

Omgaan met rapportagegrenzen van concentratiemetingen

Johan Post, Erik Liefting, Jeroen Langeveld (Partners4UrbanWater), Bert Palsma (STOWA)

In onderzoeken naar organische microverontreinigingen komt het veelvuldig voor dat een aanzienlijk deel van de metingen lager ligt dan de rapportagegrens. De wijze waarop bij het berekenen van gemiddelden wordt omgegaan met meetwaarden onder de rapportagegrens kan onbedoeld een groot effect op het berekende gemiddelde hebben. Dit artikel stelt voor om de methode Kaplan-Meier te gebruiken om onnodige systematische afwijkingen in berekende gemiddelden te voorkomen.

Decennialang stond reductie van de emissies van zuurstofbindende stoffen uit de riolering en van nutriënten uit de afvalwaterzuivering bovenaan de agenda van gemeenten en waterschappen. Inmiddels zijn hier de nodige stappen in gezet en krijgen opkomende stoffen steeds meer aandacht. Veel van deze opkomende stoffen, zoals bestrijdingsmiddelen, medicijnresten, industriële chemicaliën en sommige zware metalen en polycyclische aromatische koolwaterstoffen (PAKs), komen vaak in zulke lage concentraties voor dat ze niet goed meetbaar zijn. Rapportages van metingen van organische microverontreinigingen in hemelwater bevatten vaak de aanduiding dat een concentratie 'kleiner dan' (<) een bepaalde waarde is, de zogenoemde rapportagegrens (RG). De detectiegrens (DG) is de laagste concentratie waarbij de aanwezigheid van een bepaalde stof in het monster kan worden opgemerkt. De rapportagegrens is de laagste getalswaarde van een bepaling van een stof in een laboratorium die nog kwantitatief goed kan worden vastgesteld. De detectiegrens is per definitie gelijk aan of lager dan de rapportagegrens. De hoogte van de rapportagegrens verschilt per analysemethode. Ook is bekend dat sommige organische microverontreinigingen al toxisch zijn of andere nadelige milieueffecten hebben bij zeer lage concentraties, soms ook onder de (met de huidige technieken haalbare) rapportagegrenzen.

Bij het vertalen van de onderzoeksresultaten naar een gemiddelde of een jaarvrucht ontstaat het probleem van de keuze voor de omgang met de waarnemingen die lager liggen dan de rapportagegrens (<RG). In het vakgebied van stedelijk water, riolering en afvalwaterzuivering zijn verschillende methodes in gebruik, variërend van weglaten of substitutie door 0%, 50% of 75% van de rapportagegrens, tot geavanceerdere methodes, zoals de Volkert-Bakkermethode [1], [2]. Voor opkomende stoffen, die problematisch zijn bij lage concentraties en gekenmerkt worden door veel waarnemingen nabij en onder de rapportagegrens, is de keuze voor een methode bepalend voor de betrouwbaarheid van het afgeleide gemiddelde.

In het kader van twee STOWA-projecten, de hemelwaterdatabase 2020 en metingen aan organische microverontreinigingen in influent, is onderzoek gedaan naar methoden om om te gaan met waarnemingen <RG.

Aanpak

Als eerste stap is een overzicht gemaakt van de verschillende aanpakken. De verschillende methodes geven allemaal een andere invulling aan de vraag hoe een waarneming <RG getalsmatig kan worden meegenomen. Dit overzicht is weergegeven in tabel 1.

Tabel 1. Overzicht methoden voor omgang met rapportagegrenzen

Naam methode	Omschrijving
Weglating	Alle waarnemingen <RG niet meer meenemen
Eenvoudige substitutiemethodes	Substitutie waarneming <RG door: 0% van RG [1] 50% van RG [1] 1/√2 van RG [3] 75% van RG [4]
Volkert-Bakkermethode [2]	Het aantal waarnemingen <RG wordt uitgedrukt in een percentage ten opzichte van het totaal aantal waarnemingen. Substitutie vindt plaats met de waarde die volgt uit: $(100\% - \text{'percentage < RG'}) * RG$
Variant-Baltussen op Volkert-Bakker [2]	Nadeel van Volkert-Bakker is dat deze niet kan omgaan met meerdere getalswaarden voor de RG. De variant-Baltussen houdt in dat de verschillende getalswaarden van de RG rekenkundig worden gemiddeld. Vervolgens wordt de Volkert-Bakkermethode toegepast

De verschillen in aanpak zijn enorm en bestrijken het interval tussen substitutie door 0% en 75% van de rapportagegrens. Uit eerder onderzoek [2] volgt dat alleen methodes die substitutie afhankelijk maken van het percentage waarnemingen <RG (Volkert-Bakker en varianten daarop), leiden tot bruikbare resultaten. De uitdaging ligt daarbij wel in situaties waarbij meerdere rapportagegrenzen bestaan per stof. Dit is bij rioolwater gebruikelijk, omdat het monster flink moet worden verdund vanwege de vuile matrix. De variant van Baltussen op de methode-Volkert-Bakker voorziet hierin.

