



Verbeterde kunstmatige neurale netwerken breed toepasbaar in integraal waterbeheer

J. VAN DER WIELEN, WITTEVEEN+BOS

F. SCHULZE, WITTEVEEN+BOS / TU DELFT-CITG

M. BERGMAN, WITTEVEEN+BOS

In de wereld van integraal waterbeheer spelen kunstmatig neurale netwerken een steeds belangrijkere rol. Samen met de vergroting van toepassingsgebieden (drinkwaterverbruik, zoutgehaltemodellering, stadsdruk in waterleidingnet) staat de ontwikkeling van deze netwerken op het gebied van automatisering, kalibratieproces en modelstructuurontwikkeling zelf ook niet stil. Vernieuwende technieken zorgen voor hogere mate van nauwkeurigheid en verbeterde inzichten in deze relatief nieuwe statistische modelleringstechniek. Naast de modelleringmogelijkheid van het kunstmatig neurale netwerk kan het ook steeds vaker gebruikt worden als kwaliteitstoetsing voor (bestaande) lineaire en fysische modellen.

De statistische modelleringstechniek kunstmatig neurale netwerken (KNN)³⁾ wordt, vanwege de ruime toepasbaarheid, steeds vaker gebruikt in de vele disciplines van integraal waterbeheer. Terwijl de mogelijkheden van KNN in vergelijking tot conventionele lineaire modellen en zelfs fysische modellen binnen de waterwereld steeds duidelijker worden is de techniek van KNN zelf nog steeds onderhevig aan verdere ontwikkeling. De potentie van KNN, de universele modelleringstechniek die door kalibratie alle relevante (fysische) relaties tussen willekeurige verklarende en te verklaren grootheden kan 'leren', blijkt ook in praktijk steeds reëler. Buiten de poten-

tiële kwaliteit van het KNN is ook de simulatierekening ten opzichte van fysische modellen een belangrijke factor. Een eenmaal gekalibreerd kunstmatig neurale netwerk heeft een dermate eenvoudige structuur dat praktisch momentaan gesimuleerd kan worden, in tegenstelling tot vaak lange reketijden bij veel fysische benaderingen.

Een KNN is een wiskundige techniek die door leren van voorbeelden zelf relaties kan leggen tussen een willekeurig aantal verklarende en te verklaren factoren zonder daarbij het systeem zelf in beschouwing te nemen⁷⁾. De structuur van een KNN is opgebouwd uit

een aantal basiselementen, naar analogie van het menselijk brein neuronen genoemd, welke via verbindingen aaneengeschakeld zijn tot een netwerk³⁾ (zie afbeelding 1). Een neuron bestaat uit een sommatie van invoeren Σ en een interne transferfunctie f . Aan elke verbinding tussen neuronen wordt een 'gewicht' w toegekend, een maat voor de invloed van het ene neuron op het andere. De voeding van het netwerk wordt gevormd door de geselecteerde verklarende variabelen, P_1 t/m P_q en de uiteindelijke output door een sommatie van de neuronen uit de laatste laag. Door het aanpassen van gewichten binnen of tussen neuronen met behulp van een trainingsalgoritme en vergelijking met de werkelijkheid wordt het KNN 'getraind'.

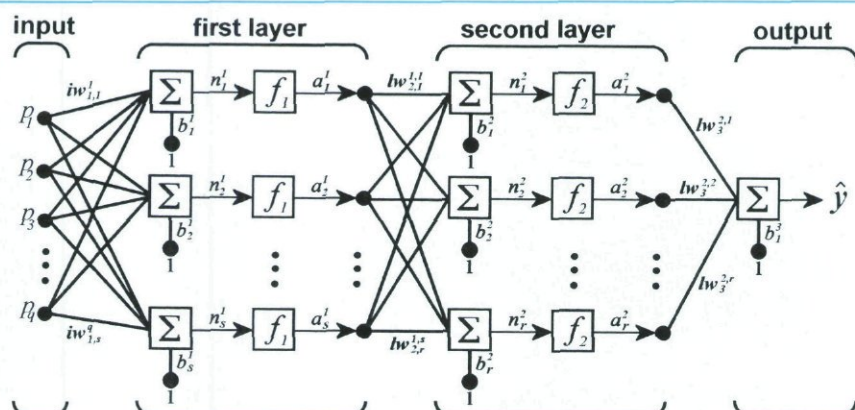
Dit artikel beschrijft de nieuwste ontwikkelingen op het gebied van kunstmatig neurale netwerken en presenteert de universele toepasbaarheid en kwaliteit van deze 'nieuwe generatie KNN'⁸⁾ op een drietal uiteenlopende toepassingsgebieden in integraal waterbeheer. Tevens wordt een vergelijking met de 'conventionele' modellen gemaakt.

