

De grondwaterstand in dijken voorspellen met machine learning

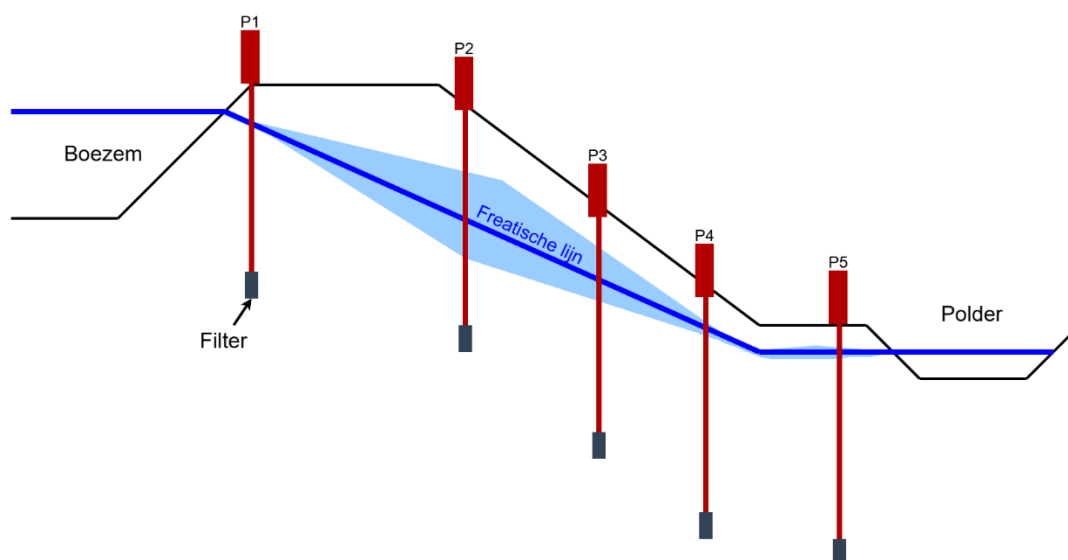
Bram van Eerden, Mitra Baratchi, Jan van Rijn (Universiteit Leiden), Erik Vastenburg (Hoogheemraadschap Hollands Noorderkwartier)

Automated Machine Learning kan een efficiënte en overdraagbare manier bieden om grondwaterstanden in dijken te voorspellen. Door automatisch modellen te kiezen en de juiste peilbuizen te combineren, verbetert de voorspelkwaliteit zonder handmatige afstemming per locatie. Een aanvullende piekbewuste aanpak vergroot de nauwkeurigheid juist tijdens kritieke toenames van de stijghoogte. Daarmee ontstaat een praktisch toepasbare basis voor betere monitoring, vroegtijdige signalering en risicogestuurd dijkbeheer in Nederland.

De gevolgen van klimaatverandering worden ook in het Nederlandse dijkbeheer steeds zichtbaarder. Langdurige droogte en perioden met extreme neerslag hebben een negatieve invloed op de stabiliteit van dijken, die cruciaal zijn voor de waterveiligheid.

De dijken in Nederland worden periodiek beoordeeld op hun veiligheid door de faalkans van de dijk onder extreme omstandigheden te berekenen. Voor de berekening van de grondmechanische stabiliteit is de grondwaterstand, of stijghoogte, in de waterkering een belangrijke parameter. Een snelle of langdurige toename van deze stijghoogte leidt tot een afname van de stabiliteit van de dijk, met mogelijke afschuiving van het grondlichaam tot gevolg [1].

Moderne monitoringssystemen meten de grondwaterstanden dagelijks met peilbuizen. Daarmee kunnen trends en het gedrag van het freatisch vlak worden geanalyseerd. Het voorspellen van dit gedrag onder toekomstige of extreme omstandigheden blijft echter complex, doordat de stijghoogte wordt beïnvloed door neerslag, verdamping, seizoensinvloeden en lokale bodemopbouw [2]. Hierdoor reageert elke dijk anders en zijn modellen vaak beperkt overdraagbaar tussen locaties.



