 (NIST) - Babynahrungsgemisch. Das ZRM ist eine Mischung aus verschiedenen Lebensmitteln pflanzlichen und tierischen Ursprungs - und die bei der Analyse des ZRM festgestellte Unsicherheit ist möglicherweise nicht identisch mit derjenigen, die bei der Analyse von Babybrei-Pulver festgestellt wird. Die Labordaten für das ZRM 2383 sind in der nachstehenden Tabelle enthalten.

ZRM 2383	Mittelwert mg kg ⁻¹	U (%) _{95%}	Laborbias (%)
zertifiziert	0,80 ± 0,15	18,8	-
analysiert	0,77 ± 0,14	18,2	-3,75

Die Messunsicherheit und der für das ZRM ermittelte Bias könnten in der analytischen Messunsicherheit berücksichtigt werden (wie im NordTest UFS-Leitfaden, Beispiel 2), aber da sich die Matrix in der Validierungsstudie von der für das verwendete ZRM unterscheidet, haben wir uns entschieden, es nicht in diese Studie aufzunehmen.

Gesamt-Messunsicherheit

Berechnungen aus der Validierungsstudie ergaben eine erweiterte Messunsicherheit U_{meas} (%) = 20 % (40 g Prüfprobe - siehe Tabelle A4.7).

Systematische Abweichung

Das Labor berichtet eine Wiederfindung von normalerweise 90 - 110 %. Wiederfindung auf der Grundlage von Laboreignungsprüfungen 1999-2005: 88 - 113 %. Die Ergebnisse der Eignungsprüfungen deuten auf keinen (oder nur sehr geringen) systematischen Bias hin. Die Analyse von ZRM 2383 im Labor ergibt einen mittleren analysierten Wert von 96,3 % des zertifizierten Wertes, was auf eine geringe Abweichung (-3,7 %) hinweist. Da sich die Matrix des ZRM "Baby-Nahrungsmischung" von der des Babybreis unterscheidet und die Analyseverfahren eine Extraktion beinhaltet, ist der bei der Analyse des ZRM ermittelte Bias möglicherweise nicht repräsentativ für die Analysen von Babybrei.

In der Validierungsstudie wurde der Mittelwert für Retinol mit $348 \mu\text{g } 100 \text{ g}^{-1}$ bestimmt (bei Verwendung einer Probe von 40 g). Nach Angaben des Herstellers (siehe Tabelle A4.2) wurde der "wahre Wert" für Retinol auf $349 \mu\text{g } 100 \text{ g}^{-1}$ Breipulver berechnet. Dies ergibt eine Wiederfindung von 99,7 % des "wahren Wertes". Dies deutet darauf hin, dass der systematische Fehler aufgrund von Probenahme und Analysen gering ist und bei der Analyse von Babybrei-Pulver, das gemahlene Getreide und Früchte enthält, vernachlässigbar sein könnte - unter der Voraussetzung, dass eine Probe von 40 - 50 g verwendet wird.

8 Anmerkungen

Wenn eine Probe von etwa 40 g verwendet wird, sollte die Retinol-Konzentration C in Babybrei, der gemahlene Getreide und Früchte enthält, mit der erweiterten Messunsicherheit angegeben werden, d. h. $C \pm 20\%$ des gemessenen Wertes C (95 % Konfidenz).

Wenn Babybrei-Pulver, das gemahlene Getreide und Früchte enthält, analysiert werden soll, wird empfohlen, eine relativ große Prüfprobe von etwa 40 - 50 g zu verwenden und nicht 2 - 10 g, wie in der offiziellen CEN-Methode (EN-12823-1) angegeben. Da die analytische Unsicherheit (40 g Prüfprobe) größer war als die normale analytische Unsicherheit des Labors, können auch größere Proben als 40 g in Betracht gezogen werden.

9 Zusammenfassung

Messunsicherheit für 40-g-Prüfproben				Probe
	Probenahme	Analytisch	Gesamt	Typische Abweichung zwischen den Zielobjekten ²
Unsicherheit u (%) = RSD (%)	4,95	8,3	9,7	6,1
¹ Erweiterte Unsicherheit U (%) = $2 \cdot u$	9,9	16,6	19,4	12,2

¹ Mit einem Erweiterungsfaktor von 2 (d. h. 95 % Konfidenz)

² berechnet als $RSD_B(\%)$ aus den Daten in Tabelle A4.3 unter Anwendung der klassischen ANOVA

10 Danksagung

Nestlé (Norwegen) danken wir für die engagierte Zusammenarbeit und für die Bereitstellung von Proben für das Projekt (Validierungs- und Qualitätskontrollstudie). Auch Smaafolk - Tine

Norske Meierier wird dafür gedankt, dass sie uns freundlicherweise Proben für die Qualitätskontrollstudie zur Verfügung gestellt haben. Dem Nationalen Institut für Ernährungs- und Meeresfrüchtelforschung (NIFES) wird für den analytischen Beitrag gedankt (Analysen und Informationen über das Qualitätssicherungssystem des Labors). Die Studie wurde mit finanzieller Unterstützung des Nordischen Innovationszentrums und der norwegischen Behörde für Lebensmittelsicherheit durchgeführt.

11 ANOVA-Berechnung, Vitamin A in Babybrei – Details

Berechnung der Unsicherheit der Analysen, einseitige ANOVA, Prüfprobe 40 g

Tabelle A4.14: ANOVA-Berechnungen - Unsicherheit der Analysen - Summe der Quadrate der Differenzen, innerhalb der Gruppen (SS-Error)

Probe	Analysen ($\mu\text{g } 100 \text{ g}^{-1}$)		Mittelwert – jede Probe ($\mu\text{g } 100 \text{ g}^{-1}$)	Quadrat der Differenzen – innerhalb der Gruppen ($\mu\text{g } 100 \text{ g}^{-1}$) ²
	A1 = $x_{ij} = x_{i1}$	A2 = $x_{ij} = x_{i2}$	$\bar{X}_1 = (x_{i1} + x_{i2})/2$	$(x_{i1} - \bar{X}_1)^2$
B1-S1	402	325	363,5	1482,25
B2-S1	382	319	350,5	992,25
B3-S1	332	291	311,5	420,25
B4-S1	280	278	279	1
B5-S1	370	409	389,5	380,25
B6-S1	344	318	331	169
B7-S1	297	333	315	324
B8-S1	336	372	353	362,5
B9-S1	372	353	362,5	90,25
B10-S1	407	361	384	529
B1-S2	361	351	356	25
B1-S2	349	362	355,5	42,25
B3-S2	397	348	372,5	600,25
B4-S2	358	321	339,5	342,25
B5-S2	378	460	419	1681
B6-S2	381	392	386,5	30,25
B7-S2	341	315	328	169
B8-S2	292	306	299	49
B9-S2	332	337	334,5	6,25
B10-S2	322	382	352	900
Mittelwert der Messungen: $\bar{X}_a = 1/20 * \sum_{i=1}^{20} \bar{x}_i$ $= 347,85 \mu\text{g } 100 \text{ g}^{-1}$			SS-Fehler (SS _E):	

			$= \sum_{i=1}^{20} [(x_{i1} - \bar{X}_i)^2 + (x_{i2} - \bar{X}_i)^2]$ $= \sum_{i=1}^{20} 2 * (x_i - \bar{X}_i)^2$	
SS_E $(\mu\text{g } 100 \text{ g}^{-1})^2$	Freiheits- grade (df) $(N*2-N)=20$	Varianz = SSE/df $(\mu\text{g } 100 \text{ g}^{-1})^2$	Standard- abweichung SD_{anal} $= \sqrt{SS_E/df}$ $(\mu\text{g } 100 \text{ g}^{-1})$	Relative Standardabweichung $RSD_{\text{anal}}(\%)$ $= (SD/\bar{X}_a) * 100 \%$
16595	20	829,75	28,80538	8,280978

Anmerkungen zur Tabelle A4.14

1. Berechnung des SS-Fehlers - in diesem Fall werden für jede Laborprobe zwei Prüfproben analysiert, daher:

$$(x_{i1} - \bar{x}_i)^2 = (x_{i2} - \bar{x}_i)^2 \Rightarrow SS_E = \sum_{i=1}^{20} [(x_{i1} - \bar{X}_i)^2 + (x_{i2} - \bar{X}_i)^2] = 2 \sum_{i=1}^{20} (x_i - \bar{X}_i)^2$$

Wenn die Anzahl der analysierten Prüfproben größer als zwei ist, sind die Quadrate der Differenzen nicht gleich, und es ist folgende Berechnung vorzunehmen:

$$SS_E = \sum_{i=1}^{20} \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2$$

2. $df = (N*n-N)=(20*2-20)= 20$, wobei N die Anzahl der Proben und n die Anzahl der analysierten Prüfproben jeder Charge ist.

Berechnung der Unsicherheit der Probenahme, einfache ANOVA, Prüfprobe 40 g ($\mu\text{g } 100 \text{ g}^{-1}$ Pulver)

Tabelle A4.15: ANOVA-Berechnungen - Unsicherheit der Probenahme - Summe der Quadrate der Differenzen

S1A1= x_{i1}	S1A1= x_{i2}	S1A1= x_{i3}	S1A1= x_{i4}	\bar{x}_i	$\left(\frac{x_{i1} + x_{i2}}{2} - \bar{x}_i\right)^2$	$\left(\frac{x_{i3} + x_{i4}}{2} - \bar{x}_i\right)^2$
402	325	361	351	359,75	14,0625	14,0625
382	319	349	352	353	6,25	6,25
332	291	397	348	342	930,25	930,25
280	278	358	321	309,25	915,0625	915,0625
370	409	378	460	404,25	217,5625	217,5625
344	318	381	392	358,75	770,0625	770,0625
297	333	341	315	321,5	42,25	42,25
336	320	292	306	313,5	210,25	210,25
372	353	332	337	348,5	196	196
407	361	322	382	368	256	256
$SS_{samp} = \sum_{i=1}^{10} \left[\left(\frac{x_{i1} + x_{i2}}{2} - \bar{x}_i\right)^2 + \left(\frac{x_{i3} + x_{i4}}{2} - \bar{x}_i\right)^2 \right]$ $= \sum_{i=1}^{10} \left[2 * \left(\frac{x_{i1} + x_{i2}}{2} - \bar{x}_i\right)^2 + 2 * \left(\frac{x_{i3} + x_{i4}}{2} - \bar{x}_i\right)^2 \right] = 14231$						
Mittelwert aller Messungen $\bar{X} = 347,85$				RSD _{samp} (%) = (SD _{samp} / \bar{X})*100 % = 4,95 %		
SS _{E-Anal} = 16595 (siehe Tabelle A4.14)				df _S = 10 (siehe Tabellenhinweis) df _A = 20 (siehe Tabelle A4.14)		
Varianz V _{samp} = (SS _S /df _S - SS _A /df _A)/2 = (14231/10-16595/20)/2 = 296,675				SD _{samp} = $\sqrt{V_{samp}}$ = 17,224		

Anmerkungen zur Tabelle A4.15.

1. Die Differenz d zwischen dem Mittelwert \bar{x} der zwei Werte $\left(\frac{x_{i1}+x_{i2}}{2}\right)$ und $\left(\frac{x_{i3}+x_{i4}}{2}\right)$ zu jedem der Werte ist gleich. Die Gleichung könnte also dargestellt werden als:

$$SS_{samp} = \sum_{i=1}^{10} 4 * d_i^2 = \sum_{i=1}^{10} \left[4 * \left(\frac{x_{i1} + x_{i2}}{2} - \bar{x}_i\right)^2 \right]$$

2. df_S = (NB*n-NB) = (10*2- 10) = 10, wobei NB die Anzahl der Chargen und n die Anzahl der für jede Charge analysierten Primärproben (= Laborproben) ist.

Beispiel A5: Enzyme in Hühnerfutter

Messgröße				Unsicherheitsschätzung		
Analyt/ Technik	Einheit ¹	Sektor/ Matrix	Probe- nahme- zielobjekt	Zweck	Design	Statistik
Enzyme/ HPLC	% m/m (i.e. Massen- anteil)	Lebens- mittel u. Futtermi- tel/Hühner- futter	25 kg Sack	Gesamt Unsicherheit (Schwach- stellen in der Messkette)	Modellie- rung mit der Stichproben- theorie (von Gy)	Summierung der Komponenten- abweichungen in relativen Einheiten

¹ Einschließlich Berichtsgrundlage

1 Anwendungsbereich

Es geht darum, die Unsicherheit der Probenahme mit der gegebenen Probenahmenvorschrift durch Anwendung der Gy'schen Stichprobentheorie (Abschnitt 10.2) abzuschätzen. Bei dem Analyten handelt es sich um einen dem Futter zugesetzten Enzymbestandteil. Die Probenahmetheorie liefert nur dann realistische Schätzungen, wenn alle Probenahme- und Probenteilungsvorgänge unter Beachtung der Regeln der korrekten Probenahme durchgeführt werden; in diesem Beispiel wird davon ausgegangen, dass keine groben Fehler vorliegen und dass "Fehler aus falscher Probenahme" vernachlässigbar sind.

2 Szenario und Probenahmezielobjekt

Ein Enzymprodukt wird als Zusatzstoff in Hühnerfutter (Dichte = $0,67 \text{ g cm}^{-3}$) verwendet. Die nominale Konzentration des Enzyms beträgt $0,05 \text{ % m/m}$. Das Enzympulver hat eine Dichte von $1,08 \text{ g cm}^{-3}$. Die Pulver werden sorgfältig gemischt. Die Größenverteilung der Enzympartikel war bekannt und es wurde geschätzt, dass die charakteristische Partikelgröße $d = 1,00 \text{ mm}$ und der Größenfaktor $g = 0,5$ beträgt. Zweck dieser Übung ist die Schätzung der Gesamtunsicherheit der Vorschrift (d. h. als fundamental sampling error, Abschnitt 10.2.7 und Abbildung 4), die für die Schätzung des durchschnittlichen Gehalts in jedem 25-kg-Sack für den Versand des Produkts an die Kunden verwendet wird.

