



Bedrijfstakonderzoek
BTO 2022.041 | Juli 2022

Langetermijnvoorspelling van waterstanden met machine learning

Bedrijfstakonderzoek

KWR

Bridging Science to Practice

Rapport

Langetermijnvoorspelling van waterstanden met machine learning

BTO 2022.041 | Juli 2022

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

Opdrachtnummer

BTO 2022.041

Projectmanager

I. (Ina) Vertommen MSc., Dr. G.J. (Geertje) Pronk

Opdrachtgever

BTO - Thematisch onderzoek - Hydroinformatica

Auteurs

E.A. (Esther) Brakkee MSc., Dr. X. (Xin) Tian en Dr. M.J.H. (Marjolein) van Huijgevoort

Kwaliteitsborgers

Dr. P. (Peter) van Thienen, Dr. ir. R.P. (Ruud) Bartholomeus

Verzonden naar

Dit rapport is verspreid onder BTO-participanten.

Een jaar na publicatie is het openbaar.

Keywords

machine learning, deep learning, droogte

Jaar van publicatie
2022

Meer informatie
Esther Brakkee MSc.
T +31 6 52825934
E esther.brakkee@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



Juli 2022 ©

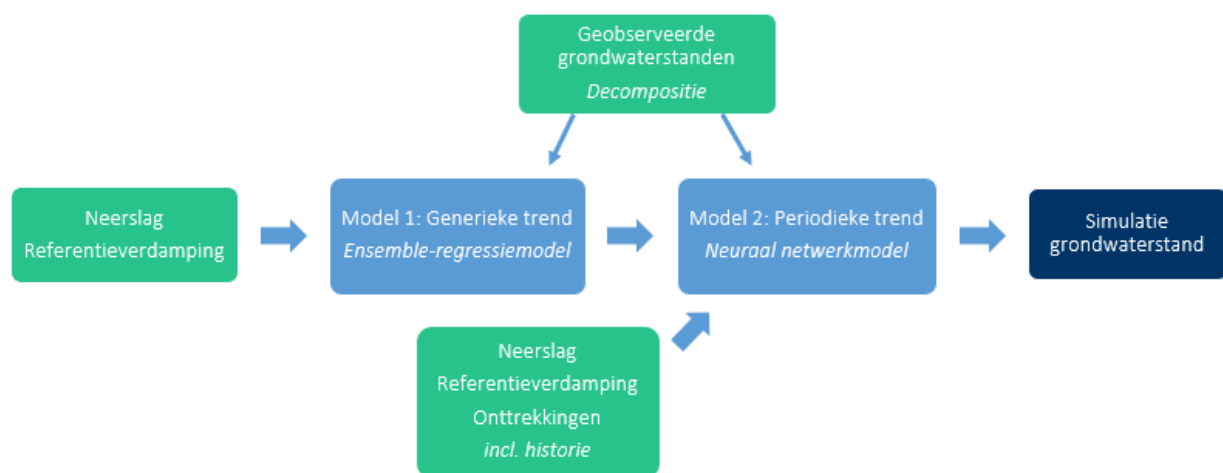
Alle rechten voorbehouden aan KWR. Niets uit deze uitgave mag - zonder voorafgaande schriftelijke toestemming van KWR - worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen,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

Managementsamenvatting

Nieuwe methodes uit machine learning en deep learning voor modelleren van (grond)waterdynamiek en -beschikbaarheid verdienen nader onderzoek

Auteurs Esther Brakkee, Xin Tian, Marjolein van Huijgevoort

Recent ontwikkelde methodes op basis van machine learning en deep learning tonen potentie voor het modelleren van waterdynamiek en -beschikbaarheid. In dit onderzoek is verkend hoe recente machine learning-technieken kunnen worden toegepast voor het maken van voorspellingen van grondwaterstanden en welke voor- en nadelen en aandachtspunten hierbij komen kijken. Er is een vernieuwende machine learning-modelleringsmethode ontwikkeld, die is toegepast voor het maken van langetermijnvoorspellingen voor een set grondwatermeetlocaties in Drenthe en uitgebreid is geëvalueerd. Het onderzoek heeft duidelijk gemaakt dat modelleringsmethodes op basis van machine learning en deep learning waarschijnlijk bruikbaar zijn voor het modelleren van (grond)waterdynamiek en -voorraden onder huidige condities, mits de modellen verder worden verbeterd. Om deze methodes te kunnen gebruiken voor langetermijnvoorspellingen en gedetailleerde analyses van droogteperiodes, is een beter begrip nodig van hun werking en betrouwbaarheid.



Visualisatie van de ontwikkelde modelleringsmethode op basis van machine learning en deep learning.

Belang: nog weinig bekend over potentie van machine learning voor waterbeschikbaarheidsvoorspellingen

Door klimaatverandering worden verschuivingen in weerspatronen en droogteperiodes verwacht, die op complexe manieren kunnen doorwerken op de waterbeschikbaarheid. Voorspellingen hiervan zijn niet makkelijk te maken. Recent ontwikkelde modelleringsmethodes op basis van machine learning en deep learning tonen potentie voor het modelleren van complexe systemen, waaronder watersystemen. De toepassing en het nut van dit soort methodes voor het voorspellen van (grond)waterdynamiek en waterbeschikbaarheid, in

vergelijking met conventionele modelleringsmethodes, is echter nog slecht bekend.

Aanpak: ontwikkeling nieuwe methode en test voor voorbeeldgebied Drenthe

Het onderzoek heeft zich gericht op grondwaterdynamiek, met een doorkijk naar oppervlaktewater. Er is een vernieuwende modelleringsmethode opgezet, die tijdreeksdecompositie combineert met decision tree-regressiemodellering en neurale-netwerkmodellering. De methode is toegepast op een set grondwatermeetlocaties in de provincie

Drenthe, waarbij simulaties zijn gemaakt voor de huidige situatie en verschillende klimaatscenario's. Ook is de toepassing voor oppervlaktewater getest. De modelleringsresultaten zijn samengevat in verschillende waterregime-indicatoren die de waterstandsdynamiek op een locatie kenschetsen. De modelleringsmethode is geëvalueerd op gedrag en betrouwbaarheid voor het reproduceren van de huidige grondwaterdynamiek en het maken van voorspellingen.

Resultaten en discussie: modelleringsmethode deels succesvol, verdere ontwikkeling nodig

Het onderzoek heeft laten zien dat de ontwikkelde combinatie van decompositie en machine learning toe te passen is voor hydrologische tijdreeksen zoals grondwaterstanden. De ontwikkelde modelleringsmethode kon voor de meeste locaties goed de huidige variatie in grondwaterstanden beschrijven. Deze modelresultaten zijn gepresenteerd in het rapport en in een online interactieve viewer. De resultaten lieten zien dat de huidige versie van de ontwikkelde modelopzet nog beperkingen kende, met name het correct representeren van tijds patronen en het omgaan met situaties buiten het trainingsbereik. Deze beperkingen hangen deels samen met de manier waarop het model is opgezet en getraind en kunnen nog worden verbeterd. Er is een overzicht gegeven van mogelijke vervolganalyses om het begrip van het modelgedrag te verbeteren. Ook worden enkele aanpassingen voorgesteld die de prestaties van het model zouden kunnen verbeteren, waaronder het beter meenemen van de vertraagde invloed van omgevingsfactoren en het testen van het gedrag onder extreme condities. Het is echter nog

niet goed bekend in hoeverre datagedreven modelleringsmethodes daadwerkelijk bruikbaar zijn voor het maken van voorspellingen voor een verre toekomst met mogelijk zeer nieuwe condities, ook in vergelijking met meer conventionele methodes. Dit aspect verdient verder onderzoek.

Toepassing: machine learning bruikbaar voor waterbeschikbaarheidsvraagstukken, mits meer inzicht in betrouwbaarheid

Modelleringsmethodes op basis van machine learning en deep learning kunnen, mits ze verder ontwikkeld en verbeterd worden, waarschijnlijk breed toegepast worden voor het modelleren van grondwaterstanden en mogelijk andere watervoorraden onder de huidige condities. Dat maakt het mogelijk gebruik te maken van de groeiende beschikbaarheid van steeds complexere data. Voordat deze modelleringsmethodes echter op een goede manier kunnen worden ingezet voor het maken van langetermijnvoorspellingen en gedetailleerde analyses van droogteperiodes is het nodig eerst hun werking en betrouwbaarheid beter te begrijpen en de modelleringsmethodes verder te verbeteren. Het rapport geeft suggesties voor verdere analyses, modelontwikkelingen en verbeteringen in databeschikbaarheid die zouden bijdragen aan betere hydrologische machine learning-voorspellingen.

Rapport

Dit onderzoek is beschreven in het rapport *Langetermijnvoorspelling van waterstanden met machine learning* (BTO-2022.041).

Inhoud

Rapport	1
<i>Managementsamenvatting</i>	2
Inhoud	4
1 Inleiding	6
2 Methodes	8
2.1 Studiegebied	8
2.2 Data	8
2.3 Indicatoren voor karakterisering van waterbeschikbaarheidsregimes	10
2.4 Ontwikkeling van modellen: inleiding in machine learning en deep learning	11
2.4.1 Machine learning en deep learning	11
2.4.2 Machine learning en deep learning in de watersector	11
2.4.3 Vergelijking van methodes voor langetermijnvoorspellingen van het watersysteem	12
2.5 Toegepaste modelleringsmethode	13
2.6 Evaluatie van gedrag en betrouwbaarheid van de modellen	18
2.6.1 Betrouwbaarheid van de modelsimulaties in de historische periode	18
2.6.2 Rol van externe invloeden	18
2.6.3 Betrouwbaarheid van reproductie van waterregime-indicatoren	18
2.7 Toekomstscenario's	18
2.8 Evaluatie van veranderingen in voorraadregimes	19
2.9 Testtoepassing voor oppervlaktewater	20
3 Resultaten	21
3.1 Grondwaterdynamiek in de huidige situatie	21
3.1.1 Waterregime-indicatoren	21
3.1.2 Variatie in externe factoren	22
3.2 Gedrag en betrouwbaarheid van de modellen	23
3.2.1 Prestatie van de modelsimulaties in de historische periode	23
3.2.2 Rol van externe invloeden	25
3.2.3 Reproductie van waterregime-indicatoren in de historische periode	26
3.2.4 Vertaling van temporele patronen naar de toekomst en potentiële limitaties in bereik	29

3.3	Veranderingen in grondwaterdynamiek in de toekomst	32
3.3.1	Veranderingen in weersomstandigheden	32
3.3.2	Veranderingen in grondwaterregime voor de KNMI-scenario's	33
3.3.3	Veranderingen in grondwaterregime in scenario WH2050	33
3.3.4	Extreem scenario	35
3.4	Testtoepassing oppervlaktewater: de Drentsche Aa	36
3.4.1	Modevaluatie	36
3.4.2	Voorspellingen toekomstscenario's	37
4	Discussie: potentie van machine learning-methodes voor voorspelling van watervoorraden en verdere ontwikkeling	38
4.1	Geschiktheid van de modelleringsmethode voor beschrijven van huidige grondwaterdynamiek	38
4.1.1	Temporele dynamiek en modellering van droogteperiodes	38
4.1.2	Rol van externe invloeden in de modelsimulaties	40
4.1.3	Toepassing voor oppervlaktewater	41
4.2	Geschiktheid van de modelleringsmethode voor voorspellen van toekomstige waterdynamiek	41
4.2.1	Hydrologische voorspellingen dicht bij de historische situatie	41
4.2.2	Hydrologische voorspellingen ver buiten de historische situatie	42
4.2.3	Onzekerheid in de ontwikkeling van sturende factoren	43
4.2.4	Conclusie: gebruik van machine learning voor langetermijnvoorspelling van waterstanden	44
4.3	Aanbevelingen voor verdere ontwikkeling	44
4.3.1	Analyse van gedrag en betrouwbaarheid van machine learning-modellen	44
4.3.2	Mogelijke verbeteringen van de modelleringsmethode	45
4.3.3	Verbetering van databeschikbaarheid voor hydrologische (machine learning-)modellen	45
5	Conclusies	47
6	Referenties	48
7	Bijlagen	51
I	Additional model performance analyses	51
	Results from Menyanthes	53
	Results from the ensemble learning regression model	54
	Concluding remarks	54
II	Tabellen resultaten	55

1 Inleiding

De recente droogteperiodes van 2018-2020 hebben opnieuw duidelijk gemaakt dat de waterbeschikbaarheid in veel Nederlandse landschappen kwetsbaar is voor veranderingen en extremen in het weer. Door klimaatverandering is het waarschijnlijk dat de patronen van waterbeschikbaarheid in grond- en oppervlaktewater in de tijd en ruimte zullen veranderen. Met name wordt verwacht dat droogteperiodes vaker op zullen treden en intenser zullen worden (Philip et al., 2020). Klimaatverandering zal op verschillende locaties op verschillende manieren doorwerken op onze zoetwatervoorraden, in interactie met andere menselijke invloeden. De effecten op de waterbeschikbaarheid op lange termijn zijn dan ook niet eenvoudig te voorspellen.

Verschiedene methodes kunnen worden ingezet om inzicht te krijgen in de toekomstige veranderingen in waterdynamiek en -voorraden onder invloed van klimaatverandering en waterbeheer. Vaak worden hiervoor fysisch gebaseerde numerieke modellen ingezet (bijv. Mens et al., 2019; Hutchins et al., 2018; Atawneh et al., 2021; Prinsen et al., 2015). Nadeel hiervan is dat deze modellen veel tijd, data en gebiedskennis vergen om op te zetten. Daarom worden ook datagedreven methodes al lange tijd gebruikt voor het modelleren en voorspellen van grond- en oppervlaktewaterdynamiek. Een bekend voorbeeld is impuls-respons-tijdreeksmodellering van grondwaterstanden (Bakker en Schaars, 2019; von Asmuth et al., 2002). Als vervolg hierop zijn in recente jaren machine learning-methodes¹ snel verder ontwikkeld, waaronder 'deep learning'-methodes (Alsumaiei, 2020; Guzman et al., 2019; Li et al., 2021; Wunsch et al., 2021). Kenmerk hiervan is dat de algoritmes zelf relaties en patronen uit de data afleiden. Dit maakt deze methodes potentieel geschikt om flexibel modellen op te zetten, die om kunnen gaan met complexe situaties. De toepassing van deze methodes op watersystemen is pas recent begonnen en moet nog beter worden verkend (Ardabili et al., 2020; Tian et al., 2022).

Het doel van dit onderzoek is om te verkennen hoe recent ontwikkelde methodes vanuit machine learning en deep learning zouden kunnen worden toegepast voor het voorspellen van watervoorraad-dynamiek, met een focus op grondwaterstanden. We ontwikkelden een modelleringsmethode en passen deze als test toe om langetermijnvoorspellingen te maken voor een set grondwatermeetlocaties. Dit geeft een eerste inzicht in de mogelijke potentie van machine learning-methodes voor langetermijnvoorspellingen op het gebied van hydrologie en waterbeschikbaarheid; en in de aandachtspunten die bij het ontwikkelen en toepassen van dit soort modellen komen kijken. Het onderzoek richt zich op de volgende uitkomsten:

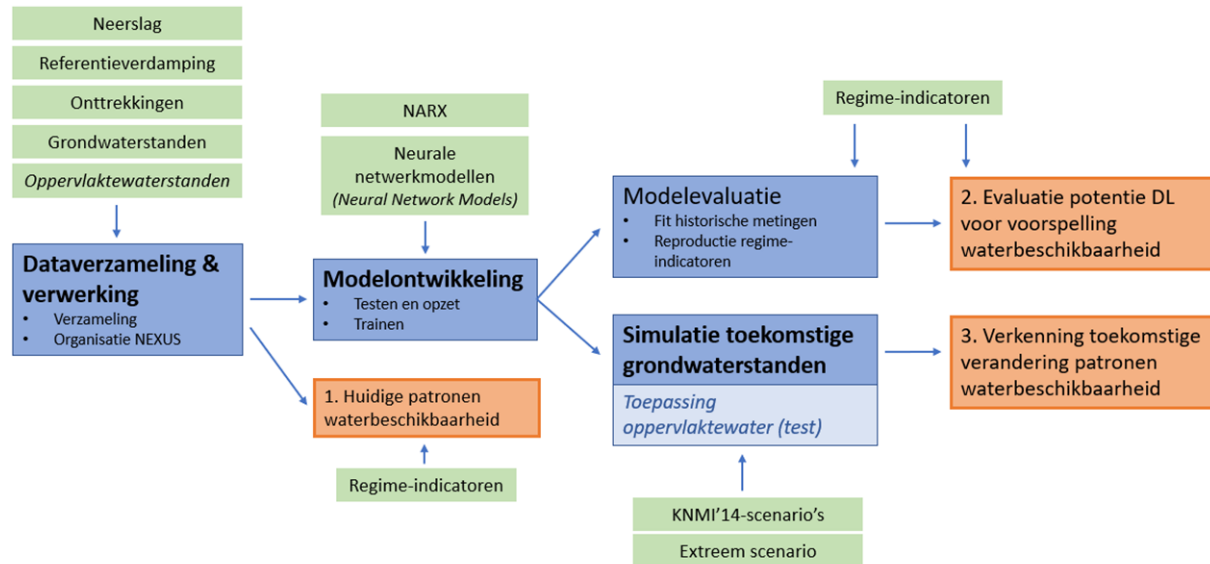
- 1 verkenning van de huidige patronen in (grond)watervoorraden in ruimte en tijd;
- 2 ontwikkeling van een modelleringsmethode voor langetermijnsimulatie van (grond-)waterstanden op basis van machine learning en deep learning;
- 3 testtoepassing van de ontwikkelde modelleringsmethode voor verkenning van de potentiële toekomstige veranderingen in (grond)waterstanden onder klimaatverandering;
- 4 evaluatie van de ontwikkelde modellen en van de potentie van recente machine learning en deep learning-methodes voor langetermijnvoorspellingen van de waterbeschikbaarheid.

Een overzicht van de activiteiten is gegeven in Figuur 1.1. De provincie Drenthe is gebruikt als studiegebied. Er is data verzameld van weersomstandigheden, grondwater- en oppervlaktewaterstanden en grondwateronttrekkingen. Op basis van deze data zijn modellen ontwikkeld voor een reeks van grondwatermeetpunten, als ook voor een testtoepassing voor oppervlaktewater. Het gedrag en de betrouwbaarheid van de modellen is geëvalueerd. Vervolgens zijn de modellen toegepast om potentiële

¹ De officiële spelling in het Nederlands is "machinelearning". We hebben ervoor gekozen om in het rapport de spelling "machine learning" aan te houden, omdat we het vaak gebruiken samen met "deep learning" en het consistent gebruiken van twee losse woorden de minste verwarring oplevert.

veranderingen in toekomstige grondwaterdynamiek te voorspellen voor verschillende klimaatscenario's. De resultaten zijn gepresenteerd in dit rapport en gevisualiseerd in een online platform:

<https://nexus.stellaspark.com/#/viewer/110>.



Figuur 1.1: Overzicht van het onderzoek.

Hoofdstuk 2 beschrijft de methodes, en omvat ook een verdere toelichting op machine learning en deep learning in vergelijking met andere modelleringsmethodes. De resultaten van de modevaluatie en de voorspellingen worden gegeven in hoofdstuk 3. Hoofdstuk 4 bevat een discussie over de bruikbaarheid van de ontwikkelde machine learning-methode voor het analyseren en voorspellen van (grond)watervoorraden. Hoofdstuk 5 sluit af met een conclusie.

2 Methodes

2.1 Studiegebied

De provincie Drenthe is gebruikt als studiegebied voor het onderzoek. De provincie Drenthe kent een samenhangend watersysteem met een variatie in hogere en lager gelegen delen, en is representatief voor grote delen van zuid- en oost-Nederland. Drenthe is opgebouwd uit een plateau dat wordt gedraineerd door verschillende beken en doorsneden door kunstmatige waterlopen. De bovengrond bestaat hoofdzakelijk uit zand, met ondiep vaak leemlagen (keileem); lokaal komen venige gronden voor. Grondwater is een belangrijke waterbron in de provincie, en is van belang voor drinkwater, landbouw en natuurgebieden. De grondwaterdynamiek varieert binnen de provincie. Hoger op het plateau zijn de grondwaterstanden dieper en meer variabel dan in de beekdalen en lager gelegen gebieden. In delen van de provincie wordt het grondwatersysteem beïnvloed door grondwateronttrekking, aanvoer en beheer van oppervlaktewater, en andere menselijke factoren.

Droogteperiodes hebben in de afgelopen jaren een sterke invloed gehad op de grondwatervoorraden in Drenthe. In de hogere gebieden en de beekdalen hebben de droge periodes geleid tot scherpe dalingen in grondwaterstanden, waardoor beken droogvielen en de druk op grondwaterafhankelijke ecosystemen vergrootte. Ook de landbouw had te lijden van beperkingen op beregening uit grondwater en droogteschade. In de lage delen van de provincie kan in droge periodes oppervlaktewater worden aangevoerd om oppervlaktewater- en grondwaterstanden op peil te houden, maar de aanvoercapaciteit loopt op dit moment in droge zomers al tegen haar grenzen aan (Aa's, 2020; Tromp et al., 2018; De Lenne en Worm, 2020).

De invloed van klimaatverandering op neerslag- en verdampingspatronen zijn de afgelopen decennia al zichtbaar geworden in het zuiden en oosten van het land (Philip et al., 2020). Volgens de huidige verwachtingen zal de neerslag in de winter toenemen en 's zomers afnemen; zal intense neerslag vaker voorkomen; en zal er een toename zijn in verdamping en droogtestress in de zomer (Philip et al., 2020; KNMI, 2014). Deze veranderingen werken vervolgens door op het watersysteem, waarbij de doorwerking sterk verschilt tussen locaties (van den Eertwegh et al., 2021).

2.2 Data

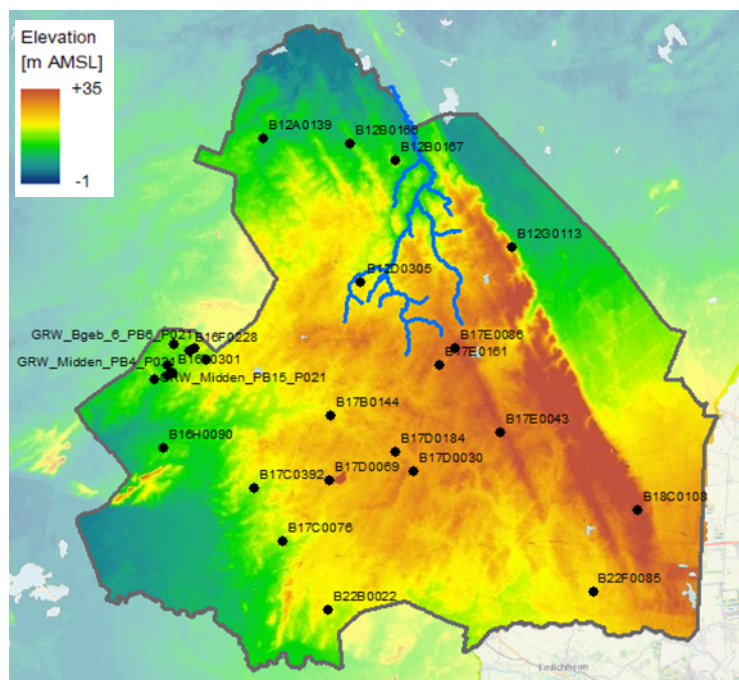
Als basis voor de opbouw van de modellen en het maken van voorspellingen zijn gegevens verzameld van de historische grond- en oppervlaktewaterdynamiek en een selectie aan sturende factoren. Data zijn verzameld van het KNMI, DINOloket, direct van waterschappen, en drinkwaterbedrijf WMD. Data van KNMI en DINOloket zijn online verzameld. Voor de aanvullende data van grond- en oppervlaktewaterstanden en waterbeheer hebben we direct contact gehad met de vier Drentse waterschappen voor de selectie en interpretatie van de belangrijkste data. Van WMD zijn gegevens ontvangen van grondwateronttrekkingen voor drinkwaterproductie.

De volgende data zijn gebruikt in de modellering:

- Historische reeksen van neerslag en referentieverdamping van KNMI-neerslag- en weerstations;
- Meetreeksen van grondwaterstijghoogten. Meetreeksen zijn verkregen via de waterschappen en aangevuld met een selectie aan DINOloket-data. Hieruit is een selectie gemaakt van reeksen met tenminste 15 jaar aan meetdata die samen een goede spreiding hadden over de provincie, zodat verschillende landschapstypen en grondwaterdieptes mee werden genomen.

- Meetreeksen van oppervlaktewaterstanden. In eerste instantie was de inzet om ook voorspellingsmodellen te ontwikkelen voor oppervlaktewaterstanden. Dit bleek echter nog niet haalbaar doordat de meeste data van oppervlaktewaterstanden bestond uit korte reeksen uit sterk beheerde watersystemen, die om een heel andere modelopzet zouden vragen dan de grondwaterreeksen. Daarom is uiteindelijk gekozen om de toepassing op oppervlaktewatersystemen te testen voor één locatie, de Drentsche Aa. De Drentsche Aa kent een relatief natuurlijke afvoer en er is een lange meetreeks van waterstanden beschikbaar. De Drentsche Aa is daarmee in principe geschikt om te modelleren met de ontwikkelde modelleringsmethode.
- Locaties van drinkwateronttrekkingen en maandelijkse onttrekkingshoeveelheden van WMD. Naast drinkwateronttrekkingen zijn er ook belangrijke onttrekkingen voor industrie en beregening. Data van grote industriële onttrekkingen worden door provincies bijgehouden, maar konden voor deze studie alleen op jaarbasis en zonder locatiegegevens verkregen worden, waarmee ze niet goed bruikbaar waren voor de modeltraining. Data van landbouwonttrekkingen zijn er in het algemeen weinig. Deze onttrekkingen zijn daarom niet meegenomen in de modellering.
- Reeksen van neerslag en referentieverdamping voor de KNMI'14-klimaatscenario's (zie par. 2.7).

Figuur 2.1 geeft de 27 grondwatermeetlocaties die zijn gebruikt in het onderzoek. In blauw is de Drentsche Aa aangegeven. De beschikbare meetreeksen starten tussen 1990 en 2005 en eindigen in 2020. De meeste reeksen hebben een dagelijkse frequentie; delen van de reeksen met een lagere frequentie zijn lineair geïnterpoleerd naar een dagelijkse frequentie. Het bleek dat voor twee locaties geen stabiel grondwaterregime aanwezig was in de historische meetreeks (B12G0113 en PB15_P021) doordat er een verschuiving of een trend aanwezig was in de metingen, waarvoor geen verklaring lag in de onttrekkingen of meteorologische invloeden. Dit betekent dat het grondwaterregime voor deze locaties niet goed gekarakteriseerd kan worden. Deze locaties zijn daarom weggelaten uit de resultaten, waarmee 25 locaties overblijven.



Figuur 2.1: Studiegebied en geselecteerde grondwatermeetpunten voor modellering.

De data zijn verzameld en georganiseerd in de NEXUS-omgeving ([Nexus - StellaSpark](#)), een platform voor het verwerken en visualiseren van ruimtelijke data. Een deel van de data van de waterschappen is toegevoegd als directe api-connectie met de datasystemen van deze waterschappen. Dit betekent dat de database en de modellen kunnen worden ge-update, en in potentie 'real-time' zouden kunnen worden gebruikt voor delen van het

studiegebied. Hoewel dit in deze studie niet is toegepast, is NEXUS wel gebruikt om een interactieve viewer te maken van de uiteindelijke voorspellingsresultaten als één van de producten van het onderzoek.

2.3 Indicatoren voor karakterisering van waterbeschikbaarheidsregimes

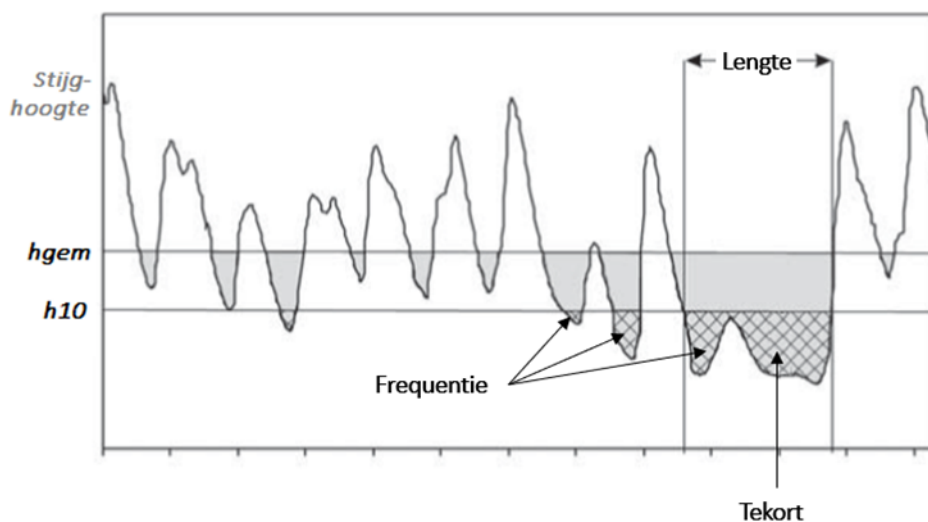
Om de grondwaterdynamiek op de studielocaties te beschrijven, en deze te kunnen vergelijken in de huidige en toekomstige situatie, zijn regime-indicatoren gebruikt. Voor het beschrijven van de grondwaterdynamiek is het wenselijk om voor alle locaties (ongeveer) dezelfde referentieperiode aan te houden. De periode 4-2003 tot 4-2020 was de maximale periode die door bijna alle reeksen werd omvat, en is daarom consistent als referentieperiode gebruikt. Dit geeft een referentieperiode van 15 tot 17 jaar voor de verschillende locaties (15 of 16 jaar voor enkele locaties die in 2004 of 2005 startten). De gebruikte regime-indicatoren voor het beschrijven van het bereik van stijghoogten zijn:

- Gemiddelde stijghoogte over de volledige geselecteerde periode **hgem**.
- Gemiddelde stijghoogte in de winter (december-februari), lente (maart-mei), zomer (juni-augustus) en herfst (september-november) **hgem_{winter}**, **hgem_{voorjaar}**, **hgem_{zomer}** en **hgem_{najaar}**.
- Percentielen van de stijghoogtemeetreeks **h5**, **h10** en **h90**. Deze 5%, 10% en 90%-percentielen zijn gekozen om de (extreem) lage en hoge stijghoogten te karakteriseren. Deze geven de grondwaterstijghoogten aan die resp. 5%, 10% en 90% van de tijd onderschreden worden (5, 10 en 90 procent laagste waarden).
- Gemiddelde laagste en hoogste grondwaterstand **GLG** en **GHG**. Deze worden berekend als het gemiddelde van de drie laagste (hoogste) stijghoogten per hydrologisch jaar over een langere aaneengesloten periode, op basis van metingen die zijn gesub-sampled naar tweewekelijkse frequentie.

Bovenstaande indicatoren zijn uitgedrukt in meters ten opzichte van het maaiveld, om verschillende locaties te kunnen vergelijken.

