

Relativiteit van analyseresultaten in het wateronderzoek en -beheer

Voordracht in het kader van de cursus 'Milieumetingen', d.d. 1 t/m 3 maart 1988 van de Stichting Postakademiale Vorming Gezondheidstechniek in de TU te Delft.

Inleiding

Ik begin met twee zinspreuken, waarvan de eerste luidt: 'Meten is weten, maar weet wel wat je meet'. Mijn boodschap spitst zich toe op het 'Weet wel wat je meet'. De rode draad in dit artikel is, dat cijfers uit analyses niet zo hard zijn als u misschien van een zogenaamd exacte wetenschap als de analytische chemie zou verwachten. Voor een juiste interpretatie is enige achtergrondinformatie nodig, onder meer statistische informatie over de mogelijke spreiding in analyseresultaten.



DR. R. VAN VEEN
Normcommissie
Afvalwater

Een tweede zinspreuk is een citaat van de Romeinse dichter Vergilius, dat ik voor mijn doel enigszins gewijzigd heb: '*Varium et mutabile, semper natura*', ofwel: De natuur is altijd iets wisselvalligs en veranderlijks. Wat wij aan de natuur denken te kunnen waarnemen is behept met onzekerheden. Deze onzekerheid is, althans ten dele, fundamenteel. Voor waarnemingen op atomair niveau geldt bijvoorbeeld de onzekerheidsrelatie van Heisenberg. Voor waarnemingen op macroscopisch niveau, bijvoorbeeld metingen in water, zijn de onzekerheden in het meetresultaat, zeker ten dele, even fundamenteel, zij het nog minder eenvoudig te voorspellen.

Laboratoriumevaluatie DBW/RIZA

Een goede methode om inzicht te krijgen in de zekerheid van waterkwaliteitsgegevens is het doen van interlaboratorium vergelijkende onderzoeken.

Sinds 1973 organiseert de Dienst Binnenwateren/RIZA deze zogenaamde ringonderzoeken, die tot doel hebben het verwerven van inzicht in en het bevorderen van de vergelijkbaarheid van analyseresultaten.

Vergelijkbaarheid is duidelijk wat anders dan juistheid. De vraag wat juist is, hangt namelijk in sterke mate af van de doelstelling van een onderzoek, bijv. moet een totaal gehalte worden bepaald of een biologisch beschikbare (= effectrelevante) fractie daarvan. De juistheid van een analysemethode kan worden nagegaan met behulp van standaarden, standaardaddities, en dergelijke in zogenaamde methodenevaluerende ringonderzoeken. In laboratoriumevaluerende ringonderzoeken daarentegen is het uitgangspunt de vergelijkbaarheid van analyseresultaten verkregen door meting aan monsters. Bij methodenevaluaties is de toe te passen analysemethode dwingend voor-

Samenvatting

Uit laboratoriumevaluerende vergelijkende onderzoeken onder Nederlandse waterlaboratoria blijkt: als één laboratorium een parameter in een monster bepaalt is de spreiding in de analyseresultaten gemiddeld 5 à 6 maal kleiner dan wanneer meerdere laboratoria die parameter in dat monster bepalen. Daarom dient bij onderzoeken van variaties in tijd en/of ruimte, waarbij vooral de relatieve analyseresultaten van belang zijn, elke te onderzoeken parameter bij één analyselaboratorium te worden ondergebracht; bij toetsing aan normen of aan voorwaarden (waarmee soms grote belangen gemoeid kunnen zijn), waarbij de absolute analyseresultaten in het geding zijn (toetsing aan een hard getal), verdient het aanbeveling meerdere onafhankelijke analyseresultaten (verschillende analyselaboratoria en/of -methoden) in te winnen. Bij vergelijking van analyseresultaten met gegevens uit andere onderzoeken of reeds aanwezige gegevensbestanden dient men de mogelijk optredende spreidingen expliciet in de beschouwing te betrekken.

geschreven; bij laboratoriumevaluatie zijn de deelnemende laboratoria vrij in de keuze van de toe te passen analysemethode en fungeren dus in zekere zin als een 'black box' waar monsters binnenkomen en analyse-resultaten uitgaan, een veel voorkomende praktijksituatie. In deze voordracht gaat het dus om de resultaten van laboratorium-evaluerende ringonderzoeken verkregen met rondzendmonsters.

