

# ILVO 367 Leeswijzer meetonzekerheid geaccrediteerde verrichtingen (versie april 2024)

*In dit stuk wordt ingegaan op het nut en gebruik van de meetonzekerheid. En op de wijze waarop de grootte van de meetonzekerheid wordt onderzocht in het periodiek uitgevoerde InterLaboratorium Vergelijkend Onderzoek (ILVO).*

## 1 De meetonzekerheid, nauwkeurigheid en precisie

Een meting van een bepaald kenmerk (bijvoorbeeld een laagdikte) geeft een schatting van de werkelijke waarde van dat kenmerk. Met schatting wordt bedoeld dat er een onzekerheid is rond de gemeten waarde. Die onzekerheid wordt aangeduid met de termen nauwkeurigheid (accuracy) en precisie (precision).

De nauwkeurigheid is een maat voor de systematische afwijking van de meetmethode. Het ringonderzoek geeft alleen inzicht in de **precisie**, dat wil zeggen in de mate van de toevallige variatie in meetuitkomsten als die herhaald door willekeurige laboratoria wordt uitgevoerd. De grootte van deze toevallige variatie wordt uitgedrukt in standaardafwijkingen  $s_r$  en  $s_R$  voor respectievelijk de herhaalbaarheid en de reproduceerbaarheid.

De **meetonzekerheid** is als volgt gedefinieerd: indien een willekeurig geaccrediteerd laboratorium een schatting van de werkelijkheid geeft, zal deze werkelijkheid met een betrouwbaarheid van 95% liggen in het interval "meetresultaat  $\pm$  meetonzekerheid". Hierbij wordt de meetonzekerheid berekend als:

$$\text{meetonzekerheid} = 2 \times s_R$$

De standaardafwijking  $s_R$  is een maat voor de spreiding van de reproduceerbaarheid, d.w.z. voor de variatie in resultaten als een willekeurig laboratorium een meetwaarde bepaalt.

Het is van belang dat de meetonzekerheid beheerst wordt. Voor bepaalde verrichtingen zijn hiervoor eisen gesteld en kan worden getoetst of de meetonzekerheid aan die eisen voldoet. Voor andere verrichtingen is het van belang dat de meetonzekerheid in de tijd niet groter wordt. Voor de beheersing van meetonzekerheid wordt periodiek een ringonderzoek uitgevoerd.

De meetonzekerheid mag niet toegepast worden als een marge op de eisen zoals die bijvoorbeeld in de "Standaard RAW Bepalingen" zijn opgenomen. Deze eisen gelden namelijk voor de meetwaarde zelf.

## 2 De rol van het InterLaboratorium Vergelijkend Onderzoek (ILVO)

Tussen de geaccrediteerde laboratoria binnen de sector Bouw worden regelmatig ringonderzoeken gehouden, het zogenoemde InterLaboratorium Vergelijkend Onderzoek (ILVO). Jaarlijks wordt 3 maal een pakket monsters verstuurd naar de deelnemende laboratoria. Op basis van de resultaten van de laatste 12 pakketten wordt via statistische analyses de spreiding in de meetresultaten geanalyseerd. Het belangrijkste resultaat van die analyses wordt gevormd door twee maten voor meetspreiding:

- $s_r$  = de **standaardafwijking voor de herhaalbaarheid** en
- $s_R$  = de **standaardafwijking voor de reproduceerbaarheid**

Hierbij verwijst het onderscheid tussen herhaalbaarheid en reproduceerbaarheid naar de variatie die optreedt als metingen worden uitgevoerd onder herhaalbaarheidsomstandigheden, dan wel reproduceerbaarheidsomstandigheden.<sup>1</sup>

Met behulp van deze standaardafwijkingen  $s_r$  en  $s_R$  worden – naast de bovengenoemde meetonzekerheid – de waarden voor de **herhaalbaarheid r** en **reproduceerbaarheid R** berekend via respectievelijk  $r = 2\sqrt{2} \times s_r$  en  $R = 2\sqrt{2} \times s_R$ .

De waarde voor de reproduceerbaarheid wordt gepubliceerd in document “ILVO 196.pdf”. De meest recente versie van dit document wordt via de website van Fenelab gepubliceerd op de volgende internetlocatie: <https://fenelab.nl/sectoren/bouw/>.

Onderstaande figuur toont een fragment van een eerdere versie van dit document. De waarden voor de reproduceerbaarheden volgens het ILVO ringonderzoek in de laatste kolom zijn vastgesteld in augustus 2021 en ze blijven gehandhaafd zolang zij binnen het 95% betrouwbaarheidsinterval van recentere resultaten liggen.

ILVO 196h

Reproduceerbaarheid ( $2\sqrt{2} \times S_R$ )

Code	Soort bepaling	Norm	Nauwkeurigheid	Reproduceerbaarheid volgens norm	Reproduceerbaarheid volgens ILVO
	ASFALTPROEFSTUK 1.1				
1 – 1	Bitumengehalte	NEN-EN 12697-1		0,5	0,2
1 – 2	Dichtheid mengsel	NEN-EN 12697-5		22	9
1 – 3	Dichtheid proefstuk	NEN-EN 12697-6		32	11

Figuur 1. Deel van de tabel met reproduceerbaarheidswaarden.

Indien voor een verrichting een eis is gesteld aan de reproduceerbaarheid, wordt de in het ringonderzoek gerealiseerde reproduceerbaarheid vergeleken met die norm. Bij het niet voldoen aan de norm worden vanuit de commissie Bouw van Fenelab acties uitgezet.

De waarde voor de herhaalbaarheid  $r$  wordt bij de statistische analyses wel berekend en gerapporteerd, maar de ILVO werkgroep heeft besloten dat aan de berekende herhaalbaarheid geen waarde moet worden gehecht. De reden hiervoor is dat de in de opzet en werkwijze van de bepalingen geen echte herhaalbaarheid wordt bepaald zoals gedefinieerd in de voetnoot<sup>1</sup>. De berekende herhaalbaarheid bevat namelijk ook variatie ten gevolge van spreiding tussen monsters (binnen een ronde) en eventuele variatie tussen apparaten en/of laboranten binnen een laboratorium.