Afbeelding 1. Dwarsprofiel van een dijklichaam. P1, ..., P5: peilbuizen

De metingen van iedere peilbuis resulteren in een tijdreeks, waarbij ieder punt de freatische grondwaterstand op een specifiek moment aangeeft, bijvoorbeeld per dag. Om de waarden van deze meetpunten in de toekomst te voorspellen, gegeven de belastingsparameters (neerslag, verdamping etc.), kan een voorspellend model worden gebruikt. Moderne voorspellende modellen zijn gebaseerd op machine learning; deze modellen leren op een slimme manier het gedrag van een of meerdere tijdreeksen te benaderen, waardoor ze zeer effectief voorspellingen kunnen doen. Machine learning-modellen kunnen zeer accuraat voorspellen, maar voor het afstemmen van het model is veel expertise nodig (zie ook [3], [4]). In deze studie is onderzocht hoe Automated Machine Learning (AutoML) [5] kan helpen bij het voorspellen van stijghoogten in dijken. AutoML-systemen automatiseren de keuze van algoritmen en het afstemmen van de modellen, zodat nauwkeurige modellen kunnen worden ontwikkeld zonder handmatige afstemming per locatie. Dit is interessant voor het dijkbeheer, aangezien er veel lokale verschillen bestaan tussen dijken in Nederland en het aantal meetpunten in de toekomst verder zal toenemen.

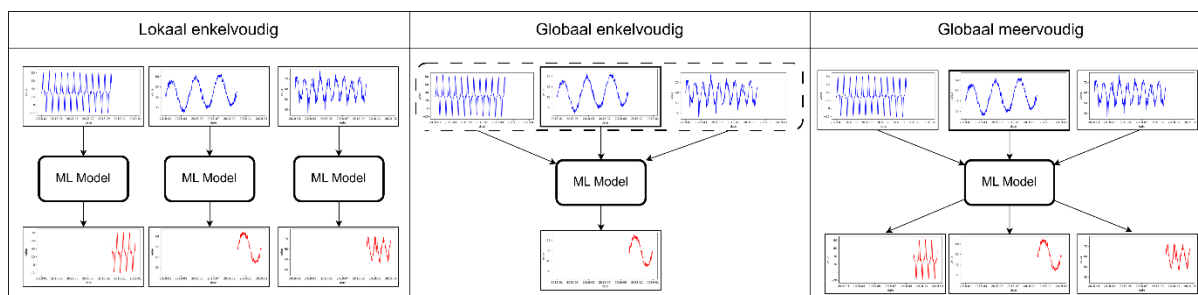
Er is al veel onderzoek gedaan naar het voorspellen van het grondwaterpeil op verschillende plekken in de wereld, maar al het eerdere onderzoek richtte zich op het afstellen van losse modellen voor één specifieke locatie [6], [7]. Deze aanpak heeft geleid tot een groot aantal specifieke modellen waarvan de resultaten niet overdraagbaar zijn naar andere locaties. Daarnaast zijn er meerdere modelleringsscenario's voor voorspellende modellen, die nog onvoldoende zijn onderzocht in de literatuur. Om deze tekortkomingen aan te pakken testen we in dit onderzoek of de AutoML-methodiek van Universiteit Leiden en Hoogheemraadschap Hollands Noorderkwartier (HHNK) een betere, schaalbare aanpak is voor het voorspellen van het freatisch vlak in dijken. We vergelijken AutoML-modellen met klassieke machine learning-modellen, testen verschillende scenario's en stellen een nieuwe methode voor die de prestaties van het best gevonden model tijdens (extreme) piekmomenten verbetert. Het doel is te onderzoeken in hoeverre AutoML, met deze uitbreidingen, nauwkeurige en operationeel bruikbare voorspellingen kan leveren voor het Nederlandse dijkbeheer.

