

Foutenclassificatie van hydrologische verwachtingen met Machine Learning

JOOST STENFERT, JOOST HEJKERS, GEERTEN HORN, DORIEN HONINGH EN MATTIJN VAN HOEK

Waterbeheerders gebruiken beslissingsondersteunende systemen (BOS) voor operationeel waterbeheer. Een belangrijke component van een BOS is een hydrologisch model dat op basis van weersverwachtingen, in combinatie met data-assimilatie, hydrologische verwachtingen genereert. In deze modellen wordt de werkelijkheid zo goed mogelijk, maar nooit perfect, nagebootst. Zowel het model zelf als de voorspelde neerslag die als invoer dient, zullen een zekere onzekerheid hebben.

Voor het operationeel beheer is het nuttig om te achterhalen wat de oorzaak is van optredende fouten in de voorspellingen van het BOS. Hiermee kan gericht worden gewerkt aan het verbeteren van het BOS. Daarnaast kan snel worden gehandeld als blijkt dat de voorspelling niet zozeer fout was, maar dat het watersysteem afwijkend gedrag vertoont door bijvoorbeeld een verstopping van een watergang. Tijdig herkennen van afwijkend gedrag in het watersysteem kan helpen schade en overlast te beperken. Wij hebben onderzocht of met Machine Learning fouten in verwachtingen herleid kunnen worden naar de oorzaak van de fout. Is het een fout in het hydrologisch model, een fout in de verwachte neerslag, of vertoont het watersysteem afwijkend gedrag?

Artikel

Inleiding

Waterschappen gebruiken een beslissingsondersteunend systeem (BOS) voor het dagelijks peilbeheer en om op tijd te kunnen anticiperen op naderende extreme gebeurtenissen (wateroverlast, storm, droogte, waterkwaliteitsproblemen). Voor een BOS zijn hydrologische verwachtingen noodzakelijk. Als er bijvoorbeeld veel regen (of juist te weinig) wordt verwacht, en daarmee een hoge of lage waterstand, kan het waterschap alvast extra water afvoeren of aanvoeren om de gevolgen van wateroverlast of droogte te beperken.

Binnen dit soort systemen wordt gebruik gemaakt van hydrologische en hydraulische modellen om een neerslagverwachting om te zetten in een waterstand- en afvoerwachting. Een voorbeeld van zo'n operationeel systeem is FEWS-BOS VIDENTE 1.0 dat op dit moment wordt gebouwd door hoogheemraadschap De Stichtse Rijnlanden (HDSR). In de modellen die binnen dit soort operationele systemen worden toegepast wordt de werkelijkheid zo goed mogelijk nagebootst. Modellen blijven echter een geïdealiseerde vorm van de werkelijkheid waarin versimpelingen en fouten aanwezig zijn. Zo kan een weersverwachting anders zijn dan de werkelijkheid, kunnen de afmetingen van kunstwerken foutief zijn

opgenomen, kan een stuw in werkelijkheid een andere stand hebben dan in het model of leidt de schematisering van bodemparameters tot afwijkingen in resultaten.

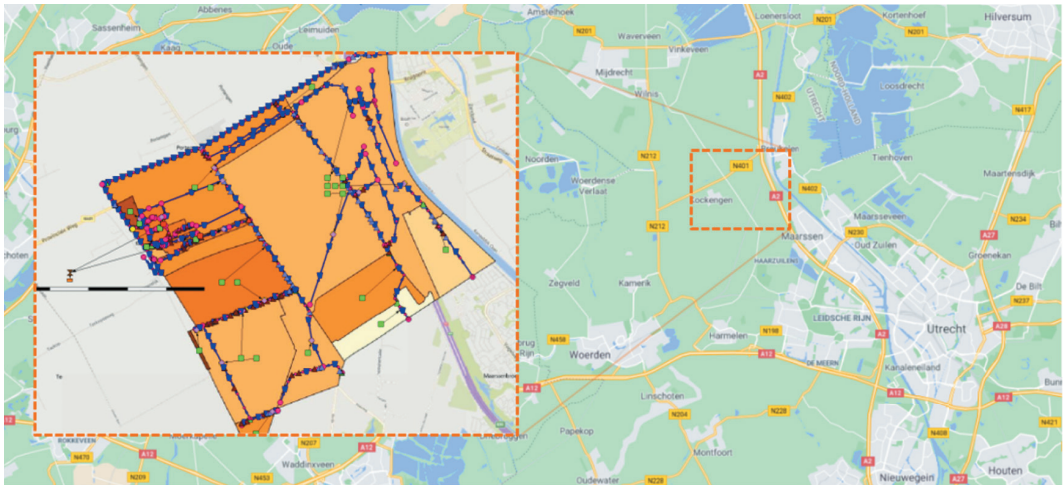
Om operationeel waterbeheer zo goed mogelijk uit te voeren, is het belangrijk om deze modellen continu te verbeteren zodat de verwachting zo nauwkeurig mogelijk is. Toch zullen modelafwijkingen blijven bestaan en is het onrealistisch om te verwachten dat weerverwachtingen perfect kloppend zullen worden in de toekomst. In een operationele setting is het daarom belangrijk om snel in te kunnen schatten of er sprake is van een modelafwijking, of van afwijkend gedrag binnen het watersysteem. In het eerste geval is het belangrijk te achterhalen hoe groot de afwijking is en waar hij door wordt veroorzaakt. In het tweede geval moet ook de oorzaak, zoals een verstopte watergang, achterhaald worden.

