quo data
Gesellschaft
für Qualitätsmanagement
und Statistik mbH

Leitfaden zur Bewertung der Gleichwertigkeit von unterschiedlichen Messverfahren für anorganische und organische Schadstoffparameter im Rahmen der Bundes-Bodenschutz- und Altlastenverordnung (BBodSchV)

Umweltbundesamt Dessau-Roßlau, März 2008

PD Dr. habil. Steffen Uhlig
quo data Gesellschaft für Qualitätsmanagement
und Statistik mbH

Kaitzer Str. 135
D-01187 Dresden

Tel: 0351 40 35 66 31
Fax: 0351 40 35 66 39

E-Mail: info@quodata.de
www.quodata.de

Im Auftrag des Fachbeirates für Bodenuntersuchungen

Fachliche Begleitung: Dr. Detlef Lück, BAM

Inhalt

1. Einführung 1

1.1. Die Einzelkriterien der Äquivalenzprüfung 1

1.2. Die Stufen der Gleichwertigkeitsprüfung 2

Stufe 1: Mittelwerte, Prüfung der Wiederfindungsrate (WFR) 2

Stufe 2: Gleichwertigkeitsteilprüfung 3

Stufe 3: Vollständige In-Haus-Gleichwertigkeitsprüfung 3

Stufe 4: Vollständige laborübergreifende Gleichwertigkeitsprüfung 3

1.3. Definition und Abgrenzung einiger Begriffe 3

Probenmessung (Run): 3

Parallelbestimmungen: 3

Unabhängige Messtage: 4

Referenzverfahren: 4

Alternativverfahren: 4

2. Mindestanforderungen 4

2.1. Mindestanforderungen bezüglich der Wiederfindungsrate 4

2.2. Mindestanforderungen bezüglich der Wiederholstandardabweichung 5

2.3. Mindestanforderungen bezüglich der Intermediate-Standard-abweichung 6

2.4. Zeitliche Abfolge der Messungen 7

3. Auswahl der Proben 10

4. Standardversuchspläne 11

4.1. WFR-Prüfung 11

Systematische Abweichung, WFR-Prüfung mit Referenzverfahren 11

4.2. Gleichwertigkeitsteilprüfung 12

4.3. Vollständige In-Haus-Gleichwertigkeitsprüfung 13

5. Referenzen 14

1. Einführung

In der Bundesbodenschutzverordnung 3.1.3 „Analysenverfahren“ ist bei der Anwendung alternativer Verfahren verlangt, dass die Ergebnisse mit den Ergebnissen der in den Tabellen 3, 4 und 5 angegebenen Verfahren gleichwertig oder vergleichbar sind. Die Bewertung der Gleichwertigkeit von unterschiedlichen Analysenverfahren stellt auch einen der Schwerpunkte der Arbeit des Fachbeirates für Bodenuntersuchungen dar.

Im Auftrag des Fachbeirates wurde 2003 ein Konzept zur statistischen Bewertung der Gleichwertigkeit von Analysenverfahren erarbeitet (Uhlig, Steffen (2003), Empfehlungen zur Charakterisierung der Gleichwertigkeit und Vergleichbarkeit von Verfahren und Methoden. In: www.umweltbundesamt.de/fbu.) Die Gleichwertigkeitsprüfung basiert auf dem in dem Softwareprogramm EquiVal implementierten Konzept des Äquivalenztests, d.h. eines Kriteriums, bei dem ein Nachweis hinreichender Ähnlichkeit der Verfahrensparameter der beiden Messverfahren zu erbringen ist.

Um eine Prüfung der Gleichwertigkeit eines Alternativverfahrens zu einem Referenzverfahren auszuführen, sind die Analysenverfahren mehrfach und parallel auf geeignete Probenmaterialien anzuwenden. Es werden damit die Gehalte der zu untersuchenden Analyten in den Proben in Abhängigkeit vom Analysenverfahren sowie die Streuungsparameter der Analysenverfahren ermittelt. Für bestimmte Fragestellungen können auch zertifizierte Referenzmaterialien (ZRM) verwendet werden, so dass auf die Anwendung des Referenzverfahrens verzichtet werden kann. Die Untersuchungen erfolgen zunächst in der Regel innerhalb eines Laboratoriums. Sind die Ergebnisse der Prüfungen positiv, erfolgt in einem zweiten Schritt die Anwendung des Alternativverfahrens unter Vergleichbarkeitsbedingungen. In diesem Fall wenden mehrere Laboratorien beide Verfahren im Rahmen eines Ringversuches an.

Mathematische Details sind der Ausarbeitung in Uhlig [2003] zu entnehmen.