De doelgroepen van laboratorium-evaluerende ringonderzoeken zijn:

- laboratoria van waterbeheerders (provincies, waterschappen, zuiveringschappen en dergelijke);
- commerciële laboratoria, ingenieursbureaus en dergelijke, voorzover opdracht-nemers die voor waterbeheerders analyses verrichten;
- voor oppervlaktewater tevens laboratoria van waterleidingbedrijven die oppervlaktewater als grondstof gebruiken.

In het kader van deze laboratorium-evaluerende ringonderzoeken ontvangen de deelnemende laboratoria triplo monsters (3 identieke monsters, ter vaststelling van de herhaalbaarheid, zie hierna), enkelvoudige monsters en standaardoplossingen.

De terug ontvangen analyse-resultaten worden statistisch verwerkt volgens een internationale ontwerpnorm ISO/DIS 5725. Na verwijdering uit de resultatenmatrix van variantie-uitschieters van de triplo-waarnemingen (herhaalbaarheid) door toepassing van de Cochran-toets met een onbetrouwbaarheid van 1% en vervolgens door toepassing van de Dixon-toets eveneens met een onbetrouwbaarheid van 1% van uitschieters in de (in het geval van de triplo-monsters: gemiddelde) uitkomsten (reproduceerbaarheid) worden van de overgebleven analyse-uitkomsten de standaardafwijkingen van de herhaalbaarheid (s_r) en van de reproduceerbaarheid (s_R) berekend. Herhaalbaarheid en reproduceerbaarheid zijn de twee uitersten van het 'onzekerheids-

spectrum' van een analyseresultaat en zijn als volgt gedefinieerd:

'Herhaalbaarheid (= duplicerbaarheid) is een kwantitatieve maat voor de spreiding tussen meetuitkomsten, verkregen op als identiek te beschouwen materiaal onder zoveel mogelijk identieke omstandigheden' (NEN 3114). In de praktijk zal met identieke omstandigheden meestal worden bedoeld, dat de metingen door dezelfde waarnemer worden verricht in één laboratorium met behulp van dezelfde apparatuur en eventuele hulpstoffen op zo dicht mogelijk bij elkaar gelegen tijdstippen. Herhaalbaarheid is dus een maat voor de binnenlaboratorium-spreiding van een analyseresultaat.

'Reproduceerbaarheid is een kwantitatieve maat voor de spreiding tussen meetuitkomsten verkregen op als identiek te beschouwen materiaal onder uiteenlopende omstandigheden, zoals: verschillende waarnemers, verschillende instrumenten, verschillende laboratoria, verschillende tijdstippen' (NEN 3114).

Aangezien het hier gaat om laboratorium-evaluerende ringonderzoeken wordt tevens een waardering toegekend aan de prestaties van de deelnemende laboratoria. De na verwijdering van statistische uitschieters overgebleven laboratoria worden daartoe ingedeeld in een A-, B- of C-klasse:

- A-klasse: resultaten binnen 1 s_R vanaf het gemiddelde
- B-klasse: resultaten tussen 1 s_R en 2 s_R vanaf het gemiddelde
- C-klasse: resultaten verder dan 2 s_R vanaf het gemiddelde.

