Development and validation of spatial distribution models of marine habitats, in support of the ecological valuation of the seabed

Ontwikkeling en validering van ruimtelijke verspreidingsmodellen van mariene habitats, ter ondersteuning van het ecologisch waarderen van de zeebodem

Els Verfaillie

Promotor: Prof. Dr. Marc De Batist
Co-promotor: Dr. Vera Van Lancker
Co-promotor: Dr. Steven Degraer

Academic year 2007-2008

Dissertation submitted for the degree of Doctor of Science: Geology

Instituut voor de Aanmoediging van Innovatie door Wetenschap en Technologie in Vlaanderen

Members of the reading committee:

Prof. Dr. Marc De Batist (Ghent University, Belgium): promotor
Dr. Vera Van Lancker (Ghent University, Belgium): co-promotor
Dr. Steven Degraer (Ghent University, Belgium): co-promotor
Prof. Dr. ir. Marc Van Meirvenne (Ghent University, Belgium)
Dr. Sytze Van Heteren (TNO Bouw en Ondergrond, the Netherlands)

Members of the examination committee:

Prof. Dr. Jean-Pierre Henriet (Ghent University, Belgium): chair
Prof. Dr. Marc De Batist (Ghent University, Belgium): promotor
Dr. Vera Van Lancker (Ghent University, Belgium): co-promotor
Dr. Steven Degraer (Ghent University, Belgium): co-promotor
Prof. Dr. ir. Marc Van Meirvenne (Ghent University, Belgium)
Dr. Sytze Van Heteren (TNO Bouw en Ondergrond, the Netherlands)
Prof. Dr. Patric Jacobs (Ghent University, Belgium)
Dr. ir. Veerle Huvenne (National Oceanography Centre, Southampton, United Kingdom)
Dr. Anthony Grehan (National University of Ireland, Galway, Ireland)

Table of contents

List of figures

List of tables

Abbreviations

a	range
BCP	Belgisch continentaal plat
BCS	Belgian continental shelf
BDNZ	Belgisch deel van de Noordzee
BPI	bathymetric position index
BPNS	Belgian part of the North Sea
BVM	biological valuation map
BWZee	biological valuation map of the Belgian continental shelf
C_0	nugget effect
C_1	sill
CCI	correctly classified instances
Chl	chlorophyl
CEC	Commission of the European Communities
C-H	Calinski-Harabasz index
DEFRA	Department for Environment, Food and Rural Affairs
DEM	digital elevation model
DF	discriminant function
DFA	discriminant function analysis
d_s10	10th percentile of the sand fraction
d50	median grain-size
d_s50	median grain-size of the sand fraction
d_s90	90th percentile of the sand fraction
DTM	digital terrain model
EGV	ecogeographical variables
ENFA	Ecological Niche Factor Analysis
EUNIS	European Nature Information System
exp	exponential model
GIS	Geographic Information System
GM	geometric mean
GPS	global positioning system
GT	ground-truthing
h	lag distance
HM	harmonic mean
HS	habitat suitability
HSM	habitat suitability model
ICES	International Council for the Exploration of the Sea
JNCC	Joint Nature Conservation Committee
KED	Kriging with an external drift
KT	Kriging with a trend model
LSD	least significant difference
LR	linear regression
MACRODAT	macrobenthos database (Marine Biology Section, Ugent – Belgium, 2008)
MAEE	mean absolute estimation error
MAREBASSE	Management, Research and Budgeting of Aggregates in Shelf Seas related to End-users
MEE	mean estimation error

MESH	Development of a Framework for Mapping European Seabed Habitats
Min	minimal distance
MLLWS	Mean Lowest Low Water at Spring
MPA	marine protected area
MSEE	mean square estimation error
MSP	marine spatial planning
MUMM	Management Unit of the North Sea, Mathematical Models and the Scheldt Estuary
OK	Ordinary Kriging
PC	principal component
PCA	principal components analysis
RCMG	Renard Centre of Marine Geology
RMSEE	root mean square estimation error
RS	remote sensing
RV	research vessel
SAC	special area of conservation
SD	standard deviation
SPA	special protection area
spp	species
TSA	transitional species assemblage
TSM	total suspended matter
TWINSPAN	Two-Way Indicator Species Analysis
UTM	Universal Transverse Mercator
WGS	World Geodetic System

Dankwoord

Allereerst bedank ik mijn co-promotor Vera. Zij is het die mij in 2001 heeft laten kennismaken met habitatkartering. Dankzij haar heb ik vele jaren ervaring kunnen opdoen en heel wat mogelijkheden gekregen om naar het buitenland te gaan, op campagne te gaan enz. Vera was op ieder moment beschikbaar, zowel voor een sterk inhoudelijke als ontspannende babbel. Zonder haar motivatie zou ik nooit aan dit doctoraat begonnen zijn of het volgehouden hebben!

Mijn andere co-promotor Steven wordt heel erg bedankt voor zijn enthousiasme en inhoudelijke visie, die me dikwijls hebben vooruitgeholpen! Dankzij de samenwerking met Steven, hebben we de laatste jaren inzichten gekregen in het biologische aspect van habitatkartering.

Mijn promotor Prof. Marc De Batist dank ik voor zijn enthousiasme en de babbels tussendoor. Ook al is ons studieterrein compleet verschillend (geen meren aan de andere kant van de wereld voor mij ☹), toch was Tist altijd bereid tot het geven van allerhande tips rond publiceren, doctoraatsopbouw, samenzitten voor een spoedvergadering...

Prof. Jean-Pierre Henriet dank ik voor de goede samenwerking van de afgelopen jaren en de ondersteuning bij alles wat ik ondernam. Het samen uitwerken van een artikel rond Piccard en allerhande duikboten, deden me echt kennis maken met zijn liefde voor de onderwaterwereld.

Prof. Van Meirvenne bedank ik voor de interessante cursus Geostatistiek die ik volgde tijdens mijn eerste doctoraatsjaar en de geslaagde samenwerking die daarop volgde.

Zonder de RCMG-collega's van de voorbije jaren zou het bijlange niet zo leuk geweest zijn om te doctoreren. Ik bedank dan ook vooral mijn onmiddellijke (ex-) collega's Vera, Samuel, Isabelle, Kristien, Sonia, Sophie en Valérie met wie het altijd leuk was om een gezellige babbel te doen aan ons 'eiland' of met wie de tijd op de Belgica superleuk was. Vooral Isabelle wil ik bedanken voor haar steun tijdens het laatste doctoraatsjaar; zij moest mijn gezaag dikwijls eens aanhoren, maar ik kon ook bij haar terecht als ik inhoudelijke vragen probeerde op te lossen.
Marc(ske) krijgt een speciaal woord van dank voor de ontspannende babbeltjes op mindere momenten, of gewoon, omdat ik weeral eens passeerde bij het naar toilet gaan en toevallig binnenstapte...
Alle andere RCMG-collega's worden ook heel erg bedankt voor de leuke middagpauzes, de fruitmomenten, de baby-babbels...

Bij de mariene biologen wil ik vooral Wouter, Daniëlle, Sofie en Julie bedanken. Zij zorgden elk op hun eigen manier voor een ruggesteun. Wouter stond op ieder moment klaar om iets na te lezen of statistische vragen op te lossen. Daniëlle heeft me heel goed geholpen bij het demonstreren van de uitwerking van de sedimentstalen. Sofie heeft me bij het beëindigen van haar doctoraat sterk gemotiveerd om verder te doen. En dankzij de wekelijkse etentjes met Julie (en Isabelle) tijdens het laatste jaar, kon ik op tijd bijpraten en ontspannen.

Bij de BMM worden Michaël Fettweis, Dries van Den Eynde, Fritz Francken, Barbara Van Mol en Kevin Ruddick bedankt voor hun bijdrages van hydrodynamische, sediment- en satellietgegevens.

De bemanning van de Belgica en de Zeeleeuw wordt bedankt voor de leuke campagnes.

De voorbije jaren heb ik veel geleerd en ervaring opgedaan tijdens de samenwerkingen in allerlei projecten (Gaufre, BWZee, Marebasse, Quest4D, MESH). Bedankt aan alle betrokken personen! Vooral het hele MESH-team wil ik bedanken voor de geslaagde samenkomsten een beetje overal in Europa. In het begin voelde ik me een echt groentje, maar dankzij al die ervaren mensen heb ik enorm veel bijgeleerd en ebde dit gevoel langzaam weg...

Het IWT Vlaanderen wordt hartelijk bedankt voor de financiering van mijn onderzoek gedurende de afgelopen vier jaren.

De vrienden van "de lopers" en "de geografen" wil ik bedanken voor de leuke momenten tussendoor, de reisjes of weekends waar ik jammer genoeg de laatste tijd niet altijd kon bijzijn, de maandelijkse etentjes... Met Griet, Ineke en Katrien was het leuk om voor de geboorte van Rien te gaan lopen en bij te kletsen. Met Tine en Eveline was het leuk om samen te zwemmen, yoga te volgen en samen zwanger te zijn.
Leen en Vinciane dank ik voor de gezellige etentjes, het samen zwanger zijn en de daarmee gepaard gaande baby-babbels.

Mireille en Filip van het Pirateneiland en Katty en Tanja van De Melkweg worden heel erg bedankt voor alle fantastische zorgen voor Rien op de vele dagen dat ik moest werken.

Mijn lieve (schoon)familie, mijn fantastische ouders en zussen worden bedankt voor alle goede zorgen voor mezelf, Bart en Rien!
Mijn liefste Bart, zonder jou zou ik het nooit volgehouden hebben. Je was er altijd voor een oppeppende babbel, voor de uitstapjes met Rien om mij te laten werken,... Bedankt voor alle liefde die je me geeft en je weet dat die dubbel en dik wederzijds is!
Dan is er nog onze liefste zoon Rien; met zijn enthousiasme over alles wat hij ontdekt in de wereld, zorgde hij altijd voor de nodige ontspanning.
Tenslotte was de baby in de buik de beste motivatie om verder te werken aan mijn doctoraat én het 'op tijd' af te werken! Binnenkort heb ik weer de tijd om mijn aandacht ten volle op je komst te richten!

Opgedragen aan Bart,
en aan Rien en zijn toekomstige broer of zus

Els Verfaillie
april 2008

Summary

The marine environment is subjected to increasing anthropogenic pressure. Economic activities, such as shipping, fisheries, aggregate extraction and windmill parks all have their own impact on the seabed and its related biodiversity. Although there is a willingness of the different users to minimize their impacts, there is a strong need for the assessment of the ecological value of the seabed, comprising both the abiotic substrate and the living organisms related to it (together called a 'habitat').
Therefore, 'habitat mapping' is crucial, not only for the assessment of the ecological value at a certain moment, but also to follow its evolution over time. Habitat mapping is used increasingly in the context of marine spatial planning, the designation of new Marine Protected Areas (MPAs), the implementation of national and international legislation and directives (e.g. European Habitats Directive), baseline studies and the planning of monitoring activities.

Because of the world-wide application of marine habitat mapping, there is currently a great variety in approaches, methodologies to use, as well as in the ways habitats are classified. Therefore, it is of utmost importance that attempts are being made to propose more 'common approaches' in marine habitat mapping.

The Belgian part of the North Sea (BPNS) has a surface area of only 3600 km², is relatively shallow (0 to -46 m) and is characterized by a highly variable topography, with a series of sandbanks and swales. The pressure of anthropogenic activities is very high, leaving little space for the 'natural users' of the seabed.
Although habitat mapping is only recently gaining importance, there is a long tradition on the BPNS of the collection of datasets of both the abiotic environment and the biotic organisms. Regarding habitat mapping, the most important datasets include sedimentological and biological grab samples; acoustical data of bathymetry and backscatter; and hydrodynamical data. The grab samples correspond with ground-truth data for the validation of abiotic coverages, being mostly acoustical and hydrodynamical datasets. Despite this huge amount of datasets, it is important that the analysis, processing and integration to habitat maps is done in a more standardized, transparant and scientifically sound manner.

The general aim of this study was to apply and develop standardized, transparant and statistically sound methodologies for highly reliable sedimentological and habitat modelling and mapping, in support of a more sustainable management of our seas.

To achieve these aims, this thesis is subdivided into 2 themes: 1) Best coverage data for habitat mapping; and 2) Integration of datasets in view of habitat mapping.
The 2 themes consist out of 2 (Chapter 2 and 3) and 3 (Chapter 4, 5 and 6) papers. These were published in or submitted to peer-reviewed international journals.

In Theme 1, emphasis is put on the creation of highly reliable sedimentological maps of the grain-size and the silt-clay percentage. The sedimentological maps are based on multivariate geostatistical techniques and more specifically Kriging with an external drift (Goovaerts 1997). Because of a linear relation between the sedimentological point data and one or more secondary datasets, this secondary dataset can assist into the interpolation and improve the final result.
In soft substrata sandy habitats, such as the BPNS, those coverages are crucial for the predictive modelling of macrobenthos (sea bottom inhabiting organisms larger than 1 mm) (e.g. Wu and Shin 1997; Van Hoey et al. 2004; and Willems et al. 2008).

In **Chapter 2**, the median grain-size of the sand fraction on the BPNS is interpolated with the bathymetry as secondary information.

Chapter 3 applies the same methodology, but in a more advanced approach. Based on data from a small study area on the BPNS, four sedimentological variables are interpolated using a whole set of secondary variables. These variables are derivatives from the bathymetry and are calculated over different spatial scales.

In both chapters, a validation against standard interpolation techniques reveals that the obtained results are significantly better than without the secondary information. The resulting maps reflect well the natural geomorphology of the seabed.

In Theme 2, the sedimentological maps and other abiotic coverages are integrated to obtain habitat maps.

In **Chapter 4**, a marine landscape map is produced, dividing the BPNS into discrete ecological units, based solely on geophysical data (Roff and Taylor 2000). For a more objective selection of the abiotic coverages, various datasets are used and clustered. Principal Components Analysis allowed reducing the datasets avoiding correlation between the coverages. The optimal number of marine landscapes is defined by a statistical index. The validation of the internal cluster consistency shows that the result of the clustering of the abiotic coverages is very reliable. The ecological validation confirms the ecological relevance of the marine landscapes on the BPNS.

In Chapter 5 and 6, habitat suitability models (HSMs) are developed, which indicate the suitability of a certain area for a specific species or community.

Chapter 5 applies Discriminant Function Analysis (DFA), resulting in a HSM for the 4 macrobenthic communities on the BPNS, being species associations (the *Macoma balthica*, *Abra alba*, *Nephtys cirrosa* and *Ophelia limacina* community; Degraer et al. 2002; Van Hoey et al. 2004). A three-fold cross-validation and two validation indices show that the agreement between the model predictions and observations is very good and consistent. The HSM of the ecologically important *A. alba* community is the most relevant model result, possibly serving as input for biological valuation (Derous et al. 2007a) or for a for the designation of potentially valuable seabed habitats.

Finally, in **Chapter 6**, Ecological Niche Factor Analysis is tested on the BPNS to predict the most suitable habitats of the species *Owenia fusiformis*. This species is capable of forming dense aggregations or patches and is one of the 10 most abundant species (ind/m²) of the *A. alba* community (Fromentin et al. 1997; Van Hoey et al. 2004). The chapter demonstrates that not only sedimentological and bathymetrical coverages, but also multi-scale topographical coverages are important predictors of the occurrence of this species.

Finally, in the Discussion, results are intercompared, overlain and integrated.

The reliability of the habitat maps is calculated following a multi-criteria approach (Foster-Smith et al. 2007b), resulting in reliability scores between 66 and 77 %.

The methodologies of Chapter 2 (multivariate geostatistics) and Chapter 5 (DFA) are tested successfully on a part of the Southern North Sea (i.e. extending beyond the BPNS) to obtain maps of the median grain-size, silt-clay percentage and HSMs of the 4 macrobenthic communities.

Based on the HSM of the ecologically important *A. alba* community (Chapter 5), combined with the potential gravel areas from the marine landscapes (Chapter 4) and from Van Lancker et al. (2007), propositions are made for potentially valuable seabed habitats on the BPNS.

Dutch summary – Nederlandse samenvatting

Het mariene milieu is onderhevig aan een steeds toenemende antropogene druk. Economische activiteiten zoals scheepvaart, visserij, aggregaatextractie en windmolenparkken hebben elk hun eigen impact op de zeebodem en de daaraan gerelateerde biodiversiteit. Alhoewel er een bereidwilligheid is van de verschillende activiteiten om hun impacten te minimaliseren, is er een sterke nood aan de opvolging van het zeebodemmilieu en zijn ecologische waarde met zowel aandacht voor het abiotische substraat als voor de levende organismen die eraan gelinkt zijn (samen 'habitat' genoemd).
Habitatkartering is aldus cruciaal; niet alleen voor de opvolging van de ecologische waarde op een bepaald moment, maar ook voor de opvolging van de evolutie in de tijd. Habitatkartering wordt dan ook toenemend toegepast in de context van mariene ruimtelijke planning, de aanduiding van nieuwe Mariene Beschermde Gebieden (*Marine Protected Areas* of MPAs), de implementatie van nationale en internationale wetgeving en richtlijnen (bv. de Europese Habitatrichtlijn), *baseline* studies en de planning van monitoringsactiviteiten.

Omwille van de vele initiatieven rond habitatkartering, verspreid over de wereld, is er een veelheid in benaderingen, methodologieën en in de classificatie van habitats ontstaan. Daarom is het van het grootste belang dat er gestreefd wordt naar meer 'gemeenschappelijke benaderingen' voor mariene habitatkartering.

Het Belgisch deel van de Noordzee (BDNZ) heeft een oppervlakte van slechts 3600 km², is relatief ondiep (0 tot -46 m) en wordt gekenmerkt door een sterk variërende topografie, met een opeenvolging van zandbanken en geulen. De druk van antropogene activiteiten is dus zeer hoog, zodat er weinig ruimte overblijft voor de 'natuurlijke gebruikers' van de zeebodem.
Alhoewel habitatkartering slechts recent aan belang heeft gewonnen, is er op het BDNZ reeds een lange traditie in de verzameling van datasets met betrekking tot zowel het abiotische milieu als de biotische organismen. In de context van habitatkartering omvatten de belangrijkste datasets sedimentologische en biologische staalnames; akoestische gegevens van de bathymetrie en *backscatter*; en hydrodynamische data. De staalnames komen overeen met *ground-truth* data voor de validatie van abiotische *coverages*, namelijk de akoestische en hydrodynamische datasets. Ondanks deze veelheid aan gegevens, is het belangrijk dat de analyse, verwerking en integratie tot habitatkaarten op een meer gestandaardiseerde en wetenschappelijk verantwoorde manier gebeurt.

Het algemene doel van dit onderzoek is de toepassing en ontwikkeling van doordachte en statistisch verantwoorde methodes om tot een betrouwbare modellering en kartering van de sedimentologie en het habitat te komen en dit ter ondersteuning van een duurzamer beheer van onze zeeën.

Om deze doelstellingen te bereiken, is de thesis opgedeeld in 2 thema's: 1) Beste *coverage* data voor habitatkartering; en 2) Integratie van datasets voor habitatkartering.
De 2 thema's omvatten respectievelijk 2 (Hoofdstuk 2 en 3) en 3 (Hoofdstuk 4, 5 en 6) artikels, die gepubliceerd zijn in of ingediend zijn bij *peer-reviewed* internationale tijdschriften.

In Thema 1 ligt de nadruk op het creeëren van heel betrouwbare sedimentologische kaarten van de korrelgrootte en het silt-klei percentage. De sedimentologische kaarten zijn gebaseerd op multivariate geostatistische technieken en meer specifiek Kriging met een externe drift (Goovaerts 1997). Omdat er een lineaire relatie bestaat tussen de sedimentologische puntgegevens en één of meerdere secundaire datasets, kan deze secundaire dataset assisteren bij de interpolatie om zo het finale resultaat te verbeteren.
In habitats van zachte zandige substraten, zoals het BDNZ, zijn deze *coverages* cruciaal voor het voorspellen van het macrobenthos (zeebodemgerelateerde organismen groter dan 1 mm) (bv. Wu en Shin 1997; Van Hoey et al. 2004; en Willems et al. 2008).

In **Hoofdstuk 2**, wordt de mediane korrelgrootte van de zandfractie op het BDNZ geïnterpoleerd, met de bathymetrie als secundaire informatie.
In **Hoofdstuk 3** wordt dezelfde methodologie toegepast, zij het op een meer geavanceerde manier. Hier worden 4 sedimentologische variabelen van een klein studiegebied op het BDNZ geïnterpoleerd, dit keer met een set van secundaire variabelen die afgeleiden zijn van de bathymetrie op verschillende ruimtelijke schalen.
In beide hoofdstukken, wordt bovendien een validatie uitgevoerd ten opzichte van standaard interpolatietechnieken. Dit wijst op significant betere resultaten voor de interpolatie mét secundaire datasets. Bovendien geven de resulterende kaarten goed de natuurlijke geomorfologie van de zeebodem weer.

In Thema 2 worden de sedimentologische kaarten en andere abiotische *coverages* geïntegreerd tot habitatkaarten.
In **Hoofdstuk 4** wordt een mariene landschapskaart aangemaakt, die het BDNZ verdeelt in discrete ecologische eenheden; louter gebaseerd op geofysische data (Roff en Taylor 2000). De nadruk ligt op de uitwerking van een objectieve methodologie. Deze kaart is gebaseerd op de clustering van een set van abiotische *coverages*, die gereduceerd zijn door een Principale Componenten Analyse om correlatie tussen de *coverages* te vermijden. Het optimale aantal van mariene landschappen wordt bepaald aan de hand van een statistische index. Bovendien toont de validatie van de interne clusterconsistentie aan dat het resultaat van de clustering van de abiotische *coverages* heel betrouwbaar is. De ecologische validatie bevestigt bovendien dat de mariene landschappen ecologisch relevant zijn.
In Hoofdstuk 5 en 6, worden habitatgeschiktheidsmodellen (*habitat suitability models* of HSMs) aangemaakt, die de geschiktheid weergeven van een bepaald gebied voor een specifieke soort of gemeenschap.
In **Hoofdstuk 5** wordt Discriminant Functie Analyse (DFA) toegepast, wat resulteert in een HSM voor de 4 macrobenthische gemeenschappen op het BDNZ, overeenkomend met soortenassociaties (de *Macoma balthica*, *Abra alba*, *Nephtys cirrosa* en *Ophelia limacina* gemeenschap; Degraer et al. 2002; Van Hoey et al. 2004). Een drievoudige kruisvalidatie en 2 validatie-indices tonen aan dat de overeenkomst tussen de modelvoorspellingen en de observaties heel goed en consistent is. De HSM van de ecologisch belangrijke *A. alba* gemeenschap is het meest waardevolle resultaat, omwille van zijn mogelijke waarde als input voor biologische waardering (Derous et al. 2007a) of voor een eerste voorstel van de aanduiding van nieuwe MPAs.

In **Hoofdstuk 6** wordt Ecologische Niche Factor Analyse uitgetest op het BDNZ om de habitatgeschiktheid te voorspellen van de soort *Owenia fusiformis.* Deze soort is één van de meest abundante soorten (ind/m²) van de *A. alba* gemeenschap (Fromentin et al. 1997; Van Hoey et al. 2004) en is in staat dense aggregaties te vormen. Het hoofdstuk toont aan dat niet alleen de sedimentologie en de bathymetrie, maar ook andere topografische *coverages* op verschillende ruimtelijke schalen belangrijke voorspellers zijn voor deze soort.

In de Discussie worden tenslotte resultaten van dit onderzoek vergeleken, over mekaar gelegd en geïntegreerd.
De betrouwbaarheid van de habitatkaarten wordt berekend op basis van een multi-criteria benadering (Foster-Smith et al. 2007b), wat resulteert in betrouwbaarheidsscores tussen 66 en 77 %.
De methodologieën van Hoofdstuk 2 (multivariate geostatistiek) en Hoofdstuk 5 (DFA) kunnen succesvol worden toegepast op een deel van de Zuidelijke Noordzee (i.e. over een ruimer gebied dan het BDNZ), om zo tot kaarten te komen van de mediane korrelgrootte, het silt-klei percentage en HSMs van de 4 macrobenthische gemeenschappen.
Bovendien wordt op basis van het HSM van de ecologisch belangrijke *A. alba* gemeenschap (Hoofdstuk 5), gecombineerd met de potentiële grindgebieden van de mariene landschappen (Hoofdstuk 4) en van Van Lancker et al. (2007), voorstellen geformuleerd voor de aanwijzing van mogelijke nieuwe MPAs.

Structure of the thesis

This thesis is structured as follows:

- **Chapter 1: Introduction**

Chapter 1 gives an overview of the objectives and the general approach of the thesis, followed by a short literature review dealing with similar topics. It gives an introduction on the Belgian part of the North Sea (BPNS), the study area of this thesis.

Chapter 2 until 6 are structured into 2 themes based on different approaches in the context of habitat mapping. All of the chapters are either published or submitted papers in peer-reviewed journals.

Theme 1: BEST COVERAGE DATA FOR HABITAT MAPPING

High quality sedimentological maps were produced, being very important coverage data in soft substrate habitats.

- **Chapter 2: Multivariate geostatistics for the predictive modelling of the surficial sand distribution in shelf seas**

This paper focuses on a methodology to produce a map of the median grain-size of the BPNS. Different interpolation techniques are compared and validated. Kriging with an External Drift (KED) that makes use of secondary information to assist into the interpolation, gave best results since a linear correlation was found between the median grain-size and the bathymetry.

→ *Verfaillie, E., Van Meirvenne, M. and Van Lancker, V., 2006. Multivariate geostatistics for the predictive modelling of the surficial sand distribution in shelf seas, Continental Shelf Research, 26 (19), 2454-2468.*

- **Chapter 3: Geostatistical modelling of sedimentological parameters using multiscale terrain variables: application along the Belgian Part of the North Sea**

As in Chapter 2, this paper uses multivariate geostatistics (KED) to interpolate different sedimentological maps of the northern part of the Vlakte van de Raan of d_s10, d_s50, d_s90 and the silt-clay percentage. This time the methodology uses different datasets as secondary information to assist into the interpolation (all secondary datasets are based on multibeam bathymetry and its topographical derivatives).

→ *Verfaillie, E., Du Four, I., Van Meirvenne, M. and Van Lancker, V., accepted. Geostatistical modelling of sedimentological parameters using multiscale terrain variables: application along the Belgian Part of the North Sea. International Journal of Geographical Information Science.*

Theme 2: INTEGRATION OF DATASETS IN THE VIEW OF HABITAT MAPPING

The coverage data are integrated with biological data, to come to habitat maps.

- **Chapter 4: A protocol for classifying ecologically relevant marine landscapes, a statistical approach**

This paper presents an objective and statistically justified methodology to classify the abiotic environment of the BPNS into 'Marine Landscapes'. Abiotic datasets are subjected to a statistical approach, using principal components analysis and a cluster analysis. The final model results classified the BPNS into 8 marine landscapes that represent well the natural variability of the seafloor.

→ *Verfaillie, E., Degraer, S., Schelfaut, K., Willems, W. and Van Lancker, V. (submitted to Estuarine, Coastal and Shelf Science) A protocol for classifying ecologically relevant marine landscapes, a statistical approach.*

- **Chapter 5: Habitat suitability as a mapping tool for macrobenthic communities: An example from the Belgian part of the North Sea**

This chapter shows for the BPNS how the map of the median grain-size (Chapter 2), combined with the map of the silt-clay percentage could be translated into four habitat suitability maps of macrobenthic communities using Discriminant Function Analysis.

→ *Degraer, S., Verfaillie, E., Willems, W., Adriaens, E., Van Lancker, V. and Vincx, M., 2008. Habitat suitability as a mapping tool for macrobenthic communities: An example from the Belgian part of the North Sea. Continental Shelf Research 28(3), 369-379.*

- **Chapter 6: The relevance of ecogeographical variables for marine habitat suitability modelling of *Owenia fusiformis***

Chapter 6 applies habitat suitability modelling of the polychaete *Owenia fusiformis* based on Ecological Niche Factor Analysis on the entire BPNS, with datasets having a resolution of 250 m. Different combinations of abiotic datasets are compared and validated.

→ *Verfaillie, E., Degraer, S., Du Four, I., Rabaut, M., Willems, W. and Van Lancker, V. (submitted to Estuarine, Coastal and Shelf Science) The relevance of ecogeographical variables for marine habitat suitability modelling of* Owenia fusiformis.

- **Chapter 7: Discussion**

The last major chapter gives a short summary of the objectives of this thesis and shows how the results of the papers gave an answer to the major questions. It stresses the importance of an objective and statistically sound approach. The habitat maps of the different chapters are overlain, compared and discussed. Furthermore, the difference between habitat maps for scientific purposes and for end users and management is discussed.

- **Chapter 8: Conclusion**

Chapter 1

Introduction

1 Introduction

1.1 Habitat mapping

1.1.1 Context

Similar to terrestrial landscapes, the underwater world is highly diverse. Various processes are responsible for the current status of the seafloor. Knowledge on how the marine landscapes look today and in the past, how they came into existence and evolved is very important for humans, because the seabed contains an enormous richness of living and non-living resources.
For a sustainable management of the marine environment, mapping the seafloor or 'habitat mapping' is crucial. It is important in the context of marine spatial planning (e.g. delineation of aggregate extraction zones, windmill parks and marine protected areas or MPAs); the protection of specific species or communities (e.g. species listed in the Habitats Directive); and for the overall improvement of the scientific knowledge base.

This thesis aims at developing **straightforward and statistically sound methodologies** for highly reliable **sedimentological and habitat modelling and mapping, in support of a more sustainable management of our seas.**

1.1.2 From a habitat to habitat mapping

'Habitat' and 'habitat mapping' are terms that have been used since decades, although different terms still exist for similar things.
For the term '**habitat**', a whole range of definitions exist. The ICES Working Group on Marine Habitat Mapping (ICES 2006) gave an overview of definitions of a habitat starting with the classical definition of Darwin (1859) that considered only *"The locality in which a plant or animal naturally lives."*. The final ICES definition was based on definitions of Allee et al. (2000), EUNIS (2002), Kostylev et al. (2001) and Valentine et al. (2005) and is the one that will be used for this study: ***"A particular environment which can be distinguished by its abiotic characteristics and associated biological assemblage, operating at particular, but dynamic spatial and temporal scales in a recognizable geographic area."***. As such, it is clear, that a habitat is the combination of both the abiotic and the biotic environment (Figure 1.1).

In the framework of the MESH Project (Development of a Framework for Mapping European Seabed Habitats, 2004-2007), a MESH Guide to Marine Habitat Mapping (MESH Project 2007) has been proposed, describing the whole process of marine habitat mapping. In the first chapter of this guide (Foster-Smith et al. 2007a), '**habitat mapping**' has been defined as: "**Plotting the distribution and extent of habitats to create a complete coverage map of the seabed with distinct boundaries separating adjacent habitats**.". Furthermore, it is stressed that a habitat map is "**a statement of our best estimate of habitat distribution at a point in time, making best use of the knowledge we have available at that time.**".

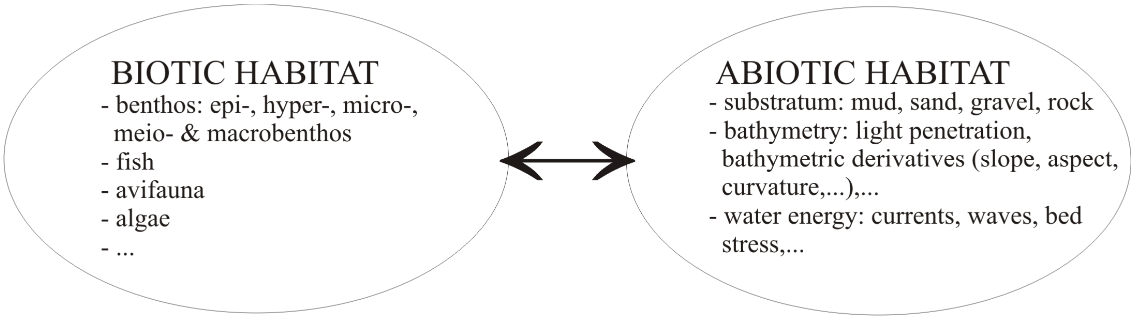


Figure 1.1: A marine habitat consists of a biotic and an abiotic part (ICES 2006). The biotic part consists of all marine fauna and flora, while the abiotic part consists of characteristics related to the substratum, bathymetry and water energy.

1.1.3 How to make a habitat map?

Van Lancker and Foster-Smith (2007) created a scheme with the main stages in the making of a habitat map by integrating sample data and full coverage abiotic data. This scheme comprises 4 steps (Figure 1.2): (1) getting the best out of the **ground-truth** data; (2) getting the best out of the **coverage** data; (3) **integration** of the ground-truth and the coverage data; and (4) habitat map **design** and **lay-out**.
Coverage data are all kinds of full coverage abiotic datasets, ranging from remote sensing data (e.g. acoustical or satellite imagery), sedimentological maps, hydrodynamical models etc. (see Coggan et al. 2007; and Van Lancker and Foster-Smith 2007 for an overview). Ground-truth data are needed for interpolating and validating the coverage data and for assigning ground types to the mapped regions. The methods for ground-truthing range from grab or core samples to video or diving datasets (see Coggan et al. 2007; and Van Lancker and Foster-Smith 2007 for an overview).
The scheme summarizes well that habitat mapping is very complex. Still, the process can even be more complex than the scheme suggests.

Figure 1.2: Scheme of the habitat mapping process, containing 4 steps (Van Lancker and Foster-Smith 2007).

This research focuses mainly on step 2 and step 3. Chapter 2 and 3 in this thesis can be categorized under step 2. Still, the result of these chapters is not a habitat map, but a high quality model of the abiotic environment, being a crucial intermediate stage in the habitat mapping process. In **Chapter 2** and **3**, a complex multivariate

geostatistical interpolation method is applied to produce high quality sedimentological maps. Still, both chapters integrate also in a way ground-truth data with coverage data (step 3), because sedimentological samples are interpolated using bathymetric data (and derivatives) as secondary information. However, in most cases the integration of coverage data and ground-truth data means that relations are sought between biotic ground-truth data and abiotic coverage data.
Chapter 4, 5 and 6 of this thesis can be categorized under step 3. In **Chapter 4**, a methodology is worked out to combine different abiotic datasets in an objective and statistically sound way. This results in a 'Marine Landscapes' map. Chapter 4 and 5 deal with habitat suitability models of macrobenthic communities and the species *Owenia fusiformis* respectively. The model in **Chapter 5** is based on Discriminant Function Analysis to predict the habitat suitability of four macrobenthic communities. In addition, a habitat suitability map, has also been translated into one classified habitat map, showing EUNIS classes (a pan-European habitat classification system, see further) (Schelfaut et al. 2007). **Chapter 6** uses Ecological Niche Factor Analysis to predict the habitat suitability of *Owenia fusiformis*.