Tijdreeksen modelleren

We bespreken drie manieren om de tijdreeksen van pijlbuisen te modelleren: lokaal eenvoudig, globaal eenvoudig en globaal meervoudig.

'Lokaal enkelvoudig' gebruikt een specifiek model voor iedere pijlbuis apart. 'Globaal enkelvoudig' gebruikt meerdere pijlbuisen, maar voorspelt alleen de tijdreeks van een zogeheten doelpijlbuis.

Het globale meervoudige model gebruikt alle tijdreeksen van een dwarsprofiel en voorspelt deze gelijktijdig.



Afbeelding 2. De drie modelleringsscenario's voor tijdreeksen

Lokaal enkelvoudig

De traditionele benadering in tijdreeksvoorspelling is lokaal: voor iedere tijdreeks wordt een afzonderlijk model getraind met historische meetgegevens. Deze aanpak sluit goed aan bij het idee dat het gedrag van de grondwaterstand binnen elke peilbuis uniek is. Methoden als ARIMA [8] of 'Exponential Smoothing' [9] gebruiken alleen de eigen waarden van de tijdreeks om toekomstige waarden te schatten.

Het voordeel van lokale modellen is dat zij specifieke patronen zoals seizoensinvloeden goed vastleggen. Een belangrijk nadeel is dat informatie uit andere peilbuizen onbenut blijft. Bij meerdere peilbuizen in één dwarsprofiel leidt dit tot versnippering van data, waardoor het model zeldzame gebeurtenissen als extreme pieken minder goed leert.

Globaal enkelvoudig

Een globaal model wordt getraind met de gegevens van meerdere tijdreeksen om gedeelde patronen te leren. In dit geval blijft het globale model enkelvoudig: elke peilbuis-tijdreeks wordt nog steeds afzonderlijk voorspeld, maar het model leert ook van andere tijdreeksen, zoals de invloed van neerslag of verdamping. Wanneer we meerdere tijdreeksen gebruiken om één enkele tijdreeks te voorspellen, spreken we van een doeltijdreeks.

Het voordeel van deze aanpak is dat schaarse data beter kunnen worden benut. Uit tijdreeksen die vergelijkbaar reageren op regenval kan het model patronen leren die het wellicht had gemist door naar een enkele tijdreeks te kijken. Het maakt met deze methode dan ook niet uit of een tijdreeks uit hetzelfde dwarsprofiel of van een compleet andere locatie komt; zolang de tijdreeks maar relevant is voor de doeltijdreeks. Het risico is echter dat niet alle reeksen relevant gedrag vertonen; sommige dijken reageren trager of juist gevoeliger. Als het model wordt blootgesteld aan niet-relevante tijdreeksen kan de prestatie juist slechter worden.

Globaal meervoudig

Bij meervoudige modellen worden meerdere peilbuizen van één dwarsprofiel gelijktijdig voorspeld. Deze benadering maakt gebruik van de samenhangende processen binnen het dijkprofiel: veranderingen in de grondwaterstand bij een van de peilbuizen kunnen ook invloed hebben op de waterstand in de andere peilbuizen. Door de reeksen samen te modelleren kan het model de samenhang tussen de peilbuizen leren. Dit kan leiden tot verbeterde prestaties. Meervoudige modellen zijn conceptueel aantrekkelijk, maar hebben alleen toegevoegde waarde als er daadwerkelijk een relatie is tussen het gedrag van meerdere peilbuizen. Omdat de meervoudige modellen in dit onderzoek zijn getraind op alle peilbuizen binnen één dwarsprofiel zonder verdere selectie, werkt afwijkend gedrag van een enkele peilbuis direct

door in het model. Dit kan ruis en prestatieverlies veroorzaken en maakt meervoudige modellen in deze opzet kwetsbaarder.

