

## Een nieuw perspectief op hydroinformatica

13 augustus 2024 – H2O-Online

De drinkwatersector staat voor complexe uitdagingen. Het is daarom tijd voor een nieuw perspectief op hydroinformatica, waarin, afhankelijk van het doel en de toepassing, de data, de modellen of het risico als uitgangspunt worden genomen.

---

*Geschreven door Peter van Thienen, Ina Vertommen, Dragan Savić (KWR)*

---

De afgelopen tien jaar is de ICT-wereld veranderd. Eerst draaide alles om applicaties (application centricity), maar nu staan data centraal (data centricity) [1], [2]. Dit betekent dat data als het belangrijkste worden gezien, en dat de processen die de data gebruiken gemakkelijk kunnen worden aangepast of vervangen. Oude programma's kunnen worden vervangen door nieuwe, maar de data blijven hetzelfde en worden steeds verder aangevuld. Deze verandering komt door de enorme toename van

de hoeveelheid data uit bronnen, zoals sensoren (denk aan satellieten en drones) en camera's, en ons gebruik van digitale media.

Ook kunstmatige intelligentie (AI) volgt deze trend. Bij AI gaat het nu meer om de kwaliteit en betrouwbaarheid van data, dan alleen maar om het verbeteren van modellen [3]. In dit artikel wordt besproken hoe deze ontwikkeling ook te zien is in de hydroinformatica en wat de voordelen daarvan zijn.

