

# Stuurfactoren voor de ecologische kwaliteit van regionaal oppervlaktewater

Een statistische analyse met  
regressiebomen voor

de Ex-ante evaluatie KRW

Achtergrondstudies



# **Stuurfactoren voor de ecologische kwaliteit van regionaal oppervlaktewater**

Een statistische analyse met regressiebomen voor de Ex-ante evaluatie KRW



PBL-rapport 500140002/2008

# Stuurfactoren voor de ecologische kwaliteit van regionaal oppervlaktewater

Een statistische analyse met regressiebomen voor de Ex-ante evaluatie KRW

**Stuurfactoren voor de ecologische kwaliteit van regionaal oppervlaktewater**  
Een statistische analyse met regressiebomen voor de Ex-ante evaluatie KRW

© Planbureau voor de Leefomgeving (PBL), Bilthoven, september 2008  
PBL-rapport 500140002/2008

H. Visser  
P.J.T.M. van Puijenbroek  
P.H.M. Janssen

*Contact*

H. Visser (IMP/PBL)  
Hans.Visser@pbl.nl

U kunt de publicatie downloaden van de website [www.pbl.nl](http://www.pbl.nl) of opvragen via [reports@mnp.nl](mailto:reports@mnp.nl) onder vermelding van het PBL-publicatienummer.

Delen uit deze publicatie mogen worden overgenomen op voorwaarde van bronvermelding:  
'Planbureau voor de Leefomgeving, de titel van de publicatie en het jaartal.'

Planbureau voor de Leefomgeving  
Postbus 303  
3720 AH Bilthoven  
T: 030 274 274 5  
F: 030 274 4479  
E: [info@pbl.nl](mailto:info@pbl.nl)  
[www.pbl.nl](http://www.pbl.nl)

# Rapport in het kort

## Stuurfactoren voor de ecologische kwaliteit van regionaal oppervlaktewater

Binnen de Kaderrichtlijn Water (KRW) wordt de ecologische kwaliteit van regionaal oppervlaktewater uitgedrukt in de vorm van ratio's die liggen tussen 0 en 1 (van een zeer slechte naar een zeer goede ecologische toestand). Deze ecologische kwaliteitsratio's (EKR's) zijn beschikbaar voor grotere waterorganismen (macrofauna), vissen, grotere waterplanten (macrofyten), bodemplanten (fytobenthos) en algen (fytoplankton). Voor het beleid is het belangrijk te weten hoe EKR-waarden afhangen van stuurfactoren (concentraties totaal fosfor, totaal stikstof, zuurstof, meandering, peildynamiek, beschaduwing, oever-inrichting, verstuwing, onderhoud). Recent is de samenhang tussen EKR's en stuurvariabelen door Royal Haskoning bepaald via neurale netwerken, dit in de Ex-ante evaluatie Kaderrichtlijn Water. Bij toepassing van neurale netwerken wordt een netwerk van 'neuronen' geprogrammeerd. Het model wordt vervolgens getraind aan de hand van de beschikbare data. Met deze methode zijn de watertypen beken, meren, kanalen en sloten onderzocht.

In het kader van de Ex-ante evaluatie Kaderrichtlijn Water is ook onderhavige studie uitgevoerd, parallel aan die met neurale netwerken. Doel is de voor de KRW verkregen resultaten via neurale netwerken te verifiëren door een heranalyse van dezelfde data met een heel andere techniek, namelijk die van *regressieboomanalyse*. Dit is een statistische techniek waarbij de variaties in EKR-waarden verklaard worden door splitsingen aan te brengen in de verschillende stuurvariabelen. De boomstructuur ('regressieboom') die door deze splitsingen ontstaat, sluit goed aan op de beleidspraktijk door het type kennisregels dat ontstaat en de hiërarchische structuur daarin. Dit type kennisregels wordt niet verkregen via neurale netwerken.

Er blijkt een grote mate van overeenkomst in resultaten te bestaan tussen beide technieken bij het voorspellen van specifieke EKR-waarden uit gegeven waarden voor de stuurvariabelen. Daarbij ligt de voorspelkracht hoog tot zeer hoog, tussen de 49 en 97%. De hoge mate van overeenkomst in voorspelresultaten geeft een onafhankelijke verificatie van de gepresenteerde berekeningen met neurale netwerken in de Ex-ante evaluatie KRW. Daarnaast is via de uitgevoerde regressieboomanalyse als extra informatie gevonden dat een goede ecologische toestand in kanalen en meren voor een belangrijk deel bepaald wordt door 'totaal fosfor' en 'totaal stikstof'. Voor beken zijn vooral een goede *meandering* en een laag biologisch zuurstofverbruik van belang. Voor sloten speelt peilbeheer vermoedelijk een belangrijke rol.

Trefwoorden: EKR, hydromorfologie, Kaderrichtlijn Water, neurale netwerken, nutriënten, oppervlaktewater, waterkwaliteit





## Abstract

### Steering factors regarding the ecological quality of regional surface waters

In the Water Framework Directive (WFD), the ecological quality of regional surface waters is expressed in terms of ratios which lie between 0 and 1 (from very poor to excellent ecological quality). These ecological quality ratios (EQRs) are available for the following groups of species: macrofauna, fish, macrophytes, phytobenthos and phytoplankton. With regard to policy, it is important to know the way in which EQR values depend on steering factors. These steering factors are partly of a physical-chemical nature (concentrations of the total of phosphorus, the total of nitrogen, oxygen) and partly hydromorphological (meandering, water level dynamics, shading, river bank design, damming, maintenance). Recently, Royal Haskoning has established the relationship between EQRs and steering variables, using neural networks. The research regarded the following water types: streams, lakes, canals and ditches.

This study has been carried out within the scope of the Ex ante evaluation of the Water Framework Directive, parallel to the study that uses neural networks. Its aim is to verify the results that were obtained through neural networks, by re-analysing the same data using a completely different technique, that of *regression tree analysis*. This is a statistical technique, whereby the variations in EQR values are explained by splitting the various steering variables. The 'regression tree' which is formed by this splitting, fits in well with policy practice because of the type of knowledge rules that is thus created and its hierarchical structure. This type of knowledge rules cannot be obtained through neural networks.

Both techniques are found to be very similar in predicting specific EQR values from given values for steering variables. Moreover, their prediction strength appears to be high, between 49 and 97 percent. This large similarity in predictions provides an independent verification of the presented calculations, obtained by using neural networks, in the Ex ante Evaluation of the WFD. Furthermore, the regression tree analysis has shown that a good ecological state of the various water types is largely determined by the 'total of phosphorus' and the 'total of nitrogen' (a limit value of 0.12 mg P/l for phosphorus and 1.7 mg N/l for nitrogen). Having a low BOD value (Biological Oxygen Demand) and a good meandering appear to be important in streams/brooks, while water level management is probably important in ditches.

Key words: EQR, hydromorphology, Water Framework Directive, neural networks, nutrients, surface water, water quality



# Inhoud

Samenvatting	11
1 Inleiding	13
1.1 Kaderrichtlijn Water	13
1.2 Statistische analyse van EKR's en stuurvariabelen	14
1.3 Leeswijzer	16
2 Data	19
2.1 Databases	19
2.2 Representativiteit	19
3 Statistiek	23
3.1 Presentatiemethoden	23
3.1.1 Correlatiematrices, scatterplotmatrices	23
3.1.2 Histogrammen en dichtheden	24
3.1.3 Boxplots	25
3.2 Regressieboomanalyse	26
3.2.1 Uitleg van de methode	26
3.2.2 Onzekerheden	29
3.2.3 Regressieboomanalyse in de context van de KRW	32
3.3 Software	34
4 Actuele EKR's en hun onderlinge relaties	35
4.1 Beken	35
4.2 Meren	39
4.3 Kanalen	43
4.4 Sloten	47
5 Resultaten regressieboomanalyse	51
5.1 Beken	51
5.2 Meren	52
5.3 Kanalen	53
5.4 Sloten	54
6 Vergelijking regressiebomen en neurale netwerken	57
6.1 Methodische overeenkomsten en verschillen	57
6.2 Voorspelkracht vergeleken	57
6.3 Doorrekenen van maatregelen	58
6.4 Conclusie	60
7 Discussie, conclusies, aanbevelingen	63
Dankwoord	67
Literatuur	69

Appendix A	Regressiebomen voor beken	71
Appendix B	Regressiebomen voor meren	91
Appendix C	Regressiebomen voor kanalen	111
Appendix D	Regressiebomen voor sloten	131
Appendix E	Verklarende kracht van de regressiebomen	147
Appendix F	Gevoeligheidsanalyse doublures	149

## Samenvatting

Binnen de Kaderrichtlijn Water (KRW) speelt de ecologische kwaliteit van oppervlaktewater een belangrijke rol. Deze kwaliteit wordt meetbaar gemaakt door voor waterplanten en waterdieren maatlaten te ontwikkelen en te schalen naar een ratio tussen 0 en 1. Hierbij staat 0 voor een zeer slechte ecologische toestand en 1 voor een zeer goede ecologische toestand. Ecologische kwaliteitsratio's (EKR's) worden routinematig bepaald voor grotere waterorganismen (macrofauna), vissen, grotere waterplanten (macrofyten), bodemplanten (fytobenthos) en algen (fytoplankton). EKR's zijn beschikbaar voor verschillende watertypen: beken, meren, kanalen en sloten.

Een belangrijke vraag voor het beleid is hoe EKR-waarden samenhangen met factoren/variabelen die deze EKR's in positieve zin kunnen bijsturen. Zulke stuurvariabelen kunnen deels fysisch-chemisch van aard zijn (concentraties totaal fosfor, totaal stikstof, biologisch zuurstofverbruik) en deels hydromorfologisch van aard zijn (meandering, peildynamiek, beschaduwing, oeverinrichting, verstuwings, onderhoud). Als de samenhang tussen EKR-waarden en stuurvariabelen bekend is, kan geëvalueerd worden welke maatregelen het effectiefst zijn om de gewenste ecologische kwaliteit te halen.

Recent is voor de Ex-ante evaluatie Kaderrichtlijn Water de samenhang tussen EKR's en stuurvariabelen bepaald door Royal Haskoning. Deze samenhang is bepaald via neurale netwerken. Dit is een wiskundige methode die de werking van de hersenen nabootst door een netwerk van 'neuronen' te programmeren en te trainen met beschikbare data. Met deze methode zijn de watertypen beken, meren, kanalen en sloten onderzocht. Vervolgens zijn voorgestelde maatregelen van waterschappen doorgerekend op haalbaarheid en doelbereik in het jaar 2027.

In het kader van de Ex-ante evaluatie Kaderrichtlijn Water is ook onderhavige studie uitgevoerd, parallel aan die met neurale netwerken. Doel is de resultaten, die voor de KRW verkregen zijn via neurale netwerken, te verifiëren door een heranalyse van dezelfde data met een heel andere techniek, namelijk die van *regressieboomanalyse*. Hiertoe is gebruikgemaakt van de databases, opgesteld door Royal Haskoning. De databases geven voor de watertypen beken, meren, kanalen en sloten een viertal EKR-deelmaatlatwaarden die steeds gelden voor *dezelfde* locatie, aangevuld met de relevante stuurvariabelen.

Regressieboomanalyse is een statistische techniek waarbij de variaties in EKR-waarden verklaard worden door splitsingen aan te brengen in de onderscheiden stuurvariabelen: als een stuurvariabele gebracht kan worden naar waarden onder (of boven) een drempel 'x', zal de EKR-waarde stijgen (of dalen) naar de waarde 'y'. Regressieboomanalyse is daarmee een vorm van modellering die goed aansluit op de beleidspraktijk: de splitsingen hebben een hiërarchisch karakter, van meest naar minst belangrijk, en zijn direct te vertalen naar grenswaarden voor de verschillende stuurfactoren.

Regressiebomen zijn geschat voor een 15-tal deelmaatlatwaarden, verspreid over de vier genoemde watertypen. De verklarende kracht van de regressiebomen is bepaald via kruisvalidatie en blijkt hoog tot zeer hoog te liggen: tussen de 49% en 97%. De voorspelfouten uit de kruisvalidatie zijn daarna vergeleken met die verkregen via *neurale netwerken*, toegepast op dezelfde vier datasets. Er blijkt een zeer goede overeenkomst te zijn tussen de resultaten van beide methoden. Daarbij is gevonden dat de voorspelfouten via regressieboomanalyse gelijk of zelfs kleiner zijn dan die verkregen via neurale netwerken. De resultaten geven een onafhanke-

lijke verificatie van de resultaten via neurale netwerken zoals gepresenteerd in de Ex-ante evaluatie KRW.

In tegenstelling tot neurale netwerken geven regressiebomen in detail aan hoe een EKR-voorspelling volgt uit waarden van de stuurvariabelen. Gevonden is dat een goede tot zeer goede ecologische toestand sterk bepaald wordt door de stuurvariabelen ‘totaal fosfor’ en ‘totaal stikstof’. Alle regionale wateren en deelmaatlatten samenvattend wordt gevonden dat de concentratie totaal fosfor lager moet zijn dan 0.12 mg P/l voor een goede tot zeer goede ecologische toestand (alle EKR's hoger dan 0.60). Idem voor totaal stikstof een concentratie lager dan 1.7 mg N/l. De hier gevonden grenswaarden liggen iets onder de MTR-waarden voor totaal fosfor en totaal stikstof (Vierde Nota Waterhuishouding). De MTR-waarde voor oppervlaktewater ligt voor totaal fosfor op 0.15 mg P/l en voor totaal stikstof op 2.2 mg N/l.

Voor beken geldt daarnaast dat een goede *meandering* van groot belang is (klasse 3 en hoger, ofwel zwak tot vrij meanderend), met daarbij een biologisch zuurstofverbruik (BZV) lager dan 4.3 mg O<sub>2</sub>/l, en een *verstuwings* in de klasse 2 of 3 (=F matig gestuwd tot ongestuwd). Voor sloten speelt *peilbeheer* een rol van betekenis: peilbeheer in klasse 3 (= natuurlijk peilbeheer) is noodzakelijk voor het halen van een goed ecologisch potentieel. Omdat het aantal verschillende sloten gering is (35), is deze conclusie indicatief. Voor meren en kanalen zijn naast totaal fosfor en totaal stikstof geen andere stuurvariabelen van belang. Kanttekening bij deze resultaten is dat ze gebaseerd zijn op de stuurvariabelen uit dit onderzoek. Niet alle mogelijke hydromorfologische maatregelen zijn opgenomen in de gebruikte databases.

De gevonden resultaten kunnen een belangrijke input zijn bij de verdere ontwikkeling van de KRW-Verkenner, een kennissysteem dat waterbeheerders ondersteunt bij het opstellen van stroomgebiedbeheerplannen en het vinden van optimale maatregelpakketten.

# I Inleiding

## I.1 Kaderrichtlijn Water

De Kaderrichtlijn Water (KRW) richt zich op de bescherming van alle Europese wateren en kent een ambitieuze taakstelling. Zo staat er in de KRW-richtlijn 2000/60/EG, dd. 23 oktober 2000 (KRW, 2000): ‘Water is geen gewone handelswaar, maar een erfgoed dat als zodanig beschermd, verdedigd en behandeld moet worden’. Daarmee heeft de KRW een grote reikwijdte. Zowel de bescherming van landoppervlaktewater, overgangswater, kustwater en grondwater als ook daarvan afhankelijke ecosystemen vallen eronder. De KRW beslaat feitelijk het gehele watersysteem, van bron tot zee en van zoet tot zout, en geeft daarvoor een samenhangend en Europees uniform wettelijk kader voor de bescherming en beheer (PBL, 2008; <http://kaderrichtlijnfilms.natuurmedia.nl>).

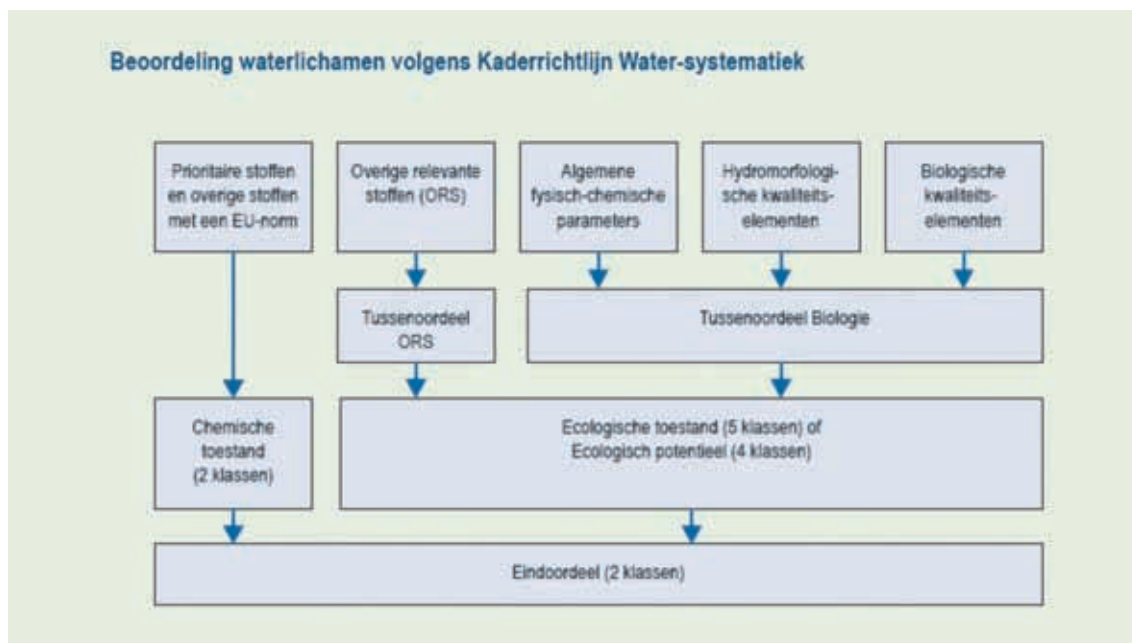
Binnen de KRW wordt de kwaliteit van het oppervlaktewater niet alleen afgemeten aan de chemische kwaliteit (onder andere prioritair stoffen, totaal fosfor, totaal stikstof) en de morfologie van het systeem (natuurlijke versus kunstmatige wateren zoals sloten, meandering, et cetera) maar ook aan de *ecologische kwaliteit*. Met het laatste wordt bedoeld hoe een aantal soortgroepen ervoor staat: grotere waterorganismen (‘macrofauna’), vissen, algen (‘fytoplankton’), bodemplanten (‘fytobenthos’) en grotere waterplanten (‘macrofyten’).

De algemene beoordeling van wateren volgens de KRW-systematiek is vrij ingewikkeld en samengevat in Figuur 1.1. De toestand van de bovengenoemde soortgroepen is gegeven door het blokje rechtsboven: ‘biologische kwaliteitselementen’. Door middel van additionele informatie over prioritair en overige relevante stoffen, nutriënten, zuurstofgehalten en het omgevingskarakter van het water (meandering, stroomsnelheden, peilbeheer, et cetera) wordt uiteindelijk een eindoordeel geformuleerd. Zie verder Van der Molen en Pot (2007) en PBL (2008).

Binnen deze complexe systematiek wordt in dit rapport alleen ingegaan op de biologische kwaliteitselementen, in relatie tot de algemene fysisch-chemische parameters en de hydromorfologische kwaliteitselementen. De biologische kwaliteitskenmerken zijn daarbij vertaald naar zogenaamde *deelmaatlat*ten voor leefgemeenschappen ‘macrofauna’, ‘vissen’, ‘fytoplankton’, ‘fytobenthos’, en ‘macrofyten’.

Het is een belangrijke vraag hoe genoemde biologische kwaliteitselementen voor regionale wateren verklaard kunnen worden uit fysisch-chemische en hydromorfologische stuurfactoren. Daartoe zullen directe relaties afgeleid moeten worden tussen deelmaatlatwaarden enerzijds en stuurvariabelen anderzijds. Stuurvariabelen zijn totaal fosfor, totaal stikstof, biologisch zuurstofverbruik (BZV), meandering, oeverbegroeiing, beschaduwing, peildynamiek en onderhoud.

Recent is door Royal Haskoning onderzocht hoe deze vraag beantwoord kan worden met de methode van *neurale netwerken* (Knoben et al., 2008). Hun resultaten zijn verwerkt in de Ex-ante evaluatie KRW (PBL, 2008). De neurale-netwerken-techniek is een wiskundige methode die gelijkenis vertoont met de werking van onze hersenen. Het netwerk bestaat uit een aantal lagen met knooppunten (‘neuronen’) die elk een eigen wiskundige overdrachtsfunctie hebben. De verbindingen tussen neuronnen hebben gewichten die door het aanbieden van data voortdurend aangepast worden. Zie verder [http://nl.wikipedia.org/wiki/Neuraal\\_netwerk](http://nl.wikipedia.org/wiki/Neuraal_netwerk) en Knoben et al. (2008, bijlage 1).



**Figuur 1.1** Overzicht beoordeling waterlichamen volgens de Europese KRW-systematiek. Het onderzoek in dit rapport richt zich op het blok rechtsboven: 'biologische kwaliteits-elementen'. Bron: PBL (2008, Figuur 2.1).

In opdracht van de projectleiding van de Ex-ante evaluatie KRW is parallel aan de studie met neurale netwerken onderzocht of de resultaten uit Knoben (2008) geverifieerd kunnen worden met een heel andere methode, die van regressieboomanalyse. Dit rapport geeft de invulling van deze opdracht.

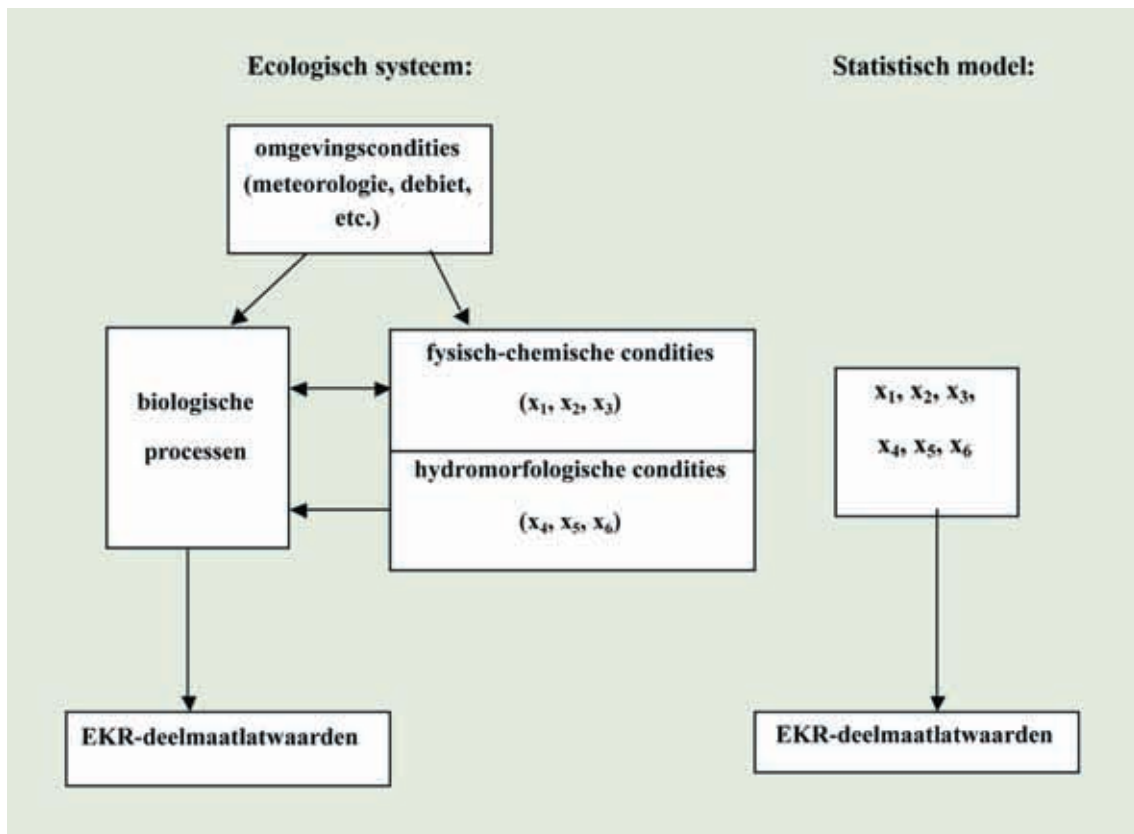
## 1.2 Statistische analyse van EKR's en stuurvariabelen

De biologische processen in een water zijn schematisch weergegeven in een stroomdiagram: het ecologisch systeem in het linkerdeel van Figuur 1.2. Het schema laat zien dat biologische processen beïnvloed worden door omgevingscondities zoals watertemperatuur en debiet, door fysisch-chemische condities, zoals de concentraties totaal fosfor en totaal stikstof, en door hydromorfologische condities, zoals meandering.

De biologische toestand wordt meetbaar gemaakt door het formuleren van zogenaamde EKR-deelmaatlaten. Hierbij staat EKR voor Ecologische KwaliteitsRatio. De EKR levert een kwaliteitsindruk in de vorm van een getal tussen 0.0 en 1.0. Vervolgens is deze ratio weer opgedeeld in vijf klassen: klasse [0.0 – 0.2] duidt op een slechte kwaliteit, klasse [0.2 – 0.4] op een ontoereikende kwaliteit, klasse [0.4 – 0.6] op een matige kwaliteit, klasse [0.6 – 0.8] op een goede kwaliteit, en klasse [0.8 – 1.0] op een zeer goede kwaliteit. Zie Figuur 1.3 voor definities.

Zo zijn er deelmaatlaten met bijbehorende EKR's afgeleid voor macrofauna, vissen, macrofyten, fyto bentos en macrofyten. Deze maatlaten zijn ook weer verschillend voor een groot aantal watertypen. Voor gedetailleerde beschrijvingen wordt verwezen naar de STOWA-rapporten van Van der Molen en Pot (2007) en Evers et al. (2007).





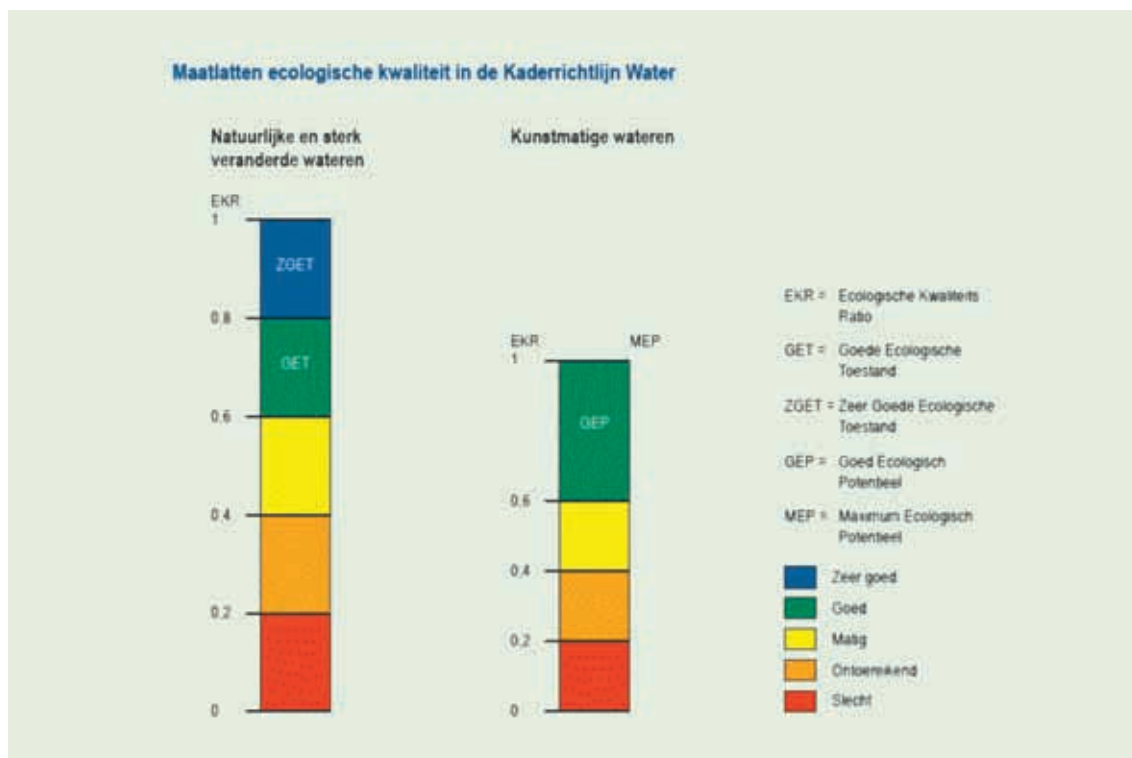
**Figuur 1.2** Stroomdiagram voor de relatie tussen EKR-deelmaatlatwaarden enerzijds en biologische processen en omgevingscondities anderzijds (links). In het statistische (meta)model (rechts) worden deze ingewikkelde interacties ‘kortgesloten’ door met regressieboonanalyse een rechtstreeks verband te leggen tussen een EKR-deelmaatlatwaarde enerzijds en de bijbehorende stuurvariabelen  $x_1$  tot  $x_6$  anderzijds.

In PBL (2008) en Knobens et al. (2008) is de biologie uit het linkerdeel van Figuur 1.2 gemodelleerd met neurale netwerken. In onderhavig rapport wordt de biologische toestand beschreven en geverifieerd met een relatief simpel statistisch model. Dit model geeft een direct verband tussen de morfologische en chemische stuurvariabelen  $x_1$  tot en met  $x_6$ , zonder aandacht te besteden aan de ingewikkelde biologisch-chemische-hydromorfologische interacties. Zo’n model wordt ook wel een *metamodel* genoemd. In metamodellen worden ingewikkelde (biologische) relaties vereenvoudigd door statistische verbanden te schatten tussen inputs van het ingewikkelde model en de output. Zie verder Janssen et al. (2005).

Het voordeel van een statistisch model is dat:

- er relaties (kennisregels) worden gevonden tussen direct meetbare grootheden in de natuur;
- de gevolgen van maatregelen (de inputs van het model) snel en eenvoudig kunnen worden doorgerekend wat betreft ecologische kwaliteitsverbetering van de verschillende watertypen;
- de voorspelkracht van het model getest kan worden middels validatie.

Een model dat dichter aansluit bij een modellering van het ecologische systeem is de KRW-Verkenner. De KRW-Verkenner is een kruising tussen een model dat de biologische processen beschrijft (linker stroomschema) en statistische relaties (rechter stroomschema). Op de KRW-Verkenner wordt teruggekomen in paragraaf 3.2.3 en Figuur 3.8.



**Figuur 1.3** Klassenindeling en kleurencode voor natuurlijke en sterk veranderde wateren (linkerbalk) en kunstmatige wateren (rechterbalk). Deze kleurencode is in dit rapport toegepast voor beken en meren (linkerbalk), en kanalen en sloten (rechterbalk). Bron: PBL (2008, Figuur 2.5).

De statistische methode die in dit rapport gehanteerd wordt, is die van regressieboomanalyse. Dit is een techniek waarbij de relatie wordt geschat tussen EKR-waarden enerzijds (de output van het model) en stuurvariabelen anderzijds (de inputs van het model). De methode is aantrekkelijk omdat het om kan gaan met sterk niet-lineaire verbanden en verder goed aansluit bij de beleidspraktijk: een regressieboom vindt kennisregels in de vorm van

*als stuurvariabele  $x_i$  onder (of boven) een drempel 'd' wordt gebracht, gaat de gemiddelde ecologische kwaliteit naar een niveau 'y'.*

Bovendien geeft de regressieboom aan met welke bijstellingen de grootste kwaliteitswinst gemaakt kan worden.

### 1.3 Leeswijzer

In hoofdstuk 2 worden de in dit rapport toegepaste datasets beschreven. De algemene statistische technieken en de werking van regressieboomanalyse worden uitgelegd in hoofdstuk 3. Vervolgens wordt in hoofdstuk 4 een overzicht gegeven van de actuele EKR-waarden voor de vier behandelde watertypen: beken (paragraaf 4.1), meren (4.2), kanalen (4.3) en sloten (4.4). Tevens wordt hier ingegaan op de verbanden (correlaties) tussen de vier deelmaatlatten per watertype.

In hoofdstuk 5 worden de resultaten van de regressieboomanalyses gegeven per watertype. Hier worden ook de bomen geëvalueerd qua voorspelkracht (validatie). Daarna worden in

hoofdstuk 6 deze resultaten vergeleken met de methode die parallel is toegepast door Royal Haskoning: neurale netwerken. Het rapport wordt afgesloten in hoofdstuk 7 met de conclusies.

In Appendix A tot en met D worden de gedetailleerde regressieboom-resultaten gegeven voor respectievelijk beken, meren, kanalen en sloten. Voor beken zijn dat bijvoorbeeld de regressiebomen voor de deelmaatlaten 'macrofauna' (paragraaf A.1), 'vissen' (A.2), 'macrofyten' (A.3) en 'fyto benthos' (A.4). In Appendix E wordt een drietal kengetallen gepresenteerd die maatgevend zijn voor de voorspelkracht van de geschatte regressiebomen (kruisvalidatie-techniek). Ten slotte wordt in Appendix F kort beschreven hoe de regressieboomanalyse uitvalt wanneer de vier databases zodanig verkleind worden dat alle doublures wegvallen.



## 2 Data

### 2.1 Databases

Voor vier watertypen zijn ecologische kwaliteitsratios's (EKR's, deelmaatlatwaarden) verzameld door Royal Haskoning. Er zijn daarbij alleen locaties geselecteerd waarvoor drie of vier deelmaatlatwaarden *gelijktijdig* beschikbaar zijn:

- voor beken de deelmaatlatwaarden 'macrofauna', 'vissen', 'macrofyten' en 'fytobenthos';
- voor meren en kanalen de deelmaatlatwaarden 'macrofauna', 'vissen', 'macrofyten' en 'fytoplankton';
- voor sloten de deelmaatlatwaarden 'macrofauna', 'vissen' en 'macrofyten'.

De watertypen met daarbij de corresponderende KRW-watertypen zijn gegeven in Tabel 2.1.

Daarnaast is als eis gesteld dat relevante stuurvariabelen voorhanden moesten zijn. Hydro-morfologische stuurvariabelen zijn: meandering, verstuwning, oeverinrichting, peilbeheer en onderhoud. Chemische parameters zijn: biologisch zuurstofverbruik (BZV), totaal fosfor en totaal stikstof. Zie Tabel 2.2 voor een samenvatting per watertype.

De methodiek om deelmaatlatwaarden te berekenen, is overgenomen uit Van der Molen en Pot (2007) en Evers et al. (2007). Zie verder ook Knobben et al. (2008, hoofdstuk 2) voor details.

### 2.2 Representativiteit

#### Representativiteit gekozen wateren

De ligging van de beken, meren, kanalen en sloten is gegeven in Figuur 2.1. Het aantal locaties per watertype is veel groter dan getoond in de figuur: voor een aantal locaties zijn data gebruikt in verschillende jaren of data op verschillende locaties langs hetzelfde water. Als locaties met dezelfde coördinaten in de vier databases samengenomen worden, dan loopt het aantal beekdata terug van 344 naar 80, het aantal meerdata van 248 naar 77, het aantal kanaaldata van 182 naar 97, en het aantal slootgegevens van 149 naar 35. De invloed van deze doublures op schattingsresultaten is onderzocht in Appendix F.

**Tabel 2.1 Clustering van watertypen. Definitie van de R- en M-typen in de rechter kolom is gegeven in Van der Molen en Pot (2007).**

Watertypencluster	Omschrijving	Bijbehorende KRW-typen
Beken	Alle permanente stromende wateren behalve bronnen en de grote rivieren.	R3, R4, R5, R6, R9, R11, R12, R13, R14, R15, R17 en R18
Zoete meren	Gebufferde stilstaande zoete vlakvormige wateren. Lijnvormige wateren, vennen en brakke wateren vallen dus niet in deze categorie.	M11, M14, M20, M21, M22, M25 en M27
Kanalen en vaarten	Zoete gegraven lijnvormige wateren breder dan 8 meter.	M3, M4, M6a/b, M7a/b en M10
Sloten	Zoete gegraven lijnvormige wateren smaller dan 8 meter. Alleen permanent watervoerende sloten worden meegenomen. Ook zeer zwak brakke sloten (<1000 mg Cl/l) vallen in deze categorie	M1a/b, M2 en M8

Bron: Knobben et al. (2008).

**Tabel 2.2 Stuurvariabelen voor verschillende typen wateren, uitgesplitst in hydromorfologische en chemische parameters.**

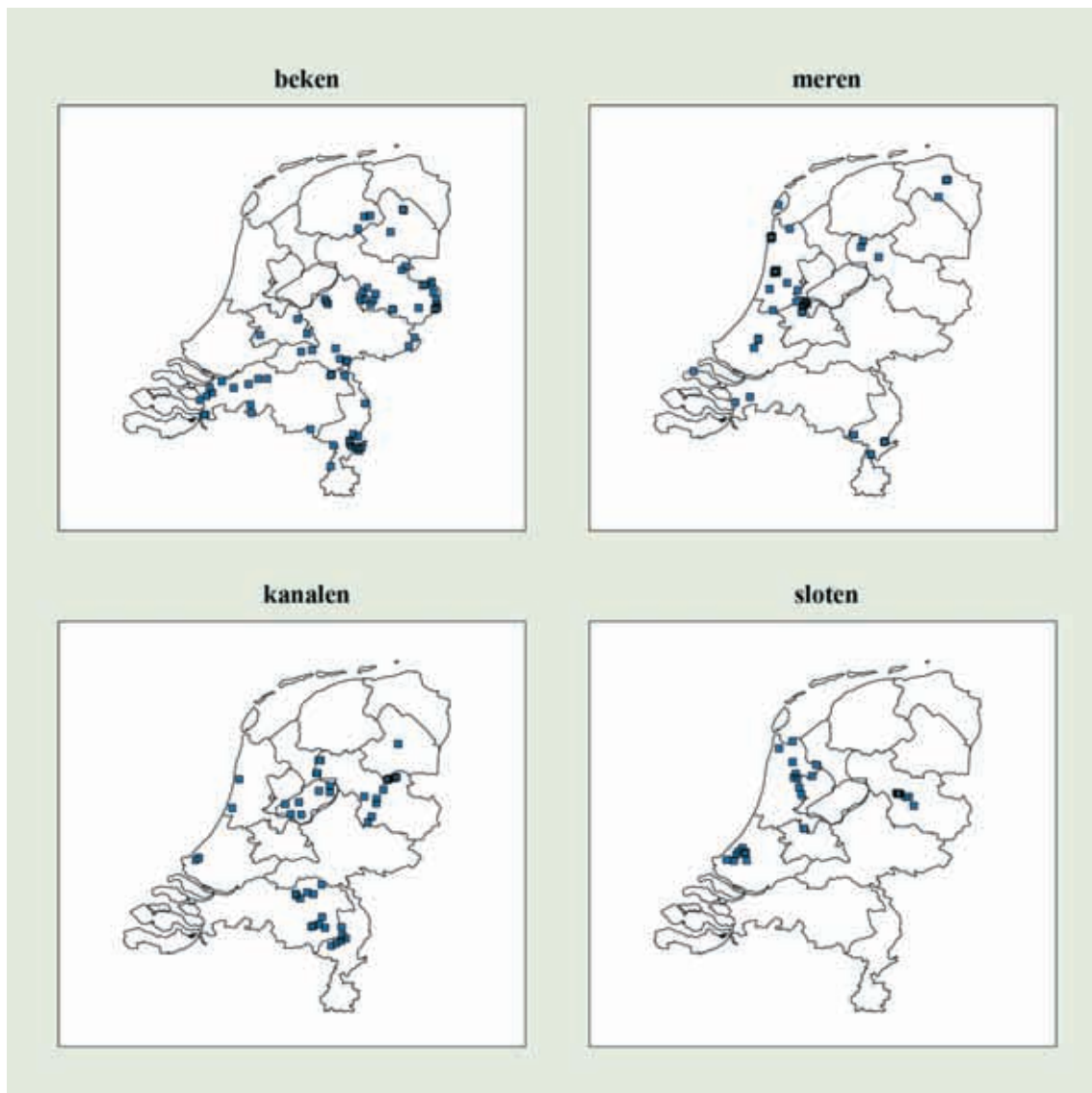
Hydromorfologische parameters beken	Klassen	Omschrijving
Meandering	1-5	1=F recht+normprofiel, 2=F gestrekt+natuurlijker dwarsprofiel, 3=F zwak slingerend, 4=F slingerend, 5=F vrij meanderend
Verstuwing	1-3	1=sterk gestuwd, 2=gestuwd, 3=ongestuwd
Beschaduwing	1-3	1=F onbeschaduwd, 2=F gedeeltelijk beschaduwd of ruigte oever en 3=F grotendeels beschaduwd (opgaande begroeiing/bos)
Chemische parameters	Eenheid	
BZV (Biologisch Zuurstof Verbruik)	mg O <sub>2</sub> /l	Zomergemiddelde (april-september)
Totaal P	mg P/l	Zomergemiddelde (april-september)
Totaal N	mg N/l	Zomergemiddelde (april-september)
Hydromorfologische parameters meren	Klassen	Omschrijving
Oeverinrichting	1-3	1=F beschoeid en/of onbegroeid, 2=F riet/helofyten, 3=F moeras+riet/helofyten
Peildynamiek	1-3	1=F tegennatuurlijk, 2=F stabiel, 3=F natuurlijk
Chemische parameters	Eenheid	
Totaal P	mg P/l	Zomergemiddelde (april-september)
Totaal N	mg N/l	Zomergemiddelde (april-september)
Hydromorfologische parameters kanalen	Klassen	Omschrijving
Oeverinrichting	1-3	1=F beschoeid, 2=F steil, 3=F flauw/moerassig (natuurvriendelijke oevers)
Peildynamiek	1-3	1=F tegennatuurlijk, 2=F stabiel, 3=F natuurlijk
Onderhoud	1-2	1=F intensief, 2=F extensief
Chemische parameters	Eenheid	
Totaal P	mg P/l	Zomergemiddelde (april-september)
Totaal N	mg N/l	Zomergemiddelde (april-september)
Hydromorfologische parameters sloten	Klassen	Omschrijving
Oeverinrichting	1-3	1=F beschoeid, 2=F steil, 3=F flauw/moerassig (natuurvriendelijke oevers)
Peildynamiek	1-3	1=F tegennatuurlijk, 2=F stabiel, 3=F natuurlijk
Onderhoud	1-2	1=F intensief, 2=F extensief
Chemische parameters	Eenheid	
Totaal P	mg P/l	Zomergemiddelde (april-september)
Totaal N	mg N/l	Zomergemiddelde (april-september)

Bron: Knoben et al. (2008).

Figuur 2.1 laat zien dat de spreiding voor elk watertype niet altijd representatief is voor Nederland als geheel. Zo ontbreken er bijvoorbeeld data voor meren, kanalen en sloten in de provincie Friesland. Op de consequenties wordt verder in gegaan in hoofdstuk 7.

### Representativiteit van stuurvariabelen

De variabelen uit Tabel 2.2 zijn belangrijke stuurfactoren voor de ecologische kwaliteit van regionale oppervlaktewateren. Natuurlijk zijn er meer stuurfactoren dan in de tabel opgenomen. Zo zijn bestrijdingsmiddelen niet meegenomen in deze analyse. In sloten worden de normen voor bestrijdingsmiddelen vaak overschreden (MNP, 2006) en zullen zeker een negatief effect hebben op de ecologische kwaliteit van sloten (De Lange et al., 2008). In hoeverre dat voor de beperkte dataset van sloten een belangrijke factor is, is niet bekend. Voor de sloten in het Westland, waar veel glastuinbouw gevestigd is, zal het wel een belangrijke factor zijn.



**Figuur 2.1** Geografische ligging van de data uit dit rapport: beken (linksboven), meren (rechtsboven), kanalen (linksonder) en sloten (rechtsonder). Voor de meeste locaties zijn meerdere data gebruikt: meerdere jaren voor hetzelfde water, of verschillende meetlocaties langs hetzelfde water.

Bij sloten is de inrichting van de oever een van de stuurvariabelen. De diepte van de sloot en de aanwezigheid van een voedselrijke waterbodem is ook een stuurvariabele. Baggeren van sloten is een beheersmaatregel die toegepast wordt om sloten een betere ecologische kwaliteit te geven.

In meren is de vissamenstelling mede bepalend voor de ecologische kwaliteit. Een hoge brasemstand zorgt voor een slechte EKR voor de soortgroep 'vissen'. Daarnaast beïnvloedt de visstand de ecologische situatie van het meer. De brasem zorgt door omwoeling van de bodem voor een troebel meer waarin geen waterplanten kunnen groeien. Bij Actief Biologisch Beheer wordt de brasemstand grotendeels verwijderd. Watervlooiën zijn dan voor korte tijd in staat om de algen te verwijderen waardoor waterplanten kunnen kiemen.

Beheersvisserij is een andere methode om meren die al een lage nutriëntbelasting hebben, helder te maken (Witteveen en Bos, 2008). Daarbij wordt vooral de witvis (zoals brasem) verwijderd. Met dezelfde nutriëntbelasting, dezelfde oeverinrichting en hetzelfde peilbeheer, kan een meer

helder zijn (goede EKR) of troebel (lage EKR) zijn. De stuurvariabele om dit proces te versnellen, actief verwijderen van een deel van de visstand, is niet opgenomen in de database. Ook het hysteresis-effect (Jaarsma et al., 2008) is niet opgenomen.

Specifiek voor vissen in beken geldt dat de bereikbaarheid nu onvoldoende is. Door stuwen en gemalen kunnen vissen sommige beken niet meer bereiken. Zelfs als de inrichting en de waterkwaliteit op orde is, is de visstand nog onvoldoende. Vispassages zijn indirect meegenomen in de onderdelen meandering en verstuwning. Echter het wel of niet bereikbaar zijn van een waterlichaam voor rheofiele (stroomminnende) soorten en soorten die migreren naar zee of grotere regionale wateren, is niet uitgewerkt.

Ondanks deze beperkingen bevat de lijst met stuurvariabelen uit Tabel 2.2 de meest belangrijke variabelen. Dat bleek uit toetsing in interne workshops, bij professor P. Goethals (Universiteit Gent) en bij externe workshops waar deze stuurvariabelen gepresenteerd zijn.



## 3 Statistiek

### 3.1 Presentatiemethoden

Bij de analyse van EKR-gegevens is gebruik gemaakt van verschillende statistisch-georiënteerde presentatiemethodes. Deze zullen in deze paragraaf kort toegelicht worden. De methodes zijn (i) correlatiematrixes en scatterplotmatrixes, en (ii) histogrammen en dichtheden.

#### 3.1.1 Correlatiematrixes, scatterplotmatrixes

De samenhang tussen twee gepaarde grootheden, zeg  $x_i$  en  $y_i$ , kan uitgedrukt worden met behulp van een correlatiecoëfficiënt  $R$  (of het kwadraat daarvan, dus  $R^2$ ).  $R$  is een getal tussen  $-1.0$  en  $1.0$ , en geeft aan in hoeverre er een *lineair* verband bestaat tussen de grootheden  $x$  en  $y$ . Bij een waarde van  $1.0$  is er een perfect lineair verband tussen  $x$  en  $y$ . Bij een  $R$ -waarde rond  $0.0$  is er geen enkel verband. Bij een  $R$ -waarde van  $-1.0$  is het lineaire verband zeer sterk, maar steeds tegengesteld: als  $x$  hoog is, is  $y$  juist laag, en omgekeerd.

Als meer dan twee grootheden aanwezig zijn, bijvoorbeeld  $x_{1,i}$ ,  $x_{2,i}$ ,  $x_{3,i}$  en  $x_{4,i}$ , dan kunnen de correlaties berekend worden tussen  $x_1$  enerzijds en  $x_2$ ,  $x_3$  en  $x_4$  anderzijds. Idem tussen  $x_2$  enerzijds en  $x_1$ ,  $x_3$  en  $x_4$  anderzijds. Et cetera. Uiteraard is de correlatie tussen een variabele en zichzelf altijd  $1.0$ . De resultaten kunnen dan weergegeven worden in een vier-bij-vier-matrix. Een voorbeeld voor de EKR-deelmaatlaten ‘macrofauna’, ‘vissen’, macrofyten’ en ‘fytoplankton’ is gegeven in Tabel 3.1. De resultaten laten zien dat sommige deelmaatlaten sterk samenhangen. Zo is de correlatie tussen ‘vissen’ en ‘macrofyten’ gelijk aan  $0.96$ .

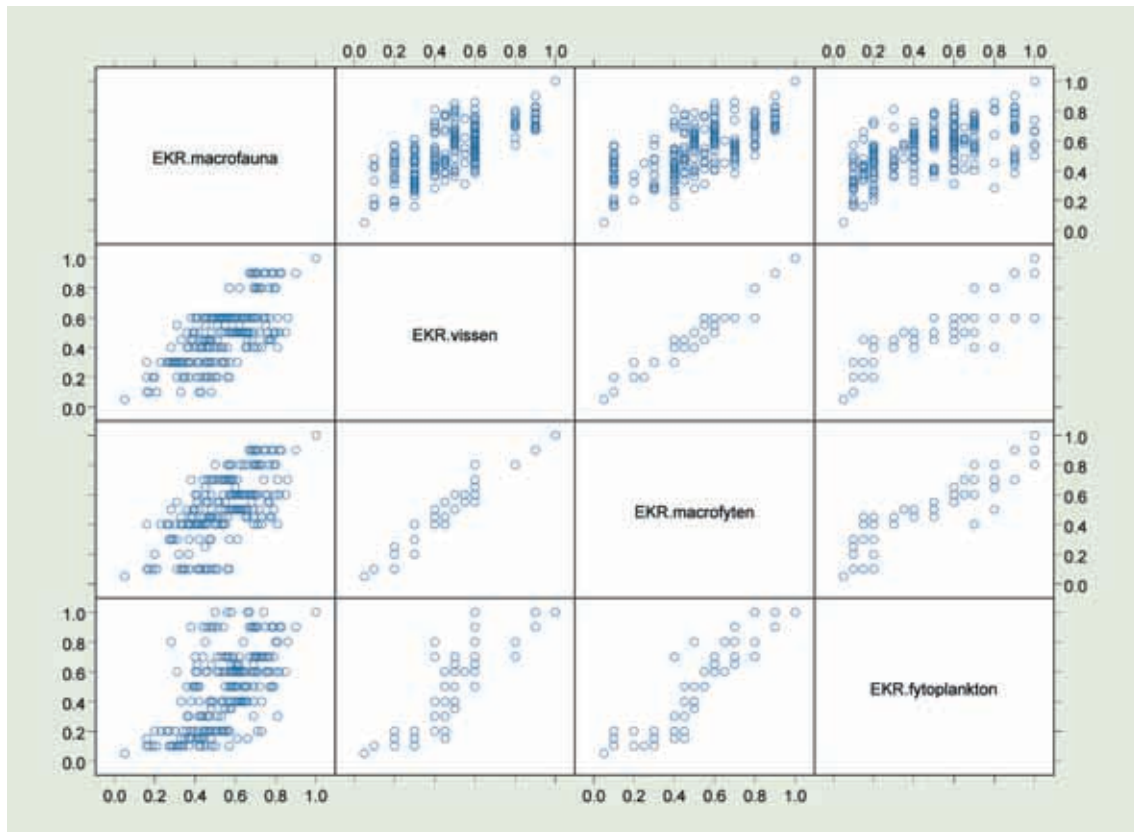
Dat betekent dat de informatie die aanwezig is in beide deelmaatlaten vrijwel uitwisselbaar is. Ook de correlatie tussen ‘macrofyten’ en ‘fytoplankton’ is hoog:  $R = 0.91$ . Idem tussen ‘vissen’ en ‘fytoplankton’:  $R = 0.87$ .

Om te zien of relaties tussen grootheden lineair zijn of niet, is het handig om getallenparen  $(x_i, y_i)$  weer te geven in een zogenaamde *scatterplot*. Dat is niets anders dan  $x_i$  op de  $x$ -as en  $y_i$  op de  $y$ -as. De patronen die dan zichtbaar worden, geven informatie over het soort relatie dat  $x$  en  $y$  hebben. Als die relatie sterk niet-lineair is, geeft het patroon informatie hoe wèl een relatie tussen  $x$  en  $y$  geformuleerd zou kunnen worden.

Als er meer dan twee variabelen aanwezig zijn, dan kunnen de scatterplots voor bijvoorbeeld  $x_1$ ,  $x_2$ ,  $x_3$  en  $x_4$  in één matrix geplaatst worden. Zie Figuur 3.1.

**Tabel 3.1 Vier-bij-vier-correlatiematrix voor EKR-bepalingen in 248 meergegevens.**

	EKR macrofauna	EKR vissen	EKR macrofyten	EKR fytoplankton
EKR macrofauna	1.00	0.66	0.63	0.58
EKR vissen	0.66	1.00	0.96	0.89
EKR macrofyten	0.63	0.96	1.00	0.91
EKR fytoplankton	0.58	0.87	0.91	1.00



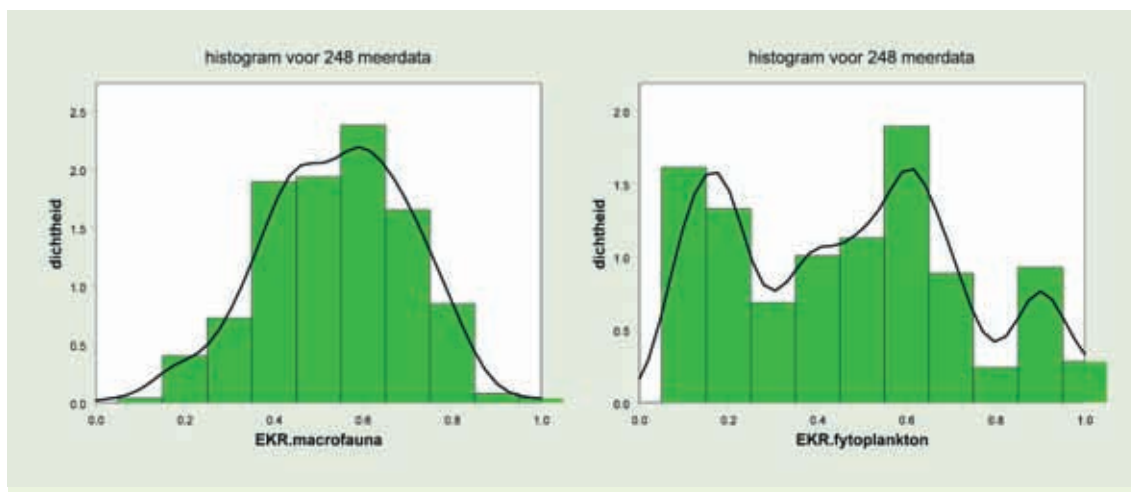
**Figuur 3.1** Scatterplotmatrix voor vier deelmaatlaten: macrofauna, vissen, macrofyten en fytoplankton. De gegevens zijn afkomstig uit de meren-dataset (N=248).

Zo'n scatterplotmatrix bevat op de eerste rij vier scatterplots: tussen  $x_1$  en zichzelf (hier als leeg blok weergegeven), tussen  $x_1$  en  $x_2$ , tussen  $x_1$  en  $x_3$ , en tussen  $x_1$  en  $x_4$ . Idem de tweede rij tussen  $x_2$  en  $x_1$ , tussen  $x_2$  en zichzelf, tussen  $x_2$  en  $x_3$ , en tussen  $x_2$  en  $x_4$ , et cetera. De data in Figuur 3.1 zijn dezelfde data als gebruikt voor de correlatiematrix in Tabel 3.1.

De scatterplotmatrix laat zien dat de relatie tussen 'vissen' en 'macrofyten' sterk lineair is (wat te verwachten was bij  $R=0.96$ ). De puntenwolk, getoond in het centrum van de scatterplotmatrix, ligt zeer nauw rond de 45-graden lijn die gaat door de coördinaten (0,0) en (1,1). Daarentegen is de relatie tussen 'macrofauna' en 'fytoplankton' slecht ( $R=0.58$ ): de puntenwolk in hun onderlinge scatterplot, gelegen in de linkeronderhoek en rechterbovenhoek van de matrix, vult een groot deel van de ruimte.

### 3.1.2 Histogrammen en dichtheden

Om een reeks van EKR-waarden te karakteriseren, kunnen voor een set data gemiddelden en standaarddeviaties berekend worden. Ook percentielwaarden zijn informatief, zoals het 25-, 50- en 75-percentiel. De data kunnen ook simpelweg gekarakteriseerd worden door ze weer te geven in een histogram met de corresponderende geschatte dichtheidscurve. Voordeel is dat bijvoorbeeld EKR-waarden visueel gemaakt worden, waarbij de dichtheid een continue weergave geeft van de histogrambalken. Hierbij staan de EKR-waarden op de x-as en de dichtheden, discreet en continu, op de y-as.



**Figuur 3.2** Overzicht van 248 macrofauna-EKR-waarden (links) en 248 fytoplankton-EKR-waarden (rechts). Data zijn voor meren.

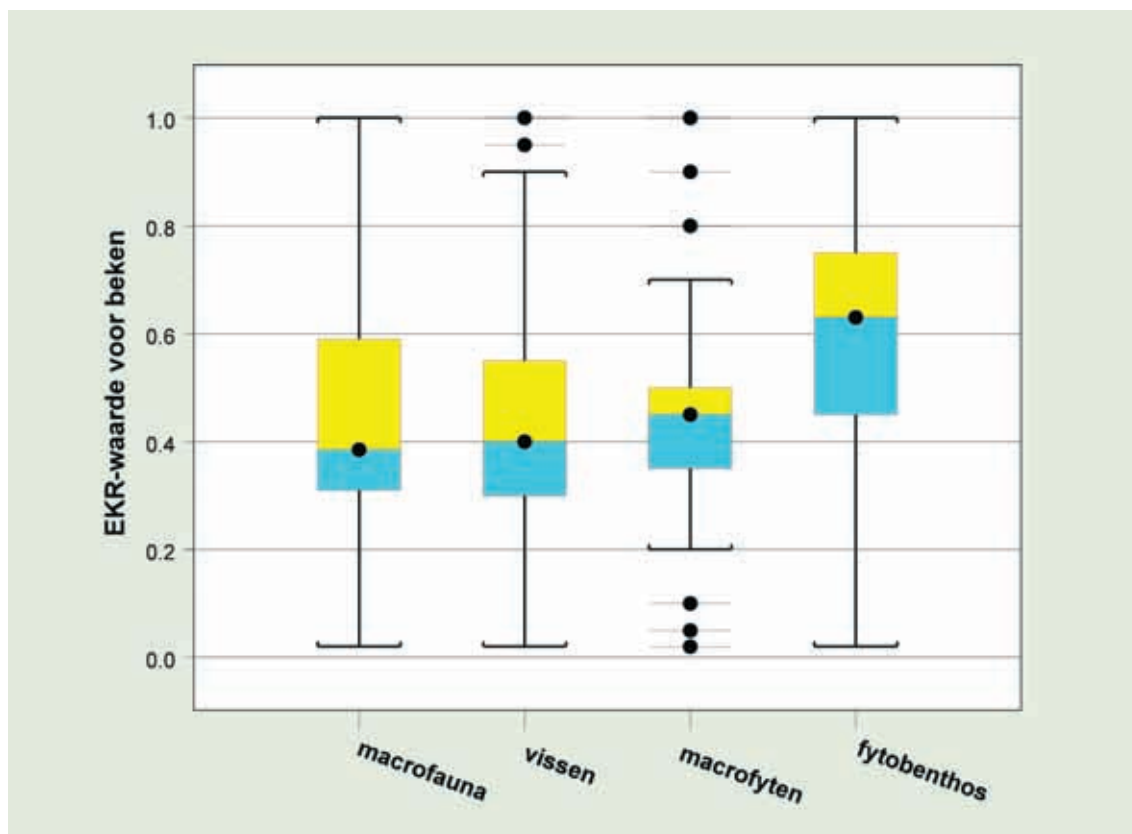
Het voordeel van dichtheden is dat het oppervlak onder de curve precies gelijk is aan 1.0, waardoor verschillende histogrammen met hun dichtheden vergeleken kunnen worden.