1.1.4 Habitat classification versus habitat suitability modelling

The classical way of habitat mapping, as defined by Foster-Smith et al. (2007a) is based on a **habitat classification**. This means that an area is subdivided into several groups or classes, 'separated by distinct boundaries'. Numerous classification systems exist, ranging from internationally accepted to very locally used classifications.
There are two kinds of habitat classification: **top-down** and **bottom-up** (Van Lancker and Foster-Smith 2007). A top-down approach is used, when an existing classification system is applied on a dataset, matching ground-truth samples to pre-defined classes. When the ground-truth data are used to determine habitat classes (finding new associations between biological and abiotic data), this is called a bottom-up approach.
An important example of a top-down habitat classification is the mapping of '**marine landscapes**'. This approach was developed for areas where biological samples are absent or very scarce (e.g. offshore or deep-sea areas). Therefore, a concept was firstly proposed for Canadian waters by Roff and Taylor (2000) and Roff et al. (2003), classifying abiotic datasets into marine landscapes. Biological data are only used passively, to validate the ecological relevance of the marine landscapes afterwards. The combination of the abiotic datasets into different classes was performed originally in a Geographic Information System (GIS). Examples of this approach can be found also for the UK (Connor et al. 2006) and for the Baltic Sea (Al-Hamdani and Reker 2007). This concept and the related issues are discussed in detail in Chapter 4, in which a more objective and statistically sound approach is suggested.
Some classification systems strive for a standardisation and a more objective, systematic approach for habitat classification. A well known example of such a habitat classification system is the **EUNIS classification** (European Nature Information System), a pan-European system, which was developed between 1996 and 2001 by the European Environment Agency, in collaboration with European-wide experts. It incorporates the classification of marine, coastal and terrestrial habitats with the marine part being based on the National Marine Habitat Classification for Britain and Ireland (Connor et al. 2004) and the North-East Atlantic classification, developed for the OSPAR Convention in 2004. The EUNIS classification consists of **six hierarchical levels**. The first level is the separation between marine, coastal and

terrestrial habitats. Level 2 and 3 are based purely on abiotic characteristics. From level 4 to 6, references to specific biological taxa are introduced (for an overview, see Foster-Smith et al. 2007a).
Other internationally accepted classification systems have been developed by Greene et al. (1999) and Valentine et al. (2005) for North America.
Although there is a strong need for a standardization of national, regional and local habitat mapping programmes (ICES 2007), most examples of marine habitat mapping in literature are based on **national, regional and local classification systems**:

- **Europe**: Sotheran et al. (1997); Service (1998); Brown et al. (2002); Freitas et al. (2003a); Freitas et al. (2003b); Bates and Oakley (2004); Kobler et al. (2006); and Brown and Collier (2008);
- **Central America**: Mumby and Harborne (1999); and Mishra et al. (2006);
- **North America**: Zacharias et al. (1999); Roff and Taylor (2000); Zajac et al. (2000), Kostylev et al. (2001); Anderson et al. (2002); Cochrane and Lafferty (2002); Edwards et al. (2003); Franklin et al. (2003); Roff et al. (2003); Zajac et al. (2003); Dartnell and Gardner (2004); Ojeda et al. (2004); Lathrop et al. (2006); and Rooper and Zimmermann (2007);
- **Oceania**: Banks and Skilleter (2002); and Porter-Smith et al. (2004);
- **Pacific Ocean**: Lundblad et al. (2006); and Gregr and Bodtker (2007).

Unlike the definition of habitat mapping, there is a trend in habitat mapping to omit distinct boundaries or habitat classes, but rather uses a continuously grading scale. In these cases, the suitability is shown (e.g. on a scale 0 – 1 or 0 – 100) for a single species or community and is often called a '**habitat suitability model**' (HSM) (for an overview, see Willems et al. in prep.). In general, HSMs are based on a range of statistical techniques, such as regression, environmental envelopes or neural networks (for an overview, see Guisan and Zimmerman 2000), integrating the biological data with the abiotic or ecogeographical variables (EGVs) from the beginning (i.e. bottom-up approach).
Although HSMs are, in general, based on statistical and thus objective methods, no or only few international standards exist. As such, most case studies are also of a national, regional or local nature. Some examples of marine HSMs are:

- **Arctic Ocean**: Jerosch et al. (2006); and Jerosch et al. (2007);
- **Europe**: Eastwood et al. (2001); Brinkman et al. (2002); Le Pape et al. (2003); Nicolas et al. (2007); Wilson et al. (2007); Degraer et al. (2008); Skov et al. (2008); and Willems et al. (2008);
- **North America**: Iampetro and Kvitek (2005); Bryan and Metaxas (2007);
- **Oceania**: Bradshaw et al. (2002).

The most important advantages and disadvantages, of both habitat classification and HSMs, are given in Table 1.1.

Chapter 2 and 3 of this thesis resulted in sedimentological coverages as input for both a habitat classification of marine landscapes (Chapter 4) and HSMs (Chapter 5 and 6). All results are mapped on a national and a local level, although the HSMs of the macrobenthic communities of Chapter 5 have been translated to a EUNIS level 5 map (Schelfaut et al. 2007), the pan-European classification system.

Table 1.1: Advantages and disadvantages of habitat classification versus habitat suitability modelling.

	Advantages	Disadvantages
Habitat classification	- simple methodologies - international classification systems exist - not limited to a single species or community	- in general no statistical methods: subjective - existing international classification systems not suitable for all datasets - sometimes artificial boundaries between habitat classes
Habitat suitability models	- statistical methods: more objective - continuous scale: no artificial boundaries	- complex methodologies - limited to a single species or community

1.1.5 Difficulties in habitat mapping

With an increasing number of habitat mapping initiatives, a growing need exists for harmonized approaches. The main difficulties that need to be dealt with relate to:

i. Numerous habitat mapping methodologies and approaches exist, but there is a strong need for **standardisation** and for a **common approach** to a more coherent mapping of wider areas and to improve consistency towards management and decision making.
ii. Methodologies are often based on 'expert judgement' and subjective decisions (e.g. traditional method of marine landscape mapping in which abiotic datasets and their class breaks have to be chosen and in which the number of landscapes is dependent on the combination of all the classes of the different datasets; e.g. 3 sediment classes and 2 bathymetry classes result already into 6 marine landscapes). As such, there is a strong need for more **objective methodologies**.
iii. The **choice of abiotic or ecogeographical variables** (EGVs), defining the abiotic habitat (substrate, bathymetry, energy and related variables), as input for habitat mapping studies, is often difficult. The variables have to be ecologically relevant, but in some cases, the relationship between the biological data and the EGVs is not known.
iv. Moreover, the reliability of EGVs and habitat maps is highly variable. EGVs based on 100 or 20 sedimentological samples logically do not have the same reliability. As such, there is a need for **tools to estimate the EGV reliability**.
v. Different spatial scales of EGVs can be related to the occurrence of species or communities (e.g. a sandbank may be superimposed by dunes, ranging from small to very-large (Ashley 1990); a species can have a preference for certain topographic locations on the sandbank or on the dunes). There is a need for a **multi-scale approach** regarding EGVs that predict the occurrence of species or communities.

Above, only those issues have been enumerated that were dealt with throughout this study. However, many other aspects remain important; these relate mainly to:

i. Habitat mapping studies performed on **different spatial scales**, going from fine- to intermediate- to broad-scale (cfr. Van Lancker and Foster-Smith 2007), are based on datasets of different spatial resolution and result into habitat maps that partly do **not overlap**.
ii. As mentioned before, most studies are of a local, regional or national nature. As such, datasets of different resolutions and qualities are used, resulting into **transborder problems**, problems of non-overlapping habitat classes and different classification systems.
iii. Moreover, different areas have **different priorities** of species and communities that are in need of protection.
iv. A last problem is related to the **temporal scale**. A habitat map is mostly represented as a final, unchanging result. Logically, habitats of species and communities change in time due to natural (e.g. seasonal) and anthropogenic (e.g. destruction of habitats by fishery impact) variation.

1.1.6 <u>Research strategy</u>

The general objective of this thesis is to develop **spatial distribution models** as input for marine habitat mapping. The spatial distribution models concern both the production of high quality **physical coverages** as well as the **integration of ground-truth data and coverages**, based on (geo)statistical methods. All models were **validated and intercompared**.

Particularly, the objectives anticipate to the following needs in habitat mapping:

A **new approach for marine landscape mapping** is proposed, which is simple, statistically sound and easy applicable to other regions. The proposed methodology is a step forward in standardising marine landscape mapping throughout Europe.

Methodologies are developed that are straightforward, **objective** and **statistically sound**. For the mapping of sediment distribution, multivariate geostatistics are applied (Chapter 2 and 3); and for the mapping of marine landscapes, the combined use of principal components and cluster analysis (Chapter 4) is proposed. Moreover, the physical coverages (sedimentological maps, multi-scale derivatives of the bathymetry and other coverages) are used as input for two kinds of HSMs, predicting the occurrence of macrobenthic communities and species (Chapter 5 and 6).

A **maximal input of EGVs**, is used for the modelling of the marine landscapes and for the HSMs (Chapter 4, 5 and 6). This is possible with factor analysis (both Principal Components Analysis and Ecological Niche Factor Analysis), transforming the correlated datasets into linear combinations of the original EGVs.

The reliability of EGVs is optimised by applying **multivariate geostatistics** for the sedimentological maps (Chapter 2 and 3). **Validation** of the sedimentological maps, marine landscapes map and the HSMs is performed (Chapter 2, 3, 4, 5 and 6).

Multi-scale topographical EGVs, derived from the bathymetry, can be used for both the modelling of sedimentological maps (Chapter 3) and for the modelling of the HSMs of the species *Owenia fusiformis* (Chapter 6).

The Belgian part of the North Sea (BPNS) served as an ideal 'test case' for all of these methodologies, because of its high amount of datasets available. The datasets concern both the abiotic and the biotic environment. The main sources of 'raw' information that were important for this research were:

- Abiotic datasets:
  - Surficial sediment data, extracted from the sedimentological database 'sedisurf@' (hosted at Ghent University, Renard Centre of Marine Geology), with samples covering the entire BPNS (of importance in Chapter 2, 3, 4, 5 and 6);
  - Bathymetric data of the BPNS (based on single beam acoustics, obtained from the Flemish Authorities, Agency for Maritime and Coastal Services, Flemish Hydrography) (of importance in Chapter 2, 3, 4, 5 and 6);
  - Bathymetric data of the study area of Chapter 3 and 6 (based on multibeam acoustics), acquired by Ghent University, Renard Centre of Marine Geology;
  - Hydrodynamical data of the BPNS, modelled by the Management Unit of the North Sea, Mathematical Models and the Scheldt Estuary (of importance in Chapter 4 and 6).
- Biological dataset:
  - Macrobenthic database 'Macrodat' (Marine Biology Section, Ugent – Belgium, 2008), with samples covering the entire BPNS (of importance in Chapter 4, 5 and 6).

1.2 Study area

This section provides an introduction to the environmental datasets that have been used in the context of habitat mapping along the Belgian part of the North Sea (BPNS). In addition, it describes the relevant legal framework.

1.2.1 General seabed characterization

The BPNS is part of the Greater North Sea and is situated on the north-west European Continental Shelf (Figure 1.3). Its surface area is 3600 km², which represents hardly 0.6 % of the north-west European shelf. The BPNS is characterized by its relative shallowness. The depth of the seabed ranges from 0 m to -46 m (Mean Lowest Low Water at Spring, MLLWS) (Figure 1.4). In the coastal zone (10-20 km), depths range between 0 m and -15 m MLLWS, followed by a central zone of -15 m to -35 m. Towards the northern part of the shelf, water depths range between -35 and -50 m MLLWS.
The seabed surface is characterized by a highly variable topography, with a series of sandbanks and swales. Sandbanks are characteristic for continental shelves with a high amount of sand and sufficiently strong currents (Stride 1982). Along the BPNS, numerous large sandbanks occur in parallel groups (Figure 1.3): the Coastal Banks and the Zeeland Banks are quasi parallel to the coastline, whereas the Flemish Banks and the Hinder Banks have a clear offset in relation to the coast. The direction of their asymmetry is mostly to the northeast for the Flemish Banks and to the southwest for the Hinder Banks, although the direction can change along the sandbanks (e.g. Buiten Ratel); the Coastal Banks and the Zeeland Banks have their steep side oriented towards the coast. Some sandbanks have a central kink (Deleu et al. 2004; and Bellec et al. in press). On the BPNS, sandbanks play an important role in natural coastal defense and as a source for marine aggregates (for an overview, see Van Lancker et al. in press).

1.2.2 Geological background

The substratum of the BPNS is composed of solid layers of various ages. The Palaeozoic basement (London-Brabant Massif), flooded since Late Cretaceous times, is covered with a series of Cretaceous, Palaeogene (Tertiary) and Pleistocene and Holocene (Quaternary) sediments (for an overview, see Le Bot et al. 2003).
The Palaeogene deposits Y1 to P1 dip gently (0.5 – 1°) towards the NNE, and the units are superposed from WSW towards ENE; the direction in which they subcrop successively (Figure 1.5) (Le Bot et al. 2003).
The Quaternary sediments are non-cemented, partly relict and partly subject to movement, caused by tidal currents and wave action. Most of the sediments are of Holocene age. Because of an important sediment reworking during the Holocene, only few sediments are considered of Pleistocene age, a period characterized by a succession of glacial and interglacial stages (Le Bot et al. 2003). However, it is possible that within deep incised scour hollows Pleistocene infillings are present (Liu et al. 1993, Stolk 1996, Trentesaux et al. 1999), although some authors assume a Holocene age (Trentesaux 1993, Berné et al. 1994).

The Holocene started 10.000 years ago and can be considered as the present interglacial. During the first part of this period, a sharp sea level rise in the Southern North Sea took place, known as the Flandrian transgression. The Holocene sediments form mainly the present tidal sandbanks (Le Bot et al. 2003).

1.2.3 Morpho-sedimentological characterization

The seabed surface is mainly sandy in nature. Sediments are sorted as a consequence of the interaction between the currents and the specific morphology of the seabed. Generally, sediments coarsen in an offshore direction (Lanckneus et al. 2001). The sand fraction (63 µm - 2 mm; Verfaillie et al. 2006) is found merely on the sandbanks, whereas in the swales, also coarser sands, gravel (> 2 mm) and higher silt-clay fractions (< 63 µm) can occur (Figure 1.6). The depth and the characteristics of the seabed sediments in the swales can differ along the two sides of a sandbank (e.g. Buiten Ratel). The sandbanks, as well as some of the swales, are covered with dune structures. The heights of the dunes differ from one region to another. Tidal action and movement of water masses under changing meteorological conditions are responsible for the displacement of these bedforms.

The coastal area around the harbour of Zeebrugge and Oostende is characterized by very high silt-clay percentages of more than 25 % (Figure 1.7). Furthermore, a gradient from high to low silt-clay percentages occurs in the whole coastal area, increasing from west to east. This zone of higher silt-clay percentage is situated mainly in an area of 20 km from the coastline. Some exceptions are the zones between the northern part of the Buiten Ratel and Oostdyck and between the Goote- and the Thorntonbank.

The Coastal and Flemish Banks are characterized by fine to medium sands with grain-sizes ranging from 63 until 350 µm (Figure 1.6). Higher grain-sizes in this area are found locally (e.g. on the Ravelingen and the Middelkerkebank). 20 km offshore from the Belgian coastline (i.e. at the northwestern side of the Akkaertbank), all sands have grain-sizes coarser than 300 µm, except for some local anomalies (note that the low grain-sizes on the Fairy Bank are due to a lack of samples). Generally, coarser sands characterize the Hinder Banks with grain-sizes of more than 350 µm. In the swales, between the Noordhinder-, Oosthinder- and Bligh Bank, sand coarser than 400 and even 500 µm is found. This is also the case for the most offshore part of the BPNS, north of the Noordhinderbank, where grain-sizes range between 350 and 600 µm (note that the highest grain-sizes are due mainly to shell fragments). Figure 1.8 shows the distribution of coarse sand, together with potential areas of gravel. The potential gravel areas are situated mainly further offshore than 20 km from the coastline; they are concentrated in the swales of the sandbanks. The most important areas of gravel occur between the Oostdyck and Buitenratel (and the continuation of this swale further to the east, between the Goote- and Akkaertbank), between the Goote- and the Thorntonbank and, locally, in the swales of all of the Hinder Banks (with a large concentration between the Westhinder- and Oosthinderbank and near the western parts of the Goote- and Thorntonbank). New insights in their occurrence and origin are described in Deleu and Van Lancker (2007).

The bedforms of the BPNS (Figure 1.9) have heights ranging from 1 to more than 6 m (although the latter are exceptional). Most of the dunes have heights between 1 and 4 m. Ashley (1990) classifies dunes as follows: small dunes, medium dunes, large dunes and very large dunes with spacings of respectively 0.6-5 m, 5-10 m, 10-100 m and

more than 100 m, with heights of respectively 0.075-0.4 m, 0.4-0.75 m, 0.75-5 m and more than 5 m.
To summarize, the trends from the coastline to further offshore (SE to NW) and parallel to the coastline (SW to NE) are given for all of the described parameters (Table 1.2).

Table 1.2: Sedimentological trends along the Belgian part of the North Sea, in an offshore direction and in a direction parallel to the coastline.

	further offshore; SE to NW	parallel to coastline; SW to NE
Median grain-size	coarser	finer
Silt-clay %	lower	higher
Gravel	more	less
Bedforms	more/higher	no trend

Further details on the origin and composition of the sediments and on the morphology can be found in Lanckneus et al. (2001) and Van Lancker et al. (2007).

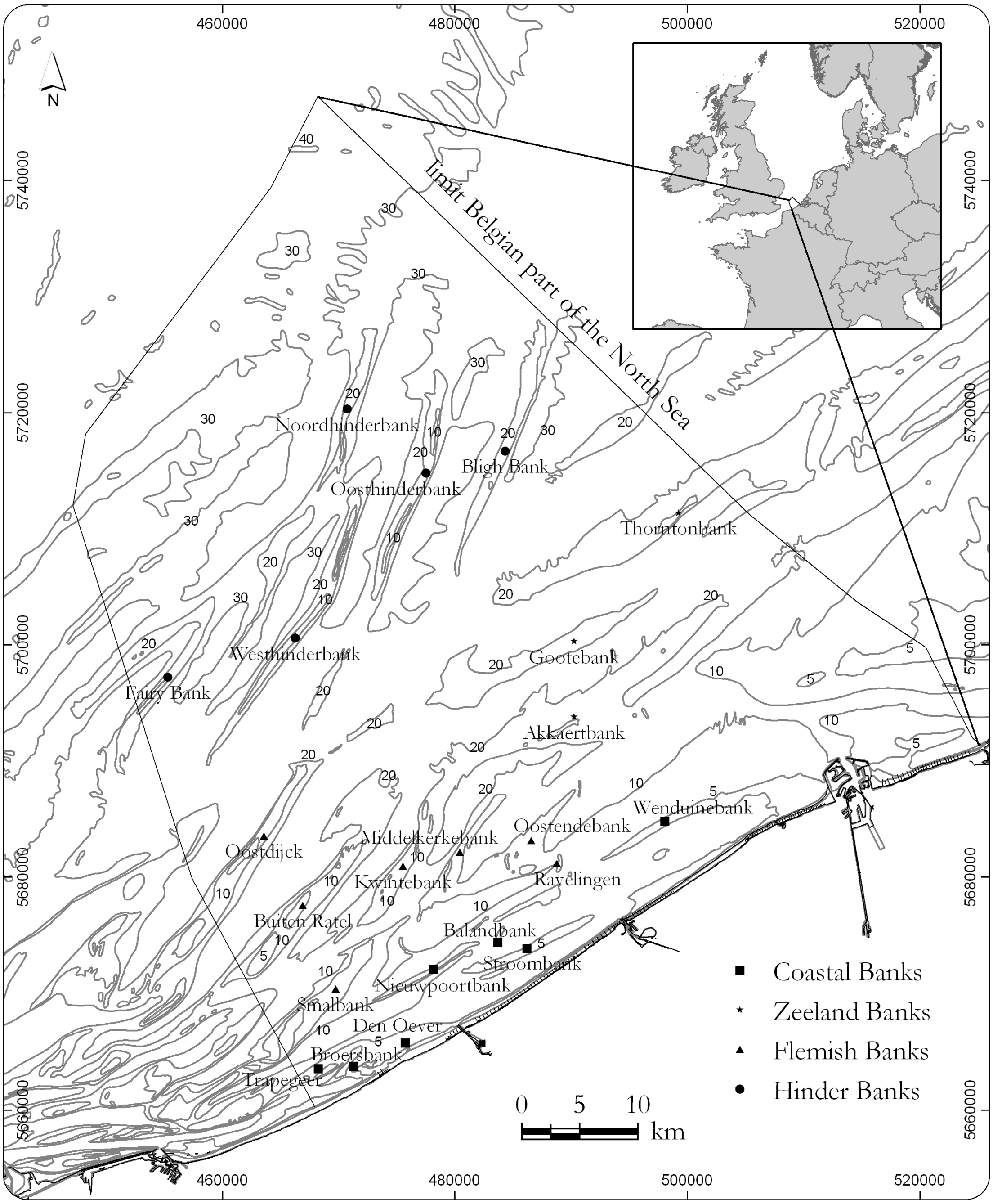


Figure 1.3: General seabed characterization of the Belgian part of the North Sea with the 4 groups of sandbanks. The contour lines of the bathymetry are marked with depth values (MLLWS).

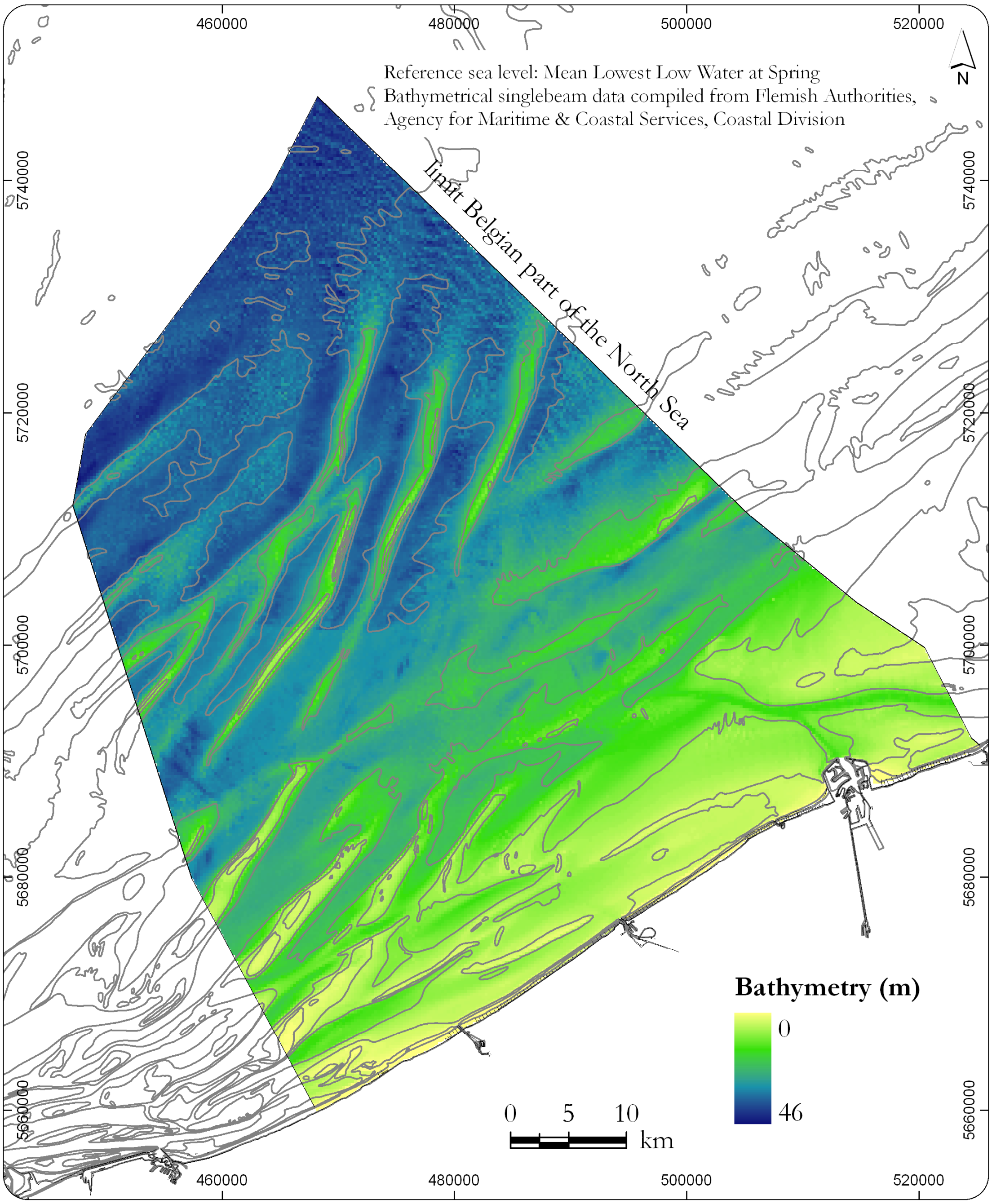


Figure 1.4: Bathymetric map of the Belgian part of the North Sea.

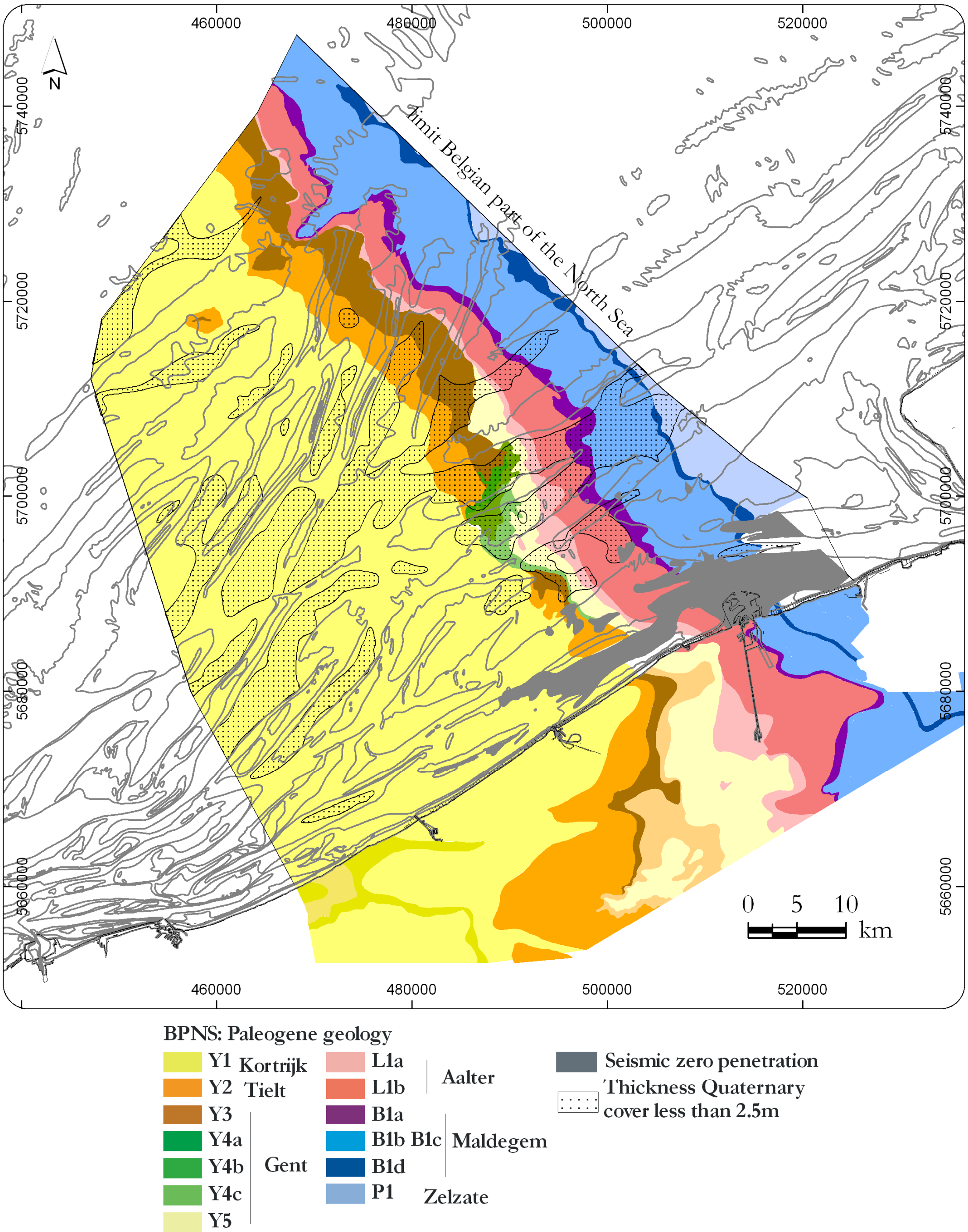


Figure 1.5: Subcrops of solid Paleogene deposits under the non-consolidated Quaternary deposits on the Belgian part of the North Sea (BPNS) (from Le Bot et al. 2003; offshore data: after Maréchal et al. 1986; De Batist 1989; and De Batist and Henriet 1995 / onland data: Jacobs et al. 2002)

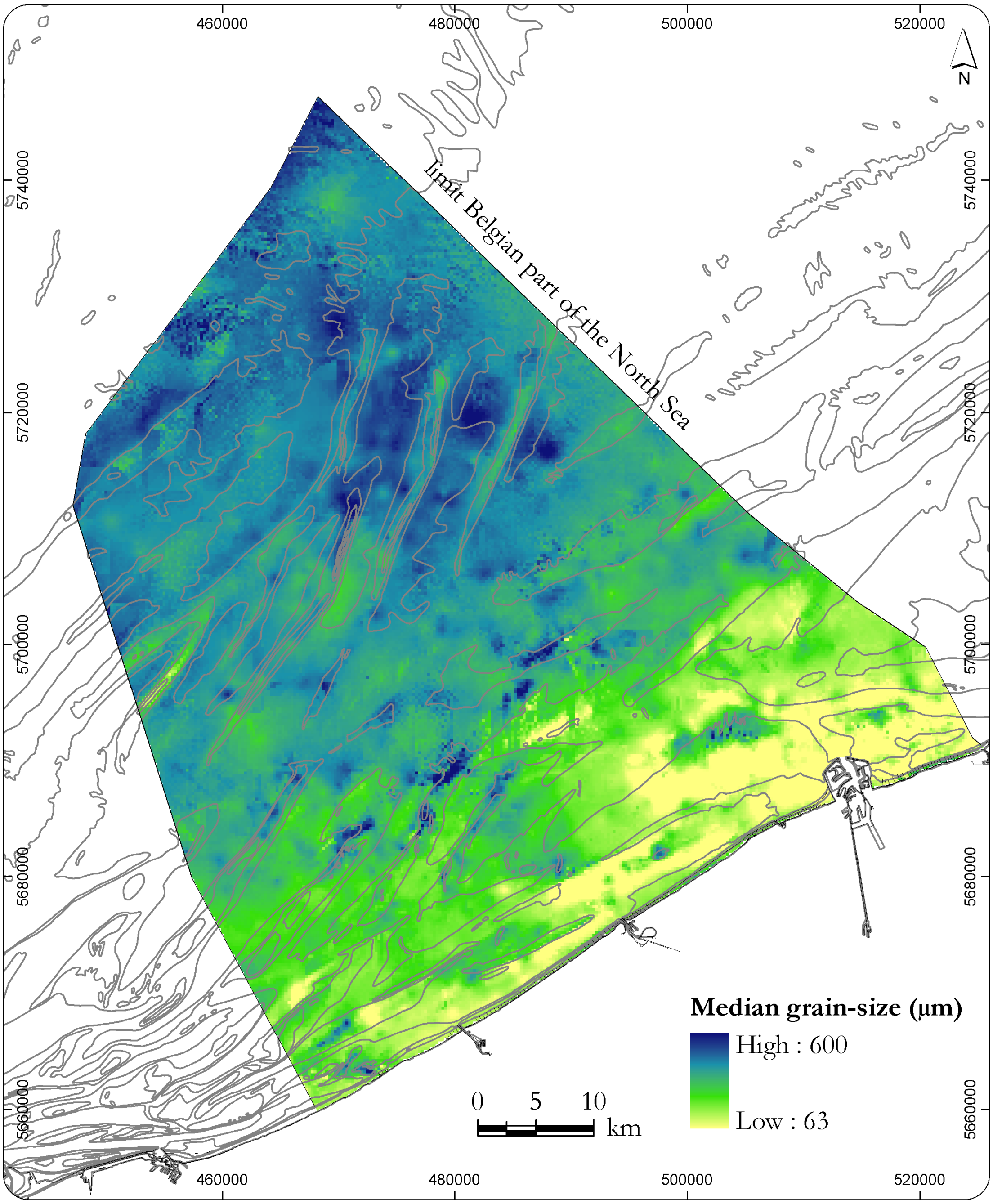


Figure 1.6: Median grain-size of the sand fraction (63 – 2000 µm) in the Belgian part of the North Sea.
The methodology to come to this map is described in detail in Chapter 2 of this thesis. This figure corresponds with Figure 2.10b. The density of point data to come to this map is presented in Figure 2.12.

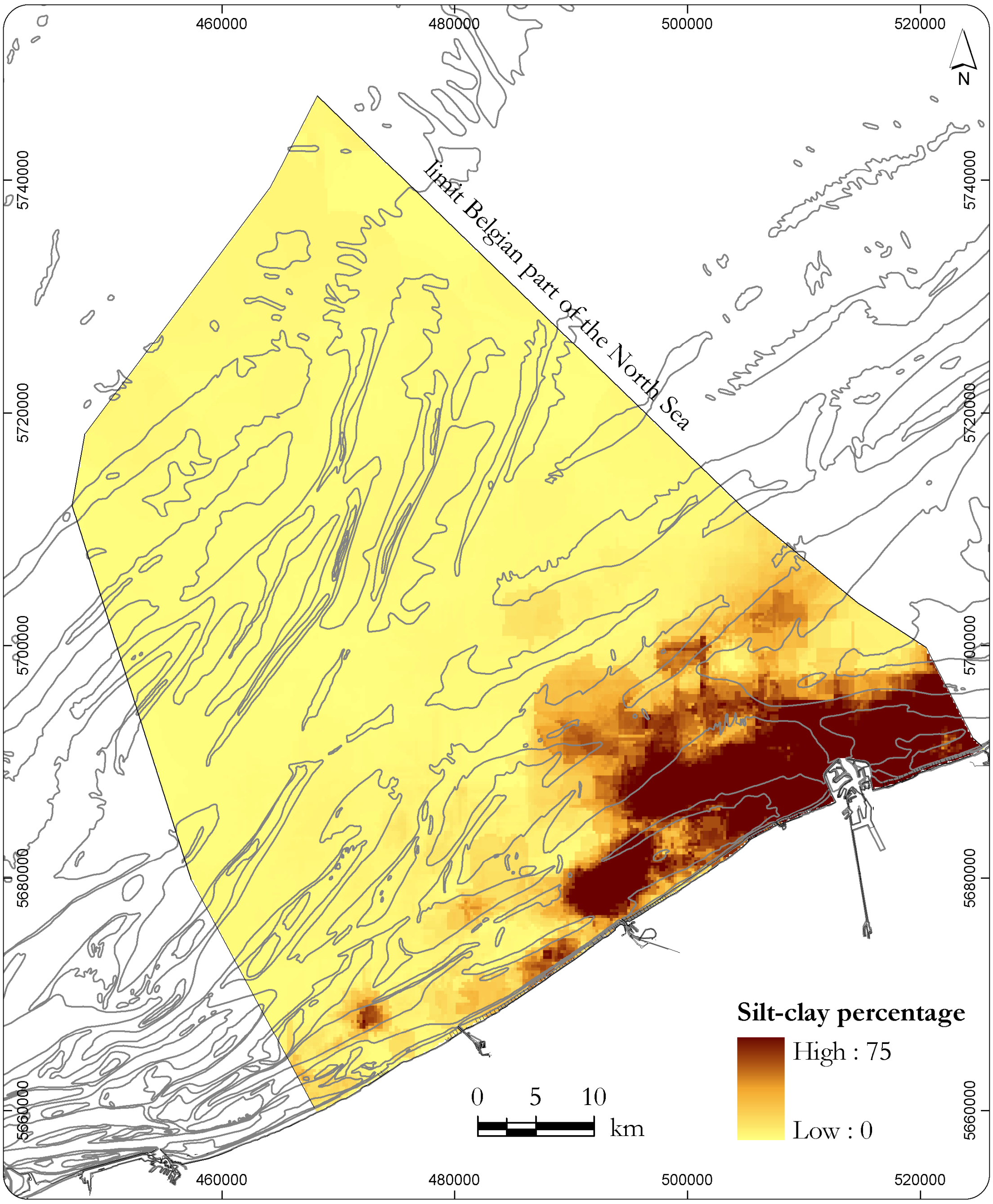


Figure 1.7: Silt-clay percentage in the Belgian part of the North Sea (Van Lancker et al. 2007). The density of point data to come to this map is presented in Figure 2.12.