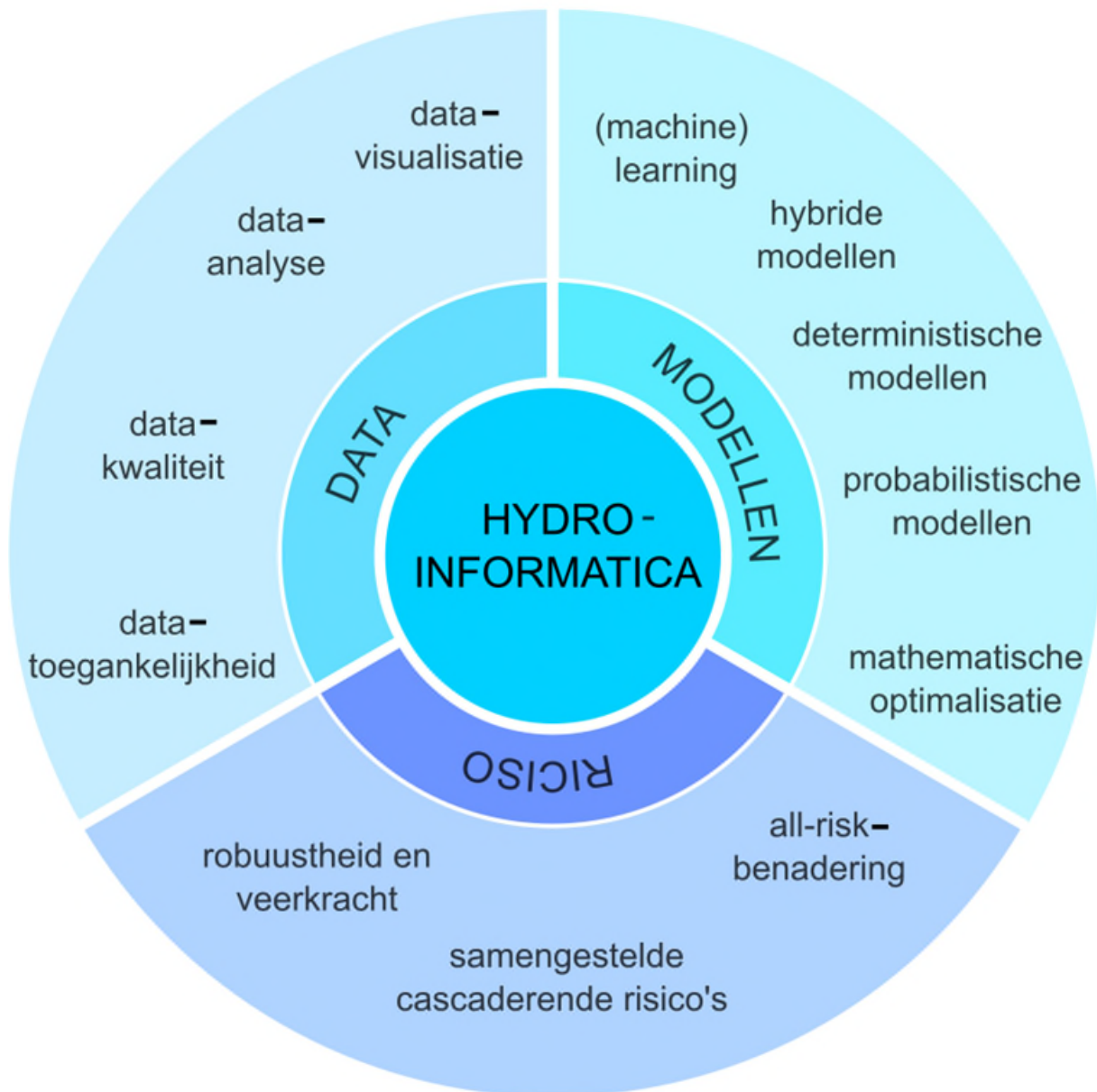
### **Hydroinformatica als ICT+water**

Hydroinformatica ontstond eind jaren '90 van de vorige eeuw als onderzoeksveld dat informatietechnologie (IT) toepast op waterbeheer. De hydroinformatica combineert technologische, sociologische en milieuperspectieven. Hoewel hydroinformatica altijd interdisciplinair is geweest, wordt het, vooral in Nederland, vaak gezien als de IT-afdeling van de watersector. Dit is een belangrijke rol, zeker nu de sector steeds meer digitaliseert. Toch doet dit beeld geen recht aan de andere belangrijke activiteiten en ontwikkelingen in de hydroinformatica. Het vakgebied moet worden gezien als een filosofie die systemen ontwikkelt om wereldwijde waterproblemen aan te pakken, mogelijk gemaakt door informatietechnologie.

### **Een nieuwe drie-eenheid**

In de ICT-wereld, en in mindere mate ook in de AI-wereld, vindt een fundamentele verschuiving plaats, van een focus op modellen naar een focus op data. In de hydroinformatica is er eerder behoefte aan een verandering in en verbreding van het perspectief – hoe kijken we naar een probleem en op welke wijze proberen we dit op te lossen? Welke aspecten van het probleem, of de gezochte oplossing, nemen we als startpunt? Verschillende auteurs hebben dit benadrukt [4], [5], omdat de focus op modellen, die lange tijd dominant was in de hydroinformatica, te weinig aandacht besteedt aan de kwaliteit van de data (betrouwbaarheid, correctheid, toegankelijkheid) die nodig is om modellen zinvol te maken. Datakwaliteit en -beschikbaarheid zijn belangrijke uitdagingen voor de watersector, maar het is niet genoeg om

alleen daarop te focussen. Daarom stellen wij voor dat hydroinformatica zich richt op drie belangrijke punten: datagerichtheid, modelgerichtheid en risicogerichtheid (zie afbeelding 1).



*Afbeelding 1. Data-, model- en risicogerichte hydroinformatica als de drie-eenheid van het werkveld*

### ***Datagerichte hydroinformatica***

Datagerichte hydroinformatica zorgt ervoor dat alle gegevens die nodig zijn voor watergerelateerde beslissingen beschikbaar, toegankelijk en begrijpelijk zijn. Dit geldt voor zowel interne als externe data. De principes van FAIR data (vindbaar, toegankelijk, interoperabel en

herbruikbaar) zijn hierbij belangrijk, vooral omdat machines steeds vaker zelfstandig datasets benaderen. Data worden vaak opgeslagen in datawarehouses of datalakes, maar gedeelde dataruimtes (shared dataspace) [6], die het delen van datasets tussen verschillende organisaties faciliteren, worden ook steeds belangrijker. Daarnaast zijn traditionele methoden, zoals automatische datavalidatie en -correctie essentieel. Deze vormen de basis van een datagerichte benadering en zijn nog niet overal beschikbaar in de watersector.

Daarom is verdere ontwikkeling en onderzoek op dit gebied nodig. Maier et al. (2021) [7] beschrijven dat de manier waarop een dataset wordt opgesplitst voor selectie, kalibratie en validatie, de kwaliteit van het model kan beïnvloeden. Dit gebeurt meestal willekeurig, maar kan wel leiden tot een systematische fout in de verwerking (systematisch onnauwkeurige voorspellingen, in dit geval als de volledige dataset de variabiliteit van het systeem niet voldoende beschrijft) of een onjuiste evaluatie van het model. Om dit te voorkomen, kunnen geavanceerde datasplitsingsmethoden worden toegepast. Daarnaast is het belangrijk om de impact van verschillende methoden te evalueren, aandacht te besteden aan een gebalanceerde toewijzing van extreme waarden in de dataset aan de trainings- en evaluatiesubsets, en rekening te houden met de variabiliteit van de data.

