

Werkdocument

Opdrachtgever:
DG Rijkswaterstaat, RIZA

Achtergronddocument voor aanvulling van het habitatinstrument: 1. Flexibilisering invoer; 2. Onzekerheidsanalyse; 3. Validatie

G.E.M. van der Lee
W.E. Penning
H. Duel
H.F.P. van den Boogaard

juni 2003

Q3433

Inhoud

1	Inleiding	1—1
2	De flexibele voorkant van het habitatinstrument	2—1
2.1	Afstemming van monitoringgegevens en habitatcriteria	2—3
2.2	De relatie tussen ecotopen en habitatbeschikbaarheid	2—3
3	Onzekerheidsanalyse van habitatmodellen	3—1
3.1	Procedure onzekerheidsanalyse	3—1
3.2	Toepassing op schedefonteinkruid.....	3—8
3.3	Resultaten en discussie.....	3—10
4	Valideren van habitatmodellen	4—1
4.1	Valideren van habitatmodellen met event detectie	4—1
4.1.1	Inleiding	4—1
4.1.2	GoF criteria voor events en niet-events.....	4—3
4.1.3	ROC-curve	4—5
4.1.4	AUC: Area under ROC-curve	4—7
4.1.5	Voorbeeld berekening ROC-curve en AUC-waarde	4—7
4.2	Toepassing: event detectie met het habitatmodel voor schedefonteinkruid.....	4—13
4.2.1	Methoden	4—13
4.2.2	Definitie van waargenomen en voorspelde events.....	4—15
4.2.3	ROC curves	4—16
4.2.4	Accuracy, specificity en sensitivity	4—16
4.2.5	Resultaten.....	4—16
4.2.6	Conclusies en aanbevelingen ten aanzien van validatie	4—22
5	Afsluitende discussie en mogelijkheden voor implementatie	5—1
5.1	Implementatie in het habitatinstrument	5—2
5.1.1	Onzekerheidsanalyse met Monte Carlo simulatie.....	5—2
5.1.2	Validatie met event detectie methode.....	5—6
6	Referenties	6—1

Bijlagen

A	Overzicht habitatmodellen	A-1
B	Algemene habitatcriteria voor verschillende soortgroepen en uitwerking tot optimumfuncties	B-1
B.1	Werkwijze.....	B-1
B.2	Evaluatie van soortcriteria	B-1
	B.2.1 Selectie van soorten	B-1
	B.2.2 Habitatfactoren per soortgroep.....	B-2
B.3	Aquatische habitatfactoren.....	B-3
	B.3.1 Paaihabitat.....	B-3
	B.3.2 Aquatisch fourageerhabitat.....	B-4
B.4	Terrestrische habitatfactoren.....	B-5
	B.4.1 Definitie van optimumfuncties.....	B-7
B.5	vergelijking habitatcriteria met ecotoopcriteria (RWES).....	B-8
B.6	Conclusies en aanbevelingen.....	B-9
B.7	Habitat-eisen voorbeeldsoorten.....	B-10
B.8	Optimumfunctie-tabellen	B-14
B.9	RWES-factoren.....	B-22
B.10	Referenties.....	B-24
C	Statistisch model voor onzekerheden	C-1
C.1	Een statistisch model voor onzekerheden.....	C-1
	C.1.1 Gebruikte verdelingsfuncties.....	C-3
	C.1.2 Opmerkingen bij de implementatie in PC-raster.....	C-6
	C.1.3 Trekkingen uit de verdelingen factorindex voor gegeven breekpunten.....	C-7
	C.1.4 Quartiepunten van uitvoerresultaten	C-8

I Inleiding

In het integraal waterbeheer zijn habitatevaluaties een geschikte methode voor de analyse van ecologische ontwikkelingsmogelijkheden van watersystemen. De afgelopen jaren is er veel aandacht voor de betrouwbaarheid van modelresultaten van habitatmodellen. In het kader hiervan is in 1999 een doelsubsidie project gestart met als onderwerp een kwaliteitstoets van het Nederlandse HEP-instrumentarium. In dit project zijn gevoeligheids- en onzekerheidsanalyses uitgevoerd voor habitatmodellen in een toepassing voor het IJsselmeergebied waarbij habitatgeschiktheidsmodellen zijn gekoppeld met ecotoop- en habitatnetwerkmodellen (Van der Lee et al., 2000).

Uit het doelsubsidie project bleek het belang van onzekerheid in parameters van habitatfuncties. De algemene conclusie was dat habitatmodellen zeer nuttig zijn om de beperking van individuele invoervariabelen in kaart te brengen, zowel voor als na inrichtingsscenario's. Met een habitatmodel is niet te voorspellen of een soort daadwerkelijk kan en zal voorkomen, aangezien niet alle relevante factoren meegenomen kunnen worden. De resultaten van het project gaven aanleiding tot een lijst aan aanbevelingen, waarvan de belangrijkste was om met hoge prioriteit te investeren in het toetsen en kalibreren van de modellen met veldgegevens om tot modelverbeteringen te komen. Daarnaast is onderstaande lijst van actiepunten naar voren gekomen:

- Definitiestudie om na te gaan hoe we op basis van de aanwezige componenten tot een generieke tool voor HEP kunnen komen (koppeling ECOMIJ-EKOS-LARCH).
- Ecotoopdefinitie is van groot belang om de koppeling met een ecotoopkaarten-tool te maken om op die manier een flexibele ecotoop-definitie mogelijk te maken (nu nog voor een deel handwerk). Daarnaast is aandacht nodig voor een flexibele voorkant van het habitatinstrument in het algemeen.
- Begin maken met de (software-) bouw van een generieke tool (inclusief de verbetering van bestaande componenten).
- Onzekerheids- en gevoeligheidsanalyse moeten tot een standaardcomponent van de generieke tool worden gemaakt (druk op de knop).
- Invloed van en omgaan met de onzekerheid in data nader beschouwen en een aanpak omschrijven.
- Het werken op verschillende schalen mogelijk maken (in ieder geval moet onderscheid tussen landelijke en regionale schaal worden gerealiseerd).
- Testen van het instrumentarium middels toepassing van de beschikbare instrumenten op een ander gebied dan het IJsselmeergebied;
- Inhoudelijke verbetering van de verschillende modellen (op basis van mening van de experts over gevoeligheid en betrouwbaarheid).
- Kennis en informatie in de modellen waar mogelijk actualiseren (PCB'S in otters en golfwerking op schedefonteinkruid).
- Uitbreiding van de modellen met meerdere soorten: nu hebben de verschillende gebruikte modellen slechts een zeer beperkt aantal overlappende soorten: de beleidsrelevante soorten zouden hiervoor een goed uitgangspunt zijn.

In een aantal andere vervolprojecten op het doelsubsidieproject is reeds een definitiestudie uitgevoerd (Van der Lee, 2003), is het belang van onzekerheid in invoergegevens onderzocht in een case studie voor het Volkerak-Zoommeer (Aarts et al., 2001) en is gestart met de bouw van een nieuw generiek habitatinstrument. In het doelsubsidie project van 2001-2002 zijn drie belangrijke nog openstaande aanbevelingen uitgewerkt. Dit zijn:

1. Het vergroten van de flexibiliteit ten aanzien van de modelinvoer door het zoeken van een alternatief voor gefixeerde ecotopen.
2. Een algemene aanpak opstellen voor het uitvoeren van onzekerheidsanalyses van habitatmodellen (onzekerheden in invoergegevens en modelparameters) en de implementatie ervan in een generiek prototype instrument.
3. Het formuleren van een procedure voor het valideren van habitatmodellen met veldgegevens.

De uitwerking van deze punten is onderwerp van de volgende hoofdstukken.

2 De flexibele voorkant van het habitatinstrument

Het huidige habitatinstrument bevat een uitgebreide set van habitatmodellen die in het kader van verschillende projecten zijn gemaakt. Het gaat hierbij om zogenaamde habitatgeschiktheidsmodellen als habitatbeschikbaarheidsmodellen. Bijlage A geeft een overzicht van bestaande habitatmodellen. Habitatbeschikbaarheidsmodellen zijn gebaseerd op de relatie tussen ecotopen en habitatbeschikbaarheid, vaak aangevuld met enkele andere factoren zoals diepte of stroomsnelheid. De habitatbeschikbaarheid geeft alleen aan of het habitat dat door een soort wordt vereist ook daadwerkelijk in een gebied aanwezig is. Het is dus een kwantitatieve maat en laat zich niet uit over de kwaliteit van het habitat. De habitatgeschiktheidsmodellen zijn gebaseerd op een uitgebreide set aan invoervariabelen, die per soort of soortgroep vaak verschilt. De toepassingsmogelijkheden van habitatmodellen zijn afhankelijk van de beschikbaarheid van monitoringgegevens en de specifieke vraagstelling. Soms is een globale verkenning naar het vóórkomen van leefgebieden voldoende (habitatbeschikbaarheid), soms is een gedetailleerde analyse naar habitatgeschiktheid nodig. Bij de toepassing van habitatmodellen is het vaak nodig om een interpretatie of vertaling te maken van de beschikbare monitoringgegevens naar definitie van invoervariabelen, aangezien deze vaak niet op elkaar afgestemd zijn.

Een belangrijk nadeel van het huidige habitatinstrument is dat het te rigide bevonden wordt, waarbij verschillende knelpunten aangewezen zijn Ducl et al. (2001):

- Er bestaat geen mogelijkheid tot vertaling/interpretatie van monitoringgegevens door middel van herclassificatie.
- Er is geen mogelijkheid voor het documenteren van modellen en voorbewerkingen op een gebruikersvriendelijke manier.
- De mee te nemen invoervariabelen kunnen niet op een gebruikersvriendelijke manier geselecteerd worden.
- Het aanpassen of definiëren van habitatfuncties is niet gebruikersvriendelijk.
- De relatie tussen habitatbeschikbaarheid en ecotopen is gebaseerd op een vaste set aan voorgedefinieerde ecotopen.

De bedoeling van een vernieuwd habitatinstrument is dat gebruiker zo flexibel mogelijk kan omgaan met invoergegevens. Het flexibel kunnen omgaan met invoergegevens betekent dat zowel een aantal informaticatechnische als inhoudelijke ontwikkelingen nodig zijn.

IT ontwikkelingen:

1. Mogelijkheid tot documentatie van habitatmodellen, invoer en aanpassingen hierop.
2. Mogelijkheid tot herclassificatie van monitoringgegevens tot geschikte invoer.
3. Flexibele selectie van mee te nemen invoervariabelen, zie voorbeeld in onderstaand schema.
4. Flexibele definitie en aanpassing van habitatfuncties, voor een optimale aansluiting op de invoergegevens (grenzen klassenverdeling flexibel of bijvoorbeeld optimumfunctie in plaats van klassenverdeling).

Inhoudelijke ontwikkelingen:

1. Afstemming van monitoringgegevens en habitatcriteria.
2. Uitwerken van de vraag hoe een gebruiker flexibel om kan gaan met de relatie tussen ecotopen en habitatbeschikbaarheid.

Soorten	Factoren							
	1	2	3	4	5	6	7	8
a								
b								
c								
d								
e								
f								
g								
h								
i								
j								
k								
l								
m								
n								
o								
p								
q								
r								
s								

Figuur 2.1 Schema voor een habitatevaluatie: de factoren 1 en 2 worden gebruikt voor een snelle algemene toets op habitatbeschikbaarheid; de factoren 1 t/m 8 worden gebruikt voor de bepaling van habitatkwaliteit voor de soorten a t/m f

De IT-ontwikkelingen worden uitgewerkt en geïmplementeerd in een parallel lopend samenwerkingsproject van RIZA, RIKZ en WL, waarin een nieuw habitatinstrument wordt gebouwd. De IT-ontwikkelingen zijn in meer detail uitgewerkt in de definitiestudie en het technisch ontwerp van dit instrument (Van der Lee et al., 2003; **Brinkman et al., 2002**). De benodigde inhoudelijke ontwikkelingen worden nader omschreven in dit hoofdstuk.

2.1 Afstemming van monitoringgegevens en habitatcriteria

De eerste aanbeveling is om de monitoring van gegevens beter aan te laten sluiten op de benodigde gegevens over invoervariabelen in habitatmodellen. Dit zal leiden tot een geringer verlies van informatie door transformatie van monitoringgegevens naar invoergegevens. Aan de andere kant is er ook winst te behalen door bij de definitie van invoervariabelen al rekening te houden met de beschikbaarheid van monitoringgegevens. Het is daarnaast nuttig om invoervariabelen op een meer éénduidige manier te definiëren, aangezien in veel modellen de definitie van invoervariabelen mede beïnvloed wordt door de maker van het model zelf. Een eerste poging hiertoe is gedaan in het kader van dit project. Voor een set van soorten, kenmerkend voor verschillende soortgroepen, is gezocht naar een serie algemene habitatcriteria die voor de meeste soorten relevant zijn.

Ook is getracht optimumfuncties op te stellen voor deze habitatcriteria met behulp van een beknopt literatuuronderzoek. De habitatcriteria en de optimumfuncties staan beschreven in bijlage B. Onderstaande tekst geeft een korte synopsis.

Synopsis

Aan de hand van 35 soorten, geselecteerd op basis van biotoopgebruik, is een overzicht gegeven van de eisen die deze soorten aan hun habitat stellen (beschikbaarheid). Hierbij is onderscheid gemaakt in aquatisch paaihabitat, aquatisch fourageerhabitat en terrestrisch habitat. De belangrijkste factoren die in de onderzochte literatuur genoemd worden zijn: waterdiepte, stroomsnelheid, substraattypen, landschapselementen en talud. Daarnaast zijn er nog een aantal invoervariabelen genoemd die meer betrekking hebben op de kwaliteit van het habitat (watertemperatuur, doorzicht en areaal).

Er blijkt een groot verschil te zijn in detailniveau waarop een invoervariabele is beschreven, waarbij terrestrische systemen veel grover en meer kwalitatief zijn beschreven, terwijl aquatische systemen veel gedetailleerder en meer kwantitatief beschreven zijn. Ook blijkt dat het vergelijken van benodigde gegevens tussen soorten niet goed gedaan kan worden, aangezien voor veel factoren niet is aangegeven hoe deze geregistreerd zijn (methodes, tijd van het jaar, gebrek aan definities).

2.2 De relatie tussen ecotopen en habitatbeschikbaarheid

In de huidige situatie is de relatie tussen ecotopen en habitatbeschikbaarheid beschreven voor een vaste set aan voorgedefinieerde ecotopen. In de praktijk gebruiken veel instituten een eigen classificatiesysteem. Dit maakt dat gebruik van habitatbeschikbaarheidsmodellen een vertaling van het ene ecotopenstelsel naar het andere nodig maakt. Een manier om dit probleem op te lossen is om in het model niet de relatie tussen een ecotoop en de habitatbeschikbaarheid te definiëren, maar alleen de relaties tussen individuele invoervariabelen en habitatbeschikbaarheid. Deze invoervariabelen zijn dan constant. Voor toepassing van deze modellen met invoer voor ecotoopkaarten van verschillende ecotopenstelsels moet dan de relatie gelegd worden tussen de invoervariabelen en de gebruikte ecotoopcriteria van de ecotopenstelsels. Dit kan wanneer het instrument de mogelijkheid biedt om op een gebruikersvriendelijke manier vertaaltabellen te maken die goed gedocumenteerd zijn en reproduceerbaar. Daarnaast is het uiteraard nodig dat de habitatcriteria en ecotoopcriteria zelf ook goed gedocumenteerd zijn.

De vertaling van ecotoopcriteria naar habitatcriteria zal niet altijd eenvoudig zijn om verschillende redenen. Ten eerste zijn ecotoopkaarten geaggregeerde kaarten, wat wil zeggen dat de ecotoopeenheden opgebouwd zijn uit informatie van verschillende criteria samen, die voor een soort allemaal belangrijke invoervariabelen kunnen zijn. Wanneer niet duidelijk is hoe de precieze combinatie is van de waarden waarmee deze factoren of ecotoopcriteria met elkaar voorkomen, levert dat problemen op voor het kwantificeren van de habitatbeschikbaarheid of –geschiktheid. Een tweede probleem bij de vertaling is dat de definitie van ecotoopcriteria vaak veel minder gedetailleerd is dan die van habitatfuncties. Ten derde zijn ecotoopcriteria in klassen ingedeeld terwijl habitatfuncties vaak uit optimumfuncties bestaan.

Hieronder zijn verschillende situaties geanalyseerd waarbij de gegevens van de ecotopenkaart niet goed aansluiten op de benodigde gegevens voor de habitatfuncties. Voor elke situatie is een richtlijn opgesteld voor het omgaan ermee. De richtlijnen zijn eenvoudig te definiëren in het habitatinstrument dat momenteel gebouwd wordt.

De uitgewerkte situaties zijn:

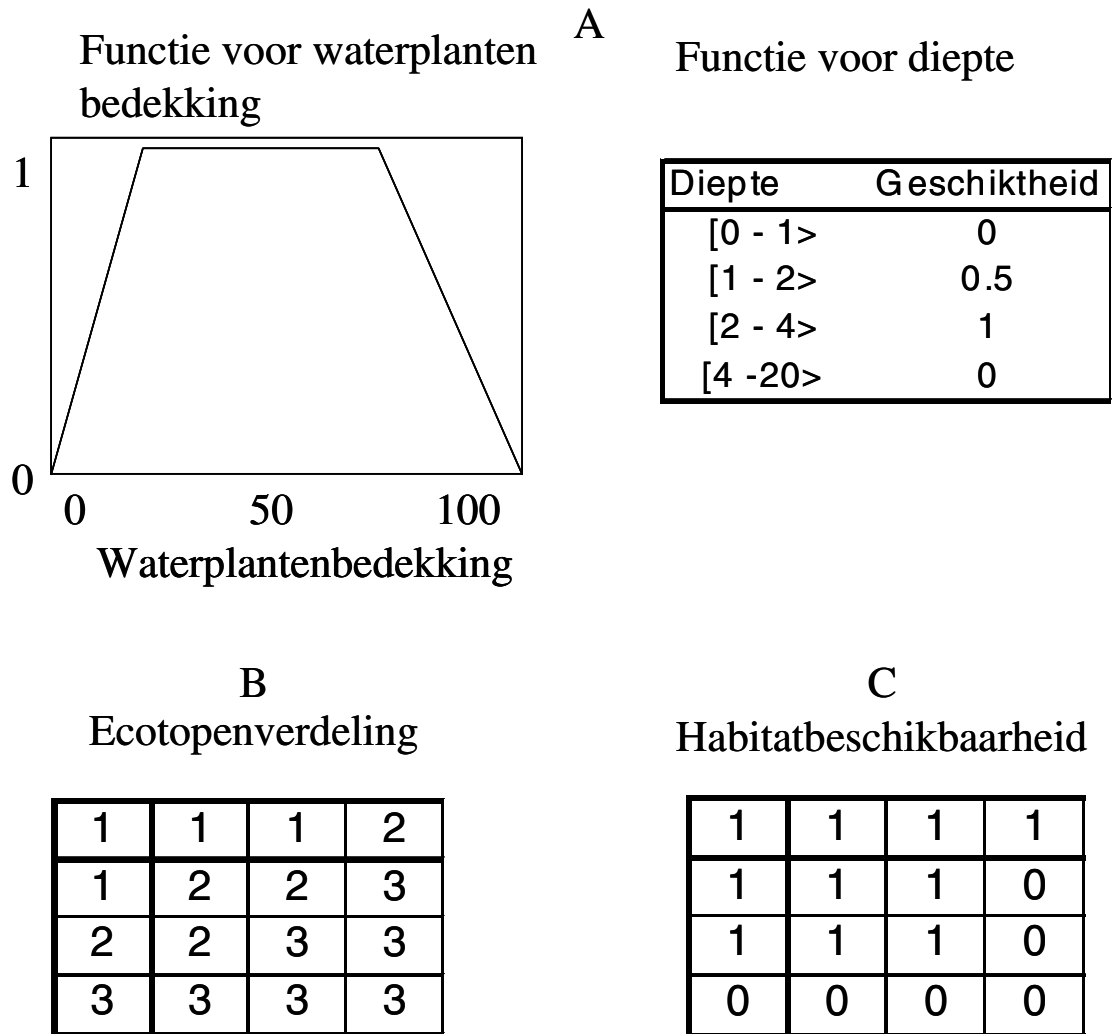
1. De ecotopenkaart is gebaseerd op criteria waarbij er meerdere beperkend kunnen zijn voor de habitat van een soort.
2. De ecotopenkaart is gebaseerd op criteria waarbij slechts één factor beperkend kan zijn voor het habitat.
3. De ecotopenkaart is gebaseerd op een indeling waarbij de klassen een veel grotere bandbreedte hebben dan de bandbreedte waarin het habitat kan voorkomen.

Ad 1. Bepalen van de habitatbeschikbaarheid met een geaggregeerde kaart in het geval dat een combinatie van meerdere factoren beperkend kan zijn.

Wanneer een combinatie van factoren beperkend is, is een exacte schatting van de habitatbeschikbaarheid of ook wel het areaal leefgebied onmogelijk. Bijvoorbeeld een suboptimale diepte, tesamen met een suboptimale bedekking van waterplanten kan het habitat uiteindelijk ongeschikt maken, terwijl de factoren op zich wel delen van het gebied als potentieel geschikt aanduiden. Er kan slechts worden aangegeven dat er potentieel leefgebied aanwezig is, zonder dat verder is aangegeven hoe groot dit areaal werkelijk is (kaartenheden krijgen waarde 1 als er potentieel leefgebied aanwezig is of waarde 0 als potentieel leefgebied niet aanwezig is). Figuur 2.2 geeft een uitgewerkt voorbeeld voor het bepalen van de habitatbeschikbaarheid met een ecotopenkaart bestaande uit de volgende ecotopen:

- Ecotoop 1: ondiep water (0-2 m) met of zonder waterplantenbedekking.
- Ecotoop 2: matig diep water (2-5 m) met of zonder waterplantenbedekking.
- Ecotoop 3: diep water (5-20 m) met of zonder waterplantenbedekking.

en met habitatbeschikbaarheidsfuncties van een soort X voor waterplantenbedekking en diepte. Volgens de functies zijn ecotoop 1 en 2 potentiële leefgebieden en krijgen de waarde 1. Of er daadwerkelijk habitat beschikbaar is, is onbekend en hangt af van de precieze combinatie van waterplantenbedekking en diepte, waar geen verdere informatie over is.



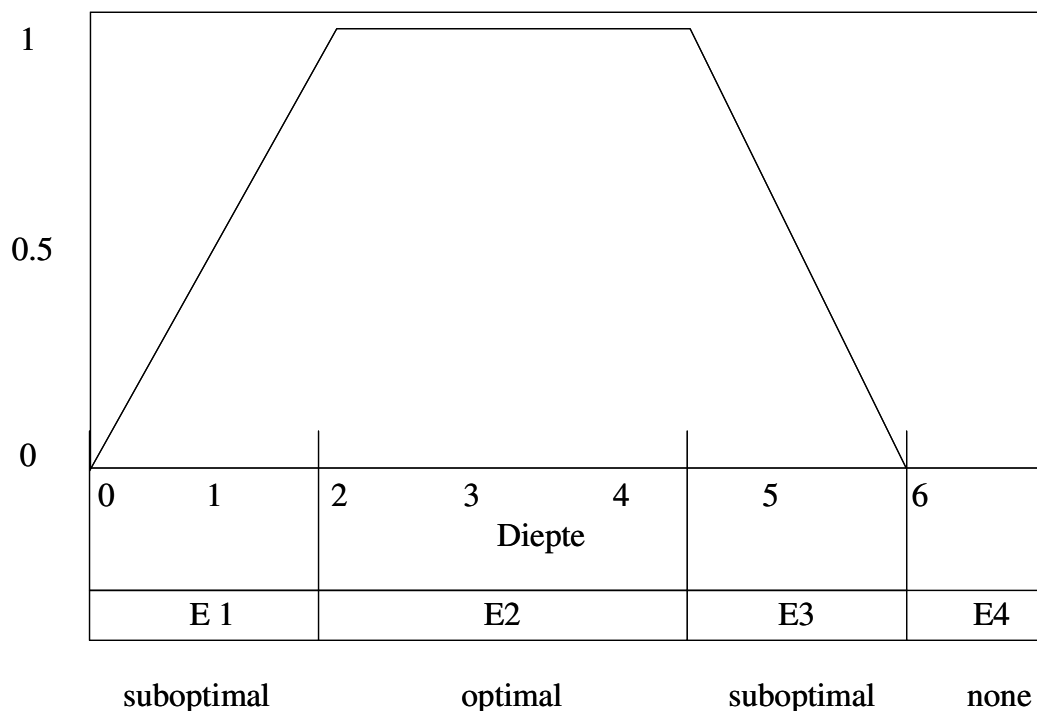
Figuur 2.2 Bepaling van de habitatbeschikbaarheid met een voorbeeld ecotopenkaart waarvan meerdere ecotoopcriteria beperkend kunnen zijn (waterplantenbedekking en diepte). A: soortmodel; B: invoerkaart; C: resultaatkaart.

Ad 2. Bepalen van de habitatbeschikbaarheid met single factor kaarten of ruimtelijk geaggregeerde kaarten waarbij slechts één van de factoren, die een kaarteenheden onderscheiden, beperkend is, waarbij:

- a) Klassengrenzen voor deze factor veel ruimer zijn dan de grenzen van een leefgebied. De klasseverdeling van ruimtelijk geaggregeerde kaarten kan zo ruim zijn dat binnen een kaarteenheden een deel van het gebied wel leefgebied vormt voor een soort en een ander deel niet. Bijvoorbeeld: Een kaarteenheden is gedefinieerd met een diepte van 0-1 m, terwijl de habitatbeschikbaarheid alleen optimaal is tussen 0.5-1 m. Er kan zo'n geval een schatting worden gemaakt van het percentage kaarteenheden dat geschikt zou kunnen zijn (bijv 50%), maar deze aanname is moeilijk hard te maken. Aan de ene kant zijn de kaartklassengrenzen zodanig gekozen dat gebieden met relatief homogene abiotische randvoorwaarden hierbinnen vallen.

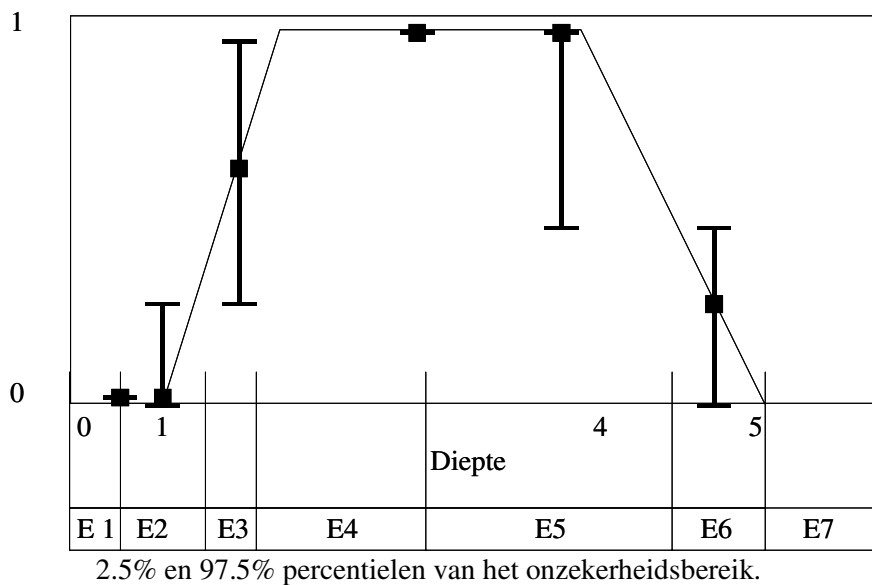
Maar aan de andere kant is de term ‘abiotisch homogeen’ beschreven op niveau van milieus en dat hoeft op de schaal van habitats niet hetzelfde te zijn. Een schatting van het areaal aan leefgebied met kaarten waarbij de klassengrenzen ruimer zijn dan die van het leefgebied is dus zeer onzeker. Een kwantitatieve schatting is dan ook ongewenst. De hele kaartenheid wordt daarom als potentieel leefgebied aangewezen, zonder dat verder aangegeven wordt wat het areaal aan leefgebied dan werkelijk is (kaartenheden krijgen waarde 0 of 1). In dit soort gevallen kan alleen een grove schifting van arealen worden gemaakt. Uiteraard moet ten alle tijden met zorg worden omgegaan met de keuze van invoermateriaal en het doel van de analyse.

- b) Klassengrenzen voor een factor kleiner zijn dan het bereik van de optimumfunctie (ordegrootte 2 keer). Wanneer binnen het bereik van een optimumfunctie slechts twee of drie klassen vallen, is het moeilijk om een kwantitatieve schatting te maken van de habitatkwaliteit. Afhankelijk van het bereik van de habitatgeschiktheidsfunctie binnen een klasse zou men soms wel het onderscheid tussen optimaal en suboptimaal habitat kunnen maken (suboptimaal: bereik >0-0.5; optimaal 0.5-1). Dit is toegelicht in Figuur 2.3.



Figuur 2.3 Opdeling in suboptimaal en optimaal als klassengrenzen niet overeen komen met de optimumfuncties. E1-E4: ecotoopnummers.

- c) Klassengrenzen voor deze factor veel kleiner zijn dan bereik optimumfunctie (ordegrootte 5 keer). In het geval dat in het bereik van een optimumfunctie een heel aantal kaartklassen vallen, kan wel een kwantitatieve schatting gedaan worden van de habitatgeschiktheid. Wel heeft deze schatting een bepaalde onzekerheid omdat op een gegeven locatie de precieze waarde van de invoervariabele binnen de klassegrenzen onbekend is. Met deze klassenverdeling kan men omgaan door te werken met een waarde precies tussen onder- en bovengrens van een klasse in, waarbij de onder- en bovengrens zelf meewegen in de onzekerheid (Figuur 2.4). Hierbij is het zaak om goede vertaaltabellen op te stellen. Bij afwezigheid van andere factoren van onzekerheid vormen de onder- en bovengrens van een klasse de



Figuur 2.4 Bepaling van de habitatbeschikbaarheid met een ecotopenkaart waarvan slechts 1 criterium beperkend is en wel zodanig dat de klassenverdeling van het ecotoopcriterium fijn is in vergelijking met de definitie van de habitatfunctie. Uit de projectie van de klassengrenzen van het ecotoopcriterium op de habitatfunctie wordt het gemiddelde, het minimum en de maximum habitatbeschikbaarheid afgeleid voor elke kaartklasse

Ad. 3 Bepalen van de habitatbeschikbaarheid met een geaggregeerde ecotopenkaart, waarbij voor één factor ook een gedetailleerde single factorkaart aanwezig is.

In dit geval kan voor de single factor kaart de optimumfunctie toegepast worden. Toepassen van deze functie resulteert in een score voor de habitatbeschikbaarheid tussen 0 en 1 in elke ruimtelijke eenheid. Voor de geaggregeerde kaart dient ook een score gedefinieerd te worden op basis van een vergelijking van de ecotoopcriteria met de habitatcriteria. De score is hier 0 of 1 wanneer meerdere factoren beperkend kunnen zijn. De waarde 1 geeft in deze kaart aan dat er potentieel leefgebied is, maar doet verder geen uitspraak over de precieze beschikbaarheid, aangezien de geaggregeerde informatie van de ecotopenkaart dit niet toelaat. De extra informatie van de single factor kaart nu kan gebruikt worden om de potentiële leefgebieden verder in te perken: dat wil zeggen, de gebieden waar volgens de ecotopenkaart potentieel leefgebied is (waarde 1), maar volgens de single factor kaart niet (waarde 0) krijgen een waarde 0 (Figuur 2.5).

De gebieden waar volgens de geaggregeerde kaart potentieel leefgebied is (waarde 1) en de score van de single factor kaart groter is dan 0, krijgen een waarde van 1.

1. beschikbaarheid volgens ecotopenkaart

1	1	1	1
1	1	1	0
1	1	1	0
0	0	0	0

2. beschikbaarheid volgens factor

0	0.2	0.5	1
0	0.7	1	0.7
0	0.3	1	0.7
0.5	0	1	1

3. beschikbaarheid volgens combinatie 1+ 2

0	1	1	1
0	1	1	0
0	1	1	0
0	0	0	0

Figuur 2.5 Bepaling van de habitatbeschikbaarheid met een ecotopenkaart (waarbij meerdere factoren beperkend kunnen zijn) en een single factor kaart

In de bijzondere situatie van de combinatie van:

1. De beschikbaarheid volgens een ecotopenkaart waar slechts 1 factor beperkend is en de beschikbaarheid met een voldoende nauwkeurigheid bepaald kan worden (situatie 2c).
2. De beschikbaarheid volgens een single factor kaart, waarin de resulterende beschikbaarheid gewoon bepaald kan worden als het minimum van beide kaarten (figuur 2.6).

1. beschikbaarheid volgens ecotopenkaart

1	1	1	1
1	1	1	0
1	1	1	0
0	0	0	0

2. beschikbaarheid volgens factor

0	0.2	0.5	1
0	0.7	1	0.7
0	0.3	1	0.7
0.5	0	1	1

3. beschikbaarheid volgens combinatie 1+ 2

0	1	1	1
0	1	1	0
0	1	1	0
0	0	0	0

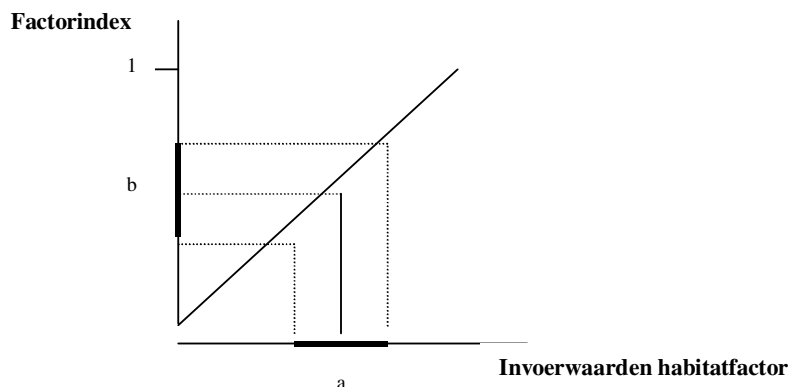
Figuur 2.6 Bepaling van de habitatbeschikbaarheid met een ecotopenkaart (waarbij slechts 1 factor beperkend is en de beschikbaarheid nauwkeurig kan worden bepaald) en een single factor kaart

3 Onzekerheidsanalyse van habitatmodellen

3.1 Procedure onzekerheidsanalyse

In het algemeen geldt dat een procedure voor onzekerheidsanalyses zeer ingewikkeld kan zijn maar ook heel eenvoudig. De mate van complexiteit die nodig is hangt af van het doel van de analyse. Eerst moet daarom bepaald worden welke informatie precies gevraagd wordt: op basis hiervan kan vervolgens de procedure vastgesteld worden. Op basis van een gebruikersbijeenkomst kwam naar voren dat met de procedure voor onzekerheidsanalyse de onzekerheid in de habitatgeschiktheid te berekenen moet zijn - in een bepaald gebied, en van een bepaalde soort -als gevolg van onzekerheid in:

- een bepaalde invoervariabele (zie Figuur 3.1);
- alle of een aantal invoervariabelen;
- een bepaalde factorindex (zie Figuur 3.1);
- alle of een aantal factorindices;
- een invoervariabele en een indexwaarde van één relatie samen;
- alle of een aantal invoervariabelen en bijbehorende indexwaarden samen.



Figuur 3.1 Habitatrelatie met onzekerheidsranges voor de modelinvoer (a) en de modelparameters uitgedrukt in de indexwaarde (b). Figuur is overgenomen uit Aarts et al., 2001

Met het model moet tevens een gevoeligheidsanalyse gedaan kunnen worden, waarbij zowel de gevoeligheid voor een afwijking in de invoergegevens als in de habitatindices te bepalen is. Met het model moeten voor elke factor kunnen worden bepaald waar in een gebied deze beperkend is voor een bepaalde soort, en in welke mate.

Op dit moment is (nog) niet voorzien in het achterhalen van de modelonzekerheid als gevolg van onzekerheid in de modelstructuur. Een voorbeeld van modelstructuur is de bepaling van de habitatkwaliteit door het minimum te berekenen van verschillende habitatindices. In werkelijkheid zal de ecologie zich niet netjes volgens deze functie gedragen.

De minimumfunctie zal dus altijd een benadering zijn en bevat dus altijd onzekerheid of voldoet soms zelfs helemaal niet.

Welke onzekerheid is meegenomen?

Voor de toepasbaarheid van het model wordt per habitatindex en per invoergegeven slechts een schatting van de onzekerheid meegenomen. Deze schatting is een gelump getal van verschillende onzekerheden. Bij invoergegevens gaat het om onzekerheid als gevolg van:

- variabiliteit in tijd;
- variabiliteit in ruimte;
- meetfouten.

Voor een aantal belangrijke factoren zal in de case studie uit hoofdstuk 4 een “best motivated guess” worden aangegeven.

Analyse methode

Er is gekozen voor de Monte Carlo simulatie omdat deze het gebruik van niet continue functies toelaat en gemakkelijk te implementeren is.

