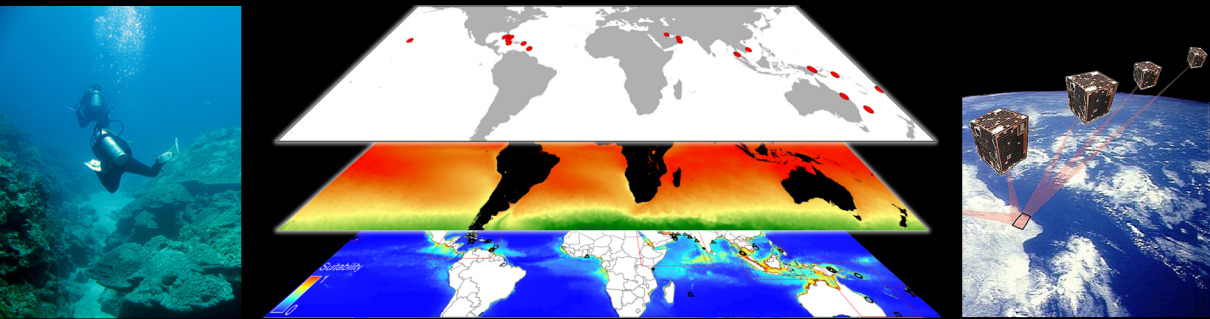


# GIS-BASED ENVIRONMENTAL ANALYSIS, REMOTE SENSING AND NICHE MODELING OF SEAWEED COMMUNITIES

Klaas Pauly





# GIS-based environmental analysis, remote sensing and niche modeling of seaweed communities

**Klaas Pauly**

Promotor: Prof. Dr. Olivier De Clerck

Thesis submitted in partial fulfillment  
of the requirements for the degree of  
Doctor (PhD) in Sciences (Biology)

Academic year 2010-2011

Research Group



Phycology



**Promotor**

Prof. Dr. Olivier De Clerck

**Members of the reading committee**

Prof. Dr. Farid Dahdouh-Guebas (Université Libre de Bruxelles)

Prof. Dr. Tom Schils (University of Guam)

Dr. Els Verfaillie (Ghent University)

**Other members of the exam committee**

Prof. Dr. Dominique Adriaens (Ghent University,  
Chairman at the pre-defense)

Prof. Dr. Paul Goetghebeur (Ghent University,  
Chairman at the public defense)

Prof. Dr. Eric Coppejans (Ghent University)

Prof. Dr. Rudi Goossens (Ghent University)

Prof. Dr. Ann Vanreusel (Ghent University)





*“Such was the region the Nautilus was now visiting, a perfect meadow, a close carpet of seaweed, fucus, and tropical berries, so thick and so compact that the stem of a vessel could hardly tear its way through it.”* – Jules Verne, *Twenty Thousand Leagues under the Sea*

*“What a magnificent spectacle was then outspread beneath the gaze of the travellers! The island of Zanzibar could be seen in its entire extent, marked out by its deeper color upon a vast planisphere; the fields had the appearance of patterns of different colors, and thick clumps of green indicated the groves and thickets.”* – Jules Verne, *Five Weeks in a Balloon*

Two quotes from a great visionary, perfectly illustrating my two passions: the biology of the seas, and looking at this from a bird’s eye perspective. I’ve been lucky enough to combine these two passions into one study during the past 6 years, and obviously this could not have been possible without the support from many, many people.

A complete stranger to seaweeds prior to college, I developed a keen interest in these often disregarded organisms during my first year in Biology, unmistakably initiated by Eric. Later, the predominantly marine-orientated courses convinced me to study zoology. However, Eric and Tom made me enthusiastic for the marine biogeographical and ecological research they were doing, and, although rather unheard of at that time in the botany department, welcomed me as a zoologist for my master’s thesis in the Phycology Research Group. And it didn’t stop there – thanking them for their open-minded research vision and giving me plenty of time to develop my own “niche” in seaweed research thereafter by hybridizing with geography would be an understatement anyhow. Likewise, Olivier supported me in every possible way when he took over as my supervisor, and, probably even more importantly, kept me focused and on track! Moreover, the opportunity they gave me to go on many a sampling trip and congress is simply invaluable. And speaking of invaluable, Christelle always brilliantly solved any administrative obstacles to these trips with a big smile! The Phycology group wouldn’t be complete without the technical help and countless fruitful and animated discussions (not just on phycology!) with my office buddies Heroen and first Tom, later Lennert (with our different backgrounds and interests we made for a very productive team I might say!), as well as Frederik, Ellen, Kenny, Aga, Ana, Frederique, Dioli, Sofie and Caroline, occasionally expanded during meetings, symposiums and the likes. I’d also like to thank the members of my PhD committee as well as anonymous article reviewers for their constructive remarks.

Our research group may have been small, the many less phycology-related discussions were cheerfully joined by the protistologists Bart, Katrijn, Lander, Griet, Jeroen, Pieter, Annick, Dagmar, Annelies, Ines, Caroline, Nicolas, Els, Evelien, Jeroen, Elie, Koen, Wim, and so on, sharing our corridor.

Down the corridor and downstairs, I found priceless technical advice to kick-start the geographical part of my research provided by Rudi,

Tony and Peter, and lately also by Geert from the Archaeology Department. Together with the people from the ground floor marine lab, I kept a broad look on the coastal environment and enjoyed many trips to Wimereux with Joke, Jelle, Jan, Karen, Katja, Nele, Marijn, Ulrike and Ann. Many people also greatly helped me out with advice, among others Bea, Dirk and Jürgen.

As I spent about a year combined abroad on sampling trips and field courses, many people were indispensable for assistance, support, help and entertainment, among which fellow students from Roscoff and Villefranche and the students who volunteered, or sacrificed, for a master's thesis under my auspices. My elite field crew was further populated by Yoshi in Japan and the Omani buddies: Five Oceans' Simon, Rob, Kris, Iain and Oli, as well as Barry and Ligaya who made Oman almost a second home for me, "Toxic" Tom and Else, Pieter, Michel, Ali, Mussallem and Toyota Land Cruiser 4500.

Throughout the years, fellow students and colleagues became friends, and living in a joyful city as Ghent is also guaranteed to get you surrounded by friends who made work and life a pure pleasure. Next to people I have already mentioned stand, in no particular order, Wolf, Pieter, Griet, Mieke, Wouter, Matthias, Maureen, Martine, Sarah, Eva, ..., the underwater hockey gang and every single Intimate Voice, as well as the Didakites and Antwerp Flyer crew who cheered up the coastal life.

Like I said in the beginning, many, many people contributed to this work one way or another. Not being mentioned in name here is a mere honor because you have made my life and work so much better without me even realizing how!

On top of that, next to all the years of unconditional support, my twin-brother-lifeguard-musician Koen and brother-expert-on-road-trip-music Jan did an amazing job as field biologists in Oman after a heroic crash course in snorkeling, unauthorized diving, make-do Arabic, better-than-me-4WD-driving and, indeed, phycology. Eternal Respect to you, and Soetkin and Delfine whom I made temporarily husbandless! I can guarantee you a very important role in Peetje's tough stories to your lovely kids(-to-be)!

Het mag duidelijk zijn dat niets van dit hele werk mogelijk geweest zou zijn zonder de immer liefdevolle steun van Moeke, Moem en Vake, net als die van Jef en Gerda. Niemand van jullie spaarde ooit kost of moeite, bloed, zweet of tranen om mij te laten verwezenlijken waarin ik geloofde. Een warm nest in de zeebries, en een warm nest in de kempen, wie zou zich meer kunnen wensen...

An, she-who-had-to-share-her-boyfriend-with-his-"mistress"/PhD, to you I dedicate this work. You made me enjoy all-things-non-biology, laugh and dance, while enjoying yourself all-things-biology I did all along. And as if that wouldn't do, without you both this classy book and its predecessor would look like crappy street paper! I couldn't dream of better!

# TABLE OF CONTENTS

<b>Acronyms</b>	<b>1</b>
<b>Summary</b>	<b>3</b>
<b>Samenvatting</b>	<b>7</b>
<b>Chapter 1</b>	<b>13</b>
<b>General Introduction</b>	
GIS and remote sensing in a nori wrap	14
Introduction	14
The (r)evolution of spatial information	14
Spatial data types	17
GIS and remote sensing: phycological applications	18
Georeferencing specimens	18
Remote sensing	20
Distribution and niche modeling	26
Future directions and research priorities	33
The quest for spatial data in seaweed biology	33
Georeferencing specimens	35
Remote sensing	36
Distribution and niche modeling	38
Aims and outline	41
Author contributions	46
<b>Chapter 2</b>	<b>47</b>
<b>Modeling the distribution and ecology of <i>Trichosolen</i> blooms on coral reefs worldwide</b>	
Abstract	48
Introduction	49
Materials and methods	51
Species occurrence data	51
Environmental data	52
ENM	53
Model evaluation	55
Results	56
<i>Trichosolen</i> potential distribution	56
Environmental response curves for <i>Trichosolen</i> on coral	57
Discussion	59
Supplementary material	64

<b>Chapter 3</b>	<b>67</b>
<b>Macroecology meets macroevolution: Evolutionary niche dynamics in the seaweed <i>Halimeda</i></b>	
Abstract	68
Introduction	69
Materials and methods	71
Species identification	71
Preprocessing observation data	72
Species phylogeny	72
Macroecological data	73
Evolutionary analysis of niche characteristics	74
Niche modeling procedure	76
Results	77
Species delimitation and phylogeny	77
Evolution of niche characteristics	78
Niche models at the global scale	80
Niche models at the regional scale	82
Discussion	83
Modeling seaweed distributions	83
Niche modeling vs. previous approaches	83
Taxonomic caveat	83
Macroevolution of the macroecological niche	84
Historical perspective	84
Niche conservatism	84
Sources of uncertainty	85
Paleobiological perspective	86
Global biogeography	87
Dispersal limitation	87
Vicariance patterns	88
Regional biogeography of tropical America	89
Supplementary material	91

<b>Chapter 4</b>	<b>97</b>
<b>Bio-ORACLE: a global environmental dataset for marine species distribution modeling</b>	
Abstract	98
Introduction	99
Materials and methods	100
Remotely sensed data	100
<i>In situ</i> measured oceanographic data	102
Pre-processing and multivariate analysis of Bio-ORACLE rasters	103
Case study: <i>Codium fragile</i> ssp. <i>fragile</i>	104
Results	105
Case study: <i>Codium fragile</i> ssp. <i>fragile</i>	105
Discussion	107
Data quality	107
Utility for marine SDM	109
Using Bio-ORACLE for marine SDM	110
Comparison to other marine environmental datasets	111
Conclusions and perspectives	113
Supplementary material	115
<b>Chapter 5</b>	<b>125</b>
<b>Spatial scale-dependent prediction in marine distribution modeling: a case study</b>	
Abstract	126
Introduction	127
Material and methods	129
Biotic data	129
Environmental data	129
Landsat data	129
Bio-ORACLE data	132
Distribution modeling	133
Data exploration & model analysis	134
Results	134
Discussion	138
Evaluation of Landsat-based SDM	138
Technical issues relating to Landsat-based SDM	141
Influence of scale on marine SDM	143
Supplementary material	145

**Chapter 6** **149**  
**Mapping coral-algal dynamics in a seasonal upwelling area  
using spaceborne high resolution sensors**

Abstract	150
Introduction	151
Material and methods	153
Study Area and Field Work	153
Remote Sensing Datasets	153
Image Preprocessing	154
PROBA/CHRIS	154
Landsat 7 ETM+	155
Image Processing	156
Supervised bottom-type classification	156
Spectral indices	157
Maxent sub-pixel modeling	158
Results and discussion	160
Supervised bottom-type classification	160
Spectral index maps	163
Maxent sub-pixel modeling	166
Conclusion	168

**Chapter 7** **169**  
**Low-cost very high resolution intertidal vegetation monitoring  
enabled by near-infrared kite aerial photography**

Abstract	170
Introduction	171
Material and methods	172
Study area	172
Kite aerial photography	173
Ground truthing	174
Image mosaicing and processing	175
Results	178
Ground truthing	178
RGB image classification	178
NDVI	179
False-color and NIR-enhanced classifications (2011 only)	182
Discussion	183
Quality of the presented data	183
Utility of KAP in monitoring and best practice	184

<b>Chapter 8</b>	<b>187</b>
<b>General discussion</b>	
Spatial information in GIS-based macroalgal studies	188
General aspects of modeling macroalgal distributions	189
Investigating environmental data integration and effects on modeling	195
Multi-sensor mapping	197
Conclusion and perspectives	199
<b>References</b>	<b>201</b>
<b>Curriculum Vitae</b>	<b>219</b>





## ACRONYMS

AUC	Area Under the Curve
AVHRR	Advanced Very High Resolution Radiometer
BEM	Bioclimate Envelope Modeling
Bio-ORACLE	Ocean Rasters for Analysis of Climate and Environment
BNDVI	Blue-substituted Normalized Difference Vegetation Index
BRT	Boosted Regression Trees
CAL	Calcite
CCD	Charge-Coupled Device
CHDK	Canon Hacker Development Kit
CHL(-a)	Chlorophyll(-a)
CHRIS	Compact High Resolution Imaging Spectrometer
DA	Diffuse Attenuation
DIVA	Data Interpolating Variational Analysis
DN	Digital Number
DOM	Dissolved Organic Matter
EEZ	Exclusive Economic Zone
EGNOS	European Geostationary Navigation Overlay
ENFA	Ecological Niche Factor Analysis
ENM	Ecological Niche Modeling
ETM+	Enhanced Thematic Mapper +
FAI	Floating Algae Index
GAD	Generalized Additive Models
GARP	Genetic Algorithm for Rule set Prediction
GCP	Ground Control Point
GEBCO	GEneral Bathymetric Chart of the Oceans
GIS	Geographic Information System(s)
GLM	Generalized Linear Models
GPS	Global Positioning System
GSD	Ground Sampling Distance
HSM	Habitat Suitability Modeling
IPCC	Intergovernmental Panel on Climate Change
ITS	Internal Transcribed Spacer
KAP	Kite Aerial Photography
KIA	Kappa Index of Agreement
LIDAR	LIght Detection And Ranging <i>or</i> Laser Imaging Detection And Ranging
Maxent	Maximum entropy
MCMC	Markov Chain Monte Carlo
MERIS	MEdium Resolution Imaging Spectrometer
ML	Maximum Likelihood
MODIS	MODERate resolution Imaging Spectroradiometer
MPB	MicroPhytoBenthos
NDVI	Normalized Difference Vegetation Index
NIR	Near-InfraRed
NOAA	National Oceanic and Atmospheric Administration
OBIS	Ocean Biogeographic Information System

PAR	Photosynthetically Available Radiation <i>or</i> Photosynthetically Active Radiation
PCA	Principal Component Analysis
PIC	Particulate Inorganic Carbon
POC	Particulate Organic Carbon
PROBA	PRoject for OnBoard Autonomy
RGB	Red Green Blue ( <i>also used separately</i> )
RMS(E)	Root Mean Square (Error)
ROC	Receiver Operating Characteristic
ROI	Region Of Interest
SAL	Salinity
SD	Standard Deviation
SDM	Species Distribution Modeling
SFM	Structure From Motion
SeaWiFS	Sea-viewing Wide Field-of-view Sensor
SLC	Scan Line Corrector
SMOS	Soil Moisture and Ocean Salinity
SST	Sea Surface Temperature
SSU	Small SubUnit
SWIR	Short-Wave InfraRed
TIR	Thermal InfraRed
TOA	Top Of Atmosphere
UAV	Unmanned Aerial Vehicle
UTM	Universal Transverse Mercator
VIM	Variational Inverse Method
VNIR	Visible and Near-InfraRed
WAAS	Wide Area Augmentation System
WOA	World Ocean Atlas
WOD	World Ocean Database
WOOD	Worldwide Ocean Optics Database

## SUMMARY

Marine habitats and environments are under increasing pressure worldwide. While human activities have long been demonstrated to impact coastal and pelagic communities, these effects will probably be aggravated by global change. It is therefore important to gain insight in spatial and temporal dynamics and patterns of primary producers such as macroalgae. However, spatial data suitable for seaweed research have only recently become available, and marine applications in general and phycological in particular have lagged behind on terrestrial applications in spatially explicit data acquisition and processing. This thesis aims at presenting case studies to identify current issues in GIS-based analysis, remote sensing and distribution modeling of algae, while suggesting workarounds, adaptations or solutions to these specific techniques to help answering biological questions.

In **chapter 1**, the concepts of GIS and remote sensing are introduced to the phycological community, and the current state of spatial data integration in phycological research is reviewed. The storage and dissemination of spatial metadata is of particular concern, while this can be remedied with minimal effort to reveal valuable information for distribution modeling studies. Promising satellite sensors have been launched or are planned to reveal coarse scale environmental information which can be used as input in distribution modeling, but satellites suitable for high resolution seaweed mapping remain underrepresented.

In **chapter 2**, following four ephemeral monotypic green algal *Trichosolen* blooms on catastrophically impacted coral reefs, from which the species was previously unknown, we use the records together with few other occurrence data in a niche modeling study. We aim to derive ecological and distributional information and delineate future bloom risk areas by overlaying the suitability map with a GIS coral reef database. A pantropical distribution was modeled, with several reef areas where the alga has not yet been recorded from delineated as bloom risk area. Response curves for chlorophyll differed markedly between blooms and non-blooms. While both blooms and non-bloom occurrences showed a strong preference for high

temperatures, blooms responded better to broader nutrient ranges than non-blooms.

In **chapter 3**, we use occurrence records and their associated sea surface temperature and nutrient (chlorophyll) values of extant *Halimeda* spp. as input in ancestral state reconstruction techniques to infer their ancestral niche and the degree of niche conservatism. We also perform niche modeling to compare the potential niche to the known distributions. Results showed that the niche of *Halimeda* is conserved for tropical, nutrient-depleted habitats, while one section of the genus invaded colder waters several times independently. Since known distribution ranges are considerably smaller than modeled potential ranges, we conclude that restricted geographical ranges are likely the result of dispersal limitation. We propose suitability hotspots in adjacent ocean basins to be targeted for fieldwork to discover sibling species.

**Chapter 4** aims to boost marine distribution modeling applications by providing Bio-ORACLE, the first global marine pre-packaged uniform set of environmental variables at 9km resolution, representing different dimensions in environmental space, freely available for download. Twenty-three raster layers were constructed based on remote sensing data products and interpolation of *in situ* measurements, for which a uniform landmask was applied. We performed a modeling test on the invasive green alga *Codium fragile* subsp. *fragile* study which shows the predictive performance of the dataset by correctly predicting the current distribution.

In **chapter 5**, a high-resolution environmental dataset for regional distribution modeling is assembled using Landsat 7 ETM+ imagery. Habitat layers relating to sea surface temperature, nutrient availability and turbidity as well as a substrate layer were based on 10 mosaiced scenes to cover a 2000km stretch along the coast of Oman at 60m resolution. Niche models for 3 species occurring in the Arabian Sea, Gulf of Oman or both were modeled based on Landsat-derived variables as well as cropped Bio-ORACLE variables equivalent to the Landsat information. It emerged that for all three species, Bio-ORACLE and Landsat-based predictions were similar for the Gulf of Oman with relatively little environmental variability.

By contrast, Bio-ORACLE showed overprediction for all three species in the more heterogeneous Arabian Sea, with Landsat models closely reflecting previously known distributions for the three species in the latter region. Hence, we conclude that spatial scale effects in heterogeneous marine environments should be considered when designing regional distribution modeling studies.

In **chapter 6**, multispectral Landsat 7 ETM+ and superspectral PROBA/CHRIS imagery, both high resolution at 30m and 18m respectively, are used in a seasonal mapping effort focusing on macroalgal stands in an upwelling-exposed area in the south of Oman. While Landsat 7 benefits from a SWIR band in a vegetation index to detect surfacing, floating and intertidal algae, classification accuracy was better for PROBA/CHRIS. Classification accuracy was also generally higher in the winter compared to the summer monsoon season, where turbid waters and dense algal stands cause bottom types to be very heterogeneous. Nonetheless, an important turnover between coral dominance in winter and algal dominance in summer was detected. Maxent models for brown algae and coral in the summer monsoon revealed extensive sub-pixel overgrowth of algae on coral, dominating the spectral signature of these mixed pixels.

**Chapter 7** presents a near-infrared enabled consumer-grade digital compact camera mounted on a kite as an effective low-cost monitoring tool for intertidal habitats, with a focus on seaweed mapping. Using computer vision software, photos were automatically aligned and a 3D terrain reconstruction was done in order to generate an orthophotomosaic. By using a red-blocking filter in front of the lens, the camera captured blue, green and near-infrared light only, and these band mosaics were used to create a segmentation image. The latter was used to aid pixel training for maximum likelihood classification. A modified normalized difference vegetation index was also calculated, substituting red by blue. We showed that a combination of blue, green, near-infrared and vegetation layers as input for the classification yielded superior results to true-color image processing. While not equaling the quality of professional UAV system-based aerial photography, this monitoring tool seems ideally suited for use in remote areas, adverse physical conditions and developing countries.

## Summary

Based on these case studies, we presented several tools and techniques that enhance distribution modeling and mapping power for seaweed applications. This makes a growing body of spatially explicit information and processing algorithms suited for the marine environment available to answer biological questions on seaweed patterns in time and space. However, some issues would benefit from further research, such as more widely accepted model selection and evaluation tools and knowledge on model parameter settings.

## SAMENVATTING

Mariene habitats en milieus staan wereldwijd onder toenemende druk. Van menselijke activiteiten is reeds lang aangetoond dat ze een impact hebben op kust- en zee gemeenschappen, en deze effecten zullen waarschijnlijk nog versterkt worden door global change. Om hierop in te spelen is het belangrijk een inzicht te verwerven in de dynamiek en patronen in ruimte en tijd van primaire producenten zoals zeewieren. Ruimtelijke gegevens die geschikt zijn voor onderzoek naar zeewieren zijn echter nog maar sinds kort beschikbaar, en mariene toepassingen in het algemeen en op wieren in het bijzonder hinken achterop in vergelijking met terrestrische toepassingen in ruimtelijke dataverwerking en –verwerking. Deze thesis heeft tot doel casestudies naar voor te brengen om de huidige moeilijkheden in GIS-gebaseerde analyses, teledetectie en verspreidingsmodellering van wieren aan te tonen, om dan aanpassingen en oplossingen voor deze technieken aan te reiken en om biologische vraagstukken te helpen beantwoorden.

In **hoofdstuk 1** worden de concepten van geografische informatiesystemen (GIS) en teledetectie voorgesteld aan de algologische gemeenschap, en wordt de huidige staat van ruimtelijke data-integratie in algologisch onderzoek nagegaan. Vooral de opslag en verspreiding van ruimtelijke metadata van waarnemingen van zeewieren lijkt een probleem te vormen, terwijl daar voor de hand liggende oplossingen voor uitgewerkt kunnen worden zodat deze kostbare informatie kan aangewend worden in studies naar verspreidingsmodellering. Een aantal van de net gelanceerde of geplande satellieten zijn veelbelovend omwille van de omgevingsinformatie op globale schaal die ze beschikbaar gaan stellen, hetgeen gebruikt kan worden voor verspreidingsmodellering, maar satellieten die geschikt zijn om zeewieren in kaart te brengen met een hoge resolutie zijn nog steeds schaars.

In **hoofdstuk 2** gebruiken we de weinige beschikbare gegevens over het voorkomen van het groenwier *Trichosolen* om de niche ervan te modelleren, nadat het in vier gevallen een snelle, kortstondige bloei had gevormd op koraalriffen die net beschadigd waren. We willen hiermee de ecologische karakteristieken en de verspreiding van het wier achterhalen, en risicogebieden voor toekomstige bloeien gaan afbakenen door de



resulterende geschiktheidswaarden in een GIS over een ruimtelijke database van koraalriffen te leggen. Het model toonde dat het zeewier een verspreiding over de volledige tropen kent, waarbinnen verschillende koraalriffen van waar het wier nog niet gerapporteerd werd als bloeirisicogebied afgelijnd werden. De gemodelleerde responscurven van *Trichosolen* waren vooral verschillend voor wat betreft chlorofylconcentraties (en dus organische nutriënten) in de waterkolom. Waarnemingen van zowel bloeigevallen als niet-bloeiende populaties toonden een sterke voorkeur voor hoge temperaturen, maar het is waarschijnlijker bloeien aan te treffen in gebieden met ver uiteenlopende nutriëntenconcentraties.

In **hoofdstuk 3** gebruiken we gegevens over het voorkomen en de geassocieerde oppervlaktetemperatuur van het zeewater en de nutriëntentoestand (a.d.h. van chlorofyl) van de huidige soorten binnen het groenwiergenus *Halimeda* om als input te dienen voor de reconstructie van de toestand in het verleden. Zo kan de ancestrale niche bepaald worden en nagegaan worden of deze niche constant bleef in de tijd (nicheconservatisme) of niet. We maken ook nichemodellen om de potentiële niche van de huidige soorten te vergelijken met hun werkelijke (gekende) verspreiding. De resultaten toonden aan dat de niche van *Halimeda* beperkt is tot tropisch, nutriëntenarm water, maar dat binnen één sectie van het genus enkele soorten verschillende malen onafhankelijk van elkaar zich in koude wateren zijn gaan verspreiden. Omdat de gekende verspreidingspatronen aanzienlijk kleiner zijn dan de gemodelleerde potentiële verspreiding, besluiten we dat de beperkte geografische verspreiding van de huidige soorten vooral het gevolg is van dispersiebeperkingen, eerder dan aan een gebrek aan geschikte milieus. We stellen voor dat de hotspots van nichegeschiktheid die in de naburige oceanen te vinden zijn voorrang moeten krijgen om bemonsterd te worden met als doel nieuwe zustersoorten te ontdekken.

**Hoofdstuk 4** heeft tot doel mariene verspreidingsmodellering te vergemakkelijken door Bio-ORACLE voor te stellen: de eerste globale, uniform gebundelde dataset van omgevingsvariabelen op 9km resolutie, die gratis online ter beschikking staat. Deze gegevens omvatten verschillende dimensies in de omgevingsruimte en bestaan uit 23 rasterkaartlagen die

samengesteld zijn op basis van teledetectieproducten en interpolaties van *in situ* metingen, om daar dan op een uniforme manier landpixels in te maskeren. We hebben de dataset getest op de modellering van de verspreiding van *Codium fragile* subsp. *fragile*, een goed bestudeerde invasieve groenwiersoort. De resultaten hiervan toonden de kracht van de dataset aan door het huidige verspreidingsgebied correct te voorspellen.

In **hoofdstuk 5** wordt op basis van beelden van Landsat 7 ETM+ een hoge-resolutie dataset van omgevingsvariabelen samengesteld voor het maken van regionale verspreidingsmodellen. Habitatkaartlagen die verband houden met de oppervlaktetemperatuur van het zeewater, beschikbaarheid van nutriënten, turbiditeit en het substraat werden samengesteld op basis van 10 satellietbeelden die een strook van 2000km langs de kust van Oman dekken met een resolutie van 60m. Nichemodellen voor drie soorten die voorkomen in de Arabische Zee, de Golf van Oman of beide werden opgesteld op basis van zowel de Landsatkaartlagen als gegevens uit de Bio-ORACLE dataset die verkleind werden tot het studiegebied, en waaruit die gegevens gehaald werden die equivalent zijn met de kaartlagen op basis van Landsat. Het bleek dat de verspreiding van de drie soorten even goed gemodelleerd werden op basis van zowel Bio-ORACLE als Landsat in de Golf van Oman die gekenmerkt wordt door een lagere variabiliteit. Anderzijds vertoonden de modellen op basis van Bio-ORACLE voor alle soorten een overpredictie in de meer heterogene Arabische Zee, terwijl op Landsat-gebaseerde modellen daar net wel de gekende verspreidingspatronen goed weergaven. We besluiten daaruit dat de resolutie van omgevingsvariabelen waarmee gewerkt wordt in de modellering een rol speelt waarmee met name in heterogene milieus rekening gehouden moet worden bij het opzetten van modelleringsstudies.

In **hoofdstuk 6** worden hoge-resolutie multispectrale Landsat 7 ETM+ (30m resolutie) en superspectrale PROBA/CHRIS (18m resolutie) beelden gebruikt in een poging zeewiergemeenschappen seizoenaal in kaart te brengen in een gebied dat gekenmerkt wordt door intense opwelling, in het zuiden van Oman. Ondanks het voordeel van de korte-golf infrarood (SWIR) band van Landsat 7 voor gebruik in vegetatie-indices om bovendrijvende en intertidale wieren te herkennen, waren de

classificatieresultaten die behaald werden met PROBA/CHRIS beelden beter. De nauwkeurigheid van classificaties was ook beter in de winter dan in de zomer, wanneer een troebele waterkolom en dichte begroeiing door wieren ervoor zorgen dat de bodembedekking heel heterogeen wordt. Ondanks deze moeilijkheden werd een grote omschakeling gevonden van koraal-dominantie in de winter naar zeewier-dominantie in de zomer. Verspreidingsmodellen voor bruinwieren en koralen in de zomer maakten duidelijk dat er een uitgebreide overgroei van wieren op koralen plaatsvindt op een schaal kleiner dan de resolutie van de pixels; op deze manier domineren wieren het spectrale signaal dat van deze gemengde pixels uitgaat.

**Hoofdstuk 7** stelt een eenvoudige nabij-infrarood-geconverteerde digitale compactcamera, opgehangen aan een windvlieger, voor als een efficiënte manier om intertidale habitats op een goedkope manier te monitoren, met de nadruk op zeewiergemeenschappen. Met behulp van computervisiesoftware werden de foto's automatisch gealigneerd en werd een 3D-reconstructie van het terrein gemaakt, wat vervolgens toeliet een orthofotomozaïek aan te maken. Door een rood-blokkerende filter voor de lens te gebruiken ontving de camera enkel blauw, groen en nabij-infrarood licht, en deze bandmozaïeken werden dan gebruikt om een beeldsegmentatie uit te voeren waarna de hele mozaïek geclassificeerd kon worden. Een aangepaste NDVI vegetatie-index kon ook berekend worden door de rode band te vervangen door de blauwe. We tonen aan dat de combinatie van blauw, groen, nabij-infrarood en de vegetatie-index betere resultaten opleverde voor classificatie in vergelijking met echte kleurenbeelden (blauw, groen, rood). Dit systeem haalt weliswaar niet de kwaliteit van camera's aan boord van professionele onbemande vliegtuigen, maar het is wel geschikt voor zeer afgelegen gebieden, voor gebruik in barre omstandigheden en in ontwikkelingslanden.

Op basis van deze casestudies hebben we verschillende methodes en technieken aangereikt die de verspreidingsmodellering en kartering van zeewieren op punt stellen. Daardoor komt een steeds groeiend deel van ruimtelijk expliciete informatie en verwerkingsmethodes binnen handbereik voor gebruik in het mariene milieu en om biologische vragen over

## Samenvatting

zeewierspreidingspatronen in ruimte en tijd te beantwoorden. Hoe dan ook, sommige problemen blijven onopgelost; zo is er meer onderzoek nodig naar een algemeen aanvaarde vorm van modelselectie en –evaluatie en is een beter begrip van het aanpassen van bepaalde instellingen in het modelleringsproces noodzakelijk.

## Samenvatting

# CHAPTER 1

## GENERAL INTRODUCTION

### **GIS-based environmental analysis, remote sensing and niche modeling of seaweed communities**

Klaas Pauly & Olivier De Clerck

Adapted from the published book chapter: Pauly K & De Clerck O (2010) GIS-based environmental analysis, remote sensing and niche modeling of seaweed communities. In: A Israel, R Einav & J Seckbach (Eds) *Seaweeds and their Role in Globally Changing Environments*. Cellular Origin, Life in Extreme Habitats and Astrobiology **15**, Springer Science & Business Media BV, pp 95-114

## **GIS AND REMOTE SENSING IN A NORI WRAP**

### INTRODUCTION

In the face of global change, spatially explicit studies or meta-analyses of published species data are much needed to understand the impact of the changing environment on living organisms, for instance by modeling and mapping species' distributional shifts. A Nature Editorial (2008) recently discussed the need for spatially explicit biological data, stating that the absence or inaccuracy of geographical coordinates associated to every single sample prohibits, or at least jeopardizes, such studies in any research field. In this chapter, we show how geographic techniques such as remote sensing and applications based on geographic information systems (GIS) are the key to document changes in marine benthic macroalgal communities.

Our aim is to introduce the evolution and basic principles of GIS and remote sensing to the phycological community and demonstrate their application in studies of marine macroalgae. Next, we review current geographical methods and techniques showing specific advantages and difficulties in spatial seaweed analyses. We conclude by demonstrating a remarkable lack of spatial data in seaweed studies to date and hence suggesting research priorities and new applications to gain more insight in global change-related seaweed issues.

### THE (R)EVOLUTION OF SPATIAL INFORMATION

The need to share spatial information in a visual framework resulted in the creation of maps as early as many thousands of years ago. For instance, an approximately 6200 year old fresco map covering the city and a nearby erupting volcano was found in Çatal Höyük, Anatolia (Turkey). Dating back even further, the animals, dots and lines on the Lascaux cave walls (France) are thought to represent animal migration routes and star groups, some 15000 years ago. Throughout written history, there has been a steady increase in both demand for and quality (i.e., the extent and amount of detail) of maps, concurrent with the ability to travel and observe one's position on earth. Like many aspects in written and graphic history, however, a revolutionary expansion took place with the introduction of (personal) computers. This new technology allowed to store maps (or any

graphics) and additional information on certain map features in a digital format using an associated relational database (attribute information). It is important to note that the creation of GIS is not a goal in itself; instead, GIS are tools that facilitate spatial data management and analysis. For instance, a Nori farmer may wonder how to quantify the influence of water quality and boat traffic on the yields (the defined goals), and use GIS as tools to create and store maps and (remotely-sensed) images, and perform spatial analyses to achieve these goals (figure 1).

At least 30000<sup>1</sup> publications dating back to 1972 involve GIS (Amsterdam et al., 1972), according to ISI Web of Knowledge<sup>2</sup>. However, twelve years went by before the first use of GIS in the coastal or marine realm was published (Ader, 1982), and since then only a meager 2257 have followed.

Parallel to the evolution of mapping and GIS, the need to observe objects without being in physical contact with the target, remote sensing, has played an important role in spatial information throughout history. In its earliest forms, it might have involved looking from a cliff to gain an overview of migration routes or cities. However, three revolutions have shaped the modern concept of remote sensing. Halfway the 19<sup>th</sup> century, the development of (balloon) flight and photography allowed to make permanent images at a higher altitude (with the scale depending on the altitude and zoom lens) and at many more times or places than were previously feasible, making remote sensing a valuable data acquisition technique in mapping. Halfway the 20<sup>th</sup> century, satellites were developed for Earth observation, allowing to expand ground coverage. At the end of the 20<sup>th</sup> century, the ability to digitally record images through the use of (multiple) CCD and CMOS sensors quickly enhanced the abilities to import and edit remote sensing data in GIS.

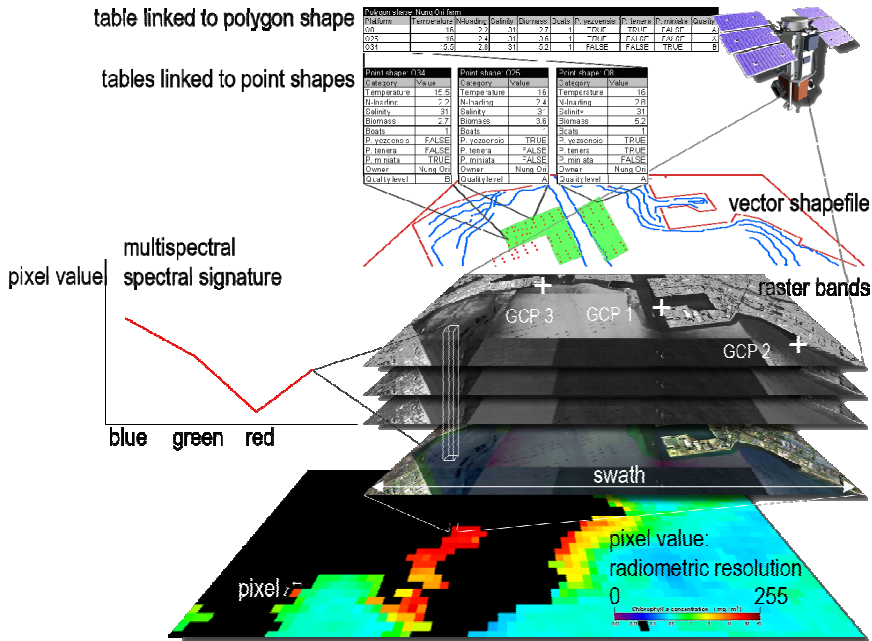
Two kinds of remote sensing have been developed. Active remote sensing involves the emission of signals with known properties, in order to analyze the reflection and backscatter, with RADAR (Radio Detecting And Ranging) as the most wide-spread and best known application.

---

<sup>1</sup> This number is based on the search term "geographic\* information system\*". The search term 'GIS' yielded 32706 records, but an unknown number of these, including the records prior to 1972, concern other meanings of the same acronym.

<sup>2</sup> All online database counts and records mentioned throughout this chapter, including ISI Web of Knowledge, OBIS and Algaebase records, refer to the status on July 1<sup>st</sup>, 2008.





**Figure 1.** Schematic overview of GIS data file types and remote sensing of a Nori farm in Tokyo Bay, Japan.

Passive remote sensing means recording radiation emitted or reflected by distant objects, and most often the reflection of sunlight by objects is investigated. This chapter will only cover passive remote sensing and laser-induced active remote sensing, as sound-based active sensing (RADAR, SONAR) is limited to (3D) geomorphological and topographical studies, rather than distinguishing benthic communities and their relevant oceanographic variables.

The first remote sensing applications are almost a decade older than the first GIS publications (Bailey, 1963), and the first coastal or marine use of remote sensing appeared only few years later, starting with oceanographical applications (Polcyn & Sattinger, 1969; Stang, 1969) and followed by mapping efforts (Egan & Hair, 1971). Out of roughly 98500 remote sensing records in ISI Web of Knowledge, however, little less than 8500 cover coastal or marine topics.

## SPATIAL DATA TYPES

Analogous to manually drawn hard-copy maps, digitized maps (hard-copies transferred to computers) or computer-designed maps most often consist of three types of geometrical features expressed as vectors (figure 1): zero-dimensional points, one-dimensional lines and two-dimensional polygons. For instance, a point could represent a tethered Nori platform in a bay, linked to a database containing quantitative fields (temperature, nutrients, salinity, biomass, number of active harvesting boats), Boolean fields (presence/absence of several species) and categorical fields (owner's name, quality level label). In turn, polygons encompassing several of these points may depict farms, regions or jurisdictions. Lines could either intersect these polygons (in case of isobaths) or border them (in case of coastal structures). Vector maps and associated databases are easy to edit, scale, re-project and query while maintaining a limited file size.

The raster data type (figure 1), also called grid or image data in which all remote sensing data come, differs greatly from vector map data. Each image (whether analogously acquired and subsequently scanned, directly digitally acquired, or computer-generated) is composed of  $x$  columns times  $y$  rows with square pixels (or cells) as the smallest unit. Each pixel is characterized by a certain spatial resolution (the spatial extent of a pixel side), typically ranging from 1m to 1000m, and an intensity ( $z$ -value). The radiometric resolution refers to the number of different intensities distinguished by a sensor, typically ranging from 8 bits (256) to 32 bits ( $4.3 \times 10^9$ ). In modern remote sensing platforms, different parts (called bands) of the incident electromagnetic spectrum are often recorded by different sensors in an array. In this case, a given scene (an image with a given length and width, the latter also termed swath, determined by the focal length and flight altitude) consists of several raster layers with the same resolution and extent, each resulting from a different sensor. The amount of sensors thus determines the spectral resolution. A "vertical" profile of a pixel or group of pixels through the different bands superimposed as layers, results in a spectral signature for the given pixel(s). The spectral signature can thus be visualized as a graph plotting radiometric intensity or pixel value against band number (figure 1). The term multispectral is used for up to ten sensors (bands), while hyperspectral means the presence of ten to hundreds

of sensors. Some authors propose the term superspectral, referring to the presence of 10 to 100 sensors, and reserve hyperspectral for more than 100 sensors. Temporal resolution indicates the coverage of a given site by a satellite in time, i.e. the time between two overpasses. In the Nori farm example, one or more satellite images might be used as background layers in GIS (figure 1) to digitize farms and the surroundings (based on large-scale imagery in a geographic sense, i.e. with a high spatial resolution) or to detect correlations with sites and oceanographical conditions (based on small-scale imagery in a geographic sense, i.e. with a low spatial resolution).

An important aspect in GIS and remote sensing is georeferencing. By indicating a limited number of tie points or ground control points (GCPs) for which geographical coordinates have been measured in the field or for which coordinates are known by the use of maps, coordinates for any location on a computer-loaded map can be calculated in seconds and subsequently instantly displayed. Almost coincidentally with GIS evolution, portable satellite-based navigation devices (Global Positioning System, GPS) have greatly facilitated accurate measurements and storage of geographical coordinates of points of interest. In the current example, a nautical chart overlaid with the satellite images covering the Nori farms might be used as the source to select GCPs (master-slave georeferencing), or alternatively, field-measured coordinates of rocky outcrops, roads and human constructions along the coast, serving as GCPs recognizable on the (large-scale) satellite images, might be used for direct georeferencing (figure 1).

## **GIS AND REMOTE SENSING: PHYCOLOGICAL APPLICATIONS**

### **GEOREFERENCING SPECIMENS**

Acquiring GPS coordinates has become self-evident, with handheld GPS devices nowadays fitting within any budget, provided that accuracy requirements are not smaller than 10-15m. Devices capable of handling publicly available differential correction signals like Wide Area Augmentation System (WAAS, covering North America), European Geostationary Navigation Overlay Service (EGNOS, covering Europe) and equivalent systems in Japan and India are slightly more expensive but offer

accuracies between 1 and 10m. However, contrary to terrestrial studies (Dodd, 2011), accuracies are almost always found to be better in coastal and marine practice, where the device would mostly be used in areas with a clear line of sight with the sky. Hence, field workers can accurately log shallow dives and snorkel tracks and revisit sampling points using consumer-grade GPS devices mounted on buoys that are dragged along. Accuracies within 1m can be obtained with commercial differential GPS systems, although this increases the cost and reduces mobility of field workers as a large portable station needs to be carried along, hence restricting use on water to boats or relatively accessible intertidal terrain. Logging GPS coordinates does however not eliminate the need for textual location information, preferably using official names or transcriptions as featured on maps, and using a hierarchical format going from more to less inclusive entities (cf. GenBank locality information; NCBI, 2008). This is vital to allow for error checking (see further). Several authors have recently independently and unambiguously stated that a lack of geographic coordinates linked to each recently and future sampled specimen can no longer be excused (Nature Editorial, 2008; Kidd & Ritchie, 2006; Kozak et al., 2008). Moreover, recommendations were made to require a standardized and publicly available deposition of spatial meta-information on all used samples accompanying each publication, including non-spatially oriented studies. This idea is analogous to most journals requiring gene sequences to be deposited in GenBank, whenever they are mentioned in a publication (Nature Editorial, 2008). For instance, the Barcode of Life project, aiming at the collection and use of short, standardized gene regions in species identifications, already requires specimen coordinates to be deposited for each sequence in its online workbench (Ratnasingham & Hebert, 2007).

Adding coordinates to existing collection databases can be a lot more challenging and time-consuming. At best, a locality description string in a certain format is already provided. In that case, gazetteers can be used to retrieve geographic coordinates. However, many coastal collections are made on remote localities without specific names, such as a series of small bays between two distant cities. Efforts have been made to develop software (e.g. GEOLocate; Rios & Bart, 1997) combining the use of gazetteers and civilian GPS databases to cope with information such as road names and distances from cities. Unfortunately, most of the existing automation efforts

are specifically designed for terrestrial collection databases, lacking proper maritime names, boundaries and functions. For instance, the software should allow specimens to be located at a certain distance from the shoreline. For relatively small collections, coordinates can also be manually obtained by identifying landmarks described in the locality fields or known by experienced field workers using Google Earth, a free GIS visualization tool with high to very high resolution satellite coverage of the entire globe (available online at <http://earth.google.com>). Manually adding specimen coordinates to database records does however increase the chance of errors in the coordinates, compared to automatically retrieving and adding coordinates.

Quality control of specimen coordinates is crucial. GIS allow for overlaying collection data with administrative boundary maps such as Exclusive Economic Zone (EEZ) boundaries, and comparing respective attribute tables to check for implausible locations. A common error, for instance, involves an erroneous positive or negative sign to a coordinate pair, resulting in locations on the wrong hemisphere, on land, or in open ocean. Additionally, when used in niche modeling studies, sample localities should be overlaid with raster environmental variable maps, to check if samples are not located on masked-out land due to the often coarse raster resolution.

## REMOTE SENSING

In documenting the consequences of global change, it is crucial to repeatedly and automatically obtain baseline thematic and change detection maps of (commercially or ecologically critical) seaweed beds. While not able to replace field work in detecting moderate changes in quantitative parameters and diversity, it has long been acknowledged that remote sensing is a powerful technique to overcome numerous problems in mapping and monitoring seaweed assemblages (Belsher et al., 1985). Accessibility of seaweed-dominated areas can be an issue if the location is remote, and the exploration of rocky intertidal shores can be hard or even hazardous. More importantly, most benthic marine macroalgal assemblages are permanently submerged, restricting their exploration to SCUBA techniques. Thus, mapping and monitoring extensive stretches on a regular basis is very time

and resource consuming when using *in situ* techniques only. This section provides an overview of different remote sensing approaches, without providing procedural information. For hands-on information on image processing techniques, see Green et al. (2000).

From a technical point of view, airborne remote sensing would seem most appropriate for seaweed mapping (Theriault et al., 2006; Gagnon et al., 2008). Light fixed-wing aircrafts are relatively easy to deploy, and sensors mounted on a light aircraft flying at low to moderate altitudes (1000m to 4000m) will typically yield datasets with a very high spatial and spectral resolution. For instance, the Compact Airborne Spectrographic Imager can resolve features measuring only 0.25m x 0.25m in up to 288 bands programmable between 400nm and 1050nm in the visible and near-infrared (VNIR) light depending on the study object characteristics. Additionally, the low acquisition altitude can result in a negligible atmospheric influence. However, light aircraft are generally not equipped with advanced autopilot capabilities and are sensitive to winds and turbulence. It takes considerable time and effort to geometrically correct images acquired from such an unstable platform. Altitude differences combined with roll and pitch (aircraft rotations around its 2 horizontal axes) all result in different ground pixel dimensions. Moreover, low altitude acquisitions result in a limited swath, increasing both acquisition time (and hence expense) through the use of multiple flight transects and processing time to geometrically correct and concatenate the different scenes. Chapter 7 in this thesis presents a tethered low-altitude aerial photography-based mapping of the intertidal using novel software that copes with these conditions. Alternatively, a more advanced (and hence again, more expensive) and stable aircraft can acquire imagery at higher altitudes covering larger areas, but this is at the cost of spatial resolution and atmospheric influence.

Overall, atmospheric and weather conditions play an important role in aerial seaweed studies, as the aircraft and the airborne and ground crew must be financed over an entire stand-by period in areas with unstable weather conditions (quite typical for coastal areas), as the weather conditions at the exact moment of acquisition cannot be forecasted long enough in advance during the planning stage of the campaign.

By contrast, satellites are more stable platforms that can cover much larger areas in one scene daily to biweekly, making these ideal monitoring

resources (table 1a and 1b). However, satellite-based studies of seaweed assemblages were suffering from a lack of spatial resolution until the late nineties. Typically, seaweed assemblages are very heterogeneous due to the morphology of rocky substrates, characterized by many differences in exposure to light, temperature fluctuations, waves, grazers and nutrients on a small area. These differences result in many microclimates and –niches, creating patchy assemblages in the scale of several meters to less than a meter, while no satellite sensor resolved features less than 15m until 2000. From that year onwards, very high resolution sensors were developed and made commercially available (table 1a), allowing for detailed subtidal seaweed mapping and quantification studies in clear coastal waters (e.g. Andréfouët et al., 2004).

With the availability of more advanced sensors in the 21<sup>st</sup> century, a trade-off between spatial and spectral resolution became apparent (figure 2) – an issue of particular relevance to seaweed studies. The trade-off situation evolved due to computer processing power and data storage capacity limitations at the time of sensor development - often 5 years prior to launch followed by another 5 years of operation. This is a long time in terms of Moore's law (Moore, 1965), describing the pace at which computer processing power doubles. These historical limitations dictated a choice between a high spatial resolution or a high spectral resolution in current sensors, but not both, while seaweed studies would arguably benefit from both. While the main macroalgal classes (red, green and brown seaweeds) are theoretically spectrally separable from each other as well as from coral and seagrass in 3 bands, this is not the case on a generic level. Additionally, information from seaweeds at below 5-10m depth can only be retrieved from blue and green bands due to attenuation of red and NIR in the water column. Hence, the availability of several blue and green bands can increase thematic resolution and the resulting classification accuracies, and this is of particular value in turbid waters, characteristic of many coastal stretches. By contrast, the absence of a blue band combined with only one green band (see several sensors in table 1a and 1b) prevents spectral discrimination of submerged seaweeds altogether and confined early remote sensing studies on seaweeds to the intertidal range (Guillaumont et al., 1993).

**Table 1a.** Current and future spaceborne passive remote sensors apt for seaweed mapping and monitoring: technical features. Sensors are sorted on decreasing resolution of the multispectral bands.

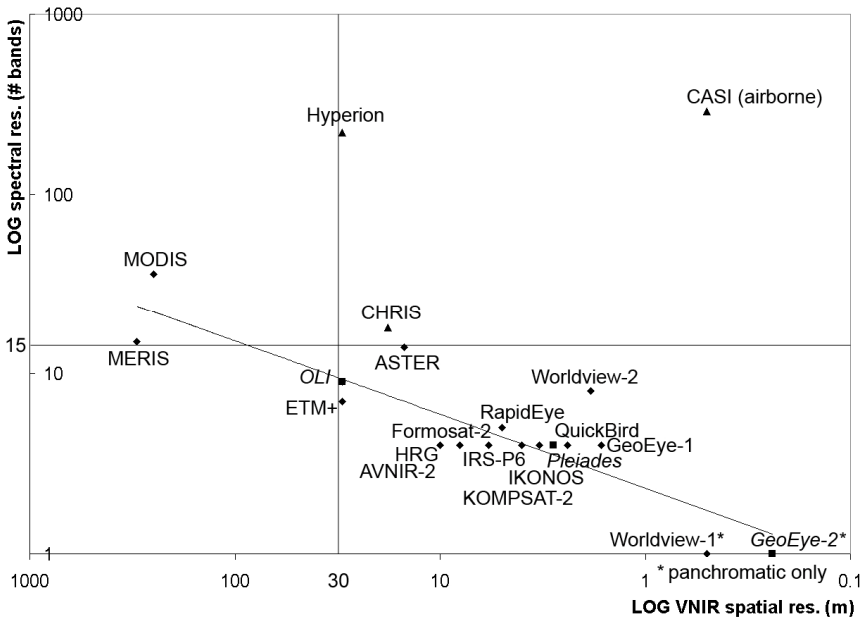
Platform	Sensor	Scene	Spatial Resolution	Spectral Resolution
Landsat 7	ETM+	185km x 170km	30m (60m TIR, 15m pan)	0.45-12.5µm, 6 bands + 1 thermal + 1 pan
EO-1	Hyperion	7.5km x 42-185km	30m	0.4-2.5µm, 220 bands
EO-1	ALI	37km x 42-185km	30m (10m pan)	0.433-2.35µm, 9 bands + 1 pan
PROBA	CHRIS	14km x 14km	18m (36m)	0.40-1.105µm, 18 bands (63 bands), programmable
TERRA	ASTER	60km x 60km	15m (30m SWIR, 90m TIR)	0.52-11.65µm, 9 bands + 5 thermal
SPO7 5	HRG	60km x 60km	10m (2.5m pan)	0.5-1.75µm, 4 bands + 1 pan
ALOS	AVNIR-2	70km x 70km	10m (2.5m pan)	0.42-0.89, 4 bands + 1 pan
FORMOSAT-2		24km x 24km	8m (2m pan)	0.45-0.9µm, 4 bands + 1 pan
IRS-P6 (ResourceSat-1)	LISS 3-4	23.9km x 23.9km	5.8m (23.5 SWIR)	0.52-1.7, 4 bands
RapidEye (5 identical satellites)	JSS 56	25km x 25km	5m	0.44-0.88µm, 5 bands
KOMPSAT-2 (=Ariang-2)		15km x 15km	4m (1m pan)	0.45-0.9µm, 4 bands + 1 pan
IKONOS		11.3km x 11.3km	3.28m (0.82m pan)	0.445-0.853µm, 4 bands + 1 pan
Quickbird		16.5km x 16.5km	2.4m (0.6m pan)	0.45-0.9µm, 4 bands + 1 pan
GeoEye-1		15.2km x 15.2km	1.64m (0.41m pan)	0.45-0.92µm, 4 bands + 1 pan
WorldView-2		16.4km x 16.4km	1.84m (0.46m pan)	0.4-1.04µm, 8 bands + 1 pan
PLEIADES-HR1/2		20km x 20km	2.8m (0.6m pan)	0.43-0.95µm, 4 bands + 1 pan
LDCM	OLI	185km x 185km	30m (100m TIR, 15m pan)	0.43-12.5µm, 8 bands + 1 thermal + 1 pan



**Table 1b.** Current and future spaceborne passive remote sensors apt for seaweed mapping and monitoring: availability, operational and quality remarks.

Platform	Sensor	Temporal Resolution	Availability	Cost	Remarks
Landsat 7	ETM+	16 days	1999-...	Free-\$	High quality earth observation data: calibration within 5%. Since the 2003 SLC failure, scenes are flawed with 25% gaps
EO-1	Hyperton	16 days	2000-...	Free-\$\$	EO-1 flies in formation with Landsat 7's (trailing by 1 min) to benefit from its high quality calibration. EO-1 has cross-track off-nadir capability. ALI is a technology verification instrument
EO-1	ALI	16 days	2000-...	Free-\$\$	
PROBA	CHRIS	7 days	2001-...	Free	Technology verification instrument; Along track +/- 55° off-nadir capability for stereo 3D imaging
TERRA	ASTER	16 days	2000-...	\$	Lack of blue band limits use to intertidal and surfacing/floating seaweeds; VNIR cross track 24° off-nadir and NIR backward looking capability for stereo 3D imaging
SPOT 5	HRG	1-3 days	2002-...	\$\$-\$\$\$	Lack of blue band limits use to intertidal and surfacing/floating seaweeds
ALOS	AVNIR-2	2 days	2006-...	\$	44° off-nadir capability; Panchromatic stereo 3D imaging
FORMOSAT-2		1 day	2004-...	\$\$-\$\$\$	Cross and along-track 45° off-nadir capability for stereo 3D imaging
IRS-P6 (ResourecSat-1)	LISS 3-4	5 days	2003-...	\$\$-\$\$\$	Lack of blue band cf. SPOT 5; 26° off-nadir capability for stereo 3D imaging
RapidEye (5 identical satellites)	JSS 56	1 day (combined)	2008-...	\$\$\$	Red edge band improves vegetation monitoring
KOMPSAT-2 (=Arirang-2)		3 days off-nadir	2006-...	\$\$	Cross-track 30° off-nadir capability
IKONOS		3-5 days off-nadir	1999-...	\$\$-\$\$\$	Cross track 60° and along-track off-nadir capability for stereo 3D imaging
Quickbird		1-3.5 days off-nadir	2001-...	\$\$-\$\$\$	Cross and along-track 30° off-nadir capability for stereo 3D imaging
GeoEye-1		3-5 days off-nadir	2008-...	\$\$\$	Cross track 60° and along-track off-nadir capability for stereo 3D imaging
WorldView-2		1.1-3.7 days	2009-...	\$\$-\$\$\$	Cross-track 40° off-nadir capability. Red edge band improves vegetation monitoring; coastal blue band improves water penetration and oceanographic observations
PLEIADES-HR/2		1 day off-nadir (combined)	2012-...	\$?	Planned successor in SPOT series. Capable of steering 30° off-track and viewing 43° off-nadir
LDCCM	OLI	16 days	2013-...	\$?	Planned successor in Landsat series. Successor of EO-1 ALI technology. coastal blue band improves water penetration and oceanographic observations

Note: table 1 represents the situation as of June 2011. \$ = less than \$500, \$\$ = \$500-2500, \$\$\$ = more than \$2500 per 50 km<sup>2</sup> or per scene. Prices may vary according to archived imagery or new acquisition tasks.



**Figure 2.** Trade-off between Log spectral resolution plotted against Log VNIR spatial resolution in current and future satellite sensors. All sensors are spaceborne, except for the airborne CASI sensor, shown here for comparison. In order to demonstrate the trade-off extremes, the panchromatic-only missions Worldview-1 and GeoEye-2 are added in the figure but not further discussed for seaweed mapping. Likewise, the ocean color sensors MODIS and MERIS are displayed for illustrative purposes. We consider sensors featuring a spatial resolution between 0 and 30m and a spectral resolution above 15 bands in the visible and NIR spectrum of high value for seaweed mapping and monitoring (upper right quadrant). We therefore recommend future satellite sensor developments towards the CASI position, but note the position of the planned earth observation missions LDCM/OLI and Pleiades along the current trade-off situation. ◆ current sensors (on which the trade-off line is based); ■ future sensors; ▲ current sensors forming an exception to the general trade-off situation between spectral and spatial resolution in satellite sensors.

Besides the intertidal, NIR bands are useful (in combination with red) to discriminate surfacing or floating seaweeds, and allow to discern decomposing macroalgae, as NIR reflection decreases with decreasing chlorophyll densities (Guillaumont et al., 1997). From figure 2, it should be noted that two high spatial resolution spectral imaging sensors have been developed recently, Hyperion (onboard EO-1) and CHRIS (onboard PROBA), with a spectral resolution approaching that of airborne sensors, hence forming an exception on the historical trade-off. Research by the first author of this chapter suggested that CHRIS imagery can be used to map and monitor benthic communities in turbid waters at the south coast of

Oman (Arabian Sea). Intertidal green, brown and red seaweeds as well as submerged mixed seaweed beds, coral and drifting decomposing seaweeds were discerned with reasonable accuracy during both monsoon seasons, as reported in Chapter 6 of this thesis.

## DISTRIBUTION AND NICHE MODELING

For centuries, biogeographical patterns have been studied in a descriptive way by delineating provinces and regions based on observed species presences and degrees of endemism, rather than quantifying and explaining these patterns based on environmental variables (Adey & Steneck, 2001). The question as to which environmental variables best explain seaweed species' niches and distributions, is however one of the most important in global change research. Biogeographical models based on these variables could allow for predicting range shifts and directing field work to discover unknown seaweed species and communities.

It is widely recognized that temperature is a major forcing environmental variable for coastal macrobenthic communities in general and seaweeds in particular. Temperature plays a significant role in biochemical processes, and generally species have evolved to tolerate only a (small) portion of the entire range of temperatures in coastal waters. Thus, it is evident that sea surface temperature (SST, often used as a proxy for water column temperature in shallow coastal waters) plays a prominent role in seaweed niche distribution models. Furthermore, while temperature is often measured in a time-averaged manner (daily, monthly, yearly), it is important to note that the timing of seasons differs globally (even within hemispheres due to seasonal upwelling phenomena). As some seaweed species or specific life cycles are limited by maximum and others by minimum temperatures, it is obviously essential to base models on biologically more relevant maximum, minimum and related derived variables rather than on raw time-averaged measurements.

van den Hoek et al. (1990) gave an overview of how generalized or annual temperature isotherm maps could be used to explain the geographic distribution of seaweed species in the context of global change.

Adey & Steneck (2001) later described a quantitative model based on maximum and minimum temperatures as the main variables, combined with

area and isolation, to explain coastal benthic macroalgal species distributions. Additionally, their thermogeographic model was integrated over time as they incorporated temperatures from glacial maxima, allowing biogeographical regions to dynamically shift in response to two historical stable states of temperature regimes (glacial maxima and interglacials). In this respect, their study is of significant value in global change research, although their graphic model outputs were not based on GIS and not straight-forward to interpret. Moreover, using analogous or vector isothermal SST maps, both studies suffered from a lack of resolution in SST input data, consequently compromising the resolution and accuracy of the model outputs.

Recently, two major studies demonstrated how seaweed distribution models can benefit greatly from the extensive and free availability of environmental variables on a global scale through the use of satellite data. These data are not only geographically explicit and readily usable in GIS, but also provide much more accuracy than isotherm maps due to their continuity. Schils & Wilson (2006) used Aqua/MODIS three-monthly averaged SST data in an effort to explain an abrupt macroalgal turnover around the Arabian Peninsula. Their results pointed to a threshold of 28°C, defined by the average of the three warmest seasons, explaining diversity patterns of the seaweed floras across the entire Indian Ocean. They stressed that a single environmental factor can thus dominate the effect of other, potentially interacting and complex variables. On the other hand, Graham et al. (2007) took several other variables in consideration to build a global model predicting the distribution of deep-water kelps. Their study was essentially a 3D mapping effort to translate the fundamental niche of kelp species, as determined by ecophysiological experiments, from environmental space into geographical space, based on global bathymetry, photosynthetically active radiation (PAR), optical depth and thermocline depth stored in GIS. The latter was based on the interpolation of vertical profiles, while the former three variables were derived from satellite datasets (table 2).