### ***Modelgerichte hydroinformatica***

Met de opkomst van verschillende blackbox-technieken op basis van machine learning is het verleidelijk om 'de data voor zichzelf te laten spreken'. Dit betekent dat men probeert om statistische relaties uit grote hoeveelheden data te halen en deze als basis voor beslissingen te gebruiken. Hoewel dit waardevol kan zijn, is het voor de watersector niet voldoende om alleen op deze manier te werken. Hier zijn enkele belangrijke redenen voor:

1. Veel watervraagstukken hebben gezondheids- en milieuaspecten, waardoor het belangrijk is om de onderliggende relaties te begrijpen.
2. De omgeving verandert snel (denk aan klimaat, water en landgebruik),

waardoor het niet zeker is dat historische relaties in de toekomst blijven gelden.

3. De complexiteit van verschillende systemen is momenteel nog te groot voor kunstmatige intelligentie om volledig te begrijpen en te beheersen.

4. Cyberbeveiliging en de noodzaak om mensen betrokken te houden bij het proces, om bijvoorbeeld onverwachte en onwenselijke besluitvorming in uitzonderlijke situaties te voorkomen.

5. Een correlatie is een statistisch verband, wat nog niet wil zeggen dat er ook een oorzakelijk verband is.

Aan het derde punt wordt in de hydroinformatica hard gewerkt. De eerste twee punten vereisen dat we blijven werken aan procesmodellen en nieuwe technologische ontwikkelingen van buiten de watersector toepassen op de specifieke eisen van de watersector. Deze activiteiten vallen onder wat we modelgerichte hydroinformatica noemen.

### ***Risicogerichte hydroinformatica***

Uiteindelijk draait het niet alleen om data of modellen, maar om het nemen van goede beslissingen op basis van de best mogelijke informatie. Deze beslissingen zijn bedoeld om schaarse middelen efficiënt te gebruiken en om onvoorziene situaties in een snel veranderende wereld op te vangen zonder de kwaliteit van processen en dienstverlening te verminderen. Dit betekent dat gegevens moeten worden gecombineerd om ze betekenisvol te maken, de prestaties van systemen onder verschillende omstandigheden gemeten moeten worden, er relevante scenario's moeten worden gegenereerd en kansen moeten worden ingeschat. Soms moeten er ook scenario's worden bekeken zonder exacte kansschattingen.

Dit alles gebeurt het liefst binnen een kader dat helpt bij het nemen van beslissingen onder grote onzekerheid, waar waterbedrijven vaak mee te maken hebben. Risicogerichte hydroinformatica ontwikkelt denkkaders en besluitvormingsstrategieën voor deze beslissingen. Het bouwt voort

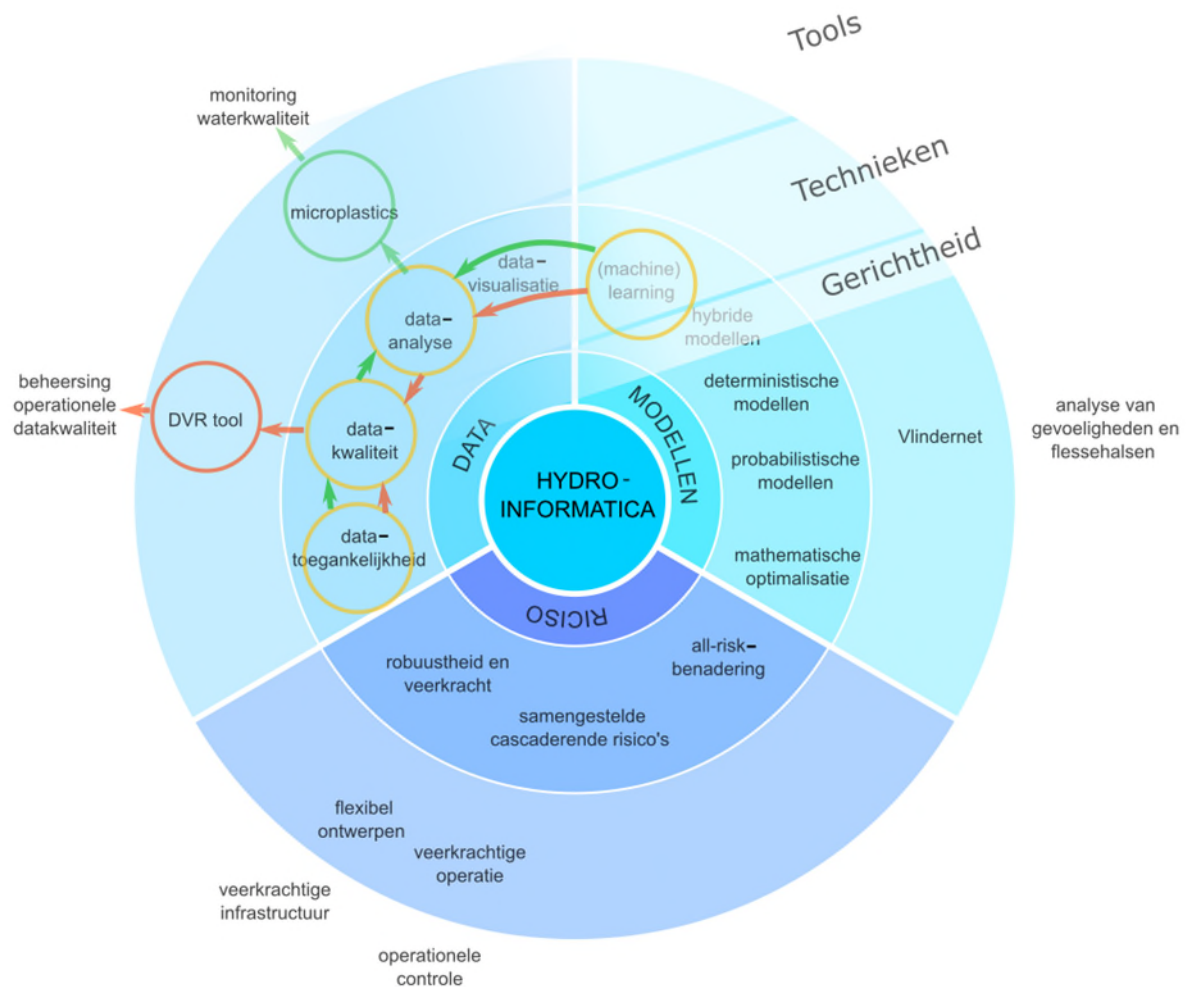
op de goede data en modellen die met data- en modelgerichte hydroinformatica worden ontwikkeld en toegepast.