Methoden

De drie modelleringsscenario's – lokaal enkelvoudig, globaal enkelvoudig en globaal meervoudig – kunnen worden uitgevoerd door een breed scala aan algoritmes. De keuze van een passend algoritme vraagt tijd en expertise. Daarnaast moet elk gekozen algoritme zorgvuldig worden afgestemd op basis van de beschikbare gegevens om goede voorspellingen te kunnen doen. AutoML-systemen als AutoGluon-Timeseries [10] automatiseren dit proces. In plaats van één model te optimaliseren, traint AutoGluon-Timeseries automatisch meerdere modellen van verschillende modeltypen, waaronder klassieke tijdreeksmodellen, complexe beslisbomen en neurale netwerken.

AutoGluon-Timeseries is ontworpen om efficiënt tot goed presterende voorspelmodellen te komen, zonder dat per tijdreeks model- en parameterselectie nodig is. De aanpak gebruikt bewezen standaardinstellingen die in uiteenlopende toepassingen goede prestaties laten zien. Een uitgebreide technische onderbouwing van de keuzes waarop AutoGluon-Timeseries is gebaseerd, evenals de bijbehorende code, is te vinden in het masterthesis-onderzoek waarop dit artikel is gebaseerd [11].

Een nadeel van automatische modelkeuze is dat de transparantie voor niet-experts kan afnemen. Daartegenover staat dat het systeem direct toepasbaar is op tijdreeksen van verschillende locaties en consistente, nauwkeurige voorspellingen kan leveren zonder specialistische modelkennis. In dit onderzoek vergelijken we de drie modelleringsscenario's, met lokaal enkelvoudige modellen als referentie. Vervolgens analyseren we globaal enkelvoudige modellen met selectie van informatieve tijdreeksen en evalueren we globaal meervoudige modellen binnen één dwarsprofiel. Tot slot onderzoeken we hoe de prestaties van het best presterende model kunnen worden verbeterd tijdens piekmomenten, waarbij de grondwaterstand plotseling sterk stijgt, wat waarschijnlijk een belangrijke bijdrage levert aan de faalkans van de dijk.

Relevante tijdreeksen voor globale modellen

Uit eerder onderzoek blijkt dat globale modellen alleen goed werken wanneer ze worden getraind op een passende selectie van tijdreeksen. Tijdreeksen die sterk afwijken in hun gedrag, bijvoorbeeld in hoe zij reageren op neerslag, kunnen de prestaties van het model negatief beïnvloeden. De kern van het probleem is daarom het bepalen welke tijdreeksen gebruikt moeten worden bij het trainen van een globaal model.

Om deze selectie te automatiseren hebben we 'Bayesian Optimization for Selective Pooling' (BOSP) ontwikkeld. Deze aanpak bepaalt per meetpunt welke andere tijdreeksen het meest informatief zijn. Hiervoor berekenen we eerst kenmerken van elke tijdreeks, zoals variatie in de waterstand en de gemiddelde stijghoogte over de tijd, en vatten deze samen tot een compacte representatie. Vervolgens schatten we met een probabilistisch model in welke combinaties van tijdreeksen naar verwachting de beste voorspellingen opleveren. In een iteratief proces evalueren we steeds nieuwe combinaties, waarbij de methode gericht zoekt naar verbeteringen ten opzichte van eerder geteste selecties. Een uitgebreide beschrijving van de methode en de technische uitwerking is te vinden in eerdergenoemde thesis.

