

Vergelijkende studie van analysetechnieken voor het meten van concentraties in bodemstalen.

Joost Vanoverbeke

Inhoudsopgave

1	Context	1
2	Onduidelijkheden in vraagstelling en datastructuur	1
3	Vraagstelling	2
4	Methode	2
5	Resultaten en bespreking	3
5.1	Analyse op afwijkingen tussen de referentieconcentratie en de gemeten concentratie.	4
5.2	Korte inspectie van precisie	7
6	Conclusie	7
7	Aanbevelingen voor verdere analyse	7

1 Context

Het laboratorium overweegt een nieuwe analysetechniek te gebruiken om bepaalde concentraties in bodemstalen te meten. Ze hebben een experiment opgezet waarbij ze een aantal stalen gemeten hebben met de oude en nieuwe techniek. Op basis van de opgeleverde gegevens vragen ze een uitspraak te doen over welke techniek het beste is.

2 Onduidelijkheden in vraagstelling en datastructuur

De vraagstelling zoals doorgegeven is niet helemaal duidelijk aangezien ‘het beste’ verschillende deelaspecten kan omvatten. In de eerste plaats is er **precisie**, waarmee bedoeld wordt hoeveel verschil (variatie) er zit tussen de verschillende replicate metingen voor een bepaalde techniek. In de tweede plaats kan er gekeken worden naar systematische **afwijkingen** (bias) in de metingen ten opzichte van de ‘ware’ concentratie. Meet een techniek systematisch een te hoge of te lage concentratie?

Verder is ook de betekenis van ‘analysetechniek’ niet duidelijk beschreven. Gaat het hier om verschillende technieken die kunnen uitgevoerd worden met dezelfde toestellen? Of betekent een verschil in techniek, een verschil in type meettoestellen? Dit kan gevolgen hebben voor de manier waarop een experiment opgesteld en geanalyseerd wordt.

Ook omtrent de gegevens zijn er een aantal onduidelijkheden, vooral met betrekking tot kolom *Staal*. De kolom *Staal* geeft volgens de beschrijving de unieke naam van elk staal. Dit zou dus betekenen dat bijvoorbeeld het staal ‘ref_laag_100_a’ voor *Toestel* = ‘1’, *Type* = ‘OUD’ en *Prep* = ‘CR’ hetzelfde staal betreft als ‘ref_laag_100_a’ voor *Toestel* = ‘11’, *Type* = ‘NIEUW’ en *Prep* = ‘ML’, en dat de gemeten concentraties voor de verschillende behandelingen gebaseerd zijn op substalen uit dit staal. Hierbij

wordt dan verondersteld dat de aanduiding ‘_a|_b|_c’ onafhankelijke stalen aangeeft. Anderzijds is het mogelijk dat juist die aanduiding (‘_a|_b|_c’) substalen aangeeft genomen uit hetzelfde staal en dat de stalen ‘ref_laag_100|ref_mid_320|ref_hoog_1000’ uniek en onafhankelijk zijn voor elke combinatie van *Toestel*, *Type* en *Prep*. Zonder verdere bijkomende informatie is het tenslotte niet uitgesloten (zij minder waarschijnlijk) dat alle stalen in de dataset onafhankelijk zijn, of dat alle stalen voor een bepaald niveau ‘ref_laag_100|ref_mid_320|ref_hoog_1000’ substalen zijn (dus inclusief ‘_a|_b|_c’) van hetzelfde staal.

Daarnaast lijken de aanduidingen ‘ref_laag_100|ref_mid_320|ref_hoog_1000’ een vooraf beoogde of opgelegde concentratie aan te geven in de stalen, namelijk concentraties van 100, 320 en 1000 respectievelijk. Dit is belangrijk omdat dergelijke informatie nodig is voor het bepalen van de afwijking van de metingen ten opzichte van de ‘ware’ concentratie.