Een voorbeeld voor de deelmaatlatten ‘macrofauna’ en ‘fytoplankton’ in 248 meren is gegeven in Figuur 3.2. Het voorbeeld laat zien dat de vorm van de dichtheidscurve aanzienlijk kan afwijken voor de verschillende deelmaatlatten, bepaald in dezelfde wateren. Dat betekent de gevoeligheid van verschillende soortgroepen voor de ecologische omstandigheden sterk variabel kan zijn.

### 3.1.3 Boxplots

Een zogenaamde *boxplot* geeft een grafische samenvatting van zes kengetallen behorend bij een groep getallen. Deze kengetallen zijn: het absolute minimum en maximum, het rekenkundig gemiddelde, het 25-, 50- en 75-percentiel. Het verschil tussen het 25- en 75-percentiel wordt aangeduid als de interkwartielafstand (IKA) en wordt in een boxplot weergegeven met een rechthoek. De maximum- en minimumwaarde worden weergegeven via dwarsstreepjes, de zogenaamde ‘whiskers’. Tenslotte worden, indien aanwezig, uitbijters weergegeven. Uitbijters zijn getallen die verder weg liggen van het 25- of 75-percentiel dan 1.5 maal de IKA. Zie voor meer details bijvoorbeeld: <http://nl.wikipedia.org/wiki/Boxplot>.

Figuur 3.3 geeft een voorbeeld voor vier EKR-deelmaatlatten, behorend bij de dataset ‘beken’. De 25- en 75-percentielen geven de onder- en bovenwaarde van de rechthoek; het bolletje in de rechthoek is de mediaan. Het rekenkundig gemiddelde is de horizontale lijn in de rechthoek en valt in dit voorbeeld vrijwel samen met de mediaan. De soortgroep fytoobenthos heeft duidelijk de hoogste ecologische toestand. Spreiding, gezien als het verschil tussen het 25- en 75-percentiel, is het geringst voor ‘macrofyten’. Maar deze deelmaatlat heeft wel weer de meeste uitbijters, zowel naar boven als naar beneden. Voor ‘macrofauna’ en ‘fytoobenthos’ zijn de ‘whiskers’ van de boxplots gelijk aan het absolute minimum en maximum van de aanwezige data. Voor ‘vissen’ zijn twee uitbijters gedetecteerd, en voor ‘macrofyten’ drie uitbijters aan beide zijden.



Figuur 3.3 Voorbeeld van een boxplot voor vier deelmaatlaten (beken).

## 3.2 Regressieboomanalyse

### 3.2.1 Uitleg van de methode

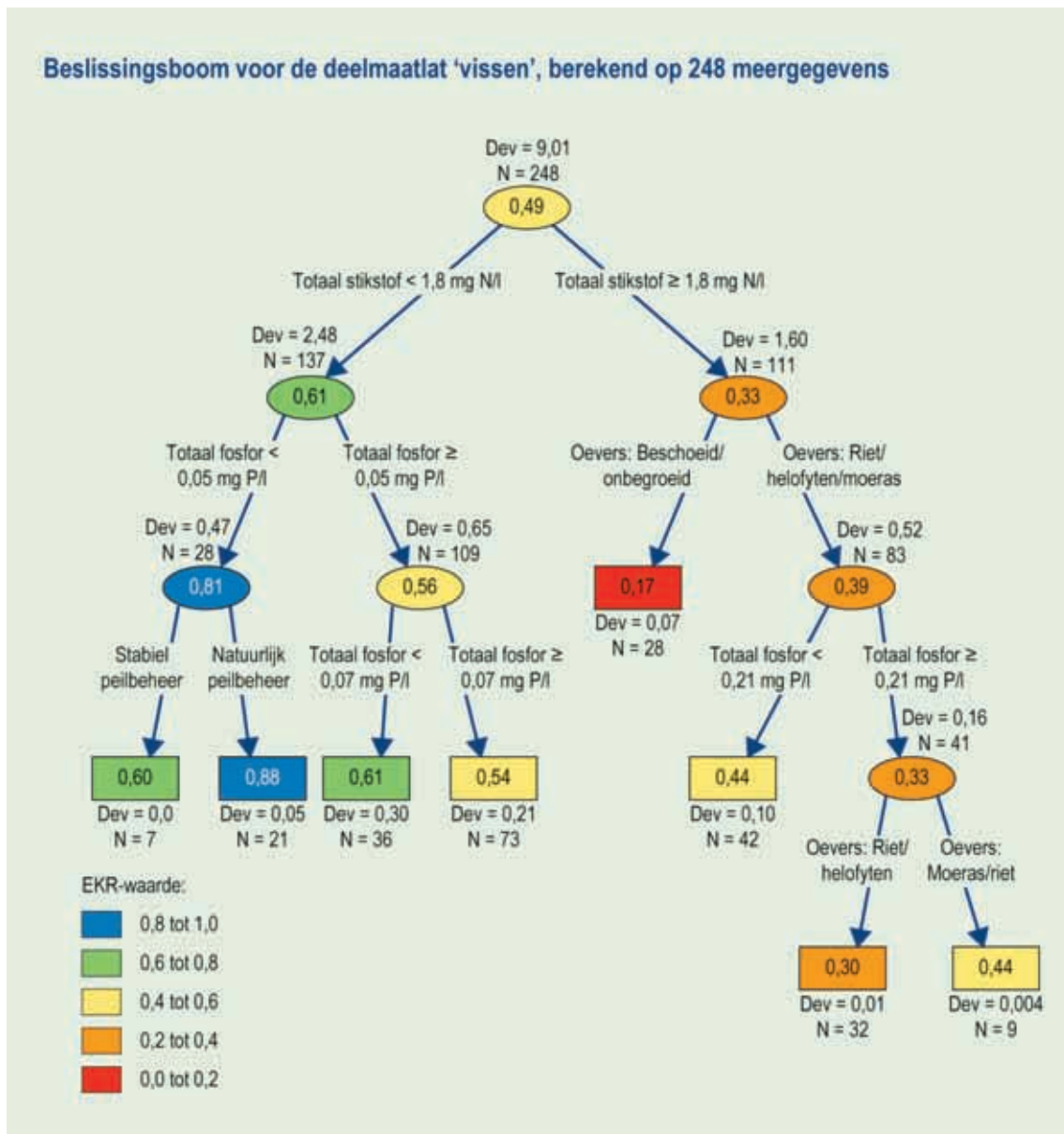
Regressiebomen of *regression trees* vallen onder de methodiek die in het Engels afgekort wordt als 'CART', ofwel Classification And Regression Trees. De methode wordt ook wel aangeduid met de term *decision tree learning*, zie bijvoorbeeld

[http://en.wikipedia.org/wiki/Decision\\_tree\\_learning](http://en.wikipedia.org/wiki/Decision_tree_learning) .

De methode is een inzichtelijke techniek waarmee de relaties tussen een variabele  $y_i$  en verklarende variabelen  $x_{1,i}$ ,  $x_{2,i}$ ,  $x_{3,i}$ , ...,  $x_{m,i}$  geschat kunnen worden. Hierbij mogen de relaties tussen een  $y_i$ - en een  $x_i$ -variabele niet-lineair zijn. Overigens wordt met de term 'classification tree' de situatie bedoeld waarbij de  $y$ -variabele een nominaal karakter heeft, dus bijvoorbeeld 'o' of '1'. Bij regressiebomen is de  $y$ -variabele een continue grootheid.

In dit rapport wordt voor 'y' steeds een EKR-deelmaatlat gekozen en voor de  $x$ -en een aantal stuurvariabelen, zoals peilbeheer, oeverinrichting, totaal fosfor of totaal stikstof. Omdat EKR-waarden continu zijn, leidt dit tot regressiebomen.

Hoe werkt het schatten van een regressieboom? Bij een regressieboom wordt gekeken of de variabiliteit van de  $y$ -variabele sterk verminderd kan worden door te 'knippen' op de waarde van een van de  $x$ -variabelen. De set van  $y$ -waarden kan bijvoorbeeld verdeeld worden over twee deelsets: deelset I bevat alle  $y$ -waarden waarvoor  $x_3 > 4.0$ , en deelset II alle  $y$ -waarden waarvoor



**Figuur 3.4** Regressieboom voor deelmaatlat 'vissen'. De boom is geschat voor 248 meergegevens. Het gemiddelde van EKR-waarden na een knip op een van de stuurvariabelen staat in een ellips of rechthoek. Dit worden de knopen van de boom genoemd. Bij elke knoop staat de zogenaamde deviantie (= som van gekwadrateerde verschillen met het gemiddelde) en het aantal wateren dat in die knoop terecht komt.

$x_3 \leq 4.0$ . Zo'n knip heeft alleen voordeel als de variabiliteit van  $y$  in beide deelsets *veel minder* is dan de variabiliteit zonder de knip. Het regressieboom-algoritme zoekt dan naar die  $x$ -variabele welke de grootste vooruitgang boekt in het reduceren van de variabiliteit in  $y$ . Als die variabele dan  $x_3$  blijkt te zijn, zoekt het algoritme naar de optimale knip, bijvoorbeeld  $x_3 = 4.0$ , om de  $y$ -waarden over twee deelsets te verdelen.

Na zo'n eerste knip wordt gezocht naar een tweede knip, een derde knip, et cetera. Het algoritme stopt wanneer er minder dan een van te voren opgegeven aantal  $y$ -waarden overblijft na een knip. Het splitsproces kan weergegeven worden door een boomstructuur (Figuur 3.4). De gemiddelde waarde van een groep  $y$ -data wordt hierbij aangegeven door een ellips (een knoop)

of een rechthoek (een eindknoop). Knopen zijn verbonden door een splitswaarde op een van de x-en. Als de boom te groot wordt, kan hij ‘teruggesnoeid’ worden (*pruning*).

De methode wordt geïllustreerd aan de hand van een voorbeeld. In Figuur 3.4 is een regressieboom gegeven waarbij de te verklaren variabele de EKR-waarden van de deelmaatlat ‘vissen’ is (in meren), waarbij vier stuurvariabelen voorhanden zijn. Dit zijn oeverinrichting, peilbeheer, totaal fosfor en totaal stikstof.

De boom in Figuur 3.4 moet van boven naar beneden gelezen worden: de belangrijkste vooruitgang wordt geboekt in de eerste knip, hier op de variabele ‘Totaal stikstof’, bij een knipwaarde van 1.8 mg N/l. De gemiddelde EKR-waarde van alle 248 EKR-waarden is 0.49 en staat in de bovenste ellips (knoop). Door nu te knippen op ‘Totaal stikstof’ worden de EKR-waarden verdeeld over twee groepen: de linker knoop heeft een gemiddelde waarde van 0.61 en de rechter knoop van 0.33. Dit betekent dat als het stikstof-niveau van meren gebracht kan worden naar waarden onder de 1.8 mg N/l, dat dan, gemiddeld gezien, de ecologische kwaliteit sterk vooruitgaat: van gemiddeld 0.49 (matige ecologische toestand) naar 0.61 (een goede ecologische toestand).

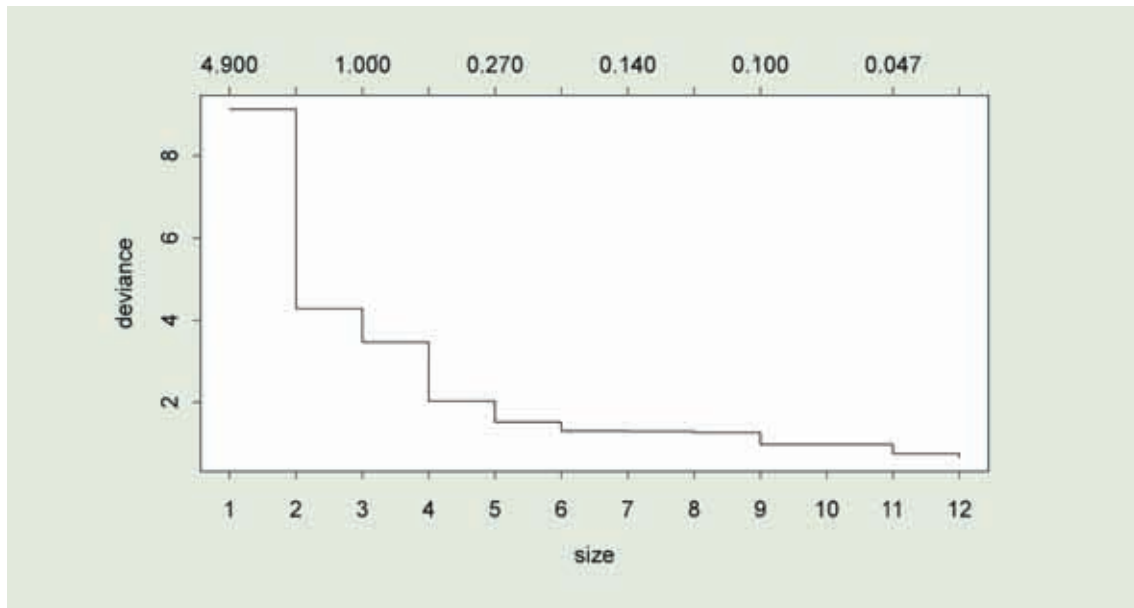
Elk van beide knopen kan nu weer gesplitst worden, op dezelfde manier als de eerste splitsing, et cetera. De boom laat zien dat de *hoogste* gemiddelde EKR-waarde (0.88) gevonden wordt voor meren waarvoor ‘Totaal stikstof’ kleiner is dan 1.8 mg N/l, ‘Totaal fosfor’ lager is dan 0.05 mg P/l en een peilbeheer in klasse 3. Omgekeerd geldt voor de meren met de *laagste* EKR-waarden (0.17) een ‘Totaal stikstof’ groter dan 1.8 mg N/l en tegelijkertijd een oeverinrichting in klasse 1 (= beschoeid/onbegroeid).

Hoe kan de vooruitgang na elke knip bepaald worden? De vooruitgang kan berekend worden uit de getallen onder de knopen: de zogenaamde *devianties*. Dit is de som van de gekwadrateerde verschillen tussen de EKR-waarde van een meer en het gemiddelde in de knoop waartoe de EKR-waarde behoort (het getal in de ellips of rechthoek). De bovenste knoop heeft deviantie 9.01 en dat is de som van alle 248 kwadratische afwijkingen van 0.49. N.B.: als een deviantie gedeeld wordt door het aantal meren, dan wordt de variantie van alle meren gevonden die tot een knoop behoren.

Na de eerste splitsing is de deviantie teruggelopen naar  $2.48 + 1.60 = 4.08$ . Dat is ten opzichte van de bovenste knoop een reductie van  $100\% * (9.01 - 4.08) / 9.01 = 55\%$ . Wordt één laag in de boom ‘gezakt’, dan is de totale deviantie gedaald naar  $0.47 + 0.65 + 0.07 + 0.52 = 1.71$ . Dit is een reductie ten opzichte van de bovenste knoop van  $100\% * (9.01 - 1.71) / 9.01 = 81\%$ . Als alle 8 eindknopen (de rechthoeken) samengenomen worden, bedraagt de reductie 92%.

In het schattingsproces van de boom uit Figuur 3.4 is teruggesnoeid naar een boom met acht knopen. Om dit getal van 8 knopen te vinden, wordt een terugsnoei-grafiek gebruikt die gegeven is in Figuur 3.5. Op de y-as staat de totale deviantie van een boom en op de x-as het aantal bijbehorende knopen van de geschatte boom. Omdat de deviantie niet veel meer afneemt bij een boom met meer dan acht knopen, is voor dit getal gekozen. De boom kan wel groter gekozen worden (bijvoorbeeld twaalf of meer knopen), maar de winst, in termen van lagere devianties, wordt dan zeer gering.

Hoe ziet nu de verdeling van EKR-waarden in elke eindknoop eruit? Dat is weergegeven in Figuur 3.6A; idem de verdeling van alle 248 EKR-waarden in de beginknoop, versus de verde-



**Figuur 3.5** Het snoei-algoritme berekent een figuur met devianties op de y-as en boomgrootte op de x-as (*size*). Hieruit kan afgelezen worden tot welk aantal eindknoten de boom moet worden teruggesnoeid.

ling van de EKR-waarden in elk van de acht eindknoten, in Figuur 3.6B. Het blijkt dat de oorspronkelijke EKR-waarden verspreid liggen over het hele EKR-traject: van 0.07 tot bijna 1.0. De variabiliteit rond de eindknoop-waarden in Figuur 3.6B, gegeven in de oranje balk boven de histogrammen, is veel kleiner geworden.

Een uitvoeriger beschrijving van de regressieboom-techniek is gegeven in Visser en Noordijk (2002). Zie verder ook Venables en Ripley (2002, hoofdstuk 9).

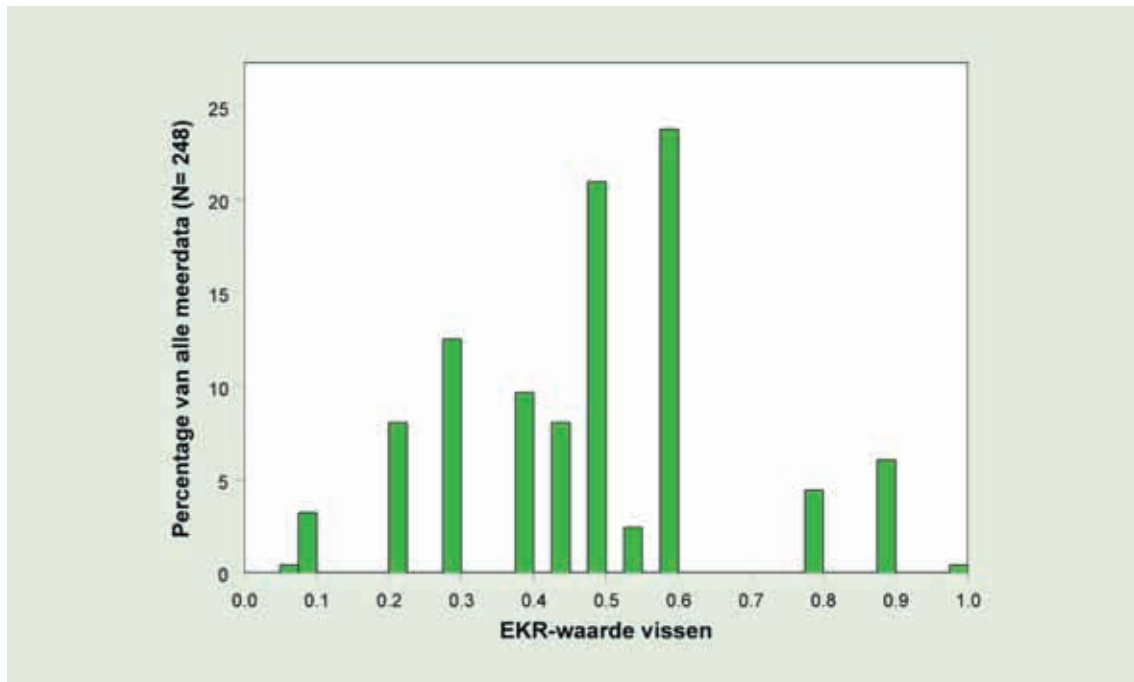
Samengevat geeft de regressieboomtechniek een gestructureerde methode om inzichtelijke relaties te schatten tussen EKR-waarden enerzijds en stuurvariabelen anderzijds. Door de hiërarchische structuur is direct duidelijk wat de belangrijkste splitsingen zijn in de boom. De omvang van de regressieboom kan bepaald worden via een snoei-algoritme.

De rol van onzekerheden komt in de volgende paragraaf aan bod.

### 3.2.2 Onzekerheden

Een maat voor de spreiding binnen een eindknoop van een geschatte regressieboom is de variantie of standaarddeviatie van de data die in zo'n knoop vallen. Die varianties volgen eenvoudig uit de bij de knoop behorende deviantie:  $\text{variantie} = \text{deviantie}/N$ . De standaarddeviatie is te gebruiken als een maat voor de voorspelnaauwkeurigheid van de betreffende EKR-waarde.

Een heel andere vraag is hoe stabiel of 'stevig' de geschatte regressieboom eigenlijk is. En wat is zijn voorspelkracht voor data waarop de boom niet geschat is? Deze vragen kunnen beantwoord worden door de data te splitsen in een trainingsset (in dit rapport 80% van de beschikbare data) en een testset (20% van de data). De boom die geschat is op basis van de trainingsset, wordt dan gebruikt om de data van de testset te voorspellen. Deze methode wordt wel aangeduid met *cross validation* (kruisvalidatie).



**Figuur 3.6A** Histogram van alle 248 EKR-waarden van 'vissen'. De hoogte van alle balken samen sommeert tot 100%.

De nauwkeurigheid van de voorspellingen kan vervolgens uitgedrukt worden als de *root mean squared error* (RMSE) of de *mean absolute deviation* (MAD). Dit is in het eerste geval de gemiddelde waarde van alle gekwadrateerde voorspelfouten voor alle testset-voorspellingen, en in het tweede geval het gemiddelde van alle absolute voorspelfouten. Het MAD-criterium is minder gevoelig voor grote voorspelfouten.

Als een deelmaatlatwaarde voor water 'i' aangeduid wordt met  $y_i$  en een kruisvalidatie-voorspelling daarvoor als  $y_{cross,i}$ , bij een testsetgrootte van M wateren, dan zijn de RMSE en de MAD als volgt gedefinieerd:

$$RMSE = \sqrt{\frac{1}{M} \sum_{i=1}^M (y_i - y_{cross,i})^2} \quad (1a)$$

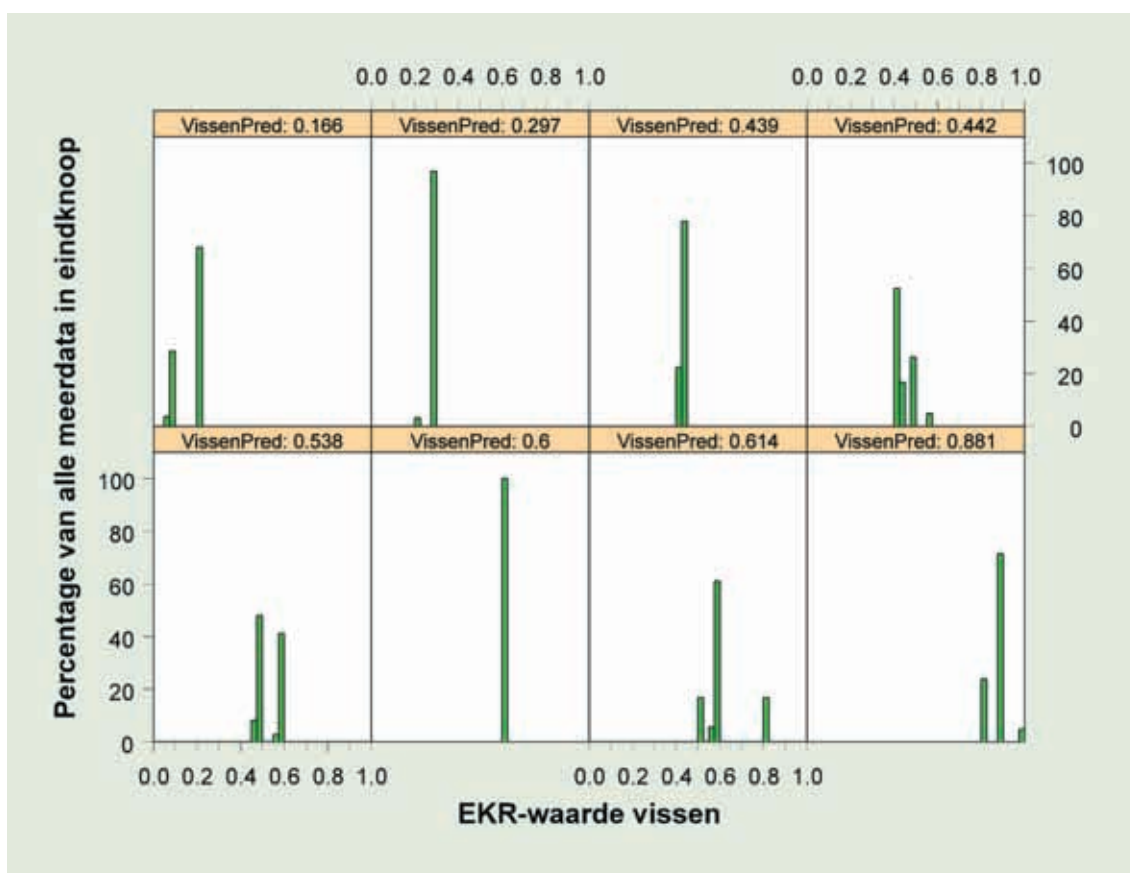
en

$$MAD = \frac{1}{M} \sum_{i=1}^M |y_i - y_{cross,i}| \quad (1b)$$

Daarnaast wordt ook een maat voor het lineaire verband tussen  $y_i$  en  $y_{cross,i}$  gebruikt die lijkt op de traditionele correlatiecoëfficiënt  $R^2$ . Bij een perfecte voorspelling voor alle M testset-wateren geldt dat  $R_{adj}^2$  gelijk is aan 1.0:

$$R_{adj}^2 = \frac{\text{var}(y_i) - RMSE^2}{\text{var}(y_i)} \quad (2)$$





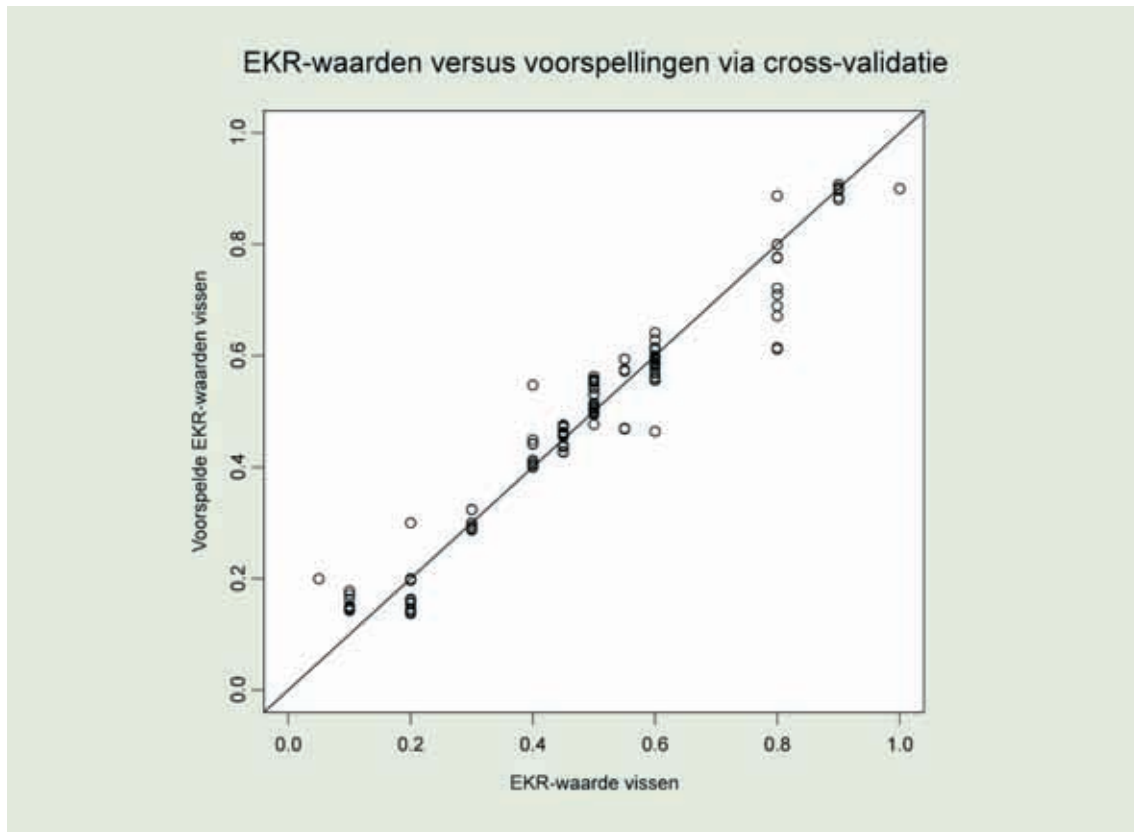
**Figuur 3.6B** Idem als Figuur 3.6A maar nu histogrammen voor elk van de 8 eindknoopen. Het aantal data per eindknoop is (van links boven naar rechts onder): 28, 32, 9, 42, 73, 7, 36 en 21. De som van deze cijfers is gelijk aan het totaal aantal data, namelijk 248.

In principe kan  $R_{adj}^2$  negatief worden. Daarom wordt  $R_{adj}^2$  begrensd op nul.

Als nu niet éénmaal een trainingsset en een testset worden kozen, maar bijvoorbeeld 500 maal, en dus 500 regressiebomen geschat worden, dan wordt een robuuste schatting gevonden voor de voorspelkracht van de regressieboombenadering. Een dergelijke kruisvalidatie-benadering is in dit rapport uitgevoerd met behulp van het pakket ‘randomForest’ van Leo Breiman, de ‘vader’ van de regressieboom-benadering (Breiman, 2001). Met zijn software worden de kengetallen RMSE, MAD en  $R_{adj}^2$  op  $500 \cdot M$  testset-wateren geschat.

Bij elke geschatte regressieboom in dit rapport worden de kruisvalidatie-voorspellingen getoond op basis van 500 *random* gekozen trainingssets in de vorm van een scatterplot. Elke trainingsset is dus een deelverzameling van de volledige dataset.

Een voorbeeld voor de deelmaatlat ‘vissen’ in meren is gegeven in Figuur 3.7 (identiek aan Figuur B.7). De figuur geeft een scatterplot van  $y_i$  (EKR-waarde vissen, op de x-as) tegen  $y_{cross,i}$  (voorspelde EKR-waarden ‘vissen’, op de y-as). Voor 248 meerdata worden de kruisvalidatie-voorspellingen  $y_{cross,i}$  gegeven. Omdat elk van de 500 regressiebomen geschat is op 80% van de beschikbare data (dus 198 meerdata), en de kruisvalidatie op de overblijvende 20% van de data (dus 50 meerdata), bedraagt het aantal kruisvalidatie-voorspellingen in de figuur  $500 \cdot 50 = 25000$ .



**Figuur 3.7** EKR-voorspellingen met *randomForest* (500 bomen en een trainingsset van 80% van het totale aantal meren). Kengetallen: MAD= 0.019, RMSE= 0.036 en  $R_{adj}^2= 0.96$ . N.B.: in de figuur vallen veel cirkelsymbolen over elkaar heen.

Voor details over *randomForest* wordt verwezen naar Breiman (2001) en naar

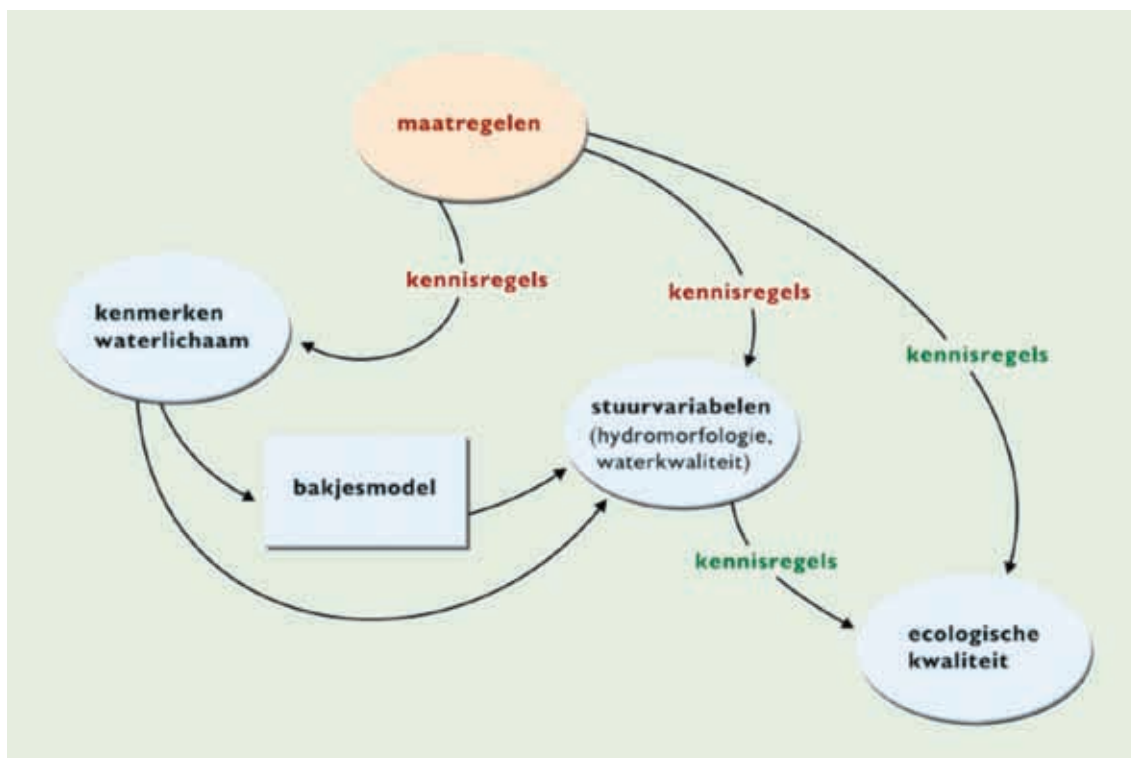
[http://en.wikipedia.org/wiki/Random\\_forest](http://en.wikipedia.org/wiki/Random_forest) .

Samengevat is gevonden dat individuele voorspellingen van een regressieboom gegeven kunnen worden met bijbehorende onzekerheden. Deze onzekerheden volgen uit de deviantie per eindknoop in de boom. Daarnaast kunnen onzekerheden in het splitsgedrag van de regressieboom geschat worden door kruisvalidatie. Kruisvalidatie levert kengetallen (de RMSE of de MAD) waarmee de *overall*-onzekerheid van regressieboomvoorspellingen wordt weergegeven.

### 3.2.3 Regressieboomanalyse in de context van de KRW

Er bestaat een reeks aan methoden, al dan niet wiskundig van karakter, waarmee de relatie tussen EKR-waarden (de afhankelijke variabele  $y_i$ ) en stuurvariabelen (verklarende variabelen  $x_{1,i}$ ,  $x_{2,i}$ , ...) geschat kan worden. Binnen de statistiek komen multiple-regressie-analyse en regressieboomanalyse in aanmerking. Beide methoden hebben het woord 'regressie' gemeen, maar zitten wiskundig gezien heel anders in elkaar. Buiten de statistiek worden vooral *expert judgement* en neurale netwerken toegepast als methode om de invloed van stuurvariabelen te beschrijven.

Een aparte plaats neemt de KRW-Verkenner in (zie <http://www.krwwerken.nl/> onder het kopje 'handleidingen'). Dit software-pakket is ontwikkeld om waterbeheerders te ondersteunen bij de



**Figuur 3.8** Schematische weergave van de opzet van de KRW-Verkenner.

Bron: <http://www.krwwerken.nl/>

opstelling van stroomgebiedbeheerplannen. Het pakket kan EKR-deelmaatlatwaarden berekenen op basis van kennisregels die deels gebaseerd zijn op *expert judgement* en deels op enkelvoudige regressiemodellen. Zie Figuur 3.8 voor een schematisch overzicht.

Het effect van maatregelen op rijkswateren is voor de Ex-ante evaluatie KRW doorgerekend met de KRW-Verkenner. Zie PBL (2008) en Engelbertink et al. (2008).

Voor- en nadelen van de genoemde methoden zijn als volgt:

- Het voordeel van multiple-regressie is dat er een mooi statistisch raamwerk is ontwikkeld om de gezochte relaties te schatten, inclusief de onzekerheden in die relaties. Nadeel van de methode is dat de gezochte relaties *lineair* moeten zijn, dus een model van de vorm

$$y_i = \text{constante} + a_1 * x_{1,i} + a_2 * x_{2,i} + \dots + \text{ruis}_i \quad (3)$$

Deze eis van lineariteit maakt de methode minder geschikt voor toepassing binnen de KRW. EKR-waarden blijken meestal in meer of mindere mate niet-lineair gekoppeld te zijn aan stuurvariabelen.

- *Expert judgement* kan een heel goede methode zijn. Experts, zoals hydrobiologen, bezitten veel ervaring op het terrein van ecologische waterkwaliteit en kunnen wellicht voorspellen of de kwaliteit van een beek, meer of sloot in de toekomst zal verbeteren of verslechteren. Op basis van *expert judgement* kunnen vuistregels/kennisregels geformuleerd worden waarop beleid kan worden gevoerd. Nadeel van deze benadering is dat (i) er altijd een mate van subjectiviteit meespeelt in de benadering, (ii) dat daardoor de reproduceerbaarheid vaak gering is, en dat (iii) biologische processen heel ingewikkeld kunnen zijn, met veel inter-

acties en terugkoppelingen. Een aantal voorbeelden van kennisregels is te vinden op de KRW-Verkenner site: [http://www.krwverkenner.nl/files/4\\_2\\_EcologischeKennisregels.pdf](http://www.krwverkenner.nl/files/4_2_EcologischeKennisregels.pdf)

- Toepassing van *neurale netwerken* heeft in het verleden bewezen een krachtige methode te zijn om op basis van een dataset voorspellingen te genereren. Op het karakter en de toepassing van neurale netwerken wordt teruggekomen in hoofdstuk 6.
- De KRW-Verkenner geeft een invulling voor het linker stroomdiagram in Figuur 1.2. Door de gemodelleerde biologische interacties en brede set van maatregelen is de Verkenner in principe een krachtiger benadering dan die volgens een metamodel (rechterstroomdiagram in Figuur 1.2). Nadeel van de KRW-Verkenner in zijn huidige vorm is dat (i) het model nog in ontwikkeling is, (ii) dat er door waterbeheerders veel werk verricht moet worden om het model te kunnen draaien en (iii) dat elke waterbeheerder zelf een validatie van de voorspelkracht van het model moet uitvoeren. Bij dat laatste punt moet opgemerkt worden dat de KRW-Verkenner niet ontwikkeld is om voor één specifieke beek, kanaal of sloot een nauwkeurige voorspelling te genereren. De opzet is om beheerders meer algemeen te laten zien wat de invloed kan zijn van het wijzigen van individuele stuurvariabelen. De uitkomsten zijn dus *indicatief* bedoeld.

### 3.3 Software

Alle berekeningen in dit rapport zijn uitgevoerd met het statistische softwarepakket S-PLUS 8.0. Zie bijvoorbeeld <http://en.wikipedia.org/wiki/S-PLUS>, Millard en Neerchal (2001), en Venables en Ripley (2002) voor meer informatie over statistiek en S-PLUS.

De R-software van Breiman, *randomForest*, is beschikbaar in een versie die aan S-PLUS 8.0 kan worden gekoppeld. Deze koppeling is gebruikt in deze studie. De site vanwaar het pakket *randomForest* (gratis) gedownload kan worden, is <http://csan.insightful.com/>.

## 4 Actuele EKR's en hun onderlinge relaties

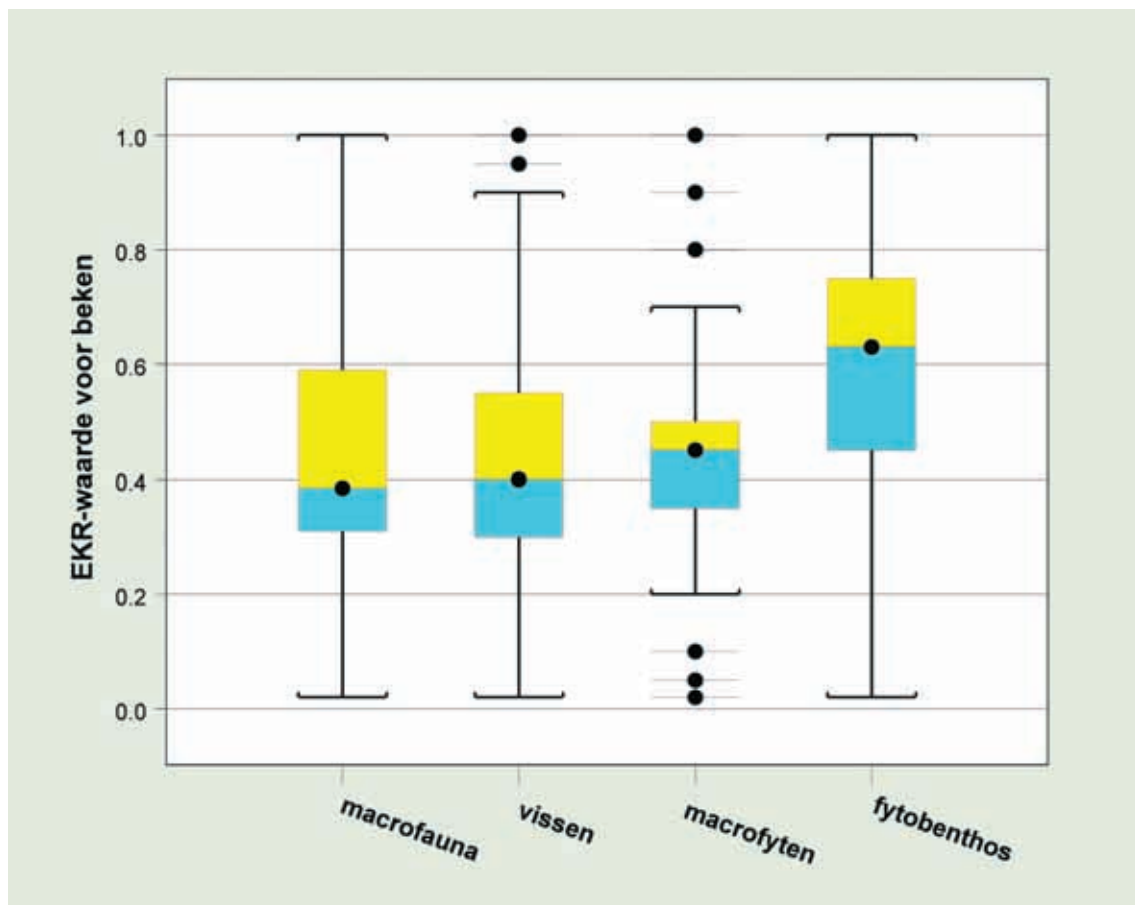
Dit hoofdstuk geeft een overzicht van de huidige EKR-waarden en hun onderlinge relaties in de onderzochte watertypen beken, meren, kanalen en sloten. De waarden worden gepresenteerd in de vorm van boxplots, histogrammen met dichtheden en scatterplotmatrices (zie respectievelijk paragraaf 3.1.3, 3.1.2 en 3.1.1 voor een methodische uitleg).

### 4.1 Beken

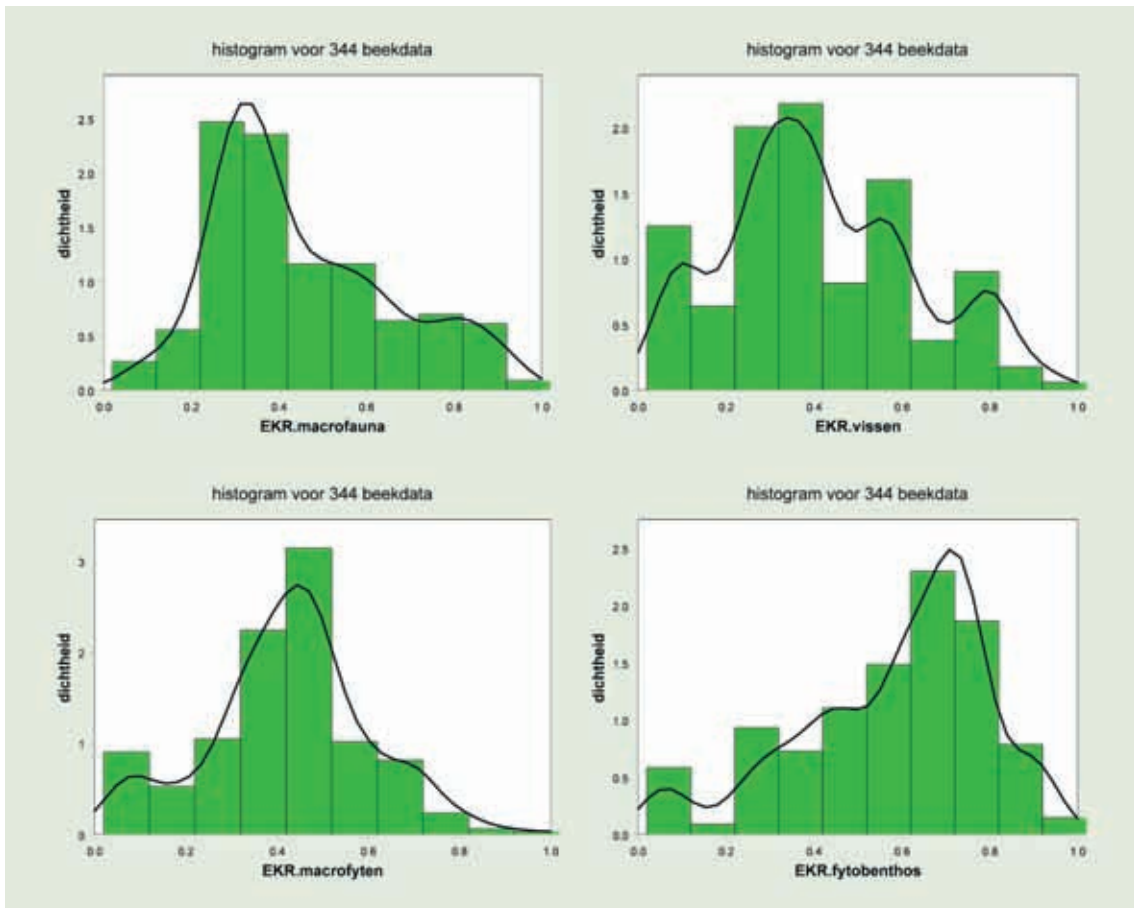
Figuur 4.1A geeft boxplots voor de vier deelmaatlatten. Het blijkt dat EKR-waarden voor fyto-benthos duidelijk het hoogst zijn en vooral liggen in de klassen matige en goede ecologische toestand. De overige deelmaatlatten liggen overwegend in de klassen 'ontoereikend' en 'matig'. De boxplots laten verder zien dat er incidenteel waarden bij 0.0 en 1.0 voorkomen (whiskers en uitbijters bij 'macrofyten').

Figuur 4.1A Overzicht ecologische toestand van 344 beekgegevens: boxplots. Figuur 4.1B geeft de histogrammen met dichtheden. Ook nu valt op dat de hoogste EKR-waarden gevonden worden voor 'fytobenthos' (dichtheid het meest naar rechts verschoven).

Figuur 4.1C geeft een indruk hoe beken met een wisselende ecologische kwaliteit er in de praktijk uit kunnen zien. Bij de beoordeling zijn alle deelmaatlatten betrokken (de slechtste EKR-waarde is daarbij maatgevend).



Figuur 4.1A Overzicht ecologische toestand van 344 beekgegevens: boxplots.



Figuur 4.1B Overzicht ecologische toestand van 344 beekgegevens: histogrammen met dichtheden.

### Ecologische kwaliteit van oppervlaktewater - beken



Kwaliteit: goed  
EKR: 0,6 - 0,8

Kenmerk van een kwalitatief goede beek is vooral afwisseling. Beschut en open, dieper en ondieper, zowel zand als stenen etc.



Kwaliteit: matig  
EKR: 0,4 - 0,6

Slechts 1 oever heeft een natuurlijk profiel; gewenste soorten zijn aanwezig, maar in mindere mate.

### Ecologische kwaliteit van oppervlaktewater - beken



Kwaliteit: ontoereikend  
EKR: 0,2 - 0,4

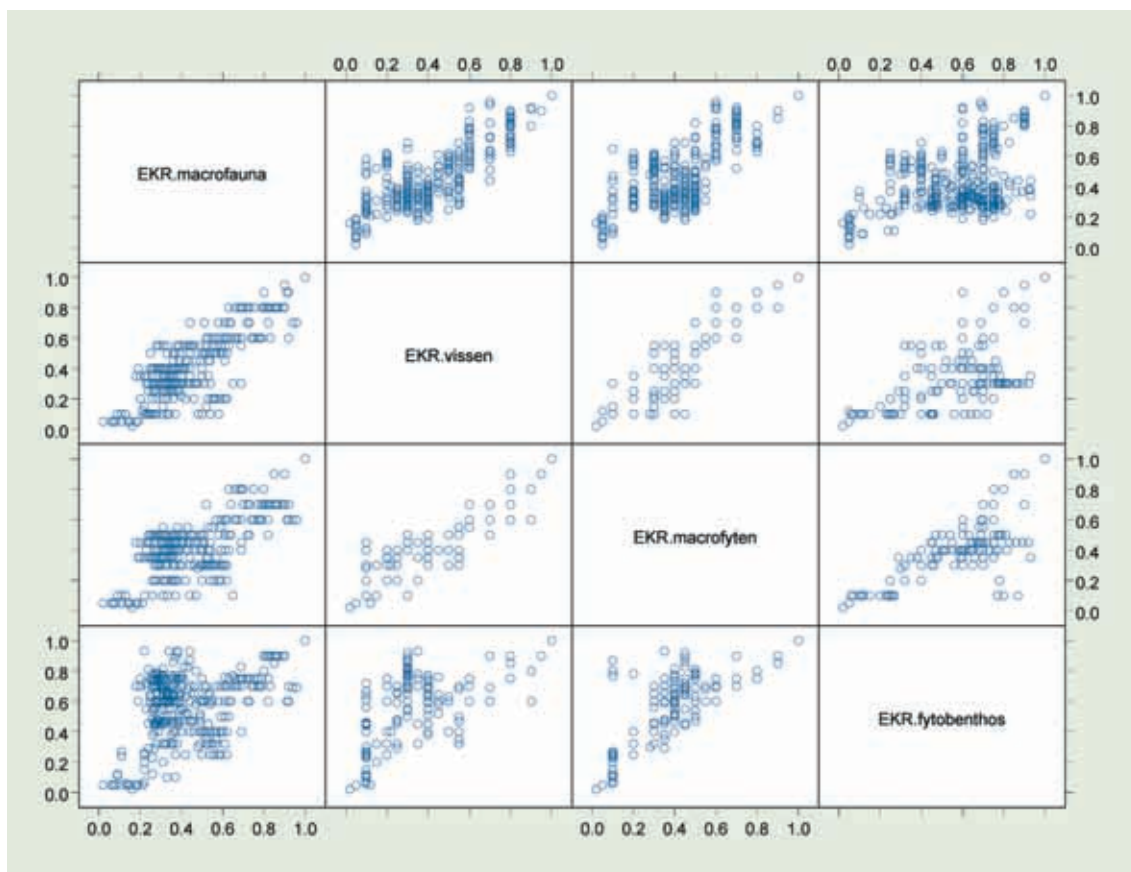
Sommige beken zijn functionele kanalen waardoor de variatie in oeverstructuur is verdwenen.



Kwaliteit: slecht  
EKR: < 0,2

In deze categorie beken is de waterkwaliteit slecht: dieren kunnen hier niet of nauwelijks in overleven.

Figuur 4.1C Voorbeelden van beken in vier verschillende ecologische toestanden (zoals gegeven in Figuur 1.1, derde laag). Genoemde EKR-klassen zijn berekend volgens de KRW-methodiek (zie ook Figuur 1.3). Bron: PBL (2008).



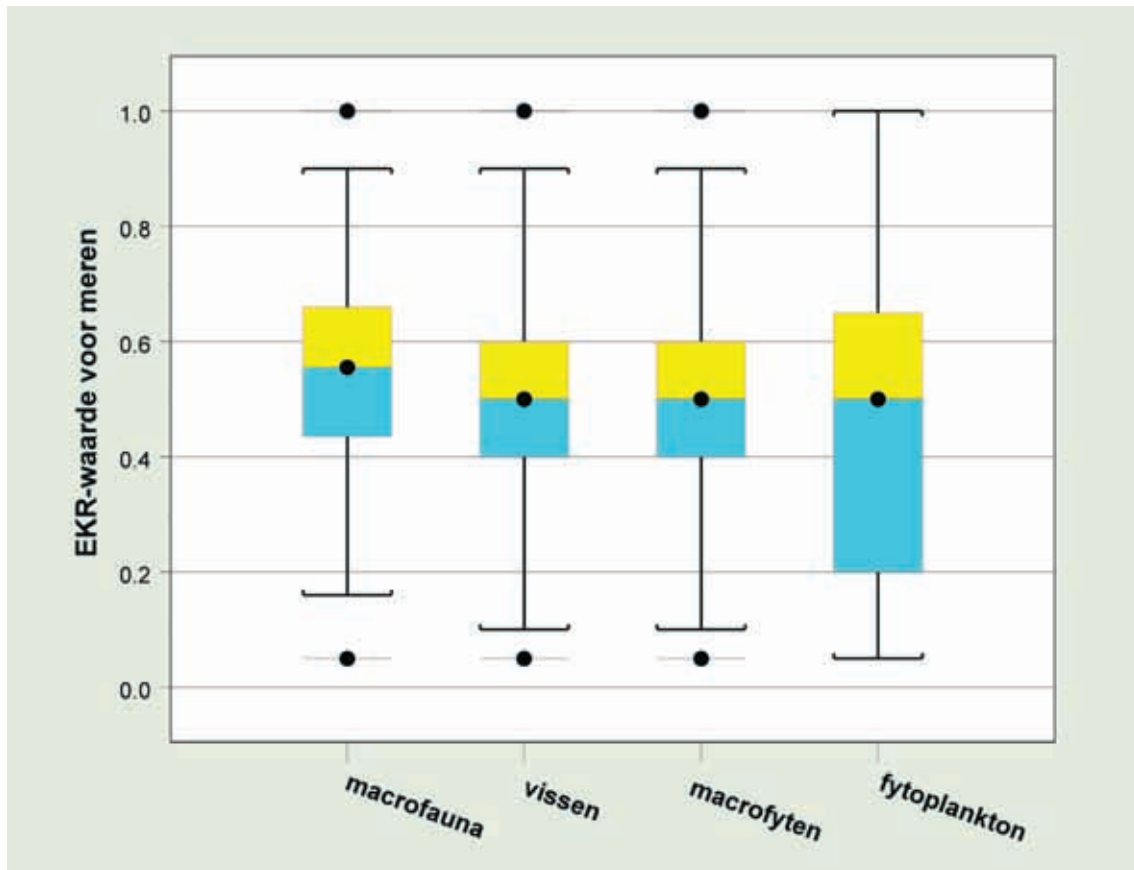
**Figuur 4.1D** Scatterplotmatrix voor vier deelmaatlaten: 'macrofauna', 'vissen', 'macrofyten' en 'fytobenthos'. De correlatiematrix voor de doelmaatlaten is:

	EKR macrofauna	EKR vissen	EKR macrofyten	EKR fytobenthos
EKR macrofauna	1.00	0.81	0.68	0.43
EKR vissen	0.81	1.00	0.82	0.56
EKR macrofyten	0.68	0.82	1.00	0.78
EKR fytobenthos	0.43	0.56	0.78	1.00

De relatie tussen de vier deelmaatlaten in beken is gegeven in Figuur 4.1D. In het onderschrift bij deze figuur is ook de correlatiematrix gegeven. Het blijkt uit deze correlatiematrix dat de deelmaatlaten 'vissen' en 'macrofyten' hoog gecorreleerd zijn ( $R=0.82$ ). Dat blijkt ook uit hun gezamenlijke scatterplot in het centrum van de scatterplotmatrix: de puntenwolk ligt zeer nauw rond de 45-graden-lijn (die loopt van (0,0,0,0) naar (1,0,1,0)). De deelmaatlaten 'macrofauna' en 'fytobenthos' zijn zeer laag gecorreleerd ( $R=0.43$ ). Dat blijkt ook uit hun scatterplot die de grootste puntenwolk laat zien (scatterplot linksonder en rechtsboven).

De scatterplotmatrix laat zien dat een aantal deelmaatlaten gelijksoortige informatie bevatten (correlaties hoger dan 0,80), maar anderen juist een heel verschillend gedrag (correlaties lager dan 0,50).





