Voorspelling van effecten van ingrepen in het waterbeheer op aquatische gemeenschappen

Dit project is uitgevoerd in opdracht van RISTORI: RIZA, STOWA, RIVM

Voorspelling van effecten van ingrepen in het waterbeheer op aquatische gemeenschappen

De ontwikkeling van cenotypenvoorspellingsmodellen voor beken en sloten in Nederland

P.F.M. Verdonschot[1]
P.W. Goedhart[2]
R.C. Nijboer[1]
H.E. Vlek[1]

Alterra-rapport 738

Alterra, Research Instituut voor de Groene Ruimte, Wageningen, 2003

REFERAAT

Verdonschot P.[1], P. Goedhart[2], R. Nijboer[1] en H. Vlek[1], 2003. *Voorspelling van effecten van ingrepen in het waterbeheer op aquatische gemeenschappen; De ontwikkeling van voorspellingsmodellen voor beken en sloten in Nederland*. Wageningen, Alterra, Research Instituut voor de Groene Ruimte. Alterra-rapport 738. .. blz.; 21 fig.; .105 tab.; 60 ref.

Dit rapport geeft de resultaten weer van de ontwikkeling van twee cenotypenvoorspellings-modellen, één voor beken en één voor sloten, en is uitgevoerd in het kader van het project RISTORI fase 3. Beide modellen borduren voort op de modellen die ontwikkeld zijn in fase 1 en 2 van genoemd project. De ontwikkelde modellen zijn gebaseerd op een multinomiale logistische regressie benadering. De gegevens waarop de modellen zijn gebouwd, zijn afkomstig van de ecologisch-typologische netwerken ontwikkeld voor beken en sloten in de afgelopen jaren in opdracht van LNV. De modellen gaan van twee of drie verschillende hiërarchische niveaus uit: hoofdgroepen en/of groepen en cenotypen. Hiermee zijn de modellen ook bruikbaar op verschillende toepassingsniveaus. Met kruisvalidatie zijn de modellen geoptimaliseerd. De modellen zijn voorzien van een betrouwbaarheidsmaat.
Tevens is onderzoek gedaan naar het gebruik van waarderingindices en zeldzaamheid als maten om de cenotypen te beoordelen. Er is een aanzet voor dergelijke benaderingen gegeven, maar nader onderzoek is noodzakelijk.

Trefwoorden: beken, sloten, ecologische typen, multinomiale regressie, voorspelling, cenotype, waterbeheer, scenario, beoordeling, waardering, zeldzaamheidsindex, multimetric, indices, watertype

ISSN 1566-7197

Auteurs:
[1] Alterra, Research Instituut voor de Groene Ruimte
Afdeling Ecologie en Milieu
Team Zoetwaterecosystemen
Wageningen UR

[2] Plant Research International
Centrum voor Biometrie
Wageningen UR

Dit rapport kunt u bestellen door € 31,- over te maken op banknummer 36 70 54 612 ten name van Alterra, Wageningen, onder vermelding van Alterra-rapport 738. Dit bedrag is inclusief BTW en verzendkosten.

Projectnummer: 11694.01 [Alterra-rapport 738/JATW/05-2003]

Inhoud

Woord vooraf

In het kader van verschillende beleidsstudies worden door het RIZA beleidsanalyses uitgevoerd op het gebied van integraal waterbeheer. Voor die beleidsanalyses worden diverse modelinstrumenten ingezet waarmee effecten van maatregelen in het waterbeheer voorspeld kunnen worden. In deze keten van modellen mist nog een ingreep-effect-model voor aquatische ecosystemen in regionale wateren. Dat gemis wordt niet alleen ervaren in landelijke analyses maar ook in de (aanstaande) regionale stroomgebiedsstudies in het kader van de EU-KRW en WB21. Het RIVM heeft eveneens behoefte aan een dergelijk model voor de natuur- en milieuverkenningen waarmee ook de effecten van ingrepen moeten kunnen worden voorspeld. Vanuit haar positie als coördinator voor het onderzoek naar regionale watersystemen heeft ook de STOWA haar belangstelling voor een ingreep-effect-model.

Door de opdrachtgevers (RIZA, RIVM en STOWA), verenigd in het project RISTORI, is de bovenstaande behoefte vertaald in een project met als doel het ontwikkelen van een ingreep-effect-model met behulp waarvan afgewogen (op ecologisch inzicht gebaseerde) besluiten genomen kunnen worden over effecten van ingrepen in watersystemen. Ook moet met het model een afwegingskader beschikbaar komen voor de onderbouwing van gebiedspecifieke normdoelstellingen. Het model zal zoveel mogelijk moeten aansluiten op beschikbare gegevensbestanden, modellen en lopende ontwikkelingen.

Het overkoepelende project volgt twee benaderingen:
Benadering A: de soortsbenadering waarin het ontwikkelen van een effectmodel op basis van soort-factor relaties voortvloeiend uit de bestaande data-bases is beoogd.
Benadering B: de gemeenschapsbenadering: het ontwikkelen van een prototype effectmodel op basis van levensgemeenschappen (cenotypen) in relatie tot factorencomplexen.

De gemeenschapsbenadering binnen het project wordt in fasen uitgevoerd.
Definitiefase: Als eerste stap op weg naar een effectmodel voor aquatische natuurwaarden in regionale wateren is een definitie-studie verricht. De resultaten van deze fase hebben geleid tot een plan van aanpak waarin drie fasen en twee benaderingen zijn onderscheiden:
Fase 1: richtte zich op het ontwikkelen van een prototype van het effectmodel, ingevuld voor twee subtypen oppervlaktewateren: veensloten en middenlopen van laaglandbeken;
Fase 2: richtte zich op een verdere gegevens-optimalisatie, een veldvalidatie door toepassing van het model op nieuwe gegevenssets en op het aanbrengen van een verdere schaalverfijning in het effectmodel;
Fase 3: in de derde fase worden beken en sloten in het effectmodel ingebracht en wordt het model geoperationaliseerd.

Dit rapport bevat de resultaten van *Fase3, benadering B*; de gemeenschapsbenadering: "het ontwikkelen van ingreep-effectmodellen op basis van de gemeenschapstypen zoals die ontwikkeld zijn in het sloten- en bekenproject. Deze typen omvatten allerlei toestanden waarin sloten en beken momenteel in Nederland voorkomen.
Het rapport is een technisch verslag van het operationeel maken van de voorspellingsmodellen. Voor achtergrondinformatie wordt verwezen naar Nijboer *et al.* (1998) en Verdonschot & Goedhart (2000).
Het rapport valt in twee delen uiteen. In de hoofdstukken 1 tot en met 6 is de ontwikkeling van de cenotypenvoorspellingsmodellen voor beken en sloten uitgewerkt. In de hoofdstukken 7 tot en met 9 is uitgebreid toegelicht hoe de cenotypen, het product van de voorspelling, gewaardeerd kunnen worden.

Het project is begeleid door Bas van der Wal (STOWA), Rick Wortelboer (RIVM) en Francisco Leus (RIZA en direct aanspreekpunt en opdrachtgever). Francisco wordt bedankt voor zijn intensieve betrokkenheid gedurende het gehele proces, Bas en Rick worden bedankt voor hun commentaren op het rapport.

Samenvatting

Aanleiding

De uitvoering van integraal waterbeheer vraagt meer en meer om voorspelling van de effecten van voorgenomen ingrepen. Voor dergelijke voorspelling zijn diverse modelinstrumenten noodzakelijk waaronder een ingreep-effect-model voor aquatische ecosystemen in regionale wateren. Een dergelijk instrument is onder anderen nodig bij de Waterverkenningen (WVK), de Natuur- en Milieuverkenningen, de regionale watersysteemrapportages en -verkenningen en de daadwerkelijke evaluatie van waterbeheerprojecten zoals beekherstel, herinrichting van oevers of natuurontwikkeling. De noodzaak van een dergelijk instrument zal door de recente ontwikkelingen in het waterbeleid (gebiedsgericht, integraal, stroomgebied-aanpak, EU Kader Richtlijn Water, etc.) alleen nog maar groter worden.

Het RIZA, het RIVM en de STOWA hebben gezamenlijk het project RISTORI opgepakt, met als doel het ontwikkelen van een ingreep-effect-model met behulp waarvan afgewogen (op ecologisch inzicht gebaseerde) besluiten genomen kunnen worden over effecten van ingrepen in watersystemen. Tevens moet dit model een afwegingskader bieden voor de onderbouwing van gebiedsspecifieke norm-doelstellingen. Als randvoorwaarde gold dat het model zoveel mogelijk zou aansluiten op bestaande gegevensbestanden, modellen en lopende ontwikkelingen.

Ontwikkeling van een effectmodel

In 1997 is een definitiestudie uitgevoerd als eerste stap op weg naar een voorspellingsmodel voor aquatisch-ecologische effecten van ingrepen in regionale wateren. Het resulterende plan van aanpak onderscheidde twee uitgangsbenaderingen:

(i) een soortsbenadering waarin het effectmodel gebaseerd is op soort-factor relaties

(ii) een gemeenschapsbenadering: waarin het effectmodel gebaseerd is op soortencombinaties in relatie tot factorencomplexen (cenotypen).

Deze laatste benadering wordt hier gerapporteerd.

Gemeenschapsbenadering

De gemeenschapsbenadering is in fasen uitgevoerd:

De eerste fase (1998) richtte zich op het ontwikkelen van een prototype van het effectmodel, ingevuld voor twee subtypen oppervlaktewateren: veensloten en middenlopen van laaglandbeken op landelijke schaal (data uit de Stowabase).

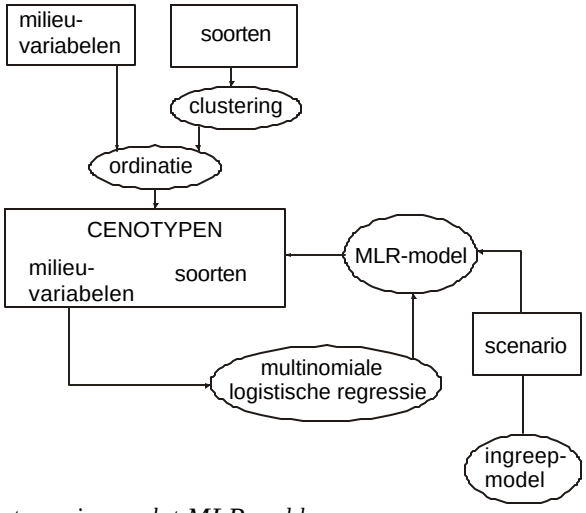
Fase 2 (1999-2000) richtte zich op drie onderdelen van het prototype van het effectmodel: (i) het verbeteren van de modelbasis door uit te gaan van betere en bredere gegevens (de EKOO data van de provincie Overijssel die verschillende watertypen omvatten), (ii) het valideren van het model door toepassing op nieuwe gegevens (case-studies voor drie projecten van regionale waterbeheerders) en (iii) op het aanbrengen van een schaalverfijning (regio Overijssel).

In de derde fase (2001-2002) zijn aparte landelijke modellen voor beken en sloten (gebaseerd op typologieën opgesteld in opdracht van LNV) ingebracht en is het

model geoperationaliseerd. Daarnaast zijn methoden van waarderen als attribuut bij het model verkend.
De ontwikkeling van de gemeenschapsbenadering is in detail terug te vinden in Roos *et al.* 1997, Nijboer *et al.* 1998 en Verdonschot & Goedhart 2000.

Het gemeenschapsmodel

De gemeenschapsbenadering (oftewel cenotypenbenadering) is gebaseerd op het gebruik van een soortencombinatie of gemeenschap als responsvariabele. Op basis van resultaten van ingreep-modellen worden scenarios in termen van milieuvariabelen geformuleerd. Het cenotypen-effectmodel berekent vervolgens de daaruit voortvloeiende kans op het voorkomen van een bepaalde aquatische gemeenschap.

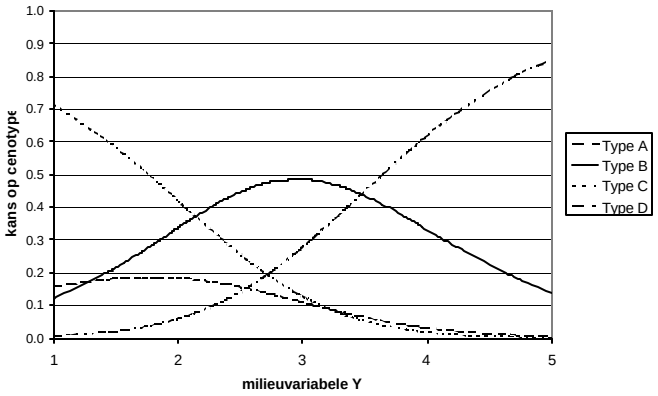


Figuur a. Ontwikkeling en toepassing van het MLR-model

De modelontwikkeling bestaat uit drie stappen (figuur a):

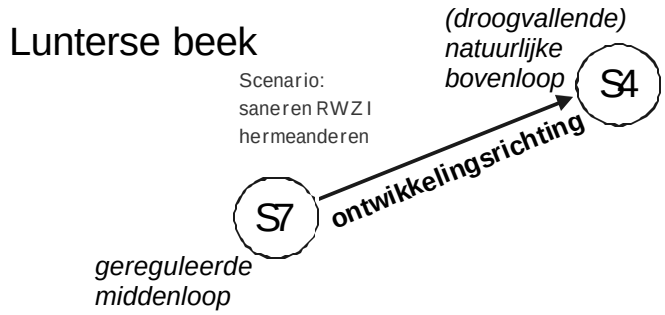
1. Ten eerste worden gemeenschappen geformuleerd op basis van macrofauna-monsters met behulp van clustering. Clustering ordent macrofauna-monsters naar mate van overeenkomst in soortensamenstelling.
2. Ten tweede wordt de gemeenschap verfijnd door het toepassen van directe ordinatie op een combinatie van de macrofauna-monsters en de daarbij behorende milieuvariabelen. Hieruit volgt de definitieve beschrijving van gemeenschappen; cenotypen genaamd (Verdonschot 1990). De cenotypen tezamen vormen een typologie en staan in onderling verband in een netwerk. Deze verbanden zijn gedefinieerd in zich tussen de typen wijzigende milieuvariabelen. Op deze wijze worden typologieën met bijbehorende netwerken geformuleerd voor de betreffende oppervlaktewateren. (oppervlakte wateren in Overijssel (fase 2), beken en sloten van Nederland (fase 3: Nijboer & Verdonschot 2002, Verdonschot & Nijboer 2002).
3. Ten derde wordt op de abiotische beschrijving van de cenotypen multinomiale logistische regressie (MLR) toegepast. In multinomiale logistische regressie wordt een complex van milieuvariabelen gebruikt om de kans op een cenotype te voorspellen. Voor een nieuw monster met een bekende set aan milieuvariabelen

(afkomstig uit inputmodellen zoals hydrologische en eutrofiëringsmodellen) wordt dan de kans op elk cenotype voorspeld, bijvoorbeeld een voorspelde kans van 0.6 op type A, 0.4 op type B en 0.0 op alle andere cenotypen (figuur b).



Figuur b. De kans op voorkomen van vier cenotypen A, B, C en D langs een milieuvariabele Y. De som van de kans op type A + B + C + D is gelijk aan 1 voor een bepaalde waarde van de milieuvariabele Y. In multinomiale regressie wordt niet met één maar met meerdere milieuvariabelen tegelijk gerekend

Een model gebaseerd op multinomiale logistische regressie bevat regressie-coëfficiënten voor de milieuvariabelen, die uit de gegevens geschat zijn. Het model is sterk afhankelijk van deze gegevens, met andere woorden het is van het grootste belang dat gegevens betrouwbaar en consistent gemeten zijn en worden om dergelijke modellen te bouwen en toe te passen.
Met het resulterende cenotypen-voorspellingsmodel wordt in feite geen nieuwe situatie voorspeld, maar worden richtingen binnen een bestaand netwerk van cenotypen aangegeven (figuur c).



Figuur c. Voorbeeld van de ontwikkelingsrichting in een mogelijke scenario dat ontwikkeld is voor de Lunterse beek

De modellen gaan steeds uit van meerdere aggregatieniveaus (van fijn naar grof): cenotypen, eventueel groepen en hoofdgroepen. Hiermee zijn de modellen ook bruikbaar op verschillende toepassingsschalen van landelijk beleid tot regionaal beheer. Met kruisvalidatie worden de modellen geoptimaliseerd. Bij kruisvalidatie wordt allereerst een groep waarnemingen weggelaten, het model wordt vervolgens

aangepast aan de overige waarnemingen en het aangepaste model wordt gebruikt om een voorspelling te berekenen voor de weggelaten waarnemingen.
In een voorspelling is de variantie vergelijkbaar met de Mahalanobis afstand van het te voorspellen monster met de dataset aan cenotypen: feitelijk de betrokken milieuvariabelen. De Mahalanobis afstands is een goede maat voor de betrouwbaarheid van een voorspelling en is aan de modellen toegevoegd.

Resultaat RISTORI fase 2: MLR-EKOO

Een totaal van 664 macrofauna-monsters door de provincie Overijssel/Alterra genomen in alle in de provincie Overijssel voorkomende oppervlaktewatertypen in de tachtiger jaren zijn gebruikt voor de bouw van het voorspellingsmodel. Daarnaast zijn op basis van multivariate analyse 23 milieuvariabelen (tabel b) voldoende differentiërend bevonden om in het model betrokken te worden.
Het model is gebaseerd op drie niveau's (tabel a): hoofdgroepen (4), groepen (12) en cenotypen (40). Het model is bruikbaar voor alle watertypen in Overijssel en mogelijk in vergelijkbare typen daarbuiten.

Ontwikkeling van MLR-SLOOT

Gedurende fase 3 is een sloten- en een bekenmodel ontwikkeld. Op basis van de landelijke slotentypologie is uit een totaal van 7787 macrofauna-monsters, door de regionale waterbeheerders genomen in sloten in de negentiger jaren, een selectie van 1204 geschikte monsters gemaakt. Na voorbewerking en pre-analyse zijn hiervan 493 monsters opgenomen in een definitieve slotentypologie van Nederland. Deze monsters zijn gebruikt bij de bouw van het voorspellingsmodel. Daarnaast zijn op basis van multivariate analyse 32 milieuvariabelen voldoende differentiërend bevonden om in het model betrokken te worden. Voor de sloten zijn twee modellen ontwikkeld: Model I en II, die enigszins verschillen in indeling in groepen van cenotypen. Onder beide modellen zijn twee niveaus gebruikt.
Een praktijktest moet uitwijzen welke van beide modellen verder ontwikkeld wordt. Hiertoe worden bij de regionale waterbeheerders enkele geschikte datareeksen verzameld. MLR-SLOOT is toepasbaar voor sloten in heel Nederland.

Tabel a. Aantal benodigde milieuvariabelen, hoofdgroepen, groepen en cenotypen per model

	EKOO	**SLOOT-I**	**SLOOT-II**	**BEEK-I**	**BEEK-II**
aantal milieuvariabelen	23	20	20	18	17
hoofdgroepen	4	7	6	6	3
groepen	12	-	-	-	9
cenotypen	40	13	13	25	25

Ontwikkeling van MLR-BEEK

Uit een totaal van 3259 macrofauna-monsters door de regionale waterbeheerders genomen in beken in de negentiger jaren is een selectie van 949 geschikte monsters gemaakt. Na voorbewerking en pre-analyse zijn hiervan 617 monsters opgenomen in een definitieve bekentypologie van Nederland. Deze monsters zijn gebruikt bij de bouw van het voorspellingsmodel. Daarnaast zijn op basis van multivariate analyse 21 milieuvariabelen voldoende differentiërend bevonden om in het model betrokken te worden.

Voor de beken zijn eveneens twee modellen ontwikkeld: Model I en II, die verschilen in verdeling van hoofdgroepen en groepen. Onder Model I zijn de cenotypen opgedeeld in 6 hoofdgroepen en 24 cenotypen; het groepsniveau ontbreekt. Model II heeft 3 hoofdgroepen en additioneel 9 groepen als tussenlaag en eveneens 24 cenotypen.
Een praktijktest moet uitwijzen welke van beide modellen verder ontwikkeld wordt. MLR-BEEK is toepasbaar voor beken in heel Nederland.

Tabel b. Model-invoerparameters, hun eenheid en de invoercode

parameteromschrijving	**code**	**eenheid**	**EKOO**	**SLOOT-I**	**SLOOT-II**	**BEEK-I**	**BEEK-II**
totaal fosfaat	t-P	mgP/l	+ (gem)	+	+	+ (90-p)	+ (90-p)
nitraat	NO3	mgN/l	+ (gem)	+ (gem)		+ (10-p)	+ (10-p)
ammonium	NH4	mgN/l	+ (gem)	+ (gem)	+ (gem)	+ (90-p)	+ (90-p)
zuurstof	O2	mg/l	+ (gem)			+ (10-p)	
breedte	b	m	+	+	+	+	+
diepte	d	m	+	+	+	+	+
stroomsnelheid	s	m/s	+			+	+
verval	verval	m/km	+				
zuurgraad	pH	-	+ (gem)	+	+	+ (med)	+ (med)
geleidendheid	EGV	mS/m	+ (gem)	+ (gem)	+ (gem)	+ (med)	
droogval	droogval	0/1	+	+	+		
laagveen	laagveen	0/1	+				
% drijvende vegetatie	%drijf	%	+	%	%		
% ondergedoken vegetatie	%onder	%	+	%	%		
% totale vegetatiebedekking	totb	%	+				
niet lijnvormig regelmatig profiel	ISRE	nominaal	+				
niet lijnvormig onregelmatig profiel	IRIR	nominaal	+				
lijnvormig regelmatig profiel	LSRE	nominaal	+				
lijnvormig onregelmatig profiel	LSIR	nominaal	+				
calcium	Ca	mg/l	+				
chloride	Cl	mg/l	+				
normalisatie niet	REGULNT	0/1	+				
normalisatie sterk	REGUL	0/1	+				
meandering	meander	0/1				+	+
natuurlijk dwarsprofiel	dwarsnat	0/1				+	+
permanentie	perman	0/1				+	+
stuwing	stuw	0/1				+	+
winter	winter	0/1				+	+
beschaduwing	schaduw	%				+	+
substraat slib	subslib	%				+	+
substraat zand	subzand	%				+	+
chloride	clmed	mg/l		+	+	+ (med)	+ (med)
Kjeldahl-stikstof	nkjel90	mgN/l					+ (90-p)
totaal stikstof	totaaln	mgN/l		+	+		
% emergente vegetatie	vemers	%		+	+		
% flab	vflab	%		+	+		
bodem zand	zand	0/1		+	+		
grondgebruik natuur	gnatuur	0/1		+	+		

parameteromschrijving	code	eenheid	EKOO	SLOOT-I	SLOOT-II	BEEK-I	BEEK-II
grondgebruik weiland	gweii	0/1		+	+		
inlaat van gebiedsvreemd water	inlaat	0/1		+	+		
functie natuur	fnatuur	0/1		+	+		
kwel	kwel	0/1		+			
bodem klei	klei	0/1			+		

Waarderen

Om te kunnen aangeven wat de ecologische waarde van een toekomstige voorspellingsresultaat is, kunnen de cenotypen worden gewaardeerd. In het rapport wordt uitgebreid toegelicht hoe de cenotypen, het product van de voorspelling, in de toekomst en naar de eisen van de Europese Kader Richtlijn Water gewaardeerd kunnen gaan worden. De waardering wordt verdeeld in vijf kwaliteitsklassen en is steeds afhankelijk van het verschil ten opzichte van de referentietoestand. Voor het waarderen is gebruik gemaakt van de resultaten van het Europese onderzoeksproject AQEM. AQEM heeft een Kader Richtlijn proof beoordelingssysteem ontwikkeld voor stromende wateren in Europa. Het waarderen van typen in de modellen kan hier in de nabije toekomst goed bij aansluiten. Daadwerkelijke invulling is vooralsnog moeilijk wegens het ontbreken van degelijke referentiebeschrijvingen.
Daarnaast is extra aandacht gegeven aan zeldzaamheid. Dit biedt de mogelijkheid om in de toekomst naast een waterkwaliteitsbeoordeling ook een natuurwaardering in het model op te nemen.

Toepassingsmogelijkheden

Het model MLR-EKOO is in samenspraak met drie waterbeheerders voor praktijksituaties gevalideerd. Uit deze praktijkvalidatie is gebleken dat het model bruikbaar is voor het regionale waterbeheer. De andere modellen gaan in 2002 getoetst worden.
De werking van een scenario-analyse bestaat uit de volgende stappen:

1. Het vaststellen van één of meer scenario's (bijvoorbeeld saneren RWZI en hermeanderen beekloop: figuur c).
2. Het berekenen (bijvoorbeeld van de vermindering in de basisafvoer bij het saneren van de RWZI), voorspellen (bijvoorbeeld van de nutriëntenconcentraties na sanering) of vaststellen op basis van expert judgement (bijvoorbeeld de dimensies en het lengteprofiel van de beekloop) van de waarden van invoerparameters (conform de lijst in tabel b).
3. Het uitvoeren van de berekening van de effecten met behulp van het voorspellingsmodel (bijvoorbeeld het MLR-EKOO in het voorbeeld Lunterse beek).
4. Het interpreteren van de voorspellingsresultaten op de ontwikkelingsrichting na uitvoeren van de maatregel, de betrouwbaarheid van de uitkomst en het cenotype dat verwacht gaat worden (in het voorbeeld de ontwikkeling van een gereguleerde middenloop naar een (droogvallende) natuurlijke bovenloop).

De modellen zijn nog niet in de praktijk toepasbaar. Er is uitgegaan van de eerder ontwikkelde prototypen met het accent op de functionaliteit daarvan en niet op uitvoering en gebruikersvriendelijkheid. De modellen zijn in fortran geprogrammeerd.

Conclusies

Het blijkt mogelijk om op basis van cenotypologieën, geordend in netwerken, voorspellingssystemen te bouwen. Cenotypologieën stellen echter hoge eisen aan de beschikbare gegevens.

Multinomiale logistische regressie blijkt een degelijke techniek voor cenotypen modellering. MLR voorspelt echter alleen binnen de gedefineerde ruimte (dus de gebruikte gegevens).

Eén regionaal en twee nationale centypenvoorspellingsmodellen zijn inmiddels beschikbaar.

De resultaten van de eerste praktijkvalidaties blijken voldoende vertrouwenwekkend om een bredere toepassing uit te testen. In 2002 volgen nog eens 4-5 case-studies.

Het project is begeleid door Bas van der Wal (STOWA), Rick Wortelboer (RIVM) en Francisco Leus (RIZA en direct aanspreekpunt en opdrachtgever). Francisco wordt bedankt voor zijn intensieve betrokkenheid gedurende het gehele proces, Bas en Rick worden bedankt voor hun commentaren op het rapport.

1 Inleiding

1.1 Doelstelling

Het doel van dit deelproject is het ontwikkelen van de effectmodule voor het toekomstige ingreep-effect-model voor aquatische ecosystemen in regionale wateren op basis van relaties tussen gemeenschappen en groepen van milieufactoren. Met het model moeten ecologische effecten van maatregelen en ingrepen kunnen worden voorspeld op het niveau van gemeenschappen. Hiermee moeten afgewogen besluiten genomen kunnen worden over voorgenomen ingrepen in een watersysteem. Ook moeten met het model gebiedspecifieke normdoelstellingen onderbouwd kunnen worden.

1.2 Gegevens

Alterra is in 1997 begonnen met de modelontwikkeling ten behoeve van het project RISTORI. Als eerste gegevensbestand zijn de middenlopen uit de Stowabase gebruikt. Dit betrof een beperkt gegevensbestand met een geringe variatie in milieuomstandigheden. Hierop is een eerste prototype model gebouwd (Nijboer *et al*. 1998).
In 1999 is het model gevuld met het EKOO-gegevensbestand (Ecologische Karakterisering van Oppervlaktewateren in Overijssel: Verdonschot 1990a/b), een bestand dat allerlei watertypen uit de regio Overijssel bevat. Het gegevensbestand EKOO omvat meer watertypen en monsters, is kwalitatief goed onderbouwd en is homogeen. Het gegevensbestand omvat een ruim aantal milieu-omstandigheden binnen één regio. Met deze gegevens is een nieuw model gebouwd en in 2000 is dit nieuwe model getoetst door middel van een pilotstudie uitgevoerd voor drie verschillende beheersgebieden Verdonschot & Goedhart 2000).
Parallel zijn in het kader van programma en stimuleringsgelden van LNV, typologieën ontwikkeld voor sloten en beken. Beide typologieën worden voor het einde van 2002 afgerond. Beide gegevensbestanden zijn landsdekkend en bevatten een zekere variatie aan milieu-omstandigheden. De typologieën zijn in eerste instantie voor het nationale niveau geformuleerd. In 2001 zijn referentietypologieën voor beken en sloten ontwikkeld.

1.3 Modelontwikkeling

In de eerste fase van het project RISTORI is een prototype van een effectmodel ontwikkeld (Nijboer *et al.* 1998). Dit effectmodel is inde tweede fase gevuld met de gegevens zoals die verzameld zijn in het project “Ecologische Karakterisering van Oppervlaktewateren in Overijssel” (EKOO) (Verdonschot 1990a, 1990b; Verdonschot & Goedhart 2000).

In fase 3 van het project zijn twee bouwstenen toegevoegd; de modelontwikkeling voor beken en sloten en het waarderen van de opgenomen gemeenschappen. Deze bouwsteen in het project heeft tot doel te komen tot operationele versies van het EKOO- (alle watertypen Overijssel), het beken- (landelijke typologie) en het sloten- (landelijke typologie) cenotypenvoorspellingsmodel.
Het EKOO-cenotypenvoorspellingsmodel is inmiddels operationeel. De volgende stap in de modelontwikkeling (als onderdeel van fase 3) is het gereed maken en opnemen in afzonderlijke modellen van de gemeenschappen van respectievelijk beken en sloten volgens de eerder hiervoor opgestelde typologieën.
Daarna zijn beide modellen in enkele testgebieden verspreid over Nederland, aan enkele hoofdvragen in het regionale waterbeheer onderworpen.
Ook de nieuw te ontwikkelen beken- en sloten-modellen zijn bedoeld om aan te tonen dat met de benadering reproduceerbare relaties tussen ingrepen in de watersystemen (in termen van gewijzigde milieu-omstandigheden) en de respons van gemeenschappen weergegeven kunnen worden.
Door middel van een validatie met onafhankelijke veld- en scenario-gegevens zijn de voorspellingen vergeleken met gemeten responsies in het veld.
Tot op heden zijn de modelformulering geprogrammeerd in Genstat. De prototypen zijn niet bruikbaar voor derden. Bij voldoende slagingskans wordt een operationeel en gebruikersvriendelijke rekenmodule in een gestandaardiseerde schil ontwikkeld.
Het effectmodel moet aansluiten op het dosismodel. Het dosismodel berekent de abiotische gevolgen van een voorgenomen ingreep (deze beoogde modellen bestaan nog slechts ten dele, dienen toekomstige milieu-omstandigheden te voorspellen en worden in dit proces verder niet meegenomen). Deze abiotische variabelen dienen als invoer voor het effectmodel. Het dosismodel gaat in de toekomst voldoen aan de eisen die het effectmodel stelt. Vooralsnog is zonder dosismodel gewerkt.
Voor de ontwikkeling van zowel het beken- als het slotenmodel zijn de volgende stappen doorlopen:
stap 1 het voorbewerken van de gegevens; het geschikt maken van de gegevensbestanden van beken en sloten en de multivariate analyse resultaten van de typologieën voor de selectie van predictoren (voorspellende milieuparameters);
stap 2 het opstellen van een hiërarchie; het uit beide typologieën en de ecologische overeenkomsten tussen typen opstellen van een hiërarchische structuur ten behoeve van de modelbouw;
stap 3 het selecteren van modelparameters; het selecteren van een minimum aantal predictoren op basis van a priori kennis en testen van alle mogelijke modellen
stap 4 het opstellen van modellen; het definitief formuleren van de modellen door het uitvoeren van verschillende typencombinaties in de vorm van een iteratief proces;
stap 5 het evalueren van de modellen; het met behulp van kruisvalidatie bepalen van de voorspelkracht van de modellen;
stap 6 het in fortran pogrammeren van de geschikte modellen; het vertalen van de Genstat formulering van het model in een fortran of C programma met een eenvoudige in- en output voor de EKOO, beken en sloten modellen en het toetsen of de programma's daadwerkelijk aan de eisen voldoen door validatie;
stap 7 het toevoegen van een betrouwbaarheidsmaat; het door extrapolatie een waarschuwing genereren bij de voorspelling door middel van het indiceren van een betrouwbaarheidswaarde;

Over de laatste stap:
stap 8 het valideren aan de hand van enkele praktijkvoorbeelden; het uitvoeren van een praktijkvalidatie waarbij beheerders/gebruikers bij voorkeur gegevens aanleveren die opgenomen zijn vooraf aan een ingreep/dosis en gegevens die na enkele jaren zijn verzameld; het betreft zowel milieugegevens als macrofauna.
zal apart worden gerapporteerd.

1.4 Waardering

In 1999 zijn de cenotypen uit EKOO in een voorspellingsmodel gebracht. Deze cenotypen zijn opgenomen in een netwerk, waarin verschillende ontwikkelingsreeksen in te onderscheiden zijn. Een ontwikkelingsreeks kan gezien worden als een waarderingsschaal. Afhankelijk van de keuze van het referentiepunt kan aan ieder cenotype een waardering worden gekoppeld. Deze "open" benadering maakt het mogelijk om ook doel en functie afhankelijk te waarderen.
Om de uitkomsten van een voorspelling te kunnen waarderen, met andere woorden om te kunnen aangeven wat de ecologische waarde van betreffende voorspellingsuitkomst is, moeten allereerst de cenotypen worden gewaardeerd. Voor het opstellen van een waardering wordt, indien mogelijk, aangesloten bij de ontwikkelingen in het kader van de Europese Kaderrichtlijn Water en de benaderingen van natuurwaarde, soortgroep trend index, rode lijst indicator en EHS-doelrealisatiegraadmeter van het Natuurplanburo (Brink *et al.* 2000). Alhoewel deze laatste groep nog niet in graadmeters is uitgewerkt komen de principes overeen met de benadering volgens de Kaderrichtlijn Water en de aquatische natuurdoeltypen.
Waarderen gebaseerd op de gemeenschapsbenadering betekent het verbinden van een waarde-oordeel aan ieder van de gemeenschapstypen in een model en het hiermee toekennen van deze zelfde waarde aan een nieuw voorspelde toestand. De Kaderrichtlijn Water gaat uit van een schaling ten opzichte van de referentietoestand. De mate van afwijking ten opzichte van de referentie bepaalt de kwaliteitstoestand van een water. In dit licht dient de waardering in RISTORI gekoppeld te worden aan de afstand ten opzichte van de referentie. De in de modellen opgenomen gemeenschapstypen dienen ten opzichte van de referentietoestanden te worden gewaardeerd.
Voor de huidige toepassing en met het oog op de EU-Kaderrichtlijn ligt het voor de hand de waarderingsschaal te koppelen aan referentietypen. De aquatische natuurdoeltypen voor beken en sloten in het Aquatisch Supplement vormen hiervoor een geschikt uitgangspunt. Door een koppeling te leggen met de natuurdoeltypen en deze te versterken met een zeldzaamheidsindex zou een geschikt waarderingssysteem worden verkregen. Echter de beschrijvingen voor beken en sloten in het Aquatisch Supplement zijn vooralsnog onvoldoende gekwantificeerd (de taxa duiden alleen een aan-/afwezigheid aan) en gecompleteerd (er zijn alleen indicatieve taxa in opgenomen).
Om de gemeenschapstypen uit de ecologisch-typologische netwerken toch te waarderen is onderzocht of:

- √ het opstellen van een waarderingssystematiek door het vergelijken en, indien mogelijk, integreren van ‘afstandsmaten’ en ‘zeldzaamheidstechnieken’ tot een waarderingssystematiek mogelijk is;
- √ een toets op bruikbaarheid van de gekozen techniek(en) plaats kan vinden.

1.5 Projectomgeving

De bouwstenen modelontwikkeling en waarderen staan niet los van elkaar. Ten eerste is er de onderlinge verwevenheid, want bij het voorspellen van een mogelijk effect behoort ook het waarderen van dat effect. Want wat heeft een beleidsmaker aan een voorspelling als hij niet ook meteen weet of het resultaat ook een kwaliteitsverbetering inhoudt. Ten tweede is gebruik gemaakt van de directe projectomgeving waarbij de volgende onderwerpen van belang zijn:
Aquatische natuurdoeltypen: In de periode 1998-2000 hebben Alterra/RIZA in opdracht van het EC gewerkt aan de formulering van referentietoestanden in regionale en rijkswateren. Eind 2000 zijn deze referentiebeschrijvingen beschikbaar gekomen. Deze beschrijvingen dienen als het natuurlijke referentiekader waartegen in de toekomst de huidige toestand (kwaliteit) kan worden afgezet. Bij beide bouwstenen spelen deze aquatische natuurdoeltypen een grote rol.
EU-Kaderrichtlijn: In de EU-Kaderrichtlijn wordt uitgegaan van het beoordelen van oppervlaktewateren middels een systeem waarbij de huidige toestand wordt afgezet tegen de referentie. De bouwsteen waarderen sluit hier direct op aan.
EU 5de Kader R&O onderzoeksprojecten: Alterra is betrokken in de EU-projecten AQEM en PAEQANN. Beide EU-projecten hebben een direct verband met RISTORI. In AQEM wordt een beoordelingssysteem ontwikkeld voor stromende wateren in Europa, dat geschikt is voor beoordeling volgens de EU-Kaderrichtlijn. De bouwsteen waarderen is hier nauw aan verwant en heeft gebruik gemaakt van de gegevens en resultaten. Het project PAEQANN werkt aan een nieuwe generatie voorspellingsmodellen (Artificiële Neurale Netwerken, Bayesiaanse technieken) waar RISTORI eind 2002 ook haar voordeel mee kan doen. Daarnaast vindt in de nabije toekomst vergelijking van de nieuwe met de in RISTORI ontwikkelde modellen plaats.
LNV programma’s: Voor RISTORI is het LNV programma 324 “Aquatische Ecosystemen en Visserij” direct van belang. De ecologisch-typologische netwerken, de basis voor de RISTORI gemeenschapsmodellen, zijn in het kader van dit programma ontwikkeld.

2 De cenotypenbenadering in kort bestek

De paragrafen 2.1 tot en met 2.3 in dit hoofdstuk zijn overgenomen uit Verdonschot & Goedhart (2000). Er is voor overname gekozen om lezers die niet bekend zijn met de cenotypenbenadering te informeren. Ieder ander kan deze paragrafen overslaan.

2.1 De gemeenschapsbenadering

Soorten zijn de expressie van hun omgeving. Deze omgeving is een samenspel van vele soorten en milieuvariabelen. Tussen één soort en één milieuvariabele bestaat een verband dat als optimumcurve kan worden weergegeven. Bij relaties tussen soorten spelen concurrentie, predatie, mutualisme en parasitisme een belangrijke rol. Theoretisch heeft iedere soort voor iedere milieuvariabele een optimum, een tolerantie-range en een lethaal gebied (een teveel of een te weinig). Echter dit is slechts theorie. Soorten leven in een omgeving opgebouwd uit een complex van milieuvariabelen (Pianka 1978, Karr 1991), inclusief andere soorten. Ook onder optimale milieuomstandigeheden voor een soort kan een predator of een concurrent deze soort in aantal doen afnemen. Een soort kan dus in lagere aantallen voorkomen bij de optimale waarde voor een milieuvariabele indien gelijktijdig een andere milieufactor, die ook biotisch van aard kan zijn, een negatief effect heeft op deze soort. Ook is het mogelijk dat de ene milieuvariabele het negatieve effect van een andere opheft, waardoor de soort juist in grotere aantallen voorkomt dan verwacht. In verschillende watertypen kan een soort dus verschillende optima hebben. De relatie tussen een soort en één milieuvariabele kan in de praktijk dus moeilijk worden vastgesteld. De interacties tussen soorten is een extra argument om alle soorten gezamenlijk te beschouwen. Vanaf het begin van deze eeuw zijn groepen van soorten gebruikt voor typologieën (Thienemann 1925, Hartog 1963) en beoordelingssystemen (Armitage *et al.* 1983, Verdonschot 1990a, 1990b).
In de toepassing zullen verschillende combinaties van soorten te onderscheiden zijn bij verschillende complexen van milieuvariabelen. Veranderen de omstandigheden dan zullen er verschuivingen optreden in de aanwezige levensgemeenschap. Deze verschuivingen zijn te herkennen aan het verdwijnen van soorten en het verschijnen van andere soorten. Bij kleine veranderingen zullen slechts verschuivingen in aantallen individuen waar te nemen zijn. Grote veranderingen in milieu kunnen echter ook leiden tot verschuivingen in aantallen als gevolg van verschuivende concurrentie of predatie verhoudingen of omgekeerd kleine veranderingen in grote veranderingen in aantallen of soorten als gevolg van wijzigingen in biotische interacties. De wijze waarop deze veranderingen verlopen is vaak nog niet wetenschappelijk onderbouwd. Zeker is dat de relaties veelal niet lineair verlopen.

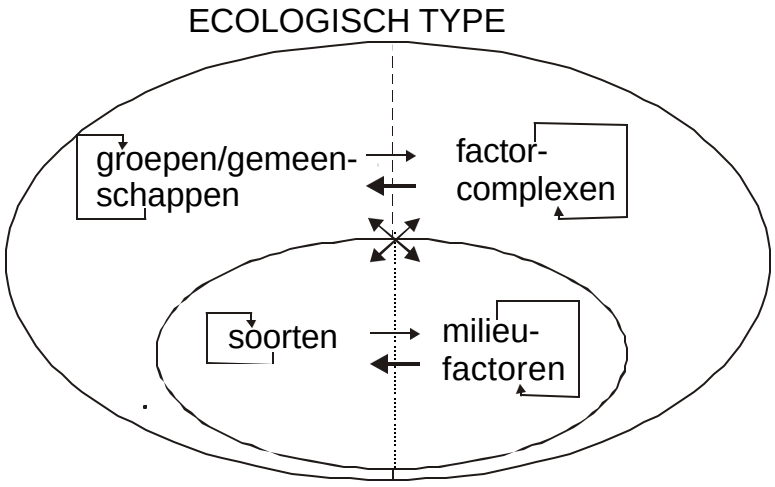
2.2 Ecologische typologie of cenotypologie

Milieuomstandigheden kunnen, vooral daar waar geologie, geomorfologie en klimaat vergelijkbaar zijn, in redelijk vergelijkbare samenstelling van combinaties van waarden voor alle milieuvariabelen voorkomen. Dit is vooral het geval binnen een watertype en een regio. Tussen verschillende watertypen of tussen verschillende regio's kunnen de verschillen beduidend groter zijn. Ook menselijke beïnvloeding kan de variatie vergroten. Het is belangrijk dat de effecten van menselijke verandering (ingrepen) van milieuvariabelen binnen een regio en/of watertype op soorten inzichtelijk gemaakt kunnen worden (Preston & Bedford 1988). In Nederland komen nog maar weinig natuurlijke situaties voor. Hierdoor bevinden zich in de meeste wateren vooral algemene soorten met brede toleranties. Deze soorten zijn op zich niet erg geschikt als indicator (Rosenberg & Resh 1993). Door echter alle soorten samen te beschouwen kan een beter beeld verkregen worden van onderliggende milieu-omstandigheden (Higler & Tolkamp 1984, Zonneveld 1984).
Omdat het vrijwel onmogelijk is om voor ieder oppervlaktewater apart een beoordeling uit te voeren en een beheersplan op te stellen worden wateren onderverdeeld in typen. Per type kan dan een beoordelings- en beheersinstrument gebouwd worden. Dit geldt ook voor het voorspellen van effecten van ingrepen. Een ingreep zal in het ene watertype een ander effect hebben dan in het andere. Het is dan ook noodzakelijk om een ingreep-effectmodel te baseren op een typologie. Volgens de doelstelling van het project moet het model gebaseerd zijn op ecologische inzichten. Daarom is voor de bouw van het model uitgegaan van een ecologische typologie.
Een type is de gemeenschappelijke grondvorm van een bepaald aantal verschijnselen. In dit geval gaat het om ecologische verschijnselen: een netwerk van organismen, milieuvariabelen en hun onderlinge relaties. Een type wordt gekarakteriseerd door een complex van milieuvariabelen en het voorkomen van een bepaalde "levensgemeenschap" (soortengroep of -associatie; in het engels als 'species-assemblage' aangeduid: figuur 2.1). Binnen een type is variatie mogelijk. Het type bepaalt slechts de algemene overeenkomst. Typen lopen vaak geleidelijk in elkaar over. Duidelijke grenzen zijn niet aan te geven. Overgangen worden geïnitieerd door veranderingen in bepaalde milieuvariabelen. Verschillen tussen typen worden bepaald door abiotische hoofdfactoren, dit zijn de milieuvariabelen die de grootste variatie tussen twee verschillende typen verklaren.
Het doel van een dergelijke ecologische typologie, ook wel cenotypologie genoemd (Verdonschot 1990a) is de multidimensionele relatie tussen soorten en milieuvariabelen te reduceren tot een weinig dimensionele praktisch hanteerbare en inzichtelijke relatie (Verdonschot 1983).

2.3 Een effectmodel op basis van een cenotypologie

Het ingreep-effectmodel is gebaseerd op een ecologische typologie. Uitgangspunt voor het model zijn de cenotypen (Verdonschot 1990b). Cenotypen zijn gekarakteriseerd aan de hand van soortengroepen en milieuvariabelen. De relaties tussen de cenotypen zijn weergegeven door de typen in een netwerk te plaatsen en de

cenotypen onderling met pijlen te verbinden. De pijlen geven weer welke milieuvariabelen als stuurvariabelen beschouwd kunnen worden voor de overgang van het ene cenotype naar het andere.



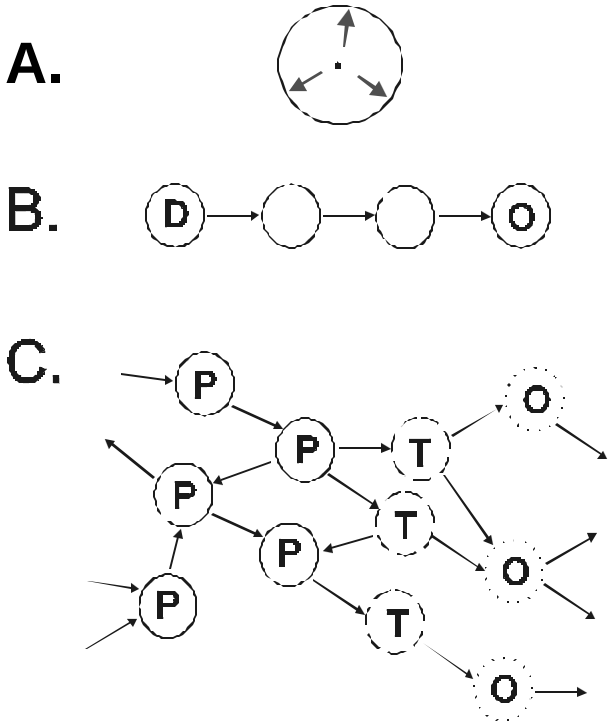
Figuur 2.1 Soort, factor, gemeenschap en factorcomplex in een ecologisch type

Ingrepen leiden in eerste instantie tot een kleine verschuiving binnen een cenotype (figuur 2.2A). Pas als er soorten verdwijnen en andere verschijnen is sprake van een overgang naar een ander cenotype (figuur 2.2B). De werkelijkheid is echter complexer dan een enkelvoudige ontwikkelingsreeks. Toestanden kunnen zich in meerdere richtingen ontwikkelen als gevolg van veranderingen in het milieu. Een hydrologische ingreep zal tot andere veranderingen in het systeem leiden als bijvoorbeeld het terugdringen van nutriëntengehalten. Voor een bepaalde actuele toestand zijn dus verschillende ontwikkelingsrichtingen mogelijk. Om deze ontwikkelingsmogelijkheden te omvatten is een netwerk van ontwikkelingsreeksen een goed instrument om dit eenvoudig te presenteren (figuur 2.2C). Complexe netwerken kunnen ook in meerdere reeksen worden omgezet. Een voorbeeld is uitgewerkt door Verdonschot, Gerritsen en Koopmans (1999). Uit een netwerk is eenvoudig af te lezen in welke toestand een cenotype zich bevindt, welke ontwikkelingsmogelijkheden er voor het cenotype zijn (tot welke andere cenotypen het zich kan ontwikkelen) en welke stuurvariabelen beïnvloed moeten worden om deze ontwikkeling te sturen.
Het model kan aan de hand van het netwerk voorspellen in welke richting een cenotype zich zal ontwikkelen. Een ingreep heeft direct effect op één of meer milieuvariabelen. De effecten kunnen echter ook doorwerken van de ene milieuvariabele op de andere. De effectmodule die in dit onderzoek is ontwikkeld, richt zich op de effecten die ontstaan als gevolg van een verandering in een milieuvariabele of een combinatie van milieuvariabelen. Voor veel ingrepen geldt het laatste en bij de ontwikkeling van de input- of dosismodellen dient hier rekening mee te worden gehouden. De milieuvariabelen die veranderen als gevolg van een bepaalde ingreep worden bekend verondersteld; deze komen voort uit het dosismodel. De effectmodule start dus bij de milieuvariabele(n) en niet bij de ingreep zelf. De

effectmodule heeft aan de andere kant ook een beperking. De voorspelling kan niet verder reiken dan de typen die vooraf in het model zijn opgenomen.
Het model kan voor twee doeleinden worden gebruikt:

- toedeling van de huidige situatie aan één van de bestaande cenotypen op basis van abiotiek
- voorspelling van de ontwikkelingsrichting (het toekomstige cenotype) aan de hand van de voorspelde waarde(n) (uit het dosismodel) voor de betreffende milieuvariabele(n).

Beide delen zijn gebaseerd op een techniek waarmee aan de hand van de abiotische variabelen een kansverdeling weergegeven wordt voor het voorkomen van een monster in elk van de cenotypen in het model.



Figuur 2.2 Relaties binnen (A: · = centroïd) en tussen (B, C) cenotypen die uiteindelijk leiden tot een netwerk (C). de pijlen stellen de stuurvariabelen voor. O = optimale toestand of referentie, P = huidig voorkomende toestand, T = toestand streefbeeld en D = extreem beïnvloed (dood) water

2.4 Ecologisch-typologisch netwerk voor beken en sloten

Voor zowel beken als sloten is getracht een nieuw, duurzaam en multifunctioneel raamwerk voor de ontwikkeling van ecologische instrumenten voor oppervlaktewateren te formuleren: een ecologisch-typologisch netwerk (zie paragraaf 2.2) . De netwerkbenadering hanteert een bottom-up werkwijze. Dit wil zeggen dat het instrument op regionale schaal gebouwd moet worden en voor hogere schalen wordt geaggregeerd. Vanuit het netwerk kunnen instrumenten worden gebouwd voor verschillende doeleinden/-groepen. Waterbeheerders prefereren instrumenten

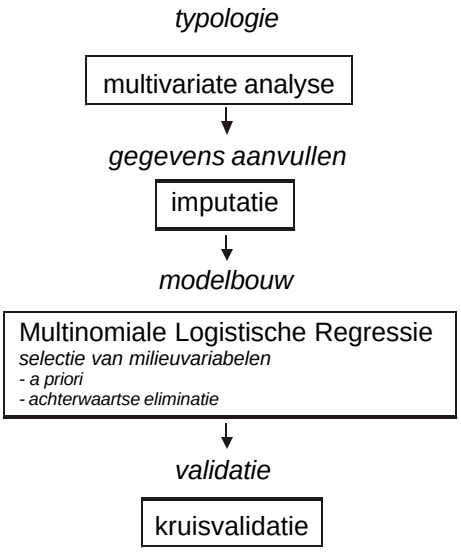
op regionaal niveau onder andere ten behoeve van beoordeling en herinrichting. Beleidsmakers op landelijk niveau wensen instrumenten onder andere ten behoeve van diagnose en verkenning. Ook de natuurbeheerders hebben hun specifieke wensen voor gebruik bij monitoring en beheer. De bouw van het netwerk en het afleiden van instrumenten ten behoeve van de praktijk van beleid en beheer, water en natuur en nationaal en regionaal zijn dus verschillende acties.

De basis van het netwerk vormt een typologie gebaseerd op de analyse van macrofauna en milieuvariabelen uit het gehele land. Een beek- of sloottype wordt gekarakteriseerd door een complex van milieuvariabelen en het voorkomen van een bepaalde levensgemeenschap (soortencombinatie). Het belangrijkste aspect in het opstellen van het netwerk is het extraheren van de sturende factoren (milieuvariabelen). Welke factoren zorgen ervoor dat een bepaalde levensgemeenschap in een beek of sloot voorkomt? Factoren zijn te verdelen in natuurlijke factoren en beïnvloedingsfactoren. Natuurlijke factoren veroorzaken ecologische verschillen en zijn van belang voor het bepalen welke typen van nature voorkomen. Beïnvloedingsfactoren zoals eutrofiëring, intensief beheer et cetera bepalen waar op de gradiënt van een bepaalde beïnvloeding een type zich bevindt: het ontwikkelings- of beïnvloedingsstadium. Kennis van deze factoren biedt de mogelijkheid de wateren te beheren en te sturen in een gewenste richting.

3 Methoden en gegevens

3.1 Inleiding

Voor de ontwikkeling van de voorspellingsmodellen voor beken en sloten is gebruikt gemaakt van de typologieën van beken en sloten zoals ontwikkeld door Alterra (Verdonschot & Nijboer 2002, Nijboer & Verdonschot 2002). In paragraaf 3.2 is in het kort uiteengezet hoe deze typologieën tot stand zijn gekomen.
In figuur 3.1 is schematische weergegeven hoe het proces doorlopen om te komen tot een model, is opgebouwd. Dit geeft tevens de paragraafindeling van dit hoofdstuk weer.



Figuur 3.1 Stroomschema gegevensbewerking ten behoeve van modelbouw en validatie

Om een optimaal model te ontwikkelen zijn met milieuvariabelen gevulde gegevensmatrices noodzakelijk. De bestanden met milieuvariabelen van beken en sloten zijn daarom compleet gemaakt op basis van zogenaamde imputatie (paragraaf 3.3).
Het in de eerdere fase van RISTORI ontwikkelde multinomiale logistische regressie model is in paragraaf 3.4 kort samengevat. In paragraaf 3.5 is kort aandacht besteed aan de selectie van milieuvariabelen, hetgeen onderdeel uitmaakt van de modelbouw.
Tenslotte is in paragraaf 3.6 uiteengezet hoe de modellen zijn geevalueerd met behulp van kruisvalidatie.

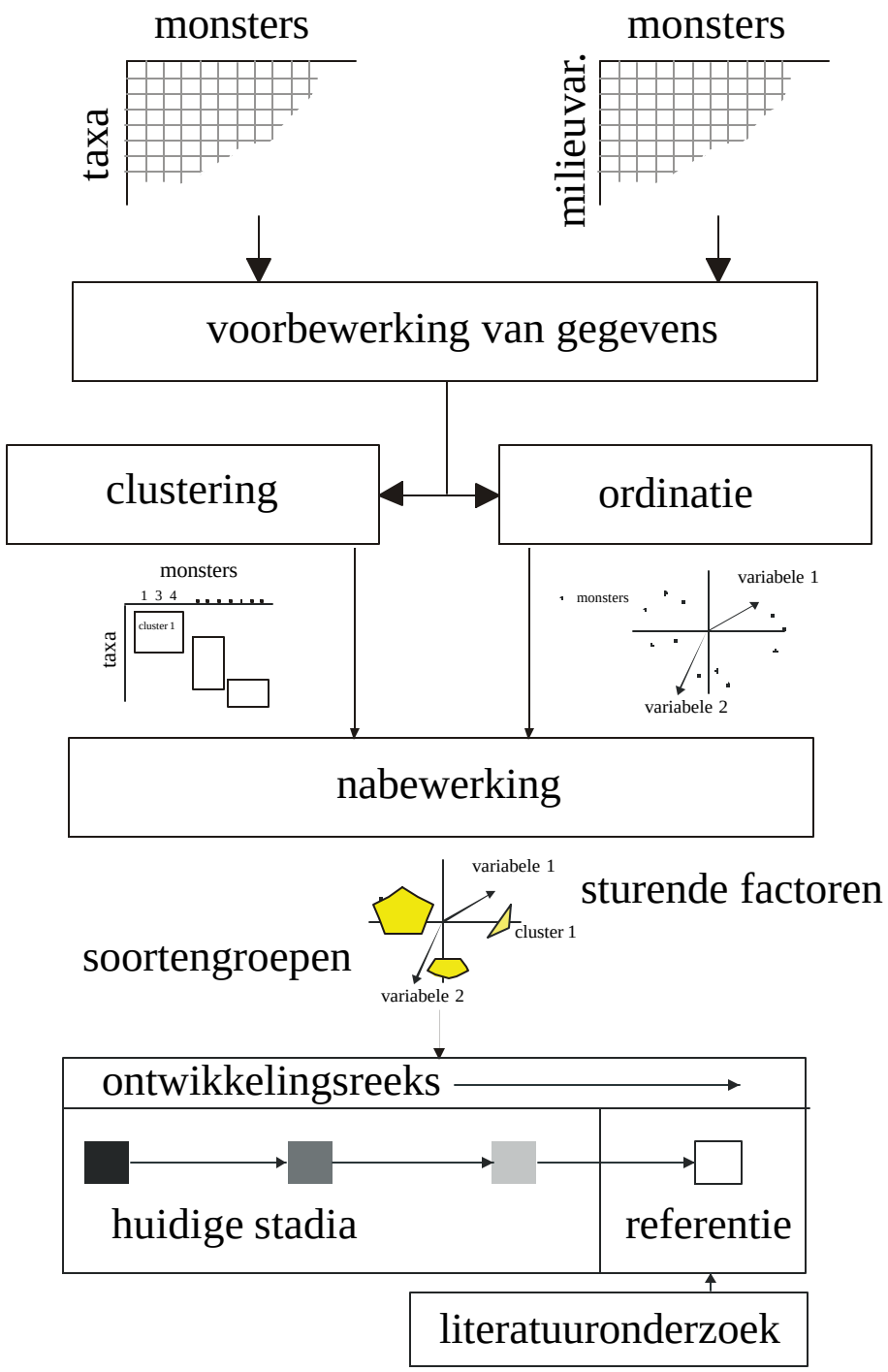
3.2 Beschrijving beken en sloten gegevens

3.2.1 Inleiding

In 1998 is gestart met het opbouwen van gegevensbestanden voor beken en sloten in Nederland. Als basis voor het ontwikkelen van instrumentarium voor ecologisch waterbeheer (bijvoorbeeld evalueren, waarderen en voorspellen) zijn daarna op basis van de gegevensbestanden ecologisch-typologische netwerken voor beken en sloten opgesteld. Deze netwerken dienen als basis voor het toekomstige (water- en natuur-) beleid en beheer van beken en sloten op nationaal, provinciaal en regionaal niveau.

3.2.2 Gegevensverzameling en -analyse

Gegevens van macrofaunabemonsteringen, vegetatie-opnamen en fysisch-chemische metingen zijn verzameld bij de waterbeheerders. Alle gegevens zijn opgeslagen in ACCESS databases. Bijna alle regio's van Nederland zijn in de databases vertegenwoordigd. Voor de sloten ontbreken Limburg en de Achterhoek. In deze regio's richt de waterbeheerder zich voornamelijk op stromende wateren en vennen. Voor de beken ontbreken de westelijke regio's omdat zich daar geen beken bevinden. Het aantal monsters per regio verschilt sterk, zo ook het aantal taxa.
Alle taxoncodes zijn gecontroleerd. Indien afwijkend van de standaard Alterra macrofaunacodelijst zijn de codes omgezet in deze standaardcodes. Onbekende codes zijn opgezocht. Alle macrofaunamonsters zijn gestandaardiseerd naar bemonstering met een standaard macrofaunanet over een lengte van 5 meter. Voor de analyses was het noodzakelijk de oorspronkelijke taxa in de gegevensbestanden taxonomisch op elkaar af te stemmen. Dit is afzonderlijk voor de beken en sloten gebeurd.
De milieuvariabelen zijn voor zowel de beken als de sloten pas na de eerste analyseronde geselecteerd. De selectie van milieuvariabelen is voor de sloten uitgevoerd tezamen met de betrokken waterbeheerders tijdens een workshop. Voor de beken zijn de variabelen in bilaterale overleggen geselecteerd en daar waar nodig en mogelijk aangevuld. Eenheden en klassen zijn, daar waar nodig, op elkaar afgestemd, opnieuw voor beken en sloten afzonderlijk.
Alle monsters van sloten die meer dan 15 m breed zijn of meer dan 1.50 meter diep zijn uit het gegevensbestand verwijderd. Voor de beken is geen aanvullend criterium opgenomen.
Het stroomschema in figuur 3.2 geeft in grote stappen de methodiek weer om van ruwe gegevens te komen tot typologieën voor beken en sloten in Nederland.
Op de beken en slotengegevens zijn afzonderlijk multivariate analyses uitgevoerd. Clustering heeft tot doel het indelen van de monsterpunten in verschillende groepen met monsters met een gelijkende soortensamenstelling. Dit is uitgevoerd met het programma FLEXCLUS.
Om de variatie binnen een gegevensbestand te bepalen is een DCA (Detrended Correspondence Analysis) op segmenten uitgevoerd. Resultaat: de gradiëntlengte van verschillende ordinatieassen.



Figuur 3.2 Stroomschema gegevensanalyse ten behoeve van de ontwikkeling van een typologie (naar Nijboer & Verdonschot 2002 en Verdonschot & Nijboer 2002)

Ordinatie plaatst, weergegeven in een tweedimensionaal ordinatiediagram, op elkaar gelijkende monsters (wat betreft soortensamenstelling) bij elkaar in een multidimensionele ruimte. Monsters die van elkaar verschillen komen juist ver uit elkaar te liggen. De monsters zijn ten opzichte van elkaar geplaatst in het ordinatiediagram, waarin de milieufactoren zijn geprojecteerd. Vervolgens zijn de clusters in de diagrammen ingetekend.
Typerende soorten zijn de soorten die het verschil tussen alle levensgemeenschappen binnen een gegevensbestand uitmaken. Voor deze soorten zijn typerende gewichten berekend met het programma NODES.

3.3 Inschatten van ontbrekende waarnemingen

3.3.1 Imputation: het compleet maken van het gegevensbestand

Onder "imputation" wordt verstaan het compleet maken van een gegevensbestand door het invullen van de ontbrekende waarnemingen op basis van een model. Dit is slechts toe te passen onder de veronderstelling van "missing at random", kortweg MAR. De volgende heuristische definitie van deze veronderstelling is te vinden op pagina 10 van Schafer (1997): "Let U and V be any two variables. Suppose that we restrict attention to units for which U is observed and equal to a specific value, say u. Missing at random means that among these units with $U=u$, the distribution of V is, apart from ordinary sampling variability, the same among the cases for which V is observed as it is among the cases for which V is missing". Een sterkere aanname is "missing completely at random" (MCAR), waarbij verondersteld wordt dat de ontbrekende waarnemingen een random steekproef vormen. Helaas is er geen methodologie om de MAR veronderstelling te controleren. Er is daarom vanuit gegaan dat aan de MAR veronderstelling voldaan wordt.
In de klassieke imputation methode van Rubin (Little & Rubin 1987, Rubin 1996) wordt allereerst een model opgesteld voor de complete gegevens ($Y_{observed}$, $Y_{missing}$). De parameters van dit model (zeg θ) worden vervolgens geschat via maximum likelihood, veelal met behulp van het EM-algorithme. Voor continue gegevens met ontbrekende waarnemingen kan bijvoorbeeld de multivariaat normale verdeling gebruikt worden als model. De parameter θ bestaat dan uit een vector van gemiddelden en een variantie-covariantie matrix. Gegeven het model én parameterschattingen T voor θ kunnen verschillende methoden gebruikt worden om te imputeren:

1. Imputation door het conditionele gemiddelde van ($Y_{missing} \mid Y_{observed}$). Hierin worden ontbrekende waarnemingen in een unit vervangen door het conditionele gemiddelde gegeven de geobserveerde waarnemingen in de unit én de parameterschattingen T. Voor de compleet gemaakte gegevensbestand geldt dan dat de gemiddelden zuivere schatters zijn, maar de varianties worden onderschat. Stel bijvoorbeeld dat $x \sim N(m, s^2)$ en dat van 100 waarnemingen er 10 ontbreken (completely at random). De conditionele verdeling van elke ontbrekende waarneming is dan $N(\hat{m}, \hat{s}^2)$, met de gebruikelijke schatters op basis van de 90

waarnemingen. Het conditionele gemiddelde voor elk van de ontbrekende waarnemingen is dus $\hat{m}$, en dat wordt in deze methode ingevuld voor de 10 ontbrekende waarnemingen. Na invullen is het gemiddelde van de 100 waarnemingen nog steeds $\hat{m}$, maar de variantie geschat uit de 100 waarnemingen wordt onderschat. Deze methode is dus correct als het alléén gaat om het schatten van het gemiddelde. Toetsen en betrouwbaarheidsintervallen zijn echter te optimistisch.

2. Imputation door één enkele random trekking uit ($Y_{missing} \mid Y_{observed}$) conditioneel op T. Hierin worden ontbrekende waarnemingen in een unit vervangen door een trekking uit de verdeling van de ontbrekende waarnemingen gegeven de geobserveerde waarnemingen in de unit én de parameterschattingen T. Voor het bij 1. besproken voorbeeld geldt dan dat, op basis van het complete gegevensbestand, zowel het gemiddelde als de variantie zuiver geschat worden, conditioneel op T. Als alleen het gemiddelde geschat moet worden, dus zonder betrouwbaarheidsinterval en/of toetsen, dan is deze methode minder efficiënt dan 1. Door immers te trekken uit de conditionele verdeling wordt extra spreiding geïntroduceerd.
3. meerdere random imputations conditioneel op T. Dan wordt een gering aantal (10 is in het algemeen genoeg) imputations voor de ontbrekende waarnemingen gesimuleerd analoog aan 2. De zeg 10 complete gegevensbestanden worden separaat geanalyseerd en de parameterschattingen van deze 10 analyses worden dan gecombineerd tot één enkele parameterschatting. In deze methode wordt rekening gehouden met variatie als gevolg van het trekken uit de conditionele verdeling van de ontbrekende waarnemingen. Er wordt echter geen rekening gehouden met variatie als gevolg van het schatten van de parameter θ. Immers alle trekkingen zijn conditioneel op T.
4. multipele imputation. Hierin wordt variatie als gevolg van het schatten van θ ook verdisconteerd. Met een Bayesiaanse methode wordt, gegeven een prior voor θ, allereerst een trekking gegenereerd uit de posterior van θ. Gegeven deze trekking wordt een nieuwe imputation gegenereerd. Ook dit wordt een aantal malen herhaald waarbij steeds nieuwe trekkingen uit de posterior gebruikt worden. De individuele schattingen uit de separate analyses worden weer gecombineerd.

Methode 4 is veruit superieur omdat daarin **alle** variatie als gevolg van het simuleren van de ontbrekende waarnemingen verdisconteerd wordt. Voor het combineren van de resultaten van de separate complete gegevensbestanden wordt de rekenregel gebruikt dat de marginale variantie gelijk is aan de variantie van de conditionele verwachting + de verwachting van de conditionele variantie. Hiervoor zijn dus de separate parameterschattingen én hun standaardafwijkingen nodig.
Multiple imputation wordt uitgebreid besproken door Schafer (1997). Schafer (http://www.stat.psu.edu/~jls/) heeft hiervoor ook de volgende sets van S-Plus routines geschreven:

A. NORM - multiple imputation of multivariate continuous data (normsp40.exe, 49K). Het complete data model voor het inschatten van de ontbrekende waarnemingen is hierin de multivariaat normale verdeling.

B. CAT - multiple imputation of multivariate categorical data (catsp40.exe, 45K). Het complete data model voor het inschatten van de ontbrekende waarnemingen is hierin een log-lineair model.
C. MIX - multiple imputation of mixed continuous and categorical data (mixsp40.exe, 55K). Het complete data model voor de categorische gegevens is hierin een log-linear model en voor de continue gegevens een multivariaat normale verdeling. De gemiddelden van de multivariaat normale verdeling kunnen hierbij afhangen van de categorische gegevens. De variantie-covariantie matrix hangt in dit model niet af van de categorische gegevens; er wordt een gemeenschappelijke variantie-covariantie matrix gebruikt. Dit model wordt ook wel het "general location model" genoemd.
D. PAN - multiple imputation of multivariate panel or clustered data (pansp40.exe, 67K).

3.3.2 Imputation: een eenvoudig voorbeeld

Beschouw als eenvoudig voorbeeld de drie continue variabelen Y, X1 en X2, elk met enkele ontbrekende waarnemingen (tabel 3.1).

Tabel 3.1 Voorbeeldtabel met drie continue variabelen Y, X1 en X2 en enkele ontbrekende waarnemingen (*)

Y	*	*	55	78	*	65	43	69	61	*	35	55	63	76	*	43	61	*
X1	62	11	*	63	*	51	*	59	45	58	21	26	50	*	47	32	59	50
X2	*	29	*	66	49	49	37	*	53	51	*	54	55	53	52	*	50	*

Gevraagd wordt schattingen van de regressiecoëfficiënten van de multipele regressie van Y op X1 en X2. Er zijn slechts 6 waarnemingen compleet. Een geschikt model om de ontbrekende gegevens in te vullen is de trivariate normale verdeling voor (Y, X1, X2). Deze verdeling heeft 9 parameters: 3 gemiddelden, 3 varianties en 3 correlaties (tabel 3.2) . De S-Plus routines van Schafer zijn allereerst gebruikt om met EM de parameters van de trivariate normale verdeling te schatten (zie ook het programma in bijlage 1B).

Tabel 3.2 Gemiddelden, standaardafwijkingen en correlatiematrix van de drie continue variablen uit tabel 3.1

	gemiddelde	**stand. afw.**	**correlatiematrix**		
			Y	**X1**	**X2**
Y	58.28864	13.278060	1.0		
X1	45.84912	16.087091	0.9189137	1.0	
X2	48.76934	9.093728	0.9066616	0.7380220	1.0

Vervolgens zijn schattingen voor de regressiecoëfficiënten berekend op basis van 10, 100, 1000 en 10000 imputations, via methode 4. uit de vorige paragraaf. Daarbij is ook variatie als gevolg van het schatten van de 9 parameters van de trivariate normale verdeling verdisconteerd. Hiervoor is de standaard niet-informatieve prior voor de trivariate normale verdeling gebruikt, zoals geïmplementeerd in de S-Plus routines. De schattingen voor de regressiecoëfficiënten zijn opgesomd in tabel 3.3, waarbij de standaardafwijkingen tussen haakjes gegeven zijn.

Tabel 3.3 De schattingen voor de regressiecoëfficiënten met de standaardafwijkingen tussen haakjes.

methode	**coëff. X1**	**coëff. X2**
regressie op 6 complete units	0.351 (0.120)	0.724 (0.255)
10 imputations	0.428 (0.122)	0.747 (0.268)
100 imputations	0.465 (0.114)	0.730 (0.178)
1000 imputations	0.453 (0.108)	0.737 (0.188)
10000 imputations	0.455 (0.109)	0.728 (0.189)

De schattingen voor de regressiecoëfficiënten zijn na 10 imputations al redelijk stabiel. De standaardafwijkingen zijn pas redelijk stabiel na 100 imputations. Na imputation is de schatting voor de regressiecoëfficiënt voor X1 behoorlijk verschillend in vergelijking met de regressie op de 6 complete waarnemingen. De S-Plus routines geven ook nog een schatting van het percentage ontbrekende informatie. Dit percentage is voor de regressiecoëfficiënten voor X1 en X2 respectievelijk 54% en 53% (berekend over 1000 imputations).
De likelihood kan meerdere lokale optima hebben en dus is het EM-algorithme gestart vanaf diverse beginwaarden voor de parameters (zie het S-Plus programma onderaan bijlage 1B). Deze beginwaarden zijn verkregen door voor de parameters een niet-informatieve prior te specificeren en vervolgens één trekking te doen uit de posterior van de parameters. Vanuit deze trekking is het EM-algorithme dan opnieuw opgestart. In de volgende stap is vanuit de verkregen trekking uit de posterior een nieuwe trekking gegenereerd van waaruit opnieuw het EM-algorithme is gestart. Op deze manier is het EM-algorithme vanuit 50 verschillende beginwaarden gestart; het bleek dat het EM-algorithme steeds naar dezelfde oplossing convergeerde. Het lijkt er dus op dat voor dit gegevensbestand de likelihood één globaal optimum heeft.
De selectie van variabelen is erg tijdrovend en daarom is van multipele imputation afgezien. Alleen de hiervoor besproken varianten 1 (conditioneel gemiddelde) en 2 (één random imputation) zijn daarom in aanmerking gekomen.
Het uiteindelijke doel van dit project is het schatten van de kansen op de verschillende cenotypen gegeven de relevante milieuvariabelen. Daarbij wordt, in ieder geval tot nu toe, geen variatie in geschatte kansen gerapporteerd. Voor puntschattingen is methode 1 efficiënter dan methode 2, en daarom is voor methode 1 gekozen. Daarbij moet wel bedacht worden dat in de modelselectie herhaaldelijk getoetst wordt of milieuvariabelen significant zijn of niet, en deze toetsen zijn onder methode 1 mogelijk te optimistisch. Dit effect wordt enigszins gecompenseerd door een gewogen multinomiale regressie uit te voeren, met als gewicht [(aantal niet ontbrekende milieuvariabelen)/aantal aanwezige milieuvariabelen]. Een monster zonder ontbrekende milieuvariabelen doet dan volledig mee in de regressie, terwijl monsters waarvoor alle milieuvariabelen ontbreken niet meedoen in de regressie.
De ontbrekende gegevens zijn uiteindelijk ingeschat met behulp van het zogenaamde "general location model". Dit model specificeert:

1. een log-lineair model voor de aantallen waarnemingen in een kruistabel uitgesplitst naar alle variabelen, en
2. conditioneel op de variabelen, een multivariaat normale verdeling voor de continue variabelen, waarbij gemiddelden af kunnen hangen van niveaus van de factoren.

De variantie-covariantie matrix hangt in dit model niet af van de factoren, maar is gemeenschappelijk. Dit model is uitgebreid beschreven in Hoofdstuk 9 van Schafer

(1997) en kan aangepast worden met de MIX software van Schafer. De MIX software heeft overigens als eigenaardigheid dat de continue variabelen allen positief moeten zijn. Een goed "general location model" kan gevonden worden met behulp van modelselectie in twee stappen:

1. selectie van factoren en hun interacties in het log-lineaire model, en
2. selectie van factoren en hun interacties waarvan de gemiddelden van de continue variabelen afhangen.

Daarbij geldt dat het cenotype een belangrijke term kan zijn in het log-lineaire model én in het model voor de gemiddelden. Aangezien het cenotype een factor is op zoveel niveaus als er cenotypen zijn, met zeer weinig waarnemingen voor sommige cenotypen, is in beide modellen de hoofdgroep gebruikt als verklarende term in plaats van het cenotype.

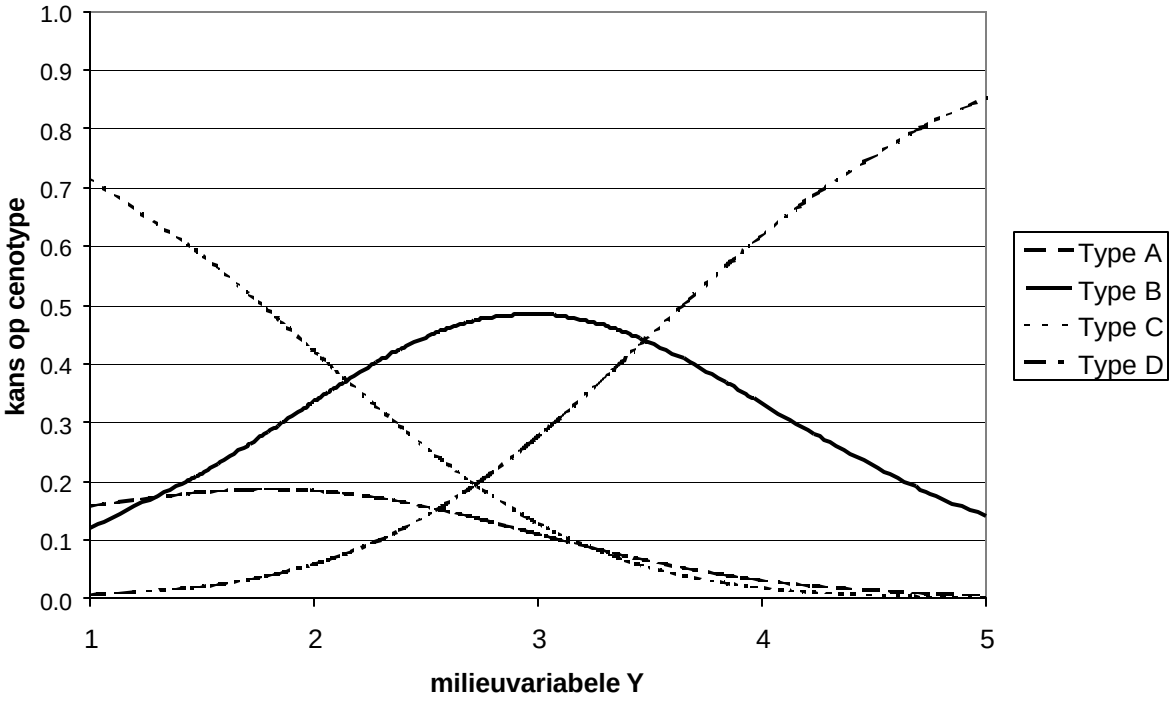
3.4 Multinomiale logistische regressie-analyse

In het effectmodel is een complex van milieuvariabelen gebruikt om het cenotype te voorspellen. Uit eerdere modellen waarin meerdere milieuvariabelen gelijktijdig zijn gebruikt, is multinomiale logistische regressie-analyse ontstaan (Hosmer & Lemeshow 1989). Bij multinomiale logistische regressie wordt de kans op een cenotype, gegeven een set milieuvariabelen, rechtstreeks gemodelleerd. Voor een nieuw monster met bekende milieuvariabelen wordt dan de kans op elk cenotype voorspeld, bijvoorbeeld een voorspelde kans van 0.6 op type A, 0.4 op type B en 0.0 op alle andere cenotypen (figuur 3.3). Voor het multinomiale logistische model is het hele scala aan methoden en technieken voor het gegeneraliseerde lineaire model beschikbaar, waarbij bijvoorbeeld gedacht kan worden aan subset selectie van milieuvariabelen. Tevens wordt er niet uitgegaan van een multivariaat normale verdeling van de milieuvariabelen, zoals bij discriminant analyse. Dit is een voordeel, omdat er nominale variabelen in de gegevensbestanden voorkomen.
Een model gebaseerd op multinomiale logistische regressie bevat regressiecoëfficiënten voor de milieuvariabelen, die uit de gegevens geschat zijn.

3.5 Selectie van voorspellende milieuvariabelen

Er zijn verschillende manieren om tot een voorspellingsmodel te komen. Het eenvoudigste is om alle milieuvariabelen in het model op te nemen. Dit wordt het ‘volledige model’ genoemd. Maar in de beschikbare gegevens is het aantal milieuvariabelen zo groot, dat een vorm van selectie van variabelen in de regressie noodzakelijk is. Immers, in de gegevens zijn ook onderling gecorreleerde milieuvariabelen aanwezig en dat geeft een instabiele schatting van het regressiemodel en daarmee onnauwkeurige voorspellingen. Met behulp van selectie van variabelen wordt geprobeerd een zo klein mogelijk model te vinden dat bijna evenveel verklaart als het volledige model. In de regel geeft dat een beter voorspellingsmodel. Een veel gebruikte methode van selectie van variabelen is het aanpassen van alle mogelijke modellen, dat wil zeggen alle mogelijke combinaties van

variabelen, om vervolgens het 'beste' model te kiezen. Deze methode is echter zeer rekenintensief en daarom niet altijd haalbaar. Alternatieven zijn selectie van variabelen op basis van 'a priori' kennis en (1) voorwaartse selectie of (2) achterwaartse eliminatie van milieuvariabelen.
In de analyses is gekozen voor selectie op basis van a priori kennis gevolgd door een eenvoudige achterwaartse eliminatie van variabelen.



Figuur 3.2 De kans op voorkomen van vier cenotypen A, B, C en D langs een milieuvariabele Y. De som van de kans op type A + B + C + D is gelijk aan 1 voor een bepaalde waarde van de milieuvariabele Y. In multinomiale regressie wordt niet met één maar met meerdere milieuvariabelen tegelijk gerekend

3.6 Evaluatie van de modellen door middel van kruisvalidatie

Resubstitutie of terugvoorspellen geeft in het algemeen een te optimistisch beeld van de voorspelkracht van een model. Immers, dezelfde gegevens worden dan gebruikt om het model aan te passen én om de voorspelkansen te berekenen. Kruisvalidatie is daarom beter om de voorspelkracht in kaart te brengen. Er is kruisvalidatie met aselecte groepen van 10 waarnemingen toegepast, maar ook met individuele waarnemingen. Voor kruisvalidatie met aselecte groepen wordt allereerst de eerste groep waarnemingen weggelaten, het model wordt vervolgens aangepast aan de overige waarnemingen en het aangepaste model wordt gebruikt om een voorspelling te berekenen voor de weggelaten waarnemingen. Op dezelfde manier worden alle groepen van waarnemingen successievelijk weggelaten en worden voorspelkansen verkregen. De selectie van variabelen, zoals beschreven in de vorige paragraaf, was geen onderdeel van de kruisvalidatie.