3 Studiendesign unter Verwendung eines Modellierungsansatzes ("bottom-up")

Ein Modell wird mit Hilfe der Probenahmetheorie, wie in Abschnitt 10.2 beschrieben, erstellt. Die Parameter werden entweder direkt gemessen oder geschätzt, wobei davon ausgegangen wird, dass es sich um Einzelwerte handelt und diese innerhalb und zwischen den einzelnen Säcken konstant sind.

4 Probenahme und Analyse im Labor

Die tatsächliche Konzentration des Enzyms im Probenahmezielobjekt, das als 25-kg-Sack identifiziert wird, wird durch die Entnahme einer 500-g-Primärprobe aus diesem Beutel geschätzt.

Das Material der Primärprobe wird auf eine Partikelgröße von $<0,5$ mm gemahlen. Anschließend wird das Enzym mit einem geeigneten Lösungsmittel aus einer 2 g Prüfmengung extrahiert und die Konzentration mit Hilfe der Flüssigkeitschromatographie bestimmt. Die relative Standardabweichung der chromatographischen Messung, die anhand der Qualitätskontrolldaten des Labors geschätzt wird, beträgt 5 %.

5 Ergebnisse

Um den fundamental sampling error (FSE, Abschnitt 10.2.7, Abbildung 4) der beiden Probenahmeschritte abzuschätzen, müssen wir die Materialeigenschaften bewerten (Tabelle A5.1).

Tabelle A5.1: Eingabewerte für die Schätzung der Probenahmeunsicherheit durch den Modellierungsansatz, unter Verwendung der Probenahmetheorie

Primärprobe	Sekundärprobe	Anmerkung
$M_1 = 500$ g	$M_2 = 2,0$ g	Probenmenge
$M_{L1} = 25.000$ g	$M_{L2} = 500$ g	Gesamtmasse (Probenahmezielobjekt)
$d_1 = 0,1$ cm	$d_2 = 0,05$ cm	Partikelgrößen
$g_1 = 0,5$	$g_2 = 0,25$	Geschätzte Größenfaktoren
Beide Stichproben		
$a_L = 0,05$ % m/m		Mittlere Konzentration des Enzyms in der Menge
$\alpha = 100$ % m/m		Enzymkonzentration in Enzympartikeln
$\rho_c = 1,08$ g cm ⁻³		Dichte der Enzympartikel
$\rho_m = 0,67$ g cm ⁻³		Dichte der Matrixpartikel
$f = 0,5$		Standardformfaktor für sphäroidische Partikel
$\beta = 1$		Freisetzungsfaktor für freigesetzte Partikel

Diese Materialeigenschaften ergeben für den Strukturfaktor (Gleichung 8) den Wert $c = 2160$ g cm⁻³ und für die Probenahmekonstanten (Gleichung 7) C-Werte

$$C_1 = 540 \text{ g cm}^{-3} \text{ und } C_2 = 270 \text{ g cm}^{-3}$$

Gleichung 6 kann verwendet werden, um Schätzungen der Standardabweichung für jeden Probenahmeschritt (als Schätzungen der Standardunsicherheit) zu erhalten.

$$s_{r1} = 0,033 = 3.3 \%$$

Primärprobe

$$s_{r2} = 0,13 = 13 \%$$

Sekundärprobe

$$s_{r3} = 0,05 = 5 \%$$

Analytische Bestimmung

Die gesamte relative Standardabweichung (s_t , kombinierte Unsicherheit) kann nun durch Anwendung der Regel der Fehlerfortpflanzung geschätzt werden; für i Fehler gilt:

$$s_t = \sqrt{\sum s_{ri}^2} = 0,143 = 14,3 \%$$

Die relative erweiterte Unsicherheit mit einem Erweiterungsfaktor von 2 beträgt daher 28,6 % (ohne analytische Unsicherheiten im Zusammenhang mit systematischen Effekten, wie z. B. analytischer Bias).

6 Anmerkungen

Als größte Unsicherheitsquelle im gesamten Messprozess wird die Vorbereitung der Prüfmenge (2 g) für die Extraktion des Enzyms identifiziert.

Unsicherheiten im Zusammenhang mit systematischen Effekten während der Analyse wurden nicht zusätzlich berücksichtigt, und es wurde davon ausgegangen, dass Fehler aus falscher Probenahme (und Probenahme-Bias) vernachlässigbar sind.

7 Bewertung der Zwecktauglichkeit dieser Messungen

Wenn entschieden wird, dass die Gesamtunsicherheit von 28,6 % nicht angemessen ist (Abschnitt 16), dann muss der Schritt, in dem die Prüfmenge vorbereitet wird, geändert werden, um die Gesamtunsicherheit zu verringern. Entweder sollte eine größere Probe für die Extraktion verwendet oder die Primärprobe auf eine feinere Partikelgröße pulverisiert werden, je nachdem, was in der Praxis wirtschaftlicher ist. Das Modell kann auch verwendet werden, um entweder die Erhöhung der Masse oder die Verringerung der Partikelgröße vorherzusagen, die erforderlich ist, um die Unsicherheit zu erreichen, die als zwecktauglich angesehen wird (z. B. Anhang E).

8 Auswertung und Interpretation

Die Messung der für jeden 25-kg-Sack angegebenen Enzymkonzentration sollte mit einer Unsicherheit von 28,6 % des Konzentrationswerts verbunden sein. Die weitere Verwendung dieses Unsicherheitswerts hängt von der regelmäßigen Überprüfung der Gültigkeit der in seiner Berechnung verwendeten Werte und Annahmen ab.

9 Zusammenfassung

Messunsicherheit*		
Probenahme	Analyse	Gesamt
26,8 % (rel.)	10,0 % (rel.)	28,6 % (rel.)

*mit einem Erweiterungsfaktor von 2 (d. h. 95 % Vertrauensniveau)

Beispiel A6: Cadmium und Phosphor in landwirtschaftlichem Oberboden mit dem Modellansatz

Messgröße				Unsicherheitsschätzung		
Analyt/ Technik	Einheit	Sektor/ Matrix	Probenahmezielobjekt	Zweck	Design	Statistik
Cd: GF-ZAAS direkte Feststoffanalyse	mg kg ⁻¹ auf luftgetrockneter Basis	Umwelt/ landwirtschaftlicher Oberboden	Ackerboden – 143 x 22 m, Tiefe 30 cm	Gesamtmessunsicherheit (mit Beiträgen von allen Probenahmeeffekten)	Modellierungsansatz (unter Verwendung von Sondiermessungen für einzelne Effekte)	Summierung der Varianzen der Komponenten in relativen Einheiten
P: Ca-Acetat-Lactat (CAL)-Methode						

1 Anwendungsbereich

Schätzung der Gesamtmessunsicherheit durch Summierung der einzelnen Unsicherheitsbeiträge von Probenahme, Probenvorbereitung und Analyse unter Verwendung des Modellierungsansatzes.

2 Szenario und Probenahmezielobjekt

Ziel der Untersuchung ist die Schätzung der mittleren Cadmium- und Phosphorkonzentration im Oberboden einer Ackerfläche in der Größe von 0,32 ha (Angabe der Messgröße). Die Probenahme erfolgte anhand von Mischproben nach einer Vorschrift, die üblicherweise bei landwirtschaftlichen Kontrollen angewandt wird.

3 Probenahmenvorschrift

Das Zielgebiet wurde anhand eines geschichteten Probenahmeplans mit einer Beprobungsdichte von etwa 20 Einzelproben pro Hektar und einer Tiefe von 30 cm mit einem Erdbohrer beprobt. Die geschichtete Zufallsstichprobe ist in Anhang C2.3 definiert.

4 Studiendesign – Ursache-Wirkungs-Modellierungsansatz

4.1 Identifizierung von Auswirkungen bei der Messung

Die folgenden Quellen können im allgemeinen Fall als potenziell signifikante Beiträge zur Unsicherheit angesehen werden.

4.1.1 Probenahme

Die räumliche Verteilung des Analyten über ein zweidimensionales Objekt erzeugt zwei verschiedene Unsicherheitskomponenten "long range point selection error" (Anhang C2.3):

- Die *Probenahmevarianz* des Analytgehalts zwischen Mischproben von verschiedenen Stellen charakterisiert die "statistische Verteilung" des Analyten über das Zielgebiet.

Dieser Wert hängt häufig von der Entfernung zwischen den Probenahmepunkten/Probenahmestellen ab.

- Wenn das räumliche Muster des Analyten auf der Fläche nicht durch das Probenahmemuster repräsentiert wird (Probenahmestrategie), kann es zu einem Probenahmebias kommen.

Bei der Verwendung eines Probenahmewerkzeugs können verschiedene Effekte auftreten, wie z. B. Point Materialization Error (Abb. 3). Dieser kann durch eine schlecht definierte Bezugsebene des Bodens (z. B. aufgrund der Welligkeit der Bodenoberfläche oder Schwierigkeiten bei der Definition der Horizonte), durch Schwankungen in der tatsächlichen Probentiefe oder der Bodendichte (z. B. durch den Feuchtigkeitsgehalt) oder durch selektiven Verlust von Bodenmaterials aus der Probenahmeverrichtung auftreten.

Diese Effekte führen nur dann zu einem Unsicherheitsbeitrag, wenn es einen Tiefengradienten im Analytgehalt gibt (eine "dritte Dimension" des Zielkörpers). Aus diesem Grund werden diese Effekte, die einzeln nur schwer zu bestimmen sind, unter dem Begriff "*Tiefeneffekt*" zusammengefasst.

4.1.2 Probenvorbereitung

Die physikalische Probenvorbereitung umfasst den Schritt von der Feldprobe zur Laborprobe. *Mechanische Behandlung*, wie z. B. Auflockerung, Sieben, Mahlen und Teilen, reduzieren die Menge des Bodenmaterials. Bei diesen Arbeitsschritten können Fehler auftreten aufgrund von Variationen in Dauer und Kräfte der mechanischen Behandlung, Heterogenität, Entmischung verschiedener Boden (Partikel-)Fraktionen und der Partikelgrößenverteilung entstehen. Ein periodic point selection error (Abbildung 3) kann aufgrund von *Schwankungen des Feuchtigkeitsgehalts* der getrockneten Bodenprobe auftreten durch Sorption/Desorption von Wasser aus der Luft bis zu einem Gleichgewichtszustand (abhängig von der Feuchtigkeit und den Eigenschaften des Probenmaterials, z. B. der Partikelgröße).

4.1.3 Analyse

Die Analyse ist der dritte Schritt des Messprozesses, der mit verschiedenen Arten von Effekten verbunden ist, die zu Unsicherheitsbeiträgen führen. Die analytische Unsicherheit der Laborproben kann mittels bereits veröffentlichter Verfahren geschätzt werden [1, 40]. Die Aufteilung der Laborprobe in analytische Teilproben trägt zur Probenahmeunsicherheit bei; insbesondere kann ein weiterer "fundamentaler Fehler" auftreten. Die Zufallskomponente dieses Probenahmeeffekts ist jedoch in der analytischen Wiederholpräzision zwischen den Prüfproben enthalten. Eine erheblicher systematischer Beitrag sollte durch ordnungsgemäßes Mischen der pulverförmigen Proben vermieden werden.

4.2 Ursache-Wirkungs-Diagramm

Abbildung A6.1 zeigt das "Ursache-Wirkungs-Diagramm" für den Messprozess. Für die Schritte der Probenahme und Probenvorbereitung sind die Quellen der Unsicherheitsbeiträge angegeben; für die Analyse sind nur die analytischen Qualitätsparameter angegeben.