1.1. Die Einzelkriterien der Äquivalenzprüfung

Die Überprüfung der Gleichwertigkeit unterschiedlicher Messverfahren im Rahmen des Äquivalenzkonzeptes beinhaltet die Bewertung der im folgenden dargestellten vier Teilkomponenten, d.h. (1) der Mittelwerte mit systematischen Messabweichungen, (2) der zufälligen Messabweichungen unter Wiederholbedingungen, (3) der zufälligen Messabweichungen unter Intermediate-Bedingungen und (4) der zufälligen Messabweichungen unter Vergleichbedingungen:

- Beurteilung der Gleichwertigkeit bezüglich des Mittelwertes mit Hilfe von systematischen Messabweichungen (Wiederfindungsrate): Wenn das zu überprüfende Messverfahren im Durchschnitt z.B. erheblich höhere Werte liefert als das Referenzverfahren, ist eine Gleichwertigkeit der Resultate nicht gewährleistet. Systematische Messabweichungen zwischen den Messverfahren sollten daher statistisch signifikant unterhalb einer vorgegebenen Schranke liegen, damit die beiden Verfahren als gleichwertig angesehen werden können. Für die vorgegebene relative Schranke wird ein Wert von **15%** als noch vertretbar angesehen. Als Faustregel kann gelten, dass diese Schranke die rela-

tive Vergleichstandardabweichung des Referenzverfahrens nicht deutlich überschreiten sollte. Falls die Präzisionsanforderungen höher oder niedriger sind, kann die Bewertungsschranke angepasst werden.

- Beurteilung der Gleichwertigkeit bezüglich zufälliger Messabweichungen unter Wiederholbedingungen, d.h. im gleichen Labor, mit dem gleichen Bearbeiter, in kurz hintereinander und unter identischen Bedingungen durchgeführten Messungen: Wenn das zu überprüfende Messverfahren eine wesentlich höhere Streuung unter Wiederholbedingungen aufweist als das Referenzverfahren, ist der Vergleich von Messwerten innerhalb einer Messreihe erheblich schwieriger als beim Referenzverfahren. Daher sollte der Quotient aus der Wiederholstandardabweichung des zu überprüfenden Messverfahrens (Alternativverfahren) zur Wiederholstandardabweichung des Referenzverfahrens signifikant unter einer vorgegebenen Schranke liegen (siehe Uhlig [2003]).
- Beurteilung der Gleichwertigkeit bezüglich zufälliger Messabweichungen unter Intermediate-Bedingungen (ISO 5725), d.h. unter unterschiedlichen Bedingungen und zu unterschiedlichen Zeitpunkten im gleichen Labor. Das zu überprüfende Messverfahren sollte auch bei Messungen unter unterschiedlichen Bedingungen keine deutlich stärkeren Schwankungen aufweisen als das Referenzverfahren. Daher sollte der Quotient aus der Intermediate-Standardabweichung des Vergleichsverfahrens zur Intermediate-Standardabweichung des Referenzverfahrens signifikant unter einer vorgegebenen Schranke liegen (siehe Uhlig [2003]).
- Beurteilung der Gleichwertigkeit bezüglich zufälliger Messabweichungen unter Vergleichbedingungen, d.h. unter unterschiedlichen Bedingungen in unterschiedlichen Labors. Das zu überprüfende Messverfahren sollte auch bei Messungen in unterschiedlichen Labors zu keinen deutlich stärkeren Schwankungen führen als das Referenzverfahren. Daher sollte der Quotient der beiden Vergleichstandardabweichungen statistisch signifikant unterhalb einer vorgegebenen Schranke liegen (siehe Uhlig [2003]). Es ist festzuhalten, dass die Überprüfung der Gleichwertigkeit der Vergleichstandardabweichung im Gegensatz zu den drei vorher genannten Gleichwertigkeitskriterien nur im Rahmen eines Ringversuchs möglich ist.

1.2. Die Stufen der Gleichwertigkeitsprüfung

Die im vorigen Abschnitt genannten Kriterien können entweder schrittweise oder in kombinierter Form überprüft werden. Hierzu bieten sich folgenden Varianten an:

Stufe 1: Mittelwerte, Prüfung der Wiederfindungsrate (WFR)
Eine Überprüfung der systematischen Messabweichungen kann in der Regel bereits durch einige Vergleichsmessungen oder durch Bestimmung geeigneter (zertifizierter) Referenzmaterialien vorgenommen werden. Wenn bei dieser WFR-Prüfung nicht der Nachweis der Gleichwertigkeit erbracht wird, kann man die Anzahl untersuchter Proben erhöhen und die Prüfung wiederholen. Bei einer WFR-Prüfung kann auf Messungen mittels des Referenzverfahrens verzichtet werden, wenn geeignete (zertifizierte) Referenzmaterialien zur Verfügung stehen.

Stufe 2: Gleichwertigkeitsteilprüfung
Bei ausreichender Anzahl von Messwiederholungen kann sich an eine erfolgreiche WFR-Prüfung sogleich eine Überprüfung der Gleichwertigkeit bezüglich der Wiederholstandardabweichung anschließen. Die In-Haus-Untersuchung kann jedoch auch so angelegt werden, dass die Gleichwertigkeit gleichzeitig bezüglich Wiederfindungsrate und Wiederholstandardabweichung überprüft wird. Die Gleichwertigkeitsteilprüfung kann nur unter gleichzeitiger Verwendung des Referenz- und des Alternativverfahrens realisiert werden.

