



waddenacademie

Object-Based Remote Sensing van Benthische Macrofauna in de Waddenzee

Een verkennend onderzoek



Object-Based Remote Sensing van Benthische Macrofauna in de Waddenzee

Een verkennend onderzoek



waddenacademie

Colofon

Dit rapport is het resultaat van een studie naar de mogelijkheid om met remote sensing verschillende ecologisch relevante eigenschappen van de Waddenzee te monitoren. Het is uitgevoerd in opdracht van de Waddenacademie als samenwerking tussen het NIOZ en de Universiteit Utrecht.

Opgesteld door Departement Fysische Geografie,
Universiteit Utrecht

Princetonlaan 8A
3584 CB Utrecht
Postbus 80115
3508 TC Utrecht
Telefoon: +31302532749
Website: www.uu.nl/onderzoek/departement-fysische-geografie

Auteurs

Job de Vries, Henk W. van der Veer en Elisabeth A. Addink

Grafisch ontwerp omslag

BW H ontwerpers

Fotografie omslag

Sentinel-2A, ESA (26 september 2016)

ISBN

978-94-90289-47-8

Position paper 2019-02

Gepubliceerd door Waddenacademie
© Waddenacademie oktober 2019

Citatie: De Vries, J., Van der Veer, H.W. en Addink E.A., 2019.
Object based remote sensing van benthische macrofauna in de
Waddenzee, een verkennend onderzoek. Rapport Waddenacademie
2019 02.

Contactpersoon

Klaas Deen
Secretaris
T 058 233 90 31
E klaas.deen@waddenacademie.nl
www.waddenacademie.nl

De basisfinanciering van de Waddenacademie
is afkomstig van het Waddenfonds.

VOORWOORD

De Basismonitoring Wadden is opgezet als een integrale, systeemgerichte gebiedsmonitor die alle informatie in en over het Waddengebied samenbrengt. Deze monitoring levert waardevolle langjarige meetreeksen op, van waterkwaliteit tot aantallen zeehonden. De Basismonitoring Wadden draagt daarmee bij aan de duurzame bescherming en ontwikkeling van de Waddenzee als natuurgebied en het behoud van het unieke open landschap, de hoofddoelstelling voor de Waddenzee.

Sinds 2008 wordt er in het kader van project SIBES (Synoptic Intertidal Benthic Survey) een jaarlijkse, Waddenzee brede, bemonstering van bodemdieren en sediment uitgevoerd op de platen van de Nederlandse Waddenzee, inclusief Eems-Dollard. Hiervoor worden volgens een 500x500 meter grid op ongeveer 4500 locaties monsters genomen. Bodemdieren in het sediment vormen in de Waddenzee de belangrijkste schakel tussen primaire productie en de rijkdom aan vogels en vissen.

SIBES is daarmee een meetprogramma dat de sedimentkarakteristieken en de bodemdieren jaarlijks voor de hele Waddenzee letterlijk in kaart brengt. Deze bron aan gegevens wordt ook gebruikt voor het volgen van effecten van menselijke activiteiten (zoals bodemdaling door gas- en zoutwinning) voor het leven op de wadplaten.

Het Kernteam Basismonitoring Wadden verkent in welke mate en vorm het meetprogramma antwoorden kan geven op de beheervragen die in het kader van de Basismonitoring zijn geïnventariseerd. Gebleken is dat beheerders van het Waddengebied meer en meer behoefte hebben aan een meetprogramma dat ook de seizoendynamiek van de wadplaten in kaart brengt. De vraag is of informatie uit aardobservaties, met een automatische herkenning van patronen op wadplaten met een hoge resolutie in tijd en ruimte, hierbij de helpende hand kan bieden. De Waddenacademie heeft de Universiteit Utrecht (J. De Vries MSC en Dr. Elisabeth A. Addink) en het Koninklijk NIOZ (Dr ir Henk W. Van der Veer) verzocht om deze vraag als pilot voor de Basismonitoring Wadden uit te voeren.

De eerste resultaten van deze pilot laten zien dat er zeker mogelijkheden zijn om aardobservaties als beheerinstrument in te zetten. Zowel de patronen in de mediane korrelgrootte van het sediment als de biomassa van de Amerikaanse zwaardschede kon aan de hand van een combinatie van verschillende informatie uit satellietbeelden (zowel textuur als kleur) vrij nauwkeurig in beeld worden gebracht.

De Waddenacademie wil de auteurs van de pilot van harte danken voor hun bijdragen aan de totstandkoming van deze studie en hoopt dat de bevindingen bijdragen aan een verdere versterking van de Basismonitoring Wadden.

Prof. dr ir Katja Philippart
Waddenacademie
Portefeuillehouder Ecologie

Samenvatting

Kennis over de bodemfauna van de internationale Waddenzee is cruciaal voor inzicht in de draagkracht voor vissen en vogels, en monitoring is daarom ook onderdeel van trilaterale en Europese richtlijnen. Gebiedsdekkende bemonstering is echter zeer arbeidsintensief, en wordt daarom (als het al gebeurt) beperkt tot een eenmalige survey per jaar. Vanwege de grote variatie in tijd wordt hierdoor slechts een “snapshot” van de biomassa en dichtheden van de verschillende soorten bodemfauna verkregen. Gezocht wordt naar methoden om de seizoendynamiek af te leiden van satellietbeelden, waarbij de algoritmen voor deze voorspelling worden gekalibreerd aan de hand van de data van een jaarlijkse survey in het Nederlands deel van het waddengebied. Dit rapport beschrijft een verkennend onderzoek naar de mogelijkheid om Geografische object-gebaseerde beeldanalyse (GEOBIA) van satellietbeelden te gebruiken bij het karteren van bodemfauna in de Waddenzee.

Geografische object-gebaseerde beeldanalyse (GEOBIA) is niet alleen gebaseerd op kleurkenmerken (zoals de intensiteit van reflectie van een specifieke band binnen het spectrum) maar ook op zogenaamde textuurkenmerken die kunnen worden berekend aan de hand van de verdeling van de pixelwaarden in een bepaald gebied (zoals de heterogeniteit van de pixelwaarden binnen een bepaalde band). Door een combinatie van 9 kenmerken (2x kleur + 7x textuur) voor elk van de 10 banden van de Sentinel2 satelliet zijn er in principe 90 variabelen beschikbaar om te vergelijken met de veldgegevens over de bodemdieren en het sediment van droogvallende wadplaten. In deze studie zijn die 90 variabelen uit een Sentinel2 beeld van september 2016 vergeleken met de biomassa en dichtheden van een aantal soorten bodemdieren en met de sedimentsamenstelling (slibgehalte, mediane korrelgrootte) van een wadplaat in het oostelijk deel van de Nederlandse Waddenzee.

Voor deze analyses is eerst per soort een verspreidingskaart gemaakt, waarna dichtheid en biomassa binnen het verspreidingsgebied met een objectset zijn voorspeld. Deze analyse was gericht op de 11 meest voorkomende en/of de meest kenmerkende soorten bodemdieren, waaronder kokkels, wadpieren en wadslakjes.

Hierbij bleek dat satellietinformatie het beste correleerde met de biomassa van de Amerikaanse zwaardschede (*Ensis directus*) en met het slibgehalte en de mediane korrelgrootte van het sediment. Voor de andere soorten lijken de eerste resultaten nog niet bruikbaar voor adequate voorspellingen aan de hand van satellietbeelden wanneer er geen velddata voorhanden zijn. De meest veelbelovende volgende stap om de voorspellingen voor deze soorten te verbeteren is om de object sets te optimaliseren voor iedere stap. Dus per soort eerst een optimale objectset voor de verspreidingskaart, vervolgens een binnen het gebied van voorspeld voorkomen optimale objectsets voor dichtheid resp. biomassa waarbij de objecten voor dichtheid en biomassa niet aan elkaar gelijk hoeven te zijn; de voorspellingen van biomassa en dichtheid worden dan niet noodzakelijkerwijs op dezelfde objecten gedaan zoals nu wel het geval was.

Inhoudsopgave

Contents

	Samenvatting	
1.	Introductie.....	1
2.	Methode.....	3
3.	Resultaten	11
4.	Discussie	20
5.	Conclusies en Aanbevelingen.....	22
	Literatuur.....	23

1. Introductie

De Waddenzee is een ecosysteem dat van groot belang is voor de omliggende landen door biologische, ecologische, economische en natuurlijke waarden (Compton et al., 2013, QSR 2017). Benthische macrofauna, zoals tweekleppigen, polychaeten en schaaldieren (zoals garnalen en krabben), spelen een belangrijke rol in het ecosysteem van de Waddenzee omdat ze bij voorbeeld organische stoffen afbreken, voedingsstoffen recycleren en een belangrijke primaire voedselbron zijn. Op de wadplaten wordt de verspreiding van macrofauna, voor elke soort op unieke wijze, bepaald door een verscheidenheid aan ecosysteemp parameters zoals sedimentsamenstelling, temperatuur, duur van droogligging, hoogte en lokale morfologie (Van der Wal et al., 2008; Choi et al., 2011a, Compton et al., 2013). De soortensamenstelling van macrofauna wordt gebruikt als een indicator voor de kwaliteit en gezondheid van kustgebieden zoals het waddengebied, zowel op trilateraal (Drent et al., 2017) als op Europees (Ysebaert et al., 2008) niveau.

De mechanismen en ecosysteemp parameters die de verdeling van leefomgeving, over tijd en in de ruimte, bepalen zijn niet altijd bekend (Kraan et al., 2013). Daarnaast is de dynamiek van het Waddensysteem groot, zo verandert de waterdiepte en bathymetrie gedurende ieder getij. De beschikbare data is vaak beperkt tot een momentopname van individuele soorten of tot één locatie, waardoor er weinig bekend is over de totale ecosystemedynamiek in ruimte en tijd. De aanwezigheid van soorten voor een groter gebied wordt daarom vaak voorspeld aan de hand van regressie-analyses of interpolatietechnieken (Van der Wal et al., 2008). Hierbij wordt vaak een voorkomenskans per soort voor het gebied bepaald (Lyashevsk et al., 2016). Door voorkomenskansen en waarnemingen te combineren kunnen voorkomenskaarten worden gemaakt.

Satellietwaarnemingen worden succesvol gebruikt voor het karteren van ecosystemeigenschappen die van invloed zijn op de ruimtelijke verdeling van soorten en spectrale eigenschappen (Dronova et al., 2015; Choi et al., 2011b; Jung et al., 2015). Hiervoor wordt een verscheidenheid aan sensoren gebruikt, waaronder optische remote sensing, LIDAR, hyperspectraal en radar. Gerealiseerde niches, met bijbehorende dynamiek, kunnen daarmee geschat worden aan de hand van relaties tussen de gevonden variabelen en de lokale verdeling van de macrofauna (Kraan et al., 2013). Het succes hangt hierbij samen met de mogelijkheid om relevante variabelen te schatten aan de hand van spectrale informatie. Biomassa of dichtheden van macrozoobenthos anders dan bovengrondse schelpdierbanken in het intergetijdengebied van de Waddenzee zijn nog niet of nauwelijks bestudeerd met remote sensing (Winter et al., 2017). Deels heeft dit te maken met de relatief homogene wadbodem en dus de relatief uniforme uiting in satellietbeelden (Kraan et al., 2013). Deze homogeniteit in satellietbeelden, ook veroorzaakt door de relatief lage ruimtelijke resolutie van de beelden t.o.v. de ruimtelijke variatie in de geomorfologie, leidt ertoe dat er slechts kleine spectrale verschillen zijn. Daarnaast is het aantal variabelen waarmee voorspeld kan worden beperkt. De ecologische informatie die kan worden afgeleid uit satellietbeelden voor relatief homogene wadplaten lijkt daarmee beperkt.

Hier biedt geografische object-gebaseerde beeldanalyse (GEOBIA) van satellietbeelden nieuwe mogelijkheden. De beeldobjecten, die gevormd worden door segmentatie van het beeld op basis van homogeniteit, kunnen aangevuld worden met andere informatie dan spectraal (Duro et al., 2012). Een deel van de lokale variantie, en dus potentiële ruis, wordt daarmee verwijderd door deze aggregatie van informatie op objectniveau (Addink et al., 2007). Daarnaast kunnen er verschillende

segmentatieniveaus worden gebruikt; dit resulteert in objecten die overeenkomen met de schaal van verschillende landschapseenheden binnen een ecosysteem. Gevariëerde landschappen, en vooral leefgebieden van verschillende soorten, kunnen namelijk niet altijd adequaat gekarakteriseerd worden op één segmentatieniveau. Het optimale segmentatieniveau hangt daarbij af van het schaalniveau van interesse voor iedere soort, maar ook de heterogeniteit in het landschap (Addink et al., 2007; Anders et al., 2011). Zowel abiotische variabelen als biodiversiteit en soorten voorkomen moeten dus bestudeerd worden op unieke schaalniveaus (Kraan et al., 2015), iets wat gefaciliteerd kan worden door het gebruik van GEOBIA i.p.v. pixelanalyses en -classificaties van satellietbeelden.

