



Bedrijfstakonderzoek
BTO 2024.010 | April 2024

Zeer Zorgwekkende Stoffen (deel 1) – clustering, bemonstering en toxiciteit

Rapport



Zeer Zorgwekkende Stoffen (deel 1) – clustering, bemonstering en toxiciteit

BTO 2024.010 | April 2024

Dit onderzoek is onderdeel van het collectieve Bedrijfstakonderzoek van KWR, de waterbedrijven en Vewin.

Opdrachtnummer

402045-233

Projectmanager

P.S. Bäuerlein

Opdrachtgever

BTO - Thematisch onderzoek - Chemische veiligheid

Auteur(s)

B.A. Wols, R.P.J. Hoondert, P.S. Bäuerlein

Kwaliteitsborger(s)

T. ter Laak, M. M. L. Dingemans (H4, Appendix I-III), T. E. Pronk (H2, H4, Appendix I-III)

Verzonden naar

Dit rapport is verspreid onder BTO-participanten.

Een jaar na publicatie is het openbaar.

Keywords

Zeer Zorgwekkende Stoffen, clustering, toxiciteit

Jaar van publicatie
2024

Meer informatie
Dr. Ir. Bas Wols
T +31 30 606 9604
E Bas.Wols@kwrwater.nl

PO Box 1072
3430 BB Nieuwegein
The Netherlands

T +31 (0)30 60 69 511
E info@kwrwater.nl
I www.kwrwater.nl

KWR

April 2024 ©

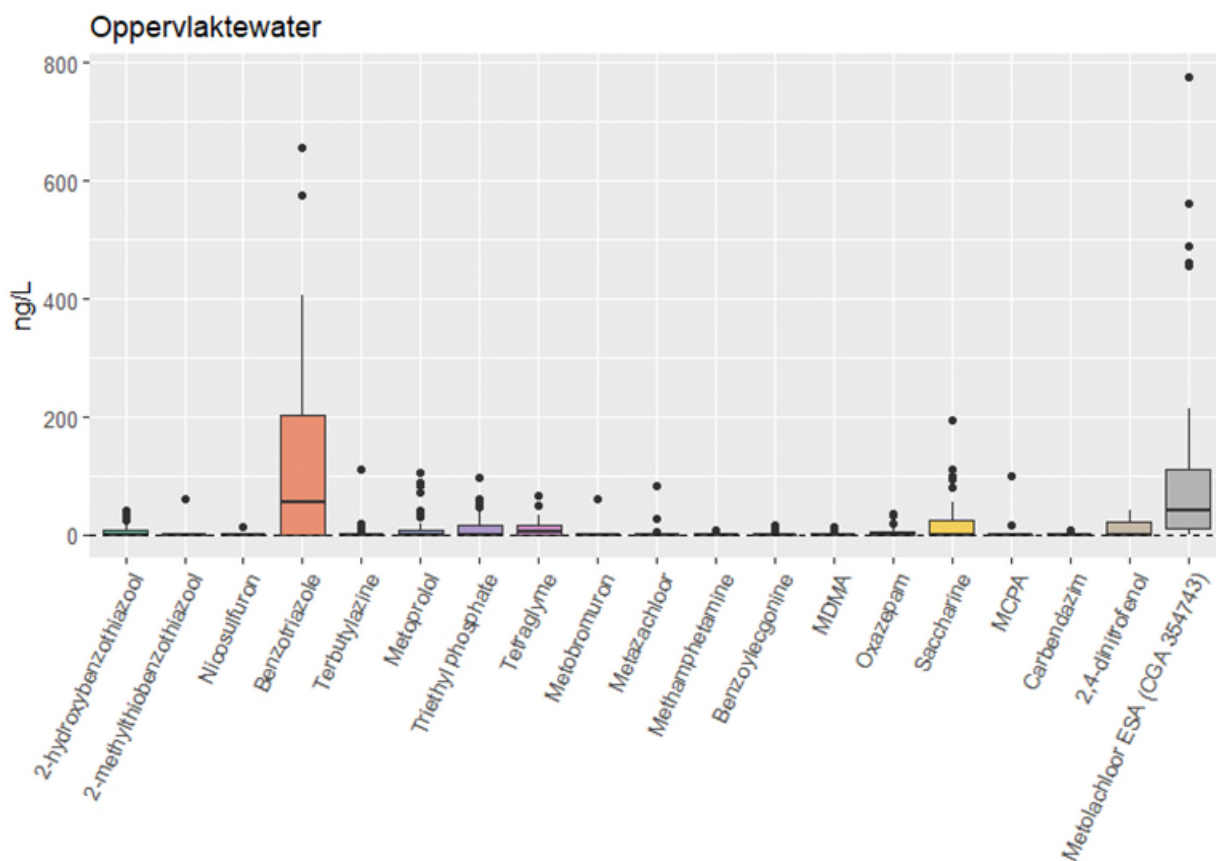
Alle rechten voorbehouden aan KWR. Niets uit deze uitgave mag - zonder voorafgaande schriftelijke toestemming van KWR - worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen, of enig andere manier.

Managementsamenvatting

Meetcampagne op basis van clustering en nieuwe analysemethode toont aanwezigheid Zeer Zorgwekkende Stoffen in drinkwaterbronnen, toxiciteitsmodellen ontwikkeld

Auteur(s) B.A. Wols, R.P.J. Hoondert, P.S. Bäuerlein

Zeer zorgwekkende stoffen (ZZS) vormen een risico voor mens en milieu. Grondige kennis van de chemische eigenschappen, emissieroutes, toxiciteit, gedrag in de ondergrond en verwijderbaarheid tijdens de drinkwaterproductie van deze stoffen is daarom nodig. Om hierin meer inzicht te krijgen, zijn de ZZS geclusterd op basis van chemische eigenschappen. Bovendien is een LC-MS analysemethode ontwikkeld voor 36 van deze ZZS op basis. Uit een meetcampagne waarin deze methode is toegepast, blijkt dat meer dan de helft van deze 36 ZZS voorkomt in drinkwaterbronnen, voornamelijk in oppervlaktewater. Om het risico van ZZS beter te kunnen beoordelen zijn toxiciteitsmodellen ontwikkeld voor ZZS en enkele niet-ZZS stoffen. Deze zijn toegevoegd aan de webtool AquaPriori, waarmee een inschatting gemaakt kan worden van het gedrag van een stof in de ondergrond en in de zuivering.



Gemeten concentraties van een aantal stoffen in oppervlaktewater

Belang: zeer zorgwekkende stoffen

Zeer zorgwekkende stoffen (ZZS) zijn stoffen die gevaarlijk kunnen zijn voor mens en milieu. De watersector krijgt steeds vaker te maken met ZZS. Om adequaat te kunnen reageren, is kennis en informatie noodzakelijk over de chemische eigenschappen van deze stoffen, emissieroutes, (humane) toxiciteit, aanwezigheid in drinkwaterbronnen en verwijderbaarheid tijdens de drinkwaterproductie.

Aanpak: clusteren en meten van zeer zorgwekkende stoffen; toxiciteit modelleren

De clustering van ZZS is uitgevoerd door stoffen te groeperen op basis van chemische eigenschappen en structuur. Verschillende stoffenlijsten, zoals de ZZS en potentiële ZZS-lijst van het RIVM, zijn gebruikt als basis voor deze clustering. Vervolgens werd voor elk cluster een specifieke stof gekozen om als vertegenwoordiger van het cluster op te nemen in de methode voor chemische doelstofanalyse. Op deze manier kan een brede range aan ZZS geanalyseerd worden in een enkele doelstofanalysemethode. Bij de clustering zijn ruim 900 stoffen verdeeld over 51 clusters. Er is uiteindelijk een directe injectie UHPLC-MS/MS analysemethode ontwikkeld voor 36 van die stoffen.

Op basis van toxicologische gegevens uit de ToxCast database zijn voorspellingsmodellen gemaakt voor de toxiciteit van ZZS, zowel aan de hand van functionele groepen als op basis van fysisch-chemische eigenschappen van stoffen zijn Random Forest en multivariate regressiemodellen ontwikkeld.

Resultaten: ruim de helft gemeten ZZS in oppervlaktewaterbronnen; toxiciteitsmodellen

Op diverse locaties bij Nederlandse en Vlaamse drinkwaterbedrijven zijn monsters genomen en geanalyseerd op de geclusterde ZZS stoffen. Uit deze meetcampagne blijkt dat ruim de helft van deze 36 ZZS voorkomt in de bronnen van drinkwater (in oppervlaktewater vaker en meer dan in grondwater). Het grootste gedeelte is niet teruggevonden in drinkwater. Alleen zes verbindingen zijn aangetoond in concentraties nagenoeg altijd onder de 0.1 µg/L.

De ontwikkelde Random Forest en multivariate regressiemodellen worden gebruikt om toxiciteitsvoorspellingen te doen voor de complete lijst van (potentiële) ZZS (900 stoffen) en niet-ZZS (duizenden stoffen).

Toepassing: inschatting gedrag stoffen

De resultaten van dit onderzoek dragen bij aan het begrip van de aanwezigheid, het gedrag van ZZS in de ondergrond, zuivering en toxiciteit. Deze informatie is verwerkt in de webtool AquaPriori. Daarmee kan een gebruiker een inschatting maken van de verwijdering in de zuivering en/of ondergrond, en toxiciteit van een stof.

Rapport

Dit onderzoek is beschreven in het rapport *Zeer Zorgwekkende Stoffen (deel 1) – clustering, bemonstering en toxiciteit* (BTO 2024.010). Lees meer over ZZS in de BTO rapporten:

- Zeer zorgwekkende stoffen (deel 2) – zuivering, BTO 2024.011.
- Zeer zorgwekkende stoffen (deel 3) – literature mining, BTO 2024.012.
- Verslag van veldmetingen en historische meetgegevens over afbraak van organische microverontreinigingen in grondwater, BTO2024.013.

Inhoud

Rapport	2
<i>Managementsamenvatting</i>	3
Inhoud	5
1 Inleiding	7
2 Clusteren van stoffen	9
2.1 Inleiding	9
2.2 Stoffenlijsten	9
2.3 Aanpak clustering	10
2.4 Resultaten	12
2.4.1 Aantal clusters	12
2.4.2 Selectie stoffen in ieder cluster	14
2.5 Definitieve stoffenlijst meetmethode	17
2.6 Herstructurering clusters	19
3 Meetprogramma	20
3.1 Analysemethode	20
3.2 Locaties monsternamen	20
3.3 Resultaten meetprogramma	21
3.3.1 Meetgegevens per watermatrix	22
4 Toxiciteitsvoorspellingen	25
4.1 Inleiding	25
4.2 Methode	26
4.2.1 Verzameling van toxiciteitsdata	26
4.2.2 Combineren van data met functionele groepen en fysisch-chemische descriptoren	28
4.2.3 Modelbeschrijving	29
4.2.4 Modevaluatie	30
4.2.5 Groeperen van <i>in vitro</i> assays	32
4.3 Resultaten en discussie	34
4.3.1 AC ₅₀ -data per cluster	34
4.3.2 Random forest-analyses	34
4.3.3 Lineaire regressieanalyses	36
4.3.4 Sensitiviteit, specificiteit, nauwkeurigheid, NPV en PPV	39
4.3.5 Subsets van <i>in vitro</i> toxiciteitstesten gebaseerd op annotaties	40
4.3.6 Algemene discussie m.b.t. toxiciteit	42
5 Conclusies en aanbevelingen	44

6	Literatuur	46
I	Appendix: Voorspellingsplots voor individuele clusters	49
II	Appendix: Clustering in vitro assayeindpunten	53
III	Appendix: Ontwikkelingen en beleidscontext cumulatie	66
	Concentratie-additie (CA):	66
	Responsadditie (RA) of Independent Action (IA):	67
	Interactie:	67
	Groeperen van stoffen	67
	Water	68
	ABM.	68
	Immissietoets.	68
	Lucht	69
	Cumulatiebenadering in omliggende landen	69
	Kaderrichtlijn Water (KRW)	71
	REACH	71
	Gewasbeschermingsmiddelen (inclusief gewasbeschermingsmiddelen die ZZS bevatten)	71
	Toelating.	71
	Onderzoek oppervlaktewater	71
	Onderzoek Toxische druk	72
IV	Details meetmethode	73

1 Inleiding

Zeer zorgwekkende stoffen (ZZS) zijn stoffen die gevaarlijk voor mens en milieu omdat ze bijvoorbeeld mogelijk kankerverwekkend of zich in de voedselketen ophopen (zeer persistent, zeer bioaccumulatief, persistent, bioaccumulatief en toxisch), zie RIVM (z.d.). Daarmee voldoen ZZS aan de Europese criteria (artikel 57 van REACH) en kunnen ze worden opgenomen in de lijst met autorisatie-plichtige stoffen (bijlage XIV van (European Commission, 2006)). Het overheidsbeleid richt zich op het voorkomen en/of beperken van ZZS in de leefomgeving door aanpak van ZZS emissies bij bedrijven (Ministerie van Infrastructuur en Waterstaat (z.d.)). De sector wordt meer en meer geconfronteerd met deze stoffen en nieuwe emissieroutes (Timmer en Bannink, 2023). Denk hierbij aan PFAS, drugsafval, medicijnresten, etc. Dit vraagt om gedegen kennis en informatie over het voorkomen en het gedrag van stoffen, emissieroutes, impact op de bronnen en verwijderbaarheid gedurende de productie van drinkwater.

In de RIVM-lijst zijn ca. 1600 ZZS en 300 potentiële ZZS (pZZS) opgenomen. Dit zijn stoffen die via het milieu (lucht, water of bodem), voedsel, werkplek of producten (zoals huishoudchemicaliën) in contact kunnen komen met mensen en ecosystemen (RIVM, z.d.). Deze verbindingen zijn in de EU verplichte gevaarsindeling, etikettering en verpakking (CLP-classificatie) van stoffen gelabeld als kankerverwekkend (mutageen, toxisch voor de reproductie klasse 1A/1B). Een belangrijke vraag hierbij is welke ZZS relevant zijn voor drinkwaterproductie. Daarvoor zijn de volgende aspecten van belang: in welke mate komen deze stoffen voor in de waterbronnen, wat is het gedrag van deze stoffen in de ondergrond en zuivering en wat is de aard van de toxiciteit van deze stoffen?

De molecuulstructuur bepaalt de eigenschappen van stoffen. Een eerste stap in de omgang met het grote aantal ZZS is het toepassen van clustering van de verschillende chemicaliën. Door het clusteren van stoffen uit de RIVM-lijst, inclusief (p)ZZS en referentiestoffen, kan worden geanalyseerd hoe deze chemisch gerelateerd zijn. Dit houdt in dat we onderzoeken of ze gelijkaardig zijn en hoe hun verhouding is tot de referentiestoffen. Daardoor kan ook meer inzicht worden gekregen in welke eigenschappen van een stof een (p) ZZS maken. (p)ZZS en referentiestoffen kunnen op basis van eigenschappen van het molecuul geclusterd worden. Uit ieder cluster kan dan vervolgens een stof worden gekozen die als ZZS-modelstoffen kunnen dienen in een meetcampagne. Het RIVM heeft een zoekstelsel voor ZZS¹, waarbij een willekeurige stof ingevoerd kan worden en waar vervolgens op basis van molecuulstructuur de meest gelijkende ZZS gevonden kan worden. Dit kan als een selectiemethode gebruikt worden om te kijken of een bepaalde stof lijkt op een ZZS en daarmee mogelijk vergelijkbare toxiciteitsprofielen heeft. In de clustering van ZZS beschreven in dit rapport wordt gebruik gemaakt van vergelijkbare methoden als het RIVM zoekstelsel om molecuulstructuren te beschrijven (fingerprints) en om deze molecuulstructuren met elkaar te vergelijken (similarity index). Dit is beschreven in Hoofdstuk 2. De aanpak die in dit rapport beschreven is, richt zich op individuele ZZS (ook binnen de clusters). Een alternatieve aanpak die wel onderzocht wordt, maar niet in dit onderzoek, is om te kijken naar mengsels. In Bijlage III is in meer detail beschreven hoe de aanpak van mengsels in een beleidscontext wordt onderzocht door het RIVM.

Het voorkomen van de stoffen in de waterbronnen en het gedrag in de ondergrond en zuivering kan onderzocht worden aan de hand van meetcampagnes. Diverse oppervlaktewateren worden gemonitord (rijkswateren door RWS, meetpunten drinkwaterbedrijven door RIWA Osté et al., 2022) op allerlei stoffen. Zo zijn tientallen ZZS gevonden met relevante exposure en risk-scores in de meetcampagnes van RWS, RIWA-Rijn en FOTO-NL (Osté et al., 2022), zoals bijvoorbeeld 1,4-dioxane en vinylchloride. Om een beeld te krijgen van het voorkomen van een brede range aan ZZS is een meetcampagne uitgevoerd op basis van de clustering. Hierin zijn ZZS gemeten in de bronnen (zowel

¹ <https://rvszoekstelsel.rivm.nl/ZzsSimilarityTool>

oppervlaktewater als grondwater), in de ondergrond (tijdens bodempassage), en tijdens en na de zuivering. Dit is beschreven in Hoofdstuk 3.

Dit rapport is uitgevoerd in het kader van het BTO project 'Zeer Zorgwekkende Stoffen'. Dit is een overkoepelend thematisch BTO project van de thema's Bronnen & Omgeving, Chemische Veiligheid, Zuivering en Hydroinformatics. Er verschijnen een aantal rapporten over onderzoek naar Zeer Zorgwekkende Stoffen. Dit rapport is het eerste rapport in dit ZZS-project en beschrijft het gedeelte van het project over de clustering van de stoffen, meetcampagne en toxiciteit. De waterzuiveringsaspecten van ZZS worden beschreven in BTO 2024.011. Daarnaast wordt een onderdeel over literatuur mining beschreven in BTO 2024.012. En er verschijnt een deelrapport over het gedrag van ZZS in de ondergrond (BTO 2024.013).

2 Clusteren van stoffen

2.1 Inleiding

De zeer zorgwekkende stoffen (ZZS)-lijst omvat een groot aantal stoffen en stofgroepen (>1500). Deze lijst is te lang en de stoffen zijn chemisch/fysisch te verschillend om allemaal in een doelstofanalysemethode op te nemen. De ZZS-lijst is samengesteld voor een breed publiek, met name voor mens en milieu die via verschillende wegen in aanraking komen met deze stoffen, zoals door het milieu (lucht, water en bodem), voedsel, de werkomgeving, en huishoudchemicaliën. Dus niet alle ZZS zullen in het water voorkomen. Naast de lijst met ZZS zijn er ook stoffenlijsten gebruikt die wel specifiek van belang zijn voor de watersector, zoals bijvoorbeeld de persistente, mobiele en toxische (PMT) stoffen. Vanwege het grote aantal relevante stoffen zijn clustertechnieken toegepast om stoffen die op elkaar lijken te groeperen. Op deze manier kunnen groepen van stoffen gedefinieerd worden waarvan vergelijkbaar gedrag valt te verwachten in ondergrond, zuivering of qua toxiciteit. Voor elk cluster is vervolgens één specifieke stof gekozen om te gebruiken in een doelstofanalysemethode. Dit is gedaan om een meetprogramma te ontwikkelen dat zo representatief mogelijk is voor de gehele lijst van stoffen.

2.2 Stoffenlijsten

De lijst van in dit BTO onderzoek onderzochte ZZS stoffen is gebaseerd op een bestaande lijst van het RIVM (RIVM, z.d.,²). Stoffen op deze lijst worden als zeer zorgwekkend aangemerkt als ze voldoen aan één of meer van onderstaande criteria: Kankerverwekkend (C), Mutageen (M), Reprotoxisch (R), PBT, vPvB, soortgelijke zorg (bijvoorbeeld hormoonverstorende stoffen).

Naast de ZZS lijst van het RIVM zijn een aantal andere stoffenlijsten gebruikt:

- Lijst met potentieel Zeer Zorgwekkende Stoffen (pZZS) van RIVM, te vinden op de website van het RIVM (RIVM, z.d.,³);
- Lijst met persistente, mobiele en toxische (PMT) volgens NGI en UBA (Deens onderzoek), Holmberg et al. (2021);
- Lijst met persistente, mobiele en toxische (PMT) gebruikt door Vitens (met gemeten concentraties) (persoonlijke communicatie);
- Lijst met stoffen gedetecteerd tijdens screening van waterbronnen door de HPLC-UV (intern rapport);
- Lijst met Gadolinium (Gd) gebaseerde röntgen (MRI) contrastmiddelen (ref). Dit is een groep stoffen waar weinig over bekend is en die steeds meer worden toegepast (Kools et al., 2013; RIWA-Rijn, 2022).

Op deze stoffenlijsten zijn een aantal bewerkingen uitgevoerd om tot een algehele lijst met stoffen te komen:

- Stofgroepen zijn verwijderd;

² <https://rvs.rivm.nl/onderwerpen/zeer-zorgwekkende-stoffen>

³ <https://rvs.rivm.nl/onderwerpen/Zeet-Zorgwekkende-Stoffen/Potentiele-ZZS>

- De stoffen die niet in werelds grootste database van stoffen PubChem⁴ voorkomen worden verwijderd omdat zonder deze database chemische descriptoren en/of fingerprints niet kunnen worden bepaald (zie Tabel 1);
- De dubbelingen uit verschillende lijsten worden verwijderd.
- Anorganische verbindingen zijn verwijderd (van de verbindingen die in PubChem staan zijn dit 135 anorganische verbindingen in de ZZS-lijst en 15 anorganische verbindingen in de pZZS lijst);
- Stoffen met een logP groter dan 8 zijn verwijderd, omdat deze zeer slecht oplosbaar zijn in water en dus niet relevant zijn voor waterbronnen. Deze grens is vrij ruim genomen om niet per ongeluk (bijvoorbeeld door onnauwkeurigheden in logP waarden) waterrelevante stoffen te verwijderen.

Tabel 1 geeft een overzicht van het aantal stoffen per lijst en hoeveel stoffen er overgebleven zijn na de diverse bewerkingen. In totaal zijn er 1079 stoffen.

Tabel 1 Overzicht van de verschillende lijsten en aantal stoffen na diverse bewerkingen

	ZZS	pZZS	PMT (UBA/NGI)	Gd-MRI contrast	HPLC_UV	Vitens PMT
# stoffen/stofgroepen	1566	261	157	9	153	121
# stoffen in pubchem	795	217	156	9	149	121
# stoffen na verwijdering dubbelingen	741	216	156	9	149	121
# stoffen na verwijdering anorganische stoffen	611	200	156	9	149	120
# stoffen (cumulatief)	611	811	932	941	1074	1135
# stoffen na verwijdering log P>8						1079

2.3 Aanpak clustering

Voor de analyse worden de volgende stappen uitgevoerd:

- Identificatie van stof op basis van CAS/stofnaam. Nadat de stof is opgezocht in PubChem wordt de isomerische smiles codering (isomeric_smiles) uit PubChem gehaald;
- Bepalen van fingerprints, dit is een code die een stof krijgt waarbij iedere bit in de code aangeeft of een bepaalde substructuur wel of niet aanwezig is in het molecuul. De volgende fingerprints zijn beschouwd (het aantal bits geeft het aantal substructuren aan):
 - CACTVS fingerprint uit pubchem (881 bits, opgehaald met behulp van pubchempy⁵);
 - MACCS fingerprint (167 bits, met behulp van rdkit⁶);
 - Morgan fingerprint (2048 bits, met behulp van rdkit), deze is uiteindelijk niet meegenomen, omdat die niet onderscheidend genoeg was;
- Kiezen van descriptoren, dit zijn numerieke waarden van een bepaalde eigenschap van het molecuul die op basis van de structuurformule eenvoudig af te leiden zijn met chemische informaticatechnieken. De volgende descriptoren zijn beschouwd:
 - Aan de hand van de fingerprint: iedere bit wordt als aparte descriptor beschouwd in de data-analyse. Zowel de CACTUS als de MACCS fingerprints zijn gebruikt;
 - Uit pubchem: charge, tpsa, rotatable_bond_count, molecular_weight, heavy_atom_count, h_bond_acceptor_count, h_bond_donor_count;

⁴ <https://pubchem.ncbi.nlm.nih.gov/>

⁵ <https://pubchempy.readthedocs.io/en/latest/api.html>

⁶ <https://www.rdkit.org/docs/GettingStartedInPython.html>

- Via rdkit: xlogp. Via rdkit kunnen er ook nog veel meer descriptoren berekend worden, deze zijn echter niet voor alle stoffen bekend en daarom niet gebruikt;
- Verwijderen van missende data. Voordat de clustering gedaan kan worden moet eerst alle ontbrekende descriptordata worden verwijderd. Het selecteren van descriptoren, welke numerieke waarden van specifieke moleculaire eigenschappen vertegenwoordigen, wordt vereenvoudigd door chemische informaticatechnieken. Deze waarden kunnen direct worden afgeleid uit de structuurformule van het molecuul;
- Clustering, dit is gebeurd via K-means clustering (met de python module scikit-learn) op basis van bovengenoemde descriptoren. Hier kan van te voren ingesteld worden hoeveel clusters er gemaakt moeten worden. Het clusteringsalgoritme is toegepast voor verschillende aantallen clusters, variërend van 2 t/m 150 clusters, om het aantal clusters voor deze stoffenlijsten te optimaliseren;
- Evaluatie van clustering om te bepalen hoe goed de stoffen passen binnen een cluster en hoeveel clusters er nodig zijn om de lijsten te representeren. Hierbij is van twee parameters gebruik gemaakt:
 - Silhouette score is een maat hoe goed een stof past binnen een cluster, tussen de -1 en 1. Groter dan 0 betekent dat de stof beter bij zijn eigen cluster past dan bij andere clusters, kleiner dan 0 betekent het omgekeerde.
 - Similarity index. Hier worden de (MACCS) fingerprints van twee moleculen vergeleken en wordt een similarity index berekend tussen 0 en 1 (0 ongelijk en 1 gelijk). Er wordt gebruik gemaakt van de Tanimoto similarity index (Bajusz et al., 2015). In een cluster wordt voor elke stof berekend wat de gemiddelde similarity is t.o.v. de andere stoffen in het cluster. RIVM maakt in de ZZS tool (Wassenaar et al., 2022) ook gebruik van similarity indices om stoffen te vergelijken.

Nadat de clusters gemaakt zijn, wordt aan de hand van de evaluatie het optimale aantal clusters gekozen. Vervolgens zijn zeer vluchtige stoffen (met een kookpunt < 20 °C) en zeer slecht oplosbare stoffen (met een oplosbaarheid < 0,1 mg/L) verwijderd, omdat deze niet in drinkwaterbronnen worden verwacht⁷. Tevens zijn clusters met heel weinig stoffen (1 of 2) verwijderd om het aantal stoffen in de doelstofanalyse methode te verminderen. Voor deze stoffen is tevens onderzocht hoe goed ze in een ander cluster passen (Paragraaf 2.6).

Vervolgens wordt er per cluster een stof geselecteerd voor in de doelstofanalyse methode. Dit gebeurt door de stoffen in ieder cluster te rangschikken op (in volgorde):

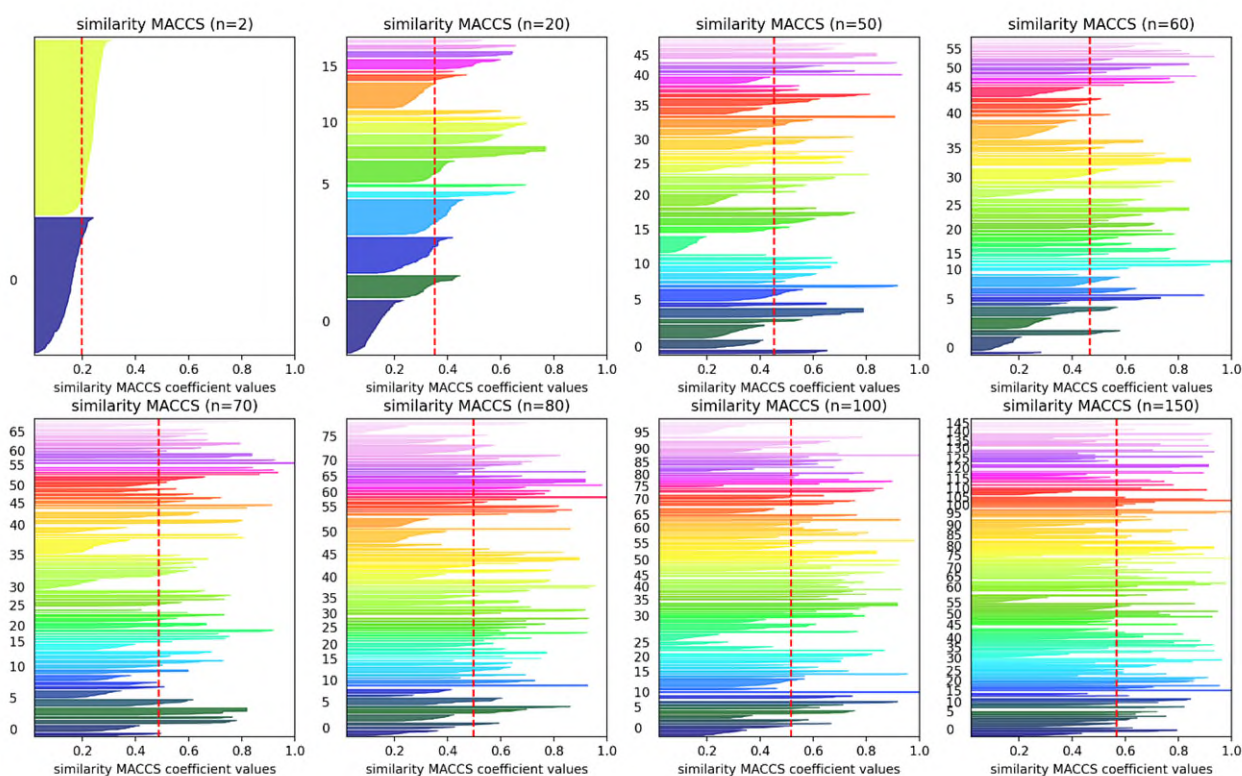
1. Beschikbare analysetechnieken
2. Reeds gemeten concentraties bij de Rijn en Maas (RIWA), HWL en Vitens (PMT lijst)
3. Afkomst van de stof uit een specifieke PMT lijst heeft hier de hoogste prioriteit, omdat die het meest relevant zijn voor drinkwaterbedrijven.
4. Eigenschappen: vluchtigheid (modelberekening kookpunt in EPI suite, USEPA, 2012) en oplosbaarheid (modelberekening EPI suite, USEPA, 2012). Hoogste rangschikking gebeurt voor hogere oplosbaarheid en lage vluchtigheid (hoog kookpunt).
5. Verwacht gebruik (Reach database – tonnages per jaar, beschikbaar voor 1/3 van de stoffen)

⁷ Hierbij is geen rekening gehouden met ionisatie van stoffen (bijvoorbeeld stoffen die als neutraal molecuul slecht oplosbaar zijn maar na ionisatie wel oplossen)

