

# Analyse van relaties tussen sedimentkarakteristieken en macroinvertebratengemeenschappen a.d.h.v. data mining technieken

<sup>1</sup> Universiteit Gent, Vakgroep  
Toegepaste Ecologie en  
Milieubiologie

<sup>2</sup> VMM,  
Afdeling Meetnetten en Onder-  
zoek

<sup>3</sup> VMM,  
Afdeling Ecologisch Toezicht

*Dit onderzoek beoogde het bepalen van de dominante sedimenteigenschappen voor rivierorganismen, door beslissingsbomen en artificiële neurale netwerken toe te passen op de VMM-databank van onbevaarbare waterlopen in Vlaanderen. De gebruikte modelleringstechnieken zijn beiden louter gebaseerd op de trends in de gegevens. Bij de discussie werden de resultaten van de gegevensgebaseerde modellen echter wel via sensitiviteitsanalyses geëvalueerd a.d.h.v. expertregels uit de literatuur. Deze benadering laat toe om nieuwe habitatrelaties af te leiden uit grote gegevensbanken, en op die manier een beter inzicht in rivierecosystemen te krijgen om het beheer van waterlopen te helpen ondersteunen.*

## Inleiding

De Europese Kaderrichtlijn Water (KRW) 2000/60/EC stelt voor alle lidstaten van de Europese Unie een goede ecologische status van alle waterlichamen tegen 2015 voorop (EU, 2000). Het overgrote deel van deze waterlichamen kan als stromende wateren of rivieren aanzien worden. Volgens de KRW moet de waterkwaliteit van de rivieren beoordeeld worden door de actuele condities te vergelijken met de referentiecondities. Daarom moeten eerst referentiecondities, die een goede ecologische status voorstellen, beschreven worden. Bijkomend moet de relatie tussen aquatische gemeenschappen en de menselijke activiteiten die deze watersystemen aantasten beter begrepen worden om zo een representatieve set van indices voor ecologisch rivierbeoordeling te kunnen ontwikkelen (De Pauw *et al.*, 2006). Kennis over deze relaties kan eveneens nuttig zijn bij zowel het opsporen van oorzaken van bepaalde riviercondities (milieueffectrapportering) als bij beslissingsondersteuning inzake rivierherstel en beheer om zo aan de eisen van de KRW te voldoen.

Tot nu toe werden ecologische modellen zelden gebruikt bij de ondersteuning van rivierbeheer en waterbeleid. Modellen kennen in deze context nochtans verscheidene interessante toepassingen. Ten eerste kunnen deze modellen bijdragen tot een betere interpretatie van de huidige riviercondities, de oorzaken van bepaalde riviercondities kunnen achterhaald worden en beoordelingsmethoden kunnen geoptimaliseerd worden. Ten tweede kunnen deze modellen het effect van toekomstige rivierherstelmaatregelen op aquatische ecosystemen doorrekenen en de selectie van de beste herstelopties ondersteunen. Ten derde kunnen deze modellen de belangrijkste hiaten in onze kennis over riviersystemen opvullen en helpen bij het opzetten van kostenefficiënte monitoringsprogramma's.

Dit onderzoek beoogde het bepalen van de dominante sedimenteigenschappen voor rivierorganismen, door beslissingsbomen en artificiële neurale netwerken toe te passen op de VMM-databank van onbevaarbare waterlopen in Vlaanderen.

## Materiaal en methoden

Beslissingsbomen (Figuur 1) en neurale netwerken (Figuur 2) werden gebruikt om relaties tussen gegevens van de TRIADE-databank van de Vlaamse Milieumaatschappij te zoeken. De databank bestond uit 360 bemonsteringen, uitgevoerd in de periode 1996-1998. In totaal werden 24 milieuv variabelen gebruikt als basis voor de voorspelling van de macroinvertebratentaxa.