Naast de verdeling van stijghoogten zijn de patronen in de tijd van belang om de waterbeschikbaarheid te karakteriseren. De gevolgen van periodes met lage waterstanden hangen immers af van hoe lang ze aanhouden en hoe groot het watertekort is dat wordt opgebouwd. Daarom is een set indicatoren gebruikt om de kenmerken van periodes met lage grondwaterstanden (en dus eventuele droogtestress) te karakteriseren. Deze zijn grof gebaseerd op eenvoudige 'vaste-grenswaarde' droogte-indicatoren zoals eerder toegepast in onderzoek naar grondwaterdroogte (o.a. Peters et al., 2006; Van Loon, 2015). Omdat een vaste grenswaarde is gebruikt zonder seizoensvariatie mee te nemen, en de referentieperiode vrij kort is, kan niet goed worden gesproken van 'droogte-indicatoren' maar hebben we gekozen om te spreken van 'laag-niveau-periodes'. Als grens om deze periodes af te bakenen is het 10% percentiel van de historische meetreeksen gebruikt (2003-2020), waarmee dus 10% van de meetreeks als laag-niveau wordt gekarakteriseerd. Als indicatoren zijn gebruikt (zie ook Figuur 2.2):

- **Lengte**: de gemiddelde lengte (dagen) van periodes waarin de grondwaterstijghoogte onder de grenswaarde valt.
- **Frequentie**: kans van voorkomen van een laag-niveau-periode in een jaar (eenheid n/jaar)
- **Tekort**: totale geïntegreerde onderschrijding van de grenswaarde over alle laag-niveau-periodes per jaar, als maat voor de ernst van de uitzakkingen van de grondwaterstand.



Figuur 2.2: Kenmerken van periodes met lage stijghoogteniveaus. Aangepast op basis van Peters et al. (2006).

2.4 Ontwikkeling van modellen: inleiding in machine learning en deep learning

2.4.1 Machine learning en deep learning

Machine learning is een breed veld van kunstmatige intelligentie dat zich bezighoudt met de ontwikkeling van algoritmen en technieken waarmee computers kunnen leren. Machine learning-technieken zijn bedoeld om patronen en relaties uit (vaak grote hoeveelheden) data af te leiden en hiermee voorspellingen te maken (Lange en Sippel, 2020). Machine learning kent verschillende algoritme-families die zijn ontwikkeld voor verschillende doeleinden. *Deep learning* is een subtak van machine learning die de werking van het menselijk brein imiteert door het simuleren van lagen van neuronen bij het verwerken van gegevens en het creëren van patronen voor gebruik bij besluitvorming. Het adjectief "diep" komt voort uit het gebruik van meerdere lagen in het netwerk. Deep learning is een moderne variant die zich bezighoudt met een onbegrensd aantal lagen van begrensde grootte, wat praktische toepassing en geoptimaliseerde implementatie mogelijk maakt. Op dit moment worden binnen deep learning veel nieuwe algoritmes ontwikkeld voor verschillende toepassingen.

2.4.2 Machine learning en deep learning in de watersector

In het verkennend onderzoek 'Deep Explorations' (Tian et al., 2022) is een overzicht gemaakt van de huidige toepassingen van machine learning en deep learning in de watersector en de omgevingswetenschappen. Bestaande toepassingen omvatten onder andere early-warning-systemen, classificatie van landgebruik uit luchtfoto's, detecteren van lekken in waterpijpen, en voorspelling van drinkwatergebruik. Dit overzicht liet zien dat de toepassingen op gebied van hydrologie nog beperkt zijn.

Toepassingen van machine learning en deep learning voor het voorspellen van hydrologische variabelen, zoals rivierafvoeren en grondwaterstanden, zijn de laatste jaren aan het opkomen in de wetenschappelijke literatuur (bijv. Alsumaiei, 2020; Guzman et al., 2019; Li et al., 2021; Wunsch et al., 2021). Een belangrijke uitdaging hierbij is het omgaan met tijdreeksen, die voor veel machine learning-technieken lastiger te modelleren zijn dan ongeordende data (Lange en Sippel, 2020). Vaak kunnen met machine learning en deep learning echter preciezere simulaties worden gemaakt dan met meer conventionele tijdreeksmethodes zoals ARMA-type-methodes. Toepassing van deze methodes in de praktijk in de watersector staat nog in de kinderschoenen. Het is dan ook nog niet goed bekend hoe geschikt deep learning-methodes zijn voor het maken van hydrologische voorspellingen in de praktijk, met welke factoren bij toepassing rekening gehouden moet worden, en hoe dit vergelijkt met meer conventionele modelleringsmethodes.

2.4.3 Vergelijking van methodes voor langetermijnvoorspellingen van het watersysteem

Bij het kiezen van een modelleringsmethode voor het maken van hydrologische voorspellingen maakt het een groot verschil of het gaat om kortetermijnvoorspellingen (uren-weken) of langetermijnvoorspellingen (jaren-decennia). Voorspellingen voor de korte termijn worden gemaakt voor operationeel waterbeheer. In dit geval is de startsituatie goed bekend en de geaccumuleerde fout door stap voor stap vooruit te rekenen is klein. Er kan worden aangenomen dat de kenmerken van het systeem niet veranderen. Wel moeten kortetermijnfluctuaties in de hydrologische tijdreeks (bijv. afvoer) precies worden voorspeld.

Voorspellingen voor de lange termijn worden meestal gemaakt voor strategische doeleinden. In dit geval is de precieze timing van gebeurtenissen minder belangrijk (of het droog wordt in 2037, 2047 of 2057 is niet te voorspellen) maar gaat het vooral om de algemene ontwikkeling (bijv. de gemiddelde grondwaterstand) en de kans op extreme situaties. Bij voorspellingen ver de toekomst in neemt de onzekerheid echter toe, doordat fouten en onzekerheden over de tijd kunnen accumuleren omdat geen recente waarden van waterstanden beschikbaar zijn. Ook is in de verdere toekomst onzeker in hoeverre de processen en relaties die in de modellen zijn aangenomen (of aangeleerd) nog geldig zijn (Thirel et al., 2015).

Tabel 2.1 geeft een vergelijking van verschillende methodes voor het maken van langetermijnvoorspellingen van waterbeschikbaarheid, specifiek de grondwaterdynamiek. In vergelijking met andere meer conventionele methodes hebben op machine learning en deep learning gebaseerde methodes potentie voor het modelleren van complexe niet-lineaire systemen die (nog) moeilijk direct in vergelijkingen te vatten zijn. Bovendien kunnen ze goed van verschillende (nieuwe) databronnen gebruik maken en de onzekerheid van voorspellingen eenduidig in beeld brengen. Daarbij kunnen ze zo opgezet worden dat de accumulatie van fouten richting de toekomst wordt beperkt. Nadelen zijn dat de modellen voor het leren van relaties in het watersysteem zeer afhankelijk zijn van de data waarmee ze gevoed worden, geen mechanistische relaties bevatten en dus in principe moeite hebben met situaties die sterk afwijken van het trainingsbereik (dit aspect wordt uitgebreid besproken in de discussie, zie par. 4.2). Bovendien is nog steeds kennis nodig van het hydrologische systeem om (handmatig) een geschikte modelopzet te bouwen en te finetunen. Deze voor- en nadelen worden in dit onderzoek verder verkend.

Tabel 2.1: Algemene kenmerken van verschillende modelleringsmethodes voor voorspellingen van grondwaterdynamiek.

	Fysisch gebaseerde ruimtelijke modellen (bijv. ModFlow)	Impuls-responsmodellering (bijv. Hydromonitor/PASTAS (von Asmuth et al., 2002; Bakker en Schaars, 2019)	Deep learning, o.a. NARX (neural network modelling with exogenous inputs)
Datavraag	Groot	Relatief klein	Flexibel; wel afhankelijk van goede kwaliteit en representativiteit data
Benodigde kennis van systeem vooraf	Groot	Gemiddeld: van tevoren keuze voor responsfunctie en type responsmodel	Gemiddeld: keuzes nodig bij modelopzet
Flexibiliteit	Hydrologische relaties 'hard' ingebouwd; betrouwbaarheid model afhankelijk van (subjectieve) modelaannames	Meestal klein aantal generieke responsfuncties gebruikt; wel eenvoudig te updaten met nieuwe data	Flexibel voor veel verschillende typen gedrag en eenvoudig te updaten met nieuwe data; handmatig tunen van modelopzet blijft nodig
Omgang met complex / niet-lineair gedrag	Kan omgaan met complexe niet-lineaire systemen, mits processen juist ingebouwd	Meestal lineaire respons gebruikt, er zijn echter veel opties voor externe invloeden en responsfuncties.	In principe geschikt voor uiteenlopende niet-lineaire relaties

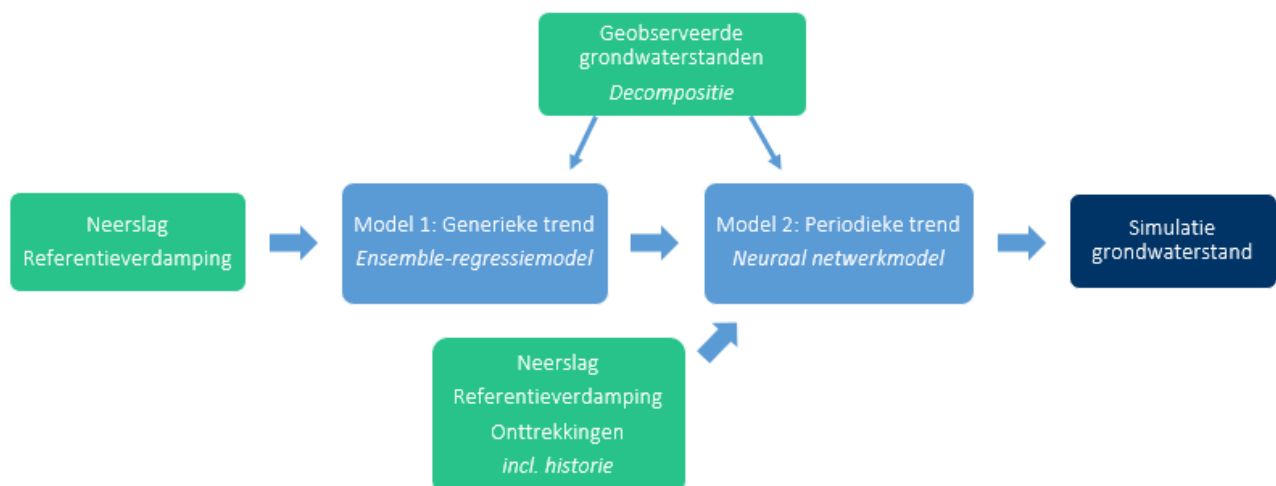
Toepassing binnen huidige condities	Onzekerheid uit modelinvoer en modelaannames	Onzekerheid uit modelinvoer en ruis in model	Onzekerheid uit de mate van determinisme in de data; te kwantificeren
Toepassing buiten huidige condities (lange termijn) (zie verder par. 4.2)	Kan door fysische basis tot op zekere hoogte situaties buiten het historische bereik modelleren; feedbacks en verandering systeemgedrag vaak niet meegenomen	Kan tot zekere hoogte situaties buiten historisch bereik modelleren; bij ver extrapoleren grote onzekerheid	Kan afhankelijk van opzet tot zekere hoogte situaties buiten historisch bereik modelleren; voor lange termijn echter grote onzekerheid door sterke afhankelijkheid historische data
Transparantie/eenvoudigheid van uitleggen en controleren uitkomsten	Gemiddeld/laag: white-box maar meestal complexe modellen	Groot	Laag: black-box-karakter

2.5 Toegepaste modelleringsmethode

Een generieke modelleringsmethode is opgezet, die is toegepast voor elk van de 25 locaties met grondwatermeetreeksen (Figuur 2.1). De invoer gebruikt voor het trainen van de verschillende modellen bestaat uit:

- gemeten neerslag van het dichtstbijzijnde meetstation (dagelijks);
- (door het KNMI berekende) referentieverdamping (hierna kortweg 'verdamping') van het dichtstbijzijnde meetstation (dagelijks);
- totale maandelijkse grondwateronttrekkingen van alle drinkwaterwinputten in een straal van 10 kilometer (data waren alleen op maandbasis beschikbaar). De afstand tot de winputten en hun diepte zijn niet expliciet aan het model meegegeven. Voor enkele locaties lagen geen drinkwaterwinningen binnen een afstand van 10 km; voor deze locaties is alleen weerdata als invoer gebruikt.

Figuur 2.3 geeft een versimpeld overzicht van de modelleringsmethode. De methode bestaat uit een decompositie, waarbij de grondwatermeetreeks wordt ontleed in een trager en een sneller signaal. Deze signalen worden vervolgens apart gemodelleerd met een regressiemodel en een neurale netwerkmodel. Gecombineerd geven deze de uiteindelijke simulatie van de grondwaterstanden.



Figuur 2.3: Vereenvoudigd overzicht van de modelleringsmethode.

Boxen 1-3 geven een volledige uitleg van de modelleringsmethode. De eerste stap is de decompositie van de originele grondwaterstandsmeetreeks met *Singular Spectrum Analysis* (Bogalo et al., 2021). Deze ontleedt de reeks in een traag signaal, de Generieke Trend (maanden – jaren), en een sneller signaal, de Periodieke Trend (dagen – weken). Figuur B1.1 geeft een voorbeeld van decompositie van een grondwatermeetreeks.

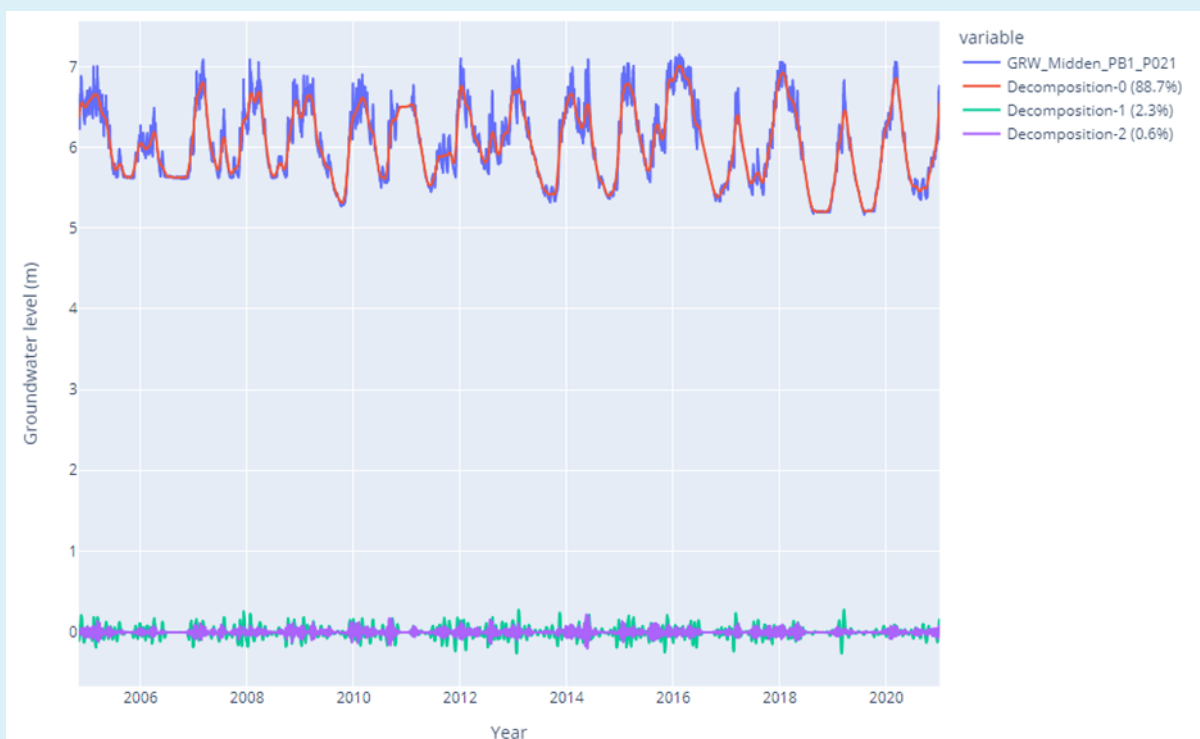
Het eerste model (voor de generieke trend) gebruikt een regressiemethode, de ‘Ensemble Decision Trees’-methode, om dagelijkse neerslag en verdamping te relateren aan het niveau van de generieke trend van de grondwaterstand. Deze methode is gekozen na tests van een reeks regressiemethodes op de data. Deze methode gebruikt alleen neerslag en verdamping van één dag terug voor de modellering. Naast neerslag en verdamping zijn ook drinkwateronttrekkingen getest als invoer voor model 1; deze verbeterden echter de simulatie van de generieke trend niet. Model 1 is daarom gebaseerd op alleen neerslag en verdamping.

Het tweede model (voor de periodieke trend) maakt gebruik van de NARX-methode, ‘Nonlinear Autoregressive Network with Exogenous Inputs’, voor het simuleren van de totale grondwaterdynamiek (generieke en periodieke trend). Dit model modelleert de grondwaterstand op dag x met de (gemeten of voorspelde) neerslag en verdamping en drinkwateronttrekkingen van de afgelopen 7 dagen; en de gemodelleerde generieke trend vanuit model 1, ook inclusief de afgelopen 7 dagen. Op deze manier wordt de vertraging en autocorrelatie in de grondwaterdynamiek meegenomen.

Box 1: Decompositie van tijdreeksen

Een tijdreeks is een signaal dat meestal uit meerdere sub-signalen met verschillende frequenties bestaat. Grondwaterstanden worden bijvoorbeeld vaak bepaald door klimaatvariaties op lange termijn (bijv. een stijgende trend van de gemiddelde temperatuur en verdamping) en door variaties op korte termijn (bijv. seizoensvariatie in verdamping of een korte regenbui). De gemeten grondwaterstanden weerspiegelen de geïntegreerde signalen van al deze factoren als één enkele tijdreeks. Het analyseren en voorspellen van dergelijke tijdreeksen is vaak een uitdaging, met grote voorspellingsfouten. Een recente studie heeft aangetoond dat het nauwkeuriger en eenvoudiger kan zijn om een keten van modellen te bouwen, door de grondwaterstanden te ontleden in sub-signalen met verschillende frequenties, elk sub-signaal afzonderlijk te voorspellen, en tenslotte de sub-signalen te integreren tot de totale voorspelling (Li et al., 2021).

In deze studie is *Singular Spectrum Analysis* (SSA) toegepast om de grondwatertijdreeks te ontleden in drie signalen: een generieke trend, een periodieke trend en ruis. SSA legt de nadruk op de scheidbaarheid van de onderliggende subsignalen en kan gemakkelijk periodiciteiten scheiden die op verschillende tijdschalen voorkomen, ongeacht of de tijdreeks ruis bevat of niet. De oorspronkelijke tijdreeks kan worden gereconstrueerd door alle deelsignalen (gewoonlijk meer dan 10) bij elkaar op te tellen. Deze sommatie moet echter handmatig worden uitgevoerd, door de eigenwaarden en periodiciteiten van de ontlede signalen te vergelijken. Een recente studie, gepubliceerd in 2021, heeft SSA sterk verbeterd door de procedure voor signaalextractie en -reconstructie te automatiseren (Bógalo et al., 2021). De nieuwe methode, Circulant SSA (CiSSA), wordt in deze studie toegepast om grondwaterstanden automatisch te ontleden naar verschillende trends (Bógalo, 2022). Bij de decompositie van de tijdreeksen is het belangrijk om de 'explainability ratio' van de subsignalen ten opzichte van de oorspronkelijke tijdreeks te kwantificeren, of de verhouding tussen de variantie van een signaal en de totale variantie. Figuur B1.1 laat bijvoorbeeld zien dat de signalen #0, #1, en #2 respectievelijk 88,7%, 2,3%, en 0,6% van de originele tijdreeks kunnen verklaren. Op basis van de frequentie van de deelsignalen staat deelsignaal #0 voor de generieke trend, deelsignalen #1 en #2 staan voor de periodieke trend (we sommeren ze dus als één trend). De resterende 8,4% (=1-88,7%-2,3%-0,6%) wordt gewoonlijk beschouwd als ruis, die moeilijk te voorspellen is. Merk op dat SSA een andere methode is dan (snelle) Fouriertransformatie (FT).



Figuur B1.1: Voorbeeld van decompositie van een grondwatermeetreeks.

Box 2: Modelleren van deelsignalen

Na de decompositie bouwen we twee afzonderlijke modellen voor het voorspellen van de generieke trend en de periodieke trend. Het eerste model voorspelt de generieke trend op basis van twee sturende factoren, neerslag en verdamping. Dit gebeurt met een regressiemodel dat de generieke trendwaarde van de grondwaterstand rechtstreeks relateert aan de weerscondities van de vorige dag. Het tweede model voorspelt de periodieke trend op basis van vier sturende factoren, namelijk neerslag, verdamping, onttrekking, en de voorspelde generieke trend van het eerste model. Dit model is een neuraal netwerkmodel, dat ook voorgaande waarden van de invoervariabelen meeneemt. Op deze manier wordt model 1 gebruikt om de generieke trend te voorspellen, zodat we een idee krijgen van de verandering van de grondwaterstand over een heel jaar; en wordt model 2 gebruikt om de periodieke variaties te voorspellen op basis van de voorspelde generieke trend, zodat we de verandering binnen een jaar kunnen onderzoeken.

Bij toepassing van neurale netwerken die de informatie uit de historie gebruiken om recursief voorspellingen te maken, moet de historische informatie (d.w.z. voorgaande grondwaterstanden) bij elke voorspellingsstap beschikbaar zijn. In de praktijk beschikken we echter niet over informatie van bijvoorbeeld 2049 om de voorspelling voor 2050 te maken (merk op dat dit nadeel ook geldt voor de meeste klassieke modellen). Om dit probleem op te lossen, hebben we een alternatieve methode onderzocht, waarbij eerst een regressiemodel wordt getraind voor de generieke trend, enkel gebaseerd op neerslag en verdamping; en deze trend wordt gebruikt als "historische informatie" om de neurale netwerken voor de volledige voorspellingen te trainen. Hierdoor wordt de afhankelijkheid van voorafgaande waterstandsinput voor de voorspellingen weggenomen.

Voor het eerste model (generieke trend) is een grove screening uitgevoerd waarin een aantal potentiële regressiemodellen is getest. Machine learning (inclusief deep learning) omvat een aantal families van algoritmen met hun eigen subalgoritmen. Volgens de 'No free lunch'-theorie kunnen we niet precies weten welk(e) algoritme(n) het beste zou(de) presteren zonder deze te testen op de dataset (Brownlee, 2021). De geteste regressiemodellen voor model 1 omvatten lineaire en kwadratische *support vector machines*, *k-nearest neighbours* en *ensemble decision trees* (EDT). Een grove screening houdt in dat we de standaardwaarden voor model(hyper)parameters gebruiken en de prestaties van verschillende modellen vergelijken. In de praktijk moeten we vaak de (hyper)parameters voor elk algoritme/model optimaliseren om de optimale prestatie te verkrijgen en vervolgens de optimale prestatie vergelijken tussen alle geselecteerde modellen. Dit is een typisch probleem dat zich voordoet in de meeste operationele machine learning-toepassingen. In deze studie hebben we alleen grove screening toegepast en EDT als de beste methode voor model 1 gekozen. Naast de selectie van modellen is de grondwateronttrekking getest als extra invoer; dit leidde echter niet tot een verbetering van het model. Dit suggereert dat drinkwaterwinning op de generieke trend een geringere invloed heeft dan meteorologische factoren. De onttrekkingen zijn daarom niet in het eerste model opgenomen, maar wel in het tweede model, dat de periodieke trend voorspelt.

Het tweede model (periodieke trend) maakt gebruik van een neuraal netwerk (NN) om voorspellingen te maken, met het zogenaamde "niet-lineair autoregressief netwerk met exogene inputs" (NARX). Het model neemt informatie uit de afgelopen 7 dagen als invoer om recursief de voorspelling voor de volgende stap te maken. De gebruikte invoeren zijn neerslag, verdamping, onttrekking en de generieke trend van de grondwaterstand (uit het eerste model). De voorspelde grondwaterstand van de volgende dag hangt dus af van 28 waarden (4 kenmerken * 7 tijdstappen) voor de eerste invoerlaag van de NN-architectuur. Om het model te trainen, wordt de voorspelling van de grondwaterstand bij elke stap op een voorwaartse manier uitgevoerd. Na de eerste invoerlaag gebruiken we 4 (verborgen) lagen met 10 neuronen en een uitvoerlaag met een sigmoïde-functie ($=1/(1+\exp(-x))$) als verliesfunctie. Meer informatie over de NN-architectuur is gegeven in het recente BTO-rapport 'Deep Explorations' (Tian et al., 2022). De NARX-modellen zijn gebouwd in Python met de fireTS-package (Xie, 2020).

De beschreven ‘modeltrein’ is qua opzet vaststaand voor alle locaties. De parameters die de relaties tussen de invoervariabelen en grondwaterstanden beschrijven zijn getraind voor iedere locatie. Het eerste model voor de Generieke Trend, een beslisboom-regressiemodel, is getraind en gevalideerd met kruisvalidatie, waarbij de meetreeks willekeurig wordt opgedeeld in 5 delen, en het model telkens wordt getraind op 4/5 van de data en gevalideerd op het overgebleven deel (Bengio en Grandvalet, 2004). Het tweede model voor de Periodieke Trend is telkens getraind op de laatste 70% van de meetreeks en gevalideerd op de eerste 30%. Hiervoor is gekozen omdat het laatste deel van de meetreeksen de droogte van 2018-2020 omvatte, en het belangrijk was deze periodes in de modeltraining mee te nemen. Er is een test gedaan om te controleren wat het effect is van het omwisselen van deze periodes; zie bijlage I.II.

Box 3 Training en validatie

De modelopzet is deels constant gehouden voor de verschillende locaties, een deel van de parameters is getraind voor de individuele modellocaties. *Hyperparameters* zijn overkoepelende parameters die buiten de modeltraining worden gebruikt om de modelstructuur van een algoritme te controleren (vergelijkbaar met bijv. de keuze voor een lineaire of niet-lineaire responsfunctie bij impuls-reponsmodellering). De *parameters* worden tijdens de modeltraining gebruikt om, binnen de vaststaande grotere modelstructuur, de modelprestaties te controleren. De in deze studie gebruikte hyperparameters zijn gegeven in Tabel B2.1.

Tabel B2.1 Gebruikte hyperparameters voor de modelopzet.

Hyperparameters	Waarde	Model	Toelichting
Aantal <i>weak learners</i>	30	1	Weak learners verwijzen naar een verzameling modellen die tot een goed model kunnen worden samengevoegd, een ‘strong learner’. Weak learners worden vaak gebruikt in ensemble-methoden voor machine learning, waarbij meerdere leeralgoritmen worden geïntegreerd om betere voorspellende prestaties te verkrijgen dan die van de individuele leeralgoritmen.
Minimale <i>leaf size</i>	8	1	De eis voor het minimum aantal punten dat een knoop in de beslisboom van het regressiemodel moet bevatten..
<i>Learning rate</i>	0.001	1	De learning rate is een parameter in een optimalisatiealgoritme, die de stapgrootte bij elke iteratie bepaalt bij het minimaliseren van de verliesfunctie binnen het machine learning-model.
Aantal <i>folds</i> voor kruisvalidatie	5	1	De dataset wordt verdeeld in 5 <i>folds</i> voor training en kruisvalidatie.
Aantal neuronen in verborgen NN-lagen	10	2	Deel van modelarchitectuur
Aantal (verborgen) NN-lagen	4	2	Deel van modelarchitectuur
Aantal dagen aan historie voor modelinvoer	7	2	Bij iedere stap in de modellering wordt informatie van de afgelopen 7 dagen als invoer gebruikt om de grondwaterstand van de volgende dag te voorspellen
Training-validatiefracties	0.7/0.3	2	70% van de data is gebruikt voor training en 30% voor validatie.

De validatie van de twee deelmodellen verschilt door hun verschillende modelstructuur. Het eerste model op basis van EDT maakt gebruik van een typische 5-voudige kruisvalidatie, die wordt verkregen door de dataset willekeurig te verdelen in 5 delen van gelijke grootte en de prestatie van het model (Root Mean Squared Error en R^2) te berekenen voor alle deelverzamelingen. Dit geeft een ruwe schatting van de fout bij toepassing van het model op nieuwe gegevens (Bengio en Grandvalet, 2004). Voor het tweede model, een voorwaarts tijdreeksmodel, kan geen kruisvalidatie worden uitgevoerd, omdat voorlopende en afsluitende data niet kunnen worden gevalideerd. In plaats daarvan zijn de modellen telkens getraind op de laatste 70% van de tijdreeksen en gevalideerd op de eerste 30%.

2.6 Evaluatie van gedrag en betrouwbaarheid van de modellen

2.6.1 Betrouwbaarheid van de modelsimulaties in de historische periode

De mate waarin de modellen in staat waren de historische meetreeksen te beschrijven, is beschreven met de RMSE (root mean squared error) die een indicatie geeft van de gemiddelde afwijking van het model ten opzichte van de metingen (hoe lager hoe beter); en de determinatiecoëfficiënt R^2 , die aangeeft welk deel van de variantie in de grondwaterstandsreeks door het model wordt verklaard (hoe hoger hoe beter). De RMSE en R^2 zijn berekend voor de gehele historische periode (startend tussen 1990 en 2005, eindigend in 2020); en apart voor de eerste 30% van deze historische periode (die niet in de training van model 2 is meegenomen) en de laatste 70% van de periode (wel meegenomen in training model 2). De precisie van het model in de eerste 30% geeft een eerste indicatie van hoe goed het model werkt in het beschrijven van nieuwe periodes, die het nog niet in de training heeft gezien. Wel geeft dit alleen de voorspellingwaarde van het tweede model, dat de Generieke Trend al als invoer gebruikt.

Om het gedrag en de betrouwbaarheid van de modellen nog beter in beeld te brengen, zijn nog enkele toegevoegde analyses gedaan (zie bijlage I):

- Effect van verder finetunen van de modellen (parameters die eerder als vaststaand zijn genomen, zie Box 3) op de uitkomsten. Dit is getest voor enkele locaties.
- Testen van twee verschillende trainings- en validatieperiodes voor model 1 (dus voor de Generieke Trend). Voor één locatie is het regressiemodel zowel getraind op de eerste 70% en gevalideerd op de laatste 30%, als omgekeerd. Dit geeft een idee van de voorspellingwaarde van het regressiemodel (die in de verdere training-validatievergelijking niet is meegenomen). Ook geeft het een eerste idee van hoe de trainingsdataset de uitkomsten beïnvloedt.
- Vergelijking van de resultaten van de huidige opzet van model 1 met impuls-responsmodellering (Menyanthes) voor het voorspellen van de Generieke Trend. Deze analyse is uitgevoerd voor één locatie als test, maar geeft geen conclusies over de prestatie van machine learning versus impuls-responsmodellering in het algemeen.

2.6.2 Rol van externe invloeden

De rol van de verschillende externe invloeden in de grondwaterstandssimulaties is op verschillende manieren verkend. De 'feature importance' of 'expandability' van invoervariabelen (Guyon en Eliseeff, 2003) is één van de uitkomsten die worden geproduceerd door de trainingsroutine. Deze waarde geeft een indicatie van het belang van een factor in de 'beslisroute' van het model naar de grondwaterstandswaarde. Belangrijk is dat deze waarden niet geïnterpreteerd kunnen worden als een verklaarde variantie. Wanneer een model een factor kan gebruiken om simulaties te maken, betekent dit nog niet dat de factor ook in werkelijkheid een belangrijke invloed heeft.

2.6.3 Betrouwbaarheid van reproductie van waterregime-indicatoren

De toegepaste modelleringsmethode zal waarschijnlijk niet alle aspecten van de grondwaterdynamiek op een locatie goed kunnen reproduceren. Het is nog niet bekend voor welke aspecten van de waterdynamiek de toegepaste methode goed werkt en voor welke deze minder geschikt is. Dit is verkend door de waterregime-indicatoren (par. 2.3) te berekenen voor de historische modelsimulaties, en deze te vergelijken met de werkelijke waarden op basis van de metingen. Dit geeft een beeld van hoe goed de modellen verschillende aspecten van de spreiding van grondwaterstanden en het gedrag over de tijd kunnen reproduceren.

