

---

# Aquapodium promovendis

---

## Hydrologische proceskennis in 'data-driven' voorspelling van rivierafvoeren

### *Inleiding*

Toen ik laatst '2001: A Space Odyssey', de tijdloze science-fictionklassieker van regisseur Stanley Kubrick en schrijver Arthur C. Clarke, weer eens zag, werd ik me wederom angstvallig bewust van het vertrouwen van de mens in de door hem geschapen technologie. Niet dat wij al hoeven te vrezen voor moordende, zelfbewuste computers zoals HAL9000, maar een grote afhankelijkheid brengt altijd risico's met zich mee.

De film uit 1968 is duidelijk geïnspireerd door het ambitieuze onderzoeksveld van de Kunstmatige Intelligentie (KI). Onderzoek op het gebied van de KI heeft inmiddels vruchten afgeworpen in de vorm van enkele nuttige praktische technieken, waaronder Artificial Neural Networks (ANNs). ANNs worden heden ten dage toegepast op onderwerpen variërend van economische voorspelling, spraakherkenning, medische data-analyse, militaire objectherkenning, en, u raadt het al, hydrologische modellering.

### *ANNs als neerslag-afvoermodellen*

Nadat ik in 2003 afstudeerde aan de Technische Universiteit Delft, als gedeeld student van de Secties Civieltechnische Informatica en Hydrologie & Ecologie, ben ik bij de laatstgenoemde (tegenwoordig: Sectie Waterhuishouding) een promotieonderzoek gestart onder begeleiding van Tom Rientjes en promotor Huub Savenije. In navolging van mijn afstudeerwerk over neerslag-afvoermodellering met ANNs (De Vos, 2003) ben ik nu onder andere bezig met een gede-

tailleerd onderzoek naar de beperkingen en mogelijkheden van deze modelleertechniek.

ANNs zijn geïnspireerd op de werking van het zenuwstelsel en brein. Ze bestaan uit eenvoudige rekenkundige elementen (neuronen) die niet-lineaire transformaties verrichten, maar welke onderling dicht verbonden zijn (zie figuur 1). De invoer wordt laag voor laag via de neuronen getransformeerd naar uitvoer van het netwerk, waarbij alle tussenliggende stappen worden vermenigvuldigd met gewichten in de connecties. Deze gewichten kunnen vervolgens worden geoptimaliseerd door automatische algoritmen die de output van het netwerk proberen overeen te laten komen met gemeten waarden. Voor een uitstekende behandeling van ANN-theorie en voorbeelden van algoritmen raad ik u het boek van Simon Haykin (1999) aan.

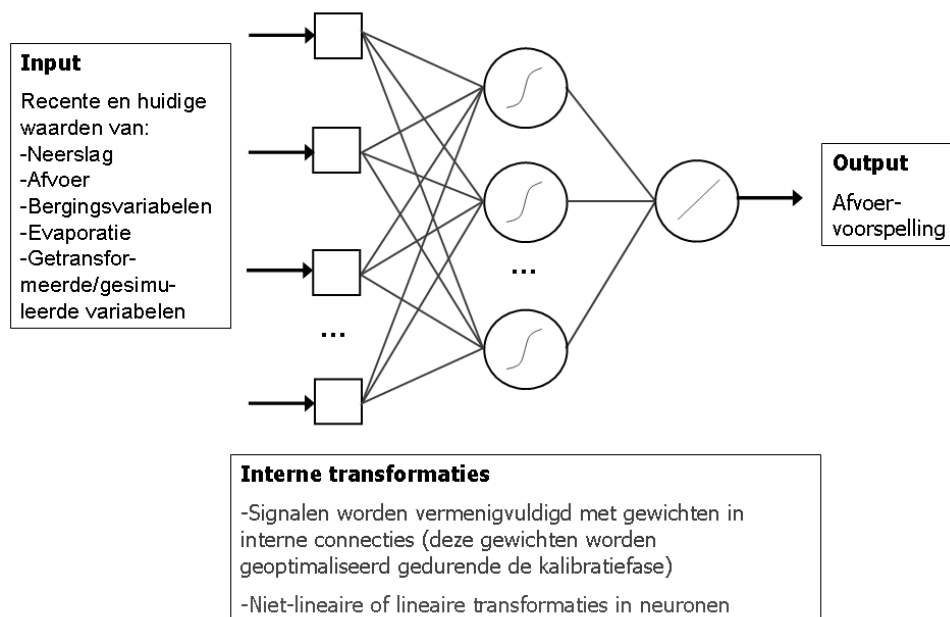
Figuur 1 toont tevens hoe ANNs gebruikt kunnen worden voor het simuleren van de respons van een stroomgebied op meteorologische impulsen en daarmee het voorspellen van rivierafvoeren. Voorbeelden van relevante inputvariabelen zijn:

- recente waarden van regenval op één of meerdere locaties in een stroomgebied,
- recente waarden van verdamping op één of meerdere locaties in een stroomgebied,
- recente waarden van lokale of bovenstroomse afvoeren,
- gemeten of gesimuleerde waarden van bodemvocht of grondwater.

Tevens kunnen getransformeerde vormen van bovenstaande variabelen gebruikt worden, zoals bijvoorbeeld voortschrijdende gemiddelden. Een van de grote voordelen van ANNs is derhalve hun flexibiliteit: als de modelleur een interpolateerbare informatiebron weet te verzinnen die in een tijdreeks kan worden uitgedrukt, kan deze gebruikt worden als input! Een goed voorbeeld van een dergelijke informatiebron is temperatuur. Indirect bevat temperatuur veel informatie over verliezen in de massabalans in de vorm van verdamping en

transpiratie. Daarnaast bevat temperatuursdata duidelijke vormen van periodici-

teit, die ook in afvoerdata te herkennen



**Figuur 1:** Illustratie van een ANN als neerslag-afvoermodel.

zijn. Het mooie van ANNs is dus dat dit soort relaties niet expliciet hoeven te worden gemaakt, maar het netwerk zelf op zoek gaat naar deze informatie. Als output van het model wordt een afvoervoorspelling met een bepaalde tijdshorizon gebruikt.

De configuratie van het ANN in termen van aantal neuronen is niet theoretisch a priori te bepalen en wordt over het algemeen via trial-and-error gevonden. Hoe meer neuronen, hoe meer parameters en dus hoe meer vrijheidsgraden in het model. Voor dit type modeltoepassingen biedt een trechtersvormige structuur met een afnemend aantal neuronen van input naar output vaak genoeg, maar niet teveel vrijheidsgraden voor een goed werkend en betrouwbaar model. De transformaties in de neuronen, welke ANNs hun niet-lineaire karakter geven, zijn over het algemeen eenvoudige niet-lineaire functies zoals de tangens hyperbolicus.

In feite voert een ANN dus slechts een

niet-lineaire transformatie van de ene getallenreeks in een andere uit; het is aan de modelleur om te zorgen dat de transformatie fysische betekenisvol is. Aan de andere kant: ANNs zijn flexibel en eenvoudig te gebruiken, en bovendien vereisen ze slechts korte rekentijden. Voor sommige hydrologische problemen is deze techniek dan ook bruikbaar dan het ontwikkelen en gebruiken van een groot en/of complex model. Zie het literatuuroverzicht door de ASCE (2000) voor een goed overzicht van succesvolle studies en toepassingen van ANNs in hydrologie.