Als tweede stap is naast de gangbare methodes uit tabel 1 ook gezocht naar alternatieven. Het alternatief dat naar voren kwam uit de literatuur is toepassing van de methode-Kaplan-Meier [5]. De methode is oorspronkelijk ontwikkeld voor de analyse van de overlevingskans als functie van de tijd ('survival functions'), maar kan ook worden gebruikt om om te gaan met waarnemingen <RG [6]. De Kaplan-Meiermethode is breed toepasbaar voor de statistiek van concentratiemetingen, omdat:

- in deze zgn. non-parametrische methode geen a priori aanname hoeft te worden gedaan over de statistische verdelingsfunctie;
- de methode zowel geschikt is voor 'left-censored' als 'right-censored' data – respectievelijk concentraties onder de detectielimiet en concentraties boven het meetbereik; beide komen voor in concentratiemetingen.

De Kaplan-Meierfunctie is beschikbaar in gangbare software voor statistische bewerkingen, zoals MATLAB, Python en R. De methode is nader omschreven in het kader.

De Kaplan-Meiermethode

De Kaplan-Meiermethode kan worden gebruikt voor het bepalen van samenvattingsstatistieken, zoals het gemiddelde of de standaarddeviatie, voor datasets waarin een deel van de gemeten concentraties lager dan de rapportagegrens (RG) liggen. Het voordeel van Kaplan-Meier ten opzichte van soortgelijke methodes is het kunnen omgaan met meerdere rapportagegrenzen en dat geen verdere informatie over de verdeling van de gemeten concentraties nodig is. Dit laatste betekent dat het niet uitmaakt welke vorm het histogram van concentraties voor een bepaalde stof heeft. De enige voorwaarde voor toepasbaarheid is dat minimaal twee gemeten concentraties hoger dan de rapportagegrens liggen.

Het principe van de methode berust op het bijhouden van een boekhouding waarbij steeds alle meetwaarden onder een gemeten concentratie geteld worden. Dit geschiedt door meetwaarden te sorteren van groot naar klein, waarna het aandeel concentraties lager dan elke unieke waarde wordt berekend. Wanneer dit in een grafiek wordt uitgezet is het gemiddelde gelijk aan de oppervlakte onder de grafieklijn. In het volgende voorbeeld wordt de methode stapsgewijs toegepast.

Voorbeeld

Voor stof X zijn de volgende concentraties gemeten in afstromend hemelwater: 9, <6, 12, <5, 9, <10, 11 mg/l.

- 1) Sorteert de gemeten concentraties van groot naar klein. Indien de laatste waarde <RG is, wordt deze gedeeld als 'gemeten concentratie' in plaats van <RG, om een systematische afwijking van het gemiddelde te voorkomen (zie ook de vetgedrukte waarde in de laatste rij van tabel k.1). Maak vervolgens een tabel van alle uniek gemeten concentraties.
- 2) Voeg een kolom toe met het aantal keer dat deze of een kleinere concentratie gemeten is.
- 3) Voeg een kolom toe met daarin het aantal keer dat een is concentratie gemeten, mits boven de RG.

Dit geeft de volgende tabel voor dit voorbeeld. Zie ook hoe de laatste waarde ontdaan is van het RG label (<-teken).

Tabel k.1. Overzicht gemeten concentraties in afstromend hemelwater

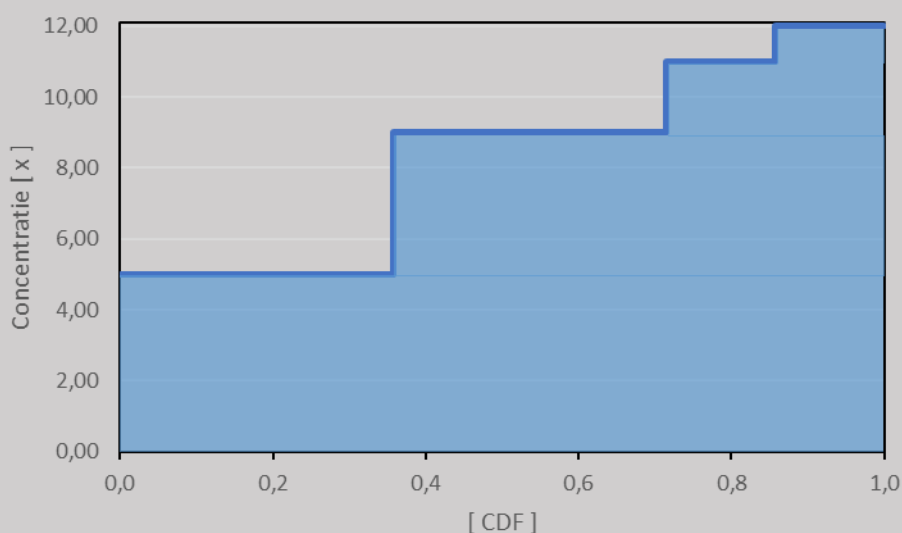
Unieke concentraties	Aantal meetwaarden met deze of lagere concentratie	Aantal meetwaarden boven RG
[x]	[n]	[d]
12	7	1
11	6	1
<10	5	0
9	4	2
<6	2	0
5	1	1