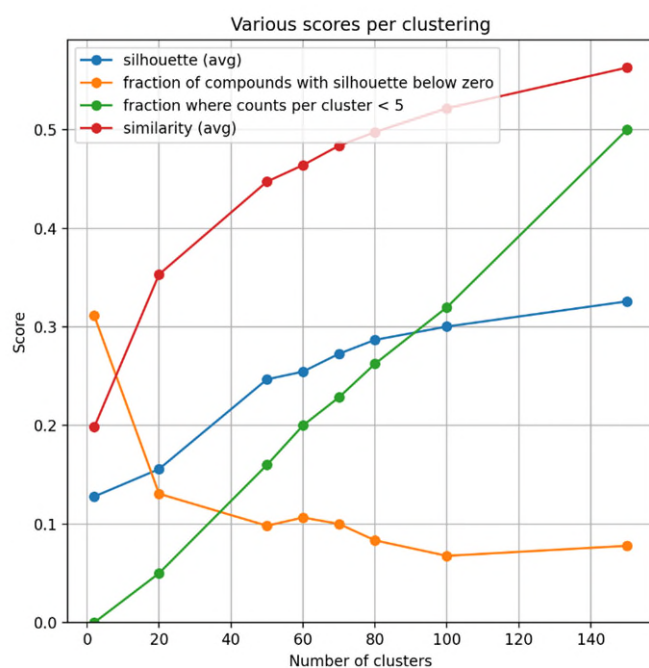
2.4 Resultaten

2.4.1 Aantal clusters

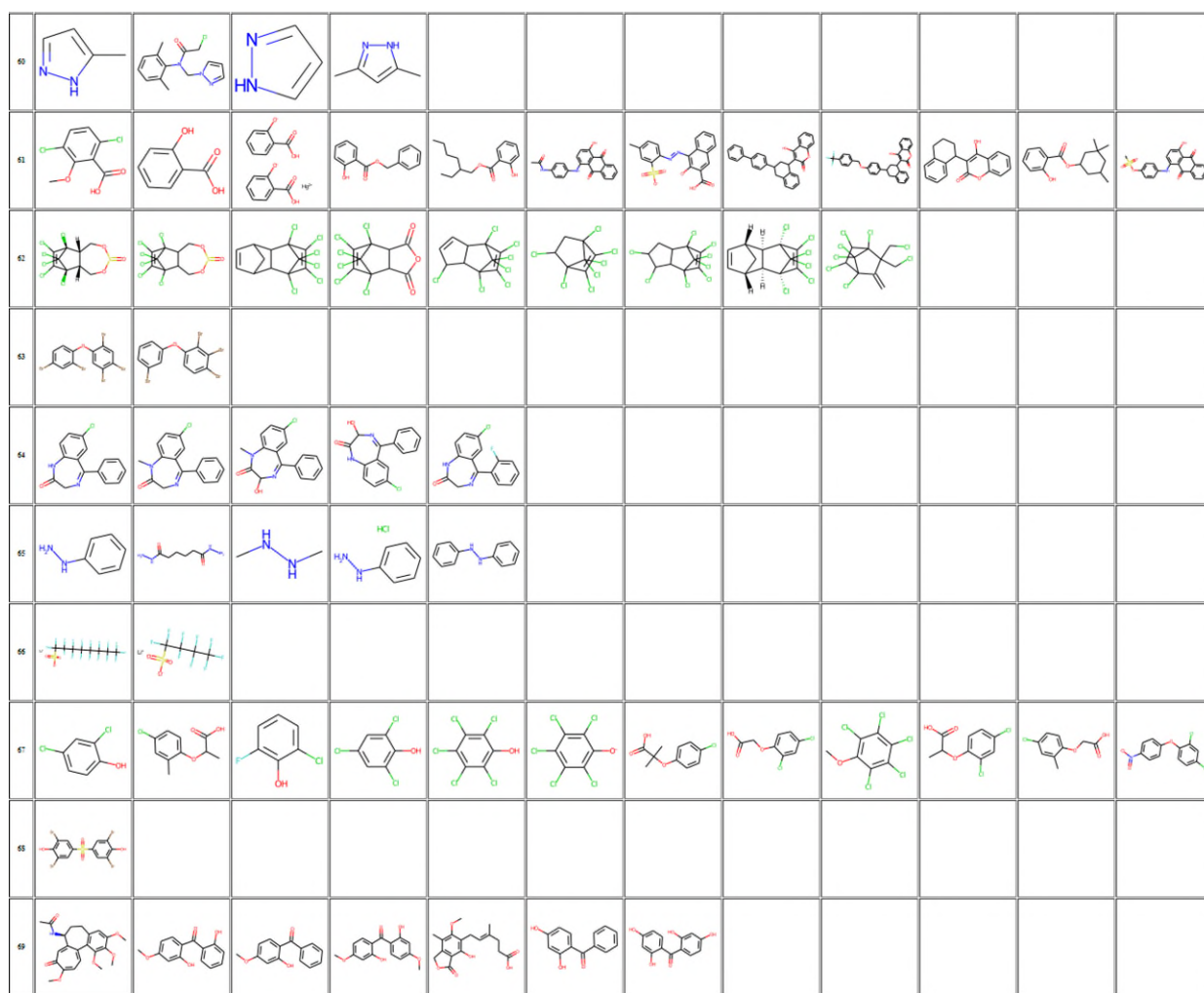
De 'similarity' scores van iedere stof in een cluster is weergegeven in Figuur 2-1 voor verschillende aantallen clusters. In Figuur 2-2 is de gemiddelde 'similarity' en 'silhouette' score als functie van het aantal clusters weergegeven. Deze scores lopen op bij een toenemend aantal clusters. Voor de silhouette score is tevens het aantal clusters weergegeven waarin stoffen zitten met een score lager dan 0 (dit betekent dat deze stoffen niet goed binnen dat cluster passen). Vanaf ongeveer 50 clusters blijft dit aandeel nagenoeg constant (bij een toenemend aantal clusters). Daarnaast neemt het aantal clusters met heel weinig stoffen erin (minder dan 5 stoffen) logischerwijs toe bij een toenemend aantal clusters. Op basis van deze vergelijking van het aantal clusters is gekozen voor 70 clusters, omdat bij een groter aantal clusters de toename van similarity en silhouette relatief klein is.



Figuur 2-1 Evaluatie van de 'similarity' scores van iedere stof in een clusters voor diverse aantal clusters (de subplots geven van links naar rechts een oplopend aantal clusters weer). Stoffen in hetzelfde cluster hebben dezelfde kleur.



Figuur 2-2 Gemiddelde evaluatie score van de stoffen in een cluster bij verschillende aantallen clusters



Figuur 2-3 Voorbeeld van stoffen in cluster 60 t/m 69. Clusters 63, 66 en 68 zijn later verwijderd omdat ze te klein waren.

2.4.2 Selectie stoffen in ieder cluster

Uit elk cluster dient één stof gekozen te worden voor toepassing in een analysemethode. Voordat dit gedaan wordt, zijn een aantal nabewerkingen uitgevoerd om het aantal clusters te verkleinen (zie Paragraaf 2.3). Het verwijderen van hele vluchtige stoffen reduceert het totaal aantal stoffen tot 919 in 69 clusters. Daarna resulteert het verwijderen van heel slecht oplosbare stoffen het aantal stoffen tot 726 in 63 clusters. Tenslotte zijn clusters met 2 stoffen of minder niet meegenomen voor de analysemethode, wat resulteert in 710 stoffen in 51 clusters (de afgevalen stoffen zijn later toegevoegd aan andere clusters, zie Paragraaf 2.6).

De stoffen in de 51 overgebleven clusters zijn gerangschikt volgens de criteria gegeven in Paragraaf 2.3. De geaggregeerde waarde per cluster van een aantal van deze criteria zijn weergegeven in Tabel 2. De hoogste geprioriteerde stof (op basis van de criteria in Paragraaf 2.3) per cluster is weergegeven in

Tabel 3. Hierbij is door het KWR laboratorium een inschatting gemaakt in hoeverre deze stoffen te analyseren zijn met een methode. Veel stoffen zijn te analyseren met behulp van een C18-kolom, daarom is gekozen om een methode met juist deze type kolom te ontwikkelen.

Tabel 2 Overzicht van 51 clusters

Cluster	# stoffen	# stoffen in NL aangetroffen /gemeten	Max. Reach gebruik (ton)	Gem. Oplosbaarheid (mg/L)	Kookpunt (K)	Gemiddelde concentratie NORMAN (µ/L)	Gemiddelde concentratie in NL (µg/L)	# stoffen in cluster met beschikbare KWR methoden
0	10	1	10k-100k	1.14E+05	479		0.109	0
1	11	5		6.36E+03	428	0.123	0.113	3
2	29	12	1M-10M	1.27E+05	143		0.287	0
3	8	3		8.96E+01	245	0.008	0.010	3
4	12	0	1k-10k	8.13E+02	291			0
5	11	7		4.55E+00	370	0.032	0.016	0
6	6	1		2.09E+03	503	0.116	0.116	0
8	6	4		3.20E+03	451	0.008	0.007	0
9	47	23		1.38E+04	392	0.083	0.061	16
10	4	2		4.53E+02	420	0.155	0.089	2
11	3	1		4.18E+03	584	0.084	0.043	1
12	4	3		3.63E+03	526	0.118	0.095	2
13	10	2	1k-10k	3.26E+03	174			0
14	5	5		3.02E+03	313	0.215	0.149	5
16	5	4		5.39E+02	368	0.066	0.024	2
17	13	11		7.80E+04	332	0.034	0.188	7
18	3	3		1.63E-01	340		0.000	0
19	4	3		5.50E+01	847	0.302	0.199	3
20	51	10		7.96E+03	337	0.121	0.080	5
21	27	2		1.56E+04	353			0
24	6	1		1.26E+05	273	0.455	0.157	1
25	11	5	0.1-1k	1.83E+05	364	0.052	0.050	3
26	21	3		2.21E+02	358	0.068	0.068	0
28	11	6		2.76E+01	447	0.129	0.071	2
29	18	7		2.21E+03	369	0.078	0.030	4
30	21	4	0.1-1k	1.26E+05	239	0.003	0.019	3
34	5	0		5.35E+00	443			0
35	9	5		1.09E+04	159	0.010	0.151	3
37	40	18		2.40E+04	250	0.114	0.070	5
39	3	0		1.12E+05	249			0
40	25	9		1.20E+04	348	1.028	0.839	3
41	13	3	1-10	6.32E+05	307	0.030	1.272	3
42	3	1		5.43E+01	494			0
43	8	0		3.34E+04	357			0
44	11	8		6.69E+02	358	0.706	0.319	2
46	39	6	1k-10k	3.66E+05	195	0.197	0.416	2
47	3	0	10k-100k	1.08E+04	64			0

48	44	9	1M-10M	1.77E+05	165	1.079	0.746	0
51	14	12		2.68E+02	359	0.124	0.038	8
52	8	5		4.08E+00	305		0.004	0
55	61	9	10k-100k	3.84E+05	287	0.058	0.196	4
56	14	4		7.82E+04	410	0.066	0.110	1
57	8	7		4.43E+03	347	0.039	0.024	5
59	3	0		1.00E+06	480			0
60	4	2		3.44E+04	250	0.123	0.505	1
61	7	2		6.12E+02	400	0.036	0.022	0
62	4	1		9.94E-01	361		0.000	0
64	5	4		1.06E+02	461	0.011	0.008	4
65	4	0		1.83E+05	251			0
67	11	9		1.09E+04	300	0.083	0.047	6
69	7	0		1.75E+02	439			0

Tabel 3 Geselecteerde stoffen per cluster en mogelijkheden met betrekking tot analysemethoden (C18, Vluchtigen of GC-MS).

cluster	Stofnaam (hoogst geprioriteerd per cluster)	C18	VLU	GC-MS
0	trichloro(fluoro)methane		x	
1	Furosemide	x		
2	chloroform		x	
3	1,1,2,2,3,3,4,4,4-nonafluorobutane-1-sulfonic acid	x		
4	cycloheximide	x		
5	phenanthrene			x
6	flumioxazin	x		
8	Climbazole	x		
9	2-hydroxybenzothiazool	x		
10	2-methylbenzothiazool	x		
11	Nicosulfuron	x		
12	pentoxifylline	x		
13	1,2-dibroomethaan		x	
14	benzotriazole	x		
16	2,4-dinitrofenol	x		
17	terbutylazine	x		
18	endrin			x
19	iopamidol	x		
20	metoprolol	x		
21	2,4,5-trimethylaniline	x		
24	carbendazim	x		
25	triethyl phosphate	x		
26	parathion-methyl	x		
28	Prothioconazool	x		
29	thiacloprid	x		
30	phoxim	x		
34	2-(benzotriazol-2-yl)-4-methylphenol (Drometrizole)	x		
35	pentadecafluorooctanoic acid	x		

37	methamphetamine	x		
39	N-[bis[(butan-2-ylideneamino)oxy]-methylsilyl]oxybutan-2-imine			
40	methadon	x		
41	CGA 354743	x		
42	warfarin	x		
43	3-(oxiran-2-ylmethoxy)-N,N-bis(oxiran-2-ylmethyl)aniline			
44	benzoylecgonine (metabolië cocaine)	x		
46	tetraglyme	x		
47	dimethoxy(dimethyl)silane			
48	tetrahydrofuran		x	
51	metobromuron	x		
52	1,2,4-trichlorobenzene		x	
55	bromacil	x		
56	saccharine	x		
57	MDMA	x		
59	[bis[6-[bis(phosphonomethyl)amino]hexyl]amino]methylphosphonicacid			
60	metazachloor	x		
61	Salicylzuur	x		
62	beta-endosulfan			
64	oxazepam	x		
65	hexanedihydrazide			
67	4-chloor-2-methylfenoxiazijnzuur (MCPA)	x		
69	oxybenzone	x		

2.5 Definitieve stoffenlijst meetmethode

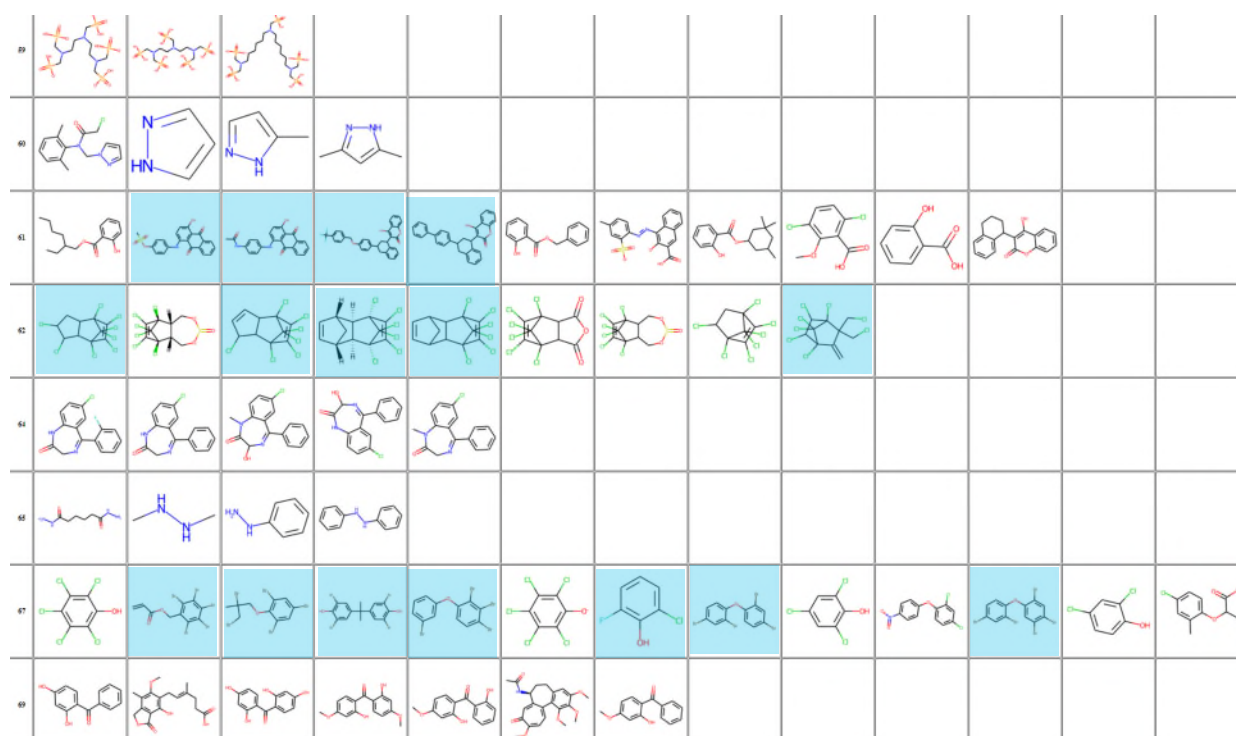
De stoffen en bijbehorende rapportagegrenzen voor de ontwikkelde methode zijn weergegeven in Tabel 4. De stoffen prothioconazool en iopamidol zijn afgevalen, omdat deze niet teruggevonden werden in de ontwikkelde analysemethode. Voor deze methode worden de volgende twee kolommen gebruikt: een analytische kolom (Phenomenex Luna Omega Polar C18 2.1x100 1,6 µm) en een Guardkolom (Phenomenex securityGuard ultra cartridges C18). De stoffen zijn gedetecteerd met een SCIEX Triple Quad 6500+. Details van de methode zijn te vinden in bijlage IV.

Tabel 4: Definitieve lijst met stoffen in de C18 methode en bij behorende rapportagegrens (ng/L) in vier verschillende matrices. DW = Drinkwater, OW = Oppervlaktewater, GW = Grondwater. Vanwege matrixeffecten of blanco achtergrond kunnen voor de roodgekleurde verbindingen geen prestatiekenmerken worden vastgesteld.

Component	Cas nummer	Rapportagegrens (ng/L)			
		Ultrapuur water	DW	OW	GW
Furosemide	54-31-9	25	25	25	25
1,1,2,2,3,3,4,4,4-nonafluorobutane-1-sulfonic acid	375-73-5	50	50	50	100
cycloheximide	66-81-9	25	50	50	50
Climbazole	38083-17-9	5	5	5	5
2-hydroxybenzothiazool	934-34-9	5	10	10	10
2-methylthiobenzothiazool	615-22-5	50	50	50	100
Nicosulfuron	111991-09-4	5	5	5	5
Pentoxifylline	6493-05-6	5	5	5	5
Benzotriazole	95-14-7	10	25	25	25
2,4-dinitrofenol	51-28-5	10	10	10	10
Terbutylazine	5915-41-3	5	5	5	5
Metoprolol	37350-58-6	5	5	5	5
2,4,5-trimethylaniline	137-17-7	5	5	5	5
Carbendazim	10605-21-7	5	5	5	5
Triethyl phosphate	78-40-0	10	50	50	50
Thiacloprid	111988-49-9	5	5	5	5
Phoxim	14816-18-3	5	10	10	10
Drometrizole	2440-22-4	50	50	50	50
Metolachloor ESA (CGA 354743)	171118-09-5	25	25	25	25
Warfarin	81-81-2	5	5	5	5
Tetraglyme	143-24-8	5	5	5	5
Metobromuron	3060-89-7	10	25	25	25
Bromacil	314-40-9	10	10	10	10
Saccharine	81-07-2	50	50	50	50
Metazachlor	67129-08-2	5	5	5	5
4-chloor-2-methylfenoxyazijnzuur (MCPA)	94-74-6	25	25	25	25
Oxybenzone	604-75-1	10	10	10	10
Methamphetamine	537-46-2	5	5	5	5
Benzoylcegonine	519-09-5	5	5	5	5
MDMA	42542-10-9	5	5	5	5
Oxazepam	604-75-1	5	5	5	5
Methadon	76-99-3	5	5	5	5
Flumioxazin	103361-09-7				
Parathion-methyl	298-0-0				
PFOA	335-67-1				
Salicylzuur	69-72-7				

2.6 Herstructurering clusters

Een aantal stoffen en clusters zijn om diverse redenen afgefallen. Een aanvullende bewerking is daarom uitgevoerd om de afgefallen stoffen of de stoffen in afgefallen clusters mee te nemen in de overgebleven clusters. Dit is gedaan door voor iedere stof in een afgefallen cluster te berekenen wat de 'similarity' is met de stoffen in de overgebleven clusters, en vervolgens is deze stof toegekend aan het cluster waarvan de gemiddelde 'similarity' score het hoogst is. Figuur 2-4 laat een voorbeeld zien van een aantal clusters waaraan stoffen op deze manier zijn toegevoegd.



Figuur 2-4 Voorbeeld van stoffen (cluster 59-69) uit weggefallen clusters (blauw gearceerd) die aan de andere clusters zijn toegevoegd.

3 Meetprogramma

3.1 Analysemethode

De meetmethode zoals beschreven in Paragraaf 2.5 en in bijlage IV is toegepast om de stoffen op de diverse locaties te meten.

3.2 Locaties monsternamen

Bij alle 11 drinkwaterbedrijven zijn op verschillende momenten in de drinkwaterproductie monsters genomen en geanalyseerd. Er zijn vijf verschillende categorieën: oppervlaktewater (OW), grondwater (GW), zuivering GW, zuivering OW en drinkwater. In Tabel 1 staan hoeveel monsters per bedrijf in elke categorie zijn gemeten.

Table 1: Aantallen monsters per type water. OW = oppervlaktewater, GW = grondwater, Zuivering = monsters genomen tijdens het zuiveringsproces.

Drinkwaterbedrijf	OW	Zuivering OW	GW	Zuivering GW	Drinkwater
Waternet	2	9	0	2	2
PWN	3	2	0	0	2
Evides	4	1	1	0	1
Vitens	1 (oeverfiltraat)	6	3	7	2
Oasen	3	0	4	0	1
Dunea	4	1	3	0	0
De Watergroep	1	0	1	0	2
Brabant Water	1	0	14	2	3
WML	2	6	4	4	1
WBG	1	3	1	2	2
WMD	0	0	2	2	2

3.3 Resultaten meetprogramma

Hieronder zijn de meetgegevens van de verschillende monsters weergegeven. De data is zowel per type water als per drinkwaterbedrijf weergegeven. Van alle 36 gemeten verbindingen zijn 19 boven hun rapportagegrens aangetoond in de monsters (Tabel 5). In de figuren hieronder worden enkel de 19 gedetecteerde stoffen getoond.

Tabel 5: Tabel met verbindingen die boven de rapportagegrens zijn aangetoond.

Component	Cas nummer	Aangetoond
Furosemide	54-31-9	
1,1,2,2,3,3,4,4,4-nonafluorobutane-1-sulfonic acid	375-73-5	
cycloheximide	66-81-9	
Climbazole	38083-17-9	X
2-hydroxybenzothiazool	934-34-9	X
2-methylthiobenzothiazool	615-22-5	X
Nicosulfuron	111991-09-4	X
Pentoxifylline	6493-05-6	
Benzotriazole	95-14-7	X
2,4-dinitrofenol	51-28-5	X
Terbutylazine	5915-41-3	X
Metoprolol	37350-58-6	X
2,4,5-trimethylaniline	137-17-7	
Carbendazim	10605-21-7	X
Triethyl phosphate	78-40-0	X
Thiacloprid	111988-49-9	
Phoxim	14816-18-3	
Drometrizole	2440-22-4	
Metolachloor ESA (CGA 354743)	171118-09-5	X
Warfarin	81-81-2	
Tetraglyme	143-24-8	X
Metobromuron	3060-89-7	X
Bromacil	314-40-9	
Saccharine	81-07-2	X
Metazachlor	67129-08-2	X
4-chloor-2-methylfenoxiazijnzuur (MCPA)	94-74-6	X
Oxybenzone	604-75-1	
Methamphetamine	537-46-2	
Benzoyllecgonine	519-09-5	X
MDMA	42542-10-9	X
Oxazepam	604-75-1	X
Methadon	76-99-3	

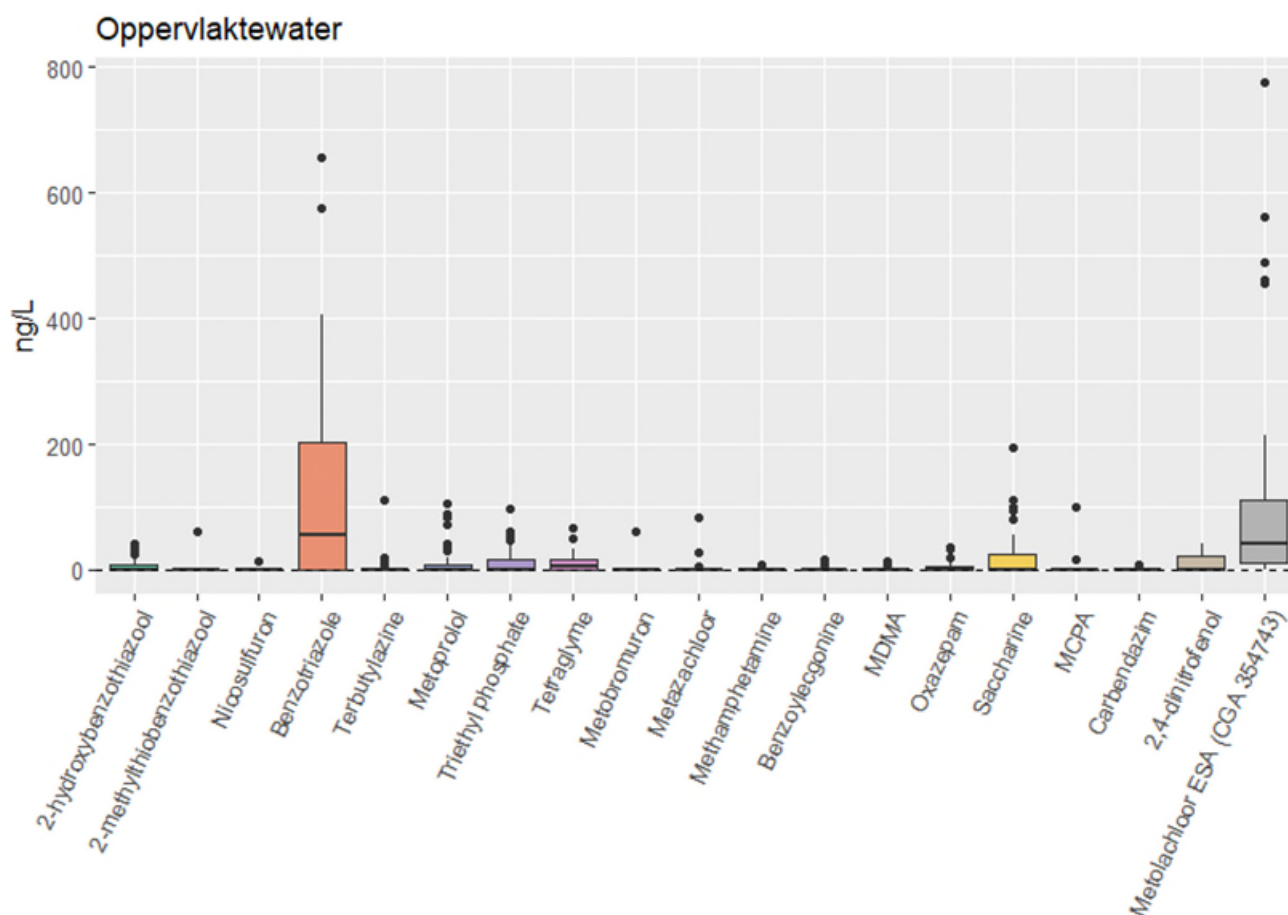
Flumioxazin	103361-09-7	
Parathion-methyl	298-0-0	
PFOA	335-67-1	
Salicylzuur	69-72-7	

3.3.1 Meetgegevens per watermatrix

Voor oppervlaktewater, grondwater en drinkwater zijn boxplots gemaakt en gemiddelde concentraties berekend (Figuur 6, Figuur 7, **Error! Reference source not found.**).

Oppervlaktewater

In totaal werden in het oppervlaktewater 19 verbindingen boven de rapportagegrens aangetoond. Twee verbindingen, namelijk benzotriazool en metolachloor vallen op vanwege je incidentele hoge concentraties ten opzichte van de andere verbindingen (> 600 ng/L) en de hogere gemiddelde concentraties. Benzotriazool kent een concentratie van ca. 207 ng/L en metolachloor ESA van ca. 110 ng/L. De andere concentratie liggen gemiddeld onder de 50 ng/L.

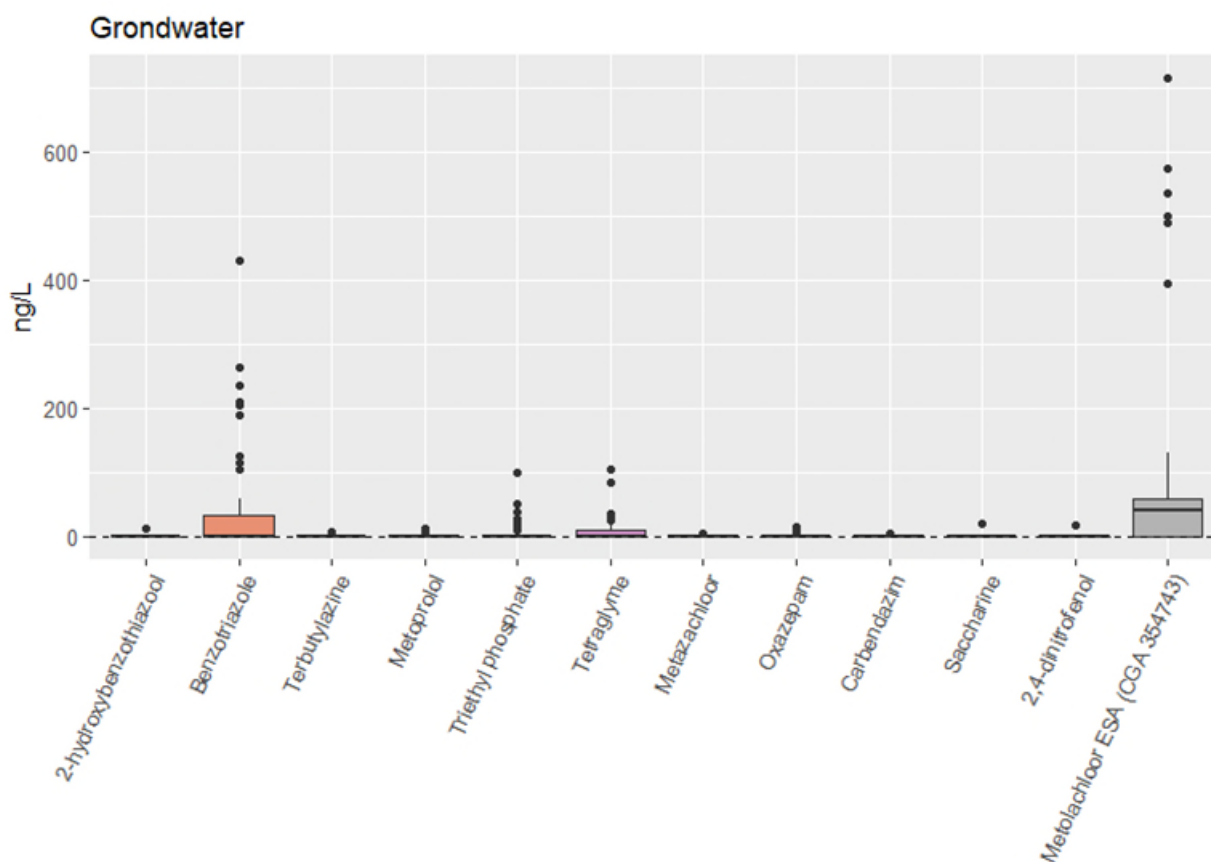


Figuur 5: Boxplot van de aangetoonde verbindingen in alle oppervlaktewatermonsters.