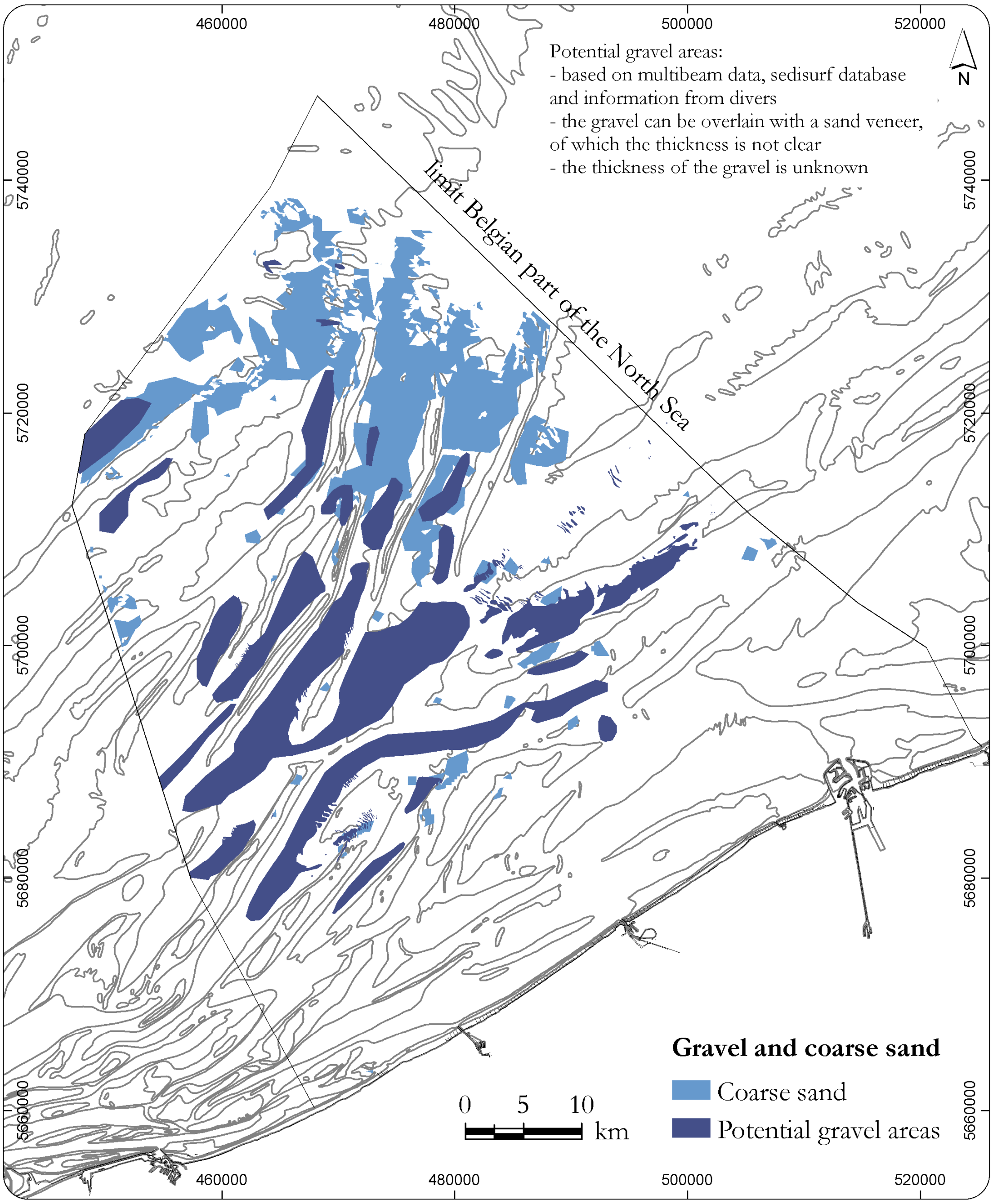


Figure 1.8: Gravel and coarse sand in the Belgian part of the North Sea (Van Lancker et al. 2007).

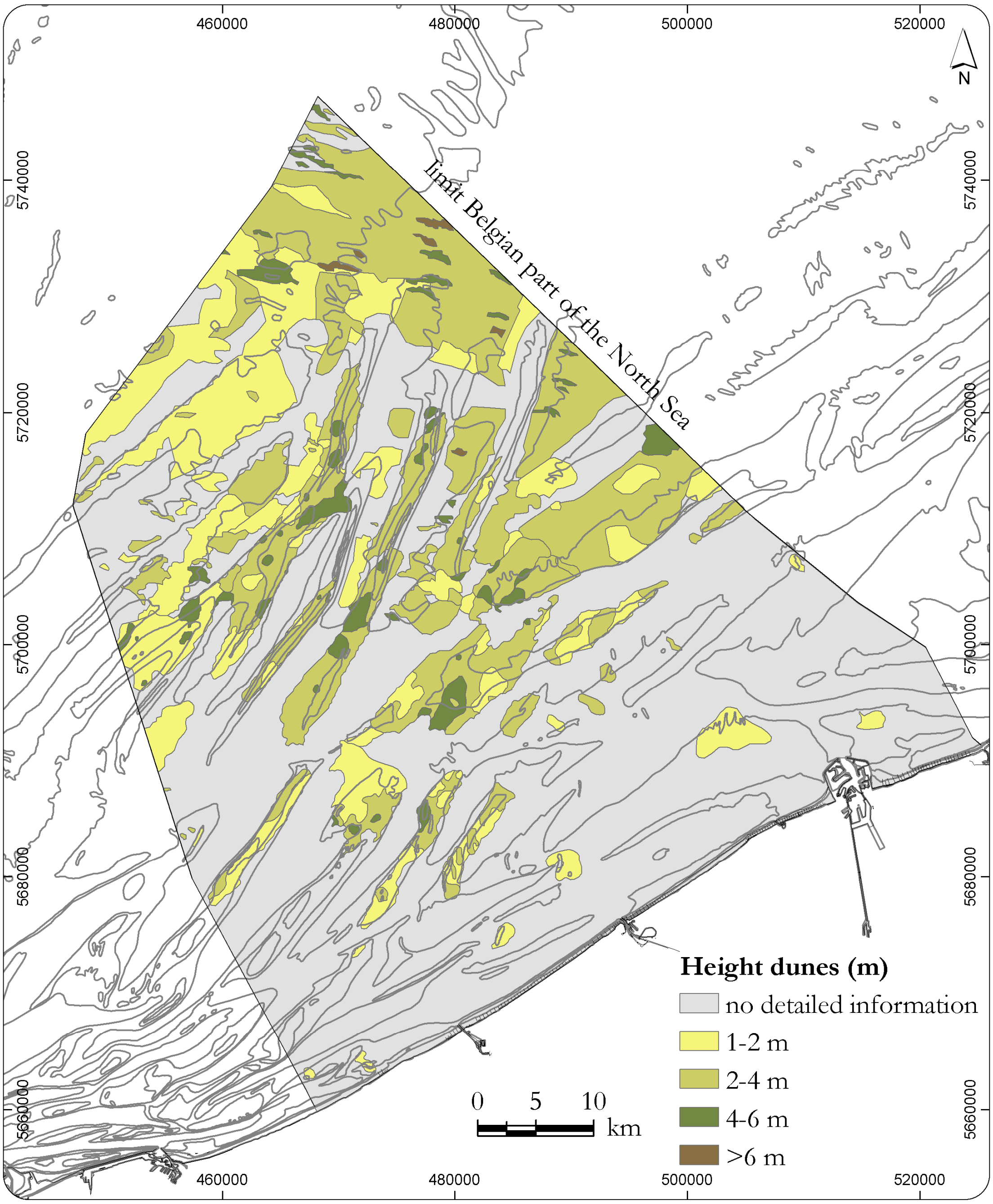


Figure 1.9: Bedforms and the height of the large to very large dunes in the Belgian part of the North Sea.
This map is based on singlebeam, side-scan sonar and multibeam data (Van Lancker et al. 2007).

1.2.4 Hydrodynamical characterization

Tides, wind and wave activity are the main hydrodynamic agents of the BPNS. Tides are semi-diurnal and slightly asymmetrical with a mean spring tidal range of 4.3 m at Zeebrugge and 2.8 m at neap tide. At spring tide, current velocities can be more than 1 m/s. Winds and waves originate mainly from the SW or from the NE. Winds are for almost 90% of the time below 5 Bft, while the significant wave height at the Westhinder is for 87% of the time below 2.0 m. The residual transport of the water masses is mainly to the NE (Van den Eynde 2004).
The Management Unit of the North Sea Mathematical Models and the Scheldt Estuary (MUMM) modelled a whole set of hydrodynamical variables such as the maximal current velocity (m/s) and the maximum bottom shear stress (N/m²). The maximum bottom shear stress is the maximal frictional force exerted by the flow per unit area of the seabed. Further details can be found in Van Lancker et al. (2007).
For both the maximum bottom stress (Figure 1.10) as well as the maximum current velocity (Figure 1.11), the same trend is visible on the BPNS: very high values occur around the harbour of Zeebrugge and along the northwestern part of the BPNS (mainly along the Oostdyck and in particular along its northern side). The lowest values occur along the western Coastal Banks. Intermediate values are found mainly along the central part of the BPNS (near the western part of the Gootebank) and to the east of the BPNS.

1.2.5 Biological characterization of the BPNS

Marine bottom fauna (or benthos) can be subdivided into five ecosystem components: species living just above and on the seafloor (hyperbenthos and epibenthos, respectively) and the fauna that lives inside of the sea bottom (infauna: micro-, meio- and macrobenthos).
Hyperbenthos are smaller species, like amphipods or larvae of epibenthos. Epibenthos are large, active benthos species, including sea stars, brittle stars, crabs, lobsters, bottom fish and cephalopods.
The microbenthos species are unicellular and bacterial organisms that live between and on the sand and silt grains. Meiobenthos species are multicellular organisms, smaller than 1 mm, e.g. copepod crustaceans and round worms inhabiting the interstitial spaces of the sediment. The macrobenthic species are all multicellular organisms and organisms larger than 1 mm. Examples of macrobenthos are bivalves, bristle worms, small crustaceans, such as amphipods and isopods and echinoderms.
Cattrijsse and Vincx (2001) give a summary of data on the BPNS for the five ecosystem components; for this research, only relationships between the physical environment and the macrobenthos has been dealt with.
A large amount of biological data were collected from the BPNS (Marine Biology Section, Ugent – Belgium, 2008). Between 1976 and 2001, over 1500 biological samples have been collected on the BPNS. The data were gathered in the framework of different research projects, resulting in an uneven distribution throughout the area. The sandbanks are mostly well-sampled, whereas almost no samples are available from the open sea and the eastern part of the Flemish Banks (Van Hoey et al. 2004). In general, the highest sample density (number of samples per km²) is found in inshore areas, decreasing steadily in an offshore direction.

As the BPNS is characterized by a highly variable topography, several macrobenthic communities and assemblages are distinguished. A community is defined as a group of organisms, occurring at a particular place (a physico-chemical environment) and time, interacting with each other and the environment. Distribution and diversity patterns of communities are therefore linked to a specific habitat type.
Up till now, five subtidal soft-bottom macrobenthic communities and six transitional communities (three subtidal and three intertidal species associations) are discerned (Degraer et al., 1999b; Degraer et al. 2002; Van Hoey et al. 2004). The occurring species associations differ drastically in habitat and species composition. The *Macoma baltica* community is bound to fine sandy, shallow locations characterized by high mud contents. These locations are only found close to estuarine environments (De Waen 2004). In nearshore muddy sands, species of the *Abra alba* community occur. The assemblage is characterized by a high species abundance, as well as a high diversity. Within the community, bivalve species occur in high densities (Van Hoey et al. 2004). These serve as an important food resource for epibenthic predators and benthic eating diving sea ducks (Degraer et al. 2002). Van Hoey et al. (2005) demonstrated that the ecological variation within the *A. alba* community is significant, as the BPNS can be considered as a major transition from the rich southern to the relatively poorer northern distribution area of this community. Furthermore, the community is due to temporal variations as well (Van Hoey et al. 2007). The *Nephtys cirrosa* community is characterized by a low species abundance and diversity and is typical for sandy areas. The *Ophelia limacina* community is found in medium to coarse sediments, often associated with gravel and shell fragments. However, this community is also represented in fine to medium sands with very low mud content. The distribution and description of the different communities and a selection of their macrobenthic species, is given in Degraer et al. (2006). The last community is the *Barnea candida* community (Degraer et al. 1999b). This community has a low diversity and density and is typically found in places where compact, tertiary clay layers outcrop. The rarity of this community is directly linked to the rarity of its habitat.

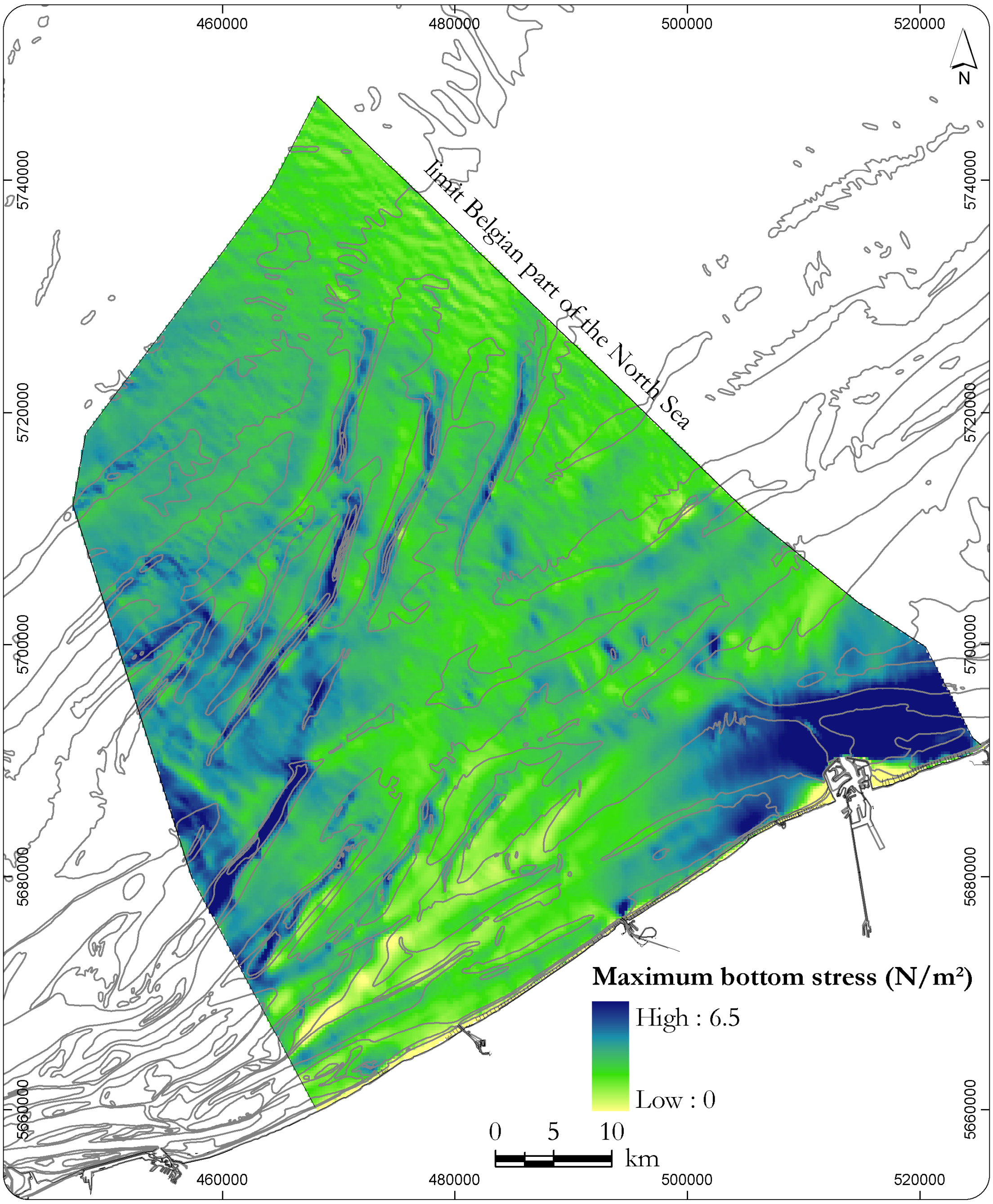


Figure 1.10: Maximum bottom stress in the Belgian part of the North Sea (Van Lancker et al. 2007).

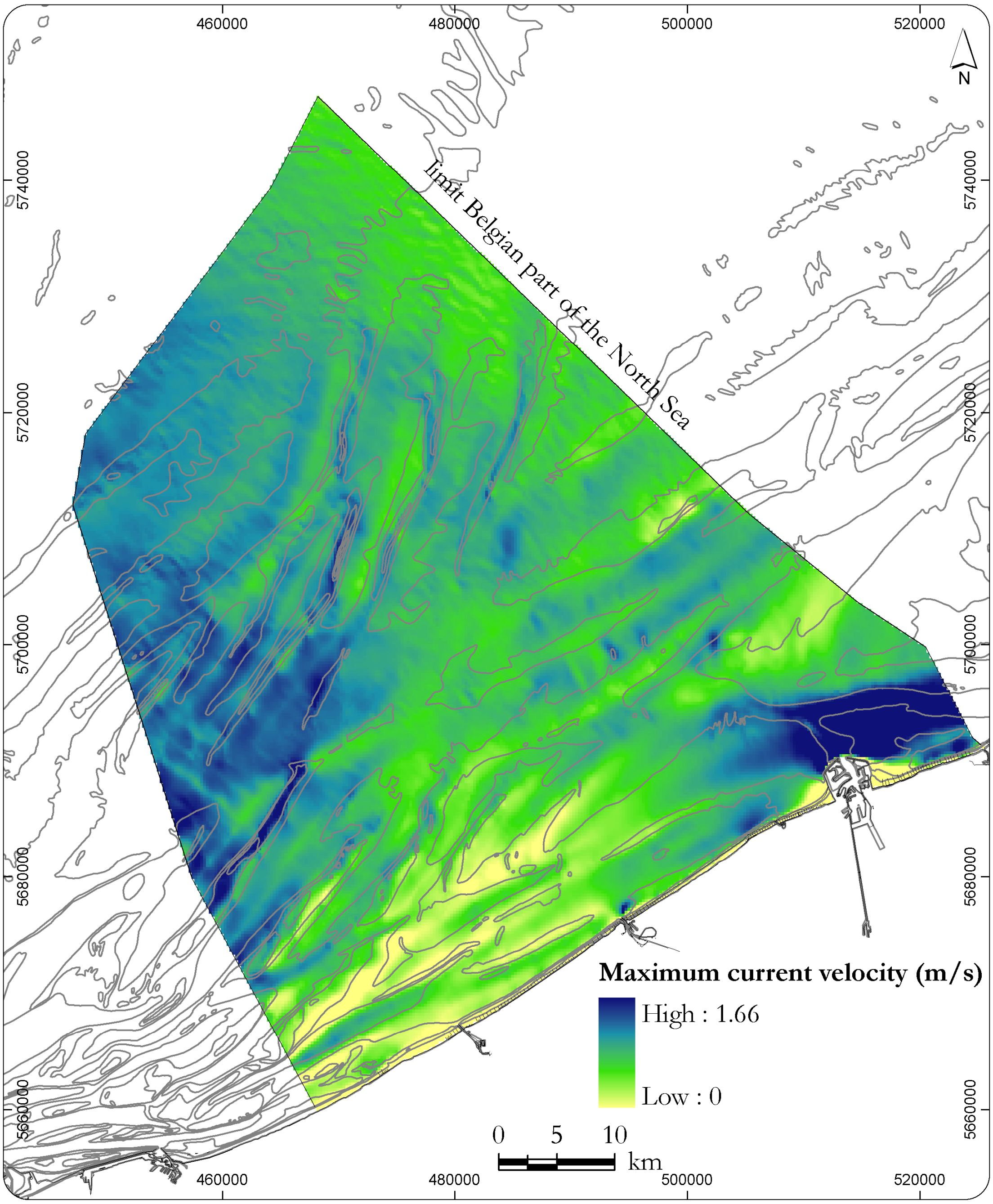


Figure 1.11: Maximum current velocity in the Belgian part of the North Sea (Van Lancker et al. 2007).

1.2.6 Legal characterization of the BPNS

Habitat maps are relevant in the context of **policy making**, although the link between science and policy is often difficult and not at all evident.
Table 1.3 gives an overview of international, European and Belgian obligations, commitments and laws related to habitat mapping on the BPNS.

Table 1.3: Overview of the most important regulations concerning the designation of marine protected areas on the BPNS in an international, European and Belgian context (modified from Cliquet et al. 2007).

INTERNATIONAL	
- Obligations	- Convention on Wetlands of International Importance, especially as Waterfowl Habitat (Ramsar 1971)[1] - Convention for the Protection of the Marine Environment of the North-East Atlantic (OSPAR 1992)[2] - Convention on Biological Diversity of Rio the Janeiro (1992)[3]
- Commitments	- designation and management of marine protected areas, such as those agreed at the World Summit on Sustainable Development, to establish a representative system of marine protected areas by 2012[4] - decision from the 7th conference of state parties to the Biodiversity Convention to establish and maintain (by 2012) marine and coastal protected areas that are effectively managed, ecologically based and contribute to a global network of marine and coastal protected areas[5]
EUROPEAN	
- Obligations	- Birds Directive 79/409/CEE (1979)[6] - Habitats Directive 92/43/CEE (1992)[7]
- Commitments	- EU Biodiversity Action Plan has as objective to complete a network of Special Protection Areas by 2008 for marine areas, adopt lists of Sites of Community Importance by 2008 for marine areas, designate Special Areas of Conservation and establish management priorities and necessary conservation measures for Special Areas of Conservation by 2012, and establish similar management and conservation measures for Special Protection Areas by 2012 for marine areas[8]
BELGIAN	
- Legislation	- Law on the protection of the marine environment in marine areas under Belgian jurisdiction on the marine environment[9] - Royal Decree of 14 October 2005[10]

The Marine Protection Law of 1999[11] was the first Belgian law, enabling the federal government to designate marine protected areas (MPAs). This law permits the designation of 5 types of marine protected areas: 1) integral marine reserves; 2) specific marine reserves; 3) Special Protection Areas (SPAs) or Special Areas of Conservation (SACs) for specific habitats or species; 4) closed zones for certain activities during certain periods; and 5) buffer zones (Cliquet et al. 2007). However, because of an initial lack of public participation or consultation of stakeholders, it took the Belgian government until **2005** to legally designate the first **5 MPAs on the BPNS**[12] (Cliquet et al. 2007):

- **3 SPAs** in the framework of the **Birds Directive**[13]:
  - SBZ-V1 Nieuwpoort;
  - SBZ-V2 Oostende;
  - SBZ-V3 Zeebrugge.
- **2 SACs** in the framework of the **Habitats Directive**[14]:
  - SBZ-H1 Trapegeer Stroombank;
  - SBZ-H2 Vlakte van de Raan.

The designation of the SPAs is based on Haelters et al. (2004). Together, the SACs and the SPAs will create a network of protected areas across the EU, known as **Natura 2000** (Douvere et al. 2007).
However, in February 2008, the Belgian Council of State annulated the designation of the Vlakte van de Raan as SAC[15] (after a complaint by the energy company Electrabel).

In 2006, a 6th area was designated: the **specific marine reserve** Bay of Heist[16].
All Belgian MPAs are presented in Figure 1.12.

Moreover, other **European Directives** for the conservation or protection of marine environments exist (**Water Framework Directive; 2000/60/EC**) or are in development (**Marine Strategy Directive**) and have to be implemented by the European member states (Derous et al. subm. b). The Water Framework Directive establishes a framework for the protection and improvement of all European surface and ground waters, with a 'good ecological water status' by 2015 (Derous et al. subm. b). The future Marine Strategy Directive (included into the EU Marine Thematic Strategy) will establish a framework for the protection and preservation of the marine environment, the prevention of its deterioration and the restoration of that environment in areas where it has been affected adversely (Derous et al. subm. b), with a 'good environmental status' in the marine environment as ultimate objective by 2021 (DEFRA 2006). The overall aim of this strategy is to promote sustainable use of the seas and to conserve marine ecosystems against certain threats (e.g. loss of habitats, degradation of biodiversity) and pressures (e.g. physical degradation of habitat from dredging and extraction of sand and gravel) (European Commission 2006). The EU Maritime Policy calls in its **Green Paper** (Commission of the European Communities 2006) for a system of ecosystem-based marine spatial planning for a growing maritime economy aiming to manage the increasingly competing economic activities, while at the same time safeguarding biodiversity (Douvere et al. 2007).
In the context of **marine spatial planning** (MSP), there is a growing need to meet these international and national commitments regarding biodiversity conservation. Until recently, MSP on the BPNS was done on an ad hoc basis with legal driving forces (Law of the Sea and Belgian legislation) and economic driving forces (e.g. aggregate extraction and fisheries) (Douvere et al. 2007). The BPNS has a very limited surface (3600 km²) and is, regarding the anthropogenic activities, one of the most occupied shelf seas of the world (Figure 1.13) (see Maes et al. 2005, for an overview).

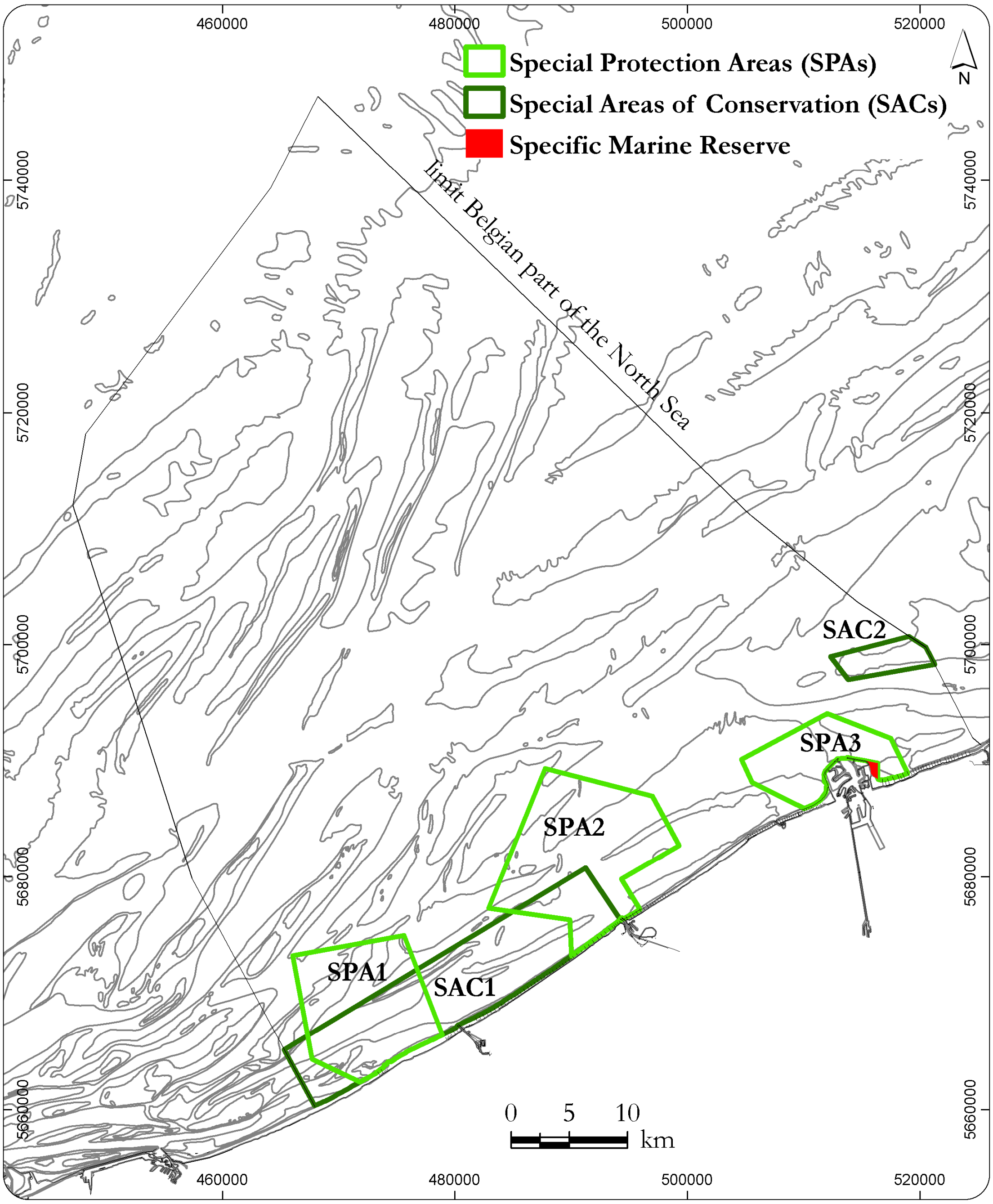


Figure 1.12: Marine Protected Areas on the Belgian Part of the North Sea: 3 Special Protection Areas (SPA1 = SBZ-V1 Nieuwpoort; SPA2 = SBZ-V2 Oostende; and SPA3 = SBZ-V3 Zeebrugge); 2 Special Areas of Conservation (SAC1 = SBZ-H1 Trapegeer Stroombank; and SAC2 = SBZ-H2 Vlakte van de Raan)[17] and 1 specific marine reserve (Bay of Heist)[18]. In February 2008, SAC2 was annulated by the Belgian Council of State, after a complaint by the energy company Electrabel.

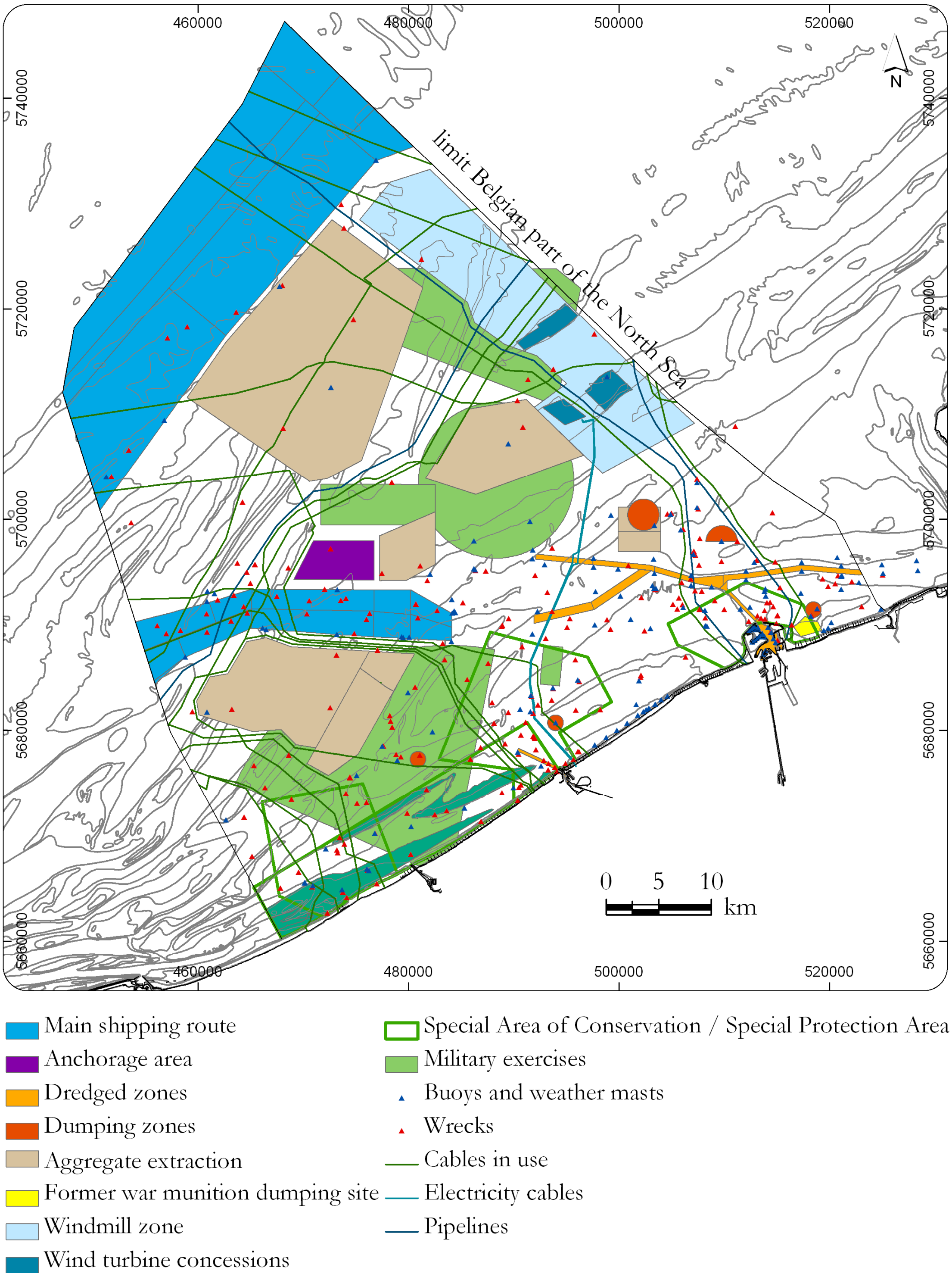


Figure 1.13: Combination of all activities on the Belgian part of the North Sea (Maes et al. 2005).

Chapter 2

Multivariate geostatistics for the predictive modelling of the surficial sand distribution in shelf seas

PUBLISHED AS:

Verfaillie, E., Van Lancker, V. and Van Meirvenne, M. 2006. Multivariate geostatistics for the predictive modelling of the surficial sand distribution in shelf seas. Continental Shelf Research 26, 2454-2468.

2 Multivariate geostatistics for the predictive modelling of the surficial sand distribution in shelf seas

Abstract

Multivariate geostatistics have been used to obtain a detailed and high-quality map of the median grain-size distribution of the sand fraction at the Belgian Continental Shelf. Sandbanks and swales are the dominant geomorphological features and impose a high-spatial seafloor variability. Interpolation over complex seafloors is difficult and as such various models were investigated. In this paper, linear regression and ordinary kriging (OK) were used and compared with kriging with an external drift (KED) that makes use of secondary information to assist in the interpolation. KED proved to be the best technique since a linear correlation was found between the median grain-size and the bathymetry. The resulting map is more realistic and separates clearly the sediment distribution over the sandbanks from the swales. Both techniques were also compared with a simple linear regression of the median grain-size against the bathymetry. An independent validation showed that the linear regression yielded the largest average prediction error (almost twice as large as with KED).
Unlike most static sedimentological maps, our approach allows for defining grain-size classes that can be adapted according to the needs of various applications. These relate mainly to the mapping of soft substrata habitats and of the most suitable aggregates for extraction. This information is highly valuable in a marine spatial planning context.

Keywords: Multivariate geostatistics; Median grain-size; Bathymetry; Habitat mapping; Resource maps; Belgian continental shelf

2.1 Introduction

Seabed habitats are subject to increasing pressures from human developments such as fisheries, aggregate extraction, dredging/dumping and windmill farms. In this context, the mapping of habitats and their prediction becomes crucial, both at the level of baseline studies as during the monitoring and decommitment phase. There is a difference between the physical (or abiotic) and the biological (or biotic) part of a seabed habitat (Figure 2.1). However, if a full coverage map of the physical habitat is available and if the relations between the physical and the biological habitat are known, it is possible to create a full coverage map of the biological habitat. Nowadays there is an increasing demand for full coverage information. 'Filling the gaps' and 'predictive modelling' or the prediction of physical and biological information in areas with gaps, is a hot topic (e.g. ICES 2005) in the framework of habitat mapping and nature protection. This is one of the aims of the project MESH (Development of a framework for Mapping European Seabed Habitats) and BWZee (Biological Valuation Map of the Belgian Continental Shelf), in which the current research plays an important role.

