

## Zoutgehalte in de voordelta van het Haringvliet voorspellen met machine learning

*Yannick Steenman, Sam Bom, Bas van Leeuwen (Svašek Hydraulics), Thomas Stolp, Vincent Vuik (HKV)*

**Het zoutgehalte in de voordelta speelt een belangrijke rol bij het operationeel beheer van de Haringvlietsluizen. Om de zoutindringing op het Haringvliet te beperken is een betrouwbare voorspelling van het zoutgehalte noodzakelijk. Er zijn operationele hydrodynamische modellen, maar die hebben aanzienlijke rekentijd. Daarom is onderzocht hoe nauwkeurig een machine learning-model het chloridegehalte bij meetpunt Stellendam Buiten tot 48 uur vooruit kan voorspellen. De resultaten laten zien dat het model aanzienlijk nauwkeuriger kan voorspellen dan het momenteel gebruikte operationele hydrodynamische model en daarmee potentieel geschikt is voor operationele ondersteuning.**

Het Haringvliet was oorspronkelijk een dynamisch estuarium met een belangrijke ecologische functie voor trekvis, zoals zalm en forel. Met de voltooiing van de Haringvlietsluizen in 1970 werd de verbinding tussen rivier en zee grotendeels afgesloten. Om de vismigratie te herstellen wordt sinds enkele jaren het zogenoemde Kierbesluit uitgevoerd, waarbij de sluisen bij vloed onder bepaalde omstandigheden gedeeltelijk worden geopend. Hierdoor stroomt er nu een deel van de tijd water landinwaarts door de Haringvlietsluizen, wat de migratiemogelijkheden voor veel soorten vis verbetert.

Het openen van de sluisen bij vloed leidt echter tot een verhoogde zoutindringing in het westelijke deel van het Haringvliet. Deze zoutindringing kan problematisch zijn voor zowel het waterschap Hollandse Delta als de innamepunten van drinkwater in het Haringvliet. Hierdoor is het noodzakelijk om te weten hoe ver het zoute water het Haringvliet binnendringt, en dus hoe zout het water is dat wordt binnengelaten.

Operationele hydrodynamische modellen kunnen dergelijke voorspellingen genereren, maar hebben relatief veel rekentijd en modelinvoer nodig. Tegelijkertijd komen er steeds meer data beschikbaar uit meetnetwerken en operationele modellen. Dit biedt mogelijkheden om datagedreven modellen toe te passen.

In deze studie onderzoeken we hoe nauwkeurig het chloridegehalte in de voordelta van het Haringvliet 48 uur vooruit voorspeld kan worden door een machine learning-model. In dit onderzoek worden de meetdata bij meetpunt Stellendam Buiten in de voordelta gebruikt voor de opzet van het model. Bij dit meetpunt wordt het chloridegehalte representatief geacht voor het zoutgehalte in de voordelta. De prestaties van het model worden vergeleken met het Operationeel Stromingsmodel Rotterdam (OSR). OSR levert operationele stroming- en zoutvoorspellingen in de haven van Rotterdam en is niet ontworpen om zoutgehaltes te voorspellen in andere regio's. Het wordt hier momenteel echter wel voor gebruikt.