- 4) Bereken de statistiek $F = 1 - \frac{d}{n}$
- 5) Voeg een kolom toe met de naam CDF, waarin rij één ($i=1$) de waarde 1,000 bevat. De overige rijen worden ingevuld met de formule $CDF_i = CDF_{i-1} \times F_{i-1}$, oftewel voor rij twee ($i=2$) volgt $0,857 \times 1,000 = 0,857$ en voor rij drie ($i=3$) $0,833 \times 0,857 = 0,714$

Tabel k.2: Uitwerking berekening voor gemiddelde concentratie

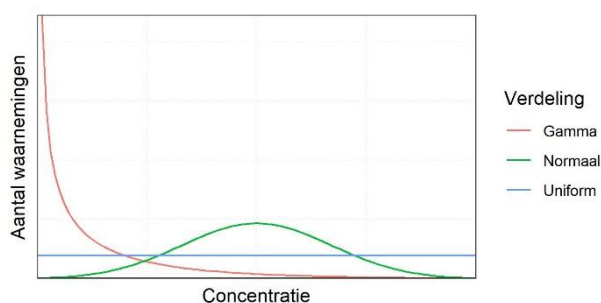
Unieke concentraties	Aantal meetwaarden met deze of lagere concentratie	Aantal meetwaarden boven RG	Oppervlakte kolom		
[x]	[n]	[d]	[F]	[CDF]	[Opp]
12	7	1	0,857	1,000	1,714
11	6	1	0,833	0,857	1,571
<10	5	0	1,000	0,714	0,000
9	4	2	0,500	0,714	3,214
<6	2	0	1,000	0,357	0,000
5	1	1	0,000	0,357	1,786
				0,000	

- 6) De grafiek van unieke concentraties x uitgezet tegen kolom CDF is in afbeelding k.1. weergegeven. De gemiddelde concentratie, aangepast voor meetwaarden onder de rapportagegrens, wordt gegeven door het oppervlak onder deze grafiek. Door het oppervlak in figuur k.1. op te delen in kolommen en het oppervlak van de gevormde rechthoeken (zie ook de laatste kolom van tabel 2) op te tellen wordt het gemiddelde berekend. In formulevorm is dit $\sum x \cdot (CDF_{i+1} - CDF_i)$, oftewel $12 \cdot (1 - 0,857) + 11 \cdot (0,857 - 0,714) + \dots + 5 \cdot (0,357 - 0,000) = 8,286$



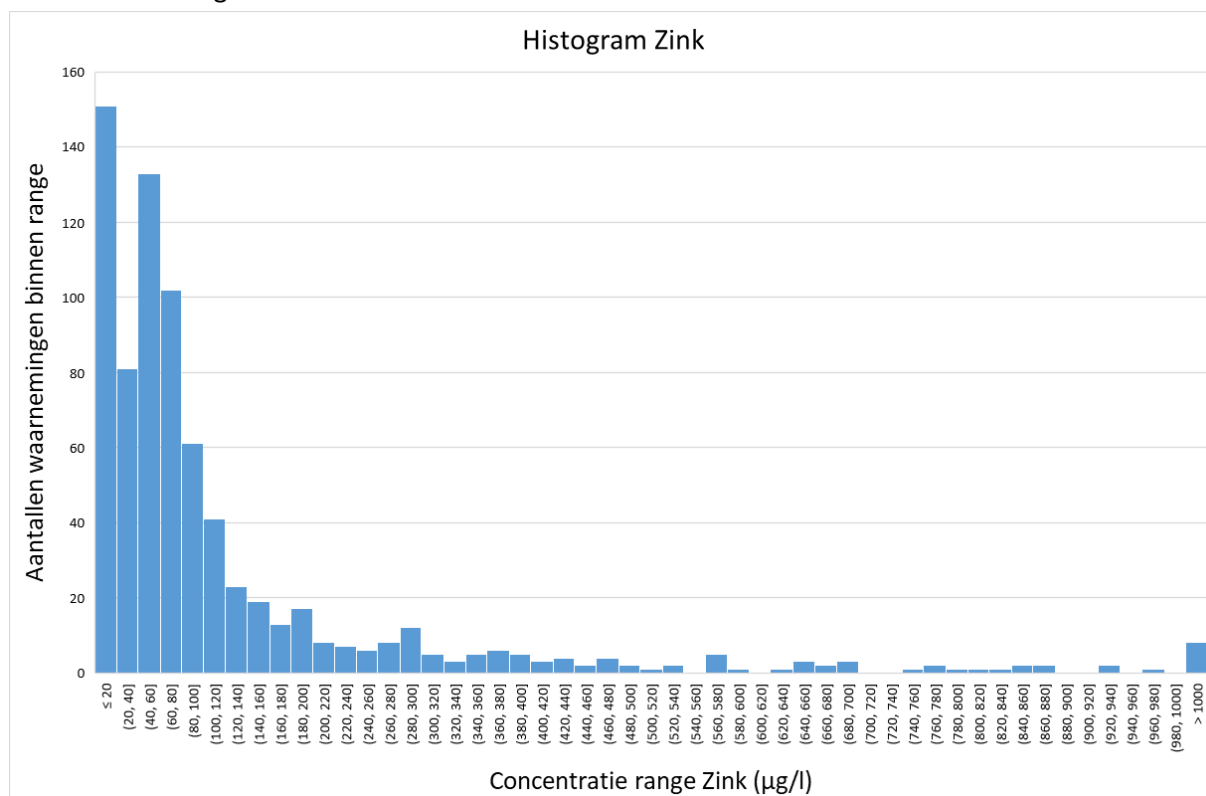
Afbeelding k.1: Spreidingsgrafiek van de gemeten concentratie x uitgezet tegen CDF . De gemiddelde concentratie wordt gegeven door het oppervlak onder de grafiek