Grondwater

In grondwater werden in totaal 12 stoffen boven hun rapportagegrens aangetoond. Benzotriazool en metolachloor ESA vallen op door hun incidenteel hoge concentraties, respectievelijk ongeveer 400 en 700 ng/L, en hun gemiddelde

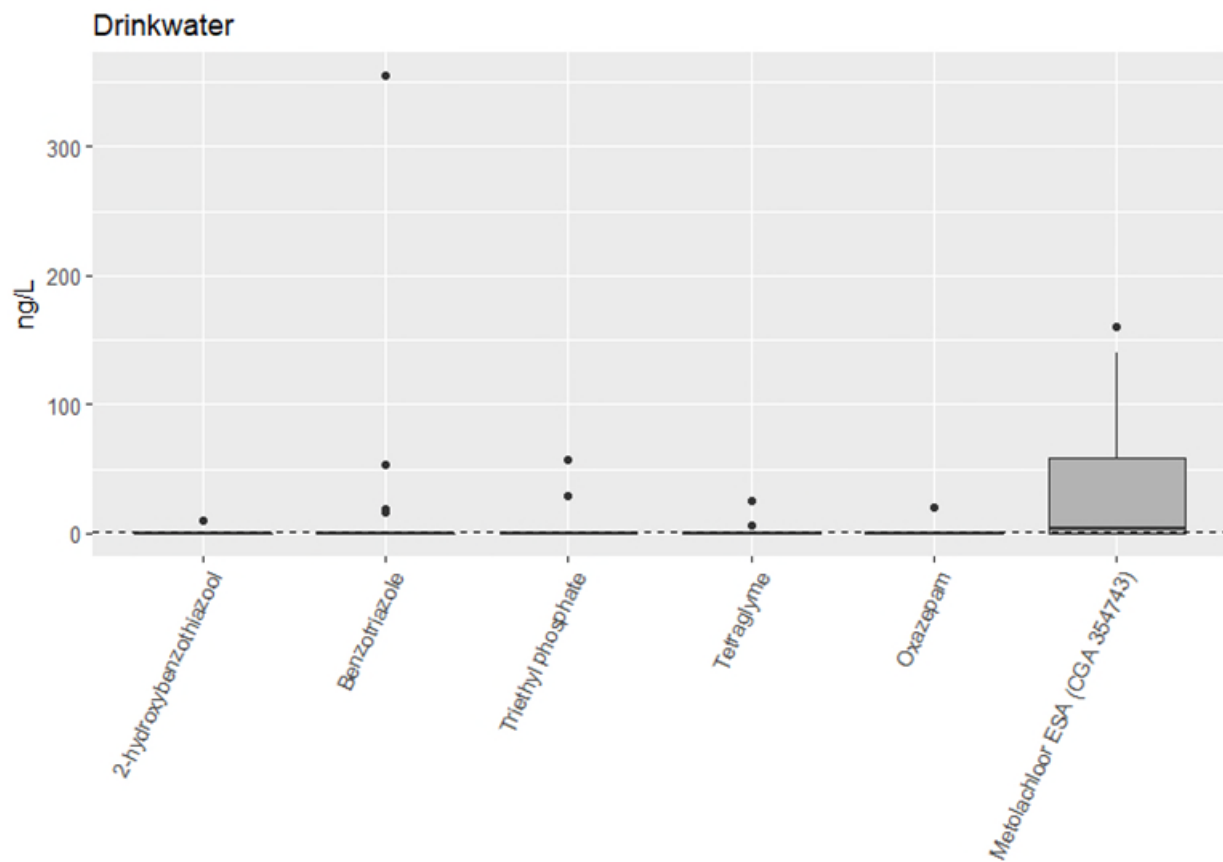
concentraties. De gemiddelde concentraties liggen rond 26 en 46 ng/L (wat voornamelijk wordt bepaald door een aantal hoge pieken).



Figuur 6: Boxplot van de aangetoonde verbindingen in alle grondwatermonsters.

Drinkwater

In de drinkwatermonsters worden 6 stoffen boven de rapportagegrens aangetoond. De meest opvallende zijn benzotriazool en metolachloor ESA met concentratie soms boven 400 en 600 ng/L. De gemiddelde concentraties van deze verbindingen zijn ca. 4 en 30 ng/L. Deze concentraties zijn overigens onder de drinkwaternorm van 1000 ng/l (1 µg/L) voor metolachloor ESA (humaan toxicologisch niet-relevante metaboliet van een pesticide), en onder de indicatieve drinkwaterrichtwaarde van 700 ng/L voor benzotriazool.



Figuur 7: Boxplot van de aangetoonde verbindingen in alle drinkwatermonsters.

4 Toxiciteitsvoorspellingen

4.1 Inleiding

Het is van groot belang om risico's verbonden aan ZZS-emissies in te kunnen schatten. Het is echter niet haalbaar of wenselijk om toxiciteitsexperimenten uit te voeren voor alle stoffen, vanwege b.v. ethische, financiële en tijdsbeperkingen. *In vitro*-onderzoeken kunnen geschikter zijn dan *in vivo*-onderzoeken, maar zijn vaak niet haalbaar voor *high-throughput* screening (Stucki et al., 2022). *In silico*-benaderingen kunnen helpen bij het inschatten van de gezondheidkundige risico's die aan dergelijke stoffen zijn verbonden (Reus et al., 2022). Deze modellen kunnen gebruikt worden om snel een uitspraak te kunnen doen over of een stof mogelijk kan worden aangemerkt als ZZS (Moon et al., 2020; Wassenaar et al., 2021). Een nadeel van deze modellen is echter het ontbreken van geschikte, grote datasets met toxicologische gegevens die als basis kunnen dienen voor voorspellende modellen (Hemmerich et al., 2020). Met name voor *machine learning*-modellen zijn grote hoeveelheden data nodig. Mechanistische modellen hebben vaak minder data nodig en maken het makkelijker om slechte data te herkennen en te extrapoleren buiten de chemische ruimte van de bestudeerde stoffen omdat het mechanisme duidelijk is, maar deze zijn vaak onvolledig en beperkt. In het huidige onderzoek is daarom een grote beschikbare dataset met toxicologische data (ToxCast) geschikt gemaakt als trainingsdataset in modellen om toxiciteit te voorspellen. Hoewel er recent al eerder modellen zijn ontwikkeld om de toxiciteit van ZZS te kunnen voorspellen (Moon et al., 2020; Wassenaar et al., 2021), worden hiervoor vaak lijsten van stoffen gebruikt waarvan bekend is dat deze toxisch zijn (ZZS-lijsten). In deze huidige studie worden de modellen gebaseerd op een breed scala aan stoffen en wordt de toepasbaarheid van de modellen getest op basis van de clusters van ZZS zoals deze zijn beschreven in Hoofdstuk 2.

Om de chemische structuur van nieuwe, onbekende stoffen te kunnen correleren aan stoffeigenschappen, is een conventioneel QSAR-model (*quantitative structure-activity relationships*) gebouwd. Deze is gebaseerd op statistische regressiemodellen voor moleculaire descriptoren (met name *structural alerts* en functionele groepen/structurele fragmenten). *Structural alerts* zijn structurele patronen binnen moleculen waarvan de aanwezigheid mogelijk verband houdt met bepaalde nadelige effecten in organen, en worden daarom veel gebruikt als input in QSAR-modellen (Yang et al., 2020). Het zijn moleculaire fragmenten met een hoge chemische reactiviteit of fragmenten die via bioactivering door menselijke enzymen kunnen worden omgezet in fragmenten met een hoge chemische reactiviteit (Limban et al., 2018). Nitrobenzeen, aniline, aziridine, epoxiden, zuurhalogeniden en onverzadigde carbonylen zijn allemaal bekende voorbeelden van structurele alerts in verband met mutageniteit (Yang et al., 2020). Functionele groepen daarentegen zijn delen (een deel van een molecuul dat ook in andere moleculen voorkomt) die de karakteristieke fysische, chemische en biologische eigenschappen van dat molecuul bepalen en niet noodzakelijkerwijs verband houden met toxiciteit (Ertl et al., 2020). Deze groepen zijn echter beide opgenomen in het QSAR model, omdat de clusters van verbindingen op de lijst van ZZS onder andere zijn gebaseerd op hun vergelijkbare functionele groepen.

Naast structurele eigenschappen is bekend dat ook fysisch-chemische eigenschappen en partitievoëfficiënten (vaak afgeleid op basis van chemische structuren) van verbindingen verband houden met (eco)toxiciteit. De correlatie tussen de hydrofobiciteit van chemicaliën - gewoonlijk uitgedrukt als de log van de octanol/water-verdelingscoëfficiënt ($\log K_{ow}$) - en hun toxische potentie is al meer dan veertig jaar geleden vastgesteld (Könemann, 1981; Lambert et al., 2022; McCarty et al., 1985). Over het algemeen hebben verbindingen met een hogere $\log K_{ow}$, en dus een hogere hydrofobiciteit, een hogere toxische potentie vanwege hun neiging tot interactie met en accumulatie in cellen en weefsels (Verhaar et al., 1992). Dit werd onlangs opnieuw bevestigd met behulp van een uitgebreide database van 617 organische chemicaliën met samengestelde en gestandaardiseerde gegevens over acute toxiciteit, waaruit blijkt dat K_{ow} omgekeerd evenredig is met ecotoxiciteit (Lambert et al., 2022). Hoewel er veel goede software tools bestaan om

de toxiciteit van stoffen in te schatten op basis van read-across en QSARs (b.v. VEGA, de OECD QSAR Toolbox en Case Ultra), zijn deze tools niet geschikt om (snel) meerdere voorspellingen te doen van de toxiciteit van stoffen waarvoor data ontbreken. Om deze reden is gekozen om zelf QSARs op te stellen op basis van fysisch-chemische parameters en functionele groepen. In BTO-rapport (BTO 2023.060; *Are PMOCs less toxic?*) werd al eerder aangetoond dat de relatie tussen fysisch-chemische parameters gerelateerd aan PMOCs (persistente mobiele organische verbindingen; log K_{oc} , log K_{ow} en biodegradatie) en toxiciteit duidelijk zichtbaar is. De molecuulstructuur van stoffen beïnvloedt de blootstelling echter op twee manieren:

- I) hun chemische eigenschappen bepalen de mate van blootstelling: door hun (slechte) oplosbaarheid, het feit dat ze allerlei biologische structuren (wel of niet) kunnen passeren of (wel of niet) ophopen in bepaalde weefsels.
- II) Daarnaast hebben ze ook een intrinsieke toxiciteit omdat ze een biologisch systeem op een bepaalde specifieke (receptor) of aspecifieke (oxidatieve stress, ophoping in membranen) manier beïnvloeden, afhankelijk van het organisme en de omgevingscondities.

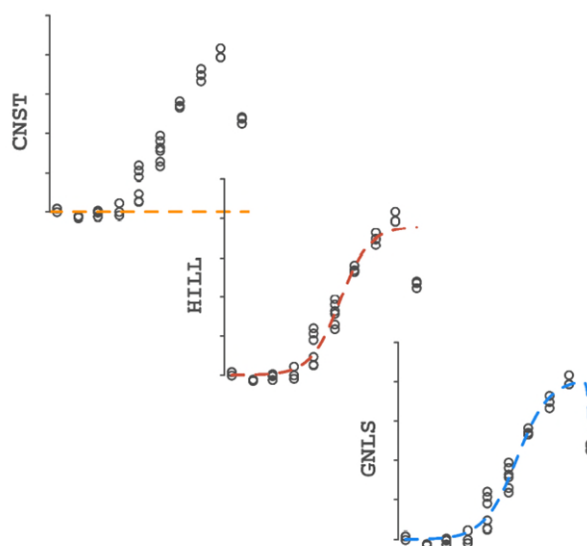
Blootstelling en toxiciteit lopen in feit dus door elkaar en de ToxCast database is juist aan de blootstellingskant niet altijd representatief. In het huidige onderzoek is onderzocht of *in silico*-technieken gebruikt kunnen worden om stofeigenschappen (moleculaire functionele groepen en/of *structural alerts*) direct te koppelen aan de toxiciteit van (ZZS) stoffen.

4.2 Methode

4.2.1 Verzameling van toxiciteitsdata

Om inzicht te verkrijgen in de toxiciteit van (ZZS) stoffen is de complete ToxCast-dataset gedownload, bestaande uit 21 databases met daarin meer dan 3,5 miljoen toxiciteitsgegevens (U.S. EPA, 2015). Wanneer alleen actieve stoffen worden meegenomen (met een *active hit call*), blijven er 357.000 datarecords over. In ToxCast wordt de actieve concentratie waarbij 50% van het effect wordt waargenomen (AC_{50} in μM) berekend met behulp van experimentele concentratie-responsreeksen voor een breed scala aan *in vitro* assays en drie modeltypen (zie Figuur 8):

- *Constant* model (CNST): een model met twee parameters met een helling en snijpunt gelijk aan 0 (lineair);
- *Hill* model (HILL): een model met drie parameters met een S-curve en vloerparameter gelijk aan 0.
- *Gain-Loss* model (GNLS): een model met vijf parameters.



Figuur 8: Model fitting (constant, hill en gain-loss model)

Log AC_{50} s voor het best voorspellende model worden automatisch berekend. In ToxCast krijgen concentratie-responsreeksen alleen een actieve *hit-call* (en hoge kwaliteitsbeoordeling) wanneer ze aan drie criteria voldoen (Feshuk et al., 2023):

1. Het Hill-model (de standaard S-curve die wordt gebruikt in dosis-responsrelaties) is het beste model.
2. De bovenkant van de gemodelleerde curve moet boven de efficiëntiedrempel liggen (de doeltreffendheidsgrenswaarde).
3. Voor ten minste één concentratie moet de mediane respons boven de werkzaamheidsdrempel liggen

Naast de gemodelleerde AC_{50} 's van het Hill-model, wordt de AC_{10} (de actieve concentratie waarbij 10% van het effect wordt waargenomen) ook afgeleid op basis van de modelparameters (d.w.z. de helling en intercept van de curve). Bovenstaande criteria zijn ook gevolgd in het huidige onderzoek. Na het opschonen op basis van deze criteria bestaat de resulterende dataset uit 1388 individuele unieke eindpunten uit *in vitro* assays uit 20 bronnen/laboratoria (Tabel 7; [ToxCast Assay Platform Descriptions 08112017.pdf \(epa.gov\)](#)), met 235.207 dataregels voor 7566 unieke stoffen (gebaseerd op CAS-nummer), zowel ZZS als niet-ZZS. Deze eindpunten zijn echter niet specifiek gerelateerd aan ZZS-eigenschappen en beslaan een breed spectrum. De ZZS-lijst die in de vorige hoofdstukken is gebruikt (zie Hoofdstuk 2.2), omvat 911 stoffen, waarvan 517 stoffen (57% van de lijst, 45.142 dataregels) zijn opgenomen in de ToxCast-database voor 1349 unieke *in vitro* assays.

Tabel 6: Partners van de ontwikkelde *in vitro* assays die zijn opgenomen in de ToxCast Database ([ToxCast Assay Platform Descriptions 08112017.pdf \(epa.gov\)](#)).

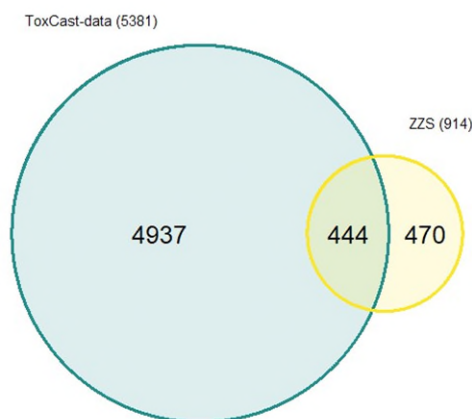
Afkorting in ToxCast	
ACEA	ACEA BioSciences, Inc.
APR	Apredica (Evotec)
ATG	Attagene, Inc.
BSK	BioSeek
NVS	Novascreen
OT	Odyssey Thera
TOX21	NCGC
CEETOX	CeeTox/OpAns
LTEA	Life Technologies / Expression Analysis
CLD	CellzDirect

CCTE	Center for Computational Toxicology and Exposure
CCTE-SHAFFER	Center for Computational Toxicology and Exposure (Shafer – neurodevelopmental toxicity)
CPHEA-STOKER	Center for Public Health and Environmental Assessment
CCTE-GLTED	Great Lakes Toxicology and Ecology Division Laboratory (GLTED)
UPITT	University of Pittsburgh
ERF	Eurofins
CCTE-MUNDY	Center for Computational Toxicology and Exposure (Ian Mundy)

4.2.2 Combineren van data met functionele groepen en fysisch-chemische descriptoren

Toxiciteitsgegevens (AC_{50} 's) uit de complete ToxCast-database (inclusief ZZS) werden gecombineerd met EPI Suite-gegevens, die informatie bevatten over fysisch-chemische kenmerken en verdelingscoëfficiënten/partitiecoëfficiënten die worden geassocieerd met bioaccumulatie en mobiliteit ($\log K_{oc}$ en $\log K_{ow}$). Er is gekozen om deze parameters mee te nemen als verklarende variabelen in de modelontwikkeling, aangezien zij geassocieerd worden met verhoogde acute toxiciteit vanwege hun relevantie voor bioconcentratie in membranen (Lambert et al., 2022). De volledige dataset werd vervolgens opgeschoond voor analyse op basis van meerdere criteria:

1. Aangezien een lage oplosbaarheid en – vaak samenvallende – hoge hydrofobiciteit van een verbinding vaak de daadwerkelijke blootstelling in een toxiciteitstest beïnvloedt (wat doorgaans leidt tot een onderschatting van het effect ervan) (Groothuis et al., 2015), zijn slecht oplosbare verbindingen (d.w.z. met een oplosbaarheid in μM onder de corresponderende AC_{50}) verwijderd uit de dataset (Jonker and Van der Heijden, 2007). Oplosbaarheid ($\log_{10} S$ in M) werd geschat met behulp van het WSKOWWIN v1.42-model. Het inkorten van de gegevens resulteerde in een dataset van 5.426 chemische stoffen voor 1588 *in vitro* toxiciteitstesten (162.743 dataregels), waarvan 453 individuele ZZS (23.169 dataregels).
2. Alle *in vitro* toxiciteitstesteindpunten waarvoor minder dan 50 dataregels (geteste stoffen) beschikbaar zijn, werden buiten beschouwing gelaten in de verdere analyse om onbevooroordeelde (*unbiased*) modellering te garanderen. Deze drempelwaarde van 50 dataregels was gebaseerd op de centrale limietstelling, waarbij aannames over de populatieverdeling zinloos zijn als de steekproefomvang groter is dan 30, omdat de steekproefverdeling dan de standaardnormale verdeling benadert (Kwak and Kim, 2017). Aangezien 70% van de data wordt gebruikt als trainingsdataset en de overige 30% van de dataregels dienst zal doen als testdataset, moet de totale subset uit minimaal $(30/0.7 = 43)$, afgerond naar 50 dataregels bestaan. Het inkorten van de gegevens op basis van dit criterium resulteerde in een dataset van 5.101 stoffen (waarvan 444 ZZS-stoffen – 48,6% van de initiële CAS-lijst, zie Figuur 20) voor 603 *in vitro* toxiciteitstesten.



Figuur 9: Venndiagram met alle stoffen in de ToxCastdatabase (5381 stoffen) en de lijst met ZZS (zie hoofdstuk 2.2)

De resulterende dataset werd gecombineerd met gegevens over functionele groepen (organische functionele groepen – structurele fragmenten) en structurele alerts (DNA-binding en eiwitbinding) uit de OECD QSAR Toolbox, gebaseerd op eerdere ervaringen (zie BTO.015 - “*Evaluation of QSAR tools in combination with bioassays for transformation products and emerging substances*”). Vervolgens zijn deze gegevens per *in vitro* toxiciteitstest verder onderverdeeld op basis van het volgende criterium:

3. Sterk gecorreleerde functionele groepen en structurele alerts ($r > 0,97$) werden verwijderd uit de dataset om multicollineariteit⁸ en voorspellingsbias te voorkomen (Næs and Mevik, 2001). Als waarden van paarsgewijze twee fysisch-chemische descriptoren een hoge correlatie hebben (hoger dan 0,97), werd de variabele met de grootste gemiddelde absolute correlatie verwijderd. De waarde van 0,97 werd gekozen op basis van een *trade-off* tussen het behouden van zoveel mogelijk fysisch-chemische descriptoren en het verwijderen van de vrijwel vergelijkbare descriptoren. Het verwijderen van teveel verklarende variabelen kan namelijk leiden tot een verlies van voorspellend vermogen in het model.

De toepassing van deze drie criteria resulteerde in een lijst van 603 toxiciteitstesten (bestaande uit in totaal 139.364 dataregels) die afzonderlijk werden geanalyseerd en die betrekking hebben op 5.101 unieke stoffen (waarvan 444 ZZS).

4.2.3 Modelbeschrijving

Om toxiciteitsvoorspellingen te kunnen doen voor ZZS en niet-ZZS, maakten we gebruik van twee modeltypen; Random Forestanalyse en multivarabele lineaire regressieanalyse. De Random Forest-analyse is een algoritme voor *supervised learning* dat gebruikmaakt van een ensemble van beslissingsbomen. Deze analyse is in staat is om zowel regressie- als classificatietaken uit te voeren, waarbij respectievelijk data geclusterd of voorspeld kunnen worden. Het algoritme selecteert voortdurend willekeurig een subset van fysisch-chemische descriptoren of functionele groepen en verdeelt de data op basis van deze descriptoren, totdat een volledige boom is ontwikkeld. Deze bomen worden geanalyseerd op voorspellend vermogen op basis van de fysisch-chemische descriptoren en structurele fragmenten (verklarende variabele) en toxiciteit (responsvariabele). Het algoritme komt tot de beste verklarende eigenschappen door altijd prioriteit te geven aan de beslisbomen met de eigenschappen die het beste presteren om toxiciteit te verklaren.

Het randomisatieproces binnen het algoritme vermindert *bias* en vermindert de variantie tussen en binnen bomen. Random Forest is, afgezien van het vermogen om nauwkeurige classificaties te bouwen, een goede en niet-parametrische methode voor het selecteren van chemische *features*, die voorspellend zijn voor toxiciteit. Om meer inzicht te krijgen in

⁸ Multicollineariteit: er is een sterk lineair verband tussen verklarende variabelen

de toxiciteit van ZZS, werden Random Forest-analyses uitgevoerd op basis van gegevens over functionele groepen, *structural alerts* en drie fysisch-chemische descriptoren van ZZS, waarbij toxicologische eindpunten als responsvariabele werden genomen.

Een vast aantal van 500 beslissingsbomen werd gebruikt in de Random Forest-analyses en de top 10 structurele alerts die de meeste variantie in AC₅₀ verklaren, werden gerapporteerd en als verklarende variabelen gebruikt in de meervoudige regressieanalyse. Het verklarend vermogen van variabelen binnen een Random Forest analyse wordt bepaald door de %IncMSE (*increase in mean-squared error* van de voorspelde waarden) te berekenen. Dit is de meest robuuste en informatieve waarde binnen de analyse. Deze waarde wordt per verklarende variabele (*j*) berekend door:

$$\%IncMSE = \frac{(MSE_j - MSE_0)}{MSE_0} * 100\% \quad [1]$$

waarbij MSE staat voor *mean squared-error* (oftewel gemiddelde kwadratische fout; het gemiddelde kwadratische verschil tussen de voorspelde waarde en de werkelijke waarde) voor beslisboom *j* t.o.v. de uitgangssituatie. Aangezien de %IncMSE in dit geval laat zien in hoeverre het model *minder* presteert wanneer een desbetreffende variabele *niet* wordt meegenomen, betekent een hogere %IncMSE-waarde een beter voorspellende variabele.

Random Forest-analyse vereist niet dat de responsvariabele en/of de voorspellende variabelen normaal verdeeld zijn. Binaire variabelen (0 / 1), zoals de aan- of afwezigheid van bepaalde functionele groepen en structurele fragmenten, zijn per definitie niet normaal verdeeld. Een nadeel van elk *machine learning*-model, inclusief Random Forest, is dat het zo ingewikkeld is dat het alleen als computermodel kan worden toegepast. De resultaten zijn daarom niet intuïtief gemakkelijk te interpreteren, wat het risico op het niet herkennen van biased data groot maakt. Bovendien bevat de Random Forest-output geen kwantitatieve regressiecoëfficiënten en geeft deze analyse daarom geen inzicht in de omvang of richting van het waargenomen effect. Om deze reden zijn de meest voorspellende verklarende variabelen uit de Random Forest-analyse gebruikt als input in een lineaire regressieanalyse (QSARs).

De QSAR's voor toxiciteit zijn afgeleid door het volgende conceptuele model toe te passen op de dataset.

$$y = x_1 + x_2 + x_3 + \dots + x_n \quad [2]$$

waarin X_1 tot X_n de 1^{ste} tot en met de n^{de} functionele groep vertegenwoordigen, en y het toxiciteitseindpunt (AC₅₀) vertegenwoordigt.

Om de modellen zo ongecompliceerd mogelijk te houden, werden geen interacties of kwadratische functies opgenomen in de QSAR-modelafleiding. Daarna werden de meest invloedrijke voorspellers van activiteit in de assays (AC₅₀) geïdentificeerd door de regressiecoëfficiënten te standaardiseren met behulp van de Relaimpo package in R statistics, Ver. 4.1.1.

4.2.4 Modevaluatie

Na het combineren van de experimentele gegevens (zie 4.2.2.) met de functionele groepen, werden subsets gemaakt van de data per *in vitro* assay en werd 70% van de data gebruikt in een trainingsdataset en de resterende 30% gebruikt als testdataset. Dataregels werden hierbij willekeurig geselecteerd. De modellen werden geëvalueerd met behulp van de R^2 en de Q^2 bij het voorspellen van toxiciteit voor stoffen die zijn weggelaten uit de initiële dataset (testdataset) (zie vergelijking 3 en 4):

jj

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad [3]$$

waarbij de R^2 wordt berekend als $1 -$ de resterende som van de kwadraten (RSS) en de totale som van de kwadraten (TSS), is y_i de waargenomen AC_{50} voor verbinding i , $\hat{y}_i$ is de voorspelde AC_{50} voor verbinding i , en $\bar{y}$ is het gemiddelde AC_{50} in de trainingsdataset. De R^2 -statistieken verklaren de variantie in de responsvariabele, verklaard door de verklarende variabele(n). Door de jaren heen is er veel discussie geweest over de R^2 -drempel waarboven een model als een goed voorspellend model kan worden beschouwd. In dit onderzoek zullen R^2 -waarden van 0,75, 0,50 of 0,25 voor responsvariabelen worden beschreven als respectievelijk substantieel, matig of zwak, volgens Hair et al. (2013) en Sarstedt et al. (2021). De voorspellende kracht van het model (voor stoffen in de testdataset) wordt geëvalueerd door de Q^2 voor de testdataset te berekenen:

$$Q^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_{ext})^2}{\sum_{i=1}^n (y_{ext} - \bar{y}_{ext})^2} \quad [4]$$

waarbij de Q^2 wordt berekend als $1 -$ de resterende som van de kwadraten (RSS) en de totale som van de kwadraten (TSS), y_{ext} is de waargenomen AC_{50} voor verbinding i , $\hat{y}_{ext}$ is de voorspelde AC_{50} voor verbinding i , en $\bar{y}_{ext}$ is het gemiddelde AC_{50} in de testdataset. De Q^2 -statistiek weerspiegelt de voorspellende relevantie en meet of een model voorspellende relevantie heeft. Q^2 -waarden boven nul geven aan dat de experimentele waarden goed zijn voorspeld en dat het model voorspellende relevantie heeft.

Daarnaast is de *sensitivity* (sensitiviteit), *specificity* (specificiteit) en *accuracy* (nauwkeurigheid) van de modellen berekend (Formule 5, 6 en 7). De sensitiviteit is het vermogen van de test om toxische stoffen en dus potentiële ZZS te herkennen (ware positieven), de specificiteit het vermogen van het model om niet-toxische stoffen te herkennen (ware negatieven) en de nauwkeurigheid het algemene voorspellend vermogen (ware positieven + ware negatieven). Daarbovenop zijn de *negative predictive value* (het aantal ware negatieven ten opzichte van alle negatieve waarden) en *positive predictive value* (het aantal ware positieven ten opzichte van alle positieve waarden) gerapporteerd (Formule 8 en 9). Omdat deze evaluatiecriteria zijn gebaseerd op nominale responsvariabelen (de stof is wel of niet toxisch), zijn de continue AC_{50} -waarden per individuele *in vitro* toxiciteitstest omgezet naar toxiciteitsklassen (lage toxiciteit of hoge toxiciteit) op basis van de distributie van AC_{50} -waarden (per toxiciteitstest), dus of de AC_{50} in de laagste 25% valt (= hoge toxiciteit) of de hoogste 75% (lage toxiciteit).

$$\text{Sensitiviteit} = \frac{\text{Ware positieven}}{\text{Ware positieven} + \text{Vals negatieven (alle geobserveerde positieven)}} \quad [5]$$

$$\text{Specificiteit} = \frac{\text{Ware negatieven}}{\text{Ware negatieven} + \text{Vals positieven (alle geobserveerde negatieven)}} \quad [6]$$

$$\begin{aligned} \text{Nauwkeurigheid} \\ = \frac{\text{Ware positieven} + \text{Ware negatieven}}{\text{Ware positieven} + \text{Ware negatieven} + \text{Vals positieven} + \text{Vals negatieven (alle waarden)}} \quad [7] \end{aligned}$$

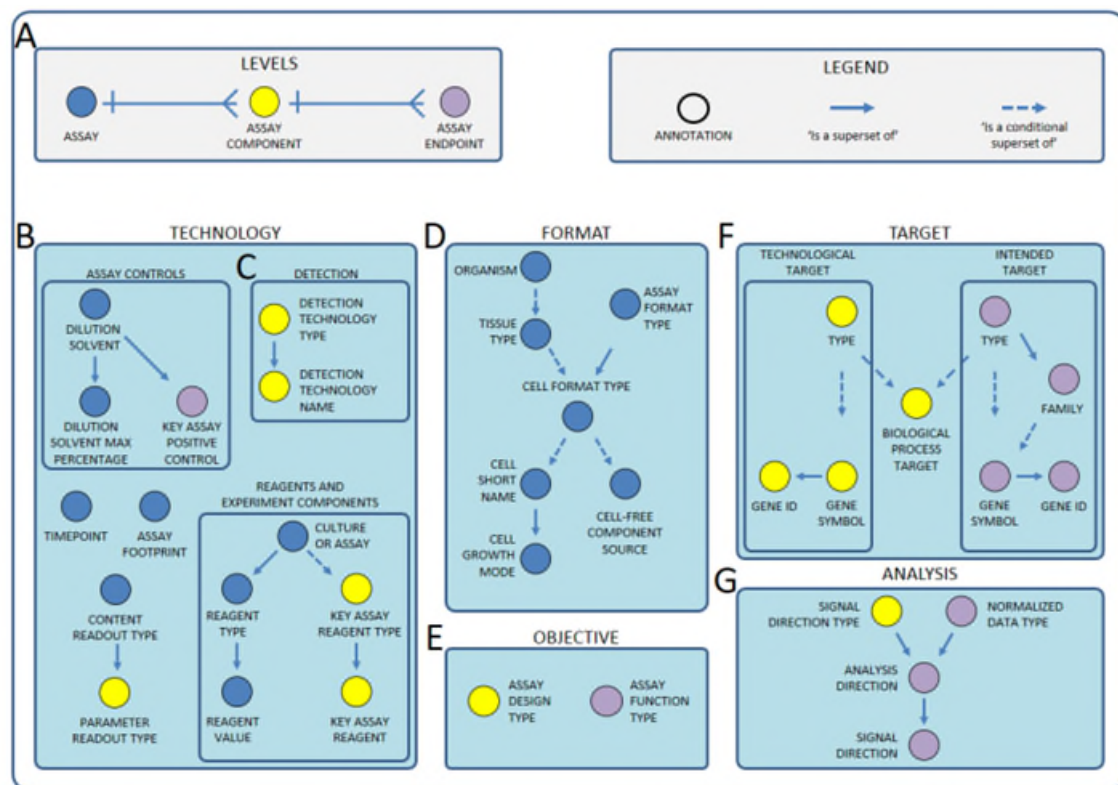
$$NPV = \frac{\text{Ware negatives}}{\text{Ware negatieven} + \text{Vals negatieven (alle voorspelde negatieven)}} \quad [8]$$

$$PPV = \frac{\text{Ware positieven}}{\text{Ware positieven} + \text{Vals positieven (alle voorspelde positieven)}} \quad [9]$$

4.2.5 Groeperen van *in vitro* assays

In vitro toxiciteitstesten werden gegroepeerd op basis van eigenschappen gerelateerd aan het toxicologische werkingsmechanisme (b.v. beoogde doeltypes (*intended target family*)). Elk datapunt in de ToxCast-database heeft een unieke set aan annotaties, beschrijvende kenmerken die een bepaald aspect van de gebruikte *in vitro* toxiciteitstest beschrijven. De meeste van de 38 annotaties zijn gerelateerd aan ten minste één andere annotatie (Figuur 4). Aangezien de dataset uit een groot aantal soorten *in vitro* assays bestaat, werden deze assays geclassificeerd op basis van assay-type en -kenmerken (d.w.z. doelfamilie (*intended target type*), technologisch doeltypes (*technological target type*), type testontwerp (*assay design type*), signaalrichting (*signal direction*), en organisme en modeltype gebaseerd op weefsel).