Gewenste eigenschappen van een onzekerheidsverdeling:

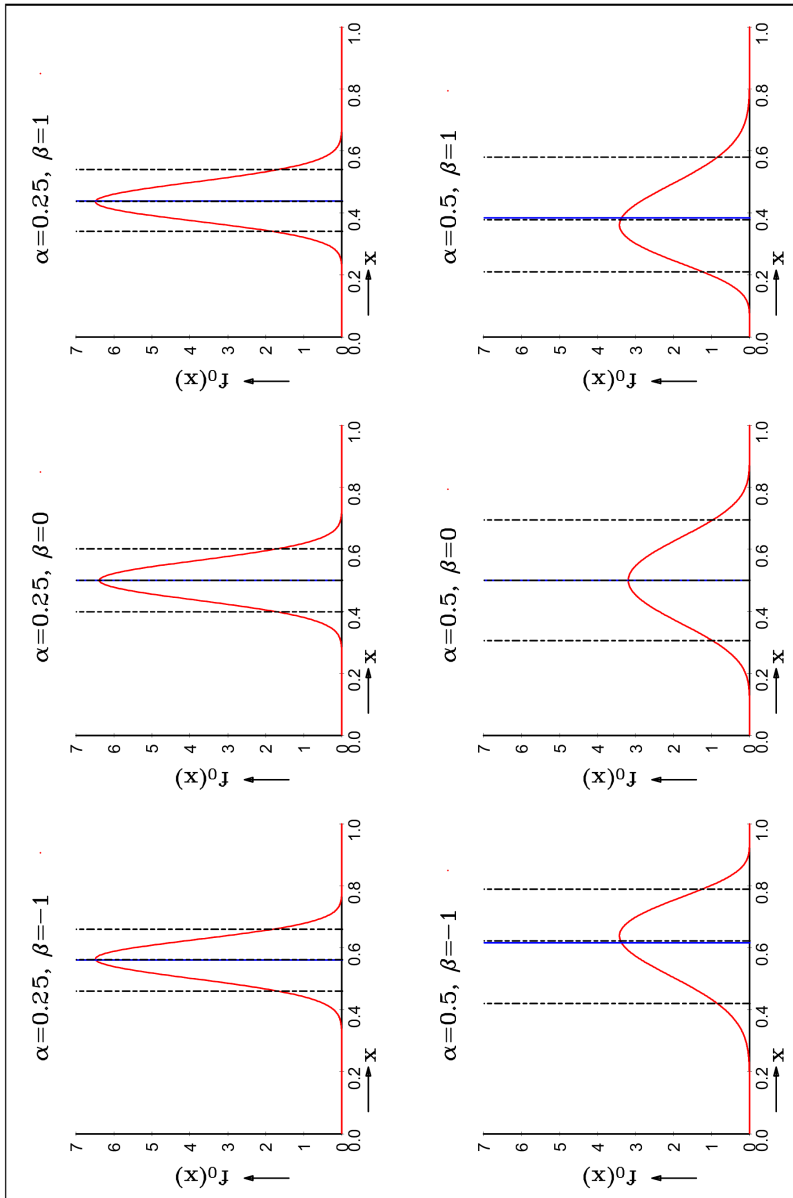
Wanneer zowel onzekerheid in de modelparameters (indexwaarden habitatfuncties) als in de invoergegevens meegenomen moet worden, is het wenselijk dat de onzekerheid voor beide op dezelfde manier gekwantificeerd is. Voor de factorindex geldt dat deze per definitie tussen 0 en 1 ligt, en dus binnen een begrens domein gedefinieerd is. Daarnaast moet de onzekerheidsverdeling zo flexibel zijn dat deze een a priori gewenste gemiddelde en variantie heeft. Dit vereist dat de functie zogenaamde vormparameters moet bevatten. Tenslotte moet de verdeling zowel symmetrische vormen kunnen hebben als asymmetrische vormen. Asymmetrische vormen zijn vooral gewenst bij factorindexwaarden nabij 0 of 1 omdat deze waarden een harde onder- of bovengrens vormen, waardoor de verdeling tussen het gemiddelde en deze onder- of bovengrens heel steil kan zijn. De andere flank van de verdeling daarentegen kan best veel breder zijn.

Technisch gesproken zijn de gewenste eigenschappen van de onzekerheidsverdeling:

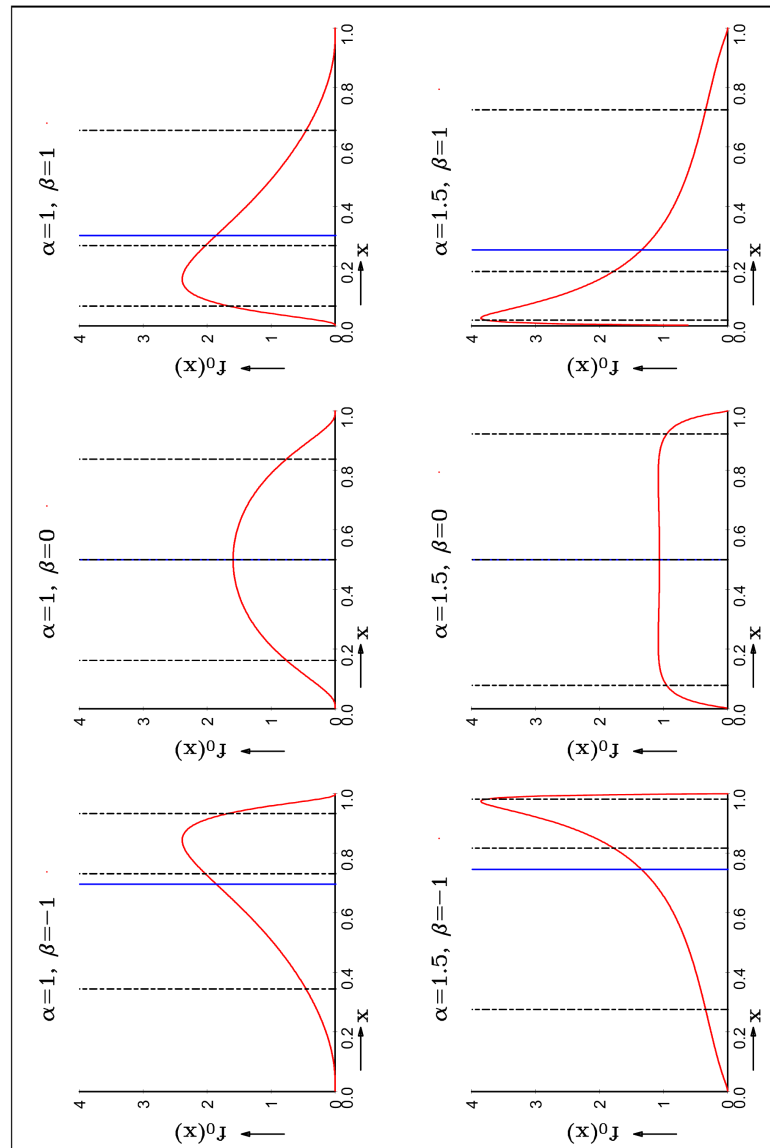
1. De verdeling is op een begrens domein $[a,b]$ gedefinieerd, waarbij we zelf nog keuzes kunnen maken t.a.v. de ondergrens a als de bovengrens b . De keuzes van die onder- en bovengrens kunnen afgeleid worden van de onzekerheidsranges die in de studie van 1999 geïnventariseerd zijn (‘integratie’ van schattingen van diverse experts?). Voor de invoergegevens kan een analyse van datasets resulteren in schattingen van a en b per invoerfactor.
2. De verdelingen bevatten naast de bereikparameters a en b eveneens twee vormparameters α , β die gebruikt kunnen worden om er voor te zorgen dat de verdeling een a priori gewenste gemiddelde en variantie (of gelijkwaardig: een spreiding) heeft.

3. Deze klasse van verdelingsfuncties omvat symmetrische vormen (als de vormparameters α , β gelijk zijn: $\alpha = \beta$) maar ook scheve verdelingen. Voor die scheve verdelingen kunnen zowel vormen met '*skewed to the left*' (de linkerflank is langer en minder steil dan de rechterflank, en het gemiddelde ligt dan rechts van het midden van het interval $[a, b]$) optreden (namelijk als $\alpha > \beta$), als vormen met '*skewed to the right*' (de rechterflank is langer en minder steil dan de linkerflank, en het gemiddelde ligt links van het midden) optreden (namelijk als $\beta > \alpha$). In de definitie van de onzekerheid in een habitatindex als functie van een invoervariabele zal een '*skewed to the left*' vorm het meest geschikt zijn voor een habitatindex die dicht bij 1 ligt, terwijl een '*skewed to the right*' vorm beter geschikt lijkt voor een habitatindex die dicht bij 0 ligt.
4. De uniforme verdeling is als een speciaal geval bevat in de klasse, namelijk wanneer $\alpha = \beta = 1$.

Na enig puzzelen is een verdeling gevonden die aan bovenstaande eisen voldoet. Voorbeelden van verdelingen staan in onderstaande figuur. Bijlage C geeft een uitgebreide beschrijving van de gebruikte onzekerheidsverdeling.



Figuur 3.2a Voorbeelden van symmetrische onzekerheidsverdelingen bij verschillende waarden van de vormparameters α en β



Figuur 3.2b Voorbeelden van asymmetrische onzekerheidsverdelingen bij verschillende waarden van de vormparameters α en β

Werking van het testmodel

Het model is een raster GIS model, waarbij berekeningen worden uitgevoerd voor elke onderscheiden ruimtelijke eenheid, in dit geval voor elke rastercel. Bij een onzekerheidsanalyse wordt dit model met een Monte Carlo simulatie een groot aantal keer gedraaid.

Invoer

Voor het uitvoeren van een onzekerheidsanalyse met onzekerheden in habitatfuncties en invoervariabelen zijn de volgende gegevens nodig:

habitatfuncties:

- tabellen met gemiddelde indexwaardes van de breekpunten van habitatfuncties;
- tabellen met uitersten en standaarddeviatie van de indexwaarden van deze breekpunten.

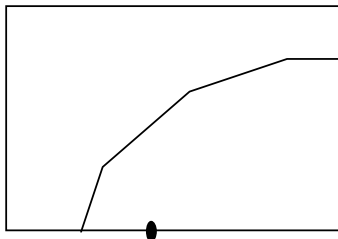
invoervariabelen:

- rasterkaarten met waarden voor invoervariabele (bijvoorbeeld dieptekaart, of kaart met fosfaatconcentraties);
- tabellen met uiterste waarden en standaarddeviatie van de invoervariabelen.

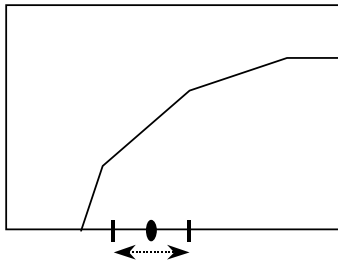
Berekeningen

Met het testmodel worden verschillende rekenstappen doorlopen:

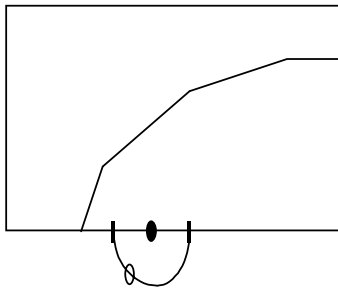
Stap 1: Berekening van de habitatindices per habitatfunctie. Dit gaat volgens onderstaand schema:



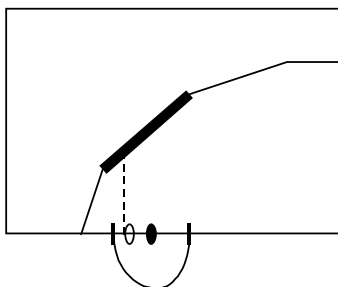
Invoer: kaart met waarden habitatfactor



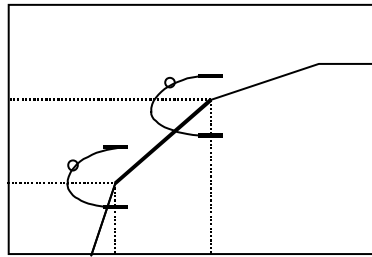
Lees in: onzekerheidsrange rond waarden



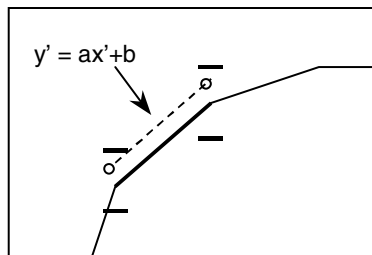
Geef verdeling van onzekerheid
Trek random getal uit deze verdeling



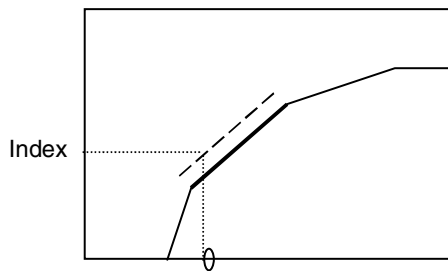
Bepaal onder welk deel van de functie de
getrokken waarde voor de habitatfactor valt



Lees in: onzekerheidsrange van breekpunten
Geef verdeling van onzekerheid voor breekpunten
Trek random getal uit deze verdeling



Bepaal a en b van functie $y' = ax' + b$



Bepaal index voor getrokken
waarde x met functie $y' = ax' + b$

Figuur 3.3 Presentatie van de verschillende stappen in de onzekerheidsanalyse

Stap 2. De berekening van de habitatkwaliteit

De habitatkwaliteit wordt berekend door een functie toe te passen op de habitatindices. In ons testmodel wordt het minimum berekend van de habitatindices, uitgaande van het principe dat de factor met de laagste habitatindex bepalend is voor de habitatgeschiktheid. Sommige habitatmodellen maken geen gebruik van een minimumfunctie. Hier moet nog naar gekeken worden.

Stap 3. Berekening van de beperkende factoren

Voor elke invoervariabele geldt dat deze beperkend is als de habitatindex gelijk is aan de habitatkwaliteit. In de Monte Carlo simulatie wordt berekend hoe vaak een bepaalde factor beperkend is (bijvoorbeeld 100 keer van het totale aantal runs van de simulatie). Hieruit volgt de kans dat een bepaalde factor beperkend is. Voor het hele gebied als totaal kan in een tabel ook berekend worden wat per factor het percentage is van het totale oppervlak waar de factor beperkend is ($\text{kans} \cdot 100 \cdot \text{areaal} / \text{totaal areaal}$).

Uitvoer van het prototype model

Op de resultaten van de runs kunnen statistische analyses worden toegepast, zoals bepaling van het gemiddelde, de variantie, het minimum en maximum, bepaalde fractielen etc. De resultaten kunnen zowel in kaartvorm als in tabelvorm gepresenteerd worden.

Er is in het prototype model voor gekozen om de onzekerheidsverdeling van de berekende habitatkwaliteit te beschrijven met hetzelfde type onzekerheidsverdeling als die van de invoergegevens en modelparameters. Dit is een verdeling begrensd binnen domein $[a, b]$ en met een bepaald gemiddelde en een variantie. De waarden van a , b , gemiddelde en variantie worden berekend met de resultaten van de Monte Carlo simulatie.

3.2 Toepassing op schedefonteinkruid

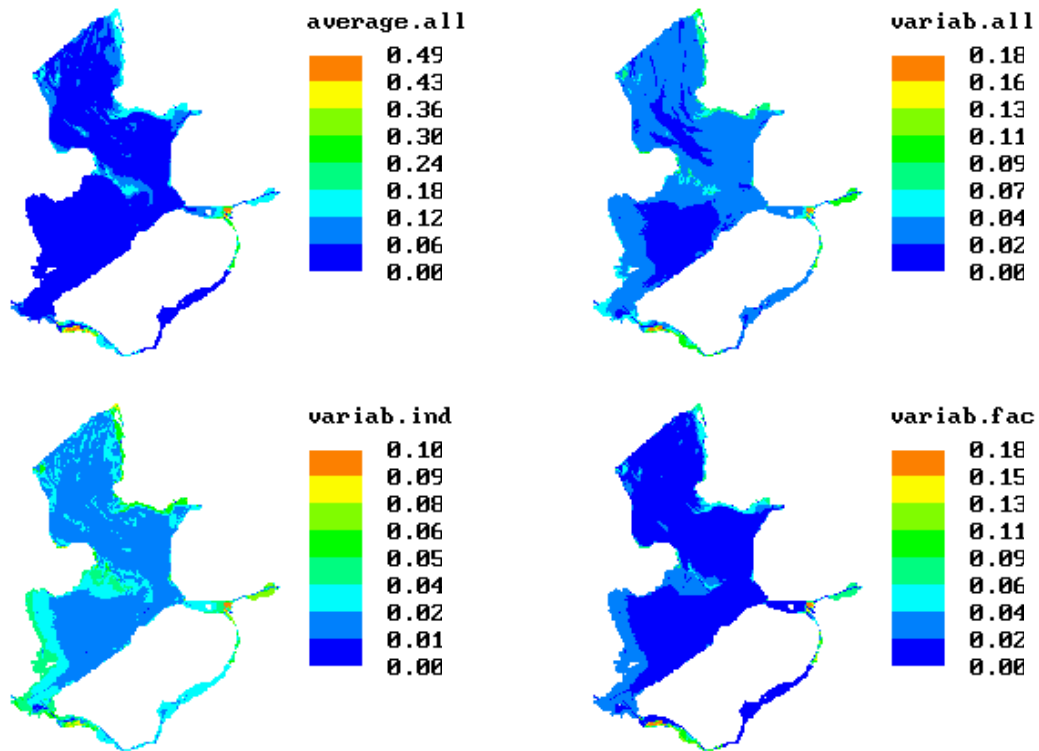
Het habitatgeschiktheidsmodel voor schedefonteinkruid is beschreven in Laane (1996). De belangrijkste beperkende factoren voor de habitatgeschiktheid zijn eerder onderzocht in Van der Lee et al. (2000). De onzekerheidsanalyse gaf als meest gevoelige en onzekere factoren de orthofosfaatconcentratie, de diepte en de strijklengte aan. Van de meegenomen factoren waren het doorzicht, bodemtype en de slibrijkdom de minst gevoelige en onzekere factoren. In de huidige studie is de onzekerheidsanalyse uitgebreid door naast de onzekerheid in de afzonderlijke habitatfuncties ook de onzekerheid in invoergegevens mee te nemen. Het is vooralsnog alleen mogelijk om de onzekerheid in continue factoren (invoervariabelen) mee te nemen, zoals bijvoorbeeld diepte. Voor het meenemen van onzekerheid in nominale of ordinale factoren, zoals bodemtype en slibgehalte, is een extra uitbreiding van de methode nodig¹. Voor de analyse in dit onderzoek zijn daarom alleen de continue factoren orthofosfaatconcentratie, diepte, strijklengte en doorzicht meegenomen.

Voor het uitvoeren van een onzekerheidsanalyse naar onzekerheden in habitatfuncties en invoergegevens zijn de volgende gegevens nodig:

- gemiddelde indexwaarden van de breekpunten van habitatfuncties;
- uitersten (bandbreedte) en standaarddeviatie van de indexwaarden van deze breekpunten;
- gemiddelde waarden van de invoergegevens diepte, orthofosfaatconcentratie, strijklengte en doorzicht;
- uiterste waarden (bandbreedte) en standaarddeviatie van deze invoergegevens.

Gemiddelde indexwaarden, uiterste waarden en standaarddeviaties van de habitatfuncties zijn bepaald in Van der Lee et al. (2000) middels een expertmeningen-enquête. Gemiddelden en uiterste waarden (minimum en maximum) van de functies staan gegeven in Tabel 3.1. Merk in deze tabel de grote bandbreedtes op die de experts aangeven voor de functies strijklengte en diepte. Uit de expertmeningen-enquête bleek dat de standaarddeviatie voor de indexwaarde van een breekpunt samenhangt met de indexwaarde van de functie zelf en in veel mindere mate met het type invoervariabele of de waarde van de invoervariabele (invoergegevens).

¹ Op dit moment ligt er een ontwerp voor deze uitbreiding dat eenvoudig geïmplementeerd zou kunnen worden. Dit is nog niet uitgevoerd, mede omdat de onzekerheden in de bodemkaart en slibgehaltekaart nog niet bekend zijn.



Figuur 3.1 Kaarten met de gemiddelde invoervariabelen die gebruikt zijn als invoergegevens in de onzekerheidsanalyses: diepte, orthofosfaatconcentratie, strijk lengte en doorzicht. NB negatieve waarden voor de diepte geven de drooggelegen locaties aan (hoogte boven gemiddeld waterpeil)

De standaarddeviatie bleek het hoogst te zijn bij indexwaarden tussen 0.4 en 0.6 en het laagst bij heel lage of heel hoge indexwaarden. De standaarddeviaties die gebruikt zijn, zijn gemiddelde standaarddeviaties per range in index (zie Tabel 3.2).

De bij de analyse gebruikte kaarten met gemiddelde waarden voor de invoervariabelen zijn dezelfde als gebruikt in Van der Lee et al., 2000 en zijn gebaseerd op gegevens van RIZA (Figuur 3.1). De strijk lengtekaart is dezelfde als gebruikt wordt in het ecotopenmodel ECOMIJ. De dieptekaart is berekend met de hoogtekaart en het gemiddelde waterpeil. De doorzicht- en orthofosfaatkaart zijn gebaseerd op maandelijkse puntmetingen per watersysteem. De waarden van de puntmetingen in een watersysteem zijn toegekend aan het gehele betreffende watersysteem. De standaarddeviaties van de factoren staan in Tabel 3.3. Bij de diepte is de standaarddeviatie gebaseerd op de meetfout van hoogtemetingen (bron: Meetkundige Dienst). Bij doorzicht en orthofosfaat zijn de standaarddeviaties bepaald uit de spreiding van seizoenwaardes voor doorzicht en orthofosfaat over de jaren heen. De onzekerheid in strijk lengte is niet goed bekend. Deze studie heeft een deskundigenschatting opgeleverd die vrij hoog is. De strijk lengte is namelijk gebaseerd op de meest voorkomende windrichtingen en de gemiddelde windrichting in het groeiseizoen varieert behoorlijk van jaar tot jaar afhankelijk van heersende weerpatronen. Van de vier invoervariabelen is de standaarddeviatie van de diepte duidelijk het laagst.

Dit komt overeen met de verwachting aangezien de diepte een min of meer permanente factor is terwijl bijvoorbeeld de orthofosfaatconcentratie en het doorzicht samenhangen met de groei van algen, die zowel in ruimte als tijd zeer sterk varieert.

De uitersten in de verwachte bandbreedte van invoergegevens voor een invoervariabele zijn standaard ingesteld op -50 en $+50$ % van de gemiddelde waarden. De keuze van deze bandbreedte is tot op zekere hoogte arbitrair. Uitgangspunt is dat de uitersten ook werkelijk net buiten de verwachting voor meetwaarden liggen. Mogelijk is de bandbreedte in sommige gevallen te klein gekozen, bijvoorbeeld bij lage waarden van doorzicht en orthofosfaat waar de bandbreedte waarschijnlijk hoger ligt dan $+50$ of -50 %. Dit zou pleiten voor een flexibele instelling van de bandbreedte afhankelijk van de invoervariabele en van de waarden van invoergegevens voor een invoervariabele. De instelling van de bandbreedte voor onzekerheid in de onzekerheidsanalyse (Monte Carlo simulatie) kan in vervolganalyses eenvoudig aangepast worden.

3.3 Resultaten en discussie

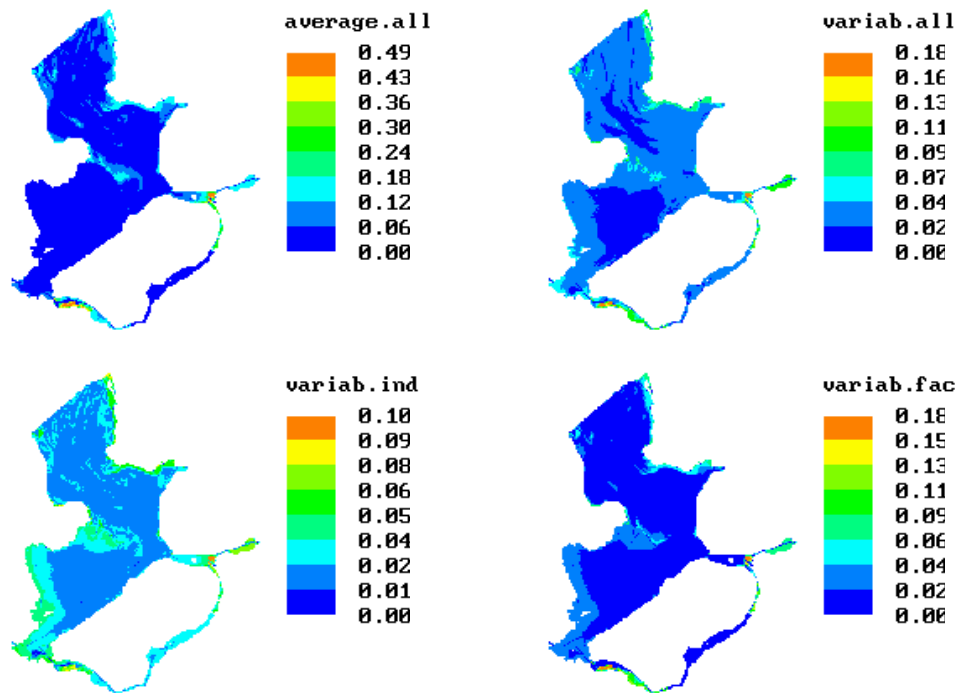
De vraagstelling bij de opzet van de onzekerheidsanalyse middels een Monte Carlo simulatie betrof het aandeel van respectievelijk de onzekerheid in habitatfuncties en in invoergegevens in de totale onzekerheid en hoe deze elkaar beïnvloeden. Figuur 3.2 laat de resultaten zien van de onzekerheidsanalyse met meenemen van zowel onzekerheid in habitatfuncties als in factorwaarden. De gemiddelde habitatgeschiktheid (average.all) is alleen hoger dan ca. 0.1 in de oeverzones van het IJsselmeergebied, en alleen hoger dan ca. 0.4 in enkele randmeren. De waarden voor de factoren strijklengte, doorzicht en orthoP samen zorgen dat de habitatgeschiktheid nergens optimaal is, ook niet op qua diepte geschikte standplaatsen.

In Figuur 3.2 is de standaarddeviatie van de berekende habitatgeschiktheid gegeven voor drie situaties:

- bij meenemen van onzekerheid in habitatfuncties en factorwaarden (variab.all);
- van alleen onzekerheid in habitatfuncties (variab.ind); en
- van alleen onzekerheid in factorwaarden (variab.fac).

De standaarddeviaties blijken samen te hangen met de berekende habitatgeschiktheid, zijn hoog in de oeverzones van het gebied en bereiken de hoogste waarden daar waar de habitatgeschiktheid ook het hoogst is (in de randmeren).

Een vergelijking van de kaarten met standaarddeviaties in de drie onzekerheidsscenario's levert een aantal relevante waarnemingen op voor een antwoord op de vraag naar het aandeel van respectievelijk de onzekerheid in functies en in factorwaarden in de totale onzekerheid (Figuur 3.2). Zo blijkt dat de hoogste standaarddeviatie die voorkomt bij meenemen van beide onzekerheden 0.18 is. Deze waarde is iets hoger dan de standaarddeviatie bij meenemen van alleen onzekerheid in de factorwaarde. Op locaties met deze hoogste standaarddeviatie is het aandeel van onzekerheid in factorwaarde dus dominant. Het algemene beeld voor de kaart als geheel is dat de totale onzekerheid aanzienlijk kleiner is dan de som van de afzonderlijke standaarddeviaties voor onzekerheid in factorwaarde en habitatfunctie.

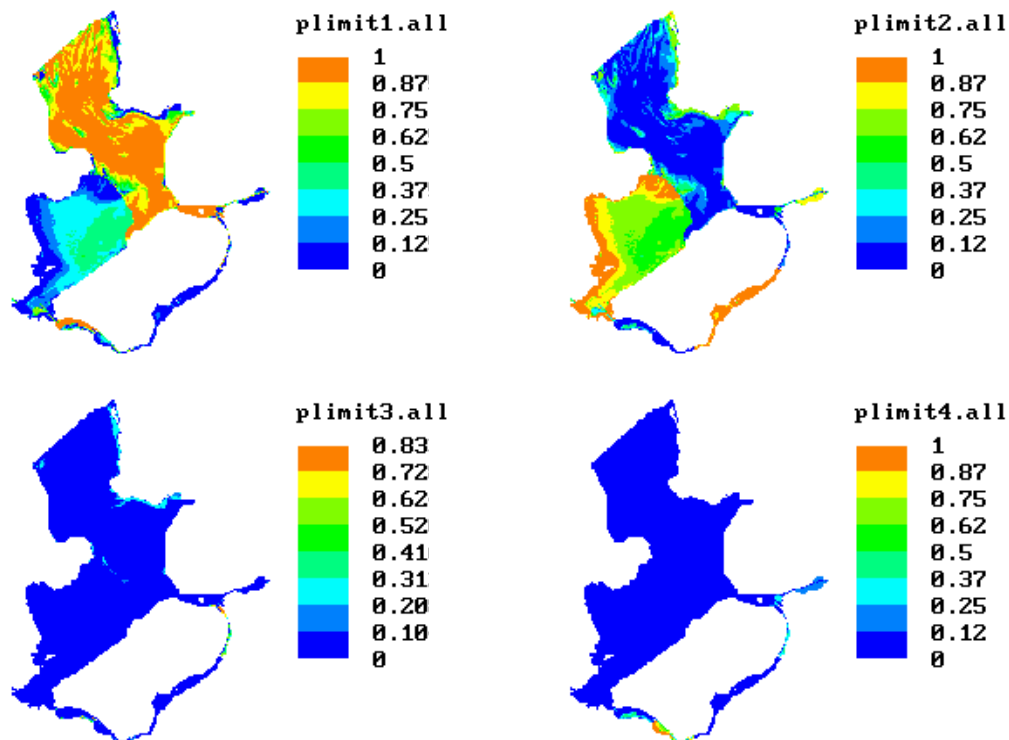


Figuur 3.2 De gemiddelde habitatgeschiktheid van de Monte Carlo simulatie met meenemen van onzekerheden in invoerkaarten en habitatfuncties (average.all), standaarddeviatie van de Monte Carlo simulatie met meenemen van onzekerheden in invoer en habitatfuncties (variab.all), standaarddeviatie met meenemen van alleen onzekerheid in habitatfuncties (variab.ind) en standaarddeviatie met meenemen van alleen onzekerheid in invoergegevens (variab.fac)

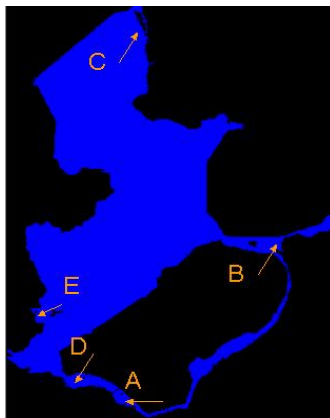
De onzekerheid in respectievelijk de habitatfunctie en de invoervariabele versterken elkaar dus niet, zoals theoretisch denkbaar is. Dit komt doordat de bandbreedtes van beide onzekerheden elkaar sterk overlappen.

De berekende ruimtelijke verdeling van habitatgeschiktheid en de standaarddeviatie is het gevolg van de mate van beperking van de geschiktheid door de waarden van de meegenomen invoervariabelen diepte, orthofosfaat, strijklengte en doorzicht in het gebied. Figuur 3.3 geeft voor elke invoervariabele in een kaart weer wat de kans is dat deze factor beperkend is op een bepaalde plek. De diepte blijkt beperkend in de diepe delen van IJsselmeer en Markermeer. Orthofosfaat kan zowel beperkend zijn in het Markermeer, de oeverzones van het IJsselmeer en in sommige randmeren. De strijklengte is beperkend in de oostelijke oeverzones van het IJsselmeer en sommige randmeren en het doorzicht is alleen lokaal beperkend in sommige randmeren.

De sterke ruimtelijke verdeling van habitatgeschiktheid als gevolg van beperking door verschillende invoervariabelen roept de vraag op wat de bijdrage is van een individuele invoervariabele aan de berekende totale onzekerheid en hoe de onzekerheden elkaar beïnvloeden. Om deze vraag te beantwoorden is een detailanalyse uitgevoerd voor 5 locaties in het IJsselmeergebied waarvoor de habitatgeschiktheid, de standaarddeviatie en de beperkende factoren verschillend zijn.



Figuur 3.3 De kans dat één van de meegenomen invoervariabelen beperkend is, volgend uit de Monte Carlo simulatie, waarbij onzekerheden in zowel invoer als habitatfuncties zijn meegenomen ($0 < \text{kans} < 1$). Plimit1.all: de kans dat diepte beperkend is; plimit2.all: de kans dat orthofosfaat beperkend is; plimit3.all: de kans dat strijklengte beperkend is en plimit4.all: de kans dat doorzicht beperkend is



Figuur 3.4 Puntlocaties (A t/m E) in het IJsselmeergebied waar de resultaten van de onzekerheidsanalyses in detail zijn uitgewerkt

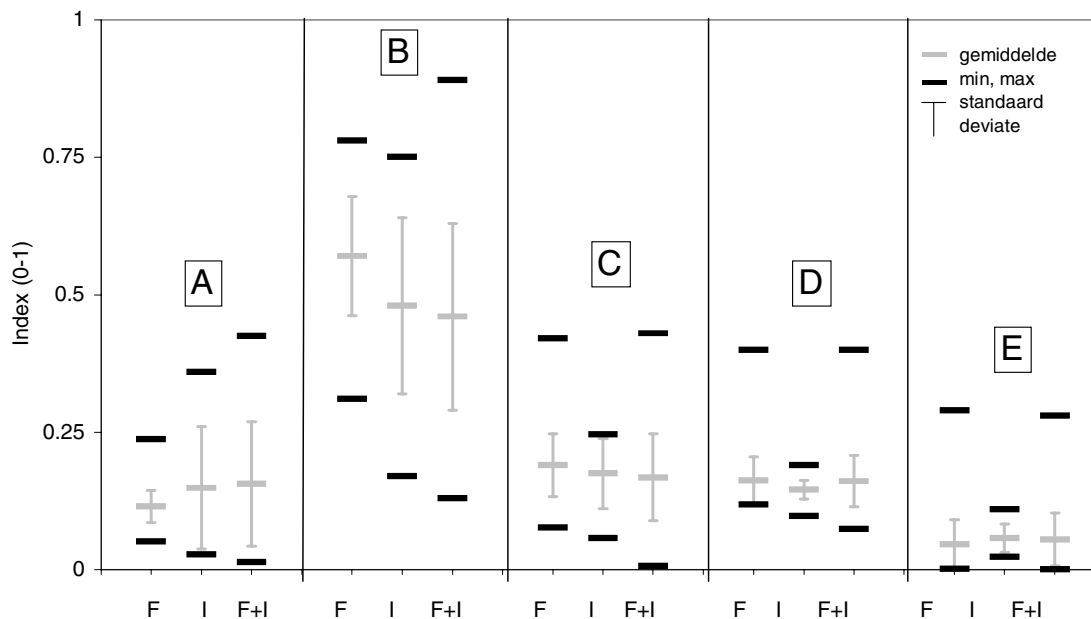
De locaties zijn met de letters A t/m E aangegeven in Figuur 3.5. In Tabel 3.4 is voor elke puntlocatie aangegeven wat de kans is dat één van de invoervariabelen beperkend is voor elk van de drie onzekerheidsscenario's:

- index: alleen onzekerheid in habitatfuncties;
- fac: allen onzekerheid in factorwaarden; en
- index + fac: beide onzekerheden meegenomen.

In punt A is doorzicht volledig beperkend, in punt B orthoP en doorzicht beide, in C orthoP, doorzicht en strijklengte, in D alleen de diepte en in punt E alleen orthoP.

Uit Tabel 3.4 blijkt dat wanneer zowel onzekerheid in functies als in factorwaarde meegenomen wordt, de kans groter wordt dat meerdere factoren beperkend zijn. Dit is het gevolg van de toenemende bandbreedtes in de resulterende index van individuele habitatfuncties.

Figuur 3.5 laat de gemiddelde, de standaarddeviatie en de bandbreedte zien van de habitatgeschiktheid in punt A t/m E. Opvallend is dat terwijl de gemiddelde habitatgeschiktheid voor punt A, C en D ongeveer gelijk is, de standaarddeviaties en bandbreedtes sterk uiteenlopen. De standaarddeviatie en bandbreedte zijn relatief hoog in punt A en B, gemiddeld in C, en relatief laag in punt D en E. In punt A en B is het voornamelijk de onzekerheid in de factorwaarde die de totale onzekerheid bepaalt (F), en in punt D en E voornamelijk de onzekerheid in habitatfunctie (I). Deze standaarddeviaties en bandbreedtes nu zijn het uiteindelijke resultaat van de interactie van onzekerheden in de vier invoervariabelen. In Figuur 3.6 zijn de onzekerheden per individuele invoervariabele uitgewerkt.