The latest advance in distribution modeling approaches concerns several Species' Distribution Modeling (SDM) algorithms, also termed Ecological Niche Modeling (ENM), Bioclimatic Envelope Modeling (BEM) or Habitat Suitability Mapping (HSM). While the names are often mixed in

the same context, a slight difference in meaning exists (see box 1 for additional background). Many different algorithms and software implementations exist [Maxent (see box 2), GARP, ENFA, BioClim, GLM, GAD, BRT, but see Elith et al. (2006) for a review], but two fundamental properties are combined in these techniques, clearly separating them from the studies described in the previous paragraphs, which showed at most one of these properties. Firstly, input data are a combination of a vector point file, representing georeferenced field observations of a species (as opposed to ecophysiological experimental data) on the one hand, and climatic variables stored in a raster GIS on the other hand. The modeling algorithms then read the data out of GIS and use statistical functions to calculate the realized niche (as opposed to the fundamental niche; Hutchinson, 1957) in environmental space, subsequently projecting the niche back into geographical space in GIS. Secondly, instead of a binary identification of suitable and unsuitable areas, ENM output is a continuous probability distribution, which makes more sense from a biological point of view.

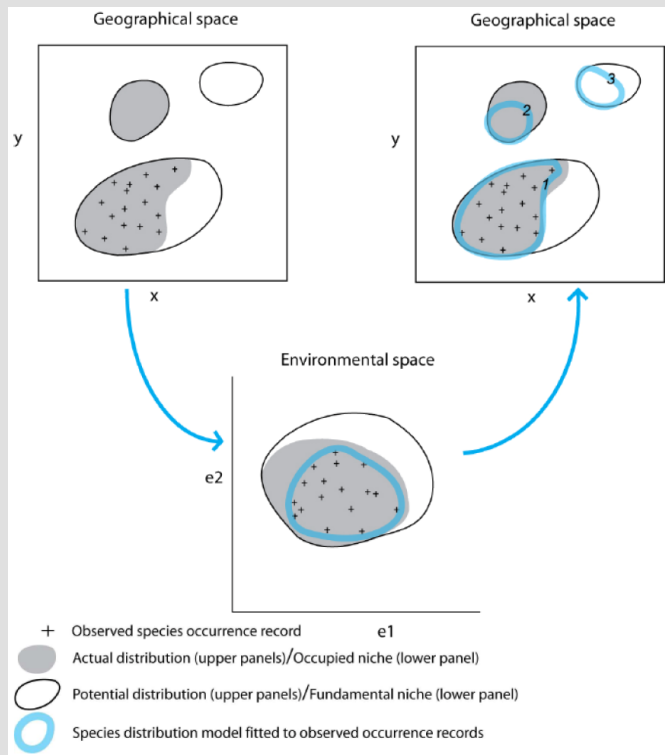
**Table 2.** Current environmental variables retrievable from selected satellite data on a global scale.

Variable	Source	Resolution (level 3)	Available
Sea Surface Temperature (SST)	NOAA/AVHRR/1-3	4-9km (2-5arcmin)	1985-...
	Terra/MODIS	4km (2arcmin)	2000-...
	Aqua/MODIS	4km (2arcmin)	2002-...
Chlorophyll-a concentration (CHL)	SeaWiFS	9km (5arcmin)	1997-2011
	Envisat/MERIS	9km (5arcmin)	2002-...
	Terra or Aqua/MODIS	4km (2arcmin)	2000-...
Photosynthetically active Radiation (PAR)	SeaWiFS	9km (5arcmin)	1997-2011
	Terra or Aqua/MODIS	4km (2arcmin)	2000-...
Diffuse Attenuation (DA)	SeaWiFS	9km (5arcmin)	1997-2011
	Terra or Aqua/MODIS	4km (2arcmin)	2000-...
Dissolved Organic Matter (DOM)	SeaWiFS	9km (5arcmin)	1997-2011
	Terra or Aqua/MODIS	4km (2arcmin)	2000-...
Particulate Organic Carbon (POC)	SeaWiFS	9km (5arcmin)	1997-2011
	Terra or Aqua/MODIS	4km (2arcmin)	2000-...
Particulate Inorganic Carbon (PIC)	SeaWiFS	9km (5arcmin)	1997-2011
	Terra or Aqua/MODIS	4km (2arcmin)	2000-...
	QuikSCAT/SeaWinds	25-110km	1999-2009
Surface winds	MetOp-A/ASCAT	12.5-25km	2006-...
	SAC-D/Aquarius	110km (1arcdegree)	2011-...
	SAC-D/Aquarius	110km (1arcdegree)	2011-...

**Box 1. Niches and niche modeling**

(after Peterson, 2006; Pearson, 2007; Soberón, 2007 and Colwell & Rangel, 2009)

Species' niches have been described in various ways since Grinnell's first definition. The Grinnellian niche can be defined by fundamentally non-interacting habitat variables and abiotic environmental conditions on broad scales (so-called scenopoetic variables), relevant to understanding coarse-scale ecological and geographic properties of species. The Eltonian niche focuses on biotic interactions and resource-consumer dynamics essentially acting at local scales (so-called bionomic variables). Grinnell and Elton viewed these niche concepts as an abstract set of characteristics belonging to a certain place in geographical space. By contrast, the Hutchinsonian niche is viewed as a multidimensional volume in hyperspace defined by axes consisting of conditions and resources, attributed to a species or population. The fundamental niche is then the full set of conditions under which a species can exist, while this is often constrained by species interactions and/or dispersal limitations to the realized niche. In this concept, a species' niche defined in environmental space can hence be translated to geographical space and vice versa. This duality forms the basis for ecological niche modeling, as shown below.



The diagram above illustrates how a hypothetical species distribution model may be fitted to observed species occurrence records (Pearson, 2007). A modeling technique is used to characterize the species' niche in environmental space by

relating observed occurrence localities to a suite of environmental variables. Notice that, in environmental space, the model may not identify either the species' occupied niche or fundamental niche; rather, the model identifies only that part of the niche defined by the observed records. When projected back into geographical space, the model will identify parts of the actual distribution and potential distribution. For example, the model projection labeled 1 identifies the known distributional area. Projected area 2 identifies part of the actual distribution that is currently unknown; however, a portion of the actual distribution is not predicted because the observed occurrence records do not identify the full extent of the occupied niche (i.e. there is incomplete sampling). Similarly, modeled area 3 identifies an area of potential distribution that is not inhabited (the full extent of the potential distribution is not identified because the observed occurrence records do not identify the full extent of the fundamental niche due to, for example, incomplete sampling, biotic interactions, or constraints on species dispersal).

Inherently, biotic samples from the physical world yield occurrence records in the realized part of a species' niche, as opposed to its fundamental niche, since knowledge on all species interactions cannot be obtained in regular sampling campaigns and absence of interactions cannot be assumed. Moreover, it is very hard to generate raster layers accounting for biotic interactions such as grazing, which would reflect a species' Eltonian niche. Hence, all the modeling case studies in this thesis are exclusively based on abiotic (so-called scenopoetic) variables, defining the (realized) Grinnellian niche (Colwell & Rangel, 2009; Rödder & Engler, 2011). It is important to note that although Eltonian niche factors are responsible for confining a species to its realized niche within its fundamental (Grinnellian) niche and that although exactly those Eltonian variables are not included in most niche modeling approaches, the modeled niche is still part of the realized niche because of the inherent input sample characteristics. This concept is especially important for macroalgae, because in many cases (for instance, on coral reefs), the only responsible factor for local absences (or lack of occurrence data) may be grazing pressure.

In the literature, the terms niche modeling, bioclimate envelope modeling, habitat suitability modeling and species distribution modeling are often mixed or used differently in a range of contexts and goals. One distinction is sometimes made between presence-only modeling techniques, which have been referred to as niche or envelope modeling, and modeling techniques using actual presence and absence data, often termed distribution or suitability modeling. In the framework of this thesis, a scale- and purpose-related definition would be more appropriate. When modeling in a broad biogeographic context where environmental variables on a global scale are considered and issues of dispersal capacity play a role, ecological niche modeling would be the more appropriate term. By contrast, distribution or suitability modeling focuses on a much finer scale, where dispersal is not the limiting factor and species interactions would be more important. As such, Chapters 2 and 3 would be niche modeling studies, with Chapter 4 expanding the dimensions of variables that can be included, while Chapters 5 and 6 would be more concerned with actual distribution and suitability modeling (although as explained earlier, biotic variables are not included). However, when relating to other studies, a strict distinction cannot be maintained.

Continuous probability maps may then be converted to binary maps using arbitrary thresholds. Additionally, ENM algorithms typically use several statistics to pinpoint the most important environmental variable in terms of model explanation, giving its percent contribution to the model output. Also, response curves can be calculated for the different variables, defining the niche optima.

However, care must be taken to restrict model input to uncorrelated environmental variables to obtain valid results. With a growing availability of (global, gridded) environmental datasets which are often correlated or redundant, a data reduction strategy should be considered. One may perform a species-environment correlation analysis or ordination to make a first selection of relevant variables and perform a subsequent Pearson correlation test between environmental variables to get rid of redundant information. Alternatively, spatial principal component analysis (PCA, see box 2) may be performed to obtain uncorrelated variables, using PCA components as input variables (Verbruggen et al., 2009), although resulting variable contributions and response curves might be hard to calculate back to original variables.

## Box 2. Spatial techniques

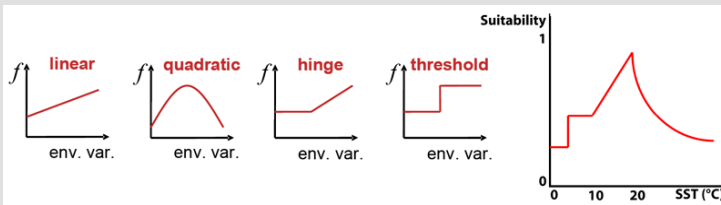
*Maxent distribution modeling algorithm* (after Pearson, 2007)

Maxent is a general-purpose method for characterizing probability distributions from incomplete information. In estimating the probability distribution defining a species' distribution across a study area, Maxent formalizes the principle that the estimated distribution must agree with everything that is known (or inferred from the environmental conditions where the species has been observed) but should avoid making any assumptions that are not supported by the data. The approach is thus to find the unique probability distribution  $P(x)$  of maximum entropy (the distribution that is most spread-out, or closest to uniform) subject to constraints imposed by the information available regarding the observed distribution of the species and environmental conditions across the study area.

$$P(x) = \frac{e^{\sum_{i=1}^n c_i \cdot f_i(x)}}{Z}, \text{ with } c = \text{scaling constants and } Z = \text{regularization parameter}$$

In the above equation,  $f_i(x)$  denote functions of environmental variables and their interactions terms (so-called features). Features can take several forms, allowing to fit more or less complex models by combining features as exemplified in the figure below. More occurrence records are needed to fit increasingly complex functions.





The Maxent method does not require absence data for the species being modeled; instead it uses background environmental data for the entire study area. The method can utilize both continuous and categorical variables and the output is a continuous prediction (most commonly depicted as logistically scaled relative suitability from 0 to 1). The freely available software implementation developed by Phillips et al. (2006, 2008, 2009, to which the reader is referred for further mathematical details) also calculates a number of alternative thresholds to convert the continuous probability distribution to discrete maps, computes model validation statistics, and enables the user to run a jackknife procedure to determine which environmental variables contribute most to the model prediction. Maxent has been shown to perform well in comparison with alternative methods (Elith et al., 2006). One drawback of the Maxent approach is that it uses an exponential model that can predict high suitability for environmental conditions that are outside the range present in the study area (i.e. extrapolation). To alleviate this problem, when predicting for variable values that are outside the range found in the study area, these values can be set (or ‘clamped’) to match the upper or lower values found in the study area. Alternatively, extrapolation can be used to identify potential suitable areas under new climate scenarios.

#### *Spatial PCA* (after Clark Labs IDRISI manual)

Principal Component Analysis (PCA) is a linear transformation technique related to Factor Analysis. Given a set of image bands, PCA produces a new set of images, known as components that are uncorrelated with each other and explain progressively less of the variance found in the original set of bands. Both standardized and unstandardized principal component analyses can be performed. In the standardized case, the correlation matrix is used for input rather than the usual variance/covariance matrix.

PCA has traditionally been used in remote sensing as a means of data compaction. For a typical set of raster layers covering the same area, it is common to find that the first two or three components are able to explain virtually all of the original variability in the pixel values. Later components thus tend to be dominated by noise effects. By rejecting these later components, the volume of data is reduced with no appreciable loss of information.

#### *DIVA interpolation* (after Troupin et al., 2010)

Data-Interpolation Variational Analysis (DIVA) is an implementation of the Variational Inverse Method (VIM). It is a geostatistical approach to the interpolation of point data to gridded climatological fields, allowing the consideration of coastlines and bottom topography without additional parameterization, and has a numerical cost almost independent on the number of

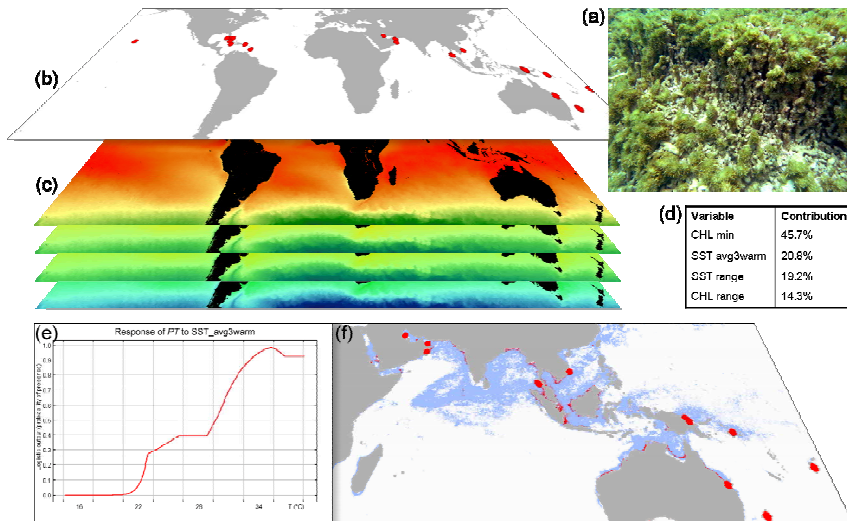
observations. Moreover, only a few parameters need to be determined, which is done in an objective way. The algorithm is able to deal with a great number of individual data, without needing to work with averaged values or data bins. It also generates coherent error maps, while the computational time required for both climatological and error maps is kept to a reasonable level, allowing routine runs on several depth levels. For elaborate mathematical details, the reader is referred to Troupin et al. (2010).

Pauly et al. (2009) applied preliminary ENM using Maxent (Phillips et al., 2006; see box 2) to gain insight in worldwide blooms of the siphonous green alga *Trichosolen* growing on physically damaged coral (figure 3). A correlation analysis was applied to identify the two least correlated biologically meaningful variables derived from SST and CHL (based on monthly datasets), adequately describing the position and extent of the distribution in environmental space. The model delineated the potential global distribution of *Trichosolen* occurring on coral based on a 95% training confidence threshold, including areas where the bloom had previously occurred. This allowed identifying areas with a high potential risk for future blooms based on environmental response curves. For instance, the response curve for the average of the three warmest months [included as a variable based on the conclusions of Schils & Wilson (2006)] shows that *Trichosolen* populations are only viable above 22°C, but only environments above 28°C are likely to sustain blooms. Chapter 2 of this thesis presents this approach more elaborately.

## **FUTURE DIRECTIONS AND RESEARCH PRIORITIES**

### **THE QUEST FOR SPATIAL DATA IN SEAWEED BIOLOGY**

In its simplest form, “spatially explicit” seaweed data would refer to the availability of georeferenced species occurrences. While we discussed the practice of georeferencing and dissemination of spatially explicit seaweed data in depth in the second section of this chapter, we briefly show a couple of examples to demonstrate the dramatic state of the current availability of this information. For instance, looking at a random Nori species, *Porphyra yezoensis* Ueda, AlgaeBase (Guiry & Guiry, 2008) mentions 13 references to occurrence records throughout the northern hemisphere.



**Figure 3.** (a) A *Pseudobryopsis/Trichosolen* (*PT*) bloom on physically damaged coral. (b) Worldwide occurrence points of *PT* on coral. (c) Environmental grids used for model training in Maxent. (d) Relative importance of each variable in the model as identified by the algorithm. (e) Response curve of *PT* to the average of the three warmest months. (f) Binary habitat suitability map for *PT*. The grey (blue) shade represents suitable environment, while the dark (red) shade along the coast delineates bloom risk areas.

By contrast, the Ocean Biogeographic Information System (OBIS, an on-line integration of marine systematic and ecological information systems; Costello et al., 2007) contains no *P. yezoensis* records. Another random example, the siphonous (sub)tropical green alga *Codium arabicum* Kützing illustrates this further: out of 55 direct or indirect occurrence references in Algaebase, 17 are georeferenced in OBIS. However, two of the specimens wrongly have zero longitudes, hence locating the records some 400km inland from the coast of Ghana, instead of at the Indian coast. Five out of the 17 are recorded to no better than 0.1 degree in both longitude and latitude, making their position uncertain within up to 120km<sup>2</sup>. Fifteen out of the 17 make no mention of the collector's name or publication, preventing to check the integrity of the identification. Eleven lack sub-country level locality name information, and none mention sub-state locality names, making it impossible to verify geographical coordinates through the use of gazetteers.

If the amount of coastal or marine publications using GIS, mapping or remote sensing can be called minimal, averaging 8% of the total publications using these geographic techniques as previously shown, the proportion of these records mentioning seaweeds or macroalgae is statistically speaking barely existing, attaining 0.5-1% of the spatial marine studies. Studies investigating the other two best-known benthic marine communities, coral (reefs) and seagrasses, constitute up to 10%, while the remainder covers (in no particular order) mangroves and other supratidal coastal communities and structures, coastal or marine topography and geomorphology or nautical issues. Some of the reasons accounting for this disproportion are obvious: for a start, relatively few investigate seaweeds. However, out of 12074 studies mentioning seaweeds or macroalgae in ISI Web of Knowledge, a potential 7279 in the fields of ecology, biogeography, phylogeography or ecophysiology could benefit from some sort of spatial explicit information, while only 177 (2.5%) actually mention to do so in their title, abstract or keywords. Other problems concern the nature of seaweed communities: while coral reefs and seagrass meadows usually form large and relatively homogenous assemblages, seaweeds are spatially and spectrally very heterogeneous. This is particularly difficult to cope with in remote sensing studies, already challenged by the properties of the water column in comparison to terrestrial vegetation studies.

#### GEOREFERENCING SPECIMENS

Previous sections of this chapter demonstrate the need to prioritize the standardization of disseminating and linking geographical seaweed specimen information. Investigating the consequences of global change requires the availability of correct and complete global datasets. Therefore, we support the requirement of the dissemination of sample coordinates not only from geographically oriented studies, but from every study using *in situ* sampled seaweeds, to allow for informative and accurate meta-analyses. Coordinate pairs should be deposited in already existing global biodiversity databases such as OBIS, but minimal geographic accuracy and complete specimen information including collector's name should be required in order to allow vigorous quality control. Additionally, the geodetic datum in which coordinates are shared should be part of this standardization effort.

Positional data recorded in a local datum by national institutions or organizations are hard to translate correctly to global projections. Hence, we recommend the use of WGS-84, the default global datum on consumer-grade GPS devices.

The use of global biodiversity databases as a main depositing centre for specimen coordinates rather than dedicated seaweed databases also opens perspectives to investigate potential correlations between seaweed and faunal distribution shifts in response to global change. However, it should also be investigated how general geographical biodiversity databases such as OBIS could be related to and synchronized with specific databases such as Algaebase and GenBank to optimize the dissemination of all kinds of specimen information.

## REMOTE SENSING

No significant time gap exists between the development and deployment of airborne sensors; due to an optimal use of the most recent technologies, airborne sensors thus represent the best technical characteristics desirable for seaweed mapping to date. As time goes on, the most recent satellite sensors can benefit from the evolution in technologies to more closely resemble the properties of airborne sensors. Vahtmäe et al. (2006) used a simulation study to demonstrate that submerged seaweeds in turbid coastal waters could well be mapped using hyperspectral satellite sensors like CHRIS and Hyperion, featuring 10nm wide bands in the visual wavelengths. However, they also postulated a signal-to-noise ratio of 1000:1, an image quality not met by these existing sensors. It is thus vital that similar new hyperspectral, very high resolution satellite sensors should be developed for seaweed mapping and monitoring in the framework of global change research. Similarly, studies on seagrass meadows have shown that spectral separability on the species level becomes problematic on depths greater than 3 m due to lack of combined very high spatial and spectral resolutions, and best results to date were achieved using airborne hyperspectral sensors that do meet these requirements (Phinn et al., 2008). Ferwerda et al. (2007, also including a review of seagrass mapping studies) suggest a workaround to the lack of suitable sensors for seagrass mapping by using environmental information such as suspended sediment, water

constituents, nutrient concentration and temperature that can be retrieved from multiple current sensors to model seagrass productivity, rather than directly mapping it. However, this approach is not directly applicable to seaweed mapping because of the less clear relationships between those environmental factors and seaweed growth and survival. For direct mapping, Phinn et al. (2008) also recommend using imagery with both very high spatial ( $< 5$  m) and spectral resolution. By contrast, figure 2 shows that planned sensors for the next years follow the historical trade-off towards multispectral very high resolution systems. Nowadays, this seems to be motivated by two elements: the huge thrust for coral reef research, in which macroalgae are often lumped into one or few functional classes and spatial resolution is considered more important than spectral resolution, and disaster event monitoring, focusing on a near-one day site revisiting time through the use of extensive off-nadir or off-track pointing capabilities (table 1b).

The latter technique also generates huge amounts of data, adding a new dimension to the historical trade-off situation: current data storage capacities allow for two image characteristics out of three (spectral, spatial and temporal resolution) to be optimized, but not all three. Unfortunately, no significant thrust seems to exist in order to develop sensors ideally suitable for large-scale algal mapping and monitoring to date, explaining the characteristics of the missions in development. As a means to deal with the lack of very high resolution hyperspectral imagery, efforts have been made to combine the information from several sensors with different characteristics into one dataset. This is analogous to pan-sharpening techniques, which use the high spatial detail of a panchromatic band to spatially enhance the multispectral imagery from the same sensor (see also figure 2). Although useful in current conditions, we suspect these techniques to become less important as more advanced sensors would be developed, since processing information from one sensor evidently is less time- and resource-consuming and more accurate than using multi-sensor information.

Light-based active remote sensing (known as LIDAR, Light Detection And Ranging) involves the emission of laser pulses with a known frequency and subsequently detecting fluorescence in certain wavelengths. Kieleck et al. (2001) proved this technique to be successful in discerning

submerged green, brown and red seaweeds in lab conditions. Mazel et al. (2003) used a similar prototype in-water laser multispectral fluorescence imaging system to map different coral reef bottom structures, including macroalgae, on a 1cm resolution. Airborne laser imaging has been used extensively to provide very high resolution imagery in terrestrial applications such as forestry. In the marine realm, its applications are mostly limited to in-water (boat-mounted) transect mapping strategies, although further research to develop aerial systems could prove useful to obtain very high resolution imagery of individual seaweed patches, e.g. to map the spreading of macroalgae on coral reefs.

## DISTRIBUTION AND NICHE MODELING

Presence-only data are mostly inherent to seaweed niche distribution modeling due to sampling locality bias (caused by difficulty of coastal and submerged terrain access), small sizes or seasonally microscopic life stages of seaweed species, or cryptic species. Under the title of Species' Distribution Modeling, Pearson (2007) published a general manual including (presence-only) niche modeling, mostly based on Maxent. However, the manual is based on terrestrial experiences, as niche modeling algorithms have rarely been applied to seaweed distribution to date, and some issues characteristic of marine benthic niche modeling are not elaborated. For instance, there are more global environmental GIS data available for the terrestrial realm compared to the marine environment. Table 2 lists marine environmental variables currently available from global satellite imagery, along with data that will become available in the near future. Especially globally gridded, remotely sensed salinity data have only recently become available. Other variable datasets (pH, nutrients, salinity, turbidity, etc.) may also be compiled from the interpolation of *in situ* data, e.g. from the Worldwide Ocean Optics Database (Freeman et al., 2006) or the World Ocean Database (NOAA, 2008). These data, consisting of vertical profiles, can be advantageous for 3D modeling, but the interpolation techniques necessary to obtain gridded maps may be challenging. An attempt to kick-start this integration of satellite data and *in situ* data interpolated using DIVA is presented in Chapter 4 in this thesis (see box 2). Furthermore, global change climate extrapolations resulted in the production of low resolution

global gridded maps of environmental variables for future scenarios in the terrestrial realm as well as the marine realm, distributed by the Intergovernmental Panel on Climate Change (IPCC). The projection of calculated niches on future distributions can greatly enhance our understanding of global change consequences, and it is therefore crucial that future research is aimed at composing similar gridded maps of future scenarios for marine environmental variables. More research should also be aimed at setting model parameters to account for spatial autocorrelation and clustering of species occurrence data. Lastly, model validation and output comparison statistics are under scrutiny in recent literature (e.g. Peterson et al., 2008), and more research is needed to agree on the best statistics suitable for marine data.

Modeling on a local scale allows for including high-resolution environmental variables that are not available for the entire globe. This is particularly the case where environmental variables not available from satellite data have been measured *in situ* and can be interpolated locally. In other cases, one or several (very) high resolution satellite scenes can be used to provide substrate data, not relevant on a global scale with 1km gridded environmental variables. For instance, De Oliveira et al. (2006) included substrate, flooding frequency and wave exposure to model the distribution of several intertidal and shallow subtidal brown seaweeds along a 20km coastal stretch in Brittany, France. Thus, it can be expected that multiscale modeling approaches will gain importance in the near future. A comparison between models based on equivalent coarse- and fine-resolution environmental data is presented in Chapter 5.

While human-induced effects on habitats are thought to drive short-term species dynamics, it is often stated that global climate change will influence the capacity of alien species to invade new areas on a medium to long term. Range shifts of individual species in an assemblage under climate change are based on largely the same processes driving the spread of alien species. Hence, the two can be addressed using the same approach (Thuiller et al., 2007). Although very complex processes are involved, the geographic component of species' invasions can be very well predicted using niche modeling techniques (Peterson, 2003). Once a comprehensive marine environmental dataset is compiled, invaded areas and areas at risk of invasion can be successfully predicted based on the native niche of alien



species (Pauly et al., 2009; Peterson, 2005), although Brönnimann et al. (2007) warn that a niche shift may occur after invasion. Nevertheless, niche modeling approaches are promising in future research of seaweeds' range shifts and invasions. Verbruggen et al. (2009) also applied niche modeling techniques to unravel the evolutionary niche dynamics in the green algal genus *Halimeda*, concluding that globally changing environments may allow certain macroalgae to invade neighboring niches and subsequently to form a divergent lineage. They also used Maxent to identify key areas to be targeted for future field work in search for new sister species – an application in biodiversity considered important in the light of global change. This approach is outlined in detail in Chapter 3 in this thesis.

To date, seaweed assemblages have often been characterized using quantitative vegetation analyses and multivariate statistics in order to delineate different community types and to establish the link between environmental variables and communities. With quickly developing niche modeling algorithms now regarded as the most advanced way to accomplish the latter, community niche modeling will be of particular value in global change-related seaweed research in the coming years. Ferrier & Guisan (2006) defined three ways to predict the niche of communities as a whole, rather than the niches of individual species. The assemble-first, predict-later strategy seems the most promising for seaweed data, since existing floristic data have often been statistically assembled into communities. We suggest that prioritizing the development of community niche modeling algorithms can greatly speed up our insight in seaweed community response to future climate change.

## AIMS AND OUTLINE

The general aim of this thesis is to investigate the applicability of several spatially explicit data acquisition and processing techniques that have been developed and used to map and predict the distribution of mainly terrestrial species to the study of seaweeds. In order to identify hands-on problems related to applying these techniques in the marine realm in general and seaweeds in particular, we present case studies (see box 3 for an outline of the species discussed in this thesis), rather than theoretical, experimental or simulation approaches. We suggest workarounds, adaptations and extensions to the existing methods to optimise the performance of distribution or niche modeling (chapters 2-5) and mapping (chapters 6-7) of marine species. Our goal is to show how results from these spatially explicit analyses add new insights in the distribution and ecology of macroalgal species and communities, and how they can be used in future monitoring projects.

In **chapter 2** we investigate the use of Maxent niche modeling to predict the potential distribution of the green algal genus *Trichosolen* and to derive ecological triggers of global ephemeral blooms of the species on impacted coral reefs. The aim is to complete scarce knowledge on the distribution and ecology of this genus characterized by a partly microscopic life cycle and to delineate potential future bloom risk conditions and areas.

In **chapter 3** we present an application of macroecological niche modeling in a macroevolutionary context. We use GIS to retrieve environmental information for georeferenced records of extant species, and use these data as input in an ancestral state reconstruction to infer ancestral niches. Niche modeling using Maxent is applied to compare known distributions with potential distributions in order to evaluate speciation scenarios and to delineate areas for potential discoveries of sister species.

Chapters 2 and 3 clearly demonstrate that a lot of effort in marine niche modeling is put in the compilation of (global) marine environmental data. Therefore, **chapter 4** aims to boost marine distribution and niche modeling applications by presenting the first comprehensive standardized and

uniform global marine environmental dataset, readily downloadable and usable for predictive studies. We present the methods used to integrate global remote sensing products and oceanographic data obtained by interpolation of *in situ* measurements. Additionally, we demonstrate global applicability through a case study of the invasive green alga *Codium fragile* subsp. *fragile*, predicting its potential spread.

While chapters 2 to 4 focus on the global scale, chapters 5 and 6 show the applicability and caveats of distribution modeling approaches on a regional and local scale. In chapters 6 and 7, local mapping techniques are presented. In **chapter 5**, we focus on regional processes by comparing Maxent distribution models for three key macroalgal species with previously characterized differing distribution patterns along the coast of Oman, where two seas with distinct environmental characteristics border each other with a sharp biogeographical turnover. Moderate resolution models are based on a cropped subset of the macroecological data presented in chapter 4. These are then compared to models based on genuine high-resolution macroecological and habitat layers derived from 10 mosaiced Landsat 7 ETM+ scenes. First, we assess the potential of Landsat imagery as a novel environmental data source for regional marine distribution modeling. Second, we evaluate the influence of differences in spatial scale between both environmental datasets on model outcomes in different environments, and compare our results to outcomes from similar terrestrial studies.

In **chapter 6**, we investigate the feasibility of satellite-based local shallow habitat mapping, with a focus on discerning macroalgal communities, to assess temporal turnover in a highly seasonal environment. To this end, we compare high-resolution multispectral Landsat 7 ETM+ imagery from the Arabian Sea with high-resolution superspectral PROBA/CHRIS imagery through supervised classification and vegetation indices. Next, in order to deal with mixed pixels characteristic of heterogeneous environments, we propose a novel sub-pixel approach by using the Maxent distribution modeling algorithm to calculate a habitat class probability surface in spectral space which is then analysed in its spatial context. We discuss the results from these approaches in the framework of seasonally changing dominances between coral and algae.

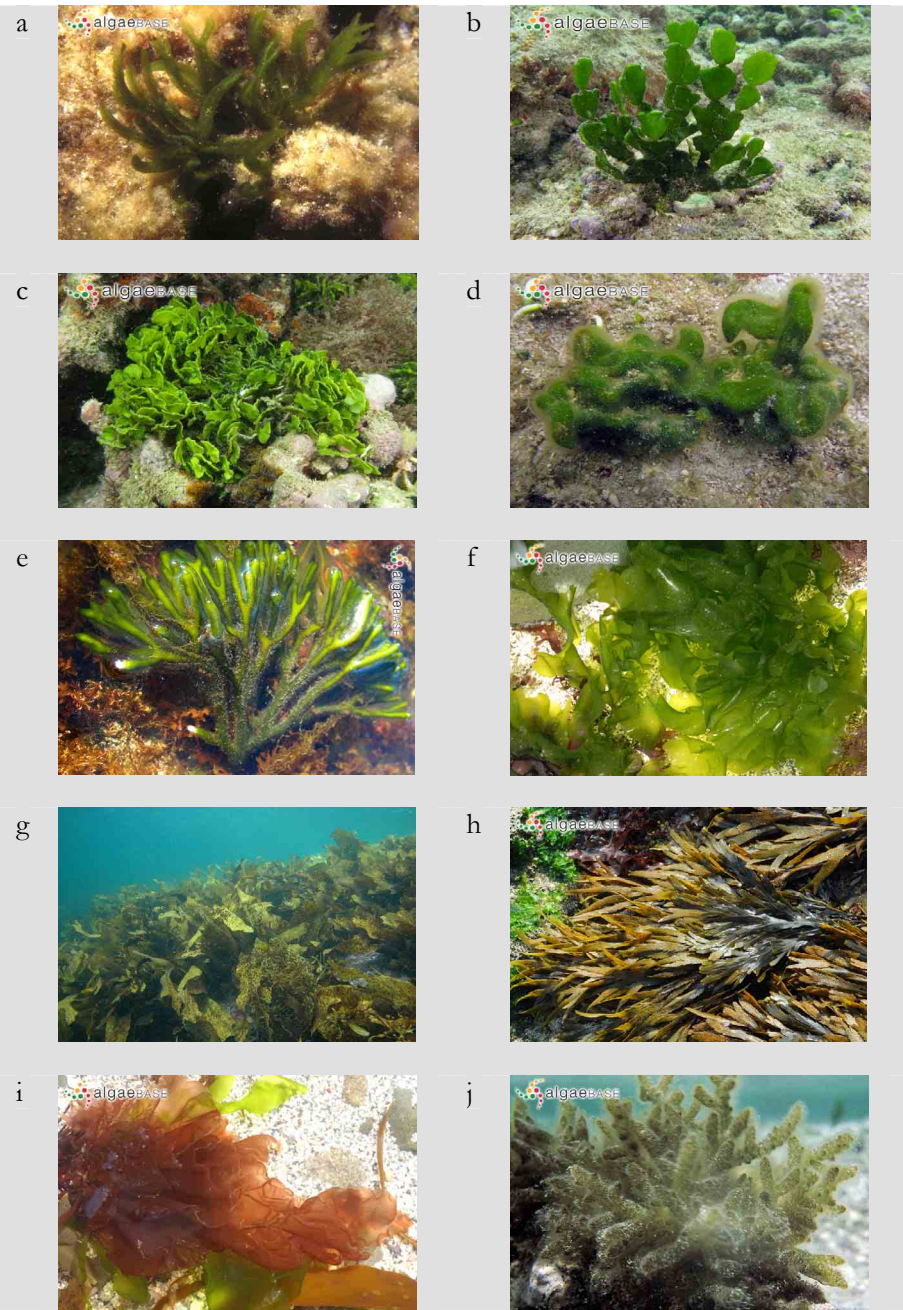
In **chapter 7**, we demonstrate a low-cost very high-resolution imaging tool to map and monitor intertidal communities, with a focus on seaweeds. We use a modified near-infrared sensitive consumer-grade digital compact camera mounted on a kite to obtain aerial photographs, which are then mosaiced using computer vision software and further processed similar to multispectral imagery.

In a **general discussion** we review the particular techniques used in this thesis in light of conceptual frameworks and discuss the present-day achievements and limitations of the applied methodologies.

**Box 3. Species**

The macroalgal species used in case studies throughout this thesis are collected from the intertidal or subtidal using SCUBA to a depth of maximum 40 m. Overall, these species represent a range of environmental conditions, habitats, growth forms, life cycles and taxonomic groups, detailed in the respective chapters. While demonstrating the wide applicability of the presented methods, the main reasons for inclusion of these species were availability of georeferenced data and confidence of identification.

<i>Taxon (following AlgaeBase)</i>	<i>Chapter - Plate</i>
Chlorophyta	
Bryopsidophyceae	
Bryopsidales	
Bryopsidaceae	
<i>Pseudobryopsis</i> Berthold emend. Henne et Schnetter spp.	Ch. 2 - Pl. a
<i>Trichosolen</i> Montagne emend. Henne et Schnetter spp.	Ch. 2
Halimedaceae	
Genus <i>Halimeda</i> J.V. Lamourou	
Section <i>Halimeda</i> spp.	Ch. 3
<i>Halimeda discoidea</i> Decaisne	Ch. 5 - Pl. b
Section <i>Micronesicae</i> spp.	Ch. 3
Section <i>Opuntia</i> spp.	Ch. 3 - Pl. c
Section <i>Pseudo-opuntia</i> spp.	Ch. 3
Section <i>Rhipsalis</i> spp.	Ch. 3
Codiaceae	
<i>Codium arabicum</i> Kützinger	Ch. 1 - Pl. d
<i>Codium fragile</i> subsp. <i>fragile</i> (Suringar) Hariot	Ch. 4 - Pl. e
Ulvophyceae	
Ulvales	
Ulvaceae	
<i>Ulva</i> Linnaeus spp.	Ch. 7 - Pl. f
Heterokontophyta	
Phaeophyceae	
Fucales	
Sargassaceae	
<i>Nizamuddinia zanardinii</i> (Schiffner) P.C. Silva	Ch. 5,6 - Pl. g
Fucaceae	
<i>Fucus</i> Linnaeus spp.	Ch. 7 - Pl. h
Rhodophyta	
Bangiophyceae	
Bangiales	
Bangiaceae	
<i>Porphyra yezoensis</i> Ueda	Ch. 1
<i>Porphyra</i> C. Agardh spp.	Ch. 7 - Pl. i
Florideophyceae	
Ceramiales	
Rhodomelaceae	
<i>Tolytiocladia glomerulata</i> (C. Agardh) F. Schmitz	Ch. 6 - Pl. j



## AUTHOR CONTRIBUTIONS

**Chapter 1.** The review was outlined, performed and written by Klaas Pauly. Olivier De Clerck commented on the manuscript.

**Chapter 2.** Klaas Pauly carried out specimen identifications, study design and analyses, as well as writing the manuscript. The initial observation of the bloom, sample collection and suggestion of a wider global pattern was made by Barry Jupp. Olivier De Clerck commented on the manuscript

**Chapter 3.** The first three authors have equally contributed to the study and manuscript. Heroen Verbruggen designed the phylogenetic and ancestral state analyses and revised the manuscript for these topics. Lennert Tyberghein carried out all analyses and wrote an initial draft. Klaas Pauly designed the GIS analyses, niche modeling and model evaluation and revised the manuscript for these topics. The discussion was written in collaboration between the three first authors. Other authors have carried out technical work and commented on the manuscript.

**Chapter 4.** The conceptual framework of the dataset was designed by the first three authors. Lennert Tyberghein carried out all GIS, interpolation and data integration analyses and wrote the manuscript. Heroen Verbruggen provided biological information for the case study and commented on the manuscript. Klaas Pauly assisted in model design and evaluation of the case study and commented on the manuscript. Other authors have also commented on the manuscript and provided biological information.

**Chapter 5.** Klaas Pauly designed the study, carried out field work and all data analyses and wrote the manuscript. Lennert Tyberghein commented on Bio-ORACLE data integration and Olivier De Clerck commented on the biological aspect of the study.

**Chapter 6.** Klaas Pauly designed the study, carried out all field work and data analyses and wrote the manuscript. Rudi Goossens suggested the use of PROBA/CHRIS and provided access to imagery and advice on data processing. Olivier De Clerck commented on the manuscript.

**Chapter 7.** Klaas Pauly designed the study, carried out field work and all data analyses and wrote the manuscript. Olivier De Clerck commented on the manuscript.

## CHAPTER 2

# MODELING THE DISTRIBUTION AND ECOLOGY OF *TRICHOSOLEN* BLOOMS ON CORAL REEFS WORLDWIDE

Klaas Pauly, Barry P. Jupp<sup>1</sup> & Olivier De Clerck

<sup>1</sup> GEO-Resources Consultancy  
PO Box 1127, PC 111  
Muscat, Sultanate of Oman

Adapted from published article: Pauly K, Jupp BP & De Clerck O (2011) Modeling the distribution and ecology of *Trichosolen* blooms on coral reefs worldwide. *Marine Biology*, DOI 10.1007/s00227-011-1729-0.



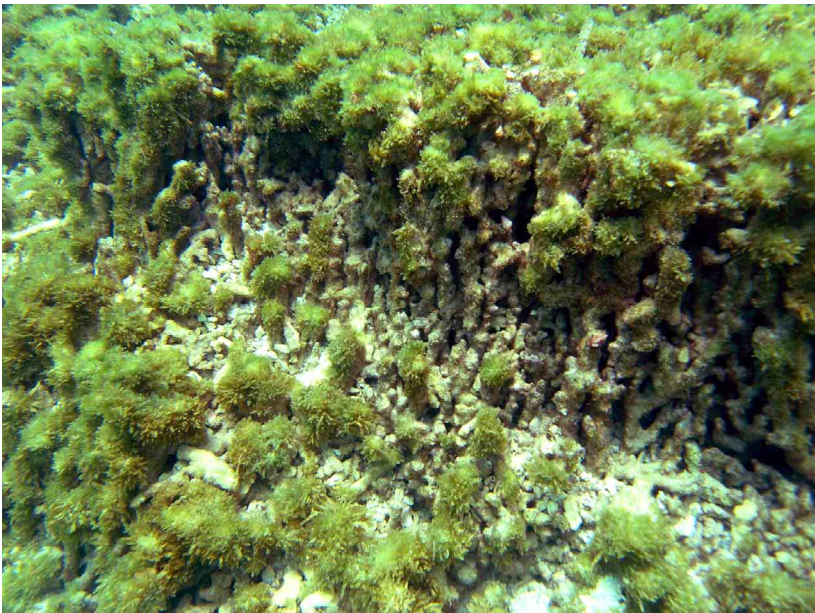
**ABSTRACT**

Worldwide blooms of the green alga *Trichosolen* have been reported on damaged coral reefs following catastrophic events. However, the global distribution of *Trichosolen* and the factors triggering such blooms remained elusive because of a paucity of occurrence records. This study presents a presence-only niche modeling approach to map the potential distribution and delineate bloom risk areas as well as to identify environmental response optima for non-blooming occurrences and blooms. The modeled suitability map revealed a pantropical to subtropical distribution, while high suitability values delineated bloom risk areas including important tropical reef systems where *Trichosolen* has not yet been reported from. While both blooms and non-bloom occurrences show a strong preference for high temperatures, blooms responded better to broader nutrient ranges than non-blooms.

## INTRODUCTION

Coral communities are subject to extensive mechanical damage from both natural and anthropogenic causes, including cyclonic wave action, earthquakes and tsunamis, *Acanthaster* outbreaks, construction, ship groundings and removals, dynamite fishing, boat anchoring, diving and coral removal. The resulting lesions and cleared substrates on coral reef systems are susceptible to quick overgrowth by rapid colonizers such as benthic diatoms, cyanobacteria and turf algae (Rogers et al., 1991; Russ & McCook, 1999; Schroeder et al., 2008; Titlyanov et al., 2008). These rapid colorizations (sometimes involving only one taxonomic entity) either mark the start of a complete phase shift (e.g. Diaz-Pulido et al., 2007), or the opportunistic algae disappear over time, resulting in the restoration of healthy coral reefs.

The present paper focuses on remarkable pantropical monotypic blooms on coral reefs following catastrophic events, involving species in the coenocytic green algal genus *Trichosolen* (Bryopsidales, Chlorophyta; figure 1). This genus belongs to the *Pseudobryopsis/Trichosolen* (PT) complex.



**Figure 1.** Mechanically damaged *Pocillopora damicornis* colonies at Fahal Island (Muscat, Oman) covered by a *Trichosolen* sp. bloom on 11 July 2007, five weeks after Cyclone Gonu in the Gulf of Oman. The picture depicts an area 1m across.

Species belonging to *Trichosolen* have often been attributed to *Pseudobryopsis* or vice versa, based on variable combinations of morphological and anatomical characters. The complex was revised by Henne & Schnetter (1999), who segregated the genera based on the presence or absence of a basal plug in the gametangia and pyrenoids in the chloroplast. Since *PT* species usually grow inconspicuously and are characterized by a partly microscopic life-cycle, worldwide distribution patterns of both *Pseudobryopsis* and *Trichosolen* remained largely elusive. Based on the revision of Henne and Schnetter (1999), distribution patterns emerged on a generic level, with *Pseudobryopsis* characterized by a subtropical and warm-temperate distribution accounting for the highest latitudes of the *PT* complex (including the Mediterranean Sea, Japan, eastern South Africa and southern Australia), while *Trichosolen* appears to have a more tropical worldwide distribution. Hence, most of the *PT* species reported to grow on coral reef systems are *Trichosolen*, some of which were exclusively found on coral. Moreover, within the complex, only *Trichosolen* has been documented to smother corals following physical damage. Woodley et al. (1981) were the first to report extensive blooms of *T. duchassaingii* one week after hurricane Allen hit Jamaican coral reefs in August 1980. Similarly, Littler et al. (1987) observed rapid colonisation by *T. molassensis* (a previously undescribed species) following the grounding of the freighter Wellwood on Molasses Reef in the Key Largo National Marine Sanctuary (Florida, US) in August 1984. Later, Littler and Littler (1999) described an extensive bloom of *Trichosolen* sp. on impacted corals on the Great Astrolabe Reef (Fiji) within 2-3 days after cyclone Gavin in March 1997. Lastly, a remarkably similar *Trichosolen* sp. bloom occurred on broken coral colonies and rubble immediately following cyclone Gonu in the Gulf of Oman in June 2007 (Jupp 2007, internal report but see Foster et al. (2008) and Wang & Zhao (2008) for reports on the cyclone and its biological consequences). All the blooms persisted for at least several weeks, and are in agreement with Littler & Littler's (1999) observations on the Florida and Fiji blooms: they are (1) monotypic, (2) widespread in time and space, (3) following severe physical impact, (4) persisting for weeks regardless of herbivore abundance, and (5) previously unknown from the sites despite elaborate prior floristic work.

Besides a sudden availability of suitable substrate, nutrient inputs originating from land-based run-off or sediment resuspension have been

suggested as possible factors in the development of rapid overgrowths (Steneck & Dethier, 1994; Condie, 2009). Overall however, little (quantitative) information is available about the factors triggering such blooms because of a paucity of observations with relevant environmental data. This paper provides insights in the macroecological niche and the potential distribution of both blooming and non-blooming *Trichosolen* species growing on (damaged) coral reef systems, using a maximum entropy ecological niche modeling (ENM) technique based on known occurrence records and environmental layers related nutrient availability, sea surface temperature (SST) and coral distribution. An attempt is made to delineate key bloom risk areas and to identify environmental response optima for both non-bloom occurrences and bloom cases.

## **MATERIALS AND METHODS**

### **SPECIES OCCURRENCE DATA**

Literature and, where available, databases of publicly deposited specimens were searched for worldwide occurrence records of *Trichosolen* species sensu Henne & Schmitter (1999) reported to grow on coral reef systems, both in natural conditions and after disturbances. Subsequently, Google Earth was used to assign lat/long coordinates matching the locality descriptions as closely as possible in the case where no coordinates were mentioned. Occurrence records lacking sufficient geographical accuracy (e.g. only mentioning a country) were excluded for the ENM. The resulting 21 worldwide occurrence records (17 non-blooming records and 4 bloom cases, table 1) were manually checked for georeferencing errors against an administrative maritime boundaries map in Esri ArcGIS 9.2 and manually relocated where necessary to fall within sea areas, given the resolution of environmental data. Other data points such as *Trichosolen* records found on substrates outside coral reef systems or *Pseudobryopsis* records on coral reefs (representing a minority of the *PT* records previously often misreported as *Pseudobryopsis* and *Trichosolen* respectively) were retained for display in figure 2. A complete list of georeferenced *PT* records is shown in supplement S1.

**Table 1.** Records of *Trichosolen* occurring on coral reef systems used in the Maxent models. Bloom records are indicated in bold. The references are included in the general list at the end of this thesis.

Species (H&S '99)	Reported as	Source	Year	Country	Lat	Long
<i>T. duchassaingii</i>	<i>Bryopsis</i>	Taylor	1928	Florida	24.63	-82.93
<i>T. duchassaingii</i>	<i>Bryopsis</i>	Taylor	1960	Bahamas	24.5	-77.7
<i>T. duchassaingii</i>	<i>Bryopsis</i>	Taylor	1960	Barbados	13.2	-59.64
<i>T. duchassaingii</i>	<i>Trichosolen</i>	Richardson	1975	Trinidad	10.76	-61.44
<b><i>T. duchassaingii</i></b>	<b><i>Trichosolen</i></b>	<b>Woodley et al.</b>	<b>1981</b>	<b>Jamaica</b>	<b>18.47</b>	<b>-77.41</b>
<i>T. duchassaingii</i>	<i>Bryopsis</i>	Suárez	2005	Cuba	21.5	-79
<i>T. gracilis</i>	<i>Pseudobryopsis</i>	Womersley and Bailey	1970	Solomon	-8.75	158.07
<i>T. gracilis</i>	<i>Pseudobryopsis</i>	Coppejans et al.	2001	Papua N. Guinea	-4.77	145.7
<b><i>T. molassensis</i></b>	<b><i>Trichosolen</i></b>	<b>Littler et al.</b>	<b>1987</b>	<b>Florida</b>	<b>25.01</b>	<b>-80.37</b>
<i>T. parva</i>	<i>Pseudobryopsis</i>	Pham-Hoàng	1969	Vietnam	12.21	109.22
<i>T. parva</i>	<i>Pseudobryopsis</i>	Cribb	1984	Australia	-23.5	151.28
<i>T. solomonensis</i>	<i>Pseudobryopsis</i>	Womersley and Bailey	1970	Solomon	-8.45	158.13
<i>T. solomonensis</i>	<i>Pseudobryopsis</i>	Egerod	1975	Thailand	8.088	98.29
<i>T. solomonensis</i>	<i>Pseudobryopsis</i>	Egerod	1975	Thailand	7.803	98.41
<i>T. solomonensis</i>	<i>Pseudobryopsis</i>	Egerod	1975	Thailand	7.727	98.79
<i>T. solomonensis</i>	<i>Pseudobryopsis</i>	Egerod	1975	Thailand	7.745	98.74
<i>T. solomonensis</i>	<i>Pseudobryopsis</i>	Coppejans et al.	2001	Papua N. Guinea	-4.18	144.87
<i>T. solomonensis</i>	<i>Pseudobryopsis</i>	Coppejans et al.	2001	Papua N. Guinea	-5.13	145.82
<i>T. sp.</i>	<i>Trichosolen</i>	De Clerck & Coppejans	1996	Jubail	27.32	49.53
<b><i>T. sp.</i></b>	<b><i>Trichosolen</i></b>	<b>Littler &amp; Littler</b>	<b>1999</b>	<b>Fiji</b>	<b>-19.1</b>	<b>178.4</b>
<b><i>T. sp.</i></b>	<b><i>Trichosolen</i></b>	<b>Jupp</b>	<b>2007</b>	<b>Muscat</b>	<b>23.68</b>	<b>58.5</b>

## ENVIRONMENTAL DATA

Global SST and chlorophyll-a data (CHL, as a proxy for nutrient concentrations) available from the Aqua-MODIS satellite were assumed to be the major relevant ecophysiological variables driving survival and growth of *Trichosolen*. Three-monthly averaged (seasonal) level-3 pre-processed gridded maps for SST and CHL were downloaded at 5 arcmin (approximately 9 km) spatial resolution from OceanColor Web (<http://oceancolor.gsfc.nasa.gov>) and were subsequently cropped to include the highest latitudes at which *Trichosolen* is found to grow on coral reef systems (33°N-33°S). Raw pixel values (16bit) were recalculated to °C and mg m<sup>-3</sup> using the respectively linear and exponential scaling equations provided in the metadata of the downloaded files. A uniform landmask was applied to all maps, masking out all freshwater bodies. These temporally averaged maps were not used as environmental variables for the modeling process as such. Instead, the following biologically meaningful variables

were included, calculated using Clark Labs Idrisi Andes and Perl scripts: minimum, maximum, range and mean CHL (CHLmin, CHLmax, CHLrng and CHLavg) and SST (SSTmin, SSTmax, SSTrng and SSTavg). Additionally, as Schils & Wilson (2006) identified the average SST of the three warmest seasons as a powerful variable in explaining macroalgal biogeographical patterns, this variable was also included (SSTavg3warm). Two approaches were taken to reduce data redundancy using Idrisi Andes, as this can bias interpretation of the response curves in the niche modeling step. First, an unstandardized principal component analysis (PCA) was performed on the variables, and the first 3 PC grids (accounting for 99.58% of the overall variance) were used as environmental variable input in the ENM. This approach was adopted to produce the most accurate ENM maps while keeping the loss of information at a minimum (Verbruggen et al., 2009). Second, a subset of variables was selected based on a correlation analysis (Riordan & Rundel, 2009), and this subset was used to model environmental response curves. This approach may lead to a reduction of data but allows for easy interpretation of the resulting response curves, as opposed to using principal components. SSTavg3warm and the two remaining least correlated variables per category (SSTmin, SSTrng, CHLmin and CHLrng; Pearson correlation,  $0.48 < r < 0.8$ ) were retained for analyses in order to estimate the position of the niche in environmental space. Other potentially relevant variables available from OceanColor Web such as photosynthetically available radiation (PAR) and diffuse attenuation (DA, a measure of turbidity) are strongly correlated ( $r > 0.8$ ) with SST and CHL, respectively, and were therefore not included.

Using ArcGIS 9.2, environmental data were extracted for all *Trichosolen* occurrences found on coral systems for subsequent multivariate analyses. Kruskal-Wallis tests were carried out in Statsoft Statistica 7 using the subset of five least-correlated variables as independent variables in order to evaluate different ecological characteristics of blooms and non-bloom records.

## ENM

Because of the rarity, small size and semi-cryptic life-cycle of *Trichosolen* and the resulting lack of reliable absence data, a presence-only niche modeling

algorithm was required. Therefore, Maxent v3.3.2, a software implementation based on the maximum entropy principle (Phillips et al., 2006; Phillips & Dudík, 2008) was used to model the potential distribution of *Trichosolen* on coral reefs and to assess the species' response to the environmental variables. The maximum entropy algorithm is a general-purpose method to infer a probability distribution from incomplete information. This probability distribution is most uniformly spread out (showing the highest degree of entropy) given the constraints represented by the environmental data observed at the given occurrence localities. Besides the raw environmental data input, Maxent also calculates transformations and interactions of these variables (so-called features) to constrain the distribution. Eventually, unlike other algorithms, Maxent converges to a unique, least-biased solution in estimating the unknown distribution. The resulting probability distribution is then translated from ecological space into geographical space. In the so-called logistic format, the output map shows pixel values ranging from 0 to 1, indicating niche suitability. Elith et al. (2006) have shown Maxent to perform well in comparison to other niche modeling algorithms, and Pearson et al. (2007) demonstrated that Maxent was able to model species' potential distributions based on as few as 5 to 25 occurrence records more accurately than other presence-only modeling algorithms such as GARP.

Maxent uses a large random sample of background environmental data (potentially including occurrence pixels, as opposed to generating pseudo-absences) to model the potential niche relative to the available environment. Most studies let Maxent randomly take 10000 points from the entire study area (a default option in the software). However, Phillips et al. (2009) pointed out that presence-only occurrence data often show a spatial sampling bias. This is often translated in an environmental sampling bias which, when combined with an evenly distributed background sample, leads to inaccurate models. To account for spatial sampling bias, background points with the same kind of bias as the occurrence records were selected: using Hawth's analysis tools for ArcGIS (Beyer, 2004), 10000 background points were randomly sampled only from the first two pixels adjacent to the landmask in the study area, so as to approximate a typical coastal sampling bias. Other model training parameters were left at their default settings (convergence threshold =  $10^{-5}$ , maximum iterations = 500, automatic

feature selection). Response curves of the modeled species to each of the 5 selected variables were also modeled by the Maxent algorithm.

The resulting continuous probability map based on the model using principal components as environmental variables was converted to a discrete map using two different thresholds to facilitate ecological interpretation. A 10<sup>th</sup> percentile training presence threshold was used to visualize the potential distribution of the macroecological niche of *Trichosolen* species growing on coral reef systems, while accounting for some potential uncertainty in occurrence data input. In order to delineate bloom risk areas, a more stringent fixed threshold of 0.8 was applied to the logistic suitability map. This threshold was chosen because VanDerWal et al. (2009) have shown that a suitability value at a given pixel is correlated with the maximum achievable biomass at that pixel, and as such a high suitability value represents a pixel where a bloom could potentially be sustained. In order to refine the macroecological niche distribution map to incorporate habitat suitability, the resulting discrete model output map was overlaid with 16700 coral distribution data points down to 40 m depth downloaded from ReefBase (Tupper et al., online publication) for which a 3-pixel buffer was created prior to raster conversion.

While a suitability map was modeled for all *Trichosolen* occurrence records on coral reefs, response curves were modeled both jointly and separately for blooms and non-blooms to gain insight in their ecology, i.e., blooms, non-blooms and all *Trichosolen* records on coral reef systems were treated as distinct “species” in the modeling process. The model based on all occurrences on coral systems was created to make full use of all possible data in examining the worldwide potential niche, as blooms necessarily develop from already present (although previously unnoticed) life stages. In contrast, the separate models for blooms and non-bloom cases were created to evaluate whether conditions specific for developing blooms could be distilled from the data.

## MODEL EVALUATION

Models were run using approximately 75% of the occurrence records as training points and 25% as independent test points. Threshold-independent receiver operating characteristic (ROC) analysis allowed for model



performance evaluation. The ROC curve plots the proportion of correctly predicted presences against the fractional area predicted present (in the case of presence-only models). The measure of predictive performance is the area under the ROC curve (AUC), varying between 0 and 1, where values above 0.5 (the area under the null expectations line) are indicative of models better than random. A four-fold cross-validation was implemented in order to obtain a standard deviation for the AUC values.

## RESULTS

### *TRICHOSOLEN* POTENTIAL DISTRIBUTION

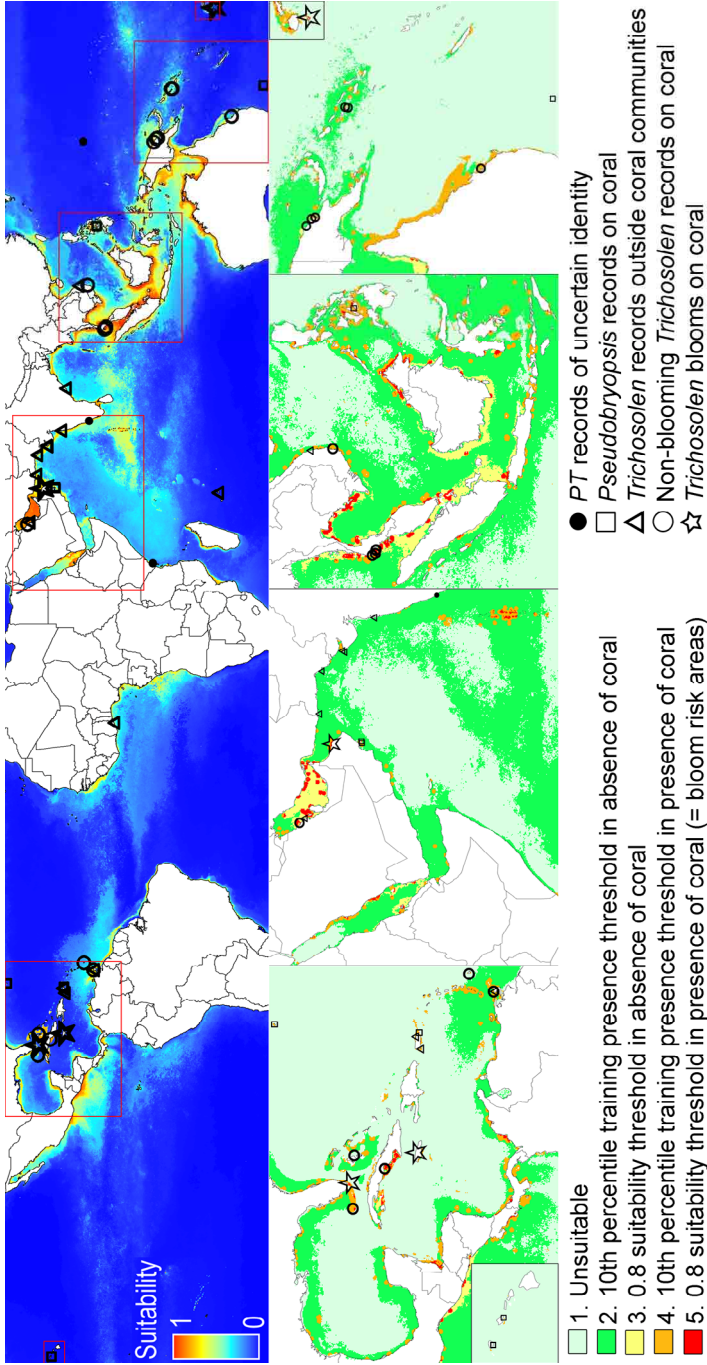
The AUC values of the resulting Maxent model based on all *Trichosolen* occurrences on coral reef systems and principal components as environmental variables (table 2) indicate adequate model performance without overfitting to training data. The map shows a worldwide tropical to subtropical distribution of the macroecological niche (figure 2, global map), excluding cool waters (e.g. the Arabian Sea). When applying the 10<sup>th</sup> percentile training presence threshold (suitability value = 0.296) to the cross-validation replicate with the highest test AUC (0.969), the model was found to predict significantly better than random ( $p = 0.00003$ ), with no records used for modeling falling outside the predicted suitable area (figure 2 selected regions: all open circles and stars fall within one of the four suitably colored areas). Although most of the macroecologically suitable areas are not located on coral reef systems with suitable substrate (dark green and yellow in figure 2), some other *Trichosolen* records have been found there (open triangles) as well as *PT* records of uncertain generic identity (filled circles). Narrowing down the distribution of the macroecological niche to areas where coral reef systems are present (orange and red in figure 2) results in the identification of several reef areas where *Trichosolen* has not yet been recorded, such as Costa Rica and Panama, the Red Sea and the Maldives, northeast Malaysia and the Philippines and Queensland (Australia). Areas with a suitability value exceeding 0.8 combined with coral availability (red pixels) represent the true bloom risk areas and cover all the recorded blooms.