Een *target* kan verschillende vormen hebben (bijvoorbeeld een gen). De doelfamilie geeft de objectieve vorm weer van het beoogde *target* (de representatieve genetische familie of het biologische proces van het *target* (bijv. celcyclus of DNA-binding)), terwijl het technologische doeltypes de gemeten vorm van het *target* betreft dat wordt gebruikt in de experimentele methoden (bijv. embryonale ontwikkeling in zebravis assays) (Phuong et al., 2014). Het type testontwerp van een toxiciteitstest is gerelateerd aan de technologie die wordt gebruikt om een biologisch of fysisch proces te vertalen naar een detecteerbaar signaal (bijv. enzymreporter of groeireporter), en de signaalrichting komt overeen met de verwachte richting van het gedetecteerde signaal in relatie tot de negatieve controle (bijv. toename of afname van activiteit) (Phuong et al., 2014) (Figuur 20).



Figuur 10: Assay annotatiestructuur. De 38 verschillende annotaties zijn gegroepeerd in A. assayinformatie, B. technologie-informatie, C. detectie-informatie, D: functie-informatie, E: ontwerp-informatie, F: doelinformatie en G. analyse-informatie (Phuong et al., 2014).

4.3 Resultaten en discussie

4.3.1 AC₅₀-data per cluster

De ZZS stoffen werden op basis van MACCS fingerprints geclusterd (zie Hoofdstuk 2.3). Boxplots met informatie over de AC₅₀-data per cluster worden in Figuur 22 weergegeven. Dit zijn zowel stoffen die in meerdere assayeindpunten een duidelijk activiteit vertonen als stoffen die specifiek aanslaan in specifieke assays en dus mogelijk alleen toxisch zijn als gekeken worden naar bepaalde eindpunten (bijvoorbeeld neurale ontwikkeling). Hier is echter geen onderscheid in gemaakt. De (ZZS-)stoffen binnen cluster 18 bleken over het algemeen de laagste AC₅₀'s – en dus de hoogste toxiciteit – te hebben (gemiddelde = 2,03 µM, S.E. = 8,08, voor alle soorten *in vitro* assays). De gegevens binnen dit cluster waren echter niet normaal verdeeld, wat resulteerde in een mediaanwaarde (12,9 µM) die significant afwijkt van de gemiddelde waarde. De hoogste AC₅₀'s (gemiddelde = 36,1 mM, mediaan = 33,3 µM, S.E. = 8,64), en dus de laagste toxiciteit, zijn waargenomen voor stoffen in cluster 61. Deze waarden verschillen echter sterk tussen de verschillende typen assays. Bovendien kan een laag aantal stoffen binnen een cluster de resultaten vertekenen. Cluster 18 bestaat bijvoorbeeld uit slechts vier stoffen (iopamidol, iopromide, natriumdiatrizoaat en iomeprol, allen radiocontrastmiddelen), terwijl cluster 61 uit vijf stoffen bestaat (benzotriazool, 4-methylbenzotriazool, 5-methyl-1H-benzotriazool, 5,6-dimethylazimidobenzeen en 5-chloorbenzotriazool, allen corrosieremmers). Uit enkel de AC₅₀ data is overigens niet op te maken of deze stoffen toxischer zijn dan andere stoffen in de dataset, omdat dit sterk afhankelijk is van de verdeling van de stoffen in de subsets en de distributie van de corresponderende AC₅₀-waarden. Om deze reden is per stof-AC₅₀ combinatie het percentiel van de waarde binnen de AC₅₀-distributie bepaald, uitgaande van normaal verdeelde AC₅₀ data per assayeindpunt. De resulterende boxplots (in Figuur 23) geven vervolgens aan hoe de data binnen het cluster zich verhoudt tot de AC₅₀s van andere stoffen (van andere clusters, maar ook niet-ZZS). De percentielen van de ZZS zijn over het algemeen uniform verdeeld (van 0 tot 1), wat betekent dat ZZS niet in alle gevallen (assayeindpunten) de meest toxische stoffen zijn.

Aangezien onze Random Forest-analyse een minimale steekproefomvang van 50 gegevensinvoeren vereist, was het niet mogelijk om deze analyse voor alle clusters afzonderlijk uit te voeren (de meeste clusters bestaan over het algemeen uit minder dan 50 dataregels (Tabel 8). Een Random Forest-analyse is uitgevoerd op de functionele groepen die het meest bijdragen aan toxiciteit (Zie Hoofdstuk 4.2.2.), waarbij toxiciteitsgegevens (AC₅₀) voor alle soorten *in vitro* assays zijn gecombineerd.

4.3.2 Random forest-analyses

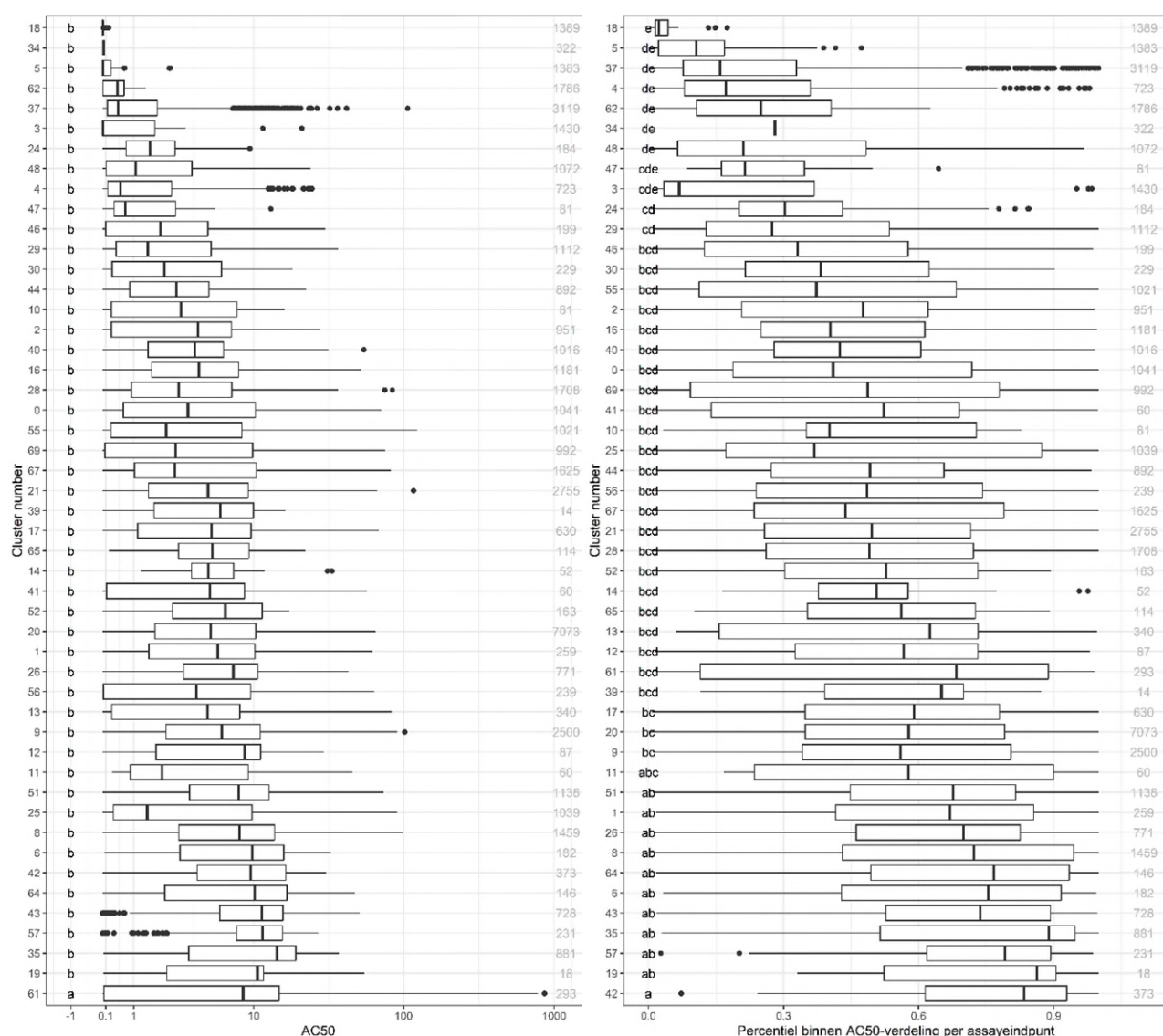
In 26,1% van alle random forest-analyses (475 *in vitro* assayeindpunten) is organotin geïdentificeerd als de belangrijkste functionele groep voor de voorspelling van de toxiciteit van stoffen (dit is echter slechts gebaseerd op een kleine dataset van stoffen die daadwerkelijk een organotingroep bevatten), gevolgd door de acetoxygroep (in 8,2% van alle random forests) en verzadigde carbocycles (8%), waarbij de laatste wordt geïdentificeerd als een belangrijke structurele alerts in verband met mutageniteit (Yang et al., 2020). Slechts vijf stoffen (waar 950 van de 37.169 AC₅₀'s bij horen) bevatten echter een functionele organotingroep, terwijl slechts twee en drie stoffen respectievelijk een acetoxygroep en een verzadigde carbocyclusgroep bevatten. Er wordt ook gezien dat 391 van de 517 stoffen meer dan één functionele groep bevatten (Figuur 23). Dit specifieke kenmerk verklaart de variatie in toxiciteit in iets meer dan 4% van alle random forest analyses. Dit resultaat is dus helaas erg onzeker en (nog) niet bruikbaar in de risicobeoordeling van stoffen waarvoor geen gegevens beschikbaar zijn.

Dit geeft ook mogelijk een vertekend beeld van het werkelijk voorspellend vermogen van de Random Forest analyses, aangezien de samenstelling van *in vitro* assays waarvoor data beschikbaar zijn in de clusters verschilt tussen de clusters. De hoogste variatie die wordt verklaard door de random forest analyses, wanneer naar alle *in vitro* assays samen wordt gekeken, is bijna 9%, voor cluster 69. Dit cluster bestaat volledig uit (oxy)benzo(pheno)nen, UV-filters.

Tabel 7: Aantal *in vitro* testeindpunten waarvoor data beschikbaar zijn per cluster.

Clusternummer	# eindpunten	Clusternummer	# eindpunten	Clusternummer	# eindpunten	Clusternummer	# eindpunten
---------------	--------------	---------------	--------------	---------------	--------------	---------------	--------------

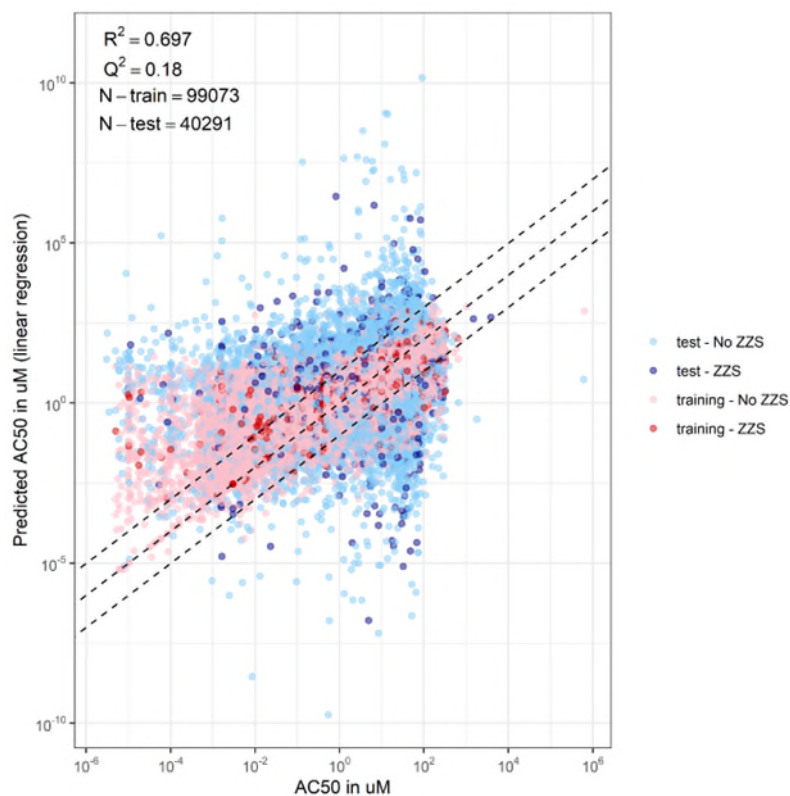
0	10	17	13	40	29	61	11
1	13	18	17	41	13	62	9
2	31	19	4	42	5	64	5
3	16	20	64	43	8	65	4
4	18	21	28	44	22	67	18
5	27	24	6	46	40	69	7
6	6	25	12	47	13		
8	8	26	27	48	55		
9	49	28	13	51	14		
10	4	29	21	52	32		
11	4	30	21	55	64		
12	5	34	10	56	17		
13	15	35	25	57	8		
14	5	37	48	59	3		
16	5	39	9	60	4		



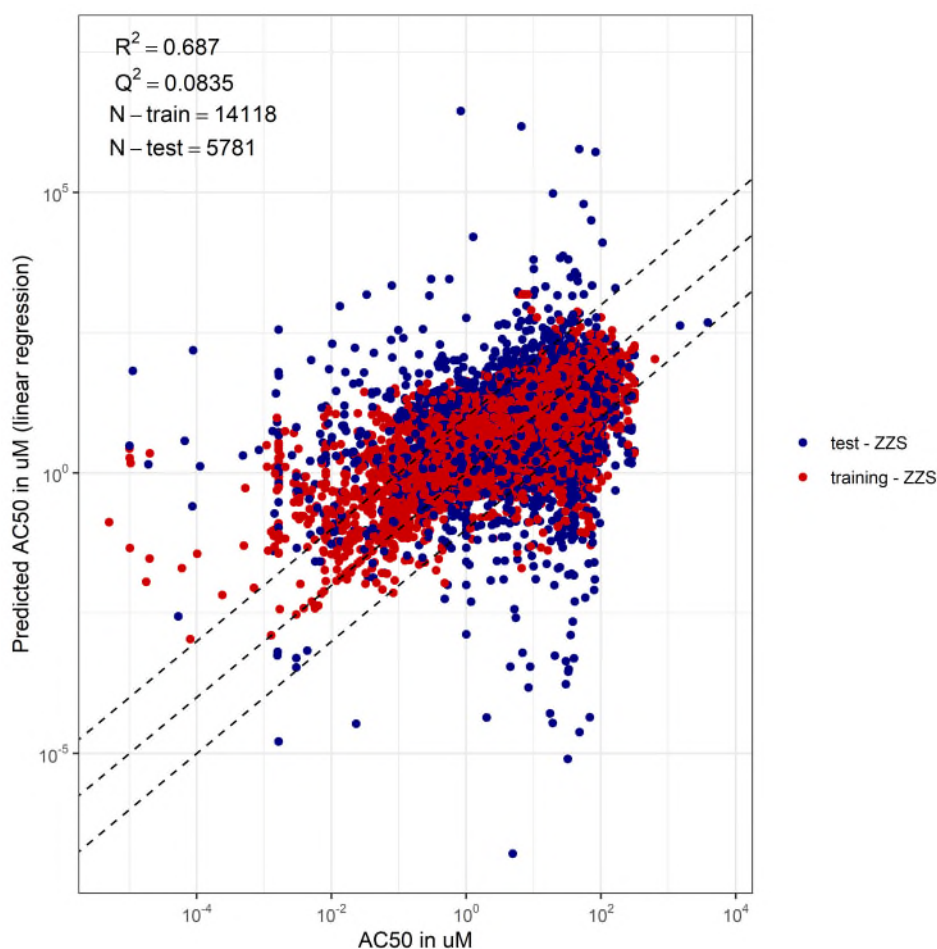
Figuur 11: Boxplots met toxiciteitswaarden (AC_{50} ; actieve concentraties waarbij 50% van het geobserveerde effect is waargenomen) voor de clusters die zijn meegenomen in deze studie (links) en boxplots met percentielen van stoffen binnen de ZZS-clusters t.o.v. alle AC_{50} -waarden binnen assay-eindpunten. De zwarte letters (a,b,c,...) geven significante verschillen tussen clusters aan. Gelijke letters geven aan dat er geen significante verschillen zijn tussen clusters. De grijze cijfers links geven de grootte van de dataset aan (aantal dataregels per cluster).

4.3.3 Lineaire regressieanalyses

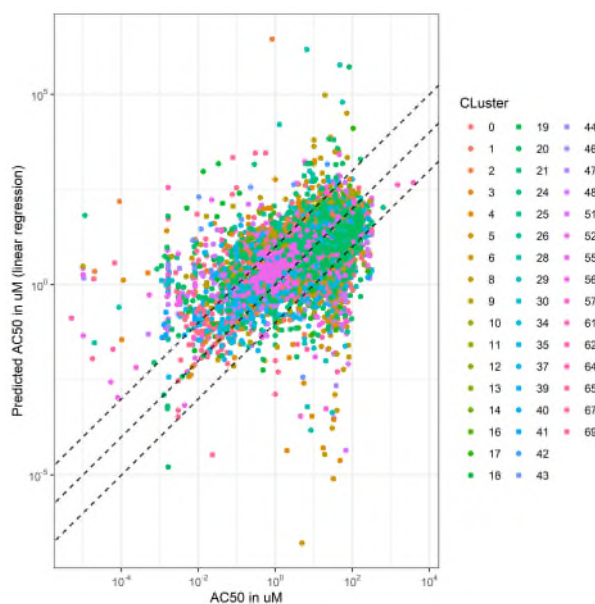
Wanneer we kijken naar het voorspellend vermogen van de analyses en de voorspelde AC_{50} s uitzetten tegen de gemeten AC_{50} s voor alle stoffen in onze geformatteerde dataset op basis van de criteria beschreven in 4.2.1. (meer dan 5000 stoffen; Figuur 23), zien we dat 75,5% van de voorspelde datapunten binnen een factor 5 vallen van de geobserveerde AC_{50} s. Van de datapunten is 2,1% overgefit ($\hat{u}-u=0$), 12,9% van de datapunten onderschat ($u/5 > \hat{u}$), en 11,5% van de datapunten overschat ($u*5 < \hat{u}$). Dit impliceert dat de modellen (afgeleid per individueel assay-eindpunt) een relatief goede indicatie geven van de toxiciteit van een stof, maar niet erg nauwkeurig zijn. Van de ZZS stoffen binnen deze dataset vallen 73,5% van de voorspelde datapunten binnen een factor 5 van de waargenomen AC_{50} . Van de voorspelde toxiciteitswaarden voor ZZS zijn 2,1% overgefit ($\hat{u}-u=0$; de data bevatten te weinig informatie om uitspraak te kunnen doen over de toxiciteit van stoffen waarvoor niet alle informatie beschikbaar is), 12,9% van de datapunten onderschat ($u/5 > \hat{u}$), 11,5% van de datapunten overschat ($u*5 < \hat{u}$). Over het algemeen wordt 69,7% van de variatie in de AC_{50} -waarden in de trainingsdataset verklaard door het lineaire regressiemodel (R^2). Wanneer het model wordt toegepast op de testdataset, wordt echter slechts 18% van de variatie verklaard door het model (Q^2). De verklaarde variatie in de testdataset is echter nog lager (8,3%) wanneer alleen gekeken wordt naar ZZS (Figuur 24). Ook binnen de eerder gedefinieerde ZZS-clusters bestaan verschillen (Figuur 25; zie Figuur 29a-d in Appendix I voor een overzicht van de voorspellingen voor de clusters waarvoor meer dan 50 dataregels beschikbaar zijn). In Figuur 26 staat beschreven in hoeverre de voorspelde AC_{50} -waarden binnen een factor vijf van de waargenomen AC_{50} -waarden liggen (*between*) en of de AC_{50} -waarden onderschat of overschat worden ten opzichte van de waargenomen waarden. Cluster 34 (benzotriazolen, o.a. gebruikt als stabilisatoren) blijkt het slechtst voorspelbaar op basis van functionele groepen, wanneer per *in vitro* assay gemodelleerd wordt. Voor dit cluster wordt 100% van de dataregels in de trainingsdataset overschat (de voorspelde waarden liggen meer dan een factor vijf hoger dan de waargenomen waarden). Deze conclusie wordt echter getrokken op basis van een kleine dataset. De beste voorspellingen voor de trainingsdataset worden gedaan voor cluster 47 (siloxanen, o.a. gebruikt als bouwstenen voor siliconenrubbers). Voor deze groep vallen 100% van de voorspelde data binnen een factor vijf van de waargenomen data. Wanneer de modellen worden toegepast op de testdataset, valt echter slechts 45% van de voorspelde data binnen een factor vijf van de waargenomen data. Hoewel voor cluster 39 (silicaten) 60% van de voorspelde data in de trainingsdataset binnen een factor vijf van de ware toxiciteitsdata ligt, valt 100% van de testdataset voor dit cluster binnen een factor vijf van de waargenomen toxiciteitsdata. Dit wordt mogelijk veroorzaakt door de kleine steekproefgrootte van dit cluster, waardoor slechts één of twee stoffen binnen dit cluster in de testdataset belanden. Wellicht dat grotere clusters voor betere voorspellingen van de test dataset zorgen, hoewel het aanpassen van het criterium voor *similarity* van stoffen er ook voor kan zorgen dat de stoffen in de testdataset minder overeenkomsten vertonen, resulterend in een lagere voorspelbaarheid van de toxiciteit van de stof binnen dit cluster. Daarbovenop zorgt het vergroten van de clusters er mogelijk voor dat nieuwe potentiële ZZS sneller in een bepaald cluster vallen. Eventueel kan bij grotere ZZS-clusters per cluster-eindpuntcombinatie QSARs afgeleid worden, wat wellicht zorgt voor gerichtere AC_{50} -voorspellingen voor stoffen die mogelijk minder vergelijkbaar zijn dan binnen de clusters die nu zijn voorgesteld. Gerichte QSARs afleiden voor de ZZS-clusters in deze studie was immers niet mogelijk vanwege datalimitaties.



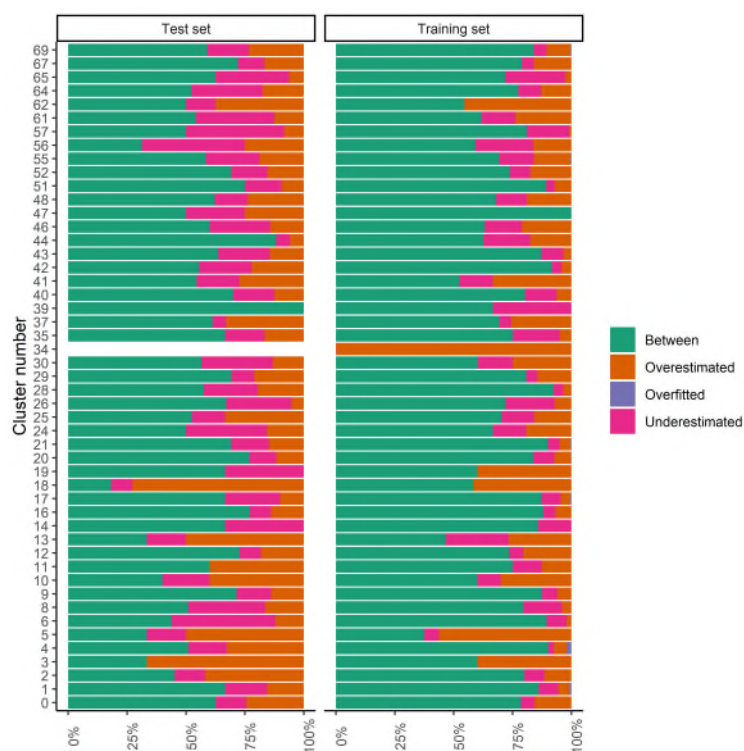
Figuur 12: Voorspelde toxiciteit (AC₅₀ in μM) door het meervoudige lineaire regressiemodel versus waargenomen toxiciteit, gebaseerd op functionele groepen, voor zowel de training- en testdataset, als stoffen die als ZZS staan aangemerkt en stoffen die niet als ZZS staan aangemerkt.



Figuur 13: Voorspelde toxiciteit (AC₅₀ in μM) door het meervoudige lineaire regressiemodel versus waargenomen toxiciteit, gebaseerd op functionele groepen, voor stoffen die als ZZS staan aangemerkt en stoffen die niet als ZZS staan aangemerkt.



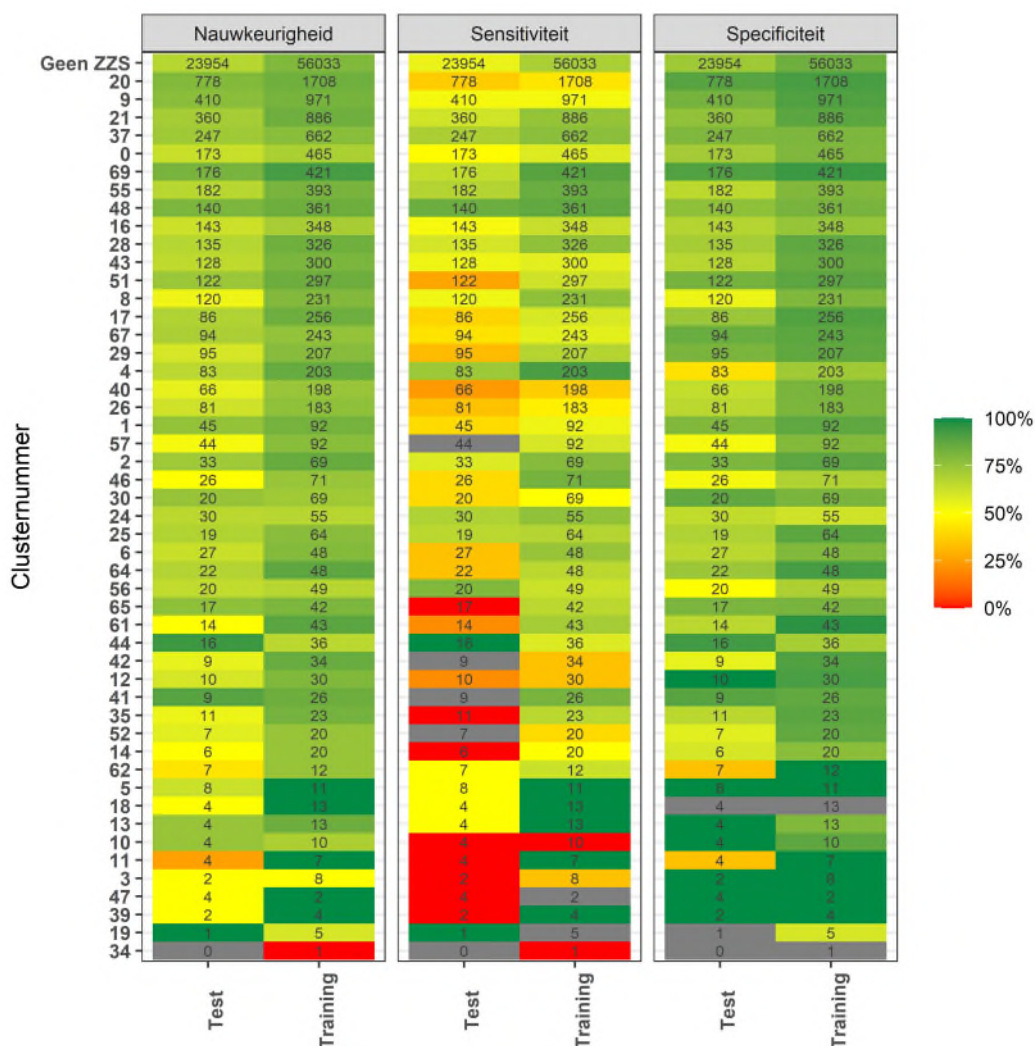
Figuur 14: Voorspelde toxiciteit (AC₅₀ in μM) door het meervoudige lineaire regressiemodel versus waargenomen toxiciteit, gebaseerd op functionele groepen, voor stoffen binnen de clusters zoals eerder beschreven. Zie Figuur 29a-d in Appendix I voor de individuele plots per ZZS-cluster.



Figuur 15: Voorspellingen van ZZS (percentages van de voorspelde toxiciteitsgegevens die binnen – of buiten – een factor vijf van de ware toxiciteitsgegevens liggen).

4.3.4 Sensitiviteit, specificiteit, nauwkeurigheid, NPV en PPV

Naast de regressieanalyses, zijn de stoffen ook geclassificeerd op basis van hun AC_{50} (zie Paragraaf 4.2.4.) en gelabeld als 'laag', 'medium' en 'hoog' toxisch. Ook de voorspelde waarden zijn op basis van de toxiciteitsdistributies beschreven in Paragraaf 4.2.4 gelabeld. Vervolgens is de sensitiviteit, specificiteit en nauwkeurigheid van de modellen bepaald. De sensitiviteit van alle individuele modellen (het percentage correct voorspelde "high toxicity"-klassen – ware positieven) was 65,6%. De specificiteit van de modellen (het percentage correct voorspelde "low toxicity"-klassen – ware negatieven) was 81,4%. Dit betekent dat over het algemeen de modellen beter zijn in het voorkomen van vals negatieven, dan van vals positieven, wat weer impliceert dat de modellen conservatief zijn. De *negative predictive value* (NPV) voor alle modellen was 84,1%, wat impliceert dat 84,1% van alle voorspelde stoffen met een "low toxicity"-categorisatie daadwerkelijk een lage toxiciteit had. De *positive predictive value* (PPV) voor alle modellen was 61,1%, wat impliceert dat 61,1% van alle stoffen waarvoor een "high toxicity"-categorisatie werd voorspeld, daadwerkelijk een hoge toxiciteit had. De algehele nauwkeurigheid van de modellen was 78,7%, wat betekent dat 78,7% van alle toxiciteitsklassen correct werden voorspeld. Als gekeken wordt naar verschillen tussen de ZZS-clusters, zien we dat de nauwkeurigheid, specificiteit en sensitiviteit van de modellen onderling erg verschilt. Merk hierbij op dat het afleiden van de modellen niet alleen op basis van data voor ZZS is gebeurd. Wanneer we dit vergelijken met modelindicatoren voor stoffen die niet zijn aangemerkt als ZZS (zo'n 80% van de complete dataset), dan zien we dat de toxiciteit van stoffen binnen een aantal clusters beter te voorspellen is dan de toxiciteit van stoffen die niet aangemerkt staan als ZZS. Voorbeelden van voldoende grote(!) clusters waarvoor de nauwkeurigheid, sensitiviteit en specificiteit hoger zijn dan de indicatoren voor niet-ZZS data zijn cluster 69 en cluster 48. Er zijn ook kleinere ZZS-clusters waarbinnen toxiciteit beter te voorspellen is dan voor niet-ZZS (zoals cluster 5), maar gezien de kleine datasets kan dit ook op toeval berusten.