Stufe 3: Vollständige In-Haus-Gleichwertigkeitsprüfung
Die Einbeziehung der Intermediate-Standardabweichung erfordert wiederholte Messungen der Probenmaterialien über einen längeren Zeitraum. Im Rahmen dieser In-Haus Untersuchung erfolgt die parallele Gleichwertigkeitsprüfung von Wiederfindungsrate, Wiederholstandardabweichung und Intermediate-Standardabweichung.

Stufe 4: Vollständige laborübergreifende Gleichwertigkeitsprüfung
Sofern in einem Ringversuch beide Messverfahren jeweils durch eine ausreichende Anzahl von Laboratorien angewandt wurden, kann eine vollständige laborübergreifende Gleichwertigkeitsprüfung durchgeführt werden, welche Wiederfindungsrate, Wiederholstandardabweichung und Vergleichstandardabweichung berücksichtigt.

1.3. Definition und Abgrenzung einiger Begriffe

Um die im Folgenden beschriebene Vorgehensweise eindeutig abzugrenzen, werden zunächst einige Begriffe definiert:

Probenmessung (Run):
Eine Probenmessung (Run) umfasst die Messung eines Probenmaterials zur Ermittlung des Analyten mit einem Messverfahren. In einer Probenmessung wird wahlweise das Probenmaterial entweder nur einmal oder mehrfach in Parallelproben unter Wiederholbedingungen, d.h. unter identischen Laborbedingungen kurz hintereinander mit dem gleichen Bearbeiter gemessen werden. Dabei wird vorausgesetzt, dass es sich um unabhängige Probenmessungen handelt, d.h. jede Parallelprobe durchläuft den gesamten Analysengang von der Extraktion oder dem Aufschluss über eventuelle Anreicherungs-/Abtrennungsschritte bis zur Messung der Probenlösung im Messgerät. Nur eine wiederholte Ablesung des Messgerätes mit einer Probenlösung ist nicht ausreichend.

Parallelbestimmungen:
Parallelbestimmungen umfassen grundsätzlich nicht nur die Ablesung, sondern das Gesamtverfahren, d.h. bei 5 Parallelbestimmungen ist das Gesamtverfahren fünfmal in allen Einzelschritten zu wiederholen. Parallelbestimmungen können entweder innerhalb eines Runs unter Wiederholbedingungen, oder an unterschiedlichen Messtagen unter Intermediate-Bedingungen, durchgeführt werden.

Versuchsplan:
Ein Versuchsplan besteht aus der wiederholten Messung einer oder mehrerer Probenmaterialien mit einer Einfach- oder Mehrfachbestimmung (Parallelproben) unter Wiederholbedingungen. Die Wiederholungsmessungen dienen zur Ermittlung der Intermediate-Standardabweichung und erfolgen an unabhängigen Messtagen. Die Anzahl der statistisch erforderlichen Wiederholungen von Probenmessungen für ein oder mehrere Probenmaterialien wird später näher ausgeführt. In jedem Fall sollte die Anzahl der Wiederholungsmessungen in einem Versuchsplan für einen Satz von Probenmaterialien bei der Messung mit dem Alternativ- oder Referenzverfahren gleich sein. Vorausgesetzt ist dabei auch, dass es sich jeweils um die gleichen Proben handelt, die durch die beiden Messverfahren untersucht werden.

Unabhängige Messtage:
Zwei oder mehrere Messtage können als unabhängig bezeichnet werden, wenn die Unterschiede der zugehörigen Laborbedingungen typisch sind für das betreffende Labor. Unterschiedliche Laborbedingungen und unabhängige Messtage sind am ehesten zu realisieren, wenn (1) die Messtage nicht direkt aufeinander folgen, (2) das Verfahren von unterschiedlichen Bearbeitern ausgeführt wird sind, und (3) deutliche Unterschiede in der Wetterlage (Temperatur, Feuchte) zu verzeichnen sind. Weiterhin wird vorausgesetzt, dass für jeden Messtag das Messgerät mit neu angesetzten Kalibrierungslösungen kalibriert wird.

Referenzverfahren:
Generell wird bei der Durchführung einer Gleichwertigkeitsprüfung unterstellt, dass eines der beiden Messverfahren genormt, amtlich oder allgemein akzeptiert ist. Dieses Messverfahren wird als Referenzverfahren bezeichnet.

Alternativverfahren:
Jenes Messverfahren, für das der Nachweis der Gleichwertigkeit mit dem Referenzverfahren erbracht werden soll, wird als Alternativverfahren bezeichnet.

