Kennisimpuls
WATERKWALITEIT

Verbeteringsmogelijkheden voor regionale kennisregels KRW-Verkenner
Pilot: macrofauna in stromende wateren

Verbeteringsmogelijkheden voor regionale ecologische kennisregels KRW-Verkenner Pilot: macrofauna in stromende wateren

Auteur(s)
Gertjan Geerling (Deltares), Mijke van Oorschot (Deltares), Jasmijn Rost (RHDHV), Niels Evers (RHDHV), Hans Korving (Deltares)

Samenvatting

Dit rapport bevat aanbevelingen voor verbetering van de regionale ecologische kennisregels van de KRW-Verkenner. Het is onderdeel van het thema ecologie van de Kennisimpuls Waterkwaliteit en bevat een uitgewerkte case voor macrofauna in stromende wateren. Het doel van deze studie was om te onderzoeken of een meer gedetailleerde dataset met meer parameters en meer trajecten een verbeterde voorspelling oplevert van EKR-scores. Daarnaast is gekeken welke typen voorspelmodellen de beste resultaten opleveren en waar rekening mee gehouden moet worden bij het trainen en testen van de voorspelkracht van de modellen. Hiervoor is een gedetailleerde dataset gebruikt met informatie van verschillende waterschappen. Deze dataset is uitgebreider dan de dataset die gebruikt is om de huidige kennisregels af te leiden omdat deze kleinere trajecten bevat met meer parameters. Dit biedt een goede basis voor het beantwoorden van de onderzoeksvragen.

Resultaten laten zien dat de modelresultaten statistisch significant verbeteren als er meer gedetailleerde stuurvariabelen worden gebruikt. Hiervoor zijn de huidige samengevoegde stuurvariabelen uitgesplitst en afzonderlijk gebruikt in het trainen van het model. De aanbeveling is dan ook om de modelresultaten te verbeteren door stuurvariabelen op te splitsen, niet relevante stuurvariabelen te verwijderen en aanvullende relevante stuurvariabelen toe te voegen. Dit maakt ook dat de aansluiting op maatregelen beter kwantitatief te maken is.

Het toevoegen van kleinere trajectgrootte leidt niet automatisch tot een betere modeltraining. Het opknippen van trajecten in kleinere trajecten lijkt alleen zinvol als de grote trajecten heterogeen zijn of op basis van gebieden waar wel en geen maatregelen getroffen worden. In dit geval levert een opsplitsing dus een andere range van stuurvariabelen op en kan het een bijdrage leveren aan een lokaal betere voorspelling of een meer gericht voorspelling van maatregel-effecten.

Het voorspelmodel dat het beste presteert op deze dataset is de Gradient Boosting Regression. Dit is een andere methode dan de Ranger Random Forest die op dit moment in de verkenner gebruikt wordt. De EKR wordt het beste voorspeld tussen EKR-scores 0.3 en 0.6. Boven een EKR-score van 0.6 worden ze meestal onderschat en onder de 0.3 worden ze overschat, dit is op dit moment ook het geval met de KRW-Verkenner. Dit kan mogelijk verbeterd worden door gericht meer waarnemingen toe te voegen die aan de uiteinden van het EKR-bereik zitten, bijvoorbeeld waarnemingen uit het buitenland of gegenereerde waarden op basis van expert-oordelen met behulp van de referentiebeschrijvingen in de maatlatdocumenten.

De voorspelkracht van het model verbetert niet als er extra waarnemingen worden toegevoegd die in de gemiddelde range van EKR-scores zitten. Ook lijken de huidige stuurvariabelen minder goed onderscheid te kunnen maken tussen mediane en hoge EKR-scores. Om het model beter te laten presteren over de gehele EKR-range kan “physics-based learning” toegepast worden in de modeltraining. Dit houdt in dat je kennis over het systeem toevoegt in de training zodat het model beter “op de hoogte is van bij ons bekende (causale) relaties”.

Verschillende verdelingen in de training- en testset van het model leveren verschillende uitkomsten op. Deze variatie kan inzichtelijk gemaakt worden door meerdere random trekkingen te doen bij verschillende machine learning modellen, deze paarsgewijs met elkaar te vergelijken en de verschillen te toetsen op significantie. Op basis hiervan kan het beste model gekozen worden en vervolgens getraind op de volledige dataset.

Inhoud

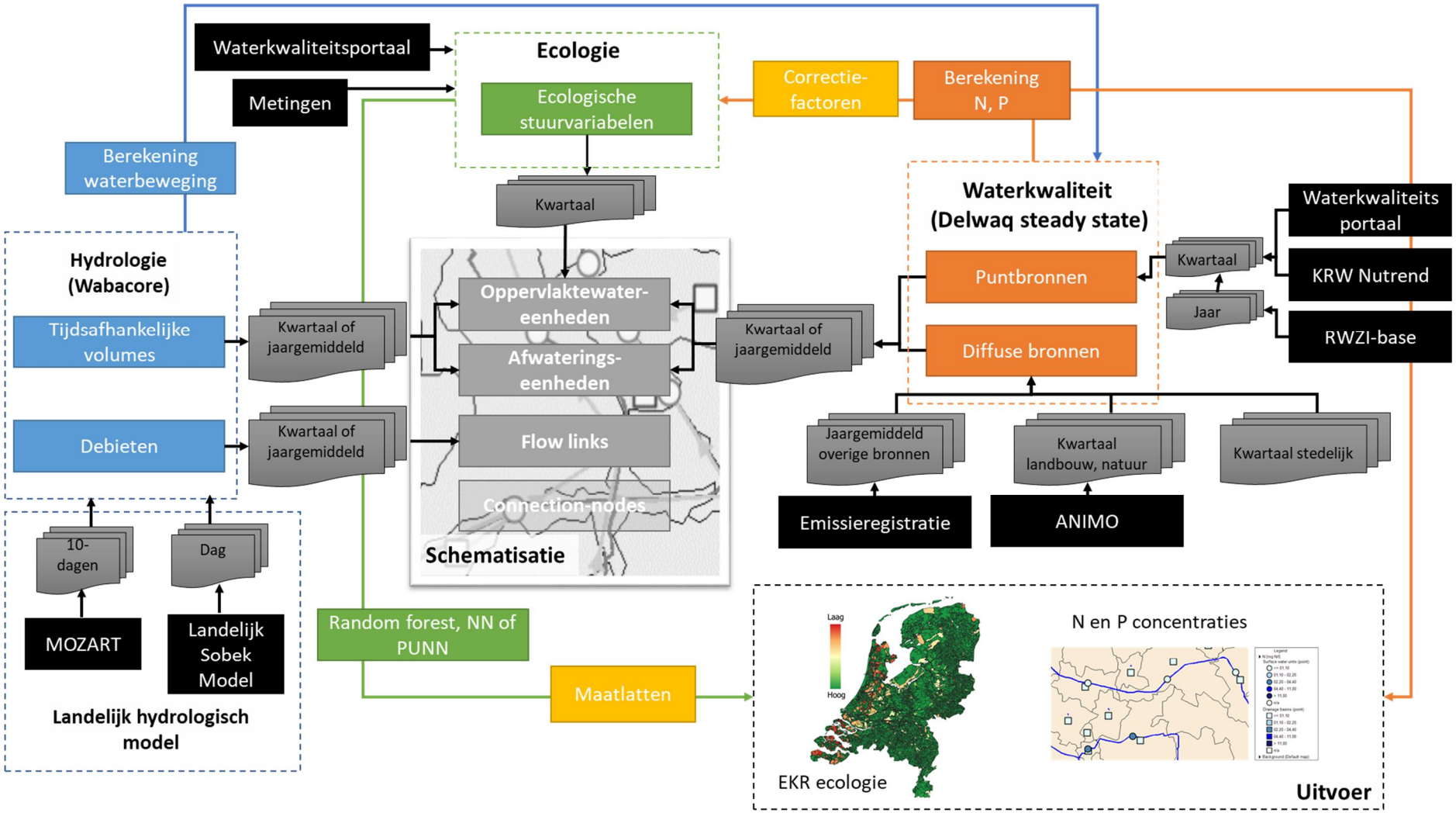
1 Introductie

1.1 KIWK

Binnen het thema ecologie van de Kennisimpuls Waterkwaliteit wordt bestaande en nieuwe kennis bijeengebracht over het beoordelen van de ecologische toestand en de effectiviteit van herstelmaatregelen en wordt deze kennis (beter) toepasbaar gemaakt voor de praktijk. Een van de toepassingen die hiervan gebruik kan maken is de Nationale Analyse Waterkwaliteit (NAWK) (van Gaalen et al. 2020). Middels NAWK wordt op landelijk niveau een prognose gegeven voor de ecologische verbetering op basis van de voorgenomen maatregelen. De KRW-Verkenner is een van de modules in het modelinstrumentarium van de NAWK. In een eerder product is de aanpak geëvalueerd en is verkend wat de mogelijkheden voor verbetering zijn (Buijse & van Geest 2021). Deze rapportage beschrijft een concreet uitgewerkte casus van verbeteringsmogelijkheden voor macrofauna in stromende wateren.

1.2 KRW-Verkenner

De KRW-Verkenner wordt gebruikt om het effect van maatregelen in watersystemen te voorspellen. Dit is gebeurd in landelijke projecten zoals de Ex-ante evaluatie uit 2021 en de Nationale Analyse Waterkwaliteit uit 2019 om het doelbereik na uitvoering van de geplande maatregelen en alternatieve scenario's te onderzoeken (van Gaalen et al. 2020). Daarnaast is de KRW-Verkenner veel toegepast in regionale KRW-watersysteemanalyses. Hierbij is de KRW-Verkenner vooral toegepast om de effecten van maatregelen op de biologische EKR's in beeld te brengen en waar nodig doelen (technisch) aan te passen. De ecologische rekenregels in de KRW-Verkenner bieden daarmee kwantitatieve inzichten in effecten van maatregelen op het niveau van EKR's waarvoor verder feitelijk nog geen andere operationele tools beschikbaar zijn.



Figuur 1-1 Flow diagram met alle processen en koppelingen in de KRW-Verkenner.

Figuur 1-1 geeft een overzicht van de verschillende modules in de KRW-Verkenner met de bijbehorende verbindingen en informatiebronnen. Als eerste wordt de waterbeweging uitgerekend, die invoer levert aan het waterkwaliteitsmodel, dat vervolgens nutriëntconcentraties doorgeeft aan het ecologische model. Als uitvoer kunnen de berekende stofconcentraties weergegeven worden en in combinatie met EKR-maatlatten ook de EKR-scores van alle biologische kwaliteitselementen.

In dit rapport concentreren we ons op de Ecologie module (zie groene elementen in Figuur 1-1) waarbij we zowel de ecologische stuurvariabelen als de rekenmodule, weergegeven als “Random forest, neuraal netwerk (NN) of PUNN” onderzoeken.

1.3 Probleembeschrijving

Uit eerder werk komt een aantal aspecten naar voren waarin de KRW-Verkenner kan worden verbeterd (Buijse & Van Geest, 2021). Op basis van beschikbare data onderzoeken we een aantal mogelijke aanpassingen op hun toegevoegde waarde voor de voorspelling van de ecologie (EKR-scores) in de KRW-verkenner voor regionale wateren.
Een aantal van de tot nu toe gebruikte stuurvariabelen zijn samengesteld uit verschillende elementen. De samengestelde stuurvariabelen zijn:

- *Meandering* is samengesteld uit een lateraal en longitudinaal profiel.
- *Oeverinrichting* is samengesteld uit beschoeiing en type vegetatie.
- *Beschaduwing* is samengesteld uit mate van beschaduwing en type oevervegetatie.
- *Verstuwing* is samengesteld uit een kwalitatieve maat voor verstuwing en de aanwezigheid van vistrappen.