De betrouwbaarheid van deze modellen werd bepaald aan de hand van het procentueel correct geclassificeerde voorbeelden (*Correctly Classified Instances of CCI*) in de evaluatiedataset en tevens op basis van de Kappa coëfficiënt (voor details, zie Goethals, 2005). De resultaten van de neurale netwerken werden bovendien via sensitiviteitsanalyses mee geëvalueerd a.d.h.v. expertregels uit de literatuur.

De dataset werd opgedeeld in drie, waarbij 2/3 gebruikt werd voor de ontwikkeling van de relaties, en het andere 1/3 voor het testen van de nauwkeurigheid. Voor details betreffende de gebruikte trainingsalgoritmen, modelontwikkelingsparameters en validatiemethoden, wordt verwezen naar Goethals (2005).

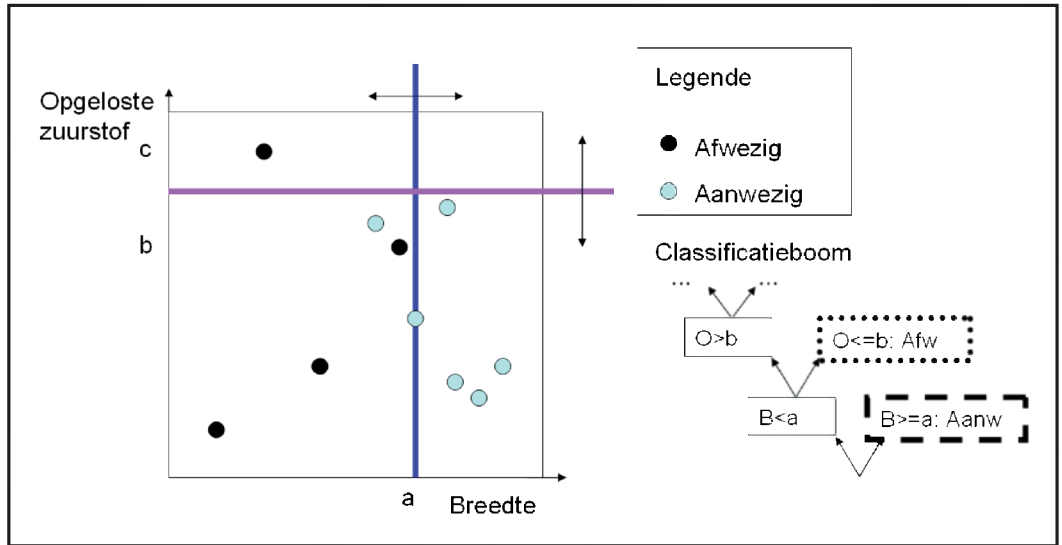
## Resultaten en discussie

In Figuur 3 wordt een voorbeeld gegeven van het type van als-dan-relaties in beslissingsbomen. Door combinatie van verschillende abiotische condities, kan men nagaan of Gammarus al dan niet aanwezig is op een site. Zo blijkt Gammarus in het algemeen afwezig te zijn op sites die gekarakteriseerd worden door een kleigehalte dat hoger is dan 11%. Ook zuurstofgehalten lager dan 6 mg per liter worden niet door dit taxon verdragen. Analoge relaties kunnen achterhaald worden betreffende de breedte, stroomsnelheid en pollutentconcentraties voor zowel dit taxon als andere dieren.