## Invoerparameters

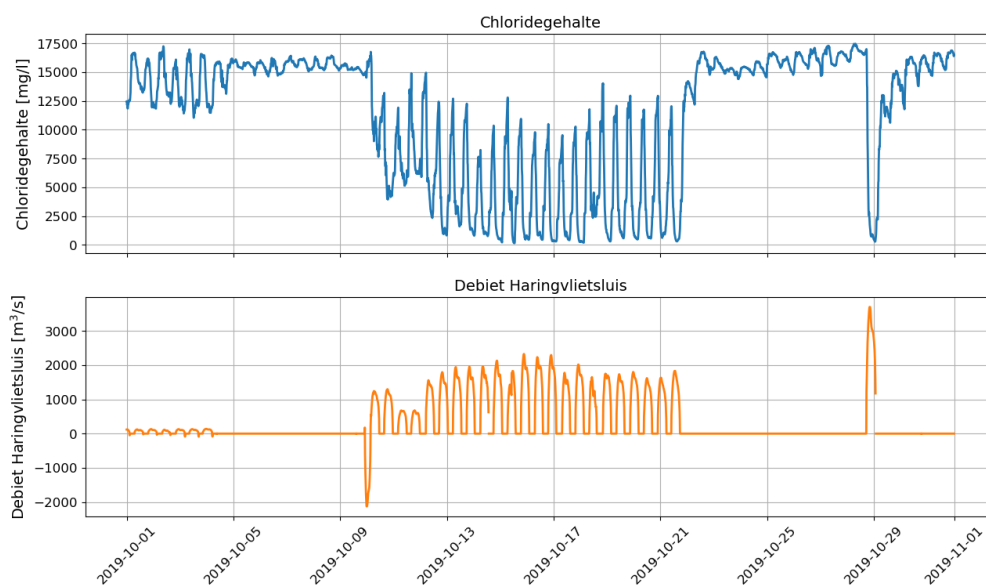
Eerder onderzoek heeft aangetoond dat het chloridegehalte bij Stellendam Buiten voornamelijk wordt bepaald door drie factoren:

- het spuidebiet door de Haringvlietsluizen
- de waterstand (getij)
- de windrichting en windsnelheid

Van deze parameters zijn zowel gemeten historische datapunten als voorspelde toekomstdata beschikbaar. Naast de bovengenoemde parameters speelt de recente historie van het chloridegehalte zelf een belangrijke rol. De recente historie van het zoutgehalte geeft context aan het model, zoals de trend in het chloridegehalte in de afgelopen 48 uur en het meest recent gemeten chloridegehalte. Doordat we de gemeten historie gebruiken garanderen we dat het model het meest nauwkeurige startpunt kiest voor de toekomstvoorspellingen.

Het spuidebiet heeft hiervan de grootste invloed. Bij hoge spuidebieten wordt zoet rivierwater naar zee afgevoerd, waardoor het chloridegehalte in de voordelta sterk daalt. Bij lage spuidebieten kan zout zeewater verder landinwaarts in de voordelta doordringen. Afbeelding 1 illustreert het effect van het spuidebiet op het chloridegehalte in de voordelta.

Naast het spuidebiet zorgt het getij voor een periodieke variatie in het chloridegehalte. Tijdens perioden zonder spuien zijn de chloridegehalten tijdens de transitie naar laagwater doorgaans hoger dan bij de transitie naar hoogwater. Ook wind kan het zouttransport beïnvloeden, met name wanneer de windrichting loodrecht op de Haringvlietsluizen gericht is.



Afbeelding 1. Relatie tussen chloridegehalte (bovenste figuur) en spuidebiet door de Haringvlietsluizen (onderste figuur). Negatief spuidebiet betekent dat er water wordt ingenomen

## Methode

Om de chlorideconcentraties te voorspellen is een machine learning-model ontwikkeld op basis van een *long short-term memory* (LSTM)-netwerk. Dit type model is geschikt voor het analyseren van tijdreeksen en wordt veel toegepast voor hydrologische en hydrodynamische voorspellingen.

Het model gebruikt historische meetdata van het chloridegehalte, het spuidebiet, de waterstand en de windrichting en windsnelheid. Deze historische meetdata worden gebruikt tot het moment van de voorspelling. Daarnaast zijn voorspellingen van spuidebiet, waterstand en windrichting/snelheid gebruikt om toekomstige chloridegehalten te berekenen.