Als derde stap zijn vervolgens verschillende methodes getest door deze toe te passen op een dataset van kunstmatige waarnemingen, bestaande uit tien trekkingen uit een kansverdeling, waarna vervolgens het aantal waarnemingen <RG is gevarieerd door de fictieve RG te leggen op het eerste tot en met het achtste getrokken getal. Dit is 1000 keer gedaan voor drie verschillende kansverdelingen: een gammaverdeling, de 'scheve' kansverdeling die het beste past op een dataset met veel waarnemingen met lage waarden en enkele hogere waarnemingen, een normale verdeling, die bestaat uit een symmetrische kansverdeling rondom het gemiddelde, en een uniforme verdeling, waarbij elke waarde binnen een bereik dezelfde kans van voorkomen heeft (zie ook afbeelding 1).

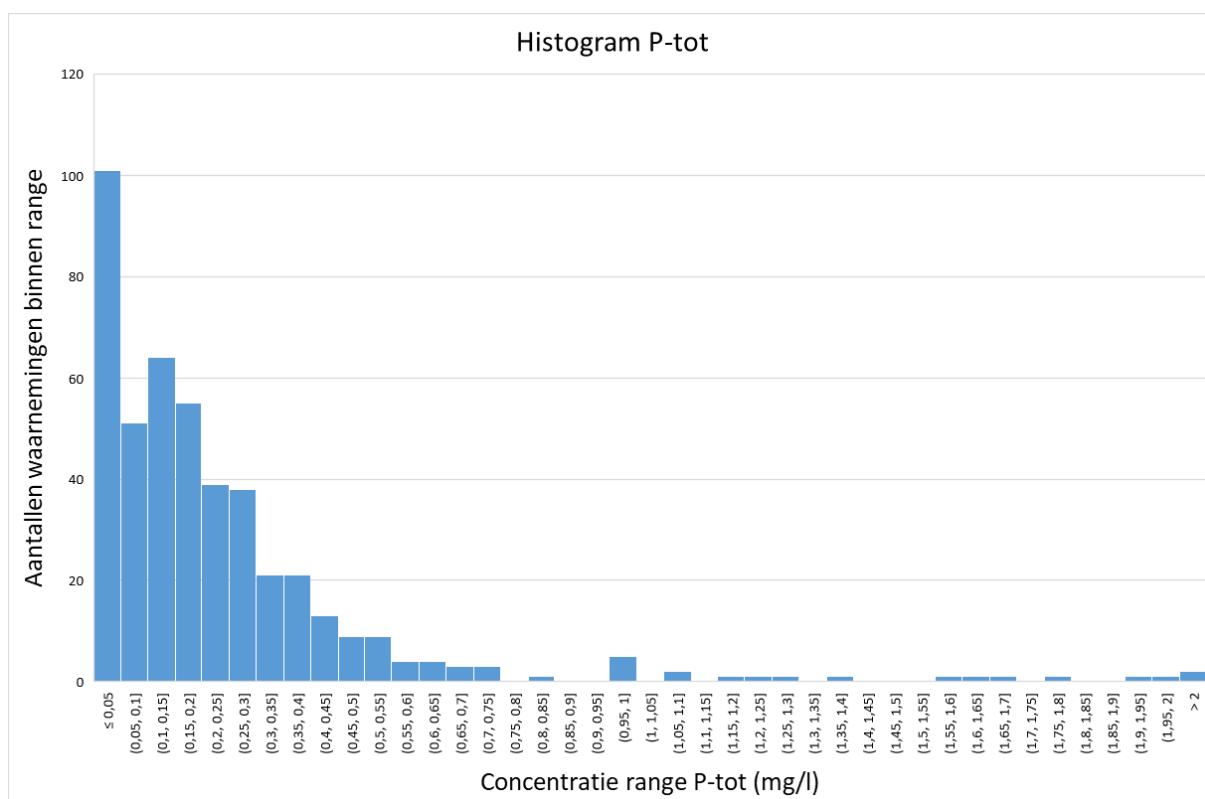


Afbeelding 1. Opbouw van de kansverdelingen waarmee kunstmatige waarnemingen zijn geproduceerd