Piekmomenten voorspellen

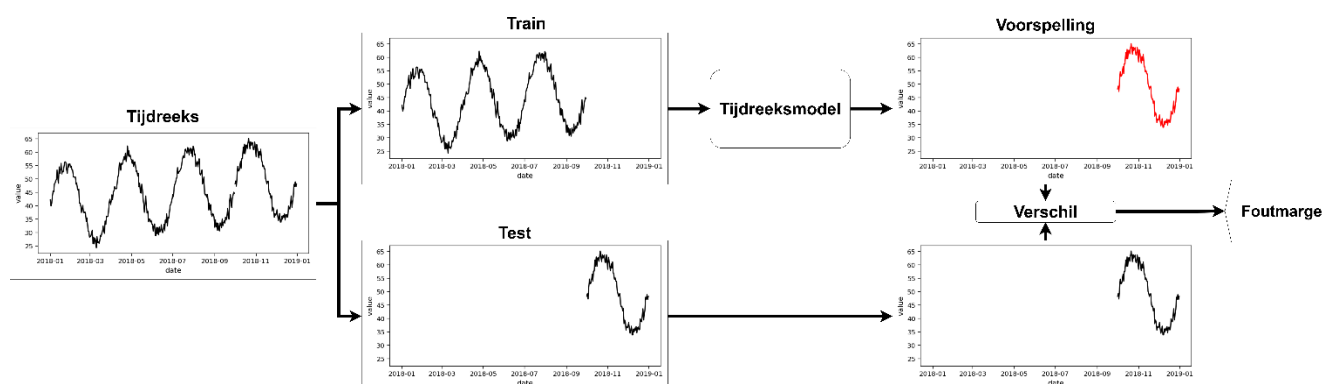
Een tweede ontwikkeling richt zich op een praktisch probleem: plotselinge pieken in de stijghoogte. Zo'n piek kan optreden wanneer de ondergrond van de dijk vrijwel volledig verzadigd is, waarbij bijna alle poriën met water zijn gevuld. Een hevige periode van regenval kan het grondwaterpeil dan zeer snel doen stijgen, wat problematisch is voor de dijkveiligheid. Er bestaan echter veel verschillende oorzaken voor deze pieken, wat voorspellen lastig maakt [12]. Bovendien zijn pieken zeldzaam in de trainingsdata, waardoor standaardmodellen ze vaak onderschatten.

Onze Piekbewuste aanpak bestaat uit twee stappen. Eerst wordt een apart model (een XGBoost-classifier [13]) getraind met de specifieke taak om per dag te voorspellen of er een piek zal optreden. Dit gebeurt op basis van historische trends, neerslag- en verdampingsgegevens. Eerder onderzoek laat zien dat dit de dominante factoren zijn voor de stijghoogte [12]. Andere mogelijke verklarende variabelen, zoals de buitenwaterstand, zijn daarom in dit onderzoek buiten beschouwing gelaten. We gebruiken dit model vervolgens om voor iedere toekomstige dag de kans op een piek te voorspellen. Deze voorspelde piekkans wordt toegevoegd als extra invoervariabele aan het AutoML-systeem. Op deze manier krijgt het voorspelmodel ook een expliciete indicatie dat een piek waarschijnlijk is. Dit helpt het model om beter rekening te houden met plotselinge stijgingen in de stijghoogte.

Een belangrijk aandachtspunt bij deze aanpak is dat ook het piekmodel fouten kan maken. Onjuiste piekvoorspellingen zouden in theorie de kwaliteit van de uiteindelijke voorspelling juist kunnen verslechteren. Dit is getest. De resultaten worden hier beschreven.

Resultaten

Om de prestaties van de gebruikte tijdreeksmodellen te beoordelen, passen we een vaste evaluatieprocedure toe. Afbeelding 3 illustreert deze werkwijze aan de hand van één voorbeeldtijdreeks.