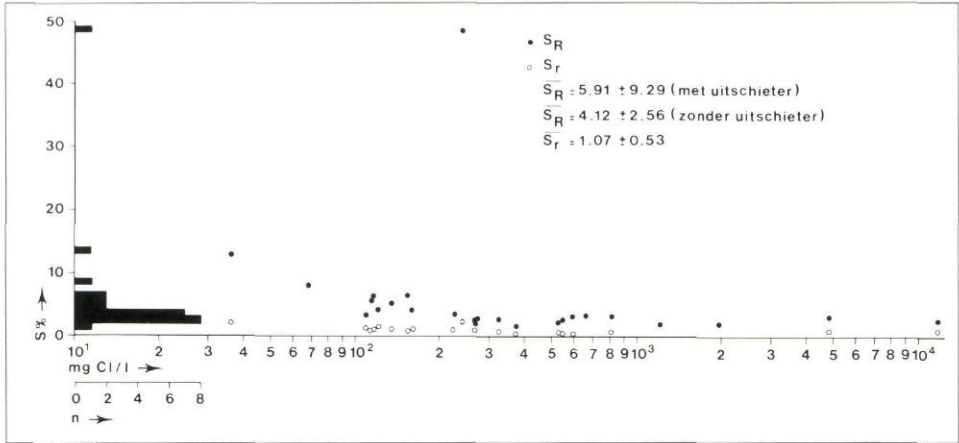
Aangezien het uitgangspunt vergelijkbaarheid van analyseresultaten is, is het 'gemiddelde = juist' aangenomen. De waardering van een A-klasse resultaat is een 10, van een B-klasse-resultaat een 7 en van een C-klasse resultaat een 3. In alle overige gevallen (Cochran-uitschieter, Dixon-uitschieter, geen of minder dan het aantal gevraagde analyseresultaten ingediend) wordt een nul gescoord.

Bespreking resultaten laboratorium-evaluaties 80 t/m '87

Met het oog op deze voordracht heb ik de gemiddelde standaardafwijkingen van de herhaalbaarheid ($\overline{s_r}$, in %) en van de reproduceerbaarheid ($\overline{s_R}$, in %) berekend. Een samenvatting is in de overzichtstabel gegeven (voor een nadere specificatie per parameter per matrix, zie referentie). De standaardafwijking van de herhaalbaarheid $\overline{s_r}$ is steeds, gemiddeld een factor 5 à 6, kleiner dan de overeenkomstige standaardafwijking van de reproduceerbaarheid $\overline{s_R}$. De tabel geeft een indruk van de onzekerheid welke kleeft aan de doorsnee analyse-resultaten van een doorsnee Nederlands waterlaboratorium. In het hiernavolgende zullen een aantal kanttekeningen worden gemaakt, welke voor een juiste interpretatie van de overzichtstabel noodzakelijk zijn.

– Monstername:

Onderzoek van monsters omvat een aantal aspecten. Naast analyse in het laboratorium van de aangeboden monsters, op welk aspect de overzichtstabel betrekking heeft, is monstername in het veld een belangrijk aspect. Doorgaans is monstername qua spreiding een zwakkere schakel dan de eigenlijke analyse in het laboratorium zeker voor een inhomogene matrix zoals baggerspecie. Monstername dient daarom bij elk onderzoek expliciet in de beschouwing te worden betrokken. Het door DBW/RIZA georganiseerde ringonderzoek betreft laboratoriumevaluatie. De overzichtstabel heeft dus betrekking op de prestaties van laboratoria vanaf de binnenkomst van de monsters tot en met de rapportage van de analyseresultaten. Hierbij blijft de variabiliteit van de monstername in het veld buiten beschouwing. Aangezien in het algemeen het aan een laboratorium aangeboden monster niet in zijn geheel in bewerking kan worden genomen, moet in het laboratorium hiervan een deelmonster worden getrokken: deze monstername in het laboratorium draagt



Afb. 1 - Standaardafwijkingen van de herhaalbaarheid en van de reproduceerbaarheid als functie van de chloride-concentratie. In het histogram is de frequentie (n) weergegeven van de standaardafwijkingen van de reproduceerbaarheid.

uiteraard wel bij in de waargenomen spreidingen. Evenals voor analyses in het laboratorium dient voor de monstername te worden uitgegaan van de doelstelling van het onderzoek. Richtlijnen met betrekking tot monstername zijn onder meer neergelegd in NEN 6600 en NPR 6601, in de internationale normenreeks ISO 5667 en in de Duitse normenreeks DIN 38402. Het aspect monstername, hoewel dus van eminent belang, laat ik in het kader van mijn voordracht verder buiten beschouwing.