Daarnaast ontbreken stuurvariabelen voor hydrologie zoals afvoerdynamiek, droogval, piekafvoer etc. De hydrologische variabele die nu in de KRW-Verkenner is opgenomen is verstuwing, wat een samengestelde maat is en een kwalitatieve maat voor peilbeheer.

Onze verwachting is dat het uit elkaar halen van deze samengestelde stuurvariabelen een eenduidigere stuurvariabele-EKR-respons geeft, tegelijkertijd beter te interpreteren is en directer te koppelen is aan inrichtingsmaatregelen. Hiermee wordt het voorspellen en interpreteren van het effect van een maatregelpakket verbeterd. Een sterkere correlatie tussen opgesplitste stuurvariabelen en EKR-score kan worden getest met al beschikbare datasets voor de macrofauna EKR.

Het aantal datapunten bepaalt hoe goed een op machine learning gebaseerd model getraind en getest kan worden. Te weinig punten of een slechte verdeling van punten over de gehele range van model input en output levert een suboptimale voorspelling. De modelprestaties kunnen bij een kleine dataset, en dus een kleine test- en training set sterk variëren bij verschillende random trekkingen. De gevoeligheid hiervoor kan worden onderzocht op basis van bestaande en nieuwe datasets, met verschillende trekkingen van test- en training data.

De segmenten van de waterlopen hebben een variabele ruimtelijke grootte. Bij grotere segmenten representeren de waarden van stuurvariabelen ook een groter areaal, ze zijn meer ruimtelijk gemiddeld. We verwachten dat de correlatie tussen de stuurvariabelen en de EKR-score van biologische elementen (macrofauna, waterplanten, vis) groter is in kleinere en hierdoor meer homogene segmenten.

Uit het bovenstaande volgen de onderzoeksvragen:

1. Verbeteren de modelprestaties als de samengestelde stuurvariabelen worden opgesplitst in meerdere enkelvoudige stuurvariabelen die een eenduidiger relatie hebben met de maatregelen?
    a) Welke parameters zijn belangrijk?
    b) Hoe goed kan de EKR-score voorspeld worden?
    c) Welk model presteert het best?
    d) Is dit significant beter dan de andere modellen?
    e) Presteert dit model significant beter dan het beste bestaande model?

2. Verbeteren de modelprestaties als meer waarnemingen worden toegevoegd aan de dataset met samengestelde stuurvariabelen?
    a) Verschillen de resultaten van een model getraind op de trainingsset uit Visser et al. (2021) plus de nieuwe KIWK-dataset van het beste model uit onderzoeksvraag 1?

3. Zijn modeluitkomsten gecorreleerd met de ruimtelijke omvang van segmenten?

Voor het beantwoorden van de onderzoeksvragen zijn in dit rapport de macrofauna EKR-scores gebruikt van stromende wateren onder beheer van waterschappen. Deze dataset bevat relatief veel datapunten, omdat er met kleinere trajecten is gewerkt. Daarnaast zijn er meer parameters gemeten. In vergelijking met de oorspronkelijke dataset bevat deze selectie dus meer gedetailleerde data en biedt daarom een goede basis om de hierboven gestelde vragen mee te beantwoorden. De datasets inclusief de stuurvariabelen zijn uitgebreid beschreven in hoofdstuk 2 “Data en methoden”.

1.4 Leeswijzer

Het rapport is opgezet als een technische rapportage met voor leken leesbare conclusies en aanbevelingen. De hoofdstukken 2 (Data en methoden) en 3 (Resultaten) bevatten een technische samenvatting van de resultaten uit het onderzoek. Hoofstuk 4 bevat heel concreet per onderzoeksvraag: de conclusies, wat het betekent voor de KRW-Verkenner, en aanbevelingen. De volledige set resultaten wordt geleverd in een afzonderlijke PowerPoint bijlage.

2 Data en methoden

2.1 Beschrijving van de dataset

De KRW-verkenner maakt gebruik van een dataset waarin de relaties tussen ecologische kwaliteitsratio's (EKR-scores) en verschillende waterkwaliteit-, beheer- en inrichtingsvariabelen (stuurvariabelen) voor een groot aantal waterlichamen in Nederland zijn opgenomen (van der Linden et al., 2021). De stuurvariabelen zijn op dit moment een combinatie van direct afgeleide variabelen (bijvoorbeeld zomergemiddelde concentratie totaal stikstof) en samengestelde variabelen (bijvoorbeeld verstuwing met onderliggend de mate van opstuwing en de vispasseerbaarheid).

Voor het afleiden van de vernieuwde kennisregels voor de KRW-verkenner is een nieuwe dataset opgebouwd, waarin samengestelde variabelen opgedeeld zijn in direct afgeleide variabelen. Daarbij is voor dit project gekozen om de dataset te beperken tot slechts één watertypecluster en één biologisch kwaliteitselement (t.o.v. tien watertypeclusters en vier biologisch kwaliteitselementen in de KRW-verkenner). De keuze is gevallen op het cluster langzaam stromende beken (R4a, R4b, R5, R6 en R12), omdat hiervoor een groot aantal datapunten beschikbaar is. Als biologisch kwaliteitselement is macrofauna gekozen vanwege de relatief sterke relatie tussen de EKR-score en de waterkwaliteit en beheer en inrichting van een beek (Evers et al., 2017). Daarnaast is voor macrofauna in deze beken meer informatie beschikbaar dan voor vis en overige waterflora.

2.1.1 Datatype en -bronnen van de dataset

De dataset is samengesteld met data van vijf waterschappen: waterschap Aa en Maas, waterschap Brabantse Delta, waterschap de Dommel, waterschap Limburg en waterschap Rijn en IJssel. Deze waterschappen hebben gebruik gemaakt van een vergelijkbare methode voor de watersysteemanalyse. Hierdoor waren stuurvariabelen op een gedetailleerder niveau beschikbaar dan in de landelijke dataset. Daarnaast hebben deze waterschappen hun KRW-waterlichamen opgedeeld in kleinere homogene deeltrajecten als het ruimtelijk schaalniveau waarop de gegevens van de stuurvariabelen zijn verzameld. Tabel 2-1 toont het aantal trajecten met een volledige dataset per waterschap en KRW-type. De trajecten waarvoor alle variabelen beschikbaar waren, zijn meegenomen in de dataset.

Tabel 2-1 aantal records per waterschap en KRW-watertype waarvoor een EKR-score en de stuurvariabelen volledig beschikbaar waren.

Waterschap	R4a/R4b	R5	R6
Waterschap Aa en Maas	29	17	6
Waterschap Brabantse Delta	27	10	12
Waterschap de Dommel	49	39	11
Waterschap Limburg	24	31	1
Waterschap Rijn en Ijssel	0	33	13

Voor de oorspronkelijke dataset van de huidige rekenregels in de KRW-verkenner was een sub-selectie gemaakt. Deze selectie was gemaakt doordat er aan de ene kant een te hoge dichtheid van records in bepaalde ranges van EKR-scores voorkwam (voornamelijk tussen

0.30 en 0.50) en aan de andere kant om de overlap tussen meerdere ruimtelijke schaalniveaus eruit te filteren (trajecten versus waterlichamen).

De dichtheid in de hier samengestelde dataset is lager in de range tussen EKR-scores 0.30 en 0.50 en er is slechts één ruimtelijk schaalniveau gebruikt (traject) waardoor selectie van datapunten niet nodig was. Alle beschikbare informatie uit tabel 2-1 is dus meegenomen in de uiteindelijke set.

2.1.2 Stuurvariabelen

De stuurvariabelen van de huidige rekenregels in de KRW-verkenner zijn een combinatie van direct afgeleide stuurvariabelen voor de waterkwaliteit en samengestelde stuurvariabelen voor inrichting (Tabel 2-2).

Tabel 2-2 Definitie van de huidige stuurvariabelen in KRW-verkenner.

Stuurvariabele	Klassen/Eenheid	Waarden en omschrijving
Meandering	-	1 = recht + normprofiel, 2 = gestrekt + natuurlijk dwarsprofiel, 3 = zwak slingerend, 4 = slingerend, 5 = vrij meanderend
Beschaduwing	-	1 = onbeschaduwd zonder ruigte op de oevers, 2 = gedeeltelijk beschaduwd of ruigte op de oever, 3 = grotendeels of geheel beschaduwd
Verstuwing	-	1 = sterk gestuwd zonder vistrappen, 2 = gestuwd met vistrappen, 3 = ongestuwd
BZV	Mg/l	Zomergemiddelde BZV
Stikstof totaal (Ntot)	Mg N/l	Zomergemiddelde concentratie stikstof totaal
Fosfor totaal (Ptot)	Mg P/l	Zomergemiddelde concentratie fosfor totaal
Ammonium (NH4)	Mg/l	Maximale concentratie ammonium
Toxiciteit (msPAF)	%	Maximale fractie msPAF

De samengestelde variabelen in de oude dataset zijn samengesteld op basis van of twee onderliggende stuurvariabelen (beschaduwing en verstuwing), of zes onderliggende stuurvariabelen (meandering).

In de nieuwe dataset zijn de samengestelde stuurvariabelen vervangen door de onderliggende stuurvariabelen (tabel 2-3). De variabelen sinuositeit, zomergemiddelde stroomsnelheid 0,05Q, debietfluctuatie, percentage gemaaid profiel (slechtste), percentage beschaduwing en mate van opstuwing zijn door de waterbeheerders op een vergelijkbare manier bepaald en zijn direct overgenomen.

Voor de andere drie onderliggende stuurvariabelen (profielvorm, ruimtelijke variatie in stroomsnelheid en vispasseerbaarheid) was nog een standaardisatie nodig (tabellen 2.4-2.6).

Tabel 2-3 Onderliggende stuurvariabelen voor de samengestelde stuurvariabelen in KRW-verkenner

Samengestelde stuurvariabele KRW-Verkenner	Onderliggende stuurvariabele	Eenheid	Waarden en omschrijving
Meandering	Sinuositeit	-	Waarde voor mate van meandering
	Gemiddelde stroomsnelheid zomer 0,05Q	cm/s	
	Debietfluctuatie	-	
	Profielvorm	-	Categorische variabele met categorieën: natuurlijk, accoladeprofiel, genormaliseerd en vervallen genormaliseerd
	Ruimtelijke variatie in stroomsnelheid	-	Categorische variabele met categorieën: geen, weinig, matig en veel
	Percentage gemaaid profiel (meest intensieve maaibeurt, vaak in het najaar)	%	Percentage
Beschaduwing	Percentage beschaduwing	%	Percentage
Verstuwing	Mate van opstuwing	%	Percentage
	Vispasseerbaarheid	-	Categorische variabele met categorieën: wel en niet

Voor profielvorm hadden alle waterschappen de stuurvariabele bepaald met een aantal categorieën. Deze categorieën kwamen niet exact overeen, daarom zijn ze gestandaardiseerd in vier categorieën op basis van de omschrijvingen van de waterschappen. In Tabel 2-4 is de indeling weergegeven van de standaardisatie. Waterschap Brabantse Delta had nog de categorie “moeras”. Deze is traject specifiek naar natuurlijk of vervallen genormaliseerd op basis van gebiedskennis.