**Table 2.** Area under the ROC curve (AUC) values of all Maxent model runs.

Model	Variables	N(train)	AUC train	N(test)	AUCtest
		per crossval run	mean ( $\pm$ SD)	per crossval run	mean ( $\pm$ SD)
All <i>Trichosolen</i> on coral	PCA	16	0.958 ( $\pm$ 0.007)	5	0.945 ( $\pm$ 0.02)
All <i>Trichosolen</i> on coral	selected	16	0.942 ( $\pm$ 0.012)	5	0.922 ( $\pm$ 0.067)
Non-blooms	selected	13	0.916 ( $\pm$ 0.013)	4	0.837 ( $\pm$ 0.16)
Blooms	selected	3	0.845 ( $\pm$ 0.091)	1	0.804 ( $\pm$ 0.114)

ENVIRONMENTAL RESPONSE CURVES FOR *TRICHOSOLEN* ON CORAL

Kruskal-Wallis tests failed to reveal a significant difference between environmental variables for bloom or non-bloom occurrences found on coral reef systems (table 3). The Maxent model based on all *Trichosolen* occurrences on coral reefs and the subset with the 5 least correlated environmental variables performed only slightly worse for the training and test records, hence indicating a marginally lower predictive performance compared to the model based on principal components (table 2). The separate models for blooms and non-blooms have progressively lower AUC values. Although still indicating a better performance than random predictions, this effect is probably due to the low number of occurrence records (only 4 in the case of blooms). Rather than producing accurate maps, this approach is used here to analyze environmental variable response curves, summarized in figure 3. Response curves for SSTavg3warm, SSTmin and SSTrng exhibit a similar pattern across the models for all *Trichosolen* on coral systems, non-blooming occurrences and blooms, in favor of the warmest temperatures but with a tolerance for high ranges. Models for all *Trichosolen* records on coral reef systems and non-blooms correspond in the response curves for CHLmin and CHLrng, suggesting high support for the most oligotrophic conditions, corresponding to Jerlov class I oceanic waters. However, the model for blooms shows no response to CHLmin, while indicating an elevated response for an elevated CHLrng.



**Figure 2.** Niche suitability map based on all *Trichosolen* occurrences on coral and principal components as environmental variables. The global continuous probability map shown above represents the median suitability values derived from the 4 cross-validation runs. Selected regions indicated by the red boxes are shown below from west to east, converted to discrete maps based on the 10th percentile training presence threshold, the 0.8 threshold and coral distribution data. Note that some occurrence symbols may overlap each other or cover colored pixels due to their size in this representation.

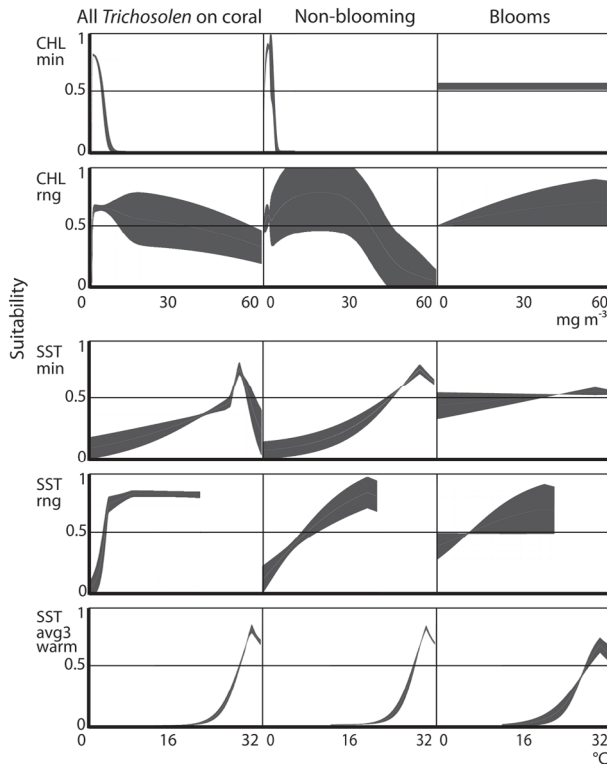
**Table 3.** Results of Kruskal-Wallis tests showing the difference between the environmental characteristics of all *Trichosolen* on coral reef systems, blooms and non-blooming occurrences.

Variable	All <i>Trichosolen</i> on coral (N=21)	Non-blooms (N=17)	Blooms (N=4)	p
	mean ( $\pm$ SD)	mean ( $\pm$ SD)	mean ( $\pm$ SD)	
CHL <sub>min</sub> (mg m <sup>-3</sup> )	0.51 ( $\pm$ 0.71)	0.53 ( $\pm$ 0.78)	0.37 ( $\pm$ 0.21)	0.98
CHL <sub>rng</sub> (mg m <sup>-3</sup> )	0.97 ( $\pm$ 1.43)	0.9 ( $\pm$ 1.41)	1.39 ( $\pm$ 1.68)	0.84
SST <sub>min</sub> (°C)	25.94 ( $\pm$ 3.07)	26.23 ( $\pm$ 3.29)	24.72 ( $\pm$ 1.57)	0.39
SST <sub>rng</sub> (°C)	4.61 ( $\pm$ 2.93)	4.45 ( $\pm$ 3.08)	5.44 ( $\pm$ 2.17)	0.44
SST <sub>avg3warm</sub> (°C)	29.11 ( $\pm$ 2.74)	28.56 ( $\pm$ 2.09)	32.01 ( $\pm$ 4.17)	0.36

## DISCUSSION

In recent years, ENM has been suggested as an effective, promising means to investigate species' biogeographical patterns and potential distribution, especially when occurrence records are scarce and reliable absence data cannot be obtained (Soberón & Peterson, 2004; Bentlage et al., 2009). This makes ENM a suitable tool for the investigation of ephemeral aquatic blooms which, by nature, offer limited sampling possibilities. Ephemeral macroalgal blooms on coral reefs have been reported to occur throughout the tropics, but the conditions enabling the formation of blooms often remain unclear (Burgess, 2006; Vroom et al., 2009). The bloom-forming species are usually known to be present in low abundances prior to the actual bloom formation. By contrast, *Trichosolen* bloom records were the only data source for some regions despite earlier floristic work around the bloom sites, probably because the present populations were in the microscopic part of the life cycle.

The failure of the Kruskal-Wallis tests to distinguish between blooms and non-blooming *Trichosolen* occurrences on coral reef systems based on the selected environmental variables illustrates the limitations of classical biogeographical approaches when few data are available. In contrast, different environmental responses are apparent for modeled blooms and non-blooming occurrences, while AUC values are acceptable for all these models. The apparent optima in the response curves are in agreement with the observed values (table 3), and support these data in a more reliable and quantitative way.



**Figure 3.** Response curves to the least-correlated environmental variables calculated by Maxent. Niche suitability ranges from 0 to 1 on the y-axis, while the range of environmental variables is shown on the x-axis. Response curves in the left column are based on all *Trichosolen* records on coral reef systems; the centre column shows response curves based on non-blooming records only; models based on bloom records only are shown in the right column. Environmental variable acronyms are listed in the materials & methods section. The light grey centre line represents the median value derived from the 4 crossvalidation runs, while the dark grey band delineates the standard deviation.

A first condition for blooms to develop is a sudden availability of suitable substrate (dead or cleared surfaces on coral reef systems) through a mechanically damaging event. Second, the peak at low concentrations in the CHLmin response curve for the models based on all occurrences and non-blooming *Trichosolen* on coral reefs, combined with the response optima for a low to intermediate CHLrng, demonstrates a preferred occurrence of *Trichosolen* in clear, oligotrophic waters. These are mostly tropical waters, since the response curve of all models to SSTmin favors the highest values. On the other hand, the lack of response for bloom cases to CHLmin and an increasing tolerance towards elevated CHLrng values indicate that blooms

are more likely to occur in areas where chlorophyll concentrations can peak to extreme values (beyond seasonal fluctuations, i.e. on rare events), whereas non-blooming *Trichosolen* occurrences are more likely to be restricted to areas with continuously low chlorophyll concentrations. Moreover, the oceanographic observations by Wang & Zhao (2008) suggest the importance of sudden nutrient inputs caused by cataclysmic events, e.g. by sediment resuspension. In fact, peak nutrient concentrations during and shortly after these events might be an important factor besides available substrate and a reduction in herbivore density for *Trichosolen* to develop into a bloom. Thirdly, since the response curves of all *Trichosolen* models to SSTavg3warm favor the highest temperatures, *Trichosolen* growth and blooms are most likely to occur in areas characterized by an average temperature during the three warmest seasons that exceeds 28 °C (characteristic of tropical floras, see Schils & Wilson (2006)). Moreover, the confidence intervals of SSTavg3warm are very small, indicating the explanatory power of this variable. Since all the blooms were exclusively observed in the hottest season, this also suggests that this is the risk season. However, since the response of *Trichosolen* blooms to SSTmin is less pronounced than that of normal occurrence events, it could be hypothesized that blooms may also develop following a catastrophic event with sudden nutrient input even when the SST is suboptimal, whereas otherwise the background population (present in the microscopic part of the life cycle) would not develop in macroscopic growth before optimal temperatures are reached when the water column is oligotrophic.

The discrete maps in figure 2 show the geographical distribution of the macroecological niche of *Trichosolen* based on the thresholds discussed above, while isolating areas within its niche that also feature a suitable habitat (shallow coral substrate). Only areas with both a higher than 0.8 niche suitability value and the presence of coral reef systems are delineated as bloom risk areas which could be prioritized for targeted bloom monitoring when damaging events occur on the reefs. Likewise, other macroecologically suitable areas coinciding with coral reef presence (orange) might be used to guide exploration of new populations to expand biogeographical knowledge of the genus *Trichosolen*. From these maps, it is also apparent that habitat suitability is often controlled at a fine spatial scale, for which the resolution of the present model may not always be sufficient

(illustrated by isolated suitable pixels around reef islands). For guidance of field work, it is important to note that biotic and abiotic microhabitat variables such as wave exposure and grazing pressure could evidently not be incorporated in a modeling effort at a global scale. To further refine similar modeling efforts, coral reef impact variable maps might be included at different scales (post-hoc); these could include meteorological variables such as cyclone frequency and intensity, shipping intensity as a proxy for ship grounding risk (Halpern et al., 2008), and *Acanthaster* distribution maps.

Besides the selection of the modeling algorithm, some measures were taken in order to optimize the modeling process based on a low number of occurrence data. This study addresses the algorithm's sensitivity to sample selection bias by adopting a target-group background selection which is likely to show the same environmental bias as in the species occurrence data. Table 4 shows that the spread of environmental data for the occurrence records reasonably matches the averaged spread of background environmental data in ecological space (although less for CHL extremes). Additionally, suitability maps were modeled using principal components as environmental variables to ensure model performance was not affected by data redundancy. Raw variables were used to calculate environmental response curves for pragmatic reasons in interpreting the results, but again data redundancy was avoided by selecting the least correlated variables. Burgess (2006) noted that ephemeral blooms on tropical reef systems might be far more widely spread in time and space than currently observed, and stressed that the paucity of distributional and water quality data hampered understanding of the factors related to these blooms.

**Table 4.** Overview of the spread in ecological space of occurrence records compared to the target-group selected background data.

Variable	All <i>Trichosolen</i> on coral (N=25)			Target-group background (N=10000)		
	min	max	mean ( $\pm$ SD)	min	max	mean ( $\pm$ SD)
CHLmin (mg m <sup>-3</sup> )	0.06	1.87	0.42 ( $\pm$ 0.48)	0.01	64	1.84 ( $\pm$ 3.01)
CHLrng (mg m <sup>-3</sup> )	0.04	8.8	1.05 ( $\pm$ 1.43)	0	64	3.03 ( $\pm$ 5.96)
SSTmin (°C)	18.68	29.12	25.94 ( $\pm$ 3.07)	5.92	30.59	24.61 ( $\pm$ 4.45)
SSTrng (°C)	1.68	14.59	4.66 ( $\pm$ 2.93)	0.48	23.19	5.25 ( $\pm$ 3.74)
SSTavg3warm (°C)	25.72	30.62	29.11 ( $\pm$ 2.74)	14.67	36	29.44 ( $\pm$ 3.35)

The approach outlined in this paper yielded quantitative ecological information based on few occurrence records and can potentially be used to shed new light on similar ephemeral blooms, which are known to be increasing in frequency and intensity with global change (Rabalais, 2009).

#### **ACKNOWLEDGEMENTS**

We are grateful to the Ministry of Environment and Climate Affairs (Muscat, Oman) and Five Oceans Environmental Services LLC (Muscat, Oman) for assistance in the field. Heroen Verbruggen, Frederik Leliaert, Wouter Willems and two anonymous reviewers are acknowledged for their useful comments on the manuscript. Heroen Verbruggen also provided Perl scripts for GIS analyses. Mark and Diane Littler kindly provided specimens. This research is embedded in project grant G.0142.05 from the Fund for Scientific Research – Flanders (FWO), Belgium.



## SUPPLEMENTARY MATERIAL

S1. Complete list of georeferenced *PT* records.

GENUS (H&S '99)	SPECIES	REPORTED AS	AUTHOR/HERBARIUM	YEAR	COUNTRY/REGION	ECOLOGICAL REMARKS	LAT	LONG
<i>Pseudobryopsis</i>	<i>Blomquistii</i>	<i>Pseudobryopsis</i>	Diaz-Piñferrer	1965	Puerto Rico	on dead coral, 10-15m	18.1100	-65.5870
<i>Pseudobryopsis</i>	cf. <i>hainanensis</i>	<i>Trichosolen</i>	Norris	1992	South Africa/KZN	in intertidal rock pools on reef	-26.8833	32.8833
<i>Pseudobryopsis</i>	cf. <i>P. mucronata</i>	<i>Trichosolen</i>	Norris	1992	South Africa/KZN	in intertidal rock pools on reef	-28.1333	32.5667
Uncertain	<i>duchassaingii</i>	<i>Trichosolen</i>	Vandenheede	1994	Kenya	on wall of lower intertidal rock pool	-4.0560	39.7060
<i>Trichosolen</i>	<i>duchassaingii</i>	<i>Trichosolen</i>	Woodley	1981	Jamaica	on shallow corals after hurricane	18.4700	-77.4100
<i>Trichosolen</i>	<i>duchassaingii</i>	<i>Bryopsis</i>	Taylor	1960	Bermuda	*no locality information*	32.3100	-64.8000
<i>Trichosolen</i>	<i>duchassaingii</i>	<i>Bryopsis</i>	Taylor	1928	Florida	*no ecological remarks*	24.6290	-82.9300
<i>Trichosolen</i>	<i>duchassaingii</i>	<i>Bryopsis</i>	Taylor	1960	Bahamas	*no locality information*	24.5000	-77.7000
<i>Trichosolen</i>	<i>duchassaingii</i>	<i>Bryopsis</i>	Taylor	1960	Barbados	*no locality information*	13.2000	-59.6400
<i>Trichosolen</i>	<i>duchassaingii</i>	<i>Bryopsis</i>	Suárez	2005	Cuba	*no locality information*	21.5000	-79.0000
<i>Trichosolen</i>	<i>duchassaingii</i>	<i>Trichosolen</i>	Richardson	1975	Trinidad	rare; intertidal rock pool; under shallow rock ledge	10.7600	-61.4350
<i>Trichosolen</i>	<i>gracilis</i>	<i>Pseudobryopsis</i>	Coppejans	2001	PNG	on coral rubble between reef flat & inner slope, 3m	-4.7667	145.7000
<i>Trichosolen</i>	<i>hainanensis</i>	<i>Pseudobryopsis</i>	Womersley	1970	Solomon	on shallow coral heads; on sandy stones, 11m	20.2730	58.7800
<i>Pseudobryopsis</i>	<i>hainanensis</i>	<i>Trichosolen</i>	GENT (unpubl.)	1999	Oman/Masirah East	on intertidal stem of <i>Rizophora</i> sp.	19.4800	110.8000
<i>Pseudobryopsis</i>	<i>hainanensis</i>	<i>Trichosolen</i>	Tseng	1936	China/Hainan	*no locality information*	36.2480	133.3760
<i>Pseudobryopsis</i>	<i>hainanensis</i>	<i>Trichosolen</i>	Yoshida	1995	Japan/OkI	*no locality information*	-31.5090	159.0650
<i>Pseudobryopsis</i>	<i>hainanensis</i>	<i>Pseudobryopsis</i>	Kraft	2007	Australia/LHI	in turf assemblage on coral bommies, 7-21m	-20.6620	116.7000
<i>Pseudobryopsis</i>	<i>hainanensis</i>	<i>Pseudobryopsis</i>	Huisman	2003	Australia/Western	*no locality information*	-28.7000	113.8000
<i>Pseudobryopsis</i>	<i>hainanensis</i>	<i>Pseudobryopsis</i>	Huisman	2000	Australia/Western	*no locality information*	-31.9900	115.5600
<i>Pseudobryopsis</i>	<i>hainanensis</i>	<i>Pseudobryopsis</i>	Huisman	2000	Australia/Western	*no locality information*	-30.2400	153.1790
<i>Pseudobryopsis</i>	<i>hainanensis</i>	<i>Pseudobryopsis</i>	Millar	1994	Australia/NSW	only on islands in NSW	18.5000	-66.0000
<i>Trichosolen</i>	<i>longipedicellata</i>	<i>Pseudobryopsis</i>	Blomquist	1961	Puerto Rico	*no ecological remarks*	18.0000	-67.2000
<i>Trichosolen</i>	<i>longipedicellata</i>	<i>Pseudobryopsis</i>	Blomquist	1948	Mauritius	free-floating after storm	-20.1667	57.4833
<i>Trichosolen</i>	<i>mauritanicus</i>	<i>Pseudobryopsis</i>	Boergesen	1979	Saudi-Arabia	on pier	26.7000	50.1000
<i>Trichosolen</i>	<i>mauritanicus</i>	<i>Trichosolen</i>	Basson	1987	Florida	on coral after ship grounding	25.0117	-80.3733
<i>Trichosolen</i>	<i>molassensis</i>	<i>Trichosolen</i>	Littler	1987	Florida	on coral after ship grounding	25.0117	-80.3733
<i>Trichosolen</i>	<i>micronatus</i>	<i>Pseudobryopsis</i>	Pham-Hoang	1969	Vietnam	*no locality information*	14.5800	109.0820
<i>Trichosolen</i>	<i>micronatus</i>	<i>Pseudobryopsis</i>	Umapaheswararao	1970	India/Visakhapatnam	on sand-covered rocks	17.7187	83.3365
<i>Trichosolen</i>	<i>micronatus</i>	<i>Pseudobryopsis</i>	Srinivasan	1969	India/S Gujarat	on walls of intertidal rock pools; in shallow subtidal	22.4790	69.0755
<i>Trichosolen</i>	<i>micronatus</i>	<i>Pseudobryopsis</i>	Srinivasan	1969	India/N Gujarat	on walls of intertidal rock pools; in shallow subtidal	22.2385	68.9577
<i>Trichosolen</i>	<i>micronatus</i>	<i>Pseudobryopsis</i>	Boergesen	1930	India/Bombay	in shallow water	18.8920	72.8080
<i>Trichosolen</i>	<i>micronatus</i>	<i>Trichosolen</i>	Shameel	1996	Pakistan/Makran	subtidal benthic drift	25.0865	61.8630
<i>Trichosolen</i>	<i>micronatus</i>	<i>Pseudobryopsis</i>	Shameel	1992	Pakistan/Karachi	benthic subtidal	24.8415	66.7745
<i>Pseudobryopsis</i>	<i>myura</i>	<i>Pseudobryopsis</i>	Wynne	2001	Oman/Dhofar	in sand in exposed bay, 8.5m	17.0052	54.0153
<i>Pseudobryopsis</i>	<i>myura</i>	<i>Pseudobryopsis</i>	Rayns	1955	Israel	in deep shaded calm tide pools	33.0649	35.1039
<i>Pseudobryopsis</i>	<i>myura</i>	<i>Trichosolen</i>	GENT (unpubl.)	1976	Turkey/South	*no ecological remarks*	36.2663	30.1420
<i>Pseudobryopsis</i>	<i>myura</i>	<i>Trichosolen</i>	Ballesteros	1990	Spain/Catalonia	*no locality information*	42.3190	3.3225
<i>Pseudobryopsis</i>	<i>myura</i>	<i>Trichosolen</i>	Meniez	1987	Tunisia	occasional; in April and May on subtidal rocks	33.7900	11.0600
<i>Pseudobryopsis</i>	<i>myura</i>	<i>Pseudobryopsis</i>	Feldmann	1929	France/Southeast	some years only, restricted to August-September	42.4760	3.1560
<i>Pseudobryopsis</i>	<i>myura</i>	<i>Trichosolen</i>	GENT (unpubl.)	1974	France/Southeast	*no ecological remarks*	43.0060	6.3765

S1. (continued)

GENUS (H&S '99)	SPECIES	REPORTED AS	AUTHOR/HERBARIUM	YEAR	COUNTRY/REGION	ECOLOGICAL REMARKS	LAT	LONG
<i>Pseudobryopsis</i>	<i>myura</i>	<i>Trichosolen</i>	GENT (unpubl.)	1975	France/Southeast	*no ecological remarks*	43.4118	6.8534
<i>Pseudobryopsis</i>	<i>myura</i>	<i>Trichosolen</i>	GENT (unpubl.)	1976	France/Corsica	*no ecological remarks*	42.5835	8.7250
<i>Pseudobryopsis</i>	<i>myura</i>	<i>Trichosolen</i>	Giaccone	1985	Italy/W Sicily	intertidal, subtidal	38.0200	12.5000
<i>Pseudobryopsis</i>	<i>myura</i>	<i>Trichosolen</i>	Giaccone	1985	Italy/N Sicily	intertidal, subtidal	38.4200	14.9500
<i>Pseudobryopsis</i>	<i>myura</i>	<i>Trichosolen</i>	Giaccone	1985	Italy/E Sicily	intertidal, subtidal	37.5600	15.1660
<i>Pseudobryopsis</i>	<i>myura</i>	<i>Trichosolen</i>	Giaccone	1985	Italy/E Sicily	intertidal, subtidal	37.0600	15.3000
<i>Pseudobryopsis</i>	<i>myura</i>	<i>Pseudobryopsis</i>	Haroun	2002	Canary Is	*no locality information*	28.1500	-15.5000
<i>Pseudobryopsis</i>	<i>myura</i>	<i>Pseudobryopsis</i>	Haroun	2002	Canary Is	*no locality information*	29.3000	-13.5000
<i>Pseudobryopsis</i>	<i>myura</i>	<i>Trichosolen</i>	Phillips	1997	Australia/Queensland	*no locality information*	-23.4830	151.2350
<i>Pseudobryopsis</i>	<i>myura</i>	<i>Trichosolen</i> sp.	GENT (unpubl.)	2006	Madeira	on large round boulders, 6-9m	32.6357	-16.9360
<i>Pseudobryopsis</i>	<i>cahuensis</i>	<i>Pseudobryopsis</i>	Abbott	2004	Hawaii/Oahu	in sand or on eroded coral, high intertidal to 5m	21.2530	-157.8100
<i>Pseudobryopsis</i>	<i>cahuensis</i>	<i>Pseudobryopsis</i>	GUAM (unpubl.)	1959	Hawaii/Oahu	growing on wave-washed bench with Gelidium	21.2687	-157.6920
<i>Pseudobryopsis</i>	<i>cahuensis</i>	<i>Pseudobryopsis</i>	GUAM (unpubl.)	1959	Hawaii/Oahu	in rapid current, 0-2m; on vertical concrete pier	21.2728	-157.8303
<i>Pseudobryopsis</i>	<i>cahuensis</i>	<i>Pseudobryopsis</i>	GUAM (unpubl.)	1959	Hawaii/Kauai	*no ecological remarks*	21.8785	-159.4779
<i>Pseudobryopsis</i>	<i>pambanensis</i>	<i>Pseudobryopsis</i>	Iyengar	1938	India/S	*no locality information*	9.2750	79.2040
<i>Pseudobryopsis</i>	<i>papillata</i>	<i>Trichosolen</i>	Papenfuss	1968	Egypt/Red Sea	*no locality information*	27.7500	34.2200
<i>Trichosolen</i>	<i>parva</i>	<i>Pseudobryopsis</i>	Pham-Hoang	1969	Vietnam	on surface of dead coral, 2-3m	12.2110	109.2170
<i>Trichosolen</i>	<i>parva</i>	<i>Pseudobryopsis</i>	Cribb	1984	Australia/Queensland	*no ecological remarks*	-23.5300	151.2800
<i>Trichosolen</i>	<i>retrorsa</i>	<i>Trichosolen</i>	John	1977	Ghana	on cobbles in waves	5.6750	0.1217
<i>Trichosolen</i>	<i>solomonensis</i>	<i>Pseudobryopsis</i>	Coppeljans	2001	PNG	on strongly insulated sand covered reef flat, 1.5m	-4.1833	144.8667
<i>Trichosolen</i>	<i>solomonensis</i>	<i>Pseudobryopsis</i>	Coppeljans	2001	PNG	on slightly sand covered coral on reef flat, 1.5m	-5.1333	145.8167
<i>Trichosolen</i>	<i>solomonensis</i>	<i>Pseudobryopsis</i>	Womersley	1970	Solomon	on acropora in waves	-8.4500	158.1333
<i>Trichosolen</i>	<i>solomonensis</i>	<i>Pseudobryopsis</i>	Egerod	1975	Thailand	on intertidal sandy coral	8.0875	98.2900
<i>Trichosolen</i>	<i>solomonensis</i>	<i>Pseudobryopsis</i>	Egerod	1975	Thailand	on silted state on coral in upper intertidal	7.8030	98.4110
<i>Trichosolen</i>	<i>solomonensis</i>	<i>Pseudobryopsis</i>	Egerod	1975	Thailand	in shady cove	7.7265	98.7900
<i>Trichosolen</i>	<i>solomonensis</i>	<i>Pseudobryopsis</i>	Egerod	1975	Thailand	on sandy coral, 2-3m	7.7445	98.7638
<i>Trichosolen</i>	sp.	<i>Trichosolen</i>	De Clerck	1996	Jubail	on dead coral, 2m	27.3147	49.5344
<i>Trichosolen</i>	sp.	<i>Trichosolen</i>	Littlers	1999	Fiji	on coral after cyclone	-19.0700	178.4000
<i>Trichosolen</i>	sp.	<i>Trichosolen</i>	Jupp	2007	Muscat	on coral after cyclone	23.6791	58.5013
Uncertain	sp.	<i>Trichosolen</i>	GENT (unpubl.)	2006	Spain/Catalonia	intertidal, shallow subtidal	41.8419	3.1279
Uncertain	sp.	<i>Trichosolen</i>	GUAM (unpubl.)	1989	Guam	on reef slope, ca. 10m	13.4223	144.7904
Uncertain	sp.	<i>Trichosolen</i>	GUAM (unpubl.)	1999	Oman/Masirah West	on sand covered rockland coral heads, 5.5m	20.215	58.632
Uncertain	sp.	<i>Trichosolen</i>	GENT (unpubl.)	1998	Philippines/Macan East	intertidal	10.301	124.0175
Uncertain	<i>thrikokolensis</i>	<i>Pseudobryopsis</i>	Anil Kumar	1993	India/Kerala	*no locality information*	11.5200	75.6000
<i>Pseudobryopsis</i>	<i>venezoleana</i>	<i>Trichosolen</i>	Ganesan	1990	Venezuela	*no locality information*	10.6100	-66.8930

Excluded: *Trichosolen planktonica*; *Pseudobryopsis hainanensis* (Korea)



# CHAPTER 3

## MACROECOLOGY MEETS MACROEVOLUTION: EVOLUTIONARY NICHE DYNAMICS IN THE SEAWEED *HALIMEDA*

Heroen Verbruggen<sup>1,3</sup>, Lennert Tyberghein<sup>1,3</sup>, Klaas Pauly<sup>1,3</sup>, Caroline Vlaeminck<sup>1</sup>, Katrien Van Nieuwenhuyze<sup>1</sup>, Wiebe H.C.F. Kooistra<sup>2</sup>, Frederik Leliaert<sup>1</sup>, Olivier De Clerck<sup>1</sup>

<sup>1</sup> Phycology Research Group and Center for Molecular Phylogenetics and Evolution, Biology Department, Ghent University  
Krijgslaan 281/S8, 9000 Ghent, Belgium

<sup>2</sup> Stazione Zoologica 'Anton Dohrn', Villa Comunale, 80121 Naples, Italy

<sup>3</sup> these authors contributed equally to this work

Article published as: Verbruggen H, Tyberghein L, Pauly K, Vlaeminck C, Van Nieuwenhuyze K, Kooistra WHCF, Leliaert F & De Clerck O (2009) Macroecology meets macroevolution: evolutionary niche dynamics in the seaweed *Halimeda*. *Global Ecology and Biogeography* **18**, 393-405

## ABSTRACT

**Aim** Because of their broad distribution in geographic and ecological dimensions, seaweeds (marine macroalgae) offer great potential as models for marine biogeographic inquiry and exploration of the interface between macroecology and macroevolution. This study aims to characterize evolutionary niche dynamics in the common green seaweed genus *Halimeda*, use the observed insights to gain understanding of the biogeographic history of the genus, and predict habitats that can be targeted for discovery of species of special biogeographic interest.

**Location** Tropical and subtropical coastal waters.

**Methods** The evolutionary history of the genus is characterized using molecular phylogenetics and relaxed molecular clock analysis. Niche modeling is carried out with Maximum Entropy techniques and uses macroecological data derived from global satellite imagery. Evolutionary niche dynamics are inferred through application of ancestral character state estimation.

**Results** A nearly comprehensive molecular phylogeny of the genus was inferred from a six-locus dataset. Macroecological niche models showed that species' distribution ranges are considerably smaller than their potential ranges. We show strong phylogenetic signal in various macroecological niche features.

**Main conclusions** The evolution of *Halimeda* is characterized by conservatism for tropical, nutrient-depleted habitats, yet one section of the genus managed to invade colder habitats multiple times independently. Niche models indicate that the restricted geographic ranges of *Halimeda* species are not due to habitat unsuitability, strengthening the case for dispersal limitation. Niche models identified hotspots of habitat suitability of Caribbean species in the East Pacific Ocean. We propose that these hotspots be targeted for discovery of new species separated from their Caribbean siblings since the Pliocene rise of the Central American Isthmus.

## INTRODUCTION

Various interacting features influence the distribution of a species. The niche of a species is commonly defined as the set of biotic and abiotic conditions in which it is able to persist and maintain stable population sizes (Hutchinson, 1957). Further distinction is made between a species' fundamental niche, which consists of the set of all conditions that allow for its long-term survival and the realized niche, which is a subset of the fundamental niche that a species actually occupies. Species' tolerances are determined by their morphological, reproductive and physiological traits, which are in turn susceptible to evolutionary forces. Hence, niche characteristics can be interpreted as evolutionary phenomena. Understanding niche evolution yields valuable insights into biogeography, biodiversity patterns and conservation biology (Wiens & Graham, 2005; Rissler *et al.*, 2006; Wiens *et al.*, 2007).

The niche concept provides a conceptual framework to predict geographical distributions of species. Niche models establish the macroecological preferences of a given species based on observed distribution records and a set of macroecological variables, and these preferences can subsequently be used to predict geographic areas with suitable habitat for the species (e.g., Guisan & Thuiller, 2005; Raxworthy *et al.*, 2007; Rissler & Apodaca, 2007). The availability of macroecological data, either in the form of remote sensing or interpolated measurement data, is increasing and has already provided many biological studies with environmental information (Kozak *et al.*, 2008). To date, most ecological niche modeling studies have focused on terrestrial organisms. A notable exception is the study by Graham *et al.* (2007), which used a synthetic oceanographic and ecophysiological model to identify known kelp populations and predict the existence of undiscovered kelp habitats in deep tropical waters.

Integration of niche models, macroecological data and phylogenetic information yields information on niche shifts and insights in the evolution of environmental preferences across phylogenetic trees. So far, evolutionary niche dynamics have been studied almost exclusively in terrestrial organisms (e.g., Graham *et al.*, 2004; Knouft *et al.*, 2006; Yesson & Culham, 2006) and little information is available on niche evolution of the organisms inhabiting

the world's oceans. Seaweeds appear to be an excellent model system to study the evolutionary dynamics of the macroecological niche in coastal marine organisms. Individual seaweed specimens are fixed in one location, yielding a direct link to georeferenced macroecological data. As a whole, seaweeds occur in a wide range of coastal habitats and many seaweed genera or families have a worldwide distribution, resulting in sufficient variability in macroecological dimensions and biogeographic patterns. Evolutionary relationships between and within seaweed genera are being characterized in increasing detail as a result of molecular phylogenetic research during the past few decades. Finally, seaweeds are straightforward to collect and process, making them easy targets for this kind of research.

The marine green algal genus *Halimeda* is among the better studied seaweeds from a phylogenetic perspective and is therefore an obvious candidate for studies of niche evolution and biogeography. *Halimeda* consists of segmented, calcified thalli and abounds on and around coral reefs and in lagoons throughout the tropics and subtropics up to depths in excess of 150 m (Hillis-Colinvaux, 1980). *Halimeda* species are important primary producers and provide food and habitat for small animals and epiflora (Jensen *et al.*, 1985; Naim, 1988). After the algae reproduce, they die and their calcified segments are shed. *Halimeda* segments account for up to 90% of tropical beach sand and carbonate rock of tropical reefs (e.g., Drew, 1983; Freile *et al.*, 1995). The biogeography of *Halimeda* has been described in some detail. All but one species are restricted to a single ocean basin (Indo-Pacific or Atlantic), and biogeography has a strong phylogenetic imprint: each of the five sections of the genus is separated into Atlantic and Indo-Pacific sublineages, suggestive of a strong vicariance event. Even though the species' distribution ranges and the historical biogeographic patterns have been identified, questions about what causes them remain (Kooistra *et al.*, 2002; Verbruggen *et al.*, 2005b). Are species restricted to one ocean basin because of habitat unsuitability in the other basin or should the limited distribution ranges be attributed to dispersal limitation? It is also not known with certainty which vicariance event may be responsible for the phylogenetic separation of Indo-Pacific and Atlantic lineages. So far, two geological events have been implied: the Miocene closure of the Tethys seaway in the Middle East and the Pliocene shoaling of the Central American Isthmus (Kooistra *et al.*, 2002; Verbruggen *et al.*, 2005b).

The first goal of the present study is to investigate evolutionary niche dynamics of the seaweed genus *Halimeda*, focusing on niche dimensions relevant to global geographic distributions rather than local distributional issues such as microhabitat preferences. The second goal is to investigate two aspects of the biogeography of the genus: why are species restricted to a single ocean basin and what caused the historical biogeographic splits. Our approach consists of a combination of molecular phylogenetics, niche modeling, optimization of models of macroecological trait evolution, and ancestral state estimation.

## **MATERIALS AND METHODS**

### **SPECIES IDENTIFICATIONS**

Species delimitation was based on a combination of DNA sequence data and morphological knowledge, with molecular data serving as the primary source of information used to define species boundaries and morphological species boundaries being assessed secondarily, using the species groups determined with DNA data. We used this approach because traditional morphological species definitions are often inaccurate in seaweeds due to morphological plasticity, convergence and cryptic speciation (e.g., Saunders & Lehmkuhl, 2005). The proposed approach has previously been applied to define species boundaries more accurately (Verbruggen *et al.*, 2005a).

The DNA datasets initiated by Kooistra *et al.* (2002) and Verbruggen (2005) were extended for this study using previously described protocols (Verbruggen, 2005), resulting in 264 UCP7 sequences, 337 ITS sequences and 106 *tufA* sequences belonging to a total of 444 specimens. These three datasets were subjected to Neighbor Joining analysis to detect species-level clusters. Using this approach, the sequenced specimens were attributed to 52 *Halimeda* species. If easily recognizable combinations of morphological features could be identified for species by studying the sequenced specimens, these features were used for identification of additional collections from various herbaria (BISH, GENT, L, PC, UPF, US) that were not suitable for sequencing. See the Index Herbariorum web site (<http://sweetgum.nybg.org/ih/>) for definitions of herbarium acronyms.



## PREPROCESSING OBSERVATION DATA

Recent collections had accurate coordinates recorded with a global positioning device. Older collections with detailed locality information were geocoded (latitude and longitude) using Google Earth (<http://earth.google.com>). Points that fell ashore when plotted on coarse resolution environmental grids were manually moved to the adjacent coastal waters using IDRISI Andes (<http://www.clarklabs.org/>). Data were examined for georeferencing errors by checking for geographical outliers with visual and overlay methods in ArcGIS (<http://www.esri.com>). Errors were identified by creating an overlay between the point locality layer and a maritime boundaries layer (exclusive economic zones and coastlines) provided by the Flanders Marine Institute (<http://www.vliz.be>). Any mismatch between these layers were indicative for a potential georeferencing error and outlying points were removed if their origin could not be confirmed.

## SPECIES PHYLOGENY

The evolutionary history underlying the 52 species of *Halimeda* included in the study was inferred from a multi-locus DNA dataset using Bayesian phylogenetic inference (Holder & Lewis, 2003). Bayesian phylogenetic inference techniques make explicit use of models of sequence evolution, an approach that has been shown to outperform methods that do not assume such models (Swofford *et al.*, 2001). Sequence data from four chloroplast loci (*rbcL*, *tufA*, UCP3, UCP7) and two nuclear markers (SSU nrDNA, ITS region) were obtained following previously described protocols (Famà *et al.*, 2002; Kooistra *et al.*, 2002; Provan *et al.*, 2004; Lam & Zechman, 2006) or from previously published studies (Kooistra *et al.*, 2002; Verbruggen *et al.*, 2005a; Verbruggen *et al.*, 2005b). Individual loci were aligned by eye and ambiguous regions were removed. Data for a few loci were missing mainly for recently discovered species but the concatenated data matrix was 90% filled. The final alignment can be obtained from [www.phycoweb.net](http://www.phycoweb.net) and [www.treebase.org](http://www.treebase.org). All new sequences generated in this study have been submitted to Genbank (accession numbers FJ624485-FJ624863).

In order to identify a suitable model of sequence evolution for our dataset, we used model selection procedures based on the second order

Akaike Information Criterion (AICc) (Sullivan, 2005). The phylogenetic analysis was carried out with the model of sequence evolution that yielded the lowest AICc score. This model contained 14 partitions: SSU nrDNA, the ITS region and 3 codon positions per protein-coding gene. The GTR+ $\Gamma_8$  substitution models yielded the best fit to the data for all partitions. Bayesian phylogenetic inference was carried out with MrBayes v.3.1.2 (Ronquist & Huelsenbeck, 2003). Five runs of four incrementally heated chains were run for 10 million generations using default priors and chain temperature settings. Convergence of the MCMC runs was assessed with Tracer v.1.4 (Rambaut & Drummond, 2007). An appropriate burn-in was determined with the automated method proposed by Beiko *et al.* (2006) and a majority rule consensus tree was built from the post-burn-in trees. The tree was rooted at the point where root-to-tip path length variance was minimal.

The branch lengths of the obtained consensus phylogram are proportional to the estimated amount of molecular evolution occurring on the branches. In order to model character evolution, in our case evolutionary niche dynamics, branch lengths should be proportional to evolutionary time rather than amounts of molecular evolution. To obtain a chronogram (i.e. a phylogram with branch lengths proportional to evolutionary time), penalized likelihood rate smoothing (Sanderson, 2002) was carried out on the consensus tree with r8s (Sanderson, 2003), using both the additive and the log-additive penalty settings. The root of the phylogeny was assigned an age of 147 my, following the molecular clock result from Verbruggen *et al.* (2009). We refer to the latter paper for details regarding the dating of the phylogeny.

## MACROECOLOGICAL DATA

Macroecological variables were selected to represent the major environmental dimensions assumed to influence seaweed distribution at a global scale and subject to data availability (Lüning, 1990) (Table 1).

**Table 1.** Geophysical parameters included in the macroecological dataset.

Macroecological parameter	Units	Original spatial resolution	Date	Source	Derived parameters
Sea Surface Temperature (SST)	°C	2.5 arcmin	2003-2007	Aqua-MODIS (NASA)	max, min, average (day & night)
Diffuse Attenuation (DA)	m <sup>-1</sup>	2.5 arcmin	2003-2007	Aqua-MODIS (NASA)	max, min, average
Calcite concentration (Ca)	moles/m <sup>3</sup>	2.5 arcmin	2006	Aqua-MODIS (NASA)	average
Chlorophyll A (chlA)	mg/m <sup>3</sup>	5 arcmin	1998-2007	SeaWiFS (NASA)	max, min, average
Photosynthetically Active Radiation	<u>Einstein/m<sup>2</sup></u> day	5 arcmin	1998-2007	SeaWiFS (NASA)	max, min, average

The base macroecological data included geophysical, biotic and climate variables derived from level-3 reprocessed satellite data (Aqua-MODIS and SeaWiFS) available at OceanColor Web (<http://oceancolor.gsfc.nasa.gov>).

We downloaded grids representing monthly averages at a 5 arcmin ( $\approx 9.2$  km) spatial resolution. These geometrically corrected images are two-dimensional arrays with an Equidistant Cylindrical (Platte Carre) projection of the globe. Yearly minimum, maximum and average values were calculated from the monthly averages with MATLAB (<http://www.mathworks.com/>). To achieve this, average monthly images were generated by averaging images of the same month across years (e.g. average SST of July from 2003–2007). Subsequently, yearly minimum and maximum images were composed by selecting the minimum and maximum pixels from these monthly averages. Finally, yearly average images were created by taking the mean value for every grid cell of the monthly averages. All images were cropped to the latitudinal range 50°N–40°S, which includes the highest latitudes at which *Halimeda* can be found.

## EVOLUTIONARY ANALYSIS OF NICHE CHARACTERISTICS

The evolutionary dynamics of niche features were studied by inferring their patterns of change along the chronogram in a maximum likelihood (ML) framework. The macroecological niche features included in our study are continuous variables and we inferred their evolution with common models of continuous trait evolution. Brownian motion models, also known as constant-variance random walk models, assume that traits vary naturally

along a continuous scale and that variation is accumulated proportionally to evolutionary time, as measured by the branch lengths in a chronogram (Martins & Hansen, 1997; Pagel, 1999). Two branch length scaling parameters (lambda and kappa) were used to extend this model and better describe the mode and tempo of trait evolution (Pagel, 1999). Lambda ( $\lambda$ ) transformations measure the amount of phylogenetic signal present in a continuous character. The transformation consists of multiplying all internal branch lengths of the tree by lambda, leaving tip branches their original length. When the ML estimate of  $\lambda$  is close to one, the internal branches retain their original length, indicating strong phylogenetic signal in the trait. If  $\lambda$  approaches zero, the evolution of the trait is virtually independent of phylogeny. Kappa ( $\kappa$ ) transformations measure the degree of punctuational versus gradual evolution of characters on a phylogeny, by raising all branch lengths to the power kappa. If the ML estimate of  $\kappa$  is close to zero, all branch lengths approach unity, and path lengths become proportional to the number of lineage splitting events, suggesting that the evolution of the trait approximates punctuated evolution associated with speciation events. If the ML estimate of  $\kappa$  is close to one, branch lengths remain unchanged, indicating that the amount of change in the character is proportional to evolutionary time. In other words,  $\kappa$  values close to one indicate gradual evolution.

In order to fit the models above and infer changes of the macroecological niche along the species phylogeny, a species  $\times$  variables matrix had to be constructed. To achieve this, the values of the macroecological data layers were extracted for each sample locality. For each species, the minimum, maximum and average of each macroecological parameter were stored in the species  $\times$  variables matrix. To reduce the influence of geographical sampling bias on the average values, they were calculated by weighted averaging. The Euclidean distance from the sample location to the centre of gravity for the species in question was used as the sample weight. The centre of gravity for the species was determined by averaging the three-dimensional Cartesian coordinates of all sample locations for that species.

The models of continuous trait evolution listed above were optimized along the phylogenetic tree for the minimum, average and maximum values

of a selection of niche variables using the ML optimization of the GEIGER package (Harmon *et al.*, 2008). Ancestral character values for macroecological niche features were estimated by ML inference (Schluter *et al.*, 1997) with the APE package (Paradis *et al.*, 2004). Resulting ancestral state values were plotted on the phylogeny with TreeGradients v1.03 (Verbruggen, 2008).

#### NICHE MODELING PROCEDURE

The macroecological niches of species were modeled with Maxent, a presence-only niche modeling technique based on the maximum entropy principle (Phillips *et al.*, 2006). We used a presence-only technique because only specimen collection data are available and absence data cannot be reliably obtained for seaweed species on a global scale. Maxent has shown remarkably good performance in a comparative study of presence-only niche modeling techniques (Elith *et al.*, 2006). It estimates the probability distribution of maximum entropy (i.e. that is most spread out, or closest to uniform) of each macroecological variable across the study area. This distribution is calculated with the constraint that the expected value of each macroecological variable under the estimated distribution matches the empirical average generated from macroecological values associated with species occurrence data. The model output consists of a spatially explicit probability surface that represents an ecological niche (habitat suitability) translated from macroecological space into geographical space. The output grid is in the logistic format, where each pixel value represents the estimated probability that the species can be present at that pixel (Phillips & Dudik, 2008).

To avoid using redundant and correlated macroecological layers for niche modeling, an unstandardized principal components analysis was performed on the original variables in IDRISI Andes. The first, second and third principle component grids, which together accounted for 98.82% of the overall variance in the original variables, were exported for subsequent use in Maxent.

Global species' niches were modeled for all *Halimeda* species for which more than ten distribution records were available, while excluding species with distribution records suffering from high spatial autocorrelation.

Additionally, a single regional model was generated using pooled occurrence data of six Caribbean species (*H. goreauii*, *H. simulans*, *H. incrassata*, *H. monile*, *H. discoidea.atl* and *H. tuna.car*).

The Maxent algorithm was run with default parameters (convergence threshold =  $10^{-5}$ , maximum iterations = 500, regularization multiplier = 1, maximum number of background points = 10000, and use of linear, quadratic, product and hinge features). Models were created using 80% of the localities for model training and 20% for model testing.

Statistical evaluation of the models was based on threshold-independent receiver operating characteristic (ROC) analysis (Phillips *et al.*, 2006). For presence-only modeling, the ROC curve is a plot of sensitivity (proportion of correctly predicted presences) against the fractional area predicted present. The area under the ROC curve (AUC) is subsequently compared to the area under the null expectations line connecting the origin and (1,1), thus providing a measure of predictive model performance. An AUC approximating 1 would mean optimal discrimination of suitable versus unsuitable sites, whereas an AUC between 0 and 0.5 is indicative of predictions no better than random. Additionally, we use a modified AUC based on partial ROC curves as proposed by Peterson *et al.* (2008). This approach accounts for a user-defined maximum acceptable omission error, which we set at 0.1, and takes only the range of acceptable models in terms of omission error into account. The partial AUC is then rationed to the partial area under the null expectations line. Hence, the AUC ratio equals one for models performing no better than random, and increases with improving model accuracy. All partial AUC calculations were performed in the R statistical computing environment (R Development Core Team, 2008).

## RESULTS

### SPECIES DELIMITATION AND PHYLOGENY

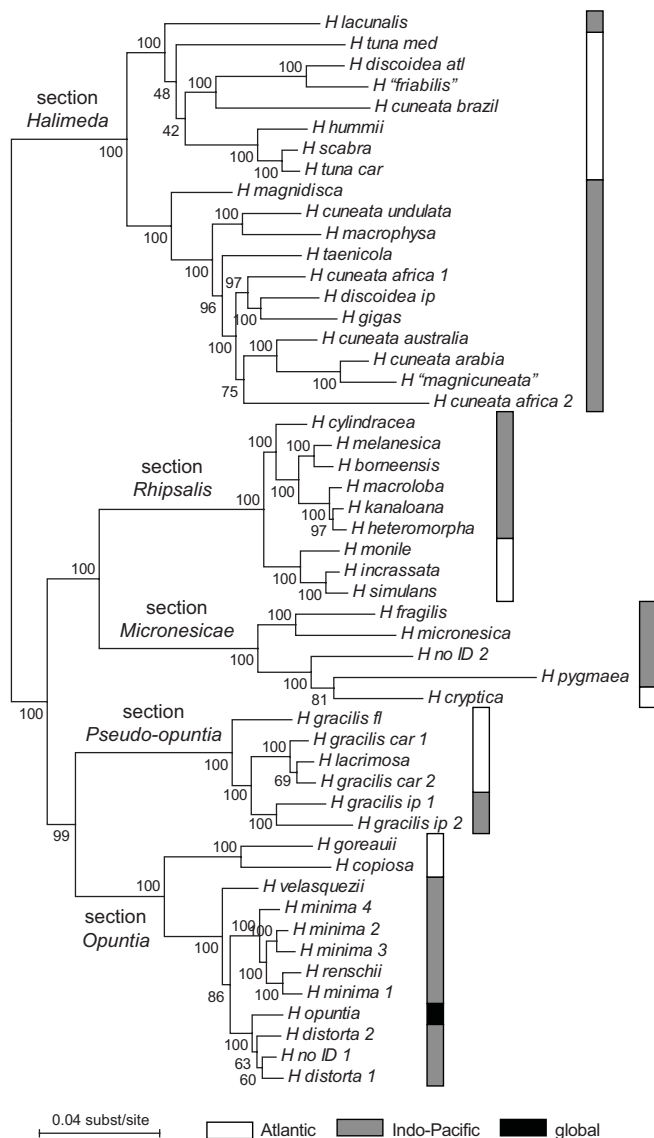
Neighbor Joining analysis of the UCP7, ITS and *tufA* sequence alignments pointed out 52 clusters with low sequence divergence within clusters and relatively high divergence between clusters, as is typically found at the species boundary (Hebert *et al.*, 2004; Verbruggen *et al.*, 2005a). Not all

clusters corresponded to described, named species. The undescribed clusters represent cryptic or pseudo-cryptic species (Kooistra *et al.*, 2002; Verbruggen *et al.*, 2005a; Verbruggen *et al.*, 2005b). The clusters inferred from DNA data formed the basis of the species definitions used in the remainder of the paper. After addition of morphologically identified herbarium specimens, the database consisted of 1080 samples from 538 unique localities. Analysis of the concatenated alignment of *rbcL*, *tufA*, UCP3, UCP7, 18S and ITS sequences (4965 characters) yielded a well-resolved species phylogeny in which five lineages, corresponding to the genus' five sections, could be recognized (Fig. 1).

### EVOLUTION OF NICHE CHARACTERISTICS

A few niche features contained considerable amounts of phylogenetic signal, as indicated by the high ML estimates of lambda values using Pagel's lambda branch length modifier (Table 2). A general observation was that average trait values contained more phylogenetic signal than minimum and maximum trait values (e.g., average temperature, not minimum or maximum temperature). High kappa values for the average trait values indicate that change of these traits is proportional to evolutionary time; in other words, change is gradual (Table 2). Some traits that also contained phylogenetic signal were not included in the table because of significant correlation with the listed variables. This is the case for photosynthetically active radiation, which is correlated with sea surface temperatures, and diffuse attenuation, which is correlated with chlorophyll values (caused by phytoplankton).

Figure 2 illustrates the estimated evolutionary patterns of average annual temperature and chlorophyll values. Estimated ancestral trait values are shown at the internal nodes and visualized using a color gradient. An average annual temperature of 27.4°C (95% confidence interval: 25.6–29.2) is inferred at the basal split (Figure 2A), indicating a tropical origin for the genus. The tree clearly shows that evolution along the SST niche dimension is not homogeneous throughout the tree. Whereas the sections *Rhipsalis*, *Micronesicae*, *Pseudo-opuntia* and *Opuntia* barely deviate from typical tropical temperatures, evolution along the temperature axis has been common in section *Halimeda*.



**Figure 1.** Phylogenetic tree of 52 *Halimeda* species inferred from six molecular loci using Bayesian techniques, rooted at the point where root-to-tip path length variance is minimal. Numbers at nodes indicate statistical support (Bayesian posterior probabilities, in percentages).



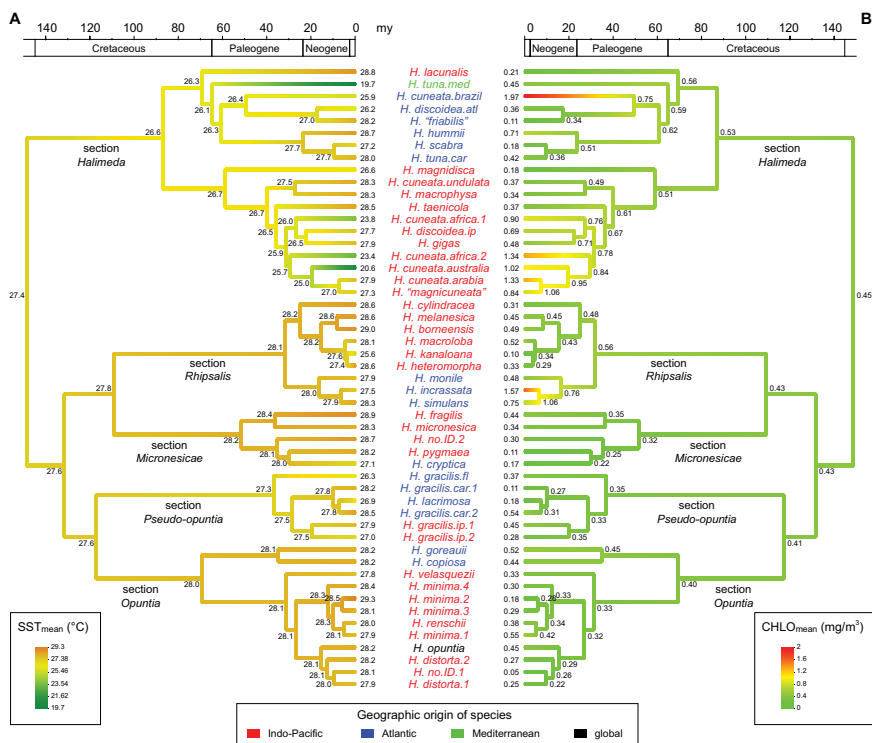
**Table 2.** Optimum values for the branch length scaling parameters  $\lambda$  and  $\kappa$  used to test the mode and tempo of evolution of niche features. The niche traits are sea surface temperature (SST) and chlorophyll A values. The high optimal  $\lambda$  values inferred for average trait values indicate strong phylogenetic signal in these traits whereas the low  $\lambda$  values obtained for the minimum and maximum traits suggest a lack of phylogenetic signal. The relatively high optimum values for  $\kappa$  for average SST values suggest that evolution of this niche feature was more or less gradual (proportional to time). The lower value for average chlorophyll A suggests that there is a non-negligible punctuated component to the evolution of nutrient preferences. The first two columns used the tree smoothed with the additive penalty; the last two columns used the tree smoothed with the log-additive penalty. The  $\kappa$  parameter was not optimized when there was poor phylogenetic signal in the data (low  $\lambda$ ).

Trait	Optimal $\lambda$	Optimal $\kappa$	Optimal $\lambda$	Optimal $\kappa$
max SST	0.07262		0.05448	
average SST	0.90087	0.83616	0.15159	0.66764
min SST	0.09371		0.05308	
min chlA	0.10672		0.06804	
average chlA	0.78528	0.47084	0.80792	0.37894
max chlA	0.01364		0.00000	

More specifically, the lineages leading to *H. tuna.med*, *H. cuneata.africa.1*, *H. cuneata.africa.2* and *H. cuneata.australia* have evolved a preference for colder water. Chlorophyll values were mapped onto the phylogeny as a proxy for nutrient preferences (figure 2B). Deviations from the average (low) nutrient preference values are present in *Halimeda* section *Halimeda* (*H. cuneata.brazil*, *H. cuneata.africa.1*, *H. cuneata.africa.2*, *H. cuneata.australia*, *H. cuneata.arabia* & *H. magnicuneata*) and in *Halimeda* section *Rhipsalis* (*H. incrassata*).

#### NICHE MODELS AT THE GLOBAL SCALE

Niche models indicating the areas where macroecological conditions are suitable for species to occur were generated for all species (figure 3 and figure S1). The average AUC across all models with 20% test localities was 0.917 (SD = 0.046) for the training data and 0.906 (SD = 0.054) for the test data. The corresponding average AUC ratios were 1.576 (SD = 0.209) for the training data and 1.615 (SD = 0.234) for the test data.

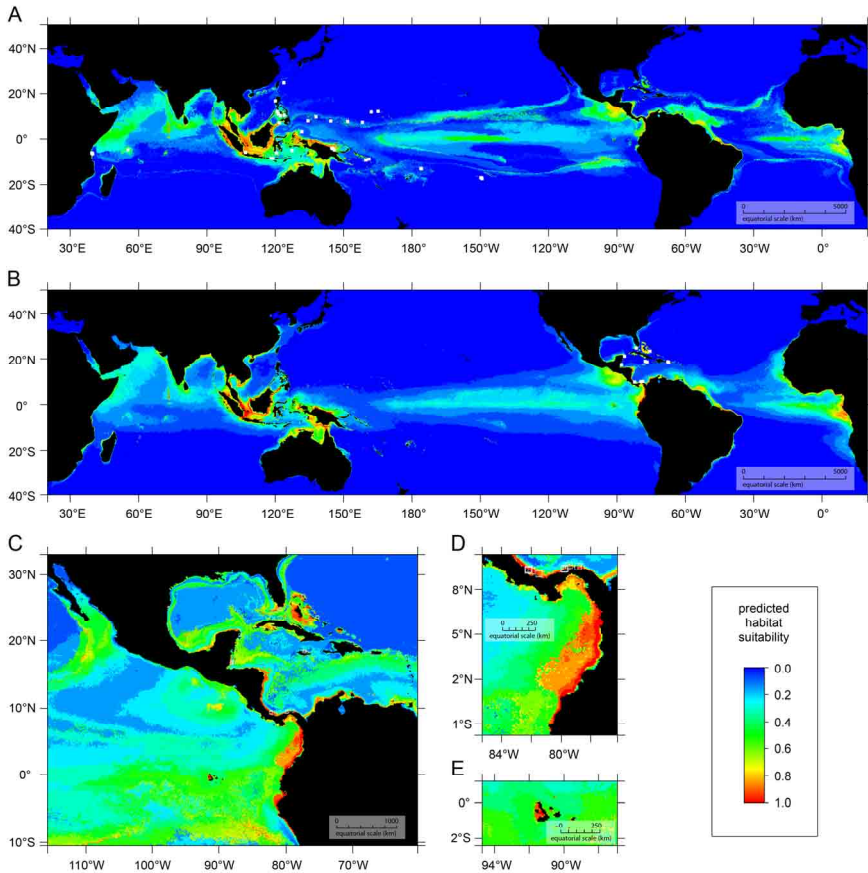


**Figure 2.** Inferred evolutionary history of niche features in *Halimeda*. Ancestral values for (A) mean sea surface temperature and (B) mean chlorophyll concentration are plotted along the phylogeny. Numbers plotted at nodes indicate the inferred ancestral values. These values were obtained using a maximum likelihood approach as described in the text. Values are also drawn along a color gradient to allow rapid visual assessment of evolutionary patterns. Green indicates low values, red stands for high values, and yellowish colors indicate intermediate values. The geographic origin of species is indicated with colored taxon names.

The high AUC values and ratios indicate that the most essential macroecological variables determining species distributions were accounted for in the dataset. The high scores for the test data indicate adequate model performance rather than overfitting of the model on the training data. The predicted distributions are clearly broader than the species known distributions. For example, the distribution model of the exclusively Caribbean species *H. simulans* (figure 3B) predicts habitat suitability in parts of the Indo-Pacific basin. Similarly, the model of the Indo-Pacific species *H. borneensis* (figure 3A) predicts habitat suitability in parts of the Atlantic Ocean. In general, there was a stronger tendency of predicting Atlantic species into the Indo-Pacific than vice versa.

## NICHE MODEL AT THE REGIONAL SCALE

The model predicting suitable habitat for a suite of six Caribbean species is shown in figure 3C (AUC ratio = 1.783). Potentially suitable habitats of these Caribbean species in the East Pacific are mainly predicted along the southern coast of Panama, the western coast of Colombia and in the Galapagos Islands (figure 3C–E).