Naast de stalen ‘ref_laag_100|ref_mid_320|ref_hoog_1000’ bevat de dataset ook stalen gelabeld als ‘staal_123’. Hierbij is er geen verdere aanduiding van ‘_a|_b|_c’, wat volgens de meest eenvoudige interpretatie aangeeft dat het om slechts één staal gaat waaruit substalen zijn genomen voor concentratiemetingen in de verschillende behandelingen. Zonder verdere informatie is het niet duidelijk wat het beoogde doel is van deze (sub)stalen in de dataset. Er kunnen eventueel wel verschillen tussen behandelingen aangetoond worden (voor dit ene staal!) maar welke vraagstelling moet hier dan aan gekoppeld worden? (vb. de vraag naar verschillen in afwijking kan hiermee niet beantwoord worden, omdat er geen referentiepunt is om de concentraties mee te vergelijken)

Verder is, zoals hoger al vermeld, de betekenis van de variabele *Type* niet geheel duidelijk. Gaat dit om twee types methodologie die op dezelfde toestellen kunnen uitgevoerd worden, of gaat het om verschillende types toestellen? Met andere woorden is er een vaste link tussen kolom *Toestel* en kolom *Type*?

Om deze onduidelijkheden in vraagstelling en datastructuur te verhelderen, en er zeker van te zijn dat de vraagstelling en de gegevens correct geïnterpreteerd worden, is terugkoppeling met het labo nodig. In afwachting hiervan worden hieronder, op basis van de eigenschappen van de dataset, een aantal aannames gedaan omtrent vraagstelling en gegevens. Op basis van deze aannames wordt dan een analyse gedaan van de gegevens om de vermoedelijke vraagstelling te beantwoorden.

- De hoofddoelstelling is om na te gaan of er een verschil is in **afwijking** van de gemeten concentraties ten opzichte van de ‘ware’ concentratie.
- Metingen met hetzelfde label in de kolom *Staal* zijn effectief gebaseerd op substalen uit hetzelfde staal. Elk uniek label geeft dus een uniek staal aan.
- de ‘ware’ concentraties in de stalen kunnen afgeleid worden uit de labels in kolom *Staal*, met respectievelijk concentraties van 100, 320 en 1000 voor ‘ref_laag_100’, ‘ref_mid_320’ en ‘ref_hoog_1000’.
- Voor het staal gelabeld als ‘staal_123’, is er geen verdere informatie voorhanden in verband met de echte of beoogde concentratie in dit staal. Bovendien is er in dit geval geen replicatie, en kunnen er geen uitspraken gedaan worden die verder rijken dan de eigenschappen voor dit specifieke staal. Deze gegevens worden daarom tot nader order uit de dataset verwijderd. Alvorens dit staal eventueel verder te behandelen is er terugkoppeling nodig met het labo.
- *Type* betekent een verschillende techniek die gelinkt is aan een verschillend type toestel. (cf. De toestellen zijn wederzijds exclusief voor de verschillende levels van *Type*)

3 Vraagstelling

Zoals hierboven vermeld, wordt nagegaan of er een verschil tussen de analysetechnieken is in **afwijking** van de gemeten concentraties ten opzichte van de echte concentraties (referentieniveaus).

4 Methode

De volledige analyse werd uitgevoerd in R.

Zoals hierboven vermeld wordt in de analyse van **afwijking** enkel gebruik gemaakt van de referentiestalen, en worden de stalen gelabeld als ‘staal_123’ uit de dataset verwijderd.

Om een vergelijking met de referentie of ‘ware’ concentraties te maken, wordt op basis van de informatie in de labels een nieuwe continue variabele *Ref_num* gecreëerd, die de referentieconcentraties weergeeft. Op basis hiervan wordt dan de afwijking in concentraties ten opzichte van de referentieconcentratie berekend als: $Bias_log = \log(Concentratie/Ref_num)$ (de log transformatie is hier nodig om afwijkingen groter en kleiner dan de referentiewaarde hetzelfde gewicht te geven). In vergelijking met een directe analyse op de gemeten concentraties, heeft een analyse van de afwijking het voordeel dat er rechtstreeks kan worden nagegaan hoe groot het verschil is tussen de gemeten concentraties en de referentie concentratie, en of deze afwijking verschilt tussen de behandelingen.