## **Praktijkvoorbeelden**

### ***Datakwaliteit in datagedreven hydroinformatica***

Robuust gegevensbeheer en geautomatiseerde kwaliteitscontrole zijn essentieel voor datagerichte hydroinformatica en bedrijfsvoering. Bij afvalwaterbehandeling en drinkwaterproductie worden bijvoorbeeld sensoren gebruikt om gegevens te verzamelen over volumestromen, waterkwaliteitsindicatoren (zoals pH, totaal stikstof, zuurstof en geleidbaarheid) en temperatuur. Procestechnologen en waterbeheerders gebruiken deze gegevens om het proces optimaal te sturen.

Als afwijkende sensorgegevens direct worden gebruikt voor procesbesturing, kan dit tot verkeerde beslissingen leiden. Soms worden sensorgegevens handmatig gecontroleerd en aangepast, maar dit is erg tijdrovend en vaak onhaalbaar. Daarom is er behoefte aan geautomatiseerde systemen voor datavalidatie en correctie van foutieve sensorsignalen.



Afbeelding 2. Beheersing van datakwaliteit en monitoring van waterkwaliteit met data (oranje pijlen) respectievelijk modellen (groene pijlen) als eerste uitgangspunt. Relevante voorbeelden van concrete tools die voortkomen uit onderzoek bij KWR worden in de Tools-laag getoond

In het collectieve bedrijfstakonderzoek voor de drinkwatersector (BTO) [8] zijn geautomatiseerde procedures ontwikkeld voor datavalidatie en -correctie (DVR) [9]. Tijdens de datavalidatiestap worden anomaliedetectiemethoden uitgevoerd, waarbij elk sensorgegevenspunt wordt gelabeld als 'afwijkend' of 'niet afwijkend'. Dit omvat het detecteren van nul- of negatieve waarden (voor variabelen waarbij deze geen fysische betekenis hebben), drempelwaardeoverschrijdingen, flatlines (opeenvolgende observaties die gedurende een onrealistische periode dezelfde waarde hebben) en plotselinge pieken of dalingen (spike drops).

Voor het corrigeren van de gedetecteerde afwijkingen in tijdreeksen, worden methoden zoals het voorwaarts invullen (Forward filling), ARIMA-modellen (AutoRegressive Integrated Moving Average) en neurale netwerkmodellen gebruikt. Deze routines zijn verpakt in software, gedeeld en getest in samenwerking met waterbedrijven.

Afbeelding 2 toont deze benadering in het kader van de eerder beschreven perspectieven (oranje pijlen): aan de ene kant vanuit de datatoegankelijkheid (datagericht perspectief – over welke data hebben we het eigenlijk?) en aan de andere kant vanuit machinelearningtechnieken (modelgericht perspectief – welke machinelearningtechnieken kunnen we toepassen?), via de data-analyse waar deze in worden toegepast, naar de verbetering van de datakwaliteit, geïmplementeerd in de DVR-tool.