– Verdelingen van s_r en s_R :

Aangezien standaardafwijkingen gedefinieerd zijn als de wortel uit de variantie, zijn standaardafwijkingen steeds groter dan of gelijk aan nul. Als de standaardafwijkingen gemiddelde voldoende van nul verschillen, kan de verdeling van de standaardafwijkingen naderen tot een normale verdeling. Naarmate de gemiddeld standaardafwijking nadert tot nul wordt de kans op een scheve verdeling groter. Een voorbeeld van zo'n scheve verdeling ziet u in afb. 1. In de verdeling van standaardafwijkingen van de reproduceerbaarheid is er

duidelijk sprake van een uitschieter ($s_R = 49\%$). Dit betrof een sterk gekleurd monster, waarvoor eindpuntsbepalingen met een kleurindicator ongeschikt zijn; in dit geval was de groep van deelnemende laboratoria die toch een dergelijke chloridebepaling uitvoerden zodanig groot dat deze resultaten niet statistisch werden verworpen. In de afb. is tevens het gemiddelde van de standaardafwijkingen van de reproduceerbaarheid met ($\overline{s_R} = 5,91 \pm 9,29\%$) en zonder ($\overline{s_R} = 4,12 \pm 2,56\%$) uitschieter weergegeven. Een relatief hoge waarde achter het $\pm$ teken is een indicatie dat er sprake is van een scheve verdeling en dat relatief ongunstige uitschieters (kunnen) voorkomen.

– s_R als functie van het gehalte:

Een aspect dat reeds in de overzichtstabel voor de metalen en tevens in afb. 1 naar voren komt is dat de procentuele spreiding groter wordt naarmate de gehalten lager worden. Voor metalen in oppervlaktewater is de spreiding ($\overline{s_R} = 41\%$) een factor 2 groter dan in afvalwater, zuiveringsslib of baggerspecie ($\overline{s_R} \approx 20\%$). In het algemeen is in het gebied van de laagste gehalten de absolute spreiding (bijv. in mg/l)

TABEL – Overzicht laboratoriumevaluaties 1980 t/m 1987: gemiddelde standaardafwijkingen⁹ van de reproduceerbaarheid ($\overline{s_R}$), in %.

Parametergroep	Oppervlaktewater	Afvalwater	Zuiveringsslib	Baggerspecie
Macroparameters	14 ± 6 ¹	6 ± 1 ²	—	30 ± 10 ³
Zwevende stof	23 ± 12	—	—	—
Metalen (totaal)	41 ± 14 ⁴	19 ± 17 ⁵	23 ± 15 ⁵	17 ± 8 ⁵
PAK	—	—	—	52 ± 3 ⁶
PCB's	—	—	—	59 ± 6 ⁷
OCB's	—	—	—	92 ± 17 ⁸

¹ Kjeldahl-N, NH₄⁺, NO₃⁻, ortho-P, totaal-P, Si, SO₄²⁻, Cl⁻

² Kjeldahl-N, totaal-P, Cl⁻, CZV

³ Kjeldahl-N, totaal-P, CZV, organische stof, CaCO₃, minerale delen 0-2µm en 2-16 µm, minerale olie

⁴ As, Cd, Cr, Cu, Hg, Ni, Pb, Zn, Fe, Mn, Ba, Be, Co, Se, V; bepaling van de opgeloste gehalten in dezelfde monsters leverde slechtere resultaten: 66 ± 26% (extra analysestap)