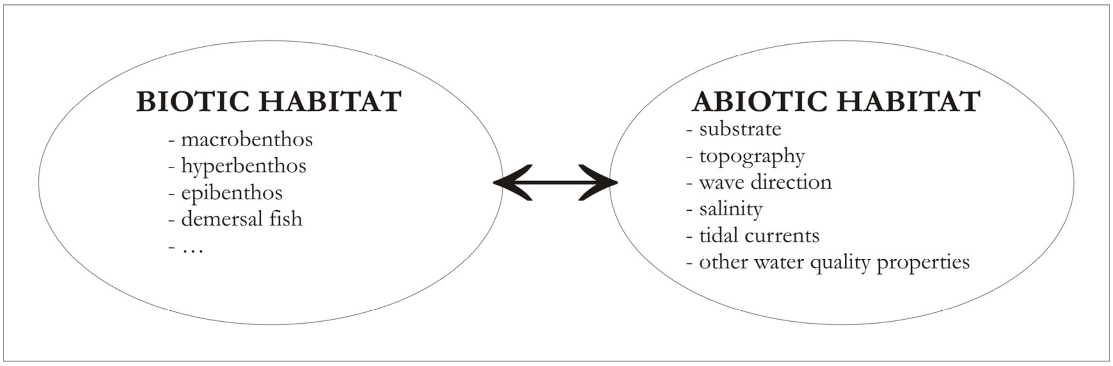


Figure 2.1: A seabed habitat consists of a biotic and an abiotic part. ICES (2005) defined a habitat as: "A recognizable space which can be distinguished by its abiotic characteristics and associated biological assemblage, operating at particular spatial and temporal scales." In this paper only the abiotic part is considered.

Data, describing the physical habitat, are available as point information (e.g. sediment samples), as full coverage information (e.g. Digital Elevation Model or DEM) or as full coverage information from a model (e.g. current data, shear stress). The datasets available in this study were sediment samples and a DEM. For the mapping of soft substrata habitats, it has been shown that the sedimentology (mainly the median grain-size and the silt-clay percentage) is an important parameter to explain and predict the occurrence of (macro)benthos (seabed organisms larger than 1mm) (e.g. Wu and Shin 1997; Leecaster 2003; and Van Hoey et al. 2004). Although sediment samples are generally more available than biological samples, it remains difficult to predict (or interpolate) their distribution and this particularly over complex seafloors. As such, a sound methodology for the interpolation of these data is necessary.

The general aim of this paper was to produce a high-quality map of the median grain-size at the Belgian Continental Shelf (BCS), looking for the best interpolation method. This map is a valuable product in the context of aggregate extraction, habitat mapping, ecological valuation, spatial planning and sediment transport.

2.2 Material and methods

2.2.1 Data description

Grain-size data was derived from a sedimentological database (‘sedisurf@’) hosted by Renard Centre of Marine Geology, Ghent University. The dataset is a compilation of sample information since 1976 and contains more than 6000 samples.
As a second variable, a high-resolution DEM was compiled based on data from the Ministry of the Flemish Community (Department of Environment and Infrastructure, Waterways and Marine Affairs Administration, Division Coast, Hydrographic Office) and completed with data from the Hydrographic Office of the Netherlands and the United Kingdom. Regarding the Belgian shelf, this is a very valuable source of information as its very large density allowed an interpolation to a resolution of 80 m, using a simple inverse distance algorithm. From the DEM, a slope map was derived. Based on the DEM and the slope map, homogeneous zones at the BCS could be defined (Figure 2.2). These zones allow a distinction between sandbanks, swales and foreshore zones. The delimitation of the zones was done by alternatively inspecting the DEM and the slope map and the visual drawing of polygons in a geographical information system (GIS). From each zone the amount of samples, the variation of the grain-size (e.g. mean value, variance, sum,…) can be queried in GIS. In this way it is possible to get an impression of the variation of the samples within each zone and to carry out a quality control of the grain-size values. The quality control of the sediment samples was done assuming that samples inside the same zone, are more similar than samples from different zones. Extreme values can be identified and if necessary, removed. For that purpose a sound knowledge of the sedimentological data is needed. On this basis 83 points were removed out of the dataset.

2.2.2 Linear regression

A well-known approach consists of modelling the relation between the median grain-size and the depth using a linear function of the type:

$$z(\mathbf{x}) = \mathrm{a}_0 * + \mathrm{a}_1 * y(\mathbf{x})$$

with $z(\mathbf{x})$ equal to the measurement of median grainsize at location **x**, $\mathrm{a}_0 *$ the intercept constant value, $\mathrm{a}_1 *$ the slope constant value; $y(\mathbf{x})$ the measurement of depth at location **x**.
With this relation, each depth value can be converted into a median grain-size value. This type of regression has the major shortcoming that the median grain-size is only derived from the depth at the same location **x**, regardless of the surrounding values (Goovaerts 1999).

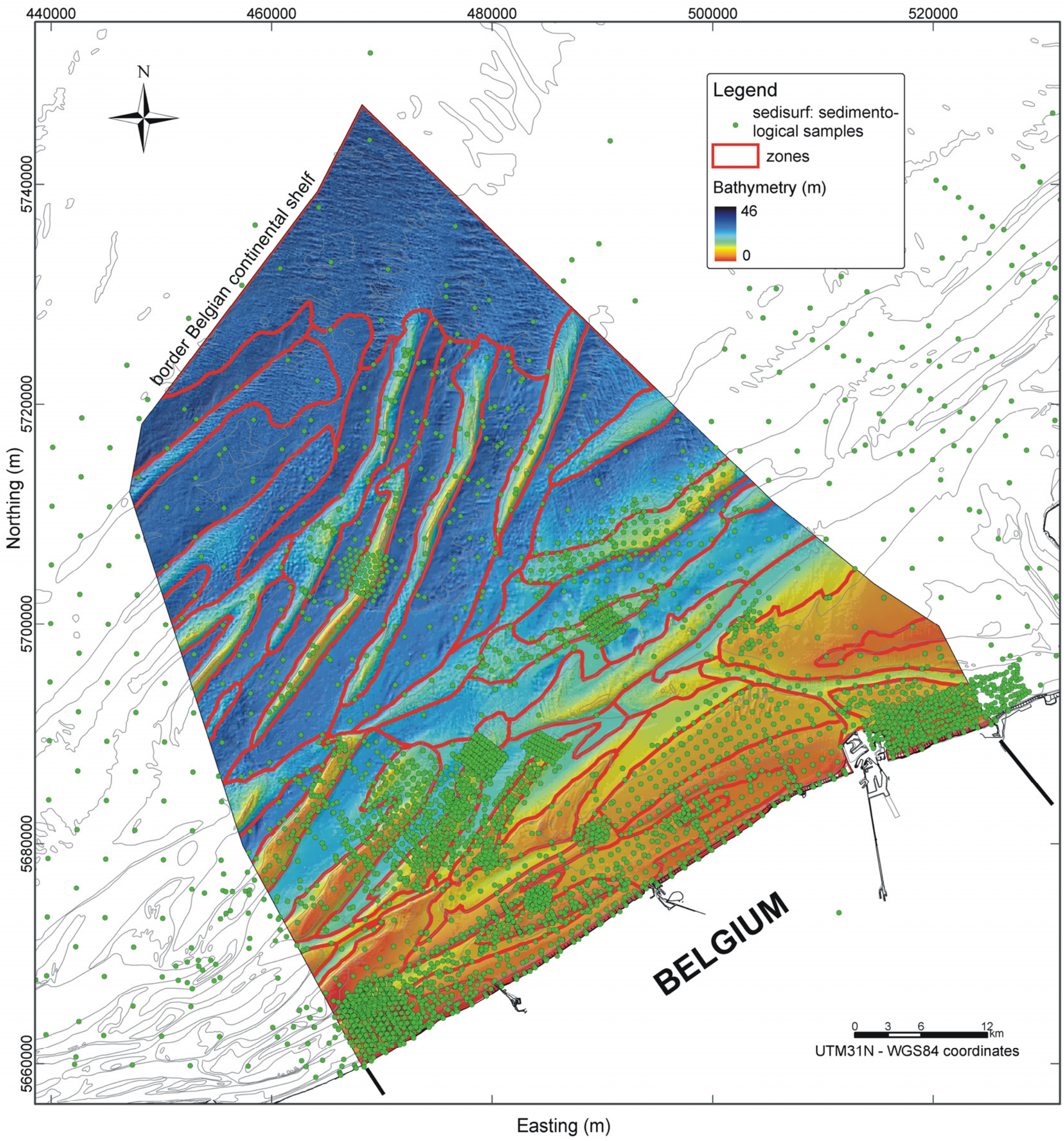


Figure 2.2: Large-scale zonation, distinguishing swales, sandbanks and foreshores.

2.2.3 Geostatistical approach

Geostatistical interpolation techniques (generally known as kriging) have the advantage that they are stochastic in contrast with deterministic techniques like trend surfaces. The latter predict an unknown value in a unique way without an associated measure of uncertainty. Stochastic techniques provide a number of possible values, with a probability of occurrence. A unique solution cannot be expected (Goovaerts 1997). Moreover, geostatistical techniques have the advantage that they make use of the spatial correlation between neighbouring observations, to predict values at unsampled places (Goovaerts 1999). These techniques give an indication of the errors and uncertainties associated with the interpolated values, based on a variance surface of the estimated values (Burrough and McDonnell 1998). Multivariate geostatistics can be used if there is a relation between the predicted variable (e.g. median grain-size) and a secondary variable (e.g. bathymetry). It is possible to include this secondary information into the interpolation. This additional information results in a more accurate and complete prediction of the variable than without the secondary information. In practice, secondary information is often cheaper or easier to obtain, and as such can complement the sparsely sampled (primary) observations. More details on the applied geostatistical analysis can be found in Goovaerts (1997), Deutsch and Journel (1998) and Wackernagel (1998).

Variogram analysis

The variogram $\gamma(\mathbf{h})$ represents the average variance between observations separated by a distance $\mathbf{h}$. The value plays an important role in the description and interpretation of the structure of the spatial variability of the investigated regionalized variable. It is estimated by (Journel and Huijbregts 1978):

$$\gamma(\mathbf{h}) = \frac{1}{2\mathrm{N}(\mathbf{h})} \sum_{\alpha=1}^{\mathrm{N}(\mathbf{h})} \left\{ z(\mathbf{x}_\alpha) - z(\mathbf{x}_\alpha + \mathbf{h}) \right\}^2 \qquad (2.1)$$

with $z(\mathbf{x}_\alpha)$ equal to the measurement at location $\mathbf{x}_\alpha$, $z(\mathbf{x}_\alpha+\mathbf{h})$ the measurement at location $\mathbf{x}_\alpha+\mathbf{h}$, $\gamma(\mathbf{h})$ the variogram for distance vector (=lag) $\mathbf{h}$ between measurements $z(\mathbf{x}_\alpha)$ and $z(\mathbf{x}_\alpha+\mathbf{h})$, and $\mathrm{N}(\mathbf{h})$ the number of couples of measurements separated by $\mathbf{h}$.

A variogram is presented as a graph (Figure 2.3), where the calculated variogram values (dots) represent the experimental variogram. The fitting of a theoretical variogram (curve) is an important step in the variogram analysis. Hereby, the ‘sill’ is the total variance s^2 of the variable, the ‘range’ is the maximal spatial extent of spatial correlation between observations of the variable and the ‘nugget’ is the random error.

The theoretical variogram can be composed of nested models or structures. Common models are the nugget model, spherical model, exponential model, Gaussian model and power model. Direction dependant variograms can be set up in the case of anisotropic variability. The formulas of these models can be found in e.g. Journel and Huijbregts 1978; Wackernagel 1998.

For the variogram analysis the programme Variowin 2.21 (Pannatier 1996) was used.

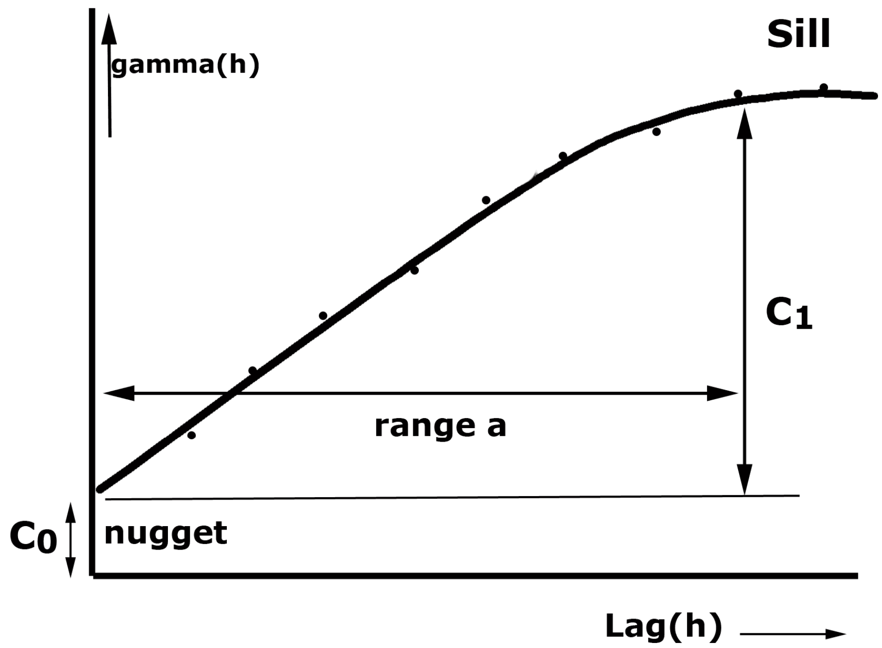


Figure 2.3: Experimental variogram (black dots) and theoretical variogram (curve)
(from Burrough and McDonnell 1998).

Interpolation with kriging

A univariate and a multivariate variant of kriging were compared:

- Ordinary Kriging (OK) of median grain-size (d50) using directional variograms;
- Kriging with an external drift (KED) of d50 with bathymetrical values as secondary information and with an omnidirectional variogram.

For the geostatistical analysis the software GSLIB 1998 (Deutsch and Journel 1998) was used.
OK is the most frequently used kriging technique. The OK algorithm uses a weighted linear combination of sampled points situated inside a neighbourhood (or interpolation window) around the location $\mathbf{x}_0$, where the interpolation is conducted. An underlying assumption is that the mean value (m) is locally stationary (i.e. that it has a constant value inside the interpolation neighbourhood). The algorithm can be written as:

$$Z^*(\mathbf{x}_0) = \sum_{\alpha=1}^{n(\mathbf{x}_0)} \{\lambda_\alpha \cdot [Z(\mathbf{x}_\alpha) - m]\} + m = \sum_{\alpha=1}^{n(x_0)} \{\lambda_\alpha Z(\mathbf{x}_\alpha)\} + \left[1 - \sum_{\alpha=1}^{n(\mathbf{x}_0)} \lambda_\alpha\right] \cdot m \qquad (2.2)$$

with λ_α equal to the weights attributed to the $n(\mathbf{x}_0)$ observations $z(\mathbf{x}_\alpha)$; n the total number of observations $z(\mathbf{x}_\alpha)$; $n(\mathbf{x}_0)$ the subset of n, lying inside the interpolation window.
The weights λ_α are obtained by solving a set of equations (the kriging system) involving knowledge of the variogram (see e.g. Goovaerts 1997). These weights are

constrained to sum to one, leading to the elimination of the parameter m from the estimator which is thus written as:

$$Z^*_{OK}(\mathbf{x}_0) = \sum_{\alpha=1}^{n(\mathbf{x}_0)} \lambda_\alpha Z(\mathbf{x}_\alpha) \qquad \text{with} \qquad \sum_{\alpha=1}^{n(\mathbf{x}_0)} \lambda_\alpha = 1 \qquad (2.3)$$

KED is a multivariate variant of 'Kriging with a Trend Model' (KT), formerly called 'Universal Kriging'. KED and KT are non-stationary methods, meaning that the statistical properties of the variable are not constant in space (i.e. no constant mean within the interpolation neighbourhood). With KT, the trend is modelled as a function of the spatial coordinates, whilst for KED the trend $m(\mathbf{x}_0)$ is derived as a local linear function of the secondary variable $z_2(\mathbf{x}_0)$, which is formulated in each interpolation window (Goovaerts 1997):

$$m(\mathbf{x}_0) = b_0 + b_1 z_2(\mathbf{x}_0) \quad (2.4)$$

with $m(\mathbf{x}_0)$ the trend at location $\mathbf{x}_0$; b_0, b_1 the unknown parameters of the trend, calculated in each interpolation window from a fit to the observations; $z_2(\mathbf{x}_0)$ the secondary variable at location $\mathbf{x}_0$.
The KED estimator has the same form as the OK estimator.
For non-stationary geostatistics such as KED, $Z(\mathbf{x})$ can be decomposed into a deterministic function or a drift $m(\mathbf{x})$ and a residual random function $Y(\mathbf{x})$ (Wackernagel 1998):

$$Y(\mathbf{x}) = Z(\mathbf{x}) - m(\mathbf{x}) \qquad (2.5)$$

The underlying variogram associated with Y is directly accessible, when the drift is not active in a particular direction of space. The variogram in this direction can be extended to the other directions under an assumption of isotropic behaviour of the underlying variogram (Wackernagel 1998).
KED is a multivariate geostatistical technique, as it makes use of secondary information. However, this secondary data must be available at all primary data locations as well as at all locations being estimated. A more complex multivariate geostatistical technique is cokriging, which does not require this secondary information to be known at all locations being estimated. Cokriging is much more demanding than other kriging techniques because both direct and cross variograms must be inferred and jointly modelled and because a large cokriging system must be solved (Goovaerts 1997).

Validation

To enable a thorough quality control of the geostatistical analysis, the sedisurf@ database was divided into two subsets: a prediction and an independent validation dataset. The proportion of both datasets is respectively, 70% and 30% of the whole dataset. The validation dataset was selected using a random selection of data points.
Several indices are suitable to evaluate the interpolation. These indices are all a measure of the estimation error that is the difference between the estimated and the observed value:

$$z^*(\mathbf{x}_\alpha) - z(\mathbf{x}_\alpha).$$

- The mean estimation error (MEE), which has to be about zero to have an unbiased estimator.

$$\text{MEE} = \frac{1}{n}\sum_{\alpha=1}^{n}\left(z^*(\mathbf{x}_\alpha) - z(\mathbf{x}_\alpha)\right) \qquad (2.6)$$

- The mean square estimation error (MSEE), which has to be as low as possible and is useful to compare different procedures. The root mean square estimation error (RMSEE) is used to obtain the same units as the variable. This parameter has to be compared to the variance or the standard deviation of the dataset.

$$\text{MSEE} = \frac{1}{n}\sum_{\alpha=1}^{n}\left(z^*(\mathbf{x}_\alpha) - z(\mathbf{x}_\alpha)\right)^2 \qquad (2.7)$$

- The mean absolute estimation error (MAEE), which is analogous to the MSEE, but less sensitive to extreme deviations.

$$\text{MAEE} = \frac{1}{n}\sum_{\alpha=1}^{n}\left|z^*(\mathbf{x}_\alpha) - z(\mathbf{x}_\alpha)\right| \qquad (2.8)$$

- The Pearson correlation coefficient between $z^*(\mathbf{x}_\alpha)$ and $z(\mathbf{x}_\alpha)$, which indicates the degree of linear correlation between observed and estimated values. This value always has to be considered in combination with the MEE. The correlation coefficient is itself a measure of the proportion of variance explained, hence is related to MSEE.

2.3 Results

2.3.1 Linear regression

The relation between median grain-size and depth was modelled as:

$$\text{d50} = 179.84 + 5.94 \cdot \text{depth}$$

resulting in the map of the median grain-size shown at Figure 2.4. This map is a simple rescaling of the DEM, converted into grain-size values between 179 and 508. Linear regression is not an exact interpolator, meaning that the interpolated map does not honour observations; the measurements are only used to calculate a linear regression function. As the map is a transformation of the DEM, it shows very clearly the anisotropy, but the typical, more patchy pattern of the grain-size is completely lost (compare with Figure 2.10a and b).

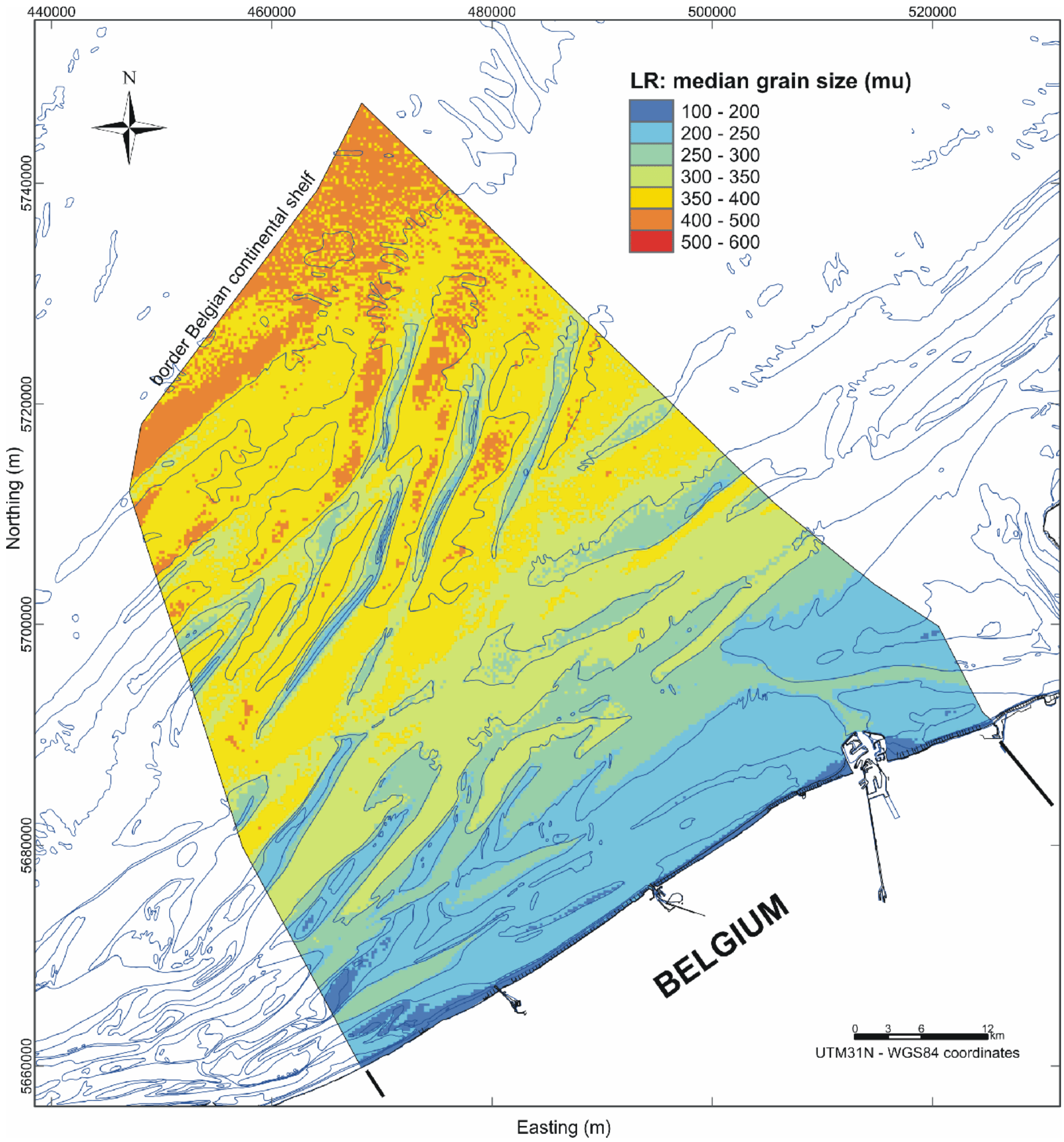


Figure 2.4: Map of median grain-size, on the basis of linear regression.

2.3.2 <u>Geostatistical approach</u>

Exploratory data analysis

The histogram of the grain-size data (Figure 2.5) shows a symmetric distribution.
At every location where the median grain-size d50 is known the depth is also known from the DEM (Figure 2.6). The Pearson correlation coefficient r_{ij} between both variables is 0.46, indicating a moderately strong correlation. The Spearman rank correlation is slightly larger (0.52) indicating the presence of some outliers (as can be seen at Figure 2.6) reducing the Pearson correlation coefficient.
The scatterplot suggests the existence of two populations (one parallel with and one perpendicular to the *X* axis). However after splitting the two populations, the correlations did not improve. To preserve the added value of the secondary variable in the geostatistical analysis, the decision was made to keep the dataset as a single entity.

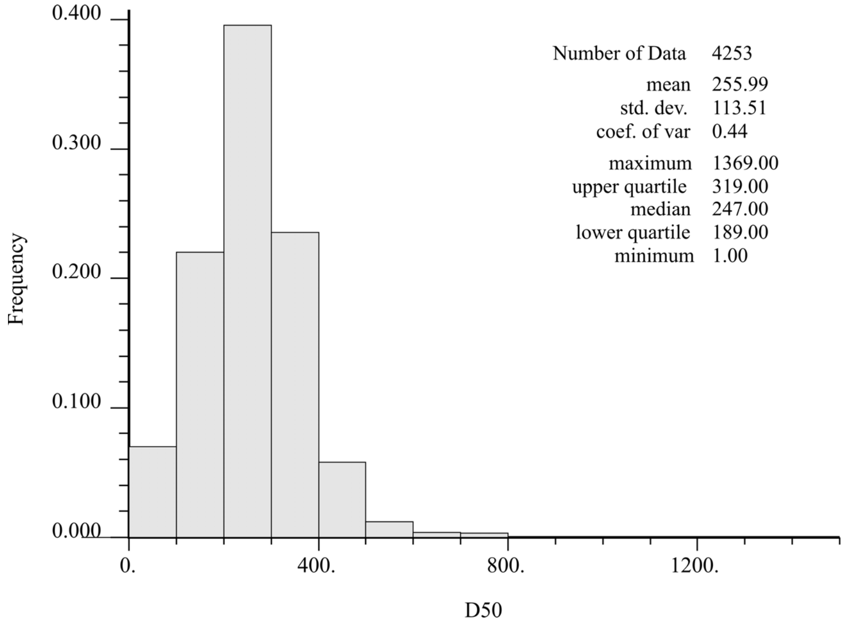


Figure 2.5: Histogram of the prediction dataset of the d50 values.

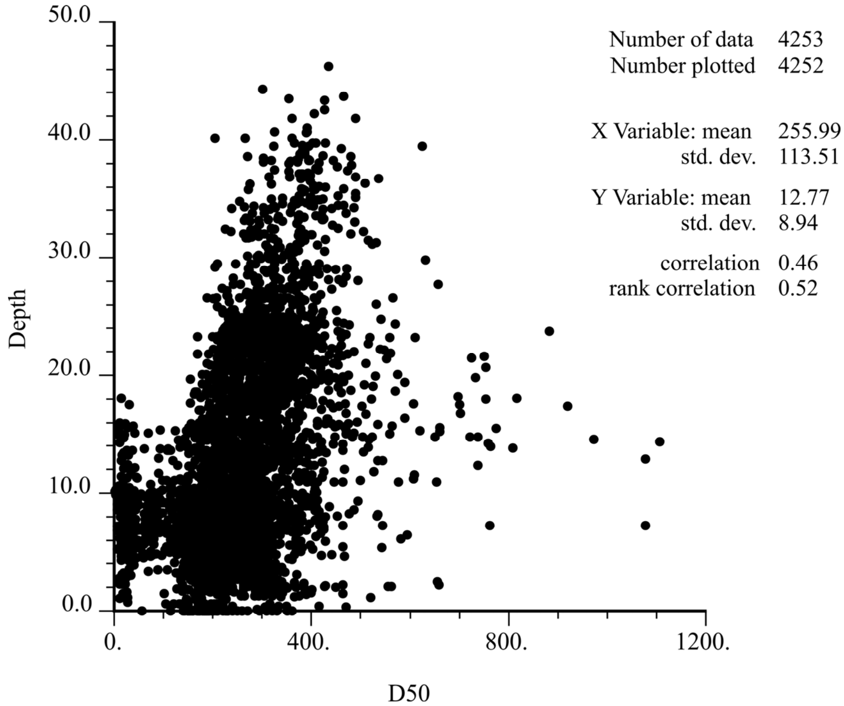


Figure 2.6: Scatter plot of d50 value compared to the depth.

Variogram analysis

The maximal diagonal distance at the BCS is about 90 km. Following a rule of thumb, the product of the lag interval distance and the number of lags should not exceed half of this largest dimension: i.e. between 30 and 45 km. Consequently, the variogram surface was calculated using 11 lags of 3000 m. This variogram surface (Figure 2.7) shows a clear anisotropy. The direction of the largest continuity is about 50° (expressed as a trigonometric angle), corresponding to the direction of the sandbanks at the BCS and to the smallest variogram values. This indicates that the sandbanks have a strong influence on the spatial variability of the data. This is the case for the median grain-size (Figure 2.7, left), but it is stronger with the depth values (Figure 2.7, right). The direction of the largest discontinuity is about 130°, corresponding to the direction perpendicular to the sandbanks. To characterize the spatial variability in different directions directional variograms were calculated in the directions: 40°, 85°, 130° and 175°.

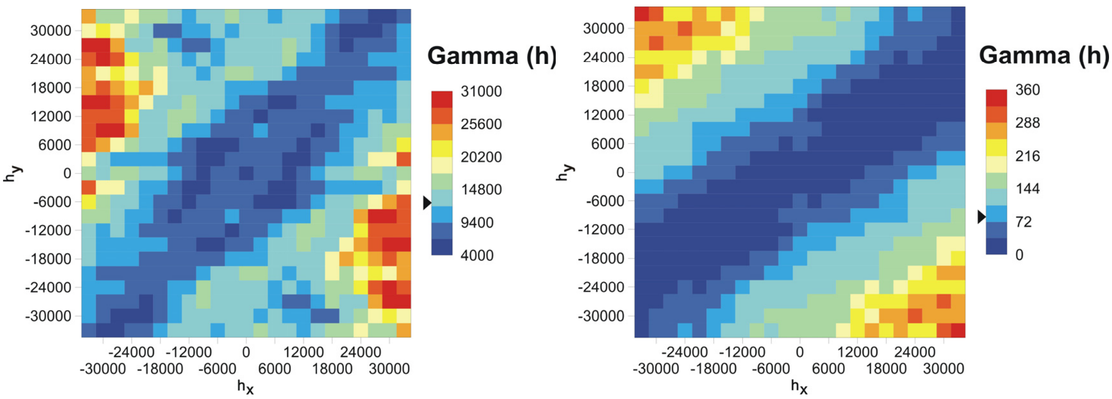
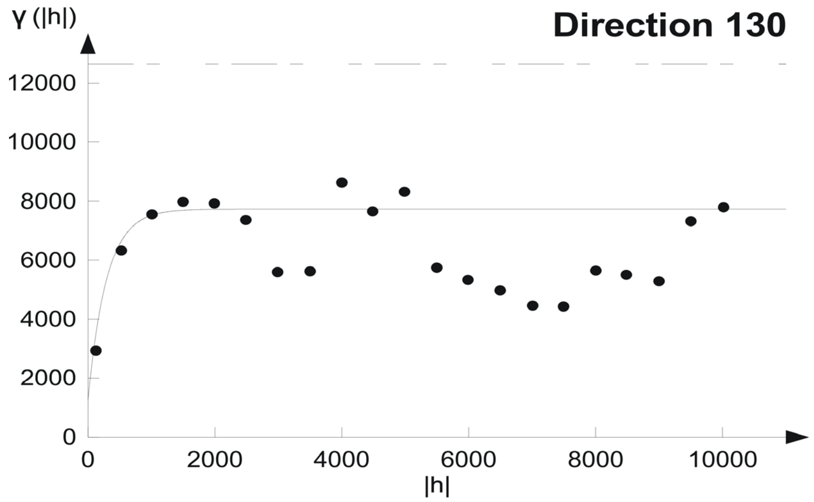


Figure 2.7: Variogram surface of d50 value (left) and depth value (right) with 11 lags of 3000 m.

For the directional variograms computed over a large distance (i.e. over a distance of 33 km or 11 lags of 3000 m.), a sill was reached in the direction of the largest continuity (40°). In the directions perpendicular to this direction, no sill was reached. This indicates a spatial trend, i.e. an increasing variability with increasing distance, which is caused by a non-stationary mean (i.e. a non-constant mean median grain-size over the BCS). Therefore, for OK the variogram (Figure 2.8) was restricted to a distance of 10 km (with 20 lags of 500 m), which is large enough to cover the interpolation window.

For KED the experimental variogram (Figure 2.9) was estimated using increasing lag spacings between 500 and 1000 m. In this way it was possible to model accurately both the short and long distance patterns during the variogram analysis. The short distance variability is important for the fitting of the nugget and initial behaviour of the variogram, while the long distance variability is important for the fitting of the range and eventually compound or 'nested' models. Only the variogram in the direction of the largest continuity (50°) was calculated, because we consider it as representative for stationary conditions without a trend. For KED a linear trend with the depth (causing the anisotropy) was calculated within each interpolation window. The variogram is also shown at a distance of 10 km, to make it comparable with the directional variograms of OK, although the sill would not change anymore over a larger distance (as 50° is the direction of the largest continuity).



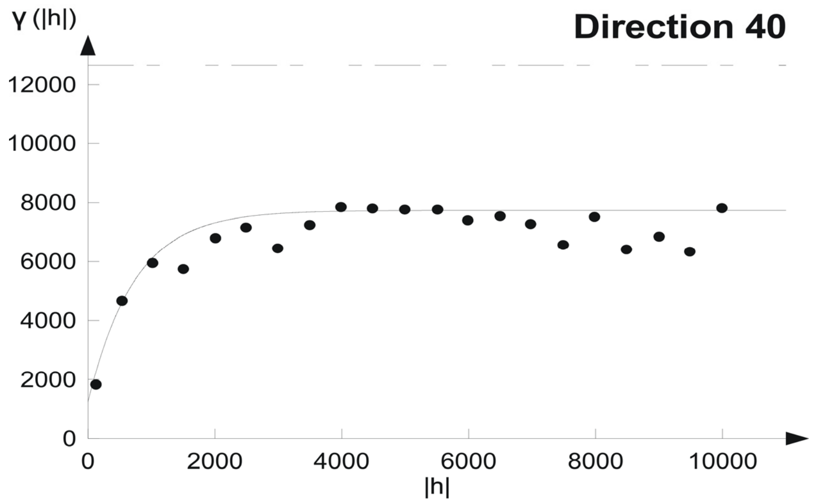


Figure 2.8: Directional variograms in the direction of largest discontinuity (130°) and in the direction of largest continuity (40°), corresponding to respectively the direction perpendicular to the sandbanks and parallel to the sandbanks.

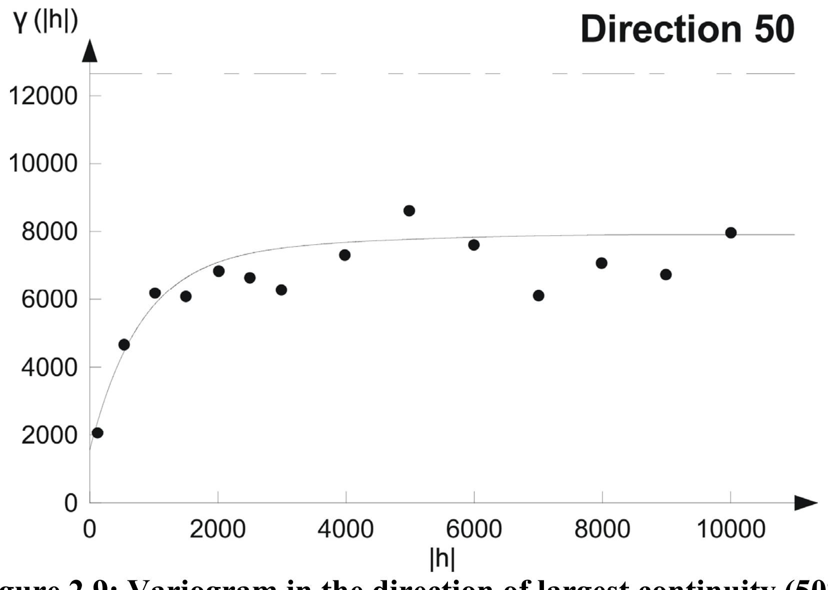


Figure 2.9: Variogram in the direction of largest continuity (50°), considered as an omnidirectional variogram.