Voor de eigenlijke analyse wordt er geopteerd voor een frequentist benadering en worden de data geanalyseerd aan de hand van een LMM (Linear Mixed-effects Model). Onafhankelijke categorische variabelen zijn hierbij *Ref_cat* (referentieconcentraties als categorische variabele), *Type* en *Prep*. *Toestel* (genest in *Type*) en *Staal* (genest in *Ref*) worden toegevoegd als categorische random variabelen.

Uit een eerste visuele verkenning van de gegevens en uit preliminaire analyses (zie Rmarkdown script) blijkt dat er ook rekening moet worden gehouden met interacties tussen de onafhankelijke variabelen. Om het éénvoudig te houden worden alle mogelijke interacties tussen de fixed factoren meegenomen in het model en worden er geen interacties tussen random factoren onderling of tussen random en fixed factoren meegenomen.

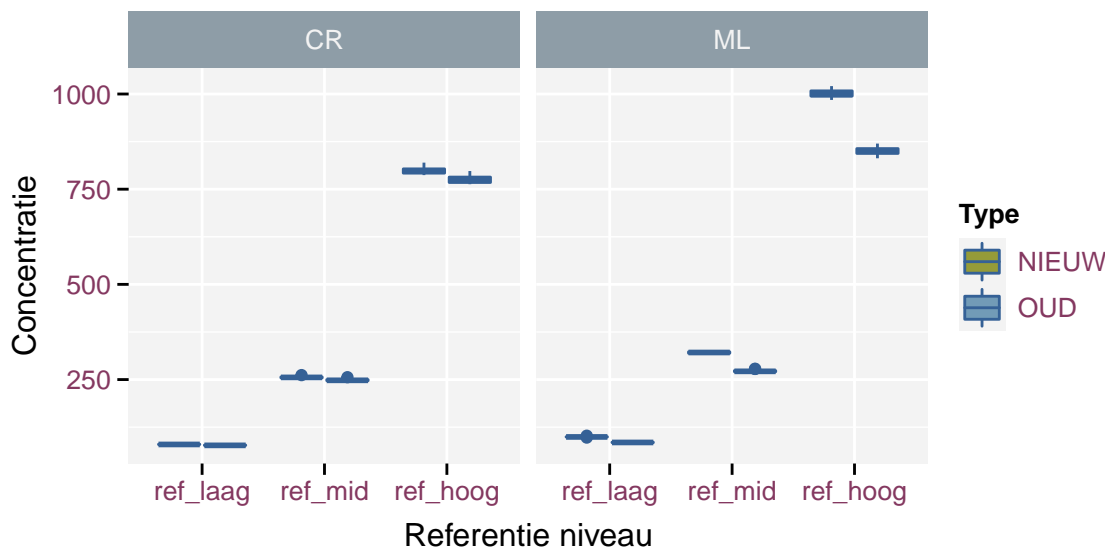
De formule voor het model ziet er tenslotte als volgt uit:

$$Bias_log \sim Ref_cat * Type * Prep + (1|Toestel) + (1|Staal)$$

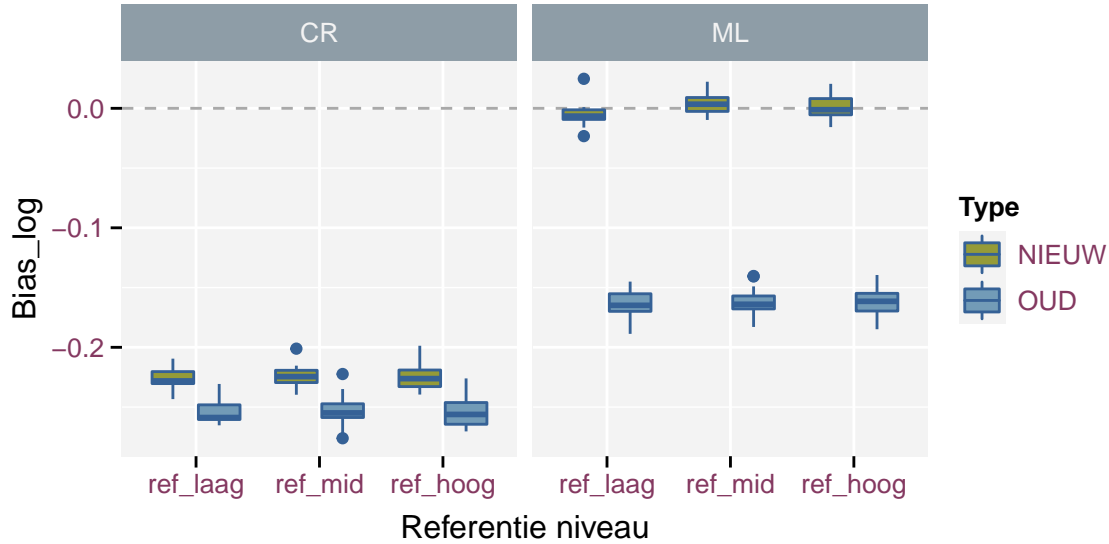
waarbij $(1|x)$ aangeeft dat variabele x wordt meegenomen als random (geneste) factor (categorisch). De analyses werden uitgevoerd aan de hand van de lmer functie (package lme4). Verschillen tussen behandelingen werden beoordeeld op basis F-testen met Kenward-Roger correcties op de vrijheidsgraden (packages car, pbkrtest). Betrouwbaarheidsintervallen rond de voorspelde waarden voor elke combinatie van factoren werden berekend via simulaties aan de hand van de functie bootMer (package lme4).

5 Resultaten en bespreking

Figuren 1 en 2 geven een overzicht van de concentraties en afwijking per referentieniveau, type analyse en preparatie (gepoold over *Toestel* en *Staal*). Bij de afwijkingen is te zien dat deze bijna uitsluitend negatief zijn, wat aangeeft dat de gemeten concentraties een onderschatting zijn van de referentiewaarden.



Figuur 1: Gemeten concentratie voor elk type analyse, preparatie en referentieniveau. Resultaten gepoold over *Toestel* en *Staal*.



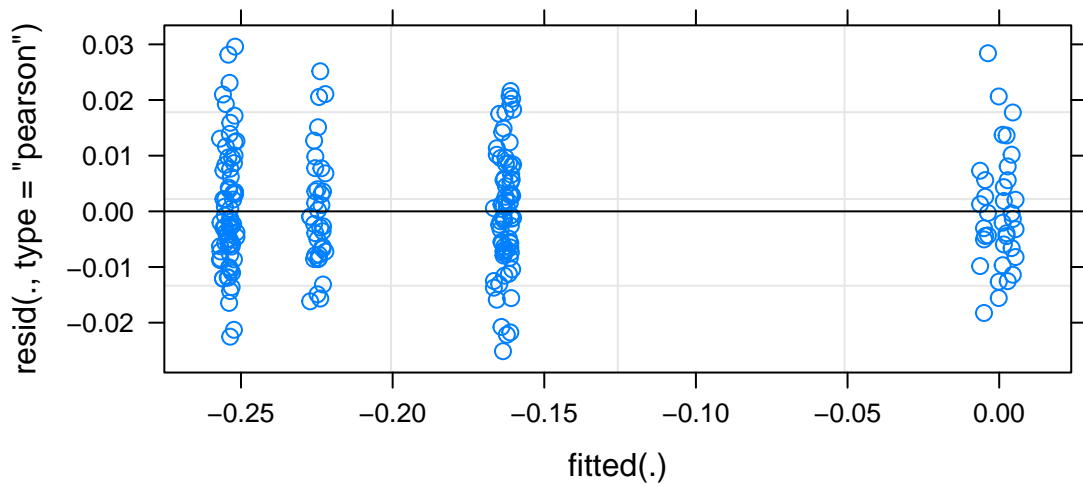
Figuur 2: Afwijking (berekend als log ratio) tussen gemeten concentratie en referentieconcentratie voor elk type analyse, preparatie en referentie niveau. Een bias van nul (grijze stippellijn) geeft aan dat er geen afwijking is tussen de gemeten concentraties en de referentiewaarden.