⁵ As, Cd, Cr, Cu, Hg, Ni, Pb, Zn

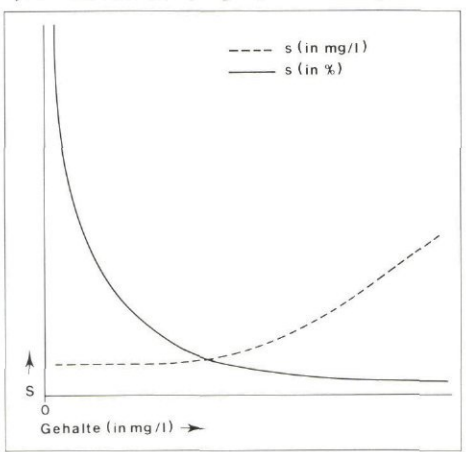
⁶ Flu, BkF, BbF, BghiPe, BaP, I123cdP

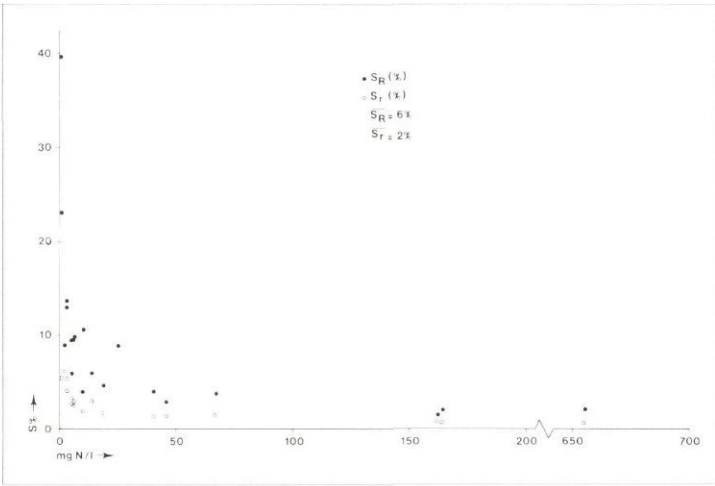
⁷ PCB28, 52, 101, 138, 153, 180

⁸ HCB, α-HCH, γ-HCH, Hepta, Aldrin, Hepo, ppDDE, dieldrin, endrin, opDDT, ppDDD, ppDDT, EOX

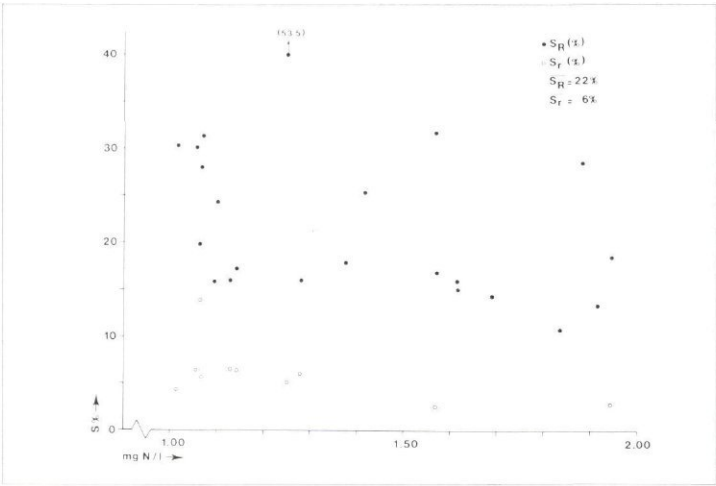
⁹ Tevens zijn weergegeven de standaardafwijkingen van $\overline{s_R}$. Een relatief hoge waarde achter het $\pm$ teken is een indicatie dat ongunstige uitschieters (kunnen) voorkomen.

Afb. 2 - Standaardafwijking als functie van het gehalte.





Afb. 3a - Verband tussen S en het gehalte voor Kjeldahl stikstof in monsters afvalwater.



Afb. 3b - Verband tussen S en het gehalte voor Kjeldahl stikstof in monsters oppervlaktewater.