Het toevoegen van textuureigenschappen op objectniveau zorgt ervoor dat er een grote verscheidenheid ontstaat aan variabelen die gebruikt kunnen worden bij het verklaren en voorspellen van soortenvoorkomen en -dichtheid. Hierdoor is het mogelijk om soortenvoorkomen en biomassa direct te schatten zonder gebruik te maken van ecosysteemeigenschappen die vaak ook een benadering zijn of vastgesteld aan de hand van (spectrale) informatie. In de wetenschap dat soorten een effect kunnen hebben op verschillende landschapseenheden (en bijbehorende schaalniveaus), proberen we hier proof of concept te leveren dat soorten op basis van textuur- en spectrale informatie uit een Sentinel-2-beeld voorspeld kunnen worden op unieke schaalniveaus.

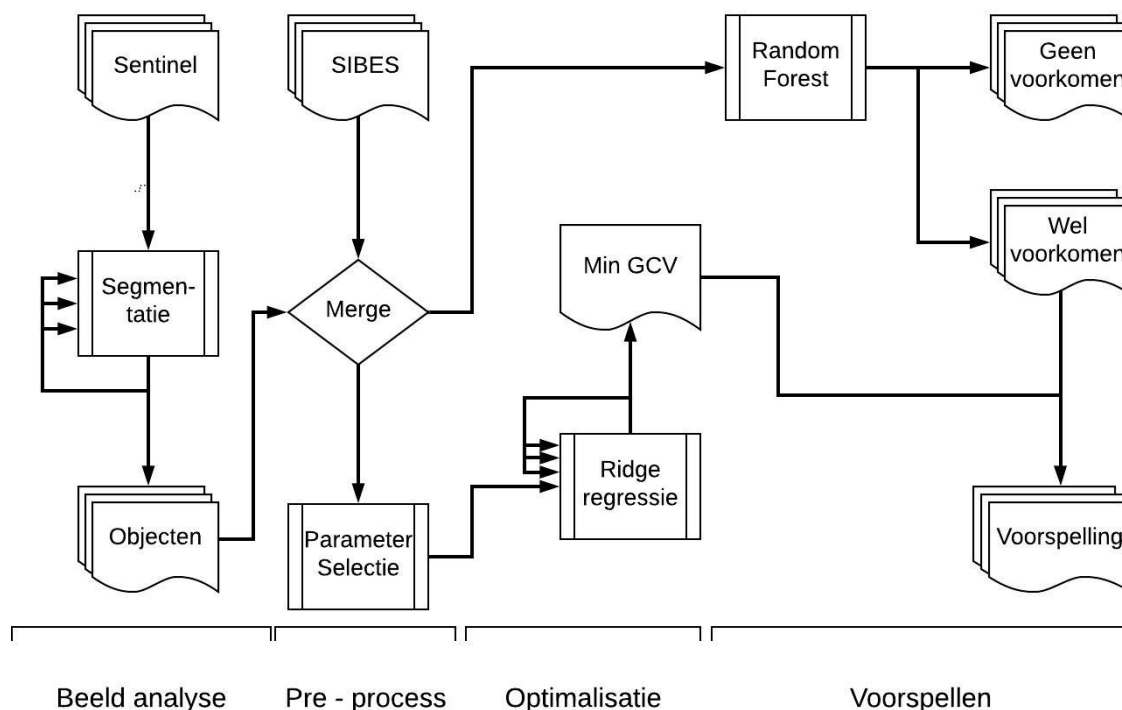
Als studiegebied hebben we het Pinkegat en de Zoutkamperlaag geselecteerd (figuur 1). Criteria waren de beschikbaarheid van zowel uitgewerkte SIBES-data van bodemdieren en sediment (zie 2.2) als een Sentinelbeeld dat wolkenvrij moest zijn en opgenomen tijdens laag water. Met behulp van GEOBIA hebben we de ruimtelijke variatie in gerealiseerde niches geschat van soorten in het onderzoeksgebied, om dit later voor de gehele Waddenzee te kunnen doen. Het doel van het verkennend onderzoek was de mogelijkheden van GEOBIA te onderzoeken voor het vlakdekkend karteren van bodemfaunasoorten (voorkomen, dichtheid, en biomassa), en verbeterpunten van de methodologie op te sporen.



Figuur 1. Sentinel-2 beeld van het Pinkegat ten noordwesten en de Zoutkamperlaag ten noorden van het Lauwersmeer. Het studiegebied is afgebeeld in Colour-Infrared, met op de achtergrond het Sentinel-2 beeld in RGB.

2. Methode

Er zijn gebiedsdekkende kaarten gemaakt voor zowel afzonderlijke soorten en als voor gebiedseigenschappen. De hoofdlijnen van de analyse bestaan uit vier fases stappen (figuur 2), namelijk Beeldanalyse, Preprocessing, Optimalisatie en Voorspellen. In Beeldanalyse wordt het beeld geselecteerd en gesegmenteerd naar de verschillende objectsets. In Preprocessing worden de variabelen uit SIBES geselecteerd en gekoppeld aan de objecten, waarna per SIBES-variabele de parameters worden bepaald die voor de voorspelling gebruikt gaan worden. Om te corrigeren voor de soms lage verhouding tussen het aantal observaties en variabelen (weinig observaties tov veel variabelen) worden in Optimalisatie correctieparameters voor de statistische voorspelling vastgesteld, waarna in Voorspelling de soortspecifieke en de gebiedskaarten worden gemaakt. De vier fases worden hieronder verder uitgewerkt.



Figuur 2. Procesbeschrijving toegepast om tot een voorspelling te komen van biomassa en dichtheid van soorten. Onderverdeeld in de stappen Beeld analyse, pre-proces, optimalisatie en uiteindelijk voorspellen.

2.1. Beeldanalyse

2.1.1. Sentinelbeeld

Sentinel2 heeft in 2016 twee wolkenvrije beelden opgenomen van het Pinkegat en de Zoutkamperlaag, beide op 25-09-2016 om 10:41 UTC. Het absoluut laagtij was die dag om 11:50 lokale tijd, dwz 9:50 UTC nabij Schiermonnikoog met -99 cm. Sentinel bevat vier banden met spectrale informatie op 10 m resolutie: Blue (490 nm), Green (560 nm), Red (665 nm) en Near-Infrared (NIR, 842 nm). Drie *Red Edge* banden (705, 740 en 783 nm), een tweede smallere NIR (865 nm) en twee *shortwave* infrarood banden (SWIR, 1610 en 2190 nm) hebben een resolutie van 20 m (tabel 1). De laatste banden, die voornamelijk voor verschillende correcties gebruikt worden, hebben een resolutie van 60 m: Aerosol (443 nm), waterdamp (945 nm) en een cirrus band (1375 nm). Het beeld is opgehaald via de Earthexplorer (USGS) en bevat *top-of-atmosfeer* (TOA) reflectie data in *digital numbers* (DN). Het gaat hier om niveau 1C data die al radiometrisch en geometrisch gecorrigeerd waren en daarmee direct klaar voor gebruik waren.

Tabel 1. Spectrale informatie van Sentinel2-beeld

Spectrale banden	Gem. golflengte [nm]	Resolutie [m]	Bandbreedte [nm]
Band 1 – Aerosol	0.443	60	20
Band 2 – Blauw ^{*^}	0.490	10	65
Band 3 – Groen ^{*^}	0.560	10	35
Band 4 – Rood ^{*^}	0.665	10	30
Band 5 – Red Edge 1 [^]	0.705	20	15
Band 6 – Red Edge 2 [^]	0.740	20	15
Band 7 – Red Edge 3 [^]	0.783	20	20
Band 8 – NIR ^{*^}	0.842	10	115
Band 8A – Smalle NIR [^]	0.865	20	20
Band 9 – Waterdamp	0.945	60	20
Band 10 – SWIR – Cirrus	1.375	60	20
Band 11 – SWIR 1 [^]	1.610	20	90
Band 12 – SWIR 2 [^]	2.190	20	180

* Banden die gebruikt zijn voor beeldsegmentatie

[^] Banden die gebruikt zijn voor objecteigenschappen

De twee Sentinel2-beelden bleken oorspronkelijk samen één beeld te zijn geweest en konden daardoor naadloos worden samengevoegd. Een uitsnede is gemaakt van het gehele onderzoeksgebied tussen Schiermonnikoog in het oosten, Ameland in het westen en Lauwersoog in het zuiden (figuur 1). Het uiteindelijke beeld heeft een omvang van 1429 rijen en 4126 kolommen wat, met de pixelgrootte van 10 bij 10 m (100 m²), neerkomt op 14,29 km bij 41,26 km.

2.1.2. Beeldsegmentatie

In plaats van de spectrale informatie op pixelniveau te analyseren maken we gebruik van object-gerichte beeldverwerking. Hiervoor wordt het beeld eerst gesegmenteerd tot objecten (Addink et al., 2012). De segmentatie is uitgevoerd in eCognition 9.2. (Definiens, 2003) met als input de vier hoge-resolutie banden (groen, blauw, rood en NIR) uit het Sentinel2 beeld (Copernicus, 2015 verwerkt door ESA). Dit is een bottom-up benadering die begint met willekeurig geselecteerde pixels die vervolgens worden geclusterd met aangrenzende pixels, tot de heterogeniteit binnen het nieuwe object te groot wordt. De maximale heterogeniteit wordt bepaald door de schaalparameter, een maat van de totale toegestane variantie binnen het ontstane object. In deze studie varieerde de schaalparameter tussen 10 en 200. Een hogere schaalparameter leidt tot grotere, en minder homogene, segmenten. Als de parameters geoptimaliseerd worden voor specifieke geomorfologische eenheden, dan kunnen deze nauwkeuriger het oppervlak weergeven (Anders et al., 2011). Resulterende beeldobjecten worden gelinkt aan textuurinformatie uit de inputlagen: contrast, correlatie, verschil, entropie, gemiddelde, standaarddeviatie en homogeniteit voor alle banden met een resolutie van 10 of 20m (tabel 1). Door het koppelen van de beeldeigenschappen aan de segmenten ontstaan de objecten; deze hebben dus zowel ruimtelijke als thematische eigenschappen.

Om textuur te beschrijven is gebruik gemaakt van de *Gray Level Co-occurrence Matrix* (GLCM) (Haralick et al., 1973) die de frequentie van pixelwaarden buurcombinaties weergeeft binnen objecten voor geselecteerde banden. De matrix kan op verschillende manieren worden geanalyseerd en geeft zo informatie over de textuur (tabel 2). Deze zeven textuurvariabelen voor elk van de tien spectrale banden hebben het voordeel dat ze ruimtelijke patronen beschrijven en daarmee extra onderscheid toevoegen, of zelfs compenseren voor gebrekkige variatie in spectrale informatie (Dronova et al., 2015). Voor de 10 banden is naast de gemiddelde spectrale reflectie ook de standaardafwijking binnen

Tabel 2. Definities van de gebruikte textuur- en spectrale variabelen.

Objectvariabelen	Beschrijving
GLCM-textuurvariabelen	
Contrast	Verschil tussen de hoogste en laagste pixelwaarde (intensiteit) binnen het object
Correlatie	Maat voor lineaire afhankelijkheid tussen combinaties van pixelwaarden binnen het object
Ongelijkheid	Lineair hogere pixelwaarden met toenemend contrast
Entropie	Maat voor wanorde versus uniformiteit in pixelwaarden binnen het object
Gemiddelde	Gemiddelde van de pixelwaarden die bijdragen aan de GLCM
StD	Maat voor heterogeniteit van pixelwaarden binnen matrix
Homogeniteit	Maat voor geslotenheid of nabijheid voor de verdeling van unieke pixelwaardes.
Spectrale variabelen	
Gemiddelde	Gemiddelde reflectie per object
StD	Standaarddeviatie reflectie per object

objecten toegevoegd, resulterend in twee spectrale kenmerken per object per band. In totaal leverde dat dus 90 objectvariabelen op: voor elk voor de 10 banden die gebruikt zijn voor objecteigenschappen (Tabel 1) in totaal 9 variabelen bestaande uit 7 GLCM-textuurvariabelen plus 2 spectrale variabelen (Tabel 2) die beschouwd kunnen worden in de analysestappen.

De focus bij het bepalen van de optimale beeldsegmentatie ligt op geometrische match tussen landschapseenheden waarin men is geïnteresseerd en de gecreëerde segmenten (Aguilar et al., 2016; Drăguț et al., 2010). Idealiter hebben de segmenten dezelfde grootte of zijn ze kleiner dan de landschapseenheid die ze dienen weer te geven. In het geval van complexe landschappen, zoals de Waddenzee, zijn de eenheden vaak lastig te begrenzen in het veld en daarmee in het beeld. Het helpt dan een hiërarchie op te zetten met verschillende segmentatieniveaus. Dit faciliteert uiteindelijk het onderscheid in relevant schaalniveau voor verschillende leefomgeving of groepen (Blaschke, 2010).