2. Mindestanforderungen

2.1. Mindestanforderungen bezüglich der Wiederfindungsrate

Wie oben schon dargelegt, ist die Überprüfung der Gleichwertigkeit bezüglich der WFR sowohl mittels (zertifizierten) Referenzmaterials als auch durch Vergleich mit einem Referenzverfahren, möglich.

Das hier verwendete Äquivalenzprinzip ermöglicht bei sehr günstig ausgefallenen Vergleichsmessungen u.U. bereits bei einer sehr geringen Anzahl von Probenmessungen den Nachweis der Gleichwertigkeit. Dieser Nachweis ist bezüglich der Wiederfindung erbracht, wenn die tatsächliche relative Abweichung der Wiederfindungsrate statistisch signifikant unterhalb eines vorgegebenen Wertes liegt. Vorläufig wird diese Toleranzgrenze mit **15%** festgelegt. Diese Grenze ist eine theoretische Grenze, jedoch keine Entscheidungsgrenze. Sie markiert jene theoretische, mittlere Abweichung, die bei einem gleichwertigen Messverfahren als noch akzeptabel angesehen wird. Die tatsächliche Entscheidungsgrenze ist kleiner, da hierbei auch die analytische Messunsicherheit berücksichtigt werden muss. Diese Grenze, unterhalb derer eine statistische signifikante Differenz zu den vorgeschlagenen **15%** besteht, wird als die „maximal tolerierte relative Abweichung“ bezeichnet. Damit eine realistische Chance besteht, das Gleichwertigkeitskriterium zu erfüllen, sollte diese Grenze, d.h. die maximal tolerierte relative Abweichung, nicht unter 7,5% liegen. Ihr Wert hängt in starkem Maße von der Intermediate-Standardabweichung ab, in geringerem Maße auch von der Wiederholstandardabweichung. Daneben hängt die maximal tolerierte relative Abweichung auch davon ab, ob die Gleichwertigkeitsprüfung entweder anhand eines (zertifizierten) Referenzmaterials vorgenommen wird oder durch Vergleichsuntersuchung mittels eines Referenzverfahrens. Damit die maximal tolerierte relative Abweichung nicht unter 7,5% liegt, sind bei einer relativen Intermediate-Standardabweichung von 10% mindestens die in der folgenden Tabelle dargestellte Versuchsumfänge erforderlich. Bei den Angaben, die sich auf das (zertifizierte) Referenzmaterial beziehen, wurde eine relative Standardabweichung des angegebenen Gehaltes von 2% unterstellt. Dies entspricht einer erweiterten Ergebnisunsicherheit von +/-4%. Alle Angaben gelten auch für Ringversuche, wobei die Anzahl der Runs durch die Anzahl der Laboratorien zu ersetzen ist.

Die Tabelle gilt für eine beliebige Anzahl von Parallelproben einer Probenmessung. Je Messtag ist mindestens eine Messung pro Probenmaterial auszuführen, d.h. die an jedem der Messtage zu realisierende Probenmessung muss aus mindestens einer Messung bestehen. Die Ausführung von Mehrfachmessungen (z.B. Doppelbestimmungen) empfiehlt sich, wenn zugleich eine Überprüfung der Gleichwertigkeit hinsichtlich der Wiederholstandardabweichung erfolgen soll. Weil die hier dargestellten Versuchspläne von der zugrunde liegenden Intermediate- oder Vergleichstandardabweichung abhängig sind, empfiehlt sich eine adaptive Vorgehensweise: Wenn sich während der Durchführung der Messungen zeigt, dass die tatsächlich ermittelte Intermediate- oder Vergleichstandardabweichung deutlich höher ist als der hier unterstellte Wert von 10%, sollte die Anzahl der untersuchten Probenmessungen mit neuen oder bereits gemessenen Probenmaterialien erhöht werden. Wenn jedoch das Messverfahren zeitlich stabil ist, reicht es unter Umständen aus, wenn anstelle der Einzelbestimmungen Doppelbestimmungen ausgeführt werden.

Tabelle 1: Versuchsplan zur Wiederfindungsrate

Versuchsumfang zur Überprüfung der Gleichwertigkeit bezüglich der Wiederfindung Annahme: Messung der Proben mit einer relativen Intermediate-Standardabweichung von 10%		
Anzahl Probenmaterialien	Anzahl Probenmessung (Run) zu unterschiedlichen Zeitpunkten (unabhängige Messtage)	
	Vergleich mit Referenzverfahren	Vergleich mit ZRM (Annahme: Erweiterte Messunsicherheit = ± 4%)*
1	11	6
2	6	4
3	4	3
4	3	2
5	3	2
6 oder mehr	2	2

*In Zertifikaten der ZRM werden Unsicherheiten des zertifizierten Wertes häufig als Vertrauensintervalle angegeben. Diese müssen dann nach s = Wurzel(n) *VI / t in die Standardabweichung umgerechnet werden.