De DVR-procedure is toegepast in de datagestuurde besturing van de rioolwaterzuivering (RWZI) van Amsterdam-West. Dit systeem, eigendom van het waterschap Amstel, Gooi en Vecht en geëxploiteerd door Waternet, gebruikt DVR om defecte sensorsignalen te herstellen, voornamelijk anomalieën die ontstaan door problemen met gegevensverwerking en sensorfouten [10].

### ***Identificatie van microplastics***

Het eerste praktijkvoorbeeld richtte zich op datakwaliteit. In dit tweede voorbeeld ligt de focus op data-analyse, specifiek op de identificatie van microplastics. Microplastics vormen een groeiend milieuprobleem omdat ze wijdverspreid zijn en langzaam afbreken. Het identificeren van deze kleine plasticdeeltjes is belangrijk om de omvang van het probleem te begrijpen en effectieve oplossingen te vinden. Microplastics variëren in grootte, vorm, staat en materiaal, wat de identificatie lastig maakt. Voor de analyse van microplastics wordt vaak infraroodspectroscopie gebruikt. Hierbij worden deeltjes bestraald met een infraroodlaser, wat een karakteristiek infraroodspectrum oplevert. Dit spectrum wordt vergeleken met spectra in een database. Deze vergelijking is vaak moeilijk omdat afgebroken microplastics in het milieu sterk afwijkende



eigenschappen kunnen vertonen in vergelijking met nieuwe plastics. Machinelearningmodellen kunnen helpen bij de nauwkeurige identificatie en telling van microplastics in monsters [11].

Er zijn verschillende modellen getest en de resultaten tonen aan dat zelfs relatief eenvoudige modellen microplastics nauwkeurig kunnen identificeren. Zelfs als de deeltjes al door het milieu zijn aangetast en moeilijk te identificeren zijn met andere methoden. Afbeelding 2 toont ook deze benadering in het kader van de eerder beschreven perspectieven (groene pijlen). Aan de ene kant wordt de benadering vanuit de datatoegankelijkheid (datagericht perspectief) via de datakwaliteit getoond, aan de andere kant vanuit machinelearningtechnieken (modelgericht perspectief) naar de data-analyse die de analyse van microplastics en haar toepassing in de monitoring van waterkwaliteit mogelijk maakt.

### ***VlinderNET***

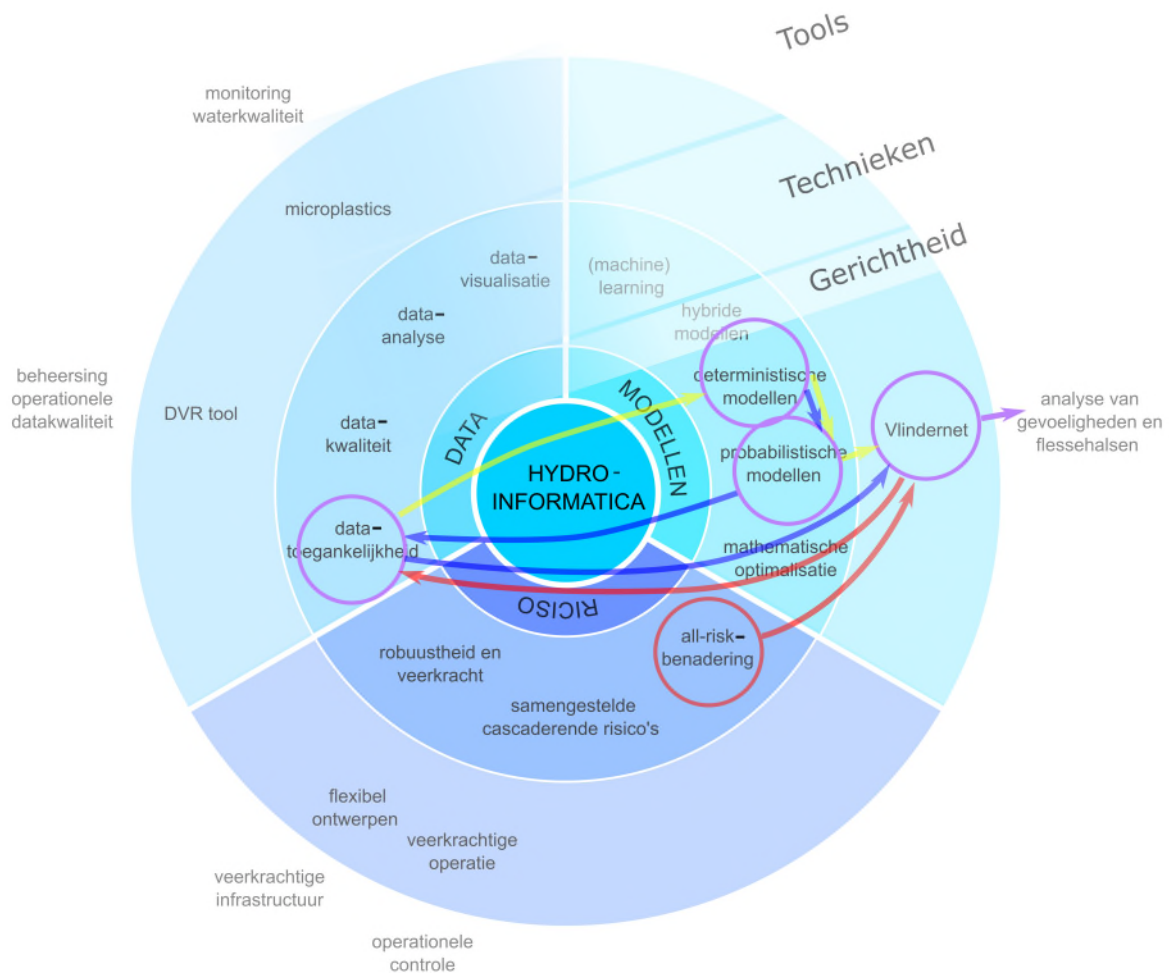
Het rekenprogramma VlinderNET [12] is een voorbeeld van een modelgerichte benadering in de hydroinformatica. Dit programma maakt gebruik van probabilistische modellering om drinkwaterdistributienetwerken te analyseren. In tegenstelling tot de gangbare deterministische benadering, waarbij netwerkeigenschappen en -gebruik als gegeven worden beschouwd (zoals leidingdiameters en watervraag), neemt VlinderNET alle onzekerheden in deze parameters mee. Dit omvat onjuiste historische gegevensregistratie, aangroei in leidingen en stochastische variaties in watervraag. Door deze onzekerheden te integreren in hydraulische berekeningen, genereert het programma voorspellingen met kansverdelingen voor hydraulische parameters als druk en volumestroom. Hiermee kan worden bepaald hoe zeker het is dat een aanpassing in het netwerk het gewenste effect zal hebben, zoals het behouden van voldoende druk in een wijk na het buiten gebruik stellen van een leiding.