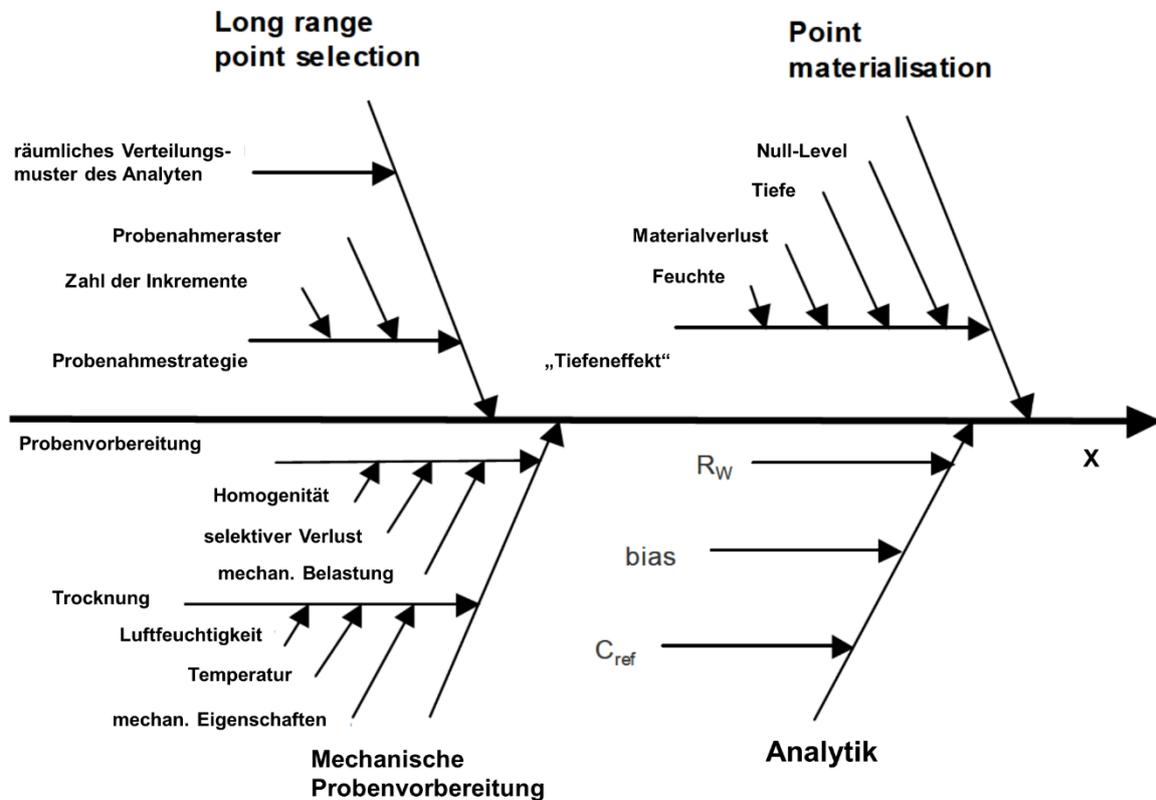


Abbildung A6.1: Ursache-Wirkungs-Diagramm für Bodenproben auf Ackerland (R_w ist die Reproduzierbarkeit innerhalb des Labors)

4.3 Modellgleichung

Die "Eingangsgrößen" der oben beschriebenen Probenahmeeffekte sind nicht Bestandteil der Gleichung, mit der das Messergebnis berechnet wird. Eine geeignete Modellgleichung für den gesamten Messprozess kann jedoch durch die Einführung entsprechender nominaler Korrekturfaktoren in das analytische Ergebnis aufgestellt werden:

$$x_{site} = \bar{x}_{analy} \times f_{b-loc} \times f_{strat} \times f_{depth} \times f_{prep} \times f_{dry}$$

mit

x_{site} = Messergebnis

\bar{x}_{analy} = Mittelwert aus der Analyse der Prüfproben

f_{b-loc} = Korrekturfaktor für die Abweichung 'zwischen Stellen'

f_{strat} = Korrekturfaktor für Bias aufgrund der Probenahmestrategie

f_{depth} = Korrekturfaktor für den 'Tiefeneffekt'

f_{prep} = Korrekturfaktor für Fehler während der mechanischen Probenvorbereitung

f_{dry} = Korrekturfaktor für die Abweichung des Feuchtigkeitsgehalts

Wird kein signifikanter Bias festgestellt, können alle Korrekturfaktoren auf 1 gesetzt werden, so dass die beste Schätzung für die Messgröße gegeben ist mit:

$$X_{site} = \bar{x}_{analy}$$

Aufgrund der Einfachheit der Modellgleichung (nur Faktoren) und angenommener Unabhängigkeit zwischen den Faktoren kann die kombinierte Unsicherheit durch Varianzaddition der relativen Standardunsicherheiten aus den verschiedenen Effekten berechnet werden:

$$u_{site} = \sqrt{u_{analy}^2 \times u_{b-loc}^2 \times u_{strat}^2 \times u_{depth}^2 \times u_{prep}^2 \times u_{dry}^2}$$

5 Probenahme und Analyse im Labor

Die Probenmasse wurde durch Kegeln und Vierteln reduziert, an der Luft getrocknet und gesiebt, um die Korngröße <2 mm zu selektieren.

Die Analyse wurde mit folgenden Methoden durchgeführt: für Cadmium mit Zeeman-GF-AAS ("direkte Feststoffanalyse") und für Phosphor mit der Calcium-Acetat-Lactat-Methode (CAL).

6 Ergebnisse der Bewertung der einzelnen Effekte in dieser Fallstudie

Die Schätzung der Standardunsicherheit aus der Verteilung der Analyten über das Zielgebiet basiert auf einer modifizierten inkrementellen Probenahme auf der Grundlage der Probenahmevervorschrift. Zur Klärung des Ergebnisses von Einzeleffekten wurden zusätzliche Sondierungsmessungen durchgeführt.

6.1 Abweichung 'zwischen Stellen'

Die Fläche wurde in neun Quadrate (A, B, C x 1, 2, 3) unterteilt, und aus jedem der fünf Quadrate (kreuzweise über die Fläche) werden drei Einzelproben entnommen. Die Einzelproben aus jedem Quadrat werden vermengt, so dass sich fünf separate Mischproben ergeben. Diese Proben werden getrennt behandelt und ausgewertet. Der Mittelwert der Einzelergebnisse stellt das Messergebnis dar, das mit der Spezifikation der Messgröße übereinstimmt.

Die Analyseergebnisse für beide untersuchten Analyten sind in Tabelle A6.1 aufgeführt. Die Standardabweichung dieser Werte (s_{sqr}) spiegelt die Variation zwischen den Mischproben für jedes nominierte Quadrat wider.

Die Standardunsicherheit des Gesamtmittelwerts (d. h. des Messergebnisses) aufgrund dieses Effekts kann geschätzt werden, indem die Anzahl der Proben "zwischen den Stellen" unter Verwendung des Standardfehlers des Mittelwerts berücksichtigt wird:

$$u_{b-loc} = \frac{s_{sqr}}{\sqrt{n_{b-loc}}}$$

Tabelle A6.1: Gemessene Cadmium- und Phosphorkonzentration in fünf Feldern

Quadrat	Cd mg kg ⁻¹	P mg kg ⁻¹
A1	0,270	124
A3	0,285	112
B2	0,343	120
C1	0,355	118
C3	0,343	105
\bar{x}_{sqr}	0,319	116
s_{sqr}	0,039 (12 %)	8,0 (6,5 %)
u_{b-loc}	5,4 %	2,9 %

Die Tabelle zeigt den Mittelwert über die fünf Quadrate (das Messergebnis), die aus diesen Werten berechnete Standardabweichung (s_{sqr}) und den geschätzten Unsicherheitsanteil aus dem Standardfehler des Mittelwertes (u_{b-loc}).

6.2 Probenahmestrategie

Die Untersuchung der Analytgehalte zwischen den Quadraten (Tabelle A6.1) zeigt für Phosphor keine nennenswerten Unterschiede in irgendeiner Richtung (weder vertikal noch horizontal noch diagonal). Daher ist für diesen Analyten aus dieser Quelle kein signifikanter Bias (z. B. $\leq 0,5$ %) des Messergebnisses zu erwarten.

Für Cadmium weisen die beiden A-Quadrate einen deutlich niedrigeren Gehalt des Analyten auf als die B- und C-Quadrate. Ein solcher Gradient war für dieses spezielle Gebiet nicht unerwartet, da die C-Quadrate an einer Waldgrenze liegen, während die A-Quadrate an Weideland grenzen und die Quadrate 1 und 3 zwischen anderen Ackerflächen liegen. Es ist bekannt, dass es im oberen Horizont von Waldböden zu einer Anreicherung von Schwermetallen kommt, die die angrenzenden Gebiete beeinflussen können.

Um nach einem solchen Effekt zu suchen, wurde ein "hypothesenbasiertes" Probenahmemuster angewandt. Bei den mit dieser Probenahmestrategie gemessenen Werten wurde jedoch nur ein geringer systematischer Effekt festgestellt. Daher wird für die Probenahmestrategie eine Standardunsicherheit von ≤ 1 % in das Unsicherheitsbudget für Cadmium aufgenommen.

6.3 'Tiefeneffekt'

Um die als "Tiefeneffekt" bezeichneten Effekte zu ermitteln, wurde das folgende Sondierungsexperiment durchgeführt.

Innerhalb der fünf "Testquadrate" werden Bohrkern in einer Tiefe von 35 cm entnommen. Von diesen Bohrkernen werden Segmente von 25 - 30 cm und von 30 - 35 cm abgetrennt und vermengt. Tabelle A6.2 zeigt die Analysenergebnisse für diese Proben.

Tabelle A6.2: Tiefenexperimente

	Cd mg kg ⁻¹	P mg kg ⁻¹
c ₋ (25 – 30 cm)	0,14	47
c ₊ (30 – 35 cm)	0,10	35
c ₋	0,34	124
c ₊	0,30	109
Δx	0,04	15
u _{depth}	3,5 %	3,7 %

Die Tabelle zeigt die durchschnittlichen Gehalte der Tiefenhorizonte aus fünf Bohrkernen von verschiedenen Stellen, die berechneten Grenzwerte und den geschätzten Unsicherheitsbeitrag

Beide Analyten weisen einen statistisch signifikanten negativen Gradienten in Bezug auf die Tiefe auf. Die Unsicherheit aufgrund der Tiefenwirkung wurde geschätzt, indem der Analytgehalt der Bodenschichten unterhalb und oberhalb der Referenztiefe (c₋, c₊) mit dem folgenden Modell berücksichtigt wurde.

Es wird angenommen, dass die maximale Abweichung in der Beprobungstiefe nicht mehr als 10 % beträgt (d. h. 27 bis 33 cm). Aus diesen Daten werden die unteren und oberen Grenzwerte (x₋, x₊), die sich auf den mittleren Gehalt eines Bohrkerns der Nenntiefe beziehen, nach folgender Formel geschätzt:

$$x_{-} = \frac{\bar{x} - 0,1 c_{-}}{0,9} \quad x_{+} = \frac{\bar{x} + 0,1 c_{+}}{1,1}$$

Die Differenz zwischen x₋ und x₊ (Δx_{depth}) wird als die maximale Abweichung vom mittleren Gehalt aufgrund der Tiefenvariation der Einzelproben angenommen. Wenn eine Rechteckverteilung für die Tiefenabweichung angenommen wird, kann die Standardunsicherheit des Mittelwertes (Tabelle A6.2) wie folgt geschätzt werden:

$$u_{depth} = \frac{\Delta x_{depth}/2}{\sqrt{3}}$$

6.4 Teilung

Die primären Feldproben wurden durch die Methode des Kegels und Viertels siebenmal zweigeteilt, was zu einer Laborprobe führte, die 1/64 der ursprünglichen Masse entsprach.

Um den "Teilungseffekt" zu ermitteln, wurde das folgende Sondierungsexperiment durchgeführt.

Beim ersten Teilungsschritt wurde die zweite Hälfte des Materials nicht verworfen, sondern als Doppelprobe betrachtet, die wie die ursprüngliche Probe behandelt und separat analysiert wurde. Tabelle A6.3 zeigt die relativen Standardabweichungen zwischen den Doppelbestimmungen jedes der fünf Quadrate für beide Analyten.

Als einfache Näherung wird der Mittelwert der relativen Standardabweichungen als Standardunsicherheit des Teilungsschritts angenommen

$$u_{split} = s_{split}$$

Anmerkung: Die beobachtete große Streuung der Standardabweichungen zwischen den Doppelbestimmungen muss erwartet werden. Die X^2 -Verteilung für $df = 1$ zeigt eine hohe Wahrscheinlichkeit für sehr niedrige Werte und eine moderate Wahrscheinlichkeit für große Werte.

Tabelle A6.3: Relative Standardabweichungen zwischen doppelten Teilproben und der Mittelwert dieser Standardabweichungen für beide Analyten

Quadrat	Cd mg kg ⁻¹	P mg kg ⁻¹
A1	0,44	1,49
A3	9,17	2,80
B2	5,32	0,84
C1	3,24	8,88
C3	0,44	1,81
\bar{s}_{split}	3,7	3,3

6.5 Trocknung

Für den Trocknungseffekt wurde kein Versuch durchgeführt, sondern Informationen aus der Literatur verwendet, um den Effekt abzuschätzen. Für eine große Anzahl von luftgetrockneten Bodenproben wurde ein Feuchtigkeitsgehalt zwischen 1 und 3 % ermittelt [62]. Gemäß der Probenahmenvorschrift bezieht sich die Messgröße auf luftgetrocknetes Bodenmaterial. Folglich ist für die Konzentrationsmessungen keine Korrektur für den Feuchtigkeitsgehalt erforderlich. Es muss jedoch ein Bereich von $\Delta x_{dry} = 2$ % Unterschied im Feuchtigkeitsgehalt berücksichtigt werden. Unter der Annahme einer Rechteckverteilung über diesen Bereich kann die Standardunsicherheit für beide Analyten wie folgt geschätzt werden:

$$u_{dry} = \frac{\Delta x_{dry}/2}{\sqrt{3}} = 0,6 \%$$

6.6 Analyse

Die Unsicherheit des Analyseverfahrens für Cadmium und Phosphor (Tabellen A6.4 und A6.5) wurde anhand von Qualitätskontrolldaten unter Verwendung des Nordtest-Ansatzes geschätzt [37].

Tabelle A6.4: Standardunsicherheitsbeiträge und kombinierte Unsicherheit bei der Analyse der Bodenprobe auf Cadmium

R_w	Unsicherheit der laborinternen Reproduzierbarkeit, bewertet anhand der Wiederholstandardabweichung des Mittelwerts von $n=10$ Prüfproben und der Stabilität des Geräts während eines Arbeitstages	$u_{Rw} = 3,6 \%$
C_{ref}	Unsicherheit des zertifizierten Wertes eines ZRM	$u_{ref} = 2,7 \%$
Bias	Kein Unsicherheitsanteil durch Laborbias, da die Ergebnisse um den täglichen Bias der ZRM-Messungen korrigiert werden	-
S_{bias}	Unsicherheitsbeitrag der Standardabweichung des Mittelwertes ($n=3$) aus der täglichen Analyse des ZRM	$u_{bias} = 2,7 \%$
	Kombinierte analytische Unsicherheit	$u_{analy} = 5,2 \%$

Tabelle A6.5: Standardunsicherheitsbeiträge und kombinierte Unsicherheit bei der Analyse der Bodenprobe auf Phosphor

R_w	Unsicherheit der laborinternen Reproduzierbarkeit, bewertet anhand der Wiederholstandardabweichung des Mittelwerts von $n=1$ Prüfprobe	$u_{Rw} = 1,7 \%$
C_{ref} Bias S_{bias}	Unsicherheit für die Richtigkeit der Ergebnisse, geschätzt aus der Vergleichpräzision s_R aus einem einzelnen Ringversuch (Worst-Case-Schätzung)	$u_{bias} = 9,5 \%$
	Kombinierte analytische Unsicherheit	$u_{analy} = 9,7 \%$

6.7 Unsicherheitsbudget und Messergebnis

Tabelle A6.6 listet die bewertete Standardunsicherheit aus den betrachteten Effekten auf. Die kombinierte Unsicherheit wird aus diesen Beiträgen berechnet.