### 3 De opzet van het ILVO – ringonderzoek

#### 3.1 Pakketten en clusters van variabelen

Per jaar wordt 2 tot 3 maal een pakket monsters verstuurd naar momenteel 16 laboratoria. Voor elk pakket moet een aantal variabelen worden gemeten. Niet alle laboratoria voeren metingen uit aan alle variabelen. Daarom zijn de variabelen ingedeeld in clusters, waarin in principe steeds dezelfde laboratoria meedoen. Het komt echter voor dat een laboratorium bepaalde verrichtingen niet uitvoert en

<sup>1</sup> Herhaalbaarheidsomstandigheden zijn: hetzelfde monster, hetzelfde laboratorium en meetsysteem, dezelfde laborant, dezelfde testomstandigheden. Reproduceerbaarheidsomstandigheden zijn: hetzelfde monster, ander laboratorium of meetsysteem, andere laborant

dus geen meetwaarden levert voor de betreffende variabele. Een overzicht van alle variabelen per cluster is opgenomen in Bijlage 2. Deze bijlage geeft de stand van zaken weer op het moment van schrijven van deze leeswijzer en kan van tijd tot tijd wijzigen. Een actueel overzicht is verkrijgbaar via de sector Bouw.

De frequentie van het aantal pakketten per jaar kan per cluster verschillen en ook binnen een cluster kan de frequentie verschillen. Een en ander is afhankelijk van de resultaten binnen het ringonderzoek. Een overzicht van de huidige indeling is weergegeven in Tabel 1.

Cluster	Aantal pakketten per jaar	Aantal laboratoria
1	3	5 tot 7
5	2 tot 3	6
8	3	3 tot 8
9	3	2 tot 8
10	3	4 tot 6
11	2	2 tot 4
12	3	4 tot 5
13	3	2 tot 3
14	3	2 tot 3
15	3	2 tot 11
16	3	3

Tabel 1. Overzicht van de aantallen pakketten en laboratoria per cluster.

De pakketten bestaan uit monsters die zeer zorgvuldig zijn samengesteld. Zo zorgvuldig dat verwacht mag worden dat de berekende standaardafwijkingen voor de herhaalbaarheid en reproduceerbaarheid kunnen worden toegeschreven aan de uitvoering van de meetmethode en niet aan materiaalspreiding (ILVO 192).

### 3.2 Numerieke en niet-numerieke meetwaarden

De meeste verrichtingen leveren numerieke meetwaarden op. Voor deze verrichtingen kan de mate van overeenstemming (precisie) worden berekend in termen van standaardafwijkingen  $s_r$  en  $s_R$  voor de herhaalbaarheid en reproduceerbaarheid.

Daarnaast zijn er in Cluster 15 ook niet-numerieke meetwaarden, zoals bijvoorbeeld variabele “1a soort laag 1” waar de asfaltsoort in termen van bijvoorbeeld DAB 0/11 wordt aangegeven. Voor dergelijke variabelen kan niet worden gewerkt met standaardafwijkingen, maar moet een alternatieve methode voor het kwantificeren van de mate van overeenstemming worden gevolgd. Aan deze methode wordt momenteel gewerkt.

In het vervolg van deze leeswijzer wordt uitgegaan van numerieke meetwaarden.

### 3.3 Meetresultaten per variabele (verrichting)

Voor elke variabele (verrichting) wordt per pakket slechts één meetresultaat bepaald door de deelnemende laboratoria. Een voorbeeld hiervan – in dit geval voor variabele “12\_Zand\_Dichtheid\_Rho\_Rd” binnen cluster 9 – is weergegeven in Tabel 2.

lab_nr	19	20	24	25	29	Mean
pakket						
2016-3	*	2.62	2.66	2.62	2.63	2.63
2017-1	2.65	2.62	2.64	2.62	2.63	2.63
2017-2	2.63	2.63	*	2.62	2.63	2.63
2017-3	2.60	2.62	2.65	2.63	2.63	2.63
2018-1	2.63	2.61	2.60	2.62	2.63	2.62
2018-2	2.63	2.63	2.61	2.64	2.63	2.63
2018-3	2.62	2.65	2.64	2.65	2.65	2.64
2019-1	2.63	2.63	2.59	2.65	2.65	2.63
2019-2	2.64	2.63	2.65	2.65	2.63	2.64
2019-3	2.61	2.64	2.59	2.65	2.63	2.62
2020-1	2.63	2.65	2.69	2.60	2.66	2.65
2020-2	2.63	2.65	2.63	2.61	2.63	2.63
Mean	2.63	2.63	2.63	2.63	2.64	2.63

Uit de analyse zijn de volgende metingen weggelaten:

lab	pakket	reden
19	2016-3	* uitbijter 2,17
24	2017-2	* uitbijter 2,36

Tabel 2. Voorbeeld van gerapporteerde meetgegevens.

Het voordeel van deze aanpak met steeds één meting is dat daarmee de hoeveelheid werk voor de deelnemende laboratoria beperkt wordt. De keerzijde is dat de spreiding  $s_r$  voor de herhaalbaarheid – per laboratorium afzonderlijk – niet rechtstreeks uit duplo-bepalingen kan worden berekend.

Berekening van de  $s_r$  is echter wel mogelijk door gebruik te maken van de statistische methodiek variantieanalyse. In de praktijk komen regelmatig ontbrekende waarden voor (aangeduid met een  $\times$  in de tabel). Dit gebeurt als een laboratorium incidenteel een variabele niet heeft gemeten, of als de meetwaarde tijdens de analyses als uitbijter werd aangemerkt (zoals in bovenstaand voorbeeld). Om in die situaties toch nog correcte berekeningen van de spreiding te doen wordt gebruik gemaakt van de zogenaamde *restricted maximum likelihood methode* (REML).