Wanneer we deze informatie voor handen hebben zijn we in staat de hydrologische verwachtingsfout snel te reduceren en handelingen bij te sturen.

Het doel van deze studie is het ontwikkelen en uitwerken van een methode om automatisch een 'fout' in een hydrologisch model te detecteren. Hierbij wordt gebruik gemaakt van een numeriek laboratorium, gebaseerd op duizenden modelruns met het SOBEM-model Polder de Tol. Met gebruik van Machine Learning (Self-Organizing Maps) is een algoritme ontwikkeld dat op basis van waterstandsafwijkingen ten opzichte van een referentiesituatie in staat is een afwijking (of fout) te herleiden naar de bron van de afwijking. Met Machine Learning algoritmes is het mogelijk om in deze studie complexe trends en verbanden te ontdekken zonder dat ze expliciet geprogrammeerd te zijn. Dit geeft inzicht in de fysica in het stroomgebied onafhankelijk van menselijke subjectiviteit. Binnen dit onderzoek hebben we gewerkt met drie typen afwijkingen: een fout in het beheerregister, een onnauwkeurigheid in de bodemkarakteristieken (kalibratieparameters van het model) of een fout in de neerslag. Een afwijking of fout is in deze studie gedefinieerd als een afwijking tussen de verwachte en daadwerkelijk opgetreden waterstand.

Casus Polder de Tol

Polder de Tol is onderdeel van het beheergebied van Hoogheemraadschap De Stichtse Rijnlanden en is weergegeven in Afbeelding 1. In de STOWA-studie 'Onzekerheden bij wateroverlast' (Van Dijk e.a., 2019 en in meer detail beschreven door Van Dijk en Botterhuis, 2019) is Polder de Tol als numeriek laboratorium gebruikt om aannames en onzekerheden in het hydrologisch model en het effect daarvan op wateroverlast te onderzoeken. Deze database biedt een schat aan informatie over het effect van onzekerheden met circa 100.000 SOBEM-berekeningen waarin in iedere berekening bewust een parameter in het model is aangepast. Ten opzichte van de referentiesituatie is in iedere berekening op een willekeurige locatie dus een 'fout' toegevoegd. In deze studie zijn resultaten uit deze database gebruikt waarbij 'fouten' zijn geïntroduceerd die effect hebben op de bodemparameters, het beheerregister en de neerslagstatistiek. Dit zijn 'fouten' die concreet aantoonbaar zijn en een effect blijken te hebben op wateroverlast.



Afbeelding 1 Visualisatie SOBEK-model Polder de Tol

Materiaal

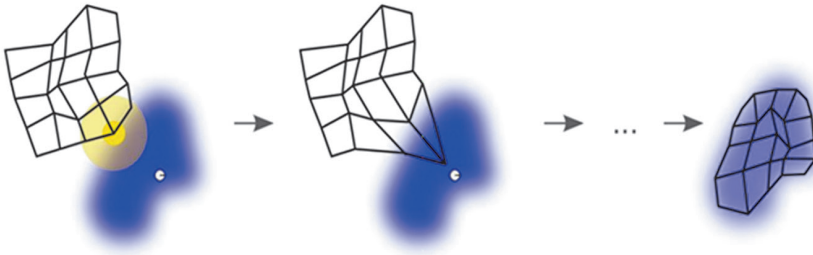
Machine Learning algoritme

In dit project zijn de te onderzoeken patronen op voorhand niet duidelijk. Daarnaast is de beschikbare database ongebalanceerd. Dit betekent dat er voor de ene soort fout veel meer datapunten beschikbaar zijn dan voor de andere soort fout. Op basis van deze kenmerken is gekozen voor de zogenoemde methode van Self-organizing Maps.