2.2. Mindestanforderungen bezüglich der Wiederholstandardabweichung

Wie bei der Wiederfindungsrate genügt auch bei der Wiederholstandardabweichung bei sehr günstig ausgefallenen Vergleichsmessungen bereits eine sehr geringe Anzahl von Probenmessungen, um den Nachweis der Gleichwertigkeit erbringen zu können. Dieser Nachweis ist bezüglich der Wiederholstandardabweichung erbracht, wenn der Quotient der Wiederholstandardabweichungen aus Alternativ- und Referenzverfahren den Wert Δ statistisch signifikant unterschreitet. Die Grenze, unterhalb derer eine statistisch signifikante Differenz zu dem vorgegebenen Wert Δ besteht, wird als „maximal tolerierter Quotient“ bezeichnet. Damit eine realistische Chance besteht, das Gleichwertigkeitskriterium zu erfüllen, sollte diese Grenze, d.h. der maximal tolerierte Quotient, nicht unter 2 liegen. Sein Wert hängt nicht von der Wiederholstandardabweichung selbst ab, sondern nur von der Anzahl der untersuchten Probenmaterialien, der Anzahl der Probenmessungen sowie der Anzahl der Parallelbestimmungen (Messwiederholungen). Damit der maximal tolerierte Quotient nicht unter 2 liegt, sind die in der folgenden Tabelle dargestellten Versuchsumfänge erforderlich.

Tabelle 2: Versuchsplan zur Wiederholstandardabweichung

Versuchsumfang zur Überprüfung der Gleichwertigkeit bezüglich der Wiederholstandardabweichung			
Anzahl Probenmaterial	Anzahl Probenmessung (Run) ...		
	... bei 2 Wiederholmessungen (2 Parallelproben)	... bei 3 Wiederholmessungen (3 Parallelproben)	... bei 4 Wiederholmessungen (4 Parallelproben)
1	15	8	5
2	8	4	3
3	5	3	2
4	4	2	2
5	3	2	2
6	3	2	2
7	3	2	2
8 oder mehr	2	2	2

Analoge Regeln lassen sich auch für die aus Ringversuchen ermittelte Wiederholstandardabweichung aufstellen, wobei anstelle der Anzahl der Runs die Anzahl der Laboratorien tritt. Die Anzahl der Laboratorien sollte dabei jedoch ca. 20% höher gewählt werden als die hier angegebene Zahl der Proben.

2.3. Mindestanforderungen bezüglich der Intermediate-Standardabweichung

Auch bei der Intermediate-Standardabweichung genügt unter Umständen bereits eine sehr geringe Anzahl von Probenmessungen, um den Nachweis der Gleichwertigkeit erbringen zu können. Dieser Nachweis ist erbracht, wenn der Quotient der Intermediate-Standardabweichungen aus Alternativ- und Referenzverfahren den Wert Δ statistisch signifikant unterschreitet. Die Grenze, unterhalb derer eine statistisch signifikante Differenz zu dem vorgegebenen Wert Δ besteht, wird als „maximal tolerierter Quotient“ bezeichnet. Damit eine realistische Chance besteht, das Gleichwertigkeitskriterium zu erfüllen, sollte diese Grenze, d.h. der maximal tolerierte Quotient, nicht unter 2 liegen. Auch dieser Wert hängt nicht von der Intermediate-Standardabweichung selbst ab, sondern nur von der Anzahl der untersuchten Probenmaterialien und der Anzahl der Probenmessungen. Damit der maximal tolerierte Quotient nicht unter 2 liegt, sind mindestens die in der folgenden Tabelle dargestellten Proben und Runs erforderlich.

Tabelle 3: Versuchsplan zur Intermediate-Standardabweichung

Versuchsumfang zur Überprüfung der Gleichwertigkeit bezüglich der Intermediate-Standardabweichung	
Anzahl Probenmaterialien	Anzahl Probenmessung (Run)
1	16
2	9
3	6
4	5
5	4
6	4
7	4
8	3
9	3
10	3
11	3
12	3
13	3
14	3
15 oder mehr	2

Analoge Regeln lassen sich auch für die aus Ringversuchen ermittelte Vergleichstandardabweichung aufstellen, wobei anstelle der Anzahl der Runs die Anzahl der Laboratorien tritt. Die Anzahl der Laboratorien sollte dabei jedoch ca. 20% höher gewählt werden als die hier angegebene Zahl der Proben.

2.4. Zeitliche Abfolge der Messungen

Im Rahmen dieses Leitfadens bezeichnet eine Probenmessung (Run) alle Messungen ein und derselben Probe, die hintereinander und ohne zeitliche Verzögerung durchgeführt werden. Aus Gründen der Einheitlichkeit wird auch dann von einer Probenmessung gesprochen, wenn es sich um eine Einfachbestimmung handelt, wenn also auf Parallelproben verzichtet wird. Bei der Festlegung der zeitlichen Abfolge der Probenmessung sind drei Grundsätze zu beachten, (1) der Grundsatz der Randomisierung, (2) der Grundsatz der zeitlichen Unabhängigkeit, und (3) der Grundsatz der Repräsentativität der Messbedingungen.