Tabel 2-4 Invulling van categorische ecologische stuurvariabele profielvorm per waterschap

Gestandaar-diseerde variabele	Waterschap Aa en Maas	Waterschap Brabantse Delta	Waterschap de Dommel	Waterschap Limburg	Waterschap Rijn en IJssel
Natuurlijk	Vrij meanderend	Natuurlijk, moeras	Natuurlijke situatie; vrij meanderend	Natuurlijk	Natuurlijk
Accoladeprofiel	Accolade	Twee fasen profiel	Accolade profiel 1 (smal/diep), Accoladeprofiel 2 (breed) 2de fase, 2 fasen met los winterbed, V-vormig profiel		Accoladeprofiel
Genormaliseerd	Genormaliseerd, flauwe oever	Genormaliseerd	Genormaliseerd, cultuurtechnisch profiel	Normprofiel, natuurlijk & Normprofiel	Civieltechnisch profiel
Vervallen genormaliseerd		Vervallen genormaliseerd, moeras	Vervallen cultuurtechnisch profiel	Verwaarloosd normprofiel	

Ook voor vispasseerbaarheid kwamen de categorieën die de waterschappen hadden niet overeen. Daarom zijn ze gestandaardiseerd in twee categorieën (wel of niet) op basis van de omschrijvingen van de waterschappen. In Tabel 2-5 is de indeling weergegeven van de standaardisatie.

Tabel 2-5 Invulling van categorische ecologische stuurvariabele vispasseerbaarheid en visoptrekbaarheid / bereikbaarheid per waterschap

Gestandaar-diseerde variabele	Waterschap Aa en Maas	Waterschap Brabantse Delta	Waterschap de Dommel	Waterschap Limburg	Waterschap Rijn en IJssel
Wel	Barrieres met voorzieningen, Geen barrieres	Goed, Matig	Goed, Matig, NVT	Ja	1
Niet	Geen voorzieningen	Slecht	Niet, slecht	Nee	0

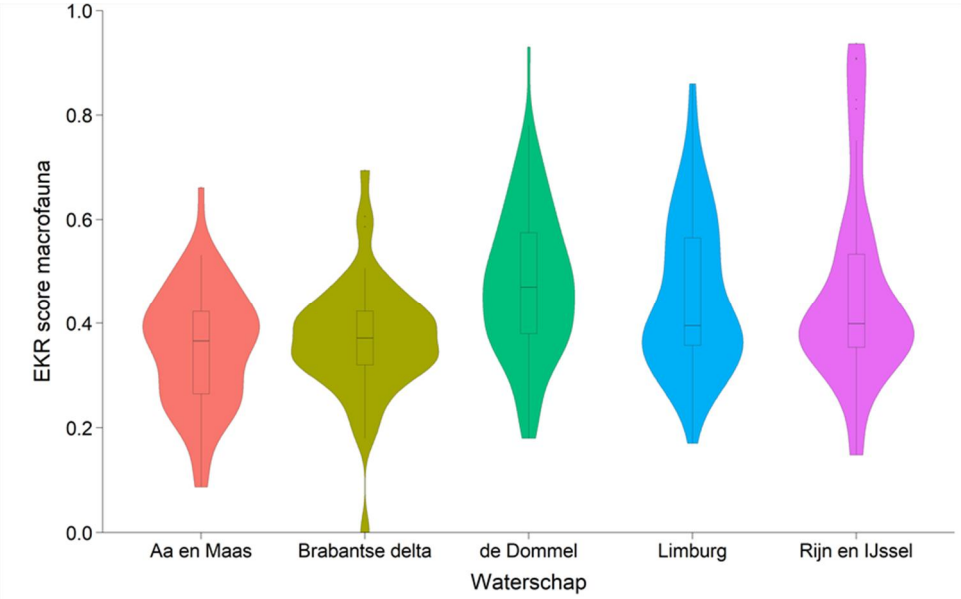
De ruimtelijke variatie in stroomsnelheid is samengesteld uit twee stuurvariabelen die door de waterschappen was bepaald, waarbij de categorische verdelingen en numerieke verdelingen zijn vertaald naar een categorische verdelingen (vier categorieën). Waterschap Aa en Maas, waterschap Brabantse Delta en waterschap Limburg hadden de ruimtelijke variatie in stroomsnelheid bepaald terwijl waterschap de Dommel en waterschap Rijn en IJssel de actieve sedimentatie/erosie aanwezig over percentage van de lengte hadden bepaald. Op basis van de onderstaande verdeling is de ecologische stuurvariabele Ruimtelijke variatie in stroomsnelheid gestandaardiseerd (Tabel 2-6).

Tabel 2-6 Invulling van categorische ecologische stuurvariabele Ruimtelijke variatie in stroomsnelheid per waterschap

Gestandaar-diseerde variabele	Waterschap Aa en Maas	Waterschap Brabantse Delta	Waterschap de Dommel	Waterschap Limburg	Waterschap Rijn en IJssel
Geen	Geen	Geen	<10	Geen/gering	<10
Weinig	Weinig	Weinig	10-25		10-25
Matig	Matig	Matig	25-75	Matig	25-75
Veel		Veel	>75	Veel	>75

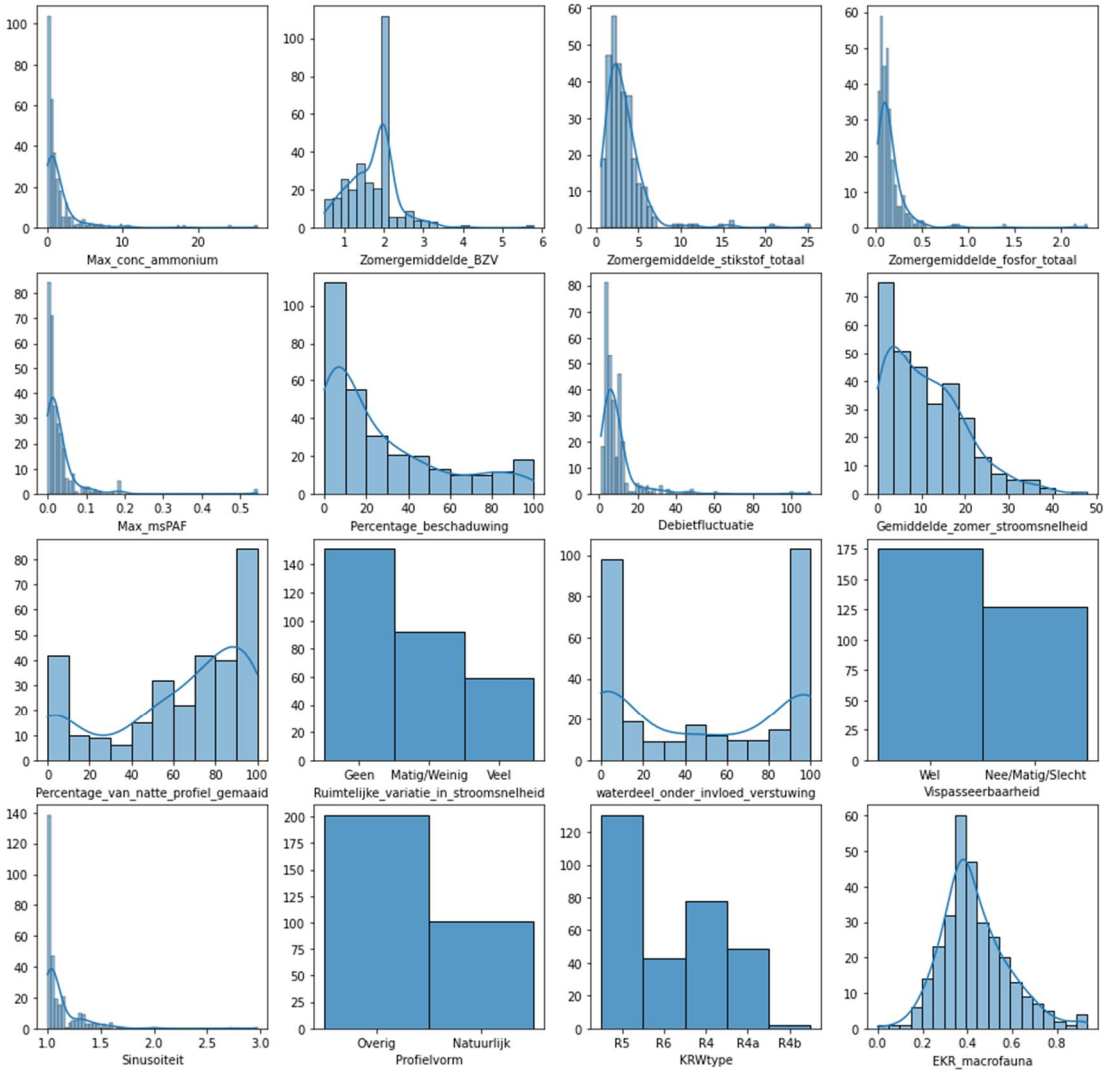
2.1.3 Overzicht van de resulterende dataset

In Figuur 2-1 is de verdeling van de EKR-scores voor macrofauna tussen de waterschappen weergegeven. Het overgrote deel van de records heeft een EKR-score tussen de 0.20 en 0.60. Waardes onder de 0.20 komen voornamelijk voor bij waterschap Aa en Maas en waterschap Brabantse Delta terwijl waardes boven de 0.60 voornamelijk voorkomen zijn waterschap de Dommel, waterschap Limburg en waterschap Rijn en IJssel.



Figuur 2-1 Verdeling van EKR-scores voor macrofauna tussen de vijf waterschappen.

De waarden van de numerieke stuurvariabelen in de dataset zijn weergegeven in Figuur 2-2, inclusief de verdeling van EKR-scores van de gezamenlijke waterschapsdata. De histogrammen laten veelal een scheve verdeling zien van de waarden. De scheve verdeling is het best zichtbaar in de stuurvariabele “waterdeel_onder_invloed_verstuwing” waarin blijkbaar óf het waterlichaam niet verstuwd is, óf in zijn totaliteit wel. Zomergemiddeld Biologisch Zuurstofverbruik (BZV) heeft rond de waarde 2 veel waarnemingen. Dit komt doordat deze waarde is ingevuld voor de trajecten waarvan geen BZV-metingen beschikbaar waren (Rost et al., 2019; Rost et al., 2020).



Figuur 2-2 Histogrammen van de nieuwe set stuurvariabelen (alle waterschappen samen). Op de y-as is het aantal datapunten zichtbaar met waarden zoals op de x-as per stuurvariabele weergegeven. Op EKR_macrofauna na, zijn de meeste verdelingen niet gelijk verdeeld (scheef).

2.2 Training methoden

Er zijn verschillende algoritmen beschikbaar voor het trainen van modellen. De random forest Ranger methode is de huidige methode die nu gebruikt wordt in de KRW-Verkenner. De laatste update van de kennisregels voor de KRW-verkenner (van der Linden et al., 2021) en het wetenschappelijk artikel van Visser et al. (2021) wezen uit dat het random forest algoritme het beste model is voor de hier gekozen toepassing. Omdat er nu een meer gedetailleerde dataset beschikbaar is dan in het onderzoek van Visser et al. (2021) wordt de Random forest Ranger methode vergeleken met een aantal andere algoritmen.