constant door een allesoverheersende invloed van de spreiding in de zogenaamde blanco (ook aangeduid met: ruis, achtergrond, nuleffect); bij hogere gehalten zien we steeds het beeld dat de absolute spreiding tevens een functie wordt van het gehalte; vaak zien we bij de hoge gehalten uiteindelijk een lineair verband tussen de absolute spreiding en het gehalte. In afb. 2 is een en ander weergegeven. Uit het voorgaande volgt dat de procentuele standaardafwijking bij lage gehalten asymptotisch zal naderen tot de Y-as en bij heel hoge gehalten asymptotisch zal naderen tot een constant percentage. Het blijkt dat deze theoretische beschouwingen door de praktijk worden bevestigd, hetgeen voor Kjeldahl stikstof in afvalwater is geïllustreerd in afb. 3a en voor totaalfosfaat in afb. 4a. Ook voor Cr, Ni en Pb in afvalwater worden dergelijke duidelijke relaties gevonden. In andere matrices, waarin de gehalten veelal minder variëren dan in afvalwater, is dit verband veel minder uitgesproken (vergelijk afb. 3a met 3b en 4a met 4b). In afb. 5 zijn de resultaten voor het chemisch zuurstofverbruik (CZV) in monsters afvalwater weergegeven. U ziet hier dat s een

betrekkelijk geringe afhankelijkheid van het CZV-gehalte vertoont. Twee uitschieters worden veroorzaakt door zoute afvalwatermonsters met een laag CZV-gehalte, vooral doordat de meeste deelnemende laboratoria onvoldoende ervaring hebben met dergelijke monsters (chloride en bromide storen de CZV bepaling).

– Gunstige invloed normalisatie en regelgeving

Een uitschieter in positieve zin vormen enkele macroparameters in afvalwater met een standaardafwijking van de reproduceerbaarheid van gemiddeld 6%. Twee parameters, namelijk de bepaling van Kjeldahl stikstof en de bepaling van het chemisch zuurstofverbruik, zijn gerelateerd aan geld (heffingsparameters). Om aan het gelijkheidsbeginsel inhoud te geven is de wijze van uitvoering van de analyse door de wetgever hier imperatief voorgeschreven. Een gunstige invloed van normalisatie gecombineerd met controle regelgeving op de vergelijkbaarheid van analyseresultaten is hier duidelijk aantoonbaar. Een keerzijde van de medaille is dat een dergelijke regelgeving ook verstarrend kan werken.

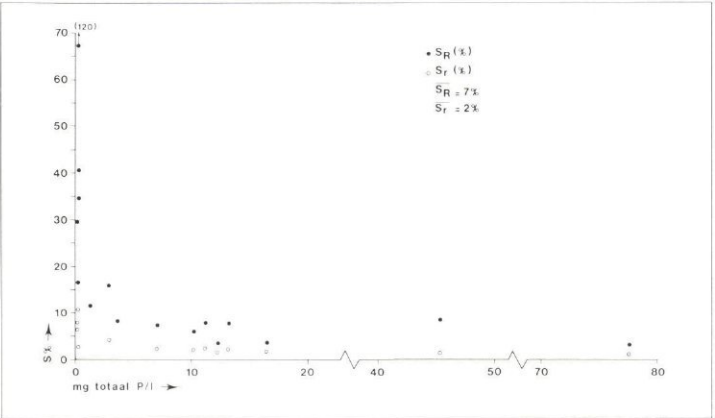
– Organische microverontreinigingen in baggerspecie

De bepaling van organische microverontreinigingen in baggerspecie springen er in de overzichtstabel in negatieve zin uit. De organochloorbestrijdingsmiddelen in baggerspecie scoren het slechtst met een $\overline{s}_R = 92\%$. U moet zich eens proberen voor te stellen wat dit resultaat inhoudt: Als het gemiddelde gevonden gehalte 100 is, dan vindt ruwweg $2/3$ van het aantal deelnemende laboratoria waarden tussen 8 en 192. Dit wil zeggen dat er tussen het hoogste en het laagste analyseresultaat een spreiding met een factor 24 zit. We hebben het dan nog alleen over die laboratoria die met hun analyseresultaat een 10 scoren! Voor de organochloorbestrijdingsmiddelen kunnen we dan ook eigenlijk niet meer spreken van gehaltebepalingen, er is eerder sprake van een bepaling van de ordegrrootte van een gehalte.