Tabelle A6.6: Relative Standardunsicherheiten aus den betrachteten Effekten und die kombinierte Unsicherheit für beide Analyten

Effekt	Relative Standardunsicherheit (%)	
	Cd	P
Abweichung 'zwischen Stellen'	5,4	2,9
Probenahmestrategie	1,0	0,5
Tiefe	3,5	3,7
Teilung	3,7	3,3
Trocknung	0,6	0,6
Analyse	5,2	9,7
Kombinierte Unsicherheit	9,1	11,3

Messergebnis:

Cd: 0,32 ± 0,06 mg kg⁻¹

P: 116 ± 26 mg kg⁻¹

(Erweiterungsfaktor 2 für ca. 95 % Vertrauensniveau)

7 Anmerkungen

7.1 Beiträge der Effekte

Tabelle A6.6 zeigt, dass der Prozess der Probenahme/Probenvorbereitung erheblich zur Gesamtmessunsicherheit beiträgt. Um die Relevanz einzelner Effekte/Prozessschritte zu erkennen und zu bewerten, müssen mehrere Aspekte berücksichtigt werden:

7.1.1 Der "*Zwischen-Stellen-Effekt*" hängt von der Homogenität des Zielgebiets und der Gesamtzahl der Einzelproben ab, die von jedem Quadrat genommen werden. Frühere Untersuchungen zeigen, dass 20 Einzelproben pro Hektar Ackerland einen Unsicherheitsanteil in der Größenordnung der analytischen Unsicherheit liefern.

7.1.2 Der *auf die Probenahmestrategie zurückzuführende Fehler* lässt sich nur schwer quantifizieren, kann aber oft viel größer sein als der in dieser Fallstudie beobachtete. In der Praxis kann er nur durch eine "Expertenbeurteilung" der großräumigen Verteilung des Analyten über das Gebiet und die Wahl einer geeigneten Probenahmestrategie kontrolliert werden.

7.1.3 Bei der Modellberechnung des *Tiefeneffekts* wird dieser als unbekannter systematischer Fehler behandelt, d.h. die Abweichung in der Tiefe tritt bei allen Einzelproben (mehr oder weniger) in der gleichen Richtung auf. Dies kann unter bestimmten Bedingungen realistisch sein; zum Beispiel neigt ein trockener Sandboden dazu, am unteren Ende des Bohrers abzufallen, so dass die durchschnittliche Einzelprobentiefe zu gering wäre. Wird ein solcher Effekt festgestellt, so ist die Korrektur der systematischen Abweichung möglich und es muss

nur die zufällige Fehlerkomponente berücksichtigt werden (d. h. die Unsicherheit nimmt mit dem Faktor $1/\sqrt{n_{\text{incr}}}$ ab). Ein Training des Probennehmers kann diesen "point materialisation error" verringern.

7.1.4 Der *Teilungseffekt* ist schwer zu kontrollieren, da die erste Massenreduzierung oft vor Ort vorgenommen wird. Er kann erheblich beitragen, wenn die Methode der Massenreduzierung ungeeignet ist oder nachlässig durchgeführt wird. Daher ist die Schulung des Probenahmepersonals von großer Bedeutung.

7.1.5 Der *Einfluss des Feuchtigkeitsgehalts* von luftgetrockneten Bodenproben scheint in diesem Fall vernachlässigbar zu sein.

7.1.6 Die Unsicherheit des *Analyseverfahrens* kann den dominierenden Anteil an der kombinierten Messunsicherheit ausmachen (z.B. bei Cadmium). Sie kann kontrolliert werden, wenn die Standardmethoden der analytischen Qualitätssicherung eingehalten werden (z. B. regelmäßige Verwendung von ZRM und Teilnahme an Ringversuchen). Die Unsicherheit aus dieser Quelle kann vorherrschend sein, wenn die Konzentration des Analyten nahe der analytischen Nachweisgrenze liegt.

7.1.7 Zu den Effekten, die in dieser Fallstudie nicht berücksichtigt wurden, gehören die Dauer und das Ausmaß der Kräfte innerhalb des Mahl- und Siebprozesses sowie die Nässe des Bodenkörpers während der Probenahme. Der Einfluss dieser Effekte wurde als nicht signifikant erachtet, obwohl diese Annahmen überprüft werden sollten.

8 Bewertung der Zwecktauglichkeit dieser Messungen

Für eine Routinemessung gemäß der Probenahmenvorschrift muss eine Mischprobe aus etwa 10 Einzelproben zweifach analysiert werden.

In dieser Fallstudie werden zur Schätzung der Unsicherheitsbeiträge von Einzeleffekten 10 zusätzliche Einzelproben genommen und insgesamt 20 (Misch-)Proben vorbereitet und analysiert.

Dieser zusätzliche Aufwand und diese Kosten sind für Routinemessungen nicht angemessen. Wenn jedoch Messungen auf Ackerland die Hauptuntersuchungsart des Labors sind, könnte eine solche Sondierungsuntersuchung wertvoll sein, um einen typischen Wert der "Probenahme"-Komponente für diese Messungen zu erhalten. Darüber hinaus ist eine Bewertung der Fehlerkomponenten (d. h. des Unsicherheitsbudgets) auch für die Optimierung des Messverfahrens nützlich.

9 Berichterstattung und Interpretation

Die Messungen der mittleren Konzentration für dieses Gebiet des Oberbodens weisen erweiterte Unsicherheitswerte auf, die für Cadmium mit $0,06 \text{ mg kg}^{-1}$ oder 18,2 % des Konzentrationswertes und für Phosphor mit 26 mg kg^{-1} oder 22,6 % angegeben werden können.

10 Zusammenfassung

	Messunsicherheit*		
Analyt	Probenahme	Analytisch	Insgesamt
Cd	15,0 %	10,4 %	18,2 %
P	11,6 %	19,4 %	22,6 %

* mit einem Erweiterungsfaktor von 2 (d.h. für 95 % Vertrauen)

Anhang B: Terminologie

Genauigkeit	<p>Grad der Übereinstimmung zwischen einem Prüf- oder Messergebnis und dem wahren Wert</p> <p><i>Anmerkungen:</i></p> <p>1 In der Praxis wird der wahre Wert durch den akzeptierten Referenzwert ersetzt.</p> <p>2 Der Begriff "Genauigkeit" beinhaltet eine Kombination von Zufallskomponenten und einer gemeinsamen systematischen Fehler- oder Bias-Komponente, wenn er auf eine Reihe von Prüf- oder Messergebnissen angewandt wird.</p> <p>3 Genauigkeit bezieht sich auf eine Kombination aus Richtigkeit und Präzision (3.3.4)</p> <p>ISO 3534-2: (2006) [63]</p>
Bias	<p>Differenz zwischen dem erwarteten und dem tatsächlichen Wert eines Prüf- oder Messergebnisses</p> <p><i>Anmerkungen:</i></p> <p>1 Der Bias ist der gesamte systematische Fehler im Gegensatz zum zufälligen Fehler. Es kann eine oder mehrere systematische Fehlerkomponenten geben, die zum Bias beitragen. Eine größere systematische Abweichung vom wahren Wert spiegelt sich in einem größeren Bias-Wert wider.</p> <p>2 Der Bias eines Messgeräts wird in der Regel durch Mittelung des Anzeigefehlers über eine angemessene Anzahl von Messwiederholungen geschätzt. Der Anzeigefehler ist die: "Anzeige eines Messgeräts abzüglich eines wahren Werts der entsprechenden Eingangsgröße".</p> <p>3 In der Praxis wird der wahre Wert durch den akzeptierten Referenzwert ersetzt.</p> <p>ISO 3534-2: (2006) [63]</p>
Mischprobe (auch Durchschnitt und Aggregat)	<p>Zwei oder mehr Einzelproben/Teilproben, die in geeigneten Portionen entweder einzeln oder kontinuierlich gemischt werden (verschnittene Mischprobe) und aus denen der Durchschnittswert eines gewünschten Merkmals ermittelt werden kann.</p> <p>AMC (2005) [64]</p>
Doppelprobe Replikatprobe	<p>Eine der beiden (oder mehreren*) Proben oder Teilproben, die zur gleichen Zeit durch dasselbe Probenahmeverfahren oder Teilprobenahmeverfahren getrennt entnommen wurden.</p> <p>*für die Wiederholungsprobe</p> <p><i>Anmerkung: Jede Doppelprobe wird an einem separaten "Probenahmepunkt" innerhalb der "Probenahmestelle" entnommen.</i></p>
Ergebnisfehler	<p>Prüfergebnis oder Messergebnis abzüglich des wahren Wertes</p> <p><i>Anmerkung:</i></p>

	<p><i>1 In der Praxis wird der wahre Wert durch den vereinbarten Referenzwert ersetzt.</i></p> <p><i>2 Ergebnisfehler ist die Summe aus zufälligen und systematischen Fehlern.</i></p> <p>Basierend auf ISO 3534-2: (2006) [63] (Anmerkung entfällt)</p>
Zwecktauglichkeit	<p>Das Ausmaß, in dem die durch ein Messverfahren erzeugten Daten einen Benutzer in die Lage versetzen, technisch und administrativ korrekte Entscheidungen für einen bestimmten Zweck zu treffen.</p> <p><i>Anmerkung: Hier für die analytische Wissenschaft definiert.</i></p> <p>Thompson and Ramsey (1995) [16]</p>
Homogenität Heterogenität	<p>Grad, zu dem eine Eigenschaft oder ein Bestandteil gleichmäßig über eine Materialmenge verteilt ist.</p> <p><i>Anmerkung:</i></p> <p><i>1 Ein Material kann in Bezug auf einen Analyten oder eine Eigenschaft homogen, in Bezug auf eine andere jedoch heterogen sein.</i></p> <p><i>2 Der Grad der Heterogenität (das Gegenteil von Homogenität) ist der entscheidende Faktor für den Probenahmefehler.</i></p> <p>IUPAC (1990) [9]</p>
Einzelprobe	<p>Einzelne Materialportion, die durch einen einzigen Vorgang eines Probenahmegeräts gesammelt wird.</p> <p>IUPAC (1990) [9], AMC (2005) [64]</p>
Laborprobe	<p>Probe, wie sie für den Versand an das Labor vorbereitet und für die Kontrolle oder Prüfung bestimmt ist.</p> <p>ISO 78-2 (1999) [65]</p>
Messgröße	<p>Die zu messende Größe</p> <p>JCGM 200:2012 [43]</p>
Messunsicherheit	<p><i>siehe Unsicherheit der Messung</i></p>
Präzision	<p>Der Grad der Übereinstimmung zwischen unabhängigen Test-/Messergebnissen, die unter festgelegten Bedingungen erzielt wurden.</p> <p><i>Anmerkungen:</i></p> <p><i>1 Die Präzision hängt nur von der Verteilung der zufälligen Fehler ab und bezieht sich nicht auf den wahren Wert oder den vorgegebenen Wert.</i></p> <p><i>2 Das Maß für die Präzision wird in der Regel als Unpräzision ausgedrückt und als Standardabweichung der Prüf- oder Messergebnisse berechnet. Eine geringere Präzision spiegelt sich in einer größeren Standardabweichung wider.</i></p> <p><i>3 Quantitative Präzisionsmessungen hängen entscheidend von den festgelegten Bedingungen ab. Wiederholbarkeits- und</i></p>

Reproduzierbarkeitsbedingungen sind besondere Umstände von extrem festgelegten Bedingungen.

ISO 3534-2: (2006) [63]

Primärprobe Die Sammlung von einer oder mehreren Einzelproben oder Einheiten, die ursprünglich einer Grundgesamtheit entnommen wurden.

Anmerkung: Der Begriff primär bezieht sich in diesem Fall nicht auf die Qualität der Probe, sondern auf die Tatsache, dass die Probe in der frühesten Phase der Messung entnommen wurde.

IUPAC (1990) [9], AMC (2005) [64]

Zufälliger Ergebnisfehler Komponente des Ergebnisfehlers, die im Verlauf einer Reihe von Prüf- oder Messergebnissen für dasselbe Merkmal oder dieselbe Größe in nicht vorhersehbarer Weise schwankt.

Anmerkung: Es ist nicht möglich, den Zufallsfehler zu korrigieren.

ISO 3534-2: (2006) [63]

Zufallsprobe Durch Zufallsprobenahme ausgewählte Probe

ISO 3534-2: (2006) [63] (Anmerkung entfällt)

Zufallsprobenahme Probenahme, bei der eine Probe aus n Probenahmeeinheiten einer Grundgesamtheit in der Weise entnommen wird, dass alle möglichen Kombinationen von n Probenahmeeinheiten mit einer bestimmten Wahrscheinlichkeit entnommen werden.

ISO 3534-2: (2006) [63]

Einfache Zufallsprobenahme Probenahme, bei der aus einer Grundgesamtheit eine Probe aus n Probenahmeeinheiten so entnommen wird, dass alle möglichen Kombinationen von n Probenahmeeinheiten mit gleicher Wahrscheinlichkeit entnommen werden.

