

Homogenität und Stabilität von Ringversuchspräparaten des NLGA - Chemische Parameter -

Einleitung

Der Ausrichter eines Ringversuchs hat ein starkes Bestreben, Ringversuchsproben zu präparieren, die den Zweck des Ringversuchs in bestmöglichem Maß erfüllen. Dazu gehört insbesondere die Herstellung von Proben für die Teilnehmer, die untereinander (nahezu) identisch, sprich ausreichend homogen, und über einen hinreichend langen Zeitraum unverändert im jeweils zu bestimmenden Merkmal, sprich ausreichend stabil, sind. Um dies zu erreichen sind sorgsame Planungen und Durchführungen bei Dingen wie der Auswahl von Probengefäßen, der Auswahl von zu dotierenden Substanzen und Matrix, der Konservierung und Stabilisierung, der Präparation und Abfüllung sowie bei Lagerung und Versand der Proben unerlässlich.

Vor der Ausrichtung eines Ringversuchs werden zu den genannten Punkten üblicherweise entsprechende Vorversuche durch den Ausrichter durchgeführt, die die erforderliche Qualität der Ringversuchsproben, insbesondere im Hinblick auf ausreichende Homogenität und Stabilität bestätigen.

Homogenität und Stabilität aus der Sicht einschlägiger Normen

Die DIN EN ISO 17043 [1] fordert, dass jeder Teilnehmer vergleichbare Prüfgegenstände erhält und dass diese Prüfgegenstände während der gesamten Eignungsprüfung stabil bleiben. Es müssen Kriterien für geeignete Homogenität und Stabilität festgelegt werden, und sie müssen auf dem Einfluss basieren, den Inhomogenität und Instabilität auf die Bewertung der Leistung der Teilnehmer haben werden. Das Verfahren zur Prüfung auf Homogenität und Stabilität muss dokumentiert sein und, wo anwendbar, in Übereinstimmung mit geeigneten statistischen Modellen durchgeführt werden.

Die Auswahl einer geeigneten statistischen Methode ist für den Ausrichter von Ringversuchen nicht banal. Verschiedene Aspekte jeder statistischen Methode müssen im Hinblick auf den Zweck der hier betrachteten Prüfung berücksichtigt werden. Denkbar wäre eine statistische Methode, die rein die Werte der Qualitätskontrollen betrachtet (Trendtest, Ausreißertest, Test auf Normalverteilung ...) oder eine, die eine Varianz der Qualitätskontrollen mit den aus den Teilnehmerergebnissen erzielten Kennwerten des Ringversuchs vergleicht [2,3]. Zudem muss das statistische Modell gewährleisten, dass trotz Anwendung einer analytischen Methode mit hoher Präzision für die Qualitätskontrollmessungen deren analytische Varianz und die Varianz der Ringversuchsproben angemessen unterschieden werden und eine zutreffend richtige Aussage über Inhomogenität bzw. Instabilität macht [3]. Da es keine statistische Methode gibt, die dies zu 100% erfüllt, ist die Wahl der statistischen Methode stets ein Kompromiss, und die Ergebnisse müssen - insbesondere im Hinblick auf Auswirkungen bei der Bewertung der Leistung der Teilnehmer - dahingehend interpretiert werden.

Die DIN EN ISO 17043 schreibt oder schlägt (bewusst) kein statistisches Modell vor, sondern verweist diesbezüglich auf andere Normen wie z.B. die ISO 13528 [2].

Unter Umständen können Prüfgegenstände, die nicht ausreichend homogen oder stabil sind, die besten verfügbaren sein. Ist dies der Fall, so ist das Ausmaß an Inhomogenität bzw. Instabilität als zusätzliche Komponente bei der Bewertung der Teilnehmerergebnisse zu berücksichtigen [1].

Die Ringversuche des NLGA werden in der Matrix Wasser durchgeführt. Dabei ist an sich davon auszugehen, dass bei sorgsamer Präparation der Proben diese ausreichend homogen sind. Dass dies zutrifft und zudem ausreichende Stabilität vorliegt, konnte anhand einer umfangreichen Qualitätskontrollstudie, die zu den chemischen Ringversuchspräparaten des RV 1-2013 durchgeführt wurde, bestätigt werden. Die Ergebnisse der Studie sind im folgenden Kapitel zusammengefasst aufgeführt.

Nachweis von Homogenität und Stabilität chemischer Ringversuchspräparate des NLGA

Während des Ringversuchs 1-2013 wurden umfangreiche Qualitätskontrollmessungen der Ringversuchspräparate durchgeführt. Dabei wurden das in ISO 13528, Anhang B [2] beschriebene Verfahren und die beschriebene statistische Methode angewandt. Parallel erfolgte zusätzlich eine statistische Auswertung hinsichtlich Homogenität nach dem im Harmonized Protocol for proficiency testing [3] beschriebenen Verfahren. Letztgenannte berücksichtigt stärker die analytische Varianz einer Methode und erkennt analytische Ausreißer. Andererseits scheint die Statistik nach ISO 13528 etwas sensitiver zu reagieren.

Zu jedem Parameter und jeder Niveaugruppe wurden gleichmäßig in die Abfüllreihenfolge der Teilnehmerproben jeweils 10 QK-Proben für die Homogenitätsmessungen und jeweils 2x 5 QK-Proben für die Stabilitätsmessungen abgefüllt. Die Messungen auf Homogenität erfolgten am Tag der Präparation und Abfüllung. Die Messungen auf Stabilität erfolgten am 1. und am 4. Tag nach Versand der Ringversuchsproben an die Teilnehmer. Jede Probe wurde geteilt und als Doppelbestimmung gemessen. Doppelbestimmungen ermöglichen, ggf. einen analytischen Ausreißer innerhalb einer Messreihe leichter zu identifizieren.