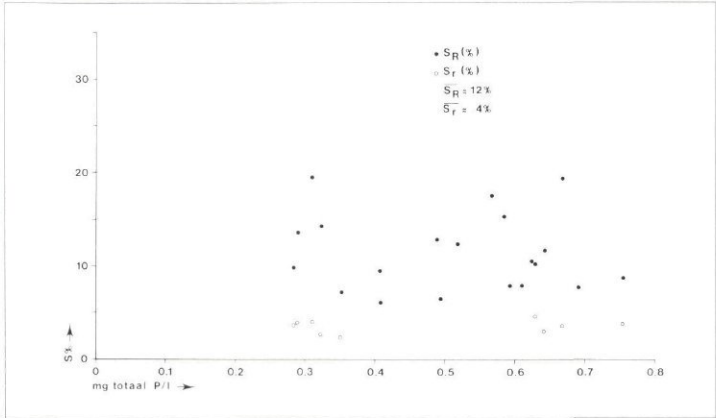
– Invloed extra analysestap

In de overzichtstabel ziet u voor het totaalgehalte van metalen in oppervlaktewater een $\overline{s}_R = 41\%$. In principe is elke extra stap in een analyse een potentiële bron van fouten.

Afb. 4a - Verband tussen S en het gehalte voor totaal fosfaat in monsters afvalwater.



Afb. 4b - Verband tussen S en het gehalte voor totaal fosfaat in monsters oppervlaktewater.



Een extra analysestap, namelijk afseparatie van de zwevende stof en vervolgens bepaling van de opgeloste gehalten in het filtraat resulteert dan ook in een verslechterd beeld, nl. $\overline{S_R} = 66\%$.

– Zwevende stof

In de overzichtstabel is de parameter zwevende stof apart gezet. Dit is een belangrijke parameter met betrekking tot de verspreiding van vele organische en anorganische microverontreinigingen, die zich aan de zwevende stof hechten. Als analytisch instrumentarium is de huidige zwevende stof bepaling daartoe naar mijn mening ontoereikend.

– Laboratorium certificatie

Laboratoriumevaluatie geven slechts een momentopname van de prestaties van de deelnemende laboratoria: de op basis van ringonderzoek aan te bevelen laboratoria kunnen voor het beschouwde parameterpakket en de beschouwde matrix (oppervlaktewater, afvalwater, zuiveringsslib, baggerspecie) bevredigende resultaten leveren. Benadrukt moet worden dat dit niet hoeft te betekenen dat ze dat in de praktijk van alledag ook doen, dat wil zeggen dat er geen garantie is dat een willekeurig te

tussen verschillende laboratoria. Als dus één laboratorium een parameter in een monster bepaalt, is de spreiding in de analysesresultaten gemiddeld 5 à 6 maal kleiner dan wanneer meerdere laboratoria die parameter in dat monster bepalen.

Deze resultaten hebben voor de praktijk van het wateronderzoek en -beheer wel enige consequenties. Er zijn hierbij een aantal aspecten te onderscheiden.

– Absolute analysesresultaten

In het geval van toetsing bijvoorbeeld aan normen of aan de voorwaarden van een lozingsvergunning zijn de absolute cijfers in het geding. Bij dergelijke toetsingen kunnen grote belangen gemoeid zijn. Te denken valt hierbij aan de berging van baggerspecie of aan het nemen van maatregelen tegen de overtreder van lozingsvoorwaarden.

In dergelijke gevallen kan het momenteel aanbeveling verdienen meerdere van elkaar onafhankelijke analysesresultaten in te winnen. Denk hierbij aan verschillende als goed te kwalificeren laboratoria en/of aan meerdere onafhankelijke analysemethoden, bijv. voor metalen: atoomabsorptiespectrometrie, atoomemissiespectrometrie, röntgenfluorescentiespectrometrie, polarografie, neutronenactiveringsanalyse, enz.

processen) en/of ruimte (bijv. lengte- en dwarsraaien in een rivier; variaties als functie van de diepte) worden onderzocht. In deze gevallen verdient het aanbeveling per te onderzoeken parameter één laboratorium in te schakelen. Men heeft dan te maken met de binnenlaboratoriumspreiding, die zoals hiervoor is aangegeven veel kleiner is dan de tussenlaboratoriumspreiding.