Voor het vaststellen van de ruimtelijke eenheden voor de verschillende variabelen (aanwezigheid, dichtheden en biomassa van de soorten, totale biomassa en dichtheid, en sedimenteigenschappen) is gekozen voor een range van schaalparameter tussen 10 en 200. Daarnaast wordt telkens de compactness van de objecten voor 10% meegenomen in de schaalparameter; naast de spectrale heterogeniteit wordt zo ook de grilligheid van de vorm beperkt. De objecten blijven zo enigszins compact en geven de directe omgeving van het SIBESpunt beter weer.

Aangezien we geïnteresseerd waren in de wadplaten is getracht alle water- en vaste-landgebieden uit te sluiten van de analyse. Hiervoor is een masker gemaakt in eCognition op basis van twee spectrale waterindexen: NDWI en MNDWI. NDWI is een verhouding tussen de NIR-band en Groene band (McFeeters, 1996) die gebruikt kan worden om water te onderscheiden. Alle objecten met NDWI waarden lager dan -0.42 worden geclassificeerd als water terwijl alle pixels met een waarde groter dan 0.35 worden geclassificeerd als land. Voor de overgebleven objecten (NDWI-waarde tussen -0.42 en 0.25) is vervolgens de MNDWI bepaald. Deze geeft een verhouding tussen de MIR en Groene band weer (Xu, 2006). Alle overgebleven objecten met een MNDWI waarde kleiner dan -0.25 zijn geclassificeerd als droogliggende wadplaat.

2.2. SIBES data

Het geselecteerde gebied wordt jaarlijks gemonitord door het ‘*synoptisch intertidal benthic survey*’ (SIBES) onderzoek. Dit netwerk bestond in 2016 in totaal uit 4029 vaste waarnemingspunten op de wadplaten in de Waddenzee (Bijleveld et al., 2012; Compton et al., 2013), inmiddels is het gebied verder uitgebreid met delen van de Duitse Waddenzee met in totaal meer dan 6000 monsterpunten (H.W. van der Veer, pers. comm.). Alle organismen groter dan 1mm worden met behulp van een kernmeting tot 25 centimeter diepte verzameld en vervolgens geregistreerd, geïdentificeerd en gewogen. Dit leidt tot een gecombineerd monster met een oppervlakte van 177 cm² of 173 cm² waarbij ook de sedimentkarakteristieken slibpercentage en mediane korrelgrootte worden gemeten (Lyashevsk et al., 2016). Als laatste wordt de biomassa van individuele soorten, of meerdere individuen van dezelfde soort, gemeten als asvrij drooggewicht (gADW m⁻², berekend als drooggewicht – asgewicht). Hierbij is het dus mogelijk dat dezelfde soort meerdere keren per punt werd geregistreerd en voor deze analyse geaggregeerd tot gemiddelde dichtheid en totale waargenomen biomassa per m². In 2016 is het SIBES veldwerk uitgevoerd in juli en augustus.

2.2.1. Soortenselectie

Om te voorkomen dat de analyse wordt gedomineerd door of alleen numeriek soortenvoorkomen of juist door biomassa, is er een soortenselectie toegepast. Voor de vier voorkomende soorteneenheden (Bivalven, Crustaceeën, Polychaeten en Overigen) hebben telkens één soort geselecteerd die op minimaal 50 punten was waargenomen in het onderzoeksgebied en die de hoogste bijdrage had aan de totale biomassa binnen de soorteneenheid. Dit waren *Cerastoderma edule* (B), *Urothoe spec.* (C), *Hydrobia ulvae* (O) en *Arenicola marina* (P). De selectie is vervolgens uitgebreid met zeven soorten waarvan bekend is dat ze kenmerkende leefomgevingen hebben in het onderzoeksgebied (tabel 3). In de lopende tekst zullen ze benoemd worden aan de hand van de gebruikte afkortingen. Voor iedere soort is gekeken naar de soorteigenschappen biomassa (gADW m⁻²) en dichtheid (ind. m⁻²).

Tabel 3. De geselecteerde soorten met de gebruikte afkorting.

Soort	Gebruikte afkorting
<i>Cerastoderma edule</i> (kokkel)	Ceredu
<i>Ensis directus</i> (Amerikaanse zwaardschede)	Enmdir
<i>Macoma balthica</i> (nonnetje) ^a	Macbal
<i>Urothoe spec.</i> (bulldozerkreeftjes)	Urospe
<i>Hydrobia ulvae</i> (wadslakje)	Hydulv
<i>Aphelochaeta marioni</i>	Aphmar
<i>Arenicola marina</i> (wadpier)	Aremar
<i>Capitella capitata</i> (slangpier)	Capcap
<i>Hediste diversicolor</i> (zeeduizendpoot)	Heddiv
<i>Lanice conchilega</i> (zandkokerworm)	Lancon
<i>Scoloplos armiger</i> (wapenworm)	Scoarm

^a Tegenwoordig *Limecola balthica* genoemd

2.2.2. Puntobservaties

Alle SIBES-punten die ten tijde van de Sentinel opname binnen het studiegebied op een drooggevalen plaat lagen (dus niet overlappen met het gemaskeerde water) worden gebruikt in de verder analyse. De SIBES-data leveren naast de soortspecifieke informatie ook de informatie voor de gebiedseigenschappen waar kaarten van zijn gemaakt: de gemiddelde korrelgrootte en slibpercentage van het sediment, de totale dichtheid, biomassa en biodiversiteit van het macrozoobenthos. Mediane korrelgrootte en slibpercentage waren al beschikbaar per puntobservatie.

Monsternamen op de meetpunten op de wadplaten is gedaan met behulp van steekbuizen, tijdens laagwater te voet met een grote steekbuis (oppervlak 177 cm²) en tijdens hoogwater vanaf een boot met een kleinere steekbuis (verwerkt als mengmonster van 2 steken, met een gesommeerd oppervlak van 173 cm²) (Bijleveld et al., 2012).

De volgende variabelen zijn in deze studie berekend en meegenomen in de voorspellingen:

- Biomassa per geselecteerde soort (gADW m⁻²)
- Dichtheid per geselecteerde soort (ind. m⁻²)
- Biodiversiteit: Aantal unieke soorten binnen de selectie (min = 0, max = 22 soorten m⁻²)
- Mediane korrelgrootte (µm)
- Slibgehalte sediment (gewichts%)

Uiteindelijk resulteerde dit in een dataset met voor alle puntobservaties de locatie, biomassa en dichtheid per soort, slibgehalte, korrelgrootte en biodiversiteit, waarbij biomassa en dichtheid per soort alleen bij aanwezigheid van een soort op een punt beschikbaar zijn.

Tabel 4. Eigenschappen van de geselecteerde soorten uit de SIBES-data. R² is de correlatiecoëfficiënt tussen totale biomassa en dichtheid per meetpunt waar de soort is aangetroffen. Slibgehalte en mediane korrelgrootte zijn de waarden op de locaties waar de soort is aangetroffen in het studiegebied in 2016.

Soort		Aantal Meetpunten	Totaal aantal individuen	Gem. Dichtheid	Totale biomassa	Gem. Biomassa	Slib (< 0.63 µm)	Mediane Korrelgrootte	R ² (biomassa dichtheid)
		[-]	[-]	[ind. m ⁻²]	[gADW]	[gADW m ⁻²]	[%]	[µm]	[#]
Bivalven	Ceredu	196	722	2105	6056	31	20	135	0.49
	Ensdir	51	81	908	93	2	17	147	0.03
	Macbal	285	1170	2346	969	3	18	139	0.18
Crustaceeën	Urospe	164	169	589	60	0	7	165	0.96
Overigen	Hydulv	55	221	2296	132	2	24	124	0.58
Polychaeten	Aphmar	300	310	590	85	0	18	137	1.00
	Arema	197	297	861	1643	8	13	147	0.12
	Capcap	178	193	620	15	0	12	150	0.56
	Heddiv	209	319	872	793	4	21	132	0.30
	Lancon	128	206	920	1017	8	16	149	0.53
	Scoarm	287	454	904	296	1	8	161	0.65

2.2.3. Optimalisatie

Verschillende landschapseenheden, bijvoorbeeld habitats van kokkels versus die van wadpiëren, kunnen niet altijd correct weergegeven worden met het gebruik van één segmentatieniveau (Dronova et al., 2015). De zoektocht naar de optimale segmentatieniveaus is in dit onderzoek soortspecifiek gemaakt met gebruik van de RR-analyse en resulterende minimale variantie. In totaal zijn er voor elke soort 191 objectsets gemaakt, voor elke schaalparameter van 10 t/m 200 één, met daarin alle geselecteerde puntobservaties van de SIBES-dataset met bijbehorende biomassa en dichtheid. Deze puntinformatie is gelinkt aan de corresponderende objecten en de 90 bijbehorende spectrale- en textuureigenschappen.

2.2.4. Correlatiecoëfficiënten

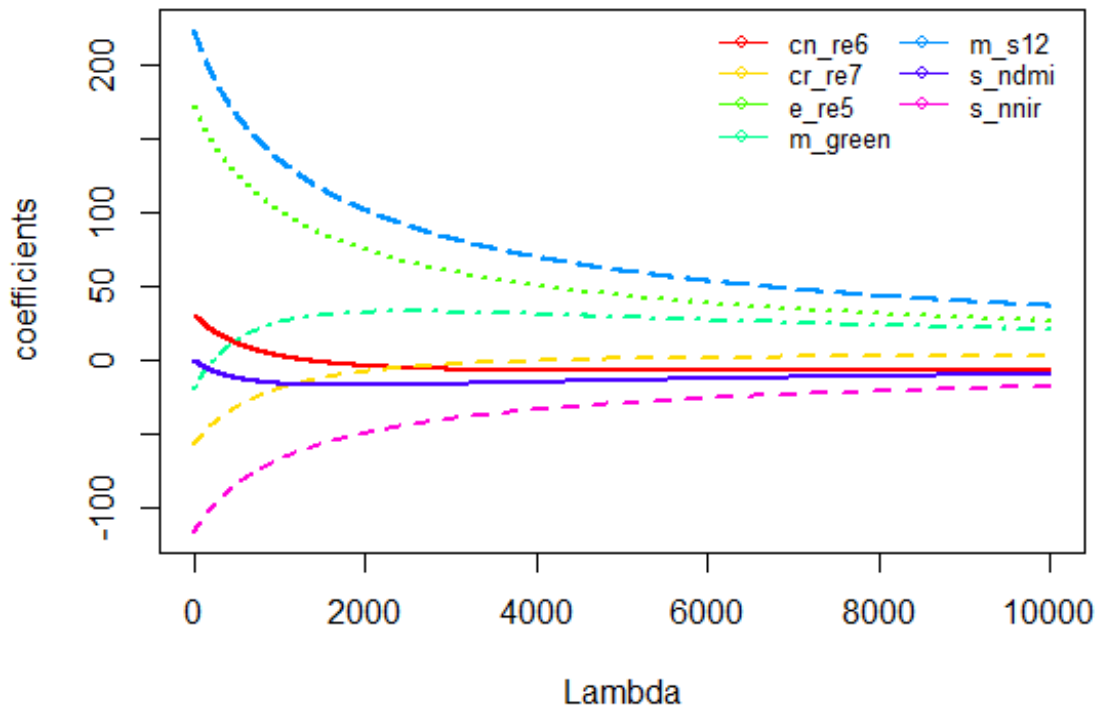
Voor elke soort zijn de objectvariabelen (zie sectie 2.1) met de hoogste correlatie geselecteerd aan de hand van Pearsons correlatiecoëfficiënten. Er is apart voor biomassa en dichtheid gekeken welke van de variabelen relevant zijn voor de voorspelling van beiden. Hierbij is de correlatie bepaald tussen de spectrale en textuurinformatie van de objecten en de biomassa- of dichtheidsgegevens van de soort. Een gevaar bij veel verklarende variabelen (hier textuur- & spectrale informatie) is het overfitten van het model dat de biomassa of dichtheid verklaart, omdat veel van de gebruikte textuurvariabelen een hoge onderlinge correlatie kennen (Clausi, 2002). Om dit te voorkomen is ook de onderlinge correlatie van de objectvariabelen bepaald. Bij objectvariabelen met een correlatie van > 0.9 is de variabele met de laagste correlatie met de te voorspellen grootte verwijderd om zo te corrigeren voor overfitten.

2.2.5. Ridge regressie

De relatie tussen de volledige set aan spectrale- en textuurinformatie en de soort-/gebiedseigenschappen is bepaald met behulp van de Ridge regressie (RR). Dit is een lineaire regressiemethode waarbij gezocht wordt naar het minimum van de voorspellingsfout, terwijl tegelijkertijd een limiet wordt gesteld aan regressiecoëfficiënten (Dijkstra, 2014). Vooral in gevallen waarbij voorspellende variabelen onderling gecorreleerd zijn (zoals bij spectrale informatie), volgt een hoge variantie voor de coëfficiënten bij gewone lineaire regressie. Dit wordt opgelost door een groeilimiet λ vast te stellen voor de coëfficiënten (Addink et al., 2007), die relatief is ten aanzien van de som van de regressiefoutmarges (geïllustreerd in figuur 3). Er is een inverse, non-lineaire relatie tussen λ en de vrijheidsgraden van de regressiecoëfficiënten, waarbij een toename van λ zorgt voor een afname van de vrijheidsgraden. Deze vrijheidsgraden worden gezien als het aantal onafhankelijke regressiecoëfficiënten dat wordt gebruikt voor de voorspelling.