De modellering is uitgevoerd in met behulp van specialistische bibliotheken (packages) van Python en R (versie py 3.7.12 en r 3.6.3, Table 2-1). Belangrijkste Python bibliotheken voor de modellen zijn sklearn, statsmodels, xgboost, shap, umap en rpy2. Daarnaast zijn de volgende R- bibliotheken gebruikt: ranger, tuneRanger, mlr en caret.

Table 2-1. Overzicht van training methoden met de gebruikte packages.

Algoritme	Python / R	Package(s)
Linear regression	Python	sklearn, statsmodels
Random forest	Python	sklearn
'Ranger' random forest	R	ranger
Extra trees regression	Python	sklearn
Gradient boosting regression	Python	sklearn
Extreme gradient boosting	Python	xgboost
KNN regression	Python	sklearn
Support vector regression	Python	sklearn
Decision trees regression	Python	sklearn

2.3 Hoe te testen en vergelijken?

2.3.1 Statistische methoden (metrics)

Om de kwaliteit van het getrainde model te bepalen, zijn een vijftal statistische methoden of metrics gebruikt. Deze metrics zijn gangbare methoden om getrainde modellen met elkaar te vergelijken, en komen grotendeels overeen met de gebruikte evaluatiemethodes in van der Linden et al. (2021) en Visser et al. (2021). De gebruikte vijf metrics zijn:

- Het kwadraat van de correlatiecoëfficiënt (R^2). Een grotere R^2 betekent een betere model prestatie.
- Fractie binnen een bandbreedte van ±0.10 EKR. Een grotere fractie binnen de bandbreedte betekent een betere modelprestatie.
- Root Mean Square Error (RMSE). Een kleinere RMSE betekent een betere modelprestatie.
- Coefficient of Determination (CoD). Een grotere CoD betekent een betere modelprestatie.
- Mean absolute Scaled Error (MASE). Een kleinere MASE betekent een betere modelprestatie.

Meer achtergrondinformatie over de gebruikte metrics is te vinden in Hyndman & Athanasopoulos (2018) en Hyndman & Koehler (2006).

2.3.2 Data-analyse

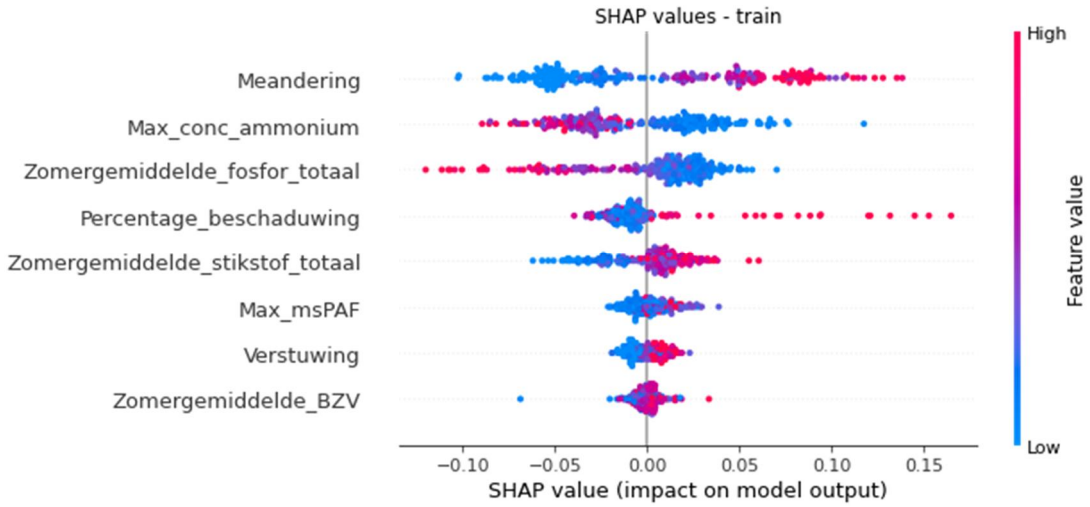
Om de modelresultaten beter te begrijpen is een aantal illustratieve figuren gemaakt.

Feature importance plot
De feature importance plot geeft weer in welke mate een stuurvariabele bijdraagt aan de voorspelde EKR-score. De zogenaamde 'permutation feature importance' is een maat voor de toename in voorspelfout van het model wanneer de relatie tussen een stuurvariabele en de werkelijke uitkomst (EKR-score) verbroken ('gepermuteerd') wordt. Hierbij moet wel bedacht worden dat de andere stuurvariabelen in het model een deel van de verslechtering in voorspelresultaat kunnen opvangen. Feature importance is dus geen 1-op-1 relatie tussen stuurvariabele en EKR-score.

Partial dependence plot
Deze plot geeft de relatie van de EKR-score met de stuurvariabelen weer. Hiervoor wordt voor iedere stuurvariabele 20 regelmatige punten bepaald tussen de minimale en maximale waarde in de trainingset, waarbij de andere stuurvariabelen op een referentiewaarde zijn gezet. Vervolgens worden de berekende EKR weergegeven voor de range van de stuurvariabele. Dit geeft dus een beeld van hoe de EKR-score reageert op het veranderen van een enkele stuurvariabele.

SHAP value plot
De SHAP value plot (Figuur 2-3) vormt een aanvulling op de feature importance en laat de positieve en negatieve relaties tussen stuurvariabelen en EKR-score zien. De plot laat alle punten in de dataset zien en verschaft de volgende informatie:

1. *Feature importance*: variabelen zijn geordend op volgorde van belangrijkheid;
2. *Impact*: de horizontale locatie van een waarde laat zien of het effect van die waarneming (datapunt) verbonden is aan een hogere of lagere voorspelling;
3. *Originele waarde*: de kleur van elk punt laat zien of de originele waarde hoog (rood) of laag (blauw) is;
4. *Correlatie*: een hoge waarde van 'Meandering' heeft een grote positieve invloed op de EKR-score. 'Groot' komt uit de rode kleur, de 'positieve invloed' volgt uit de waarden op de x-as. 'Max_conc_ammonium' is juist negatief gecorreleerd met de EKR-score.



Figuur 2-3 Voorbeeld SHAP value plot. Stuurvariabelen (links) staan gerangschikt op belangrijkheid. De SHAP value (x-as) laat zien of het datapunt verbonden is aan een hogere of lagere voorspelling. De kleurstelling van de datapunten laat zien of het datapunt zelf een hoge of lage waarde heeft binnen de range van een stuurvariabele.

2.3.3 Vergelijkingen van datasets

Om een beeld te krijgen van het effect van het toevoegen van meer en gedetailleerde variabelen en het toevoegen van meer waarnemingen zijn verschillende datasets met elkaar vergeleken. Deze staan beschreven in Tabel 2-7. In Tabel 2-8 staat beschreven welke datasets met elkaar zijn vergeleken om verschillende onderzoeksvragen te beantwoorden.

Tabel 2-7 Gebruikte datasets in de analyse.

Naam dataset	Beschrijving
Visser	Dataset zoals gebruikt en beschreven in Visser et al. (2021) voor macrofauna in langzaam stromende beken. Deze wordt gebruikt in de huidige versie van de KRW-Verkenner voor het voorspellen van EKR-scores voor regionale wateren en bevat samengestelde variabelen voor meandering en verstuwing.
KIWK-dataset	De gedetailleerde dataset die beschreven staat in paragraaf 2.1 met meer variabelen en meer waarnemingen. De variabelen voor meandering en verstuwing zijn opgesplitst in afzonderlijke variabelen.
KIWK-getrimd	De KIWK-dataset maar met de samengestelde variabelen Meandering en Verstuwing zoals die in de Visser dataset zijn meegenomen.
Visser aangevuld	Dataset 'Visser' aangevuld met waarnemingen uit de KIWK-getrimd dataset waarbij de dubbelingen verwijderd zijn.

Tabel 2-8 Overzicht van welke datasets met elkaar vergeleken zijn voor welk doeleinde.

Getest	Referentie set	Nieuwe set	Opmerking
Meer en gedetailleerdere stuurvariabelen	KIWK-getrimd	KIWK-dataset	
Toevoegen van meer waarnemingen	Visser et al. (2021)	Visser aangevuld	Hiervoor is dezelfde testset gebruikt en bij Visser aangevuld een grotere trainingset
Vergelijken modelprestaties	Visser et al. (2021)	KIWK-getrimd & KIWK-dataset	

3 Resultaten

We hebben de vragen zoals gesteld in de introductie kunnen beantwoorden. Voor de leesbaarheid tonen we hier alleen de resultaten die de meest directe antwoorden op de hoofdonderzoeksvragen laten zien. We realiseren ons dat dit een vrij technisch hoofdstuk is. In het hoofdstuk conclusies en aanbevelingen is getracht de resultaten eenvoudig weer te geven en direct conclusies te trekken. Een uitgebreide visuele verslaglegging staat in een afzonderlijke bijlage (PowerPoint). Hierin worden ook resultaten van voorbereidend onderzoek getoond, zoals hoe we met bestaande instrumenten en data de vragen het eerlijkst kunnen beantwoorden.

3.1 Prestatie van verschillende machine-learning methoden

3.1.1 Welk machine learning model presteert het best?

Conclusie: op de dataset met samengevoegde variabelen presteert Extreme Gradient Boosting (XGB) significant beter dan de andere algoritmes. De prestaties van Gradient Boosting Regression (GBR), Random Forest Breiman (RF) en Ranger Random Forest (RNG) zijn vergelijkbaar. Op de gedetailleerde KIWK-dataset presteert Gradient Boosting Regression (GBR) significant beter dan de andere algoritmes.De prestatie van Extreme Gradient Boosting (XGB) is vergelijkbaar.

Tijdens deze studie zijn verschillende machine-learning methoden onderzocht op hun voorspellende kracht, deze staan beschreven in paragraaf 2.2. In eerste instantie is gekeken of dezelfde algoritmen het beste uit de bus komen als in de Visser dataset (Visser et al., 2021). Hiervoor is de getrimde KIWK-dataset gebruikt. De drie best scorende modellen zijn gebaseerd op gradient boosting (XGB; Extreme gradient boosting) en random forest (RF en RNG; resp. Breiman en Ranger). In de studie van Visser et al. (2021) werd geconcludeerd dat een Ranger random forest de beste prestaties levert.

Om te komen tot de resultaten in de tabel is de volgende procedure toegepast:

- 5-voudige crossvalidatie om hyperparameters van de modellen te optimaliseren;
- Gestratificeerde train-test split om een zo evenwichtig mogelijke verdeling van de punten te bereiken;
- Beoordeling van totale prestatie (overall performance) op basis van de mediaan van 500 random train-test splits om de invloed van verschillende train-test splits te verdisconteren.

Voor de meeste metrics presteert XGB het beste, maar de prestaties van GBR, RF en RNG zijn redelijk vergelijkbaar. De prestaties van de modellen zijn op basis van een sample van 500 modelevaluaties met verschillende train en test sets paarsgewijs vergeleken. Hierbij is de Conover test voor meervoudige paarsgewijze vergelijking gebruikt. De modelprestatie heeft in termen van RMSE dezelfde orde van grootte als de modellen in Visser et al. (2021).

Tabel 3-1 Model performance voor alle geteste modellen voor de getrimde KIWK-dataset. LR = linear regression, RF = random forest (Breiman), XT = Extreme trees regression, GBR = Gradient boosting regression, XGB = Extreme gradient boosting, KNN = KNN regression, SVR = Support vector regression, DT = decision tree, RNG = Ranger random forest.