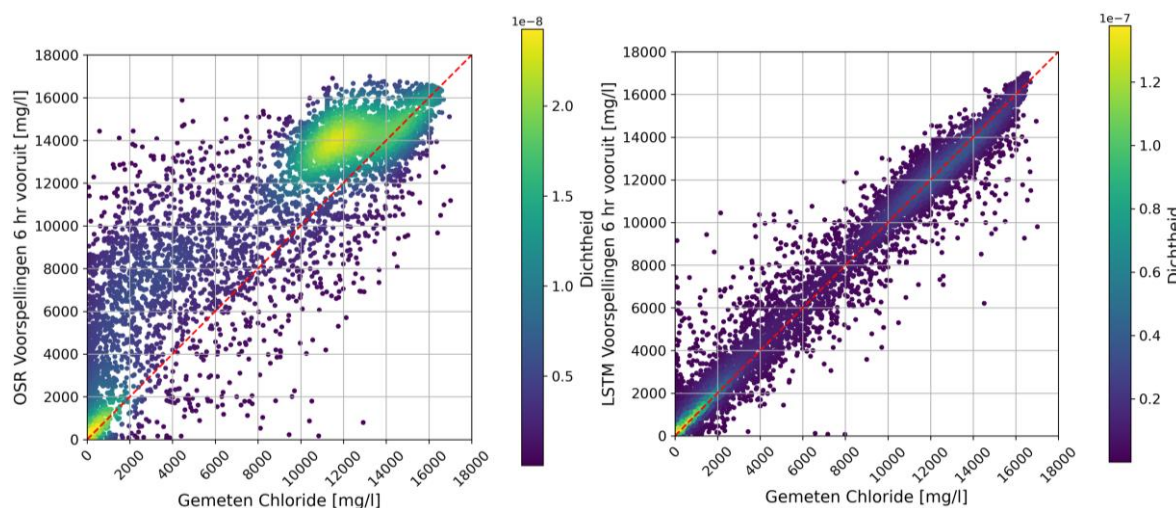
De beschikbare meetreeksen beslaan de periode van 2018 tot 2025 en hebben een tijdsresolutie van tien minuten. Voor het model zijn deze gegevens geaggregeerd naar uurlijkse waarden.

Het model gebruikt steeds de gegevens van de voorgaande 48 uur als invoer en genereert voorspellingen tot 48 uur vooruit. Door de nauwkeurigheid van het model te evalueren in een validatie- en testperiode, voorkomen we dat het model alleen goede resultaten levert in de trainingsperiode door over-optimalisatie op de trainingsdata (overfitting).

## Resultaten

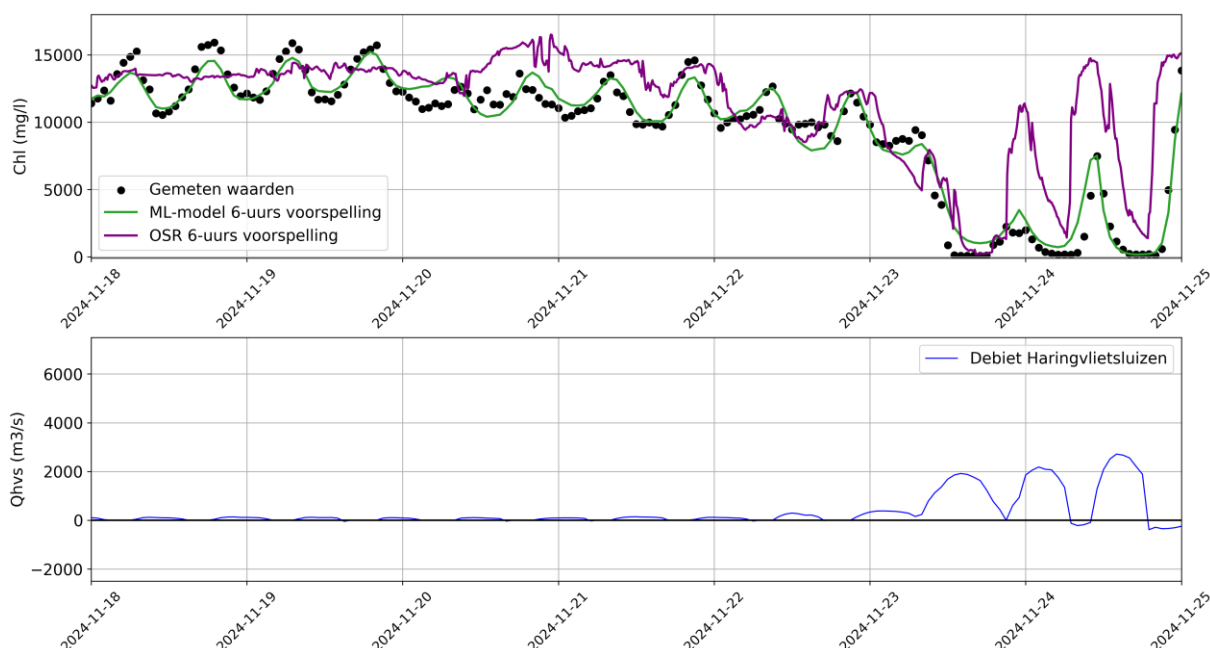
Het model genereert elk uur een voorspelling voor het chloridegehalte tot 48 uur vooruit. Voor operationeel beheer is vooral de voorspelling zes uur vooruit relevant, zodat deze beschikbaar is rond het moment waarop beslissingen over de schuifconfiguratie van de Haringvlietsluizen worden genomen.