4 Ontwikkeling bekenmodel

4.1 Beschrijving bekengegevens

Het bekengegevensbestand bevat 617 beken. ‘A priori’ zijn op basis van de multivariate analyse resultaten 21 milieuvariabelen geselecteerd. Voor 54 beken ontbreken alle milieuvariabelen en deze zijn uit het gegevensbestand verwijderd; er resteerde een gegevensbestand met 563 beken en 21 milieuvariabelen.
Zeven milieuvariabelen zijn nominaal, de resterende 14 variabelen zijn continue (tabel 4.1). In tabel 4.1 is voor de 7 nominale factoren het aantal ontbrekende waarnemingen (Nmissing), en de scores 0 (N=0) en 1 (N=1) gegeven. De som van "Nmissing", "N=0" en "N=1" is steeds gelijk aan het aantal beken (563). De kolom "P=0" is gelijk aan de fractie beken met score 0. De continue variabelen zijn logaritmisch of logit getransformeerd aan de hand van een Box-plot. Percentages zijn allen logit getransformeerd. De variabelen die in het eerder opgestelde EKOO-model logaritmisch getransformeerd zijn (totaal fosfaat, nitraat, ammonium, breedte, diepte, stroomsnelheid, verval, elektrisch geleidend vermogen) zijn ook hier logaritmisch getransformeerd. Box-plots van de oorspronkelijke continue variabelen en de getransformeerde continue variabelen zijn gegeven in bijlage 1A. De nomiale milieuvariabelen voorjaar, zomer, herfst en winter zijn gecombineerd tot een factor seizoen met de 3 niveaus, namelijk voorjaar, winter en zomer/herfst. Deze factor seizoen is in eerste instantie gebruikt. Later zijn enekele seizoenen toch apart nomiaal meegenomen.
De 563 beken zijn op basis van de multivariate analyses onderverdeeld in 25 cenotypen:

Cenotype nr.	Omschrijving
8	Droogvallende, zure, natuurlijke bovenloopjes
7	Droogvallende, bijna natuurlijke bovenloopjes
12	(Droogvallende), zure, half-natuurlijke bovenloopjes-lopen
27	Zwak zure, stromende, half-natuurlijke bovenloopjes
2	Zwak zure, natuurlijke bovenlopen
21	Snel stromende, bijna natuurlijke bovenloopjes
3b	Snel stromende, half-natuurlijke bovenlopen
3a	Snel stromende, half-natuurlijke boven-middenlopen
25	Snel stromende, belaste boven-middenlopen
24b	(Snel) stromende, natuurlijke bovenloopjes-lopen
24c	(Snel) stromende, bijna natuurlijke bovenloopjes
24a	(Snel) stromende, bijna natuurlijke bovenlopen
26	Half-natuurlijke bovenlopen
20	Natuurlijke midden-benedenlopen
16	Half-natuurlijke middenlopen
19	Half-natuurlijke benedenlopen
1	Plantenrijke, genormaliseerde bovenlopen
9	Belaste bovenlopen

Cenotype nr.	Omschrijving
31	Belaste, genormaliseerde middenlopen
6	Belaste midden-benedenlopen
14b	Belaste, plantenrijke midden-benedenlopen
14a	Belaste, genormaliseerde riviertjes
15	Sterk belaste bovenloopjes-lopen
13	Sterk belaste, stromende boven-middenlopen
10	Sterk belaste, langzaam stromende boven-middenlopen

Tabel 4.1 Overzicht van kenmerken van de voor het te ontwikkelen bekenmodel geselecteerde milieuvariabelen

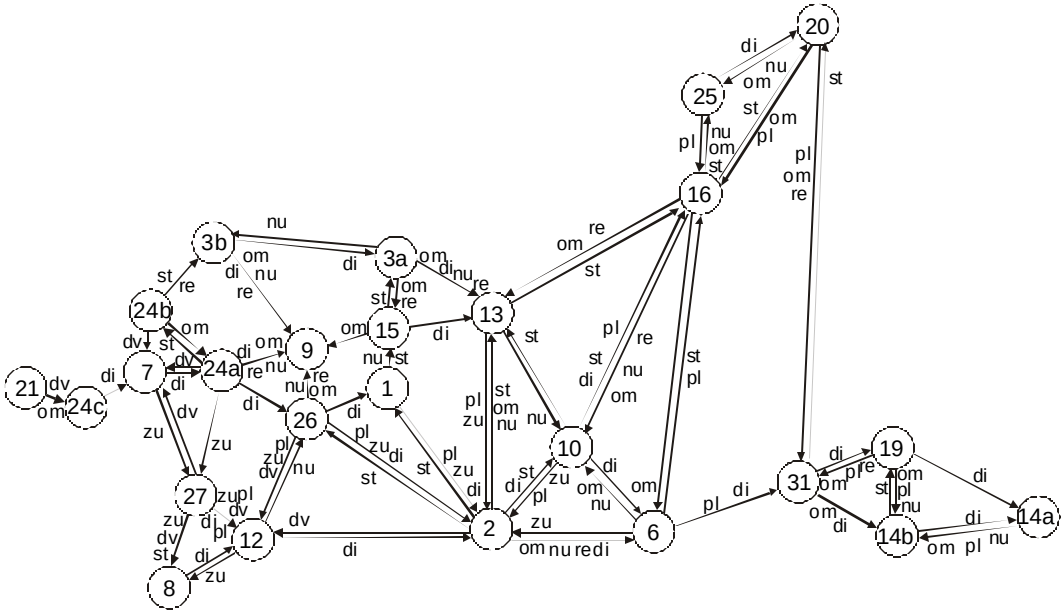
nr	code milieuvariabele	type	Nmissing	N=0	N=1	P=0
1	meandering	nominaal	170	221	172	0.56
2	natuurlijk dwarsprofiel	nominaal	101	240	222	0.52
3	grondgebruik natuur	nominaal	132	238	193	0.55
4	permanentie	nominaal	133	42	388	0.10
5	stuwing	nominaal	142	317	104	0.75
6	voorjaar	nominaal	39	294	230	0.56
7	winter	nominaal	39	513	11	0.98

nr	code milieuvariabele	type	Nmissing	trans-formatie	gemid-delde	mini-mum	maxi-mum
8	beschaduwing	continue	86	logit	-1.67	-5.30	5.30
9	substraat slib	continue	105	logit	-3.27	-5.30	5.30
10	diepte	continue	79	log	-1.26	-4.61	1.39
11	stroomsnelheid	continue	114	log (+0.001)	-2.22	-6.91	0.10
12	breedte	continue	98	log (+0.02)	0.88	-3.91	3.69
13	substraat zand	continue	105	logit	-1.02	-5.30	5.30
14	pH-mediaan	continue	82	-	7.18	4.30	9.00
15	chloride-mediaan	continue	95	log	3.65	1.95	5.45
16	egv-mediaan	continue	118	log	5.99	2.97	7.14
17	ammonium-90-percentiel	continue	102	log	-0.48	-4.61	3.51
18	Nkjeldal-90-percentiel	continue	156	log	0.86	-2.12	3.75
19	O_2-gehalte-10-percentiel	continue	165	-	7.21	0.88	16.90
20	totaal fosfaat-90-percentiel	continue	96	log	-1.19	-3.96	2.31
21	nitraat-10-percentiel	continue	98	log (+0.001)	0.68	-6.91	3.87

Deze 25 cenotypen zijn in een netwerk geplaatst (figuur 4.1). In het netwerk wordt de samenhang tussen de typen duidelijk. Het netwerk bevat een overgang van klein naar groot gaande van links naar rechts in de figuur en van snel naar langzaam stromend gaande van boven naar onder.

De 25 cenotypen zijn vervolgens op basis van expert judgement ingedeeld in 6 hoofdgroepen A tot en met F. De hoofdgroep indeling is gebruikt om het aantal parameters in het multinomiale logistische regressiemodel te beperken: er zijn aparte regressie modellen aangepast voor de hoofdgroepen en voor de cenotypen binnen

elke hoofdgroep. De hoofdgroepen A tot en met F bestaan uit de in tabel 4.2 gegeven cenotypen, met tussen haakjes de aantallen waarnemingen.



Figuur 4.1 Het beekcenotypennetwerk met de overgang van klein naar groot gaande van links naar rechts in de figuur en van snel naar langzaam stromend gaande van boven naar onder. De pijlen wijzen in de richting van de toename van een factor. Legenda: om=organisch materiaal, di=dimensie, pl=vegetatie, st=stroomsnelheid, nu=nutriënten, zu=zuurgraad, dv=droogval, re=regulatie

Tabel 4.2 De hoofdgroepen A tot en met F en bijbehorende cenotypen, met tussen haakjes de aantallen waarnemingen

hoofdgroep		**cenotypen**								
A	(159)	3a (39)	3b (6)	24a (24)	24b (39)	24c (43)	25 (8)			
B	(270)	1 (14)	6 (89)	10 (84)	14a (10)	14b (9)	16 (3)	19 (48)	20 (4)	31 (9)
C	(32)	7 (29)	27 (3)							
D	(29)	2 (15)	8 (2)	12 (12)						
E	(60)	9 (22)	13 (11)	15 (16)	26 (11)					
F	(13)	21 (13)								

In onderstaande tabellen 4.3 en 4.4 zijn de aantallen beken in elke hoofdgroep en in elk cenotype uitgesplitst naar aantallen met respectievelijk géén, 1-5, 6-10, 11-15 en 16-21 ontbrekende milieuvariabelen.

Tabel 4.3 Het aantal beken met aantal ontbrekende milieuvariabelen per hoofdgroep (zie tabel 4.2)

hoofdgroep	**aantal beken met ontbrekende milieuvariabelen**					**totaal**
	0	1-5	6-11	11-15	16-21	
A	47	51	32	27	2	159
B	96	102	31	26	15	270
C	20	7	3	2	-	32
D	2	18	4	5	-	29
E	6	44	6	3	1	60
F	-	11	1	1	-	13
totaal	**171**	**233**	**77**	**64**	**18**	**563**

Tabel 4.4 Het aantal beken met aantal ontbrekende milieuvariabelen per cenotype

cenotype	aantal beken met ontbrekende milieuvariabelen					totaal
	0	1-5	6-11	11-15	16-21	
1	3	4	2	5	-	14
2	1	8	3	3	-	15
3a	14	9	4	11	1	39
3b	1	2	1	2	-	6
6	22	32	9	13	13	89
7	20	5	2	2	-	29
8	1	1	-	-	-	2
9	4	14	1	3	-	22
10	32	40	6	6	-	84
12	-	9	1	2	-	12
13	2	9	-	-	-	11
14a	7	-	2	1	-	10
14b	-	7	-	-	2	9
15	-	12	3	-	1	16
16	1	2	-	-	-	3
19	27	8	12	1	-	48
20	3	1	-	-	-	4
21	-	11	1	1	-	13
24a	4	15	5	-	-	24
24b	7	12	6	13	1	39
24c	21	7	14	1	-	43
25	-	6	2	-	-	8
26	-	9	2	-	-	11
27	-	2	1	-	-	3
31	1	8	-	-	-	9
totaal	**171**	**233**	**77**	**64**	**18**	**563**

Een aantal cenotypen komt in zeer lage aantallen voor: 3b (6x), 8 (2x), 16 (3x), 20 (4x) en 27 (3x). Aangezien dit zeer interessante cenotypen zijn is weglaten uit het gegevensbestand geen optie. Gelukkig geldt voor de beken van deze cenotypen dat het aantal ontbrekende milieuvariabelen niet extreem hoog is. Het grote aantal ontbrekende milieuvariabelen bemoeilijkt de analyse van de bekengegevens aanzienlijk. In de huidige implementaties van modelselectie kunnen slechts beken gebruikt worden waarvoor alle milieuvariabelen beschikbaar zijn. Aangezien de ontbrekende waarnemingen kriskras door de gegevens staan zijn er slechts 171 beken met een complete set milieuvariabelen. Een eenvoudige methode om het gegevensbestand compleet te maken is het verwijderen van één of meerdere milieuvariabelen. Door respectievelijk 1, 2, 3, 4 en 5 milieuvariabelen weg te laten komt het maximale aantal complete beken op respectievelijk 199, 231, 259, 278 en 292. Daardoor wordt het gegevensbestand te klein; daarom is gekozen voor het compleet maken van het gegevensbestand via "imputation" (paragraaf 3.3).

4.2 Imputation toegepast op de bekengegevens

De bekengegevens bevatten 6 nominale, 1 categorische en 14 continue milieuvariabelen. De ontbrekende gegevens zijn ingeschat met behulp van het

zogenaamde "general location model". Het complete gegevensbestand is geanalyseerd met behulp van multinomiale logistische regressie met selectie van milieuvariabelen, zowel op hoofdgroepniveau als binnen de 6 hoofdgroepen.
In de eerste stap is met een traditionele methode een log-lineair model geselecteerd. Als startpunt voor het log-lineaire model is gekozen voor een analyse op de deelset van beken waarvoor alle factoren beschikbaar zijn. Voor deze deelset is een kruistabel gemaakt met aantallen waarnemingen uitgesplitst naar hoofdgroep én de 6 nominale milieuvariabelen. Deze kruistabel is geanalyseerd met een log-lineair model, waarbij modelselectie is gebruikt om te bepalen welke 2-factor interacties de aantallen het best beschrijven. De 7 hoofdeffecten (hoofdgroep en 6 nominale variabelen) zaten daarbij steeds vast in het model. De modelselectie is uitgevoerd met Genstat procedure SELECT. Deze past alle mogelijke modellen aan voor maximaal 16 termen en geeft aan welke van de aangepaste modellen de beste zijn. Aangezien 7 factoren 21 2-factor interacties hebben, is SELECT iteratief aangepast. Er zijn geen 3-factor interacties aan het model toegevoegd. Als belangrijkste 2-factor interacties kwamen naar voren hoofdgroep $*$ meandering, hoofdgroep $*$ natuurlijk dwarsprofiel, hoofdgroep $*$ permanentie, hoofdgroep $*$ stuwing, meandering $*$ natuurlijk dwarsprofiel, natuurlijk dwarsprofiel $*$ grondgebruik natuur, meandering $*$ stuwing en tenslotte grondgebruik natuur $*$ permanentie. Daarnaast zitten zoals gezegd alle 7 hoofdeffecten in het model.
In de tweede stap is dit log-lineaire model verder verfijnd met de MIX software voor het "general location model". MIX gebruikt het gehele gegevensbestand om, met maximum likelihood via het EM-algorithme, de parameters van het "general location model" te schatten. MIX vereist zowel een log-lineair model alsook een model voor de gemiddelden van de continue variabelen. Als startmodel voor de gemiddelden is gekozen voor een model met alle hoofdeffecten (inclusief de hoofdgroep) én de zes 2-factor interacties tussen meandering, natuurlijk dwarsprofiel, grondgebruik natuur en stuwing. Dit startmodel voor de gemiddelden heeft in totaal $19 \times 14 = 266$ parameters; 19 omdat de hoofdeffecten in totaal 13 parameters hebben en de zes 2-factor interacties 6 parameters; 14 omdat dit het aantal continue milieuvariabelen is. Vanuit het log-lineaire model uit de eerste stap en het model voor de gemiddelden is aan het log-lineaire model steeds één 2-factor interactie toegevoegd, en wel voor alle 2-factor interacties die nog niet in het log-lineaire model opgenomen waren. Daarbij werd de likelihood ratio toets gebruikt om te toetsen of de toegevoegde term significant was. Het bleek dat de interactie hoofdgroep$*$grondgebruik natuur zeer significant was ($p<0.001$). Na toevoeging van deze term aan het model is de procedure herhaald en bleken 4 termen significant te zijn met $0.01<p<0.05$. Vanwege de beperkte significantie is geen van deze termen aan het log-lineaire model toegevoegd. Tenslotte is nog gecontroleerd of alle geselecteerde 2-factor interacties wel significant zijn. Daartoe is elke 2-factor interactie uit het model verwijderd; alle likelihood ratio toetsen waren zeer significant met $p<0.001$.
In de derde stap is, gegeven het geselecteerde log-lineaire model, een model voor de gemiddelden geselecteerd. Als startpunt is het model voor gemiddelden genomen waarbij de gemiddelden afhangen van alle hoofdeffecten (inclusief de hoofdgroep) én de zes 2-factor interacties tussen meandering, natuurlijk dwarsprofiel, grondgebruik natuur en stuwing. Aan dit model is steeds één 2-factor interactie toegevoegd, en wel

voor alle 2-factor interacties die nog niet in het model opgenomen waren. Significantie van toegevoegde termen is weer bepaald met de likelihood ratio toets. Voor relatief veel 2-factor interacties convergeerde het EM-algorithme niet, ook niet als het aantal iteratiestappen opgehoogd werd tot 2500. Deze 2-factor interacties zijn verder buiten beschouwing gelaten, hoewel met name de term hoofdgroep∗seizoen (met 140 parameters!) belangrijk leek. Achtereenvolgens zijn aan het model toegevoegd stuwing∗seizoen ($p=0.000$) en permanentie∗seizoen ($p<0.00007$). Na toevoeging van deze termen zijn er nog een aantal termen significant, echter allen met $p>0.001$. Geen van deze termen is toegevoegd. Vervolgens is getoetst of er nog 2-factor interacties verwijderd konden worden. Daarbij bleek dat de meeste vooraf geselecteerde 2-factor interacties tussen meandering, natuurlijk dwarsprofiel, grondgebruik natuur en stuwing verwijderd konden worden. Alleen de interactie natuurlijk dwarsprofiel∗ grondgebruik natuur bleef significant met $p=0.004$ en deze term is niet verwijderd. Wat resteert is het model met alle hoofdeffecten (inclusief hoofdgroep) en de interacties natuurlijk dwarsprofiel∗ grondgebruik natuur, stuwing∗seizoen en permanentie∗seizoen. Aan dit model zijn tenslotte nog één voor één alle niet opgenomen interacties toegevoegd. Geen van de termen waarvoor het EM-algorithme convergeerde bleek daarbij zeer significant ($p<0.001$); alle p-waarden waren groter dan 0.003.

Gegeven het in de derde stap geselecteerde model voor de gemiddelden kan het log-lineaire model wellicht nog verbeterd worden, en gegeven een verbeterd log-lineair model kan het model voor de gemiddelden mogelijk weer verbeterd worden, enzovoort. Bovenstaande modelselectie kost echter erg veel rekentijd, en daarom is er na de derde stap gestopt. Het geselecteerde model waarmee de ontbrekende gegevens ingevuld gaan worden is dan:

log-lineair: alle 7 hoofdeffecten (inclusief de hoofdgroep) en de 9 interacties hoofdgroep∗meandering, hoofdgroep∗natuurlijk dwarsprofiel, hoofdgroep∗grondgebruik natuur, hoofdgroep∗permanentie, hoofdgroep∗stuwing, meandering∗natuurlijk dwarsprofiel, natuurlijk dwarsprofiel∗grondgebruik natuur, meandering∗stuwing en grondgebruik natuur∗permanentie;

gemiddelden: alle 7 hoofdeffecten (inclusief de hoofdgroep) en de 3 interacties natuurlijk dwarsprofiel∗grondgebruik natuur, stuwing∗seizoen en permanentie∗seizoen.

Dit geselecteerde "general location model" is vervolgens gebruikt om de ontbrekende gegevens te vervangen door conditionele gemiddelden. De MIX software van Schafer, meer specifiek de functie imp.mix, kan gebruikt worden om, volgens imputation methode 2, één enkele random imputatie te genereren. Door een groot aantal van deze random imputations te genereren en deze vervolgens te middelen worden goede schattingen van de conditionele gemiddelden verkregen. Voor de ontbrekende categorische milieuvariabelen worden op deze manier kansen op factor niveau 1 verkregen. Stel bijvoorbeeld dat van de 100 imputations voor een ontbrekende waarneming in de variabele meandering, er 10 factor niveau 0 hebben en 90 factor niveau 1. Het gemiddelde is dan 0.9 en de ontbrekende waarneming wordt hierdoor vervangen. Ontbrekende categorische milieuvariabelen worden dus

in het algemeen niet vervangen door 0 of 1, maar door een kans. Voor de bekengegevens zijn 100.000 random imputations gebruikt om het conditionele gemiddelde te benaderen.

4.3 Modelselectie

Multinomiale logistische regressie is gebruikt om het vóórkomen van de cenotypen te relateren aan de milieuvariabelen. Hierbij zijn waarnemingen gewogen naar rato van het aantal milieuvariabelen dat niet ontbreekt. Het aantal milieuvariabelen is zo groot dat een vorm van selectie van variabelen noodzakelijk is; daarbij is gezocht naar een model met een zo gering mogelijk aantal parameters. Bij multinomiale logistische regressie op cenotype-niveau is het aantal parameters al snel groot omdat elke variabele leidt tot 24 regressiecoëfficiënten (één minder dan het aantal cenotypen). Daarom is weer gekozen voor hiërarchische modellering, waarbij aparte modellen aangepast zijn voor de hoofdgroepen en voor de cenotypen binnen de hoofdgroepen. Met 6 hoofdgroepen en maximaal 9 cenotypen binnen een hoofdgroep geeft dit een belangrijke reductie in het aantal parameters. Immers op hoofdgroepniveau heeft een variabele in het model 5 parameters, en binnen hoofdgroepen maximaal 8 parameters. Tevens kunnen verschillende milieuvariabelen op elk van de niveaus een rol spelen. Bij hiërarchische modellering blijft het aantal te schatten parameters dus steeds binnen de perken omdat het aantal klassen binnen elk niveau van de hiërarchie gering is. Een uitgebreide zorgvuldige modelselectie met alle predictoren is dan steeds mogelijk en dus ook uitgevoerd.
Met behulp van het aangepaste model zijn daarna de kansen op het voorkomen van de cenotypen berekend.
Voor een nieuw monster met een bepaalde set van milieuvariabelen worden dan eerst, via het hoofdgroepen model, kansen op elk van de hoofdgroepen berekend. Dan worden kansen op elk van de cenotypen binnen de hoofdgroepen berekend. De zo verkregen kansen worden doorvermenigvuldigd. Veronderstel bijvoorbeeld dat er twee hoofdgroepen zijn (1 en 2), met binnen hoofdgroep 1 drie cenotypen (A, B en C) en binnen hoofdgroep 2 twee cenotypen (D en E). Veronderstel verder dat voor een nieuw monster de voorspelde kans op hoofdgroep 1 en 2 gelijk is aan respectievelijk 0.8 en 0.2; dat de voorspelde kans op A, B en C binnen hoofdgroep 1 gelijk is aan respectievelijk 0.9, 0.1 en 0.0, en op D en E binnen hoofdgroep 2 gelijk aan respectievelijk 0.4 en 0.6. De kans op elk van de cenotypen A - E is dan

A:	$0.8 \times 0.9 = 0.72$	D:	$0.2 \times 0.4 = 0.08$
B:	$0.8 \times 0.1 = 0.08$	E:	$0.2 \times 0.6 = 0.12$
C:	$0.8 \times 0.0 = 0.00$		

De som van de kansen over de cenotypen A tot en met E is dan weer 1.0. Merk op dat de voorspelde kans op een cenotype nooit hoger kan zijn dan de voorspelde kans op de bijbehorende hoofdgroep.
Modelselectie is steeds uitgevoerd met behulp van de Genstat procedure SELECT. Deze past alle mogelijke modellen aan voor maximaal 16 predictoren en geeft aan welke van de aangepaste modellen de beste zijn. Aangezien er in totaal 21

predictoren zijn én modelselectie met meer dan 12 predictoren te rekenintensief is, is SELECT iteratief toegepast. Voor een eerste selectie zijn de predictoren ingedeeld in 3 groepen (1…7, 8…14, 15…21) en is een aparte modelselectie uitgevoerd voor de drie groepen. Met de beste predictoren uit de drie groepen is een nieuwe modelselectie uitgevoerd en daar zijn weer de beste predictoren uit geselecteerd. Met deze beste predictoren vast in het model worden de resterende predictoren weer onderverdeeld in 2 of 3 groepen en begint het spel opnieuw. Uiteindelijk blijven zo een gering aantal kandidaatmodellen over. Deze kandidaatmodellen zijn vervolgens geëvalueerd met behulp van kruisvalidatie. Hierbij is een individuele waarneming weggelaten (leave-one-out), is het model opnieuw aangepast en zijn de voorspelde kansen voor de weggelaten waarneming berekend. Door op deze wijze voorspelde kansen voor elke waarneming te berekenen, is een goed beeld verkregen van de voorspelkracht van een model. De gemiddelde kruisvalidatie terugvoorspelkans, **met** doorvermenigvuldiging van kruisvalidatie kansen hoger in de hiërarchie, is gebruikt als criterium om te kiezen uit de kandidaatmodellen. Bij berekening van het gemiddelde zijn de terugvoorspelkansen weer gewogen naar rato van het aantal milieuvariabelen dat niet ontbreekt. Merk op dat weglating van waarnemingen uit een bepaalde groep géén gevolgen heeft voor de kruisvalidatie voorspelling binnen een andere groep. Resubstitutie is gebruikt om het voorspelgedrag voor het geheel in kaart te brengen; resubstitutie is in het algemeen te optimistisch.

4.4 Model I: Modelselectie, kruisvalidatie en resubstitutie

4.4.1 Hoofdgroepen

Voor het hoofdgroepenmodel zijn in de eerste ronde de variabelen 12 en 14 (breedte, pH-mediaan) geselecteerd en in de twee ronde 15 en 21 (Cl-mediaan, NO_3-10-percentiel). Vervolgens zijn de predictoren 4, 2, 20, 10, 17, 8, 11 en 18 geselecteerd als mogelijke kandidaten (respectievelijk permanentie, natuurlijk dwarsprofiel, totaal fosfaat, diepte, ammonium, beschaduwing, stroomsnelheid en Kjeldal-stikstof). Kruisvalidatie is uitgevoerd voor een aantal modellen met een oplopend aantal milieuvariabelen. Voor deze modellen staan in onderstaande tabel 4.5 de gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie. Het gekozen model is vet gedrukt.

Tabel 4.5 Kandidaatmodellen aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de hoofdgroepn. Het gekozen model is vet gedrukt

nr	**milieuvariabelen in het model**												**kruis-validatie**	**resub-stitutie**
1	12	14	15	21									0.631	0.642
2	12	14	15	21	4								0.664	0.678
3	12	14	15	21	4	2							0.687	0.703
4	12	14	15	21	4	2	20						0.709	0.730
5	12	14	15	21	4	2	20	10					0.725	0.748
6	12	14	15	21	4	2	20	10	17				0.734	0.761
7	12	14	15	21	4	2	20	10	17	8			0.743	0.773
8	**12**	**14**	**15**	**21**	**4**	**2**	**20**	**10**	**17**	**8**	**11**		**0.753**	**0.784**
9	12	14	15	21	4	2	20	10	17	8	11	18	0.749	0.789

Volgens het kruisvalidatie-criterium is model 8 het beste met een gemiddelde kruisvalidatie terugvoorspelkans van 0.753. Dit model heeft milieuvariabelen natuurlijk dwarsprofiel, permanentie, beschaduwing, diepte, stroomsnelheid, breedte, pH-mediaan, chloride-mediaan, ammonium-90-percentiel, totaal fosfaat 90-percentiel en nitraat-10-percentiel. In onderstaande tabel 4.6 is de gemiddelde kruisvalidatie voorspelkans uitgesplitst naar de verschillende hoofdgroepen. De kolom "In" bevat de som van de kansen in een rij, in dit geval overal 1. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen.

Tabel 4.6 De gemiddelde kruisvalidatie voorspelkans uitgesplitst naar de verschillende hoofdgroepen. De kolom "In" bevat de som van de kansen in een rij, in dit geval overal 1. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen

hoofdgroep	**A**	**B**	**C**	**D**	**E**	**F**	**In**	**Uit**	**Ny**
A	*0.802*	0.061	0.060	0.005	0.033	0.040	1.000	0.000	159
B	0.021	*0.846*	0.017	0.012	0.103	0.000	1.000	0.000	270
C	0.302	0.082	*0.494*	0.021	0.101	0.000	1.000	0.000	32
D	0.004	0.088	0.053	*0.815*	0.040	0.000	1.000	0.000	29
E	0.117	0.405	0.050	0.028	*0.382*	0.018	1.000	0.000	60
F	0.357	0.002	0.027	0.000	0.001	*0.614*	1.000	0.000	13

In de rij voor hoofdgroep A staan bijvoorbeeld de gemiddelde kruisvalidatie-kansen dat een A monster terugvoorspeld wordt als respectievelijk A (0.802), B (0.061), C (0.060), D (0.005), E (0.033) en tenslotte F (0.040). De hoofdgroepen A (0.802), B (0.846) en D (0.815) zijn goed voorspeld onder dit model, de hoofdgroepen C (0.494), E (0.382) en F (0.614) minder goed. Hoofdgroep C en F zijn tevens terugvoorspeld als hoofdgroep A, en hoofdgroep E heeft zelfs een grotere gemiddelde kans om terugvoorspeld te worden als hoofdgroep B. Dit biedt dus ruimte voor verbetering.

4.4.2 Cenotypen binnen hoofdgroep A

In de eerste en tweede ronde zijn respectievelijk de variabelen (12 (breedte), 20 (totaal fosfaat)) en (5 (stuwing), 11 (stroomsnelheid)) geselecteerd. Met deze vier variabelen vast in het model zijn vervolgens de variabelen 2, 4, 7, 14, 8 en 9 (respectievelijk natuurlijk dwarsprofiel, permanentie, winter, pH, beschaduwing en substraat slib) geselecteerd. Gemiddelde kruisvalidatie en resubstitutie kansen, waarbij voorspelkansen hoger in de hiërarchie verdisconteerd zijn, zijn voor de verschillende modellen gegeven in tabel 4.7.

Tabel 4.7 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de cenotypen binnen hoofdgroep A. Het gekozen model is vet gedrukt

nr	milieuvariabelen in het model										kruisvalidatie	resubstitutie
1	11	12	20	2							0.299	0.355
2	5	11	12	20							0.305	0.350
3	5	11	12	20	2						0.311	0.374
4	5	11	12	20	2	14					0.325	0.396
5	5	11	12	20	2	4	7				0.339	0.415
6	5	11	12	20	2	4	7	14			0.357	0.442
7	**5**	**11**	**12**	**20**	**2**	**4**	**7**	**14**	**8**	**9**	**0.372**	**0.480**

Volgens het kruisvalidatie criterium is model 7 het beste met milieuvariabelen natuurlijk dwarsprofiel, permanentie, stuwing, winter, beschaduwing, substraat slib, stroomsnelheid, breedte, pH-mediaan en totaal fosfaat-90-percentiel. Dit model heeft een kruisvalidatie-terugvoorspelkans van 0.372, en dit is veel lager dan de 0.753 voor de hoofdgroepen. In onderstaande tabel 4.8 is de gemiddelde kruisvalidatie-voorspelkans uitgesplitst naar de verschillende cenotypen. De kolom "In" is weer de som van de kansen in een rij, en dus is bijvoorbeeld 0.752 de gemiddelde kruisvalidatie kans dat een monster van cenotype 3a terugvoorspeld wordt in groep A.

Tabel 4.8 De gemiddelde kruisvalidatie voorspelkans uitgesplitst naar de verschillende cenotypen in hoofdsgroep A. De kolom "In" bevat de som van de kansen in een rij, in dit geval overal 1. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen

cenotype	3a	3b	24a	24b	24c	25	In	Uit	Ny
3a	*0.389*	0.024	0.122	0.097	0.087	0.032	0.752	0.248	39
3b	0.216	*0.052*	0.170	0.197	0.057	0.191	0.882	0.118	6
24a	0.161	0.052	*0.153*	0.211	0.202	0.035	0.814	0.186	24
24b	0.147	0.022	0.099	*0.404*	0.168	0.006	0.845	0.155	39
24c	0.056	0.007	0.138	0.097	*0.502*	0.022	0.822	0.178	43
25	0.058	0.128	0.084	0.003	0.007	*0.356*	0.636	0.364	8

De terugvoorspelkansen op de hoofddiagonaal zijn laag, met name voor cenotypen 3b [snelstromende, half-natuurlijke bovenlopen] (0.052) en 24a [(snel) stromende, bijna natuurlijke bovenlopen] (0.153). Er is nog getracht dit model te verbeteren door allereerst een model op te stellen voor 3b, 25 [snel stromende, belaste boven-middenlopen] en vervolgens een model voor de resterende typen. De achterliggende gedachte hierbij was dat 3b en 25 relatief weinig voorkomen. Dit bleek echter geen verbetering te geven.

4.4.3 Cenotypen binnen hoofdgroep B

Vanwege het relatief grote aantal cenotypen in hoofdgroep B en het sterk wisselende aantal waarnemingen per cenotype is een verdere opsplitsing in drie groepen (6, 10, 19), (1, 14a, 14b, 31) en (16,20) uitgeprobeerd met aparte modellen per groep. Dit bleek echter niet beter te zijn dan één enkel model voor alle cenotypen. Hetzelfde geldt voor een opsplitsing in de twee groepen (6, 10, 19, 1, 14a, 14b, 31) en (16, 20). Deze exercities gaven aan dat de milieuvariabelen 8, 12, 14, 20, 2, 15, 16 en 1 belangrijk zijn. Van de volgende twee uitgeprobeerde modellen is model 2 de beste,

met predictoren meandering, natuurlijk dwarsprofiel, beschaduwing, breedte, pH-mediaan, chloride-mediaan, egv-mediaan en totaal fosfaat-90-percentiel (tabel 4.9).

Tabel 4.9 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de cenotypen binnen hoofdgroep B. Het gekozen model is vet gedrukt

nr	milieuvariabelen in het model								kruisvalidatie	resubstitutie
1	8	12	14	20	2	15	16		0.405	0.452
2	**8**	**12**	**14**	**20**	**2**	**15**	**16**	**1**	**0.425**	**0.474**

In onderstaande tabel 4.10 is te zien dat met name de cenotypen 1 (0.040), 14b (0.252), 16 (0.000) en 31 (0.189) slecht terugvoorspeld worden. Cenotype 20, met slechts 4 waarnemingen, is daarentegen zeer goed terugvoorspeld (0.932). Deze terugvoorspelkans is echter slechts gebaseerd op 4 waarnemingen.

Tabel 4.10 De gemiddelde kruisvalidatie voorspelkans uitgesplitst naar de verschillende cenotypen in hoofdsgroep B. De kolom "In" bevat de som van de kansen in een rij, in dit geval overal 1. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen

cenotype	1	6	10	14a	14b	16	19	20	31	In	Uit	Ny
1	*0.040*	0.198	0.287	0.000	0.012	0.006	0.046	0.000	0.008	0.597	0.403	14
6	0.035	*0.408*	0.184	0.012	0.030	0.014	0.141	0.015	0.042	0.880	0.120	89
10	0.035	0.175	*0.479*	0.008	0.012	0.003	0.046	0.000	0.017	0.775	0.225	84
14a	0.000	0.081	0.017	*0.588*	0.012	0.001	0.296	0.000	0.001	0.997	0.003	10
14b	0.003	0.288	0.136	0.000	*0.252*	0.010	0.114	0.000	0.170	0.973	0.027	9
16	0.010	0.292	0.091	0.000	0.040	*0.000*	0.225	0.000	0.012	0.670	0.330	3
19	0.004	0.221	0.091	0.070	0.031	0.023	*0.470*	0.007	0.011	0.927	0.073	48
20	0.000	0.000	0.000	0.000	0.000	0.000	0.001	*0.932*	0.000	0.934	0.066	4
31	0.024	0.291	0.264	0.000	0.029	0.010	0.051	0.000	*0.189*	0.858	0.142	9

4.4.4 Cenotypen binnen hoofdgroep C

Hoofdgroep C bevat slechts 2 cenotypen: 7 en 27 met respectievelijk 29 en 3 monsters. Een eerste selectie geeft aan dat met name de milieuvariabelen 9, 13, 15, 16, 19, 21 (respectievelijk substraat slib, substraat zand, chloride, EGV, zuurstof en nitraat) van belang zijn. Modelselectie met deze variabelen laat zien dat er diverse modellen met 3 variabelen zijn die een volledig onderscheid maken tussen de twee cenotypen. Dit kan ook verwacht worden omdat cenotype 27 slechts 3 monsters bevat. Gemiddelde kruisvalidatie en resubstitutie kansen voor de verschillende uitgeprobeerde modellen zijn gegeven in tabel 4.11.

Tabel 4.11 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de cenotypen binnen hoofdgroep C. Het gekozen model is vet gedrukt

nr	milieuvariabelen in het model			kruisvalidatie	resubstitutie
1	21			0.473	0.530
2	15			0.464	0.523
3	16			0.450	0.512
4	14			0.440	0.496
5	16	19		0.482	0.538
6	**13**	**15**	**19**	**0.494**	**0.550**

Het model met de drie predictoren substraat zand, chloride-mediaan en O_2-gehalte-10-percentiel is het beste. Uitgesplitst naar cenotype zijn de kruisvalidatie kansen onder dit model als in onderstaande tabel 4.12. De kansen buiten de hoofdiagonaal zijn verwaarloosbaar klein, en dus maakt het geselecteerde model een duidelijk onderscheid tussen cenotypen 7 en 27 binnen hoofdgroep C.

Tabel 4.12 De gemiddelde kruisvalidatie voorspelkans uitgesplitst naar de verschillende cenotypen in hoofdsgroep C. De kolom "In" bevat de som van de kansen in een rij. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen

cenotype	7	27	In	Uit	Ny
7	*0.504*	0.000	0.504	0.496	29
27	0.009	*0.366*	0.375	0.625	3

4.4.5 Cenotypen binnen hoofdgroep D

Een eerste groepsgewijze selectie geeft aan dat de variabelen 2, 4, 5, 10, 11, 12, 13, 15 en 19 (respectievelijk natuurlijk dwarsprofiel, permanentie, stuwing, diepte, stroomsnelheid, breedte, substraat zand, chloride en zuurstof) de belangrijkste zijn. Modelselectie met deze variabelen geeft een geschikt model met de vier variabelen 4, 10, 13 en 15. De gemiddelde kruisvalidatie kans voor dit model en voor meer beperkte modellen is gegeven in tabel 4.13.

Tabel 4.13 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de cenotypen binnen hoofdgroep D. Het gekozen model is vet gedrukt

nr	milieuvariabelen in het model				kruisvalidatie	resubstitutie
1	4	10			0.625	0.730
2	2	12			0.597	0.704
3	2	4	10		0.563	0.792
4	2	10	12		0.588	0.777
5	**4**	**10**	**13**	**15**	**0.677**	**0.898**

Het model 5, met predictoren permanentie, diepte, substraat zand en chloride-mediaan is het beste model. Uitgesplitst naar cenotype zijn de kruisvalidatie kansen onder dit model als in onderstaande tabel. De kansen op de hoofddiagonaal zijn relatief hoog.

Tabel 4.14 De gemiddelde kruisvalidatie voorspelkans uitgesplitst naar de verschillende cenotypen in hoofdsgroep D. De kolom "In" bevat de som van de kansen in een rij, in dit geval overal 1. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen

cenotype	2	8	12	In	Uit	Ny
2	*0.617*	0.000	0.146	0.763	0.237	15
8	0.000	*0.488*	0.000	0.488	0.512	2
12	0.113	0.045	*0.783*	0.941	0.059	12

4.4.6 Cenotypen binnen hoofdgroep E

Een eerste selectie geeft als belangrijke variabelen 5, 6, 9, 10, 12, 14, 17, 19 en 21 (respectievelijk stuwing, voorjaar, substraat slib, diepte, breedte, pH, ammonium, zuurstof en nitraat). Hiervan zijn 9, 14 en 21 veruit de belangrijkste. Met deze vast in het model zijn vervolgens de variabelen 1 (meandering) en 2 (natuurlijk dwarsprofiel) geselecteerd, en in een volgende iteratieslag blijken ook nog 5, 8, 10, 17 en 12 (respectievelijk stuwing, beschaduwing, diepte, ammonium en breedte) van belang te zijn. Een uiteindelijke modelselectie met 9, 14, 21, 1, 2, 5, 8, 10, 17 en 12 gaf aanleiding om kruisvalidatie uit te voeren met de modellen in tabel 4.15.
Model 2 is hiervan het beste, met overigens een lage gemiddelde kruisvalidatie kans. In dit model zitten de milieuvariabelen meandering, substraat slib, pH-mediaan en nitraat-10-percentiel. Uitgesplitst naar de cenotypen zijn de vrij lage kansen als in tabel 4.16.

Tabel 4.15 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de cenotypen binnen hoofdgroep E. Het gekozen model is vet gedrukt.

nr	milieuvariabelen in het model							kruisvalidatie	resubstitutie
1	9	14	21					0.148	0.197
2	**9**	**14**	**21**	**1**				**0.166**	**0.224**
3	9	14	21	1	2			0.161	0.237
4	9	14	21	1	2	12		0.162	0.250
5	9	14	21	1	2	12	8	0.163	0.265

Tabel 4.16 De gemiddelde kruisvalidatie voorspelkans uitgesplitst naar de verschillende cenotypen in hoofdsgroep E. De kolom "In" bevat de som van de kansen in een rij. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen

cenotype	9	13	15	26	In	Uit	Ny
9	*0.167*	0.071	0.076	0.059	0.373	0.627	22
13	0.087	*0.194*	0.065	0.057	0.403	0.597	11
15	0.099	0.061	*0.202*	0.037	0.399	0.601	16
26	0.177	0.058	0.038	*0.079*	0.352	0.648	11

4.4.7 Totaal voorspelgedrag en conclusies Model I

De enige milieuvariabelen die niet in één van de modellen gebruikt worden zijn grondgebruik natuur, voorjaar en Nkjel-90-percentiel. Op hoofdgroepniveau is de voorspelling goed voor de hoofdgroepen A, B en D, en matig voor C, E en F. Op cenotype-niveau is de voorspelkracht in de regel matig. Om het totale voorspelgedrag te kunnen beoordelen zijn de gemiddelde resubstitutie voorspelkansen berekend van cenotype naar cenotype. Deze zijn opgenomen in bijlage 1C. Hieruit blijkt dat veel cenotypen bij terugvoorspelling uitgesmeerd zijn over een groot aantal cenotypen, ook buiten de eigen hoofdgroep. Een wisseling van cenotype van de ene naar de andere hoofdgroep ligt niet meteen voor de hand, immers buiten de blokdiagonaal zijn veel kansen klein. In de volgende paragraaf is daarom een nieuwe indeling in hoofdgroepen uitgeprobeerd. Hierbij zijn de hoofdgroepen A, C en F samengevoegd tot één hoofdgroep, en zo ook voor de hoofdgroepen B en E. Hoofdgroep D blijft geheel intact. Cenotype 25, die geheel apart staat in de huidige hoofdgroep A, is

verplaatst naar de gecombineerde hoofdgroep B en E. Omdat de gecombineerde hoofdgroepen uit veel cenotypen bestaan, is tussen de hoofdgroepen en de cenotypen nog een tussenlaag, genaamd groepen, geplaatst.

4.5 Model II: Modelselectie, kruisvalidatie en resubstitutie

In de nieuwe indeling zijn de in tabel 4.17 opgesomde hoofdgroepen, groepen en cenotypen onderscheiden.
Hoofdgroep C bestaat uit slechts één groep en groepen A3 en B1 bestaan uit slechts één cenotype. Bij deze indeling behoren 10 modellen. Eén model op hoofdgroepniveau, twee modellen op groepsniveau (voor hoofdgroepen A en B) en zeven modellen op cenotype-niveau (voor groepen A1, A2, B2, B3, B4, B5, C1). Deze modellen zijn hieronder besproken.
De ontbrekende waarnemingen zijn niet opnieuw geïmputeerd, met name omdat het vinden van een geschikt model met de huidige software veel tijd kost. De imputations van model I, waarbij gebruikt is gemaakt van de hoofdgroepindeling van model I, zijn dus opnieuw gebruikt. Dit is zeker niet optimaal omdat de hoofdgroepen indeling van Model I een grote rol speelt in het imputationmodel.

Tabel 4.17 De samenstellling van hoofdgroepen, groepen en cenotypen in Model II

hoofdgroepen	**groepen**	**cenotypen**	**aantal waarnemingen**		
A			196		
	A1	24a 24b 24c 7 27		138	(24, 39, 43, 29, 3)
	A2	3a 3b		45	(39, 6)
	A3	21		13	(13)
B			338		
	B1	6		89	(89)
	B2	1 10 15		114	(14, 84, 16)
	B3	25 20		12	(8, 4)
	B4	14a 14b 16 19		70	(10, 9, 3, 48)
	B5	31 9 13 26		53	(9, 22, 11, 11)
C			29		
	C1	2 8 12		29	(15, 2, 12)

4.5.1 Hoofdgroepen

Met het iteratieve selectie proces zijn achtereenvolgens de volgende variabelen geselecteerd (12, 14, 17), (15, 21), en (2, 20) (respectievelijk breedte, pH, ammonium; chloride, nitraat; natuurlijk dwarsprofiel, totaal fosfaat) . Met al deze variabelen vast in het model zijn dan nog van belang 3, 4 en 10 (respectievelijk grondgebruik natuur, permanentie en diepte). Een uiteindelijke modelselectie met deze variabelen gaf aanleiding om kruisvalidatie uit te voeren met de in tabel 4.18 gegeven modellen.

Tabel 4.18 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de hoofdgroepen in Model II. Het gekozen model is vet gedrukt

nr	milieuvariabelen in het model										kruis-validatie	resub-stitutie
1	12	14	17								0.824	0.828
2	12	14	17	21							0.859	0.863
3	12	14	17	21	15						0.874	0.880
4	**12**	**14**	**17**	**21**	**15**	**2**					**0.890**	**0.898**
5	12	14	17	21	15	2	20				0.894	0.905
6	12	14	17	21	15	2	20	3			0.894	0.909
7	12	14	17	21	15	2	20	3	4		0.897	0.917
8	12	14	17	21	15	2	20	3	4	10	0.902	0.920

Na model 4 neemt de gemiddelde kruisvalidatie kans nog slechts marginaal toe, en daarom is gekozen voor model 4 met milieuvariabelen natuurlijk dwarsprofiel, breedte, pH-mediaan, chloride-mediaan, ammonium-90-percentiel en nitraat-10-percentiel. Deze variabelen zijn ook op hoofdgroep niveau onder model I geselecteerd. In onderstaande tabel 4.19 is de gemiddelde kruisvalidatie voorspelkans uitgesplitst naar de verschillende hoofdgroepen. Alle hoofdgroepen worden goed terugvoorspeld, de kleinste hoofdgroep C met 29 waarnemingen nog het minste. Hoofdgroepen A en B worden nauwelijks terugvoorspeld als hoofdgroep C. In vergelijking met de vorige hoofdgroep indeling is deze indeling beduidend beter, mede door het geringere aantal hoofdgroepen.

Tabel 4.19 De gemiddelde kruisvalidatie voorspelkans uitgesplitst naar de hoofdsgroepen in Model II. De kolom "In" bevat de som van de kansen in een rij, in dit geval overal 1. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen

hoofdgroep	A	B	C	In	Uit	Ny
A	*0.866*	0.126	0.008	1.000	0.000	196
B	0.071	*0.914*	0.015	1.000	0.000	338
C	0.050	0.185	*0.764*	1.000	0.000	29

Tabel 4.20 De gemiddelde kruisvalidatie terugvoorspelkansen uitgesplitst naar groep en cenotype onder Model II. De kolom "Ny" bevat het aantal waarnemingen

groep	A	B	C	Ny
A1	*0.892*	0.097	0.011	138
A2	*0.735*	0.263	0.001	45
A3	*0.984*	0.016	0.000	13
B1	0.030	*0.958*	0.012	89
B2	0.068	*0.912*	0.020	114
B3	0.297	*0.697*	0.006	12
B4	0.018	*0.973*	0.009	70
B5	0.151	*0.828*	0.021	53
C1	0.050	0.185	*0.764*	29

cenotype	groep	A	B	C	Ny
24a	A1	*0.876*	0.111	0.013	24
24b	A1	*0.908*	0.089	0.003	39
24c	A1	*0.953*	0.044	0.003	43
7	A1	*0.830*	0.146	0.024	29
27	A2	*0.599*	0.336	0.065	3
3a	A2	*0.741*	0.258	0.001	39
3b	A2	*0.703*	0.297	0.001	6
21	A3	*0.984*	0.016	0.000	13
6	B1	0.030	*0.958*	0.012	89
1	B2	0.219	*0.730*	0.051	14
10	B2	0.029	*0.956*	0.015	84
15	B2	0.180	*0.794*	0.027	16
25	B3	0.416	*0.584*	0.000	8
20	B3	0.081	*0.903*	0.016	4
14a	B4	0.001	*0.999*	0.000	10
14b	B4	0.001	*0.999*	0.000	9
16	B4	0.135	*0.863*	0.002	3
19	B4	0.017	*0.971*	0.012	48
31	B5	0.070	*0.927*	0.003	9
9	B5	0.143	*0.849*	0.008	22
13	B5	0.129	*0.795*	0.076	11
26	B5	0.266	*0.732*	0.002	11
2	C1	0.070	0.259	*0.671*	15
8	C1	0.002	0.000	*0.998*	2
12	C1	0.038	0.139	*0.822*	12

In tabel 4.20 zijn de gemiddelde kruisvalidatie terugvoorspelkansen uitgesplitst naar groep en cenotype. Hieruit blijkt dat groep A2 een overlap heeft met hoofdgroep B, en dat geldt in meer of mindere mate voor alle 3 de typen in groep A2. Ook groep B3 heeft enige overlap met hoofdgroep A en dat wordt met name veroorzaakt door cenotype 25. Groep B5 komt enigszins terug als hoofdgroep A, met name door cenotype 26. Groep C1 is ook terugvoorspeld als hoofdgroep B, met name door cenotype 2.

4.5.2 Groepen binnen hoofdgroep A

Achtereenvolgens zijn de variabelen (12, 14) en (4, 2) (respectievelijk breedte, pH; permanentie, natuurlijk dwarsprofiel) geselecteerd. Met deze termen in het model zijn vervolgens de variabelen 13, 7, 6 en 8 (substraat zand, winter, voorjaar, beschaduwing) belangrijk. Modelselectie met deze milieuvariabelen gaf aanleiding om de kruisvalidatie uit te voeren voor de modellen in tabel 4.21.

Tabel 4.21 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de groepen in hoofdgroep A. Het gekozen model is vet gedrukt

nr	milieuvariabelen in het model								kruis-validatie	resub-stitutie
1	12	14	4	2					0.636	0.651
2	12	14	4	2	13				0.643	0.663
3	12	14	4	2	7				0.644	0.661
4	**12**	**14**	**4**	**2**	**13**	**7**			**0.653**	**0.672**
5	12	14	4	2	13	6			0.646	0.672
6	12	14	4	2	13	8			0.645	0.670
7	12	14	4	2	13	6	8		0.654	0.685
8	12	14	4	2	13	6	8	7	0.660	0.690

Model 4 is het beste model; toevoeging van extra milieuvariabelen geeft slechts een marginale verbetering. Model 4 heeft als milieuvariabelen natuurlijk dwarsprofiel, permanentie, winter, breedte, substraat zand en pH-mediaan. De gemiddelde kruisvalidatie terugvoorspelkans onder dit model, met doorberekening van kansen hoger in de hiërarchie, is gegeven in onderstaande tabel 4.22. Daaruit blijkt dat A1 goed voorspeld wordt. A2 en A3 hebben beide een behoorlijke overlap met A1. De groep A1 met de meeste waarnemingen trekt dus erg aan de voorspellingen van de andere groepen A2 en A3.

Tabel 4.22 De gemiddelde kruisvalidatie terugvoorspelkansen binnen hoofdgroep A onder Model II. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen

groep	A1	A2	A3	In	Uit	Ny
A1	*0.747*	0.105	0.040	0.892	0.108	138
A2	0.365	*0.362*	0.009	0.735	0.265	45
A3	0.377	0.076	*0.530*	0.984	0.016	13

cenotype	groep	A1	A2	A3	Ny
24a	A1	*0.635*	0.177	0.064	24
24b	A1	*0.702*	0.156	0.050	39
24c	A1	*0.844*	0.070	0.038	43
7	A1	*0.767*	0.049	0.014	29
27	A2	0.523	*0.035*	0.040	3
3a	A2	0.357	*0.373*	0.010	39
3b	A2	0.414	*0.284*	0.004	6
21	A3	0.377	0.076	*0.530*	13

4.5.3 Groepen binnen hoofdgroep B

In de eerste ronde van de modelselectie werden de variabelen (12, 11, 20; breedte, stroomsnelheid, totaal fosfaat) geselecteerd, en in de tweede ronde werden daar (4, 1; permanentie, meandering) aan toegevoegd. Gegeven deze variabelen zijn nog belangrijk 8, 10, 12 en 2 (respectievelijk beschaduwing, diepte, breedte en natuurlijk dwarsprofiel). Vervolgens is kruisvalidatie uitgevoerd met de volgend modellen.

Tabel 4.23 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de groepen in hoofdgroep B. Het gekozen model is vet gedrukt

nr	milieuvariabelen in het model									kruis-validatie	resub-stitutie
1	12	11								0.357	0.367
2	12	11	20							0.387	0.400
3	12	11	20	4						0.404	0.420
4	12	11	20	4	1					0.417	0.435
5	12	11	20	4	1	21				0.427	0.448
6	12	11	20	4	1	21	8			0.435	0.459
7	12	11	20	4	1	21	8	10		0.439	0.467
8	**12**	**11**	**20**	**4**	**1**	**21**	**8**	**10**	**2**	**0.445**	**0.476**

het geselecteerde model 8 bevat de milieuvariabelen meandering, natuurlijk dwarsprofiel, permanentie, beschaduwing, diepte, stroomsnelheid, breedte, totaal fosfaat-90-percentiel en nitraat-10-percentiel. De gemiddelde kruisvalidatie terugvoorspelkansen voor dit model zijn gegeven in onderstaande tabel 4.24. Hieruit blijkt dat B1 voorspeld wordt als B1, B2 en B4, dat B2 beken ook ten dele terechtkomen bij B1 en B5, dat B3 ook voorspeld wordt als B4 (met name cenotype 20), dat B4 ook voorspeld wordt als B1 (met name cenotypen 14b en 16), en dat B5 ook voorspeld wordt als B2. Er is dus nogal wat overlap tussen de verschillende groepen.

Tabel 4.24 De gemiddelde kruisvalidatie terugvoorspelkansen binnen hoofdgroep B onder Model II. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen

groep	B1	B2	B3	B4	B5	In	Uit	Ny
B1	*0.417*	0.272	0.004	0.197	0.069	0.958	0.042	89
B2	0.175	*0.478*	0.011	0.069	0.179	0.912	0.088	114
B3	0.025	0.047	*0.369*	0.203	0.053	0.697	0.303	12
B4	0.217	0.109	0.045	*0.561*	0.040	0.973	0.027	70
B5	0.115	0.359	0.017	0.055	*0.282*	0.828	0.172	53

cenotype	groep	B1	B2	B3	B4	B5	Ny
6	B1	*0.417*	0.272	0.004	0.197	0.069	89
1	B2	0.232	*0.327*	0.002	0.050	0.119	14
10	B2	0.166	*0.514*	0.013	0.082	0.181	84
15	B2	0.180	*0.387*	0.007	0.009	0.210	16
25	B3	0.008	0.056	*0.365*	0.087	0.068	8
20	B3	0.055	0.031	*0.376*	0.415	0.026	4
14a	B4	0.101	0.040	0.002	*0.845*	0.011	10
14b	B4	0.403	0.221	0.000	*0.313*	0.062	9
16	B4	0.278	0.121	0.122	*0.187*	0.156	3
19	B4	0.208	0.104	0.056	*0.567*	0.035	48
31	B5	0.309	0.300	0.002	0.135	*0.181*	9
9	B5	0.072	0.458	0.032	0.017	*0.272*	22
13	B5	0.056	0.324	0.015	0.049	*0.351*	11
26	B5	0.090	0.251	0.003	0.066	*0.320*	11

4.5.4 Cenotypen binnen groep A1

Met gegroepeerde selectie van variabelen zijn achtereenvolgens de variabelen (4 [permanentie], 5 [stuwing]) en (15 [chloride], 17 [ammonium]) aan het model toegevoegd. Gegeven deze variabelen zijn dan nog belangrijk 8, 9, 10, 13, 14 en 20 (respectievelijk beschaduwing, substraat slib, diepte, substraat zand, pH en totaal fosfaat). Van de onderstaande aangepaste modellen is model 6 het beste, met milieuvariabelen permanentie, stuwing, beschaduwing, substraat slib, chloride-mediaan, ammonium-90-percentiel en totaal fosfaat-90-percentiel (tabel 4.25). De terugvoorspelkansen laten met name een onderlinge overlap zien voor de cenotypen 24a, 24b en 24c (tabel 4.26).

Tabel 4.25 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de cenotypen in groep A1. Het gekozen model is vet gedrukt

nr	**milieuvariabelen in het model**									**kruis-validatie**	**resub-stitutie**
1	4	5								0.304	0.316
2	4	5	15							0.321	0.343
3	4	5	15	17						0.360	0.387
4	4	5	15	17	20					0.380	0.411
5	4	5	15	17	8	9				0.390	0.423
6	**4**	**5**	**15**	**17**	**8**	**9**	**20**			**0.405**	**0.447**
7	4	5	15	17	8	9	20	10		0.408	0.461
8	4	5	15	17	8	9	20	10	13	0.405	0.466

Tabel 4.26 De gemiddelde kruisvalidatie terugvoorspelkansen binnen hoofdgroep A onder Model II. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen

cenotype	**24a**	**24b**	**24c**	**7**	**27**	**In**	**Uit**	**Ny**
24a	*0.170*	0.208	0.204	0.070	0.001	0.652	0.348	24
24b	0.119	*0.313*	0.166	0.092	0.022	0.712	0.288	39
24c	0.141	0.127	*0.517*	0.070	0.004	0.859	0.141	43
7	0.077	0.061	0.113	*0.519*	0.000	0.769	0.231	29
27	0.005	0.009	0.008	0.001	*0.510*	0.533	0.467	3

4.5.5 Cenotypen binnen groep A2

In de eerste stap is variabele 20 (totaal fosfaat) geselecteerd en in de tweede stap 11 (stroomsnelheid). Dan zijn nog van belang 4, 7, 10, 14, 17, 16, 5 en 3 (respectievelijk permanentie, winter, diepte, pH, ammonium, EGV, stuwing en grondgebruik natuur). Modelselectie met de genoemde variabelen gaf aanleiding om onderstaande modellen aan te passen. Hiervan is model 3 het beste met milieuvariabelen stuwing, diepte, stroomsnelheid en totaal fosfaat-90-percentiel (tabel 4.27). Onder dit model is het cenotype 3b naar het grotere cenotype 3a getrokken (tabel 4.28).

Tabel 4.27 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de cenotypen in groep A2. Het gekozen model is vet gedrukt

nr	milieuvariabelen in het model						kruis-validatie	resub-stitutie
1	20						0.312	0.322
2	20	11	5				0.325	0.342
3	20	11	5	10			0.341	0.360
4	20	11	5	16			0.341	0.358
5	20	11	4	7	10		0.336	0.370
6	20	11	14	17	16	5	0.340	0.384

Tabel 4.28 De gemiddelde kruisvalidatie terugvoorspelkansen binnen groep A2 onder Model II. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen

cenotype	3a	3b	In	Uit	Ny
3a	*0.367*	0.022	0.389	0.611	39
3b	0.126	*0.171*	0.297	0.703	6

4.5.6 Cenotypen binnen groep B2

In de eerste iteratie zijn de variabelen 12 (breedte) en 21 (nitraat) geselecteerd, en in de tweede iteratie komen daar de variabelen 17 (ammonium) en 18 (Kjeldal-stikstof) bij. Gegeven deze vier zijn dan nog van belang 14, 15, 8, 9 en 10 (respectievelijk pH, chloride, beschaduwing, substraat slib en diepte). Modelselectie met de genoemde variabelen gaf aanleiding om kruisvalidatie uit te voeren met onderstaande modellen. Model 2, met breedte, ammonium-90-percentiel, Nkjel-90-percentiel en nitraat-10-percentiel, is dan het beste model (tabel 4.29). Onder dit model is cenotype 10 goed terugvoorspeld, cenotype 1 slecht en cenotype 15 komt voor een deel terecht bij 10 (tabel 4.30).

Tabel 4.29 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de cenotypen in groep B2. Het gekozen model is vet gedrukt

nr	milieuvariabelen in het model							kruis-validatie	resub-stitutie
1	12	21	18					0.372	0.382
2	**12**	**21**	**18**	**17**				**0.387**	**0.398**
3	12	21	18	17	14			0.387	0.402
4	12	21	18	17	15	10		0.388	0.407
5	12	21	18	17	14	15	10	0.388	0.410

Tabel 4.30 De gemiddelde kruisvalidatie terugvoorspelkansen binnen groep B2 onder Model II. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen

cenotype	1	10	15	In	Uit	Ny
1	*0.047*	0.211	0.093	0.351	0.649	14
10	0.036	*0.464*	0.034	0.533	0.467	84
15	0.063	0.159	*0.205*	0.427	0.573	16

4.5.7 Cenotypen binnen groep B3

Deze kleine groep met slechts 2 cenotypen en weinig waarnemingen geeft voor een aantal modellen met twee predictoren een volledige uitsplitsing. Met name variabele

21 (nitraat-10-percentiel) is hierbij betrokken. Van de onderstaande modellen blijkt het model met 21 en 18 (nitraat-10-percentiel en Nkjel-90-percentiel) het beste kruisvalidatie resultaat te hebben (tabel 4.31). Ook de kruisvalidatie terugvoorspelkansen laten een volledige uitsplitsing zien (tabel 4.32).

Tabel 4.31 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de cenotypen in groep B3. Het gekozen model is vet gedrukt

nr	milieuvariabelen in het model		kruis-validatie	resub-stitutie
1	21		0.337	0.389
2	20		0.335	0.377
3	**21**	**18**	**0.421**	**0.431**
4	21	20	0.400	0.431
5	21	10	0.398	0.431

Tabel 4.32 De gemiddelde kruisvalidatie terugvoorspelkansen binnen groep B3 onder Model II. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen

cenotype	25	20	In	Uit	Ny
25	*0.401*	0.000	0.401	0.599	8
20	0.000	*0.456*	0.456	0.544	4

4.5.8 Cenotypen binnen groep B4

In een eerste selectie is variabele 12 (breedte) geselecteerd. Daarnaast zijn de volgende variabelen van belang 8, 10, 11, 14, 15, 17, 18 en 20 (respectievelijk beschaduwing, diepte, stroomsnelheid, pH, chloride, ammonium, Kjeldal-stikstof en totaal fosfaat). Modelselectie met deze variabelen gaf aanleiding om onderstaande modellen aan de kruisvalidatie te onderwerpen. Model 6 heeft duidelijke het beste kruisvalidatie criterium; dit model heeft als variabelen diepte, stroomsnelheid, breedte, pH-mediaan, chloride-mediaan en ammonium-90-percentiel (tabel 4.33). Cenotype 14a komt onder dit model gedeeltelijk terug als type 19. Type 16, met slecht 3 waarnemingen, wordt niet als zodanig voorspeld (tabel 4.34).

Tabel 4.33 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de cenotypen in groep B4. Het gekozen model is vet gedrukt

nr	milieuvariabelen in het model									kruis-validatie	resub-stitutie
1	12									0.375	0.386
2	12	20								0.420	0.440
3	12	14	17							0.441	0.467
4	12	8	10	18						0.449	0.487
5	12	8	10	18	15					0.453	0.515
6	**12**	**10**	**11**	**14**	**15**	**17**				**0.465**	**0.550**
7	12	10	11	14	15	17	18			0.445	0.559
8	12	10	11	14	15	17	18	8		0.440	0.587
9	12	10	11	14	15	17	18	8	20	0.407	0.587

Tabel 4.34 De gemiddelde kruisvalidatie terugvoorspelkansen binnen groep B4 onder Model II. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen

cenotype	14a	14b	16	19	In	Uit	Ny
14a	*0.626*	0.000	0.000	0.226	0.852	0.148	10
14b	0.000	*0.266*	0.000	0.088	0.354	0.646	9
16	0.000	0.005	*0.000*	0.222	0.226	0.774	3
19	0.052	0.019	0.035	*0.495*	0.600	0.400	48

4.5.9 Cenotypen binnen groep B5

In de eerste iteratie zijn de variabelen 12 (breedte) en 14 (pH) geselecteerd. Dan blijken de variabelen 1, 2, 4, 9, 10, 16, 17, 20 en 21 (respectievelijk meandering, natuurlijk dwarsprofiel, permanentie, substraat slib, diepte, EGV, ammonium, totaal fosfaat en nitraat) van belang te kunnen zijn. Modelselectie met deze variabelen gaf aanleiding om onderstaande modellen aan de kruisvalidatie te onderwerpen. Model 4, met natuurlijk dwarsprofiel, substraat slib, breedte, pH-mediaan en ammonium-90-percentiel, is dan het beste model (tabel 4.35). De terugvoorspelkansen onder dit model zijn vrij laag, met name voor cenotype 26 (tabel 4.36).

Tabel 4.35 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de cenotypen in groep B5. Het gekozen model is vet gedrukt

nr	milieuvariabelen in het model									kruis-validatie	resub-stitutie
1	12	14								0.118	0.135
2	12	14	17							0.139	0.160
3	12	14	17	9						0.147	0.181
4	**12**	**14**	**17**	**9**	**2**					**0.162**	**0.201**
5	12	14	1	2	9	4				0.163	0.223
6	12	14	1	2	9	4	10			0.048	0.269
7	12	14	1	2	9	16	20	21	10	0.176	0.285

Tabel 4.36 De gemiddelde kruisvalidatie terugvoorspelkansen binnen groep B5 onder Model II. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen

cenotype	31	9	13	26	In	Uit	Ny
31	*0.137*	0.044	0.011	0.022	0.214	0.786	9
9	0.013	*0.173*	0.057	0.064	0.307	0.693	22
13	0.037	0.073	*0.231*	0.041	0.382	0.618	11
26	0.021	0.191	0.048	*0.085*	0.345	0.655	11

4.5.10 Cenotypen binnen groep C1

In de eerste iteratie is variabele 10 (diepte) geselecteerd. Met 10 in het model bleken de variabelen 2, 4, 9, 12, 13, 14 en 20 (natuurlijk dwarsprofiel, permanentie, substraat slib, breedte, substraat zand, pH en totaal fosfaat) potentieel belangrijk te zijn. Volledige modelselectie met deze milieuvariabelen gaf aanleiding om onderstaande modellen aan te passen. Model 6 is het beste model met milieuvariabelen natuurlijk dwarsprofiel, permanentie, substraat slib en diepte (tabel 4.37). Dit model geeft hoge kruisvalidatie terugvoorspelkansen (tabel 4.38).

Tabel 4.37 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de cenotypen in groep C1. Het gekozen model is vet gedrukt

nr	milieuvar. in het model				kruis-validatie	resub-stitutie
1	4	10			0.596	0.655
2	2	12			0.590	0.637
3	2	9	10		0.605	0.741
4	2	4	10		0.563	0.703
5	2	10	12		0.604	0.694
6	**2**	**4**	**9**	**10**	**0.679**	**0.801**

Tabel 4.38 De gemiddelde kruisvalidatie terugvoorspelkansen binnen groep C1 onder Model II. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen

cenotype	2	8	12	In	Uit	Ny
2	*0.621*	0.000	0.050	0.671	0.329	15
8	0.000	*0.998*	0.000	0.998	0.002	2
12	0.000	0.144	*0.679*	0.822	0.178	12

4.5.11 Totaal voorspelgedrag en conclusies Model II

De enige milieuvariabelen die niet in één van de modellen gebruikt zijn, betreffen grondgebruik natuur, voorjaar en O_2-gehalte-10-percentiel. Op hoofdgroepniveau is de voorspelling goed voor de hoofdgroepen A (0.866) en B (0.914) en wat minder voor C (0.764). Binnen hoofdgroep A functioneert de groepsindeling A1, A2 en A3 zeer matig, met name voor de groepen A2 en A3 gaat voorspellend vermogen verloren. Dit geldt ook voor de opsplitsing van hoofdgroep B in de 5 groepen B1, B2, B3, B4 en B5. Op cenotype-niveau binnen groepen vindt een verder verlies aan voorspellend vermogen plaats, waardoor de voorspelkracht op cenotype-niveau in de regel matig is. Om het totale voorspelgedrag te kunnen beoordelen zijn de gemiddelde resubstitutie voorspelkansen berekend van groep naar groep. Deze zijn in tabel 4.39 opgenomen, waarbij kansen kleiner dan 0.01 niet vermeld zijn. Tabel 4.39 geeft geen aanleiding om groepen van hoofdgroep te laten wisselen.

Tabel 4.39 De gemiddelde resubstitutie voorspelkansen berekend van groep naar groep onder Model II. De kolom "Ny" bevat het aantal waarnemingen

groep	A1	A2	A3	B1	B2	B3	B4	B5	C1	Ny
A1	**0.764**	0.100	0.035	-	0.044	-	-	0.033	-	138
A2	0.353	**0.384**	-	0.038	0.081	0.036	0.031	0.068	-	45
A3	0.344	0.071	**0.570**	-	0.010	-	-	-	-	13
B1	0.019	-	-	**0.443**	0.261	-	0.187	0.066	0.011	89
B2	0.039	0.022	-	0.165	**0.504**	-	0.067	0.173	0.015	114
B3	0.099	0.182	-	0.022	0.044	**0.431**	0.169	0.048	-	12
B4	-	0.015	-	0.205	0.104	0.033	**0.594**	0.040	-	70
B5	0.101	0.043	-	0.110	0.345	0.013	0.052	**0.319**	0.017	53
C1	0.027	0.013	-	0.051	0.075	-	0.016	0.018	**0.801**	29

De gemiddelde resubstitutie voorspelkansen van cenotype naar cenotype zijn opgenomen in bijlage 1D. Hoge gemiddelde misclassificatie kansen komen met name voor in de kolommen voor cenotypen 6 en 10. De cenotypen 3b, 1, 14b, 16, 31, 9,

13, en 26 lijken in aanmerking te komen voor indeling in een andere groep. Cenotypen in hoofdgroep C worden goed terugvoorspeld.

4.6 Vergelijking Model I en Model II

Onder Model I zijn de cenotypen slechts opgedeeld in 6 hoofdgroepen. Model II heeft 3 hoofdgroepen en additioneel 9 groepen als tussenlaag. Daarmee heeft model II als mogelijk voordeel dat het op drie niveaus gebruikt kan worden, hoewel het hoogste niveau met slechts drie hoofdgroepen mogelijk weinig relevant is voor een voorspelling. De verschillende indeling maakt beide modellen onderling slecht vergelijkbaar, behalve op cenotype-niveau. Daarom zijn in onderstaande tabel 4.40 de gemiddelde kruisvalidatie kansen voor beide modellen nogmaals opgenomen.
Het met het aantal Ny gewogen gemiddelde van bovenstaande kruisvalidatie kansen is voor model I en II respectievelijk gelijk aan 0.401 en 0.396. Voor geheel verschillende indelingen van de cenotypen is de globale voorspelkracht op cenotype niveau dus zeer vergelijkbaar. Dit is bemoedigend omdat het aangeeft dat de precieze indeling in hoofdgroepen en groepen, althans voor de bekengegevens, weinig uitmaakt. Echter op basis van de kruisvalidatie resultaten is een keuze voor òf model I òf model II niet mogelijk.
De soms hoge kruisvalidatie terugvoorspelkansen voor cenotypen met weinig waarnemingen, zoals 20 en 8, zijn waarschijnlijk te optimistisch. Er kan immers toevallig een milieuvariabele zijn die een volledige uitsplitsing geeft voor een cenotype met weinig waarnemingen.

Tabel 4.40 De gemiddelde kruisvalidatie kansen voor beide modellen

cenotype	**kruisvalidatie Model I**	**kruisvalidatie Model II**	**Ny**
24a	0.153	0.170	24
24b	0.404	0.313	39
24c	0.502	0.517	43
7	0.504	0.519	29
27	0.366	0.510	3
3a	0.389	0.367	39
3b	0.052	0.171	6
21	0.614	0.530	13
6	0.408	0.417	89
1	0.040	0.047	14
10	0.479	0.464	84
15	0.202	0.205	16
25	0.356	0.401	8
20	0.932	0.456	4
14a	0.588	0.626	10
14b	0.252	0.265	9
16	0.000	0.000	3
19	0.470	0.495	48
31	0.189	0.137	9
9	0.167	0.173	22
13	0.194	0.231	11
26	0.079	0.085	11
2	0.617	0.621	15
8	0.488	0.998	2
12	0.783	0.679	12
gewogen gemiddelde	0.401	0.396	

5 Ontwikkeling slotenmodel

5.1 Beschrijving slotengegevens

Het slotengegevensbestand bevat 410 sloten. 'A priori' zijn op basis van de multivariate analyse resultaten 26 milieuvariabelen geselecteerd. Er zijn slechts 2 sloten met de waarde 1 (aanwezig) voor de milieuvariabele leem, alle andere sloten hebben de waarde 0 (afwezig). Omdat deze twee sloten ook in andere opzichten enigszins extreem zijn, zijn deze twee sloten uit het bestand verwijderd. Er resteert dan een bestand met 408 sloten en 25 milieuvariabelen. De eerste 9 milieuvariabelen zijn continue, de volgende 4 zijn percentages en de laatste 12 milieuvariabelen zijn nominaal (0/1). Een samenvatting van de milieuvariabelen over de 408 sloten is in tabel 5.1 weergegeven. Voor de continue gegevens zijn achtereenvolgens gegeven het aantal ontbrekende waarnemingen (Nmissing), de gebruikte transformatie, en gemiddelde, minimum en maximum na transformatie. De transformaties zijn gekozen aan de hand van een Box-plot (bijlage 2A). Percentages zijn allen logit getransformeerd en alle andere continue variabelen, met uitzondering van pH, zijn logaritmisch getransformeerd. De transformaties komen overeen met die voor de EKOO- en de beken-analyses. Voor de factoren zijn achtereenvolgens gegeven het aantal ontbrekende waarnemingen (Nmissing), het aantal sloten met factor niveau 0 (N=0), het aantal sloten met factor niveau 1 (N=1), en de fractie sloten met niveau 0 (P=0). De som van Nmissing, N=0 en N=1 is steeds gelijk aan het aantal sloten (408). Slechts 3.7% van de milieuvariabelen ontbreekt, en dat is gering ten opzichte van het bekenbestand.
De factoren veen, klei en zand hebben alle drie betrekking op het bodemtype in de omgeving van de sloot. Deze drie factoren zijn niet exclusief, dat wil zeggen dat in de omgeving van een sloot meerdere bodemtypes kunnen voorkomen. De factoren grondgebruik-akker, grondgebruik-natuur, grondgebruik-stedelijk, grondgebruik-tuinbouw en grondgebruik-weiland intensief hebben betrekking op het grondgebruik en ook deze zijn niet exclusief.
De sloten zijn op basis van de multivariate analyses onderverdeeld in 13 cenotypen:

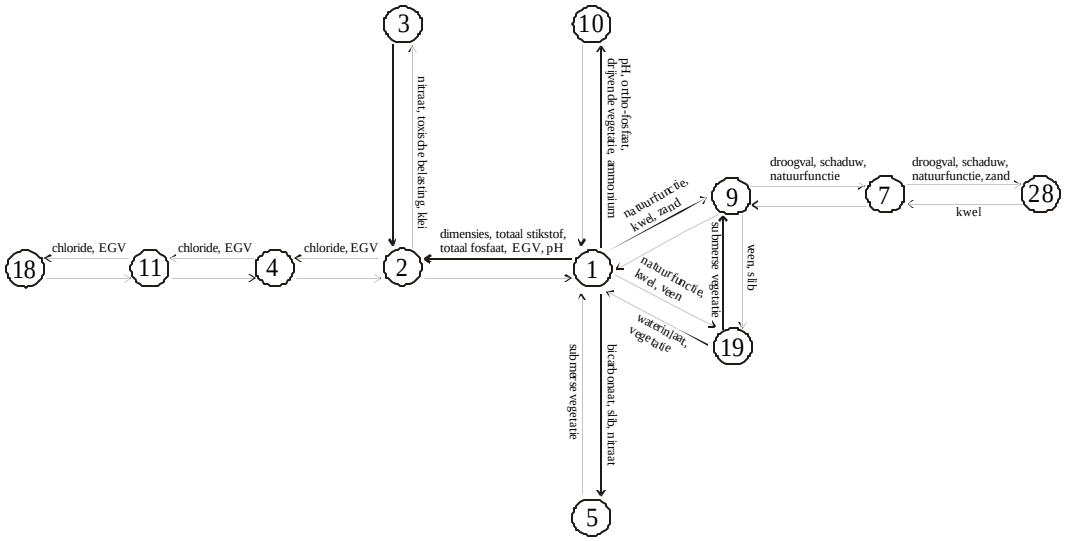
Cenotype nr.	Omschrijving
1	Plantenrijke eutrofe sloten
2	Matig grote sloten met enige belasting
3	Matig grote sloten op klei met toxische beïnvloeding
4	Grote sloten
5	Organisch belaste sloten
7	Natuurlijke zandsloten
9	Natuurlijke sloten
10	Hypertrofe klei/veensloten
11	Licht brakke sloten
18	Brakke sloten
19	(Oligo-)mesotrofe, kleine sloten
28	Droogvallende zandsloten

Tabel 5.1 Overzicht van kenmerken van de voor het te ontwikkelen slotenmodel geselecteerde milieuvariabelen (v=vegetatie, gg=grondgebruik, f=functie, int.=intensief)

nr	milieuvar	type	N-missing	trans-formatie	gemid-deld	mini-mum	maxi-mum
1	breedte	continue	8	log (+0.03)	1.26	-3.51	2.71
2	diepte	continue	4	log	-0.72	-3.22	0.41
3	chloride	continue	19	log	5.05	1.87	9.11
4	EGV	continue	47	log	4.02	1.27	7.22
5	ammonium	continue	18	log (+0.00167)	-0.86	-6.40	2.94
6	nitraat	continue	18	log (+0.001)	-0.33	-6.91	3.33
7	totaal stikstof	continue	47	log (+0.015)	1.53	-4.20	3.43
8	pH	continue	9	-	7.71	5.80	9.50
9	totaal fosfaat	continue	18	log	-0.74	-3.22	2.33
10	vegetatie-drijflaag	continue (%)	12	logit	-2.69	-5.30	5.30
11	vegetatie-flab	continue (%)	8	logit	-4.05	-5.30	2.15
12	vegetatie-emers	continue (%)	13	logit	-2.26	-5.30	5.30
13	vegetatie-submers	continue (%)	14	logit	-2.59	-5.30	5.30

nr	milieuvar	type	N-missing	N=0	N=1	P=0
14	veen	factor 2	9	294	105	0.74
15	klei	factor 2	9	230	169	0.58
16	zand	factor 2	9	227	172	0.57
17	grondgebruik-akker	factor 2	16	268	124	0.68
18	grondgebruik-natuur	factor 2	16	309	83	0.79
19	grondgebruik-stedelijk	factor 2	16	354	38	0.90
20	grondgebruik-tuinbouw	factor 2	16	352	40	0.90
21	grondgebruik-weiland intensief	factor 2	16	214	178	0.55
22	droogval	factor 2	9	388	11	0.97
23	inlaat	factor 2	10	237	161	0.60
24	functie natuur	factor 2	5	316	87	0.78
25	kwel	factor 2	7	289	112	0.72

De 13 slootcenotypen zijn in een netwerk geplaatst (figuur 5.1).
De 13 cenotypen zijn vervolgens ingedeeld in 7 hoofdgroepen A tot en met G. De hoofdgroep indeling is, evenals voor de EKOO- en de beken-analyses, gebruikt om het aantal parameters in het multinomiale logistische regressiemodel te beperken: er zijn aparte regressie-modellen aangepast voor de hoofdgroepen en voor de cenotypen binnen elke hoofdgroep. De hoofdgroepen A tot en met G bestaan uit de in tabel 5.2 benoemde cenotypen, met tussen haakjes de aantallen waarnemingen.



Figuur 5.1 Het netwerk van slootcenotypen

Tabel 5.2 De hoofdgroepen A tot en met F en bijbehorende cenotypen, met tussen haakjes de aantallen waarnemingen

hoofdgroep	**cenotypen**					
A: Normaal (160)	1	(105)	1a	(55)		
B: Groot (80)	2	(37)	3	(14)	4	(29)
C: Droogvallend (17)	7	(12)	28	(5)		
D: Natuurlijk (20)	9	(15)	19	(5)		
E: Brak (36)	11	(22)	18	(14)		
F: Belast (42)	5	(42)				
G: Belast (53)	10	(53)				

De hoofdgroepen F en G bestaan beide uit slechts één cenotype en zijn beide belast. De cenotypen 5 en 10 worden toch verschillend genoeg geacht om twee hoofdgroepen te rechtvaardigen. In tabel 5.3 zijn de aantallen sloten in elk cenotype uitgesplitst naar aantallen met respectievelijk géén, 1, 2 - 5, 6 - 10 en 11 - 25 ontbrekende milieuvariabelen.

Er zijn dus 303 sloten zonder ontbrekende waarnemingen en 63 sloten met slechts één ontbrekende waarneming. Slechts 7 sloten hebben meer dan 10 ontbrekende waarnemingen. De weinig voorkomende cenotypen 19 en 28 hebben relatief de meeste ontbrekende waarnemingen. Het ontbrekende waarnemingen patroon voor de individuele sloten met meer dan 5 ontbrekende waarnemingen is weergegeven in tabel 5.4. Hierin staat een 1 voor een aanwezige waarneming en een streepje (-) voor een ontbrekende waarneming. In de drie kolommen staan de continue milieuvariabelen, de percentages en de factoren gescheiden weergegeven. Voor sloten met 6-11 ontbrekende waarnemingen, ontbreken met name de continue en percentage milieuvariabelen. Voor de 5 sloten met 17 of meer ontbrekende waarnemingen, allen van type 11, ontbreken alle percentages en alle factoren.

Tabel 5.3 Het aantal sloten met aantal ontbrekende milieuvariabelen per hoofdgroe.

cenotype		aantal sloten met ontbrekende milieuvariabelen					totaal
		0	1	2 - 5	6 -10	11 - 25	
1	(A)	81	23	1	-	-	105
1a	(A)	42	6	3	4	-	55
2	(B)	35	-	2	-	-	37
3	(B)	13	1	-	-	-	14
4	(B)	23	6	-	-	-	29
7	(C)	11	-	1	-	-	12
28	(C)	-	-	-	3	2	5
9	(D)	10	1	4	-	-	15
19	(D)	3	-	-	2	-	5
11	(E)	4	10	3	-	5	22
18	(E)	4	8	2	-	-	14
5	(F)	35	4	2	1	-	42
10	(G)	42	4	-	7	-	53
Totaal		**303**	**63**	**18**	**17**	**7**	**408**

Tabel 5.4 Het ontbrekende waarnemingen patroon voor de individuele sloten met meer dan 5 ontbrekende waarnemingen (1 = aanwezige waarneming en - = ontbrekende waarneming)

cenotype		aantal sloten	N-missing	continue variabelen	percentage gegevens	factoren
1a	(A)	4	6	11-----1-	1111	111111111111
10	(G)	7	6	11-----1-	1111	111111111111
5	(F)	1	7	111111111	----	11111111--1-
19	(D)	1	7	11-------	1111	111111111111
19	(D)	1	8	11-------	111-	111111111111
28	(C)	1	8	11-------	111-	111111111111
28	(C)	2	10	11-------	-1--	111111111111
28	(C)	2	11	11-------	-1--	11111111-111
11	(E)	2	17	-11111111	----	----------
11	(E)	3	19	--1-11111	----	----------

5.2 Imputation toegepast op de slotengegevens

Vanwege het geringe aantal ontbrekende waarnemingen, mag verwacht worden dat de uiteindelijke conclusies ongevoelig zijn voor het te kiezen imputatie-model. Naar verwachting hebben grondgebruik en bodemtype weinig relatie met de andere milieuvariabelen. De factoren inlaat, kwel en functie natuur zouden daarentegen een grote invloed kunnen hebben op de chemie en op de waterkwaliteit. De factor droogval is bijna overal 0, dat wil zeggen afwezig, en zal daarom ook weinig relatie hebben met de overige milieuvariabelen. Deze overwegingen geven aanleiding om de ontbrekende waarnemingen in de factoren die betrekking hebben op grondgebruik en bodemtype en ook in de factor droogval niet via een model te imputeren. In plaats daarvan zijn ontbrekende waarnemingen in deze factoren vervangen door het gemiddelde binnen hetzelfde cenotype. Dit heeft bovendien als voordeel dat slechts drie factoren, namelijk inlaat, kwel en functie natuur, gebruikt behoeven te worden in het "general location model", en dat is een aanzienlijke reductie ten opzichte van het totale aantal factoren van 12. Daarnaast is ook de factor hoofdgroep in het "general location model" gebruikt.

Als startpunt voor het imputation model is gebruikt het hoofdeffecten model "hoofdgroep + inlaat + functie natuur + kwel", zowel voor het log-lineaire deel van het model als ook voor de gemiddelden van de multivariaat normale verdeling voor de continue gegevens. Vervolgens zijn aan beide delen van het model alle mogelijke twee-factor interacties tussen hoofdgroep, inlaat, functie natuur en kwel iteratief toegevoegd, en wel tot géén van de toegevoegde termen significant waren. Dit resulteerde in het volgende model:

log-lineair: hoofdgroep + inlaat + functie natuur + kwel + hoofdgroep.inlaat + hoofdgroep.kwel + hoofdgroep.functie natuur

gemiddelden: hoofdgroep + inlaat + functie natuur + kwel + hoofdgroep.kwel + hoofdgroep.functie natuur + inlaat.kwel

Dit geselecteerde model is vervolgens gebruikt om de ontbrekende waarnemingen te vervangen door conditionele gemiddelden gebaseerd op 100.000 random imputations. Voor de ontbrekende categorische milieuvariabelen resulteerde dat in kansen op factor niveau 1.

5.3 Model I: Modelselectie, kruisvalidatie en resubstitutie

5.3.1 Inleiding

De toegepaste methodologie voor modelselectie en kruisvalidatie is dezelfde als eerder toegepast in de EKOO- en de bekenanalyses. Bij het aanpassen van het multinomiale logistische regressie model zijn de waarnemingen gewogen naar rato van het aantal milieuvariabelen dat niet ontbreekt. De gemiddelde kruisvalidatie kansen zijn op dezelfde manier gewogen, met doorvermenigvuldiging van kruisvalidatie kansen hoger in de hiërarchie.

5.3.2 Hoofdgroepen

Voor het hoofdgroepen model zijn in de eerste iteratie de variabelen 3 en 4 (chloride en EGV) geselecteerd en in de twee iteratie 1, 5 en 21 (breedte, ammonium en grondgebruik weiland intensief). Daarna zijn geselecteerd 2, 8, 10 en 23 (diepte, pH, vegetatie-drijf en inlaat). Vervolgens zijn nog de predictoren 6, 7, 12, 13, 16, 18 en 22 geselecteerd als mogelijke kandidaten. Kruisvalidatie is uitgevoerd voor een aantal modellen met een oplopend aantal milieuvariabelen. Voor deze modellen staat in onderstaande tabel 5.5 de gemiddelde terugvoorspelkans berekend met kruisvalidatie en met resubstitutie. Het gekozen model is vet gedrukt.
Volgens het kruisvalidatie criterium is model 11 met 14 milieuvariabelen het beste met een gemiddelde kruisvalidatie terugvoorspelkans van 0.555. Dit model heeft milieuvariabelen breedte, diepte, chloride, EGV, ammonium, nitraat, totaal stikstof, pH, vegetatie-emers, vegetatie-submers, grondgebruik natuur, grondgebruik weiland intensief, droogval en inlaat. In tabel 5.6 is de gewogen gemiddelde kruisvalidatie voorspelkans uitgesplitst naar de verschillende hoofdgroepen. De terug-voorspelkansen van hoofdgroep naar hoofdgroep zijn cursief weergegeven.