**Figure 3.** Predictive ecological niche models of *Halimeda* species inferred from environmental data and species occurrence records. (A) Niche model of the exclusively Indo-Pacific species *H. borneensis* indicating habitat suitability in some Atlantic regions. (B) Niche model of the exclusively Caribbean species *H. simulans* predicting habitat suitability in several Indo-Pacific regions. (C) Pooled niche model of six Caribbean *Halimeda* species predicting habitat suitability along parts of the Pacific coast of Central America. (D) Detailed view of the areas along the Pacific coastlines of Panama, Colombia and Ecuador predicted by the model from panel C. (E) Detailed view of the Galapagos archipelago as predicted by the model from panel C. Predicted habitat suitability is indicated with colors along a gradient, warmer colors indicating areas with better predicted conditions. White squares indicate specimen localities used for model training. All maps are equidistant cylindrical projections.

## DISCUSSION

The obtained results invite discussion about several issues related to the macroecological niche of seaweeds, how it evolves and how it relates to patterns of biogeography.

### MODELLING SEAWEED DISTRIBUTIONS

#### *Niche modeling vs. previous approaches*

Our niche models indicate areas where the macroecological conditions are likely to be suitable for various *Halimeda* species to establish populations. They reflect the marked tropical nature of most species and show that many species occupy only part of the potentially suitable habitat (see below). Previous knowledge about the macroecological niche of seaweeds mainly stemmed from comparing distribution ranges with isotherms (isotherm fitting), studying survival and growth under various culture conditions, or the combination of both (e.g., van den Hoek, 1982). These approaches and the niche modeling approach presented here differ from each other in a number of aspects. Whereas the fundamental niche is investigated with in vitro studies of survival and growth, the realized niche is central in modeling techniques and isotherm fitting. A fundamental difference between niche modeling and both other approaches is that the former yields probabilistic output whereas the latter usually propose hard thresholds. The ease with which a niche modeling study can be carried out has benefits as well as drawbacks. The advantage is obvious when targeting species that are difficult to grow in culture. A disadvantage of niche modeling is that the choice of a specific niche modeling algorithm and the parameter settings may influence niche predictions and predictive model performance (Elith *et al.*, 2006; Peterson *et al.*, 2008). The maximum entropy method with ROC modifications appeared to be the most suitable option for our goals.

#### *Taxonomic caveat*

An additional concern about the application of niche models in seaweed research is the ease with which heterogeneous distribution records can be used to generate models. As mentioned earlier, morphological species delimitation is troublesome in algae and, as a consequence, published

species occurrence records based on morphological identifications are not always meaningful. We have taken great caution to avoid identification errors through DNA-guided species delimitation. All methods share the drawback of being sensitive to specimen sampling. In this respect, the absolute number of samples is likely to be of inferior importance compared to the spread of samples across relevant macroecological dimensions (Pearson, 2007).

## MACROEVOLUTION OF THE MACROECOLOGICAL NICHE

### *Historical perspective*

Evolutionary processes are influenced by environmental variation in space and time (Kozak *et al.*, 2008). Many studies taking a niche modeling approach to study environmental variation in a phylogenetic framework have shown strong heritability of macroecological preferences (e.g., Martinez-Meyer & Peterson, 2006; Yesson & Culham, 2006). To our knowledge, these studies have all focused on terrestrial organisms. The evolutionary dynamics of the niche of seaweeds have hardly been studied in the past. Breeman *et al.* (2002) investigated the evolution of temperature responses in the seaweed genus *Cladophora*. Their approach consisted of measuring cold tolerance, heat tolerance, and growth of various culture strains at different temperature regimes. The response variables (tissue damage and growth rates) were interpreted along a phylogenetic tree, leading to the conclusion that the two main lineages of the *Cladophora vagabunda* complex had divergent cold tolerances. Although the experimental data of this study differs from ours as discussed above, the approach taken to infer niche dynamics in both studies is not fundamentally different. However, thanks to the advances in models describing the evolution of continuous characters that have taken place since the publication of Breeman *et al.* (2002) and their implementation in user-friendly packages for the R statistical computing environment (Paradis *et al.*, 2004; Harmon *et al.*, 2008), much more detailed inferences can now be made.

### *Niche conservatism*

Our study shows that the macroecological niche in the seaweed genus *Halimeda* has a strong phylogenetic imprint and that niches appear to change

gradually with time. The results clearly indicate the phylogenetic heritability of macroecological preferences: four out of five sections (*Rhipsalis*, *Micronesica*, *Opuntia* and *Pseudo-opuntia*) demonstrate conserved preference for high temperatures and low nutrient levels, confirming the association of these sections with tropical coral reefs and shallow lagoons (figure 2). Adaptation to colder and more nutrient-rich water only occurred in section *Halimeda*. Remarkably, the transition into colder water seems to have taken place four times independently (in *H. tuna.med*, *H. cuneata.africa.1*, *H. cuneata.africa.2* and *H. cuneata.australia*). The species *H. tuna.med* is the only one inhabiting the Mediterranean Sea and can maintain populations at sites with yearly sea surface temperature minima around 10°C. The species *H. cuneata.africa.1* and *H. cuneata.africa.2* occur in SE Africa. *H. cuneata.australia* is found along the shores of SW Australia. Chlorophyll values, used as a proxy for the trophic status of the surface water (Duan *et al.*, 2007), are above average for certain species in section *Halimeda*, often in the subtropical species. It is known that nutrient levels increase with latitude in the latitudinal range studied here (Sasai *et al.*, 2007). *Halimeda cuneata.brazil* occurred in waters with high average chlorophyll values due to an overall high concentration along the Brazilian coast. The high average chlorophyll value of waters in which *H. incrassata* was recorded is largely due to an outlier observation in Florida.

### ***Sources of uncertainty***

Our study of evolutionary niche dynamics involves several subsequent analyses, hence a discussion of the potential sources of uncertainty affecting the final result is in place. The first source of uncertainty is in the species phylogeny. A lack of support of phylogenetic relationships will have direct repercussions on the accuracy of downstream analyses. In our study, the use of a multi-locus alignment yielded very high statistical support for the great majority of branches in the tree. Therefore we have used the tree resulting from the Bayesian analysis (figure 1) in subsequent analyses as if it were known without uncertainty. Second, inferences of trait evolution also depend on branch lengths, which are affected by two potential sources of uncertainty: branch length estimation error in the phylogenetic analysis and error from the rate smoothing process that transforms the phylogram into a

chronogram. Especially rate smoothing can lead to variation in branch lengths if different settings are used. We followed the recommendations in the manual of the r8s program. Third, the values used as character states of the terminal taxon influence the results. We used distance-weighted averages as fixed character states for the terminal taxa whereas in reality there is variation around the average. Taking this variation into account is expected to broaden confidence intervals on inferred ancestral states (Martins & Hansen, 1997). A fourth source of error could result from the inability of Brownian motion models to capture the complexity of historical forces affecting niche evolution, a source of error inherent to using simple models to describe a more complex reality. The last element of uncertainty lies in the ancestral character estimation, which infers values for ancestral taxa based on values of recent taxa. These analyses, however, report the 95% confidence intervals around the inferred value. If a character evolves fast, this will be reflected in broader confidence intervals on ancestral character states (Martins, 1999). We have not attempted to quantify the accumulation of uncertainty throughout our sequence of analyses due to practical limitations but the reader should be aware of the assumptions that were made.

### ***Paleobiological perspective***

Despite the relatively high levels of uncertainty usually associated with ancestral state estimation of continuous characters (Schluter *et al.*, 1997), the observed conservatism for environmental preferences yields a relatively narrow 95% confidence interval for the average sea surface temperature characterizing the habitat of the most recent common ancestor of extant *Halimeda* species (25.6–29.2°C). The ML estimate of 27.4°C appears to be in agreement with the tropical Tethyan origin of *Halimeda* that was previously derived from the fossil record. The earliest known fossil that is considered to belong to the genus is *Halimeda soltanensis* from the Upper Permian ( $\pm 250$ –270 my) of Djebel Tebaga in South Tunisia (Poncet, 1989), which was at that time located at a low latitude along the western shore of the Tethys Ocean (Smith *et al.*, 1994). A more diverse assemblage of species with a markedly tropical distribution had evolved by the Upper Cretaceous ( $\pm 100$ –65 my) (Dragastan & Herbig, 2007). The invasion of *Halimeda* into higher latitudes has not been documented in the fossil record. Our

chronogram suggests that the invasion occurred during late Paleogene and Neogene times, a period characterized by global cooling (Zachos *et al.*, 2001). This finding confirms earlier hypotheses that at least parts of the warm-temperate seaweed floras originated from tropical ancestry during this period of globally decreasing temperatures (van den Hoek, 1984; Lüning, 1990).

## GLOBAL BIOGEOGRAPHY

### ***Dispersal limitation***

*Halimeda* species have previously been shown to be geographically restricted to either the Atlantic Ocean or the Indo-Pacific basin (Kooistra *et al.*, 2002; Verbruggen *et al.*, 2005a; Verbruggen *et al.*, 2005b). One could ask whether the absence of Atlantic species in the Indo-Pacific (and vice versa) is a consequence of dispersal limitation or if habitat differences may be responsible for the limited distributions. The niche model of the Indo-Pacific species *H. borneensis* clearly indicates that some parts of the Caribbean Sea would be suitable habitat (figure 3A) and the niche model of the Atlantic species *H. simulans* suggests that it could survive in large parts of the Indo-Pacific tropics (figure 3B). Similar patterns were observed for other species (figure S1). So, unless *Halimeda* species are limited by habitat differences between the Atlantic and Indo-Pacific basins that are not represented in our macroecological data, it can be concluded that dispersal limitation is the most likely explanation for the strong separation of Atlantic and Indo-Pacific species. Dispersal limitation of benthic tropical marine organisms between oceans is not uncommon (Lessios *et al.*, 2001; Teske *et al.*, 2007) and can be explained by the North–South orientation of the African and American continents, prohibiting marine dispersal between the Atlantic and Indo-Pacific basins through tropical waters. *Halimeda opuntia* is the only species that occurs in both ocean basins. It is part of a clade of Indo-Pacific species, indicating that it originated in the Indo-Pacific basin and subsequently dispersed to the Atlantic Ocean and spread throughout its tropical regions. It was previously suggested that *H. opuntia* was introduced in the Atlantic Ocean by early interoceanic shipping (Kooistra & Verbruggen, 2005). If this scenario is correct, our models' prediction that



parts of the tropical Atlantic Ocean form suitable habitat for Indo-Pacific species and the conclusion of dispersal limitation between ocean basins would be confirmed.

### ***Vicariance patterns***

Geographic distribution patterns show a clear phylogenetic signal: each section separates largely into an Atlantic and an Indo-Pacific lineage (figure. 1), confirming previous observations (Kooistra *et al.*, 2002; Verbruggen *et al.*, 2005b). This pattern indicates ancient lineage splitting through vicariance and subsequent diversification within the Atlantic and Indo-Pacific basins. A number of geological events are commonly invoked to explain sister relationships between strictly Atlantic and strictly Indo-Pacific lineages. The first is the spreading of the Atlantic Ocean, which started during the Jurassic ( $\pm 170$ – $160$  my) (Smith *et al.*, 1994). The second is the collision of the African and Eurasian plates in the Middle East during the Miocene ( $\pm 15$ – $12$  my) (Rögl & Steininger, 1984). The third event is the closure of the Central American Seaway in the Pliocene ( $\pm 3$  my) (Coates & Obando, 1996). Different events have been hypothesized to be at the basis of the geographic splits in *Halimeda* but results have remained inconclusive (Kooistra *et al.*, 2002; Verbruggen *et al.*, 2005b). Our chronogram suggests that the splits between Atlantic and Indo-Pacific lineages originated at various times during the Paleogene (65–25 my). In other words, the time frame of initial divergence does not correspond closely with either one of the geological events. During the Paleogene, however, an important oceanographic event that may have limited dispersal between the Atlantic and Indo-Pacific ocean basins took place: the circum-equatorial current that homogenized the tropical marine biome during the Cretaceous was deflected to south of Africa (Lawver & Gahagan, 2003). This result suggests that geological barriers may not be the initial cause of divergence between populations but instead act as barrier reinforcements after divergence has been initiated by oceanographic events. A similar conclusion was reached in molecular and paleontological studies of species across the Central American Isthmus (e.g., Collins *et al.*, 1996; Knowlton & Weigt, 1998). The generality of this pattern requires additional study. At least for some organisms, divergence times between Atlantic and Indo-Pacific lineages obtained with a molecular clock match more closely with the timing of the

collision of the African and Eurasian plates in the Middle East (e.g., Teske *et al.*, 2007).

#### REGIONAL BIOGEOGRAPHY OF TROPICAL AMERICA

As an alternative to the molecular clock, one would also be able to infer which geological events were involved in species partitioning between Atlantic and Indo-Pacific through a thorough study of eastern Pacific *Halimeda* species. The Caribbean and East Pacific formed a single tropical marine biota that was separated by the shoaling of the Central American Isthmus during the Pliocene, resulting in the formation of many trans-isthmian sister species (Knowlton & Weigt, 1998). The emergence of a land bridge has been dated at approximately three million years (Coates & Obando, 1996). The presence of trans-isthmian species pairs with a distribution limited to the tropical Americas (i.e., not in the wider Indo-Pacific) can be taken as evidence for vicariance across the Central American Isthmus.

Only *Halimeda discoidea* has been reported from the East Pacific and, curiously, molecular analyses have shown these populations not to be related to the Caribbean species *H. discoidea.atl* as one may expect but to the Indo-Pacific species *H. discoidea.ip* (Verbruggen *et al.*, 2005b). So either *Halimeda* does not have trans-isthmian species pairs in the tropical Americas or they have not been discovered yet. The seaweed flora of the tropical East Pacific Ocean has not been studied in great detail in the past and recent inventories have shown lots of new discoveries (Wysor, 2004). We aimed to facilitate the discovery of trans-isthmian sister pairs by identifying geographic regions in the East Pacific Ocean that are hotspots of habitat suitability for Caribbean species. The niche model of pooled distribution data of six Caribbean species predicted parts of the East Pacific Ocean as suitable habitat (figure 3C) and identified three hotspots of habitat suitability: the western Galapagos Islands (figure 3E), the West coast of Colombia and parts of the South coast of Panama (figure 3D). We suggest that these areas should be targeted in future research expeditions aiming to discover trans-isthmian species pairs. The utility of ecological niche models to guide discovery has already been documented. Unexplored deep-water

kelp forests were recently found in the Galapagos archipelago based on predictions of a synthetic oceanographic and ecophysiological model (Graham *et al.*, 2007). Similarly, expeditions directed by niche models of chameleons led to the discovery of additional populations of known species and several species new to science (Raxworthy *et al.*, 2003). It should be noted that the niche model presented here predicts habitat suitability only as a function of the macroecological variables included in the dataset. It is beyond doubt that factors not included in our dataset (e.g., micro-habitat characteristics, tidal amplitudes, grazing pressure and other biotic interactions) affect the actual distribution of species. If such data were available, it could be used to create a more specific model and would likely result in smaller hotspots, allowing even more targeted expeditions.

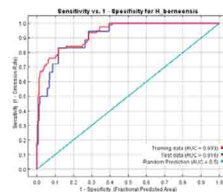
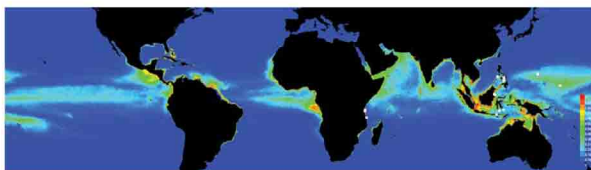
## ACKNOWLEDGEMENTS

We thank W. Willems for providing the R script to calculate partial AUC values and for discussion of techniques. We are grateful to M. Accioly, K. Arano, M. Bandeira-Pedrosa, C. Battelli, B. Brooks, K. Clifton, M. Coffroth, P. Colinvaux, R. Collin, E. Coppejans, O. Dargent, Y. de Jong, G. De Smedt, E. Demeulenare, R. Diaz, E. Drew, S. Fredericq, C. Galanza, S. Guimaraes, F. Gurgel, O. Gussmann, R. Haroun, I. Hendriks, J. Hernandez, L. Hillis, J. Huisman, M. Kaufmann, L. Kirkendale, L. Liao, D. Littler, M. Littler, G. Llewellyn, P. Marshall, J. Maté, A. Maypo, A. N'Yeurt, D. Olandesca, C. Ortuno, K. Page, F. Parrish, C. Payri, G. Procaccini, W. Prud'homme van Reine, L. Raymundo, T. Schils, E. Tronchin, M. Van Veghel, P. Vroom, S. Williams, S. Wilson, B. Wysor and J. Zuccarello for providing specimens. Funding was provided by the Research Foundation – Flanders (research grant G.0142.05 and post-doctoral fellowships to HV and FL) and IWT (doctoral fellowship to LT). We thank two anonymous referees for their constructive comments on a previous version of the manuscript.

## SUPPLEMENTARY MATERIAL

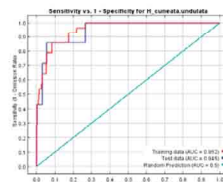
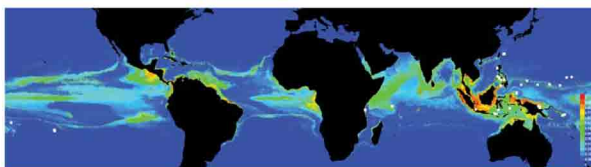
Figure S1. Model output for *Halimeda* species

### *Halimeda borneensis*



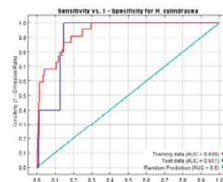
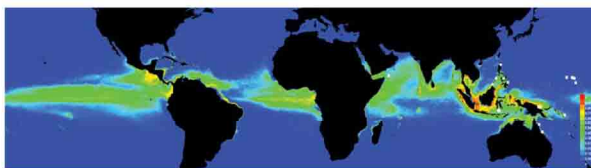
# Training samples = 76 # Test samples = 18 AUC (training) = 0.933 AUC (test) = 0.916 AUC ratio (training) = 1.540 AUC ratio (test) = 1.537

### *Halimeda cuneata.undulata*



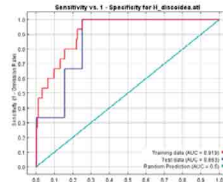
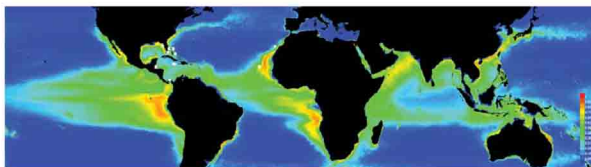
# Training samples = 28 # Test samples = 7 AUC (training) = 0.952 AUC (test) = 0.945 AUC ratio (training) = 1.691 AUC ratio (test) = 1.576

### *Halimeda cylindracea*



# Training samples = 22 # Test samples = 5 AUC (training) = 0.936 AUC (test) = 0.917 AUC ratio (training) = 1.662 AUC ratio (test) = 1.737

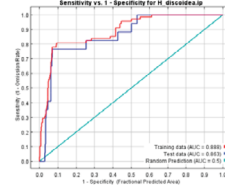
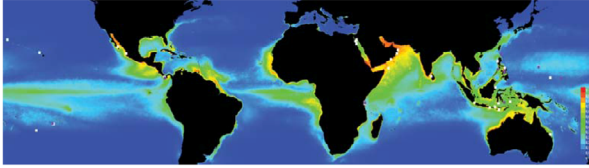
### *Halimeda discoidea.atl*



# Training samples = 12 # Test samples = 3 AUC (training) = 0.919 AUC (test) = 0.863 AUC ratio (training) = 1.564 AUC ratio (test) = 1.880

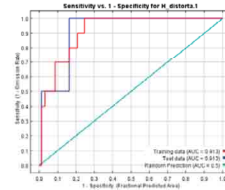
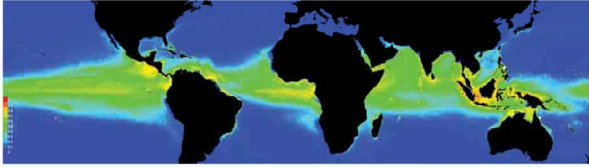
# Chapter 3

## *Halimeda discoidea.ip*



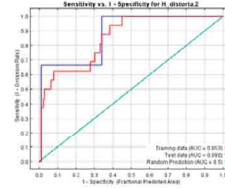
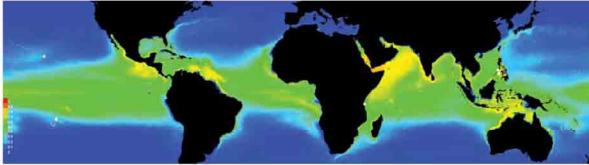
# Training samples = 68    # Test samples = 17    AUC (training) = 0.888    AUC (test) = 0.863    AUC ratio (training) = 1.392    AUC ratio (test) = 1.312

## *Halimeda distorta.1*



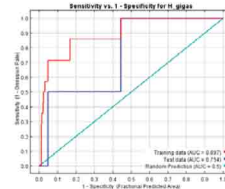
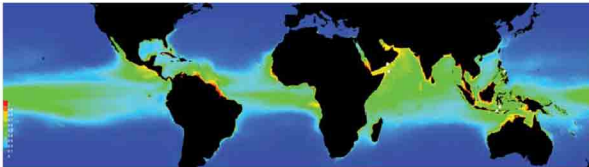
# Training samples = 10    # Test samples = 2    AUC (training) = 0.913    AUC (test) = 0.913    AUC ratio (training) = 1.652    AUC ratio (test) = 1.541

## *Halimeda distorta.2*



# Training samples = 16    # Test samples = 3    AUC (training) = 0.853    AUC (test) = 0.880    AUC ratio (training) = 1.423    AUC ratio (test) = 1.854

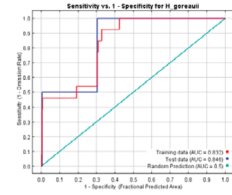
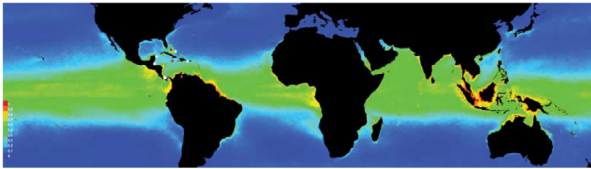
## *Halimeda gigas*



# Training samples = 14    # Test samples = 2    AUC (training) = 0.897    AUC (test) = 0.754    AUC ratio (training) = 1.384    AUC ratio (test) = 1.708

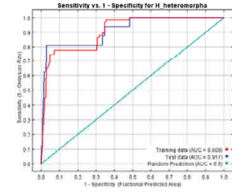
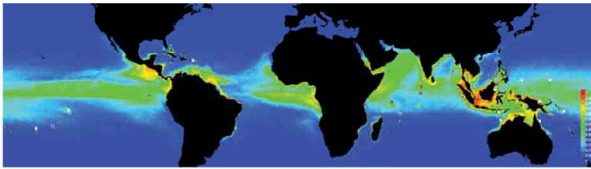
# Evolutionary niche dynamics in the seaweed *Halimeda*

## *Halimeda goreauii*



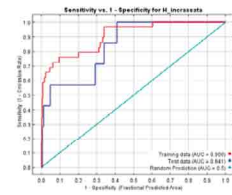
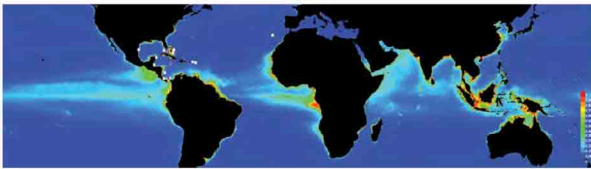
# Training samples = 13   # Test samples = 2   AUC (training) = 0.832   AUC (test) = 0.846   AUC ratio (training) = 1.488   AUC ratio (test) = 1.991

## *Halimeda heteromorpha*



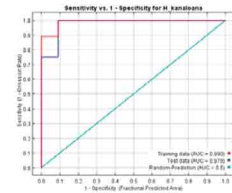
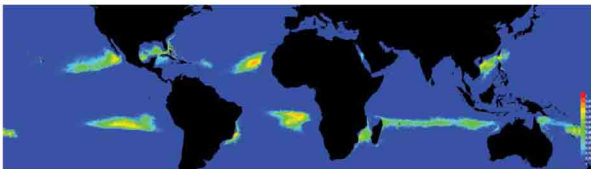
# Training samples = 67   # Test samples = 16   AUC (training) = 0.908   AUC (test) = 0.917   AUC ratio (training) = 1.470   AUC ratio (test) = 1.458

## *Halimeda incrassata*



# Training samples = 29   # Test samples = 7   AUC (training) = 0.906   AUC (test) = 0.841   AUC ratio (training) = 1.464   AUC ratio (test) = 1.411

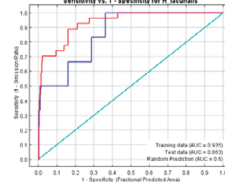
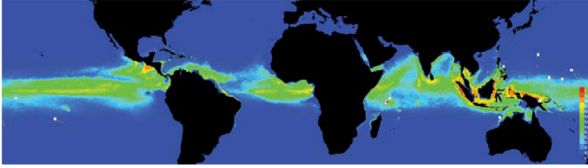
## *Halimeda kanaloana*



# Training samples = 18   # Test samples = 4   AUC (training) = 0.990   AUC (test) = 0.978   AUC ratio (training) = 0.940   AUC ratio (test) = 0.940

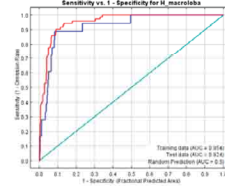
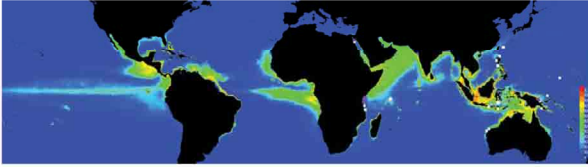
# Chapter 3

## *Halimeda lacunalis*



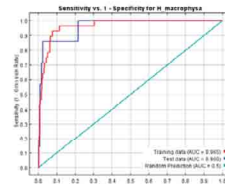
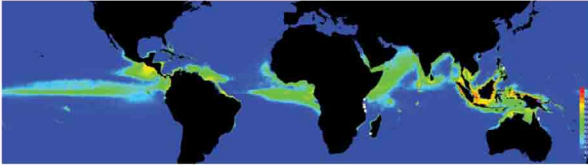
# Training samples = 27    # Test samples = 6    AUC (training) = 0.935    AUC (test) = 0.863    AUC ratio (training) = 1.569    AUC ratio (test) = 1.381

## *Halimeda macroloba*



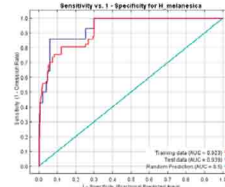
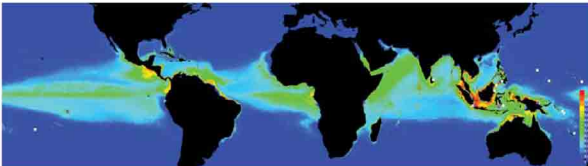
# Training samples = 72    # Test samples = 18    AUC (training) = 0.954    AUC (test) = 0.924    AUC ratio (training) = 1.806    AUC ratio (test) = 1.582

## *Halimeda macrophysa*



# Training samples = 28    # Test samples = 7    AUC (training) = 0.965    AUC (test) = 0.960    AUC ratio (training) = 1.837    AUC ratio (test) = 1.640

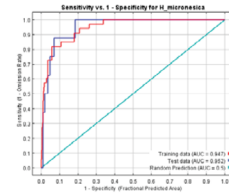
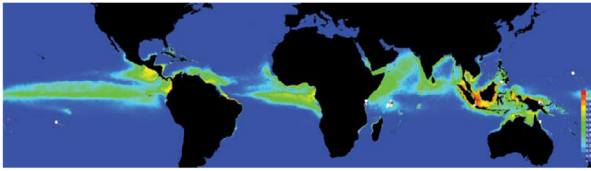
## *Halimeda melanesica*



# Training samples = 57    # Test samples = 14    AUC (training) = 0.923    AUC (test) = 0.939    AUC ratio (training) = 1.535    AUC ratio (test) = 1.583

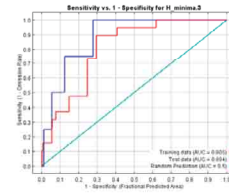
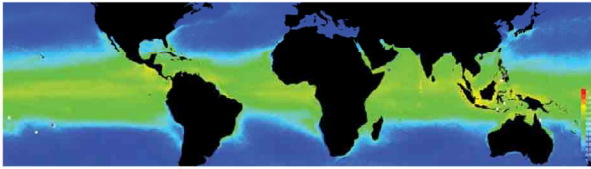
# Evolutionary niche dynamics in the seaweed *Halimeda*

## *Halimeda micronesica*



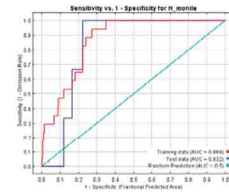
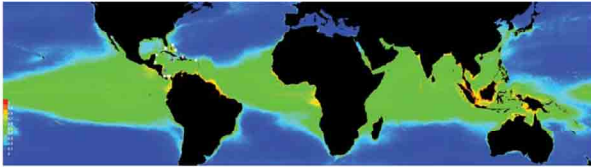
# Training samples = 33 # Test samples = 8 AUC (training) = 0.947 AUC (test) = 0.852 AUC ratio (training) = 1.675 AUC ratio (test) = 1.683

## *Halimeda minima*



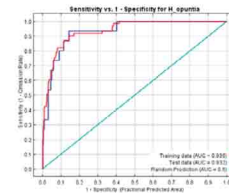
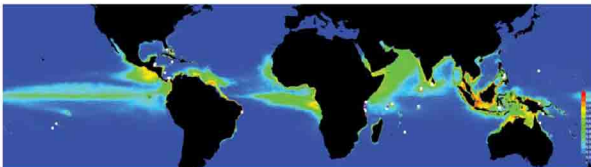
# Training samples = 19 # Test samples = 4 AUC (training) = 0.805 AUC (test) = 0.884 AUC ratio (training) = 1.388 AUC ratio (test) = 1.565

## *Halimeda monile*



# Training samples = 17 # Test samples = 3 AUC (training) = 0.866 AUC (test) = 0.832 AUC ratio (training) = 1.499 AUC ratio (test) = 1.531

## *Halimeda opuntia*

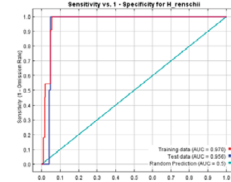
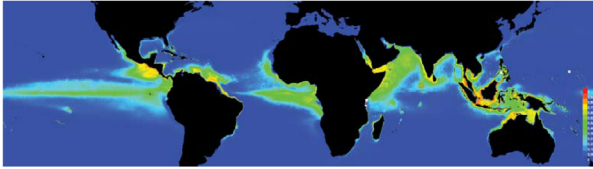


# Training samples = 61 # Test samples = 15 AUC (training) = 0.936 AUC (test) = 0.932 AUC ratio (training) = 1.707 AUC ratio (test) = 1.711



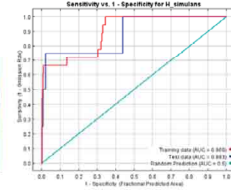
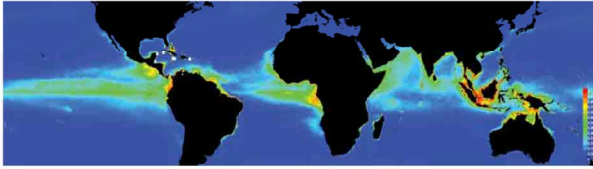
# Chapter 3

## *Halimeda renschii*



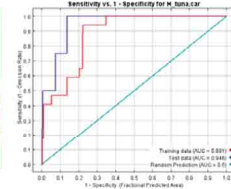
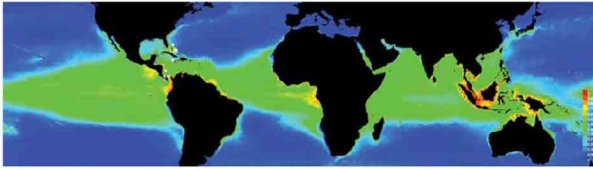
# Training samples = 11    # Test samples = 2    AUC (training) = 0.970    AUC (test) = 0.956    AUC ratio (training) = 1.902    AUC ratio (test) = 1.907

## *Halimeda simulans*



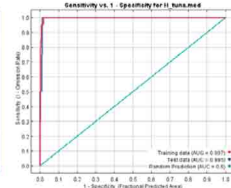
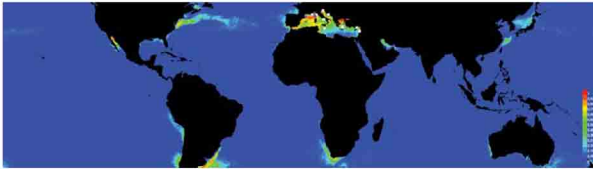
# Training samples = 18    # Test samples = 4    AUC (training) = 0.900    AUC (test) = 0.883    AUC ratio (training) = 1.495    AUC ratio (test) = 1.376

## *Halimeda tuna.car*



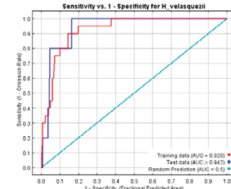
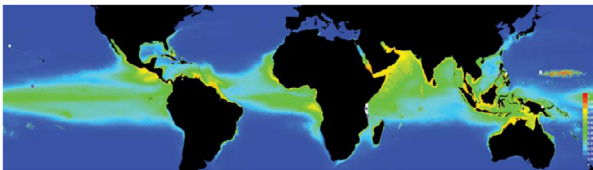
# Training samples = 17    # Test samples = 4    AUC (training) = 0.881    AUC (test) = 0.946    AUC ratio (training) = 1.621    AUC ratio (test) = 1.757

## *Halimeda tuna.med*



# Training samples = 20    # Test samples = 4    AUC (training) = 0.997    AUC (test) = 0.995    AUC ratio (training) = 1.984    AUC ratio (test) = 1.977

## *Halimeda velasquezii*



# Training samples = 20    # Test samples = 5    AUC (training) = 0.926    AUC (test) = 0.942    AUC ratio (training) = 1.719    AUC ratio (test) = 1.721

# CHAPTER 4

## **BIO-ORACLE: A GLOBAL ENVIRONMENTAL DATASET FOR MARINE SPECIES DISTRIBUTION MODELING**

Lennert Tyberghein<sup>1</sup>, Heroen Verbruggen<sup>1</sup>, Klaas Pauly<sup>1</sup>, Charles Troupin<sup>2</sup>, Frederic Mineur<sup>3</sup>, Olivier De Clerck<sup>1</sup>

<sup>1</sup> Phycology Research Group, Biology Department, Ghent University, Krijgslaan 281 S8, 9000 Ghent, Belgium

<sup>2</sup> GeoHydrodynamics and Environment Research (B5a), Université de Liège, Allée du 6 Août 17, B4000 Liège, Belgium

<sup>3</sup> School of Biological Sciences, Queen's University Belfast, 97 Lisburn Road, Belfast BT9 7BL, UK

Article published as: Tyberghein L, Verbruggen H, Pauly K, Troupin C, Mineur F & De Clerck O (2011) Bio-ORACLE: a global environmental dataset for marine species distribution modeling. *Global Ecology and Biogeography*, DOI: 10.1111/j.1466-8238.2011.00656.x

## ABSTRACT

**Aim** The oceans harbor a great diversity of organisms whose distribution and ecological preferences are often poorly understood. Species distribution modeling (SDM) could improve such knowledge and inform marine ecosystem management and conservation. Although marine environmental data are available from various sources, there are currently no user-friendly, high-resolution, global datasets designed for SDM applications. This study aims to fill this gap by assembling a comprehensive, uniform, high-resolution and readily usable package of global environmental rasters.

**Location** Global, marine

**Methods** We compiled global coverage data, e.g. satellite based and *in situ* measured data, representing various aspects of the marine environment relevant for species distributions. Rasters were assembled at a resolution of 5 arcmin (ca. 9.2 km) and a uniform landmask was applied. The utility of the dataset is evaluated by maximum entropy SDM of the invasive seaweed *Codium fragile* subsp. *fragile*.

**Results** We present Bio-ORACLE (Ocean Rasters for Analysis of CLimate and Environment), a global dataset consisting of 23 geophysical, biotic and climate rasters. This user-friendly data package for marine species distribution modeling is available for download at <http://www.bio-oracle.ugent.be>. The high predictive power of the distribution model of *Codium fragile* subsp. *fragile* clearly illustrates the potential of the data package for SDM of shallow-water marine organisms.

**Main conclusions** The availability of this global environmental data package has the potential to stimulate marine SDM. The high predictive success of the presence-only model of a notorious invasive seaweed shows that the information contained in Bio-ORACLE can be informative about marine distributions and permits building highly accurate species distribution models.

## INTRODUCTION

During the last two decades, interest in predicting species distributions has grown substantially. Species distribution modeling has become an important tool in ecology, evolution, biogeography and conservation biology (Peterson, 2006; Graham *et al.*, 2004). Common applications include predicting the spread of invasive species (Thuiller *et al.*, 2005), forecasting impacts of climate change (Thomas *et al.*, 2004), inferring spatial patterns of species diversity (Graham *et al.*, 2006) and reconstructing ancestral niches (Martinez-Meyer *et al.*, 2004).

Considering the strong interest in species distribution models (SDM) and their wide application in terrestrial ecosystems, remarkably few studies infer SDM of marine species (Robinson *et al.*, in press). Notable exceptions where robust predictions of geographic distributions of marine fauna and flora were made include studies on fish (Maravelias & Reid, 1997; Wiley *et al.*, 2003; Guinotte *et al.*, 2006), cold-water corals (Davies *et al.*, 2008; Tittensor *et al.*, 2009), jellyfish (Bentlage *et al.*, 2009), crabs (Compton *et al.*, 2010) and seaweeds (Graham *et al.*, 2007; Verbruggen *et al.*, 2009). Although these examples illustrate the utility of SDM for marine studies, several issues have restricted the application of SDM in the marine realm compared to the terrestrial environment. One challenge is that the extensive spatiotemporal variability characterizing the oceans can hinder SDM (Valavanis *et al.*, 2008; Franklin, 2009). A second obstacle is the restricted availability of marine data (Kaschner *et al.*, 2006). SDM algorithms require both high quality species occurrence records and environmental information to infer the macroecological preferences of species (Elith & Leathwick, 2009). Gathering reliable species occurrence records is not straightforward as collecting can be impeded for highly mobile, circumglobal or deep-sea organisms (Kaschner *et al.*, 2006). Furthermore, global marine environmental data, although increasingly available on the internet, are often challenging to use with popular SDM applications. The available data often have coarse spatial resolution and suffer from missing data in coastal regions. Data are frequently provided in different file formats and spatial resolutions, making the assembly of a dataset one of the most time-consuming aspects of marine SDM. WorldClim (Hijmans *et al.*, 2005), a freely available set of global high-resolution climate layers, has served this purpose for the terrestrial SDM community for the past five years but marine species distribution modelers have not had access to a similar pre-packaged dataset

(Robinson *et al.*, in press). Although some environmental datasets exist, such as Aquamaps (Kaschner *et al.*, 2008a), the Hexacoral project environmental data (Fautin & Buddemeier, 2008) and a set of layers related to human impact on marine ecosystems (Halpern *et al.*, 2008), they have not been widely applied in SDM studies.

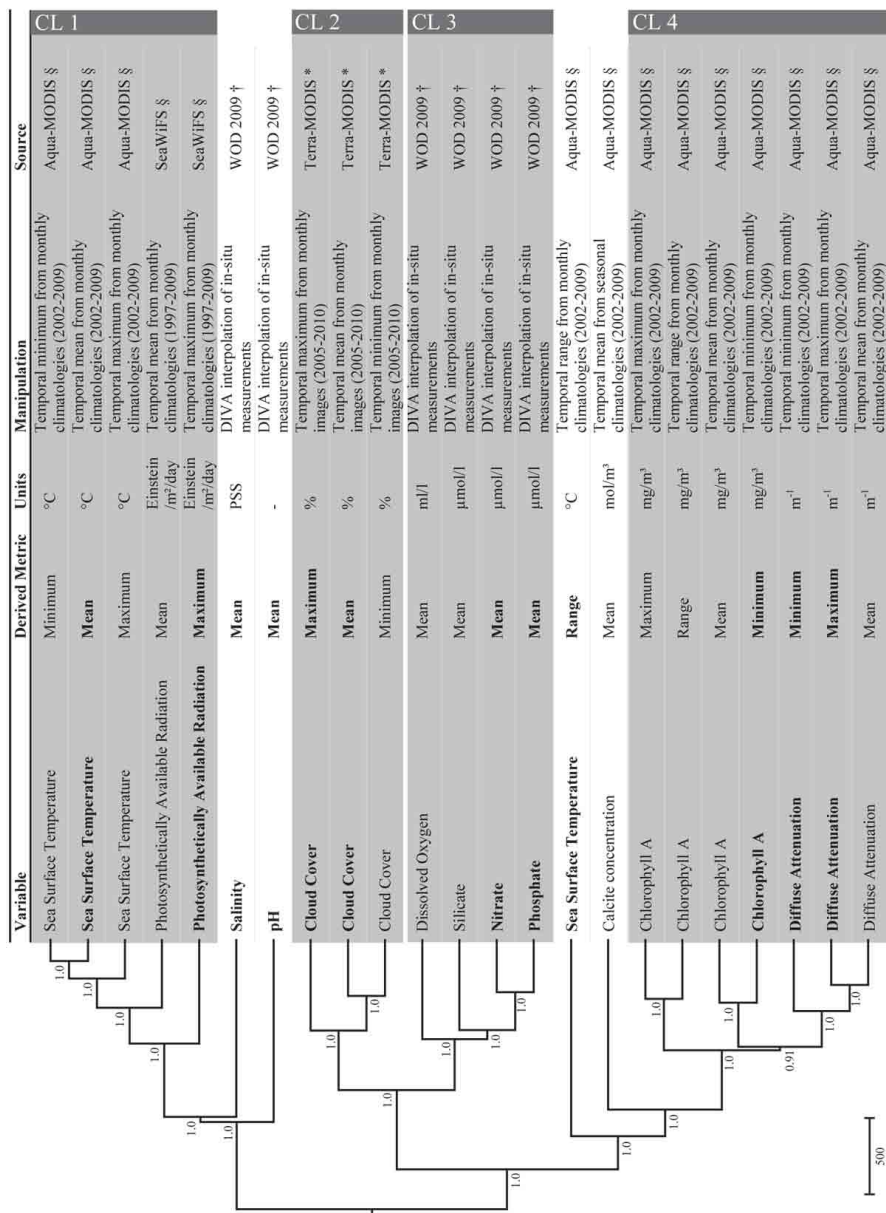
This study aims to facilitate marine species distribution modeling by assembling a comprehensive collection of global environmental rasters and supplying it ready for use in common species distribution modeling software. A broad set of macroecological variables representing environmental dimensions assumed to influence the distribution of marine shallow water organisms are packaged in the Bio-ORACLE database (Ocean Rasters for Analyses of CLimate and Environment). The utility of Bio-ORACLE for marine SDM is evaluated by modeling the distribution of *Codium fragile* subsp. *fragile*, a notorious invasive seaweed.

## MATERIALS AND METHODS

We compiled preprocessed remotely sensed and in situ measured oceanographic data representing various quantitative environmental predictors of species distributions. In the first place, we looked for proximal variables, i.e. those with a recognized physiological or ecological relevance for marine organisms. Secondly, we included several other variables that may serve as proxies for species' environmental requirements.

### REMOTELY SENSED DATA

Remotely sensed data were taken from various ocean observing satellite sensors (figure 1). We acquired monthly level-3 preprocessed satellite data (Aqua-MODIS and SeaWiFS; <http://oceancolor.gsfc.nasa.gov/>) at a 5 arcminute (ca. 9.2 km) spatial resolution. These geometrically corrected images are two-dimensional arrays with an equidistant cylindrical projection of the globe. Climatological composites, i.e. images summarizing information from the same month across several years, were used to generate three relevant metrics: annual maximum, minimum and mean. For sea surface temperature and chlorophyll A, the annual range (difference between maximum and minimum) was calculated as well. The latter metrics are biologically important as they represent proxies of seasonality and temporal variation in nutrient supply, respectively.



**Figure 1.** Dendrogram depicting agglomerative clustering of Bio-ORACLE data layers. Numbers at nodes indicate statistical support (bootstrap probabilities). Numbers range from 0 (no support) to 1 (maximal support). The table includes Bio-ORACLE variables, derived metrics, units, manipulation and source. Variables used to build the *Codium fragile* SDM are depicted in boldface. Grey shaded areas represent distinct clusters. Legend: † (Boyer *et al.*, 2009); § (Feldman & Mcclain, 2010); \* (Nasa, 2010).

We also included the Terra-MODIS-derived cloud fraction data (<http://modis-atmos.gsfc.nasa.gov/>) available in monthly composites at a resolution of 6 arcminutes. The monthly data over 10 years (2000-2009) were used to create average monthly composites and then further processed to produce the three standard metrics (annual mean, maximum and minimum). Eventually these layers were resampled to a resolution of 5 arcminutes using bilinear interpolation.

#### *IN SITU* MEASURED OCEANOGRAPHIC DATA

In addition to remotely sensed data, spatially interpolated data layers were developed from oceanographic in-situ surface measurements gathered from the World Ocean Database 2009 (WOD09) (Boyer *et al.*, 2009). We rejected all data flagged as erroneous in the WOD09 (Johnson *et al.*, 2009).

Different statistical approaches have been used to generate interpolated environmental surfaces (Daly, 2006). We have used DIVA (Data Interpolating Variational Analysis), a method developed for gridding in situ data using the variational inverse method (Brasseur & Haus, 1991) that has previously been applied to temperature, salinity and phosphate records of the Mediterranean Sea and the North-East Atlantic (Brasseur *et al.*, 1996; Karafistan *et al.*, 2002; Troupin *et al.*, 2010). Compared to other interpolation techniques, DIVA has a number of features that makes it very attractive for our application, including: (i) the ability to work with large amounts of data without intermediate averaging; (ii) the consideration of coastlines and topography; (iii) the generation of coherent error maps useful for identifying regions of uncertainty in the resulting environmental rasters; (iv) the use of a limited amount of parameters estimated in an objective way (Troupin *et al.*, 2010).

Coastlines were extracted from the global GEBCO one arcminute bathymetry (November 2009; <http://www.gebco.net/>). In order to perform DIVA analyses, two parameters have to be determined: the correlation length  $L$  and the signal to noise ratio  $\lambda$ . The software provides tools to estimate these parameters from the data. The correlation length ( $L$ ) was estimated by fitting the correlation between the data and a theoretical kernel function. As suggested by Troupin *et al.* (2010) and based on sensitivity, the signal-to-noise ratio ( $\lambda$ ) was assigned the constant value of one. The variance of the background field was also assigned a value of one to have DIVA generate relative error fields.

As a means of quality control, the outlier detection option was activated and error maps were calculated. Outlier detection is based on a comparison between the data-analysis residual and the expected standard deviation. The removal of data outside the realistic range was not necessary as this part of the quality control was carried out by the World Ocean Data center. The error estimates reflect the confidence in interpolated surfaces depending both on the data coverage and on their quality. Continuity across the 180° meridian was achieved by running two DIVA analyses, one ranging from -180° to 180° and a second starting and stopping at the 0° meridian (i.e. crossing the 180° meridian). The two resulting rasters were weight-smoothed into one final raster, with pixel weights decreasing linearly from the center of the input images to their sides.

#### PREPROCESSING AND MULTIVARIATE ANALYSIS OF BIO-ORACLE RASTERS

In order to arrive at a ready-for-use data package for SDM applications, a uniform landmask was applied to all data layers. This procedure consisted of correcting discrepancies between environmental data and the coastline, masking data pixels that were on land and calculating values for marine pixels without data by cubic extrapolation. Finally, the Polar Regions, which suffer from missing and imprecise data, were excluded by cropping the Bio-ORACLE rasters to latitudes between 70° N/S.

To better grasp the major environmental dimensions present in Bio-ORACLE, we explored the dataset with multivariate statistics. After the variables had been standardized, they were subjected to hierarchical clustering using 'hclust' of the 'stats' library in R (<http://www.r-project.org/>). This clustering technique was performed on a matrix of all pixel values in all rasters (7,257,600 pixels × 23 variables). We used the Euclidean distances and the average distance agglomeration method. Confidence in the hierarchical clustering was assessed by multiscale bootstrapping (Shimodaira, 2004) using the R package 'pvclust' (Suzuki & Shimodaira, 2006). We used 100 bootstrap replicates and the default relative sample sizes of bootstrap replications.

Environmental maps for SDM are usually provided in equidistant projections (north-south distances neither stretched nor compressed), but such maps may bias distribution models of species that span a large latitudinal range



(Tittensor *et al.*, 2009). For this reason and for applications like species richness analysis, equal-area cells may be preferred. To accommodate this, we remapped all Bio-ORACLE layers onto a Behrmann equal-area projection using ArcGIS 9.2 (<http://www.esri.com>). Coordinate transformation formulas are provided in Appendix S3. Both these equal-area and the original equidistant projections are provided to the public.

#### CASE STUDY: *CODIUM FRAGILE* SUBSP. *FRAGILE*

The ability of Bio-ORACLE to predict the distribution of marine species was evaluated by inferring a presence-only species distribution model of the invasive species *Codium fragile* subsp. *fragile* (hereafter *C. fragile*). This green alga is considered to be native in the NW-Pacific (i.e. Japan and surrounding areas) and has been introduced to Europe, the west and east coasts of North America, Australasia, South Africa and South America (Trowbridge, 1998; Provan *et al.*, 2008). Being a shallow-water marine organism, this species is a suitable candidate to evaluate the utility of Bio-ORACLE.

We collected occurrence records of the species in its native range and the invaded European range. Occurrences for Japan were acquired by georeferencing herbarium collections housed in the TNS (Tsukuba, Japan), SAP (Sapporo, Japan) and GENT (Ghent, Belgium) herbaria. For the European invaded range, occurrences were obtained from the primary literature and collecting activities. The resulting database consisted of 94 records for Japan and 284 for Europe.

Species distribution models were inferred with Maxent version 3.3.2. (Phillips *et al.*, 2006), a machine-learning algorithm for SDM with superior performance among presence-only algorithms (Elith *et al.*, 2006). To avoid modeling issues relating to overparameterization and multicollinearity of environmental layers, we adopted a predictor selection procedure (Guisan & Zimmermann, 2000). Variable reduction was achieved with performance-based forward-stepwise selection. Model performance was measured in terms of the area under the curve (AUC) of the receiver operating characteristic for test data as implemented in Maxent (Phillips *et al.*, 2006). All analyses were replicated five times with random training and test sets (both 50%). The test AUC can be expected to yield a good trade-off between underfitting and overfitting the model. Underfitting is avoided because this naturally leads to low AUC values,

and overfitting is countered by using the test AUC instead of the training AUC as the metric to be maximized. A recent study shows that this approach selects models of appropriate complexity (Warren & Seifert, in press).

The best subset of predictors, producing the highest test AUC value, was used to carry out the final modeling step. Final models were inferred by running the same subsampling procedure (50% training, 50% test data) with ten replicate runs. All analyses used linear and quadratic features (leaving the other Maxent settings at their default values).

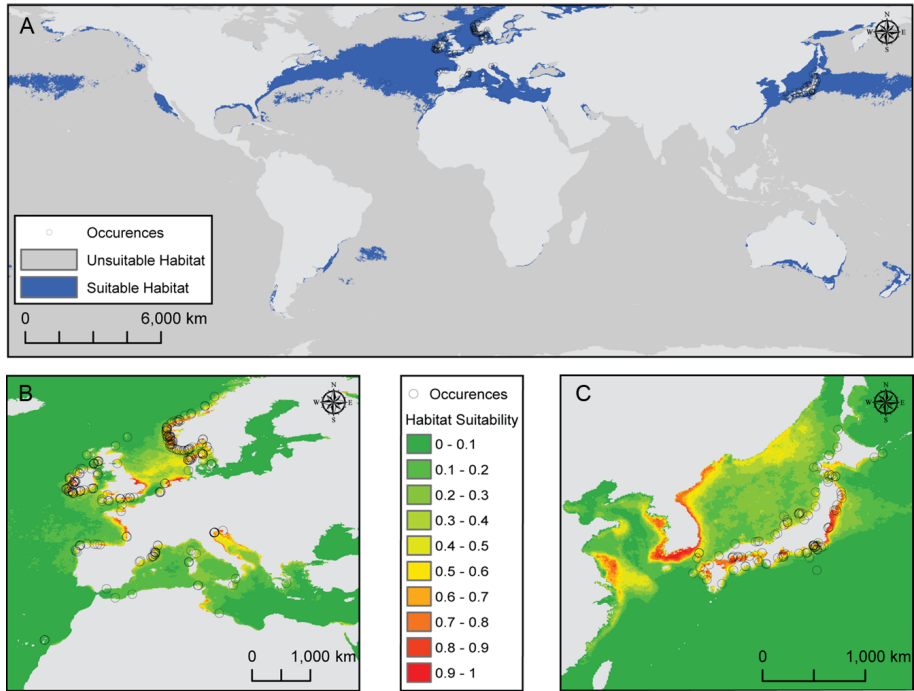
## RESULTS

Twenty-three global rasters of marine environmental predictors were generated either from remotely sensed or in situ measured oceanographic data. The gridded fields of 5 arcmin spatial resolution (9.2 km) are available for download on the Bio-ORACLE website (<http://www.bio-oracle.ugent.be>). Elaborate descriptions of the layers, their data sources and quality maps for interpolated layers can be found on the Bio-ORACLE website and in the supplementary material (Appendix S1).

Agglomerative clustering of the data layers resulted in a strongly structured dendrogram with high bootstrap values for nearly all clusters. Four clusters (CL1-4), representing major macroecological axes, were clearly discernable. They represent ocean color bio-optical parameters (CL4), nutrients and dissolved oxygen (CL3), clouds (CL2) and temperature and light resources denoting latitudinal patterns (CL1). Besides these four groups, four singletons were present in the dendrogram. These are the mean salinity, calcite concentration, water pH and the annual range of sea surface temperature, a measure of seasonality.

### CASE STUDY: *CODIUM FRAGILE* SUBSP. *FRAGILE*

The variable selection procedure resulted in thirteen sets of environmental layers that obtained the maximum test AUC of 0.989. From these sets, we opted to use the one with the fewest predictor variables for the final SDM, in order to avoid data redundancy and model overparameterization. This set consisted of 12 variables representing all six environmental clusters (variable names in boldface in figure 1).



**Figure 2.** Inferred species distribution model of *Codium fragile* subsp. *fragile* based on occurrence records from the native (Japan) and invaded range (Europe). (a) Binary prediction map applying a minimum training presence threshold (0.019). (b, c) Habitat suitability maps of European and Asian ranges, respectively. Warmer colors represent areas with better predicted conditions. Black circles indicate occurrence records used to build models.

The resulting presence-only model for *C. fragile* achieved high classification success (average AUC of ten replicate runs: 0.990). The predicted habitat suitability maps of the model are depicted in figure 2.

The SDM predicted the realized distributions of the species in both the native and European invaded ranges very well. Other predicted areas include South Australia, New Zealand, North America and parts of South America, where the species has also been introduced and is spreading. The annual range of sea surface temperature, annual maximum of diffuse attenuation and mean of phosphate concentration were the most important variables explaining recorded observations (see Appendix S2 for more comprehensive SDM results, including a full list of the variable contributions).

## DISCUSSION

Species distribution models play a central role in many fundamental and applied aspects of ecology. Besides improving our understanding of biogeography and dispersal barriers, they allow us to narrow in on species' ecological requirements, predict the effects of species invasions, habitat loss and climate change and can even lead to the discovery of new species (Peterson, 2006). The world's oceans harbor a high biodiversity (Costello *et al.*, 2010), and despite the importance of marine organisms for global biogeochemical cycles and human exploitation, their distribution and specific ecological needs are not nearly as well documented as their terrestrial counterparts. The application of SDM to marine species can also inform us about marine reserve design and conservation, and has the potential to predict how future climate and ocean acidification scenarios will affect the distribution and abundance of keystone species in the biogeochemical cycles.

While these examples clearly illustrate the need of SDM in the marine realm, the number of studies applying these methods to marine taxa is nearly an order of magnitude lower than those applying them to terrestrial taxa (Robinson *et al.*, in press). By introducing Bio-ORACLE, we hope to alleviate the need for user-friendly marine environmental data packages for global SDM applications. The development of a dataset of this type brings up several questions related to data quality, utility of the data for marine SDM, how the data should be used in practical applications, and how Bio-ORACLE compares to other marine environmental datasets. In what follows, we will touch on each of these topics.

## DATA QUALITY

Several precautions were taken to produce a dataset of the highest possible quality. These included selection of input data with a high level of a priori quality check, the use of state of the art interpolation techniques and an assessment of uncertainty about the resulting data rasters.

For the interpolated maps based on WOD09 data, quality is spatially variable and depends on local environmental variability, the quality and density of the observations and the interpolation method (Hijmans *et al.*, 2005). The error maps computed by DIVA (see Appendix S1) result from an analysis on the covariance field of the data with respect to the true field. Pixel values in

these error maps represent relative error (ranging from 0 to 1) and give an idea about the level of confidence in the pixel values in the corresponding environmental raster. The overall error is small, and the highest uncertainty, i.e. the highest predicted error, occurred in regions with low data coverage such as high latitude areas (e.g. the northern polar seas, Hudson Bay, Antarctica) and some unsampled areas in the middle of the oceans.

The rasters derived from remotely sensed data only included information with the highest possible quality pixels (resulting from Level-3 quality maps). However, an inevitable source of error results from the irregular temporal sampling of ocean color sensors (MODIS & SeaWiFS) (Gregg & Casey, 2007). Daily data gaps exist due to clouds, thick aerosols, inter-orbit gaps, sun glint and high solar zenith angles (Gregg & Casey, 2007). Binning these data into climatologies makes these gaps disappear but could lead to unpredictable biases. These biases (and resulting uncertainties) are most pronounced at high latitudes. For example, chlorophyll A, photosynthetically available radiation and diffuse attenuation, which are measured at relatively short wavelengths (in the visible spectrum), cannot be accurately measured during the winter season at high latitudes due to high solar zenith angles (Gregg & Casey, 2007). Sea surface temperature data do not suffer from this effect because they are measured in the thermal infrared part of the spectrum (longer wavelengths). A second source of bias affecting the quality of the remotely sensed data rasters follows from intra- and extrapolation of data for pixels with missing data. A simulation experiment (Appendix S1) indicates that extrapolations do lead to errors, but that these are small (generally < 1%) and that extrapolation into coastal pixels was somewhat more error-prone than interpolation of pixels in the open ocean.

The majority of potential errors in the datasets are due to the absence of in situ measurements, bias in remotely sensed data and extrapolation towards coastal areas. All of these problems are more pronounced in high-latitude areas. Therefore, we advise against the use of Bio-ORACLE in latitudes above 70° latitude N/S. We provide data rasters cropped at 70°N – 70°S as well as rasters spanning the entire latitudinal range, the latter only to be used with great vigilance and judicious consideration of results.

## UTILITY FOR MARINE SDM

The distribution of marine organisms is controlled by the interplay of a multitude of physical, chemical and biological variables. Bio-ORACLE includes both variables deemed important physiological determinants as well as potentially useful proxies. Each of the clusters and singletons in the dataset (figure 1) represents a distinct aspect of the marine macroecological environment. Temperature is thought to be the most important physical oceanographic variable determining the abundance, the spatial distribution and diversity of marine ectotherms (Schmidt-Nielsen, 1990; Lüning, 1990). Sea surface temperature clusters with photosynthetically available radiation (CL1), an essential and potentially limiting variable for photosynthetic organisms as it provides in their energy needs. Chlorophyll A (CL4) was included because it is a useful proxy for the trophic status of the surface waters (Duan *et al.*, 2007). However, the chlorophyll A metrics do not group with the actual nutrient layers (nitrate, phosphate and silicate) in the cluster analysis and consequently capture another dimension of the marine environment. We presume that this is due to an organic vs. inorganic nutrients dichotomy, chlorophyll A being a proxy of primary production and the WOD09 layers representing inorganic nutrients. Cloud cover variables (CL2) were included because of their potential to indirectly influence marine organisms. Clouds can block the transmission of light and harmful UV radiation and affect intertidal communities and organisms abounding in the ocean surface layer (Karentz & Bosch, 2001; Mangel *et al.*, 2010; Roleda *et al.*, 2005; Dring *et al.*, 1996). Four environmental entities did not cluster in one of the previously mentioned groups: salinity, calcite concentration, pH and sea surface temperature range. Besides temperature, salinity is known to be among the most important factors influencing marine life (Lüning, 1990; Gogina & Zettler, 2010). Ocean acidity (pH) plays a critical role in mediating physiological reactions (Wootton *et al.*, 2008) and numerous important groups of marine organisms have calcium carbonate skeletons that dissolve when pH drops (Doney *et al.*, 2009). The range in sea surface temperature is a measure of temperature seasonality. Our case study clearly shows that this variable can be an important determinant (or proxy) of marine species' distributions as it had the highest variable contribution to the SDM of *Codium fragile* (See Appendix S3). In contrast to the terrestrial environment where seasonal climatic variability increases with latitude (Chown *et al.*, 2004),

the seasonal variability in sea surface temperature is highest at intermediate latitudes (Clarke, 2009).