5.1 Analyse op afwijkingen tussen de referentieconcentratie en de gemeten concentratie.

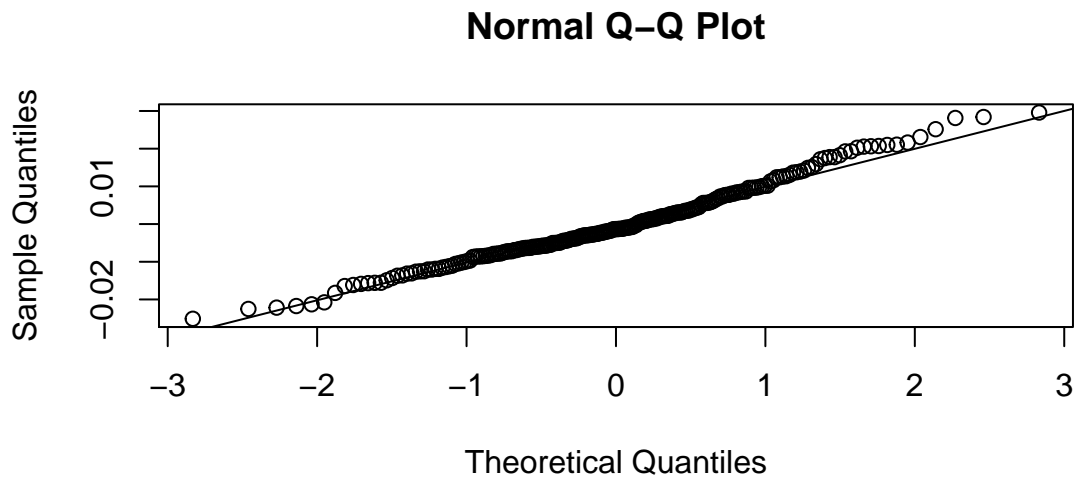
Figuren 3 en 4 tonen aan dat de gegevens voldoen aan de assumpties voor normaal verdeelde gegevens (residuen). Bij de berekening van het model wordt wel een waarschuwing gegeven dat het model singulair is. Dit betekent dat de complexiteit van het model op de rand zit van wat kan worden berekend met de huidige dataset. Meer bepaald is er een complexe afhankelijkheidsstructuur, voornamelijk doordat *Staal* genest is in referentie waarde (*Ref_cat*). Op 216 gemeten waarden zijn er eigenlijk maar 9 onafhankelijke waarnemingen, namelijk de drie stalen voor elk referentieniveau. De singulariteit is dan ook zichtbaar in de geschatte variantie toegekend aan *Staal*, die gelijk is aan nul. Aangezien *Staal* deel uitmaakt van het ontwerp van het experiment kan de random factor *Staal* in principe niet zomaar uit de analyse worden weggelaten. Indien we dit wel doen, zijn de resultaten echter nagenoeg identiek aan het model inclusief *Staal*, maar vervalt de waarschuwing van een singulair model (zie Rmarkdown script).

De anova tabel van de analyse wordt weergegeven in Tabel 1. Uit de analyse blijkt dat het referentieniveau geen effect heeft op de afwijking tussen gemeten concentraties en referentie waarden. Er is wel een significant effect van Type, Prep en hun interactie op de afwijking. In Figuur 5 is te zien dat *Type* = 'OUD' systematisch grotere (negatieve) afwijkingen vertoont dan *Type* = 'NIEUW'. Bovendien vertoont *Prep* methode = 'CR' ook systematisch grotere afwijkingen dan *Prep* methode = 'ML'. De interactie weerspiegelt zich in het feit dat de verschillen tussen *Type* ('OUD' en 'NIEUW') groter zijn voor *Prep* = 'ML' dan voor *Prep* = 'CR'. De beste (= kleinste afwijking) combinatie van type analyse en preparatie, is *Type* = 'NIEUW' met *Prep* = 'ML'. Hier wordt voor geen enkel van de referentieniveaus een afwijking waargenomen tussen de gemeten en de referentie concentratie (0 ligt voor elk van de referentieniveaus binnen het 95% betrouwbaarheidsinterval op de geschatte afwijking; zie Figuur 5).