Tabel 5.5 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de hoofdgroepen onder Model I. Het gekozen model is vet gedrukt

nr	**milieuvariabelen in het model**														**kruis-validatie**	**resub-stitutie**
1	1	3	4	5											0.449	0.465
2	1	3	4	5	21										0.468	0.488
3	1	3	4	5	21	23									0.481	0.503
4	1	3	4	5	21	23	10								0.495	0.523
5	1	3	4	5	21	2	8	10							0.507	0.542
6	1	3	4	5	21	2	8	10	23						0.517	0.554
7	1	3	4	5	21	2	8	10	23	13					0.527	0.569
8	1	3	4	5	21	2	8	10	23	13	18				0.535	0.583
9	1	3	4	5	21	10	13	18	22	23	12	16			0.530	0.594
10	1	3	4	5	21	7	8	10	18	22	23	6	16		0.539	0.606
11	**1**	**3**	**4**	**5**	**21**	**2**	**7**	**8**	**13**	**18**	**22**	**23**	**6**	**12**	**0.555**	**0.628**
12	1	3	4	5	21	2	7	8	10	13	18	22	23	6	12	0.551	0.638

Tabel 5.6 De gemiddelde kruisvalidatie voorspelkans uitgesplitst naar de verschillende hoofdgroepen. De kolom "In" bevat de som van de kansen in een rij, in dit geval overal 1. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen

hoofd-groep	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**In**	**Uit**	**Ny**
A	*0.622*	0.122	0.016	0.044	0.000	0.069	0.127	1.000	0.000	160
B	0.198	*0.525*	0.013	0.002	0.092	0.105	0.065	1.000	0.000	80
C	0.101	0.010	*0.699*	0.050	0.038	0.066	0.035	1.000	0.000	17
D	0.368	0.021	0.061	*0.467*	0.000	0.019	0.065	1.000	0.000	20
E	0.015	0.179	0.031	0.000	*0.772*	0.002	0.001	1.000	0.000	36
F	0.292	0.176	0.000	0.026	0.001	*0.388*	0.117	1.000	0.000	42
G	0.385	0.091	0.020	0.015	0.000	0.096	*0.393*	1.000	0.000	53

Tabel 5.7 De gemiddelde kruisvalidatie voorspelkans uitgesplitst naar de cenotypen. De kolom "In" bevat de som van de kansen in een rij, in dit geval overal 1. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen

cenotype	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**In**	**Uit**	**Ny**
A - 1	*0.614*	0.161	0.010	0.039	0.000	0.033	0.143	1.000	0.000	105
A - 1a	*0.636*	0.046	0.028	0.055	0.000	0.140	0.095	1.000	0.000	55
B - 2	0.211	*0.530*	0.000	0.004	0.004	0.132	0.119	1.000	0.000	37
B - 3	0.166	*0.623*	0.000	0.000	0.000	0.199	0.013	1.000	0.000	14
B - 4	0.197	*0.470*	0.035	0.000	0.248	0.026	0.024	1.000	0.000	29
C - 7	0.073	0.013	*0.728*	0.059	0.000	0.082	0.044	1.000	0.000	12
C - 28	0.211	0.000	*0.587*	0.016	0.186	0.000	0.000	1.000	0.000	5
D - 9	0.307	0.018	0.034	*0.559*	0.000	0.003	0.079	1.000	0.000	15
D - 19	0.566	0.029	0.148	*0.169*	0.000	0.069	0.019	1.000	0.000	5
E - 11	0.006	0.271	0.054	0.000	*0.664*	0.003	0.002	1.000	0.000	22
E - 18	0.028	0.058	0.000	0.000	*0.914*	0.000	0.001	1.000	0.000	14
F - 5	0.292	0.176	0.000	0.026	0.001	*0.388*	0.117	1.000	0.000	42
G - 10	0.385	0.091	0.020	0.015	0.000	0.096	*0.393*	1.000	0.000	53

In de rij voor hoofdgroep A staan bijvoorbeeld de gemiddelde kruisvalidatie kansen dat een A sloot terugvoorspeld is als respectievelijk A (0.622), B (0.122), C (0.016), D (0.044), E (0.000), F (0.069) en tenslotte G (0.127). De hoofdgroepen F (0.388) en G (0.393) zijn het slechtst terugvoorspeld. Voor zowel F als G, maar ook voor B en D, geldt dat er een relatief grote kans is om terugvoorspeld te worden als A. Verder

geldt dat E en F deels terugvoorspeld zijn als hoofdgroep B. Uitgesplitst naar cenotypen zijn de gemiddelde kruisvalidatie terugvoorspelkansen gegeven in tabel 5.7.
Met name voor de cenotypen 19, 5 en 10 zijn de terugvoorspelkansen op hoofdgroepniveau laag. Voor cenotype 18 is deze kans juist erg groot, hetgeen niet verwonderlijk is gezien het afwijkende, sterk brakke karakter van dit cenotype.

5.3.3 Cenotypen binnen hoofdgroep A

Tabel 5.8 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de cenotypen binnen hoofdgroep A. Het gekozen model is vet gedrukt

nr	**milieuvariabelen in het model**										**kruis-validatie**	**resub-stitutie**
1	3	4	23								0.494	0.536
2	3	4	23	9							0.504	0.545
3	3	4	23	12							0.498	0.542
4	3	4	12	23	9						0.510	0.551
5	3	4	12	23	9	24					0.516	0.560
6	3	4	12	23	9	16	24				0.518	0.563
7	3	4	12	23	9	16	19	24			0.517	0.565
8	**3**	**4**	**12**	**23**	**9**	**5**	**8**	**16**	**24**		**0.528**	**0.576**
9	3	4	12	23	9	5	8	16	24	19	0.526	0.576

In de eerste en tweede ronde zijn respectievelijk de variabelen 3, 4 (chloride, EGV) en 12, 23 (vegetatie-emers, inlaat) geselecteerd. Met deze vier variabelen vast in het model zijn vervolgens de variabelen 5, 8, 9, 16, 19 en 24 als potentieel belangrijk geselecteerd. Gemiddelde kruisvalidatie en resubstitutie kansen zijn voor de verschillende modellen als in tabel 5.8, hierin zijn voorspelkansen hoger in de hiërarchie verdisconteerd. Volgens het kruisvalidatie criterium is model 8 het beste met de 9 milieuvariabelen chloride, EGV, ammonium, pH, totaal fosfaat, vegetatie emers, zand, inlaat en functie natuur. Dit model heeft een kruisvalidatie terugvoorspelkans van 0.528, en dit is niet veel lager dan de 0.622 voor hoofdgroep A. Op dit niveau gaat dus niet veel voorspelkracht verloren, met andere woorden binnen hoofdgroep A is er een redelijk onderscheid tussen de cenotypen 1 en 1a. In tabel 5.9 is de gemiddelde kruisvalidatie voorspelkans uitgesplitst naar de verschillende cenotypen. De kolom "In" is weer de som van de kansen in een rij, en dus is bijvoorbeeld 0.614 de gemiddelde kruisvalidatie kans dat een monster van cenotype 1 terugvoorspeld wordt in groep A.

Tabel 5.9 De gemiddelde kruisvalidatie terugvoorspelkansen voor cenotypen binnen hoofdgroep A. De kolom "In" bevat de som van de kansen in een rij, in dit geval overal 1. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen

cenotype	**1**	**1a**	**In**	**Uit**	**Ny**
1	*0.546*	0.068	0.614	0.386	105
1a	0.145	*0.491*	0.636	0.364	55

5.3.4 Cenotypen binnen hoofdgroep B

In de eerste iteratie zijn de milieuvariabelen 3, 4, 5 (chloride, EGV, ammonium) geselecteerd, in de tweede iteratie zijn daar 1 en 11 bij (breedte, vegetatie-flab) aan toegevoegd. Met deze variabelen vast in het model zijn 6, 10, 12, 16, 18 en 24 (respectievelijk nitraat, vegetatie-drijf, vegetatie-emers, zand, grondgebruik natuur, functie natuur) nog potentieel belangrijk. Gemiddelde kruisvalidatie en resubstitutie kansen voor de verschillende uitgeprobeerde modellen zijn gegeven in tabel 5.10.

Tabel 5.10 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de cenotypen binnen hoofdgroep B. Het gekozen model is vet gedrukt

nr	**milieuvariabelen in het model**											**kruis-validatie**	**resub-stitutie**
1	1	3	4	5								0.359	0.442
2	1	3	4	5	11							0.386	0.485
3	1	3	4	5	11	12						0.383	0.497
4	1	3	4	5	11	18	24					0.391	0.513
5	1	3	4	5	11	12	18	24				0.304	0.523
6	1	3	4	5	11	12	18	24	10			0.419	0.534
7	**1**	**3**	**4**	**5**	**11**	**12**	**18**	**24**	**10**	**16**		**0.432**	**0.544**
8	1	3	4	5	11	12	18	24	10	16	6	0.416	0.547

Model 7 is het beste model met milieuvariabelen breedte, chloride, EGV, ammonium, vegetatie-drijf, vegetatie-flab, vegetatie-emers, zand, grondgebruik natuur en functie natuur. De kruisvalidatie terugvoorspelkans is 0.432, niet veel lager dan de eerder gevonden 0.525 voor hoofdgroep B. Ook binnen hoofdgroep B is het onderscheid tussen de cenotypen dus redelijk. De gemiddelde terugvoorspelkansen onder dit model zijn opgenomen in tabel 5.11.

Tabel 5.11 De gemiddelde kruisvalidatie terugvoorspelkansen voor cenotypen binnen hoofdgroep B. De kolom "In" bevat de som van de kansen in een rij, in dit geval overal 1. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen

cenotype	**2**	**3**	**4**	**In**	**Uit**	**Ny**
2	*0.428*	0.005	0.097	0.530	0.470	37
3	0.000	*0.622*	0.000	0.623	0.377	14
4	0.114	0.013	*0.343*	0.470	0.530	29

5.3.5 Cenotypen binnen hoofdgroep C

Hoofdgroep C bevat de 2 relatief weinig voorkomende cenotypen 7 en 28. Diverse modellen met 2 predictoren geven een perfecte uitsplitsing. Kruisvalidatie is duidelijk minder optimistisch, zoals te zien is in tabel 5.12.
Model 4 met diepte en kwel presteert het beste onder kruisvalidatie. De gemiddelde kruisvalidatie terugvoorspelkans is 0.699 en dit is gelijk aan die van hoofdgroep C. De uitsplitsing onder kruisvalidatie is blijkbaar nog volledig. Uitgesplitst naar de cenotypen zijn de gemiddelde kruisvalidatie voorspelkansen weergegeven in tabel 5.13.

Tabel 5.12 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de cenotypen binnen hoofdgroep C. Het gekozen model is vet gedrukt

nr	milieuvariabelen in het model		kruis-validatie	resub-stitutie
1	2	3	0.662	1.000
2	2	6	0.530	1.000
3	2	8	0.662	1.000
4	**2**	**25**	**0.699**	**1.000**
5	3	11	0.578	1.000

Tabel 5.13 De gemiddelde kruisvalidatie terugvoorspelkansen voor cenotypen binnen hoofdgroep C. De kolom "In" bevat de som van de kansen in een rij, in dit geval overal 1. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen

cenotype	7	28	In	Uit	Ny
7	*0.728*	0.000	0.728	0.272	12
28	0.000	*0.587*	0.587	0.413	5

Door het geringe aantal waarnemingen in deze hoofdgroep kan het gevonden model een toevalstreffer zijn. Dat wil zeggen dat voor nieuwe sloten de voorspelling slecht kan zijn. Het betreft echter wel zeer specifieke typen.

5.3.6 Cenotypen binnen hoofdgroep D

Ook Hoofdgroep D bevat 2 relatief weinig voorkomende cenotypen, namelijk 9 en 19. Twee modellen met 2 milieuvariabelen, namelijk (12,18) en (12,7), geven een volledige uitsplitsing naar cenotype. De resultaten van deze modellen, tezamen met 2 andere modellen staan in tabel 5.14.

Tabel 5.14 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de cenotypen binnen hoofdgroep D. Het gekozen model is vet gedrukt

nr	milieuvariabelen in het model		kruisvalidatie	resubstitutie
1	12		0.407	0.513
2	7	12	0.381	0.615
3	**12**	**18**	**0.466**	**0.615**
4	12	17	0.414	0.556

Het kruisvalidatie resultaat van model 3 (0.466) is nagenoeg gelijk aan dat van hoofdgroep D (0.467); ook onder kruisvalidatie is er dus nagenoeg een volledige uitsplitsing van de cenotypen. Model 3 heeft milieuvariabelen vegetatie-emers en grondgebruik natuur. In tabel 5.15 is te zien dat cenotype 19 slecht terugvoorspeld is, maar dit was al het geval op hoofdgroepniveau. Ook hier geldt dat, door het geringe aantal waarnemingen in deze hoofdgroep, het gevonden model een toevalstreffer kan zijn.

Tabel 5.15 De gemiddelde kruisvalidatie terugvoorspelkansen voor cenotypen binnen hoofdgroep D. De kolom "In" bevat de som van de kansen in een rij, in dit geval overal 1. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen

cenotype	9	19	In	Uit	Ny
9	*0.559*	0.000	0.559	0.441	15
19	0.003	*0.166*	0.169	0.831	5

5.3.7 Cenotypen binnen hoofdgroep E

Bij de eerste selectie is de meest belangrijke variabele 3 (chloride). In de volgende iteratieslag zijn ook nog 1, 6, 7, 12, 13 en 16 potentieel van belang. Een uiteindelijke modelselectie met deze milieuvariabelen gaf aanleiding om kruisvalidatie uit te voeren met zes modellen (tabel 5.16).

Tabel 5.16 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de cenotypen binnen hoofdgroep E. Het gekozen model is vet gedrukt

nr	milieuvariabelen in het model				kruis-validatie	resub-stitutie
1	3				0.607	0.711
2	3	6			0.644	0.767
3	3	16			0.638	0.747
4	3	12	16		0.654	0.859
5	3	6	7		0.645	0.799
6	**3**	**12**	**13**	**16**	**0.678**	**0.889**

Model 6 is hiervan het beste, met milieuvariabelen chloride, vegetatie-emers, vegetatie-submers en zand. Uitgesplitst naar de cenotypen zijn de terugvoorspelkansen als in tabel 5.17, en deze zijn weer niet veel lager dan op hoofdgroep niveau.

Tabel 5.17 De gemiddelde kruisvalidatie terugvoorspelkansen voor cenotypen binnen hoofdgroep D. De kolom "In" bevat de som van de kansen in een rij, in dit geval overal 1. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen

cenotype	11	18	In	Uit	Ny
11	*0.504*	0.160	0.664	0.336	22
18	0.008	*0.905*	0.914	0.086	14

5.3.8 Totaal voorspelgedrag en conclusies Model I

De enige milieuvariabelen die niet in één van de modellen gebruikt worden zijn veen, klei, grondgebruik akker, grondgebruik stedelijk en grondgebruik tuin. Op hoofdgroepniveau is de voorspelling goed voor de hoofdgroepen A, C en E, met een gemiddelde kruisvalidatie terugvoorspelkans hoger dan 0.6. Voor de hoofdgroepen F en G is de voorspelling matig met een gemiddelde kans onder de 0.4. Op cenotype-niveau wordt er in de regel weinig extra voorspelkracht ingeleverd. Om het totale voorspelgedrag te kunnen beoordelen zijn de gemiddelde resubstitutie voorspelkansen berekend van cenotype naar cenotype. Deze zijn opgenomen in tabel 5.18, waarin kansen kleiner dan 0.01 niet opgenomen zijn. Merk overigens op dat deze resubstitutie kansen soms veel hoger zijn dan de kansen volgens de kruisvalidatie. De resubstitutie kansen zijn dus te optimistisch.
Opvallend in deze tabel is dat cenotype 19 met grote kans (0.46) terugvoorspeld is als cenotype 1a, en cenotype 10 met kans 0.30 terugvoorspeld is als cenotype 1. De andere cenotypen met een relatief lage terugvoorspelkans (1, 1a, 2, 4 en 5) zijn uitgesmeerd over een groter aantal cenotypen. Op basis hiervan is een nieuwe

indeling in hoofdgroepen uitgeprobeerd: 1 en 10 vormen een nieuwe hoofdgroep, en 1a, 9 en 19 vormen een nieuwe hoofdgroep. De andere hoofdgroepen blijven intact.

Tabel 5.18 De gemiddelde resubstitutie voorspelkansen berekend van cenotype naar cenotype onder Model I voor de sloten. De kolom "Ny" bevat het aantal waarnemingen

hoofdgroep	**A**	**A**	**B**	**B**	**B**	**C**	**C**	**D**	**D**	**E**	**E**	**F**	**G**	
ceno-type	**1**	**1a**	**2**	**3**	**4**	**7**	**28**	**9**	**19**	**11**	**18**	**5**	**10**	**Ny**
A - 1	**.58**	.07	.09	.02	.03	-	-	.01	.02	-	-	.03	.14	105
A - 1a	.14	**.56**	.02	.02	-	-	-	.03	.01	-	-	.13	.09	55
B - 2	.18	.02	**.51**	-	.05	-	-	-	-	-	-	.12	.11	37
B - 3	.12	.03	-	**.65**	-	-	-	-	-	-	-	.18	.01	14
B - 4	.18	.04	.07	-	**.53**	-	-	-	.01	.08	.03	.02	.03	29
C - 7	-	-	-	-	-	**1.00**	-	-	-	-	-	-	-	12
C - 28	-	-	-	-	-	-	**1.00**	-	-	-	-	-	-	5
D - 9	.17	.03	.01	-	-	-	-	**.71**	-	-	-	-	.06	15
D - 19	.09	.46	.02	-	-	-	-	-	**.31**	-	-	.07	.04	5
E - 11	-	-	.01	-	.15	-	-	-	-	**.83**	-	-	-	22
E - 18	-	-	-	-	.02	-	-	-	-	-	**.97**	-	-	14
F - 5	.10	.17	.07	.07	.01	-	-	-	-	-	-	**.45**	.10	42
G - 10	.30	.08	.05	-	.03	-	-	-	-	-	-	.09	**.43**	53

5.4 Model II: Modelselectie, kruisvalidatie en resubstitutie

5.4.1 Inleiding

In de nieuwe indeling zijn de in tabel 5.19 benoemde hoofdgroepen en cenotypen onderscheiden.

Tabel 5.19 De hoofdgroepen K tot en met P en bijbehorende cenotypen, met tussen haakjes de aantallen waarnemingen

hoofdgroep			**cenotypen**					
K: Normaal	(158)		1	(105)	10	(53)		
L: Natuurlijk	(75)		1a	(55)	9	(15)	19	(5)
M: Belast	(42)	(was F)	5	(42)				
N: Groot	(80)	(was B)	2	(37)	3	(14)	4	(29)
O: Droogvallend	(17)	(was C)	7	(12)	28	(5)		
P: Brak (36)	(was E)		11	(22)	18	(14)		

In vergelijking met de vorige indeling zijn de groepen A (1, 1a), D (9, 19) en G (10) opnieuw gegroepeerd in de groepen K en L. Bij deze indeling behoren 6 modellen. Eén model op hoofdgroepniveau en vijf modellen op cenotype-niveau (voor de groepen K, L, N, O ,P), M heeft maar 1 cenotype. De potentiële modellen voor de hoofdgroepen N, O en P wijzigen niet, mogelijk wel de keuze uit deze modellen omdat in de beoordeling van de modellen ook de voorspelling op hoofdgroep niveau een rol speelt. Alle 6 de modellen zijn hieronder besproken.
De ontbrekende waarnemingen zijn niet opnieuw geïmputeerd, omdat er relatief weinig ontbrekende waarnemingen zijn en omdat de hoofdgroep indeling niet

drastisch gewijzigd is. De imputations van model I, waarbij gebruikt is gemaakt van de hoofdgroep indeling van model I, zijn dus opnieuw gebruikt.

5.4.2 Hoofdgroepen

In de eerste iteratie zijn de variabelen 1, 3, 4 en 5 geselecteerd (breedte, chloride, EGV en ammonium). Daarna zijn achtereenvolgens 10, 23 (vegetatie-drijf, inlaat) en 8, 21 (pH, grondgebruik weiland intensief) geselecteerd. Met deze 8 predictoren vast in het model zijn de volgende predictoren potentieel belangrijk 6, 9, 13, 14, 15, 16, 18, 22 en 25. Gegeven de in de eerste iteratie geselecteerde milieuvariabelen 1, 3, 4 en 5, zijn vervolgens alle modellen aangepast met de overige genoemde variabelen. Dit is aanleiding om de volgende 10 modellen aan de kruisvalidatie te onderwerpen (tabel 5.20).

Tabel 5.20 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de hoofdgroepen onder Model II. Het gekozen model is vet gedrukt

nr	**milieuvariabelen in het model**																**kruis-validatie**	**resub-stitutie**
1	1	3	4	5	8	10	16	23									0.543	0.571
2	1	3	4	5	8	10	16	21	23								0.555	0.590
3	1	3	4	5	8	9	10	16	21	23							0.569	0.606
4	1	3	4	5	8	9	10	16	21	22	23						0.571	0.613
5	1	3	4	5	8	9	10	16	18	21	22	23					0.575	0.626
6	1	3	4	5	8	9	10	15	16	18	21	22	23				0.585	0.646
7	**1**	**3**	**4**	**5**	**8**	**9**	**10**	**13**	**15**	**16**	**18**	**21**	**22**	**23**			**0.590**	**0.661**
8	1	3	4	5	6	8	9	10	13	15	16	18	21	22	23		0.589	0.669
9	1	3	4	5	6	8	9	10	13	14	15	16	18	21	22	23	0.591	0.674
10	1	3	4	5	6	8	9	10	13	14	15	16	18	21	22	23	0.591	0.681

Na model 7 stabiliseert de gewogen gemiddelde kruisvalidatie kans en daarom is gekozen voor model 7 met milieuvariabelen breedte, chloride, EGV, ammonium, pH, totaal fosfaat, vegetatie-drijf, vegetatie-submers, klei, zand, grondgebruik natuur, grondgebruik weiland intensief, droogval en inlaat. Dit model heeft 14 predictoren, evenveel als het hoofdgroepen model I. In totaal 10 predictoren zijn onder beide indelingen geselecteerd. De gemiddelde kruisvalidatie kans van deze indeling is 0.590, en dat is hoger dan de 0.555 onder indeling I. Uitgesplitst naar de hoofdgroepen zijn de gemiddelde terugvoorspelkansen als in tabel 5.21.

Tabel 5.21 De gemiddelde kruisvalidatie terugvoorspelkansen voor de hoofdgroepen onder Model II. De kolom "In" bevat de som van de kansen in een rij, in dit geval overal 1. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen. In de eerste kolom is tussen haakjes de overeenkomstige terugvoorspelkans onder indeling I gegeven

hoofdgroep		**K**	**L**	**M**	**N**	**O**	**P**	**In**	**Uit**	**Ny**
K		*0.675*	0.132	0.053	0.133	0.006	0.001	1.000	0.000	158
L		0.283	*0.461*	0.127	0.029	0.101	0.000	1.000	0.000	75
M	(0.388)	0.237	0.221	*0.381*	0.161	0.000	0.000	1.000	0.000	42
N	(0.525)	0.230	0.033	0.092	*0.561*	0.000	0.086	1.000	0.000	80
O	(0.699)	0.049	0.235	0.023	0.004	*0.676*	0.013	1.000	0.000	17
P	(0.772)	0.028	0.002	0.001	0.196	0.000	*0.773*	1.000	0.000	36

De kleine winst zit dus vooral in de nieuwe hoofdgroepen K (0.675) en L (0.461) ten opzichte van de oude hoofdgroepen A (0.622), D (0.467) en G (0.393). De hoofdgroepen L, M en N hebben met name een sterke overlap met hoofdgroep K. Daarnaast hebben O en M een zekere overlap met L, en hoofdgroepen M en P met N. Een uitsplitsing naar cenotypen geeft onderstaande gemiddelde kansen (tabel 5.22). Binnen hoofdgroepen zijn de kansen min of meer gelijk, behalve binnen hoofdgroep O met een kleinere kans voor cenotype 28 (0.400) en een grotere kans voor cenotype 7 (0.746). De winst en verlies rekening ten opzichte van indeling I laat een kleine winst zien voor het grote cenotype 1. Er is winst voor 10 en een vergelijkbaar verlies voor 1a, waarbij geldt dat beide cenotypen ongeveer evenveel waarnemingen hebben. Daarbuiten zijn de plussen en minnen klein, behalve voor de kleine cenotypen 19 (positief) en 28 (negatief).

Tabel 5.22 De gemiddelde kruisvalidatie terugvoorspelkansen voor cenotypen binnen de hoofdgroepen onder Model II. De kolom "In" bevat de som van de kansen in een rij, in dit geval overal 1. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen. In de eerste kolom is tussen haakjes de overeenkomstige terugvoorspelkans onder indeling I gegeven

cenotype		**K**	**L**	**M**	**N**	**O**	**P**	**In**	**Uit**	**Ny**
K - 1	(0.614)	*0.682*	0.125	0.029	0.154	0.010	0.001	1.000	0.000	105
K - 10	(0.393)	*0.661*	0.146	0.103	0.090	0.000	0.000	1.000	0.000	53
L - 1a	(0.636)	0.269	*0.433*	0.162	0.028	0.108	0.000	1.000	0.000	55
L - 9	(0.559)	0.340	*0.558*	0.007	0.037	0.056	0.001	1.000	0.000	15
L - 19	(0.169)	0.270	*0.475*	0.086	0.015	0.155	0.000	1.000	0.000	5
M - 5	(0.388)	0.237	0.221	*0.381*	0.161	0.000	0.000	1.000	0.000	42
N - 2	(0.530)	0.278	0.032	0.107	*0.579*	0.000	0.003	1.000	0.000	37
N - 3	(0.623)	0.151	0.026	0.192	*0.631*	0.000	0.000	1.000	0.000	14
N - 4	(0.470)	0.205	0.037	0.024	*0.502*	0.000	0.232	1.000	0.000	29
O - 7	(0.728)	0.011	0.209	0.029	0.005	*0.746*	0.000	1.000	0.000	12
0 - 28	(0.587)	0.200	0.337	0.000	0.000	*0.400*	0.063	1.000	0.000	5
P - 11	(0.664)	0.010	0.001	0.001	0.286	0.000	*0.703*	1.000	0.000	22
P - 18	(0.914)	0.053	0.002	0.000	0.079	0.000	*0.866*	1.000	0.000	14

5.4.3 Cenotypen binnen hoofdgroep K

In de eerste ronde zijn de variabelen 1 en 5 geselecteerd (breedte, ammonium). Daarnaast zijn mogelijk belangrijk 2, 4, 9, 11, 13, 15, 17, 18, 20, 21 en 22. Volledige modelselectie met deze predictoren gaf aanleiding om de kruisvalidatie uit te voeren voor acht modellen (tabel 5.23).

Tabel 5.23 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de cenotypen binnen hoofdgroep K. Het gekozen model is vet gedrukt

nr	**milieuvariabelen in het model**									**kruis-validatie**	**resub-stitutie**
1	1	5								0.445	0.466
2	1	5	18							0.460	0.483
3	1	5	11	18						0.463	0.488
4	1	2	5	11	18					0.469	0.496
5	1	2	5	11	18	22				0.471	0.501
6	1	2	4	5	11	18	22			0.474	0.505
7	**1**	**2**	**4**	**5**	**11**	**15**	**18**	**22**		**0.482**	**0.514**
8	1	2	4	5	11	13	15	18	22	0.480	0.515

Model 7 is het beste model met milieuvariabelen breedte, diepte, EGV, ammonium, vegetatie-flab, klei, grondgebruik natuur en droogval. Uitgesplitst naar cenotypen zijn de kansen gegeven in tabel 5.24 met tussen haakjes de kansen onder indeling I. Deze indeling geeft dus een klein verlies in voorspelkracht ten opzichte van de vorige indeling.

Tabel 5.24 De gemiddelde kruisvalidatie terugvoorspelkansen voor cenotypen binnen hoofdgroep K. De kolom "In" bevat de som van de kansen in een rij, in dit geval overal 1. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen. In de eerste kolom is tussen haakjes de overeenkomstige terugvoorspelkans onder indeling I gegeven

cenotype	1	10	In	Uit	Ny
1 (0.546)	*0.531*	0.150	0.682	0.318	105
10 (0.393)	0.280	*0.381*	0.661	0.339	53

5.4.4 Cenotypen binnen hoofdgroep L

In de eerste ronde van de modelselectie is de variabele 4 (EGV) geselecteerd. Daarnaast zijn potentieel belangrijk 1, 3, 7, 9, 8, 12, 14, 15, 16, 18, 24 en 25. Volledige modelselectie met deze predictoren, met 4 vast in het model, geeft één model met 5 predictoren met een nagenoeg volledige uitsplitsing (model nummer 6 in tabel 63). Vervolgens is kruisvalidatie uitgevoerd met acht modellen (tabel 5.25).

Tabel 5.25 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de cenotypen binnen hoofdgroep L. Het gekozen model is vet gedrukt

nr	milieuvariabelen in het model						kruis-validatie	resub-stitutie
1	4						0.379	0.484
2	4	3					0.376	0.491
3	4	3	16				0.391	0.514
4	4	14	15	16			0.388	0.519
5	4	14	24	25			0.392	0.514
6	**4**	**7**	**9**	**18**	**25**		**0.406**	**0.579**
7	4	3	7	9	25		0.399	0.579
8	4	7	14	18	24	25	0.404	0.579

Model 6 is het beste met variabelen EGV, totaal stikstof, totaal fosfaat, grondgebruik natuur en kwel. De kansen uitgesplitst naar cenotype laten zien dat er een winst is voor cenotype 19, maar een verlies voor de grotere cenotypen 1a en 9 (tabel 5.26).

Tabel 5.26 De gemiddelde kruisvalidatie terugvoorspelkansen voor cenotypen binnen hoofdgroep L. De kolom "In" bevat de som van de kansen in een rij, in dit geval overal 1. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen. In de eerste kolom is tussen haakjes de overeenkomstige terugvoorspelkans onder indeling I gegeven

cenotype	1a	9	19	In	Uit	Ny
1a (0.491)	*0.384*	0.045	0.004	0.433	0.567	55
9 (0.559)	0.043	*0.498*	0.017	0.558	0.442	15
19 (0.166)	0.082	0.020	*0.372*	0.475	0.525	5

5.4.5 Cenotypen binnen hoofdgroep N

Voor de selectie van predictoren wordt verwezen naar de beschrijving onder hoofdgroep B. De kruisvalidatie resultaten zijn opgenomen in tabel 5.27.

Tabel 5.27 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de cenotypen binnen hoofdgroep N. Het gekozen model is vet gedrukt

nr	milieuvariabelen in het model											kruis-validatie	resub-stitutie
1	1	3	4	5								0.373	0.451
2	1	3	4	5	11							0.410	0.503
3	1	3	4	5	11	12						0.402	0.519
4	1	3	4	5	11	18	24					0.407	0.531
5	1	3	4	5	11	12	18	24				0.325	0.545
6	1	3	4	5	11	12	18	24	10			0.434	0.557
7	**1**	**3**	**4**	**5**	**11**	**12**	**18**	**24**	**10**	**16**		**0.451**	**0.567**
8	1	3	4	5	11	12	18	24	10	16	6	0.426	0.571

Model 7 is weer het beste model met milieuvariabelen breedte, chloride, EGV, ammonium, vegetatie-drijf, vegetatie-flab, vegetatie-emers, zand, grondgebruik natuur en functie natuur. De kruisvalidatie terugvoorspelkans is 0.451, vergelijkbaar met de eerder gevonden 0.432 onder indeling I. De gemiddelde terugvoorspelkansen uitgesplitst naar cenotype zijn onder dit model ook vergelijkbaar met die onder indeling I (tabel 5.28).

Tabel 5.28 De gemiddelde kruisvalidatie terugvoorspelkansen voor cenotypen binnen hoofdgroep N. De kolom "In" bevat de som van de kansen in een rij, in dit geval overal 1. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen. In de eerste kolom is tussen haakjes de overeenkomstige terugvoorspelkans onder indeling I gegeven

cenotype	2	3	4	In	Uit	Ny
2 (0.428)	*0.448*	0.016	0.115	0.579	0.421	37
3 (0.622)	0.000	*0.631*	0.000	0.631	0.369	14
4 (0.343)	0.109	0.025	*0.368*	0.502	0.498	29

5.4.6 Cenotypen binnen hoofdgroep O

Voor de selectie van predictoren wordt verwezen naar de beschrijving bij hoofdgroep C. De kruisvalidatie resultaten zijn opgenomen in tabel 5.29.

Tabel 5.29 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de cenotypen binnen hoofdgroep O. Het gekozen model is vet gedrukt

nr	milieuvariabelen in het model		kruis-validatie	resub-stitutie
1	2	3	0.676	0.990
2	2	6	0.506	0.990
3	**2**	**8**	**0.676**	**0.990**
4	2	25	0.651	0.990
5	3	11	0.554	0.990

Hier is model 3 gekozen met diepte en pH, maar model 1 met diepte en chloride geeft hetzelfde kruisvalidatie resultaat. De gemiddelde kruisvalidatie kans van 0.676 is iets lager dan de 0.699 onder indeling I. Deze marginaal lagere kruisvalidatie kans kan op het conto geschreven worden van cenotype 28 (tabel 5.30).

Tabel 5.30 De gemiddelde kruisvalidatie terugvoorspelkansen voor cenotypen binnen hoofdgroep O. De kolom "In" bevat de som van de kansen in een rij, in dit geval overal 1. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen. In de eerste kolom is tussen haakjes de overeenkomstige terugvoorspelkans onder indeling I gegeven

cenotype	7	28	In	Uit	Ny
7 (0.728)	*0.746*	0.000	0.746	0.254	12
28 (0.587)	0.000	*0.400*	0.400	0.600	5

5.4.7 Cenotypen binnen hoofdgroep P

Voor de selectie van predictoren wordt verwezen naar de beschrijving bij hoofdgroep E. De kruisvalidatie resultaten zijn opgenomen in tabel 5.31.

Tabel 5.31 Kandidaatmodellen, aantallen milieuvariabelen en gemiddelde terugvoorspelkansen berekend met kruisvalidatie en met resubstitutie voor de cenotypen binnen hoofdgroep P. Het gekozen model is vet gedrukt

nr	milieuvariabelen in het model				kruis-validatie	resub-stitutie
1	3				0.620	0.709
2	3	6			0.654	0.763
3	3	16			0.642	0.745
4	**3**	**12**	**16**		**0.693**	**0.855**
5	3	6	7		0.670	0.796
6	3	12	13	16	0.695	0.874

Nu is model 4 in plaats van 6 geselecteerd met een hogere gemiddelde terugvoorspelkans (0.693 versus 0.645). De uitsplitsing naar cenotypen is opgenomen in tabel 5.32.

Tabel 5.32 De gemiddelde kruisvalidatie terugvoorspelkansen voor cenotypen binnen hoofdgroep P. De kolom "In" bevat de som van de kansen in een rij, in dit geval overal 1. De kolom "Uit" (= 1 - "In") is het complement van "In". De kolom "Ny" bevat het aantal waarnemingen. In de eerste kolom is tussen haakjes de overeenkomstige terugvoorspelkans onder indeling I gegeven

cenotype	11	18	In	Uit	Ny
11 (0.504)	*0.573*	0.130	0.703	0.297	22
18 (0.905)	0.014	*0.852*	0.866	0.134	14

Er is dus een kleine winst voor cenotype 11 en een klein verlies voor cenotype 18.

5.4.8 Totaal voorspelgedrag en conclusies Model II

De enige milieuvariabelen die niet in één van de modellen gebruikt worden zijn nitraat, veen, grondgebruik akker, grondgebruik stedelijk en grondgebruik tuin. Op hoofdgroepniveau is de voorspelling goed voor de hoofdgroepen K, O en P, en

matig voor hoofdgroep M. De resubstitutie kansen van cenotype naar cenotype zijn opgenomen in tabel 5.33. Buiten de hoofddiagonaal zijn er geen hele hoge kansen. Veel cenotypen, met uitzondering van die in hoofdgroepen O en P, zijn deels terugvoorspeld als cenotype 1 en voor een geringer deel als cenotype 10. Op basis van deze tabel ligt een andere indeling niet direct voor de hand.

Tabel 5.33 De gemiddelde resubstitutie voorspelkansen berekend van cenotype naar cenotype onder Model II voor de sloten. De kolom "Ny" bevat het aantal waarnemingen

hoofd-groep	**K**	**K**	**L**	**L**	**L**	**M**	**N**	**N**	**N**	**O**	**O**	**P**	**P**	
cenotype	**1**	**10**	**1a**	**9**	**19**	**5**	**2**	**3**	**4**	**7**	**28**	**11**	**18**	**Ny**
K - 1	**.57**	.15	.07	.02	.03	.03	.08	.03	.03	-	-	-	-	105
K - 10	.27	**.41**	.13	.01	-	.09	.04	-	.05	-	-	-	-	53
L - 1a	.21	.05	**.56**	-	-	.15	.01	.01	-	-	-	-	-	55
L - 9	.20	.12	-	**.65**	-	-	.03	-	-	-	-	-	-	15
L - 19	.21	.08	-	-	**.60**	.08	.01	-	-	-	-	-	-	5
M - 5	.13	.08	.16	.02	.02	**.45**	.07	.06	.01	-	-	-	-	42
N - 2	.12	.13	.03	-	-	.09	**.56**	-	.06	-	-	-	-	37
N - 3	.12	.03	.02	-	-	.17	-	**.66**	-	-	-	-	-	14
N - 4	.18	.03	.04	-	-	.02	.07	-	**.53**	-	-	.10	.03	29
O - 7	-	-	-	-	-	-	-	-	-	**.99**	-	-	-	12
O - 28	-	-	-	-	-	-	-	-	-	-	**.98**	-	-	5
P - 11	-	-	-	-	-	-	.01	-	.16	-	-	**.80**	.02	22
P - 18	.02	-	-	-	-	-	-	-	.04	-	-	.02	**.92**	14

5.5 Vergelijking Model I en Model II

Onder Model I zijn de cenotypen onderverdeeld in 7 hoofdgroepen, en onder Model II in 6 hoofdgroepen. De verschillende indeling maakt een vergelijking op hoofdgroepniveau lastig. Wel geldt dat Model II het over alle sloten gemiddeld iets beter doet dan Model I, met een gemiddelde kruisvalidatie terugvoorspelkans van 0.590 versus 0.555. Op hoofdgroep niveau is de voorspelling iets beter dan onder indeling I. Op cenotype-niveau kunnen de modellen eenvoudig vergeleken worden door de kruisvalidatie terugvoorspelkansen te vergelijken (tabel 5.34).

Tabel 5.34 De gemiddelde kruisvalidatie kansen voor beide sloten modellen. De kolom "Ny" bevat het aantal waarnemingen

cenotype	**Model I**	**Model II**	**Ny**
1	.546	.531	105
10	.393	.381	53
1a	.491	.384	55
9	.559	.498	15
19	.166	.372	5
5	.388	.381	42
2	.428	.448	37
3	.622	.631	14
4	.343	.368	29
7	.728	.746	12
28	.587	.400	5
11	.504	.573	22
18	.905	.852	14
totaal	**.492**	**.475**	

Het met het aantal Ny gewogen gemiddelde van bovenstaande kruisvalidatie kansen is voor model I en II respectievelijk gelijk aan 0.492 en 0.475. De globale voorspelkracht op cenotype niveau is dus onder beide modellen vergelijkbaar. Model I voorspelt op cenotype-niveau iets beter dan model II.

6 Modelbouw; programmatuur, betrouwbaarheid en conclusies

6.1 Inleiding

In de voorgaande hoofdstukken zijn in totaal 4 modellen ontwikkeld voor het voorspellen van cenotypen uit milieuvariabelen voor respectievelijk een beken en een sloten in Nederland. In dit hoofdstuk is een Fortran programma beschreven waarin deze modellen zijn geïmplementeerd. Het hier beschreven programma is een licht gewijzigde versie van het eerder beschreven programma voor de EKOO-gegevens. De belangrijkste wijziging is dat het nieuwe programma overweg kan met een groepering van cenotypen op twee én op drie niveaus. Tevens is aan de uitvoer file enige annotatie toegevoegd, zodat in de uitvoer file te lezen is welk model, welke milieuvariabelen en welke invoergegevens zijn gebruikt.
In paragraaf 6.2 is beschreven hoe het programma gebruikt kan worden. In paragraaf 6.3 zijn de milieuvariabelen gegeven die gebruikt worden in de 4 ontwikkelde modellen. In de paragraaf 6.4 volgt een technische beschrijving van het Fortran programma.

6.2 Berekenen van voorspelkansen voor nieuwe locaties

Het FORTRAN programma Scenario.EXE moet aangeroepen worden met drie parameters:
1) Een ongeformatteerde invoer file met constanten en parameterschattingen behorende bij het model.
2) Een ascii invoerfile met milieuvariabelen voor locaties waarvoor een voorspelling berekend moet worden. De invoer file start met een identificatie-regel. Dan volgen locatie voor locatie de milieuvariabelen. Een nieuwe locatie begint met een code ter identificatie (maximaal 20 karakters), dan volgen de milieuvariabelen, en -99 sluit de locaties af. Een voorbeeld van een invoerfile is gegeven in bijlage 3A.
3) Een ascii uitvoerfile met informatie over het model, gevolgd door de voorspelde kansen op hoofdgroep-, groep- en cenotype-niveau. In bijlage 3B is de bij de invoer in bijlage 3A behorende uitvoerfile opgenomen. De uitvoerfile bevat achtereenvolgens de volgende onderdelen:

- de identificatie regel uit de invoerfile;
- een regel die het gebruikte model specificeert;
- informatie over de hiërarchie; achtereenvolgens het aantal niveaus in de hiërarchie, dan het aantal hoofdgroepen op niveau 1, gevolgd door de namen van de hoofdgroepen (strings ter lengte 12, met 4 spaties aan het begin van een regel). Vervolgens het aantal groepen op niveau 2, gevolgd door de namen van de groepen. Tenslotte het aantal cenotypen op niveau 3, gevolgd door de namen van de cenotypen. Het aantal cenotypen op niveau 3 kan gelijk zijn aan 0.

- informatie over de milieuvariabelen; met allereerst het aantal milieuvariabelen in het model. Vervolgens per milieuvariabele één regel met naam, code voor de transformatie van de milieuvariabele (0 voor geen, 1 voor logaritme, 2 voor logit), welke constante bij de milieuvariabele opgeteld moet worden vóór een logaritmische transformatie, toegestaan minimum voor de milieuvariabele en tenslotte toegestaan maximum voor de milieuvariabele.
- dan worden de voorspelkansen locatie voor locatie afgedrukt; de eerste regel voor elk scenario bestaat uit de identificatie code uit de invoer file, het volgnummer van de locatie, een fout code en twee daarbij horende getallen. Dan volgen achter elkaar de voorspelde kansen voor hoofdgroepen, groepen en cenotypen.

De fout code in de uitvoer file heeft de volgende betekenis

0	geen fout; deze code wordt gevolgd door 0 en 0.0000000000e+00.
-1	locatie wordt overgeslagen omdat een 0/1 variabele een verkeerde waarde heeft; deze code wordt gevolgd door het volgnummer van de variabele en de waarde van die variabele. In dit geval volgen géén voorspelde kansen;
-2	locatie wordt overgeslagen omdat een continue variabele niet in het interval [min , max] ligt; deze code wordt gevolgd door het volgnummer van de variabele en de waarde van die variabele. In dit geval volgen géén voorspelde kansen;
-3	gegevens eindigen niet met -99. Het programma wordt afgebroken.

Eventuele andere fouten worden weggeschreven naar de command prompt.

Tabel 6.1 Abiotische variabelen die gebruikt worden in de voorspelling voor beken volgens Model I

volgnr	variabele-code	eenheid	trans	constante log	waarden of [min,max]
1	meander	nominaal	0		0/1 variabele
2	dwarsnat	nominaal	0		0/1 variabele
3	perman	nominaal	0		0/1 variabele
4	stuw	nominaal	0		0/1 variabele
5	winter	nominaal	0		0/1 variabele
6	schaduw	%	2		[0, 100]
7	subslib	%	2		[0, 100]
8	diepte	m	1	0.0	[0.01, 4]
9	strmsn	m/s	1	0.001	[0, 1.1]
10	breedte	m	1	0.02	[0, 40]
11	subzand	%	2		[0, 100]
12	phmed	-	0		[4.3, 9.0]
13	clmed	mg/l	1	0.0	[7, 232]
14	egvmed	uS/cm	1	0.0	[19.5, 1265]
15	nh490	mg N/l	1	0.0	[0.01, 33.5]
16	o2geh10	mg/l	0		[0.88, 169]
17	ptot90	mg P/l	1	0.0	[0.019, 10.1]
18	no310	mg N/l	1	0.001	[0.0, 48]

verklaring kolom variabele-code: meander = meandering, dwarsnat = natuurlijk dwarsprofiel, perman = permanentie, stuw = stuwing, winter = seizoen winter, schaduw = beschaduwing, subslib = substraat slib, diepte = diepte locatie, strmsn = stroomsnelheid, breedte = breedte locatie, subzand =substraat zand, phmed = zuurgraad mediaan, clmed = chloride-gehalte mediaan, egvmed = elektrisch geleidend vermogen mediaan, nh490 = ammoniumgehalte 90-percentiel, o2geh10 = zuurstofgehalte 10-percentiel, ptot90 = totaal fosfaatgehalte 90-percentiel, no310 = nitraatgehalte 10-percentiel

6.3 Invoerfile voor de modellen voor MLR-BEEK en MLR-SLOOT

In deze paragraaf is voor elk model een tabel gegeven met milieuvariabelen die in dat specifieke model gebruikt worden. De tabel bevat de volgende kolommen:

1. *Volgnr* bevat het volgnummer van de milieuvariabele.
2. *Variabele-code* bevat de naam van de milieuvariabele. De invoerfile met milieuvariabelen van nieuwe beken/sloten dient slechts deze milieuvariabelen te bevatten en wel in de volgorde van de tabel.
3. *Eenheid* bevat de fysische eenheid waarin de milieuvariabele opgegeven moet zijn.
4. *Trans* bevat een transformatie code (0 - geen transformatie, 1 - logaritmische transformatie, 2 - empirische logit transformatie).
5. *Constante Log* bevat een getal dat vóór logaritmische transformatie bij de variabele wordt opgeteld
6. *Waarden of [min,max]* bevat de toegestane waarden voor de milieuvariabelen. Een 0/1 variabel moet de waarde 0 of 1 aannemen, Andere variabelen moeten in de range [min, max] liggen. De minima en maxima zijn berekend voor het gegevensbestand waarop het model is aangepast. Merk op dat de transformatie in het programma wordt uitgevoerd. De invoer moet dus de milieuvariabelen in oorspronkelijke eenheden bevatten.

De abiotische variabelen die gebruikt zijn in **Model I voor beken,** zijn gegeven in tabel 6.1.

De parameterconstanten en de parameterschattingen voor het **Model I voor beken** zijn opgeslagen in de ongeformatteerde file Beken-1.unf. Tevens zijn voorbeeld in- en uitvoerfiles: Beken-1.in en Beken-1.out beschikbaar. De Beken-1.in file bevat voor elke beek slechts de 18 milieuvariabelen uit tabel 6.1. In Beken-1.in zijn alle beken opgenomen waarvoor in het oorspronkelijke gegevensbestand de milieuvariabelen compleet zijn.

Tabel 6.2 Abiotische variabelen die gebruikt worden in de voorspelling voor beken volgens Model II

volgnr	variabele-code	eenheid	trans	constante log	waarden of [min,max]
1	meander	nominaal	0		0/1 variabele
2	dwarsnat	nominaal	0		0/1 variabele
3	perman	nominaal	0		0/1 variabele
4	stuw	nominaal	0		0/1 variabele
5	winter	nominaal	0		0/1 variabele
6	schaduw	%	2		[0, 100]
7	subslib	%	2		[0, 100]
8	diepte	m	1	0.0	[0.01, 4]
9	strmsn	m/s	1	0.001	[0, 1.1]
10	breedte	m	1	0.02	[0, 40]
11	subzand	%	2		[0, 100]
12	phmed	-	0		[4.3, 9.0]
13	clmed	mg/l	1	0.0	[7, 232]
14	nh490	mg N/l	1	0.0	[0.01, 33.5]
15	nkjel90	mg N/l	1	0.0	[0.12, 42.5]
16	ptot90	mg P/l	1	0.0	[0.019, 10.1]
17	no310	mg N/l	1	0.001	[0.0, 48]

verklaring kolom variabele-code: zie tabel 6.1 en nkjel90 = Kjeldal-stikstofgehalte 90-percentiel

De abiotische variabelen die gebruikt zijn in **Model II voor beken,** zijn gegeven in tabel 6.2. De corresponderende files zijn Beken-2.unf, Beken-2.in en Beken-2.out. De Beken-2.in file bevat voor elke beek slechts de 17 milieuvariabelen uit tabel 6.2. Merk op dat dit andere milieuvariabelen zijn dan voor Model I.

De abiotische variabelen die gebruikt zijn in **Model I voor sloten**, zijn gegeven in tabel 6.3. De corresponderende files zijn Sloten-1.unf, Sloten-1.in en Sloten-1.out. De Sloten-1.in file bevat voor elke beek slechts de 20 milieuvariabelen uit tabel 6.3.

Tabel 6.3 Abiotische variabelen die gebruikt worden in de voorspelling voor sloten volgens Model I

volgnr	variabele	eenheid	trans	constante log	waarden of [min,max]
1	breedte	m	1	0.03	[0.0, 15.0]
2	diepte	m	1	0.0	[0.04, 1.5]
3	chloride	mg/l	1	0.0	[6.5, 9013.3]
4	egv	mS/m	1	0.0	[3.56, 1360.6]
5	ammonium	mg N/l	1	0.00167	[0.0, 18.99]
6	nitraat	mg N/l	1	0.001	[0.0, 27.8]
7	totaaln	mgN /l	1	0.015	[0.0, 30.7]
8	ph	-	0		[5.8, 9.5]
9	totaalp	mg P/l	1	0.0	[0.04, 10.26]
10	vdrijf	%	2		[0, 100]
11	vflab	%	2		[0, 100]
12	vemers	%	2		[0, 100]
13	vsubmers	%	2		[0, 100]
14	zand	nominaal	0		0/1 variabele
15	gnatuur	nominaal	0		0/1 variabele
16	gweii	nominaal	0		0/1 variabele
17	droogval	nominaal	0		0/1 variabele
18	inlaat	nominaal	0		0/1 variabele
19	functie natuur	nominaal	0		0/1 variabele
20	kwel	nominaal	0		0/1 variabele

verklaring kolom variabele-code: breedte = breedte locatie, diepte = diepte locatie, chloride = chloride-gehalte, egv = electrisch geleidend vermogen, ammonium = ammoniumgehalte, nitraat = nitraatgehalte, totaaln = totaal stikstofgehalte, ph = zuurgraad, totaalp = totaal fosfaatgehalte, vdrijf = bedekking drijflaag vegetatie, vflab = bedekking flablaag, vemers = bedekking emerse laag vegetatie, vsubmers = bedekking submerse laag vegetatie, zand = substraat zand, gnatuur = grondgebruik natuur, gweii = grondgebruik weiland intensief, droogval = droogval locatie, inlaat = inlaatwater, functie natuur = functie natuur van locatie, kwel = kwel op locatie aanwezig

De abiotische variabelen die gebruikt zijn in **Model II voor sloten,** zijn gegeven in Tabel 6.4. De corresponderende files zijn Sloten-2.unf, Sloten-2.in en Sloten-2.out. De Sloten-2.in file bevat voor elke beek slechts de 20 milieuvariabelen uit Tabel 6.4. Merk op dat dit andere milieuvariabelen zijn dan voor Model I.

Tabel 6.4 Abiotische variabelen die gebruikt worden in de voorspelling voor sloten volgens Model II

volgnr	variabele	eenheid	trans	constante log	waarden of [min,max]
1	breedte	m	1	0.03	[0.0, 15.0]
2	diepte	m	1	0.0	[0.04, 1.5]
3	chloride	mg/l	1	0.0	[6.5, 9013.3]
4	egv	mS/m	1	0.0	[3.56, 1360.6]
5	ammonium	mg N/l	1	0.00167	[0.0, 18.99]
6	totaaln	mg N/l	1	0.015	[0.0, 30.7]
7	ph	-	0		[5.8, 9.5]
8	totaalp	mg P/l	1	0.0	[0.04, 10.26]
9	vdrijf	%	2		[0, 100]
10	vflab	%	2		[0, 100]
11	vemers	%	2		[0, 100]
12	vsubmers	%	2		[0, 100]
13	klei	nominaal	0		0/1 variabele
14	zand	nominaal	0		0/1 variabele
15	gnatuur	nominaal	0		0/1 variabele
16	gweii	nominaal	0		0/1 variabele
17	droogval	nominaal	0		0/1 variabele
18	inlaat	nominaal	0		0/1 variabele
19	functie natuur	nominaal	0		0/1 variabele
20	kwel	nominaal	0		0/1 variabele

verklaring kolom variabele-code: zie tabel 6.3 en klei = substraat klei

6.4 Korte beschrijving Fortran programmatuur

In deze paragraaf is allereerst een Fortran programma beschreven dat de ongeformatteerde file maakt met de constanten en parameter schattingen van een model. Tevens is een korte beschrijving gegeven van het Fortran programma dat de voorspellingen berekent. Op hoofdgroep-, groep- en cenotype-niveau spelen steeds andere predictoren een rol. Een predictor is volledig benoemd als elke categorie een eigen regressiecoëfficiënt heeft en partieel als sommige categorieën een eigen regressiecoëfficiënt hebben. Voor een partiële predictor moet dus bekend zijn welke categorie het betreft. De parameters van het model zijn middels een Genstat programma weggeschreven naar een ascii file EstiMaak.out. Een voorbeeld van deze file, voor beken Model II, is opgenomen in bijlage 3C. Deze file bevat achtereenvolgens de volgende secties:

- Een regel ter identificatie van het model, deze komt terug in de uitvoer.
- Indeling info: Eerst het aantal niveaus in de hiërarchie (2 of 3). Dan het aantal hoofdgroepen op niveau 1, gevolgd door de namen van de hoofdgroepen (strings met maximale lengte 12). Vervolgens het aantal groepen op niveau 2, gevolgd door de namen van de groepen (strings met maximale lengte 12). Tenslotte het aantal cenotypen op niveau 3, gevolgd door de namen van de cenotypen (strings met maximale lengte 12). Het aantal cenotypen op niveau 3 mag gelijk zijn aan 0.
- Predictor info: Allereerst het aantal predictoren. Vervolgens volgt per predictor één regel met a) een code voor de transformatie van de predictor (0 voor geen, 1 voor log, 2 voor logit), b) welke constante bij de predictor opgeteld moet worden vóór een log transformatie, c) eerste schalingsconstante m_i voor de predictor, d) tweede schalingsconstante s_i voor de predictor, e) minimum voor de predictor, f)

maximum voor de predictor, g) naam van de predictor en tenslotte h) volgnummer van de predictor. Indien het minimum en maximum gelijk zijn (hier is -99 gebruikt) dan is de predictor een 0/1 variabele.

- Hoofdgroep info: Er zijn in het voorbeeld 3 hoofdgroepen en het onderscheid tussen deze hoofdgroepen wordt gemaakt met behulp van 6 volledige predictoren en 0 partiële predictoren. Dan volgt een nummering van de hoofgroepen. Daarna volgen de regressiecoëfficiënten voor de volledige predictoren: de eerste kolom geeft het volgnummer van de predictor aan, waarbij 0 staat voor het intercept. De volgende drie kolommen bevatten de regressiecoëfficiënten voor elk van 3 hoofdgroepen. Merk op dat voor de vierde hoofdgroep alle regressiecoëfficiënten gelijk zijn aan 0. Omdat er geen partiële predictoren zijn, volgt er geen informatie over de partiële regressiecoëfficiënten. Als er wel partiële predictoren zijn dan volgt per partiële predictor één regel met het volgnummer van de partiële predictor, de groep waarvoor de regressiecoëfficiënt geldt en de regressiecoëfficiënt zelf.
- Hoofdgroep 1 info: Er zijn 3 groepen binnen hoofdgroep 1 met 6 volledige en 0 partiële predictoren. Dit wordt gevolgd door een nummering van de groepen en de regressiecoëfficiënten voor de volledige predictoren. Idem dito voor de andere hoofdgroepen 2 en 3. De parameters voor Hoofdgroep 3 reflecteren het feit dat er binnen de derde hoofdgroep geen verdere onderverdeling in groepen is.
- Groep 1, 2, ... info: Deze hebben dezelfde structuur als hiervoor.

Er is een apart FORTRAN programma, EstiMaakUnf.FOR, geschreven dat deze file inleest en de gegevens wegschrijft naar een zogenaamde ongeformatteerde file. Voordeel van een ongeformatteerde file is dat hij niet leesbaar is en dat het dus minder waarschijnlijk is dat een gebruiker hem wijzigt. De parameters behorende bij de volledige predictoren worden ingelezen in een 2 of 3 dimensionaal array (estfull1, estfull2 en estfull3) waarbij de laatste index van het array de predictoren afloopt, de voorlaatste de categorieën en de eerste de hoofdgroepen of groepen.

Het programma Scenario.FOR berekend voorspelde kansen op de hoofdgroepen, groepen en cenotypen uit de milieuvariabelen. De voorspelkansen worden berekend in een loop die start met het commentaar "Loop over data". In deze loop wordt allereerst een volgend locatie ingelezen, wordt gecontroleerd of elke waarde binnen de range van mogelijke waarden ligt, wordt de transformatie uit Tabel 1 toegepast en tenslotte wordt elke waarde gestandaardiseerd volgens de formule $x_i = (x_i - m_i) / s_i$. Dan worden achtereenvolgens de voorspelde kansen op elk van de niveaus berekend met de aanroep van subroutines level1, level2 en level3. Tenslotte worden in mult12 en mult23 de kansen op respectievelijk niveau 1 en 2 en op niveau 2 en 3 van de hiërarchie doorvermenigvuldigd. De subroutines level2 en level3 zijn vergelijkbaar met level1, het enige verschil is dat er binnen level2 en level3 een loop wordt uitgevoerd over respectievelijk de hoofdgroepen en de groepen. In subroutine level1 wordt allereerst een lineaire predictor, genaamd pred1, berekend voor elk van de hoofdgroepen (loop met label 100). Vervolgens worden deze, indien nodig, aangevuld met het effect van de partiële predictoren (loop met label 102). Dan wordt het maximum maxpred van pred1 bepaald in loop 200. In loop 201 en 202 worden de voorspelde kansen berekend volgens de formule:

$$p_i = \exp(pred1_i) \Big/ \sum_i \exp(pred1_i).$$

Door het bijtellen van (10.0d0 - maxpred) en het controleren van het argument van de exp functie, worden fouten tijdens de executie van het programma vermeden. De subroutines mult12 en mult23 spreken voor zich.
Een overzicht van alle files is gegeven in bijlage 3D.

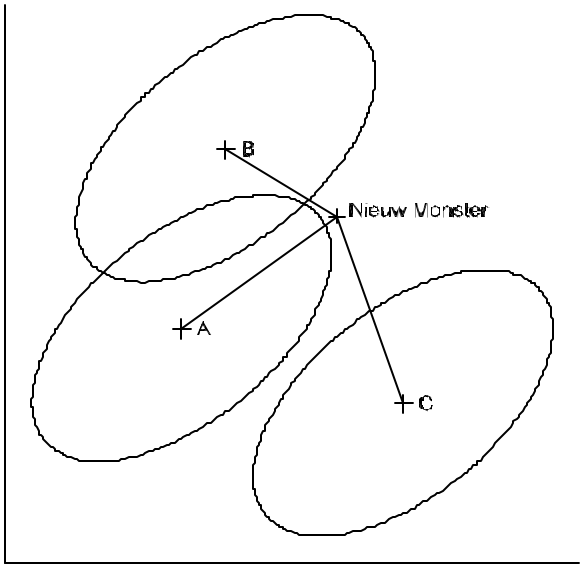
6.5 Betrouwbaarheid

6.5.1 Een maat voor de betrouwbaarheid

In de hoofdstukken 4 en 5 zijn statistische modellen beschreven voor de voorspelling van cenotypen uit milieuvariabelen voor respectievelijk beken en sloten. In de voorgaande paragrafen is een Fortran programma beschreven waarin deze modellen zijn geïmplementeerd. In deze paragraaf wordt een kwantitatieve maat beschreven voor de betrouwbaarheid van de voorspelling, alsmede de implementatie hiervan in het Fortran programma.
In principe kan voor elk regressiemodel, en dus ook voor een multinomiaal logistisch regressiemodel, de standaardafwijking van een voorspelling voor een nieuw datapunt berekend worden. Deze standaardafwijking is een maat voor de nauwkeurigheid van de voorspelling en kan bijvoorbeeld gebruikt worden om een betrouwbaarheidsinterval te construeren. Slechts de variantie-covariantie matrix van de parameterschattingen is hiervoor nodig. Echter in veel van de aangepaste multinomiaal logistische modellen zijn er een aantal parameterschattingen die naar plus of min oneindig gaan. De bijbehorende standaardafwijkingen zijn overeenkomstig erg groot. In dergelijke grensgevallen kunnen de standaardafwijkingen van de voorspellingen wel berekend worden, maar deze zijn in alle gevallen zeer groot en hebben weinig praktische waarde. Dit probleem treedt overigens ook op bij gewone logistische regressie. In het algemeen is de bootstrap een goed alternatief. De parametrische bootstrap kan echter om dezelfde reden als hierboven niet uitgevoerd worden. De niet-parametrische bootstrap, waarbij nieuwe datasets met teruglegging getrokken worden uit de oorspronkelijke dataset, stuit ook op bezwaren. Het is immers op voorhand niet duidelijk hoe de nieuwe datasets getrokken moeten worden. Zelfs als dit wel duidelijk zou zijn, dan is het zeer waarschijnlijk dat het algoritme voor het aanpassen van het multinomiaal logistische model niet zal convergeren voor een aantal nieuwe datasets. De gebruikelijke methoden om een betrouwbaarheidsmaat van de voorspelling te berekenen kunnen dus niet gebruikt worden.
Met behulp van een analogie redering kan het volgende alternatief verkregen worden. Veronderstel dat de vector van parameterschattingen $\hat{b}$ verkregen is uit het regressiemodel $Y = Xb + e$. Voor een nieuw monster met milieuvariabelen x_0 wordt de voorspelling gegeven door $\hat{y} = x_0^T \hat{b}$ met variantie $Var(\hat{y}) \propto x_0^T \left(X^T X \right)^{-1} x_0$. De variantie kan ook geschreven worden als

$Var(\hat{y}) \propto 1/n + y^T \left(\Psi^T \Psi\right)^{-1} y$, waarbij y en Ψ gecentreerde versies zijn van x_0 en X, waarbij gecentreerd is ten opzichte van het gemiddelde in X. Afgezien van de term $1/n$ is de variantie van de voorspelling dus evenredig met de zogenaamde Mahalanobis afstand van het nieuwe monster ten opzichte van de dataset, waarbij $\left(\Psi^T \Psi\right)^{-1}$ de metriek definieert ten opzichte waarvan de afstand berekend wordt. Met andere woorden, in gewone regressie is de Mahalanobis afstand een goede maat voor de betrouwbaarheid van de voorspelling. Voor gegeneraliseerde lineaire modellen kan een, met de iteratieve gewichten, gewogen versie van de Mahalanobis afstand gebruikt worden. Vanwege zijn eenvoud wordt hier de niet-gewogen versie van de Mahalanobis afstand gebruikt.

Bij multinomiale logistische regressie is de voorspelling een set van kansen, namelijk voor elk cenotype één voorspelkans. Het ligt dan voor de hand om de Mahalanobis afstand van een nieuw monster ten opzicht van elk cenotype te berekenen. Indien er genoeg monsters zijn per cenotype, dan zou elk cenotype zijn eigen metriek kunnen definiëren, maar dat is hier niet het geval. Daarom wordt een gemeenschappelijke metriek voor elk cenotype verondersteld. Er wordt dus verondersteld dat de cenotypen alleen verschillen in locatie. Een grafische voorstelling van de Mahalanobis afstanden van een nieuw monster ten opzichte van 3 cenotypen in een 2-dimensionale ruimte wordt in figuur 6.1 gegeven.



Figuur 6.1Grafische voorstelling van de Mahalanobis afstanden van een nieuw monster/locatie ten opzichte van 3 cenotypen in een 2-dimensionale ruimte

De metriek ten opzichte waarvan de afstanden berekend worden, wordt voor elk cenotype gegeven door dezelfde ellips. In gedachten liggen de waargenomen milieuvariabelen binnen elke ellips. Het gemiddelde van de milieuvariabelen per cenotype definieert de centra van de ellipsen. De afstanden van het nieuwe monster

(locatie) worden berekend ten opzichte van de ellips. De afstand ten opzichte van cenotype A is dan ook kleiner dan ten opzichte van cenotype B, ofschoon het lijnstuk van het centrum van B naar het nieuwe monster korter is. Merk op dat de ellipsen elkaar kunnen overlappen, waardoor sommige nieuwe monsters een korte afstand kunnen hebben ten opzichte van meerdere cenotypen.
De ellipsvorm heeft als onderliggend model de multivariaat normale verdeling en is dus minder geschikt voor kwalitatieve variabelen zoals grondgebruik, natuurlijkheid en kwel. Deze variabelen nemen immers slechts een beperkt aantal waarden aan. De formule voor de variantie van een voorspelling geldt echter ook voor kwalitatieve variabelen en daarom wordt de Mahalanobis afstand verder toch gebruikt.
De cenotypen gemiddelden en de metriek kunnen in Genstat eenvoudig uit de beschikbare data geschat worden middels een zogenaamde binnen groepen SSPM data structuur. Om te corrigeren voor het inschatten van ontbrekende waarnemingen worden monsters weer gewogen met het quotiënt van het aantal niet ontbrekende waarnemingen in een monster en het totaal aantal milieuvariabelen.

6.5.2 Toepassing op de beken

De Mahalanobis afstanden zijn berekend voor Model I voor beken. Daarbij zijn alle milieuvariabelen gebruikt met uitzondering van grondgebruik natuur, voorjaar en Kjeldal stikstof-90-percentiel omdat deze drie in geen van de modellen voorkomen. Dit geeft afstanden ten opzichte van elk van de 25 cenotypen. In bijlage 4 wordt per cenotype een histogram gegeven van alle afstanden, terwijl de histogrammen in bijlage 5 alleen betrekking hebben op afstanden ten opzichte van het eigen cenotype. De gebruikte klassen zijn 0-10, 10-20, … 90-100 en de laatste klasse is voor afstanden groter dan 100. De meeste verdelingen in bijlage 4 zijn enigszins scheef naar rechts, met als grote uitzondering het histogram voor cenotype 8. Dit cenotype heeft echter slechts 2 beken die blijkbaar erg extreem liggen ten opzichte van de andere cenotypen. Ook voor cenotype 14a, met 10 beken, zijn er relatief veel grote afstanden. Uit bijlage 5 blijkt dat de Mahalanobis afstanden ten opzichte van het eigen type nagenoeg allemaal kleiner zijn dan 50. Slechts 18 van de 563 afstanden zijn groter dan 50, met een maximale afstand ten opzichte van het eigen type van 78.9. In bijlagen 6 en 7 zijn de voorspelkansen uitgezet tegen de Mahalanobis afstanden, respectievelijk voor alle data en voor cenotype specifieke data. Zoals verwacht corresponderen grote afstanden met kleine voorspelkansen. Model II voor de bekengegevens geeft vergelijkbare resultaten. Dit mag ook verwacht worden omdat in Model II nagenoeg dezelfde milieuvariabelen gebruikt worden als in Model I.

6.5.3 Toepassing op de sloten

De histogrammen van de Mahalanobis afstanden voor Model II voor sloten worden gegeven in bijlagen 8 en 9, terwijl grafieken van voorspelkansen versus afstanden gegeven worden in bijlagen 10 en 11. Voor de Mahalanobis afstanden zijn alle milieuvariabelen gebruikt met uitzondering van nitraat, veen, grondgebruik akker, grondgebruik stedelijk en grondgebruik tuin. De histogrammen van bijlage 8 zijn wat

meer symmetrisch dan voor de sloten, met name doordat er relatief weinig korte afstanden zijn. Dit hangt mogelijk samen met het relatief grote aantal kwalitatieve milieuvariabelen. Cenotype 28, met 5 sloten, ligt erg extreem in vergelijking met de andere cenotypen. Voor afstanden ten opzichte van het eigen cenotype geldt dat slechts 13 van 408 afstanden groter zijn dan 50, met een maximale afstand van 97.7. De grafieken van voorspelkansen versus afstanden geven weer aan dat grote afstanden voornamelijk corresponderen met lage voorspelkansen.

6.5.4 Aanpassing van het Fortran programma

Het in paragraaf 6.4 beschreven Fortran programma voor het berekenen van voorspelkansen uit milieuvariabelen is aangepast zodat nu ook de Mahalanobis afstanden ten opzichte van hoofdgroepen, groepen en cenotypen worden berekend. De modificaties van het Fortran programma zijn zo triviaal, dat ze verder hier niet beschreven worden. In de uitvoerfile zijn de Mahalanobis afstanden opgenomen na de voorspelkansen, waarbij voorspelkansen en afstanden gescheiden worden door het woord "Mahalanobis". Een voorbeeld van een voorspelling voor de twee beken met codes respectievelijk "daa491" en "dad995" is gegeven in tabel 6.5.

Tabel 6.5 Een voorbeeld van een voorspelling voor de twee beken met codes respectievelijk "daa491" en "dad995" en waarbij de Mahalanobis afstanden zijn opgenomen na de voorspelkansen, waarbij voorspelkansen en afstanden gescheiden worden door het woord "Mahalanobis"

```
daa491                1    0    0    0.0000000000E+00
 0.00000003  0.99943687  0.00000008  0.00000000  0.00056302  0.00000000  0.00000000  0.00000000
 0.00000003  0.00000000  0.00000000  0.00000000  0.00025796  0.36976890  0.00787608  0.01783603
 0.00885136  0.00168656  0.59314775  0.00000000  0.00001222  0.00000008  0.00000000  0.00000000
 0.00000000  0.00000000  0.00036029  0.00004676  0.00003412  0.00012185  0.00000000
 Mahalanobis
  0.5288E+02  0.2857E+02  0.5275E+02  0.6237E+02  0.3995E+02  0.6246E+02  0.5317E+02
0.6223E+02
  0.6392E+02  0.6564E+02  0.7447E+02  0.5934E+02  0.4484E+02  0.2968E+02  0.3722E+02
0.2999E+02
  0.2534E+02  0.4614E+02  0.2854E+02  0.5716E+02  0.3333E+02  0.6410E+02  0.6740E+02
0.6268E+02
 0.1448E+03  0.8409E+02  0.4633E+02  0.4574E+02  0.5270E+02  0.5165E+02  0.7549E+02
dad995                2    0    0    0.0000000000E+00
 0.00000000  0.99956170  0.00000001  0.00000000  0.00043829  0.00000000  0.00000000  0.00000000
 0.00000000  0.00000000  0.00000000  0.00000000  0.00017205  0.35418358  0.00130995  0.01047346
 0.00771375  0.00534811  0.62033633  0.00000000  0.00002447  0.00000001  0.00000000  0.00000000
 0.00000000  0.00000000  0.00000003  0.00000001  0.00043815  0.00000009  0.00000000
 Mahalanobis
  0.7753E+02  0.4201E+02  0.7439E+02  0.7290E+02  0.5322E+02  0.8887E+02  0.8334E+02
0.9060E+02
  0.9448E+02  0.9422E+02  0.1091E+03  0.9365E+02  0.5796E+02  0.4097E+02  0.6210E+02
0.5074E+02
  0.4647E+02  0.6357E+02  0.4805E+02  0.7410E+02  0.5517E+02  0.9442E+02  0.8002E+02
0.7744E+02
 0.1796E+03  0.1004E+03  0.7439E+02  0.7063E+02  0.5763E+02  0.7037E+02  0.1075E+03
```

Tevens is een Genstat programma geschreven dat gebruikt kan worden als een gebruiksvriendelijke schil om het Fortran programma Scenario.exe. Invoer voor het

Genstat programma is een Excel-bestand met per regel een identificatiecode gevolgd door alle relevante milieuvariabelen. Uitvoer is een Excel-bestand met de voorspelkansen op hoofdgroep, groep en cenotype niveau, gevolgd door de Mahalanobis afstanden op de drie niveaus.
Per model zijn de volgende files beschikbaar, voorbeeld voor Model I voor beken:

Beken-1.unf	File met alle benodigde informatie over het model, zoals parameterschattingen voor de voorspelkansen en SSPM structuur voor de Mahalanobis afstanden.
Beken-1.in	Invoer file voor het Fortran programma Scenario.exe.
Beken-1.out	Uitvoer file van het Fortran programma Scenario.exe. De structuur is reeds beschreven in paragraaf 6.4, waarbij aangetekend moet worden dat de voorspelkansen gevolgd worden door de Mahalanobis afstanden per niveau in de hiërarchie.
Beken-1.gen	Genstat programma waarmee voorspellingen gedaan kunnen worden voor data in een Excelbestand. De eerste regels van het programma zijn specifiek voor gegevensbestand en model.
Beken-1.lis	Uitvoer van het Genstat programma.
Beken-1-In.xls	Invoer file voor het Genstat programma zoals gespecificeerd in het begin van het Genstat programma. Het Excelbestand bevat per regel een identificatiecode gevolgd door alle relevante milieuvariabelen.
Beken-1-Out.xls	Uitvoer file van het Genstat programma zoals gespecificeerd in het begin van het Genstat programma. Het Excelbestand bevat per regel een identificatiecode gevolgd door de voorspelkansen en de Mahalanobis afstanden op elk niveau in de hiërarchie.

Voor Model II voor beken en voor de twee modellen voor sloten zijn overeenkomstige files beschikbaar.

6.6 Discussie modelontwikkeling

1. De bekengegevensset bevat 25 cenotypen. Het eenvoudigste voorspellingsmodel is dan om elk cenotype een gelijke kans op voorkomen te geven (4%) of om aan te nemen dat de beken een random steekproef vormen en dat de waargenomen frequenties van voorkomen van de cenotypen goede schattingen zijn van de werkelijke kansen op voorkomen. In dat laatste geval is bijvoorbeeld de kans op voorkomen van cenotype 3a 39/563 = 0.07 en cenotype 6 89/536=0.16. In dat licht bezien is een voorspelkans van 50% zeer hoog. De gewogen gemiddelde voorspelkans voor de beken is 40% en dat is lager dan voor de EKOO dataset, maar toch nog vrij redelijk gezien het grote aantal cenotypen. De vraag of ze goed zijn is weinig relevant. Ze zijn met de beschikbare niet optimale gegevens zo goed mogelijk gemaakt (natuurlijk met inachtneming van de min of meer arbitraire keuzes die onderweg zijn gemaakt).
2. In het modelselectie proces zijn de belangrijkste milieuvariabelen geselecteerd. Daardoor komen sommige milieuvariabelen in geen enkel model voor, of in

slechts een gering aantal modellen (bijvoorbeeld alleen binnen een bepaalde hoofdgroep). Dat wil niet zeggen dat een ingreep in een van deze milieuvariabelen geen of weinig invloed zou hebben op de kans van voorkomen van een cenotype. De milieuvariabelen hebben immers een soms grote mate van samenhang. Door 1 milieuvariabele te veranderen, veranderen ook op korte of langere termijn andere milieuvariabelen waardoor de voorspelkansen gaan schuiven. De modellen mogen dan ook eigenlijk niet gebruikt worden om het effect van 1 milieuvariabele bij gelijkblijvende andere milieuvariabelen te bepalen. Een ingreep moet dus in zijn samenhang gezien worden.

3. In het verleden zijn ontbrekende waarnemingen ingevuld met het gemiddelde van de niet ontbrekende waarnemingen. Deze methode is redelijk als er weinig ontbrekende waarnemingen zijn. Een alternatief is het weglaten van de eenheden met 1 of meer ontbrekende waarnemingen, maar omdat deze veelal kris kras door de dataset staan, blijft er niet veel data over. In deze studie is het aantal ontbrekende milieuwaarnemingen dermate groot dat besloten is extra zorg te besteden aan het inschatten. In hoofdstuk 3 is ingegaan op de gemaakte keuze bij het imputeren en de consequenties daarvan.
4. Bij een niet-hierarchische multinomiale regressie analyse geeft een extra milieuvariabele in het model gelijk veel extra parameters (aantal cenotypn -1). Hierdoor bleek bij de EKOO dataset een zorgvuldige modelselectie niet goed mogelijk. Bij een hierarchische modellering blijft het aantal te schatten parameters per milieuvariabele steeds gering omdat het aantal klassen binnen elk niveau van de hierarchie gering is. Een uitgebreide zorgvuldige modelselectie met alle milieuvariabelen kan dan worden uitgevoerd. Het uiteindelijke hierarchische model zal in het algemeen minder parameters hebben dan een niet hierarchische model. Dit is voordelig omdat modellen met minder parameters vrijwel altijd beter voorspellen dan modellen met veel parameters. Additioneel voordeel van de hierarchische modellering is dat het model gebruikt kan worden op verschillende schaalniveau's.

7 Referenties

7.1 Inleiding

Bij de ontwikkeling van de cenotypenvoorspellingsmodellen is vanuit de RISTORI begeleidingsgroep de behoefte gegroeid om de uitkomsten van de voorspelling te kunnen kwalificeren. Het voorspellen van de effecten van een ingreep in het waterbeheer is nauw verbonden met de vraag of de ingreep leidt tot een verandering van de kwaliteit van het betreffende watersysteem. Aangezien de cenotypenvoorspellingsmodellen feitelijk voorspellen in de bestaande ruimte van reeds beschreven cenotypen is het kwalificeren van het effect van de ingreep het meten van het kwaliteitsverschil tussen de huidige toestand en de voorspelde toestand. In termen van cenotypen betekent dat het verschil in kwaliteit tussen het aanwezige en voorspelde cenotype. Door de cenotypen vooraf te voorzien van een kwaliteitsklasse, in termen van de Europese Kaderrichtlijn Water een schaal van klasse 1 (goed) tot 5 (slecht), kunnen uitspraken over kwaliteitsveranderingen worden gedaan. Omdat de Kaderrichtlijn kwalificatie uitgaat van een meting ten opzichte van de referentie en de referentie direct gerelateerd is aan de mate van natuurlijkheid van een water is uitvoerig ingegaan op de mogelijkheden om de referentie (hoofdstuk 7), een maatlat (hoofdstuk 8) en respectievelijk zeldzaamheid als maat voor natuurlijkheid (hoofdstuk 9), bij de kwalificatie te betrekken. Hiermee is onderzocht of de kwalificatie naar de eisen van de Kaderrichtlijn kan worden opgenomen.

7.2 De referentietoestand

Een cenotype beschrijft de ecologische situatie van (een deel van) een watersysteem. Dit kan zowel een natuurlijk als afgeleid beïnvloedingsstadium betreffen. De beschrijvingen kunnen daarmee fungeren als kwaliteitsreeks voor zo'n watersysteem. In het licht van de Kaderrichtlijn Water bestaat een dergelijke kwaliteitsreeks uit een referentietoestand en verschillende beïnvloedingsstadia. De referentietoestand wordt hier gedefinieerd als de ecologisch optimale situatie: een situatie waarin alleen menselijk handelen dat nodig is om het watertype in stand te houden aanwezig is en de soortensamenstelling een afspiegeling is van een gezonde leefomgeving (Verdonschot 2000). Dit kan daarmee zowel de natuurlijke als de best ecologische toestand zijn. Het beschrijven van referenties, die de oorspronkelijke natuurlijke situaties zo dicht mogelijk benaderen, is van het allergrootste belang. Referenties vormen immers de basis van een beoordelingssysteem. De keuze voor de wijze waarop een referentietoestand wordt beschreven is essentieel. Een referentietoestand kan worden beschreven aan de hand van een soortenlijst, abiotische variabelen, functionaliteit, et cetera.
Om te komen tot een goede beschrijving van een referentietoestand kan gebruik worden gemaakt van (Verdonschot 1991, United States EPA 1993):

1. bestaande wateren die de ecologisch optimale situatie zo dicht mogelijk benaderen;
2. historische gegevens;
3. empirische modellen;
4. expert-judgement van natuurlijke of potentiële systeembeschrijvingen.

(1) Wanneer gebruik wordt gemaakt van bestaande beken als referentie, moet met een aantal dingen rekening worden gehouden. In sommige landen zijn alle beken zodanig beïnvloed, dat deze onmogelijk als referentie kunnen dienen. Door minder beïnvloede beken uit naburige landen als referentie te hanteren, kan dit probleem worden opgelost. Hierbij gaat de voorkeur uit naar beken uit landen binnen dezelfde ecoregio, omdat deze het meest overeen zullen komen met de nationale beken. Verder moet rekening worden gehouden met de natuurlijke variatie binnen levensgemeenschappen. Sommige soorten zullen bijvoorbeeld alleen gedurende bepaalde jaargetijden in de beek worden gevonden. Dit kan overigens ook weer anders zijn in het buitenland. Deze natuurlijke variatie zal eerst goed in beeld moeten worden gebracht alvorens te komen tot de beschrijving van een referentietoestand. Een nog verdergaande stap is vergelijking op basis van functionele kenmerken uit te voeren. Dergelijke functionele en procesmatige kenmerken kunnen de constructie van de referentie goed ondersteunen.

(2) Aan het gebruik van historische gegevens kleven een aantal grote nadelen. Gegevensbestanden bevatten vaak specifieke informatie afhankelijk van met welk doel is bemonsterd. Vaak geeft de historische spreiding niet accuraat het voorkomen van organismen weer, maar eerder waar mensen monsterden. Dit betekent dat historische gegevens vaak afkomstig zullen zijn van meer beïnvloede (gemakkelijk te bereiken) plaatsen. Vaak was ook de taxonomie nog niet zo ver ontwikeld.

(3) Bij het gebruik van empirische modellen (bijvoorbeeld voorspellingsmodellen zoals NATLESS) moet in het achterhoofd worden gehouden, dat deze methode vaak tekortschiet op het gebied van de beschrijving van processen op gemeenschaps- danwel systeemniveau.

(4) Om te komen tot een referentietoestand die de werkelijkheid zo dicht mogelijk benadert kunnen de bovenstaande methoden worden gecombineerd met expert-judgement. De hoge mate van subjectiviteit die samengaat met expert-judgement is echter een beperking. Bij expert-judgement speelt de landschapsecologische component vaak een rol

Door de combinatie van alle methoden, wordt een groot deel van de nadelen bij het gebruik van één enkele methode ondervangen.

7.3 Referenties voor Nederlandse beken en sloten

Referenties zoals gedefinieerd in de Kaderrichtlijn Water ontbreken (nagenoeg) volledig in Nederland. Daarbij is het duidelijk dat sloten geen natuurlijke referentie bezitten, ze zijn van oorsprong kunstmatig. De referentie voor een sloot is de ecologisch optimale toestand met minimaal beheer om de sloot in stand te houden.