Het vaststellen van de optimale schaalparameter is vervolgens gedaan met behulp van *Generalized Cross Validation* (GCV). Dit is een validatiemethode waarbij voor dezelfde range aan inputparameters (λ) alle waarnemingen één voor één worden weggelaten en voorspeld met de coëfficiënten verkregen met de overige waarnemingen (Addink et al., 2007). De resulterende GCV-waarde is gelijk aan de onverklaarde variantie, wat inhoudt dat lagere GCV-waardes een betere werking van het model aangeven. Hier hebben we voor iedere gebiedseigenschap en voor iedere soort biomassa en dichtheid met de bijbehorende geselecteerde variabelen voor 191 objectsets de GCV bepaald. Zo konden we bepalen welk segmentatieniveau optimaal is voor elke te voorspellen variabele. De corresponderende optimale λ -waarde bepaalt vervolgens de regressiecoëfficiënten die gebruikt worden voor de voorspelling van biomassa en dichtheid van verschillende soorten.

RR wordt op deze manier toegepast om vast te stellen welke schaalparameter resulteert in de kleinste variantie (GCV) en dus de hoogste voorspellingskracht voor elke soortspecifieke soorteigenschap. Voor biomassa en dichtheid zijn de GCV-waardes genormaliseerd ten opzichte van de respectievelijke laagste waardes. Om te voorkomen dat er bij één soort verschillende objectsets zijn voor biomassa en voorkomendichtheid is de som van de genormaliseerde GCV-waardes voor deze soorteigenschappen genomen. Door hiervan het minimum te kiezen komt er per soort één unieke schaalparameter uit die de dekking van de soort in het gebied bepaalt.



Figuur 3. Illustratie van de afhankelijkheid van de regressiecoëfficiënten van Lambda voor de voorspelling van de dichtheid van Macbal. De zeven variabelen zijn Contrast RedEdge 6 (cn_re6), correlatie RedEdge 7 (cr_re7), entropie RedEdge 5 (e_re5), gemiddelde groene (m_green) en SWIR12 (m_s12) banden en de standaard deviatie in NDWI (s_ndmi) en de Smaller NIR-band (s_nnir). Naar mate Lambda groter wordt komen de coëfficiënten dichter bij elkaar te liggen.

2.3. Voorspellen

Na het vaststellen van de optimale voorspellingsparameters (variabelen, schaalparameter, λ) wordt voor elke soort de objectset gebruikt die correspondeert met de schaalparameter. Met behulp van Random Forest (RF) wordt eerst voor alle objecten het aldanniet voorkomen van een soort voorspeld op basis van alle textuur- en spectrale variabelen. Dit resulteert in een binaire voorkomenskaart per soort. Voor de gebiedseigenschappen wordt deze stap overgeslagen aangezien deze nergens afwezig zullen zijn. Vervolgens wordt met behulp van de RR voor ieder object de correlatiecoëfficiënten voor alleen de geselecteerde variabelen gebruikt om zowel dichtheid als biomassa van de soort te voorspellen.

RF (Breiman, 2001) is toegepast om te achterhalen waar in het gebied een soort wel of niet voorkomt. Hiervoor is gebruik gemaakt van het randomForest pakket in R (Liaw & Wiener, 2002). Een Random Forest bestaat uit een groot aantal beslisbomen, hier 10.000. Voor elke boom wordt een subset genomen van de objecten met bekende klasse (hier de objecten waar een SIBESpunt in ligt) en van de voorspellende variabelen, waarna de subset stapsgewijs wordt onderverdeeld in homogenere groepen totdat uiteindelijk de overgebleven groepen elk tot één klasse behoren. De klassen hier zijn wel of niet voorkomen van een soort. Bij iedere stap worden alle voorkomende variabelenwaardes van alle variabelen in de subset beschouwd en de waarde met de meest homogene onderverdeling wordt geselecteerd. Het resultaat is een beslisboom waar op ieder knooppunt een beslissing wordt genomen die optimaal is voor de subset van waarnemingen en met eindpunten waar alle waarnemingen tot dezelfde klasse behoren. Er wordt een groot aantal bomen gebouwd, waardoor een RandomForest ontstaat. Vervolgens worden alle objecten (dus ook die waar geen SIBESpunt in ligt) door het RandomForest geleid met als resultaat even veel voorkomensvoorspellingen als er bomen waren. De meerderheid van de voorspelling bepaalt vervolgens voor ieder object het wel of niet voorkomen van die soort. Voor alle objecten waar een soort wordt voorspeld zijn dichtheid en biomassa voorspeld.

Het vermogen van de RF om de klassen (wel/geen voorkomen) van elkaar te scheiden wordt beschreven door de gemiddelde *Out Of Bag* (OOB) error. Deze error beschrijft voor iedere beslisboom de fout van de classificatie op de subset. Deze foutmarge voorkomt de noodzaak om nog een aparte validatie groep uit de dataset te halen (Breiman, 2001). Het gemiddelde van de OOB-foutmarges voor alle beslisbomen kan gebruikt als indicator voor hoe goed het wel of niet voorkomen van een soort bepaald kan worden (Wilschut et al., 2013). Hier is het RF gebouwd aan de hand van de helft van de punten, waarna het is toegepast op de andere helft om de nauwkeurigheid van de classificatie vast te stellen.

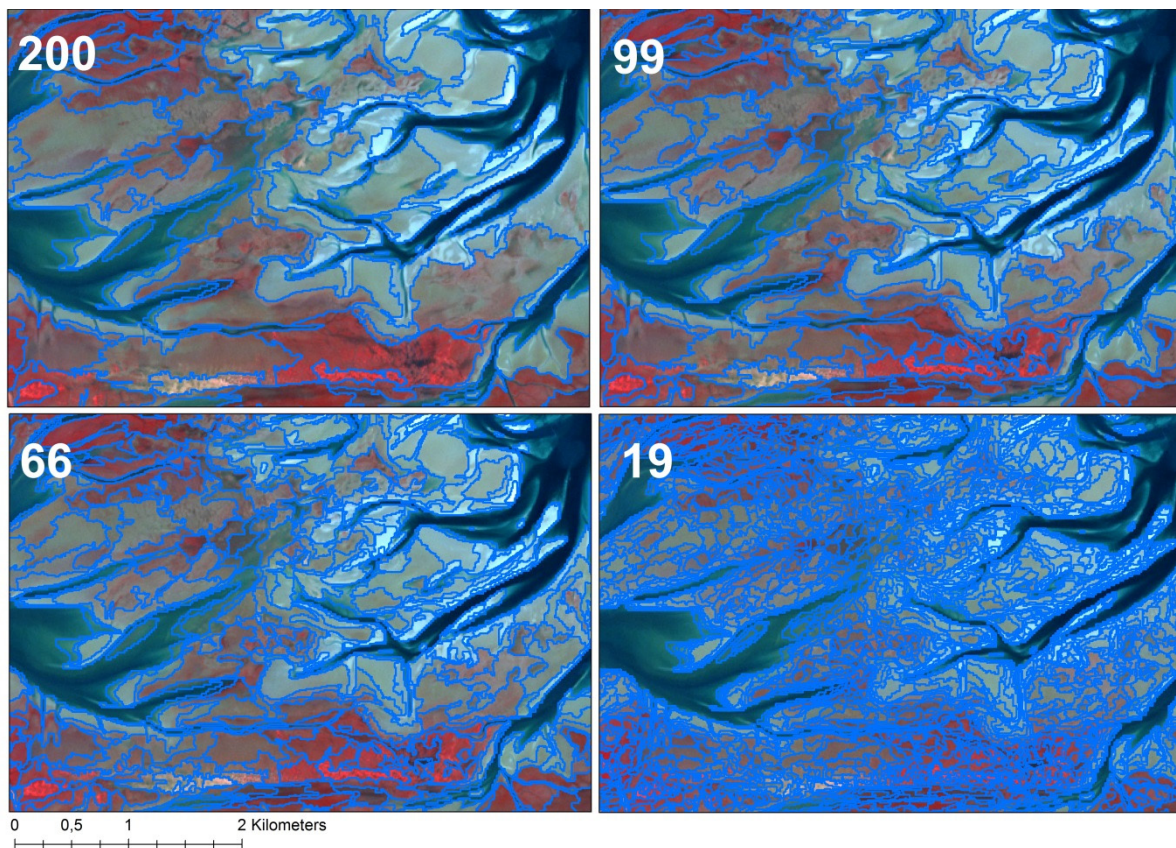
Na het vaststellen van de relevante variabelen, de bijbehorende coëfficiënten, de optimale objectset en het bepalen van soortenvoorkomen zijn vervolgens de gebiedseigenschappen en voor iedere soort biomassa en dichtheid voorspeld met de Ridge regressie.

3. Resultaten

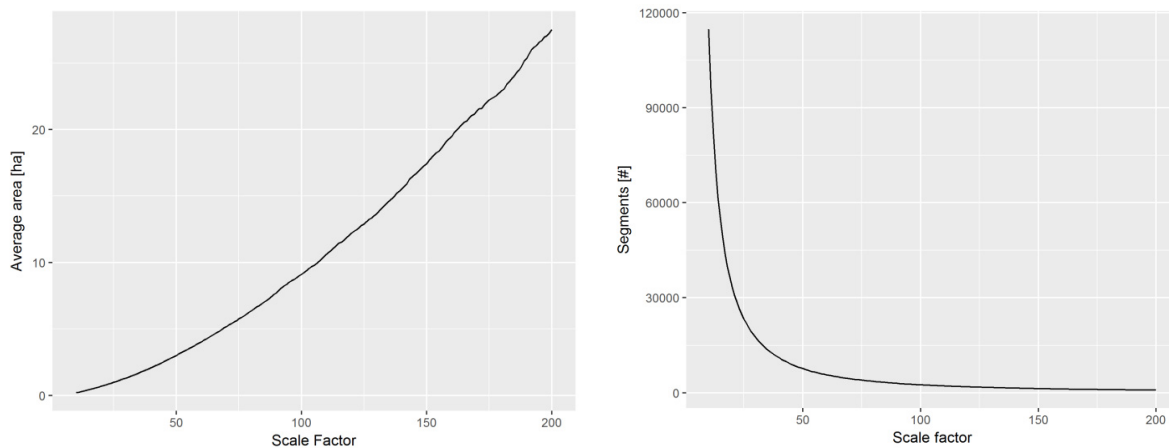
Na de selectie van de 11 soorten macrofauna (Tabel 3) bleven er 481 SIBES-punten over, waarvan er nog tien afvielen omdat ze in het water lagen op het Sentinelbeeld. In de selectie is de gemiddelde biomassa en dichtheid van soorten hoger dan voor de hele dataset geldt. In totaal zijn er 4142 individuen van de elf soorten verspreid over 471 punten geteld, met een gemiddelde dichtheid van 8,8 individuen per m².

3.1. Segmentatie

De resultaten van de segmentatie worden gebruikt om de ruimtelijke spreiding van soorten, bijbehorende dichtheid en de biomassa te voorspellen. Optimale voorspellingskracht van de beschreven methode zijn gevonden tussen een schaalparameter van 10 – 200 (figuur 4). De gemiddelde optimale objectgrootte neemt toe van 0.01 ha bij een schaalparameter van 19 (Hydulv) tot 27.5 ha bij een schaalparameter van 200 (Capcap). Het aantal objecten neemt snel af van 36441 naar 837 bij de hoogste schaalparameter van 200 (tabel 6 & figuur 5).



Figuur 4. Segmentatieresultaten voor een deel van het onderzoeksgebied. Linksboven staat telkens de schaalparameter.



Figuur 5. Toename van gemiddelde objectgrootte en de afname van het aantal segmenten bij verschillende schaalparameters

3.2. Correlatiecoëfficiënten

In elke objectset zijn aan ieder object 90 textuur- en spectrale variabelen toegekend, afgeleid van de verschillende Sentinelbanden (tabel 1). In totaal zijn er 24 te voorspellen variabelen (22 soorteigenschappen en drie gebiedseigenschappen), waarvoor telkens de relevante variabelen zijn bepaald. Van de 90 variabelen komen er 75 terug in de top 10 van meest relevante variabelen voor een soorteigenschap. De Correlatie- en Homogeniteitsvariabelen van de verschillende banden worden het meest gebruikt, respectievelijk 61 en 53 keer. De variatie aan relevante variabelen toont dat zowel spectral als textuurinformatie belangrijk is (tabel 5).