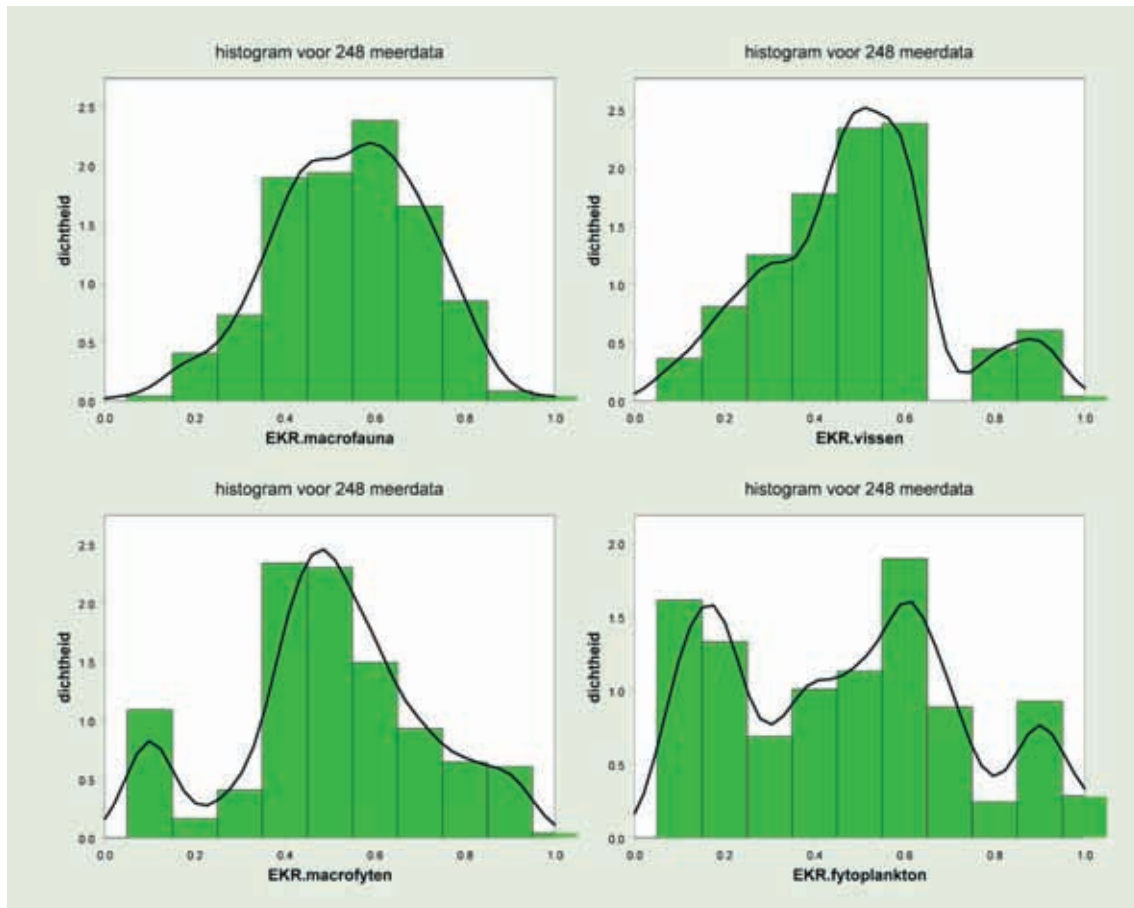
Figuur 4.2A Overzicht ecologische toestand van 248 meergegevens: boxplots.

## 4.2 Meren

Figuur 4.2A geeft boxplots voor de vier deelmaatlaten. Het blijkt dat EKR-waarden voor 'fytoplankton' de hoogste spreiding vertonen en vooral liggen in de klassen ontoereikende, matige en goede ecologische toestand. De overige deelmaatlaten liggen overwegend in de klasse 'matig'. De boxplots laten verder zien dat er incidenteel waarden bij 0.1 en 1.0 voorkomen (de whiskers en uitbijters).

Figuur 4.2B geeft de histogrammen met dichtheden. Ook nu valt op dat de hoogste spreiding in EKR-waarden gevonden worden voor 'fytoplankton'.

Figuur 4.2C geeft een indruk hoe meren met een wisselende ecologische kwaliteit er in de praktijk uit kunnen zien. Bij de beoordeling zijn alle deelmaatlaten betrokken (de slechtste EKR-waarde is daarbij maatgevend).



Figuur 4.2B Overzicht ecologische toestand van 248 meergegevens: histogrammen en dichtheden.

### Ecologische kwaliteit van oppervlaktewater - ondiepe plassen



Kwaliteit: goed  
EKR: 0,6 - 0,8

Kenmerkend is helder water en een diverse samenstelling van planten en dieren.



Kwaliteit: matig  
EKR: 0,4 - 0,6

Een overgang tussen land en water met soms groen en soms stenen is matig bij een natuurlijk meer.

### Ecologische kwaliteit van oppervlaktewater - ondiepe plassen



Kwaliteit: ontoereikend  
EKR: 0,2 - 0,4

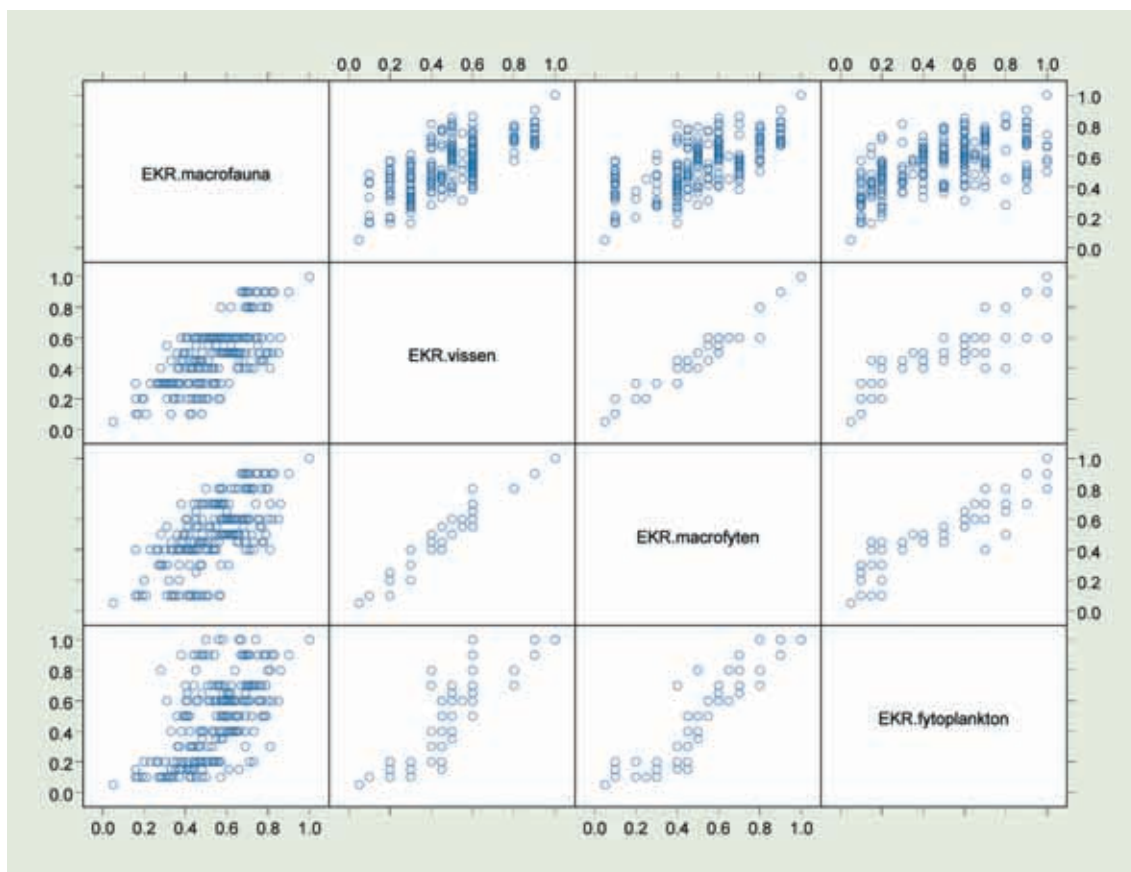
Kenmerkend is troebel water: algen en geen waterplanten. De visstand bestaat vooral uit brasem.



Kwaliteit: slecht  
EKR: < 0,2

Kenmerkend is bovenmatige algengroei waardoor verstikking optreedt.

Figuur 4.2C Voorbeelden van meren in vier verschillende ecologische toestanden (zoals gegeven in Figuur 1.1, derde laag). Genoemde EKR-klassen zijn berekend volgens de KRW-methodiek (zie ook Figuur 1.3). Bron: PBL (2008).

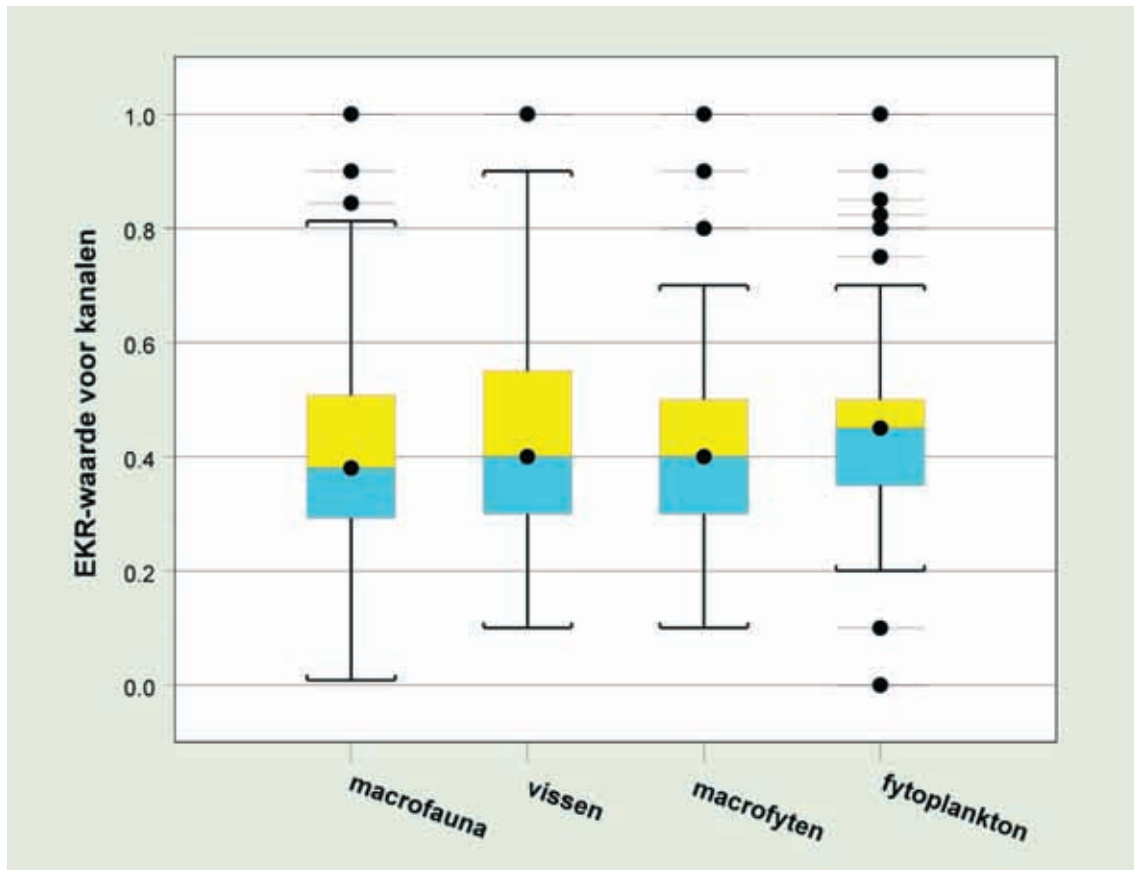


**Figuur 4.2D** Scatterplotmatrix voor vier deelmaatlaten: 'macrofauna', 'vissen', 'macrofyten' en 'fytoplankton'. De correlatiematrix voor de doelmaatlaten is:

	EKR macrofauna	EKR vissen	EKR macrofyten	EKR fytoplankton
EKR macrofauna	1.00	0.66	0.63	0.58
EKR vissen	0.66	1.00	0.96	0.87
EKR macrofyten	0.63	0.96	1.00	0.91
EKR fytoplankton	0.58	0.87	0.91	1.00

De relatie tussen de vier deelmaatlaten in meren is gegeven in Figuur 4.2D. In het onderschrift bij deze figuur is ook de correlatiematrix gegeven. De correlatiematrix laat zien dat de deelmaatlaten 'vissen' en 'macrofyten' zeer hoog gecorreleerd zijn (= 0.96). Dat blijkt ook uit hun gezamenlijke scatterplot in het centrum van de scatterplotmatrix: de puntenwolk ligt zeer nauw rond de 45-graden-lijn (die loopt van (0.0,0.0) naar (1.0,1.0)).

De deelmaatlaten 'macrofauna' en 'fytoplankton' zijn laag gecorreleerd (=0.58). Dat blijkt ook uit hun scatterplot die de grootste puntenwolk laat zien (scatterplot linksonder en rechtsboven).



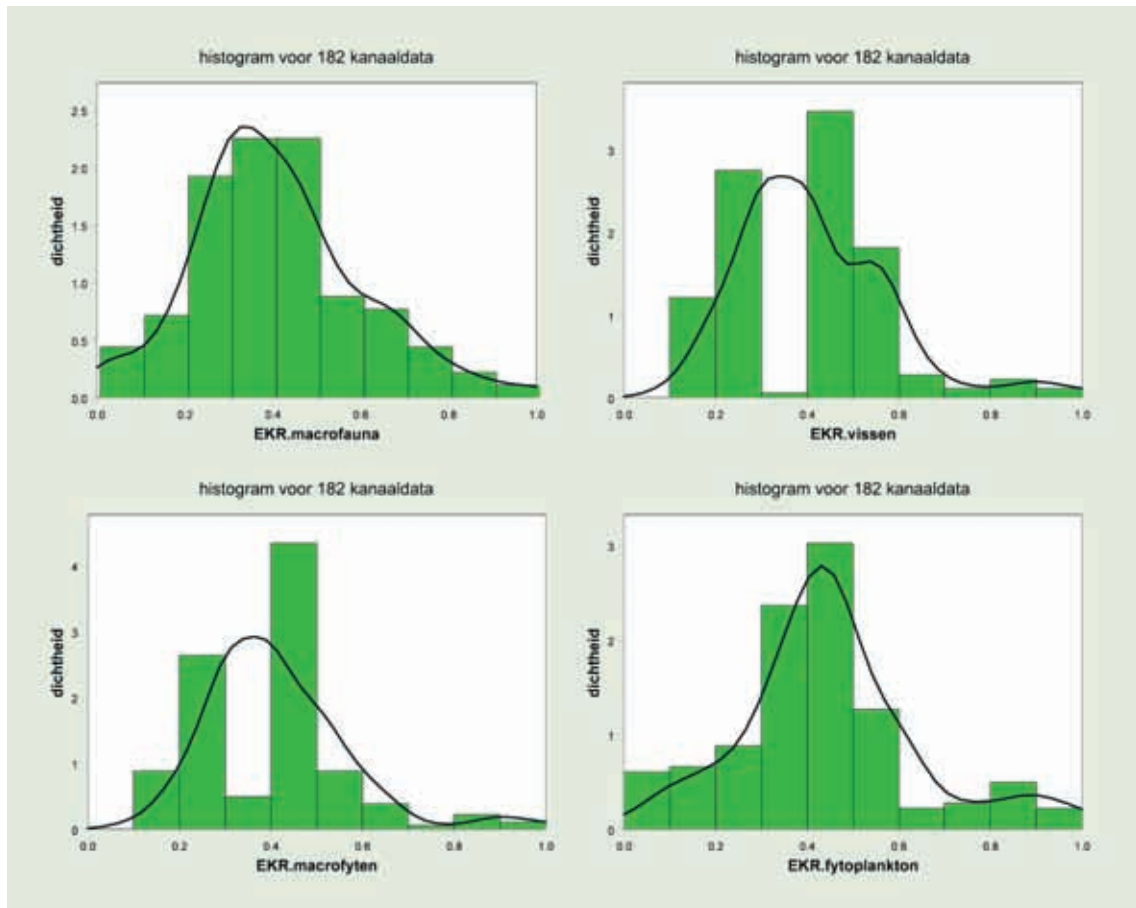
Figuur 4.3A Overzicht ecologische toestand van 182 kanaalgegevens: boxplots.

### 4.3 Kanalen

Figuur 4.3A geeft boxplots voor de vier deelmaatlatten. Het blijkt dat voor alle deelmaatlatten de ecologische toestand overwegend ligt in de klassen 'ontoereikend' en 'matig'. De boxplots laten verder zien dat er incidenteel waarden bij 0.1 en 1.0 voorkomen (whiskers en uitbijters). Vooral de grote gelijkens van de vier boxplots valt op.

Figuur 4.3B geeft de histogrammen met dichtheden. De dichtheden vertonen een sterk gelijkend patroon, overeenkomstig de boxplots in Figuur 4.3A.

Figuur 4.3C geeft een indruk hoe kanalen met een wisselende ecologische kwaliteit er in de praktijk uit kunnen zien. Bij de beoordeling zijn alle deelmaatlatten betrokken (de slechtste EKR-waarde is daarbij maatgevend).



Figuur 4.3B Overzicht ecologische toestand van 182 kanaalgegevens: histogrammen en dichtheden.

### Ecologische kwaliteit van oppervlaktewater - kanalen



Kwaliteit: goed  
EKR: 0,6 - 0,8

Kenmerkend voor kanalen zijn rechte lijnen. De helderheid van het water en de ontwikkeling van de oevers hangt af van het gebruik.



Kwaliteit: matig  
EKR: 0,4 - 0,6

Een stalen vooroever is niet optimaal: liever natuurlijk materiaal en een geleidelijke overgang tussen land en water.

### Ecologische kwaliteit van oppervlaktewater - kanalen



Kwaliteit: ontoereikend  
EKR 0,2 - 0,4

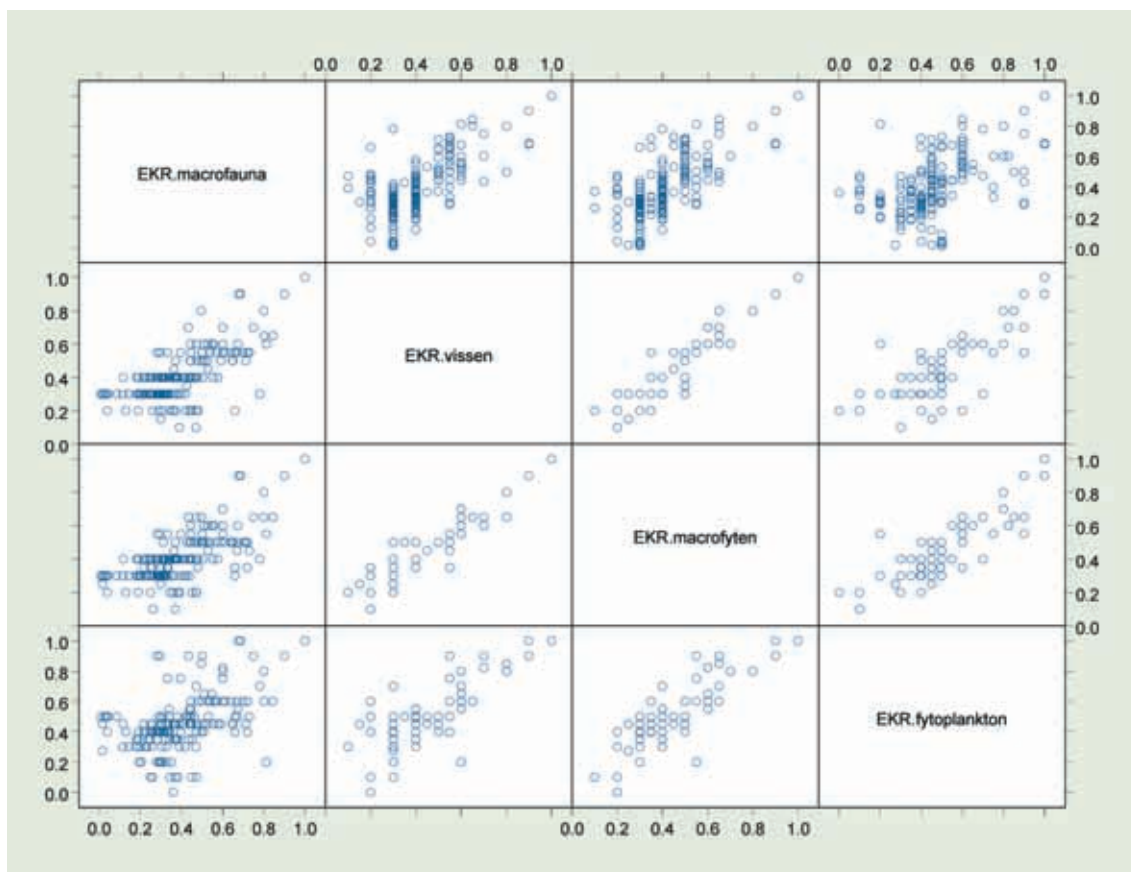
Stalen damwanden met groene oevers bieden geringe kansen voor de waternatuur.



Kwaliteit: slecht  
EKR: < 0,2

Stalen damwanden zonder groene oevers bieden nauwelijks kansen voor de natuur.

Figuur 4.3C Voorbeelden van kanalen in vier verschillende ecologische toestanden (zoals gegeven in Figuur 1.1, derde laag). Genoemde EKR-klassen zijn berekend volgens de KRW-methodiek (zie ook Figuur 1.3). Bron: PBL (2008).



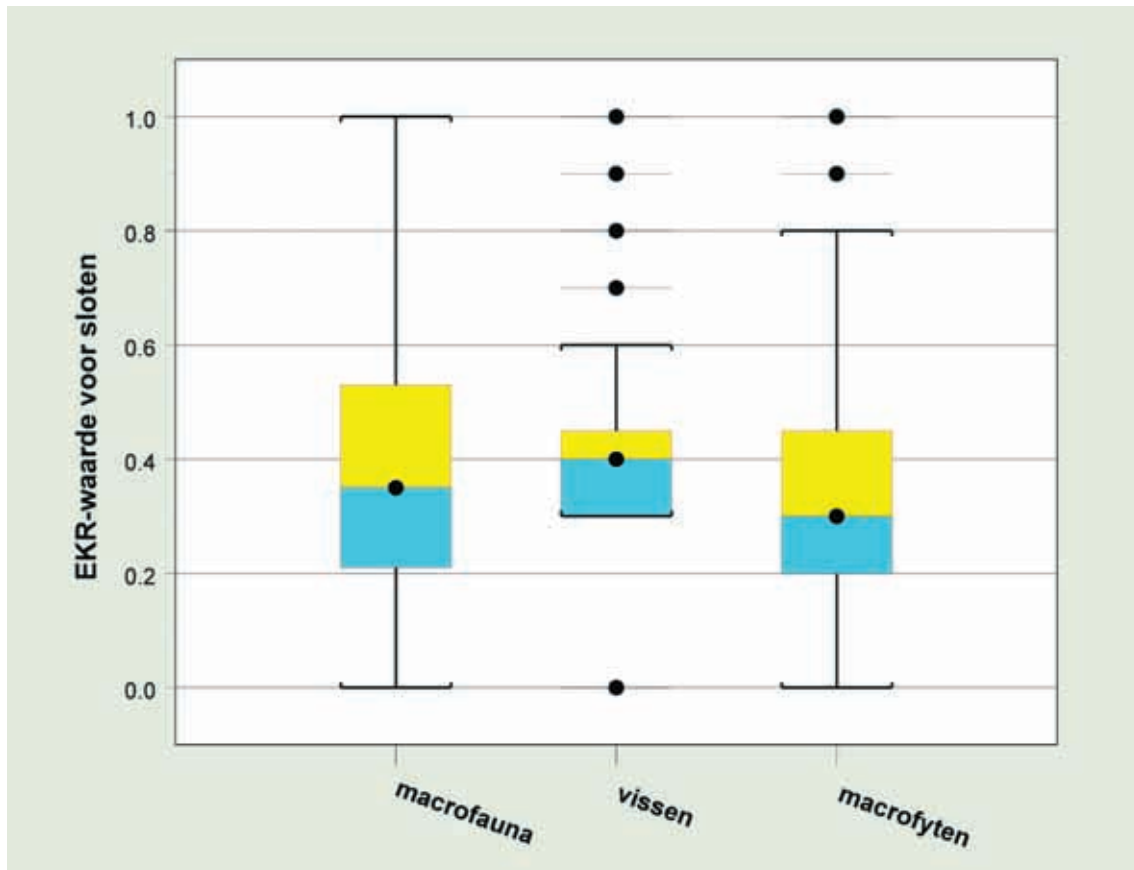
**Figuur 4.3D** Scatterplotmatrix voor vier deelmaatlaten: 'macrofauna', 'vissen', 'macrofyten' en 'fytoplankton'. De correlatiematrix voor de doelmaatlaten is:

	EKR macrofauna	EKR vissen	EKR macrofyten	EKR fytoplankton
EKR macrofauna	1.00	0.70	0.69	0.51
EKR vissen	0.70	1.00	0.96	0.78
EKR macrofyten	0.69	0.96	1.00	0.84
EKR fytoplankton	0.51	0.78	0.84	1.00

De relatie tussen de vier deelmaatlaten in kanalen is gegeven in Figuur 4.3D. In het onderschrift bij deze figuur is ook de correlatiematrix gegeven. Het blijkt dat de deelmaatlaten 'vissen' en 'macrofyten' zeer hoog gecorreleerd zijn ( $R=0.96$ ). Dat blijkt ook uit hun gezamenlijke scatterplot in het centrum van de scatterplotmatrix: de puntenwolk ligt zeer nauw rond de 45-gradenlijn (die loopt van (0.0,0.0) naar (1.0,1.0)).

De deelmaatlaten 'macrofauna' en 'fytoplankton' zijn laag gecorreleerd ( $R=0.51$ ). Dat blijkt ook uit hun scatterplot die de grootste puntenwolk laat zien (scatterplot linksonder en rechtsboven).





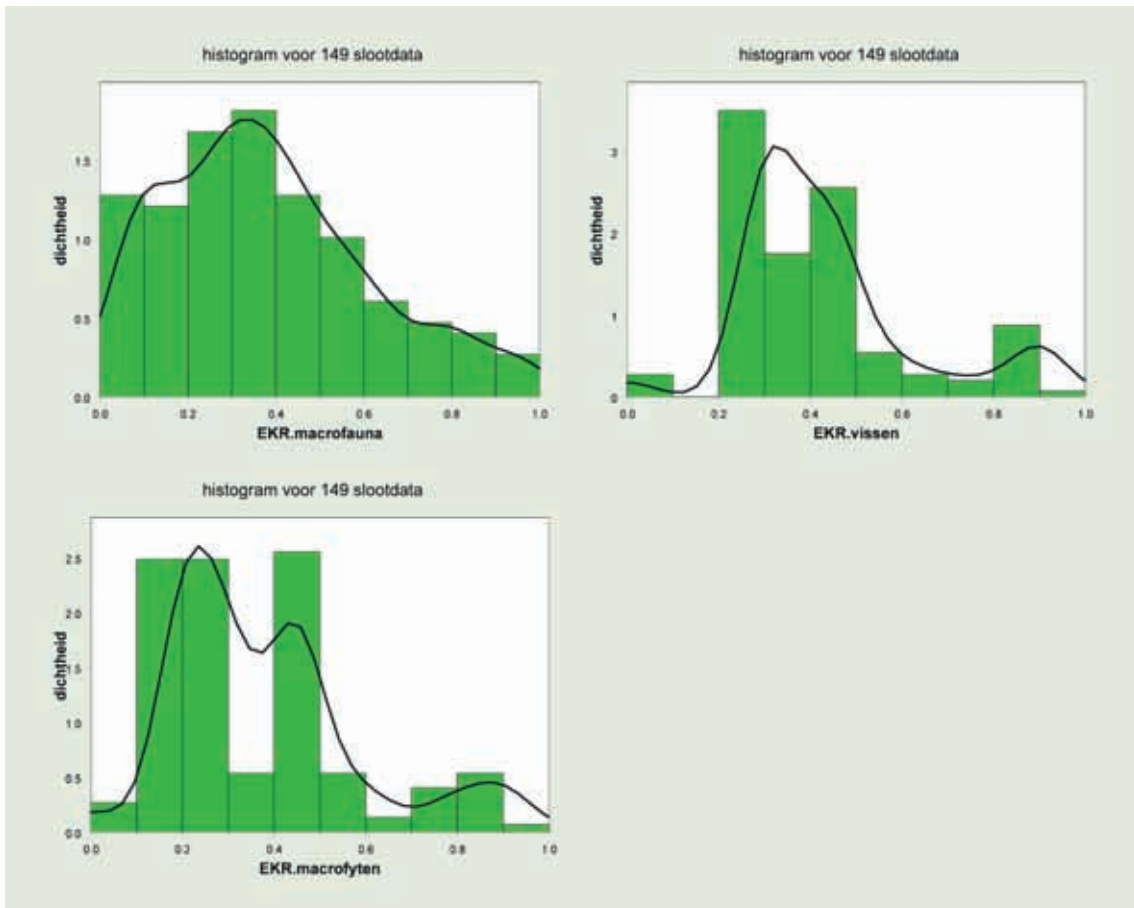
Figuur 4.4A Overzicht ecologische toestand van 149 slotgegevens: boxplots.

#### 4.4 Sloten

Figuur 4.4A geeft boxplots voor de drie beschikbare deelmaatlatten. Het blijkt dat voor elke deelmaatlat de EKR-waarden liggen in de klassen 'ontoereikend' en 'matig'. De grootste spreiding geven de gevonden waarden voor 'macrofauna'. De boxplots laten verder zien dat er incidenteel waarden bij 0.0 en 1.0 voorkomen (whiskers en uitbijters).

Figuur 4.4B geeft de histogrammen met dichtheden. De EKR-waarden voor 'macrofauna' vertonen de grootste spreiding van de drie deelmaatlatten.

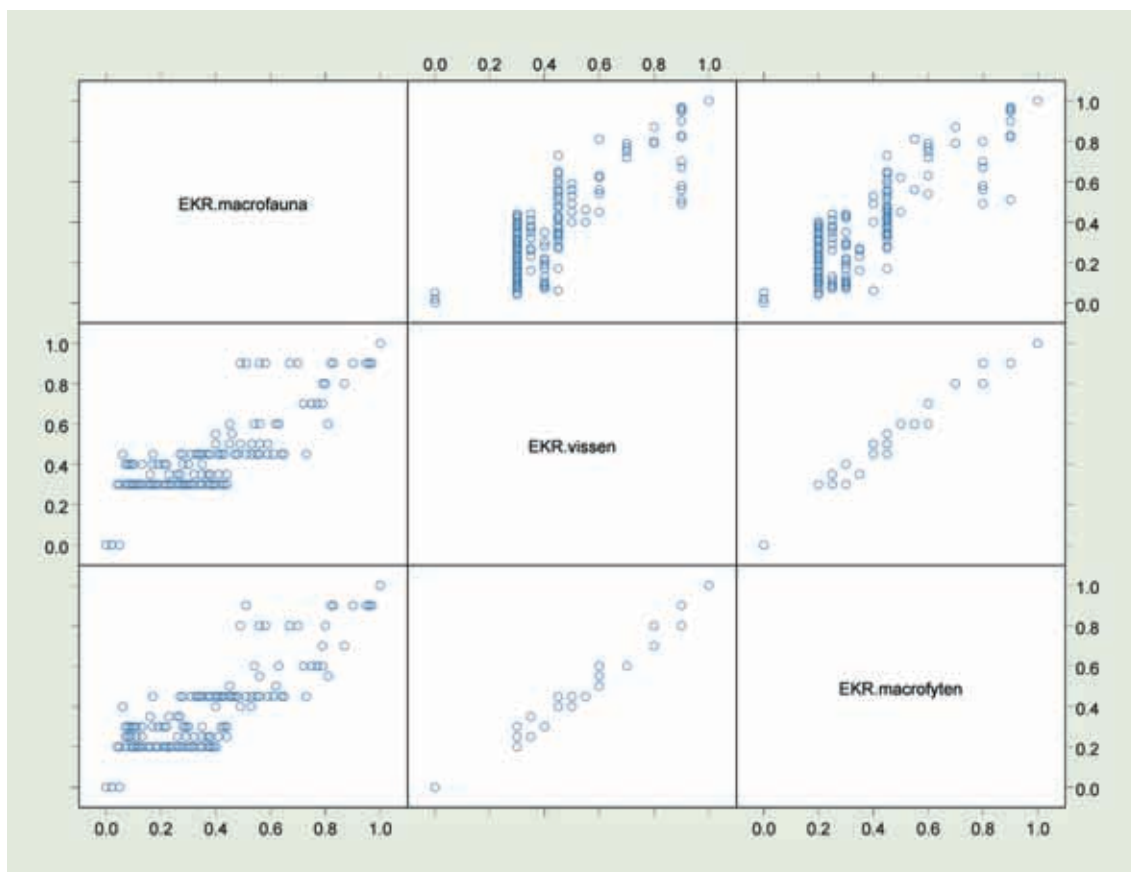
Figuur 4.4C geeft een indruk hoe sloten met een wisselende ecologische kwaliteit er in de praktijk uit kunnen zien. Bij de beoordeling zijn alle deelmaatlatten betrokken (de slechtste EKR-waarde is daarbij maatgevend).



**Figuur 4.4B** Overzicht ecologische toestand van 149 slootgegevens: histogrammen met dichtheden.



**Figuur 4.4C** Voorbeelden van sloten in vier verschillende ecologische toestanden (zoals gegeven in Figuur 1.1, derde laag). Genoemde EKR-klassen zijn berekend volgens de KRW-methodiek (zie ook Figuur 1.3). Bron: PBL (2008).



**Figuur 4.4D** Scatterplotmatrix voor vier deelmaatlaten: 'macrofauna', 'vissen' en 'macrofyten'. Voor een methodische uitleg van de figuur zie paragraaf 3.1.1. De correlatiematrix voor de deelmaatlaten is:

	EKR macrofauna	EKR vissen	EKR macrofyten
EKR macrofauna	1.00	0.81	0.83
EKR vissen	0.81	1.00	0.97
EKR macrofyten	0.83	0.97	1.00

De relatie tussen de vier deelmaatlaten in sloten is gegeven in Figuur 4.4D. In het onderschrift bij deze figuur is ook de correlatiematrix gegeven. Het blijkt dat de deelmaatlaten 'vissen' en 'macrofyten' zeer hoog gecorreleerd zijn ( $R=0.97$ ). Dat blijkt ook uit hun gezamenlijke scatterplot, rechtsonder in de scatterplotmatrix: de puntenwolk ligt zeer nauw rond de 45-graden-lijn (die loopt van (0.0,0.0) naar (1.0,1.0)).

Ook de overige correlaties zijn hoog:  $R=0.81$  en  $R=0.83$ . De puntenwolken zijn wel duidelijk groter dan voor 'vissen' en 'macrofyten'.

## 5 Resultaten regressieboomanalyse

In dit hoofdstuk worden de schattingsresultaten met regressiebomen samengevat. Gedetailleerde resultaten zijn gegeven in de Appendix A (voor beken), B (voor meren), C (voor kanalen) en D (voor sloten). Opgemerkt zij dat alle resultaten gelden voor de complete datasets, inclusief de doublures zoals beschreven aan het begin van paragraaf 2.2. De invloed van doublures op de geschatte regressiebomen word kort toegelicht in Appendix F.

### 5.1 Beken

De resultaten van de regressiebomen voor beken zijn in detail gegeven in Appendix A en voor de belangrijkste splitsingen samengevat in Tabel 5.1. De tabel moet zo gelezen worden dat de opeenvolgende stuurvariabelen alle opgevolgd moeten worden om uiteindelijk te komen tot de hoogst gemiddelde EKR-deelmaatlatwaarde. Dus om bijvoorbeeld voor ‘vissen’ een gemiddelde ecologisch toestand van 0.81 te halen, moet een beek vallen in de meanderings-classes 3, 4 of 5, een BZV-waarde hebben van minder dan 4.3 ml O<sub>2</sub>/l, vallen in de verstuwingsclasses 2 of 3, en, ten slotte, een concentratie van totaal fosfor hebben van lager dan 0.06 mg P/l.

De verklarende kracht van de vier regressiebomen blijkt zeer hoog te liggen, namelijk tussen de 81% en 90%, zoals blijkt uit de laatste kolom van Tabel 5.1.

Uit Tabel 5.1 kan verder afgelezen worden waar beken aan moeten voldoen om voor elk van de vier deelmaatlaten een goede tot zeer goede kwaliteit te halen. Anders gezegd, onder welke voorwaarden geldt dat alle EKR's  $\geq 0.60$ ? De tabel laat zien dat aan de volgende vier voorwaarden voldaan moet worden:

**Tabel 5.1 Samenvatting van de regressiebomen uit Appendix A. De getallen tussen [.] geven de gemiddelde deelmaatlatwaarde na genoemde splitsing. De eerste splitsing geeft de grootste verbetering (dus op de hoogste EKR-waarde), et cetera. Bovendien gaat de tweede splitsing verder op het beste alternatief uit de eerste splitsing, et cetera. De eenheid voor 'Totaal P' is mg P/l, voor 'Totaal N' mg N/l en voor het biologisch zuurstof verbruik (BZV) mg O<sub>2</sub> /l. De kleurcode is gegeven in de linker balk uit Figuur 1.3. De verklarende kracht in de laatste kolom geldt voor de gehele regressieboom (zie Figuur A.2, A.6, A.10 en A.14) en is berekend als de percentuele reductie in deviantie in de beginknoop van de boom ten opzichte van de som der devianties in de eindknoten.**

EKR deelmaatlat beken	Eerste splitsing	Tweede splitsing	Derde splitsing	Vierde splitsing	Verklarende kracht
<b>Macrofauna</b> [0.45]	Meandering=1,2 [0.32] Meandering=3,4,5 [0.62]	Totaal P > 0.07 [0.58] Totaal P < 0.07 [0.84]			84%
<b>Vissen</b> [0.42]	Meandering=1,2 [0.29] Meandering=3,4,5 [0.58]	BZV > 4.3 [0.24] BZV < 4.3 [0.64]	Verstuwung=1 [0.31] Verstuwung=2,3 [0.66]	Totaal P > 0.06 [0.63] Totaal P < 0.06 [0.81]	81%
<b>Macrophyten</b> [0.43]	Totaal P > 0.83 [0.17] Totaal P < 0.83 [0.48]	Verstuwung=1,2 [0.42] Verstuwung=3 [0.59]	Totaal P > 0.16 [0.36] Totaal P < 0.16 [0.66]	Totaal P > 0.10 [0.58] Totaal P < 0.10 [0.74]	84%
<b>Fytobenthos</b> [0.58]	Totaal P > 0.27 [0.33] Totaal P < 0.27 [0.69]	Totaal P > 0.14 [0.58] Totaal P < 0.14 [0.74]			90%

1. Meandering in de klassen 3, 4 of 5;
2. BZV lager dan 4.3 ml O<sub>2</sub>/l ;
3. Verstuwings in de klasse 3;
4. Totaal fosfor lager dan 0.16 mg P/l .

Het is interessant om bovenstaande grenswaarde voor totaal fosfor te vergelijken met de gestelde werknormen en MTR-waarde (volgens de Vierde Nota Waterhuishouding). Zie Tabel 2.3 in PBL (2008). Voor beken ligt de GET-norm tussen de 0.12 en 0.14 mg P/l. De MTR-waarde voor regionale wateren bedraagt 0.15 mg P/l. De gevonden grens van 0.16 mg P/l komt zeer goed overeen met deze normwaarden.

Conclusie uit Tabel 5.1 is dat een goede meandering en lage totaal-fosforconcentraties leiden tot een goede ecologische kwaliteit. In iets mindere mate spelen verstuwings en lage BZV-waarden een rol.

## 5.2 Meren

De resultaten van de regressiebomen voor meren zijn in detail gegeven in Appendix B en samengevat in Tabel 5.2. De tabel moet zo gelezen worden dat de opeenvolgende stuurvariabelen alle opgevolgd moeten worden om uiteindelijk te komen tot de hoogst gemiddelde deelmaatlatwaarde. Dus om bijvoorbeeld voor vissen een gemiddelde ecologisch toestand van [0.68] te halen, moet een meer een Totaal-N-concentratie hebben lager dan 1.8 mg N/l en een concentratie van Totaal P van minder dan 0.05 mg P/l.

De verklarende kracht van de vier regressiebomen blijkt zeer hoog te liggen, namelijk tussen de 71% en 93%, zoals blijkt uit de laatste kolom van Tabel 5.2.

**Tabel 5.2 Samenvatting van de regressiebomen uit Appendix B. De getallen tussen [.] geven de gemiddelde deelmaatlatwaarde na genoemde splitsing. De eerste splitsing geeft de grootste verbetering, et cetera. Bovendien gaat de tweede splitsing verder op het beste alternatief uit de eerste splitsing, et cetera. De eenheid voor 'Totaal P' is mg P/l en voor 'Totaal N' mg N/l. De kleurcode is gegeven in de linker balk uit Figuur 1.3. De verklarende kracht in de laatste kolom geldt voor de gehele regressieboom (zie Figuur B.2, B.6, B.10 en B.14) en is berekend als de percentuele reductie in deviantie in de beginknoop van de boom ten opzichte van de som der devianties in de eindknoten.**

EKR deelmaatlat meren	Eerste splitsing	Tweede splitsing	Voorspelkracht
Macrofauna [0.54]	Totaal N > 2.1 [0.41] Totaal N < 2.1 [0.62]	Totaal P > 0.08 [0.57] Totaal P < 0.08 [0.68]	71%
Vissen [0.49]	Totaal N > 1.8 [0.33] Totaal N < 1.8 [0.61]	Totaal P > 0.05 [0.56] Totaal P < 0.05 [0.81]	92%
Macrofyten [0.51]	Totaal N > 1.7 [0.36] Totaal N < 1.7 [0.68]	Totaal P > 0.05 [0.62] Totaal P < 0.05 [0.85]	93%
Fytoplankton [0.47]	Totaal N > 1.8 [0.26] Totaal N < 1.8 [0.68]	Totaal P > 0.04 [0.63] Totaal P < 0.04 [0.90]	92%

Uit Tabel 5.2 kan verder afgelezen worden waar meren aan moeten voldoen om voor elk van de vier deelmaatlaten een goede tot zeer goede kwaliteit te halen. Ofwel, onder welke voorwaarden geldt dat alle EKR's  $\geq 0.60$ ? De tabel laat zien dat slechts aan één voorwaarde voldaan moet worden:

1. Totaal stikstof lager dan 1.7 mg N/l.

De hier gevonden grenswaarde voor totaal stikstof kan vergeleken worden met de gestelde werknormen en MTR-waarde (volgens de Vierde Nota Waterhuishouding). Zie Tabel 2.3 in PBL (2008). Voor meren ligt de stikstof-GET-norm tussen de 1.0 en 1.3 mg N/l. De MTR-waarde voor regionale wateren bedraagt 2.2 mg N/l. De hier gevonden grens van 1.7 mg N/l is dus soepeler dan de GET-waarden, maar strenger dan de geldende MTR-waarde.

Conclusie uit Tabel 5.2 is dat lage totaal-stikstofconcentraties voldoende zijn om EKR-waarden hoger dan 0.60 te bereiken. In veel mindere mate speelt totaal fosfor een rol.

### 5.3 Kanalen

De resultaten van de regressiebomen voor kanalen zijn in detail gegeven in Appendix C en samengevat in Tabel 5.3. De tabel moet zo gelezen worden dat de opeenvolgende stuurvariabelen alle opgevolgd moeten worden om uiteindelijk te komen tot de hoogst gemiddelde deelmaatlatwaarde. Dus om bijvoorbeeld voor vissen een gemiddelde ecologisch toestand van [0.91] te halen, moet een kanaal een Totaal-N-concentratie hebben lager dan 1.3 mg N/l.

De verklarende kracht van de vier regressiebomen blijkt hoog te liggen, namelijk tussen de 67% en 94%, zoals blijkt uit de laatste kolom van Tabel 5.3.

Uit Tabel 5.3 kan verder afgelezen worden waar kanalen aan moeten voldoen om voor elk van de vier deelmaatlaten een goede tot zeer goede kwaliteit te halen. Ofwel, onder welke voorwaarden geldt dat alle EKR's  $\geq 0.60$ ? De tabel laat zien dat aan de volgende twee voorwaarden voldaan moet worden:

1. Totaal stikstof lager dan 2.8 mg N/l ;
2. Totaal fosfor lager dan 0.12 mg P/l .

Het is interessant om de hier gevonden grenswaarden voor totaal stikstof en totaal fosfor te vergelijken met de gestelde werknormen en MTR-waarden (volgens de Vierde Nota Waterhuishouding). Zie Tabel 2.3 in PBL (2008). Voor kanalen ligt de fosfor-GEP-norm tussen de 0.15 en 0.25 mg P/l. De MTR-waarde voor regionale wateren bedraagt 0.15 mg P/l. De hier gevonden grens van 0.12 mg P/l ligt iets onder de normwaarden.

De stikstof-GEP-norm voor kanalen ligt tussen de 2.8 en 3.8 mg N/l. De MTR-waarde voor regionale wateren bedraagt 2.2 mg N/l. De hier gevonden grens van 2.8 mg N/l ligt boven de MTR-norm en is gelijk aan de ondergrens van de GEP-norm. Deze resultaten stemmen dus goed overeen.

**Tabel 5.3 Samenvatting van de regressiebomen uit Appendix C. De getallen tussen [.] geven de gemiddelde deelmaatlatwaarde na genoemde splitsing. De eerste splitsing geeft de grootste verbetering, et cetera. Bovendien gaat de tweede splitsing verder op het beste alternatief uit de eerste splitsing, et cetera. De eenheid voor 'Totaal P' is mg P/l en voor 'Totaal N' mg N/l. De kleurcode is gegeven in de rechter balk uit Figuur 1.3. De verklarende kracht in de laatste kolom geldt voor de gehele regressieboom (zie Figuur C.2, C.6, C.10 en C.14) en is berekend als de percentuele reductie in deviantie in de beginknoop van de boom ten opzichte van de som der devianties in de eindknoten.**

EKR deelmaatlat kanalen	Eerste splitsing	Tweede splitsing	Voorspelkracht
Macrofauna [0.41]	Totaal N > 2.8 [0.36]	Totaal N > 1.3 [0.57]	67%
	Totaal N < 2.8 [0.62]	Totaal N < 1.3 [0.85]	
Vissen [0.41]	Totaal N > 3.0 [0.35]	Totaal N > 1.3 [0.58]	87%
	Totaal N < 3.0 [0.63]	Totaal N < 1.3 [0.91]	
Macrofyten [0.41]	Totaal N > 3.0 [0.35]	Totaal N > 1.3 [0.56]	86%
	Totaal N < 3.0 [0.62]	Totaal N < 1.3 [0.91]	
Fytoplankton [0.45]	Totaal P > 0.12 [0.40]	Totaal P > 0.05 [0.75]	94%
	Totaal P < 0.12 [0.85]	Totaal P < 0.05 [0.93]	

Conclusie uit Tabel 5.3 is dat lage totaal-stikstofconcentraties voldoende zijn om EKR-waarden hoger dan 0.60 te bereiken voor drie van de vier deelmaatlaten. Voor 'fytoplankton' speelt totaal fosfor de hoofdrol.

## 5.4 Sloten

De resultaten van de regressiebomen voor beken zijn in detail gegeven in Appendix D en samengevat in Tabel 5.4. De tabel moet zo gelezen worden dat de opeenvolgende sturingsvariabelen alle opgevolgd moeten worden om uiteindelijk te komen tot de hoogst gemiddelde deelmaatlatwaarde. Dus om bijvoorbeeld voor vissen een gemiddelde ecologisch toestand van [0.89] te halen, moet een sloot een peilbeheer hebben in klasse 3 en een concentratie van totaal P van minder dan 0.23 mg P/l.

De verklarende kracht van de vier regressiebomen blijkt zeer hoog te liggen, namelijk tussen de 80% en 95%, zoals blijkt uit de laatste kolom van Tabel 5.4.

Uit Tabel 5.1 kan verder afgelezen worden waar sloten aan moeten voldoen om voor elk van de vier deelmaatlaten een goede tot zeer goede kwaliteit te halen. Ofwel, onder welke voorwaarden geldt dat alle EKR's  $\geq 0.60$ ? De tabel laat zien dat er slechts aan één voorwaarde voldaan moet worden:

3. Peilbeheer in klasse 3.

Een verdere verbetering kan daarna nog worden bereikt via totaal-fosfor-concentraties lager dan 0.17 mg P/l en totaal- stikstof-concentraties lager dan 2.2 mg N/l . Deze grenswaarden voor totaal stikstof en totaal fosfor kunnen vergeleken worden met de gestelde werknormen en MTR-waarden (volgens de Vierde Nota Waterhuishouding). Zie Tabel 2.3 in PBL (2008). Voor sloten ligt de fosfor-GEP-norm tussen de 0.22 en 0.50 mg P/l. De MTR-waarde voor regionale wateren



**Tabel 5.4 Samenvatting van de regressiebomen uit Appendix D. De getallen tussen [.] geven de gemiddelde deelmaatlatwaarde na genoemde splitsing. De eerste splitsing geeft de grootste verbetering, et cetera. Bovendien gaat de tweede splitsing verder op het beste alternatief uit de eerste splitsing, et cetera. De eenheid voor 'Totaal P' is mg P/l en voor 'Totaal N' mg N/l. De kleurcode is gegeven in de rechter balk uit Figuur 1.3. De verklarende kracht in de laatste kolom geldt voor de gehele regressieboom (zie Figuur D.2, D.6 en D.10) en is berekend als de percentuele reductie in deviantie in de beginknoop van de boom ten opzichte van de som der devianties in de eindknoten.**

EKR deelmaatlat sloten	Eerste splitsing	Tweede splitsing	Voorspelkracht
Macrofauna [0.38]	Peilbeheer=1,2 [0.31]	Totaal N > 2.2 [0.64]	80%
	Peilbeheer=3 [0.74]	Totaal N < 2.2 [0.85]	
Vissen [0.44]	Peilbeheer=1,2 [0.37]	Totaal P > 0.23 [0.64]	95%
	Peilbeheer=3 [0.81]	Totaal P < 0.23 [0.89]	
Macrofyten [0.38]	Peilbeheer=1,2 [0.31]	Totaal P > 0.17 [0.58]	94%
	Peilbeheer=3 [0.75]	Totaal P < 0.17 [0.87]	

bedraagt 0.15 mg P/l. De hier gevonden grens van 0.17 mg P/l is dus vrijwel gelijk aan de MTR-waarde maar strenger dan de GEP-norm-range.

De stikstof-GEP-norm voor sloten ligt op 2.4 mg N/l. De MTR-waarde voor regionale wateren bedraagt 2.2 mg N/l. De hier gevonden grens van 2.2 mg N/l is identiek aan de MTR-waarde.

Conclusie uit Tabel 5.4 is dat natuurlijk peilbeheer voldoende is EKR-waarden hoger dan 0.60 te bereiken. In veel mindere mate spelen totaal fosfor en totaal stikstof een rol. N.B.: deze conclusie is *indicatief* zoals blijkt uit Appendix F.



## 6 Vergelijking regressiebomen en neurale netwerken

### 6.1 Methodische overeenkomsten en verschillen

Zoals uiteengezet in paragraaf 3.2.3 schatten neurale netwerken en regressiebomen relaties tussen stuurvariabelen en een EKR-waarden voor een specifieke deelmaatlat. Beide methoden geven dus een invulling voor het metamodel in het rechter stroomdiagram van Figuur 1.2. Na training op een gegeven dataset kunnen beide methoden EKR-waarden voor specifieke wateren berekenen vanuit een gegeven set van stuurvariabelen. Ook kunnen beide methoden EKR's voorspellen als bestaande waarden van stuurvariabelen worden aangepast om de invloed van maatregelen in de toekomst door te rekenen. Ten slotte blinken beide methoden uit in het omgaan met niet-lineaire relaties tussen EKR-deelmaatlatwaarden enerzijds en stuurvariabelen anderzijds, dit in tegenstelling tot een statistische techniek als multiple regressie.

Tot zover de overeenkomsten. Er zijn ook belangrijke verschillen te noemen. In de eerste plaats geven neurale netwerken EKR-voorspellingen als een *continu* getal tussen 0.0 en 1.0. Daarbij worden geen onzekerheden gegenereerd. Daarentegen geeft een regressieboom geen voorspelling als een *continu* getal, maar als een *discreet* getal, met waarden gelijk aan een van de eindknoten in de regressieboom. Dus in geval van de regressieboom voor 'vissen' in meren uit Figuur 3.4 is elke voorspelling gelijk aan één van de waarden 0.60, 0.88, 0.61, 0.54, 0.17, 0.44, 0.30 of 0.44 (eindknoten van links naar rechts in de figuur). Bij de eindknoten zijn onzekerheden te berekenen vanuit de devianties en het aantal wateren per eindknoop.

Een tweede verschil is het karakter van de geschatte modellen. Neurale netwerken geven wel voorspellingen voor EKR-waarden, maar ze geven niet aan hoe die voorspellingen tot stand komen. Met andere woorden, het is een *black box*-model. Regressiebomen daarentegen geven juist wel aan hoe tot een voorspelling is gekomen, namelijk via de splitsingen op de stuurvariabelen. Het is daarmee dus een *white box*-model.

### 6.2 Voorspelkracht vergeleken

De vier datasets zoals in dit onderzoek onderzocht met regressiebomen, zijn ook geanalyseerd met neurale netwerken (Knoben et al., 2008). Het is daarom mogelijk om de voorspellingen en de voorspelkracht van neurale netwerken te verifiëren met die bereikt via regressieboom-analyse. Daartoe zijn voor beide methoden twee informatiebronnen beschikbaar:

- de scatterplots voor voorspellingen versus werkelijke waarden, zoals getoond in Figuur 3.7, en
- een samenvatting van de voorspelkracht in termen van '*root mean squared errors*' (RMSE's, zie definitie (1a)). Hierbij geldt voor beide modellen dat er een trainingsset is gebruikt van 80% van de data en dat de voorspellingen zijn geëvalueerd op de resterende 20%. Het gaat voor beide modellen dus om echte voorspellingen, om een *validatie*.

Een visuele vergelijking van de kruisvalidatie-scatterplots per deelmaatlat laat zien dat er een grote overeenkomst bestaat tussen de scatterplots. De ligging van de voorspellingen is zeer vergelijkbaar. Omdat de RMSE-waarden volgens elke methode beschikbaar is voor elk van de

**Tabel 6.1 Samenvatting verklarende kracht (kruisvalidatie) voor regressieboomanalyse (geschat met *randomForest*) en neurale netwerken (Knoben et al. 2008). Beide methoden zijn toegepast op dezelfde datasets en de voorspelkracht is geëvalueerd op basis van de *Root Mean Squared Error* (RMSE, zie definitie (1a)).**

Soort waterlichaam	Deelmaatlat	RMSE regressieboom-analyse	RMSE neurale netwerken
Beken	macrofauna	0.081	0.088
	vissen	0.056	0.086
	macrofyten	0.054	0.074
	fytobenthos	0.068	0.086
Meren	macrofauna	0.094	0.110
	vissen	0.036	0.080
	macrofyten	0.036	0.089
	fytoplankton	0.053	0.091
Kanalen	macrofauna	0.140	0.130
	vissen	0.064	0.075
	macrofyten	0.066	0.079
	fytoplankton	0.052	0.068
Sloten	macrofauna	0.120	0.120
	vissen	0.050	0.048
	macrofyten	0.044	0.060
	---	---	---

vijftien geschatte EKR-modellen, kan de voorspelkracht van beide methoden in detail vergeleken worden. Zie Tabel 6.1 (de RMSE-waarden voor regressiebomen in Tabel 6.1, derde kolom, zijn gelijk aan die getoond in Tabel E.1, derde kolom).

Het interval ‘ $\pm$  RMSE’ kan geïnterpreteerd worden als een 68%-betrouwbaarheidsinterval voor een willekeurige voorspelling. Dus bij een RMSE-waarde van 0.81 voor ‘macrofauna’ in beken geldt dat een 68%-betrouwbaarheidsinterval gegeven wordt door  $\pm 0.08$ . Een 95%-betrouwbaarheidsinterval is dan  $\pm 0.16$ .

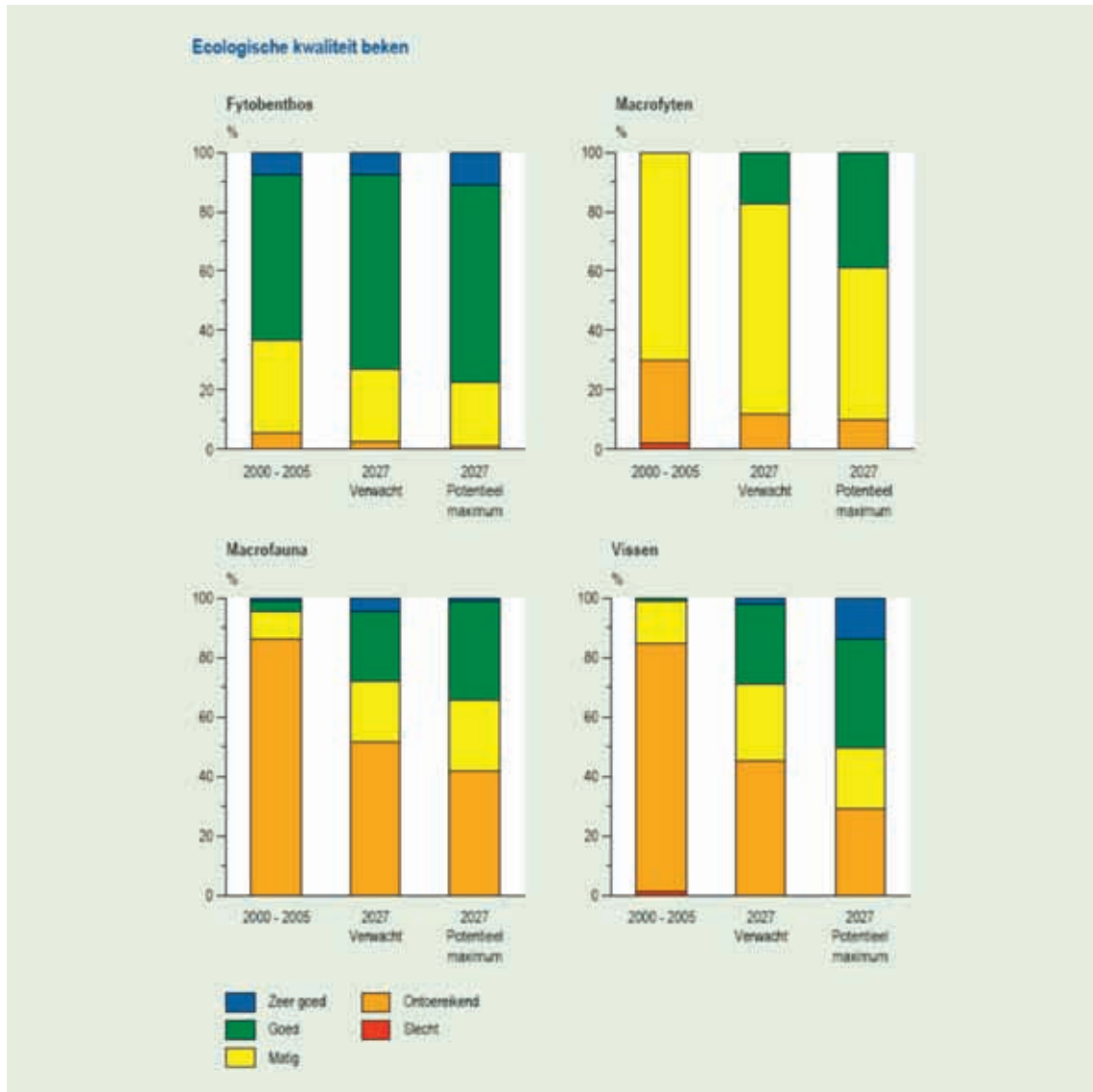
De tabel laat zien dat de RMSE-waarden voor regressiebomen over het algemeen iets lager en dus gunstiger liggen dan die voor neurale netwerken. Alleen de neurale RMSE-waarde voor ‘macrofauna’ in kanalen ligt beter (RMSE=0.13 voor neurale netwerken, 0.14 voor regressiebomen). Voor sommige deelmaatlaten bedragen de RMSE-waarden volgens regressiebomen de helft van die volgens neurale netwerken (‘vissen’ en ‘macrofyten’ in meren); dus in die gevallen is de regressieboomtechniek aanmerkelijk beter.