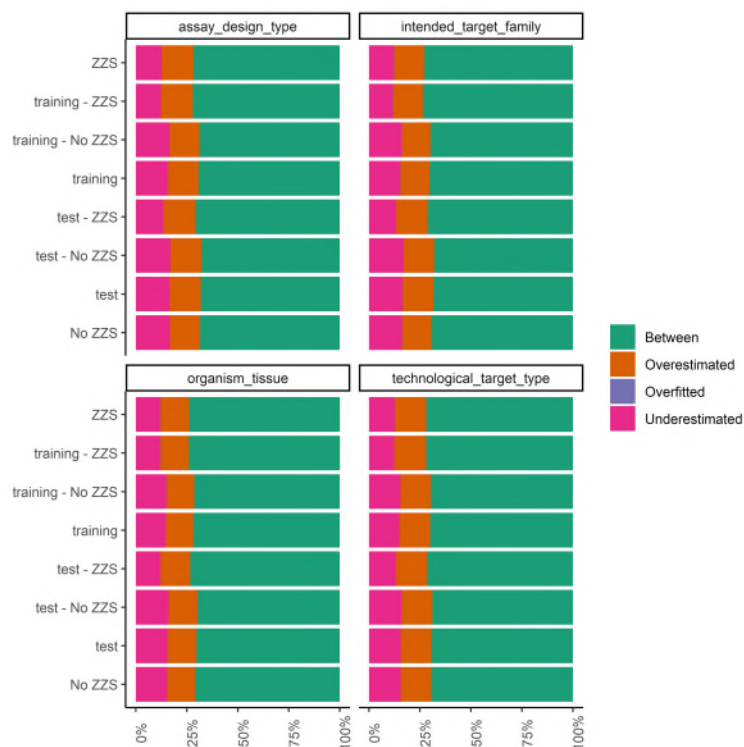
Figuur 16: Modelindicatoren (sensitiviteit, specificiteit en nauwkeurigheid) voor de resultaten per ZZS-cluster voor de test- en trainingsdataset (de dataset waarop de modellen niet, respectievelijk wel zijn gebaseerd). De zwarte cijfers in de tegels geven de grootte van de subset aan. De dataset is geordend op datasetgrootte (grootste bovenaan).

Hoewel voorspellingen voor continue data (AC_{50} s) relatief onzeker waren (zie Hoofdstuk 4.3.3.), zijn de voorspellingen voor de toxiciteitsklassen veel nauwkeuriger. Er zijn echter grote verschillen in de bepaalde sensitiviteit specificiteit en nauwkeurigheid tussen meegenomen *in vitro* assays en ZZS-clusters (en combinaties daarvan). Verschillende ZZS zorgen namelijk niet altijd voor een verhoogde activiteit in alle assays. Om een potentieel toxische stof (waarvoor geen toxiciteitsgegevens beschikbaar zijn) aan te kunnen merken als ZZS moet deze stof voldoen aan een aantal criteria. Deze zijn vastgelegd in artikel 57 van de REACH Verordening (EG) 1907/2006. Stoffen met een of meer van de volgende toxicologische eigenschappen voldoen aan deze criteria: de stoffen zijn kankerverwekkend, mutageen of giftig voor de voortplanting. Om een stof aan te kunnen merken als potentiële ZZS op basis van modellen met data uit de ToxCast-database is dus van belang om de juist *in vitro* assays te selecteren met eindpunten die het sterkst correleren met de ZZS-criteria. Mogelijk blijft dat een stof niet mutageen is, maar vertoont deze waarschijnlijk wel biologische activiteit in een assay gerelateerd aan reproductie (zoals een hormoonverstoringssay). Ook binnen de ZZS-clusters zoals bepaald in Hoofdstuk 2.3. kunnen stoffen voorkomen die afwijken ten opzichte van andere stoffen in het cluster als we kijken naar de biologische activiteit van de stof in diverse assays.

4.3.5 Subsets van *in vitro* toxiciteitstesten gebaseerd op annotaties

In vitro assayeindpunten werden geclusterd op basis van diverse annotaties (Zie Hoofdstuk 4.2.5.). De samenstelling van deze clusters aan assayeindpunten kan worden teruggevonden in de ToxCast database (U.S. EPA, 2015). Resultaten met

betrekking tot modellen die zijn getraind zijn op subsets van data op basis *in vitro* annotaties zijn weergegeven in Appendix II. In een vervolgstap werden Random Forest-analyses en de meervoudige lineaire regressieanalyses gebaseerd op de functionele groepen, aangezien deze de meest consistente verklarende parameters bleken te zijn.



Figuur 17: Voorspellingen van toxiciteitsgegevens (percentages van de voorspelde toxiciteitsgegevens die binnen – of buiten – een factor vijf van de werkelijke toxiciteit liggen), wanneer in vitro testen worden geclusterd op basis van assay-annotaties.

4.3.6 Algemene discussie m.b.t. toxiciteit

In dit hoofdstuk is gekeken naar de voorspelbaarheid van de toxiciteit van stoffen (ZZS en niet-ZZS) op basis van functionele groepen en fysisch-chemische parameters. Over het algemeen was het voorspellend vermogen gebaseerd op functionele groepen voor de Random Forest bijzonder laag, en was het voorspellend vermogen een stuk hoger voor lineaire regressiemodellen. Bij het meenemen van de individuele functionele groepen is gebruik gemaakt van dummyvariabelen in het regressiemodel. Bij een dummyvariabele (een variabele met een 0 of 1 welke de af- of aanwezigheid van een structureel element weerspiegelt) met een waarde van 0 zal in het regressiemodel ervoor zorgen dat de coëfficiënt ervan uit de vergelijking verdwijnt, wat impliceert dat bij afwezigheid van het structurele element deze niet wordt geacht bij te dragen aan de toxiciteit van de stof. Dit betekent dat ook niet wordt geacht dat deze structurele elementen een omgekeerd effect kunnen hebben op activiteit binnen de assay (de aanwezigheid van het element zorgt voor een verlaagde toxiciteit). Omgekeerd zorgt de waarde van 1 ervoor dat de coëfficiënt functioneert als een aanvullend snijpunt, aangezien deze waarde wordt vermenigvuldigd met 1. Van enkele functionele groepen is echter bekend dat ze deactiverend zijn, wat betekent dat van stoffen met deze structurele groepen bekend is dat ze minder toxisch zijn (Honma, 2020; Myatt et al., 2018). Deactiverende features (-1), structuren die mogelijk zorgen voor minder biologische activiteit, zoals in sommige QSAR-software wordt meegenomen, zijn op dit moment niet meegenomen in de modellen, wat betekent dat de toxiciteit van stoffen mogelijk ook overschat kan worden. Dit is met het oog op beleidsvorming echter gunstiger dan dat de toxiciteit van stoffen wordt onderschat door de modellen. Ook bestaan er mogelijk interacties tussen functionele groepen die toxische effecten kunnen versterken of afzwakken (synergisme of antagonisme), welke naar verwachting niet lineair schalen met toxiciteit. Andere *machine learning* en *deep learning*-modellen (zoals KNN (k-nearest neighbors) en andere neurale netwerken) zouden kunnen worden gebruikt om dergelijke complexe relaties tussen de functionele groepen en toxiciteit bloot te leggen. Waar Random Forest in deze studie echter minimaal 50 dataregels vereist (waarvan 30% wordt gebruikt in de testdata), vereisen neurale netwerken veel meer data (Roßbach, 2018), welke niet altijd beschikbaar zijn.

Over het algemeen zijn er geen grote verschillen zichtbaar tussen de voorspelbaarheid van ZZS ten opzichte van niet-ZZS in beide modellen, hoewel de verschillen in voorspelbaarheid wel groter zijn tussen de training- en testdataset bij niet-ZZS. De ZZS-data beslaan een stuk minder datapunten ten opzichte van de niet-ZZS data en de stoffen in deze data zijn beperkt tot stoffen met een relatief lage AC₅₀. Dit betekent dat de stoffen een minder breed toxiciteitsdomein hebben en daarom wellicht eenvoudiger te voorspellen zijn. De stoffen beslaan echter ook slechts een beperkte set aan functionele groepen. Tussen clusters van ZZS werden over het algemeen wel grote verschillen waargenomen in de voorspelbaarheid van toxiciteit wat mogelijk te wijten is aan verschillen in datasetgrootte, combinatie van *in vitro* assays en combinaties van functionele groepen in de dataset. Wellicht is de voorspelbaarheid van toxiciteit hoger bij het meenemen van diverse stofgroepen met gelijke toxische werkingsmechanismen (zie BTO-rapport *Zijn PMOCs minder giftig?*). Hoewel de clusters in het huidige onderzoek zijn gebaseerd op “similarity indices” (gelijkheidsindices, zie Hoofdstuk 2.3), hebben waarschijnlijk niet alle stoffen binnen de clusters precies hetzelfde toxicologische werkingsmechanisme, wat kan zorgen voor spreiding van uitkomsten van *in vitro* assays. In het huidige rapport is daarom gekozen om *in vitro* assays te clusteren om toxiciteitsvoorspellingen van stoffen op basis van gelijke werkingsmechanismen te beoordelen. Hieruit blijkt dat met name toxiciteit op basis van een celmorfologisch doeltype en neurologische ontwikkeling goed te voorspellen zijn, terwijl andere eindpunten (zoals membraameiwit-toxiciteit) minder goed te voorspellen zijn. Daarnaast kan covariantie bestaan in *in vitro* assay typen tussen diverse categorieën, zoals bijvoorbeeld het geval is voor neurologische ontwikkeling (beoogd doelfamilie) en elektrische activiteit (technologisch doeltype). Dit betekent dat het afleiden van een model waarbij *in vitro* assays zijn geclusterd op basis van een subcategorie op basis van de ene annotatie (bijvoorbeeld neurologische ontwikkeling binnen de annotatie beoogde doelfamilie) zal leiden tot exact dezelfde voorspelde toxiciteit als een model waarbij assays zijn geclusterd op een andere subcategorie binnen een andere annotatie. Het meenemen van een combinatie van zowel diverse stofgroepen als diverse eindpunten vereist een behoorlijke rekenkracht en hoeveelheden data, aangezien dit mogelijk duizenden combinaties van stofgroep en categorie betreft. Slechts een klein deel van deze combinaties bevat voldoende toxicologische data om een Random Forest of regressieanalyse uit te voeren (> 50 dataregels). Per model blijft er dan een relatief klein aantal datapunten over voor regressie, wat resulteert in een erg lage R² en p-waarde van de individuele regressies. Daarnaast maakt dit het model minder geschikt om in de praktijk toe te passen, aangezien de voorspelbaarheid van toxiciteit voor stoffen buiten de groep stoffen waarop het model gebaseerd is waarschijnlijk nog veel lager ligt. De algehele conclusie van deze studie is dan ook dat de huidige analyses hoogstwaarschijnlijk onvoldoende in staat zijn om de structurele eigenschappen (op basis van functionele groepen gerapporteerd in de QSAR Toolbox) te relateren aan de

toxiciteit van stoffen en dat op basis van het merendeel van de modellen onbekende stoffen niet als ZZS kunnen worden aangemerkt, zeker niet op basis van modellen waarvoor een groot aantal assayeindpunten is gecombineerd. Er bestaan echter grote verschillen tussen het voorspellend vermogen voor individuele *in vitro* assay eindpunten, waardoor er voor een deel van de individuele *in vitro* assay eindpunten wel realistische toxiciteitsvoorspellingen gedaan kan worden. Daarnaast is in de huidige studie enkel gekeken naar de individuele toxiciteit van stoffen, terwijl stoffen in mengsels mogelijk extra of juist minder toxisch kunnen zijn. Ondanks dat de synergistische (of antagonistische) werking van bepaalde stofcombinaties op het gebied van toxiciteit niet altijd bekend is, zijn er diverse manieren om om te gaan met mengseltoxiciteit (ook in beleid). Het RIVM heeft in een recent rapport beschreven welke methoden kunnen worden gebruikt door overheden om de effecten van stoffenmengsels voor mens en milieu te kunnen inschatten (Bodar et al., 2022). Een samenvatting hiervan (inclusief implicaties voor het huidige BTO-rapport) staat in bijlage III. Hoewel de toxiciteitsvoorspellingen in de huidige studie niet altijd even goed waren, zijn de voorspellingen naar verwachting wel betrouwbaar genoeg om een uitspraak te kunnen doen over de toxiciteitsklasse van een stof, wat blijkt uit de hoge berekende sensitiviteit en specificiteit van de modellen in de huidige studie. Dit is iets wat in de online applicatie AquaPriori verder wordt uitgewerkt.

5 Conclusies en aanbevelingen

Aan de hand van clustering kan een indeling gemaakt worden van Zeer Zorgwekkende Stoffen (ZZS) op basis van moleculaire eigenschappen. Hiermee is het voor een brede range aan ZZS mogelijk om met een enkele doelstofanalysemethode een beeld te vormen van het gedrag van ZZS in de (drink)watercyclus. Van de 36 ZZS die gemeten zijn in een meetcampagne bij Nederlandse en Vlaamse drinkwaterbedrijven zijn 19 ZZS aangetroffen boven de rapportagegrenzen. De meeste stoffen en de hoogste concentraties werden aangetroffen in oppervlaktewater. Op alle locaties zijn benzotriazole en metolachloor ESA gevonden. Ook zijn deze stoffen aangetroffen in de hoogste concentraties van alle gemeten stoffen.

In het toxiciteitshoofdstuk is gekeken naar de voorspelbaarheid van de toxiciteit van stoffen (ZZS en niet-ZZS) op basis van functionele groepen en fysisch-chemische parameters. Over het algemeen was het voorspellend vermogen gebaseerd op functionele groepen voor de Random Forest bijzonder laag, en was het voorspellend vermogen een stuk hoger voor lineaire regressiemodellen. Ook zijn er geen grote verschillen zichtbaar tussen de voorspelbaarheid van ZZS ten opzichte van niet-ZZS in beide modellen, hoewel de verschillen in voorspelbaarheid wel groter zijn tussen de training- en testdataset bij niet-ZZS. Tussen clusters van ZZS werden wel grote verschillen waargenomen in de voorspelbaarheid van toxiciteit wat mogelijk te wijten is aan verschillen in datasetgrootte, combinatie van *in vitro* assays en combinaties van functionele groepen in de dataset. De algehele conclusie van het toxiciteitshoofdstuk is dat voor het merendeel van de assayeindpunten de huidige analyses hoogstwaarschijnlijk onvoldoende in staat zijn om de toxiciteit van stoffen te voorspellen op basis van structuureigenschappen. Ook kunnen op basis van het merendeel van de modellen onbekende stoffen niet als ZZS kunnen worden aangemerkt, zeker niet op basis van modellen waarvoor een groot aantal assayeindpunten is gecombineerd. Wel zijn reageren stoffen binnen hetzelfde cluster over het algemeen wel op een zelfde manier in de bioassays, wat betekent dat het clusteren van stoffen op basis van fingerprints zorgt voor clusters met stoffen met een vergelijkbaar toxisch werkingsmechanisme.

Er bestaan echter grote verschillen tussen het voorspellend vermogen voor individuele *in vitro* assay eindpunten, waardoor er voor een deel van de individuele *in vitro* assay eindpunten wel realistische toxiciteitsvoorspellingen gedaan kan worden. Daarnaast zijn de toxiciteitsvoorspellingen naar verwachting wel betrouwbaar genoeg om een uitspraak te kunnen doen over de toxiciteitsklasse van een stof, wat blijkt uit de hoge berekende sensitiviteit en specificiteit van de modellen in de huidige studie.

De volgende aanbevelingen worden gedaan:

- Ondanks dat de verschillen tussen moleculen klein zijn in een cluster, kan er echter nog wel een verschil in werkingsmechanisme voor bijvoorbeeld toxiciteit of de zuivering zijn. Aanbevolen wordt om nader te onderzoeken hoe omgegaan moet worden met de verschillen binnen een cluster.
- Regelmatig wordt de lijst met ZZS aangevuld met nieuwe ZZS-stoffen. RIVM identificeert deze nieuwe ZZS, onder meer op basis van zelfclassificatie en gegevens van het ECHA (zo zijn in september 2023 een aantal ZZS stoffen toegevoegd, zie <https://rvs.rivm.nl/nieuws/nieuwe-zeer-zorgwekkende-stoffen-september-2023>). Ook zijn er recentelijk gewasbeschermingsmiddelen toegevoegd aan de ZZS. Het wordt aanbevolen om de clustering na elke grote herziening van de ZZS-lijst te herhalen.
- De huidige aanpak volgt een individuele stofaanpak. Het kan echter zo zijn dat een combinatie van stoffen samen zorgt voor een verhoogde toxiciteit, of dat er juist minder toxiciteit wordt waargenomen bij een bepaalde combinatie van stoffen. In het milieu en in ons (drink)water zal het nooit voorkomen dat slechts één stof toxische activiteit vertoont. Om deze reden is er steeds meer aandacht voor een risicogestuurde aanpak

door middel van mengseltoxiciteit. Het RIVM heeft in een recent rapport beschreven welke methoden kunnen worden gebruikt door overheden om de effecten van stoffenmengsels voor mens en milieu te kunnen inschatten. Een samenvatting hiervan (inclusief implicaties voor het huidige BTO-rapport) staat in bijlage III. In een toekomstig project kan dieper gekeken worden naar de mengseltoxiciteit van ZZS (en andere stoffen) door gebruik te maken van methoden die in deze bijlage staan beschreven.

- Één van de conclusies in het huidige rapport is dat het betrouwbaarder is om toxiciteitsklassen te voorspellen dan om toxiciteitswaarden te voorspellen (AC_{50} s). In een vervolgproject kan mogelijk dieper ingegaan worden op het gebruik van toxiciteitsklassen in de prioriteren van stoffen (zowel ZZS als niet-ZZS). In dat geval kan dan ook opnieuw gekeken worden naar de toxische werkingsmechanismen van clusters van ZZS. Waarschijnlijk zijn stoffen in bepaalde ZZS-clusters enkel toxisch voor specifieke eindpunten en vertonen ze niet allemaal dezelfde activiteit in alle *in vitro* assayeindpunten. Aanbevolen wordt om meer inzicht te verkrijgen in de (verschillende) toxische werkingsmechanismen van de ZZS-clusters.
- In het huidige onderzoek is niet gekeken naar de mogelijkheden om mengseltoxiciteit te voorspellen met het oog op beleid. In een vervolgonderzoek na het huidige BTO-project ZZS kan gekeken worden naar het groeperen van mengsels van ZZS en het effecten van mengsels en het onderbouwen van het MAF-concept (het gebruik van zogenaamde mixture allocation factors - een extra risicofactor om alle mogelijke risico's voor alle producten af te dekken) voor humane effecten bij immissietoetsen. Ook kan in vervolgonderzoek gekeken kunnen worden naar ecotoxiciteit door SSDs (soortgevoeligheidsverdelingen) voor ZZS clusters op te stellen, bijvoorbeeld op basis van data uit de ECOTOX-database van de US EPA, en (ms)PAFs op specifieke locaties op basis van monitoringsdata.
- Het RIVM maakt gebruik van een ZZS similarity tool, waarmee stoffen vergeleken kunnen worden met de meest gelijkende ZZS. Aanbevolen wordt om te onderzoeken of en hoe de in dit rapport gevolgde aanpak d.m.v. clustering en voorspellingsmodellen in AquaPriori gecombineerd kunnen worden de ZZS similarity tool.

6 Literatuur

- Bajusz, D., Rácz, A. & Héberger, K., 2015, Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations?. *J Cheminform* 7, 20. <https://doi.org/10.1186/s13321-015-0069-3>
- Bodar, C., de Boer, L., ter Burg, W., Janssen, N. and Smit, E. 2022. Cumulatie en vergunningverlening ZZS.
- Brinkmann, T., Santonja, G.G., Yükseler, H., Roudier, S. and Sancho, L.D. 2016. Best available techniques (BAT) reference document for common waste water and waste gas treatment/management systems in the chemical sector. Publications Office of the European Union.
- Ertl, P., Altmann, E. and McKenna, J.M. 2020. The Most Common Functional Groups in Bioactive Molecules and How Their Popularity Has Evolved over Time. *J Med Chem* 63(15), 8408-8418.
- Escher, B., Braun, G. and Zarfl, C. 2020. Exploring the concepts of concentration addition and independent action using a linear low-effect mixture model. *Environmental Toxicology and Chemistry* 39(12), 2552-2559.
- European Comission 2006 VERORDENING (EG) Nr. 1907/2006 VAN HET EUROPEES PARLEMENT EN DE RAAD: ANNEX XIV LIJST VAN AUTORISATIEPLICHTIGE STOFFEN.
- Feshuk, M., Kolaczowski, L., Dunham, K., Davidson-Fritz, S.E., Carstens, K.E., Brown, J., Judson, R.S. and Paul Friedman, K. 2023. The ToxCast pipeline: updates to curve-fitting approaches and database structure. *Front Toxicol* 5, 1275980.
- Groothuis, F.A., Heringa, M.B., Nicol, B., Hermens, J.L.M., Blaauboer, B.J. and Kramer, N.I. 2015. Dose metric considerations in in vitro assays to improve quantitative in vitro–in vivo dose extrapolations. *Toxicology* 332, 30-40.
- Hair, J.F., Ringle, C.M. and Sarstedt, M. 2013. Partial least squares structural equation modeling: Rigorous applications, better results and higher acceptance. *Long range planning* 46(1-2), 1-12.
- Hemmerich, J., Troger, F., Füzi, B. and G, F.E. 2020. Using Machine Learning Methods and Structural Alerts for Prediction of Mitochondrial Toxicity. *Molecular informatics* 39(5), e2000005.
- Holmberg, R., Wedeby, E. B., Nikolov, N. G., & Tyle, H. (2021). How many potential vPvM/PMT substances have been registered under REACH? - vPvM/PMT-screening by using the Danish (Q)SAR database. Danmarks Tekniske Universitet.
- Honma, M. 2020. An assessment of mutagenicity of chemical substances by (quantitative) structure–activity relationship. *Genes and Environment* 42(1), 1-13.
- Jonker, M.T. and Van der Heijden, S.A. 2007. Bioconcentration factor hydrophobicity cutoff: An artificial phenomenon reconstructed. *Environmental science & technology* 41(21), 7363-7369.
- Könemann, H. 1981. Quantitative structure-activity relationships in fish toxicity studies Part 1: Relationship for 50 industrial pollutants. *Toxicology* 19(3), 209-221.
- Kwak, S.G. and Kim, J.H. 2017. Central limit theorem: the cornerstone of modern statistics. *Korean journal of anesthesiology* 70(2), 144-156.
- Lambert, F.N., Vivian, D.N., Raimondo, S., Tebes-Stevens, C.T. and Barron, M.G. 2022. Relationships Between Aquatic Toxicity, Chemical Hydrophobicity, and Mode of Action: Log Kow Revisited. *Archives of Environmental Contamination and Toxicology*.
- Limban, C., Nuță, D.C., Chiriță, C., Negreș, S., Arsene, A.L., Goumenou, M., Karakitsios, S.P., Tsatsakis, A.M. and Sarigiannis, D.A. 2018. The use of structural alerts to avoid the toxicity of pharmaceuticals. *Toxicology Reports* 5, 943-953.
- Maas, J., Van De Plassche, E., Straetmans, A., Vethaak, A. and Belfroid, A. 2003 Normstelling voor bioassays, Rijkswaterstaat, RIKZ.
- McCarty, L., Hodson, P., Craig, G. and Kaiser, K. 1985. The use of quantitative structure-activity relationships to predict the acute and chronic toxicities of organic chemicals to fish. *Environmental Toxicology and Chemistry: An International Journal* 4(5), 595-606.
- Moon, J., Lee, B., Ra, J.-S. and Kim, K.-T. 2020. Predicting PBT and CMR properties of substances of very high concern (SVHCs) using QSAR models, and application for K-REACH. *Toxicology Reports* 7, 995-1000.
- Myatt, G.J., Ahlberg, E., Akahori, Y., Allen, D., Amberg, A., Anger, L.T., Aptula, A., Auerbach, S., Beilke, L., Bellion, P., Benigni, R., Bercu, J., Booth, E.D., Bower, D., Brigo, A., Burden, N., Cammerer, Z., Cronin, M.T.D., Cross, K.P.,

- Custer, L., Dettwiler, M., Dobo, K., Ford, K.A., Fortin, M.C., Gad-McDonald, S.E., Gellatly, N., Gervais, V., Glover, K.P., Glowienke, S., Van Gompel, J., Gutsell, S., Hardy, B., Harvey, J.S., Hillegass, J., Honma, M., Hsieh, J.-H., Hsu, C.-W., Hughes, K., Johnson, C., Jolly, R., Jones, D., Kemper, R., Kenyon, M.O., Kim, M.T., Kruhlak, N.L., Kulkarni, S.A., Kümmerer, K., Leavitt, P., Majer, B., Masten, S., Miller, S., Moser, J., Mumtaz, M., Muster, W., Neilson, L., Oprea, T.I., Patlewicz, G., Paulino, A., Lo Piparo, E., Powley, M., Quigley, D.P., Reddy, M.V., Richarz, A.-N., Ruiz, P., Schilter, B., Serafimova, R., Simpson, W., Stavitskaya, L., Stidl, R., Suarez-Rodriguez, D., Szabo, D.T., Teasdale, A., Trejo-Martin, A., Valentin, J.-P., Vuorinen, A., Wall, B.A., Watts, P., White, A.T., Wichard, J., Witt, K.L., Woolley, A., Woolley, D., Zwickl, C. and Hasselgren, C. 2018. In silico toxicology protocols. *Regulatory Toxicology and Pharmacology* 96, 1-17.
- Næs, T. and Mevik, B.H. 2001. Understanding the collinearity problem in regression and discriminant analysis. *Journal of Chemometrics: A Journal of the Chemometrics Society* 15(4), 413-426.
- Nisbet, I.C. and Lagoy, P.K. 1992. Toxic equivalency factors (TEFs) for polycyclic aromatic hydrocarbons (PAHs). *Regulatory toxicology and pharmacology* 16(3), 290-300.
- Phuong, J., Sipes, N., Truong, L., Connors, K., Houck, K. and Martin, M. 2014. ToxCast assay annotation version 1.0 data user guide. US Environmental Protection Agency, 36.
- Post, V., van der Grift, B., van der Schans M. (2024), Verslag van veldmetingen en historische meetgegevens over afbraak van organische microverontreinigingen in grondwater, BTO 2024.013, KWR Water Research Institute, Nieuwegein, The Netherlands.
- Posthuma, L., De Zwart, D., Keijzers, R. and Postma, J. 2016. Ecologische Sleutelfactor Toxiciteit. Deel 2: Kalibratie: toxische druk en ecologische effecten op macrofauna. STOWA rapport (2016-15).
- Reus, A., Dingemans, M., Hoondert, R., Houtman, C. and van der Oost, R. 2022. QSAR en read-across-modellen voor waterkwaliteit. H2O-online.
- Rorije, E., Wassenaar, P.N., Slootweg, J., van Leeuwen, L., van Broekhuizen, F.A. and Posthuma, L. 2022. Characterization of ecotoxicological risks from unintentional mixture exposures calculated from European freshwater monitoring data: Forwarding prospective chemical risk management. *Science of The Total Environment*, 153385.
- Roßbach, P. 2018. Neural Networks vs. Random Forests—Does it always have to be Deep Learning? Germany: Frankfurt School of Finance and Management.
- Sarstedt, M., Ringle, C.M. and Hair, J.F. (2021) Handbook of market research, pp. 587-632, Springer.
- Stucki, A.O., Barton-Maclaren, T.S., Bhuller, Y., Henriquez, J.E., Henry, T.R., Hirn, C., Miller-Holt, J., Nagy, E.G., Perron, M.M., Ratzlaff, D.E., Stedeford, T.J. and Clippinger, A.J. 2022. Use of new approach methodologies (NAMs) to meet regulatory requirements for the assessment of industrial chemicals and pesticides for effects on human health. *Frontiers in Toxicology* 4.
- Tian, X., Seshan, S., Wols B.A. (2024), , Zeer Zorgwekkende Stoffen (deel 3) – literature mining, BTO 2024.012, KWR Water Research Institute, Nieuwegein, The Netherlands.
- Timmer H. en Bannink (2023), Op weg naar Europese regels om bronnen van drinkwater te beschermen tegen zeer zorgwekkende stoffen, H2O Podium, 23 november 2023, [Op weg naar Europese regels om bronnen van drinkwater te beschermen tegen Zeer Zorgwekkende Stoffen \(h2owaternetwerk.nl\)](https://www.h2owaternetwerk.nl)
- U.S. EPA 2015 ToxCast & Tox21 Summary Files from invitrodb_v3.
- Van den Berg, M., Birnbaum, L., Bosveld, A.T., Brunström, B., Cook, P., Feeley, M., Giesy, J.P., Hanberg, A., Hasegawa, R., Kennedy, S.W., Kubiak, T., Larsen, J.C., van Leeuwen, F.X., Liem, A.K., Nolt, C., Peterson, R.E., Poellinger, L., Safe, S., Schrenk, D., Tillitt, D., Tysklind, M., Younes, M., Waern, F. and Zacharewski, T. 1998. Toxic equivalency factors (TEFs) for PCBs, PCDDs, PCDFs for humans and wildlife. *Environ Health Perspect* 106(12), 775-792.
- Verhaar, H.J., Van Leeuwen, C.J. and Hermens, J.L. 1992. Classifying environmental pollutants. *Chemosphere* 25(4), 471-491.
- Wassenaar, P.N.H., Rorije, E., Vijver, M.G. and Peijnenburg, W.J.G.M. 2021. Evaluating chemical similarity as a measure to identify potential substances of very high concern. *Regulatory Toxicology and Pharmacology* 119, 104834.
- Wols, B.A., Siegers, W., Vries D. (2024), Zeer Zorgwekkende Stoffen (deel 2) – zuivering, BTO 2024.011, KWR Water Research Institute, Nieuwegein, The Netherlands.
- Yang, H., Lou, C., Li, W., Liu, G. and Tang, Y. 2020. Computational Approaches to Identify Structural Alerts and Their Applications in Environmental Toxicology and Drug Discovery. *Chemical Research in Toxicology* 33(6), 1312-1322.
- Zeilmaker, M., Fragki, S., Verbruggen, E., Bokkers, B. and Lijzen, J. 2018. Mixture exposure to PFAS: A relative potency factor approach.