Afbeelding 3. Het evaluatieproces van tijdreeksmodellen

We beginnen met een complete tijdreeks. Van deze tijdreeks houden we het laatste deel apart als testperiode, zodat het model hier geen kennis van heeft. Het eerdere deel vormt de trainingsset en wordt aangeboden aan het voorspelmodel. Op basis van deze trainingsdata genereert het model voorspellingen voor de periode die bewust is achtergehouden. Deze

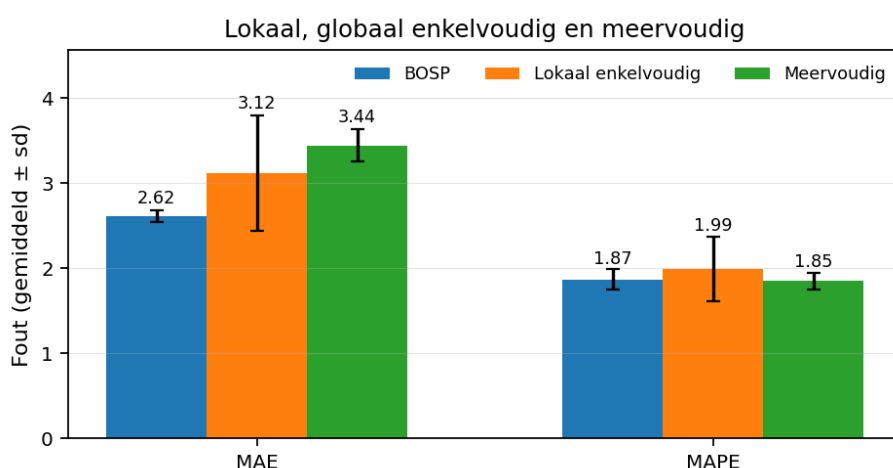
voorspellingen vergelijken we daarna met de daadwerkelijk gemeten waarden in de testperiode.

Het verschil tussen de gemeten en voorspelde waarden wordt de fout of foutmarge genoemd. Hoe kleiner deze fout, hoe beter het model de tijdreeks heeft weten te voorspellen. In dit onderzoek gebruiken we twee veel toegepaste foutmaten. De Mean Absolute Error (MAE) geeft het gemiddelde absolute verschil tussen gemeten en voorspelde waarden en drukt dit uit in centimeters. De Mean Absolute Percentage Error (MAPE) relateert de fout aan de grootte van de gemeten waarden. Wanneer de variatie in het waterpeil klein is, is het bijvoorbeeld interessant om naar de relatieve fout te kijken, aangezien de absolute fout altijd kleiner is dan bij een peilbuis waar het waterpeil zeer sterk fluctueert.

Deze evaluatie herhalen we systematisch voor alle tijdreeksen, waarna we de resultaten samenvatten en vergelijken tussen de verschillende modelvarianten.

Vergelijking van modelleringsstrategieën

In afbeelding 4 is te zien hoe de drie verschillende manieren van modelleren zich tot elkaar verhouden. De experimenten zijn gestart met lokale modellen, die per peilbuis apart worden getraind. Deze aanpak levert stabiele resultaten, maar benut geen informatie van andere meetpunten langs de dijk.



Afbeelding 4. Overzicht van de verschillende modelleringsscenario's. Waarden tonen de resultaten van de modelleringsscenarios in twee foutmaten (MAE en MAPE), berekend over een driedaagse voorspelhorizon

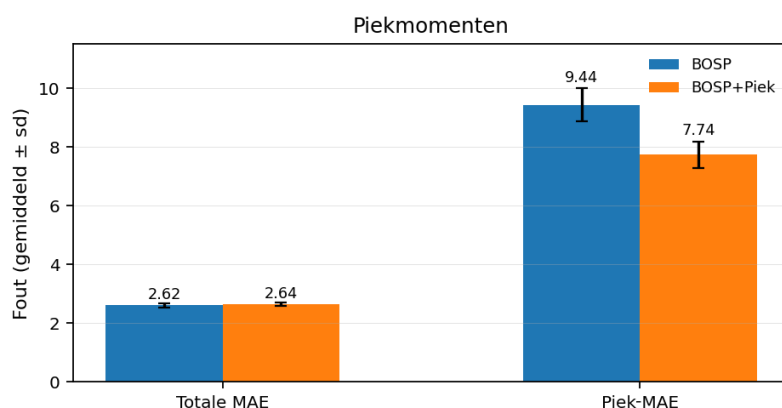
Door meerdere tijdreeksen gezamenlijk te gebruiken in één model kunnen gedeelde patronen worden herkend, bijvoorbeeld in de reactie op neerslag of seizoensinvloeden. Dit werkt echter alleen goed wanneer de juiste tijdreeksen worden gecombineerd. De hier voorgestelde BOSP-aanpak selecteert automatisch welke meetreeksen informatief zijn voor elkaar. Deze globale enkelvoudige aanpak resulteert in de kleinste gemiddelde fout en presteert duidelijk beter dan zowel de lokale enkelvoudige als de meervoudige modellen.