Conclusie is dat de voorspellingen en de voorspelkracht volgens regressieboomanalyse en neurale netwerken zeer goed vergelijkbaar zijn. Daarbij blijkt de voorspelkracht van regressiebomen zelfs nog iets gunstiger te liggen dan die via neurale netwerken.

### 6.3 Doorrekenen van maatregelen

#### Neurale netwerken

Met neurale netwerken kunnen door waterschappen voorgestelde maatregelen doorgerekend worden nadat het netwerk per deelmaatlat is getraind en gevalideerd. Een reeks van voorbeelden



**Figuur 6.1** Potentiële verandering ecologische kwaliteit in beken na uitvoering van de voorgestelde regiomaatregelen in 2027, ten opzichte van de huidige situatie (2000-2005). Bron: PBL (2008, Figuur 4.6).

is gegeven in PBL (2008, paragraaf 4.4). Figuur 6.1 is uit dit rapport overgenomen: een voorbeeld voor de ecologische kwaliteit van beken.

### Regressieboomanalyse

Aan de hand van de geschatte regressieboom is direct af te lezen met welke maatregelen een gewenste EKR-deelmaatlatwaarde bereikt kan worden. Onderstaande tekstbox geeft een uitgewerkt voorbeeld.

Analoog aan de aanpak uit de tekstbox is onderstaand voorbeeld uitgewerkt. In de praktijk is het vaak niet mogelijk om concentraties van totaal fosfor en totaal stikstof omlaag te krijgen. De vraag rijst dan wat er bereikt kan worden als er *uitsluitend hydromorfologische maatregelen* getroffen zouden worden. Is het dan mogelijk om EKR-waarden van 0.60 of beter te halen?

### Tekstbox Doorrekenen van maatregelen met een regressieboom

Hoe kan het effect van maatregelen bepaald worden in de context van regressiebomen?

Als voorbeeld wordt een meer gekozen waarvoor de EKR-waarde van 'vissen' 0.38 blijkt te bedragen, ofwel een ontoereikende ecologische toestand. De oevers horen tot de categorie 'moeras/riet', de totaal-fosfor-concentratie bedraagt 0.25 mg P/l en de totaal-stikstofconcentratie 4.0 mg N/l. Het peilbeheer is stabiel.

Het verantwoordelijke waterschap wil de EKR-waarde van het meer brengen naar de waarde 0.60, ofwel een goede ecologische toestand. Daartoe stelt zij de volgende maatregelen voor:

- natuurlijk peilbeheer;
- totaal-fosforconcentratie verlagen naar 0.15 mg P/l (MTR-waarde voor regionale wateren);
- totaal-stikstofconcentratie verlagen naar 2.2 mg N/l (MTR-waarde voor regionale wateren).

Zal het doel met deze maatregelen bereikt worden?

Om deze vraag te beantwoorden, hoeven geen modelberekeningen te worden uitgevoerd. Het antwoord kan rechtstreeks uit de geschatte regressieboom worden afgelezen. De regressieboom voor de deelmaatlat 'vissen' in meren is getoond in Figuur 3.4 (identiek aan Figuur B.6). Merk op dat de huidige waarden van stuurvariabelen consistent zijn met de meest rechtse eindknoop van de regressieboom.

De voorgestelde maatregelen vallen in het rechter deel van de boom en komen in de derde eindknoop van rechts terecht, die met een EKR-waarde van 0.44. De onzekerheid in deze waarde kan bepaald worden uit de variantie 'Dev/N' ofwel  $0.10/42 = F 0.0024$ . Dit levert een standaarddeviatie van 0.05. Als een 2- $\sigma$ -betrouwbaarheidsinterval gekozen wordt, is de verwachte EKR-waarde met onzekerheid:  $0.44 \pm 0.10$ . Daarmee zal het gestelde doel van 0.60 zeer waarschijnlijk niet gehaald worden.

Voor een overzicht en literatuurstudie naar dit type maatregelen zij verwezen naar Nijboer (2004) en Nijboer et al. (2004).

Met de hydromorfologische maatregelen kunnen verbeteringen geboekt worden zoals af te lezen is uit de 15 geschatte regressiebomen (Appendix A tot en met D). Deze verbeteringen of juist het ontbreken daarvan zijn samengevat in Tabel 6.2. De resultaten kunnen rechtstreeks afgelezen worden uit de geschatte regressiebomen.

De tabel laat zien dat er alleen verbeteringen te boeken zijn voor beken en sloten. Door meandering aan te brengen in beken zijn de EKR-waarden voor 'macrofauna' en 'vissen' te verhogen naar respectievelijk 0.62 en 0.58. Door stuwings achterwege te laten, is de EKR-waarde voor 'macrofyten' te verhogen naar 0.59. Omdat 'fyto benthos' al hoog ligt zonder maatregelen (0.58), kan gesteld worden dat de EKR-waarden van beken dicht bij de gewenste waarde van 0.60 te krijgen zijn.

Voor sloten geldt dat EKR-waarden boven de 0.60 ruim gehaald kunnen worden bij natuurlijk peilbeheer. Voor de haalbaarheid en betaalbaarheid van dit type peilbeheer zij verwezen naar Brouwer et al. (2004).

Voor kanalen geldt dat verbeteringen vooral zullen moeten plaatsvinden door het omlaagbrengen van totaal stikstof. Hydromorfologische maatregelen spelen geen rol van betekenis. Hetzelfde geldt voor meren.

## 6.4 Conclusie

Binnen de Ex-ante evaluatie KRW (PBL, 2008) zijn via neurale netwerken een reeks van voorgestelde maatregelen doorgerekend. De maatregelen zijn samengevat in hoofdstuk 3 van deze evaluatie en de neuraal-netwerkvoorspellingen voor het jaar 2027 staan in hoofdstuk 4. Parallel aan de studie van Royal Haskoning is in opdracht van de projectleiding van de Ex-ante evaluatie KRW de in dit rapport gepresenteerde benadering uitgewerkt. Doel is de robuustheid van de neurale-netwerkvoorspellingen te verifiëren.

**Tabel 6.2 Samenvatting invloed van hydromorfologische maatregelen op de verbetering van EKR-waarden.**

Soort waterlichaam	Deelmaatlat	Gemiddelde EKR-waarde zonder maatregelen	Gemiddelde EKR-waarde met hydro-morfologische maatregelen	Maatregel (indien van toepassing)
<b>Beken</b>	macrofauna	0.45	0.62	maak beek zwak slingerend tot vrij meanderend
	vissen	0.42	0.58	maak beek zwak slingerend tot vrij meanderend
	macrofyten	0.43	0.59	ongestuwd (bij totaal-fosforconcentraties lager dan 0.83 mg P/l)
	fytobenthos	0.58	idem	---
<b>Meren</b>	macrofauna	0.54	idem	---
	vissen	0.49	idem	---
	macrofyten	0.51	idem	---
	fytoplankton	0.47	idem	---
<b>Kanalen</b>	macrofauna	0.41	idem	---
	vissen	0.41	idem	---
	macrofyten	0.41	idem	---
	fytoplankton	0.45	idem	---
<b>Sloten</b>	macrofauna	0.38	0.74	voer natuurlijk peilbeheer in
	vissen	0.44	0.81	voer natuurlijk peilbeheer in
	macrofyten	0.38	0.75	voer natuurlijk peilbeheer in
	---	---	---	---

Om deze robuustheid te testen, zijn voor beide methoden modellen geschat op 80% van de beschikbare wateren binnen een watertype (de ‘trainingsset’). Daarna zijn EKR-voorspellingen gegenereerd voor de overige 20% (de ‘testset’) waarbij het model niet de EKR krijgt per specifiek water maar wel de waarden van de stuurvariabelen. Via deze benadering kunnen EKR-modelvoorspellingen dus vergeleken worden met de echte, gerealiseerde EKR-waarden, en dit onafhankelijk voor neurale netwerken en regressiebomen.

Gevonden is dat (i) de voorspellingen via beide methoden grote gelijkenis vertonen (visuele inspectie van voorspelscatterplots), en (ii) dat beide methoden een grote verklarende kracht bezitten. Als de voorspelfouten van alle validatie-runs samengenomen worden, is gevonden dat de voorspellingen via regressieboomanalyse zelfs nog iets nauwkeuriger zijn dan die via neurale netwerken.

Geconcludeerd wordt dat beide methoden zeer geschikte instrumenten zijn om maatregelen door te rekenen ter verbetering van biologische kwaliteitselementen. De verschillen en overeenkomsten, beschreven in paragraaf 6.1, bepalen daarom welke van de twee methoden de voorkeur geniet voor een specifieke toepassing.





## 7 Discussie, conclusies, aanbevelingen

De volgende conclusies kunnen uit het hier gepresenteerde onderzoek getrokken worden.

### Over de waarde van de EKR-datasets voor regionale wateren

Door Royal Haskoning zijn vier datasets samengesteld met EKR-deelmaatlatten en vier tot zes stuurvariabelen. De datasets zijn voor beken (N= 344), meren (N= 248), kanalen (N= 182) en sloten (N= 149). Deze datasets zijn een zeer belangrijke eerste stap om de relatie tussen waterkwaliteit enerzijds en stuurvariabelen anderzijds te bepalen. De datasets zijn om twee redenen uniek: (i) er is voor het eerst een koppeling gemaakt tussen EKR-waarden enerzijds en corresponderende stuurvariabelen anderzijds en (ii) per watertype zijn juist die locaties verzameld waarvoor *gelijktijdig* alle vier de deelmaatlatten bemeten zijn.

Toch moeten er enkele kanttekeningen geplaatst worden bij toepassing van deze vier datasets. In de eerste plaats is de spreiding van de wateren niet altijd mooi verdeeld over Nederland (zie Figuur 2.1). Zo ontbreken alle wateren in Friesland voor meren, kanalen en sloten. Vooral voor *sloten* is de dataset niet representatief voor Nederland als geheel: de provincies Friesland, Groningen, Drente, Gelderland, Noord-Brabant en Limburg hebben geen data.

In de tweede plaats bevatten de datasets veel ‘doublures’. Daarmee wordt bedoeld dat bijvoorbeeld de dataset meren met N=F 248 niet bestaat uit 248 *verschillende* meren, maar een kleiner aantal. Hierbij kan het zijn dat hetzelfde meer op verschillende plaatsen bemonsterd is, maar soms ook op dezelfde locatie in een ander jaar. De invloed van doublures is toegelicht in Appendix F. Gevonden is dat de doublures voor sloten de schattingsresultaten beïnvloeden. Het verdient daarom aanbeveling om vooral deze dataset uit te breiden in vervolgonderzoek.

In de derde plaats zijn stuurvariabelen soms geïnterpoleerd uit omliggende meetwaarden wanneer ze niet aanwezig waren voor een bepaald water, op het moment waarop de deelmaatlatwaarden bepaald werden. De invloed van deze interpolaties is in dit rapport niet onderzocht. Vermoedelijk is de invloed gering.

Ten slotte zijn niet alle mogelijke stuurvariabelen in de databases vertegenwoordigd. Zo ontbreken het aanwezig zijn van bestrijdingsmiddelen of bijvoorbeeld het baggeren in sloten. In meren is de vissamenstelling mede bepalend voor de ecologische kwaliteit. Een hoge brasemstand zorgt voor een slechte EKR voor het onderdeel vissen. Nijboer et al. (2004, p. 15) geven voor beken en sloten een reeks van hydrologische maatregelen die hier niet opgenomen zijn.

Ondanks deze vier beperkingen van de datasets moeten ze beschouwd worden als zeer waardevol en geven ze wel degelijk een indruk van de ecologische kwaliteit van de verschillende watertypen in relatie tot stuurvariabelen. Voor verder onderzoek, in de lijnen van dit rapport, verdient het aanbeveling om de datasets uit te breiden naar een betere dekking over Nederland.

### Over de huidige EKR-kwaliteit en de relatie tussen verschillende deelmaatlatten

De EKR-waarden van de toegepaste deelmaatlatten schommelen voor alle typen wateren rond de waarde 0.40, ofwel in de klassen ‘ontoereikend’ tot ‘matig’ (zie Figuur 1.3 voor definities). Hierbij vallen de EKR-waarden in *meren* gemiddeld het hoogst uit: rond de 0.50, ofwel een ‘matige ecologische toestand’. Er is slechts één deelmaatlat die valt in de klasse ‘goede ecologische toestand’, namelijk fyto-benthos in beken, met een gemiddelde EKR-waarde van 0.62.

**Tabel 7.1** Correlaties tussen de deelmaatlat 'vissen' enerzijds en de deelmaatlaten 'macrofauna', 'macrofyten', 'fytobenthos' en 'fytoplankton' anderzijds. De code '---' duidt aan dat de betreffende maatlat niet aanwezig is.

Correlatie-coëfficiënten	EKR-waarde 'macrofauna'	EKR-waarde 'macrofyten'	EKR-waarde 'fytobenthos'	EKR-waarde 'fytoplankton'
EKR-waarde 'vissen' in beken	0.81	0.82	0.56	---
EKR-waarde 'vissen' in meren	0.66	0.96	---	0.87
EKR-waarde 'vissen' in kanalen	0.70	0.96	---	0.78
EKR-waarde 'vissen' in sloten	0.81	0.97	---	---

De laagste EKR-waarden worden gevonden voor macrofauna en macrofyten in *sloten*: een gemiddelde EKR-waarde rond de 0.30, dus de klasse 'ontoereikende ecologische toestand'.

De ecologische kwaliteit van water moet binnen de Kaderrichtlijn Water bepaald worden aan de hand van vier deelmaatlaten. Toch is het vaak zo dat met een enkel biologisch gegeven de kwaliteit van een water al goed is in te schatten. Het is daarom interessant te kijken in hoeverre deelmaatlaten in hetzelfde type water aan elkaar gecorreleerd zijn. Bevatten ze dezelfde informatie (correlatiecoëfficiënten van 0.80 of hoger) of juist niet?

De correlatiecoëfficiënten tussen de EKR-waarden voor 'vissen' enerzijds en de overige EKR's anderzijds zijn samengevat in Tabel 7.1. De deelmaatlat 'vissen' blijkt voor alle watertypen zeer hoog te correleren aan de deelmaatlat 'macrofyten', met correlaties tot aan 0.97. De relatie tot 'macrofauna' en 'fytoplankton' is minder sterk, maar toch nog rond de 0.80. De relatie tussen 'vissen' en 'fytobenthos' is daarentegen gering:  $R=F$  0.56.

Algehele conclusie uit de correlatiematrix is dat de EKR's van de meeste soortgroepen, gezien binnen hetzelfde watertype, sterk samenhangen ( $R$ -waarden groter dan 0.80). Dit is een resultaat dat biologen niet zal verbazen, maar dat hier wel op een kwantitatief-statistische manier gevonden is. De hoge correlaties suggereren dat men in feite toe zou kunnen met minder deelmaatlaten per watertype. Dit aspect is echter in dit rapport niet verder uitgewerkt.

### Over het vergelijken van regressieboomanalyse met neurale netwerken

Zowel regressieboomanalyse als neurale netwerken blijken aantrekkelijke technieken om de invloed van stuurvariabelen op EKR-waarden te schatten. In de eerste plaats mogen relaties tussen een EKR-deelmaatlatwaarde enerzijds en een stuurvariabele anderzijds *niet-lineair* zijn. Een voordeel van regressieboomanalyse is dat de methode goed aansluit bij beleidsvragen: een regressieboom geeft door zijn structuur precies aan welke EKR-waarde bereikt kan worden als een bepaalde stuurvariabele veranderd zou worden: *als een bepaalde stuurvariabele  $x_i$  onder de drempel 'd' blijft, zal de gemiddelde EKR-waarde verbeteren (of verslechteren) naar het niveau 'y'*. Ook geeft de boom een rangorde voor het belang van toepassing van de verschillende stuurvariabelen. In de 'single regression'-benadering zoals toegepast in bijvoorbeeld Heinis en Evers (2007a) of de KRW-Verkenner, wordt een dergelijke ordening niet gevonden.

De voorspelkracht van regressiebomen is in hoofdstuk 6 in detail vergeleken met die van neurale netwerken. Beide methoden zijn toegepast op dezelfde data. Gevonden is dat (i) de voorspellingen via beide methoden grote gelijkheid vertonen (visuele inspectie van voorspel-scatterplots), en (ii) dat beide methoden een grote verklarende kracht bezitten. Als de voorspelfouten van alle

**Tabel 7.2** Maatregelen per watertype om een goede tot zeer goede ecologische toestand te bereiken. Daarmee geldt voor alle deelmaatlaten een EKR van 0.60 of hoger (de laagste EKR-deelmaatlatwaarden per watertype is leidend).

Watertype	Maatregelen waarmee gemiddeld genomen een EKR-waarde van 0.60 of hoger wordt bereikt	MTR-norm	GET- of GEP-norm
Beken	Zwak slingerend tot vrij meanderend	---	---
	Geen verstuwing	---	---
	Concentratie totaal fosfaat lager dan 0.16 mg P/l	0.15 mg P/l	0.12 tot 0.14 mg P/l
	Biologisch zuurstofverbruik lager dan 4.4 ml O <sub>2</sub> /l	---	---
Meren	Concentratie totaal stikstof lager dan 1.7 mg N/l	2.2 mg N/l	1.0 tot 1.3 mg N/l
Kanalen	Concentratie totaal stikstof lager dan 2.8 mg N/l	2.2 mg N/l	2.8 tot 3.8 mg N/l
	Concentratie totaal fosfor lager dan 0.12 mg P/l	0.15 mg P/l	0.15 tot 0.25 mg P/l
Sloten	Natuurlijk peilbeheer	---	---

validatieruns samengenomen worden, is gevonden dat de voorspellingen via regressieboom-analyse zelfs nog iets gunstiger liggen dan die via neurale netwerken.

Geconcludeerd wordt dat beide methoden zeer geschikte instrumenten zijn om maatregelen door te rekenen ter verbetering van biologische kwaliteitselementen. De verschillen en overeenkomsten, beschreven in paragraaf 6.1, bepalen daarom welke van de twee methoden de voorkeur geniet voor een specifieke toepassing.

#### Over de invloed van stuurvariabelen

Overall-conclusie uit hoofdstuk 5 is dat een goede tot zeer goede ecologische toestand (alle EKR-waarden 0.60 of hoger) vooral bepaald wordt door de stuurvariabelen 'totaal fosfor' en 'totaal stikstof'. Alle regionale wateren samenvattend is gevonden dat de concentratie totaal fosfor lager moet zijn dan 0.12 mg P/l. Idem voor totaal stikstof een concentratie lager dan 1.7 mg N/l. Zie Tabel 7.2 voor een overzicht.

De hier gevonden grenswaarden liggen steeds iets onder de MTR-waarden voor totaal fosfor en totaal stikstof (Vierde Nota Waterhuishouding). De MTR-waarde voor totaal fosfor bedraagt 0.15 mg P/l en voor totaal stikstof 2.2 mg N/l.

Voor **beken** geldt daarnaast dat een goede *meandering* cruciaal is (klasse 3 en hoger, ofwel zwak tot vrij meanderend), met daarbij een *BZV* lager dan 4.3 mg O<sub>2</sub>/l, en een *verstuwing* in de klasse 2 of 3 (= matig gestuwd tot ongestuwd). Voor **sloten** is *peilbeheer* de cruciale factor.

Peilbeheer in klasse 3 (= natuurlijk peilbeheer) is noodzakelijk voor het halen van een goed ecologisch potentieel (GEP). Opgemerkt zij dat de resultaten uit Appendix F laten zien dat het aantal *verschillende* sloten gering is (als doublures gemiddeld worden in de database blijven er van de 149 data 35 over). Deze conclusie is daarom indicatief.

Voor meren en kanalen spelen hydromorfologische stuurvariabelen geen rol van betekenis.

#### Verdere ontwikkeling

De resultaten uit deze studie laten zien dat regressieboomanalyse een krachtig instrument is om relaties tussen biologische kwaliteit en stuurvariabelen te bepalen. Hoe zou dit onderzoek voortgezet kunnen worden?

Wat betreft de methode om regressiebomen te schatten, kan gesteld worden dat de techniek, zoals hier toegepast, wellicht nog iets verfijnd kan worden. Een aantal instelmogelijkheden van de techniek zijn niet uitgetest, zoals het bepalen van de invloed van individuele stuurvariabelen in *randomForest*.

De vier databases uit deze studie zouden uitgebreid en wellicht verbeterd kunnen worden. In de huidige databases hebben veel regionale wateren doublures (metingen op dezelfde locatie, maar genomen in verschillende jaren), met stuurvariabelen die hier en daar geïnterpoleerd zijn. Gestreefd zou moeten worden naar meer verschillende wateren per watertype. Voor sloten is dit het belangrijkste. Verder zou toepassing van interpolaties zoveel mogelijk vermeden moeten worden (ook al blijkt uit Appendix F dat het effect voor beken, meren en kanalen gering is). En voor alle watertypen is het van belang te streven naar een goede dekking over Nederland.

Daarnaast verdient het aanbeveling de databases aan te vullen met meer stuurfactoren dan nu aanwezig zijn in de databases. Zie de discussie in het tweede deel van paragraaf 2.2. Welke aanvullende stuurfactoren het meest in aanmerking komen, is hier niet onderzocht.

Het is een interessante vraag of de hier voorgestelde methodiek ook toepasbaar is op Rijkswateren. Het antwoord is vermoedelijk ontkennend. Het aantal Rijkswateren is hooguit zestien en dat aantal lijkt te weinig om regressieboomanalyse toe te passen. Toch kan dit punt nader onderzocht worden. Resultaten zouden dan vergeleken kunnen worden met die bereikt met de *KRW-Verkenner*.

Hoge correlaties tussen deelmaatlaten suggeren dat men wellicht toe zou kunnen met minder *EKR*'s. Dit punt zou nader onderzocht kunnen worden.

Ten slotte kan het interessant zijn om de hier gevonden kennisregels te vergelijken en wellicht te integreren met de kennisregels zoals die nu toegepast worden in de *KRW-Verkenner*. Een vergelijking tussen kennisregels heeft nog niet plaatsgevonden.

## Dankwoord

De auteurs bedanken Royal Haskoning voor het beschikbaar stellen van de databases voor beken, kanalen en sloten, met daarbij de informatie uit Tabel 2.1 en 2.2.

De auteurs bedanken verder Reinier van den Berg (LDL/PBL), Guus Beugelink (RIM/PBL), Ron Franken (RIM/PBL) en Anton van der Giessen (IMP/PBL) voor het kritisch lezen van eerdere versies van dit rapport. De zestien figuren met regressiebomen zijn vormgegeven door Filip de Blois (IMP/PBL). Arnold Dekkers (EMI/RIVM) heeft voor deze studie de software van *randomForest* beschikbaar gemaakt als module binnen S-PLUS 8.0.



## Literatuur

- Breiman, L., 2001. Random Forests. *Machine Learning* 45, pp. 5-32.
- Brouwer, R., Claessen, T.H.L., Coops, H. en Van der Veeren, R., 2004. De economische waarde van natuurlijk peilbeheer in de KRW. *H<sub>2</sub>O* 37 (25/26), pp. 25-27.
- De Lange, M., Arts, G. en Clevering, O., 2008. Versterkt belasting watersysteem met bestrijdingsmiddelen de gevolgen van eutrofiëring? *H<sub>2</sub>O* 14/15, pp. 47-48.
- Engelbertink, R., Driesprong, A., Van de Ven, K. en Van den Roovaart, J., 2008. Worden in 2027 de doelen voor stoffen gehaald in rijkswateren? *H<sub>2</sub>O* 14/15, pp. 43-46.
- Evers, C.H.M., Van den Broek, A.J.M., Buskens en Van Leerdam, A., 2007. Omschrijving MEP en conceptmaatlaten voor sloten en kanalen voor de Kaderrichtlijn Water. Rapport Royal Haskoning 9S3656.
- Heinis, F. en Evers, C.H.M., 2007a. Afleiding getalswaarden voor nutriënten voor de goede ecologische toestand voor natuurlijke wateren. STOWA-rapport nr. 2007-02, Utrecht.
- Heinis, F. en Evers, C.H.M., 2007b. Toelichting op ecologische doelen voor nutriënten in oppervlaktewater. STOWA-rapport nr. 2007-18, Utrecht.
- Jaarsma, N., Klinge, M. en Lamers, L. (2008) Van helder naar troebel.... en weer terug. Een ecologische systeemanalyse en diagnose van ondiepe meren en plassen voor de kaderrichtlijn water. STOWA-rapport nr. 2008-04, Utrecht.
- Janssen, P.H.M., Heuberger, P.S.C. en Tiktak, A., 2005. Meta-modellieren bij het MNP-RIVM. Naar praktische toepassingen. MNP-rapport 500013001, Bilthoven.
- Knoben, R.A.E., Evers en C.H.M., Jansen, 2008. Ontwikkeling en toepassing ecologisch expertsysteem voor regionale wateren. Achtergrondrapport Ex-ante evaluatie KRW. Rapport Royal Haskoning, referentie: 9S9605/R00002/901530/AH/DenB.
- KRW, 2000. Europees Parlement en Raad van de Europese Unie, 2000: richtlijn 2000/60/EG 'Kaderrichtlijn Water'. Publicatieblad van de Europese Gemeenschappen.
- PBL, 2008 (red: Ligtoet, W., Beugelink, G, en Brink, C.). Kwaliteit voor later. Ex-ante evaluatie Kaderrichtlijn Water. PBL-publicatienummer 500140001. Te downloaden van: <http://www.mnp.nl/bibliotheek/rapporten/500140001.pdf>
- Millard, S.P. en Neerchal, N.K., 2001. *Environmental statistics with S-PLUS*. CRC Press, Washington D.C.
- MNP, 2006. Tussenevaluatie van de nota Duurzame gewasbescherming. MNP-rapport nr. 500126001, Milieu- en Natuurplanbureau, Bilthoven.
- Nijboer, R.C., 2004. Een expertsysteem voor de keuze van hydrologische maatregelen. I: Literatuurstudie naar hydrologische maatregelen en de effecten op sloot- en beekecosystemen. Alterra-rapport 1066, Wageningen.
- Nijboer, R.C., Van Diepen, L.T.A. en Higler, L.W.G., 2004. Een expertsysteem voor de keuze van hydrologische maatregelen. III: Inventarisatie van hydrologische herstelprojecten. Alterra-rapport 1067, Wageningen.
- Van der Molen, D.T. en Pot, R. (red.), 2007. Referenties en maatlaten voor natuurlijke watertypen voor de Kaderrichtlijn Water. STOWA 2007-32.
- Venables, W.N. en Ripley, B.D., 2002. *Modern applied statistics with S*. Springer Verlag.
- Visser, H. en Noordijk, H., 2002. Correcting air pollution time series for meteorological variability. With an application to regional PM<sub>10</sub> concentrations. RIVM report 722601007. Te downloaden van: <http://www.rivm.nl/bibliotheek/rapporten/722601007.html>
- Witteveen en Bos, 2008. Kosten en baten van actief visstandbeheer. Rapportage in het kader van de Ex-ante evaluatie van de Europese Kaderrichtlijn Water (EA-KRW), Deventer.



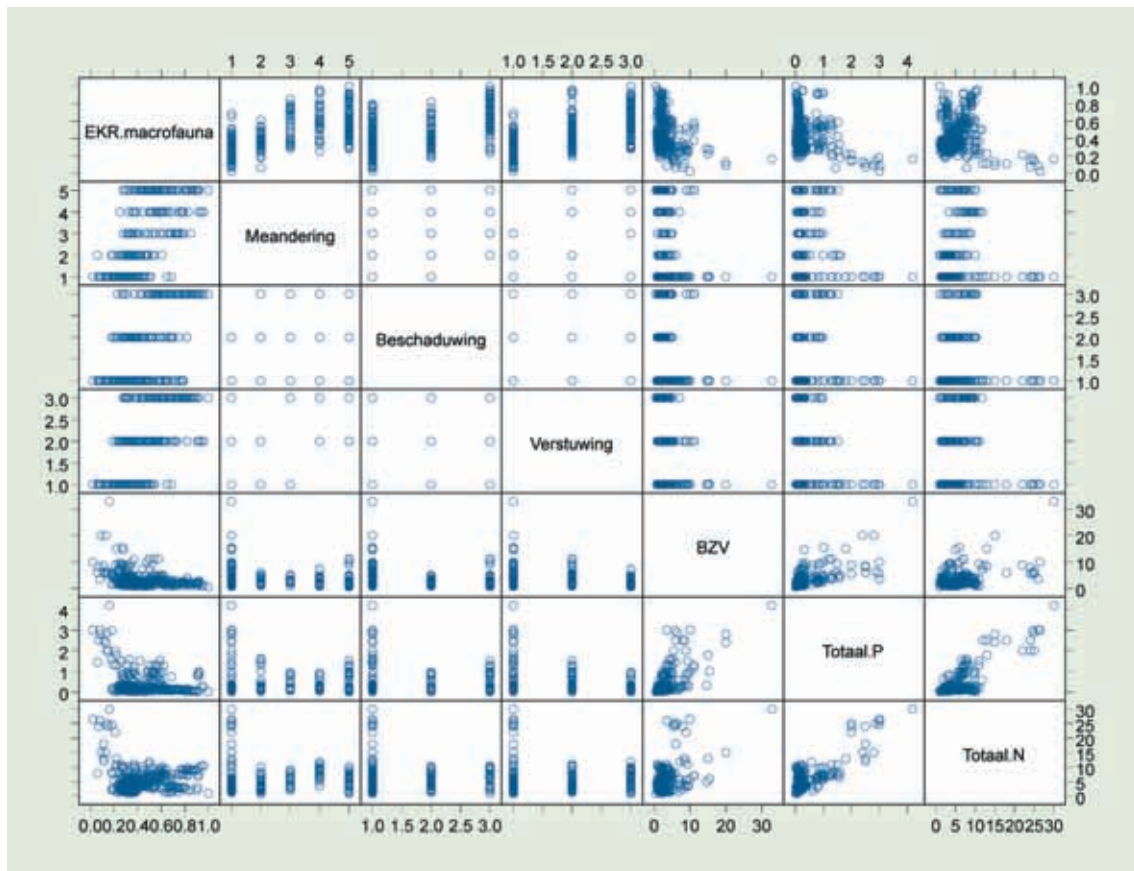


## Appendix A Regressiebomen voor beken

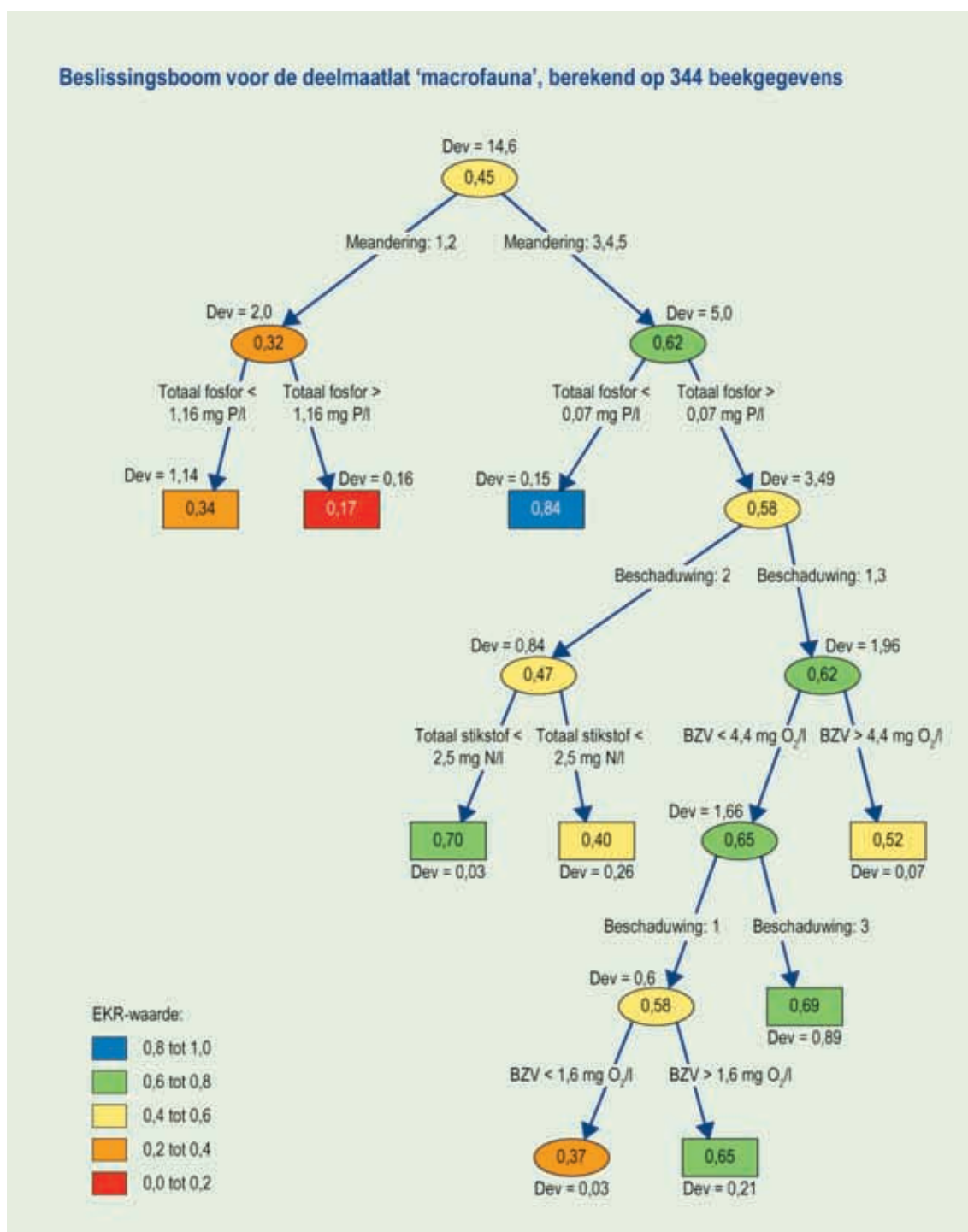
In deze appendix worden de regressieboomschattingen gegeven voor de deelmaatlaten ‘macrofauna’ (paragraaf A.1), ‘vissen’ (A.2), ‘macrofyten’ (A.3) en ‘fytobenthos’ (A.4) in beken. Fysisch- chemische stuurvariabelen zijn voor al deze deelmaatlaten: BZV, totaal fosfor en totaal stikstof. Hydromorfologische variabelen zijn: meandering, beschaduwing en verstuwing. Zie voor een uitleg van deze zes variabelen Tabel 2.2, het eerste paneel.

### A.1 Deelmaatlat macrofauna

Figuur A.1 geeft de relatie tussen de deelmaatlat ‘macrofauna’ enerzijds en de beschikbare stuurvariabelen anderzijds in de vorm van een scatterplotmatrix. De figuur laat in de bovenste rij zien dat EKR-waarden stijgen bij toenemende waarden van meandering, beschaduwing en verstuwing, en dalen bij toenemende waarden van BZV, Totaal P en Totaal N. Figuur A.2 en Tabel A.1 geven de geschatte regressieboom. Meandering blijkt de belangrijkste stuurvariabele te zijn. Figuur A.3 geeft het resultaat voor een validatiecheck op 500 regressiebomen. Figuur A.4 tenslotte geeft histogrammen met dichtheden voor de verschillende waarden van de belangrijkste stuurvariabele, namelijk meandering.



Figuur A.1 Scatterplotmatrix voor ‘macrofauna’ (afhankelijke variabele) en zes stuurvariabelen. Zie voor een methodische uitleg van de figuur paragraaf 3.1.1.



**Figuur A.2** Regressieboom met 9 eindknoten (rechthoeken). De gemiddelde EKR-waarde van een knoop staat in de rechthoek of ellips. De bijbehorende deviantie staat steeds onder de knoop. Het belang van de splitsingen is het grootst voor de eerste splitsing en neemt af naarmate 'gezakt' wordt in de boom. Knoopwaarden tussen 0,6 en 0,8 vallen in de klasse 'goede ecologische toestand' (GET), waarden tussen 0,8 en 1,0 in de klasse 'zeer goede ecologische kwaliteit' (ZGET). Zie voor een methodische uitleg paragraaf 3.2. De kleurcode correspondeert met Figuur 1.3. Door regressieboom verklaarde variantie bedraagt 80%.

**Tabel A.1** Kwantitatieve informatie bij geschatte regressieboom.

Number of terminal nodes: 9

Residual mean deviance: 0.008704 = 2.916 / 335

Distribution of residuals:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-3.222e-001	-6.390e-002	-1.217e-002	7.810e-017	4.940e-002	3.461e-001

node), split, n, deviance, yval

\* denotes terminal node

1) root 344 14.61000 0.4515

2) Meandering:1,2 191 2.02500 0.3187

4) Totaal.P&lt;1.1575 164 1.13500 0.3439 \*

5) Totaal.P&gt;1.1575 27 0.15590 0.1659 \*

3) Meandering:3,4,5 153 5.01800 0.6172

6) Totaal.P&lt;0.0730357 23 0.15420 0.8422 \*

7) Totaal.P&gt;0.0730357 130 3.49400 0.5774

14) Beschaduwng:2 38 0.84010 0.4637

28) Totaal.N&lt;2.52083 8 0.02549 0.6987 \*

29) Totaal.N&gt;2.52083 30 0.25470 0.4010 \*

15) Beschaduwng:1,3 92 1.95900 0.6243

30) BZV&lt;4.38456 75 1.66300 0.6481

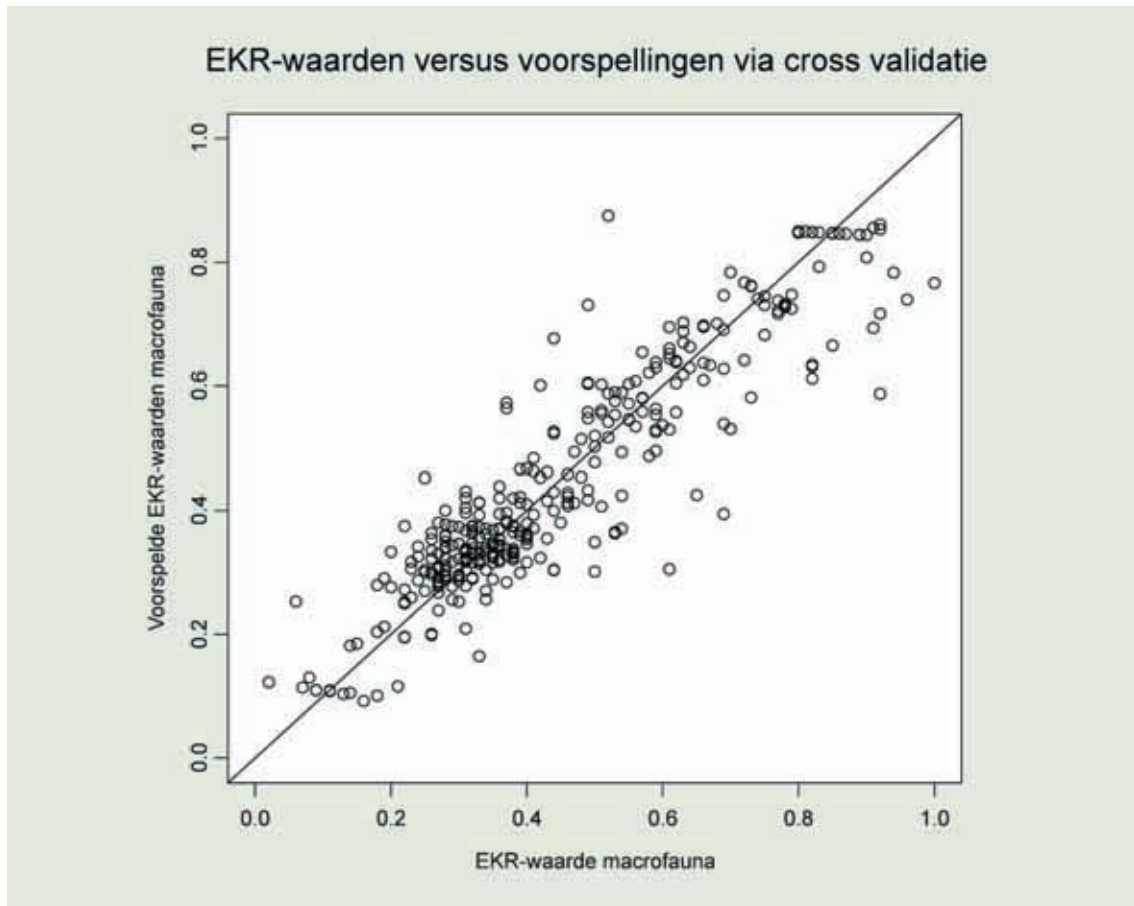
60) Beschaduwng:1 27 0.60020 0.5833

120) BZV&lt;1.58333 6 0.03333 0.3667 \*

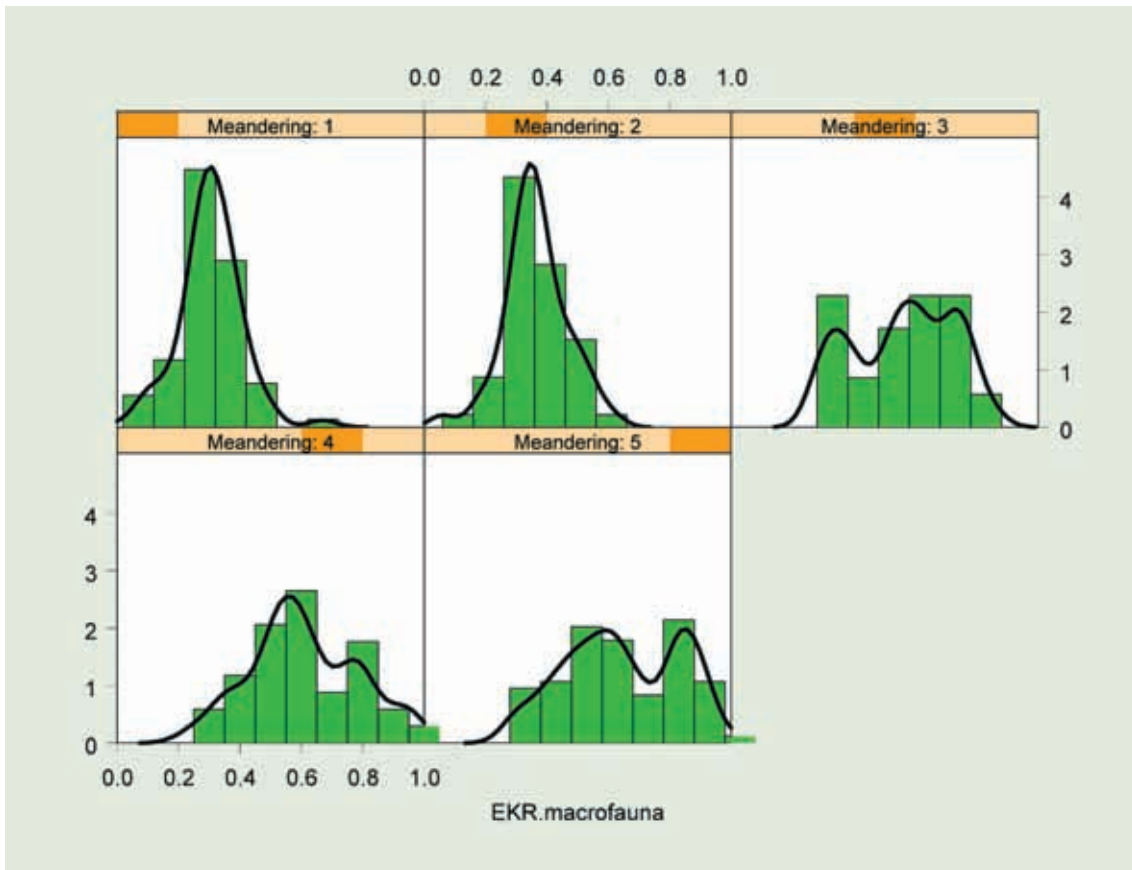
121) BZV&gt;1.58333 21 0.20470 0.6452 \*

61) Beschaduwng:3 48 0.88540 0.6846 \*

31) BZV&gt;4.38456 17 0.06689 0.5194 \*



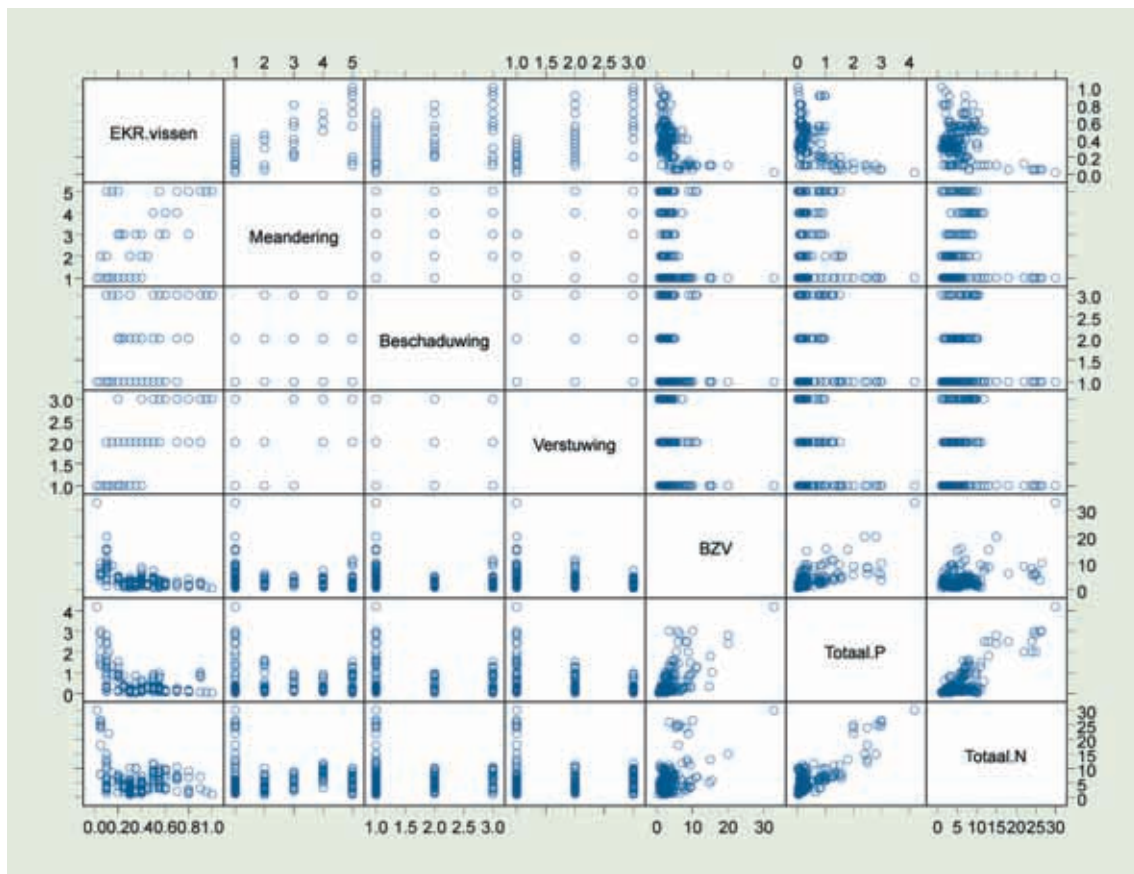
**Figuur A.3** EKR-voorspellingen met *randomForest* (500 bomen en een trainingsset van 80% van het totaal aantal beken). Kengetallen: MAD=0.058, RMSE=0.081 en  $R_{adj}^2=0.85$ . Zie voor een methodische uitleg paragraaf 3.2.2



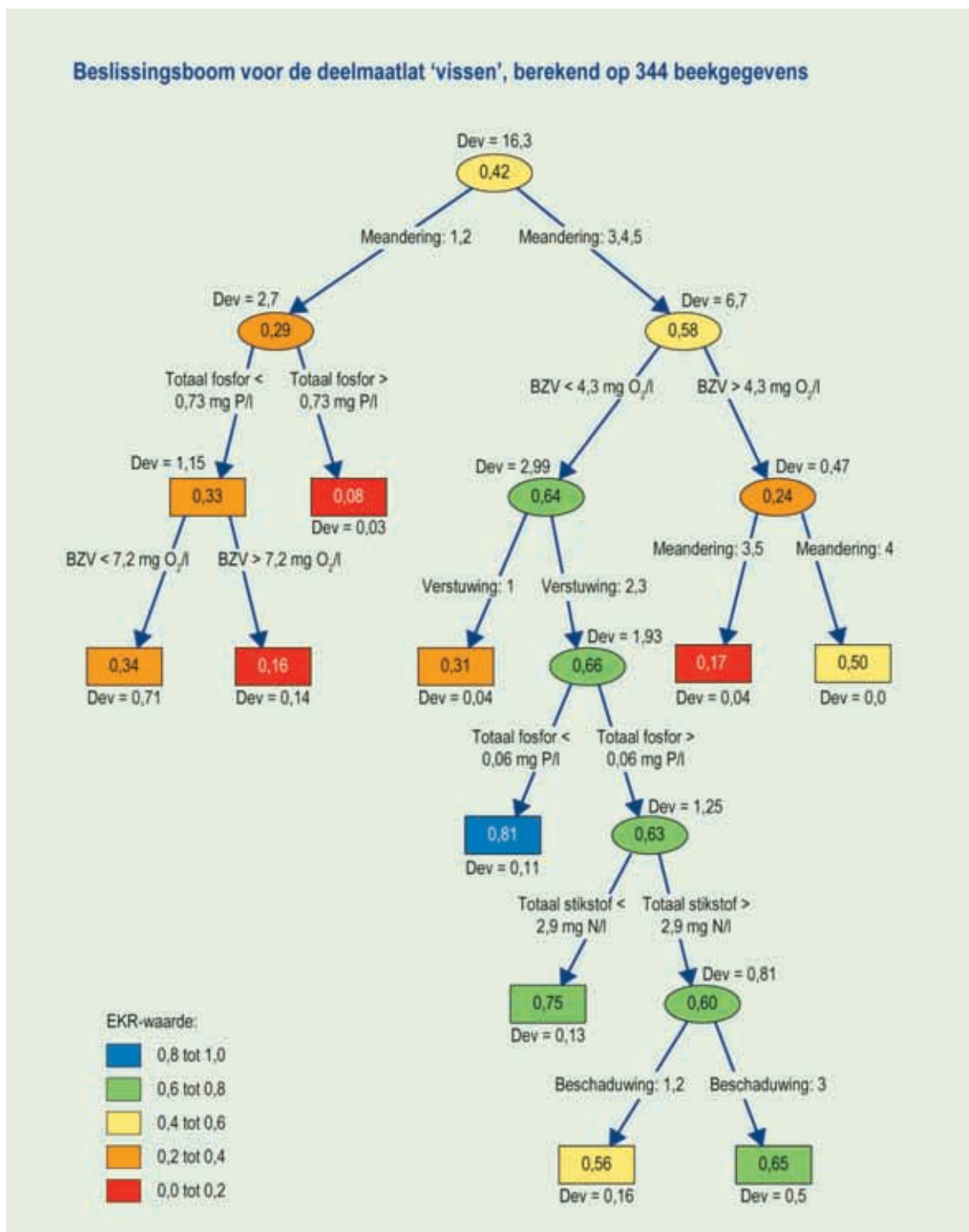
**Figuur A.4** Histogrammen en dichtheden voor de belangrijkste stuurvariabele 'Meandering' (5 discrete klassen). Zie voor een methodische uitleg paragraaf 3.1.2.

## A.2 Deelmaatlat vissen

Figuur A.5 geeft de relatie tussen de deelmaatlat ‘vissen’ enerzijds en de beschikbare stuurvariabelen anderzijds in de vorm van een scatterplotmatrix. De figuur laat in de bovenste rij zien dat EKR-waarden stijgen bij toenemende waarden van meandering, beschaduwing en verstuwing, en dalen bij toenemende waarden van BZV, Totaal P en Totaal N. Figuur A.6 en Tabel A.2 geven de geschatte regressieboom. Meandering blijkt de belangrijkste stuurvariabele te zijn. Figuur A.7 geeft het resultaat voor een validatiecheck op 500 regressiebomen. Figuur A.8 geeft histogrammen met dichtheden voor de verschillende waarden van de twee belangrijkste stuurvariabelen, namelijk meandering en BZV.



Figuur A.5 Scatterplotmatrix voor ‘vissen’ (afhankelijke variabele) en zes stuurvariabelen. Zie voor een methodische uitleg van de figuur paragraaf 3.1.1.



**Figuur A.6** Regressieboom met 10 eindknoten (rechthoeken). De gemiddelde EKR-waarde van een knoop staat in de rechthoek of ellips. De bijbehorende deviantie staat steeds onder de knoop. Het belang van de splitsingen is het grootst voor de eerste splitsing en neemt af naarmate 'gezakt' wordt in de boom. Knoopwaarden tussen 0,6 en 0,8 vallen in de klasse 'goede ecologische toestand' (GET), waarden tussen 0,8 en 1,0 in de klasse 'zeer goede ecologische kwaliteit' (ZGET). Zie voor een methodische uitleg paragraaf 3.2. De kleurcode correspondeert met Figuur 1.3. Door regressieboom verklaarde variantie bedraagt 89%.