To the variogram of OK an exponential model was fit with a nugget of 1240 μm^2, a range of 2200 m in the direction of the largest continuity and a range of 880 m in the direction of the lowest continuity. This represents a geometrical anisotropy, meaning that there are different ranges in different directions. This anisotropy is modelled using an ellipse, with the largest and the smallest range as, respectively the main axis and the side axis. The ratio between the largest and the smallest ranges, is the anisotropy ratio. The anisotropy ratio is 0.40. The sill of the structure has a value of 7740 μm^2.
The theoretical variogram of KED was best modelled as a nested structure. The nugget is 1560 μm^2, the first structure is an exponential model with a range of 2400 m and the second structure is a spherical model with a range of 9000 m, the sill is 7410 μm^2.

Interpolation with kriging

For the calculation of the final OK map, the fitted variogram parameters (nugget, range, sill, anisotropy ratio) were used. Minimum two and maximum 16 observations were required for the interpolation. Quadrants (i.e. circles divided in four equal parts) were used, with a maximal amount of observations of four per quadrant. The search radius was 5000 m. So, points further than 2200 m (i.e. maximal range) were also involved in the interpolation. This is advantageous for locations with a low density of data (e.g. in the northern part of the BCS). These observations obtain very low weights, because they are located outside of the distance of the maximal range, but still carry some information.
The result of the OK (Figure 2.10a) is an almost full coverage map. A strip in the northeast of the BCS is not covered, because data for interpolation are lacking. This map appears quite continuous (except for the three spots in the northern part of the BCS), without the 'bull's eyes' or concentric patterns around data points, typical for 'inverse distance' interpolations. However, the map shows grain-sizes with continuous values across the sandbanks. As no secondary bathymetry information was used for this map, the topography of the seabed cannot be observed inside of the pattern of the median grain-size values.
For the calculation of the KED map, the parameters from the variogram analysis were also used. Besides minimum two and maximum 16 observations and quadrants with maximum four observations are used. The maximal search radius is 9000 m, corresponding to the maximal range.
The result of KED (Figure 2.10b) looks much more realistic than the result of OK. Moreover the median grain-size varies in proportion to the depth. This is very clear in the Hinderbanks region (northern part of BCS). Values between 400 and 500 µm are mainly found in the swales, while the values between 350 and 400 µm are dominantly found at the sandbanks. This pattern is also clear closer to the coast. Unlike OK, KED made use of the secondary information of the bathymetry. The topography pattern of the seabed can be clearly seen inside of the median grain-size map.

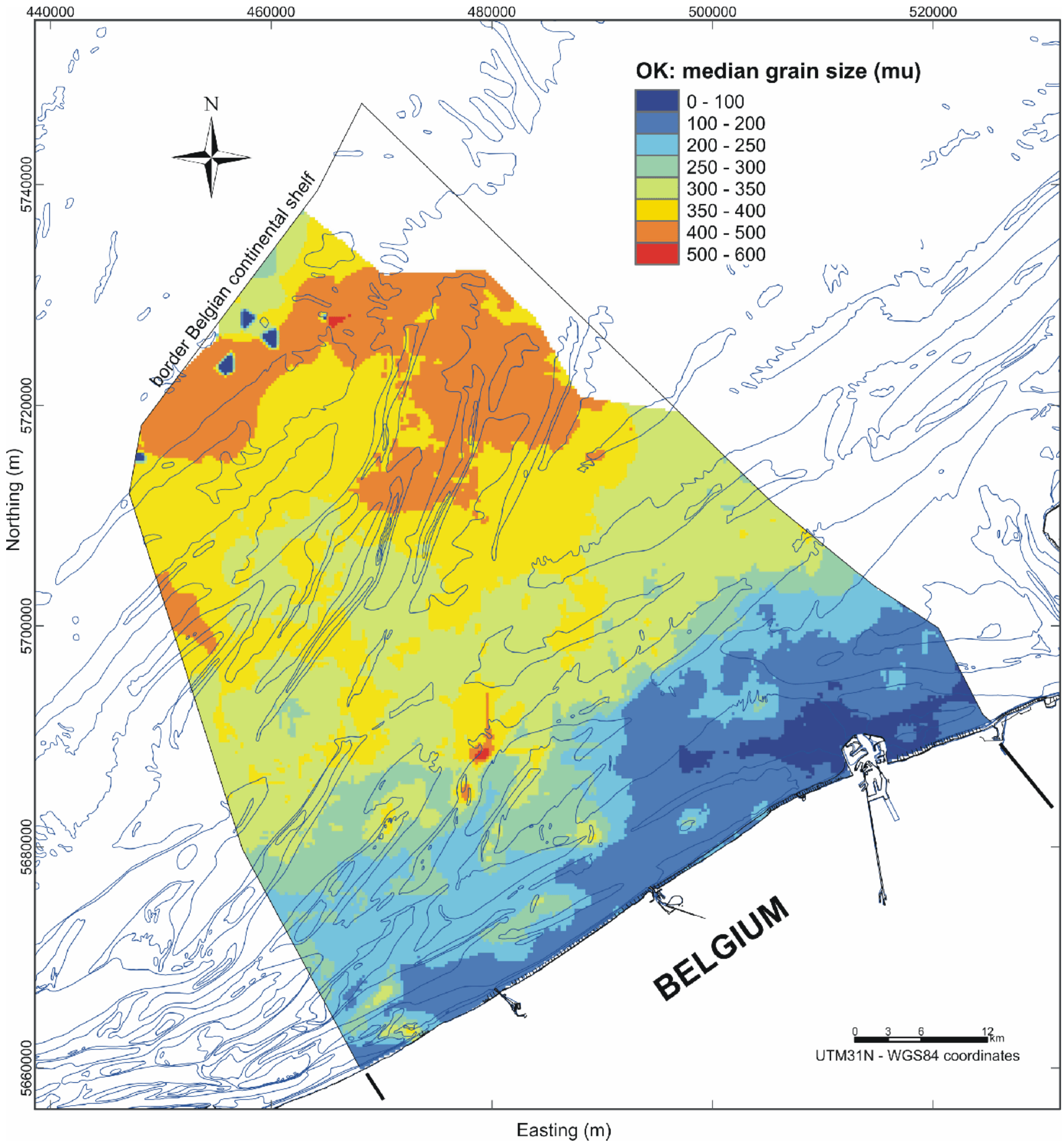


Figure 2.10a: Map of median grain-size, on the basis of ordinary kriging.

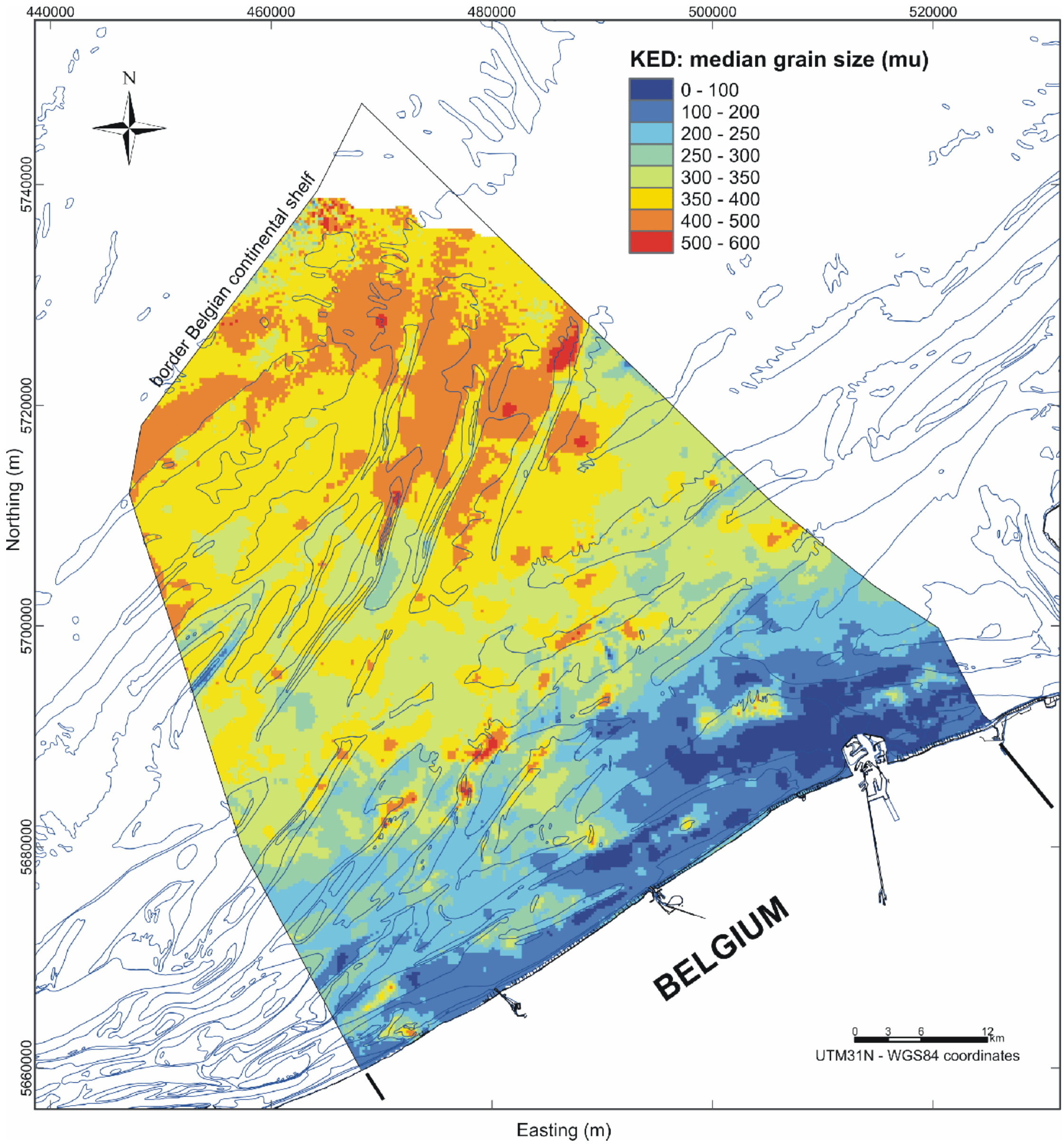


Figure 2.10b: Map of median grain-size, based on kriging with an external drift. The topography of the seabed can be recognized inside of the map, because this methodology uses the bathymetry to assist with the interpolation.

Validation

The scatter plots (Figure 2.11) of the observed versus the estimated values give a first indication in the validation of different techniques. For linear regression the correlation coefficient is much lower than the values for both OK and KED, which demonstrates the inefficiency of this technique compared with both kriging techniques. The correlation coefficient between both values is slightly larger for KED than for OK, indicating that KED gives better results.

Table 2.1: Validation indices of independent validation for linear regression, OK and KED.

	LR	OK	KED
MEE: mean estimation error	-9.17	-8.09	-5.71
MSEE: mean square estimation error	12469.29	7409.95	6745.60
RMSEE: root mean square estimation	111.67	86.09	82.13
MAEE: mean absolute estimation error	74.89	54.97	50.29
Pearson correlation coefficient r	0.42	0.72	0.75

The different parameters are explained in the section on Material and Methods.

However, scatter plots have to be considered in combination with validation indices (Table 2.1). Linear regression yields the largest error for each validation index. KED provides a better result compared to OK, next to a visually more realistic map.

The estimation variance of the kriging analysis gives an indication of the overall reliability of the kriging. This is not an absolute measure of reliability of the kriging estimate (Journel 1993; Armstrong 1994; Goovaerts 1997), but it gives more an indication of the sampling density (a high sampling density means logically a high quality). This is valuable information as it can be used to guide future sampling campaigns. Where the variance reaches high values, new samples are preferably taken. This allows filling gaps and monitoring on a purposive and efficient manner. Figure 2.12 shows the estimation variance of KED. As the KED and the OK map are based on the same samples, only the result of KED is given in Figure 2.12. For the interpolation the extreme minimum of two observations was used, to obtain a map that approaches a full coverage map. Figure 2.12 indicates clearly where this minimum of two observations is too low to give a reliable grain-size value.

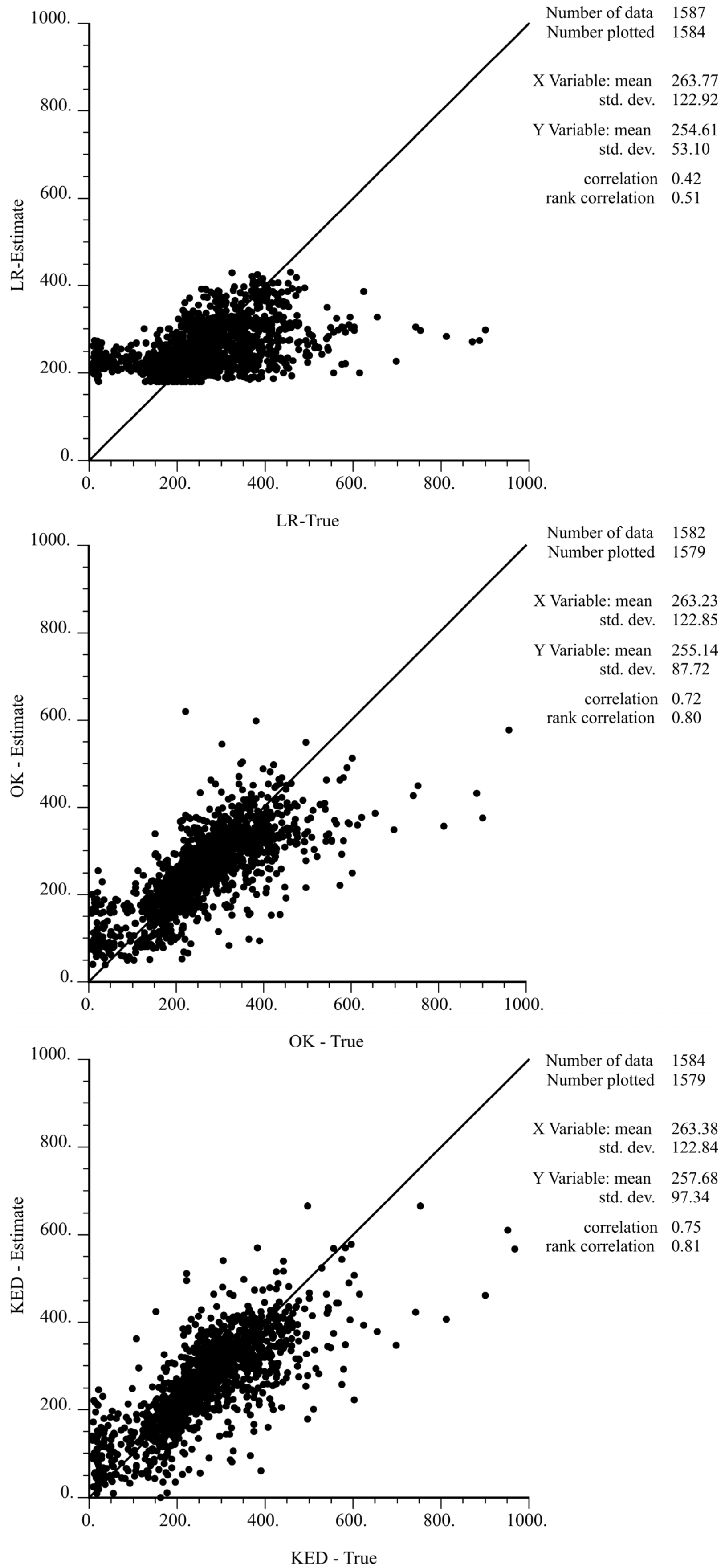


Figure 2.11: Independent validation:
scatter plot of true compared to estimated value using linear regression (top), OK (middle) and KED (bottom).

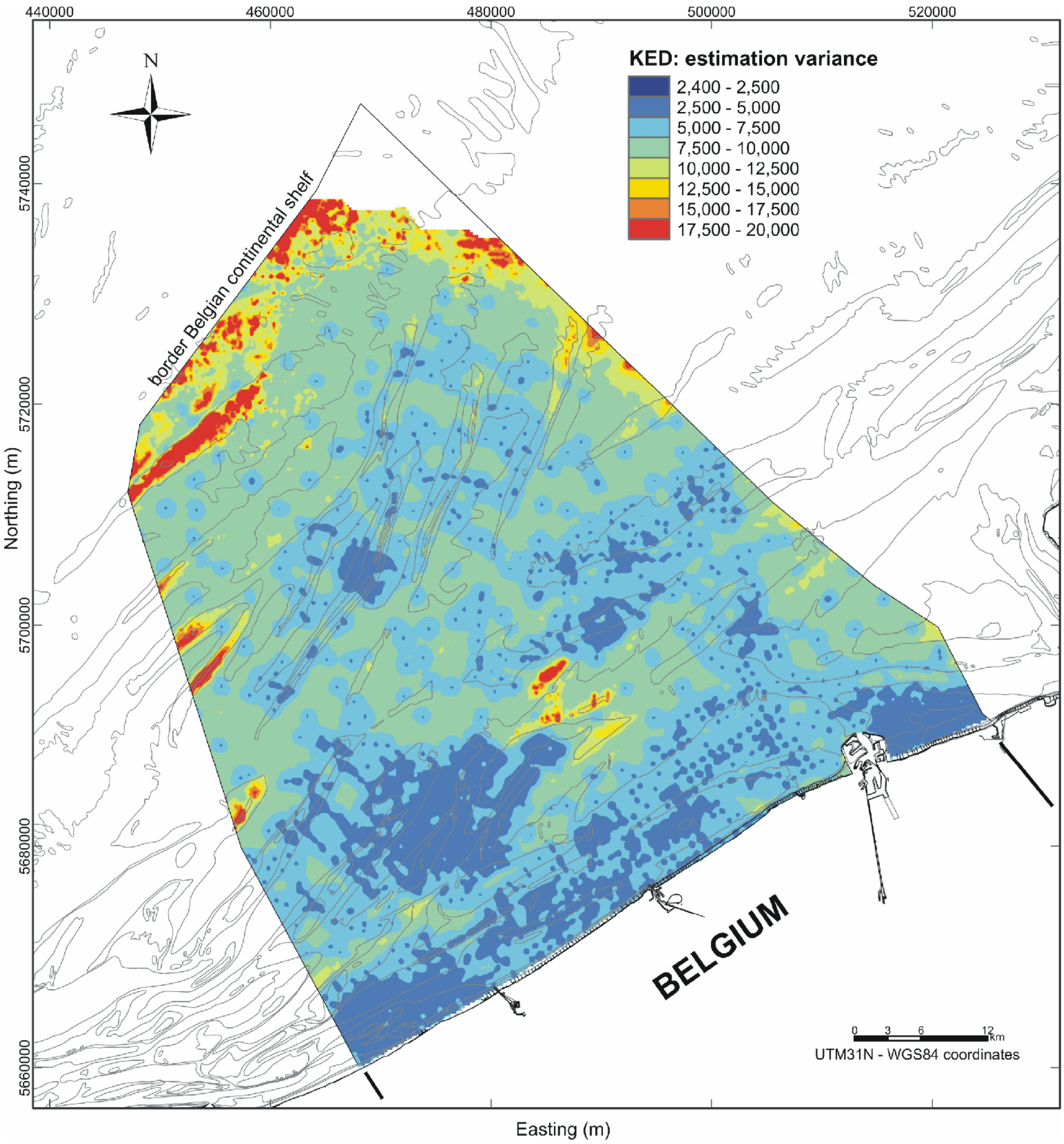


Figure 2.12: KED estimation variance of the median grain size.

2.4 Discussion

The result of KED is a high-quality and high-resolution map (250 x 250 m) of the median grainsize at the BCS, using the bathymetry as secondary information (Figure 2.10b). Leecaster (2003) also used a multivariate kriging technique in combination with bathymetry for the mapping of the grainsize in Santa Monica Bay, California. She showed that the inclusion of depth in the model improved the prediction in the depth-defined areas like canyons, canyon lips and shortbanks. Most applications of multivariate geostatistics using a DEM as secondary information are, however, found in the soil science (Bourennane et al. 2000; Bourennane and King 2003; and Hengl et al. 2004) and climatology (Goovaerts 1999; Hudson and Wackernagel 1994; and Martinez-Cob and Cuenca 1992). As such, it was a challenge to apply and test these techniques in a complex marine environment, dominated by a high spatial variability imposed by sandbanks. Moreover, the technique was used over a large area (3600 km²) comprising a nearshore, coastal and offshore zone, each with different

morphologies. Although the data availability drastically decreased in an offshore direction, the results were very satisfactory.
By comparing and validating linear regression and the two-kriging techniques (OK and KED), it is obvious that kriging is a better interpolation method. Moreover, for data which are unevenly distributed (such as the samples of the median grain-size), kriging has a declustering effect because of its well-known 'screening effect' (see e.g. Goovaerts 1997). This results from the fact that kriging considers both the distance to the interpolation point as the sampling configuration (i.e. distance between observations). Consequently, kriging is preferable to non-declustering techniques (such as linear regression or inverse distance interpolation) for situations with unevenly distributed data.
In cases of a general anisotropy or trend (drift), one solution is to use a small search neighbourhood so that one can assume local stationary conditions within it. This is clearly not a solution in this situation where quite abrupt local changes of the bathymetry occur which needs to be modelled as a local trend. So even locally stationary conditions cannot be assumed over the entire study area. Therefore a non-stationary method like KED should be used. A discussion on this topic is given in Meul and Van Meirvenne (2003).
For applications outside the Belgian shelf, some precautions are needed as the correlation between the bathymetry and the grain-size will depend on the morphology, topography and on the substrate type. However, it is likely that some level of correlation will exist (e.g. Leecaster 2003). The study area of this paper has a very definite presence of the sandbanks dominating the topography of the seabed. The grain-size is expected to vary following the alternation of sandbanks and swales. However, it remains open to discussion whether the environmental setting of the study area defines the benefit of KED and whether this should be investigated first. Where the bathymetry is not a dominating characteristic of the study area, other secondary information (e.g. current and wave parameters) might control the pattern of grainsizes. At a local scale, it can be expected that the grain-size is more related with the geomorphology than with the bathymetry as such. However, this kind of data is more difficult to obtain and is more vulnerable to subjectivity than bathymetry data. However, future research will test the potential correlation as, nowadays, several algorithms exist to estimate the morphological variance from bathymetric-derived features (such as depressions, crests, flats and slopes, both on a large- and small-scale). The calculation of the bathymetric position index is the most widely available technique and is a measure of where a location, with a defined elevation, is relative to the overall landscape (Weiss 2001; Iampietro and Kvitek 2002; and Lundblad et al. 2006). Further research will focus on the relation between bathymetric-derived features and physical datasets such as the median grain-size. Furthermore the geomorphology will be analysed in the context of marine habitat mapping, as topographic features are assumed to be important possible habitats for marine organisms.
If a good correlation can be found between the grain-size distribution and the bathymetry (or other environmental variables such as bathymetric-derived features), detailed grain-size maps can be produced. These have considerable advantages compared to the traditional static sedimentological mapping. Most sedimentological maps are based on the Folk classification (Folk 1974) giving a percentage of gravel, sand and mud. These maps remain highly valuable, but are rather difficult to use for detailed purposes. Nowadays, there is however a need for detailed maps that give a direct reflection of the grain-size itself. Median grain-size values become more

widespread available as also bathymetry data that can assist the interpolation. This combination of information is crucial to define the most suitable areas for aggregate extraction and to reserve these areas in a spatial planning context. Moreover, numerical sediment maps are needed to serve as an input layer for various modelling initiatives. These relate to sediment transport modelling or to the predictive modelling of the distribution of soft substrata habitats. In literature, it has been shown that macrobenthic communities in sandy shelf environments have a clear relationship with well-defined ranges of median grain-size and silt-clay percentage (Van Hoey et al. 2004; Lu 2005; and Willems et al. 2008). As such a mapping of these variables, and their querying, enables direct predictions that are biologically relevant. This calls however for detailed sedimentological maps and these are rarely available. In an international context, there is also a growing interest in sedimentological maps and this related to the concept of 'Marine Landscapes' (Roff and Taylor 2000; Roff et al. 2003; and Golding et al. 2004). Generally, marine landscape modelling is an approach that uses geophysical data as a surrogate for biological mapping. Biological data are only used for the validation of the marine landscapes in terms of their biological relevance. In most cases, the approach remains rather broad-scale, mostly because of the limited detail of the sedimentological maps that are used in the analysis. Schelfaut (2005) used however the detailed grain-size map, described in this paper, and was able to obtain very detailed marine landscapes with high relevance towards the biological value. Future research will focus on the mapping of other target variables, such as the silt-clay percentage, because of its importance for the mapping of the occurrence of the macrobenthos in soft substrata (Van Hoey et al. 2004; Lu 2005; and Willems et al. 2008).

2.5 Conclusions

There is a growing need for a detailed mapping of the seafloor and this is required at a full coverage basis. Apart from the bathymetry, the most crucial variable is sedimentology, as it rules sediment transport processes and it is often the missing link for the prediction of the occurrence of soft substrata habitats or macrobenthic communities/species. The median grain-size was chosen as environmental parameter, as this parameter is the most calculated by a wide variety of scientists and the most frequently used in modelling studies. Hence, a sound interpolation of these data is highly valuable for a wide range of disciplines.

Kriging techniques proved to be the most promising tools to obtain a detailed and high quality map of the median grain-size distribution. These techniques differ from other linear estimation techniques in their aim to minimize the error variance. In addition, kriging with an external drift allowed using correlated secondary information such as bathymetry to assist in the interpolation. Linear regression, ordinary kriging (OK) and kriging with an external drift were compared and validated using an independent dataset. Several validation indices were involved. The independent validation showed that the KED map of the median grain-size is much better than the results obtained using linear regression and better than using OK.

KED enabled to obtain a high-quality and high resolution map (250 x 250 m) of the median grainsize at the BCS, using the bathymetry as secondary information. The resulting map is more realistic and separates clearly the sediment distribution over a complex of sandbanks and swales.

2.6 Acknowledgements

This paper is part of the Ph.D. research of the first author, financed by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). The study frames into the research objectives of the project MESH (''Development of a framework for Mapping European Seabed Habitats''), financed by the INTERREG IIIB North West Europe Programme. It also frames into the research objectives of the projects MAREBASSE (''Management, Research and Budgeting of Aggregates in Shelf Seas related to End-users'', EV/02/18A) and BWZee (''A Biological Valuation Map for the Belgian Continental Shelf'', EV/88/37B), financed by the Belgian Science Policy.
The bathymetric data were delivered by the Ministry of the Flemish Community, Department of Environment and Infrastructure, Waterways and Marine Affairs Administration, Division Coast, Hydrographic Office. The sedisurf@ database has been compiled by the Renard Centre of Marine Geology and consists out of data from various institutes.
Finally, the authors would like to thank the reviewers for their very useful comments and suggestions.

Chapter 3

Geostatistical modelling of sedimentological parameters using multi-scale terrain variables: application along the Belgian Part of the North Sea

ACCEPTED FOR PUBLICATION AS:

Verfaillie, E., Du Four, I., Van Meirvenne, M. and Van Lancker, V., accepted. Geostatistical modelling of sedimentological parameters using multiscale terrain variables: application along the Belgian Part of the North Sea. International Journal of Geographical Information Science.

3 Geostatistical modelling of sedimentological parameters using multi-scale terrain variables: application along the Belgian Part of the North Sea

Abstract

The sediment nature and processes are the key to the understanding of the marine ecosystem, and can explain particularly the presence of soft-substrata habitats. For predictions of the occurrence of species and habitats, detailed sedimentological information is often crucial.

This paper presents a methodology to create high quality sedimentological data grids of grain-size fractions and the percentage of silt-clay. Based on a multibeam bathymetry terrain model, multiple sources of secondary information (multi-scale terrain variables) were derived. Through the use of the geostatistical technique, Kriging with an external drift (KED), this secondary information was used to assist in the interpolation of the sedimentological data. For comparison purposes, the more commonly used Ordinary Kriging technique, was also applied. Validation indices indicated that KED gave better results for all of the maps.

Keywords: Multivariate geostatistics; sedimentology; topography; ecogeographical variables; Belgian part of the North Sea

3.1 Introduction

For marine habitat mapping and spatial planning purposes, high quality maps of ecogeographical variables (EGVs), that assist in the prediction of the occurrence of biological species or communities are invaluable (Derous et al. 2007 and Degraer et al. 2008). For soft substrata habitats, the grain-size and the silt-clay percentage are often the most determining EGVs for the modelling of macrobenthic species (Wu and Shin 1997; Van Hoey et al. 2004; Willems et al. 2008). As such, interpolated data of these sedimentological variables are required, if full-coverage maps of macrobenthos are needed for scientific or management purposes. However, the occurrence of macrobenthic species or communities is known to be patchy or bound to topographic variation (Rabaut et al. 2007); as such, more detailed sedimentological information is required if targeted predictions of macrobenthos are to be made (e.g. impact assessments). Consequently, (multi-scale) terrain characteristics are believed to be important EGVs also (Guisan and Thuiller 2005; Baptist et al. 2006 and Wilson et al. 2007).

EGVs that cover entire parts of the seafloor (e.g. derived from high-resolution multibeam bathymetry), represent well the topographical and morphological variation; however, this is seldom the case when the sedimentological variability is considered. Mostly, sedimentological data are interpolated from poorly distributed sediment sampling points and most often inadequate techniques are being used for the interpolation. Verfaillie et al. (2006) and Pesch et al. (2008) argumented already that the quality of the sedimentological maps can be improved significantly, if complex geostatistical interpolation methods are applied.

Multivariate geostatistics can be considered when there is a linear correlation between the variable and a secondary dataset. In Verfaillie et al. (2006), one secondary dataset (Digital Terrain Model or DTM) was used to create a high quality map of the median grain-size of the sand fraction (fraction between 63 and 2000 µm), based on Kriging with an External Drift (KED). However, if more than one secondary dataset is available, that correlates with the sedimentological variable, improved results can be obtained (e.g. Kyriakidis et al. 2001; Bourennane and King 2003; Reinstorf et al. 2005; Hengl et al. 2007a and Miras-Avalos et al. 2007). Furthermore, Verfaillie et al. (2006) demonstrated that interpolations based on linear regression and Ordinary Kriging (OK), resulted in respectively bad and relatively good results, compared to KED.

Our aim was to produce high quality maps of d_s10 (10th percentile of the sand fraction), d_s50 (median grain-size of the sand fraction), d_s90 (90th percentile of the sand fraction) and silt-clay% (fraction below 63 µm) using KED (Goovaerts 1997) with multiple secondary datasets, derived from multibeam bathymetry. For unimodal sandy sediments, maps of the d_s10 and d_s90 are in principle very similar to those of d_s50. Still, for skewed grain-size distributions, with extreme fine or coarse fractions, those variables can be important to explain presences of certain species or communities.

This paper will demonstrate particularly the strength of advanced geostatistical techniques to model a suite of sedimentological variables, using multiple secondary EGVs.

3.2 Materials and methods

3.2.1 Study area and datasets

The study area (Figure 3.1) was situated on the Belgian part of the North Sea (BPNS), at about 16 km away from the harbour of Zeebrugge and very close to the Belgian-Dutch border. Depths were between 15 and 24 m MLLWS (Mean Lowest Low Water at Spring tide). Important geomorphological and ecological values characterise this area. Large- to very large sand dunes (sensu Ashley 1990) were present in the area, reaching heights of 2.5 m, with wavelengths of a few hundred meters.

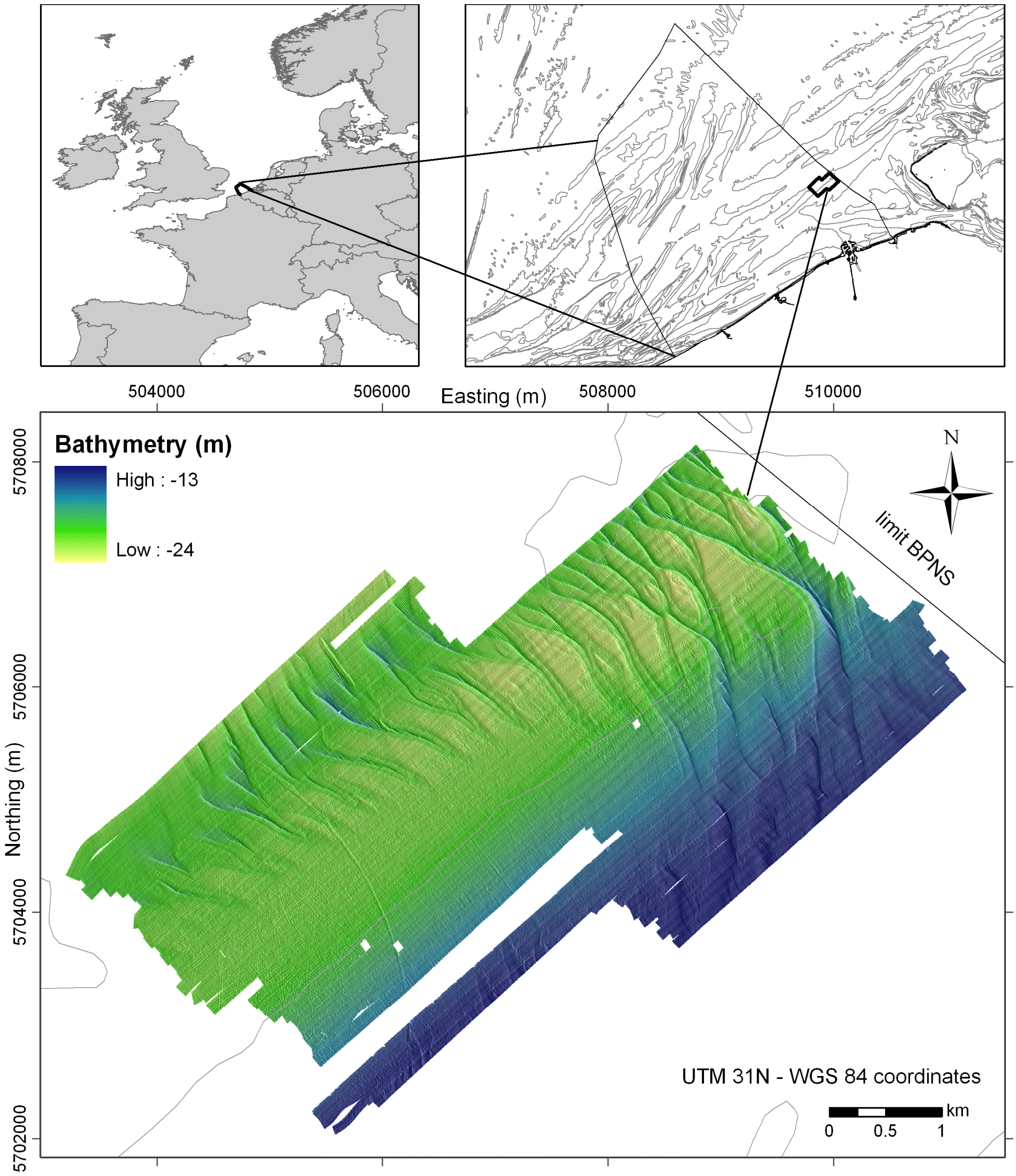


Figure 3.1: Study area (bottom), located in Europe (top left) and the Belgian part of the North Sea (BPNS) (top right).
Large- to very large sand dunes (sensu Ashley 1990) are present in the area.