Bij meervoudige modellering worden meerdere peilbuizen binnen één dwarsprofiel gelijktijdig voorspeld. Hoewel dit intuïtief aantrekkelijk klinkt, blijkt het in de praktijk geen structurele verbetering op te leveren. De onderlinge samenhang tussen de peilbuizen is vaak te beperkt, waardoor de extra complexiteit niet wordt terugverdiend in betere voorspellingen. Wel laten

de MAPE-resultaten zien dat meervoudige modellen relatief goed presteren op dijken met kleine waterstandsfluctuaties. Dit suggereert dat deze aanpak potentie heeft in specifieke situaties, maar niet als algemene oplossing.

Voorspellen van piekmomenten

Afbeelding 5 toont het effect van het toevoegen van piekinformatie aan het voorspelsysteem. In de linkerkolom is te zien dat de foutmarge over de complete testset vrijwel gelijk blijft. Dit laat zien dat het extra piekmodel de algemene prestaties niet verslechtert, ondanks het feit dat het ook onzekerheden introduceert.



Afbeelding 5. Overzicht van de resultaten van het model, gericht op piekmomenten. Waardes tonen de gemiddelde MAE in centimeters (cm), berekend over een dieraagse voorspelhorizon

Tijdens piekmomenten is echter wel een duidelijke verbetering zichtbaar: in de rechterkolom zien we dat de foutmarge tijdens de specifieke piekmomenten significant lager is. De fout neemt daar met ongeveer 18 procent af. In driekwart van alle tijdreeksen worden pieken beter benaderd wanneer deze extra informatie wordt gebruikt.

Dit betekent dat de uitbreiding met piekinformatie een gerichte verbetering oplevert voor de meest kritieke situaties in termen van dijkstabiliteit, zonder dat dit ten koste gaat van de algemene voorspelkwaliteit. Daarmee vormt deze aanpak een veelbelovende stap richting praktisch inzetbare voorspelmodellen voor dijkmonitoring en risicogestuurde inspecties en beoordelingen.

Discussie en betekenis voor het dijkbeheer

De resultaten laten zien dat AutoML zeer waardevol kan zijn om stijghoogten in dijken te voorspellen. In tegenstelling tot methoden waar een enkel model moet worden geoptimaliseerd per locatie, biedt AutoML de mogelijkheid om één systeem te gebruiken dat automatisch de beste combinatie van modellen kiest per tijdreeks.

Het onderzoek toont drie belangrijke inzichten voor de praktijk:

- Efficiëntie en overdraagbaarheid: AutoGluon-Timeseries kan zonder handmatige afstemming direct op verschillende dijklocaties worden toegepast. Dit maakt de techniek interessant voor operationele monitoring.
- Selective pooling belangrijk voor prestaties: de resultaten laten zien dat slimme

selectie van vergelijkbare tijdreeksen de prestaties van het model verbetert. Met BOSP kan een systeem automatisch bepalen welke tijdreeksen informatief zijn, zodat de doeltijdreeks beter kan worden voorspeld.

- Betere prestaties bij pieken: door de voorspelde kans op een piek te integreren als aanvullende informatie, presteert het model beter tijdens piekmomenten. De piekbewuste uitbreiding is eenvoudig te implementeren en kan direct worden gekoppeld aan bestaande AutoML-modellen, zonder dat de algemene prestaties eronder lijden.