In ieder geval is het dus mogelijk om voor ieder water een ecologisch optimale toestand te beschrijven.
Tot op heden worden de watertypen zoals beschreven in het Aquatisch Supplement (bijvoorbeeld sloten (Nijboer 2000) en beken (Verdonschot 2000)) als referenties voor de Nederlandse binnenwateren beschouwd. Voor ieder watertype is in principe de natuurlijke ecologische toestand van (een deel van) het watersysteem beschreven. Deze beschrijving fungeert daarmee als referentie voor zo'n watersysteem. In deze beschrijving is de biotiek opgenomen als een opsomming van de meest kenmerkende macrofyten, macrofauna en vissen. De abiotische omstandigheden zijn richtinggevend voor de milieu-omstandigheden waaronder het type zich optimaal ontwikkelt.
Echter de watertypen in het Supplement zijn voor de biotiek niet gekwantificeerd noch is de volledige gemeenschap opgenomen. Iedere gemeenschap bevat namelijk naast de meest kenmerkende (vaak zeldzame) soorten ook een aantal meer algemene. De laaste categorie ontbreekt. Het is daarom onmogelijk om de referenties uit het Supplement te betrekken in de typologieën voor beken en sloten, behalve een kwalitatieve koppeling gebaseerd op kenmerkende soorten.
Dit betekent dat een rekenkundige waardering gebaseerd op een afstand ten opzichte van de referentie, binnen dit project, nog niet haalbaar is omdat daarvoor de referenties compleet en gekwantificeerd moeten zijn. Een alternatief is een benadering gebaseerd op andere systeemkenmerken. Karr *et al*. (1986) groepeerden hiertoe de belangrijkste milieuvariabelen in vijf groepen:

1. Afvoerhydrologie (Stroming in termen van het 5-S-model (Verdonschot *et al.* 1995)); waaronder variabelen zoals stroomsnelheid, runoff, stroomgebiedskarakteristieken, waterkwantiteit, afvoerdynamiek (piek- en dalafvoeren), neerslag en grondwaterstroming genoemd zijn;
2. Habitatstructuren (Structuren in het 5-S-model); waartoe oeverstabiliteit, oeverbegroeiing, breedte, diepte, verval, dwarsdoorsnede vorm, watervegetatie, substraten, beschaduwing en sinuositeit behoren;
3. Waterchemie (Stoffen in het 5-S-model); waaronder organische verbindingen, nutriënten, zware metalen, opgeloste organische verbindingen, pH, troebelheid, hardheid, temperatuur, alkaliniteit en andere opgeloste stoffen gerekend worden;
4. Energie-bronnen (processen die in het 5-S-model in de interacties zijn opgenomen); waarbij beschikbaarheid van voedingsstoffen, zonlicht, toevoer van organisch materiaal (bijvoorbeeld bladinval), primaire en secundaire productie en seizoenen genoemd zijn;
5. Biotische interacties (Soorten in het 5-S-model); met competitie, reproductie, ziekten, parasitismne, voedingswijze en predatie als voorbeelden.

Een ecosysteemwaardering dient niet in te zoomen op één van de genoemde kenmerken maar dient alle vijf de aspecten van het systeem mee te nemen. Dit vraagt om een beoordelingssysteem waarin die indices zijn opgenomen die de belangrijkste processen verantwoordelijk voor de natuurlijke samenstelling van een water weergeven. In hoofdstuk 8 is een eerste stap in deze richting gezet gebaseerd op beschikbare informatie.
De watertypen in het Supplement zijn wat betreft de milieuvariabelen slechts richtinggevend geschetst. Deze richtinggevende getallen kunnen betrokken worden

in de waardering. Tenslotte vormen in RISTORI de milieuvariabelen de ingang. Om de mogelijkheden nader te onderzoeken worden de beschikbare waarden van de variabelen van de watertypen uit het Supplement tezamen met nieuwe voor het voorspellingsmodel benodigde waarden getest.

7.4 Abiotische referenties en waardering

7.4.1 Afstand tot de referentie bepaald met het voorspellingsmodel

Voor de beken zijn de voorspellende variabelen uit het Aquatisch Supplement genomen (Verdonschot 2000) en gekoppeld aan de bekentypologie. Daarbij zijn de waarden waar nodig enigszins bijgesteld en de ontbrekende waarden zijn ten behoeve van het model aangevuld (tabel 7.1).

Tabel 7.1 Voor het model geselecteerde abiotische toestandsvariabelen van de beekreferentietypen

parameter	een-heid	dvbj	dvbl	zbj	zbl	zml	ssbj	ssbl	ssml	ssbb	ssri	lsbj	lsbl	lsml	lsbb	lsri
pf-meander	nom	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
pf-dwarsnat	nom	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
perman	nom	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
pf-stuw	nom	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sei-winter	nom	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
schaduw	%	75	75	50	50	30	75	50	30	10	5	75	50	30	10	5
subslib	%	0	0	0	0	20	0	0	10	20	20	0	5	10	20	30
diepte	m	0.15	0.4	0.15	0.40	0.70	0.1	0.2	0.7	1	1.2	0.15	0.4	0.7	1	2
strmsn	m/s	0.2	0.2	0.1	0.1	0.1	0.4	0.4	0.4	0.5	0.5	0.15	0.15	0.1	0.1	0.1
breedte	m	0.75	2	0.75	2	5	0.75	1.5	4	7.5	20	0.75	2	5	10	20
subzand	%	30	30	30	30	40	30	30	50	60	60	30	40	50	60	70
pHmed	-	5.5	5.5	5.5	5.5	6.5	7	7	7.25	7.25	7.5	6.8	6.8	7	7.25	7.5
Clmed	mg/l	40	40	10	20	20	20	20	30	40	50	20	30	40	40	50
egvmed	µS	250	250	125	150	175	250	250	275	300	375	200	200	225	250	250
NH490	mgN/l	0.40	0.40	0.40	0.40	0.40	0.1	0.1	0.1	0.4	0.4	0.4	0.4	0.4	0.4	0.4
O2geh10	mg/l	8	7	8	7	6	9	9	8	7	7	9	8	7	6	6
Ptot90	mgP/l	0.04	0.04	0.02	0.02	0.02	0.02	0.02	0.02	0.04	0.04	0.04	0.04	0.04	0.04	0.04
NO310	mgN/l	0.50	0.50	0.50	0.50	0.50	0.35	0.35	0.35	0.5	0.5	0.35	0.35	0.35	0.5	0.5
Nkjel90	mgN/l	3	3	3.00	3.00	3.00	1.5	1.5	1.5	3	3	1.5	1.5	1.5	3	3

Legenda: referentietypen: dv=droogvallend, z=zuur, ss=snelstromend, ls=langzaam stromend, bj=bovenloopje, bl=bovenloop, ml=middenloop, bb=benedenloop, ri=riviertje; de paramteterscodes zijn verklaard in tabel 4.1.

Met het model zijn op basis van de parameterwaarden (tabel 7.1) de afstanden ten opzichte van de verschillende beektypen berekend. De kleinste afstand geeft aan op welk beektype het referentietype het meest gelijkt en hoeveel verschil nog tussen beide typen resteert (tabel 7.2).
Uit tabel 7.2 volgt dat deze benadering een grote afstand laat zien tussen de huidige beektypen en de referenties, behalve voor de cenotypen 2 (met (zwak) zure bovenlopen) en 8 (met droogvallende bovenlopen/loopjes). Dit is een gevolg van de

in het model opgenomen parameters ten opzichte van de belangrijke in de referenties. Deze parameters zijn wel onderscheidend voor de huidige typen maar minder voor de referenties. In de referenties zijn (mogelijk) andere parameters van veel groter belang. Daarnaast zijn de voor de referenties aangenomen waarden slechts schattingen die nauwelijks gebaseerd zijn op daadwerkelijke metingen.

Tabel 7.2 De afstand van de referenties ten opzichte van de meest verwante huidige beektypen (voor de verklaring van codes zie tabel 7.1)

referentietype	meest gelijkend beekcenotype(n)	gekarakteriseerd als:	afstand (schaal 0-1)
dvbj	8	dvbj/bl	0.131
dvbl	8	dvbj/bl	0.000
zbj	27	z/lsbj	1.000
zbl	2	zbl	0.000
zml	31	zml	1.000
ssbj	24b	ssbj/bl	0.477
ssbl	24b; 3b	ssbj/bl; ssbl	0.553; 1.000
ssml	16	ssml	0.999
ssbb	20	ssml/bb	0.768
ssri	14a	lsri	1.000
lsbj	21	lsbj	0.999
lsbl	1	lsbl	0.991
lsml	6	lsml/bb	0.997
lsbb	19	ls/ssbb	0.999
lsri	14a	lsri	1.000

7.4.2 Afstand tot de referentie bepaald met PCA

Om de afstand tussen de referentietypen uit het Aquatisch Supplement te vergelijken met de cenotypen uit de bekentypologie is een Principale Componenten Analyse (PCA) uitgevoerd. Deze lineaire ordinatie methode is uitgevoerd onder “centrering” en “standaardisatie” van de variabelen waardoor een correlatiediagram is verkregen. Het diagram (figuur 7.1) laat een duidelijke scheiding zien tussen de bekencenotypen en de referentietypen. Van links naar rechts loopt door het diagram een dimensiegradiënt van bovenloopjes naar riviertjes. In z’n algemeenheid kan geconcludeerd worden dat;

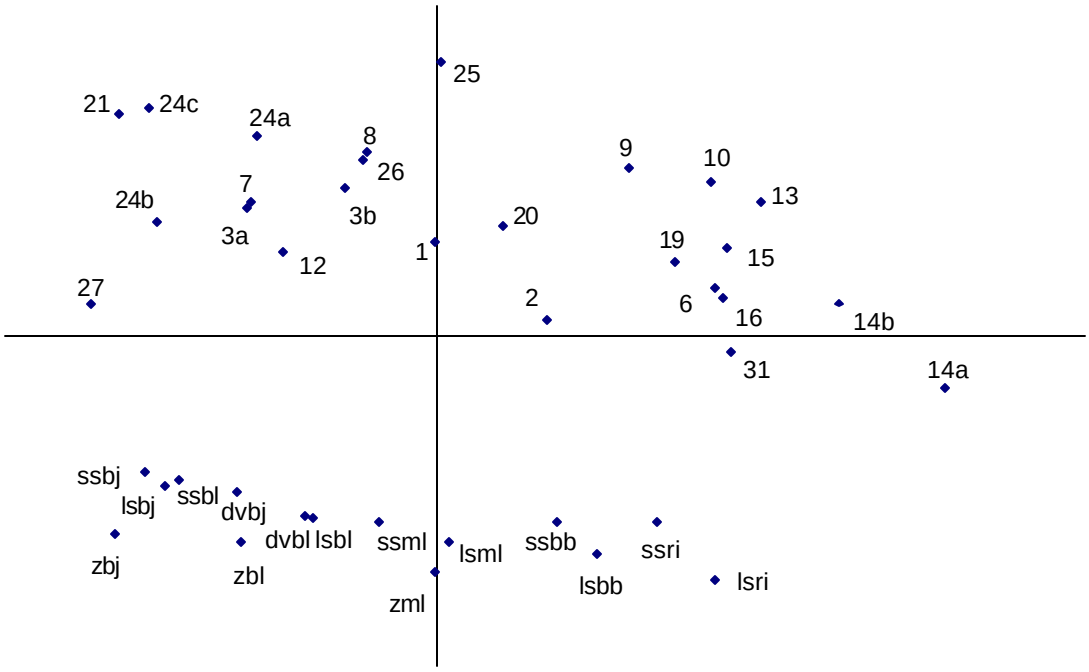
- de huidige bekencenotypen allemaal nagenoeg even ver van de referenties staan;
- de gehanteerde milieuvariabelen in het model niet per sé de onderscheidende variabelen voor de referenties behoeven te zijn;
- het gebruik van afstandsmaten zonder gekwantificeerde kennis (metingen) van referentietoestanden minder zinvol lijkt.

Om deze redenen is vooralsnog niet overgegaan tot het toepassen van deze methoden op de slotengegevens.

7.5 Discussie en conclusies

De Europese Kaderrichtlijn Water gaat uit van de beoordeling van de toestand van een water ten opzichte van de referentie. Wanneer gewerkt wordt met een index

(bijvoorbeeld de compleetheid van de gemeenschap) dan wordt de score op de maatlat (uitgedrukt in termen van de afstand ten opzichte van de referentie) berekend door de huidige toestand (in het voorbeeld een huidige toestand met een minder complete gemeenschap) te delen door de referentie (in het voorbeeld de complete gemeenschap). De uitkomst is een waarde die loopt van 0 (beide gemeenschappen zijn volledig verschillend) tot één (beide gemeenschapppen zijn even compleet). Door vervolgens deze schaal van 0-1 op te delen in vijf stukken worden de kwaliteitsklassen gedefinieerd.



Figuur 7.1PCA-diagram voor as 1 en 2 met de bekencenotypen (cijfers) en de referentietypen (codes)

Uit de paragrafen 7.3 en 7.4 volgt dat de referentie vooralsnog voor de Nederlandse oppervlaktewateren onvoldoende is gekwantificeerd om in dit kader toe te passen. Een dergelijke afstandsbenadering ontwikkelen valt buiten de doelstellingen van dit onderzoek. Wel zijn de keuzen die gemaakt zijn in de hoofdstukken 8 en 9, geleid door de overwegingen die voortvloeien uit de Kaderrichtlijn.
De koppeling tussen de watertypen uit het Aquatisch Supplement en de cenotypen is wel te maken, echter daarvoor is een completerings- en kwantificeringsslag van het eerste nodig.

8 Waarderen

8.1 Inleiding

Een aan voorspelling gekoppeld waarderingssysteem stelt gebruikers in staat het kwalitatieve effect van een voorgenomen maatregel op een watersysteem (het resultaat van een voorspelling) te waarderen. Waarderen is het kwalificeren van de voorspelde toestand; in termen van de Kaderrichtlijn het bepalen van de afstand tussen de feitelijke toestand en de beste toestand. Voor het bepalen van deze afstand is een maatlat nodig. De uiteinden van een dergelijke maatlat kunnen overeenkomen met de uiteinden van een ecologische ontwikkelingsreeks die een water kan doorlopen. Het ene uiteinde van de ontwikkelingsreeks is dood water, terwijl het andere de referentie betreft (Gardeniers 1976; Verdonschot 1983). Echter ook andere uiteinden zijn mogelijk. De referentie betreft een omschrijving van de natuurlijke of meest gewenste toestand, te bepalen aan de hand van ecologische karakteristieken. Goed gedefinieerde referenties zijn volgens de Europese Kaderrichtlijn Water onmisbaar in de ontwikkeling van een beoordelingssysteem. Afhankelijk van de wijze waarop een referentie wordt omschreven zal een kwantitatieve c.q. kwalitatieve maatlat ontwikkeld kunnen worden (zie hoofdstuk 7). Een aan de referentie gerelateerd waarderingssysteem heeft slechts één ijkpunt, de referentie, op de maatlat nodig.

In dit hoofdstuk zijn de eerste aanzetten gepresenteerd om te komen tot een nieuw waarderingssysteem voor de Nederlandse beken. In paragraaf 8.2 zijn de bestaande beoordelingsmethoden beschreven en zijn de stappen en keuzes om te komen tot een op een referentie gebaseerd waarderingssysteem uiteengezet. In paragraaf 8.3 is het proces uit paragraaf 8.2 toegespitst op de ontwikkeling van een nieuw waarderingssysteem voor de Nederlandse beken. In paragraaf 8.4 zijn de beperkingen gegeven die aan het ontwikkelde waarderingssysteem kleven. Tot slot zijn in paragraaf 8.5 conclusies getrokken en aanbevelingen gedaan.

Deze systematiek is alleen voor de beken uitgewerkt. Deze uitwerking heeft plaats gevonden in het kader van het, door de EU gefinancierd, project AQEM. Binnen RISTORI was een dergelijke uitwerking niet haalbaar, toch zijn niet alleen de resultaten maar is ook het proces dat in AQEM is doorlopen, in dit rapport opgenomen. Het proces geeft inzicht in de mogelijke aanpak en de plussen en minnen van het ontwikkelen van waarderingssystemen. Deze kennis is relevant voor het vervolg en eveneens bruikbaar voor de sloten.

8.2 De ontwikkeling van een beoordelingssysteem

In paragraaf 8.2.1 is de noodzaak van een typologisch raamwerk voor het ontwikkelen van een waarderingssysteem uitgelegd. Daarnaast is beschreven op welke wijze wateren aan een type kunnen worden toegedeeld en welke haken en ogen

hieraan zijn verbonden. In paragraaf 8.2.2 is de van constructie van de maatlat beschreven. Tot slot is in paragraaf 8.2.3 een opsomming gepresenteerd van de bestaande beoordelingsystemen en is de werking van de maatlat toegelicht.

8.2.1 Het typologisch raamwerk

Het is onmogelijk beoordeling, beheer en normstelling op alle wateren tezamen dan wel op ieder oppervlaktewater afzonderlijk af te stemmen. Het afstemmen op afzonderlijke wateren is vanuit financieel oogpunt niet haalbaar. Het afstemmen op alle wateren tezamen is geen optie, omdat dan voorbij wordt gegaan aan het feit dat biotische en abiotische variabelen van wateren onderling sterk van elkaar kunnen verschillen. Daarom moet een beoordelingssysteem een typologisch raamwerk bevatten waarbinnen de beoordeling plaatsvindt. Een typologisch raamwerk zou per type een omschrijving moeten bevatten van de referentietoestand, om zo het uiteinde van de beoogde ontwikkelingsreeks duidelijk in beeld te krijgen. Het toekennen van een water aan een type kan gebeuren op basis van abiotische of biotische variabelen of een combinatie van deze twee. De wijze waarop toekenning plaatsvindt verschilt per beoordelingsmethode. Belangrijk is het opstellen van eenduidige criteria voor het toekennen van wateren aan typen.
Het probleem met het toekennen van een water aan het juiste type is, dat naarmate een water meer beïnvloed is de abiotische en biotische variabelen steeds minder gaan lijken op de variabelen in de referentietoestand. Hoe groter de mate van beïnvloeding, hoe meer ook de levensgemeenschappen op elkaar zullen gaan lijken (Verdonschot, 1983). Er zijn verschillende manieren waarop dit probleem kan worden aangepakt. Wel moet worden opgemerkt dat geen van de methoden een sluitende oplossing biedt voor het probleem en dat de methode ook doelafhankelijk is. De meest eenvoudige benadering is het gebruik van de breedte van een beek. Wateren kunnen ook toegekend worden op basis van expert-judgement. Hierbij wordt vaak een landschapsecologische benadering gevolgd. De huidige ligging van een beek in het landschap in combinatie met het hydrologisch potentieel van een beek kan inzicht geven in hoe de beek er vroeger uit moet hebben gezien. Ook kan toedeling plaatsvinden aan de hand van gemeenschappen of cenotypen. Voorbeelden van toedelingstechnieken zijn opgenomen in EKOO (Verdonschot, 1990a) en ASSOCIA (Tongeren, 2000), waarbij EKOO is gebaseerd op cenotypen en ASSOCIA op vegetatietypen. Voor de ontwikkeling van een netwerk van cenotypen zoals toegepast in EKOO worden wateren met een zo groot mogelijke spreiding in milieuomstandigheden onderzocht. Bepaalde rekentechnieken (clusteranalyse en canonische ordinatie-technieken) maken het mogelijk om groepen wateren te beschrijven in termen van overeenkomende soortensamenstelling en gemiddelde milieuomstandigheden. De kern van de toegepaste cenotypologie is het om praktische redenen onderscheiden van ecologische groepen wateren in het van nature aanwezige continuüm aan combinaties van soorten en milieuomstandigheden (Verdonschot, 1990b). De relaties tussen cenotypen worden weergegeven in het netwerk van cenotypen. Dit netwerk geeft de onderlinge positie van de cenotypen, de overgangen tussen de cenotypen (een continuüm) en de belangrijkste milieugradiënten weer. Een monster kan met behulp van een programma zoals

EKOO achteraf worden toegedeeld aan een bestaand netwerk van cenotypen aan de hand van de soortensamenstelling en de abundantie van de macrofaunagemeenschap. Het type waaraan het monster wordt toegedeeld zegt wat over de mate van beïnvloeding van het water en de mate van overeenkomst met de referentietoestand van dat type. Opvallend is dat EKOO en ASSOCIA wateren of vegetaties toedelen aan typen op basis van de soortensamenstelling van gemeenschappen, terwijl in de literatuur toedeling voornamelijk geschiedt op basis van abiotische variabelen.

8.2.2 Constructie van de maatlat

Om te kunnen waarderen moet allereerst een ijkpunt op de maatlat, in dit geval de ecologische referentie, worden vastgesteld. Bij de constructie van de maatlat kon voor de referentie in dit onderzoek gekozen worden uit verschillende opties. Een optie is om de maatlat te baseren op de beste bestaande veldsituaties. In dit geval worden de referenties gevormd door de nog meest natuurlijke of optimale nog aanwezige toestanden van wateren in Nederland. Het nadeel hiervan is dat het doel voor Nederland laag wordt gesteld. Een tweede optie is om de referentietoestand voor de verschillende watertypen op te stellen zoals in hoofdstuk 7 beschreven is. Een derde optie is de referentie niet te gebruiken wanneer de beste toestand niet optimaal is en om deze beste maar beïnvloede toestand dan als tweede beste klasse te beschouwen. In feite wordt hiermee het ijkpunt tijdelijk verlaagd.
Wanneer het ijkpunt, de referentie, op de maatlat is vastgesteld, moet worden bepaald hoe gescoord wordt op de maatlat. In de benadering van de Europese Kaderrichtlijn Water is ervoor gekozen om de waarde van een bepaalde maat voor een monster te delen door de waarde voor die maat in de referentietoestand (EU, 2000). In het geval dat een referentietoestand wordt omschreven door ranges van variabelen kunnen scores worden toegekend aan de hand van afwijkingen van de mediaan (hoe groter de afwijking van de mediaan, hoe lager de score).

8.2.3 De waardering

Wanneer de afstand tussen de feitelijke en de gewenste toestand met behulp van de maatlat kan worden berekend, dan kan ook een waardering plaats vinden. Waarderen komt neer op het normeren van deze berekende afstand. De afstand op de maatlat kan worden onderverdeeld in klasse of anderzins. De grenzen tussen de klassen kunnen worden vastgesteld op basis van expert-judgement of ecologisch worden onderbouwd.

<u>Bestaande beoordelingsmethoden</u>
Er zijn inmiddels veel verschillende methoden waarop de beoordeling van een watersysteem kan worden gebaseerd beschreven. De beoordelingsmethoden zijn in te delen in acht categorieën:

1. kwalitatieve maten voor taxonomische rijkdom (kwalitatieve soortenrijkdom);
Deze maten beschrijven het aantal gevonden taxonomische eenheden en daarmee de structuur van de gemeenschap in een water. De redenering is dat de taxonomische rijkdom geleidelijk afneemt met een verslechtering van de waterkwaliteit (Weber 1973, Resh & Grodhaus 1983) en dat het ene taxon gevoeliger is voor beïnvloeding dan de ander. Opgemerkt moet worden dat de gedachte achter deze indices nogal gedateerd is. Het is inmiddels algemeen bekend dat bijvoorbeeld relatief schone middenlopen minder soorten kunnen bevatten in vergelijking tot meer vervuilde, maar plantenrijke middenlopen. Dit laatste maakt deze maten minder geschikt.
Voorbeelden: aantal taxa, aantal EPT taxa, aantal families

2. kwantitatieve maten voor taxonomische rijkdom (kwantitatieve soortenrijkdom);
Deze maten omvatten het tellen van alle verzamelde organismen in een monster tot het schatten van de relatieve abundantie van verschillende taxonomische groepen (verhoudingsmaten) of het bepalen van de biomassa. Aangenomen wordt, dat afhankelijk van het type stress het aantal individuen of de totale biomassa toe- of afneemt (Weber 1973, Resh & Grodhaus 1983, Plafkin *et al.*, 1989, Bode, 1988). Ook in deze benaderingen wordt ervan uitgegaan dat het ene taxon gevoeliger is voor vervuiling dan het andere. Er zijn overigens ook opvattingen die ervan uitgaan dat bij stress verschuivingen optreden binnen de gemeenschap, waarbij sommige soorten in aantal of biomassa zullen toenemen terwijl anderen in aantal of biomassa zullen afnemen. Dit laatste maakt deze maten minder geschikt.
Voorbeelden: aantal individuen , ratio [EPT abundantie] : [Chironomidae abundantie], ratio [individuen van numeriek dominante taxa] : [totaal aantal individuen]

3. diversiteitsindices;
Deze indices combineren soortenrijkdom, evenness (evenredige verdeling van de individuen over de aanwezige soorten in een levensgemeenschap) en abundantie. De beoordeling van watersystemen op basis van de diversiteit berust op de aanname, dat onbeïnvloede wateren worden gekenmerkt door een hoge soortenrijkdom, een grote evenness en hoge aantallen individuen (Mason *et al.* 1985). Er bestaan zeer veel diversiteitsindices, maar de meeste worden slechts sporadisch gebruikt. De Shannon's Index (Shannon & Weaver 1949) wordt het meest gebruikt. Het is inmiddels gebleken dat de juistheid van de diversiteitsindices slechts geldig is binnen een (regionaal) watertype.
Voorbeeld: Shannon's Index

4. similariteitsindces;
Deze indices zijn een maat voor de overeenkomst of verschil in de samenstelling van twee gemeenschappen. De similariteitsindices kunnen worden onderverdeeld in twee categorieën: de associatiecoëfficiënten en de afstandsmaten. Associatiecoëfficiënten geven de overeenkomst tussen de soortensamenstelling van twee monsters weer. Er bestaan zeer veel associatiecoëfficiënten, de meeste zijn echter wiskundig aan elkaar gerelateerd (Williams & Dale 1965). De kwalitatieve associatiecoëfficiënten houden geen rekening met de relatieve abundantie van soorten in verschillende monsters, waardoor het belang van zeldzame soorten wordt overschat en het belang van algemene soorten wordt onderschat. Dit kan worden vermeden door het gebruik van

kwantitatieve associatiecoëfficiënten (Nijboer 1996). De overeenkomst tussen monsters kan ook worden uitgedrukt in een afstandsmaat. De relatie tussen monsters kan worden weergegeven in een geometrisch model. De afstand tussen de monsters kan dan gebruikt worden als een maat voor de overeenkomst tussen de monsters. De meest bekende afstandsmaat is de Euclidische afstand. Een nadeel van dergelijke ruimtelijke maten is dat ze weinig rekening houden met kwalitatieve verschillen.
Voorbeelden: Kothe's species deficit (Kothe 1962), similariteits-index van Sørensen, Jaccard-index (Jaccard 1912), associatie-index van Whittaker , Czekanowski coëfficiënt (Czekanowski 1913), Euclidische afstand, Mahanalobis afstand

5. milieu-indices;
Deze indices indiceren bepaalde dominante milieukenmerken. Ze maken gebruik van vastgestelde waarden voor de milieu-indicaties van taxa. De relatieve abundantie van een taxon, gewogen voor indicatiewaarden, wordt soms gebruikt in de berekening van de index. Ook hier wordt dus uitgegaan van het verschil in respons voor een specifieke milieufactor van verschillende taxa. De meest klassieke is de saprobie-index gebaseerd op het ammoniumgehalte (bijvoorbeeld Zelinka en Marvan 1961).
Voorbeelden: stroomsnelheidsindex, saprobie-index, zuurgraad-index, droogval-index

6. biotische indices;
Deze indices maken gebruik van vastgestelde waarden voor de vervuilingstolerantie van taxa. De relatieve abundantie van een taxon, gewogen voor tolerantiewaarden, wordt soms gebruikt in de berekening van een biotische index. Ook hier wordt dus uitgegaan van het verschil in tolerantie voor vervuiling van verschillende taxa.
Voorbeelden: Emperical Biotic Index (Chutter 1972), Biotic Score (Chandler 1970), BMWP Score (Armitage et al. 1983), Trent Biotic Index (Woodiwiss 1964), K_{135}*-index (Tolkamp & Gardeniers 1971)*
Een andere vorm van biologische indices zijn het gebruik van aantal indicatortaxa, compleetheid, kenmerkendheid en zeldzaamheid.

7. functionele en proces-indices, zoals functionele voedingsgroepen;
Indices afgeleid van functionele voedingsgroepen zijn gebaseerd op de morfologische structuren en gedragingen van soorten bij het verzamelen van voedsel. Bepaalde functionele groepen zijn gevoeliger voor vervuiling dan anderen. Bovendien zeggen de groepen wat over de aanwezige bron van voedsel en daarmee over de aanwezige organische verontreiniging of de mate van eutrofiëring.
Voorbeelden: ratio [aantal knippers] : [totaal aantal individuen] (Plafkin et al. 1989), ratio [aantal schrapers] : [aantal vergaarders-filtreerders]

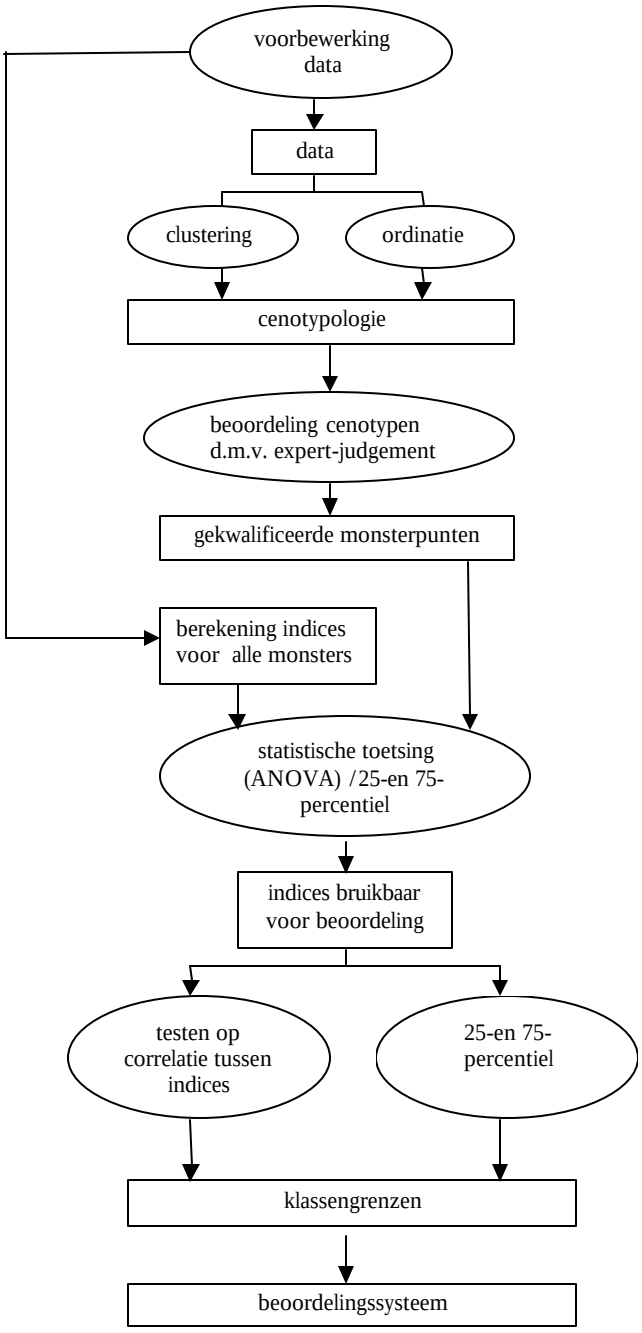
8. multimetrics of meervoudige indices
De meervoudige indices vormen een andere manier om de problematiek van de beoordelingssystemen te benaderen. Deze methode maakt gebruik van meerdere maten, die elk andere complementaire informatie verschaffen over een levensgemeenschap. De combinatie van deze maten functioneert als een overall indicator van de biologische conditie van een levensgemeenschap. De kracht van deze benadering is de mogelijkheid om informatie op individu, levensgemeenschap en ecosysteem niveau te integreren (Karr *et al.* 1986, Pafkin *et al*. 1989, Karr 1991).

Beoordeling op basis van meerdere maten geeft de mogelijkheid van detectie over een grotere range en aard van stressfactoren en geeft een completer beeld van de biologische conditie van een watersysteem dan individuele biologische indicatoren. Ohio EPA (1987) suggereert, dat de kracht van de gecombineerde maten, de zwakheden van de individuele maten minimaliseert.
Voorbeelden: Index of Biotic Integrity (Karr et al. 1986), de Mean Biometric Score (Shackleford 1988), de Biological Condition Score (Plafkin et al. 1989), tot op zekere hoogte EBEOSWA (STOWA 1992)

Bij gebruik van de besproken indices moet er rekening mee worden gehouden, dat verschillende indices vergelijkbaar kunnen scoren op basis van verschillende combinaties van stressoren. Verder kan de range van de indexwaarden gelimiteerd zijn en de schaal vertoont nog wel eens verminderde gevoeligheid bij extremen of in het centrum. Het is daarom belangrijk om het gedrag van indices te bepalen, door ze toe te passen in bekende omstandigheden zodat een goed inzicht kan worden verkregen in hun eigenschappen en vooral hun beperkingen (Hellawell 1978). Om een aanzet te kunnen geven voor de ontwikkeling van een nieuw beoordelingssysteem moeten indices worden toegepast op wateren van verschillende beïnvloedingsstadia van alle watertypen. Op deze wijze kan worden bepaald welke indices het meest geschikt zijn voor gebruik in een nieuw beoordelingssysteem. Belangrijk is om op te merken dat de similariteitsindices, in tegenstelling tot de overige maten, monsters op een objectieve manier met elkaar vergelijken (de soortensamenstelling komt meer of minder overeen). Alle overige maten zijn echter gebaseerd op aannames over het functioneren van levensgemeenschappen. Welke relaties de indices met de zes hoofdcomponenten van het ecosysteem functioneren hebben dient vooraf te worden vastgesteld. Tot slot moet worden vermeld dat aan een kwalitatieve maat het nadeel kleeft, dat zeldzame soorten over- en dominante soorten ondergewaardeerd kunnen worden (Nijboer 1996). Bij kwantitatieve indices is hiervan geen sprake, omdat rekening wordt gehouden met de abundantie van de aanwezige taxa. Een groot voordeel van de kwalitatieve indices is echter, dat de gegevensverzameling minder tijdrovend is en dus kostenbesparend.

8.3 De ontwikkeling van een waarderingssysteem voor Nederlandse beken

Op basis van de in paragraaf 8.2 genoemde voordelen is ervoor gekozen om een multimetric benadering te ontwikkelen. In paragraaf 8.3.1 zijn de hiervoor te onderzoeken indices vermeld. In paragraaf 8.3.2 is beschreven op welke wijze de meest geschikte indices zijn geselecteerd en hoe deze indices gecombineerd zijn tot een waarderingssysteem. De bruikbaarheid van deze indices voor de waardering van Nederlandse beken is afgeleid van de werking onder verschillende aard en mate van beïnvloedingen en bij verschillende beektypen. Tot slot is in paragraaf 8.3.3 inzicht gegeven in de betrouwbaarheid van het ontwikkelde systeem. De verschillende stappen die zijn genomen en technieken die zijn gebruikt bij de ontwikkeling van het waarderingssysteem staan schematisch weergegeven in figuur 8.1.



Figuur 8.1Stroomschema waarin de stappen staan weergeven, die zijn doorlopen bij de ontwikkeling van het beoordelingssysteem. De ovalen in het stroomschema staan voor de gebruikte technieken en de vierkanten voor het bereikte resultaat

8.3.1 De onderzochte indices

De te onderzoeken indices (bijlage 12) zijn afkomstig van het Europese onderzoeksproject AQEM (AQEM consortium, 2002). In tabel 8.1 is een overzicht geven van het aantal indices per categorie. Hieruit blijkt dat alle categorieën vermeld in paragraaf 8.2 zijn vertegenwoordigd, met uitzondering van de similariteitsindices en de multimetrics. De laatste groepvormt een combinatie van indices uit de voorgaande groepen. De keuze voor de indices is vooral gebaseerd op de mate waarin de indices in gebruik zijn bij de aan het AQEM-project deelnemende landen.

Tabel 8.1 Weergave van het aantal indices per categorie geselecteerd voor toetsing op bruikbaarheid

categorieën	aantal indices
Kwalitatieve soortenrijkdom	33
Kwantitatieve soortenrijkdom	32
Diversiteitsindices	3
Milieu-indicaties	14
a. stroming	7
b. microhabitat	7
Biotische indices	16
Maten op basis van proceskenmerken	26
a. functionele voedingsgroepen	11
b. bewegingsgedrag	5
c. zonatie	10

8.3.2 Selectie indices geschikt voor de waardering van Nederlandse beken

In eerste instantie, was het uitgangspunt voor dit project het selecteren van geschikte indices op basis van de uitkomsten van het onderzoek aan het AQEM gegevensbestand (Vlek *et al.* 2002). De resultaten hiervan bleken weinig bemoedigend. Daarom is besloten de selectie van de indices te baseren op de monsters/cenotypen uit de bekentypologie (Verdonschot & Nijboer 2002). Op basis van expert-judgement zijn aan de beekcenotypen kwaliteitsklassen toegekend (een zogenaamde post-classificatie). In tabel 8.2 is het resultaat van de post-classificatie weergegeven. Bij de post-classificatie is op basis van de samenstelling van de macrofaunagemeenschap en de abiotische kenmerken, aan ieder beekcenotype een kwaliteitsklasse toegekend. Het toekennen van kwaliteitsklassen bleek niet eenvoudig. Tussen de monsters in het gegevensbestand bestonden namelijk naast kwaliteitsverschillen ook typologische verschillen. Om dit probleem gedeeltelijk te omzeilen is het gegevensbestand gesplitst in twee typen: langzaam stromende beken (N01) en snelstromende beken (N02). Beken met een stroomsnelheid lager dan 30 cm/sec zijn beschouwd als langzaam stromend, beken met een hogere stroomsnelheid als snel stromend. Cenotypen, die voor meer dan de helft bestonden uit monsters van langzaam stromende beken, zijn beschouwd als langzaam stromend. Cenotypen, die voor meer dan de helft bestonden uit monsters van snel stromende beken, zijn beschouwd als snel stromend. In het geval dat het aantal langzaam en snel stromende beken behorend tot één cenotype elkaar niet veel ontliep, zijn de monsters van het betreffende cenotype zowel in het gegevensbestand langzaam

stromend als snel stromend opgenomen. Alle droogvallende en zure beken zijn buiten beschouwing gelaten. Op deze wijze hebben de belangrijkste typologische verschillen (namelijk stroomsnelheid, droogval en zuurgraad) in het gegevensbestand geen rol gespeeld bij het toekennen van de kwaliteitsklassen. De beken zijn gewaardeerd van 1 (zeer slechte kwaliteit) tot 4 (goede kwaliteit). Het gegevensbestand bevatte geen beken van zeer goede kwaliteit, ofwel de referentietoestand, aangezien beken van dergelijke kwaliteit in Nederland niet of nauwelijks meer voorkomen en in ieder geval in het bekenbestand ontbraken. De onderscheiden kwaliteitsklassen komen overeen met de kwaliteitsklassen beschreven in de Europese Kaderrichtlijn Water.
Er is vooralsnog niet gekozen om uit te gaan van een referentie gerelateerde benadering omdat;

1. een werkelijke referentie ontbrak,
2. de kwaliteitsklassen mogelijk niet lineair zijn gerelateerd aan de afstand tot de referentie,
3. een dergelijke benadering om redenen van 1 en 2 meer onderzoek vraagt.

Tabel 8.2 Resultaten van de post-classificatie op basis van expert-judgement

cenotype	**post-classificatie (N01)**	**post-classificatie (N02)**
3a	4	
21	4	4
24a	4	4
24b		4
24c	4	
1	3	
3b		3
9	3	1
19	3	1
20		3
25		3
26	3	
10	2	2
13	2	
15	2	2
6	1	
14a	1	
14b	1	
16		1
31	1	

Na het toekennen van de kwaliteitsklassen, is gekeken met een ANOVA of de indexwaarden tussen de kwaliteitsklassen significant van elkaar verschilden ($p < 0.05$). In bijlage 13 is per index vermeld welke klassen al of niet significant van elkaar verschillen. Indien een index significant verschil aantoont tussen bepaalde kwaliteitsklassen is de volgende stap, het opstellen van klassengrenzen, doorlopen. Gekozen is om de klassengrenzen vast te stellen aan de hand van het 25- en 75-percentiel van de indexwaarden voor de kwaliteitsklassen (ITFM, 1993). Bij het vaststellen van het 25- en 75-percentiel voor de verschillende indices, bleek dat in een groot aantal gevallen overlap bestond tussen het 25-percentiel van één klasse en het 75-percentiel van een andere klasse, of tussen het 25-percentiel c.q. 75-percentiel van één klasse en de mediaan van een andere klasse. Hoe groter de overlap van

indexwaarden tussen kwaliteitsklassen, hoe groter de kans dat een monster foutief wordt beoordeeld. In bijlage 13 is een overzicht gegeven van alle indices met eventuele overlap.
Vervolgens is onderzocht of er indices zijn die alle kwaliteitsklassen significant van elkaar onderscheiden zonder overlap van 25- en 75-percentiel of overlap van 25- of 75-percentiel met de mediaan. Uit bijlage 13 blijkt dat geen van de indices aan deze voorwaarden voldoet. Daarom zijn uiteindelijk indices geselecteerd die één (of meerdere) klasse(n) van alle andere klassen onderscheiden. In tabel 8.3 en 8.4 zijn de geselecteerde indices weergegeven. Omdat elke index weer andere kwaliteitsaspecten van de macrofaunagemeenschap benadrukt, is ervoor gekozen om minimaal twee indices te selecteren, die één kwaliteitsklasse onderscheiden. Als maximum is een aantal van vier indices per klasse gesteld. Bij voorkeur zijn indices uit verschillende categorieën geselecteerd, dit om het diagnostisch karakter van de multimetric te behouden. Als laatste criterium is gesteld dat indices, discriminerend voor dezelfde kwaliteitsklasse, niet gecorreleerd mogen zijn (correlatie <0.7).

Tabel 8.3 Indices geselecteerd voor de beoordeling van langzaam stromende beken (N01)

indices	**discriminerend voor klasse:**	**klassengrens**
Saprobie-index (Zelinka en Marvan)	4	<2.12
hypopotamal [%]	4	<0.55
type PEL [%]	4	<8.4
type RP [%]	4	>29.4
EPT/OL [%]	2	<0.51
EPT/OL	2	<0.67
hypopotamal [%]-EPT/OL [%]	3	<3.22 - >0.91 en >1.3
Gastropoda-EPT/OL [%]	3	<=6 - >=2 en >1.3
grazers+schrapers/vergaarders+ filtreerders	1	>2
Gastropoda [%]	1	>9.92

Tabel 8.4 Indices geselecteerd voor de beoordeling van snel stromende beken (N02)

indices	**discriminerend voor klasse:**	**klassengrens**
Saprobie-index (Zelinka & Marvan)	4	<2.02
	2	>=2.46
	1	>2.27 - <2.46
hypopotamal [%]	4	<0.12
	3	>=0.12 - <0.9
aantal taxa	4	<=24
passieve filtreerders [%]	3	>1.65
EPT/OL [%]	3	>22
	2	<0.61
EPT/OL	2	<0.71
Trichoptera [%]	2	<0.23
EPT/OL [%]-Gastropoda [%]	1	<16.4 - >1.19 en >=0.12
EPT/OL [%]-type RP [%]	1	<16.4 - >1.19 en <=53.5
EPT/OL [%]-type PEL [%]	1	<16.4 - >1.19 en >=4.36

Uit tabel 8.3 en 8.4 blijkt dat in sommige gevallen (klasse 3 van N01 en klasse 1 van N02) een combinatie van twee indices is gebruikt om één klasse te onderscheiden. In dit geval konden de betreffende klassen niet met één index worden onderscheiden (bijlage 13).

Om nieuwe monsters te kunnen classificeren zijn per index klassengrenzen vastgesteld (tabel 8.3 en 8.4). De klassengrenzen zijn gevormd door het 25-percentiel, 75-percentiel of een combinatie van beide. In het geval van een combinatie van twee indices zijn voor beide indices de klassengrenzen bepaald. Na het vaststellen van de klassengrenzen is het mogelijk om aan de individuele indices een score toe te kennen. Wanneer de indexwaarde van een nieuw monster binnen de klassengrenzen valt, scoort een monster (afhankelijk van de klasse waarvoor de index onderscheidt) een 1 (zeer slecht), 2 (slecht), 3 (gemiddeld) of 4 (goed). Indien een combinatie van twee indices is gebruikt om een klasse te onderscheiden, dan moeten beide indexwaarden binnen de gestelde klassengrenzen vallen om te scoren. Als de indexwaarde niet binnen de klassengrenzen valt, is niet gescoord en is de betreffende index buiten beschouwing gelaten.
Omdat een combinatie van indices meer zegt over de kwaliteit van een macrofaunagemeenschap dan een individuele index, zijn de indices gecombineerd tot een multimetric. Het berekenen van het multimetric resultaat komt neer op het gewogen middelen van de individuele scores voor de indices. Omdat niet voor elke kwaliteitsklasse een vergelijkbaar aantal indices is geselecteerd, is gecorrigeerd voor het aantal indices per kwaliteitsklasse. Hiervoor zijn de volgende multimetric formules opgesteld:

<u>Langzaam stromende beken</u>

$$S = \frac{T_1 * \frac{1}{2} + T_2 * \frac{1}{2} + T_3 * \frac{1}{2} + T_4 \frac{1}{4}}{n_1 * \frac{1}{2} + n_2 * \frac{1}{2} + n_3 * \frac{1}{2} + n_4 * \frac{1}{4}}$$

met:
S : score
T_1 : som van alle scores voor klasse 1
T_2 : som van alle scores voor klasse 2
T_3 : som van alle scores voor klasse 3
T_4 : som van alle scores voor klasse 4
n_1 : aantal indices scorend voor klasse 1
n_2 : aantal indices scorend voor klasse 2
n_3 : aantal indices scorend voor klasse 3
n_4 : aantal indices scorend voor klasse 4

Snel stromende beken

$$S = \frac{T_1 * \frac{1}{4} + T_2 * \frac{1}{4} + T_3 * \frac{1}{3} + T_4 \frac{1}{3}}{n_1 * \frac{1}{4} + n_2 * \frac{1}{4} + n_3 * \frac{1}{3} + n_4 * \frac{1}{3}}$$

met:

S : score
T_1 : som van alle scores voor klasse 1
T_2 : som van alle scores voor klasse 2
T_3 : som van alle scores voor klasse 3
T_4 : som van alle scores voor klasse 4
n_1 : aantal indices scorend voor klasse 1
n_2 : aantal indices scorend voor klasse 2
n_3 : aantal indices scorend voor klasse 3
n_4 : aantal indices scorend voor klasse 4

Aan de hand van de met de bovenstaande formules berekende scores, wordt de kwaliteitsklasse van een monster met behulp van tabel 8.5 bepaald. De kwaliteitsklassen in tabel 8.5 komen overeen met de kwaliteitsklassen beschreven in de Europese Kaderrichtlijn Water.

Tabel 8.5 Klassengrenzen voor de verschillende kwaliteitsklassen

kwaliteitsklasse	**score**
5 (zeer goed)	niet aanwezig
4 (goed)	>=3.5 – 4
3 (gemiddeld)	>=2.5 – 3.4
2 (slecht)	>=1.5 – 2.4
1 (zeer slecht)	<1.5

8.3.3 Validatie van het waarderingssysteem aan de hand van post-classificatie

Om te bepalen of het in paragraaf 8.3.2 beschreven beoordelingssysteem ook daadwerkelijk goed functioneert, zijn de resultaten van het beoordelingssysteem vergeleken met de post-classificatie van de monsters uit de bekentypologie. Het beoordelingssysteem is alleen gevalideerd voor de monsters waarop het systeem betrekking heeft, monsters uit snel en langzaam stromende beken. Per index is de score van ieder monster voor de betreffende index bepaald. De scores voor de individuele indices zijn vergeleken met de kwaliteitsklassen die eerder aan de cenotypen zijn toegekend (post-classificatie). Vervolgens is voor iedere index het percentage foutief geclassificeerde monsters van het totaal berekend (tabel 8.6 en 8.7). Bij de classificatie kunnen twee soorten fouten worden gemaakt:

fout 1: een monster scoort niet voor een index, terwijl het monster wel in de kwaliteitsklasse valt die de index behoort te onderscheiden (percentage van het aantal monsters met de betreffende kwaliteitsklasse).

fout 2: een index scoort voor monsters, die bij post-classificatie kwalitatief beter of slechter zijn gewaardeerd (percentage van het aantal monsters niet behorend tot de betreffende kwaliteitsklasse).

Opgemerkt moet worden dat fout 1 en fout 2 niet met elkaar kunnen worden vergeleken, omdat fout 1 betrekking heeft op een veel kleiner aantal waarnemingen dan fout 2. Alleen verschillen tussen de indices onderling kunnen worden vergeleken.

Tabel 8.6 Percentage foutief geclassificeerde monsters per index voor N01

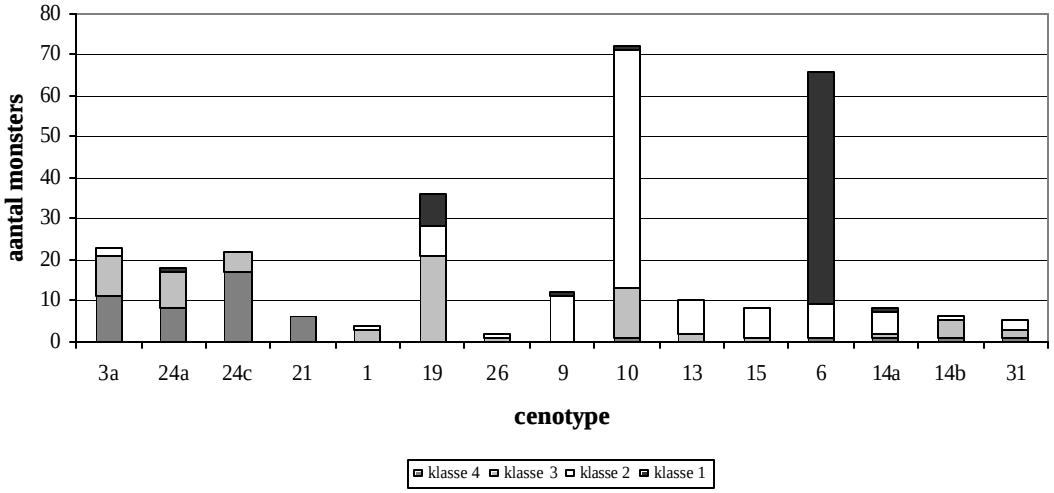
indices	klasse	% fout 1	% fout 2	% totaal foutief
saprobie-index	4	24	3	9
hypopotamal [%]	4	25	9	14
type RP [%]	4	24	4	10
type Pel [%]	4	25	17	19
Gastropoda-EPT/OL [%]	3	61	10	20
hypopotamal [%]-EPT/OL [%]	3	70	7	19
EPT/Ol [%]	2	25	21	22
EPT/Ol	2	26	18	21
grazers+schrapers/ verzamelaars+filtreerders	1	25	6	11
Gastropoda [%]	1	25	13	16

Tabel 8.7 Percentage foutief geclassificeerde monsters per index voor N02 (opmerking: de fout is afhankelijk van de gekozen klassengrenzen)

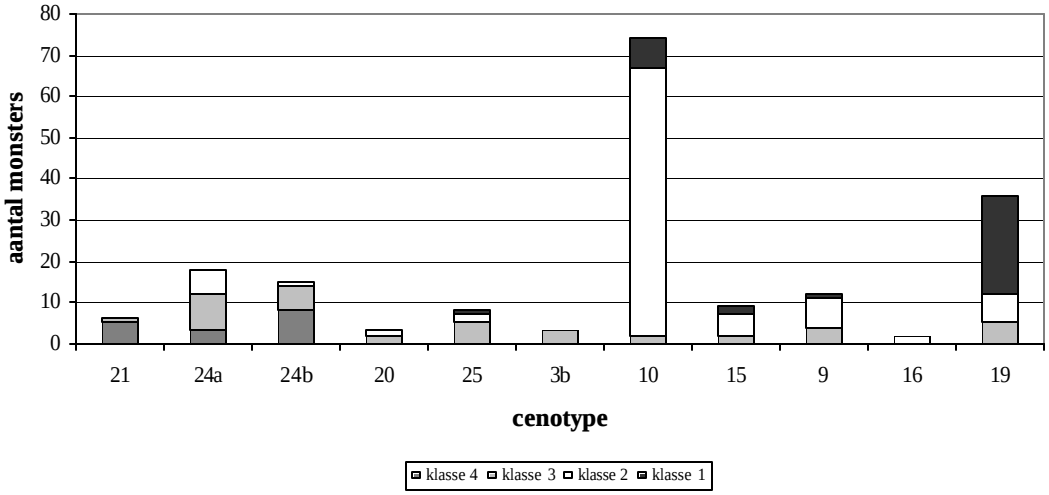
indices	klasse	% fout 1	% fout 2	% totaal foutief
aantal taxa	4	24	12	15
hypopotamal [%]	4	24	4	9
Saprobie-index	4	24	3	9
EPT/OL [%]	3	29	12	13
hypopotamal [%]	3	21	13	13
passieve filtreerders [%]	3	26	16	17
Saprobie-index	2	25	13	18
EPT/OL [%]	2	25	18	22
EPT/OL	2	25	18	22
Trichoptera [%]	2	25	14	18
Saprobie-index	1	30	12	20
EPT/OL [%]-Gastropoda [%]	1	52	15	25
EPT/OL [%]-type RP [%]	1	52	16	26
EPT/Ol [%]-type PEL [%]	1	52	15	25

Naast de foutenpercentages voor de individuele indices is tevens het percentage fouten in de uiteindelijke beoordeling vastgesteld. Met behulp van de formules uit paragraaf 8.2.2 is de uiteindelijke kwaliteitsklasse van elk monster bepaald. De uiteindelijke kwaliteitsklasse is per cenotype weergegeven in figuur 8.2 voor de

langzaam stromende beken en in figuur 8.3 voor de snel stromende beken. In tabel 8.2 is de post-classificatie per cenotype
vermeld.



Figuur 8.2 Verdeling van de uiteindelijke kwalificaties over de cenotypen voor N01



Figuur 8.3 Verdeling van de uiteindelijke kwalificaties over de cenotypen voor N02

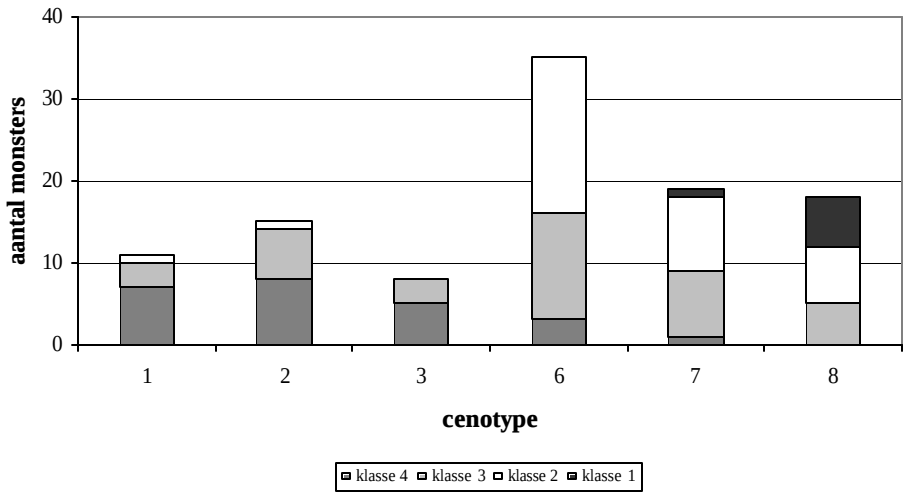
Uit figuur 8.2 en 8.3 en tabel 8.2 kan worden opgemaakt dat voor de meeste cenotypen geldt, dat er niet veel monsters zijn die meer dan één klasse afwijken van de post-classificatie. Tevens blijkt dat in de meeste gevallen de helft van alle monsters (of meer) behorend tot een cenotype worden geclassificeerd overeenkomstig de post-classificatie. Feit blijft dat de uiteindelijke kwalificatie in een aantal gevallen afwijkt van de post-classificatie. Hierbij zijn twee soorten afwijkingen c.q. fouten mogelijk:
type I fout: een monster is lager geclassificeerd dan in de post-classificatie.
type II fout: een monster is hoger geclassificeerd dan in de post-classificatie.
In tabel 8.9 is het percentage correct geclassificeerde monsters en het percentage type I en type II fouten gepresenteerd.

Tabel 8.9 Percentage correct en foutief beoordeelde bekentypologie monsters

beektype	% correct	% type I fout	% type II fout
N01	67	18	15
N02	65	19	16

Uit tabel 8.6 en 8.7 blijkt dat vooral bij het gebruik van gecombineerde indices fout 1 relatief vaak optreedt. In totaal zijn 67% van de monsters uit N01 en 65% van de monsters uit N02 juist geclassificeerd. Wanneer een afwijking van één klasse nog als acceptabel is beschouwd, zijn 92% van de monsters uit N01 en 91% van de monsters uit N02 juist geclassificeerd. Het percentage type I en type II fouten ontloopt elkaar niet veel en ligt tussen de 19% en 15%. Het percentage type II fouten is relatief laag vergeleken met het percentage type II fouten uit de studie van Johnson (1998). Johnson (1998) heeft voor vier indices (algemeen gebruikt in de Scandinavische landen om stress door verzuring te beoordelen) het percentage type I en type II fouten vastgesteld. Het percentage type I fouten lag tussen de 10 en 23% en het percentage type II fouten tussen de 27 en 55%.

Naast de monsters uit de bekentypologie zijn ook de monsters uit het AQEM gegevensbestand gebruikt om het waarderingssysteem te testen. Hierbij is alleen gekeken naar de fouten in de uiteindelijke beoordeling. De uiteindelijke kwaliteitsklasse is per cenotype weergegeven in figuur 8.4. In tabel 8.10 staat de post-classificatie per cenotype vermeld.



Figuur 8.4 Verdeling van de uiteindelijke kwalificaties over de cenotypen voor het AQEM gegevensbestand

Tabel 8.10 Resultaten van de post-classificatie op basis van expert-judgement voor het AQEM gegevensbestand

cenotype	post-classificatie
1	4
2	4
3	4
6	1
7	2
8	1

Uit figuur 8.4 en tabel 8.10 kan worden opgemaakt dat voor de cenotypen 1, 2, 3 en 7 geldt, dat er niet veel monsters zijn die meer dan één klasse afwijken van de post-classificatie. Tevens blijkt dat in de meeste gevallen de helft van alle monsters (of meer) behorend tot deze cenotypen worden geclassificeerd overeenkomstig de post-classificatie. In totaal worden slechts 33% van de monsters correct geclassificeerd. Het percentage type I en type II fouten bedraagt respectievelijk 14 en 53%. In figuur 8.4 is te zien dat voornamelijk cluster 6 en 8 verantwoordelijk zijn voor het lage percentage correcte beoordelingen.

8.4 Beperkingen van en aanbevelingen bij het beken waarderingssysteem

Voordat het beoordelingssysteem daadwerkelijk wordt toegepast in de praktijk is een uitgebreider onderzoek noodzakelijk. De validatie op basis van de monsters uit de bekentypologie gaf redelijke resultaten. Hierbij moet namelijk worden bedacht dat de post-classificatie ook een algemene noemer is waaronder alle monsters van een typen zijn geschaard. De uitkomsten van de test op basis van het AQEM gegevensbestand zijn minder vertrouwenwekkend. Het probleem is in ieder geval mede veroorzaakt door de wijze van opstellen van de post-classificatie. De post-classificatie voor de AQEM monsters is onafhankelijk van de post-classificatie voor de bekentypologie monsters uitgevoerd. Het vermoeden bestaat dat de bemonsterde AQEM locaties met de slechtste kwaliteit beter van gekwalificeerd zijn dan de monsters met de slechtste kwaliteit uit de beetenypologie. Simpel omdat de post-classificatie per bestand voor de variatie aanwezig binnen het betreffende bestand is uitgevoerd. Om deze reden kunnen bij de post-classificatie de AQEM monsters uit de clusters 6, 7 en 8 te laag zijn geclassificeerd. Het valideren van het waarderingssysteem met een dataset die alle kwaliteitsklassen goed dekt is dus noodzakelijk om harde uitspraken te kunnen doen over de consistentie en toepasbaarheid.
Om een waarderingsysteem te ontwikkelen en te testen is post-classificatie noodzakelijk. Zoals vermeld in paragraaf 8.3.2 is de post-classificatie in dit project gebaseerd op het kwalificeren, op basis van expert-judgement, van cenotypen. Deze vorm van kwalificeren is echter subjectief en arbitrair. Een alternatieve optie is het vergroten van de rol van de milieu-informatie bij de post-classificatie. Doordat het systeem is gebaseerd en gevalideerd op de post-classificatie is er ook sprake van een cirkelredenering. Een dergelijke cirkelredenering is echter onvermijdelijk. Wel kan bij een reproduceerbare post-classificatie-techniek de validatie onafhankelijk worden uitgevoerd.

Naast deze onvolkomenheden bij het gepresenteerde systeem zijn nog andere aanbevelingen van belang om te komen tot een verbeterd en vernieuwd waarderingssysteem:

1. Het verdient aanbeveling om de gecombineerde indices door enkelvoudige indices te vervangen. De gecombineerde indices zijn geselecteerd omdat binnen de kaders van dit onderzoek geen betere maten voorhanden waren.
2. De beschikbare indices blijken in het algemeen niet lineair gecorreleerd te zijn met het verloop van stressoren. Dit is mogelijk een gevolg van de aanwezige typologische variatie in de gegevens, van de zwakten van de gebruikte indices, van het ontbreken van voldoende gegevens van de betreffende klasse of van een combinatie van genoemde argumenten.
3. De typologische verschillen indiceren dat waarderingssystemen in ieder geval typegedifferentieerd opgesteld moeten worden. Met typedifferentiatie lijkt in veel gevallen ook een gebiedsdifferentiatie samen te hangen. Hiermee kan worden voorkomen dat typologische verschillen worden aangemerkt als kwaliteitsverschillen. Naast opsplitsing naar stroomsnelheid zou daarom opsplitsing naar dimensie (beekzones) en regio / gebied aan te bevelen zijn.
4. Met de zwakten van de gebruikte indices worden de inhoudelijke indicatiewaarden van de index bedoeld. Deze waarden zijn gebaseerd op een combinatie van Nederlandse en buitenlandse kennis. De waarden kunnen daardoor afwijken van de Nederlandse optima omdat ook mediterrane, droogvallende beken of hooggebergte beken invloed hadden op de indicatiewaarde van bepaalde taxa. Een degelijke autecologische optimalisatie voor de Nederlandse situatie kan de benadering sterk verbeteren.
5. Het aanbod aan indices was niet dekkend. Het is aanbevelingswaardig om indices met een groter onderscheidend vermogen te testen dan wel te ontwikkelen.
6. Al eerder is de nadruk gelegd op beoordelen aan de hand van de afstand tot de referentie. Deze manier van beoordelen ligt ten grondslag aan de EU Kaderrichtlijn Water. Het gepresenteerde beoordelingssysteem is niet gebaseerd op de afstand tot de referentie. Doordat niet wordt beoordeeld ten opzichte van de referentie blijft er altijd sprake van een bepaalde mate van subjectiviteit (wat is goede kwaliteit?). Natuurlijke beeksystemen ontbreken echter in Nederland. Referenties kunnen slechts worden omschreven aan de hand van historische data of door informatie te vergaren over de natuurlijke situatie in vergelijkbare buitenlandse beeksystemen. Wanneer referenties op deze wijze tot stand komen, kan nog steeds niet worden nagegaan of de beschrijving overeenkomt met toestand van een beeksysteem voor beïnvloeding. Het gevolg hiervan is, dat in Nederland ook bij beoordelen ten opzichte van een referentie sprake is van een zekere mate van subjectiviteit. Kortom, het beoordelen ten opzichte van de referentie heeft in Nederland nog weinig toegevoegde waarde. Een voordeel is wel, dat wanneer referentiebeelden beschikbaar zijn, doelen scherper kunnen worden gesteld. Ook de optie om de beste Nederlandse situatie als referentie te nemen dient nader te worden onderzocht.
7. Het opstellen van meer stressor-specifieke waarderingssystemen verdient veel meer aandacht. Indien een beek in klasse 1, 2 of 3 valt, worden de waterbeheerders door de Europese Kaderrichtlijn Water verplicht de kwaliteit van de beek te verbeteren naar een kwaliteitsklasse 4 of 5. Om de kwaliteit van

een beek te verbeteren moeten maatregelen worden getroffen. Waterbeheerders zouden er veel baat hebben, om te weten met welke maatregelen zij de grootste toename in kwaliteit zouden kunnen realiseren. In het geval van een stressor-specifiek waarderingssysteem (waar stressoren een eigen effect hebben op gemeenschappen) kan worden bepaald welke stressor een slechte indicatie veroorzaakt, daaruit kan weer worden afgeleid welke maatregelen het meeste effect zouden sorteren. Het hier ontwikkelde systeem bepaalt alleen de algehele kwaliteit van een beek. Hieruit valt niet af te leiden welke maatregelen de voorkeur verdienen bij beekherstel. In integrale benadering waarin waardering en diagnose in één systeem zijn vervlochten verdient de voorkeur. Met een multimetric waarin diagnostische indices zijn opgenomen is een dergelijke aanpak haalbaar.

Tot slot zijn de volgende praktische beperkingen van kracht om incorrect gebruik van het gepresenteerde systeem te voorkomen:

- monsters uit droogvallende of zure beken kunnen niet worden beoordeeld;
- alleen voor- en najaarsmonsters kunnen worden beoordeeld;
- de berekening van de indexwaarden gaat uit van met aantallen per 1.25 m^2.

8.5 Toepassing van de EBEO-systemen op de beken en sloten gegevens

De EBEO-systemen (EBEO=Ecologische BEOordelingsmethoden) zijn in de negentiger jaren in opdracht van STOWA ontwikkeld. Voor de CUWVO-typen stromende wateren en sloten zijn respectievelijk EBEOSWA (Ecologische BEOordelingsmethode voor Stromend Water) en EBEOSLO (Ecologische BEOordelingsmethode voor SLOten) beschikbaar. De methoden maken gebruik van karakteristieken. Iedere karakteristiek geeft de invloed van vooral één (beïnvloedings-)factor weer. In de methode wordt de (relatieve) abundantie van indicatorsoorten als maatstaf voor de respectievelijke karakteristiek gebruikt. De maatlat is de grafische presentatie van iedere karakteristiek. Voor EBEOSWA bevat de maatlat karakteristieken voor saprobie, stroming, trofie, substraat (4 typen) en voedselstrategie (3 typen). Voor EBEOSLO (macrofauna) bevat de maatlat karakteristieken voor brak, zuur, saprobie, permanentie en toxiciteit. De maatlatten geven indicesscores op een procentuele schaal voor de respectievelijke kenmerken weer. Tenslotte zijn de maatlatten in vijf delen opgedeeld. Ieder deel geeft een kwaliteitsniveau aan. Het hoogste kwaliteitsniveau is afgeleid uit de beschikbare gegevens verzameld door de regionale waterbeheerders in de tachtiger jaren.

In bijlage 14 zijn de maatlatscores per beken- en slotencluster in boxplots uitgezet. Daarna is het aantal monsters per maatlat per kwaliteitsniveau per cluster uitgezet. Over het algemeen bevat een bekencluster 3 kwaliteitsniveaus (variërend van 1 tot en met 5). De spreiding van de scores over de maatlat of kewaliteitsniveaus is groot. Een mogelijke oorzaak is gelegen in het beschouwen van slechts één (beïnvloedings-)factor per karakteristiek. Aangezien bijna altijd beïnvloeding zich niet als één maar als een combinatie van factoren aandient leidt dit tot een variatie in scores per cluster. Met andere woorden een beetje meer van de ene factor tezamen met een beetje

minder van de andere of juist omgekeerd leidt tot eenzelfde gemeenschap maar tot verschillende score.
Voor de sloten is de spreiding minder groot voor de maatlatten voor zuur, brak en permanentie (alle vaak type-gebonden kenmerken die zich vooral in extreme waarden als dominant manifesteren) maar is de spreiding vergelijkbaar voor saprobie en toxiciteit. Dit laatste om dezelfde redenen als bij de beken.
Omdat de scores niet gecombineerd worden tot een eindscore is vergelijking per type moeilijk.

8.6 Conclusies

Met het in dit hoofdstuk beschreven waarderingssysteem is het mogelijk een ecologische waarde toe te kennen aan de uitkomsten van de voorspellingsmodules voor beken. Het waarderingssysteem is gebaseerd op cenotypen, waaraan een kwalificatie is toegekend door middel van expert-judgement.
Met het waarderingssysteem blijken tussen de 65-67% van de monsters overeenkomstig de post-classificatie te worden gekwalificeerd. Met de mogelijkheid dat monsters tot twee aangrenzende klassen kunnen behoren (een post-classificatie van cenotypen sluit dit zeker iet uit), waarbij de correcte klassificatie oploopt tot 91-92%, lijkt de waardering in ieder geval valide. De waardering van de AQEM monsters gaf veel slechtere resultaten. Mogelijk is dit een gevolg van de wijze van classificeren, maar dat behoeft zeker nader onderzoek. Het gebruik van het waarderingssysteem dient vooralsnog met in achtneming van bovenstaande aanbevelingen en conclusies te worden gedaan. Temeer daar het waarderingssysteem zonder meer kan worden verbeterd door onder andere de gebruikte buitenlandse indicatiewaarden te vertalen naar de Nederlandse situatie en de typologie achter het beoordelingssysteem te verfijnen.
Tot slot zal het waarderingssysteem in de toekomst zodanig moeten worden aangepast dat beoordeling ten opzichte van de referentie (of ‘ second best’) mogelijk wordt. Beoordeling ten opzichte van de referentie is immers opgelegd vanuit de Europese Kaderrichtlijn Water. De toegevoegde waarde van deze manier van beoordelen is voor in Nederland echter twijfelachtig. Het is een rekenkundig verschil dat noch het referentie-probleem nog de niet-lineariteit in ecologische responsies oplost.

9 Zeldzaamheid als waarderingsmaat

9.1 Inleiding

Zeldzaamheid van macrofauna is een belangrijk aspect in het water- en natuurbeheer. Zeldzame soorten worden door de meeste onderzoekers gedefinieerd als soorten die een lage abundantie hebben en/of een klein verspreidingsareaal (Gaston 1994). Onderzoeken naar aquatische macrofauna zijn meestal uitgevoerd op regionale schaal of voor een stroomgebied. Met zeldzame soorten worden in dat geval soorten bedoeld die of op veel plekken voorkomen maar in lage abundanties of soorten die op slechts enkele locaties voorkomen in lage of hoge abundanties (Gauch 1982).
Een soort kan om verschillende redenen zeldzaam zijn:

- Intrinsieke zeldzaamheid (*afhankelijk van zichzelf*): een soort kan van zichzelf altijd in zeer lage aantallen voorkomen. Soortskenmerken die vaak genoemd worden in relatie tot zeldzaamheid zijn het kolonisatievermogen (dispersie en vestiging) en de lichaamsgrootte. Andere mogelijkheden zijn levensstrategie (K-strategen, soorten die weinig nakomelingen voortbrengen en een lange levensduur hebben lijken vaker zeldzaam te zijn), of voedingswijze (predatoren komen bijvoorbeeld in lagere aantallen voor dan prooidieren). Autecologische kennis van soorten kan een beeld geven van de zeldzaamheid ervan.
- De zeldzaamheid van een soort kan ook gebonden zijn aan de zeldzaamheid van het habitat waarin de soort leeft: Milieu/habitat gerelateerde zeldzaamheid. Het habitat wordt bepaald door de combinatie van alle milieuvariabelen op een bepaalde plek. De geografische grenzen, waar vanuit gegaan wordt bij bepaling van zeldzaamheid, bepalen of een habitat veel of weinig voorkomt en daarmee of een soort wel of niet zeldzaam is.
- Zeldzaamheid kan ook veroorzaakt worden doordat een soort om te overleven gebonden is aan het voorkomen van andere soorten: Interspecifieke zeldzaamheid. Het betreft hier biotische interacties, zoals concurrentie, predator-prooi relaties en parasitisme. Indien een systeem rijper is, treden meer biotische interacties op en komen meer zeldzame soorten voor. Hierover is echter nog weinig bekend.

Het is onmogelijk te voorspellen waarom een soort zeldzaam is. Een aantal belangrijke punten zijn (Gaston 1994):

- Er is weinig bewijs dat zeldzaamheid samenhangt met soortskenmerken;
- Het onderscheiden van de oorzaken van zeldzaamheid van de effecten is moeilijk. Effecten kunnen de oorzaak worden en wat voor de ene soort de oorzaak is, is voor de ander het effect (bijvoorbeeld dispersievermogen);
- De oorzaken voor zeldzaamheid kunnen afhangen van ruimtelijke en temporele schaal. Bijvoorbeeld over een groot gebied hangt het voorkomen van een soort af van de temperatuur, binnen een regio van het aanwezige substraat;
- De lage abundantie en het kleine verspreidingsareaal van veel soorten die nu als zeldzaam worden beschouwd, zijn het resultaat van menselijke activiteiten.

Soorten die zijn aangepast aan menselijke of verstoorde habitats zijn nu algemeen;
- Veel studies naar de oorzaken van zeldzaamheid hebben slechts een aantal van de mogelijke factoren onderzocht.

Het voorkomen van zeldzame soorten in een oppervlaktewater kan een indicatie zijn voor de kwaliteit (natuurwaarde of biodiversiteit) van het betreffende water. In monitoringsprogramma's bijvoorbeeld, kan onderzocht worden of het aantal zeldzame soorten en daarmee de natuurwaarde toeneemt na herstelmaatregelen. Zeldzame soorten kunnen dienen als indicator voor een verbetering van de kwaliteit van een oppervlaktewater of voor een toename aan habitatdiversiteit. Beheer en beoordeling van wateren met behulp van zeldzame soorten is van groot belang. Het meenemen van zeldzame soorten in beoordeling en monitoring geeft de mogelijkheid om sneller verstoring van het milieu te kunnen detecteren en om het type verstoring te kunnen aanduiden (Cao *et al.* 2001). Zeldzame soorten zijn vaak de soorten die het eerste verdwijnen bij verstoring. Ze vormen daarmee indicatoren die snel kunnen aangeven of een bepaalde ingreep effect heeft op een levensgemeenschap of dat een maatregel werkelijk leidt tot verbetering van de situatie. In dit opzicht is het voorspellen van zeldzame soorten bij het uitvoeren van een herstelingreep een goede indicator voor kwaliteitsverbetering.

Doelstelling & leeswijzer

Het voorspellingsmodel voorspelt de kans op het voorkomen van cenotypen aan de hand van milieuvariabelen. Aan deze voorspelling moet een waardering gekoppeld worden. Zeldzame soorten kunnen van belang zijn bij het waarderen van oppervlaktewateren. Er wordt vanuit gegaan dat zeldzame soorten vooral voorkomen als de situatie natuurlijk is. Het is echter van belang om dit aan te tonen voordat zeldzame soorten gebruikt kunnen worden in de voorspelling en waardering van gemeenschappen binnen RISTORI. De doelstelling van dit hoofdstuk is het evalueren van de mogelijkheid om een oppervlaktewater te waarderen op basis van de zeldzaamheid van de soorten, zowel voor sloten als voor beken.
Drie onderzoeksvragen zijn van belang:
1. Zijn er grote verschillen tussen het aantal zeldzame soorten in sloten en beken?
2. Bestaat er een relatie tussen het totale aantal soorten, het aantal zeldzame soorten en de fractie zeldzame soorten in een monster enerzijds en bepaalde milieuvariabelen anderzijds? Zo ja, duiden deze milieuvariabelen op een natuurlijke situatie?
3. Is er een verband tussen het totaal aantal soorten, het aantal zeldzame soorten en de fractie zeldzame soorten enerzijds en de natuurlijkheid van de cenotypen anderzijds?

Vraag 1 zal worden beantwoord in paragraaf 1.2, vraag 2 in paragraaf 1.3, vraag 3 in paragraaf 1.4 Paragraaf 1.5 bevat de conclusies.

9.2 Sloten en beken: uiten de verschillen zich in zeldzaamheid?

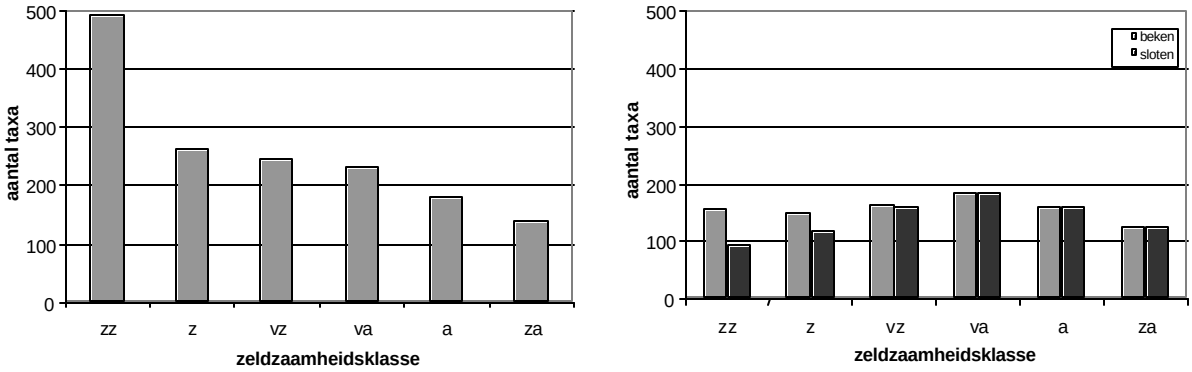
Onlangs is een zeldzaamheidslijst opgesteld voor de Nederlandse macrofauna (Nijboer & Verdonschot (red) 2001). Deze lijst is toegepast op 6127 sloten- en 2242 bekenmonsters. Hiertoe zijn de zeldzaamheidsklassen uit de zeldzaamheidslijst gekoppeld aan de macrofaunagegevens van de beken en sloten. Per klasse is berekend hoeveel taxa hiertoe behoren in de sloten en de beken.

Tabel 9.1 Overzicht van aantallen taxa per zeldzaamheidsklasse die ofwel in sloten ofwel in beken ofwel in beide watertypen voorkomen

watertype	zeldzaamheidsklasse	aantal taxa
beken	zz	105
beken	z	73
beken	vz	28
beken	va	12
beken	a	6
Beken	za	0
sloten	zz	42
sloten	z	39
sloten	vz	25
sloten	va	9
sloten	a	5
Sloten	za	0
beken & sloten	zz	51
beken & sloten	z	78
beken & sloten	vz	134
beken & sloten	va	174
beken & sloten	a	155
beken & sloten	za	127

In totaal zijn er 1063 taxa waarvoor een zeldzaamheidsklasse bekend is in de sloten- en bekengegevens (tabel 9.1). Daarvan komen 719 taxa in zowel sloten als beken voor. Daarnaast zijn er 224 taxa die alleen in beken gevonden zijn en 120 taxa die alleen in sloten voorkomen. Het blijkt dat de soorten die in beide watertypen voorkomen vrijwel altijd zeer algemeen tot vrij zeldzaam zijn en de soorten die slechts in één van beide watertypen voorkomen meestal zeldzaam of zeer zeldzaam. Hieruit kan geconcludeerd worden dat een soort algemener is naarmate deze in meer watertypen kan voorkomen. Sloten en beken zijn zeer verschillend van karakter. De soorten die in beide watertypen voorkomen, betreffen algemene soorten die tolerant zijn voor 'enige' stroming maar daar niet van afhankelijk zijn. In stromende wateren komen daarnaast soorten voor die zich niet alleen hebben aangepast aan het stromende water maar daar van afhankelijk zijn, bijvoorbeeld voor de aanvoer van voedsel of voor hun ademhaling. Deze soorten komen niet in stilstaand water voor. In stilstaande wateren daarentegen komt een groep soorten voor die geen stroming verdraagt (bijvoorbeeld doordat deze soorten zich niet kunnen vasthouden) en daarom in beken ontbreekt. Hierbij moet echter wel in gedachten gehouden worden dat veel van de beken in het gegevensbestand genormaliseerd zijn en daardoor een

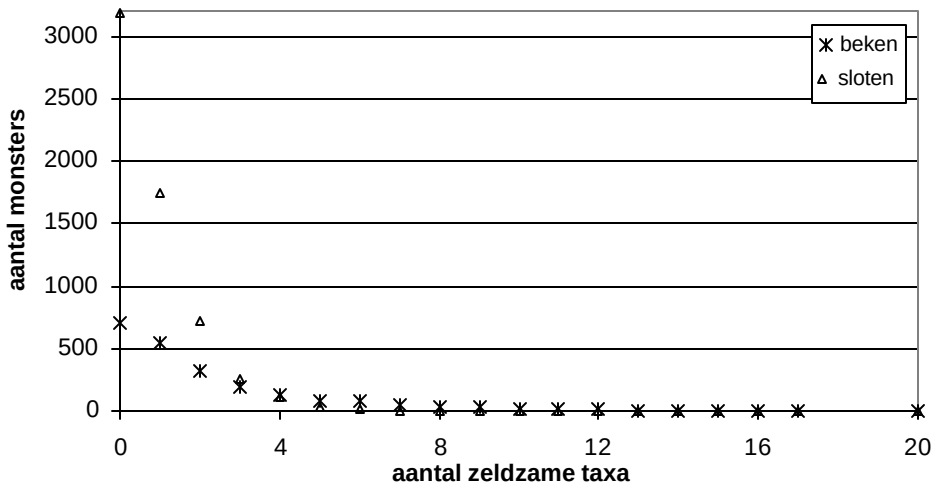
lage stroomsnelheid hebben en ook veel van de sloten niet natuurlijk zijn. Dit kan een vertekend beeld geven.
Als naar het algemene patroon van de verdeling over de zeldzaamheidsklassen in sloten en beken wordt gekeken (figuur 9.1) dan valt eveneens op dat er in beken meer soorten voorkomen dan in sloten en dat dit zich vooral uit in een hoger aantal zeldzame soorten. Het aantal algemenere soorten komt vrijwel overeen, en dit komt doordat het voor het grootste deel precies dezelfde soorten betreft. Het totale aantal zeldzame taxa in sloten en bekengegevens is veel lager dan het totale aantal zeldzame taxa in Nederland volgens de zeldzaamheidslijst (figuur 9.1), terwijl vrijwel de meeste algemene taxa in sloten en beken gevangen zijn.



Figuur 9.1 Verdeling van alle taxa van Nederland (die een zeldzaamheidsaanduiding hebben) (links) en de taxa in de sloten en beken (rechts) over de zeldzaamheidsklassen

Dit kan verklaard worden doordat enerzijds de zeldzame taxa in bijzondere milieus voorkomen die niet bemonsterd zijn, zoals vennen en moerassen. Ook wateren in natuurgebieden zijn sterk onderbemonsterd. Een andere oorzaak is dat zeldzame taxa vaak ook in lage aantallen voorkomen. Hierdoor is de kans dat een zeldzame soort die ergens wel voorkomt ook daadwerkelijk gevangen wordt klein in vergelijking tot de kans dat een algemene soort gevangen wordt.

Ondanks een optimale efficiënte bemonstering is het nooit zeker dat alle aanwezige soorten gevangen wordt. Een monster is altijd een steekproef uit een water, slechts een kwart van de soorten die aanwezig zijn wordt gevangen (Verdonschot 1990). De zeldzame soorten zijn dan de eersten die ontbreken omdat de kans nu eenmaal kleiner is om een soort met een lage abundantie te vangen dan een soort met een hoge abundantie. Uit de analyses van de sloten- en bekengegevens blijkt dat zeldzame taxa niet veel zijn gevangen. In de meeste monsters zijn dan ook geen of slechts één of twee zeldzame taxa aangetroffen (figuur 9.2).