Deze methode levert nog betrouwbare resultaten op als de datatabel niet te veel ontbrekende waarden heeft per lab (in een kolom) of per pakket (in een rij). Het minimaal aantal vereiste meetwaarden per lab of per pakket is in de analysescripts aanpasbaar en staat gewoonlijk op 2 ingesteld.

Praktisch betekent deze aanpak dat de verschillen ten opzichte van het pakketgemiddelde worden beschouwd (zie Tabel 3). Een lab draagt bij aan een grote standaardafwijking  $s_r$  voor de herhaalbaarheid als in betreffende kolom de verschillen sterk variëren. Een lab draagt bij aan een grote standaardafwijking  $s_R$  voor de reproduceerbaarheid als zijn overall gemiddelde afwijking (onderste regel) veel van 0 verschilt.

lab_nr	19	20	24	25	29	Mean
pakket						
2016-3	*	-0.01	0.03	-0.01	0.00	0.00
2017-1	0.02	-0.01	0.01	-0.01	0.00	0.00
2017-2	0.00	0.00	*	-0.01	0.00	0.00
2017-3	-0.03	-0.01	0.02	0.00	0.00	0.00
2018-1	0.01	-0.01	-0.02	0.00	0.01	0.00
2018-2	0.00	0.00	-0.02	0.01	0.00	0.00
2018-3	-0.02	0.01	0.00	0.01	0.01	0.00
2019-1	0.00	0.00	-0.04	0.02	0.02	0.00
2019-2	0.00	-0.01	0.01	0.01	-0.01	0.00
2019-3	-0.01	0.02	-0.03	0.03	0.01	0.00
2020-1	-0.02	0.00	0.04	-0.05	0.01	0.00
2020-2	0.00	0.02	0.00	-0.02	0.00	0.00
Mean	0.00	0.00	0.00	0.00	0.00	0.00

Tabel 3. Verschillen ten opzichte van het pakketgemiddelde (uit data van Tabel 2).

## 4 De rapportage in het ILVO – ringonderzoek

Telkens na een nieuwe ronde (pakket) worden de resultaten in het ringonderzoek gerapporteerd op basis van berekeningen over de laatste 12 pakketten. In dit hoofdstuk wordt toegelicht hoe deze rapportage gelezen dient te worden. In Bijlage 1 worden de achtergronden bij de statistische berekeningen toegelicht.

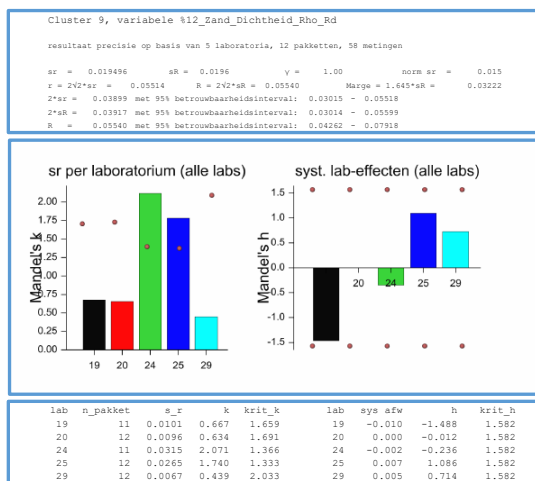
### 4.1 Rapportageblad

De ILVO-rapportage geeft per variabele (verrichting) een overzicht van de meetgegevens en een tabel met verschillen ten opzichte van het pakketgemiddelde, zoals hierboven al genoemd (paragraaf 3.3). Bovendien wordt nog aangegeven welke metingen eventueel niet in de analyse zijn betrokken en om welke reden.

Vervolgens wordt een rapportage opgesteld met de resultaten van de statistische analyses. Deze rapportage dient in feite twee doelen. In de eerste plaats wordt aangegeven hoe nauwkeurig de meetmethode voor elke variabele (verrichting) is in termen van herhaalbaarheid en reproduceerbaarheid. Daarnaast wordt ook aangegeven wat de bijdrage van ieder afzonderlijk laboratorium is aan die meetnauwkeurigheid. Voor elke variabele wordt een pagina vervaardigd met een indeling zoals verkleind is weergegeven in Figuur 2. De onderdelen uit deze rapportage worden hierna uitvergroot en toegelicht.

Hierbij worden drie onderdelen onderscheiden:

- Berekening van de overall meetspreiding (zie 4.2) in het bovenste deel van de rapportage
- Grafische weergave van de meetonzekerheid per individueel lab (zie 4.3)
- Overzicht van binnen-lab variatie en systematische afwijkingen per laboratorium (zie 4.4)



Figuur 2. Voorbeeld van een rapportage

### 4.2 Rapportage overall meetspreiding

De prestaties van de meetmethode kunnen worden samengevat in twee maten voor meetspreiding:

- $s_r$  = de **standaardafwijking voor de herhaalbaarheid** en
- $s_R$  = de **standaardafwijking voor de reproduceerbaarheid**

Hierbij verwijst het onderscheid tussen herhaalbaarheid en reproduceerbaarheid naar de variatie die optreedt als metingen worden uitgevoerd door hetzelfde laboratorium (binnen korte tijd en met dezelfde laborant op hetzelfde meetinstrument), respectievelijk door willekeurige laboratoria. Uit deze standaardafwijkingen kunnen nog diverse andere kentallen worden berekend, zoals weergegeven in Figuur 3. Ze worden hieronder toegelicht.

```
Cluster 9, variabele %12_Zand_Dichtheid_Rho_Rd
resultaat precisie op basis van 5 laboratoria, 12 pakketten, 58 metingen

sr = 0.019496      sR = 0.0196      γ = 1.00      norm sr = 0.015
r = 2√2*sr = 0.05514      R = 2√2*sR = 0.05540      Marge = 1.645*sR = 0.03222
2*sr = 0.03899 met 95% betrouwbaarheidsinterval: 0.03015 - 0.05518
2*sR = 0.03917 met 95% betrouwbaarheidsinterval: 0.03014 - 0.05599
R = 0.05540 met 95% betrouwbaarheidsinterval: 0.04262 - 0.07918
```