2.7 Toekomstscenario's

Om de potentiële veranderingen in watervoorraden richting de toekomst in beeld te brengen zijn de KNMI'14-scenario's gebruikt. Deze vier scenario's geven een bereik aan mogelijke ontwikkelingen van het klimaat rond 2050, met een gematigde of sterke stijging in temperatuur (G of W) en zwakke of sterke veranderingen in luchtstromingspatronen (L of H). Tijdreeksen van neerslag en referentieverdamping voor de verschillende KNMI-

scenario's zijn opgehaald van het KNMI (2015c, 2015b) voor weerstation Eelde en Hoogeveen. Het gaat om tijdreeksen van de periode 1981-2010, die zijn getransformeerd naar het klimaat rond 2050 volgens de verschillende scenario's (zie Figuur 3.11 - Figuur 3.13). Ze geven dus een mogelijke uitkomst van de weersomstandigheden in de periode 2035-2065 (KNMI, 2014, 2015a). Van de drinkwateronttrekkingen waren geen gegevens beschikbaar over de te verwachten toekomstige ontwikkeling in hoeveelheden onttrekking en seizoenspatronen voor de verschillende locaties. Van de winningen in Drenthe heeft een deel een vrij constante onttrekking over de tijd, een deel variabel over het jaar, soms met een sprong of trend, met verschillende patronen per locatie (Figuur 3.3). Als vereenvoudiging is daarom voor de toekomstige simulaties een constante onttrekking aangehouden, gelijk aan het gemiddelde van de historische periode. Verwacht wordt dat het effect van het verwaarlozen van de variatie klein is, omdat de (procentuele) variatie in grondwateronttrekking over de tijd in het algemeen veel kleiner is dan in de neerslag en verdamping. Modelleren van de historische situatie met werkelijke en constante situatie (paragraaf 3.2.2) laat zien dat de constante onttrekking de resultaten wel beïnvloedt, maar dat het effect op de uitkomsten klein is.

De KNMI'14-scenario's onderschatten waarschijnlijk de toename in extreme zomerdroogte, die in recenter onderzoek naar voren is gekomen (Philip et al., 2020; van Dorland et al., 2021). Ook is, vanzelfsprekend, de potentiële toename in (drink)watervraag in deze scenario's niet meegenomen. Om toch een groter bereik van mogelijke toekomstige grondwaterdynamiek in beeld te brengen, is een hypothetisch 'extreem' scenario getest voor één van de studielocaties. Dit scenario is gebaseerd op WH50, waarbij de verdamping is verhoogd met 25% en de neerslag verlaagd met 25% in de zomerperiode (juni-augustus). Tegelijk zijn de drinkwateronttrekkingen verhoogd met 50% over het gehele jaar; dit is een extreme versie van de verwachte toename in drinkwateronttrekking vanuit de Deltascenario's (+35% voor het meest groeiende scenario STOOM; Wolters et al. 2018).

Dit levert vijf toekomstscenario's op:

- KNMI'14 G_L2050: milde temperatuurstijging, zwakke veranderingen in luchtstroming
- KNMI'14 G_H2050: milde temperatuurstijging, sterke veranderingen in luchtstroming
- KNMI'14 W_L2050: sterke temperatuurstijging, zwakke veranderingen in luchtstroming
- KNMI'14 W_H2050: sterke temperatuurstijging, sterke veranderingen in luchtstroming
- Extreem scenario (één locatie): W_H2050 met extremere zomerdroogte en meer onttrekking.

In de resultaten is de meeste aandacht gegeven aan het WH50-scenario.

2.8 Evaluatie van veranderingen in voorraadregimes

Om de voorspelde veranderingen in grondwaterdynamiek in de toekomst in beeld te brengen, zijn de indicatoren gebruikt zoals beschreven in paragraaf 2.3. Deze zijn berekend over de laatste 17 jaar van de simulaties (technisch gezien 4-2048 – 4-2065, gebaseerd op de historische periode 1994-2010) om zo een periode van dezelfde lengte te bekijken als de historische referentieperiode. Deze indicatoren zijn vergeleken met indicatoren berekend voor historische modelsimulaties met constante onttrekkingen. Op deze manier geeft de vergelijking puur de toekomstige veranderingen door de verschillen in klimaatinvloed, los van het effect van de simulatie zelf of het verwaarlozen van de variatie in de onttrekkingen. Opgemerkt moet worden dat de toekomstscenario's dus gebaseerd zijn op de transformatie van een iets vroegere periode (1994-2010) dan de historische periode die hier als referentie is gebruikt (2003-2020). Er waren in de periode 1994-2010 twee belangrijke droogtejaren (1996 en 2003) die in de toekomstscenario's in versterkte vorm terug zullen komen. Deze periode bevat echter niet zo'n extreme zomerdroogte als 2018 (zie Figuur 3.12) waardoor de veranderingen in zomerdroogte mogelijk onderschat worden.

Voor het berekenen van de karakteristieken van laag-niveau-periodes is voor de historische modelsimulaties het 10% percentiel van de reeks zelf als grenswaarde gebruikt om periodes van lage stijghoogte af te bakenen. Voor de toekomstige simulaties is diezelfde grenswaarde van de historische simulaties gebruikt. Op die manier wordt weergegeven of lage grondwaterstanden zoals die in de huidige situatie voorkomen in de toekomst mogelijk vaker

en langer op zullen treden, en er dus meer droogteschade kan zijn. Het gebruik van de historische referentie om droogte in de toekomst te kwantificeren wordt in droogte-onderzoek vaak op deze manier gedaan; de redenering is dat waterafhankelijke functies zich op korte termijn (nog) niet hebben aangepast aan het nieuwe waterregime en dus de grens van wat 'droog' is niet direct meebeweegt met de regimeveranderingen zelf (Van Loon et al., 2016).

2.9 Testtoepassing voor oppervlaktewater

De opzet van het onderzoek was in eerste instantie om ook oppervlaktewaterstanden mee te nemen in de modellering. De meeste beschikbare reeksen van oppervlaktewater waren echter te kort om een goed model te trainen (enkele jaren); bovendien worden de waterstanden in de meeste oppervlaktewateren in het studiegebied sterk gecontroleerd, waarmee ze moeilijker te modelleren zijn met de opgezette methode op basis van weersinvloeden. Om de bruikbaarheid van de ontwikkelde methode voor oppervlaktewatertoepassingen te testen, is deze toegepast op de meetreeks van waterstanden van de Drentsche Aa. Dit riviertje kent een redelijk natuurlijk afvoerregime en wordt al lange tijd bemeten, waardoor hier wel een goede lange meetreeks beschikbaar is. Hoewel de processen zeer verschillend zijn, wordt de waterstand in de beek waarschijnlijk met name bepaald door neerslag en verdamping in voorgaande dagen en maanden, die doorwerken op de grondwaterstanden in verschillende delen van het stroomgebied en vervolgens op de oppervlaktewaterafvoer. Het oppervlaktewaterdebiet zou daarmee in principe met een zelfde type model als de grondwaterlocaties gemodelleerd moeten kunnen worden. Wel zijn de algemene modelopzet en de 'hyperparameters' (zie kader 3) voor de grondwatermodellen, die bijvoorbeeld het aantal modellagen en de tijdschaal waarover weersvariabelen worden meegenomen bepalen, gelijk gehouden. Dit kan de toepassing voor oppervlaktewater beperken.

Op dezelfde manier als voor de grondwaterlocaties is voor de Drentsche Aa een model getraind; is dit model geëvalueerd op betrouwbaarheid; en zijn de potentiële veranderingen in het waterstandsregime voor de toekomst in beeld gebracht.

3 Resultaten

3.1 Grondwaterdynamiek in de huidige situatie

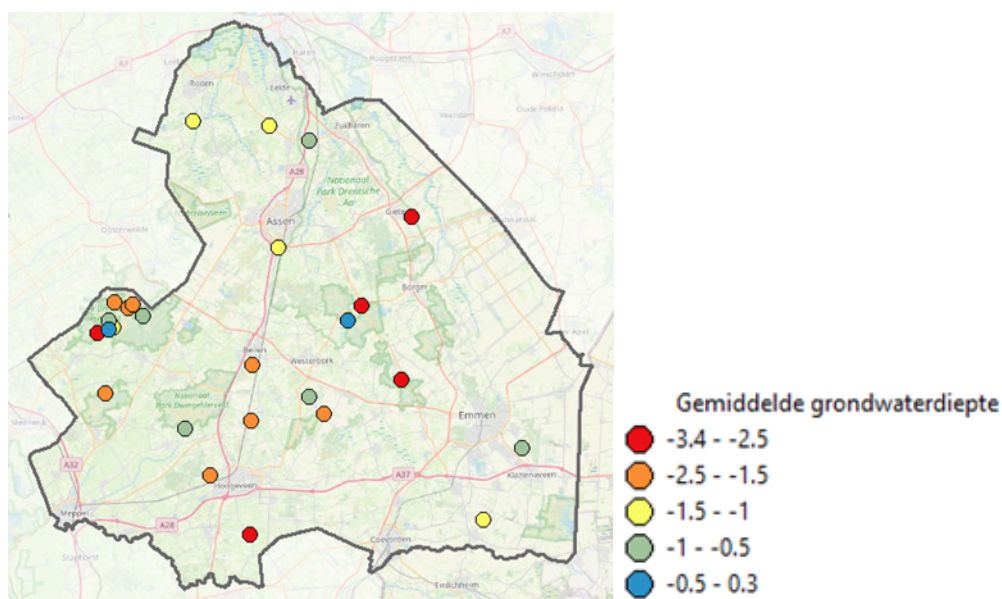
3.1.1 Waterregime-indicatoren

De karakteristieken van het waterregime in de huidige situatie zijn gegeven in Tabel 3.1. Figuur 3.1 geeft de gemiddelde grondwaterdiepte voor de verschillende locaties. De gemiddelde grondwaterstand varieert tussen 3 en 22 m+NAP, en tussen -0.3 en -3.4 m ten opzichte van maaiveld (hierna +MV). De GLG varieert tussen -4 en -0.8 meter +MV, de GHG tussen -3 en +0.1 m. De studielocaties beslaan dus een ruime variatie aan grondwaterregimes, waarbij de drogere en nattere locaties waarschijnlijk verschillend zullen reageren op weersextremen en klimaatveranderingen (zie bijv. Klimaat-effectatlas, 2022).

Periodes met lage grondwaterstanden, dus wanneer de grondwaterstand onder het 10% percentiel valt, duren gemiddeld ongeveer 60 dagen, met een grote variatie tussen locaties. Periodes met lage grondwaterstanden komen gemiddeld 0,7 keer per jaar voor, ofwel zo'n 2 keer in 3 jaar, variërend van één per drie jaar tot twee keer per jaar. In het algemeen herstellen grondwaterstanden veel sneller op locaties met ondiepe grondwaterstanden of invloed van oppervlaktewater (Peters et al., 2006; Brakkee et al., 2022) en zullen periodes met lage waterstanden hier dus korter zijn, maar wel frequenter, dan voor locaties met diepe grondwaterstanden. Het jaarlijkse tekort, of de opgetelde uitzakking onder de grenswaarde, is ongeveer 5 m*d.

Tabel 3.1: Waterregime-indicatoren voor de historische meetreeksen.

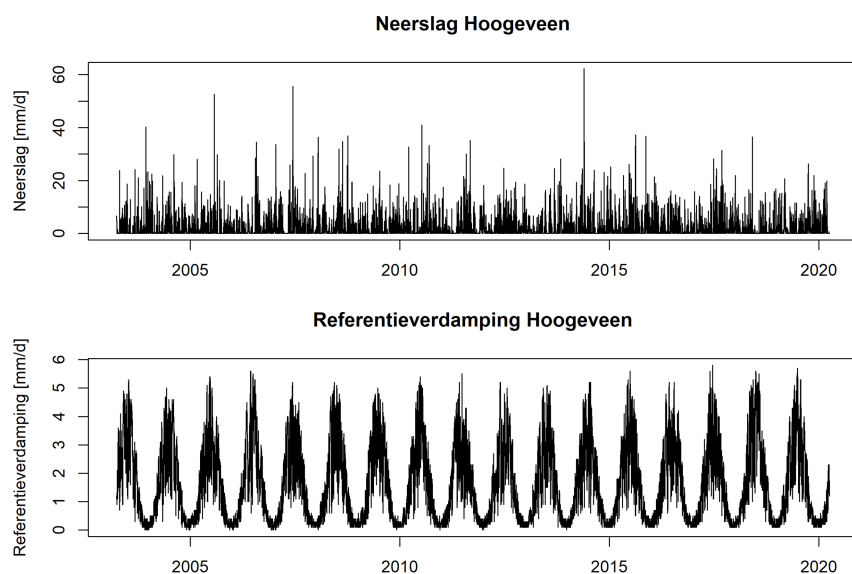
	Spreiding in stijghoogte [m+MV]					Periodes met laag niveau		
	Gemiddelde <i>h_{gem}</i>	10% percentiel <i>h₁₀</i>	90% percentiel <i>h₉₀</i>	GLG	GHG	Lengte [d]	Frequentie [n/jaar]	Tekort [m*d/j]
gemiddelde	-1.5	-2.0	-1.1	-2.0	-1.0	60.7	0.7	-5.2
min	-3.4	-3.9	-2.8	-3.8	-2.8	17.1	0.3	-9.8
max	-0.3	-0.8	0.1	-0.8	0.1	113.0	2.1	-0.8



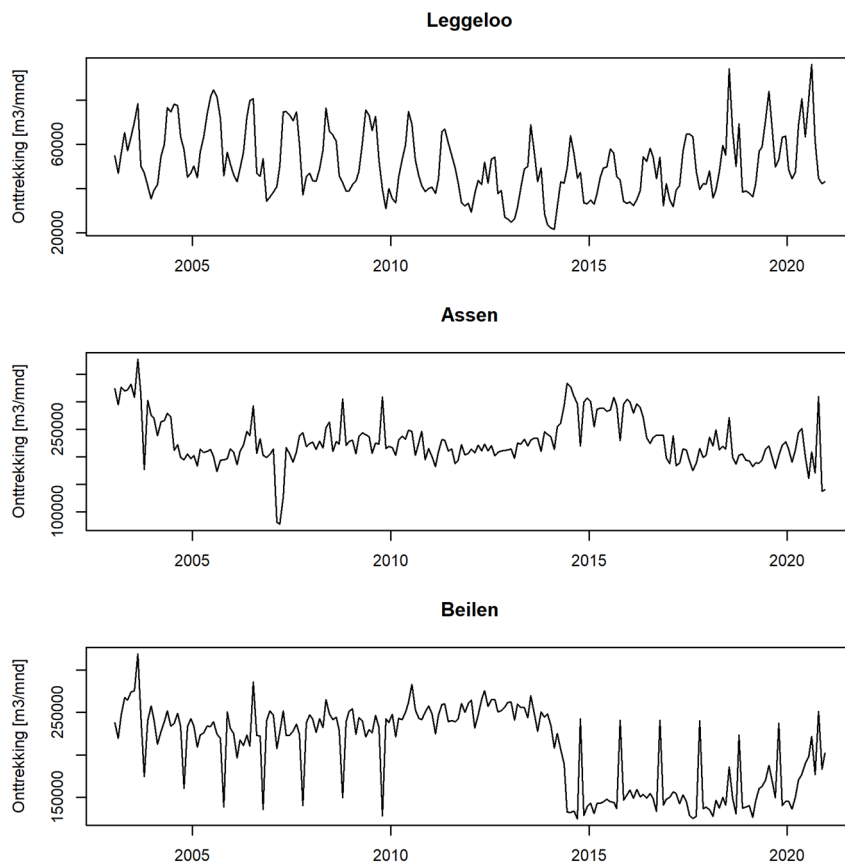
Figuur 3.1: Gemiddelde grondwaterdiepte (in meters) voor de studielocaties.

3.1.2 Variatie in externe factoren

De externe factoren die als invoer zijn gebruikt, zijn de neerslag, referentieverdamping en drinkwateronttrekkingen. De variatie over de tijd in deze factoren is gevisualiseerd in Figuur 3.2 en Figuur 3.3. De neerslag toont een vrij willekeurige spreiding zonder duidelijke seizoenspatronen; de verdamping laat een karakteristieke seizoensfluctuatie zien. Van de winlocaties voor drinkwater hebben sommige een duidelijke seizoensvariatie in de onttrekkingen, terwijl dit voor veel andere locaties afwezig is. Sommige onttrekkingen zijn ongeveer constant gebleven over de afgelopen 15 jaar, terwijl in andere grotere verschuivingen zijn geweest, mogelijk door bijvoorbeeld veranderingen in vergunningen. Er zijn geen consistente patronen over de verschillende locaties. De meeste winningen hebben een variatie in onttrekking van 10-20%, relatief klein ten opzichte van de variatie in neerslag en verdamping. Figuur 3.3 geeft de dynamiek voor drie kenmerkende winlocaties; meer details staan in Bijlage II.II.



Figuur 3.2: Variatie in sturende factoren voor de modellen: neerslag en referentieverdamping van weerstation Hogeveen (KNMI).



Figuur 3.3: Variatie in sturende factoren voor de modellen: onttrekkingen van drie voorbeelden van drinkwaterwinlocaties.

3.2 Gedrag en betrouwbaarheid van de modellen

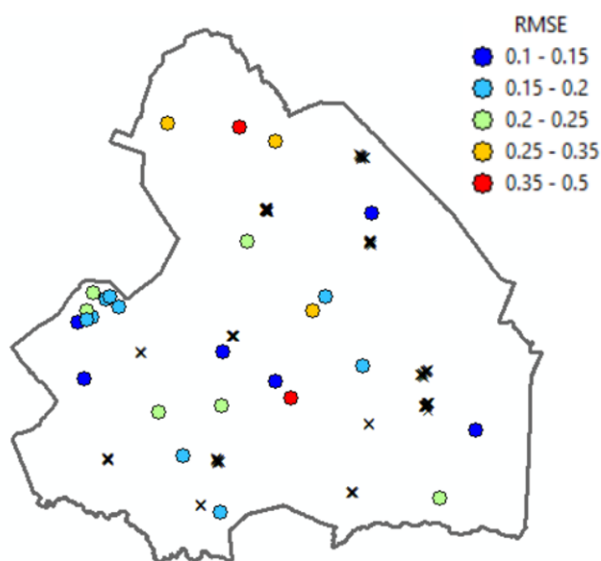
3.2.1 Prestatie van de modelsimulaties in de historische periode

Tabel 3.2 laat de modelfit zien van de getrainde modellen over de gehele historische meetreeks. De modellen zijn gegroepeerd voor de locaties waarin onttrekkingen wel als verklarende variabele zijn meegenomen (onttrekking aanwezig binnen 10 km) en locaties waar het model alleen op neerslag en verdamping is gebaseerd (geen drinkwateronttrekking binnen 10 km). De gemiddelde fout van de simulaties ten opzichte van de metingen is gegeven door de RMSE. Deze is in de tabel weergegeven als absolute waarde en als fractie van de gemiddelde grondwaterdiepte, omdat kleine fouten op locaties met ondiepe grondwaterstanden een veel grotere invloed hebben dan op locaties met diepe standen. De R^2 geeft aan in hoeverre het model de variatie in de meetwaarden kan verklaren. De modellen konden in het algemeen redelijk de historische dynamiek reproduceren, met een gemiddelde fout van zo'n 20 cm en verklaarde variantie van 76%. Er is geen duidelijk ruimtelijk patroon in de modelfout (Figuur 3.4). Voor locaties waar onttrekkingen niet in de modellen zijn meegenomen, doen de modellen het minder goed, met een lage verklaarde variantie. Dit betekent niet direct dat drinkwateronttrekkingen ook daadwerkelijk een belangrijke invloed vormen op de grondwaterstanden in de provincie²; dit wordt verder besproken in paragraaf 3.2.2 en 4.1.2.

² Het meenemen van onttrekking als extra variabele voegt een extra vrijheidsgraad toe aan een model, waardoor een model puur door het toevoegen van een parameter al een betere fit kan opleveren.

Tabel 3.2: Modelfit voor de 25 gemodelleerde locaties met de NARX-decompositiemethode. RMSE: root mean squared error, of de gemiddelde fout, in meters en als fractie van de grondwaterdiepte. R^2 : verklaarde variantie.

Locaties met onttrekking meegenomen	RMSE [m]	RMSE, fractie van grondwaterdiepte	R^2
Gemiddelde	0.19	0.17	0.76
Min	0.12	0.05	0.69
Max	0.29	0.45	0.83
Locaties zonder onttrekking meegenomen	RMSE	RMSE, fractie grondwaterdiepte	R^2
Gemiddelde	0.36	0.38	0.34
Min	0.24	0.17	0.25
Max	0.46	0.93	0.45



Figuur 3.4: Gemiddelde fout RMSE [m] voor de verschillende modellocaties. De zwarte kruisjes geven de locaties van drinkwateronttrekkingsputten; voor iedere grondwatermodellocatie zijn de onttrekkingen in een straal van 10 km als invoer gebruikt.

Het bekijken van de modelfit in de validatieperiode, die niet in de training is meegenomen, geeft inzicht in de betrouwbaarheid van de modellen voor het voorspellen van nieuwe situaties. Model 1 is gevalideerd via kruisvalidatie. Model 2 (dat de uitkomst van model 1 als invoer gebruikt) is telkens getraind op de laatste 70% van de meetperiode en gevalideerd op de eerste 30%. Tabel 3.3 geeft de validatieresultaten van model 2 voor de Periodieke Trend. De tabel geeft een vergelijking van de verklaarde variantie (R^2) in de kalibratieperiode en de validatieperiode. De verklaarde variantie is in het algemeen iets kleiner voor de validatieperiode dan voor de kalibratieperiode, maar het verschil is relatief klein. In Bijlage I.II is voor één locatie als test een zelfde 30%-70%-validatie uitgevoerd voor model 1. De resultaten suggereren dat model 1 mogelijk iets meer last heeft van overfitting. Ook is bekeken wat het effect is van het omwisselen van de kalibratie- en validatieperiodes voor model 1 (Generieke Trend). Voor de geteste locatie maakte dit geen duidelijk verschil, ook doordat zowel het begin als het einde van de historische meetreeks een hevige droogte omvatten (2003 en 2018).

Tabel 3.3: Verklaarde variantie (R^2) voor de 25 gemodelleerde locaties, vergelijking tussen de kalibratieperiode (laatste 70% van de periode beschikbare data) en de validatieperiode (laatste 30%, niet meegenomen in de modeltraining).

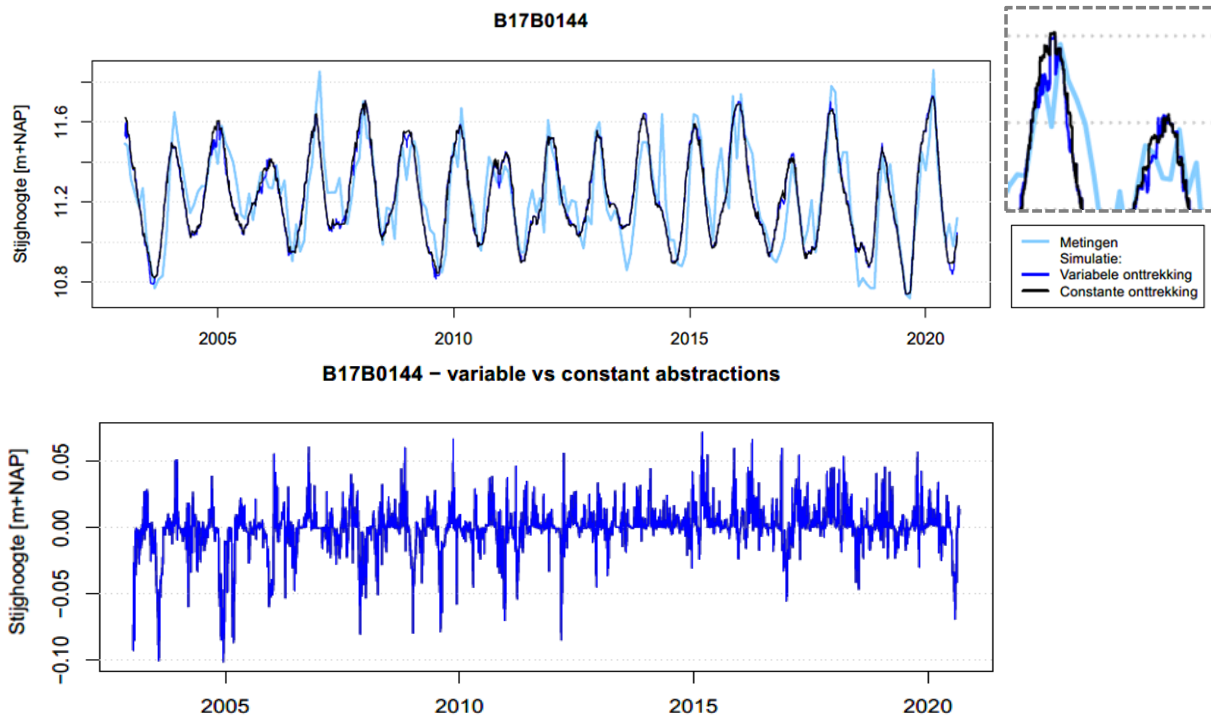
Locaties met onttrekking meegenomen	R^2	R^2 kalibratie (laatste 70%)	R^2 validatie (eerste 30%)	Vershil R^2 validatie - kalibratie
Gemiddeld	0.76	0.76	0.71	-0.05
Min	0.69	0.63	0.56	-0.21
Max	0.83	0.80	0.86	0.06
Locaties zonder onttrekking meegenomen	R^2	R^2 kalibratie (laatste 70%)	R^2 validatie (eerste 30%)	Vershil R^2 validatie - kalibratie
Gemiddeld	0.34	0.37	0.33	-0.05
Min	0.25	0.27	0.29	-0.11
Max	0.45	0.45	0.37	0.08

3.2.2 Rol van externe invloeden

De rol van externe factoren in een model is voor een deep learning-model moeilijker te interpreteren dan voor een standaard statistisch model. Model 1, voor de Generieke Trend, werkt met invoer van neerslag en verdamping en kan hiermee goed de algehele patronen van de meeste reeksen simuleren. Toevoegen van drinkwateronttrekkingen had geen duidelijk verbeterend effect op deze modellen. De Generieke Trend maakte meestal het grootste deel van de variatie in de reeksen uit; dit betekent dat verdamping en neerslag de belangrijkste factoren zijn voor de modellen als geheel. Dit is in overeenstemming met de verwachting, aangezien de meeste modellocaties niet erg dicht bij drinkwateronttrekkingslocaties liggen.

Model 2 voor de Periodieke Trend omvat wel de neerslag, verdamping én drinkwateronttrekkingen. Het belang van deze factoren voor model 2 wordt gegeven door hun relatieve “feature importance”-waarden. In het algemeen kent model 2 een ruwweg even groot belang toe aan de verschillende externe invloeden. Merk op dat dit alleen gaat om model 2, dat enkel de variatie bovenop de Generieke Trend modelleert. De “feature importance” van drinkwaterwinningen toont echter weinig ruimtelijke patronen, terwijl verwacht zou worden dat deze dicht bij drinkwaterwinningen groter zou zijn (Bijlage II.IV). Wanneer de historische periode opnieuw wordt gesimuleerd met constante onttrekkingen (Figuur 3.5) blijkt dat de verschillen minimaal zijn. Dit suggereert dat de invloed van de externe factoren in model 2 relatief klein is voor de modelresultaten als geheel. Dit wordt verder besproken in paragraaf 4.1.2.

Niet alle factoren die de grondwaterstand op een locatie kunnen beïnvloeden konden in dit onderzoek worden meegenomen, zoals landbouwonttrekkingen en oppervlaktewaterbeheer. De mogelijke rol van deze factoren wordt in de discussie (4.1.2) verder besproken.



Figuur 3.5: Vergelijking van modelsimulaties voor de historische periode met werkelijke en constante onttrekkingen. Boven: Vergelijking van beide simulatiereeksen met de metingen, rechts bovenin ingezoomd op de winters van 2005 en 2006. Onder: verschil tussen simulaties met constante en daadwerkelijke onttrekkingen.

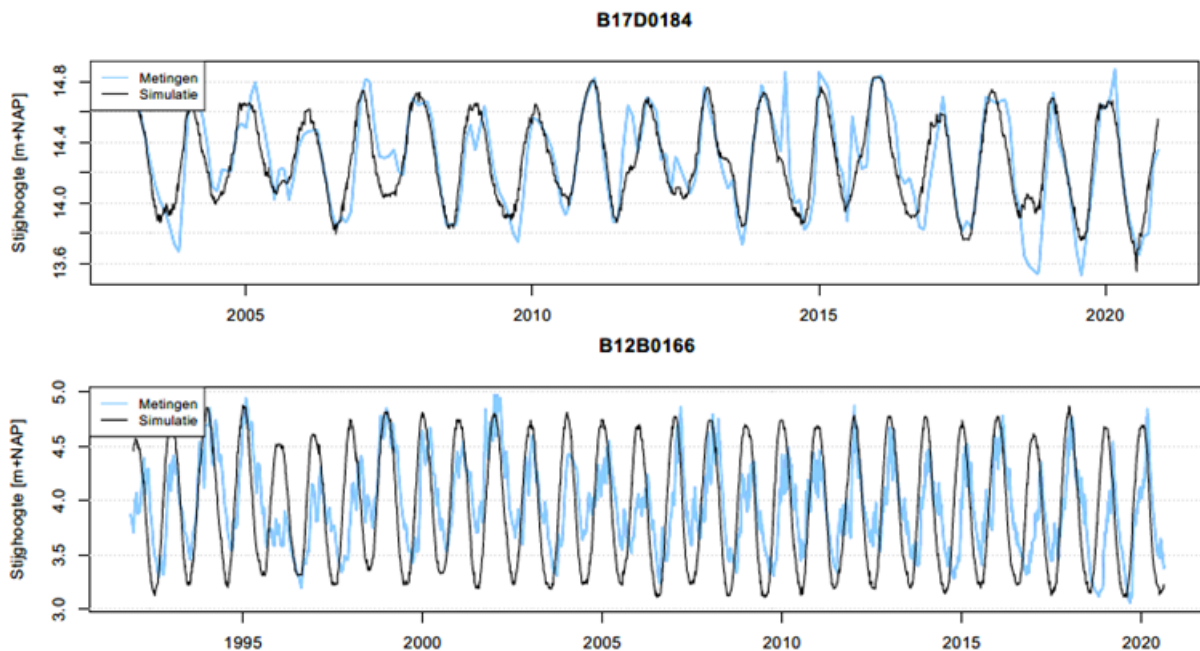
3.2.3 Reproductie van waterregime-indicatoren in de historische periode

Gemiddelde stijghoogten

Tabel 3.4 geeft weer in hoeverre de modellen in de historische periode in staat waren de regime-indicatoren gerelateerd aan jaar- en seizoensstijghoogten te reproduceren. De tabel geeft de afwijkingen weer van de regime-indicatoren op basis van de simulaties ten opzichte van de echte indicatoren op basis van de meetreeksen, in absolute waarden en percentages. De langjarig gemiddelde grondwaterstand wordt goed gereproduceerd, met een fout van op zijn hoogst enkele centimeters. De seizoensgemiddelde grondwaterstanden worden minder goed gereproduceerd. De standen in het najaar en de winter worden consistent overschat, terwijl de waarden in de standen in het voorjaar en de zomer juist worden onderschat. Voor sommige locaties leidt dit tot grote afwijkingen in de seizoensgemiddelde grondwaterdieptes. Het blijkt dat voor veel reeksen naast de over- en onderschattingen in winter en zomer een lichte verschuiving in de tijd plaatsvindt in de simulaties, waardoor de timing van hoge en lage waterstanden niet goed wordt gesimuleerd (Figuur 3.6).