**Tabel A.2 Kwantitatieve informatie bij geschatte regressieboom.**

Number of terminal nodes: 10

Residual mean deviance: 0.005549 = 1.853 / 334

Distribution of residuals:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-2.374e-001	-3.742e-002	-1.190e-002	-3.744e-017	5.000e-002	2.500e-001

node), split, n, deviance, yval

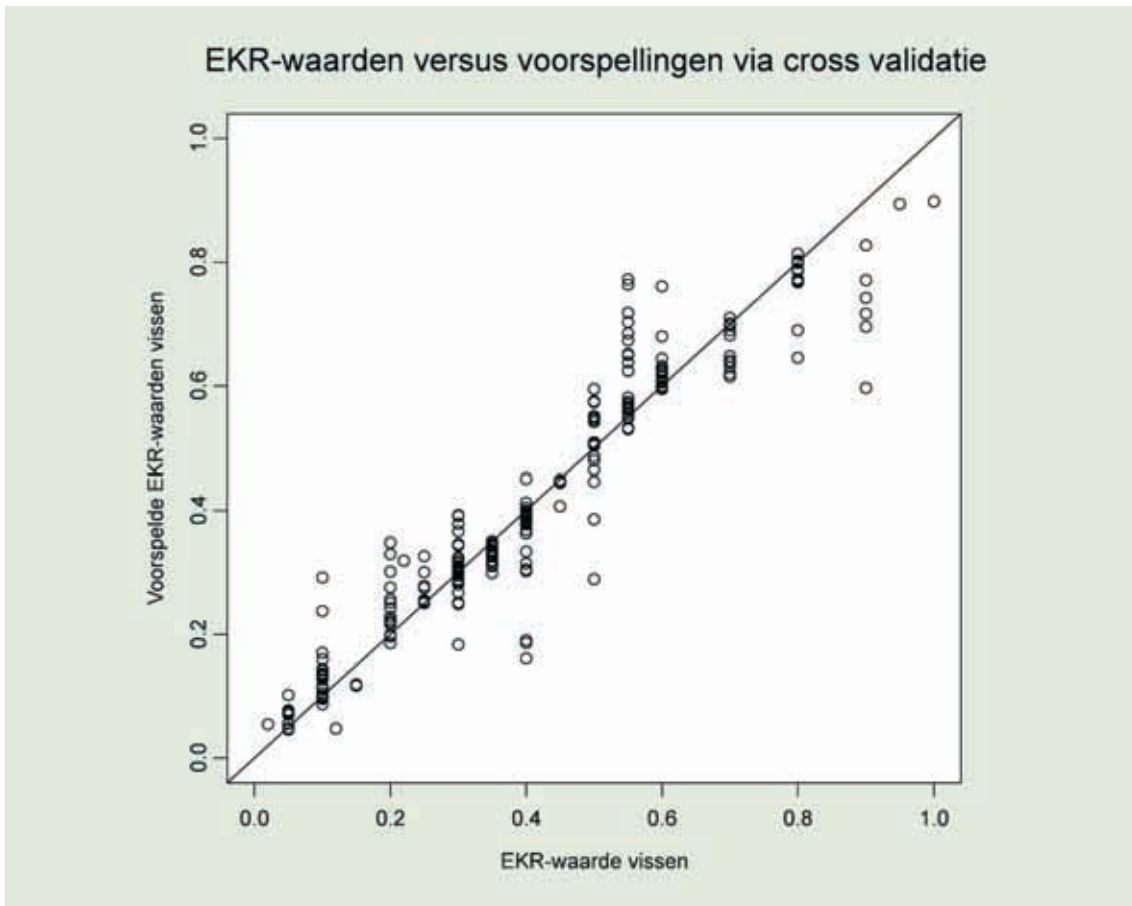
\* denotes terminal node

```

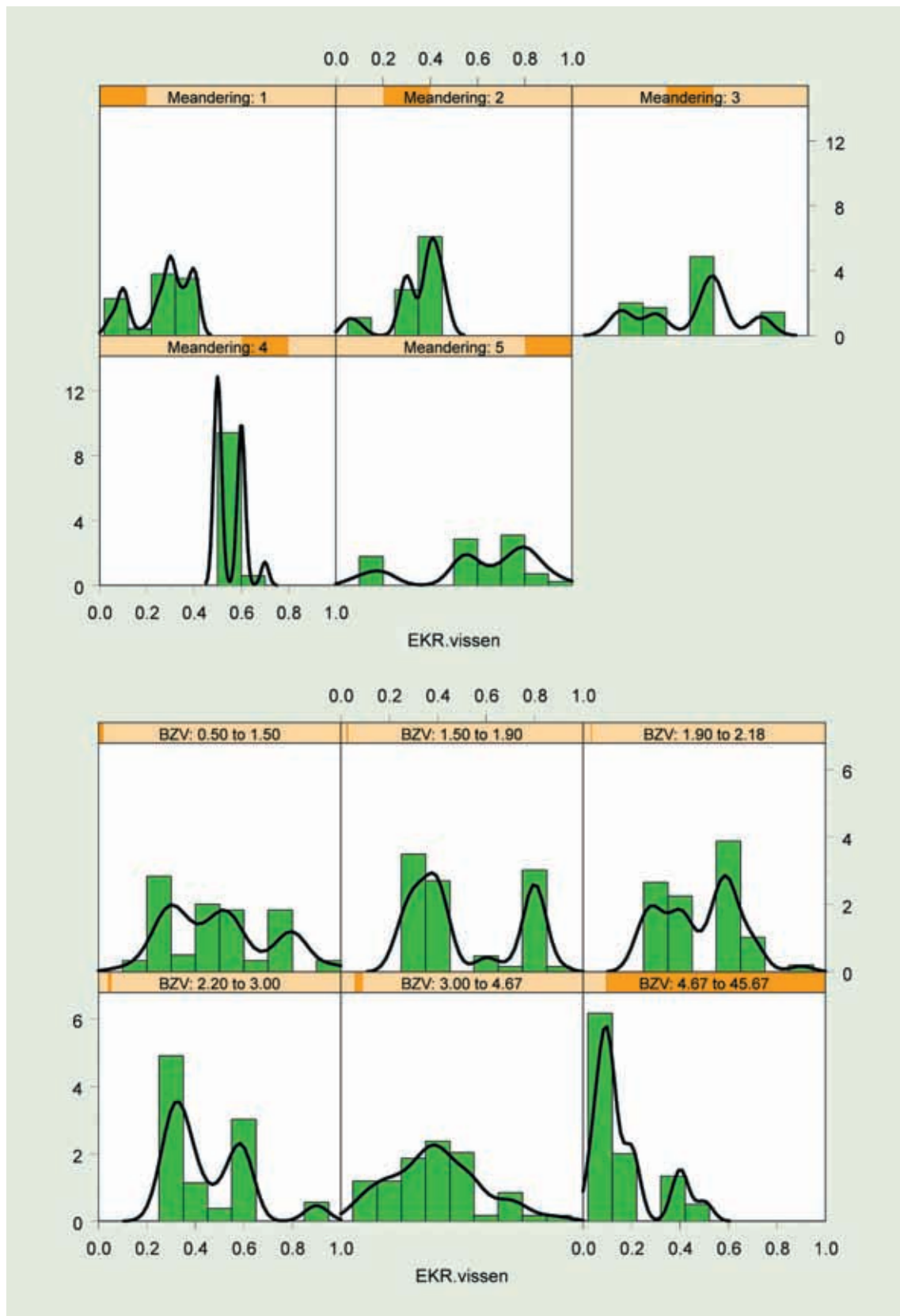
1) root 344 16.32000 0.4157
2) Meandering:1,2 191 2.67900 0.2882
4) Totaal.P<0.73 161 1.15000 0.3264
8) BZV<7.23333 151 0.71110 0.3374 *
9) BZV>7.23333 10 0.14400 0.1600 *
5) Totaal.P>0.73 30 0.03063 0.0830 *
3) Meandering:3,4,5 153 6.65600 0.5750
6) BZV<4.25 129 2.98900 0.6374
12) Verstuwing:1 9 0.03980 0.3133 *
13) Verstuwing:2,3 120 1.93400 0.6617
26) Totaal.P<0.063869 21 0.10950 0.8119 *
27) Totaal.P>0.063869 99 1.25000 0.6298
54) Totaal.N<2.91917 18 0.12500 0.7500 *
55) Totaal.N>2.91917 81 0.80670 0.6031
110) Beschaduwung:1,2 44 0.16180 0.5636 *
111) Beschaduwung:3 37 0.49500 0.6500 *
7) BZV>4.25 24 0.46490 0.2396
14) Meandering:3,5 19 0.03658 0.1711 *
15) Meandering:4 5 0.00000 0.5000 *

```





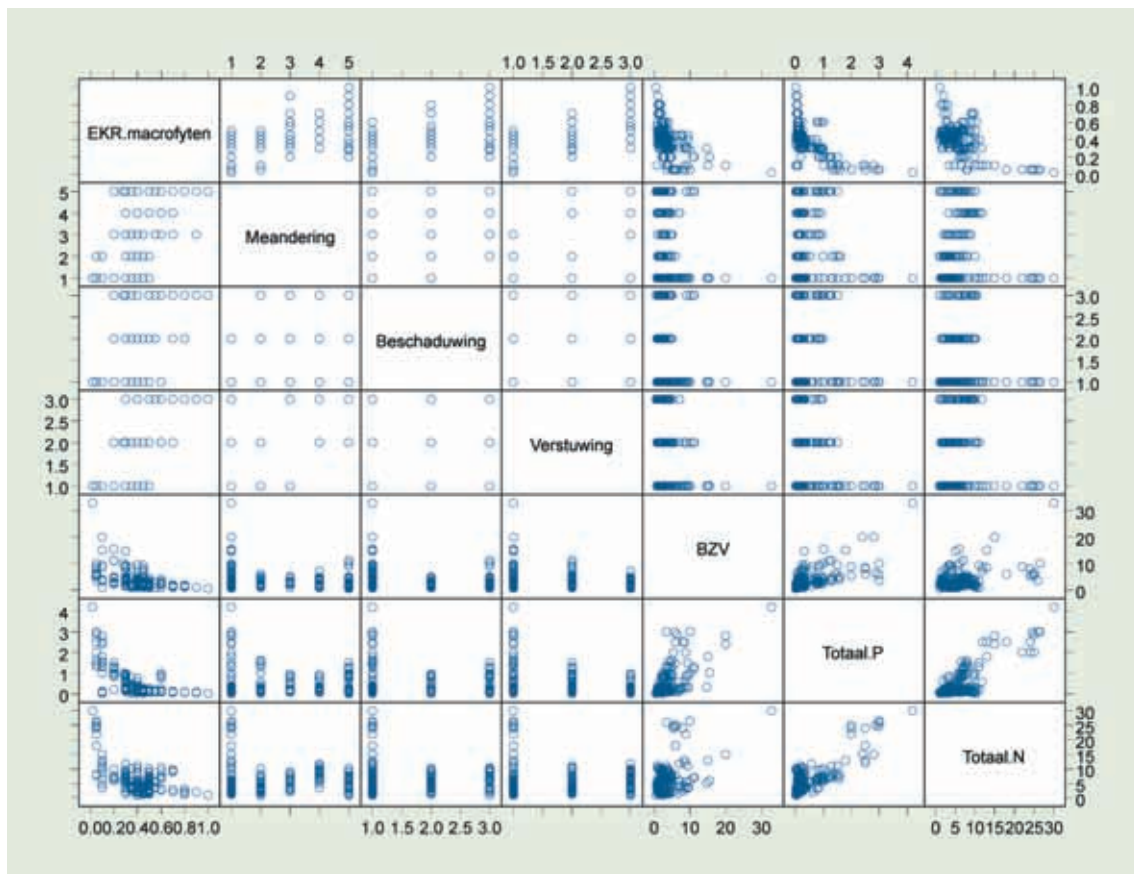
**Figuur A.7** EKR-voorspellingen met *randomForest* (500 bomen en een trainingsset van 80% van het totaal aantal beken). Kengetallen: MAD=0.030, RMSE=0.056 en  $R_{adj}^2=0.93$ . Zie voor een methodische uitleg paragraaf 3.3.2.



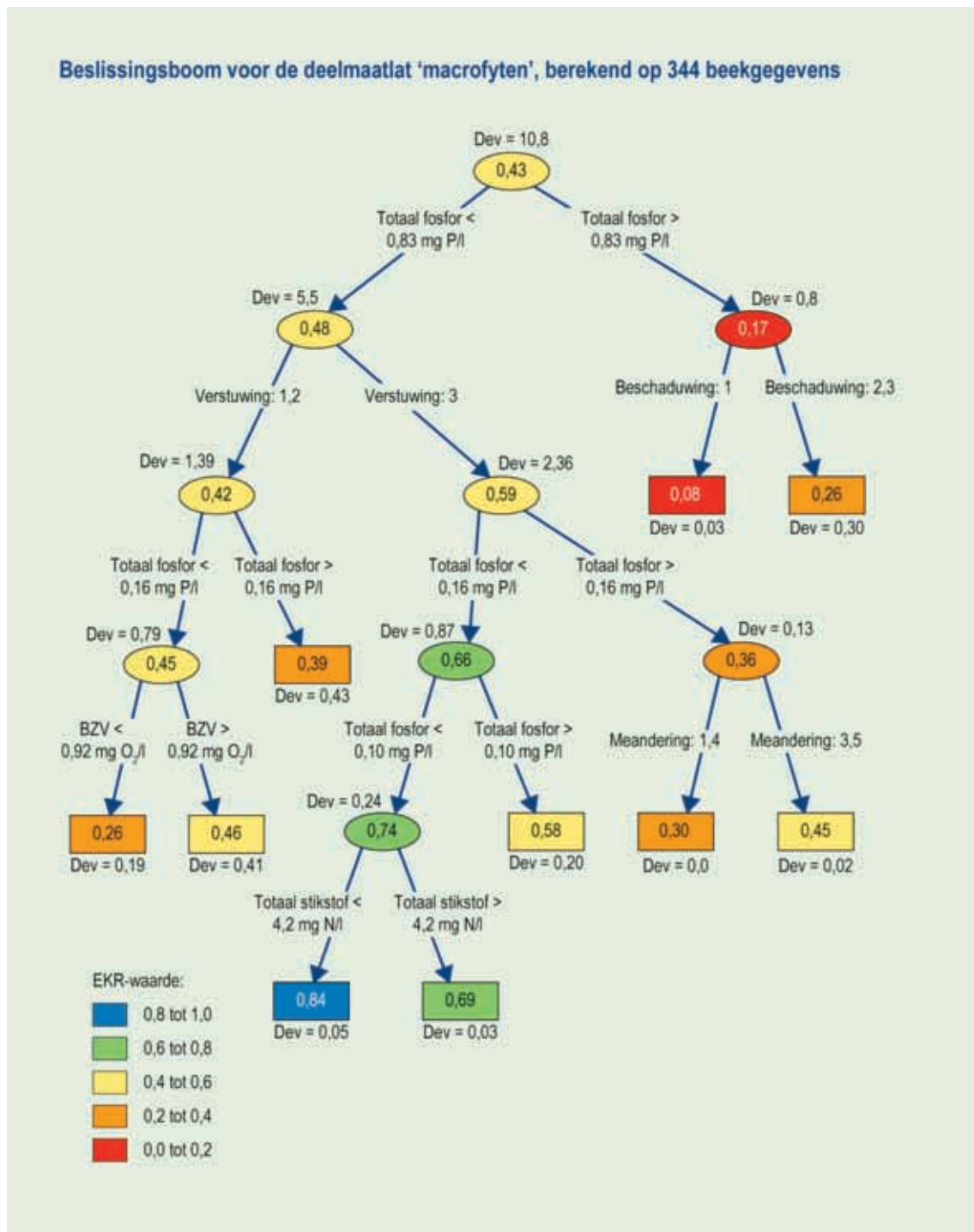
Figuur A.8 Histogrammen en dichtheden voor de twee belangrijkste stuurvariabelen: ‘meandering’ (boven, 5 discrete klassen) en ‘BZV’ (onder). Zie voor een methodische uitleg paragraaf 3.1.2.

### A.3 Deelmaatlat macrofyten

Figuur A.9 geeft de relatie tussen de deelmaatlat ‘macrofyten’ enerzijds en de beschikbare stuurvariabelen anderzijds in de vorm van een scatterplotmatrix. De figuur laat zien in de bovenste rij dat EKR-waarden stijgen bij toenemende waarden van meandering, beschaduwing en verstuwung, en dalen bij toenemende waarden van BZV, Totaal P en Totaal N. Figuur A.10 en Tabel A.3 geven de geschatte regressieboom. ‘Totaal P’ blijkt de belangrijkste stuurvariabele te zijn. Figuur A.11 geeft het resultaat voor een validatiecheck op 500 regressiebomen. Tenslotte geeft Figuur A.12 histogrammen met dichtheden voor de verschillende waarden van de belangrijkste stuurvariabele, namelijk ‘Totaal P’.



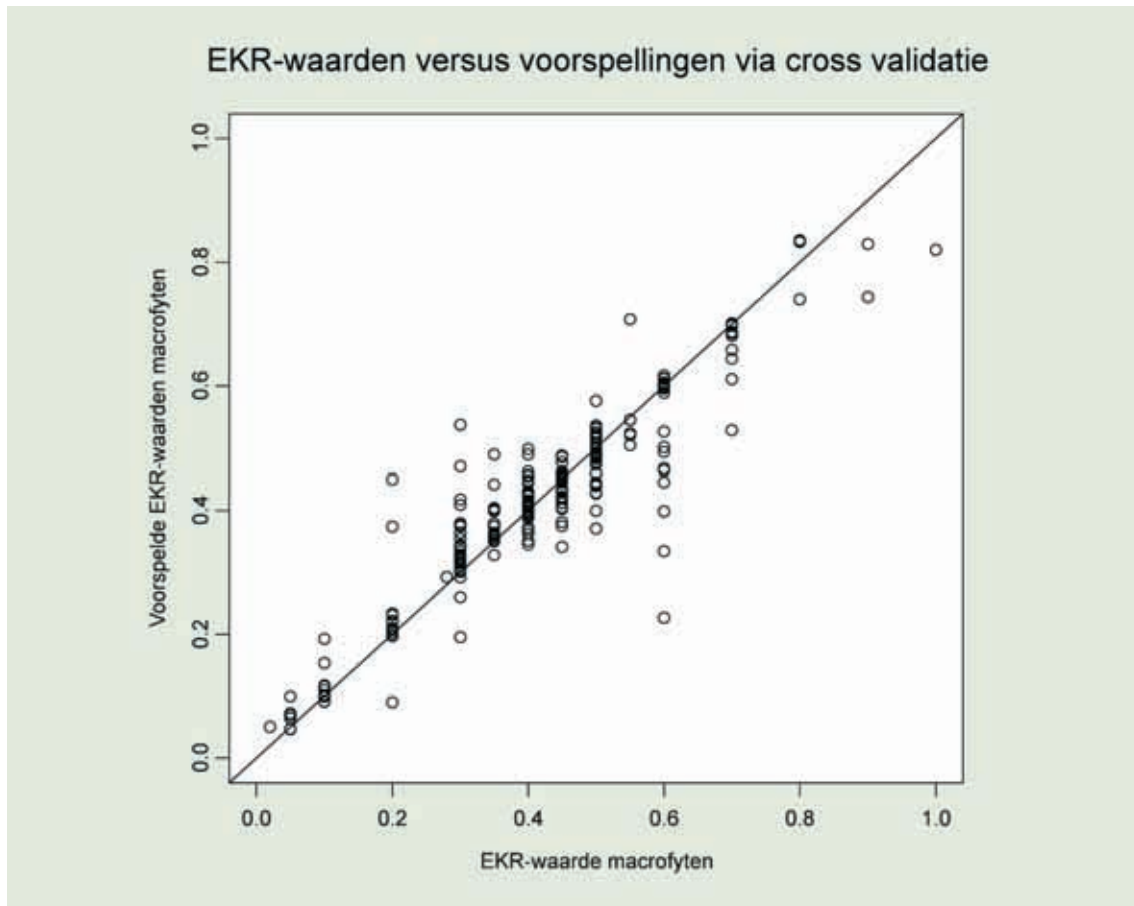
Figuur A.9 Scatterplotmatrix voor ‘macrofyten’ (afhankelijke variabele) en zes stuurvariabelen. Zie voor een methodische uitleg van de figuur paragraaf 3.1.1.



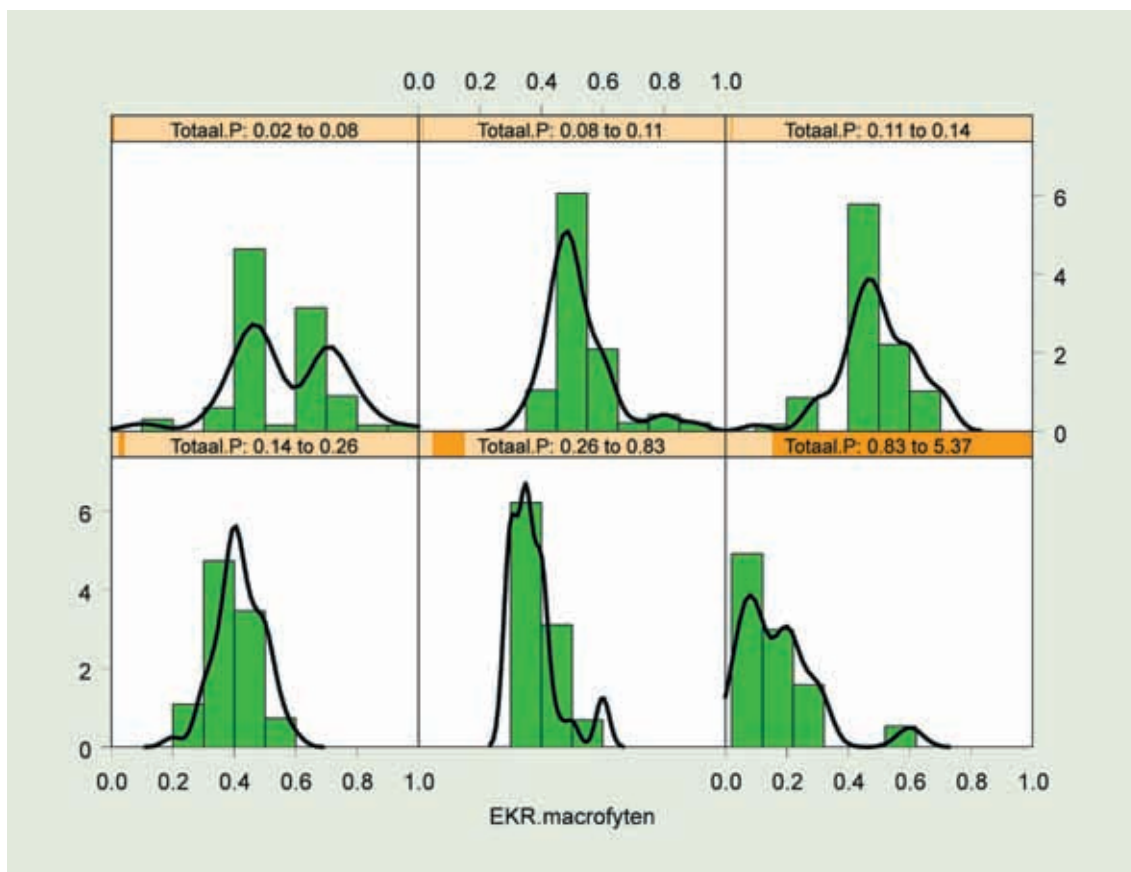
**Figuur A.10** Regressieboom met 10 eindknoten (de rechthoeken). De gemiddelde EKR-waarde van een knoop staat in de rechthoek of ellips. De bijbehorende deviantie staat steeds onder de knoop. Het belang van de splitsingen is het grootst voor de eerste splitsing en neemt af naarmate 'gezak' wordt in de boom. Knoopwaarden tussen 0.6 en 0.8 vallen in de klasse 'goede ecologische toestand' (GET), waarden tussen 0.8 en 1.0 in de klasse 'zeer goede ecologische kwaliteit' (ZGET). Zie voor een methodische uitleg paragraaf 3.2. De kleurcode correspondeert met Figuur 1.3. Door regressieboom verklaarde variantie bedraagt 85%.

**Tabel A.3** Kwantitatieve informatie bij geschatte regressieboom.

Number of terminal nodes: 10					
Residual mean deviance: 0.004993 = 1.668 / 334					
Distribution of residuals:					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-2.816e-001	-3.782e-002	1.110e-016	-7.068e-017	2.000e-002	3.378e-001
node), split, n, deviance, yval					
* denotes terminal node					
1) root 344 1.078e+001 0.4269					
2) Totaal.P<0.831714 288 5.489e+000 0.4773					
4) Verstuwing:1,2 196 1.387e+000 0.4240					
8) Totaal.P<0.15875 118 7.870e-001 0.4479					
16) BZV<0.916667 5 1.920e-001 0.2600 *					
17) BZV>0.916667 113 4.107e-001 0.4562 *					
9) Totaal.P>0.15875 78 4.309e-001 0.3878 *					
5) Verstuwing:3 92 2.360e+000 0.5908					
10) Totaal.P<0.161667 72 8.708e-001 0.6549					
20) Totaal.P<0.095 34 2.365e-001 0.7368					
40) Totaal.N<4.22083 11 4.545e-002 0.8364 *					
41) Totaal.N>4.22083 23 2.978e-002 0.6891 *					
21) Totaal.P>0.095 38 2.021e-001 0.5816 *					
11) Totaal.P>0.161667 20 1.280e-001 0.3600					
22) Meandering:1,4 12 2.231e-032 0.3000 *					
23) Meandering:3,5 8 2.000e-002 0.4500 *					
3) Totaal.P>0.831714 56 8.009e-001 0.1679					
6) Beschaduwng:1 29 3.480e-002 0.0800 *					
7) Beschaduwng:2,3 27 3.019e-001 0.2622 *					



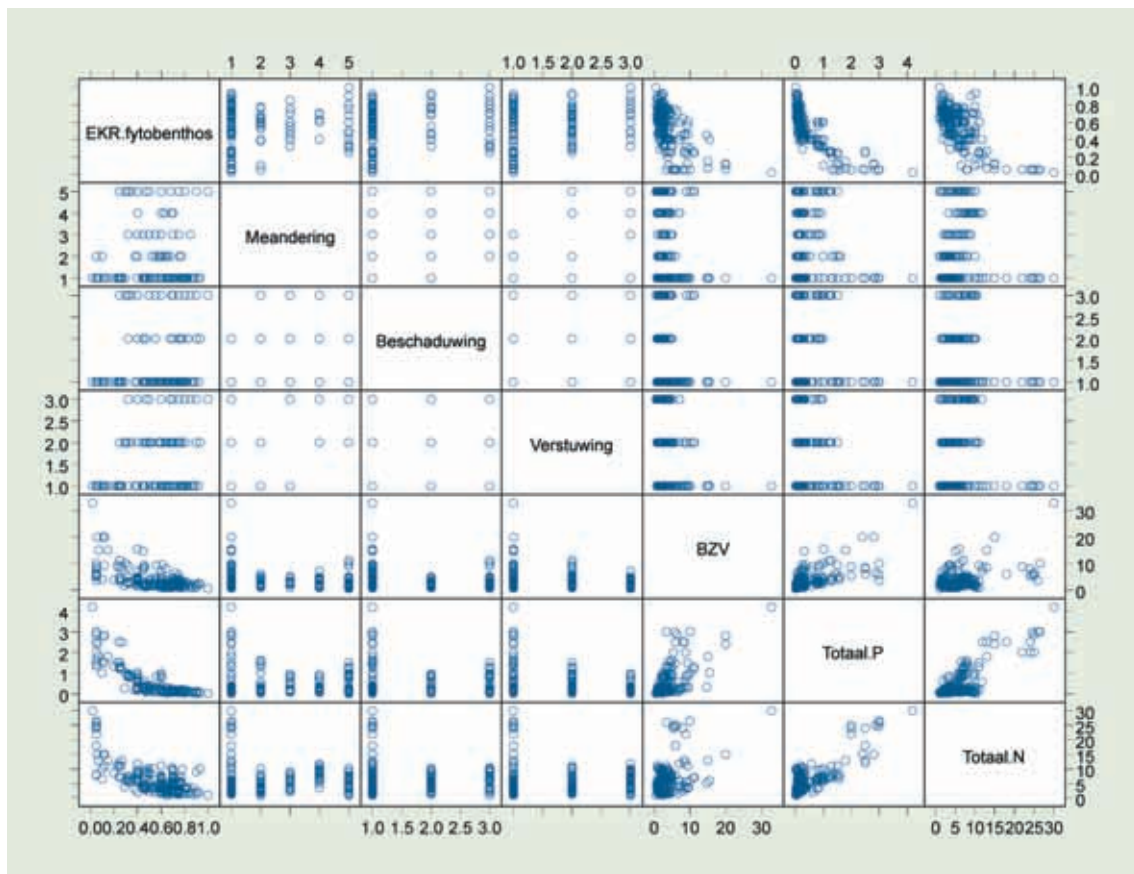
**Figuur A.11** EKR-voorspellingen met *randomForest* (500 bomen en een trainingsset van 80% van het totaal aantal beken). Kengetallen: MAD=0.029, RMSE=0.054 en  $R_{adj}^2=0.91$ . Zie voor een methodische uitleg paragraaf 3.2.2.



Figuur A.12 Histogrammen en dichtheden voor de belangrijkste stuurvariabele 'Totaal P'. Zie voor een methodische uitleg paragraaf 3.1.2.

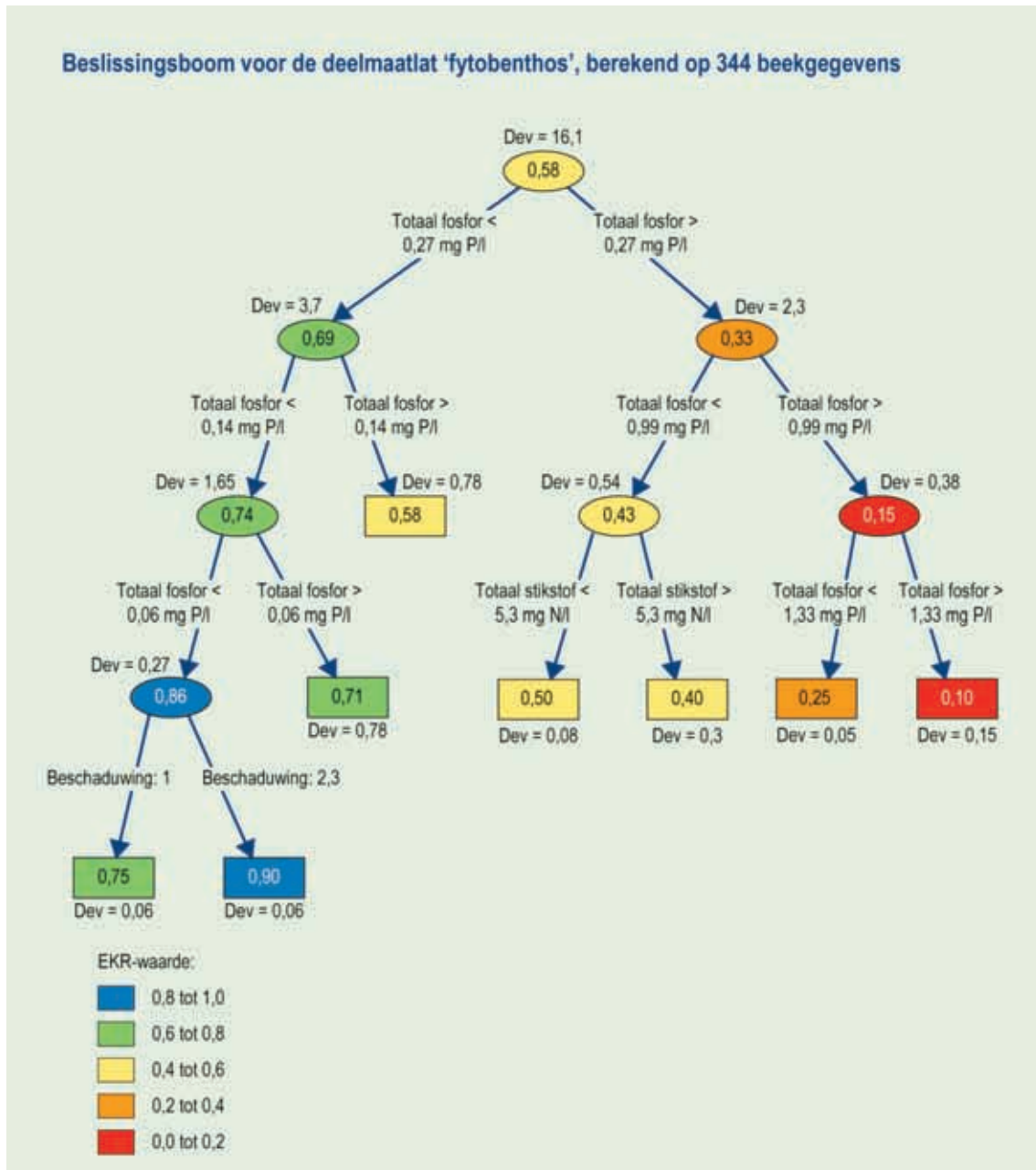
## A.4 Deelmaatlat fyto benthos

Figuur A.13 geeft de relatie tussen de deelmaatlat ‘fyto benthos’ enerzijds en de beschikbare stuurvariabelen anderzijds in de vorm van een scatterplotmatrix. De figuur laat in de bovenste rij zien dat EKR-waarden stijgen bij toenemende waarden van meandering, beschaduwing en verstuwing, en dalen bij toenemende waarden van BZV, Totaal P en Totaal N. Figuur A.14 en Tabel A.4 geven de geschatte regressieboom. ‘Totaal P’ blijkt de belangrijkste stuurvariabele te zijn. Figuur A.15 geeft het resultaat voor een validatiecheck op 500 regressiebomen. Figuur A.16 tenslotte geeft histogrammen met dichtheden voor de verschillende waarden van de belangrijkste stuurvariabele, namelijk ‘Totaal P’.



Figuur A.13 Scatterplotmatrix voor ‘fyto benthos’ (afhankelijke variabele) en zes stuurvariabelen. Zie voor een methodische uitleg van de figuur paragraaf 3.1.1.





**Figuur A.14** Regressieboom met 8 eindknopen (de rechthoeken). De gemiddelde EKR-waarde van een knoop staat in de rechthoek of ellips. De bijbehorende deviantie staat steeds onder de knoop. Het belang van de splitsingen is het grootst voor de eerste splitsing en neemt af naarmate 'gezakt' wordt in de boom. Knoopwaarden tussen 0.6 en 0.8 vallen in de klasse 'goede ecologische toestand' (GET), waarden tussen 0.8 en 1.0 in de klasse 'zeer goede ecologische kwaliteit' (ZGET). Zie voor een methodische uitleg paragraaf 3.2. De kleurcode correspondeert met Figuur 1.3. Door regressieboom verklaarde variantie bedraagt 86%.

**Tabel A.4 Kwantitatieve informatie bij geschatte regressieboom.**

Number of terminal nodes: 8

Residual mean deviance: 0.006756 = 2.27 / 336

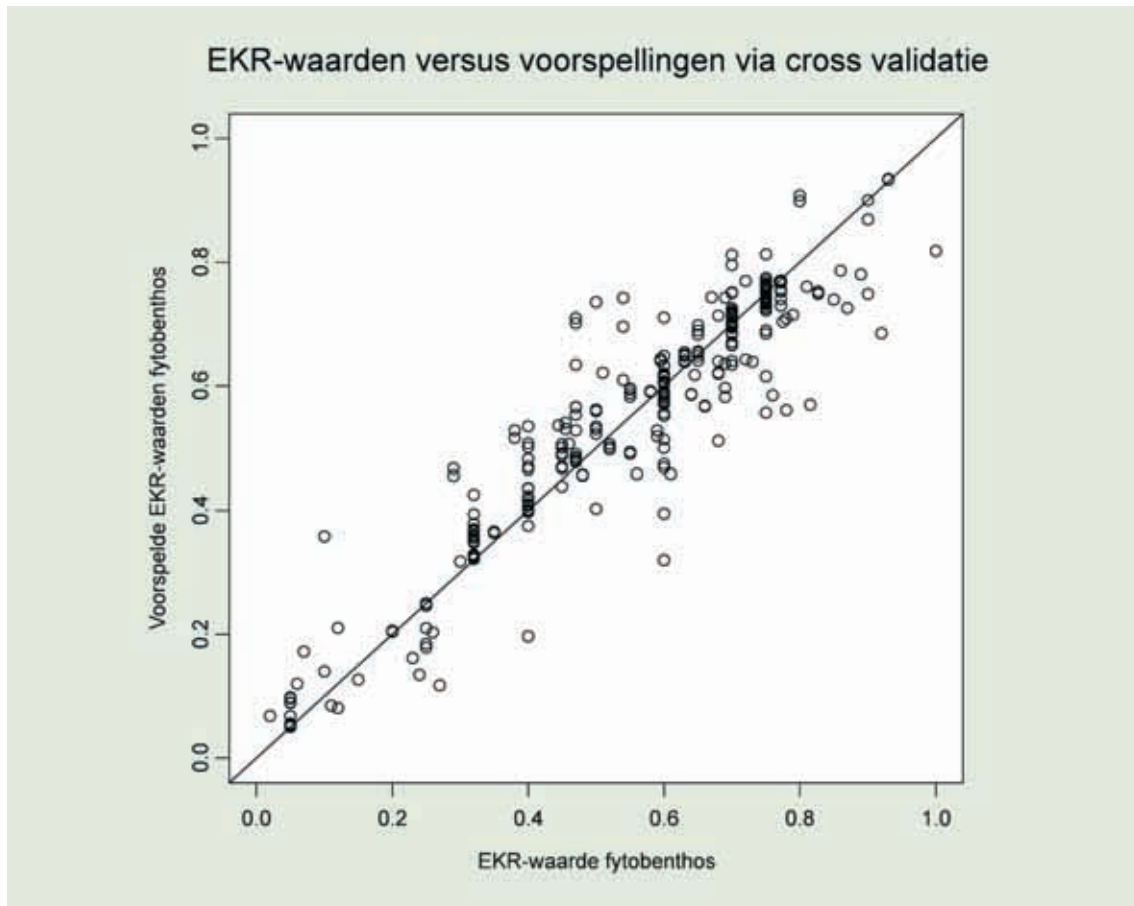
Distribution of residuals:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-2.863e-001	-4.875e-002	3.200e-003	-5.453e-016	4.078e-002	2.387e-001

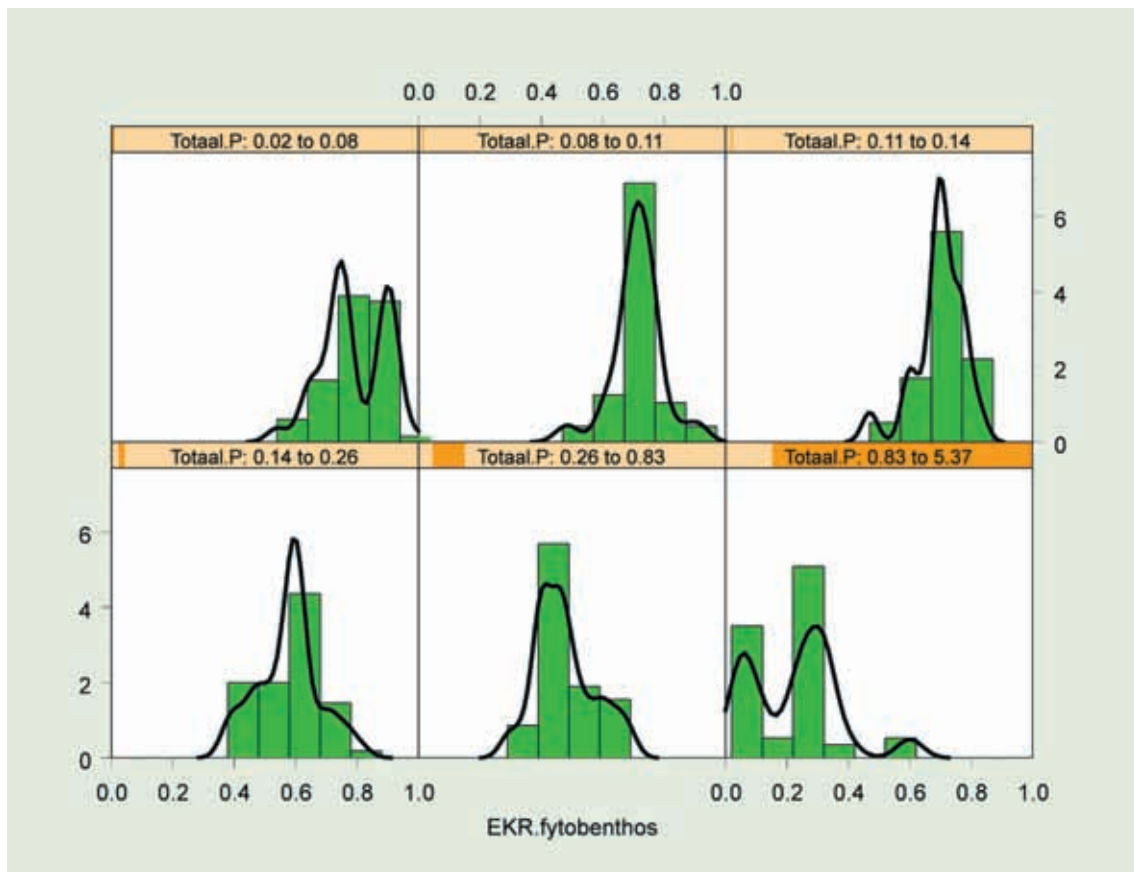
node), split, n, deviance, yval

\* denotes terminal node

- 1) root 344 16.07000 0.5822
- 2) Totaal.P<0.271667 239 3.71600 0.6926
- 4) Totaal.P<0.143333 171 1.65300 0.7388
- 8) Totaal.P<0.0645833 34 0.27020 0.8579
- 16) Beschaduwing:1 9 0.06420 0.7500 \*
- 17) Beschaduwing:2,3 25 0.06334 0.8968 \*
- 9) Totaal.P>0.0645833 137 0.78030 0.7092 \*
- 5) Totaal.P>0.143333 68 0.77890 0.5763 \*
- 3) Totaal.P>0.271667 105 2.82100 0.3311
- 6) Totaal.P<0.994167 67 0.54300 0.4324
- 12) Totaal.N<5.26583 24 0.08226 0.4987 \*
- 13) Totaal.N>5.26583 43 0.29610 0.3953 \*
- 7) Totaal.P>0.994167 38 0.38050 0.1526
- 14) Totaal.P<1.3255 13 0.05248 0.2469 \*
- 15) Totaal.P>1.3255 25 0.15240 0.1036 \*



**Figuur A.15** EKR-voorspellingen met randomForest (500 bomen en een trainingsset van 80% van het totaal aantal beken). Kengetallen: MAD=0.043, RMSE=0.068 en  $R_{adj}^2=0.89$ . Zie voor een methodische uitleg paragraaf 3.2.2.



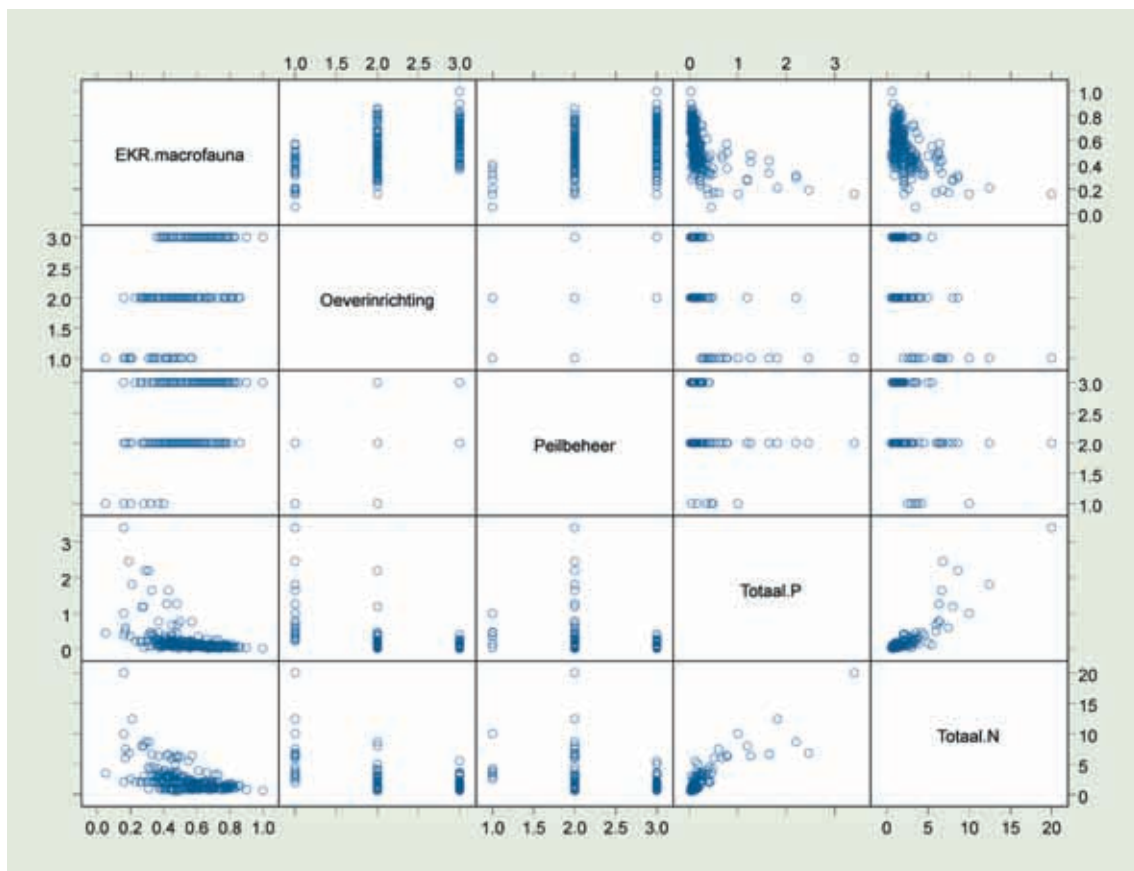
Figuur A.16 Histogrammen en dichtheden voor de belangrijkste stuurvariabele 'totaal P'. Zie voor een methodische uitleg paragraaf 3.1.2.

## Appendix B Regressiebomen voor meren

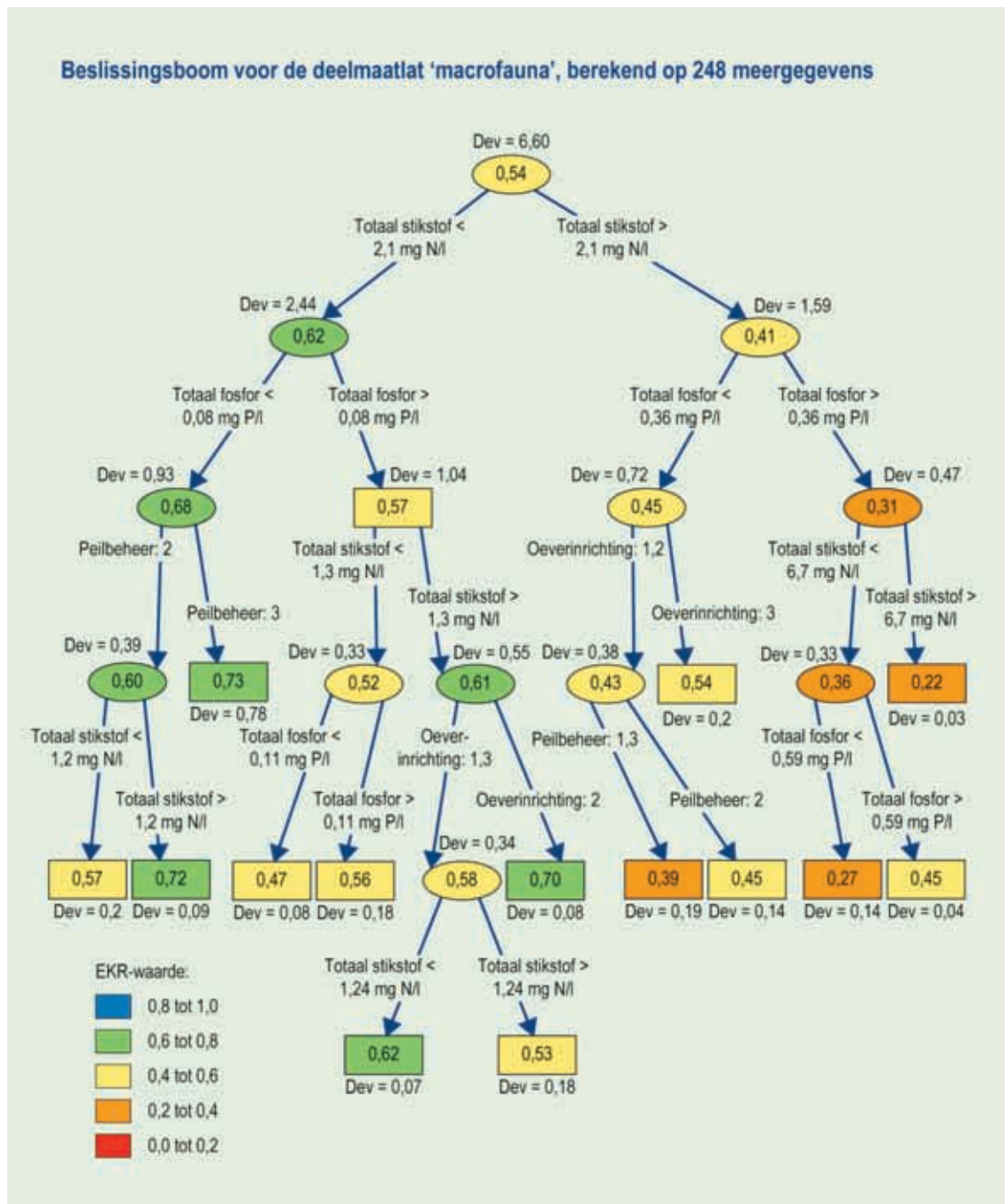
In deze appendix worden de regressieboom-schattingen gegeven voor de deelmaatlaten 'macrofauna' (paragraaf B.1), 'vissen' (paragraaf B.2), 'macrofyten' (paragraaf B.3) en 'fytoplankton' (paragraaf B.4) in meren. Fysisch-chemische stuurvariabelen zijn voor al deze deelmaatlaten: totaal fosfor en totaal stikstof. Hydromorfologische variabelen zijn: oeverinrichting en peilbeheer. Zie voor een uitleg van deze vier variabelen Tabel 2.2, het tweede paneel.

### B.1 Deelmaatlat macrofauna

Figuur B.1 geeft de relatie tussen de deelmaatlat 'macrofauna' enerzijds en de beschikbare stuurvariabelen anderzijds in de vorm van een scatterplotmatrix. De figuur laat in de bovenste rij zien dat EKR-waarden stijgen bij toenemende waarden van oeverinrichting en peilbeheer, en dalen bij toenemende waarden van Totaal P en Totaal N. Figuur B.2 en Tabel B.1 geven de geschatte regressieboom. 'Totaal N' blijkt de belangrijkste stuurvariabele te zijn. Figuur B.3 geeft het resultaat voor een validatiecheck op 500 regressiebomen. Tenslotte geeft Figuur B.4 histogrammen met dichtheden voor de verschillende waarden van de belangrijkste stuurvariabele, namelijk 'Totaal N'.



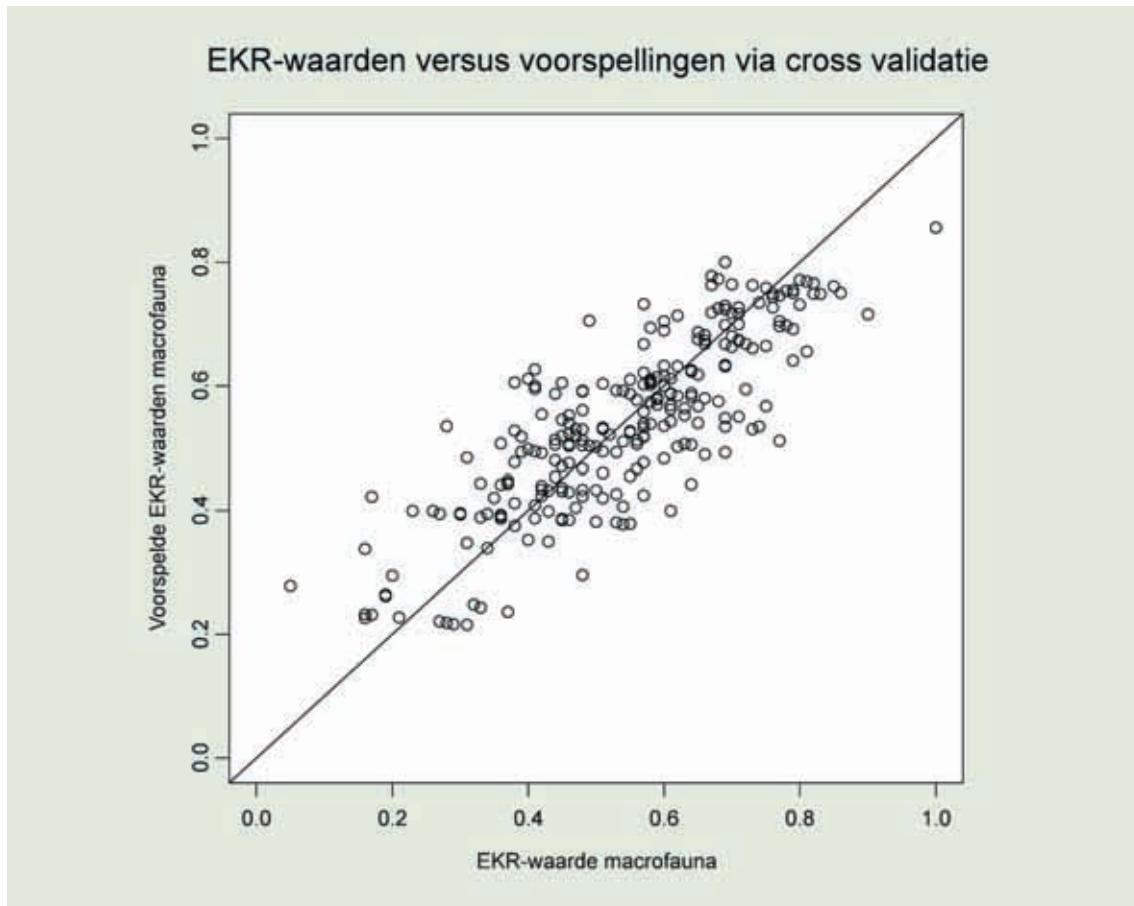
Figuur B.1 Scatterplotmatrix voor 'macrofauna' (afhankelijke variabele) en vier stuurvariabelen. Zie voor een methodische uitleg van de figuur paragraaf 3.1.1.



**Figuur B.2** Regressieboom met 14 eindknoten (rechthoeken). De gemiddelde EKR-waarde van een knoop staat in de rechthoek of ellips. De bijbehorende deviantie staat steeds onder de knoop. Het belang van de splitsingen is het grootst voor de eerste splitsing en neemt af naarmate ‘gezakt’ wordt in de boom. Knoopwaarden tussen 0.6 en 0.8 vallen in de klasse ‘goede ecologische toestand’ (GET), waarden tussen 0.8 en 1.0 in de klasse ‘zeer goede ecologische kwaliteit’ (ZGET). Zie voor een methodische uitleg paragraaf 3.2. De kleurcode correspondeert met Figuur 1.3. Door regressieboom verklaarde variantie bedraagt 71%.

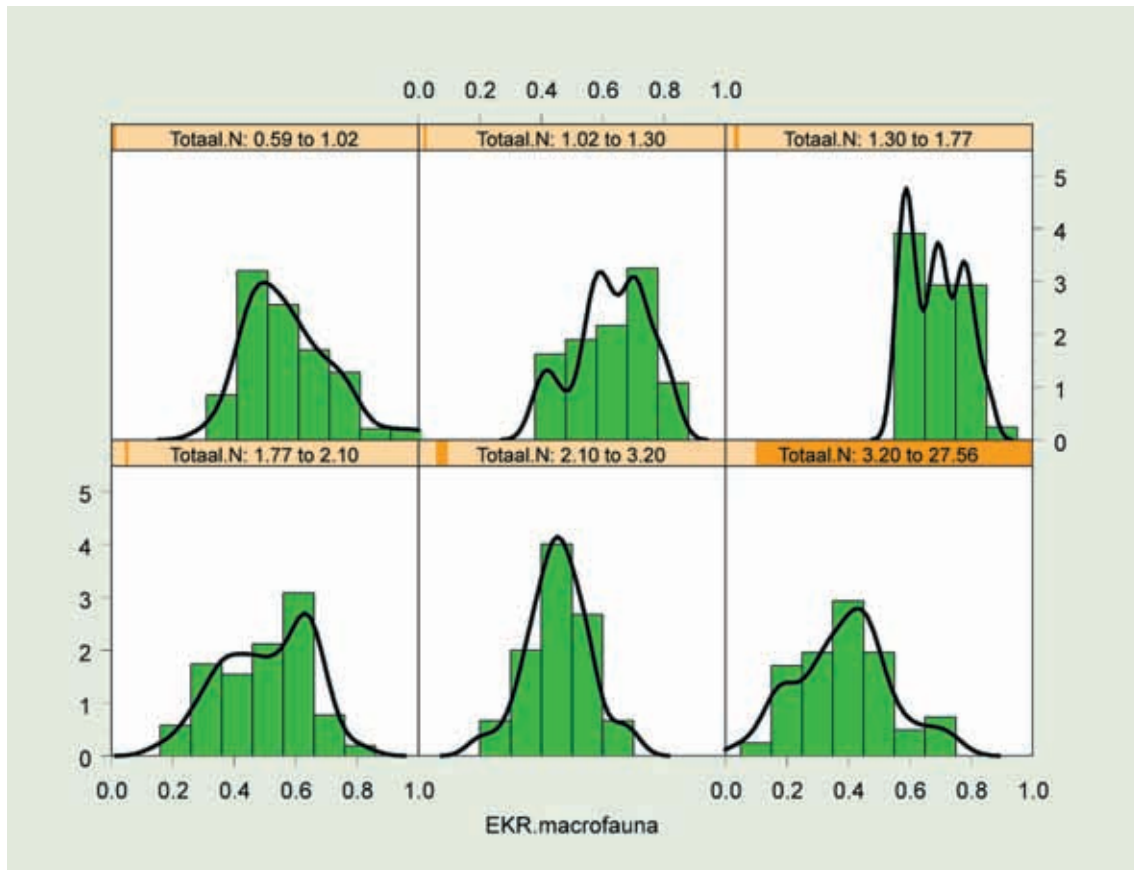
**Tabel B.1** Kwantitatieve informatie bij geschatte regressiebom.

Number of terminal nodes: 14					
Residual mean deviance: 0.008094 = 1.894 / 234					
Distribution of residuals:					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-2.230e-001	-5.306e-002	-8.333e-003	-2.194e-017	5.805e-002	2.680e-001
node), split, n, deviance, yval					
* denotes terminal node					
1) root 248 6.59600 0.5412					
2) Totaal.N<2.05 158 2.44400 0.6179					
4) Totaal.P<0.0758333 66 0.92740 0.6829					
8) Peilbeheer:2 25 0.38910 0.6024					
16) Totaal.N<1.24 19 0.20060 0.5663 *					
17) Totaal.N>1.24 6 0.08533 0.7167 *					
9) Peilbeheer:3 41 0.27760 0.7320 *					
5) Totaal.P>0.0758333 92 1.03800 0.5713					
10) Totaal.N<1.3325 38 0.32810 0.5216					
20) Totaal.P<0.10625 16 0.08454 0.4731 *					
21) Totaal.P>0.10625 22 0.17870 0.5568 *					
11) Totaal.N>1.3325 54 0.54990 0.6063					
22) Oeverinrichting:1,3 42 0.33960 0.5800					
45) Totaal.N>1.80667 19 0.18030 0.5279 *					
23) Oeverinrichting:2 12 0.07957 0.6983 *					
3) Totaal.N>2.05 90 1.58800 0.4064					
6) Totaal.P<0.358333 61 0.71630 0.4523					
12) Oeverinrichting:1,2 46 0.37630 0.4252					
24) Peilbeheer:1,3 20 0.18560 0.3875 *					
25) Peilbeheer:2 26 0.14040 0.4542 *					
13) Oeverinrichting:3 15 0.20280 0.5353 *					
7) Totaal.P>0.358333 29 0.47360 0.3100					
14) Totaal.N<6.74 19 0.32750 0.3558					
28) Totaal.P<0.5875 10 0.14200 0.2730 *					
29) Totaal.P>0.5875 9 0.04076 0.4478 *					
15) Totaal.N>6.74 10 0.03061 0.2230 *					



**Figuur B.3** EKR-voorspellingen met *randomForest* (500 bomen en een trainingsset van 80% van het totaal aantal meren). Kengetallen: MAD=0.074, RMSE=0.094 en  $R_{adj}^2=0.67$ . Zie voor een methodische uitleg paragraaf 3.2.2.