	LR	RF	XT	GBR	SVR	KNN	DT	XGB	RNG
RMSE	0.11895	0.11660	0.11860	0.11680	0.12820	0.12060	0.13010	0.11525	0.11600
COD	0.35300	0.38245	0.35405	0.37435	0.26780	0.33395	0.22960	0.39270	0.38310
R2	0.36360	0.39255	0.36465	0.38460	0.27980	0.34485	0.24225	0.40265	0.39325
MASE	0.78540	0.77505	0.79460	0.78075	0.86045	0.79965	0.86240	0.76685	0.77390
CORR	0.62565	0.64200	0.61970	0.63970	0.57140	0.61075	0.52795	0.64935	0.64165
+/-0.1	0.63930	0.65570	0.62300	0.63930	0.59020	0.62300	0.59020	0.65570	0.63930

Vervolgens is gekeken welk algoritme het beste presteert als de volledige gedetailleerde KIWK-dataset gebruikt wordt. Voor deze dataset presteerde Gradient Boosting Regression (GBR) het beste.

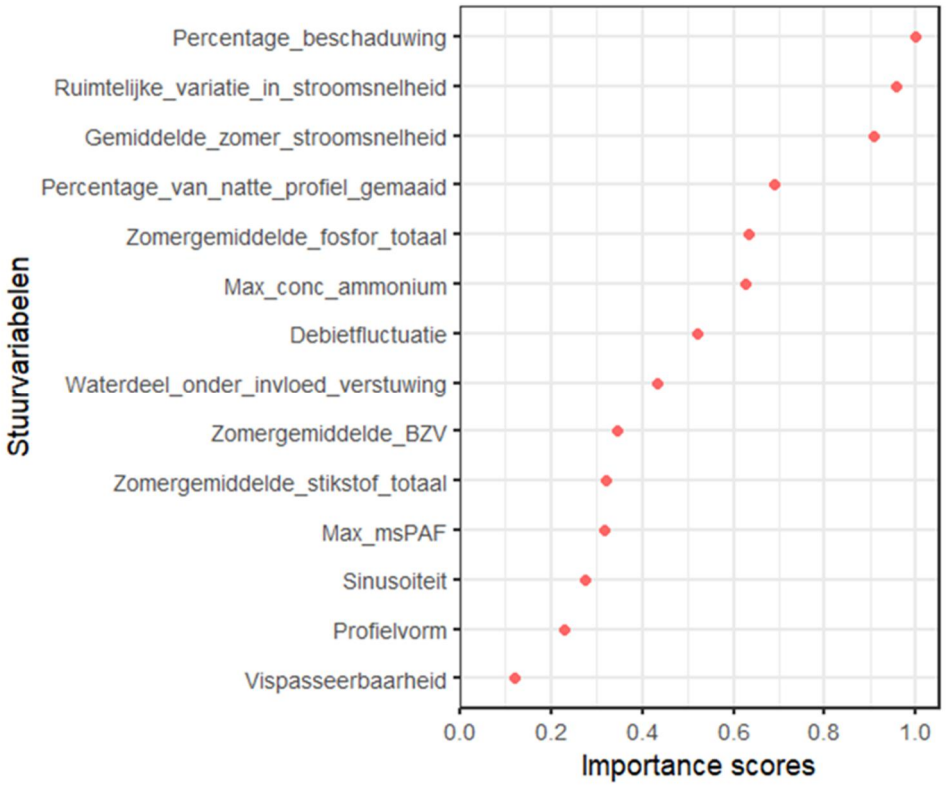
Tabel 3-2 Model performance voor alle geteste modellen voor de KIWK-dataset. LR = linear regression, RF = random forest (Breiman), XT = Extreme trees regression, GBR = Gradient boosting regression, XGB = Extreme gradient boosting, KNN = KNN regression, SVR = Support vector regression, DT = decision tree, RNG = Ranger random forest.

	LR	RF	XT	GBR	SVR	KNN	DT	XGB	RNG
RMSE	0.1098	0.1052	0.1068	0.1022	0.1116	0.111	0.1197	0.1021	0.1039
CoD	0.4461	0.4767	0.4701	0.512	0.4039	0.4189	0.329	0.5104	0.4911
R2	0.4552	0.4852	0.4788	0.52	0.4137	0.4284	0.34	0.5184	0.4995
MASE	0.7429	0.711	0.7198	0.6959	0.7724	0.7534	0.8082	0.6975	0.7018
CORR	0.6916	0.708	0.7018	0.7337	0.6785	0.6742	0.6178	0.7304	0.7194
+/-0.1	0.6557	0.6885	0.6721	0.6885	0.623	0.6557	0.623	0.6885	0.6885

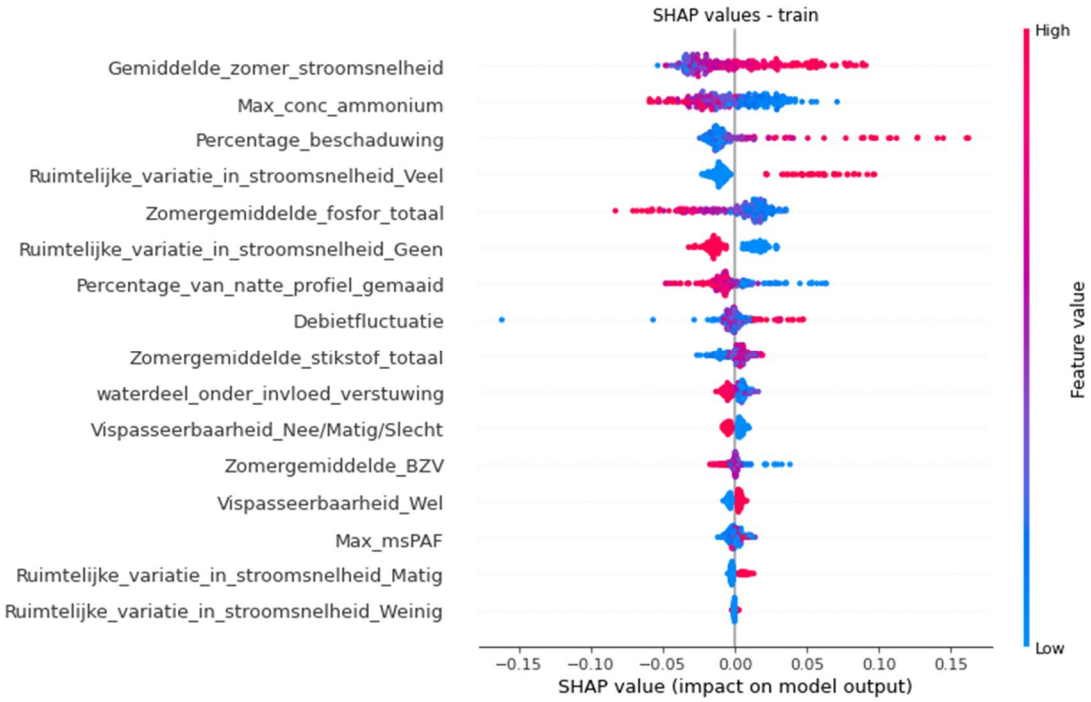
3.2 Welke parameters zijn belangrijk in de nieuwe en oude dataset?

<u>Conclusie: de stuurvariabelen beschaduwing, stroomsnelheid, fosfor, ammonium en beheer leveren de belangrijkste bijdrage aan het model.</u>

In Figuur 3-1 en Figuur 3-2 zijn de stuurvariabelen gerangschikt op hun bijdrage aan de training en het testen van het model. Het berekenen van de 'belangrijkheid' van stuurvariabelen kan op verschillende manieren en geeft dan een iets ander resultaat. Op hoofdlijn is het beeld hetzelfde in beide figuren. Van de top 7 stuurvariabelen zijn er 6 hetzelfde. De stuurvariabelen 'percentage beschaduwing', 'de ruimtelijke variatie in stroomsnelheid', de 'gemiddelde zomer stroomsnelheid' en in mindere mate 'zomergemiddeld fosfor', 'max concentratie ammonium' en het 'percentage van natte profiel gemaaid' leveren een belangrijke bijdrage aan het model. Vispasseerbaarheid en profielvorm dragen weinig bij. Vispasseerbaarheid is vanzelfsprekend voor macrofauna natuurlijk ook minder relevant. Profielvorm betreft een categorische variabele die mogelijk niet onderscheidend genoeg is en inhoudelijk ook grotendeels wordt gedekt door de ruimtelijke variatie in stroomsnelheid die wel continu is en duidelijk relevanter is. Sinuositeit scoort ook laag (en is daarom niet meegenomen in Figuur 3-2).

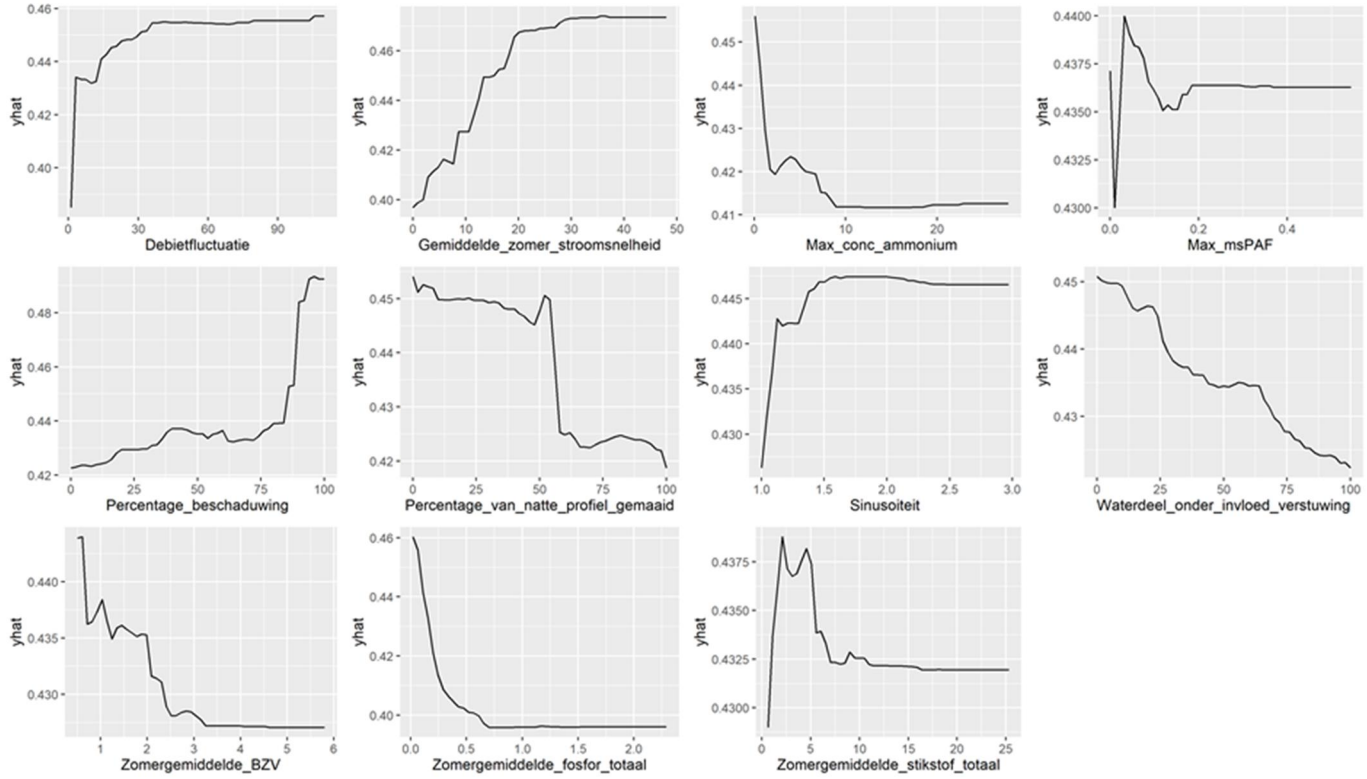


Figuur 3-1 De bijdrage van iedere stuurvariabele aan het model op basis van importance scores uit de Random Forest training. De stuurvariabelen zijn gerangschikt van grootste naar kleinste bijdrage.