Naast de frequentie van gebruik varieert ook het voorspellend vermogen van de onafhankelijke variabelen (tabel 2) sterk, gegeven de variatie in correlatiecoëfficiënten (laagste 0,00 hoogste absolute waarde 0,57) (tabel 5). Voor individuele soorten varieert de absoluut hoogste correlatiecoëfficiënt van 0,09 voor beide typen variabelen tot 0,39 resp 0,44 voor de textuur- en spectrale variabelen. Hierbij zijn ook de abiotische factoren en totale biomassa, gemiddelde totale dichtheid en biodiversiteit meegenomen. Zo heeft bijvoorbeeld de biomassa van de wadpier (“Aremar”) een relatief grote range aan correlatiecoëfficiënten binnen de contrastgroep, met een minimum van -0.14 en een maximum van 0.1. Gemiddeld voor alle banden waarvan een textuur variant is berekend liggen de correlatiecoëfficiënten in deze groep op -0.01. Voor dichtheid van Aremar is de range kleiner van -0.07 tot 0.01 maar ligt het gemiddelde op -0.04 voor deze groep variabelen.

Een aantal soorten heeft opvallend sterke correlatie met een aantal groepen variabelen, maar andere groepen hebben ook relatief vaak een sterke correlatie met de biomassa of dichtheid van meerdere soorten. Zo heeft bijvoorbeeld de Amerikaanse zwaarschede (“Ensdir”) dichtheid sterke correlatie met de entropie van verschillende objectvariabelen (-0.24 tot -0.32, met een gemiddelde van -0.29) en is de spreiding in correlatie met homogeniteit variabelen groter (-0.39 tot 0.28 met een gemiddelde van -0.09). Over het algemeen is ook de gemiddelde reflectie (kolom gem in tabel 5) van banden een sterke variabele voor de verklaring van zowel biomassa als dichtheid van deze soort. De variabelen zijn relatief vaak gebruikt in de voorspelling maar hebben ook de hoogste correlatiecoëfficiënten. De textuurvariabelen verschil en contrast hebben voornamelijk een sterke negatieve correlatie met biomassa en dichtheid van soorten, Ensdir en de slangpier (“Capcap”) zijn hierop een uitzondering met positieve correlatiecoëfficiënten. Deze twee soorten hebben over alle groepen variabelen altijd een tegengestelde gemiddelde correlatie dan voor de biomassa en dichtheid van de overige soorten.

Tabel 5. Statistisch overzicht van de correlatiecoëfficiënten voor elk van de soort- en gebiedseigenschappen voor de negen variabelen telkens op basis van de 10 Sentinelbanden (min, gem en max geven spreiding voor deze banden weer). De kleur geeft aan of de coëfficiënt positief (blauw) of negatief (rood) is, en de intensiteit van de kleur neemt toe met grotere afstand van de 0. De eerste zeven kolommen zijn textuurvariabelen, de laatste twee (Gem. en StD) spectrale variabelen.

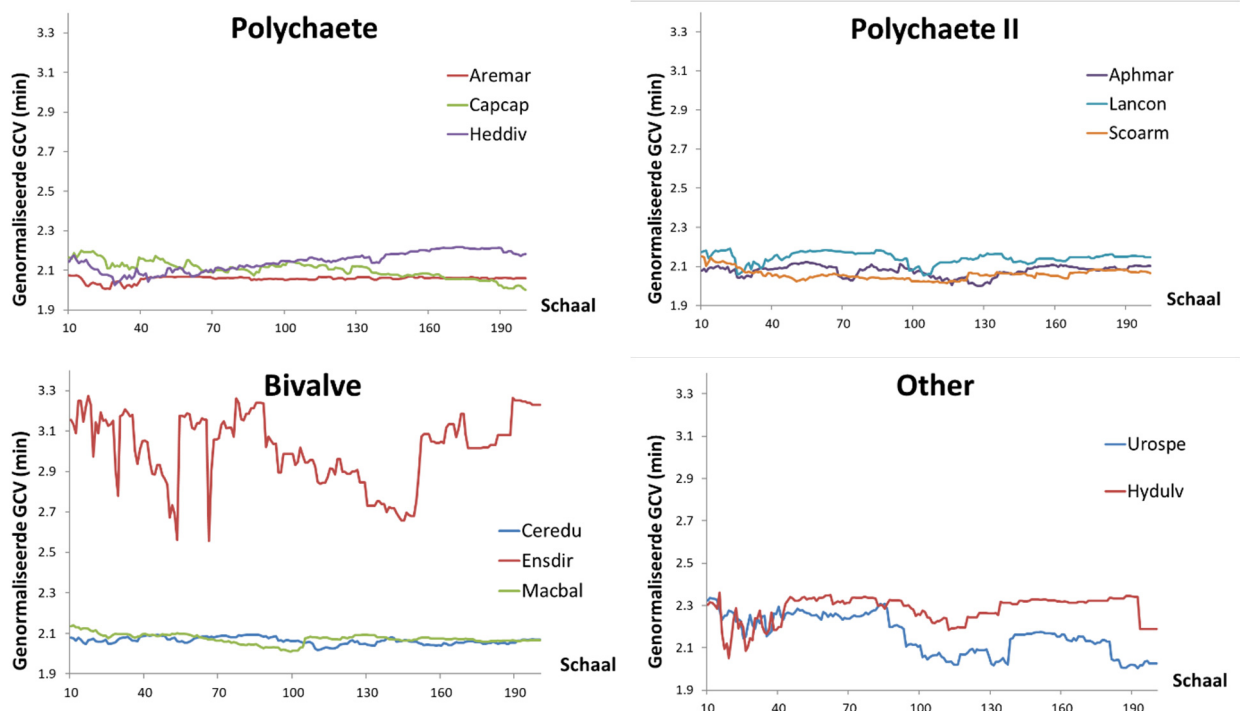
	Contrast			Correlatie			Verschil			Entropie			GLCM_Gem.			GLCM_StD			Homogeniteit			Gem.			StD			
	Min	Gem.	Max	Min	Gem.	Max	Min	Gem.	Max	Min	Gem.	Max	Min	Gem.	Max	Min	Gem.	Max	Min	Gem.	Max	Min	Gem.	Max	Min	Gem.	Max	
Abiotisch	0.06	0.09	0.15	-0.11	-0.04	0.08	0.04	0.08	0.14	-0.15	-0.12	-0.08	-0.15	-0.10	-0.06	0.04	0.14	0.25	-0.07	-0.04	0.06	-0.41	0.17	0.58	-0.06	0.11	0.28	
Silt	-0.12	-0.05	0.00	-0.10	0.03	0.12	-0.12	-0.04	0.02	0.01	0.07	0.12	0.03	0.08	0.13	-0.17	-0.07	0.02	-0.10	0.00	0.05	-0.57	-0.11	0.50	-0.35	-0.09	0.06	
Aphmar	Biomassa	-0.18	-0.13	-0.06	-0.02	0.13	0.20	-0.17	-0.12	-0.05	0.13	0.20	0.23	-0.03	0.01	0.09	-0.15	-0.08	-0.04	-0.02	0.05	0.13	-0.21	-0.04	0.23	-0.20	-0.09	-0.03
	Dichtheid	-0.18	-0.13	-0.06	-0.02	0.13	0.20	-0.17	-0.12	-0.05	0.13	0.20	0.23	-0.03	0.01	0.09	-0.15	-0.08	-0.04	-0.02	0.05	0.13	-0.21	-0.04	0.23	-0.20	-0.09	-0.03
Aremer	Biomassa	-0.14	-0.01	0.10	-0.13	0.00	0.13	-0.14	-0.02	0.08	-0.02	0.01	0.05	-0.01	0.04	0.09	-0.10	-0.04	0.03	-0.04	0.03	0.20	-0.07	-0.03	0.05	-0.11	-0.07	0.01
	Dichtheid	-0.07	-0.04	0.01	-0.06	0.03	0.09	-0.07	-0.03	0.02	0.00	0.03	0.05	0.00	0.02	0.04	-0.08	-0.04	-0.01	0.01	0.03	0.06	-0.11	-0.02	0.13	-0.10	-0.06	-0.03
Capcap	Biomassa	-0.06	0.05	0.22	-0.24	-0.04	0.08	-0.04	0.07	0.21	0.03	0.05	0.07	-0.09	-0.05	0.02	-0.04	0.05	0.13	-0.16	-0.07	-0.01	-0.14	0.13	0.25	-0.08	0.02	0.13
	Dichtheid	-0.06	0.04	0.13	-0.18	-0.03	0.08	-0.04	0.06	0.16	0.06	0.07	0.09	-0.03	0.01	0.03	0.00	0.06	0.10	-0.15	-0.05	0.03	-0.06	0.09	0.21	-0.02	0.05	0.16
Ceredu	Biomassa	-0.16	-0.09	-0.06	0.01	0.08	0.15	-0.19	-0.10	-0.06	-0.01	0.03	0.07	-0.06	0.01	0.06	-0.11	-0.09	-0.06	-0.01	0.07	0.16	-0.16	0.01	0.16	-0.11	-0.05	0.00
	Dichtheid	-0.08	-0.04	-0.01	-0.01	0.03	0.08	-0.09	-0.04	0.01	-0.10	-0.06	-0.03	0.03	0.09	0.16	-0.11	-0.05	0.01	-0.04	0.04	0.13	-0.25	-0.05	0.25	-0.18	-0.07	0.01
Ensdire	Biomassa	-0.07	0.09	0.16	-0.23	-0.10	0.08	-0.08	0.06	0.17	-0.13	-0.11	-0.08	-0.23	-0.18	-0.12	-0.11	0.07	0.19	-0.17	-0.06	0.03	-0.03	0.04	0.10	-0.01	0.05	0.16
	Dichtheid	-0.04	0.15	0.27	-0.21	-0.11	0.14	-0.11	0.12	0.29	-0.32	-0.29	-0.24	-0.19	-0.16	-0.11	0.13	0.23	0.32	-0.39	-0.09	0.28	-0.12	0.08	0.24	0.08	0.26	0.44
Heddiv	Biomassa	-0.14	-0.07	-0.03	0.02	0.08	0.17	-0.14	-0.07	-0.02	-0.05	0.03	0.06	-0.04	0.01	0.08	-0.08	-0.02	0.11	-0.05	0.07	0.14	-0.23	-0.05	0.26	-0.12	-0.03	0.04
	Dichtheid	-0.14	-0.05	0.04	-0.06	0.07	0.17	-0.16	-0.06	0.01	-0.14	-0.07	-0.03	-0.07	-0.02	0.04	-0.04	0.02	0.07	-0.08	0.04	0.18	-0.14	-0.05	0.17	-0.03	0.05	0.11
Hydulv	Biomassa	-0.15	-0.10	-0.03	-0.01	0.12	0.18	-0.17	-0.10	-0.04	0.09	0.12	0.18	-0.04	0.02	0.11	-0.16	0.00	0.09	-0.14	0.05	0.17	-0.21	0.00	0.15	-0.07	0.12	0.30
	Dichtheid	-0.15	-0.09	0.01	-0.05	0.11	0.18	-0.15	-0.10	-0.01	0.09	0.13	0.18	-0.08	-0.02	0.06	-0.15	-0.01	0.08	0.00	0.08	0.22	-0.29	0.05	0.31	-0.01	0.11	0.35
Lancon	Biomassa	-0.16	-0.02	0.03	-0.01	0.05	0.20	-0.18	-0.03	0.03	-0.02	0.05	0.11	-0.06	-0.03	0.01	-0.02	0.07	0.14	-0.05	0.06	0.21	-0.07	-0.03	0.08	0.04	0.12	0.24
	Dichtheid	-0.21	-0.10	-0.04	0.04	0.12	0.23	-0.24	-0.12	-0.03	-0.02	0.04	0.10	-0.01	0.03	0.08	-0.11	0.00	0.07	-0.07	0.10	0.28	-0.10	-0.04	0.04	-0.05	0.04	0.16
Macbal	Biomassa	-0.12	-0.06	0.06	-0.09	0.06	0.16	-0.12	-0.06	0.07	0.10	0.12	0.14	-0.07	-0.01	0.04	-0.12	-0.04	0.00	-0.09	0.03	0.14	-0.10	-0.02	0.10	-0.14	-0.07	-0.02
	Dichtheid	-0.13	-0.05	0.03	-0.06	0.06	0.16	-0.11	-0.04	0.04	0.14	0.15	0.17	-0.06	-0.02	0.00	-0.06	-0.02	0.00	-0.06	-0.03	0.03	-0.16	0.05	0.18	-0.12	-0.08	-0.01
Scoarm	Biomassa	-0.06	0.00	0.10	-0.08	0.01	0.08	-0.05	0.01	0.11	0.07	0.08	0.10	0.00	0.03	0.06	-0.02	0.04	0.11	-0.10	-0.04	0.00	-0.04	0.11	0.23	-0.08	-0.03	0.01
	Dichtheid	-0.14	-0.05	0.08	-0.10	0.06	0.18	-0.12	-0.03	0.09	0.11	0.12	0.15	0.00	0.03	0.06	-0.06	0.00	0.07	-0.09	-0.03	0.00	-0.09	0.17	0.33	-0.16	-0.08	0.00
Urospe	Biomassa	-0.13	-0.01	0.07	-0.15	-0.02	0.15	-0.14	-0.02	0.05	-0.12	-0.09	-0.07	-0.05	0.00	0.04	-0.14	-0.05	0.01	-0.04	0.05	0.12	-0.11	-0.03	0.05	-0.02	0.03	0.07
	Dichtheid	-0.11	-0.02	0.06	-0.16	-0.01	0.13	-0.12	-0.03	0.04	-0.12	-0.10	-0.07	-0.06	-0.01	0.03	-0.18	-0.07	-0.01	-0.03	0.06	0.12	-0.09	-0.03	0.04	-0.02	0.03	0.07
Totaal	Σ Biomassa	-0.07	0.02	0.07	-0.08	-0.01	0.10	-0.06	0.02	0.09	-0.05	-0.01	0.01	-0.12	-0.01	0.14	-0.08	0.02	0.11	-0.06	-0.03	0.03	-0.31	0.04	0.43	-0.14	0.03	0.22
	Dichtheid	-0.18	-0.13	-0.08	0.02	0.12	0.18	-0.17	-0.13	-0.08	0.15	0.17	0.19	-0.01	0.01	0.03	-0.15	-0.10	-0.06	0.03	0.07	0.14	-0.06	0.02	0.14	-0.12	-0.08	-0.02
	Biodiversiteit	-0.21	-0.16	-0.05	-0.05	0.15	0.24	-0.20	-0.14	-0.03	0.14	0.17	0.19	0.03	0.09	0.14	-0.16	-0.10	-0.04	-0.02	0.04	0.09	-0.29	-0.06	0.29	-0.25	-0.14	-0.01
Aantal keer geslecteerd	13			61			37			15			15			37			53			46			44			

