



## INTRODUCTIE

Een van de kerntaken van een waterschap is het voeren van een oppervlaktewaterpeil dat past bij de gebruiksfuncties. Het te voeren peil legt de waterbeheerder vast in een peilbesluit. Het peilbesluit wordt zowel inhoudelijk als procedureel geëvalueerd met als doel het verbeteren van het peilbesluit. Inhoudelijk wordt geanalyseerd hoe vaak en in welke mate het daadwerkelijke peil afwijkt van het streefpeil.

## ONS IDEE

Het polderlabel. We willen een soort van peilbesluitevaluatie 2.0 op basis van open data en machine learning. We geven elke polder een label, vergelijkbaar met energielabel. Het polderlabel geeft informatie over de robuustheid van een polder. Om uiteindelijk verschillende polders met elkaar te kunnen vergelijken hebben we een heldere definitie nodig van het begrip robuustheid. De polderrobuustheid index wordt gevormd door 1) een neerslagevent en 2) als gevolg daarvan de integraal van de (gemeten) waterstanden tijdserie(s) onder de streefpeilondergrens en boven de streefpeilbovengrens (zie afbeelding 1).

## WAAROM WILLEN WE DIT?

Als je een systeem gaat evalueren wil je in feite ook weten in welke mate het onder druk stond. Een polder (of eigenlijk haar operationele regeling) die weinig is belast en goed presteert, volgens de huidige peilevaluatie, kan minder robuust zijn dan een polder die veel is belast en onvoldoende presteert volgens de huidige peilevaluatie. Door rekening te houden met forceringen (bijv. neerslag, verdamping, voorgeschiedenis bodem, droogte, etc.) en omgevingsfactoren (gewastype, bodemtype, drooglegging, etc.) wordt de analyse verrijkt. Door de analyse over langjarige periode uit te voeren, vergroten we het inzicht in hoe goed een polder presteert in tijden van watertekort en wateroverlast.

## WAAROM IS DIT IDEE RELEVANT?

- 1) beleidsmatig anticiperen:** door het vergroten van het inzicht in de robuustheid van een polder kan beleidsmatig (in een eerder stadium) worden geanticipeerd op droogte voor de landbouw, verdrogen van dijken maar ook het voorkomen van bosbranden. Door het tijdig ingrijpen zijn geen of veel minder extra maatregelen en inzet van mensen nodig. Oogsten worden daardoor niet waardeloos, informeren of opschalen is mogelijk voor gebieden die een hogere brandgevoeligheid hebben zoals bossen, heide en duingebieden maar ook de veiligheid van dijken kan geborgd worden.
- 2) Ruimte voor samenwerking:** als een robuuste polder relatief dichtbij een minder robuuste polder ligt, dan kunnen de polders wellicht voor elkaar compenseren in de waterbalans. Dit kunnen twee polders zijn binnen één waterschap, maar dit kan ook waterschapsgrensoverschrijdend voorkomen (elkaar helpen in tijden van droogte of wateroverlast).
- 3) Datamogelijkheden:** er wordt veel (frequent en met hoog detailniveau) gemeten in Nederland. Het is zonde om al deze beschikbare data niet te gebruiken. Door het slim combineren van allerlei data ontstaat informatie. Een beleidsmedewerker kan die informatie omzetten in kennis waarmee betere peilbesluiten worden genomen.

## METHODE

Het polderlabel is tweedelig opgebouwd: 1) analyse van het robuustheid in het verleden, ook rekening houdend met neerslagtijdseries, en 2) een voorspelling van hoe robuust de polder in de toekomst zal zijn.