Anmerkung: Handelt es sich bei der Probenahmeeinheit bei einer Beprobung aus einer größeren Menge um eine Einzelprobe, so werden die Positionierung, die Abgrenzung und die Entnahme der Einzelproben so vorgenommen, dass alle Probenahmeeinheiten mit gleicher Wahrscheinlichkeit ausgewählt werden.

ISO 3534-2: (2006) [63]

Referenzprobenahme Eingehende Charakterisierung eines Gebiets mit einem einzigen Probenahmegerät und einem einzigen Labor, das die Erstellung eines Verteilungsmodells ermöglicht, um die Elementkonzentrationen mit bekannter Unsicherheit an jedem Probenahmepunkt vorherzusagen.

IUPAC (2005) [66]

Referenz-Probenahmezielobjekt Die Entsprechung bei der Probenahme zu einem Referenzmaterial oder zertifiziertem Referenzmaterials (in der chemischen Analyse).

Anmerkung: Ein Probenahmezielobjekt, bei dem die Konzentrationen eines oder mehrerer Elemente in Bezug auf die räumliche/zeitliche Variabilität gut charakterisiert sind. Die Entsprechung bei der Probenahme zu einem Referenzmaterial oder einem zertifizierten Referenzmaterial (in der chemischen Analyse) (Anmerkungen angepasst an den Empfehlungsentwurf der IUPAC (2003); ursprünglich definiert in ISO Guide 30: 1992).

Thompson and Ramsey (1995) [16]

Repräsentative Stichprobe

Stichprobe, die sich aus einem Probenahmeplan ergibt, von dem erwartet werden kann, dass er die Eigenschaften der Grundgesamtheit, die von Interesse sind, angemessen widerspiegelt.

IUPAC (1990) [9], AMC (2005) [64]

Stichprobe

Ein Teil des Materials, der aus einer größeren Materialmenge ausgewählt wird.

IUPAC (1990) [9], AMC (2005) [64]

Probenvorbereitung

Gesamtheit der Vorgänge, die erforderlich sind, um eine Sammel- oder Gesamtprobe in eine Labor- oder Prüfprobe umzuwandeln.

BEISPIEL Verringerung der Partikelgröße, Mischen und Teilen einer Probe.

Anmerkungen:

1 Bei partikelförmigen Materialien definiert der Abschluss jedes Arbeitsgangs der Probenteilung den Beginn der nächsten Probenvorbereitungsstufe.

2 Die Probenvorbereitung sollte die Fähigkeit der Probe, die Grundgesamtheit, aus der sie entnommen wurde, zu repräsentieren, so weit wie möglich nicht verändern.

Basierend auf ISO 3534-2: (2006) [63] (der Wortlaut wurde geändert, um Sammel- oder Gesamtproben und die Relevanz für chemische Messungen einzubeziehen)

Probenvorbehandlung

Sammelbegriff für alle Verfahren, die dazu dienen, eine Probe in einen definierten Zustand zu bringen, der eine spätere Untersuchung oder Analyse oder eine langfristige Lagerung ermöglicht.

Anmerkung: Die Vorbehandlung der Proben umfasst z. B. Mischen, Teilen, Trocknen, Zerkleinern und Stabilisieren.

Basierend auf ISO 11074:2015 [67] (Beschränkung auf Bodenproben aufgehoben)

Probengröße

Anzahl der Objekte oder die Menge des Materials, aus dem eine Probe besteht.

ISO 11074:2015 [67], ISO 7002: A.40 (1986) [68]

Probenehmer Probenahmepersonal	<p>Person oder Gruppe von Personen, die die Probenahme am Ort der Probenahme durchführt.</p> <p>Anmerkung 1: Werkzeuge und andere Geräte zur Probenahme werden manchmal auch als "Probenehmer" bezeichnet. Diese sollten als "Probenahmegeräte" oder "Probenahmeausrüstung" bezeichnet werden.</p> <p>ISO 11074:2015 [67]</p>
Probenahme	<p>Entnehmen oder Erzeugen einer Probe.</p> <p>ISO 3534-2: (2006) [63]</p>
Probenahme-Bias	<p>Der Teil des gesamten Mess-Bias, der auf die Probenahme zurückzuführen ist.</p> <p>AMC (2005) [64]</p>
Probenahmestelle	<p>Der Ort, an dem die Probenahme innerhalb des Probenahmezielobjekts erfolgt. Kann für die Stelle verwendet werden, an der Duplikatproben (oder Doppelproben) an bestimmten Probenahmepunkten genommen werden.</p>
Probenahmeplan	<p>Vorgegebenes Verfahren für die Auswahl, die Rücknahme, Konservierung, Transport und Aufbereitung der aus einer Grundgesamtheit als Stichprobe zu entnehmenden Menge.</p> <p>UPAC (1900) [9], AMC (2005) [64]</p>
Probenahmepunkt	<p>Der Ort, an dem die Probenahme innerhalb der Probenahmestelle erfolgt. Kann für einen bestimmten Punkt innerhalb einer Probenahmestelle verwendet werden, an dem eine Doppel- (oder Replikat-) Probe genommen wird.</p> <p><i>Anmerkung: Die Genauigkeit, mit der ein Probenahmepunkt räumlich oder zeitlich lokalisiert wird, hängt von der Vermessungsmethode ab. Von Probenahmepunkten, die diese Genauigkeit widerspiegeln, werden Doppelproben genommen.</i></p>
Probenahmepräzision	<p>Der Teil der Gesamtmesspräzision, der auf die Probenahme zurückzuführen ist.</p> <p>AMC (2005) [64]</p>
Probenahmeverfahren	<p>Betriebliche Anforderungen und/oder Anweisungen für die Anwendung eines bestimmten Probenahmeplans (d. h. die Anweisungen für die Anwendung des Plans)</p> <p>AMC (2005) [64]</p>
Probenahmezielobjekt	<p>Anteil des Materials zu einem bestimmten Zeitpunkt, den die Probe repräsentieren soll.</p> <p><i>Anmerkungen:</i></p> <p><i>1 Das Probenahmezielobjekt sollte vor der Erstellung des Probenahmeplans festgelegt werden.</i></p>

	<p>2 Das Probenahmezielobjekt kann durch Vorschriften festgelegt werden (z. B. Losgröße).</p> <p>3 Wenn die Eigenschaften und Merkmale (z. B. die chemische Zusammensetzung) eines bestimmten Gebiets oder eines bestimmten Zeitraums von Interesse sind und bekannt sein müssen, kann es als Probenahmezielobjekt betrachtet werden.</p>
Probenahme-unsicherheit	Siehe Unsicherheit der Probenahme
Teilprobe	<p>Ausgewählter Teil einer Probe</p> <p>Anmerkung: Die Teilprobe kann, muss aber nicht, nach der gleichen Methode ausgewählt werden wie die ursprüngliche Probe.</p> <p>ISO 3534-2: (2006) [63] ("Teilprobe")</p>
Probenteilung	<p>Prozess der Auswahl einer oder mehrerer Teilproben aus einer Stichprobe einer Grundgesamtheit.</p> <p>ISO 11074:2015 [67]</p>
Systematischer Ergebnisfehler	<p>Komponente des Ergebnisfehlers, die im Verlauf einer Reihe von Prüfergebnissen oder Messergebnissen für dasselbe Merkmal oder dieselbe Größe konstant bleibt oder in vorhersehbarer Weise variiert.</p> <p>Anmerkung: Systematische Fehler und ihre Ursachen können bekannt oder unbekannt sein.</p> <p>ISO 3534-2: (2006) [63]</p>
Systematische Probenahme	<p>Probenahme nach einem systematischen Plan.</p> <p>Basierend auf ISO 3534-2: (2006) [63] (Anmerkung entfällt)</p>
Prüfmenge	<p>Aus der Probe entnommene Materialmenge, die eine für die Messung der Konzentration oder einer anderen Eigenschaft, die von Interesse ist, geeignete Größe aufweist.</p> <p>IUPAC (1990) [9], AMC (2005) [64]</p>
Prüfprobe	<p>Aus der Laborprobe aufbereitete Probe, aus der die zu prüfende Prüfmenge für die Prüfung oder Analyse entnommen werden.</p> <p>IUPAC (1990) [9], ISO 11074:2015 [67], AMC (2005) [64]</p>
Richtigkeit	<p>Grad der Übereinstimmung zwischen dem erwarteten Ergebnis eines Tests oder einer Messung und dem wahren Wert.</p> <p>Anmerkungen:</p> <p>1 Das Maß für die Richtigkeit wird in der Regel in Form von Bias ausgedrückt.</p> <p>2 Die Richtigkeit wird manchmal auch als "Genauigkeit des Mittelwerts" bezeichnet. Diese Verwendung wird nicht empfohlen.</p>

	<p>3 <i>In der Praxis wird der wahre Wert durch den akzeptierten Referenzwert ersetzt.</i></p> <p>ISO 3534-2: (2006) [63]</p>
Unsicherheit (Messunsicherheit)	<p>Dem Messergebnis zugeordneter Parameter, der die Streuung der Werte kennzeichnet, die sinnvollerweise der Messgröße zugeordnet werden kann.</p> <p>Anmerkungen:</p> <p>1 <i>Der Parameter kann z. B. eine Standardabweichung (oder ein bestimmtes Vielfaches davon) oder die halbe Breite eines Intervalls mit einem bestimmten Konfidenzniveau sein.</i></p> <p>2 <i>Die Messunsicherheit besteht im Allgemeinen aus vielen Komponenten. Einige dieser Komponenten lassen sich anhand der statistischen Verteilung der Ergebnisse von Messreihen abschätzen und können durch experimentelle Standardabweichungen charakterisiert werden. Die anderen Komponenten, die ebenfalls durch Standardabweichungen charakterisiert werden können, werden aus angenommenen Wahrscheinlichkeitsverteilungen auf der Grundlage von Erfahrungen oder anderen Informationen abgeschätzt.</i></p> <p>3 <i>Es wird davon ausgegangen, dass das Messergebnis die beste Schätzung des Wertes der Messgröße ist und dass alle Unsicherheitskomponenten, einschließlich derjenigen, die sich aus systematischen Effekten ergeben, wie z. B. Komponenten im Zusammenhang mit Korrekturen und Referenzstandards, zur Streuung beitragen.</i></p> <p>4 <i>(ergänzt) Wenn die Messgröße als Größe innerhalb des Probenahmezielobjekts definiert ist, dann ist die Unsicherheit der Probenahme in der Messunsicherheit enthalten.</i></p> <p>JCGM 100 (2008) / ISO/IEC Guide 98-3:2008 [2]</p>
Unsicherheitsfaktor	<p>Der Faktor, mit dem der Messwert multipliziert und dividiert wird, um die Grenzen eines Unsicherheitsintervalls zu ermitteln.</p> <p>Ramsey and Ellison (2015) [23].</p>
Unsicherheit aus der Probenahme	<p>Der Teil der gesamten Messunsicherheit, der auf die Probenahme zurückzuführen ist.</p> <p><i>Anmerkung: Auch Probenahmeunsicherheit genannt</i></p> <p>IUPAC (2005) [66]</p>

Anhang C: Nützliche statistische Verfahren

C.1 Schätzung des Bias zwischen zwei Probenahmeverfahren unter Verwendung gepaarter Proben

Bei der Methode der gepaarten Proben wird nach beiden betrachteten Probenahmeverfahren jeweils eine Probe einer Anzahl (vorzugsweise $n > 20$) von Zielobjekten entnommen.

Die Methode eignet sich besonders für den Vergleich einer neuen zu prüfenden Vorschrift mit einer routinemäßig verwendeten Vorschrift, ist aber auch allgemein anwendbar. Bei beiden Verfahren muss das Probenahmeverfahren in irgendeiner Weise randomisiert werden, indem beispielsweise die Sammlung von Einzelproben an einer zufälligen Position innerhalb des Zielobjekts beginnt und das Raster in einer zufälligen Richtung ausgerichtet wird. Die entnommenen Proben werden unter randomisierten Wiederholbedingungen analysiert, so dass ein analytischer Bias eliminiert wird.

Das unten in Abbildung C1.1 dargestellte Design gewährleistet ein Minimum an zusätzlicher Arbeit an jedem Zielobjekt, so dass das Experiment mit geringen Kosten durchgeführt werden kann, ohne den Fluss der Routineprobenahme zu unterbrechen. Das Ergebnis ist auch robust, da es aus Daten abgeleitet ist, die an vielen typischen, aber unterschiedlichen Zielobjekten gesammelt wurden. Es stellt daher den durchschnittlichen Bias zwischen den Ergebnissen der beiden Vorschriften und nicht den Bias bei einem einzelnen Zielobjekt dar, das sich als untypisch erweisen könnte.

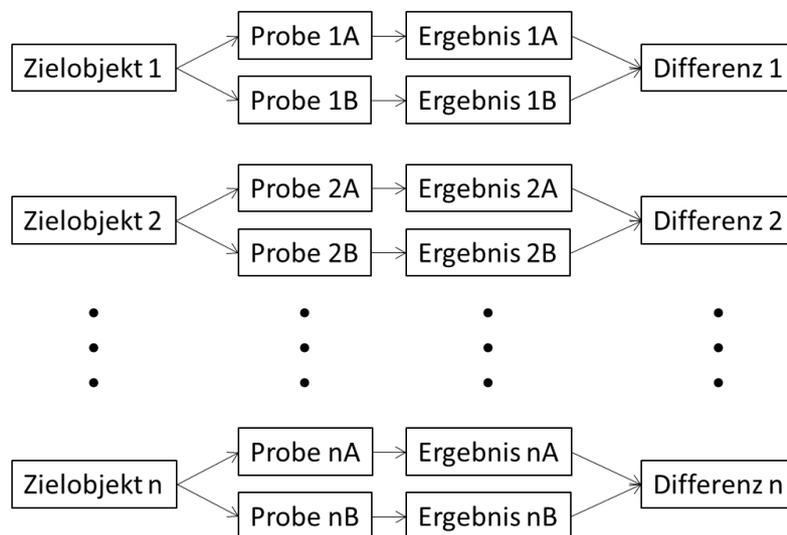


Abbildung C1.1: Design eines Experiments zur Schätzung des Bias zwischen zwei Probenahmeverfahren

Design eines Experiments zur Schätzung des Bias zwischen zwei Probenahmeverfahren A und B durch Entnahme von gepaarten Proben an jedem Zielobjekt.