The sedimentological dataset consisted out of 97 samples, collected during 2 campaigns (RV/*Belgica* 2006/11/20-24 and 2007/11/26-30). A stratified random sampling approach was chosen, based on previously acquired multibeam bathymetry. Sedimentological samples were analyzed with a Malvern Mastersizer 2000 laser particle size analyzer (Malvern Instruments 2008). New multibeam bathymetry (Kongsberg Simrad EM1002S) data were acquired also during the 2 sampling campaigns. For this study, the bathymetry datasets were processed at a resolution of 5 m.
Software used was Variowin 2.21 (Pannatier 1996) for the variogram analysis of the sedimentological datasets; gstat 0.9-42 (Pebesma 2004), implemented in R 2.6.1 (R version 2.6.1 2007) for the geostatistical analysis; ArcGIS 9.2 for GIS analyses and modelling; Biomapper 3.2 (Hirzel et al. 2002b; Hirzel et al. 2006) for the Principal Component Analyses (PCA); and SPSS 15.0 for the correlation analysis of the sedimentological data with the EGVs.

3.2.2 Research strategy

The research strategy consisted out of three steps (Figure 3.2): (1) the selection of relevant EGVs as secondary variables for KED; (2) geostatistical interpolation, based on KED and OK; and (3) comparison of the results.

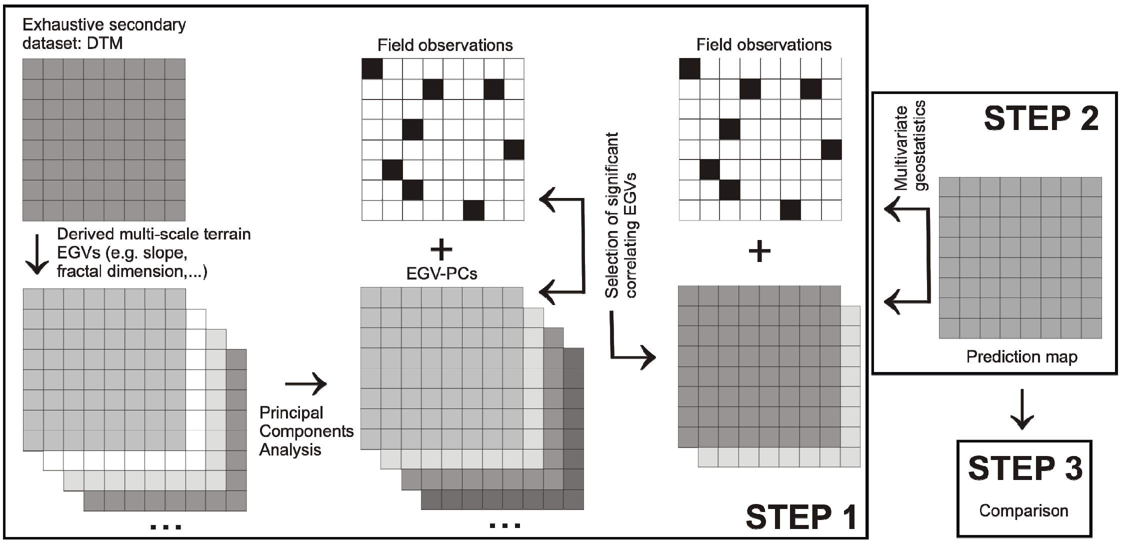


Figure 3.2: Research strategy:
Step 1: The full coverage Digital Terrain Model (DTM) was subjected to a multi-scale terrain analysis, resulting in a set of derived Ecogeographical Variables (EGVs). After a Principal Components Analysis, a Pearson correlation between the field observations and the secondary datasets was calculated. Only significantly ($p \leq 0.05$) correlating Principal Components (PCs or EGV-PCs) were retained as secondary variables for Kriging with an external drift (KED); Step 2: Field observations were interpolated using KED with the selected EGV-PCs as secondary information. Ordinary Kriging (OK) was also applied on the field observations without secondary information (not shown in the scheme); and Step 3: Results of KED and OK are compared and evaluated.

3.2.3 Selection of EGVs as secondary variables for KED

Based on the DTM, a range of multi-scale characteristics were derived that could be used as secondary datasets for KED (slope, eastness, northness, profile curvature, plan curvature, mean curvature and fractal dimension; cfr. Wilson et al. 2007, for an overview and description). Each variable was calculated on 5 different spatial scales, ranging from fine- (15 m) to large-scale (155 m). Window sizes of 3, 7, 13, 21 and 31 cells were applied (with a resolution of 5 m, this corresponded respectively to lengths of 15, 35, 65, 105 and 155 m). In this paper, the dataset of multi-scale characteristics were called 'terrain EGVs'.

To avoid multicollinearity (i.e. high degree of linear correlation) of the terrain EGVs, a PCA was applied. The PCA is based on a correlation matrix, implying that the Kaiser-Guttman criterion can be applied (Legendre and Legendre 1998). This means that Principal Components (PCs) with eigenvalues larger than 1 were preserved as meaningful components for the analysis.

A Pearson correlation coefficient was calculated between the PCs (or EGV-PCs) and the sedimentological point data (d_s10, d_s50, d_s90 and silt-clay%). The selection of EGV-PCs as secondary datasets for the geostatistical modelling was based on statistically significant correlations ($p \leq 0.05$) and the visual inspection of linearity on a scatter plot.

3.2.4 Interpolation with OK and KED

Kriging requires a variogram analysis. The variogram $\gamma(\mathbf{h})$ represents the average variance between observations, separated by a distance $\mathbf{h}$. This value is important in the description and interpretation of the structure of the spatial variability of the investigated regionalized variable (Journel and Huijbregts 1978). The 'sill' is the total variance s^2 of the variable, the 'range' is the maximal spatial extent of spatial correlation between observations of the variable and the 'nugget variance' represents random error or small-distance variability.

Geostatistics is based on the concept of Random Functions, whereby the set of attribute values $z(\mathbf{x})$ at all locations $\mathbf{x}$ are considered as a particular realization of a set of spatially dependent Random Variables $Z(\mathbf{x})$ (Meul and Van Meirvenne 2003).

To compare the resulting maps of predictions of the sedimentological data, the datasets were interpolated, both with OK and KED.

OK is the most frequently used kriging technique. The OK algorithm uses a weighted linear combination of sampled points, situated inside of a neighbourhood (or interpolation window) around the location $\mathbf{x}_0$ where the interpolation is conducted. An underlying assumption is that the mean value (m) is locally stationary (i.e. that it has a constant value inside the interpolation neighbourhood). The algorithm can be written as:

$$Z^*(\mathbf{x}_0) = \sum_{\alpha=1}^{n(\mathbf{x}_0)} \{\lambda_\alpha \cdot [Z(\mathbf{x}_\alpha) - m]\} + m = \sum_{\alpha=1}^{n(x_0)} \{\lambda_\alpha Z(\mathbf{x}_\alpha)\} + \left[1 - \sum_{\alpha=1}^{n(\mathbf{x}_0)} \lambda_\alpha\right] \cdot m \qquad (3.1)$$

with λ_α equal to the weights attributed to the $n(\mathbf{x}_0)$ observations $z(\mathbf{x}_\alpha)$; n the total number of observations $z(\mathbf{x}_\alpha)$; $n(\mathbf{x}_0)$ the subset of n, lying inside the interpolation window. The weights λ_α are obtained by solving a set of equations (the kriging system), involving knowledge of the variogram (see e.g. Goovaerts, 1997). These weights are constrained to sum to one, leading to the elimination of the parameter m from the estimator which is thus written as:

$$Z^*_{OK}(\mathbf{x}_0) = \sum_{\alpha=1}^{n(\mathbf{x}_0)} \lambda_\alpha Z(\mathbf{x}_\alpha) \qquad \text{with} \qquad \sum_{\alpha=1}^{n(\mathbf{x}_0)} \lambda_\alpha = 1 \qquad (3.2)$$

KED is a multivariate variant of 'Kriging with a Trend Model' (KT), formerly called 'Universal Kriging'. KED and KT are non-stationary methods, meaning that the statistical properties of the variable are not constant in space (i.e. no constant mean within the interpolation neighbourhood). With KT, the trend is modelled as a function of the spatial coordinates, whilst for KED, the trend $m(\mathbf{x}_0)$ is derived from a local linear function of the secondary variable, which is formulated in each interpolation window (Goovaerts 1997):

$$m(\mathbf{x}_0) = b_0 + b_1 u_2(\mathbf{x}_0) \quad (3.3)$$

with $m(\mathbf{x}_0)$ the trend on location $\mathbf{x}_0$; b_0, b_1 the unknown parameters of the trend, calculated in each interpolation window from a fit to observations; $u_2(\mathbf{x}_0)$ the secondary variable on location $\mathbf{x}_0$.
In the case of more than one secondary variable $u_i(\mathbf{x}_0)$, this formula can be extended to:

$$m(\mathbf{x}_0) = b_0 + b_1 u_2(\mathbf{x}_0) + b_2 u_3(\mathbf{x}_0) + \ldots + b_{i-1} u_i(\mathbf{x}_0) \qquad (3.4)$$

with $m(\mathbf{x}_0)$ the trend at location $\mathbf{x}_0$; b_0, b_1, b_2, b_{i-1} the unknown parameters of the trend, calculated in each interpolation window from a fit to the observations ; $u_2(\mathbf{x}_0)$, $u_3(\mathbf{x}_0)$, …, $u_i(\mathbf{x}_0)$ the secondary variables at location $\mathbf{x}_0$, depending on the number of secondary variables i-1.
The KED estimator has the same form as the OK estimator.
At each location where the primary sedimentological variable $z(\mathbf{x}_\alpha)$ was observed, the residual $r(\mathbf{x}_\alpha)$ was computed:

$$r(\mathbf{x}_\alpha) = z(\mathbf{x}_\alpha) - m(\mathbf{x}_\alpha) \quad (3.5)$$

A major problem concerning KED is that the underlying (trend-free) variogram is assumed to be known. This means that the variogram, estimated from the raw data, is biased if the mean changes from place to place. As such, it is necessary to remove the local mean and estimate the residual variogram (Lloyd 2005). A solution to estimate the underlying variogram, associated with $r(\mathbf{x}_\alpha)$, is to use the variogram in a direction where the drift is not active (Goovaerts 1997; Wackernagel 1998; Hudson and Wackernagel 1994; Lloyd 2005 and Verfaillie et al. 2006). The variogram in this direction can be extended to other directions under the assumption of isotropic behavior of the underlying variogram.
For KED, the secondary data must be available at all primary data locations as well as at all locations being estimated. A more complex multivariate geostatistical technique is cokriging, which does not require this secondary information to be known at all locations being estimated. Cokriging is much more demanding than other kriging techniques because both direct and cross variograms must be inferred and jointly modelled and because a large cokriging system must be solved (Goovaerts, 1997).
The selected EGV-PCs were used as secondary datasets for KED, resulting into sedimentological data grids of d_s10, d_s50, d_s90 and silt-clay%.

KED was computed in R, based on Hengl (2007b) and Hengl (pers. comm.).

3.2.5 Comparison of OK and KED

To enable a thorough quality control of the geostatistical analysis, based on both OK and KED, a 5-fold cross validation was performed (Fielding and Bell 1997), meaning that the sedimentological dataset was split into 5 partitions and that each partition was withheld one after the other. Several indices are suitable to evaluate the interpolation. These indices are all a measure of the estimation error, which is the difference between the estimated and the observed value:

$$z^*(\mathbf{x}_\alpha) - z(\mathbf{x}_\alpha).$$

(a) The *mean estimation error* (MEE), which has to be around zero to have an unbiased estimator.

$$\text{MEE} = \frac{1}{\text{n}}\sum_{\alpha=1}^{\text{n}}\left(z^*(\mathbf{x}_\alpha) - z(\mathbf{x}_\alpha)\right) \qquad (3.6)$$

(b) The *mean square estimation error* (MSEE), which has to be as low as possible and is useful to compare different procedures. The *root mean square estimation error* (RMSEE) is used to obtain the same units as the variable. This parameter has to be compared to the variance or the standard deviation of the dataset.

$$\text{MSEE} = \frac{1}{\text{n}}\sum_{\alpha=1}^{\text{n}}\left(z^*(\mathbf{x}_\alpha) - z(\mathbf{x}_\alpha)\right)^2 \qquad (3.7)$$

(c) The *mean absolute estimation error* (MAEE), which is similar to the MSEE, but is less sensitive to extreme deviations.

$$\text{MAEE} = \frac{1}{\text{n}}\sum_{\alpha=1}^{\text{n}}\left|z^*(\mathbf{x}_\alpha) - z(\mathbf{x}_\alpha)\right| \qquad (3.8)$$

(d) The *Pearson correlation coefficient* between $z^*(\mathbf{x}_\alpha)$ and $z(\mathbf{x}_\alpha)$, indicates the degree of linear correlation between observed and estimated values. This value has to be considered in combination with the MEE. The correlation coefficient is, in itself, a measure of the proportion of variance explained, hence is related to MSEE.
The validation indices permit comparing the results of OK and KED.

3.3 Results

3.3.1 Selection of EGVs as secondary variables for KED

PCA resulted in 9 PCs, explaining 81.4 % of the total variance. Table 3.1 gives an overview of the selected PCs with the corresponding EGVs with high factor loads (-0.5 < r and r > 0.5). The Pearson correlation coefficients of all 9 PCs with the values of d_s10, d_s50, d_s90 and silt-clay% and the significant linear correlations are presented in Table 3.2. All of the sedimentological variables showed a significant correlation with PC2 and PC6. A selection of scatter plots is presented in Figure 3.3. As the scatter plots of d_s10, d_s50 and d_s90 are very similar for PC2 and PC6, only the scatter plots of d_s90 are given. The correlation coefficient between the silt-clay% and PC2 and PC6 is very weak and only significant at the 0.05 level (Table 3.2). As such, these

scatter plots are not presented in Figure 3.3 and it is expected that the secondary variables PC2 and PC6 will not contribute significantly to the KED interpolation of the silt-clay%. PC2 was mainly explained by multi-scale slope and fractal dimension, while PC6 by multi-scale plan curvature (Table 3.1). Those PCs were the major contributors for the KED analysis. Moreover, d_s90 correlated weakly with PC1 as well, mainly explained by multi-scale mean and profile curvature. This means that the sediment variation was mainly correlated with the combined pattern of slope, fractal dimension and plan curvature and this on different spatial scales.
The correlation coefficient between the sedimentological variables and the other 6 PCs (PC3, PC4, PC5, PC7, PC8 and PC9) were not given, as they were not statistically significant and thus not having a linear relation.

Table 3.1: Principal Components (PCs) showing significant correlations with the sedimentological variables (cfr. Table 3.2), with their corresponding ecogeographical variables (EGVs) and factor loads (between brackets). Only those EGVs are given with factor loads < -0.5 or > 0.5, being the EGVs that are most explaining the PCs.

PC1	PC2	PC6
mcurv_13 (-0.89)	slp_13 (-0.89)	plcurv_21 (-0.67)
mcurv_21 (-0.88)	slp_21 (-0.87)	plcurv_13 (-0.56)
prcurv_13 (-0.83)	slp_7 (-0.79)	plcurv_31 (-0.55)
prcurv_21 (-0.82)	slp_31 (-0.76)	
mcurv_7 (-0.74)	fd_13 (0.65)	
mcurv_31 (-0.72)	slp_3 (-0.62)	
prcurv_31 (-0.67)	fd_7 (0.56)	
prcurv_7 (-0.67)	fd_21 (0.54)	

(mcurv = mean curvature, prcurv = profile curvature, slp = slope, plcurv = plan curvature, fd = fractal dimension, 3, 7, 13, 21 and 33 are multi-scale indices).

Table 3.2: Pearson correlation coefficients between the sedimentological variables and the Principal Components (PCs) and their statistical significance values (p). Only those PCs and correlation coefficients are given that have a statistical significant correlation. Those PCs were used as secondary variables for the Kriging with an external drift analysis.

		PC1	PC2	PC6
d_s10	Pearson correlation		-.537**	.355**
	p		.000	.001
d_s50	Pearson correlation		-.524**	.377**
	p		.000	.000
d_s90	Pearson correlation	-.284**	-.537**	.387**
	p	.008	.000	.000
Silt-clay%	Pearson correlation		.260*	-.263*
	p		.012	.011

**** Correlation is significant at the 0.01 level, * Correlation is significant at the 0.05 level.**

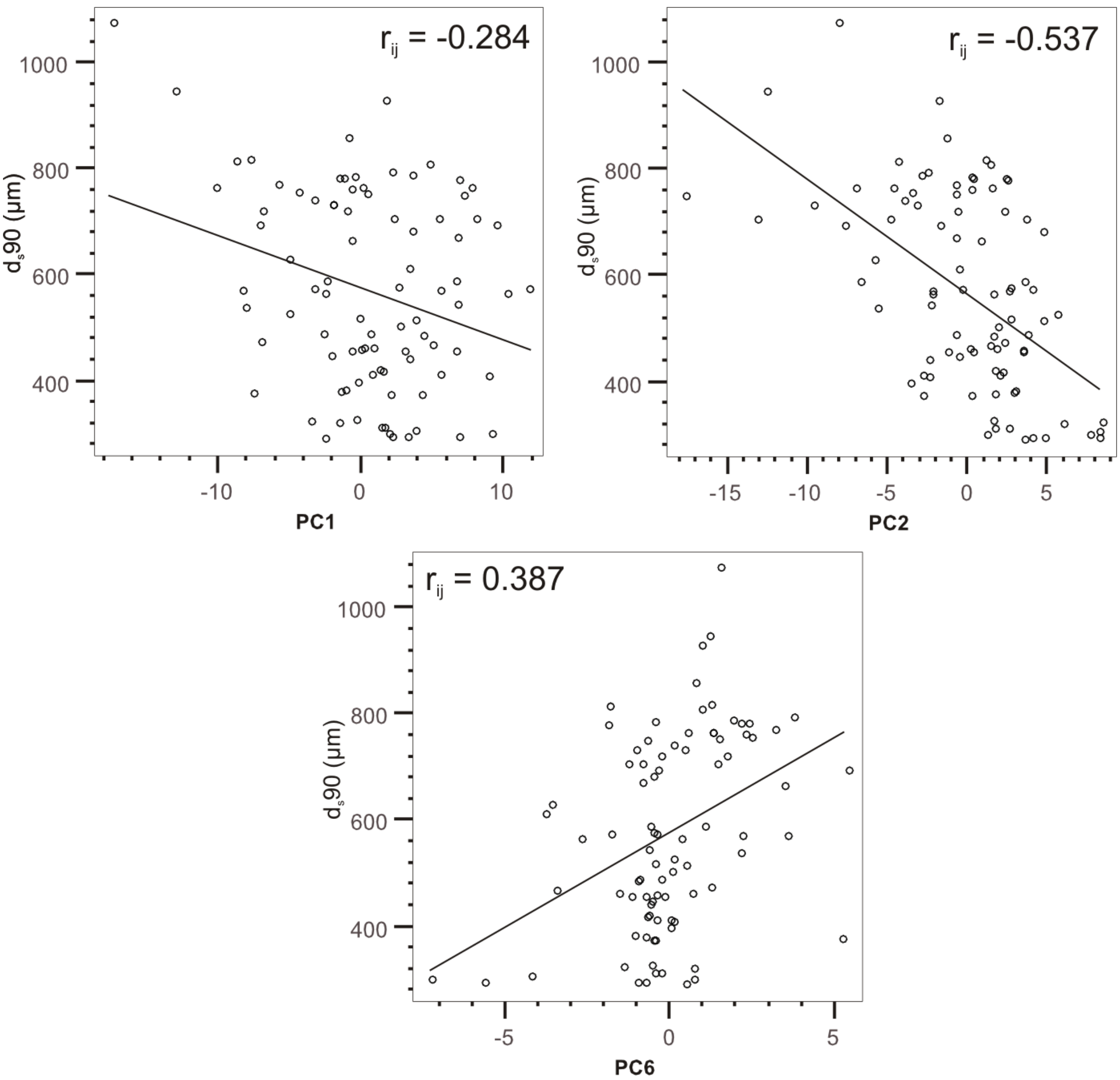


Figure 3.3: Scatter plots showing the Pearson correlation coefficients (r_{ij}) of Table 3.2 between d_s90 and the Principal Components (PCs). Correlation coefficients and scatter plots between d_s10, d_s50 and PC2 and PC6 are very similar; as such scatter plots are not presented. Correlation coefficients between the silt-clay% and PC2 and PC6 are very weak. As such, those scatter plots are not presented.

3.3.2 Interpolation with OK and KED

The variograms for OK and KED are presented in respectively Figure 3.4 and 3.5. All variograms of the sedimentological variables could be fit in a relatively straightforward way, except that of the silt-clay%, which behaved more unstable, due to the relative small values of this variable and the impact of a larger-scale trend.
The variogram surface for each sedimentological variable did not show any obvious anisotropy, still the direction of the strike of the sand dunes (120°, expressed as a trigonometric angle) was considered as the direction of the highest continuity. This means that, in this direction, it was expected that the sedimentological variables were more continuous than in other directions. It is logical that in the direction of the strike of a sand dune, similar sedimentological characteristics are found, while those characteristics are different in a perpendicular direction. Two OK variograms and data

grids per sedimentological variable were created, with an omnidirectional and a directional variogram (being the direction of the strike of the sand dunes). The two results were compared, based on their validation indices: for d_s10 and silt-clay%, a directional variogram gave the best result, whilst for d_s50 and d_s90, an omnidirectional variogram scored best.
For KED, the direction of the strike of the sand dunes, was considered as a drift-free direction. As such, the variogram of this direction was considered as omnidirectional and was used for the analysis.

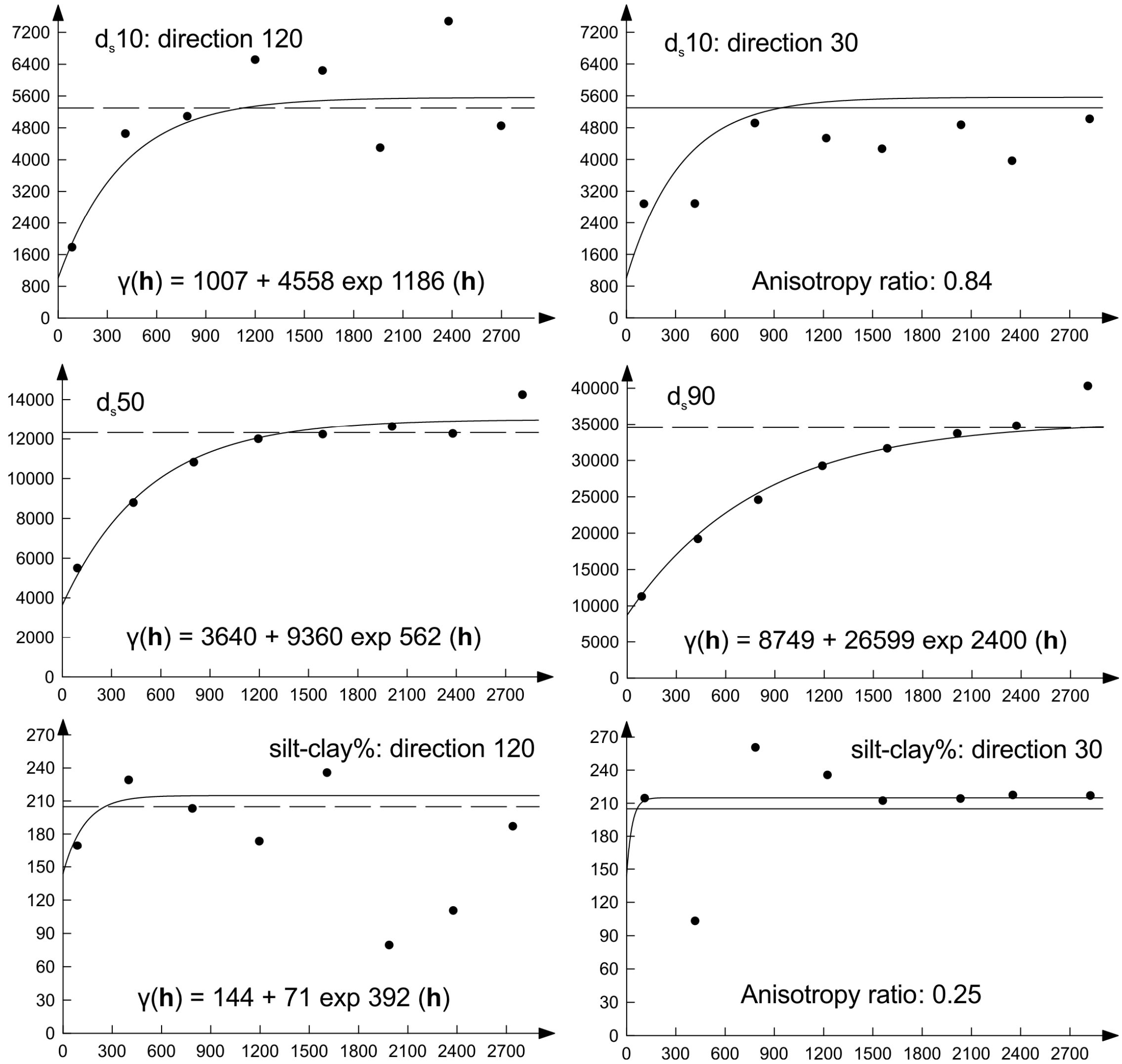


Figure 3.4: Experimental and fitted variograms for Ordinary Kriging (OK): X-axis represents lag distance (m) and the Y-axis is the semivariance (units are µm² for d_s10, d_s50, d_s90 and %² for silt-clay%). Variogram models are expressed as $\gamma(h) = C_0 + C_1$ exp a (h), with C_0 = nugget effect, C_1 = sill, exp = exponential model and a(h) = practical range. Practical ranges are equal to the distance at which 95% of the sill has been reached. Directions are expressed as trigonometric angles (zero degrees = east increasing counter clock wise).

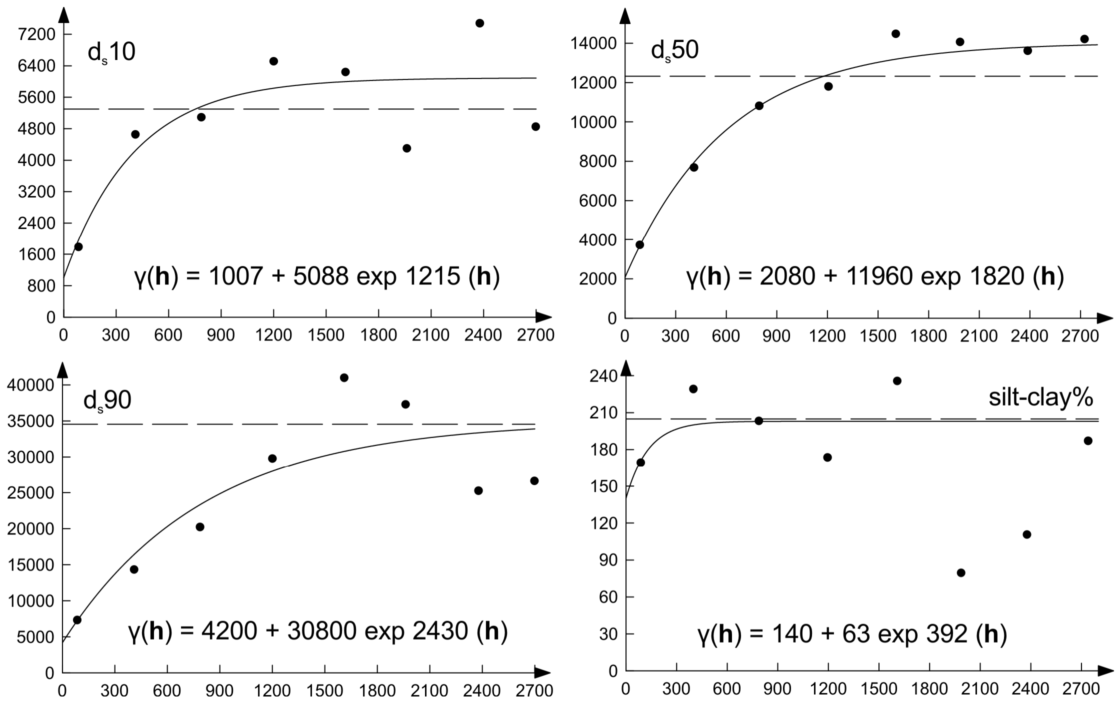


Figure 3.5: Experimental and fitted variograms for Kriging with an external drift (KED),
in the direction of the strike of the sand dunes (120° expressed as a trigonometric angle; zero degrees = east increasing counter clock wise); they are considered omnidirectional, because of the assumption that this direction is drift-free. The X-axis represents the lag distance (m) and the Y-axis is the semi-variance (units are µm² for d_s10, d_s50, d_s90 and %² for silt-clay%). Variogram models are expressed as γ(h) = C_0 + C_1 exp a (h), with C_0 = nugget effect, C_1 = sill, exp = exponential model and a(h) = practical range. Practical ranges are equal to the distance at which 95% of the sill has been reached.

Figure 3.5 shows the maps of the resulting sedimentological data grids, modelled with OK and KED. The blanked zones are due to missing data; their surface area has been enlarged due to the multi-scale analysis (with window sizes of maximum 31 cells).
The results of d_s10, d_s50 and d_s90 are very similar. As such, no outliers of extreme fine or coarse fractions are present; the sediment is very homogeneous and well sorted. The OK maps are smooth and rather unnatural, in the sense that they show concentric patterns around the data points, whilst the KED maps reflect well the variation of the natural environment. Still, the two methodologies showed the same trend: coarser grain-sizes on the sand dunes and finer grain-sizes between and away from the sand dunes. The influence of the underlying topography was very clear in the results from KED. The same trend, showing a difference between the sand dunes (low silt-clay%) and the area away from the dunes (higher silt-clay%), holded true for the silt-clay%. The rough, mottled pattern away from the dunes, and visible on all of the KED maps, was due to the presence of dense colonies of tube worms; their existence was validated with extensive terrain verification.

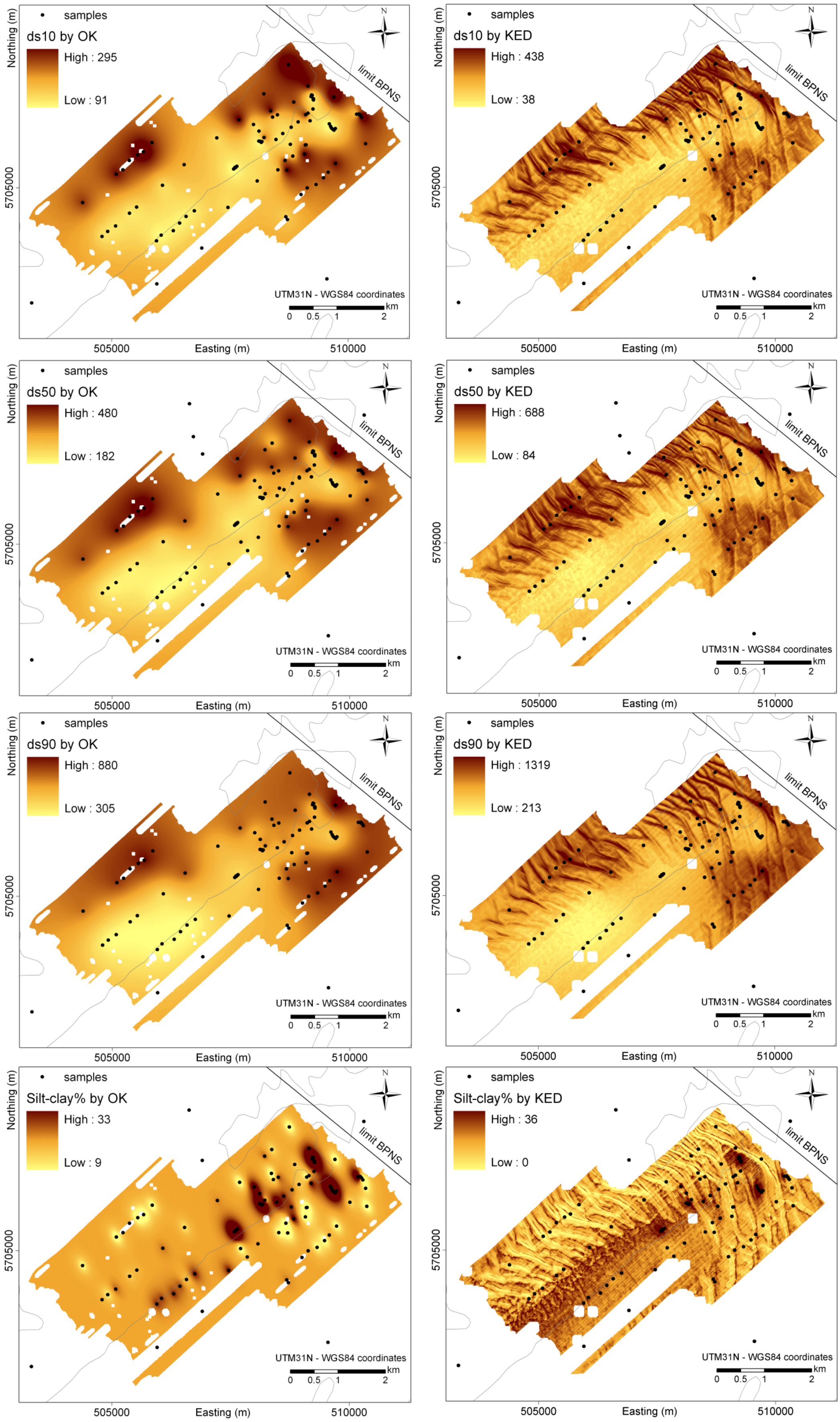


Figure 3.6: Sedimentological maps, based on Ordinary Kriging (OK) (left) and Kriging with an external drift (KED) (right).

3.3.3 Comparison of OK and KED

The validation indices are given in Table 3.3. KED provided a better result, compared to OK for all of the indices of d_s10, d_s50 and d_s90. From this, the KED results of d_s10, d_s50 and d_s90 could be considered better than those of OK.
For the silt-clay%, the result of OK was highly comparable to the result of KED. The MEE and Pearson correlation coefficient between the observed and the estimated values were better for OK compared to KED. The other validation indices were slightly better for KED compared to OK. This was due to the low correlation coefficient between silt-clay% and PC2 and PC6 (Table 3.2), meaning that the contribution of the secondary variables for KED was limited. The significant correlation coefficients between d_s10, d_s50, d_s90 and the PCs were all significant at the 0.01 level, while for silt-clay%, the correlation was significant at the 0.05 level (the lower the significance level, the stronger the evidence) (Table 3.2).
Next to the better validation indices, KED gave visually more natural maps.

Table 3.3: Validation indices (cfr. Materials and Methods) of different sedimentological data grids.
Except for the MEE and the Pearson correlation coefficient of the silt-clay%, all validation indices give better results for Kriging with an external drift (KED) compared to Ordinary Kriging (OK).

	d_s10_{OK}	d_s10_{KED}	d_s50_{OK}	d_s50_{KED}	d_s90_{OK}	d_s90_{KED}	$Sc\%_{OK}$	$Sc\%_{KED}$
MEE	2.44	**-0.55**	6.85	**-1.22**	3.09	**2.48**	**-0.42**	-0.51
RMSEE	63.01	**56.50**	93.51	**82.78**	134.78	**121.68**	13.09	**13.04**
MAEE	46.32	**40.47**	71.99	**64.69**	104.04	**93.82**	9.90	**9.82**
r	0.52	**0.64**	0.55	**0.67**	0.68	**0.75**	**0.50**	0.46

Sc% = silt-clay%, in bold are the best results.

3.4 Discussion

The aim of this paper was to create high quality sedimentological data grids, using multiple sources of secondary information. Next, the following items will be discussed: the secondary variables for KED and the comparison between OK and KED.