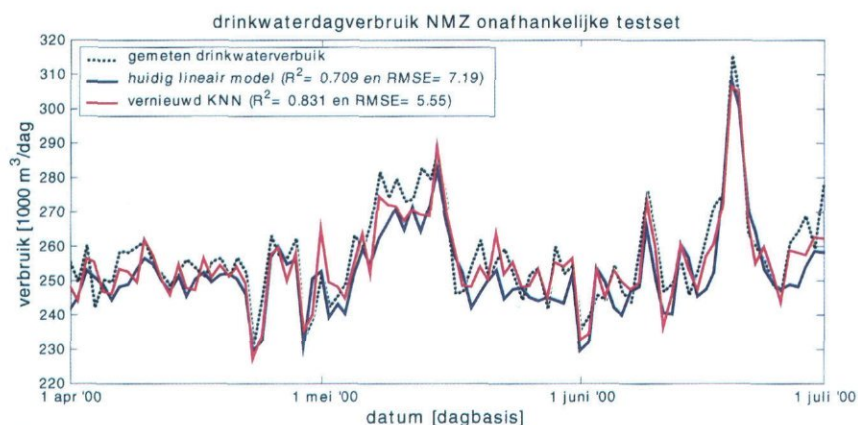
Meest recente ontwikkelingen

Modelleren met KNN is relatief nieuw; de methode is, in 1958 naar aanleiding van het menselijk brein ontwikkeld. De toepassing van KNN op het gebied van integraal waterbeheer is nog jonger⁷⁾. Het is duidelijk dat de in potentie universele modelleringstechniek nog lang niet is uitontwikkeld en voortdurend onderhevig is aan vernieuwingen en verbeteringen. Hierdoor wordt het KNN nog te vaak gezien als een minder betrouwbare en minder beproefde techniek in vergelijking met lineaire of fysische modellering, hoewel men van de grote potentie wel bewust is. Juist op het gebied van doorzichtigheid en verbeteringen van de techniek en de vergelijking met lineaire modellen wordt veel onderzoek gedaan.

Een lineair probleem is perfect te beschrijven door een lineair model ook al is ruis aanwezig. Wanneer echter het lineaire model niet optimaal presteert, rijst de volgende vraag: is de rest alleen ruis of zit er nog een component in welke door niet-lineaire relaties of zelfs interacties van de verklarende grootheden afhangt. Vanuit deze situatie is het onmogelijk af te tasten wat de maximaal mogelijke prestatie is. Het KNN daarentegen leert deze relaties zelf aan waardoor in potentie de maximaal haalbare prestatie behaald wordt. Presteert het KNN identiek aan het lineaire model, dan kan alsnog voor het naar het blijkt optimale lineaire model gekozen worden⁸⁾. Bij een duidelijke meerwaarde in prestatie van het KNN kan nu met terugwerkende kracht gezocht worden naar de belangrijkste factoren welke zorgdra-

Afb. 1: De algemene structuur van een kunstmatig neurale netwerk^{3),8)}





Afb. 2: Verschillende voorspellingsmodellen voor het drinkwaterverbruik per dag in de regio's noord, midden en zuid.

gen voor deze meerwaarde en naar de specifieke (kwantitatieve) relatie. Hierdoor is het mogelijk geworden het 'black-box' karakter van het KNN doorzichtiger te maken en het KNN model toetsbaar te maken op fysische wetmatigheden⁸.