VlinderNET is vanuit het modelgerichte perspectief ontwikkeld, waarbij bestaande methoden voor deterministische modellering zijn

gecombineerd met probabilistische benaderingen (afbeelding 3, blauwe pijlen). Tijdens het gebruik van de tool bleek echter dat de beschikbaarheid van data over onzekerheden in modelparameters een knelpunt vormt. Dit betekent dat er onderbouwde schattingen moeten worden gemaakt van bepaalde onzekerheden en dat er duidelijkheid moet komen over de onzekerheden die beter gekwantificeerd moeten worden.

Maar omdat er vanuit het modelperspectief is gestart met de veronderstelling van volledige databeschikbaarheid, is de tool dus voorbereid op een toekomst met een betere databeschikbaarheid. Door zijn generieke opzet vanuit het modelperspectief biedt VlinderNET ook ruimte voor nieuwe toepassingen en inzichten die niet vooraf zijn voorzien. Als het vanuit een datagericht perspectief was ontwikkeld, zou VlinderNET zijn gebouwd op basis van de beschikbare (en wellicht gemakkelijk te verkrijgen) data, gericht op de huidige situatie (afbeelding 3, gele pijlen).

Een risicocentrische benadering zou eerst de grootste risico's hebben geïdentificeerd die voortkomen uit onzekerheden, om vervolgens de benodigde model- en databehoeften vast te stellen (afbeelding 3, rode en gele pijlen). Dit zou een tool hebben opgeleverd die voorbereid is op een toekomst met betere databeschikbaarheid, maar vooral gericht zou zijn op het minimaliseren van huidige risico's. De flexibiliteit voor nieuwe, onvoorziene toepassingen en inzichten zou echter beperkter zijn geweest.