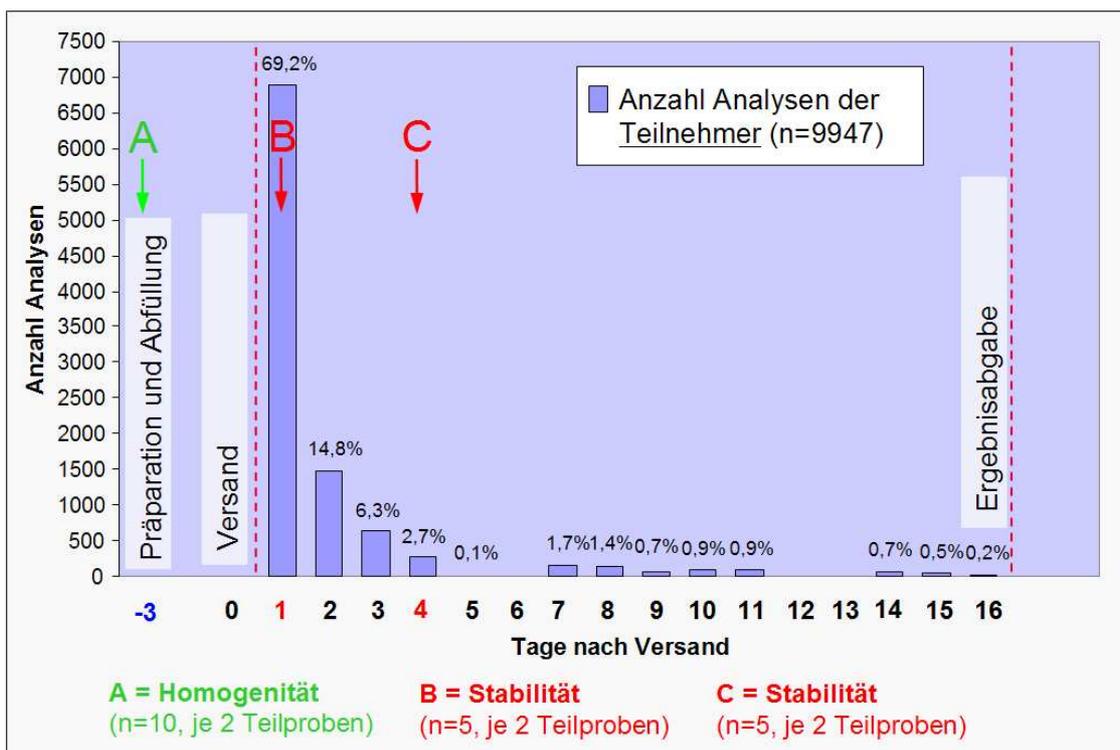


Abb. 1: Zeitlicher Ablauf der Homogenitäts- und Stabilitätsmessungen im RV 1-2013

Die Balken in Abb. 1 zeigen die Anzahl der Teilnehmer-Analysen im zeitlichen Verlauf der Ringversuche 2-2011 bis 1-2013. Die Frist zur Ergebnisabgabe eines Ringversuchs des NLGA läuft i.d.R. 16 Tage nach Versanddatum ab. Die Teilnehmer sind durch die Rahmenbedingungen aufgefordert, die Proben unmittelbar nach deren Erhalt zu analysieren. Abb. 1 zeigt, dass am Tag der 2. Stabilitätsmessungen (4. Tag nach Versand) im Durchschnitt bereits rund 93% aller Analysen durch die Teilnehmer abgeschlossen sind.

Kriterien für Homogenität und Stabilität nach ISO 13528

$$s_s \leq 0,3 \sigma$$

Homogenitätskriterium

s_s ... Standardabweichung der Probenmittelwerte (Berechnung: s. [2])

σ ... Vergleichsstandardabweichung im Ringversuch

Die Standardabweichung zwischen den Proben s_s berechnet sich für die wie oben beschrieben ermittelten Messwerte nach ISO 13528 wie folgt:

$$s_s = \sqrt{s_x^2 - (s_w^2/2)}$$

s_x ... Standardabweichung der Probenmittelwerte (Berechnung: s. [2])

s_w ... Standardabweichung innerhalb der Proben (Berechnung: s. [2])

$$|x - y| \leq 0,3 \sigma$$

Stabilitätskriterium

x ... Mittelwert der Homogenitätsmessungen (Berechnung: s. [2])

y ... Mittelwert der Stabilitätsmessungen (Berechnung: s. [2])

Für die Homogenitätsbetrachtung wird die Standardabweichung zwischen den Proben s_s mit den Ergebnissen der Teilnehmer verglichen (σ). Es liegt ausreichend Homogenität vor, wenn s_s das 0,3-fache der Vergleichsstandardabweichung nicht übersteigt.

Für die Stabilitätsbetrachtung wird die absolute Differenz zwischen den Mittelwerten aus Homogenität und Stabilität mit den Ergebnissen der Teilnehmer verglichen (σ). Es liegt ausreichend Stabilität vor, wenn $|x - y|$ das 0,3-fache der Vergleichsstandardabweichung nicht übersteigt.