*Figuur 3. Rapportage van kentallen voor de meetspreiding.*

Voor de **herhaalbaarheid** – dat wil zeggen voor de meetonzekerheid bij metingen binnen een laboratorium, in een klein tijdsbestek, bij één laborant en één apparaat – worden de volgende kentallen vermeld:

- De standaardafwijking  $s_r$ . Deze waarde wordt ook wel de standaardonzekerheid ( $u$ ) genoemd.
- De waarde  $2*s_r$ . Dit wordt de uitgebreide onzekerheid ( $U$ ) onder herhaalbaarheidscondities genoemd. Met een betrouwbaarheid van 95% kan worden gesteld dat de werkelijke waarde zal liggen in het interval “meetresultaat  $\pm U$ ”. Met het 95% betrouwbaarheidsinterval wordt aangegeven binnen welke grenzen deze  $2*s_r$  met 95% betrouwbaarheid zal liggen. De breedte van het betrouwbaarheidsinterval hangt af van enerzijds de  $s_r$  en anderzijds van het aantal meetwaarden (aantal laboratoria en aantal metingen per laboratorium) waarop de berekening plaatsvond.
- De waarde voor de herhaalbaarheid  $r = 2\sqrt{2} * s_r$ , ook wel herhaalbaarheidsgrens genoemd. Met een betrouwbaarheid van 95% kan worden gesteld dat het absolute verschil van twee herhaalde meetresultaten kleiner zal zijn dan  $r$ .
- De normwaarde voor  $s_r$  indien die voor betreffende verrichting bekend is.

Zoals al opgemerkt in hoofdstuk 2 heeft de ILVO werkgroep besloten dat aan de berekende herhaalbaarheid geen waarde moet worden gehecht. De berekende herhaalbaarheid bevat namelijk ook variatie ten gevolge van spreiding tussen monsters (binnen een ronde) en eventuele variatie tussen apparaten en/of laboranten binnen een laboratorium.

Voor de **reproduceerbaarheid** – dat wil zeggen voor de meetonzekerheid bij metingen door verschillende laboratoria – worden de volgende kentallen vermeld:

- De standaardafwijking  $s_R$ . Deze waarde wordt ook wel de standaardonzekerheid ( $u$ ) genoemd.
- De waarde  $2*s_R$ . Deze waarde is al vermeld in hoofdstuk 1 en wordt de uitgebreide onzekerheid ( $U$ ) onder reproduceerbaarheidscondities genoemd. Met een betrouwbaarheid van 95% kan worden gesteld dat de werkelijke waarde zal liggen in het interval “meetresultaat  $\pm U$ ”. Met het 95% betrouwbaarheidsinterval wordt aangegeven binnen welke grenzen deze  $2*s_R$  met 95% betrouwbaarheid zal liggen. De breedte van het betrouwbaarheidsinterval hangt af van enerzijds de  $s_R$  en anderzijds van het aantal meetwaarden (aantal laboratoria en aantal metingen per laboratorium) waarop de berekening plaatsvond.
- De waarde voor de reproduceerbaarheid  $R = 2\sqrt{2} * s_R$ , ook wel reproduceerbaarheidsgrens genoemd. Met een betrouwbaarheid van 95% kan worden gesteld dat het absolute verschil van twee herhaalde

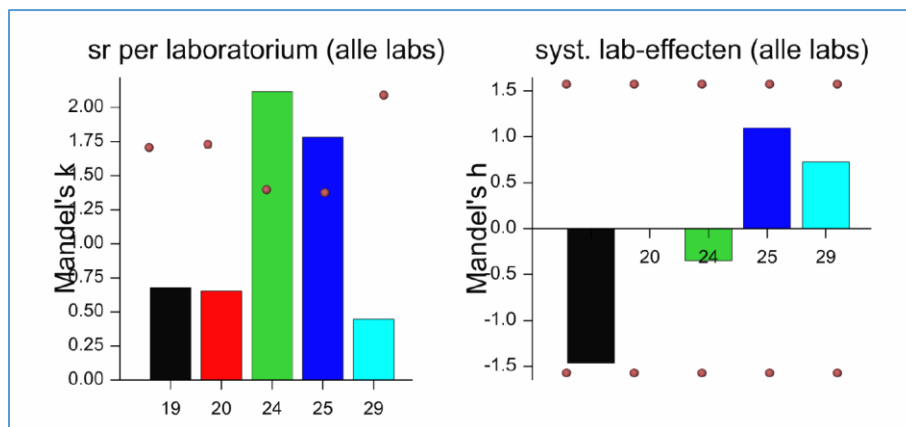
meetresultaten kleiner zal zijn dan R. Merk op dat in de output het  $\sqrt{\quad}$  teken onbedoeld als een S is overgekomen. Ook voor de R is het 95% betrouwbaarheidsinterval vermeld.

- Marge  $M = 1,645 \cdot s_R$ . De marge wordt gebruikt om te toetsen aan een grenswaarde voor het gemeten kenmerk. Met 95% betrouwbaarheid kan worden gesteld dat de werkelijke waarde kleiner zal zijn dan de grenswaarde, indien de gemeten waarde meer dan M onder de grenswaarde ligt. Of anderszijds dat de werkelijke waarde groter zal zijn dan de grenswaarde, indien de gemeten waarde meer dan M boven de grenswaarde ligt. Merk op dat de marge een andere betekenis heeft dan de meetonzekerheid. Met de meetonzekerheid wordt een *tweezijdig* interval gegeven rond de gemeten waarde waarbinnen de werkelijke waarde met 95% betrouwbaarheid zal liggen. Met de marge wordt een *eenzijdig* interval gegeven om te toetsen of de gemeten waarde aantoont dat de werkelijke waarde met 95% betrouwbaarheid onder (danwel boven) de grenswaarde ligt.