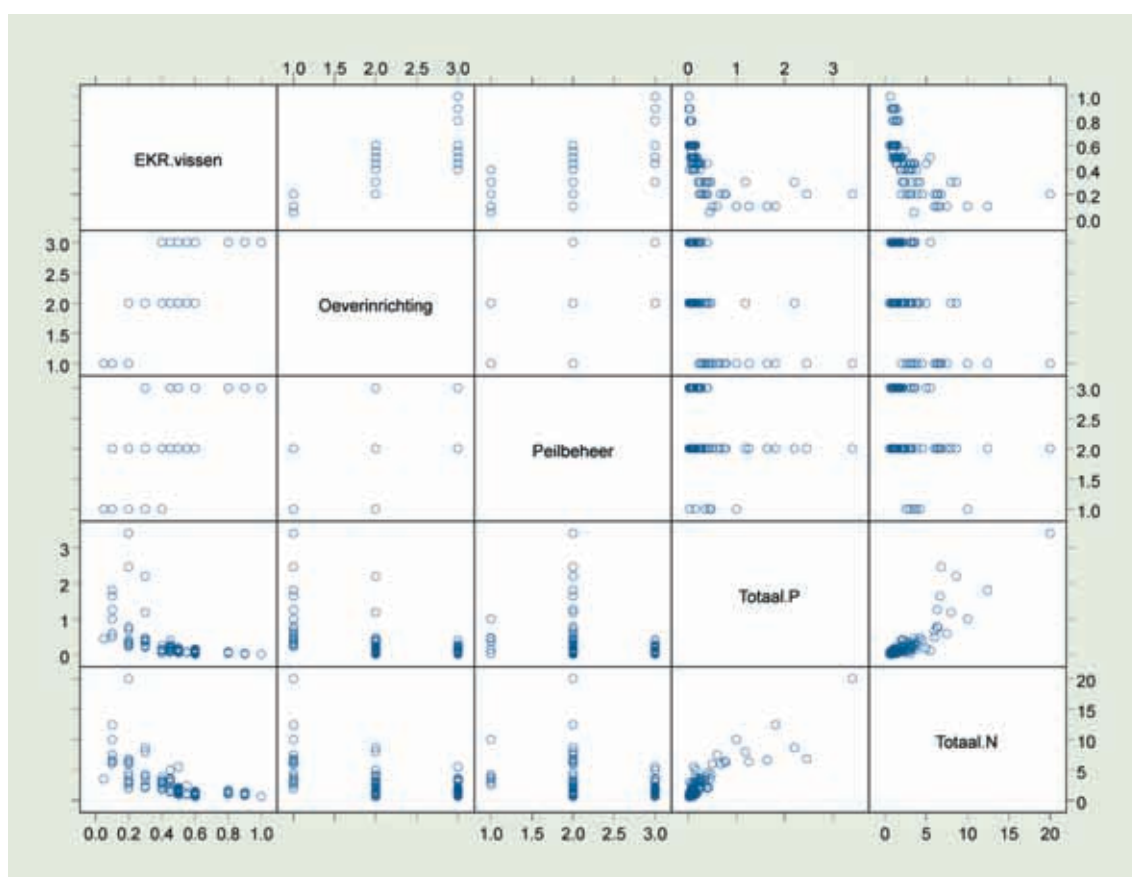




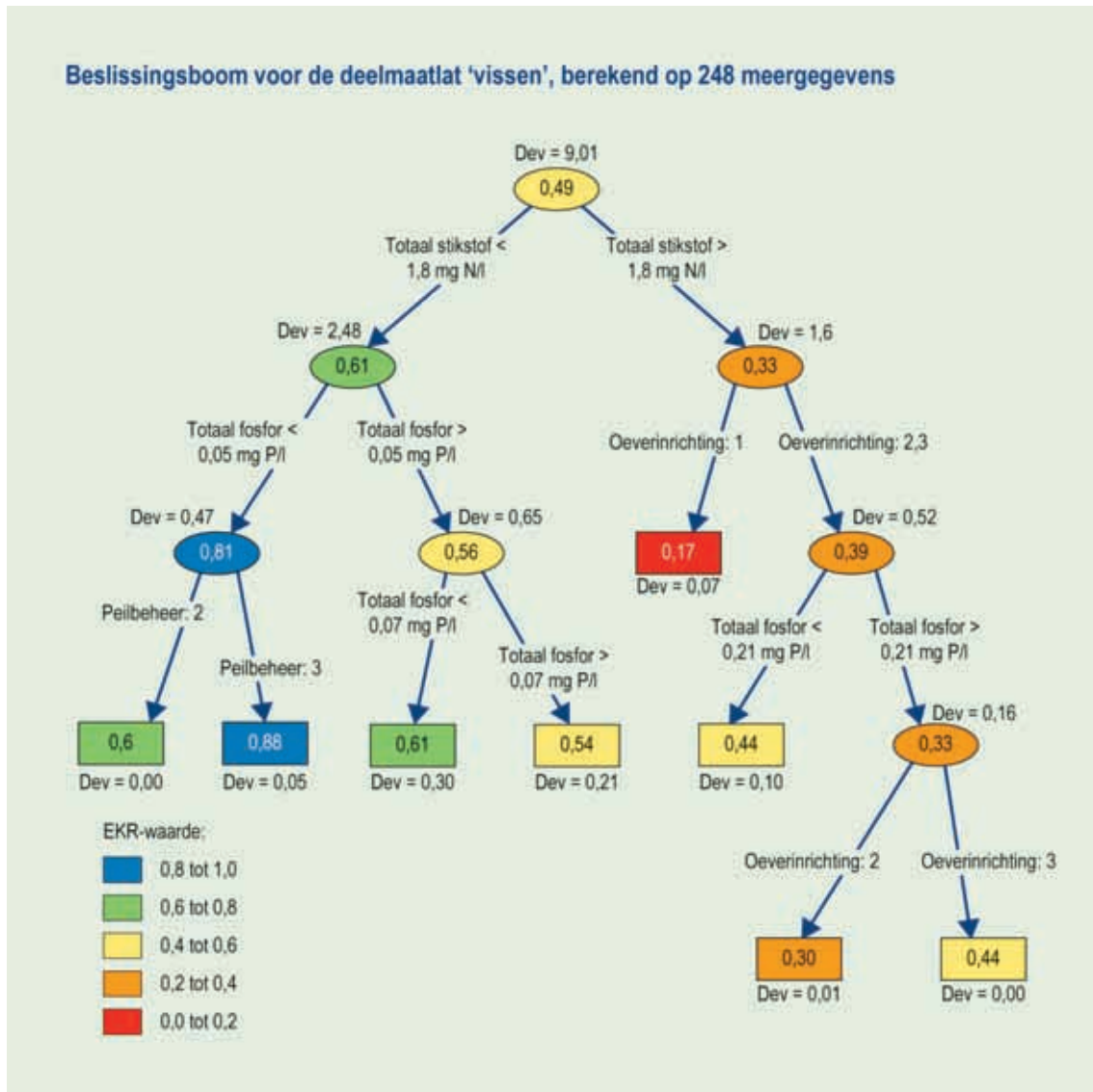
Figuur B.4 Histogrammen en dichtheden voor de belangrijkste stuurvariabele 'Totaal N'. Zie voor een methodische uitleg paragraaf 3.1.2.

## B.2 Deelmaatlat vissen

Figuur B.5 geeft de relatie tussen de deelmaatlat ‘vissen’ enerzijds en de beschikbare stuurvariabelen anderzijds in de vorm van een scatterplotmatrix. De figuur laat in de bovenste rij zien dat EKR-waarden stijgen bij toenemende waarden van oeverinrichting en peilbeheer, en dalen bij toenemende waarden van Totaal P en Totaal N. Figuur B.6 en Tabel B.2 geven de geschatte regressieboom. ‘Totaal N’ blijkt de belangrijkste stuurvariabele te zijn. Figuur B.7 geeft het resultaat voor een validatiecheck op 500 regressiebomen. Figuur B.8, ten slotte, geeft histogrammen met dichtheden voor de verschillende waarden van de belangrijkste stuurvariabele, namelijk ‘Totaal N’.



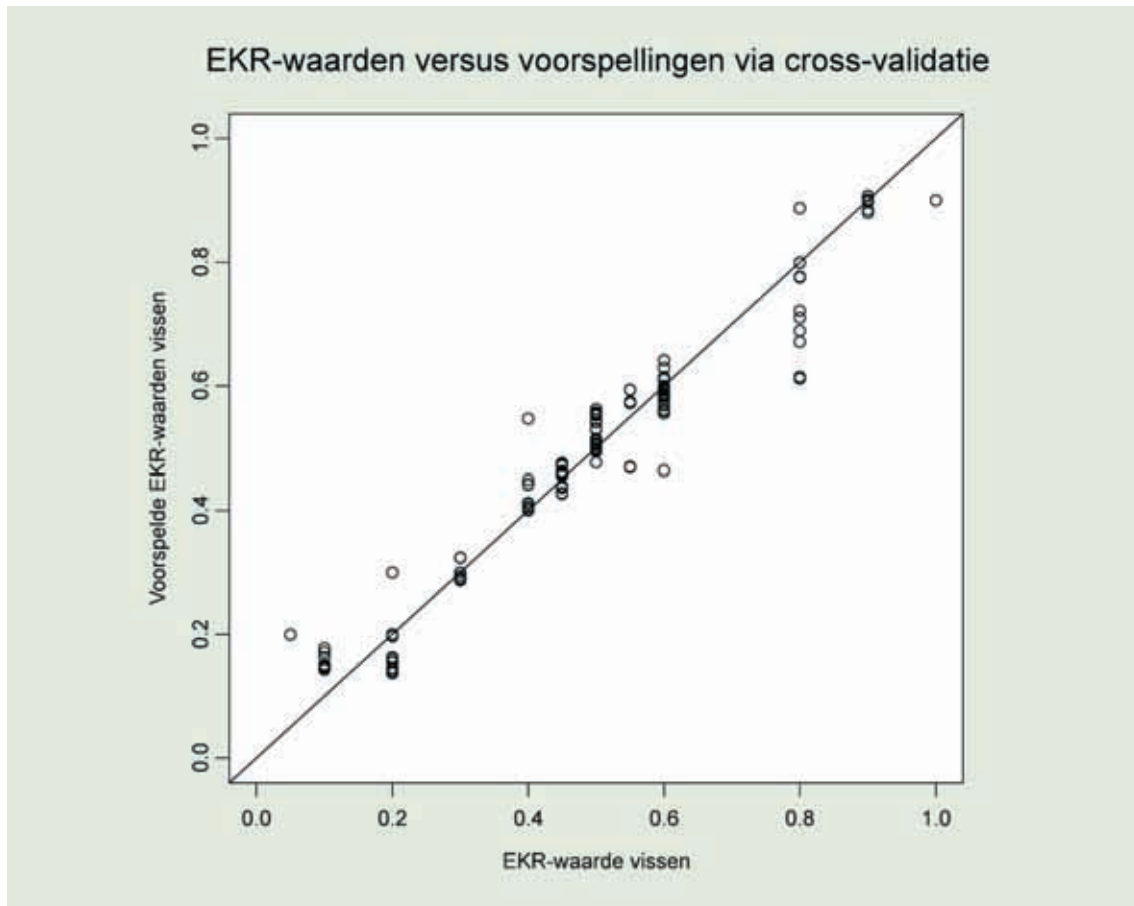
Figuur B.5 Scatterplotmatrix voor ‘vissen’ (afhankelijke variabele) en vier stuurvariabelen. Zie voor een methodische uitleg van de figuur paragraaf 2.2.1.



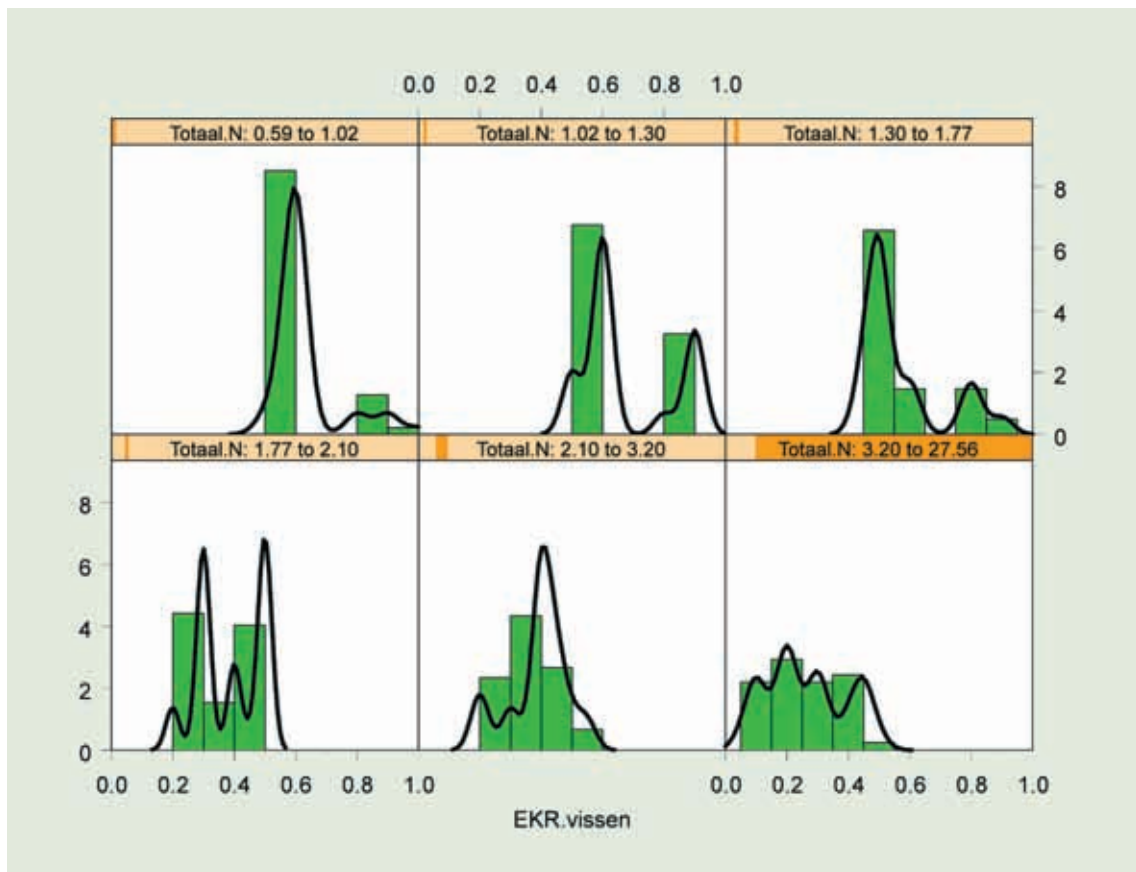
Figuur B.6 Regressieboom met 8 eindknoten (rechthoeken). De gemiddelde EKR-waarde van een knoop staat in de rechthoek of ellips. De bijbehorende deviantie staat steeds onder de knoop. Het belang van de splitsingen is het grootst voor de eerste splitsing en neemt af naarmate 'gezakt' wordt in de boom. Knoopwaarden tussen 0.6 en 0.8 vallen in de klasse 'goede ecologische toestand' (GET), waarden tussen 0.8 en 1.0 in de klasse 'zeer goede ecologische kwaliteit' (ZGET). Zie voor een methodische uitleg paragraaf 3.2. De kleurcode correspondeert met Figuur 1.3. Door regressieboom verklaarde variantie bedraagt 92%.

**Tabel B.2 Kwantitatieve informatie bij geschatte regressieboom.**

Number of terminal nodes: 8					
Residual mean deviance: 0.00311 = 0.7465 / 240					
Distribution of residuals:					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-1.161e-001	-3.836e-002	3.125e-003	-1.542e-016	3.393e-002	1.861e-001
node), split, n, deviance, yval					
* denotes terminal node					
1) root 248 9.015000 0.4869					
2) Totaal.N<1.8425 137 2.479000 0.6139					
4) Totaal.P<0.0491667 28 0.466800 0.8107					
8) Peilbeheer:2 7 0.000000 0.6000 *					
9) Peilbeheer:3 21 0.052380 0.8810 *					
5) Totaal.P>0.0491667 109 0.648200 0.5633					
10) Totaal.P<0.0741667 36 0.298100 0.6139 *					
11) Totaal.P>0.0741667 73 0.212600 0.5384 *					
3) Totaal.N>1.8425 111 1.601000 0.3302					
6) Oeverinrichting:1 28 0.070270 0.1661 *					
7) Oeverinrichting:2,3 83 0.522700 0.3855					
14) Totaal.P<0.205833 42 0.099580 0.4417 *					
15) Totaal.P>0.205833 41 0.155200 0.3280					
30) Oeverinrichting:2 32 0.009687 0.2969 *					
31) Oeverinrichting:3 9 0.003889 0.4389 *					



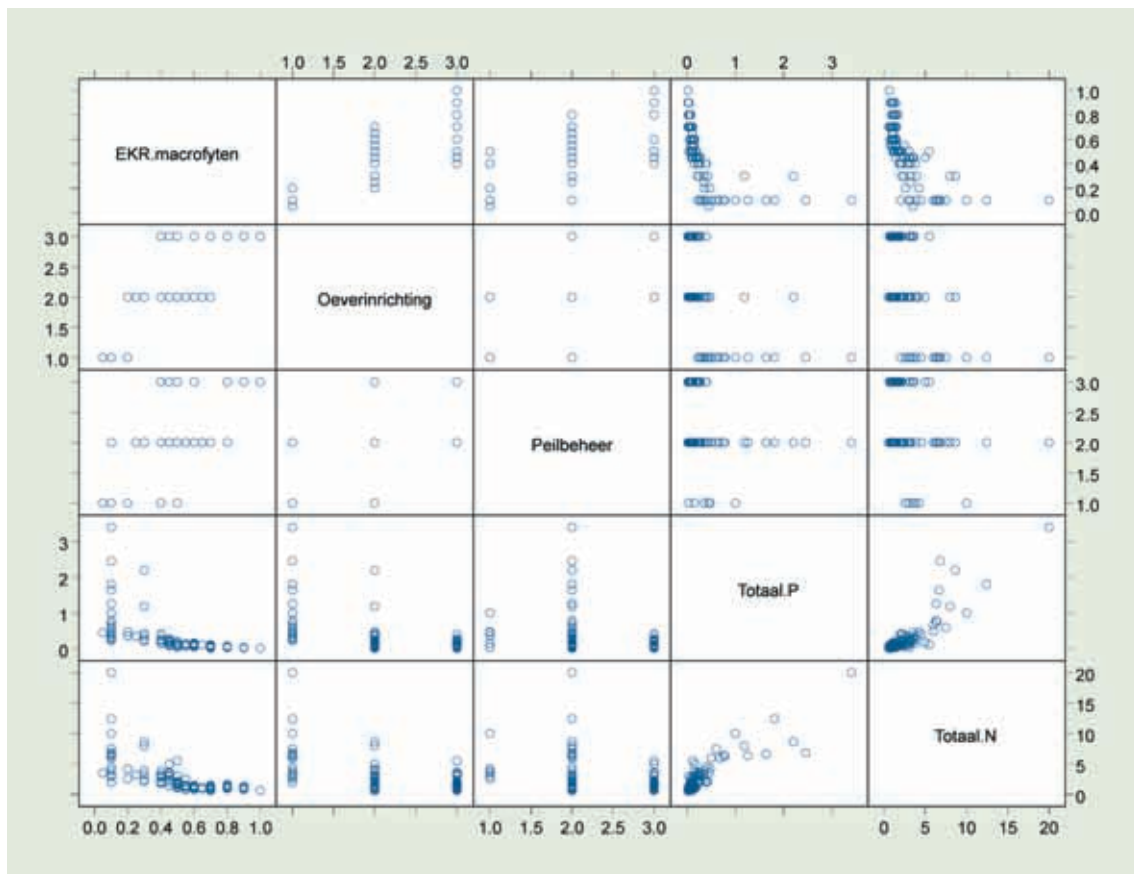
Figuur B.7 EKR-voorspellingen met *randomForest* (500 bomen en een trainingsset van 80% van het totaal aantal meren). Kengetallen: MAD=0.019, RMSE=0.036 en  $R_{adj}^2=0.96$ . Zie voor een methodische uitleg paragraaf 3.2.2.



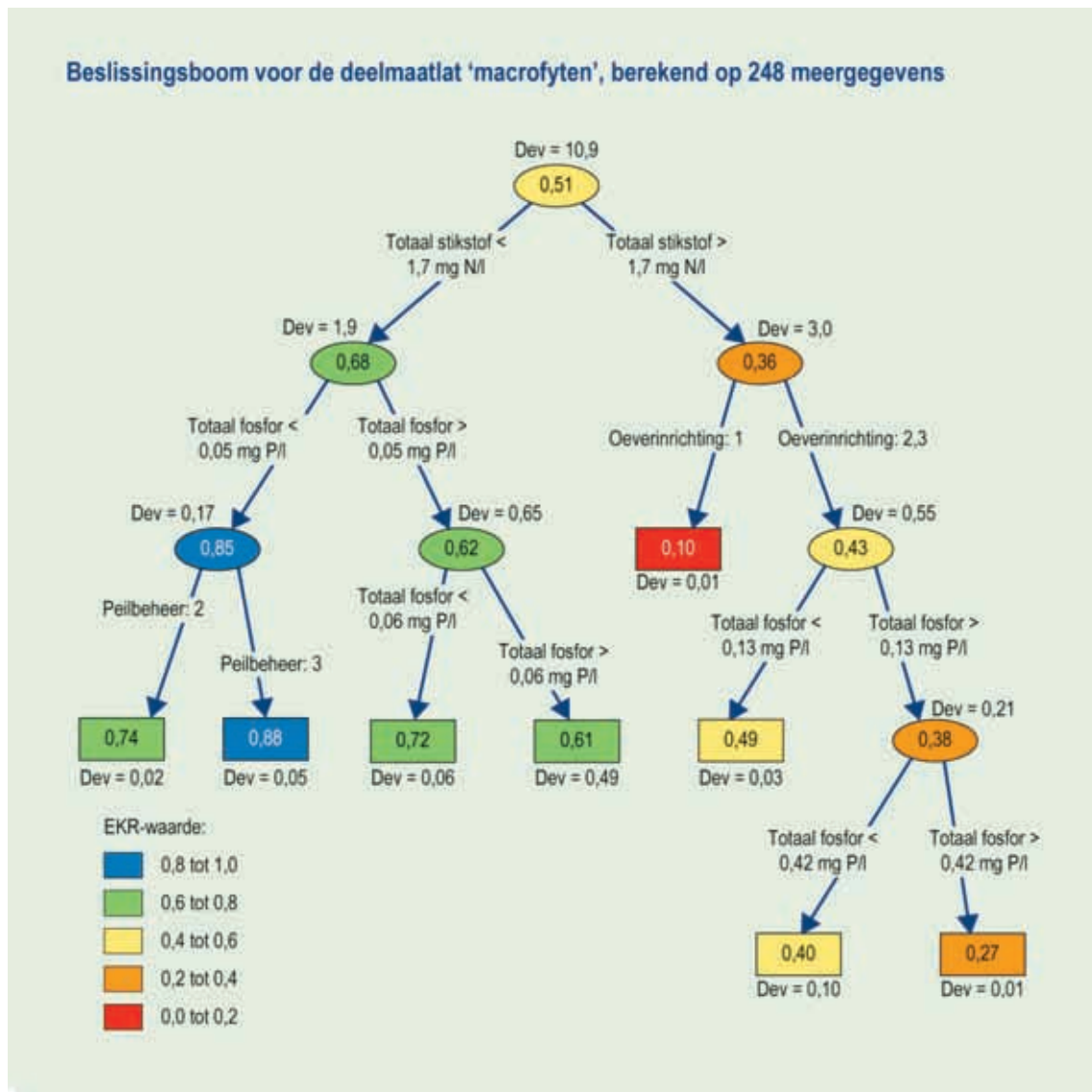
Figuur B.8 Histogrammen en dichtheden voor de belangrijkste stuurvariabele 'Totaal N'. Zie voor een methodische uitleg paragraaf 3.1.2.

### B.3 Deelmaatlat macrofyten

Figuur B.9 geeft de relatie tussen de deelmaatlat ‘macrofyten’ enerzijds en de beschikbare stuurvariabelen anderzijds in de vorm van een scatterplotmatrix. De figuur laat in de bovenste rij zien dat EKR-waarden stijgen bij toenemende waarden van oeverinrichting en peilbeheer, en dalen bij toenemende waarden van Totaal P en Totaal N. Figuur B.10 en Tabel B.3 geven de geschatte regressieboom. ‘Totaal N’ blijkt de belangrijkste stuurvariabele te zijn. Figuur B.11 geeft het resultaat voor een validatiecheck op 500 regressiebomen. Figuur B.12, ten slotte, geeft histogrammen met dichtheden voor de verschillende waarden van de belangrijkste stuurvariabele, namelijk ‘Totaal N’.



Figuur B.9 Scatterplotmatrix voor ‘macrofyten’ (afhankelijke variabele) en vier stuurvariabelen. Zie voor een methodische uitleg van de figuur paragraaf 3.1.1.

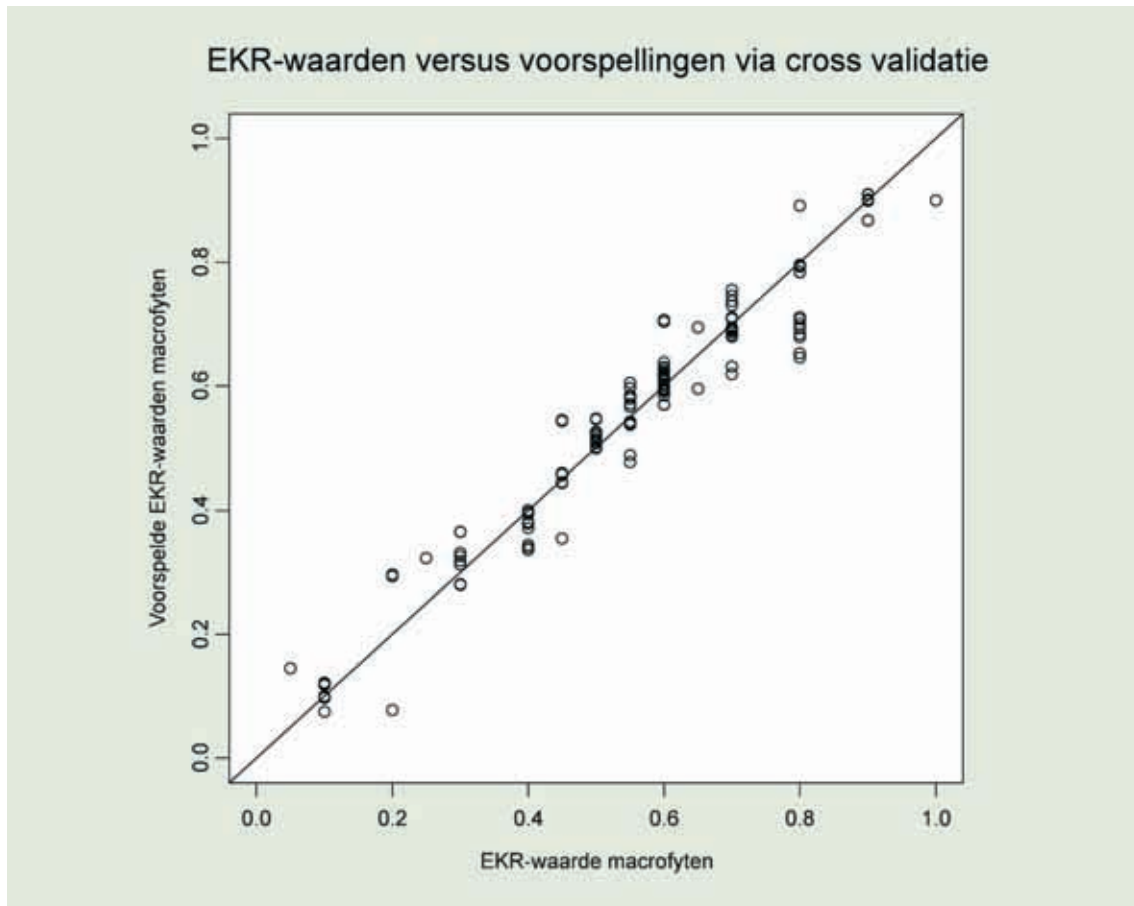


**Figuur B.10** Regressieboom met 8 eindknoten (de rechthoeken). De gemiddelde EKR-waarde van een knoop staat in de rechthoek of ellips. De bijbehorende deviantie staat steeds onder de knoop. Het belang van de splitsingen is het grootst voor de eerste splitsing en neemt af naarmate 'gezakt' wordt in de boom. Knoopwaarden tussen 0,6 en 0,8 vallen in de klasse 'goede ecologische toestand' (GET), waarden tussen 0,8 en 1,0 in de klasse 'zeer goede ecologische kwaliteit' (ZGET). Zie voor een methodische uitleg paragraaf 3.2. De kleurcode correspondeert met Figuur 1.3. Door regressieboom verklaarde variantie bedraagt 93%.

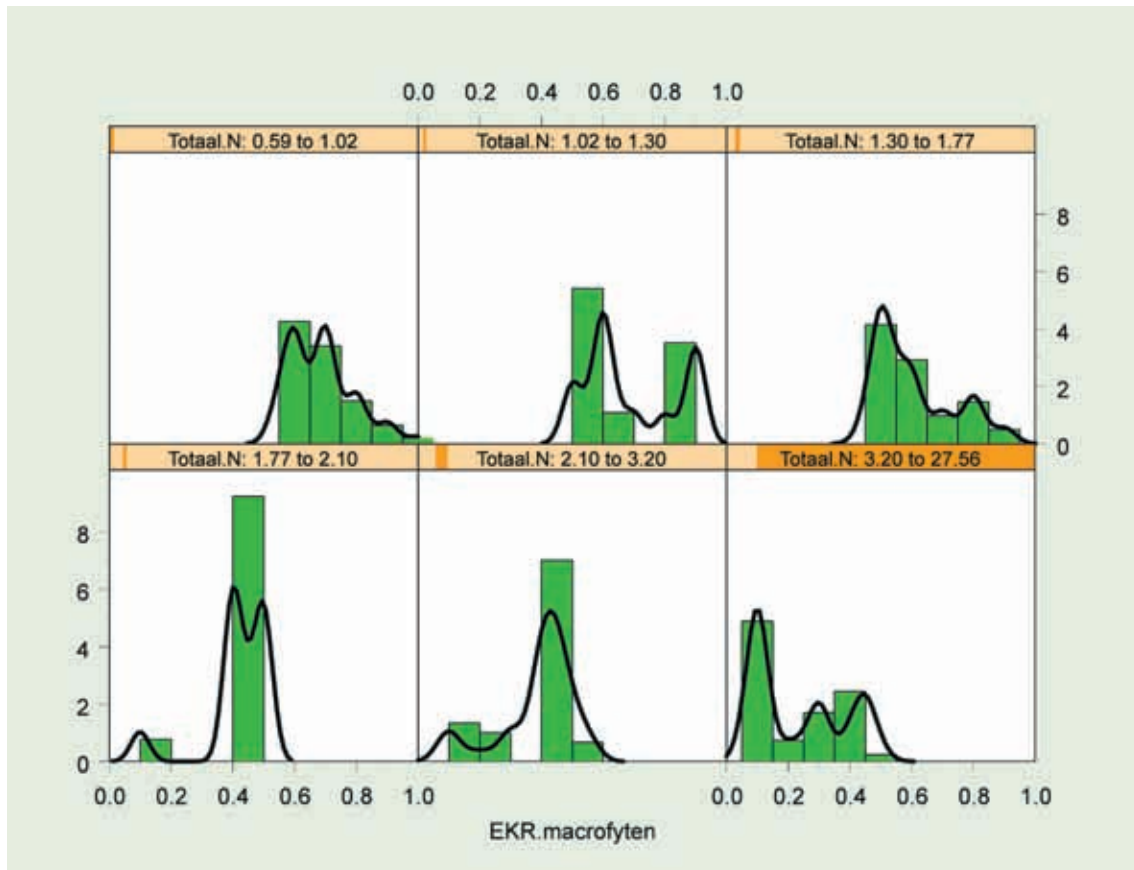


**Tabel B.3** Kwantitatieve informatie bij geschatte regressiebom.

Number of terminal nodes: 8					
Residual mean deviance: 0.003202 = 0.7685 / 240					
Distribution of residuals:					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-1.591e-001	-9.091e-003	1.961e-003	-1.217e-016	1.905e-002	1.909e-001
node), split, n, deviance, yval					
* denotes terminal node					
1) root 248 10.92000 0.5069					
2) Totaal.N<1.69714 114 1.89100 0.6763					
4) Totaal.P<0.0491667 28 0.16960 0.8464					
8) Peilbeheer:2 7 0.01714 0.7429 *					
9) Peilbeheer:3 21 0.05238 0.8810 *					
5) Totaal.P>0.0491667 86 0.64730 0.6209					
10) Totaal.P<0.0645 9 0.05556 0.7222 *					
11) Totaal.P>0.0645 77 0.48860 0.6091 *					
3) Totaal.N>1.69714 134 2.96800 0.3627					
6) Oeverinrichting:1 28 0.01241 0.1018 *					
7) Oeverinrichting:2,3 106 0.54660 0.4316					
14) Totaal.P<0.132917 48 0.02828 0.4906 *					
15) Totaal.P>0.132917 58 0.21280 0.3828					
30) Totaal.P<0.415 51 0.09980 0.3980 *					
31) Totaal.P>0.415 7 0.01429 0.2714 *					



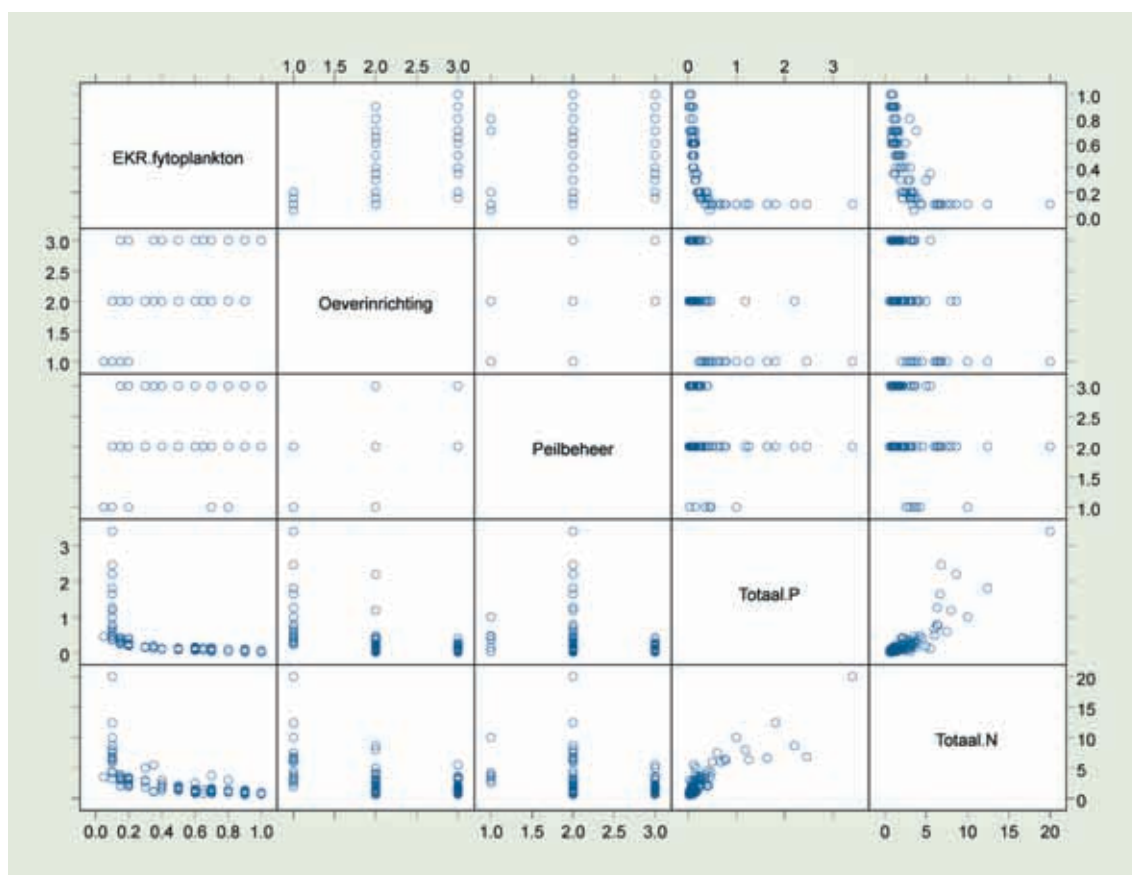
**Figuur B.11** EKR-voorspellingen met *randomForest* (500 bomen en een trainingsset van 80% van het totaal aantal meren). Kengetallen: MAD=0.020, RMSE=0.036 en  $R_{adj}^2=0.97$ . Zie voor een methodische uitleg paragraaf 3.2.2.



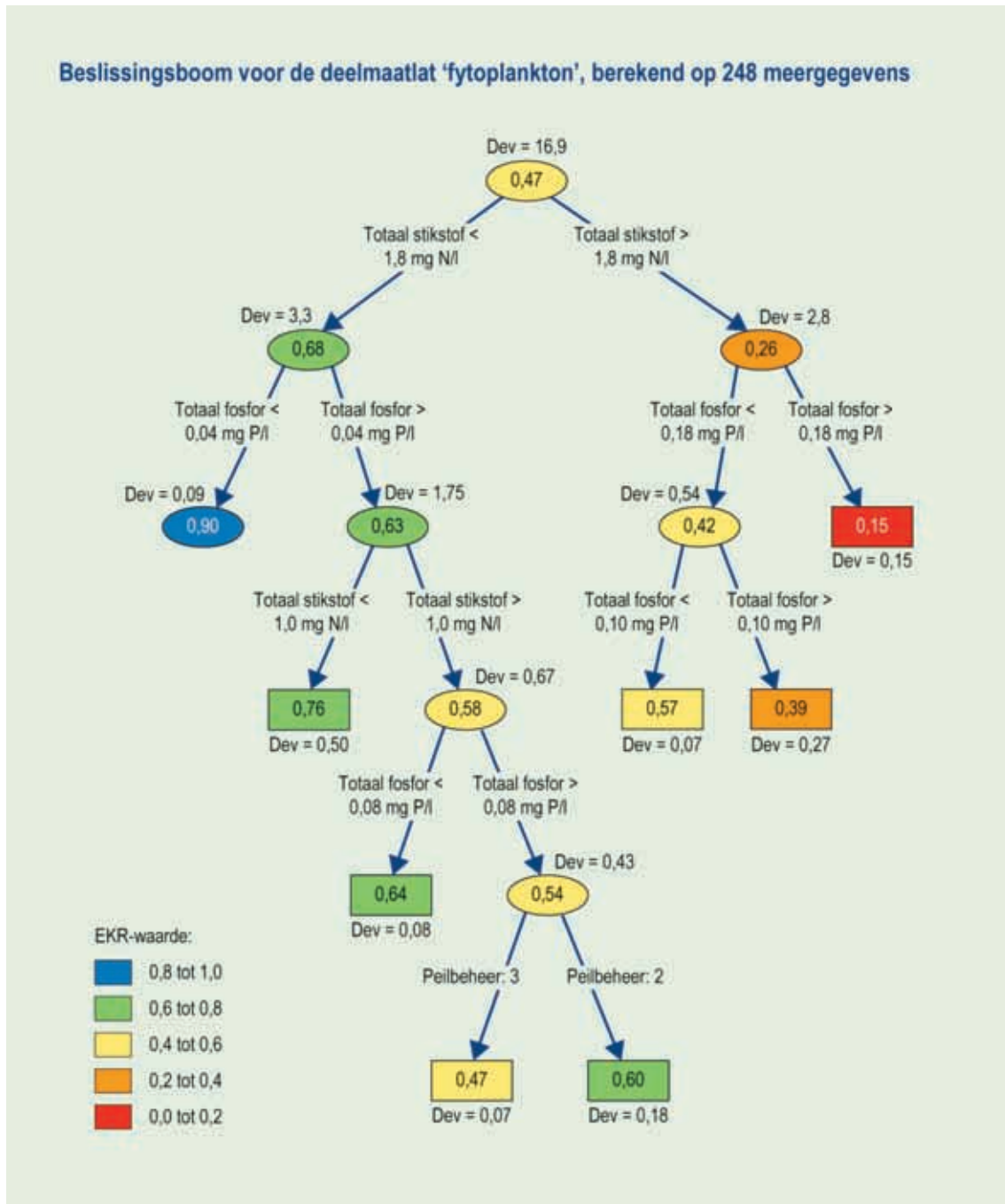
Figuur B.12 Histogrammen en dichtheden voor de belangrijkste stuurvariabele 'Totaal N'. Zie voor een methodische uitleg paragraaf 3.1.2

## B.4 Deelmaatlat fytoplankton

Figuur B.13 geeft de relatie tussen de deelmaatlat ‘fytoplankton’ enerzijds en de beschikbare stuurvariabelen anderzijds in de vorm van een scatterplotmatrix. De figuur laat in de bovenste rij zien dat EKR-waarden stijgen bij toenemende waarden van oeverinrichting en peilbeheer, en dalen bij toenemende waarden van Totaal P en Totaal N. Figuur B.14 en Tabel B.4 geven de geschatte regressieboom. ‘Totaal N’ blijkt de belangrijkste stuurvariabele te zijn. Figuur B.15 geeft het resultaat voor een validatiecheck op 500 regressiebomen. Figuur B.16, ten slotte, geeft histogrammen met dichtheden voor de verschillende waarden van de belangrijkste stuurvariabele, namelijk ‘Totaal N’.



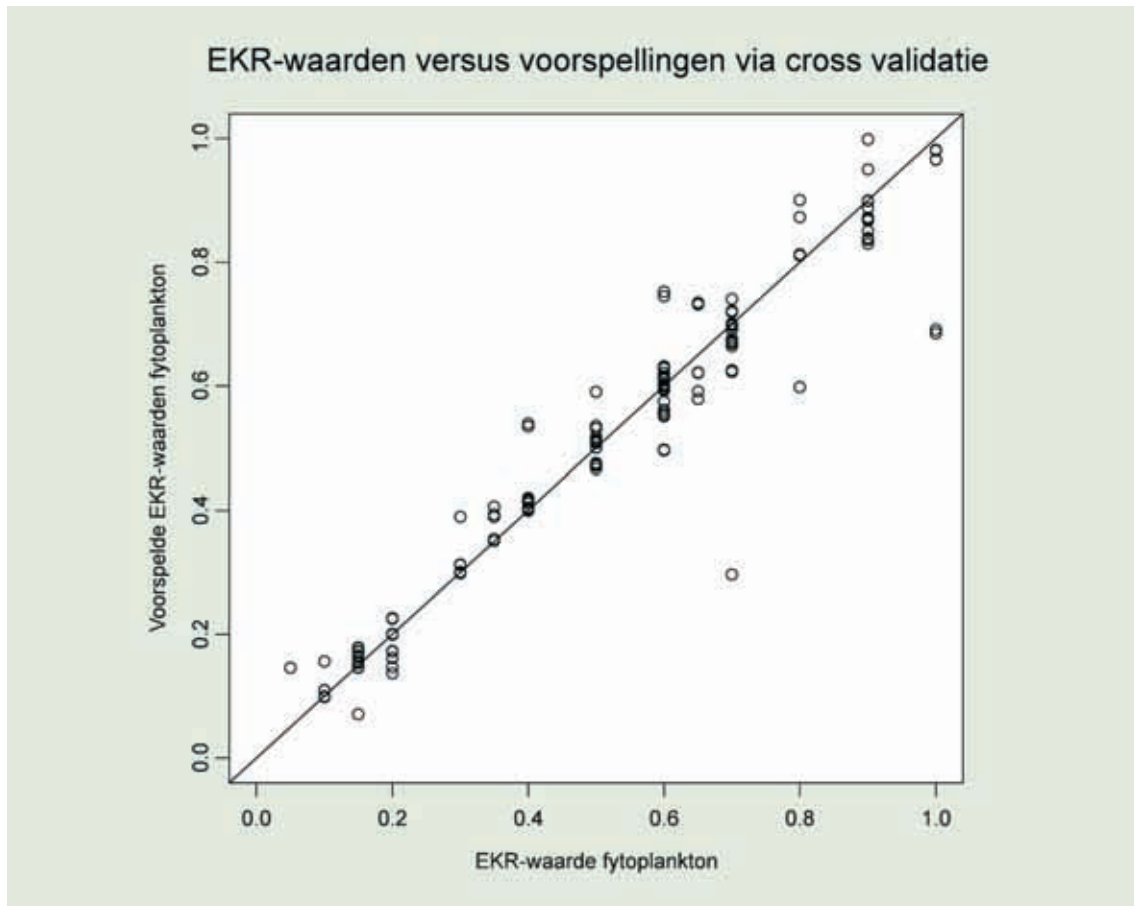
Figuur B.13 Scatterplotmatrix voor ‘fytoplankton’ (afhankelijke variabele) en vier stuurvariabelen. Zie voor een methodische uitleg van de figuur paragraaf 3.1.1.



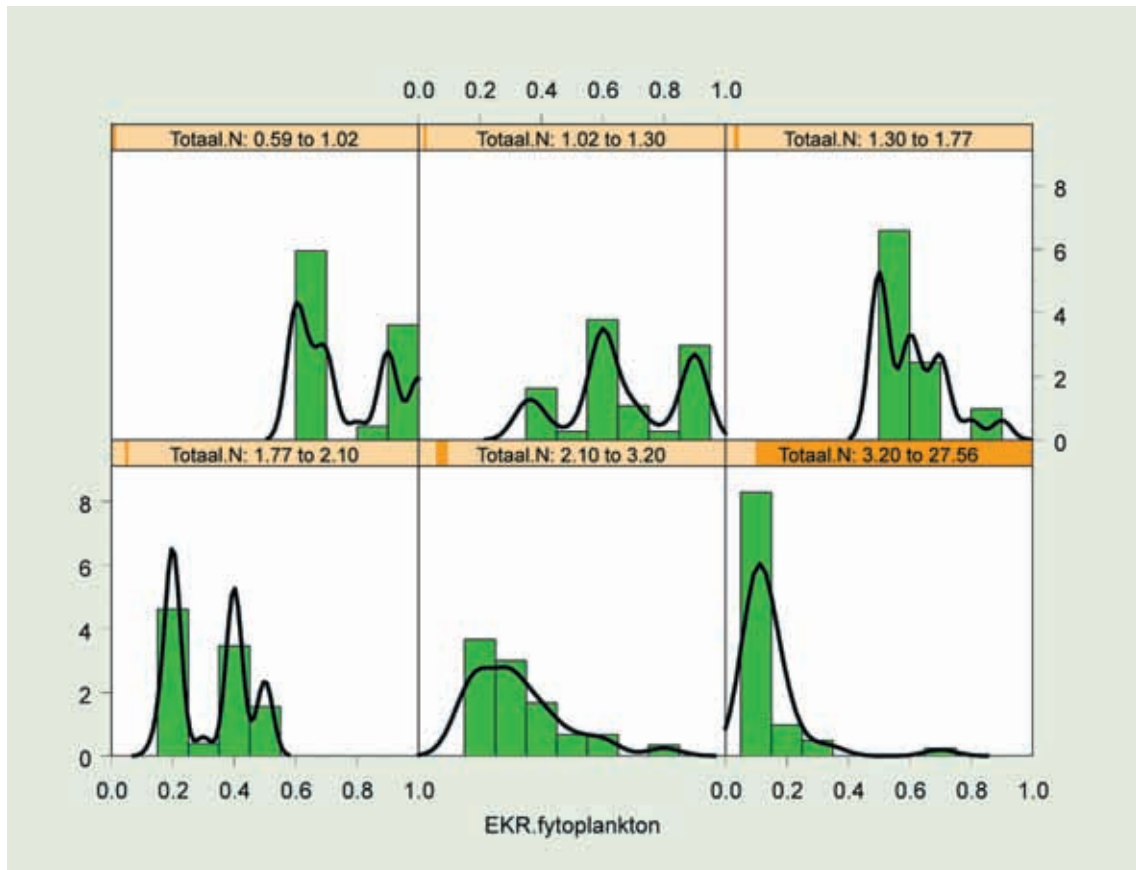
**Figuur B.14** Regressieboom met 8 eindknoten (de rechthoeken). De gemiddelde EKR-waarde van een knoop staat in de rechthoek of ellips. De bijbehorende deviantie staat steeds onder de knoop. Het belang van de splitsingen is het grootst voor de eerste splitsing en neemt af naarmate 'gezakt' wordt in de boom. Knoopwaarden tussen 0,6 en 0,8 vallen in de klasse 'goede ecologische toestand' (GET), waarden tussen 0,8 en 1,0 in de klasse 'zeer goede ecologische kwaliteit' (ZGET). Zie voor een methodische uitleg paragraaf 3.2. De kleurcode correspondeert met Figuur 1.3. Door regressieboom verklaarde variantie bedraagt 92%.

**Tabel B.4 Kwantitatieve informatie bij geschatte regressieboom.**

Number of terminal nodes: 8					
Residual mean deviance: 0.005872 = 1.409 / 240					
Distribution of residuals:					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-2.000e-001	-5.411e-002	2.220e-016	2.115e-017	4.589e-002	3.081e-001
node), split, n, deviance, yval					
* denotes terminal node					
1) root 248 16.94000 0.4720					
2) Totaal.N<1.77988 125 3.33500 0.6796					
4) Totaal.P<0.0425 24 0.08958 0.9042 *					
5) Totaal.P>0.0425 101 1.74800 0.6262					
10) Totaal.N<1.0095 25 0.49840 0.7580 *					
11) Totaal.N>1.0095 76 0.67280 0.5829					
22) Totaal.P<0.0775 34 0.07765 0.6353 *					
23) Totaal.P>0.0775 42 0.42620 0.5405					
46) Peilbeheer:3 19 0.07105 0.4684 *					
47) Peilbeheer:2 23 0.17500 0.6000 *					
3) Totaal.N>1.77988 123 2.74300 0.2610					
6) Totaal.P<0.180833 50 0.53810 0.4170					
12) Totaal.P<0.0983333 7 0.07429 0.5714 *					
13) Totaal.P>0.0983333 43 0.26970 0.3919 *					
7) Totaal.P>0.180833 73 0.15380 0.1541 *					



**Figuur B.15** EKR-voorspellingen met *randomForest* (500 bomen en een trainingsset van 80% van het totaal aantal meren). Kengetallen: MAD=0.025, RMSE=0.053 en  $R_{adj}^2=0.96$ . Zie voor een methodische uitleg paragraaf 3.2.2.



Figuur B.16 Histogrammen en dichtheden voor de belangrijkste stuurvariabele 'Totaal N'. Zie voor een methodische uitleg paragraaf 3.1.2.

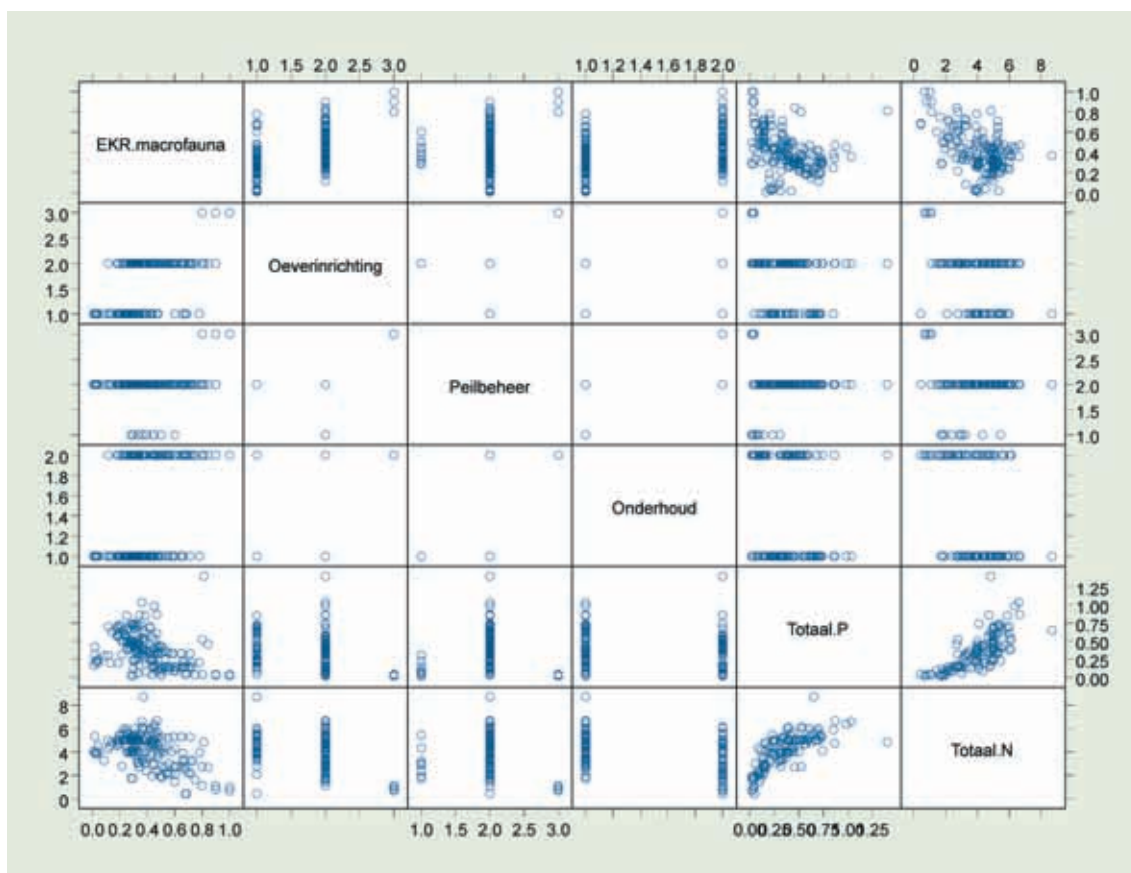


## Appendix C Regressiebomen voor kanalen

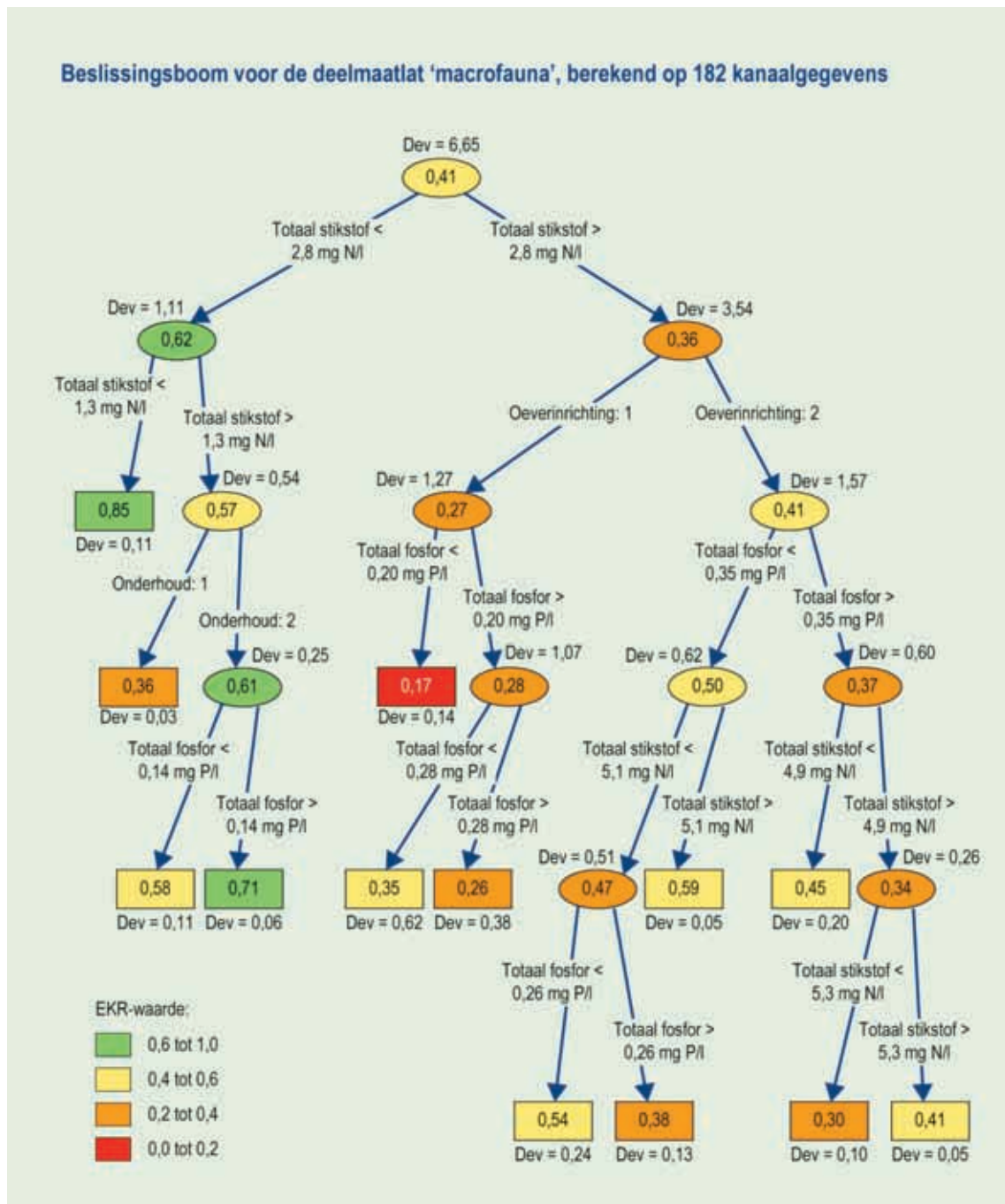
In deze appendix worden de regressieboom-schattingen gegeven voor de deelmaatlaten ‘macrofauna’ (paragraaf C.1), ‘vissen’ (C.2), ‘macrofyten’ (C.3) en ‘fytoplankton’ (C.4) in kanalen. Fysisch-chemische stuurvariabelen zijn voor al deze deelmaatlaten: totaal fosfor en totaal stikstof. Hydromorfologische variabelen zijn: oeverinrichting, peilbeheer en onderhoud. Zie voor een uitleg van deze vijf variabelen Tabel 2.2, het derde paneel.

### C.1 Deelmaatlat macrofauna

Figuur C.1 geeft de relatie tussen de deelmaatlat ‘macrofauna’ enerzijds en de beschikbare stuurvariabelen anderzijds in de vorm van een scatterplotmatrix. De figuur laat in de bovenste rij zien dat EKR-waarden stijgen bij toenemende waarden van oeverinrichting, peilbeheer en onderhoud, en dalen bij toenemende waarden van Totaal P en Totaal N. Figuur C.2 en Tabel C.1 geven de geschatte regressieboom. ‘Totaal N’ blijkt de belangrijkste stuurvariabele te zijn. Figuur C.3 geeft het resultaat voor een validatiecheck op 500 regressiebomen. Figuur C.4, ten slotte, geeft histogrammen met dichtheden voor de verschillende waarden van de belangrijkste stuurvariabele, namelijk ‘Totaal N’.



Figuur C.1 Scatterplotmatrix voor ‘macrofauna’ (afhankelijke variabele) en vijf stuurvariabelen. Zie voor een methodische uitleg van de figuur paragraaf 3.1.1.



**Figuur C.2** Regressieboom met 13 eindknoten (rechthoeken). De gemiddelde EKR-waarde van een knoop staat in de rechthoek of ellips. De bijbehorende deviantie staat steeds onder de knoop. Het belang van de splitsingen is het grootst voor de eerste splitsing en neemt af naarmate ‘gezakt’ wordt in de boom. Knoopwaarden tussen 0.6 en 1.0 vallen in de klasse ‘goed ecologisch potentieel’ (GEP). Zie voor een methodische uitleg paragraaf 3.2. De kleurcode correspondeert met Figuur 1.3. Door regressieboom verklaarde variantie bedraagt 67%.