Bij slijbgehalte en korrelgrootte zijn de gemiddelde correlatiecoëfficiënten telkens tegengesteld, als de een positief is is de ander negatief. Bij GLCM gemiddeld treedt dit bij een aantal soorten ook op voor dichtheid en biomassa. Zo heeft de slangpier (“Capcap”) een gemiddelde correlatiecoëfficiënt van -0.05 (met een range van -0.09 tot -0.02) bij biomassa, voor de dichtheid is deze juist 0.01 (met een range van -0.03 tot 0.03). Ook de zeeduizendpoot (“Heddiv”), het wadslakje (“Hydulv”) en de zandkokerworm (“Lancon”) tonen vergelijkbare patronen bij GLCM gemiddelde. Naast de groep GLCM gemiddeld, ontstaat een vergelijkbaar patroon bij de kokkel (“Ceredu”) voor Entropie en Gemiddelde reflectie. Bij Heddiv gebeurt hetzelfde bij Entropie, GLCM StD en StD reflectie. Het nonnetje (“Macbal”) heeft tegengestelde gemiddelde correlatiecoëfficiënten in de gemiddelde-reflectie en homogeniteitgroep. Bij Scoarm wordt dit alleen waargenomen bij de correlatiegroep verschil. De correlatiecoëfficiënten in tabel 5 laten zien dat de te voorspellen variabelen elk hun eigen kenmerken hebben. Het is daardoor aannemelijk dat ze ook een specifiek optimum hebben bij de objectsets.

3.3. Optimalisatie

Voor de 11 soorten is telkens voor biomassa en dichtheid apart voor elk van de 191 objectsets de minimale GCV waarde bepaald. Deze minimumwaarden zijn genormaliseerd naar de laagste waarde voor elke specifieke eigenschap en vervolgens per soort opgeteld. Dit gaf de gecombineerde, genormaliseerde GCV-waardes. De minimumwaarde is 2 en wordt verkregen wanneer voor een soort biomassa en dichtheid bij dezelfde objectgrootte de laagste voorspellingsfout (GCV waarde) hebben. Voor de elf soorten zijn de gecombineerde GCV-waardes meestal kleiner dan 2.4 (figuur 6), behalve bij Enmdir. Hier treden grote schommelingen op tussen 2.7 en 3.3.

Hoge gecombineerde en gestandaardiseerde GCV-waardes (zie hoofdstuk 2.3.2.) suggereren een groot verschil tussen de optimale voorspellingswaarde voor biomassa en dichtheid (zie bijvoorbeeld de wadpier “Enmdir” in figuur 6). Sommige soorten zoals Heddiv zien een toenemende GCV-waarde, wat duidt op meer variantie bij toenemende schaalparameter. Bij bijvoorbeeld de slangpier (“Capcap”) en het wadslakje (“Hydulv”) is dit juist het tegenovergestelde en neemt variantie af met een hogere schaalparameter.



Figuur 6. Gestandaardiseerde minimale GCV voor de volledige range aan schaalparameters bij de verschillende soorten. Hierbij zijn per soort de minimale GCV-waardes voor biomassa en dichtheid gestandaardiseerd en bij elkaar opgeteld.

De laagste gecombineerde GCV-waarde correspondeert met de optimale schaalparameter die gebruikt wordt om de ruimtelijke verspreiding, biomassa en dichtheid van soorten te voorspellen (tabel 6). Zo correspondeert de absolute minimale GCV bij de slangpier ("Capcap") met een optimale schaalparameter van 200. Het aantal vrijheidsgraden (λ) is per eigenschap bepaald en net als het aantal variabelen voor in de Ridge regressie. De correlatie geeft aan hoe biomassa en dichtheid per soort correleren voor de voorspellingen en in de SIBESdata.

Het aantal variabelen varieert van 2 tot 44. De schaal van de optimale objectset voor biomassa en dichtheid samen varieert ook sterk, van 19 tot 200. De schaal grootte van de objectset is voor sommige soorten sterk verschillend van de optimale schaal van de aparte eigenschappen (bv Aremar dichtheid), terwijl deze bij andere soorten nauwelijks verschilt (bv Heddiv biomassa en dichtheid).

Voorspelde (Pred.) en geobserveerde (Obs.) correlatie in tabel 6 geven een indicatie van verhouding biomassa en dichtheid bij de veldobservaties. Zo heeft de wadpier ("Arema") een geobserveerde correlatiecoëfficiënt van 0.12 en een voorspelde van 0.04. Tegelijkertijd is bij deze soort de schaalfactor vastgesteld op 25 terwijl deze voor dichtheid 113 was geweest indien er geen gecombineerde objectset was geselecteerd. Het verschil in het aantal variabelen verschilt ook tussen de voorspelling van biomassa en dichtheid. Soorten waarbij de correlatie relatief groter is hebben ook relatief gezien een vergelijkbaar aantal variabelen als input voor de voorspelling. Zo hebben bulldozerkreeftjes ("Urospe") een correlatie tussen biomassa en dichtheid van 0.96 die bij de voorspelling afneemt naar 0.78, het aantal bijbehorende verklarende variabelen is 8 en 9. Voor de wapenworm ("Scoarm") geldt deze redenering niet, die heeft namelijk een correct voorspelde correlatie tussen biomassa en dichtheid van 0.65, het aantal voorspellende variabelen verschilt alleen met "slechts" 6 voor biomassa en 14 voor dichtheid voorspellingen. Tegelijkertijd is de gekozen schaalparameter ook niet optimaal voor biomassa met 161 t.o.v. de toegepaste schaal van 114. Naarmate de verschillen tussen optimale schaal en toegepaste schaal, en de bijbehorende verschillen in λ , dus toenemen verandert er ook meer in de voorspelde correlatie.

Tabel 6. Statistieken voor de voorspellingen van de geselecteerde soorten en gebiedseigenschappen. Obs geeft aan op hoe veel punten een soort is aangetroffen. Variabelen geeft het aantal textuur en spectrale variabelen aan dat als input is gebruikt voor de bepaling van de optimale schaal en de uiteindelijke voorspelling. Schaal geeft aan welke objectset is gebruikt bij de voorspelling van soorteigenschappen, en λ is de maat voor het aantal vrijheidsgraden; de waarden tussen haakjes geven aan wat de optimale schaal en vrijheidsgraden waren wanneer de optima voor de twee soorteigenschappen verschillend waren. Correlatie toont de correlatie tussen biomassa en dichtheid in de voorspellingen (Pred.) en in de SIBESdata (Obs.).

Soort	Eigenschap	Obs	Variabelen	Schaal	λ	Correlatie	
		[N]	[N]	[-]	[-]	Pred.	Obs.
Aphmar	Biomassa	300	21	127	46	1	1
	Dichtheid		21	127	46		
Aremar	Biomassa	195	6	25	67	0.04	0.12
	Dichtheid		2	25 (113)	267 (170)		
Capcap	Biomassa	176	16	200	156	0.89	0.56
	Dichtheid		16	200 (194)	136 (137)		
Ceredu	Biomassa	197	12	110 (56)	312 (408)	0.29	0.49
	Dichtheid		14	110 (160)	43 (32)		
Ensdir	Biomassa	50	23	66 (144)	1	0.86	0.03
	Dichtheid		44	66 (53)	156 (6)		
Heddiv	Biomassa	209	14	29 (23)	116 (37)	0.76	0.30
	Dichtheid		12	29 (34)	44 (73)		
Hydulv	Biomassa	56	23	19 (17)	9	0.52	0.58
	Dichtheid		21	19 (26)	3 (28)		
Lancon	Biomassa	127	12	104	26	0.09	0.53
	Dichtheid		15	104 (25)	610		
Macbal	Biomassa	283	8	99	115 (191)	0.0008	0.18
	Dichtheid		7	99 (88)	21 (39)		
Scoarm	Biomassa	277	6	114 (161)	158 (149)	0.65	0.65
	Dichtheid		14	114	8		
Urospe	Biomassa	158	8	192 (131)	41 (23)	0.78	0.96
	Dichtheid		9	192 (186)	35 (30)		
	Korrelgrootte	470	22	149	1	-	-
	Slibgehalte	470	21	171	4.2	-	-
	Biodiversiteit	471	30	110	111.3	-	-

3.4. Random Forest

De RF-resultaten geven een indicatie hoe goed de spreiding van soorten voorspeld kan worden met behulp van objecten en de volledige set objectvariabelen (tabel 7). De algemene nauwkeurigheid met de correct voorspelde aan- en afwezigheid van soorten, is voornamelijk hoog voor de wadpier ("Ensdire") en het wadslakje ("Hydulv") met 0.89 en 0.87. De bijbehorende Kappa-waardes zijn negatief, wat duidt op geen overeenstemming tussen de voorspelde en waargenomen waardes. Het lage aantal waarnemingen zal hier debet aan zijn. Voornamelijk de wapenworm ("Scoarm"), de kokkel ("Ceredu"), het nonnetje ("Macbal") en een borstelworm soort ("Aphmar") zien wel een redelijke ruimtelijke overeenstemming tussen observatie en voorspelling en hebben ook relatief hoge nauwkeurigheden (kappa waardes van 0.48, 0.41 en 0.39).

Tabel 7. Nauwkeurigheid van de verspreidingskaarten. De OOB-error wordt gebruikt als indicator voor hoe goed het wel of niet voorkomen van een soort bepaald kan worden met een individuele beslisboom; lagere waardes geven een hogere betrouwbaarheid aan. De percentages bij het aantal punten geven aan hoe vaak een soort is aangetroffen in de totale dataset.