Ministerie van Infrastructuur en Waterstaat (z.d.), ZZS-beleid, geraadpleegd op 9-11-2023,

<https://www.infomil.nl/onderwerpen/lucht-water/zeer-zorgwekkende/zzs-beleid/>

Rijksinstituut voor Volksgezondheid en Milieu (z.d.), Zeer zorgwekkende Stoffen, geraadpleegd op 9-11-2023,

<https://rvs.rivm.nl/onderwerpen/zeer-zorgwekkende-stoffen>

RIWA-Rijn, 2023, Jaarrapport 2022, De Rijn, Vereniging van rivierwaterbedrijven. <https://www.riwa-rijn.org/wp-content/uploads/2023/09/RIWA-2023-NL-Jaarrapport-2022-De-Rijn.pdf>

Kools, S.A.E, Roskam, G.D., Verheul, M.R.A., Pieters, B.J., 2013, MRI contrast media, Magnetic Resonance Imaging (MRI) contrast media in het aquatisch milieu, Vereniging van rivierwaterbedrijven, https://www.riwa-rijn.org/wp-content/uploads/2015/04/RIWA_Rapport_MRI_POD_01.pdf

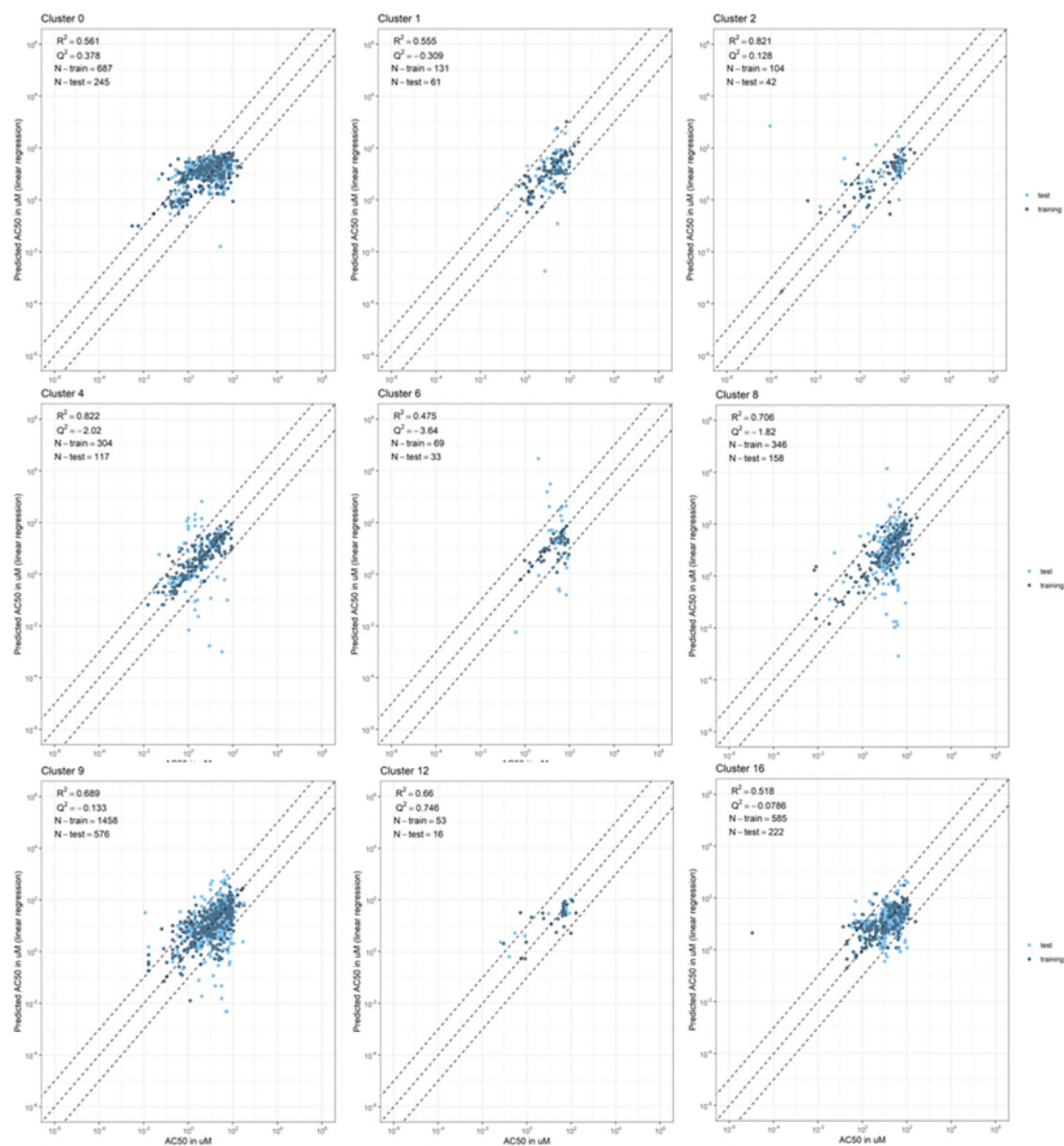
RDKit (z.d.), RDKit: Open-source cheminformatics. <https://www.rdkit.org>, geraadpleegd op 9-11-2023.

US EPA, 2012. Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.11. United States Environmental Protection Agency, Washington, DC, USA.

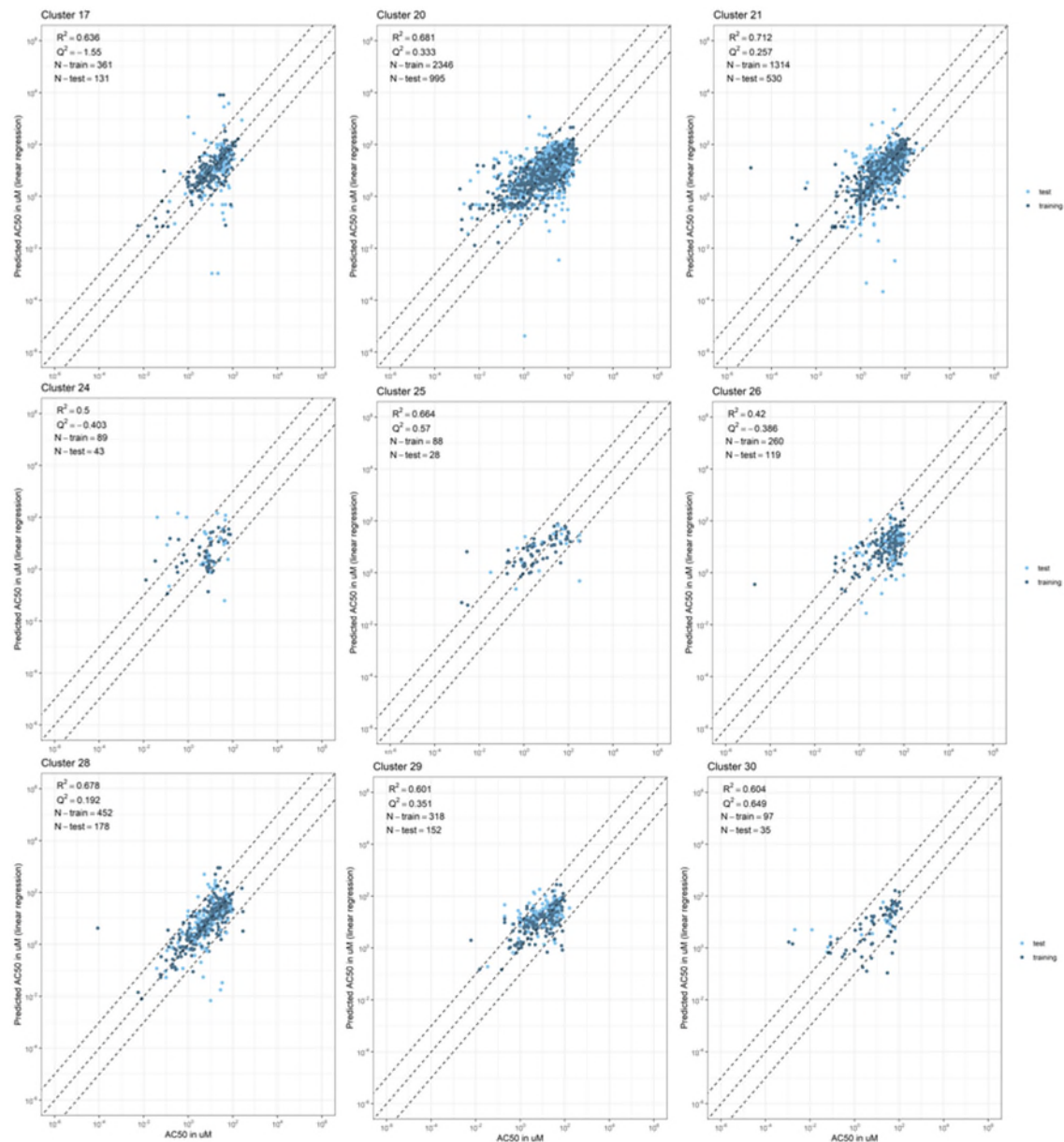
Wassenaar, P. N. H., Rorije, E., Vijver, M. G., Peijnenburg, W. J. G. M., 2022, ZZS similarity tool: The online tool for similarity screening to identify chemicals of potential concern, J. Comput. Chem. 2022, 43(15), 1042. <https://doi.org/10.1002/jcc.26859>

Osté, L.A., Ouwerkerk, K., Derksen, A., 2022, NORMAN prioritering Nederlandse Waterkwaliteitsdata, Deltares, https://publications.deltares.nl/11206216_010_0001.pdf

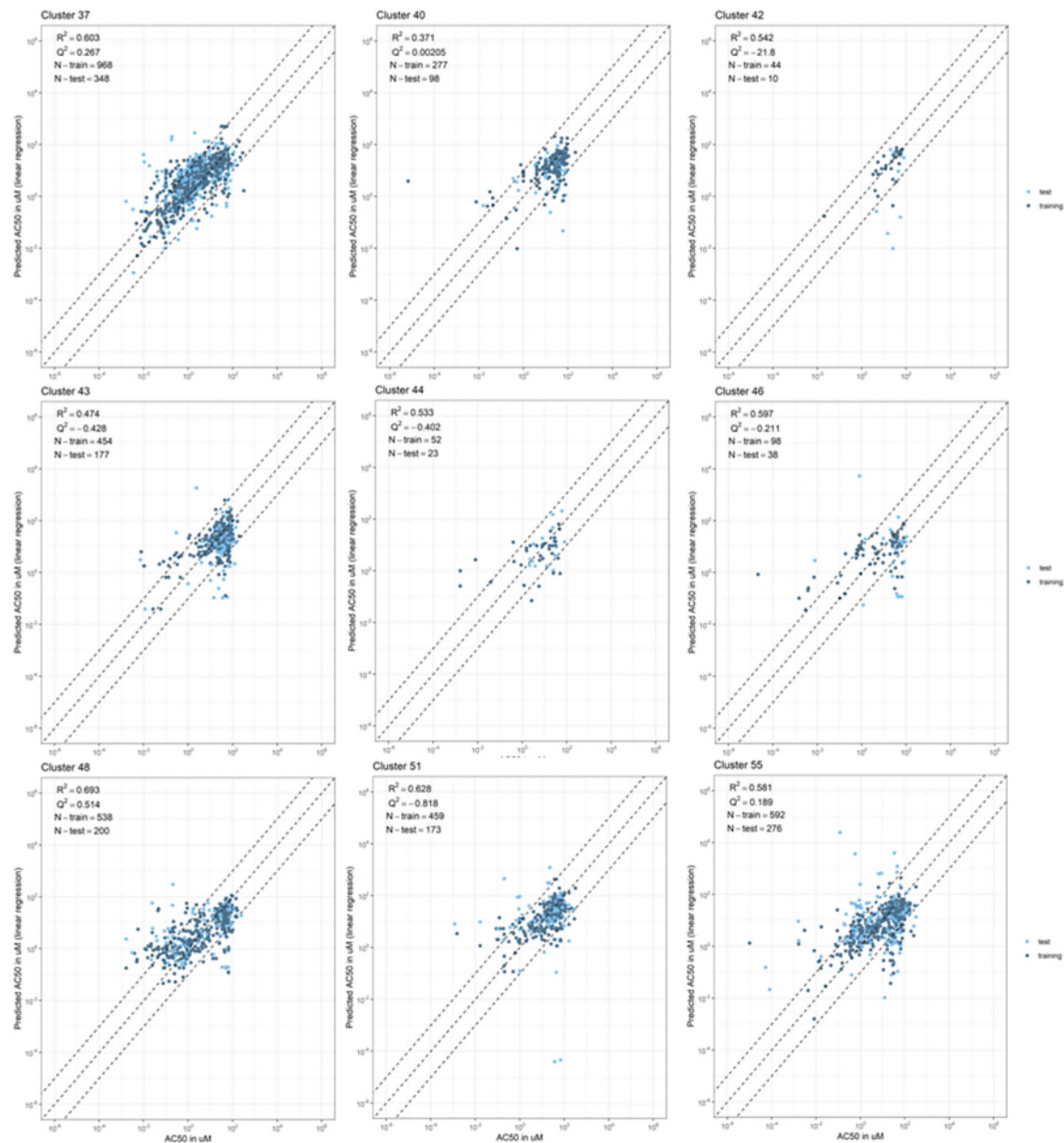
I Appendix: Voorspellingsplots voor individuele clusters



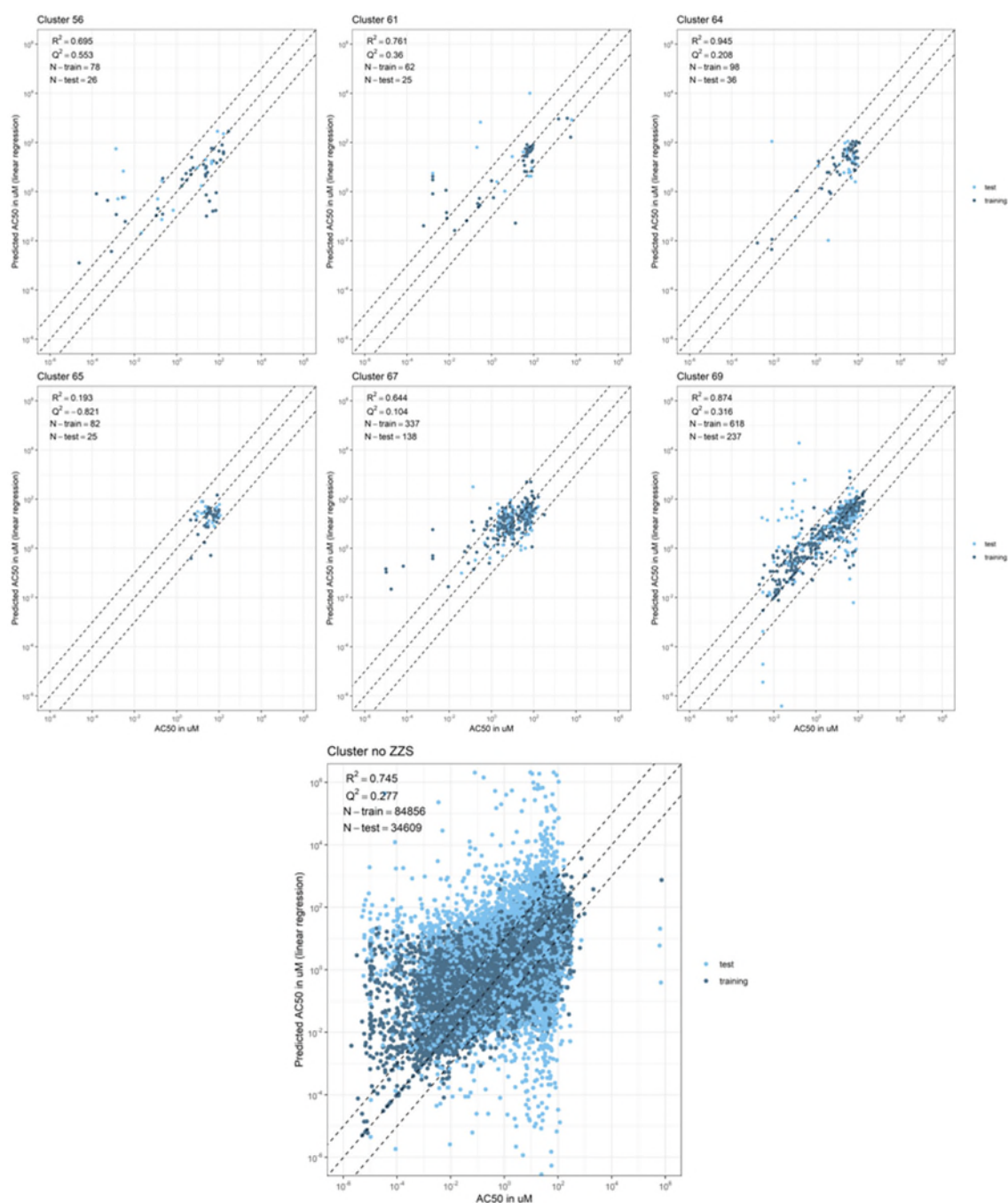
Figuur 18-a: Voorspelde toxiciteit (AC_{50} in μM) door het meervoudige lineaire regressiemodel versus waargenomen toxiciteit, gebaseerd op functionele groepen, geclusterd per assaytype (gebaseerd op beoogde doelfamilie), voor zowel de training- en testdataset, per ZZS-cluster.



Figuur 29-b: Voorspelde toxiciteit (AC_{50} in μM) door het meervoudige lineaire regressiemodel versus waargenomen toxiciteit, gebaseerd op functionele groepen, geclusterd per assaytype (gebaseerd op beoogde doelfamilie), voor zowel de training- en testdataset, per ZZS-cluster.



Figuur 29-c: Voorspelde toxiciteit (AC_{50} in μM) door het meervoudige lineaire regressiemodel versus waargenomen toxiciteit, gebaseerd op functionele groepen, geclusterd per assaytype (gebaseerd op beoogde doelfamilie), voor zowel de training- en testdataset, per ZZS-cluster.



Figuur 29-d: Voorspelde toxiciteit (AC_{50} in μM) door het meervoudige lineaire regressiemodel versus waargenomen toxiciteit, gebaseerd op functionele groepen, geclusterd per assaytype (gebaseerd op beoogde doelfamilie), voor zowel de training- en testdataset, per ZZS-cluster.

II Appendix: Clustering in vitro assayeindpunten

II.1 Beoogde doelfamilie (*intended target type*)

De beoogde doelfamilie geeft de individuele doelwitten voor alle assay-eindpunten weer. Deze families hebben betrekking op genfamilies en omvatten morfologische en celcyclusconcepten (U.S. EPA, 2015).

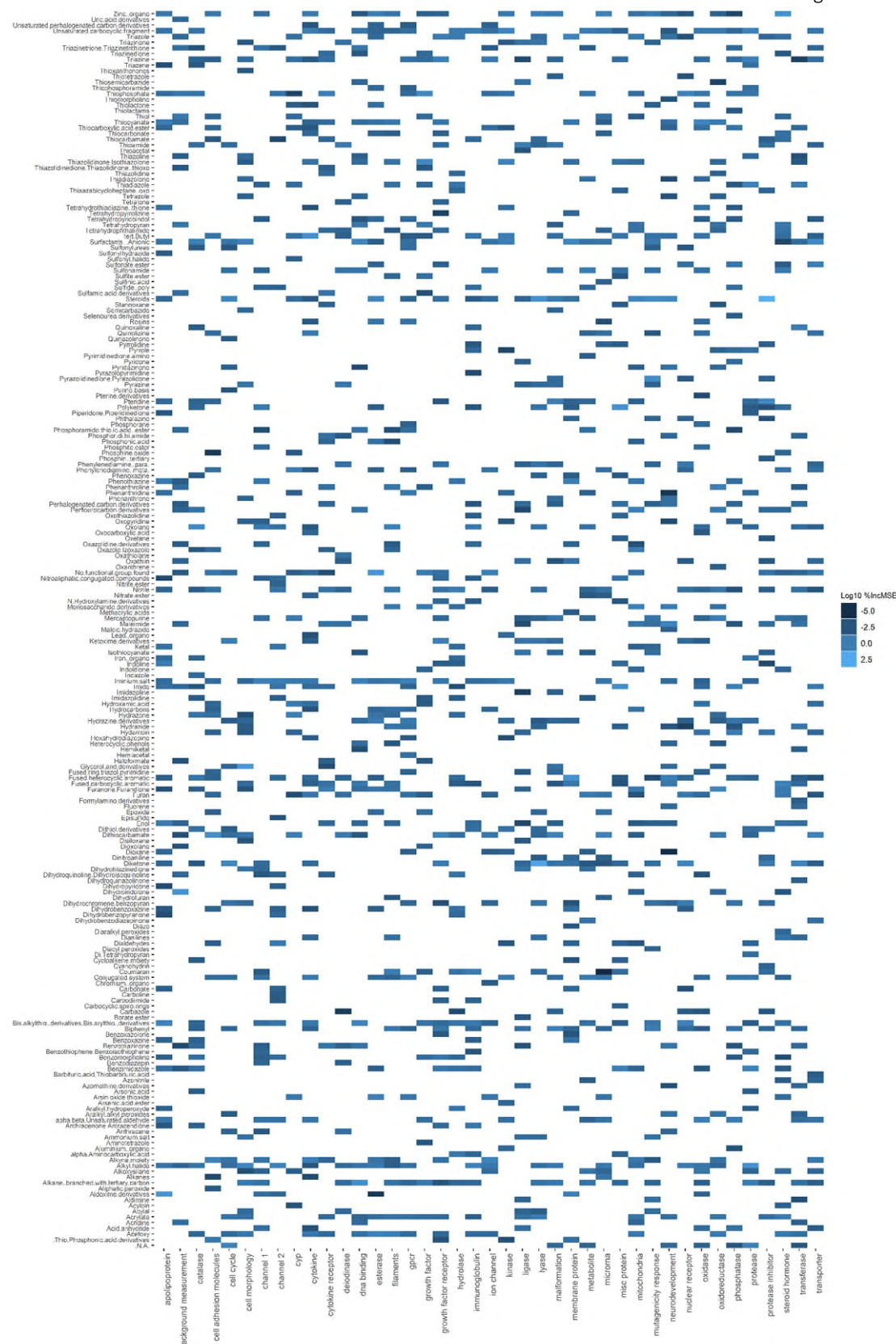
Random Forest model

Over het algemeen werd bij het groeperen van *in vitro* assay's op basis van de beoogde doelfamilie het hoogste percentage verklaarde varianties bepaald voor *in vitro* assay's gerelateerd aan nucleair receptors (50,84%), terwijl het laagste % variantie verklaard door het Random Forest-model werd gevonden voor *in vitro* assay's gerelateerd aan membraaneiwwitten (- 4,96%). Dit impliceert dat molecuulstructuren betere voorspellers zijn voor uitkomsten van toxiciteitstesten gerelateerd aan nucleair receptors dan voor testen gerelateerd aan eiwitmembranen. In Figuur 29 is een heatmap te zien die visualiseert in welke mate de functionele groepen correleren met toxiciteit voor assays binnen een van de beoogde doelfamilies. De toename in MSE (%IncMSE) (vergelijking 1) komt overeen met de mate waarin de functionele groepen de variantie in het Random Forest-model verklaren.

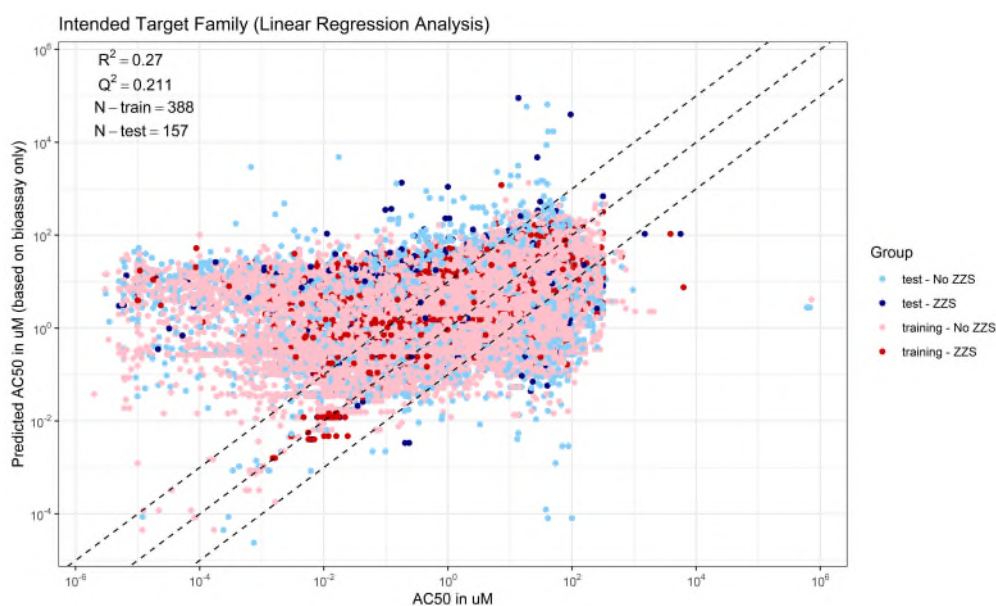
Lineaire regressiemodel

Over het algemeen resulteerde de meervoudige regressieanalyse, waarbij functionele groepen als verklarende variabelen worden meegenomen in een verklaring van 26,1% (mediaan: 26.1%, S.E.: 0,06) van alle variantie in de toxiciteitsgegevens (AC_{50} 's) bij het categoriseren van *in vitro* assay's op basis van de beoogde doelfamilie. Over het algemeen werd bij het groeperen van *in vitro* assay's op basis van de beoogde doelfamilie het hoogste percentage verklaarde varianties in de testdata bepaald voor *in vitro* assay's gerelateerd aan neurologische ontwikkeling (87,84%), terwijl het laagste % variantie verklaard door het regressiemodel werd gevonden voor assays gerelateerd aan membraaneiwwitten (- 23,18%). Dit laat zien dat er grote verschillen zitten tussen de voorspelbaarheid van toxiciteit binnen de verschillende doelfamilies. Figuur 30 laat de voorspelde effectconcentraties ($\log_{10} AC_{50}$ s) zien, uitgezet tegen de waargenomen effectconcentraties, gebaseerd op het meervoudige lineaire regressiemodel, met de functionele groepen als verklarende variabelen (vergelijking 2), voor alle afzonderlijke beoogde doelfamilies afzonderlijk. In totaal lag 69,83% van alle individueel voorspelde AC_{50} 's binnen een factor 5 van de waargenomen AC_{50} 's; 15.7% van de voorspelde datapunten lag meer dan een factor vijf onder de waargenomen datapunten (onderschat), terwijl 14.4% van alle datapunten meer dan een factor vijf boven de waargenomen data lag (overschat). Geen van de voorspelde datapunten waren identiek aan de waargenomen datapunten. Over het algemeen zijn er geen grote verschillen zichtbaar tussen de voorspelbaarheid van AC_{50} -waarden in de trainingsdataset (waarop het model is gebaseerd) en de testdataset (met stoffen waarop het model niet is getraind). 70,4% van de voorspelde datapunten binnen de trainingsdataset lag binnen een factor vijf van de waargenomen waarden, tegenover 68.5% van de voorspelde datapunten in de testdataset (Figuur 28). Ook is de gemiddelde *adjusted* R^2 (0.27) van alle modellen getraind op de trainingsdataset vergelijkbaar met de gemiddelde Q^2 (0.21) wanneer de modellen worden toegepast op de testdataset. Ook zijn er geen grote verschillen zichtbaar tussen de voorspelbaarheid van ZZS ten opzichte van niet-ZZS, hoewel de verschillen in voorspelbaarheid wel groter zijn tussen de training- en testdataset bij niet-ZZS. 72,9% van de voorspelde datapunten binnen de ZZS-data lag binnen een factor vijf van de waargenomen waarden, tegenover 69.5% van de voorspelde datapunten in de niet-ZZS-data (Figuur 28). De ZZS hebben mogelijk een heel specifieke soort toxiciteit en de clusters zijn de manier om dit te onderscheiden. Deze clusters moeten dan gelinkt worden aan een specifiek bioassaytype dat hier geschikt voor is. De ZZS-data beslaan echter een stuk minder datapunten en de stoffen in deze data zijn beperkt tot stoffen met een relatief lage AC_{50} . Dit betekent dat de

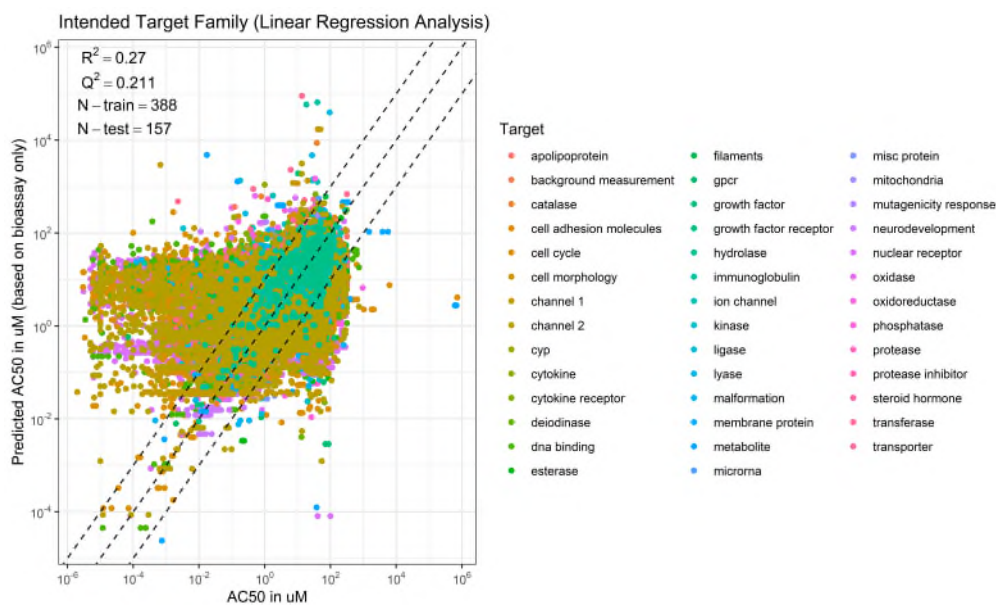
stoffen een minder breed toxiciteitsdomein hebben en daarom wellicht eenvoudiger te voorspellen zijn.



Figuur 19: Heatmap die de toename in MSE (Mean Squared Error) visualiseert (= in welke mate de variabele de variantie in toxiciteit in het Random Forest-model verklaart) voor de functionele groepen (y-as), bij het groeperen van de in vitro toxiciteitstesten op basis van de beoogde doelfamilie (x-as).



Figuur 20: Voorspelde toxiciteit (AC_{50} in μM) door het meervoudige lineaire regressiemodel versus waargenomen toxiciteit, gebaseerd op functionele groepen, geclusterd per assaytype (gebaseerd op beoogde doelfamilie), voor zowel de training- en testdataset, als stoffen die als ZZS staan aangemerkt en stoffen die niet als ZZS staan aangemerkt.



Figuur 21: Voorspelde toxiciteit (AC_{50} in μM) door het meervoudige lineaire regressiemodel versus waargenomen toxiciteit, gebaseerd op functionele groepen, geclusterd per beoogde doelfamilie.

II.II Technologisch doeltype (*technological target type*)

Het technologische doeltype probeert de individuele doelen voor alle assay-eindpunten weer te geven. Deze families hebben betrekking op genfamilies en omvatten morfologische en celcyclusconcepten (U.S. EPA, 2015).