De prestaties van het machine learning-model (ML) zijn vergeleken met die van het numerieke OSR (zie boven). De gemiddelde foutmarge van het ML-model voor een voorspelling zes uur vooruit is ongeveer 620 mg/l, terwijl OSR een gemiddelde fout van zo'n 2360 mg/l laat zien. Afbeelding 2 toont de afwijking ten opzichte van de meetdata voor OSR (links) en het ML-model (rechts). Bij een voorspelling 48 uur vooruit neemt de gemiddelde foutmarge van het ML-model toe tot 840 mg/l.



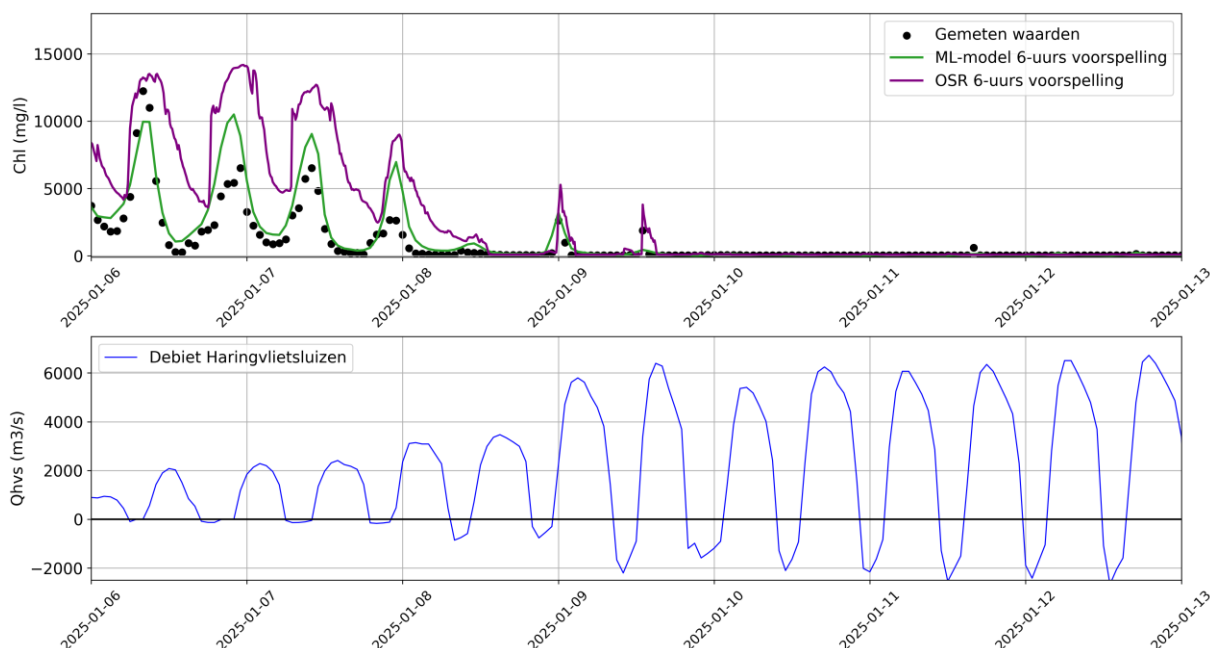
Afbeelding 2. Afwijking in voorspelling van het chloridegehalte ten opzichte van de meetdata voor OSR (links) en het ML-model (rechts). De rode stippellijn geeft een perfecte voorspelling weer waarin  $y=x$