Figuur 3-2 De bijdrage van iedere stuurvariabele aan het model op basis van SHAP values. De stuurvariabelen zijn gerangschikt van meeste naar minste bijdrage. SHAP feature values laten zien hoe 'belangrijk' een waarneming is voor de training. In deze figuur zijn categorische stuurvariabelen opgesplitst per categorie. Bijvoorbeeld, de 'ruimtelijke variatie in stroomsnelheid' is nu in 4 aparte variabelen weergegeven '..._geen', '..._weinig', '..._matig' en '..._veel'.

De partial dependence plots laten het verwachte beeld zien voor de ecologische respons op de meeste stuurvariabelen (Figuur 3-3). Uitzondering daarop is zomergemiddelde stikstof totaal waar een toename in concentratie eerst leidt tot een toename in de EKR waarna deze weer afneemt. Dit is een relatie die ook in de KRW-verkenner werd gevonden en waarschijnlijk het gevolg is van enkele datapunten in de dataset. Er zijn enkele beken waar zowel de zomergemiddelde totaal-stikstof concentraties als de EKR-scores hoog zijn. Dit zijn beken op zandgronden waar het verhang groot is. Dat is waarschijnlijk de reden is voor de hoge EKR-score zoals zichtbaar in de partial dependence plot voor gemiddelde zomer stroomsnelheid. In het model wordt deze verhoging daardoor naar alle waarschijnlijkheid onterecht ook deels aan de hogere stikstofconcentratie toegeschreven.



Figuur 3-3 Partial dependence plot voor alle numerieke stuurvariabelen met op de y-as de EKR-score. De lijn geeft het verloop per stuurvariabele aan waarbij de andere stuurvariabelen op referentiewaarde worden gehouden.

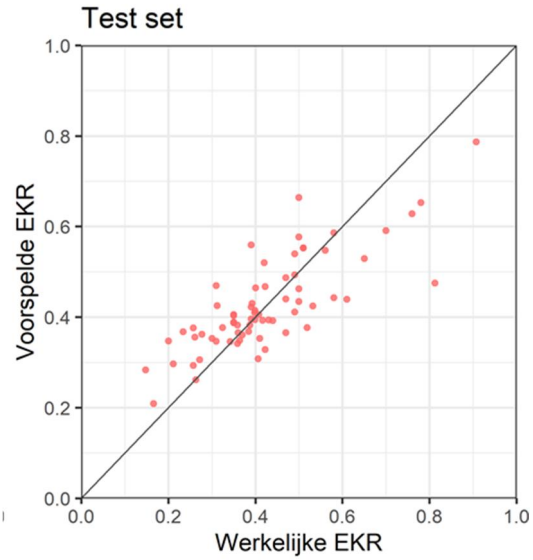
3.3 Verbeteren de modelprestaties als de samengestelde stuurvariabelen worden opgesplitst in variabelen die een eenduidiger relatie hebben met de maatregelen?

Conclusie: de modelresultaten verbeteren statistisch significant als de samengestelde stuurvariabelen worden opgesplitst. Voor alle prestatiecriteria is een verbetering te zien.

We presenteren eerst een vergelijking tussen de Ranger random forest resultaten op basis van de dataset uit Visser et al. (2021) en de nieuwe KIWK-dataset. Dit betreft een enkelvoudige berekening op basis waarvan geen uitspraken over significantie van verschillen gedaan kunnen worden. Daarna laten we de resultaten zien van de vergelijking van het beste algoritme voor de KIWK-dataset met samengestelde en de KIWK-dataset met opgesplitste stuurvariabelen op basis van een sample van 500 train-test splits.

3.3.1 Prestaties Ranger random forest

De voorspelde EKR-scores zijn uitgezet tegen de werkelijke EKR-scores (Figuur 3-4). Als het model toegepast wordt op KIWK-dataset dan blijkt het model de EKR-scores goed te kunnen voorspellen in de midden-range, maar de voorspelling aan de uiteindes is minder goed. De hoge EKR-scores worden onderschat en de lage EKR-scores worden overschat. Dit komt voor een deel doordat in uiteindes van de range van de EKR-scores (tussen 0.00 en 0.20 en tussen 0.80 en 1.00) weinig waardes zijn, waardoor het model sterk door slechts enkele waardes wordt gestuurd in deze zones.



Figuur 3-4 Werkelijke EKR's (x-as) en voorspelde EKR's (y-as) voor de testset. De zwarte lijn geeft de 1-op-1 lijn aan.

In tabel 3.3 zijn de kwaliteitsscores weergegeven van de uitgebreide KIWK-dataset en van de dataset in Visser et al. (2021). Voor alle prestatiecriteria is een verbetering te zien. Alleen kan op basis van dit resultaat nog niet geconcludeerd worden of het nieuwe model met opgesplitste stuurvariabelen beter presteert dan het model met samengestelde variabelen. Daar zijn verschillende redenen voor: (1) de dataset van Visser et al. (2021) verschilt van de KIWK-dataset, (2) de train-test split die op beide datasets is toegepast is verschillend en (3) de random seed waarmee de random forests zijn doorgerekend is ook anders. Dit is ondervangen door bij de best presterende methode op de uitgebreide KIWK-dataset (GBR, paragraaf 3.3.2) een mediaan te nemen van 500 train-test samples.

Tabel 3-3 Overzicht van kwaliteitsscores van Random Forest voor de testdata.

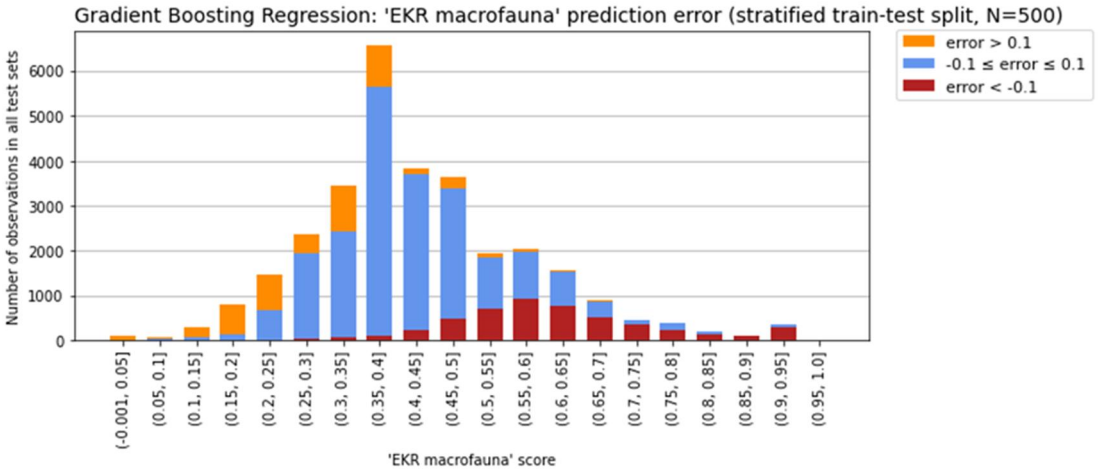
	R^2	Fractie binnen ±0.10	RMSE	CoD
Visser et al. (2021)	0.589	0.591	0.122	0.548
KIWK-dataset	0.625	0.725	0.091	0.616

3.3.2 Prestaties gradient boosting regression

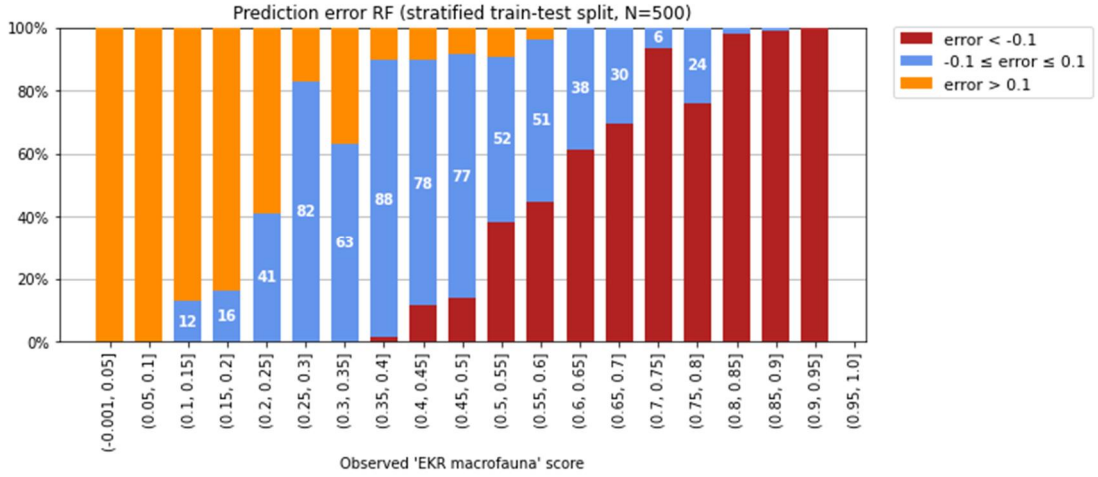
Figuur 3-5 en Figuur 3-6 geven een overzicht van de voorspelfout van het best presterende model (GBR) als de gedetailleerde dataset wordt gebruikt. Dit resultaat is gebaseerd op een sample van 500 modelevaluaties met verschillende train en test sets. Dit laat zien dat de EKR-scores in de range van 0.3 en 0.6 (blauwe balken) goed voorspeld kunnen worden, maar dat het model een onderschatting geeft bij lagere EKR-scores (oranje balken) en een overschatting bij hogere EKR-scores (rode balken). Het resultaat is vergelijkbaar met Figuur

3-4 waarin ook de overschatting bij lage waarden en de onderschatting bij hogere waarden zichtbaar is.

Dit betekent dat waterlichamen die een lage gemeten EKR-score hebben op basis van de waardes van hun stuurvariabelen automatisch een relatief hoge EKR-score voorspeld zullen krijgen. Omgekeerd zullen waterlichamen die een hoge gemeten EKR-score hebben op basis van hun stuurvariabelen, juist een lagere EKR-score voorspeld krijgen dan in werkelijkheid het geval is.



Figuur 3-5 De voorspelfout (+/- 0.1 EKR) per EKR-score klasse van 0.05. De voorspelfout is bepaald door 500 keer random een gestratificeerde test/training set te kiezen. Het aantal waarnemingen per klasse is uitgezet (y-as) waarbij in kleur is aangegeven in welke foutklasse de waarnemingen vallen.



Figuur 3-6 Voorspeelfout (=/- 0.1 EKR) als percentages weergegeven voor Random Forest. De voorspelfout is bepaald door 500 keer random een gestratificeerde test/training set te kiezen. In tegenstelling tot figuur 3-5 is het aantal waarnemingen naar 100% geschaald. In de blauwe balken staat het percentage dat per klassebreedte correct voorspeld wordt.