Merk op dat, hoewel referentieniveau geen effect heeft op de afwijking (berekend als verhouding), het natuurlijk wel een effect heeft op de gemeten concentraties zelf (zie Figuur 1). Bovendien zijn de afwijkingen, indien uitgedrukt als verschil ($Concentratie - Ref_num$), groter (in absolute waarde) voor hogere referentiewaarden (6). Het feit dat dit effect van referentieniveau niet wordt weerspiegeld in de analyse (op de verhouding tussen gemeten concentraties en referentie) wijst erop dat de afwijkingen proportioneel zijn met de grootte van de referentiewaarden.



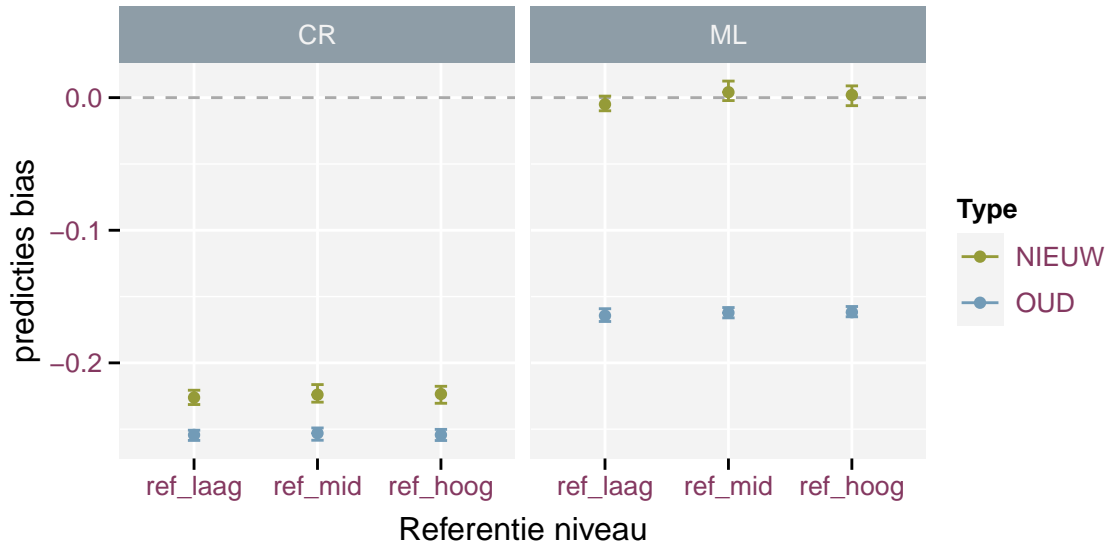
Figuur 3: residuen ten opzichte van voorspelde waarden voor de analyse op afwijking van concentraties



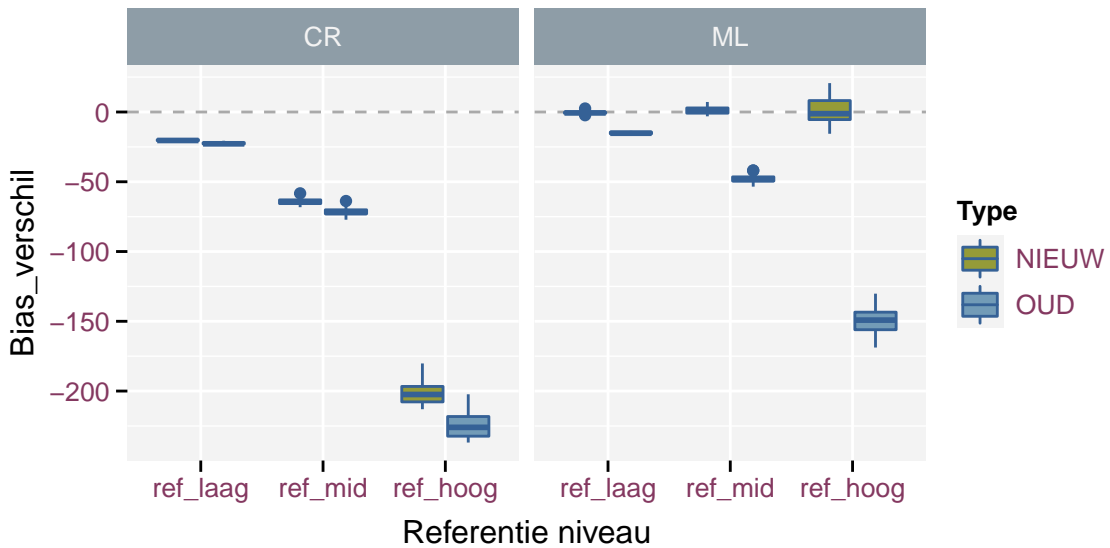
Figuur 4: qqplot voor de analyse op afwijking van concentraties