Randomisierung bedeutet, dass die Reihenfolge der für die verschiedenen Probenmaterialien durchgeführten Probenmessungen zufällig gewählt werden sollte. Es ist also nicht zulässig, dass zunächst alle Probenmessungen einer Probe durchgeführt werden, dann alle Probenmessungen der zweiten Probe usw. Vielmehr sind die Probenmessungen in geeigneter Weise zu mischen. Diese Randomisierung ist erforderlich, um bei der statistischen Auswertung überprüfen zu können, ob etwaige Abweichungen bei der Wiederfindung einer Probe auf die Probe selbst oder auf spezifische Messbedingungen zurückzuführen sind, die während der Durchführung der Probenmessungen herrschten.

Zeitliche Unabhängigkeit bedeutet, dass zwischen je zwei Messungen ein und desselben Probenmaterials ein zeitlicher Mindestabstand von mindestens 24 Stunden bestehen sollte, d.h. nach einem Messtag sollte nicht gleich ein weiterer Messtag folgen. Idealerweise sollten die Probenmessungen in unterschiedlichen Wochen durchgeführt werden. Außerdem sollte berücksichtigt werden, dass für jeden Tag der Probenmessungen eine neue Kalibrierung anzusetzen und das Messgerät erneut zu kalibrieren ist.

Repräsentativität der Messbedingungen bedeutet im Idealfall, dass die Messtage zufällig über das ganze Jahr verteilt werden. Diese Forderung dürfte in der Praxis kaum jemals erfüllbar sein, doch sollten die Probenmessungen auf insgesamt mindestens 6 unterschiedliche Messtage verteilt werden. Wenn nur eine WFR-Prüfung oder eine Gleichwertigkeitsteilprüfung vorgenommen werden soll, genügen bereits 4 unterschiedliche Messtage.

Ein Beispiel soll die genannten Anforderungen verdeutlichen: Sollen 5 Probenmaterialien in jeweils in einem Versuchsumfang von 4 Wiederholungsmessungen untersucht werden, ergeben sich insgesamt 20 Probenmessungen, die gemäß der Forderung der Repräsentativität auf mindestens 6 Messtage zu verteilen sind. Auf jeden Messtag entfallen somit 3 bzw. 4 Probenmessungen. Die 6 Messtage sind, um der Forderung der zeitlichen Unabhängigkeit gerecht zu werden, so auf mindestens zwei Wochen zu verteilen, dass nicht zwei Messtage aufeinander folgen. Schließlich ist noch zu gewährleisten, dass die Auswahl der Probenmaterialien zu den Messtagen dem Grundsatz der Randomisierung folgt. Der folgende Plan erfüllt die genannten Forderungen:

Tabelle 4: Zeitplan der für jede der 5 Probenmaterialien zu realisierenden 4 Wiederholungsmessungen. Dargestellt ist die laufende Nummer der Probenmessungen pro Probenmaterial.

	Probe A	Probe B	Probe C	Probe D	Probe E
01.01.2006 So					
02.01.2006 Mo	1		1	1	
03.01.2006 Di					
04.01.2006 Mi	2	1		2	1
05.01.2006 Do					
06.01.2006 Fr	3	2			2
07.01.2006 Sa					
08.01.2006 So					
09.01.2006 Mo		3	2	3	
10.01.2006 Di					
11.01.2006 Mi	4		3	4	3
12.01.2006 Do					
13.01.2006 Fr		4	4		4

Bei komplexen, zeitlich aufwändigen Analysenverfahren können die Messpausen zur Herstellung der Probenlösungen nach Anwendung von Extraktions- oder Aufschlussprozeduren verwendet werden. In jedem Falle sind die Probenlösungen frisch herzustellen und an dem darauf folgenden Tag zu messen.

3. Auswahl der Proben

Auf der Grundlage des Anwendungsbereichs des Referenzverfahrens sind die Untersuchungsproben oder ZRM repräsentativ hinsichtlich der Matrixzusammensetzung, die Gehaltsbereiche der Analyten sowie möglicher Störkomponenten auszuwählen. Die Auswahl der Probenmaterialien sollte so erfolgen, dass Proben mit Gehalten des oberen und unteren Konzentrationsbereichs enthalten sind. Weiterhin sind verschiedene Bodenarten zu berücksichtigen. Sandige, schluffige, tonhaltige und humusreiche Böden trocken oder feldfrisch können Extraktionsausbeuten stark beeinflussen. Sind für einige Analyten bestimmte Störkomponenten bekannt, so sollten auch diese in der Auswahl der Bodenproben vertreten sein.
Damit eine probenübergreifende Auswertung möglich ist, sollten die Konzentrationen nicht stärker als um den Faktor 10-50 variieren. Grundsätzlich sind ab einem Faktor 2-3 die Berechnungen mittels standardisierter Messwerte durchzuführen. Dazu sind vor Durchführung der statistischen Berechnungen alle Messwerte durch den jeweiligen Probenmittelwert der Referenzmessungen zu dividieren. Diese Standardisierung ändert nichts an den relativen Unterschieden bei den zufälligen und systematischen Abweichungen.