Nachfolgend ist die Auswertung der QK-Werte exemplarisch am Beispiel Trübung (Gruppe C) aufgeführt:

Trübung (Gruppe C)															
Homogenitätstest (ISO 13528)					Stabilitätstest (ISO 13528)				Homogenitätstest (Harmonized Protocol 2006 (IUPAC))						
Messdatum: 18.01.13					Messdatum: 22.01.13										
	Wert 1	Wert 2	Probenmittelwert \bar{x}_i	Spannweite w_i	Wert 1	Wert 2			D=a-b	S=a+b	D ² =(a-b) ²				
Probe 1	✓	1,49	1,48	1,485	0,01	✓	1,46	1,47	1,465	0,01	2,97	0,0001			
Probe 2	✓	1,49	1,49	1,49	0	✓				0	2,98	0			
Probe 3	✓	1,5	1,5	1,5	0	✓	1,46	1,46	1,46	0	3	0			
Probe 4	✓	1,5	1,49	1,495	0,01	✓				0,01	2,99	0,0001			
Probe 5	✓	1,49	1,49	1,49	0	✓	1,46	1,45	1,455	0	2,98	0			
Probe 6	✓	1,48	1,48	1,48	0	✓				0	2,96	0			
Probe 7	✓	1,5	1,5	1,5	0	✓	1,45	1,47	1,46	0	3	0			
Probe 8	✓	1,48	1,49	1,485	0,01	✓				-0,01	2,97	0,0001			
Probe 9	✓	1,5	1,48	1,49	0,02	✓	1,45	1,44	1,445	0,02	2,98	0,0004			
Probe 10	✓	1,49	1,5	1,495	0,01	✓				-0,01	2,99	0,0001			
Probe 11	✓					✓									
Probe 12	✓					✓									
Anzahl der Proben				10					Höchstes D ²	0,0004					
Allgemeiner Mittelwert $\bar{x}_..$				1,491				Allgemeiner Mittelwert $\bar{y}_..$	1,457		Summe D ²	0,0008			
Standardabw. der Probenmittelwerte s_x				0,006583				$x_{i..} - y_{i..}$	0,034		2,3 %	Cochran's test C	0,5000000		
Standardabw. innerh. der Proben s_w				0,006325								krit. Wert (99%)	0,718	no outlying pair	
Standardabw. zwischen den Proben s_s				0,004830	0,3 %							krit. Wert (95%)	0,602	no outlying pair	
Kenndaten im Ringversuch												Test auf ausreichende Homogenität:			
Mittelwert nach Hampel				1,427								Varianz der Summen S (V_S)	0,0001733		
Vergleichsstandardabweichung σ				0,176 NTU								Analytische Varianz s_{an}^2	0,000040		
Beurteilungskriterium Homogenität ($s_s \leq 0,3\sigma$)				0,0528				Beurteilungskriterium Stabilität ($ x_{i..} - y_{i..} \leq 0,3\sigma$)	0,0528			Varianz der Proben s_{sam}^2	0,000023		
Beurteilung:								Beurteilung:	✓			\hat{s}	2,982		
Die Proben sind				hinreichend homogen.				Die Proben sind	hinreichend stabil.			Vergleichsstandardabweichung	0,176 mg/l		
Bemerkung:												Erlaubte Varianz zwischen Proben σ_{all}^2	0,0027878	Kritischer F1	1,88
												Kritischer Wert	0,0052815	Kritischer F2	1,01
												Die Proben sind	hinreichend homogen.		

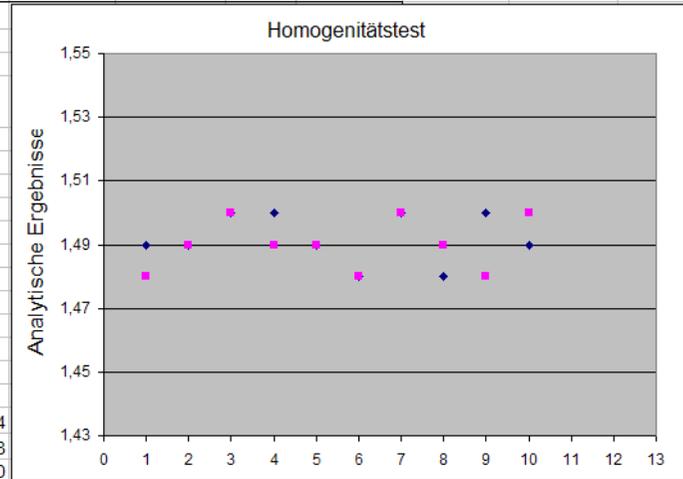


Abb. 2: Excel-sheet Auswertung Homogenität und Stabilität (Beispiel: Trübung, Gruppe C, RV 1-2013)

Auswertung

Das nach ISO 13528 herangezogene Kriterium ($0,3 \sigma$) stellt - selbst bei einer angenommenen „absoluten“ Homogenität bzw. Stabilität der Proben - analytisch eine sehr hohe Anforderung für das Labor, welches die QK-Messungen durchführt, dar. Die Vergleichsstandardabweichung σ der Teilnehmerergebnisse wird nach robuster Statistik ermittelt. Dabei fallen Ausreißer im Teilnehmerfeld somit kaum ins Gewicht bei der Ermittlung von σ , was bei einem analytisch sehr gut aufgestellten Teilnehmerfeld, oder einem Parameter, der analytisch sehr präzise bestimmt werden kann, in niedrigen Vergleichsstandardabweichungen resultiert. Die Analytik der QK-Proben erfordert daher Methoden mit hoher Präzision, um das gestzte Kriterium schlüssig beurteilen zu können .