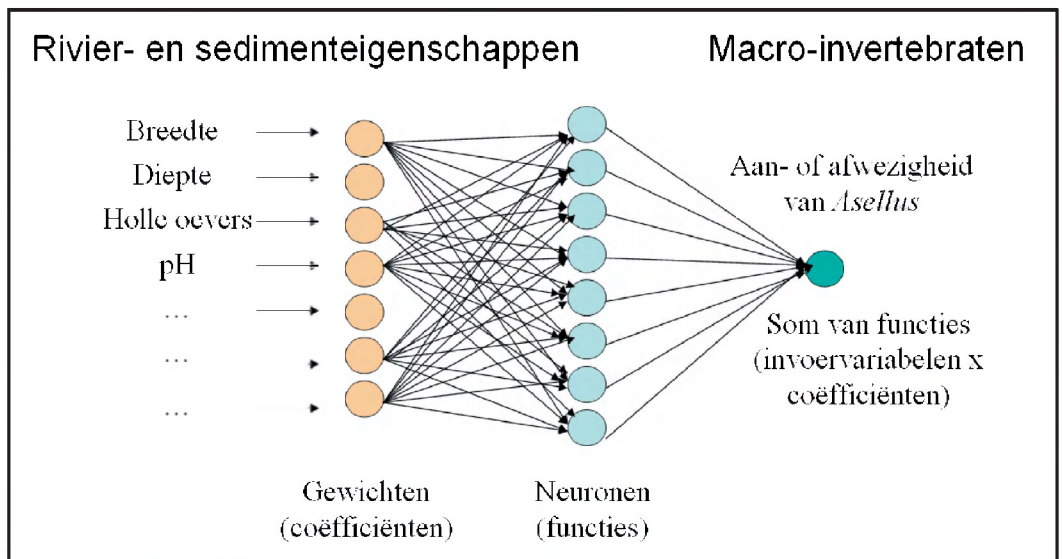
Bij vergelijking van de betrouwbaarheid van beide methoden, valt op dat zowel bij beslissingsbomen als neurale netwerken met een beperkt aantal



Figuur 1. Illustratie die de generatie van classificatiebomen voor de voorspelling van aan- en afwezigheid van macro-invertebraten op basis van rivier- en sedimentgegevens aantoont. Op basis van de gegevensset worden als-dan-regels getest op hun algemeenheid en betrouwbaarheid. In eerste instantie wordt vertrokken van de meest betrouwbare en algemene relatie (regel), in dit geval 'Als de breedte van de rivier groter dan of gelijk is aan a, dan is het taxon aanwezig', door dit proces systematisch te herhalen wordt een boom met classificatieregels bekomen die zo nauwkeurig mogelijk de aan- en afwezigheid van het taxon voorspelt op basis van rivier- en sedimentgegevens.

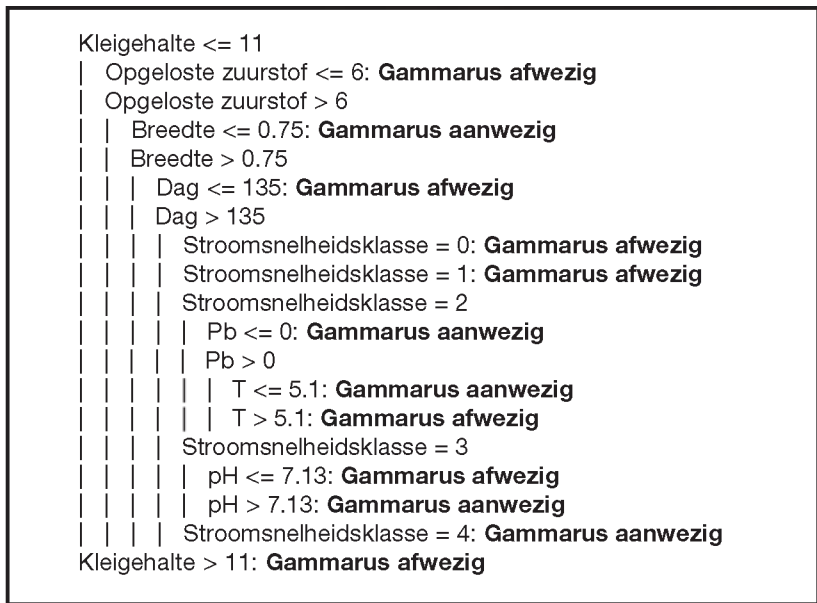


Figuur 2. Architectuur van de neurale netwerken voor de voorspelling van aan- en afwezigheid van macro-invertebraten op basis van rivier- en sedimentgegevens.