Tabel 3.4: Afwijkingen in de gemodelleerde regime-indicatoren voor jaar- en seizoensstijghoogten, als verschil tussen de indicator voor de modelsimulaties en voor de meetreeks. Afwijkingen worden gegeven voor *hgem* (gemiddelde grondwaterstand) en de vier seizoenen (*hwinter* – *hnajaar*). Voor iedere afwijking wordt het gemiddelde gegeven over 25 locaties, het gemiddelde van de absolute afwijkingen, en de laagste en hoogste afwijkingen. Afwijkingen zijn gegeven in meters en als percentage van de meetwaarde. *Hoge percentages ontstaan voor enkele locaties met zeer ondiepe grondwaterstand (grondwaterdiepte bijna nul).

	Afwijking <i>hgem</i>		Afwijking <i>hwinter</i>		Afwijking <i>hvoorjaar</i>		Afwijking <i>hzomer</i>		Afwijking <i>hnajaar</i>	
	abs. [m]	%	abs. [m]	%	abs. [m]	%	abs. [m]	%	abs. [m]	%
Gemiddeld	0	0%	0.12	15%	-0.14	-28%	-0.13	-9%	0.14	10%
Gemiddeld absoluut	0.01	1%	0.11	14%	0.13	25%	0.12	8%	0.13	9%
Minimum	-0.03	-6%	-0.05	-2%	-0.33	-348%*	-0.38	-22%	0.05	3%
Maximum	0.03	2%	0.43	43%	-0.04	-4%	0.02	1%	0.35	41%



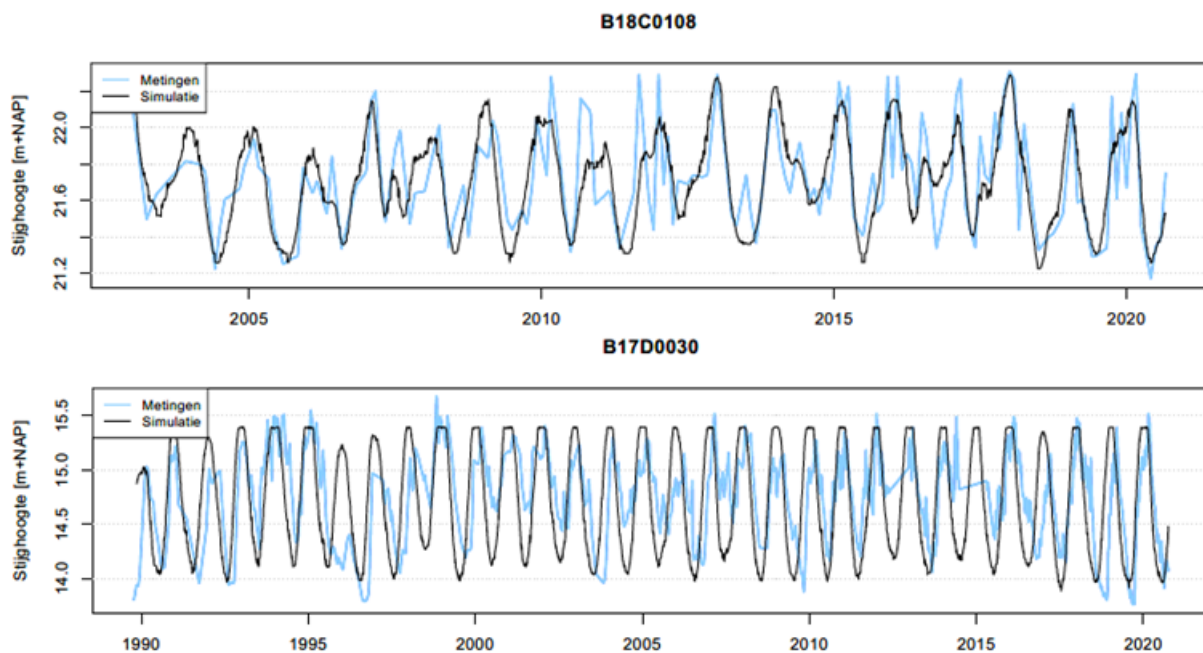
Figuur 3.6: Voorbeeld van een locatie waarvoor de historische periode goed wordt gereproduceerd (boven) en een waar de seizoenswaarden slecht terugkomen door over- en onderschattingen in winter en zomer en een verschuiving in de timing (onder).

Spreiding van grondwaterstanden: GxGs en percentielen

Tabel 3.5 geeft weer in hoeverre de spreiding in grondwaterstijghoogten in de tijd wordt gereproduceerd door de modellen voor de verschillende locaties. De tabel geeft de afwijkingen weer van de percentielen en de GxGs op basis van de simulaties ten opzichte van de echte percentiel- en GxG-waarden op basis van de meetreeksen, in absolute waarden en percentages. Te zien is dat de spreiding in stijghoogten in het algemeen goed wordt gereproduceerd, met afwijkingen van slechts enkele centimeters. Voor sommige locaties zijn de afwijkingen wel groter (Figuur 3.7). Ook de GLG en GHG worden in het algemeen goed gesimuleerd in de historische periode door de getrainde modellen, waarbij er geen duidelijk verschil is tussen de GLG en GHG.

Tabel 3.5: Afwijkingen in de gemodelleerde regime-indicatoren voor de spreiding van grondwaterstanden, als verschil tussen de indicator voor de modelsimulaties en die voor de meetreeks. Afwijkingen worden gegeven voor de 5%, 20%, 50% en 90% percentielen (h5 – h90) en de GLG en GHG. Voor iedere afwijking wordt het gemiddelde gegeven over 25 locaties, het gemiddelde van de absolute afwijkingen, en de laagste en hoogste afwijkingen over alle locaties. Afwijkingen zijn gegeven in meters en als percentage van de meetwaarde. Merk op dat de procentuele afwijkingen voor de hoge standen automatisch hoger worden, doordat deze als percentage van de meetwaarde in m-MV zijn weergegeven.

	Afwijking h5		Afwijking h20		Afwijking h50		Afwijking h80		Afwijking GLG		Afwijking GHG	
	abs. [m]	%	abs. [m]	%	abs. [m]	%	abs. [m]	%	abs. [m]	%	abs. [m]	%
Gemiddeld	0.05	3%	-0.02	-1%	-0.03	-5%	0.04	-41%	-0.01	0%	-0.02	7%
Gemiddeld absoluut	0.09	5%	0.07	4%	0.04	6%	0.10	53%	0.07	4%	0.09	22%
Minimum	-0.17	-9%	-0.25	-14%	-0.15	-63%	-0.15	-112%	-0.24	-13%	-0.21	-69%
Maximum	0.37	37%	0.16	23%	0.04	5%	0.36	45%	0.19	24%	0.27	200%



Figuur 3.7: Voorbeeld van een locatie waarvoor de spreiding in stijghoogten in de historische periode goed wordt gesimuleerd (boven) en een waar de spreiding slecht wordt gereproduceerd door over- en onderschattingen in winter en zomer (onder).

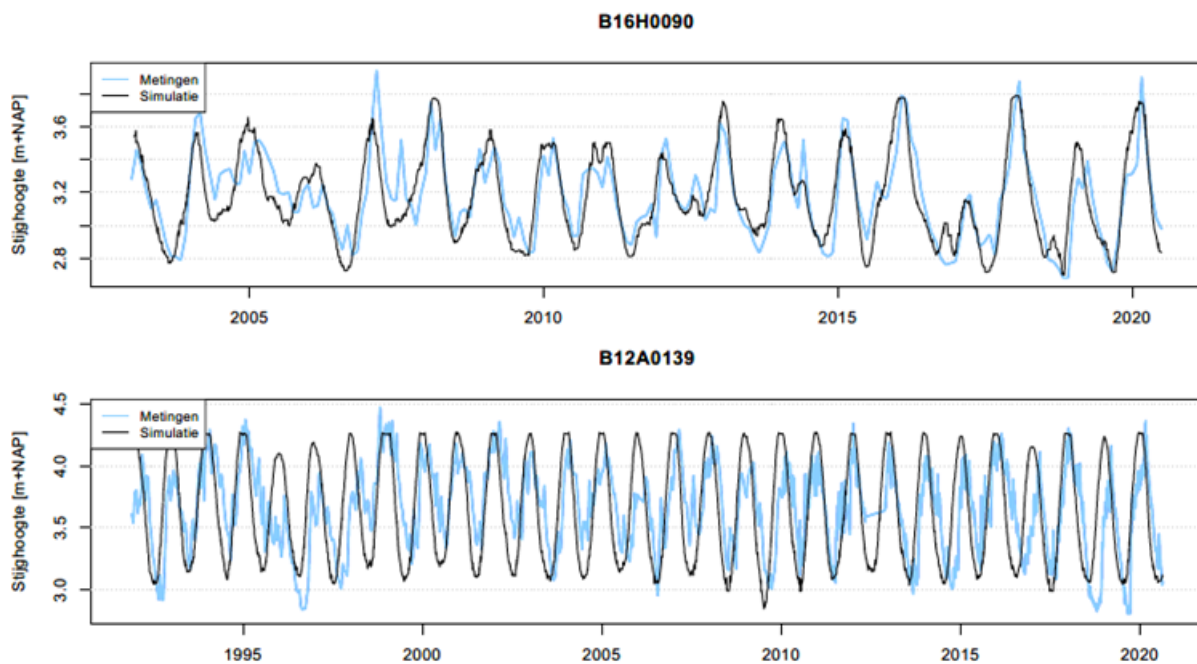
Periodes met lage grondwaterstanden

Tabel 3.6 geeft weer in hoeverre de kenmerken van periodes met lage grondwaterstanden (zie par. 2.3) worden gereproduceerd door de modellen voor de historische periode. De tabel geeft de afwijkingen weer in de lengte en frequentie van periodes met lage grondwaterstanden en het jaarlijkse tekort, in absolute waarden en percentages.

Het blijkt dat de kenmerken van laag-niveau-periodes moeilijker zijn om goed te reproduceren dan de spreiding in waarden die hiervoor is besproken. De lengte en frequentie van laag-niveau-periodes worden nog redelijk gemodelleerd, maar de tekorten (oppervlak onder het 10% percentiel) veel slechter. De lengte van laag-niveau-periodes lag in de metingen gemiddeld rond de 65 dagen. De gemiddelde absolute afwijking hierin is substantieel met 22 dagen. De frequentie van periodes met een laag niveau, die in de metingen rond 0.7/jaar ligt, wordt iets beter gesimuleerd met een gemiddelde fout van 0.23/jaar. Voor alle drie de indicatoren zijn er grote uitschieters. De kenmerken van laag-niveau-periodes zijn erg gevoelig voor de temporele patronen in de meet- en simulatiereeksen, waardoor de fouten in deze indicatoren voor sommige locaties al snel groot zijn als de temporele patronen niet goed kloppen (Figuur 3.8).

Tabel 3.6: Afwijkingen in de gemodelleerde kenmerken van laag-niveau-periodes, als verschil tussen de indicator voor de modelsimulaties en die voor de meetreeks. Afwijkingen worden gegeven voor de lengte, frequentie en tekort van laag-niveau-periodes. Voor iedere afwijking wordt het gemiddelde gegeven over 25 locaties, het gemiddelde van de absolute afwijkingen, en de laagste en hoogste afwijkingen over alle locaties. Afwijkingen zijn gegeven in werkelijke waarden en als percentage van de meetwaarde.

	Periodes met laag niveau					
	Afwijking Lengte		Afwijking Frequentie		Afwijking Tekort	
	abs. [d]	%	abs. [n/j]	%	abs. [m*d/j]	%
Gemiddeld	-9.2	1%	0.06	13%	0.67	4%
Gemiddeld absoluut	22	43%	0.23	33%	3.0	61%
Minimum	-53	-68%	-1.3	-95%	-14	-308%
Maximum	58	199%	0.71	125%	6.9	99%



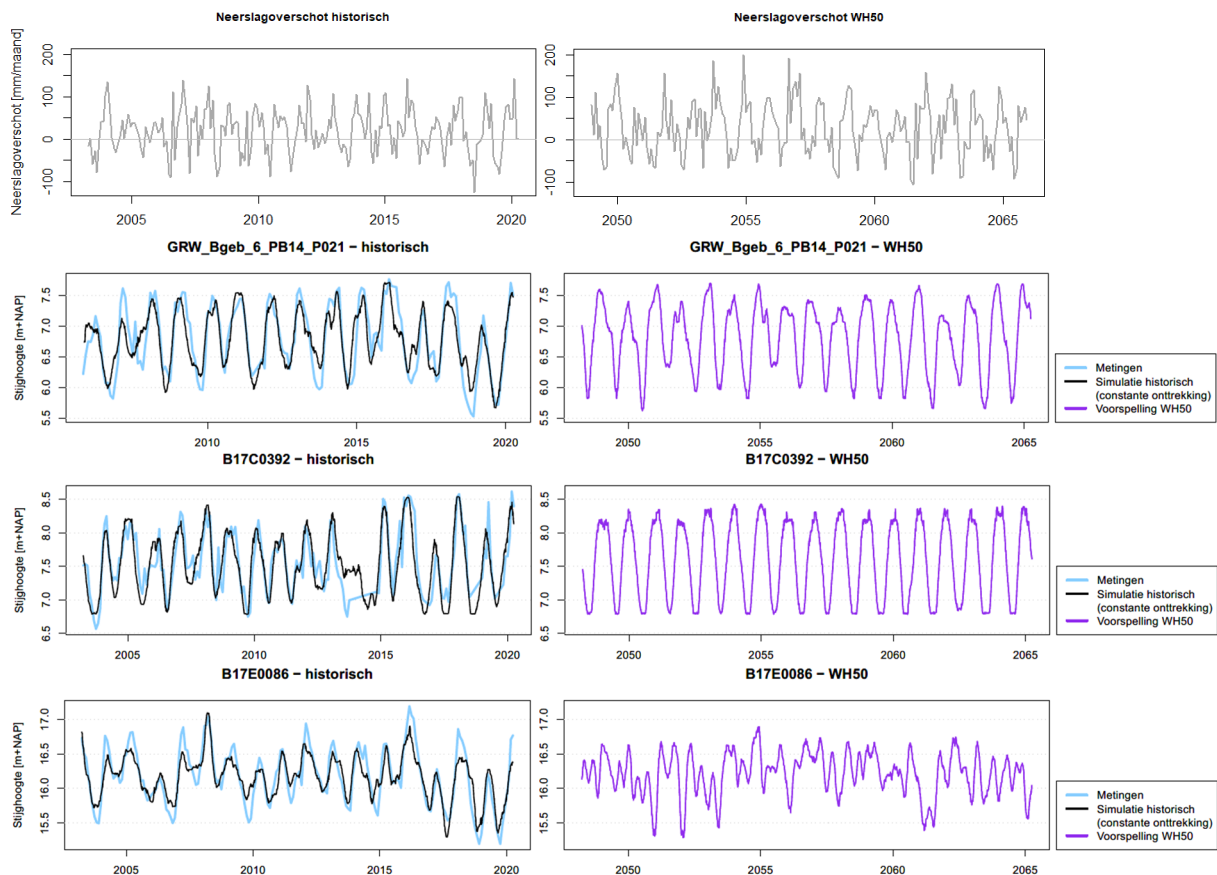
Figuur 3.8: Voorbeeld van een locatie waarvoor de kenmerken van periodes met lage stijghoogten goed worden gereproduceerd (boven) en een waar de kenmerken slecht worden gesimuleerd (onder). Voor de onderste locatie is de simulatie regelmatig dan de meetreeks, waardoor periodes met lage waterstanden vaker voorkomen en langer duren, terwijl het totale tekort wel redelijk wordt gereproduceerd.

Selectie van indicatoren

Op basis van voorgaande resultaten zijn enkele indicatoren geselecteerd voor het evalueren van de veranderingen in waterregime richting de toekomst. Er is gekozen om de gemiddelde grondwaterstijghoogte *hgem*, de gemiddelde laagste en hoogste grondwaterstand *GLG* en *GHG*, en het 5% percentiel *h5* als indicatoren te gebruiken als beschrijving van de huidige en voorspelde spreiding in stijghoogten. De gemiddelden per seizoen worden slecht gereproduceerd en zijn daarom niet betrouwbaar genoeg. Ook zijn de kenmerken van periodes met lage stijghoogten voor de toekomst niet betrouwbaar, omdat de modellen minder goed werken om de temporele patronen te vertalen naar de toekomst (zie par. 3.2.4). Hierdoor worden voorspelde veranderingen in de kenmerken van laag-niveau-periodes waarschijnlijk eerder veroorzaakt door afwijkingen in de simulatie van tijdspatronen dan door veranderingen in droogtekenmerken zelf.

3.2.4 Vertaling van temporele patronen naar de toekomst en potentiële limitaties in bereik

De modelfit en indicatoren in de historische periode geven niet alle informatie over de betrouwbaarheid van de voorspellingen voor toekomstscenario's. Belangrijk is ook of de temporele patronen in de reeksen op een goede manier worden vertaald naar de voorspellingen, wanneer geen observaties meer beschikbaar zijn. Uit de simulaties voor de toekomstscenario's bleek dat de voorspelde grondwaterstandsreeksen vaak onwaarschijnlijke temporele patronen lieten zien die afweken van de historische meet- en simulatiereeksen. Het gaat met name om een verandering in temporele fluctuaties, waarbij bij veel reeksen de toekomstige voorspelde reeks regelmatig was dan de historische. Ook leken de voorspelde waarden soms te 'blokkeren' in het hoge of lage bereik. Figuur 3.9 geeft voorbeelden van reeksen waarbij dit optreedt (zie ook paragraaf 3.3.1 voor de variatie in weersomstandigheden). De verandering in regelmatigheid en het 'plateau-effect' voor sommige reeksen worden hieronder toegelicht; een discussie van oorzaken volgt in de discussiesectie.



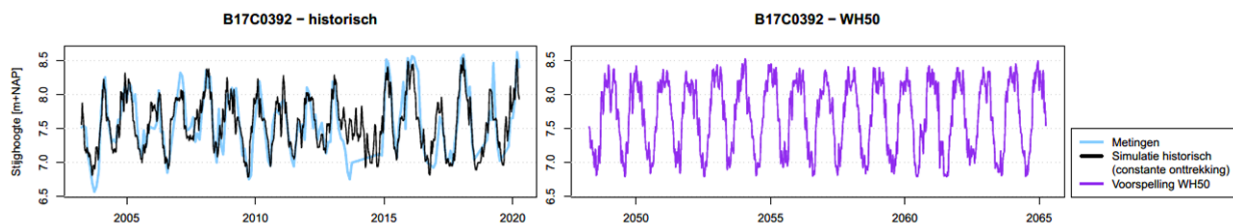
Figuur 3.9: Voorbeelden van de vertaling van temporele patronen naar de toekomstvoorspellingen. Het bovenste paneel geeft het maandelijkse neerslagoverschot voor de historische en toekomstige periodes, de gekleurde panelen de gemodelleerde grondwaterstanden. Voor de bovenste grondwaterlocatie worden de temporele patronen goed vertaald naar de voorspellingen; voor de middelste locatie wordt de reeks zeer regelmatig (exemplarisch voor een groot deel van de locaties). Voor enkele locaties (onderste paneel) neemt de temporele fluctuatie juist toe.

Verandering in regelmaat

De voorspelde reeksen voor de toekomstscenario's waren vaak veel regelmatiger dan de historie en lieten weinig variatie zien tussen jaren. Voor een enkele locatie namen de temporele fluctuaties juist toe (Figuur 3.9). De variatie in standen tussen jaren is weergegeven in Tabel 3.7. De historische simulaties laten al minder variatie tussen jaren zien dan de observaties. Deze variatie neemt verder af in de meeste voorspelde reeksen. Dit betekent dat de voorspellingen een regelmatiger karakter hebben dan de historie. Voor het voorspellen van het toekomstige bereik aan waarden is dit geen heel groot probleem, maar wel voor het weergeven van de karakteristieken van extreme periodes zoals droge zomers. Dit punt wordt verder besproken in de discussie. Voor de toekomstvoorspellingen is gebruik gemaakt van constante extractiewaarden. Dit heeft echter maar een zeer klein effect op de voorspelde reeksen (zie Figuur 3.5) en kan dus de regelmaat niet verklaren. In Appendix I.I is getest wat voor effect toegevoegde fine-tuning van de modellen heeft op de toekomstvoorspellingen. Dit leek voor sommige locaties de regelmatigheid te verminderen, maar de verandering in temporele patronen in de toekomstvoorspellingen ten opzichte van de historie en het verlies aan verschillen tussen jaren bleven bestaan (zie Figuur 3.10). De regelmaat in voorspelde reeksen zou versterkt kunnen zijn door het voorkomen van plateaus, doordat, wanneer de reeksen elke zomer op dezelfde waarde blijven hangen, deze ook elk najaar op dezelfde waarde starten en op deze manier extra regelmaat wordt afgedwongen.

Tabel 3.7: Variatie in jaarlijkse minimum- en maximumstijghoogten tussen verschillende jaren. De variatie is weergegeven als standaarddeviatie van de minimum- en maximumstijghoogten per jaar voor de historische meetreeksen, de simulaties voor de historische periode, en de voorspelde reeksen voor het WH50-scenario in de 17-jarige studieperiodes.

Interjaarlijkse variatie in minimum stijghoogten (standaarddeviatie, m)		Interjaarlijkse variatie in maximum stijghoogten (standaarddeviatie, m)	
Historische observaties	0.24	Historische observaties	0.20
Historische simulaties	0.21	Historische simulaties	0.14
Voorspellingen WH50	0.19	Voorspellingen WH50	0.06



Figuur 3.10: Simulatieresultaten voor locatie B17C0392 na extra fine-tuning van de (hyper)parameters. Zie Figuur 3.9 voor het origineel met generieke parameters.

Plateau-effect en potentiële limitaties in simulatiebereik

De gesimuleerde reeksen voor de historie en de toekomst lijken in sommige gevallen aan de onderkant of de bovenkant te worden gelimiteerd, waarbij ze op bepaalde waarden hangen (Figuur 3.9). Dit 'plateau'-effect is soms al te zien in de historische simulaties, en komt in de toekomstige situaties vaker terug. Er kunnen verschillende fenomenen spelen, die we hieronder verder onderzoeken:

- Het plateau-effect zou veroorzaakt kunnen worden doordat de modellen niet volledig zijn ge-fine-tuned (een deel van de hyperparameters is constant gehouden voor alle locaties, zie par. 2.5).
- De modelsimulaties zouden direct gelimiteerd kunnen worden door de simulatiewaarden in het trainingsbereik. In principe komt een verlies aan variatie in simulaties ten opzichte van de originele metingen vaak voor in (statistische) modellen, doordat de ruis in de originele metingen niet wordt meegenomen (bijv. Brakkee et al., 2022). Sommige machine learning-methodes (m.n. classificatie) kennen echter een harde beperking in de mogelijke simulatiewaarden die geproduceerd kunnen worden tot de historisch geobserveerde waarden. De gebruikte decision tree-methode voor model 1 zou beperkt kunnen zijn tot een discrete set aan waarden; het gebruik van een ensemble zou hier echter minder last van moeten hebben. Ook de NARX-methode van model 2 zou in principe geen harde waardelimitaties moeten kennen omdat deze het accumulerende effect van het weer over meerdere dagen meeneemt. Toch onderzoeken we het bereik van de simulaties om mogelijke beperkingen hierin te herkennen.
- Ook als de modellen geen harde beperking kennen in mogelijke outputwaarden, zouden ze moeite kunnen hebben met goede simulaties buiten het trainingsbereik en hier slechter presteren.

Het effect van de fine-tuning van de modellen is bekeken in Appendix I.I. Extra fine-tuning leek het voorkomen van plateaus in de simulaties te verhelpen (zie Figuur 3.10) en lijkt daarmee een belangrijke oorzaak van dit probleem. Het waardebereik van de simulaties werd door de optimalisatie echter niet beïnvloed.

Om het bereik van de gesimuleerde grondwaterstanden te controleren, zijn de minimum- en maximumwaarden van de observaties, historische simulaties en toekomstige voorspellingen bekeken. Vergelijking van het bereik van historische simulaties en historische observaties per modellocatie liet zien dat het bereik van de historische simulaties altijd beperkt was tot het bereik van de historische observaties, waarbij de laagste gesimuleerde stijghoogten consistent gelijk aan of hoger waren dan de gemeten laagste stijghoogten, met verschillen tot 40 cm. Vergelijking tussen de toekomstige voorspellingen voor het WH50-scenario en de historische simulaties liet zien dat de toekomstvoorspellingen niet gelimiteerd worden door het bereik van de historische simulaties, en dat de laagste voorspelde stijghoogten in WH50 tot 14 cm lager uit konden komen dan de laagste gesimuleerde waarden in de

historische periode. Vergelijking van de toekomstvoorspellingen met de historische *observaties* liet wel zien dat de voorspellingen altijd binnen het bereik van de historische *observaties* bleven. Voor de simulaties met de verder geoptimaliseerde modellen (bijlage I.I) gold hetzelfde: de voorspelde waarden voor de toekomst gingen wel buiten het bereik van de historische *simulaties*, maar niet buiten het bereik van de historische *observaties*. Dit gaat puur om de absolute waarden van de grondwaterstanden op dagbasis: de modellen voorspellen wel duidelijk veranderingen in de frequentie en patronen van de grondwaterstanden (zie par. 3.3).

Om de beperking in voorspelde waarden ten opzichte van het trainingsbereik te testen, is ook bekeken wat het effect is van het trainen van model 1 op het laatste deel van de historische periode (mét de droogte van 2018) en het eerste deel van de periode (zonder 2018) voor één locatie (Appendix I.II). Voor deze locatie maakte dit weinig verschil, mogelijk doordat de droogte van 2003 een soortgelijk effect had op de grondwaterstanden als die van 2018. Om de omgang met situaties buiten het historisch bereik beter te testen, zijn uitgebreidere analyses nodig voor meer locaties, bijvoorbeeld het trainen van de modellen op natte periodes en testen op droge periodes.

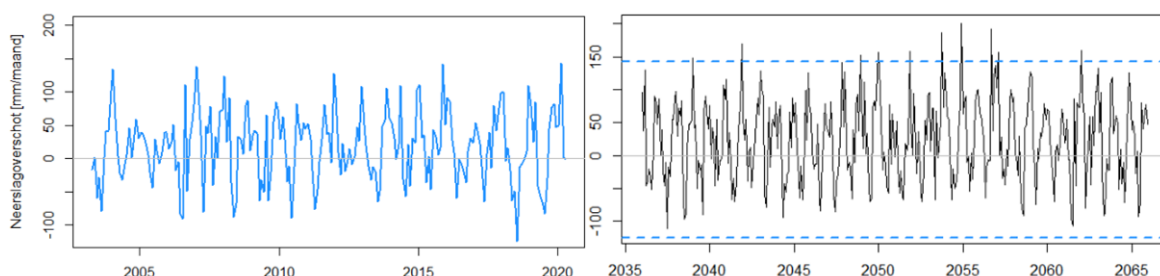
Alles bij elkaar lijkt de modelopzet dus een bepaalde beperking te vertonen in wat het kan voorspellen, en heeft waarschijnlijk moeite met het voorspellen van situaties buiten het historisch bereik. Dit betekent dat de voorspellingen voor de toekomstscenario's onderschattingen kunnen zijn. Het zou kunnen dat een aangepaste modelleringsmethode wel beter is in het maken van voorspellingen buiten het historisch bereik. In par. 4.2 wordt dit punt verder besproken.

3.3 Veranderingen in grondwaterdynamiek in de toekomst

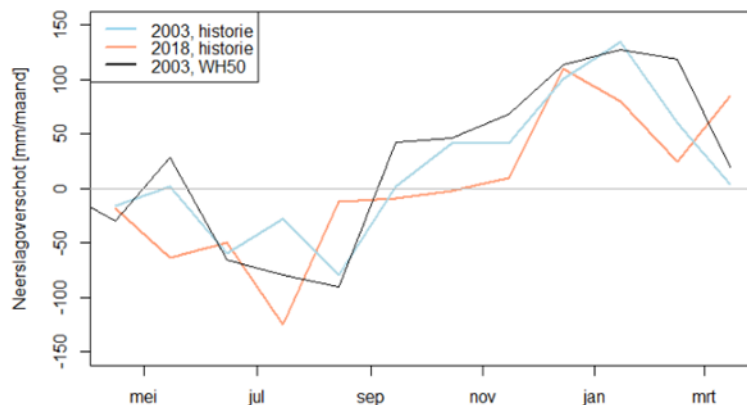
In deze paragraaf worden de veranderingen in weersomstandigheden volgens de KNMI-scenario's gepresenteerd en hoe deze volgens de uitgevoerde modelberekeningen doorwerken op het grondwater. De vorige twee paragrafen hebben duidelijk gemaakt dat er nog onzekerheid is over het gedrag van de modellen wanneer het gaat om toekomstvoorspellingen. In deze paragraaf worden alleen de resultaten gepresenteerd voor die locaties waar de voorspelde reeksen visueel geen duidelijke problemen vertoonden. Ook worden telkens de veranderingen besproken ten opzichte van de historische simulaties, zodat de getoonde effecten in principe alleen door het verschil in invoer (klimaatverandering) worden veroorzaakt.

3.3.1 Veranderingen in weersomstandigheden

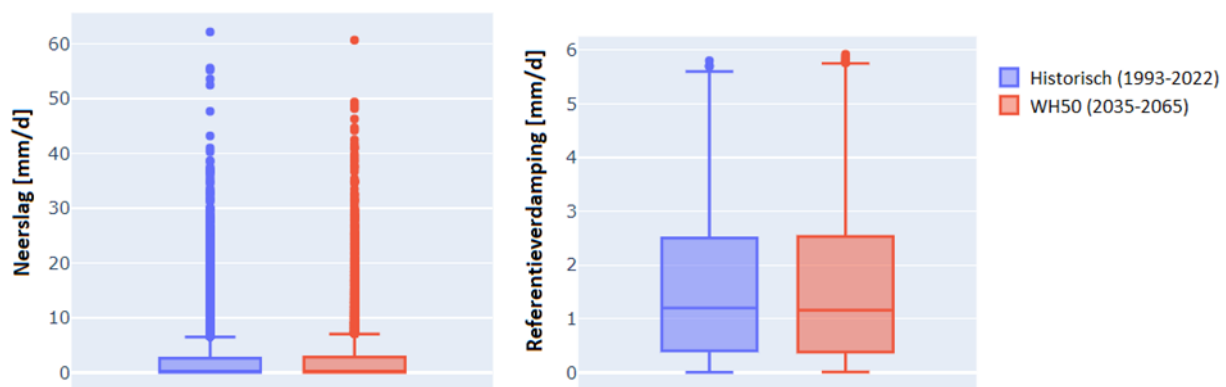
In Figuur 3.11 t/m Figuur 3.13 worden de veranderingen in de sturende weersomstandigheden volgens het KNMI WH50-scenario weergegeven. Figuur 3.11 en Figuur 3.12 geven het maandelijks neerslagoverschot voor de gehele historische en toekomstige periode; en specifiek voor droge jaren. De KNMI'14-scenario's zijn gebaseerd op de periode 1981-2011. Het droge jaar 2003 is in het WH50-scenario wel droger dan de originele 2003, maar niet zo droog als het jaar 2018, dat bij het maken van de KNMI'14-scenario's nog onbekend was. Er lijken in WH50 met name meer natte periodes voor te komen. Wanneer puur naar de spreiding in dagwaarden van de meteorologische variabelen wordt gekeken (Figuur 3.13) dan zijn de verschillen tussen de huidige situatie en de klimaatscenario's zeer klein. Dit is van belang voor de veranderingen die met de modelleringsmethode kunnen worden gesimuleerd, aangezien Model 1 voor de Generieke Trend puur met dagwaarden van de invoervariabelen werkt (geen historie).