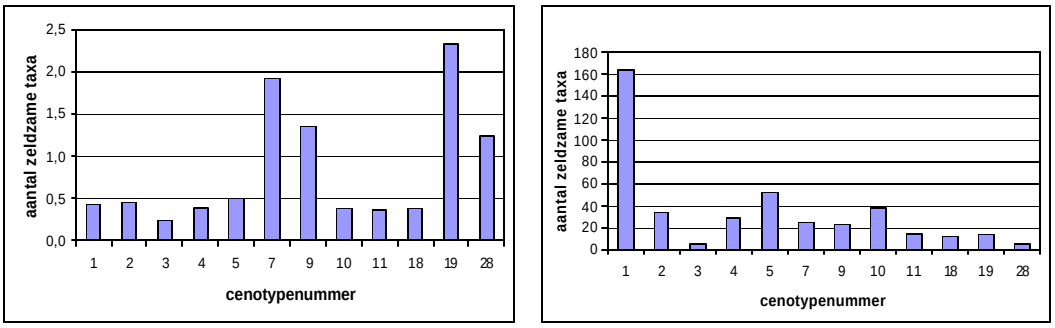


Figuur 9.2 Aantal monsters waarin een bepaald aantal zeldzame taxa is aangetroffen

9.3 Zeldzame soorten in de cenotypen

9.3.1 Sloten

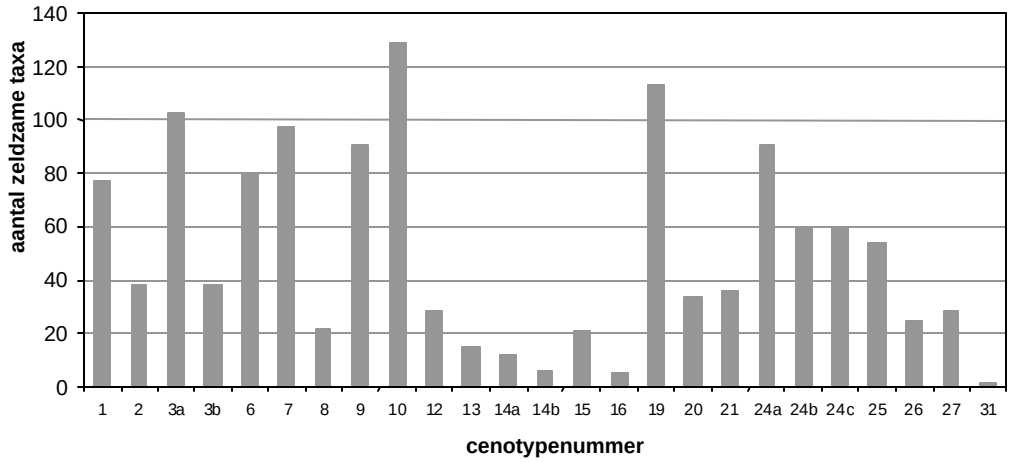
Het aantal zeldzame soorten verschilt per cenotype (figuur 9.3). In cenotype 1 is het aantal met 164 zeldzame soorten verreweg het hoogste, gevolgd door cenotype 5 en cenotype 10 met 52 respectievelijk 39 zeldzame soorten. Dit zijn ook de cenotypen met het grootste aantal monsters. Het hoge aantal zeldzame soorten in deze cenotypen wordt dan ook zeer waarschijnlijk veroorzaakt doordat de kans op het vangen van zeldzame soorten toeneemt als meer monsters genomen zijn in verschillende wateren binnen een cenotype. Daarom is vervolgens het gemiddelde aantal zeldzame soorten per monster voor ieder cenotype weergegeven in de tweede grafiek van figuur 9.3. Dit levert een heel ander patroon op. Het aantal zeldzame soorten per monster is in alle cenotypen laag. Voor de meeste cenotypen wordt minder dan 1 zeldzame soort gevonden per twee monsters. In de cenotypen 7, 9, 19 en 28 blijkt het aantal zeldzame soorten per monster het hoogste te zijn (meer dan 1 zeldzame soort per monster). Dit hangt waarschijnlijk samen met de natuurlijkheid van de sloten in deze cenotypen. Het zijn vrijwel allemaal sloten met een natuurfunctie. Ook is er sprake van bijzondere omstandigheden zoals, droogval in de cenotypen 7 en 28, duinwateren in cenotype 9 en natuurlijke laagveensloten in cenotype 19. In cenotype 3 komen de minste zeldzame soorten voor. Dit is waarschijnlijk te verklaren door de toxische beïnvloeding van de sloten in dit cenotype. In de overige cenotypen is de kans om één zeldzame soort te vangen in een monster minder dan 50 %.



Figuur 9.3 Aantal zeldzame soorten per cenotype (links het totale aantal zeldzame taxa in het cenotype, rechts het gemiddeld aantal zeldzame taxa per monster)

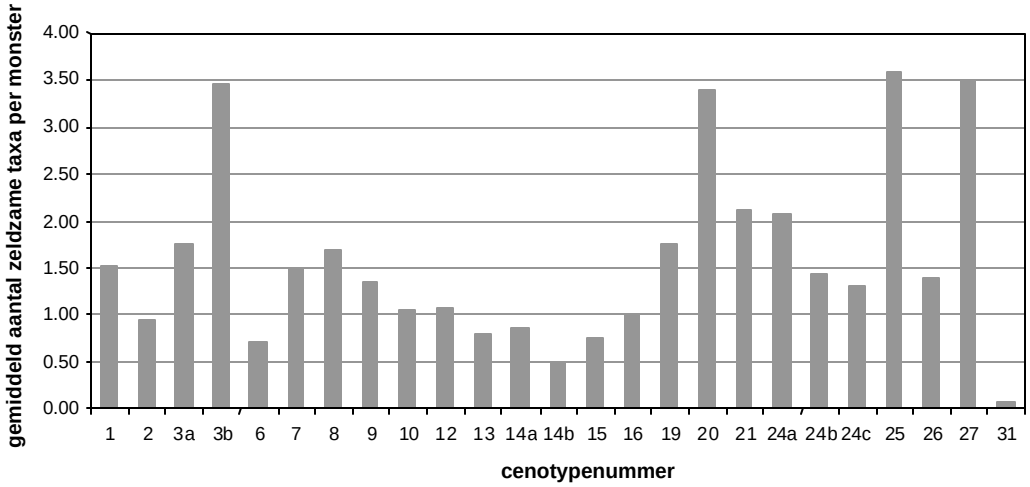
9.3.2 Beken

Het aantal zeldzame soorten in de beken verschilt per cenotype (figuur 9.4). In cenotype 10 is het aantal met 129 zeldzame soorten verreweg het hoogste, gevolgd door de cenotypen 19 en 3a met 113 respectievelijk 103 zeldzame soorten. Cenotype 10 is ook het cenotype met het grootste aantal monsters. Een groot aantal monsters geldt niet voor de cenotypen 19 en 3a. Het hoge aantal zeldzame soorten in cenotype 10 is dan ook zeer waarschijnlijk veroorzaakt doordat de kans op het vangen van zeldzame soorten toeneemt als meer monsters genomen zijn. Daarom is ook het gemiddelde aantal zeldzame soorten per monster voor ieder cenotype weergegeven (figuur 9.5).
Dit levert een heel ander patroon op. Het aantal zeldzame soorten per monster is in alle cenotypen laag en schommelt tussen 0.5 tot 3.5 soorten. Het maximum aantal zeldzame soorten per monster is hoger dan in de sloten. In de cenotypen 25, 27, 3b en 20 blijkt het aantal zeldzame soorten per monster het hoogste te zijn (circa 3.5 zeldzame soort per monster). Dit hangt waarschijnlijk samen met de bijzonderheid van de beken in deze cenotypen. Drie van de vier cenotypen betreffen (zeer) snelstromende typen (afwijkend milieu) terwijl het vierde zwak zure, oligo-β-mesosaprobe, langzaam stromende beken (natuurlijke toestand) betreft.



Figuur 9.4 Het totaal aantal zeldzame taxa per bekencenotype

In cenotype 31 komen de minste zeldzame soorten voor. Dit is waarschijnlijk te verklaren door de organische en morfologische beïnvloeding van de beken in dit cenotype. De cenotypen 14a en 13 (beide organisch belast), 14b, 6 en 15 (alle drie organisch en morfologisch belast), 2 (zwak zuur) en 16 (gering aantal monsters) bevatten tot gemiddeld 1 zeldzame soort per monster. In het algemeen is er een zwakke relatie tussen kwaliteit van de beken in een cenotype en het aantal zeldzame soorten. De grens schommelt rond de gemiddeld één soort per monster, echter met dit getal is voorzichtigheid geboden.



Figuur 9.5 Het gemiddeld aantal zeldzame taxa per monster in de bekencenotypen

9.4 Zeldzame soorten in relatie tot milieuvariabelen

9.4.1 Inleiding

De abundantie en ruimtelijke verspreiding van alle soorten wordt beperkt door milieuvariabelen. Er zijn veel voorbeelden van zeldzame soorten die beperkt worden doordat een of meer milieuvariabelen de abundantie tot lage aantallen beperken of het verspreidingsareaal klein maken. De vraag is of er een specifieke variabele of een combinatie van variabelen is die ervoor zorgt dat soorten zeldzaam zijn. Het is echter moeilijk om dit aan te tonen. Causale verbanden moeten worden aangetoond met behulp van experimenten met soorten. Om een algemene uitspraak te kunnen doen is het van belang dergelijk onderzoek met veel zeldzame soorten uit te voeren. Vaak is gebleken dat de zeldzaamheid van veel soorten niet het resultaat is van een enkele factor. Verder lijkt het onwaarschijnlijk dat het relatieve belang van milieuvariabelen in het veroorzaken van zeldzaamheid van een soort in een levensgemeenschap kan worden veralgemeniseerd naar andere levensgemeenschappen.

9.4.2 Methode

Voor elke beek respectievelijk sloot is het totale aantal soorten waargenomen, alsmede de uitsplitsing naar aantal zeer zeldzame soorten (zz), aantal zeldzame (z), aantal vrij zeldzame (vz), aantal vrij algemene (va), aantal algemene (a) en aantal zeer algemene soorten (za). Er is onderzocht hoe het aantal zeldzame soorten en de fractie zeldzame soorten samenhangen met de waargenomen milieuvariabelen, en wel apart voor beken en voor sloten. De gebruikte milieuvariabelen staan respectievelijk beschreven in bijlagen 16 en 17. De ontbrekende milieuvariabelen zijn ingevuld volgens de methodiek beschreven in paragraaf 4.2 en 5.2 en de gegevens zijn getransformeerd (paragraaf 4.2 en 5.2). De seizoensvariabelen voorjaar en winter in de bekendataset zijn niet in deze analyse meegenomen. Voor de sloten is de variabele grondgebruik buiten beschouwing gelaten (gakker, gnatuur, gstedelijk, gtuin, gweii). De bekendata hebben dan in totaal 19 milieuvariabelen en de slotendata 20 variabelen.
Er zijn analyses uitgevoerd op de volgende twee variabelen:

1. totaal aantal zeldzame soorten : zz + z + vz
2. fractie zeldzame soorten : (zz + z + vz) / (zz + z + vz + va + a + za)

De eerste analyse is, zoals gebruikelijk, uitgevoerd met Poisson regressie met overdispersie. De tweede analyse is uitgevoerd met logistische regressie, eveneens met overdispersie. Om te compenseren voor het inschatten van ontbrekende waarnemingen zijn in alle regressies de waarnemingen gewogen met het aantal niet ontbrekende waarnemingen. Wateren zonder ontbrekende milieuvariabelen doen dan volledig mee in de regressie, terwijl wateren waarvoor alle milieuvariabelen ontbreken, niet meedoen in de regressie. Allereerst zijn univariate analyses uitgevoerd, waarbij het effect van elke milieuvariabele apart is geschat, met een grafische weergave van de geschatte relatie. Deze univariate analyses geven ook zogenaamde marginale toetsen, waarbij getoetst wordt of een individuele term aan het model toegevoegd moet worden. Vervolgens zijn zogenaamde conditionele toetsen berekend voor elke milieuvariabele. Hierbij wordt getoetst of een variabele uit het model met alle milieuvariabelen kan worden verwijderd. Milieuvariabelen met

een significante marginale én conditionele toets zijn in het algemeen de belangrijkste variabelen. Tenslotte is een uitgebreide modelselectie met alle milieuvariabelen uitgevoerd, waarbij gebruik is gemaakt van Genstat procedure SELECT. Deze past alle mogelijke modellen aan voor maximaal 16 predictoren en geeft aan welke van de aangepaste modellen de beste zijn. Aangezien er in totaal 19 of 20 predictoren zijn én modelselectie met meer dan 12 predictoren rekenintensief is, is SELECT iteratief toegepast. Voor een eerste selectie zijn de predictoren ingedeeld in 2 groepen (1…10, 11…20) en is een aparte modelselectie uitgevoerd voor de twee groepen. Met de beste predictoren uit de twee groepen is een nieuwe modelselectie uitgevoerd en daar zijn weer de beste predictoren uit geselecteerd. Met deze beste predictoren vast in het model worden de resterende predictoren weer onderverdeeld in 2 groepen en begint het spel opnieuw. Uiteindelijk blijven zo een gering aantal kandidaat milieuvariabelen over en met deze is een finale modelselectie uitgevoerd. Het "beste" model dat vervolgens geselecteerd is, is het model met de kleinste mean deviance en heeft milieuvariabelen met p-waarden die allen kleiner zijn dan 0.01. In alle gevallen zijn er alternatieve modellen met een vergelijkbare fit, deze zijn in de betreffende bijlagen gegeven. Alle modellen zijn lineair in de predictoren, dat wil zeggen dat noch interacties noch kwadratische termen zijn aangepast.
Een analyse die (nog) niet is uitgevoerd beschrijft de verdeling van de soorten over de 6 categorieën in relatie tot de milieuvariabelen. Een dergelijke analyse is vergelijkbaar met de voorspelling van de cenotypen uit de milieuvariabelen en wordt ook uitgevoerd met multinomiale logistische regressie. Daarbij moet wel bedacht worden dat het aantal zeer zeldzame en zeldzame soorten gering is, en dat deze twee categorieën in een dergelijke analyse waarschijnlijk beter samengevoegd kunnen worden.

9.4.3 Beken

Aantal zeldzame soorten

Voor het aantal zeldzame soorten zijn nagenoeg alle marginale (univariate) toetsen zeer significant (tabel 9.2). De variabelen met de hoogste regressiecoëfficiënt zijn: meander, dwarsnat, grgenatu (positief verband met het aantal zeldzame soorten) en stuw en clmed (negatief verband met het aantal zeldzame soorten). De conditionele toetsen zijn slechts significant zijn voor dwarsnat, diepte, phmed, clmed, egvmed en no310. Meandering en stuw zijn gecorreleerd met dwarsprofiel natuurlijk en komen daardoor niet tot uiting in de conditionele toetsen. De mean deviance laat zien dat er overdispersie is. De grafieken van de univariate analyses zijn opgenomen in bijlage 18.

Tabel 9.2 De resultaten van de univariate analyse met de milieuvariabelen en het totale aantal zeldzame soorten (beta=geschatte regressiecoëfficiënt, sebeta=standaardafwijking, tbeta=t-waarde, meandev=de mean deviance van het model), inclusief de marginale (p-marg) en conditionele (p-cond) toetsen

nr	x	p-marg	p-cond	beta	sebeta	tbeta	meandev
1	meander	0.000	0.086	1.103	0.089	12.41	2.18
2	dwarsnat	0.000	0.000	1.163	0.094	12.41	2.15
3	grgenatu	0.000	0.784	0.709	0.093	7.62	2.55
4	perman	0.026	0.214	-0.275	0.149	-1.84	2.80

5	stuw	0.000	0.619	-0.787	0.134	-5.87	2.62
6	schaduw	0.000	0.245	0.153	0.019	8.25	2.51
7	subslib	0.000	0.904	-0.142	0.022	-6.59	2.59
8	diepte	0.000	0.019	-0.230	0.039	-5.94	2.65
9	strmsn	0.000	0.307	0.232	0.043	5.47	2.65
10	breedte	0.000	0.240	-0.294	0.042	-7.03	2.59
11	subzand	0.266	0.801	0.019	0.022	0.89	2.81
12	phmed	0.000	0.000	0.304	0.083	3.65	2.75
13	clmed	0.000	0.006	-0.651	0.087	-7.52	2.56
14	egvmed	0.000	0.000	-0.378	0.064	-5.90	2.67
15	nh490	0.000	0.504	-0.289	0.031	-9.24	2.44
16	nkjel90	0.000	0.372	-0.433	0.054	-8.06	2.52
17	o2geh10	0.000	0.539	0.135	0.017	7.96	2.54
18	ptot90	0.000	0.603	-0.237	0.041	-5.75	2.66
19	no310	0.000	0.000	0.226	0.035	6.48	2.60

De uiteindelijke modelselectie is opgenomen in bijlage 19; hieruit blijkt dat met name de variabelen dwarsnat, phmed, clmed, egvmed en no310 van belang zijn. De bijbehorende geschatte regressiecoëfficiënten zijn opgenomen in tabel 9.3.
Het conditionele effect van de geselecteerde milieuvariabelen is weergegeven in bijlage 20. Een grafiek van residuen tegen gefitte waarden geeft aan dat de Poisson veronderstelling met overdispersie adequaat is.

Tabel 9.3 Geschatte regressiecoëfficiënten (estimate), standaardafwijking (s.e.) en t-waarde (t) voor het geselecteerde model voor het totaal aantal zeldzame soorten in beken

	estimate	**s.e.**	**t(557)**
Constant	0.532	0.565	0.94
dwarsnat	0.8253	0.0928	8.89
phmed	0.4344	0.0756	5.75
clmed	-0.4093	0.0837	-4.89
egvmed	-0.3065	0.0630	-4.87
no310	0.1793	0.0295	6.08

Fractie zeldzame soorten

De resultaten van de univariate analyse zijn opgenomen in tabel 9.4. De grafieken van de univariate analyses zijn opgenomen in bijlage 21. Dezelfde variabelen als bij het aantal zeldzame soorten heeft een hoge regressiecoëfficiënt maar nu zijn er meer variabelen die een hoge waarde hebben, zoals diepte en breedte. De conditionele toetsen geven aan dat met name dwarsnat, breedte, phmed, clmed, egvmed en no310 belangrijke milieuvariabelen zijn, en dit komt goed overeen met de milieuvariabelen die gevonden zijn bij het aantal zeldzame soorten. Het verschil is dat diepte vervangen is door breedte. Deze variabelen zijn echter met elkaar gecorreleerd. De mean deviance laat zien dat er sprake is van overdispersie ten opzichte van de binomiale verdeling.

Tabel 9.4 De resultaten van de univariate analyse met de milieuvariabelen en de fractie zeldzame soorten (beta=geschatte regressiecoëfficiënt, sebeta=standaardafwijking, tbeta=t-waarde, meandev=de mean deviance van het model), inclusief de marginale (p-marg) en conditionele (p-cond) toetsen

nr	**x**	**p-marg**	**p-cond**	**beta**	**sebeta**	**tbeta**	**meandev**
1	meander	0.000	0.280	1.435	0.100	14.29	2.57
2	dwarsnat	0.000	0.000	1.536	0.104	14.77	2.47
3	grgenatu	0.000	0.600	1.111	0.105	10.59	2.96

4	perman	0.000	0.754	-0.830	0.181	-4.60	3.46
5	stuw	0.000	0.717	-1.361	0.146	-9.29	2.98
6	schaduw	0.000	0.072	0.269	0.022	12.02	2.80
7	subslib	0.000	0.332	-0.247	0.025	-9.91	2.95
8	diepte	0.000	0.727	-0.604	0.042	-14.48	2.60
9	strmsn	0.000	0.206	0.330	0.049	6.80	3.25
10	breedte	0.000	0.000	-0.714	0.047	-15.26	2.46
11	subzand	0.953	0.834	0.001	0.025	0.04	3.57
12	phmed	0.554	0.000	0.039	0.102	0.38	3.57
13	clmed	0.000	0.000	-1.128	0.108	-10.47	2.98
14	egvmed	0.000	0.009	-0.429	0.072	-5.93	3.38
15	nh490	0.000	0.116	-0.541	0.042	-12.97	2.69
16	nkjel90	0.000	0.085	-0.736	0.075	-9.80	3.02
17	o2geh10	0.000	0.879	0.174	0.019	9.02	3.13
18	ptot90	0.000	0.259	-0.293	0.051	-5.73	3.37
19	no310	0.000	0.000	0.346	0.042	8.17	3.13

Na een uitgebreide modelselectie komt daar nog de predictor nh490 bij. De laatste modelselectie is opgenomen in bijlage 22. Het geselecteerde model bevat dezelfde variabelen als voor het aantal zeldzame soorten, aangevuld met breedte en nh490 (tabel 9.5). De geschatte parameters van het geselecteerde model zijn opgenomen in onderstaande tabel en het conditionele effect van de geselecteerde milieuvariabelen is weergegeven in bijlage 23. Een grafiek van residuen tegen gefitte waarden geeft aan dat de Bionimale verdelingsveronderstelling met overdispersie adequaat is.

Tabel 9.5 Geschatte regressiecoëfficiënten (estimate), standaardafwijking (s.e.) en t-waarde (t) voor het geselecteerde model voor de fractie zeldzame soorten in beken

	estimate	**s.e.**	**t(555)**
Constant	-1.778	0.576	-3.09
dwarsnat	0.7981	0.0923	8.65
breedte	-0.3583	0.0427	-8.38
phmed	0.3790	0.0790	4.80
clmed	-0.6673	0.0922	-7.24
egvmed	-0.2179	0.0620	-3.52
nh490	-0.1512	0.0360	-4.20
no310	0.2141	0.0284	7.54

9.4.4 Sloten

Aantal zeldzame soorten

De resultaten van de univariate analyse zijn opgenomen in tabel 9.6. De grafieken van de univariate analyses zijn opgenomen in bijlage 24. De negatief gecorreleerde variabelen met de grootste regressiecoëfficiënt zijn: klei, egv, totaaln en inlaat. Variabelen met een positief verband met het aantal zeldzame soorten zijn: fnatuur, kwel, veen, zand en droogval. De conditionele toetsen geven aan dat met name diepte en fnatuur van belang zijn. Dit betekent dat de andere positief gecorreleerde variabelen waarschijnlijk correleren met fnatuur.

Tabel 9.6De resultaten van de univariate analyse met de milieuvariabelen en het totale aantal zeldzame soorten (beta=geschatte regressiecoëfficiënt, sebeta=standaardafwijking, tbeta=t-waarde, meandev=de mean deviance van het model), inclusief de marginale (p-marg) en conditionele (p-cond) toetsen

nr	x	p-marg	p-cond	beta	sebeta	tbeta	meandev
1	breedte	0.000	0.017	-0.185	0.055	-3.36	1.91
2	diepte	0.578	0.001	0.051	0.107	0.48	1.95
3	chloride	0.000	0.629	-0.278	0.053	-5.20	1.82
4	egv	0.000	0.381	-0.405	0.071	-5.72	1.81
5	ammonium	0.000	0.112	-0.306	0.046	-6.69	1.77
6	nitraat	0.000	0.581	-0.167	0.031	-5.45	1.83
7	totaaln	0.000	0.243	-0.418	0.057	-7.39	1.78
8	ph	0.000	0.702	-0.351	0.102	-3.44	1.90
9	totaalp	0.000	0.921	-0.375	0.065	-5.81	1.80
10	vdrijf	0.097	0.028	-0.043	0.031	-1.41	1.94
11	vflab	0.409	0.921	0.029	0.040	0.72	1.95
12	vemers	0.068	0.137	0.053	0.033	1.60	1.94
13	vsubmers	0.000	0.125	0.098	0.029	3.33	1.90
14	veen	0.000	0.037	0.705	0.136	5.18	1.84
15	klei	0.000	0.790	-0.602	0.150	-4.02	1.87
16	zand	0.000	0.077	0.443	0.138	3.22	1.90
17	droogval	0.221	0.773	0.397	0.355	1.12	1.95
18	inlaat	0.000	0.220	-0.457	0.148	-3.09	1.91
19	fnatuur	0.000	0.000	1.176	0.127	9.26	1.63
20	kwel	0.000	0.158	0.614	0.137	4.48	1.86

De uiteindelijke modelselectie is opgenomen in bijlage 25. Het beste model heeft slechts predictoren totaaln, klei en fnatuur. Daarnaast is er een groot aantal termen dat significant is met p-waarden tussen 0.01 en 0.05. De geschatte parameters van het geselecteerde model zijn opgenomen in tabel 9.7 en het conditionele effect van de geselecteerde milieuvariabelen is weergegeven in bijlage 26. Hierin zijn de 4 grafieken allen een functie van totaaln, maar achtereenvolgens voor (klei, fnatuur) = (0,0), (0,1), (1,0), (1,1). Een grafiek van residuen tegen gefitte waarden geeft aan dat de Poisson veronderstelling met overdispersie adequaat is.

Tabel 9.7Geschatte regressiecoëfficiënten (estimate), standaardafwijking (s.e.) en t-waarde (t) voor het geselecteerde model voor het totale aantal zeldzame soorten in sloten

	estimate	s.e.	t(404)
Constant	0.336	0.134	2.50
totaaln	-0.3608	0.0660	-5.47
klei	-0.368	0.139	-2.64
fnatuur	0.926	0.129	7.18

Fractie zeldzame soorten

De resultaten van de univariate analyse zijn opgenomen in tabel 9.8. De grafieken van de univariate analyses zijn opgenomen in bijlage 27. In tegenstelling tot de beken is voor de sloten bij het gebruik van de fractie zeldzame soorten het aantal variabelen met een hoge regressiecoëfficiënt afgenomen. De belangrijkste variabelen zijn: fnatuur, kwel, veen, zand, droogval, klei, inlaat, totaaln en pH. Het zijn dus nog steeds dezelfde variabelen die van belang zijn. De conditionele toetsen geven aan dat met name fnatuur van belang is. Daarnaast is er een groot aantal milieuvariabelen met p-waarden kleiner dan 0.05, maar groter dan 0.008.

Tabel 9.8De resultaten van de univariate analyse met de milieuvariabelen en de fractie zeldzame soorten (beta=geschatte regressiecoëfficiënt, sebeta=standaardafwijking, tbeta=t-waarde, meandev=de mean deviance van het model), inclusief de marginale (p-marg) en conditionele (p-cond) toetsen

nr	x	p-marg	p-cond	beta	sebeta	tbeta	meandev
1	breedte	0.008	0.090	-0.134	0.053	-2.51	1.78
2	diepte	0.760	0.015	-0.028	0.103	-0.27	1.80
3	chloride	0.002	0.081	-0.173	0.061	-2.81	1.77
4	egv	0.006	0.297	-0.187	0.076	-2.46	1.77
5	ammonium	0.000	0.066	-0.208	0.047	-4.45	1.72
6	nitraat	0.000	0.586	-0.105	0.032	-3.29	1.76
7	totaaln	0.000	0.260	-0.309	0.067	-4.62	1.72
8	ph	0.000	0.090	-0.375	0.107	-3.51	1.75
9	totaalp	0.000	0.479	-0.238	0.064	-3.71	1.74
10	vdrijf	0.002	0.011	-0.089	0.032	-2.76	1.76
11	vflab	0.207	0.270	-0.045	0.041	-1.11	1.79
12	vemers	0.019	0.188	0.072	0.034	2.14	1.78
13	vsubmers	0.124	0.435	0.042	0.030	1.38	1.79
14	veen	0.000	0.046	0.503	0.135	3.74	1.74
15	klei	0.001	0.515	-0.424	0.146	-2.90	1.76
16	zand	0.000	0.010	0.450	0.133	3.38	1.75
17	droogval	0.068	0.319	0.615	0.348	1.77	1.79
18	inlaat	0.000	0.056	-0.648	0.142	-4.56	1.71
19	fnatuur	0.000	0.000	1.067	0.126	8.49	1.54
20	kwel	0.000	0.008	0.592	0.134	4.43	1.72

De modelselectie is opgenomen in bijlage 28. Het beste model heeft de 4 predictoren totaaln, vdrijf, fnatuur en kwel (tabel 9.9). Hiervan komen totaaln en fnatuur ook voor in het model voor het aantal zeldzame soorten. De geschatte parameters van het geselecteerde model zijn opgenomen in onderstaande tabel en het conditionele effect van de geselecteerde milieuvariabelen is weergegeven in bijlage 29. Hierin zijn de bovenste 4 grafieken allen een functie van totaaln, maar achtereenvolgens voor (fnatuur, kwel) = (0,0), (0,1), (1,0), (1,1). De onderste 4 grafieken zijn voor vdrijf. Een grafiek van residuen tegen gefitte waarden geeft aan dat de Bionimale verdelingsveronderstelling met overdispersie adequaat is.

Tabel 9.9 Geschatte regressiecoëfficiënten (estimate), standaardafwijking (s.e.) en t-waarde (t) voor het geselecteerde model voor de fractie zeldzame soorten in sloten

	estimate	s.e.	t(403)
Constant	-4.039	0.166	-24.28
totaaln	-0.2046	0.0766	-2.67
vdrijf	-0.0930	0.0304	-3.05
fnatuur	0.896	0.130	6.92
kwel	0.363	0.128	2.84

9.5 Conclusies

9.5.1 Gebruik van zeldzaamheid voor natuurwaardering

Zeldzame soorten kunnen goed gebruikt worden voor de waardering van de natuur van wateren. In beken zal het gebruik van zeldzame soorten gemakkelijker zijn omdat er meer zeldzame soorten voorkomen dan in sloten maar ook in sloten is er een redelijk aantal aanwezig. Het gebruik van algemene soorten voor typegerichte beoordeling zoals dat binnen de Kaderrichtlijn water gewenst is, is niet geschikt. Het is namelijk gebleken dat deze soorten niet specifiek zijn voor een watertype en dezelfde soorten zowel in sloten als in beken voorkomen. Vanaf de klasse vrij zeldzaam is er al een redelijk aantal soorten dat specifiek is voor sloten dan wel beken. Er zijn ook zeldzame soorten die in beide watertypen voorkomen. Dit zijn waarschijnlijk soorten die niet zeldzaam zijn vanwege hun gebondenheid aan een specifiek watertype of milieu-omstandigheid maar soorten die of van nature altijd in lage aantallen voorkomen of soorten die zeer kritisch zijn wat betreft waterkwaliteit en daar zowel in sloten als in beken door achteruit gegaan zijn.
Als naar de verdeling van de zeldzame soorten over de cenotypen gekeken wordt, is gebleken dat het totale aantal zeldzame soorten per cenotype niet zoveel zegt, omdat dit sterk afhankelijk is van het aantal monsters in het cenotype. Het beschouwen van het gemiddeld aantal zeldzame soorten per monster in een cenotype geeft een veel beter beeld. Ondanks dat deze aantallen laag zijn en dicht bij elkaar liggen komen hierin wel verschillen naar voren. Deze verschillen zijn voor de sloten duidelijk te relateren aan natuurlijke omstandigheden. Ook in beken is dit het geval maar hier geldt dat een extreem hoog aantal zeldzame soorten vaak ook gerelateerd is aan bijzondere typologische omstandigheden zoals een hoge stroomsnelheid. Natuurlijke beken zonder verdere bijzondere omstandigheden hebben een gemiddeld aantal zeldzame soorten tussen 1 en 2.5. Verstoorde beken (organisch belast of morfologisch aangetast) hebben duidelijk een lager aantal zeldzame soorten (minder dan 1 soort per monster).
In dit onderzoek is slechts gekeken naar het aantal zeldzame soorten per cenotype. Hierbij zijn de drie klassen vrij zeldzaam, zeldzaam en zeer zeldzaam samengevoegd. Dit is gebeurd op basis van de resultaten in paragraaf 1.2 waaruit bleek dat binnen deze klassen verschillen zichtbaar zijn tussen watertypen. Het samenvoegen van de klassen levert als voordeel dat de aantallen groter worden en de verschillen daardoor duidelijker naar voren komen. Dit is nodig, omdat het gemiddeld aantal zeldzame soorten per monster laag ligt. Een nadeel is echter dat geen specifieke waarde gehecht kan worden aan de mate van zeldzaamheid van de soorten. Een analyse waarin alle klassen afzonderlijk gewogen worden geeft meer detail. Het meenemen van de algemene klassen lijkt niet echt zinvol, aangezien deze soorten toch overal kunnen voorkomen.
De regressie-analyses hebben eveneens aangetoond dat er een verband is tussen zowel het aantal als de fractie zeldzame soorten in een monster en de milieuvariabelen die natuurlijkheid representeren. Voor beken hebben de variabelen natuurlijk dwarsprofiel, grondgebruik natuur en meandering een positieve relatie met het aantal en de fractie zeldzame soorten. Negatieve variabelen zijn stuwing, een hoog chloridegehalte, een grote diepte, hoge pH en hoog EGV. Deze variabelen zijn

met elkaar gecorreleerd. Voor voorspelling van het aantal zeldzame soorten is het opnemen van de variabelen dwarsprofiel natuurlijk (+), pH (-), chloridegehalte (-), EGV (-) en nitraat (-) voldoende. Voor het voorspellen van de fractie zeldzame soorten zouden daar nog bij komen de variabelen breedte (-) en ammonium (-).
Voor sloten zijn de belangrijkste variabelen: klei (-), EGV (-), totaal stikstof (-), waterinlaat (-), functie natuur (+), kwel (+), veen (+), zand (+) en droogval (+). Hieruit blijkt dat ook in sloten duidelijk de verstoringsfactoren en de factoren die samenhangen met natuurlijkheid naar voren komen als variabelen die het aantal en de fractie zeldzame soorten respectievelijk negatief en positief beïnvloeden.
Met de variabelen totaal stikstof (-), klei (-) en functie natuur (+) kan het aantal zeldzame soorten voorspeld worden. Voor de fractie zeldzame soorten zijn dit de variabelen drijvende vegetatie (-), functie natuur (+), kwel (+) en totaal stikstof (-).
Voor zowel de sloten als de beken geldt dat voor voorspelling van het aantal zeldzame soorten minder variabelen nodig zijn dan voor voorspelling van de fractie zeldzame soorten. Dit betekent dat het aantal zeldzame soorten een betere maat is om te gebruiken.
Uit zowel de analyse van de cenotypen van sloten en beken als de regressie-analyse komt naar voren dat er een relatie is tussen factoren die in relatie staan tot de mate van natuurlijkheid en het aantal zeldzame soorten. Hieruit kan worden geconcludeerd dat het gebruik van zeldzame soorten waardevol is voor het beoordelen van de natuurwaarde van sloten en beken.

9.5.2 Bemonstering

Zeldzame soorten komen vaak voor in lage abundanties. Uit de analyses is gebleken dat het aantal zeldzame soorten dat in beken en sloten gevonden is, laag is. Voor het gebruik van zeldzame soorten in beoordeling is een goede bemonstering van een sloot of beek dan ook onontbeerlijk. Het aantal zeldzame soorten dat gevangen wordt is afhankelijk van:

- Het aantal bemonsterde (micro)habitats. Als alle habitats van een sloot of beek worden bemonsterd wordt een beter beeld verkregen van de totale soortensamenstelling en is de kans op het vinden van zeldzame soorten groter;
- De grootte van het monster;
- De spreiding van de monsters over het seizoen. Gedurende bepaalde perioden in het jaar kunnen soorten in diapauze zijn en niet gevonden worden, soorten kunnen te klein zijn om te worden gedetermineerd of soorten kunnen een terrestrisch levensstadium hebben en niet in het water voorkomen.
- De bemonsteringsmethode, bijvoorbeeld de maaswijdte van het net;
- De tijd die genomen wordt om een monster uit te zoeken;
- Het niveau waarop de dieren gedetermineerd worden.

Optimaal is het bemonsteren van alle microhabitats in een sloot of beek in tenminste twee seizoenen (voorjaar en herfst). Hoe meer monsters van een water hoe hoger de kans dat er meer soorten gevonden worden. Door de bemonstering te spreiden over twee seizoenen kan het aantal soorten toenemen doordat niet alle soorten tegelijkertijd aanwezig zijn. In de sloten en bekendataset is ervoor gekozen om niet de seizoenen samen te voegen omdat voor de meeste monsters geen twee seizoenen

aanwezig waren. Niet alle waterbeheerders bemonsteren een water twee keer per jaar. Vooral in de wat oudere gegevens is dit het geval. Tegenwoordig gebeurt het wel meer maar nog steeds niet ieder waterschap doet dit. Aangezien het niet mogelijk is om voor een deel van de dataset twee seizoensmonsters op te tellen en voor het andere deel maar een monster beschikbaar te hebben is ervoor gekozen om alle monsters apart mee te nemen.
Als de monsters gedetailleerd levend worden uitgezocht en alle groepen worden meegenomen en gedetermineerd tot op soort zullen de meeste zeldzame soorten gevonden worden en kan een goed beeld van de situatie verkregen worden.

10 Beoordelen en waarderen: een voorlopige invulling

10.1 Beoordelen

Om te kunnen aangeven wat de ecologische waarde van een toekomstige voorspellingsresultaat, een cenotype, is dienen de cenotypen vooraf van een kwaliteitsoordeel te worden voorzien. Het is echter te vroeg om een beoordelingssysteem in de modellen op te nemen omdat dit (i) geen onderdeel vormt van deze studie en (ii) omdat een breed draagvlak nodig is voor een landelijk gedragen systeem naar EU-KRW eisen. Voor dit laatste worden lopen inmiddels initiatieven.
Om toch vooruit te kunnen met de modellering in de praktijk is aan ieder cenotype, op basis van de resultaten van hoofdstuk 8 en expert judgement een voorlopig kwaliteitsoordeel geplakt (tabel 9.1).

Tabel 9.1Kwaliteitsoordeel voor de beken- en slotencenotypen (kwaliteitsklasse 1 = slecht tot 4 = goed conform de KRW kwaliteitsklassen 1 tot 4 uit de schaal van 1 tot 5: zie ook paragraaf 8.3.2)

beek-cenotype	**kwaliteitsoordeel N01 - (N02) ***	**sloot-cenotype**	**kwaliteitsoordeel**
3a	4-3	1	3-4
21	4	2	3
24a	4-3 (3-2)	3	1
24b	4-3	4	3
24c	4	5	2
1	3	7	4
3b	3	9	4
9	2	10	2
19	3 (1)	11	3-4
20	3	18	3-4
25	3	19	4
26	3-2	28	4
10	2		
13	2		
15	2		
6	1		
14a	2		
14b	3		
16	2		
31	1-3		

* N01 = langzaam stromende beken; N02 snelstromende beken

10.2 Waarderen

Ten Brink et al. (2000) hebben vier operationele graadmeters uitgewerkt waarmee het nationale natuurbeleid kan worden ondersteund: natuurwaarde, soortgroep trend index, rode lijst indicator en EHS-doelrealisatiegraadmeter. De natuurwaarde is gebaseerd op het areaal (=percentage van het oppervlak van Nederland) en kwaliteit gemeten aan natuurlijkheid (=percentage soorten van de referentie). Aangezien

KRW-referenties voor Nederlandse aquatische systemen nog niet zijn beschreven en daarmee meetbare oppervlakten nog onbekend zijn is deze graadmeter nog niet aan de orde. De soortgroep trend index beschrijft de trend van afzonderlijke soortgroepen of deelsets daarvan vanaf een vast vergelijkingsjaar. Bij het vast stellen van de macrofauna doelsoortenlijst ten behoeve van het aquatisch Supplement is een inschatting van de trend gemaakt. De doelsoortenlijst is vooralsnog niet in de waardering betrokken maar dit is wel mogelijk nadat ze formeel zijn vastgesteld. Dit laatste dient te geschieden door de minister van LNV.
De rode lijst indicator maakt gebruik van zeldzaamheid en wordt berekend voor een hele soortengroep. Hier is gebruik gemaakt van de zeldzaamheid per soort.
De EHS-doelrealisatie meet de mate van succes van de implementatie van het EHS-natuurbeleid. Dit geschiedt naar oppervlak en kwaliteit. Kwaliteit wordt gemeten aan de presentie van doelsoorten per natuurdoeltypen.
Onderdelen van bovengenoemde graadmeters zijn inmiddels als bouwstenen voorhanden maar nog niet formeel vast gesteld voor aquatische systemen. Vooralsnog is zeldzaamheid ook het meest meetbare criterium.

In hoofdstuk 9 is reeds geconcludeerd dat zeldzame soorten goed gebruikt kunnen worden voor de waardering van de natuur. Ook is geconcludeerd dat het gemiddeld aantal zeldzame soorten een veel beter beeld geeft.
Voor zowel sloten als beken kan het gemiddeld aantal zeldzame soorten als basis dienen voor een waardering (tabel 9.2). De klassen in tabel 9.2 zijn vastgesteld op basis van expert judgement. Validatie van deze klassen is nog noodzakelijk. Bij deze benadering moet worden opgemerkt dat zeldzaamheid zich vaak niet aan typologische verschillen houdt en beter per monster berekend en toegepast kan worden. Hierin verschillen kwaliteitsbeoordeling en waardering van elkaar.

Tabel 9.2 Waardering op basis van zeldzaamheid voor sloten en beken gebaseerd op het gemiddeld aantal zeldzame soorten

waardering	**sloten**	**beken**
hoog	>1.60	>2.00
redelijk	0.81-1.60	1.01-2.00
matig	0.41-0.80	0.51-1.00
laag	<= 0.40	<= 0.50

Op basis van de 4 waarderingsklassen, zoals gedefinieerd in tabel 9.2, kan aan ieder cenotype op basis van het cenotype-gemiddelde, een voorlopige waardering worden gekoppeld. In tabel 9.3 is de voorlopige waardering van de sloten- en bekencenotypen gegeven.

Tabel 9.3 Waarde voor natuur voor de beken- en slotencenotypen

beek-cenotype	**waarde**	**sloot-cenotype**	**waarde**
3a	redelijk	1	matig
21	hoog	2	matig
24a	hoog	3	laag
24b	redelijk	4	laag
24c	redelijk	5	matig
1	redelijk	7	hoog
3b	hoog	9	redelijk
9	redelijk	10	laag
19	redelijk	11	laag
20	hoog	18	laag
25	hoog	19	hoog
26	redelijk	28	redelijk
10	redelijk		
13	matig		
15	matig		
6	matig		
14a	matig		
14b	laag		
16	matig		
31	laag		

Referenties

AQEM consortium 2002. Manual for the application of the AQEM system. A comprehensive method to assess European streams using benthic macroinvertebrates, developed for the purpose of the Water Framework Directive. Version 1.0, February 2002.

Armitage P.D., D. Moss, J.F. Wright & M.T. Furse 1983. The performance of a new biological water quality score system based on macroinvertebates over a wide range of unpolluted running-water sites. Water Res. Vol. 17, no 3: 333-347.

Bode R.W. 1988. Quality Assurance Work Plan for Biological Stream Monitoring in New York State. Stream Biomonitoring Unit, Bureau of Monitoring and Assessment, Division of Water, New York State Department of Environmental Conservation, Albany, NY.

Chandler J.R. 1970. A biological approach to water quality management. Wat. Pollut. Control, 69, 415-422.

Chutter F.M. 1972. An empirical biotic index of the quality of water in South African streams and rivers. Wat. Res., 6, 19-30.

Czekanowski J. 1913. Zarys metod statystycznych Die grundzuge der statistischen Metoden. S.n., Warsaw.

Hartog C. den 1963. The anphipods of the Caltaic region of the rivers Rhine, Meuse and Schelde in relation to the hydrography of the area. I: Introduction and hydrography. Netherlands Journal of Sea Research 2: 29-39.

EU 2000. Richtlijn 2000/60/EG Van het Europees Parlement en de raad tot vaststelling van een kader voor communautaire maatregelen betreffende het waterbeleid. Publicatieblad van de Europese Gemeenschappen, L 327, 1-72.

Gardeniers J.J.P. 1976. Problematiek en waarde van de biologische beoordeling van waterkwaliteit. In: Practische aspecten van de hydrobiologie. Landbouwhogeschool Wageningen, Vakgroep Waterzuivering, Wageningen.

Gaston, K.J. 1994. Rarity. Population and Community Biology Series, no 13, Chapman & Hall, London.

Gauch, H.G. Jr. 1982. Multivariate Analysis in Community Ecology. Cambridge University Press, Cambridge.

Hellawell J.M. 1978. Biological surveillance of rivers: A biological monitoring handbook. Water Research Centre, Stevenage, 331p.

Higler L.W.G. & H.H. Tolkamp 1984. Karakterisering van stromende wateren met behulp van bioindicatoren: het geslacht Hydropsyche (Trichoptera). In E.P.H. Best & J. Haeck (eds). Ecologische indicatoren voor de kwaliteitsbeoordeling van lucht, water, bodem en ecosysteem. Symposium van de Oecologische Kring, Utrecht, 14-15 october 1982. Pudoc, Wageningen.

Hosmer D.W. & S. Lemeshow 1989. Applied logistic regression. Wiley, New York.

Intergovernmental Task Force on Monitoring Water Quality 1993. The Multimetric Benadering for describing ecological condition. EPA, Position Paper No. 2

Jaccard P. 1912. The distribution of the flora in the alpine zone. New. Phytol., 11, 37-50.

Johnson R.K. 1998. Benthic macroinvertebrates. pp 85-166 In: Sjöar och vattendrag. Bakgrunds rapport, Biologiska parameter t. Wiederholm, ed., Naturvärdsverket Rapport 4921.

Karr J.R., K.D. Fausch, P.L. Angermeier, P.R. Yant & I.J. Schlosser 1986. Assessing Biological Integrity in Running Waters: A Method and its Rationale. Special Publication 5. Illinois Natural History Survey, Urbana, Illinois.

Karr J.R. 1991. Biological integrity: a long-neglected aspect of water resource management. Ecological Applications, 1(1): 66-84.

Kothe P. 1962. Der 'Artenfehlbetrag', ein einfaches Gütekriterium und seine Anwendung bei biologischen Vorfluteruntersuchungen. Dtsch. Gewässerkundl. Mitt., 6, 60-65.

Linde N. ter. & P.B. Worm 1996. Indices in de auquatische ecologie: een overzicht. Tauw, Deventer, 43p.

Little R.J.A. & Rubin D.B. 1987. Statistical analysis with missing data. Wiley. New York.

Mason W.T., P.A. Lewis & C.I. Weber 1985. An evaluation of benthic macroinvertebrate biomass methodology. Environmental Monitoring and Assessment, 5, 399-422.

Nijboer R.C. 1996. Ecologische karakterisering van oppervaktewateren in Overijssel: Toetsing van een expertsysteem voor regionaal waterbeheer. Afstudeerverslag, Vakgroep Milieukunde, Katholieke Universiteit Nijmegen, Nijmegen, 75p.

Nijboer R.C. 2000. Natuurlijke levensgemeenschappen van de Nederlandse binnenwateren. Deel 6. Sloten. Achtergronddocument bij het 'Handboek Natuurdoeltypen in Nederland'. Rapport AS-06, EC-LNV. Alterra, Wageningen. 80 blz.

Nijboer R.C., Goedhart P.W., Verdonschot P.F.M. & Ter Braak C.J.F. 1998. Effecten van ingrepen in het waterbeheer op aquatische levensgemeenschappen. cenotypenbenadering, fase 1: ontwikkeling van het prototype. RIZA werkdoc. 98.141X, STOWA werkrapportnr 98-W-03, RIVM rapportnr 70 37 18 004. 140 pp.

Nijboer R.C. & Verdonschot P.F.M. 2002. Vegetatie en macrofauna van de Nederlandse sloten. Van typologie tot beoordeling. Alterra-rapport 000, Wageningen. 000 blz. (in voorbereiding)

Ohio Enviromental Protection Agency 1987. Biological Criteria for the Protection of Aquatic Life: Volume I-III. Ohio EPA, Division of Water Quality Monitoring and Assessment, Surface Water Section, Columbus Ohio.

Pianka E.R. 1978. Evolutionary ecology. Harper and Row Publishers, New York.

Plafkin J.L., M.T. Barbour, K.D. Porter, S.K. Gross & R.M. Hughes 1989. Rapid Bioassessment Protocols for Use in Streams and Rivers: Benthic Macroinvertebrates and Fish. EPA/44/4-89-001. U.S. EPA, Office of Water, Washington, D.C.

Preston E.M. & B.L. Bedford 1988. Evaluating cumulative effects on wetland functions: a conceptual overview and general framework. Environmental Management 12: 515-583.

Resh V.H. & G. Grodhaus 1983. Aquatic insects in urban enviroments. In Urban Entomology: Interdisciplinary Perspectives, eds. G.W. Frankie & C.S. Koehler, 247-76. Praeger Pubs., New York.

Rosenberg D.M. & V.H. Resh 1993. Freshwater bio-monitoring and benthic macro-invertebrates. Chapman & Hall, New York.

Rubin D.B. 1996. Multiple imputation after 18+ years. Journal of the American Statistical Association, 91, 473-489.

Schafer J.L. 1997. Analysis of incomplete multivariate data. Chapman and Hall. London.

Shackleford B. 1988. Rapid Bioassessments of Lotic Macroinvertebrate Communities: Biocriteria Development. Biomonitoring Section, Arkansas Department of Pollution Control and Ecology, Little Rock, AR.

Shannon C.E. & W. Weaver 1949. The mathematical theory of communication. The University of Illinois Press., Urbana, 107p.

Sørensen T 1948. A method of establishing groups of equel amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. Biol. Skr., 5, 1-34.

Stichting Toegepast Onderzoek Waterbeheer 1992. Ecologische beoordeling en beheer van oppervlaktewater: Beoordelingssysteem voor stromende wateren op basis van macrofauna. STOWA, Utrecht, 58p.

Ten Brink B.J.E., Strien A. van, Hinsberg A. van, Reijnen M.S.J.M., Wiertz J., Alkemade J.R.M., Dobben H.F. van, Higler L.W.G., Koolstra B.J.H., Ligtvoet W., Peijl M. van der & Semmekrot S. 2000. Natuurgraadmeters voor de behoudoptiek. RIVM, Alterra, CBS. RIVM rapport 408657005.

Thienemann A. 1925. Die Binnengewässer Mitteleuropas. Eine Limnologische Einführung. In: Die Binnengewässer, BD. 1. Schweizerbart, Stuttgart.

Tolkamp H.H. & J.J.P. Gardeniers 1971. Hydrobiological survey of lowland streams. PhD Thesis, Standaard boekhandel, Tilburg, 286p.

Tongeren O. van 2000. Programma Associa: Gebruikershandleiding en voorwaarden. Data-Analyse Ecologie, s.l.

USEPA 1993. An SAB report: evaluation of draft technical guidance on biological criteria for streams and small rivers. ited States Environmental Protection Agency, Washington, 24 p.

Verdonschot P.F.M. 1983. Ecologische karakterisering van oppervlaktewateren in Overijssel. Provinciale Waterstaat in Overijssel, Zwolle. 1-57.

Verdonschot P.F.M. 1983. Ecologische karakterisering van oppervlaktewateren in Overijssel. H_2O, 16, 547-579.

Verdonschot P.F.M. 1990a. Ecological characterization of surface waters in the province of Overijssel (The Netherlands). Thesis, Wageningen, 1-255.

Verdonschot P.F.M. 1990b. Ecologische karakterisering van oppervlaktewateren in Overijssel. Het netwerk van cenotypen als instrument voor ecologisch beheer, inrichting en beoordeling van oppervlaktewateren. Provincie Overijssel, Zwolle, Rijksinstituut voor Natuurbeheer, Leersum, 301p.

Verdonschot, P.F.M. 1991. The web-approach: a tool in water management. In: Ecological Water Managemenet in Practice (Proceedings of the technical meeting held in Ede, The Netherlands, 3 October 1990). Proc. and Inf. CHO-TNO, 45: 59-76.

Verdonschot P. et al. (red.) 1995. Beken stromen. Leidraad voor ecologisch beekherstel. Werkgroep Ecologisch Waterbeheer, subgroep Beekherstel, WEW-06. Stichting Toegepast Onderzoek Waterbeheer, STOWA 95-03, Utrecht. 1-236.

Verdonschot P.F.M. 2000. Natuurlijke levensgemeenschappen van de Nederlandse binnenwateren. Deel 2, Beken. Achtergronddocument bij het 'Handboek Natuurdoeltypen in Nederland'. Rapport AS-02, EC-LNV. Alterra, Wageningen. 128 blz.

Verdonschot P.F.M., M. Koopmans & R.C. Gerritsen 1999. Ecologisch maatweb stromende wateren. Waterschap Veluwe en Waterschap Vallei & Eem. 142 pp.

Verdonschot P.F.M. & P.W. Goedhart 2000. RISTORI. Effecten van ingrepen in het waterbeheer op aquatische levensgemeenschappen. Cenotypenbenadering, fase 2: Verfijning van het prototype. Alterra rapport.

Verdonschot P.F.M. & Nijboer R.C. 2002. Beken in Nederland. Macro-invertebraten en beekplanten in Nederlandse beeksystemen. Alterra-rapport 000, Wageningen. 000 blz. (in voorbereiding)

Vlek H., Verdonschot P.F.M., Nijboer R.C., Hoek van den Tj.H. & Hoorn M.W. van der 2002. De ontwikkeling en toetsing van de Europese AQEM beoordelingsmethode voor stromende wateren in Nederland. Alterra-rapport 000, Wageningen. 000 blz. (in voorbereiding)

Weber C.I. 1973. Biological Field and Laboratory Methods for Measuring the Quality of Surface Waters and Effluents. EPA-670/4-73-001. U.S. Enviromental Protection Agency, Cincinnati, OH.

Williams W.T. & M.B. Dale 1965. Fundamental problems in numerical taxonomy. Advance. Bot. Res., 2, 35-68.

Woodiwiss F.S. 1964. The biological system of stream classification used by the Trent River Board. Chemistry & Industry, 1, 1443-1447.

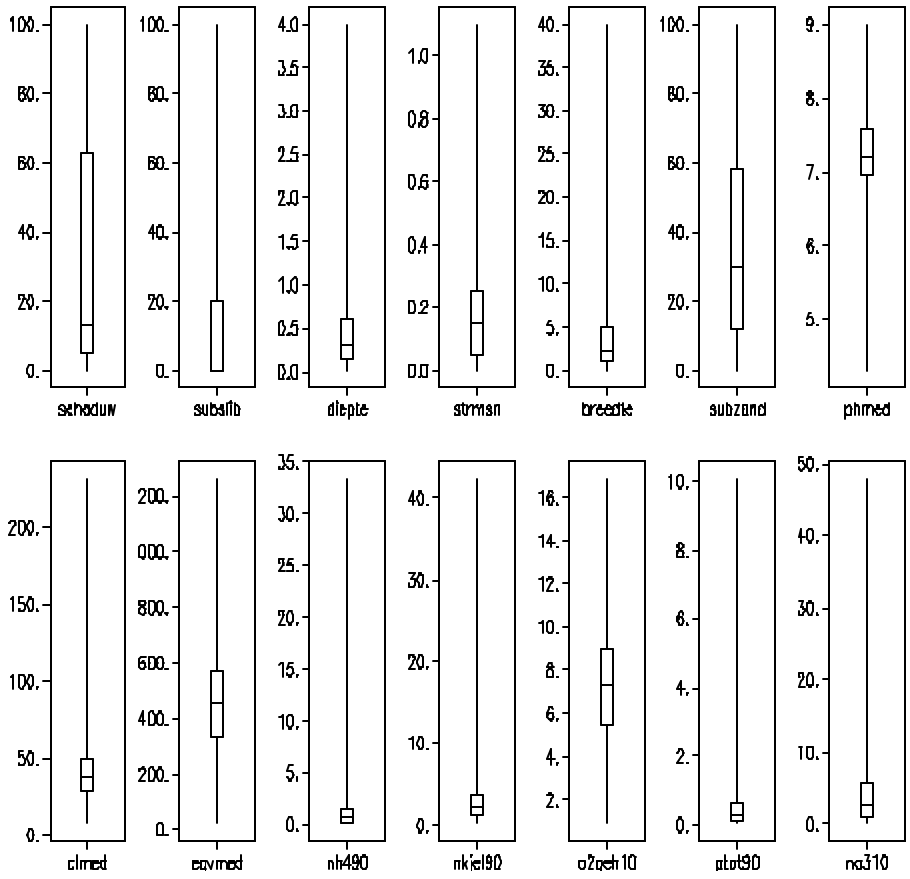
Zelinka M. & P. Marvan 1961. Zur Prazisierung der biologischen Klassifikation der Reinheit fliessender Gewasser. Arch. Hydrobiol., 57, 389-407.

Zonneveld I.S. 1984. Grondslagen voor de bioindicatie. In: E.P.H. Best & J. Haeck (eds). Ecologische indicatoren voor de kwaliteitsbeoordeling van lucht, water, bodem en ecosystemen.

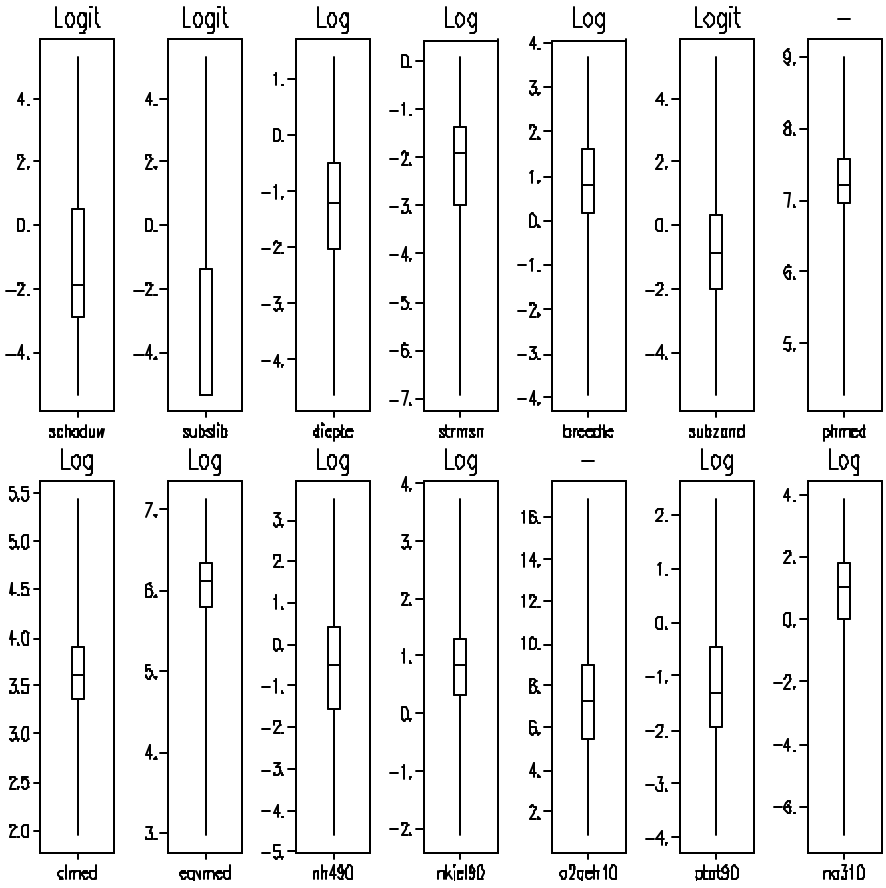
Bijlagen

Bijlage 1

Bijlage 1A1. Box plots van ongetransformeerde continue predictoren.



Bijlage 1A2. Box plots van getransformeerde continue predictoren (log of logit).



```
# DEFINE DATAMATRIX AND PERFORM REGRESSION ON COMPLETE UNITS
x <- matrix(nrow=18, ncol=3)
x[,1] <- c(NA,NA,55,78,NA,65,43,69,61,NA,35,55,63,76,NA,43,61,NA)
x[,2] <- c(62,11,NA,63,NA,51,NA,59,45,58,21,26,50,NA,47,32,59,50)
x[,3] <- c(NA,29,NA,66,49,49,37,NA,53,51,NA,54,55,53,52,NA,50,NA)
summary.lm(lm(x[,1] ~ x[,2] + x[,3], na.action=na.omit))
# ESTIMATE PARAMETERS OF TRIVARIATE NORMAL BY MEANS OF EM
library(norm)
s <- prelim.norm(x)
theta <- em.norm(s, maxits=100, crit=1.0e-8)
param <- getparam.norm(s, theta, corr=T)
param
# DO IMPUTATIONS IN A LOOP
nimputation <- 100      # NUMBER OF IMPUTATIONS
stepsda <- 500          # NUMBER OF STEPS IN DATA AUGMENTATION
rngseed(329031)
alfa    <- list(1:nimputation)
sealfa  <- list(1:nimputation)
beta1   <- list(1:nimputation)
sebeta1 <- list(1:nimputation)
beta2   <- list(1:nimputation)
sebeta2 <- list(1:nimputation)
for (i in 1:nimputation) {
  cat(i, "..", sep="")
  theta <- da.norm(s, theta, steps=stepsda)     # GET DRAW FROM POSTERIOR
  imp <- imp.norm(s, theta, x)                  # IMPUTE GIVEN THIS DRAW
  summary <- summary.lm(lm(imp[,1] ~ imp[,2] + imp[,3]))
  alfa[i] <- summary$coefficients[1,1]
  sealfa[i]   <- summary$coefficients[1,2]
  beta1[i] <- summary$coefficients[2,1]
  sebeta1[i]   <- summary$coefficients[2,2]
  beta2[i] <- summary$coefficients[3,1]
  sebeta2[i]   <- summary$coefficients[3,2]
}
c0 <- mi.inference(alfa, sealfa)
c1 <- mi.inference(beta1, sebeta1)
c2 <- mi.inference(beta2, sebeta2)
result <- matrix(nrow=3, ncol=6, dimnames=list(c("alfa","beta1","beta2"),
          c("Est","Se","Tval","Df","Signif","Minf")))
result[1,] <- c(c0$est, c0$std.err, NA, c0$df, c0$signif, c0$fminf)
result[2,] <- c(c1$est, c1$std.err, NA, c1$df, c1$signif, c1$fminf)
result[3,] <- c(c2$est, c2$std.err, NA, c2$df, c2$signif, c2$fminf)
result[,3] <- result[,1]/result[,2]
result



# START EM FROM DIFFERENT INITIAL VALUES OBTAINED BY DATA AUGMENTATION
nstart <- 50              # NUMBER OF DIFFERENT STARTING POINTS
rngseed(329031)
start <- matrix(nrow=nstart, ncol=7)    # TO SAVE LOGLIK AND STARTING VALUES
theta.new <- theta
for (i in 1:nstart) {
  cat(i, "..", sep="")                              # MONITORING
  theta.new <- da.norm(s, theta.new, steps=500)    # GET DRAW FROM POSTERIOR
  param.new <- getparam.norm(s, theta.new, corr=T)
  theta <- em.norm(s, start=theta.new, maxits=200, crit=1.0e-8, showits=F)
  start[i,1] <- loglik.norm(s, theta)     # SAVE LOGLIK
  start[i,c(2:4)] <- param.new$mu         # SAVE STARTING MEANS
  start[i,c(5:7)] <- param.new$sdv        # SAVE STARTING STANDARD ERRORS
}
var(start[,1])       # VARIANCE OF LOGLIK SHOULD BE CLOSE TO ONE
start
```

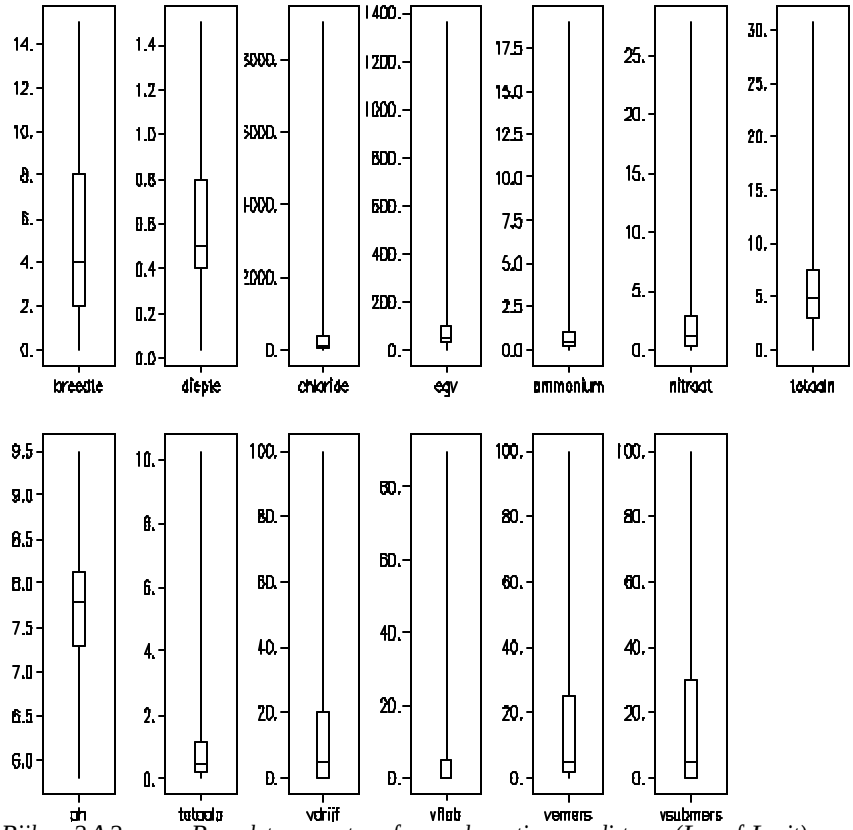
Bijlage 1C. Gemiddelde resubstitutie terugvoorspelkans van cenotype naar cenotype voor Model I. Hoofdgroepen zijn weergegeven door een vette omlijning, en de diagonaal is vet afgedrukt. Kansen kleiner dan 0.01 zijn niet afgedrukt en kansen groter dan 0.99 zijn afgedrukt als 0.99.

Cenot.	Ny	A	A	A	A	A	A	B	B	B	B	B	B	B	B	B	C	C	D	D	D	E	E	E	E	F
		3a	3b	24a	24b	24c	25	1	6	10	14a	14b	16	19	20	31	7	27	2	8	12	9	13	15	26	21
A-3a	39	**.46**	.02	.12	.10	.08	-	.01	.05	.02	-	-	-	.03	-	-	.05	-	-	-	-	.01	-	-	.01	.02
A-3b	6	.19	**.38**	.08	.18	.05	-	-	-	.02	-	-	-	-	-	-	.02	-	-	-	-	.03	-	-	.02	-
A-24a	24	.15	.04	**.29**	.17	.19	-	-	-	.02	-	-	-	-	-	.01	.06	-	.01	-	-	.02	-	-	-	-
A-24b	39	.14	.02	.10	**.48**	.15	-	-	-	-	-	-	-	-	-	-	.04	-	-	-	-	-	-	-	.01	.03
A-24c	43	.05	-	.14	.09	**.58**	-	-	-	-	-	-	-	-	-	-	.07	-	-	-	-	-	-	-	.02	.02
A-25	8	-	-	-	-	-	**.66**	-	-	.01	-	.03	.02	.06	.12	.01	-	-	-	-	-	.02	-	-	.04	-
B-1	14	-	-	.02	.01	-	-	**.11**	.19	.26	-	.01	-	.04	-	-	.10	.01	.02	-	-	.05	.04	.08	.02	-
B-6	89	-	-	-	-	-	-	.03	**.44**	.18	.01	.03	.01	.14	-	.04	.01	-	-	-	-	.02	.01	.04	.02	-
B-10	84	.02	-	-	-	-	-	.03	.17	**.51**	-	-	-	.04	-	.02	.01	-	-	-	-	.09	.02	.03	.03	-
B-14a	10	-	-	-	-	-	-	-	.07	.01	**.67**	-	-	.23	-	-	-	-	-	-	-	-	-	-	-	-
B-14b	9	-	-	-	-	-	-	-	.25	.13	-	**.37**	-	.09	-	.12	-	-	-	-	-	-	-	.01	-	-
B-16	3	.04	.02	-	.08	-	-	.01	.28	.09	-	.04	**.04**	.21	-	.01	.04	-	-	-	-	.04	.02	-	.08	-
B-19	48	.01	-	-	-	-	-	-	.21	.09	.04	.02	.02	**.55**	-	-	-	-	-	-	-	-	-	-	.01	-
B-20	4	.01	-	-	-	-	-	-	-	-	-	-	-	-	**.94**	-	.02	-	-	-	-	-	-	-	.02	-
B-31	9	-	-	-	-	-	.01	.02	.26	.24	-	.02	-	.05	-	**.26**	.03	-	-	-	-	.05	.01	-	.03	-
C-7	29	.05	-	.07	.07	.07	-	.02	-	.04	-	-	-	-	-	-	**.56**	-	-	-	-	.03	.01	.01	.03	-
C-27	3	.03	-	.09	.04	.18	-	.08	.01	.03	-	-	-	-	-	-	-	**.41**	-	-	-	.02	.06	.03	.02	-
D-2	15	-	-	-	-	-	-	.03	.04	.03	-	-	-	-	-	-	-	-	**.84**	-	-	-	-	.02	-	-
D-8	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.13	-	-	**.85**	-	-	.02	-	-	-
D-12	12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**.97**	-	-	-	-	-
E-9	22	.04	.03	.03	-	.01	-	.02	.07	.27	-	.03	-	-	-	.03	.03	-	.01	-	-	**.22**	.05	.08	.07	-
E-13	11	.01	-	.04	.01	-	-	.02	.08	.14	-	.02	.02	.02	-	.15	.02	-	-	.01	-	.08	**.26**	.05	.05	-
E-15	16	-	-	-	-	.03	-	.07	.13	.19	-	.01	-	-	-	.02	.02	.03	-	-	-	.10	.06	**.27**	.04	.03
E-26	11	.03	-	.04	.07	.09	-	.03	.11	.07	.05	.02	-	.01	-	-	.05	-	-	-	-	.18	.05	.04	**.13**	-
F-21	13	.03	.01	.12	.06	.04	-	-	-	-	-	-	-	-	-	-	.01	-	-	-	-	-	-	-	-	**.72**

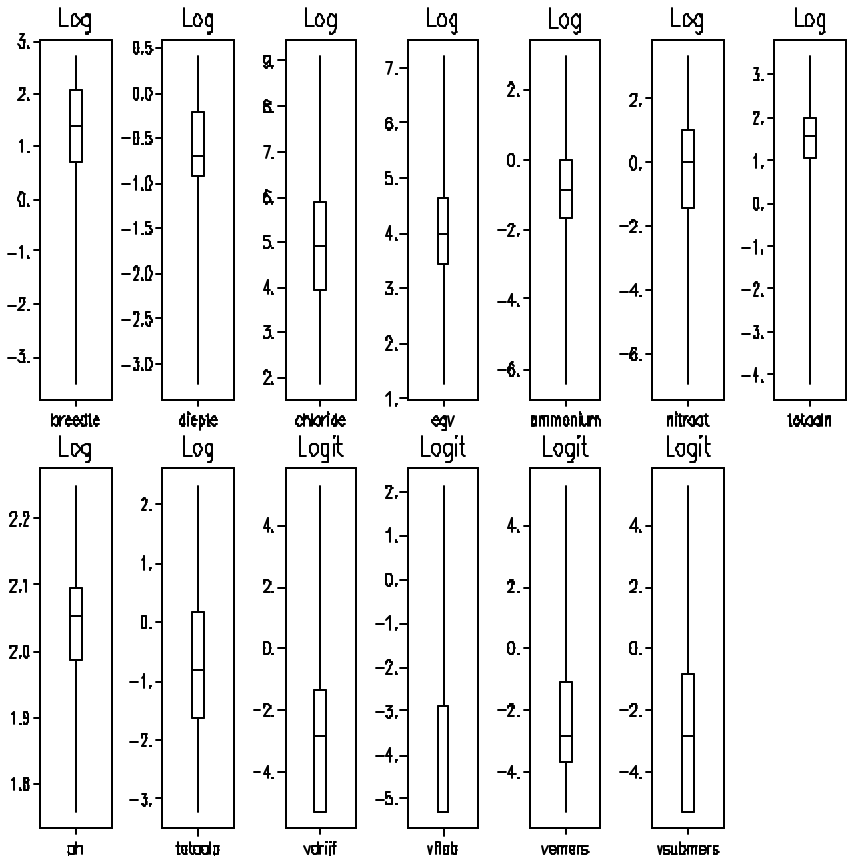
Bijlage 1D. Gemiddelde resubstitutie terugvoorspelkans van cenotype naar cenotype voor Model II. Hoofdgroepen en groepen zijn weergegeven door een vette omlijning, en de diagonaal is vet afgedrukt. Kansen kleiner dan .01 zijn niet afgedrukt en kansen groter dan .99 zijn afgedrukt als .99

Cenot.	Ny	A1	A1	A1	A1	A1	A2	A2	A3	B1	B2	B2	B2	B3	B3	B4	B4	B4	B4	B5	B5	B5	B5	C1	C1	C1
		24a	24b	24c	7	27	3a	3b	21	6	1	10	15	25	20	14a	14b	16	19	31	9	13	26	2	8	12
A1-24a	24	**.22**	.18	.19	.07	-	.14	.03	.06	.01	.01	.03	.01	-	-	-	-	-	-	-	.02	-	.01	-	-	-
A1-24b	39	.12	**.36**	.16	.08	-	.13	.02	.05	.01	-	.02	-	-	-	-	-	-	-	-	.01	-	-	-	-	-
A1-24c	43	.13	.12	**.54**	.06	-	.06	-	.03	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A1-7	29	.06	.06	.10	**.57**	-	.05	-	.01	-	.01	.03	.03	-	-	-	-	-	-	-	.02	-	.03	.01	-	-
A1-27	3	-	-	-	-	**.57**	.04	-	.03	.09	.05	.06	.04	-	-	-	-	-	-	.01	.02	.03	-	-	-	.05
A2-3a	39	.07	.10	.08	.06	.02	**.38**	.01	-	.04	.02	.05	-	.03	-	-	-	-	.03	.02	.01	-	.03	-	-	-
A2-3b	6	.18	.09	.11	.03	-	.10	**.21**	-	.01	.03	.07	.02	-	.03	-	-	-	-	-	.06	-	.06	-	-	-
A3-21	13	.12	.11	.08	.02	.01	.07	-	**.57**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
B1-6	89	-	-	-	-	-	-	-	-	**.44**	.03	.19	.03	-	-	.03	.02	.02	.12	.02	.02	-	.01	-	-	-
B2-1	14	-	-	-	.12	-	.07	-	-	.22	**.07**	.21	.08	-	-	-	-	.03	.01	.03	.04	.02	.02	.04	-	-
B2-10	84	-	-	-	-	-	.01	-	-	.16	.04	**.47**	.03	-	-	.01	-	.02	.04	.03	.10	.02	.03	.01	-	-
B2-15	16	-	.01	.02	.09	-	-	-	.04	.15	.06	.15	**.23**	-	-	-	-	-	-	-	.14	.02	.03	-	-	.01
B3-25	8	.07	-	-	-	.07	.07	.17	-	-	-	.04	-	**.42**	-	-	-	.05	.01	-	.03	-	.03	-	-	-
B3-20	4	-	-	-	-	-	.07	-	-	.05	-	.02	-	-	**.46**	-	-	.09	.26	.02	-	-	-	.02	-	-
B4-14a	10	-	-	-	-	-	-	-	-	.10	-	.04	-	-	-	**.75**	-	-	.10	.01	-	-	-	-	-	-
B4-14b	9	-	-	-	-	-	-	-	-	.38	-	.20	-	-	-	-	**.31**	-	.04	.04	.01	-	-	-	-	-
B4-16	3	-	-	-	-	.05	.07	-	-	.26	.01	.09	.01	.05	.05	-	-	**.23**	-	.02	.05	.01	.07	-	-	-
B4-19	48	-	-	-	-	-	.01	-	-	.20	-	.09	-	.03	.01	.02	.01	-	**.57**	.02	-	-	-	-	-	-
B5-31	9	-	-	.02	.05	-	-	-	-	.30	.02	.27	-	-	-	-	.01	.09	.03	**.15**	.04	-	.02	-	-	-
B5-9	22	-	-	.01	.05	-	.04	.02	-	.07	.03	.32	.08	.02	-	-	-	.01	-	-	**.21**	.03	.05	-	-	-
B5-13	11	-	.04	-	.06	-	.01	-	-	.05	.02	.25	.05	-	.01	-	.02	.02	-	.03	.06	**.28**	.02	-	-	.06
B5-26	11	.01	.02	.09	.04	.02	.05	.02	-	.09	.05	.15	.04	-	-	.04	.01	-	-	.01	.17	.03	**.13**	-	-	-
C1-2	15	-	-	-	.03	-	.03	-	-	.07	.01	.06	.03	-	-	-	-	.03	-	-	-	.02	-	**.72**	-	-
C1-8	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**.99**	-
C1-12	12	-	-	-	.01	-	-	-	-	.04	.01	.04	.01	-	-	-	-	-	-	-	-	-	-	-	-	**.86**

Bijlage 2 Box plots van ongetransformeerde continue predictoren.



Bijlage 2A2. Box plots van getransformeerde continue predictoren (Log of Logit).