Tabel 3-4 laat de resultaten zien voor de met GBR getrainde modellen van de dataset met oorspronkelijke stuurvariabelen (KIWK-getrimd) en de meer gedetailleerde variabelen (KIWK-dataset).

Tabel 3-4 GBR resultaten vergeleken tussen KIWK getrimd en KIWK-dataset (stratified train-test spilt, N=500).

	R^2	Fractie binnen ±0.10	RMSE	CoD	MASE
KIWK-getrimd	0.38460	0.63930	0.11680	0.37435	0.78075
KIWK-dataset	0.51995	0.68850	0.10215	0.51195	0.69585

M.b.v. de Kruskal-Wallis test is getoetst of de GBR-modellen getraind met de oorspronkelijke samengestelde stuurvariabelen (KIWK-getrimd) en de dataset met meer en gedetailleerdere stuurvariabelen (KIWK-dataset) al dan niet significant verschillen. De nul-hypothese (H0) is dat 'beide kansverdelingen identiek zijn'. In onderstaande tabel zijn de resultaten van de Kruskal-Wallis test samengevat. Hieruit blijkt dat op alle metrics de modelprestatie significant verbetert bij toepassing van de dataset met meer en gedetailleerdere stuurvariabelen (KIWK-dataset).

Tabel 3-5 Overzicht van Kruskal-Wallis test met als H0 dat de kansverdeling identiek is. De H0 wordt verworpen als p < 5.0E-2.

metric	KIWK-getrimd	KIWK-dataset	reject_HO	pval
RMSE	GBR	GBR	True	2.63710E-75
CoD	GBR	GBR	True	5.43609E-68
R2	GBR	GBR	True	5.52003E-68
MASE	GBR	GBR	True	1.33090E-63
+/- 0.1	GBR	GBR	True	1.04611E-33

3.4 Verbeteren de modelresultaten als meer datapunten worden toegevoegd?

Conclusie: het toevoegen van meer datapunten verbetert de modelprestatie niet als deze datapunten van vergelijkbare typen waterkwaliteit komen (i.e. in dezelfde range zitten).

Om dit te testen is een model getraind met dezelfde trajecten en variabelen uit Visser et al. (2021) en een model getraind met de dataset uit Visser et al., gecombineerd met de getrimde gedetailleerde dataset uit deze studie (KIWK-getrimd). Er is gecontroleerd of datapunten niet dubbel in de gecombineerde dataset zitten. De gecombineerde dataset bevat dus dezelfde parameters als Visser et al. (2021), maar meer trajecten.

De modelprestaties van beide modellen zijn met elkaar vergeleken, zie Tabel 3-6. Zowel voor Gradient boosting regression als voor Ranger random forest verbeteren de modelprestaties niet of nauwelijks. Alle metrics laten nagenoeg hetzelfde beeld zien.

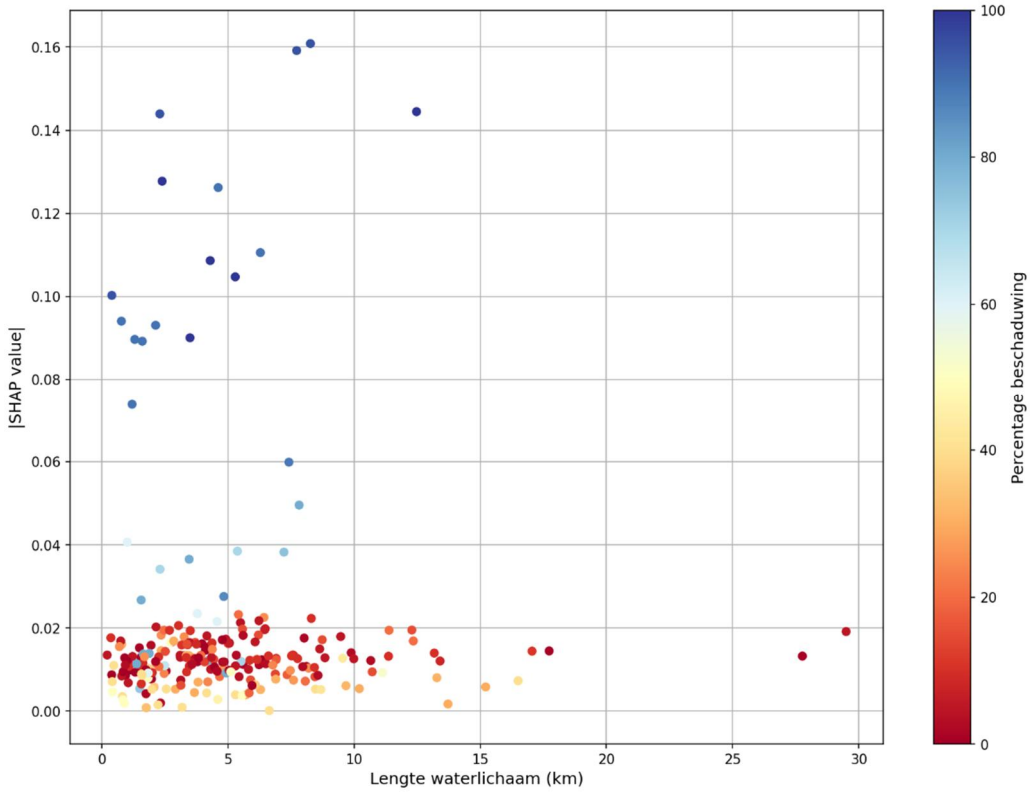
Tabel 3-6 Vergelijk van trainingsresultaten tussen modellen met kleine dataset (Visser) en grote dataset (Visser+KIWK-getrimd) voor metrics RMSE, CoD, R2, MASE en bandbreedte +/- 0.1.

	RMSE	CoD	R^2	MASE	+/- 0.1
GBR (Visser)	0.1151	0.5869	0.5963	0.6155	0.5909
GBR (Visser+KIWK-getrimd)	0.1162	0.5789	0.5884	0.6149	0.5682
RNG (Visser)	0.1182	0.5646	0.5745	0.6165	0.5455
RNG (Visser+KIWK-getrimd)	0.1179	0.5668	0.5766	0.6353	0.5

3.5 Zijn modeluitkomsten gecorreleerd met de ruimtelijke omvang van segmenten?

Conclusie: een kleinere segmentlengte leidt niet automatisch tot een betere modeltraining.

De waarden van stuurvariabelen zijn gekoppeld aan delen van beken en kleine rivieren. De indeling en grootte van segmenten is bepaald door de waterschappen en is een gegeven. Als de dataset uit kleinere segmenten bestaat, dan beschrijven de stuurvariabelen een kleiner ruimtelijk deel en wellicht minder 'ruimtelijke middeling' van de stuurvariabel-waarde. Het zou een betere koppeling kunnen zijn met de gemeten biota, in ons geval macrofauna.



Figuur 3-7 Relatie tussen de 'bijdrage van een waarneming aan de model training' (SHAP waarden) en de lengte van een waterlichaam (segmentgrootte).

Figuur 3-7 laat de relatie zien tussen segmentlengte en de SHAP-value van waarnemingen van de stuurvariabele 'percentage_beschaduwing'. De SHAP-value is een maat voor de bijdrage van een segment aan de totale modeltraining voor een enkele stuurvariabele. De figuur laat een grote spreiding zien zonder een duidelijke correlatie met de lengte van een waterlichaam (segment). Voor andere stuurvariabelen geldt hetzelfde resultaat (niet afgebeeld). Hieruit leiden we af dat een kleinere segmentlengte niet automatisch tot een betere modeltraining leidt. Daar tegenover staat dat kleinere segmenten ook leidt tot een meer datapunten waarmee het model wel beter getraind kan worden, maar dat is hier niet getest. Los daarvan laat de figuur wel zien dat lagere waarden van de stuurvariabele "percentage_beschaduwing" minder waarde hebben voor de modeltraining (op basis van macrofauna data), en onderschrijft de conclusie in paragraaf 3.2 dat beschaduwing een relevante ecologische parameter is.

4 Conclusies en aanbevelingen

Wat zijn de hoofdconclusies uit de resultaten? En wat betekenen de resultaten en conclusies voor de verdere ontwikkeling van de ecologische kennisregels van de KRW-Verkenner? Hieronder staat per onderzoeksvraag de conclusie, de betekenis van het resultaat voor de KRW-verkenner, en de concrete aanbeveling (doorgenummerd). De voorgestelde aanbevelingen volgen zowel direct uit de conclusies als uit discussie tussen de auteurs over het resultaat.

4.1 Verbetert de voorspelprestatie na toevoegen van meer en eenduidiger stuurvariabelen?

Sommige van de stuurvariabelen van de huidige rekenregels in de KRW-Verkenner zijn samengesteld en niet eenduidig, zie hiervoor de beschrijvingen in de introductie (H1) en methoden (H2). De hypothese is dat het model beter kan worden getraind en de modelprestaties vooruitgaan, als elke stuurvariabele daadwerkelijk één type sturing beschrijft. De gebruikte stuurvariabelen zijn te vinden in paragraaf 2.1.

- Uit de analyse blijkt dat inderdaad de modelprestatie significant verbetert door het toevoegen van meer en gedetailleerdere variabelen.
- EKR-scores worden het beste voorspeld tussen waardes van 0.3 en 0.6. Boven een EKR-score van 0.6 worden ze meestal onderschat en onder de 0.3 overschat.

Wat betekent dit voor de verkenner?

Met eenduidigere stuurvariabelen is het ook gemakkelijker uit te leggen aan gebruikers wat precies bedoeld wordt. Daarnaast is de aansluiting van eenduidigere stuurvariabelen op maatregelen eenvoudiger en gemakkelijker te kwantificeren. Ze vragen weliswaar meer informatie over het systeem, maar de meeste waterbeheerders verzamelen al veel informatie op dit niveau. Bijvoorbeeld voor KRW-watersysteemanalyses met de Ecologische Sleutelfactoren van de STOWA of met SESA (Verdonschot & Verdonschot, 2021).

Wat zijn de aanbevelingen die hieruit volgen?

1. De gebruikte stuurvariabelen bij andere watertypen en kwaliteitselementen goed screenen op samengesteldheid en eenduidigheid. Wanneer andere stuurvariabele data ook daadwerkelijk beschikbaar is, dan de set stuurvariabelen verbeteren en bijbehorende modellen trainen en valideren. Een belangrijk voorbeeld hiervan is oeverinrichting bij de clusters van M-typen.

2. In de huidige rekenregels hanteren we voor alle biologische kwaliteitselementen binnen een watertypecluster dezelfde set stuurvariabelen. Het verdient aanbeveling om dat te heroverwegen. Vismigreerbaarheid is bijvoorbeeld minder van belang voor macrofauna en planten. Verbanden zonder causale relatie worden door het meenemen van niet relevante stuurvariabelen ongewenst toch gelegd.

3. Het effect van maatregelen opnieuw koppelen aan de nieuwe set stuurvariabelen.

4. In zijn algemeenheid is het aan te bevelen voor alle watertypen en kwaliteitselementen te laten zien welke stuurvariabelen vanuit de ecologie gewenst zijn. Deze lijst stuurvariabelen kan dan vergeleken worden met de huidige implementatie om vervolgens met de waterschappen te overleggen of monitoring op dit vlak kan worden verbeterd om betere koppelingen te kunnen leggen met ecologie. De rapportage van Van der Lee et al. (2022) geeft richtlijnen voor doelgerichtere monitoringsopzet.