Tabelle 1 zeigt die zusammenfassende Übersicht der statistischen Auswertung der Qualitätskontrollmessungen im RV 1-2013. Sowohl das statistische Modell nach ISO 13528 als auch das nach dem IUPAC Report (2006) belegen ausreichend Homogenität für alle Parameter in jeder Gruppe (A, B und C).

Parameter	Gruppe	Homogenität (nach ISO 13528)	Homogenität (Harmon. Protocol 2006 IUPAC)	Stabilität (+1 Tag) (nach ISO 13528)	Stabilität (+4 Tage) (nach ISO 13528)
Aluminium	A	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
	B	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
	C	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
Eisen	A	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
	B	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
	C	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
Kupfer	A	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
	B	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
	C	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
Mangan	A	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
	B	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
	C	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
Ammonium	A	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
	B	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
	C	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
Nitrat	A	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
	B	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
	C	hinreichend homogen.	hinreichend homogen.	nicht hinreichend stabil.	nicht hinreichend stabil.
Nitrit	A	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
	B	hinreichend homogen.	hinreichend homogen.	nicht hinreichend stabil.	nicht hinreichend stabil.
	C	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
pH-wert	A	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
	B	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
	C	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
Leitfähigkeit	A	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
	B	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
	C	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
Trübung	A	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
	B	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
	C	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
Färbung	A	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
	B	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
	C	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
Oxidierbarkeit	A	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
	B	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.
	C	hinreichend homogen.	hinreichend homogen.	hinreichend stabil.	hinreichend stabil.

Tab. 1: Übersicht der statistischen Auswertung der QK-Messungen im RV 1-2013

Die Auswertung der Stabilitätsmessungen zeigt sowohl am 1. als auch am 4. Tag nach Versand ausreichend Stabilität für alle Parameter in jeder Gruppe (A, B und C) der

Ringversproben. Davon ausgenommen sind die Gruppe C beim Parameter Nitrat und die Gruppe B beim Parameter Nitrit.

Instabilität wäre an sich in allen Gruppen der betreffenden Parameter zu erwarten. Hierbei zeigt sich jedoch, dass der statistische Ansatz in diesen Fällen an seine Grenzen hinsichtlich der Beurteilung ausreichender Stabilität stößt. Bei der Betrachtung der Stabilität muss unter Wiederholbedingung mit einer hohen Richtigkeit gemessen werden. Kommt es dabei zu Schwankungen zwischen den Mittelwerten der Homogenitäts- und der Stabilitätsmessungen, die nicht auf die Zusammensetzung der Proben sondern auf die Analytik zurückzuführen sind (bei gleichzeitig niedriger σ), kommt es leicht zu Fehlinterpretationen hinsichtlich der Stabilität von Proben.

Anhand der nachfolgenden näheren Betrachtung von Nitrit Gruppe B soll belegt werden, dass die lt. Statistik vorliegende Instabilität keinerlei Auswirkungen bzw. Einfluss auf die Bewertung der Teilnehmer hat.

RV-Kennwerte	mg/l		
Sollwert (Einw.)	0,31		
HAMPEL	0,307		
Vergl.stdabw (σ)	0,01233	(4,01%)	
AU	0,277	Limitiert auf 5%	
AO	0,339		
Mittelwert Homogenität			
x	0,3116		
Mittelwerte Stabilität		x - y _i	
y ₁	0,3077	0,0039	
y ₂	0,3162	0,0046	
Stabilitätskriterium:		0,0037	(0,3 σ)

Tab. 2: RV-Kennwerte und statistische Auswertung Homogenität/Stabilität Parameter Nitrit (Gruppe B), RV 1-2013

Aus der Vergleichsstandardabweichung σ von 4,01% resultiert, dass die abs. Abweichung der Mittelwerte der QK-Messungen nicht größer als 1,2% ($0,3 \times 4,01\%$) betragen ($\approx 0,00372$ mg/l) darf. Dies ist analytisch extrem anspruchsvoll.

Die angewandte Statistik kann in solchen Fällen offensichtlich nicht sinnvoll angewandt werden und stößt an die Grenzen ihrer Eignung für die betrachtete Prüfung. Trotz einer lt. Statistik scheinbar vorliegenden Instabilität, hat diese keine Auswirkung auf die Beurteilung der Teilnehmer. Dies belegt auch Abbildung 3 die die Ergebnisse der Teilnehmer und die Ergebnisse der QK-Messungen gegenüber stellt.

Hinzu kommt, dass die für die statistische Betrachtung von Homogenität und Stabilität herangezogene Vergleichsstandardabweichung σ dann, wenn sie zu niedrig ist, für die Festlegung der Ausschlussgrenzen bei der Auswertung des Ringversuchs auf 5% limitiert wird. Damit werden die Beurteilungsgrenzen bei der Bewertung der Teilnehmer noch größer (s. auch Abb. 3).