Log
breedte
Log
diepte
Log
chloride
Log
egv
Log
ammonium
Log
nitraat
Log
totaalN
Log
ph
Log
totaalp
Logit
vdrijf
Logit
vflab
Logit
vemers
Logit
vsubmers

Bijlage 3

Bijlage 3A. Voorbeeld van een invoer file voor Scenario.Exe

```
Voorspelling voor beken
gaf496
   1.000   1.000   1.000   0.000   0.000   0.000  10.000   0.500   0.100   2.000  90.000 7.150
  75.000   2.008   3.920   0.420   0.106     -99
rbw591
   0.000   0.000   1.000   1.000   0.000  38.000   9.709   1.500   0.090   7.200  58.252 7.350
 150.000  21.550  23.000   2.965   5.850     -99
rbxn90
   0.000   0.000   1.000   1.000   1.000  13.000   5.000   0.500   0.250   6.000  70.000 7.500
 190.000  25.730  29.330   5.552   2.330     -99
rcm090
   0.000   0.000   1.000   1.000   0.000  13.000  10.000   0.500   0.050   5.000  80.000 7.300
  82.500  20.170  22.820   2.970   4.540     -99
rcod90
   0.000   1.000   1.000   0.000   0.000  63.000  15.000   0.200   0.180   1.900  60.000 7.500
 139.000  24.000  30.600   3.782   5.840     -99
rcv494
   1.000   1.000   1.000   0.000   0.000  63.000   5.155   0.090   0.200   1.600  63.918 7.225
  37.500   0.400   2.250   0.166   8.250     -99
reb592
   1.000   1.000   1.000   0.000   0.000  13.000   0.000   0.060   0.350   0.250  35.000 6.200
  26.000   0.010   1.118   0.372   8.480     -99
rfa095
   1.000   1.000   1.000   0.000   0.000  88.000   0.000   0.050   0.000   0.260   2.000 6.970
  30.000   0.200   1.620   0.276   2.070     -99
yayv96
   1.000   1.000   1.000   0.000   0.000  80.000   0.000   0.800   0.400   7.000  40.000 7.700
  42.000   3.160   6.710   1.281   3.650     -99
ybmn93
   0.000   1.000   1.000   1.000   0.000  10.000  25.000   1.250   0.010  14.000  25.000 7.400
  42.000   1.200   2.840   0.524   2.200     -99
ybov93
   0.000   0.000   1.000   1.000   0.000  10.000  25.000   1.500   0.050  14.000  25.000 7.900
  49.000   0.440   1.100   0.410   1.060     -99
ycmn93
   0.000   0.000   0.000   0.000   0.000  40.000   8.333   0.300   0.200   1.200  41.667 7.100
  29.000   0.400   2.700   0.130   5.660     -99
yelv96
   1.000   1.000   1.000   0.000   0.000  60.000   0.000   0.250   0.250   1.000  22.222 7.100
  33.500   0.800   2.490   0.262   6.920     -99
```

```
 Voorspelling voor beken
 Beken Model II (zie PGO-2001-05)
  3 (# Aantal niveaus in hierarchie)
  3 (# Niveau 1)
     A            B            C
  9 (# Niveau 2)
     A1           A2           A3           B1           B2           B3           B4
     B5           C1
 25 (# Niveau 3)
     24a          24b          24c          7            27           3a           3b
     21           6            1            10           15           25           20
     14a          14b          16           19           31           9            13
     26           2            8            12
 17 (# Milieuvariabelen)
    meander        0    0.0000000000E+00   -0.9900000000E+02   -0.9900000000E+02
    dwarsnat       0    0.0000000000E+00   -0.9900000000E+02   -0.9900000000E+02
    perman         0    0.0000000000E+00   -0.9900000000E+02   -0.9900000000E+02
    stuw           0    0.0000000000E+00   -0.9900000000E+02   -0.9900000000E+02
    winter         0    0.0000000000E+00   -0.9900000000E+02   -0.9900000000E+02
    schaduw        2    0.5000000000E+00    0.0000000000E+00    0.1000000000E+03
    subslib        2    0.2105263160E+00    0.0000000000E+00    0.1000000000E+03
    diepte         1    0.0000000000E+00    0.1000000000E-01    0.4000000000E+01
    strmsn         1    0.1000000000E-02    0.0000000000E+00    0.1100000000E+01
    breedte        1    0.2000000000E-01    0.0000000000E+00    0.4000000000E+02
    subzand        2    0.2000000000E+00    0.0000000000E+00    0.1000000000E+03
    phmed          0    0.0000000000E+00    0.4300000000E+01    0.9000000000E+01
    clmed          1    0.0000000000E+00    0.7000000000E+01    0.2320000000E+03
    nh490          1    0.0000000000E+00    0.1000000000E-01    0.3350000000E+02
    nkjel90        1    0.0000000000E+00    0.1200000000E+00    0.4250000000E+02
    ptot90         1    0.0000000000E+00    0.1900000000E-01    0.1010000000E+02
    no310          1    0.1000000000E-02    0.0000000000E+00    0.4800000000E+02
gaf496                    1    0    0    0.0000000000E+00
 0.0007755 0.9375229 0.0617015 0.0003047 0.0004708 0.0000000 0.3683046 0.3063310 0.0127971
0.0934378
 0.1566526 0.0617015 0.0000051 0.0000027 0.0000047 0.0002922 0.0000000 0.0004707 0.0000001
0.0000000
 0.3683046 0.0303426 0.1917775 0.0842109 0.0000000 0.0127971 0.0000000 0.0000000 0.0934378
0.0000000
 0.0002102 0.0604648 0.0778635 0.0181140 0.0617015 0.0000000 0.0000000
rbw591                    2    0    0    0.0000000000E+00
 0.0000032 0.9999702 0.0000265 0.0000000 0.0000032 0.0000000 0.1293587 0.6486249 0.0007360
0.1445902
 0.0766604 0.0000265 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000032
0.0000000
 0.1293587 0.0020352 0.6448407 0.0017490 0.0007360 0.0000000 0.0000000 0.0000000 0.1445902
0.0000000
 0.0005023 0.0726343 0.0016676 0.0018562 0.0000265 0.0000000 0.0000000
rbxn90                    3    0    0    0.0000000000E+00
 0.0000008 0.9999803 0.0000190 0.0000005 0.0000003 0.0000000 0.0381714 0.6040849 0.0171617
0.0339271
 0.3066352 0.0000190 0.0000000 0.0000000 0.0000000 0.0000005 0.0000000 0.0000000 0.0000003
0.0000000
 0.0381714 0.0024576 0.5888521 0.0127751 0.0171617 0.0000000 0.0000000 0.0000000 0.0339271
0.0000000
 0.0000826 0.3007738 0.0005117 0.0052671 0.0000190 0.0000000 0.0000000
rcm090                    4    0    0    0.0000000000E+00
 0.0000192 0.9999697 0.0000111 0.0000001 0.0000191 0.0000000 0.0655708 0.6827287 0.0001097
0.0366186
 0.2149419 0.0000111 0.0000000 0.0000000 0.0000000 0.0000001 0.0000000 0.0000000 0.0000191
0.0000000
 0.0655708 0.0034285 0.6706756 0.0086246 0.0001097 0.0000000 0.0000001 0.0366185 0.0000000
0.0000000
 0.0001517 0.2076433 0.0028796 0.0042673 0.0000111 0.0000000 0.0000000
rcod90                    5    0    0    0.0000000000E+00
 0.0003125 0.9990817 0.0006059 0.0001221 0.0001900 0.0000003 0.0020359 0.7197368 0.0001715
0.0019110
 0.2752264 0.0006059 0.0000014 0.0000001 0.0000000 0.0001206 0.0000000 0.0000050 0.0001850
0.0000003
 0.0020359 0.0050564 0.4229422 0.2917382 0.0001715 0.0000000 0.0000000 0.0000000 0.0019110
0.0000000
 0.0000000 0.2613142 0.0103015 0.0036107 0.0006059 0.0000000 0.0000000
rcv494                    6    0    0    0.0000000000E+00
 0.8610237 0.1373806 0.0015957 0.5567812 0.3038848 0.0003578 0.0035559 0.0523967 0.0009958
0.0041362
 0.0762958 0.0015957 0.2537510 0.1462875 0.0294430 0.1272997 0.0000000 0.2959021 0.0079827
0.0003578
```

```
 0.0035559 0.0148707 0.0337475 0.0037785 0.0009958 0.0000000 0.0000000 0.0027509 0.0000000
0.0013853
 0.0012296 0.0236056 0.0057018 0.0457588 0.0015956 0.0000001 0.0000000
reb592                     7    0    0    0.0000000000E+00
 0.9992664 0.0000041 0.0007295 0.9987984 0.0002674 0.0002006 0.0000000 0.0000012 0.0000002
0.0000000
 0.0000027 0.0007295 0.0195447 0.0269421 0.9522777 0.0000339 0.0000000 0.0000202 0.0002472
0.0002006
 0.0000000 0.0000000 0.0000000 0.0000012 0.0000002 0.0000000 0.0000000 0.0000000 0.0000000
0.0000000
 0.0000001 0.0000001 0.0000003 0.0000022 0.0000000 0.0000000 0.0007295
rfa095                     8    0    0    0.0000000000E+00
 0.9953972 0.0041177 0.0004851 0.8036228 0.0004442 0.1913302 0.0000675 0.0038470 0.0000000
0.0000006
 0.0002026 0.0004851 0.2375740 0.1661414 0.3790300 0.0208773 0.0000000 0.0004442 0.0000000
0.1913302
 0.0000675 0.0001524 0.0000244 0.0036702 0.0000000 0.0000000 0.0000000 0.0000006 0.0000000
0.0000000
 0.0000000 0.0001078 0.0000002 0.0000945 0.0000000 0.0000000 0.0004851
yayv96                     9    0    0    0.0000000000E+00
 0.0036257 0.9962711 0.0001032 0.0000700 0.0035556 0.0000000 0.0300457 0.1882592 0.3047900
0.3989723
 0.0742039 0.0001032 0.0000143 0.0000039 0.0000176 0.0000342 0.0000000 0.0027081 0.0008476
0.0000000
 0.0300457 0.0033835 0.1845857 0.0002901 0.3047900 0.0000000 0.0657468 0.1631339 0.0000000
0.1700915
 0.0035868 0.0494691 0.0005209 0.0206271 0.0001032 0.0000000 0.0000000
ybmn93                    10    0    0    0.0000000000E+00
 0.0028582 0.9948632 0.0022787 0.0000264 0.0028318 0.0000000 0.0643711 0.0703569 0.0000000
0.8499980
 0.0101372 0.0022787 0.0000211 0.0000000 0.0000000 0.0000053 0.0000000 0.0027891 0.0000427
0.0000000
 0.0643711 0.0014115 0.0689423 0.0000031 0.0000000 0.0000000 0.0003242 0.0741113 0.0000000
0.7755625
 0.0100921 0.0000005 0.0000430 0.0000016 0.0022787 0.0000000 0.0000000
ybov93                    11    0    0    0.0000000000E+00
 0.0004830 0.9995114 0.0000056 0.0000001 0.0004828 0.0000000 0.3200169 0.0608319 0.0000079
0.6110013
 0.0076534 0.0000056 0.0000001 0.0000000 0.0000000 0.0000000 0.0000000 0.0001819 0.0003009
0.0000000
 0.3200169 0.0027158 0.0581154 0.0000006 0.0000000 0.0000079 0.0142041 0.0000001 0.0000000
0.5967972
 0.0076501 0.0000004 0.0000000 0.0000030 0.0000056 0.0000000 0.0000000
ycmn93                    12    0    0    0.0000000000E+00
 0.6276269 0.3722952 0.0000779 0.6254919 0.0021350 0.0000000 0.0000000 0.1247069 0.0000001
0.0000000
 0.2475882 0.0000779 0.0000000 0.0063093 0.0039888 0.6151938 0.0000000 0.0021344 0.0000006
0.0000000
 0.0000000 0.0327894 0.0471923 0.0447252 0.0000001 0.0000000 0.0000000 0.0000000 0.0000000
0.0000000
 0.0017445 0.1260408 0.0003762 0.1194267 0.0000000 0.0000052 0.0000726
yelv96                    13    0    0    0.0000000000E+00
 0.8512271 0.1472201 0.0015528 0.8022863 0.0463831 0.0025577 0.0060625 0.0902547 0.0055326
0.0022342
 0.0431360 0.0015528 0.1375673 0.1117740 0.3211936 0.2317514 0.0000000 0.0458684 0.0005147
0.0025577
 0.0060625 0.0197115 0.0603486 0.0101947 0.0055326 0.0000000 0.0000000 0.0000000 0.0022342
0.0000000
 0.0000026 0.0297727 0.0005283 0.0128325 0.0015528 0.0000000 0.0000000
```

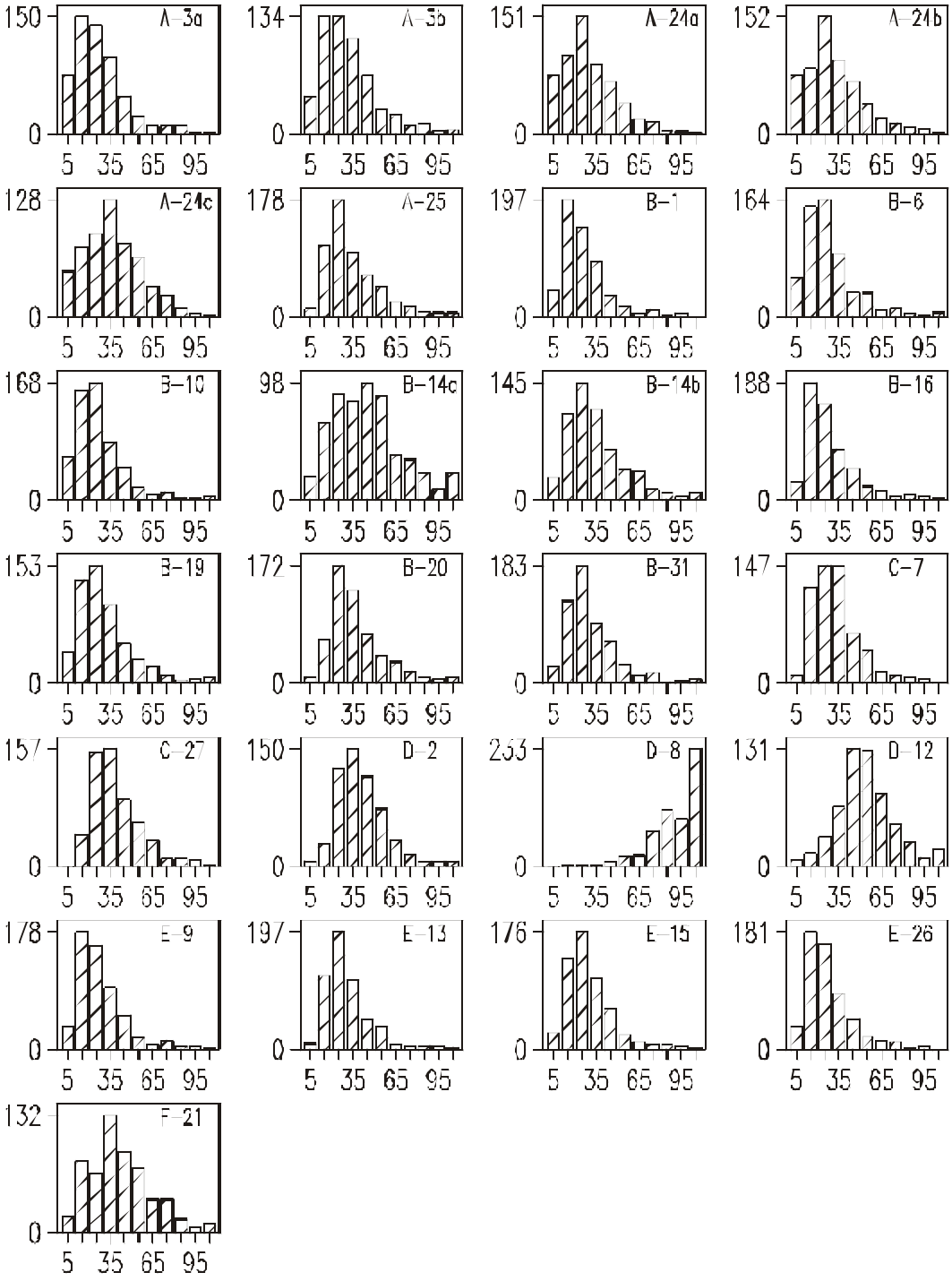
```
Beken Model II (zie PGO-2001-05)
 Indeling info **********
  3  (#aantal niveaus in hierarchie)
  3  (#niveau 1,  labels:)
         A         B         C
  9  (#niveau 2,  labels:)
        A1        A2        A3        B1        B2        B3        B4        B5
        C1
 25  (#niveau 3,  labels:)
       24a       24b       24c         7        27        3a        3b        21
         6         1        10        15        25        20       14a       14b
        16        19        31         9        13        26         2         8
        12
 :
 Predictor info **********
 17  (#predictoren)
 0  0.000000000E+00  0.000000000E+00  1.000000000E+00 -9.900000000E+01 -9.900000000E+01
meander     1
 0  0.000000000E+00  0.000000000E+00  1.000000000E+00 -9.900000000E+01 -9.900000000E+01
dwarsnat     2
 0  0.000000000E+00  0.000000000E+00  1.000000000E+00 -9.900000000E+01 -9.900000000E+01
perman     3
 0  0.000000000E+00  0.000000000E+00  1.000000000E+00 -9.900000000E+01 -9.900000000E+01
stuw     4
 0  0.000000000E+00  0.000000000E+00  1.000000000E+00 -9.900000000E+01 -9.900000000E+01
winter     5
 2  0.500000000E+00  0.000000000E+00  1.000000000E+00  0.000000000E+00  1.000000000E+02
schaduw     6
 2  0.210526316E+00  0.000000000E+00  1.000000000E+00  0.000000000E+00  1.000000000E+02
subslib     7
 1  0.000000000E+00  0.000000000E+00  1.000000000E+00  0.100000000E-01  4.000000000E+00
diepte     8
 1  0.100000000E-02  0.000000000E+00  1.000000000E+00  0.000000000E+00  1.100000000E+00
strmsn     9
 1  0.200000000E-01  0.000000000E+00  1.000000000E+00  0.000000000E+00  4.000000000E+01
breedte    10
 2  0.200000000E+00  0.000000000E+00  1.000000000E+00  0.000000000E+00  1.000000000E+02
subzand    11
 0  0.000000000E+00  0.000000000E+00  1.000000000E+00  4.300000000E+00  9.000000000E+00
phmed    12
 1  0.000000000E+00  0.000000000E+00  1.000000000E+00  7.000000000E+00  2.320000000E+02
clmed    13
 1  0.000000000E+00  0.000000000E+00  1.000000000E+00  0.100000000E-01  3.350000000E+01
nh490    14
 1  0.000000000E+00  0.000000000E+00  1.000000000E+00  0.120000000E+00  4.250000000E+01
nkjel90    15
 1  0.000000000E+00  0.000000000E+00  1.000000000E+00  0.190000000E-01  1.010000000E+01
ptot90    16
 1  0.100000000E-02  0.000000000E+00  1.000000000E+00  0.000000000E+00  4.800000000E+01
no310    17
 :
 Hoofdgroepen info **********
  3  (#hoofdgroepen)
  6  (#predictoren volledig)
  0  (#predictoren partieel)
  1  2  3  (hoofdgroep nummers)
   0 -0.21968581E+002 -0.29659049E+002  0.00000000E+000
   2 -0.23231069E+001 -0.43709873E+001  0.00000000E+000
  10 -0.19599283E+001  0.10185655E-001  0.00000000E+000
  12  0.60314763E+001  0.64932392E+001  0.00000000E+000
  13 -0.42596779E+001 -0.22477646E+001  0.00000000E+000
  14 -0.54335233E+000  0.10301003E+001  0.00000000E+000
  17  0.13700539E+001  0.31154452E+000  0.00000000E+000
 :
 Hoofdgroep  1  info **********
  3  (#groepen)
  6  (#predictoren volledig)
  0  (#predictoren partieel)
  1  2  3  (groep nummers)
   0  0.19174156E+003  0.16813089E+003  0.00000000E+000
   2 -0.54859507E+002 -0.57220865E+002  0.00000000E+000
   3 -0.82137979E+002 -0.76112151E+002  0.00000000E+000
   5  0.80287400E+001  0.14391925E+001  0.00000000E+000
  10  0.28217736E+001  0.53864314E+001  0.00000000E+000
  11  0.63568133E+000  0.10513738E+001  0.00000000E+000
  12 -0.67979316E+001 -0.43248227E+001  0.00000000E+000
 :
```

```
 Hoofdgroep  2  info **********
  5  (#groepen)
  9  (#predictoren volledig)
  0  (#predictoren partieel)
  4  5  6  7  8  (groep nummers)
   0 -0.16762216E+002  0.57016373E+000 -0.75193318E+001 -0.69189609E+002  0.00000000E+000
   1  0.28117575E+000 -0.76710442E+000  0.52486444E+001  0.68429553E+000  0.00000000E+000
   2 -0.15612081E+001 -0.23761734E+000 -0.17348060E+001  0.35512895E+000  0.00000000E+000
   3  0.17559392E+002  0.23978784E+001  0.83840234E+001  0.65986041E+002  0.00000000E+000
   6  0.21788502E+000  0.35306274E+000 -0.16982049E+000  0.22037517E+000  0.00000000E+000
   8  0.16274077E+001  0.90893333E+000  0.14158716E+001  0.14113489E+001  0.00000000E+000
   9 -0.47116672E+000 -0.34952484E+000  0.27825306E+001 -0.28452776E+000  0.00000000E+000
  10  0.18106500E+000 -0.59940408E+000 -0.33630041E-002  0.21959953E+001  0.00000000E+000
  16 -0.10159789E+001 -0.67140884E-001  0.43973321E+000 -0.96552796E+000  0.00000000E+000
  17 -0.68565930E+000 -0.34524685E+000  0.17623566E-001 -0.34027284E+000  0.00000000E+000
 :
 Hoofdgroep  3  info **********
  1  (#groepen)
  0  (#predictoren volledig)
  0  (#predictoren partieel)
  9  (groep nummers)
   0  0.00000000E+000
 :
 Groep  1  info **********
5  (#cenotypen)
  7  (#predictoren volledig)
  0  (#predictoren partieel)
  1  2  3  4  5  (cenotype nummers)
   0 -0.10379175E+003  0.94041463E+000 -0.10254832E+001 -0.52249630E+001  0.00000000E+000
   3  0.80092296E+002 -0.24696173E+002 -0.26139425E+002 -0.29925588E+002  0.00000000E+000
   4  0.44164462E+004  0.44082223E+004  0.29897371E+004  0.44150071E+004  0.00000000E+000
   6  0.14680195E+002  0.14521269E+002  0.14092540E+002  0.14339796E+002  0.00000000E+000
   7  0.17818503E+002  0.17610407E+002  0.16618181E+002  0.17428294E+002  0.00000000E+000
  13  0.36733356E+002  0.36207608E+002  0.36312677E+002  0.39219422E+002  0.00000000E+000
  14 -0.10188054E+002 -0.10234446E+002 -0.10527814E+002 -0.86837280E+001  0.00000000E+000
  16  0.16914622E+001  0.12101386E+001  0.20013175E+001  0.46285297E+000  0.00000000E+000
 :
 Groep  2  info **********
  2  (#cenotypen)
  4  (#predictoren volledig)
  0  (#predictoren partieel)
  6  7  (cenotype nummers)
   0 -0.93388486E+000  0.00000000E+000
   4 -0.15513647E+002  0.00000000E+000
   8  0.31718984E+001  0.00000000E+000
   9 -0.39500856E+001  0.00000000E+000
  16 -0.32558120E+001  0.00000000E+000
 :
 Groep  3  info **********
  1  (#cenotypen)
  0  (#predictoren volledig)
  0  (#predictoren partieel)
  8  (cenotype nummers)
   0  0.00000000E+000
 :
 Groep  4  info **********
  1  (#cenotypen)
  0  (#predictoren volledig)
  0  (#predictoren partieel)
  9  (cenotype nummers)
   0  0.00000000E+000
 :
 Groep  5  info **********
  3  (#cenotypen)
  4  (#predictoren volledig)
  0  (#predictoren partieel)
 10 11 12  (cenotype nummers)
   0  0.35746356E+001  0.50900533E+001  0.00000000E+000
  10  0.23690142E+001  0.33506650E+001  0.00000000E+000
  14  0.13344491E+001  0.25541377E+001  0.00000000E+000
  15 -0.42435541E+001 -0.48130541E+001  0.00000000E+000
  17  0.62385340E+000  0.81810390E+000  0.00000000E+000
 :
 Groep  6  info **********
  2  (#cenotypen)
  2  (#predictoren volledig)
  0  (#predictoren partieel)
 13 14  (cenotype nummers)
   0 -0.92152460E+002  0.00000000E+000
  15  0.17423205E+002  0.00000000E+000
  17  0.63488755E+002  0.00000000E+000
```

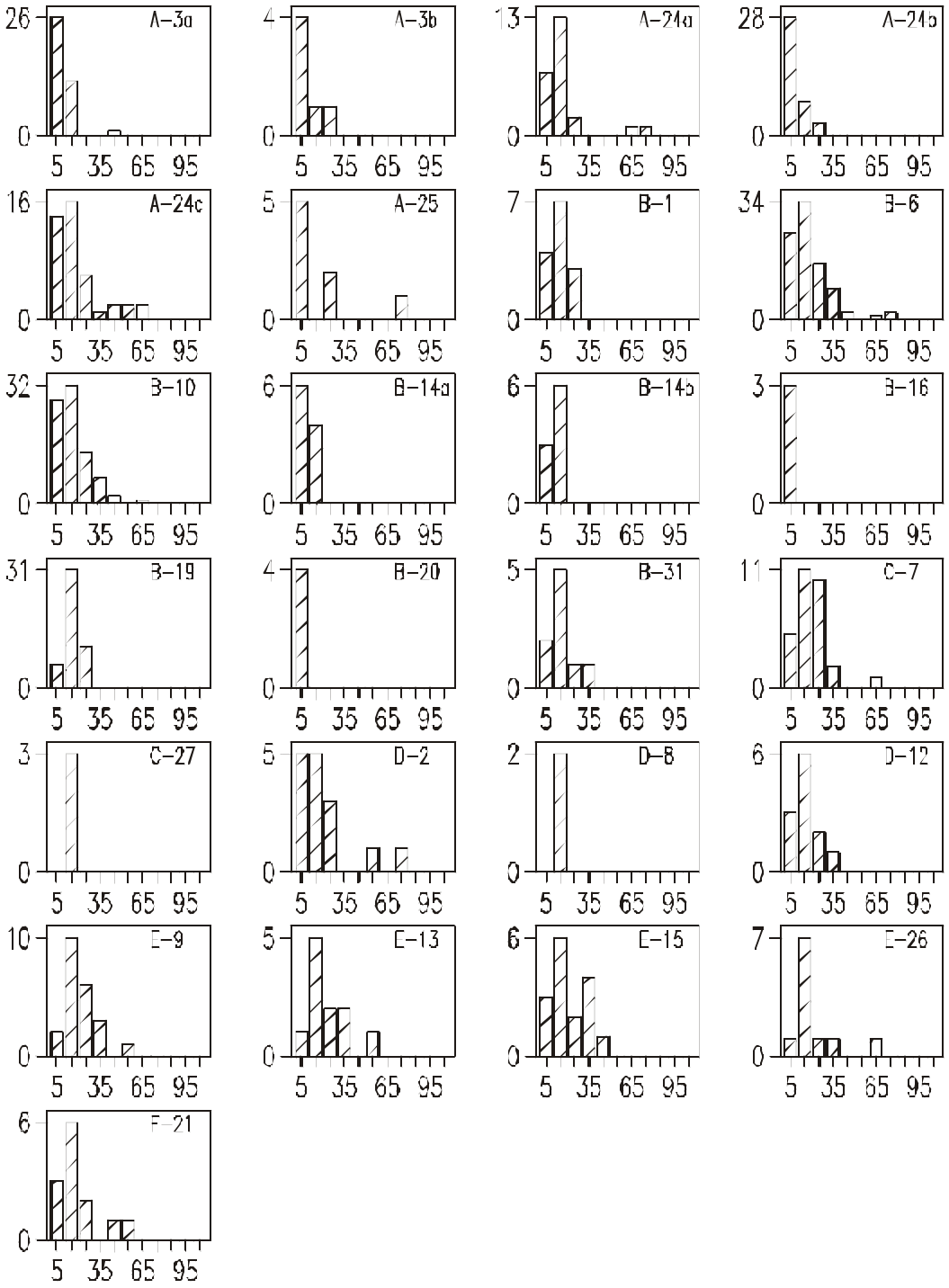
```
 :
 Groep  7  info **********
  4  (#cenotypen)
  6  (#predictoren volledig)
  0  (#predictoren partieel)
 15 16 17 18  (cenotype nummers)
   0 -0.83609619E+002  0.30950859E+002  0.42818511E+004  0.00000000E+000
   8  0.52037314E+001 -0.85528534E+001  0.99272439E+002  0.00000000E+000
   9  0.83148054E+000 -0.14681719E+001  0.13285471E+003  0.00000000E+000
  10 -0.79422059E+000  0.39771455E+001 -0.25236922E+003  0.00000000E+000
  12  0.85757520E+001 -0.18842370E+001 -0.60132142E+003  0.00000000E+000
  13  0.44098046E+001 -0.96059856E+001  0.18300266E+003  0.00000000E+000
  14  0.31394301E+001  0.73175478E+001  0.51608803E+002  0.00000000E+000
 :
 Groep  8  info **********
  4  (#cenotypen)
  5  (#predictoren volledig)
  0  (#predictoren partieel)
 19 20 21 22  (cenotype nummers)
   0  0.38566248E+002  0.35347351E+001  0.66458123E+002  0.00000000E+000
   2 -0.31915619E+000 -0.48848283E+000  0.47517075E+001  0.00000000E+000
   7  0.46588200E+000 -0.82164617E-001  0.87372813E+000  0.00000000E+000
  10  0.68472992E+001 -0.79552021E+000  0.20100823E+001  0.00000000E+000
  12 -0.63159159E+001 -0.32928053E+000 -0.97972511E+001  0.00000000E+000
  14 -0.19442055E+001  0.12850017E+001  0.11009122E+001  0.00000000E+000
 :
 Groep  9  info **********
  3  (#cenotypen)
  4  (#predictoren volledig)
  0  (#predictoren partieel)
23 24 25  (cenotype nummers)
   0  0.36697255E+002  0.81497032E+002  0.00000000E+000
   2  0.12523353E+003  0.86236752E+002  0.00000000E+000
   3  0.10417125E+003 -0.95113247E+001  0.00000000E+000
   7  0.19248089E+002  0.19207923E+002  0.00000000E+000
   8 0.73199694E+002  0.32462118E+002  0.00000000E+000
 :
```

File naam	Grootte	Datum	Tijd
AArun.bat	223	06-11-01	10:41
Beken-1.in	60.402	05-11-01	14:16
Beken-1.out	70.220	06-11-01	10:41
Beken-1.unf	34.116	05-11-01	14:17
Beken-2.in	57.324	05-11-01	14:17
Beken-2.out	80.482	06-11-01	10:41
Beken-2.unf	316.928	05-11-01	14:18
Scenario.exe	405.808	05-11-01	13:29
Sloten-1.in	118.816	05-11-01	14:19
Sloten-1.out	88.234	06-11-01	10:41
Sloten-1.unf	34.116	05-11-01	14:19
Sloten-2.in	118.816	05-11-01	14:19
Sloten-2.out	85.192	06-11-01	10:41
Sloten-2.unf	34.116	05-11-01	14:20

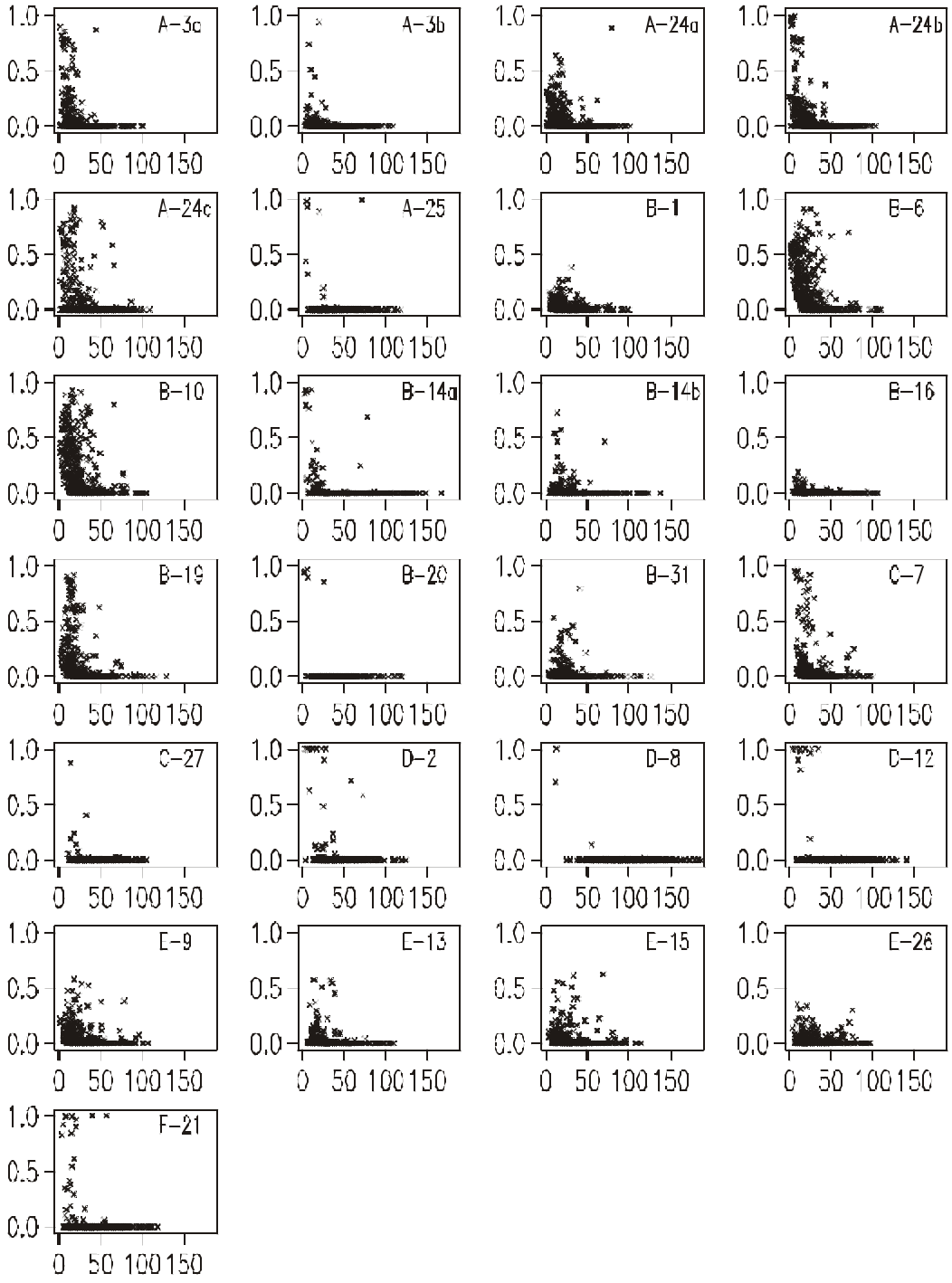
Bijlage 4 Beken Model I: alle afstanden ten opzichte van het cenotype



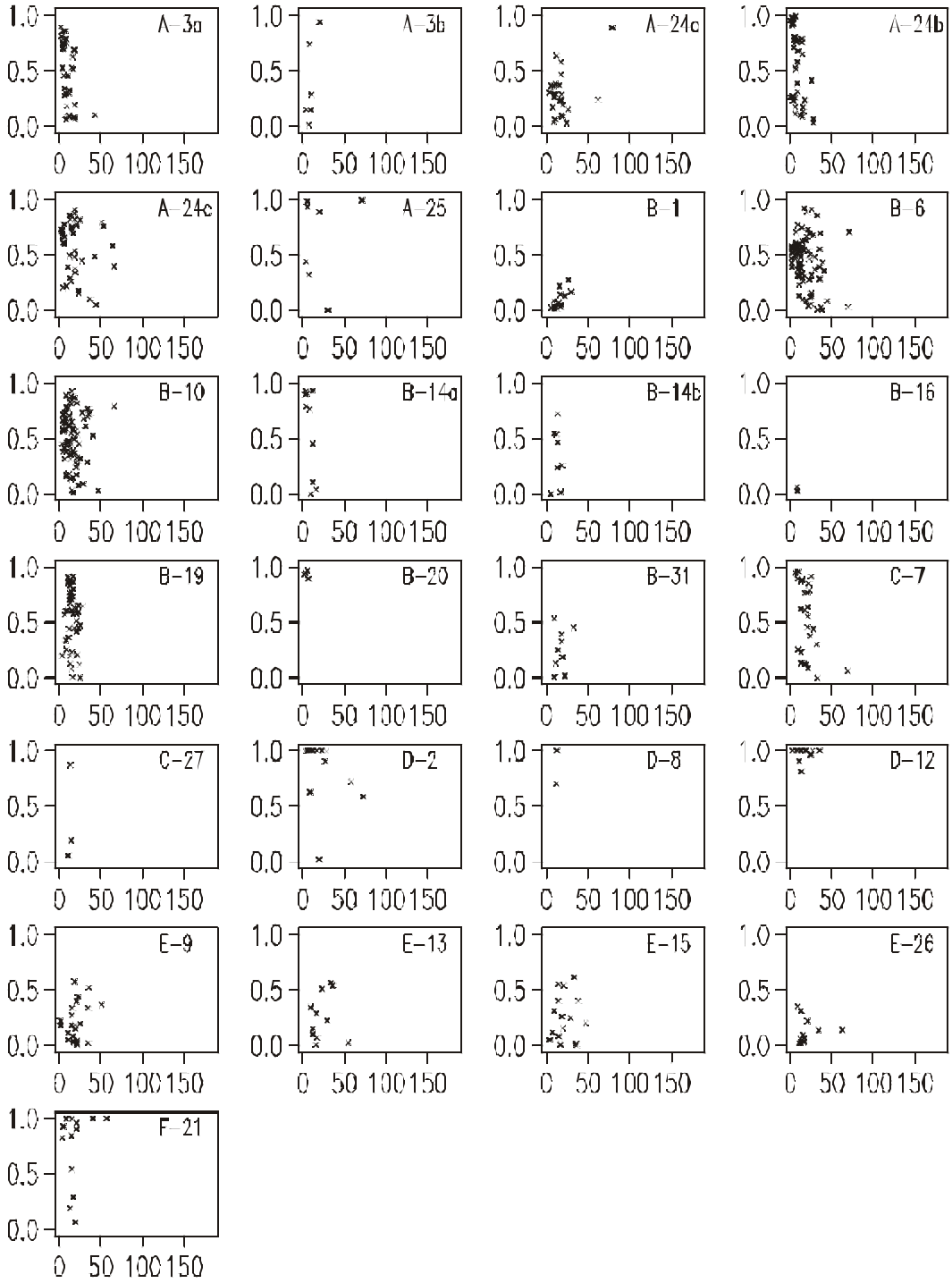
Bijlage 5 Beken Model I: afstanden ten opzichte van het eigen cenotype.



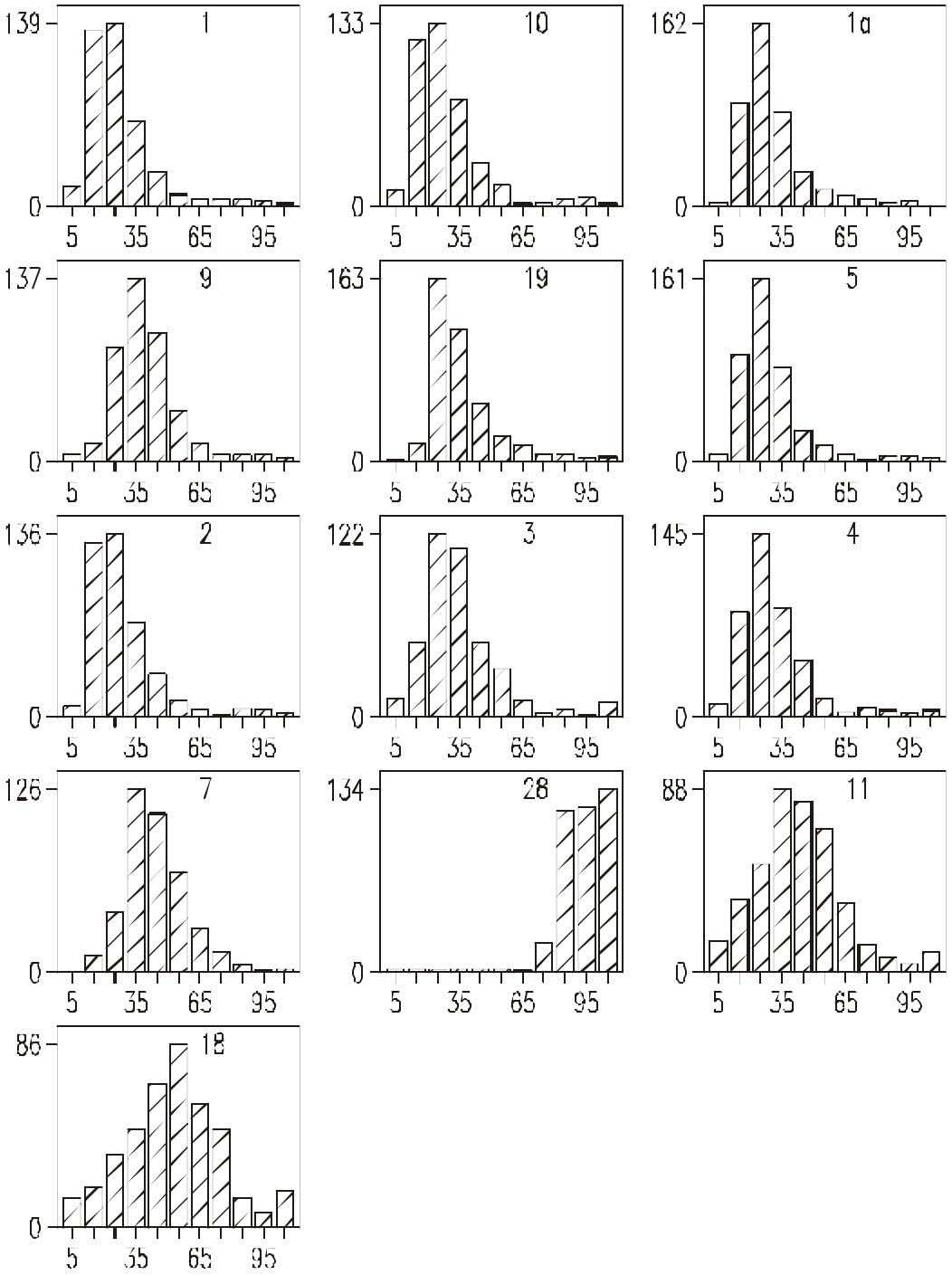
Bijlage 6 Beken Model I: voorspelkans versus afstand, alle data.



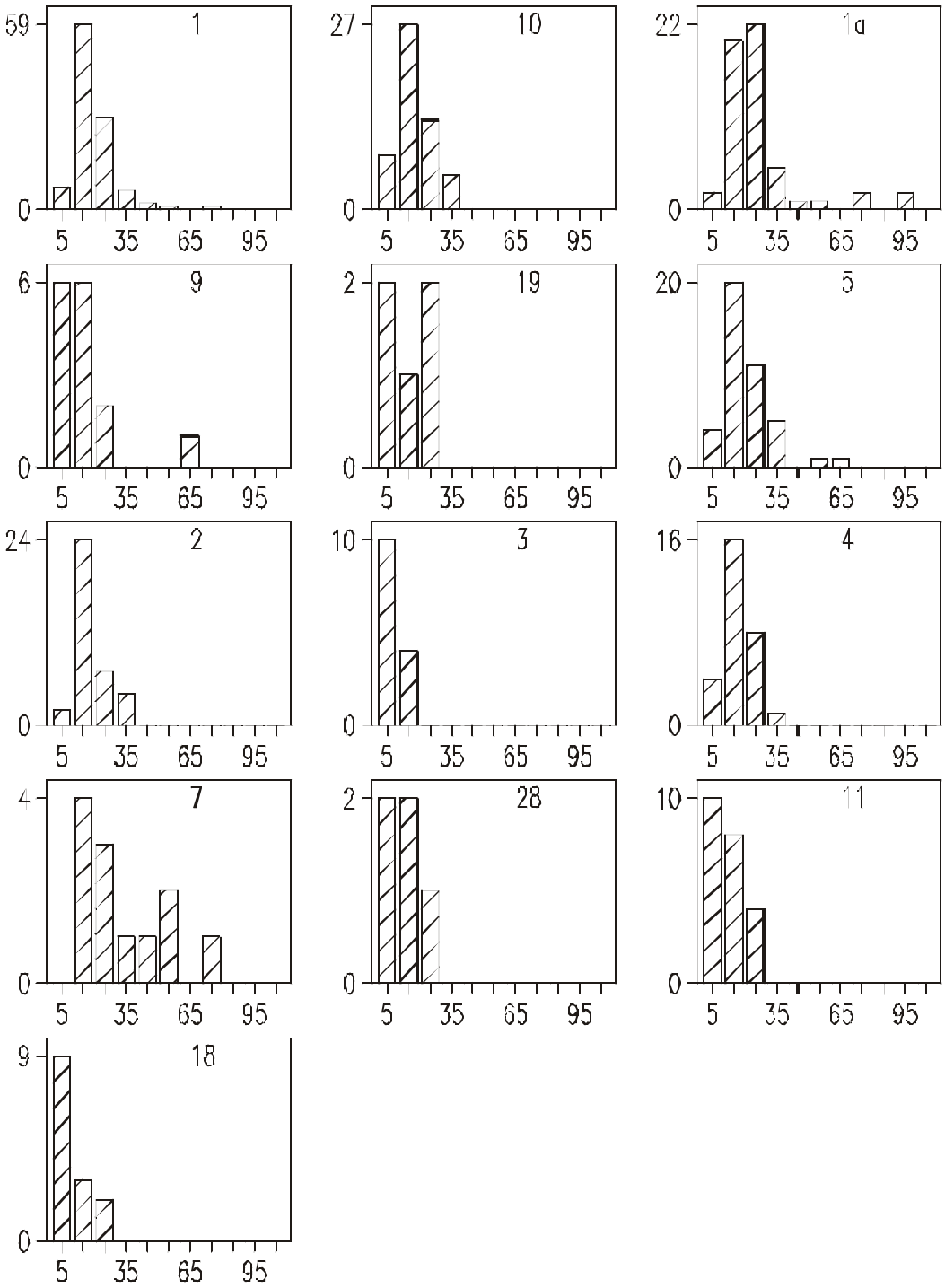
Bijlage 7 Beken Model I: voorspelkans versus afstand, cenotype specifiek.



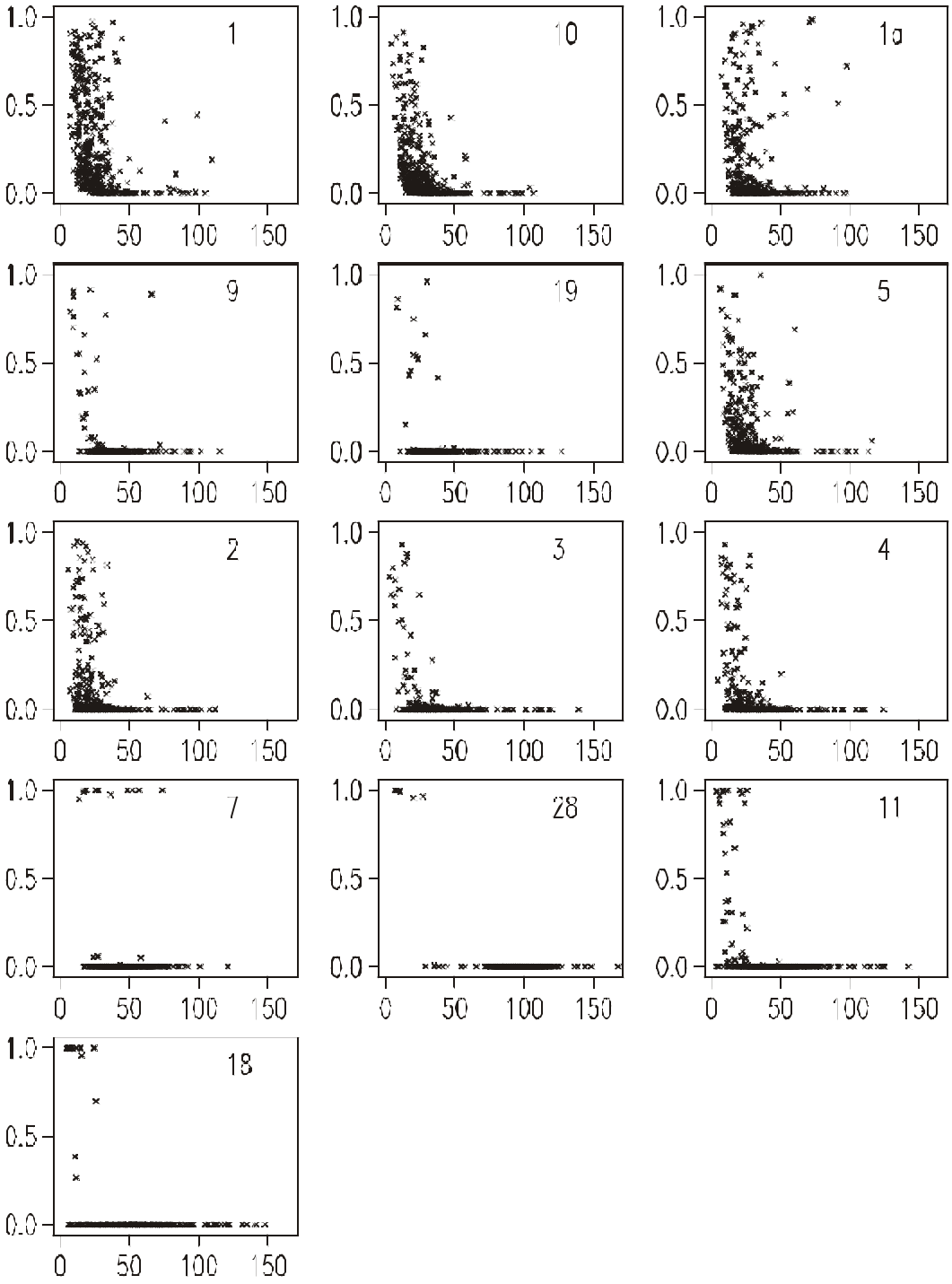
Bijlage 8 Sloten Model II: alle afstanden ten opzichte van de cenotypen.



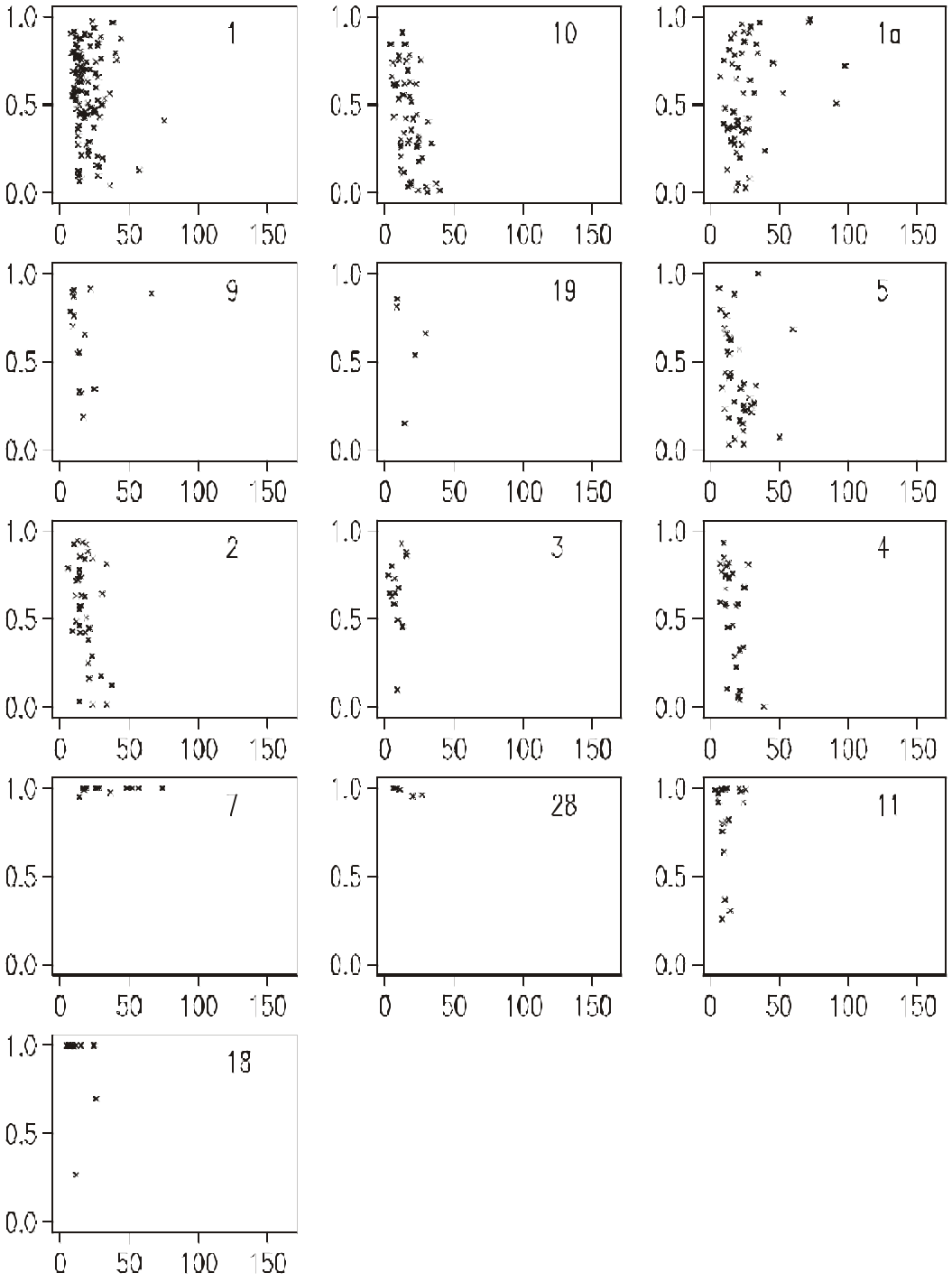
Bijlage 9 Sloten Model II: afstanden ten opzichte van het eigen cenotype.



Bijlage 10 Sloten Model II: voorspelkans versus afstand, alle data.



Bijlage 11 Sloten Model II: voorspelkans versus afstand, cenotype specifiek.



Bijlage 12 Indices geselecteerd voor toetsing op bruikbaarheid voor de beoordeling van Nederlandse beeksystemen

Kwalitatieve soortenrijkdom

- Porifera
- Coelenterata
- Cestoda
- Trematoda
- Turbellaria
- Nematoda
- Nematomorpha
- Gastropoda
- Bivalvia
- Polychaeta
- Oligochaeta (OL)
- Hirudinea
- Crustacea
- Araneae
- Ephemeroptera
- Odonata
- Plecoptera
- Heteroptera
- Planipennia
- Megaloptera
- Trichoptera
- Lepidoptera
- Coleoptera
- Diptera
- Bryozoa
- EPT-taxa (Ephemeroptera, Plecoptera en Trichoptera)
- Diptera + Oligochaeta
- ratio Diptera:Oligochaeta
- ratio EPT-taxa:Oligochaeta
- ratio EPT-taxa:Diptera
- ratio EPT-taxa:Diptera + Oligochaeta
- aantal taxa
- aantal families

Kwantitatieve soortenrijkdom

- Porifera
- Coelenterata
- Cestoda
- Trematoda
- Turbellaria
- Nematoda
- Nematomorpha
- Gastropoda

- Bivalvia
- Polychaeta
- Oligochaeta
- Hirudinea
- Crustacea
- Araneae
- Ephemeroptera
- Odonata
- Plecoptera
- Heteroptera
- Planipennia
- Megaloptera
- Trichoptera
- Lepidoptera
- Coleoptera
- Diptera
- Bryozoa
- EPT-taxa
- Diptera + Oligochaeta

- totaal aantal individuen
- ratio Diptera individuen:Oligochaeta individuen
- ratio EPT-taxa individuen:Oligochaeta individuen
- ratio EPT-taxa individuen:Diptera individuen
- ratio EPT-taxa individuen:Diptera + Oligochaeta individuen

Diversiteitsindices
- Simpson-index
- Shannon-Wiener-index
- evenness-index

Milieu-indices
a. saprobie
- Saprobie-index van Zelinka en Marvan
- percentage xeno-saprobe soorten
- percentage oligo-saprobe soorten
- percentage beta-meso-saprobe soorten
- percentage alpha-meso-saprobe soorten
- percentage poly-saprobe soorten
- Duitse Saprobie-index
- Nederlandse Saprobie-index

b. stroming (percentage individuen van de totale levensgemeenschap die een bepaalde stroming prefereren):
- limnobiont (alleen in stilstaand water)
- limnofiel (bij voorkeur in stilstaand water, zelden gevonden in langzaam stromende beken)

- limno- tot rheofiel (bij voorkeur in stromend water, regelmatig gevonden in langzaam stromende beken)
- rheo- tot limnofiel (bij voorkeur langzaam stromende beken met lentische zones, maar wordt ook gevonden in stilstaand water)
- rheofiel (in beken, bij voorkeur met gemiddelde tot hoge stroomsnelheid)
- rheobiont (gebonden aan beken met hoge stroomsnelheden)
- indifferent (geen voorkeur voor een bepaalde stroomsnelheid)

c. microhabitat (percentage individuen van de totale levensgemeenschap die een bepaald microhabitat prefereren):
- pelal (modder; korrelgrootte < 0.063mm)
- argyllal (silt,leem, klei: korrelgrootte < 0.063 mm)
- psammal (zand; korrelgrootte 0.063-2 mm
- akal (fijn tot middelgroot grind; korrelgrootte 0.2-2 cm)
- lithal (grof grind, stenen, keien; korrelgrootte > 2 cm)
- phytal (algen, mos en macrofyten inclusief levende delen van terrestrische planten)
- particulair organisch materiaal (zoals hout, CPOM, FPOM)

Biotische indices
- BMWP Score (Biological Monitoring Working Party)
- ASPT (Average Score per Taxon)
- Spaanse BMWP Score
- Danish Stream Fauna Index
- Belgian Biotic Index
- IBE (Indice Biotico Esteso)
- MAS (Mayfly Average Score)
- MAS (gote rivier)

Functionele en procesindices
a. functionele voedingsgroepen (percentage individuen van de totale levensgemeenschap):
- schrapers
- mineerders
- xylophagen
- knippers
- vergaarders
- actieve filtreerders
- passieve filtreerders
- predatoren
- parasieten
- ratio aantal individuen (grazers + schrapers):aantal individuen (vergaarders + filtreerders)
- Rhithron Feeding Type Index

b. bewegingsgedrag (percentage individuen van de totale levensgemeenschap):
- zwemmers/schaatsers
- zwemmers/duikers

- gravers / boorders
- spartelaars / wandelaars
- (semi)sessiel

c. zonatie (percentage individuen van de totale levensgemeenschap die een bepaald beek- of riviertraject prefereren):

- crenal (bron)
- hypocrenal (bron-beek)
- epirhithral (bovenste zone van de beekforel)
- metarhithral (onderste zone van de beekforel)
- hypohithral (vlagzalmzone)
- epipotamal (barbeelzone)
- metapotamal (brasemzone)
- hypopotamal (brakwaterzone)
- litoraal
- profundaal

Bijlage 13 Overzicht significantie en overlap indexwaarden tussen kwaliteitsklassen AQEM typen voor de verschillende indices

Langzaam stromende beken

Indices	2-3	2-4	2-5	3-4	3-5	4-5
Duitse Saprobie-index	x	x	x	x		x
xeno [%]	x	x	x			
Saprobie-index (Zelinka en Marvan)	x	x	x	x		x
oligo [%]						
poly [%]	x	x	x	x		x
alpha-meso [%]	x	x	x	x	x	x
Nederlandse Saprobie-index		x		x		x
beta-meso [%]	x	x	x	x	x	x
ASPT	x	x	x	x	x	
BMWP	x	x		x	x	x
BMWP Spain	x	x		x	x	x
IBE		x	x	x		x
BBI		x	x	x	x	x
DSFI	x	x	x	x	x	x
MAS	x	x	x			x
MAS (grote rivier)	x	x	x			x
hypocrenal [%]	x	x	x			
epirhithral [%]	x	x	x			
crenal [%]	x	x	x			
metapotamal [%]	x	x	x		x	x
hypopotamal [%]	x	x	x	x	x	x
metarhithral [%]	x	x	x	x	x	
profundaal [%]		x	x	x	x	x
hyporhithral [%]	x	x	x	x	x	
epipotamal [%]	x	x	x	x		x
litoraal [%]	x	x	x		x	x
RETI	x	x	x	x		x
knippers [%]	x	x	x	x	x	
grazers+schrapers/vergaarders+filtreerders [%]	x		x	x	x	x
grazers en schrapers [%]		x	x	x		
actieve filtreerders [%]	x	x	x		x	x
passieve filtreerders [%]		x	x		x	
mineerders %]	x		x		x	x
vergaarders [%]	x	x	x			x
xylophagen [%]	x	x	x			
predatoren [%]	x			x	x	
parasieten [%]			x			
type Pel [%]	x	x	x	x		
type POM [%]		x	x	x	x	
type Lit [%]	x	x	x		x	
type Aka [%]	x	x	x		x	
type Arg [%]	x				x	x
type Psa [%]		x	x	x	x	x
type Phy [%]	x		x		x	x
type RP [%]	x	x	x	x	x	
type LR [%]	x	x	x	x	x	x
type IN [%]	x	x	x	x		x
type RB [%]			x			x
type LP [%]	x	x	x			
type LB [%]						
type RL [%]		x	x	x	x	x
Trichoptera	x	x	x	x	x	
Gastropoda	x	x	x		x	
Bivalvia	x	x	x		x	x
Diptera	x		x	x	x	x
EPT-Taxa		x	x	x	x	
Hirudinea	x	x	x	x	x	x
aantal taxa	x	x	x	x		
Turbellaria	x	x	x	x	x	

Indices	2-3	2-4	2-5	3-4	3-5	4-5
EPT/Ol+Di	x	x	x	x		x
aantal families	x	x	x			x
EPT/Ol		x	x	x	x	x
Megaloptera	x		x	x		
Odonata	x		x	x	x	
EPT/Di	x	x	x	x		x
Ephemeroptera	x	x	x	x	x	x
Coleoptera	x	x	x			x
Di+Ol	x	x	x		x	x
Crustacea	x	x	x		x	
Heteroptera	x	x	x	x		x
Planipennia	x	x	x			
Lepidoptera	x		x	x		x
Di/Ol	x	x	x	x	x	x
Plecoptera	x	x	x			
Oligochaeta		x		x		x
Araneae			x		x	
abundantie		x	x	x		x
Diptera [%]	x	x			x	x
Plecoptera [%]	x	x	x			
Crustacea [%]				x	x	
Di+Ol %	x	x		x	x	x
Trichoptera [%]	x	x	x	x	x	
Bivalvia [%]			x		x	x
Turbellaria [%]	x	x	x			
Oligochaeta [%]		x		x		x
EPT-Taxa [%]		x		x		x
Megaloptera [%]				x		x
Ephemeroptera [%]	x		x		x	x
Heteroptera [%]	x		x	x	x	x
EPT/Di [%]			x		x	
EPT/Ol+Di [%]			x		x	x
EPT/Ol [%]		x		x		x
Gastropoda [%]	x		x		x	x
Coleoptera [%]		x	x	x		x
Planipennia [%]						
Hirudinea [%]	x	x	x	x	x	
Lepidoptera [%]	x		x	x		x
Odonata [%]	x			x	x	
Di/Ol [%]	x			x	x	
Araneae [%]						
gravers/boorders [%]	x	x	x	x	x	
zwemmers/duikers [%]	x	x	x			
(semi)sessiel [%]	x	x			x	x
zemmers/schaatsers [%]				x	x	
spartelaars/wandelaars [%]	x	x			x	x
diversity (Shannon-Wiener-Index)	x	x	x	x		x
evenness	x	x	x	x		x
diversity (Simpson-Index)	x	x	x	x		x

Snel stromende beken

Indices	2-3	2-4	2-5	3-4	3-5	4-5
Duitse Saprobie-index	x	x	x	x	x	x
xeno [%]	x	x	x			
Saprobie-index (Zelinka en Marvan)	x	x	x	x	x	x
oligo [%]	x	x	x	x	x	
poly [%]		x	x	x		x
alpha-meso [%]	x	x			x	x
Nederlandse Saprobie-index		x		x		x
beta-meso [%]		x	x	x	x	
ASPT	x	x	x	x	x	x
BMWP	x	x	x		x	x
BMWP Spain	x		x			x
IBE	x	x	x			
BBI			x			
DSFI		x		x		
MAS						
MAS (grote rivier)						
hypocrenal [%]		x	x			
epirhithral [%]		x	x	x	x	
crenal [%]	x	x	x			
metapotamal [%]	x	x	x	x	x	
hypopotamal [%]		x	x	x	x	x
metarhithral [%]	x	x	x	x	x	
profundaal [%]		x		x		x
hyporhithral [%]		x	x	x	x	
epipotamal [%]	x	x	x	x	x	
litoraal [%]		x	x	x	x	x
RETI		x	x	x	x	
knippers [%]	x	x	x	x	x	
grazers+schrapers/vergaarders+filtreerders [%]			x	x	x	
grazers en schrapers [%]	x			x	x	
actieve filtreerders [%]		x	x	x	x	
passieve filtreerders [%]	x			x	x	
mineerders %]			x			x
vergaarders [%]		x	x	x	x	
xylophagen [%]		x	x			
predatoren [%]						
parasieten [%]						
type Pel [%]		x	x	x	x	
type POM [%]		x	x	x	x	x
type Lit [%]		x	x	x	x	
type Aka [%]	x	x	x	x	x	
type Arg [%]		x	x	x	x	
type Psa [%]	x	x	x		x	x
type Phy [%]	x	x	x			
type RP [%]	x	x	x	x	x	
type LR [%]		x	x	x	x	x
type IN [%]		x	x	x	x	x
type RB [%]						
type LP [%]		x	x	x	x	
type LB [%]						
type RL [%]	x		x	x	x	x
Trichoptera	x	x	x	x		x
Gastropoda		x	x	x	x	x
Bivalvia		x	x	x	x	
Diptera	x	x	x		x	x
EPT-Taxa	x	x	x	x		x
Hirudinea	x	x	x	x		x
aantal taxa	x	x	x	x	x	x
Turbellaria			x		x	x
EPT/OI+Di		x	x	x	x	x
aantal families						
EPT/OI	x	x		x	x	x

Indices	2-3	2-4	2-5	3-4	3-5	4-5
Megaloptera		x	x	x	x	
Odonata		x	x		x	x
EPT/Di		x	x	x		x
Ephemeroptera	x	x	x	x		x
Coleoptera		x	x	x	x	
Di+Ol	x	x	x	x	x	
Crustacea			x			
Heteroptera		x	x	x	x	x
Planipennia						
Lepidoptera						
Di/Ol	x	x		x	x	x
Plecoptera		x	x			
Oligochaeta		x	x	x		x
Araneae		x				
abundantie						
Diptera [%]		x	x		x	x
Plecoptera [%]		x	x			
Crustacea [%]	x	x	x	x	x	
Di+Ol %		x	x	x	x	x
Trichoptera [%]	x	x		x	x	x
Bivalvia [%]		x				
Turbellaria [%]	x	x	x			
Oligochaeta [%]		x		x		x
EPT-Taxa [%]	x		x	x	x	x
Megaloptera [%]						
Ephemeroptera [%]	x	x	x	x		x
Heteroptera [%]			x		x	x
EPT/Di [%]			x			
EPT/Ol+Di [%]		x	x	x		x
EPT/Ol [%]	x	x		x	x	x
Gastropoda [%]		x	x		x	x
Coleoptera [%]						x
Planipennia [%]						
Hirudinea [%]		x	x	x		x
Lepidoptera [%]			x		x	x
Odonata [%]	x		x	x		x
Di/Ol [%]						x
Araneae [%]						
gravers/boorders [%]		x	x	x	x	x
zwemmers/duikers [%]		x	x	x	x	
(semi)sessiel [%]		x	x		x	
zemmers/schaatsers [%]						
spartelaars/wandelaars [%]	x	x	x			
diversity (Shannon-Wiener-Index)	x	x	x		x	x
evenness	x	x	x		x	x
diversity (Simpson-Index)	x	x	x		x	x

x = significant verschil tussen de betreffende klassen
overlap 25-percentiel van een klasse met het 75-percentiel van een andere klasse
overlap 25- of 75-percentiel van een klasse met de mediaan van een andere klasse

Bijlage 14 Overzicht significantie en overlap indexwaarden tussen bekentypologie kwaliteitsklassen voor de verschillende indices

Langzaam stromende beken

Indices	4-3	4-2	4-1	3-2	3-1	2-1
Duitse Saprobie-index	x	x	x	x		x
xeno [%]	x	x	x			
Saprobie-index (Zelinka en Marvan)	x	x	x	x		x
oligo [%]						
poly [%]	x	x	x	x		x
alpha-meso [%]	x	x	x	x	x	x
Nederlandse Saprobie-index		x		x		x
beta-meso [%]	x	x	x	x	x	x
ASPT	x	x	x	x	x	
BMWP	x	x		x	x	x
BMWP Spain	x	x		x	x	x
IBE		x	x	x		x
BBI		x	x	x	x	x
DSFI	x	x	x	x	x	x
MAS	x	x	x			x
MAS (grote rivier)	x	x	x			x
hypocrenal [%]	x	x	x			
epirhithral [%]	x	x	x			
crenal [%]	x	x	x			
metapotamal [%]	x	x	x		x	x
hypopotamal [%]	x	x	x	x	x	x
metarhithral [%]	x	x	x	x	x	
profundaal [%]		x	x	x	x	x
hyporhithral [%]	x	x	x	x	x	
epipotamal [%]	x	x	x	x		x
litoraal [%]	x	x	x		x	x
RETI	x	x	x	x		x
knippers [%]	x	x	x	x	x	
grazers+schrapers/vergaarders+filtreerders [%]	x		x	x	x	x
grazers en schrapers [%]		x	x	x		
actieve filtreerders [%]	x	x	x		x	x
passieve filtreerders [%]		x	x		x	
mineerders %]	x		x		x	x
vergaarders [%]	x	x	x			x
xylophagen [%]	x	x	x			
predatoren [%]	x			x	x	
parasieten [%]			x			
type Pel [%]	x	x	x	x		
type POM [%]		x	x	x	x	
type Lit [%]	x	x	x		x	
type Aka [%]	x	x	x		x	
type Arg [%]	x				x	x
type Psa [%]		x	x	x	x	x
type Phy [%]	x		x		x	x
type RP [%]	x	x	x	x	x	
type LR [%]	x	x	x	x	x	x
type IN [%]	x	x	x	x		x
type RB [%]			x			x
type LP [%]	x	x	x			
type LB [%]						
type RL [%]		x	x	x	x	x
Trichoptera	x	x	x	x	x	
Gastropoda	x	x	x		x	
Bivalvia	x	x	x		x	x
Diptera	x		x	x	x	x
EPT-Taxa		x	x	x	x	

Indices	4-3	4-2	4-1	3-2	3-1	2-1
Hirudinea	x	x	x	x	x	x
aantal taxa	x	x	x	x		
Turbellaria	x	x	x	x	x	
EPT/Ol+Di	x	x	x	x		x
aantal families	x	x	x			x
EPT/Ol		x	x	x	x	x
Megaloptera	x		x	x		
Odonata	x		x	x	x	
EPT/Di	x	x	x	x		x
Ephemeroptera	x	x	x	x	x	x
Coleoptera	x	x	x			x
Di+Ol	x	x	x		x	x
Crustacea	x	x	x		x	
Heteroptera	x	x	x	x		x
Planipennia	x	x	x			
Lepidoptera	x		x	x		x
Di/Ol	x	x	x	x	x	x
Plecoptera	x	x	x			
Oligochaeta		x		x		x
Araneae			x		x	
abundantie		x	x	x		x
Diptera [%]	x	x			x	x
Plecoptera [%]	x	x	x			
Crustacea [%]				x	x	
Di+Ol %	x	x		x	x	x
Trichoptera [%]	x	x	x	x	x	
Bivalvia [%]			x		x	x
Turbellaria [%]	x	x	x			
Oligochaeta [%]		x		x		x
EPT-Taxa [%]		x		x		x
Megaloptera [%]				x		x
Ephemeroptera [%]	x		x		x	x
Heteroptera [%]	x		x	x	x	x
EPT/Di [%]			x		x	
EPT/Ol+Di [%]			x		x	x
EPT/Ol [%]		x		x		x
Gastropoda [%]	x		x		x	x
Coleoptera [%]		x	x	x		x
Planipennia [%]						
Hirudinea [%]	x	x	x	x	x	
Lepidoptera [%]	x		x	x		x
Odonata [%]	x			x	x	
Di/Ol [%]	x			x	x	
Araneae [%]						
gravers/boorders [%]	x	x	x	x	x	
zwemmers/duikers [%]	x	x	x			
(semi)sessiel [%]	x	x			x	x
zemmers/schaatsers [%]				x	x	
spartelaars/wandelaars [%]	x	x			x	x
diversity (Shannon-Wiener-Index)	x	x	x	x		x
evenness	x	x	x	x		x
diversity (Simpson-Index)	x	x	x	x		x

Snel stromende beken

Indices	4-3	4-2	4-1	3-2	3-1	2-1
Duitse Saprobie-index	x	x	x	x	x	x
xeno [%]	x	x	x			
Saprobie-index (Zelinka en Marvan)	x	x	x	x	x	x
oligo [%]	x	x	x	x	x	
poly [%]		x	x	x		x
alpha-meso [%]	x	x			x	x
Nederlandse Saprobie-index		x		x		x
beta-meso [%]		x	x	x	x	
ASPT	x	x	x	x	x	x
BMWP	x	x	x		x	x
BMWP Spain	x		x			x
IBE	x	x	x			
BBI			x			
DSFI		x		x		
MAS						
MAS (grote rivier)						
hypocrenal [%]		x	x			
epirhithral [%]		x	x	x	x	
crenal [%]	x	x	x			
metapotamal [%]	x	x	x	x	x	
hypopotamal [%]	x	x	x	x	x	x
metarhithral [%]	x	x	x	x	x	
profundaal [%]		x		x		x
hyporhithral [%]		x	x	x	x	
epipotamal [%]	x	x	x	x	x	
litoraal [%]		x	x	x	x	x
RETI		x	x	x	x	
knippers [%]	x	x	x	x	x	
grazers+schrapers/vergaarders+filtreerders [%]			x	x	x	
grazers en schrapers [%]	x			x	x	
actieve filtreerders [%]		x	x	x	x	
passieve filtreerders [%]	x			x	x	
mineerders %]			x			x
vergaarders [%]		x	x	x	x	
xylophagen [%]		x	x			
predatoren [%]						
parasieten [%]						
type Pel [%]		x	x	x	x	
type POM [%]		x	x	x	x	x
type Lit [%]		x	x	x	x	
type Aka [%]	x	x	x	x	x	
type Arg [%]		x	x	x	x	
type Psa [%]	x	x	x		x	x
type Phy [%]	x	x	x			
type RP [%]	x	x	x	x	x	
type LR [%]		x	x	x	x	x
type IN [%]		x	x	x	x	x
type RB [%]						
type LP [%]		x	x	x	x	
type LB [%]						
type RL [%]	x		x	x	x	x
Trichoptera	x	x	x	x		x
Gastropoda		x	x	x	x	x
Bivalvia		x	x	x	x	
Diptera	x	x	x		x	x
EPT-Taxa	x	x	x	x		x
Hirudinea	x	x	x	x		x
aantal taxa	x	x	x	x	x	x
Turbellaria			x		x	x
EPT/Ol+Di		x	x	x	x	x
aantal families						

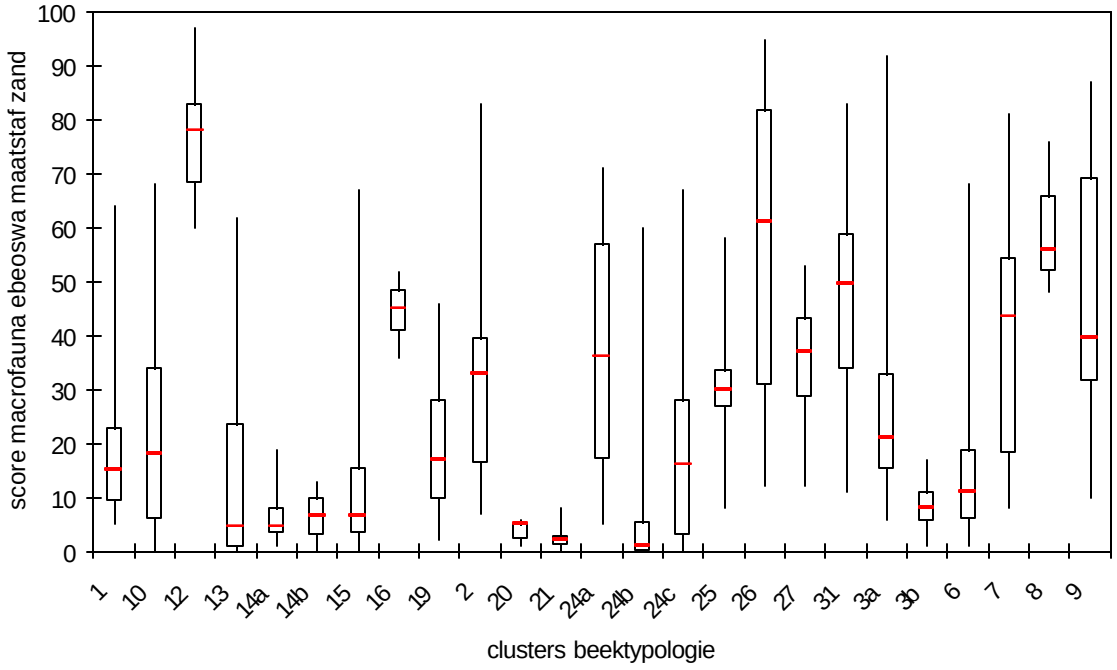
Indices	2-3	2-4	2-5	3-4	3-5	4-5
EPT/Ol	x	x		x	x	x
Megaloptera		x	x	x	x	
Odonata		x	x		x	x
EPT/Di		x	x	x		x
Ephemeroptera	x	x	x	x		x
Coleoptera		x	x	x	x	
Di+Ol	x	x	x	x	x	
Crustacea			x			
Heteroptera		x	x	x	x	x
Planipennia						
Lepidoptera						
Di/Ol	x	x		x	x	x
Plecoptera		x	x			
Oligochaeta		x	x	x		x
Araneae		x				
abundantie						
Diptera [%]		x	x		x	x
Plecoptera [%]		x	x			
Crustacea [%]	x	x	x	x	x	
Di+Ol %		x	x	x	x	x
Trichoptera [%]	x	x		x	x	x
Bivalvia [%]		x				
Turbellaria [%]	x	x	x			
Oligochaeta [%]		x		x		x
EPT-Taxa [%]	x		x	x	x	x
Megaloptera [%]						
Ephemeroptera [%]	x	x	x	x		x
Heteroptera [%]			x		x	x
EPT/Di [%]			x			
EPT/Ol+Di [%]		x	x	x		x
EPT/Ol [%]	x	x		x	x	x
Gastropoda [%]		x	x		x	x
Coleoptera [%]						x
Planipennia [%]						
Hirudinea [%]		x	x	x		x
Lepidoptera [%]			x		x	x
Odonata [%]	x		x	x		x
Di/Ol [%]						x
Araneae [%]						
gravers/boorders [%]		x	x	x	x	x
zwemmers/duikers [%]		x	x	x	x	
(semi)sessiel [%]		x	x		x	
zemmers/schaatsers [%]						
spartelaars/wandelaars [%]	x	x	x			
diversity (Shannon-Wiener-Index)	x	x	x		x	x
evenness	x	x	x		x	x
diversity (Simpson-Index)	x	x	x		x	x

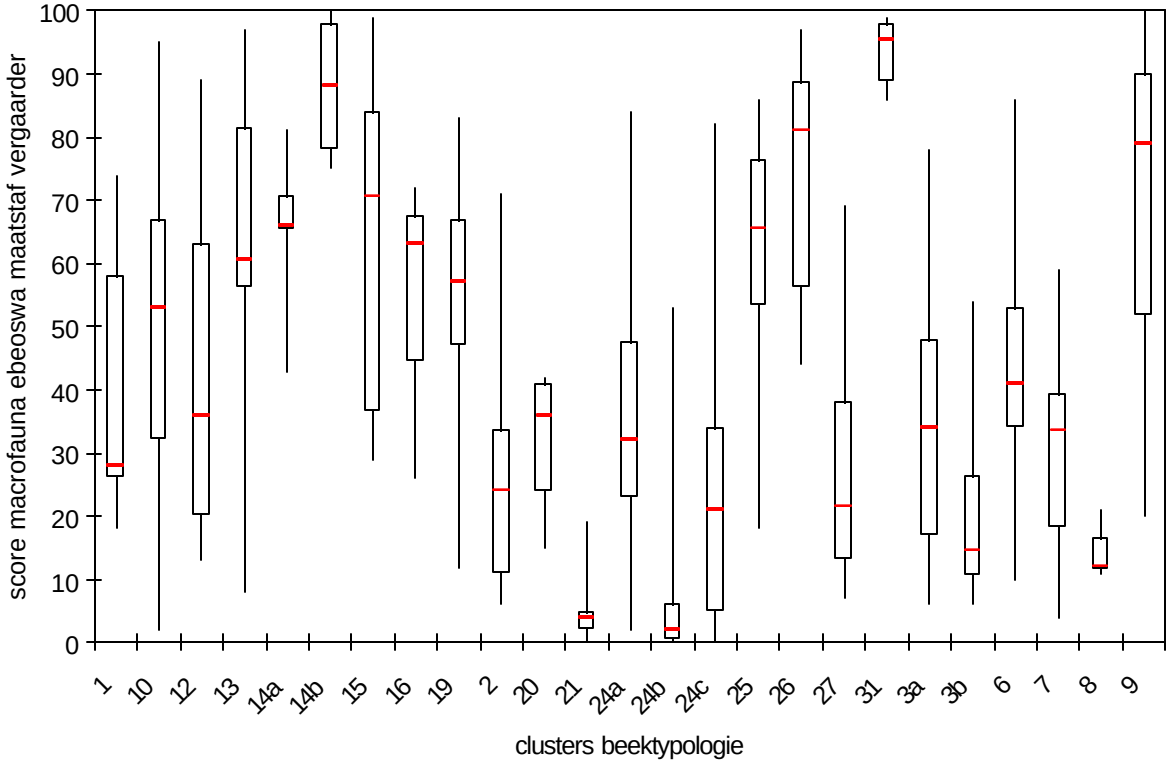
x = significant verschil tussen de betreffende klassen

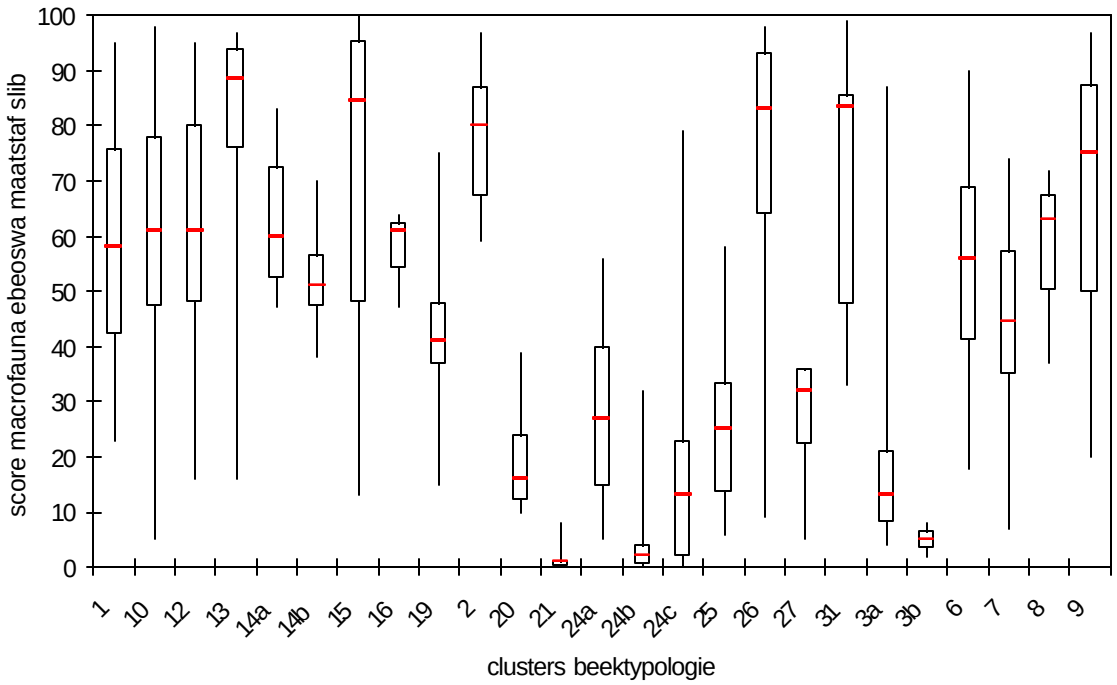
overlap 25-percentiel van een klasse met het 75-percentiel van een andere klasse

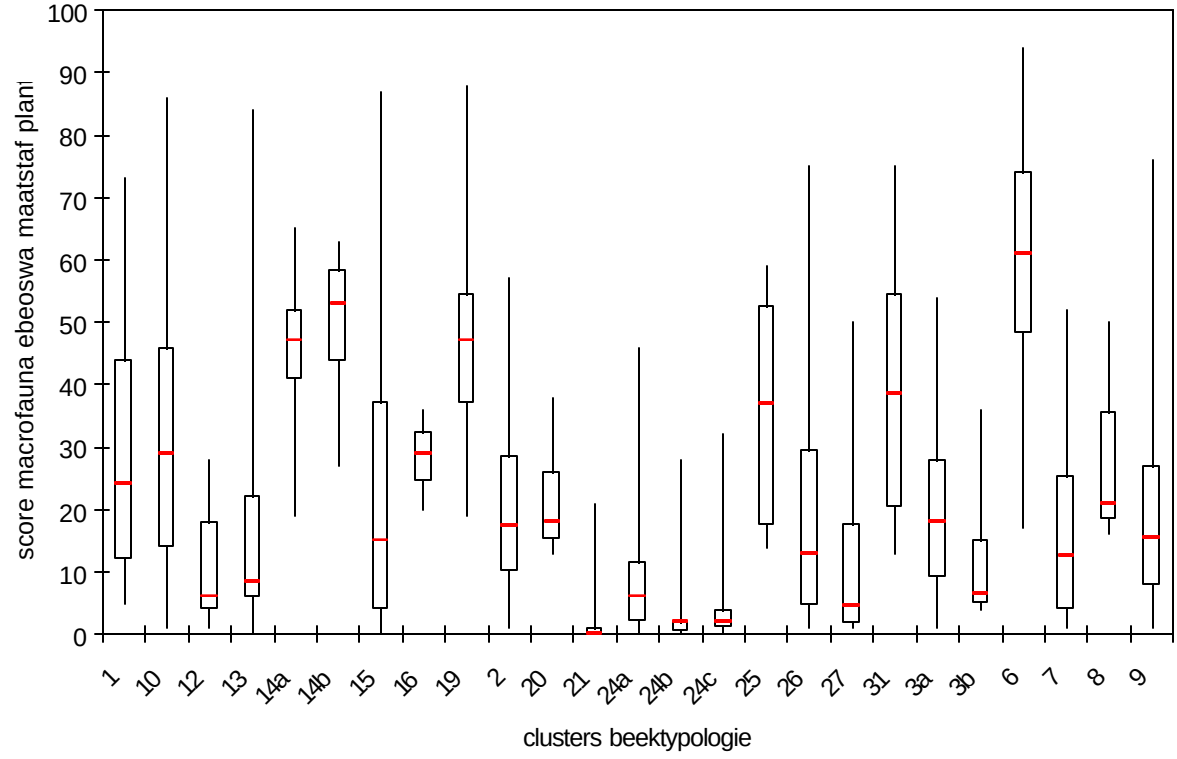
overlap 25- of 75-percentiel van een klasse met de mediaan van een andere klasse

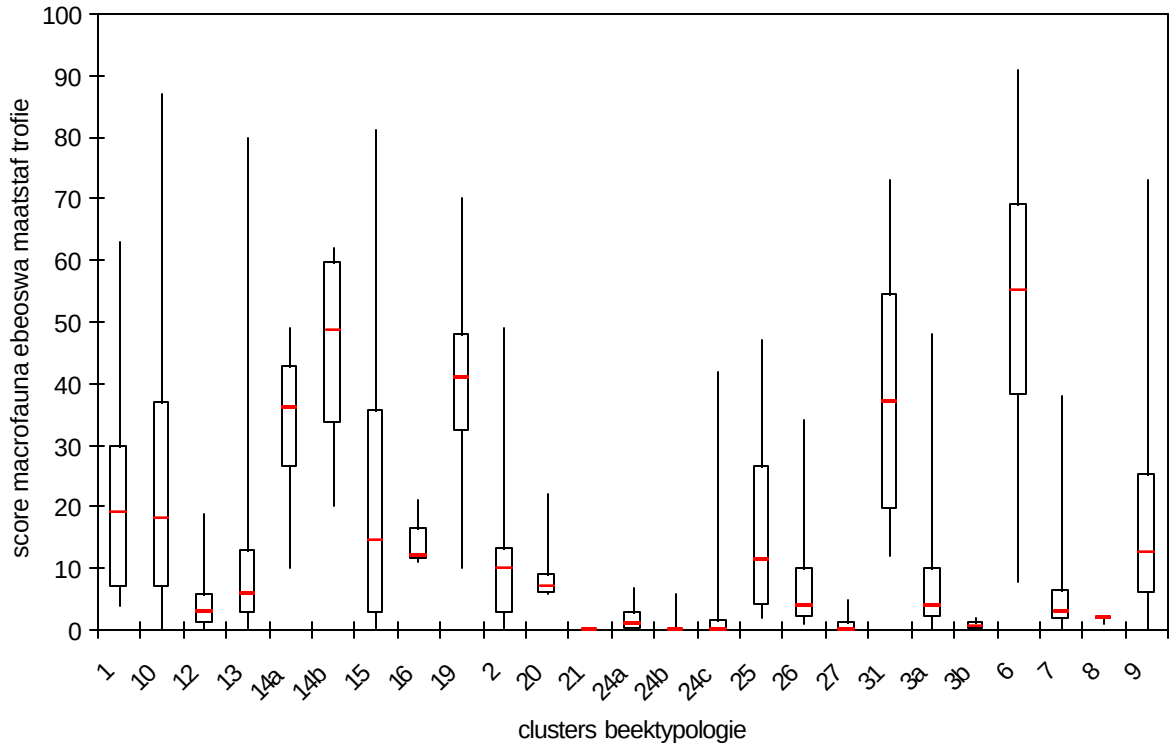
Bijlage 15 EBEOSWA en EBEOSLO figuren voor beken- en slotentypen.