3.4.1 Secondary variables for KED

The proposed methodology allowed using a whole set of secondary variables. Here, 34 multi-scale terrain EGVs were derived from the DTM (slope, eastness, northness, profile curvature, plan curvature, mean curvature and fractal dimension). All of them were calculated on 5 different spatial scales, ranging from fine- to large-scale. A PCA reduced the large number of secondary variables to 9 PCs. Three of these PCs correlated significantly with the sedimentological variables. The PCA allowed maintaining a maximum of information, but avoided redundancy of correlating data.
For all of the sedimentological variables, there was a similar subset of PCs and EGVs, correlating significantly with the sedimentology (Table 3.1): mean, profile and plan curvature; slope and fractal dimension, on all different spatial scales. This means that a combination of different spatial scales was important in explaining the

sedimentological variation. Mainly the larger window sizes of 13, 21 and 31 (or 65, 105 and 155 m) were well represented, but also the smaller window sizes of 3 and 7 cells (or 15 and 35 m) were important. Mainly the larger distances were well suited to explain the sedimentological variability imposed by bedforms having wavelengths of around 100 m (very large dunes sensu Ashley 1990), but the smaller distances corresponded more with the smaller dunes (large dunes sensu Ashley 1990). Mainly the EGVs, associated with PC2 and PC6 (multi-scale slope, fractal dimension and plan curvature), were responsible for the overall sedimentological variation, as all of the sedimentological variables were correlated with those PCs. Such a slope – grain-size correlation has also been detected on sandy beaches (McLachlan 1996), while Azovsky et al. (2000) detected a correlation between grain-size and fractal dimension. Fractal dimension (Mandelbrot 1983) is often referred to as a measure of the surface complexity; as such it can be linked to habitat complexity of macrofauna (Kostylev et al. 2005).

Besides topography, possibly other EGVs correlate with the sedimentology and could be valuable secondary datasets for a multivariate geostatistical interpolation: e.g. the correlation between silt and nutrient richness (Greulich et al. 2000); between sand and organic matter content (Mantelatto and Fransozo 1999); and between grain-size and bottom current strength (Revel et al. 1996). Still, no high resolution datasets, other than the DTM, were available for this study area.

Categorical EGVs could be valuable secondary datasets as well (Hengl et al. 2007c). An example of such a dataset could be acoustic seabed classes of the sediment, derived from the classification of multibeam backscatter strength (Van Lancker et al. 2007) or side-scan sonar classes. Still, this information was not available for this study area.

3.4.2 Comparison of KED and OK

Validation indices, as presented in Table 3.3, are a valuable tool, though they permit only a comparison of different interpolation methods, applied on the same dataset. A d_s50_{OK} and a d_s50_{KED} map can be compared and the best result can be evaluated. It is more difficult to compare results from e.g. the d_s10_{KED}, d_s50_{KED}, d_s90_{KED} and silt-$clay_{KED}$ data grids. To overcome this issue, the correlation coefficients of the observed versus the estimated values can be compared. For this study, the coefficient indicates that d_s90_{KED} map is the most reliable.

The validation indices can be compared with the accuracy of the sedimentological variables. The accuracy of the sedimentological analyses is in the range of 1 % (Malvern Instruments 2008). The differences between OK and KED were well above this analytical accuracy. For example, the RMSEE of d_s50 reduced with 10.73 µm (Table 3.3), which represents a relative gain of 11.45 %. For the silt-clay%, where the RMSEE only reduced with 0.05 %, the difference in accuracy between OK and KED was negligible. The interpolation of the silt-clay% was less straightforward than the interpolation of the d_s10, d_s50 and d_s90. This poor increase in accuracy between both interpolation methods was mainly due to the small correlation coefficients between the silt-clay% and the PCs.

3.5 Conclusion

This paper proposed a multivariate geostatistical approach to obtain high quality sedimentological data grids of d_s10, d_s50, d_s90 and silt-clay%. KED was used with multiple secondary variables on different spatial scales, all derived from a DTM of the bathymetry. The sedimentological data were interpolated also with OK, and validation indices enabled to compare both results. For all of the sedimentological variables, KED gave the best result, although the results for the silt-clay% for both OK and KED, were very similar. The maps, based on KED, showed a different pattern on the sand dunes and away from and between the sand dunes. The sand dunes are composed of coarser sand, whilst the zones away from them have finer grain-sizes. The same difference can be observed for the silt-clay%: a high silt-clay% away from the dunes is observed and a low silt-clay% on the sand dunes. This pattern is not at all clear when the results, obtained with OK, were evaluated.

These highly detailed sedimentological data grids are the key for the adequate prediction of biological species, communities or habitats. This is especially the case for the predictive modelling of soft-substrata macrobenthos, of which the occurrence relates highly with sedimentological gradients (e.g. Degraer et al. 2008).

3.6 Acknowledgements

This paper is part of the PhD research of the first author, financed by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). It contributes also to the EU INTERREGIIIB project MESH (‘‘Development of a framework for Mapping European Seabed Habitats’’). In addition, it frames into the research objectives of the project MAREBASSE (‘‘Management, Research and Budgeting of Aggregates in Shelf Seas related to End-users’’, Belgian Science Policy, SPSDII, contract EV/02/18A). The sedisurf@ database has been compiled at the Renard Centre of Marine Geology and consists of data from various institutes.

Tomislav Hengl is acknowledged for his contribution to the R-gstat computation.

Chapter 4

A protocol for classifying ecologically relevant marine landscapes, a statistical approach

SUBMITTED FOR PUBLICATION AS:

Verfaillie, E., Degraer, S., Schelfaut, K., Willems, W. and Van Lancker, V. (submitted to Estuarine, Coastal and Shelf Science). A protocol for classifying ecologically relevant marine landscapes, a statistical approach.

4 A protocol for classifying ecologically relevant marine landscapes, a statistical approach

Abstract

Mapping ecologically relevant zones in the marine environment has become increasingly important. However, biological data are scarce and alternatives are being sought in optimal classifications of abiotic variables. The concept of 'marine landscapes' is based on a hierarchical classification of geological, hydrographic and other physical data. However, this approach is subject to many assumptions and subjective decisions.

Here, an objective protocol is being proposed where abiotic variables are subjected to a statistical approach, using principal components analysis (PCA) and a cluster analysis. The optimal number of clusters is being defined using the Calinski-Harabasz criterion. The methodology has been applied on datasets of the Belgian part of the North Sea (BPNS), a shallow sandy shelf environment with a sandbank-swale topography.

The BPNS was classified into 8 marine landscapes that represent well the natural variability of the seafloor. The internal cluster consistency was validated with a split-run procedure, with more than 99% correspondence between the validation and the original dataset. The ecological relevance of 6 out of the 8 clusters was demonstrated, using indicator species analysis.

The proposed protocol, as exemplified for the BPNS, can easily be applied to other areas and provides a strong knowledge basis for environmental protection and management of the marine environment. A SWOT-analysis, showing the strengths, weaknesses, opportunities and threats of the protocol was performed.

Keywords: marine landscapes, abiotic variables, macrobenthic species, Principal Components Analysis, cluster analysis, Belgian part of the North Sea

4.1 Introduction

Biodiversity is of utmost importance to maintain the long-term stability of ecosystems, certainly with changing environmental conditions, such as global warming (Keytsman and Jones 2007). This applies to both terrestrial and marine habitats, species or communities, many of which are threatened by the ever-growing pressure on their environment.

Several techniques to map the environment are in place and if we consider predictive modelling and classification techniques for habitat mapping, all of them are based on the assumption that the biological value of an area is related to its abiotic characteristics. Generally, species and communities are linked to their substrate type, topographic position and energy regime. The information on the abiotic environment is generally more widely available than biological information itself; as such, it is aimed at distinguishing ecological landscapes on the basis of specific combinations of these abiotic variables.

Terrestrial examples of classifications of abiotic variables can be found in Fairbanks et al. (2000), Jobin et al. (2003), Rosa-Freitas et al. (2007) and Svoray et al. (2007). Similar methodologies can be applied to the marine environment.

In the framework of marine protection or management, available biotic data (e.g. absence/presence of benthic organisms) are often patchy and highly variable in nature. Moreover, offshore areas are generally devoid of samples. In response, the mapping of "marine landscapes" was developed, as a surrogate of biologically driven habitat mapping. If reliable, this methodology would facilitate the development of management measures for offshore areas in the absence of biological data (obtaining biological data in offshore areas is extremely expensive and time consuming). This hierarchical abiotic classification was first proposed for Canadian waters by Roff and Taylor (2000) and Roff et al. (2003). In this concept, biological data are used only passively, as a validation tool afterwards.

The integration of abiotic datasets (e.g. seabed substrata, depth, slope) that lead to a classification of seabed features, can be performed in a Geographic Information System (GIS) and is now applied widely (e.g. Golding et al. 2005; Connor et al. 2006; and Al-Hamdani and Reker 2007). The advantages are that abiotic parameters and processes are relatively easy to observe and monitor. Moreover, they can often be correlated with biological species or communities (Zacharias and Roff 2000). Another advantage is that the GIS process is quite simple, compared to statistical techniques (e.g. clustering as proposed in this study).

Unfortunately, this approach still lacks objectivity: in several stages of the methodology, subjective decisions have to be made: (1) 'Ecologically relevant abiotic variables' have to be selected as input for the GIS analyses. However, since biological data are generally sparse, this selection is not straightforward; (2) the analysis needs abiotic variables, classified into 'relevant classes in terms of biology'. It is very hard to define both the relevant class breaks and the number of classes; and (3) the 'Queries' step that combines the predefined classes of the abiotic variables into new combinations, being the final marine landscapes. As such, there is a strong need for a more objective and repeatable methodology.

This paper proposes a protocol to increase the objectivity of the marine landscapes approach, based on a statistical analysis for the grouping of full coverage abiotic data. The performance of a combination of PCA and cluster analysis will be demonstrated.

The proposed protocol aims for an unsupervised classification of purely abiotic variables. The ecological validation is done afterwards, independently from the PCA and the cluster analysis to test the ecological relevance of the marine landscapes. The Belgian part of the North Sea (BPNS) is an ideal case study area, because of its extensive availability of both abiotic and biotic variables. However, in most cases, abiotic datasets will be available and only a few or no biological datasets.

4.2 Material and Methods

4.2.1 Study area

The BPNS (3600 km²) is situated on the North-West European Continental Shelf. The shelf is relatively shallow and dips gently from 0 to 50 m. The seabed surface is characterized by a highly variable topography, with a series of sandbanks and swales. The sandbanks can be subdivided into four major groups: the Coastal Banks and the Zeeland Banks are quasi-parallel to the coastline, whereas the Flemish Banks and the Hinder Banks have a clear offset in relation to the coast (Lanckneus et al. 2001). The seabed is sandy; the sand fraction (0.063 - 2 mm) is merely found on the sandbanks, whereas coarser sands, gravel (> 2 mm) and higher silt-clay fractions (< 0.063 mm) are found also in the swales (Lanckneus et al. 2001). The sandbanks and the swales are both covered with ripples and dunes. The height of the dunes commonly ranges between 2 and 4 m, though dune heights of up to 11 m are found in the most offshore areas.

Five macrobenthic communities (four subtidal and one intertidal community) are discerned within the mobile substrates of the BPNS (Degraer et al. 2003; and Van Hoey et al. 2004).

On the BPNS, various abiotic datasets are widely available (Van Lancker et al. 2007). In addition, a large dataset of 741 macrobenthic samples (Marine Biology Section, Ugent – Belgium, 2008) can be used for an ecological validation. As such, the BPNS is an ideal test area to develop a new classification method and to validate its ecological relevance.

16 abiotic variables are available for the BPNS (Table 4.1). All of them have a resolution of 250 m, except maximum Chlorophyl a (Chl a) concentration and maximum Total Suspended Matter (TSM) with a resolution of 1000 m. All data grids of the abiotic variables were resampled to 54307 pixels with a resolution of 250 m. Although other abiotic variables (e.g. salinity, temperature, stratification) could be important as well for explaining the presence of benthic species, they were not available for this study. Still, the current dataset represents well the abiotic variability. In the Discussion Section, this topic is discussed in more detail.

4.2.2 Research strategy

The protocol starts with a PCA for data reduction (step 1). The resulting components are then subjected to a hierarchical cluster analysis (step 2) and the cluster centres from step 2 are used as starting positions for a *K*-means partitioning (step 3). In step 4, the optimal number of clusters is calculated; in step 5, a validation of the internal cluster consistency is performed; and in step 6, a species indicator analysis (INDVAL) is done (Dufrêne and Legendre 1997), defining for each cluster a number of

significant indicator species and as such offering the possibility for an ecological validation of the classification.
Software used is SPSS version 12 for PCA, ClustanGraphics version 8.03 for the hierarchical and *K*-means clustering, R version 2.5.1 for the calculation of the Calinski and Harabasz (1974) indices (called C-H in this paper) and PC-ORD 4.41 (McCune and Mefford 1999) for the INDVAL analysis.

4.2.3 Step 1: PCA analysis

For data reduction and to avoid multicollinearity (i.e. high degree of linear correlation) of the abiotic variables, a PCA was performed (theoretical background e.g. in Jongman et al. 1987; Legendre and Legendre 1998). PCA computes a reduced set of new, linearly independent variables, called principal components (PCs) that account for most of the variance of the original variables. The PCs are a linear combination of the original variables. The PCA was based on a correlation matrix, implying that the Kaiser-Guttman criterion could be applied (Legendre and Legendre 1998). This means that PCs with eigenvalues larger than 1 were preserved as meaningful components for the analysis. To maximize the independence of each PC, a Varimax rotation of the PCs was computed. The PCs were the input variables for the cluster analysis.
Similar applications of PCA for data reduction of abiotic variables are found in Cardillo et al. (1999), Fairbanks (2000), Moreda-Piñeiro et al. (2006), and Frontalini and Coccioni (2008).

4.2.4 Step 2: Hierarchical cluster analysis based on Ward’s method

To group the pixels with abiotic data on a statistical basis, a hierarchical clustering, based on Ward’s (1963) or Orlóci 's (1967) minimum variance method was applied on the PCs (theoretical background e.g. in Jongman et al. 1987; Legendre and Legendre 1998). This method is an agglomerative clustering algorithm that minimizes an objective function which is the same “squared error” criterion that is used in multivariate analysis of variance and results into clusters with a minimal variance between each cluster. At each clustering step, this method finds the pair of objects or clusters whose fusion increases as little as possible the sum, over all objects of the squared Euclidean distances between objects and cluster centroids (Legendre and Legendre 1998). The Euclidean distance is an appropriate model for the relationships among abiotic variables (Legendre and Legendre 1998). Applications of Ward’s method for the clustering of abiotic variables can be found in Cao et al. (1997) and Frontalini and Coccioni (2008).

Table 4.1: Abiotic variables as input for the PCA and cluster analysis.

Abiotic variable	Unit	Reference or procedure
Sedimentology		Reference: sedimentological database ('sedisurf@') hosted at Ghent University, Renard Centre of Marine Geology.
• Median grain-size of sand fraction (63-2000 µm) or d_s50	µm	Reference: Verfaillie et al. (2006)
• Silt-clay percentage (0-63 µm)	%	Reference: Van Lancker et al. (2007)
• Sand percentage (63-2000 µm)	%	Reference: Van Lancker et al. (2007)
• Gravel percentage (> 2000 µm)	%	Reference: Van Lancker et al. (2007)
Topography		
• Digital terrain model (DTM) of bathymetry	m	Reference: Flemish Authorities, Agency for Maritime and Coastal Services, Flemish Hydrography All other topographic variables are derived from the DTM
• Slope = a first derivative of the DTM	°	Procedure: Evans (1980); Wilson et al. (2007)
Aspect = a first derivative of the DTM Indices of northness and eastness provide continuous measures (−1 to +1) describing orientation of the slopes.		Procedure: Wilson et al. (2007); Hirzel et al. (2002a)
• Eastness = sin (aspect)	/ (no unit)	
• Northness = cos (aspect)	/	
• Rugosity = ratio of the surface area to the planar area across the neighbourhood of the central pixel	/	Procedure: Jenness (2002); Lundblad et al. (2006); Wilson et al. (2007)
Bathymetric Position Index (BPI) = measure of where a location, with a defined elevation, is relative to the overall landscape		Procedure: Lundblad et al. (2006); Wilson et al. (2007)
• BPI (large scale)	/	
• BPI (small scale)	/	
Hydrodynamics		
• Maximum bottom shear stress = frictional force exerted by the flow per unit area of the seabed	N/m²	Reference: Management Unit of the North Sea Mathematical Models and the Scheldt estuary (MUMM)
• Maximum current velocity	m/s	
Satellite derived variables		
• Maximum near-surface Chlophyl a (Chl a) concentration over a 2-year period (2003-2004)	mg/m³	Reference: MERIS data processed by MUMM in the framework of the BELCOLOUR-2 project (ESA ENVISAT AOID3443)
• Maximum near-surface Total Suspended Matter (TSM): measure for turbidity over a 2-year period (2003-2004)	mg/l	
• Distance to coast	m	Computed in GIS

4.2.5 Step 3: K-means partitioning

Although the result of a hierarchical cluster analysis on its own is prone to multiple errors, a hierarchical clustering, based on Ward's method, can generate excellent starting positions (i.e. cluster centroids used as cluster seeds) for a *K*-means partitioning (Milligan 1980; Legendre and Legendre 1998; and Wishart 1987). Partitioning clustering methods produce clusters in a predefined number of groups (K). *K*-means is the most widely used numerical method for partitioning data (examples from the marine environment are found in Legendre et al. (2002); Legendre (2003); Preston and Kirlin (2003); Hewitt et al. (2004); and Zharikov et al. (2005)). Pixels from clusters are allocated to a cluster in which the distance to its centre is minimal. The procedure stops if all pixels have been allocated. A *K*-means procedure exists of 3 steps: the initiation of the starting cluster centres, the allocation of pixels to the initial clusters and the re-allocation of pixels to another cluster. The starting positions and the allocation of the pixel to the initial clusters were taken from the hierarchical clustering, based on Ward's method. Those pixels are clustered that show the smallest increase in the Euclidean Sum of Squares.

4.2.6 Step 4: Number of clusters

The most difficult and most subjective decision in the cluster analysis is the number of clusters. Several indices to calculate the optimal number of clusters exist. From a simulation study comparing 30 indices, Milligan and Cooper (1985), proposed the C-H criterion as giving the best results. C-H is the *F*-statistic of multivariate analysis of variance and canonical analysis. *F* is the ratio of the mean square for the given partition, divided by the mean square for the residuals. The number of clusters corresponding with the highest C-H value is the optimal solution in the least-squares sense. C-H was also used as stopping criterion for cluster analysis in the marine environment in Legendre et al. (2002); Hewitt et al. (2004); and Orpin and Kostylev (2006).

4.2.7 Step 5: Validation of internal cluster consistency

A cluster analysis automatically allocates each individual to a cluster. To evaluate the internal consistency of the cluster composition, a validation with a split-run procedure was performed. For this procedure, the cluster analysis was first done for the whole dataset. After that, the optimal number of clusters was computed with the C-H criterion. Next, the dataset was randomly split into 2 equal validation parts to which the cluster analysis was applied with the same number of clusters. Finally, the cluster compositions from both validation parts were compared with the original cluster composition by calculating the number of differently classified pixels.

4.2.8 Step 6: Indicator species analysis of the clusters

To evaluate whether the obtained clusters have an ecological relevance, a species indicator analysis or INDVAL (Dufrêne and Legendre 1997) was performed. This method identifies indicator species for each of the clusters: if indicator species are identified then the cluster should have an ecological relevance, whereas if no indicator species can be identified, (most probably) the cluster has no ecological significance. The index is maximum when all individuals of a species are found in a single group of sites and when the species occurs in all sites of that group. The INDVAL index is defined as follows:

$\text{INDVAL}_{ij} = A_{ij} \times B_{ij} \times 100$

with A_{ij} = $\text{Nindividuals}_{ij}/\text{Nindividuals}_{i}$ or the mean abundance of species *i* in the sites of group *j*, compared to all groups in the study. A_{ij} is a measure of specificity and is maximum when species *i* is only present in cluster *j*.

B_{ij} = $\text{Nsites}_{ij}/\text{Nsites}_{j}$ or the relative frequency of occurrence of species *i* in the sites of group *j*. B_{ij} is a measure of fidelity and is maximum when species *i* is present in all sites of cluster *j*.

The index is maximal when all individuals of a species are found in a single group of sites and when the species occurs in all sites of that group. The statistical significance for the species indicator values is evaluated using a Monte Carlo permutation procedure. 1000 random permutations were used for this study.
Examples of applications of INDVAL to test the ecological relevance of predefined clusters can be found in Mouillot et al. (2002); Heino and Mykrä (2006); and Perrin et al. (2006).

4.3 Results

4.3.1 Step 1: PCA analysis

Retaining only those PCs with eigenvalues larger than 1; PCA resulted in 6 PCs, explaining 78.0% of the total variance. The rotated component matrix (Table 4.2) shows the factor loads, being the correlations between the rotated PCs and the original variables.
In decreasing order, PC 1 has high loads ($r < -0.5$ or $r > 0.5$) for the variables distance to coast, DTM, maximum TSM, d_s50, maximum Chl a, silt-clay % and gravel %; PC 2 for maximum bottom shear stress and maximum current velocity; PC 3 for slope and rugosity; PC 4 for BPI large scale and BPI small scale; PC 5 for eastness and northness; and PC 6 for sand % and gravel %. Gravel % is the only variable that has a high load for 2 PCs, meaning that this relationship is not exclusive.

4.3.2 Step 2: Hierarchical cluster analysis based on Ward's method

The 54307 cases with 6 PC variables were clustered to achieve a hierarchical partition tree. This tree is not at all appropriate as end result of the clustering, but the partitions are very useful as starting positions for the *K*-means partitioning.

4.3.3 Step 3: K-means partitioning

The cluster centres of the partition tree based on Ward's method, were used as input for the *K*-means partitioning. Subsequently, new cluster centres based on the *K*-means algorithm were computed forming a cascade from 2 to 20 clusters. Those centres were used to compute the C-H criterion.

Table 4.2: Component matrix showing correlations between the Varimax rotated PCs and the original variables.
High factor loads (r < -0.5 or r > 0.5) are indicated in bold. Information of the variables can be found in Table 4.1.

	Principal component					
	1	2	3	4	5	6
d_s50	**-0.894**	-0.094	0.088	0.052	0.064	0.133
silt-clay %	**0.668**	0.467	-0.135	-0.069	-0.079	-0.285
sand %	-0.230	-0.487	0.149	0.098	0.064	**0.748**
gravel %	**-0.514**	0.071	-0.069	-0.034	-0.016	**-0.665**
DTM	**0.932**	-0.075	0.028	0.212	-0.048	0.039
slope	-0.054	0.014	**0.958**	0.031	0.019	0.045
eastness	-0.021	0.041	-0.005	0.004	**0.828**	0.040
northness	0.105	-0.027	-0.050	0.000	**-0.798**	0.046
rugosity	-0.184	0.037	**0.909**	0.186	0.037	-0.042
BPI large scale	0.074	0.048	0.170	**0.862**	-0.001	0.048
BPI small scale	-0.116	0.003	0.023	**0.851**	0.005	-0.048
Max. bottom shear stress	-0.029	**0.918**	0.040	0.089	0.142	0.032
max current velocity	-0.184	**0.912**	0.055	-0.005	-0.029	0.021
max chl a	**0.794**	-0.133	-0.035	-0.079	-0.028	0.081
max TSM	**0.921**	-0.097	-0.151	-0.052	-0.034	-0.009
distance to coast	**-0.944**	0.091	0.074	0.068	0.043	-0.032

4.3.4 Step 4: Number of clusters and resulting clusters

Applying the C-H criterion (Figure 4.1), an optimum of 8 clusters was found. The result of the 8 cluster solution is presented in Figure 4.2 and Table 4.3. The clusters or marine landscapes represent well the natural environment and clear relationships with the original abiotic variables are visible.

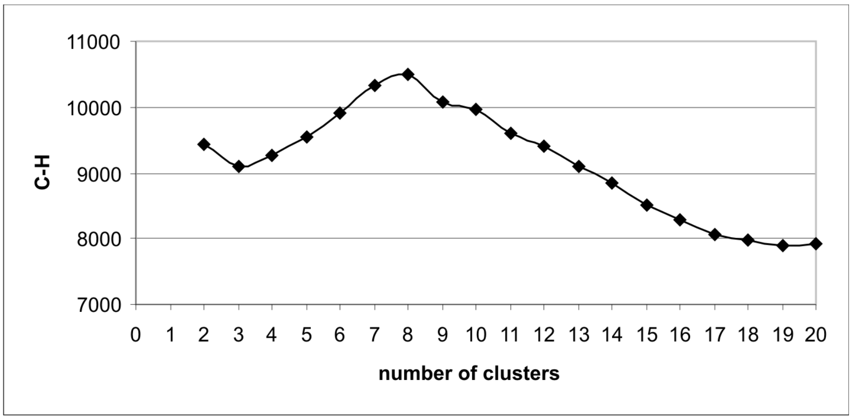


Figure 4.1: The number of clusters versus the C-H criterion. C-H reaches an optimum for 8 clusters.

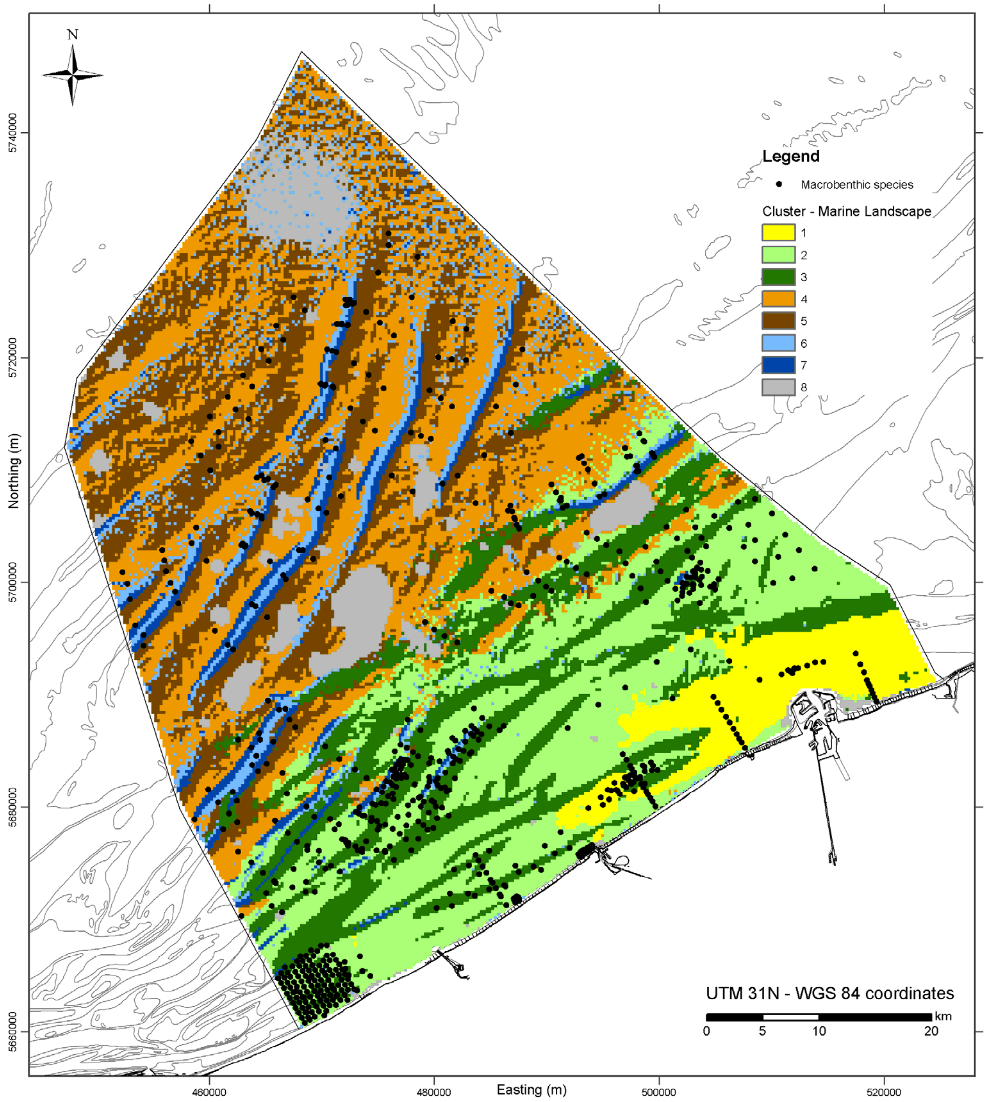


Figure 4.2: Belgian part of the North Sea with 8 clusters. The location of macrobenthic community samples are plotted for validation. Important patterns of the original abiotic variables are clearly visible on the map: e.g. high silt-clay % in cluster 1, alternation of sandbanks and flats-depressions in clusters 2, 3, 4, 5, 6 and 7; patches of gravel and shell fragments in cluster 8.

Boxplots (Figure 4.3) show the contribution of the original variables against the clusters. A clear example is the boxplot representing slope. This variable is approximately the same for all of the clusters, except for cluster 7 with higher values.

Table 4.3: The 8 clusters and their characteristics based on the boxplots (Figure 4.3).

Cluster	Characteristics
1	Shallow, high silt-clay percentage, high current velocity, high bottom shear stress, turbid, high Chl a concentration
2	Shallow NW orientated flats and depressions, fine sand, slightly turbid, high Chl a concentration
3	Shallow SE orientated sandbanks, fine to medium sand, slightly turbid, high Chl a concentration
4	Deep NW orientated flats and depressions, medium sand
5	Deep SE orientated flats and depressions, medium sand
6	Crests of sandbanks, medium sand
7	Slopes of sandbanks, medium sand
8	High percentage of gravel – shell fragments

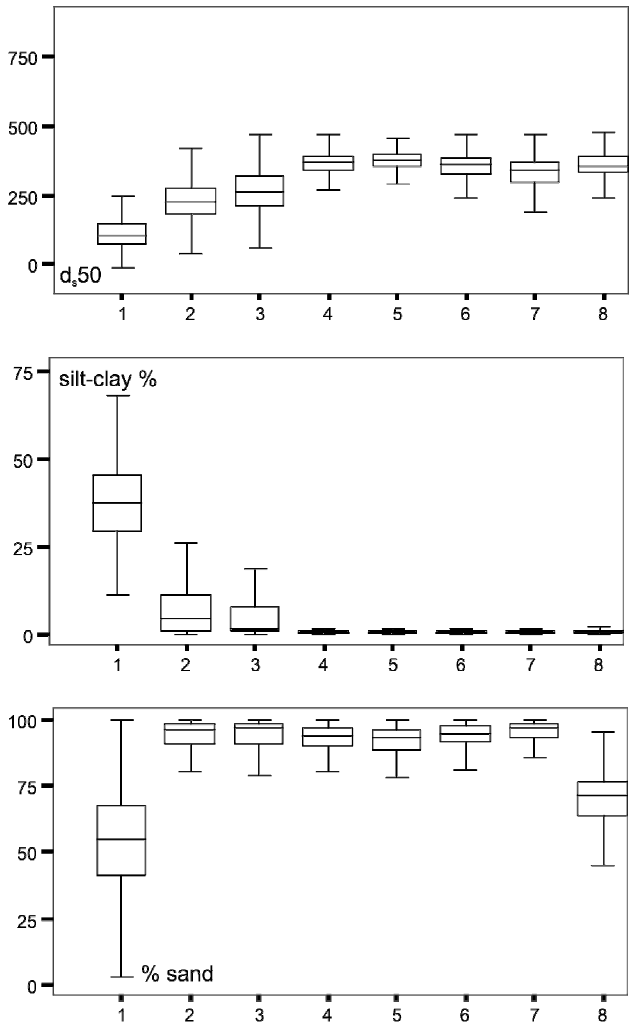
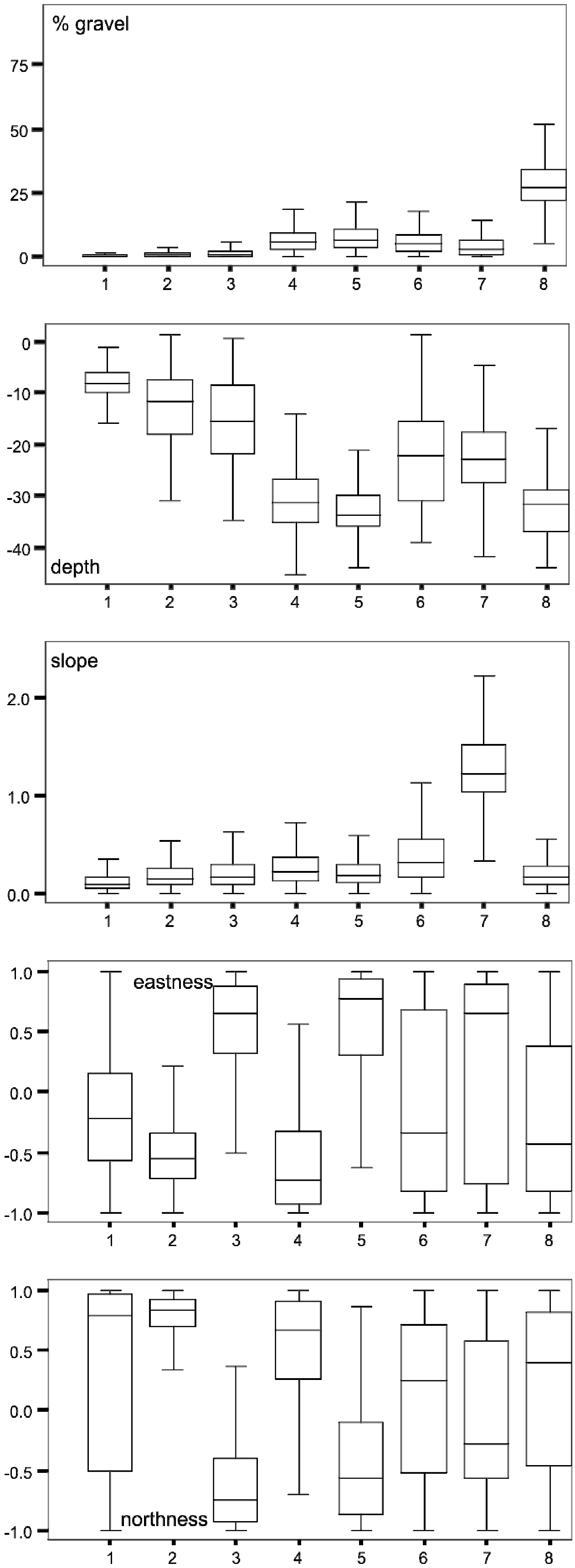


Figure 4.3a



% gravel
75
50
25
0
1
2
3
4
5
6
7
8
0
-10
-20
-30
-40
depth
slope
2.0
1.0
0.0
eastness
0.5
-0.5
-1.0
northness

Figure 4.3b

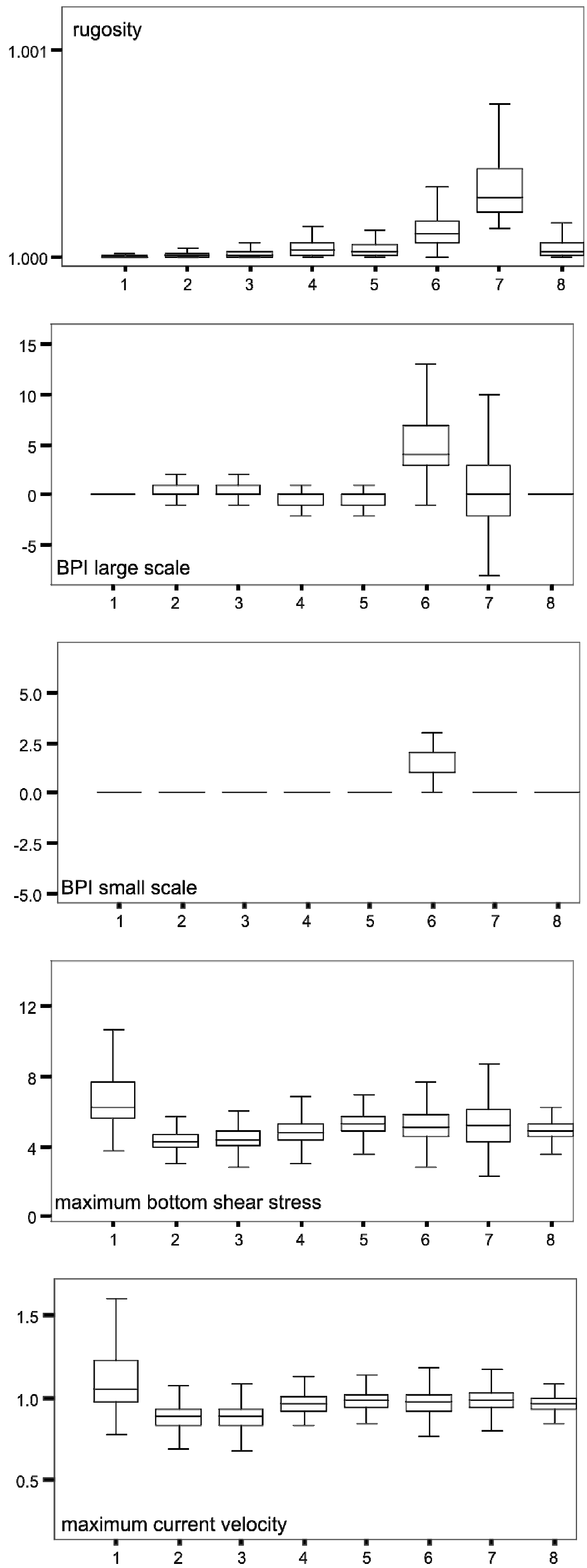

rugosity
1.001
1.000
1 2 3 4 5 6 7 8
15
10
5
0
-5
BPI large scale
1 2 3 4 5 6 7 8
5.0
2.5
0.0
-2.5
-5.0
BPI small scale
1 2 3 4 5 6 7 8
12
8
4
0
maximum bottom shear stress
1 2 3 4 5 6 7 8
1.5
1.0
0.5
maximum current velocity
1 2 3 4 5 6 7 8

Figure 4.3c

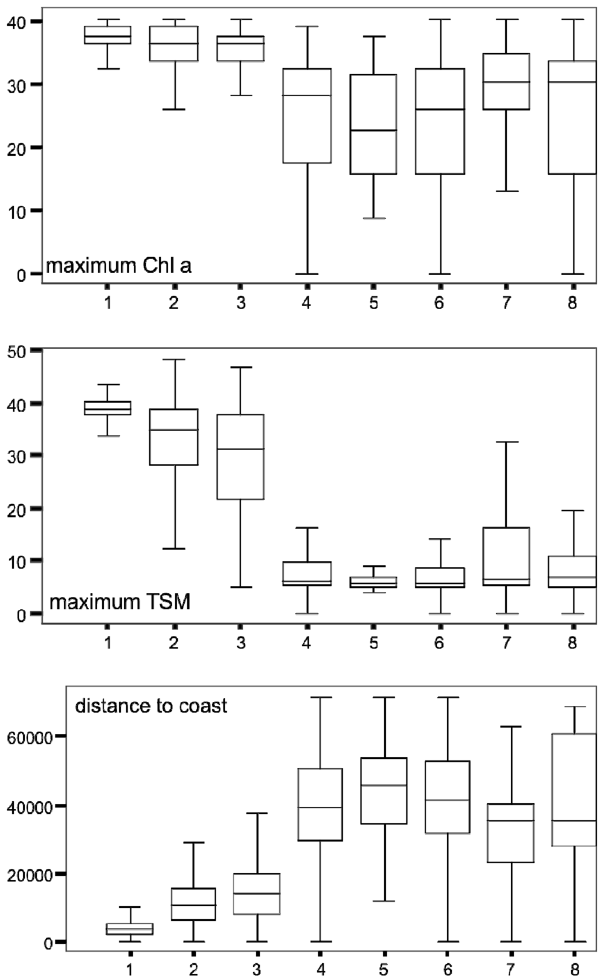


Figure 4.3d: Boxplots of clusters (X-axis) against abiotic variables (Y-axis). An overview of the abiotic variables and their units is given in Table 4.1. The middle line in the box is the median, the lower and upper box boundaries mark the first and third quartile. The whiskers are the vertical lines ending in horizontal lines at the largest and smallest observed values that are not statistical outliers (values more than 1.5 interquartile range).