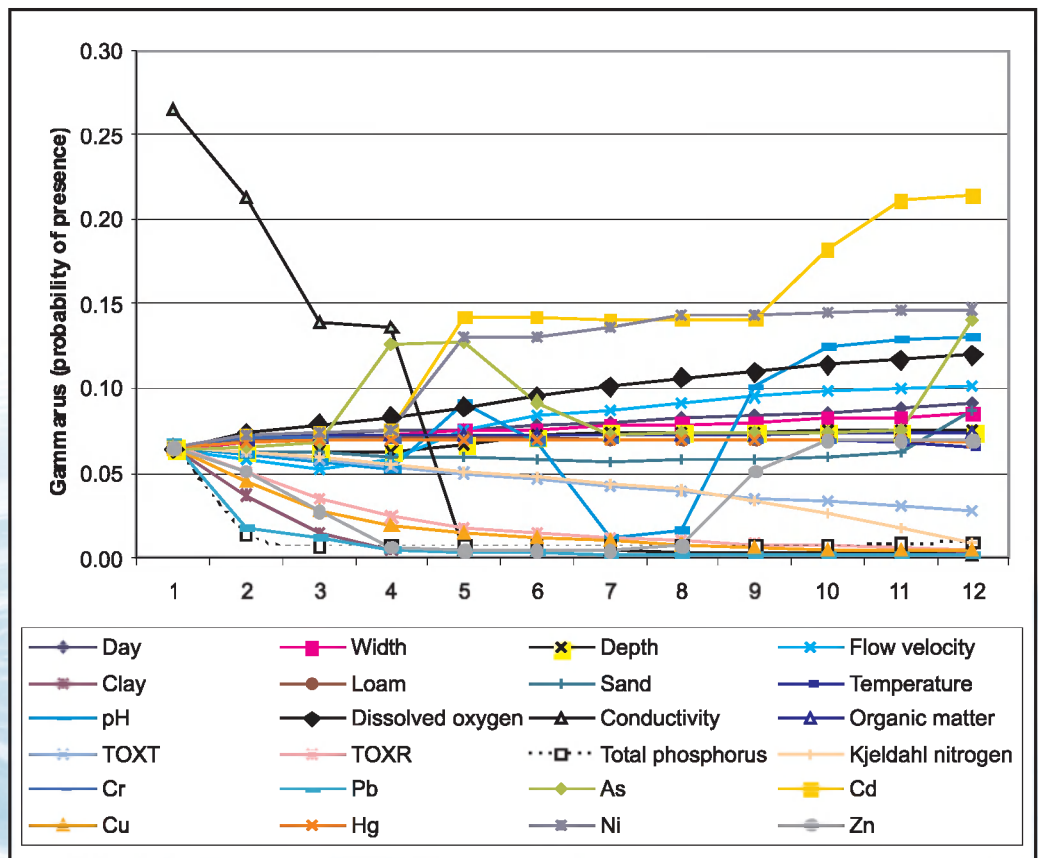


variabelen relatief hoge betrouwbaarheden konden bekomen worden. Voor de beslissingsbomen was de optimale CCI = 90,4 % en K = 0,64, terwijl voor neurale netwerken een maximale CCI = 86,8 en K = 0,49 werd gehaald. Voor andere taxa lag de betrouwbaarheid in het algemeen veel lager. Lage betrouwbaarheden kunnen zowel te wijten zijn aan relatief weinig gegevens (of een ongelijke verdeling van aan- en afwezigheden), maar tevens aan ontbrekende of foutieve meetgegevens (bv. ontbreken van cruciale milieuvariabelen) en dynamica van de betreffende systemen.

Figuur 3. Voorbeeld van een beslissingsboom voor Gammarus. Hoe meer de regels links staan, hoe belangrijker de relaties die worden weergegeven. Regels dienen meestal gecombineerd te worden om de aan- of afwezigheid van het taxon te kunnen voorspellen. Voorbeeld: 'Als het kleigehalte kleiner of gelijk is aan 11 % en het zuurstofgehalte lager of gelijk is aan 6 mg/liter, dan is Gammarus afwezig'.