Self-organizing Maps zijn kunstmatige neurale netwerken die in staat zijn om multidimensionale data op een slimme manier te clusteren (Milkovic, 2017). Het verdeelt een verzameling van objecten in verschillende groepen op basis van meerdere eigenschappen per object. De Self-organizing Map maakt gebruik van competitive learning, in tegenstelling tot veel andere kunstmatige neurale netwerken waarin gebruik wordt gemaakt van error-correction learning. Competitive learning is een vorm van unsupervised machine learning waarin output nodes in het netwerk concurreren op basis van de gegeven input. De node die gedurende de training het meest geactiveerd wordt door de gegeven input, wordt als winnaar beschouwd en verplaatst zich meer richting de gegeven input, terwijl de rest van de nodes ongewijzigd blijven. Deze techniek is goed toepasbaar op ongebalanceerde datasets waarbij de te onderzoeken patronen op voorhand niet bekend zijn.

Afbeelding 2 toont het concept van training van het algoritme. De paarse vorm illustreert de input data en het raster de neuronen. Na een aantal iteraties vormt het raster zich naar de structuur van de inputdata. Vervolgens kan het gevormde raster gebruikt worden om de data te ordenen. De Self-organizing Map clustert data in groepen die vervolgens worden gerepresenteerd door een gemiddeld gewicht; de gemiddelde waarde per object of per eigenschap. Vervolgens is het ook mogelijk deze complexe multidimensionale clustering te visualiseren op een tweedimensionale kaart. Deze kaart helpt bij het interpreteren van de resultaten en verlaagt het 'black box'-gehalte. Zie voor een uit-

gewerkt voorbeeld van deze methode en visualisatie: <http://ai.hkvservices.nl/european-catchments/>.



Afbeelding 2 Training van een Self-organizing map. De paarse vorm illustreert de inputdata en het raster de neuronen. Na verscheidene iteraties vormt het raster zich naar de structuur van de inputdata en kan het raster gebruikt worden om een 2D weergave van de input data te ontwikkelen.

Data

In de STOWA-studie 'Onzekerheden bij wateroverlast' is voor het genereren van de database voor Polder de Tol het modelprogramma SOBEK toegepast met daarin een gekoppeld hydrologisch model. Een gedetailleerde beschrijving van het model is beschikbaar in Van Dijk en Botterhuis (2019). Iedere modelrun heeft dezelfde voorgeschiedenis en heeft een doorlooptijd van 14 dagen, waarin één neerslagevent optreedt. De modelruns realiseren output voor 14 dagen met een frequentie van 15 minuten.

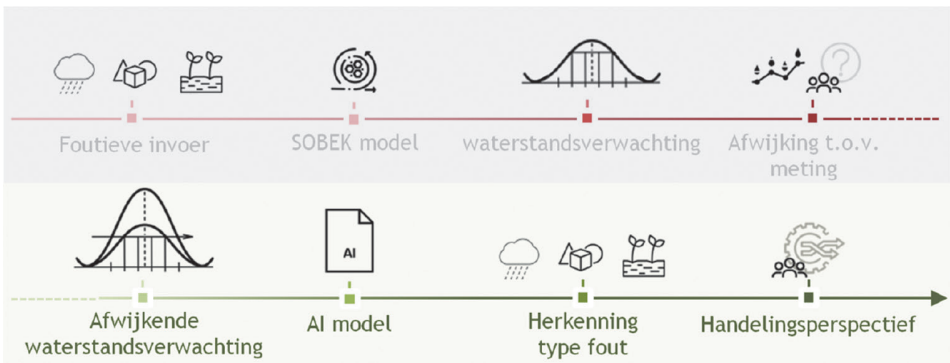
Tabel 1 toont de gebruikte soorten en hoeveelheden data uit Van Dijk en Botterhuis (2019). Ten opzichte van een referentiesituatie zijn per modelrun willekeurige aanpassingen gedaan in het beheerregister, zijn aanpassingen gedaan aan de bodemkarakteristieken én is onzekerheid toegevoegd in een extreme bui (T100).

Tabel 1 Beschikbare data vanuit STOWA-studie 'Onzekerheden bij wateroverlast' (Dijk van, D. e.a., 2019)

Soorten fouten	Aantal modelruns	Informatie
Beheerregister	1525	Willekeurige aanpassing in formaat van duikers, gemalen, stuwen en watergangen.
Bodemkarakteristieken	500	Variatie in dikte bodemlaag en stroomsnelheid in bodem in range van bekende onzekerheid
Neerslagstatistiek	25	Aanpassing van gemiddelde neerslag, 95% interval en standaarddeviatie.