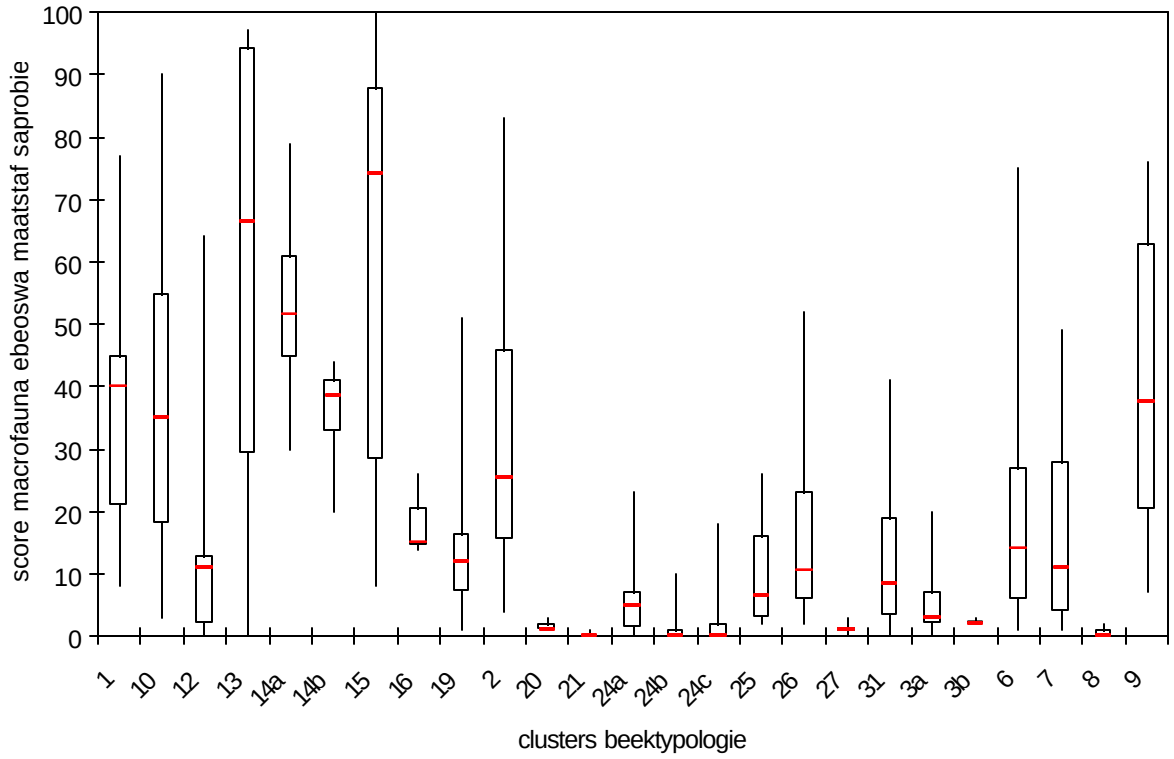


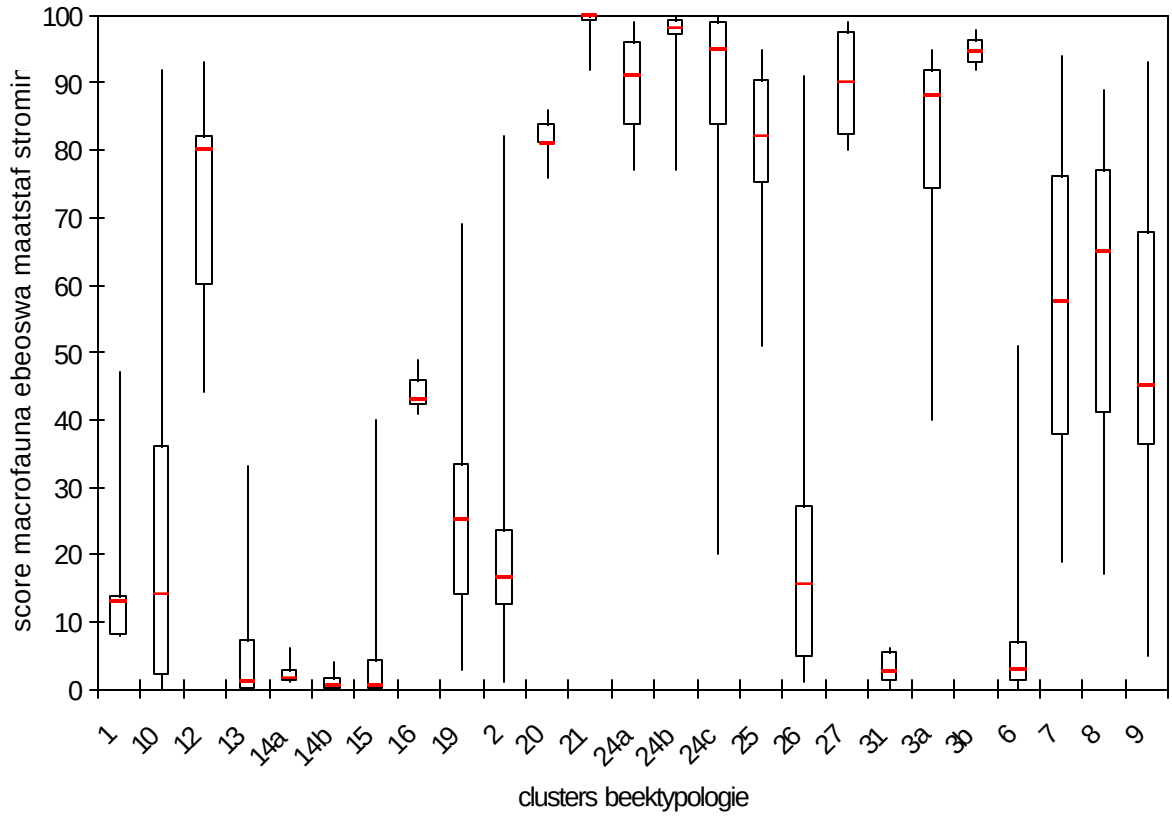


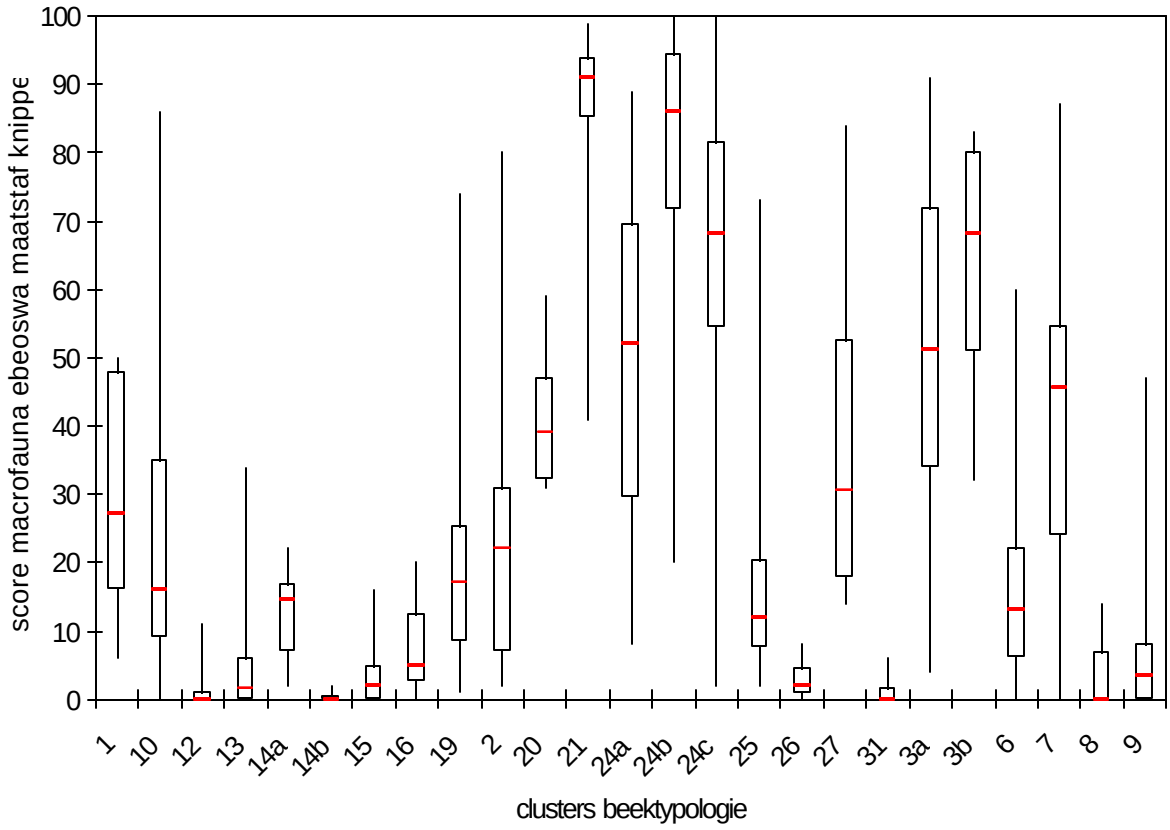

score macrofauna ebeoswa maatstaf slib
100
90
80
70
60
50
40
30
20
10
0
1
10
12
13
14a
14b
15
16
19
2
20
21
24a
24b
24c
25
26
27
31
3a
3b
6
7
8
9
clusters beektypologie

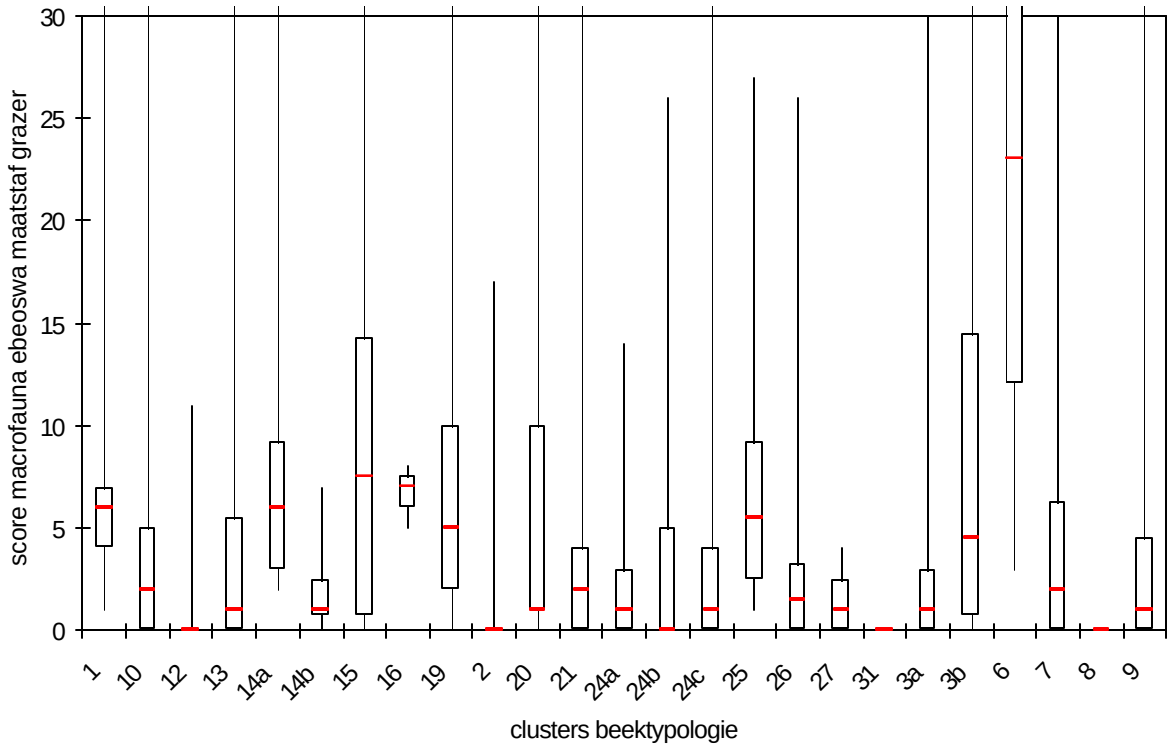

score macrofauna ebeoswa maatstaf plan
100
90
80
70
60
50
40
30
20
10
0
1
10
12
13
14a
14b
15
16
19
2
20
21
24a
24b
24c
25
26
27
31
3a
3b
6
7
8
9
clusters beektypologie

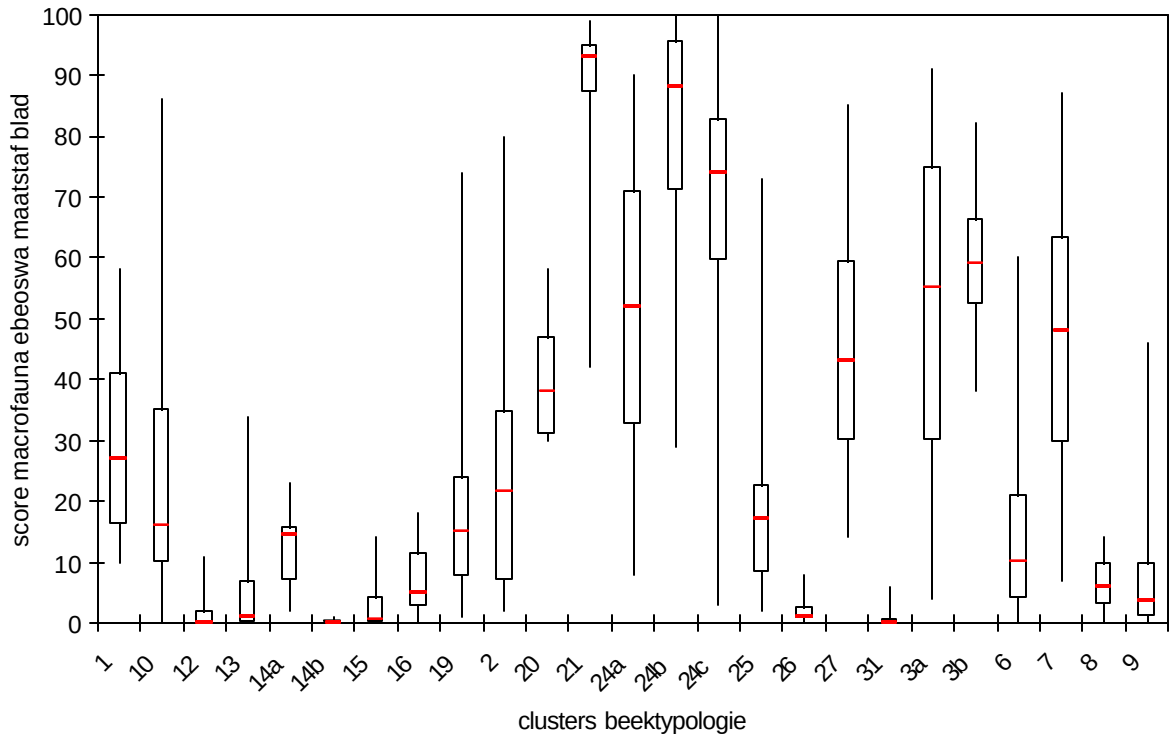

score macrofauna ebeoswa maatstaf trofie
100
90
80
70
60
50
40
30
20
10
0
1
10
12
13
14a
14b
15
16
19
2
20
21
24a
24b
24c
25
26
27
31
3a
3b
6
7
8
9
clusters beektypologie

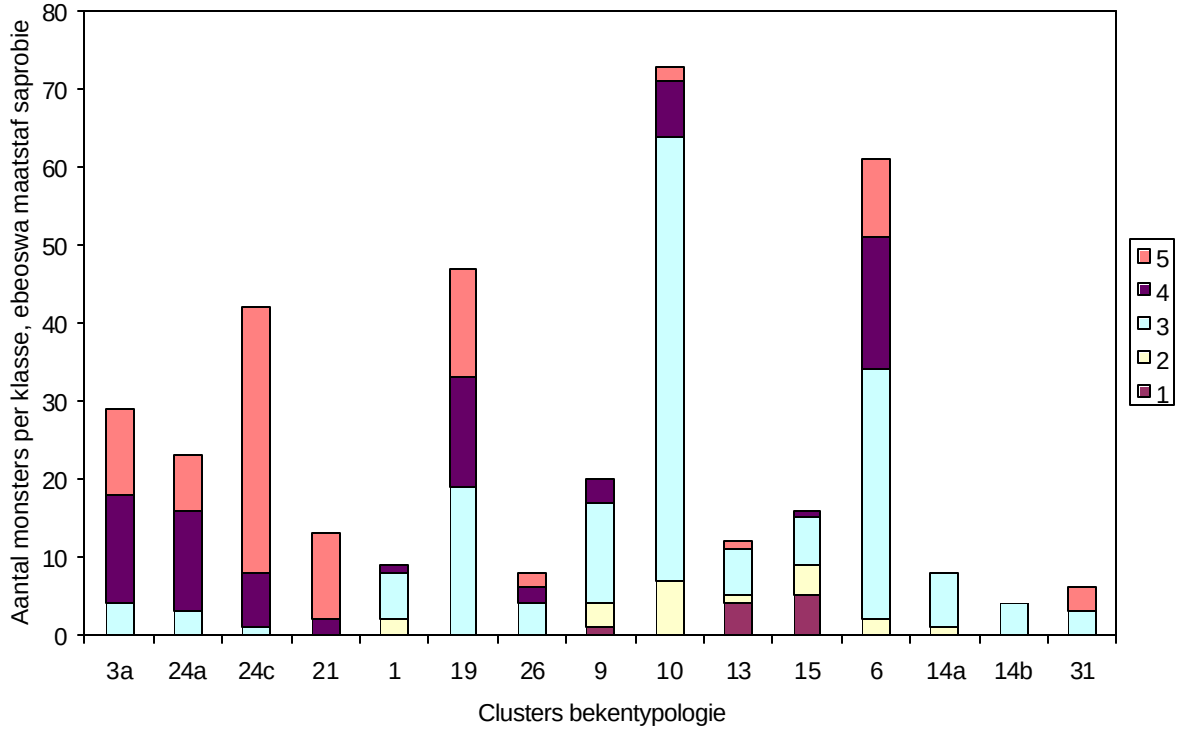

score macrofauna ebeoswa maatstaf saprobie
100
90
80
70
60
50
40
30
20
10
0
1
10
12
13
14a
14b
15
16
19
2
20
21
24a
24b
24c
25
26
27
31
3a
3b
6
7
8
9
clusters beektypologie

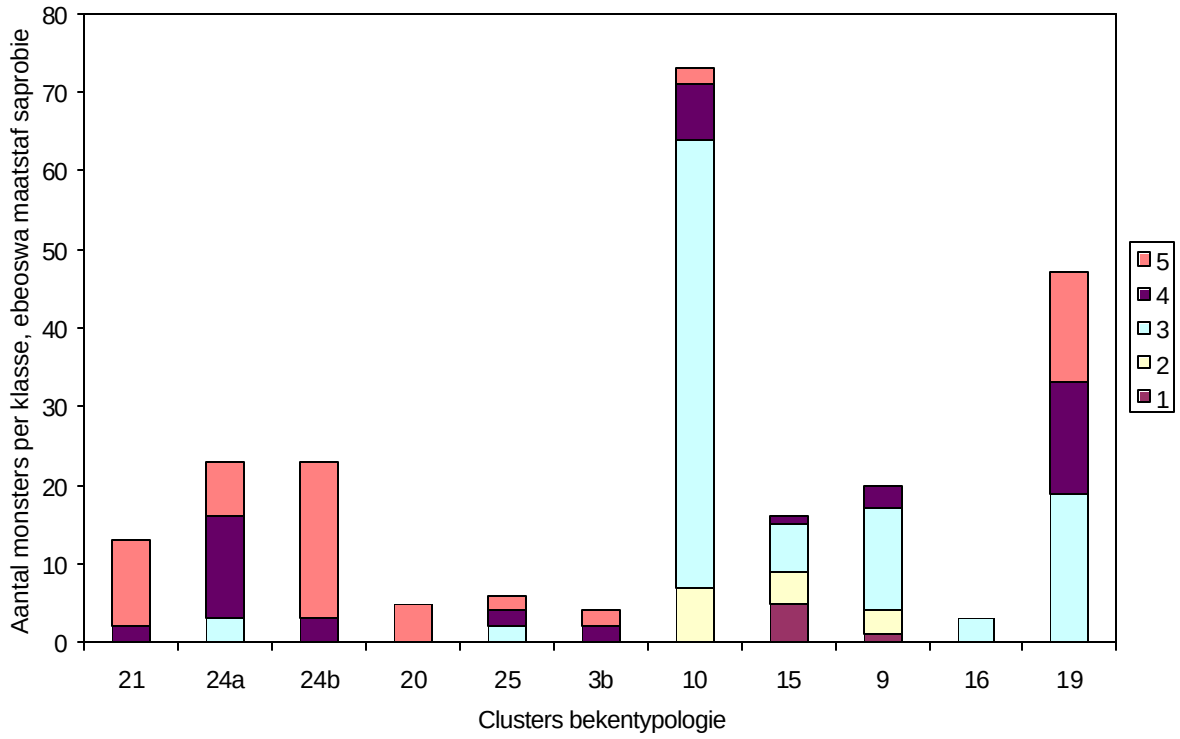

score macrofauna ebeoswa maatstaf stromir
100
90
80
70
60
50
40
30
20
10
0
1
10
12
13
14a
14b
15
16
19
2
20
21
24a
24b
24c
25
26
27
31
3a
3b
6
7
8
9
clusters beektypologie

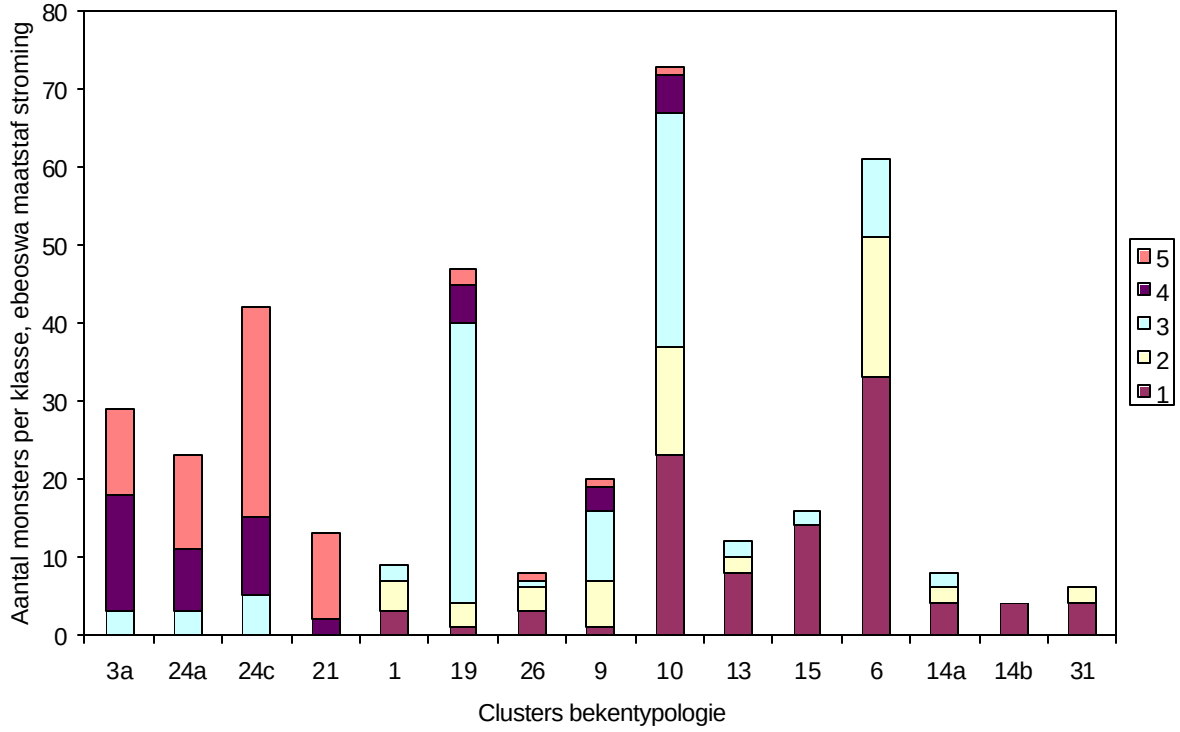

score macrofauna ebeoswa maatstaf knippe
100
90
80
70
60
50
40
30
20
10
0
1
10
12
13
14a
14b
15
16
19
2
20
21
24a
24b
24c
25
26
27
31
3a
3b
6
7
8
9
clusters beektypologie

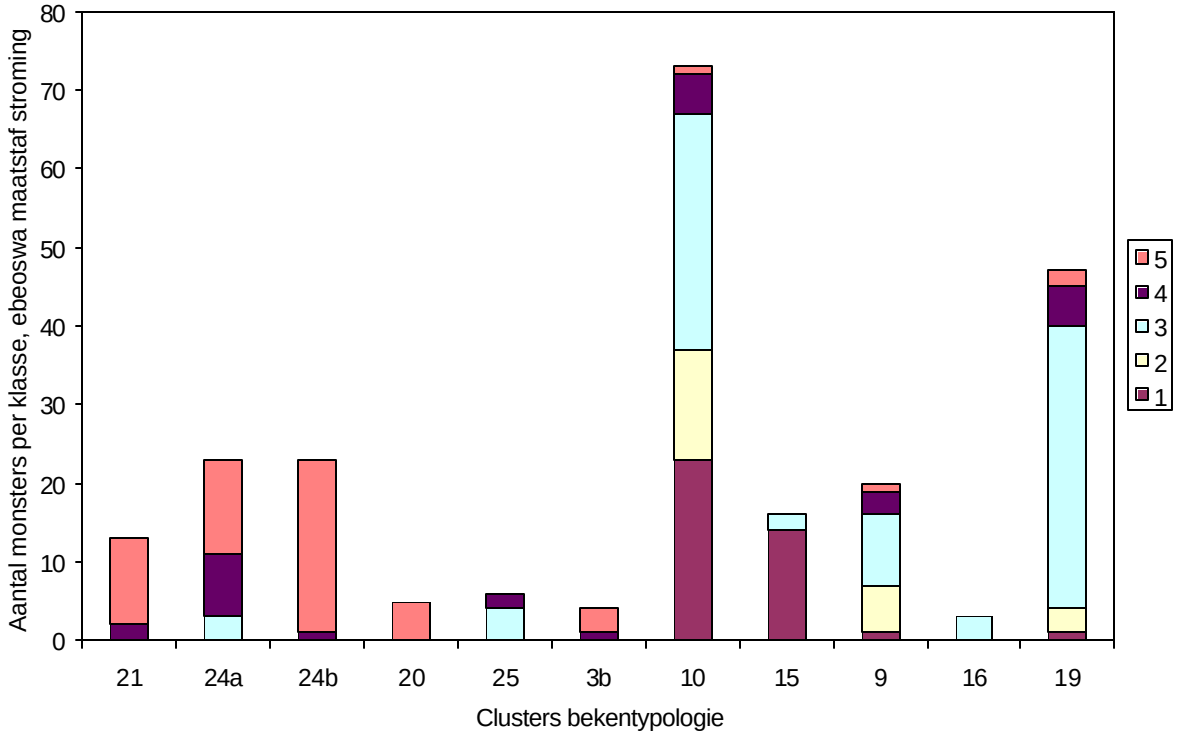

score macrofauna ebeoswa maatstaf grazer
30
25
20
15
10
5
0
1
10
12
13
14a
14b
15
16
19
2
20
21
24a
24b
24c
25
26
27
31
3a
3b
6
7
8
9
clusters beektypologie

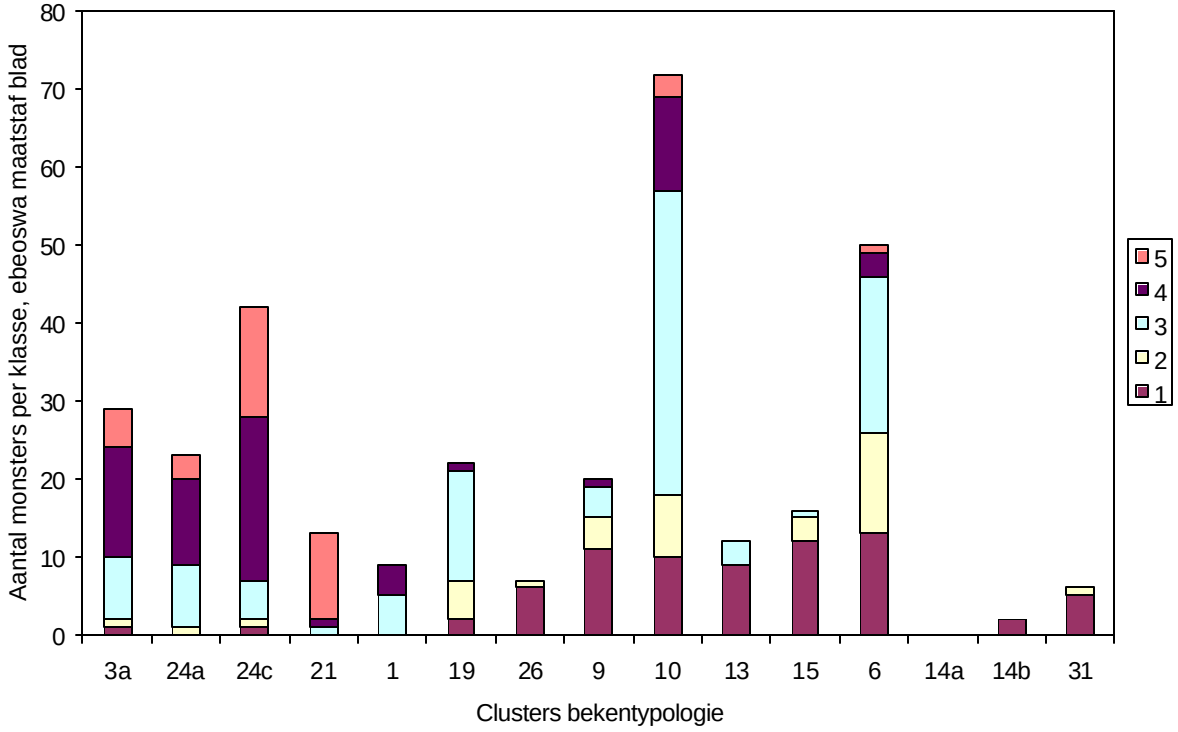

score macrofauna ebeoswa maatstaf blad
100
90
80
70
60
50
40
30
20
10
0
1
10
12
13
14a
14b
15
16
19
2
20
21
24a
24b
24c
25
26
27
31
3a
3b
6
7
8
9
clusters beektypologie

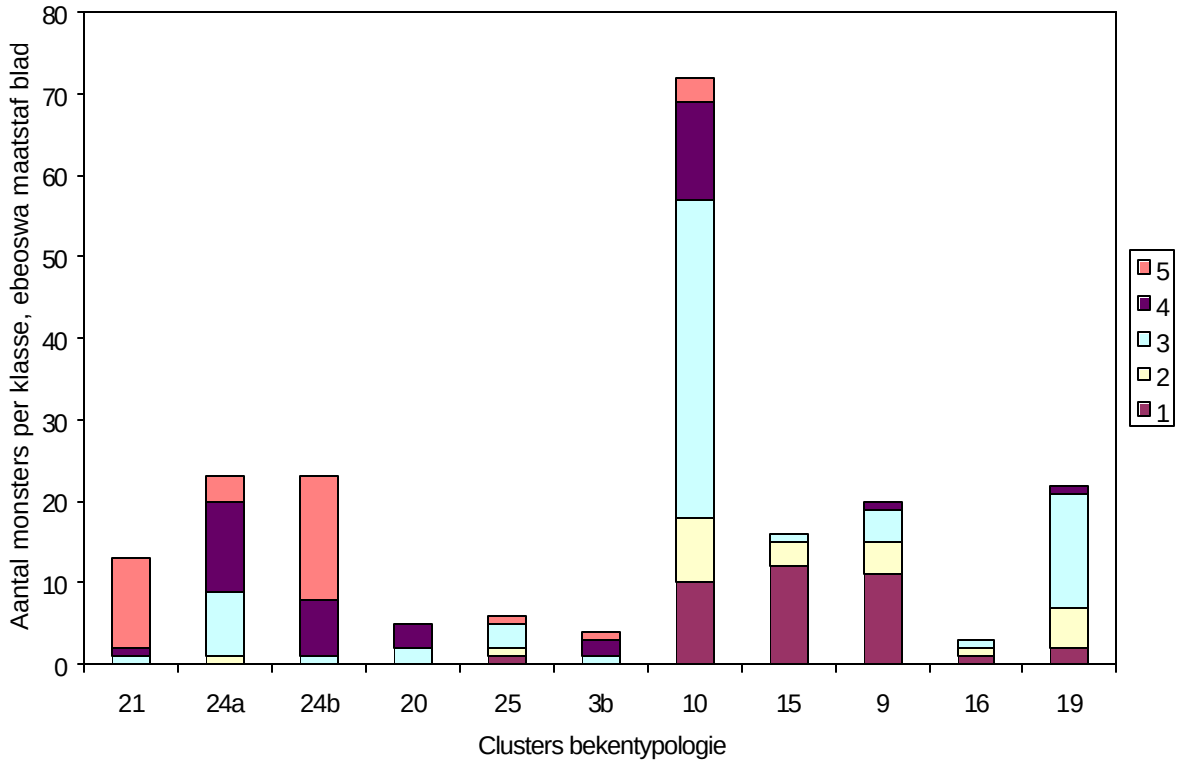

Aantal monsters per klasse, ebeoswa maatstaf saprobie
80
70
60
50
40
30
20
10
0
3a
24a
24c
21
1
19
26
9
10
13
15
6
14a
14b
31
Clusters bekentypologie
5
4
3
2
1

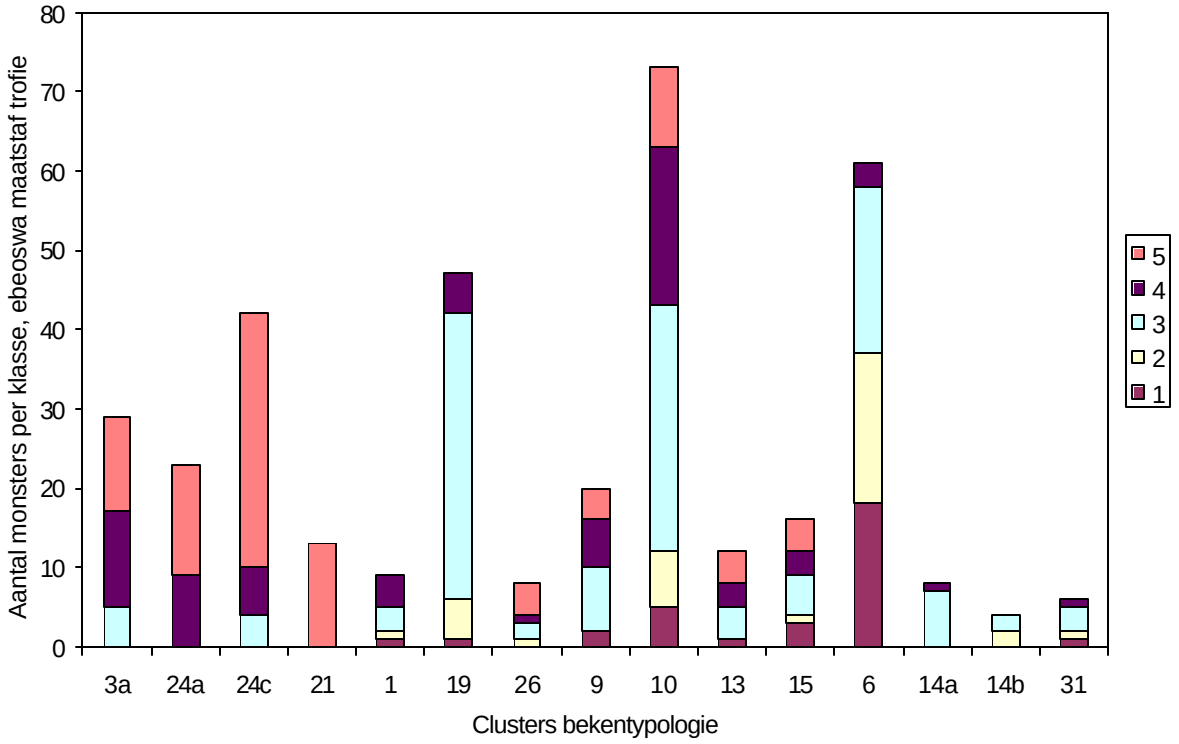

Aantal monsters per klasse, ebeoswa maatstaf saprobie
0
10
20
30
40
50
60
70
80
21
24a
24b
20
25
3b
10
15
9
16
19
Clusters bekentypologie
5
4
3
2
1

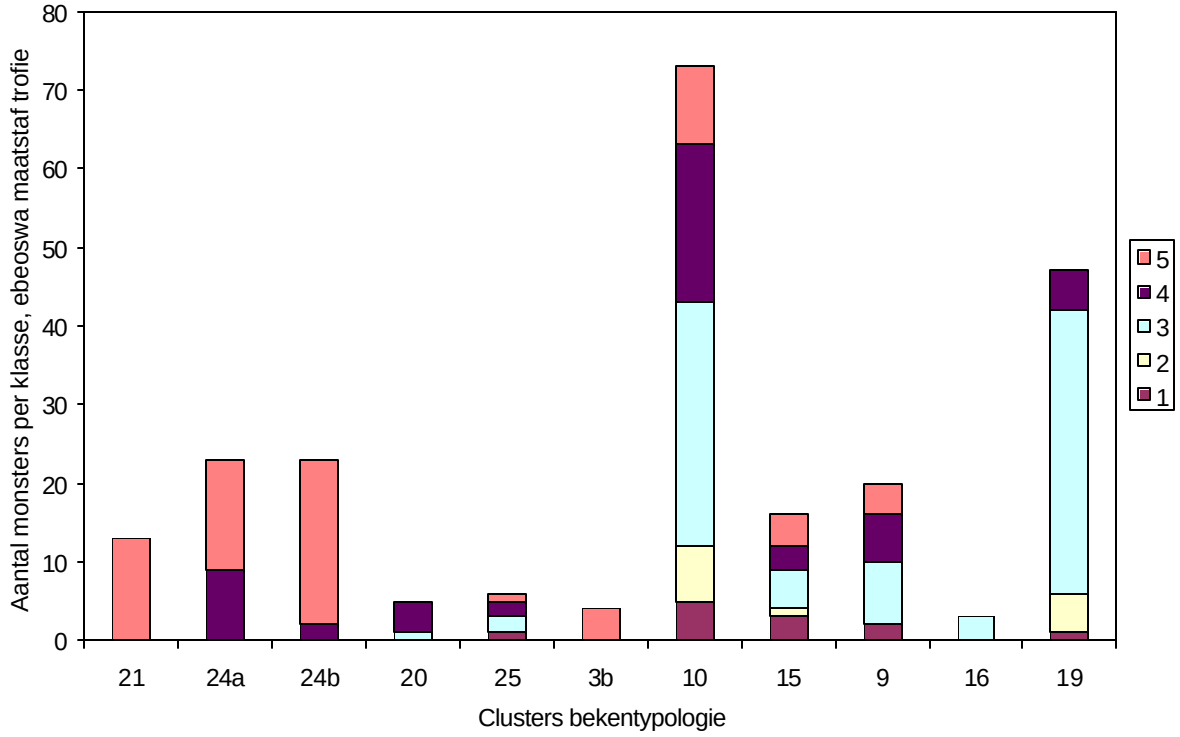

Aantal monsters per klasse, ebeoswa maatstaf stroming
0
10
20
30
40
50
60
70
80
3a
24a
24c
21
1
19
26
9
10
13
15
6
14a
14b
31
Clusters bekentypologie
5
4
3
2
1

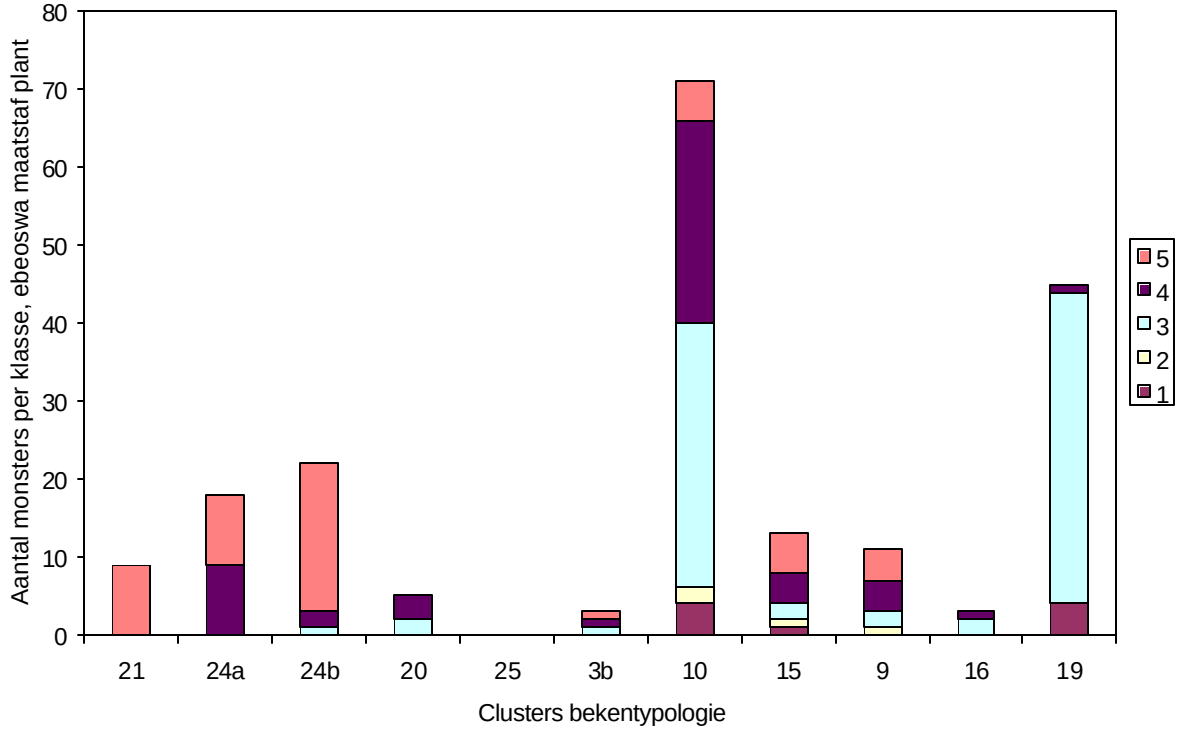

Aantal monsters per klasse, ebeoswa maatstaf stroming
0
10
20
30
40
50
60
70
80
21
24a
24b
20
25
3b
10
15
9
16
19
Clusters bekentypologie
5
4
3
2
1

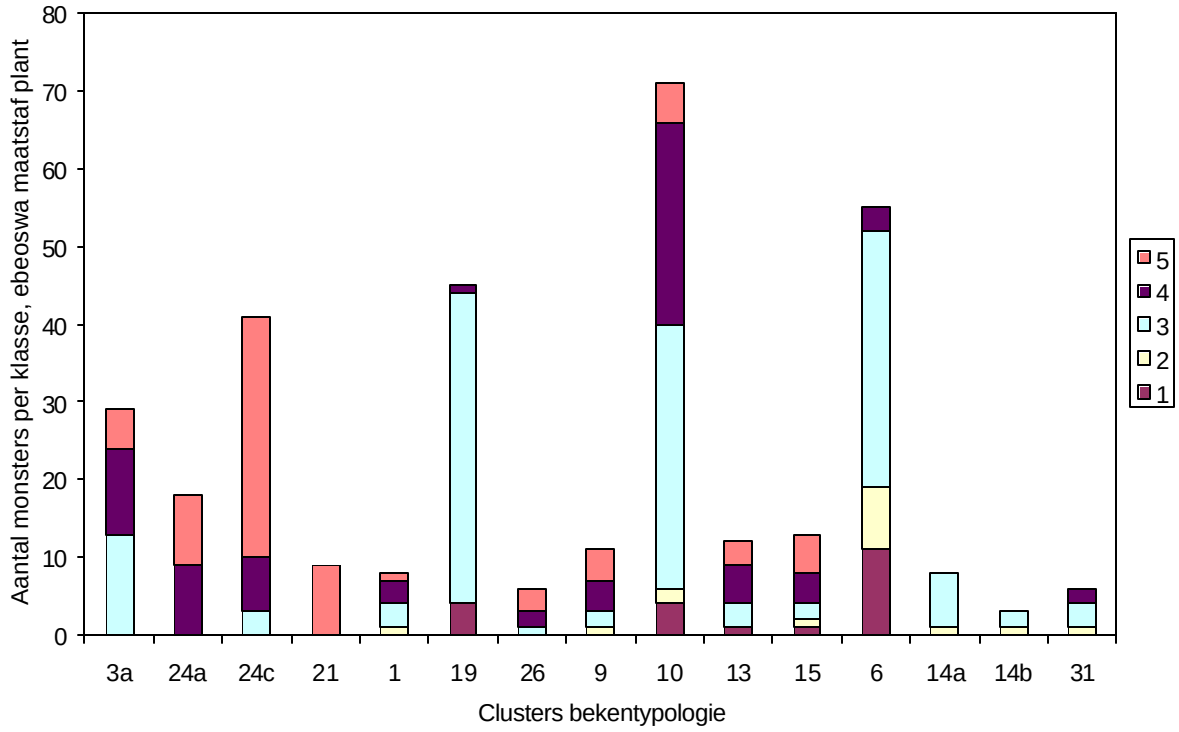

Aantal monsters per klasse, ebeoswa maatstaf blad
0
10
20
30
40
50
60
70
80
3a
24a
24c
21
1
19
26
9
10
13
15
6
14a
14b
31
Clusters bekentypologie
5
4
3
2
1

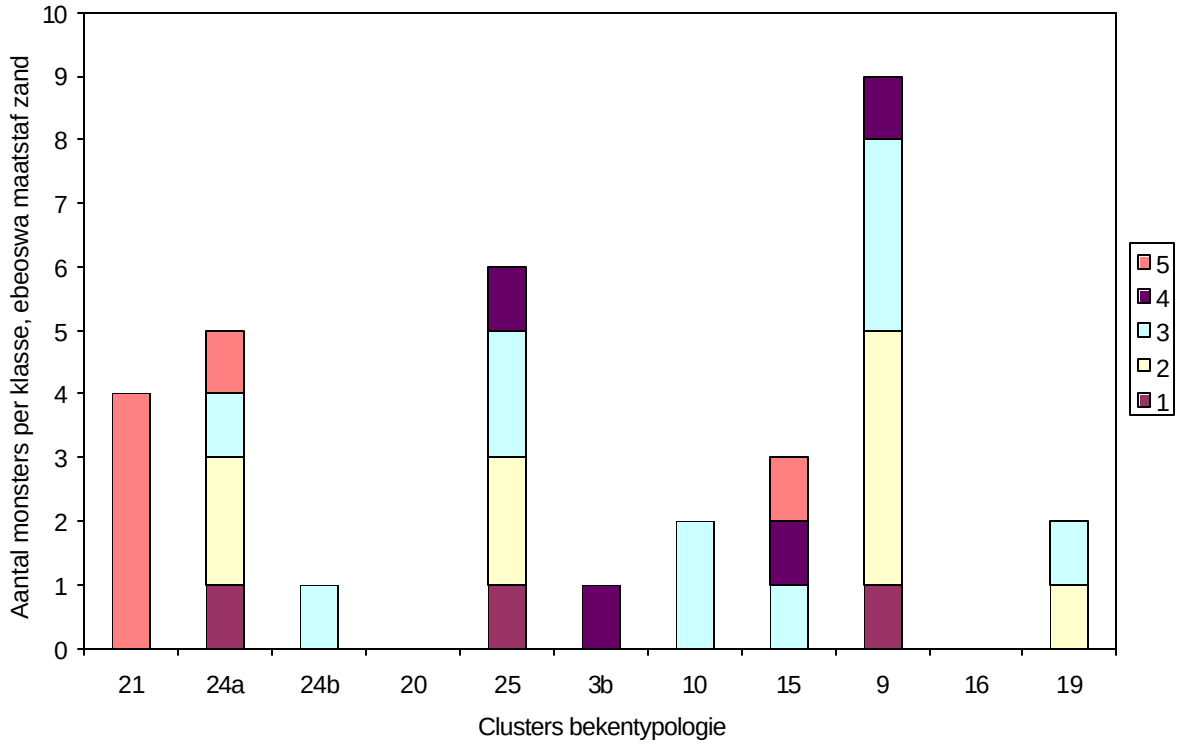

Aantal monsters per klasse, ebeoswa maatstaf blad
0
10
20
30
40
50
60
70
80
21
24a
24b
20
25
3b
10
15
9
16
19
Clusters bekentypologie
5
4
3
2
1

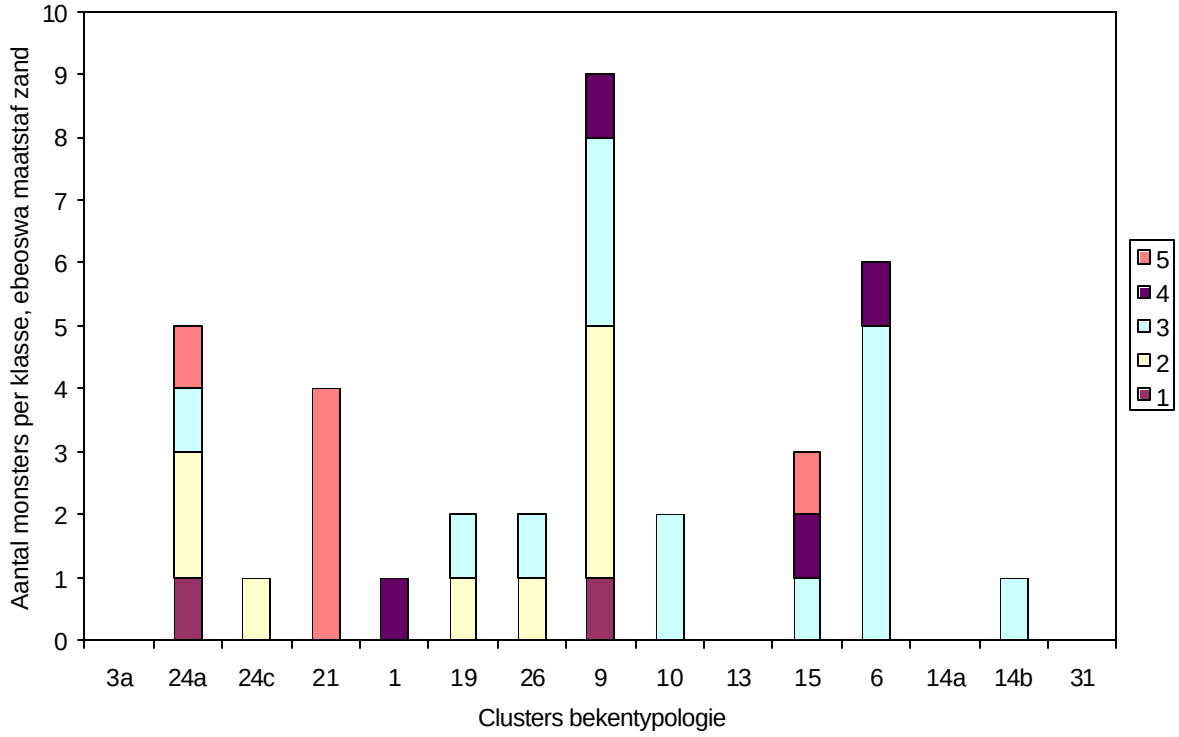

Aantal monsters per klasse, ebeoswa maatstaf trofie
0
10
20
30
40
50
60
70
80
3a
24a
24c
21
1
19
26
9
10
13
15
6
14a
14b
31
Clusters bekentypologie
5
4
3
2
1

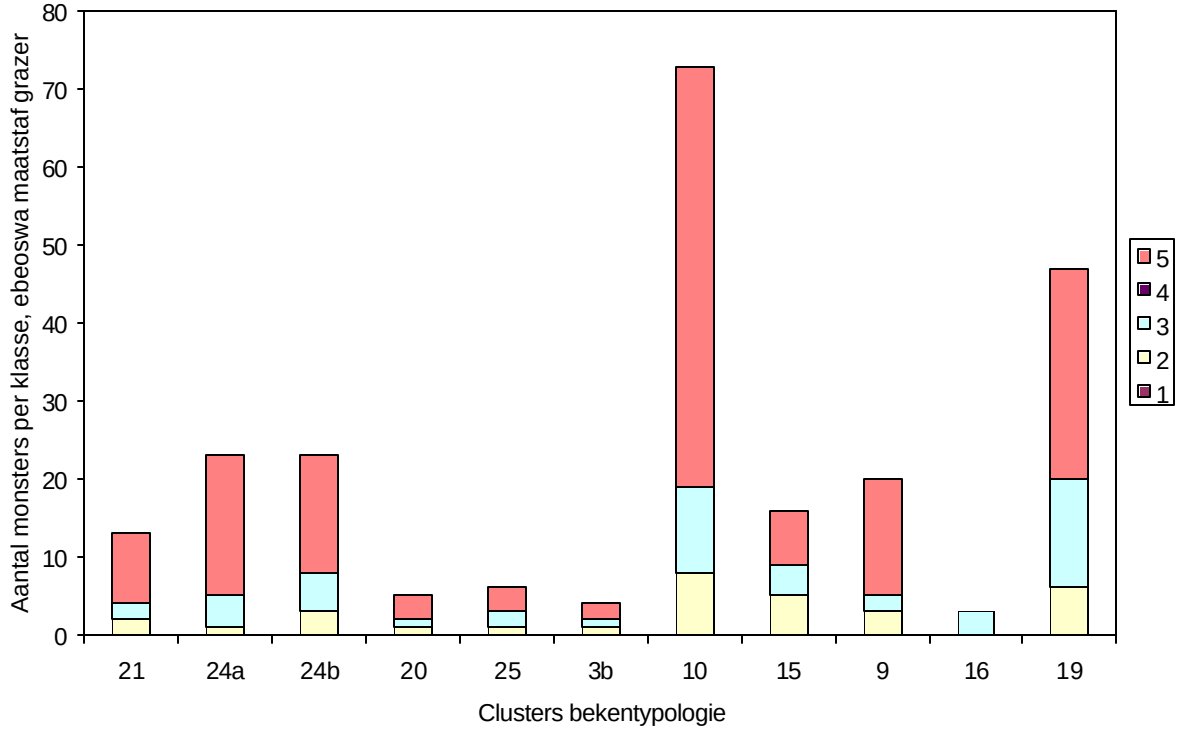

Aantal monsters per klasse, ebeoswa maatstaf trofie
0
10
20
30
40
50
60
70
80
21
24a
24b
20
25
3b
10
15
9
16
19
Clusters bekentypologie
5
4
3
2
1

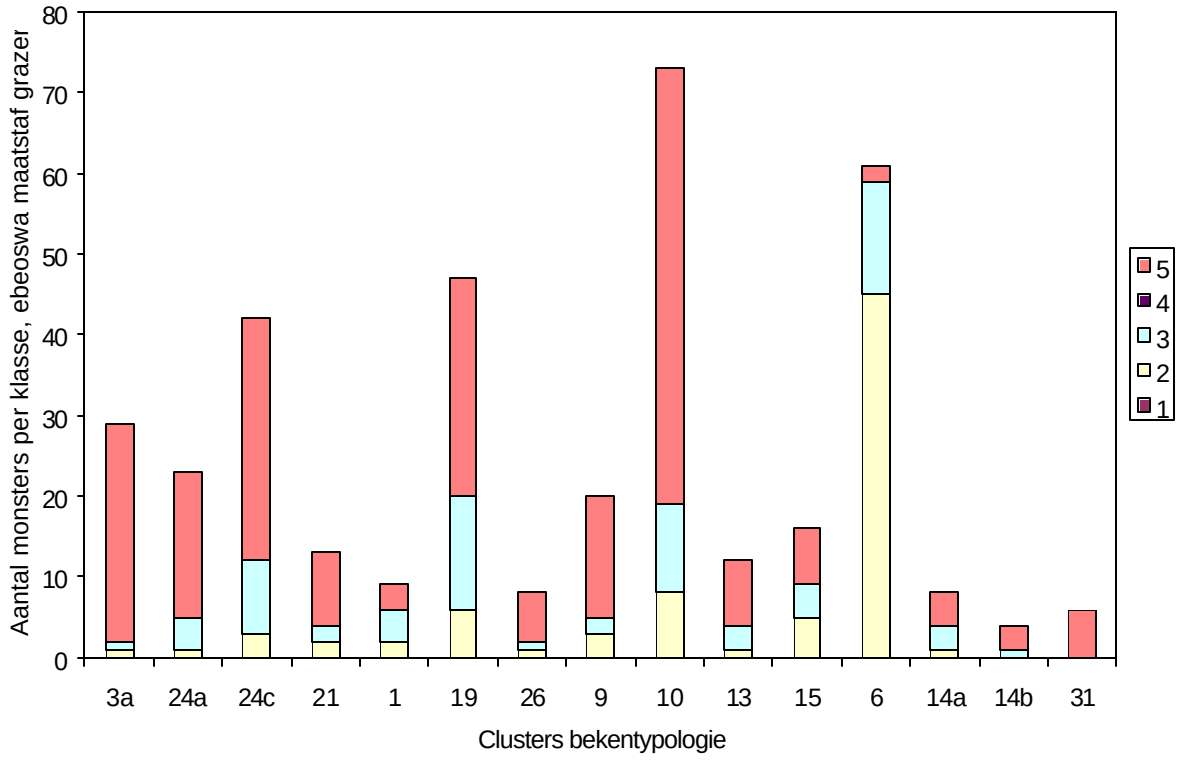

Aantal monsters per klasse, ebeoswa maatstaf plant
0
10
20
30
40
50
60
70
80
21
24a
24b
20
25
3b
10
15
9
16
19
Clusters bekentypologie
5
4
3
2
1

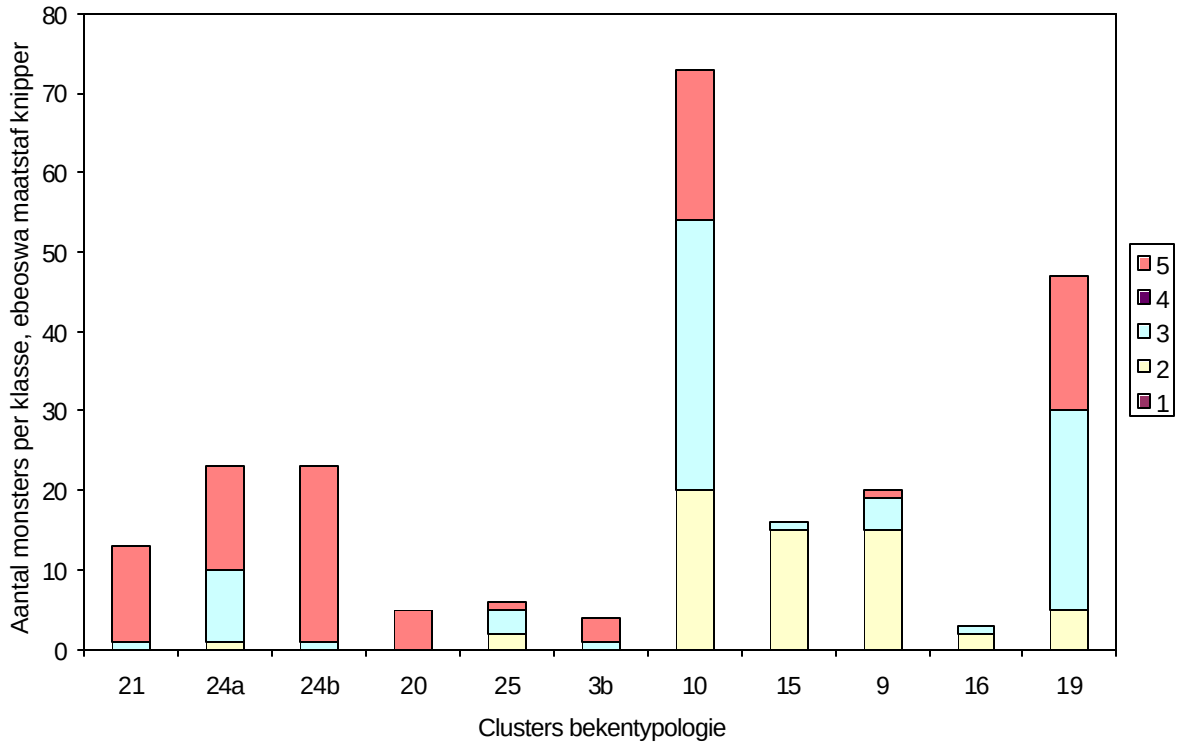

Aantal monsters per klasse, ebeoswa maatstaf plant
80
70
60
50
40
30
20
10
0
3a
24a
24c
21
1
19
26
9
10
13
15
6
14a
14b
31
Clusters bekentypologie
5
4
3
2
1

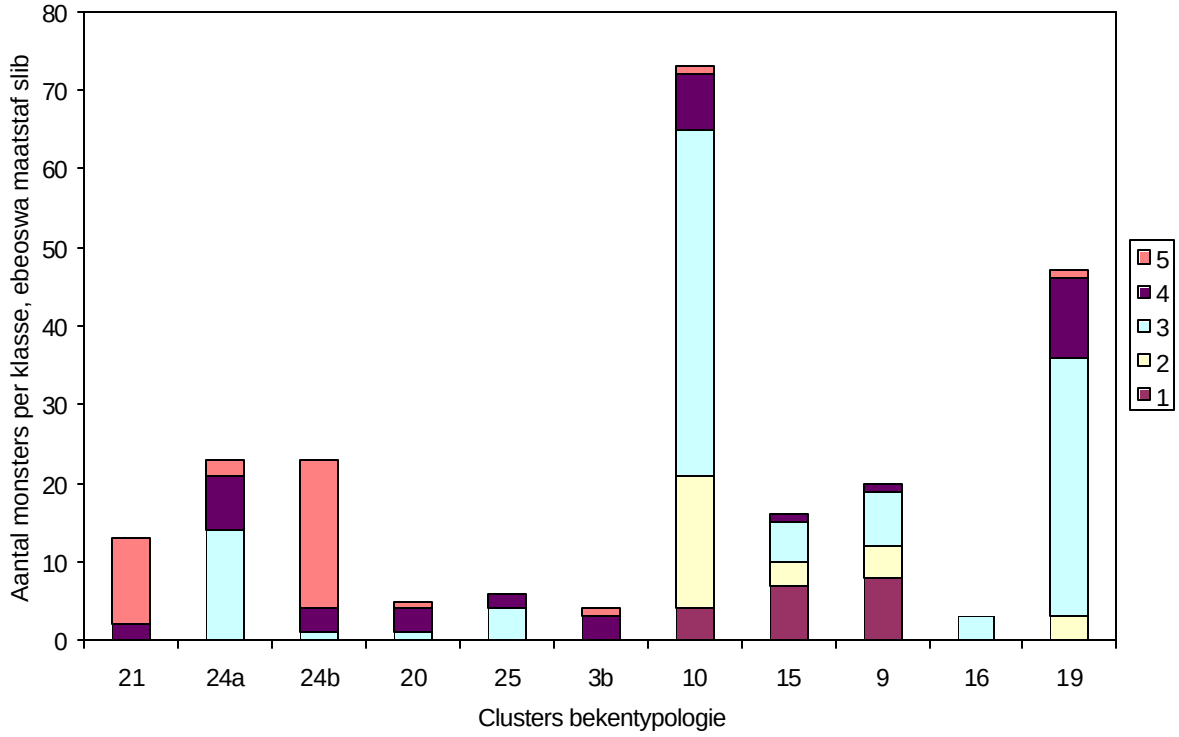

Aantal monsters per klasse, ebeoswa maatstaf zand
10
9
8
7
6
5
4
3
2
1
0
21
24a
24b
20
25
3b
10
15
9
16
19
Clusters bekentypologie
5
4
3
2
1

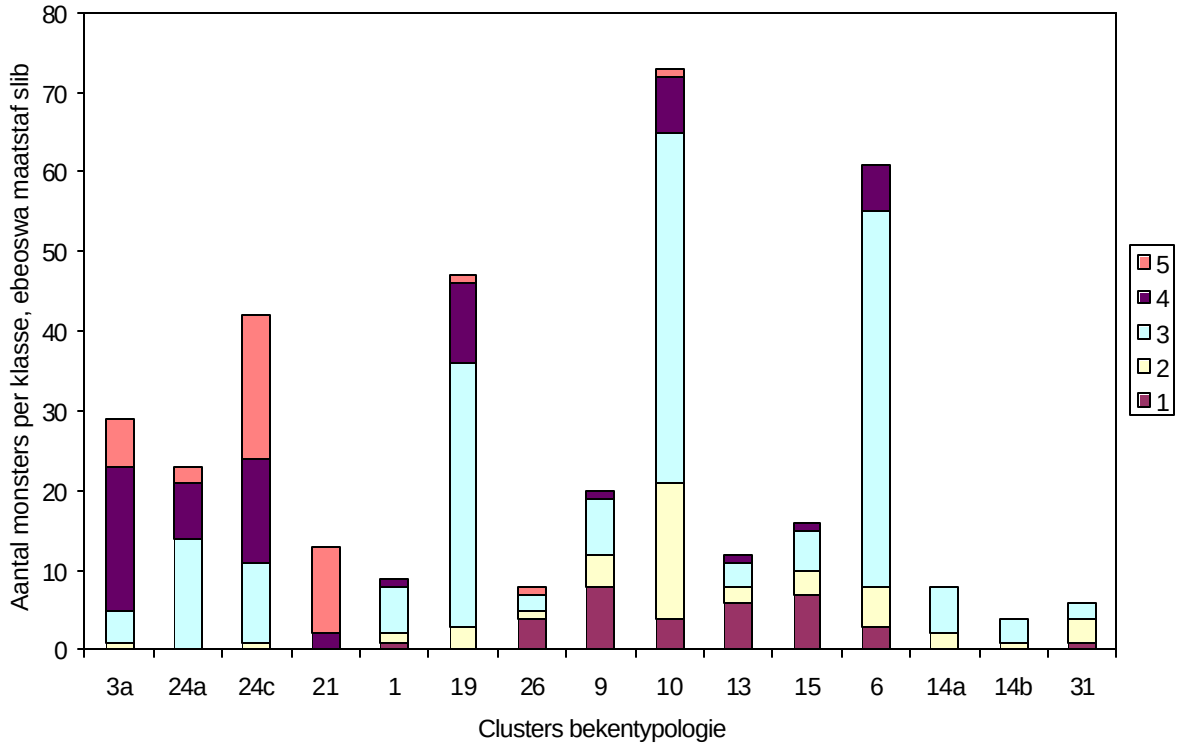

Aantal monsters per klasse, ebeoswa maatstaf zand
0
1
2
3
4
5
6
7
8
9
10
3a
24a
24c
21
1
19
26
9
10
13
15
6
14a
14b
31
Clusters bekentypologie
5
4
3
2
1

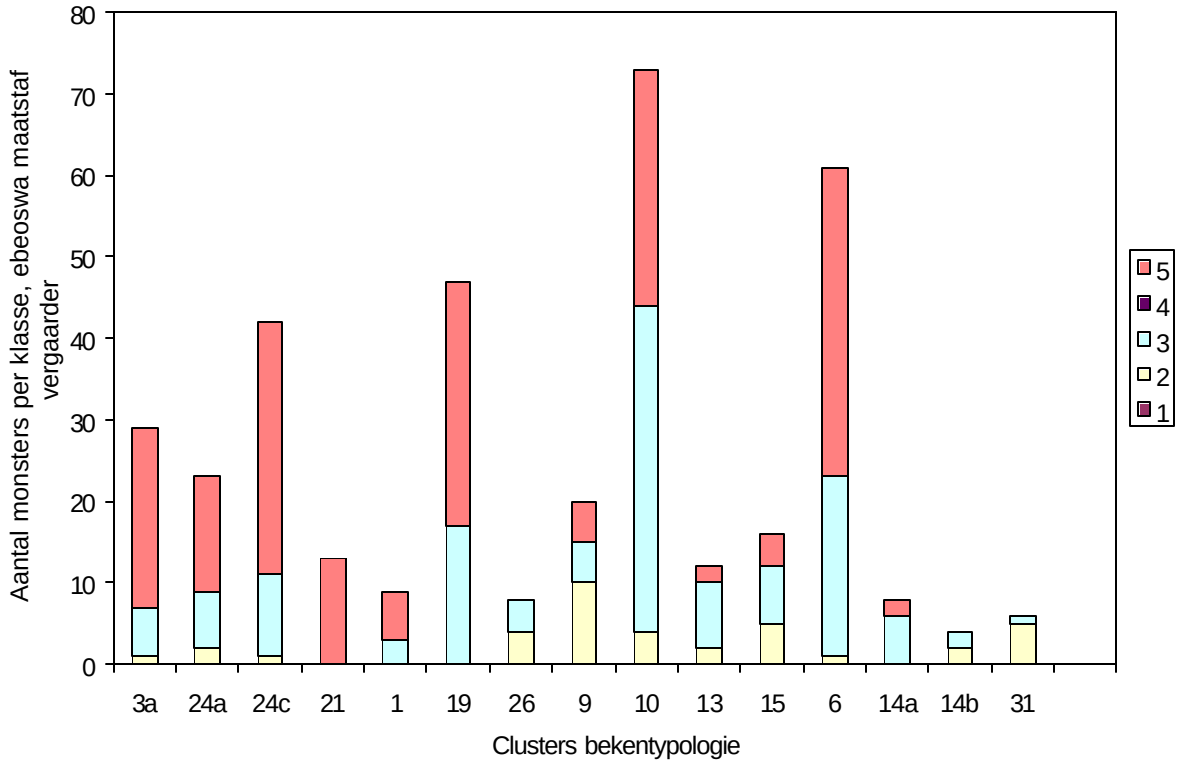

Aantal monsters per klasse, ebeoswa maatstaf grazer
0
10
20
30
40
50
60
70
80
21
24a
24b
20
25
3b
10
15
9
16
19
Clusters bekentypologie
5
4
3
2
1

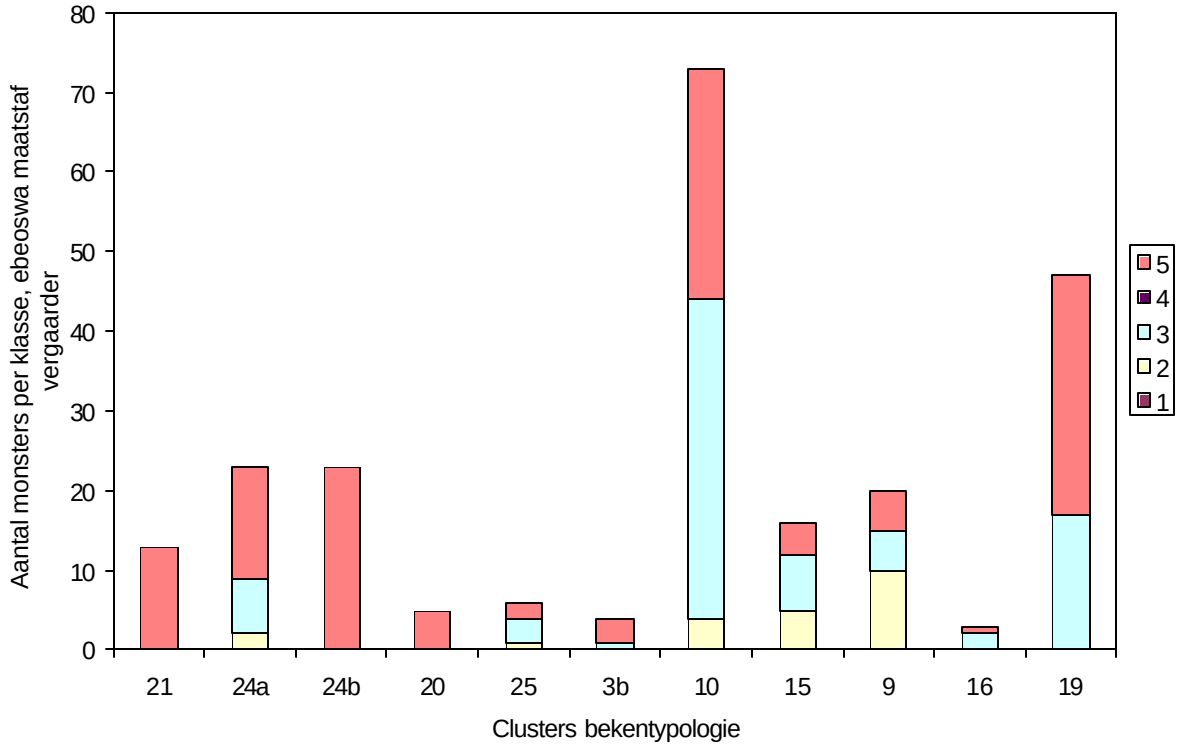

Aantal monsters per klasse, ebeoswa maatstaf grazer
80
70
60
50
40
30
20
10
0
3a
24a
24c
21
1
19
26
9
10
13
15
6
14a
14b
31
Clusters bekentypologie
5
4
3
2
1

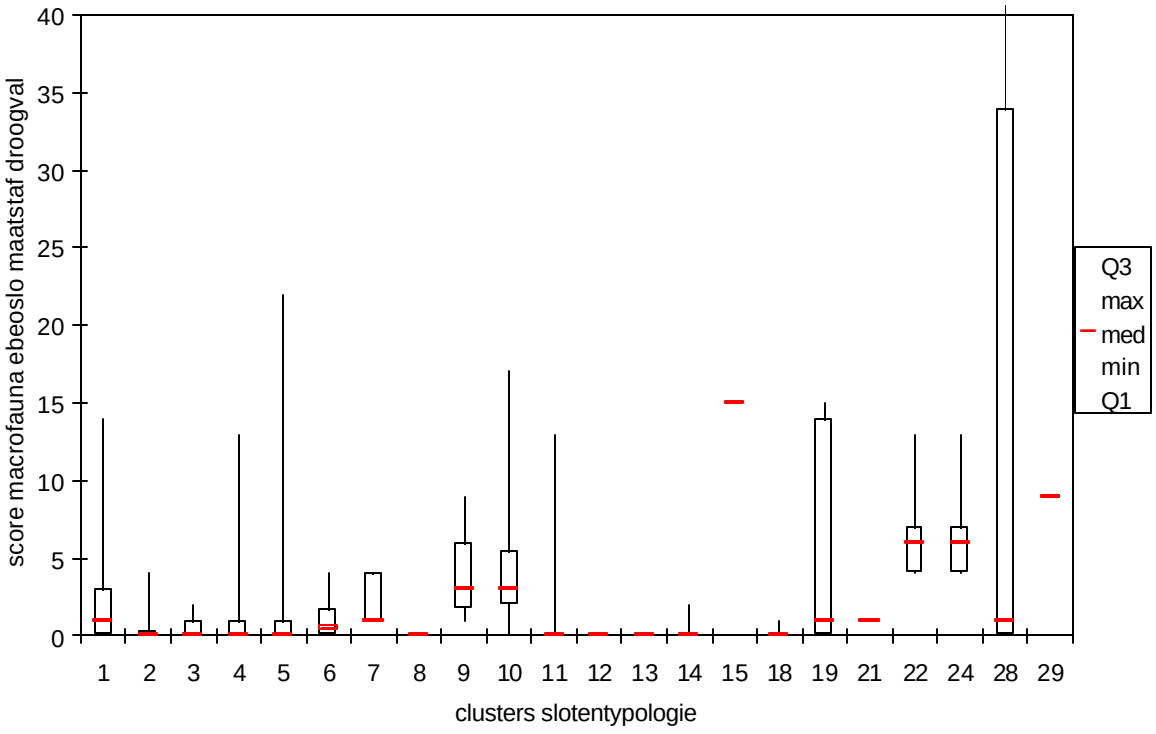

Aantal monsters per klasse, ebeoswa maatstaf knipper
80
70
60
50
40
30
20
10
0
21
24a
24b
20
25
3b
10
15
9
16
19
Clusters bekentypologie
5
4
3
2
1

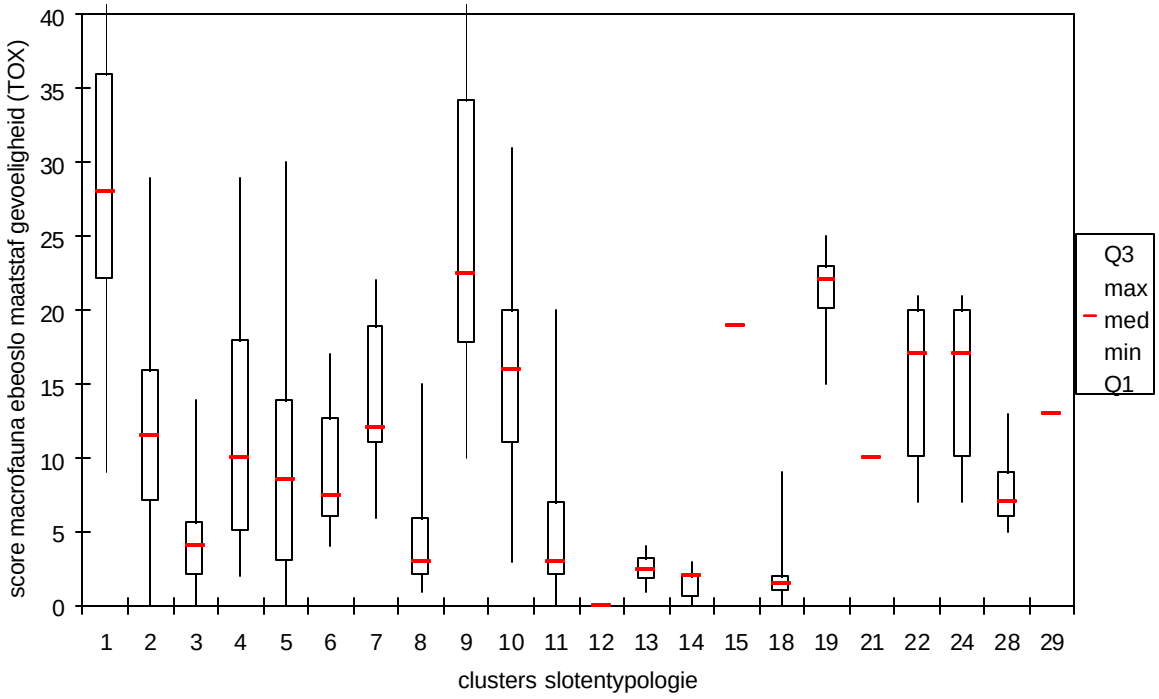

Aantal monsters per klasse, ebeoswa maatstaf slib
0
10
20
30
40
50
60
70
80
21
24a
24b
20
25
3b
10
15
9
16
19
Clusters bekentypologie
5
4
3
2
1

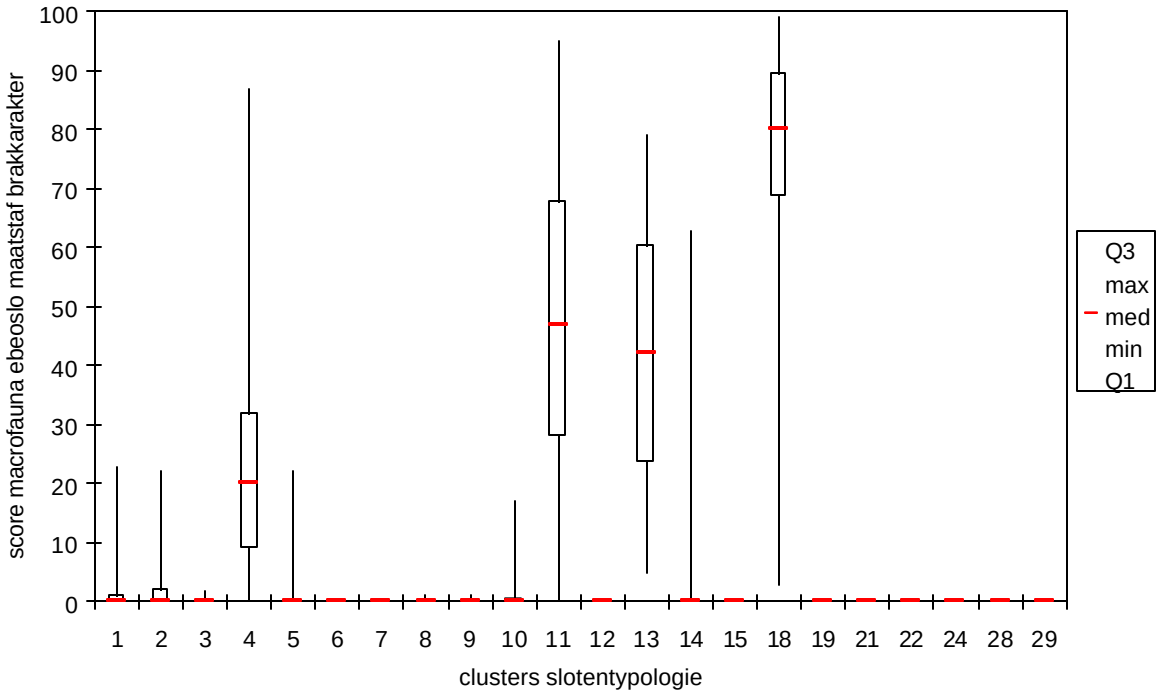

Aantal monsters per klasse, ebeoswa maatstaf slib
0
10
20
30
40
50
60
70
80
3a
24a
24c
21
1
19
26
9
10
13
15
6
14a
14b
31
Clusters bekentypologie
5
4
3
2
1

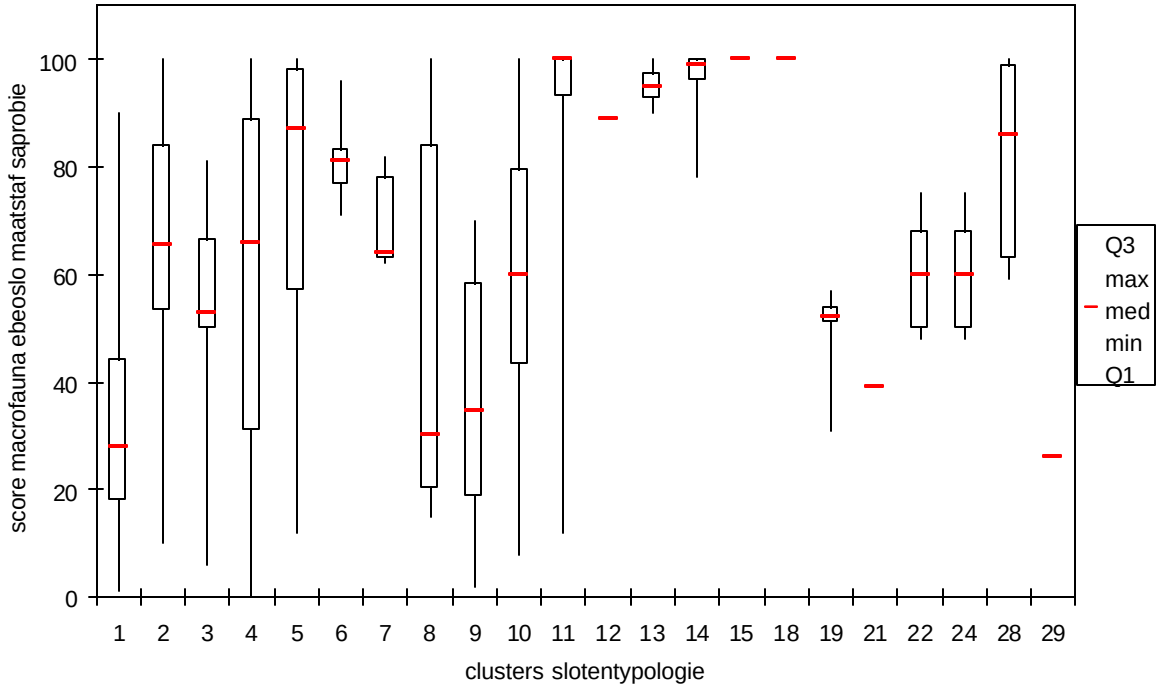

Aantal monsters per klasse, ebeoswa maatstaf vergaarder
80
70
60
50
40
30
20
10
0
3a
24a
24c
21
1
19
26
9
10
13
15
6
14a
14b
31
Clusters bekentypologie
5
4
3
2
1

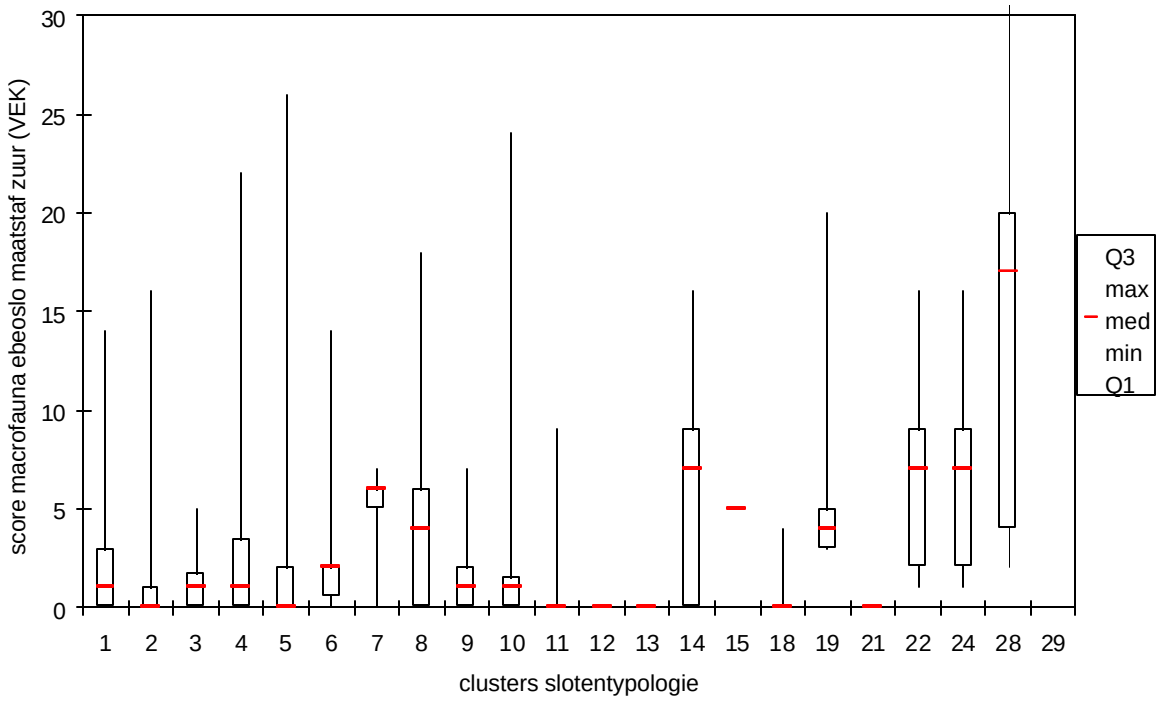

Aantal monsters per klasse, ebeoswa maatstaf vergaarder
80
70
60
50
40
30
20
10
0
21
24a
24b
20
25
3b
10
15
9
16
19
Clusters bekentypologie
5
4
3
2
1