In der ersten Stufe der Untersuchung der Ergebnisse wird geprüft, ob die gepaarten Unterschiede von der Konzentration des Analyten abhängen. Dies ist besonders wahrscheinlich, wenn der Konzentrationsbereich bei aufeinanderfolgenden Zielobjekten groß ist. Ein Streudiagramm bietet eine nützliche visuelle Überprüfung. Liegt keine Abhängigkeit vor, ist der Schätzwert für den Bias der Mittelwert der vorzeichenbehafteten gepaarten Differenzen, und dieser Mittelwert kann in der üblichen Weise auf eine signifikante Abweichung von Null getestet werden. In dem in Abbildung C1.2 gezeigten Beispiel gibt es offensichtlich keine signifikante Abhängigkeit zwischen der vorzeichenbehafteten Differenz und der Konzentration, und der Bias zwischen den Verfahren ist, gemäß eines 2-Stichproben-t-Test mit einem Vertrauensniveau von 95 %, nicht signifikant von Null verschieden. Liegt ein eindeutiger konzentrationsabhängiger Bias vor, wie in Abbildung C1.3, so sollte der Bias als Funktion der Konzentration ausgedrückt werden. In dem dargestellten Fall wurde (ermittelt durch die „funktionale Zusammenhangsmethode“ [59]) ein signifikanter Rotationsbias nachgewiesen, mit einem Trend, der durch die Gleichung Ergebnis (B) = Ergebnis (A) x 1,2 ausgedrückt wird.

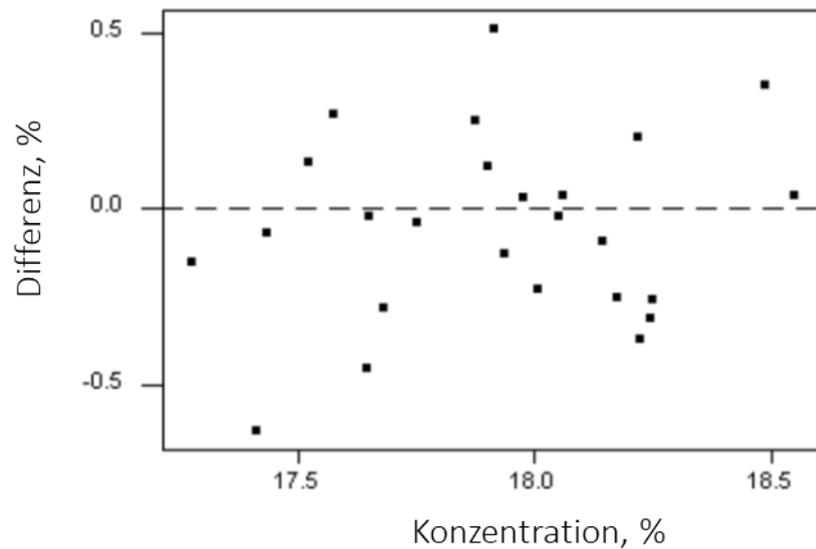


Abbildung C1.2: Kein signifikanter Bias oder Trend

Differenzen zwischen den Ergebnissen aus zwei Probenahmevorschriften angewandt auf 25 Zielobjekte als Funktion der Konzentration. Es gibt keinen signifikanten Bias und kein Anzeichen für die Abhängigkeit des Bias von der Konzentration.

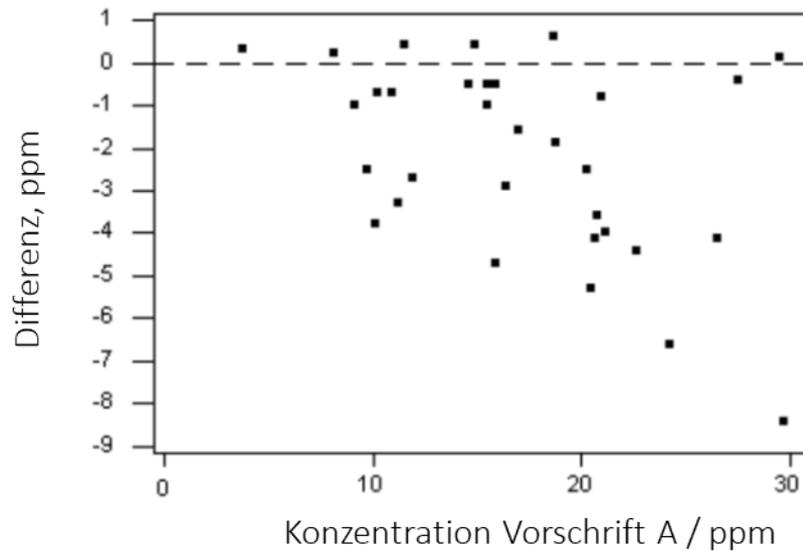


Abbildung C1.2: Signifikanter Bias und Trend

Differenzen zwischen den Ergebnissen aus zwei Probenahmeverfahren angewandt auf 33 Zielobjekte dargestellt als Funktion der Konzentration. Es gibt einen signifikanten Bias (da 27/33 der Ergebnisse negativ sind) und der Betrag des Bias nimmt mit steigender Konzentration zu.

C.2 Weitere Beschreibung des Probennahmefehlers durch die Probenahmetheorie

C.2.1 Der Weighting error (SWE; Gewichtungsfehler) stellt eine eigene Klasse dar. Er entsteht beispielsweise, wenn der Batch (Probenahmezielobjekt) aus Teilbatches unterschiedlicher Größe besteht, aber die mittlere Konzentration als einfacher Mittelwert geschätzt wird, ohne die Größe der Teilbatches zu berücksichtigen. Die korrekte Methode besteht darin, den gewichteten Mittelwert zu berechnen, indem die Größe der Teilbatches als Gewichtung verwendet wird. Bei der Analyse von bewegtem Material entsteht ein Gewichtungsfehler, wenn die Durchflussmenge schwankt, aber bei der Berechnung des Mittelwerts nicht berücksichtigt wird; in diesem Fall sollten die Durchflussmengen gleichzeitig mit der Probenahme erfasst und bei der Berechnung des Mittelwerts als Gewichtung verwendet werden. Eine andere Möglichkeit ist, ein Probenahmegerät zu verwenden, das Proben entnimmt, deren Größe proportional zur Durchflussmenge ist, und das damit die Probengröße als Gewichtung bei der Berechnung des Mittelwerts verwendet. Es ist zu beachten, dass, wenn eine zusammengesetzte Probe aus Teilproben gebildet wird, eine proportionale Probe verwendet werden sollte, da sonst ein Gewichtungsfehler in der zusammengesetzten Probe entsteht.

C.2.2 Der Grouping and segregation error (GSE) ist der zweite Fehlerterm aus den Fehlern mit kurzer Reichweite. Er wird durch die Tatsache verursacht, dass die Probe normalerweise nicht Fragment für Fragment, sondern als eine Gruppe von Fragmenten genommen wird. Liegt eine Entmischung des Materials vor, führt dies zu dieser Art von Fehler. Dieser Fehler wird

normalerweise nicht geschätzt. Gy hat jedoch gezeigt, dass der GSE bei korrekter Probenahme kleiner oder maximal gleich dem fundamental sampling error (FSE) ist.

C2.3 Point selection error (PSE). Wenn der Mittelwert eines kontinuierlichen Objekts (z. B. Prozess, Strom, Fluss, verschmutzter Standort, ...) mit Hilfe diskreter Stichproben geschätzt wird, hängt die Unsicherheit des Mittelwertes von der Probenahmestrategie ab, da die Ergebnisse in der Regel autokorreliert sind. Dieser Fehler wird als Point selection error (PSE) bezeichnet und hängt von der Probenahmestrategie ab. Drei grundlegende Strategien können für die Auswahl der Probenahme angewandt werden (siehe Abbildung C2.1):

- 1) **Zufallsprobenahme:** Zeit oder Ort der N Probenahmepunkte werden zufällig über das Zielobjekt verteilt.
- 2) **Geschichtete (Zufalls-)Probenahme:** Die Partie wird zunächst in N Teilpartien unterteilt und innerhalb jeder Teilpartie werden die Probenahmepunkte nach dem Zufallsprinzip zugewiesen.
- 3) **Systematische (geschichtete) Probenahme:** Alle N Proben werden in gleichen Abständen entnommen (eindimensionaler Fall) oder auf einem festen symmetrischen Raster (für Zieleobjekte, die aus Sicht der Probenahme zwei oder mehr Dimensionen haben).

Schätzung der Standardabweichung des Mittelwerts der Partie

$$\text{Zufallsprobenahme: } s(a_L) = \frac{s_p}{\sqrt{N}}$$

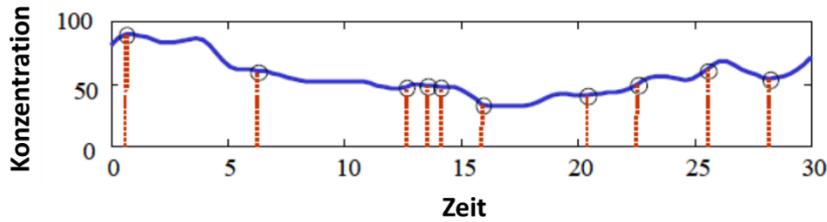
$$\text{Geschichtete Probenahme: } s(a_L) = \frac{s_{strat}}{\sqrt{N}}$$

$$\text{Systematische Probenahme: } s(a_L) = \frac{s_{sys}}{\sqrt{N}}$$

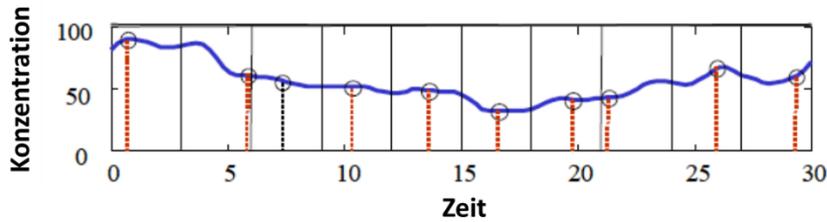
s_{strat} und s_{sys} sind Schätzungen von Standardabweichungen, bei denen die Autokorrelation berücksichtigt wurde.

Anmerkung: Diese Gleichungen gehen von einer gleichen Größe der Inkremente aus. Für andere Bedingungen siehe (zum Beispiel) [69].

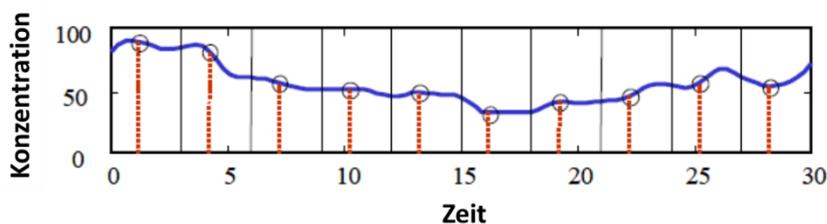
Normalerweise ist die Reihenfolge $s_p > s_{strat} > s_{sys}$, außer wenn bei der systematischen Probenahme die Probenahmefrequenz ein Vielfaches der Prozessfrequenz ist. In diesem Fall ist die systematische Probenahme die schlechteste Wahl und der Mittelwert kann systematisch abweichen.



Zufällige Auswahl



Geschichtete Auswahl



Systematische Auswahl

Abbildung C2.1: Probenmestrategien

10 ausgewählte Proben aus dem Zielobjekt mit *zufälliger*, *geschichteter* und *systematischer* Probenauswahl

Schätzung des PSE

Die Verteilungsheterogenität einer eindimensionalen Partie kann durch die Durchführung eines variographischen Experiments charakterisiert werden, d.h. es werden N Proben aus dem Zielobjekt durch systematische Stichprobenauswahl genommen. N sollte mindestens 30, besser 60...100 betragen. Proportionale Probenahme über den Querschnitt sollte verwendet werden, oder wenn dies nicht möglich ist (wenn große Gas- oder Flüssigkeitsströme beprobt werden), sollte die Durchflussmenge gleichzeitig mit der Probenahmezeit erfasst werden. Aus diesen Ergebnissen kann die experimentelle Heterogenität h_i als die relative Abweichung vom Mittelwert der Partie (oder Mittelwert des Probenahmeziels) berechnet werden, wenn N Proben der Größe M_i entnommen und analysiert werden (Ergebnisse sind a_i). M_i kann auch die Durchflussmenge sein, wenn keine proportionale Probenahme durchgeführt werden kann.

$$h_i = \frac{a_i - a_L}{a_L} \frac{M_i}{\bar{M}} \quad (i = 1, 2, \dots, N)$$

Wobei a_L das gewichtete Mittel aus der Partie ist:

$$a_L = \frac{\sum M_i a_i}{\sum M_i} = \frac{1}{N} \sum \left(\frac{M_i}{\bar{M}} \right) a_i$$

Die Standardabweichung der Heterogenität h ist gleich der *relativen Standardabweichung* der Partie oder des Prozesses, s_p .

Um die Variabilität des Prozesses zu charakterisieren, wird ein experimentelles Variogramm aus den Heterogenitäten berechnet:

$$V_j = \frac{1}{2(N-j)} \sum_{i=1}^{N-j} (h_{i+j} - h_i)^2, \quad j = 1, 2, \dots, \frac{N}{2}$$

Das Variogramm muss integriert werden, um den PSE für verschiedene Probenahmestrategien zu schätzen. Gy verwendet eine robuste numerische Integration.