**Tabel C.1** Kwantitatieve informatie bij geschatte regressiebboom.

Number of terminal nodes: 13

Residual mean deviance: 0.01315 = 2.223 / 169

Distribution of residuals:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-3.230e-001	-5.890e-002	-3.122e-003	-1.342e-017	6.239e-002	4.320e-001

node), split, n, deviance, yval

\* denotes terminal node

1) root 182 6.64600 0.4109

2) Totaal.N&lt;2.76833 36 1.10500 0.6223

4) Totaal.N&lt;1.27167 7 0.10890 0.8520 \*

5) Totaal.N&gt;1.27167 29 0.53740 0.5668

10) Onderhoud:1 5 0.03185 0.3626 \*

11) Onderhoud:2 24 0.25380 0.6093

22) Totaal.P&lt;0.141429 18 0.10980 0.5753 \*

23) Totaal.P&gt;0.141429 6 0.06079 0.7113 \*

3) Totaal.N&gt;2.76833 146 3.53600 0.3587

6) Oeverinrichting:1 55 1.26800 0.2695

12) Totaal.P&lt;0.201667 5 0.14110 0.1697 \*

13) Totaal.P&gt;0.201667 50 1.07200 0.2795

26) Totaal.P&lt;0.278333 12 0.61580 0.3480 \*

27) Totaal.P&gt;0.278333 38 0.38160 0.2579 \*

7) Oeverinrichting:2 91 1.56600 0.4127

14) Totaal.P&lt;0.34625 30 0.62440 0.4996

28) Totaal.N&lt;5.10125 23 0.50570 0.4736

56) Totaal.P&lt;0.2575 14 0.24210 0.5354 \*

57) Totaal.P&gt;0.2575 9 0.12710 0.3775 \*

29) Totaal.N&gt;5.10125 7 0.05198 0.5851 \*

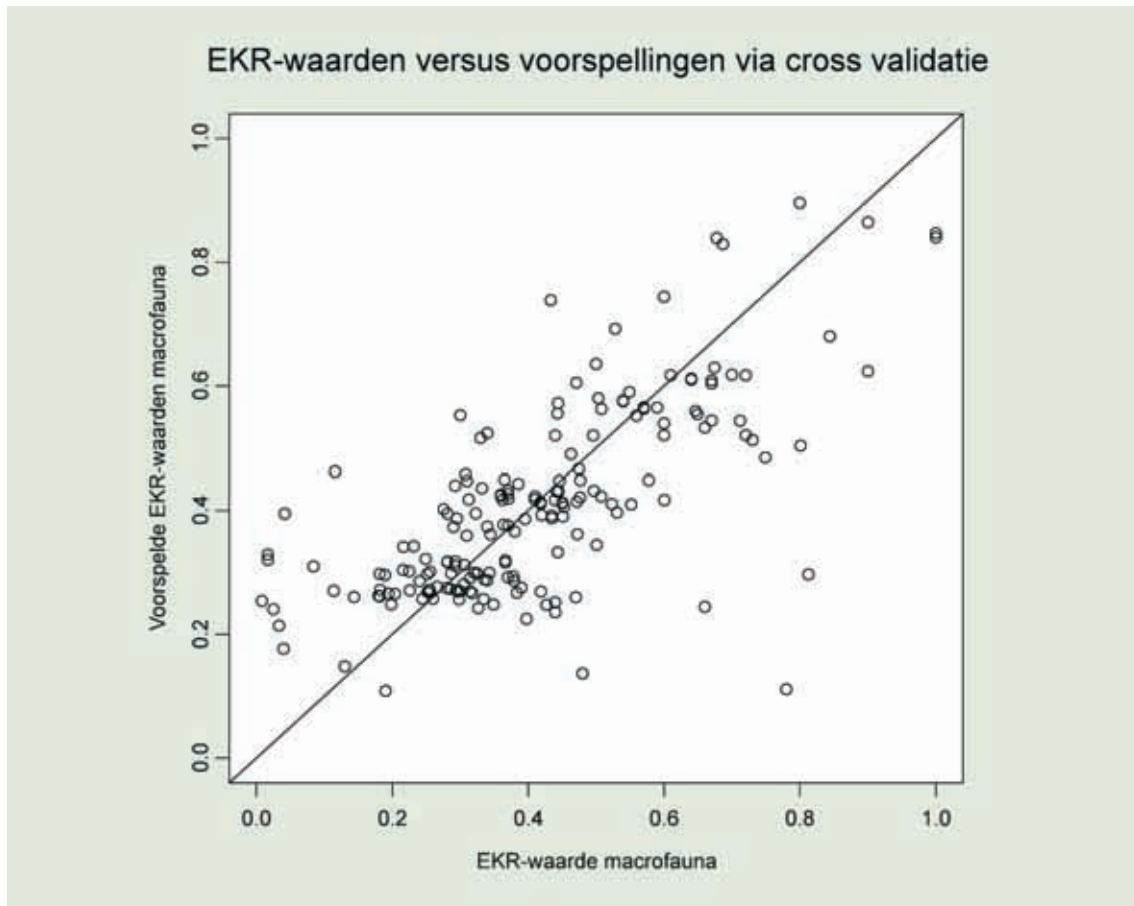
15) Totaal.P&gt;0.34625 61 0.60300 0.3699

30) Totaal.N&lt;4.9175 17 0.19950 0.4468 \*

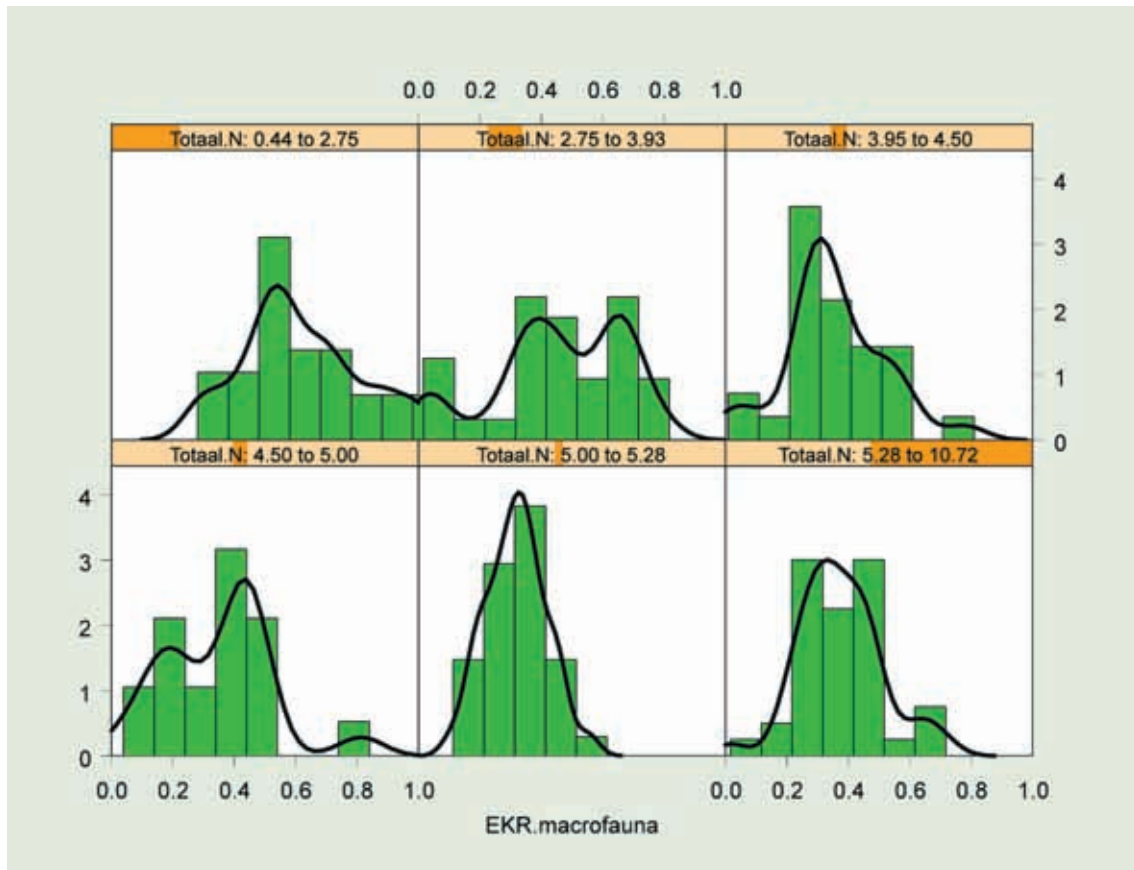
31) Totaal.N&gt;4.9175 44 0.26420 0.3402

62) Totaal.N&lt;5.29167 29 0.10200 0.3039 \*

63) Totaal.N&gt;5.29167 15 0.04999 0.4104 \*



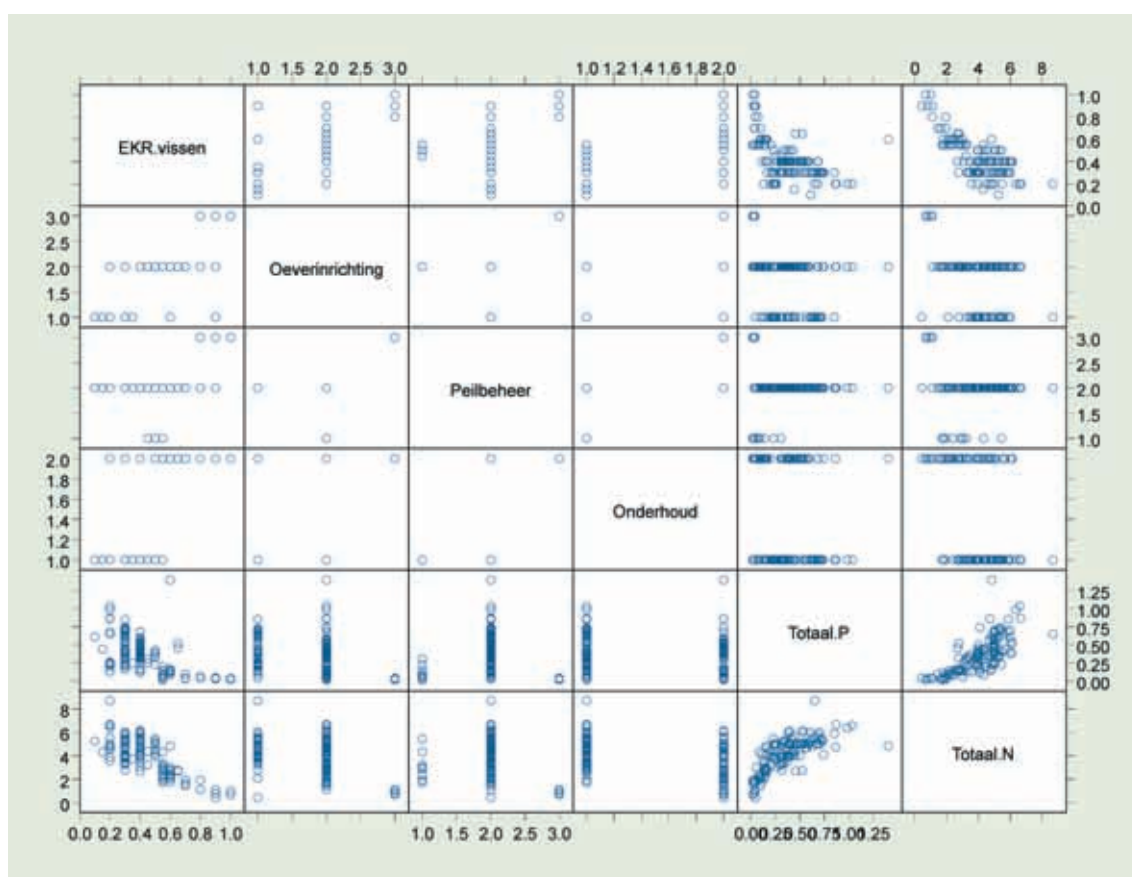
**Figuur C.3** EKR-voorspellingen met *randomForest* (500 bomen en een trainingsset van 80% van het totaal aantal kanalen). Kengetallen: MAD=0.097, RMSE=0.14 en  $R_{adj}^2=0.49$ . Zie voor een methodische uitleg paragraaf 3.2.2.



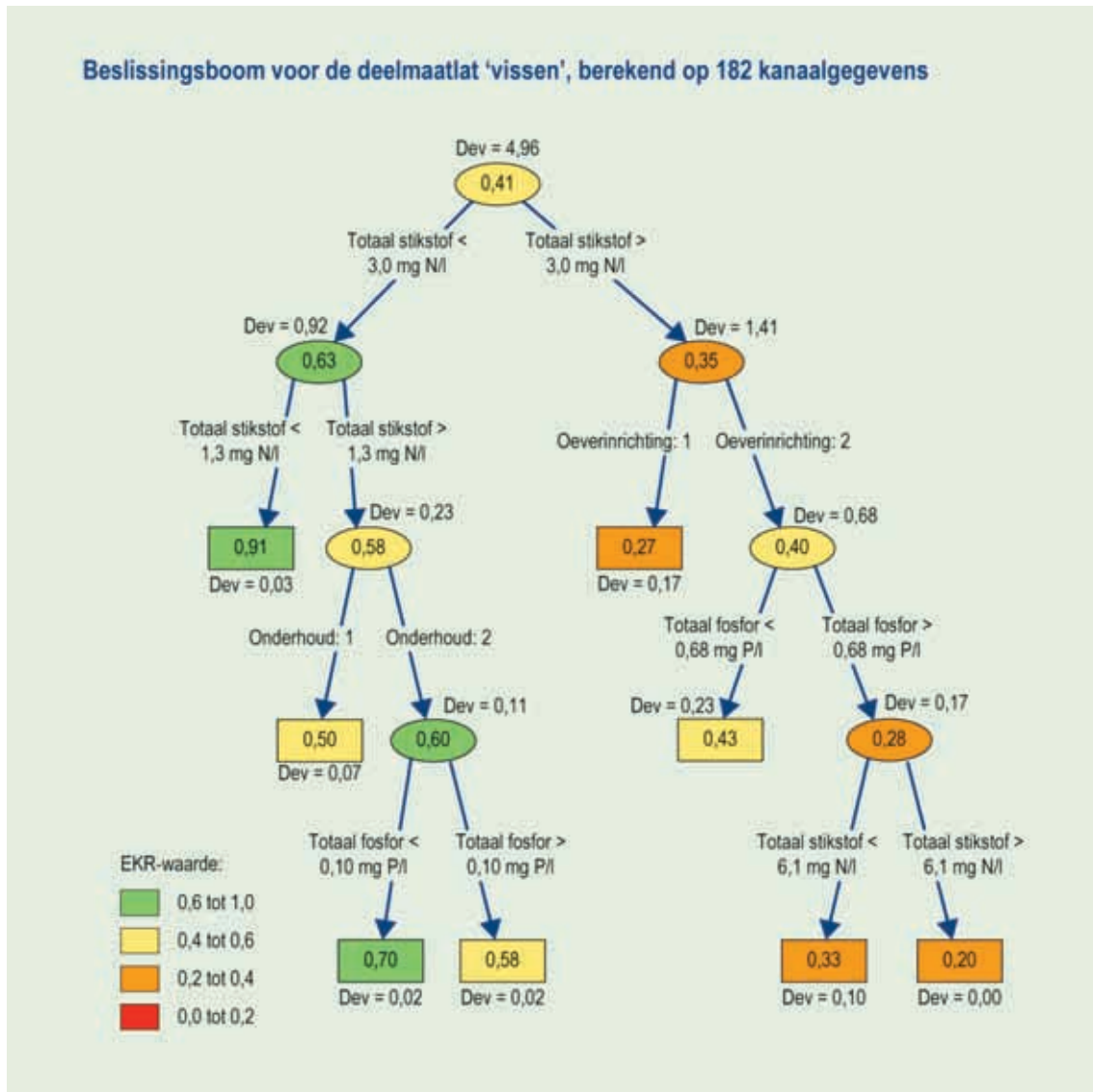
**Figuur C.4** Histogrammen en dichtheden voor de belangrijkste stuurvariabele 'Totaal.N'. Zie voor een methodische uitleg paragraaf 3.1.2.

## C.2 Deelmaatlat vissen

Figuur C.5 geeft de relatie tussen de deelmaatlat ‘vissen’ enerzijds en de beschikbare stuurvariabelen anderzijds in de vorm van een scatterplotmatrix. De figuur laat in de bovenste rij zien dat EKR-waarden stijgen bij toenemende waarden van oeverinrichting, peilbeheer en onderhoud, en dalen bij toenemende waarden van Totaal P en Totaal N. Figuur C.6 en Tabel C.2 geven de geschatte regressieboom. ‘Totaal N’ blijkt de belangrijkste stuurvariabele te zijn. Figuur C.7 geeft het resultaat voor een validatiecheck op 500 regressiebomen. Figuur C.8, ten slotte, geeft histogrammen met dichtheden voor de verschillende waarden van de belangrijkste stuurvariabele, namelijk ‘Totaal N’.



Figuur C.5 Scatterplotmatrix voor vissen (afhankelijke variabele) en vijf stuurvariabelen. Zie voor een methodische uitleg van de figuur paragraaf 3.1.1.



**Figuur C.6** Regressieboom met 8 eindknoten (rechthoeken). De gemiddelde EKR-waarde van een knoop staat in de rechthoek of ellips. De bijbehorende deviantie staat steeds onder de knoop. Het belang van de splitsingen is het grootst voor de eerste splitsing en neemt af naarmate 'gezakt' wordt in de boom. Knoopwaarden tussen 0.6 en 1.0 vallen in de klasse 'goed ecologisch potentieel' (GEP). Zie voor een methodische uitleg paragraaf 3.2. De kleurcode correspondeert met Figuur 1.3. Door regressieboom verklaarde variantie bedraagt 87%.

**Tabel C.2** Kwantitatieve informatie bij geschatte regressieboom.

Number of terminal nodes: 8

Residual mean deviance: 0.003669 = 0.6384 / 174

Distribution of residuals:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-2.000e-001	-2.681e-002	-1.429e-002	8.083e-017	3.148e-002	2.727e-001

node), split, n, deviance, yval

\* denotes terminal node

1) root 182 4.957e+000 0.4140

2) Totaal.N&lt;3.0375 42 9.233e-001 0.6333

4) Totaal.N&lt;1.27167 7 2.857e-002 0.9143 \*

5) Totaal.N&gt;1.27167 35 2.317e-001 0.5771

10) Onderhoud:1 8 6.500e-002 0.5000 \*

11) Onderhoud:2 27 1.050e-001 0.6000

22) Totaal.P&lt;0.0969048 5 2.000e-002 0.7000 \*

23) Totaal.P&gt;0.0969048 22 2.364e-002 0.5773 \*

3) Totaal.N&gt;3.0375 140 1.407e+000 0.3482

6) Oeverinrichting:1 54 1.715e-001 0.2685 \*

7) Oeverinrichting:2 86 6.772e-001 0.3983

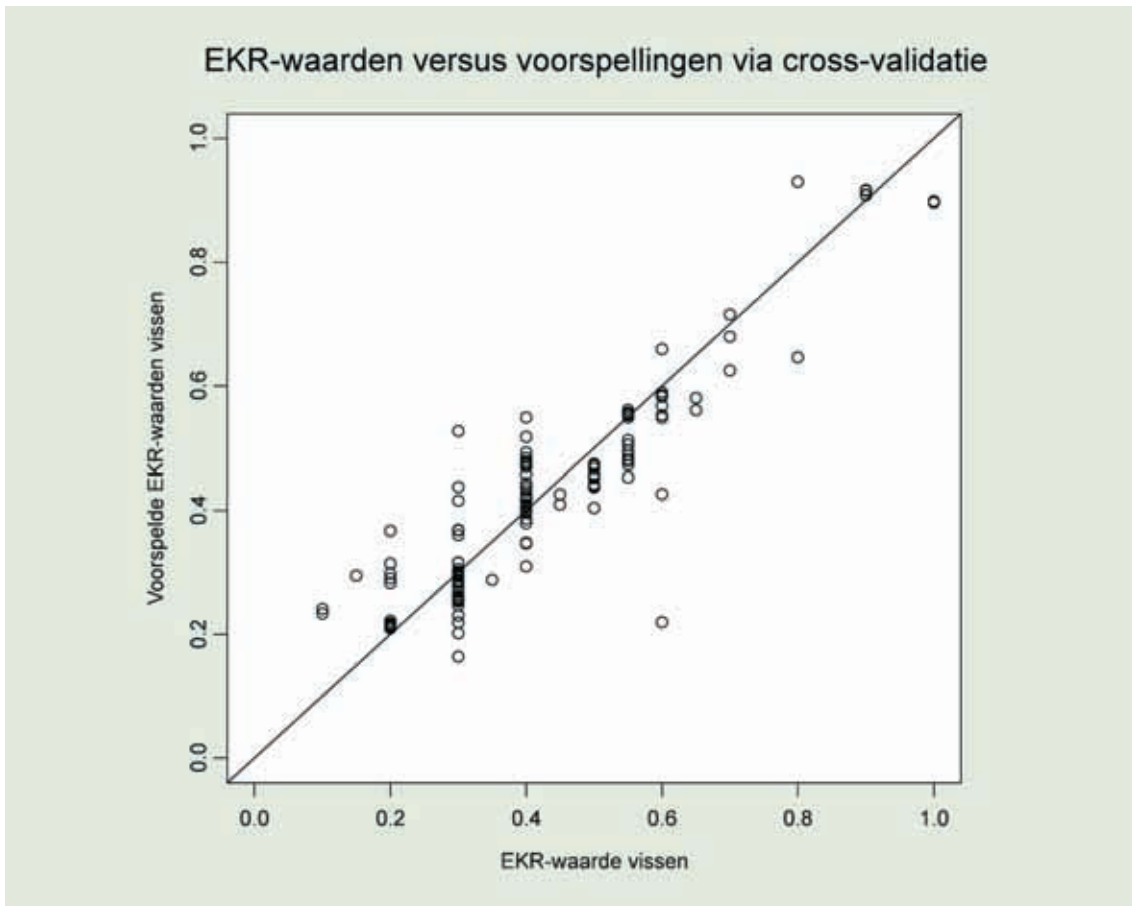
14) Totaal.P&lt;0.6825 69 2.279e-001 0.4268 \*

15) Totaal.P&gt;0.6825 17 1.647e-001 0.2824

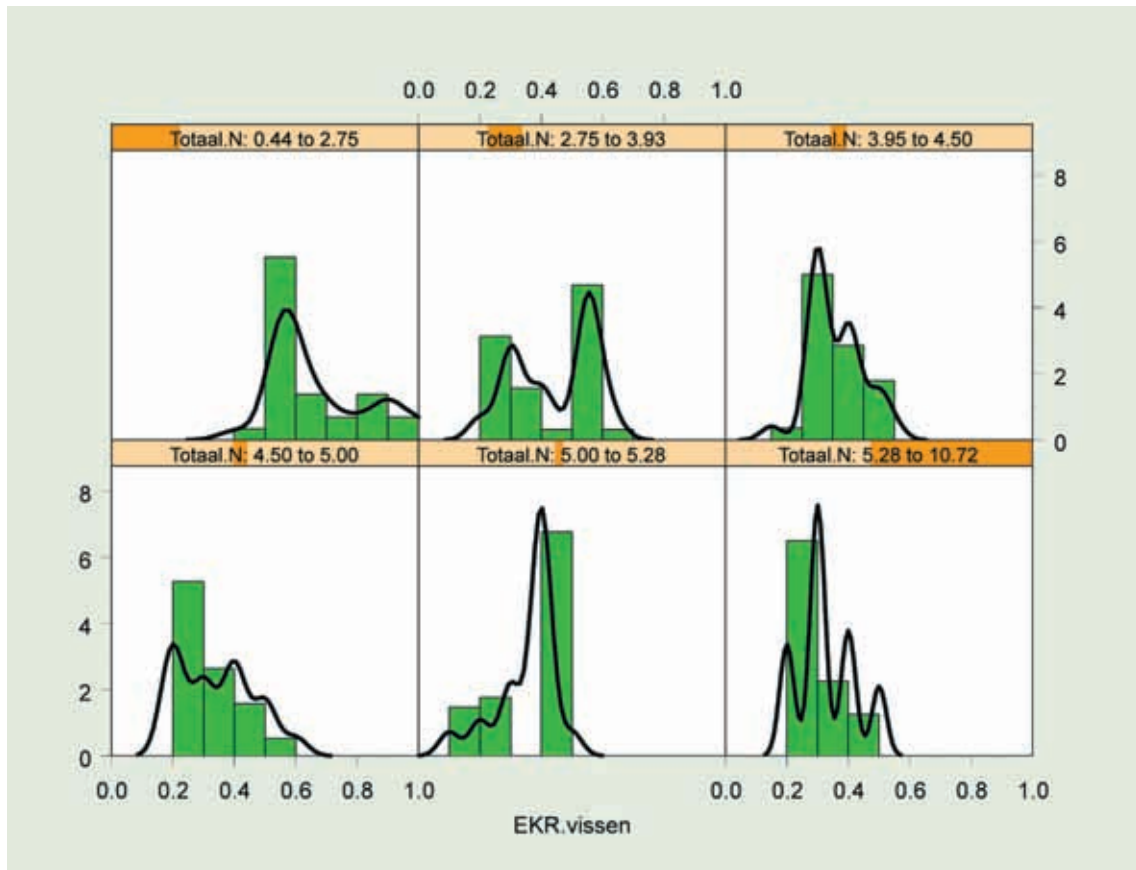
30) Totaal.N&lt;6.1 11 1.018e-001 0.3273 \*

31) Totaal.N&gt;6.1 6 5.778e-034 0.2000 \*





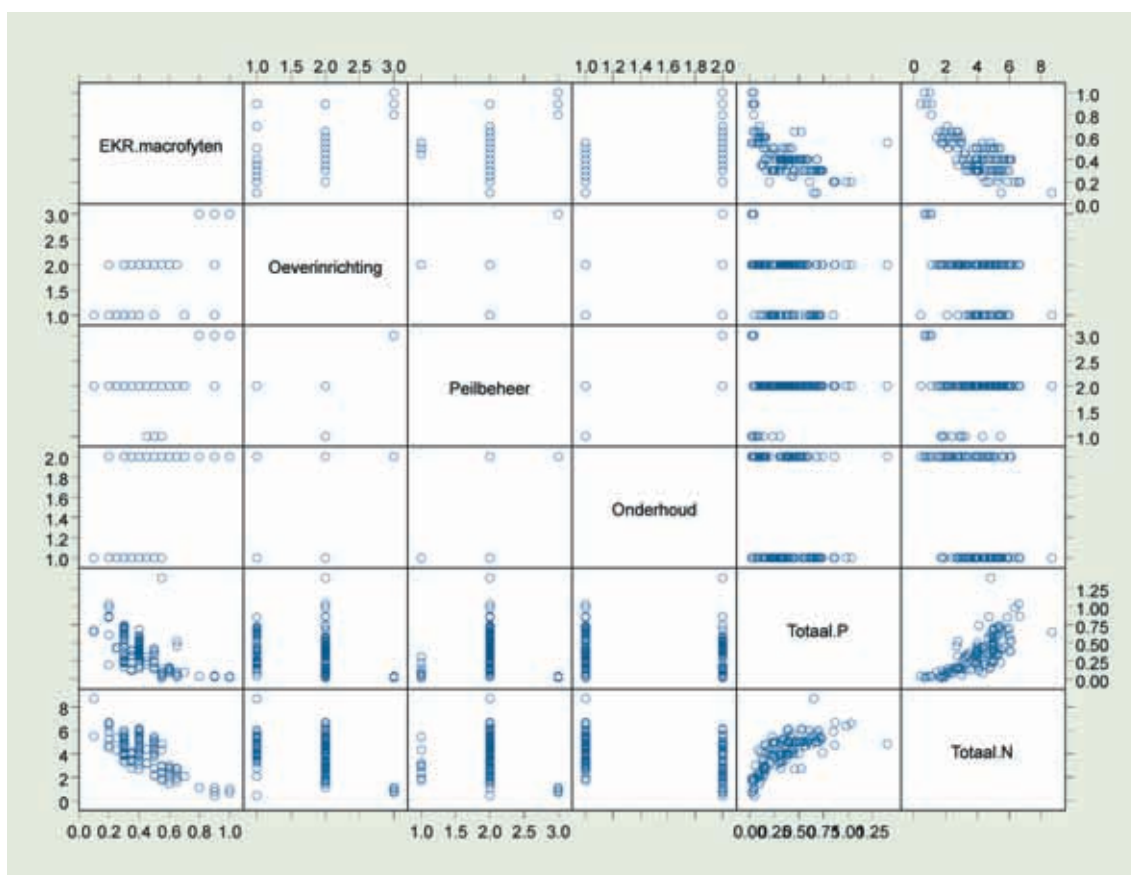
Figuur C.7 EKR-voorspellingen met *randomForest* (500 bomen en een trainingsset van 80% van het totaal aantal kanalen). Kengetallen: MAD=0.040, RMSE=0.064 en  $R_{adj}^2=0.85$ . Zie voor een methodische uitleg paragraaf 3.2.2.



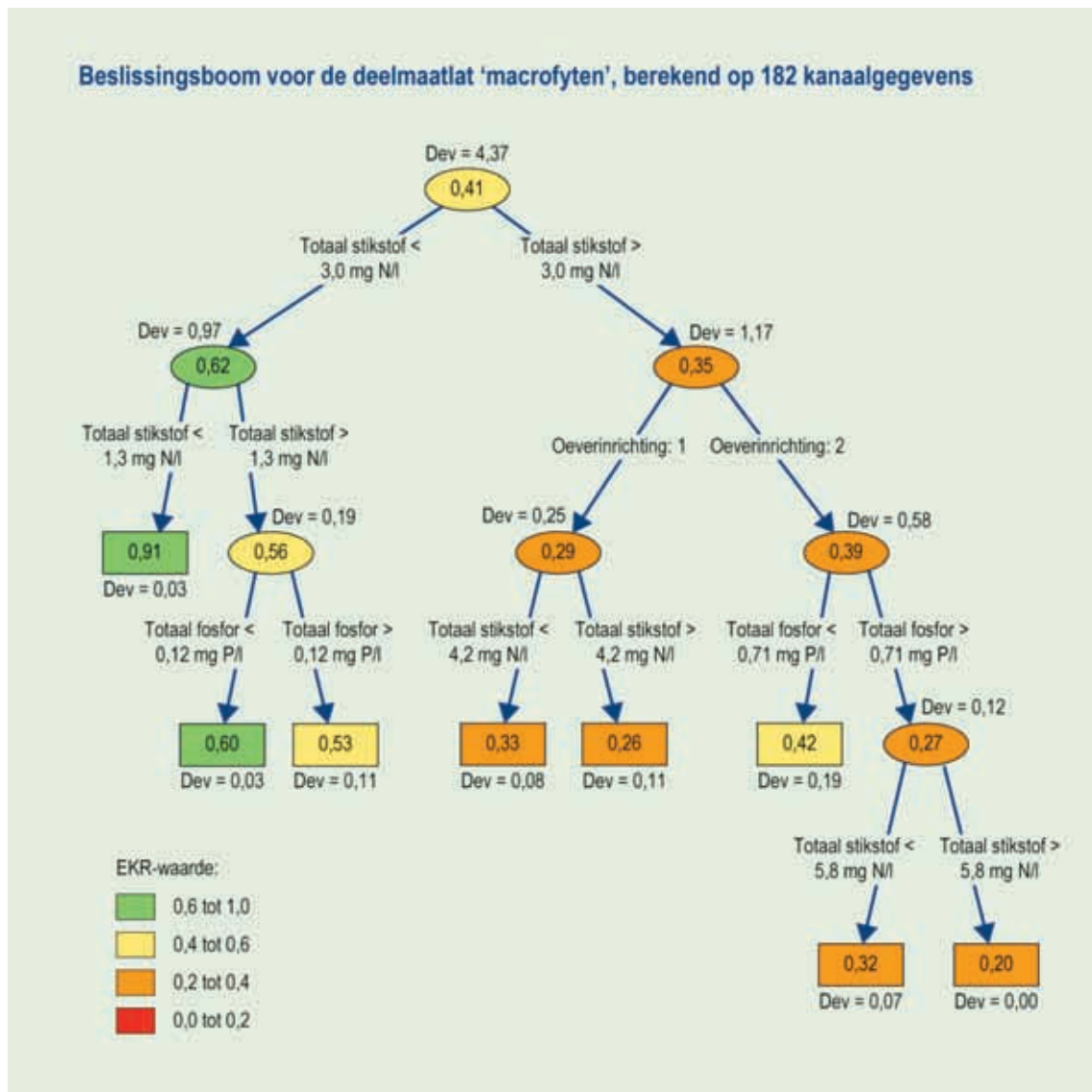
**Figuur C.8** Histogrammen en dichtheden voor de belangrijkste stuurvariabele 'Totaal.N'. Zie voor een methodische uitleg paragraaf 3.1.2.

### C.3 Deelmaatlat macrofyten

Figuur C.9 geeft de relatie tussen de deelmaatlat ‘macrofauna’ enerzijds en de beschikbare stuurvariabelen anderzijds in de vorm van een scatterplotmatrix. De figuur laat in de bovenste rij zien dat EKR-waarden stijgen bij toenemende waarden van oeverinrichting, peilbeheer en onderhoud, en dalen bij toenemende waarden van Totaal P en Totaal N. Figuur C.10 en Tabel C.3 geven de geschatte regressieboom. ‘Totaal N’ blijkt de belangrijkste stuurvariabele te zijn. Figuur C.11 geeft het resultaat voor een validatiecheck op 500 regressiebomen. Figuur C.12, ten slotte, geeft histogrammen met dichtheden voor de verschillende waarden van de belangrijkste stuurvariabele, namelijk ‘Totaal N’.



Figuur C.9 Scatterplotmatrix voor ‘macrofyten’ (afhankelijke variabele) en vijf stuurvariabelen. Zie voor een methodische uitleg van de figuur paragraaf 3.1.1.



**Figuur C.10** Regressieboom met 8 eindknoten (de rechthoeken). De gemiddelde EKR-waarde van een knoop staat in de rechthoek of ellips. De bijbehorende deviantie staat steeds onder de knoop. Het belang van de splitsingen is het grootst voor de eerste splitsing en neemt af naarmate 'gezakt' wordt in de boom. Knoopwaarden tussen 0,6 en 1,0 vallen in de klasse 'goed ecologisch potentieel' (GEP). Zie voor een methodische uitleg paragraaf 3.2. De kleurcode correspondeert met Figuur 1.3. Door regressieboom verklaarde variantie bedraagt 86%.

**Tabel C.3 Kwantitatieve informatie bij geschatte regressiebom.**

Number of terminal nodes: 8

Residual mean deviance: 0.003581 = 0.623 / 174

Distribution of residuals:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-1.773e-001	-2.727e-002	-1.761e-002	1.284e-016	3.636e-002	2.333e-001

node), split, n, deviance, yval

\* denotes terminal node

1) root 182 4.370e+000 0.4135

2) Totaal.N&lt;3.0375 42 9.724e-001 0.6155

4) Totaal.N&lt;1.27167 7 2.857e-002 0.9143 \*

5) Totaal.N&gt;1.27167 35 1.939e-001 0.5557

10) Totaal.P&lt;0.124167 13 3.231e-002 0.6038 \*

11) Totaal.P&gt;0.124167 22 1.136e-001 0.5273 \*

3) Totaal.N&gt;3.0375 140 1.169e+000 0.3529

6) Oeverinrichting:1 54 2.504e-001 0.2907

12) Totaal.N&lt;4.1925 21 7.667e-002 0.3333 \*

13) Totaal.N&gt;4.1925 33 1.114e-001 0.2636 \*

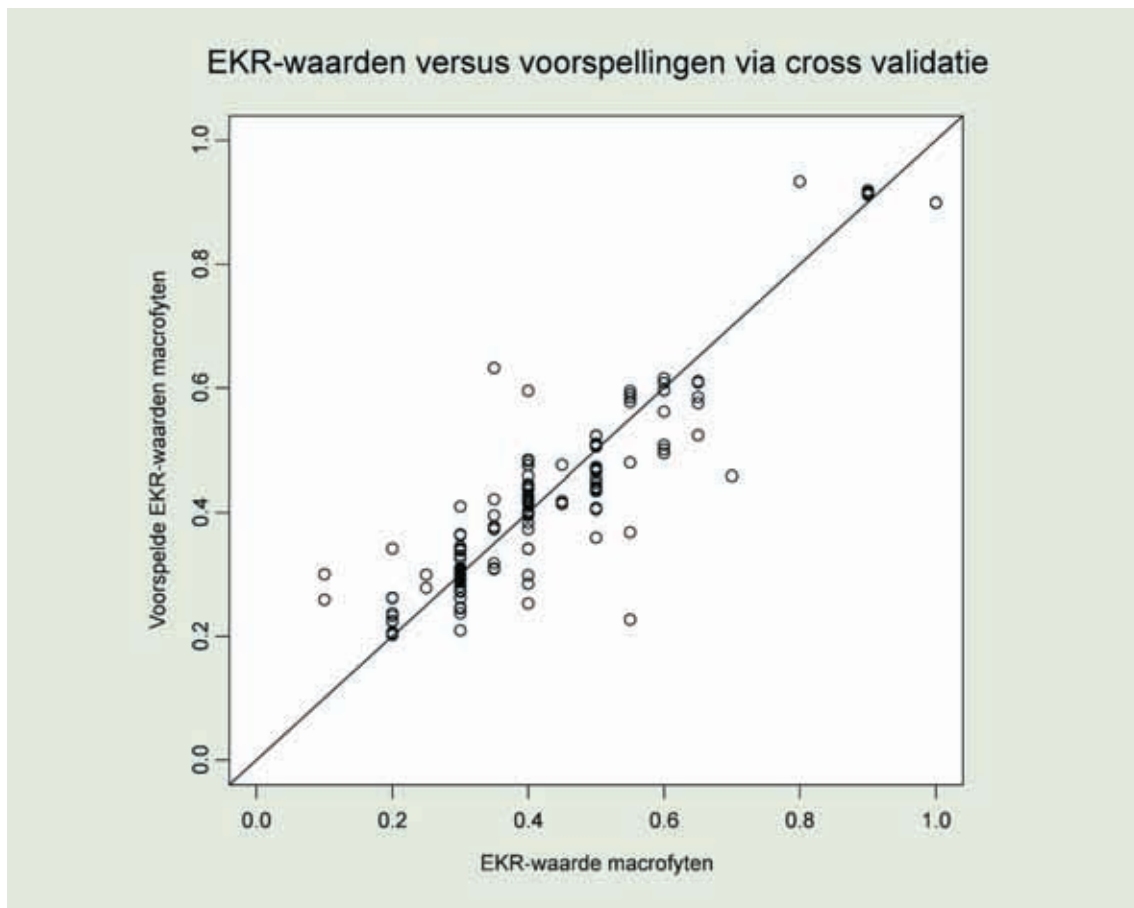
7) Oeverinrichting:2 86 5.793e-001 0.3919

14) Totaal.P&lt;0.709167 71 1.905e-001 0.4176 \*

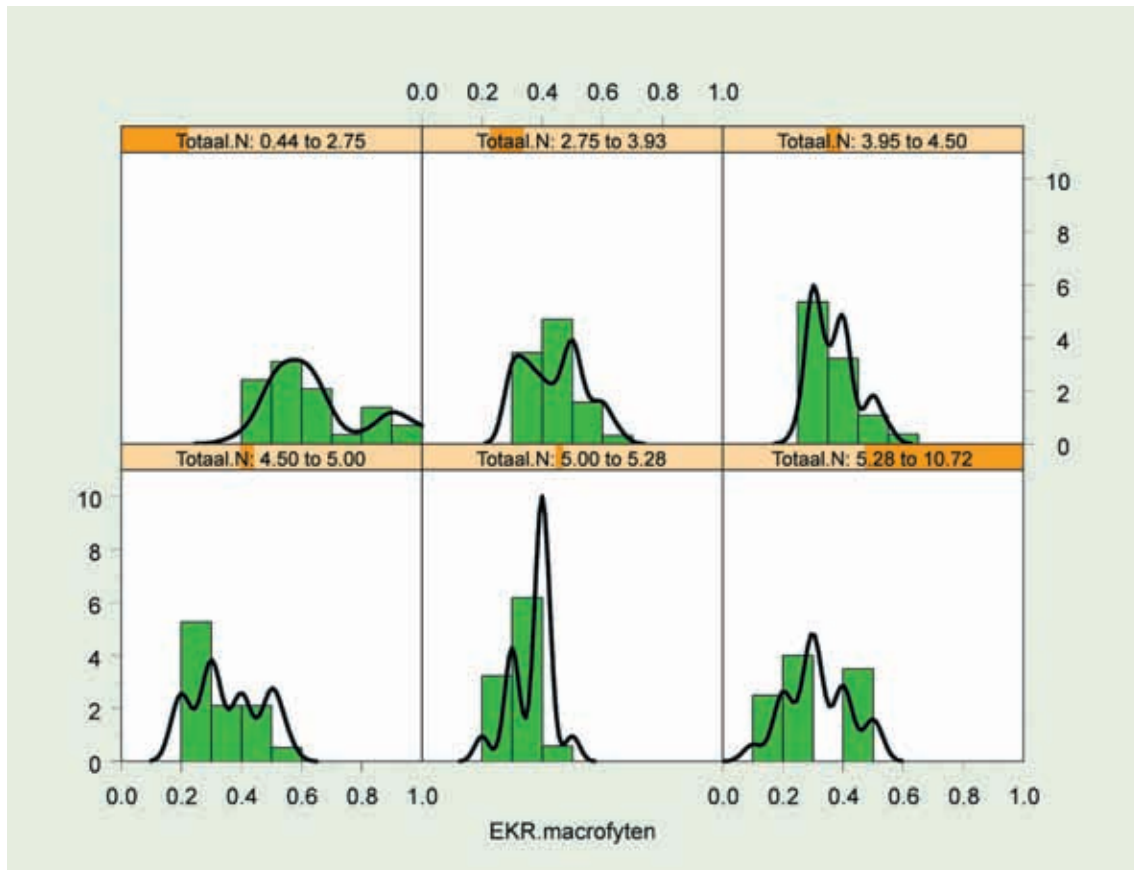
15) Totaal.P&gt;0.709167 15 1.190e-001 0.2700

30) Totaal.N&lt;5.825 9 7.000e-002 0.3167 \*

31) Totaal.N&gt;5.825 6 5.778e-034 0.2000



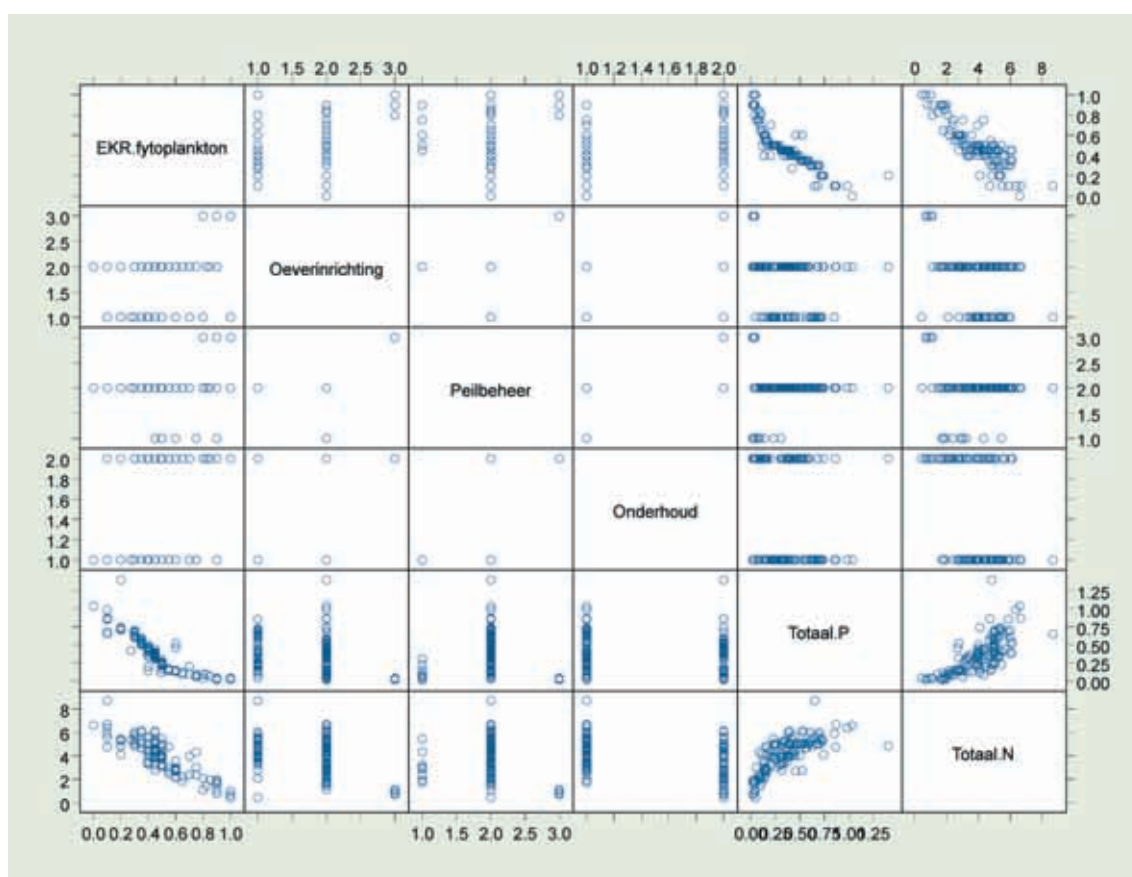
**Figuur C.11** EKR-voorspellingen met *randomForest* (500 bomen en een trainingsset van 80% van het totaal aantal kanalen). Kengetallen: MAD=0.042, RMSE=0.066 en  $R_{adj}^2=0.82$ . Zie voor een methodische uitleg paragraaf 3.2.2.



Figuur C.12 Histogrammen en dichtheden voor de belangrijkste stuurvariabele 'Totaal.N'. Zie voor een methodische uitleg paragraaf 3.1.2

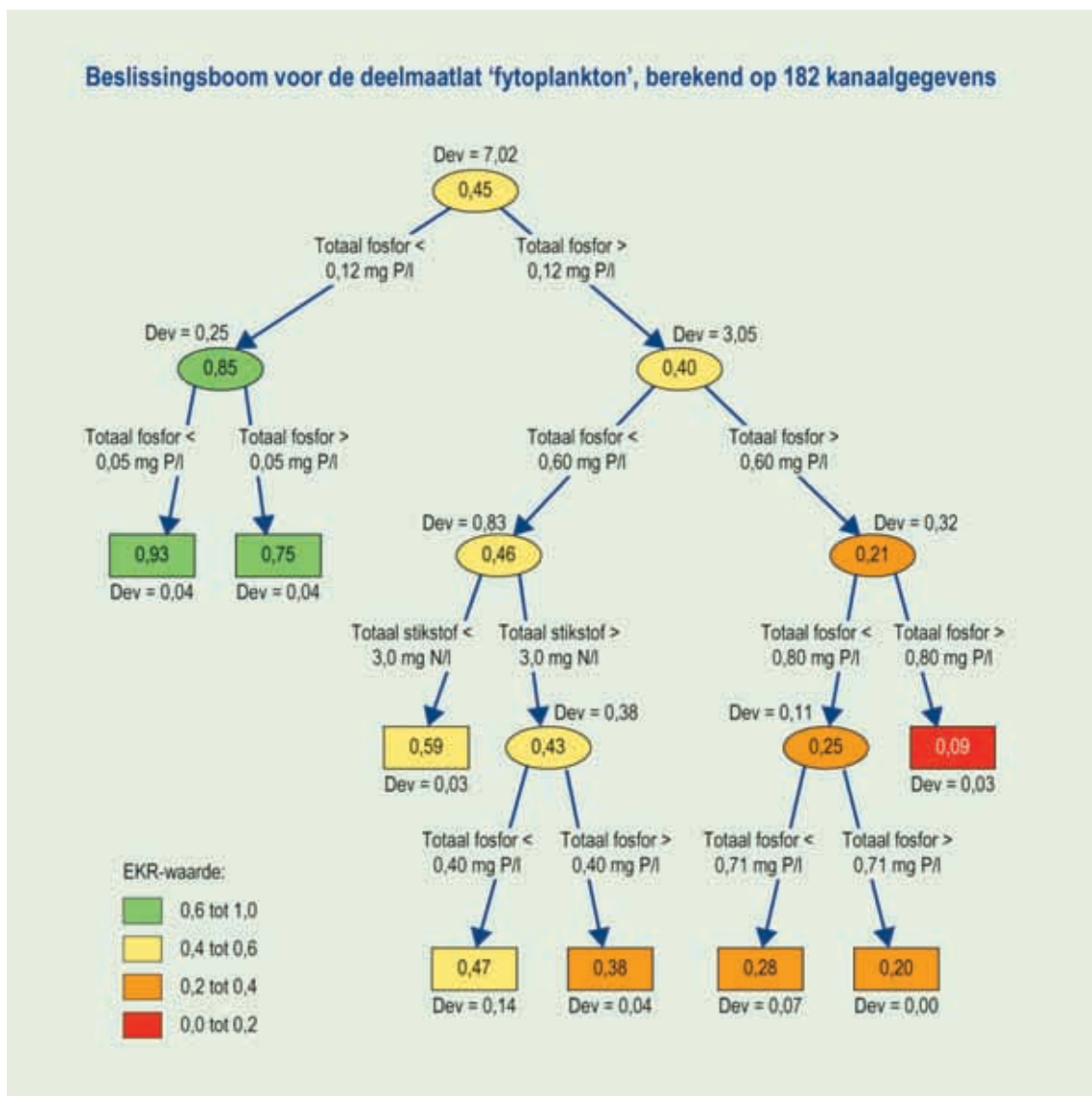
## C.4 Deelmaatlat fytoplankton

Figuur C.13 geeft de relatie tussen de deelmaatlat ‘fytoplankton’ enerzijds en de beschikbare stuurvariabelen anderzijds in de vorm van een scatterplotmatrix. De figuur laat in de bovenste rij zien dat EKR-waarden stijgen bij toenemende waarden van oeverinrichting, peilbeheer en onderhoud, en dalen bij toenemende waarden van Totaal P en Totaal N. Figuur C.14 en Tabel C.4 geven de geschatte regressieboom. ‘Totaal P’ blijkt de belangrijkste stuurvariabele te zijn. Figuur C.15 geeft het resultaat voor een validatiecheck op 500 regressiebomen. Figuur C.16, ten slotte, geeft histogrammen met dichtheden voor de verschillende waarden van de belangrijkste stuurvariabele, namelijk ‘Totaal P’.



Figuur C.13 Scatterplotmatrix voor ‘fytoplankton’ (afhankelijke variabele) en vijf stuurvariabelen. Zie voor een methodische uitleg van de figuur paragraaf 3.1.1.

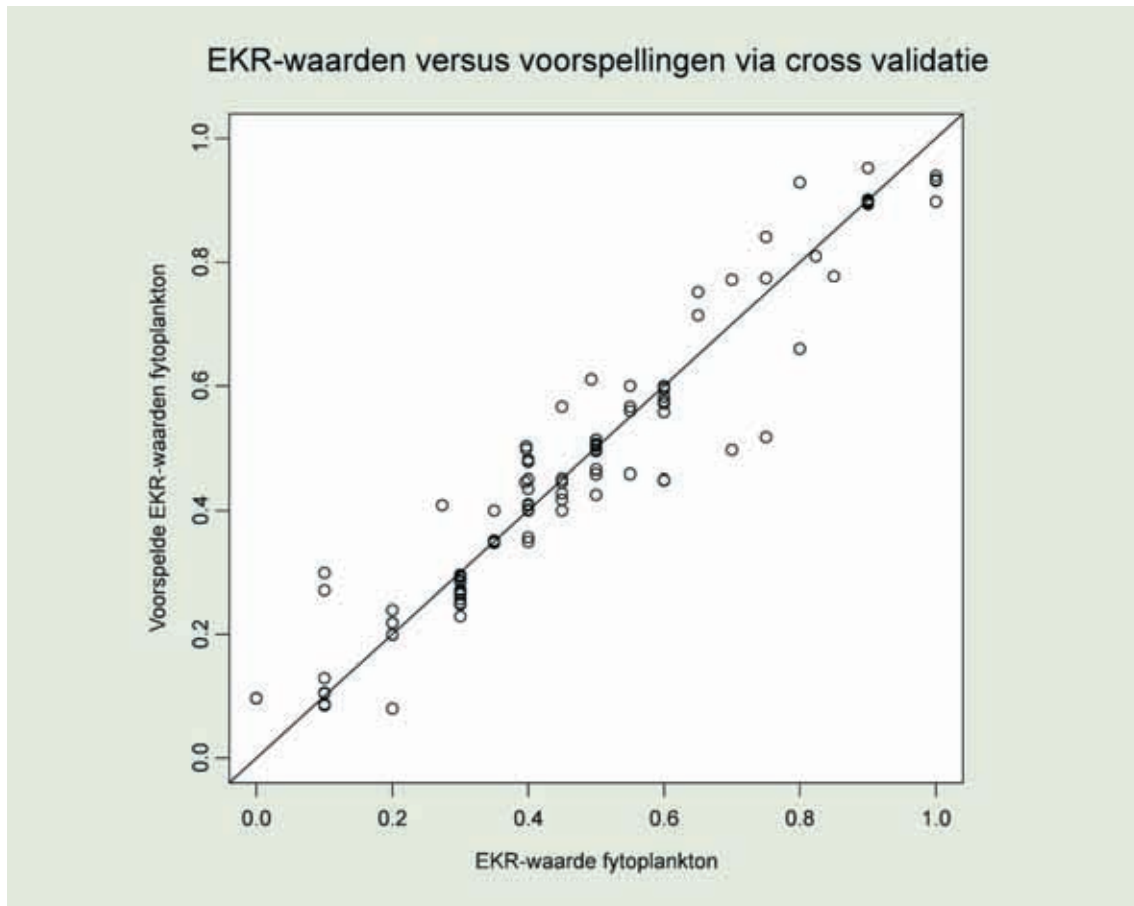




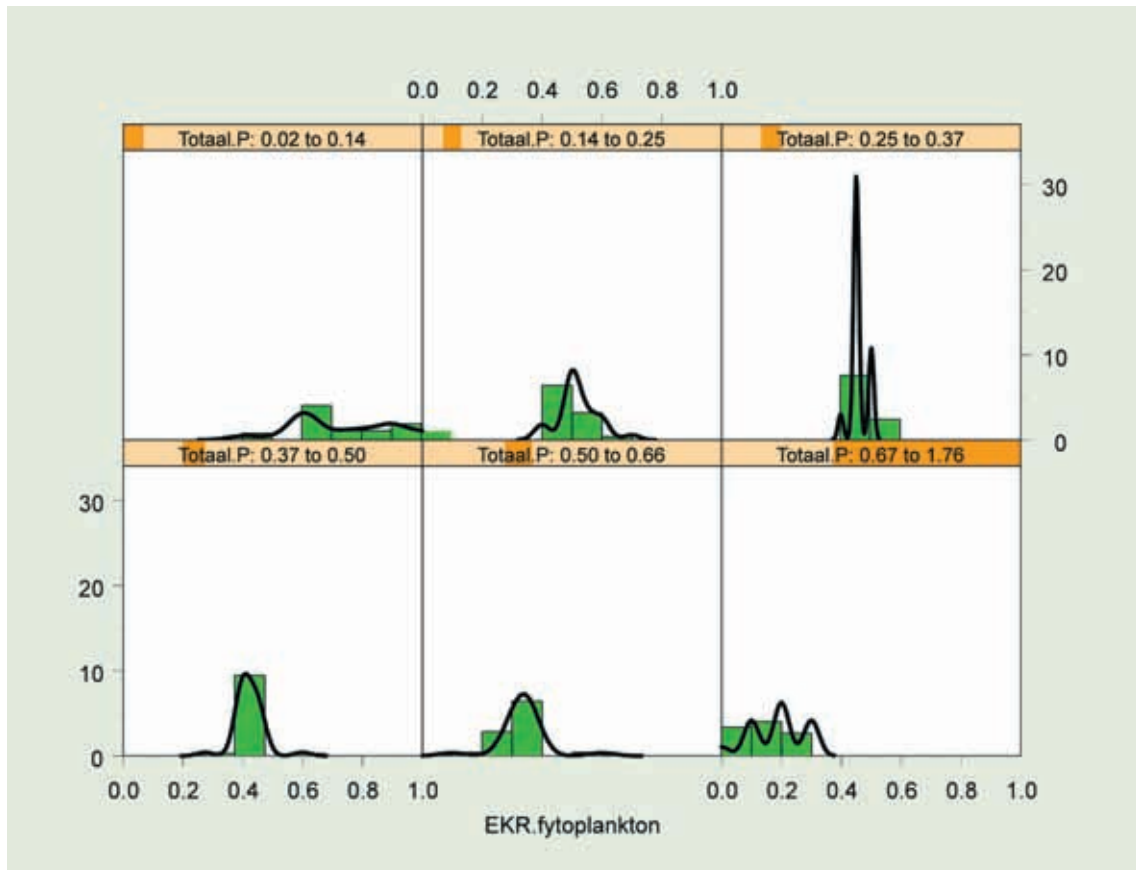
**Figuur C.14** Regressieboom met 8 eindknoten (de rechthoeken). De gemiddelde EKR-waarde van een knoop staat in de rechthoek of ellips. De bijbehorende deviantie staat steeds onder de knoop. Het belang van de splitsingen is het grootst voor de eerste splitsing en neemt af naarmate 'gezakt' wordt in de boom. Knoopwaarden tussen 0.6 en 1.0 vallen in de klasse 'goed ecologisch potentieel' (GEP). Zie voor een methodische uitleg paragraaf 3.2. De kleurcode correspondeert met Figuur 1.3. Door regressieboom verklaarde variantie bedraagt 94%.

**Tabel C.4 Kwantitatieve informatie bij geschatte regressieboom.**

Number of terminal nodes: 8					
Residual mean deviance: 0.00224 = 0.3898 / 174					
Distribution of residuals:					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-1.765e-001	-1.878e-002	1.000e-002	6.497e-017	2.353e-002	2.312e-001
node), split, n, deviance, yval					
* denotes terminal node					
1) root 182 7.020e+000 0.4526					
2) Totaal.P<0.115 21 2.457e-001 0.8487					
4) Totaal.P<0.0508333 12 4.250e-002 0.9250 *					
5) Totaal.P>0.0508333 9 4.030e-002 0.7470 *					
3) Totaal.P>0.115 161 3.050e+000 0.4010					
6) Totaal.P<0.596667 123 8.255e-001 0.4614					
12) Totaal.N<3.03 22 2.591e-002 0.5864 *					
13) Totaal.N>3.03 101 3.814e-001 0.4342					
26) Totaal.P<0.400833 63 1.432e-001 0.4688 *					
27) Totaal.P>0.400833 38 3.829e-002 0.3769 *					
7) Totaal.P>0.596667 38 3.189e-001 0.2053					
14) Totaal.P<0.795833 28 1.096e-001 0.2464					
28) Totaal.P<0.705 17 7.059e-002 0.2765 *					
29) Totaal.P>0.705 11 3.991e-033 0.2000 *					
15) Totaal.P>0.795833 10 2.900e-002 0.0900 *					



Figuur C.15 EKR-voorspellingen met *randomForest* (500 bomen en een trainingsset van 80% van het totaal aantal kanalen). Kengetallen: MAD=0.027, RMSE=0.052 en  $R_{adj}^2=0.93$ . Zie voor een methodische uitleg paragraaf 3.2.2.



Figuur C.16 Histogrammen en dichtheden voor de belangrijkste stuurvariabele 'Totaal.P'. Zie voor een methodische uitleg paragraaf 3.1.2.