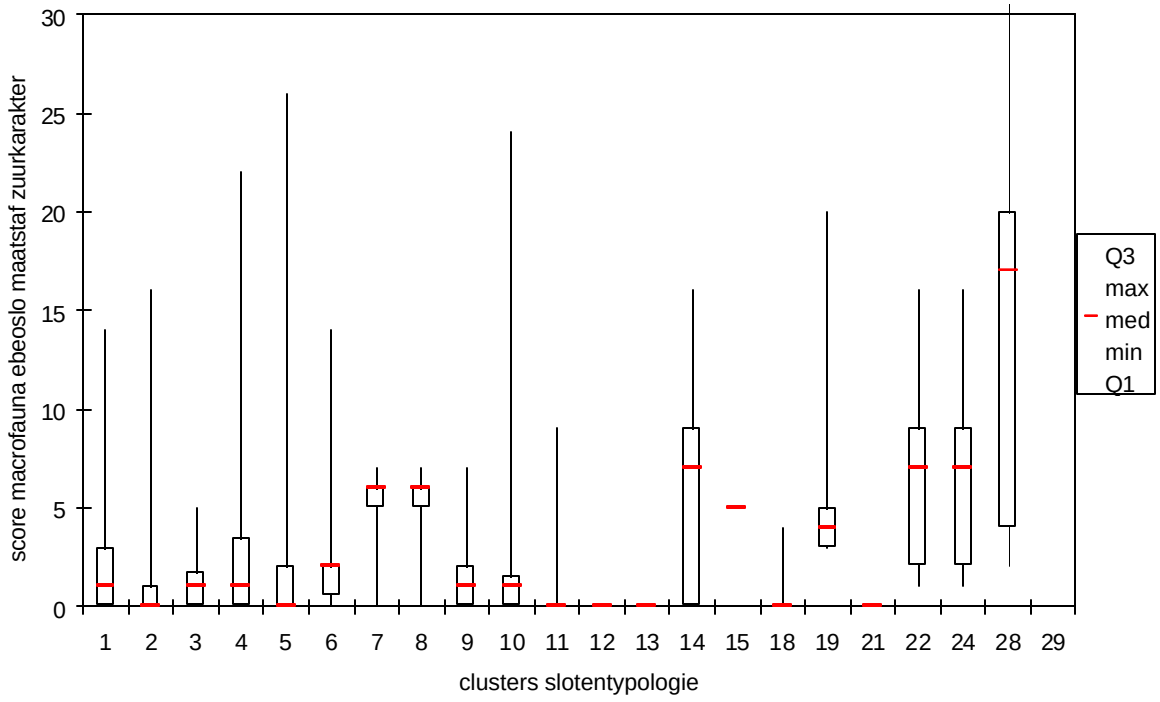

score macrofauna ebeoslo maatstaf droogval
clusters slotentypologie
Q3
max
med
min
Q1

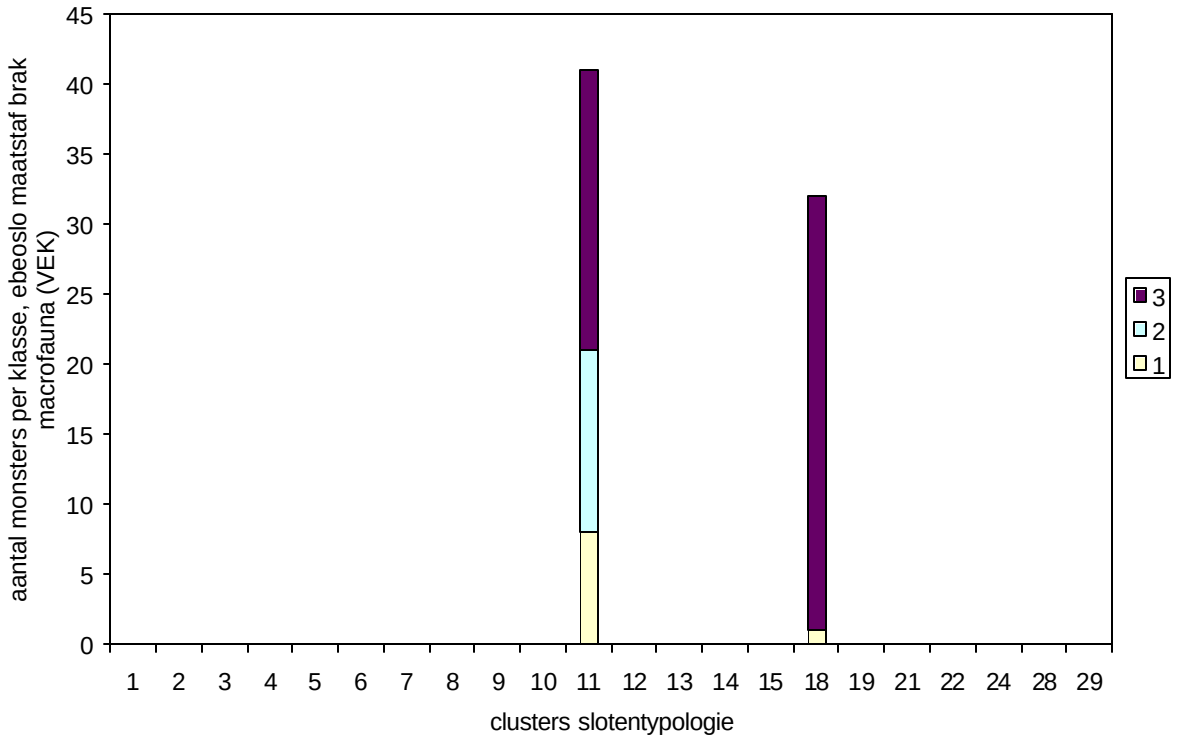

score macrofauna ebeoslo maatstaf gevoeligheid (TOX)
clusters slotentypologie
Q3
max
med
min
Q1

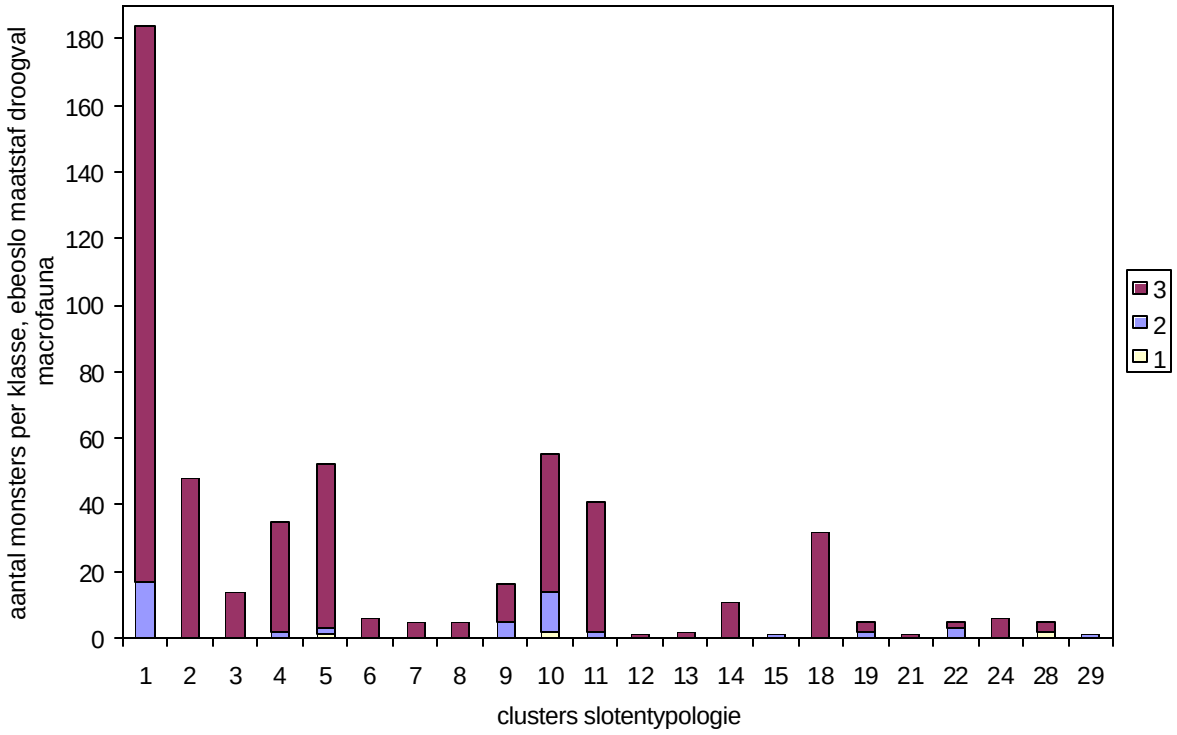

score macrofauna ebeoslo maatstaf brakkarakter
100
90
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 18 19 21 22 24 28 29
clusters slotentypologie
Q3
max
med
min
Q1

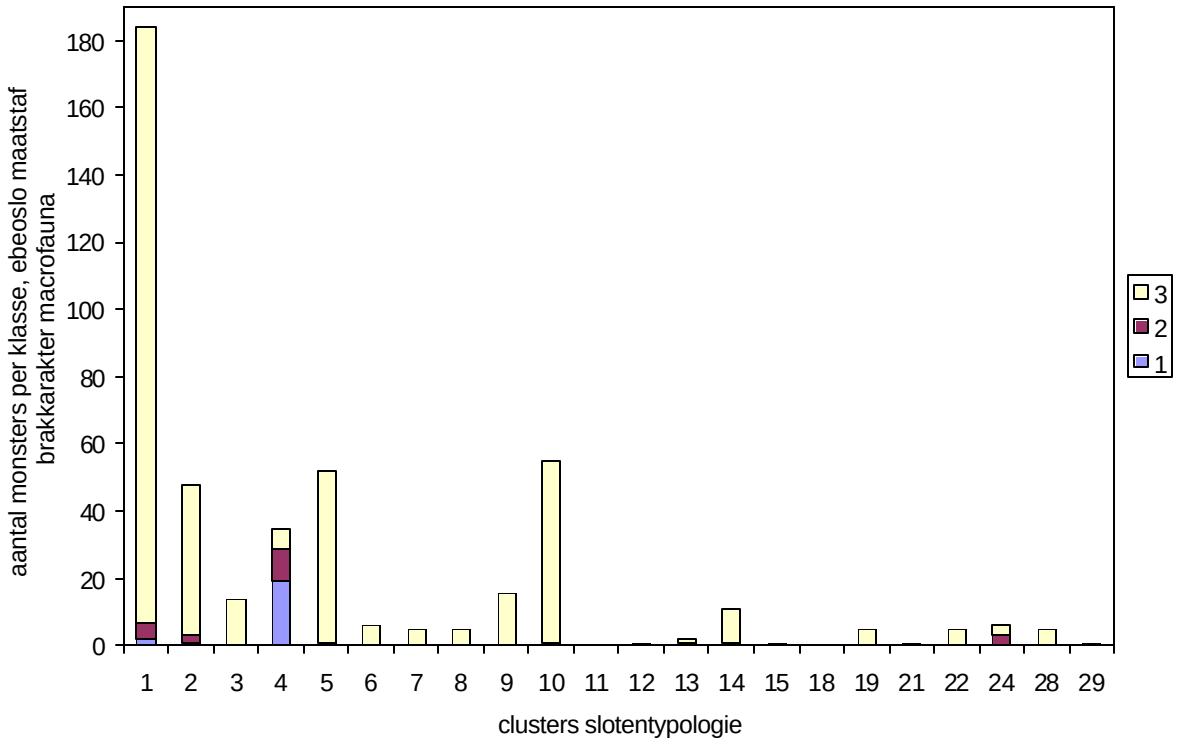

score macrofauna ebeoslo maatstaf saprobie
100
80
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 18 19 21 22 24 28 29
clusters slotentypologie
Q3
max
med
min
Q1

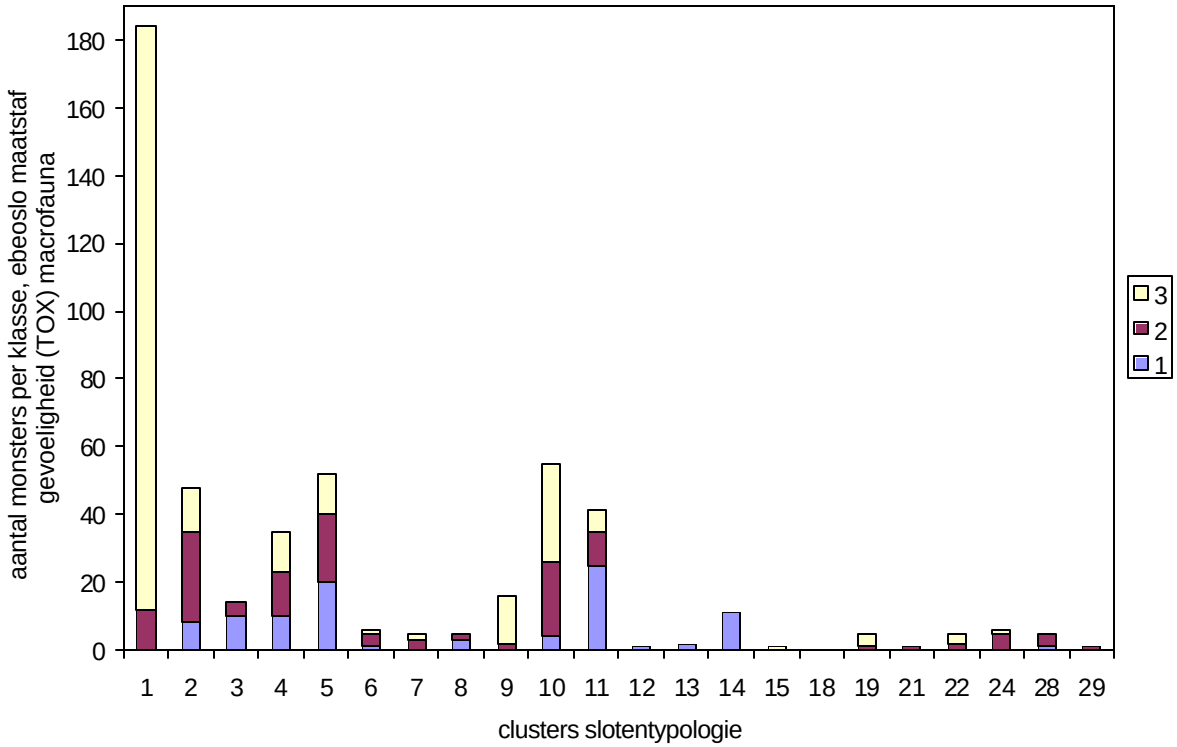

score macrofauna ebeoslo maatstaf zuur (VEK)
30
25
20
15
10
5
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 18 19 21 22 24 28 29
clusters slotentypologie
Q3
max
med
min
Q1

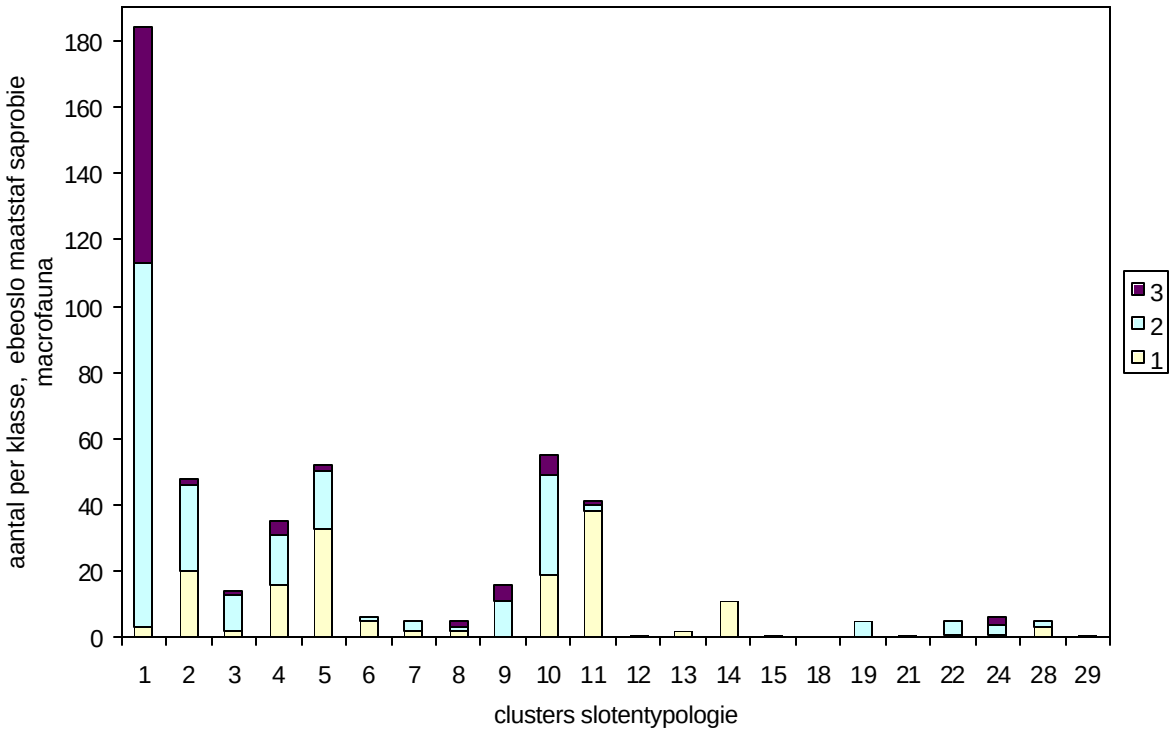

score macrofauna ebeoslo maatstaf zuurkarakter
30
25
20
15
10
5
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 18 19 21 22 24 28 29
clusters slotentypologie
Q3
max
med
min
Q1

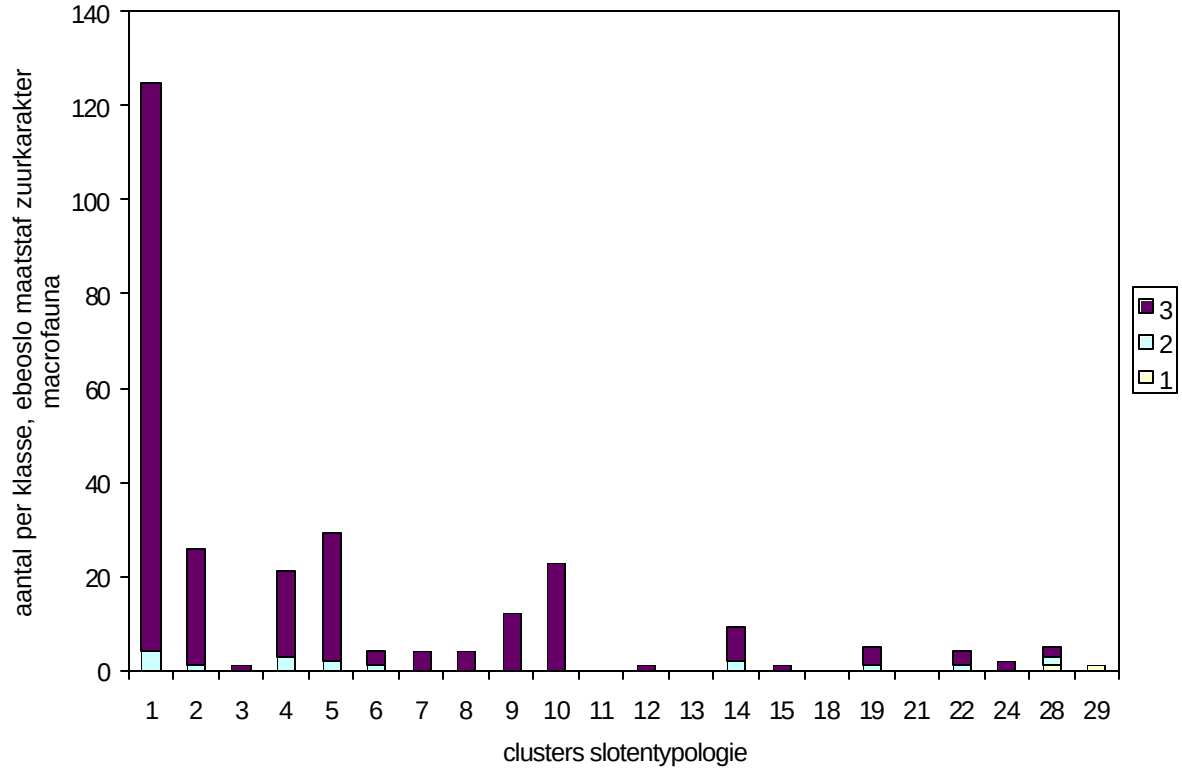

aantal monsters per klasse, ebeoslo maatstaf brak macrofauna (VEK)
0
5
10
15
20
25
30
35
40
45
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 18 19 21 22 24 28 29
clusters slotentypologie
3
2
1


aantal monsters per klasse, ebeoslo maatstaf droogval macrofauna
0
20
40
60
80
100
120
140
160
180
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 18 19 21 22 24 28 29
clusters slotentypologie
3
2
1

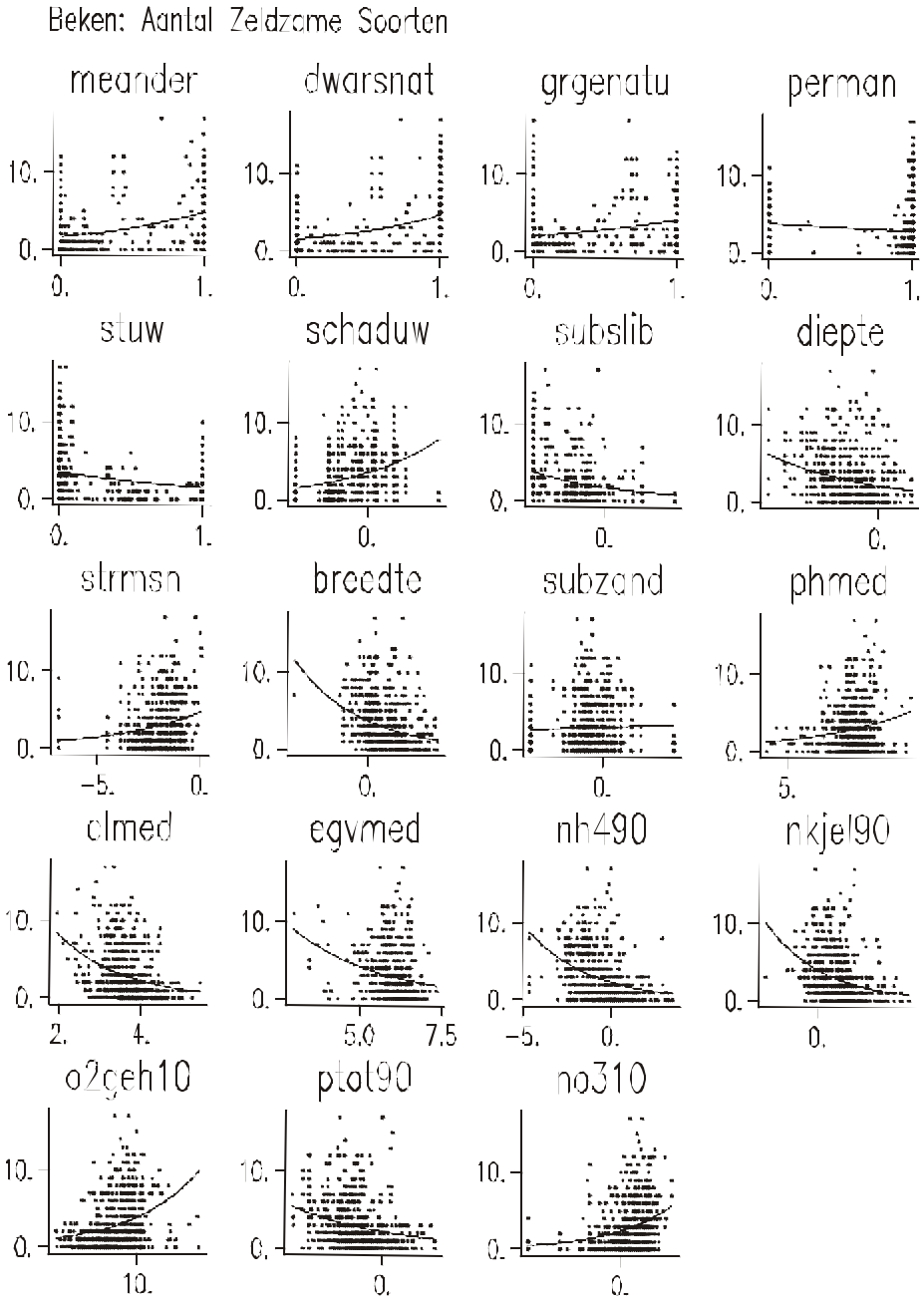

aantal monsters per klasse, ebeoslo maatstaf
brakkarakter macrofauna
0
20
40
60
80
100
120
140
160
180
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 18 19 21 22 24 28 29
clusters slotentypologie
3
2
1


aantal monsters per klasse, ebeoslo maatstaf
gevoeligheid (TOX) macrofauna
0
20
40
60
80
100
120
140
160
180
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 18 19 21 22 24 28 29
clusters slotentypologie
3
2
1


aantal per klasse, ebeoslo maatstaf saprobie macrofauna
0
20
40
60
80
100
120
140
160
180
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 18 19 21 22 24 28 29
clusters slotentypologie
3
2
1


aantal per klasse, ebeoslo maatstaf zuurkarakter macrofauna
0
20
40
60
80
100
120
140
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 18 19 21 22 24 28 29
clusters slotentypologie
3
2
1

Bijlage 16 Univariate analyses voor beken op basis van aantal zeldzame soorten.



Bijlage 17 Modelselectie voor beken op basis van aantal zeldzame soorten.

Beken: Aantal Zeldzame Soorten

Best subsets with 1 term

Adjusted	MeanDev	Df	(1)	(2)	(8)	(10)	(12)	(13)	(14)	(15)	(16)	(19)
23.66	2.1457	2	-	.000	-	-	-	-	-	-	-	-
22.52	2.1776	2	.000	-	-	-	-	-	-	-	-	-
13.25	2.4381	2	-	-	-	-	-	-	-	.000	-	-
10.42	2.5177	2	-	-	-	-	-	-	-	-	.000	-
8.92	2.5600	2	-	-	-	-	-	.000	-	-	-	-
7.80	2.5914	2	-	-	-	.000	-	-	-	-	-	-
7.59	2.5973	2	-	-	-	-	-	-	-	-	-	.000
5.61	2.6530	2	-	-	.000	-	-	-	-	-	-	-

Best subsets with 2 terms

Adjusted	MeanDev	Df	(1)	(2)	(8)	(10)	(12)	(13)	(14)	(15)	(16)	(19)
27.42	2.0400	3	-	.000	-	-	-	.000	-	-	-	-
27.05	2.0504	3	-	.000	-	-	-	-	-	-	.000	-
26.74	2.0591	3	.000	.000	-	-	-	-	-	-	-	-
26.65	2.0617	3	-	.000	-	-	-	-	.000	-	-	-
26.53	2.0649	3	-	.000	-	-	-	-	-	.000	-	-
26.11	2.0767	3	-	.000	-	-	-	-	-	-	-	.000
25.65	2.0898	3	.000	-	-	-	-	-	-	-	-	.000
25.32	2.0990	3	.000	-	-	-	-	-	-	-	.000	-

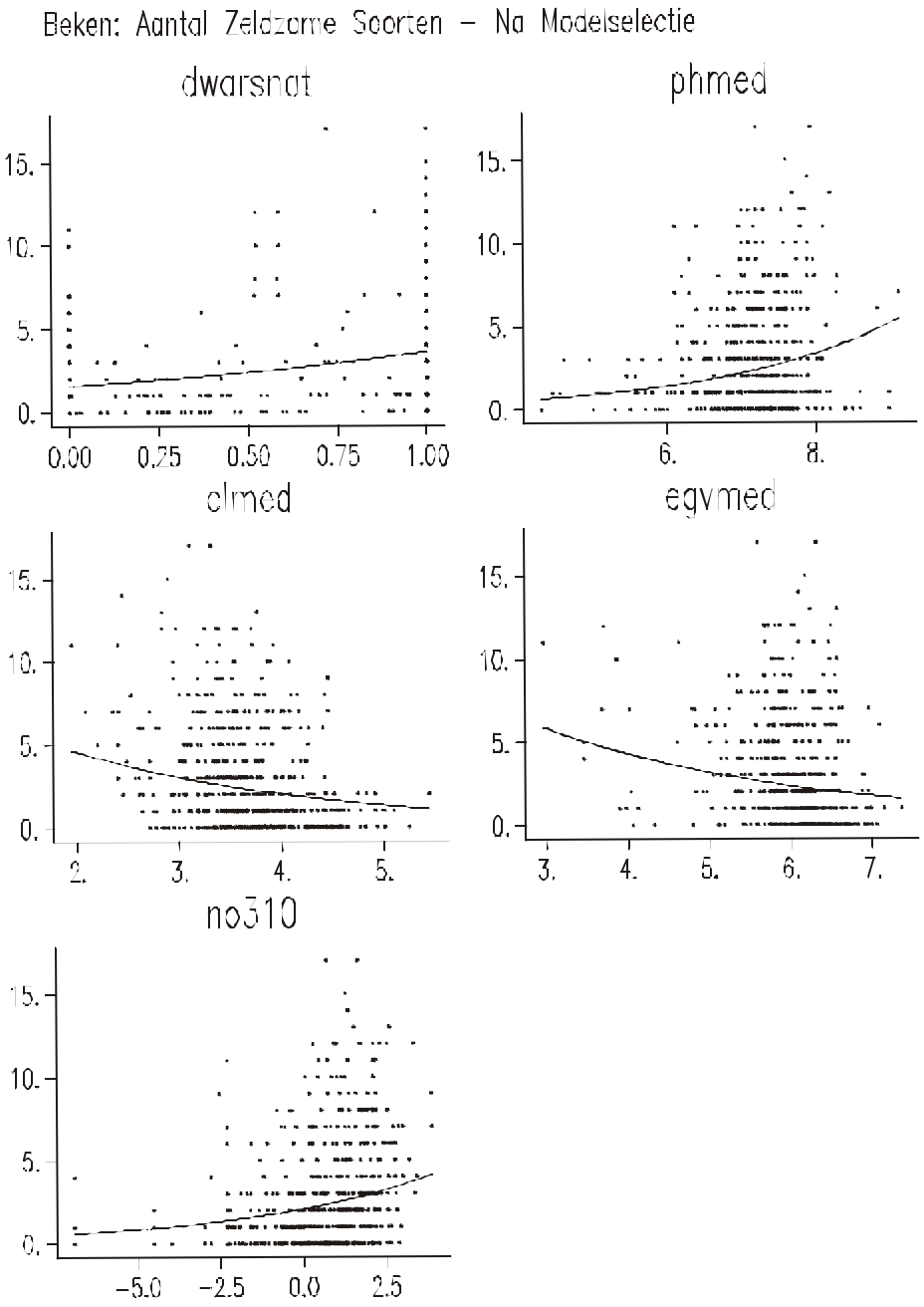
Best subsets with 3 terms

Adjusted	MeanDev	Df	(1)	(2)	(8)	(10)	(12)	(13)	(14)	(15)	(16)	(19)
30.61	1.9503	4	-	.000	-	-	-	.000	-	-	-	.000
29.50	1.9814	4	-	.000	-	-	-	-	-	-	.000	.000
29.43	1.9836	4	-	.000	-	-	-	-	.000	-	-	.000
29.22	1.9892	4	-	.000	-	-	.000	.000	-	-	-	-
29.07	1.9937	4	.000	.000	-	-	-	-	-	-	.000	-
29.01	1.9953	4	-	.000	-	-	.000	-	.000	-	-	-
28.99	1.9958	4	.000	.000	-	-	-	-	-	-	-	.000

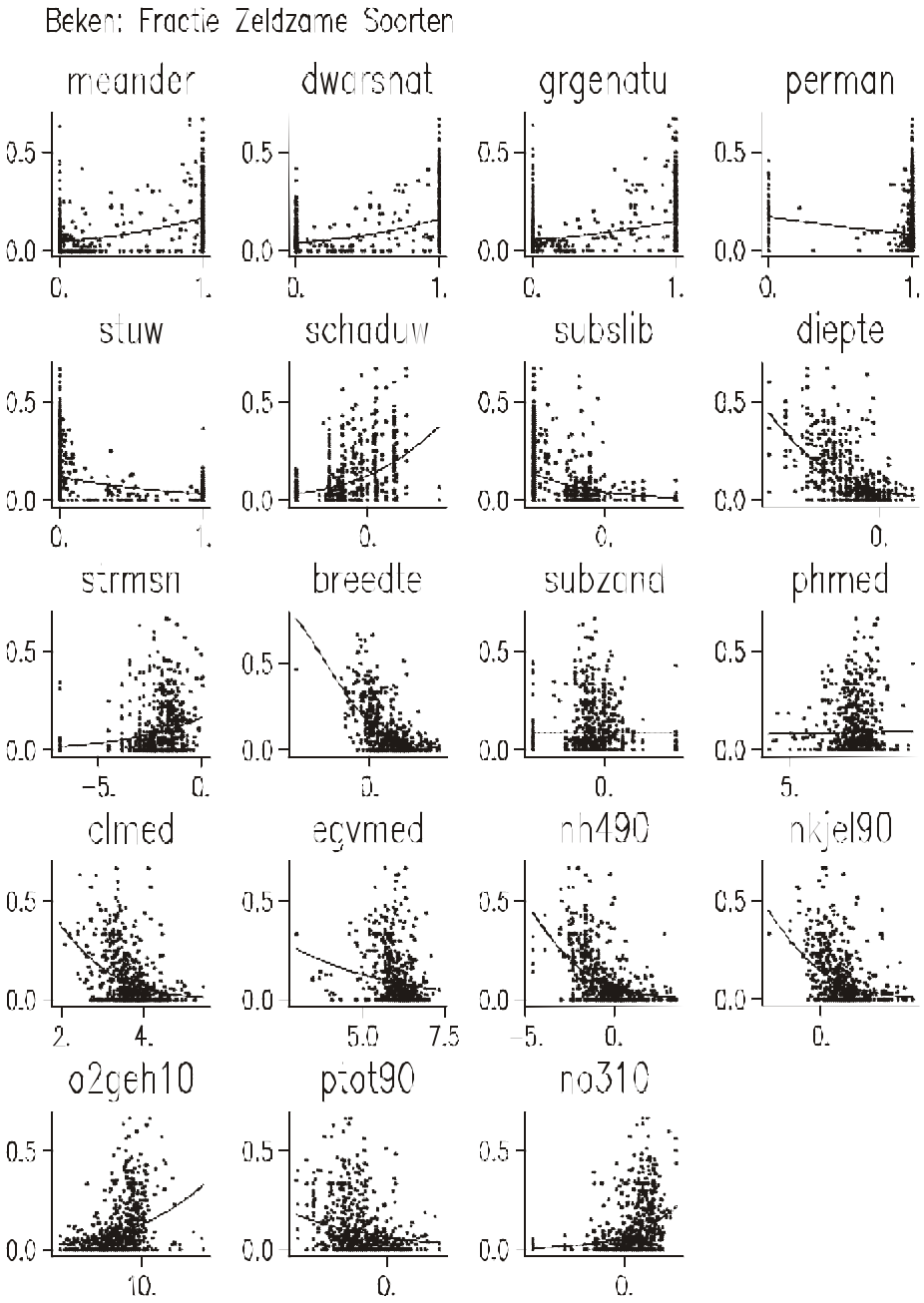
```
     28.89    1.9985   4  .000  .000    -     -     -     -    .000    -     -
-

 Best subsets with 4 terms

  Adjusted   MeanDev  Df   (1)   (2)   (8)  (10)  (12)  (13)  (14)  (15)  (16)
(19)
     33.37    1.8726   5    -   .000    -     -   .000  .000    -     -     -
.000
     32.88    1.8866   5    -   .000    -     -   .000    -   .000    -     -
.000
     31.82    1.9162   5  .001  .000    -     -     -   .000    -     -     -
.000
     31.67    1.9205   5    -   .000    -     -     -   .000    -     -   .002
.000
     31.55    1.9239   5  .000  .000    -     -     -     -   .000    -     -
.000
     31.50    1.9253   5    -   .000    -     -     -   .000  .004    -     -
.000
     31.41    1.9278   5  .000  .000    -     -     -     -     -     -   .000
.000
     31.32    1.9302   5    -   .000    -     -     -   .000    -   .009    -
.000

 Best subsets with 5 terms

  Adjusted   MeanDev  Df   (1)   (2)   (8)  (10)  (12)  (13)  (14)  (15)  (16)
(19)
     35.59    1.8104   6    -   .000    -     -   .000  .000  .000    -     -
.000
     34.72    1.8349   6    -   .000    -     -   .000    -   .000    -   .000
.000
     34.39    1.8442   6    -   .000    -     -   .000    -   .000  .000    -
.000
     34.36    1.8450   6    -   .000    -     -   .000  .000    -     -   .002
.000
     34.28    1.8472   6    -   .000    -     -   .000  .000    -   .003    -
.000
     34.21    1.8492   6  .000  .000    -     -   .000    -   .000    -     -
.000
     33.89    1.8581   6  .021  .000    -     -   .000  .000    -     -     -
.000
     33.36    1.8729   6    -   .000    -   .337  .000  .000    -     -     -
.000

 Best subsets with 6 terms

  Adjusted   MeanDev  Df   (1)   (2)   (8)  (10)  (12)  (13)  (14)  (15)  (16)
(19)
     36.18    1.7937   7    -   .000    -     -   .000  .000  .000    -   .013
.000
     36.17    1.7941   7    -   .000    -     -   .000  .000  .000  .014    -
.000
     35.99    1.7991   7  .034  .000    -     -   .000  .000  .000    -     -
.000
     35.57    1.8109   7  .004  .000    -     -   .000    -   .000    -   .000
.000
     35.57    1.8109   7    -   .000    -   .356  .000  .000  .000    -     -
.000
     35.53    1.8121   7    -   .000  .481    -   .000  .000  .000    -     -
.000
     35.05    1.8256   7  .010  .000    -     -   .000    -   .000  .004    -
.000
     34.82    1.8321   7  .027  .000    -     -   .000  .000    -     -   .003
.000

 Best subsets with 7 terms

  Adjusted   MeanDev  Df   (1)   (2)   (8)  (10)  (12)  (13)  (14)  (15)  (16)
(19)
```

```
     36.54    1.7837   8  .043  .000    -     -    .000  .002  .000    -    .016
.000
     36.42    1.7870   8    -   .000  .074    -    .000  .000  .000  .003    -
.000
     36.38    1.7881   8  .092  .000    -     -    .000  .000  .000  .036    -
.000
     36.28    1.7910   8    -   .000  .174    -    .000  .000  .000    -    .006
.000
     36.17    1.7939   8    -   .000    -     -    .000  .000  .000  .337  .308
.000
     36.10    1.7960   8    -   .000    -   .587   .000  .000  .000    -    .018
.000
     36.05    1.7973   8    -   .000    -   .938   .000  .000  .000  .023    -
.000
     36.02    1.7982   8  .022  .000  .262    -    .000  .000  .000    -     -
.000

 Best subsets with 8 terms

  Adjusted    MeanDev  Df   (1)   (2)   (8)  (10)  (12)  (13)  (14)  (15)  (16)
(19)
     36.78    1.7770   9  .021  .000  .080    -    .000  .002  .000    -    .006
.000
     36.73    1.7783   9  .054  .000  .044    -    .000  .000  .000  .007    -
.000
     36.63    1.7811   9    -   .000  .018  .044  .000  .000  .000    -    .005
.000
     36.61    1.7818   9    -   .000  .016  .105  .000  .000  .000  .006    -
.000
     36.46    1.7860   9  .063  .000    -     -    .000  .002  .000  .606  .200
.000
     36.43    1.7868   9  .050  .000    -   .883   .000  .002  .000    -    .019
.000
     36.40    1.7874   9    -   .000  .083    -    .000  .000  .000  .149  .352
.000
     36.27    1.7912   9  .091  .000    -   .862   .000  .000  .000  .042    -
.000

 Best subsets with 9 terms

  Adjusted    MeanDev  Df   (1)   (2)   (8)  (10)  (12)  (13)  (14)  (15)  (16)
(19)
     37.03    1.7698  10  .034  .000  .012  .073  .000  .004  .000    -    .005
.000
     36.86    1.7746  10  .072  .000  .013  .143  .000  .001  .000  .012    -
.000
     36.79    1.7767  10  .037  .000  .049    -    .000  .001  .000  .297  .222
.000
     36.65    1.7805  10    -   .000  .013  .076  .000  .000  .000  .275  .236
.000
     36.34    1.7892  10  .066  .000    -   .990  .000  .002  .000  .621  .204
.000
     36.08    1.7965  10  .006  .000  .007  .044  .000    -   .000  .633  .021
.000
     34.95    1.8283  10  .034  .000  .052  .263  .000  .000    -   .435  .082
.000
     34.03    1.8541  10  .001  .000  .005  .254    -   .020  .004  .708  .046
.000

 Best subsets with 10 terms

  Adjusted    MeanDev  Df   (1)   (2)   (8)  (10)  (12)  (13)  (14)  (15)  (16)
(19)
     36.98    1.7712  11  .049  .000  .010  .100  .000  .003  .000  .451  .152
.000
```

Het conditionele effect van de geselecteerde milieuvariabelen voor beken op basis van aantal zeldzame soorten.



Bijlage 18 Univariate analyses voor beken op basis van de fractie zeldzame soorten.



Bijlage 19 Modelselectie voor beken op basis van de fractie zeldzame soorten.

Beken: Fractie Zeldzame Soorten

Best subsets with 1 term

Adjusted	MeanDev	Df	(1)	(2)	(6)	(7)	(10)	(12)	(13)	(14)	(15)	(16)	(19)
31.02	2.4574	2	-	-	-	-	.000	-	-	-	-	-	-
30.67	2.4699	2	-	.000	-	-	-	-	-	-	-	-	-
27.86	2.5700	2	.000	-	-	-	-	-	-	-	-	-	-
24.37	2.6942	2	-	-	-	-	-	-	-	-	.000	-	-
21.44	2.7989	2	-	-	.000	-	-	-	-	-	-	-	-
17.17	2.9510	2	-	-	-	.000	-	-	-	-	-	-	-
16.48	2.9756	2	-	-	-	-	-	-	.000	-	-	-	-
15.36	3.0154	2	-	-	-	-	-	-	-	-	-	.000	-

Best subsets with 2 terms

Adjusted	MeanDev	Df	(1)	(2)	(6)	(7)	(10)	(12)	(13)	(14)	(15)	(16)	(19)
46.22	1.9160	3	-	.000	-	-	.000	-	-	-	-	-	-
44.19	1.9882	3	.000	-	-	-	.000	-	-	-	-	-	-
40.89	2.1057	3	-	.000	-	-	-	-	-	-	.000	-	-
39.99	2.1380	3	-	.000	-	-	-	-	.000	-	-	-	-
38.83	2.1792	3	-	-	-	-	.000	-	-	-	.000	-	-
38.71	2.1834	3	-	.000	.000	-	-	-	-	-	-	-	-
38.48	2.1917	3	-	.000	-	-	-	-	-	-	-	.000	-
37.72	2.2189	3	-	-	-	-	.000	-	-	-	-	.000	-

Best subsets with 3 terms

Adjusted	MeanDev	Df	(1)	(2)	(6)	(7)	(10)	(12)	(13)	(14)	(15)	(16)	(19)
50.81	1.7525	4	-	.000	-	-	.000	-	.000	-	-	-	-
49.81	1.7880	4	-	.000	-	-	.000	-	-	-	-	.000	-
49.27	1.8072	4	-	.000	-	-	.000	-	-	-	.000	-	-
48.73	1.8265	4	-	.000	-	-	.000	-	-	-	-	-	.000
48.17	1.8467	4	-	.000	-	-	.000	-	-	.000	-	-	-
48.01	1.8520	4	-	.000	-	-	-	-	.000	-	-	-	.000
47.99	1.8528	4	-	.000	.000	-	.000	-	-	-	-	-	-
47.97	1.8536	4	.000	.000	-	-	.000	-	-	-	-	-	-

Best subsets with 4 terms

Adjusted	MeanDev	Df	(1)	(2)	(6)	(7)	(10)	(12)	(13)	(14)	(15)	(16)	(19)
54.90	1.6066	5	-	.000	-	-	.000	-	.000	-	-	-	.000
53.17	1.6685	5	-	.000	.000	-	.000	-	.000	-	-	-	-
52.71	1.6846	5	-	.000	-	-	.000	-	.000	-	.000	-	-
52.63	1.6877	5	-	.000	-	-	.000	-	-	-	-	.000	.000
52.62	1.6878	5	-	.000	-	-	-	-	.000	-	.000	-	.000
52.48	1.6930	5	-	.000	-	-	.000	-	.000	-	-	.000	-
52.43	1.6946	5	-	.000	-	.000	.000	-	.000	-	-	-	-
52.36	1.6973	5	.000	-	-	-	.000	-	.000	-	-	-	.000

Best subsets with 5 terms

Adjusted	MeanDev	Df	(1)	(2)	(6)	(7)	(10)	(12)	(13)	(14)	(15)	(16)	(19)
56.65	1.5443	6	-	.000	-	-	.000	-	.000	-	.000	-	.000
56.61	1.5457	6	-	.000	-	-	.000	-	.000	-	-	.000	.000
56.33	1.5557	6	-	.000	-	-	.000	.000	.000	-	-	-	.000
55.73	1.5773	6	-	.000	.001	-	.000	-	.000	-	-	-	.000
55.58	1.5827	6	.002	.000	-	-	.000	-	.000	-	-	-	.000
55.40	1.5890	6	-	.000	-	.008	.000	-	.000	-	-	-	.000
55.34	1.5910	6	-	.000	-	-	.000	-	.000	.011	-	-	.000
54.75	1.6119	6	-	.000	.000	-	.000	-	.000	-	.000	-	-

Best subsets with 6 terms

Adjusted	MeanDev	Df	(1)	(2)	(6)	(7)	(10)	(12)	(13)	(14)	(15)	(16)	(19)
57.88	1.5006	7	-	.000	-	-	.000	.000	.000	-	.000	-	.000
57.78	1.5041	7	-	.000	-	-	.000	.000	.000	-	-	.000	.000
57.39	1.5182	7	-	.000	-	-	.000	.000	.000	.000	-	-	.000
57.33	1.5203	7	-	.000	.002	-	.000	-	.000	-	.000	-	.000
57.25	1.5231	7	-	.000	.002	-	.000	-	.000	-	-	.000	.000
57.15	1.5267	7	.005	.000	-	-	.000	-	.000	-	-	.000	.000
57.04	1.5306	7	-	.000	-	.015	.000	-	.000	-	.000	-	.000
57.02	1.5311	7	.016	.000	-	-	.000	-	.000	-	.000	-	.000

Best subsets with 7 terms

Adjusted	MeanDev	Df	(1)	(2)	(6)	(7)	(10)	(12)	(13)	(14)	(15)	(16)	(19)
58.65	1.4732	8	-	.000	-	-	.000	.000	.000	.001	.000	-	.000
58.50	1.4784	8	-	.000	-	-	.000	.000	.000	.001	-	.000	.000
58.40	1.4822	8	-	.000	.005	-	.000	.000	.000	-	.000	-	.000
58.33	1.4845	8	-	.000	-	.008	.000	.000	.000	-	.000	-	.000
58.28	1.4864	8	-	.000	.006	-	.000	.000	.000	-	-	.000	.000
58.17	1.4903	8	-	.000	-	.013	.000	.000	.000	-	-	.000	.000
58.01	1.4960	8	.046	.000	-	-	.000	.000	.000	-	-	.000	.000
57.99	1.4967	8	-	.000	-	-	.000	.000	.000	-	.053	.118	.000

Best subsets with 8 terms

Adjusted	MeanDev	Df	(1)	(2)	(6)	(7)	(10)	(12)	(13)	(14)	(15)	(16)	(19)
59.03	1.4597	9	-	.000	.014	-	.000	.000	.000	.002	.000	-	.000
58.93	1.4631	9	-	.000	-	.029	.000	.000	.000	.003	.000	-	.000
58.87	1.4654	9	-	.000	.015	-	.000	.000	.000	.003	-	.000	.000
58.75	1.4697	9	.127	.000	-	-	.000	.000	.000	.001	.000	-	.000
58.74	1.4700	9	-	.000	-	.042	.000	.000	.000	.003	-	.000	.000
58.71	1.4711	9	.054	.000	-	-	.000	.000	.000	.001	-	.000	.000
58.70	1.4715	9	-	.000	-	-	.000	.000	.000	.001	.058	.199	.000
58.60	1.4749	9	-	.000	.032	.053	.000	.000	.000	-	.000	-	.000

Best subsets with 9 terms

Adjusted	MeanDev	Df	(1)	(2)	(6)	(7)	(10)	(12)	(13)	(14)	(15)	(16)	(19)
59.14	1.4558	10	-	.000	.052	.115	.000	.000	.000	.004	.000	-	.000
59.06	1.4585	10	-	.000	.015	-	.000	.000	.000	.003	.058	.228	.000
59.05	1.4587	10	.240	.000	.024	-	.000	.000	.000	.002	.000	-	.000
58.98	1.4613	10	.191	.000	-	.041	.000	.000	.000	.003	.000	-	.000
58.98	1.4615	10	.115	.000	.031	-	.000	.000	.000	.003	-	.000	.000
58.95	1.4625	10	-	.000	-	.036	.000	.000	.000	.004	.050	.263	.000
58.95	1.4626	10	-	.000	.051	.153	.000	.000	.000	.005	-	.000	.000
58.88	1.4648	10	.085	.000	-	.066	.000	.000	.000	.003	-	.000	.000

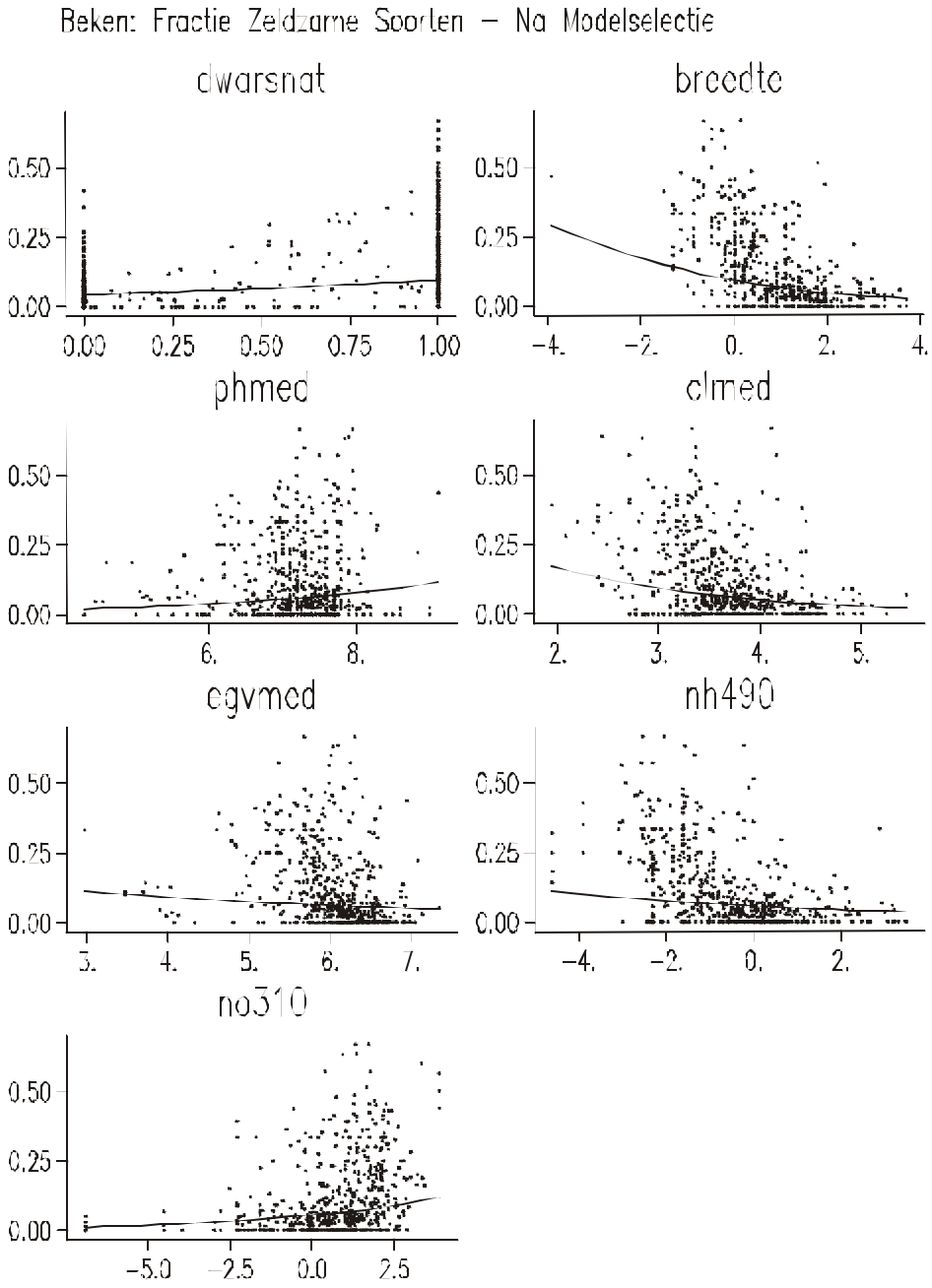
Best subsets with 10 terms

Adjusted	MeanDev	Df	(1)	(2)	(6)	(7)	(10)	(12)	(13)	(14)	(15)	(16)	(19)
59.15	1.4553	11	-	.000	.054	.135	.000	.000	.000	.006	.052	.272	.000
59.15	1.4555	11	.288	.000	.074	.136	.000	.000	.000	.004	.000	-	.000
59.11	1.4566	11	.190	.000	.028	-	.000	.000	.000	.003	.093	.182	.000
59.03	1.4595	11	.140	.000	.084	.187	.000	.000	.000	.005	-	.000	.000
59.03	1.4597	11	.151	.000	-	.055	.000	.000	.000	.004	.087	.205	.000
58.68	1.4719	11	.244	.000	.051	.086	.000	.000	.000	-	.076	.149	.000
57.91	1.4995	11	.040	.000	.038	.226	.000	-	.000	.105	.106	.106	.000
56.78	1.5398	11	.000	-	.028	.080	.000	.000	.000	.007	.030	.327	.000

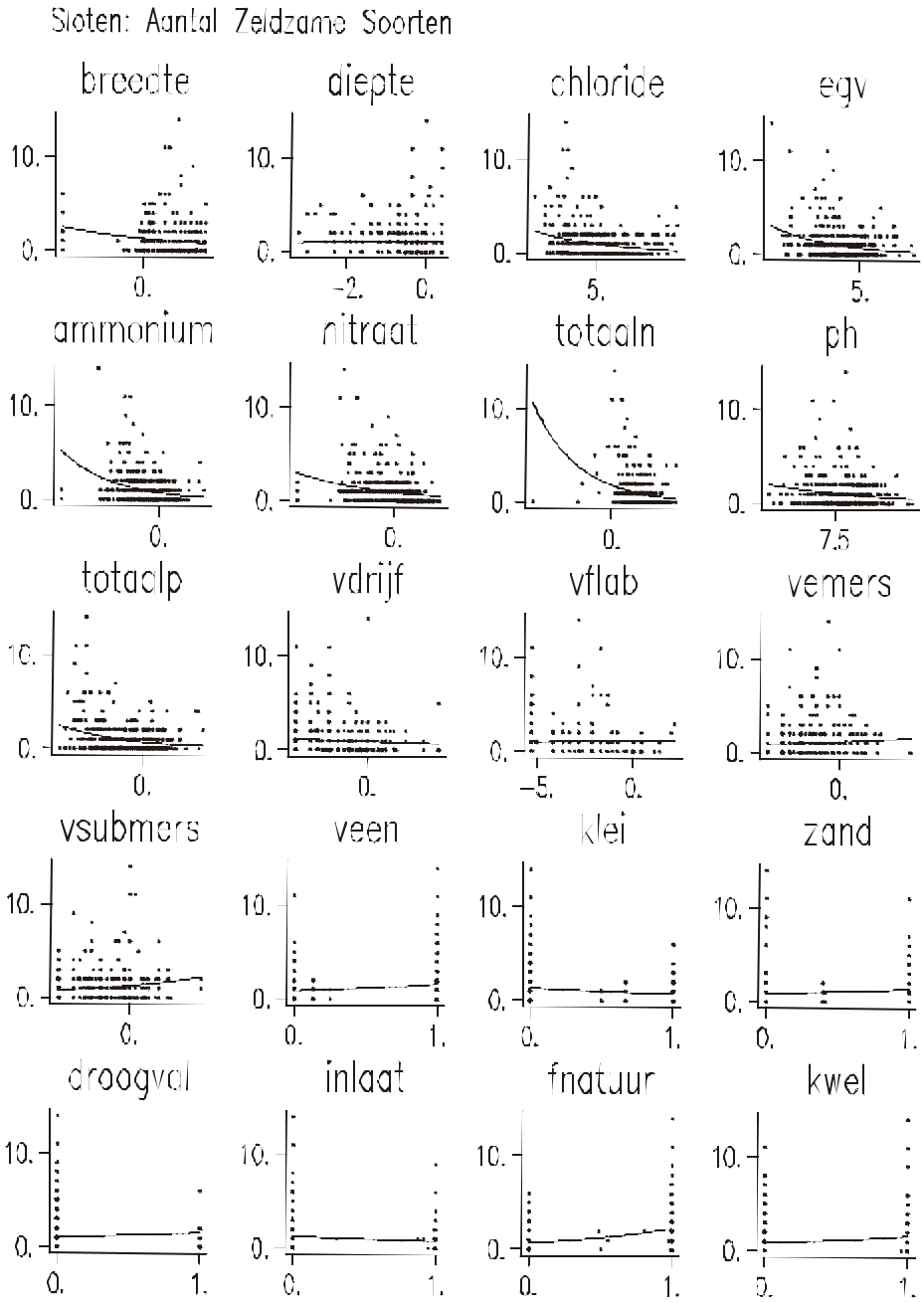
Best subsets with 11 terms

Adjusted	MeanDev	Df	(1)	(2)	(6)	(7)	(10)	(12)	(13)	(14)	(15)	(16)	(19)
59.18	1.4542	12	.234	.000	.080	.164	.000	.000	.000	.006	.082	.221	.000

Bijlage 20 Het conditionele effect van de geselecteerde milieuvariabelen voor beken op basis van de fractie zeldzame soorten.



Bijlage 21 Univariate analyses voor sloten op basis van aantal zeldzame soorten.



Bijlage 22 Modelselectie voor sloten op basis van aantal zeldzame soorten.

Sloten: Aantal Zeldzame Soorten - Na modelselctie

Best subsets with 1 term

Adjusted	MeanDev	Df	(1)	(2)	(3)	(4)	(5)	(7)	(10)	(13)	(15)	(19)	(20)
16.11	1.6335	2	-	-	-	-	-	-	-	-	-	.000	-
8.89	1.7741	2	-	-	-	-	.000	-	-	-	-	-	-
8.59	1.7801	2	-	-	-	-	-	.000	-	-	-	-	-
7.22	1.8067	2	-	-	-	.000	-	-	-	-	-	-	-
6.50	1.8208	2	-	-	.000	-	-	-	-	-	-	-	-
4.26	1.8644	2	-	-	-	-	-	-	-	-	-	-	.000
3.85	1.8724	2	-	-	-	-	-	-	-	-	.000	-	-
2.33	1.9020	2	-	-	-	-	-	-	-	.001	-	-	-

Best subsets with 2 terms

Adjusted	MeanDev	Df	(1)	(2)	(3)	(4)	(5)	(7)	(10)	(13)	(15)	(19)	(20)
20.49	1.5482	3	-	-	-	-	-	.000	-	-	-	.000	-
19.39	1.5697	3	-	-	-	-	.000	-	-	-	-	.000	-
18.74	1.5823	3	-	-	-	.000	-	-	-	-	-	.000	-
18.04	1.5959	3	-	-	.001	-	-	-	-	-	-	.000	-
18.00	1.5968	3	-	-	-	-	-	-	-	.001	-	.000	-
17.29	1.6107	3	-	-	-	-	-	-	-	-	-	.000	.010
17.23	1.6118	3	-	-	-	-	-	-	-	-	.011	.000	-
16.53	1.6253	3	.082	-	-	-	-	-	-	-	-	.000	-

Best subsets with 3 terms

Adjusted	MeanDev	Df	(1)	(2)	(3)	(4)	(5)	(7)	(10)	(13)	(15)	(19)	(20)
21.70	1.5248	4	-	-	-	-	-	.000	-	-	.007	.000	-
21.53	1.5281	4	-	-	-	.012	-	.000	-	-	-	.000	-
21.42	1.5302	4	-	-	-	-	.017	.001	-	-	-	.000	-
21.38	1.5309	4	-	-	-	-	-	.000	-	.019	-	.000	-
21.28	1.5329	4	-	-	.025	-	-	.000	-	-	-	.000	-
21.08	1.5368	4	-	-	-	-	-	.000	-	-	-	.000	.046
21.02	1.5379	4	-	-	-	-	-	.000	.054	-	-	.000	-
20.87	1.5408	4	.086	-	-	-	-	.000	-	-	-	.000	-

Best subsets with 4 terms

Adjusted	MeanDev	Df	(1)	(2)	(3)	(4)	(5)	(7)	(10)	(13)	(15)	(19)	(20)
22.68	1.5056	5	-	-	-	-	.014	.001	-	-	.006	.000	-
22.66	1.5060	5	-	-	-	-	-	.000	-	.014	.006	.000	-
22.44	1.5102	5	-	-	-	-	-	.000	.011	.004	-	.000	-
22.21	1.5148	5	-	-	-	.022	-	.001	-	.034	-	.000	-
22.21	1.5149	5	-	-	-	.008	-	.000	.034	-	-	.000	-
22.20	1.5150	5	.004	.005	-	-	-	.000	-	-	-	.000	-
22.19	1.5151	5	-	-	.025	-	.017	.004	-	-	-	.000	-
22.19	1.5151	5	-	-	-	-	-	.000	-	-	.009	.000	.059

Best subsets with 5 terms

Adjusted	MeanDev	Df	(1)	(2)	(3)	(4)	(5)	(7)	(10)	(13)	(15)	(19)	(20)
23.55	1.4887	6	-	-	-	-	-	.000	.017	.004	.009	.000	-
23.50	1.4896	6	-	-	-	-	.020	.003	-	.022	.005	.000	-
23.42	1.4913	6	-	-	-	.014	-	.001	.007	.007	-	.000	-
23.27	1.4940	6	.004	.008	-	-	-	.000	-	-	.010	.000	-
23.20	1.4955	6	-	-	-	-	.027	.003	.011	.007	-	.000	-
23.19	1.4957	6	-	.013	.004	-	.012	.011	-	-	-	.000	-
23.18	1.4959	6	.008	.003	-	.014	-	.000	-	-	-	.000	-
23.14	1.4966	6	-	-	-	-	.012	.001	.065	-	.008	.000	-

Best subsets with 6 terms

Adjusted	MeanDev	Df	(1)	(2)	(3)	(4)	(5)	(7)	(10)	(13)	(15)	(19)	(20)
24.40	1.4721	7	-	-	-	-	.019	.003	.017	.006	.007	.000	-
24.10	1.4780	7	.007	.009	-	-	.021	.001	-	-	.008	.000	-
24.04	1.4792	7	.008	.009	-	-	-	.000	-	.026	.008	.000	-
24.00	1.4799	7	-	-	-	.067	-	.000	.012	.005	.044	.000	-
23.95	1.4809	7	.005	.003	-	-	-	.000	-	-	.016	.000	.033
23.94	1.4812	7	-	-	-	-	-	.000	.016	.005	.011	.000	.082
23.93	1.4812	7	.082	-	-	-	-	.000	.013	.006	.008	.000	-
23.93	1.4812	7	-	-	-	.028	.054	.017	.007	.011	-	.000	-

Best subsets with 7 terms

Adjusted	MeanDev	Df	(1)	(2)	(3)	(4)	(5)	(7)	(10)	(13)	(15)	(19)	(20)
24.79	1.4645	8	.008	.019	-	-	-	.000	.025	.008	.013	.000	-
24.75	1.4653	8	.013	.010	-	-	.029	.005	-	.035	.007	.000	-
24.70	1.4663	8	.108	-	-	-	.025	.003	.012	.008	.006	.000	-
24.70	1.4663	8	-	-	.108	-	.017	.010	.015	.009	.022	.000	-
24.70	1.4664	8	.014	.009	-	.017	-	.001	.012	.015	-	.000	-
24.66	1.4670	8	-	-	-	-	.028	.005	.014	.008	.008	.000	.122
24.64	1.4674	8	-	-	-	.131	.036	.012	.012	.008	.030	.000	-
24.64	1.4675	8	-	.027	.007	-	.018	.026	.019	.014	-	.000	-

Best subsets with 8 terms

Adjusted	MeanDev	Df	(1)	(2)	(3)	(4)	(5)	(7)	(10)	(13)	(15)	(19)	(20)
25.52	1.4504	9	.012	.021	-	-	.028	.005	.024	.011	.009	.000	-
25.39	1.4528	9	.008	.007	-	-	-	.000	.021	.011	.020	.000	.041
25.21	1.4564	9	.012	.014	-	.073	-	.001	.019	.012	.054	.000	-
25.16	1.4573	9	.052	.005	.028	-	.028	.022	.018	.020	-	.000	-
25.15	1.4576	9	-	.066	.038	-	.014	.020	.026	.011	.054	.000	-
25.14	1.4577	9	.014	.005	-	-	.046	.009	-	.047	.010	.000	.079
25.12	1.4580	9	.020	.010	-	.031	.070	.025	.011	.021	-	.000	-
25.11	1.4583	9	.014	.004	-	.047	-	.002	.011	.018	-	.000	.074

Best subsets with 9 terms

Adjusted	MeanDev	Df	(1)	(2)	(3)	(4)	(5)	(7)	(10)	(13)	(15)	(19)	(20)
25.98	1.4415	10	.013	.008	-	-	.042	.009	.020	.015	.014	.000	.063
25.78	1.4452	10	.036	.011	.119	-	.023	.017	.025	.016	.037	.000	-
25.75	1.4458	10	.016	.016	-	.135	.049	.018	.018	.016	.037	.000	-
25.58	1.4492	10	.012	.006	-	.158	-	.001	.018	.015	.062	.000	.085
25.47	1.4514	10	.044	.002	.071	-	.045	.026	.016	.022	-	.000	.105
25.46	1.4515	10	.020	.004	-	.072	.090	.031	.011	.024	-	.000	.095
25.42	1.4523	10	.019	.005	.288	-	-	.000	.023	.013	.053	.000	.072
25.37	1.4533	10	-	.040	.085	-	.022	.023	.023	.012	.060	.000	.141

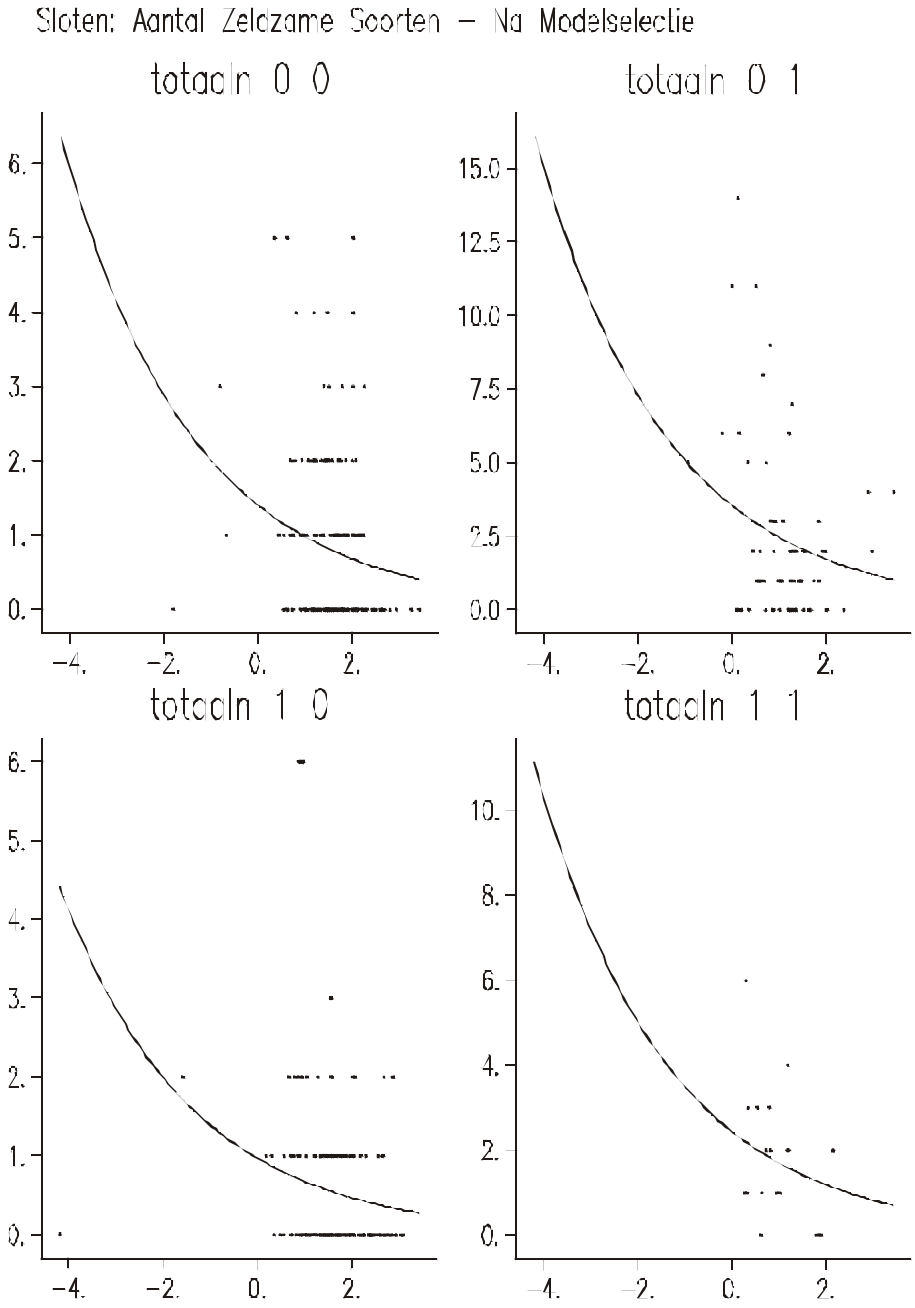
Best subsets with 10 terms

Adjusted	MeanDev	Df	(1)	(2)	(3)	(4)	(5)	(7)	(10)	(13)	(15)	(19)	(20)
26.05	1.4400	11	.031	.006	.236	-	.036	.021	.021	.018	.042	.000	.120
26.04	1.4401	11	.017	.008	-	.244	.062	.024	.016	.019	.043	.000	.109
25.74	1.4460	11	.033	.011	.332	.384	.038	.026	.021	.018	.058	.000	-
25.49	1.4510	11	.040	.003	.287	.291	.069	.041	.013	.026	-	.000	.132
25.44	1.4519	11	.018	.006	.613	.290	-	.001	.019	.015	.081	.000	.099
25.26	1.4554	11	.031	.003	.326	.284	.002	-	.017	.008	.085	.000	.122
25.26	1.4554	11	-	.044	.212	.525	.032	.032	.020	.013	.084	.000	.159
25.13	1.4579	11	.020	.004	.412	.400	.039	.012	.055	-	.081	.000	.127

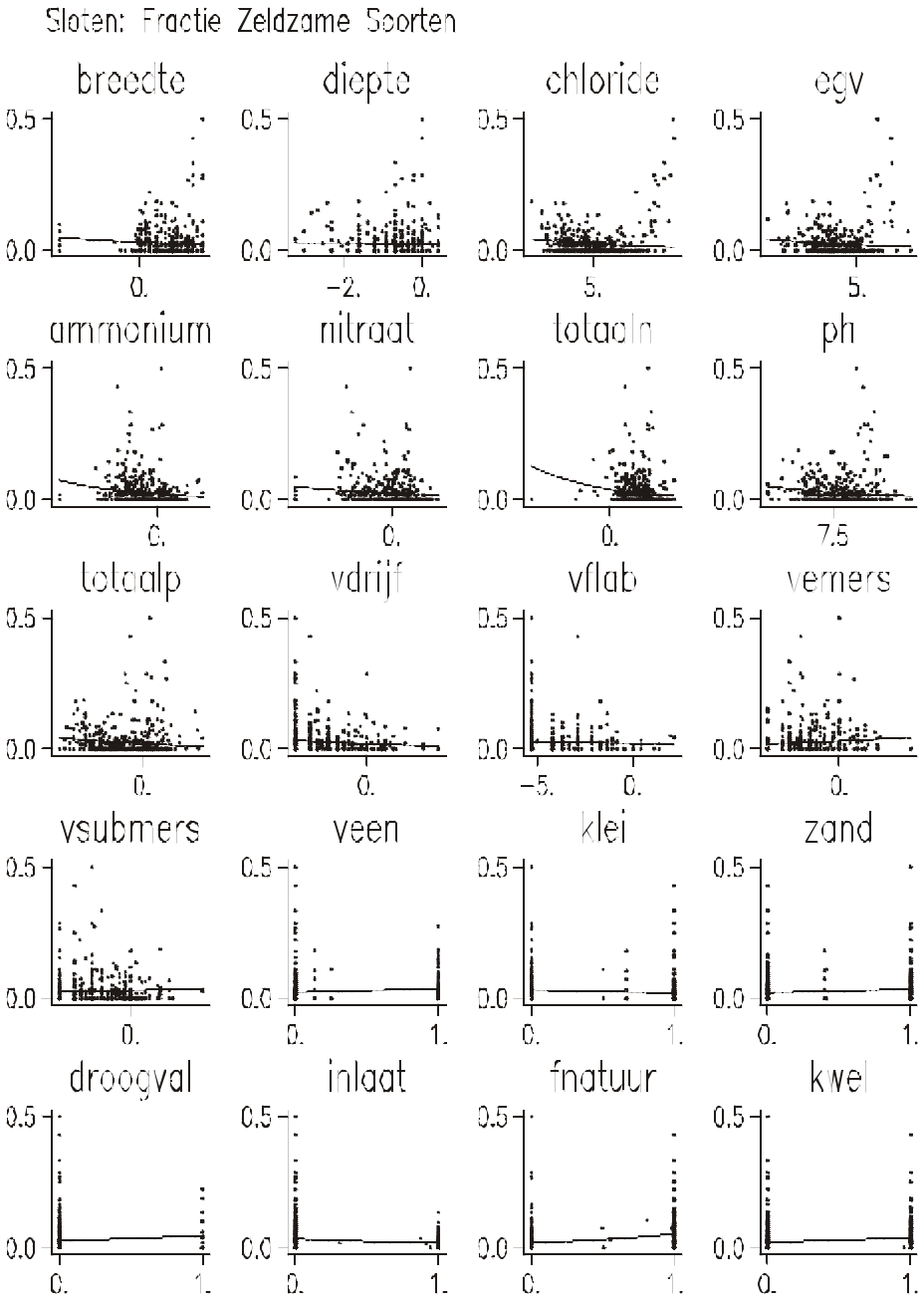
Best subsets with 11 terms

Adjusted	MeanDev	Df	(1)	(2)	(3)	(4)	(5)	(7)	(10)	(13)	(15)	(19)	(20)
25.96	1.4418	12	.030	.006	.456	.475	.052	.030	.019	.020	.061	.000	.140

Bijlage 23 Het conditionele effect van de geselecteerde milieuvariabelen voor sloten op basis van aantal zeldzame soorten.



Bijlage 24 Univariate analyses voor sloten op basis van de fractie zeldzame soorten.



Bijlage 25 Modelselectie voor sloten op basis van de fractie zeldzame soorten.

Sloten: Fractie Zeldzame Soorten

Best subsets with 1 term

Adjusted	MeanDev	Df	(3)	(5)	(7)	(8)	(10)	(14)	(16)	(18)	(19)	(20)
14.08	1.5429	2	-	-	-	-	-	-	-	-	.000	-
4.95	1.7069	2	-	-	-	-	-	-	-	.000	-	-
4.24	1.7196	2	-	.000	-	-	-	-	-	-	-	-
4.18	1.7208	2	-	-	-	-	-	-	-	-	-	.000
4.04	1.7232	2	-	-	.000	-	-	-	-	-	-	-
2.97	1.7424	2	-	-	-	-	-	.000	-	-	-	-
2.66	1.7481	2	-	-	-	.001	-	-	-	-	-	-
2.50	1.7510	2	-	-	-	-	-	-	.001	-	-	-

Best subsets with 2 terms

Adjusted	MeanDev	Df	(3)	(5)	(7)	(8)	(10)	(14)	(16)	(18)	(19)	(20)
15.73	1.5134	3	-	-	-	-	.003	-	-	-	.000	-
15.61	1.5154	3	-	-	-	-	-	-	-	-	.000	.004
15.55	1.5165	3	-	-	.005	-	-	-	-	-	.000	-
15.31	1.5209	3	-	.009	-	-	-	-	-	-	.000	-
15.17	1.5233	3	-	-	-	-	-	-	-	.013	.000	-
14.86	1.5289	3	-	-	-	-	-	-	.031	-	.000	-
14.53	1.5349	3	-	-	-	.078	-	-	-	-	.000	-
14.09	1.5427	3	-	-	-	-	-	.306	-	-	.000	-

Best subsets with 3 terms

Adjusted	MeanDev	Df	(3)	(5)	(7)	(8)	(10)	(14)	(16)	(18)	(19)	(20)
17.42	1.4830	4	-	-	-	-	.002	-	-	-	.000	.002
17.17	1.4874	4	-	-	.005	-	.003	-	-	-	.000	-
17.01	1.4903	4	-	.007	-	-	.002	-	-	-	.000	-
16.76	1.4948	4	-	-	.011	-	-	-	-	-	.000	.009
16.75	1.4951	4	-	-	-	-	-	-	-	.011	.000	.003
16.69	1.4961	4	-	-	.004	-	-	-	-	.011	.000	-
16.63	1.4972	4	-	-	-	-	.005	-	-	.021	.000	-

```
     16.49    1.4996   4     -    .022     -      -      -      -      -      -    .000
.010

 Best subsets with 4 terms

  Adjusted    MeanDev  Df    (3)    (5)    (7)    (8)   (10)   (14)   (16)   (18)   (19)
(20)
     18.56    1.4626   5     -      -    .010     -    .002     -      -      -    .000
.005
     18.39    1.4655   5     -    .016     -      -    .001     -      -      -    .000
.005
     18.38    1.4657   5     -      -      -      -    .003     -      -    .017  .000
.002
     18.13    1.4702   5     -      -    .004     -    .005     -      -    .017  .000
-
     17.88    1.4748   5     -      -    .011     -      -      -      -    .011  .000
.009
     17.82    1.4757   5     -    .009     -      -    .004     -      -    .026  .000
-
     17.69    1.4781   5     -    .004     -      -    .007     -    .038     -   .000
-
     17.61    1.4796   5     -      -      -      -    .005     -    .167     -   .000
.004

 Best subsets with 5 terms

  Adjusted    MeanDev  Df    (3)    (5)    (7)    (8)   (10)   (14)   (16)   (18)   (19)
(20)
     19.50    1.4456   6     -      -    .010     -    .003     -      -    .017  .000
.005
     19.24    1.4502   6     -    .022     -      -    .002     -      -    .022  .000
.005
     18.78    1.4585   6     -    .009     -      -    .004     -    .087     -   .000
.012
     18.72    1.4597   6     -    .181   .107     -    .002     -      -      -   .000
.007
     18.67    1.4605   6   .209     -    .006     -    .001     -      -      -   .000
.003
     18.62    1.4614   6     -      -    .015     -    .004     -    .251     -   .000
.008
     18.50    1.4635   6     -      -      -      -    .006     -    .206   .020  .000
.004
     18.49    1.4638   6     -      -    .014     -    .001   .412     -      -   .000
.006

 Best subsets with 6 terms

  Adjusted    MeanDev  Df    (3)    (5)    (7)    (8)   (10)   (14)   (16)   (18)   (19)
(20)
     19.60    1.4438   7     -    .224   .096     -    .002     -      -    .020  .000
.007
     19.55    1.4447   7     -    .013     -      -    .005     -    .113   .029  .000
.011
     19.55    1.4448   7   .271     -    .006     -    .002     -      -    .021  .000
.003
     19.52    1.4452   7     -      -    .014     -    .006     -    .293   .019  .000
.008
     19.46    1.4464   7     -      -    .014     -    .002   .383     -    .016  .000
.005
     19.34    1.4485   7     -      -    .010   .656   .003     -      -    .021  .000
.014
     19.17    1.4515   7   .424   .018     -      -    .002     -      -    .026  .000
.003
     19.16    1.4517   7     -    .017     -    .447   .003     -      -    .031  .000
.019

 Best subsets with 7 terms

  Adjusted    MeanDev  Df    (3)    (5)    (7)    (8)   (10)   (14)   (16)   (18)   (19)
(20)
```

```
     19.78    1.4405   8     -    .023     -      -    .005   .142   .042   .026   .000
.014
     19.78    1.4406   8  .131     -   .007     -    .006     -    .140   .028   .000
.004
     19.76    1.4410   8  .155  .007     -      -    .005     -    .048   .042   .000
.005
     19.75    1.4411   8     -  .145  .157     -    .005     -    .185   .025   .000
.013
     19.74    1.4413   8     -     -   .026     -    .005   .151   .121   .018   .000
.012
     19.71    1.4419   8  .032  .010     -      -    .003   .048   .005     -    .000
.004
     19.70    1.4420   8  .096     -   .004   .184   .003     -      -    .043   .000
.010
     19.64    1.4430   8  .269  .222   .068     -    .002     -      -    .025   .000
.004

 Best subsets with 8 terms

  Adjusted    MeanDev  Df    (3)    (5)    (7)    (8)   (10)   (14)   (16)   (18)   (19)
(20)
     20.37    1.4300   9  .042     -   .012     -    .005   .048   .025   .030   .000
.004
     20.34    1.4305   9  .052  .013     -      -    .005   .048   .008   .042   .000
.004
     20.13    1.4343   9  .006  .003     -    .078   .005   .049   .007     -    .000
.020
     20.07    1.4353   9  .107  .118   .109     -    .005     -    .077   .038   .000
.005
     19.97    1.4371   9  .065  .124   .057   .104   .003     -      -    .061   .000
.017
     19.95    1.4375   9  .050  .003     -    .160   .007     -    .061   .085   .000
.020
     19.93    1.4378   9  .022  .111   .145     -    .004   .056   .009     -    .000
.004
     19.93    1.4378   9  .006     -   .006   .111   .005   .042   .024     -    .000
.012

 Best subsets with 9 terms

  Adjusted    MeanDev  Df    (3)    (5)    (7)    (8)   (10)   (14)   (16)   (18)   (19)
(20)
     20.61    1.4257  10  .035   .138   .126     -    .005   .055   .014   .037   .000
.005
     20.53    1.4271  10  .017   .006     -    .163   .006   .049   .010   .084   .000
.018
     20.47    1.4282  10  .018     -    .007   .220   .006   .045   .031   .056   .000
.012
     20.46    1.4284  10  .003   .057   .104   .057   .005   .055   .013     -    .000
.023
     20.33    1.4307  10  .029   .068   .089   .130   .007     -    .096   .081   .000
.023
     20.02    1.4362  10  .012   .044   .081   .025   .012   .057   .013   .085   .000
-
     19.94    1.4377  10  .052   .158   .054   .100   .002   .362     -    .057   .000
.016
     19.72    1.4416  10     -   .167   .196   .884   .005   .190   .101   .025   .000
.029

 Best subsets with 10 terms

  Adjusted    MeanDev  Df    (3)    (5)    (7)    (8)   (10)   (14)   (16)   (18)   (19)
(20)
     20.88    1.4209  11  .009   .082   .098   .127   .006   .054   .017   .079   .000
.022
```

Bijlage 26 Het conditionele effect van de geselecteerde milieuvariabelen voor sloten op basis van de fractie zeldzame soorten.