Figuur 3.5 Gemiddelde, standaarddeviatie, minimum en maximum van de habitatgeschiktheid op de 5 puntlocaties A t/m E (zie Figuur 3.4) voor 3 scenario's: I: alleen onzekerheid in habitatfunctie is meegenomen; F: alleen onzekerheid in invoervariabele is meegenomen en 3: onzekerheid in zowel habitatfunctie als invoerfactor is meegenomen

Van elk punt geeft de bandbreedte en standaarddeviatie in verticale richting de resulterende onzekerheid in de index aan als beide onzekerheden meegenomen zijn. De bandbreedte en standaarddeviatie in horizontale richting geven de onzekerheid in de invoergegevens aan.

Punt A:

In punt A is het doorzicht de enige beperkende factor (Tabel 3.4). Figuur 3.6d geeft de onzekerheid aan voor punt A voor de indexwaarde van de factor doorzicht. Het blijkt dat de standaarddeviatie en bandbreedte groot zijn, wat dus de hoge standaarddeviatie in de uiteindelijke habitatgeschiktheid verklaart. De onzekerheid in de functie doorzicht is groot omdat punt A op een steil gedeelte van de functie ligt. Een kleine bandbreedte op de x-as (doorzicht) zorgt dan voor veel grotere bandbreedte in verticale richting. Merk op dat de onzekerheid in de habitatgeschiktheid door onzekerheid in de habitatfunctie zelf relatief gering is (Figuur 3.5). Dit komt doordat de waarde voor doorzicht in punt A op een gedeelte van de curve ligt waar doorzicht sterk beperkend werkt, en waar de experts ook vrij zeker van zijn (zie Tabel 3.2, Figuur 3.6d).

Punt B:

In punt B zijn de factoren doorzicht en orthofosfaat beperkend. Voor beide habitatfuncties ligt punt B op een steil gedeelte van de curve waardoor onzekerheid in factorwaarde tot een relatief grote onzekerheid in indexwaarde leidt. Daarnaast heeft punt B in beide functies een indexwaarde van ca. 0.5. Bij deze waardes geven experts een hoge onzekerheid in index aan (Tabel 3.2).

Punt C:

In punt C zijn de factoren orthofosfaat, strijklengte en doorzicht beperkend. De standaarddeviatie in de index van orthofosfaat is vrij groot, maar kleiner dan in punt B. Dit komt doordat de standaarddeviatie in de waarde van orthofosfaat kleiner is dan bij punt B. De standaarddeviatie in de index van strijklengte is relatief klein. Dit komt doordat punt C op een vlak gedeelte van de habitatfunctie ligt, waardoor onzekerheid in de waarde van strijklengte nauwelijks effect heeft op de onzekerheid in indexwaarde. De onzekerheid in doorzicht is wel weer groot vanwege de hoge index en de positie op het steile gedeelte van de curve. De factoren doorzicht en orthofosfaat zullen gezamenlijk dus de grootste bijdrage leveren aan de uiteindelijke onzekerheid in habitatgeschiktheid (Figuur 3.5).

Punt D:

In punt D is de diepte hoofdzakelijk beperkend. Punt D ligt op een relatief vlak gedeelte van de habitatfunctie waardoor de – toch al geringe – standaarddeviatie in de waarde van de diepte weinig invloed heeft op de onzekerheid in de index. Daarnaast geven experts op dit deel van de curve een lage onzekerheid aan in de habitatfunctie, aangezien de diepte hier sterk beperkend is (relatief lage indexwaardes).

Punt E:

In punt E is orthofosfaat hoofdzakelijk beperkend. De onzekerheid in de indexwaarde is relatief gering omdat punt E op een positie in de functie ligt waar orthofosfaat sterk beperkend is en waar de experts ook vrij zeker van zijn (Tabel 3.4). Daarnaast is de standaarddeviatie in de waarde van orthofosfaat gering, waardoor het effect op de standaarddeviatie in de index ook gering zal zijn.

Uit de analyse van punt A t/m E is de bijdrage van individuele invoervariabelen aan de onzekerheid in habitatgeschiktheid en de interactie tussen onzekerheid in index en in factorwaarde als volgt samen te vatten.

Het blijkt dat het effect van onzekerheid in factorwaarde en index sterk afhankelijk is van de positie van het punt op de habitatfunctie van de beperkende factoren. Op steile delen van de habitatfunctie heeft onzekerheid in factorwaarde het grootste effect op de onzekerheid in indexwaarde, en wanneer deze beperkend is, dus ook op onzekerheid in de habitatgeschiktheid. Verder is op punten met indexwaardes rond de 0.5 de onzekerheid in de functie het grootst (Tabel 3.4), waardoor de bijdrage van deze onzekerheid in habitatgeschiktheid in potentie ook groot is (namelijk waar de factor beperkend is). De maximale standaarddeviaties worden bereikt wanneer bovengenoemde situaties tegelijk voorkomen, zoals in punt B.

Wanneer we de resultaten van de detailanalyse projecteren op het algemene beeld komt voor de onzekerheid in de habitatgeschiktheid naar voren dat de factoren doorzicht en orthofosfaat een belangrijke bijdrage leveren aan de onzekerheid. Dus ruimtelijke en temporele verschillen in waterkwaliteit zorgen voor een hoge onzekerheid in de habitatgeschiktheid. Dit sluit aan bij de kennis over de aanwezigheid van schedefonteinkruid in het IJsselmeergebied, waarbij jaarlijks grote variaties optreden in de vegetatiebedekking en de locaties van schedefonteinkruidvelden.

Met de geschatte standaarddeviaties in de berekende habitatgeschiktheid kan de vraag gesteld worden wat een zinvolle klassenverdeling is voor een woordelijke omschrijving van de habitatgeschiktheid.

Een verdeling met te smalle klassen maakt dat de kans dat een habitatgeschiktheid in een bepaalde klasse valt soms net zo groot is als de kans op indeling in bovenliggende of onderliggende klassen. Het feit dat de onzekerheid in habitatgeschiktheid sterk afhangt van de waarde van de geschiktheidindex, pleit voor een variabele indeling in klassenbreedte. De standaarddeviatie bij indices rond de 0.5 is ongeveer 0.2, waardoor een klasse van 0.3-0.7 geschikt lijkt als klasse voor “matig geschikt”. De verdere indeling kan zijn: een klasse van 0 voor “ongeschikt” en 0-0.3 voor “nauwelijks geschikt” en 0.7-1 voor “optimaal geschikt”.

Tabel 3.1 Parameters van de habitatfuncties die gebruikt zijn in de Monte Carlo simulatie van de habitatgeschiktheid van scheidfonteinkruid. De waarde van de factor is de x-waarde van het breekpunt in de gebroken lineaire optimumfunctie; de gemiddelde index is de y-waarde van de breekpunten en het minimum en maximum geven de uitersten aan van de bandbreedte waarbinnen een index kan variëren. De gemiddelden, minima en maxima zijn afgeleid uit resultaten van een expertmeningen-enquête naar onzekerheden in o.a. het habitatgeschiktheidsmodel van scheidfonteinkruid (Van der Lee et al., 2000).

Diepte (m)				Strijk lengte (m)			
waarde factor	gemid. index	minimum	maximum	waarde factor	gemid. index	minimum	maximum
0.00	0.00	0.00	0.00	0	0.30	0.15	1.00
0.10	0.05	0.03	0.34	500	0.30	0.15	1.00
0.40	0.50	0.27	0.83	1500	0.70	0.30	1.00
0.80	0.95	0.53	1.00	3500	0.95	0.30	1.00
1.25	0.50	0.39	0.86	4500	0.50	0.16	1.00
1.75	0.20	0.17	0.68	5500	0.25	0.19	1.00
4.00	0.05	0.02	0.32	15000	0.00	0.00	0.00
6.00	0.00	0.00	0.00				
Orthofosfaat (mg/l)				Doorzicht (m)			
waarde factor	gemid. index	minimum	maximum	waarde factor	gemid. index	minimum	maximum
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05
0.02	0.05	0.00	0.54	0.20	0.05	0.00	0.26
0.06	0.50	0.21	0.92	0.60	0.95	0.48	0.98
0.13	0.80	0.36	1.00	10.00	0.95	0.79	0.96
0.21	0.95	0.41	1.00				
0.38	0.50	0.16	0.81				
0.55	0.05	0.00	0.41				
0.70	0.00	0.00	0.00				

Tabel 3.2 Gemiddelde standaarddeviatie van de index (y-waarde breekpunt) van een habitatfunctie, afhankelijk van de range voor de gemiddelde waarde van de index. De gemiddelde standaarddeviaties zijn afgeleid uit resultaten van een expertmeningen-enquête naar onzekerheden in o.a. het habitatgeschiktheidsmodel van scheidfonteinkruid (Van der Lee et al., 2000)

Index	
range (-)	stdev (-)
0.00-0.15	0.023
0.15-0.25	0.056
0.25-0.35	0.083
0.35-0.45	0.099
0.45-0.55	0.116
0.55-0.65	0.099
0.65-0.75	0.083
0.75-0.85	0.056
0.85-1.00	0.023

Tabel 3.3 Standaarddeviaties voor de waarden van de invoervariabelen, afhankelijk van de range. De gemiddelden van de invoervariabelen zijn weergegeven in Figuur 3.1. De uitersten van de invoervariabelen zijn berekend als + en - 0.5* het gemiddelde

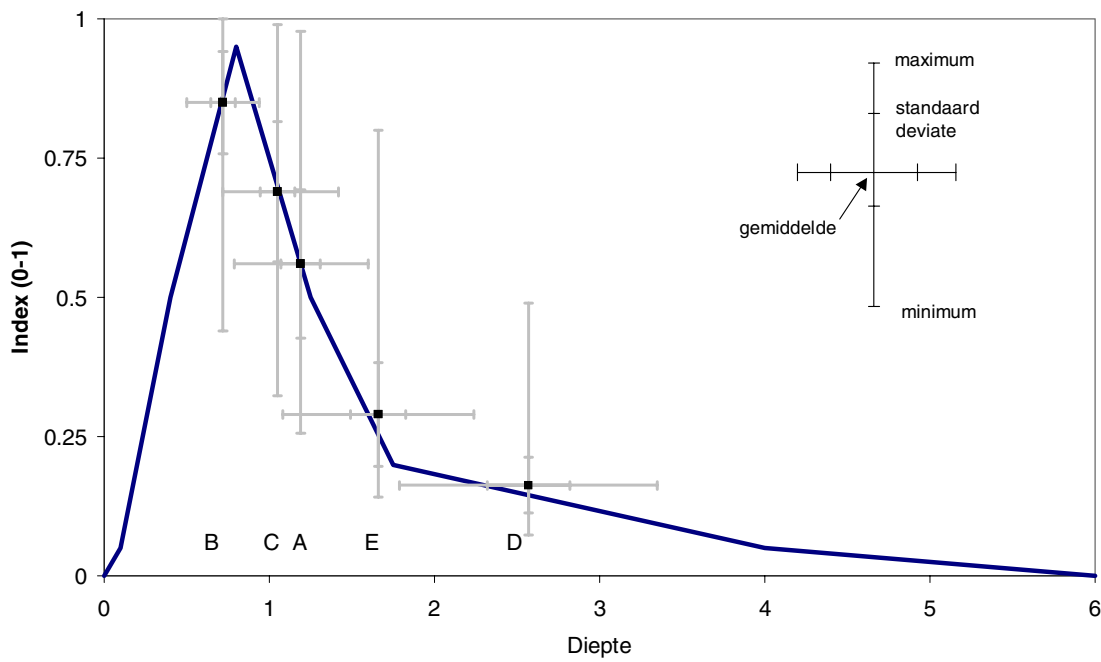
Diepte (m)		Orthofosfaat (mg/l)		Strijk lengte (m)		Doorzicht (m)	
range (-)	stdev (-) ^a	range (-)	stdev (-)	range(-)	stdev(-) ^b	range (-)	stdev (-)
0.0-0.1	0.008	0-1*10 ⁻⁴	0	0-200	0	0.0-0.1	1*10 ⁻⁶
0.1-0.2	0.008	1*10 ⁻⁴ -0.025	0.01	200-500	200	0.1-0.2	0.05
0.2-0.3	0.008	0.025-0.05	0.03	500-1500	200	0.2-0.4	0.2
0.3-0.4	0.017	0.05-0.10	0.07	1500-3500	500	0.4-0.5	0.2
0.4-0.5	0.017	0.10-0.15	0.12	3500-4500	800	0.5-0.6	0.3
0.5-0.6	0.017	0.15-0.25	0.17	4500-5500	1000	0.6-0.8	0.4
0.6-0.7	0.017			5500-200000	1400	0.8-1.0	0.5
0.7-0.8	0.017						
0.8-3.0	0.017						
3.0-50	0.033						

^aBij standaarddeviaties lager dan de minimumwaarde in de Monte Carlo simulatie wordt de standaarddeviatie naar boven afgerond op deze minimumwaarde. De minimumwaarde is gedefinieerd als standaarddeviatie_{min} = (maximum - minimum) * 0.1.

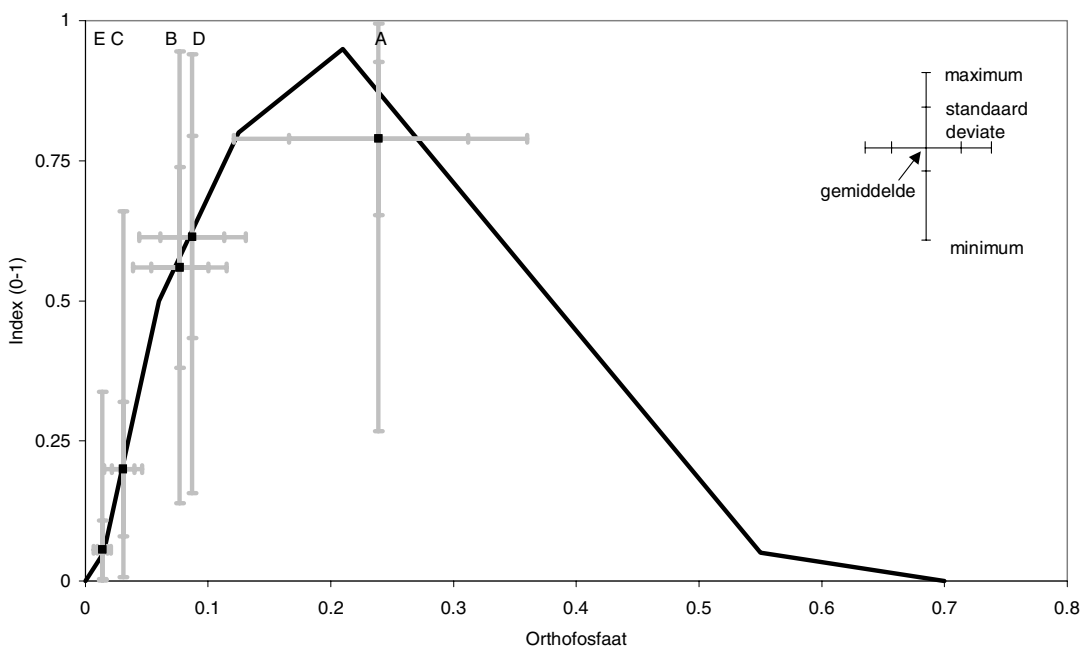
^bDe opgegeven standaarddeviaties zijn in sommige gevallen hoger dan mogelijk is bij een bepaald gemiddelde en uitersten in bandbreedtes. De standaarddeviatie wordt in die gevallen gelijk gesteld aan de maximaal mogelijke standaarddeviatie.

Tabel 3.4 Voor 3 scenario's van onzekerheid de kans dat een invoervariabele beperkend is op de 5 puntlocaties (0 < kans < 1),: 1. alleen de onzekerheid in habitatfuncties is meegenomen (index); 2. alleen de onzekerheid in de waarde van de invoervariabele is meegenomen (factor); 3. zowel de onzekerheid in habitatfunctie als in invoervariabele is meegenomen (index + factor)

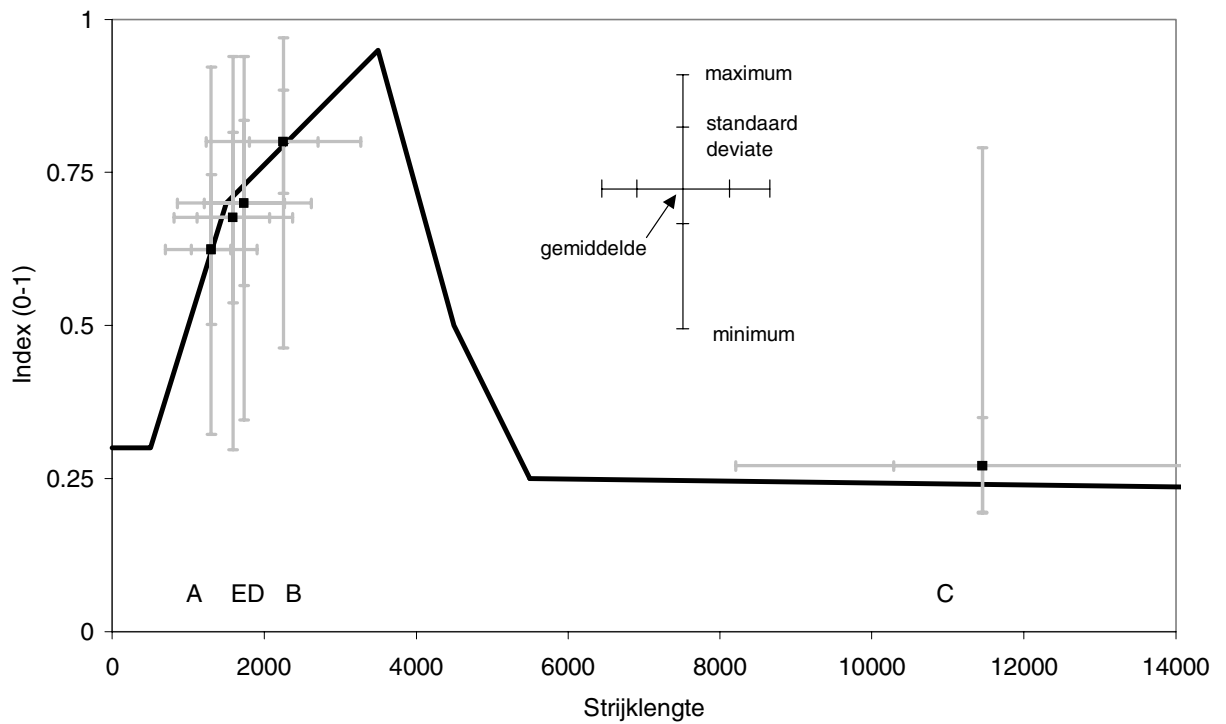
Puntlocatie	Onzekerheid in	Diepte	Orthofosfaat	Strijk lengte	Doorzicht
A	index	0	0	0	1
	factor	0	0	0	1
	index+factor	0.004	0	0.003	0.99
B	index	0.01	0.85	0.035	0.11
	factor	0.005	0.62	0.02	0.36
	index+factor	0.027	0.58	0.05	0.35
C	index	0	0.75	0.25	0
	factor	0	0.56	0.33	0.11
	index+factor	0	0.61	0.25	0.14
D	index	1	0	0	0
	factor	1	0	0	0
	index+factor	0.97	0.004	0	0.023
E	index	0.004	1	0	0
	factor	0	0.98	0	0.016
	index+factor	0.007	0.97	0	0.023



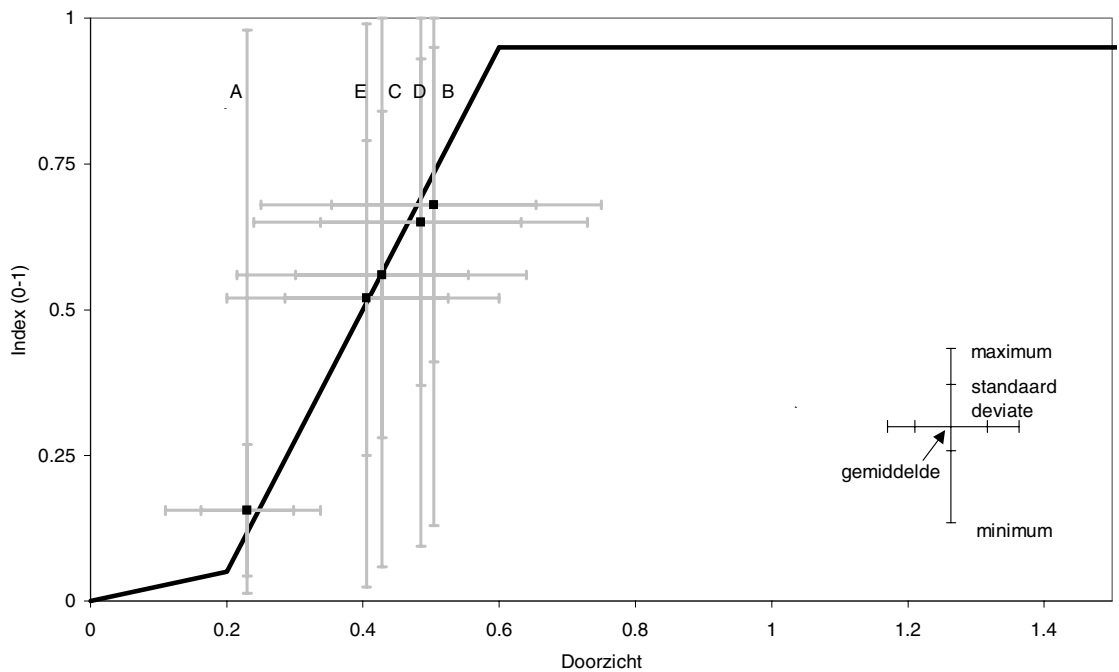
Figuur 3.6a Gemiddelde, standaarddeviatie, minimum en maximum voor de index van de invoervariabele diepte van de puntlocaties A t/m E in het IJsselmeergebied (zie Figuur 3.4). Behalve deze puntgegevens is ook de habitatfunctie als geheel weergegeven (gemiddelde index waarden) Zie Tabel 3.1 en 4.2 voor de minima, maxima en standaarddeviatie



Figuur 3.6b Gemiddelde, standaarddeviatie, minimum en maximum voor de index van de invoervariabele orthofosfaat van de puntlocaties A t/m E in het IJsselmeergebied (zie Figuur 3.4). Behalve deze puntgegevens is ook de habitatfunctie als geheel weergegeven (gemiddelde index waarden) Zie Tabel 3.1 en 4.2 voor de minima, maxima en standaarddeviatie. NB De gemiddelde index resulterend uit de onzekerheidsanalyse op punt A wijkt af van de habitatfunctie. Dit komt doordat punt A vlakbij het optimum van de functie ligt, waardoor zowel een afwijking lager als een afwijking hoger dan de gemiddelde orthofosfaatconcentratie resulteert in een lagere index dan de indexwaarde van de functie op punt A



Figuur 3.6c Gemiddelde, standaarddeviatie, minimum en maximum voor de index van de invoervariabele strijklengte van de puntlocaties A t/m E in het IJsselmeergebied (zie Figuur 3.4). Behalve deze puntgegevens is ook de habitatfunctie als geheel weergegeven (gemiddelde index waarden). Zie Tabel 3.1 en 4.2 voor de minima, maxima en standaarddeviatie



Figuur 3.6d Gemiddelde, standaarddeviatie, minimum en maximum voor de index van de invoervariabele doorzicht van de puntlocaties A t/m E in het IJsselmeergebied (zie Figuur 3.4). Behalve deze puntgegevens is ook de habitatfunctie als geheel weergegeven (gemiddelde index waarden). Zie Tabel 3.1 en 4.2 voor de minima, maxima en standaarddeviatie

4 Valideren van habitatmodellen

4.1 Valideren van habitatmodellen met event detectie

4.1.1 Inleiding

Bij het valideren of kalibreren van (numerieke) modellen is het gebruikelijk om aan de hand van één of ander Goodness of Fit (GoF) criterium de overeenkomst tussen modeluitkomsten en metingen te kwantificeren. Bij aanpassingen in het model (zoals bijvoorbeeld variaties van instellingen van onzekere modelparameters, of vervangingen van een procesformulering door een andere) of bij het vergelijken van twee of meer modellen kan dan aan de hand van zo'n GoF de performance van de modelvarianten vergeleken worden, en/of beoordeeld worden welk een betere of zelfs beste performance vertoont. Het 'beter' of 'beste' zijn wordt op deze manier in sterke, zo niet volledige mate, bepaald door het gehanteerde criterium en dit moet derhalve met zorg en kennis van zaken (dat wil zeggen inzicht in de fysische processen en hun representatie in het model, representativiteit van de metingen, onzekerheden in model en metingen, etc.) gekozen worden.

In kwantitatieve modellen met 'continue uitkomsten' (b.v. waterbeweging, stoftransport, of regressiemodellen) worden vaak criteria gebruikt waarin modelvoorspellingen en metingen puntsgewijs vergeleken worden, en het criterium een functie is van hun verschillen. Dat leidt vaak tot (al dan niet gewogen) kleinste kwadraten criteria. Afhankelijk van specifieke doelstellingen van een gebruiker worden daarvan soms varianten afgeleid, en kunnen criteria soms meer geaggregeerde grootheden bevatten.

Bij het valideren van habitatmodellen moet rekening gehouden worden met het feit dat een habitatmodel een potentiële geschiktheid voorspelt. Deze potentiële geschiktheid hoeft niet direct overeen te komen met een bepaalde waargenomen dichtheid van voorkomen in het veld, omdat andere factoren dan die in het habitatmodel beperkend kunnen zijn zoals populatiedynamische factoren. De door het model voorspelde geschiktheid hoeft dus niet één op één te corresponderen met de waargenomen dichtheid. Op basis hiervan ligt het voor de hand om een modeltest te kiezen die niet zozeer de voorspellende waarde voor de actuele dichtheid test, maar slechts of het model kan voorspellen of de soort wel of niet zal voorkomen. Dit zou dus een test zijn voor modellen met binaire uitkomsten: gegeven bepaalde invoergegevens moet het model voorspellen of een bepaalde event wel of niet optreedt.

Als illustratief voorbeeld wordt genoemd het aan de hand van een medisch beeld (CT, MRI, of Röntgen) beslissen of een pathologie wel of niet optreedt. Het model kan hierbij uit een geformaliseerde procedure bestaan (één of andere feature detectie methode) of een oordeel van een specialist (perceptueel model).

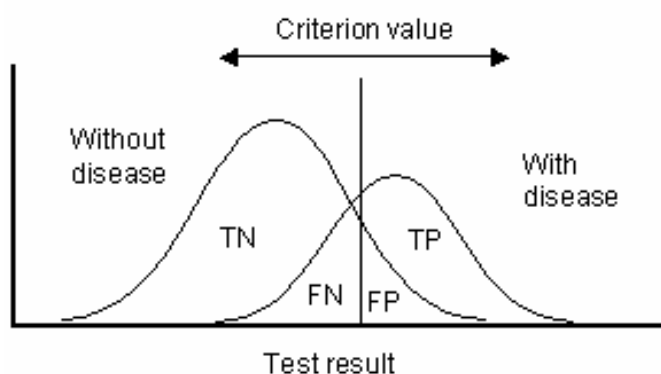
Het resultaat is een continue of een discrete beslissingsvariabele en als deze een drempel D overschrijdt zal het oordeel zijn dat een pathologie aanwezig is (of meer algemeen: event is aanwezig), en in het andere geval afwezig.

In dit soort 'decision making' kunnen diverse bronnen van onzekerheden optreden. Dat zijn in het algemeen vooral onzekerheden in het model dat de beslissingsvariabele berekent, maar kunnen ook (systematische of random) fouten in de metingen van de beslissingsfactoren omvatten. Daarnaast is het de vraag of het beslissingscriterium überhaupt in staat is om events te detecteren. Het criterium kan bijvoorbeeld niet representatief zijn, of de onzekerheden in model en data zijn van een omvang dat de signaal/ruis ratio te groot is om goed te kunnen discrimineren tussen wel of niet optreden van een event. Bovendien is er nog de vraag hoe een drempel D moet worden gekozen.

In zo'n beslissingsprocedure kunnen dan de volgende vier varianten optreden:

- True Positive (TP): een event is aanwezig en wordt als zodanig correct voorspeld.
- False Positive (FP): er wordt ten onrechte een event voorspeld.
- False Negative (FN): een event is aanwezig maar wordt niet voorspeld.
- True Negative (TN): een event is niet aanwezig en wordt als zodanig correct voorspeld.

In onderstaande Figuur 4.1 wordt één en ander nader geïllustreerd. De linkergrafiek geeft de kansverdelingfunctie voor cases zonder event, en de rechter voor cases met aanwezigheid van een event. De verticale lijn geeft de drempel voor een beslissingsvariabele. De bijhorende TP, TN, FP en FN zijn de aangegeven oppervlakten links en rechts van de verticale lijn. Hier hebben de verdelingsfuncties een aanzienlijke overlap zodat er inderdaad misclassificaties kunnen optreden, zowel bij deze drempel als bij een andere. Is bijvoorbeeld de drempel heel laag dan zullen opgetreden events altijd herkend worden, echter ten koste van veel FP (veel of alle niet-events worden toch als event geclassificeerd). Is de drempel heel hoog, dan worden vrijwel alle niet-events correct herkend, maar dit gaat ten koste van een grote fractie van FN.



Figuur 4.1

Het is dan de vraag hoe aan de hand van een tabel met TP, TN, FP en FN waarden, en diens afhankelijkheid van een drempel D , geschikte criteria of kentallen beschikbaar zijn voor het beoordelen van de geschiktheid van een model (of test) voor het detecteren van events, en detecteren van niet-events. Hierop wordt in de volgende paragraaf verder ingegaan.

4.1.2 GoF criteria voor events en niet-events

Gegeven een procedure voor het testen op het wel of niet optreden van een event kan een volgende tabel worden opgesteld:

Tabel 4.1

		<i>Event</i>		Total
		Present	Absent	
<i>Test</i>	Positive	<i>a</i> TP	<i>c</i> FP	<i>a+c</i>
	Negative	<i>b</i> FN	<i>d</i> TN	<i>b+d</i>
Total		<i>a+b</i>	<i>c+d</i>	<i>N=a+b+c+d</i>

Dat betekent dat van een populatie van N uitgevoerde experimenten:

- het aantal cases is waarvan een event correct werd voorspeld;
- het aantal is waarvoor geen events werden voorspeld terwijl die in werkelijkheid wel optraden;
- het aantal cases waarvoor ten onrechte een event werd voorspeld; en
- het aantal cases waarvoor terecht geen event werd voorspeld.

Merk op dat deze tabel is gegeven bij één bepaalde drempel D . Op variaties van die drempel die o.a. tot een ROC (receiver operating characteristic) curve leiden wordt later teruggekomen. Aan de hand van de uitkomsten in de tabel kunnen dan de volgende GoF kentallen gedefinieerd worden:

- Sensitivity (ook wel TP rate of TP-ratio genoemd);
- Specificity (ook wel TN rate of TN-ratio genoemd);
- Accuracy;
- Positive Predictive Value (PPV);
- Negative Predictive Value (NPV);
- Positive Likelihood ratio;
- Negative Likelihood ratio.

De definitie en betekenis van deze kentallen is als volgt:

- **Sensitivity (True Positive Rate):** De kans dat de test een event voorspelt wanneer (dat wil zeggen onder de conditie) dat event in werkelijkheid ook optreedt. In een formule: $TPR = \frac{a}{a+b}$. Vaak wordt deze met een percentage weergegeven. Hoe dichter bij 100% des te beter het model het optreden van events voorspelt (en des te minder events over het hoofd ziet).

- **Specificity (True Negative Rate):** De kans dat de test de afwezigheid van een event voorspelt wanneer in werkelijkheid zo'n event inderdaad afwezig is. In een formule: $TNR = \frac{d}{c+d}$. Vaak wordt deze met een percentage weergegeven. Hoe dichter bij 100% des te beter het model de afwezigheid van events voorspelt (en des te minder events voorspelt die er in werkelijkheid niet zijn).
- **Accuracy:** De kans dat het model het juiste antwoord geeft op het al dan niet optreden van een event. In een formule: $ACC = \frac{a+d}{a+b+c+d}$. Vaak wordt deze met een percentage weergegeven. Hoe dichter bij 100% des te beter het model events en niet-events voorspelt (en des te minder 'false alarms' produceert en events over het hoofd ziet).
- **Positive Predictive Value (PPV):** De kans dat in werkelijkheid een event is opgetreden wanneer dat als zodanig door het model wordt voorspeld. In een formule: $PPV = \frac{a}{a+c}$. Vaak wordt deze met een percentage weergegeven. Hoe dichter bij 100% des te minder het model 'false alarms' produceert.
- **Negative Predictive Value (NPV):** De kans dat in werkelijkheid geen event is opgetreden wanneer dat als zodanig door het model wordt voorspeld. In een formule: $NPV = \frac{d}{b+d}$. Vaak wordt deze met een percentage weergegeven. Hoe dichter bij 100% des te minder het model 'false negatives' produceert.
- **Positive Likelihood Ratio (PLR):** Het quotiënt van
 - de kans dat de test een event voorspelt wanneer dat event in werkelijk er is geweest (TPR: true positive rate, ofwel sensitivity) en
 - e kans dat de test een event voorspelt wanneer dat event er in werkelijkheid niet is geweest (FPR: false positive rate). In een formule: $PLR = \frac{TPR}{FPR} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$. Deze uitdrukking is ook te schrijven als $\frac{\frac{a}{c}}{\frac{a+b}{c+d}}$ en dan is dit het quotiënt van
 - de fractie van de door de test voorspelde events die ook in werkelijkheid een event waren en
 - de verhouding van het werkelijk aantal events en werkelijke aantal niet-events. Wanneer het optreden van events altijd correct voorspeld wordt, dat wil zeggen geen FP, dan is $c = 0$. In dit geval is PLR oneindig groot. Voor een goed presterend model moet de PLR zo groot mogelijk zijn (TPR is maximaal 1; FPR is minimaal 0).
- **Negative Likelihood Ratio (NLR):** Het quotiënt van
 - (i)de kans dat de test een niet-event voorspelt wanneer er in werkelijk een event is geweest (FNR: False Negative Rate) en
 - (ii)de kans dat de test een niet-event voorspelt wanneer dat event er in werkelijkheid ook niet is geweest (TNR: true negative rate). In een formule: $NLR = \frac{FNR}{TNR} = \frac{\frac{b}{a+b}}{\frac{d}{c+d}} = \frac{\frac{b}{d}}{\frac{a+b}{c+d}}$. De NLR wordt onbegrensd als $d=0$, dat wil zeggen geen TN, ofwel niet-events worden nooit correct voorspeld. De NLR moet zo klein mogelijk zijn.

Bovenstaande kentallen of criteria zijn opgesteld voor een gegeven situatie waarin onderscheid is gemaakt tussen events en niet-events op basis van één bepaalde drempelwaarde (zie voorgaande paragraaf). In de praktijk is het onderscheid tussen events en niet-events niet altijd gemakkelijk te maken. Een modelresultaat levert bijvoorbeeld een getal tussen 0 en 1 en er moet een keuze gemaakt worden wanneer dit als ‘positive’ en ‘negative’ wordt geïnterpreteerd. Bij een event speelt iets dergelijks. Je kunt niet altijd met zekerheid zeggen of iemand wel of niet ziek is en voor de aanwezigheid van waterplanten heb je te maken met verschillende bedekkingklassen. De keuze tussen ‘positive’ en ‘negative’ en tussen ‘present’ en ‘absent’ (event, waarneming) wordt drempel genoemd. Voor een optimale keuze van de drempelwaarde kunnen bovenvermelde kentallen als functie van de drempel worden berekend. Zo zullen bij een afnemende drempel de waarden van b en d naar nul gaan, zodat de sensitivity naar 1 gaat (gewenst) en tegelijkertijd de specificity naar 0 (ongewenst). Omgekeerd zullen bij een toenemende drempel a en c naar nul gaan zodat de sensitivity naar 0 gaat (ongewenst) en de specificity naar 1 (gewenst). Iets dergelijks geldt voor de andere combinaties van kentallen, en er moet dus één of ander optimum worden gezocht (bijvoorbeeld niet te lage waarden voor zowel sensitivity als selectivity). Voor die keuze is slechts op basis van de vraagstelling een criterium te definiëren.

4.1.3 ROC-curve

Het variëren van de drempel brengt ons bij het begrip ROC (Receiver Operating Characteristic). Hierin wordt voor variërende drempel de *sensitivity* (ofwel $TPR = \frac{a}{a+b}$) uit gezet tegen $1 - specificity$ (ofwel de $FPR = \frac{c}{c+d}$). De keuze van $1 - specificity$ in plaats van *specificity* heeft te maken met het feit dat de eerste net zoals de *sensitivity* stijgt met toenemende drempel zodat een monotone grafiek mag worden verwacht. Deze stijgt van het punt (0,0) (bij zeer hoge drempel) naar het punt (1,1) (bij zeer lage drempel) of het punt (100,100) als met percentages wordt gewerkt. In onderstaande Figuur 4.2 zijn drie voorbeelden van ROC-curves gegeven (bijvoorbeeld voor verschillende criteria; niet voor andere drempels).