Een tweede ontwikkeling is een vernieuwde modelleringmethode met behulp van KNN. De huidige kunstmatig neurale netwerken blijken een aantal belangrijke knelpunten te bevatten, onder andere in de softwaremogelijkheden. Ten eerste is het zeer lastig een optimale modelstructuur voor het specifieke systeem te bepalen. Nog weinig houvast bestaat bijvoorbeeld voor de relatie van het aantal neuronen, het aantal lagen, mogelijke verbindingen en de transferfuncties in relatie tot het modelleringsprobleem. Hierdoor komt het modelleren met KNN in de praktijk neer op de ervaring van de modelleur. Voor het optimaal ontwerpen van een probleemspecifieke structuur is een methode van gestructureerd modelontwerp ontwikkeld⁸ welke op geautomatiseerde wijze de invloeden van de verschillende factoren op de kwaliteit van de uitvoer onderzoekt en hieruit de optimale modelstructuur bepaalt. Verder blijken de meest gebruikte trainingsmethoden, voor het kalibreren van een KNN op het benaderen en/of voorspellen van complexe systemen, in veel gevallen niet in staat de potentiële prestatie te behalen. Voor het zuiverder trainen van een KNN is een hybride trainingsalgoritme ontwikkeld⁸ die de sterke kanten van een aantal algoritmen combineert en een oplossing biedt voor de knelpunten.

Toepassingen in de praktijk

Drinkwaterverbruik Noord, Midden en Zuid

In het onderzoek 'Prognose dagverbruik drinkwater nadere uitwerking alternatieven'⁵ in opdracht van PWN Waterleidingbedrijf Noord-Holland is een verbetering aangebracht in de huidige lineaire verbruiksmodellen voor het drinkwaterdagverbruik in het betreffende gebied in Noord-Holland (inclusief Texel) door

het ontwikkelen van KNN. Er is een prestatieverbetering van bijna twaalf procent behaald door het optimaliseren van de verklarende variabelen (zie afbeelding 2). De toegevoegde waarde van het vernieuwde KNN ten opzichte van een lineair model op basis van de geoptimaliseerde set van verklarende variabelen bleek echter niet zo groot. De fractie verklarende variantie* (R^2) van het KNN (83%) verschilt maar weinig van het lineair model (82%). Puur alleen de prestatie van het lineaire model zou aanleiding hebben gegeven tot de gedachte dat nog andere relaties van de verklarende grootheden een relevante rol kunnen spelen voor meer prestatieverbetering. Door het KNN is echter aangetoond dat het lineaire model nagenoeg maximaal verklaard. Naast de modellering van het drinkwaterdagverbruik in de regio's noord, midden en zuid is voor Texel eenzelfde vergelijking gemaakt. Hierbij gaf het KNN, zeker op het modelleren van extremen, een duidelijkere verbetering.

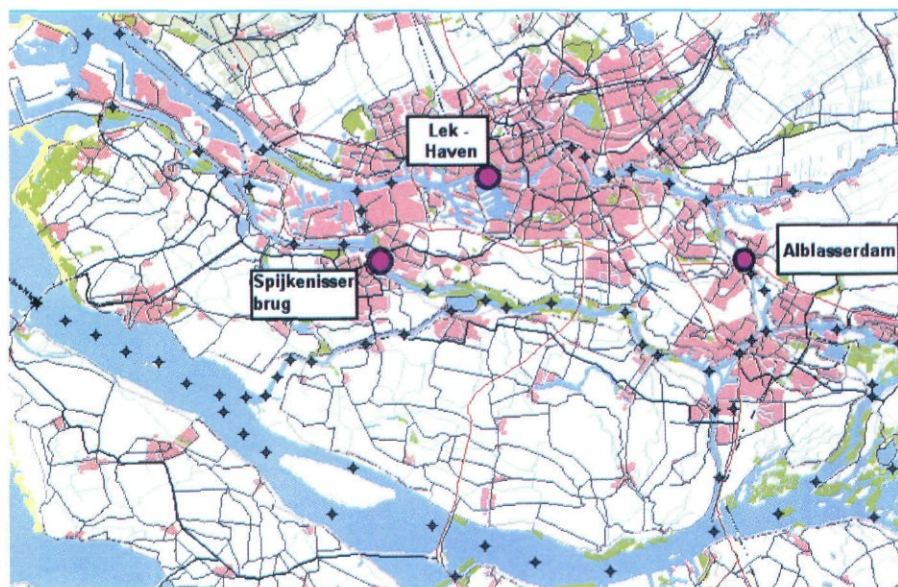
Chlorideconcentratie noordelijk deltabekken
In het kader van de implementatie van De