Figuur 3.11: Maandelijks neerslagoverschot (neerslag – referentieverdamping) voor de huidige situatie (links) en het WH50-scenario (rechts) voor weerstation Hoogeveen. Rechts zijn ook het minimum en maximum uit de referentieperiode (2003-2020) weergegeven.



Figuur 3.12: Vergelijking van droge jaren in de huidige situatie en het WH50-scenario voor weerstation Hoogeveen. Maandelijkse neerslagoverschot in de observaties voor het jaar 2003, het jaar 2018, en het getransformeerde jaar 2003 in het WH50-scenario.



Figuur 3.13: Spreiding van dagelijkse neerslag (links) en verdamping (rechts) in de historische meetdata en het WH50-scenario voor weerstation Hoogeveen.

3.3.2 Veranderingen in grondwaterregime voor de KNMI-scenario's

De ontwikkelde modellen zijn na training en validatie toegepast om voorspellingen te maken van de grondwaterdynamiek in toekomstige klimaatscenario's. De modellen voorspellen in het algemeen weinig verandering in de gemiddelde grondwaterstand richting de toekomst. Voor de laagste grondwaterstanden worden dalingen voorspeld, voor de hogere grondwaterstanden juist lichte stijgingen. De verschillen in de voorspellingen tussen de scenario's zijn zeer klein. Dit wijkt af van verwachtingen uit berekeningen met het Landelijk Hydrologisch model, een fysisch-ruimtelijk model (Klimaat-effectatlas, 2022 n.d.), waarin wel duidelijke verschillen tussen mildere en sterke klimaatscenario's werden gevonden. De kleine verschillen geproduceerd door de modelopzet komen overeen met observaties gemaakt in par. 3.2.2, en worden verder besproken in par. 4.1.2. Deze geobserveerde beperkingen van de modelleringsmethode betekenen dat de gesimuleerde veranderingen richting de toekomst met voorzichtigheid bekeken moeten worden.

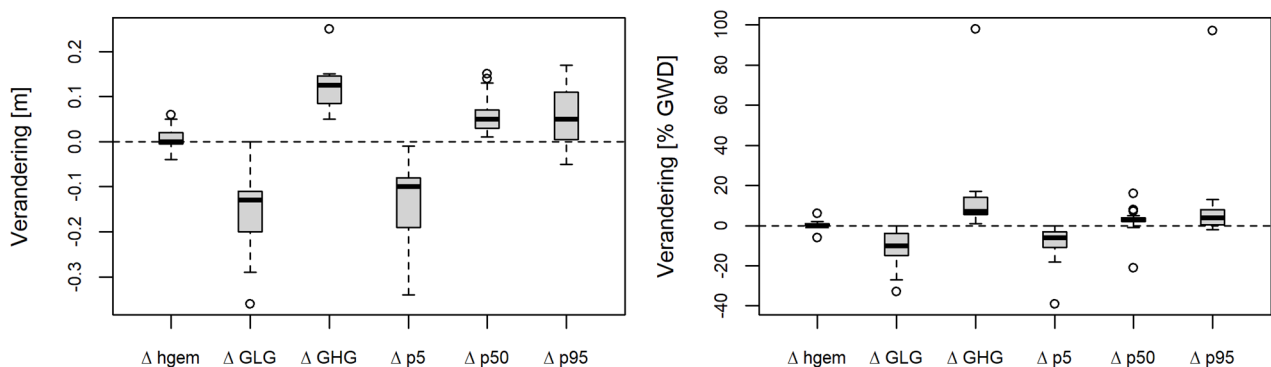
3.3.3 Veranderingen in grondwaterregime in scenario WH2050

De voorspelde veranderingen in grondwaterregime voor het WH50-scenario, met relatief sterke klimaatverandering, zijn in meer detail gegeven in Figuur 3.14 t/m Figuur 3.16. De gemiddelde grondwaterstanden blijven vrijwel gelijk, met weinig variatie tussen de locaties. Voor de lage grondwaterstanden (GLG en 5% percentiel h5) wordt door de modellen voor alle locaties een daling voorspeld, met dalingen tot meerdere decimeters. Er zijn hierin geen duidelijke ruimtelijke patronen te zien. Voor de hogere grondwaterstanden (GHG, 95% percentiel h95) worden voornamelijk stijgingen voorspeld, met opnieuw een orde grootte van enkele cm tot meerdere decimeters. Ook voor de hoge grondwaterstanden zijn geen ruimtelijke patronen te zien in de voorspelde veranderingen.

Deze resultaten komen in grote lijnen overeen met eerdere berekeningen met het Landelijk Hydrologisch model (Klimaat-effectatlas, 2022). Deze berekeningen gaven voor het grootste deel van de provincie Drenthe dalingen in GLG van 10-25 cm, met op de Hondsrug sterkere dalingen tot een meter. Bij mildere klimaatveranderingen lagen de dalingen voor het grootste deel van de provincie tussen 0 en 25 cm. Voor de GHG gaven deze modellen bij sterke klimaatverandering een stijging van 0 tot 25 cm voor het grootste deel van de provincie, met op de Hondsrug stijgingen tot 1 meter. De voorspellingen met de Deep Learning-methode geven dus soortgelijke trends en een soortgelijke ordegrrootte, maar geven iets zwakkere veranderingen. Ook komen de verschillen tussen scenario's en de patronen in het landschap nauwelijks terug in de voorspellingen, terwijl deze in de voorspellingen met het LHM wel duidelijk zijn.

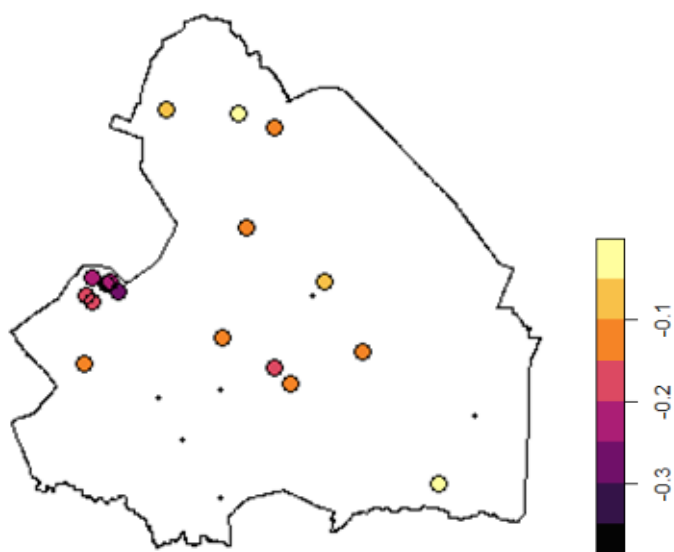
De veranderingen in de kenmerken van periodes met lage grondwaterstijghoogten zijn weergegeven in Bijlage II.VI. Zoals besproken in paragraaf 3.2.3 zijn de modellen op dit moment nog minder geschikt om de timing van grondwaterstandfluctuaties over het jaar goed te simuleren. Daarom zijn de indicatoren in relatie tot temporele kenmerken van periodes met lage standen waarschijnlijk niet voldoende betrouwbaar en worden daarom niet verder besproken.

Voorspelde verandering WH50 t.o.v. historische periode



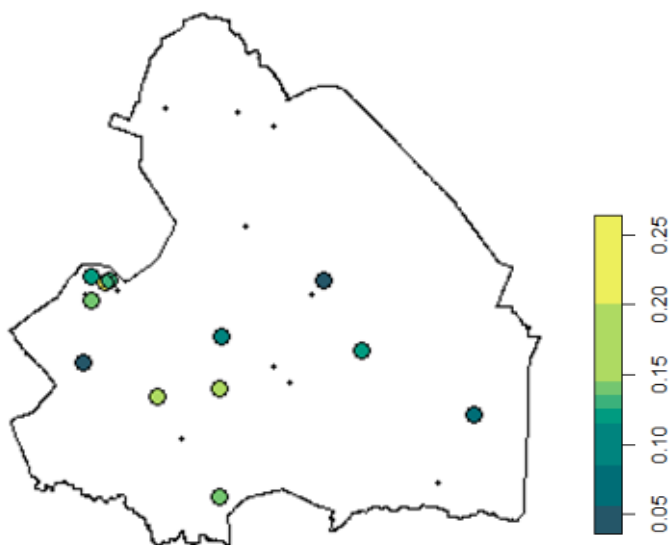
*Figuur 3.14: Voorspelde veranderingen in regime-indicatoren voor het WH50-scenario ten opzichte van de huidige situatie (historische simulaties). Gegeven zijn de verandering in de indicatoren, uitgedrukt in grondwaterdiepte (GWD) in het toekomstscenario [2047-2065] ten opzichte van historische modelsimulaties (2003-2020) in meters en in percentages van de historische grondwaterdiepte in m-MV. **Hgem**: gemiddelde stijghoogte; **GLG** en **GHG**: gemiddelde laagste en hoogste grondwaterstand; **h5**, **h50**, **h95**: 5%, 50% en 95% percentiel van de stijghoogtereeks. Voor de verschillende indicatoren zijn alleen die locaties weergegeven waarvoor de fout in de historische indicator minder was dan 20%; en waarvoor de voorspelde tijdreeks niet duidelijk tegen een minimum of maximum leek te blokkeren. Merk op dat alle voorspelde tijdreeksen gelimiteerd zijn tot de historische observaties en ze vrijwel allemaal een versterkte regelmaat vertonen.*

Voorspelde verandering in GLG [m], WH50



Figuur 3.15: Voorspelde verandering in de gemiddelde laagste grondwaterstand GLG in het WH50-scenario ten opzichte van de huidige situatie.

Voorspelde verandering in GHG [m], WH50

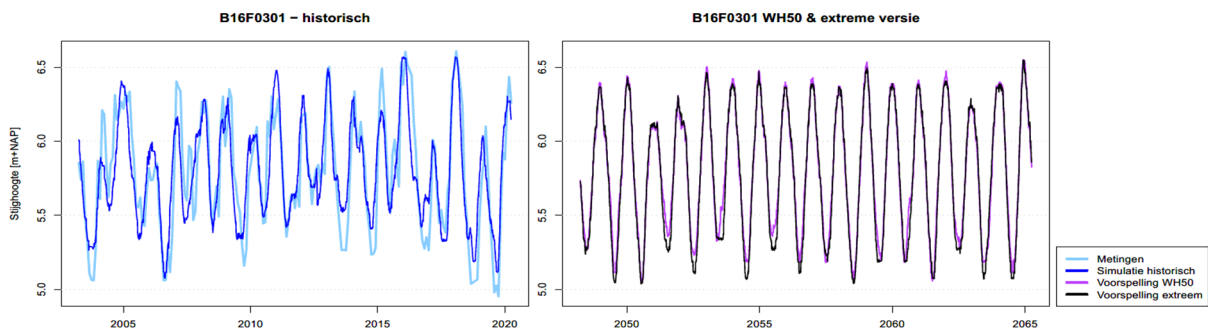


Figuur 3.16: Voorspelde verandering in de gemiddelde hoogste grondwaterstand GHG in het WH50-scenario ten opzichte van de huidige situatie.

3.3.4 Extreem scenario

Om een volledig beeld te krijgen van de potentiële toekomstige veranderingen in grondwaterstand is een extreem scenario toegepast voor één van de locaties, B16F0301 in het westen van het gebied. Deze locatie is gekozen omdat het model hier een goede betrouwbaarheid liet zien in de historische situatie. Het linker paneel in Figuur 3.17 geeft de historische meet- en simulatiereeksen. Het rechter paneel geeft de voorspelde dynamiek voor het WH50-scenario (paars) en het extreme scenario (zwart), waarin ten opzichte van WH50 de verdamping in de zomer met 25% is verhoogd, de neerslag met 25% verlaagd en de drinkwateronttrekking met 50% verhoogd in het gehele jaar. Tabel 3.8 geeft het bereik van de voorspelde waarden in vergelijking met de andere simulaties. De voorspelde veranderingen in regime-indicatoren zijn gegeven in Bijlage II.VII (vanwege de lage betrouwbaarheid zijn deze resultaten niet in de hoofdttekst opgenomen).

Onder het extreme scenario wordt een scherpe daling in de lagere grondwaterstanden voorspeld, die nog sterker uitvalt dan onder WH50. De hogere grondwaterstanden stijgen onder het extreme scenario nog steeds, maar iets minder dan onder WH50. De voorspellingen onder het extreme scenario verschillen relatief weinig van het WH50-scenario. Dit komt overeen met de eerdere observaties in par. 3.2.2 en 3.3.2. Net als de gewone voorspellingen voor de toekomstscenario's blijven de voorspelde grondwaterstanden voor het extreme scenario binnen het bereik van de historische observaties, maar gaan ze wel buiten het bereik van de historische simulaties (Tabel 3.8).



Figuur 3.17: Historische (links) en voorspelde dynamiek (rechts) voor de locatie waar een extreem toekomstscenario is toegepast.

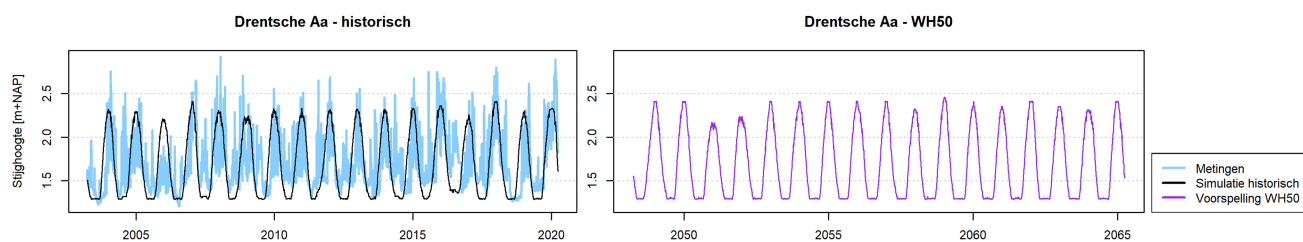
Tabel 3.8: Minimum-en maximum grondwaterstanden in de historische en toekomstige observaties en simulaties voor locatie B16F0301.

	Historisch observatie	Historisch simulatie	WH50 simulatie	Extreem scenario simulatie
Minimum [m+NAP]	4.95	5.08	5.01	5.01
Maximum [m+NAP]	6.61	6.57	6.55	6.55

3.4 Testtoepassing oppervlaktewater: de Drentsche Aa

3.4.1 Modevaluatie

Om een beeld te vormen van de toepasbaarheid van de ontwikkelde modelopzet voor oppervlaktewatervoorspellingen, is deze als test toegepast op de waterstanden van de Drentsche Aa. Daarbij is de modelopzet, ontwikkeld voor grondwaterstandsreeksen, niet aangepast. De waterstandsreeks voor de historische periode werd gesimuleerd met een RMSE van 0.31 m en een R^2 van 0.38. Figuur 3.18 (links) en Tabel 3.9 laten zien dat het model de dynamiek van de waterstandsreeks niet erg goed kan simuleren in vergelijking met de grondwaterreeksen. Met name de snelle fluctuaties in de metingen worden niet gereproduceerd. Voor toepassing op gebied van het voorspellen van waterbeschikbaarheid zijn echter de gemiddeldes en dynamiek van periodes met lage standen meer van belang. Net als voor de grondwatersimulaties wordt de langjarig gemiddelde waterstand goed gereproduceerd, maar worden de standen in de zomer en het voorjaar gemiddeld onderschat en in de herfst en winter overschat, deels door een lichte verschuiving in tijd ten opzichte van de gemeten fluctuatie. In de zomers produceert het model onnatuurlijke, regelmatige waterstanden die op een consistente ondergrens lijken te blokkeren. Als gevolg wijken de kenmerken van periodes met lage waterstanden (Tabel 3.9) sterk af van de gemeten waarden. Het lijkt erop dat de ontwikkelde modelopzet voor het simuleren van grondwaterstijghoogten niet direct toepasbaar is voor de oppervlaktewaterstanden van de Drentsche Aa en dat hiervoor aanpassing van de modelopzet en -parameters nodig is.



Figuur 3.18: Simulaties voor de Drentsche Aa in de historische periode (links) en de voorspellingen voor WH50 (rechts).

Tabel 3.9: Reproductie van regime-indicatoren voor de Drentsche Aa. De waarden van de indicatoren zijn gegeven voor de meetreeks en voor de modelsimulaties over 2003-2020.

Indicator	Historische observaties	Historische simulaties
Gemiddelde waterstand [m]	1.68	1.70
Gemiddelde winter [m]	1.93	2.22
Gemiddelde voorjaar [m]	1.65	1.53
Gemiddelde zomer [m]	1.56	1.31
Gemiddelde najaar [m]	1.59	1.77
5% percentiel h5 [m]	1.32	1.29
20% percentiel h20 [m]	1.44	1.30
50% percentiel h50 [m]	1.60	1.60
80% percentiel h80 [m]	1.89	2.16
Periodes met lage standen:		
Lengte [d]	12.6	71.9
Frequentie [n/j]	2.88	1.65
Tekort [m*d/j]	-1.91	-7.37

3.4.2 Voorspellingen toekomstscenario's

De voorspelde veranderingen in het waterstandsregime van de Drentse Aa voor de verschillende toekomstscenario's worden vanwege het beperkte functioneren van het model hier niet verder besproken. De resultaten zijn te vinden in bijlage II.VIII. Op dit moment wordt door KWR voor Waterbedrijf Groningen gewerkt aan fysische modellering van het Drentsche Aa-stroomgebied. Dit onderzoek zal meer begrip opleveren van het Drentsche Aa-systeem en de mogelijke toekomstige ontwikkelingen.

4 Discussie: potentie van machine learning-methodes voor voorspelling van watervoorraden en verdere ontwikkeling

4.1 Geschiktheid van de modelleringsmethode voor beschrijven van huidige grondwaterdynamiek

Dit onderzoek richtte zich op het verkennen van de toepassing van machine learning en deep learning voor het voorspellen van de waterbeschikbaarheid in de toekomst, met de nadruk op grondwaterstanden. Dit is een nieuw onderzoeksgebied met nog maar een klein aantal gepubliceerde studies, waarin simulatiemethodes nog volop worden ontwikkeld en getest. Deze studie richtte zich daarom op een eerste verkenning van hoe deze methodes zouden kunnen worden toegepast voor hydrologische, drinkwatergerelateerde vragen, en welke aandachtspunten bij de modelontwikkeling van belang zijn. Een perfecte modelopzet hebben we in dit stadium nog niet kunnen bereiken, en dit was ook niet de bedoeling van het onderzoek.

Voor deze studie is een modelleringsmethode ontwikkeld in twee stappen, waarbij de gemeten grondwaterstandsreeks wordt opgedeeld in een langzaam en een sneller signaal, en deze afzonderlijke componenten worden gemodelleerd met een regressiemodel (*ensemble decision trees*) en een neuriaal netwerkmodel (*NARX*). Het onderzoek heeft laten zien dat het inderdaad mogelijk is om voor hydrologische tijdreeksen de decompositie van reeksen te combineren met modellering op basis van machine learning en deep learning. De ontwikkelde aanpak bleek in de meeste geteste gevallen in staat de historische grondwaterstanden goed te reproduceren. Er is één standaard modelopzet toegepast en getraind voor alle locaties, die een brede set aan landschapsposities en grondwaterdieptes omvatten. Dat de grondwaterdynamiek in zulke verschillende condities meestal goed kon worden gereproduceerd, betekent dat machine learning-deep learning-modelleringsmethodes in principe breed toepasbaar kunnen zijn voor het modelleren van huidige grondwaterstanden in (hoger) Nederland.

4.1.1 Temporele dynamiek en modellering van droogteperiodes

Niet alle aspecten van de grondwaterdynamiek werden even goed gereproduceerd door de gebruikte modelopzet (par. 3.2.3). In het algemeen reproduceerden de modellen de verdeling van grondwaterstanden op een locatie goed, dus de gemiddelde, hoogste en laagste standen en de regimekenmerken (GxG's). De modellen functioneerden minder goed in het simuleren van de timing van grondwaterstandsfluctuaties. De gesimuleerde seizoensdynamiek vertoonde vaak een verschuiving ten opzichte van de variaties in de meetreeks, en ook de temporele dynamiek van uitzakking en herstel in de zomer werd vaak niet goed gereproduceerd.

Voor toepassingen die gericht zijn op het begrijpen van het bereik van grondwaterstanden en de langetermijntrends in bijvoorbeeld de laagste niveaus is de timing van grondwaterstandfluctuaties minder belangrijk. Wanneer echter een completer inzicht nodig is in de mogelijke effecten van droogte en watertekort, is het wel belangrijk om de seizoensdynamiek correct te simuleren (een bepaalde waterstand kan bijvoorbeeld duiden op droogteproblemen als deze in het voorjaar optreedt, maar niet in de zomer) en de dynamiek van verdroging en vernatting (bijvoorbeeld de duur van een periode met lage waterstanden) (Van Loon, 2015). Meer gedetailleerde kenmerken van perioden met lage waterstanden, zoals drempelwaarde-droogte-indicatoren (par. 2.3), verwante systeemindicatoren (Hashimoto et al., 1982) of seizoensgebonden droogte-indicatoren (Bloomfield en Marchant, 2013) konden in deze studie dan ook niet worden gekwantificeerd.

De beperking in het reproduceren van de tijdsdynamiek in de huidige situatie is waarschijnlijk te wijten aan verschillende oorzaken:

- De modellen zijn getraind op de gehele historische set van grondwaterstanden in plaats van specifieke periodes in het jaar. Om specifiek de grondwaterstanden in bepaalde seizoenen te modelleren, is een andere modeltraining nodig die specifiek traint op deze seizoenen. Dit is binnen de beperkte tijd van het huidige project niet uitgevoerd, maar zou een nuttige vervolgstap kunnen zijn.
- De gebruikte modelopzet kan waarschijnlijk niet goed genoeg omgaan met de vertraagde reactie van het grondwater op externe factoren. Het eerste model voor de generieke (langzame) trend is een decision tree-regressiemodel dat de grondwaterstand relateert aan de neerslag en verdamping van één dag terug. Dit betekent dat dit deelmodel het effect van opeenvolgende droge of natte dagen niet kan meenemen, waardoor het waarschijnlijk moeite heeft om het werkelijke effect van weersinvloeden te 'leren' uit de data. Het tweede model voor de periodieke (snelle) trend neemt voorgaande waarden van de sturende variabelen, inclusief de gemodelleerde grondwaterstand vanuit model 1, wel expliciet als invoer mee. Waarschijnlijk kan model 2 echter de beperking van model 1 niet compenseren, omdat de generieke trend meestal het grootste deel van de variatie in grondwaterstanden vertegenwoordigde. Het gebrek aan 'historisch besef' in model 1 zou ook de verschuivingen in de tijd voor de modelresultaten ten opzichte van de gemeten grondwaterstanden (paragraaf 3.2.3) kunnen verklaren. Hierbij moet worden benadrukt dat de nog beperkte omgang van het huidige model met tijdsfluctuaties en vertraging niet intrinsiek is aan machine learning op zich, maar waarschijnlijk ligt in de specifieke gebruikte modelopzet en -parameters. Dit punt kan worden verbeterd door in model 1 ook historische waarden van neerslag en verdamping mee te nemen. De gemeten historische grondwaterstand kan in principe ook als invoer worden meegenomen, maar dit is voor voorspellingen langer vooruit niet wenselijk omdat dit een accumulatie van onzekerheid oplevert over de tijd. Daarnaast kan in plaats van het decision tree-model een alternatieve modelleringsmethode worden gebruikt zoals 'Long Short-term Memory' (bijv. Li et al., 2021) dat mogelijk beter met het tijdsaspect omgaat, of impuls-responsmodellering.
- Voor sommige locaties kunnen andere invloeden van belang zijn naast de gebruikte weersinvloeden en (drinkwater)onttrekking. Voor die locaties waar het model niet goed presteert, is een nadere beschouwing nuttig om te zien of er mogelijk andere factoren in het model moeten worden opgenomen.
- De gebruikte periode voor training is voor sommige locaties mogelijk te kort (kortste gebruikte periodes zijn 10 jaar) voor het krijgen van een goed beeld van extremere situaties. Voor sommige locaties zijn er ook gaten in de gemeten grondwaterreeks (die door interpolatie worden omgezet in rechte lijnen of plateaus), die de temporele patronen en daarmee de modeltraining verstoren.
- De modellen zijn niet volledig geoptimaliseerd: één set standaard (hyper)parameters is gebruikt voor alle locaties en enkel een subset van de parameters is getraind. Uit de analyse in bijlage I.I blijkt dat fine-tuning van de modellen en toevoeging van een extra modellaag in het netwerkmodel de modelprestaties voor sommige locaties kan verbeteren, maar dit lijkt de problemen met de simulatie van temporele patronen niet op te lossen. Bovendien laat deze test zien dat de temporele patronen in de geproduceerde modelsimulaties zeer gevoelig zijn voor de modelopzet. Het zou nuttig zijn om het effect van de modelopzet en de tijdschaal op de modeluitkomsten verder te onderzoeken.

In paragraaf 4.3 worden enkele mogelijke aanpassingen in de modellen gegeven om de genoemde punten te verbeteren.

Opgemerkt moet worden dat de toegevoegde waarde van de ontwikkelde machine learning - deep learning-benadering mogelijk relatief klein is voor de toepassing in deze studie, modellering van grondwaterstandsvariatie als gevolg van regenval, verdamping en onttrekking. Dit is immers een situatie waar meer conventionele instrumenten zoals impuls-responsmodellering meestal nog goed mee overweg kunnen (von Asmuth et al., 2002; Bakker en Schaars, 2019). De huidige opzet met twee of drie inputvariabelen geeft de modellen relatief weinig variatie om mee te werken. De toegevoegde waarde van een machine learning- of deep learning-benadering is mogelijk groter in complexere situaties, waarin grondwaterstanden door meerdere factoren worden beïnvloed

(zoals waterbeheer, landgebruik, etc.) en de relaties in het systeem nog niet goed bekend zijn. Dit vereist dan wel dat er (goede) data beschikbaar is van deze factoren (zie paragraaf 4.3.3).

4.1.2 Rol van externe invloeden in de modelsimulaties

De modelopzet die hier is toegepast, gebruikt gegevens van neerslag, referentieverdamping en grondwateronttrekkingen voor drinkwater als invoer. Model 1 gebruikt neerslag en verdamping van één dag terug om een initiële voorspelling te maken van de generieke trend in de grondwaterstand; model 2 gebruikt neerslag, verdamping, onttrekking en de grondwaterstanden uit model 1, alle tot 7 dagen terug, om de uiteindelijke grondwaterstandsvoorspelling te maken.

Gebaseerd op kennis van het grondwatersysteem en eerdere modelstudies hadden we bij aanvang enkele verwachtingen met betrekking tot externe invloeden in de modellen:

- Neerslag en verdamping hebben waarschijnlijk een dominante invloed voor de meeste locaties;
- Klimaatverandering heeft een sterk effect op de grondwaterstanden, waarbij de effecten ook sterk verschillen tussen klimaatscenario's. De KNMI-scenario's omvatten een stijging in temperatuur, verdamping en een toename in warme, droge zomerperiodes, waarbij de veranderingen sterk verschillen tussen de verschillende scenario's. Kijkend naar de sterke grondwaterstands dalingen als gevolg van de droogteperiodes in 2018-2019, zou verwacht mogen worden dat zulke uitzakkingen in de toekomst vaker en nog sterker op zullen treden, met in de mildere scenario's weinig toename en in het WH50-scenario veel. Eerdere LHM-berekeningen voor de KNMI-scenario's laten voor Drenthe inderdaad duidelijke veranderingen zien in de GHG en GLG van enkele tientallen cm tot meer dan 1 m, waarbij er duidelijke verschillen zijn tussen de klimaatscenario's (Klimaat-effectatlas, 2022).
- De invloed van drinkwateronttrekkingen is niet overal aanwezig is en hangt samen met de afstand tot winningen.

Uit de uitgevoerde scenarioberekeningen (par.3.2.2, 3.3.2) blijkt dat veranderingen in de externe invloeden minder invloed hebben op de modeluitkomsten dan verwacht. De voorspellingen voor de verschillende klimaatscenario's lagen zeer dicht bij elkaar en ook het extreme scenario had een relatief mild effect op de grondwaterstanden. Dit kan gerelateerd zijn aan de hierboven besproken beperking in de temporele modellering in model 1. Dit model neemt enkel waarden van neerslag en verdamping van één dag terug mee, waardoor het effect van opeenvolgende droge of natte dagen niet kan worden herkend (zie par. 3.3.1). De drinkwaterwinningen werden in bijna alle modellen meegenomen, maar leken weinig effect te hebben op de simulaties; bovendien vertoonde de "importance value" die door de modellen aan de drinkwaterwinning werd toegekend, geen verband met de daadwerkelijke afstand tot winputten of de diepte. Hier kan meespelen dat het meenemen van grondwaterwinning als extra variabele puur door het toevoegen van een extra vrijheidsgraad de modelfit al verbeterde. Een open vraag is ook nog of de drinkwateronttrekking nog daadwerkelijk invloed kan hebben in het model, wanneer deze als constante aan het model wordt meegegeven, zoals is gedaan voor de toekomstscenario's en het extreme scenario. Dit aspect hebben we in dit stadium niet verder kunnen onderzoeken.

Deze resultaten geven aan dat het analyseren van het fysisch realisme van de modelopzet en -uitkomsten van cruciaal belang is bij het toepassen van machine learning en deep learning voor hydrologische voorspellingen. Een nadeel van de machine learning-opzet is dat de invloed van externe factoren (verdamping, neerslag, onttrekking) moeilijk uit de modellen te interpreteren is. Een mogelijke stap om dit type modellen beter uit te leggen is het visualiseren van hun architectuur en het illustreren van de verandering van waarden bij veranderende inputs, zoals onderzocht door Wang et al. (2020). Dit houdt verband met het grotere onderzoeksveld van *Explainable AI (XAI)*.

In dit onderzoek zijn alleen neerslag, verdamping en grondwateronttrekking voor drinkwater als sturende factoren meegenomen. Neerslag en verdamping zijn meestal de dominante invloeden op de grondwaterstand, en op basis van deze twee kunnen veel grondwatermeetpunten in Nederland redelijk tot goed worden gemodelleerd (Zaand Noordijk et al., 2019; Brakkee et al., 2022). Op sommige locaties zullen echter ook andere factoren van invloed

zijn die in dit onderzoek niet zijn meegenomen, zoals oppervlaktewaterstanden, regionale kwelinvloed of andere onttrekkingen dan die voor drinkwater, met name voor industrie en landbouw. Dit kan de modelprestaties beperken in laaggelegen gebieden waar oppervlaktewater belangrijk is, dichtbij grote industriële winningen en in droge zomers, als landbouwonttrekkingen een belangrijke invloed kunnen hebben op het grondwatersysteem (De Louw et al., 2022). Van deze factoren zijn meestal minder gegevens beschikbaar dan voor grondwaterstanden en drinkwateronttrekkingen. Het zou nuttig kunnen zijn om voor een kleiner aantal locaties waar gegevens beschikbaar zijn verder te testen wat het toevoegen van deze variabelen doet voor de modelprestaties.

4.1.3 Toepassing voor oppervlaktewater

In principe kan de hier ontwikkelde modelleringsmethode voor grondwaterstanden (al dan niet in combinatie met seriedecompositie) ook worden toegepast op andere hydrologische variabelen, zoals oppervlaktewaterstanden. Een moeilijkheid bij het toepassen van machine learning en deep learning voor oppervlaktewaterdynamiek is dat het verzamelen van de benodigde invoergegevens hiervoor lastiger kan zijn. Beschikbare meetreeksen van oppervlaktewaterniveaus zijn vaak korter dan beschikbare grondwaterreeksen in een regio, en metingen van oppervlaktewaterafvoer zijn vaak relatief schaars. Bovendien worden oppervlaktewaterstanden en -debieten vaak sterk beïnvloed door waterbeheer; de benodigde informatie van streefpeilen en sturingsregels kan moeilijker te verkrijgen zijn dan gegevens van het weer en (grond)waterwinning. Dit betekent ook dat het toepassen van de in deze studie ontwikkelde methode voor het voorspellen van grondwaterstanden in laag Nederland mogelijk lastig zal zijn, omdat hier het grondwater sterk beïnvloed wordt door oppervlaktewaterdynamiek en -beheer.