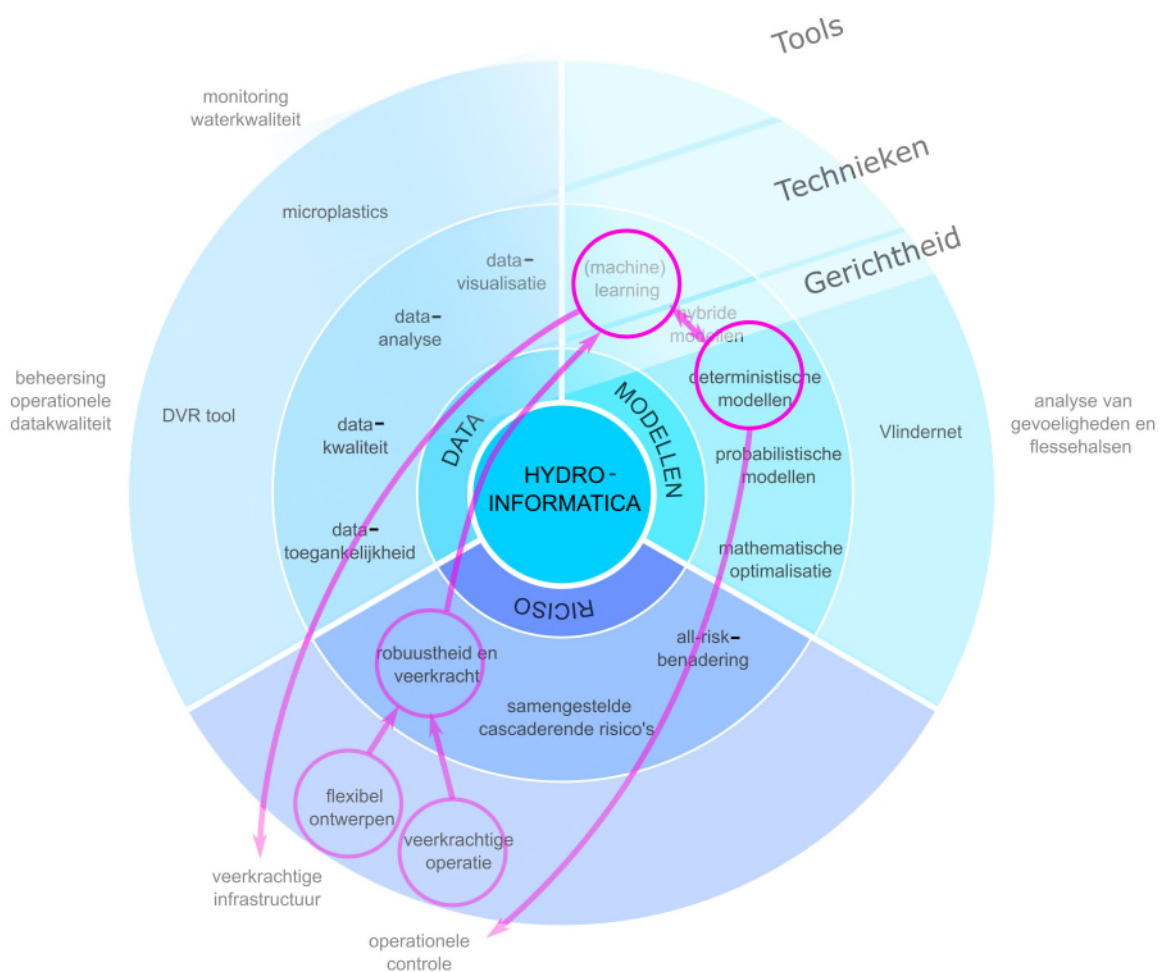


Afbeelding 3. Probabilistische modellering van drinkwaternetwerken met respectievelijk het risicoperspectief (rood+geel, paars), het modelperspectief (blauw, paars), en het dataperspectief (geel, paars) als uitgangspunt

## Water-Futures

Het project Water-Futures [13] richt zich op het ontwerpen van flexibele drinkwaternetwerken die zich kunnen aanpassen en uitbreiden naarmate de toekomstige omstandigheden veranderen. De traditionele benadering van het ontwerpen van infrastructuur met grote marges om toekomstige onzekerheden op te vangen, wordt hier vervangen door een slimmere aanpak. In plaats daarvan wordt gebruik gemaakt van reinforcement learning, waarbij het netwerk continu leert en zich aanpast aan veranderende omstandigheden. Zowel het ontwerp als de operationele aspecten worden in dit proces geoptimaliseerd.

Dit project illustreert hydroinformatica als een systeemfilosofie die is ontwikkeld om te reageren op risico's die veerkracht vereisen, mogelijk gemaakt door ICT-technologie (afbeelding 4). Het richt zich op diepe onzekerheden, zoals bevolkingsgroei, menselijk gedrag en klimaatverandering, die het moeilijk maken om infrastructuurbehoefte op lange termijn nauwkeurig te voorspellen. De flexibele benadering van Water-Futures [14], [15] stelt waterbedrijven in staat om adaptief en efficiënt te zijn in een snel veranderende wereld, zonder vast te zitten aan rigide ontwerpen die mogelijk snel verouderd raken.