C3. Quellen für Berechnungssoftware

Die klassische Varianzanalyse (ANOVA) ist in den meisten allgemeinen Tabellenkalkulationsprogrammen für einseitige ANOVA verfügbar, aber die Ergebnisse enthalten normalerweise keine expliziten Werte für alle Komponenten-Varianzen. F-Tests und andere statistische Standardtests für die Normalverteilung sind auch in den meisten Tabellenkalkulationen implementiert.

Programme für klassische und robuste statistische Methoden im Allgemeinen sind bei der RSC/AMC erhältlich. Programme, die die Anwendung des in diesem Leitfaden beschriebenen empirischen Ansatzes unterstützen, umfassen RANOVA für ausgewogene und nicht-ausgewogene Designs und RANOVA2 für dieselben, aber auch die Schätzung des Unsicherheitsfaktors und optional für Designs mit mehr als zwei Stichprobenwiederholungen.

(<http://www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/Software/index.asp>).

Ausreißertests (z. B. Grubb's oder Dixon's) sind weniger allgemein verfügbar, ebenso wie Software für die Spannweitenmethode. Die Spannweitenmethode lässt sich jedoch relativ einfach mit Hilfe der Maximum und Minimum-Funktionen in einer Tabellenkalkulation implementieren.

Die Spannweitenberechnungen (in Abschnitt 7 des Anhangs A3) lassen sich mit einem Standard-Tabellenblatt leicht durchführen.

Anhang D: Alternative experimentelle Designs für empirische Unsicherheitsschätzungen

1. Mehrstufige Designs zur Schätzung sonstiger Komponenteneffekte

Das allgemeine ausgewogene Design für die empirische Schätzung der Unsicherheit (Abbildung 2) schließt die Unsicherheit aus der physikalischen Probenvorbereitung in dem Schritt "Probe" ein. Ein alternatives experimentelles Design (Abbildung D.1) kann verwendet werden, um eine separate Schätzung der Unsicherheit aus dieser Quelle (s_{prep}) vorzunehmen. Zwei Teilproben der beiden Primärproben werden getrennt vorbereitet (graue Kästen in Abbildung D.1). Aus diesen Teilproben werden zweifache Prüfmengen entnommen, so dass auch der analytische Beitrag geschätzt werden kann. Die standardmäßige robuste ANOVA kann verwendet werden, um alle diese Varianzquellen zu trennen (Abbildung A1.2 und Anhang C3), indem zwei verschiedene Teilgruppen von vier Messungen ausgewählt werden, wie in Abbildung D.1 dargestellt. Ausführliche Informationen über die Anwendung dieses Designs auf die Probenahme von Lebensmitteln sind an anderer Stelle zu finden [22].

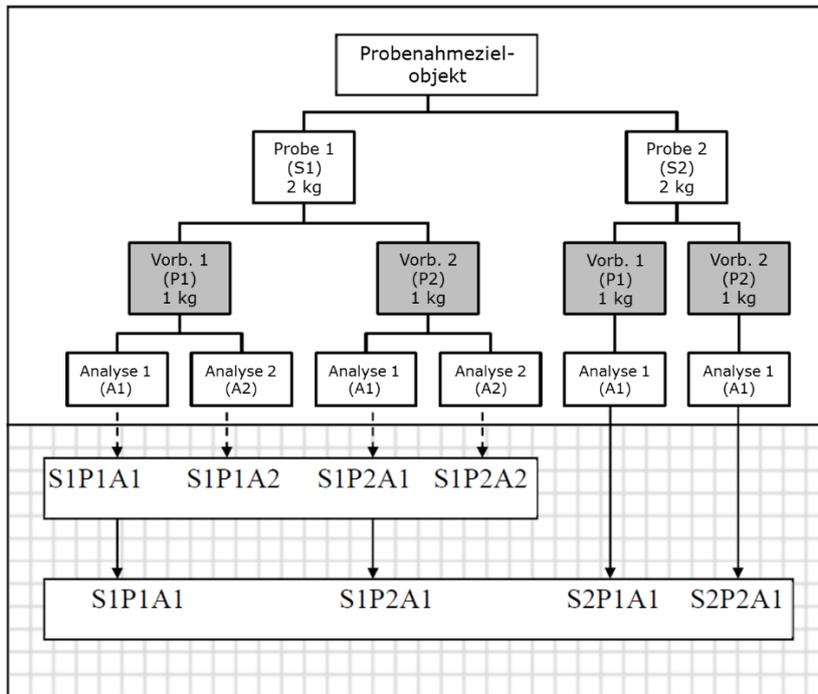


Abbildung D.1: Experimentelles Design zur Schätzung der Unsicherheit bei der Probenvorbereitung sowie bei der Probenahme und Analyse

*Im oberen Teil ist das dreistufige und nicht-ausgewogene Versuchsdesign dargestellt. Die zusätzliche Stufe in diesem Versuchsplan, die für die Bewertung von s_{prep} erforderlich ist, wird durch die grauen Kästen dargestellt. Der untere Abschnitt (schattiert) zeigt die Datengruppierungen, die für die Anwendung der ANOVA erforderlich sind, um Schätzungen für s_{samp} , s_{prep} und s_{anal} , d. h. das statistische Versuchsdesign, zu erhalten. Die Abbildung ist mit Genehmigung der Royal Society of Chemistry aus [22] entnommen.

2. Vereinfachtes und nicht-ausgewogenes Design zur Reduzierung der Kosten

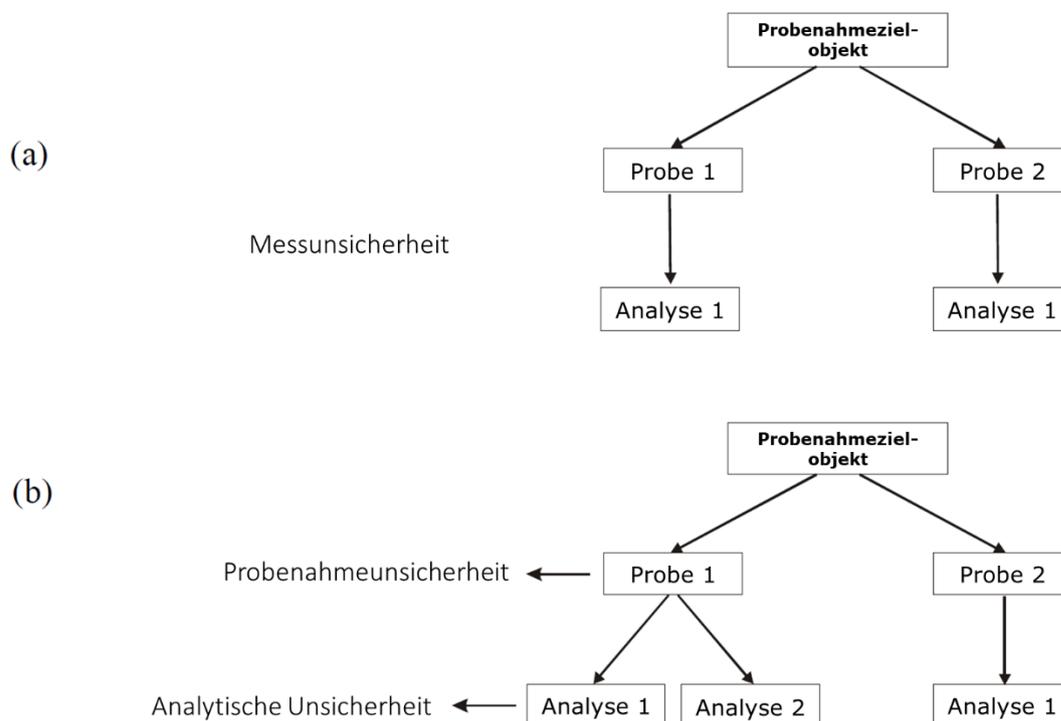


Abbildung D.2: Zwei vereinfachte Alternativen zum vollständig ausgewogenen Design (Abbildung 2), die zur Reduzierung der Kosten der Messunsicherheitsschätzung unter Verwendung des empirischen Ansatzes angewendet werden können: (a) das vereinfachte ausgewogene Design und (b) das nicht-ausgewogene Design

Das vereinfachte Design (Abbildung D.2a) enthält die gleichen Doppelproben wie das vollständig ausgewogene Design (Abbildung 2), enthält aber keine doppelten chemischen Analysen. Die Unsicherheit, die mit diesem Design geschätzt wird, gibt die gesamte Messunsicherheit an, ohne Werte für die Unsicherheitskomponenten aus der Probenahme oder der Analyse. Wenn diese Komponenten erforderlich sind, kann die analytische Unsicherheit extern vom Laboratorium geschätzt und von der Gesamtunsicherheit abgezogen werden, um, unter Verwendung von Gleichung 1, eine separate Schätzung der Probenahmeunsicherheit zu erhalten. Der Hauptvorteil dieses Designs ist, dass bei der gleichen Anzahl von Doppelproben die analytischen Kosten für die Anwendung nur halb so hoch sind wie beim vollständig ausgewogenen Design. Alternativ kann die doppelte Anzahl von Doppelproben genommen werden, und zwar aus der doppelten Anzahl von Zieleobjekten, um deren Repräsentativität bei gleichem Kostenaufwand für die chemische Analyse zu erhöhen.

Das nicht-ausgewogene Design (Abbildung D.2b) liegt zwischen diesen beiden Designs, wobei eine Doppelanalyse nur an einer der Doppelproben durchgeführt wird. Dies hat den Vorteil Schätzungen für die Komponenten aus Probenahme- und Analysenunsicherheit zu erhalten

sowie für die gesamte Messunsicherheit (mit den gleichen Vorbehalten wie beim vollständig ausgewogenen Design in Abschnitt 9.4.2). Die zusätzlichen analytischen Kosten werden im Vergleich zu denen für das vollständig ausgewogene Modell um 33 % reduziert. Die Freiheitsgrade sind in diesem Fall für die Schätzung der Varianz sowohl für die Analyse wie für die Probenahme ähnlich, was kostengünstiger ist als die zusätzlichen Freiheitsgrade für die analytische Unsicherheit im vollständig ausgewogenen Modell.

Die klassische ANOVA kann mit vielen verschiedenen Tabellenkalkulationssoftwarepaketen auf die Ergebnisse beider Designs angewandt werden (Anhang C3). Die robuste ANOVA wurde sowohl für das ausgewogene wie für das nicht-ausgewogene Design entwickelt [53]. Referenz [20] enthält auch ein ausgearbeitetes Beispiel, das die Vorteile des nicht-ausgewogenen Designs gegenüber dem ausgewogenen Design zeigt.

Anhang E: Änderung der Probenahmeunsicherheit durch Vorhersagen aus der Probenahmetheorie

Wenn die Unsicherheit der Stichproben geschätzt wurde und sich herausstellt, dass sie für den Zweck nicht ausreicht, kann es notwendig sein, diesen Unsicherheitsgrad zu ändern. Vorhersagen darüber, wie diese Änderung zu erreichen ist, können mit Hilfe der Probenahmetheorie gemacht werden (Abschnitt 10.2). Mehrere Theorien sagen voraus, dass die Probenahmevarianz umgekehrt proportional zur Masse der entnommenen Probe ist (z. B. Gleichung 6). Dies führt zu der Vorhersage, dass jede erforderliche Änderung der Unsicherheit der Probenahme (von $u_{s\text{amp}1}$ zu $u_{s\text{amp}2}$) berechnet werden kann, indem die Masse der Probe (von m_{s1} auf m_{s2}) verändert wird unter Verwendung der Beziehung

$$m_{s2} = (u_{s\text{amp}1}/u_{s\text{amp}2})^2 \cdot m_{s1} \quad (\text{Gleichung E1})$$

Dieser Ansatz lässt sich anhand der Fallstudie über Nitrat in Salat in Beispiel A1 illustrieren. Es zeigte sich (nach der Methode in Abschnitt 16.3), dass die Probenahmeunsicherheit nicht für den Zweck geeignet war, und es wurde berechnet, dass die erforderliche optimale Unsicherheit um einen Faktor von etwa 2 niedriger ist. Gleichung E1 sagt voraus, dass dies durch eine Erhöhung der Probenmasse um den Faktor 4 (d. h. 2^2) erreicht werden sollte. Die Umsetzung dieser Vorhersage durch Erhöhung der Anzahl der Inkremente von 10 auf 40 Salatköpfe pro Charge führte in diesem Fall zu der vorhergesagten Verringerung der Stichprobenunsicherheit (d. h. um den Faktor 1,80, was sich nicht statistisch signifikant von der vorhergesagten Verbesserung von 2,0 unterscheidet) [55]. Solche erfolgreichen Vorhersagen werden in der Praxis nicht immer erreicht. In einem anderen Beispiel für die Bestimmung des Feuchtigkeitsgehalts in Butter wurde berechnet, dass eine vorhergesagte Verringerung in $u_{s\text{amp}}$ um 3,7 eine Erhöhung der m_s um den Faktor 14 erfordert. In der Praxis führte diese Erhöhung der Probenmasse aber nur zu einer experimentellen Verbesserung von 1,3. Die Unfähigkeit dieses Modells zur Vorhersage der Änderung der Probenahmeunsicherheit war wahrscheinlich auf die Art der Heterogenität der Analyten in diesem speziellen Material zurückzuführen [70].