Tabel 1: anova tabel van de analyse op afwijkingen in gemeten concentraties

	F	Df	Df.res	Pr(>F)
Ref_cat	1.522	2	6	0.292
Prep	8283.359	1	188	0.000
Type	2284.588	1	10	0.000
Ref_cat:Prep	0.463	2	188	0.630
Ref_cat:Type	0.542	2	188	0.583
Prep:Type	1776.267	1	188	0.000
Ref_cat:Prep:Type	0.305	2	188	0.737



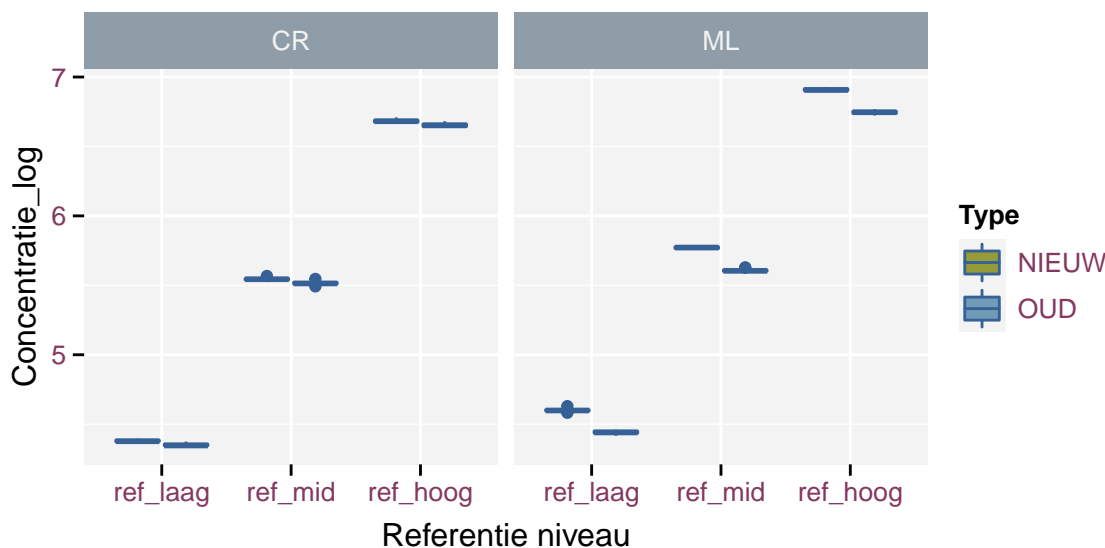
Figuur 5: Voorspelde afwijking (+/- 95% betrouwbaarheidsinterval) per type analyse, preparatie en referentieniveau. Een bias van nul (grijze stippellijn) geeft aan dat er geen afwijking is tussen de gemeten concentraties en de referentiewaarden.



Figuur 6: Afwijking (berekend als verschil: Concentratie - Ref_num) tussen gemeten concentratie en referentieconcentratie voor elk type analyse, preparatie en referentieniveau. Een bias van nul (grijze stippellijn) geeft aan dat er geen afwijking is tussen de gemeten concentraties en de referentiewaarden.