Hierbij is de veronderstelling dat het testcriterium zo is opgezet dat naarmate de beslissingsvariabele groter wordt de kans op een event ook groter wordt. Voor de verdelingen van Figuur 4.1 betekent dit dat de verdeling van “without disease”, noem deze $F_-(\cdot)$ (cumulatieve verdeling), zich ‘links’ van de verdeling $F_+(\cdot)$ van “with disease” concentreert. Met die $F_+(\cdot)$ en $F_-(\cdot)$ kan dan het volgende voorschrift van de bijhorende ROC-curve (TPR als functie van FPR) worden afgeleid:

$$TPR = 1 - F_+(F_-^{inv}(1 - FPR)) \quad (4.1)$$

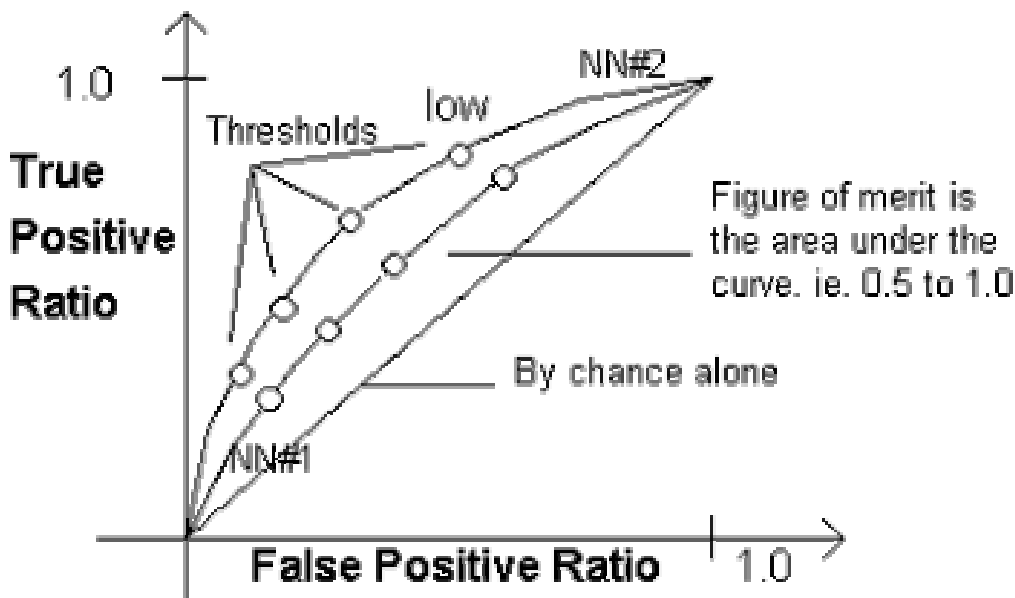
Deze relatie leidt dan inderdaad tot grafieken van de vorm zoals die in Figuur 4.2 zijn geïllustreerd.

De interpretatie die aan de positie van een ROC moet worden gegeven is als volgt.

Naarmate een test beter in staat is om zowel events als niet-events te classificeren zal de ROC-curve zich dichterbij de linker bovenhoek (dus nabij de verticale as $FPR=0$ en de horizontale lijn $TPR=1$) bevinden. Dit correspondeert met de situatie dat events en niet-events zeer goed onderscheiden kunnen worden.

Voor de verdelingsfuncties van Figuur 4.1 betekent dat deze geheel gescheiden zijn. Uit Formule 4.1 volgt dan dat in dit geval $TPR=1$ voor elke FPR .

Als ander uiterste is er de situatie dat de ROC-curve op, zeer nabij of zelfs onder de diagonaal ligt. Dat correspondeert met de situatie dat events en niet-events niet of vrijwel niet onderscheiden kunnen worden. Voor de verdelingsfuncties van Figuur 4.1 betekent dat deze geheel samenvallen. Inderdaad volgt uit Formule 1 dat $TPR=FPR$ als $F_+(\cdot) = F_-(\cdot)$. Wanneer de ROC-curve onder de diagonaal ligt voorspelt het model het optreden van events en niet-events stelselmatig verkeerd. Dit betekent dat events en niet-events waarschijnlijk wel in enige mate te onderscheiden zijn, maar dat het model niet goed is.



Figuur 4.2

Op basis van de ROC-curve kan derhalve de performance van de test voor het kunnen onderscheiden van events en niet-events worden beoordeeld. Daarnaast kan door het vergelijken van de ROC-curves horende bij verschillende testcriteria een idee worden gekregen of de ene test meer onderscheidend is dan de andere.

Er moet echter worden opgemerkt dat als een test een beperkt onderscheidend vermogen heeft dat niet betekent dat deze test niet deugt of verbeterd kan worden.

Immers de informatie waaruit de beslissingsvariabele wordt afgeleid kan dermate onvolledig zijn, of dermate verstoord met ruis, dat op geen enkele manier events en niet-events te onderscheiden zijn. Bovendien is het de vraag of events en niet-events überhaupt gescheiden zijn: als events en niet-events bijvoorbeeld door een volledig random proces gegenereerd worden zal geen enkel criterium het al dan niet optreden van events kunnen voorspellen.

4.1.4 AUC: Area under ROC-curve

Bij een ROC-curve wordt vaak het oppervlak onder de kromme berekend: AUC (Area Under Curve). Deze AUC is een getal tussen 0.5 en 1, en moet als volgt geïnterpreteerd worden:

Een waarde $AUC \in [0,1]$ betekent dat in $AUC \cdot 100\%$ van de gevallen de uitkomst van de beslissingsvariabele voor een a-select gekozen individu uit de populatie der events groter zal zijn dan de uitkomst van de beslissingsvariabele voor een a-select gekozen individu uit de populatie der niet-events.

Als de beslissingsvariabele niet kan onderscheiden, zal de ROC-curve op de diagonaal liggen, en is de $AUC=0.5$. Bij volledige onderscheidbaarheid, en de ROC-curve overal gelijk is aan 1 is, is de $AUC=1$.

4.1.5 Voorbeeld berekening ROC-curve en AUC-waarde

In dit voorbeeld worden voor de volgende uitkomsten van een diagnostische test de punten van de ROC-curve berekend. In deze case zijn 100 experimenten uitgevoerd. In elk van de experimenten is een medisch beeld geregistreerd, en voor elk beeld is door een specialist beoordeeld of er wel of niet een pathologie aanwezig is. Hier is die beoordeling (dat wil zeggen de uitkomst van de beslissingsvariabele) in discrete vorm uitgevoerd (een beoordeling in 5 klassen, en dus niet met een continu verlopende beslissingsvariabele). Per beeld neemt de specialist namelijk één van de volgende beslissingen ten aanzien van het vóórkomen (event) of niet voorkomen (niet-event) van een pathologie:

- ‘pathologie zeker niet aanwezig is’ (- -); of
- ‘pathologie waarschijnlijk niet aanwezig’ (-); of
- ‘weet het niet’ (0); of
- ‘pathologie waarschijnlijk aanwezig’ (+); of
- ‘pathologie zeker aanwezig’ (++)

Van de populatie bevatten 50 individuen de pathologie, en de andere 50 niet. Als uitkomst van de beoordeling werd (bijvoorbeeld) de volgende tabel met getelde events en niet-events gevonden:

Tabel 4.2

	- -	-	0	+	++	Totaal
D+	2	4	10	14	20	50
D-	28	14	5	2	1	50
Totaal	30	18	15	16	21	100

Dit betekent bijvoorbeeld dat 4 van pathologische gevallen (D+) door de specialist als ‘waarschijnlijk niet pathologisch’ geïnterpreteerd werden. Of als ander voorbeeld: Bij 5 van de niet-pathologische gevallen (D-) werd als beoordeling ‘weet het niet’ gegeven.

Op basis van Tabel 4.2 kunnen dan de volgende 6 punten van de ROC-curve berekend worden:

Het eerste punt van de ROC-curve:

Bij het leggen van de drempel D onder de laagste klasse ‘- -’ (dat wil zeggen aan alle ‘hogere’ uitkomsten -, -, 0, +, en ++ van de beslissingsvariabele wordt een pathologie toegekend, dus het optreden van een event) ziet Tabel 4.1 er als volgt uit:

$D < --$		<i>Event</i>		
		Present	Absent	Total
<i>Test</i>	Positive	$a=50$ TP	$c=50$ FP	100
	Negativ e	$b=0$ FN	$d=0$ TN	0
Total		50	50	$N=100$

Dit geeft voor de $sensitivity=TPR = \frac{a}{a+b}$ de uitkomst 1, en voor de $FPR = \frac{c}{c+d}$ is het resultaat dan ook 1. Hiermee wordt het punt (1,1) van de ROC-curve gevonden, ofwel het rechter-boven punt dat altijd in een ROC-curve aanwezig is.

Het tweede punt bij de ROC-curve:

Bij het leggen van de drempel D bij de klasse ‘- -’ (dat wil zeggen aan alle ‘hogere’ uitkomsten -, 0, +, en ++ van de beslissingsvariabele wordt een pathologie toegekend, dus het optreden van een event) ziet Tabel 4.1 er als volgt uit:

$D=--$		<i>Event</i>		
		Present	Absent	Total
<i>Test</i>	Positive	$a=48$ TP	$c=22$ FP	70
	Negative	$b=2$ FN	$d=28$ TN	30
Total		50	50	$N=100$

Dit geeft voor de $sensitivity=TPR = \frac{a}{a+b}$ de uitkomst 0.96, en voor de $FPR = \frac{c}{c+d}$ is het resultaat dan 0.44. Hiermee wordt het punt (0.44, 0.96) van de ROC-curve gevonden.

Het derde punt bij de ROC-curve:

Bij het leggen van de drempel D bij de klasse ‘-’ (dat wil zeggen aan de ‘hogere’ uitkomsten 0, +, en ++ van de beslissingsvariabele wordt een pathologie toegekend, dus het optreden van een event) ziet Tabel 4.1 er als volgt uit:

$D=-$		<i>Event</i>		
		Present	Absent	Total
<i>Test</i>	Positive	$a=44$ TP	$c=8$ FP	52
	Negative	$b=6$ FN	$d=42$ TN	48
Total		50	50	$N=100$

Dit geeft voor de $sensitivity=TPR = \frac{a}{a+b}$ als uitkomst 0.88, en voor de $FPR = \frac{c}{c+d}$ is het resultaat dan 0.16. Hiermee wordt het punt (0.16, 0.88) van de ROC-curve gevonden.

Het vierde punt van de ROC-curve:

Bij het leggen van de drempel D bij de klasse '0' (dat wil zeggen aan de 'hogere' uitkomsten +, en ++ van de beslissingsvariabele wordt een pathologie toegekend) ziet Tabel 4.1 er als volgt uit:

$D=0$		<i>Event</i>		
		Present	Absent	Total
<i>Test</i>	Positive	$a=34$ TP	$c=3$ FP	37
	Negative	$b=16$ FN	$d=47$ TN	63
Total		50	50	$N=100$

Dit geeft voor de $sensitivity=TPR = \frac{a}{a+b}$ de uitkomst 0.88, en voor de $FPR = \frac{c}{c+d}$ is het resultaat dan 0.16. Hier wordt het punt (0.06, 0.68) van de ROC-curve gevonden.

Het vijfde punt van de ROC-curve:

Bij het leggen van de drempel D bij de klasse '+' (dat wil zeggen alleen aan de 'hogere' uitkomst ++ van de beslissingsvariabele wordt een pathologie toegekend) ziet Tabel 4.1 er dan als volgt uit:

$D=+$		<i>Event</i>		
		Present	Absent	Total
<i>Test</i>	Positive	$a=20$ TP	$c=1$ FP	21
	Negative	$b=30$ FN	$d=49$ TN	79
Total		50	50	$N=100$

Dit geeft voor de $sensitivity=TPR = \frac{a}{a+b}$ als uitkomst 0.40, en voor de $FPR = \frac{c}{c+d}$ is het resultaat 0.02. Hiermee wordt het punt (0.02, 0.40) van de ROC-curve gevonden.

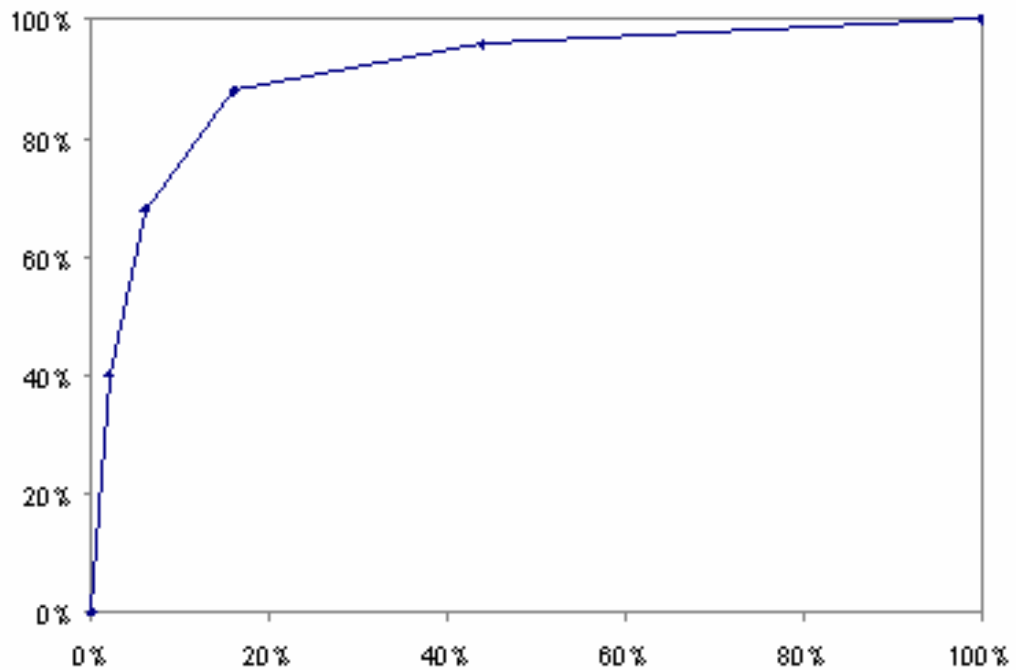
Het zesde punt van de ROC-curve:

Bij het leggen van de drempel D bij de klasse ‘++’ (dat wil zeggen elk beeld wordt als niet pathologisch geïdentificeerd) ziet Tabel 4.1 er als volgt uit:

$D=+$ +		<i>Event</i>		
		Present	Absent	Total
<i>Test</i>	Positive	$a=0$ TP	$c=0$ FP	21
	Negative	$b=50$ FN	$d=59$ TN	79
Total		50	50	$N=100$

Dit geeft voor de $sensitivity=TPR = \frac{a}{a+b}$ de uitkomst 0, en voor de $FPR = \frac{c}{c+d}$ is het resultaat dan ook 0, zodat het punt (0,0) van de ROC-curve wordt gevonden. Dat is het linker onder-punt dat in elke ROC-curve aanwezig is.

In onderstaande grafiek, zie Figuur 4.3, is de zo gevonden ROC-curve geïllustreerd, met de FPR langs de horizontale as en de TPR langs de verticale as. Hier zijn deze grootheden in percentages weergegeven.



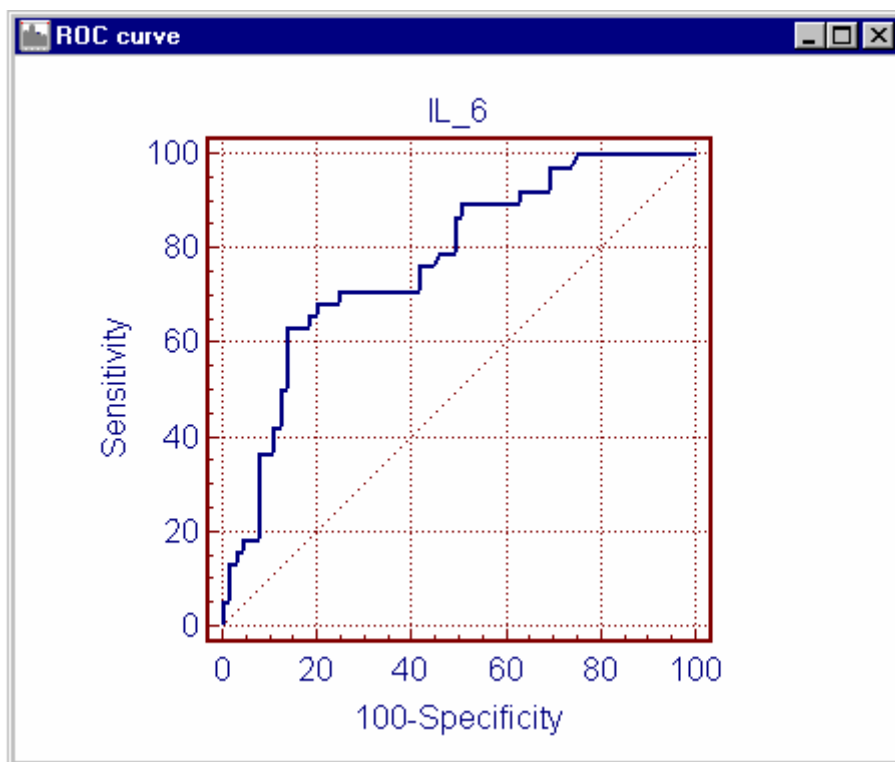
Figuur 4.3 ROC curve met de FPR op de x-as en TPR op de y-as.

In dit geval kunnen vanwege de indeling van de beslissingsvariabele in 5 klassen ‘slechts’ 6 punten van de ROC-curve berekend worden. Via lineaire interpolatie kan de ROC-curve dan in continue vorm berekend en getekend worden.

Bij een *continue-variërende* beslissingsvariabele is het recept om op een voldoende fijne partitie van die beslissingsvariabele de punten van de ROC te berekenen.

Is de beslissingsvariabele bijvoorbeeld een habitatgeschiktheid (die we h zullen noemen) met uitkomsten in het interval $[0,1]$ dan zou men een serie drempels van de vorm $D_k = \frac{k-1}{K-1}$ kunnen kiezen, voor $0 \leq k \leq K$. Bij een gegeven drempel D moet men het wel of niet optreden van een event dan als volgt definiëren: wèl een event als $h > D$ en gèèn event als $h \leq D$. Op deze manier worden dan $K+1$ punten van de ROC-curve gevonden. Voor $k=0$ geeft dat ‘vanzelf’ het punt $(1,1)$ en voor $k=K$ is dat het punt $(0,0)$. Voor K kan men bijvoorbeeld de waarde 101 kunnen nemen zodat de afstand tussen opvolgende drempels $\frac{1}{100}$ is. Er moet hierbij worden opgemerkt dat opvolgende punten van de ROC-curve op deze manier kunnen samenvallen. Immers, bij kleine veranderingen van de drempel kunnen de getallen a , b , c en d uit Tabel 4.1 hetzelfde blijven en pas bij grotere verschillen van de drempel gaan afwijken.

Op deze manier kunnen er tevens sprongen optreden in de ROC-curve. Dat is echter geen bezwaar. Een voorbeeld van zo’n ROC-curve is in onderstaande Figuur 4.4 weergegeven.



Figuur 4.4 ROC curve met de FPR op de x-as en TPR op de y-as. De curve is trapsgewijs opgebouwd door de berekening met stappen in de gekozen drempelwaarde.

Tenslotte wordt het recept gegeven voor het berekenen van de AUC (area under curve) gegeven de volgens bovenstaand recept berekende $K+1$ punten $\{(x_k, y_k)\}_{k=0}^K$ van de ROC-curve. Hierin zijn x_k en y_k de bij drempel D_k berekende FPR (ofwel *1-specificity*) en TPR (*sensitivity*) respectievelijk. Beide moeten zijn uitgedrukt in een fractie (dus als $FPR = \frac{c}{c+d}$ en $TPR = \frac{a}{a+b}$) en niet als percentage.

De puntenparen $\{(x_k, y_k)\}_{k=0}^K$ moeten geordend worden zodat de x_k stijgend (of althans niet dalend) zijn met toenemende index k (de bijhorende y_k zijn dan vanzelf ook niet-dalend). Dan geldt voor de AUC de volgende formule (trapezeumformule voor integratie):

$$AUC = \frac{1}{2} \cdot \sum_{k=1}^K (x_k - x_{k-1}) \cdot (y_k + y_{k-1}) \quad (4.2)$$

Deze AUC moet dan tussen 0.5 en 1 liggen. Voor een percentage moet de uitkomst met 100 vermenigvuldigd worden.

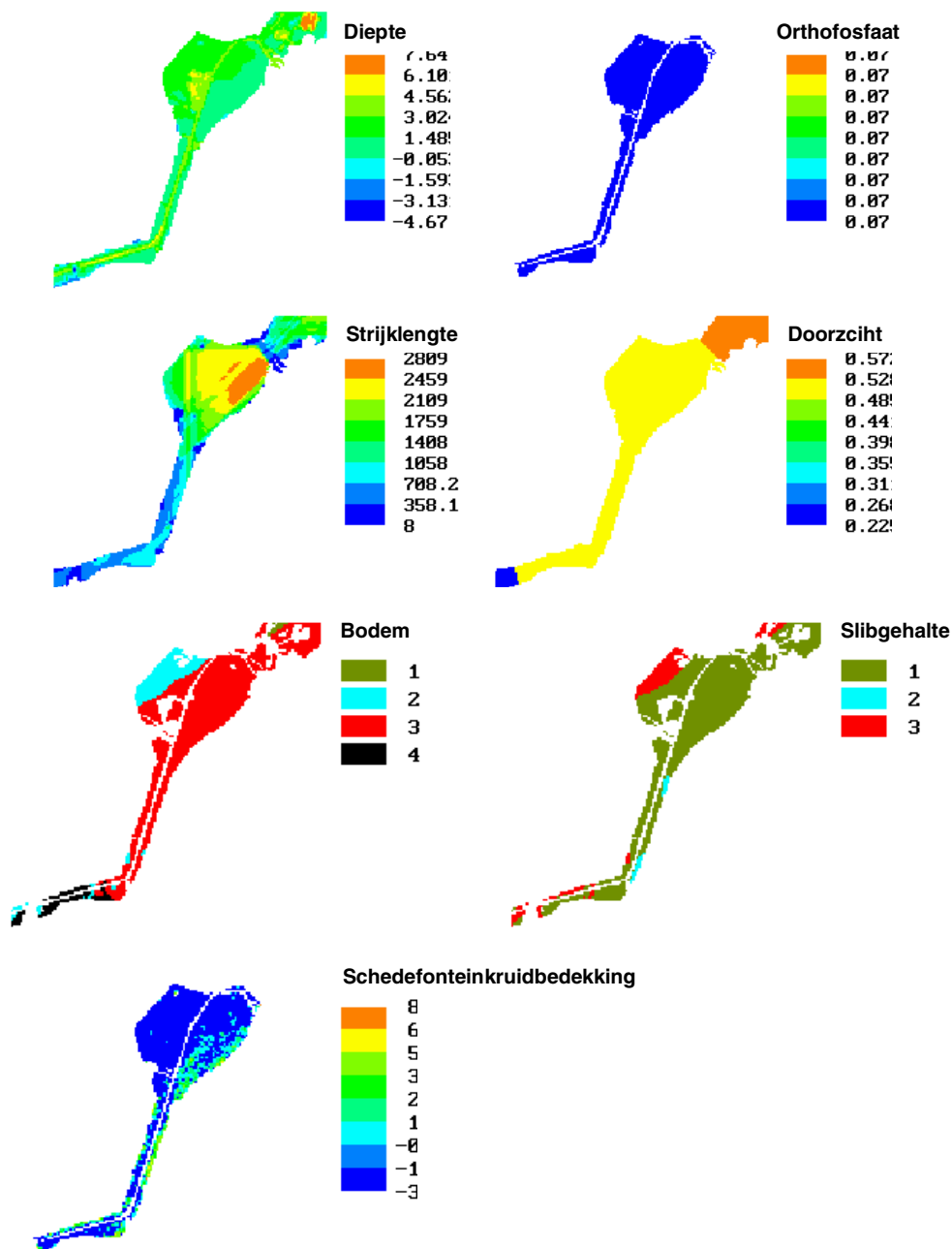
Voor het bovenstaande voorbeeld [dat wil zeggen de ROC van Figuur 4.3 met de zes punten $\{(0,0), (0.02,0.4), (0.06,0.68), (0.16,0.88), (0.44,0.96), (1,1)\}$] levert de AUC op deze manier het getal 0.91 op, ofwel 91%.

4.2 Toepassing: event detectie met het habitatmodel voor schedefonteinkruid

In hoofdstuk 3 is de onzekerheidsanalyse besproken voor het habitatmodel voor schedefonteinkruid. Deze onzekerheidsanalyse gaf inzicht in de eigenschappen van het habitatmodel en de betrouwbaarheid van de verschillende habitatfuncties en invoergegevens. Om een antwoord te krijgen op de vraag of de geschatte modelresultaten vervolgens goed overeenkomen met waarnemingen is een modelvalidatiestap nodig. In hoofdstuk 4 is beargumenteerd dat een methode van event detectie daarvoor geschikt is. Een geschikt criterium om de voorspellende waarde mee te kwantificeren is de Receiving Operating Characteristic (zie hoofdstuk 4).

4.2.1 Methoden

De event-detectiemethode is net als de onzekerheidsanalyse toegepast in het IJsselmeergebied voor het habitatmodel voor schedefonteinkruid (Laane, 1996). In het model zijn dezelfde invoervariabelen meegenomen als in de onzekerheidsanalyse: diepte, strijklengte, doorzicht en orthofosfaatconcentratie. Zie Figuur 4.5 voor een presentatie van de gebruikte invoerkaarten en zie hoofdstuk 4 voor de parameters van de habitatfuncties. Het habitatmodel voor schedefonteinkruid bevat ook klassefuncties voor de factoren bodemtype en slibrijkdom. Deze zijn in de onzekerheidsanalyse buiten beschouwing gelaten. In de event detectie methode zijn ze als aparte functies meegenomen buiten het habitatmodel om met als doel om de voorspellende waarde van bodemtype en slibrijkdom te analyseren. De gebruikte invoerkaarten voor bodemtype en slibrijkdom staan ook afgebeeld in Figuur 4.5.



Figuur 4.5 Kaarten van de invoervariabelen (diepte, orthofosfaatconcentratie, strijklengte, doorzicht, bodem en slibgehalte) en de schedefonteinkruidebedekking (klassen: -3 = niet gekarteerd).

In de event detectie methode wordt de voorspelde habitatgeschiktheid vergeleken met waarnemingen van schedefonteinkruide. De beschikbare veldgegevens over het optreden van schedefonteinkruide bestaat uit een kartering van schedefonteinkruide in 8 verschillende bedekkingklassen. De kartering van het hele IJsselmeergebied is opgebouwd uit karteringen van subgebieden uit verschillende jaren (Figuur 4.5).

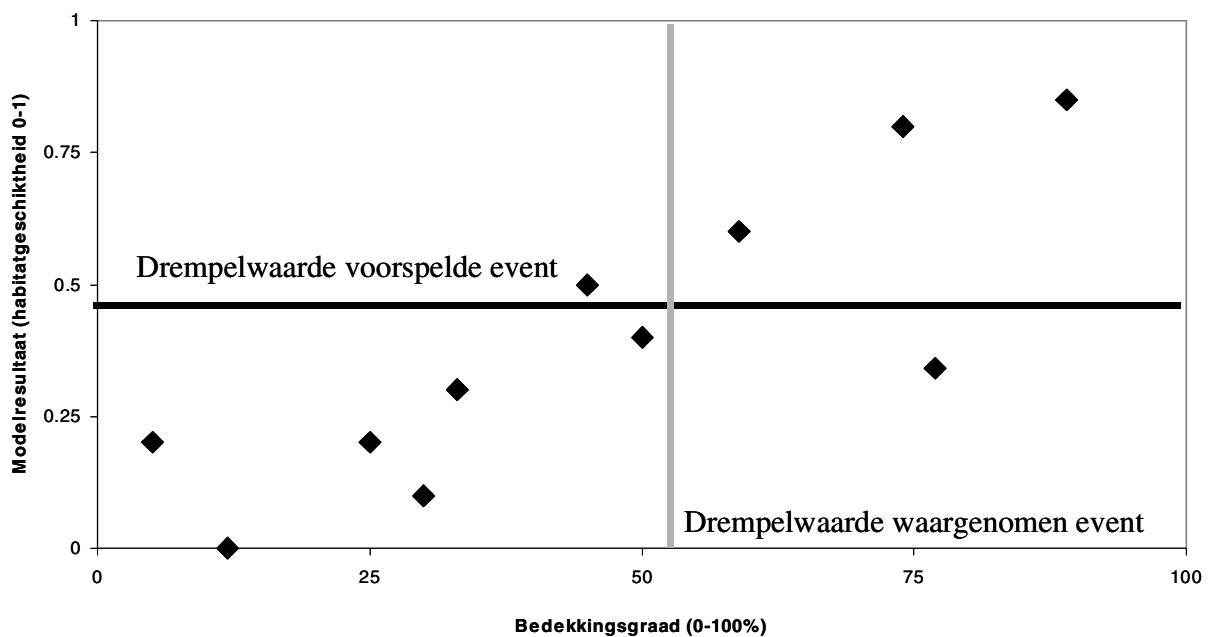
4.2.2 Definitie van waargenomen en voorspelde events

Om de event detectie methode toe te passen, is een definitie nodig van wanneer een event optreedt (waargenomen is) en wanneer een event voorspeld is. Deze definitie is toegelicht in Figuur 4.6. Waargenomen events zijn locaties waar schedefonteinkruid daadwerkelijk voorkomt. In deze case studie is een waargenomen event te definiëren als een locatie waar de bedekkingsgraad gelijk is aan of groter is dan een bepaalde drempelwaarde in bedekking. De event detectie is toegepast bij verschillende drempelwaardes in bedekkingsklasse.

Het optreden van een voorspelde event moet ook gedefinieerd worden omdat het modelresultaat van het habitatgeschiktheidsmodel een kaart met continue waardes is, variërend van 0 tot 1. Een voorspelde event is te definiëren als een locatie waar de berekende habitatgeschiktheid hoger is dan een bepaalde drempelwaarde.

In de event detectie methode wordt de drempelwaarde voor een voorspelde event gevarieerd tussen 0 en 1.

Voor de combinatie van de definitie voor een voorspelde en die van een waargenomen event worden de waarden bepaald van de goodness of fit criteria voor het onderscheidend/voorspellend vermogen van het habitatmodel. De gebruikte goodness of fit criteria die toegepast zijn in de event detectie methode zijn de ROC, de accuracy, de sensitivity en de specificity (zie hoofdstuk 4).



Figuur 4.6 Definitie van een waargenomen en van een voorspelde event.

4.2.3 ROC curves

ROC curves zijn bepaald bij verschillende drempelwaardes voor de waargenomen bedekkingsklasse van sedefonteinkruid. De gebruikte drempelwaardes voor de sedefonteinkruidbedekking zijn: klasse 1 (0-2.5% bedekking); klasse 2 (2.5-5% bedekking); klasse 3 (5-10% bedekking) en klasse 5 (25-50% bedekking).

De drempelwaarde in bedekkingsklasse waarbij het oppervlak onder de ROC curve het hoogst is, is de bedekkingsklasse waarbij het onderscheidend vermogen van het habitatmodel tussen events en niet-events het grootst is.

De ROC curve van het habitatgeschiktheidsmodel geeft het onderscheidend/voorspellend vermogen weer op basis van alle invoervariabelen die meegenomen zijn in het model (diepte, strijklengte, doorzicht, orthofosfaat). Om de performance van individuele habitatfuncties te testen zijn ook ROC curves berekend per habitatfunctie. Als extra functies zijn ook die van bodemtype en slibrijkdom getest.

Uit deze analyse blijkt welke functies een voorspellende waarde hebben en welke factoren dus bepalend zijn voor het voorkomen van sedefonteinkruid.

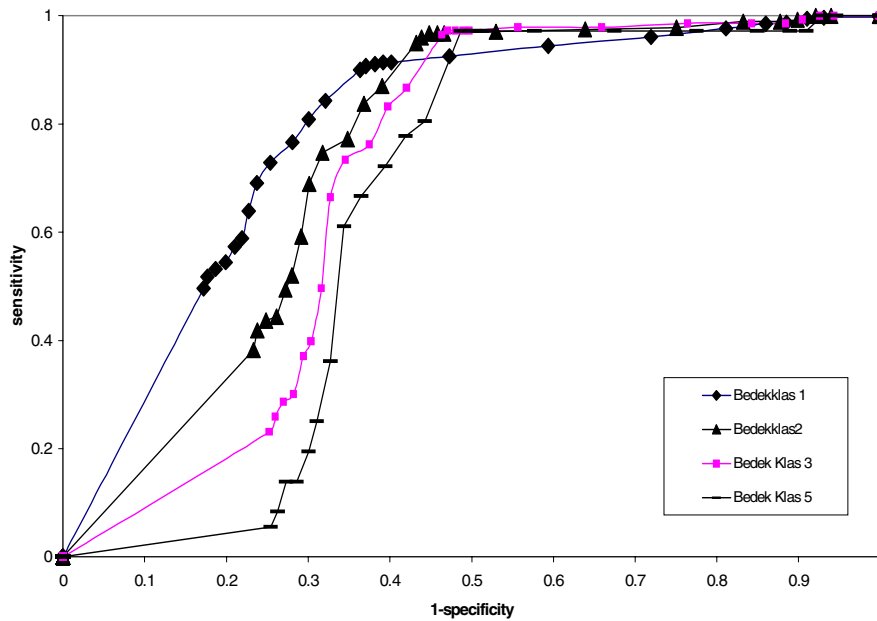
4.2.4 Accuracy, specificity en sensitivity

De criteria accuracy, specificity en sensitivity zijn gebruikt om de drempelwaarde in habitatgeschiktheid te kiezen waarbij het onderscheidend vermogen van het model tussen events en niet-events het hoogst is. Dit is de drempelwaarde waarbij accuracy, sensitivity en specificity alle zo hoog mogelijk zijn. Hierbij wegen de sensitivity en de accuracy het zwaarst aangezien een lage specificity (correcte voorspelling van niet-events) niet per se betekent dat het model niet correct is, dit omdat factoren die niet meegenomen zijn in het model beperkend kunnen zijn (bijvoorbeeld zaadverspreiding, vraat door vogels en anderen, zie inleiding hoofdstuk 4).

4.2.5 Resultaten

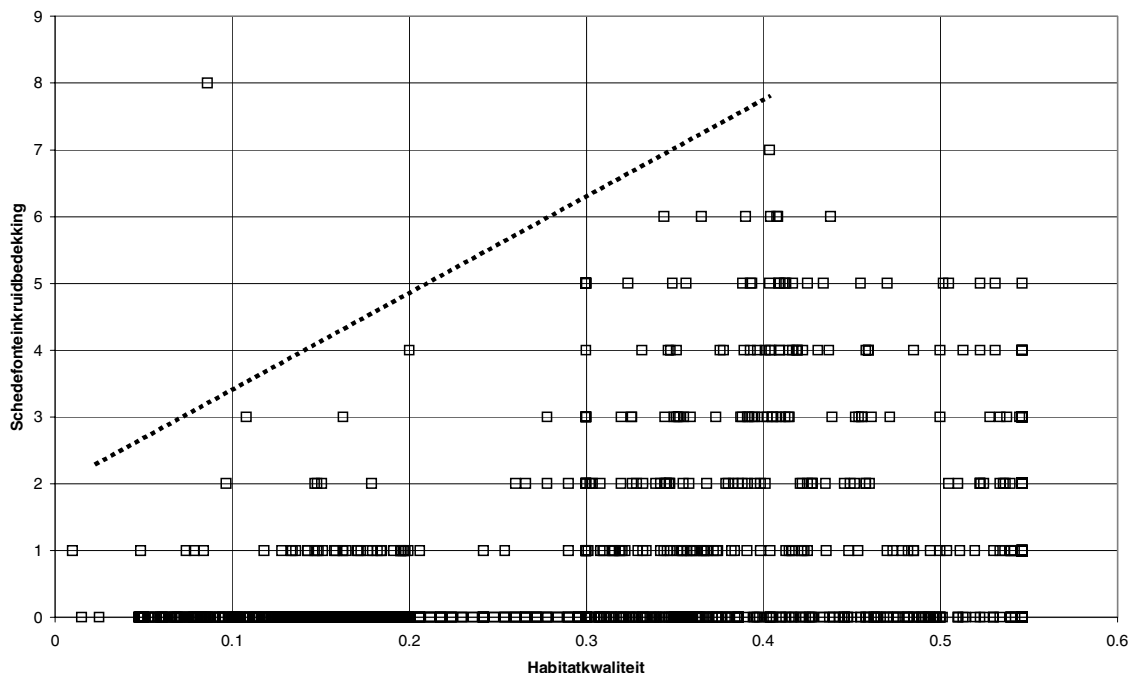
Figuur 4.7 toont de ROC curves van het habitatmodel bij verschillende drempelwaardes in bedekkingsklasse voor het optreden van een waargenomen event. In een ROC curve is de sensitivity uitgezet tegen 1-specificity (zie hoofdstuk 4.1). De ROC curve zou een 1:1 lijn zijn bij een model zonder enige voorspellende waarde en zou (als andere uiterste) convergeren naar een stapfunctie bij een perfect voorspellend model (volledige scheidbaarheid van de events en non-events). Het oppervlak onder deze curve (de zogenaamde AUC: Area Under Curve) is een kwantitatieve maat voor het voorspellend vermogen. De AUC is 50% bij een random model en 100% bij een perfect model.