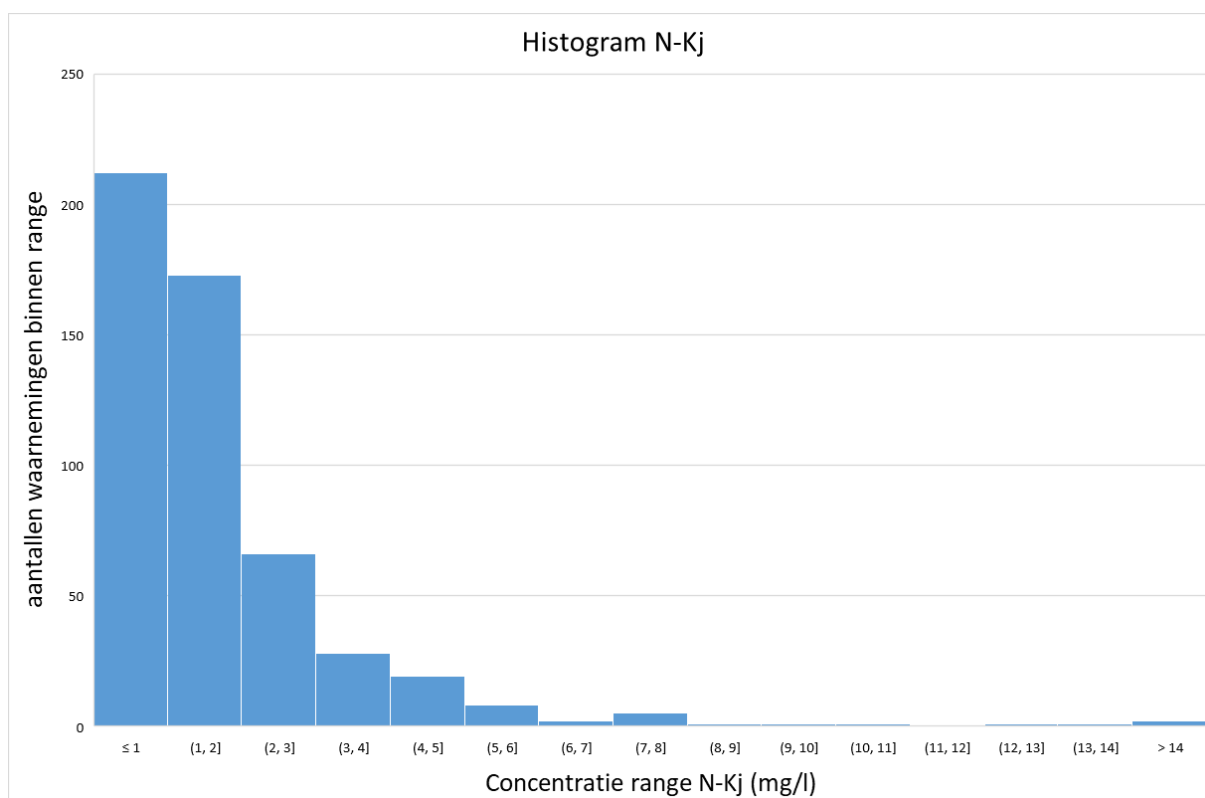
De afbeeldingen 2 tot en met 5 geven een histogram van de gemeten concentraties uit de STOWA-hemelwaterdatabase voor zink, fosfaat, Kjeldahl-stikstof (N-Kj) en nitraat. Voor zink, fosfaat en Kjeldahl-stikstof lijkt de verdeling op een gammaverdeling, terwijl deze voor nitraat meer lijkt op een normale verdeling.