Soorten		Nauwkeurigheid			OOB-error
Soort	Punten	Fractie	Kappa	P-waarde	
Aphmar	300 (63%)	0.74	0.39	0.00	0.27
Aremar	195 (41%)	0.58	0.11	0.77	0.39
Ceredu	197 (37%)	0.71	0.41	0.00	0.39
Capcap	176 (42%)	0.67	0.21	0.27	0.27
Ensdire	50 (11%)	0.89	-0.01	0.63	0.12
Heddiv	209 (44%)	0.66	0.30	0.01	0.28
Hydulv	56 (12%)	0.87	-0.01	0.63	0.12
Lancon	127 (27%)	0.73	0.12	0.53	0.24
Macbal	283 (60%)	0.72	0.39	0.00	0.30
Scoarm	277 (58%)	0.75	0.48	0.00	0.27
Urospe	158 (33%)	0.69	0.20	0.23	0.28

3.5. Voorspelling

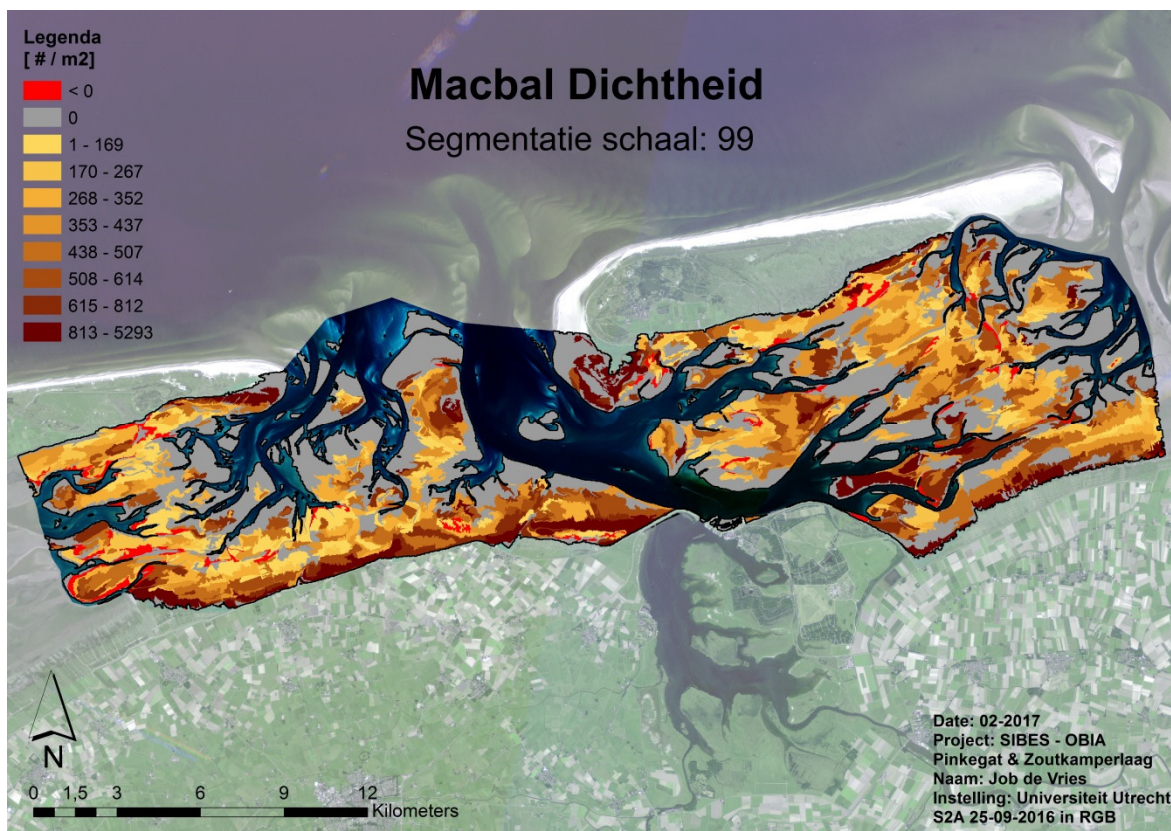
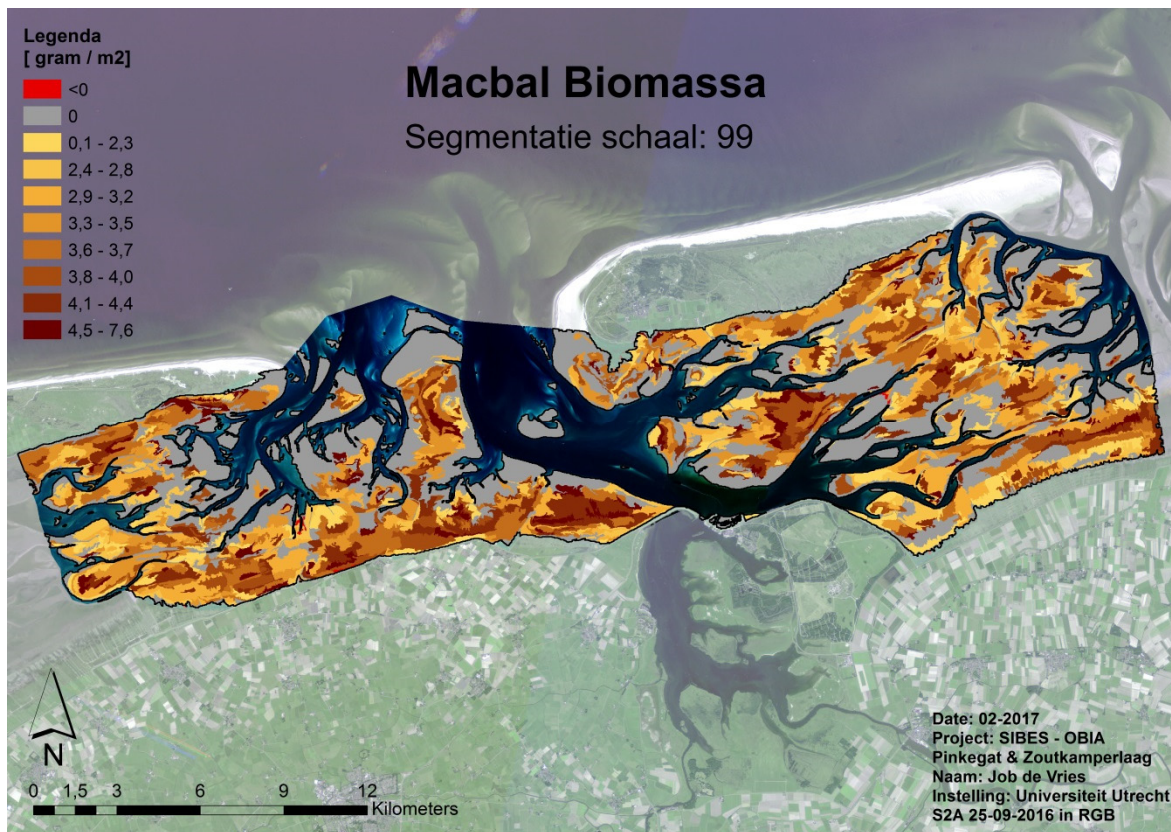
Uiteindelijk zijn alle soorten op een eigen optimaal schaalniveau voorspeld, variërend van 25 tot 200 met bijbehorende toename in gemiddeld oppervlak en afname in absoluut aantal van segmenten (tabel 8). De RF-classificatie van soorten voorkomen leidt ertoe dat soorten een dekkingsgraad laten zien in het gebied variërend van 1.2% tot 63% voor de wadpier ("Ensdire") en een borstelworm soort ("Aphmar"). Dit is een direct gevolg van de RF-output met het aantal polygonen geclassificeerd als aanwezig en de bijbehorende gemiddelde objectgrootte van het optimale schaalniveau. Zo is bijvoorbeeld Aphmar in 39% van de polygonen aanwezig, wat overeenkomt met tot een gebiedsdekking van 63% als gevolg van de objectgrootte.

De negatieve waardes zijn polygonen waarvoor een negatieve biomassa of dichtheid is voorspeld. Het gaat het om minder dan 5% van de objecten bij soorten, behalve bij de dichtheid van de wapenworm ("Scoarm"), waarbij het om bijna 21% van de polygonen gaat. Het verschil in het aantal negatieve voorspellingen tussen biomassa en dichtheid van dezelfde soort past goed bij soorten met een lage correlatie tussen beide of een groot verschil in het aantal gebruikte variabelen (tabel 6). Zo heeft Scoarm 54 polygonen met een negatieve voorspelling voor biomassa en 426 voor dichtheid (tabel 8). De correlatie tussen biomassa en dichtheid is 0.65 (tabel 6) met 6 verklarende variabelen voor biomassa en 14 voor dichtheid.

De overeenstemming tussen de observaties en voorspelling varieert van 0.02 voor de dichtheid van Aremar tot 0.77 voor de biomassa van Heddiv. Naarmate de schaalfactor groter wordt zijn er ook meer polygonen die overlappen met meer dan enkele puntobservaties. Hierbij gaat het wel om alle punten, niet specifiek om de puntobservaties waar de betreffende soort is waargenomen.

Tabel 8. Unieke schaalparameters die horen bij de unieke soorten. De voorspelling is uitgesplitst voor biomassa en dichtheid met negatief voorspelde polygonen en de overeenstemming tussen de voorspelling en observaties (R^2). Punt > 1 geeft een indicatie bij hoeveel objecten op het betreffende segmentatieniveau meerdere SIBESpuntobservaties binnen de grenzen vallen, ongeacht de aanwezigheid van de soort. Gem. opp. geeft de gemiddelde objectgrootte aan. Spreiding geeft aan hoe groot het gebied is waar de soort wordt voorspeld.

Soorten		Polygonen / objecten						Spreiding [ha]
Schaal	Variabele	Aanwezig [N]	Negatief [N]	Totaal [N]	R^2 [#]	Punt > 1 [N]	Gem opp. [ha]	
Aphmar 127	Biomassa Dichtheid	678 (39%)	68	1740	0.16	115	1324	1451433 (63%)
Aremar 25	Biomassa Dichtheid	4021 (17%)	846 970	23270	0.10 0.02	15	99	538978 (23.4%)
Capcap 200	Biomassa Dichtheid	170 (20.3%)	5 6	837	0.21 0.17	119	2753	426386 (18.5%)
Ceredu 110	Biomassa Dichtheid	404 (18.6%)	23 14	2167	0.08 0.15	108	1063	812830 (35.3%)
Ensdire 66	Biomassa Dichtheid	43 (1%)	10 4	4893	0.77 0.45	54	471	27859 (1.2%)
Heddiv 29	Biomassa Dichtheid	3865 (21%)	543 783	18410	0.14 0.18	15	125	821162 (35.6%)
Hydulv 19	Biomassa Dichtheid	268 (1%)	156 168	36441	0.58 0.56	9	63	10518 (5%)
Lancon 104	Biomassa Dichtheid	253 (11%)	22 3	2381	0.23 0.08	102	968	297839 (12.9%)
Macbal 99	Biomassa Dichtheid	966 (38 %)	39 105	2563	0.10 0.12	96	899	1346982 (58.4%)
Scoarm 114	Biomassa Dichtheid	1153 (56%)	54 426	2047	0.07 0.28	113	1125	1085434 (47.1%)
Urospe 192	Biomassa Dichtheid	145 (16%)	5 2	885	0.22 0.25	117	2604	481212 (20.9%)
149	Korrelgrootte	-	1	1333	0.57	122	1729	-
171	Slibgehalte	-	316	1069	0.52	120	2156	-
110	Biodiversiteit	-	158	2167	0.26	108	1063	-



*Figuur 7. Voorspelde biomassa (gADW m⁻²) en dichtheid (ind. m⁻²) van het schelpdier *Macoma balthica* (Macbal) met objecten die volgen uit een schaalparameter van 99.*

4. Discussie

De geobserveerde variatie aan verklarende textuur en spectrale variabelen, samen met de unieke schaalniveaus, bevestigt dat een onderscheid in gerealiseerde niches gemaakt kan worden op basis van de hier gekozen variabelen. De verspreiding van de elf geselecteerde soorten werd met een redelijke tot bruikbare nauwkeurigheid van 58 tot 89% voorspeld. Biomassa en dichtheid van de soorten werden aanzienlijk minder nauwkeurig voorspeld met correlatiecoëfficiënten van 0.02 tot 0.77, dwz variërend van ruis tot redelijke nauwkeurigheid.

Ridge Regressie is, in combinatie met Random Forest, een bruikbare methode om tegelijkertijd de verspreiding van soorten, samen met biomassa en dichtheid, te voorspellen op unieke schaalniveaus. De optimale objectgrootte is namelijk afhankelijk van de unieke ruimtelijke spreiding van soorten (Compton et al., 2013). Dit wordt ook hier bevestigd door de variatie in voorspellingsnauwkeurigheid (GCV-waardes, figuur 6) voor de verschillende schaalparameters die objectgrootte bepalen (zie methode). RR minimaliseert de variantie via een groeilimiet (λ) voor de gekozen variabelen en bepaalt zo de optimale objectgroottes.

In eerste instantie werden voor biomassa en dichtheid verschillende schaalparameters gebruikt, geoptimaliseerd voor de desbetreffende variabele. Dit leidde tot de onrealistische situatie dat er op sommige locaties wel biomassa werd voorspeld en geen dichtheid, of vice versa. Hierop is besloten één objectset te gebruiken voor beide soorten die werd geselecteerd door het normaliseren van de GCV-curves voor biomassa en dichtheid en de schaalparameter te kiezen met de laagste som van deze twee GCV-curves. Voor sommige soorten lagen de afzonderlijke schaalparameter-optima van biomassa en dichtheid dicht bij elkaar, zoals bij de Amerikaanse zwaardschede (*Ensis directus*), maar bij andere was dit niet het geval en verslechterde de voorspellingen aanzienlijk. Voordeel was dat voor deze objectset de voorkomenskaart gemaakt kon worden die voor beide variabelen leidend was. Het lijkt essentieel om voor iedere soort voor het voorkomensgebied, biomassa en dichtheid drie afzonderlijke objectsets te definiëren.