*Afbeelding 4. Risicocentrische benadering in het project Water-Futures, die filosofieën over ontwerp en operatie naar modellen en machine learning concreet vormgeven.*

## **Afsluitende gedachten**

Het vernieuwde perspectief op hydroinformatica laat zien dat de gereedschapskist aanzienlijk is uitgebreid en dat de tools veel breder toepasbaar zijn dan voorheen gedacht. Het wordt duidelijk dat de keuze van het perspectief waarmee wordt begonnen cruciaal is voor het uiteindelijke resultaat. De drinkwatersector staat voor complexe uitdagingen op het gebied van waterkwaliteit en -beschikbaarheid, terwijl de watervraag blijft toenemen. Deze uitdagingen vragen om een integrale benadering, wat vaak leidt tot een complexiteit die moeilijk te bevatten is. Het is nu hoog tijd om meer diverse tools uit de gereedschapskist van hydroinformatica te benutten om deze uitdagingen aan te pakken.

## **Dankwoord**

Dit artikel is een resultaat van het project Water-Futures, dat is gefinancierd door de Europese Onderzoeksraad (ERC) in het kader van het Horizon 2020-programma voor onderzoek en innovatie van de Europese Unie (subsidieovereenkomst nr. 951424).

---

## **Samenvatting**

De drinkwatersector staat voor complexe uitdagingen, zoals waterkwaliteit, -beschikbaarheid en toenemende watervraag, die vanwege de complexiteit ervan een integrale benadering vereisen. Dit vraagt om het gebruik van diverse tools uit de hydroinformatica – feitelijk een uitgebreide gereedschapskist - waarbij de keuze van het startperspectief sterk het eindresultaat bepaalt. Vanwege dit laatste is het tijd voor een nieuw perspectief op hydroinformatica, waarin, afhankelijk van het doel en de toepassing, de data, de modellen, of het risico als uitgangspunt worden genomen.

---

## **REFERENTIES**

1. McComb, D. (2015). *The Data-Centric Revolution*. <https://tdan.com/the-data-centric-revolution/18780>, geraadpleegd op 7 maart 2024
2. Bhageshpur, K. (2016). *The Emergence of Data-Centric Computing*. <https://www.nextplatform.com/2016/10/06/emergence-data-centric->

[computing/](#), geraadpleegd op 7 maart 2024

3. Zha, D. et al. (2023). 'Data-centric ai: Perspectives and challenges'. In *Proceedings of the 2023 SIAM International Conference on Data Mining (SDM)*. Society for Industrial and Applied Mathematics.
4. Zolghadr-Asli, B. et al. (2024). 'A call for a fundamental shift from model-centric to data-centric approaches in hydroinformatics'. *Cambridge Prisms: Water*, Vol. 2.
5. Fu, G. et al. (2024). 'Making Waves: towards data-centric water engineering'. *Water Research*, Vol. 256, 121585.
6. IDS2024 International Data Spaces (2024). Terminology. <https://docs.internationaldataspaces.org/ids-knowledgebase/v/dataspace-protocol/overview/terminology>, geraadpleegd op 7 maart 2024
7. Maier, H.R. et al. (2023). 'On how data are partitioned in model development and evaluation: Confronting the elephant in the room to enhance model generalization'. *Environmental Modelling & Software*, 167.
8. KWR Water Research Institute (2024). *Bedrijfstakonderzoek voor waterbedrijven*. <https://www.kwrwater.nl/samenwerkingen/bedrijfstakonderzoek-voor-waterbedrijven>
9. Seshan, S. et al. (2024). 'AI enabled data validation tool: from development to deployment'. *KWR Water Research Institute*, rapport BTO 2024.004
10. Seshan, S. et al. (2024). 'LSTM-based autoencoder models for real-time quality control of wastewater treatment sensor data'. *Journal of Hydroinformatics*; 26 (2): 441–458. doi: <https://doi.org/10.2166/hydro.2024.167>
11. Bäuerlein, P.S. et al. (2023). 'Nauwkeurige identificatie polymeren en microplastics met machine learning'. *H2O Water Matters*, juni 2023.
12. Morley, M.S. et al. (2023). *VlinderNET – a tool for Probabilistic Hydraulic Water Distribution Modelling and Visualization*. *EGU General Assembly 2023*, Vienna, Austria, EGU23-17062, <https://doi.org/10.5194/egusphere-egu23-17062>
13. Savic, D.A. et al. (2022). *Long-Term Transitioning of Water Distribution Systems: ERC Water-Futures Project*. In *Proceedings of the 2nd International Joint Conference on Water Distribution System Analysis (WDSA) & Computing and Control in the Water Industry (CCWI)*, Valencia, Spain.
14. Tsiami, L. et al. (2024). 'Staged Design of Water Distribution Networks: A Reinforcement Learning Approach'. *Geaccepteerd voor een presentatie op de derde gezamenlijke WDSA&CCWI-conferentie*, Ferrara, Italië.
15. Zanutto, D. et al. (2024). 'Incorporating flexibility in the long-term design of water distribution systems using operational variables'. *Geaccepteerd voor een presentatie op de derde gezamenlijke WDSA&CCWI-conferentie*, Ferrara, Italië