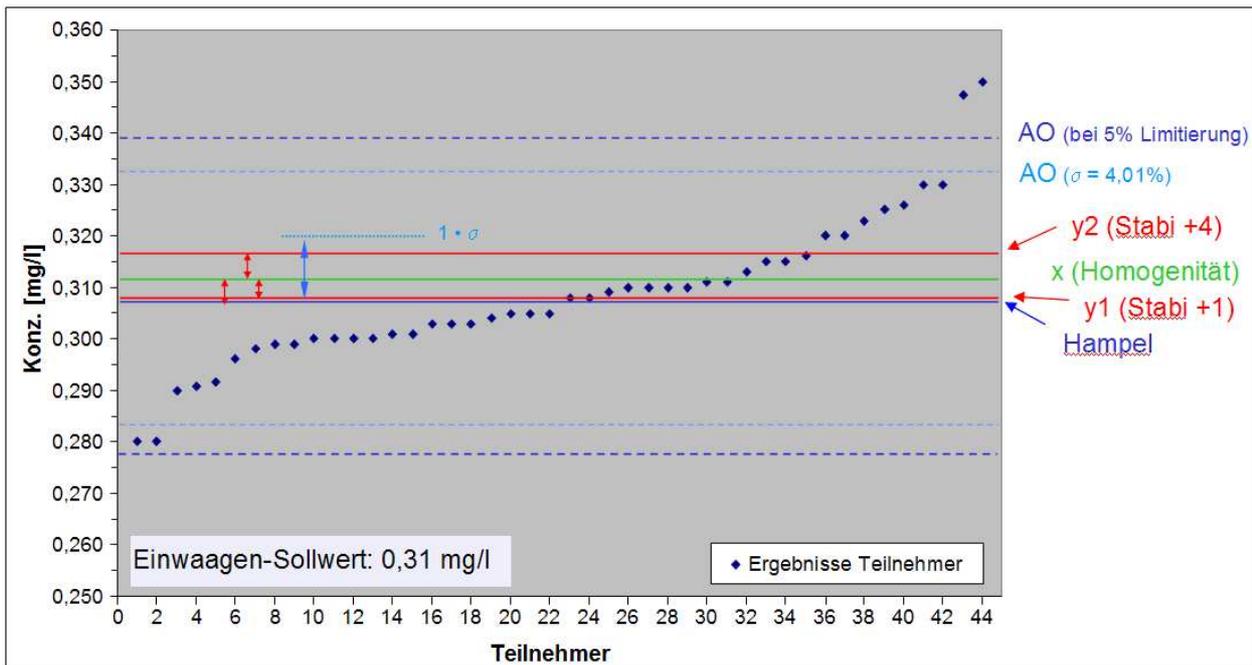


Abb. 3: Graphische Darstellung der RV-Ergebnisse und der QK-Messungen Parameter Nitrit (Gruppe B), RV 1-2013

Schlussfolgerung

Die Qualitätskontrollmessungen der chemischen Ringversuchspräparate zu RV 1-2013 haben gezeigt, dass die Proben ausreichend homogen und stabil sind. Bei identischer Probenpräparation kann für das betroffene Parameterspektrum davon ausgegangen werden, dass keine Inhomogenität und Instabilität in den Ringversuchsproben des NLGA vorliegen, die einen Einfluss auf die Bewertung der Teilnehmer haben könnten.

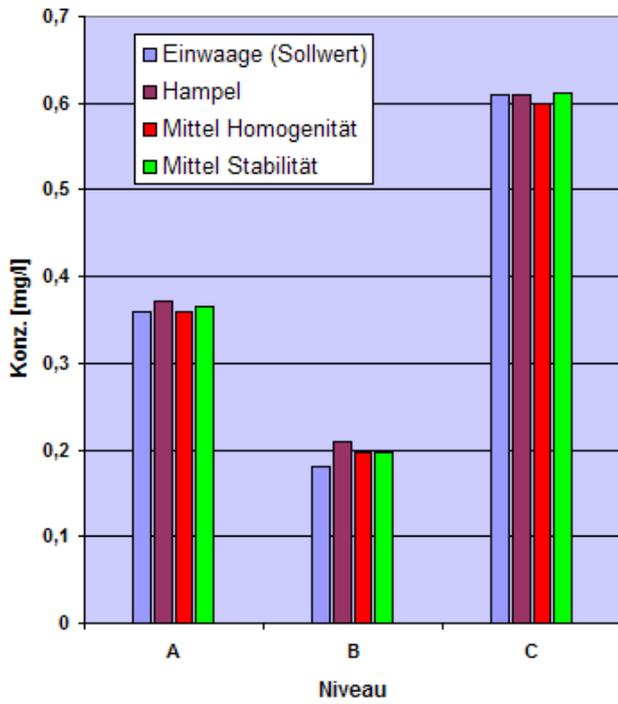
Graphische Darstellungen von Sollwert (Einwaage), HAMPEL und den QK-Mittelwerten für Homogenität und Stabilität aller Parameter und Gruppen des RV 1-2013 sind im Anhang abgebildet. Insbesondere bei Parametern wie z.B. den Metallen oder den Stickstoffen, deren Sollwerte durch exakte Einwaagen definiert werden können, ist eine sehr gute Wiederfindung in Form des HAMPEL zu erkennen. Dies ist - unabhängig von den Ergebnissen der QK-Messungen - ebenfalls ein Beleg für Homogenität und Stabilität der Ringversuchspräparate.