In Figuur 4.7 is te zien dat bij gebruik van bedekkingsklasse 1 (0-2.5%) het oppervlak onder de curve het grootst is. Het oppervlak onder de ROC curves varieert van 64 tot 80% (Tabel 4.3). De maximale waarde van 80% is te interpreteren als een situatie waarin in 80% van de gevallen de habitatkwaliteit voor een a-selecte trekking uit de populatie der events groter zal zijn dan de habitatkwaliteit van een a-selecte trekking uit de populatie der niet-events.



Figuur 4.7 ROC (reciever operating characteristic) curves van het habitatmodel voor sedefonteinkruid, bij verschillende definities van een event: 1 de bedekking is groter dan 0%, 2 de bedekking is groter dan 2.5%; 3 de bedekking is groter dan 5% en 5 de bedekking is groter dan 25%. Ter vergelijking zijn de ROC van een random model en een perfect model gegeven. In het habitatmodel zijn de invoervariabelen diepte, strijklengte, doorzicht en orthofosfaat meegenomen.

Dat de prestatie van het model het hoogst is bij een bedekkingsklasse van 1 is uit te leggen met een vergelijking tussen Figuur 4.7 en 4.8. Het traject in Figuur 4.4 waarbij de sensitivity sterk afneemt begint vanaf een waarde voor '1-specificity' die lager is dan circa 0.45. Deze waarde komt overeen met een drempelwaarde in habitatgeschiktheid van circa 0.3 (resultaten niet gepresenteerd). Dat de sensitivity sterker daalt bij gebruik van hogere bedekkingsklasse als drempelwaarde voor een waargenomen event, is te begrijpen met Figuur 4.5. Voor een drempelwaarde van bedekkingsklasse 5 bijvoorbeeld zien we dat punten met deze bedekkingsklasse gaan voorkomen bij habitatgeschiktheden vanaf 0.3. Bij drempelwaardes in habitatgeschiktheid hoger dan 0.3 gaan dus waargenomen events voorkomen die niet voorspeld worden. Dit aantal is al snel relatief hoog ten opzichte van het aantal voorspelde en ook waargenomen events. De sensitivity als percentage correct voorspelde events neemt dus snel af. Bij bedekkingsklasse 1 als drempelwaarde is dit anders. Uiteraard neemt het aantal niet voorspelde maar wel waargenomen events ook geleidelijk toe als de drempelwaarde in habitatgeschiktheid groter wordt. Maar het aantal voorspelde en ook waargenomen aantal events blijft echter vrij groot, waardoor de sensitivity (het % correct voorspelde events) relatief hoog blijft.



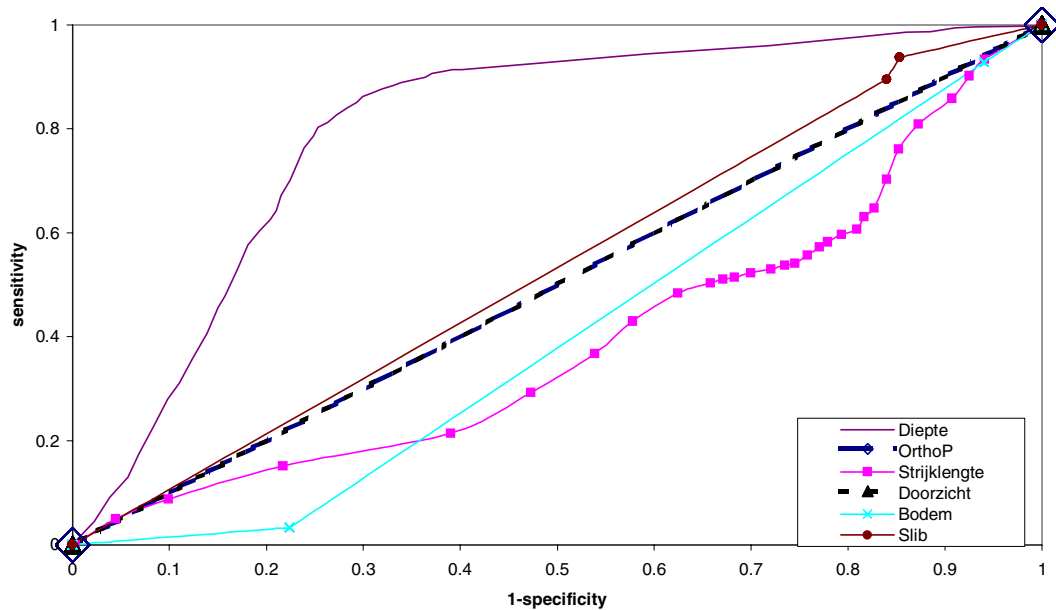
Figuur 4.8 XY-diagram van de schedefonteinkruiddedekking in gekarteerde gridcellen, afhankelijk van de voorspelde habitatkwaliteit.

Aangezien de prestatie van het model het beste is bij een drempelwaarde van bedekkingsklasse 1, wordt deze waarde in de onderstaande analyses verder gebruikt. Zoals gezegd is de waarde van de AUC 80% bij deze bedekkingsklasse (Tabel 4.3).

Deze waarde is aanzienlijk groter dan de 50% van een random model. Het habitatmodel voor schedefonteinkruiddedekking is op basis van de gebruikte invoergegevens dus tot op zekere hoogte in staat om het optreden van events en non-events te onderscheiden. Op dit moment is echter nog niet duidelijk of dit percentage het maximum haalbare is doordat 1. de invoerinformatie van het habitatmodel onvolledig is en/of te veel ruis bevat en/of 2. het AUC-percentage nog vergroot kan worden door de optimumfuncties van de invoervariabelen te verbeteren. Een methode om vast te stellen wat de maximale voorspelkwaliteit is die met de huidige informatie in de invoergegevens haalbaar is zou kunnen bestaan uit een neurale netwerk analyse. In zo'n neurale netwerk analyse worden relaties gelegd die een optimale fit geven tussen veldgegevens en invoerinformatie. Ofwel: het beste (maar op de data gebaseerd multivariaat regressie) model voor het voorspellen van de bedekkingsgraad als functie van de invoervariabelen. Wanneer een neurale netwerk niet in staat is om het % onder de ROC curve te vergroten, dan is duidelijk dat de invoerinformatie beperkend is voor het model en niet zozeer de kennisregels. In het andere geval is er de indicatie dat (de kennisregels in) het model of zelfs de modelstructuur verbeterd kunnen worden.

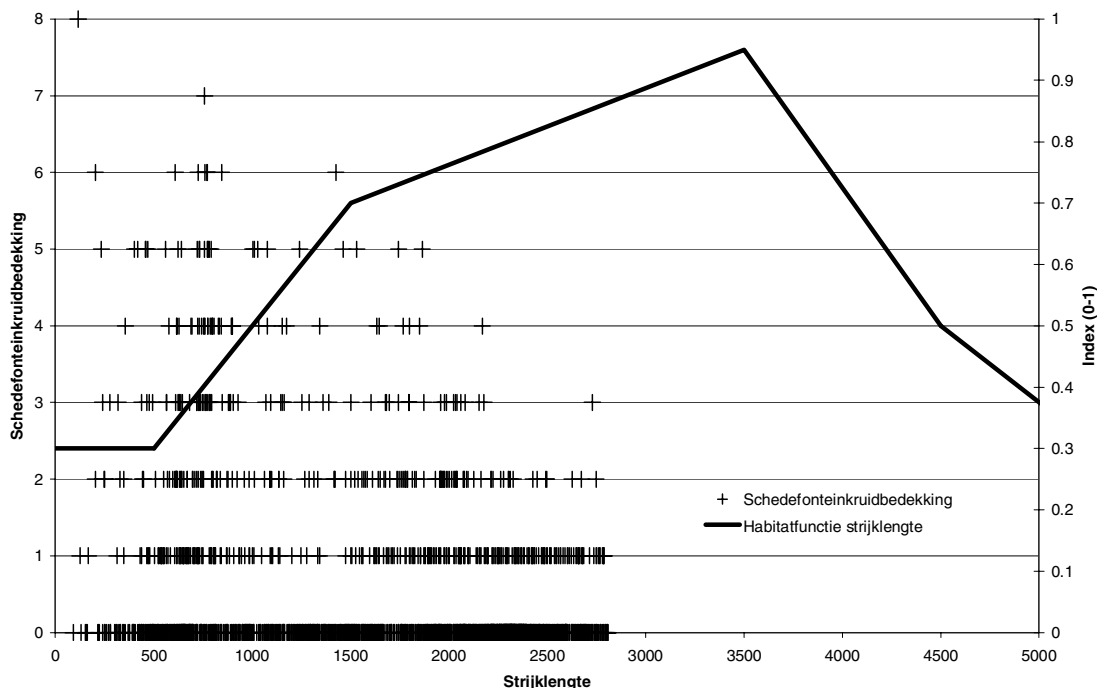
Figuur 4.9 geeft de ROC curves van individuele habitatfuncties. In deze figuur zijn ook de functies voor bodemtype en slibrijckdom opgenomen. Tabel 4.4 geeft de oppervlakken onder de ROC curves. De AUC van de functies orthofosfaat en doorzicht is 50%. Dit komt doordat de invoerkaarten voor doorzicht en fosfaat een constante waarde voor het gehele meer aangeven vanwege de beperkte meetpunten.

Receiver Operating Characteristic voor individuele habitatfuncties



Figuur 4.9 ROC (reciever operating characteristic) curves van individuele habitatfuncties van het habitatmodel voor schedefonteinkruid. Ter vergelijking zijn de ROC's gegeven van een random model en een perfect model.

En met maar één waarde is het onderscheidend vermogen van de habitatfunctie per definitie gelijk aan dat van een random model. Van de overige functies is de AUC van de habitatfunctie diepte het hoogst. Deze is zelfs even hoog als de AUC van het complete habitatmodel (zie Tabel 4.4), waaruit volgt dat de factor diepte geheel bepalend is voor het onderscheidend vermogen van het habitatmodel. De AUC van de functie strijklengte is significant lager dan 50%, wat betekent dat de functie de events en non events trendmatig fout onderscheidt. Hieruit valt af te leiden dat de strijklengte wel samenhangt met het voorkomen van schedefonteinkruid maar in een relatie die afwijkt van die in de gebruikte habitatfunctie. Figuur 4.10 geeft de relatie tussen strijklengte en de schedefonteinkruidbedekking, tezamen met de habitatfunctie voor strijklengte. Inderdaad is te zien dat schedefonteinkruidbedekkingen van 50-100% voorkomen bij strijklengtes van 0-1000 m waar de habitatfunctie aangeeft dat de geschiktheid laag is (0.3).



Figuur 4.10 XY-diagram van de schedefonteinkruidebedekking, afhankelijk van de striijklengte (eerste y-as), in combinatie met de geschiktheid afhankelijk van de striijklengte (tweede y-as).

De AUC van de functies bodemtype en slibrijkdome ligt rond de 50%, waaruit volgt dat de habitatfuncties voor bodemtype en slibrijkdome de events en non events niet significant van elkaar kunnen scheiden, ondanks dat de factoren lokaal wel beperkend zijn voor de habitatgeschiktheid (resultaten niet gepresenteerd). Een kanttekening bij de schijnbaar geringe invloed van bodemtype en slibrijkdome is dat het habitatmodel nu alleen is toegepast op het Wolderwijd. Een toepassing op het hele IJsselmeergebied leidt mogelijk tot een wat grotere bijdrage van bodemtype en slibrijkdome aan het onderscheidend vermogen van het model, door een hogere ruimtelijke differentiatie van de factoren en een mogelijk betere relatie met de schedefonteinkruidebedekking.

Een analyse van de sensitivity en specificity voor verschillende deelgebieden van het IJsselmeergebied zou meer inzicht kunnen geven in hoe de performance van het model afhangt van de combinaties van invoervariabelen.

Op deze manier kunnen fouten in kennisregels achterhaald worden, en het model mogelijk verbeterd.

Tabel 4.3 AUC van ROC curves voor de habitatgeschiktheid bij verschillende drempelwaarden in bedekkingsklasse van schedefonteinkruide. Alleen de klassen 1, 2, 3 en 5 zijn gebruikt.

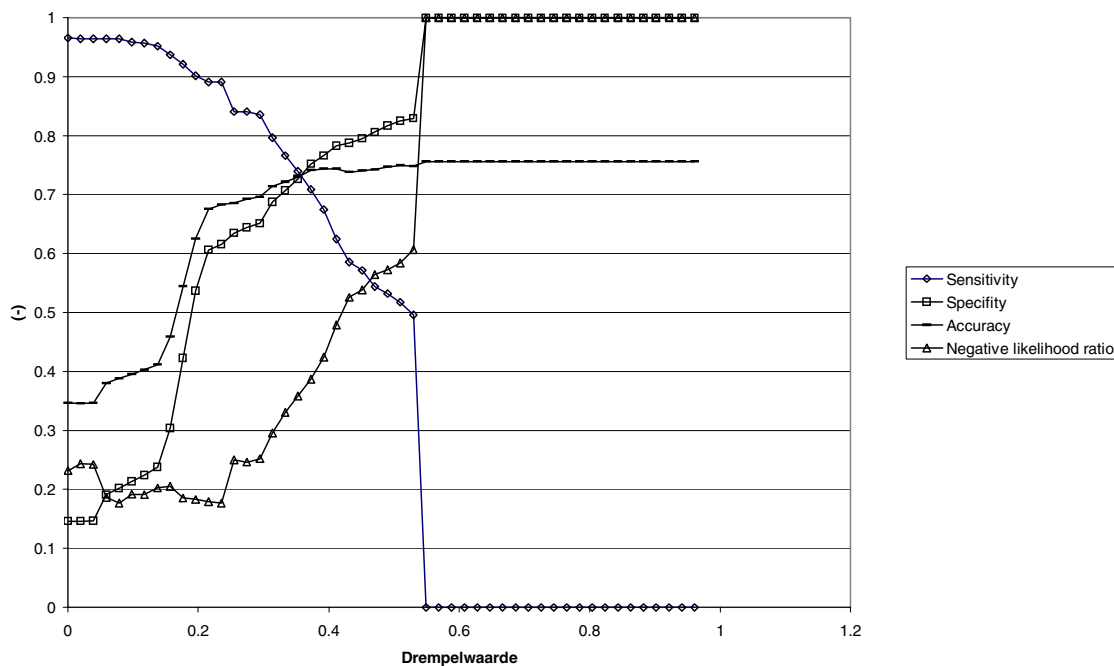
ROC curve	AUC
bedekkingsklasse 1 (0-2.5%)	0.79
bedekkingsklasse 2 (2.5-5%)	0.74
bedekkingsklasse 3 (5-10%)	0.70
bedekkingsklasse 5 (25-50%)	0.64

Tabel 4.4 AUC van ROC curves individuele habitatfuncties. De gebruikte drempelwaarde in bedekkingsklasse voor een event van schedefonteinkruid is 1

ROC curve	AUC
diepte	0.80
orthoP	0.50
strijk lengte	0.39
doorzicht	0.50
bodemtype	0.41
slibrijksdom	0.53

Wanneer het habitatmodel in de praktijk toegepast zou worden voor het voorspellen van events en non-events in het voorkomen van schedefonteinkruid moet men een drempelwaarde vaststellen voor de habitatgeschiktheid waarboven sprake is van een event. Deze drempelwaarde zou zodanig gekozen moeten worden dat het onderscheidend vermogen van het model het beste is. Een keuze van een drempelwaarde is mogelijk op basis van de criteria accuracy, sensitivity en specificity. Figuur 4.11 toont de accuracy, sensitivity, specificity voor het habitatmodel van schedefonteinkruid, afhankelijk van de (systematisch gevarieerde) drempelwaarde in habitatkwaliteit. De accuracy, bijvoorbeeld het aantal goed voorspelde events en non-events ten opzichte van het totaal neemt toe met de drempelwaarde. Tot een waarde van circa 0.2 is de toename groot maar voor nog hogere waarden van de drempel neemt de accuracy nog maar langzaam toe.

Van de juist voorspelde situaties (events + non-events) neemt het aantal juist voorspelde events (dat wil zeggen de sensitivity) af met de drempelwaarde, en het juist voorspelde aantal non-events (dat wil zeggen de specificity) toe. Bij de meest geschikte drempelwaarde zijn de accuracy, de sensitivity en specificity zo hoog mogelijk. De sensitivity als maat voor het juist voorspelde aantal events zou hierbij wat zwaarder moeten wegen dan de specificity (zie methode). Op grond hiervan en de eerder genoemde afwegingen zou een geschikte drempelwaarde dan circa 0.2-0.25 zijn. Hierbij is de sensitivity nog steeds relatief hoog (>0.9), en de accuracy en specificity zijn gestegen tot boven de 0.6-0.7 (Figuur 4.11). Bij hogere drempelwaarden neemt de sensitivity snel af terwijl accuracy en specificity relatief langzamer gaan toenemen.



Figuur 4.11 Accuracy, sensitivity en specificity van het habitatmodel voor schedefonteinkruid, afhankelijk van de drempelwaarde in voorspelde habitatkwaliteit. In het habitatmodel zijn de invoervariabelen diepte, strijklengte, doorzicht en orthofosfaat meegenomen.

4.2.6 Conclusies en aanbevelingen ten aanzien van validatie

De ROC methode geeft een kwantificering van het onderscheidend vermogen van het habitatmodel voor schedefonteinkruid en is een handvat om een goede drempelwaarde te selecteren. De drempelwaarde is belangrijk: de accuracy van het habitatmodel voor schedefonteinkruid neemt in traject van de habitatgeschiktheid tussen 0 en 0.3 bijvoorbeeld zeer sterk toe, hierna is de toename minder snel.

Het is aanbevelenswaardig om een neurale netwerkanalyse toe te passen op de dataset van waarnemingen voor schedefonteinkruid en de voorspelde habitatgeschiktheid. Met een dergelijke analyse is de maximale overeenkomst tussen modelresultaat en veldwaarnemingen vast te stellen die je mag verwachten op basis van de invoergegevens. Wanneer de maximale overeenkomst niet veel hoger is dan het onderscheidend vermogen van het model, kunnen de habitatfuncties niet significant verbeterd worden op basis van de gebruikte dataset.

In toekomstige toepassingen van de ROC methode zouden onzekerheden in de modelvoorspellingen meegenomen moeten worden bij de bepaling van events en non-events.

De invoergegevens voor waterkwaliteit zijn zeer beperkt, waardoor de potentiële voorspellende waarde voor de habitatgeschiktheid ook beperkt is. Per watersysteem is slechts één waarde beschikbaar.

De ruimtelijke verdeling van waterkwaliteit binnen een watersysteem zou vergroot kunnen worden met behulp van waterkwaliteitsmodellen of ruimtelijke interpolatietechnieken.

NB: De onzekerheidsanalyse gaf aan dat de onzekerheid in het modelresultaat lokaal in belangrijke mate door onzekerheid in de expertfuncties voor waterkwaliteit veroorzaakt werd.

Nauwkeuriger voorspellingen van de waterkwaliteit zelf hoeven dus niet per se te leiden tot een significante afname van de onzekerheid, omdat de onzekerheid in de functies ook groot is.

5 Afsluitende discussie en mogelijkheden voor implementatie

De conclusie uit de onzekerheidsanalyse voor het habitatmodel voor schedefonteinkruid was dat de onzekerheid in de voorspelde habitatgeschiktheid een onderscheid in drie geschiktheidsklassen toelaat. De validatie van het habitatmodel voor schedefonteinkruid gaf aan dat het model het voorkomen van events en non-events redelijk kan voorspellen, maar dat de prestatie van het model niet zodanig goed is dat nog een verder significant onderscheid in verschillende habitatgeschiktheidsklassen verantwoord is. Voor een deel kan dit liggen aan fouten in de habitatfuncties. Zo bleek de habitatfunctie voor strijklengte aantoonbaar fout te zijn. Daarnaast heeft de discrepantie tussen de resultaten van beide analyses ook te maken met de betekenis van de habitatgeschiktheid als modelresultaat. De geschiktheid geeft een potentie aan en niet de actuele te verwachten bedekking van schedefonteinkruid. Het niet of beperkt voorkomen van schedefonteinkruid op plaatsen met een hoge geschiktheid zegt niet perse dat de gebruikte parameters in het model fout zijn, maar kan ook te maken met onvolledigheid. Er kunnen relevante factoren niet zijn meegenomen in het model, zoals bijvoorbeeld vraat door vogels of zaadverspreiding.

Dit is een argument om de geadviseerde onderverdeling van drie klassen in de habitatgeschiktheid voor schedefonteinkruid te handhaven. Omdat de onzekerheidsanalyse en validatie alleen voor het habitatmodel voor schedefonteinkruid zijn uitgevoerd, is onduidelijk of een onderverdeling in drie habitatgeschiktheidsklassen ook te adviseren is bij andere modellen. Van der Lee et al. (2000) voerden onzekerheidsanalyses uit voor meerdere modellen en dan alleen met onzekerheid in de kennisregels. De resultaten van deze studie gaven aan dat de onzekerheid in de kennisregels van het schedefonteinkruidmodel relatief groot is, en dat een onderscheid in in ieder geval drie geschiktheidsklassen verantwoord lijkt. Het is niet bekend wat de toepassing van de validatiemethode op andere habitatmodellen oplevert. Het is aan te bevelen om de validatiemethode op meerdere modellen toe te passen met een zo goed mogelijke set met invoergegevens.

De validatie met de event detectie methode is alleen geschikt om de functie van één habitatfactor te toetsen, namelijk de factor die beperkend is voor de habitatgeschiktheid. Daar waar de ROC curve significant en in negatieve zin afwijkt van de curve van een random model, is duidelijk dat de relatie tussen deze habitatfactor en geschiktheid niet correct is. Gerichte verbeteringen in het habitatmodel zijn mogelijk voor individuele habitatfuncties in gebieden waar: [ook hier zit weer wat systeemanalyse: validatie is bedoeld om te zeggen dat een model wel/niet voor zijn taak berekend is, kalibratie gebruik je veel meer voor aanpassing aan de model parameters en eventueel –structuur:

1. de invoergegevens voor de habitatfunctie zorgen dat deze beperkend is voor de habitatgeschiktheid.
2. de habitatfunctie een niet-event voorspelt, terwijl wel een event is waargenomen.

Voor de situaties waar een parameter in een habitatfunctie verbeterd kan worden, zou een beperkte calibratie toegepast kunnen worden waarbij het oppervlak van de ROC curve wordt gemaximaliseerd of de richtingscoëfficiënt van de ROC curve wordt geminimaliseerd.

Hierbij kan gebruik gemaakt worden van Monte Carlo-achtige simulaties met verschillende parameterwaarden van de habitatfuncties. Of de gebruikte dataset voldoende informatie bevat om een verbetering van functies mogelijk te maken is af te leiden uit een eerste, snelle neurale netwerk analyse van de dataset.

5.1 Implementatie in het habitatinstrument

De onzekerheidsanalyse met Monte Carlo simulatie en de validatie met de event detectie methode bleken goed bruikbaar te zijn en kunnen daarom in het habitatinstrument geïmplementeerd worden. Op termijn kunnen ook methodes voor een gevoeligheidsanalyse en een calibratie ingebouwd worden, maar deze methodes zijn in dit rapport niet verder uitgewerkt. Implementatie van de methodes voor onzekerheidsanalyse en validatie vergen de volgende activiteiten:

- functioneel en technisch ontwerp;
- bouw van user interface schermen;
- vertaling van door de gebruiker ingevoerde gegevens tot modelscripts;
- inbouwen van Monte Carlo simulatie en gebruikte verdelingen voor de onzekerheid (zie hoofdstuk x) als standaard analyses in PCRaster;
- vormgeven en bouwen van tabellen voor uitvoergegevens;
- inbouwen van presentatiemogelijkheden voor tabelgegevens.

In de onderstaande tekst wordt alleen ingegaan op het ontwerp van de user interface. Gegevens over kostenschattingen zijn te vinden in de notitie over gewenste uitbreidingen van het habitatinstrument (Wesseling et al., 2003).

5.1.1 Onzekerheidsanalyse met Monte Carlo simulatie

De gevolgde methode voor de onzekerheidsanalyse van habitatmodellen in hoofdstuk 3 bleek goed bruikbaar en kan in het habitatinstrument geïmplementeerd worden. Voor het uitvoeren van een onzekerheidsanalyse zijn de volgende gegevens nodig (zie hoofdstuk 3):

- gemiddelde indexwaarden van de breekpunten van habitatfuncties;
- minimum, maximum en standaarddeviatie van de indexwaarden van deze breekpunten;
- gemiddelde waarden van de invoergegevens (zoals diepte, orthofosfaatconcentratie etc.);
- minimum, maximum en standaarddeviatie van deze invoergegevens.

Daarnaast willen gebruikers de keuzemogelijkheid hebben voor de selectie van invoervariabelen en kennisregels waarvoor wel of geen onzekerheid moet worden meegenomen. De onzekerheid in het modelresultaat kan met verschillende parameters gekwantificeerd worden: standaarddeviatie, minimum, maximum en fractielwaarden. Het is wenselijk dat gebruikers kunnen aangeven welke resultaatkaarten ze nodig hebben.

Bij de implementatie van de onzekerheidsanalyse gaat het dus om de volgende functionaliteit:

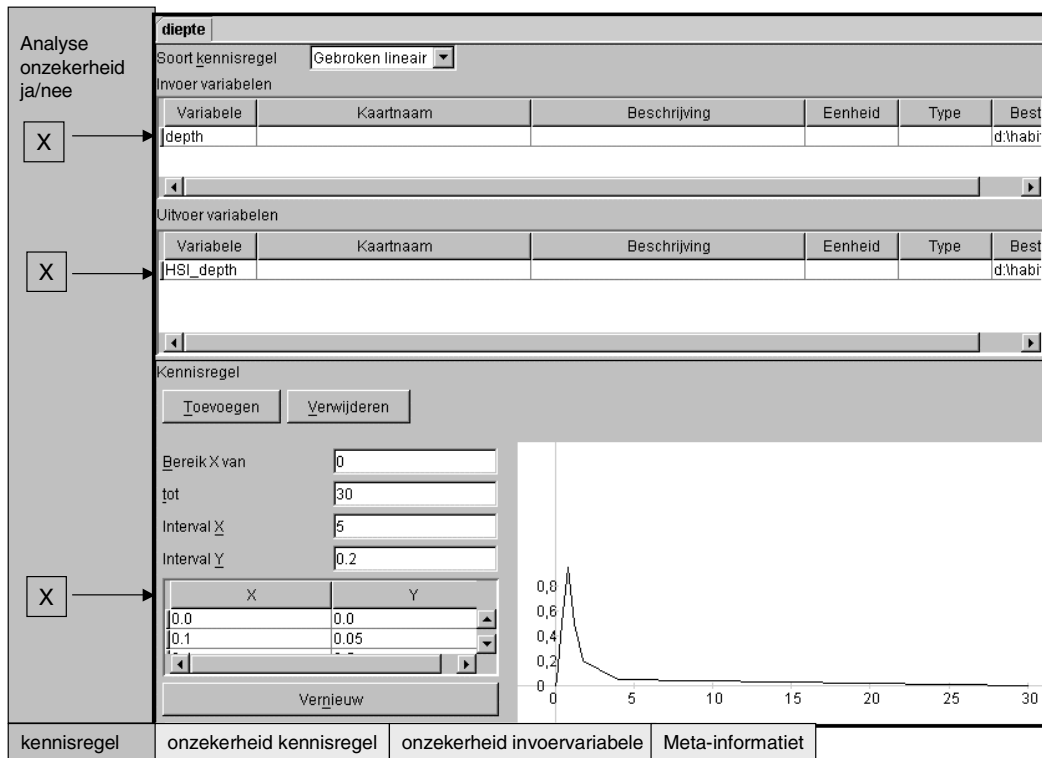
- het kunnen kiezen welke onzekerheden meegenomen moeten worden;
- het opgeven van de benodigde invoergegevens; en
- het opgeven van de gewenste resultaatkaarten.

Het ontwerp van de user interface schermen voor bovenstaande functionaliteit moet gebruikersvriendelijk zijn en aansluiten bij de huidige structuur en vormgeving van het habitatinstrument. Onderstaande figuren geven een eerste schets van het ontwerp voor de benodigde werkschermen voor de onzekerheidsanalyse. NB De figuren betreffen de vormgeving van het ‘werkscherm’ van de user interface. In het habitatinstrument bevindt zich links van het werkscherm een scherm met de structuur van de project- en kennisdatabases en erboven het menu en de knoppenbalk. Een werkscherm kan uit verschillende tabbladen bestaan die door het aanklikken ervan in beeld komen. Voor de onzekerheidsanalyse gaat bij het werkscherm voor een kennisregel om het aanpassen van het tabblad van de kennisregel zelf en om het toevoegen van 2 tabbladen: een voor onzekerheid in de invoervariabele en een voor onzekerheid in de kennisregel zelf. In het werkscherm van het model (combinatiescherm van kennisregels) gaat het om het toevoegen van een tabblad waar de gebruiker de gewenste uitvoergegevens specificeert en opgeeft uit hoeveel runs de Monte Carlo simulatie moet bestaan. Naast de functionele tabbladen is het nuttig om een extra tabblad meta-informatie toe te voegen waarin de methode staat beschreven. Dit tabblad is verder niet uitgewerkt.

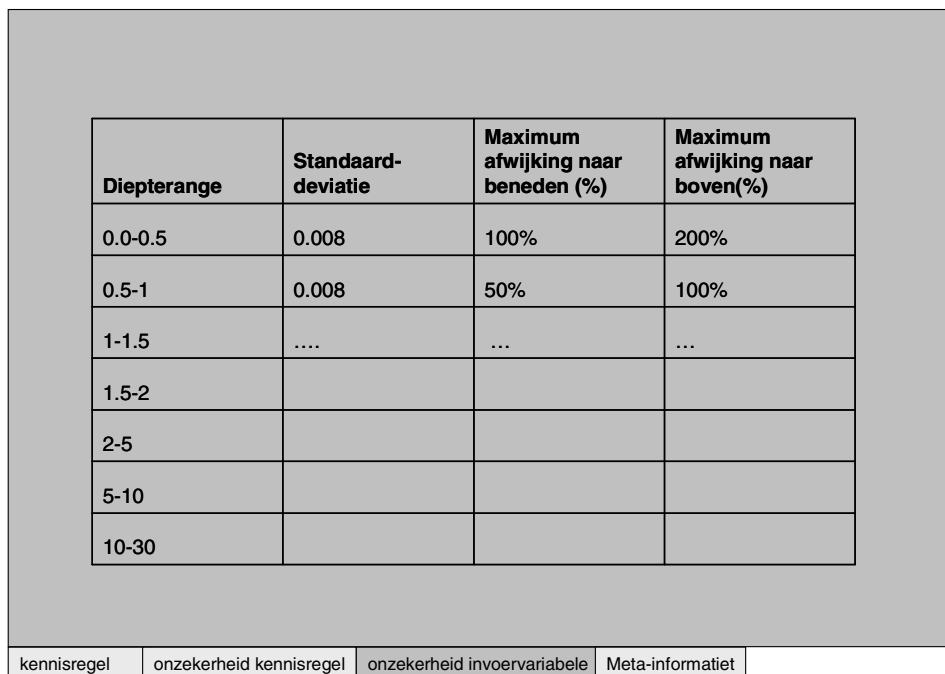
De resultaten van de onzekerheidsanalyse zijn kaarten met voor iedere gridcel de gemiddelde geschiktheid, de standaarddeviatie en andere kentallen. Gebruikers willen deze gegevens vaak aggregeren naar waardes voor hele gebieden. Eén van de gewenste uitbreidingen in het instrument is een extra werkscherm waarin de gebruiker kaartresultaten —bijvoorbeeld van verschillende scenario’s— kan aggregeren tot resultaten voor deelgebieden en uitvoeren in tabelvorm. Een dergelijk werkscherm is hier niet verder uitgewerkt.

Een eerste kostenschatting maakt duidelijk dat de implementatie van de onzekerheidsanalyse een relatief grote financiële inspanning vraagt die in de orde grootte ligt van tientallen procenten van de produktiekosten van het huidige nieuwgebouwde habitatinstrument. De grootste kostenpost is de implementatie van de Monte Carlo simulatie zelf. De kosten voor het toevoegen van flexibiliteit door middel van verschillende keuzemogelijkheden voor de gebruiker zijn in vergelijking daarmee relatief laag.

Een beslissing voor implementatie hangt af van de frequentie waarmee een onzekerheidsanalyse gebruikt zal worden. Bij incidenteel gebruik kan door gebruikers met PCRASTER ervaring ook de standalone methode gebruikt worden die in dit project ontwikkeld is. Het is wel zo dat het habitatinstrument een generiek instrument is waarin allerlei verschillende ecologische modellen gehangen kunnen worden. Indien de onzekerheidsanalyse zo wordt ingebouwd dat deze te gebruiken is voor alle ecologische modellen die in de toekomst in het instrument komen, betekent de implementatie ervan zeker een meerwaarde. Het strekt daarom tot aanbeveling de onzekerheidsanalyse via Monte Carlo simulatie op een generieke manier te implementeren.



Figuur 5.1. Aangepast tabblad in het werkscherm van een kennisregel zodat de gebruiker kan opgeven of onzekerheid moet worden meegenomen.



Figuur 5.2. Tabblad voor de opgave van onzekerheden in de kaart van de invoervariabele van een kennisregel. NB Met de standaarddeviatie voor een diepterange worden de standaarddeviaties bedoeld van alle werkelijke diepten die binnen die range vallen.

NB Omdat de onzekerheid gedefinieerd wordt voor alle voorkomende diepten binnen een range wordt in bovenstaande figuur gesproken over een minimum en maximum als percentage van het gemiddelde. Dit in tegenstelling tot de onderstaande figuur waar het minimum en maximum gedefinieerd zijn per breekpunt van de gebroken lineaire kennisregel.

Diepte	Gemiddelde Index	Standaarddeviatie	Minimum	Maximum
0.0	0.0	0	0	0
0.1	0.05	0.01	0	0.15
0.4	0.5		
0.8	0.95			
1.25	0.5			
1.75	0.2			
4.0	0.05			
30.0	0.0			

kennisregel onzekerheid kennisregel onzekerheid invoervariabele Meta-informatie

Figuur 5.3. Tabblad voor de opgave van onzekerheden in de kennisregel zelf (onzekerheid in de expertkennis).

Aantal runs Monte Carlo simulatie	300																																				
Uitvoer op puntlocaties																																					
	<table border="1"> <thead> <tr> <th>Kaartnaam</th> <th>Beschrijving</th> <th>Bestand</th> </tr> </thead> <tbody> <tr> <td>puntenkaart</td> <td>meetlocaties IJss</td> <td>c:\maps\puntenIJss.bil</td> </tr> <tr> <td></td> <td>Bestand</td> <td></td> </tr> <tr> <td>Uitvoertabel</td> <td>c:\tabellen\punt.tbl</td> <td></td> </tr> </tbody> </table>	Kaartnaam	Beschrijving	Bestand	puntenkaart	meetlocaties IJss	c:\maps\puntenIJss.bil		Bestand		Uitvoertabel	c:\tabellen\punt.tbl																									
Kaartnaam	Beschrijving	Bestand																																			
puntenkaart	meetlocaties IJss	c:\maps\puntenIJss.bil																																			
	Bestand																																				
Uitvoertabel	c:\tabellen\punt.tbl																																				
Uitvoerkaarten studiegebied																																					
<table border="1"> <thead> <tr> <th>Variabele</th> <th>Gemiddelde</th> <th>Standaarddeviatie</th> <th>Mediaan</th> <th>25% en 75% fractiel</th> <th>Minimum/Maximum</th> </tr> </thead> <tbody> <tr> <td>HSI_totaal</td> <td>c:\maps\hsitot_avg.bil</td> <td>c:\maps\...</td> <td></td> <td></td> <td>c:\maps\...</td> </tr> <tr> <td>HSI_regel1</td> <td>c:\maps\...</td> <td>c:\maps\...</td> <td></td> <td></td> <td>c:\maps\...</td> </tr> <tr> <td>HSI_regel2</td> <td>c:\maps\...</td> <td>c:\maps\...</td> <td></td> <td></td> <td>c:\maps\...</td> </tr> <tr> <td>HSI_regel3</td> <td>c:\maps\...</td> <td>c:\maps\...</td> <td></td> <td></td> <td>c:\maps\...</td> </tr> <tr> <td>HSI_regel4</td> <td>c:\maps\...</td> <td>c:\maps\...</td> <td></td> <td></td> <td>c:\maps\...</td> </tr> </tbody> </table>	Variabele	Gemiddelde	Standaarddeviatie	Mediaan	25% en 75% fractiel	Minimum/Maximum	HSI_totaal	c:\maps\hsitot_avg.bil	c:\maps\...			c:\maps\...	HSI_regel1	c:\maps\...	c:\maps\...			c:\maps\...	HSI_regel2	c:\maps\...	c:\maps\...			c:\maps\...	HSI_regel3	c:\maps\...	c:\maps\...			c:\maps\...	HSI_regel4	c:\maps\...	c:\maps\...			c:\maps\...	
Variabele	Gemiddelde	Standaarddeviatie	Mediaan	25% en 75% fractiel	Minimum/Maximum																																
HSI_totaal	c:\maps\hsitot_avg.bil	c:\maps\...			c:\maps\...																																
HSI_regel1	c:\maps\...	c:\maps\...			c:\maps\...																																
HSI_regel2	c:\maps\...	c:\maps\...			c:\maps\...																																
HSI_regel3	c:\maps\...	c:\maps\...			c:\maps\...																																
HSI_regel4	c:\maps\...	c:\maps\...			c:\maps\...																																
Meta-informatie model	model	Meta-informatie onzekerheidsanalyse	onzekerheid resultaat																																		

Figuur 5.4. Tabblad voor de opgave van het aantal runs in de Monte Carlo simulatie en de uitvoergegevens.