In deze studie is de modelleringsmethode als test toegepast op de Drentsche Aa, een riviersysteem waarvoor wel een lange meetreeks beschikbaar was van goede kwaliteit, en dat weinig wordt beïnvloed door direct waterbeheer. Uit de analyse die is uitgevoerd voor de Drentsche Aa (par. 3.4.1) is gebleken dat de ontwikkelde modelopzet niet direct vertaald kan worden naar een oppervlaktewatertoepassing, en dat hiervoor aanpassing van de algemene parameters nodig is. In 2022 wordt voor Waterbedrijf Groningen door KWR een procesgebaseerd hydrologisch model ontwikkeld voor de Drentsche Aa. Een vergelijking van dit model met machine learning-methodes zou interessante lessen op kunnen leveren over de sterke en zwakke punten van een machine learning-opzet.

4.2 Geschiktheid van de modelleringsmethode voor voorspellen van toekomstige waterdynamiek

4.2.1 Hydrologische voorspellingen dicht bij de historische situatie

Zowel fysische als datagedreven modellen zijn gebaseerd op historische observaties van het watersysteem en de sturende factoren. Bij het opzetten van een fysisch model zijn keuzes gemaakt over te in te bouwen processen, op basis van kennis van het huidige watersysteem; bovendien bevatten ze meestal meerdere parameters die zijn gekalibreerd op historische metingen. Een datagedreven model moet het gedrag van het watersysteem in verschillende situaties volledig leren uit de historische data. Dit betekent dat alle typen hydrologische modellen het meest betrouwbaar zijn voor combinaties van sturende factoren die dicht liggen bij de historische situaties waarop de modellen ontwikkeld zijn.

In hoeverre een hydrologisch model in staat is om voorspellingen te maken voor dit soort “normale” situaties, kan worden bekeken met validatie. Bij validatie wordt een deel van de data uit de training weggelaten en wordt bekeken hoe goed het model deze data kan ‘voorspellen’. Dit is een controle of het model niet over-gefit is op de trainingsdata. In deze studie hebben we voor model 1 (dat de langzame grondwatervariaties beschrijft) kruisvalidatie en ensemble-regressiemodellering toegepast. Dit zorgt ervoor dat alle gegevens zowel voor training als voor validatie kunnen worden gebruikt en dat het uiteindelijke voorspellingsmodel een combinatie is van verschillende modellen die elkaar compenseren. Dit zou over-fitting van de modellen moeten voorkomen. Model 2 (dat de snellere grondwaterfluctuaties beschrijft) is gevalideerd door steeds het eerste deel van de data buiten de training te houden en de prestaties van het model op deze gegevens te testen (par. 3.2.1). Deze validatie gaf

vergelijkbare prestaties te zien in de trainings- en validatieperiode, waarmee er dus ook geen directe tekenen zijn van over-fitting in het tweede model.

Bij het maken van voorspellingen door modellering vooruit in de tijd kan het daarnaast gebeuren dat fouten zich over de tijd gaan opstapelen, omdat gegevens over voorgaande grondwaterstanden niet langer beschikbaar zijn. In deze studie is een methode toegepast waarbij de voorspelde generieke trend van model 1 (uitsluitend en rechtstreeks gebaseerd op weergegevens) wordt gebruikt als "voorafgaande grondwaterstand"-invoer voor het neurale netwerk van model 2. Dit zou het probleem van accumulerende fouten moeten omzeilen.

Een basaler probleem bij het maken van voorspellingen speelt wanneer het model door zijn opzet beperkt is in de waarden die het kan produceren tot de geobserveerde waarden in de trainingsset. Een machine learning-classificatiemodel, dat met discrete categorieën werkt, zal in principe nooit uitkomsten buiten het historisch bereik produceren. Dit type limitaties zou kunnen spelen voor model 1, dat is gebaseerd op decision tree-regressie (Lange en Sippel, 2020). Door het gebruik van een ensemble-regressie-opzet in model 1 en de NARX-opzet voor model 2 zou echter worden verwacht dat zulke directe beperkingen niet meer zouden moeten optreden. De in deze studie ontwikkelde modelopzet leek echter toch bepaalde beperkingen te vertonen in het bereik van de gesimuleerde grondwaterstanden (par 3.2.4). Dit gold voor alle toekomstscenario's, ook het extreme scenario, en ook voor de verder geoptimaliseerde modellen (bijlage I.I). Dit gaat alleen om de absolute range in grondwaterstanden; het model was wel in staat om veranderingen in frequenties, en daarmee bijvoorbeeld verschuivingen in GHG en GLG, te produceren. Dit probleem zou mogelijk te verhelpen zijn door een alternatieve methode te gebruiken voor model 1. Het geobserveerde 'plateaugedrag' in de toekomstige voorspellingen, het blijven hangen op plateaus in het hoge of lage bereik, leek mogelijk verholpen te kunnen worden door de extra optimalisatie van de modellen.

Behalve deze gedeeltelijke beperkingen in bereik vertoonden de voorspelde reeksen vaak een sterke regelmatigheid, met beperkte verschillen tussen verschillende jaren; in een enkel geval was er juist een toegenomen onregelmatigheid. Dit gedrag leek niet (volledig) te kunnen worden verklaard door de beperkte lokale fine-tuning of het gebruik van een constante onttrekking voor de toekomst, en lijkt daarom door de modellen zelf te worden veroorzaakt. Het kan zijn dat het 'plateaugedrag' in de modellen bij heeft gedragen aan deze regelmatigheid; immers, als de simulaties ieder jaar vanaf dezelfde minimum- of maximumstand beginnen, zou dit een herhaling van grondwaterstandspatronen kunnen veroorzaken. De beperkingen van het model bij het omgaan met temporele patronen en vertraging in de grondwaterreeksen, zoals in de vorige paragraaf beschreven, zouden ook deels verantwoordelijk kunnen zijn voor deze problemen in de temporele patronen voor de toekomst. Om de tijdspatronen en het modelbereik te verbeteren, zouden aanpassingen in de modelopzet kunnen helpen, bijvoorbeeld een andere opzet van model 1 om beter om te gaan met tijdseffecten en externe factoren, zoals boven besproken. Daarvoor is echter nog een beter begrip van de oorzaken van het probleem nodig.

4.2.2 Hydrologische voorspellingen ver buiten de historische situatie

De onzekerheden in voorspellingen van het watersysteem worden veel groter wanneer situaties worden gemodelleerd die ver buiten het historisch bereik van gebeurtenissen liggen. Voor fysisch gebaseerde modellen kan het gebeuren dat de gekalibreerde parameters in de nieuwe, zeer veranderde situatie niet meer geldig zijn. Ook kan het zijn dat de aangenomen processen in het model het veranderde watersysteem niet meer goed beschrijven - bijvoorbeeld doordat de grondwaterstanden zo ver dalen dat de koppeling met oppervlaktewater verdwijnt of het transpiratiegedrag van de vegetatie zich aanpast (Lange en Sippel, 2020; Blöschl et al., 2019; Thirel et al., 2015; Bouaziz et al., 2022). Fysische modellen kunnen een deel van de processen omvatten die nodig zijn om extreme omstandigheden en een veranderend watersysteem te modelleren, maar zullen mogelijk niet alle relevante terugkoppelingen omvatten.

Datedreven modellen zijn waarschijnlijk echter nog gevoeliger voor onzekerheden in voorspellingen voor zeer veranderde situaties, omdat deze modellen afhankelijk zijn van de historische gegevens om te leren hoe het watersysteem zich onder extreme omstandigheden ontwikkelt, en geen kennis van hydrologische processen

bevatten. Machine learning-methodes zijn al wel regelmatig toegepast om toekomstvoorspellingen te maken voor het gedrag van watersystemen onder klimaatverandering, dus onder een veranderend regime (bijv. Ardabili et al., 2020; Zhu et al., 2019). Ook in dit onderzoek was de globale voorspelde trend onder klimaatverandering (daling GLG en stijging GHG) in principe plausibel en in overeenstemming met fysisch gebaseerde modelvoorspellingen (par. 3.3.3).

In het algemeen is de consensus dat voor het werkelijk kunnen voorspellen van omstandigheden ver buiten het historisch bereik modellen nodig zijn die zo veel mogelijk procesgebaseerd zijn (Blöschl et al., 2019). De discussie over de potentie van machine learningmodellen voor het voorspellen van nieuwe situaties en het bijdragen aan systeembegrip is echter nog in volle gang; er lopen nog veel ontwikkelingen op dit gebied in machine learning en hybride modellering (Nearing et al., 2021; Lange en Sippel, 2020).

Om het gedrag van de ontwikkelde modellen onder nieuwe, zeer veranderde omstandigheden beter te begrijpen, zouden aanvullende tests kunnen worden gedaan door bijvoorbeeld de modellen te trainen op een periode zonder extremen en deze te valideren op een bepaalde extreme gebeurtenis. In dit onderzoek is een eerste aanzet gedaan van een dergelijke analyse voor één locatie, door het model te testen op de droogte van 2018-2019 (bijlage I.II). Dit leverde voor deze locatie geen duidelijke problemen op met de voorspelling van deze droogteperiode, maar om het probleem volledig te begrijpen zouden uitgebreidere tests voor meer locaties moeten worden gedaan. Het begrip van de waarde van de modellen voor toekomstvoorspellingen zou ook nog kunnen worden verbeterd door een referentiescenario te modelleren. De KNMI-14-scenario's zijn gebaseerd op de jaren 1981-2010, terwijl de historische simulaties in deze studie betrekking hadden op de periode 2003-2020. Om voor dit tijdsverschil te corrigeren, zou idealiter ook een simulatie moeten worden gemaakt voor een referentiescenario over dezelfde periode 1981-2010.

Om de modellering van toekomstige situaties te verbeteren, kunnen de modellen regelmatig opnieuw worden getraind wanneer nieuwe gegevens beschikbaar komen of wanneer zich nieuwe gebeurtenissen hebben voorgedaan. Een andere mogelijkheid is om veranderingen in waterstanden te modelleren als functie van externe invloed in plaats van de grondwaterstand zelf. Dit zou de modellen in principe minder gebonden maken aan de historische patronen in grondwaterstanden. Een dergelijke aanpak kan echter wel grotere onzekerheden met zich meebrengen. Het modelleren van toestandsveranderingen in plaats van toestanden zelf vereist een ander type model dan in deze studie is gebruikt. Methodes hiervoor zijn in andere onderzoeksgebieden al wel ontwikkeld (Pande et al., 2010) maar zijn in de hydrologie nog niet veel verkend.

4.2.3 Onzekerheid in de ontwikkeling van sturende factoren

Voorspellingen van de grondwaterstanden in de verre toekomst zijn niet alleen onzeker door de modellen zelf, maar ook doordat grote onzekerheid bestaat in hoe de sturende factoren, in ons geval klimaat en grondwateronttrekkingen, zich zullen ontwikkelen. De KNMI'14-scenario's die in deze studie zijn gebruikt, geven voor elk klimaatscenario slechts één mogelijke uitkomst van de toekomstige weersomstandigheden. Om een completer beeld te krijgen is het mogelijk om een groot aantal replicaties te creëren van de KNMI'14 scenario's (Lall en Sharma, 1996) en het voorspellingsmodel te gebruiken om de spreiding van de resulterende voorspellingen in beeld te brengen. Ook kunnen de ontwikkelde modellen opnieuw worden getraind wanneer er nieuwe klimaatscenario's beschikbaar komen, zoals de KNMI'23-scenario's.

Met name de ontwikkeling van menselijke activiteiten (bijvoorbeeld het stopzetten van een waterwinning of ingrijpende veranderingen in het landgebruik) kan uiterst onzeker zijn. Bovendien kan de onzekerheid over plausibele toekomstige omstandigheden "diep" zijn, wat betekent dat de onzekerheid waarschijnlijk geen bekende statistische verdeling volgt. Zo kan bijvoorbeeld de bevolkingsgroei (en daarmee bijvoorbeeld de watervraag) nooit precies worden voorspeld voor de komende 30 jaar, of zelfs maar voor 5 jaar. Deze 'diepe onzekerheid', ook wel 'Knightiaanse onzekerheid' genoemd, is een van de typische knelpunten bij het doen van voorspellingen voor de verre toekomst (Vogel et al., 2015; Maier et al., 2016; Walker et al., 2013).

Een belangrijke onzekere menselijke factor in dit onderzoek is de toekomstige ontwikkeling van de drinkwaterwinning. In Nederland wordt in het algemeen een stijging van de drinkwatervraag verwacht als gevolg van bevolkingsgroei en frequentere droge periodes (Mens et al., 2019). De verschillende drinkwaterwinlocaties in het studiegebied verschillen echter sterk in hun winhoeveelheid, de verdeling over het jaar en de situatie in de omgeving. Er was voor deze studie geen informatie beschikbaar van hoe de winning en de verdeling ervan over het jaar zich voor de verschillende winlocaties zou kunnen ontwikkelen. De afgelopen zomers hebben laten zien dat de drinkwatervraag en winning in droge periodes sterk kunnen toenemen, maar dit effect varieert tussen locaties en droogteperiodes (van den Eertwegh et al., 2021). Daarom is in dit stadium als vereenvoudiging voor de toekomst een constante onttrekking aangehouden, die voor het extreme scenario is verhoogd. Het zou interessant zijn om mogelijke ontwikkelingen in drinkwateronttrekking en temporele patronen hierin verder te verkennen en in nieuwe voorspellingen mee te nemen.

Een onderzoeksgebied dat werkt aan het omgaan met diepe onzekerheid, is de tak van decision making-onderzoek die werkt aan 'exploratieve modellering' (Kwakkel en Pruyt, 2013; Kwakkel, 2017). Dit onderzoek richt zich op besluitvorming onder 'diepe onzekerheid', waarbij de sturende invloeden onzeker zijn, maar de werking van het systeem ook deels onbekend is. In deze aanpak wordt modellering gebruikt om een groot bereik aan plausibele toekomsten en kritieke punten voor besluitvorming te verkennen, in plaats van harde voorspellingen te maken. Onzekerheden in de (toekomstige) werking van het systeem worden expliciet meegenomen door hiervoor verschillende opties te verkennen. Het zou in toekomstig onderzoek nuttig kunnen zijn om exploratieve modellen te verkennen om scenario's van waterwinning of andere 'diep onzekere' variabelen te onderzoeken (Kwakkel en Pruyt, 2013).

4.2.4 Conclusie: gebruik van machine learning voor langetermijnvoorspelling van waterstanden

Uit de resultaten van de toekomstvoorspellingen en de hier gegeven discussie concluderen we dat we met dit onderzoek nog niet goed kunnen vaststellen of de ontwikkelde machine learning-modellen daadwerkelijk geschikt en betrouwbaar zijn voor gebruik voor langetermijnvoorspellingen, waarbij de sturende condities sterk veranderen (deels zal dit ook gelden voor veel fysisch-gebaseerde modellen). Er is nog beter inzicht nodig in het gedrag van de hier ontwikkelde en andere machine learning-methodes en de modellen moeten nog verder worden ontwikkeld. Meer algemeen geldt dat het voorspellen van het gedrag van watersystemen onder sterk veranderde, nog nooit vertoonde omstandigheden nog een belangrijke uitdaging vormt voor de praktijk en de wetenschap. De discussies en het onderzoek rond deze vraag zijn op dit moment nog volop gaande (Thirel et al., 2015; Blöschl et al., 2019; Nearing et al., 2021).

4.3 Aanbevelingen voor verdere ontwikkeling

4.3.1 Analyse van gedrag en betrouwbaarheid van machine learning-modellen

Deze studie heeft duidelijk gemaakt dat het voor het toepassen van een machine learning- en/of deep learning-methode voor hydrologische voorspellingen belangrijk is om het gedrag en de beperkingen van deze modellen goed te begrijpen. Op dit gebied liggen nog enkele vragen open voor de in deze studie ontwikkelde modelleringsmethode. Er zijn verschillende analyses die het gedrag en bruikbaarheid van de modellen beter inzichtelijk kunnen maken.

Inzicht in het gedrag van de modellen:

- verkenning van het belang van modelopzet en (hyper)parameters, door gevoeligheidsanalyses uit te voeren en effecten op modeluitkomsten te bekijken;
- betere analyse van de rol van externe factoren in de modellen;
- nader onderzoeken van locaties waar het model niet goed presteerde, om te begrijpen of mogelijk andere beïnvloedende factoren een rol spelen.

Inzicht in de bruikbaarheid van het model voor toekomstige voorspellingen:

- onderzoeken van de prestaties van de modellen bij het voorspellen van historische extreme gebeurtenissen. Dit kan worden gedaan door modellen te trainen op een periode met gemiddeld weer en de prestatie van het model te analyseren voor het voorspellen van een extreme gebeurtenis;
- simulatie van een referentiescenario dat overeenkomt met de tijdsperiode van de klimaatvoorspellingen, om een directer beeld te krijgen van het gemodelleerde effect van klimaatverandering;
- verder onderzoeken, door literatuuronderzoek of modelexperimenten, welke machine learning- en deep learning-methoden wel of niet geschikt zijn voor hydrologische modellering en voorspellingen en hoe dit vergelijkt met conventionele methodes.

4.3.2 Mogelijke verbeteringen van de modelleringsmethode

Deze studie heeft de toepassing onderzocht van een modelleringsmethode op basis van reeksdecompositie, regressiemodellering en neurale netwerkmodellering voor het maken van langetermijnvoorspellingen van watervoorraden, met name grondwater. Er is aangetoond dat de modelopzet over het algemeen goed presteert voor de huidige situatie, maar beperkingen vertoont ten aanzien van tijdspatronen, de rol van externe factoren en modellering buiten de historische bandbreedte. Mogelijkheden om de modellen te verbeteren zijn:

- Model 1 voor de generieke trend: het gebruikte regressiemodel gebruikt de neerslag en verdamping van de voorgaande tijdsstap om de waterstand van de huidige tijdsstap te voorspellen. We verwachten dat de prestatie van het model verder kan verbeteren door ook rekening te houden met de neerslag- en verdampingswaarden uit het verleden (die altijd beschikbaar zijn in de KNMI'14-scenario's). Als alternatief zouden andere technieken zoals impuls-respons-transferfunctiemodellen kunnen worden toegepast om de generieke trend te modelleren, zoals onderzocht in bijlage I.II. Ook het gecombineerde gebruik van beide methoden is mogelijk (Ljung et al., 2020).
- Model 2 voor de periodieke trend: we verwachten dat het model verder kan verbeteren door het toevoegen van 'langetermijngeheugen' (Long-Short-Term-Memory) aan de huidige NARX-architectuur, zoals bijvoorbeeld is gedaan door Massaoudi et al. (2021). Voor hydrologische voorspellingen op lange termijn bestaat een dergelijke studie nog niet, maar het is een potentieel nuttig onderwerp om te onderzoeken.
- Voorspellen van seizoensvariabiliteit: in dit project zijn de modellen nog niet getraind om voorspellingen te maken voor specifieke periodes in het jaar. Als specifiek informatie nodig is voor bijvoorbeeld de lente of de zomer, kunnen seizoens- of maandsubmodellen worden getraind. In dat geval zouden bijvoorbeeld de gegevens van alle lenteseizoenen van de afgelopen 20-30 jaar worden gebruikt en statistische verdeling hiervan worden gebruikt als voorspellingsfunctie in het regressiemodel.
- Verbeteren van voorspellingen buiten het historisch bereik: De in 4.3.1 genoemde analyses en de hierboven genoemde aanpassingen zouden de werking van het model voor toekomstige voorspellingen al kunnen verbeteren. Zoals eerder besproken zou het gebruik van concepten vanuit exploratieve modellering een richting kunnen zijn om externe en modelgerelateerde onzekerheden beter mee te nemen in voorspellingsstudies.

4.3.3 Verbetering van databeschikbaarheid voor hydrologische (machine learning-)modellen

De ontwikkeling en toepassing van machine learning-modellen voor voorspellingen van het watersysteem werden in dit onderzoek nog deels beperkt door de beschikbaarheid van sommige typen data. Voor het ontwikkelen van betere modellen en voorspellingen van watervoorraden op de lange termijn zijn enkele knelpunten aan te wijzen. Deze gelden in sterkste mate voor datagedreven methodes, maar in principe voor alle modeltypen.

- De grootste beperking ligt waarschijnlijk bij de data over andere invloeden op het watersysteem dan het weer en drinkwateronttrekkingen. Dit zijn name onttrekkingen door industrie, landbouw en particulieren en de oppervlaktewaterdynamiek, maar ook bijvoorbeeld bodemvocht en verdamping. Van deze invloeden zijn op dit moment ofwel überhaupt nauwelijks data beschikbaar, ofwel incompleet, met een lage frequentie of over een korte periode. Dit maakt het nog lastig deze factoren in datagedreven modellering mee te nemen.

- Data van grondwaterstanden zijn in Nederland met hoge frequentie en op veel locaties beschikbaar. De betrouwbaarheid van datagedreven modellering voor een verdere toekomst hangt echter af van de beschikbaarheid van data van een brede range aan gebeurtenissen in het verleden, omdat het gedrag in extreme situaties volledig moet worden geleerd uit extreme situaties in het verleden. Daarom zijn (zeer) langjarige datasets zeer waardevol voor het ontwikkelen van langetermijnvoorspellingen. Het zou bijvoorbeeld nuttig kunnen zijn om een dataset bij elkaar te brengen van enkele zeer langjarige grondwaterstandsreeksen, waarin meerdere extreme (droogte)situaties voorkomen. Deze kan dan gebruikt worden voor het verbeteren van modellen. Daarnaast zou een betere beschikbaarheid van real-time data, die door veel organisaties al worden verzameld, het mogelijk maken om modellen voortdurend te updaten, één van de sterke punten van machine learning.
- Data van drinkwateronttrekkingen zijn meestal goed beschikbaar, maar gaan vaak niet erg ver terug en moeten meestal handmatig door medewerkers aangeleverd worden.

Op dit moment komen uit veel verschillende richtingen (bijv. citizen science, remote sensing, slimme sensoren) nieuwe data beschikbaar met steeds hogere frequenties en resoluties. Machine learning en deep learning hebben dan ook een toenemend potentieel voor toepassing in hydrologie en waterbeheer. Om deze methodes echter op een goede manier in te kunnen zetten voor langjarige voorspellingen, zijn nog verbeteringen in de modelopzet en een beter begrip van het gedrag van machine learning-modellen nodig.

5 Conclusies

Het doel van dit onderzoek was om de toepassing van machine learning en deep learning te verkennen voor het voorspellen van veranderingen in watervorraden (in deze studie grondwaterstanden) richting de toekomst. Er is een nieuwe 2-staps modelopzet ontwikkeld, bestaand uit een combinatie van reeksdecompositie, een regressiemodel en een neurale netwerkmodel. Deze methode is toegepast voor 25 grondwaterlocaties en één oppervlaktewaterlocatie in het studiegebied, de provincie Drenthe. Voor deze locaties zijn vervolgens voorspellingen gemaakt van de verandering in grondwaterstanden in het klimaat rond 2050.

Het onderzoek heeft laten zien dat het mogelijk is om tijdreeksdecompositie te combineren met machine learning en deep learning voor het modelleren van hydrologische tijdreeksen zoals grondwaterstanden. De ontwikkelde modelleringsmethode was voor de meeste locaties succesvol om de huidige variatie in grondwaterstanden te beschrijven. Een verbeterde versie van de ontwikkelde modelopzet kan daarmee in principe breder worden toegepast voor het modelleren van de grondwaterstandsspreiding onder de huidige condities voor andere gebieden in Nederland. Voor toepassing voor oppervlaktewatervoorspellingen en in laag Nederland is de methode echter nog niet direct toepasbaar. De gebruikte modelleringsmethode functioneerde ook nog minder goed in het representeren van tijdspatronen en de invloed van sturende factoren en lijkt daarmee nog niet direct bruikbaar voor gedetailleerde analyse van droogteperiodes.

De betrouwbaarheid van de modelleringsmethode voor het maken van voorspellingen naar de (verre) toekomst kon nog niet volledig worden geverifieerd. De modellen produceren in principe aannemelijke trends voor de grondwaterdynamiek onder klimaatverandering, met lagere grondwaterstanden in de zomer en hogere standen in de winter onder de KNMI'14-scenario's. Uit de analyses kwamen echter enkele belangrijke beperkingen naar voren van de modellen bij het maken van voorspellingen, waaronder het correct meenemen van sturende factoren en het omgaan met situaties buiten het historisch bereik. Er is verdere analyse nodig van de (intrinsieke) voordelen en tekortkomingen van de modellen bij het maken van (langetermijn)voorspellingen, zowel voor de in het onderzoek ontwikkelde modelleringsmethode als voor machine learning in het algemeen. In het onderzoek is een overzicht gegeven van mogelijke vervolganalyses voor een beter begrip van het modelgedrag, als ook enkele aanpassingen die de modelprestaties zouden kunnen verbeteren.

Gelet op het steeds breder beschikbaar komen van verschillende soorten data van het milieu en het watersysteem, hebben op machine learning gebaseerde methodes een toenemende potentie voor hydrologische toepassingen, met name in complexe situaties. Er is echter nog werk nodig om de werking en betrouwbaarheid van deze methodes beter te begrijpen en ze nog beter geschikt te maken voor (grond)waterdynamiek, zeker wanneer toepassing voor strategische langetermijnvoorspellingen gewenst is.

6 Referenties

- Aa's, W. H. e.: Evaluatie droogte 2019, 2020.
- Alsumaiei, A. A.: A Nonlinear Autoregressive Modeling Approach for Forecasting Groundwater Level Fluctuation in Urban Aquifers, *Water*, 12, 10.3390/w12030820, 2020.
- Ardabili, S., Mosavi, A., Dehghani, M., en Várkonyi-Kóczy, A. R.: Deep Learning and Machine Learning in Hydrological Processes Climate Change and Earth Systems a Systematic Review, *Engineering for Sustainable Future*, Cham, 2020, 52-62,
- Atawneh, D. A., Cartwright, N., en Bertone, E.: Climate change and its impact on the projected values of groundwater recharge: A review, *Journal of Hydrology*, 601, 126602, <https://doi.org/10.1016/j.jhydrol.2021.126602>, 2021.
- Bakker, M., en Schaars, F.: Solving Groundwater Flow Problems with Time Series Analysis: You May Not Even Need Another Model, *Groundwater*, 57, 826-833, <https://doi.org/10.1111/gwat.12927>, 2019.
- Bengio, Y., en Grandvalet, Y.: No Unbiased Estimator of the Variance of K-Fold Cross-Validation, *The Journal of Machine Learning Research*, 5, 1089-1105, 2004.
- Bloomfield, J. P., en Marchant, B. P.: Analysis of groundwater drought building on the standardised precipitation index approach, *Hydrol. Earth Syst. Sci.*, 17, 4769-4787, 10.5194/hess-17-4769-2013, 2013.
- Blöschl, G., Bierkens, M. F. P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., . . . Zhang, Y.: Twenty-three unsolved problems in hydrology (UPH) – a community perspective, *Hydrological Sciences Journal*, 64, 1141-1158, 10.1080/02626667.2019.1620507, 2019.
- Bógalo, J., Poncela, P., en Senra, E.: Circulant singular spectrum analysis: A new automated procedure for signal extraction, *Signal Processing*, 179, 107824, <https://doi.org/10.1016/j.sigpro.2020.107824>, 2021.
- Bógalo, J.: CiSSA: Circulant SSA under Matlab, <https://github.com/jbogalo/CiSSA/releases/tag/2.1.2>, access: 2022-05-03, 2022.
- Bouaziz, L. J. E., Aalbers, E. E., Weerts, A. H., Hegnauer, M., Buiteveld, H., Lammersen, R., . . . Hrachowitz, M.: Ecosystem adaptation to climate change: the sensitivity of hydrological predictions to time-dynamic model parameters, *Hydrol. Earth Syst. Sci.*, 26, 1295-1318, 10.5194/hess-26-1295-2022, 2022.
- Brakkee, E., van Huijgevoort, M. H. J., en Bartholomeus, R. P.: Improved understanding of regional groundwater drought development through time series modelling: the 2018-2019 drought in the Netherlands, *Hydrol. Earth Syst. Sci.*, 26, 551-569, 10.5194/hess-26-551-2022, 2022.
- De Lenne, R., en Worm, B.: Droogte 2018 & 2019: steppeachtige verschijnselen op de 'Hoge Zandgronden'. De gevolgen voor beheer en beleid bij Waterschap Vechtstromen, Stromingen, 2020.
- De Louw, P., Pouwels, J., Witte, F., en Van den Eertwegh, G.: Effecten van beregening uit grondwater op het watersysteem tijdens de droogte van 2018, H2O-online, 2022.
- Guzman, S. M., Paz, J. O., Tagert, M. L. M., en Mercer, A. E.: Evaluation of Seasonally Classified Inputs for the Prediction of Daily Groundwater Levels: NARX Networks Vs Support Vector Machines, *Environmental Modeling & Assessment*, 24, 223-234, 10.1007/s10666-018-9639-x, 2019.
- Hashimoto, T., Loucks, D. P., en Stedinger, J. R.: Reliability, Resiliency, and Vulnerability Criteria For Water Resource System Performance Evaluation *Water resources research*, 18, 14-20, 1982.
- Hutchins, M. G., Abesser, C., Prudhomme, C., Elliott, J. A., Bloomfield, J. P., Mansour, M. M., en Hitt, O. E.: Combined impacts of future land-use and climate stressors on water resources and quality in groundwater and surface waterbodies of the upper Thames river basin, UK, *Science of The Total Environment*, 631-632, 962-986, <https://doi.org/10.1016/j.scitotenv.2018.03.052>, 2018.
- Klimaat-effectatlas: Klimaat-effectatlas, <https://www.klimaat-effectatlas.nl/nl/>, 2022.
- KNMI: KNMI'14: Climate Change scenarios for the 21st Century – A Netherlands perspective, De Bilt, The Netherlands, Scientific Report WR2014-01, 2014.
- KNMI: KNMI-14 Toelichting transformatie tijdreeksen, KNMI, De Bilt, 2015a.
- KNMI: Evaporation - transformed daily reference evaporation, KNMI14, Netherlands, Services, K. D. [Data set], urn:xkdc:ds:nl.knmi::knmi14_referentieverdamping/3.2/, 2015b
- KNMI: Precipitation - transformed daily precipitation (centr), KNMI14, Netherlands, Services, K. D. [Data set], urn:xkdc:ds:nl.knmi::knmi14_neerslag_cent/3.2/, 2015c