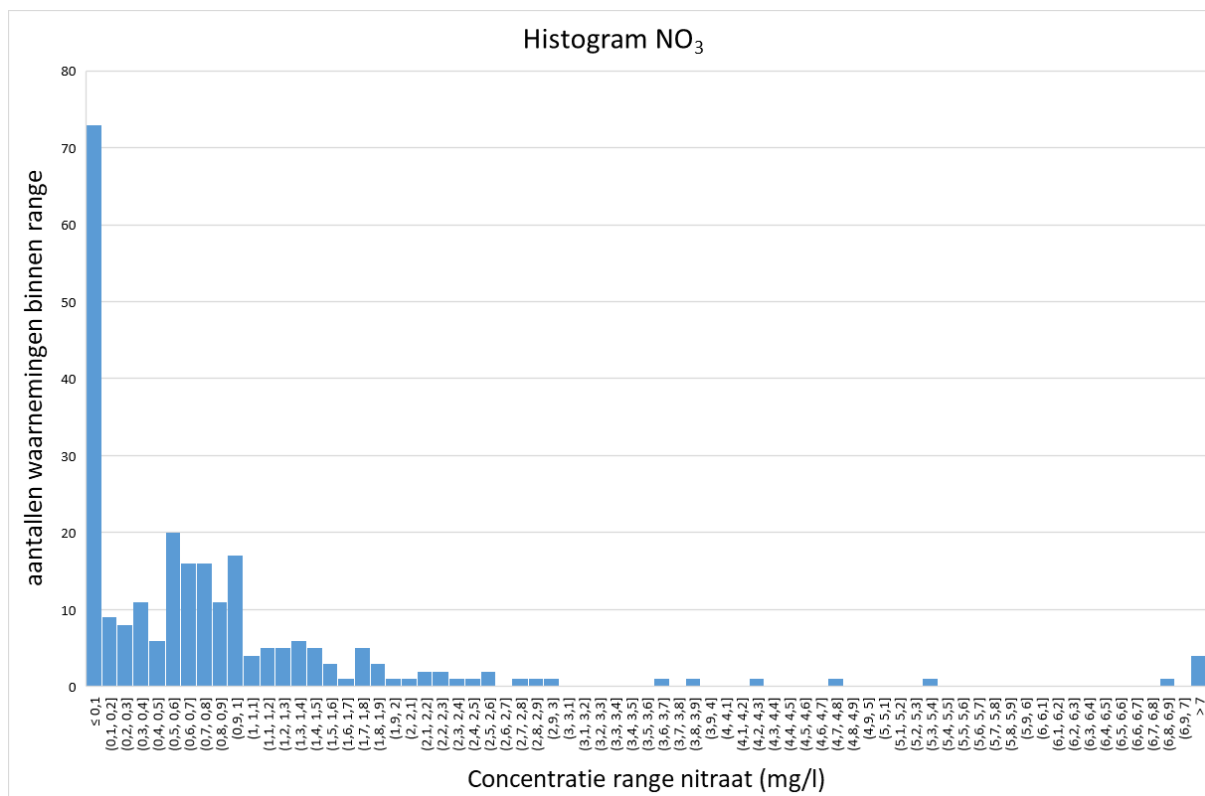
Afbeelding 2. Histogram metingen zink in afstromend hemelwater



Afbeelding 3. Histogram metingen fosfaat in afstromend hemelwater



Afbeelding 4. Histogram metingen N-Kjeldahl in afstromend hemelwater



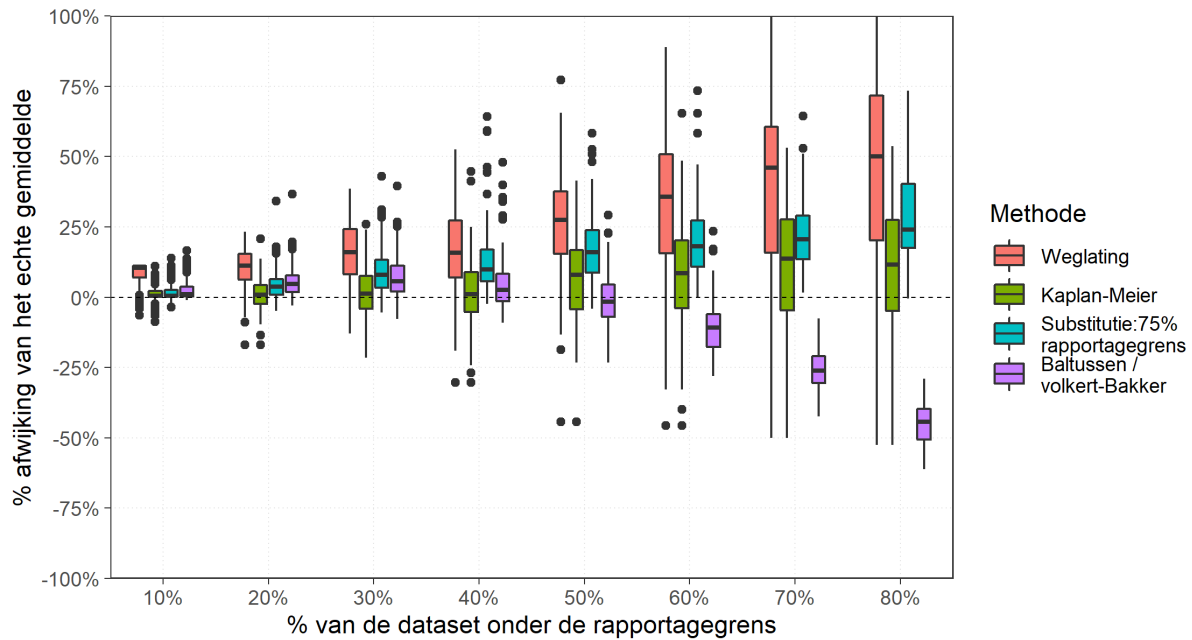
Afbeelding 5. Histogram metingen nitraat in afstromend hemelwater

Door gebruik te maken van kunstmatige waarnemingen is het ‘werkelijke’ gemiddelde bekend en is het mogelijk om te bepalen welke methode het beste omgaat met rapportagegrenzen.