Literatur:

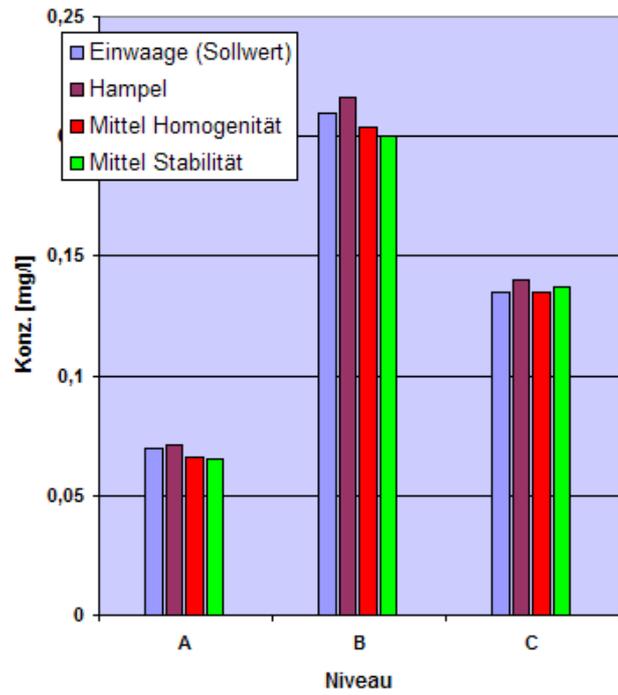
- [1] ISO/IEC 17043:2010, *Konformitätsbewertung – Allgemeine Anforderungen an Eignungsprüfungen*
- [2] ISO 13528:2005, *Statistical methods for use in proficiency testing by interlaboratory comparisons*
- [3] Thompson M., Ellison S.L.R., Wood R., "The International Harmonized Protocol for the proficiency testing of analytical chemistry laboratories" (IUPAC Technical Report), in *Pure and Applied Chemistry*, Vol. 78, No. 1, pp. 145-196, 2006
- [4] ISO Guide 34, *General requirements for the competence of reference material producers*
- [5] ISO Guide 35, *Reference materials — General and statistical principles for certification*

Anhang: Darstellungen von Sollwert (Einwaage), HAMPPEL und den QK-Mittelwerten für Homogenität und Stabilität