- Kwakkel, J. H., en Pruyt, E.: Exploratory Modeling and Analysis, an approach for model-based foresight under deep uncertainty, *Technological Forecasting and Social Change*, 80, 419-431, <https://doi.org/10.1016/j.techfore.2012.10.005>, 2013.
- Kwakkel, J. H.: The Exploratory Modeling Workbench: An open source toolkit for exploratory modeling, scenario discovery, and (multi-objective) robust decision making, *Environmental Modelling & Software*, 96, 239-250, <https://doi.org/10.1016/j.envsoft.2017.06.054>, 2017.
- Lall, U., en Sharma, A.: A Nearest Neighbor Bootstrap For Resampling Hydrologic Time Series, *Water Resources Research*, 32, 679-693, <https://doi.org/10.1029/95WR02966>, 1996.
- Lange, H., en Sippel, S.: Machine Learning Applications in Hydrology, in: *Forest-Water Interactions*, edited by: Levia, D. F., Carlyle-Moses, D. E., Iida, S. i., Michalzik, B., Nanko, K., en Tischer, A., Springer International Publishing, Cham, 233-257, 2020.
- Li, Z., Kang, L., Zhou, L., en Zhu, M.: Deep Learning Framework with Time Series Analysis Methods for Runoff Prediction, *Water*, 13, 10.3390/w13040575, 2021.
- Ljung, L., Andersson, C., Tiels, K., en Schön, T. B.: Deep Learning and System Identification**This research was financially supported by the Swedish Foundation for Strategic Research (SSF) via the project ASSEMBLE (contract number: RIT15-0012) and by the Swedish Research Council via the projects Learning flexible models for nonlinear dynamics (contract number: 2017-03807) and NewLEADS - New Directions in Learning Dynamical Systems (contract number: 621-2016-06079). Ljung's work was supported by Vinnova's Competence Center LinkSic, *IFAC-PapersOnLine*, 53, 1175-1181, <https://doi.org/10.1016/j.ifacol.2020.12.1329>, 2020.
- Maier, H. R., Guillaume, J. H. A., van Delden, H., Riddell, G. A., Haasnoot, M., en Kwakkel, J. H.: An uncertain future, deep uncertainty, scenarios, robustness and adaptation: How do they fit together?, *Environmental Modelling & Software*, 81, 154-164, <https://doi.org/10.1016/j.envsoft.2016.03.014>, 2016.
- Massaoudi, M., Chihi, I., Sidhom, L., Trabelsi, M., Refaat, S. S., Abu-Rub, H., en Oueslati, F. S.: An Effective Hybrid NARX-LSTM Model for Point and Interval PV Power Forecasting, *IEEE Access*, 9, 36571-36588, 10.1109/ACCESS.2021.3062776, 2021.
- Mens, M., Hunink, J., Delsman, J., Pouwels, J., en Schasfoort, F.: Geactualiseerde knelpuntenanalyse voor het Deltaprogramma Zoetwater: Fase II, 2019.
- Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., . . . Gupta, H. V.: What Role Does Hydrological Science Play in the Age of Machine Learning?, *Water Resources Research*, 57, e2020WR028091, <https://doi.org/10.1029/2020WR028091>, 2021.
- Pande, V. S., Beauchamp, K., en Bowman, G. R.: Everything you wanted to know about Markov State Models but were afraid to ask, *Methods*, 52, 99-105, <https://doi.org/10.1016/j.ymeth.2010.06.002>, 2010.
- Peters, E., Bier, G., van Lanen, H. A. J., en Torfs, P. J. J. F.: Propagation and spatial distribution of drought in a groundwater catchment, *Journal of Hydrology*, 321, 257-275, <https://doi.org/10.1016/j.jhydrol.2005.08.004>, 2006.
- Philip, S. Y., Kew, S. F., van der Wiel, K., Wanders, N., en Jan van Oldenborgh, G.: Regional differentiation in climate change induced drought trends in the Netherlands, *Environmental Research Letters*, 15, 094081, 10.1088/1748-9326/ab97ca, 2020.
- Prinsen, G., Sperna Weiland, F., en Ruijgh, E.: The Delta Model for Fresh Water Policy Analysis in the Netherlands, *Water Resources Management*, 29, 645-661, 10.1007/s11269-014-0880-z, 2015.
- Thirel, G., Andréassian, V., en Perrin, C.: On the need to test hydrological models under changing conditions, *Hydrological Sciences Journal*, 60, 1165-1173, 10.1080/02626667.2015.1050027, 2015.
- Tian, X., Vertommen, I., Béen, F., en Bäuerlein, P.: Deep Explorations: an explorative study for machine learning and deep learning applications in the water sector KWR, *Nieuwegein, BTO 2022.021*, 61, 2022.
- Tromp, G., Verheijen, S., Fokekma, F., Hendriks, B., Verhoeff, A., en Kaarsemaker, M.: (Beleids)inhoudelijke evaluatie droogteperiode bij WDO Delta 2018. Bijlage 2 bij agendapunt Evaluatie calamiteit droogte 2018., *Waterschap Drents Overijsselse Delta*, 2018.
- van Dorland, R., van der Wiel, K., Haarsma, R., Daniels, E., van der Schrier, G., en Aalbers, E.: KNMI Klimaatsignaal'21: hoe het klimaat in Nederland snel verandert,, *KNMI, De Bilt*, 72, 2021.
- Van Loon, A. F.: Hydrological drought explained, *WIRES Water*, 2, 359-392, <https://doi.org/10.1002/wat2.1085>, 2015.
- Van Loon, A. F., Stahl, K., Di Baldassarre, G., Clark, J., Rangecroft, S., Wanders, N., . . . Van Lanen, H. A. J.: Drought in a human-modified world: reframing drought definitions, understanding, and analysis approaches, *Hydrol. Earth Syst. Sci.*, 20, 3631-3650, 10.5194/hess-20-3631-2016, 2016.

- Vogel, R. M., Lall, U., Cai, X., Rajagopalan, B., Weiskel, P. K., Hooper, R. P., en Matalas, N. C.: Hydrology: The interdisciplinary science of water, *Water Resources Research*, 51, 4409-4430, <https://doi.org/10.1002/2015WR017049>, 2015.
- von Asmuth, J. R., Bierkens, M. F. P., en Maas, K.: Transfer function-noise modeling in continuous time using predefined impulse response functions, *Water Resources Research*, 38, 23-21-23-12, <https://doi.org/10.1029/2001WR001136>, 2002.
- Walker, W. E., Haasnoot, M., en Kwakkel, J. H.: Adapt or Perish: A Review of Planning Approaches for Adaptation under Deep Uncertainty, *Sustainability*, 5, 10.3390/su5030955, 2013.
- Wang, Z. J., Turko, R., Shaikh, O., Park, H., Das, N., Hohman, F., . . . Horng Chau, D.: CNN Explainer: Learning Convolutional Neural Networks with Interactive Visualization, <https://doi.org/10.48550/arXiv.2004.15004>, 2020.
- Wunsch, A., Liesch, T., en Broda, S.: Groundwater level forecasting with artificial neural networks: a comparison of long short-term memory (LSTM), convolutional neural networks (CNNs), and non-linear autoregressive networks with exogenous input (NARX), *Hydrol. Earth Syst. Sci.*, 25, 1671-1687, 10.5194/hess-25-1671-2021, 2021.
- Zaadnoordijk, W. J., Bus, S. A. R., Lourens, A., en Berendrecht, W. L.: Automated Time Series Modeling for Piezometers in the National Database of the Netherlands, *Groundwater*, 57, 834-843, <https://doi.org/10.1111/gwat.12819>, 2019.
- Zhu, R., Yang, L., Liu, T., Wen, X., Zhang, L., en Chang, Y.: Hydrological Responses to the Future Climate Change in a Data Scarce Region, Northwest China: Application of Machine Learning Models, *Water*, 11, 10.3390/w11081588, 2019.

7 Bijlagen

I Additional model performance analyses

I.1 Effect of model finetuning on predicted future patterns

During the training of the models (section 2.5) several of the model (hyper)parameters were kept fixed for all locations. To understand how this influences model performance, tests were performed to further finetune the models for several locations. ‘Finetuning’ consists of manually modifying model hyperparameters (e.g., the number of layers, optimization solver) and some parameters to better fit the location in question.

The example we give is for observation well B17C0392, which was found to perform well in the historical simulations, but produce a very regular outcome for the future, also seemingly being limited at the lower end (Figure I. 1). This could be partly because we use a set of model parameters with predefined values for training, which may not fit all locations, B17C0392 being one of the examples.

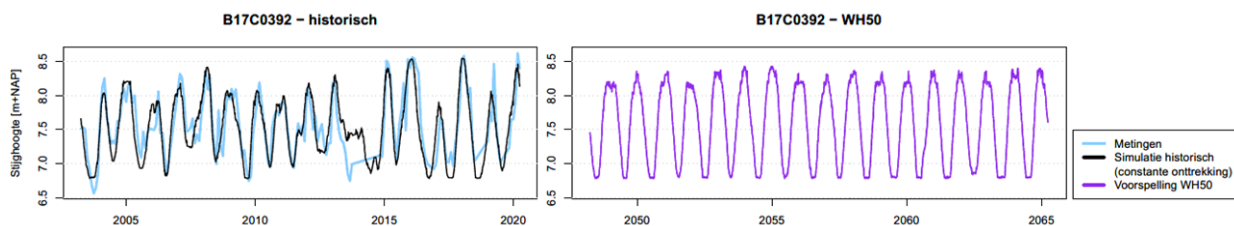


Figure I. 1: original simulation and prediction for well B17C0392.

Figure I. 2 gives the model results after fine-tuning the model parameters (so not just using the default settings), without changing the structure of the model. One of the changes made was the addition of an extra model layer in the NARX model structure of model 2. For the historical period, model errors do not improve (from 0.235 to 0.251), but explained variance does (R^2 increased from 0.75 to 0.82). The predictions for the future show more short-term variability and no longer a lower ‘plateau’ in summers. However, the overall pattern over the year is still very similar between years.

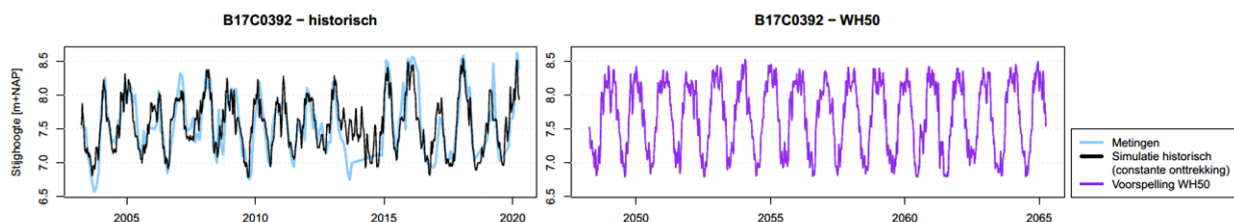


Figure I. 2: Model validation between predictions and observations. Note that the predicted water level is 5-fold cross validated.

Figure I. 3 shows two more examples of the effect of model finetuning on the historical simulations and future predictions. As for location B17C0392, the finetuned simulations show more short-term variations. However, the decreased (a) or increased (c) variability between years in the future predictions compared to the historical period does not appear to be solved, and in the latter case (d) is rather worsened. The future predicted groundwater levels with optimized parameters still stay within the range of historical observations, although, as for the non-optimised version, the future simulations do go beyond the historical simulations, even if by small amounts of a few cm at most.

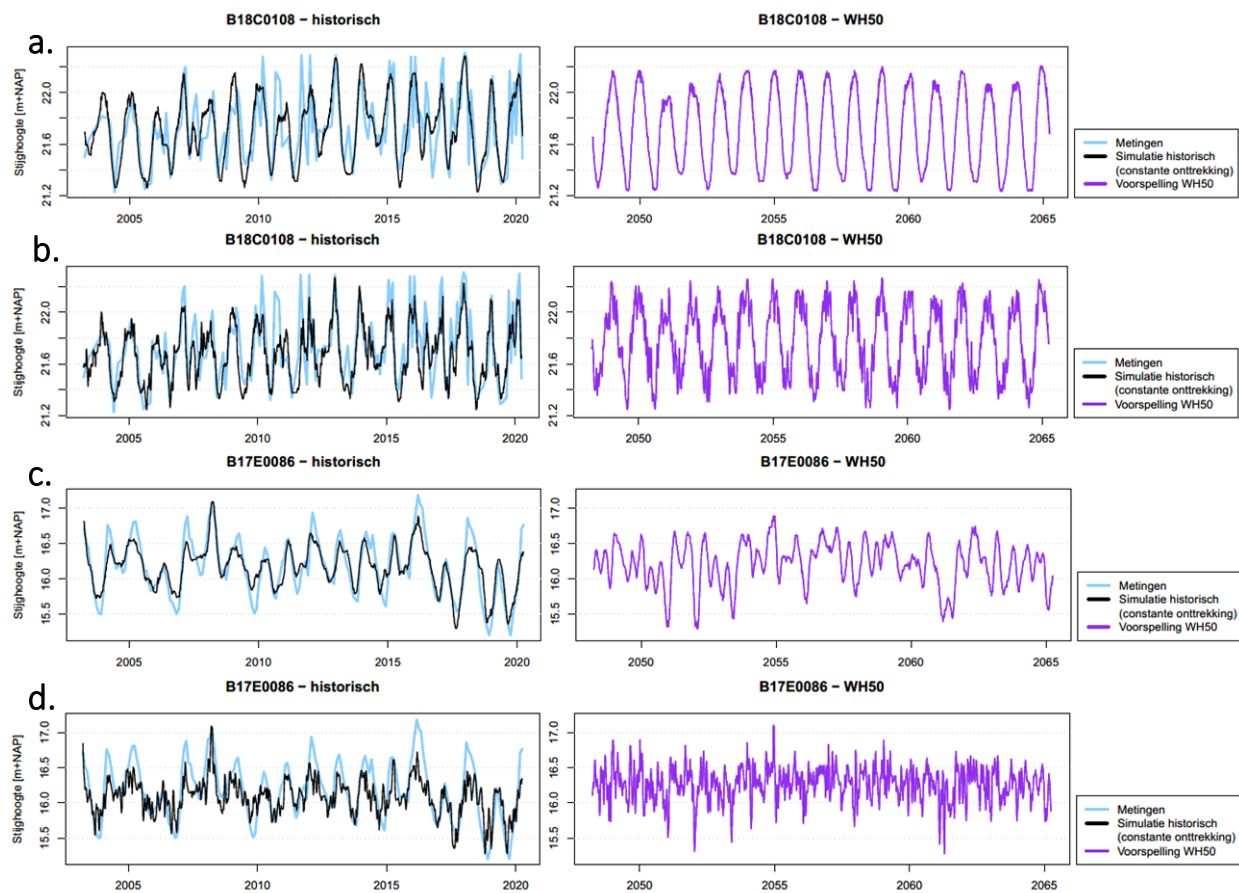


Figure 1. 3: 3 examples of the effect of model finetuning. For these locations, the original model training (a, c) produced unexpected patterns in the future predictions, with series becoming either more regular (a) or more irregular than the historical simulations. Panel b and d give the results after finetuning.

I.II Shifting training set and validation set for the General Trend model

The choice of training and validation sets may affect prediction performance, especially if the validation set contains observations beyond the range of the training set. Figure I. 4 shows the simulations for the General Trend of location B17C0392 with the ensemble regression model (model 1), first using the last 30% as validation (upper panel, yellow line) versus using the first 30% (second panel). Blue lines show the observed general trend. Table I. 1 shows the corresponding RMSE values.

It shows that the used general trend model performs generally well for both cases, but has signs of some overfitting, as the error increases strongly in the validation period. Training on the last 70% and validation on the first 30% gives a slightly poorer result in the training period, but slightly better in the validation period, than training on the last 70%. From the graph it can be seen that the General Trend of the groundwater level is actually lower in 2003 than in 2018-2020, and so excluding the last years from the training set may not affect the prediction of dry periods.

This 30-70 validation is just a test for better understanding the GT model. In the general model training, the regression model was validated using cross-validation (in our study, 5-fold). The latter is obtained by randomly partitioning the entire dataset into 5 equal-sized subsets and computing the mean value of the model performance indicator for all subsets. This typical model validation technique provides a rough estimate of the error when applying the model to new data [1].

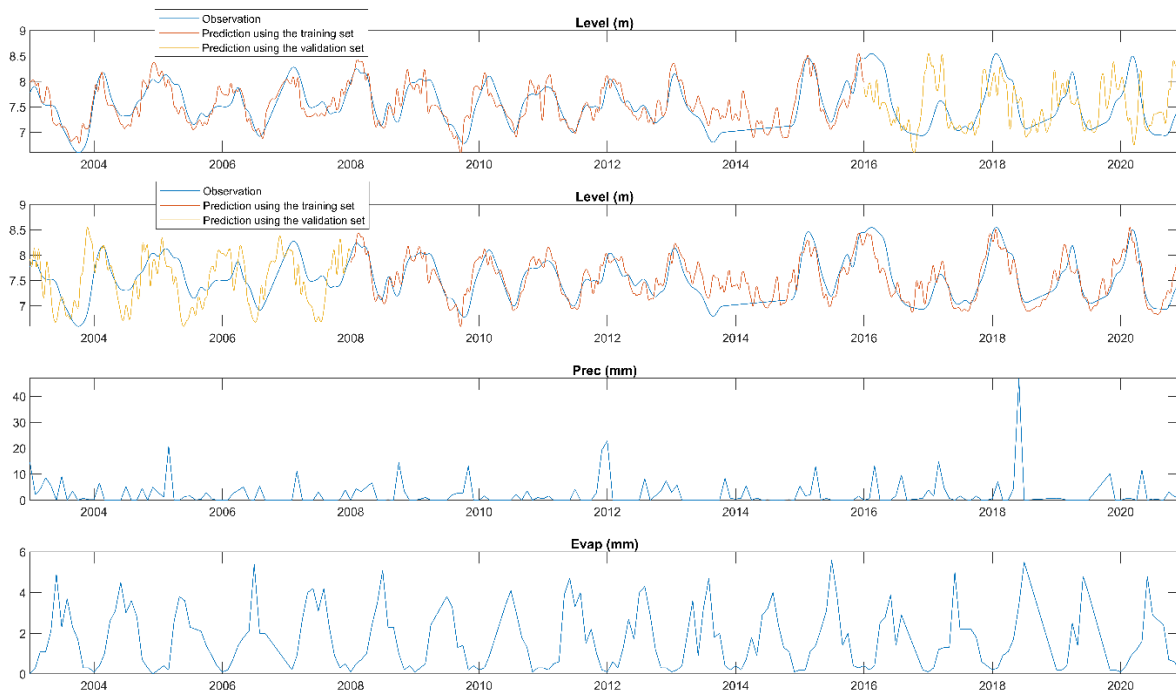


Figure I. 4: Comparison of two training-validation separations for well B170392.

Table I. 1: RMSE for training and validation periods for two separations of training and validation for the General Trend model of location B17C0392.

	Training last 70%, validation first 30%	Training first 70%, validation last 30%
RMSE training period	0.25	0.23
RMSE validation period	0.48	0.53

I.III Comparison between the regression model and the model used by Menyanthes for the General Trend

In this study, we applied an ensemble regression model to simulate the General Trend component of the groundwater level driven by precipitation and evaporation. Here we compare this method to the impulse-response transfer function method (von Asmuth et al., 2002) as applied in the well-known Menyanthes and PASTAS tools. The simulation is again based on the observation well B17C0392. Both models used 2003-2015 for training and 2016-2020 for testing. In both cases the groundwater level trained to is the General Trend component resulting from the decomposition, which in this case contributes 95% of the total variance.

Results from Menyanthes

In the training period, Menyanthes reproduces the General Trend with an RMSE of 0.25 m and explained variance R^2 of 60.5%. In general, the model performs well for retrieving the trend of the time series from the historical data. However, the amplitude of the model prediction is smaller than that of the observations, with especially a few low water level events being underestimated by the model.

Figure I. 5 shows the prediction that the model makes for 2016-2020 (dark blue), compared with the observation (light blue). The RMSE of the predictions is close to that of the training period (= 0.25 m). However, the prediction underestimates the high levels and overestimates the low levels, with the largest deviation of 0.6 m. Generally, the observed levels are higher than the predicted levels. There might be other factors than meteorological variables controlling the groundwater level variations, such as seepage.

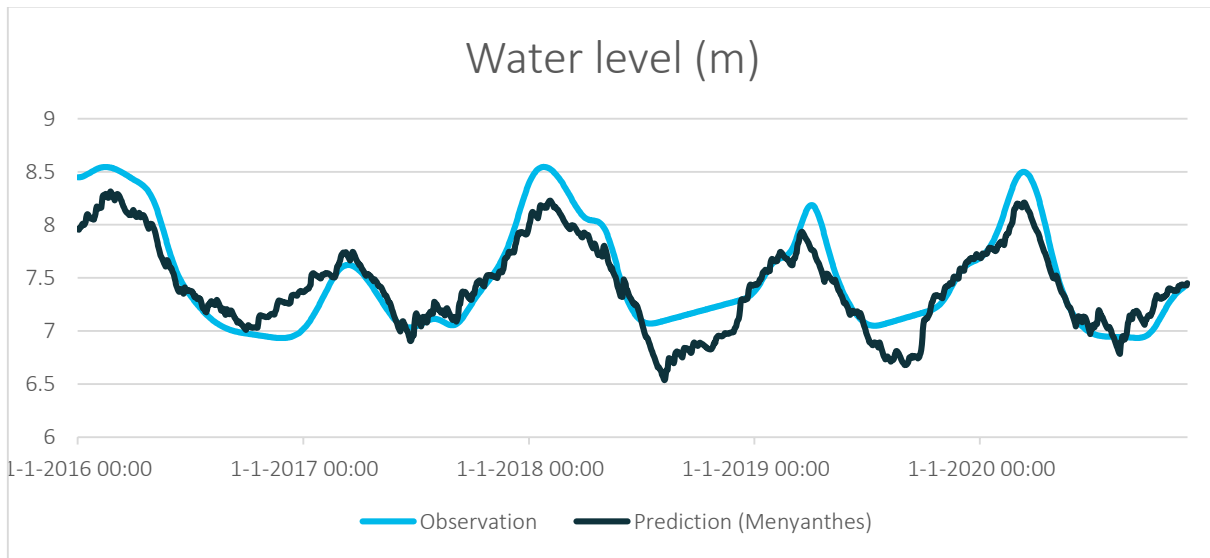


Figure I. 5: Prediction for B17C0392 using the model built in Menyanthes.

Results from the ensemble learning regression model

In this study, ensemble learning was adopted to build a regression model to predict the general trend of the groundwater levels. To compare the results with Menyanthes, the same simulation periods were set. The RMSE of training and validation sets are 0.23 and 0.5, respectively (Figure I. 6). For prediction, the ensemble learning model is therefore outperformed by the transfer function method used by Menyanthes for this specific location.

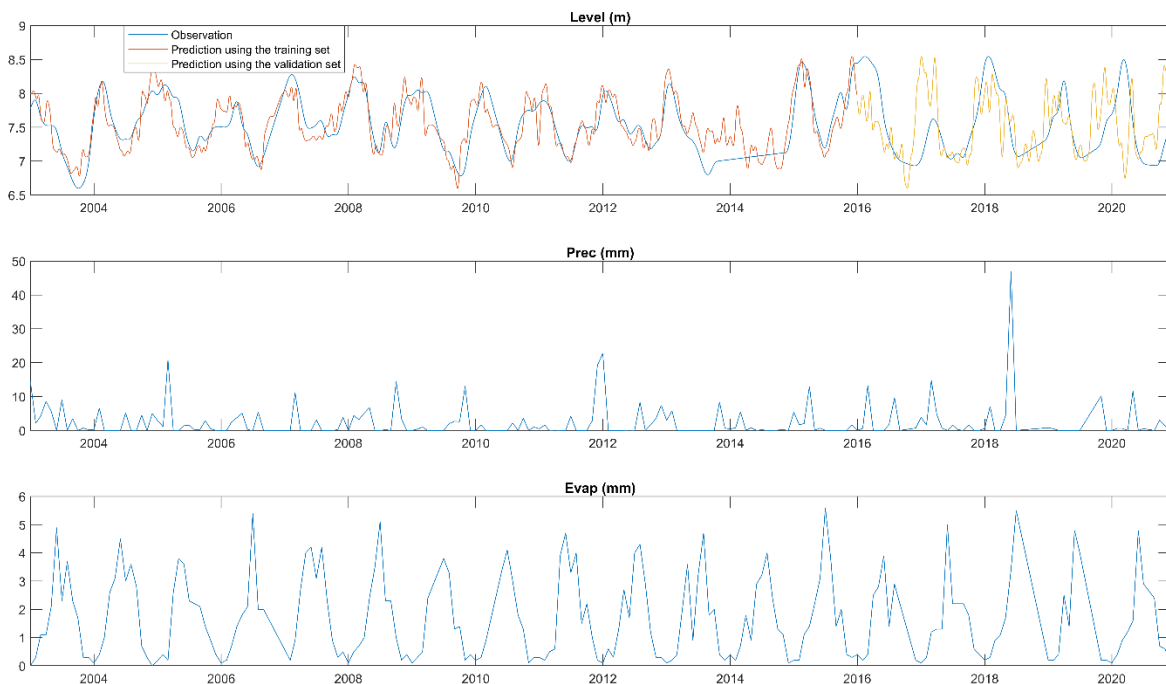


Figure I. 6: Prediction for well B17C0392 made by ensemble learning. The first panel shows the model simulations in the training period (red) and in the validation period (yellow).

Concluding remarks

We presented and discussed the general trend model for one test location using two methods. Menyanthes outperforms the regression model in catching the trend of the time series for this specific location; however, this does not need to be the case for other locations. The impulse-response method used by Menyanthes is suited specifically for reproducing temporal patterns and delay effects; however, the limitations of the currently applied

decision tree method for dealing with temporal patterns is not inherent to machine learning, and new ML-DL methods are being developed to suit temporal patterns. The ensemble learning model also has room to improve, for instance, more historical meteorological data can be used to simulate the water level (e.g., the past 30 or 60 days of precipitation and evaporation can be used to investigate the accumulated effect). On the other hand, if a particular location shows significant re-occurring seasonal or monthly patterns, we can also train seasonal or monthly sub-models by using the data, for instance, of all the spring seasons of the past 20-30 years, and use the density function of the probabilistic distribution as the prediction function in the regression model. Another line of development could be to use other ML/DL techniques or to combine impulse-response modelling together with the second model (NARX) to improve the total model performance. These can be future research directions of time series analysis for groundwater levels.

II Tabellen resultaten

II.I Regime-indicatoren historische situatie

Location	Mean level [m+N AP]	Mean level [m +surf]	p10 [m+surf]	p90 [m+surf]	GLG [m+surf]	GHG [m+surf]	LL duration [d]	LL frequency [1/y]	LL severity [m*d/y]
B12A0139	3.6	-1.3	-1.8	-0.9	-1.8	-0.9	24.1	1.5	-4.5
B12B0166	3.9	-1.4	-1.9	-0.9	-1.9	-0.9	41.3	0.9	-4.7
B12B0167	8.3	-1.0	-1.7	-0.4	-1.6	-0.3	51.8	0.7	-9.8
B12D0305	10.5	-1.1	-1.6	-0.8	-1.6	-0.7	56.5	0.6	-7.5
B12G0113	6.8	-2.8	-3.1	-2.5	-3.0	-2.6	103.5	0.4	-3.5
B16F0228	7.0	-2.1	-2.6	-1.5	-2.5	-1.5	73.7	0.5	-5.9
B16F0301	5.8	-1.2	-1.7	-0.6	-1.6	-0.6	62.1	0.6	-4.9
B16H0090	3.2	-2.1	-2.4	-1.8	-2.4	-1.7	62.1	0.6	-2.3
B17B0144	11.2	-1.7	-2.0	-1.4	-1.9	-1.3	77.7	0.5	-3.2
B17C0076	7.7	-1.7	-2.1	-1.3	-2.1	-1.3	68.9	0.5	-4.5
B17C0392	7.5	-0.9	-1.4	-0.3	-1.4	-0.3	69.0	0.5	-4.7
B17D0030	14.7	-2.0	-2.5	-1.6	-2.5	-1.5	25.9	1.4	-6.5
B17D0069	12.1	-2.5	-3.1	-1.9	-3.1	-1.9	69.0	0.5	-7.5
B17D0184	14.3	-0.6	-1.0	-0.2	-1.0	-0.2	51.8	0.7	-4.0
B17E0043	16.7	-3.3	-3.7	-2.8	-3.7	-2.7	77.5	0.5	-3.6
B17E0086	16.2	-3.4	-3.9	-2.8	-3.8	-2.8	88.7	0.4	-5.6
B17E0161	16.0	-0.3	-0.8	0.1	-0.8	0.1	27.0	1.4	-7.0
B18C0108	21.7	-0.6	-0.9	-0.2	-0.9	-0.2	47.8	0.8	-2.6
B22B0022	8.9	-2.6	-3.0	-2.2	-3.0	-2.2	69.0	0.5	-3.8
B22F0085	11.8	-1.5	-1.8	-1.2	-1.8	-1.1	17.1	2.1	-3.3
PB14_P021	6.8	-1.8	-2.6	-1.1	-2.5	-1.1	87.9	0.4	-6.9
PB20_P021	7.8	-0.9	-1.4	-0.3	-1.3	-0.3	75.2	0.5	-6.4
PB6_P021	7.3	-1.9	-2.5	-1.3	-2.4	-1.3	71.7	0.5	-4.4
PB1_P021	6.1	-0.8	-1.4	-0.2	-1.4	-0.2	80.0	0.5	-5.7
PB15_P021	4.9	-2.8	-3.2	-2.4	-3.0	-2.4	113.0	0.3	-0.8

PB4_P021	5.4	-0.4	-0.8	0.1	-0.8	0.0	80.2	0.5	-6.4
mean	9.8	-1.5	-2.0	-1.1	-2.0	-1.0	60.7	0.7	-5.2
min	3.2	-3.4	-3.9	-2.8	-3.8	-2.8	17.1	0.3	-9.8
max	21.7	-0.3	-0.8	0.1	-0.8	0.1	113.0	2.1	-0.8

II.II Variatie in onttrekkingen per locatie

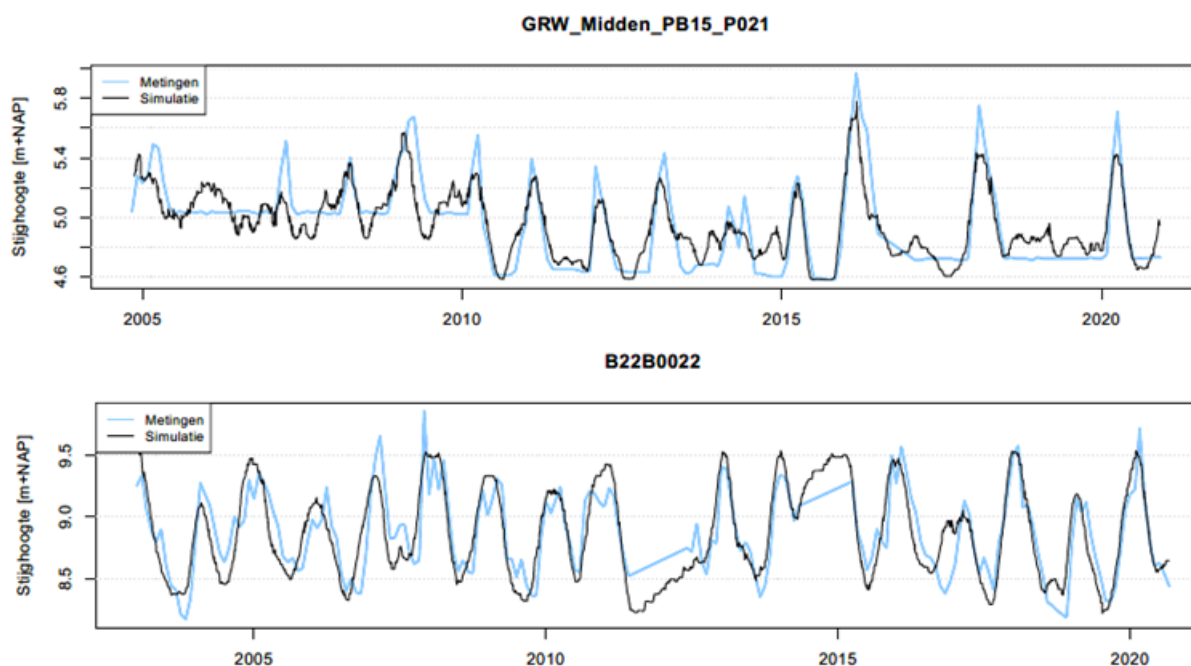
	Assen	Annen	Gasselte	Beilen	Ruinerwold	Zuidwold	Leggeloo	Hoogeveen	Valtherbos	Noordbargeres	Kruidhears	Dalen
Mean abstraction [1000 m ³ /mnd]	229	418	173	208	161	68	51	346	423	350	129	92
Standard deviation [1000 m ³ /mnd]	43	176	27	48	22	9	15	39	38	49	26	9.3
Variation fraction	0.19	0.42	0.15	0.23	0.13	0.13	0.29	0.11	0.09	0.14	0.20	0.10

II.III Modelfit

Location	Abstraction included	FilterTop [mSurf]	RMSE [m] (measured abstr.)	Fraction of GWD	R2	RMSE (constant abstr.)	Effect const. abstr. on RMSE
B12A0139	N	-3.05	0.36	0.27	0.35	0.351	0.005
B12B0166	N	-9.96	0.46	0.32	0.34	0.443	0.012
B12B0167	Y	-9.88	0.29	0.30	0.74	0.428	0.015
B12D0305	Y	-5.73	0.21	0.19	0.69	0.306	-0.003
B12G0113	Y	-3.95	0.10	0.04	0.83	0.231	0.011
B16F0228	Y	-4.51	0.20	0.10	0.77	0.284	0.009
B16F0301	Y	-1.67	0.18	0.16	0.76	0.215	-0.003
B16H0090	Y	-50.94	0.14	0.07	0.72	0.185	-0.010
B17B0144	Y	-2.97	0.12	0.07	0.74	0.26	-0.010
B17C0076	Y	-2.11	0.18	0.11	0.74	0.199	-0.013
B17C0392	Y	-1.02	0.24	0.26	0.76	0.187	-0.011
B17D0030	N	-57.17	0.44	0.22	0.29	0.164	0.001
B17D0069	Y	-6	0.22	0.09	0.77	0.235	-0.005
B17D0184	Y	-3.52	0.15	0.25	0.76	0.132	-0.003
B17E0043	Y	-12.1	0.16	0.05	0.81	0.204	-0.004
B17E0086	Y	-25.03	0.18	0.05	0.83	0.14	0.002
B17E0161	N	-1.06	0.30	0.93	0.25	0.192	-0.011
B18C0108	Y	-0.85	0.14	0.25	0.69	0.153	-0.003
B22B0022	Y	-37.83	0.19	0.07	0.74	0.122	-0.007
B22F0085	N	-25.98	0.24	0.17	0.45	0.155	-0.012
GRW_Bgeb_6_P B14_P021	Y	-2.64	0.25	0.14	0.79	0.109	-0.008
GRW_Bgeb_6_P B20_P021	Y	-1.15	0.18	0.20	0.77	0.192	0.000
GRW_Bgeb_6_P B6_P021	Y	-1.75	0.19	0.10	0.79	0.229	-0.005

GRW_Midden_P B1_P021	Y	-0.69	0.23	0.27	0.74	0.18	0.001
GRW_Midden_P B15_P021	Y	-2.16	0.13	0.05	0.77	0.233	0.002
GRW_Midden_P B4_P021	Y	-1.11	0.17	0.45	0.76	0.167	-0.007
Abstraction included							
Mean		-1.60	0.19	0.17	0.76	0.19	0.00
Min		-3.37	0.12	0.05	0.69	0.12	-0.01
Max		-0.37	0.29	0.45	0.83	0.28	0.01
Abstraction not included							
Mean		-1.31	0.36	0.38	0.34		
Min		-2.01	0.24	0.17	0.25		
Max		-0.33	0.46	0.93	0.45		

Potential overfitting is shown by series where also shifts and missing periods are reproduced. Series PB15_P021 is left out of the analysis as no stable regime is present in the observations.