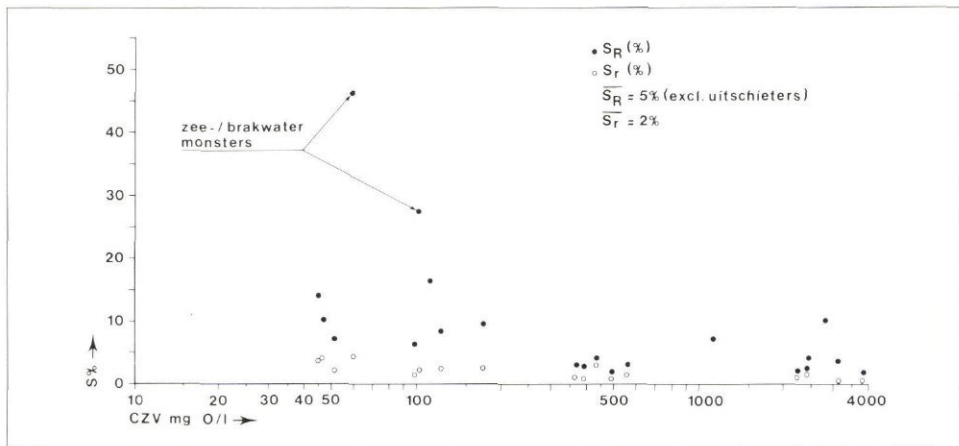
– Vergelijking met andere onderzoeken of gegevensbestanden

In vele gevallen zult u uw analysesresultaten toch willen vergelijken met die van andere onderzoeken of met reeds aanwezige gegevensbestanden. In deze gevallen is het van groot belang dat u zich hierbij eerst afvraagt met welke spreiding in de analysesresultaten u te maken kunt hebben. De tabellen welke zijn weergegeven in de desbetreffende DBW/RIZA notitie (zie referentie) geven u hiervoor enig houvast.

In de praktijk blijkt veelvuldig dat verschillen tussen analysesresultaten artefacten zijn, dit wil zeggen de geconstateerde verschillen zijn toe te schrijven aan spreidingen in het analyseproces tussen verschillende laboratoria of zelfs binnen éénzelfde laboratorium. Het belang van de vergelijkbaarheid van analysesresultaten en het zich bewust zijn van de mogelijk optredende spreidingen daarin is in toenemende mate evident bij de koppeling van watergegevensbestanden. Een computer maakt, als wij het hem niet vertellen, geen onderscheid tussen bijvoorbeeld een goed laboratorium of een minder goed laboratorium.

Literatuur

Veen, R. van (1988). *Relativiteit van analysesresultaten in het wateronderzoek en -beheer*. DBW/RIZA notitie nr. 88.023 X.



Afb. 5 - Verband tussen S en het gehalte voor het chemisch zuurstofverbruik in monsters afvalwater.

analyseren monster met dezelfde zorg wordt omgeven als een ringonderzoek rondzendmonster. Binnen de laboratoriumwereld wordt daarom gewerkt aan een systeem van laboratoriumerkenning en kwaliteitsborgingseisen voor laboratoria (Good Laboratory Practice, GLP) om de kwaliteit van de 'dagelijkse' laboratoriumwerkzaamheden met meer waarborgen te omgeven.

Consequenties voor wateronderzoek en -beheer

In het voorafgaande is reeds aangegeven dat de binnenlaboratoriumspreiding gemiddeld een factor 5 à 6 kleiner is dan de spreiding

Een dergelijke werkwijze is analoog aan de manier waarop gecertificeerde referentiematerialen worden verkregen door instellingen als het Bureau Communautaire de Référence (BCR) te Brussel, het National Bureau of Standards (NBS) in de USA en het International Atomic Energy Agency (IAEA) te Wenen.

– Relatieve analysesresultaten

In vele op zichzelf staande onderzoeken zijn veelal reeds de relatieve analysesresultaten toereikend om aan de onderzoeksdoelstelling te voldoen. Het betreft hier bijvoorbeeld onderzoeken, waarbij variaties in de tijd (bijv. trendmonitoring; het volgen van