Kier (het in de toekomst op een kier zetten van de Haringvlietssluisen. Hierover verschijnt binnenkort een uitgebreid artikel in dit blad.) is in opdracht van Rijkswaterstaat (RIKZ) onderzoek gedaan naar het beschrijven van de zoutgehalten in de huidige situatie op drie meetlocaties in het noordelijk deltabekken (Spijkenisserbrug, Lekhaven en Alblasterdam) (zie afbeelding 3), op basis van de verschillende omgevingsfactoren als rivierafvoeren, zeezoutconcentratie, waterstanden en windsnelheid.

Voor de verschillende meetlocaties zijn, na uitgebreid vooronderzoek, een set van 15 verklarende grootheden voor locatie Alblasterdam bepaald en 27 voor de overige locaties. Voor deze locaties zijn zowel een additief lineair, een conventioneel KNN als een vernieuwd KNN ontwikkeld. Hierbij is voor locatie Alblasterdam voor een set van onafhankelijk testweken een verbetering behaald van bijna 20 procent ten opzichte van de lineaire modellen (zie ook afbeelding 4).

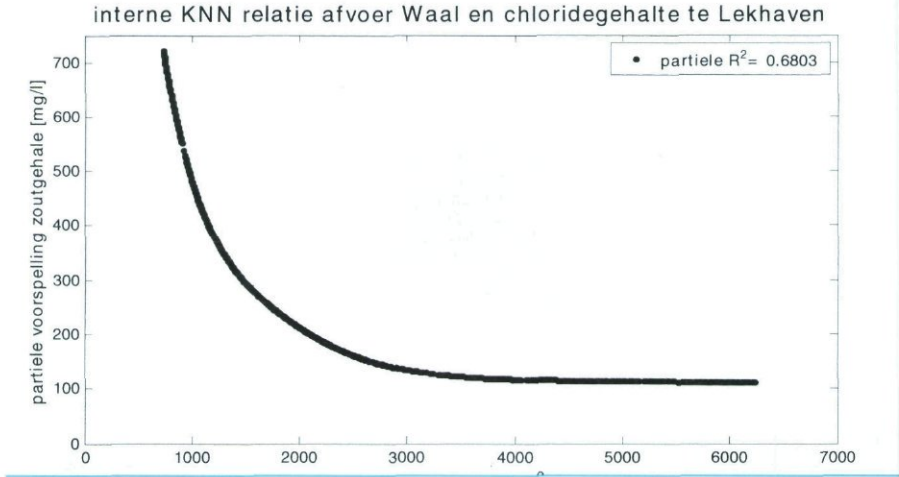
Naast de mogelijkheid van een optimaal KNN om zelf, buiten lineaire relaties, ook niet-lineaire relaties en zelfs relevante interacties tussen grootheden te kunnen detecteren is een KNN ook veel beter in staat tot het modelleren van extremen situaties. Afbeelding 4 toont in het begin van week 35 een tweetal zeer grote pieken waar het lineaire model aan voorbij gaat. Omdat deze situatie slechts op een paar plekken voorkwam, de oorzaak ervan vooraf fysisch moeilijk was te bepalen en het door het lineaire model niet wordt voorspeld, blijft de vraag of deze situatie te wijten is aan meetonnauwkeurigheden of dat het toch verklaarbaar is door de geselecteerde verklarende grootheden. Het KNN geeft het doorslaggevende antwoord door een goede voorspelling voor deze situaties te