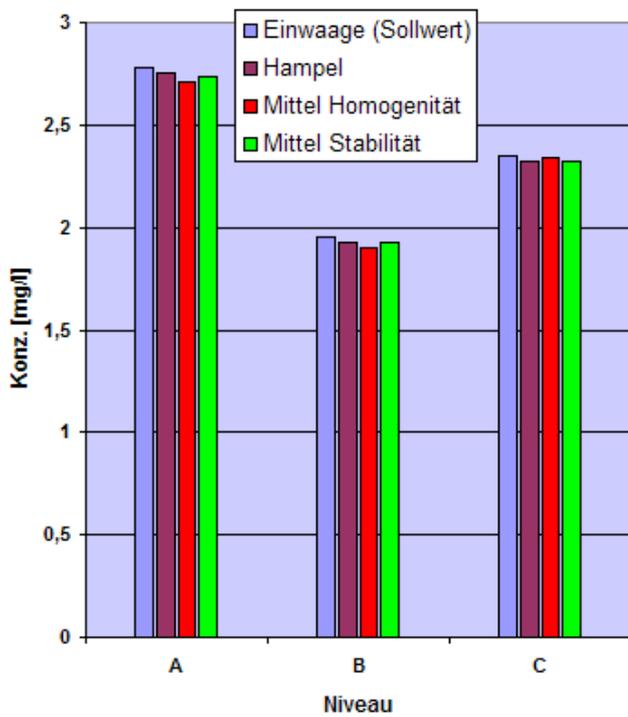
Aluminium



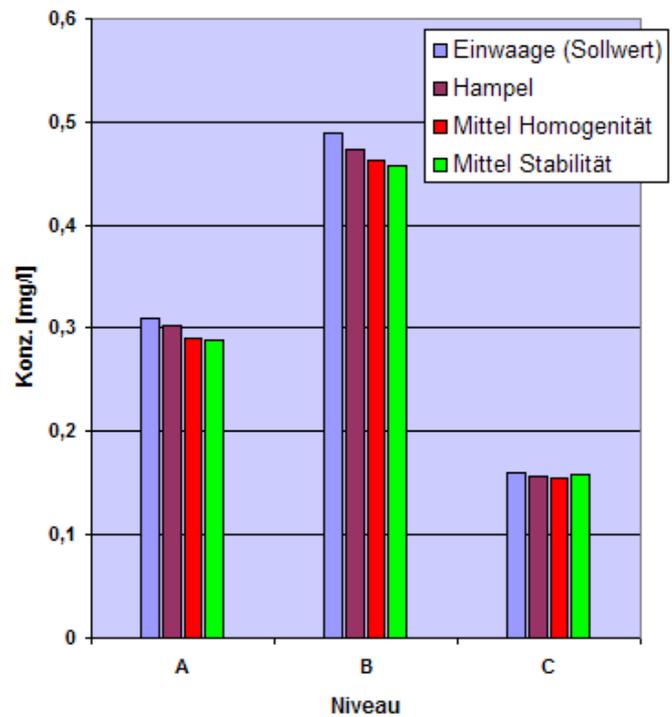
Eisen

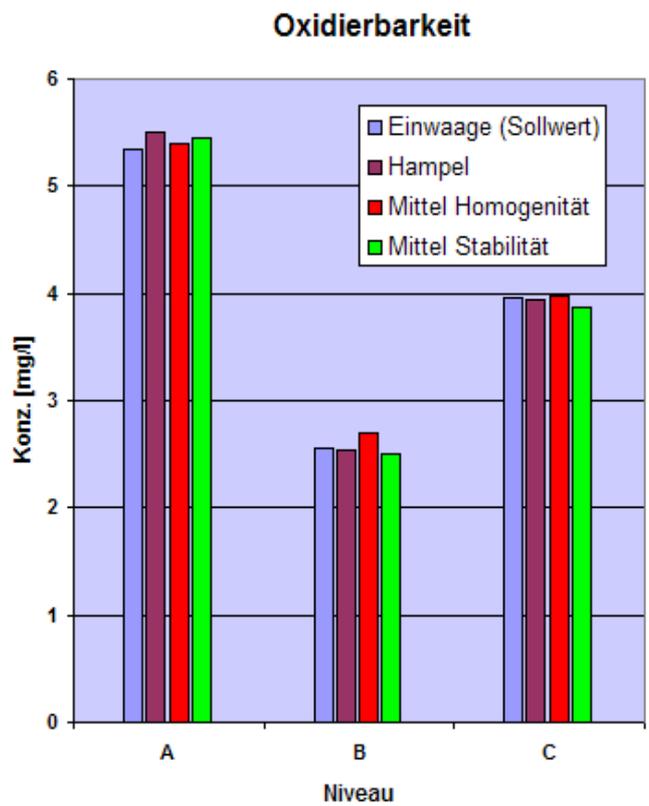
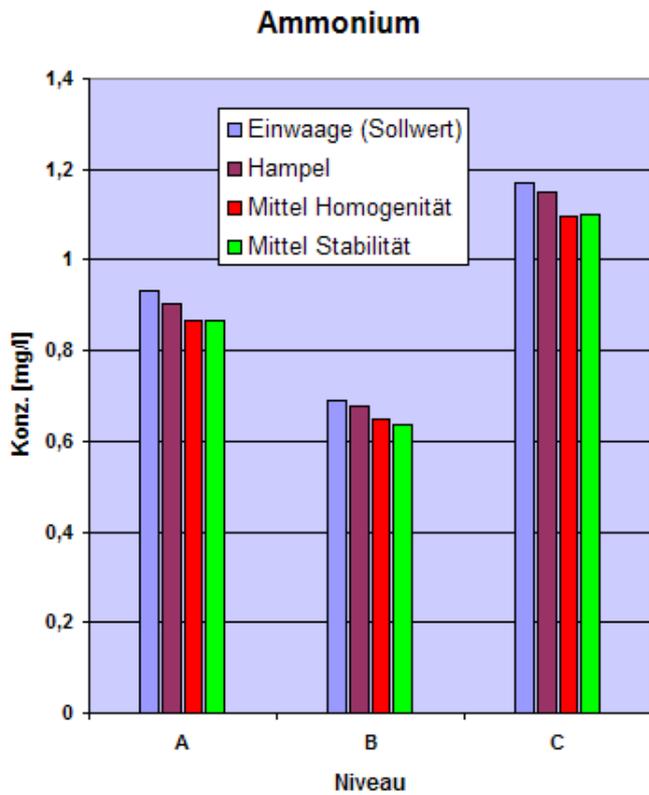
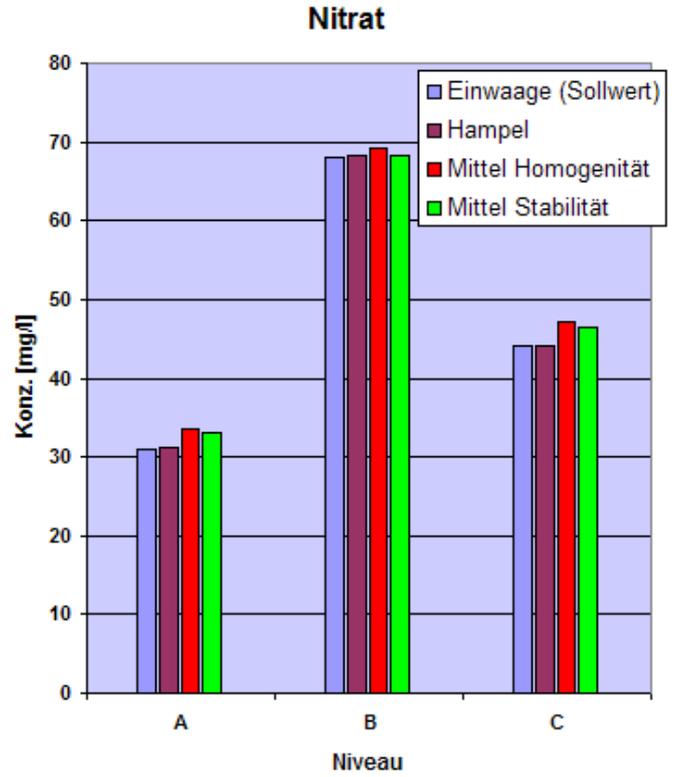
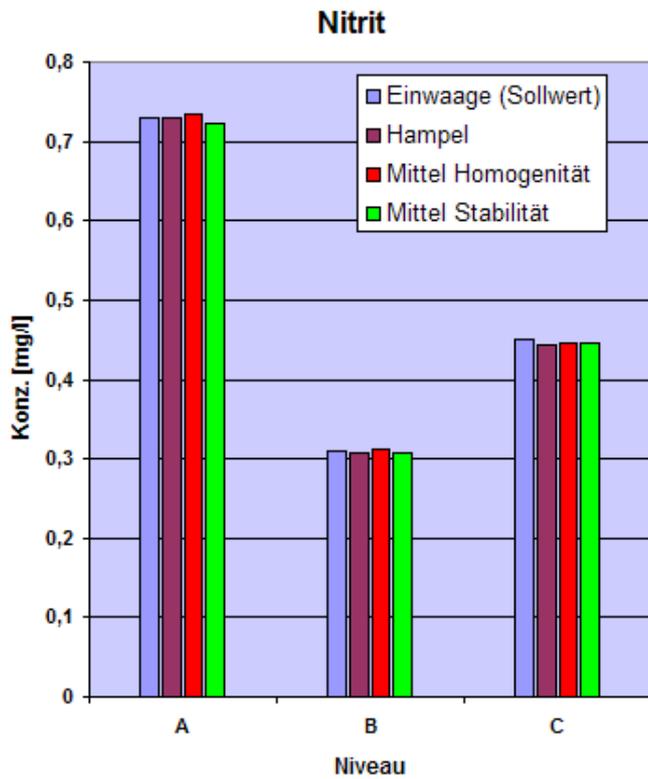