5. Al geïdentificeerde missende stuurfactoren op basis van interne discussie: waterbodemkwaliteit, sulfaat, ijzer en alkaliniteit.

4.2 Dragen kleinere trajecten meer bij aan de modeltraining?

De stuurvariabelen zijn een gemiddelde weergave van de toestand van een traject. De macrofaunasamenstelling is een afspiegeling van lokale milieuomstandigheden. Hieruit volgt de hypothese is dat het gebruik van kleinere trajecten een homogenere beschrijving kan geven van de lokale abiotische omstandigheden en zo wellicht beter correleert met ecologische indicators zoals de KRW-maatlat score, mits er ook op dezelfde kleinere trajectgrootte biotische data beschikbaar is.

- We hebben getest of kleinere trajecten een belangrijkere bijdrage in de training van een model hebben dan grotere trajecten (paragraaf 3.5). We konden met onze dataset geen verschil vinden in de bijdrage van grotere of kleinere trajecten op basis van trajectlengte op de training van het model. Dit komt omdat in de huidige set de opknipping zover is doorgevoerd dat een dusdanig hoge homogeniteit binnen het traject wordt behaald dat de koppeling met het inliggende macrofaunameetpunt (of meetpunten) vrijwel optimaal is. Let op, het indelen van gebieden in kleinere eenheden kan wel tot meer data leiden, waarmee wellicht een betere modelprestatie kan worden bereikt, dit is afzonderlijk getest, zie volgende paragraaf (4.3).

Wat betekent dit voor de verkenner?

In de praktijk is het waarschijnlijk zo dat de langere trajecten juist homogeen zijn, bijvoorbeeld zijn het trajecten met een gegraven of vastgelegde oever, en daarom door waterschappen niet zijn opgeknipt (pers. comm. Niels Evers). Let op, bij de toepassing van de KRW-verkenner kan een kleine trajectgrootte wel beter zijn, zeker als te toetsen maatregelen slechts een deel van het traject beïnvloeden. Door ruimtelijke middeling van stuurfactoren kan het effect van een maatregel ‘verdwijnen’ door wegmiddeling met delen waar geen maatregel plaatsvindt. (pers. comm. Niels Evers).

Wat zijn de aanbevelingen die hieruit volgen?

6. Het opknippen van trajecten in kleinere trajecten is alleen zinvol als de grote trajecten heterogeen zijn in de verdeling van stuurvariabelen. Ook bij toepassing van een maatregelscenario heeft opknippen van trajecten in delen met maatregelen en zonder maatregelen zin en geeft een eerlijker en meer gerichte voorspelling van het effect van maatregelen.

4.3 Verbeteren de modelprestaties als meer waarnemingen worden toegevoegd?

Machine learning is gebaseerd op het trainen van een model met predictor en response waarnemingen zoals de stuurvariabele en KRW-score. Hoe meer stuurvariabelen, hoe meer waarnemingen er nodig zijn om een model goed te trainen. Maar leidt het 'klakkeloos' toevoegen van waarnemingen ook tot een modelverbetering in onze dataset?

- We hebben twee datasets zorgvuldig gecontroleerd op identieke waarnemingen en waar aangetroffen die verwijderd. De resultaten laten zien dat het samenvoegen van deze twee datasets geen verbetering in de modelprestatie levert ten opzichte van de prestaties van de afzonderlijke datasets (paragraaf 3.4).

In eerste aanleg lijkt dit een tegen-intuïtief resultaat. Bij nadere beschouwing blijken de verschillende datasets vooral uit waarnemingen te bestaan die in het middengebied van de KRW-scores vallen en daar dezelfde verdeling hebben in stuurvariabelen. De datasets zijn niet complementair maar hetzelfde, dus geeft samenvoegen geen 'winst'.

Wat betekent dit voor de verkenner?

Bij voorspellen van de biologische kwaliteit in algemeenheid, zijn waarnemingen nodig over het hele spectrum van een 'slechte' tot 'uitmuntende' biologische kwaliteit. Wanneer waarnemingen veel op elkaar lijken, en/of slechts een deel van het spectrum bestrijken, heeft het toevoegen van meer waarnemingen geen zin om tot een betere voorspellende kracht te komen. Daarnaast lijkt het toevoegen van een paar datapunten met hoge EKR-scores weinig effect te hebben op het resultaat. De stuurvariabelen lijken geen goed onderscheid te kunnen maken tussen EKR-scores rond de mediaan en EKR-scores in het hoge spectrum.

Wat zijn de aanbevelingen die hieruit volgen?

7. Goed kijken welke stuurvariabele – EKR-scores missen in de dataset en gericht monitoren om deze te verkrijgen. De uiteinden van het kwaliteitsbereik zijn wel beperkt beschikbaar. Deze kunnen nog toegevoegd worden vanuit expert judgement, eventueel buitenland, en/of de beschrijvingen van de referentie in de maatlatmappen. Dit is ook gebeurd bij de huidige rekenregels van de KRW-Verkenner.

8. Een andere optie is "physics-based learning" in te de modeltraining te gebruiken. Hiermee gebruik je bestaande 'kennisregels' zodat de training zowel op waarnemingen als bestaande kennis is gebaseerd. Doel is een model dat beter presteert over de gehele EKR-score verdeling.

4.4 Enkele verdere methodische conclusies en aanbevelingen

Het blijkt dat de modelprestatie afhangt van de random trekking van de test/training set. Vergelijking van modellen op één random verdeling tussen training en test set zijn niet betrouwbaar.

Aanbeveling

9. Gebruik voor de validatie van het model meerdere random trekkingen (in dit rapport N=500). Toets de significantie van de verschillen in prestaties vervolgens met een geschikte test (bijv. Conover's test). Het uiteindelijk beste model dat gebruikt wordt in de voorspelling, kan worden getraind met alle data.

Voor machine learning zijn verschillende methoden beschikbaar (zie paragraaf 2.2). De vier best scorende modellen zijn gebaseerd op variaties op Gradient Boosting Regression (resp. GBR en XGB (Extreme Gradient Boosting) en Random Forests methode, RF, RNG (resp. Breiman en Ranger).

Aanbeveling

10. Kies een van de gradient boosting methoden (XGB of GBR) voor het model dat uiteindelijk gebruikt gaat worden in de voorspelling.

De KRW-verkenner wordt voor verschillende toepassingen gebruikt, elk met zijn eigen eisen en wensen. Nu voorspelt het model elke KRW-score en wegen we de kwaliteit van de voorspelling over de hele range even zwaar. Maar het model presteert slechter bij hele lage en hele hoge KRW-scores (Figuur 3-5). Voor sommige beleidsaanbevelingen is het voldoende om te weten of een KRW-score uiteindelijk boven het doel uitkomt en dit doel ligt zelden hoger dan 0,6. Hierbij maakt het voor de beleidsaanbeveling minder uit hoe hoog de score uiteindelijk is boven dit doel.

Aanbeveling

11. Een toetsing van een classificatie methode om de voorspel-prestatie te verbeteren. Klassen zouden bijvoorbeeld kunnen zijn: klasse 1: EKR-score < 0,2; gevolgd door verschillende klassen in stappen van 0,05 tot aan 0.6; laatste klasse met EKR-scores > 0.6. Deze modelopzet kan mogelijk een hogere prestatiekwaliteit leveren, omdat het niet wordt afgerekend op fouten binnen een klasse.

5 Referenties

Buijse, T. & van Geest, G. (2021) Evaluatie werkwijze ecologische prognose Nationale Analyse Waterkwaliteit. Notitie Kennisimpuls Waterkwaliteit

Evers, N., Schipper, M., Barten, I., Scheepens, M. (2017). Zoektocht naar stuurknoppen om de ecologische toestand van beken te verbeteren. H2O-online.

Gaalen, F. van, L. Osté & E. van Boekel (2020), Nationale analyse waterkwaliteit. Onderdeel van de Deltaaanpak Waterkwaliteit, Den Haag: Planbureau voor de Leefomgeving.

Hyndman, R.J. and Koehler, A.B. (2006) "Another look at measures of forecast accuracy". International Journal of Forecasting, 22(4), 679-688.

Hyndman, R.J. and Athanasopoulos, G. (2018) "Forecasting: principles and practice", 2nd ed., OTexts, Melbourne, Australia. Section 3.4 "Evaluating forecast accuracy". https://otexts.com/fpp2/accuracy.html.

Loos, S., Renaud, L., Groenendijk,P., Cleij, P., Van der Linden, A., Van der Bolt, F., Kroon, T. (2020). Rapportage Basisprognoses Waterkwaliteit. Toepassing van het Landelijk WaterKwaliteitsModel: status tussenrapportage. Deltares. Delft.

Molen, D.T. van der, R. Pot, C.H.M. Evers, F.C.J. van Herpen & L.L.J. van Nieuwerburgh[red.] (2018) Referenties en maatlatten voor natuurlijke watertypen
voor de kaderrichtlijn water 2021-2027. Stowa rapport 2018-49

Rost, J., M. Schipper, F. van Herpen, J. Lenssen, B. van Zuidam, M. de Vos & P. Kloosterman (2019). Resultaten watersysteemanalyse en voorstellen voor technische aanpassing KRW-doelen Waterschap Rijn en IJssel. Royal HaskoningDHV, 13 mei 2019.

Rost, J., M. Schipper & F. van Herpen (2020). Watersysteemrapportage KRW-Waterlichamen Aa en Maas. Royal HaskoningDHV, 24 januari 2020

Van der Lee G.H., Verdonschot R.C.M. en Verdonschot P.F.M. (2021). Advies voor het monitoren van de ecologische waterkwaliteit. Notitie Kennisimpuls waterkwaliteit (KIWK), Zoetwaterecosystemen, Wageningen Environmental Research, Wageningen UR, Wageningen. pp. 28.

Van der Linden, A. van den Roovaart, J.C., Evers, E., Rost, J., Visser, H., Vethman, P., de Niet, A.C., Nieuwhof, S., Knoben, R., Bontsma, A., van Gaalen, F. (2021). Update ecologische kennisregels KRW-verkenner. Deltares-rapport 11203728-008-BGS-0009

Verdonschot P.F.M. & Verdonschot R.C.M. (2020). Stroomgebiedsbrede Ecologische SysteemAnalyse van het stroomgebied van de Groote Molenbeek. Notitie Zoetwaterecosystemen, Wageningen Environmental Research, Wageningen UR, Wageningen.

Visser, H., Evers, N., Bontsema, A., Rost, J., de Niet, A., Vethman, P., Mylius, S., van der Linden, A., van den Roovaart, J., van Gaalen, F. and Knoben, R. (2022). What drives the ecological quality of surface waters? A review of 11 predictive modeling tools. Water Research, 208, p.117851.

6 Colofon

Utrecht, april 2022

Auteurs: Gertjan Geerling (Deltares), Mijke van Oorschot (Deltares), Jasmijn Rost (RHDHV), Niels Evers (RHDHV), Hans Korving (Deltares)

Leesgroep: Gertie Schmidt (gebruikerscommissie KIWK), Rogier Meijs (gebruikerscommissie KIWK), Hermen Klomp (gebruikerscommissie KIWK), Tom Buijse (Deltares), Joost van den Roovaart (Deltares),

Te citeren als: Geerling, G., M. van Oorschot, J. Rost, N. Evers, H. Korving (2022) Verbeteringsmogelijkheden voor regionale ecologische kennisregels KRW-Verkenner. Pilot: macrofauna in stromende wateren. Notitie Kennisimpuls Waterkwaliteit.