Self-Organizing Maps

In deze studie is een methode ontwikkeld om automatisch een type 'fout' in het model te detecteren op basis van de afwijking in de waterstand. De te volgen stappen zijn getoond in Afbeelding 3.



Afbeelding 3 Werkwijze voor Machine Learning gebaseerde foutenclassificatie van hydrologische verwachtingen met in het groen de toevoeging op de bestaande werkwijze. De stappen zijn opvolgend van linksboven naar rechtsonder.

We hebben gebruik gemaakt van Self-Organizing Maps. Afbeelding 4 toont de stappen die genomen zijn om in dit onderzoek tot een resultaat te komen. De verwachting is dat een fout in bijvoorbeeld de neerslag zich op een andere manier manifesteert in het gebied dan een fout in het beheerregister. Daarom is gekozen voor een uitwerking in twee analyses: locatiespecifiek en gebiedsbreed. Op basis van de geselecteerde en voorbereekte data zijn deze analyses uitgevoerd.



Afbeelding 4 Overzicht stappenplan in dit onderzoek.

Invoerdata worden geclusterd in groepen met vergelijkbare patronen. De modelrun met de initiële instellingen en parameters is gebruikt als referentiesituatie en wordt als 'waarheid' gepresenteerd. Een fout is gedefinieerd als een verandering in het hydrologische model ten opzichte van een referentiesituatie. Om het effect van zo'n fout te kunnen beoordelen is daarom op iedere meetlocatie gebruik gemaakt van het verschil in waterstanden ten opzichte van de referentie modelrun, in plaats van de absolute waterstanden zelf.

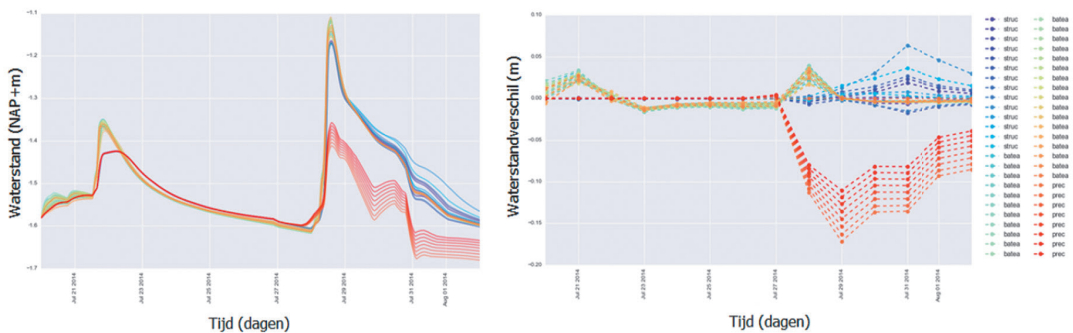
Selectie en voorbereking data

Voor iedere meetlocatie is per modelrun een waterstandsverloop beschikbaar. In dit onderzoek zijn de data geaggregeerd naar gemiddelde dagwaarden. Er is een afweging gemaakt tussen de datahoeveelheid en de karakteristieken van de waterstanden. Bij gemiddelde waarden wordt de datahoeveelheid fors gereduceerd en blijven de belangrijkste karakteristieken in het waterstandsverloop zichtbaar.

Afbeelding 5 toont het waterstandsverloop voor ongeveer 40 modelruns op één locatie in het beheergebied met links het 14-daagse waterstandsverloop met

resultaten per 15 minuten en rechts het geaggregeerde waterstandsverschil ten opzichte van de referentiesituatie ('de waarheid').

De voorbereekte dataset is klaargemaakt als training- en testset voor het Machine Learning algoritme. Hierin zijn 'input-output' koppels gegenereerd. De waterstandsverschillen (input) zijn gekoppeld aan de bijbehorende 'fout' in het model (output). De trainingsset wordt gebruikt om het algoritme te trainen en de testset wordt gebruikt om de kwaliteit van het algoritme te analyseren. De input-outputkoppeling wordt dus in dit project alleen gebruikt in de testfase (en dus niet in de trainingsfase). Het algoritme krijgt tijdens training op geen enkele manier informatie mee over de mogelijke oorzaak van de fout.



Afbeelding 5 Overzicht waterstanden voor verschillende modelruns op één locatie (links) en geaggregeerde waterstandsverschillen ten opzichte van referentie situatie op diezelfde locatie.