4. Standardversuchspläne

Zunächst ist festzulegen, ob die Gleichwertigkeit der beiden Verfahren nur bezüglich der Mittelwerte auf der Grundlage der Wiederfindungsrate, oder auch bezüglich der Wiederhol-, Intermediate- und Vergleichstandardabweichung überprüft werden soll. In der Regel ist der Aufwand zur Überprüfung der WFR-Gleichwertigkeit deutlich geringer als bei den weiteren Kriterien. Geeignete Standardversuchspläne für die unterschiedlichen Anforderungen werden in den folgenden Abschnitten vorgestellt.

4.1. WFR-Prüfung

Systematische Abweichung, WFR-Prüfung mit Referenzverfahren
Wenn auf eine Überprüfung der Gleichwertigkeit bezüglich Wiederhol-, Intermediate- und Vergleichstandardabweichung verzichtet werden kann, empfiehlt es sich bei Verwendung eines Referenzverfahrens, 6 Probenmaterialien an jeweils zwei unterschiedlichen Messtagen einfach zu bestimmen. Es handelt sich also um einen Versuch, bei dem jede der 6 Proben zweimal parallel gemessen wird. In der folgenden Tabelle ist ein geeigneter Zeitplan dargestellt. Es ist nicht erforderlich, die Messungen nach dem Referenz- und dem Alternativverfahren zugleich vorzunehmen. Allerdings kann eine solche parallele Abarbeitung die Arbeit u.U. deutlich erleichtern.

Tabelle 5: Standardversuchsplan zur WFR-Prüfung mit Referenzverfahren für 6 Proben mit je 2 Messwiederholungen an insgesamt 4 Messtagen. An jedem Messtag werden mit Lösungen des Referenz- und des Alternativverfahrens jeweils 3 Messungen durchgeführt. Die Tabelle zeigt jeweils die laufende Nummer der Probenmessung, d.h. der parallel durchgeführten Bestimmung.

	Probe A	Probe B	Probe C	Probe D	Probe E	Probe F
01.01.2006 So						
02.01.2006 Mo	1		1	1		
03.01.2006 Di						
04.01.2006 Mi		1		2	1	
05.01.2006 Do						
06.01.2006 Fr		2	2			1
07.01.2006 Sa						
08.01.2006 So						
09.01.2006 Mo	2				2	2

Bei komplexen, zeitlich aufwändigen Verfahren können die Messpausen zur Extraktion oder Aufschluss der Proben verwendet werden. In jedem Falle sind die Probenlösungen frisch herzustellen und an dem darauf folgenden Tag zu messen.

Sofern die Gleichwertigkeit nicht nachgewiesen werden kann, empfiehlt sich die Erweiterung der Versuchsreihe um z.B. 6 Messungen. Hierzu können entweder weitere Proben einbezogen werden oder die bereits eingesetzten Proben erneut gemessen werden. Sinnvoll erscheinen diese ergänzenden Messungen allerdings nur dann, wenn die relative Abweichung der Mittelwerte der beiden Verfahren nicht über 15%

liegt. Andernfalls ist zu befürchten, dass tatsächlich signifikante Unterschiede in den Mittelwerten bestehen, die größer als die maximal tolerierte Abweichung sind.

Wenn nach Durchführung des o.g. Standardversuchsplanes entschieden wird, nun auch eine vollständige Gleichwertigkeitsprüfung vorzunehmen, genügt es, den im folgenden Abschnitt beschriebenen Standardversuchsplan für eine Gleichwertigkeitsteilprüfung vorzunehmen. In der statistischen Auswertung sind dann die Ergebnisse beider Pläne zu kombinieren.

Systematische Abweichung, WFR-Prüfung mit Referenzmaterialien
Wenn geeignete (zertifizierte) Referenzmaterialien vorliegen, reduziert sich bei jeweils zweimaliger Wiederholbestimmung die mindestens erforderliche Probenmaterialanzahl von 6 auf 4, d.h. es sind dann 4 Probenmaterialien an jeweils zwei unterschiedlichen Messtagen jeweils einmal zu bestimmen. Die folgende Tabelle enthält einen entsprechenden Zeitplan.

Tabelle 6: Standardversuchsplan zur WFR-Prüfung mit (zertifiziertem) Referenzmaterial für 4 Probenmaterialien mit je 2 Wiederholungsmessungen an insgesamt 4 Messtagen. An jedem Messtag werden mit Lösungen des Referenz- und des Alternativverfahrens jeweils 2 Messungen durchgeführt. Die Tabelle zeigt jeweils die laufende Nummer der Probenmessung, die aus zwei parallel durchgeführten Messungen besteht.