We aimed to illustrate the utility of Bio-ORACLE for marine SDM by generating a distribution model of *Codium fragile*. This highly invasive species is of economic interest and various aspects of its physiology and bioactive compounds, as well as the ecology and genetic signature of its invasion have been studied (Trowbridge, 1998; Provan *et al.*, 2008). The extremely high AUC values obtained for both training and test data sets show that the Bio-ORACLE rasters capture the macroecological preferences of the species and that, when used correctly (see below), the dataset permits building highly accurate SDM of marine species. From an organismal perspective, the SDM of *Codium fragile* is of interest in that it predicts other areas where the species could thrive. In fact, the species is known to have been introduced and is spreading in some of the predicted regions (Northern America, Australia, New-Zealand and South America) (Provan *et al.*, 2008). An in-depth analysis of the *Codium fragile* models is beyond the scope of this paper and will be presented elsewhere.

#### USING BIO-ORACLE FOR MARINE SDM

It is evident that ecological preferences differ between species and not all variables are useful in predicting a species' distribution. Choosing the right predictor variables for a particular species of interest is considered to be one of the most crucial steps in the SDM procedure (Guisan & Zimmermann, 2000). Variables could be chosen based on the knowledge that they are ecologically meaningful for the target species (Guisan & Zimmermann, 2000; Austin, 2002) and/or have good explanatory power (Araujo & Guisan, 2004). The latter aspect has been given much attention in regression techniques using presence-absence data, where several methods for predictor selection are available (e.g. stepwise selection with cross-validation, ridge regression, lasso) (Guisan & Thuiller, 2005). Unfortunately, predictor selection has been getting much less attention in recent presence-only modeling approaches.

For our case study, a performance-based forward stepwise variable selection procedure resulted in the selection of 12 out of the 23 variables in Bio-ORACLE. The importance of predictor selection is confirmed by the fact that a model built with all 23 layers resulted in considerably lower predictive power, most likely as a consequence of predictor collinearity and model

overparameterization. This illustrates that datasets like Bio-ORACLE and WorldClim should not be used blindly but that SDM requires meticulous species-specific variable selection, preferably based on a combination of physiological knowledge and variable selection approaches. In this context, the development of information criterion-based model selection (e.g. Akaike and Bayesian information criteria) for use in presence-only SDM applications would be useful (Warren & Seifert, in press).

#### COMPARISON TO OTHER MARINE ENVIRONMENTAL DATASETS

Bio-ORACLE was developed to be a ready-to-use global environmental dataset for shallow-water marine species distribution modeling. Other marine datasets for SDM do exist but the uniformity and user-friendliness of Bio-ORACLE is unique. Table 1 lists strengths and weaknesses of Bio-ORACLE compared to other marine environmental datasets. Noteworthy examples of marine preprocessed datasets that contain environmental data potentially informative for SDM are AquaMaps (Kaschner *et al.*, 2008a) and HexaCoral (Fautin & Buddemeier, 2008). AquaMaps is an approach to generate predictions of the natural occurrence of marine species based on their environmental tolerances (Kaschner *et al.*, 2008b). The AquaMaps datasets represent long-term averages of temporally varying environmental variables (Ready *et al.*, 2010). The HexaCoral datasets were developed to enable environmental classification (typology) and understand spatial and temporal patterns in biogeochemistry and biogeography. Both AquaMaps and HexaCoral can be downloaded at a spatial resolution of 30 arcminutes.

Common SDM applications require a set of uniformly constructed environmental layers. Bio-ORACLE provides data with a consistent landmask across all layers. We also made the data available in the ascii raster grid format used by many popular SDM algorithms (e.g. GARP, Maxent). It has been common practice in marine environmental modeling to use data with a spatial resolution of 30 arcminutes (Guinotte *et al.*, 2006). Bio-ORACLE has a considerably higher resolution of 5 arcminutes (ca. 9.2 km). Our choice for this resolution is a trade-off between the desire for sufficient resolution in near-shore environments, manageability of the rasters on current desktop computers and avoiding unreasonable interpolations.



**Table 1.** Comparison between freely available marine environmental datasets

	WOD2009 <sup>e</sup>	OCEAN COLOR <sup>f</sup>	HEXACORAL <sup>g</sup>	AQUAMAPS <sup>h</sup>	HALPERN <sup>i</sup>	Bio-ORACLE
<b>Resolution</b>	30-60 arcminutes (~ 55-110 km)	2.5-5 arcminutes (~ 4-9 km)	30 arcminutes (~ 55 km)	30 arcminutes (~ 55 km)	0.5 arcminutes (~ 1 km)	5 arcminutes (~ 9 km)
<b>Uniform landmask<sup>a</sup></b>	Yes	No	Yes	No	No	Yes
<b>Uniform geographic range</b>	Yes	Yes	No	Yes	No	Yes
<b>Dataset suitable for fine scale coastal studies</b>	No	Yes	No	No	Yes	Yes
<b>Multiple depth levels<sup>b</sup></b>	Yes	No	Yes	Yes	No	No
<b>Uniform file format</b>	Yes	Yes	Yes	Yes	Yes	Yes
<b>Uniform file format suitable for common SDM applications<sup>c</sup></b>	No	No	No	No	Yes <sup>d</sup>	Yes
<b>Equal-area grids available</b>	No	No	No	No	Yes	Yes

<sup>a</sup> Uniformity of landmask across all data layers in package, <sup>b</sup> Layers provided at different subsurface depths, <sup>c</sup> ascii raster grid format, <sup>d</sup> Suitability for SDM hampered by large file sizes.

References: <sup>e</sup> WOD2009 (Boyer et al., 2009), <sup>f</sup> Ocean Color (Feldman & McClain, 2010), <sup>g</sup> Biogeoinformatics of the Hexacorals (Fautin & Buddemeier, 2008), <sup>h</sup> AquaMaps (Kaschner et al., 2008b), <sup>i</sup> Global Mapping of Human Impacts to Marine Ecosystems (Halpern et al., 2008)

Furthermore, this resolution makes the dataset suitable for addressing questions about distributions at a global scale while still allowing model predictions at a resolution fine enough for most management purposes. In this context, it is important to note that the variables included in the Bio-ORACLE dataset are situated at the macroecological level, and when interpreting models at a fine spatial resolution, certain aspects of the organisms' microhabitat preferences (e.g. presence of suitable substrate for benthic species) become important to consider besides the macroecological niche dimensions.

Even though the comparison in Table 1 shows that Bio-ORACLE is currently among the best datasets for marine SDM, we also want to emphasize the utility of the other datasets. Our evaluation was focused on SDM applications and Bio-ORACLE was specifically designed for this purpose whereas others were not. Nonetheless, these datasets can complement Bio-ORACLE in various ways. For example, they contain some environmental dimensions that are not included in our database and variables at multiple depth levels.

## CONCLUSIONS AND PERSPECTIVES

Species distribution models have gained importance in various biological disciplines in recent years. Remarkably, they are less commonly used in studies of marine species than of terrestrial taxa. The present study was carried out to develop a marine counterpart of the WorldClim database, which is widely used for terrestrial SDM. Bio-ORACLE is a dataset consisting of 23 environmental rasters for marine species distribution modeling at a global scale. We hope that the availability of this set of environmental rasters will bring marine SDM on par with terrestrial studies. Our species distribution model of the invasive seaweed *Codium fragile* clearly shows that the rasters contain information relevant to the distribution of marine species and permits developing very accurate species distribution models.

We consider the present version of Bio-ORACLE an important first step in the development of a more complete set of environmental data rasters. Progress can obviously still be made, for example by including depth-related variables, various other physical parameters and layers representing important limiting nutrients in the marine environment. The present dataset will hopefully

provide our colleagues and us with the necessary groundwork to move this objective forward.

## **ACKNOWLEDGEMENTS**

We are grateful to three anonymous referees and the associate editor for their constructive criticisms and suggestions. We thank Satoshi Shimada (SAP herbarium) and Taiju Kitayama (TNS herbarium) for providing *C. fragile* subsp. *fragile* records. LT is funded by the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT). HV is a postdoctoral fellow of the Research Foundation – Flanders. Part of this work was carried out using the Stevin Supercomputer Infrastructure at Ghent University.

## SUPPLEMENTARY MATERIAL

### SUPPORTING INFORMATION S1

#### Appendix 1: Extended materials and methods used to build Bio-ORACLE dataset

##### General Approach

A set of 23 macroecological variables representing environmental dimensions assumed to influence the distribution of marine shallow water organisms was assembled. Our data collecting approach consisted of the compilation of preprocessed remotely sensed and in situ measured oceanographic data. We limited our analyses to variables relevant at the macroecological level and with global coverage. This supplementary material aims to give an overview and more detailed information of the Bio-ORACLE variables.

Our data processing pipeline consisted of the following steps:

- Data acquisition (WOD09 and Ocean Color Web)
- In case of WOD09 data, raster generation by interpolation with DIVA
- Conversion of raster to a convenient format: ESRI ascii
- Calculation of metrics (e.g. rasters with mean, minimum and maximum values)
- Application of uniform landmask
- Clipping rasters to 70°N – 70°S

##### Recommendations & Availability

The authors advise against the use of the Bio-ORACLE variables at latitudes higher than 70° N/S to reduce possible biases and potential errors in the data at high latitudes (see text for explanation).

Bio-ORACLE is released under the GNU General Public License and is available for download on <http://www.bio-oracle.ugent.be>.

##### Data layers

###### Remotely sensed data

Calcite concentration (mol/m <sup>3</sup> )	
<b>Original Spatial Resolution</b>	5 arcmin (9.2 km)
<b>Sensor</b>	Aqua-MODIS
<b>Data</b>	Seasonal climatologies
<b>Temporal Range</b>	2002 - 2009
<b>Brief description</b>	Calcite concentration indicates the concentration of calcite (CaCO <sub>3</sub> ) in oceans.
<b>Manipulation</b>	Derivation of metric: Mean
<b>Source</b>	Reference: (Feldman & McClain 2010) URL: <a href="http://oceancolor.gsfc.nasa.gov/">http://oceancolor.gsfc.nasa.gov/</a>

Chlorophyll A concentration (mg/m <sup>3</sup> )	
<b>Original Spatial Resolution</b>	5 arcmin (9.2 km)
<b>Sensor</b>	Aqua-MODIS
<b>Data</b>	Monthly climatologies
<b>Temporal Range</b>	2002 - 2009
<b>Brief description</b>	Chlorophyll A concentration indicates the concentration of photosynthetic pigment chlorophyll A (the most common "green" chlorophyll) in oceans. Please note that in shallow water these values may reflect any kind of autotrophic biomass.
<b>Manipulation</b>	Derivation of metrics: Mean, Minimum, Maximum, Range
<b>Source</b>	Reference: (Feldman & McClain 2010) URL: <a href="http://oceancolor.gsfc.nasa.gov/">http://oceancolor.gsfc.nasa.gov/</a>

Cloud fraction (%)	
<b>Original Spatial Resolution</b>	6 arcmin (11 km)
<b>Sensor</b>	Terra-MODIS
<b>Data</b>	Monthly images
<b>Temporal Range</b>	2005 - 2010
<b>Brief description</b>	Cloud fraction indicates how much of the earth is covered by clouds.
<b>Manipulation</b>	Derivation of metrics: Mean, Minimum, Maximum Bilinear interpolation (10 km → 9.2 km)
<b>Source</b>	Reference: (NASA 2010) URL: <a href="http://neo.sci.gsfc.nasa.gov/Search.html">http://neo.sci.gsfc.nasa.gov/Search.html</a>

Diffuse attenuation coefficient at 490 nm ( $m^{-1}$ )	
<b>Original Spatial Resolution</b>	5 arcmin (9.2 km)
<b>Sensor</b>	Aqua-MODIS
<b>Data</b>	Monthly climatologies
<b>Temporal Range</b>	2002 - 2009
<b>Brief description</b>	The diffuse attenuation coefficient is an indicator of water clarity. It expresses how deeply visible light in the blue to the green region of the spectrum penetrates in to the water column.
<b>Manipulation</b>	Derivation of metrics: Mean, Minimum, Maximum
<b>Source</b>	Reference: (Feldman & McClain 2010) URL: <a href="http://oceancolor.gsfc.nasa.gov/">http://oceancolor.gsfc.nasa.gov/</a>

Photosynthetically Available Radiation ( $Einstein/m^2/day$ )	
<b>Original Spatial Resolution</b>	5 arcmin (9.2 km)
<b>Sensor</b>	SeaWiFS
<b>Data</b>	Monthly climatologies
<b>Temporal Range</b>	1997 - 2009
<b>Brief description</b>	Photosynthetically Available Radiation (PAR) indicates the quantum energy flux from the Sun (in the spectral range 400-700 nm) reaching the ocean surface.
<b>Manipulation</b>	Derivation of metrics: Mean, Maximum Minimum PAR was considered, but excluded due to the high level of artifacts in original data.
<b>Source</b>	Reference: (Feldman & McClain 2010) URL: <a href="http://oceancolor.gsfc.nasa.gov/">http://oceancolor.gsfc.nasa.gov/</a>

Sea Surface Temperature ( $^{\circ}C$ )	
<b>Original Spatial Resolution</b>	5 arcmin (9.2 km)
<b>Sensor</b>	Aqua-MODIS
<b>Data</b>	Monthly climatologies
<b>Temporal Range</b>	2002 - 2009
<b>Brief description</b>	Sea surface temperature is the temperature of the water at the ocean surface. This parameter indicates the temperature of the topmost meter of the ocean water column.
<b>Manipulation</b>	Derivation of metrics: Mean, Minimum, Maximum, Range
<b>Source</b>	Reference: (Feldman & McClain 2010) URL: <a href="http://oceancolor.gsfc.nasa.gov/">http://oceancolor.gsfc.nasa.gov/</a>

### In situ measured oceanographic data

Dissolved oxygen (ml/l)	
<b>Database</b>	World Ocean Database 2009
<b>Data</b>	Standard Level Data: Ocean Station Data (OSD); High-resolution Conductivity-Temperature-Depth (CTD) (Surface)
<b>Temporal Range</b>	1898 - 2009
<b>Number of data points</b>	540582
<b>Brief description</b>	Dissolved oxygen concentration [O <sub>2</sub> ]
<b>Manipulation</b>	DIVA interpolation
<b>Source</b>	Reference: (Boyer <i>et al.</i> 2009) URL: <a href="http://www.nodc.noaa.gov/">http://www.nodc.noaa.gov/</a>

Nitrate ( $\mu\text{mol/l}$ )	
Database	World Ocean Database 2009
Data	Standard Level Data: OSD (Surface)
Temporal Range	1928 - 2008
Number of data points	189530
Brief description	This layer contains both $[\text{NO}_3]$ and $[\text{NO}_3+\text{NO}_2]$ data. By this we mean chemically reactive dissolved inorganic nitrate and nitrate or nitrite. (It is important to note that data reported as $[\text{NO}_3]$ in the WOD09 should be used with caution because it is difficult to verify that the $[\text{NO}_3]$ (nitrate) data are $[\text{NO}_3+\text{NO}_2]$ or $[\text{NO}_3]$ . (Boyer <i>et al.</i> 2009))
Manipulation	DIVA interpolation
Source	Reference: (Boyer <i>et al.</i> 2009) URL: <a href="http://www.nodc.noaa.gov/">http://www.nodc.noaa.gov/</a>

pH (unitless)	
Database	World Ocean Database 2009
Data	Standard Level Data: OSD (Surface)
Temporal Range	1910 - 2007
Number of data points	117833
Brief description	Measure of acidity in the ocean.
Manipulation	DIVA interpolation
Source	Reference: (Boyer <i>et al.</i> 2009) URL: <a href="http://www.nodc.noaa.gov/">http://www.nodc.noaa.gov/</a>

Phosphate ( $\mu\text{mol/l}$ )	
Database	World Ocean Database 2009
Data	Standard Level Data: OSD (Surface)
Temporal Range	1922 - 1986
Number of data points	226816
Brief description	Reactive ortho-phosphate concentration $[\text{HPO}_4^{2-}]$ in the ocean.
Manipulation	DIVA interpolation
Source	Reference: (Boyer <i>et al.</i> 2009) URL: <a href="http://www.nodc.noaa.gov/">http://www.nodc.noaa.gov/</a>

Salinity (PSS)	
Database	World Ocean Database 2009
Data	Standard Level Data: CTD (Surface)
Temporal Range	1961 - 2009
Number of data points	532377
Brief description	Salinity indicates the dissolved salt content in the ocean.
Manipulation	DIVA interpolation
Source	Reference: (Boyer <i>et al.</i> 2009) URL: <a href="http://www.nodc.noaa.gov/">http://www.nodc.noaa.gov/</a>

Silicate ( $\mu\text{mol/l}$ )	
Database	World Ocean Database 2009
Data	Standard Level Data: OSD & CTD (Surface)
Temporal Range	1930 - 2008
Number of data points	234417
Brief description	This variable indicates the concentration of silicate or ortho-silicic acid $[\text{Si}(\text{OH})_4]$ in the ocean.
Manipulation	DIVA interpolation
Source	Reference: (Boyer <i>et al.</i> 2009) URL: <a href="http://www.nodc.noaa.gov/">http://www.nodc.noaa.gov/</a>

### References

- Boyer T.P., Antonov J.I., Baranova O.K., Garcia H.E., Johnson D.R., Locarnini R.A., Mishonov A.V., O'Brien T.D., Seidov D., Smolyar I.V. & Zweng M.M. (2009). *World Ocean Database 2009*. U.S. Gov. Printing Office, Washington D.C.
- Feldman G.C. & McClain C.R. (2010). Ocean Color Web. URL <http://oceancolor.gsfc.nasa.gov/>
- NASA (2010). NASA Earth observations (NEO). URL <http://neo.sci.gsfc.nasa.gov/Search.html>



Simulation study: Cubic inter/extrapolation

In order to produce rasters with a uniform landmask across the complete dataset two operations were necessary:

- Values that fell on land needed to be masked;
- Missing values needed to be estimated.

The latter operation was done using a cubic inter/extrapolation algorithm. We chose this algorithm because of its stable and smooth characteristics. To assess the quality of the interpolation we performed a simulation study on the remotely sensed data\*.

The evaluation consisted of:

- Random selection of a monthly climatology of each remotely sensed variable;
- Random selection of 100 'open sea' pixels (i.e. pixels that are completely surrounded by other sea pixels) from the respective climatology;
- Random selection of 100 'coastal' pixels (i.e. pixels directly adjacent to one or more no-data value) from the respective climatology;
- Removal of these 200 data points from the respective climatology;
- Application of the cubic inter/extrapolation algorithm;
- Extraction of data from both the original and the inter/extrapolated layer;
- Calculation of the average difference between both layers;
- Evaluation of that difference in relation to the total variable range.

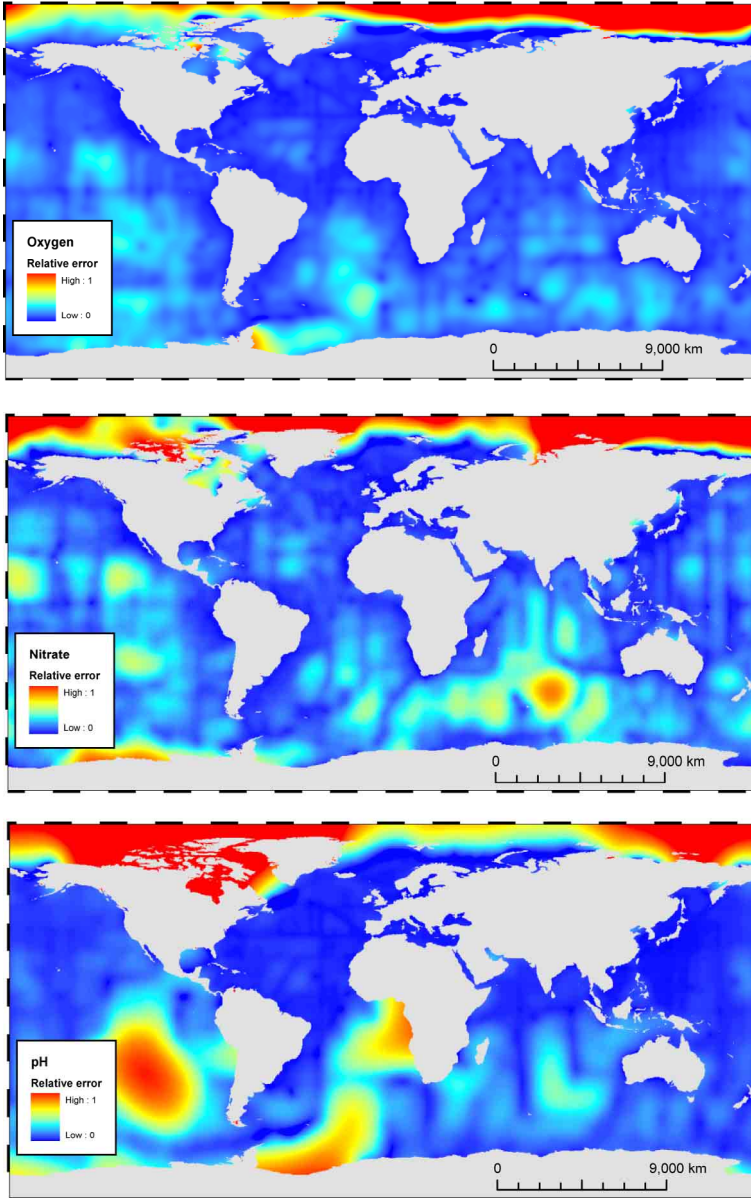
The results are shown in the table below:

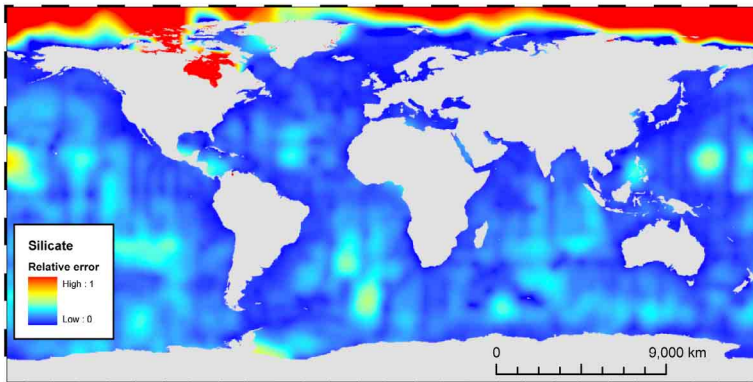
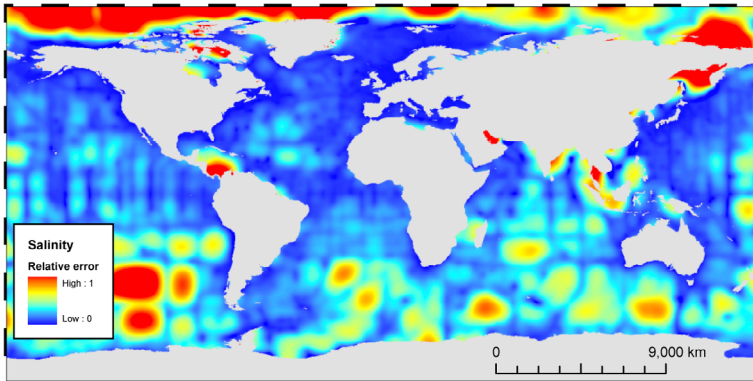
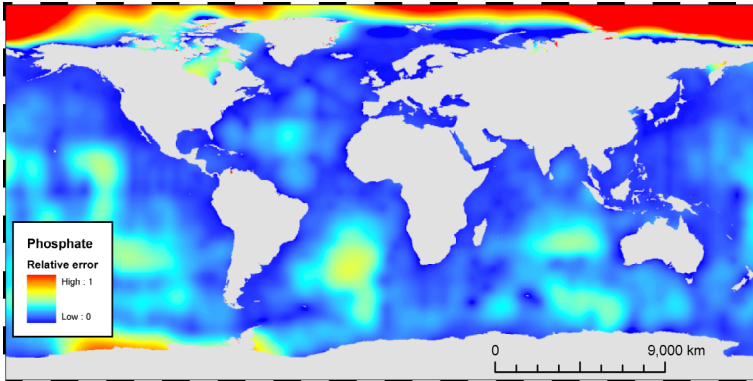
Variable	Coast		Open Sea	
	$\Delta(\text{avg}(\text{Interpol} - \text{Real}))$	% of total variable range	$\Delta(\text{avg}(\text{Interpol} - \text{Real}))$	% of total variable range
Sea Surface Temperature	0.22675 °C	0.57	0.04161 °C	0.11
Chlorophyll A Concentration	0.38901 mg/m <sup>3</sup>	0.60	0.00831 mg/m <sup>3</sup>	0.01
Diffuse Attenuation	0.00922 m <sup>-1</sup>	0.80	0.00088 m <sup>-1</sup>	0.08
Photosynthetically Available Radiation	0.49984 Einstein/m <sup>2</sup> /day	0.73	0.18093 Einstein/m <sup>2</sup> /day	0.27
Calcite Concentration	0.00086 mol/m <sup>3</sup>	1.54	0.000134 mol/m <sup>3</sup>	0.24

\* Cloud cover data was not taken into account as no pixels needed to be inter/extrapolated.



Supporting figure S1. Error maps computed by DIVA.



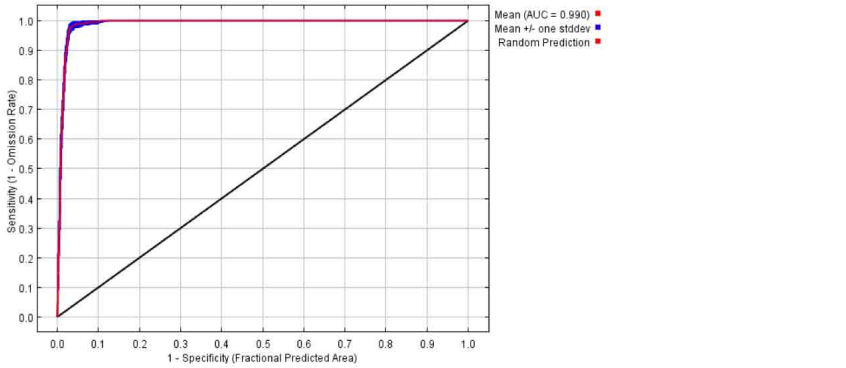


**SUPPORTING INFORMATION S2**

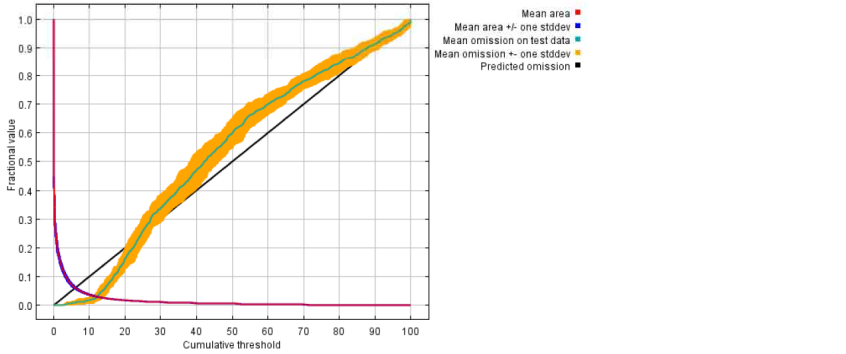
**Appendix 2: Extended results: Species Distribution Modeling: *C. fragile* subsp. *fragile***

This appendix summarizes the results of 10 replicate Maxent runs (see text for explanation). Figures are developed by Maxent version 3.3.2. (Phillips *et al.* 2006).

Analysis of omission/commission



**Figure S2.1** The following figure shows the receiver operating characteristic (ROC) curve, averaged over the replicate runs. The average test AUC for the replicate runs is 0.990, and the standard deviation is 0.001.



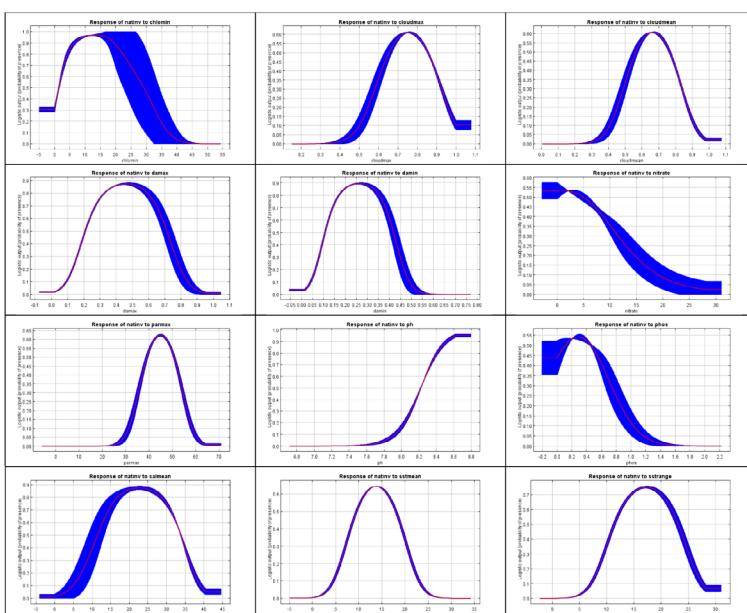
**Figure S2.2** This figure shows the test omission rate and predicted area as a function of the cumulative threshold, averaged over the replicate runs.

Analysis of variable contributions

**Table S2.1** This table represents an estimate of relative contributions of the environmental variables to the Maxent model. Values shown are averages over replicate runs.

Variable	Percent contribution
Sea surface temperature (RANGE)	36.3
Diffuse attenuation (MAX)	15.6
Phosphate concentration	12.5
Sea surface temperature (MEAN)	10.8
Salinity (MEAN)	8.8
PAR (MAX)	7
Cloud cover (MAX)	3.9
Nitrate concentration (MEAN)	1.7
Cloud cover (MEAN)	1.5
Diffuse attenuation (MIN)	1.1
Chlorophyll A concentration (MIN)	0.7
pH (MEAN)	0

Response curves



**Figure S2.3** These curves represent models created using only the corresponding variable. They reflect the dependence of predicted suitability both on the selected variable and on dependencies induced by correlations between the selected variable and other variables.

Reference

Phillips S.J., Anderson R.P. & Schapire R.E. (2006). Maximum entropy modeling of species geographic distributions. *Ecol. Model.*, 190, 231-259.

**SUPPORTING INFORMATION S3**

**Appendix 3: Transformation methods ‘Behrmann Projection’**

There exist several ways to transform geographical coordinates (lat/lon WGS84) to a Behrmann equal-area projection.

1. Using ArcMap 9.2 (<http://www.esri.com>)
  - Import pointfile as shapefile
  - Transformation: ArcToolbox → Data Management Tools → Projections and Transformations → Feature → Project
  
2. Using transformation formulae

The Behrmann projection is a cylindrical equal-area projection with fixed standard parallels at  $\pm 30^\circ$ . The transformation equations are:

$$x = R \lambda \cos \varphi_0$$
$$y = \frac{R}{\sin \varphi_0} \sin \varphi$$

where  $\lambda$  is the longitude,  $\varphi$  is the latitude and  $\varphi_0$  is the standard latitude, all expressed in radians. R represents the earth radius (6378000m).

These formulas yield approximate coordinates because differences in the earth's radius are not taken into account.

# CHAPTER 5

## **SPATIAL SCALE-DEPENDENT PREDICTION IN MARINE DISTRIBUTION MODELING: A CASE STUDY**

Klaas Pauly, Lennert Tyberghein<sup>1</sup> & Olivier De Clerck

<sup>1</sup> Phycology Research Group  
Biology Department, Ghent University  
Krijgslaan 281/S8, 9000 Ghent, Belgium

## **ABSTRACT**

Spatial scale, defined as the spatial resolution and extent of environmental data, is known to influence species distribution models. With most studies investigating spatial scale effects in the context of predicting impacts of global change in the terrestrial realm, comparative species distribution modeling based on fine-scale versus coarse-scale environmental data in the marine realm remains to be investigated. We present a regional case study for three benthic seaweed species with different well-defined distribution patterns in two adjacent but contrasting seas: the spatially and temporally more homogeneous Gulf of Oman and the highly variable Arabian Sea. Rather than starting from a single environmental dataset with subsequent up- or downscaling, a genuine sub-100m resolution environmental dataset was compiled based on 10 mosaiced Landsat scenes for winter and summer. This resulted in habitat layers pertaining to sea surface temperature, nutrient content, turbidity and substrate availability. The coarse-scale dataset (9km resolution) is based on the global environmental dataset Bio-ORACLE, cropped to the same 2000-km long coastline. Models for the three species were generated with the Maxent algorithm using both environmental datasets. The Landsat-based models performed equally well in terms of AUC compared to the Bio-ORACLE models. However, important differences in output maps could be noted, capturing the difference between coarse-scale macroecological modeling and fine-scale habitat modeling. Overall, it appears that while both coarse and fine-scale models are in good agreement for all species in the less variable Gulf of Oman, coarse-scale models suffered from overprediction for all three species in the more heterogeneous Arabian Sea. This suggests that the choice of modeling scale is important for applied marine modeling studies such as the guiding of field work.

## INTRODUCTION

Ecological niche modeling (ENM), alternatively termed species distribution modeling (SDM), habitat suitability modeling (HSM) or bioclimate envelope modeling has recently been widely applied in spatially explicit studies on forecasting climate change impacts (Wiens et al., 2009), forecasting the spread of invasive species (Thuiller et al., 2005), inferring spatial patterns of species diversity (Gillespie et al., 2008), guiding field work (Le Lay et al., 2010) and discovering new species (Raxworthy et al., 2003) or new occurrences of rare species (Williams et al., 2009). To date, most of these studies have focused on the terrestrial realm, with marine studies lagging behind in assessing influences of different factors on the modeling process (Robinson et al., 2011).

With environmental raster layers now widely available for global and regional research topics, several terrestrial studies have characterized the effects of modeling species' distributions at different spatial scales (Guisan et al., 2007). Spatial scale in SDM encompasses both spatial extent and spatial resolution (the latter often termed grain size) of the datasets used in SDM, and is most often used to refer to the environmental data. The limiting factors of data storage and computing power on the one hand and data sources such as satellite imagery on the other hand have caused a trade-off towards moderate or low-resolution environmental data used for studies on a global or continental extent, and moderate to (very) high resolution data used for studies on a regional to local extent. In order to avoid this trade-off, some studies have adopted a nested (or hierarchical) approach in capturing both local and regional patterns (Elith & Leathwick, 2009), or combined both high- and low-resolution data through resampling (Kulkarni et al., 2010). In fact, the modeled ecological processes differ over scales: whereas models using coarse resolution data yield the potential distribution of the macroecological niche, fine resolution models are indicative of (micro-) habitat-level processes (Elith & Leathwick, 2009). Next to ecophysiological or climatic driving variables, fine resolution data can also include habitat variables such as land cover, soil type, canopy structure, topography or other factors determining true habitat suitability within a suitable climate or macroecological niche (Bradley & Fleishman, 2008; Lahoz-Monfort et al., 2010). Habitat variables generally vary on a finer scale



compared to climatic variables. Some terrestrial studies have examined these factors by modeling species' distributions based on a set of moderate resolution environmental data, followed by scaling the same data up or down (Guisan et al., 2007). Others have made use of original data on both a moderate resolution and high resolution (Trivedi et al., 2008), the latter mainly facilitated by Landsat imagery processing (Kulkarni et al., 2010).

While a few terrestrial studies have compared distribution models based on small- and large scale data, scale influences on modeling studies in the marine realm remain to be explored. We present a case study along the coast of the Sultanate of Oman, covering approximately 2000km from the Gulf of Oman in the north to the Arabian Sea in the south. This region was selected because of the contrasting environmental conditions governing the two adjacent seas. The eastern cape of Ra's al Hadd forms a sharp transition between the strongly seasonal climate of the Arabian Sea and the year-round warmer Gulf of Oman. The former is characterized by clear, warm waters during winter monsoon and agitated, nutrient-rich and cold upwelling waters during summer monsoon, while the latter doesn't experience seasonal upwelling (Wilson, 2000). This difference in climatic regimes causes a sharp turnover in biotic communities, as demonstrated for seaweeds by Schils & Wilson (2006). This allows for coarse and fine-scale model comparisons on regions characterized by a differing environmental variability and on species occurring either (mostly) in one of the seas in the study area or in both.

In this paper, we present (1) a first application of Landsat scene mosaicing to extract environmental and habitat layers for regional distribution modeling in the marine realm at a sub-100m resolution over approximately 2000km of coastline and (2) a comparison of these Landsat-based habitat suitability models with niche models based on 9km resolution macroecological layers available in the Bio-ORACLE dataset (Tyberghein et al., 2011). The latter dataset was recently released in order to facilitate marine SDM studies by providing uniformly compiled global data from remote sensing products and in situ data interpolation and can be easily cropped for regional applications. The goal of this comparison is to investigate whether spatial resolution affects marine distribution modeling by revealing habitat patterns within macroecologically discerned niches for

regions with differing environmental variability, in a similar way as demonstrated in terrestrial niche modeling studies.

## **MATERIALS & METHODS**

### **BIOTIC DATA**

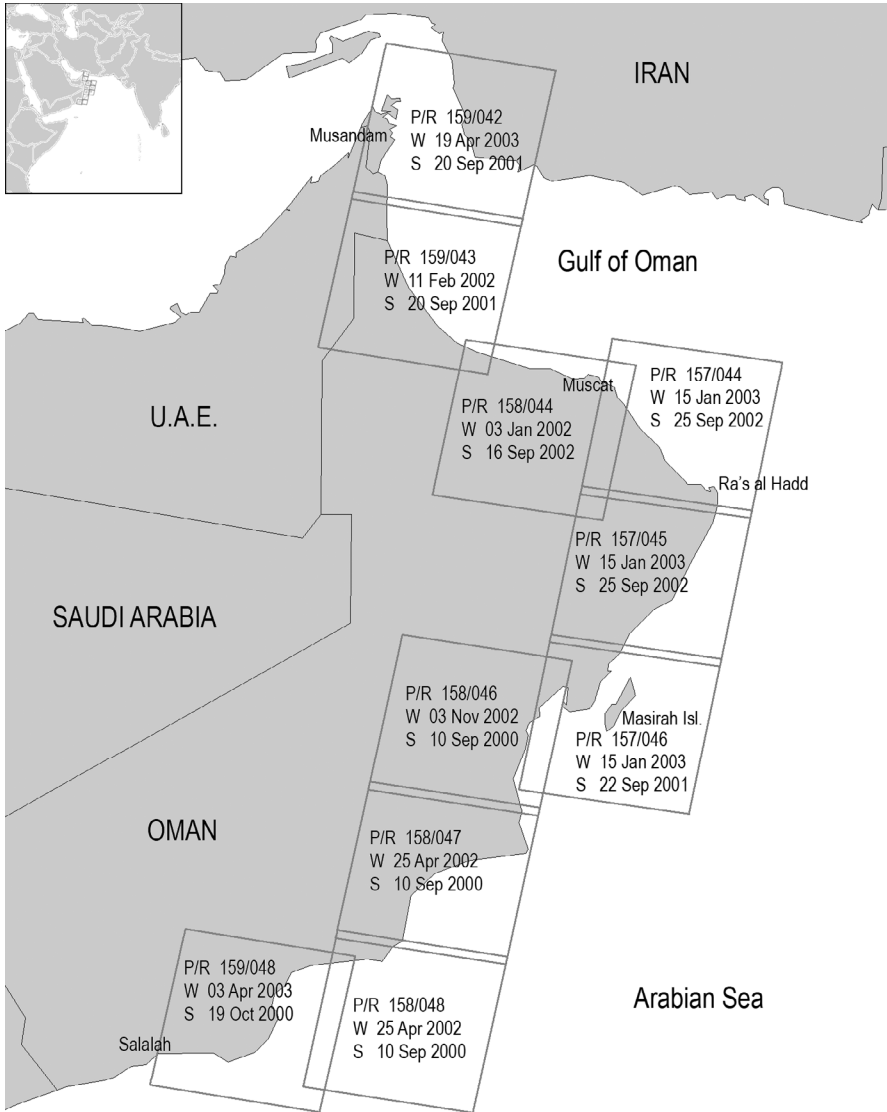
Three macroalgal species were selected for niche modeling: the conspicuous brown kelp-like alga *Nizamuddinia zanardinii* (Schiffner) P.C. Silva, the red algal turf *Tolytiocladia glomerulata* (C. Agardh) F. Schmitz and the green calcareous alga *Halimeda discoidea* Decaisne. Georeferenced records were available from 7 sampling campaigns between 2003 and 2008, covering the entire coastline of Oman and both monsoon seasons. Voucher specimens are available in the GENT herbarium (Schils & Coppejans, 2003; Schils & Wilson, 2006). Because of cryptic diversity, specimens of *H. discoidea* were only included if their identity was confirmed by genetic analyses. The species were selected because of the availability of at least 10 occurrence records per species and their distinct ecologies. *N. zanardinii* (14 occurrence records) is known to be restricted to the Arabian Sea coast of Oman and grows on exposed rocky platforms in the shallow subtidal. *T. glomerulata* (18 records), although occasionally found in sheltered localities in the Arabian Sea, is far more common and abundant in the Gulf of Oman and occurs on rocks and sand-covered rocky platforms in the intertidal and shallow subtidal. *H. discoidea* (10 records) favors sheltered sandy stretches in the shallow and deeper subtidal in both the Arabian Sea and the Gulf of Oman, although the species is somewhat less common in the latter region.

### **ENVIRONMENTAL DATA**

#### ***Landsat data***

Ten Landsat 7 ETM+ scenes from both the southeast (summer) and northwest (winter) monsoon season were selected to cover the entire coastline of Oman, with the exclusion of the Yemen-Oman border region west of Salalah due to its location in a different UTM zone (figure 1). Scene selection was restricted to the SLC-ON mode (from 2000-2003, prior to the failure of the scan line corrector or SLC) and sufficiently low cloud cover. The SLC failure after 2003 doesn't prevent single scenes from being used,

but SLC-off images cannot be used for mosaicing since approximately one-fourth of a scene is not imaged, resulting in wedge-shaped missing data chunks at the along-track edges.



**Figure 1.** Location of the study area showing WRS-2 path/row tiles from which Landsat scenes were downloaded. For each tile, a winter (W) and a summer (S) scene were selected from the given dates. Some major coastal localities are given for reference: from north to south the Musandam peninsula, the Muscat Capital Area, the Ra's al Hadd cape separating the Gulf of Oman and the Arabian Sea, Masirah Island and the southern province capital Salalah.

Due to the cloud cover criterion, suitable summer monsoon scenes were only available for September-October, when the intensity of monsoonal activity is decreasing. Winter monsoon scenes were available for November-April. By and large, one suitable scene per combination of path/row tile and season was selected. For each scene, bands 1-4 (visible and near-infrared, VNIR), band 5 (short-wave infrared, SWIR) and band 6H (high-gain thermal infrared, TIR) were downloaded at Level 1T (terrain corrected, projected to UTM 40N) from the USGS archive, freely available at <http://glovis.usgs.gov>. The VNIR and SWIR bands were first corrected to at-sensor spectral radiance and then corrected for variations in solar output, haze and atmospheric absorption and scattering based on acquisition time and the cosine of the solar zenith angle, resulting in the conversion of digital numbers to planetary reflectance using standard image-based practices (Chander et al., 2009; NASA, 2011). TIR digital numbers (DN) were converted to surface temperature following atmospheric correction detailed in appendix 1. Subsequently, individual scenes were resampled (nearest neighbor) to 60m resolution (i.e., the native resolution at which Landsat 7 ETM+ samples the thermal region) and linearly rescaled to reduce file size. Next, for each season individual scenes were mosaiced with an image grey level matching algorithm based on the average pixel values of adjacent scenes in order to account for between-scene variations due to different acquisition dates. Thus, for each of the 6 bands, both a winter and a summer monsoon mosaic were assembled, covering a coastline of about 2000km. The winter mosaic was remapped onto the summer mosaic to eliminate marginal differences in pixel positioning and image extent. UTM projection was retained throughout this study because of its precise georeferencing characteristics necessary to relate biotic occurrence data to the environmental information from the corresponding 60m grid cells, while scale differences across the study area are negligible. Lastly, a general land and cloud mask was created based on SWIR thresholds of both mosaics. The mask image was manually edited to additionally mask false water pixels in mountainous areas (due low SWIR reflectance in hill-shaded valleys) and erroneous values along scene edges.

Several macroecological and habitat variables thought to play an explanatory role in the distribution of macroalgae were derived from the band mosaics. Next to sea surface temperature (SST) which was directly

inferred from the thermal band, the band 2 (green) mosaic was directly retained as a measure for suspended matter and, hence, turbidity (Ahn et al., 2006). Additionally, a NIR/red ratio index correlated with chlorophyll-a concentration (CHL, from phytoplankton; Duan et al., 2007) was used as a proxy for primary production correlated with organic nutrient availability. Based on the winter and summer monsoon mosaics, the minimum, maximum, average and range metrics for each of these variables were calculated as biologically relevant input layers in the modeling step.

Furthermore, a substrate map was obtained by supervised maximum likelihood classification of the VNIR winter mosaic bands (yielding better classification results due to less turbid waters compared to the summer monsoon) resulting in a categorical variable. Due to the influence of original between-scene variability in the resulting mosaic and the limited availability of training and testing samples compared to the total size of the study area, a distinction was made between soft (sand, mud and seagrass meadows) and hard substrates (bare and vegetated rock, coral) only. Pixels were sampled using flood polygons based on georeferenced transects along the entire coastline of Oman (with deep water pixels included as a separate class and randomly sampled from beyond 50m depth based on a digitized nautical chart). From these samples, 50% was used for training and testing each. The overall classification accuracy was 86%, with main confusions of hard substrate close to the coast being classified as deep water and some confusion of soft substrate classified as hard substrate, mainly in silted areas with seagrass meadows and isolated coral communities.

### ***Bio-ORACLE data***

Environmental grids at 5 arcmin resolution were downloaded from <http://www.bio-oracle.ugent.be> (Tyberghein et al., 2011), cropped to the area covered by the Landsat mosaics and projected to UTM 40N. The following grids were selected to represent the equivalent of Landsat environmental data: SST, CHL and diffuse attenuation (DA, equivalent with turbidity), for which minimum, maximum, average and range data were available. Additionally, photosynthetically available radiation (PAR), calcite concentration and salinity were included in the modeling effort because these were also derived from seasonal data. Only the average was available

for calcite and salinity, while the average and maximum were available for PAR.

## DISTRIBUTION MODELING

Because of the relatively low number of occurrence records and inherent uncertainty on absences in algal sampling, a maximum entropy-based presence-only distribution modeling approach was adopted using Maxent v3.3.3. Maxent is a general purpose algorithm for inferring probability distributions based on incomplete information. In its species distribution modeling implementation, Maxent finds the most equally spread out probability distribution constrained by the environmental variables, their interactions and functions thereof at the known occurrence localities (Phillips et al., 2006; Phillips & Dudik, 2008). It has been shown to perform well in comparison with other modeling algorithms (Jane Elith et al., 2006), especially with low numbers of occurrence records (Pearson et al., 2007). For the Landsat-based models, selection of a background sample of 10,000 random points was restricted to the coastal zone defined by the penetration depth of the visible light bands in the final mosaic (i.e. classified as either hard or soft substrate, excluding deep water in the substrate layer). This approach was adopted to cope with the high number of grid cells in the study area, by providing the algorithm with a background sample that is similarly biased compared to occurrence data (Phillips et al., 2009; Elith et al., 2010). Due to the low number of grid cells in models based on Bio-ORACLE layers cropped to the region, all marine cells in the study area had to be included as background values. For each species-environmental dataset combination, preliminary Maxent model runs were performed to achieve forward stepwise variable selection based on the area under the receiver operating characteristic curve (AUC) of the test dataset as the selection criterion in order to reduce input data redundancy in the final model runs. Variable selection runs were executed with 50% of the occurrence records as test data in 10 replicate runs using subsampling, while final model runs were made with 5-fold cross-validation.

## DATA EXPLORATION & MODEL ANALYSIS

In order to assess environmental variability patterns in the input layers, environmental data values for the different input layers selected in the final models were sampled separately for the Gulf of Oman and the Arabian Sea from all cells within a 2-pixel wide coastal buffer for the Landsat-based models and a 1-pixel wide buffer for the Bio-ORACLE models. Standard deviations of these environmental values were compared between the Gulf of Oman and the Arabian Sea as a measure of variability and the significance of the relation was tested using an F-test. For the categorical variable substrate, variability was measured as the difference of the relative fraction of the substrate classes (excluding deep water) from 0.5. A smaller difference was considered to indicate a higher variability (less dominance of a single class).

The same coastal buffer points were used to extract data values of the final models. Using the R statistical computing environment (R Development Core Team, 2010), habitat suitability was plotted against the distance along the coast (roughly a north-south transect). Next, the lowest significant threshold for converting the Maxent replicate model with the highest test AUC to a binary prediction was chosen to evaluate the relative amount of predicted suitable pixels in each of the regions. Thresholds are based on different rules such as balancing sensitivity and fractional predicted area, for which Maxent provides 1-sided p-values for the null hypothesis that test points are predicted no better than random with the same fractional predicted area.

## RESULTS

Although some models performed slightly better than others, high test-AUC values for all models indicated adequate model performance based on the selected variable sets (table 1). Country-wide distribution patterns which were known from previous sampling campaigns were largely reflected in the Maxent models, although important differences between Bio-ORACLE and Landsat-based models emerged. Areas with high suitability values for *N. zanardinii* were predicted exclusively in the Arabian Sea based on Bio-ORACLE variables as well as Landsat-based models. However, while Bio-ORACLE models predicted almost the entire Arabian Sea as highly suitable,

only discrete stretches of coast along the Arabian Sea were characterized by high suitability values based on Landsat variables, mostly in the south and around Masirah Island (figure 2 and 3A). For *T. glomerulata*, the most important highly predicted areas were found in the Gulf of Oman by both the Landsat and Bio-ORACLE models, but larger areas with high suitability values were computed in the Arabian Sea by the Bio-ORACLE model (figure 3B and appendix 2). While Bio-ORACLE models predicted generally high suitability throughout the Arabian Sea and low values in the Gulf of Oman for *H. discoidea*, the inverse was true for Landsat-based models, with the exception of a few (largely sandy) embayments in the Arabian Sea (figure 3C and appendix 3).

**Table 1.** The variables selected for the final models, their contributions to the final model, the resulting average test AUC values for the 5-fold cross-validation and the lowest significant threshold values based on the replicate model with the highest test AUC for both environmental datasets. Selected variables are given in the order of variable contribution to the model. SST (sea surface temperature), CHL (chlorophyll *a*), PAR (photosynthetically available radiation), DA (diffuse attenuation), B2 (Landsat band 2), CAL (calcite) and SUBSTR (substrate).

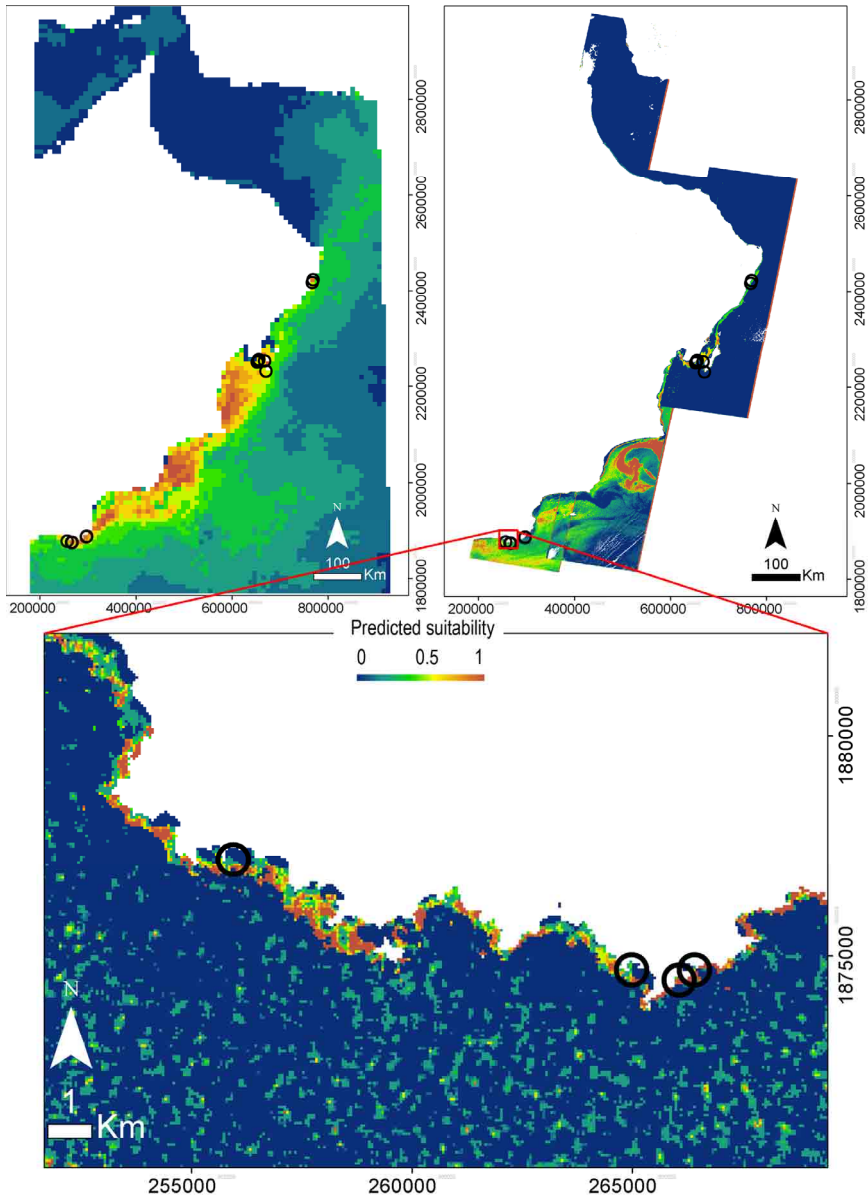
**a. Bio-ORACLE models**

Species	Selected variables	test AUC ( $\pm$ SD)	threshold
<i>T. glomerulata</i>	SSTavg, CHLmin (28.8%), DAmin (24%), CAL (19.9%), PARmax (6.7%)	0.874 (0.179)	0.063
<i>N. zanardinii</i>	SSTavg (33.8%), CHLmin (24.7%), SSTmin (21.3%), SSTmax (14.2%), CAL (5.9%)	0.965 (0.010)	0.17
<i>H. discoidea</i>	DAavg (81.9%), PARavg (9.1%), CAL (4.9%), CHLmin (4.2%), SSTavg (0.5%)	0.917 (0.095)	0.302

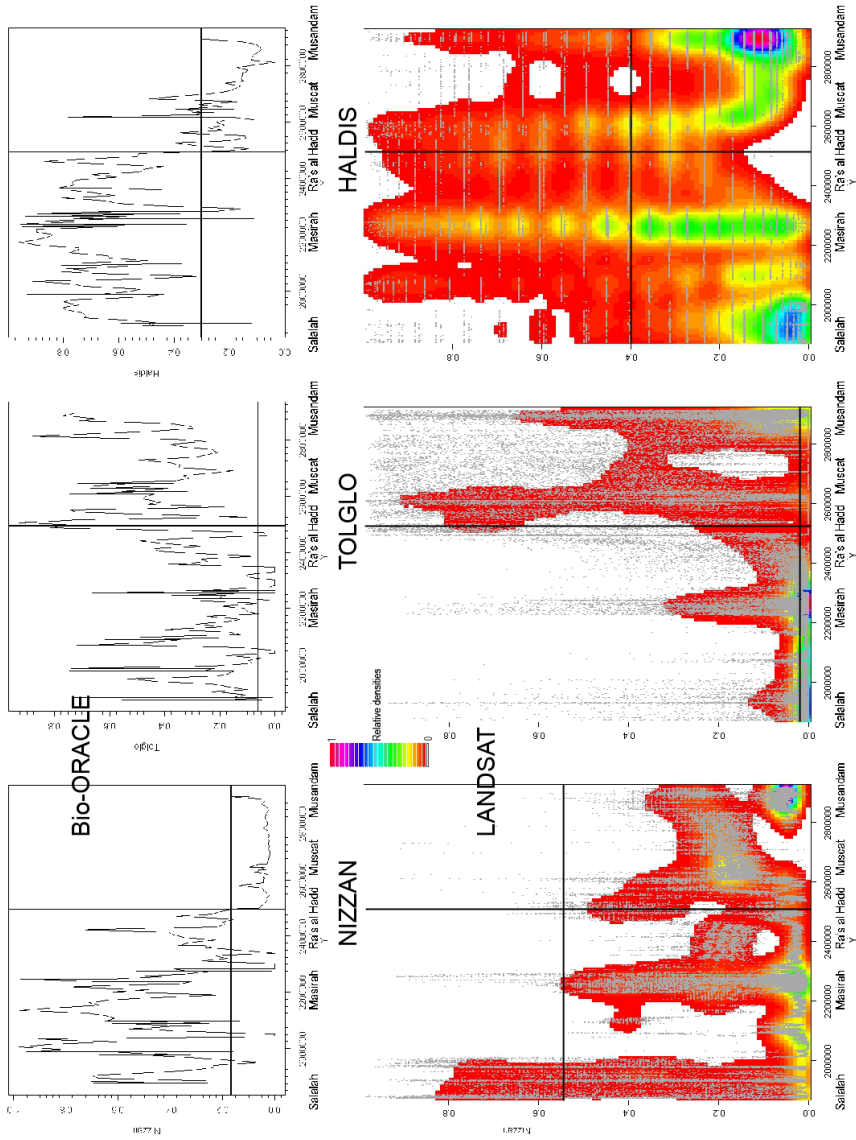
**b. Landsat-based models**

Species	Selected variables	test AUC ( $\pm$ SD)	threshold
<i>T. glomerulata</i>	SSTmin (22%), SSTmax (47.5%), CHLrng (12.4%), SUBSTR (10.4%), CHLmin (6.4%), B2avg (1.4%)	0.976 (0.024)	0.026
<i>N. zanardinii</i>	CHLmax (54.1%), SUBSTR (42.3%), SSTavg (3.6%)	0.867 (0.058)	0.534
<i>H. discoidea</i>	SSTmin (74.4%), SUBSTR (25.6%), CHLmax (0.5%)	0.953 (0.031)	0.402





**Figure 2.** Maxent logistic model output for *Nizamuddinia zanardinii* (occurrence records are indicated as pink dots in the upper maps, black circles in the lower magnified area) based on Bio-ORACLE (upper left) and Landsat-based environmental data (upper right).



**Figure 3.** Suitability values in the 1 pixel (Bio-ORACLE) or two pixels (Landsat) wide coastal buffer plotted against UTM northing. The approximate locations of coastal locality names as shown on figure 1 are plotted below the axis to aid orientation. Note that the number of data points from the Landsat models is too high to be plotted as separate points to the naked eye; hence, many points may be plotted on top of each other on discrete levels (depending on model input layers). Therefore, relative point densities are plotted as a colored background layer as a visual interpretation aid. Points plotted in white areas represent rare outliers (i.e., although there may seem many to the naked eye, they represent a very small fraction of the total number of points in the plot). Vertical lines represent the Arabian Sea / Gulf of Oman (right side) boundary at Ra's al Hadd. Horizontal lines indicate the lowest significant threshold value calculated by Maxent from which the continuous probability distribution could be converted to a binary (suitable/unsuitable) prediction.

The Landsat model for *N. zanardinii* shown in figure 2 clearly illustrated that on a fine-scale, habitat suitability values vary greatly between adjacent bays and even within areas with high overall suitability, important patches of low habitat suitability values are discerned along the coast. Figure 3 shows the same pattern where for a given high or low value in the Bio-ORACLE plot the corresponding Landsat pixels cover the whole range of suitability values.