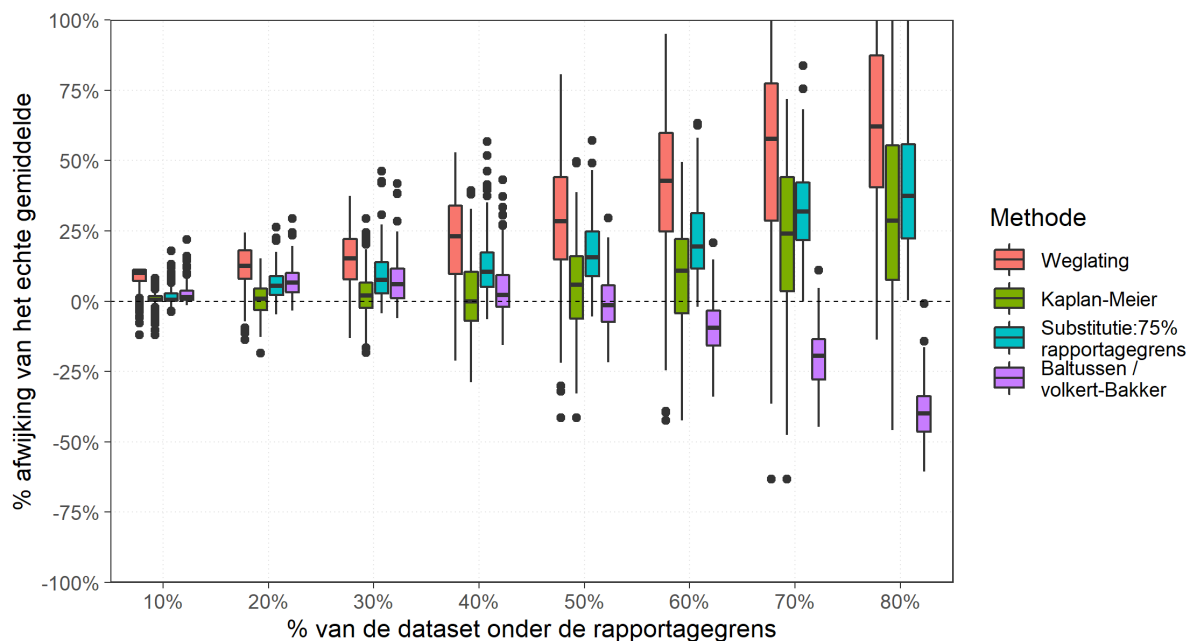
Resultaten en discussie

De resultaten voor de normale en de uniforme verdeling zijn weergegeven in afbeeldingen 6 en 7. De resultaten van de analyse zijn per methode weergegeven als een boxplot, met in de gekleurde box het gemiddelde als zwarte streep en de randen van de gekleurde box als 25- en 75-percentielwaarde.

In beide gevallen leidt de methode van substitutie met 75% van de RG tot een systematische afwijking (bias) naar boven. De variant-Baltussen van de Volkert-Bakkermethode leidt bij percentages waarnemingen <math>< RG</math> vanaf 60% tot een steeds grotere systematische onderschatting, terwijl de Kaplan-Meiermethode voor deze percentages juist leidt tot een overschatting, zij het dat deze bias kleiner is dan de onderschatting van de variant-Baltussen. De spreiding in de methode-Kaplan-Meier is wel groter dan die in de variant-Baltussen.



Afbeelding 6. Vergelijking methoden voor omgang met rapportagegrenzen bij een normale verdeling



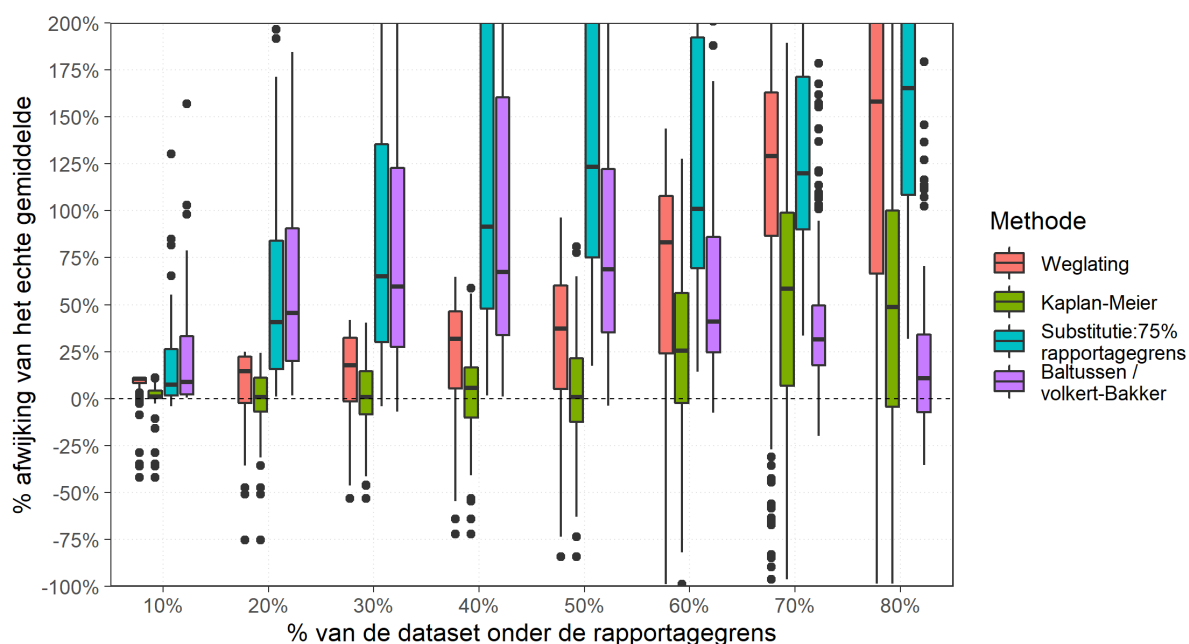
Afbeelding 7. Vergelijking methoden voor omgang met rapportagegrenzen bij uniforme verdeling

Afbeelding 8 toont de vergelijking van de resultaten bij de gammadistributie, de verdeling die het meest lijkt op de datasets in de STOWA-hemelwaterdatabase.