Afb. 3: De meetlocaties in het noordelijk deltabekken.



generen. Wanneer ook het KNN aan deze extremen voorbij was gegaan, is met grote zekerheid te stellen dat deze situaties niet door de geselecteerde variabelen verklaard kunnen worden en wellicht te wijten zijn aan (meet)onnauwkeurigheden. Wel is opvallend dat het conventionele KNN weliswaar de extremen detecteert, maar niet zo fraai benadert als het vernieuwde KNN. Ook volgt het vernieuwde KNN beter de lineaire patronen. In tabel 1 is een vergelijking in prestatie te zien tussen lineair, conventioneel KNN en de vernieuwde KNN.

Afbeelding 5 geeft een voorbeeld van een door het KNN gedetecteerde niet-lineaire relatie op meerlocatie Lekhaven tussen de afvoer van de



Afb. 4: De verschillende voorspellingsmodellen voor het chloridegehalte te Alblasterdam.

Tabel 1: Prestatievergelijking vernieuwd KNN, conventioneel KNN en lineaire modellen op onafhankelijke testset

prestatie	lineair	conventioneel KNN (verschil in %)	vernieuwd KNN (verschil in %)
R²	68.5%	77.4% (+8.9%)	85.1% (+16.6%)
RMSE	19.6	16.0 (-18.4%)	13.3 (-32.3%)
gemiddelde fout	11.0	10.6 (-3.6%)	9.31 (-15.4%)
maximale fout	248	169 (-31.9%)	132 (-46.8%)

Waal en het chloridegehalte. Bij verhoging van de afvoer neemt het chloridegehalte exponentieel af tot een minimumwaarde. Bij de hier gepresenteerde relatie is een partiële R² bepaald van 68 procent. Deze geeft de invloed weer van de afvoer van de Waal volgens het model.

Momentane stadsdruk in waterleidingnet Amsterdam

In het kader van het project ‘Kunstmatig neurale netwerk als voorspelmodel voor het waterverbruik van Amsterdam en Amstelveen’ in opdracht van Gemeentewaterleidingen Amsterdam²⁾ is de vraag gerezen of het mogelijk zou zijn om met behulp van vernieuwde

KNN voorspellingen te doen van de leidingdruk op diverse locaties in Amsterdam aan de hand van het verpompte debiet en de bijbehorende waargenomen druk op een vijftal pomp-

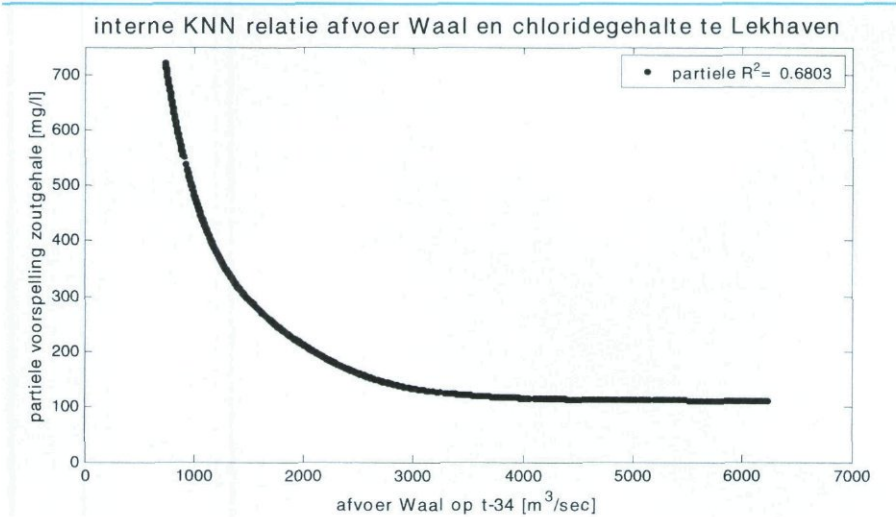
stations (zie afbeelding 6). Bij alle pompstations wordt zowel het verpompte debiet (watertransport) als de bijbehorende druk om de drie seconden gemeten en getransformeerd naar vijf minuten gegevens. Het uiteindelijke doel van de voorspelling van de druk is de ontwikkeling van een monitoringsysteem om enerzijds de kwaliteit van de inkomende drukgegevens te controleren en anderzijds het detecteren van trendmatige veranderingen en/of calamiteiten.