Random Forest model

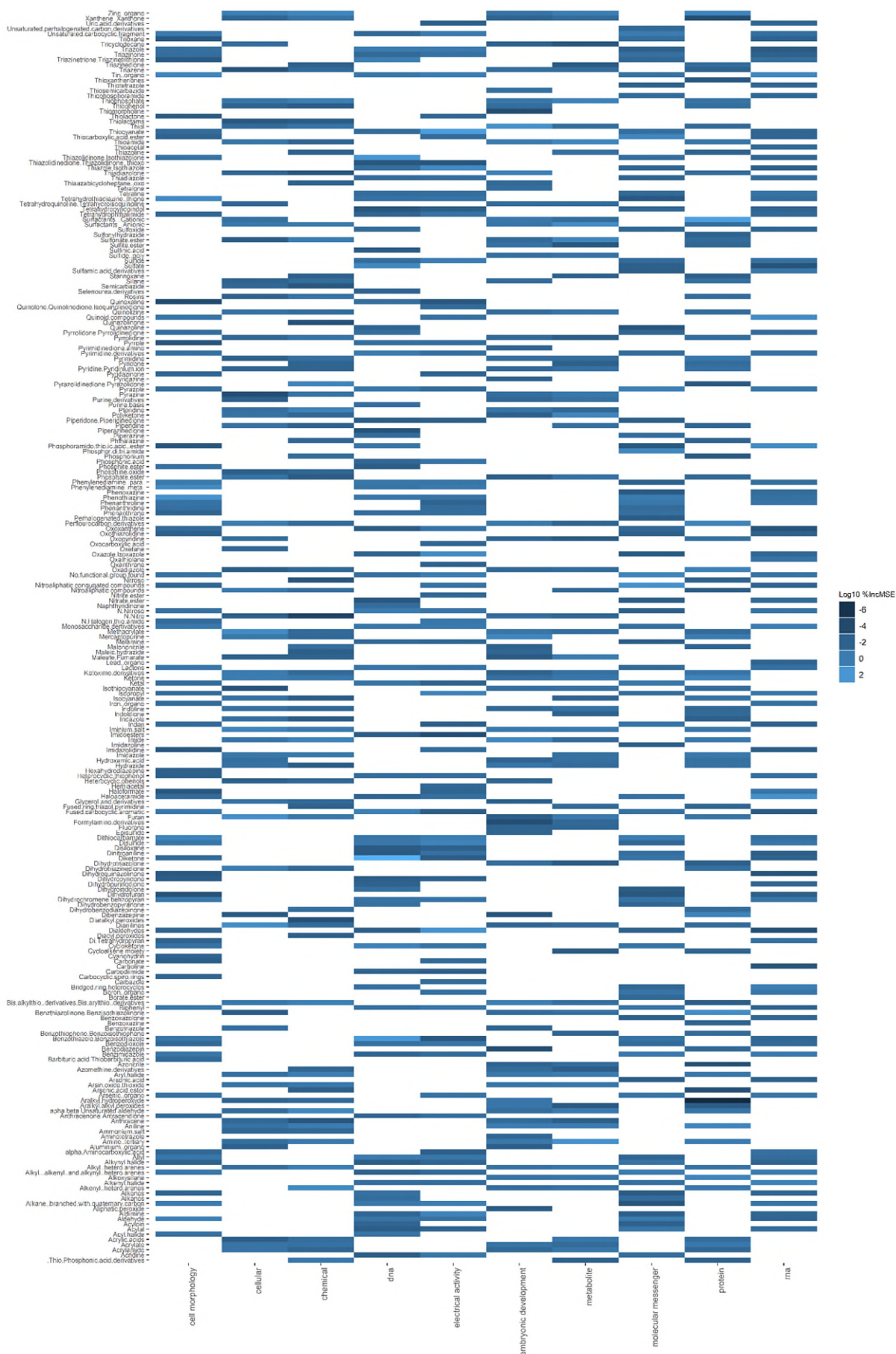
Over het algemeen werd bij het groeperen van *in vitro* toxiciteitstesten op basis van het technologische doeltype het hoogste percentage verklaarde varianties bepaald voor *in vitro* toxiciteitstesten met een chemisch technologische doeltype (6,45%), terwijl het laagste % variantie verklaard door het Random Forest-model werd gevonden voor *in vitro* toxiciteitstesten met een metabolisch technologisch doeltype (-0,75%). In Figuur 31 is een heatmap te zien die visualiseert in welke mate de functionele groepen correleren met toxiciteit voor toxiciteitstesten binnen een van de beoogde technologische doeltypes. De toename in MSE (%IncMSE) (vergelijking 1) komt overeen met de mate waarin de functionele groepen de variantie in het Random Forest-model verklaren.

Lineair regressiemodel

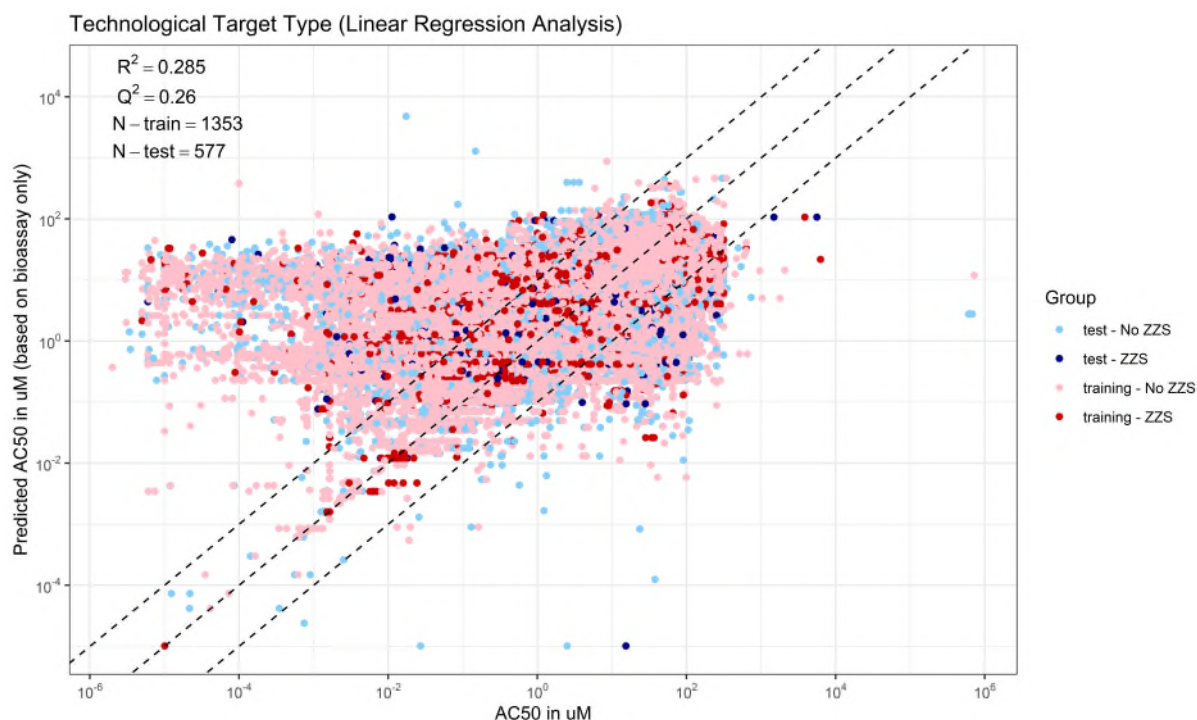
Over het algemeen resulteerde de meervoudige regressieanalyse, waarbij functionele groepen als verklarende variabelen worden meegenomen in een verklaring van 39,2% (mediaan: 36.7%, S.E.: 0,05) van alle variantie in de toxiciteitsgegevens (AC_{50} 's) bij het categoriseren van *in vitro* toxiciteitstesten op basis van het technologische doeltype. Over het algemeen werd bij het groeperen van *in vitro* toxiciteitstesten op basis van het technologisch doeltype het hoogste percentage verklaarde varianties bepaald voor *in vitro* toxiciteitstesten met een type gerelateerd aan elektrische activiteit (87,84%), terwijl het laagste % variantie verklaard door het regressiemodel werd gevonden voor testen gerelateerd aan DNA (16%). Dit laat zien dat er grote verschillen zitten tussen de voorspelbaarheid van toxiciteit binnen de verschillende technologische doeltypen, hoewel deze verschillen wel lager zijn wanneer de *in vitro* testen worden gecategoriseerd op basis van technologisch doeltype.

Figuur 14 laat de voorspelde effectconcentraties ($\log_{10} AC_{50}$ s) zien, uitgezet tegen de waargenomen effectconcentraties, gebaseerd op het meervoudige lineaire regressiemodel, met de functionele groepen als verklarende variabelen (vergelijking 2), voor alle afzonderlijke beoogde doelfamilies afzonderlijk. In totaal lag 69,77% van alle individueel voorspelde AC_{50} 's binnen een factor 5 van de waargenomen AC_{50} 's; 15% van de voorspelde datapunten lag meer dan een factor vijf onder de waargenomen datapunten (onderschat), terwijl 15.1% van alle datapunten meer dan een factor vijf boven de waargenomen data lag (overschat). Geen van de voorspelde datapunten waren identiek aan de waargenomen datapunten.

Over het algemeen zijn er geen grote verschillen zichtbaar tussen de voorspelbaarheid van AC_{50} -waarden in de trainingsdataset (waarop het model is gebaseerd) en de testdataset (met stoffen waarop het model niet is getraind). 70% van de voorspelde datapunten binnen de trainingsdataset lag binnen een factor vijf van de waargenomen waarden, tegenover 69.2% van de voorspelde datapunten in de testdataset (Figuur 28). Ook is de gemiddelde *adjusted* R^2 (0.285) van alle modellen getraind op de trainingsdataset vergelijkbaar met de gemiddelde Q^2 (0.26) wanneer de modellen worden toegepast op de testdataset. Ook zijn er geen grote verschillen zichtbaar tussen de voorspelbaarheid van ZZS ten opzichte van niet-ZZS, hoewel deze verschillen wel groter zijn dan tussen de training- en testdataset. 71,8% van de voorspelde datapunten binnen de ZZS-data lag binnen een factor vijf van de waargenomen waarden, tegenover 69.4% van de voorspelde datapunten in de niet-ZZS-data (Figuur 32). De ZZS-data beslaan echter een stuk minder datapunten en de stoffen in deze data zijn beperkt tot stoffen met een relatief lage AC_{50} . Dit betekent dat de stoffen een minder breed toxiciteitsdomein hebben en daarom wellicht eenvoudiger te voorspellen zijn.



Figuur 22: Heatmap die de toename in MSE (Mean Squared Error) visualiseert (= in welke mate de variabele de variantie in toxiciteit in het Random Forest-model verklaart) voor de functionele groepen (y-as), bij het groeperen van de in vitro toxiciteitstesten op basis van het technologisch doeltype (x-as).



Figuur 23: Voorspelde toxiciteit (AC_{50} in μM) door het meervoudige lineaire regressiemodel versus waargenomen toxiciteit, gebaseerd op functionele groepen, geclusterd per assaytype (gebaseerd op technologisch doeltype), voor zowel de training- en testdataset, als stoffen die als ZS staan aangemerkt en stoffen die niet als ZS staan aangemerkt.



Figuur 24: Voorspelde toxiciteit (AC_{50} in μM) door het meervoudige lineaire regressiemodel versus waargenomen toxiciteit, gebaseerd op functionele groepen, geclusterd per technologisch doeltype.

II.III Assayontwerptype (*assay design type*)

Het ontwerptype van de assay vertegenwoordigt de methode waarmee een biologisch of fysisch proces wordt vertaald naar een detecteerbaar signaal (U.S. EPA, 2015). De annotatie van het assayontwerptype legt de methode vast waarmee het technologische doel wordt gemeten.

Random Forest model

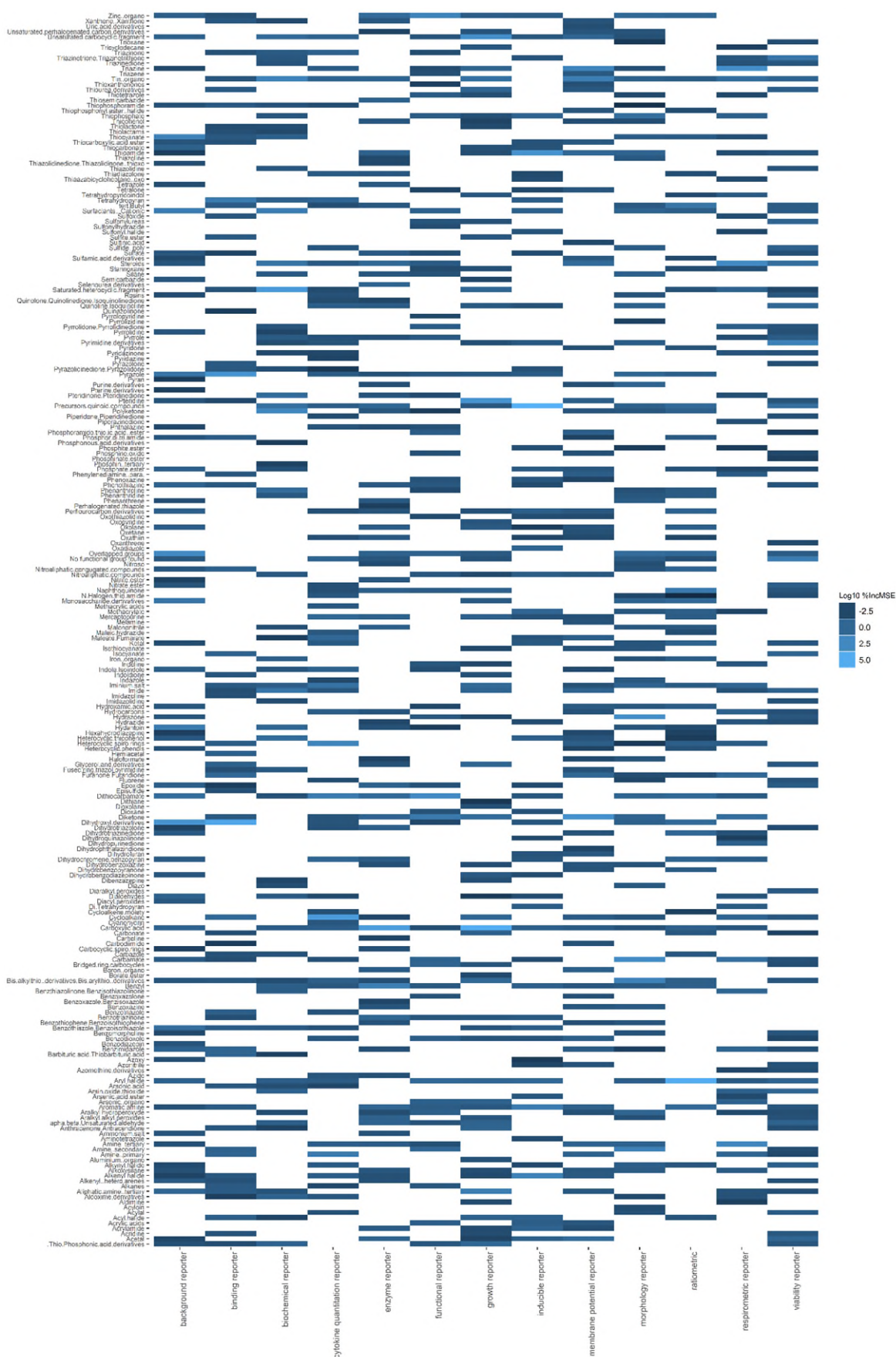
Over het algemeen werd bij het groeperen van assays op basis van het assayontwerptype het hoogste percentage verklaarde varianties bepaald voor assays met een *background reporter* assayontwerptype (5,53%), terwijl het laagste % variantie verklaard door het Random Forest-model werd gevonden voor toxiciteitstesten met een ratiometrisch assayontwerptype (- 2,89%). In Figuur 16 is een heatmap te zien die visualiseert in welke mate de functionele groepen correleren met toxiciteit voor testen binnen een van de assayontwerptypes. De toename in MSE (%IncMSE) (vergelijking 1) komt overeen met de mate waarin de functionele groepen de variantie in het Random Forest-model verklaren.

Lineair regressiemodel

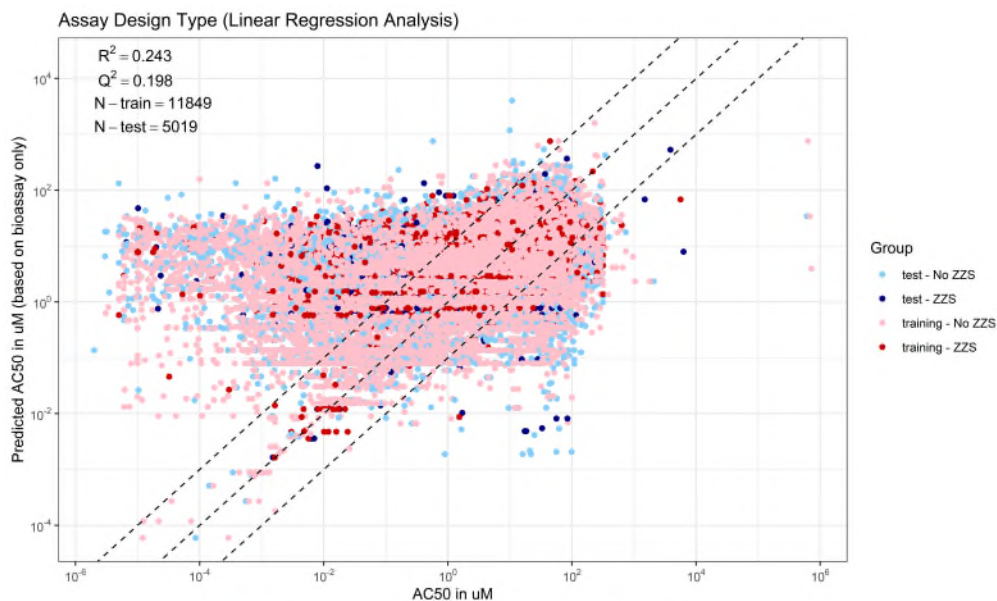
Over het algemeen resulteerde de meervoudige regressieanalyse, waarbij functionele groepen als verklarende variabelen worden meegenomen in een verklaring van 37,1% (mediaan: 30.5%, S.E.: 0,05) van alle variantie in de toxiciteitsgegevens (AC_{50} 's) bij het categoriseren van *in vitro* assays op basis van het technologische doeltype. Over het algemeen werd bij het groeperen van *in vitro* assay's op basis van het assayontwerptype het hoogste percentage verklaarde varianties bepaald voor *in vitro* assay's met een *functional reporter* ontwerptype (87,87%), terwijl het laagste % variantie verklaard door het regressiemodel werd gevonden voor assay's met een *enzyme reporter* ontwerptype (14.9%) (Tabel 1). Dit laat zien dat er grote verschillen zitten tussen de voorspelbaarheid van toxiciteit binnen de verschillende assayontwerptypen, hoewel deze verschillen wel lager zijn wanneer de *in vitro* toxiciteitstesten worden gecategoriseerd op basis van ontwerptype.

Figuur x laat de voorspelde effectconcentraties ($\log_{10} AC_{50}$ s) zien, uitgezet tegen de waargenomen effectconcentraties, gebaseerd op het meervoudige lineaire regressiemodel, met de functionele groepen als verklarende variabelen (vergelijking 1), voor alle afzonderlijke beoogde doelfamilies afzonderlijk. In totaal lag 68,92% van alle individueel voorspelde AC_{50} 's binnen een factor 5 van de waargenomen AC_{50} 's; 5,9% van de voorspelde datapunten lag meer dan een factor vijf onder de waargenomen datapunten (onderschat), terwijl 29,5% van alle datapunten meer dan een factor vijf boven de waargenomen data lag (overschat). Geen van de voorspelde datapunten waren identiek aan de waargenomen datapunten.

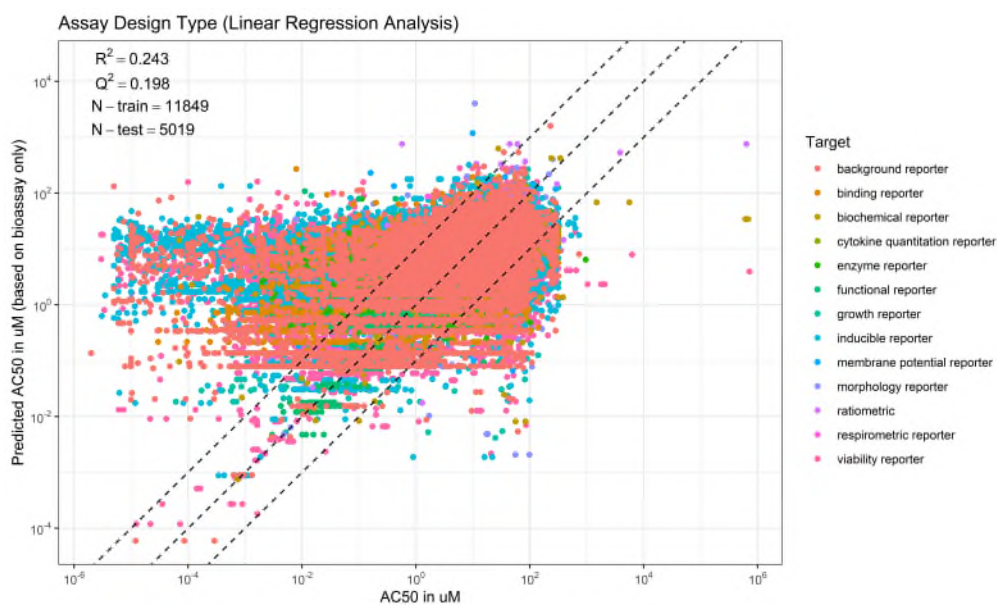
Over het algemeen zijn er geen grote verschillen zichtbaar tussen de voorspelbaarheid van AC_{50} -waarden in de trainingsdataset (waarop het model is gebaseerd) en de testdataset (met stoffen waarop het model niet is getraind). 69,2% van de voorspelde datapunten binnen de trainingsdataset lag binnen een factor vijf van de waargenomen waarden, tegenover 68,2% van de voorspelde datapunten in de testdataset (Figuur 9). Ook is de gemiddelde *adjusted* R^2 (0.243) van alle modellen getraind op de trainingsdataset vergelijkbaar met de gemiddelde Q^2 (0.198) wanneer de modellen worden toegepast op de testdataset. Ook zijn er geen grote verschillen zichtbaar tussen de voorspelbaarheid van ZZS ten opzichte van niet-ZZS, hoewel deze verschillen wel groter zijn dan tussen de training- en testdataset. 71,6% van de voorspelde datapunten binnen de ZZS-data lag binnen een factor vijf van de waargenomen waarden, tegenover 68.5% van de voorspelde datapunten in de niet-ZZS-data (Figuur 9). De ZZS-data beslaan echter een stuk minder datapunten en de stoffen in deze data zijn beperkt tot stoffen met een relatief lage AC_{50} . Dit betekent dat de stoffen een minder breed toxiciteitsdomein hebben en daarom wellicht eenvoudiger te voorspellen zijn.



Figuur 25: Heatmap die de toename in MSE (Mean Squared Error) visualiseert (= in welke mate de variabele de variantie in toxiciteit in het Random Forest-model verklaart) voor de functionele groepen (y-as), bij het groeperen van de in vitro assays op basis van de assayontwerptype (x-as).



Figuur 26: Voorspelde toxiciteit (AC_{50} in μM) door het meervoudige lineaire regressiemodel versus waargenomen toxiciteit, gebaseerd op functionele groepen, geclusterd per assayontwerptype (gebaseerd op assayontwerptype), voor zowel de training- en testdataset, als stoffen die als ZZS staan aangemerkt en stoffen die niet als ZZS staan aangemerkt.



Figuur 27: Voorspelde toxiciteit (AC_{50} in μM) door het meervoudige lineaire regressiemodel versus waargenomen toxiciteit, gebaseerd op functionele groepen, geclusterd per assayontwerptype.

II.IV Organisme en weefseltype (*model type*)

Random Forest model

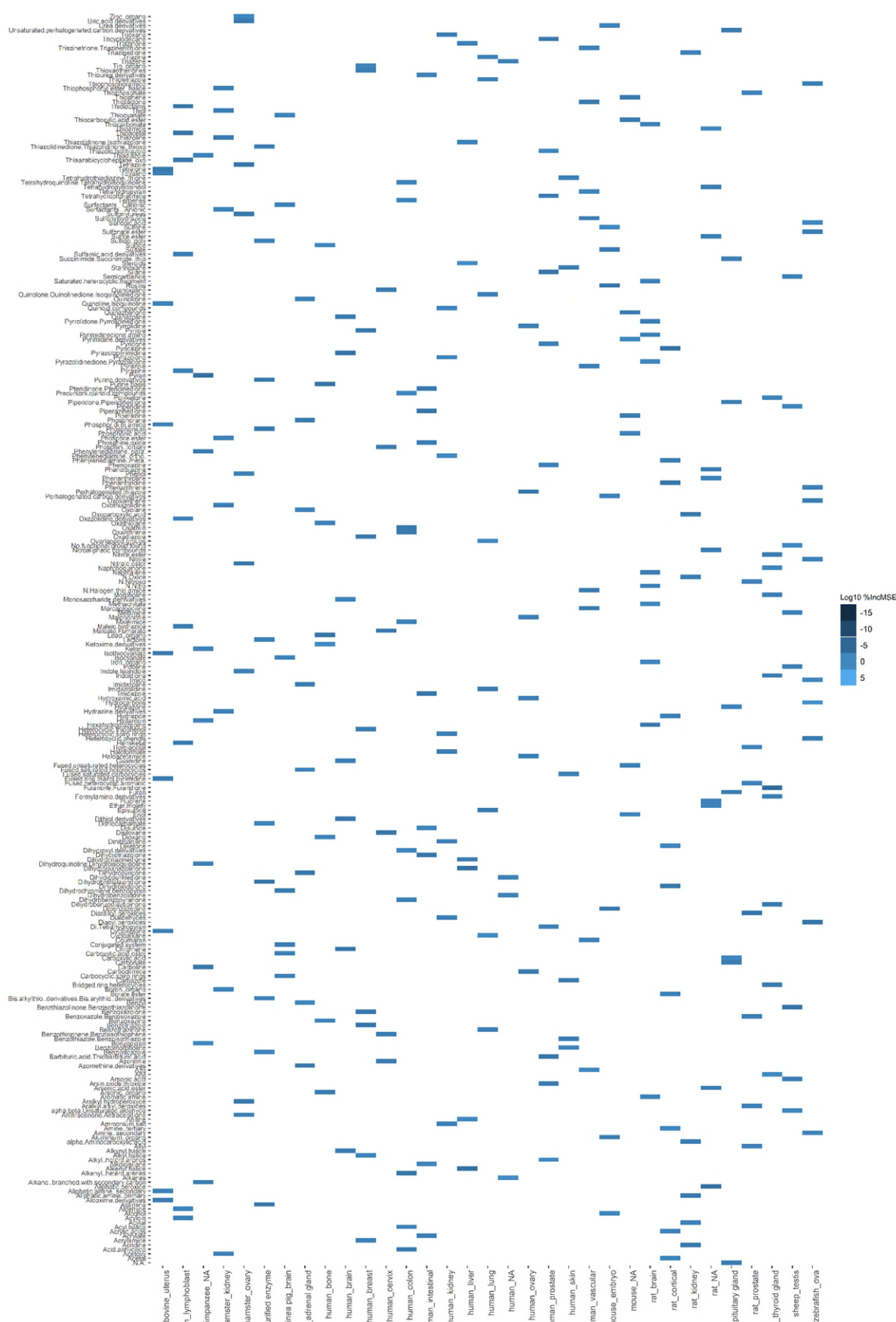
Over het algemeen werd bij het groeperen van assays op basis van het organisme en weefseltype het hoogste percentage verklaarde varianties bepaald voor assays met menselijke niercellen (5,77%), terwijl het laagste % variantie verklaard door het Random Forest-model werd gevonden voor assays met menselijke hersencellen (- 2,59%). In Figuur 37 is een heatmap te zien die visualiseert in welke mate de functionele groepen correleren met toxiciteit voor assays binnen één van de organisme-weefseltypecombinaties. De toename in MSE (%IncMSE) (vergelijking 1) komt overeen met de mate waarin de functionele groepen de variantie in het Random Forest-model verklaren.

Lineair regressiemodel

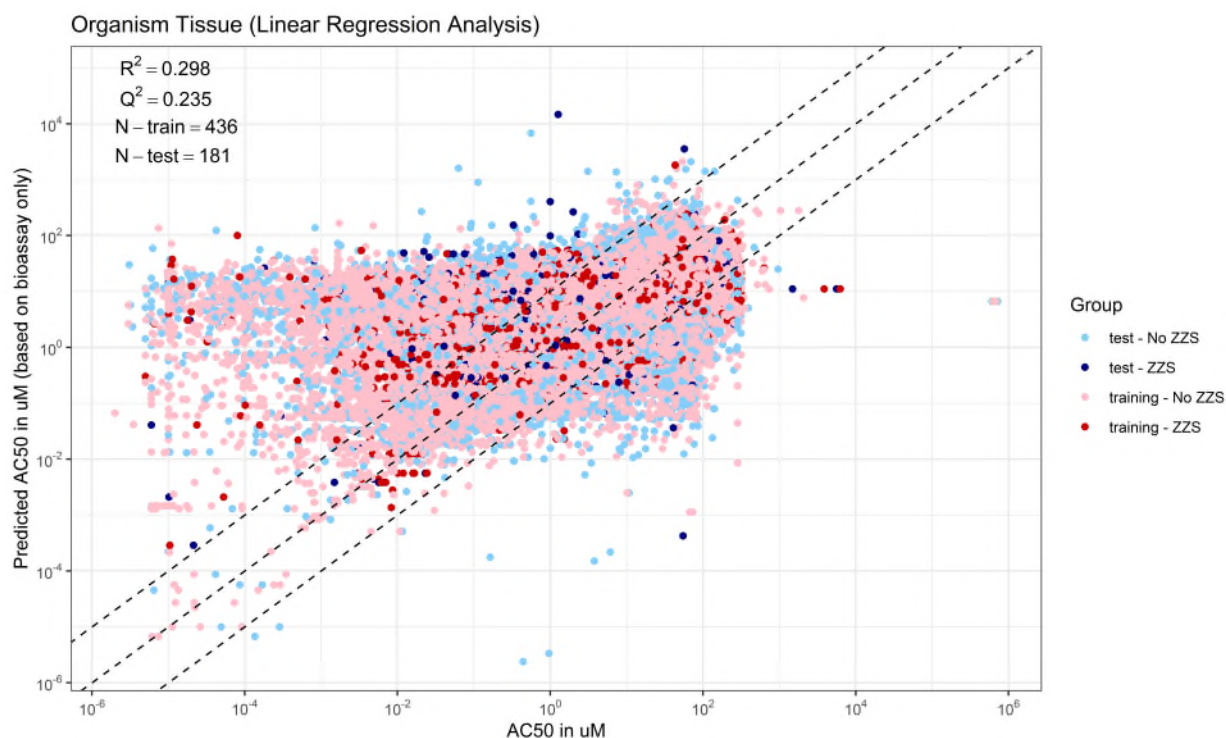
Over het algemeen resulteerde de meervoudige regressieanalyse, waarbij functionele groepen als verklarende variabelen worden meegenomen in een verklaring van 37,8% (mediaan: 34,9%, S.E.: 0,04%) van alle variantie in de toxiciteitsgegevens (AC_{50} 's) bij het categoriseren van *in vitro* toxiciteitstesten op basis van het technologische doeltype. Over het algemeen werd bij het groeperen van *in vitro* toxiciteitstesten op basis van organisme en weefseltype het hoogste percentage verklaarde varianties bepaald voor *in vitro* toxiciteitstesten met corticale cellen van ratten (85,44%), terwijl het laagste % variantie verklaard door het regressiemodel werd gevonden voor assay's met menselijke levercellen (14,6%). Dit laat zien dat er grote verschillen zitten tussen de voorspelbaarheid van toxiciteit tussen de verschillende celtypen, hoewel deze verschillen wel lager zijn wanneer de *in vitro* testen worden gecategoriseerd binnen diverse celtypen (organisme-weefseltypecombinatie).

Figuur 38 laat de voorspelde effectconcentraties ($\log_{10} AC_{50}$ s) zien, uitgezet tegen de waargenomen effectconcentraties, gebaseerd op het meervoudige lineaire regressiemodel, met de functionele groepen als verklarende variabelen (vergelijking 2), voor alle afzonderlijke beoogde doelfamilies afzonderlijk. In totaal lag 68,92% van alle individueel voorspelde AC_{50} 's binnen een factor 5 van de waargenomen AC_{50} 's; 5,9% van de voorspelde datapunten lag meer dan een factor vijf onder de waargenomen datapunten (onderschat), terwijl 29,5% van alle datapunten meer dan een factor vijf boven de waargenomen data lag (overschat). Geen van de voorspelde datapunten waren identiek aan de waargenomen datapunten.