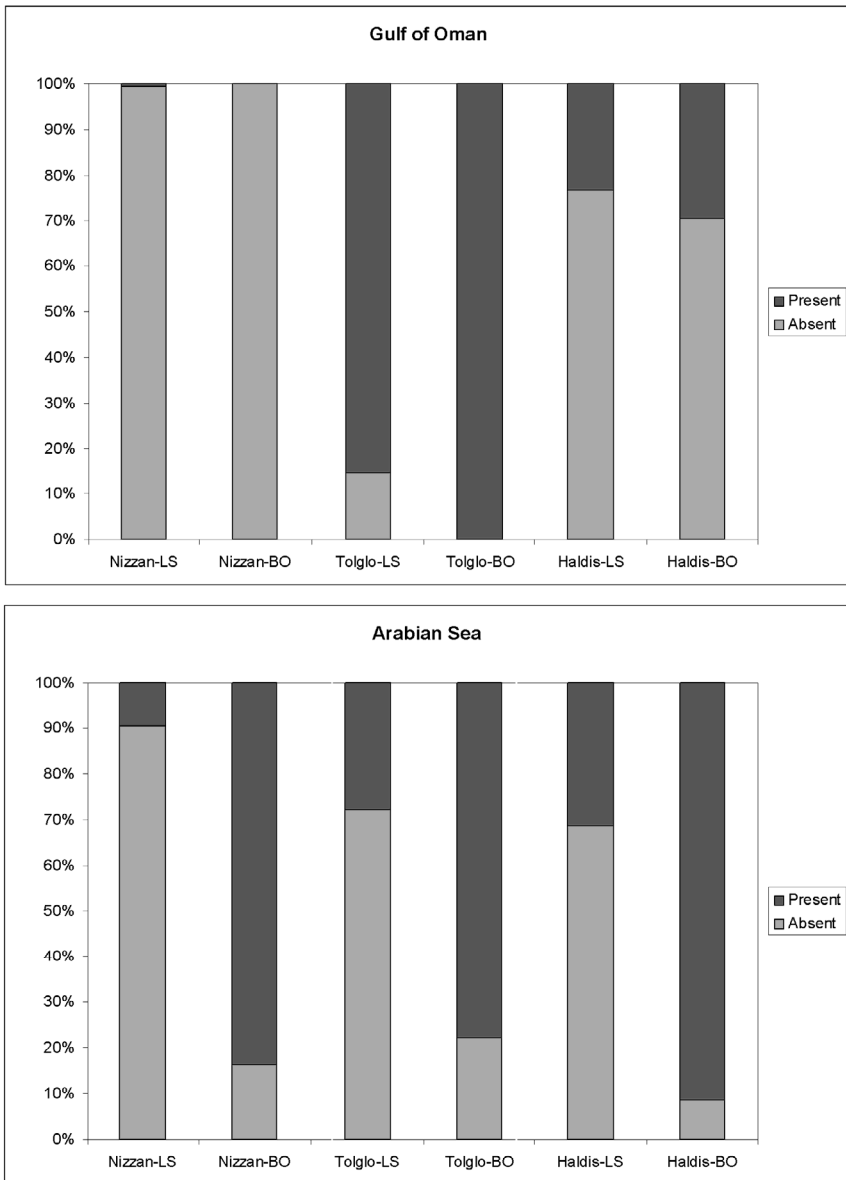
When translated to binary predictions (figure 4, with indicated threshold values plotted as horizontal lines in figure 3), previously known distributional patterns emerged more clearly for both environmental datasets, with *N. zanardinii* occurring only in the Arabian Sea, *T. glomerulata* mostly predicted in the Gulf of Oman and *H. discoidea* found on both sides of Ra's al Hadd. Although *N. zanardinii* is known to be absent from the Gulf of Oman, 0.64% of the Gulf of Oman coastal buffer pixels was erroneously predicted to be suitable for *N. zanardinii*. Furthermore, figure 3 clearly showed that while predictions based on both environmental datasets were in good agreement for the Gulf of Oman, a consistently higher portion of the pixels were predicted suitable in the Arabian Sea by the Bio-ORACLE models when compared to the Landsat-based models for the same species.

For the selected environmental variables, standard deviations of the environmental values encountered in the coastal buffers consistently indicated a higher variability in the marine environment of the Arabian Sea compared to the Gulf of Oman. This pattern was apparent from both Landsat-based and Bio-ORACLE variables (table 2). In 3 out of 9 Bio-ORACLE variables where the variability seemed higher in the Gulf of Oman, the difference was not significant. The higher environmental variability in the Arabian Sea corresponded well with the relatively higher predictions in the Arabian Sea by the coarser Bio-ORACLE models, while lower environmental variability yielded models more in mutual agreement based on environmental datasets of different spatial resolutions.

## **DISCUSSION**

### **EVALUATION OF LANDSAT-BASED SDM**

Recent years have seen a rising number of successful marine species distribution modeling efforts, facilitated by the increasing availability and ease of integration of regional and global environmental datasets.



**Figure 4.** Relative portions of suitable (present) and unsuitable (absent) pixels in the coastal buffers for each species predicted by the Bio-ORACLE (BO) and Landsat-based (LS) Maxent models for the Gulf of Oman and Arabian Sea.

Most marine studies have used remote sensing or interpolation-based datasets with a resolution of several hundred meters to kilometers (Robinson et al., 2011; Tyberghien et al., 2011), while side-scan sonar and

**Table 2.** Standard deviations (SD) of environmental variables in a 1 pixel (Bio-ORACLE) or 2 pixels (Landsat) wide coastal buffer for the Gulf of Oman and Arabian Sea. Regions with the highest variability are indicated in bold. Equivalent variables for both datasets are aligned. Asterisks indicate significant difference at the  $p=0.05$  level.

	Bio-ORACLE		Landsat 7 ETM+	
	SD(Gulf)	SD(ArabSea)	SD(Gulf)	SD(ArabSea)
SSTmin	0.414908	<b>0.930975*</b>	6.496189	<b>8.518251*</b>
SSTmax	<b>0.709888</b>	0.602955	7.407164	<b>9.360719*</b>
SSTavg	0.278297	<b>0.4153*</b>	6.179518	<b>7.534023*</b>
CHLmin	0.495206	<b>3.801434*</b>	4.81162	<b>7.901196*</b>
Chlmax	-	-	2.758753	<b>8.102843*</b>
Chlavg	-	-	3.520513	<b>6.410257*</b>
PARmax	<b>1.365203</b>	1.217916	-	-
PARavg	<b>1.125351</b>	0.948214	-	-
DAMin	0.027374	<b>0.085621*</b>	-	-
DAavg	0.02766	<b>0.074294*</b>	12.73504	<b>41.77098*</b>
CAL	0.000403	<b>0.009434*</b>	0.38418	<b>0.108657*</b>

interpolation of in situ measurements has enabled SDM on a sub-100m (Degraer et al., 2008) or sub-10m resolution (Monk et al., 2010; although in that case confined to less than 50 km<sup>2</sup>).

This study represents the first modeling effort in which sub-100m resolution spaceborne remote sensing-based marine habitat variables are shown to perform well for regional distribution modeling, especially compared to cropped global datasets at 9km resolution in terms of AUC values. However, it is important to note that the equal performance of models based on different resolutions means that equally relevant but different processes are captured by the environmental data, rather than that similar maps and variable contributions might be expected from these models. This is due to the finer resolution of Landsat data, allowing small bays to be mapped in which very local differences in chlorophyll and SST may be caused by gradients in depth and exposure to waves (mixing) and upwelling, affecting habitat suitability within bays. These are often factors governing local patchiness of benthic habitats, as opposed to coarse-scale gradients in coastal temperature and chlorophyll determining the spread of communities along the entire coastline. Furthermore, the resolution of

Landsat also enables the meaningful inclusion of a substrate layer. Substrate is known to play a key role in structuring habitat patchiness but almost always acts on a sub-km level. In the Landsat models for *N. zanardinii* and *H. discoidea*, substrate is the second highest contributing variable (table 1), favoring hard and soft substrates, respectively, in agreement with field observations. While not dominating overall variable contributions, substrate patchiness is reflected in fluctuating suitability values in otherwise macroecologically suitable areas, as demonstrated in figure 2; see also the high-resolution mapping study of this area by Pauly et al. (2011; Chapter 6 in this thesis). However, we also suspect that the categorical variable substrate was partly responsible for the erroneous suitable classification of the few coastal pixels in the Gulf coast of Oman for *N. zanardinii* (although a large portion of these erroneous suitability values are found in shallow and intertidal inland bays near Ra's al Hadd, where deviant environmental values might also have been recorded for the other variables). Additionally, very turbid waters such as encountered in eddies in the Arabian Sea also led to misclassifications in the substrate layer and hence erroneous suitability in models where substrate plays an important role. On the other hand, substrate contributed only 10% (4<sup>th</sup> out of 6 selected variables; table 1) to the Landsat model for *T. glomerulata*, a species known to occur on both substrates.

The above outlined differences between coarse-scale macroecological models and fine-scale habitat models are best illustrated by the models for *N. zanardinii*, an indicator species of Arabian Sea upwelling communities. While the Bio-ORACLE-based suitability map shows the sharp transition between the Arabian Sea and Gulf of Oman, with SSTavg as the most contributing variable (generally in agreement with Schils & Wilson (2006)), the Landsat-based suitability map shows a refined distribution pinpointing the exact boundaries of species turnover, governed mostly by nutrient availability (CHL, or exposure to upwelling) and substrate.

#### TECHNICAL ISSUES RELATING TO LANDSAT-BASED SDM

In order to compile a regional set of Landsat-derived marine habitat variables, some issues relating to data availability and computing resources were identified. First, several scenes per combination of path/row tile and

season should ideally be combined to reduce the influence of daily variability between path/row tiles and to minimize the chance of capturing extreme events. The availability of only one scene per combination of path/row tile and season introduced some visible effects such as deviant environmental values and the resulting habitat suitability in eddies (see for instance figure 2), clouds (from edge pixels that were not included in the cloud mask) and abrupt between-scene variations (both mostly visible in appendix 2 and 3). Some steps were taken to mitigate these effects as much as possible: (1) instead of using raw seasonal environmental values such as SST or chlorophyll index, derived layers such as minimum, maximum, mean and range based on the combined winter and summer mosaics were used as input in the models and (2) mosaics were created using an averaging algorithm for overlapping scenes, enabling a smoother transition between scenes. Between-scene variations due to different acquisition dates are mostly expected to occur in the Arabian Sea, as this area experiences a generally more variable environment during the summer monsoon. However, although the intensity of patterns might fluctuate widely from scene to scene, their relative effects are still discernible; see for instance the resulting continuing patterns across scenes of increased suitability offshore, located west of Masirah Island and Salalah in appendices 2 and 3, respectively, albeit in different relative values.

A second factor hampering easy integration of Landsat mosaics is the UTM projection in which scenes are delivered. The ten selected path/row tiles are located in the same UTM zone, which allows direct mosaicing. Inclusion of different zones would require the construction of a mosaic for each zone separately, followed by a reprojection of one mosaic to the other using tiepoints, greatly increasing processing time. Overall, the derivation of useful environmental layers for distribution modeling from Landsat imagery remains a time and resource-intensive process compared to the use of cropped global environmental datasets. Therefore, this paper represents a first step towards increasing the ease of integrating marine Landsat-based environmental information in regional modeling efforts.

It should be noted that the use of sub-100 m resolution data in general and a substrate layer in particular for distribution modeling necessitates biotic data that are georeferenced accurately to within half a pixel (in this case, 30m); a level of accuracy which is usually only achievable

with a surface-mounted GPS during sampling, rather than through post-hoc georeferencing of biotic occurrence records from collections.

## INFLUENCE OF SCALE ON MARINE SDM

Most studies investigating the influences of scale (usually by varying both resolution and extent) in the terrestrial realm have focused on the differences in model predictions for lowland versus mountainous habitats (Guisan et al., 2007). These studies generally reveal divergent results between fine-scale and coarse-scale models for the mountainous areas, characterized by a higher habitat variability, while showing similar results for lowland areas. More specifically, overpredictions of coarse-scale models applied mostly to mountainous species, while lowland species were similarly modeled (Trivedi et al., 2008). Especially the comparison of genuine high-resolution remote sensing datasets with coarse-scale climatic data has revealed significant differences between fine-scale and coarse-scale models. This opposes studies starting from one environmental dataset with subsequent up- or downscaling to investigate scale influences on distribution modeling (Trivedi et al., 2008). The present study, enabled by purpose-built sub-100m resolution habitat variables based on Landsat mosaics covering a 2000km coastal stretch, is the first marine application to demonstrate important effects of resolution on modeling efforts, caused by differences in habitat variability. Both Bio-ORACLE and Landsat data from the coastal buffer clearly confirmed that the Arabian Sea is characterized by a higher environmental variability, as described earlier by Wilson (2000). It emerged that, while predictions are similar for coarse and fine resolution in the less variable Gulf of Oman, the coarse resolution environmental data caused a general overprediction of all species in this study for the more variable Arabian Sea, regardless of the species' known preferred occurrence in the Gulf of Oman, the Arabian Sea or both. The effect of overprediction by coarse-scale macroecological niche models in highly variable environments appears to be similar to observations in terrestrial studies (Trivedi et al., 2008). In contrast, however, species characteristic to either less or more variable marine habitats (or both) seem to be affected in the same way by overprediction in variable habitats by coarse resolution modeling. The most likely cause for this difference is that environmental



data in marine coarse-scale models are not consistently biased towards higher or lower values. Data aggregation in coarse grid cells over mountainous terrain tends to exaggerate lowland environments due to projected surface areas, hence yielding better results for lowland species models. This is in contrast with environmental data of marine systems where data layers represent surface conditions or integrated water column data. With variability found only in one dimension, data aggregation into coarse pixels doesn't systematically favor certain values and all species are equally affected in modeling efforts.

Partly due to a lack of similar regional marine modeling studies investigating scale effects, it is yet unknown if these effects are a ubiquitous trend. However, care should be taken to select appropriate modeling scales when using distribution modeling in the marine environment, especially with studies aimed at guiding fieldwork, where models trained on a certain environment and projected on a more variable environment may result in overprediction. With Landsat imagery now more widely and freely available, we encourage further multi-scale modeling research.

**SUPPLEMENTARY MATERIAL**

## APPENDIX 1.

TIR digital numbers were converted to surface temperature in three steps (Chander, 2009): first, at-satellite radiances ( $L_{sat}$ ) were calculated using the standard NASA equation correcting for known calibration constants:

$$L_{sat} = \frac{L_{min} + (L_{max} - L_{min}) \cdot DN}{DN_{max}} \quad (1)$$

Second, surface-leaving radiances ( $L_T$ ) were inferred by correcting (1) for atmospheric interference:

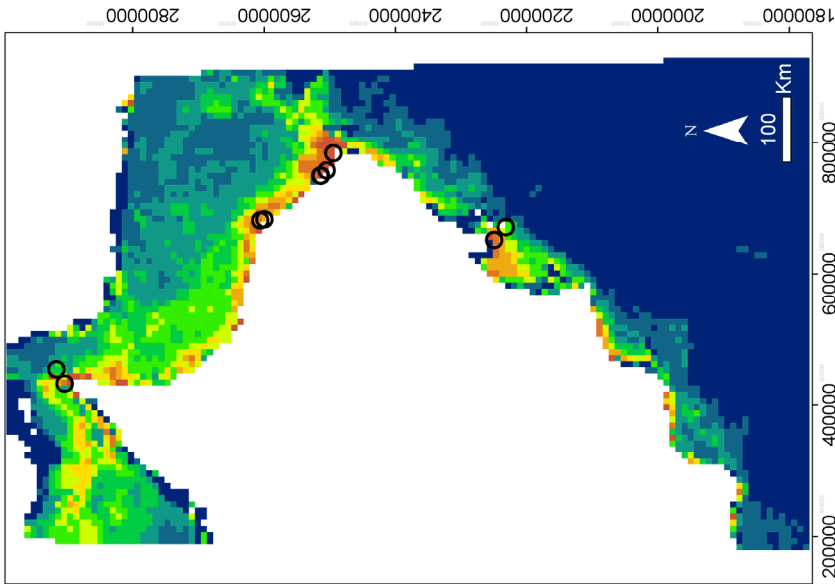
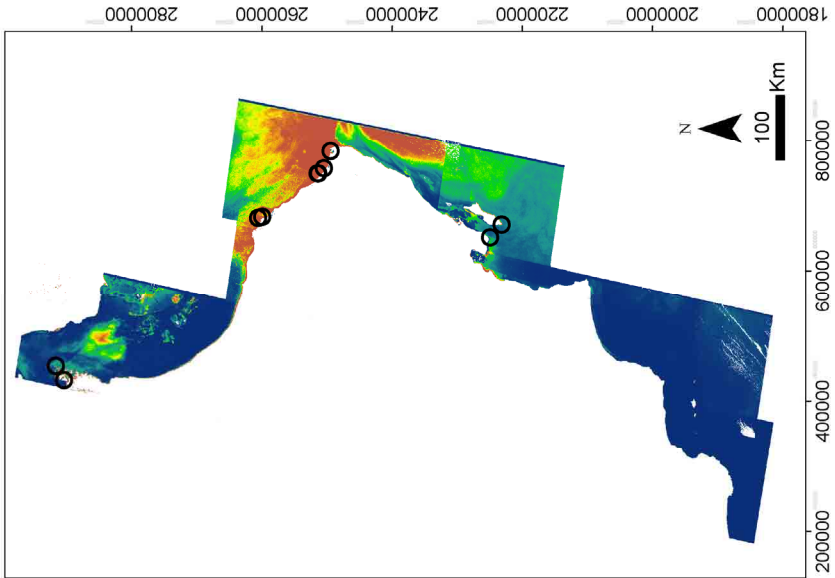
$$L_T = \frac{L_{sat} - L_u - (1 - \epsilon) \cdot \tau \cdot L_d}{\epsilon \cdot \tau} \quad (2)$$

where the emissivity ( $\epsilon$ ) for water was assumed to be constant at 0.98. The up- and downwelling radiance ( $L_u$  and  $L_d$ , respectively) and atmospheric transmission ( $\tau$ ) were calculated using NCEP atmospheric profiles and the MODTRAN radiative transfer code implemented in the online Atmospheric Correction Parametre Calculator freely available at <http://atmcorr.gsfc.nasa.gov/>. Next, sea surface temperature  $T$  ( $^{\circ}\text{C}$ ) was computed using the calibration coefficients and conversion formula given by NASA:

$$T = \frac{K_2}{\ln\left(\frac{K_1 \cdot \epsilon}{L_T} + 1\right)} - 273.15 \quad (3)$$

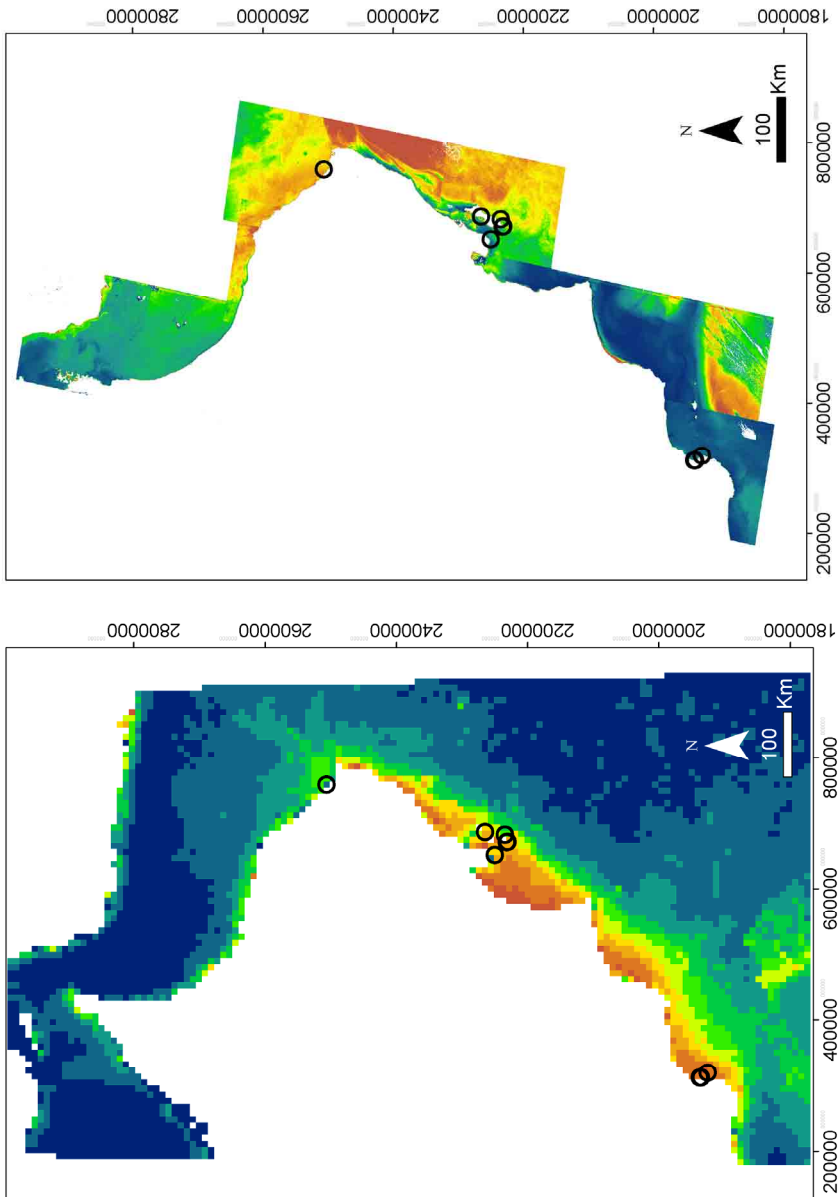
where  $K_2 = 1282.71$  and  $K_1 = 666.09$ .

APPENDIX 2.



Maxent logistic model output for *Toxoptrocladia glomerulata* based on Bio-ORACLE (lower) en Landsat-based environmental data (upper).

APPENDIX 3.



Maxent logistic model output for *Halimeda discoidea* based on Bio-ORACLE (upper) en Landsat-based environmental data (lower).



# CHAPTER 6

## **MAPPING CORAL-ALGAL DYNAMICS IN A SEASONAL UPWELLING AREA USING SPACEBORNE HIGH RESOLUTION SENSORS**

Klaas Pauly, Rudi Goossens<sup>1</sup> & Olivier De Clerck

<sup>1</sup> 3D Data Acquisition Cluster  
Geography Department, Ghent University  
Krijgslaan 281/S8, 9000 Ghent, Belgium

Adapter from published proceedings article: Pauly K, Goossens R & De Clerck O (2011) Mapping coral-algal dynamics in a seasonal upwelling area using spaceborne high resolution sensors. Proceedings of the ESA Living Planet Symposium, ESA SPS-686, 8pp.

## ABSTRACT

PROBA/CHRIS is one of the first satellite sensors to offer both high spatial and spectral resolutions. We explored the potential of this sensor to map the dynamics of seaweed and coral cover in an area influenced by seasonal upwelling in the Arabian Sea. Quantitative field assessments coincided with image acquisitions. After removal of sensor noise and atmospheric effects, maximum likelihood supervised classification yielded a *tau* accuracy of 64.09 for the summer monsoon dataset. Clearer waters and a lower spatial heterogeneity in the winter monsoon dataset resulted in a *tau* accuracy of 71.45. Post-classification comparison and vegetation indices illustrated the conspicuous turnover from dense macroalgal stands covering nearly all coral communities during summer to bare rock or turf communities during winter, with coral becoming the predominant bottom type. These results were further analyzed using a novel maximum entropy sub-pixel approach and were shown to consistently outperform results from Landsat 7 ETM+ imagery.

## INTRODUCTION

Competition between hermatypic corals and benthic macroalgae is widely believed to be a key process determining the composition and structural changes of coral reefs (McCook et al., 2001). On tropical coral reefs, a suite of interaction mechanisms between coral and algae can lead to a sharp decline in healthy coral cover when algal production exceeds a critical level, induced by (human-caused) stress, starting a positive feedback loop (Mumby et al., 2007; Smith et al. 2006). Such interactions and eventually resulting phase shifts have mostly been studied in view of natural or human-induced disturbances, but ephemeral or seasonally recurring macroalgal dominances on corals have remained largely understudied until recent years (Vroom et al., 2006; Haas et al., 2009). However, the study of such naturally changing but healthy systems is vital to understand baseline processes affecting tropical and subtropical marine benthic communities and to assess the effectiveness of monitoring techniques to capture such processes.

This study focuses on the seasonal dynamics of benthic communities along the Arabian Sea coast of the Dhofar region (southern Oman), in an area seasonally influenced by one of the world's five strongest upwelling systems. Although very limited reef growth is found in this area, coral communities typically account for on average 20% of the benthic cover, locally up to 99% (Coles, 1996). During the southwest monsoon in July-August, warm oligotrophic surface waters are displaced by cold nutrient-rich waters, leading to a boost in algal production and diversity with up to 100% coverage in the intertidal and shallow subtidal while coral communities become overshadowed or overgrown by seaweed canopies (Ormond & Banaimoon, 1994; Schils & Coppejans, 2003). One of the most productive seaweeds during summer monsoon is the regional endemic meadow-forming brown alga *Nizamuddinia zanardinii* (Schiffner) P.C. Silva. With temperatures increasing and nutrient levels dropping sharply by the end of September, the algae die back and accumulate in shallow decomposing packs in sheltered bays. Together with *Nizamuddinia* standing stock, these decomposing algae serve as major shelter and food sources for the Green turtle (*Chelonia mydas*) population in the region (Ferreira et al., 2006). During the rest of the year, the intertidal and shallow subtidal are devoid of algal growth or support only turf and small macroalgal assemblages.



Seasonal monitoring efforts are traditionally carried out using *in situ* quantitative assessments of fixed transects, which are very time and resource consuming for larger and remote areas. By contrast, satellite remote sensing has proven a valuable tool in repeatedly mapping coastal communities on a larger scale and over a wider range of variables than feasible by field research only. Although spaceborne remote sensing has been successfully used for monitoring coral reefs and seagrass beds in the last decade, macroalgae have often been overlooked or lumped into a single benthic class in many of these studies. This is mostly due to a lack of the combination of high spatial and spectral discriminating power of current satellite sensors, needed to cope with typically heterogeneous algae-dominated assemblages resulting in spectral mixing. For instance, according to Hochberg & Atkinson (2003), narrowband multispectral (hereafter called superspectral) spaceborne sensors are better suited to discern coral from algae than multispectral satellite sensors, even if the latter can have much finer spatial resolutions.

Launched in 2001, the Compact High-Resolution Imaging Spectrometer (CHRIS) onboard the PROject for On-Board Autonomy (PROBA) satellite offers up to 63 programmable bands of 6 to 33 nm width at about 17 or 34m spatial resolution (depending on the band configuration), with a 14 km swath. This sensor is therefore thought to be better suited for monitoring coral-algal dynamics than commonly used satellites.

The aim of this study was to map seasonal coral-algal dynamics in the Arabian Sea, assessing the potential of the PROBA/CHRIS sensor to discern coral and several benthic macroalgal bottom types by means of supervised classification, and results were compared to those from the multispectral high resolution Landsat 7 ETM+ sensor. Additionally, several spectral indices were used to visualize seasonal differences in intertidal and surfacing algal growth. Lastly, a maximum entropy approach was introduced to investigate sub-pixel class probabilities in order to produce abundance maps based on limited input data.

## MATERIALS AND METHODS

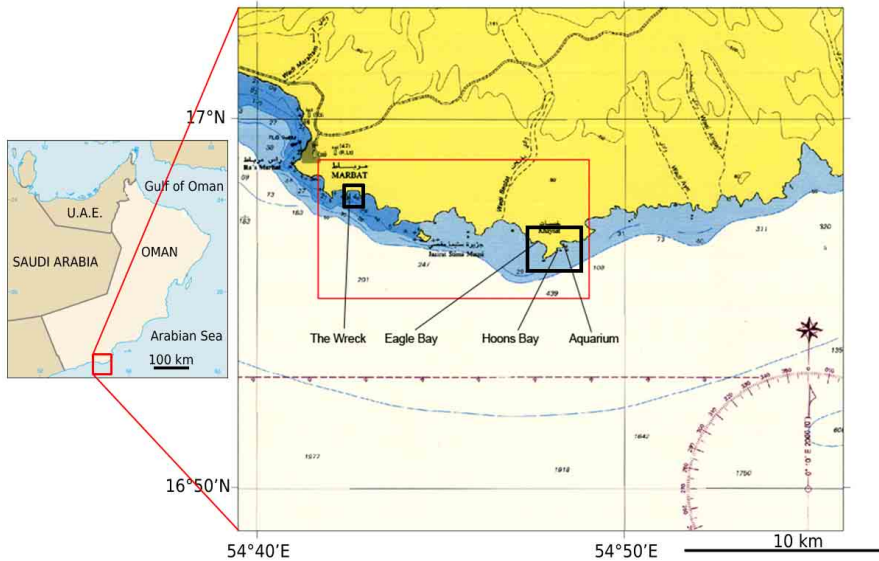
### STUDY AREA AND FIELD WORK

The study area comprised a series of rocky bays stretching out 14 km east of Mirbat, a town about 64 km east of Salalah, Dhofar province, Sultanate of Oman (figure 1). Four of these bays have been extensively studied using GPS-located transects and quadrats on a near-seasonal basis from 2003 until 2006. Field work for this study coincided with PROBA/CHRIS acquisitions in late September 2005 and late March 2006, representing summer and winter monsoon conditions, respectively. During each campaign, three 0.25 m<sup>2</sup> quadrats spaced at least 30 m apart were exhaustively assessed for algal and coral cover percentages and fresh weights of macroalgae on the species level at three different depths (intertidal,  $\pm 5$  m and  $\pm 10$  m) in each of the 4 bays. Additionally, coral and algal cover percentages were assessed on the genus level along three 15 m long transects perpendicular to the coast at  $\pm 5$  m depth, separated by at least 30 m, in each of the 4 bays. Finally, general community-level cover percentages for the four entire study bays were obtained using GPS-tracked video surveys of on average 2.5 km for which still frames were analyzed each 30 seconds (roughly 15-30 m). Ground control points (GCPs) for georeferencing were acquired by tracking rocky shorelines of the study bays. GPS locations were logged using a buoy-mounted Garmin GPSMAP 76CSx, ensuring the continuous reception of at least 6 satellite signals to keep the accuracy below half a pixel on PROBA/CHRIS imagery.

### REMOTE SENSING DATASETS

Two cloud-free PROBA/CHRIS datasets were acquired on 27 September 2005 and 26 March 2006 at mid tide to low tide in mode 2 (water bands) configuration. These datasets consisted of 18 bands on average 11.3 nm wide in the VNIR spectrum with a ground sampling distance (GSD) of 17 m at 556 km altitude. Only nadir imagery was used for analyses.

Additionally, two cloud-free Landsat 7 ETM+ SLC-on scenes from 19 October 2000 (summer monsoon) and 3 April 2003 (winter monsoon) were downloaded from the USGS EROS archive, consisting of 4 bands on average 82.2 nm wide in the VNIR and two additional on average 250 nm wide bands in the SWIR spectrum with a GSD of 30 m.



**Figure 1.** Situation of the study area with indication of the 4 study bays and the PROBA/CHRIS region of interest (ROI; red box).

The image acquisition time of the Landsat scenes coincided with mid to high tide.

## IMAGE PREPROCESSING

### *PROBA/CHRIS*

Given the experimental character of the superspectral push-broom CHRIS sensor, raw imagery suffered heavily from noise in the form of drop-out pixels in some image rows and vertical striping across all bands. Because these effects are enhanced in latter processing stages, the imagery was corrected for these anomalies in a first step using an algorithm described in detail by Gomez-Chova et al. (2008). Second, top-of-atmosphere (TOA) radiance values expressed in 16-bit digital numbers (DN) were converted to surface reflectance images using an automated image-based algorithm described by Guanter et al. (2005) & Guanter (2006). Noise removal and atmospheric correction were carried out using the CHRIS/Proba tools at default settings implemented in BEAM VISAT 4.7.1 (Brockmann Consult). Due to the atmospheric correction, saturated pixels over extensive white desert sand areas were masked out, which had no effect on later processing steps in the marine environment. However, it appeared that over 95% of the

sea area had zero reflectance values in band 1 of both summer and winter datasets, which were therefore excluded from further processing. Imagery was subsequently cropped to the coastline along the entire swath and exported to ITC ILWIS 3.7 for further processing. Georeferencing was accomplished using 23 ground control points along the entire coastline and a first-order transformation, resulting in a RMS error of 0.29 pixels for the summer monsoon dataset and 0.38 pixels for the winter dataset at 17 m and 19 m GSD, respectively (resulting from different acquisition altitudes). Finally, a landmask was applied based on NIR thresholding and manual editing. An image-based bathymetric correction of these superspectral datasets was attempted based on Lyzenga's method (Tassan, 1996), but was abandoned because of erroneous results.

### ***Landsat 7 ETM+***

Landsat scenes from the USGS EROS archive were downloaded at Level 1T (terrain corrected), meaning systematic and radiometric accuracy is derived from ground control points and DEM incorporation. The imagery was first converted from 8-bit DN to radiance values following Chander et al. (2009). Next, atmospheric correction was done using the  $\cos(t)$  model developed by Chavez (1996), accounting for haze removal by dark object subtraction while compensating for variations in solar output according to the acquisition time, and estimating atmospheric absorption and Rayleigh scattering based on the cosine of the solar zenith angle. This was accomplished using the ATMOSC module in Clark Labs IDRISI 15.0 (Andes). Subsequently, surface reflectance images were cropped to the study area and exported to ITC ILWIS 3.7 for further processing. Attempts to pansharpener the multispectral bands using the 15 m resolution panspectral band were abandoned because the stochastic noise in water areas in the panchromatic band degraded the image quality of pansharpener bands. Since errors in the Level 1T georeferencing were often significant and scenes did not accurately match GCPs measured in the field, Landsat imagery was master-slave referenced to the respective PROBA/CHRIS datasets using the respective GCPs as tie-points, resulting in RMS errors of 0.31 and 0.27 for summer and winter scenes. Lastly, a landmask was applied based on SWIR thresholding.

## IMAGE PROCESSING

***Supervised bottom-type classification***

Based on the quantitative sampling results, a hierarchical classification scheme was adopted for supervised classification of PROBA/CHRIS and Landsat 7 ETM+ datasets (table 1), in which pixels were assigned the class of the predominant bottom type in terms of percent cover. GPS-located quadrats and transects for quantitative analyses and the GPS-tracked video transects were digitized to a point file from which contiguous training pixels were added in a flood polygon within an empirical spectral tolerance. Half of the resulting pixels were used for training, while the other half was used as a post-classification test set. A specific deep-water mask was not applied; instead, deep water was included as a bottom type in the classification based on arbitrary training sites located at least 1 km off-shore. Additionally, since breaking waves are spectrally characterized by high reflectance values but could not be removed in preprocessing, whitecaps were included as a bottom type in the classification. To account for water column effects, bottom types that were encountered across the entire field study depth range were assigned to three different classes to be merged post-classification. Additionally, NIR band information was not included for classification of uniquely subtidal bottom types.

**Table 1.** General bottom type classification scheme.

Bare substrate	Intertidal sand	S1/W1
	Subtidal sand	S1/W2
	Intertidal bare rock	W3
	Subtidal bare rock	W3
Biotic communities		
- algae-dominated	Subtidal rocks with turf	W4
	Intertidal red algae	S4
	Shallow brown algae	S5
	Intertidal green algae	S6
	Subtidal mixed algal stands	S7/W5
	Accumulated dead algae	S8
- coral-dominated	Coral	S9/W6
Whitecaps		S2
Deep water		S3/W7

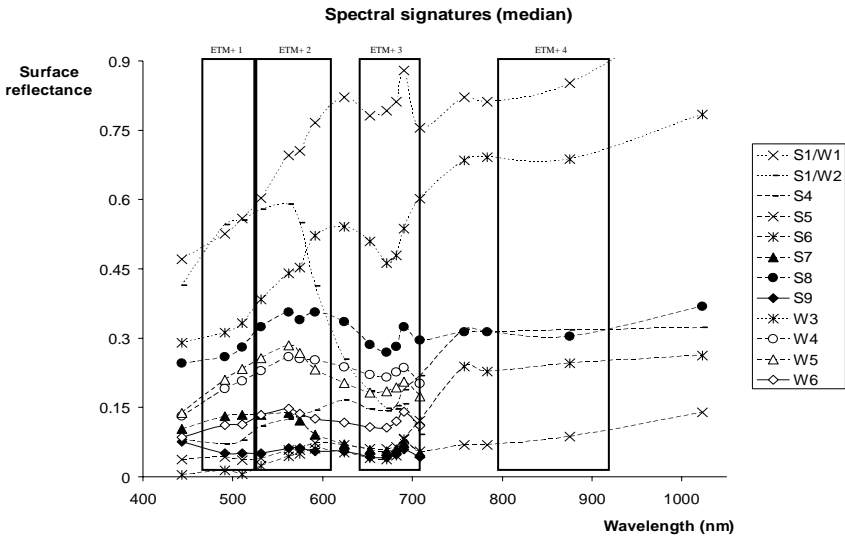
(S) = only used in summer monsoon imagery; (W) = only used in winter monsoon imagery.

Following Mumby & Edwards (2002), a maximum likelihood classifier was chosen as the most appropriate algorithm, which was implemented without assigning a threshold value. For an objective measure of classification accuracy incorporating the number of misclassifications, the *tau* coefficient was calculated besides the overall accuracy (Mumby & Edwards, 2002). The bottom type maps of different seasons based on CHRIS bands were cross-combined to produce a change map.

To assess whether different classification results between Landsat 7 ETM+ and PROBA/CHRIS datasets were mainly attributable to the difference in spectral or spatial resolution, the classification was also performed on simulated Landsat bands based on a weighted-averaging aggregation of CHRIS bands following Jarecke et al. (2001).

### *Spectral indices*

Two kinds of spectral index maps were generated to explore intertidal and surfacing vegetation density based on both PROBA/CHRIS and Landsat datasets across seasons.



**Figure 2.** Spectral signatures of the summer and winter bottom types based on PROBA/CHRIS data. Dotted lines represent abiotic bottom types; dashed lines represent algal bottom types and full lines represent coral. Spectra for exclusively subtidal bottom types exclude NIR bands. The approximate spectral coverage of Landsat 7 ETM+ bands is indicated by box outlines.

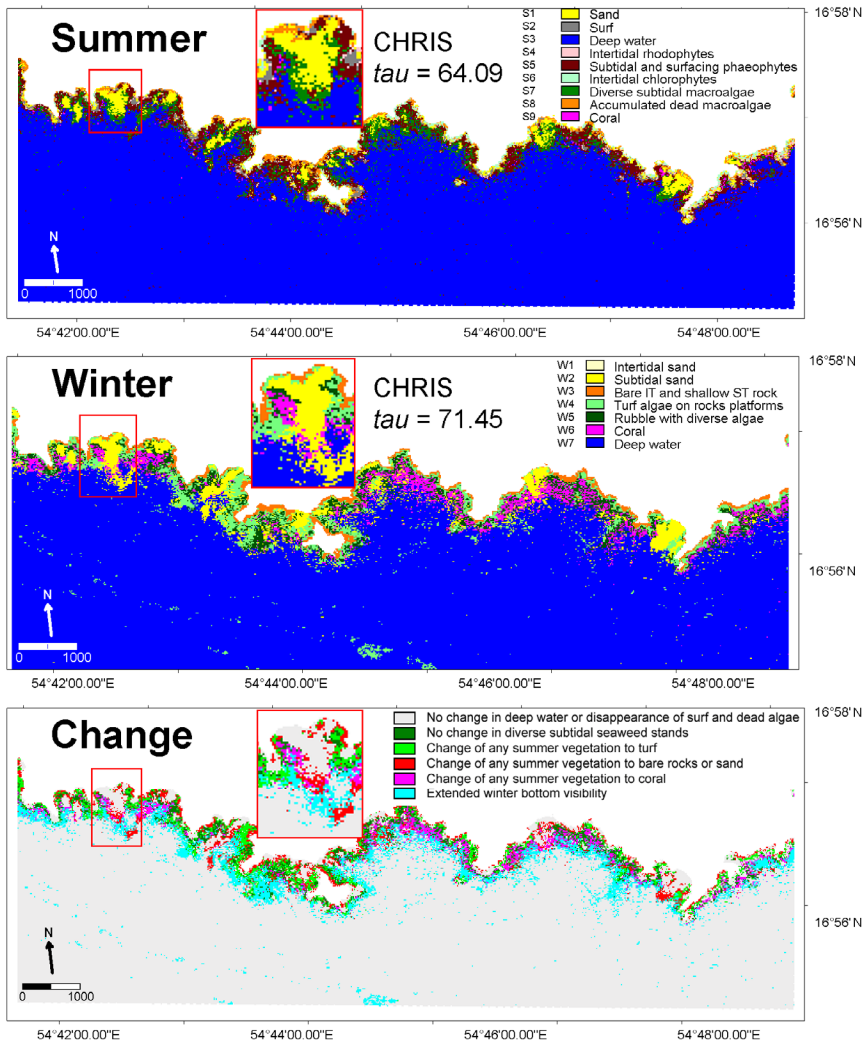
For the NDVI, the most widely applied index, CHRIS bands 11 and 16 were chosen for red and NIR reflectance, respectively, based on the difference in the spectral response for intertidal vegetation (figure 2.). Second, the floating algae index (FAI), recently developed on MODIS datasets and applied to Landsat scenes by Hu (2009), was used to detect both surfacing and intertidal algae while eliminating anomalous values over open water caused by illumination effects as often seen in the NDVI. The FAI is defined by

$$FAI = R_{NIR} - (R_R + (R_{SWIR} - R_R) \times \frac{\lambda_{NIR} - \lambda_R}{\lambda_{SWIR} - \lambda_R})$$

where  $R$  stands for the Rayleigh-corrected reflectance images. To approximate the MODIS spectral response, Landsat band 5 and CHRIS band 18 (1019 nm) were designated for SWIR. Furthermore, CHRIS bands 10 and 17 were used to represent  $R$  and NIR reflectances, respectively. Correlation analyses were performed for both FAI and NDVI with respect to total macroalgal fresh weight at the 9 intertidal quadrats during summer monsoon, tested for significance with a non-directional t-test.

### ***Maxent sub-pixel modeling***

To investigate the feasibility of modeling the extent of coral communities covered by algae, resulting in mixed pixel spectra often dominated by algal characteristics, a maximum entropy approach was taken to model separate pixel probabilities. The Maxent algorithm as implemented by Philips & Dudik (2008) was developed for species' distribution modeling based on known occurrence records (presence-only) and environmental raster data, and although hitherto not reported as a sub-pixel classifier to model bottom type probabilities across an image, the underlying reasoning is identical. The maximum entropy algorithm is a general-purpose machine-learning method to infer a probability distribution from incomplete information. This probability distribution is most uniformly spread out (showing the highest degree of entropy) given the constraints represented by the spectral data observed at the given training pixels. The resulting probability distribution is then translated from spectral space into geographical space.



**Figure 3.** ML classification results based on PROBA/CHRIS data for the summer and winter monsoon. The area surrounded by the red boxes is shown magnified to the right. A change map is shown to the left.

In the so-called logistic format, the output map shows pixel values ranging from 0 to 1, indicating bottom type probability. Five replicate Maxent runs with subsampling of a random 20% of the data for testing were carried out for the coral and shallow brown algae bottom types based on the PROBA/CHRIS summer monsoon dataset. Accuracy was assessed using the threshold-independent area under the receiver operating characteristics (ROC) curve (AUC). Ranging from 0 to 1, any value above 0.5 means better



prediction than random. Additionally, correlation analyses were performed for probability values with respect to fresh weight at the subtidal quadrats during summer monsoon, tested for significance with a non-directional t-test.

## RESULTS AND DISCUSSION

### SUPERVISED BOTTOM-TYPE CLASSIFICATION

Overall, the image classification based on PROBA/CHRIS datasets performed well (figure 3), especially for the winter monsoon where case I waters facilitated the distinction between the more homogenous bottom types ( $\tau = 71.45$ ; table 2a). The summer monsoon classification suffered more from turbid waters. Moreover, the bottom types coral, brown algae and diverse seaweeds were prone to misclassifications because of mixed assemblages and spectral resemblance. However, the overall classification accuracy was still reasonable ( $\tau = 64.09$ ; table 2b). The difference between the latter two classifications was very significant ( $Z = 3.24$ ), and was mainly the result of 101 Ha of dense summer seaweed communities changing into bare substrates and another 101 Ha being reduced to turf communities, while an additional 79 Ha of coral communities became apparent due to the die-off from overshadowing seaweeds from summer to winter. Thus, 33% of the bottom types excluding deep water and whitecaps suffered either a complete loss of seaweeds, a reduction to turf communities or a dominance shift towards coral communities from summer to winter. Also, it was apparent that an extended part of the seafloor could be mapped in the winter monsoon dataset due to clearer waters (figure 3).

By contrast, image classifications based on Landsat 7 ETM+ data (figure 4) performed poor for the given classification scheme for both summer ( $\tau = 21.51$ ) and winter ( $\tau = 33.12$ ) datasets. While not surprising for intertidal bottom types given the mid to high tidal conditions, this resulted foremost from an overestimation of coral where this spectrum was close to deep water, brown algae and diverse subtidal seaweed stands (figure 2). This overestimation of coral and underestimation of algae by ETM+ was in line with the results reported by Hochberg & Atkinson (2003).

The image classification based on simulated Landsat bands for the winter monsoon ( $\tau = 47.58$ ) differed remarkably less from the

PROBA/CHRIS winter classification than the original Landsat winter classification did (figure 4). The higher spatial resolution in the simulated Landsat bands likely allowed for a better classification result than the original Landsat bands could yield with the same multispectral resolution. This was in contrast with the summer monsoon classifications, where the simulated Landsat bands performed equally low ( $\tau = 28.34$ ) compared to the CHRIS bands as the original Landsat bands (figure 4). This was probably because the higher spatial resolution of simulated Landsat bands was not sufficient to cope with spectral information from mixed pixels consisting of coral and algal spectra in the broad multispectral bands.

**Table 2a.** ML classification error matrix based on PROBA/CHRIS data for the winter monsoon.

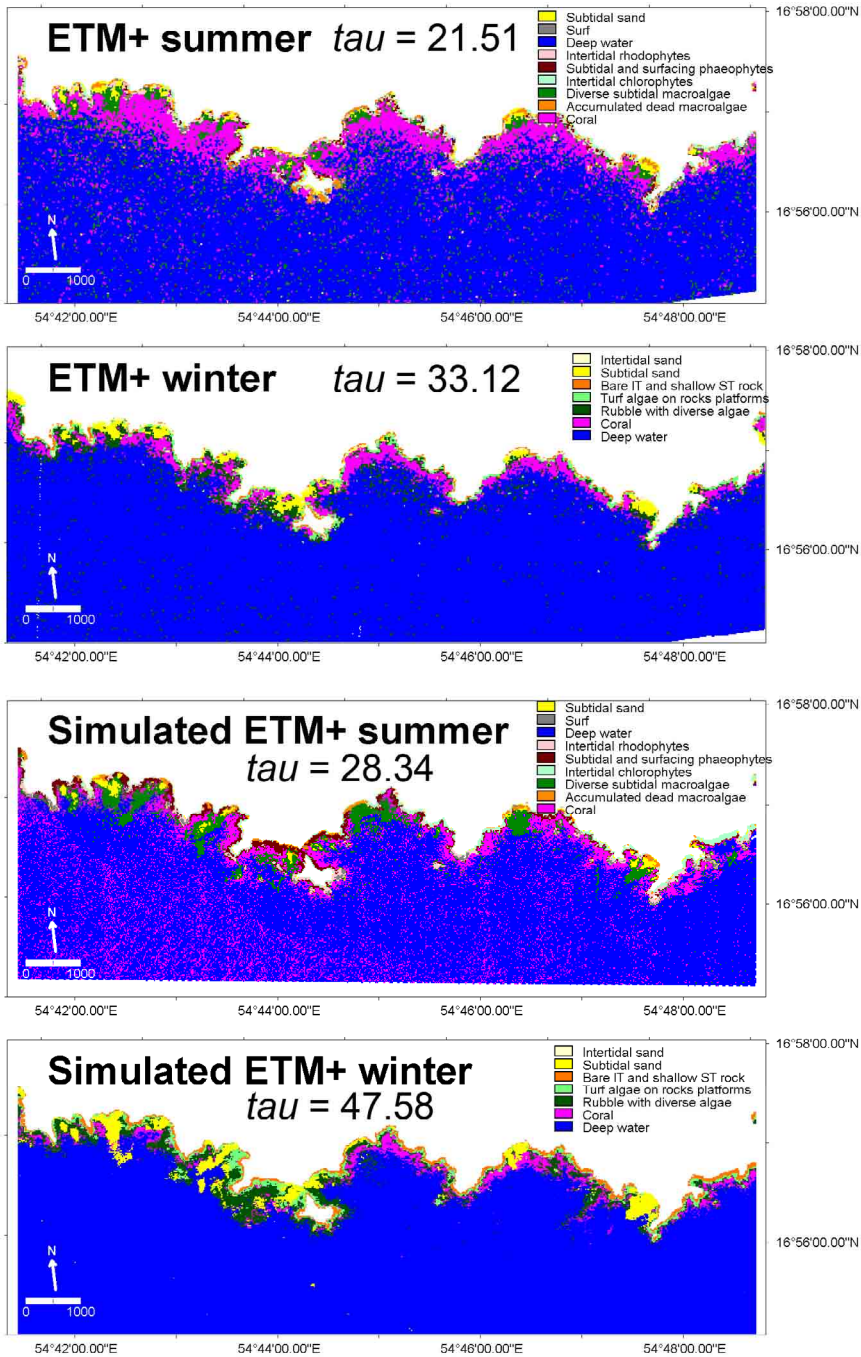
Class \ True	W1	W2	W3	W4	W5	W6	W7	total	User acc.
W1	46	4	8	0	0	0	0	58	0.79
W2	2	94	6	12	14	0	2	130	0.72
W3	6	0	82	20	0	0	0	108	0.76
W4	0	4	20	86	10	0	0	120	0.72
W5	0	14	0	0	86	12	3	115	0.75
W6	0	0	0	0	6	78	15	99	0.79
W7	0	22	0	0	20	26	218	286	0.76
total	54	138	116	118	136	116	238	916	75.57
Prod. acc.	0.85	0.68	0.71	0.73	0.63	0.67	0.92	74.14	75.33

Producer accuracy = 74.14; User accuracy = 75.57      Overall accuracy = 75.33; Tau-coefficient = 71.45

**Table 2b.** ML classification error matrix based on PROBA/CHRIS data for the summer monsoon.

Class \ True	S1	S1	S2	S3	S4	S5	S6	S7	S8	S9	total	User acc.
S1	34	0	0	0	0	0	0	0	4	0	38	0.89
S1	0	50	0	0	0	0	0	10	6	0	66	0.76
S2	0	0	40	0	10	0	8	0	0	0	58	0.69
S3	0	0	0	173	0	0	0	16	0	0	189	0.92
S4	0	0	14	0	38	20	8	0	0	0	80	0.48
S4	0	0	0	7	12	66	0	8	2	6	101	0.65
S6	0	0	4	0	6	4	40	0	4	0	58	0.69
S7	0	12	0	20	0	6	0	62	4	18	122	0.51
S8	6	10	0	0	9	0	8	10	42	0	85	0.49
S9	0	0	0	1	0	10	0	18	0	32	61	0.52
total	40	72	58	201	75	106	64	124	62	56	858	66.02
Prod. acc.	0.85	0.69	0.69	0.86	0.51	0.62	0.63	0.50	0.68	0.57	65.98	67.25

Producer accuracy = 65.98; User accuracy = 66.02      Overall accuracy = 67.25; Tau-coefficient = 64.09



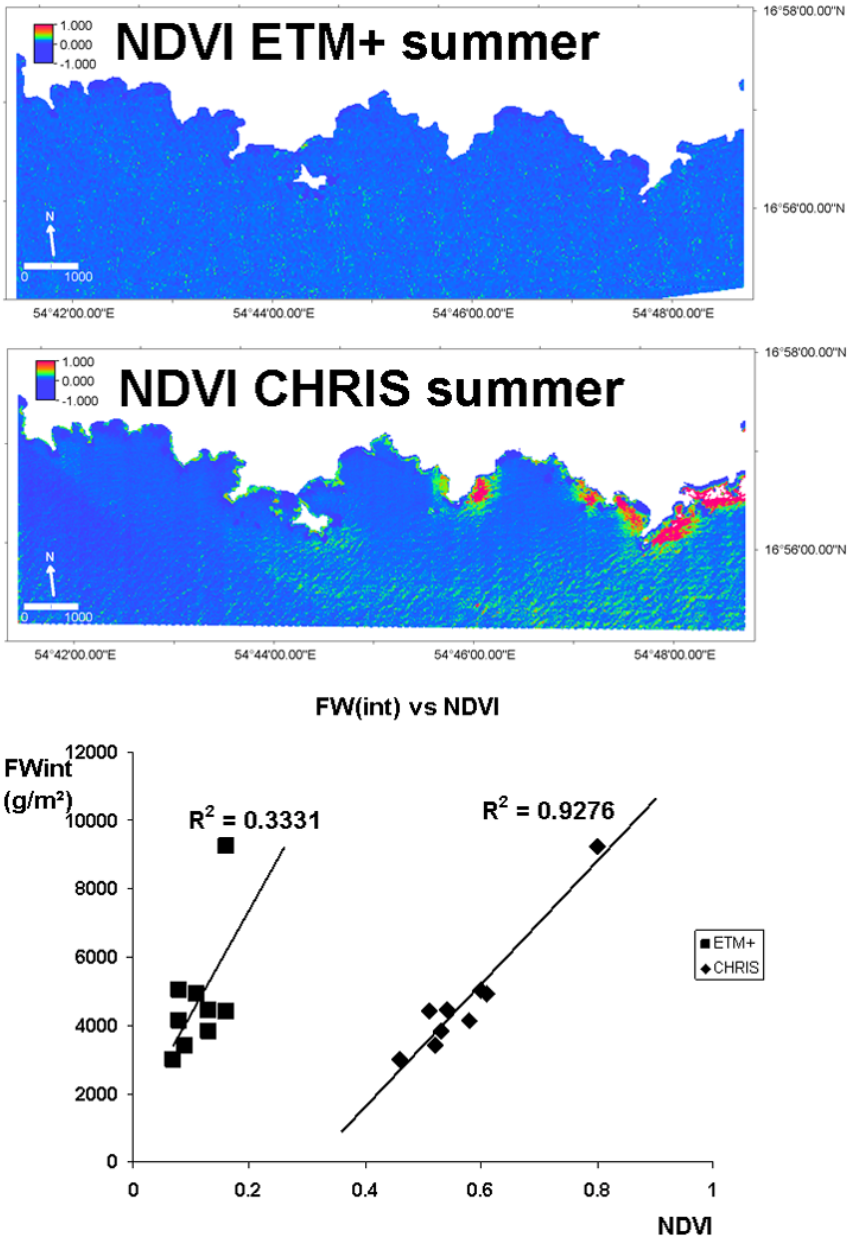
**Figure 4.** ML classification results based on Landsat 7 ETM+ data (upper) and simulated ETM+ bands based on weighted-average aggregation of PROBA/CHRIS bands (lower).

These results supported the idea that a higher spatial resolution is mostly beneficial in clear waters, and illustrated the role of environmental conditions in the need of a high spectral resolution to discern several algal and coral bottom types when dealing with a high habitat complexity (Mumby & Edwards, 2002; Capolsini et al., 2003).

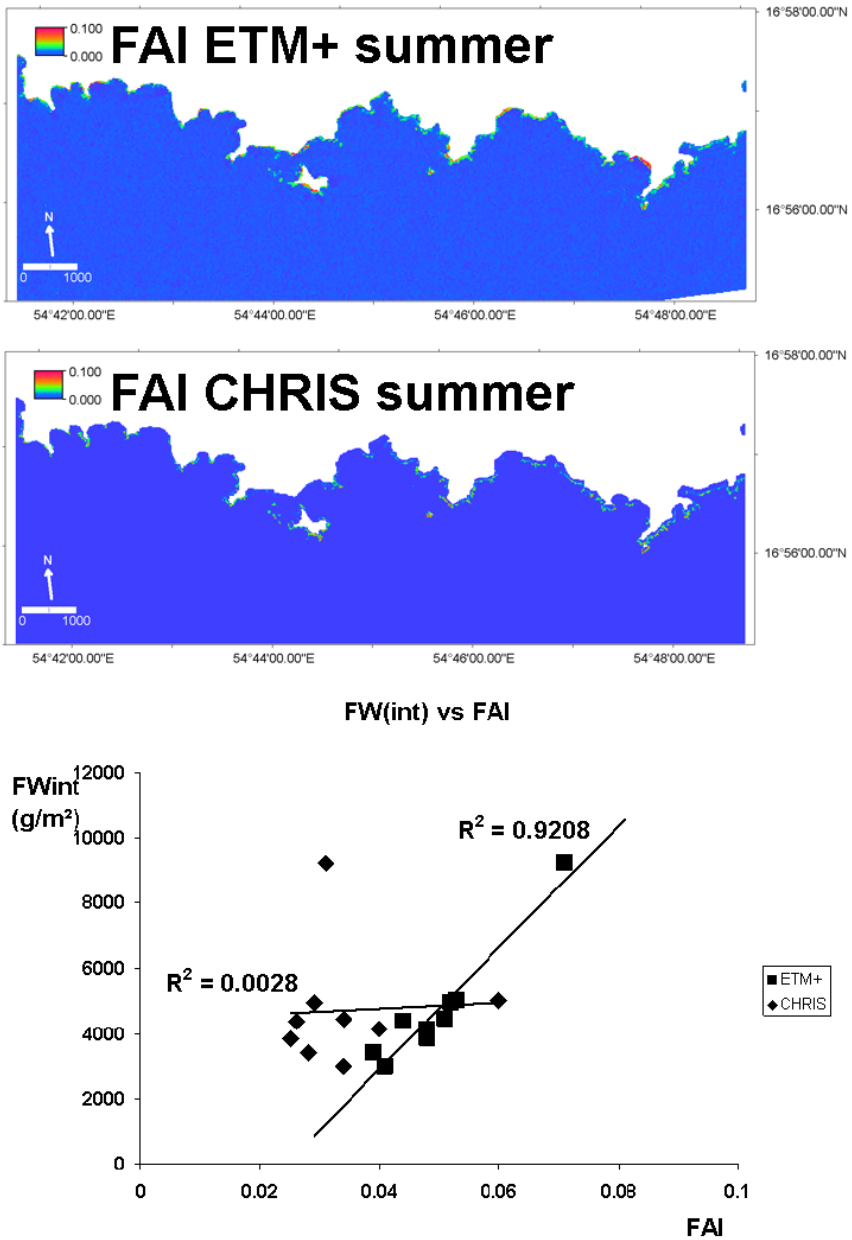
#### SPECTRAL INDEX MAPS

Although the NDVI map for the summer monsoon based on Landsat imagery showed some positive values offshore, there were only slightly positive values apparent for the intertidal pixels (figure 5a). This was probably caused by the mid to high tide in combination with the spatial resolution of 30 m, which greatly affected the NDVI signal of the few emerged algal communities. As a result, there was a very weak but non-significant overall correlation with intertidal summer biomass ( $R^2 = 0.333$ ,  $p = 0.10$ ,  $N = 9$ ; figure 5a). Elevated offshore values were either due to high surface water chlorophyll content or illumination effects caused by waves. The Landsat winter NDVI showed no positive values at all (figure 5c). While also showing strong offshore effects characterized by a clear wave pattern, the PROBA/CHRIS-based NDVI map for the summer clearly indicated the presence of intertidal seaweeds, strongly correlating with intertidal biomass ( $R^2 = 0.927$ ,  $p < 0.01$ ,  $N = 9$ ; figure 5a). The winter monsoon NDVI based on CHRIS datasets showed only very slightly elevated values along the coast, while suffering heavily from offshore effects caused by deviating illuminations (figure 5c).

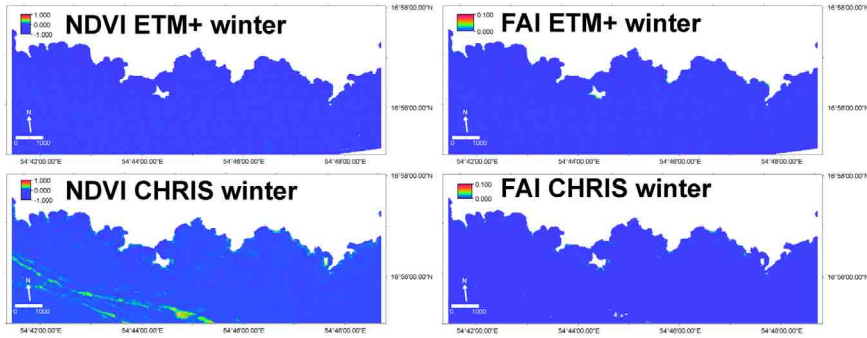
In contrast with the summer monsoon NDVI based on Landsat imagery, the Landsat FAI clearly detected the presence of intertidal seaweeds resulting in a very good correlation with intertidal summer biomass ( $R^2 = 0.921$ ,  $p < 0.01$ ,  $N = 9$ ; figure 5b), regardless of the mid to high tide. This could probably be attributed to the higher sensitivity of the SWIR band to vegetation, while the ability to eliminate illumination effects would enable the detection of very shallow and surfacing seaweeds in the intertidal zone. The Landsat winter FAI showed no indication of intertidal seaweeds (figure 5c). Since the FAI showed no elevated values offshore, it could be assumed that increased NDVI values in both CHRIS- and ETM+-based maps were caused by wave illumination effects rather than floating algae or blooms.



**Figure 5a.** NDVI maps based on Landsat 7 ETM+ data (upper) and PROBA/CHRIS data (lower) for the summer monsoon. The graph below shows the correlation between spectral index values and intertidal biomass. Squares represent ETM+; diamonds represent CHRIS data.



**Figure 5b.** FAI maps based on Landsat 7 ETM+ data (upper) and PROBA/CHRIS data (lower) for the summer monsoon. The graph below shows the correlation between spectral index values and intertidal biomass. Squares represent ETM+; diamonds represent CHRIS data.