Daarnaast blijkt dat het ML-model beter in staat is om variaties in het chloridegehalte te reproduceren tijdens perioden met weinig spuidebiet. In deze situaties vertoont het chloridegehalte een duidelijke variatie onder invloed van getij en wind. Het ML-model kan deze variatie grotendeels volgen, terwijl OSR vaak een vrijwel constant chloridegehalte voorspelt. Afbeelding 3 toont de voorspellingen van het ML-model ten opzichte van de meetdata en OSR voor een overgangsregime van lage naar gematigde spuidebieten.



Afbeelding 3. Boven: chloridegehalte bij meetpunt Stellendam buiten. De zwarte punten zijn de gemeten waarden, de groene lijn de ML voorspelling 6 uur vooruit, en de paarse lijn de OSR-voorspellingen 6 uur vooruit. Onder: debiet door het Haringvliet, 18 november 2024-25 november 2024

Tijdens perioden met hoge spuidebieten daalt het chloridegehalte sterk door de invloed van zoet rivierwater. Beide modellen zijn in staat deze overgang te reproduceren, maar het ML-model volgt de gemeten waarden doorgaans nauwkeuriger. Afbeelding 4 illustreert de voorspellingen van het ML-model en OSR bij een dergelijke overgang.



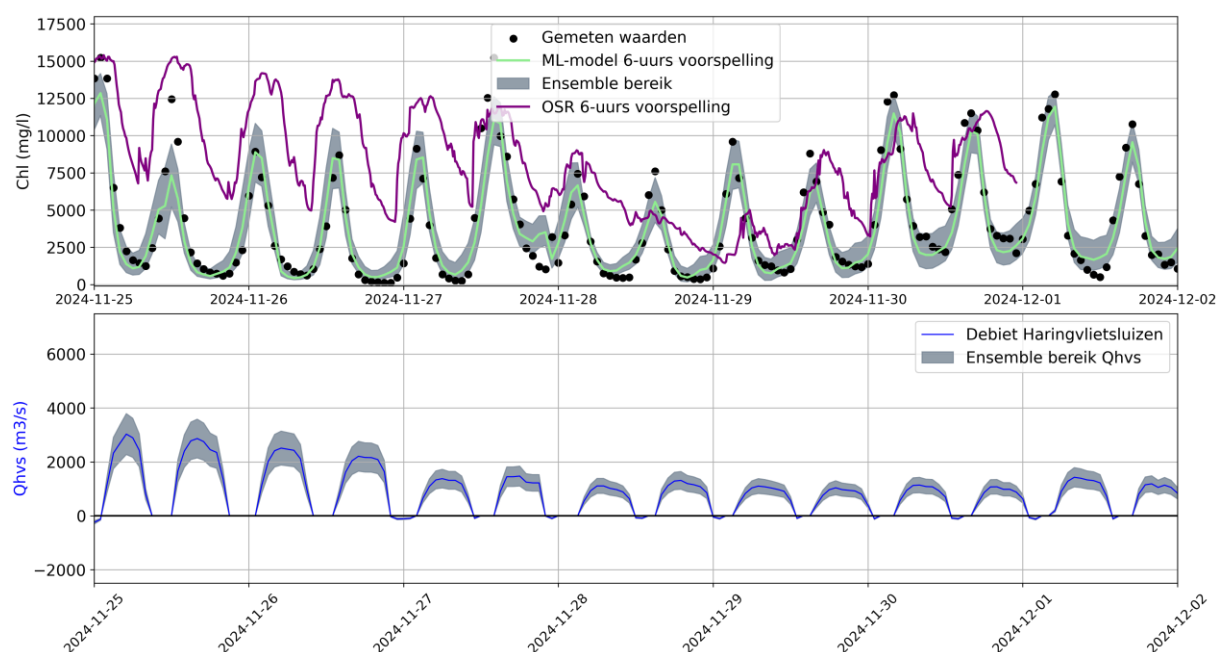
Afbeelding 4. Boven: chloridegehalte bij meetpunt Stellendam buiten. De zwarte punten zijn de gemeten waarden, de groene lijn de ML voorspelling 6 uur vooruit, de paarse lijn de OSR-voorspellingen 6 uur vooruit. Onder: debiet door het Haringvliet, 6 januari 2025-13 januari 2025