	Probe A	Probe B	Probe C	Probe D
01.01.2006 So				
02.01.2006 Mo	1		1	
03.01.2006 Di				
04.01.2006 Mi		1		1
05.01.2006 Do				
06.01.2006 Fr		2	2	
07.01.2006 Sa				
08.01.2006 So				
09.01.2006 Mo	2			2

Sofern die Gleichwertigkeit nicht nachgewiesen werden kann, empfiehlt sich die Erweiterung der Versuchsreihe um z.B. 4 Messungen. Hierzu können entweder weitere Probenmaterialien einbezogen werden oder die bereits eingesetzten Probenmaterialien erneut gemessen werden. Sinnvoll erscheinen diese ergänzenden Messungen allerdings nur dann, wenn die relative Abweichung der Mittelwerte nicht über 15% liegt.

4.2. Gleichwertigkeitsteilprüfung

Wenn auf eine Überprüfung der Gleichwertigkeit bezüglich der Intermediate- oder Vergleichstandardabweichung verzichtet werden kann, empfiehlt sich entweder die Messung von 5 Probenmaterialien mit je 3 Wiederholungen oder die Messung von 8 Probenmaterialien mit je 2 Wiederholungen. In allen Probenmessungen ist eine Doppelbestimmung (Parallelproben) auszuführen. Die folgende Tabelle zeigt einen geeigneten Versuchsplan für 8 Probenmaterialien mit jeweils 2 Wiederholungen. Die

insgesamt 16 Probenmessungen werden innerhalb von 4 Messtagen absolviert, d.h. 4 Probenmessungen oder 8 Messungen pro Messtag.

Tab.: Standardversuchsplan zur Gleichwertigkeitsteilprüfung für 8 Proben mit je 2 Wiederholungen, bestehend aus jeweils 2 Parallelmessungen. Dargestellt sind die laufenden Nummern der Probenmessungen.

	Probe A	Probe B	Probe C	Probe D	Probe E	Probe F	Probe G	Probe H
01.01.2006 So								
02.01.2006 Mo	1		1	1	1			
03.01.2006 Di								
04.01.2006 Mi		1	2			1	1	
05.01.2006 Do								
06.01.2006 Fr		2			2		2	1
07.01.2006 Sa								
08.01.2006 So								
09.01.2006 Mo	2			2		2		2

Bei komplexen, zeitlich aufwändigen Verfahren können die Messpausen zur Extraktion oder Aufschluss der Proben verwendet werden. In jedem Falle sind die Proben frisch zu extrahieren und an dem darauf folgenden Tag zu messen.

Sofern die Gleichwertigkeit bezüglich eines der beiden Kriterien nicht nachgewiesen werden kann, empfiehlt sich die Erweiterung der Versuchsreihe um 12-20 Probenmessungen. Hierzu können entweder weitere Probenmaterialien einbezogen werden oder die bereits vorliegenden Probenmaterialien erneut gemessen werden. Sinnvoll erscheinen diese ergänzenden Messungen allerdings nur dann, wenn die relative Abweichung der Mittelwerte zwischen den beiden Verfahren nicht über 15% liegt und zugleich der Quotient der beiden Wiederholstandardabweichungen nicht größer als 2 ist.

4.3. Vollständige In-Haus-Gleichwertigkeitsprüfung

Bei einer vollständigen In-Haus-Gleichwertigkeitsprüfung empfiehlt sich entweder die Messung von 5 Probenmaterialien mit je 4 Wiederholungen oder die Messung von 8 Probenmaterialien mit je 3 Wiederholungen. In allen Probenmessungen sind jeweils Doppelbestimmugen auszuführen. Ein geeigneter Versuchsplan für 5 Probenmaterialien ist in Abschnitt 2.4 beschrieben. Sofern die Gleichwertigkeit bezüglich eines der Kriterien nicht nachgewiesen werden kann, empfiehlt sich die Erweiterung der Versuchsreihe analog zum Vorgehen bei einer Gleichwertigkeitsteilprüfung.

5. Referenzen

Bundesbodenschutz- und Altlastenverordnung, Bundesgesetzblatt Jahrgang 1999 Teil 1 Nr. 36, ausgegeben zu Bonn am 16. Juli 1999.

Uhlig, Steffen (2003), Empfehlungen zur Charakterisierung der Gleichwertigkeit und Vergleichbarkeit von Verfahren und Methoden. In: „Überprüfung von Methoden des Anhanges 1 der Bundesbodenschutz- und Altlastenverordnung (BBodSchV) zur Beurteilung der Bodenqualität“ (Teil 2). Umweltbundesamt Texte 37/03.

DIN ISO 5725 1-6: Genauigkeit (Richtigkeit und Präzision) von Messergebnissen.