#### *Kalibratie van een ANN*

Een belangrijk aspect van het modelleren met ANNs is de kalibratie van de gewichten van het netwerk. De algoritmes die hiervoor gebruikt worden nemen de fout van het model ten opzichte van metingen, rekenen

de bijdrage van elk gewicht aan deze fout uit (het zogenaamde proces van 'backpropagation'), en op basis hiervan wordt een correctie van de gewichtenset uitgevoerd. De bovengenoemde fout van het model is over het algemeen een doelfunctie, welke een bewerking is van het verschil tussen modelschatting en meting. Een voorbeeld is de Mean Squared Error (MSE), welke grote afwijkingen zwaarder doet wegen en daardoor impliciet meer gericht is op het kalibreren van piekafvoeren.

De algoritmes die worden gebruikt in ANN modelleren zijn veelal gebaseerd op eerste- of tweedegraads gradienten van deze fout. Voorbeelden van de eerste groep zijn zogenaamde 'steepest-descent'-algoritmen die zoeken in de richting van de steilste neerwaartse helling. Een goed voorbeeld van de tweede groep is het populaire Levenberg-Marquardt-algoritme. Tegenwoordig worden ook globale optimalisatie-algoritmen gebaseerd op evolutionaire methoden (bijvoorbeeld Genetische Algoritmes, zie Goldberg (1989)) regelmatig gebruikt voor kalibratie van ANNs.

Als een algoritme de gewichten van een model zo aanpast dat de kalibratiedataset uitstekend wordt benaderd, betekent dit dan ook dat de ANN een goed model is? Niet noodzakelijk, want net als in andere modelbenaderingen willen we weten of ons model de voorbeelden uit de kalibratie kan generaliseren naar nieuwe situaties en metingen. Om deze reden wordt vaak een deel van de dataset ter validatie gebruikt. Een deel van deze validatiedata wordt gebruikt als cross-validatie: tegelijk met het veranderen van de gewichten tijdens de kalibratie, wordt in de gaten gehouden wanneer de fout op deze cross-validatiedata stijgt. Een algoritme kan namelijk zover gaan dat de ANN ook nietszeggende relaties in de data 'leert', zoals bijvoorbeeld de ruis in de kalibratiedata. Wanneer de kalibratie wordt stopgezet op het moment dat de cross-validatiefout toeneemt, is de kans dat

het model zijn generaliserend vermogen verliest door een te grote fixatie op de kalibratiedata theoretisch kleiner. Ten slotte worden, zoals gebruikelijk bij modeltoepassingen, onafhankelijke validatiedata gebruikt om het ANN-model te valideren.

### *Proceskennis in 'data-driven' modelleren*

ANN neerslag-afvoermodellen kunnen worden beschouwd als black-boxmodellen, welke een hydrologisch systeem simuleren aan de hand van de invoer en uitvoer en niet gebaseerd zijn op proceskennis. De kritiek van veel hydrologen is dan ook dat de combinatie van ondoorzichtigheid, flexibiliteit en het automatisch aanpassen aan informatie van dit soort modelbenaderingen leidt tot onrealistische en onbetrouwbare simulaties. En inderdaad: als er niet met voldoende kennis, inzicht in processen en ervaring wordt gemodelleerd en de modellen niet kritisch genoeg worden beschouwd, bestaat hetzelfde soort dreiging als waar HAL9000 symbool voor staat.

Echter, in plaats van de black-boxbenadering is het correcter om te spreken over een zogenaamde 'data-driven' modelbenadering. Hierbij beoogt de modelleur het informatiegehalte van de data optimaal te benutten zonder daarbij de toepassing van proceskennis uit te sluiten. Met behulp van hydrologisch inzicht bij het modelleren met ANNs kan namelijk een vorm van consistentie met de werkelijkheid worden afgedwongen. Hierdoor wordt het model betrouwbaarder en mogelijk ook accurater, en bovendien vergroot het de mogelijkheid dat het model ons iets leert over het functioneren van het werkelijke systeem. Een voorbeeld van wat mijns inziens ook in dit data-driven kader past, is het werk van Jos von Asmuth en collega's in stochastische tijdreeksanalyse (bijv. Asmuth e.a., 2002).