Over het algemeen zijn er geen grote verschillen zichtbaar tussen de voorspelbaarheid van AC_{50} -waarden in de trainingsdataset (waarop het model is gebaseerd) en de testdataset (met stoffen waarop het model niet is getraind). 71,8% van de voorspelde datapunten binnen de trainingsdataset lag binnen een factor vijf van de waargenomen waarden, tegenover 70,2% van de voorspelde datapunten in de testdataset (Figuur 28). Ook is de gemiddelde *adjusted* R^2 (0.298) van alle modellen getraind op de trainingsdataset vergelijkbaar met de gemiddelde Q^2 (0.235) wanneer de modellen worden toegepast op de testdataset. Ook zijn er geen grote verschillen zichtbaar tussen de voorspelbaarheid van ZZS ten opzichte van niet-ZZS, hoewel deze verschillen wel groter zijn dan tussen de training- en testdataset. 73,8% van de voorspelde datapunten binnen de ZZS-data lag binnen een factor vijf van de waargenomen waarden, tegenover 70,9% van de voorspelde datapunten in de niet-ZZS-data (Figuur 39). De ZZS-data beslaan echter een stuk minder datapunten en de stoffen in deze data zijn beperkt tot stoffen met een relatief lage AC_{50} . Dit betekent dat de stoffen een minder breed toxiciteitsdomein hebben en daarom wellicht eenvoudiger te voorspellen zijn.



Figuur 28: Heatmap die de toename in MSE (Mean Squared Error) visualiseert (= in welke mate de variabele de variantie in toxiciteit in het Random Forest-model verklaart) voor de functionele groepen (y-as), bij het groeperen van de in vitro assays op basis van organisme - weefseltypecombinatie (x-as).



Figuur 29: Voorspelde toxiciteit (AC_{50} in μM) door het meervoudige lineaire regressiemodel versus waargenomen toxiciteit, gebaseerd op functionele groepen, geclusterd per assaytype (gebaseerd op organisme - weefseltypecombinatie), voor zowel de training- en testdataset, als stoffen die als ZZS staan aangemerkt en stoffen die niet als ZZS staan aangemerkt.



Figuur 30: Voorspelde toxiciteit (AC_{50} in μM) door het meervoudige lineaire regressiemodel versus waargenomen toxiciteit, gebaseerd op functionele groepen, geclusterd per organisme-weefseltypecombinatie.

III Appendix: Ontwikkelingen en beleidscontext cumulatie

Het RIVM beschrijft welke methoden kunnen worden gebruikt door overheden om de effecten van stoffenmengsels voor mens en milieu te kunnen inschatten (Bodar et al., 2022). Ook wordt er gekeken naar hoe andere landen binnen de EU (Denemarken, België (Vlaanderen) en Duitsland) omgaan met mengseltoxiciteit.

Cumulatie bij de emissie van zeer zorgwekkende stoffen (ZZS) wordt momenteel niet meegenomen bij vergunningverlening. Dit maakt dat de gezondheidkundige risico's voor omwonenden momenteel onvoldoende kunnen worden ingeschat en onderschatting niet kan worden uitgesloten. Cumulatie is de optelling van milieu- en gezondheidsrisico's door gelijktijdige blootstelling aan verschillende chemische stoffen, waaronder ZZS. Het is echter lastig om deze effecten door cumulatie te voorspellen, aangezien veel verschillende, potentieel toxische, stoffen uit veel verschillende bronnen momenteel in het milieu aanwezig zijn, welke via verschillende blootstellingsroutes (huidcontact, inademing, inname via voedsel en drinkwater) in de mens terechtkomen. Dit maakt het een complexe vraagstelling. In de RIVM-studie wordt daarom nog niet naar de geaggregeerde blootstelling van de mens aan mengsels gekeken, maar focust deze zich eerst enkel op de manier waarop mengsels milieucompartimenten kunnen bereiken. Hierbij is een sterke focus op "onbedoelde" mengsels van stoffen die vrijkomen bij vergunde milieu-emissies bij industriële processen. In het onderzoek zijn andere mengsels die niet intentioneel zijn ontstaan in producten/preparaten en waar mensen aan kunnen worden blootgesteld niet meegenomen.

Uiteindelijk zijn 915 ZZS-stoffen in dit BTO-onderzoek geclusterd op basis van similarity MACCS-scores. Hierin zitten óók gewasbeschermingsmiddelen en (dier)geneesmiddelen. De hier beschreven RIVM-studie houdt enkel rekening met vergunningverlening voor niet-gewasbeschermingsmiddelen. Residuen van gewasbeschermingsmiddelen, biociden en (dier)geneesmiddelen vallen buiten de scope van dit onderzoek.

III.1 Wetenschappelijke achtergrond bij het begrip cumulatie.

Stoffenkaders zijn van oudsher gericht op een stof-per-stof benadering, waarmee interactie tussen stoffeigenschappen in mengsels niet worden meegenomen. Het is lastig om mengsels te reguleren aangezien stoffen in mengsels soms door verschillende producenten op de markt worden gebracht. Daarnaast moet men kunnen inschatten welke stoffen al in het milieu(compartment) aanwezig zijn en welke mengsels zullen ontstaan bij toevoeging van een nieuwe stof. Om een uitspraak te kunnen doen over de toxiciteit van een mengsel dient men rekening te houden met de toxicologische werkingsmechanismen van een stof. Bij een gelijk werkingsmechanisme (bijvoorbeeld twee neurotoxische stoffen) kan de toxiciteit worden opgeteld, maar vaak hebben verschillende stoffen in een mengsel verschillende werkingsmechanismen of is het werkingsmechanisme van individuele stoffen niet bekend. Wanneer stoffen in een mengsel allen hetzelfde werkingsmechanisme hebben, wordt het concentratie-additieprincipe toegepast. Bij een verschillend (of onbekend) werkingsmechanisme, wordt het responsadditieprincipe (of *independent action*) toegepast.

Concentratie-additie (CA): hierbij wordt ervan uitgegaan dat de stoffen een gelijk werkingsmechanisme hebben en alleen verschillen in potentie zullen leiden tot een verschillend effect (toxicologische potentie). Van elke stof in een mengsel kan de concentratie dan worden geconverteerd naar toxische eenheden door deze te delen door eenzelfde effectwaarde (de concentratie waarop (g)een effect wordt verwacht). Deze worden vervolgens opgeteld:

$$TU_{mix} = \sum_i \frac{C_i}{EC_{x_i}}$$

Deze toxische eenheden (toxic units: TUs) kunnen vervolgens worden omgerekend naar EC₅₀-mix middels (Escher et al., 2020):

$$EC_{x_{mix}} = \frac{1}{TU_{mix}}$$

Hiervoor zijn wel altijd data nodig m.b.t. relatieve effectwaarden (die de toxische potentie van een stof beschrijven) en concentraties in het milieucompartiment.

Responsadditie (RA) of Independent Action (IA): bij dit model wordt ervan uitgegaan dat elke stof een ander werkingsmechanisme heeft en onafhankelijk bijdraagt aan het effect van de mengsel. Ook hier kunnen berekeningen enkel worden uitgevoerd als er informatie bekend is over de concentratie van de individuele stoffen in het milieucompartiment en het effect wat de individuele stoffen hebben op mens en milieu.

$$effect_{mix} = 1 - \prod_i (1 - effect_i)$$

Het werkingsmechanisme en het effect van veel stoffen in humane toxicologie is echter vaak onbekend, waardoor dit model voornamelijk in ecotoxicologie wordt toegepast om de toxische druk van mengsels (uitgedrukt in msPAF) op levensgemeenschappen te beoordelen.

Interactie: deze term wordt gebruikt bij situaties waarin stoffen elkaars effect kunnen versterken of afzwakken. Dit is sterk afhankelijk van de samenstelling van het mengsel en behoeft per definitie een case-by-case benadering. De modellen die hierboven zijn beschreven houden geen rekening met het feit dat stoffen verschillende effecten kunnen hebben (bijvoorbeeld zowel neurotoxisch als kankerverwekkend). Bij elk model wordt uitgegaan van één type effect. Voor milieueffecten blijken het CA-model en het IA-model goed te werken, met name omdat er naar generieke effecten wordt gekeken (mortaliteit, voorplanting en overleving). Eindpunten bij het bepalen van gezondheidkundige effecten zijn erg specifiek en hierbij speelt blootstellingsroute een grote rol bij het waar te nemen effect. Om deze reden is het heel lastig om mengseleffecten te berekenen voor de mens. Dit is enkel mogelijk voor specifieke orgaaneffecten. Mengseltoxiciteit voor milieueffecten is vaak dus eenvoudiger te bepalen dan voor gezondheidkundige effecten. Om dit wel mogelijk te kunnen maken is meer informatie nodig over de interne blootstelling van stoffen in het menselijk lichaam, en de daaraan gelieerde transformatieprocessen. Daarbij worden deze effecten vaak bepaald door een beperkt aantal stoffen in het mengsel (< 10) (Posthuma et al., 2016). De aard van deze stoffen kan echter per locatie verschillen. Dit is van belang wanneer cumulatie wordt meegenomen in vergunningverlening.

Groeperen van stoffen

Stofconcentraties van stoffen met eenzelfde werkingsmechanisme binnen een mengsel kunnen worden gegroepeerd door deze te converteren naar Toxic Equivalency Factors (e.g. PCB-equivalenten (Van den Berg et al., 1998)) of relatieve potentiefactoren (e.g. bij PFAS en PAKs (Nisbet and Lagoy, 1992; Zeilmaker et al., 2018)). In een eerder BTO-project werden effectsignaalwaarden per effecttype (per bioassay) afgeleid op basis van effectwaarden voor een referentiestof. Technisch gezien werd hier dus dezelfde aanpak gehanteerd als hierboven (voor PAKs en PFASs). Stoffen worden hier niet gegroepeerd, maar worden wel zo beoordeeld (door het gebruik van referentiestoffen en relatieve potentiefactoren). Voor het gebruik van relatieve potentiefactoren kan het verstandig zijn om enkel bioassays mee te

nemen die relevant zijn in de beoordeling van ZZS, zoals bioassays gerelateerd aan genotoxiciteit. Om de toxiciteit van een mengsel van ZZS te beoordelen is informatie over de samenstelling van het mengsel essentieel. De gezamenlijke effecten van een mengsel van stoffen kunnen worden bepaald door middel van bioassays met eindpunten gerelateerd aan ZZS-classificatie (b.v. genotoxiciteit of carcinogeniteit). In de praktijk worden veel mengsels in effluentdirect getest op (eco)toxische effecten middels totaaleffluentbeoordelingen (TEB), waarbij *in vivo* (b.v. *Daphnia magna*, *Brachydanio rerio* en *Vibrio fischeri*) en *in vitro* (b.v. Umu-C) bioassays zijn meegenomen (Maas et al., 2003). Hierbij is dus niet duidelijk door welke (interactie van) individuele stoffen de effecten worden bepaald. Omdat de gevoeligheid van organismen per stof kan verschillen is het van belang om meerdere bioassays en modelorganismen te gebruiken.

In dit BTO-onderzoek naar ZZS is enkel gekeken naar gezondheidkundige effecten van deze stofgroep, waarbij is geclusterd op *similarity* van stoffen. Bij deze clusters is geen rekening gehouden met het toxicologisch werkingsmechanisme van de stof, hoewel veel stoffen binnen de clusters wel een vergelijkbaar toxicologisch werkingsmechanisme kunnen hebben (Zie Hoofdstuk 4.3.6. en Hoofdstuk 2.3). QSARs om toxiciteit te voorspellen van ZZS zijn ontwikkeld op basis van toxicologische gegevens over humaanrelevante effecten (mutageniteit, neurotoxiciteit etc.) uit ToxCast (Hoofdstuk 4). Deze data zijn gebaseerd op resultaten uit *in vitro* bioassays, waarbij voornamelijk gebruikgemaakt wordt van menselijke cellen, terwijl *in vivo* bioassays typisch worden gebruikt om milieueffecten te voorspellen.

III.II Huidige situatie cumulatie en vergunningverlening.

Op dit moment wordt in het huidig BTO-onderzoek naar ZZS amper rekening gehouden met de cumulatie van stoffen. Daarbij wordt cumulatie momenteel ook beperkt meegenomen in vergunningverlening in Nederland. In onderstaande tekst worden de manieren waarop cumulatie kan worden meegenomen in vergunningverlening verder toegelicht, op basis van eerdergenoemd rapport van het RIVM (Bodar et al., 2022). Emissies naar lucht en water worden in dit rapport apart behandeld omdat voor beiden een aparte beoordelingssystematiek geldt. Ook wordt onderscheid gemaakt tussen immissies en emissies:

- o Immissie is het binnendringen van een stof in bodem, water of lucht. Het betreft hier enkel het primaire ontvangende compartiment (oppervlaktewater of lucht)
- o Emissie is de uitstoot van stoffen uit industriële puntbronnen (lozingspijp of schoorsteen).

Water

ABM. Momenteel wordt de waterbezwaarlijkheid van een stof beoordeeld middels de Algemene Beoordelingsmethodiek (ABM). Via de [ABM-tool](#) worden stoffen ingedeeld in vier waterbezwaarlijkheidsklassen:

- o ZZS (Z)
- o Niet snel afbreekbare, waterbezwaarlijke stoffen (A)
- o Afbreekbare, waterbezwaarlijke stoffen (B)
- o Stoffen die van nature voorkomen in het lokale oppervlaktewater (C)

Bij elk van deze categorieën geldt een saneringsinspanning in de vorm van een bronaanpak met de best beschikbare technieken (BBT). Dit zijn de inspanningen die van een lozer verwacht mogen worden om de lozing te voorkomen, en komen voort uit de ABM-tool. In het RIVM-rapport worden deze technieken niet verder toegelicht. Mengsels worden ingedeeld door gebruik te maken van de systematiek van de Europese CLP-verordening, op basis van toxiciteitsklassen en regels om mengsels in te delen in deze toxiciteitsklassen. Het betreft hier enkel producten en preparaten en dus “bedoelde” mengsels i.p.v. onbedoelde mengsels waarvan de precieze samenstelling niet altijd bekend is.

Immissietoets. Bij de immissietoets wordt de toelaatbaarheid van de lozing beoordeeld in het licht van de kwaliteit van het waterlichaam wat de lozing ontvangt en de daarvoor beschikbare milieukwaliteitseisen (MKE). Deze toets stelt dan vast of er verdergaande maatregelen dan de bronaanpak en BBT (zoals beoordeeld met de ABM-tool) nodig zijn. De

immissietoets is opgebouwd uit meerdere stappen. Bij de *significantietoets* wordt gekeken naar de concentratieverhoging op de rand van de mengzone (de directe omgeving van het lozingspunt). Deze concentratieverhoging mag hierbij niet meer dan 10% bedragen om cumulatie door meerdere lozingen te voorkomen. Er wordt hierbij echter geen rekening gehouden met emissies van meerdere stoffen. Naast de significantietoets wordt middels een *normtoets* nagegaan of de concentratieverhoging, bovenop het achtergrondgehalte, geen nadelig effect heeft op de waterkwaliteit, door gebruik te maken van bestaande MKE voor oppervlaktewater. Hierbij wordt een richtwaarde van 10% van de norm (MKE) gehanteerd.

Voor innamepunten van drinkwater geldt een aparte toetsing binnen de emissietoets. Hierbij wordt gebruikgemaakt van een aparte MKE voor oppervlaktewater bij innamepunten voor drinkwater. Als deze MKE niet beschikbaar is, wordt een signaleringswaarde van 0,1 µg/L gehanteerd. Bij verwachte overschrijding van deze waarde, moeten concentraties getoetst worden aan drinkwaterrichtwaarden, onderbouwde gezondheidkundige grenzen, waarbij wel rekening wordt gehouden met de bijdrage via diverse blootstellingsroutes (er wordt uitgegaan van hooguit 20% bijdrage via drinkwater). Zoals hierboven beschreven is vergunningverlening er momenteel vooral op gericht om concentraties van individuele stoffen te toetsen aan normen en worden mengsels niet beoordeeld.

Lucht

Toetsing op luchtmissies vindt plaats door stoffen volgens het Activiteitenbesluit in te delen in stofcategorieën en -klassen, afhankelijk van hun fysisch-chemische of toxicologische eigenschappen. Voor ZZS gelden de volgende stofklassen:

- ERS: extreem risicovolle stoffen (zoals dioxine-achtigen)
- MVP 1: minimalisatie-verplichte vaste stoffen
- MVP 2: minimalisatie-verplichte gassen

Deze stofklassen zijn anders dan bij de watergerelateerde toetsen, waarin ZZS allen als gelijk worden behandeld. Bij iedere stofklasse hoort een grensmassaastroom (drempelwaarde om te zien of ongereinigde emissies relevant zijn) en emissiegrenswaarde waar een industriële puntbron aan moet voldoen. Is de omvang van alle emissies van het gehele bedrijf groter dan de grensmassaastroom (in gram per uur), dan is de emissie relevant. Daarna wordt de emissie getoetst aan de emissiegrenswaarde (in milligram per kuub lucht). Gelijktijdig optredende emissies van verschillende stoffen binnen eenzelfde stofcategorie of -klasse worden opgeteld. Gesommeerde ZZS emissies (voor ERS, MVP 1 en MVP 2) worden getoetst aan de grensmassaastroom voor MVP 2 (2,5 gram per uur). Bij de emissiegrenswaarde vindt een vergelijkbare sommatiebepaling plaats, waarbij de totale ZZS-concentraties wordt getoetst aan de emissiegrenswaarde voor MVP 2 (1 mg/m³). In beide bepalingen worden stoffen uit andere stofgroepen niet meegenomen in de sommatie. De grensmassaastroom als drempelwaarde verdwijnt in de Omgevingswet, waardoor aandacht voor ZZS-emissies naar de lucht vanuit meerdere bronnen binnen een bedrijf op de achtergrond lijken te raken. De sommatiebepaling voor de emissiegrenswaarde blijft wel in de Omgevingswet, maar geldt enkel per stofklasse. Wanneer voor een individuele stof of stofgroep de emissiegrenswaarde(n) worden overschreden, worden deze getoetst aan de maximaal toelaatbare risiconiveau (MTR) voor lucht. Bij mengsels wordt de MTR-toetsing enkel uitgevoerd voor de relevante individuele ZZS-bestanddelen in het mengsel, waarbij geen sommatie meer plaatsvindt. Bij afleiding van een MTR gaat men echter uit van 100% allocatie via de lucht, waardoor geen marge wordt ingebouwd voor blootstelling via andere routes.

Cumulatiebenadering in omliggende landen

In omliggende landen wordt vaak de SVHC-lijst (substances of very high concern) van ECHA gebruikt om ZZS aan te duiden. SVHC-criteria voor stoffen zijn echter kritischer dan ZZS-criteria, waardoor deze lijsten korter zijn. Duitsland en België (Vlaanderen) maken gebruik van biologische monitoring in effluent om impliciet de effecten van mengseltoxiciteit mee te nemen in de beoordeling. Cumulatie-effecten lijken echter niet expliciet te worden vermeld in de vergunningverlening. Er wordt echter wel rekening gehouden met de bijdrage van de lozing ten opzichte van de reeds aanwezige stoffen in het oppervlaktewater (Brinkmann et al., 2016). Ook vindt in Duitsland, net als in Nederland, sommatie plaats van stoffen uit dezelfde stofcategorieën bij luchtmissies. In Denemarken zijn richtlijnen hoe cumulatie meegenomen kan worden bij vergunningverlening. Hierbij wordt getoetst aan het gewogen gemiddelde van de normen

voor de individuele stoffen en gaat men dus uit van concentratie-additie. Ook vindt in Denemarken biologische monitoring van afvalwater plaats.

III.III Andere beleidskaders

Kaderrichtlijn Water (KRW)

De KRW is een Europese richtlijn die waterbeheerders verplicht om maatregelen te nemen die nodig zijn om een goede waterkwaliteit te waarborgen. De richtlijn bevat geen handvatten om mengseltoxiciteit (van ZZS) te voorkomen (preventie), maar de effecten van mengseltoxiciteit worden wel in retrospect meegenomen door de ecologische toestand (biodiversiteit) te monitoren. Dit zegt echter weinig over drinkwaterkwaliteit, aangezien oppervlaktewater wordt gezuiverd alvorens hier drinkwater van wordt geproduceerd. Een afwijking in biodiversiteit kan mogelijk optreden door een slechte chemische toestand, onder andere door mengseltoxiciteit. Om de beoordeling van mengsels van stoffen mogelijk te maken, is in Nederland de Ecologische Sleutelfactor (ESF) [ontwikkeld](#), waarbij de toxische druk op het systeem direct kan worden getoetst aan de hand van gemeten concentraties. Ook wordt de ESF ingezet om vóór een lozing te beoordelen of een locatie al belast is met stoffen.

REACH

REACH (Registratie, Evaluatie, Autorisatie en restrictie van Chemische stoffen) is een Europese verordening die zich richt op de preventieve risicobeoordeling van individuele (ZZS-) stoffen en intentionele mengsels (producten), waarbij modellen worden gebruikt om risico's voor de mens in te schatten, door gebruik te maken van het concept van Mixture Assessment/Allocation Factors (MAF). Het idee van MAF is dat op voorhand met een vaste factor al rekening gehouden wordt met mengsels die later kunnen ontstaan. Hierbij werd tot voorheen echter geen rekening gehouden met onbedoelde mengsels. Inmiddels is het MAF-concept ook voor onbedoelde mengsels onderbouwd door middel van diverse verkennende studies (Rorije et al., 2022). In het onderzoek van Rorije et al. (2022) werd echter enkel gekeken naar effecten van mengsels op waterorganismen en nog niet naar effecten van mengsels op de mens. Het betreft in dit onderzoek ook een groter spectrum aan stoffen (bestrijdingsmiddelen, geneesmiddelen en industriële producten) dan in de vergunningverlening van industriële emissies aan de orde is. Momenteel bestaan er vanuit de Europese Commissie ook plannen om de MAF-aanpak op te nemen in de risicobeoordeling bij REACH-registratie. Dit betreft echter een generieke risicobeoordeling die geen direct verband heeft met de daadwerkelijke risico's van lokale emissies. In een vervolgonderzoek na het huidige BTO-project ZZS kan gekeken worden naar het groeperen van mengsels van ZZS en het effecten van mengsels en het onderbouwen van het MAF-concept voor humane effecten bij immissietoetsen.

Gewasbeschermingsmiddelen (inclusief gewasbeschermingsmiddelen die ZZS bevatten)

Toelating. Risico's van gewasbeschermingsmiddelen worden beoordeeld volgens Europese wetgeving en criteria. Enkel middelen die effectief zijn en geen onacceptabele risico's hebben voor mens en milieu mogen enkele jaren in Nederland worden verhandeld en gebruikt. Bij toelating worden enkel bedoelde mengsels (producten) beoordeeld door het CTGB. EFSA (European Food Safety Authority) heeft een methode ontwikkeld om ook voor onbedoelde mengsels risico's in te schatten door stoffen in te delen in CAGs (cumulatieve assessmentgroepen), stoffen met een vergelijkbare toxisch werkingsmechanismen. Blootstelling via voedsel en water van mengsels van bestrijdingsmiddelen kan vervolgens berekend worden door gebruik te maken van MCRA (Monte Carlo Risk Assessment), ontwikkeld door het RIVM en Wageningen Universiteit. Onbekend is of het CTGB gebruikmaakt van MCRA in hun risicobeoordelingen.

Onderzoek oppervlaktewater. Milieubelasting van bestrijdingsmiddelen op oppervlaktewater kan worden berekend door gebruik te maken van het NMI-model (Nationale Milieu-Indicator), op basis van gebruik, geografische informatie, emissie-informatie, stofeigenschappen en normen. Milieubelasting wordt vervolgens uitgedrukt in milieu-indicatorpunten. Uit onderzoek blijkt echter dat in de oppervlaktewatermonsters slechts een klein deel van de (bekende) stoffen de gezamenlijke mengseltoxiciteit bepalen. Daarnaast verschilt per locatie welke stoffen dit zijn, mede afhankelijk van het schaalniveau.

Onderzoek Toxische druk. De risico's van (p)ZZS in gewasbeschermingsmiddelen kunnen worden bepaald door gebruik te maken van soortgevoeligheidsverdelingen (SSDs). Deze verdelingen geven een beeld van de spreiding van gevoeligheden van soorten op basis van *in vivo* ecotoxiciteitstesten. Door deze verdelingen te vergelijken met monitoringsdata kan worden bepaald welk percentage van soorten potentieel last ondervindt van een bepaalde concentratie van een stof (ook wel toxische druk genoemd). De toxische druk (op de y-as van de SSD) wordt uitgedrukt in de potentieel aangetaste fractie van soorten (PAF). Bij het testen van mengsels van stoffen wordt een 'meer-stoffen/multiple substances'-PAF gebruikt om mengseltoxiciteit te meten. In vervolgonderzoek na het huidige BTO-ZZS-project zouden SSDs voor ZZS clusters opgesteld kunnen worden op basis van de ECOTOX-database van de US EPA, en (ms)PAFs op specifieke locaties op basis van monitoringsdata.

III.IV Conclusies en perspectieven

Duidelijk is dat cumulatie een belangrijk onderdeel is in de risicobeoordeling van stoffen en dus ook bij vergunningverlening, omdat dit een instrument is om blootstellingen en risico's voor mens en milieu te beperken. De vraag die cruciaal is in de RIVM-studie is of het meenemen van cumulatie van stoffen in de vergunningverlening van meerwaarde is ten opzichte van de huidige situatie. Het RIVM beveelt aan om, los van wat methodologisch kan, ook te onderzoeken waar het meenemen van deze effecten het meest zinvol is. De belangrijkste punten die onder andere worden genoemd zijn:

- Hoe verhouden de onzekerheden rond cumulatie zich met onzekerheden in de risicobeoordeling en de vergunningverlening?
- In de regel draagt een klein aantal stoffen het meest bij aan mengseleffecten.

In het BTO-ZZS-project houden we geen rekening met mogelijke mengseltoxiciteit. Binnen de ZZS-clusters is deels bekend welke hiervan de hoogste toxiciteit hebben. In aanvullend onderzoek na het huidige BTO ZZS project kan worden onderzocht of deze stoffen ook verantwoordelijk zijn voor het grootste aandeel in eventuele mengseltoxiciteit.

IV Details meetmethode

IV.1 Overzicht componenten en rapportagegrenzen

Component	Cas nummer	Interne standaard	Rapportagegrens (ng/l)			
			Ultrapuur water	DRW	OPW	GW
Furosemide	54-31-9	Furosemide-D5	25	25	25	25
1,1,2,2,3,3,4,4,4-nonafluorobutane-1-sulfonic acid	375-73-5	Furosemide-D5	50	50	50	100
cycloheximide	66-81-9	Nicosulfuron-D6	25	50	50	50
Climbazole	38083-17-9	Climbazole-D4	5	5	5	5
2-hydroxybenzothiazool	934-34-9	Benzotriazole-D4	5	10	10	10
2-methylthiobenzothiazool	615-22-5	Benzotriazole-D4	50	50	50	100
Nicosulfuron	111991-09-4	Nicosulfuron-D6	5	5	5	5
Pentoxifylline	6493-05-6	Pentoxifylline-D6	5	5	5	5
Benzotriazole	95-14-7	Benzotriazole-D4	10	25	25	25
2,4-dinitrofenol	51-28-5	2,4-dinitrofenol-D3	10	10	10	10
Terbutylazine	5915-41-3	Terbutylazine-D5	5	5	5	5
Metoprolol	37350-58-6	Metoprolol-D7	5	5	5	5
2,4,5-trimethylaniline	137-17-7	Tetraglyme-D6	5	5	5	5
Carbendazim	10605-21-7	Carbendazim-D4	5	5	5	5
Triethyl phosphate	78-40-0	Triethyl phosphate-D15	10	50	50	50
Thiacloprid	111988-49-9	Thiacloprid-D4	5	5	5	5
Phoxim	14816-18-3	Phoxim-D5	5	10	10	10
Drometrizole	2440-22-4	Drometrizole-D3	50	50	50	50
Metolachloor ESA (CGA 354743)	171118-09-5	Metolachlor ESA-D6	25	25	25	25
Warfarin	81-81-2	Warfarin-D6	5	5	5	5
Tetraglyme	143-24-8	Tetraglyme-D6	5	5	5	5
Metobromuron	3060-89-7	Metobromuron-D6	10	25	25	25
Bromacil	314-40-9	Furosemide-D5	10	10	10	10
Saccharine	81-07-2	Saccharine-D4	50	50	50	50
Metazachlor	67129-08-2	Metazachlor-D6	5	5	5	5
4-chloor-2-methylfenoxiazijnzuur (MCPA)	94-74-6	Furosemide-D5	25	25	25	25
Oxybenzone	604-75-1	Oxybenzone-D3	10	10	10	10
Methamphetamine	537-46-2	Methamphetamine-D5	5	5	5	5
Benzoyllecgonine	519-09-5	Benzoyllecgonine-D3	5	5	5	5
MDMA	42542-10-9	MDMA-D5	5	5	5	5

Oxazepam	604-75-1	Oxazepam-D5	5	5	5	5
Methadon	76-99-3	Methadon-D9	5	5	5	5

IV.II Instellingen UHPLC en MS

Tabel 8: UHPLC instellingen

UHPLC pomp	Shimadzu Nexera X2 LC-30AD
Monsterwisselaar en temperatuur	Shimadzu Nexera X2 SIL-30AC, 15 °C
Analytische kolom	Phenomenex Luna Omega Polar C18 2.1x100 1,6 µm
Guardkolom	Phenomenex securityGuard ultra cartridges C18
Kolomthermostaat en temperatuur	Shimadzu Nexera X2 CTO-20AC, 20 °C
Autosampler spoelvloeistof R3	70% acetonitril, 15% methanol, 15% ultrapuur water
Autosampler spoelvloeistof R1 en R2	Wordt niet gebruikt
Autosampler spoelvloeistof R0	Ultrapuur water
Injectievolume	100 µl
Maximale pompdruk	1000 bar

Tabel 9: UHPLC gradiënt

Stap	Tijd in minuten	1-A%	2-A%	Flow (ml/min)
0	0	96	4	0.3
1	1	96	4	0.3
2	13	50	50	0.3
3	14	0	100	0.3
4	18	0	100	0.3
5	18.5	96	4	0.3
6	20	96	4	0.3

Mobiele fase 1-A: Ultrapuur water + 0,10% mierenzuur. Mobiele fase 2-A Acetonitril + 0,10% mierenzuur.

Tabel 10: Instellingen massaspectrometer

Massaspectrometer	SCIEX Triple Quad 6500+
Scanmethode	MRM
LC/MS interface	H-ESI
Ionisatie mode	Positief en Negatief
Target cycle time	0,4 s

	(across sMRM experiments)
Curtain gas (CUR)	30
Colission gas (CAD)	8
IonSpray Voltage (IS)	Positief: +4000 v Negatief: -4500 v
Temperature (TEM)	400
Ion Source gas 1 (GS1)	40
Ion source gas 2 (GS2)	40
Entrance potential (EP)*	Positief: +10 Negatief: -10
Settling time	5 ms
Pause between mass ranges	5 ms
Min. dwell tijd	3 ms
Max. dwell tijd	50 ms
Divert valve	2 – 18 minuut
Resolutie Q1	Unit
Resolutie Q3	Unit
Horizontale probe positie	3,0 mm
Verticale probe positie	5,0 mm