II.IV Invloed van sturende factoren in model 2

De tabellen geven de "invloed" van externe factoren in de modelsimulaties voor model 2. De gegeven relatieve "feature importance" geeft het relatieve belang van een factor ten opzichte van de andere externe factoren in de 'beslisboom' van het model, gemiddeld over de hele modelperiode. Het gaat alleen om model 2, die de variaties ten opzichte van de Generieke Trend modelleert, dat al voor een groot deel van de variatie verantwoordelijk is. Het gaat dus niet om een verklaarde variantie.

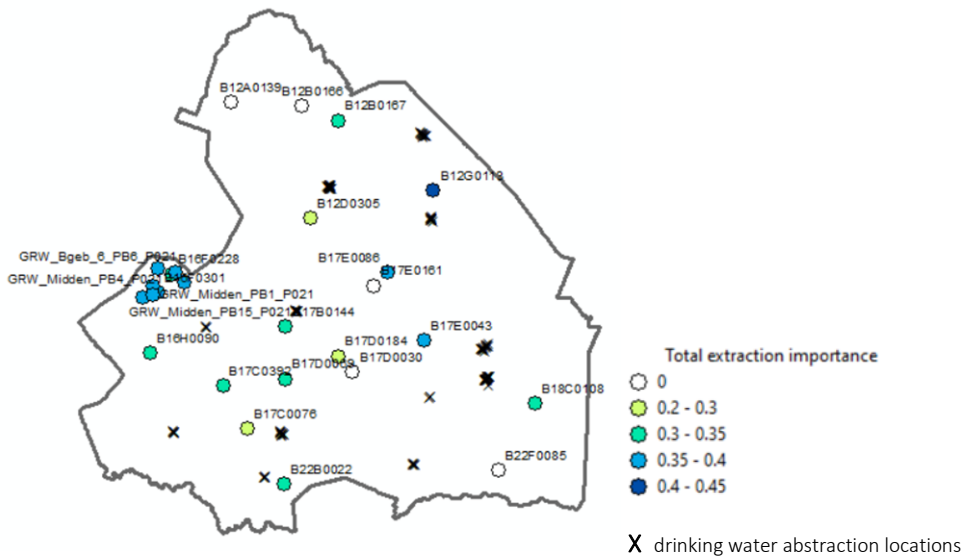
Location	Relative factor importance			FilterTop measloc mSurf	Mean total extraction	Location nearest extr	Distance nearest	Filtertop extr mSurf
	ETpot	Precip	Abstraction					

					10km [m3/d]		extr (km)	
B12A0139	0.62	0.38	NA	-3.05	0	Assen	16.1	-56.4
B12B0166	0.45	0.55	NA	-9.96	0	Assen	10.5	-88.3
B12B0167	0.28	0.37	0.35	-9.88	383	Assen	8.2	-88.3
B12D0305	0.46	0.24	0.30	-5.73	383	Assen	4.5	-70.1
B12G0113	0.33	0.22	0.45	-3.95	1510	Gasselte	3.3	-55.1
B16F0228	0.36	0.29	0.35	-4.51	561	Leggelo	7.9	-46
B16F0301	0.39	0.25	0.36	-1.67	568	Leggelo	7.4	-46
B16H0090	0.42	0.26	0.32	-50.94	566	Leggelo	7.7	-37.5
B17B0144	0.4	0.29	0.31	-2.97	1263	Beilen	2.3	-39.88
B17C0076	0.23	0.48	0.29	-2.11	2422	Hoogeveen	4.2	-105
B17C0392	0.4	0.27	0.33	-1.02	2440	Leggelo	7.7	-37.5
B17D0030	0.5	0.50	NA	-57.17	0	Kruidhaars	10.2	-41
B17D0069	0.3	0.38	0.32	-6.00	1656	Hoogeveen	6.5	-105
B17D0184	0.21	0.50	0.29	-3.52	692	Beilen	7.6	-47.2
B17E0043	0.32	0.31	0.37	-12.1	2821	Kruidhaars	7.2	-51.9
B17E0086	0.36	0.26	0.38	-25.03	443	Gasselte	8.6	-48.6
B17E0161	0.44	0.56	NA	-1.06	0	Beilen	10.1	-47.2
B18C0108	0.46	0.21	0.33	-0.85	1381	Valtherbos	6.3	-39.53
B22B0022	0.48	0.21	0.31	-37.83	1530	Zuidwolde	2.5	-60.35
B22F0085	0.71	0.29	NA	-25.98	0	Dalen	10.6	-28.85
PB14_P021	0.34	0.27	0.39	-2.64	550	Leggelo	9.4	-46
PB20_P021	0.32	0.28	0.40	-1.15	552	Leggelo	6.2	-46
PB6_P021	0.33	0.28	0.39	-1.75	552	Leggelo	7.9	-46
PB1_P021	0.38	0.26	0.36	-0.69	557	Leggelo	8.3	-46
PB15_P021	0.34	0.26	0.40	-2.16	558	Leggelo	8.7	-46
PB4_P021	0.33	0.30	0.37	-1.11	555	Leggelo	7.7	-46
Abstraction included								
Mean	0.36	0.30	0.34	-9.0	1046		6.7	-55.4
Min	0.21	0.21	0.29	-50.9	383		2.3	-105.0
Max	0.48	0.5	0.4	-0.7	2821		9.4	-37.5
Abstraction excluded								
Mean	0.544	0.456	0	-19.44	0.0		11.5	-52.4
Min	0.44	0.29	0	-57.17	0.0		10.1	-88.3
Max	0.71	0.56	0	-1.06	0.0		16.1	-28.9

Relative "feature importance"	Onttrekkingen meegenomen			Onttrekkingen niet meegenomen		
	ETref	Neerslag	Onttrekking	ETref	Neerslag	Onttrekking
Gemiddeld	0.36	0.30	0.34	0.54	0.46	-
Min	0.21	0.21	0.29	0.44	0.29	-
Max	0.48	0.5	0.4	0.71	0.56	-

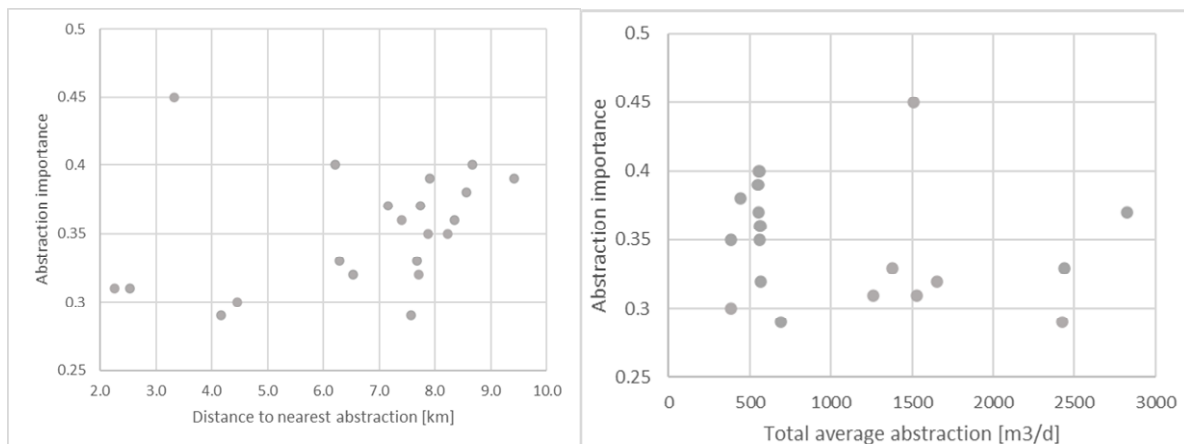
The figures below give more details on the abstraction influence.

- There is no relation between the distance to the nearest extraction well and the influence it has in the simulations
- There is no relation with the difference in depth between the measurement and abstraction wells. Measurement wells are mostly shallow, while abstraction is from deeper aquifers
- Abstraction has an influence in the models for all locations where it is included, even when abstraction and measurement are in different aquifers



Statistics of extraction influence for those locations with abstraction included:

	Depth measurement filter [m-surf]	Mean total extr. 10km [m3/d]	Distance nearest extr. (km)	Depth abstraction filter [m-surf]	Effect constant abstr. on RMSE
Mean	-9.0	1046	6.7	-55.4	0.00
Min	-50.9	383	2.3	-105.0	-0.01
Max	-0.7	2821	9.4	-37.5	0.01



II.V Reproductie van regime-indicatoren

Table: difference between the indicator based on the simulated and observed series. Both in absolute quantities (deviation in m) and as a percentage of the groundwater depth.

Location	Δ Mean year		Δ Mean Dec-Feb		Δ Mean Mar-May		Δ Mean Jun-Aug		Δ Mean Sep-Nov	
	abs. [m]	rel. [%]	abs. [m]	rel. [%]	abs. [m]	rel. [%]	abs. [m]	rel. [%]	abs. [m]	rel. [%]
B12A0139	0.01	1%	0.29	28%	-0.27	-23%	-0.24	-15%	0.28	19%
B12B0166	0.03	2%	0.43	39%	-0.29	-23%	-0.37	-22%	0.35	21%
B12B0167	-0.01	-1%	0.18	31%	-0.17	-26%	-0.22	-17%	0.18	14%
B12D0305	-0.02	-2%	0.13	15%	-0.12	-12%	-0.16	-12%	0.08	7%
B12G0113	0	0%	-0.02	-1%	-0.06	-2%	0.02	1%	0.04	1%
B16F0228	0	0%	0.06	3%	-0.12	-7%	-0.07	-3%	0.13	6%
B16F0301	0	0%	0.09	10%	-0.13	-14%	-0.05	-4%	0.11	8%
B16H0090	0	0%	0.11	6%	-0.07	-4%	-0.1	-5%	0.06	3%
B17B0144	0	0%	0.06	4%	-0.06	-4%	-0.08	-4%	0.06	3%
B17C0076	0	0%	0.15	11%	-0.06	-4%	-0.16	-8%	0.08	4%
B17C0392	0	0%	0.13	22%	-0.09	-13%	-0.13	-11%	0.08	7%
B17D0030	-0.03	-1%	0.33	19%	-0.33	-18%	-0.38	-17%	0.28	13%
B17D0069	0	0%	0.12	6%	-0.11	-5%	-0.13	-5%	0.12	4%
B17D0184	0	0%	0.03	11%	-0.09	-18%	-0.06	-7%	0.1	13%
B17E0043	0	0%	-0.02	-1%	-0.11	-4%	0	0%	0.13	4%
B17E0086	-0.01	0%	-0.05	-2%	-0.11	-4%	0.02	1%	0.12	3%
B17E0161	-0.02	-6%	0.06	43%	-0.31	-348%	-0.08	-16%	0.24	41%
B18C0108	0	0%	0.11	29%	-0.04	-7%	-0.11	-16%	0.05	8%
B22B0022	-0.01	0%	0.1	4%	-0.11	-4%	-0.13	-5%	0.12	4%
B22F0085	-0.01	-1%	0.16	13%	-0.17	-12%	-0.19	-12%	0.17	11%
PB14_P021	0	0%	0.05	3%	-0.18	-13%	-0.08	-4%	0.2	9%
PB20_P021	0	0%	0.05	8%	-0.07	-11%	-0.1	-9%	0.1	9%
PB6_P021	0	0%	0.03	2%	-0.11	-7%	-0.05	-2%	0.12	5%
PB1_P021	0	0%	0.1	20%	-0.17	-29%	-0.09	-8%	0.19	16%
PB15_P021	0	0%	0.05	2%	-0.08	-3%	-0.01	0%	0.06	2%
PB4_P021	0	0%	0.06	36%	-0.1	-53%	-0.05	-9%	0.08	14%
Mean	0.00	0%	0.12	15%	-0.14	-28%	-0.13	-9%	0.14	10%
Mean abs.	0.01	1%	0.11	14%	0.13	25%	0.12	8%	0.13	9%
Min	-0.03	-6%	-0.05	-2%	-0.33	-348%	-0.38	-22%	0.05	3%
Max	0.03	2%	0.43	43%	-0.04	-4%	0.02	1%	0.35	41%

Location	Δ p5	Δ p10		Δ p20	Δ p50		p80	Δ p90		p95	Δ GLG		Δ GHG	
	[m]	[m]	% WTD	[m]	[m]	% WTD	[m]	[m]	% WTD	[m]	[m]	% WTD	[m]	% WTD
B12A0139	0.01	-0.05	-3%	-0.14	-0.08	-6%	0.22	0.22	24%	0.16	-0.1	-6%	0.18	20%
B12B0166	-0.17	-0.21	-11%	-0.25	-0.01	-1%	0.36	0.29	31%	0.24	-0.24	-13%	0.27	31%

B12B0167	-0.05	-0.08	-5%	-0.07	-0.09	-10%	0.1	0.17	46%	0.17	-0.11	-7%	0.06	19%
B12D0305	0.08	0.02	1%	-0.08	-0.11	-11%	0.08	0.11	15%	0.07	-0.01	-1%	0.05	7%
B12G0113	0.12	0.09	3%	0.07	0.01	0%	-0.08	-0.11	-4%	-0.13	0.03	1%	-0.09	-3%
B16F0228	0.11	0.07	3%	0.06	0.02	1%	-0.13	-0.08	-5%	-0.05	0.04	2%	-0.14	-10%
B16F0301	0.19	0.1	6%	0.04	-0.03	-3%	-0.08	-0.04	-6%	-0.04	0.06	4%	-0.08	-13%
B16H0090	0	-0.01	0%	-0.01	-0.03	-1%	0.08	0.06	3%	0.04	-0.04	-2%	0	0%
B17B0144	0.05	0.02	1%	0	-0.04	-2%	0.02	0.02	1%	0.01	0	0%	-0.01	-1%
B17C0076	0	-0.07	-3%	-0.09	-0.03	-2%	0.12	0.09	7%	0.06	-0.07	-3%	0.04	3%
B17C0392	-0.13	-0.08	-6%	-0.04	0	0%	0.03	0.02	7%	-0.02	-0.06	-4%	-0.04	-15%
B17D0030	-0.01	-0.15	-6%	-0.22	-0.15	-8%	0.29	0.22	14%	0.12	-0.22	-9%	0.16	11%
B17D0069	0.12	0	0%	0.04	-0.06	-2%	0.02	0	0%	0.02	-0.01	0%	-0.02	-1%
B17D0184	0.08	0.06	6%	0.03	-0.03	-5%	-0.01	-0.02	-10%	-0.03	0.05	5%	-0.01	-5%
B17E0043	0.14	0.15	4%	0.11	0.03	1%	-0.08	-0.17	-6%	-0.26	0.09	2%	-0.21	-8%
B17E0086	0.11	0.15	4%	0.1	0.04	1%	-0.15	-0.17	-6%	-0.2	0.11	3%	-0.2	-7%
B17E0161	0.37	0.22	27%	0.16	-0.15	-63%	-0.09	-0.11	183%	-0.15	0.19	24%	-0.1	200%
B18C0108	-0.04	-0.05	-6%	-0.04	0.03	5%	0.04	0	0%	-0.03	-0.03	-3%	-0.02	-10%
B22B0022	-0.03	-0.05	-2%	-0.09	-0.03	-1%	0.1	0.15	7%	0.09	-0.03	-1%	0	0%
B22F0085	-0.05	-0.1	-6%	-0.13	-0.04	-3%	0.18	0.09	8%	0.04	-0.09	-5%	0.07	6%
PB14_P021	0.12	0.13	5%	0.1	0.01	1%	-0.14	-0.12	-11%	-0.1	0.06	2%	-0.17	-16%
PB20_P021	0.14	0.05	4%	0.01	0.01	1%	-0.01	-0.1	-30%	-0.12	0	0%	-0.09	-27%
PB6_P021	0.09	0.09	4%	0.05	0.04	2%	-0.1	-0.08	-6%	-0.08	0.04	2%	-0.18	-14%
PB1_P021	0.13	0.05	4%	0.02	-0.04	-5%	-0.05	-0.05	-25%	-0.07	0.03	2%	-0.11	-69%
PB15_P021	0	0.07	2%	0.07	-0.08	-3%	-0.01	-0.1	-4%	-0.12	-0.03	-1%	-0.15	-6%
PB4_P021	0.05	-0.02	-3%	0.01	-0.04	-11%	0.04	0.01	-12%	-0.03	0.04	5%	-0.03	75%
Mean	0.05	0.01	1%	-0.02	-0.03	-5%	0.04	0.02	10%	-0.01	-0.01	0%	-0.02	7%
Mean abs.	0.09	0.08	4%	0.07	0.04	6%	0.10	0.09	18%	0.08	0.07	4%	0.09	22%
Min	-0.17	-0.21	-11%	-0.25	-0.15	-63%	-0.15	-0.17	-30%	-0.26	-0.24	-13%	-0.21	-69%
Max	0.37	0.22	27%	0.16	0.04	5%	0.36	0.29	183%	0.24	0.19	24%	0.27	200%

Location	Δ LL occurrence		Δ LL duration		Δ LL frequency		Δ LL severity	
	Abs. [d/y]	[%]	Abs. [d]	[%]	Abs. [n/y]	[%]	Abs. [m*d/y]	[%]
B12A0139	25.63	72%	3.96	16%	0.71	48%	0.22	5%
B12B0166	74.94	206%	58.39	141%	0.24	27%	-14.44	-308%
B12B0167	11.46	31%	16.24	31%	0	0%	1.09	11%
B12D0305	-5.11	-14%	-20.86	-37%	0.24	37%	3.98	53%
B12G0113	-23.16	-63%	-27.34	-27%	-0.18	-18%	2.71	76%
B16F0228	-12.29	-34%	-34.86	-47%	0.06	12%	3.48	59%
B16F0301	-25.33	-69%	-24.11	-39%	-0.29	-49%	3.86	79%
B16H0090	8.82	24%	-2.81	-5%	0.18	31%	-0.25	-11%
B17B0144	-10.46	-29%	-53.04	-68%	0.59	125%	1.35	42%
B17C0076	27.1	74%	8.31	12%	0.29	55%	-1.3	-29%
B17C0392	25.63	70%	-2.95	-4%	0.41	78%	-3.37	-72%
B17D0030	50.08	137%	33.04	128%	0.06	4%	-3.99	-61%
B17D0069	0.53	1%	-16.52	-24%	0.18	34%	3.11	42%

B17D0184	-21.92	-60%	-20.76	-40%	-0.24	-34%	3.3	82%
B17E0043	-26.1	-72%	-18.86	-24%	-0.29	-62%	2.09	58%
B17E0086	-17.93	-49%	-25.54	-29%	-0.12	-29%	2.49	45%
B17E0161	-34.62	-95%	4.08	15%	-1.29	-95%	6.88	99%
B18C0108	21.92	60%	14.35	30%	0.18	24%	-1.78	-68%
B22B0022	14.22	39%	-7.37	-11%	0.29	55%	-0.9	-24%
B22F0085	66.24	183%	34.15	199%	-0.12	-6%	-5.43	-164%
PB14_P021	-14.49	-40%	-48.12	-55%	0.14	34%	4.09	59%
PB20_P021	-17.54	-48%	-44.83	-60%	0.14	29%	3.23	50%
PB6_P021	-17.65	-52%	-37.02	-52%	0	0%	2.21	50%
PB1_P021	-17.92	-49%	-39.31	-49%	0	0%	4.36	76%
PB15_P021	-7.71	-26%	-46.41	-41%	0.06	23%	-0.06	-8%
PB4_P021	1.56	4%	3.42	4%	0	0%	1.88	29%
mean	4.45	12%	-9.21	1%	0.06	13%	0.67	4%
mean abs.	21.13	58%	22.03	43%	0.23	33%	3.04	61%
min	-34.62	-95%	-53.04	-68%	-1.29	-95%	-14.44	-308%
max	74.94	206%	58.39	199%	0.71	125%	6.88	99%

II.VI Veranderingen in regime-indicatoren in WH50

The table below gives the changes in simulated regime indicators for the WH50 scenario relative to the historical simulations. Right-most column gives a visual evaluation of temporal patterns in the predicted series. R=more regular; RR=extremely regular; IR= more irregular; TB='blocking' values at the top of the distribution; BB='blocking' values at the bottom. In the columns with the changes, unreliable values are greyed out, where the error in the historical simulations was 20% or more compared to the historical observations; or where the predictions show signs of blocking at the bottom (GLG, p5) or top (GHG, p95). Averages were calculated without these locations.

Location	Absolute [m]						Percentage of groundwater depth						Reliability
	Δ Mean	Δ GLG	Δ GHG	Δ p5	Δ p50	Δ p95	Δ Mean	Δ GLG	Δ GHG	Δ p5	Δ p50	Δ p95	Future patterns
B12A0139	0.00	-0.05	0.01	-0.08	0.03	0.00	0%	-4%	1%	-6%	2%	0%	TB + R
B12B0166	0.01	0.00	0.03	-0.02	0.03	0.03	2%	-4%	7%	-3%	2%	8%	R
B12B0167	0.00	-0.13	0.06	-0.04	0.04	-0.01	1%	0%	2%	-2%	2%	2%	TB+R
B12D0305	-0.01	-0.13	0.04	-0.09	0.04	-0.01	0%	-13%	6%	-4%	4%	-1%	TB+R
B16F0228	0.02	-0.36	0.25	-0.24	0.10	0.17	-1%	-12%	4%	-8%	4%	0%	R
B16F0301	0.01	-0.17	0.14	-0.12	0.02	0.02	1%	-2%	7%	0%	-1%	4%	R
B16H0090	0.00	-0.12	0.05	-0.09	0.04	-0.01	1%	-17%	12%	-12%	5%	8%	RR
B17B0144	0.01	-0.13	0.11	-0.10	0.05	0.05	1%	-15%	12%	-10%	2%	1%	RR
B17C0076	0.01	-0.12	0.08	-0.09	0.07	0.03	0%	-6%	2%	-5%	2%	-1%	BB+TB+RR
B17C0392	0.06	-0.17	0.15	-0.02	0.15	-0.02	0%	-8%	7%	-6%	3%	3%	BB+RR
B17D0030	-0.02	-0.11	-0.04	-0.10	0.05	0.00	0%	-7%	5%	-5%	4%	2%	TB+RR
B17D0069	0.00	-0.24	0.15	-0.16	0.07	0.07	6%	-19%	16%	-2%	16%	-2%	BB+RR
B17D0184	0.00	-0.16	0.10	-0.11	0.02	0.08	-1%	-5%	-2%	-5%	3%	0%	TB+RR
B17E0043	0.04	-0.14	0.12	-0.14	0.07	0.13	0%	-10%	6%	-7%	3%	3%	IR
B17E0086	0.05	-0.06	0.05	-0.01	0.06	0.02	1%	-27%	17%	-18%	4%	13%	IR
B17E0161	-0.02	-0.17	0.32	-0.30	-0.07	0.32	1%	-4%	4%	-4%	2%	4%	-

B18C0108	-0.01	-0.10	0.06	-0.07	0.01	0.03	1%	-2%	1%	0%	2%	1%	BB+RR
B22B0022	-0.01	-0.17	0.14	-0.08	0.02	-0.03	-6%	-98%	98%	-93%	-21%	97%	BB+RR
B22F0085	0.00	-0.02	0.00	-0.03	0.01	0.00	-1%	-18%	11%	-12%	2%	4%	TB + RR
PB14_P021	0.03	-0.24	0.12	-0.20	0.14	0.09	0%	-7%	5%	-3%	1%	-1%	
PB20_P021	0.02	-0.29	0.14	-0.34	0.14	0.09	0%	-1%	0%	-2%	1%	0%	R
PB6_P021	0.04	-0.24	0.13	-0.21	0.13	0.15	2%	-13%	7%	-11%	8%	5%	R
PB1_P021	-0.02	-0.20	0.11	-0.19	0.07	0.00	2%	-33%	16%	-39%	16%	10%	
PB4_P021	-0.04	-0.16	0.06	-0.21	0.04	-0.05	2%	-13%	7%	-11%	7%	8%	BB
Mean	0.01	-0.15	0.12	-0.12	0.06	0.06	0%	-11%	16%	-9%	3%	0.01	
Min	-0.04	-0.36	0.05	-0.34	0.01	-0.05	-6%	-33%	1%	-39%	-21%	-0.04	
Max	0.06	0	0.25	-0.01	0.15	0.17	6%	0%	98%	0%	16%	0.06	

Indicatoren laag-niveau-periodes

Grey values = unreliable due to >50% error in historical simulations or strongly disturbed dynamics in future simulations relative to the historical simulations.

Location	Δ LL occurrence	Δ LL duration	Δ LL frequency	Δ LL severity
B12A0139	26.34	9.67	0.24	-2.50
B12A0143	52.09	11.92	0.94	-3.43
B12B0166	5.36	4.14	0.00	-0.44
B12B0167	13.70	-22.33	1.47	-1.26
B12D0305	24.87	10.92	0.29	-4.54
B12G0113	5.83	-22.30	0.41	1.12
B16F0228	40.61	19.97	0.45	-13.28
B16F0301	33.69	30.12	-0.06	-5.73
B16H0090	42.27	17.02	0.47	-4.41
B17B0144	58.32	23.44	0.59	-6.08
B17C0076	38.39	16.55	0.41	-4.87
B17C0392	44.21	9.13	0.88	-5.37
B17D0030	32.75	12.43	0.29	-4.91
B17D0069	47.62	8.58	0.65	-8.61
B17D0184	57.91	19.03	0.71	-8.49
B17E0043	28.11	-19.80	0.65	-5.72
B17E0086	-2.17	-11.55	0.12	-0.12
B17E0161	117.05	95.30	-0.65	-30.64
B18C0108	37.45	-9.91	1.47	-3.86
B22B0022	34.98	9.42	0.41	-3.98
B22F0085	15.00	9.11	-0.12	-0.78
GRW_Bgeb_6_PB14_P021	33.82	13.20	0.40	-10.95
GRW_Bgeb_6_PB20_P021	49.74	40.53	0.06	-14.67
GRW_Bgeb_6_PB6_P021	32.36	36.97	-0.01	-11.22
GRW_Midden_PB1_P021	46.53	19.07	0.46	-9.49
GRW_Midden_PB15_P021	70.71	23.36	0.75	-6.38
GRW_Midden_PB4_P021	38.09	22.19	0.30	-9.37
mean	38.1	13.4	0.37	-5.14
min	13.7	-22.3	-0.12	-9.37
max	58.3	30.1	1.47	-1.26

II.VII Extreem scenario

Tabel onder: Regime-indicatoren voor de historische simulaties en de voorspellingen volgen WH50 en het extreme scenario voor locatie B16F0301. Linker drie kolommen geven de waarden in meters boven NAP, rechter kolommen de verandering ten opzichte van de huidige situatie (historische simulaties).

Indicator	Historische simulaties	Voorspelling WH50	Voorspelling extreem scenario	Verandering WH50	Verandering extreem scenario
Gemiddelde stijghoogte <i>h_{gem}</i> [m +NAP]	5.80	5.80	5.78	0.00	-0.02
GLG [m+NAP]	5.41	5.22	5.17	-0.19	-0.24
GHG [m+NAP]	6.22	6.37	6.35	0.15	0.13
5% percentiel <i>h₅</i> [m]	5.33	5.19	5.11	-0.14	-0.21
50% percentiel <i>h₅₀</i> [m]	5.75	5.77	5.79	0.02	0.04
95% percentiel <i>h₉₅</i> [m]	6.39	6.40	6.37	0.01	-0.02

II.VIII Simulaties Drentsche Aa

Tabel onder: Voorspelde verandering in regime-indicatoren voor het WH50-scenario voor de Drentsche Aa: voorspellingen (2047-2065) in vergelijking met historische simulaties (2003-2020). Merk op dat het model slecht functioneerde en deze resultaten dus onbetrouwbaar zijn.

	WH2050
Verandering gemiddelde waterstand [m]	0.00
Verandering 5% percentiel <i>h₅</i> [m]	0.00
Verandering 20% percentiel <i>h₂₀</i> [m]	-0.01
Verandering 50% percentiel <i>h₅₀</i> [m]	-0.02
Verandering 80% percentiel <i>h₈₀</i> [m]	0.04