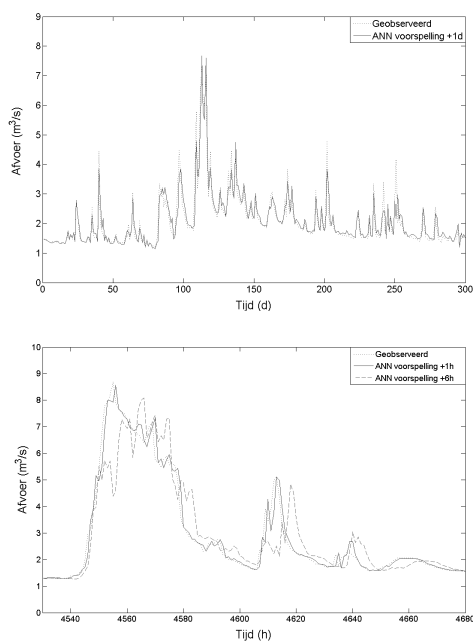
Een voorbeeld van de toepassing van proceskennis is het gebruik van toestands-

variabelen gerelateerd aan berging in een stroomgebied als input voor het model (zie De Vos en Rientjes (2005) en De Vos e.a. (2005)). Op deze manier is een belangrijke informatiebron beschikbaar voor het model, waarmee het drempelgedrag van afvoerprocessen zoals 'saturation overland flow' beter kan worden benaderd. Ook de simulatie van de basisafvoer, welke sterk is gerelateerd aan deze berging, kan derhalve worden verbeterd. De toestandsvariabelen hoeven niet per se metingen te zijn, maar kunnen ook gesimuleerd worden zoals wij gedaan hebben met een eenvoudig conceptueel model voor bodemvocht (De Vos en Rientjes, 2005). Met dit soort ontwikkelingen verwaagt de grens tussen conceptuele modellen en data-driven modellen, en het vakgebied van hydrologisch modelleren zal hierdoor wel eens een stuk minder overzichtelijk kunnen worden dan het voorheen was.

Een andere manier waarop wij trachten proceskennis in ANNs te implementeren, en welke complementair is aan het hierboven genoemde gebruik van toestandsvariabelen, is door de kalibratieprocedure aan te passen. Vaak wordt de nadruk gelegd op de algoritmes die de automatische optimalisatie uitvoeren. Echter, het belang van passende doelfuncties is tevens groot. Aangezien de afvoerrespons die we willen benaderen een aggregatie van meerdere complexe en variabele processen is, kan serieus betwijfeld worden of een ééndimensionale test zoals de MSE afdoende is om een ANN neerslag-afvoermodel kritisch te evalueren. Ik deel dan ook de mening van Gupta e.a. (1998) dat kalibratie van hydrologische modellen inherent 'multiobjective' is. Om deze reden testen wij momenteel multiobjective algoritmen voor de optimalisatie van ANNs.

Een van de tekortkomingen van huidige methoden is dat over het algemeen alleen naar het verticale verschil tussen de gemeten en gesimuleerde afvoerkromme wordt gekeken, terwijl de horizontale afwijkingen

ook van groot belang zijn (denk aan de correcte timing van afvoerpieken). ANNs kunnen moeilijkheden ondervinden met het timen van afvoeren omdat vorige afvoeren vaak als input van het model worden gebruikt. In het slechtste geval produceert het netwerk zelfs een voorspelling die niet veel meer is dan de huidige waarde welke verschoven is in de tijd (zie figuur 2). Daarom is het belangrijk dat aspecten zoals de timing en vorm van de afvoerkromme zwaarder worden meegenomen in de kalibratie van een ANN-model (zie ook De Vos en Rientjes, 2005).