Literatur

- 1 Ellison S L R, Williams A (eds) (2012) Eurachem/CITAC Guide: *Quantifying Uncertainty in Analytical Measurement*, 3rd edition, ISBN 978-0-948926-30-3. Available from www.eurchem.org/
- 2 JCGM 100:2008 *Evaluation of measurement data - Guide to the Expression of Uncertainty in Measurement (GUM)*. Sevres, (2008). <http://www.bipm.org/en/publications/guides/gum.html>. Auch verfügbar als ISO/IEC CD Guide 98-3, *Evaluation of measurement data - Part 3: Guide to uncertainty in measurement (GUM)*, International Organization for Standardization, Geneva (2008)
- 3 Ellison S L R, Roesslein M, Williams A (eds) (1995) Eurachem Guide: *Quantifying Uncertainty in Analytical Measurement*, Eurachem, 1st edition. ISBN 0-948926-08-2
- 4 Analytical Methods Committee (1995) Uncertainty of measurement: implications of its use in analytical science. *Analyst*, 120, 2303–2308.
- 5 ISO 21748:2017 Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty evaluation. International Organization for Standardization, Geneva (2017)
- 6 Gy P M (1979) *Sampling of Particulate Materials – Theory and Practice*. Elsevier, Amsterdam, 431pp.
- 7 Codex (2004) General Guidelines on Sampling. CAC/GL-2004 (www.codexalimentarius.net/download/standards/10141/CXG_050e.pdf)
- 8 Nordtest (2007) Uncertainty from sampling. A Nordtest handbook for sampling planners and sampling quality assurance and uncertainty estimation. NT tec 604/TR604 (www.nordtest.info)
- 9 IUPAC (1990) Nomenclature for sampling in analytical chemistry (Recommendations 1990), prepared for publication by Horwitz W, *Pure and Applied Chemistry*, 62, 1193–1208.
- 10 Lee J C, Ramsey M H (2001) Modelling measurement uncertainty as a function of concentration: an example from a contaminated land investigation. *Analyst*, 126 (10), 1784–1791.
- 11 Ellison, S L R, Ramsey M H, Lawrance P, Stuart B, Minguez J, Walker M J (2017) Is measurement uncertainty from sampling related to analyte concentration? *Analytical Methods*, 9, 5989-5996. DOI: 10.1039/C7AY00752C. (pubs.rsc.org/en/content/articlelanding/2017/ay/c7ay00752c#!divAbstract)
- 12 Magnusson B, Naykki T, Hovind H, Krysell M, Sahlin E, Handbook for calculation of measurement uncertainty in environmental laboratories, Nordtest Report TR 537 (ed. 4) 2017. Verfügbar von www.nordtest.info
- 13 Gy P M (1992) *Sampling of Heterogeneous and Dynamic Material Systems*. Elsevier, Amsterdam.
- 14 Magnusson B. and Ornemark U. (eds.) Eurachem Guide: *The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics*, (2nd ed. 2014). ISBN 978-91-87461-59-0.

- 15 Ramsey M H, Squire S, Gardner M J (1999) Synthetic reference sampling target for the estimation of measurement uncertainty. *Analyst*, 124 (11), 1701–1706.
- 16 Thompson M, Ramsey M H (1995) Quality concepts and practices applied to sampling – an exploratory study. *Analyst*, 120, 261–270.
- 17 Ramsey M H, Geelhoed B, Damant, AP, Wood, R (2011) Improved evaluation of measurement uncertainty from sampling by inclusion of between-sampler bias using sampling proficiency testing. *Analyst*, 136 (7), 1313 – 1321. DOI:10.1039/C0AN00705F.
- 18 Ramsey M H (1998) Sampling as a source of measurement uncertainty: techniques for quantification and comparison with analytical sources. *Journal of Analytical Atomic Spectrometry*, 13, 97–104.
- 19 Lyn J A, Ramsey M H, Coad D S, Damant A P, Wood R, Boon K A (2007) The duplicate method of uncertainty estimation: are 8 targets enough? *Analyst* 132, 1147-1152. DOI: 10.1039/b702691a.
- 20 Rostron P and Ramsey M H (2012) Cost effective, robust estimation of measurement uncertainty from sampling using unbalanced ANOVA. *Accreditation and Quality Assurance*. 17, 1, 7-14. DOI: 10.1007/s00769-011-0846-2
- 21 Ramsey M H and Ellison S L R (2017) Combined uncertainty factor for sampling and analysis. *Accreditation and Quality Assurance*, 22(4), 187-189. DOI 10.1007/s00769-017-1271-y
- 22 Lyn J A, Ramsey M H, Fussel R J, Wood R (2003) Measurement uncertainty from physical sample preparation: estimation including systematic error. *Analyst*, 128 (11), 1391 – 1398
- 23 Ramsey M H, Ellison S L R (2015) Uncertainty Factor: an alternative way to express measurement uncertainty in chemical measurement. *Accreditation and Quality Assurance*. 20, 2, 153-155. DOI:10.1007/s00769-015-1115-6
- 24 Ramsey M H, Argyraki A (1997) Estimation of measurement uncertainty from field sampling: implications for the classification of contaminated land. *Science of the Total Environment*, 198, 243–257.
- 25 De Zorzi P, Belli M, Barbizzi S, Menegon S, Deluisa A (2002) A practical approach to assessment of sampling uncertainty. *Accreditation and Quality Assurance*, 7, 182–188.
- 26 De Zorzi, Barbizzi S, Belli M, Barbina M, Fajgelj A, Jacimovic R, Jeran Z, Menegon S, Pati A, Petruzzelli G, Sansone U, Van der Perk M.(2008) Estimation of uncertainty arising from different soil sampling devices: The use of variogram parameters. *Chemosphere* 70, 745–752. DOI:10.1016/j.chemosphere.2007.07.068
- 27 Kurfürst U, Desaules A, Rehnert A, Muntau H (2004) Estimation of measurement uncertainty by the budget approach for heavy metal content in soils under different land use. *Accreditation and Quality Assurance*, 9, 64–75.
- 28 Minkkinen P (2004) Practical applications of sampling theory. *Chemometrics and Intelligent Lab. Systems*, 74, 85–94.
- 29 Gy P M (1998) *Sampling for Analytical Purposes*. John Wiley & Sons Ltd, Chichester.

- 30 Gy P M (2004) Proceedings of First World Conference on Sampling and Blending. Special Issue of Chemometrics and Intelligent Lab. Systems, 74, 7–70.
- 31 Pitard F F (1993) Pierre Gy's Sampling Theory and Sampling Practice. CRC Press, Boca Raton, 2nd edition.
- 32 Smith P L (2001) A Primer for Sampling Solids, Liquids and Gases – Based on the Seven Sampling Errors of Pierre Gy. ASA SIAM, USA.
- 33 Analytical Methods Committee (2001). Robust statistics: a method of coping with outliers. Technical Brief No.6, Royal Society of Chemistry, London.
www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/TechnicalBriefs.asp
- 34 Ramsey M H (1993) Sampling and analytical quality control (SAX) for improved error estimation in the measurement of heavy metals in the environment, using robust analysis of variance. Applied Geochemistry, 2, 149–153.
- 35 Nordtest Sampler Certification, Version 1-0, Nordic Innovation Centre (2005).
- 36 Thompson M, Wood R (1995) Harmonised guidelines for internal quality control in analytical chemistry laboratories. Pure and Applied Chemistry, 67, 649–666.
- 37 Magnusson B, Naykki T, Hovind H, Krysell M (2003) Handbook for Calculation of Measurement Uncertainty. NORDTEST report TR 537 (project 1589-02). Erhältlich von www.nordicinnovation.net/nordtest.cfm unter dem Link Rapportier.
- 38 Thompson M, Coles B J, Douglas J K (2002) Quality control of sampling: proof of concept. Analyst, 127, 174–177.
- 39 Farrington D, Jervis A, Shelley S, Damant A, Wood R, Thompson M (2004) A pilot study of routine quality control of sampling by the SAD method, applied to packaged and bulk foods. Analyst, 129, 359–363.
- 40 Codex (2004) Guidelines on measurement uncertainty CAC/GL 54-2004.
- 41 Ellison S L R and Williams A (Eds). Eurachem/CITAC guide: Use of uncertainty information in compliance assessment. (First Edition, 2007). Available from www.eurachem.org.
- 42 ISO/IEC Guide 98-4:2012(en) Uncertainty of measurement — Part 4: Role of measurement uncertainty in conformity assessment. International Organization for Standardization, Geneva (2012).
- 43 JCGM 200:2012 International vocabulary of metrology – Basic and general concepts and associated terms (VIM) 3rd edition.
- 44 Bettencourt da Silva R, Williams A (Eds), Eurachem/CITAC Guide: Setting and Using Target Uncertainty in Chemical Measurement, (1st ed. 2015). Erhältlich von www.eurachem.org.
- 45 Ramsey M H, Thompson M, Hale M (1992) Objective evaluation of precision requirements for geochemical analysis using robust analysis of variance. Journal of Geochemical Exploration, 44, 23–36.
- 46 Thompson M, Fearn T (1996) What exactly is fitness for purpose in analytical measurement? Analyst, 121, 275–278.

- 47 Boon K A and Ramsey M H (2012) Judging the fitness of on-site measurements by their uncertainty, including the contribution from sampling. *Science of the Total Environment* 419, 196–207. [dx.doi.org/10.1016/j.scitotenv.2011.12.001](https://doi.org/10.1016/j.scitotenv.2011.12.001)
- 48 Ramsey M H and Boon K A (2012) Can in situ geochemical measurements be more fit-for-purpose than those made ex situ? *Applied Geochemistry* 27, 969-976. [dx.doi.org/10.1016/j.apgeochem.2011.05.022](https://doi.org/10.1016/j.apgeochem.2011.05.022)
- 49 AMC (2018) Analytical Methods Committee Technical Brief No 84. Beam sampling: taking samples at the micro-scale, *Analytical Methods*, 10, 1100-1102
- 50 Rostron, P, Heathcote J A, Ramsey, M H (2015) Evaluation of uncertainties in in situ and ex situ gamma measurements on land areas with low contamination levels. *Journal of Radiological Protection* 35 (2015) 391-399. DOI: <https://dx.doi.org/10.1088/0952-4746/35/2/391>
- 51 Commission Regulation (EC) No 563/2002 of 2 April 2002 amending Regulation (EC) No 466/2001 setting maximum levels for certain contaminants in foodstuffs, *Official Journal of the European Communities*, L 86/5 to L 86/6.
- 52 European Directive 79/700/EEC. *OJ L 207*, 15.8.1979, p26.
- 53 Robust ANOVA programs available for balanced design (RANOVA) and both balanced and unbalanced designs (RANOVA2) from website of Royal Society of Chemistry. www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/Software/
- 54 Ramsey M H, Lyn J A, Wood R (2001) Optimised uncertainty at minimum overall cost to achieve fitness-for-purpose in food analysis. *Analyst*, 126, 1777–1783.
- 55 Lyn J A, Palestra I M, Ramsey M H, Damant A P, Wood R (2007) Modifying uncertainty from sampling to achieve fitness for purpose: a case study on nitrate in lettuce. *Accreditation and Quality Assurance: Journal for Quality, Comparability and Reliability in Chemical Measurement*, 12, 67-74.
- 56 Argyraki A, (1997) Estimation of measurement uncertainty in the sampling of contaminated land. PhD Thesis (Imperial College) awarded by University of London.
- 57 Thompson M, Walsh J N A (1989) *Handbook of Inductively Coupled Plasma Spectrometry*. Blackie, Glasgow.
- 58 Ramsey M H, Solomon-Wisdom G O, Argyraki A (2013) Evaluation of in situ heterogeneity of elements in solids: implications for analytical geochemistry. *Geostandards and Geoanalytical Research*, 37, 4, 379-391. DOI: [10.1111/j.1751-908X.2013.00236.x](https://doi.org/10.1111/j.1751-908X.2013.00236.x)
- 59 Analytical Methods Committee (2002) Fitting a linear functional relationship to data with error on both variables, Technical Brief No.10, Royal Society of Chemistry, London. www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/TechnicalBriefs.asp
- 60 Ramsey M H, Taylor P D, Lee J C (2002) Optimized contaminated land investigation at minimum overall cost to achieve fitness-for-purpose. *Journal of Environmental Monitoring*, 4 (5), 809–814.
- 61 ISO 18400-104:2018, Soil Quality – Sampling – Part 104: Strategies, Annex C

- 62 Dahinden R, Desaulles A (1994) Die Vergleichbarkeit von Schwermetallanalysen in Bodenproben von Dauerbeobachtungsflächen, Eidgenössische Forschungsanstalt für Agrikulturchemie und Umwelthygiene. Liebefeld, Switzerland, p27.
- 63 ISO 3534-2: 2006 Statistics – Vocabulary and symbols - Part 2: Applied statistics. International Organization for Standardization, Geneva (2006).
- 64 AMC (2005) Analytical Methods Committee Technical Brief No 19. Terminology – the key to understanding analytical science. Part 2: Sampling and sample preparation www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/TechnicalBriefs.asp
- 65 ISO Standard 78-2: Chemistry – Layouts for Standards – Part 2: Methods of Chemical Analysis (Second Edition, 1999).
- 66 IUPAC (2005) Terminology in Soil Sampling (IUPAC Recommendations 2005), prepared for publication by De Zorzi P, Barbizzi S, Belli M, Ciceri G, Fajgelj A, Moore D, Sansone U, and Van der Perk M. Pure and Applied Chemistry, 77 (5), 827–841.
- 67 ISO 11074:2015 Soil Quality – Vocabulary. International Organization for Standardization, Geneva (2015).
- 68 ISO 7002:1986 Agricultural food products – Layout for a standard method of sampling from a lot. First edition, International Organization for Standardization, Geneva (1986)
- 69 Cochran, W.G. (1977) Sampling Techniques, Wiley
- 70 Lyn J A, Ramsey M H, Damant A, Wood R (2005), Two stage application of the OU method: a practical assessment, Analyst, 130, 1271-1279