Verder wordt nog een waarde gamma ( $\gamma$ ) berekend, in de tekst weergegeven als een g. Er geldt  $\gamma = s_R/s_r$ . Deze grootte geeft de verhouding weer tussen de standaardafwijking van de reproduceerbaarheid en de standaardafwijking van de herhaalbaarheid weer. Hoe dichter deze waarde in de buurt van 1 ligt, hoe kleiner de systematische verschillen tussen de laboratoria zijn. Indien de meetspreiding te groot wordt bevonden kan aan de  $\gamma$  dus worden afgelezen of verbeteringen moeten worden gezocht in het verkleinen van verschillen tussen de laboratoria (indien  $\gamma$  veel groter is dan 1) of in het verkleinen van de inherente meetspreiding binnen de laboratoria (indien  $\gamma$  dicht bij 1 ligt).

### 4.3 Rapportage meetonzekerheid per individueel laboratorium

De meetonzekerheid per individueel laboratorium wordt grafisch weergegeven met twee grafieken zoals getoond in Figuur 4.



Figuur 4. Meetonzekerheden per laboratorium.

Het linker deel van de grafiek toont de verschillen in meetspreiding binnen de laboratoria. Dit wordt gedaan via de zogenaamde Mandel's k. Mandel's k is de verhouding tussen:

- De standaardafwijking voor de meetspreiding binnen betreffend laboratorium, en
- De overall standaardafwijking  $s_r$  voor de herhaalbaarheid, of – indien deze gegeven is – de normwaarde voor  $s_r$ .

Meer details over de berekeningsmethode worden gegeven in bijlage 1.

De staaf in de grafiek geeft de berekende Mandel's k per laboratorium weer. Indien die waarde groter is dan 1, geeft dit de indruk dat de spreiding binnen betreffend lab aan de hoge kant is. Er is echter ook een

statistische onzekerheid in de berekende waarde voor Mandel's k, gezien de beperkte hoeveelheid meetgegevens waarop de statistische analyses plaatsvinden. Deze onzekerheid is verdisconteerd in de grenswaarde, aangegeven met een rode punt. Het betreffende laboratorium heeft dus een significant hogere meetspreiding dan de overall waarde (of de normwaarde) indien de staaf hoger is dan de grenswaarde.

Opmerkingen:

- In incidentele gevallen ontbreekt de staaf en de grenswaarde voor Mandel's k voor een laboratorium. Dit betekent dat de spreiding binnen betreffend laboratorium veel kleiner is dan voor de overige laboratoria en daarom niet te berekenen was.
- De grenswaarden zijn verschillend per laboratorium, omdat ze afhankelijk zijn van het aantal pakketten dat een laboratorium heeft gemeten en van de grootte van de meetspreiding binnen dat laboratorium.
- Als slechts 2 laboratoria deelnemen bij een verrichting, kan geen afzonderlijke meetspreiding per laboratorium worden berekend.

Het rechter deel van de grafiek toont de verschillen in meetniveaus tussen de laboratoria. Dit wordt gedaan via de zogenaamde Mandel's h. Mandel's h is de verhouding tussen:

- Het verschil van het gemiddelde meetniveau van een specifiek laboratorium ten opzichte van het overall gemiddelde, en
- De standaardafwijking van de gemiddelde meetniveaus van de laboratoria.

Meer details over de berekeningsmethode worden gegeven in bijlage 1.

De staaf in de grafiek geeft de berekende Mandel's h per laboratorium weer. Als er weinig verschillen in meetniveau tussen laboratoria zijn, liggen de waarden rond 0. Om te toetsen of een lab-niveau statistisch significant groter (of kleiner) is dan het overall gemiddelde zijn met de rode punten de grenswaarden aangegeven. Die grenswaarden zijn voor alle laboratoria gelijk. De grenswaarden komen dichterbij 0 te liggen naarmate het aantal laboratoria dat aan betreffende verrichting meedoet groter wordt.

#### 4.4 Binnen-lab variatie en systematische afwijkingen per laboratorium

Het onderste gedeelte van de rapportage bevat een tabel met per laboratorium de waarden die zijn weergegeven in hierboven genoemde grafieken van Mandel's k en Mandel's h. Een voorbeeld hiervan is weergegeven in Figuur 5. Meer details over de berekeningsmethode worden gegeven in bijlage 1.

lab	n_pakket	s_r	k	krit_k	lab	sys afw	h	krit_h
19	11	0.0101	0.667	1.659	19	-0.010	-1.488	1.582
20	12	0.0096	0.634	1.691	20	0.000	-0.012	1.582
24	11	0.0315	2.071	1.366	24	-0.002	-0.236	1.582
25	12	0.0265	1.740	1.333	25	0.007	1.086	1.582
29	12	0.0067	0.439	2.033	29	0.005	0.714	1.582

Figuur 5. Mandel's k en Mandel's h per laboratorium.



## 5 Onderliggende normdocumenten

De hiervoor beschreven statistische analysemethoden sluiten aan bij de norm ISO 5725-2. In deze norm wordt uitgelegd hoe de herhaalbaarheid en reproduceerbaarheid dienen te worden bepaald. Als ideale situatie wordt aangegeven dat er een gebalanceerde opzet is, waarin  $p$  laboratoria op  $q$  meetniveaus  $n$  herhaalde metingen uitvoeren. In de ILVO-situatie zijn er normaal gesproken  $q = 12$  pakketten, die door een aantal laboratoria worden gemeten. Het aantal laboratoria verschilt per verrichting (variabele). Metingen gebeuren in enkelvoud, dus  $n = 1$ .