Onzekerheid kan berekend worden voor de indices van de individuele kennisregels en van de geschiktheid als geheel. Desgewenst kan de gebruiker een kaart met puntlocaties opgeven en de bestandsnaam van een uitvoertabel. Deze uitvoertabel geeft dan een overzicht van de onzekerheidsparameters op de puntlocaties.

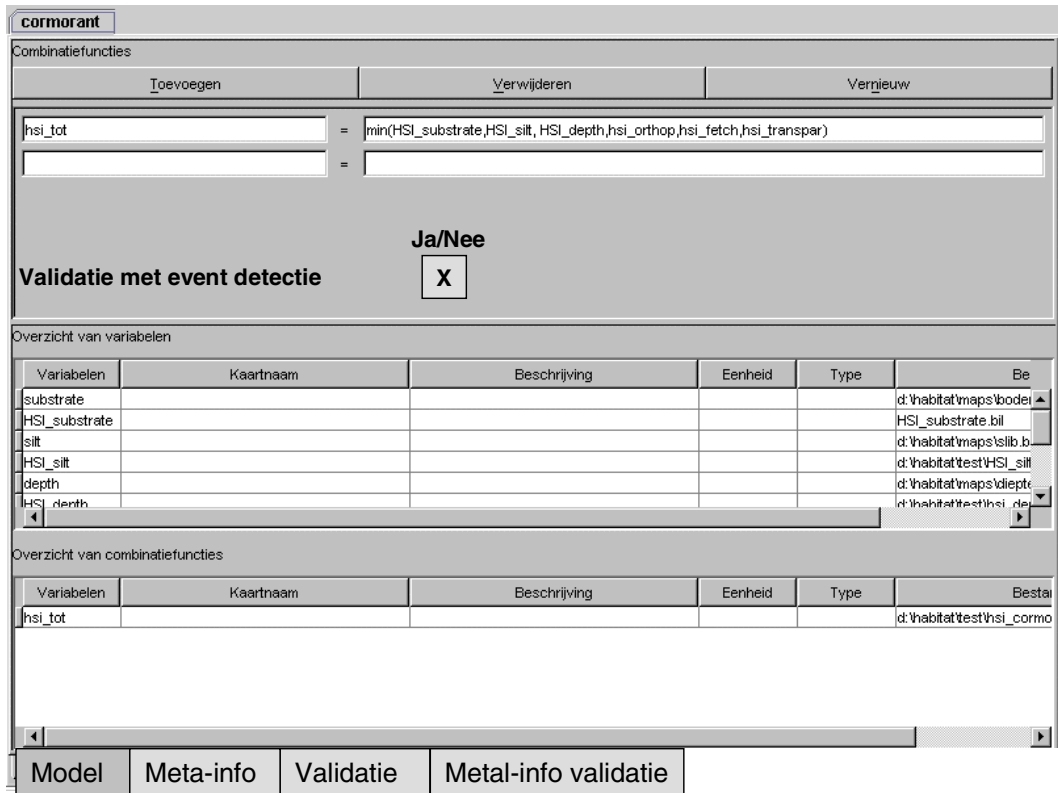
5.1.2 Validatie met event detectie methode

De validatie methode met event detectie bleek een bruikbare methode voor het testen van het voorspellend vermogen van een habitatmodel. Het voorspellend vermogen kan getest worden aan de hand van verschillende criteria. Uit de analyses bleken de criteria ROC, sensitivity, specificity en accuracy het meest inzichtelijk. Bij de implementatie van de methode zouden deze criteria dus standaard uitgevoerd kunnen worden.

De benodigde invoergegevens voor het uitvoeren van de validatiemethode zijn: een kaart met waarnemingen van een soort om de overeenkomst met de geschiktheid mee te toetsen.

de drempelwaarde in de waarnemingen (aantallen/dichtheid) waarboven sprake is van een event (de soort komt voor)

Net als bij de onzekerheidsanalyse moet de implementatie van de validatiemethode aansluiten bij de structuur en vormgeving van het habitatinstrument en de toepassing ervan moet gebruikersvriendelijk zijn. Voorgesteld wordt om in het werkscherm van een model (combinatiefuncties van kennisregels) de gebruiker de keuze te laten maken om wel of geen validatie uit te voeren. Onderstaande figuren geven een eerste schets van het ontwerp van de tabbladen in het werkscherm van een model: het eerste tabblad is een aanpassing van het huidige modelscherm waar de gebruiker de keuze maakt voor het uitvoeren van een validatie; het tweede is een extra tabblad voor het opgeven van de in- en uitvoergegevens. Voor elke testvariabele (geschiktheid en indices van kennisregels) worden de curves van de criteria sensitivity, specificity, accuracy en ROC standaard weggeschreven in een tabel. Met deze tabel kunnen grafieken gemaakt worden van de curves, in de tabel staan verder de AUC (area under the curve) van de ROC curves. Naast de functionele tabbladen is het nuttig om een extra tabblad met meta-informatie op te nemen met een beschrijving van de methode. Dit tabblad is hier verder niet uitgewerkt.



Figuur 5.5 Aangepast tabblad van het werkscherm van een model zodat de gebruiker de keuze krijgt om een validatie uit te voeren. NB De voorbeeld functie in de figuur is de berekening van de habitatgeschiktheid (hsi_tot) als minimum van de indices van de kennisregels van het model (hsi_substrate, hsi_depth etc.). Deze kennisregels zijn gedefinieerd in andere werkschermen die hier niet zijn getoond.

Invoervariabelen				
	Kaartnaam	Beschrijving	Eenheid	Bestand
Waarnemingen soort	schedef	bedekkingsgraad	klassen	c:\maps\sf.bil

	Waarde	Beschrijving
Drempelwaarde voor waargenomen event	3	klasse 10-25%

Uitvoervariabelen	
Variabele	Test
HSI_totaal	X
HSI_regel1	X
HSI_regel2	X
HSI_regel3	X
HSI_regel4	X
Bestand	
Uitvoertabel	c:\uitvoer\roc_sf.tbl

Model	Meta-info	Validatie	Meta-info validatie
-------	-----------	-----------	---------------------

Figuur 5.6. Extra tabblad in het werkscherm van een model waar de gebruiker de in- en uitvoergegevens voor de validatie op kan geven.

6 Referenties

Aarts, H.P.A., G. van der Lee & D.T. van der Molen, 2001. Toepassing MORRES op het Volkerakmeer.

Brinkman, R. 2003. Technisch ontwerp modelinstrument Habitat. Intern document bij project Q3266.

Duel, H., R. Pouwels, M. de Vries, 2001. Visie op het modelinstrumentarium voor het uitvoeren van habitatevaluaties van de Nederlandse rijkswateren. WL, Alterra, Delft.

Heeger, D., 2002: Signal Detection Theory Handout (Psych 30), <http://white.stanford.edu/~heeger/sdt/sdt.html>.

Laane, W.E.M., 1996. Habitatmodellen waterplanten Gele Plomp, Rivierfonteinkruid, schedefonteinkruid en watergentiaan. Werkdocument 96.049X. TNO-BSA Delft.

Lee, G. van der, H. Duel, S. Groot, H. Aarts en R. Pouwels, 2000. Kwaliteit van het HEP-instrumentarium voor toepassing in het IJsselmeergebied. WL rapport T2391.

Lee, G.E.M. van der, 2003. Definitiestudie naar vernieuwing van het bestaande WL-habitatinstrument.

MedCalc, 2002. ROC curves and comparison of ROC curves in MedCalc. <http://www.medcalc.be/roccurves.html>.

Wesseling, C., G. van der Lee, R. Brinkman, 2003. Raming van werkzaamheden van PCRaster Environmental Software (PES) en WL ten behoeve van gewenste uitbreidingen van het Habitat Instrument.

A Overzicht habitatmodellen

Tabel A.1 Overzicht van HIS-modellen die aanwezig zijn bij WL.

Soort	Bron		MORRES	WL doelsubsidie 2002	DWW	OVB
		TNO- WSV				
aal						x
aalscholver		x	x	x		
baardmannetje					x	
baars			x		x	
barbeel			x			
beekforel						x
bever			x	x	x	
biezen			x			
blankvoorn			x			x
blauwborst					x	
boomkikker			x			
bosrietzanger					x	
brasem			x	x		x
bruin cypergras			x			
das			x			
driehoeksmossel		x	x	x		
driekantige bies		x	x			
engelse alant			x			
essen-iepenbos		x				
fint			x			x
fonteinkruiden			x			
fuut		x	x	x		
gele plomp (T)		x				
grauwe gans					x	
helofyten			x			
ijsvogel			x	x	x	
kamsalamander			x	x		
karper						x
kleine karekiet					x	
kleine plevier					x	
kleine tanglibel			x			
kleine zwaan		x				
kokerjuffer		x				
krabbescheer			x			
kroeskarper				x		
kuifeend		x	x	x		
kwak		x	x			
kwartelkoning			x			
lepelaar		x				
maasraket			x			
mattenbies		x				
mosselwants			x			
muskusrat					x	

nonnetje		x				
noordse woelmuis					x	
oeveraas		x				
oeverzwaluw		x	x			
otter		x	x	x		
pijlkruid			x			
rietgors			x		x	
rietzanger					x	
ringslang			x			
rivierdonderpad				x		x
rivierfonteinkruid	x		x			
riviergrondel						x
rivierprik						x
rivierrombout			x			
roerdomp			x	x	x	
rugstreppad			x			
schedefonteinkruid	x					
serpeling				x		x
slobeend					x	
snoek			x	x		
snoekbaars			x	x		x
spiering						x
veldsalie			x			
visdief			x		x	
vlottende watteranonkel			x			
watergentiaan	x		x			
waterhoen					x	
waterplanten			x			
waterral			x		x	
watersnip					x	
watervleermuis					x	
wilde kruisdistel			x			
wilgen-populierenbossen	x					
winde			x			x
woelrat					x	
zandoeverdansmug			x			
zeeforel						x
zoetwaterneriet			x			
zomereik			x			
zomertaling					x	
zwarte els			x			
zwarte populier			x			
zwarte stern					x	

- DWW: beknopte HSI modellen met vaak functies van klassenverdelingen of aan/uit functies.
- TNO-WSV: uitgebreide HSI modellen met veel optimumfuncties/uitgebreide klassenverdelingen.
- OVB: HSI modellen: hetzij uitgebreide literatuurstudies, hetzij alleen een overzicht van HSI functies.
- WL-MORRES: habitatbeschikbaarheidsmodellen (enkele habitatfactoren), functies ongedocumenteerd.

- WL-doelsubsidie2002: habitatbeschikbaarheidsmodellen door middel van beknopte literatuurreview.

Tabel A.2 Overzicht van gebruikte habitatfactoren (aangegeven met 1) in HSI-modellen die gemaakt zijn binnen MORRES

Soorten	Habitatfactoren						
	getijde- dynamiek	stroom- snelheid	inundatie -duur	water- diepte	grondwater- stand	rivier- dynamiek	scheepvaart
aalscholver (nest)			1	1	1		
aalscholver (voedsel)			1	1	1		
baars (paai)		1	1	1	1		1
baars (voedsel)		1	1	1	1		
barbeel (paai)		1	1	1	1		
barbeel (voedsel)		1	1	1	1		
biezen	1	1	1	1	1	1	1
blankvoorn (paai)		1	1	1	1		1
blankvoorn (voedsel)		1	1	1	1		
boomkikker (land)			1	1	1	1	1
boomkikker (water)		1	1	1	1		
brasem (paai)		1	1	1	1		1
brasem (voedsel)		1	1	1	1	1	1
bruin cypergras			1	1	1		
das (nest)			1	1	1		
das (voedsel)			1	1	1		
driehoeksmossel		1	1	1	1		
driekantige bies	1	1	1	1	1	1	1
engelse alant			1	1	1		
fint (paai)	1	1	1	1	1		
fint (voedsel)			1	1	1		
fonteinkruiden	1	1	1	1	1	1	1
fuut (nest)	1	1	1	1	1	1	
fuut (voedsel)		1	1	1	1		
helofyten		1	1	1	1		1
ijsvogel					1	1	1
kamsalamander (water)	1	1	1	1	1		
kleine tanglibel		1	1	1	1		
krabbescheer	1	1	1	1	1		
kuifeend (voedsel)		1	1	1	1	1	1
kwak (nest)			1	1	1		
kwak (voedsel)			1	1	1		
kwartelkoning			1	1	1		
maasraket			1	1	1		
mosselwants		1	1	1	1		
oeverzwaluw				1			
otter (nest)	1		1	1			
otter (voedsel)			1	1	1		
pijlkruid	1	1	1	1	1		1
rietgors			1	1	1		
ringslang (nest)			1	1	1		
ringslang (voedsel)		1	1	1	1		1
rivierfonteinkruid	1	1	1	1	1	1	1
rivierrombout		1	1	1	1	1	1
roerdomp	1		1	1	1		
rugstreppad (land)			1	1	1		
rugstreppad (water)		1	1	1	1		

snoek (paai)		1	1	1	1		1
snoek (voedsel)		1	1	1	1		
snoekbaars (paai)			1				1
snoekbaars (voedsel)				1	1		
veldsalie			1	1	1		
visdief (nest)				1	1		
visdief (voedsel)		1	1	1	1		
vlottende waterranonkel	1	1	1	1	1	1	1
watergentiaan	1	1	1	1	1	1	1
waterplanten	1	1	1	1	1	1	1
waterral	1			1	1		
wilde kruisdistel			1	1	1		
winde (paai)		1	1	1	1		1
winde (voedsel)		1	1	1	1		
zandoeverdansmug		1	1	1	1		
zoetwaterneriet		1	1	1	1		
zomereik			1	1	1	1	1
zwarte els			1	1	1		
zwarte populier			1	1	1		

Tabel A.3 Overzicht van habitatfactoren in HSI-modellen die gemaakt zijn in het kader van het doelsubsidieprogramma Onzekerheden in habitatmodellen (Van der Lee et al.), 2002).

Soorten	Habitatfactoren								
	watertemperatuur (°C)	waterdiepte (cm)	stroom-snelheid (cm/s)	zuurstof-concentratie (%)	minimum areaal	waterplanten-bedekking (%)	doorzicht	oeverbreedte	minimale breedte water
Brasem									
paaihabitat	1	1	1	1			1		
voedselhabitat	1	1	1	1			1		
Driehoeksmossel	1	1	1	1	1		1		
Kroeskarper									
paaihabitat	1	1	1	1			1		
voedselhabitat	1	1	1	1			1		
Rivierdonderpad	1	1	1	1			1		
Serpeling									
paaihabitat	1	1	1	1			1		
voedselhabitat	1	1	1	1			1		
Snoek									
paaihabitat	1	1	1	1			1		
voedselhabitat	1	1	1	1			1		
Snoekbaars									
paaihabitat	1	1	1	1			1		
Snoekbaars									
voedselhabitat	1	1	1	1			1		
IJsvogel			1			1	1		
Fuut		1	1			1	1	1	
Kuifeend		1	1			1	1	1	
Roerdomp		1							1
nesthabitat		1							1
Aalscholver		1	1			1	1		
Bever		1				1			
Otter		1				1	1	1	1

Kamsalamander

1

Tabel A.4 Overzicht van habitatfactoren in HSI-modellen van vogelsoorten

Soort	Bron/bijzonderheden	Functie habitat	Habitatfactoren
Aalscholver	TNO-WSV	nesthabitat voedselhabitat	verstoringafstand, type vegetatie gem. waterdiepte, doorzicht, trofie, verstoringafstand, afstand voedselgebied, PCB in voedsel,
Baardmannetje	DWW	nesthabitat voedselgebied	peilfluctuatie, overstromingen, type vegetatie, ondiep water, aantal prooien
Blauwborst	DWW	broedgebied nesthabitat voedselgebied	oppervlakte, saliniteit, type vegetatie, breedte oevervegetatie, taludvorm, overstromingen, grondwater, taludvorm, vegetatiebedekking, afstand voedselgebied,
Bosrietzanger	DWW	leefgebied nesthabitat broedgebied	Breedte, oppervlakte, type vegetatie, overstromingen, grondwater, vegetatiebedekking, vegetatiehoogte, taludvorm, oppervlakte, % oeverbeschutting (overhangend), type vegetatie, breedte oevervegetatie,
Fuut	TNO-WSV	nesthabitat voedselhabitat	peilfluctuatie, verstoringafstand, type vegetatie, breedte oevervegetatie, gem. waterdiepte, saliniteit, doorzicht, trofie, PCB in voedsel, kwikgehalte,
Grauwe gans	DWW	nesthabitat voedselhabitat leefgebied	peilfluctuatie, overstromingen, verstoringafstand, type vegetatie, breedte oevervegetatie, taludvorm, oppervlakte, afstand voedselgebied, oppervlakte,
IJsvogel	DWW	nesthabitat voedselgebied broedgebied	substraattype, lengte oever, taludvorm, breedte, oppervlakte, gem. waterdiepte, doorzicht, % oeverbeschutting (overhangend), peilfluctuatie, overstromingen, afstand voedselgebied,
Kleine karekiet	TNO-WSV	nesthabitat	breedte, oppervlakte, peilfluctuatie, beheer, type vegetatie, vegetatiebedekking, vochtigheidstoestand,
Kleine plevier	DWW	nesthabitat voedselgebied broedgebied	overstromingen, grondwater, verstoringafstand, vegetatiebedekking, aantal prooien oppervlakte, saliniteit, ondiep water, % oeverbeschutting (overhangend), lengte oever ongeschikt, taludvorm,
Kleine zwaan	TNO-WSV	voedselgebied	gem. waterdiepte, waterplantenbedekking, peildynamiek, verstoringafstand, tuberbiomassa, bovengrondse biomassa schedefonteinkruid,
Kuifeend	TNO-WSV	broedhabitat-nest	type vegetatie, breedte oevervegetatie, lengte oevervegetatie,

		broedhabitat-voedsel winterhabitat	gem. waterdiepte, waterplantenbedekking, trofie, PCB in voedsel, kwikgehalte, DDE gehalte, HCB gehalte, Cadmium gehalte, gem. waterdiepte, driehoeksmossel dichtheid, PCB in voedsel, kwikgehalte, DDE gehalte, HCB gehalte, Cadmium gehalte,
Kwak	TNO-WSV	nesthabitat voedselhabitat	gem. waterdiepte, verstoringsafstand, type vegetatie, verstoringsafstand, PCB in voedsel, DDE gehalte,
Lepelaar	TNO-WSV	nesthabitat	gem. waterdiepte, peilfluctuatie, verstoringsafstand, areaal moeras, type vegetatie,
Nonnetje	TNO-WSV	voedselhabitat	gem. waterdiepte, doorzicht, trofie, windwerking, verstoringsafstand, PCB in voedsel,
Oeverzwaluw	TNO-WSV	nesthabitat	breedte, afstand voedselgebied, expositie, aanwezigheid open ruimte,
Rietgors	DWW	nesthabitat	overstromingen, grondwater, vegetatiehoogte,
Rietzanger	DWW	broedgebied nesthabitat	oppervlakte, type vegetatie, vegetatiebedekking, breedte oevervegetatie, taludvorm, breedte, oppervlakte, beheer, overstromingen, grondwater, type vegetatie, vegetatiebedekking, vochtigheidstoestand,
Roerdomp	DWW	nesthabitat voedselhabitat	breedte, peilfluctuatie, overstromingen, vegetatiebedekking, type vegetatie, aanwezigheid open water,
Slobeend	DWW	leefgebied nesthabitat	oppervlakte, gem. waterdiepte, vegetatiebedekking, taludvorm, beheer, overstromingen, dichtheid grote grazers, grondwater, breedte oevervegetatie, vegetatiehoogte, taludvorm,
Smient	DWW	leefgebied	oppervlakte, saliniteit, ondiep water, stroomsnelheid, type vegetatie,
Visdief	DWW	nesthabitat	oppervlakte, substraattypen, overstromingen, isolatie, verstoringsafstand, type vegetatie, vegetatiebedekking, breedte oevervegetatie, vegetatiehoogte, taludvorm,
Waterhoen	DWW	voedselgebied nesthabitat leefgebied	oppervlakte, vervuilingsgraad, peilfluctuatie, overstromingen, oppervlakte, gem. waterdiepte, saliniteit, ondiep water, % oeverbeschutting (overhangend), type vegetatie, breedte oevervegetatie, lengte oevervegetatie,
Waterral	DWW	nesthabitat leefgebied	gem. waterdiepte, peilfluctuatie, overstromingen, oppervlakte, ondiep water, stroomsnelheid, type vegetatie, vegetatiebedekking, breedte oevervegetatie, lengte oevervegetatie, taludvorm, micro reliëf in moeras,

Watersnip	DWW	nesthabitat voedselgebied leefgebied	beheer, overstromingen, dichtheid grote grazers, vegetatiehoogte, oppervlakte, substraatype, % zachte bodem, afstand voedselgebied, vochtigheidstoestand, oppervlakte, grondwater, type vegetatie, taludvorm,
Zomertaling	DWW	nesthabitat leefgebied	beheer, overstromingen, dichtheid grote grazers, grondwater, breedte oeervegetatie, vegetatiehoogte, taludvorm, oppervlakte, ondiep water, waterplantenbedekking, stroomsnelheid, type vegetatie,
Zwarte stern	DWW	nesthabitat voedselgebied	waterplantenbedekking, peilfluctuatie, overstromingen, type vegetatie, afstand voedselgebied, vegetatiehoogte, breedte, oppervlakte, aanwezigheid open ruimte

Tabel A.5 Overzicht van habitatfactoren in HSI-modellen van vissoorten

Soort	Bron/bijzonderheden	Specifieke functie habitat	Habitatfactoren
aal	OVB	waterkwaliteit beschutting	gem. watertemperatuur, pH, zuurstofgehalte, % bodembedekking (planten/stenen/takken), % zachte bodem,
baars	DWW	paaihabitat voedselhabitat	% paaisubstraat, waterplantenbedekking, peilfluctuatie, gem. waterdiepte, gem. watertemperatuur, saliniteit, % litoraal, stroomsnelheid, zuurstofgehalte, trofie, vervuilingsgraad,
beekforel	OVB	zomerhabitat opgroeihabitat paaihabitat	% opp. stroomkommen, % oeverbeschutting (overhangend), ratio gem. jaarlijkse peikafvoer/gem. dagelijkse piekafvoer, % opp. stroomkommen, grootte substraat, % bodembedekking (planten/stenen/takken), % slib/zand in kiezelsubstraat, % oppervlak met kiezel 1-7 cm, % ronde kiezel in kiezelfractie, dikte kiezellaag, max. watertemperatuur, % opp. stroomkommen, % oeverbeschutting (overhangend), % stabiele oever (begroeid/stenen bij gemiddelde jaarlijkse afvoer),
blankvoorn	OVB	winterhabitat paaihabitat voedselhabitat waterkwaliteit	% opp. stroomkommen, % oeverbeschutting (overhangend), grootte substraat, % paaisubstraat, gem. waterdiepte, gem. watertemperatuur, pH, saliniteit, % litoraal, waterplantenbedekking, doorzicht, stroomsnelheid, zuurstofgehalte, gem. watertemperatuur, max. watertemperatuur, pH, saliniteit, zuurstofgehalte,
brasem	OVB	voortplantingshabitat voedsel/beschutting waterkwaliteit	% paaisubstraat, gem. Waterdiepte, gem. watertemperatuur pH, saliniteit, stroomsnelheid, zuurstofgehalte, waterplantenbedekking, doorzicht, stroomsnelheid, substraatype, gem. waterdiepte, pH, saliniteit, zuurstofgehalte, trofie,
fint	OVB	beschutting en voedselhabitat	max. watertemperatuur, stroomsnelheid, zuurstofgehalte, substraatype, getijdeninvloed,
karper	OVB	waterkwaliteit paaihabitat (max. temp voor ei/larve en max. temp paai)	gem. watertemperatuur, pH, saliniteit, zuurstofgehalte, gem. waterdiepte, max. watertemperatuur, zuurstofgehalte,
riviergrondel	OVB	voortplantingshabitat groeihabitat	waterplantenbedekking, doorzicht, substraatype, gem. waterdiepte, gem. watertemperatuur. saliniteit, waterplantenbedekking, stroomsnelheid, zuurstofgehalte, bodemstructuur,

rivierdonderpad	OVB		gem. waterdiepte, max. watertemperatuur, stroomsnelheid, zuurstofgehalte, type vegetatie, aanwezigheid open ruimte,
rivierprik	OVB		gem. waterdiepte, zuurstofgehalte, substraattipe, % oppervlak met kiezel 1-7 cm,
serpeling	OVB		max. watertemperatuur, waterplantenbedekking, zuurstofgehalte, substraattipe, % oppervlak met kiezel 1-7 cm,
snoekbaars	OVB	waterkwaliteit	gem. watertemperatuur, pH, saliniteit, stroomsnelheid, zuurstofgehalte,
		beschutting	gem. waterdiepte, doorzicht, % bodembedekking (planten/stenen/takken),
		voortplantingshabitat	% paaisubstraat, gem. watertemperatuur, substraattipe,
spiering	OVB		gem. watertemperatuur, max. watertemperatuur, stroomsnelheid, zuurstofgehalte, substraattipe,
winde	OVB	voortplantingshabitat	% paaisubstraat, zuurstofgehalte, % oeverbeschutting (overhangend),
zeeforel	OVB		gem. watertemperatuur, stroomsnelheid, zuurstofgehalte,

Tabel A.6 Overzicht van habitatfactoren in HSI-modellen van macrofaunasoorten, zoogdieren en vegetaties.

Macrofauna	Bron/ bijzonderheden	Functie habitat	Habitatfactoren
Driehoeksmossel	TNO-WSV, onderscheid model stagnante wateren en stromende wateren		breedte, gem. waterdiepte, gem. temperatuur pH, saliniteit, stroomsnelheid, zuurstofgehalte, trofie, calcium,
Kokerjuffer	TNO-WSV	Substraat (eieren)	gem. waterdiepte, grootte substraat,
Oeveraas	TNO-WSV, vnl. historische literatuur en literatuur Oost-Europa	Water Substraat (larven)	gem. temperatuur, stroomsnelheid, zuurstofgehalte, gem. waterdiepte, substraattypen,
Zoogdieren		Water	gem. temperatuur, stroomsnelheid, zuurstofgehalte,
Otter	TNO-WSV, OVB, literatuur	Nest	oppervlakte, breedte, afstand tussen rustplaatsen, beheer
		Voedsel	breedte, doorzicht, gem. diepte, trofie, waterplantenbedekking, PCB in voedsel, kwik in voedsel, dieldrin in voedsel
		Dekking	taludvorm, oeverbeschutting, breedte oevervegetatie, lengte oever ongeschikt, beheer
		Verstoring	versnippering door infrastructuur, verstoringafstand
Watervleermuis	DWW		breedte, oppervlakte, waterplantenbedekking, trofie, oeverbeschutting, afstand voedselgebied, afstand lijnelement- water, verlicht oppervlak, opp. effectief jachtgebied,
Bever	DWW		gem. waterdiepte, vegetatiebedekking, afstand voedselgebied, lengte oevervegetatie, lengte oever ongeschikt,
Noordse woelmuis	DWW		peilfluctuatie, beheer, overstromingen, dichtheid grote grazers, type vegetatie, vegetatiebedekking, breedte oevervegetatie,

Woelrat	DWW	gem. waterdiepte, peilfluctuatie, vegetatiebedekking, breedte oevervegetatie, taludvorm,
Muskusrat	DWW	breedte, substraatype, beheer, lengte oever ongeschikt, vegetatiehoogte,
Vegetatie		
Gele plomp	TNO-WSV	gem. waterdiepte pH, saliniteit, stroomsnelheid, trofie, substraatype, peildynamiek, windwerking,
Schedefonteinkruid	TNO-WSV	gem. waterdiepte, gem. temperatuur pH, saliniteit, doorzicht, stroomsnelheid, trofie, substraatype, % slib/zand in kiezelsubstraat, windwerking, bicarbonaat,
Rivierfonteinkruid	TNO-WSV	gem. waterdiepte saliniteit, doorzicht, stroomsnelheid, substraatype, hydrodynamiek,
Wattergentiaan	TNO-WSV	gem. waterdiepte, pH, saliniteit, doorzicht, stroomsnelheid, trofie, substraatype, peildynamiek, windwerking, calcium,
Driekantige bies Mattenbies	TNO-WSV TNO-WSV, verschillende submodellen voor ondersoorten	gem. waterdiepte, saliniteit, getijdeninvloed, % slib/zand in kiezelsubstraat, gem. waterdiepte, pH, saliniteit, stroomsnelheid, trofie, % slib/zand in kiezelsubstraat, peildynamiek, windwerking, peilfluctuatie, beheer,
Wilgen- populierenbossen	TNO-WSV, type bos hangt af van watersysteem	substraatype, overstromingen, dichtheid grote grazers, relatieve hoogteligging, isolatie,
Essen-iepenbos	TNO-WSV, verschillende submodellen voor bostypen	substraatype, overstromingen, dichtheid grote grazers, grondwater,

B Algemene habitatcriteria voor verschillende soortgroepen en uitwerking tot optimumfuncties

B.1 Werkwijze

Evaluatie van soortcriteria

De evaluatie van soortcriteria is in de volgende stappen uitgevoerd:

1. De selectie van relevante voorbeeld soorten.
2. De selectie van relevante factoren die habitats beschrijven en het aangeven van ecologisch relevante klassen.
3. De (her-)definitie van optimumfuncties voor deze habitatfactoren per individuele soort.
4. De vergelijking met ecotoopcriteria van een bestaande ecotopenindeling (RWES).

De geselecteerde soorten zijn soorten van de zoete rijkswateren. De soorten zijn geselecteerd via soortgroepen die in sterke mate afhankelijk zijn van eenzelfde soort biotoop. Voor veel van de geselecteerde soorten zijn al habitatmodellen opgesteld die specifiek ingaan op de habitateisen, en de belangrijkste factoren zijn hieruit geselecteerd en geregistreerd. Daarnaast is er met behulp van literatuuronderzoek extra informatie toegevoegd.

Het vóórkomen van soorten uit een bepaalde soortgroep is gerelateerd aan een bepaalde waarde van de factoren. Wanneer kan worden aangegeven binnen welke range zo'n soortgroep kan voorkomen, kan worden gesproken van een ecologisch relevante klasse. Er is hierbij onderscheid gemaakt tussen klassen voor reproductie- en fourageerhabitat.

B.2 Evaluatie van soortcriteria

B.2.1 Selectie van soorten

Gedurende de afgelopen decennia zijn in binnen- en buitenland vele habitatmodellen opgesteld. De hoeveelheid soorten die hierbij is beschouwd is groot en voor dit project is een selectie gemaakt van soorten gebaseerd op karakteristieke soortgroepen. Hierbij zijn faunagroepen geselecteerd die zo divers mogelijk gebruik maken van landschapselementen. Als opstappunt is gebruik gemaakt van de AMOEBES IJsselmeergebied (Vanhemelrijk 1993) waarin soorten gerangschikt zijn naar biotoopgebruik (Tabel B.1). Per soortgroep zijn meerdere soorten gekozen. Deze soorten zijn als voorbeeld gebruikt in deze studie.

Tabel B.1 Soortselectie via biotoopgebruik (naar Vanhemelrijk 1993) (n = 39)

Biotooptype	Soortgroep fauna	Geselecteerde soort
open water zone	visetende vogels (zichtpredatie) open-watervis, stagnerend water open-watervis, stromend water bodemfauna en hun predators (duikeenden)	fuut, ijsvogel, grote zaagbek brasem, kroeskarper, barbeel, grote modderkruiper rivierdonderpad, serpeling, rivierprik, roofblei, blankvoorn, kwabaal driehoeksmossel, kuifeend
ondiepe oeverzone	waterplantengebonden vis	snoek, snoekbaars, kleine zwaan
natte oever	visetende zoogdieren slikfouragerende vogels pioniervegetatie broedende vogels	otter grutto, lepelaar kluut, oeverzwaluw kluut
aanverwante (natte) terrestrische systemen	rietzangvogels moerasbroedvogels graslandbroedvogels bosbroedvogels, 'waterminnende zoogdieren' 'waterminnende reptielen' amfibieën	grote karekiet, rietzanger roerdomp, kwak kemphaan aalscholver bever, noordse woelmuis ringslang rugstreeppad, kamsalamander, boomkikker

Deze voorbeeldset aan soorten is gebruikt om de twee verschillende methoden uit te werken en te vergelijken. In het uiteindelijke HEP-instrumentarium zal het mogelijk zijn om nieuwe soorten toe te voegen, en waar noodzakelijk informatie aan of in te passen. De eisen die deze soorten stellen aan bepaalde onderdelen van hun habitat zijn vermeld in Tabel B8-11.

Sommige soorten maken gebruik van verschillende habitats voor functies zoals fourageren en reproductie. Ook zijn er soorten met een seizoensgebonden voorkeur voor een bepaald type biotoop en kunnen soorten zeer kritisch zijn ten aanzien van de inrichting en kwaliteit van een gebied (bijvoorbeeld areaalgrootte, verstoring, mate van eutrofiëring). Dit heeft effect op de aanwezige gegevens per soort. Van alle geselecteerde soorten is getracht een zo volledig mogelijk beeld te beschrijven van de eisen ten aanzien van afzonderlijke habitatfactoren. Enkele soorten stellen zeer specifieke eisen aan hun habitat. Een voorbeeld hiervan is het verschil tussen jong riet en oud riet als geschikt habitat voor de Grote en Kleine Karekiet. Dergelijke verschillen zijn te klein om in een habitatbeschikbaarheidsmodel duidelijk op te kunnen nemen. Het is voor deze soorten beter om direct een apart habitatgeschiktheidsmodel te gebruiken.

B.2.2 Habitatfactoren per soortgroep

De eisen die een soort stelt aan zijn habitat kunnen worden teruggebracht naar eisen voor factoren die deze habitat vormen (bijvoorbeeld het habitat van een snoek kan worden beschreven aan de hand van de diepte van het water, het doorzicht, de aanwezigheid/afwezigheid van macrofyten, stroomsnelheid en eventueel watertemperatuur). Niet iedere soort is even kritisch ten opzichte van een bepaalde factor. Om een goed overzicht te krijgen van de relevante factoren die een mogelijk habitat kunnen voorspellen, is eerst bekeken welke factoren in de literatuur als kritisch worden beschouwd. De tabellen B8-11 geven een overzicht van de in de literatuur genoemde factoren, die voor de geselecteerde soorten van belang zijn. Hierbij is geen onderscheid gemaakt tussen habitatbeschikbaarheids- en habitatgeschiktheidsfactoren, omdat op deze manier een beeld kan worden gegeven van de meest noodzakelijke eisen die een soort stelt aan zijn habitat.

Het is opvallend dat een groot deel van de factoren niet voor elke soort beschreven zijn. Er zitten veel ‘gaten’ in de tabellen. Voor een deel is dit het gevolg van het feit dat bepaalde soorten geen belang hebben bij die specifieke factoren. Daarnaast kan dit ook een gevolg zijn van de hoeveelheid aan beschikbare literatuur of van de kwaliteit van de habitatbeschrijvingen op zich. Hiervoor wordt verder verwezen naar de discussie (hoofdstuk 4).

Gebaseerd op de inventarisatie van habitateisen is in onderstaande paragrafen een overzicht gegeven van de verschillende habitatfactoren die van belang zijn voor de gekozen soorten. Hierbij wordt onderscheid gemaakt in vier groepen factoren, aquatisch paaihabitat (alleen geldig voor vis en amfibieën), aquatisch fourageerhabitat, terrestrisch reproductie-habitat, terrestrisch fourageerhabitat. Er bestaat een groot verschil tussen de kwantificeerbaarheid en nauwkeurigheid van de verschillende optimumfuncties bij aquatische en terrestrische habitats. De aquatische habitatfactoren zijn veelal kwantitatief beschreven terwijl terrestrische habitats veelal kwalitatief beschreven zijn.

B.3 Aquatische habitatfactoren

Bij de aquatische habitatfactoren bestaat een groot verschil tussen de eisen die gesteld worden aan paai- en fourageergebieden. Dit komt doordat het hier ruimtelijk verschillende eenheden gaat, waarbij de paaiplaats tevens een seizoensgebonden bezoektijd kent. Ook de eisen die opgroeiende vissen aan hun habitat stellen zijn vaak anders dan de eisen die adulten van dezelfde soort stellen. De aquatische fourageergebieden gelden zowel voor vogels als voor vissen.