Op basis van de druk en debiet gegevens van twee jaar op vijf minuten basis zijn op

Tabel 2: Prestatievergelijking vernieuwd KNN, conventioneel KNN en lineaire modellen op onafhankelijk random testset.

prestatie	lineair	conventioneel KNN (verschil in %)	vernieuwd KNN (verschil in %)
R²	94.4%	96.3% (+1.9%)	99.2% (+4.8%)
RMSE	1.72	1.36 (-20.9%)	0.608 (-64.7%)
gemiddelde fout	0.639	0.991 (+55.1%)	0.433 (-32.2%)
maximale fout	222	22.0 (-90.0%)	20.5 (-90.8%)

Afb. 5: Door KNN gelegde relatie tussen de afvoer van de Waal en het chloridegehalte van de Lekhaven.

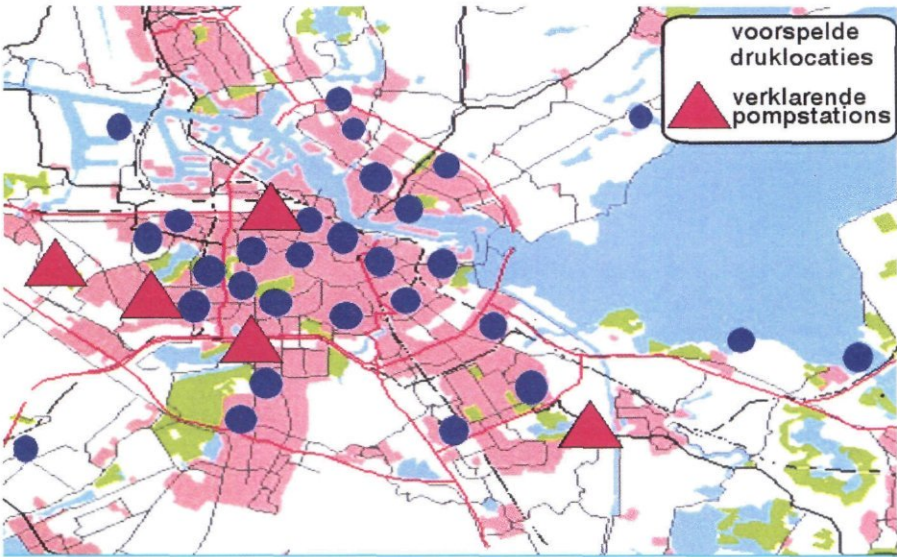


geautomatiseerde wijze 28 KNN-modellen ontwikkeld en optimaal gekalibreerd. Hierbij zijn na uitgebreid onderzoek per locatie tussen de 13 en 23 relevante verklarende grootheden bepaald. Voor een gedegen vergelijking zijn voor dezelfde input-output relaties ook additieve lineaire modellen ontwikkeld.

Afbeelding 7 presenteert een gedeelte van de onafhankelijke en gerandomiseerde testset voor de voorspelling van de stadsdruk op de locatie Savorin Lohmanstraat.

In tabel 2 is een vergelijking in prestatie te zien tussen lineair, conventioneel KNN en de vernieuwde KNN.

Wanneer alleen een additief lineaire analyse van de onderzochte druksystemen zou zijn uitgevoerd, geeft een gemiddelde verklaring



Afb. 6: Te voorspellen stadsdruk en pompstations op locaties in Amsterdam.

van 94 procent geen aanleiding tot de veronderstelling dat de overige procenten nog belangrijke informatie kan bevatten welke volgens een niet-lineaire relatie of interactie afhangt van de geselecteerde verklarende grootheden. Toch blijkt een relatief simpel, maar optimaal KNN in staat deze kleine zes procent nog zeer goed te verklaren en te voorspellen wat zelfs voor deze hoge mate van verklaring nog resulteert in een grote relevante verbetering.