De tot voor kort in de STOWA-regenwaterdatabase toegepaste methode van substitutie door 75% van de rapportagegrens geeft zeer forse afwijkingen van het echte gemiddelde. Alleen als 10% van de dataset <RG is, valt de fout nog mee.

De variant-Baltussen op de methode van Volkert-Bakker leidt bij kleine (10%) en bij hele grote percentages van de dataset <RG (80%) tot een redelijke benadering. Daar tussenin varieert de systematische afwijking gemiddeld gezien ruim 50% van het gemiddelde, hetgeen een aanzienlijke fout is. De methode-Kaplan-Meier heeft tot een percentage <RG van 50% geen systematische afwijking en

een beperkte spreiding. Bij hogere percentages <RG begint de systematische fout op te lopen, tot een maximum van ruim 50% bij 70% waarnemingen <RG. Tot 60% waarnemingen <RG scoort Kaplan-Meier het best, bij 70 of 80% waarnemingen <RG scoort de variant-Baltussen het best.



Afbeelding 8. Vergelijking methoden voor omgang met rapportagegrenzen bij gammadistributie. Percentage afwijking van het echte gemiddelde voor vier methodes: weglating, substitutie door 75% Rapportagegrens, Kaplan-Meier en Baltussen, afhankelijk van het percentage van de dataset <RG

De resultaten bevestigen dat weglating of substitutie met een vaste waarde leidt tot relatief grote afwijkingen in het berekende gemiddelde. De methode-Kaplan-Meier doet het bijna altijd redelijk tot goed, met uitzondering van situaties met heel scheve data en meer dan 60% data <RG. Alleen in deze situatie scoort de Volkert-Bakkermethode beter. Het probleem bij kleine datasets zoals gebruikt in de test (10 stuks) en 7 of meer waarnemingen <RG is dat de bijbehorende kansverdeling niet (goed) te bepalen is. Om die reden is ook in een dergelijk geval Kaplan-Meier te verkiezen, omdat deze voor de verschillende verdelingen het beste scoort.

Dé methode voor de omgang met waarnemingen <RG bestaat niet. De prestatie van de verschillende methodes is afhankelijk van de verdeling van de waarnemingen (b.v. vooral lage of extreme waarden, of meer gelijk verdeeld) en het percentage waarnemingen <RG.

Concentraties van (micro)verontreinigingen in hemelwater en rioolwater hebben doorgaans een scheve verdeling (zie ook afbeeldingen 2 tot 5), wat tot uitdrukking komt in het gegeven dat het gemiddelde van de metingen duidelijk hoger ligt dan de mediaan. De achterliggende verdeling is meestal niet eenvoudig te bepalen. In de praktijk van milieukundig onderzoek hebben meetdata zelden een eenduidige statistische verdeling [6].

Conclusie

Op basis van de resultaten van de vergelijking van methodes voor de omgang met metingen <RG, is de conclusie getrokken dat de methode-Kaplan-Meier de beste methode is voor datasets met een percentage metingen <RG tot maximaal 65%. Is het percentage waarnemingen <RG groter dan 65%, dan is het advies om zowel de methode-Kaplan-Meier als de methode-Volkert-Bakker toe te passen.

Is het verschil tussen de uitkomsten van beide methoden groot (bijvoorbeeld >20%), dan moet in de rapportage worden aangegeven dat de berekende gemiddelde waarde erg onzeker is.

Referenties

1. Baltussen, J.J.M. (2010). *Emissie onderzoek op een zestal rwzi's in het kader van de E-PRTR*. STOWA 2010-W07
2. Baltussen, J.J.M. (2013). *Watergerelateerde emissies vanuit rwzi's in het kader van de iPRTR*. STOWA 2013-W01
3. Sanford, R.T., Pierson, C.T., Crovelli, R.A. (1993) 'An objective replacement method for censored Geochemical data'. *Math Geol*; 25: 59–80.
4. Boogaard, F., Lemmen, G. (2007). *De feiten over de kwaliteit van afstromend regenwater*. STOWA 2007-21
5. Kaplan, E. L., Meier, P. (1958). 'Nonparametric estimation from incomplete observations'. *J. Amer. Statist. Assoc.* 53 (282): 457–481
6. Helsel, D., (2010). 'Much Ado About Next to Nothing: Incorporating Nondetects in Science'. *Ann. Occup. Hyg.*, Vol. 54, No. 3, pp. 257–262, 2010