Het aantal geselecteerde variabelen varieerde van twee bij de voorspelling van de wadpier (*Arenicola marina*) dichtheid tot 44 voor de dichtheid van Amerikaanse zwaardschede. Deze variabelen hebben we in dit onderzoek bepaald aan de hand van onderlinge correlatiecoëfficiënten tussen aan de ene kant biomassa en dichtheid en aan de andere kant de objectvariabelen. Door het minimaliseren van de gestandaardiseerde minimale GCV waarde is het vervolgens gelukt om voor elke soort één optimaal segmentatieniveau vast te stellen en tegelijkertijd biomassa en dichtheid zo nauwkeurig mogelijk te voorspellen. Resulterende objectgroottes varieerde van gemiddeld 63 ha bij een segmentatieniveau van 19 tot 2753 ha bij een segmentatieniveau van 200. Bij soorten waar een hoge correlatie is gevonden tussen de biomassa en dichtheid zitten de optimale λ , GCV en dus schaalfactor voor biomassa en dichtheid dichtbij elkaar. Optimale minimale schaalfactor voor biomassa en dichtheid liggen bij overige soorten verder van de gebruikte schaalfactor. Hier komt naar voren dat de gestandaardiseerde GCV waardes beïnvloed worden door de correlatie tussen biomassa en dichtheid van de betreffende soort. Waar de gestandaardiseerde GCV-waarde relatief weinig varieert zijn de verschillen in voorspellingsnauwkeurigheid tussen biomassa en dichtheid relatief kleiner.

Het verschil in correlatiecoëfficiënten tussen waargenomen biomassa en dichtheid met de voorspelde waardes bevestigt daarmee dat de gekozen methode niet voor elke soort even goed werkt (tabel 6). Zo was voor een borstelworm soort (*Aphelochaeta marioni*) en de wapenworm (*Scoloplos armiger*) de geobserveerde correlatie identiek met de voorspelde. Voor het wadslakje (*Hydrobia ulvae*) en de wadpier (*Arenicola marina*) waren de waardes vergelijkbaar terwijl voor de Amerikaanse zwaardschede (*Ensis directus*), de zandkokerworm (*Lanice conchilega*) en het nonnetje (*Macoma balthica*) de waardes verschillend waren. Dit houdt in dat de verhouding tussen biomassa en dichtheid bij deze laatste verloren is gegaan bij de voorspelling. Voor de overige soorten is er een uitgebreide (statistische) analyse nodig het verschil te verklaren.

Het verschil lijkt wel (deels) verklaard te kunnen worden door de gekozen verklarende variabelen (tabel 5 en 6). Bij voldoende identieke variabelen in de voorspelling van dichtheid en biomassa blijft het verband tussen biomassa en dichtheid beter behouden. Dit is terug te zien in de gemiddelde reflectiewaarden per groep objectvariabelen; deze vertonen soms tegenstrijdige correlaties bij biomassa en dichtheid van dezelfde soort. Zo is de gemiddelde correlatiecoëfficiënt in de groep gemiddeld (tabel 5) positief voor de voorspelling van dichtheid, voor de voorspelling van biomassa is deze juist negatief. Iets vergelijkbaars treedt op bij de zandkokerworm waar ook de verhouding tussen biomassa en dichtheid verloren gaat bij de afzonderlijke voorspellingen. Bij de Amerikaanse zwaardschede zijn de gemiddelde correlatiecoëfficiënten niet tegengesteld, wel zijn er duidelijke verschillen in de absolute gemiddelde correlatiecoëfficiënten die significant verschillen bij de voorspelling van biomassa en dichtheid. Soorten waarbij deze verhouding niet verloren gaat (*Aphelochaeta marioni* en *Scoloplos armiger*) laten dit niet zien bij de groepen verklarende variabelen. Het lijkt dus van belang dat voldoende verklarende variabelen voor beide voorspellingen gebruikt worden om te voorkomen dat het verband tussen biomassa en dichtheid verdwijnt. Dit suggereert dat de gekozen selectiemethode voor de verklarende variabelen (Pearsons) in de toekomst extra aandacht verdient omdat hij in deze vorm niet optimaal werkt. Daarnaast is het de moeite waard te kijken naar extra textuur en spectrale variabelen van bijvoorbeeld vegetatie-indices.

Ook het aantal negatief voorspelde polygonen verschilt sterk per soort, maar ook zeker voor biomassa en dichtheid. Ondanks het gebruik van RF om de spreiding in het gebied vast te stellen aan de volledige set aan object variabelen wordt bij iets minder dan 5% van de polygonen een negatieve waarde voorspeld. De voorspelling van de dichtheid van de wapenworm valt hier op als uitzondering met een relatief hoog aantal negatief voorspelde polygonen (22%, 456 van de 2047 polygonen). Wanneer het segmentatieniveau optimaal is voor alleen biomassa of dichtheid, neemt het aantal negatief voorspelde polygonen toe.

In de gekozen aanpak is het model gevalideerd met interne validatiematen zoals de OOB-error, de GCV, de RF nauwkeurigheid, vergelijking met puntobservaties, verschil in overeenstemming biomassa en dichtheid en het aantal negatief voorspelde polygonen. Hieruit blijkt dat de gekozen methode logische spreidingspatronen van de soorten laat zien en tegelijkertijd voor een aantal soorten nauwkeurige voorspellingen van biomassa en dichtheid geeft. De voorspellingskracht van de methode lijkt af te nemen, wanneer het aantal geselecteerde variabelen hoog is ten opzichte van het aantal waarnemingspunten. Het lijkt belangrijk om een goede variabelenselectie uit te overne om mogelijk overfitten te voorkomen.

5. Conclusies en Aanbevelingen

Tijdens dit onderzoek hebben we geprobeerd *proof of concept* te leveren dat van verschillende soorten de biomassa en dichtheid op basis van textuur- en spectrale informatie voorspeld kunnen worden op, voor soorten, unieke segmentatieniveaus. Dit zou ertoe kunnen leiden dat verspreiding, dichtheid en biomassa van soorten nauwkeuriger, vaker en zonder tussenkomst van indirecte relaties met omgevingsvariabelen, geschat kunnen worden aan de hand van satellietbeelden (bijvoorbeeld Sentinel2).

Beeldsegmentatie op basis van de homogeniteit binnen objecten kan hiervoor goed gebruikt worden omdat resulterende objecten gekoppeld kunnen worden aan o.a. textuur en spectrale variabelen. Daarnaast zorgt de aggregatie van individuele pixels tot objecten ervoor dat ruis verminderd wordt en daarmee verbeteringen in classificaties en voorspelling mogelijk maakt. Het gebruik van RandomForest classificatie zorgt ervoor dat er rekening wordt gehouden met observaties van afwezigheid van de betreffende soorten.

De distributiekaarten (aan- / afwezigheid) zijn redelijk betrouwbaar gebleken en variëren van 58 tot 89% overeenstemming tussen waarnemingen en voorspellingen. De voorspellingen van biomassa en dichtheid zijn minder succesvol. De biomassa van de Amerikaanse zwaardschede (*Ensis directus*) wordt het best voorspeld met een betrouwbaarheid van 77%. Mediane korrelgrootte en slibgehalte volgen met percentages van respectievelijk 57% en 52%. Voor de andere soorten en eigenschappen zijn de biomassa- en dichtheidsresultaten nog niet bruikbaar.

Belangrijke conclusies met betrekking tot de methode zijn

- 1) dat per soort aparte optimale objectsets voor distributie, biomassa en dichtheid worden vastgesteld. In het verkennend onderzoek is een grote fout geïntroduceerd bij het selecteren van één schaalparameter voor biomassa en dichtheid samen. De verwachting is dat een flinke verbetering gemaakt kan worden door eerst een distributiekaart te maken en dan voor biomassa en dichtheid apart de optimale objectset te bepalen binnen het voorkomingsgebied van iedere soort.
- 2) dat het maximum aantal variabelen dat wordt gebruikt binnen de Ridge Regressie wordt gerelateerd aan het aantal waarnemingen van een soort zodat er een realistisch aantal vrijheidsgraden blijft bestaan.
- 3) dat het cruciaal, maar moeilijk, is om relevante onafhankelijke variabelen te selecteren uit de lange lijst van potentiële variabelen.

Literatuur

- Addink, E. A., De Jong, S. M., & Pebesma, E. J. (2007). The importance of scale in object-based mapping of vegetation parameters with hyperspectral imagery. *Photogrammetric Engineering & Remote Sensing*, 73, 905-912.
- Addink, E. A., Van Coillie, F. M. B., & De Jong, S. M. (2012). Introduction to the GEOBIA 2010 special issue: From pixels to geographic objects in remote sensing image analysis. *International Journal of Applied Earth Observation and Geoinformation*, 15, 1-6.
- Adolph, W., Schückerl, U., Soo Son, C., Jung, R., Bartholomä, A., Ehlers, M., Kröncke, I., Lehner, S., & Farke, H. (2017) Monitoring spatiotemporal trends in intertidal bedforms of the German Wadden Sea in 2009–2015 with TerraSAR-X, including links with sediments and benthic macrofaunal. *Geo-Mar Lett*, 37, 79–91.
- Aguilar, M. A., Aguilar, F. J., Lorca, A. G., Guirado, E., Betlej, M., Cichon, P., A. Nemmaoui, A., Vallario, A., & Parente, C. (2016). Assessment of multiresolution segmentation for extracting greenhouses from Worldview-2 imagery. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 41.
- Anders, N. S., Seijmonsbergen, A. C., & Bouten, W. (2011). Segmentation optimization and stratified object-based analysis for semi-automated geomorphological mapping. *Remote Sensing of Environment*, 115(12), 2976-2985.
- Bijleveld, A. I., van Gils, J. A., van der Meer, J., Dekinga, A., Kraan, C., van der Veer, H. W., & Piersma, T. (2012). Designing a benthic monitoring program with multiple conflicting objectives. *Methods in Ecology and Evolution*, 3(3), 526-536.
- Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1), 2-16.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Choi, J. K., Oh, H. J., Koo, B. J., Lee, S., & Ryu, J. H. (2011a). Macro benthos habitat mapping in a tidal flat using remotely sensed data and a GIS-based probabilistic model. *Marine pollution bulletin*, 62(3), 564-572.
- Choi, J. K., Oh, H. J., Koo, B. J., Ryu, J. H., & Lee, S. (2011b). Crustacean habitat potential mapping in a tidal flat using remote sensing and GIS. *Ecological modelling*, 222(8), 1522-1533.
- Clausi, D. A. (2002). An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of remote sensing*, 28(1), 45-62.
- Compton, T. J., Holthuijsen, S., Koolhaas, A., Dekinga, A., ten Horn, J., Smith, J., Galama, Y., Brugge, M., van der Wal, D., van der Meer, J., van der Veer, H. W., Piersma, T. (2013). Distinctly variable mudscapes: distribution gradients of intertidal macrofauna across the Dutch Wadden Sea. *Journal of Sea Research*, 82, 103-116.
- Definiens, 2003. eCognition, *Object-based Image Analysis*, München.

- Dijkstra, T. K. (2014). Ridge regression and its degrees of freedom. *Quality & Quantity*, 48(6), 3185-3193.
- Duro, D. C., Franklin, S. E., & Dubé, M. G. (2012). Multi-scale object-based image analysis and feature selection of multi-sensor earth observation imagery using random forests. *International Journal of Remote Sensing*, 33(14), 4502-4526.
- Drăguț, L., Tiede, D., & Levick, S. R. (2010). ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *International Journal of Geographical Information Science*, 24(6), 859-871.
- Drent J., Bijkerk R., Herlyn M., Grotjahn M., Voß J., Carausu M.-C. & Thieltges D.W. (2017) Macrozoobenthos. In: Wadden Sea Quality Status Report 2017. Eds.: Kloepper S. et al., Common Wadden Sea Secretariat, Wilhelmshaven, Germany.
- Dronova, I. (2015). Object-based image analysis in wetland research: A review. *Remote Sensing*, 7(5), 6380-6413.
- Haralick, R.M., Shanmugam, K. & Dinstein, I. (1973). Textural features for image classification. *IEEE Transaction on Systems, Man and Cybernetics*, SMC-3(6) 610-621.
- Kraan, C., van der Meer, J., Dekinga, A., & Piersma, T. (2009). Patchiness of microbenthic invertebrates in homogenized intertidal habitats: hidden spatial structure at a landscape scale. *Marine Ecology Progress Series*, 383, 211-224.
- Kraan C, Dormann CF, Greenfield BL, Thrush SF (2015) Cross-Scale Variation in Biodiversity-Environment Links Illustrated by Coastal Sandflat Communities. *PLoS ONE* 10(11): e0142411. doi:10.1371/journal.pone.0142411
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.
- Lyashevskaya, O., Brus, D. J., & Van der Meer, J. (2016). Mapping species abundance by a spatial zero-inflated Poisson model: a case study in the Wadden Sea, the Netherlands. *Ecology and evolution*, 6(2), 532-543.
- McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International journal of remote sensing*, 17(7), 1425-1432.
- Van der Wal, D., Herman, P. M. J., Forster, R. M., Ysebaert, T., Rossi, F., Knaeps, E., Plancke, Y.M.G., & Ides, S. J. (2008). Distribution and dynamics of intertidal macro benthos predicted from remote sensing: response to microphytobenthos and environment. *Marine Ecology Progress Series*, 367, 57-72.
- Xu, H. (2006). Modification of normalized difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International journal of remote sensing*, 27(14), 3025-3033.
- Ysebaert, T., de Mesel, I. & Herman, P. (2008) Kaderrichtlijn water : achtergronddocument zoute macrofauna 2008. IMARES Wageningen.