5.2 Korte inspectie van precisie

Zoals aangegeven is de belangrijkste vraagstelling die hier behandeld wordt het achterhalen van afwijkingen van de gemeten concentraties ten opzichte van de referentiewaarden. Toch kan ook kort de **precisie** besproken worden. Op basis van Figuren 1 en 7 kan opgemaakt worden dat de precisie van de metingen over het algemeen groot is. De maximale verschillen tussen individuele metingen en het gemiddelde voor een bepaald type analyse, referentieniveau en preparatie bedraagt slechts 3% van de gemiddelde waarde.



Figuur 7: Gemeten concentratie (log getransformeerd) voor elk type analyse, preparatie en referentieniveau. Resultaten gepoold over Toestel en Staal.

Een snelle test op homogeniteit van de varianties (Levene's test) geeft aan dat er significante verschillen zijn in variantie tussen de verschillende combinaties van *Type*, *Prep* en *Ref_cat* ($F = 3.942$, $p = 0$). Wanneer homogeniteit van de varianties echter getest wordt op log getransformeerde concentraties, vervallen deze verschillen ($F = 0.718$, $p = 0.876$), wat aangeeft dat de precisie proportioneel is met de grootte van de concentraties (*Ref_Cat*), maar dat er voor concentraties binnen dezelfde grootte orde geen verschillen zijn tussen niveaus van *Type* en/of *Prep*. Met andere woorden er zijn geen verschillen waar te nemen in precisie tussen analysetechnieken. Merk hierbij wel op dat bij het testen van de homogeniteit via Levene's test geen rekening is gehouden met de afhankelijkheidsstructuur te wijten aan *Toestel* en *Staal*.

6 Conclusie

Als conclusie kan er gesteld worden dat er significante verschillen bestaan tussen de behandelingen in de grootte van de afwijkingen. De beste resultaten worden bereikt wanneer de nieuwe analysemethode (*Type* = 'NIEUW') gecombineerd wordt met *Prep* = 'ML'. Bij deze combinatie is er geen afwijking waar te nemen van de gemeten concentraties ten opzichte van de referentiewaarden.

7 Aanbevelingen voor verdere analyse

Door het gebruik van dezelfde stalen voor elke combinatie van *Toestel*, *Type* en *Prep* kampt het statistische model met singulariteit. Met andere woorden, er zijn eigenlijk niet genoeg onafhankelijke datapunten in de dataset. Om dit te voorkomen, zouden in het meest ideale geval alle rijen in de dataset onafhankelijk moeten zijn, wat neerkomt op de aanmaak van 216 onafhankelijke stalen. Dit kan gereduceerd worden door slechts twee replicate stalen per combinatie van *Toestel*, *Type* en *Prep* te nemen en het aantal toestellen

van *Type* = ‘OUD’ te reduceren van 8 naar 4. Dan komt men op een totaal van 96 onafhankelijke stalen (3 referentieniveaus x 2 types x 2 preparaties x 4 toestellen per type x 2 replicaten). Dit kan nog verder gereduceerd worden naar 48 onafhankelijke stalen indien dezelfde stalen gebruikt worden voor de twee *prep* methodes (3 referentieniveaus x 2 types x 4 toestellen per type x 2 replicaten). Verdere reductie van het aantal stalen is moeilijk indien men singulariteit wil vermijden en toch voldoende algemene uitpraken wil doen (voldoende toestellen per *Type*; *Noot: eventueel kan het aantal toestellen per type nog verlaagd worden naar 3*).

Los van de verschillen in precisie tussen technieken lijkt de variatie in metingen overal relatief klein. Indien de variatie van de geschatte gemiddeldes toch te groot is (en dus de betrouwbaarheid te klein), kan er altijd geopteerd worden om een poweranalyse te doen waarbij er wordt nagegaan hoeveel replicate metingen er op een staal moeten worden gedaan om een uitspraak te kunnen doen over de concentratie met een vooropgestelde maximale foutenmarge.