Voor de locatiespecifieke toepassing bestaat de input voor het algoritme per modelrun uit waterstandsverschillen corresponderend met de gemiddelde dagwaarden met een doorlooptijd van 14 dagen. Deze locatiespecifieke analyse is uitgevoerd voor één locatie. De output van deze analyse is een clustering van modelruns voor deze specifieke locatie.

Bij de gebiedsbrede toepassing bestaat de input voor het algoritme uit karakteristieken die de verdeling van waterstandsverschillen in het gebied beschrijven. Het waterstandsverschil voor iedere specifieke locatie wordt vertaald naar het hele gebied. Op bijvoorbeeld $t=1$, is voor alle meetlocaties in het hydrologische model een waterstandsverschil bekend t.o.v. de referentiesituatie. Daarmee is het gemiddelde waterstandsverschil op $t=1$ in het gehele gebied te bepalen. Zo worden de volgende statistieken per tijdstap afgeleid op basis van de statistische verdeling in het gebied: minimum, 1^e kwartiel, mediaan, 3^e kwartiel en maximum. Per modelrun bestaat daarmee de input uit de genoemde statistieken per tijdstap: minimum t_1 , minimum t_2 , ..., minimum t_{14} . De output van deze analyse is een clustering van modelruns op basis van het gehele gebied.

Analyse en evaluatie

De Self-organizing Map is toegepast op twee verschillende manieren: locatiespecifiek en gebiedsbreed. De hypothese is dat een fout in het hydrologische model zichtbaar is in het verloop in de tijd op één specifieke meetlocatie, maar ook in de ruimtelijke verdeling van de fout op meerdere meetlocaties in het gebied.

Neem als voorbeeld een fout in de dimensionering van een kunstwerk. Deze fout zal leiden tot stijging of daling van de waterstand en dus tot een verandering (afwijking ten opzichte van referentiesituatie) in het waterstandsverloop. Daarnaast zal deze fout zich ook relatief lokaal presenteren ten opzichte van bijvoorbeeld een fout in de neerslagstatistiek, waarvan het effect over een groter gebied zichtbaar zal zijn. De fout in de dimensionering zal eerder zichtbaar worden op één specifieke meetlocatie en de fout in de neerslagstatistiek zal zich eerder ruimtelijk presenteren, op meerdere meetlocaties tegelijk.

Self-organizing Maps worden gebruikt voor het clusteren van de inputdata. Voor de analyse en evaluatie van deze methode wordt een kleine testset gebruikt. In de evaluatie wordt beoordeeld of de gecreëerde clustering overeenkomt met de testdata. Een Self-organizing Map heeft als resultaat een tweedimensionale kaart (U-matrix). Op deze kaart wordt de testdata geprojecteerd. Op deze manier is te achterhalen of de testdata in de juiste clusters vallen.

Resultaten

De resultaten van de gebied brede en locatie specifieke analyse zijn interactief te raadplegen via:

ai.hkvservices.nl/HDSR-locatiespecifiek

ai.hkvservices.nl/HDSR-gebiedbreed

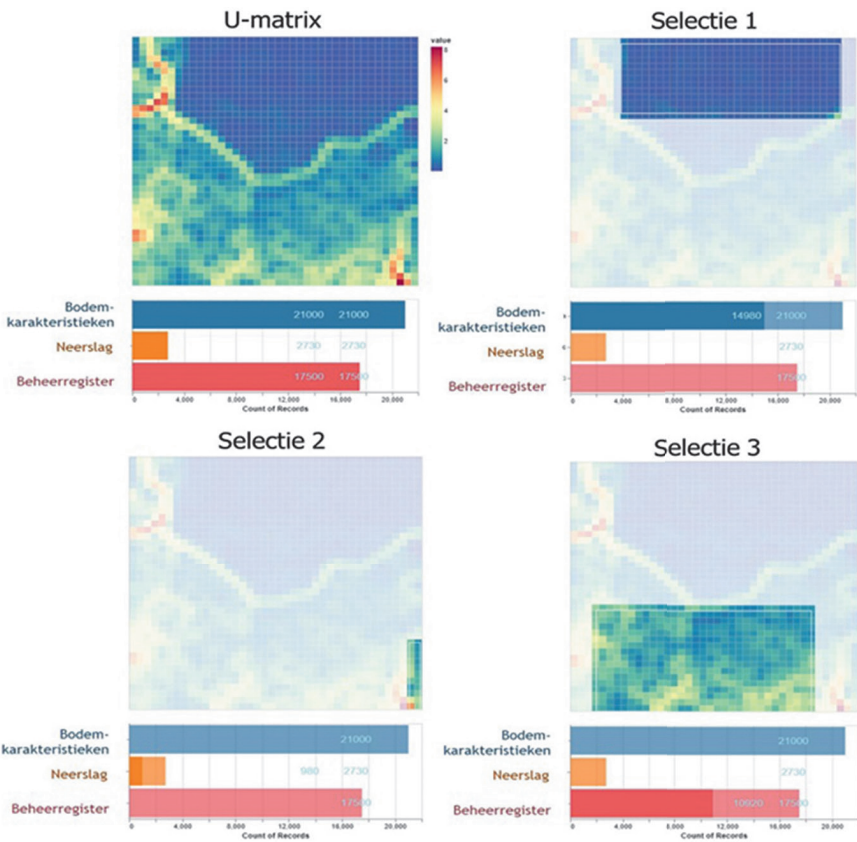
Gebiedsbreed

Afbeelding 6 en afbeelding 7 tonen de resultaten van de Self-organizing Maps waarin de input geclusterd is in een U-matrix. De U-matrix (uniforme afstandsmatrix) is een veelgebruikte weergave van de Self-organizing Map en visualiseert de afstand tussen omliggende datapunten (Ultsch, 2007). In de U-matrix zijn lage waarden - donkerblauwe kleuren - knooppunten die dicht bij elkaar liggen (clusters) en hoge waarden - donkerrode kleuren - ver uit elkaar staande knooppunten (grenzen tussen clusters). Met behulp van deze U- matrices is het mogelijk het algoritme te interpreteren.

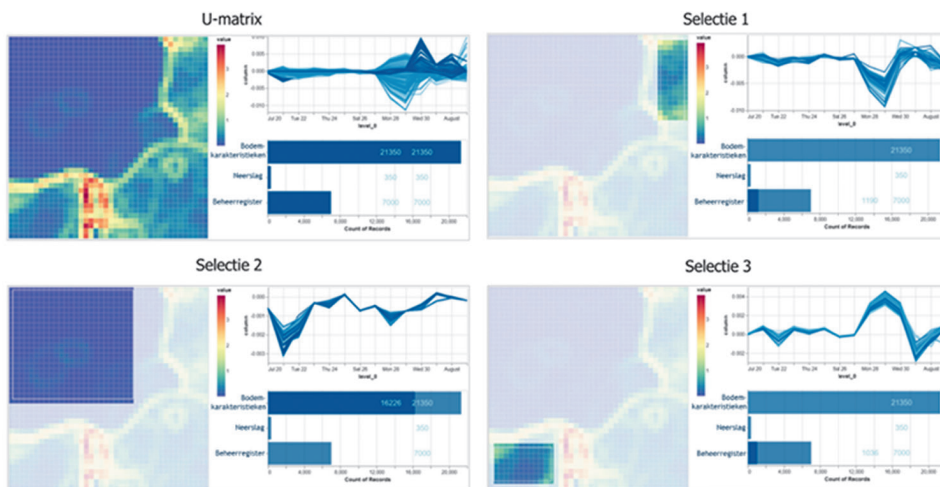
Afbeelding 6 is het resultaat van de gebiedsbrede analyse. De U-matrix toont dat er clustering optreedt: de blauwe cellen (dichtbij elkaar gelegen knooppunten) worden omringd door gele en meer rodere cellen (ver van elkaar gelegen knooppunten). Op basis van de U-matrix is alleen niet duidelijk waaruit deze clusters zijn opgebouwd. Het staafdiagram onder de U-matrix geeft ondersteuning bij de interpretatie. Door gebieden te selecteren wordt de inhoud van het cluster zichtbaar in het staafdiagram.

- Selectie 1 bestaat alleen uit modelruns met fouten in de bodemkarakteristieken (blauwe staafdiagram).
- Selectie 2 bestaat volledig uit modelruns met fouten in de neerslagstatistiek (oranje staafdiagram) en
- Selectie 3 bestaat volledig uit modelruns met fouten in het beheerregister (rode staafdiagram).

Het bijzondere aan dit resultaat is dat het algoritme op voorhand niet weet welke fout er bij een modelrun geïntroduceerd is en deze clustering dus puur



Afbeelding 6 Resultaat Self-organizing map gebiedsbreed. U-matrix met drie selecties. Staafdiagrammen tonen de inhoud van de geselecteerde data. Als alles geselecteerd is (U-matrix, links) zijn alle staafdiagrammen gevuld. Bij selectie van een deel (bijvoorbeeld Selectie 3) bestaat de selectie blijkbaar alleen uit bijvoorbeeld 'fouten' in het beheerregister.