POLDERLABEL				
Verleden: evaluatie van polder robuustheid			Toekomst: voorspelling van polderrobuustheid (middels Machine Learning)	
Centrale vraag: Hoe goed werd het peil gehandhaafd rekenend houdend met gevallen neerslag?			Centrale vraag: hoe robuust is de polder (huidige peilbesluit en operationele regeling) in de toekomst?	
Wateroverlast		Watertekort		Alle omstandigheden
Hoe vaak (aantal keer) en hoeveel dagen werd de bovengrens streefpeil bandbreedte overschreden?	Wat is integraal van de waterstand-tijdserie boven de bovengrens streefpeil bandbreedte?	Hoe vaak (aantal keer) en hoeveel dagen werd de ondergrens streefpeil bandbreedte overschreden?	Wat is de integraal van de waterstand-tijdserie boven de bovengrens streefpeil bandbreedte?	Hoe goed kunnen we het huidige waterpeil voorspellen? (= Machine Learning regression_result)  Wat is de voorspellende waarde (m.a.w. relatieve bijdrage van iedere input variabelen (bijvoorbeeld 30daagse neerslag voorafgaand)? (= Machine Learning feature importance)  Hoe vaak en in welke mate wordt het streefpeil overschreden tijdens een neerslag event in bijv. 2050?
Hoe is dat gerelateerd aan de gevallen neerslag?	Hoe is dat gerelateerd aan de gevallen neerslag?	Hoe is dat gerelateerd aan de gevallen neerslag?	Hoe is dat gerelateerd aan de gevallen neerslag?	

**Machine Learning:** met een state-of-the-art machine learning algoritme (Gradient Boosting Regression) wordt het actuele waterpeil in een polder gerelateerd aan verschillende omgevingsvariabelen. Dit komt erop neer dat de het actuele polderpeil wordt verklaard aan de hand van bijvoorbeeld landgebruikstypes, neerslagtekorten en hoogteligging. Het algoritme kan vervolgens gebruikt worden om de robuustheid van een polder te bepalen.

## PILOT

We hebben een klein experiment gedaan om te controleren of de methode werkt. Hiervoor hebben twee peilgebieden in het beheersgebied van Hoogheemraadschap Hollands Noorder Kwartier meegenomen: peilgebied Wieringermeer afd. 2 en peilgebied Hollands Kroon. Het Lizard portaal is gebruikt als databron. De input parameters en doelvariabelen voor de Machine Learning zijn beschreven in de bijlage.

## RESULTAAT (ZIE OOK BIJLAGE B TOT EN MET F)

Het peilgebied Hollands Kroon krijgt een beter label dan de Wieringermeer Afd. 2. Niet alleen komen peiloverschrijdingen in Wieringermeer Afd. 2 vaker, langer en met hogere mate voor. Ook zijn peiloverschrijding bijna een factor 2 keer gevoeliger voor neerslag dan in Hollands Kroon. Het modelresultaat is extreem nauwkeurig:  $R^2$  van 0.99. Dit komt omdat het eigenlijk eenvoudig is om de waterstanden uit de dataset te voorspellen (er is weinig variatie in gemeten waterstand). Kortom: het model hoeft alleen maar een 'toevallig' ander landgebruikstype te identificeren voor deze twee locaties, en dan kan het al bijna perfect voorspellen wat de waterstand is. Daar bovenop dan nog een neerslageffect (SPI). In de bijlage B tot en met F is een uitwerking van het resultaat opgenomen.