In tabel 3 is een vergelijking gegeven in gemiddelde prestatievoorwaarden voor alle druklo-

cationele lineaire, hydraulische en fysische modellen. Ook geven gekalibreerde operationele KNN veel snellere simulaties en voorspellingen dan fysische modellen door de zeer geringe rekentijd.

Door de vergroting van het inzicht in het KNN-systeem en de mogelijkheden modelrelaties aan het model te onttrekken wordt het mogelijk een gekalibreerd KNN kwantitatief te toetsen aan fysische wetmatigheden. Ook kan met behulp van het netwerk steeds beter de kwaliteit van bestaande lineaire of fysische modellen worden onderzocht op de potentiële

maximale prestatie. Beter prestaties van KNN, door een toegevoegde waarde in fysisch relevante relaties, ten opzichte van lineaire of fysische modellen geven aan dat het KNN als identificatietechniek gebruikt kan worden.

LITERATUUR:

- 1) Gemeentewaterleidingen Amsterdam (1998). Kunstmatig Neuraal Netwerk als voorspelmodel voor het waterverbruik van Amsterdam en Amstelveen. Witteveen+Bos.
- 2) Gemeentewaterleidingen Amsterdam (2001). Ontwikkeling analyse component. Witteveen+Bos.
- 3) Hagan M., H. Demuth en M. Beale (1996). Neural network design. PWS Publishing Company.
- 4) Masters T. (1995). Neural, novel and hybrid algorithms for time series prediction. John Wiley & Sons.
- 5) PWN Waterleidingbedrijf Noord-Holland (2001). Prognose dagverbruik drinkwater nadere uitwerking alternatieven. Witteveen+Bos.
- 6) RWS-RIKZ (2001). Opzetten van een neuraal netwerk voor het voorspellen van chlorideconcentraties in het noordelijk deltabekken. Witteveen+Bos.
- 7) Schulze F. en A. Salverda (1999). Kunstmatige Neurale Netwerken in integraal waterbeheer: fictie of de toekomst?. H₂O nr. 3.
- 8) Wielen J. van der (2001). Gestructureerd modelontwerp met kunstmatig neurale netwerken. Witteveen+Bos en Wageningen Universiteit.

NOTEN:

* R² staat voor fractie verklaarde variantie, een relatieve maat-eenheid voor het gedeelte (de fractie of percentage) van het systeem dat door het model verklaard wordt. RMSE staat voor Root Mean Square Error, een absolute maat voor de grootte van de fout.

Tabel 3: Prestatievergelijking KNN en lineaire modellen over alle locaties.

prestatie	lineair	vernieuwd KNN (verschil in %)
gemiddelde R ²	94.7%	97.7% (+3.1%)
gemiddelde RMSE	1.7	1.2 (-29.4%)
aantal locaties met R ² < 95%	12 van de 28	1 van de 28
aantal locaties met R ² < 90%	3 van de 28	0 van de 28
aantal locaties met R ² > 99%	3 van de 28	8 van de 28

ties voor zowel lineair als vernieuwde KNN modellen.

Discussie/conclusies

Met behulp van de nieuwe ontwikkelingen op het gebied geautomatiseerd modelontwerp, verbetering trainingsproces en mogelijkheden tot kwantitatieve relatie detectie van kunstmatig neurale netwerken is zowel de kwaliteit als de prestatiemogelijkheden van het KNN als statistische modelleringstechniek sterk verbeterd ten opzichte van de conventionele KNN. Het KNN is volgens een vaste werkwijze steeds meer in staat zeer uiteenlopende systemen binnen vele gebieden van integraal waterbeheer kwalitatief zeer goed te modelleren en te voorspellen. Zeker op het gebied van het voorspellen van extremen heeft het KNN een grote voorsprong op con-

Afb. 7: Verschillende voorspellingsmodellen voor de stadsdruk in de Savorin Lohmanstraat in Amsterdam.