Afbeelding 7 Resultaat Self-organizing Map locatiespecifiek. U-matrix met drie selecties. Staafdiagrammen tonen de inhoud van de selectie. Rechtsboven is per selectie het waterstandsverloop getoond.

op basis van de statistiek van de waterstandsverschillen uitvoert. Toch is het algoritme in staat om modelsimulaties met een vergelijkbare fout in eenzelfde cluster te groeperen.

Locatiespecifiek

Afbeelding 7 toont het resultaat van de locatiespecifieke analyse op één locatie. De U-matrix toont dat er clustering optreedt. De interpretatie is hetzelfde als in Afbeelding 6. Selectie 1 bestaat alleen maar uit modelruns met fouten in het beheerregister. Selectie 2 bestaat volledig uit modelruns met fouten in de bodemkarakteristieken en selectie 3 bestaat zoals selectie 1 volledig uit modelruns met fouten in het beheerregister. Wat deze resultaten goed laten zien is dat een bepaald type geïntroduceerde fout een vergelijkbaar 'uniek' waterstandsverloop geeft op basis waarvan het algoritme ze clustert.

Predictie

Na de ontwikkeling van het algoritme, is het mogelijk om het te voeden met nieuwe informatie vanuit de SOBEK-simulaties. Deze data hebben het algoritme nog niet eerder gezien. Voor iedere nieuwe modelrun wordt op basis van het getrainde algoritme een U-waarde bepaald. Deze U-waarde correspondeert met een locatie op de U-matrix en daarmee is het mogelijk om vast te stellen tot welk cluster een nieuwe simulatie behoort. Voor de gebiedsbrede Self-Organizing Map is de testset gebruikt om te beoordelen of het algoritme de nog niet eerder gebruikte simulaties in de juiste clusters plaatst. Voor 100 modelruns die niet in de trainingsset zitten is de locatie op de U-matrix bepaald (zie Figuur 7). De testdata blijken uniform en consistent op de juiste plekken in de U-matrix te belanden. Oftewel, een modelrun uit de testset met een 'fout' in de neerslagstatistiek valt op een locatie op de U-matrix waarop alleen maar modelruns aanwezig zijn met een fout in de neerslagstatistiek.

Discussie

In dit project is gebruik gemaakt van het numerieke laboratorium 'SOBEK model Polder de Tol', met een uitgebreide database met kunstmatig gegenereerde data. Dit is een aantrekkelijk grote dataset die niet overal beschikbaar is. Het is dan ook een uitdaging om dit concept in te zetten op andere locaties.

Het resultaat voor de Tol is nog niet direct toepasbaar in een verwachtingssituatie. De conclusies die getrokken zijn in het numeriek laboratorium Polder de Tol moeten vertaald worden naar een generieke methode wat gebruikt kan worden in een operationeel BOS. De vraag hierbij is of het mogelijk is om onze methode, waarin een 'fout' is gedefinieerd als een afwijking tussen de referentiesituatie en een modelrun, te vertalen naar de praktijk: in de praktijk zal de verwachte met de daadwerkelijk opgetreden waterstand (oppervlaktewater en grondwater) worden vergeleken, de verwachtingsfout worden bepaald en vervolgens worden geïdentificeerd.

De ontstane clusters zijn achteraf te koppelen aan bestaande fenomenen: Een te kleine duiker leidt bijvoorbeeld tot opstuwing en dus tot hogere waterstanden ten opzichte van de referentiesituatie. Deze conclusie is wellicht ook zon-

der Machine Learning te trekken waardoor het zogeheten 'black box'-gehalte van de analyse afneemt. Wij zien dat Self-organizing Maps wel degelijk helpen deze analyse te verbeteren. Allereerst wordt onbevooroordeeld en objectief naar de data gekeken en daarnaast lijken deze afhankelijkheden en patronen te complex voor een rechtlijnige analyse. Doordat het met Self-organizing Maps mogelijk is het resultaat van het Machine-Learning-algoritme te koppelen aan de gebruikte invoer (de fysica) wordt het resultaat goed te interpreteren en neemt het 'black box'-gehalte af. Self-organizing Maps lijken daarmee uitstekend bruikbaar voor deze hydrologische analyse.

Conclusie

Ons onderzoek laat zien dat het mogelijk is om modelfouten te classificeren en patronen inzichtelijk te maken. Op basis van de gemodelleerde waterstand, zijn we in staat gebleken een afwijking in de waterstand te herleiden tot de oorzaak van deze afwijking.