## Onzekerheidsanalyse

Een voordeel van een datagedreven model is dat het snel rekt en daardoor relatief eenvoudig kan worden gecombineerd met probabilistische invoer. De toekomstvoorspellingen van het zoutgehalte zijn zowel in datagedreven als numerieke modellen afhankelijk van voorspellingen van relevante andere, zoals het spuidebiet. In deze voorspellingen zit onzekerheid. Dit kan expliciet worden gemaakt door meerdere mogelijke toekomstscenario's (ensembles) te kwantificeren. Door deze ensembles van de invoerparameters te gebruiken in het ML-model, kan een bandbreedte van mogelijke zoutgehaltes worden bepaald.

In deze studie is dit concept getest door variaties aan te brengen in de voorspellingen van spuidebiet en waterstand. Afbeelding 5 toont het resultaat. De figuur laat zien dat de onzekerheid in de voorspelling van het chloridegehalte vooral zichtbaar wordt tijdens pieken en dalen in het chloridegehalte.

Een dergelijke probabilistische voorspelling kan waardevolle informatie bieden voor operationele besluitvorming rond het beheer van de Haringvlietsluizen.



Afbeelding 5. Boven: chloridegehalte bij meetpunt Stellendam buiten. De zwarte punten zijn de gemeten waarden, de groene lijn de ML-voorspelling 6 uur vooruit. De grijze bandbreedte toont de bandbreedte van de verschillende ensemble-voorspellingen, en de paarse lijn de OSR-voorspellingen 6 uur vooruit. Onder: debiet Haringvliet, 25 november-12 december 2024

## Conclusie

Deze studie laat zien dat een machine learning-model in staat is om het chloridegehalte in de voordelta van het Haringvliet 6 uur vooruit te voorspellen met een gemiddelde foutmarge van 620 mg/l. Deze foutmarge neemt toe tot 840 mg/l bij een voorspelling van 48 uur vooruit.

We hebben aangetoond dat het ML-model nauwkeurigere voorspellingen van het zoutgehalte bij meetpunt Stellendam Buiten genereert dan het numerieke model OSR. De toename in nauwkeurigheid is zichtbaar in een significant lagere gemiddelde foutmarge, maar ook in een betere reproductie van de trends in het zoutgehalte. Hieruit concluderen wij dat machine learning kan worden ingezet om het chloridegehalte in de voordelta van het Haringvliet operationeel te voorspellen.

Het ontwikkelde model kan met spuidebiet, waterstand, wind en historische chloridemetingen als invoer voorspellingen tot 48 uur vooruit genereren. Door de snelheid van het ML-model is het mogelijk om onzekerheden in deze invoerparameters expliciet mee te nemen via ensemblevoorspellingen. Dit kan bijdragen aan een betere onderbouwing van operationele beslissingen rond het beheer van de Haringvlietsluizen.

In algemene zin demonstreert deze studie dat datagedreven modellering zeer geschikt kan zijn voor het voorspellen van situaties waarvoor al veel meetgegevens beschikbaar zijn.