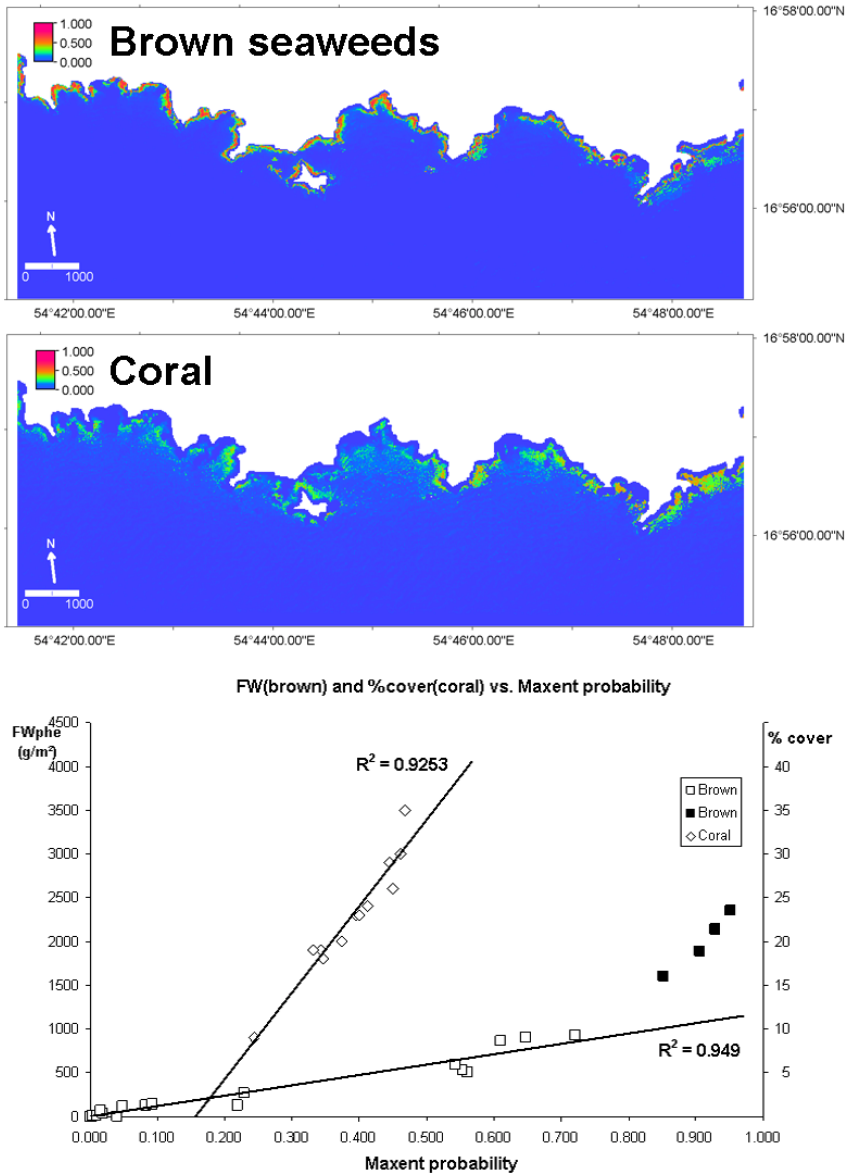


**Figure 5b.** NDVI (left) and FAI (right) maps based on Landsat 7 ETM+ data (upper) and PROBA/CHRIS data (lower) for the winter monsoon.

The correlation of the CHRIS-based FAI with intertidal biomass was extremely low and non-significant ( $R^2 = 0.003$ ,  $p = 0.89$ ,  $N = 9$ ; figure 5b), demonstrating the inability of CHRIS band 18 to substitute for SWIR wavelengths. The few elevated values along the coast were due to illumination effects caused by breaking waves. The winter FAI based on CHRIS bands lacked signal (figure 5c).

#### MAXENT SUB-PIXEL MODELING

The Maxent model for the brown algal bottom type showed high probabilities along the entire coastline of the study area, mainly situated in the shallow subtidal areas, with near-zero probabilities for sandy areas and intertidal pixels (figure 6). This closely matched the brown algal areas as delineated by the CHRIS-based maximum likelihood classification. Consequently, the model showed an AUC for the test data of 0.976 ( $SD \pm 0.023$ ). Correlation analyses between predicted probability and fresh weight data generally supported the strong model performance (figure 6), with a linear relation ( $R^2 = 0.949$ ,  $p < 0.01$ ,  $N = 17$ ) for all the data points except the four highest biomass values. This was probably because the biomass values also correlated with cover percent, except for the highest biomass values where cover is 100% but biomass can continue to increase by growth towards the surface. Brown algae in the area have been observed to reach the surface from 5 m depth. Based on the observed relation, a partly categorical biomass map could be calculated in the framework of resource conservation for the Green turtle.



**Figure 6.** Potential distribution map of the brown algal and coral bottom types produced by Maxent. Values between 0 and 1 represent probability of occurrence. Diamonds represent coral; squares represent brown seaweeds. Filled squares show values that have not been incorporated in correlation analyses.

The Maxent model for coral communities also showed probabilities along the entire coastline of the study area except for sandy areas. Generally, the suitable pixels were located in the deeper subtidal and showed overall lower



probabilities, not exceeding 0.5 (figure 6). However, the model had strong support, with an AUC for the test data of 0.988 ( $SD \pm 0.007$ ) and a strong linear correlation ( $R^2 = 0.925$ ,  $p < 0.01$ ,  $N = 12$ ) between predicted probabilities and coral cover (figure 6). This was a good indication that coral communities covered a far greater area than obvious from the maximum likelihood classification and were for the greater part overshadowed or overgrown by seaweeds during summer monsoon, which probably caused the lower predicted probabilities. Thus, when these probability maps would be combined in a post-modeling classification hardening operation, the outcome would mostly favor brown seaweeds, similar to the maximum likelihood classification outcome.

## CONCLUSION

The recurring presence of two different benthic communities in different water column properties in one place over the course of a year offered a unique opportunity to pinpoint the difficulties of high resolution satellite monitoring of algae on coral reefs. It appeared that a broadband multispectral sensor such as Landsat 7 ETM+ was only able to perform well for clear waters and low habitat complexity situations. However, the availability of a SWIR band was beneficial to neutralize illumination effects in the detection of intertidal, surfacing and floating algae. A narrowband sensor such as PROBA/CHRIS outperformed ETM+ for benthic bottom type classifications across environmental conditions, showing an extensive seasonal turnover between coral and macroalgal dominance in the study area. This demonstrated the need for development of high resolution superspectral sensors to monitor algal dynamics on coral reefs. Additionally, a recently developed Maxent implementation was successfully used to calculate sub-pixel bottom type probabilities from the superspectral information.

## ACKNOWLEDGEMENTS

Tony Vanderstraete and Alanda Savat are greatly acknowledged for assistance in image processing. Ali Al-Kiyumi, Jelle Evenepoel, Jan Pauly, Koen Pauly, Pieter Provoost, Tom Schils and Mussallem Zaidan Jaboob have assisted in the field. This research was embedded in grant G.0142.05 from the Fund for Scientific Research - Flanders (FWO), Belgium.

# CHAPTER 7

## **LOW-COST VERY HIGH RESOLUTION INTERTIDAL VEGETATION MONITORING ENABLED BY NEAR- INFRARED KITE AERIAL PHOTOGRAPHY**

Klaas Pauly & Olivier De Clerck

## **ABSTRACT**

With ecosystem services of intertidal habitats under rising pressure of human disturbance and climate change, monitoring habitat diversity is increasingly required. However, field-based surveys are time and resource-intensive and often do not provide spatially explicit information. While airborne (multi-spectral) photography and LIDAR (Laser Imaging Detecting And Ranging) offer an efficient, very high resolution and high-quality solution, the costs for skilled crew and equipment often preclude their use in remote areas, for small reserves and in developing countries. We present a simple yet robust, low-cost, low-altitude aerial photography solution using a kite and off-the-shelf camera equipment, resulting in photos covering the near-infrared part of the spectrum for vegetation monitoring. Photos can be mosaiced to generate 3D models, orthophotomosaics, vegetation indices and supervised classifications using low-cost computer vision and remote sensing software. We demonstrate the utility of kite aerial photography for intertidal monitoring in a case study in Northern France and discuss strengths and weaknesses of kite aerial photography.

## INTRODUCTION

Rocky intertidal coasts offer important habitats supporting biodiversity by providing food and shelter. However, these habitats have been observed to decline globally over the past decades, affecting the ecosystem services that they provide (Ambrose & Smith, 2004). Links to human disturbances such as collecting, trampling and turning of rocks have emerged, and these effects may be worsened by climate change in the coming decades. Many countries now require monitoring schemes for these vulnerable habitats (Chust et al., 2008). Changes to benthic communities have often been recorded qualitatively using field surveys in the past. However, inconsistent timing, detail and extent of surveys have hampered establishment of a baseline map and quantitative spatially explicit change detections (Alexander, 2008).

Advanced technologies such as remote sensing have been shown to lower the cost in monitoring schemes and increase mapping accuracy significantly (Lengyel et al., 2008). However, spatial resolution of spaceborne imagery precludes capturing the typically high intertidal rocky habitat variability. By contrast, aerial color or multispectral photography or airborne LIDAR have been shown to be effective in intertidal mapping efforts (Chust et al., 2008). Unfortunately, since many factors such as weather and remoteness are involved, the elevated costs for an aircraft together with highly trained staff and special camera equipment often rule out regular monitoring campaigns. Recent years have seen the development of low-cost alternatives, such as the use of small unmanned aerial vehicles (UAV; see Laliberte et al. (2010), although the cost for a professional UAV system still amounts to approximately \$60,000) or tethered low-altitude balloon (Planer-Friedrich et al., 2007), helikite (Verhoeven et al., 2009) or kite aerial photography using consumer-grade cameras. Additionally, recent advances in computing power and software availability have enabled low-cost processing of consumer-grade photos, including advanced classifications algorithms and image-based 3D reconstruction.

From these systems, kites provide arguably the cheapest and most simple yet robust solution. Kite aerial photography (KAP) has been around since 1887 (Archibald, 1897), but was only much later used in mapping and monitoring studies in the coastal (Scoffin, 1982) and the terrestrial

environment. Since, applications have covered archaeology (Dvorak & Dvorak, 1998), geomorphology (Marzolff & Poesen, 2009), agriculture (Oberthur et al., 2007) and vegetation monitoring (Wundram & Löffler, 2008). As several of these applications require patterns in vegetation health to be detected, imaging the near-infrared (NIR) part of the spectrum became essential to discern different vegetation types and stress factors (Lebourgeois et al., 2008). CCD and CMOS sensors found in digital cameras are inherently sensitive to NIR light, and modified cameras (see Verhoeven, 2008); obtained by removal of the internal NIR-blocking filter in front of the sensor, used by manufacturers to simulate human eye color perception) mounted for KAP have been demonstrated to yield information otherwise not achievable with a digital compact camera (Gerard et al., 1997; Siebert et al., 2004).

The aim of the present paper is to explore the utility of NIR-enabled KAP as a tool for monitoring intertidal rocky shore habitats in the Wimereux area (northern France), with a focus on seaweed communities. We assess best baseline mapping practices and show the potential for change detection using two imagery series acquired over 1 year. The rationale is to keep the design of the kite, the camera suspension and operation as well as the subsequent image analysis as simple and low-cost as possible, while using the latest technologies.

## **MATERIAL & METHODS**

### **STUDY AREA**

The study area comprises a rocky intertidal stretch running south-north between the coastal towns of Boulogne-sur-Mer and Wimereux (Nord-Pas-de-Calais, France), known as Pointe de la Crêche, located between N50.750 and N50.756. The area is known to have supported extensive and dense intertidal brown algal communities dominated by *Fucus* spp. which collapsed between 1990 and 2000 (Coppejans, pers. comm.). Since 2000, wave-exposed rocks are either bare (upper zones), dominated by limpet/barnacle communities (*Patella/Balanus*) or mussel communities (*Mytilus*; mid-tidal zones) or spionid worm reefs causing heavy siltation on rock platforms (lower zones). Intertidal seaweed communities dominated by dense *Fucus*, green algal *Ulva* spp. and red algal *Porphyra* stands (mid to upper zones) are

still found mostly on the edges of rocky platforms and vertical surfaces. Scattered mixed assemblages with mainly red algae can be found in the lower zones.

#### KITE AERIAL PHOTOGRAPHY

Kite aerial photographs were acquired on 16 April 2010 and on 7 April 2011. Depending on wind conditions, either a Rokkaku 7' or FlowForm 32' (figure 1a and 1b) were launched to a height between approximately 80m (in 2011) to 160m (in 2010). The camera was mounted on Brooxes Basic Frames tethered to a Picavet suspension system attached to the kite line approximately 20m below the kite (figure 1b). The camera rig was set to look straight down and an intervalometer was programmed on the camera's SD card using Canon Hacker Development Kit (CHDK, freely available at <http://chdk.wikia.com>) which triggered the camera every 5 seconds. Hence, no external electronic parts or remote control were used on the camera rig and all settings were made prior to the KAP session, enabling the kite pilot to walk around freely for terrain acquisition.

In 2010, photos were taken under overcast conditions with an unmodified (true-color) 12MP Canon Powershot SX200 IS set at ISO 200, 5mm (28mm equivalent) focal length and variable shutter speed and aperture. A shutter speed of at least 1/500<sup>th</sup> is needed to prevent motion blur. In 2011, both a true-color (RGB) and a false-color series were subsequently acquired. The former used the same SX200 camera, while the false-color camera was a full-spectrum modified 10MP Canon Ixus 870 IS with a red-blocking Lee 172 Lagoon Blue film filter fitted to the lens, hence capturing blue, green and NIR light (Hunt & Linden, 2009). Both cameras were set to 5mm focal length, ISO100 (because of intense direct sunlight) and variable shutter speed and aperture in the latter session.

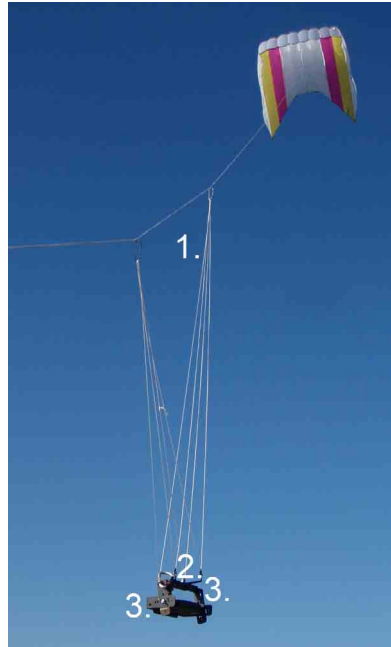
Individual image extent and inherent resolution were calculated before the KAP sessions as a guideline. Photo coverage can be calculated based on the relation between focal length ( $f$ ), acquisition height ( $H$ ) and sensor width ( $d$ ), from which the image width ( $D$ ) can be calculated as

$$D = \frac{d \cdot H}{f} \quad (1)$$

A



B



**Figure 1:** Equipment for kite aerial photography: a Rokkaku 7' framed (A) or FlowForm 32' frameless kite (B), used depending on wind speed and variability. The camera is suspended from the kite line using a Picavet suspension (1), a Picavet cross (2) and two pivoting Brooxes Basic KAP frames (3)

The spatial resolution or ground sampling distance (GSD) can be calculated based on pixel size, acquisition height and focal length as

$$GSD = \frac{H \cdot D}{\frac{P(d)}{f}} \quad (2)$$

where  $P(d)$  is the number of pixels at the long side of the sensor. For a 12MP camera at 140m flying height and a 10MP camera at 60m flying height at minimal focal length of 5mm, this results in expected coverage and resolution of 173m by 130m at 4cm GSD and 74m by 66m at 2cm GSD per photo for a 1/2.3" camera sensor, respectively.

### GROUND TRUTHING

Two days after the first acquisition date and coinciding with the second date, two separate transects measuring 50m by 2m were delineated including all major habitat types covered by the KAP, for which the outlines were drawn

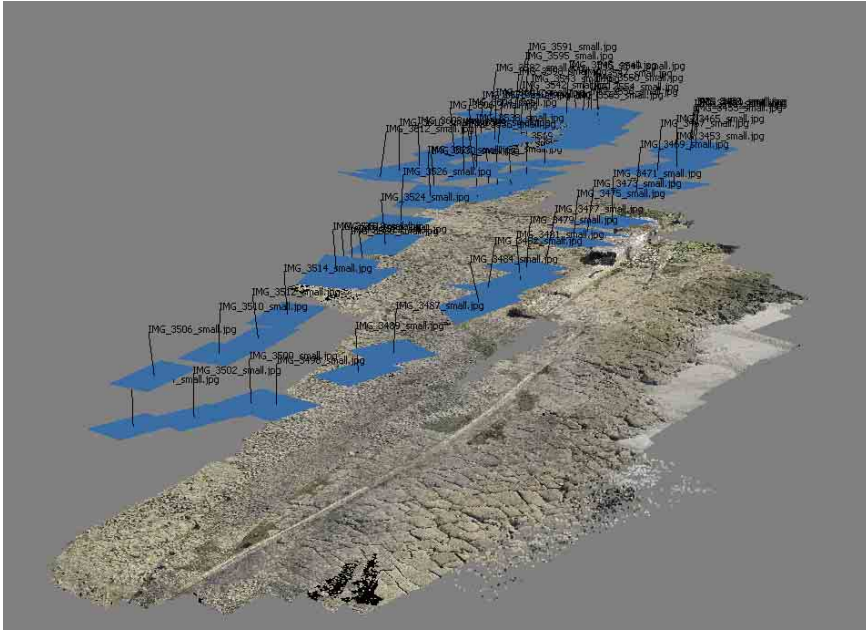
together with field identity codes. The drawings were subsequently digitized and overlaid on the image mosaics to provide classification training and test data, as well as to visualize the difference between aerial photography-based and traditional intertidal habitat monitoring. Using a handheld Garmin eTrex GPS, coordinates of transect as well as of other conspicuous landmarks were logged. Where insufficient landmarks could be found, georeferencing was aided by laminated A3-format high-contrast target cards which were spread out across the terrain prior to image acquisition.

#### IMAGE MOSAICING AND PROCESSING

From about 700 pictures acquired in a one-hour KAP session, only the sharpest pictures approaching nadir view and with reasonable overlap are retained for further analysis.

Image mosaicing can then be done in two ways. Images could either be master-slave rectified to each other in a relative coordinate system using at least 12 manually located tiepoints for every image added. After tiepoint editing, a 1<sup>st</sup> order transformation can be applied (avoiding excessive errors at the edges). However, the preferred method is to reconstruct a dense surface model of the scene with computer vision software using structure from motion (SFM) and dense stereo-reconstruction algorithms (Verhoeven, 2011). Based on the 3D model, an orthophotomosaic can then be generated. The latter approach has the advantage of automation which greatly reduces processing time and allows better handling of low-oblique imagery while also accounting for internal camera calibration parameters. Additionally, when parts of images cannot be matched due to a dynamic environment such as water and waves, these parts are automatically masked out from the resulting mosaic. However, since SFM algorithms need to retrieve thousands of feature points over the scene, each one recognized in at least 2 images, this approach may fail on low-contrast and dynamic environments such as a wet beach. In this study, the 14 selected images from the KAP session in 2010 were manually rectified and mosaiced using ClarkLabs IDRISI Taiga, while the roughly 140 selected images for the 2011 sessions (separately for the RGB and false-color acquisitions) were processed using AgiSoft PhotoScan Pro (Verhoeven, 2011); figure 2). To avoid memory issues in PhotoScan, the study area was subdivided in four blocks which were subsequently aligned.





**Figure 2.** Oblique view on 3D terrain reconstruction and image mosaicing of 63 RGB photos in PhotoScan; the southernmost of 2 blocks to be aligned. The approximate viewpoint (\*) and viewing directions are shown in figure 3B. Besides terrain reconstruction, the camera acquisition points and camera orientation for each photo are calculated, clearly showing the walking lines.

Upon importing in Idrisi Taiga, the RGB mosaic resulting from the first date and the RGB orthophoto mosaic from the second date were master-slave referenced to the false-color orthophoto mosaic to allow for easy classification signature extraction, overlay analysis and change detection. The false-color mosaic was originally georeferenced using the handheld GPS-measured waypoints, which resulted in a total Root Mean Square error (RMS) of 49.2 pixels. All mosaics were resampled to a GSD of 4cm. Resulting mosaics were subsequently color-separated to obtain blue (B), green (G), red (R) and NIR bands for classification and vegetation index retrieval. For the RGB mosaic obtained in 2010, no vegetation indices could be calculated. For 2011, two options were available to obtain a vegetation index: while the R band from the RGB mosaic could be used in combination with the NIR band from the separate acquisition to calculate the NDVI (3), preference is given to retrieve a modified NDVI based on the single false-color acquisition for pragmatic reasons, substituting the red band by the green or blue band (4; Lebourgeois et al., 2008).

$$NDVI = \frac{NIR - R}{NIR + R} \quad (3)$$

$$GNDVI = \frac{NIR - G}{NIR + G} \quad \text{or} \quad BNDVI = \frac{NIR - B}{NIR + B} \quad (4)$$

Since pixels are much smaller than benthic cover class patches in sub-decimeter resolution imagery, pixel variability within classes is as high, or even higher than between classes, hampering traditional pixel-based maximum likelihood classification approaches. Therefore, image segmentation was undertaken in Idrisi Taiga, grouping pixels based on their spectral variance and local scene structure into polygons using a 3 by 3 pixel moving window (Laliberte et al., 2010). Half of the polygons within the ground truthing transects were then assigned to the different classes found in the transects for spectral signature training, while the polygons from the other transect were kept aside for testing and error matrix construction. In order to enhance spectral separability, different band combinations of RGB, NIR and NDVI were tested for classification. Based on the generated spectral signatures, a maximum-likelihood classification was run and post-processed using a 5 by 5 pixel majority filter. Preliminary classifications were run on image mosaics from 2011 to assess end user accuracies for each class and merge the most erroneous classes (commission error > 0.5). Besides omission errors (producer accuracy) and commission errors (user accuracy), accuracy was assessed using the kappa index of agreement (KIA) between ground truth data different from training data and the classified image. KIA values generally range between 0 and 1, but can take negative values in case of no agreement. Values between 0.4 and 0.8 can be viewed as fair to good while values below 0.4 generally indicate poor agreement, although the results depend on the number of classes and other factors.

Change detection was demonstrated by reclassing the bottom types of the second date RGB-based classification to the fewer classes of the first date. The mosaics were then cropped to the best overlapping area and a cross-tabulation was carried out.

## RESULTS

### GROUND TRUTHING

Although the image acquisitions on both dates covered a certain overlap area, the session centers were some distance apart. Hence, ground truthing transects were chosen on different locations for the acquisition dates. In 2010, six bottom types were noted in the transects (table 1). Additionally, a microphytobenthos (MPB) bloom that was seen on the beach on the acquisition day had disappeared by the time the transect work was carried out. Therefore, an additional training site was delineated on the bloom visible in the image mosaic outside of the transect. In 2011, seventeen classes were discerned in the transects (table 1). As a result of the preliminary classification runs, classes defined by mixed assemblages like *Fucus/Ulva*, *Ulva/Porphyra* or *Balanus* together with *Fucus* or *Ulva* were considered as spectrally dominated by 1 species (*Ulva* or *Fucus*), thus reducing the habitat classes to 12 (table 1). Since deep crevices, large boulders and certain man-made structures cast dark shadow patches on a sunny day, shadow was also added as a separate class for the second date.

### RGB IMAGE CLASSIFICATION

The RGB mosaic from 2010 suffered from exposure differences between individual images and from low contrasts within images due to hazy and drizzling weather (figure 3A). As a result, accuracies for this KAP session were the lowest, with an overall KIA of 0.15 (table 2a). The low KIA was also influenced by the absence of the MFB class in the ground truthing data. Classification based on RGB mosaics from 2011 achieved a KIA of 0.43, where misclassification resulted mainly from the fact that intertidal pools could hardly be discerned on RGB imagery, and mixed algal/animal or substrate patches were often not recognized as containing algae (table 2b). Change detection on the best overlapping area showed an important turnover from rock to sand, and seaweed to rock classes in the mid-tidal zones, although increased algal cover was mapped in the upper tidal zone for the second date (figure 3C).

## NDVI

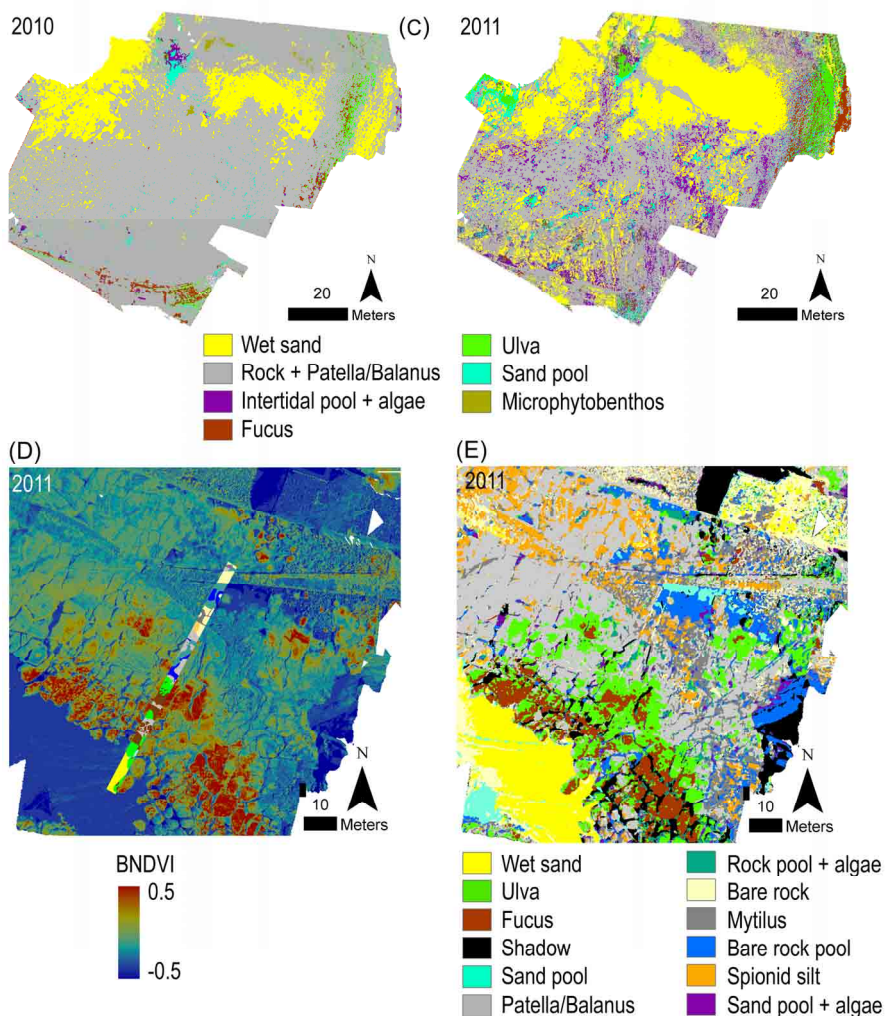
Visual inspection of the NDVI images revealed adequate indication of intertidal seaweeds by using either blue or green as a substitute for the red band. However, the blue-substituted NDVI images (BNDVI) yielded more contrast and proved more sensitive to shadow patches and water or wet surfaces; therefore, only the BNDVI was retained for further display and analysis (figure 3D). The BNDVI provides good contrast between vegetated and bare or wet surfaces, and is able to discern submerged vegetation in tidal pools. Algae in tide pools are also seen to reflect NIR on false-color imagery down to about 0.5m. A single threshold could not be applied to the entire mosaic to make a hard distinction between seaweeds and substrate, since the mosaic covers a wide tidal zone acquired at spring low tide, where the difference in exposure time varies up to 6 hours over the entire scene. This causes stress-related responses in the NDVI.

**Table 1:** Overview of corresponding bottom types recognized in transects and used for image classification. Codes between brackets refer to the error matrices in Table 2.

2010	2011 aggregated classes	2011 transect classes
Wet sand (WS)	Wet sand (WS)	Wet sand
	Spionid Silt (SS)	Spionid silt
Rock (R)	Bare rock (BR)	Bare rock
		Bare concrete
	<i>Patella/Balanus</i> (PB)	<i>Patella/Balanus</i>
	<i>Mytilus</i> (M)	<i>Mytilus</i>
Sand pool (SP)	Sand pool (SP)	Sand pool
Pool + algae (PA)	Sand pool + algae (SPA)	Sand pool + algae
	Bare rock pool (RP)	Bare rock pool
	Rock pool + algae (RPA)	Rock pool + algae
<i>Ulva</i> (U)	<i>Ulva</i> (U)	<i>Ulva</i>
		<i>Ulva + Porphyra</i>
		<i>Ulva + Balanus</i>
<i>Fucus</i> (F)	<i>Fucus</i> (F)	<i>Fucus</i>
		<i>Fucus + Ulva</i>
		<i>Fucus + Balanus</i>
Microphythobenthos (MFB; *)	Shadow (SH)	Shadow

(\*) Ephemeral bloom, not present at the time of transect monitoring and therefore only visually sampled for classification training.





**Figure 3 (continued).** (A) shows the RGB mosaic obtained by manually rectifying 14 photos from the 2010 acquisition. The digitized ground truthing transect is shown in overlay. The red box in (A) and the corresponding upper red box in (B) indicate the position of (C), where RGB classification results of the overlapping area from 2010 and 2011 allow for change detection. Note that the bottom type “microphytobenthos” is only present in the 2010 classification. (B) shows the false color mosaic (B, G and NIR) resulting from 142 photos acquired in 2011. The lower red box indicates the position of (D) and (E). The former shows the BNDVI and the overlaying ground truthing transect while the latter shows the classification based on B, G, NIR and BNDVI.

FALSE-COLOR AND NIR-ENHANCED CLASSIFICATIONS (2011 ONLY)

Expanding the RGB data with the NIR band from the false-color mosaic (figure 3B) and the BNDVI progressively improved classification KIA from 0.43 to 0.58. However, best classification results were obtained from the single sensor- based combination of B, G, NIR and BNDVI bands with a resulting KIA of 0.65 (table 2c). Although much more accurate in discerning different seaweed assemblages and (vegetated) intertidal pools, the lack of the red band decreased classification accuracies of benthic animal assemblages on rock platforms such as the heavily silted spionid worm banks, confused with sand, and *Mytilus* communities, confused with *Patella/Balanus* communities.

**Table 2.** Classification error matrices. (A) based on the 2010 RGB mosaic, (B) based on the 2011 RGB mosaic and (C) based on the 2011 B, G, NIR and BNDVI mosaics. UserAcc = user accuracy or (1 – commission error); ProdAcc = producer accuracy or (1 – omission error). Full class names of the codes are found in Table 1. The value in bold in the lower right cell is the kappa index of agreement (KIA).

(A)	WS	R	PA	F	U	SP	Total	UserAcc
WS	4074	1692	374	0	0	250	6390	0.64
R	458	2870	381	8	2	233	3952	0.73
PA	0	407	225	43	36	956	1665	0.13
F	0	1905	297	66	2	1032	3302	0.02
U	23	1499	183	24	0	1160	2889	0
SP	48	408	73	48	11	582	1170	0.5
MFB	1169	46	0	0	0	0	1215	0
<b>Total</b>	<b>5772</b>	<b>8827</b>	<b>1531</b>	<b>189</b>	<b>51</b>	<b>4213</b>	<b>20583</b>	
<b>ProdAcc</b>	<b>0.08</b>	<b>0.36</b>	<b>0.05</b>	<b>0.04</b>	<b>0</b>	<b>0.14</b>		<b>0.15</b>

(B)	WS	U	F	SH	SP	PB	RPA	BR	M	RP	SS	SPA	Total	UserAcc
WS	3704	79	170	10	20	869	12	7174	77	188	275	18	12596	0.29
U	4	5430	797	56	768	577	501	60	128	83	3	307	8714	0.62
F	0	904	3055	2008	71	390	471	108	2245	67	0	944	10263	0.3
SH	1	100	987	18842	0	108	62	18	246	7	0	468	20839	0.9
SP	0	1282	220	15	401	1054	358	540	531	365	32	1552	6350	0.06
PB	1	212	21	12	175	11338	52	2014	617	881	3279	60	18662	0.61
RPA	0	672	734	91	193	318	707	51	90	92	4	1333	4285	0.18
BR	893	123	45	4	4	2723	13	6230	1347	307	542	36	12267	0.51
M	11	653	786	336	41	1043	193	1984	12227	413	339	871	18897	0.65
RP	2	466	84	25	620	4773	420	1757	2438	2171	793	751	14300	0.15
SS	17	104	1	6	17	13615	5	1494	80	130	10467	9	25945	0.4
SPA	0	373	597	242	90	615	345	192	1039	138	31	2711	6373	0.43
<b>Total</b>	<b>4633</b>	<b>10398</b>	<b>7497</b>	<b>21647</b>	<b>2400</b>	<b>37423</b>	<b>3139</b>	<b>21622</b>	<b>21065</b>	<b>4842</b>	<b>15765</b>	<b>9060</b>	<b>159491</b>	
<b>ProdAcc</b>	<b>0.8</b>	<b>0.52</b>	<b>0.41</b>	<b>0.87</b>	<b>0.17</b>	<b>0.3</b>	<b>0.23</b>	<b>0.29</b>	<b>0.58</b>	<b>0.45</b>	<b>0.66</b>	<b>0.3</b>		<b>0.43</b>

(C)	WS	U	F	SH	SP	PB	RPA	BR	M	RP	SS	SPA	Total	UserAcc
WS	7377	0	30	15	7	80	55	1433	0	18	179	0	9194	0.8
U	3	10138	1719	160	6	1305	409	157	76	8	0	17	13998	0.72
F	0	2199	7331	170	0	56	0	20	0	0	52	0	9828	0.75
SH	0	2	0	30851	51	0	254	16	107	32	0	68	31381	0.98
SP	9	0	2	432	1644	0	396	26	0	86	7	158	2760	0.6
PB	30	1670	279	35	11	21937	864	2654	651	3	3473	40	31647	0.69
RPA	0	114	33	328	7	107	1126	69	27	64	0	80	1955	0.58
BR	328	191	117	192	41	369	214	2381	145	42	482	43	4545	0.52
M	0	333	52	643	29	328	521	1711	5099	64	109	220	9109	0.56
RP	2	55	7	513	245	9	533	192	254	2149	225	632	4816	0.45
SS	41	181	55	19	27	3940	694	3550	285	8	5195	4	13999	0.37
SPA	0	90	82	2599	118	97	294	23	356	180	0	2286	6125	0.37
<b>Total</b>	<b>7790</b>	<b>14973</b>	<b>9707</b>	<b>35957</b>	<b>2186</b>	<b>28228</b>	<b>5360</b>	<b>12232</b>	<b>7000</b>	<b>2654</b>	<b>9722</b>	<b>3548</b>	<b>139357</b>	
<b>ProdAcc</b>	<b>0.95</b>	<b>0.68</b>	<b>0.76</b>	<b>0.86</b>	<b>0.75</b>	<b>0.78</b>	<b>0.21</b>	<b>0.19</b>	<b>0.73</b>	<b>0.81</b>	<b>0.53</b>	<b>0.64</b>		<b>0.65</b>

Common for all classifications using the NIR band was misclassification of bottom types in large shaded areas such as at the basis of cliffs. Over these areas, the camera adjusted its shutter speed, making the area too bright to be recognized as shadow, while darkening the known spectra. This tricked the classification algorithm in classifying these patches as underwater, although correctly recognizing presence or absence of vegetation and rock or sand substrate.

## **DISCUSSION**

### **QUALITY OF THE PRESENTED DATA**

From a monitoring point of view, user accuracies for the target groups (i.e. 1 – commission error) are most important, besides a good overall mapping performance indicated by KIA. In that perspective, the choice for true-color or false-color aerial photography can be based on varying classification accuracies for the project target species. The present case study clearly shows that a combination of blue, green and NIR bands is outperforming true-color imagery for seaweed (and potentially other intertidal vegetation and microphytobenthos) monitoring. To compensate for the loss of information on the red part of the spectrum, the BNDVI was successfully added as an extra spectral band. Since pixels on camera sensors are covered by either blue, green or red filters (Verhoeven et al., 2009) and only red filters pass a significant amount of NIR, red light should be blocked when photographing NIR and a full VNIR image cannot be acquired at once with a single camera (Hunt & Linden, 2009). Although the red band can be added from a coinciding or subsequent KAP session with another camera or filter, this study shows that while the red band is useful to discern certain invertebrate communities, it doesn't increase overall KIA. Additionally, resampling an RGB mosaic to match the false-color mosaic is a time-consuming step that may be avoided.

For the purpose of demonstrating change detection, the 2011 RGB mosaic was used to compare to the 2010 RGB mosaic, in order to avoid much more detailed NIR-based vegetation information from 2011 to bias the change detection. Due to the large commission error of classifying sand as rock in the 2010 classification, most of the turnover from rock to sand is indicative of improved classification accuracy in 2011. However, the



increased number of scattered sand pixels in the southern part of the area may indicate ongoing siltation, an increasing phenomenon in the study area during recent years. The silt is accumulated on rock platforms by reefs of spionid worms, and prevents algal communities from developing. This is in agreement with a decrease in mapped algal cover in the south of the change detection area. In the more dynamic upper intertidal zones, located to the west in the area, algal cover on the boulders seems to have increased from 2010 to 2011.

#### UTILITY OF KAP IN MONITORING AND BEST PRACTICE

Although (tethered) low-altitude aerial photography with a consumer-grade camera has been used successfully for vegetation monitoring in the terrestrial environment, and airborne aerial photography has been applied for seaweed monitoring, this is the first successful application of low-cost, low-altitude NIR aerial photography of intertidal habitats.

This case study raised some important practical points for KAP. A flying height between 100 and 200m greatly reduces the time and effort for acquisition (allowing walking quicker or trailing the car by 4WD or boat) and mosaicing, hence also reducing potential errors in the mosaics. The relative loss in GSD is mostly not relevant for monitoring. Further, a single NIR-enabled camera combined with a red-blocking external filter yields a good accuracy for general intertidal monitoring. The added gain in accuracy from the red band will often not compare to the extra acquisition (unless both cameras are suspended from the same kite) and processing time to include this information. Additionally, preference should always be given to computer vision-based 3D reconstruction and orthophoto mosaicing, rather than rectifying images relative to each other. However, the inherent characteristics of the intertidal environment may sometimes hamper the latter approach.

Like other monitoring methods, KAP has its specific drawbacks that need to be considered when designing and implementing a monitoring scheme. First, a KAP session cannot be planned with certainty more than a couple of days in advance, especially in areas with variable winds and weather patterns. This issue is worsened when acquisitions need to coincide with (spring) low tide and sufficient daylight (of concern at high latitudes).

However, some kites can be adjusted to a range of wind conditions and with a choice of kites, a wind force range from 2 to 6 Bft allows for KAP, which is wider than acceptable for certain types of UAV. Although we intentionally didn't use external electronics on the KAP rig, gyro-servos can be added to the three axes on the rig to keep the camera pointing down when the rig starts swinging in heavy winds, or to trigger the shutter each time the camera passes nadir. Second, preferably coinciding ground truthing data are vital to successful KAP monitoring studies, with a recommended two transects per mosaic for classification training and testing. The very high resolution in relation to habitat variability requires a far larger training and test sample than usually required for VNIR remote sensing. That said, KAP is especially useful in addition to field-based surveys, with recorded data acquisition and processing times for this study amounting to 200m<sup>2</sup>/h for the transects and up to 1 Ha/h by KAP. Lastly, while KAP is a good solution for certain monitoring needs, the image quality and resulting analyses will mostly not meet standards achieved in airborne surveying or professional UAV systems. However, the simple design and robustness of KAP is specifically meant to meet certain goals, such as remote and developing areas where budgets and access to spare parts and maintenance are not available. A frameless kite is virtually unbreakable and the lack of external electronics on the camera rig is beneficial for work in a sandy and salty environment, while the potential use of a consumer-grade shockproof and waterproof camera with internal timer further eliminates any risk. With the advent of low-cost computer vision and remote sensing software such as the packages used in this study and open source or free alternatives (such as ITC Ilwis, Agisoft PhotoScan Standard, MeshLab), valuable information can be gathered based on consumer-grade photography. A basic KAP set including NIR-enabled camera for monitoring can thus be assembled from \$750 onwards, with higher budgets allowing for professional software editions and better cameras. This figure is a one-time investment with virtually no subsequent costs for use. Since using a kite for aerial photography doesn't require a piloting license (only maximum flying heights and minimum distance from air traffic apply in many countries) and can be learnt within a week, this opens up aerial photography-based monitoring to many local conservation practitioners and managers.

## **ACKNOWLEDGEMENTS**

Peter Bults (Kapshop.com) and Lien Everaerd (Didakites NV) kindly provided test equipment and valuable advice on technical aspects of KAP. Tom Benedict (CFHT, Hawaii), Nathan Craig (Analytical Cartography and GIS Lab, Pennsylvania State University) and Geert Verhoeven (Archaeology Dept., Ghent University) are greatly acknowledged for sharing their insights in KAP data processing and feedback on the present work. Pieter Provoost and Koen Pauly have assisted in KAP field work.

# **CHAPTER 8**

## **GENERAL DISCUSSION**

## SPATIAL INFORMATION IN GIS-BASED MACROALGAL STUDIES

Spatially explicit information on biotic samples and environmental data is now relatively easy to download from various sources. However, while environmental data coverage has increased significantly in recent years with now many terabytes of environmental information ready to be incorporated in GIS-based and distribution modeling studies, reliable spatial information on biotic data remains scarce. Georeferencing specimens from herbarium collections is very time-consuming and can often not be carried out with sufficient resolution or certainty to match the spatial scale and data resolution for the desired study. As a good rule of thumb in order to obtain valid species-environment relations, we recommend that each sample should be georeferenced to within half a pixel error of the environmental dataset resolution. While this may seem easy to attain for coarse resolution studies, this may not be an evident task when adding coordinates to herbarium specimens with unclear locality names. It is also worth noting that the error is the crucial factor here, and not the accuracy. While a consumer-grade hand-held GPS may for instance indicate ones position with approximately 1m accuracy (showing five decimal places in decimal degrees), the error on the positioning is typically 5-15m and hence, biotic data georeferenced in this way should not be used in relation to environmental variables with a spatial resolution of less than 30m. For records georeferenced from herbarium collections, the error is less evident and can vary greatly.

In Chapter 2, specimens lacking any locality information and coordinates on a decimal degree were *a priori* excluded for the modeling effort. For specimens with locality information but no coordinates, Google Earth was used to assign coordinates. These records were only used for modeling if this could be done to within a decimal degree based on textual descriptions or collectors' indications. Overall these prerequisites greatly reduced the availability of occurrence records for modeling, but all modeling records except one were georeferenced to at least 1 km resolution, well below the half-pixel rule to allow for some uncertainty, which cannot be calculated exactly for herbarium data. . One record that was referenced to 11 km resolution was still considered for modeling because neighboring pixels did not differ substantially, again allowing a certain degree of uncertainty

without jeopardizing the model. In Chapter 3 and 4, a similarly stringent selection procedure for occurrence records was adopted, and all records were georeferenced to at least 1 km. In Chapter 5 and 6, all occurrence records were georeferenced using a hand-held GPS. In practice, the observational error indicated on the GPS device was rarely more than 15 m, i.e. half the size of a Landsat pixel and 1/4<sup>th</sup> of the pixel size in Landsat-based environmental variables in Chapter 5. For use in image georeferencing and classifications based on CHRIS imagery (18 m resolution), a GPS reading was only logged when the error was below 9m. Similarly, ground truth points retrieved from video transects were only retained when the log file indicated an error of less than 9 m. In Chapter 7, pixel training was carried out without the use of GPS because of the very small pixel sizes. Instead, a 50 x 2 m transect delineated by measuring tapes visible on the color and NIR photographs were used to draw the habitat classes. Pixels within each image segmentation polygon in the transects were then assigned to their respective class. The segmentation approach was adopted to cope with the situation in which pixels are smaller than the objects targeted for classification, to which the rule of thumb outlined above cannot be applied.

Particularly of concern to seaweeds is the role of taxonomy in spatial information errors. This is especially true for information from public biodiversity databases such as OBIS, where not only coordinates are often erroneous and hard to check as far as seaweed entries concern, but also identifications are often impossible to check. Mixing occurrence records of different species may significantly affect model performance, depending on the study goals. Moreover, many macroalgal taxa are characterized by a high cryptic diversity, further increasing the difficulty of obtaining valid models (Lozier et al., 2009). In Chapter 3 and 5, we have made use of phylogenetic analysis to confirm species identity or delineate evolutionary significant units for species known to feature high degrees of cryptic diversity.

## **GENERAL ASPECTS OF MODELING MACROALGAL DISTRIBUTIONS**

The use of macroalgae as a marine test case in assessing niche modeling techniques throughout this thesis has some specific advantages and drawbacks, which we discuss here in the conceptual framework of niches

and their modeled distributions.

Any form of species distribution modeling or ecological niche modeling relies fundamentally on the duality between multidimensional environmental space, in which a species' niche is defined, and (in the case of the chapters presented here) two-dimensional geographical space (Colwell & Rangel, 2009). This means that, theoretically, there is a partially reciprocal relationship in which each point in geographic space translates to exactly one point in environmental space at a given time, but a point in environmental space may correspond to one or multiple geographic points, or none. Niche modeling algorithms use this relationship by reading the environmental values from  $n$  raster layers at known occurrence localities, delineating the niche using more or less complex algorithms in  $n$ -dimensional space, and subsequently projecting its potential distribution back to physically located pixels which are environmentally similar. However, in practice this relationship depends on the spatial resolution and community structure under consideration. As environmental data are composed of pixels rather than infinitely small points, one pixel can contain several microhabitats with associated differing environmental conditions. Also, conditions experienced by canopy and understory species may differ for a given locality. Hence, values in environmental space attributed to a pixel should in fact be associated with a confidence interval.

One of the most important assumptions to obtain solid relationships and valid niche models is the (pseudo-)equilibrium state of species in their sampled environment (although ways have been proposed to compensate for lack of equilibrium in an additional data layer; (Elith et al., 2010a). This is because biotic and non climatologies-based environmental samples represent a snapshot in time (and place in the case of regional modeling), and occurrence records of species not in equilibrium with their environment would result in underestimation of the potential distribution (Guisan & Thuiller, 2005). Seaweeds, by their benthic nature, are forced to adapt to local conditions, to move or to perish. Combined with their generally short life cycle allowing them do so quickly, this means that macroalgal communities which are known to be established at a locality for some decades can be assumed to have reached a near-equilibrium with their environment. This assumption was made for all niche modeling case studies in this thesis. However, predictive performance of invasive species models

can be increased by training the model on both invaded range occurrence records as well as native records (Capinha et al., 2011; de Rivera et al., 2011). Although the former might not yet be in equilibrium with their environment, the constraints are typically considered to be greatest in native ranges (Keane & Crawley, 2002; Torchin et al., 2006); causing native-range-only models to underestimate a species' Grinnellian niche in new areas. This illustrates a known limitation of ecological niche modeling caused by species' interactions that are not included in the model.

Within the above outlined assumptions, fundamental decisions need to be made to define a niche modeling strategy. Most importantly, two data types may be used to delineate the niche in environmental space, restricting the choice for modeling algorithms: environmental data from both known occurrences and confirmed absences (presence/absence modeling) or information based on occurrence records only (presence-only modeling). It has been suggested that whenever reliable absence data are available, the use of presence/absence techniques should be preferred (Brotons et al., 2004). However, although seaweeds can be more easily sampled than many animal groups in terms of localization, capture and legal issues, there is often a paucity of occurrence data for (subtidal) species because many algae are small and inconspicuous and exhaustive samplings are rather hard to achieve by SCUBA diving. Additionally, since many species are characterized by partly microscopic life stages, reliable absence data cannot be obtained for most macroalgal species. For these reasons, the choice for Maxent as the best suited modeling technique throughout this thesis was warranted, based on its demonstrated high performance as a presence-only modeling technique (Elith et al., 2006), even with low sample sizes (Pearson et al., 2007).

A second set of strategic modeling choices relate to the number of raster layers describing the environmental space (i.e., variable selection) and the functions describing the relations between species and these variables and their interactions (feature selection). Including too large a number of variables may have detrimental effects on the models due to multicollinearity in the datasets, which may hamper ecological inferences. Including too complex features may lead to overfitting of the models, which means that models are matched too closely to input data in environmental space, reducing predictive performance of the models (Phillips & Dudik, 2008).



While sample numbers were generally too low to allow complex feature fitting for the case studies presented in Chapter 2 and 3, features selection was set to allow only the two least complex relations for the model presented in Chapter 4 (whereas Maxent automatically allows the most complex features to be fitted based on the number of occurrence records by default). Variable selection was carried out in Chapter 2 using a similar approach as described by Pearson et al. (2007), Rissler & Apodaca (2007) and Saatchi et al. (2008) to obtain modeled variable contributions and response curves. A correlation analysis was performed and the least correlated variables were selected, retaining at least one variable per macroecological dimension in the environmental dataset to adequately explain ecological patterns. This approach reduces data redundancy effectively (although not completely) and ensures that input variables, and hence output variable contributions and response curves, are still straightforward to interpret. An alternative approach was proposed in Chapter 2 and 3 by performing PCA and using the principal component maps as input in the modeling step. This was done because the use of principal components virtually eliminates data redundancy, needed to obtain the most accurate macroecological suitability maps to compare to known distribution patterns, without referring to variable contributions and response curves. Chapters 4 and 5 present a novel approach in Maxent model selection by running preliminary models in a forward stepwise variable selection procedure until predictive performance stops increasing. Performance was in this case measured by the area under the curve (AUC) of receiver operating characteristics (ROC) for the test data.

Thirdly, measures need to be taken to deal with small sample sizes, preferential sampling and spatial autocorrelation. The latter often occurs for instance where remote coastlines are hardly accessible and many samples hence tend to be aggregated around few physical locations. Preferential sampling (or sample selection bias) arises when consistently only parts of the environment are represented in the biotic samples by sampling for instance along roads, or, in the case of this thesis, only along coastlines. Both issues are exacerbated when using small sample sizes, and are caused by insufficient sampling of the data in environmental space as opposed to sampling along a line or gradient in geographical space or by an over-representation of parts of the environmental space by the sample set.

Different solutions have been proposed to cope with small sample sizes and spatial autocorrelation. While we minimized the risk of erratic models caused by small or spatially autocorrelated samples in Chapter 3 by rejecting biotic datasets containing less than 10 records or obviously suffering from autocorrelation, we took several precautionary measures for the small sample set available in Chapter 2. Similar to Raxworthy et al. (2007), we manually divided the sample set in training and test data to prevent the formation of data clusters in one part of the world. Furthermore, we constructed several replicate runs by cross-validation in order to evaluate statistical properties of the resulting models and AUC values (Pearson et al., 2007). Recently, several authors have identified the issue of sampling bias or preferential sampling for niche modeling studies in the terrestrial and marine environment (Phillips et al., 2009; Elith et al., 2010b; Dambach & Rödder, 2011). In Chapter 1 and 5, we proposed a first step towards remedying preferential sampling bias in the marine environment by biasing the background sample in a similar way as the biotic input data, i.e. restricting the random sample of 10000 background points to the narrow coastal zones in which the macroalgal species under scrutiny may occur and might be sampled. Hence, we provided the modeling algorithm with a relevant background sample against which to compare the observed environmental data. This approach seems to lower the AUC value slightly, but in our opinion this reflects more realistic model outcomes suffering from less overfitting, rather than worse model performance.

Lastly, model evaluation is an important consideration in the design of modeling studies. To date, while AUC-ROC analysis is by far the most widely used in terrestrial and marine niche modeling, some drawbacks exist to this approach. There are no fixed thresholds from which to determine that a given AUC value represents poor, acceptable or very good model performance, other than that values lower than 0.5 are indicative of no better than random prediction. There is also no fixed highest possible value due to the presence-only nature of Maxent models (though maximum values will be very close to 1 in practice). Additionally, AUC values for similar models will be calculated differently for various modeling algorithms and cannot be compared for different approaches (Lobo et al., 2008). There is as of yet no consensus on the best alternative for AUC analysis and it continues to be the most widespread, although alternatives have been

suggested. In Chapter 3, we applied the partial AUC suggested by Peterson et al. (2008) to enable comparison with replicated model runs using other algorithms. Overall, we evaluated model performance internally based on replicate runs using, with or without some sort of resampling. These allow to calculate an average and standard deviation of the test AUC, which offers a good estimate of overfitting and general predictive performance when compared to the training AUC (where a test AUC exceeding the training AUC would mean excellent predictive performance and a training AUC exceeding the test AUC would indicate overfitting).

In Chapter 6 we suggested a new subpixel classification method by translating bottom type fractions per pixel to a multi-species distribution modeling framework, in which training localities for bottom types are defined as input species occurrence localities and spectral bands were considered as environmental layers. Since spectral characteristics change immediately with changing bottom types, the assumption of equilibrium was valid, as is the semi-reciprocal relationship between geographical and spectral space. In that sense, each bottom type is defined by its own Grinnellian niche in spectral space. The equivalency still holds for fundamental versus realized niches, where a bottom type could be covered from its usual spectral signature by a “competitor”. The choice for the presence-only Maxent modeling algorithm was warranted based on its high predictive performance in comparison with other algorithms including presence/absence techniques (Elith et al., 2006) and based on its tendency to extrapolate better to unsampled regions (Bentlage et al., 2009). Next to modeling the different bottom types (and subsequently hardening the different models to a single classification map), this approach is ideally suited to map the distribution probability of certain bottom types of interest since there is no need to train the model on all bottom types, which is usually necessary for other supervised classification algorithms. The proposed approach has now also been tested on a heterogeneous (urban) environment in aerial photography (Li & Guo, 2010).

## INVESTIGATING ENVIRONMENTAL DATA INTEGRATION AND EFFECTS ON MODELING

Rather than constituting a goal in itself, ecological niche modeling or species distribution modeling is mostly used as a means in the framework of other goals (Franklin, 2009). Because both species occurrence data and environmental data are generally more easily available for the terrestrial realm, both the number of studies and the range of applications for which niche modeling has been used exceeds that of marine studies to date. In Chapter 2 and 3, we adopted a niche modeling approach in applications for which terrestrial equivalents were already published, for the first time in a marine environment. For example, niche modeling in an evolutionary context has been investigated by, among others, Rissler & Apodaca (2007) and Raxworthy et al. (2007). Niche modeling based on few occurrence records as a means to delineate other probable areas of occurrence has been proposed by Pearson et al. (2007), and Saatchi et al. (2008) inferred ecological information on species based on response curves generated by niche modeling. Although standardized global environmental datasets such as WorldClim (Hijmans et al., 2005) were already available for the terrestrial modeling studies mentioned above, only Raxworthy et al. (2007) limited their variable choice to this particular dataset. Rissler & Apodaca (2007) and Pearson et al. (2007) both included climatic and geophysical layers from other sources, while Saatchi et al. (2008) compiled custom remote sensing-based climatic and land cover data. Chapter 2 and 3 took a similar approach to the latter study, by downloading global MODIS climatology data and deriving biologically meaningful variables. Hence, assembling a global marine environmental dataset proved time-consuming in these chapters, and the need for a pre-packaged, uniform environmental dataset was clearly indicated. Dambach & Rödder (2011) and Robinson et al. (2011) later also suggested the need for such a dataset to facilitate niche modeling in many applied studies.

Chapter 4 described the compilation of Bio-ORACLE, a set of biologically relevant raster layers indicative of climate averages, extremes and seasonality for variables thought to play a role in the distribution of many marine organisms. Raster layers were compiled from various sources, both remote sensing-based and resulting from interpolation of *in situ*

measurements. Remote sensing-based variables suffering from cloud cover or low coastal resolution were also interpolated using advanced algorithms, with the availability of uncertainty maps. Most importantly, variables were packaged uniformly with an equal landmask applied to all variables and data are disseminated in an equidistant and an equal area projection to meet different needs depending on the study concept. However, the *Codium fragile* case study presented in Chapter 4 as well as terrestrial studies using pre-packaged datasets (Rissler & Apodaca, 2007; Pearson et al., 2007) clearly indicated that careful selection of the available variables and, depending on the studied species, expanding the variable set with data from other sources, should still be considered in order to reduce data redundancy and to capture the relevant ecophysiological driving variables.

The resolution of approximately 9 km of the Bio-ORACLE dataset reflects both the native resolution at which most of the global level-3 mapped remote sensing data products are made available and the best meaningful resolution to which *in situ* measurements and masked pixels can be interpolated without artificial data inflation through internal resampling. At the same time, it adequately captures macroecological gradients in nutrients, sea surface temperature etc. While designed at the global scale, this dataset can easily be cropped to accommodate regional studies. On the other hand, this resolution does not allow for a meaningful inclusion of a substrate layer, a variable which often determines distribution at the regional level. Additionally, in Chapter 5 we demonstrated that 9 km coastal pixels can hide fine-scale patterns in variables other than substrate, for instance in chlorophyll and sea surface temperature which are generally assumed to vary over coarse scales, indicating among others sheltered versus exposed bays. While it is true that coarse resolution coastal pixels are biased towards open-sea average values, this seems to be of major concern only in highly variable environments. Therefore, we presented a methodology to incorporate high resolution Landsat 7 ETM+ imagery to derive environmental layers suitable for distribution modeling. While as of yet time-consuming, this approach allows capturing local processes for an entire region. The use of Landsat-based environmental layers representing Grinnellian factors similar to the Bio-ORACLE dataset including substrate showed that these factors also vary greatly on a very fine resolution in the marine environment due to differences in, among others, exposure to waves and upwelling in small bays.

The fine scale at which these factors can work is somewhat overlooked in literature, where Grinnellian variables are mostly explained to vary at coarser scales while Eltonian factors are thought to work at very fine scales (Colwell & Rangel, 2009; Soberón & Nakamura, 2009).

## **MULTI-SENSOR MAPPING**

Satellite imagery-based mapping has as of yet rarely been applied to monitoring specific macroalgal communities, except for algal cover on tropical coral reefs. Since the water column on tropical coral reefs is typically clear, these studies may benefit from very high-resolution multispectral imagery. However, most abundant and diverse seaweed communities are found in heterogeneous patches on solitary coral communities or rock outcrops in subtropical to (cold-)temperate environments, where imagery would benefit from both very high spectral and very high spatial resolution. With current satellite sensors such as the ones used in Chapter 6, mixed pixels are of major concern. We proposed a multi-sensor approach in which a multispectral resolution sensor with a similar spatial resolution contributes to data analyses from a superspectral sensor by providing SWIR data, useful in an index to detect surfacing and floating algae. However, both sensors discussed in Chapter 6 still yield mixed pixels. Besides adopting a sub-pixel modeling algorithm such as the one presented, it is also useful to consider incorporating imagery from a sensor with sub-meter resolution. With spaceborne sensors limited to a panchromatic band in sub-meter imagery, aerial acquisition techniques can yield sub-meter multispectral or hyperspectral. The tools demonstrated in Chapter 7 provide a promising addition to the current set of airplane-based technologies, although it is foremost intended for low-budget or remote studies and its use is confined to limited areas and intertidal communities (and shallow waters in case of calm, sheltered conditions). In this type of imagery, pixels are no longer mixed but instead are many times smaller than the objects of interest, necessitating an image segmentation or object-oriented classification approach. Although Chapter 7 presents a separate study, the main strength of this type of imagery would arguably be its use in combination with other remote sensing products on coarser resolution of the same area for which the data characteristics would be complementary.



## CONCLUSION AND PERSPECTIVES

A growing body of spatially explicit information and processing algorithms has become available to the biological community in recent years. Many applications involving these data and techniques have been designed and tested in the terrestrial environment, but marine and especially phycological applications remained scattered. While benthic marine algae make a good pragmatic choice to test and expand spatially explicit data processing techniques for the marine environment, they are also of key importance in marine food webs and patterns in space and time, and hence need to be mapped and monitored. The case studies presented in this thesis offer insights in real-life issues with macroalgal mapping and monitoring and are used to demonstrate the feasibility and motivated practices of these techniques in answering biological questions.

In agreement with a growing number of published case studies on distribution modeling, the results from this thesis indicate the need to consider model concept, settings and choices carefully for each species, environment and application. We assert that there is no “one way fits all” in modeling distributions and niches, not even when the modeling efforts use the same algorithm and when applications are similar. Preliminary models should be run to test for various effects, and the results from this thesis may be used to adjust models accordingly to achieve higher predictive performance. The biggest challenges remain in model evaluation metrics other than AUC and their applied use in variable and feature selection through the use of preliminary model runs in stepwise selection procedures. In this way, several model parameter settings should be investigated using simulation studies to open the “black box” of certain complex modeling algorithms, including Maxent. For instance, the setting of the regularization multiplier was until recently poorly understood, while having an important function in preventing overfitting. Warren & Seifert (2011) demonstrated an important first step towards this by using AIC-based model selection to obtain the best degree of model complexity (regularization).