Deze norm is wat betreft de statistische analyses veel uitgebreider dan de norm ISO/IEC 17043, waarnaar wordt verwezen in het document RvA-T030-NL. Het RvA document verwijst voor de toetsing op afwijkingen naar de Z-waarden en En-waarden. In de ISO/IEC 17043 wordt beschreven dat deze waarden staan voor de afwijkingen tussen het resultaat van een laboratorium en een referentiewaarde ("assigned value"), gedeeld door de standaardafwijking (bij de Z-waarde) of de meetonzekerheid (bij de En-waarde). De norm beschrijft niet hoe de referentiewaarde en de spreiding daadwerkelijk worden bepaald.

De norm ISO 5725-2 is veel specifiek en gaat uitgebreid in op de volgende zaken:

- Een onderliggend statistisch model van de vorm  $Y = M + B + \varepsilon$ ; d.w.z. een meetwaarde kan worden beschreven als de som van een overall gemiddelde, een afwijking voor een specifiek lab, en een random ruiscomponent ten gevolge van de toevallige meetfout.
- Berekeningen van de standaardafwijkingen  $s_r$  (herhaalbaarheid),  $s_L$  (spreiding tussen laboratoria) en  $s_R$  (reproduceerbaarheid)
- Toetsing op uitbijters
- Toetsing van de "between laboratory consistency" met Mandel's  $h$ ; deze Mandel's  $h$  voor toetsing op spreiding tussen labs komt in de plaats van de Z-waarde of En-waarde in ISO/IEC 17043.
- Toetsing van de "within laboratory consistency" met Mandel's  $k$ ; dit is een extra karakterisering voor een toetsing van de binnen-lab spreidingen, die niet is afgedekt in ISO/IEC 17043.

De ILVO analysescripts sluiten aan op ISO 5725-2, waarbij de methodiek zodanig is aangepast dat deze kan omgaan met de specifieke ILVO-situatie. De belangrijkste zaken hierin zijn:

- Omdat er geen herhalingen zijn, kan de  $s_r$  niet worden bepaald uit de spreiding binnen de groepjes van herhaalde metingen (per lab, per pakket). De geëigende statistische aanpak is dan dat de  $s_r$  wordt berekend uit de spreiding in de residuen, d.w.z. de variatie die overblijft in de meetwaarden na correctie voor lab-verschillen en pakket-verschillen. Extreme waarden in die residuen zijn verdacht en worden – na een check – als uitbijter aangemerkt.
- De ISO norm gaat uit van een gebalanceerde opzet, waarin voor elke combinatie van lab en pakket metingen beschikbaar zijn. In de ILVO-situatie komen echter geregeld lege plekken voor, omdat voor een bepaalde combinatie van lab en pakket de meetwaarde niet is bepaald, of als uitbijter is weggelaten. In een dergelijke niet-gebalanceerde situatie is een speciale analysetechniek (REML) gehanteerd om dan toch nog zuivere effecten van labs en pakketten te kunnen berekenen. Die zuivere effecten zijn weer belangrijk bij de berekening van de echte herhaalbaarheidsspreiding, Mandel's  $k$  en Mandel's  $h$ .

## Bijlage 1. Toelichting op de statistische analyses

### 1 Mandel's k en h bij meer dan 2 deelnemende labs

Mandel's k en Mandel's h zijn in feite gestandaardiseerde waarden voor de interne meetspreiding en de systematische afwijking van een laboratorium. Deze waarden vinden hun oorsprong in de norm ISO 57252 (1994). In dit hoofdstuk wordt aangegeven hoe deze waarden worden berekend in de situatie dat er minstens 3 laboratoria deelnemen aan betreffende verrichting.

#### 1.1 Mandel's k voor de meetspreiding

##### 1.1.1 Berekening meetspreiding voor individuele labs

Per laboratorium wordt de standaardafwijking voor de herhaalbaarheid berekend als  $s_{r,i}$  waarbij  $i$  staat voor een willekeurig laboratorium.

De standaardafwijking  $s_{r,i}$  per lab wordt berekend uit de variaties in de verschillen ten opzichte van de pakketgemiddelden voor dit lab (zie Tabel 3 in voorgaande paragraaf 3.3). Omdat in het pakketgemiddelde impliciet ook de spreiding  $s_{r,j}$  van de overige labs zit, ontstaat een stelsel van  $L$  vergelijkingen met  $L$  waarden voor de  $s_{r,i}$  voor  $i = 1, 2 \dots L$ . Waarbij  $L$  het aantal labs is. Door dit stelsel op te lossen worden de standaardafwijkingen  $s_{r,i}$  per laboratorium  $i$  berekend.

De standaardafwijkingen  $s_{r,i}$  worden berekend op basis van een beperkt aantal metingen en er is dus sprake van een onzekerheidsmarge in die standaardafwijkingen. Deze marge wordt bepaald door het zogenaamde aantal vrijheidsgraden. Bij een groter het aantal vrijheidsgraden neemt de onzekerheid in de spreiding af, wordt de marge dus kleiner en komt de grenswaarde dichter bij 1 te liggen.

Bij een simpele steekproef zou gelden dat het aantal vrijheidsgraden gelijk is aan het aantal waarnemingen minus 1. In het ILVO ringonderzoek is de situatie echter meer gecompliceerd. Praktisch geldt dat het aantal vrijheidsgraden groter wordt als de  $s_{r,i}$  voor betreffend lab relatief groot is t.o.v. de andere labs, of als het aantal metingen voor betreffend lab groter is.

##### 1.1.2 Toetsing van de meetspreiding per lab aan de overall $s_r$

Indien voor betreffende verrichting geen normwaarde voor de herhaalbaarheid beschikbaar is, wordt de meetspreiding  $s_{r,i}$  per individueel lab getoetst aan de overall waarde  $s_r$ . Deze  $s_r$  is eveneens berekend op basis van een beperkte set van meetwaarden. Dus ook de  $s_r$  heeft een betrouwbaarheidsmarge, die wordt bepaald door de vrijheidsgraden.