Figuur 4. Voorbeeld van een sensitiviteitsanalyse op een artificieel neurale netwerk model voor Gammarus. Het totale interval tussen minimum en maximum werd voor elke variabele in twaalf gelijke delen opgedeeld. Vervolgens werden modelsimulaties gemaakt voor deze twaalf verschillende waarden van elke variabele, terwijl de andere variabelen op het gemiddelde werden gehouden.



In Figuur 4 worden de resultaten van de sensitiviteitsanalyse op een neurale netwerk model weergegeven. Gezien dergelijk model bestaat uit een complexe samenstelling van wiskundige transferfuncties, wordt deze methode vaak afgedaan als een black-box modelleringstechniek. Evenwel kan via deze sensitiviteitsanalyses nagegaan worden welke ecologische relaties onder deze transferfunctiesamenstelling schuilgaan. Uit Figuur 4 kan bijvoorbeeld eenvoudig worden afgeleid dat een hogere zuurstofconcentratie duidelijk een verbetering betekent voor Gammarus. Tevens geeft het model weer dat wanneer de conductiviteit hoger wordt dan 5/12 van het maximum (ongeveer 7 mS/cm), de kans dat Gammarus nog voorkomt nagenoeg nul is.

Voor meer gedetailleerde en uitgebreidere resultaten wordt verwezen naar Goethals (2005).

### Besluit

Deze benadering laat toe om nieuwe habitatrelaties af te leiden uit grote gegevensbanken, en op die manier een beter inzicht in rivierecosystemen te krijgen om het beheer van waterlopen te helpen ondersteunen. Bovendien kan nagegaan worden hoeveel informatie in een monitoringsset aanwezig is (maximale betrouwbaarheid van de modellen). Op basis daarvan krijgt de beheerder inzicht in welke aspecten vooral van belang zijn voor de aan- en afwezigheid van aquatische diersoorten, maar tevens wordt aangegeven hoe be-

trouwbaar en relevant monitoringsgegevens zijn. Deze modellen kunnen dus zowel een rol spelen bij het uitstippelen van herstelopties en het uitwerken van meetstrategieën.

## Referenties

De Pauw, N., Gabriels, W. & Goethals, P.L.M. (2006), Chapt. 2.1. River monitoring and assessment methods based on macro-invertebrates, In: Ziglio, G., Siligardi, M. & Flaim, G. (eds.), Biological monitoring of rivers, John Wiley & Sons, Ltd., Chichester, West Sussex. p. 113-134.

EU (2000), Richtlijn 2000/60/EG van het Europees Parlement en de Raad van 23 oktober 2000 tot vaststelling van een kader voor communautaire maatregelen betreffende het waterbeleid, Publicatieblad van de Europese Gemeenschappen L327, 1-72.

Goethals, P. (2005), Data driven development of predictive ecological models for benthic macroinvertebrates in rivers, Doctoraatsproefschrift, Universiteit Gent. 400 p.

P. Goethals<sup>1</sup>,  
W. Gabriels<sup>1,2</sup>,  
A. Dedecker<sup>3</sup>,  
W. De Cooman<sup>2</sup> en  
N. De Pauw<sup>1</sup>

<sup>1</sup> Universiteit Gent,  
Vakgroep Toegepaste Ecologie en Milieubiologie,  
J. Plateastraat 22,  
B-9000 Gent,  
Tel. 09 264 37 76,  
Fax 09 264 41 99

<sup>2</sup> VMM, Afdeling Meetnetten en Onderzoek,  
Dr. De Moorstraat 24-26,  
B-9300 Aalst,  
Tel. 053 72 65 09,  
Fax 053 70 63 44

<sup>3</sup> VMM, Afdeling Ecologisch Toezicht,  
Zandvoordestraat 375,  
B-8400 Oostende,  
Tel. 059 56 28 59,  
Fax 059 56 26 00