B.3.1 Paaihabitat

De onderscheiden habitatfactoren voor paaigebieden van vissen zijn waterdiepte, substraattype en stroomsnelheid. Een belangrijke kwalitatieve factor is de watertemperatuur. Voor veel factoren zijn geen eensluidende definities aanwezig. Bijvoorbeeld, de definitie van waterdiepte kan per bron verschillen aangezien diepte bepaald kan zijn ten opzichte van gemiddelde zomer-, jaar- of winterpeil. Ook voor stroomsnelheid is niet aangegeven hoe deze is gemeten (op het sediment, halverwege de waterkolom of gemiddeld na een reeks van metingen). Aangezien veel vissen paaien bij een bepaalde vereiste watertemperatuur (gekoppeld aan tijd in het jaar) is deze fysische parameter hier opgenomen.

Tabel B.2 Habitatfactoren voor paaigebieden. Hierbij is n de hoeveelheid soorten waarvoor een waarde opgegeven is. Cursief afgebeelde factoren zijn factoren die meer als habitatgeschiktheidsfactoren worden gebruikt.

waterdiepte	substraattype	stroomsnelheid	<i>watertemperatuur C</i>
n = 12	n = 15	n = 9	<i>n = 12</i>
Onbekend/niet relevant	Onbekend/niet relevant	Onbekend/niet relevant	<i>Onbekend/niet relevant</i>
<10 cm	steen	<10 cm/s (stagnant)	<i><4</i>
10-20	grind	10-20 cm/s	<i>4-8</i>
20-30	zand	20-30 cm/s	<i>8-12</i>
30-50	klei	30-50 cm/s	<i>12-16</i>
50-80	slib	50 - 70 cm/s	<i>16-20</i>
80-100	macrofyten, dicht	70-100 cm/s	<i>20-24</i>
100-150	macrofyten, open	>100 cm/s	<i>>24</i>
>150			

B.3.2 Aquatisch fourageerhabitat

Voor het definiëren van klassen voor aquatisch fourageerhabitat zijn zowel vissen als vogels en zoogdieren meegenomen. Resultaten zijn weergegeven in Tabel B.3. De waterdiepte en stroomsnelheid verschillen van de klassen in het paaihabitat. De klassen voor de factor structuur van fourageerhabitat zijn gelijk aan de klassen in het paaihabitat. Voor vogels en zoogdieren is met name het habitat areaal en de mate van doorzicht van groot belang.

Voor de factor stroomsnelheid is in de literatuur zowel gebruik gemaakt van kwantitatieve als kwalitatieve beschrijvingen. Bijvoorbeeld: de driehoeksmossel komt in riviersystemen optimaal voor bij een waterdiepte van 1-3 meter beneden het gemiddeld laagwaterpeil (dat is het peil dat gemiddeld 350 dagen per jaar wordt overschreden), terwijl peildynamiek bij broedende futen niet hoger mag zijn dan 25 cm ten opzichte van gemiddeld waterpeil. Het is onduidelijk ten opzichte van wat voor maat de term waterpeil gedefinieerd is. Hierbij zijn voor de vissen vaak kwantitatieve gegevens bekend en voor vrijwel alle andere fauna wordt alleen het onderscheid gemaakt tussen stagnant, langzaam stromend en stromend water. Stagnant is geïnterpreteerd als <10 cm/s, matig stromend als 10-50 cm/s en stromend als >50 cm/s.

Tabel B.3 Habitatklassen voor fourageerhabitat van vissen, vogels en zoogdieren. Hierbij is n de hoeveelheid soorten waarvoor een waarde opgegeven is. Cursief afgebeelde factoren zijn factoren die meer als habitatgeschiktheidsfactoren worden gebruikt.

waterdiepte (cm)	stroomsnelheid (cm/s)	substraattype	Watertemp. zomer (°C)	Doorzicht (cm)	areaal (ha)
n = 20	n = 14	n = 21	n = 9	n = 6	n = 5
Onbekend/ niet relevant	Onbekend/ niet relevant	Onbekend/ niet relevant	Onbekend/ niet relevant	Onbekend/ niet relevant	Onbekend/ niet relevant
<10	<10	steen	<12	<40	<0, 5
10-20	10-20	grind	12-20	40-75	0, 5-10
20-30	20-30	zand	20-24	75-100	>10
30-50	30-50	klei	24-28	>100	
50-70	50-70	slib	>28		
70-100	70-100	macrofyten, dicht			
100-150	>100	macrofyten, open			
150-200		waterkolom			
200-300					
300-400					
>400					

In het huidige MORRES instrument worden habitatfactoren gebruikt die in bovenstaande tabellen nog niet genoemd worden (Tabel B.4). Deze kunnen als aanvulling gezien worden en zijn verder niet onderscheiden naar paai- of fourageerhabitat.

Tabel B.4 Habitatfactoren en hun klassengrenzen zoals gebruikt in het huidige MORRES. De factoren getijdedynamiek, rivierdynamiek en scheepvaart worden gebruikt voor modellen van habitats in riviersystemen, de factoren slibrijkdom en bedekkingsgraad waterplanten voor modellen van meersystemen. Cursief afgebeelde factoren zijn factoren die meer als habitatgeschiktheidsfactoren worden gebruikt.

Getijde dynamiek (cm)	Rivierdynamiek (m)	Slibrijkdom	<i>Bedekkingsgraad waterplanten</i>	<i>Scheepvaart</i>
niet bekend/relevant	Niet bekend/relevant	geen slib of slibarm	< 1%	<i>Niet bekend/relevant</i>
<30	<5	matig slibrijk	1-5 %	<i>geen/gering</i>
30-70	>5	slibrijk	5 - 25 %	<i>matig intensief</i>
70-150			>25%	<i>intensief</i>
>150				<i>zeer intensief</i>

B.4 Terrestrische habitatfactoren

Voor de terrestrische habitats is per factor het onderscheid in klassen een stuk minder eenvoudig kwantitatief weer te geven. Soortspecifieke eisen zijn in de literatuur veel vaker kwalitatief omschreven, waarbij met name de structuur van een habitat van belang blijkt. Daarom zal ook hier een groot deel van de habitatfactoren een beschrijvend karakter hebben. Voordat hiermee kan worden begonnen, is het noodzakelijk enkele begrippen nader te definiëren. Aangezien voor veel faunasoorten de structuur van het terrestrisch habitat het meest belangrijk is, delen we de structuur in op basis van inundatieklassen zoals weergegeven in Tabel B.5. Het aantal vegetatiestructuren is veel kleiner dan het aantal verschillende ecotopen dat op basis van een vegetatiekaart wordt onderscheiden.

Dit is met name het gevolg van het feit dat veel soorten niet zo zeer gekoppeld zijn aan een bepaald vegetatietype, maar meer aan een bepaalde vegetatiestructuur.

Deze vegetatiestructuur kan al worden aangegeven met de zeven klassen die hierboven beschreven zijn.

Tabel B.5 Terrestrische habitatgroepen

habitat	beschrijving
steiloever	langs stromende wateren, onbegroeid
pionier-vegetatie	onbegroeide, kale vlakte met enkele pioniersoorten
grasland	zeer open structuur. inundatie in het broedseizoen niet gewenst. gras in verschillende hoogtes, van kort tot vrij hoog.
ruigte	minder dan 20 dagen geïnundeerd, hoog opschietende vegetatie van kruidachtigen
moeras	permanent geïnundeerd (waterrietmoeras) 360-150 dagen per jaar geïnundeerd (ouder waterrietmoeras) 150-20 dagen geïnundeerd (landrietmoeras)
struweel	niet nader gespecificeerd
bos	frequent geïnundeerd bos (vgl. zachthoutoobos) weinig frequent geïnundeerd bos (vgl. hardhoutoobos)

Inundatieklassen zijn van groot belang voor het type moeras/oever waarin een soort voor kan komen. Met name voor grondbroeders is dit een belangrijke factor, aangezien legsels verloren kunnen gaan bij een sterk wisselend peil gedurende het broedseizoen (Van der Winden et al. 1996). Ook de grondwaterstand beneden maaiveld is een belangrijke factor. Permanent droge terrestrische systemen en veengebonden (moeras-) systemen zijn verder buiten beschouwing gelaten. In het huidige MORRES instrument worden grondwaterstand en inundatieduur als aparte habitatfactoren meegenomen, naast het vegetatietype (Tabel B.6). Wanneer de kennis over deze factoren beschikbaar is, is dit uiteraard de meest volledige benadering.

Tabel B.6 Habitatklassen voor de factoren inundatieduur en grondwaterstand beneden maaiveld.

Inundatieduur (dagen/jaar)	Grondwaterstand (cm beneden maaiveld)
Niet bekend/relevant	Niet bekend/relevant
<2	<20
2-20	20-50
20-50	50-100
50-150	>100
150-360	
>360	

Het type oever en de lengte hiervan is voor vele soorten fauna (met name zoogdieren) een belangrijke factor. Schuilplaats, fourageer areaal en migratiemogelijkheden zijn drie belangrijke functies die een oever heeft voor bijvoorbeeld de otter en de bever. Kleine landschapselementen en een bepaalde mate van afwisseling hierin is ook van belang. Kleine stagnante wateren fungeren als voortplantingshabitat voor amfibieën zoals de kamsalamander.

Reptielen en amfibieën zijn gebaat bij een afwisselend landschap, zowel wat betreft structuur van de vegetatie en de morfologie van het gebied, waarbij droge hoge delen afgewisseld worden met kleine stagnante wateren en laaggelegen ruigten. Al deze opmerkingen meegenomen komen er een aantal belangrijke criteria naar voren voor het definiëren van terrestrische habitatfactoren. Deze zijn weergegeven in Tabel B.7. De vaak voorkomende verwijzing naar het vóórkomen van een grote diversiteit aan kleinschalige landschapselementen is hierbij buiten beschouwing gelaten.

Tabel B.7 Terrestrische habitatfactoren gebaseerd op (Duel *et al.* 1996; Hollander & Reinhold 1999; Laane 1996a; Van der Winden *et al.* 1996; Vanhemelrijk 1993).

Landschapselementen	Talud	Oeverzone lengte	Oeverzone breedte
hardhoutoobos (licht)	Steil	<2 km	<2 m
zachthoutoobos (dicht)	Flauw	2-10 km	2-5 m
moerasriet struweel		>10 km	>5 m
kaal zand (open ruimte)			
grasland (open ruimte)			
kleine stagnante wateren			

B.4.1 Definitie van optimumfuncties

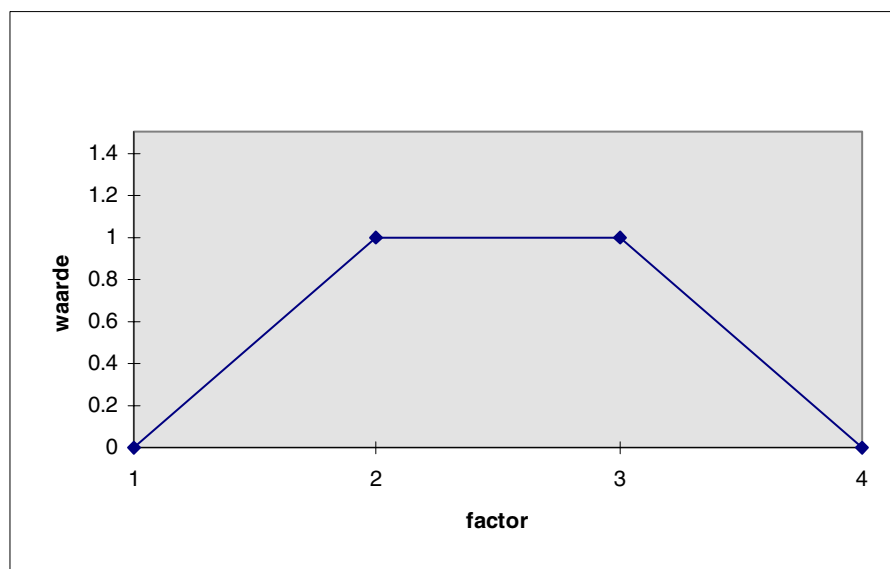
Elke habitatfactor heeft een range waarbinnen een soort optimaal kan functioneren. Door het definiëren van een optimumfunctie kan per soort en per factor een beschrijving worden gemaakt van het mogelijk vóórkomen van een soort bij een bepaalde waarde van deze factor (zie Figuur B.1). Habitatbeschikbaarheid of het areaal aan leefgebieden kan berekend worden met behulp van optimumfuncties. Ook voor het bepalen van de habitatgeschiktheid kunnen deze optimumfuncties worden gebruikt. De factoren voor de bepaling van het areaal aan leefgebied hoeven op deze wijze niet in klassen vast te liggen. Het gebruik van optimumfuncties is waar mogelijk te verkiezen boven het gebruik van klassenverdeling, omdat het gebruik van klassen een grotere onzekerheid impliceert dan dat van optimumfuncties.

Per habitatfactor zijn optimumfuncties bepaald op het moment dat bij een minimum van vijf soorten deze factor wordt genoemd.

Voor de definitie van habitatfuncties zijn de gegevens uit de Tabellen B8-11 in twee EXCEL-sheets weergegeven: een tabel voor gebroken lineaire optimumfuncties en een tabel met klassenverdelingen.

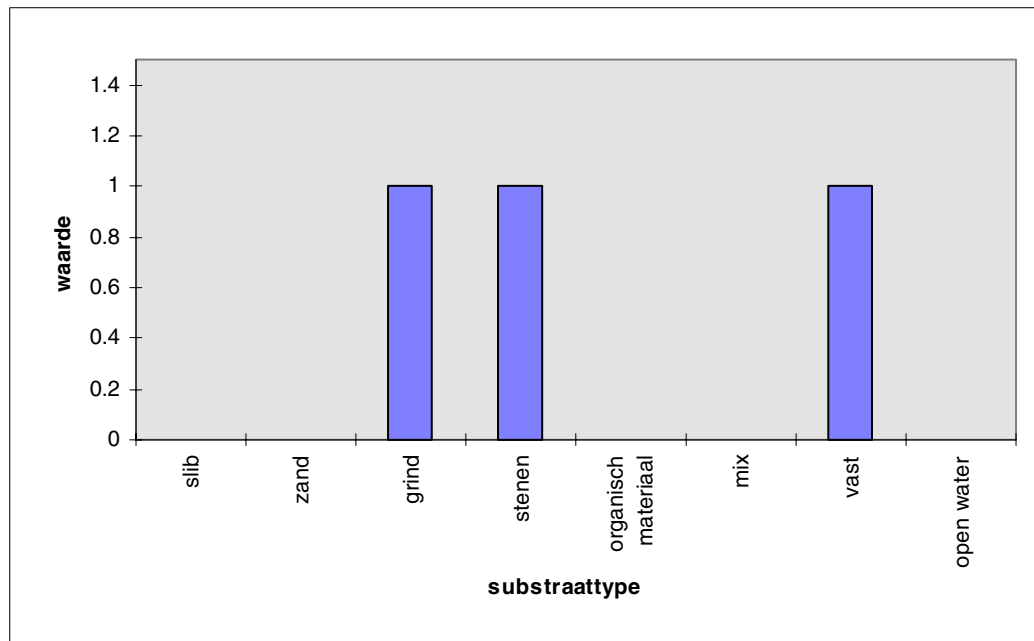
In de tabel met optimumfuncties zijn steeds 4 punten aangegeven (zie ook Figuur B.1):

- A: nulpunt (waarde op de Y-as = 0)
- B: start optimum (waarde op de Y-as = 1)
- C: einde optimum (waarde op de Y-as = 1)
- D: maximum (waarde op de Y-as = 0)



Figuur B.1 Voorbeeld van een optimumfunctie voor een kwantitatieve habitatfactor (zie tevens tekst).

De tabel met klassenverdelingen betreft factoren die slechts kwalitatief beschreven zijn (bijvoorbeeld substraattypen). Hierbij is voor elke klasse aangegeven of dit gewenst of niet gewenst is (1 of 0; zie Figuur B.2). Wanneer voldoende gegevens bekend zijn over aantallen van een soort per klasse, kunnen ook waarden lager dan 1 toegekend worden.



Figuur B.2. Voorbeeld van substraattypen voorkeuren van een soort (hier Driehoeksmossel).

Waar noodzakelijk is onderscheid gemaakt tussen de verschillende deelhabitats en is voor de afzonderlijke delen een optimumfunctie, dan wel kwalitatieve beschrijving gemaakt. Deze staan allen weergegeven in B12 en B13. Er zijn zowel optimumfuncties gemaakt voor habitatbeschikbaarheids- als habitatgeschiktheidsfactoren. Gegevens met betrekking tot afstand tussen fourageer- en reproductiegebied en gewenst areaal habitat betreffen de samenhang van habitats en zijn in deze studie buiten beschouwing gelaten.

B.5 Vergelijking habitatcriteria met ecotoopcriteria (RWES)

De habitatfactoren zoals beschreven in hoofdstuk B.2 kunnen worden vergeleken met de bestaande ecotopen en factoren die daaraan zijn gekoppeld. Het Rijkswateren Ecotopenstelsel voor aquatische systemen zal als voorbeeld worden gebruikt.

De RWES ecotopen zijn ingedeeld naar gekwantificeerde criteria. Deze kartering kan als input dienen voor habitatmodellen. In hoofdstuk B.9 is een overzicht gegeven van factoren die gebruikt worden bij de indeling in aquatische ecotopen, zoals ze in het RWES-aquatisch zijn opgenomen (Van der Molen *et al.* 2000).

Veel van de RWES-habitatfactoren verschillen per type watersysteem. De volgende watertypen zijn onderscheiden: rivieren (stromend), rivieren (stagnant), getijde wateren, meren, kanalen. De onderscheiden factoren in het RWES-aquatisch zijn mechanische dynamiek, waterdiepte, sedimenttype en saliniteit. De klassengrenzen die zijn aangegeven zijn gebaseerd op ecologische verschijnselen (bijvoorbeeld de diepte waarbij macrofyten nog voorkomen).

De definitie van waterdiepte in het RWES-aquatisch verschilt per watersysteem. Hierbij valt op dat er slechts 5 klassen worden onderscheiden, terwijl met name voor vishabitats over veel gedetailleerdere grenzen wordt gesproken. Er ligt in de soortgebonden habitatbeschrijvingen ook veel meer nadruk op de ondiepe delen van het water.

De sedimenttypen zoals in het RWES beschreven komen redelijk overeen met de substraattypen zoals onderscheiden in hoofdstuk B.2. Ook hier wordt een indeling gemaakt in verschillende klassen. Het RWES geeft zelfs een korrelgrootte weer voor de verschillende klassen. Dit is in de soortgebonden literatuur niet gevonden. De ecotopenkaart van het RWES-aquatisch kan hiermee goed gebruikt worden als invoerkaart voor de habitatfactor bodemtype. Informatieverlies hoeft niet op te treden.

Mechanische dynamiek wordt niet als zodanig gebruikt bij het beschrijven van soortgebonden habitatvoorkeuren. Slechts impliciet worden hier in de soortliteratuur eisen aan gesteld, aangezien er per watertype verschillende factoren zijn die bijdragen aan de term 'dynamisch'. Voor een aantal broedvogels is weergegeven dat peilfluctuaties ongewenst zijn en voor paaigebieden geldt vaak een bepaalde stroomsnelheid (die sedimentatie beïnvloed). Het is goed mogelijk dat mechanische dynamiek een relevante factor vormt voor soortgroepen die niet zijn meegenomen in deze studie, zoals bijvoorbeeld macrofauna. Of mechanische dynamiek gedetailleerder moet worden beschreven voor macrofauna habitats is op dit moment niet duidelijk.

Saliniteit wordt genoemd in het RWES en is in hoofdstuk B.2 buiten beschouwing gelaten aangezien alleen is gekeken naar soorten gerelateerd aan zoete wateren. De conclusie van bovenstaande is dat ecotopenkaarten geschikte informatie kunnen geven over enkele relevante habitatfactoren zoals diepte en sedimenttype. Voor een goede habitatbeschrijving van veel soorten is aanvullende informatie nodig over onder andere stroomsnelheid. Voor de factor diepte in de ecotoopindeling zijn de klassen veel ruimer gedefinieerd dan de grenzen van de habitats. Dit beperkt dus de mogelijkheid om de habitatbeschikbaarheid goed te kwantificeren.

B.6 Conclusies en aanbevelingen

Aan de hand van 35 soorten, geselecteerd op basis van biotoopgebruik, is een overzicht gegeven van de eisen die deze soorten aan hun habitat stellen. Hierbij is onderscheid gemaakt in aquatisch paaihabitat, aquatisch fourageerhabitat en terrestrisch habitat. De belangrijkste factoren die in de onderzochte literatuur genoemd worden zijn: waterdiepte, stroomsnelheid, substraatype, landschapselementen en talud. Daarnaast zijn er nog een aantal habitatfactoren genoemd die betrekking hebben op de kwaliteit van het habitat (watertemperatuur, doorzicht en areaal).

Er blijkt een groot verschil te zijn in detailniveau waarop een habitatfactor is beschreven, waarbij terrestrische systemen veel grover en kwalitatief zijn beschreven, terwijl aquatische systemen veel gedetailleerder en kwantitatief beschreven zijn. Ook blijkt dat het vergelijken van gegevens per soort met hetzelfde type gegevens van een andere soort niet goed mogelijk is, aangezien er voor veel factoren niet is aangegeven hoe deze geregistreerd zijn (methodes, tijd van het jaar, gebrek aan definities).

B.7 Habitatieisen voorbeeldsoorten

Overzicht van habitatieisen voor de voorbeeldsoorten uit diverse soortsgroepen. Verscheidende soortsgroepen zijn per tabel samengevoegd uit praktische overwegingen met betrekking tot overzichtelijkheid voor bepaalde factoren. Ook vallen sommige soorten binnen twee soortsgroepen, bijvoorbeeld de aalscholver die een typische bosbroedvogel is, maar tevens fourageert in open water (zichtpredatie).

Tabel B.8. Habitatfactoren van belang voor soortsgroepen **open water vis, waterplant gebonden vissen en bodemfauna** (Glasbergen 2001; Laane 1996b).

	Serpeling	Rivieronderpad	Snoek	Snoekbaars	Brasem	Kroeskarper	Driekhoeks- mossel
PAAI FYSISCH							
gewenste watertemp. °C	12-20	11-17	8-12	12-15	12-20	14-20 cm	
min. watertemp. °C	8	7	5	6	8		15
max. watertemp. °C	28	19	16	26	24		
gewenste diepte (cm)	25-40	10-40	30-60	50-150	20-80		
min. diepte water (cm)	5-10	10	15		9	>20	
max. diepte paaiplaats (cm)	50	60	100		120		
structuur paaigebied	grind met matig slib, een matige bedekking met macrofyten	zand, grind, stenen	dichte macrofyten-begroeiing	substraat zand grind of klei met macrofyten	macrofyten/vast substraat		
ongewenste structuur	slib		slib	slib			
gewenste stroomsnelheid (cm/s)	30-50	<50s				stagnant	
PAAI CHEMISCH							
Zuurstof concentratie (juv.) (mg/l)	>7	100%	>3	>4, 5	> 2		
FOURAGEER AQUATISCH FYSISCH							
maximum temp water °C	28	19	30	35	28-30		30
gewenste temp. water zomer		11-17					17-20
min. breedte water (m)							2
min. areaal (ha)							10
gewenst habitat diepte (cm)	70-150	10-40	matig diep	>1m (tot 25 m diep)		ondiep	>200
min. waterdiepte		10	ondiep				
gewenste stroomsnelheid (cm/s)	20-50	30		<10	3-8	stagnant	zowel stagnant als stromend
stroomsnelheid max. (cm/s)	70	120	10	25	20		
fourageergebied			vegetatie	open water	zachte bodems		
beschaduwing	min. 20%						
macrofytenbegroeiing	max. 40%	niet nodig	gewenst				
substraat	variabel	zand, grind, stenen	geschikt voor macrofyten vestiging		deeltjes groter dan 0.4 mm	hoge slib concentratie	hard substraat
ongewenste substraat				vegetatie	grind	grind	veen, slib
FOURAGEER AQUATISCH CHEMISCH							
benodigde O ₂ concentratie adulten	>7	100%	>7 mg/l	>6	> 5 mg/l	kan tegen zeer lage concentraties	>5
Cl concentratie optimaal							<300 mg/l
Ortho fosfaat							0, 02-0, 10 mg/l
calcium min.							28 mg/l
calcium optimaal							45-50 mg/l
pH							7, 8-8, 5

Tabel B.9 Habitatfactoren soortsgroepen **open water fourageerders, bos- en moerasbroedvogels** (Duel *et al.* 1996; Van der Winden *et al.* 1996; Laane 1994)

	IJsvogel	Fuut	Kuifeend	Roerdomp	Aalscholver
AQUATISCH FOURAGEER fysisch					
minimaal areaal (ha)	0,5	>0,8			10
gewenst habitat diepte (m)	0,3-1	2-5	2-4		3-4
minimale waterdiepte (m)	0,1	0,8	0,5		2
geprefereerde stroomsnelheid	stromend			stagnant	
macrofytenbegroeiing	ongewenst		<20%		ongewenst
substraat	zand				
doorzicht minimaal	matig troebel	25 cm			>40 cm
doorzicht optimaal	helder	>75 cm			>40 cm
extra opmerkingen	uitkijkposten		combinatie winter grote meren, genoeg Dreissenidae	eet geen brasem	
AQUATISCH FOURAGEER chemisch					
Cl concentratie optimaal		<300			
Cl maximaal		1000		tolerant voor brak	indifferent
Ortho fosfaat					0, 2-0, 5 mgP/l
TERRESTRISCH REPRODUCTIE					
habitatstructuur	steiloever			overjarig riet afgewisseld met open water	zacht hout oobos, moerasbos
minimale breedte habitatstructuur (m)	2	1			
optimale breedte	5	>3		>10	
broedperiode				half maart/eind augustus	april mei
kenmerken				nest in minimaal 10-25 cm water	hoge bomen in soppig/natte ondergrond
opmerking				gevoelig voor verstoring (400m.)	verstoringafstand 400 m
TERRESTRISCH FOURAGEER					
vegetatiestructuur				beschutte oevers	
mijdt				open veld	
waterdiepte helofytenzone optimaal				20-50 cm	
waterdiepte helofytenzone minimaal				10	
waterdiepte helofytenzone maximaal				70	

Tabel B.10 **Terrestrische fourageerders** (Ruitenbeek *et al.* 1990).

	grutto	kluut	kemphaan	bosrietzanger	kleine karekiet	grote karekiet
TERRESTRISCH REPRODUCTIE						
vegetatiestructuur	vochtig cultuurgrasland met hoog gras	kaal, schaars begroeide bodem	open weiland	vegetatie van natte, periodiek overstromde gebieden	dicht riet, verticale stengels nodig	overjarig riet, stevige stengels, vooral te vinden langs grote meren en plassen
TERRESTRISCH FOURAGEER						
vegetatiestructuur	vochtige graslanden op veen	slibrijke bodems ondiep	slikkige drassige plekken met weinig vegetatie	als het maar ruig is		wilgenopslag en ruitgen/verlandingszones

Tabel B.11 Habitatfactoren van belang voor **waterminnende reptielen/amfibieën/zoogdieren** (Duel *et al.* 1996;
Hollander & Reinhold 1999)

	bever	otter	noordse woelmuis	kamsalamander	ringslang	rugstreeppad
AQUATISCH REPRODUCTIE						
maximum diepte paaiplaats				50 cm		
structuur paaigebied				veel macrofyten		dynamisch, zandbodem/macrofyten begroeiing
AQUATISCH FOURAGEER						
minimale breedte water		>5 m				
minimaal areaal		10-20 km oever, >0, 5 ha				
gewenst habitat diepte	>50 cm	50-250 cm				
minimale waterdiepte	>50 cm of >1, 50 afhankelijk van literatuur bron	50 cm				
maximale waterdiepte		2-3 meter				
stroomsnelheid maximaal		stromend		stagnant		
stroomsnelheid minimaal		stagnant				
fourageergebied	oever/macrofyten					
macrofytenbegroeiing	gewenst	niet te veel		gewenst	gewenst	
substraat				zand, leem		
ongewenste condities				matig voedselrijk		
doorzicht minimaal	troebel	60 cm				
doorzicht optimaal		>100 cm				
extra opmerkingen		zeer gevoelig voor vervuiling	tolerant voor winterinundatie	ecotoop geïsoleerde strangen	ondiepe, stilstaand tot langzaam stromend	kleinschalige stagnante wateren
TERRESTRISCH REPRODUCTIE						
vegetatiestructuur	licht bos	dichte moerasvegetatie			bladhopen in bos/broeiplaatsen	
minimale breedte vegetatiestructuur		>2				
optimale breedte		>5				
reproductieperiode		gehele jaar				
kenmerken	burchten	lengte en type oever van belang voor migratie	handhaaft zich goed in dynamische omgeving		lijnstructuren voor migratie	
opmerking	geen peilfluctuaties					overwintering op hoger gelegen delen
TERRESTRISCH FOURAGEER						
vegetatiestructuur	bos	dichte oevers, dekking in moeras- en broekbossen, struwelen, rietlanden	vochtige tot zeer natte rietlanden, natte hooilanden, periodiek overstroomde terreinen		afwisselend bos, open, nat	
mijdt	niet-bos gebieden	steile taluds, open ruimtes	bos			
reliëf		flauw				
voedselbron	boom en kruid	vis	planten			
opmerking	oeverlengte >8 km, bos direct langs oever noodzakelijk	moeras, meer, rivier en beek	vegetatie bij voorkeur gevarieerd en gesloten	zand, leem		
inundatieduur				niet meer dan 10- 20 dagen		
schuilmogelijkheden		aanwezigheid van dekking, structuur van oever, daar is heel wat aan gemeten	begrazing niet wenselijk			

B.8 Optimumfunctie-tabellen

OPTIMUMFUNCTIES VOOR HEP- INSTRUMENTARIUM

Per optimumfunctie bestaan er 4 punten: ·

A: nulpunt of minimale waarde	Y-as 0
B: start optimum	Y-as 1
C: einde optimum	Y-as 1
D: maximale waarde	Y-as 0

Wanneer er voor een punt een range wordt weergegeven komt het gemiddelde in de desbetreffende kolom en wordt het maximum en minimum in naastgelegen tabel weergegeven. Per soort zullen de factoren onder elkaar worden weergegeven. Voor macrofyten wordt de bedekking in percentages weergegeven, ook als deze kwalitatief is beschreven in de literatuur. Een ~ geeft aan dat er geen maximum is aangegeven en dat alles boven het optimum in principe goed is.

	A	B	C	D
--	---	---	---	---

soort: Brasem (Abramis brama)

functioneel habitat: **paaihabitat**

watertemp (C)	8	12	20	24
waterdiepte (cm)	9	20	80	120
stroomsnelheid (cm/s)				
O2 concentratie	3	5	~	~
macrofyten (%)	10	20	30	40

soort: Brasem (Abramis brama)

functioneel habitatdeel: **fourageerdeel**

watertemp (C)				28
waterdiepte (cm)				
stroomsnelheid (cm/s)	0	3	8	20
O2 concentratie	2	5	~	~
macrofyten (%)	0	0	0	10

soort: Driehoeksmossel

functioneel habitatdeel: **totaal habitat**

watertemp (C)	2	17	20	30
waterdiepte (cm)	200	200	~	~
stroomsnelheid (cm/s)				
O2 concentratie	5	5	~	~
minimum areaal (ha)	2	2	~	~
macrofyten (%)	~	~	0	0

soort: Kroeskarper (Carassius carassius)

functioneel habitat: **paaihabitat**

watertemp (C)		14	20	
waterdiepte (cm)	20	20		
stroomsnelheid (cm/s)	0	0	3	5
O2 concentratie				
macrofyten (%)	50	70	90	100

soort: Kroeskarper (Carassius carassius)

functioneel habitatdeel: **fourageerdeel**

watertemp (C)				
waterdiepte (cm)	20			
stroomsnelheid (cm/s)	0	0	15	50
O2 concentratie	-			
macrofyten (%)				

soort: Rivierdonderpad (Cottus gobio)

functioneel habitat: totaal

watertemp (C)	7	11	17	19
waterdiepte (cm)	10	10	40	60
stroomsnelheid (cm/s)	50	50	~	~
O2 concentratie	100%	~	~	~
macrofyten (%)	0	0	60	80

soort: Serpeling (Leuciscus leuciscus)

functioneel habitatdeel: **paaihabitat**

watertemp (C)	8	12	20	28
waterdiepte (cm)	7	25	40	50
stroomsnelheid (cm/s)	0	30	50	100
O2 concentratie	6	7	7	~
macrofyten (%)	10	30	40	50

soort: Serpeling

functioneel habitatdeel: **fourageerdeel**

watertemp (C)				28
waterdiepte (cm)		70	150	
stroomsnelheid (cm/s)	0	20	50	70
O2 concentratie	6	7	7	~
macrofyten (%)	0	40	40	~

soort: Snoek (Esox lucius)

functioneel habitatdeel: **paaihabitat**

watertemp (C)	5	8	12	16
waterdiepte (cm)	15	30	60	100
stroomsnelheid (cm/s)	0	0	15	25
O2 concentratie	3	3	~	~
macrofyten (%)	50	75	~	~

soort: Snoek (Esox lucius)

functioneel habitatdeel: **fourageerdeel**

watertemp (C)				29
waterdiepte (cm)				
stroomsnelheid (cm/s)			10	10
O2 concentratie	5.7			
macrofyten (%)	40	60	100	~

soort: Snoekbaars (Stizostedion lucioperca)functioneel habitatdeel: **paaihabitat**

watertemp (C)	6	12	15	26
waterdiepte (cm)		50	150	
stroomsnelheid (cm/s)				
O2 concentratie	4.5			
macrofyten (%)	10	20	30	40

soort: Snoekbaars (Stizostedion lucioperca)functioneel habitatdeel: **fourageerhabitat**

watertemp (C)				35
waterdiepte (cm)		100	150	~
stroomsnelheid (cm/s)	0	0	10	25
O2 concentratie	6	6	~	~
macrofyten (%)	0	0	10	10

soort: IJsvogel (Alcedo atthis)

min. areaal (ha)	0.5	1	~	~
waterdiepte (cm)	10	30	100	
stroomsnelheid				
macrofytenbegroeiing			0	0

soort: Fuut (Podiceps cristatus)

min. areaal (ha)	0.8	1	~	~
waterdiepte (cm)	80	200	500	600
stroomsnelheid				
macrofytenbegroeiing				
doorzicht (cm)	25	75	~	~
oeverbreedte (cm)	100	300	~	~

soort: Kuifeend (Aythya marila)

min. areaal (ha)				
waterdiepte (cm)	50	200	400	600
stroomsnelheid				
macrofytenbegroeiing	0	0	0	20
doorzicht (cm)				
oeverbreedte (cm)				

soort: Roerdomp (Botaurus stellaris)

nesthabitat

oeverbreedte (cm)	1000	1000	~	~
waterdiepte rond nest (cm)	10	25		

soort: Aalscholver (Phalacrocorax carbo)

min. areaal (ha)	10			
waterdiepte (cm)	200	300	400	600
stroomsnelheid				
macrofytenbegroeiing			0	10
doorzicht (cm)	40			

soort: Bever (Castor fiber)

waterdiepte (cm)	50	150		
macrofytenbegroeiing	30	40	50	70

soort: Otter (Lutra lutra)

minimale breedte water	5	6	~	
min. areaal	0.5			
waterdiepte	50	50	250	350
doorzicht (cm)	60	100	~	~
min. breedte oever	200	500	~	~

soort: Kamsalamander (Cristurus cristatus)

paaiplaats diepte				50
-------------------	--	--	--	----

KLASSENVERDELINGEN VOOR KWALITATIEVE GEGEVENS

Enkele habitatfactoren worden aangegeven in beschrijvende termen:

voor deze habitatfactoren moeten de afzonderlijke opties worden geclassificeerd als gunstig (1) of ongunstig (0)

	SUBSTRAAT								WATERDIEPTE			STROOMSNELHEID			DOORZICHT			VEGETATIE/terrestrisch habitatstructuur										
	slib	zand	grind	stenen	organisch materiaal	mix	vast	open water	ondiep	matig diep	diep	stilstaand	langzaam stromend	snel stromend	troe bel	matig troebel	hel der	riet- kraag	moeras bos	riet kraag jong	riet kraag oud	steil- oever	vochtig grasland	kaal	wilgen opslag	ruigte		
Brasem; paai																												
substraat	0	0	1	1	0	0	1				1	0	0															
Brasem: fourageer																												
substraat	1	1	0	0			0																					
Driekhoeksmossel																												
substraat, stroomsnelheid	0	0	1	1	0	0	1				1	1	1															
Kroeskarper (Carassius carassius) functioneel habitatpaaihabitat																												
stroomsnelheid											1	0	0															
Kroeskarper (Carassius carassius) fourageerdeel																												
substraat, waterdiepte, stroomsnelheid	1	1	0	0	0	0			1	0	0	1	0	0														

Bosrietzanger (Acrocephalus palustris)																											
vegetatie																	1	1	1	1	0	0	0	1	1		
Kleine karekiet (Acrocephalus scirpaceus)																											
vegetatie																	1	0	1	0	0	0	0	0	0		
Grote karekiet (Acrocephalus arundinaceus)																											
vegetatie																	1	0	0	1	0	0	0	1	1		
Bever (Castor fiber)																											
waterdiepte, doorzicht								0	1	1							1	1	1	0	1	0	0	0	0	1	1
Otter (Lutra lutra)																											
stroomsnelheid, vegetatie										0	1	1					1	1	1	1	0	0	0	1	1		
Noordse woelmuis (Microtus oeconomus)																											
vegetatie																	1	0	1	1	0	1	0	0	1		
Kamsalamander (Cristurus cristatus)																											

B.9 RWES-factoren

Tabel B.14 De onderscheiden factoren die de mechanische dynamiek bepalen in het RWES zijn afhankelijk van het type watersysteem waarmee gewerkt wordt.