With continually increasing availability of environmental data, it is important to note that Bio-ORACLE, although useful in its present form, is a work in progress which can be updated with future availability of new global climate data, mainly from remote sensing. For instance, salinity is



currently provided as an interpolated map, but can be expected to be updated by monthly averaged global level 3 or 4 remotely sensed data products from the SMOS or Aquarius missions. Other new data sources may include the derivation of variables which cannot be directly measured using remote sensing but which can be modeled based on remote sensing information. For example, global raster layers of nitrate, now only available through interpolation of *in situ* measurements in Bio-ORACLE, could be calculated based on the correlation with SST and chlorophyll-a. Hence, the database may be updated with such derived layers in cases where error maps indicate that these models would outperform spatial interpolation using the DIVA algorithm. We also foresee bathymetrical information to be integrated for the derivation of environmental data layers in 3 dimensions on a short term. Coupling models based on these environmental data with mechanistic models will be an important field of research.

Landsat-derived environmental variables could be further explored for regional niche modeling studies. Our approach suffered from drawbacks in the assembling methods, but nevertheless offered promising results on a combination of extent and resolution that is new to our knowledge. Since contrasting processes can be captured by these variables compared to Bio-ORACLE, we could see this spatial scale as a useful basis to add spatially explicit Eltonian data layers, if feasible at all, since this type of factors act rather at fine scales.

Lastly, multi-scale approaches are increasingly important to capture the full range of processes and patterns at different spatial and temporal scales in the study of macroalgal communities. Very high resolution aerial photography could be used as a tool to validate and refine seasonal mapping efforts based on spaceborne sensors, which in turn could yield useful environmental data for distribution modeling efforts.

## REFERENCES

- Ader RR (1982) A Geographic Information System for addressing issues in the coastal zone. *Computers Environment and Urban Systems* **7**, 233-243.
- Adey WH & Steneck RS (2001) Thermogeography over time creates biogeographic regions: a temperature/space/time-integrated model and an abundance-weighted test for benthic marine algae. *Journal of Phycology* **37**, 677-698.
- Ahn Y, Shanmugam P, Ryu J & Jeong J (2006) Satellite detection of harmful algal bloom occurrences in Korean waters. *Harmful Algae* **5**(2), 213-231.
- Alexander D (2008) Remote sensing and the coast: development of advanced techniques to map nuisance macro-algae in estuaries. *New Zealand Geographer* **64**(2), 157-161.
- Ambrose RF & Smith J (2004) *Restoring rocky intertidal habitats in Santa Monica Bay*. Report to Santa Monica Bay Restoration Commission, Los Angeles.
- Amsterdam R, Andresen E & Lipton H (1972) Geographic Information Systems in the U.S. - an overview. *Afips Conference Proceedings* **40**, 511-22.
- Andréfouët S, Zubia M & Payri C (2004) Mapping and biomass estimation of the invasive brown algae *Turbinaria ornata* (Turner) J. Agardh and *Sargassum mangarevense* (Grunow) Setchell on heterogeneous Tahitian coral reefs using 4-meter resolution Ikonos satellite data. *Coral Reefs* **23**, 26-38.
- Araujo MB & Guisan A (2004) Five (or so) challenges for species distribution modelling. Workshop on Generalized Regression Analyses and Spatial Predictions, 1677-1688. Blackwell Publishing, Riederalp, Switzerland.
- Archibald D (1897) *The story of Earth's atmosphere*. D. Appleton & Co, New York (USA), 215pp.
- Austin MP (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling* **157**, 101-118.
- Bailey WH (1963) Remote sensing of the environment. *Annals of the Association of American Geographers* **53**, 577-578.
- Beiko R G, Keith J M, Harlow T J & Ragan MA (2006) Searching for convergence in phylogenetic Markov chain Monte Carlo. *Systematic Biology* **55**, 553-565.
- Belsher T, Loubersac L & Belbeoch G (1985) Remote sensing and mapping. In: MM Littler & DS Littler (eds.) *Ecological field methods: macroalgae*. Phycological handbook **4**, 177-197. Cambridge University Press, Cambridge.
- Bentlage B, Peterson AT, Cartwright P (2009) Inferring distributions of chirodropid box-jellyfishes (Cnidaria: Cubozoa) in geographic and ecological space using ecological niche modeling. *Marine Ecology Progress Series* **384**, 121-133.
- Beyer HL (2004) *Hawth's analysis tools for ArcGIS*. Available online at <http://www.spatial ecology.com/htools>.

## References

- Boyer TP, Antonov JI, Baranova OK, Garcia HE, Johnson DR, Locarnini RA, Mishonov AV, O'Brien TD, Seidov D, Smolyar IV & Zweng MM (2009) *World Ocean Database 2009*. U.S. Gov. Printing Office, Washington D.C.
- Bradley BA & Fleishman E (2008) Can remote sensing of land cover improve species distribution modelling? *Journal of Biogeography* **35**(7), 1158-1159.
- Brasseur P, Beckers JM, Brankart JM & Schoenauen R (1996) Seasonal temperature and salinity fields in the Mediterranean Sea: climatological analyses of a historical data set. *Deep Sea Research Part I: Oceanographic Research Papers* **43**, 159-192.
- Brasseur PP & Haus JA (1991) Application of a 3-D variational inverse model to the analysis of ecohydrodynamic data in the Northern Bering and Southern Chukchi Seas. *Journal of Marine Systems* **1**, 383-401.
- Breeman AM, Oh YS, Hwang MS & van den Hoek C (2002) Evolution of temperature responses in the *Cladophora vagabunda* complex and the *C. albida/sericea* complex (Chlorophyta). *European Journal of Phycology* **37**, 45-58.
- Brönnimann O, Treier UA, Müller-Schärer H, Thuiller W, Peterson AT & Guisan A (2007) Evidence of climatic niche shift during biological invasion. *Ecology Letters* **10**(8), 701-709.
- Burgess SC (2006) Algal blooms on coral reefs with low anthropogenic impact in the Great Barrier Reef. *Coral Reefs* **25**, 390.
- Capinha C, Leung B & Anastácio P (2011) Predicting worldwide invasiveness for four major problematic decapods: an evaluation of using different calibration sets. *Ecography* **34**, 448-459.
- Capolsini P, Andréfouët S, Rion C & Payri C (2003). A comparison of Landsat ETM+, SPOT HRV, Ikonos, ASTER and Airborne MASTER Data for coral reef habitat mapping in South Pacific Islands. *Canadian Journal of Remote Sensing* **29**(2), 187-200.
- Chander G, Markham BL, Helder DL & Ali E (2009). Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment* **113**(5), 893-903.
- Chavez PS (1996). Image-based atmospheric corrections revisited and improved. *Photogrammetric Engineering and Remote Sensing* **62**(9), 1025-1036.
- Chown SL, Sinclair BJ, Leinaas HP & Gaston KJ (2004) Hemispheric asymmetries in biodiversity - A serious matter for ecology. *PLoS Biology* **2**, 1701-1707.
- Chust G, Galparsoro I, Borja Á, Franco J & Uriarte A (2008) Coastal and estuarine habitat mapping, using LIDAR height and intensity and multi-spectral imagery. *Estuarine, Coastal and Shelf Science* **78**(4), 633-643.
- Clarke A (2009) Temperature and marine macroecology. In: JD Witman & K Roy (eds.) *Marine Macroecology*, 250-278. The University of Chicago Press, Chicago.
- Coates AG & Obando JA (1996) The geologic evidence of the Central American Isthmus. In: BC Jackson, AF Budd & AG Coates (eds.) *Evolution and environment in tropical America*, 21-56. University of Chicago Press, Chicago.
- Coles S (1996) Corals of Oman. R. Keech Publ., Thorns, Hawes, UK. 106pp.

- Collins LS, Budd AF & Coates AG (1996) Earliest evolution associated with closure of the Tropical American Seaway. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 6069-6072.
- Colwell RK & Rangel TF (2009) Hutchinson's duality: the once and future niche. *Proceedings of the National Academy of Sciences of the United States of America* **106**(Suppl. 2), 19651-19658.
- Compton TJ, Leathwick JR & Inglis GJ (2010) Thermogeography predicts the potential global range of the invasive European green crab *Carcinus maenas*. *Diversity and Distributions* **16**, 243-255.
- Condie SA, Herzfeld M, Margvelashvili N & Andrewartha JR (2009) Modeling the physical and biogeochemical response of a marine shelf system to a tropical cyclone. *Geophysical Research Letters* **36** L22603, DOI: 10.1029/2009GL039563.
- Coppejans E, Leliaert F, Dargent O & De Clerck O (2001) Marine green algae (Chlorophyceae) from the north coast of Papua New Guinea. *Cryptogamie Algologie* **22**, 375-443.
- Costello MJ, Coll M, Danovaro R, Halpin P, Ojaveer H & Miloslavich P (2010) A census of marine biodiversity knowledge, resources, and future challenges. *PLoS One* **5**(8) e12110, DOI: 10.1371/journal.pone.0012110.
- Costello MJ, Stocks K, Zhang Y, Grassle JF & Fautin DG (2007) *About the Ocean Biogeographic Information System*. Available online at <http://www.iobis.org>.
- Cribb AB (1984) Algal vegetation of the Capricornia section, Great Barrier Reef Marine Park. In: WT Ward & P Saenger (eds.) *The Capricornia section of the Great Barrier Reef: past, present and future*, 79-86. Royal Society of Queensland and Australian Coral Reef Society, Brisbane (Australia).
- Daly C (2006) Guidelines for assessing the suitability of spatial climate data sets. *International Journal of Climatology* **26**, 707-721.
- Davies AJ, Wisshak M, Orr JC & Roberts JM (2008) Predicting suitable habitat for the cold-water coral *Lophelia pertusa* (Scleractinia). *Deep Sea Research Part I: Oceanographic Research Papers* **55**, 1048 - 1062.
- De Clerck O, Coppejans E (1996) Marine algae of the Jubail Marine Wildlife Sanctuary, Saudi Arabia. In: F Krupp, AH Abuzinada & IA Nader (eds.) *A marine wildlife sanctuary for the Arabian Gulf – Environmental research and conservation following the 1991 Gulf war oil spill*, 199-289. National Commission for Wildlife Conservation and Development, Riyadh and Senckenberg Research Institute, Frankfurt a.M., Germany.
- De Oliveira E, Populus J & Guillaumont B (2006) Predictive modeling of coastal habitats using remote sensing data and fuzzy logic: a case for seaweed in Brittany (France). *EARSeL eProceedings* **5**(2), 208-223.
- de Rivera CE, Steves BP, Fofonoff PW, Hines AH & Ruiz GM (2011) Potential for high-latitude marine invasions along western North America. *Diversity and Distributions*, DOI: 10.1111/j.1472-4642.2011.00790.x.
- Degraer S, Verfaillie E, Willems W, Adriaens E, Vincs M & Van Lancker V (2008) Habitat suitability modelling as a mapping tool for macrobenthic communities: an

## References

- example from the Belgian part of the North Sea. *Continental Shelf Research* **28**(3), 369-379.
- Diaz-Pulido G, Chin A, Davidson J, McCook LJ (2007) Cyclone promotes rapid colonisation of benthic diatoms in the Great Barrier Reef. *Coral Reefs* **26**, 787.
- Dodd M (2011) Where are my quadrats? Positional accuracy in fieldwork. *Methods in Ecology and Evolution*, DOI: 10.1111/j.2041-210X.2011.00118.x.
- Doney SC, Fabry VJ, Feely RA & Kleypas JA (2009) Ocean acidification: the other CO<sub>2</sub> problem. *Annual Review of Marine Science* **1**, 169-192.
- Dragastan ON & Herbig HG (2007) *Halimeda* (green siphonous algae) from the Paleogene of Morocco – Taxonomy, phylogeny and paleoenvironment. *Micropaleontology* **53**, 1-72.
- Drew EA (1983) *Halimeda* biomass, growth rates and sediment generation on reefs in the central Great Barrier Reef province. *Coral Reefs* **2**, 101-110.
- Dring MJ, Wagner A, Boeskov J & Luning K (1996) Sensitivity of intertidal and subtidal red algae to UVA and UVB radiation, as monitored by chlorophyll fluorescence measurements: influence of collection depth and season, and length of irradiation. *European Journal of Phycology* **31**, 293-302.
- Duan H, Zhang Y, Zhang B, Song K & Wang Z (2007) Assessment of chlorophyll-a concentration and trophic state for lake Chagan using Landsat TM and field spectral data. *Environmental Monitoring and Assessment* **129**, 295-308.
- Dvorak D & Dvorak E (1998) Kite aerial photography of Easter Island's archaeological sites. In: C Stevenson, G Lee & F Morin (eds.) *Easter Island Foundation occasional paper*. Easter Island Foundation, Albuquerque, NM (USA), 4<sup>th</sup> edition, 223pp.
- Egan WG & Hair ME (1971) Automated delineation of wetlands in photographic remote sensing. *Proceedings of the 7th International Symposium on Remote Sensing of Environment*, Volume III, 2231-2251. University of Michigan, Ann Arbor, Michigan (USA).
- Egerod L (1975) Marine algae of the Andaman Sea coast of Thailand: Chlorophyceae. *Botanica Marina* **18**, 41-66.
- Elith J & Leathwick JR (2009) Species Distribution Models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* **40**(1), 677-697.
- Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberon J, Williams S, Wisz MS & Zimmermann NE (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**, 129-151.
- Elith J, Phillips SJ, Hastie T, Dudik M, Chee YE & Yates CJ (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* **17**, 43-57.

- Famà P, Wysor B, Kooistra W & Zuccarello GC (2002) Molecular phylogeny of the genus *Caulerpa* (Caulerpaceae, Chlorophyta) inferred from chloroplast *tufA* gene. *Journal of Phycology* **38**, 1040-1050.
- Fautin DG & Buddemeier RW (2008) *Biogeoinformatics of the Hexacorals*. Available online at <http://www.kgs.ku.edu/Hexacoral>.
- Feldman GC & McClain CR (2010) *Ocean Color Web*. In: N Kuring & SW Bailey (eds.) NASA Goddard Space Flight Center. Available online at <http://oceancolor.gsfc.nasa.gov>.
- Ferreira B, Garcia M, Jupp BP & Al-Kiyumi A (2006) Diet of the Green Turtle (*Chelonia mydas*) at Ra's Al Hadd, Sultanate of Oman. *Chelonian Conservation and Biology* **5**(1), 141-146.
- Ferrier S & Guisan A (2006) Spatial modeling of biodiversity at the community level. *Journal of Applied Ecology* **43**, 393-404.
- Ferwerda JG, de Leeuw J, Atzberger C & Vekerdy Z (2007) Satellite-based monitoring of tropical seagrass vegetation: current techniques and future developments. *Hydrobiologia* **591**, 59-71.
- Foster KA, Foster JG, Tourenq C, Shuriqui MK (2008) Spatial and temporal recovery patterns of coral reefs within the Gulf of Oman (United Arab Emirates) following the 2007 cyclone disturbance. *Proceedings of the 11<sup>th</sup> International Coral Reef Symposium* **1**, 731-734.
- Franklin J (2009) *Mapping species distributions: spatial inference and prediction*. Cambridge University Press, Cambridge, 338pp.
- Freeman AS, Chiu CP & Smart JH (2006) *The Office of Naval Research's Worldwide Ocean Optics Database (WOOD v4.5i), user's guide*. The Johns Hopkins University Applied Physics Laboratory, USA, available online at <http://wood.jhuapl.edu>.
- Freile D, Milliman JD & Hillis L (1995) Leeward bank margin *Halimeda* meadows and draperies and their sedimentary importance on the western Great Bahama bank slope. *Coral Reefs* **14**, 27-33.
- Gagnon P, Scheibling RE, Jones W & Tully D (2008) The role of digital bathymetry in mapping shallow marine vegetation from hyperspectral image data. *International Journal of Remote Sensing* **29**, 879-904.
- Gerard B, Buerkert A, Hiernaux P & Marschner H (1997) Non-destructive measurement of plant growth and nitrogen status of pearl millet with low-altitude aerial photography. *Soil Science and Plant Nutrition* **43**, 993-998.
- Gillespie TW, Foody GM, Rocchini D, Giorgi AP & Saatchi S (2008) Measuring and modelling biodiversity from space. *Progress in Physical Geography* **32**(2), 203-221.
- Gogina M & Zettler ML (2010) Diversity and distribution of benthic macrofauna in the Baltic Sea Data inventory and its use for species distribution modelling and prediction. *Journal of Sea Research* **64**, 313-321.
- Gomez-Chova L, Alonso L, Guanter L, Camps-Valls G, Calpe J & Moreno J (2008) Correction of systematic spatial noise in push-broom hyperspectral sensors: application to CHRIS/PROBA images. *Applied Optics* **47**(28), F46-F60.

## References

- Graham CH, Moritz C & Williams SE (2006) Habitat history improves prediction of biodiversity in rainforest fauna. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 632-636.
- Graham CH, Ron SR, Santos JC, Schneider CJ & Moritz C (2004) Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution* **58**, 1781-1793.
- Graham MH, Kinlan BP, Druehl LD, Garske LE & Banks S (2007) Deep-water kelp refugia as potential hotspots of tropical marine diversity and productivity. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 16576-16580.
- Green EP, Mumby PJ, Edwards AJ & Clark CD (2000) Remote sensing handbook for tropical coastal management. In: AJ Edwards (ed.) *Coastal Management Sourcebooks* **3**. UNESCO, Paris, x+360pp.
- Gregg WW & Casey NW (2007) Sampling biases in MODIS and SeaWiFS ocean chlorophyll data. *Remote Sensing of Environment* **111**, 25-35.
- Guanter L (2006). Spectral calibration of hyperspectral imagery using atmospheric absorption features. *Applied Optics* **45**(10), 2360-2370.
- Guanter L, Alonso L & Moreno J (2005). A method for the surface reflectance retrieval from PROBA/CHRIS data over land: application to ESA SPARC campaigns. *IEEE Transactions on Geoscience and Remote Sensing* **43**(12), 2908-2917.
- Guillaumont B, Bajjouk T & Talec P (1997) Seaweed and remote sensing: a critical review of sensors and data processing. In: FE Round & DJ Chapman (eds.) *Progress in Phycological Research* **12**, 213-282. Biopress Ltd.
- Guillaumont B, Callens L & Dion P (1993) Spatial distribution and quantification of *Fucus* species and *Ascophyllum nodosum* beds in intertidal zones using SPOT imagery. *Hydrobiologia* **260/261**, 297-305.
- Guinotte JM, Bartley JD, Iqbal A, Fautin DG & Buddemeier RW (2006) Modeling habitat distribution from organism occurrences and environmental data: case study using anemone fishes and their sea anemone hosts. *Marine Ecology-Progress Series* **316**, 269-283.
- Guiry MD & Guiry GM (2008) *AlgaeBase*. National University of Ireland, Galway, Ireland, available online at <http://www.algaebase.org>.
- Guisan A & Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters* **8**, 993-1009.
- Guisan A & Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecological Modelling* **135**, 147-186.
- Guisan A, Graham CH, Elith J & Huettmann F (2007) Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions* **13**(3), 332-340.
- Haas A, el-Zibdah M & Wild C (2009) Seasonal monitoring of coral-algae interactions in fringing reefs of the Gulf of Aqaba, Northern Red Sea. *Coral Reefs* **29**(1), 93-103.
- Halpern BS, Walbridge S, Selkoe KA, Kappel CV, Micheli F, D'agrosa C, Bruno JF, Casey KS, Ebert C, Fox HE, Fujita R, Heinemann D, Lenihan HS, Madin EMP,

- Perry MT, Selig ER, Spalding M, Steneck R & Watson R (2008) A global map of human impact on marine ecosystems. *Science* **319**, 948-952.
- Harmon LJ, Weir JT, Brock CD, Glor RE & Challenger W (2008) GEIGER: investigating evolutionary radiations. *Bioinformatics* **24**, 129-131.
- Hebert PDN, Penton EH, Burns JM, Janzen DH & Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 14812-14817.
- Henne KD, Schnetter R (1999) Revision of the *Pseudobryopsis/Trichosolen* complex (Bryopsidales, Chlorophyta) based on features of gametangial behavior and chloroplasts. *Phycologia* **38**, 114-127.
- Hijmans RJ, Cameron SE, Parra JL, Jones PG & Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**, 1965-1978.
- Hillis-Colinvaux L (1980) Ecology and taxonomy of *Halimeda*: primary producer of coral reefs. *Advances in Marine Biology* **17**, 1-327.
- Hochberg EJ & Atkinson MJ (2003) Capabilities of remote sensors to classify coral, algae, and sand as pure and mixed spectra. *Remote Sensing of Environment* **85**(2), 174-189.
- Holder M & Lewis PO (2003) Phylogeny estimation: traditional and Bayesian approaches. *Nature Reviews Genetics* **4**, 275-284.
- Hu C (2009) A novel ocean color index to detect floating algae in the global oceans. *Remote Sensing of Environment* **113**(10), 2118-2129.
- Hunt ER & Linden DS (2009) New method for acquiring digital color-infrared photographs to monitor vegetation. *Precision Agriculture*, 9pp.
- Hutchinson GE (1957) Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology* **22** (2), 415-427.
- Jarecke P, Barry P, Pearlman J & Markham B (2001) Aggregation of Hyperion hyperspectral spectral bands into Landsat-7 ETM+ spectral bands. *IGARSS 2001 Proceedings: Scanning the Present and Resolving the Future* **6**, 2822-2824.
- Jensen PR, Gibson RA, Littler MM & Littler DS (1985) Photosynthesis and calcification in four deep-water *Halimeda* species (Chlorophyceae, Caulerpales). *Deep Sea Research Part I: Oceanographic Research Papers* **32**, 451-464.
- Johnson DR, Boyer TP, Garcia HE, Locarnini RA, Baranova OK & Zweng MM (2009) *World Ocean Database 2009 Documentation*. In: S Levitus (ed.) NODC Internal Report **20**, 175pp. NOAA Printing Office, Washington D.C.
- Jupp BP (2007) Report on destruction of corals on Reef Ball site, Fahal Island, Muscat. *Rapid assessment of impacts from cyclone Gonu*, Ministry of Environment and Climate Affairs, Muscat, Oman.
- Karafistan A, Martin JM, Rixen M & Beckers JM (2002) Space and time distributions of phosphate in the Mediterranean Sea. *Deep Sea Research Part I: Oceanographic Research Papers* **49**, 67-82.



## References

- Karentz D & Bosch I (2001) Influence of ozone-related increases in ultraviolet radiation on antarctic marine organisms. *American Zoologist* **41**, 3-16.
- Kaschner K, Ready JS, Agbayani E, Rius J, Kesner-Reyes K, Eastwood PD, South AB, Kullander SO, Rees T, Close CH, Watson R, Pauly D & Froese R (2008a) AquaMaps: environmental dataset: Half-degree Cells Authority File (HCAF). Available online at <http://www.aquamaps/data.org>, Version 10/2008.
- Kaschner K, Ready JS, Agbayani E, Rius J, Kesner-Reyes K, Eastwood PD, South AB, Kullander SO, Rees T, Close CH, Watson R, Pauly D & Froese R (2008b) AquaMaps: Predicted range maps for aquatic species. Available online at <http://www.aquamaps.org>, Version 10/2008.
- Kaschner K, Watson R, Trites AW & Pauly D (2006) Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. *Marine Ecology-Progress Series* **316**, 285-310.
- Keane RM & Crawley MJ (2002) Exotic plant invasions and the enemy release hypothesis. *Trends in Ecology and Evolution* **16**, 199-204.
- Kidd DM & Ritchie MG (2006) Phylogeographic information systems: putting the geography into phylogeography. *Journal of Biogeography* **33**, 1851-1865.
- Kieleck C, Bousquet B, Le Brun G, Cariou J & Lotrian J (2001) Laser induced fluorescence imaging: application to groups of macroalgae identification. *Journal of Physics D: Applied Physics* **34**, 2561-2571.
- Knouft JH, Losos JB, Glor RE & Kolbe JJ (2006) Phylogenetic analysis of the evolution of the niche in lizards of the *Anolis sagrei* group. *Ecology* **87**, S29-S38.
- Knowlton N & Weigt LA (1998) New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society of London Series B - Biological Sciences* **265**, 2257-2263.
- Kooistra W & Verbruggen H (2005) Genetic patterns in the calcified tropical seaweeds *Halimeda opuntia*, *H. distorta*, *H. hederacea* and *H. minima* (Bryopsidales, Chlorophyta) provide insights in species boundaries and inter-oceanic dispersal. *Journal of Phycology* **41**, 177-187.
- Kooistra W, Coppejans EGG & Payri C (2002) Molecular systematics, historical ecology, and phylogeography of *Halimeda* (Bryopsidales). *Molecular Phylogenetics and Evolution* **24**, 121-138.
- Kozak KH, Graham CH & Wiens JJ (2008) Integrating GIS-based environmental data into evolutionary biology. *Trends in Ecology and Evolution* **23**, 141-148.
- Kulkarni MA, Desrochers RE & Kerr JT (2010) High resolution niche models of malaria vectors in northern Tanzania: a new capacity to predict malaria risk? *PloS one* **5**(2) e9396, DOI: 10.1371/journal.pone.0009396.
- Lahoz-Monfort JJ, Guillera-Aroita G, Milner-Gulland EJ, Young RP & Nicholson E (2010) Satellite imagery as a single source of predictor variables for habitat suitability modelling: how Landsat can inform the conservation of a critically endangered lemur. *Journal of Applied Ecology* **47**(5), 1094-1102.

- Laliberte AS, Herrick JE, Rango A & Winters C (2010) Acquisition, orthorectification, and object-based classification of unmanned aerial vehicle (UAV) imagery for rangeland monitoring. *Photogrammetric Engineering and Remote Sensing* **76**(6), 661–672.
- Lam DW & Zechman FW (2006) Phylogenetic analyses of the Bryopsidales (Ulvophyceae, Chlorophyta) based on Rubisco large subunit gene sequences. *Journal of Phycology* **42**, 669–678.
- Lawver LA & Gahagan LM (2003) Evolution of Cenozoic seaways in the circum-Antarctic region. *Palaeogeography Palaeoclimatology Palaeoecology* **198**, 11–37.
- Le Lay G, Engler R, Franc E & Guisan A (2010) Prospective sampling based on model ensembles improves the detection of rare species. *Ecography* **33**, 1015–1027.
- Lebourgeois V, Bégué A, Labbé S, Mallavan B, Prévot L & Roux B (2008) Can commercial digital cameras be used as multispectral sensors? A crop monitoring test. *Sensors* **8**(11), 7300–7322.
- Lengyel S, Déri E, Varga Z, Horváth R, Tóthmérész B, Henry PY, Kobler A, Kutnar L, Babij V, Seliskar A, Christia C, Papastergiadou E, Gruber B & Henle K (2008) Habitat monitoring in Europe: a description of current practices. *Biodiversity and Conservation* **17**(14), 3327–3339.
- Lessios HA, Kessing BD & Pearse JS (2001) Population structure and speciation in tropical seas: global phylogeography of the sea urchin *Diadema*. *Evolution* **55**, 955–975.
- Littler MM, Littler DS (1999) Disturbances due to cyclone Gavin parallel those caused by a ship grounding. *Coral Reefs* **18**, 146.
- Littler MM, Littler DS, Norris JN, Bucher KE (1987) Recolonization of algal communities following the grounding of the Freighter Wellwood on Molasses Reef, Key Largo National Marine Sanctuary - Phase 2, Survey of algae and experimental design. *NOAA Technical Memoranda* **15**, Washington DC, USA. Available online at <http://www.aoml.noaa.gov/general/lib/CREWS>.
- Lüning K (1990) *Seaweeds: their environment, biogeography, and ecophysiology*. Wiley & Sons, New York, 527pp.
- Mangel M, Richerson K, Cresswell KA & Wiedenmann JR (2010) Modelling the effects of UV radiation on the survival of Antarctic krill (*Euphausia superba* Dana) in the face of limited data. *Ecological Modelling* **221**, 2095–2101.
- Maravelias CD & Reid DG (1997) Identifying the effects of oceanographic features and zooplankton on prespawning herring abundance using generalized additive models. *Marine Ecology-Progress Series* **147**, 1–9.
- Martinez-Meyer E & Peterson AT (2006) Conservatism of ecological niche characteristics in North American plant species over the Pleistocene-to-recent transition. *Journal of Biogeography* **33**, 1779–1789.
- Martinez-Meyer E, Peterson AT & Navarro-Siguenza AG (2004) Evolution of seasonal ecological niches in the *Passerina buntings* (Aves : Cardinalidae). *Proceedings of the Royal Society of London Series B-Biological Sciences* **271**, 1151–1157.

## References

- Martins EP & Hansen TF (1997) Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *American Naturalist* **149**, 646-667.
- Martins EP (1999) Estimation of ancestral states of continuous characters: a computer simulation study. *Systematic Biology* **48**, 642-650.
- Marzolf I & Poesen J (2009) The potential of 3D gully monitoring with GIS using high-resolution aerial photography and a digital photogrammetry system. *Geomorphology* **111**(1-2), 48-60.
- Mazel CH, Strand MP, Lesser MP, Crosby MP, Coles B & Nevis AJ (2003) High-resolution determination of coral reef bottom cover from multispectral fluorescence laser line scan imagery. *Limnology and Oceanography* **48**1 (1, part 2), 522-534.
- McCook IJ, Jompa J & Diaz-Pulido G (2001). Competition between corals and algae on coral reefs: a review of evidence and mechanisms. *Coral Reefs* **19**(4), 400-417.
- Monk J, Ierodiaconou D, Versace V, Bellgrove A, Harvey E, Rattray A, Laurenson L & Quinn GP (2010) Habitat suitability for marine fishes using presence-only modelling and multibeam sonar. *Marine Ecology-Progress Series* **420**, 157-174.
- Moore GE (1965) Cramping more components onto integrated circuits. *Electronics* **38**(8), 4pp.
- Mumby PJ & Edwards AJ (2002) Mapping marine environments with Ikonos imagery: enhanced spatial resolution can deliver greater thematic accuracy. *Remote Sensing of Environment* **82**(2-3), 248-257.
- Mumby PJ, Hastings A & Edwards HJ (2007) Thresholds and the resilience of Caribbean coral reefs. *Nature* **450**(7166), 98-101.
- Naim O (1988) Distributional patterns of mobile fauna associated with *Halimeda* on the Tiahura coral-reef complex (Moorea, French Polynesia). *Coral Reefs* **6**, 237-250.
- NASA (2010) *NASA Earth observations (NEO)*. NASA Goddard Space Flight Center. Available online at <http://neo.sci.gsfc.nasa.gov/Search.html>.
- NASA (2011) *Landsat 7 Science data users handbook*. NASA Goddard Space Flight Center. Available online at <http://landsathandbook.gsfc.nasa.gov>.
- Nature Editorial (2008) A place for everything. *Nature* **453**, 2.
- NCBI (2008) *GenBank*. National Library of Medicine, USA. Available online at <http://www.ncbi.nlm.nih.gov>.
- NOAA (2008) *World Ocean Database*. National Oceanographic Data Center, USA. Available online at <http://www.nodc.noaa.gov>.
- Oberthur T, Cock J, Andersson MS, Naranjo RN, Castaneda D & Blair M (2007) Acquisition of low altitude digital imagery for local monitoring and management of genetic resources. *Computers and Electronics in Agriculture* **58**(1), 60-77.
- Ormond R & Banaimoon S (1994). Ecology of intertidal macroalgal assemblages on the Hadramout Coast of southern Yemen, an area of seasonal upwelling. *Marine Ecology-Progress Series* **105**, 105-120.

- Pagel M (1999) Inferring the historical patterns of biological evolution. *Nature* **401**, 877-884.
- Paradis E, Claude J & Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289-290.
- Pauly K, Goossens R & De Clerck O (2011) Mapping coral-algal dynamics in a seasonal upwelling area using spaceborne high resolution sensors. *Proceedings of the ESA Living Planet Symposium*, ESA SPS-686, 8pp.
- Pauly K, Verbruggen H, Tyberghein L, Mineur F, Maggs CA, Shimada S & De Clerck O (2009) *Predicting spread and bloom risk areas of introduced and invasive seaweeds*. Oral presentation, 9<sup>th</sup> International Phycological Congress, Tokyo (Japan), 2-8 August 2009. *Phycologia* **48**(4), S103.
- Pearson RG (2007) *Species' Distribution Modeling for conservation educators and practitioners: synthesis*. American Museum of Natural History, USA. Available online at <http://ncep.amnh.org>.
- Pearson RG, Raxworthy CJ, Nakamura M & Peterson TA (2007) Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* **34**(1), 102-117.
- Peterson AT (2003) Predicting the geography of species' invasions via ecological niche modeling. *The Quarterly Review of Biology* **78**(4), 419-433.
- Peterson AT (2005) Predicting potential geographic distributions of invading species. *Current Science* **89**(1), 9.
- Peterson AT (2006) Uses and requirements of ecological niche models and related distributional models. *Biodiversity Informatics* **3**, 59-72.
- Peterson AT, Papes M & Soberon J (2008) Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling* **213**, 63-72.
- Pham-Hoàng H (1969) *Marine algae of south Việt Nam*. Ministry of Education and Youth, Saigon (Vietnam).
- Phillips SJ & Dudík M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**(2), 161-175.
- Phillips SJ, Anderson RP & Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**(3-4), 231-259.
- Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J & Ferrier S (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications* **19**(1), 181-197.
- Phinn S, Roelfsema C, Dekker A, Brando V & Anstee J (2008) Mapping seagrass species, cover and biomass in shallow waters: an assessment of satellite multi-spectral and airborne hyper-spectral imaging systems in Moreton Bay (Australia). *Remote Sensing of Environment* **112**, 3413-3425.
- Planer-Friedrich B, Becker J, Brimer B & Merkel BJ (2007) Low-cost aerial photography for high-resolution mapping of hydrothermal areas in Yellowstone National Park. *International Journal of Remote Sensing* **29**(6), 1781-1794.

## References

- Polcyn FC & Sattinger IJ (1969) Water depth determinations using remote sensing techniques. *Proceedings of the 6th International Symposium on Remote Sensing of Environment*, Volume II, 1017-1028. University of Michigan, Ann Arbor, Michigan (USA).
- Poncet J (1989) Présence du genre *Halimeda* Lamouroux, 1812 (algue verte calcaire) dans le Permien supérieur du sud tunisien. *Revue de Micropaléontologie* **32**, 40-44.
- Provan J, Booth D, Todd NP, Beatty GE & Maggs CA (2008) Tracking biological invasions in space and time: elucidating the invasive history of the green alga *Codium fragile* using old DNA. *Diversity and Distributions* **14**, 343-354.
- Provan J, Murphy S & Maggs CA (2004) Universal plastid primers for Chlorophyta and Rhodophyta. *European Journal of Phycology* **39**, 43-50.
- R development core team. (2010) R: A language and environment for statistical computing. Vienna, Austria: R foundation for statistical computing. Available online at <http://www.r-project.org>.
- Rabalais NN, Turner RE, Díaz RJ, Justic D (2009) Global change and eutrophication of coastal waters. *ICES Journal of Marine Science* **66**, 1528-1537.
- Rambaut A & Drummond AJ (2007) *Tracer*. Available online at <http://beast.bio.ed.ac.uk/tracer>.
- Ratnasingham S & Hebert PDN (2007) BOLD: The Barcode of Life Data System ([www.barcodinglife.org](http://www.barcodinglife.org)). *Molecular Ecology Notes* **7**(3), 355-364.
- Raxworthy CJ, Ingram CM, Rabibisoa N & Pearson RG (2007) Applications of ecological niche modeling for species delimitation: a review and empirical evaluation using day geckos (*Phelsuma*) from Madagascar. *Systematic Biology* **56**, 907 - 923.
- Raxworthy CJ, Martinez-Meyer E, Horning N, Nussbaum RA, Schneider GE, Ortega-Huerta MA & Peterson AT (2003) Predicting distributions of known and unknown reptile species in Madagascar. *Nature* **426**(6968), 837-41.
- Ready J, Kaschner K, South AB, Eastwood PD, Rees T, Rius J, Agbayani E, Kullander S & Froese R (2010) Predicting the distributions of marine organisms at the global scale. *Ecological Modelling* **221**, 467-478.
- Richardson WD (1975) The marine algae of Trinidad, West Indies. *Bulletin of the British Museum (Natural History) Botany* **5**, 71-143.
- Riordan EC, Rundel PW (2009) Modelling the distribution of a threatened habitat: the California sage scrub. *Journal of Biogeography* **36**, 2176-2188.
- Rios NE & Bart HL Jr. (1997) *GEOLocate georeferencing software: user's manual*. Tulane Museum of Natural History, Belle Chasse LA, USA. Available online at <http://www.museum.tulane.edu/geolocate>.
- Rissler LJ & Apodaca JJ (2007) Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*). *Systematic Biology* **56**, 924 - 942.
- Rissler LJ, Hijmans RJ, Graham CH, Moritz C & Wake DB (2006) Phylogeographic lineages and species comparisons in conservation analyses: a case study of California herpetofauna. *American Naturalist* **167**, 655-666.

- Robinson LM, Elith J, Hobday AJ, Pearson RG, Kendall BE, Possingham HP & Richardson AJ (2011) Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. *Global Ecology and Biogeography*, DOI: 10.1111/j.1466-8238.2010.00636.x.
- Rogers CS, McLain LN, Tobias CR (1991) Effects of hurricane Hugo (1989) on a coral-reef in St-John, Usvi. *Marine Ecology-Progress Series* **78**, 189-199.
- Rögl F & Steininger FF (1984) Neogene paratethys, Mediterranean and Indo-Pacific seaways. In: P Brenchley (ed.) *Fossils and climate*, 171-200. John Wiley & Sons.
- Roleda MY, Wiencke C, Hanelt D, Van De Poll W & Gruber A (2005) Sensitivity of Laminariales zoospores from Helgoland (North Sea) to ultraviolet and photosynthetically active radiation: implications for depth distribution and seasonal reproduction. *Plant Cell and Environment* **28**, 466-479.
- Ronquist F & Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574.
- Russ GR, McCook LJ (1999) Potential effects of a cyclone on benthic algal production and yield to grazers on coral reefs across the central Great Barrier Reef. *Journal of Experimental Marine Biology and Ecology* **235**, 237-254.
- Sanderson MJ (2002) Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution* **19**, 101-109.
- Sanderson MJ (2003) r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301-302.
- Sasai Y, Sasaoka K, Sakaki H & Ishida A (2007) Seasonal and intra-seasonal variability of chlorophyll-a in the North Pacific: model and satellite data. *Journal of the Earth Simulator* **8**, 3-11.
- Saunders GW & Lehmkuhl KV (2005) Molecular divergence and morphological diversity among four cryptic species of *Plocamium* (Plocamiales, Florideophyceae) in northern Europe. *European Journal of Phycology* **40**, 293-312.
- Schils T & Coppejans E (2003) Phylogeography of upwelling areas in the Arabian Sea. *Journal of Biogeography* **30**(9), 1339-1356.
- Schils T & Wilson SC (2006) Temperature threshold as a biogeographic barrier in northern Indian Ocean macroalgae. *Journal of Phycology* **42**, 749-756.
- Schluter D, Price T, Mooers AO & Ludwig D (1997) Likelihood of ancestor states in adaptive radiation. *Evolution* **51**, 1699-1711.
- Schmidt-Nielsen K (1990) *Animal physiology: adaptation and environment*. Cambridge University Press, New York, 607pp.
- Schroeder RE, Green AL, Demartini EE, Kenyon JC (2008) Long-term effects of a ship grounding on coral reef fish assemblages at Rose Atoll, American Samoa. *Bulletin of Marine Sciences* **82**, 345-364.
- Scoffin TP (1982) Reef aerial photography from a kite. *Coral Reefs* **1**(1), 67-69.
- Shimodaira H (2004) Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling. *Annals of Statistics* **32**, 2616-2641.

## References

- Siebert S, Gries D, Zhang X, Runge M & Buerkert A (2004) Non-destructive dry matter estimation of *Alhagi Sparsifolia* vegetation in a desert oasis of Northwest China. *Journal of Vegetation Science* **15**(3), 365-372.
- Smith AG, Smith DG & Funnell BM (1994) *Atlas of Mesozoic and Cenozoic coastlines*. Cambridge University Press, Cambridge.
- Smith JE, Shaw M, Edwards RA, Obura D, Pantos O, Sala E, Sandin SA, Smriga S, Hatay M & Rohwer FL (2006) Indirect effects of algae on coral: algae-mediated, microbe-induced coral mortality. *Ecology Letters* **9**, 835-845.
- Soberón J (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters* **10**, 1115-1123.
- Soberón J, Peterson AT (2004) Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society of London B* **359**, 689-698.
- Stang FW (1969) Ocean and water surface temperature measurements using infrared remote sensing techniques. *Proceedings of the SPIE 14th Annual Technical Symposium: Photo-Optical Instrumentation Applications and Theory*, 77-83.
- Steneck RS, Dethier MM (1994) A functional group approach to the structure of algal-dominated communities. *Oikos* **69**, 476-498.
- Suárez AM (2005) Lista de las macroalgas marinas Cubanas. *Revista de Investigaciones Marinas* **26**, 93-148.
- Sullivan J (2005) Maximum-likelihood methods for phylogeny estimation. *Molecular Evolution: Producing the Biochemical Data, Part B. Methods in Enzymology* **395**, 757-779.
- Suzuki R & Shimodaira H (2006) Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540-1542.
- Swofford DL, Waddell PJ, Huelsenbeck JP, Foster PG, Lewis PO & Rogers JS (2001) Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Systematic Biology* **50**, 525-539.
- Tassan S (1996). Modified Lyzenga's method for macroalgae detection in water with non-uniform composition. *International Journal of Remote Sensing* **17**(8), 1601-1607.
- Taylor WR (1928) The marine algae of Florida with special reference to the Dry Tortugas. *Bibliotheca Phycologica* **2**.
- Taylor WR (1960) *Marine algae of the eastern tropical and subtropical coasts of the Americas*. University of Michigan Press, Ann Arbor (USA)
- Teske P, Hamilton H, Matthee C & Barker N (2007) Signatures of seaway closures and founder dispersal in the phylogeny of a circumglobally distributed seahorse lineage. *BMC Evolutionary Biology* **7**, 138.
- Theriault C, Scheibling R, Hatcher B & Jones W (2006) Mapping the distribution of an invasive marine alga (*Codium fragile* subsp. *tomentosoides*) in optically shallow coastal waters using the Compact Airborne Spectrographic Imager (CASI). *Canadian Journal of Remote Sensing* **32**, 315-329.

- Thomas CD, Cameron A, Green RE, Bakkenes M, Beaumont LJ, Collingham YC, Erasmus BFN, De Siqueira MF, Grainger A, Hannah L, Hughes L, Huntley B, Van Jaarsveld AS, Midgley GF, Miles L, Ortega-Huerta MA, Peterson AT, Phillips OL & Williams SE (2004) Extinction risk from climate change. *Nature* **427**, 145-148.
- Thuiller W, Richardson DM & Midgley GF (2007) Will climate change promote alien plant invasions? In: W Nentwig (ed.) *Biological invasions*. Ecological Studies **193**, 197-211. Springer Verlag Berlin Heidelberg.
- Thuiller W, Richardson DM, Pysek P, Midgley GF, Hughes GO & Rouget M (2005) Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Global Change Biology* **11**, 2234-2250.
- Titlyanov EA, Titlyanova TV, Chapman DJ (2008) Dynamics and patterns of algal colonization on mechanically damaged and dead colonies of the coral *Porites lutea*. *Botanica Marina* **51**, 285-296.
- Tittensor DP, Baco AR, Brewin PE, Clark MR, Consalvey M, Hall-Spencer J, Rowden AA, Schlacher T, Stocks KI & Rogers AD (2009) Predicting global habitat suitability for stony corals on seamounts. *Journal of Biogeography* **36**, 1111-1128.
- Torchin ME, Lafferty K, McKenzie V & Kuris A (2003) Introduced species and their missing parasites. *Nature* **421**, 628-630.
- Trivedi MR, Berry PM, Morecroft MD & Dawson TP (2008) Spatial scale affects bioclimate model projections of climate change impacts on mountain plants. *Global Change Biology* **14**, 1089-1103.
- Troupin C, Machin F, Ouberdous M, Sirjacobs D, Barth A & Beckers JM (2010) High-resolution climatology of the northeast Atlantic using Data-Interpolating Variational Analysis (DIVA). *Journal of Geophysical Research-Oceans* **115** C08005, DOI: 10.1029/2009JC005512.
- Trowbridge CD (1998) Ecology of the green macroalga *Codium fragile* (Suringar) Hariot: invasive and noninvasive subspecies. *Oceanography and Marine Biology: an Annual Review* **36**, 1-64.
- Tupper M, Tan MK, Tan SL, Radius MJ, Abdullah S (2011) ReefBase: a global information system on coral reefs. Available online at <http://www.reefbase.org>.
- Tyberghein L, Verbruggen H, Pauly K, Troupin C, Mineur F & De Clerck O (2011) Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography*, DOI: 10.1111/j.1466-8238.2011.00656.x.
- Vahtmäe E, Kutser T, Martin G & Kotta J (2006) Feasibility of hyperspectral remote sensing for mapping benthic macroalgal cover in turbid coastal waters - a Baltic Sea case study. *Remote Sensing of Environment* **101**, 342-351.
- Valavanis VD, Pierce GJ, Zuur AF, Palialexis A, Saveliev A, Katara I & Wang JJ (2008) Modelling of essential fish habitat based on remote sensing, spatial analysis and GIS. *Hydrobiologia* **612**, 5-20.
- van den Hoek C (1982) The distribution of benthic marine algae in relation to the temperature regulation of their life histories. *Biological Journal of the Linnean Society* **18**, 81-144.



## References

- van den Hoek C (1984) World-wide latitudinal and longitudinal seaweed distribution patterns and their possible causes, as illustrated by the distribution of rhodophyten genera. *Helgolander Meeresuntersuchungen* **38**, 227-257.
- van den Hoek C, Breeman AM & Stam WT (1990) The geographic distribution of seaweed species in relation to temperature: present and past. In: JJ Beukema, WJ Wolff & JJWM Brouns (eds.) *Expected effects of climatic change on marine coastal ecosystems*, 55-67. Kluwer Academic Press, Dordrecht.
- VanDerWal J, Shoo LP, Johnson CN, Williams, SE (2009) Abundance and the environmental niche: environmental suitability estimated from niche models predicts the upper limit of local abundance. *The American Naturalist* **174**, 282-291.
- Verbruggen H (2005) Resegmenting *Halimeda*. Molecular and morphometric studies of species boundaries within a green algal genus. In *Phycology Research Group*. Ghent University, Ghent, 213pp.
- Verbruggen H (2008) *TreeGradients*. Available online at <http://www.phycoweb.net>.
- Verbruggen H, Ashworth M, LoDuca ST, Vlaeminck C, Cocquyt E, Sauvage T, Zechman FW, Littler DS, Littler MM, Leliaert F & De Clerck O (2010) A multi-locus time-calibrated phylogeny of the siphonous green algae. *Molecular Phylogenetics and Evolution* **56**, 659-674.
- Verbruggen H, De Clerck O, Kooistra W & Coppejans E (2005a) Molecular and morphometric data pinpoint species boundaries in *Halimeda* section *Rhipsalis* (Bryopsidales, Chlorophyta). *Journal of Phycology* **41**, 606-621.
- Verbruggen H, De Clerck O, Schils T, Kooistra W & Coppejans E (2005b) Evolution and phylogeography of *Halimeda* section *Halimeda*. *Molecular Phylogenetics and Evolution* **37**, 789-803.
- Verbruggen H, Tyberghein L, Pauly K, Vlaeminck C, Van Nieuwenhuyze K, Kooistra WHCF, Leliaert, F & De Clerck O (2009) Macroecology meets macroevolution: evolutionary niche dynamics in the seaweed *Halimeda*. *Global Ecology and Biogeography* **18**(4), 393-405.
- Verhoeven GJ (2008) Imaging the invisible using modified digital still cameras for straightforward and low-cost archaeological near-infrared photography. *Journal of Archaeological Science* **35**(12), 3087-3100.
- Verhoeven GJ (2011) Software Review: Taking Computer Vision Aloft - Archaeological Three-dimensional Reconstructions from Aerial Photographs with PhotoScan. *Archaeological Prospection* **18**, 67-73.
- Verhoeven GJ, Loenders J, Vermeulen F & Docter R (2009) Helikite aerial photography – a versatile means of unmanned, radio controlled, low-altitude aerial archaeology. *Archaeological Prospection* **138**, 125-138.
- Verhoeven GJ, Smet PF, Poelman D & Vermeulen F (2009) Spectral characterization of a digital still camera with NIR modification to enhance archaeological observation. *IEEE Transactions on Geoscience and Remote Sensing* **47**(10), 3456-3468.
- Vroom P, Page K, Kenyon J & Brainard R (2006). Algae-dominated reefs. *American Scientist* **94**(5), 430-437.

- Vroom PS, Asher J, Braun CL, Coccagna E, Vetter, OJ, Cover WA, McCully KM, Potts DC, Marie A, Vanderlip C (2009) Macroalgal (*Boodlea composita*) bloom at Kure and Midway Atolls, Northwestern Hawaiian Islands. *Botanica Marina* **52**, 361-363.
- Wang D, Zhao H (2008) Estimation of phytoplankton responses to hurricane Gonu over the Arabian Sea based on ocean color data. *Sensors* **8**, 4878-4893.
- Warren DL & Seifert S (2011) Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications* **21**, 335-342.
- Wiens JA, Stralberg D, Jongsomjit D, Howell CA & Snyder MA (2009) Niches, models, and climate change: assessing the assumptions and uncertainties. *Proceedings of the National Academy of Sciences of the United States of America* **106** Suppl, 19729-19736.
- Wiens JJ & Graham CH (2005) Niche conservatism: integrating evolution, ecology, and conservation biology. *Annual Review of Ecology Evolution and Systematics* **36**, 519-539.
- Wiens JJ, Parra-Olea G, Garcia-Paris M & Wake DB (2007) Phylogenetic history underlies elevational biodiversity patterns in tropical salamanders. *Proceedings of the Royal Society B-Biological Sciences* **274**, 919-928.
- Wiley EO, Mcnysset KM, Peterson AT, Robins CR & Stewart AM (2003) Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. *Oceanography* **16**, 120-127.
- Williams JN, Seo C, Thorne J, Nelson JK, Erwin S, O'Brien JM & Schwartz MW (2009) Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions* **15**(4), 565-576.
- Wilson SC (2000). Northwest Arabian Sea and Gulf of Oman. In: C Sheppard (ed.) *Seas at the millennium: an environmental evaluation: 2*, xxi + 920 pp. Regional chapters: The Indian Ocean to The Pacific. Pergamon, Amsterdam.
- Womersley HBS & Bailey A (1970) Marine algae of the Solomon Islands. *Philosophical Transactions of the Royal Society of London B* **259**, 257-352.
- Woodley JD, Chornesky EA, Clifford PA, Jackson JB, Kaufman LS, Knowlton N, Lang JC, Pearson MP, Porter JW, Rooney MC, Rylaarsdam KW, Tunnicliffe VJ, Wahle CM, Wulff JL, Curtis ASG, Dallmeyer MD, Jupp BP, Koehl MAR, Neigel J, Sides EM (1981) Hurricane Allen's impact on Jamaican coral reefs. *Science* **214**, 749-755.
- Wootton JT, Pfister CA & Forester JD (2008) Dynamic patterns and ecological impacts of declining ocean pH in a high-resolution multi-year dataset. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 18848-18853.
- Wundram D & Löffler J (2008) High-resolution spatial analysis of mountain landscapes using a low-altitude remote sensing approach. *International Journal of Remote Sensing* **29**(4), 961-974.
- Wysor B (2004) An annotated list of marine Chlorophyta from the Pacific Coast of the Republic of Panama with a comparison to Caribbean Panama species. *Nova Hedvigia* **78**, 209-241.

## References

- Yesson C & Culham A (2006) Phyloclimatic modeling: combining phylogenetics and bioclimatic modeling. *Systematic Biology* **55**, 785-802.
- Zachos J, Pagani M, Sloan L, Thomas E & Billups K (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686-693.

# CURRICULUM VITAE

**NAME** **Klaas PAULY**  
**HOME ADDRESS** Akkerstraat 54,  
B-2990 WUUSTWEZEL,  
BELGIUM  
**CONTACT** [klaas.pauly@gmail.com](mailto:klaas.pauly@gmail.com)  
Mobile +32 475 826 369  
**DATE OF BIRTH** 3 April, 1981  
Oostende (Belgium)  
**NATIONALITY** Belgian  
**LAST POSITION** Teaching Assistant (15/11/2004 - 15/11/2011)  
**AFFILIATION** Phycology Research Group, Biology Department  
Ghent University (UGent), Belgium  
Krijgslaan 281/S8, B-9000 GENT, BELGIUM  
<http://www.phycology.ugent.be>



## AREA OF EXPERTISE

My PhD research covered spatio-temporal analysis of seaweed communities using remote sensing and niche modeling, with a focus on invasive species management. For sampling and ground-truthing, I organized several field campaigns in remote areas. Recently, I'm exploring the potential of low-altitude photography for coastal monitoring. I have a strong field knowledge of seaweeds from the North Sea and (sub)tropics and animal diversity from the North Sea. As a teaching assistant, I organized practical courses and field training courses in general botany, phycology and marine ecology.

## EDUCATION

### *SELECTED GRADUATE AND POSTGRADUATE COURSES AND TRAINING WORKSHOPS*

12/2010, 1 day	Demo-day: UAV mission planning, execution and automatic triangulation of aerial photos	Gatewing NV, Ghent (Belgium)
07/2010, 1 day	Training workshop BEAM: processing tools for Level-0 SMOS and MERIS satellite data	Brockmann Consult, ESA Living Planet Symposium (Norway)
2009 - 2010, 1 <sup>st</sup> semester	(Stereo)photogrammetry, DEM extraction and 3D visualization	Ghent University (Belgium)
10/2007, 1 week	Training workshop Species Distribution Modeling	American Museum of Natural History, Portal branch, AZ (USA)
2005 - 2006, 2 <sup>nd</sup> semester	Remote sensing: image interpretation	Ghent University (Belgium)

2005 - 2006, 1 <sup>st</sup> semester	Geographic Information Systems	Ghent University (Belgium)
2004 - 2005, 1 <sup>st</sup> semester	Remote sensing: image registration and processing	Ghent University (Belgium)
2004 - 2005, 1 <sup>st</sup> semester	Applied biostatistics	Ghent University (Belgium)
09/2004, 2 weeks	Field training course <i>Posidonia</i> : marine ecology and spatial analysis of seagrass meadows	Observatoire Océanologique, UPMC (Paris VI), Villefranche- sur-Mer (France)
08/2003, 1 month	Field training course Biodiversity of Megatidal Coastal Systems: marine zoology and ecology	Observatoire Océanologique, UPMC (Paris VI), Roscoff (France)
2007	Basic teaching assistance training course and research practical course education	Ghent University

#### UNIVERSITY CURRICULUM

2002 - 2004	Licentiate (MSc) Biology, option Zoology <ul style="list-style-type: none"> <li>▪ Optional subjects: Introduction to GIS; Coastal Ecosystems; Oceanography; Ecophysiology</li> <li>▪ Dissertation: “Biogeography and seasonality of macroalgal communities in the Gulf of Oman” (in Dutch): baseline study for EIA at the coastal construction site of the OMIFCO fertilizer plant</li> </ul>	Ghent University (Belgium), graduated with Great Distinction
1999 - 2002	Candidate (BSc) Biology	Ghent University (Belgium), with Distinction

#### PROFESSIONAL EXPERIENCE AND TEACHING

##### ORGANISATION OF FIELD CAMPAIGNS

2003 - 2009, 9 one-month expeditions in Oman, Tanzania and Japan for quantitative and qualitative sampling of intertidal and subtidal seaweeds along several geographical and ecological gradients for population genetics, phylogeographic analyses and ground-truthing of remote sensing studies; grazer exclusion and nutrient enrichment field experiments; deployment and retrieval of long-term temperature loggers and measurement of water column nutrient concentrations

##### SUPERVISION OF DISSERTATIONS

2004 - 2010, 8 MSc and 5 BSc dissertations concerning ecology and biogeography of (invasive) seaweeds and corals (using vegetation relevés and

multivariate analyses, culture and field experiments, PAM and growth measurements, near-infrared and kite aerial photography, remote sensing and niche modeling)

#### TEACHING

2004 - 2010, organization and supervision of weekly practical courses (Phycology, Evolution, Microscopic techniques, Botany) and annual 1-week field training courses (coastal biodiversity and ecosystems) organized in Wimereux (France) for BSc Bioscience Engineering, BSc and MSc Biology, MSc Marine & Lacustrine Sciences

#### CONSULTANCY SERVICES

- 2010, cooperation on the checklist for seaweeds along the Belgian coast, commissioned by Koninklijk Belgisch Instituut voor Natuurwetenschappen (KBIN) and Vlaams Instituut voor de Zee (VLIZ, project BeRMS, <http://www.marinespecies.org/berms>)
- 10/2005, 08/2009 and 01/2011, baseline surveying and assembling checklists of marine algae and seagrasses for environmental impact assessment in Oman and Abu Dhabi, commissioned by Five Oceans Environmental Services LLC (Oman, <http://www.5oes.com>)
- 2007, 2008 and 2010, cooperation on the seaweed monitoring of the Spuikom (Ostend, Belgium) commissioned by the Vlaamse Milieumaatschappij (VMM) and the North Sea, commissioned by the Management Unit of the North Sea Mathematical Models (MUMM)

**PUBLICATIONS** (full text copies are available on <http://www.phycology.ugent.be>)

#### A1 (INTERNATIONAL PEER-REVIEWED JOURNAL ARTICLES, ISI-LISTED)

- Pauly K., Jupp B.P. & De Clerck O. (2011) Modeling the distribution and ecology of *Trichosolen* blooms on coral reefs worldwide. *Marine Biology*, DOI: 10.1007/s00227-011-1729-0. IF: 1.999
- Tyberghein L., Verbruggen H., Pauly K., Troupin C., Mineur F. & De Clerck O. (accepted) Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography*. IF: 5.304
- Verbruggen H., Tyberghein L., Pauly K., Van Nieuwenhuyze K., Vlaeminck C., Kooistra W.H.C.F., Leliaert F. & De Clerck O. (2009) Macroecology meets macroevolution: evolutionary niche dynamics in the marine green alga *Halimeda*. *Global Ecology and Biogeography* **18**: 393-405. (First three authors have equal contributions). IF: 5.304.

#### A1 PAPERS IN PREPARATION

- Pauly K., Tyberghein L. & De Clerck O. Spatial scale-dependent prediction in marine niche modeling: a case study.
- Pauly K. & De Clerck O. Low-cost very high resolution intertidal vegetation monitoring enabled by near-infrared kite aerial photography.

## B2 (BOOK CHAPTERS)

Pauly K. & De Clerck O. (2010) GIS-based environmental analysis, remote sensing and niche modeling of seaweed communities. In: Israel A., Einav R. & Seckbach J. (eds.) *Seaweeds and their role in globally changing environments*. Kluwer Academic Publishers, Dordrecht, The Netherlands, p93-114.

## P1 (INTERNATIONAL CONFERENCE PROCEEDINGS, ISI-LISTED)

- Pauly K., Goossens R. & De Clerck O. (in press) Mapping coral-algal dynamics in a seasonal upwelling area using spaceborne high resolution sensors. *The Proceedings of the ESA Living Planet Symposium*, 28 June - 2 July 2010, Bergen (Norway), ESA Special Publication SP-686, 8pp. Orally presented in the Coastal Zones section on 1 July, 2010.
- Pauly K., Verbruggen H., Tyberghein L., Mineur F., Maggs C.A., Shimada S. & De Clerck O. (2009) Predicting spread and bloom risk areas of introduced and invasive seaweeds. *Phycologia* **48** (S4), 103, meeting abstract from the International Phycological Congress 9, 2-8 August, Tokyo (Japan). Orally presented in the Invasive Seaweeds symposium on 3 August, 2010.

## ORAL CONFERENCE PRESENTATIONS, NOT ISI-LISTED

Pauly K. (2009) *Kite Aerial Photography for monitoring remote nature reserves: a test case in a Green turtle breeding area on Masirah Island (Oman)*. 10<sup>th</sup> VLIZ Young Scientist's Day, November 27, Oostende (Belgium). Accompanied by a similarly-titled poster presentation. Both received honorable mentions by the public and jury.  
<http://www.vliz.be/EN/INTRO/&p=show&id=838>

## COMPUTING EXPERIENCE

- *GIS, remote sensing and photogrammetry*: ESRI ArcGIS 9.2, Clark Labs Idrisi Taiga, ITC ILWIS 3.7, SupreSoft Virtuozo, AgiSoft PhotoScan
- *ENM*: Maxent 3.3.3, ENMTools 1.2, ModEco 1.0, Biomapper 4.0
- *Statistics*: Statistica 6, Primer 6, PC-ORD 4, limited knowledge of the R
- *Image processing*: Adobe Photoshop and Illustrator CS3, ImageJ
- *External device control software for GPS, temperature loggers and PAM*

## LANGUAGES

*Dutch* (native), *English* (fluent), *French* (fluent), *Swedish* (limited), *German* (reading), *Arabic* (basic)

## OTHER QUALIFICATIONS

- *European standard driving licence B* (2000) and *4WD desert driving* (Oman, 2 day training, 2004)
- *Superior Life-guard, Coastal life-guard and life-guard Navigation Instructor* (1999 - 2005)
- *1\* SCUBA diver*, CMAS (2003), over 100 logged dives, 2\* in course
- *VHF maritime communication and International Certificate for Pleasure Craft* (2006)
- *Indoor maritime rescue training*, Falck Nutec (Netherlands, 1 day training, 2009)