**Figuur 2:** Voorbeelden van simulatieresultaten van ANN-modellen: (a) dagelijkse en (b) uurlijkse voorspellingen gebaseerd op data uit het Jekerstroomgebied in België. De dagelijkse voorspellingen zijn, op het oog en volgens gehanteerde doelfuncties, goed. De uurlijkse zijn volgens dezelfde criteria minder goed en worden vooral geplaagd door timingsfouten.

De huidige fase, ongeveer halverwege mijn promotietraject, geeft de mogelijkheid tot reflectie over wat ik tot nu toe geleerd en gedaan heb en wat er nog moet gebeuren om de promotie met een goed gevoel achter me te kunnen laten. Na de succesvolle publicatie van en positieve respons op mijn eerste peer-reviewed artikel (De Vos en Rientjes, 2005) voel ik me gesterkt in het vertrouwen dat ik op de goede weg ben. Daarnaast ben ik inmiddels in het bezit van goed gereedschap (lees: software) om mijn modellen makkelijk toe te passen op andere casussen. Wie weet, mogelijk zelfs wel op andere onderzoeksterreinen dan de hydrologie. Dit geldt voor mij persoonlijk dan ook als een groot voordeel van de flexibiliteit van ANNs en het helpt me breder georiënteerd te blijven dan hydrologie alleen. Bovendien is er nog een ander onderwerp dat binnen mijn promotieonderzoek gepland staat, en waar ik mij de komende jaren meer op wil richten: het onderzoeken van de toepassing van data-assimilatietechnieken in hydrologische modellering. Omdat ook deze technieken trachten optimaal gebruik te maken van hydrologische data in een modelomgeving voor afvoervoorstellingen, probeer ik ze in mijn werk samen met ANNs in één theoretisch kader te plaatsen.

Zoals ik hierboven duidelijk gemaakt hoop te hebben, ben ik van mening dat een slimme data-driven modelleur profiteert van zowel state-of-the-art technieken als proceskennis. Met name dat laatste maakt de toepassing van modellen en waardevolle terugkoppeling en daaropvolgende verbeteringen mogelijk. Wetende dat juist de combinatie van proceskennis en KI-technieken mijn modellen maakt tot wat ze zijn, voel ik mij niet opgejaagd zoals de menselijke protagonist uit 2001: A Space Odyssey, maar zie ik de computer eerder als mijn beste vriend.

- ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000)** Artificial neural networks in hydrology, II – Hydrologic applications; in: *J. Hydrol. Eng.*, vol 5 (2), pag 124–137.
- Asmuth, J.R. von, M.F.P. Bierkens en K. Maas (2002)** Transfer function-noise modeling in continuous time using predefined impulse response functions; in: *Water Resources Research*, vol 38 (12), 1287.
- Goldberg, D.E. (1989)** Genetic Algorithms in search, optimization, and machine learning; Addison-Wesley, Reading, Massachusetts.
- Gupta, H.V., S. Sorooshian en P.O. Yapo (1998)** Toward improved calibration of hydrologic models: Multiple and non-commensurable measures of information; in: *Water Resources Research*, vol 34 (4), pag 751–763.
- Haykin, S. (1999)** Neural Networks: A Comprehensive Foundation; Prentice-Hall, Upper Saddle River, New Jersey.
- Vos, N.J. de (2003)** Rainfall-Runoff Modelling Using Artificial Neural Networks; afstudeerrapport, Technische Universiteit Delft.
- Vos, N.J. de en T.H.M. Rientjes (2005)** Constraints of artificial neural networks for rainfall-runoff modelling: trade-offs in hydrological state representation and model evaluation; in: *Hydrology and Earth System Sciences*, vol 9, pag 111–126.
- Vos, N.J. de, T.H.M. Rientjes, L. Pfister (2005)** Groundwater levels as state indicator in rainfall-runoff modelling using Artificial Neural Networks; in: B. Makaske en A.G. van Os (red) *Proceedings of the NCR-days 2004*; Research for managing rivers, Netherlands Centre for River studies, Delft, pag 70–73.