Hoofdwatersysteem	Factoren die mechanische dynamiek bepalen
Rivieren	wisselende afvoeren van water en sediment in de rivier (rivierregime), golfwerking door scheepvaart en daarnaast ook in mindere mate wind
Getijdenwateren	de dagelijkse richting van wisselende getijdenstromingen en de door de rivieren veroorzaakte onregelmatig wisselende afvoeren van water en sediment en daarnaast ook golfwerking door wind en scheepvaart
Meren	wind, strijklengte, grootte van het meer, wisseling in waterstanden, waterdiepte, stroming, type substraat of bodem en de wisselende aan- en afvoer van water en sediment
Kanalen	vaarbeweging van schepen en daarnaast ook de aan- en afvoer van water, stuw- en spuiregimes, lozingen en wind

Tabel B.15 Ecologische indeling van de term dynamisch volgens het RWES-aquatich

Klassen	Begrenzende biologische factor
Zeer sterk dynamisch (z)	geen sessiele soorten; vestiging van vegetatie wordt verhinderd; bodemfauna beperkt tot enkele beweeglijke soorten; bodem is meer dan decimeters diep continu in beweging; regelmatig nieuwe standplaatsen gevormd
Sterk dynamisch (s)	enkel aangepaste sessiele soorten; weinig wormen, relatief veel kreeftachtigen; bodem is centimeters tot decimeters diep regelmatig in beweging
Dynamisch (d)	relatief hoge biomassa en diversiteit aan soorten; veel variatie aan substraat; er worden nieuwe standplaatsen gevormd; bodem in beweging maar vestiging of het voorkomen van vegetatie en/of bodemfauna wordt niet blijvend verhinderd
Laag dynamisch (l)	voorkomen van vegetatie en/of bodemfauna wordt niet of nauwelijks beïnvloed; relatief veel wormen

Tabel B.16 Afbakening van de verschillende klassen voor de factoren waterdiepte, op basis van biologische factoren

Klassen	Begrenzende biologische factor
Zeer diep (z)	langdurige stratificatie als gevolg van verschillen in temperatuur of zoutgehalte
Diep (d)	geen langdurige stratificatie en geen uitgebreide bedekking mogelijk met waterplanten
Matig diep (m)	uitgebreide bedekking met waterplanten mogelijk, doch niet met helofyten
Ondiep (o)	helofyten kunnen voorkomen

Tabel B.17 Beschrijving van de verschillende diepteklassen van de delen van hoofdwatersystemen die nagenoeg permanent met water bedekt zijn.

Klasse	Watersysteem	Begrenzing
Zeer diep water (z)	- stromend Rivieren - stagnant Getijdenwateren Meren Kanalen	niet van toepassing bij gemiddeld zomerpeil dieper dan 5 meter bij gemiddeld laagwater dieper dan 10 meter bij gemiddeld zomerpeil dieper dan 5 meter bij gemiddeld kanaalpeil dieper dan 10 meter (alleen in brak water)
Diep water (d)	- stromend Rivieren - stagnant Getijdenwateren Meren Kanalen	bij gemiddeld laagwater dieper dan 2 meter bij gemiddeld zomerpeil tussen de 3 en 5 meter diep bij gemiddeld laagwater tussen de 3 en 10 meter bij gemiddeld zomerpeil tussen de 3 en 5 meter diep bij gemiddeld kanaalpeil tussen de 2 en 10 meter diep
Matig diep water (m)	- stromend Rivieren - stagnant Getijdenwateren Meren Kanalen	bij gemiddeld laagwater tussen de 1 en 2 meter diep bij gemiddeld zomerpeil tussen de 1 en 3 meter diep bij gemiddeld laagwater tussen de 1 en 3 meter diep bij gemiddeld zomerpeil tussen de 1 en 3 meter diep bij gemiddeld kanaalpeil tussen de 1 en 2 meter diep
Ondiep water (o)	- stromend Rivieren - stagnant Getijdenwateren Meren Kanalen	nooit of zelden (gemiddeld minder dan 2 dagen per jaar) droogvallende standplaats, bij gemiddeld laagwater tussen 0 en 1 meter diep bij gemiddeld zomerpeil tussen de 0, 3 en 1 meter diep nooit of zelden droogvallende standplaats, bij gemiddeld laagwater tussen 0 en 1 meter diep nooit of zelden droogvallende standplaats, die bij een gemiddeld zomerpeil tussen de 1 en 0, 3 meter diep is nooit of zelden droogvallende standplaats, die bij een gemiddeld kanaalpeil tussen de 1 en 0, 3 meter diep is

Tabel B.18 Typering sediment

Klasse	Korrelgrootte (mm)	Sedimenttypes
Klei	<2	
Slib	0 of 2-63	slib
Zand	63-150	fijn zand
	150-250	matig fijn zand
Grof zand	250-500	matig grof zand
	500-2000	grof-zeer grof zand
Grind	>2000	grind

Tabel B.19 Saliniteit is in de habitatbeschikbaarheid niet specifiek meegenomen aangezien we alleen naar de zoete systemen gekeken hebben. Hieronder een overzicht van de klassen die worden onderscheiden in saliniteit.

Zonnering (invloed alleen NACI)	Saliniteit (g Cl/l)	Klasse
Rivier	<0, 3	zoet
Zoetwatergetijdezone	<0, 3	zoet
Oligohaliene zone	0, 3-3	(zwak) brak
Mesohaliene zone	3-10	brak
Polyhaliene zone	10-18	(sterk) brak
Euhaliene zone	18-24	zout
Hyperhaliene zone	>24	zout

B.10 Referenties

Aarts, H., G. Van der Lee & D. T. Van der Molen (2001) *Toepassing MORRES op het Volkerakmeer. Effecten van onzekerheden in de invoergegevens op de habitatgeschiktheid van de Baars, Waterral, Fuut, Otter, Woelmuis en Slobend.* Lelystad RIZA Werkdocument 2001.136X.

Baptist, M. J., H. Duel & M. B. De Vries (1999) *MORRES: Een ecologische module voor het uitvoeren van habitatevaluaties. Beknopte gebruikershandleiding.* Delft WL | Delft Hydraulics rapport T2197.

Duel, H., G. B. M. Pedroli & G. Arts (1996) *Watersysteemverkenningen. Een Stroom Natuur. Achtergronddocument B: Ontwikkelingsmogelijkheden voor doelsoorten.* RWS RIZA werkdocument 95.173X.

Duel, H., B.P.M. Specken, W.D. Denneman & C. Kwakernaak (1995) *The Habitat Evaluation Procedure as a tool for ecological rehabilitation of wetlands in the Netherlands.* Water Science and Technology, **31**, 387-391.

Glasbergen, M. J. (2001) *Fish habitat requirements in the Dutch part of the river Rhine.* Delft WL | Delft Hydraulics T2401.

Duel, H., Pouwels, R. & De Vries, M.J., 2001. *Visie op het modelinstrumentarium voor het uitvoeren van habitatevaluaties van de Nederlandse rijkswateren.* Delft WL | Delft Hydraulics Q2812.

Hollander, H. & J. O. Reinhold (1999) *HSI-modellen voor 5 oevergebonden zoogdiersoorten.* Delft Rijkswaterstaat DWW W-DWW-99-011.

Laane, W. E. M. (1994) *Habitatmodellen vogels (1) Aalscholver, Fuut, Grote Zaagbek, Kleine Zwaan en Kuifeend.* Lelystad RIZA Werkdocument 96.047X.

Laane, W. E. M. (1996a) *Habitatmodel Otter.* Lelystad RIZA Werkdocument 96.045X.

Laane, W. E. M. (1996b) *Habitatmodellen Driehoeksmossel, Hydropsyche contubernalis en Oeveraas.* Lelystad RIZA Werkdocument 96.044X.

Rademakers, J. G. M. & H. P. Wolfert (1994) *Het Rivier-Ecotopen-Stelsel. Een indeling van ecologisch relevante ruimtelijke eenheden ten behoeve van ontwerp- en beleidsstudies in het buitendijkse rivierengebied.* Lelystad RIZA Publicaties en rapporten van het project 'Ecologisch Herstel Rijn en Maas' 61-1994.

Ruitenbeek, W., Scharringa, C.J.G., Zomerdijk, P.J. (1990) *Broedvogels van Noord-Holland,* Stichting Samenwerkende Vogelwerkgroepen Noord-Holland, Assendelft.

Van der Lee, G., H. Duel, S. Groot, H. Aarts & R. Pouwels (2000) *Kwaliteit van het HEP-instrumentarium voor toepassing in het IJsselmeergebied.* RIZA, Alterra, WL | Delft Hydraulics WL | Delft Hydraulics report T2391.

Van der Meulen, Y. A. M. (1997) *Meren Ecotopen Stelsel. een ecotopenstelsel voor de meren van het IJsselmeergebied en Volkerak-Zoommeer.* Lelystad RIZA RIZA nota 97.076.

Van der Molen, D. T., H. P. A. Aarts, J. J. G. M. Brackx, E. F. M. Geilen & M. Platteeuw (2000) *Rijkswateren-Ecotopen-Stelsels. Aquatisch.* Lelystad RIZA RIZA rapport 2000.038, RWES rapport nr. 5.

Van der Winden, J., S. Dirksen & M. J. M. Poot (1996) *HSI-modellen voor 15 oevergebonden broedvogelsoorten.* DWW-RWS

Vanhemelrijk, J. A. M. (1993) *Amoebes IJsselmeergebied. Basisrapport.* RIZA RIZA nota 93.015.

C Statistisch model voor onzekerheden

C.1 Een statistisch model voor onzekerheden

Er dient voor vier aspecten in het habitatmodel een statistisch model voor onzekerheden te worden gedefinieerd. Twee hiervan hebben betrekking op de specificatie waarmee uit een habitatfactor een habitatindex (met waarde tussen 0 en 1) moet worden berekend. Deze relatie wordt aan de hand van een partitie van het domein van de habitatfactor vastgelegd. Dat betekent dat voor een beperkt aantal habitatfactoren de bijhorende habitatindex wordt gegeven (dat wil zeggen specificatie van de habitatindex op een aantal breekpunten in het domein van de habitatfactor). Als bij berekeningen een habitatfactor een tussenliggende waarde aanneemt (dat wil zeggen niet op één van de breekpunten) dan volgt diens habitatindex via lineaire interpolatie van de habitatindices die bij de breekpunten horen. De twee aspecten van de onzekerheidsmodellering in dit sub-model zijn dan:

1. Het onzekerheidsmodel van de habitatindex voor de breekpunten van de habitatfactor. Dit heeft dus betrekking op de onzekerheid in de *verticale* richting ('y-as' met een habitatindex als uitvoergrootheid) van de habitatfactor-habitatindex figuur.
2. Het onzekerheidsmodel voor een habitatfactor wanneer die in modelberekeningen als invoer aan het model wordt aangeboden om hiervoor de bijhorende habitatindex te bepalen. Dit heeft dus betrekking op de onzekerheid in de *horizontale* richting ('x-as' met een habitatfactor als invoergrootheid) van de habitatfactor-habitatindex relatie.

Een derde aspect waarvoor een onzekerheidsmodel moet worden opgesteld heeft betrekking op:

3. De onzekerheid in de referentiedichtheid. Deze referentiedichtheid wordt gebruikt om uit de habitatkwaliteit een potentiële dichtheid te berekenen.

In alle drie gevallen betekent "onzekerheidsmodel" dat een geschikte kansverdeling gedefinieerd moet worden. Bij een Monte-Carlo (MC) simulatie worden hier dan trekkingen genomen, en als zodanig een ensemble van realisaties gegenereerd (voor de habitatindex, habitatkwaliteit, beperkende factoren, potentiële dichtheid, etc.).

Bovenstaande drie aspecten hebben betrekking op onzekerheden in de definitie van het model, en/of diens *invoergrootheden*. Aan de andere kant hebben we na de MC-simulatie een ensemble van *uitvoergrootheden*, en kunnen kentallen voor de onzekerheden van die uitvoergrootheden afgeleid worden. Dit kan op basis van de empirische verdelingen van die uitvoergrootheden (en omvat opslag en sorteren van ensemble uitkomsten), òf op basis van geschikte parameterisaties van die grootheden. De laatste variant heeft als belangrijk voordeel dat er beduidend minder tussen-resultaten van de MC bewaard hoeven te worden.

Gezien het enorme grote aantal cellen dat voorzien is in PC-Raster berekeningen (bijvoorbeeld $10\text{km} \times 10\text{km}$ gebied, met cellen van bijvoorbeeld $25\text{m} \times 25\text{m}$) is zo'n geparametriseerde aanpak van de verdelingen van uitvoergrootheden eerder een praktische noodzaak dan 'gemak'. Feitelijk zijn we daarmee op een vierde aspect van de onzekerheden beland, namelijk:

4. Het onzekerheidsmodel voor, maar hier in de betekenis van "een parameterisatie van-, de diverse uitvoergrootheden van het model, zoals bijvoorbeeld de habitatkwaliteit. Hieruit moeten kentallen (bijvoorbeeld gemiddelden, spreidingen, kwartielpunten, etc.) voor onzekerheden van die uitvoergrootheden afgeleid worden.

Gewenste eigenschappen van (een klasse van) verdelingsfuncties zijn onder andere:

- i. De klasse van verdelingen is op een begrensde domein $[a, b]$ gedefinieerd, waarbij we zelf nog keuzes kunnen maken ten aanzien van de ondergrens a en de bovengrens b . De keuzes van die onder- en bovengrens kunnen afgeleid worden van de onzekerheidsranges die in de studie van 1999 geïnventariseerd zijn ('integratie' van schattingen van diverse experts?).
- ii. De verdelingen bevatten naast de bereikparameters a en b eveneens twee vormparameters α , β die gebruikt kunnen worden om er voor te zorgen dat de verdeling een a priori gewenste gemiddelde en variantie (of gelijkwaardig: een spreiding) heeft.
- iii. Deze klasse van verdelingsfuncties omvat symmetrische vormen (als de vormparameters gelijk zijn: $\alpha = \beta$) maar ook scheve verdelingen. Voor die scheve verdelingen kunnen zowel vormen met 'skewed to the left' (de linkerflank is langer en minder steil dan de rechterflank, en het gemiddelde ligt dan rechts van het midden van het interval $[a, b]$) optreden (namelijk als $\alpha > \beta$), als vormen met 'skewed to the right' (de rechterflank is langer en minder steil dan de linkerflank, en het gemiddelde ligt links van het midden) optreden (namelijk als $\beta > \alpha$). In de definitie van de onzekerheid in een habitatindex als functie van een habitatfactor zal een 'skewed to the left' vorm het meest geschikt zijn voor een habitatindex die dicht bij 1 ligt, terwijl een 'skewed to the right' vorm beter geschikt lijkt voor een habitatindex die dicht bij 0 ligt.
- iv. De uniforme verdeling is als een speciaal geval onderdeel van de klasse, namelijk wanneer $\alpha = \beta = 1$.

In de praktijk komt het er op neer dat een gebruiker de range $[a, b]$ moet opgeven, alsmede de gewenste verwachtingswaarde en spreiding, waarmee dan de verdeling van de onzekerheid kan worden vastgelegd. Hierbij moeten wel (minstens) drie waarschuwingen/beperkingen gegeven worden:

- 1°. Het gemiddelde μ moet binnen de range $[a, b]$ worden opgegeven (uiteraard).
- 2°. De spreiding is bij gegeven $[a, b]$ en gemiddelde μ niet geheel vrij te kiezen, en moet lager zijn dan een bovengrens die (op een complexe manier) van die andere drie parameters afhangt. Dat er voor de spreiding een bovengrens is, volgt natuurlijk al meteen uit het eindige domein van de verdeling. Als een verdeling op een interval van

lengte L gedefinieerd is volgt bijvoorbeeld al op intuïtieve basis dat dan de spreiding niet bijvoorbeeld $10L$ kan zijn, maar op z'n hoogst van de orde $0.5L$ of lager moet zijn.

- 3^e. In principe zijn de vormparameters $\alpha, \beta > 0$. Echter voor $0 < \alpha < 1$ is de verdeling J -shaped voor $x \downarrow a$ (d.w.z. $f(x) \uparrow \infty$ voor $x \downarrow a$) terwijl voor $0 < \beta < 1$ de verdeling J -shaped is voor $x \uparrow b$ (d.w.z. $f(x) \uparrow \infty$ voor $x \uparrow b$). In woorden: de dichtheid wordt oneindig groot nabij één of beide randen als één of beide vormparameters kleiner dan 1 is. Voor het geval dat beide vormparameters dicht bij 0 liggen kan het zelfs gebeuren dat de verdeling zich volledig concentreert op (degenereert tot) de randpunten a en b , en een trekking dus òfwel de ondergrens, òfwel de bovengrens oplevert (discrete flip-flop verdeling). 'Nette' vormen met maar één en een *eindig* maximum, dat bovendien binnen het interval $[a, b]$ ligt, worden gevonden indien $\alpha, \beta \geq 1$.

Opmerking: een J -shaped vorm bij de ondergrens a zou nog toegestaan kunnen worden als $a=0$. Opmerking: een J -shaped vorm bij de bovengrens b zou nog toegestaan kunnen worden als $b=1$.

Moraal: bij een implementatie kunnen in eerste instantie een $[a, b]$ en een μ en σ worden opgegeven. De bijhorende spreiding kan dan om twee redenen niet geaccepteerd worden, namelijk:

1. De opgegeven spreiding is te groot om bijhorende α, β te kunnen vinden.
2. Er worden bijhorende α, β gevonden, maar minstens één van de bijhorende α, β is kleiner dan 1, en de J -shapeness wordt gevonden bij een $a > 0$ en/of bij een $b < 1$.

In deze gevallen moet de opgegeven spreiding verlaagd worden, of gezet op diens theoretische minimum.

C.1.1 Gebruikte verdelingsfuncties

In deze paragraaf wordt een verdelingsfunctie beschreven met de gewenste eigenschappen gedefinieerd in C.1. Het idee is om uit te gaan van een stochast X met een standaard normale verdeling en deze X met een niet-lineair voorschrift $S(\cdot)$ te transformeren naar een stochast Y die waarden aanneemt in het interval $(0, 1)$. Deze Y is een tussenvorm omdat wij 'straks' verdelingen wensen die op een door de gebruiker voorgeschreven eindig interval $[a, b]$ gedefinieerd zijn. Een stochast Z met zo'n verdeling kan dan (op zijn beurt) gemakkelijk via een lineair voorschrift uit Y geconstrueerd worden, namelijk via:

$$Z = a + (b - a) \cdot Y \quad \leftrightarrow \quad Y = \frac{Z - a}{b - a} \quad (\text{C.1})$$

Nu is dan het probleem om een geschikt niet-lineair voorschrift voor de Y als functie van X (met de standaard normale verdeling) te vinden. Hiervoor wordt het volgende voorgesteld:

$$Y = S(X | \alpha, \beta) \quad \text{met} \quad S(x | \alpha, \beta) := \frac{e^{\alpha(x-\beta)}}{1 + e^{\alpha(x-\beta)}} \quad (\text{C.2})$$

Hierin zijn α en β twee vorm-parameters die gebruikt zullen worden om er voor te zorgen dat de verwachtingswaarde μ_0 en spreiding σ_0 van Y aan voorgeschreven waarden voldoen. Merk op dat het inverse voorschrift, namelijk X als functie van Y, eveneens met een analytische expressie kan worden weergegeven, namelijk:

$$X = S^{inv}(Y | \alpha, \beta) \quad \text{met} \quad S^{inv}(y | \alpha, \beta) := \beta + \frac{1}{\alpha} \ln\left(\frac{y}{1-y}\right) \quad (\text{C.3})$$

Als we de cumulatieve verdelingsfunctie van de standaard normale verdeling weergeven met $\Phi(\cdot)$, en de bijhorende kansdichtheidsverdeling met $\phi(\cdot)$ dan is:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad \text{en} \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}\xi^2} \cdot d\xi \quad (\text{C.4})$$

Met de gegeven $S(\cdot | \alpha, \beta)$ zien dan de cumulatieve verdeling $F_0(\cdot)$ en kansdichtheidsfunctie $f_0(\cdot)$ van Y er als volgt uit:

$$\begin{aligned} F_0(y | \alpha, \beta) &= \Phi\left(S^{inv}(y | \alpha, \beta)\right) = \Phi\left(\beta + \frac{1}{\alpha} \cdot \ln\left(\frac{y}{1-y}\right)\right) \\ f_0(y | \alpha, \beta) &= \phi\left(S^{inv}(y | \alpha, \beta)\right) \cdot \frac{dS^{inv}(y | \alpha, \beta)}{dy} = \frac{1}{\alpha} \cdot \left(\frac{1}{y} + \frac{1}{1-y}\right) \cdot \phi\left(\beta + \frac{1}{\alpha} \cdot \ln\left(\frac{y}{1-y}\right)\right) \\ &= \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\alpha} \cdot \left(\frac{1}{y} + \frac{1}{1-y}\right) \cdot \exp\left(-\frac{1}{2} \cdot \left(\beta + \frac{1}{\alpha} \cdot \ln\left(\frac{y}{1-y}\right)\right)^2\right) \end{aligned} \quad (\text{C.5})$$

Aan de hand van figuren van $f_0(\cdot | \alpha, \beta)$ voor een aantal variaties van de vormparameters α en β is in een ander kader vastgesteld dat deze klasse ('schaar') van verdelingen inderdaad de gewenste (hierboven genoemde) eigenschappen heeft.

Daarnaast moet de symmetrie-eigenschap $f(\cdot | \alpha, \beta) = f(\cdot | -\alpha, -\beta)$ worden opgemerkt zodat volstaan kan worden met $\alpha \geq 0$.

Op dezelfde manier geldt voor de cumulatieve verdeling $F(\cdot)$ en kansdichtheidsfunctie $f(\cdot)$ van Z dat die er als volgt uit zien op het interval $[a, b]$:

$$F(z|a,b; \alpha, \beta) = \Phi\left(\beta + \frac{1}{\alpha} \cdot \ln\left(\frac{z-a}{b-z}\right)\right)$$

$$f(z|a,b; \alpha, \beta) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\alpha} \cdot \left(\frac{1}{z-a} + \frac{1}{b-z}\right) \cdot \exp\left(-\frac{1}{2} \cdot \left(\beta + \frac{1}{\alpha} \cdot \ln\left(\frac{z-a}{b-z}\right)\right)^2\right) \quad (\text{C.6})$$

Dit zijn de verdelingen van de grootheid Z gedefinieerd op het interval $[a, b]$ met de gewenste eigenschappen waar het ons uiteindelijk om gaat. De vormparameters a, b, α en β moeten daarbij door de gebruiker worden opgegeven. Voor de a en b (onder en bovengrens van het interval) is de betekenis duidelijk, maar van een gebruiker kan niet verwacht worden dat deze iets zinnigs over de α en β zal kunnen zeggen. Voor zo'n gebruiker is het daarentegen gemakkelijker om een verwachtingswaarde μ en spreiding σ op te geven. Het probleem is dan om te bepalen welke α en β met die μ en σ overeenkomen. Helaas kan de relatie $(\mu, \sigma) \leftrightarrow (\alpha, \beta)$ niet met analytische expressies worden uitgedrukt en moet die relatie numeriek worden bepaald. In FORTRAN zou dat zonder exceptionele inspanning gedaan kunnen worden, maar voor de PC-Raster omgeving moet hiervoor echter een tabel worden aangemaakt (buiten PC-Raster te genereren, daarna eenmalig als invoer in PC-Raster in te lezen). De vorm van deze tabel en hoe die in de PC-raster-omgeving gebruikt moet worden, wordt in de volgende secties uitgewerkt.

De $(\mu_0, \sigma_0) \leftrightarrow (\alpha, \beta)$ tabel

De tabel waarmee uit opgegeven verwachtingswaarde en spreiding de vormparameters α en β kunnen worden gelezen kan het beste worden aangemaakt voor de 'standaardvorm' Y , d.w.z. voor de verdelingen met ondergrens $a=0$ en bovengrens $b=1$. Dit bespaart twee entries (voor a en b) in zo'n tabel. De tabel heeft dan een vorm waarbij bij gegeven (μ_0, σ_0) de bijhorende (α, β) kunnen worden opgezocht. Deze zou de volgende vorm kunnen hebben

$\mu_0^{(1)} \sigma_0^{(1)}$	$\alpha^{(1)} \beta^{(1)}$	1 ^e record
$\mu_0^{(2)} \sigma_0^{(2)}$	$\alpha^{(2)} \beta^{(2)}$	2 ^e record
$\mu_0^{(3)} \sigma_0^{(3)}$	$\alpha^{(3)} \beta^{(3)}$	3 ^e record
....	
$\mu_0^{(L)} \sigma_0^{(L)}$	$\alpha^{(L)} \beta^{(L)}$	L ^e record

In bovenstaande voorbeeldtabel zijn geen records met $\mu_0=0$ en $\mu_0=1$ opgenomen omdat hierbij σ_0 noodzakelijkerwijs 0 moeten zijn. Eveneens zijn er geen records opgenomen met $\sigma_0=0$.

De gemiddelde verwachtingswaarde μ_0 moet altijd tussen 0 en 1 liggen. Ook de spreiding σ_0 moet groter of gelijk 0 zijn, maar voor deze parameter is er een (variabele) bovengrens die afhangt van de daarbij gekozen μ_0 . Zo is bij $\mu_0=0.5$ die maximale σ_0 ongeveer 0.32, terwijl dat maximum voor σ_0 kleiner is bij alle andere instellingen van μ_0 . Bij $\mu_0=0$ en $\mu_0=1$ moet de bijhorende spreiding zelfs 0 zijn. Deze bovengrens van σ_0 als functie van μ_0 moet in een tweede afzonderlijke tabel in PC-Raster opgeslagen worden (omdat bij $\mu_0=0$ en $\mu_0=1$ de bijhorende spreiding 0 moet zijn, worden voor deze uiterste μ_0 s niet in de tabel opgenomen).

In de praktijk zullen de μ_0 en σ_0 op een rooster gekozen worden waarna de bijhorende α en β (éénmalig en buiten PC-raster) berekend moeten worden. Om de omvang van de tabel te beperken dient dat rooster niet te fijn te zijn. Men zou hierbij K_μ equidistante samples kunnen kiezen zodat $\Delta\mu=1/(K_\mu-1)$. Is K_μ bijvoorbeeld 51 dan is $\Delta\mu=0.02$. Bij $K_\mu=11$ is dan $\Delta\mu=0.1$. Voor de resolutie van de σ_0 moet eveneens een geschikte $\Delta\sigma$ worden gekozen, b.v. 0.01 of 0.02.

Op het (μ_0, σ_0) -rooster geldt dan dat de samples zijn gegeven door $\mu_0(i)=(i-1)\cdot\Delta\mu$ voor $1\leq i\leq K_\mu$, en $\sigma_0(j)=(j-1)\cdot\Delta\sigma$ voor $1\leq j\leq K_\sigma$ (waarbij j eerder dan K_σ kan ophouden omdat de maximale $\sigma_0(\mu_0)$ werd overschreden). In de tabel zouden derhalve in plaats van de 'reële' waarden μ_0 en σ_0 net zo goed de indices i en j gegeven kunnen worden. Hierover moet bij de definitieve implementatie nader beslist worden door de programmeurs in overleg met gebruikers.

Daarnaast wordt opgemerkt dat de waarden $\mu_0=0$ en $\mu_0=1$ niet in de tabel hoeven te worden opgenomen aangezien bij deze μ_0 de spreiding $\sigma_0=0$ moet zijn. Alle trekkingen uit een verdeling worden dan afgebeeld op die μ_0 (dat wil zeggen een verdeling die helemaal in die μ_0 geconcentreerd is; hiervoor hoeven geen α en β te worden opgezocht). In de tabel hoeven eveneens geen records te worden opgenomen waarbij $\sigma_0=0$. Is namelijk $\sigma_0=0$ dan is de verdeling weer helemaal geconcentreerd in de bijhorende μ_0 (die daarbij niet beperkt is tot 0 en/of 1).

C.1.2 Opmerkingen bij de implementatie in PC-raster

De invoer van de verdeling van de factorindex voor gegeven breekpunten in een habitatfactor

Bij een gegeven breekpunt van een habitatfactor moeten vier parameters worden opgegeven, namelijk a , b (die het interval gegeven waarop de verdeling van Z gedefinieerd is) en μ , σ (het gemiddelde en de spreiding van de verdeling van Z). Let op: deze μ en σ zijn (nog) niet de μ_0 en σ_0 .

Op deze invoer a , b , μ , σ moeten de volgende controles plaatsvinden:

- Stap 1 a moet kleiner dan b zijn. Als dat niet zo is: foutmelding.
- Stap 2 μ moet in het interval $[a, b]$ liggen. Als dat niet zo is: foutmelding.
- Stap 3 Is $\mu=a$ dan *moet* $\sigma=0$ zijn. Als dat niet zo is: foutmelding.
- Stap 4 Is $\mu=b$ dan *moet* $\sigma=0$ zijn. Als dat niet zo is: foutmelding.
- Stap 5 Is $a < \mu < b$ dan moet gecontroleerd worden of de opgegeven σ niet te groot is. Dat gaat als volgt:
- Stap 5a Bereken μ_0 uit a , b , en μ volgens $\mu_0 = (\mu - a) / (b - a)$. Deze μ_0 ligt dan in $(0, 1)$
- Stap 5b In de tabellen zijn de μ_0 'slechts' op een rooster gedefinieerd. Zoek de index i en $\mu_0(i)$ waarvoor geldt dat die $\mu_0(i)$ het dichtste bij de μ_0 van stap 5a ligt. Deze $\mu_0(i)$ vervangt dan de μ_0 van stap 5a (m.a.w. we kiezen het roosterpunt dat het dichtste bij de μ_0 van de gebruiker ligt).
- Stap 5c Voor de i en bijhorende $\mu_0 = \mu_0(i)$ aan het einde van stap 5b moet in de **tabel met spreidingen (C1.1)** de voor deze μ_0 horende maximale spreiding $\sigma_{0, \max}$ worden opgezocht.
- Stap 5d Bereken σ_0 van de gebruiker als volgt uit a , b en σ : $\sigma_0 = \sigma / (b - a)$.
- Stap 5e Ook de σ_0 zijn (in de tabellen) op een rooster gedefinieerd. Zoek de index j en $\sigma_0(j)$ waarvoor geldt dat die $\sigma_0(j)$ het dichtste bij de σ_0 van stap 5d ligt. Deze $\sigma_0(j)$ vervangt dan de σ_0 van stap 5d. (met andere woorden we kiezen het roosterpunt dat het dichtste bij de σ_0 van de gebruiker ligt).
- Stap 5f Is de σ_0 van stap 5e kleiner/gelijk aan de $\sigma_{0, \max}$ van stap 5d dan akkoord en hebben we voor het huidige breekpunt een acceptabele combinatie (a, b, μ_0, σ_0) (met die μ_0 , σ_0 op het rooster, en mogelijk iets afwijkend van wat de gebruiker strikt heeft opgegeven). In het andere geval: foutmelding en abort van programma.
- Stap 6 Als we nog 'in business' zijn dan moeten we nu met de gevonden μ_0 en σ_0 in de voorbeeldtabel van C1.1 de bijhorende parameters α en β opzoeken. Deze zullen we later (in de Monte Carlo procedure) nog vaak nodig hebben dus indien mogelijk deze opslaan (anders moet die telkens volgens voorafgaande procedure opnieuw berekend worden).

Samenvatting. Bovenstaande moet voor elk breekpunt van alle habitatfactor-habitatindex relaties worden uitgevoerd. Deze procedure omvat het controleren van de invoer, en het voor elk breekpunt berekenen/opzoeken van de (α, β) -parameters.

C.1.3 Trekkingen uit de verdelingen factorindex voor gegeven breekpunten

Hier wordt beschreven hoe een trekking uit een standaard normale verdeling moet worden vertaald naar de trekking uit de verdeling van een habitatindex. Hierin worden de $\{a, b, \alpha, \beta\}$ parameters gebruikt die de verdeling van de factorindex (bij gegeven breekpunt(en) van een habitatfactor) vastleggen. De α en β zijn zoals hierboven is beschreven afgeleid uit de invoercombinatie $\{a, b, \mu, \sigma\}$.

- Stap 1 Bepaal (in PC-raster) een trekking x uit de standaard normale verdeling
- Stap 2 Konverteer x naar y in $(0, 1)$ volgens

$$y = S(x | \alpha, \beta) = \frac{e^{\alpha(x-\beta)}}{1 + e^{\alpha(x-\beta)}}$$

Stap 3 Converteer y naar z volgens:

$$z = a + (b - a) \cdot y$$

Dit is de gewenste conversie. Bij één bepaalde habitatfactor-habitatindex relatie (in de vorm van enkele breekpunten, met lineaire interpolatie daartussen) wordt per breekpunt met dezelfde trekking x gewerkt. De vertaling $x \rightarrow z$ moet echter per breekpunt worden uitgevoerd, met de daar gegeven $\{a, b, \mu, \sigma\}$ of daarmee equivalent $\{a, b, \alpha, \beta\}$.

Opmerking 1: bovenstaande recept geldt alleen indien μ strikt tussen a en b ligt. Is $\mu=a$ dan is per definitie $\sigma=0$ en moet de trekking z altijd op a worden gezet. Zo ook: Is $\mu=b$ dan is per definitie $\sigma=0$ en moet de trekking z direct op b worden gezet.

Opmerking 2: Is $\sigma=0$ dan hoeft bovenstaande recept eveneens niet te worden uitgevoerd, en kan de trekking z direct op μ worden gezet.

In de situatie van bovenstaande twee opmerkingen is de verdeling geconcentreerd in één punt, namelijk $z=\mu$. In feite hoeft daar dan geen x voor getrokken te worden.

C.1.4 Quartielpunten van uitvoerresultaten

In de Monte Carlo (MC) simulatie wordt per trekking voor elke cel één (of meer?) uitvoergrootheid berekend en is in principe een ensemble beschikbaar. Het lijkt erop dat alle realisaties bewaren teveel opslag vergt en het voordeliger is om specifieke kentallen te bewaren. Aansluitend op de klasse van verdelingen die voor de habitatfactor-habitatindex relatie wordt gebruikt zouden deze kentallen als volgt kunnen zijn: gemiddelde, spreiding, ondergrens, bovengrens.

Bepaling van het gemiddelde betekent dat tijdens de MC de opvolgende uitkomsten z_ℓ van een uitvoer grootheid worden opgeteld, en na de MC gedeeld wordt door het totale aantal L van MC-simulaties. Voor de spreiding moeten de kwadraten van z_ℓ worden opgeteld, en na de MC vertaald naar een spreiding. Bij de l^e run moet het minimum en maximum geüpdatet worden. Resumerend:

$$\mu = \frac{1}{L} \sum_{\ell=1}^L z_\ell$$

$$\sigma = \sqrt{\frac{1}{L} \sum_{\ell=1}^L z_\ell^2 - \mu^2}$$

$$a = \text{Min} \{z_\ell\}_{\ell=1}^L$$

$$b = \text{Max} \{z_\ell\}_{\ell=1}^L$$

Voor het fitten van de hier gedefinieerde klasse van verdelingen moeten weer volgens het recept van paragraaf C.1.2 een α en β worden bepaald.

Dan is de vraag hoe vervolgens een kwartiel punt te bepalen, dat wil zeggen gegeven een (kansniveau) p in $(0, 1)$ welke $z=z_p$ hoort daarbij?

Met de a , b , μ , σ en de daaruit afgeleide α en β , en met $\Phi(\cdot)$ de cumulatieve verdelingsfunctie van de standaard normale verdeling, gaat dat als volgt.

Stap 1 Zoek x zodanig dat $\Phi(x) = p$, ofwel $x = \Phi^{inv}(p)$

Stap 2 Vertaal deze x naar een y volgens:

$$y = S(x | \alpha, \beta) = \frac{e^{\alpha \cdot (x - \beta)}}{1 + e^{\alpha \cdot (x - \beta)}}$$

Stap 3 Konverteer y naar z volgens:

$$z = a + (b - a) \cdot y$$

Dit geeft de gewenste z die bij niveau p hoort.

Een probleem is stap 1, namelijk de inverse van de cumulatieve verdelingsfunctie van de standaard normale verdeling. Hiervoor moeten we dan weer een derde tabel aanmaken waarin op p gezocht moet worden om de bijhorende x te vinden. Voor de p moeten we weer een rooster op $(0, 1)$ maken. Dat zouden we kunnen doen met (bijvoorbeeld) $\Delta p = 0.025$ (2.5%) voor $p = 0.025$ tot en met $p = 0.975$. Die tabel omvat dan 39 (p, x)-records en is dus beperkt van omvang.