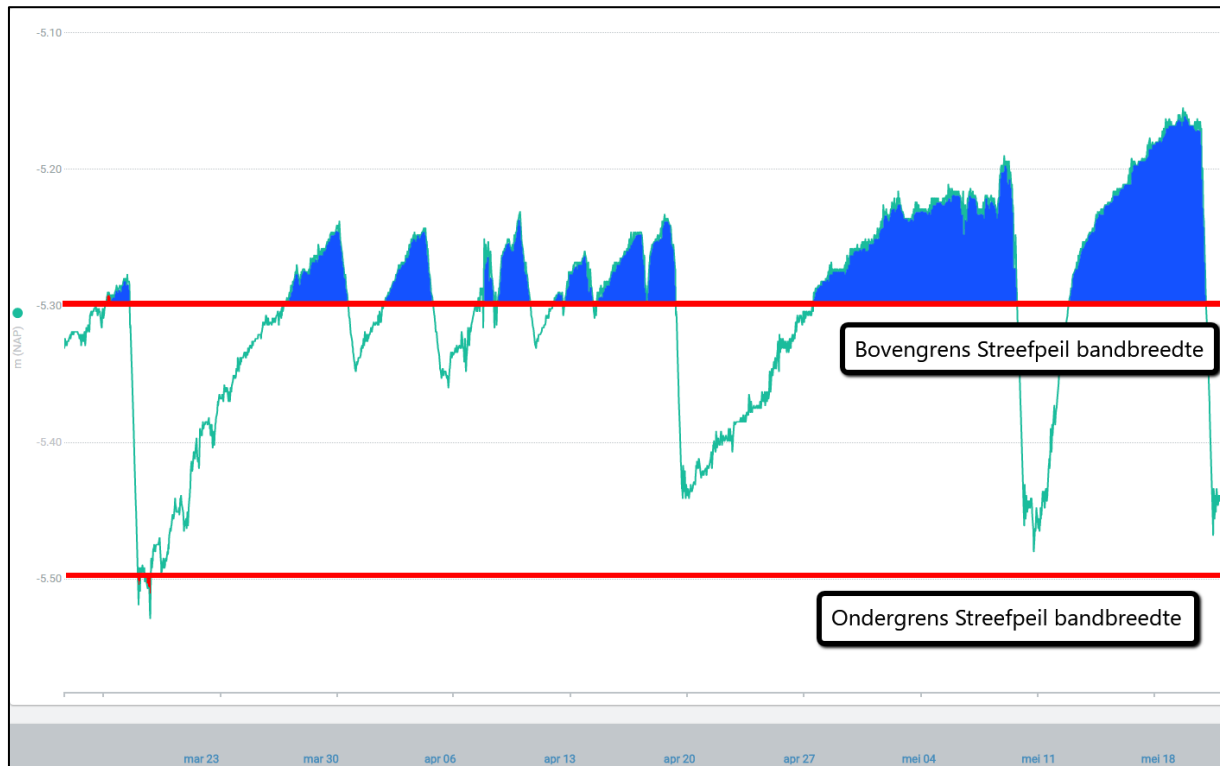
## VERVOLG

**1) Opschalen:** we hebben nu slechts twee locaties gekozen. Om uit te komen op een Machine Learning model dat écht interessante voorspellingen kan doen, moeten we opschalen. We hebben vooral meer gevarieerde data nodig: meer locaties, ook vooral locaties met geheel andere landsgebruikstypen en sterk wisselende waterstanden. Dit kunnen peilgebieden binnen één waterschap zijn, maar ook peilgebieden van verschillende

waterschappen in Nederland. Deze worden allemaal in één analyse gebruikt.

**2) Input tijdreeksen niet versimpelen:** zowel de neerslag als waterstand tijdseries zijn in dit experiment versimpeld naar uurgemiddelden. In het vervolg willen we de 15 minuten tijdserie gebruiken om het algoritme beter te kunnen trainen.

**3) Meerdere meetstation in eenzelfde peilgebied meenemen:** in deze studie hebben we één meetstation representatief gesteld voor het hele peilgebied. Echter, vaak zijn er meerdere meetstations binnen één peilgebied. Door al deze meetstations mee te nemen en ook de gemiddelde waterstand van deze meetstations kunnen we een “slecht geijkt” meetstation voorkomen.



*Afbeelding 1: De waterstandtijdserie van een meetstation in de Wieringermeerpolder met streefpeilgrenzen. De polderrobustheid index wordt gevormd door 1) een neerslagevent en 2) als gevolg daarvan de integraal van de (gemeten) waterstanden tijdserie(s) onder de streefpeilondergrens en boven de streefpeilbovengrens.*