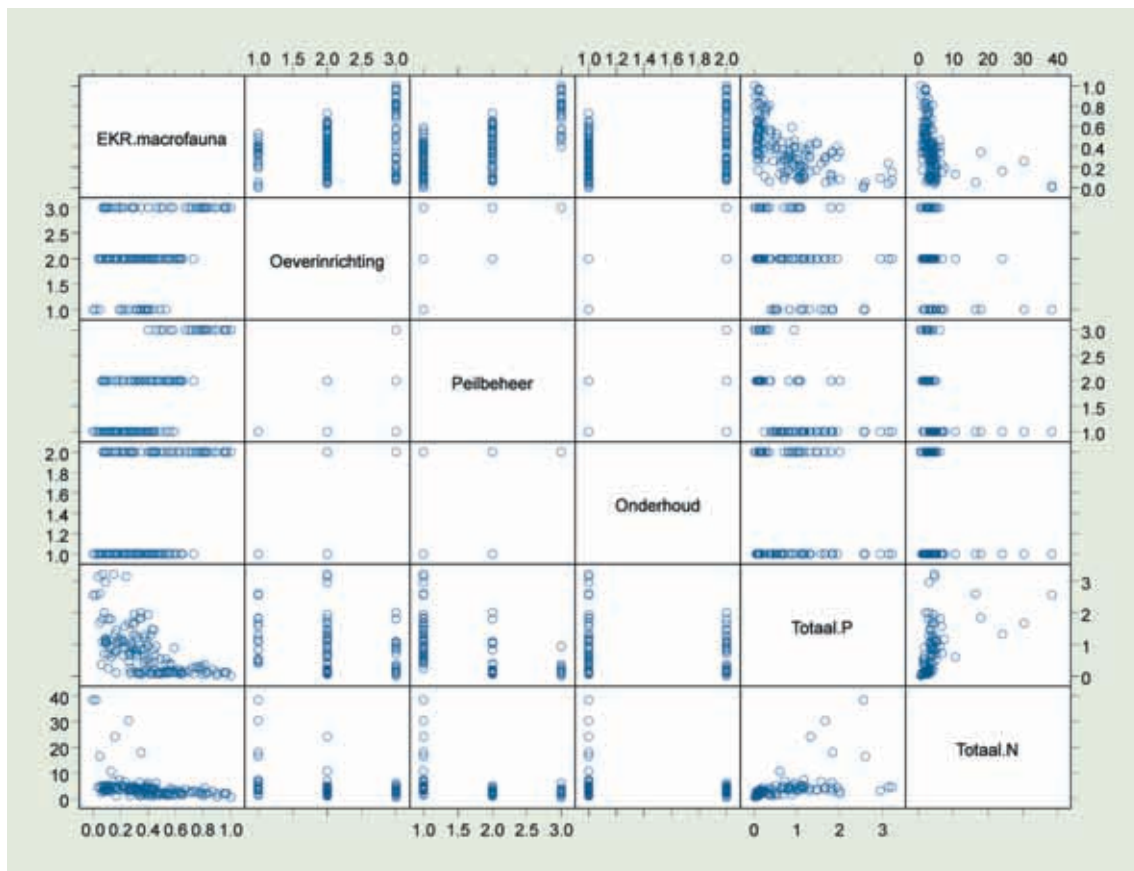
## Appendix D Regressiebomen voor sloten

In deze appendix worden de regressieboom-schattingen gegeven voor de deelmaatlaten 'macrofauna' (paragraaf D.1), 'vissen' (D.2) en 'macrofyten' (D.3) in sloten. Fysisch-chemische stuurvariabelen zijn voor al deze deelmaatlaten: totaal fosfor en totaal stikstof. Hydromorfologische variabelen zijn: oeverinrichting, peilbeheer en onderhoud. Zie voor een uitleg van deze zes variabelen Tabel 2.2, het vierde paneel.

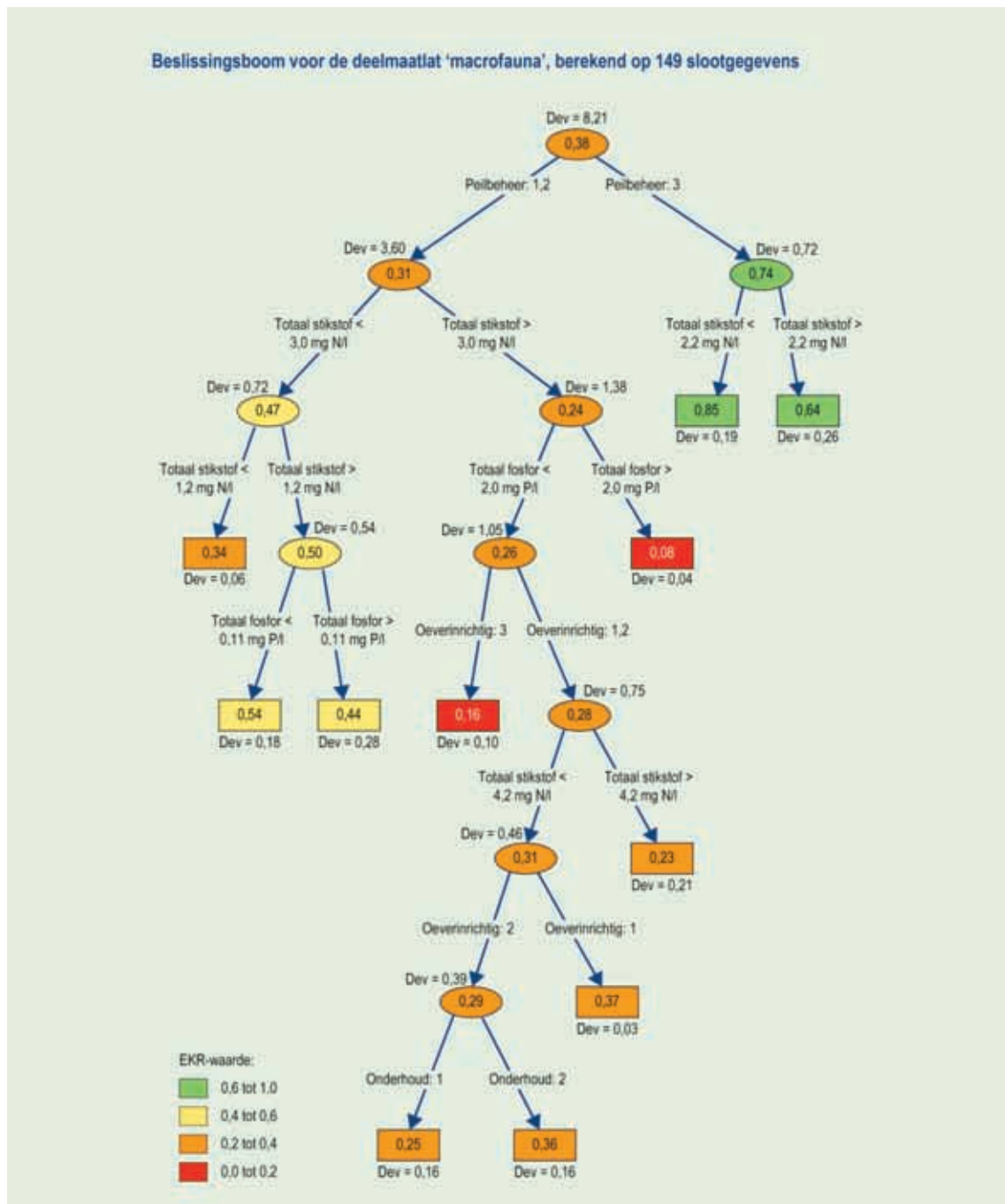
### D.1 Deelmaatlat macrofauna

Figuur D.1 geeft de relatie tussen de deelmaatlat 'macrofauna' enerzijds en de beschikbare stuurvariabelen anderzijds in de vorm van een scatterplotmatrix. De figuur laat in de bovenste rij zien dat EKR-waarden stijgen bij toenemende waarden van oeverinrichting, peilbeheer en onderhoud, en dalen bij toenemende waarden van Totaal P en Totaal N. Figuur D.2 en

Tabel D.1 geven de geschatte regressieboom. Peilbeheer blijkt de belangrijkste stuurvariabele te zijn. Figuur D.3 geeft het resultaat voor een validatiecheck op 500 regressiebomen. Figuur D.3, ten slotte, geeft histogrammen met dichtheden voor de verschillende waarden van de belangrijkste stuurvariabele, namelijk peilbeheer.



Figuur D.1 Scatterplotmatrix voor 'macrofauna' (afhankelijke variabele) en vijf stuurvariabelen. Zie voor een methodische uitleg van de figuur paragraaf 3.1.1.



Figuur D.2 Regressieboom met 11 eindknoten (rechthoeken). De gemiddelde EKR-waarde van een knoop staat in de rechthoek of ellips. De bijbehorende deviantie staat steeds onder de knoop. Het belang van de splitsingen is het grootst voor de eerste splitsing en neemt af naarmate ‘gezakt’ wordt in de boom. Knoopwaarden tussen 0.6 en 1.0 vallen in de klasse ‘goed ecologisch potentieel’ (GEP). Zie voor een methodische uitleg paragraaf 3.2. De kleurcode correspondeert met Figuur 1.3. Door regressieboom verklaarde variantie bedraagt 80%.

**Tabel D.1 Kwantitatieve informatie bij geschatte regressiebom.**

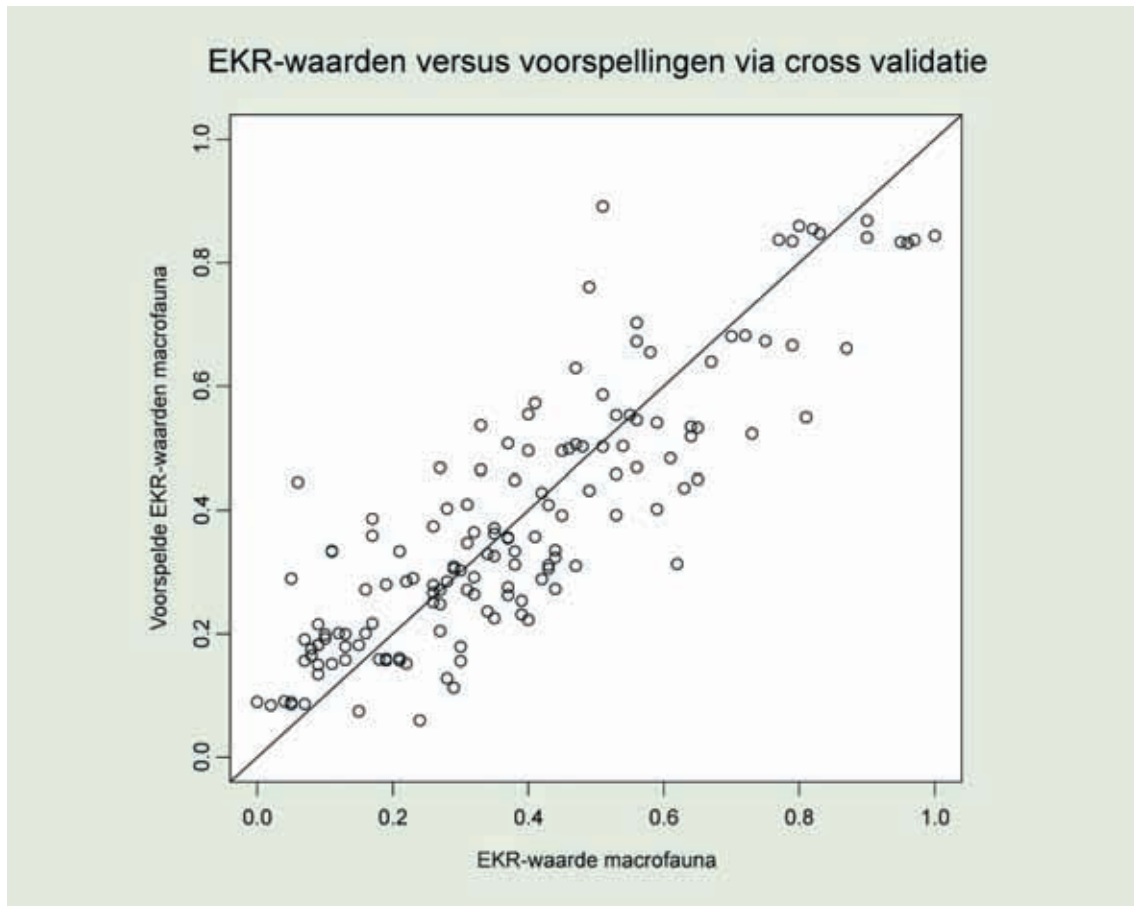
Number of terminal nodes: 11

Residual mean deviance: 0.01209 = 1.669 / 138

node), split, n, deviance, yval

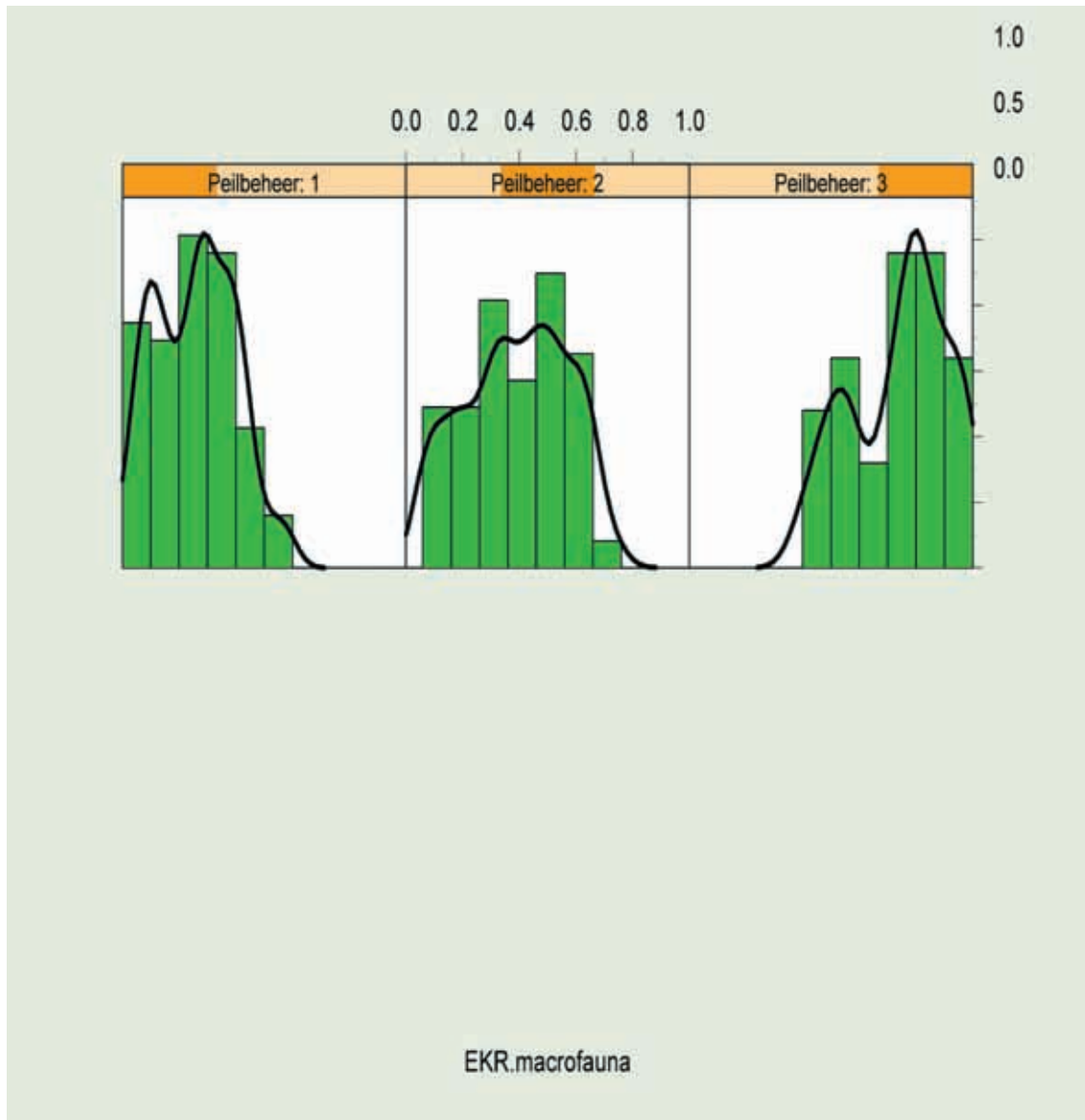
\* denotes terminal node

1) root 149 8.20600 0.3825  
 2) Peilbeheer:1,2 124 3.59800 0.3099  
 4) Totaal.N<2.9535 39 0.72170 0.4723  
 8) Totaal.N<1.21333 6 0.05888 0.3383 \*  
 9) Totaal.N>1.21333 33 0.53550 0.4967  
 18) Totaal.P<0.111667 18 0.17880 0.5417 \*  
 19) Totaal.P>0.111667 15 0.27650 0.4427 \*  
 5) Totaal.N>2.9535 85 1.37600 0.2354  
 10) Totaal.P<1.97 75 1.05500 0.2563  
 20) Oeverinrichting:3 17 0.09920 0.1600 \*  
 21) Oeverinrichting:1,2 58 0.75160 0.2845  
 42) Totaal.N<4.2125 38 0.46340 0.3111  
 84) Oeverinrichting:2 28 0.39400 0.2918  
 168) Onderhoud:1 18 0.15820 0.2539 \*  
 169) Onderhoud:2 10 0.16340 0.3600 \*  
 85) Oeverinrichting:1 10 0.02985 0.3650 \*  
 43) Totaal.N>4.2125 20 0.21050 0.2340 \*  
 11) Totaal.P>1.97 10 0.04409 0.0790 \*  
 3) Peilbeheer:3 25 0.71670 0.7424  
 6) Totaal.N<2.15 12 0.19340 0.8500 \*  
 7) Totaal.N>2.15 13 0.25610 0.6431 \*



**Figuur D.3** EKR-voorspellingen met *randomForest* (500 bomen en een trainingsset van 80% van het totaal aantal sloten). Kengetallen: MAD=0.093, RMSE=0.12 en  $R_{adj}^2=0.75$ . Zie voor een methodische uitleg paragraaf 3.2.2.



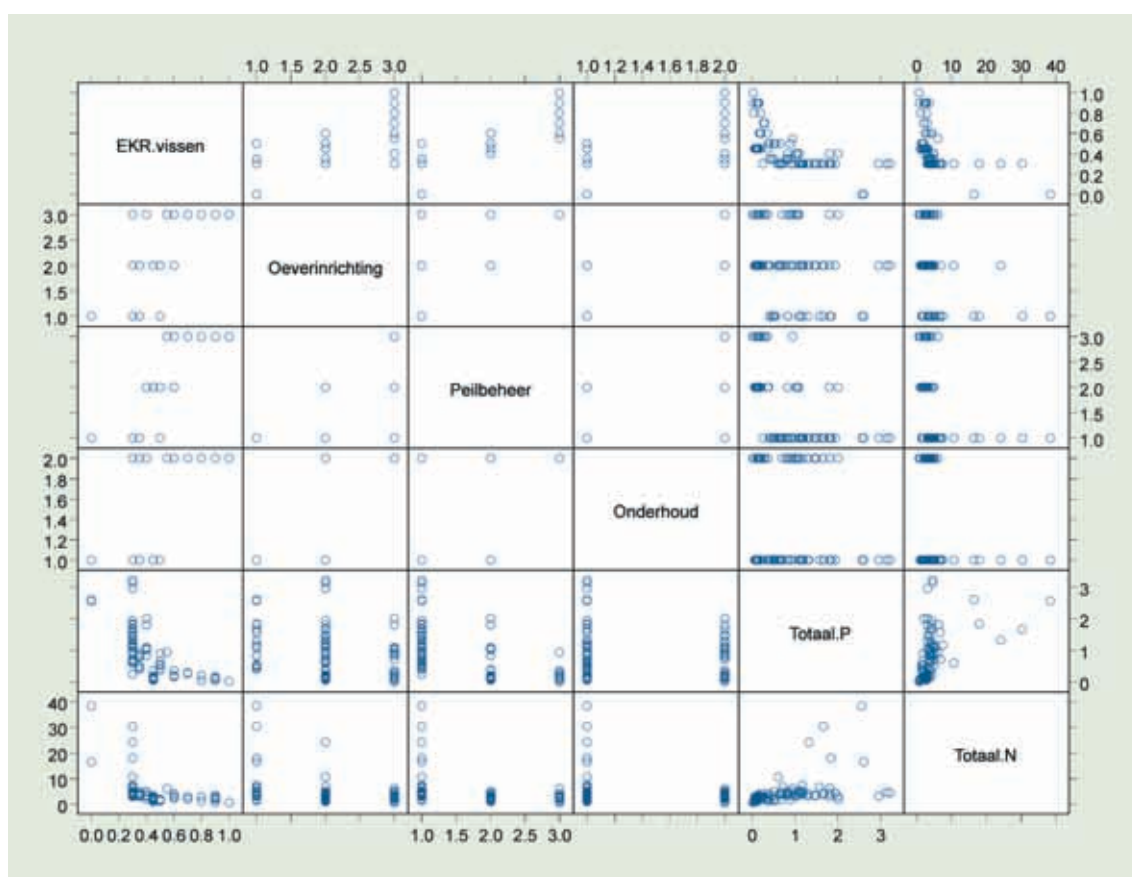


**Figuur D.4** Histogrammen en dichtheden voor de belangrijkste stuurvariabele 'peilbeheer' (drie discrete klassen). Zie voor een methodische uitleg paragraaf 3.1.2.

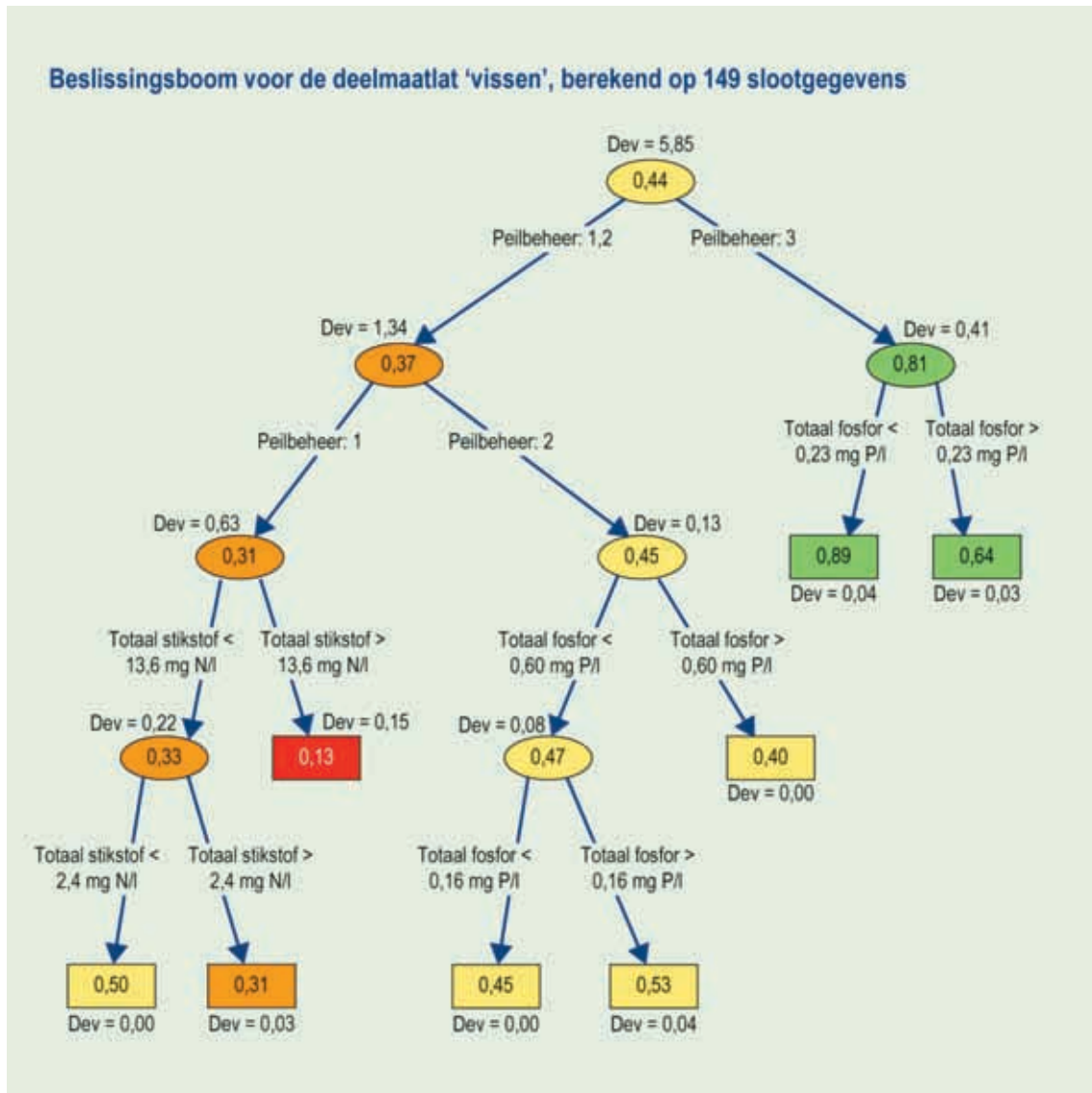
## D.2 Deelmaatlat vissen

Figuur D.5 geeft de relatie tussen de deelmaatlat ‘vissen’ enerzijds en de beschikbare stuurvariabelen anderzijds in de vorm van een scatterplotmatrix. De figuur laat in de bovenste rij zien dat EKR-waarden stijgen bij toenemende waarden van oeverinrichting, peilbeheer en onderhoud, en dalen bij toenemende waarden van Totaal P en Totaal N. Figuur D.6 en

Tabel D.2 geven de geschatte regressieboom. Peilbeheer blijkt de belangrijkste stuurvariabele te zijn. Figuur D.7 geeft het resultaat voor een validatiecheck op 500 regressiebomen. Figuur D.8, ten slotte, geeft histogrammen met dichtheden voor de verschillende waarden van de belangrijkste stuurvariabele, namelijk peilbeheer.



Figuur D.5 Scatterplotmatrix voor ‘vissen’ (afhankelijke variabele) en vijf stuurvariabelen. Zie voor een methodische uitleg van de figuur paragraaf 3.1.1.



**Figuur D.6** Regressieboom met 8 eindknoten (rechthoeken). De gemiddelde EKR-waarde van een knoop staat in de rechthoek of ellips. De bijbehorende deviantie staat steeds onder de knoop. Het belang van de splitsingen is het grootst voor de eerste splitsing en neemt af naarmate 'gezakt' wordt in de boom. Knoopwaarden tussen 0.6 en 1.0 vallen in de klasse 'goed ecologisch potentieel' (GEP). Zie voor een methodische uitleg paragraaf 3.2. De kleurcode correspondeert met Figuur 1.3. Door regressieboom verklaarde variantie bedraagt 95%.

**Tabel D.2 Kwantitatieve informatie bij geschatte regressieboom.**

Number of terminal nodes: 8

Residual mean deviance: 0.002064 = 0.2911 / 141

Distribution of residuals:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-1.286e-001	-1.048e-002	0.000e+000	-1.673e-016	1.176e-002	1.714e-001

node), split, n, deviance, yval

\* denotes terminal node

1) root 149 5.849e+000 0.4389

2) Peilbeheer:1,2 124 1.344e+000 0.3645

4) Peilbeheer:1 75 6.269e-001 0.3087

8) Totaal.N&lt;13.5583 68 2.222e-001 0.3272

16) Totaal.N&lt;2.43875 6 0.000e+000 0.5000 \*

17) Totaal.N&gt;2.43875 62 2.569e-002 0.3105 \*

9) Totaal.N&gt;13.5583 7 1.543e-001 0.1286 \*

5) Peilbeheer:2 49 1.250e-001 0.4500

10) Totaal.P&lt;0.601667 36 8.076e-002 0.4681

20) Totaal.P&lt;0.156944 28 2.248e-031 0.4500 \*

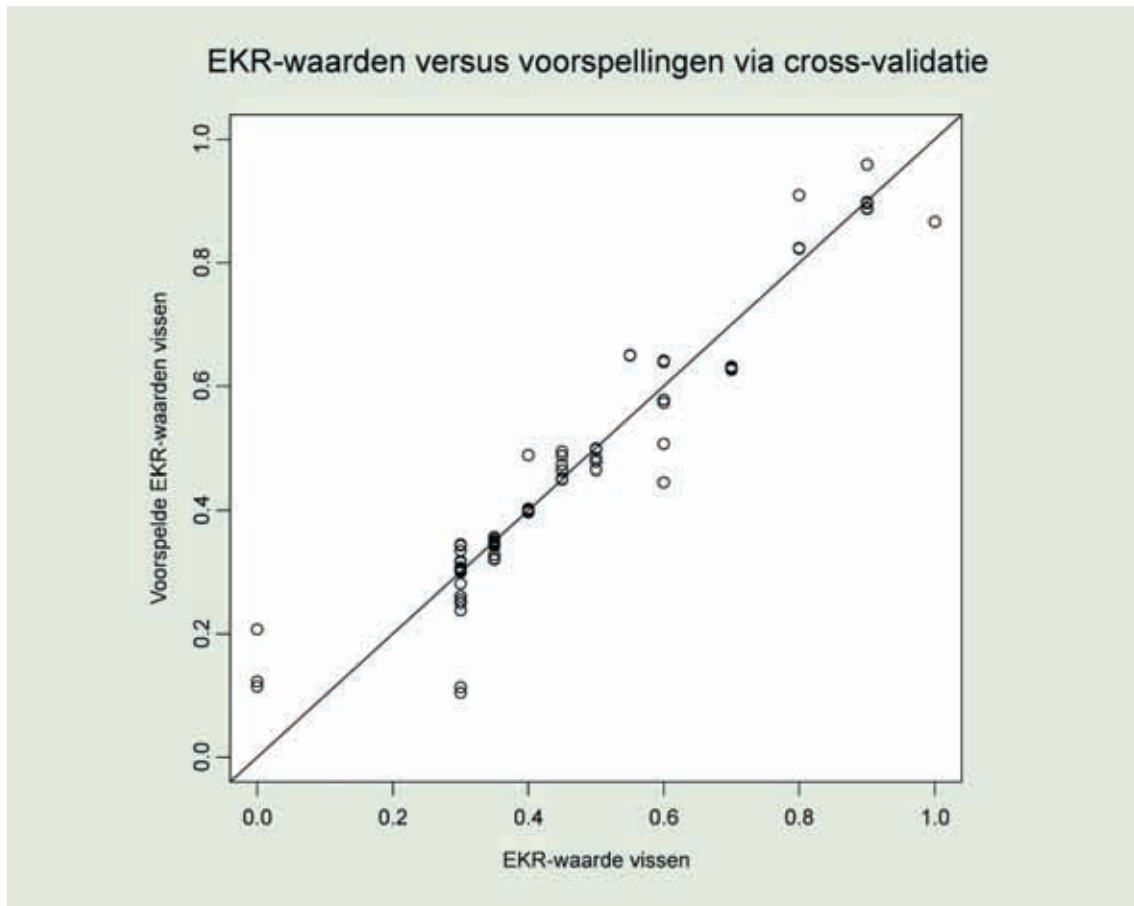
21) Totaal.P&gt;0.156944 8 3.969e-002 0.5313 \*

11) Totaal.P&gt;0.601667 13 2.163e-032 0.4000 \*

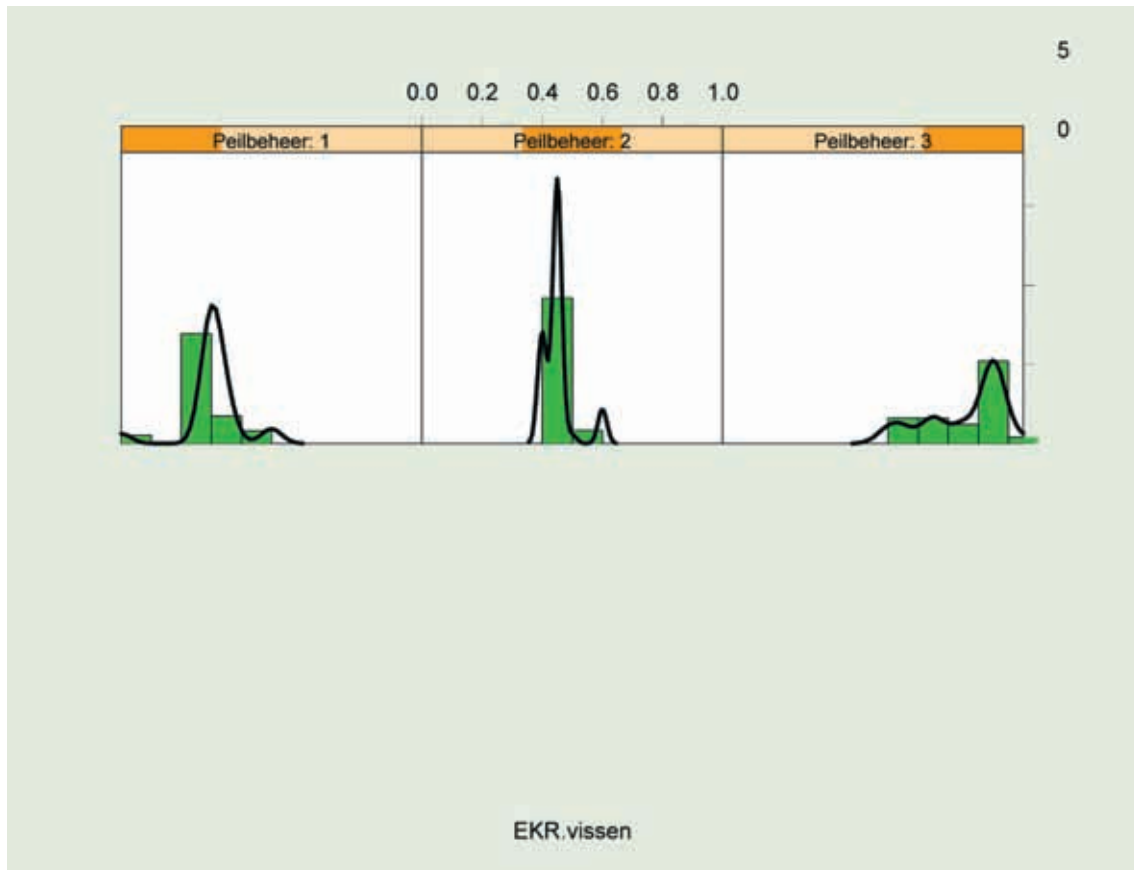
3) Peilbeheer:3 25 4.134e-001 0.8080

6) Totaal.P&lt;0.23 17 3.765e-002 0.8882 \*

7) Totaal.P&gt;0.23 8 3.375e-002 0.6375 \*



Figuur D.7 EKR-voorspellingen met *randomForest* (500 bomen en een trainingsset van 80% van het totaal aantal sloten). Kengetallen: MAD=0.023, RMSE=0.050 en  $R_{adj}^2=0.94$ . Zie voor een methodische uitleg paragraaf 3.2.2.



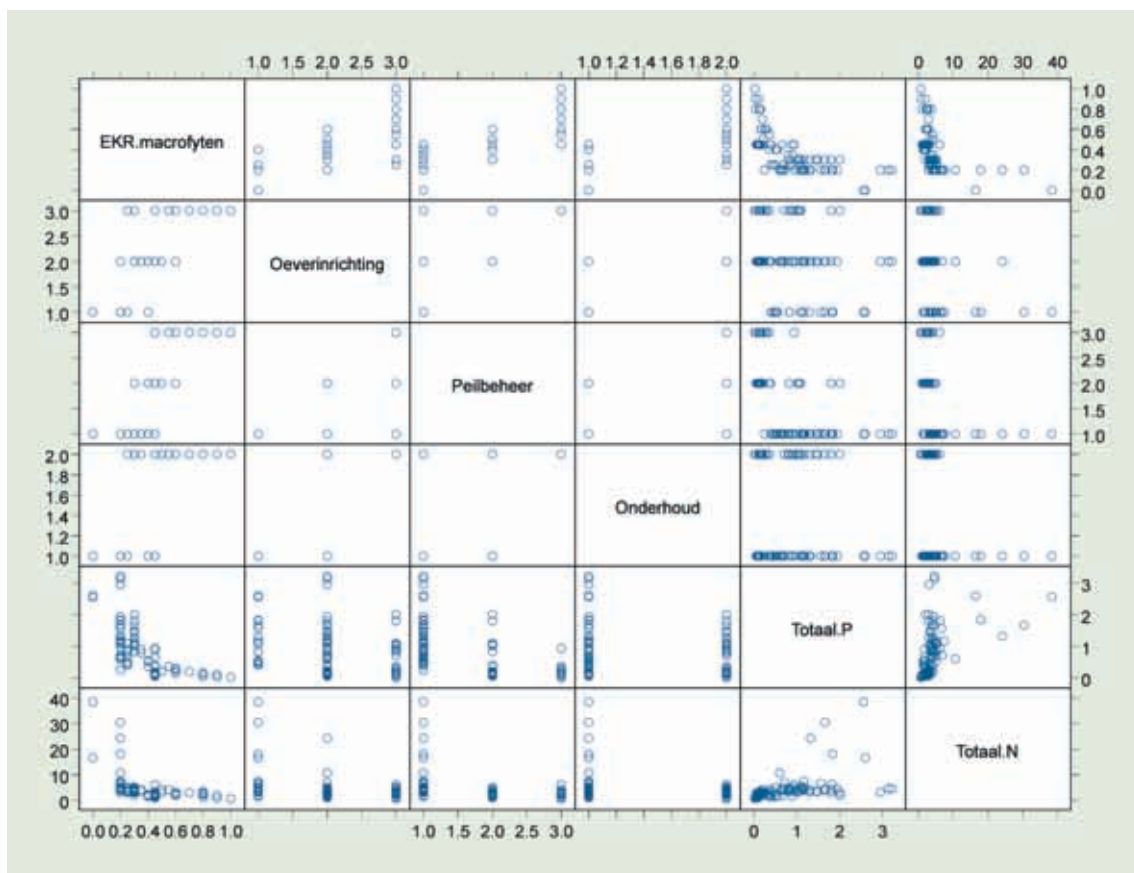
**Figuur C.8** Histogrammen en dichtheden voor de belangrijkste stuurvariabele 'Totaal.N'. Zie voor een methodische uitleg paragraaf 3.1.2.

### D.3 Deelmaatlat macrofyten

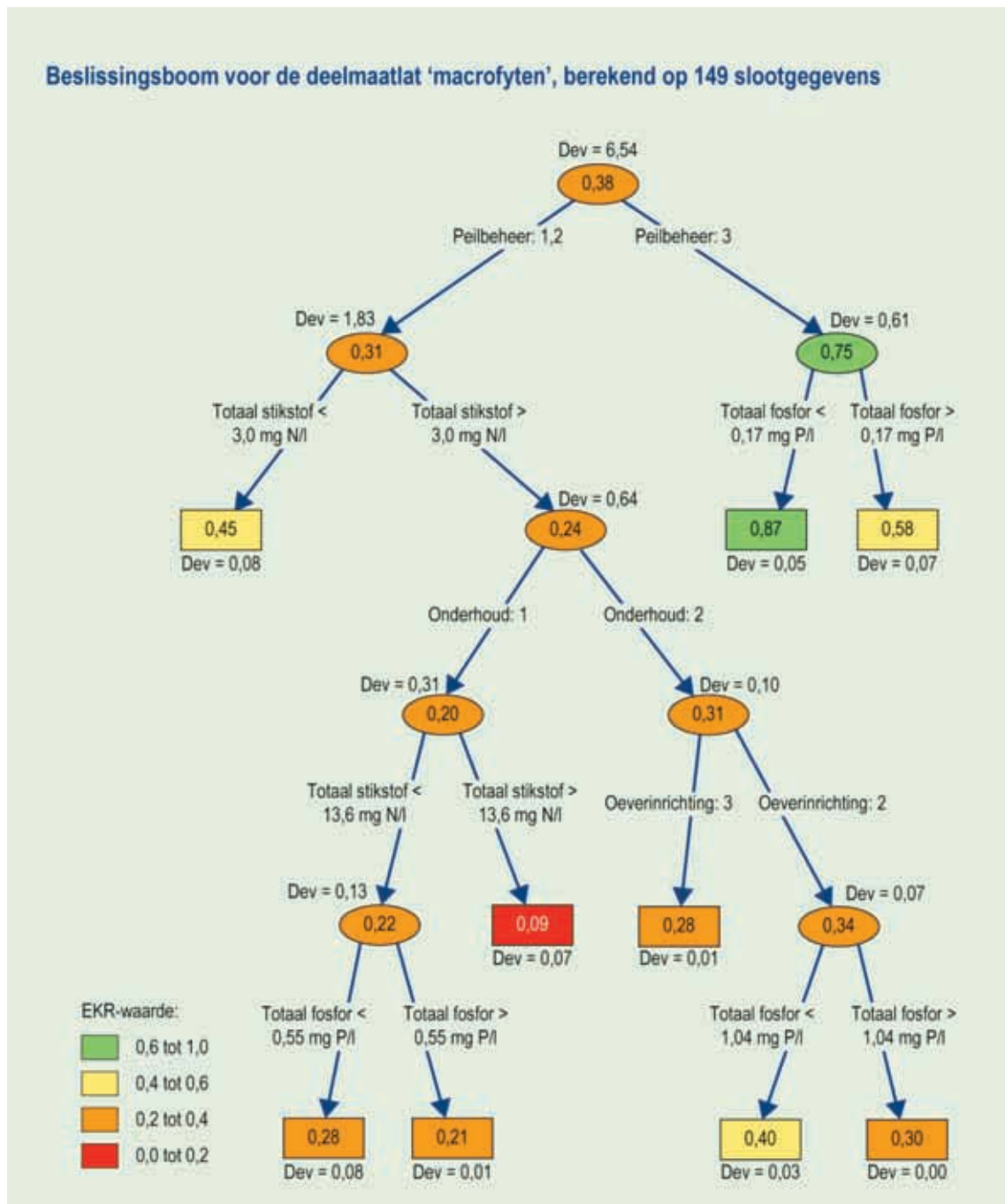
Figuur D.9 geeft de relatie tussen de deelmaatlat ‘macrofyten’ enerzijds en de beschikbare stuurvariabelen anderzijds in de vorm van een scatterplotmatrix. De figuur laat in de bovenste rij zien dat EKR-waarden stijgen bij toenemende waarden van oeverinrichting, peilbeheer en onderhoud, en dalen bij toenemende waarden van Totaal P en Totaal N. Figuur D.10 en

Tabel D.3 geven de geschatte regressieboom. Peilbeheer blijkt de belangrijkste stuurvariabele te zijn. Figuur D.11 geeft het resultaat voor een validatiecheck op 500 regressiebomen.

Figuur D.12 , ten slotte, geeft histogrammen met dichtheden voor de verschillende waarden van de belangrijkste stuurvariabele, namelijk peilbeheer.



Figuur D.9 Scatterplotmatrix voor ‘macrofyten’ (afhankelijke variabele) en vijf stuurvariabelen. Zie voor een methodische uitleg van de figuur paragraaf 3.1.1.

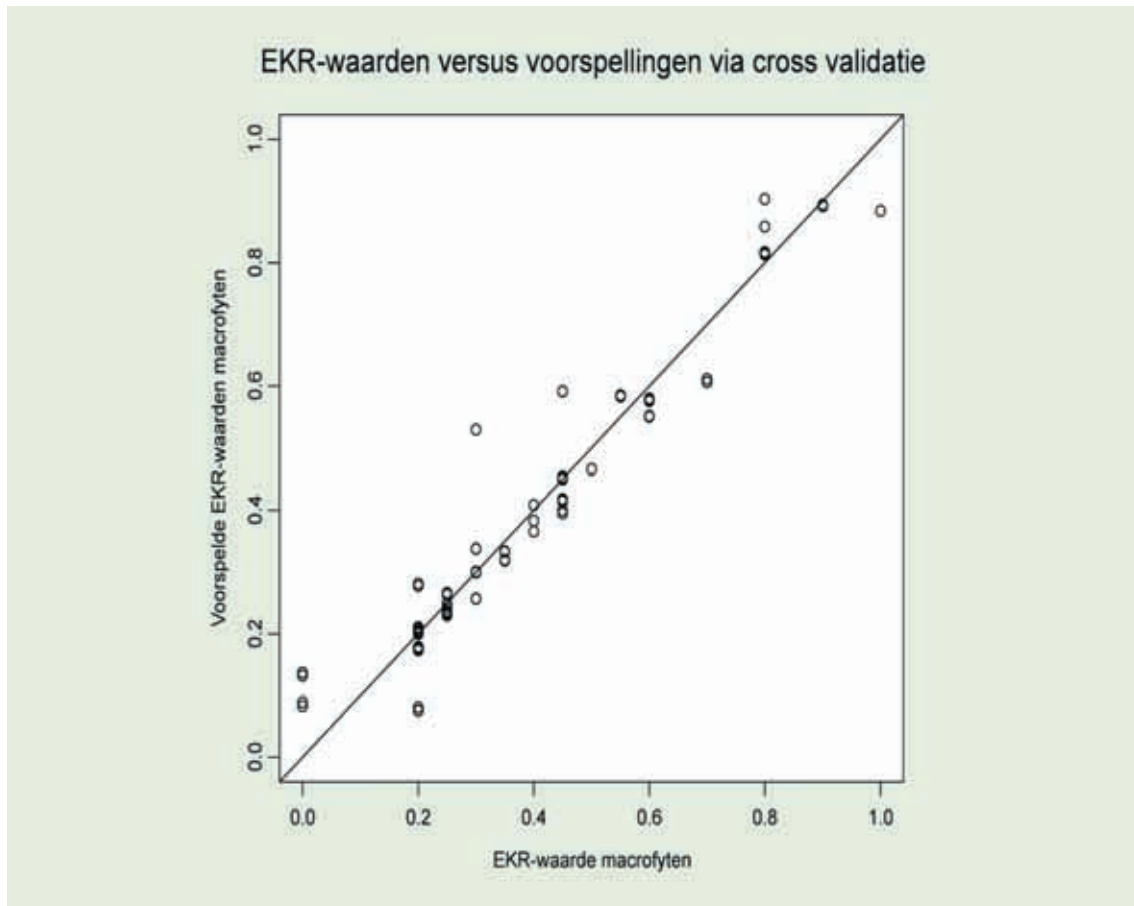


**Figuur D.10** Regressieboom met 9 eindknoten (de rechthoeken). De gemiddelde EKR-waarde van een knoop staat in de rechthoek of ellips. De bijbehorende deviantie staat steeds onder de knoop. Het belang van de splitsingen is het grootst voor de eerste splitsing en neemt af naarmate ‘gezakt’ wordt in de boom. Knoopwaarden tussen 0.6 en 1.0 vallen in de klasse ‘goed ecologisch potentieel’ (GEP). Zie voor een methodische uitleg paragraaf 3.2. De kleurcode correspondeert met Figuur 1.3. Door regressieboom verklaarde variantie bedraagt 94%.

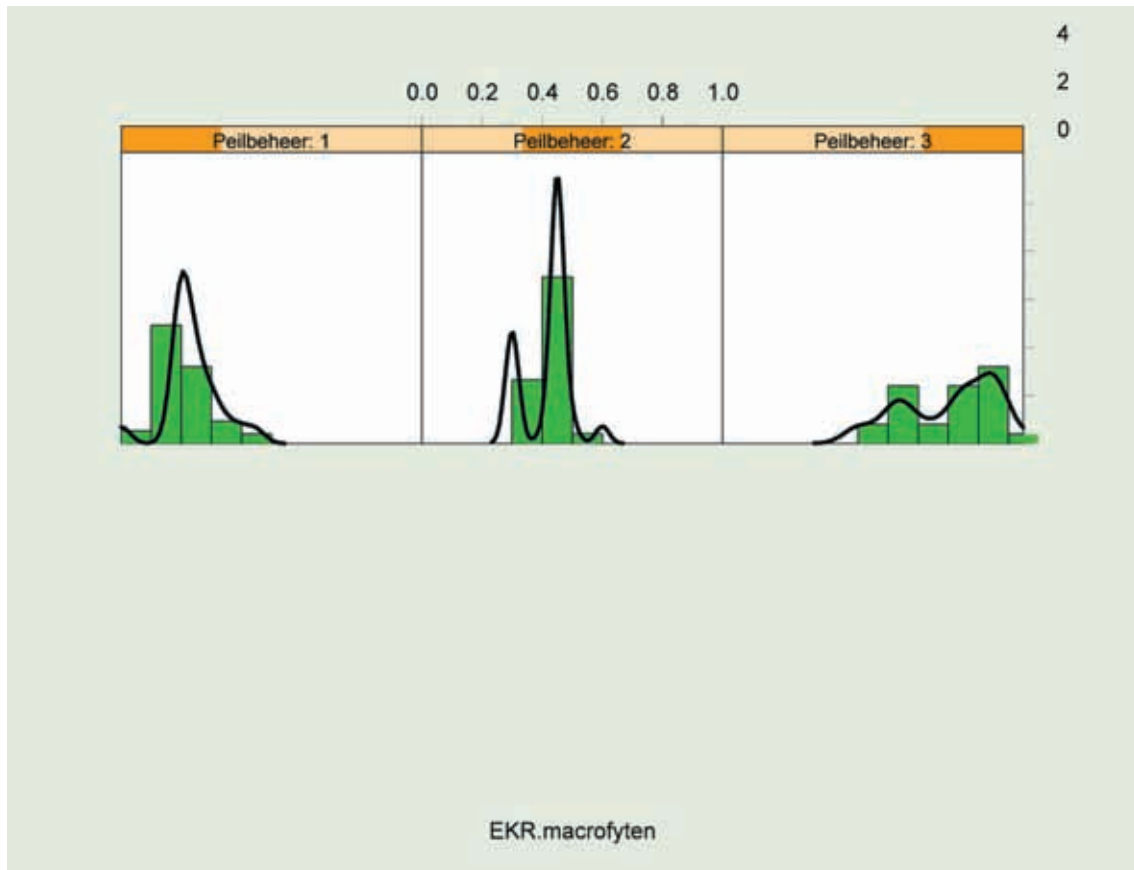


**Tabel D.3 Kwantitatieve informatie bij geschatte regressiebom.**

Number of terminal nodes: 9					
Residual mean deviance: 0.00278 = 0.3892 / 140					
Distribution of residuals:					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-1.487e-001	-5.556e-003	5.551e-017	6.073e-017	1.667e-002	1.667e-001
node), split, n, deviance, yval					
* denotes terminal node					
1) root 149 6.539e+000 0.38290					
2) Peilbeheer:1,2 124 1.834e+000 0.30850					
4) Totaal.N<2.9535 39 7.744e-002 0.44870 *					
5) Totaal.N>2.9535 85 6.371e-001 0.24410					
10) Onderhoud:1 52 3.071e-001 0.20290					
20) Totaal.N<13.5583 45 1.274e-001 0.22110					
40) Totaal.P<0.5525 9 7.500e-002 0.28330 *					
41) Totaal.P>0.5525 36 8.889e-003 0.20560 *					
21) Totaal.N>13.5583 7 6.857e-002 0.08571 *					
11) Onderhoud:2 33 1.023e-001 0.30910					
22) Oeverinrichting:3 18 1.000e-002 0.28330 *					
23) Oeverinrichting:2 15 6.600e-002 0.34000					
46) Totaal.P<1.03583 6 3.000e-002 0.40000 *					
47) Totaal.P>1.03583 9 5.435e-033 0.30000 *					
3) Peilbeheer:3 25 6.124e-001 0.75200					
6) Totaal.P<0.1725 15 5.333e-002 0.86670 *					
7) Totaal.P>0.1725 10 6.600e-002 0.58000 *					



Figuur D.11 EKR-voorspellingen met *randomForest* (500 bomen en een trainingsset van 80% van het totaal aantal sloten). Kengetallen: MAD=0.022, RMSE=0.044 en  $R_{adj}^2=0.96$ . Zie voor een methodische uitleg paragraaf 3.2.2.



Figuur D.12 Histogrammen en dichtheden voor de belangrijkste stuurvariabele 'peilbeheer' (drie discrete klassen). Zie voor een methodische uitleg paragraaf 3.1.2.



## Appendix E Verklarende kracht van de regressiebomen

In Tabel E.1 is de verklarende kracht van regressiebomen bepaald met het pakket *randomForest*. Met dit softwarepakket zijn per deelmaatlat en per watertype 500 regressiebomen geschat. Daarbij is elke boom geschat op een random-trekking van 80% van de beschikbare data. De voorspellingen zijn per regressieboom berekend over de overige 20% van de EKR-data. Gebruikmakende van de kengetallen RMSE, MAD en  $R_{adj}^2$  (zie definities (1) en (2)) worden de resultaten verkregen zoals samengevat in de tabel.

De tabel laat zien dat de verklarende kracht van de regressiebomen voor de meeste deelmaatlaten zeer hoog is:  $R_{adj}^2$  hoger dan 0.90. Er zijn enkele uitzonderingen, namelijk deelmaatlaten voor macrofauna in meren, kanalen en sloten, met resp.  $R_{adj}^2 = 0.67, 0.49$  en  $0.75$ . Dit resultaat laat zien dat EKR-waarden voor macrofauna lastiger te voorspellen zijn dan EKR-waarden voor de overige deelmaatlaten (vissen, macrofyten, fyto benthos en fytoplankton).

**Tabel E.1 Samenvatting verklarende kracht via kruisvalidatie op 500 regressiebomen per deelmaatlat (geschat met randomForest). Zie voor een methodische uitleg paragraaf 3.2.2.**

Soort waterlichaam	Deelmaatlat	RMSE	MAD	$R_{adj}^2$
<b>Beken</b>	macrofauna	0.081	0.058	0.85
	vissen	0.056	0.030	0.93
	macrofyten	0.054	0.029	0.91
	fyto benthos	0.068	0.043	0.89
<b>Meren</b>	macrofauna	0.094	0.074	0.67
	vissen	0.036	0.019	0.96
	macrofyten	0.036	0.020	0.97
	fytoplankton	0.053	0.025	0.96
<b>Kanalen</b>	macrofauna	0.14	0.097	0.49
	vissen	0.064	0.040	0.85
	macrofyten	0.066	0.042	0.82
	fytoplankton	0.052	0.027	0.93
<b>Sloten</b>	macrofauna	0.12	0.093	0.75
	vissen	0.050	0.023	0.94
	macrofyten	0.044	0.022	0.96
	---	---	---	---



## Appendix F Gevoeligheidsanalyse doublures

Zoals opgemerkt in paragraaf 2.2 bevatten de vier databases veel doublures: data voor dezelfde locatie, maar gemeten in verschillende jaren. Deze data zijn daarom niet (geheel) statistisch onafhankelijk van elkaar en kunnen de resultaten uit dit rapport wellicht beïnvloeden. Als locaties met dezelfde coördinaten in de vier databases samengenomen worden, loopt het aantal beekdata terug van 344 naar 80, het aantal meerdata van 248 naar 77, het aantal kanaaldata van 182 naar 97 en het aantal slootgegevens van 149 naar 35.

De invloed van deze doublures op de schattingsresultaten uit dit onderzoek is onderzocht door records met dezelfde x- en y-coördinaat rekenkundig te middelen wat betreft EKR-waarden en stuurvariabelen. Op deze manier zijn vier ‘ingedikte’ databases afgeleid voor beken

(80 records), meren (77 records), kanalen (97 records) en sloten (35 records). Vervolgens zijn op deze nieuwe databases scatterplotmatrices berekend (analoog aan die getoond in Figuur A.1, A.5, A.9, A.13, B.1, B.5, B.9, B.13, C.1, C.5, C.9, C.13, D.1, D.5 en D.9) en zijn opnieuw regressiebomen geschat.

Gevonden is dat de scatterplotmatrices tussen de vier soortgroepen en corresponderende stuurvariabelen voor de watertypen beken, meren en kanalen sterk lijken op die getoond in Appendix A, B en C. Ondanks het indikken blijft de vorm van de puntenwolken vrijwel gelijk. Alleen voor sloten wordt een verschil gevonden voor de relatie tussen ‘macrofauna’, ‘vissen’ en ‘macrofyten’ enerzijds en peilbeheer anderzijds: het aantal records met natuurlijk peilbeheer is sterk afgenomen in de gereduceerde database.

Bovengenoemde overeenkomsten en verschillen weerspiegelen zich ook in de geschatte regressiebomen. Voor de regressiebomen zoals getoond in Figuur A.2, A.6, A.10, A.14, B.2, B.6, B.10, B.14, C.2, C.6, C.10 en C.14 worden vergelijkbare resultaten gevonden. Wel liggen de verklaarde varianties iets lager. Ook kunnen splitsingen omdraaien: waar eerst gesplitst wordt op totaal stikstof en daarna op totaal fosfor, kan dat in de ‘ingedikte’ database precies omgedraaid zijn. Oorzaak is dat totaal fosfor en totaal stikstof onderling ook hoog gecorreleerd zijn (vooral bij beken en meren). In dat geval maakt het voor de verklarende kracht niet zoveel uit op welke variabele eerst wordt gesplitst.

De regressiebomen voor de 35 sloten wijken af van die gevonden in appendix D: peilbeheer wordt nu niet gevonden als een belangrijke verklarende variabele. Verklaring is dat in de ingedikte sloten-database er nog maar drie sloten overblijven met een peilbeheer van ‘3’. Dit aantal is te gering om voor een splitsing te kiezen.

Conclusie uit deze gevoeligheidsanalyse is dat de aanwezigheid van doublures in de datasets voor beken, meren en kanalen slechts een klein (en verklaarbaar) effect hebben op de gevonden resultaten zoals in de hoofdtekst gepresenteerd. Voor sloten blijkt de invloed van peilbeheer veel minder duidelijk wanneer de dataset teruggebracht wordt van 144 records naar 35 records. Deze stuurvariabele komt niet meer voor in de drie regressiebomen. In de conclusies in de hoofdtekst is daarom de invloed van peilbeheer vermeld als een indicatief resultaat.

Zoals vermeld aan het eind van hoofdstuk 7 verdient het aanbeveling om de datasets in vervolgonderzoek uit te breiden, met speciale aandacht voor sloten.





### **Hoe kan de ecologische kwaliteit van regionaal oppervlaktewater verbeterd worden?**

Binnen de Kaderrichtlijn Water (KRW) wordt de ecologische kwaliteit van regionaal oppervlaktewater uitgedrukt in de vorm van ratio's. Ecologische kwaliteitsratio's, afgekort als EKR's, zijn per water beschikbaar voor grotere waterorganismen, vissen, grotere waterplanten, bodemplanten en algen. Voor het beleid is het belangrijk te weten hoe de verschillende EKR-waarden verbeterd kunnen worden door het nemen van maatregelen. Kan dat door het verlagen van concentraties totaal fosfor of totaal stikstof? Of door beken beter te laten meanderen, een aangepast peilbeheer of een betere oeverinrichting?

Recent is de samenhang tussen EKR's en stuurvariabelen door Royal Haskoning bepaald via neurale netwerken. Onderhavige studie is uitgevoerd parallel aan die met neurale netwerken. Doel is de resultaten via neurale netwerken te verifiëren door een heranalyse van dezelfde data met een heel andere techniek, namelijk die van regressieboomanalyse. Dit is een statistische techniek waarbij de variaties in EKR-waarden verklaard worden door splitsingen aan te brengen in de verschillende stuurvariabelen.

Er blijkt een grote overeenkomst in resultaten te bestaan tussen beide technieken. Daarmee is een onafhankelijke verificatie gegeven van de gepresenteerde berekeningen met neurale netwerken in de Ex-ante evaluatie KRW.