De Mandel's k voor een individueel lab wordt nu berekend als  $k_i = s_{r,i}/s_r$ . Simpelweg zou men kunnen zeggen dat de herhaalbaarheidsspreiding van een lab te groot is, indien  $k_i > 1$ . Maar vanwege de onzekerheidsmarges op zowel  $s_{r,i}$  als  $s_r$  moet voor  $k_i$  een 95% grenswaarde  $krit\_k_i$  worden aangehouden. Die grenswaarde wordt berekend op basis van een F-verdeling, en ligt rond 1.5, afhankelijk van het aantal vrijheidsgraden.

De uitspraak wordt dus: een individueel laboratorium ( $i$ ) heeft een statistisch significant grotere meetspreiding dan de overige, indien voor dit lab geldt dat  $k_i > krit\_k_i$ .

### 1.1.3 Toetsing van de meetspreiding per lab aan de normwaarde voor $s_r$

Indien voor betreffende verrichting wel een normwaarde voor de herhaalbaarheid beschikbaar is, wordt de meetspreiding  $s_{r,i}$  per individueel lab getoetst aan de normwaarde  $s_{norm}$ .

De Mandel's  $k$  voor een individueel lab wordt nu berekend als  $k_i = s_{r,i}/s_{norm}$ . Simpelweg zou men kunnen zeggen dat de herhaalbaarheidsspreiding van een lab te groot is, indien  $k_i > 1$ . Maar vanwege de onzekerheidsmarge op  $s_{r,i}$  moet voor  $k_i$  een 95% grenswaarde  $krit_{k_i}$  worden aangehouden. Die grenswaarde wordt berekend op basis van een chikwadraat-verdeling. Grenswaarden liggen meestal tussen 1.3 en 2.0.

De uitspraak wordt dus: een individueel laboratorium heeft een statistisch significant grotere meetspreiding dan de norm, indien voor dit lab geldt dat  $k_i > krit_{k_i}$ .

## 1.2 Mandel's $h$ voor systematische afwijkingen per lab

Als eerste wordt per laboratorium de gemiddelde waarde  $M_i$  berekend. Als alle laboratoria voor alle pakketten een meetwaarde hebben gerapporteerd volgt  $M_i$  eenvoudig door de meetresultaten van het betreffende lab te middelen. In andere gevallen wordt bij de berekening van de  $M_i$  gecorrigeerd voor ontbrekende waarden via de methodiek van REML (restricted maximum likelihood).

Vervolgens wordt over de labgemiddelden  $M_i$  voor  $i = 1, 2 \dots L$  het overall gemiddelde  $MM$  en de standaardafwijking  $s_m$  berekend. Tenslotte wordt voor ieder individueel lab de waarde voor Mandel's  $h$  berekend als  $h_i = (M_i - MM)/s_m$ .

Bij deze  $h_i$  wordt ook een kritieke grenswaarde  $krit_h$  berekend die voor ieder laboratorium gelijk is, en qua grootte afhankelijk is van het aantal deelnemende labs.

De uitspraak wordt: een laboratorium heeft een systematische afwijking in het meetniveau, indien voor dit lab voor de absolute waarde van  $h_i$  geldt  $|h_i| > krit_h$ .

## 2 Berekeningen bij 2 deelnemende labs

In een beperkt aantal gevallen (verrichtingen) komt het voor dat slechts 2 laboratoria deelnemen. Dit heeft de volgende consequenties voor het berekenen van de Mandel's  $k$  en  $h$ :

- De berekening van de afzonderlijke meetspreiding  $s_{r,i}$  per lab is niet meer mogelijk.
- De toetsing van de systematische verschillen via Mandel's  $h$  moet worden aangepast.

### 2.1 Toetsing van de meetspreiding per lab bij deelname van 2 labs

Indien slechts 2 laboratoria deelnemen, kan voor deze 2 laboratoria niet de afzonderlijke meetspreiding  $s_{r,i}$  worden berekend. En dus ook niet de waarde  $k_i$  van Mandel's  $k$ .

De reden hiervoor is dat de meetspreiding per lab moet worden berekend uit de variatie in de verschillen van de resultaten voor dat lab ten opzichte van het pakketgemiddelde. In de situatie van 2 deelnemende labs zijn die verschillen per definitie altijd even groot voor beide labs, op het +/- teken na. Daardoor is er geen onderscheid te maken in de spreiding per lab.

Dit probleem is binnen de huidige opzet van het ringonderzoek niet oplosbaar. In de rapportage wordt in deze situaties dus vermeld dat geen toetsing op verschillen in meetspreiding tussen labs mogelijk is.

## 2.2 Toetsing van de systematische verschillen bij deelname van 2 labs

Zoals hierboven is aangegeven wordt Mandel's  $h$  voor de systematische verschillen berekend als  $h_i = (M_i - MM)/s_M$ . In de situatie van 2 labs geldt dat deze  $h_i$  voor beide labs per definitie dezelfde waarde heeft op het +/- teken na. Het is zelfs te berekenen dat die waarden altijd gelijk zijn aan  $\pm 1/\sqrt{2} = \pm 0,707$ .

Om ook in deze situatie een zinvolle toetsing op systematische verschillen te kunnen doen is de volgende alternatieve werkwijze gevolgd. De waarde van Mandel's  $h$  per laboratorium is berekend als het labgemiddelde, gedeeld door de standaardafwijking voor de reproduceerbaarheid. Dus  $h_i = M_i / s_R$ . De kritische waarde  $krit\_h$  is bepaald vanuit het uitgangspunt dat de systematische verschillen tussen de 2 laboratoria niet zo groot mogen zijn dat ze de leiden tot een grote waarde van  $\gamma = s_R / s_r$ . Dus de eis is dat  $\gamma < \gamma_{crit}$ .

Te berekenen is dat deze eis overeenkomt met  $krit\_h = \sqrt{\frac{1}{2} \left(1 - \frac{1}{\gamma_{crit}^2}\right)}$

In de praktijk blijkt dat voor 95% van de verrichtingen de  $\gamma$  lager ligt dan 1,5. Deze waarde is gekozen als  $\gamma_{crit}$ . En daarmee komt  $krit\_h$  op 0,527.