Kupfer

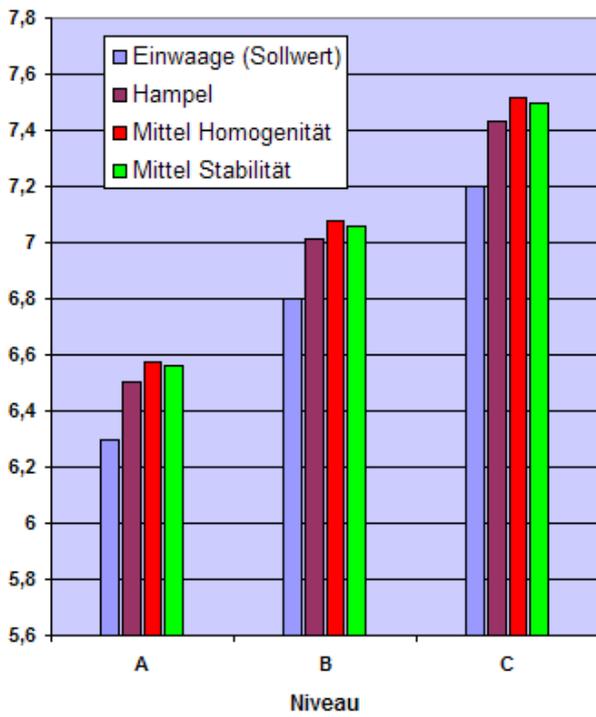


Mangan

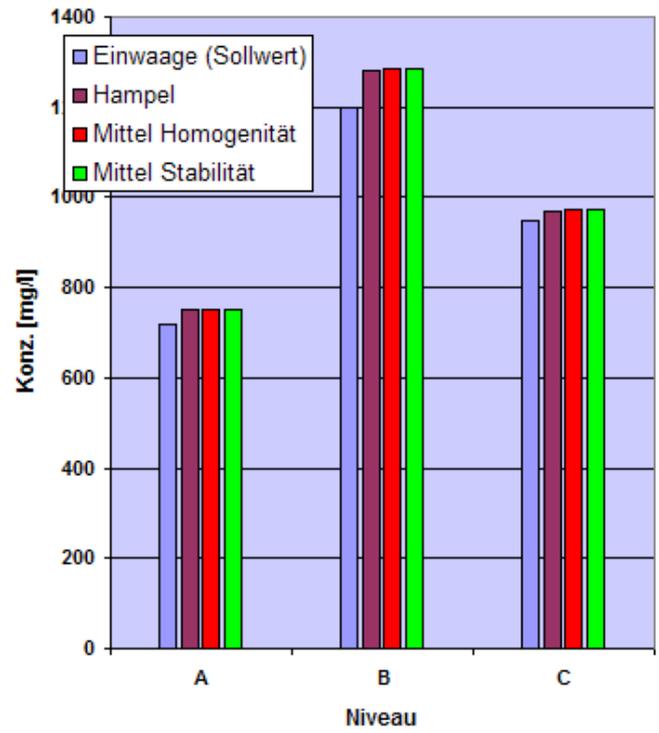




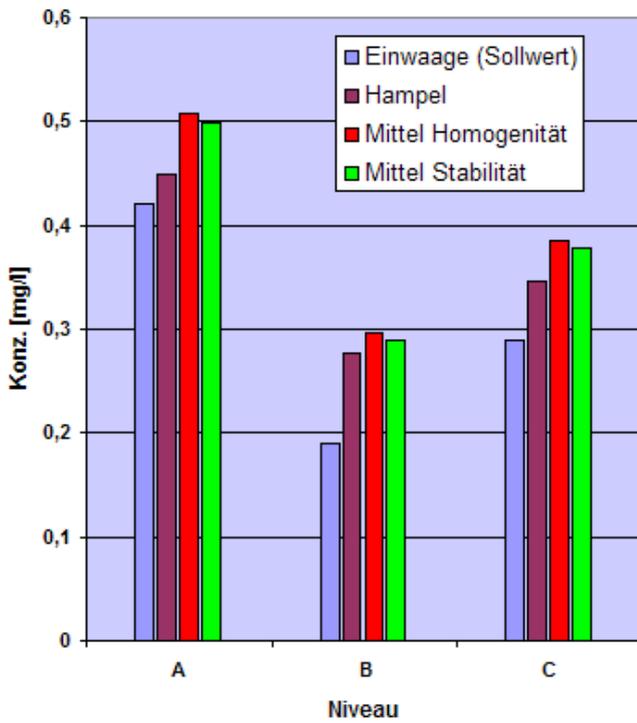
pH-Wert



el. Leitfähigkeit



Färbung



Trübung