In dit gezamenlijke innovatieproject van HDSR en HKV is Machine Learning toegepast om de herkomst van fouten in hydrologische modellen/verwachtingen te herkennen. Met gebruik van Self-organizing Maps zijn patronen inzichtelijk gemaakt en clusters gevormd van waterstandsverlopen die horen bij een specifiek soort fout in een hydrologisch model. Denk hierbij aan afwijkingen in het beheerregister, bodemkarakteristieken en de neerslag. Uiteraard zijn er andere foutenbronnen mogelijk. Juist de inzet van een unsupervised algoritme stelt ons in staat om automatisch ook andere fouten te herkennen in de structuur van verwachtingsfouten (in zowel ruimte als tijd).

De resultaten zijn veelbelovend en roepen ook nieuwe vragen op. De toegevoegde fouten in het hydrologische model worden zeer goed onderscheiden in verschillende clusters. Dit biedt op termijn de kans om deze methode in te zetten voor verbetering van hydrologische modellen. Waarbij we opmerken dat in deze studie wordt aangenomen dat het gebruikte model en de opgetreden fouten de werkelijkheid weergeven. De gemeten werkelijkheid is uiteraard anders.

In een vervolg kan deze methode in de praktijk worden gebracht. Om deze methode operationeel in te zetten is het de uitdaging de geleerde lessen (binnen dit numeriek laboratorium) te vertalen naar de praktijk van het operationele waterbeheer. Daarbij komt een aantal vragen naar boven:

1. Ten opzichte van welke situatie wordt het hydrologische model vergeleken? Insteek zou kunnen zijn de verwachting die gisteren werd afgegeven voor de waterstanden van vandaag te vergelijken met de metingen van die waterstanden vandaag;
2. Wanneer is er sprake van een fout? Is bijvoorbeeld elke discrepantie tussen verwachting en observatie een fout of passen we daar een bandbreedte toe, omdat ook observaties fouten bevatten;
3. Hoe zeker weten we dat een bepaalde, door het algoritme gevonden fout, aanwezig is?
4. En wat is het effect hiervan op de waterstanden?

En de belangrijkste vraag is natuurlijk:

Hoe kunnen we de gevonden fouten, inclusief hun grootte, inzetten om enerzijds operationeel de hydro-verwachtingen te updaten de komende dagen en parallel daaraan de fouten in de basis te corrigeren? Deze vragen worden hopelijk verder onderzocht in een vervolgtraject.

Literatuur

- Dijk, D. van en Botterhuis, T.** (2019). Beslissen onder onzekerheid, de invloed van (model)onzekerheden op uitkomsten; 2019. In *Stromingen* 25(2019)3.
- Dijk, D. van, D. Klopstra en T. Botterhuis** (2019) Onzekerheden bij wateroverlast, impact op berekende schades en investeringen; *Stowa-rapport 2019- 48*, Amersfoort.
- Miljkovic, D.** (2017) Brief Review of Self Organizing Maps; in *MIPRO*, Zagreb, pages 1252-1257.
- Utsch, A.** (2007) Emergence in Self Organizing Feature Maps; Marburg. in *Proceedings of the 6 th International Workshop on Self-Organizing Maps (WSOM)*.

Summary Deployment of a Machine Learning-based error classification method for hydrological predictions

Decision support systems (DSS) are used for operational water management. An important component within a DSS is a hydrological forecast model that generates hydrological expectations. This model is usually based on weather forecasts, in combination with data assimilation. In these models, reality is simulated as closely as possible, but like any model remains a simplification of reality with errors and imperfections. Examples include uncertainties around culvert dimensions, soil schematization, rainfall predictions. We have developed a method that uses the Machine Learning method of Self-organizing Maps to detect the cause of a deviation from a model result. This was developed in the numerical laboratory Polder de Tol using data from a previous study on uncertainties in flood risk management. Based on a large database containing the effect of errors in the management register, soil characteristics and precipitation statistics, it is now possible to use this algorithm to find out what the errors are that lead to a certain deviation in the hydrological forecast.

The results show this method is able to provide insight into patterns in water level differences. Based on the modeled water level, we have been able to trace a deviation in the water level to the cause of this deviation. The added errors in the hydrological model are very well distinguished in different clusters. This offers the opportunity to use this algorithm to improve hydrological models and the expectations generated thereby, and thus to improve operational water management.

Auteurs

JOOST STENFERT

HKV Lijn in water
stenfert@hkv.nl

JOOST HEJKERS

Hoogheemrtaadschap De Stichtse Rijnlanden
joost.heijkers@hdsr.nl

GEERTEN HORN

HKV Lijn in water
horn@hkv.nl

DORIEN HONINGH

HKV Lijn in water
honingh@hkv.nl

MATTIJN VAN HOEK

HKV Lijn in water
hoek@hkv.nl