## Bijlage 2. Overzicht van variabelen per cluster (horend bij paragraaf 3.1)

Cluster 1	Cluster 5	Cluster 8	Cluster 9
1 Bitumengehalte	1a Eenpuntproctorproef vochtgehalte	1a Gehalte < 2 mum	1a Zand, korrelg door zeef 4mm
2 Dichtheid mengsel	1b Eenpuntproctor - dichtheid, ong.	1b Gehalte < 20 mum (areometer)	1b Zand, korrelg door zeef 2mm
3 Dichtheid proefstuk	2a Cbr-waarde	2 Gehalte < 20 mum	1c Zand, korrelg door zeef 1mm
4a Door zeef 16 mm	2b Cbr-dichtheid	3 Gehalte door 63 mum	1d Zand, korrelg door zeef 0,5mm
4b Door zeef 11,2 mm	2c Cbr-vochtgehalte	4 Geleidbaarheid	1e Zand, korrelg door zeef 0,25mm
4c Door zeef 8 mm	4l Vlakheidsindex FI	5a Gloeiverlies	1f Zand, korrel door zeef 0,125mm
4d Door zeef 5,6 mm	5a Gebroken grind en steenslagbeton	5b Organisch stofgehalte	6 Zand, Door zeef 0,063mm (zeer fijn)
4e Door zeef 4 mm	5b Overige gebroken steen	5c CaCO3-gehalte	2n Zand, Gehalte aan humus (NEN1744)
4f Door zeef 2 mm	5c Gebroken metselwerkpuin	6a Proctorproef, -vochtgehalte	3n Zand, Gehalte aan humus (kleurenplaatje)
4g Door zeef 0,5 mm	5d Andersoortig gebroken steen	6b Proctorproef, -dichtheid	5n Zand, Gehalte aan fulvo (NEN1744)
4h Door zeef 0,18 mm	5e Asfalt (gebroken&gefreed en asfaltbeton)		4n Zand, Gehalte aan fulvozuur
4i Door zeef 0,063 mm	5f Gips en niet-steenachtig materiaal		8 Zand, stroomcoefficient
8a Laagdikte 1 (boven)	5g Verteerbaar organisch materiaal		11 Zand, Schijnb dichth (Rho a)
8b Laagdikte 2	5h Klei en grove organ bestanddelen		12 Zand, Dichtheid (Rho Rd)
8c Laagdikte 3	6 Los Angeles coefficient		13 Zand, Dichtheid (Rho ssd)
8d Laagdikte 4	7 Lichte deeltjes in toeslagmateriaal		14 Zand, Waterabsorptie
8e Laagdikte 5	8a FL		7a Gehalte materiaal groep Cc
	8b Rc		7b Gehalte materiaal groep Cr
	8c Ru		7c Gehalte materiaal groep Ctc
	8d Rb		7d Gehalte materiaal groep Ctr
	8e Ra		15 Grind, Schijnb dichth (Rho a)
	8f Rg		16 Grind, Dichtheid (Rho Rd)
	8g X		17 Grind, Dichtheid (Rho ssd)
			18 Grind, Waterabsorbtie
			19 zandequivalent
			20 zand methyleenblauwwaarde
			21 Rho a
			22 Rho rd
			23 Rho ssd
			24 Waterobsoptie WA24

Cluster 10	Cluster 11	Cluster 12	Cluster 13
1 Betonprisma's, vol. massa	1a Klei, plasticiteitsindex	1 Prisma mortel volumieke massa	2 Aanmaakwater, indamprest
2 Betonprisma's, buigsterkte	1b Klei, vloeigrens	2 Prisma mortel, buigsterkte	3 Aanmaakwater, chloridegehalte
3 Betonprisma's, druksterkte	2a Textuurbep. massaperc lutum	3 Prisma mortel, druksterkte	4 Aanmaakwater, sulfaten
4 Betonprisma's, splijttreksterkte	2b Textuurbep. Massaperc leem		5 Aanmaakwater, pH
5 Betonkubus, waterindringing	2c M50 cijfer		6 Aanmaakwater, carbonaten
	3 Geh organische stof van grond		7 Aanmaakwater, bicarbonaten
	4 Massaverlies zoutzuurbep van grond		
	7 Geh organische stof in grond		

Cluster 14	Cluster 15	Cluster 16
1a Betonstraatstenen lengte	1a Asfaltcil laagopb t/m 1	1 Zonder additie
1b Betonstraatstenen breedte	1b Asfaltcil laagopb t/m 2	2 Met additie
1c Betonstraatstenen dikte	1c Asfaltcil laagopb t/m 3	
2 Betonstraatstenen ongepolijste slipweerstand (USRV)	1d Asfaltcil laagopb t/m 4	
3 Betonstraatstenen totale wateropsorping	1e Asfaltcil laagopb t/m 5	
5 Betonstraatstenen slijtbestandheid	1f Asfaltcil laagopb t/m 6	
6 Betontegels vorst/dooibestandheid met dooizouten	1g Asfaltcil laagopb t/m 7	
	1h Asfaltcil laagopb t/m 8	
	1i Asfaltcil laagopb t/m 9	
	1j Asfaltcil laagopb t/m 10	
	1k aantal lagen	
	1l Fluorescentie van bovenaf	
	1m Fluorescentie van bovenaf	
	2b PAK granulaat (GCMS)	
	2c PAK granulaat (HPLC)	
	2a PAK granulaat Ja/Nee (DLC)	
	1a soort laag 1	
	1b soort laag2	
	1c soort laag3	
	1d soort laag4	
	1e soort laag5	
	1 f soort laag6	
	1 ff soort laag7	
	1 fff soort laag8	
	1 fff soort laag9	
	1 fff soort laag10	
	1 fff soort laag11	
	soort PAK-gehalte bepaling	
	extractie vloeistof	
	2a PAK granulaat 1 (DLC)	
	3 PAK granulaat 2	