Praktisch gezien betekent dit dat waterschappen in de toekomst kunnen beschikken over automatische voorspelsystemen die niet alleen trends volgen, maar ook waarschuwen voor plotselinge stijgingen in de stijghoogte. Zo kan de informatie uit peilbuizen sneller worden omgezet in beslissingen, bijvoorbeeld over extra dijksinspecties of als invoer voor toetsingen. Voor toepassing in de dagelijkse praktijk van het dijkbeheer is aanvullend onderzoek nodig, onder andere naar robuustheid bij verschillende meetfrequenties, omgang met ontbrekende metingen en aansluiting bij bestaande beoordelingsprocessen. Dit werk laat zien dat op AutoML gebaseerde tijdreeksmodellen hiervoor een geschikt uitgangspunt vormen, maar het is geen afgerond operationeel systeem. De resultaten bieden daarmee een onderbouwde basis voor vervolgonderzoek en pilotstudies in het dijkbeheer.

Dit artikel is gebaseerd op de masterscriptie van Bram van Eerden in computer science: data science aan de Universiteit Leiden.

Referenties

1. Van Woerkom, T. (2023). *Improving Dike Slope Failure Assessments related to Groundwater Hydrology*. PhD thesis, Universiteit Utrecht
2. Strijker, B., Kok, M. (2025). 'The dynamics of peak head responses at Dutch canal dikes and the impact of climate change'. *Natural Hazards and Earth System Sciences*, 25:3355–3379.
3. Lamers, C., Van Rijn, J., Beenen, T. (2020). 'Data science-technieken om regenwateroverlast in stedelijk gebied te voorspellen'. *H2O-Online*, 2 oktober 2020.
<https://www.h2owaternetwerk.nl/vakartikelen/data-science-technieken-voor-regenwateroverlast-in-stedelijk-gebied>
4. Bavelaar, C., Simons, T., Baratchi, M., Van Rijn, J., Beenen, T. (2021). 'Toepassing van machine learning-technieken op voorspelling van regenwateroverlast in stedelijk gebied'. *H2O-Online*, 10 december 2021. <https://www.h2owaternetwerk.nl/vakartikelen/toepassing-van-machine-learning-technieken-op-voorspelling-van-regenwateroverlast-in-stedelijk-gebied>
5. Baratchi, M. et al. (2024). 'Automated machine learning: past, present and future'. *Artificial Intelligence Review*, 57.
6. Nourani, V., Mogaddam, A. A., and Nadiri, A. O. (2008). 'An ANN-based model for spatio-temporal groundwater level forecasting'. *Hydrological Processes*, 22:5054–5066.
7. Dahl, M. B., Vilhelmsen, T. N., Bach, T., and Hansen, T. M. (2023). Hydraulic head change predictions in groundwater models using a probabilistic neural network. *Frontiers in Water*,

- 5.
8. Box, G.E.P., Jenkins, G.M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day.
9. Hyndman, R.J., Koehler, A.B., and Grose, S. (2002). 'A state space frame-work for automatic forecasting using exponential smoothing methods'. *International Journal of Forecasting*, 18(3):4 39–454.
10. Shchur, O. et al. (2023). 'AutoGluon–TimeSeries: Automl for probabilistic time series forecasting'. In *Proceedings of the Second International Conference on Automated Machine Learning*, volume 224, pages 9–21.
11. Van Eerden, B. (2025). *Automl for hydraulic head forecasting in dikes: A selective pooling and peak-aware approach*. Master's thesis, Leiden University.
12. Strijker, B., Heimovaara, T., Jonkman, S., Kok, M. (2024). 'Exploring subsurface water conditions in dutch canal dikes during drought periods: Insights from multiyear monitoring'. *Water Resources Research*, 60.
13. Chen, T., Guestrin, C. (2016). 'XGBoost: A scalable tree boosting system'. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794.