4.3.5 Step 5: Validation of the internal cluster consistency

The split-run analysis showed very high correlations between the clusters obtained for the subsets and the clusters obtained for the whole dataset. Subset 1 contains 27153 cases, of which 159 have been classified differently as for the complete dataset. Subset 2 contains 27154 cases of which 184 have been classified differently. This is respectively 99.4 % and 99.3 % correspondence with the complete dataset for subset 1 and subset 2. The misclassified cases of both subsets were randomly distributed.
As shown by the split-run procedure, the internal cluster consistency is very good.

4.3.6 Step 6: Indicator species analysis of the clusters

Of the 123 species present in the 741 samples, randomization identified 25 species having a significant indicator value (5% level of significance) for 6 of the 8 defined clusters (Table 4.4). No indicator species could be found for cluster 3 and 6. This

means that both clusters do not show significant ecological differences with the other clusters.
Species with indicator values higher than 20 are *Cirratulidae* spp. and *Macoma balthica* for cluster 1; *Lanice conchilega* and *Spisula subtruncata* for cluster 2; *Echinocyamus pusillus* for cluster 4; *Tellina pygmaea*, *Gastrosaccus spinifer* and *Bathyporeia spp.* for cluster 5; and *Ophiura* spp. for cluster 7.

Table 4.4: Significant indicator species analysis of the defined clusters.

species	cluster	INDVAL (%)	Randomised INDVAL (%) Mean	SD	p*	A (%)	B (%)
Cirratulidae spp.	1	**51.4**	10.2	4.49	0.001	**73**	**70**
Macoma balthica	1	**26.8**	4.4	2.37	0.001	**67**	40
Glycera alba	1	14.5	5	2.55	0.011	48	30
Nephtys hombergii	2	19.7	8	2.6	0.005	41	48
Ensis spp.	2	19.4	7.3	3.35	0.015	**65**	30
Lanice conchilega	2	**21.3**	8.8	3.86	0.016	**62**	34
Phyllodoce mucosa / Phyllodoce maculata	2	17.7	7.7	3.6	0.026	**57**	31
Eumida spp.	2	12.3	5.5	2.87	0.035	**57**	22
Donax vittatus	2	9.3	4.4	2.43	0.042	45	21
Spisula subtruncata	2	**23.3**	14.1	4.98	0.046	**75**	31
Glycera capitata = Glycera lapidum	4	17.8	4.2	2.31	0.002	41	44
Echinocyamus pusillus	4	**22.8**	3.8	2.45	0.003	**59**	39
Branchiostoma lanceolatum	4	8.9	2.1	1.56	0.005	**73**	12
Pisione remota	4	8	2.2	1.82	0.017	**76**	11
Aonides oxycephala	4	10.1	3.4	2.61	0.03	**64**	16
Hesionura elongata	4	8.3	4	2.61	0.041	40	21
Thia scutellata	4	7.1	3.3	2	0.043	31	23
Tellina pygmaea	5	**31.7**	3.2	1.96	0.001	**60**	**53**
Gastrosaccus spinifer	5	**23.3**	7.1	2.77	0.002	35	**66**
Bathyporeia spp.	5	23	11.4	4.57	0.022	30	**76**
Pisidia longicornis	5	6.7	2.1	1.88	0.028	**51**	13
Ophiura spp.	7	25	11.6	5.13	0.019	**62**	40
Nephtys cirrosa	7	18.3	13.3	2.41	0.038	20	**90**
Aonides paucibranchiata	8	8.1	2.8	1.93	0.029	45	18
Bivalvia spp.	8	6.8	2.6	1.82	0.037	**75**	9

p* Statistically significant at the 0.05 level; SD = standard deviation; A = specificity; B = fidelity; INDVAL values higher than 20% are marked in bold; A and B values higher than 50% are marked in bold.

4.4 Discussion

This paper proposes an objective protocol to define ecologically relevant zones, solely on the basis of abiotic datasets. These zones are called ‘marine landscapes’, as they show a strong correlation with the abiotic variables and, in particular, the topography.

4.4.1 An objective method to define marine landscapes

The classical Marine Landscape methodology, as proposed by Roff and Taylor (2000); and Roff et al. (2003); and applied by Golding et al. (2004); Schelfaut (2005); Connor et al. (2006); and Al-Hamdani and Reker (2007) is highly subjective because of three reasons. First, the selection of ecologically relevant abiotic variables is biased. For the present protocol, no selection is necessary as input for a PCA; because PCs are constructed as linear combinations of the available, original abiotic variables (e.g. Cardillo 1999; and Fairbanks 2000). Secondly, there is a difficulty of classifying the selected abiotic variables into relevant classes. In this paper, a solution is proposed that abandons the classification and uses the continuous abiotic variables as input for the further analysis (e.g. Wilson 2007). Thirdly, the 'Queries' step is highly subjective because new combinations (the clusters or 'marine landscapes') are chosen arbitrarily from the predefined classes of the abiotic variables. This can be overcome by combining all possible classes, but this would lead rapidly to too many classes (e.g. 6 variables with 5 classes already means 30 landscapes). As such, this paper uses the C-H criterion to define a relevant number of clusters to automatically cluster the continuous abiotic variables (e.g. Legendre et al. 2002; Hewitt et al. 2004; and Orpin and Kostylev 2006).

With the objective approach proposed in this paper, there are still some decisions to be made during the analysis. First, for the cluster analysis, the number of groups has to be decided. Out of own physical knowledge of the BPNS, the solution of 8 marine landscapes seems to represent well the natural environment and none of the clusters seems to be useless. Their relation with the overall environment is clear, which was also exemplified by boxplots indicating the contribution of each abiotic variable to the clusters (Figure 4.3). The C-index (Hubert and Levin 1976), being a very good stopping criterion comparable to C-H (Milligan and Cooper 1985), has been tried as stopping criterion on this dataset, but it does not work for large datasets as used for this study. Secondly, for the *K*-means procedure, the Euclidean Sum-of-Squares clustering criterion was used as a distance index. As Punj and Stewart (1983) demonstrated, the choice of the (dis)similarity or distance index is of minor importance, compared to the clustering algorithm.

4.4.2 Abiotic datasets

Degraer et al. (2008) already discussed the many abiotic variables that might explain the distribution of macrobenthic communities on the BPNS. For the present study, not only typical variables, such as bathymetry and sedimentological information (e.g. Wu and Shin 1997; Van Hoey et al. 2004; and Willems et al. 2008) were used, but also hydrodynamical data (e.g. Caeiro et al. 2005), turbidity (e.g. Akoumianaki and Nicolaidou 2007), topographically derived features such as BPI (e.g. Lundblad et al. 2006; Wilson et al. 2007), eastness and northness (e.g. Hirzel et al. 2002a; Wilson et al. 2007) and rugosity (e.g. Jenness 2002; Lundblad et al. 2006; Wilson et al. 2007). Still, other abiotic variables could be used, such as curvature (e.g. Wilson et al. 2007), primary productivity (e.g. Smith et al. 2006), organic matter (e.g. Verneaux et al. 2004), salinity (e.g. Al-Hamdani and Reker 2007), temperature (e.g. Connor et al. 2006) and stratification (e.g. Connor et al. 2006). In addition, Guisan and Thuiller (2005); Baptist et al. (2006) and Wilson et al. (2007) stress the importance of spatial scales for predicting the distribution of fauna.

The more abiotic variables become available as input for habitat mapping, the more potential habitats can be classified and potentially new habitats could be identified. However, the relevance of additional classes may not always be clear. It remains important that the variables can be measured or obtained easily and that a sound evaluation of the end products is guaranteed. Another difficulty is the spatial and temporal bias of both biotic and abiotic datasets. Most of the ground-truth data are taken close to the coast and harbours and are strongly biased towards topographic locations. On the BPNS, most samples were taken on the sandbanks, because of their economic potential (e.g. aggregate extraction). Here, the samples are often closely spaced, while other locations are mostly under-sampled. In the most offshore areas, samples are commonly scarce. Apart from the spatial complexity, the samples are also subject to a temporal bias. Gregr and Bodtker (2007) stress the importance of the temporal dynamics (i.e. seasonal variations) for abiotic variables.

In a short time-span, extreme events, such as storms can cause completely different situations of e.g. current regime or suspended matter, causing differences in species composition. Therefore, for the present study, it was decided to work with maximal values of abiotic variables, as those datasets are best suited to represent extreme events (maximum bottom shear stress, maximum current velocity, maximum Chlorophyl a and maximum total suspended matter; cfr. Table 4.1).

On the BPNS, sedimentological samples have been taken from 1976 until now, whilst biological samples are all from a more recent date. Most of the datasets do not cover the same period. Some abiotic datasets are the result of a compilation over many years (e.g. map of d_s50 and silt-clay %), whereas others represent a very limited time span (e.g. maximum bottom shear stress; based on data from a spring-neap tidal cycle, 14.8 days). In an ideal situation, all abiotic and biotic datasets would cover the same spatial and temporal scale.

Misleading conclusions can be drawn because of the inappropriate use of some datasets. Sedimentological samples are very suitable to define the sand and to a lesser extent the silt-clay fraction. The gravel fraction (> 2 mm) might be underestimated when grab samples only have been obtained. Gravel can be detected with acoustical classification techniques, but only minor parts of the BPNS have been covered until known (Van Lancker et al. 2007). However, gravel is a part of very interesting habitats with generally high biodiversities (e.g. relation of gravel occurrence with *Ostrea edulis* and *Clupea harengus* (Houziaux et al. 2007a; and Houziaux et al. 2007b); with scallops (Kostylev et al. 2003); and with algae and *Crepidula fornicata* (Brown et al. 2002)).

Therefore, marine landscape mapping 'suggests' only possible ecologically interesting areas, and its predictive power remains dependent on the nature, quality and stability of the abiotic variables.

4.4.3 Ecological relevance

The BPNS is an ideal test case for the proposed methodology as both abiotic and biological datasets are widely available. Since the marine landscapes in the present study are rather limited in surface area, they might be considered as habitats and the results might be similar than those that would be obtained with habitat mapping. However, the difference between them is that he Marine Landscape approach is top-down and the habitat mapping approach is bottom-up. This means that for the top-down approach biotic data are used at the end of the process for the validation (or, in the case of no samples, not at all for some marine landscapes). For the bottom-up approach, abiotic and biotic data are used from the beginning of the process to create a habitat model, centering around the relationships between both (e.g. Willems et al. 2008). Still, for the top-down approach, abiotic data have to be selected that are at least assumed to have an ecological relevance. This knowledge may be derived from literature or expert judgement, but also from a visual inspection at the beginning of the process, comparing possible abiotic input layers with the number of biotic samples. In this paper, no prior selection of abiotic variables has to be done, as all of them are used as PCs.

The ecological validation for this study was based on an indicator species analysis, defining significant indicator species for the predefined clusters. The results showed that for each cluster, except for cluster 3 and 6, significant indicator species could be found.

As such, the clusters are a good proxy for biological predictions. Still, it must be clear that it is not the absolute aim of the marine landscape mapping to predict the biology as such, therefore other and better predictive modelling techniques exist (e.g. Guisan and Zimmermann 2000). Marine landscapes give an indication about the biology, derived solely from abiotic datasets, and offer a valuable in alternative in areas where biological data are scarce or absent.

There seems to be a discrepancy between the number of landscapes (8) and the number of clusters with significant indicator species (6): if a landscape is ecologically meaningful, then this landscape should be populated by specific biota or, in other words, every landscape should be uniquely linked to the biology. Although we might conclude from this discrepancy that several identified marine landscapes have no ecological meaning, we might also explain this by the potential lack of sufficiently detailed information on the marine biota, used to validate the marine landscapes. In conclusion, the level of detail of our current knowledge on the macrobenthos might be insufficient for an unbiased validation of the marine landscapes. If such detailed information would be available, then these data could help to further unravel the ecological meaning of all eight marine landscapes. At the same time, one will never be able to completely explain the occurrence of certain species and communities on the basis of the abiotic environment alone. A biological or an abiotic point of view will never result in exact the same abstractions of the marine seabed, because both approaches are a different way of looking at the same thing.

4.4.4 SWOT analysis

A critical evaluation of the protocol to map marine landscapes is performed using a SWOT analysis (strengths, weaknesses, opportunities and threats).

The main *strengths* of the protocol are the following:

- the possibility to use all available abiotic variables as input for PCA, a technique that eliminates all redundancy of correlating data;
- the unnecessity to classify the abiotic variables before the clustering and thus the possibility to use continuous abiotic variables as input for the clustering;
- the use of the C-H criterion to help defining the optimal number of clusters of marine landscapes.
- the proposed protocol is repeatable and objective; it forms a good alternative for the currently used methodologies which imply subjective decisions to be made.

The main *weaknesses* are:

- the added value of the defined clusters or marine landscapes map is dependent on the availability of relevant abiotic datasets;
- PCA, cluster analysis and INDVAL requires statistical insight and knowledge of the user.

A possible *opportunity* is:

- the application of this protocol for mapping marine landscapes on an international scale and for a larger area than the BPNS (e.g. as contribution to the European Atlas of the Seas in the context of the future European Marine Strategy Framework Directive).

A possible *threat* is:

- for a mapping exercise over a large area, abiotic and biotic datasets, based on different techniques and with different accuracies will be merged, causing an inpredictable error propagation.

Summarizing, the protocol creates interesting opportunities for a mapping exercise on a European scale, but important considerations have to be made about the accuracy of the final result, when datasets of different qualities and origins are used. Foster-Smith et al. (2007b) describe how the accuracy and confidence of marine habitat maps can be assessed, based on a multi-criteria approach.

4.5 <u>Conclusion</u>

This paper proposes an objective statistical method for the definition of ecologically relevant marine landscapes. The zones represent well the natural environment and there are clear relationships with the original abiotic variables and the occurrence of macrobenthic species. The methodology is straightforward and allows an easy application to other areas. Marine spatial planning, environmental protection and management of marine zones can benefit from the definition of ecologically relevant marine landscapes (e.g. definition of most important ecological zones, to be protected from dredging and dumping activities or aggregate extraction).

4.6 <u>Acknowledgements</u>

This paper is part of the PhD research of the first author, financed by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). It also contributes to the EU INTERREGIII B project MESH (''Development of a framework for Mapping European Seabed Habitats''. In addition, it frames into the research objectives of the projects MAREBASSE (''Management, Research and Budgeting of Aggregates in Shelf Seas related to End-users'', EV/02/18A) and BWZee (''A Biological Valuation Map for the Belgian Continental Shelf'', EV/88/37B), financed by the Belgian Science Policy. For the bathymetric data, the Flemish Authorities, Agency for Maritime and Coastal Services, Flemish Hydrography are acknowledged and the hydrodynamical data are provided by the Management Unit of the North Sea Mathematical Models and the Scheldt estuary (MUMM). The MERIS data on Chlorophyll a and Total Suspended Matter are processed by MUMM in the framework of the BELCOLOUR-2 project (ESA ENVISAT AOID3443). The sedisurf@ database has been compiled by the Renard Centre of Marine Geology and consists of data from various institutes. The Macrodat

database (Marine Biology Section, Ugent – Belgium, 2008) has been compiled by the Marine Biology Section.
This paper contributes also to the Ghent University BBSea project (n° GOA 01600705).

Chapter 5

Habitat suitability as a mapping tool for macrobenthic communities: An example from the Belgian part of the North Sea

PUBLISHED AS:

Degraer, S., Verfaillie, E., Willems, W., Adriaens, E., Van Lancker, V. and Vincx, M., 2008. Habitat suitability as a mapping tool for macrobenthic communities: An example from the Belgian part of the North Sea. Continental Shelf Research 28(3), 369-379.

5 Habitat suitability as a mapping tool for macrobenthic communities: An example from the Belgian part of the North Sea

---

Abstract

Being ecologically important and well-known, the spatial distribution pattern of the macrobenthos is often used to support an ecologically sustainable marine management. Though in many cases the macrobenthic spatial distribution is relatively well-known, this information is merely restricted to point observations at the sampling stations: although being increasingly demanded, full coverage spatial distribution maps are generally lacking. This study therefore aimed at demonstrating the usefulness of habitat suitability modelling as a full coverage mapping tool with high relevance for marine management through (1) the construction of a habitat suitability model for the soft sediment macrobenthic communities in the Belgian part of the North Sea (BPNS) and (2) predicting the full coverage spatial distribution of macrobenthic communities within the BPNS. The BPNS was selected as a case study area because of the high data availability on both macrobenthos and environmental characteristics. Discriminant function analysis (DFA) objectively selected median grain size and sediment mud content and omitted bathymetry, slope and distance to the coast to represent the most important environmental variables determining the macrobenthic community distribution. The consequent crossvalidated, empirical habitat suitability model, using both median grain size and mud content, showed an a posteriori average correctly classified instances (CCI) of 79% (community-dependent CCI ranging from 72% to 86%) and a Cohen's kappa of 0.71, pointing towards a very good agreement between model predictions and observations. The application of the habitat suitability model on the full coverage maps of median grain size and sediment mud content, taken from literature, allowed to reliably assess the distribution of the macrobenthic communities within 96.3% of the 53,297 BPNS grid cells with a resolution of 250 m. Next to its applicability to the BPNS, the model is further anticipated to potentially perform well in the full Southern Bight of the North Sea: testing is advised here. Since the habitat suitability is considered far more stable through time compared to the permanently fluctuating macrobenthic communities, information on the habitat suitability of an area is considered highly important for a scientifically sound marine management.

Keywords: Benthos; aquatic communities; habitat selection; mathematical modelling; habitat suitability; discriminant function analysis

---

5.1 Introduction

Due to its ecological importance and obvious presence within the marine ecosystem, the macrobenthos is one of the most intensively investigated marine ecosystem components. Data on the spatial distribution of macrobenthic species and species assemblages are available for many areas worldwide (e.g. North Sea: Rees et al. 2002). Being ecologically important and well-known, the spatial distribution patterns of the macrobenthos is often used to support an ecologically sustainable marine management (e.g. Borja et al. 2000).

Though in many cases the macrobenthic spatial distribution is relatively well-known, this information is merely restricted to point observations at the sampling stations: although being increasingly demanded, full coverage spatial distribution maps are generally lacking (Young 2007). In general, two strategies could be followed to attain full coverage distribution maps: (1) spatial interpolation based on sampling point information (e.g. Dutch part of the North Sea: Holtmann et al. 1996) or (2) the development of habitat suitability models that predict the presence of macrobenthos based on the suitability of the physical habitat. Though being attractive, spatial interpolation is perilous since often community structure might change over very short distances. Another drawback of spatial interpolation is that the resulting map is highly dependent on the density of the samples. Degraer et al. (2002) demonstrated that – for instance in the geomorphologically highly diverse Belgian coastal zone – even a dense grid of sampling stations (120 sampling stations in 5x5 km area) did not allow to spatially extrapolate the macrobenthic community distribution patterns. Spatial interpolation further has the disadvantage that a rather static map is produced: whenever new data become available, the whole interpolation exercise has to be repeated. Predictive habitat suitability modelling, on the other hand, allows to objectively produce distribution maps at a level of detail limited only by the availability and resolution of environmental data. Being generally less costly to gather, compared to the collection of the labour-intensive biological information, environmental data is detailedly available in many areas. In such areas, small-scale patchiness within the macrobenthos will be detected as such. Once the predictive model is developed, this strategy further allows to easily update the spatial distribution whenever more detailed abiotic habitat data become available. If full coverage maps of the environmental variables (e.g. physical habitat) are available, it is even possible to create a full coverage map of the macrobenthos' spatial distribution.

This study aims at demonstrating the usefulness of habitat suitability modelling as a mapping tool with high relevance for marine management. This exercise will be performed using data from the well-investigated Belgian Part of the North Sea (BPNS) and dealt with in two steps: (1) the construction of a habitat suitability model for the macrobenthic communities in the BPNS (i.e. modelling) and (2) an extension of the knowledge of the spatial distribution of macrobenthic communities on the BPNS to the level of full coverage community distribution maps.

5.2 Material and methods

5.2.1 The Belgian part of the North Sea: current knowledge

The BPNS has a surface area of only 3600 km², but comprises a wide variety of soft sediment habitats (Verfaillie et al. 2006). Due to the presence of several series of sandbanks, the area is characterized by a highly variable and complex topography. Consequently, sediment types are highly variable throughout the area. Since the spatial distribution of the macrobenthos is largely dependent on the physical environment, a high macrobenthic diversity can be expected (Degraer et al. 1999a).

Because of the limited spatial extent of the BPNS in combination with the large interest in marine research, detailed knowledge on the macrobenthos' spatial distribution became available through several Flemish and Belgian research projects. Based on a combination of these datasets, Degraer et al. (2003) and Van Hoey et al. (2004) summarized the soft sediment macrobenthic community structure. They discerned between four subtidal communities: (1) the *Macoma balthica* community, (2) the *Abra alba – Mysella bidentata* community (or *A. alba* community; Van Hoey et al. 2005), (3) the *Nephtys cirrosa* community and (4) the *Ophelia limacina – Glycera lapidum* community (further abbreviated as *O. limacina* community). Next to these communities, several transitional species assemblages, connecting the four communities, were defined.

Because of its high macrobenthic diversity, in combination with a detailed knowledge of the macrobenthic community structure, the BPNS represents an ideal case study area for the development of a predictive model to attain a (full coverage) spatial distribution map of the macrobenthos.

5.2.2 Research strategy

Two major steps can be distinguished within the research strategy: (1) habitat suitability modelling and (2) full coverage mapping of the macrobenthic habitat suitability (Figure 5.1). The first step comprised to model the link between the biological point data and the accompanying physical data, aiming at creating a solid mathematical habitat suitability model. In the second step the habitat suitability model was applied to the full coverage maps of the ecologically most relevant physical variables in order to attain a full coverage habitat suitability map.

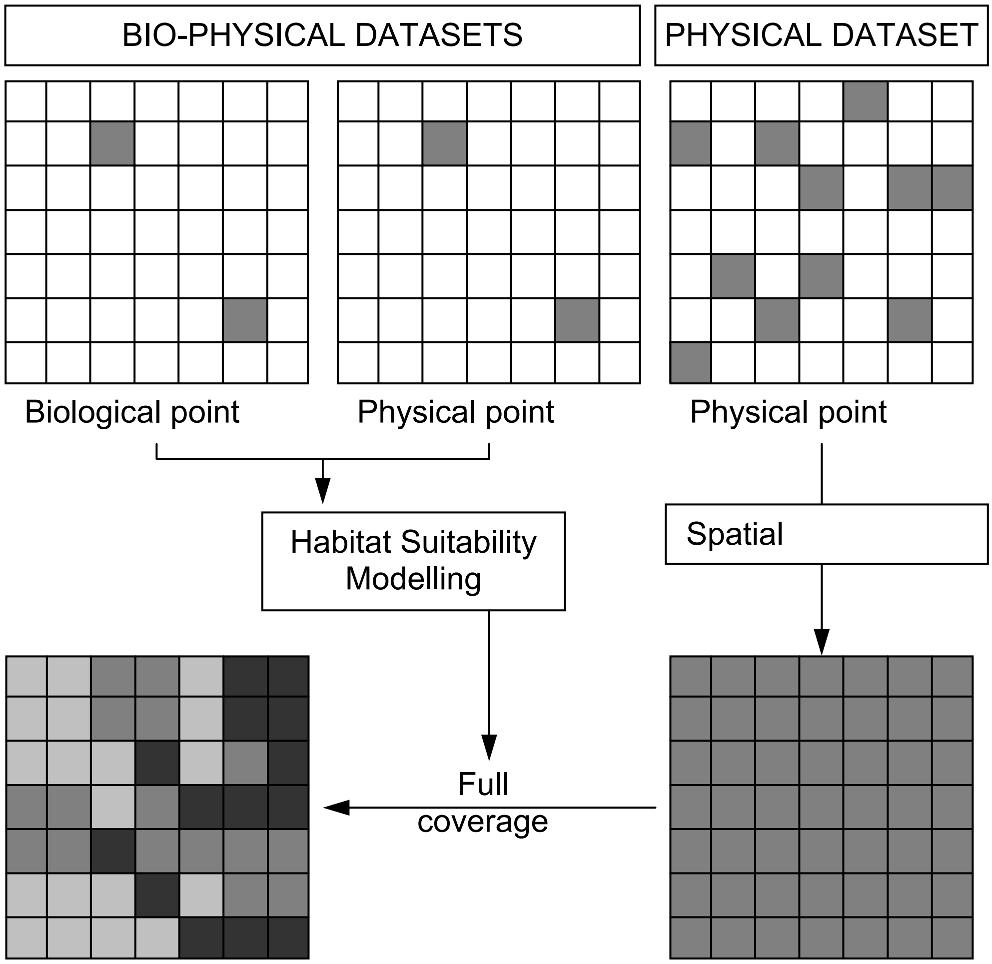


Figure 5.1: Schematic presentation of the research strategy, starting from bio-physical and physical point data to a full coverage macrobenthic habitat suitability map.

5.2.3 Data availability

Biological data

Within the framework of several projects 1197 macrobenthos samples were collected at the BPNS between 1994 and 2004. The samples were all collected with a Van Veen grab (sampling surface area: 0.1 m²) and sieved over a 1 mm mesh-sized sieve. All organisms were identified to species level, whenever possible, and species-specific densities (ind./m²) were determined.
Before analysis, a thorough data quality control was performed. Non-representatively sampled species were excluded from the dataset. A first set of non-representatively sampled species consisted of non-macrobenthic species, such as hyperbenthic mysids, fish and pelagic larvae, which cannot representatively be sampled with a Van Veen grab. A second set consisted of rare species, here defined as any species with a frequency of occurrence of less than 2% and encountered with a maximum of three individuals per sample. Because datasets, derived from different research projects, were combined, the dataset was further checked for inconsistent species

identifications. In case of inconsistent species identifications (e.g. *Bathyporeia* spp., *Capitella* spp. and *Ensis* spp.), the species were lumped to the taxonomically highest common denominator. To avoid temporal autocorrelation, temporal series were excluded from the analysis. Given (1) a distance of at least 350 m between any pairs of sampling stations and (2) the high spatial heterogeneity in macrobenthos (Degraer et al. 2002), spatial autocorrelation was considered negligible in our data set. After data quality control the final dataset comprised 773 samples and 123 species.

Environmental data

Habitat suitability model input data

To maximise the applicability of the habitat suitability model, only frequently measured and/or widely available environmental variables were offered in the modelling exercise. A first set of environmental data was composed of variables measured *in situ*, i.e. median grain size, sediment mud content and depth. Slope was calculated on the basis of detailed bathymetric maps. Finally, distance to the coast, calculated from the geographic position of the sampling points, was included in the list of potentially explanatory variables.

Full coverage maps

The bathymetric map of the BPNS is based on single beam echosounder data from the Maritime Services and Coast Agency, Flemish Hydrography and completed with data from the Hydrographic Office of the Netherlands and the United Kingdom. This dataset was interpolated using a simple inverse distance algorithm to a digital terrain model with a resolution of 80 m. The slope map is the first derivative of the bathymetric map. It is expressed in degrees and has a resolution of 80 m. Full coverage median grain size and mud content maps with a resolution of 250 m were derived from the 'sedisurf@' database (UGent-RCMG), containing more than 6000 data points, spread throughout the BPNS and collected since 1976. At first, the database was cleaned using a 'zonation approach' and extreme or unrealistic data points were removed. To create full coverage median grain size maps, Kriging with an external drift was used, taking into account bathymetry as a secondary variable to assist in the interpolation (for more detailed information: Verfaillie et al. 2006). The map of the mud content was created, using Ordinary Kriging with directional variograms for the anisotropy of the data (for more detailed information: Van Lancker et al. 2007).

5.2.4 Habitat suitability modelling

Modelling strategy

Since the relevance for marine management is a major aim of this paper, the outcome of the modelling and mapping exercise should be easy to communicate to politicians, policy-makers and managers (Olsson and Andersson 2007). Hence, although we acknowledge macrobenthos to be structured along gradients, for sake of an easy communication an abstraction of this complexity was set (Turney 1995): instead of modelling the detailed macrobenthic gradients, we deliberately focused our model on the prediction of the chance of occurrence of each of the four macrobenthic communities, given a set of environmental factors. As such, the macrobenthos was modelled and mapped at the community level (i.e. clearly delineated entities), a level of detail allowing an easy communication and interpretation of the final outcome within a management perspective. To assure the incorporation of only well-delimited macrobenthic communities into the model (i.e. distinct sample groups from the multivariate analyses), transitional species assemblages were hence excluded from the predictive modelling exercise. To allow an easy communication of the model outcome, continuous variables are often converted into discrete variables (Turney 1995). The biological variation for certain endpoints may be too large to make reasonable predictions, therefore the modeller may decide to convert the data into two or more categories (Worth and Cronin 2003).

Biological data exploration: Community analysis

The community structure was investigated by several multivariate techniques: Group-averaged cluster analysis based on Bray-Curtis similarity (Clifford and Stephenson 1975), Detrended Correspondence Analyses (DCA) (Hill and Gauch 1980) and Two-Way Indicator Species Analysis (TWINSPAN) (Hill 1979; Gauch and Whittaker 1981), based on the final dataset with 773 samples and 123 taxa. For cluster analysis and DCA the data were fourth-root transformed prior to analysis. TWINSPAN was run using both the species density data as well as the presence/absence data.
The outcome of each multivariate analysis was compared to extract consistent groups of samples. Samples that were placed in different sample groups by the different multivariate analyses were considered as inconsistently grouped and were excluded from further analysis. This strategy assured that atypical observations (i.e. inconsistently grouped samples) did not bias any further analysis.
To designate the newly defined multivariate sample groups to the macrobenthic communities identified in previous research in the BPNS (Van Hoey et al., 2004) (i.e. *A. alba*, *N. cirrosa* and *O. limacina* communities), the relative distribution (%) of the samples over the macrobenthic communities was calculated per sample group. Because samples, belonging to the *M. balthica* community, were not present in the database, used by Van Hoey et al. (2004), sample group designation to the latter community was based on Degraer et al. (2003). Each sample group was designated to the community or transitional species assemblage (TSA) with the highest relative distribution value. For a detailed description (biology and environment) of all communities and TSAs one is referred to Degraer et al. (2003) (*M. balthica* community) and Van Hoey et al. (2004) (*A. alba*, *N. cirrosa* and *O. limacina* communities).

Discriminant Function Analysis

Discriminant function analysis (DFA) was used (1) to objectively select abiotic habitat variables that allow to discriminate between the four macrobenthic communities and (2) to develop a habitat suitability model. Finally, the habitat suitability model was applied to the full coverage environmental maps, generating full coverage distribution maps for the macrobenthic communities.

The forward procedure was used to detect the best set of abiotic habitat variables. The Wilk's Lambda statistics was applied to test the significance of the discriminant functions. The standardized coefficients for the discriminant functions allow to determine the contribution of each abiotic habitat variable to the separation of the macrobenthic communities: the larger the standardized coefficient, the greater is the contribution of the respective variable to the discrimination between groups. DFA assumes low multicollinearity of the independents and the same within-group variance covariance matrix for all groups. Variables are redundant when the pooled within-groups absolute correlation is equal or higher then 0.75, when this is the case one of the correlated variables is excluded from the analysis. The homogeneity of the variance covariance matrix was assessed by the log determinants.
To test the predictive performance of this approach on test data, not used to construct the model, a three-fold crossvalidation was applied. First, the data was split up in three parts. Care has been taken to assure that the proportion of each community in the three parts resembled the proportion in the whole data set (Witten and Frank 2000). Then two parts of the data set were used as a training set to develop a DFA habitat suitability model. This model was then applied to the third part of the data set. The predictions for the third part, not used to develop the model, were compared with the actual observations. This procedure was iteratively repeated, each part of the data set being used to train or test the model. If the performance of the three models is good and consistent we can conclude that the modelling approach is appropriate: a final model could then be constructed using all data points.
Two model performance indices were calculated: the % Correctly Classified Instances (CCI) and the Cohen's kappa. The formulae and a discussion on these model performance indices is given by Fielding and Bell (1997). Cohen's kappa is compensated for the prevalence of the entity to predict. It takes into account the chance that a sample would be attributed to a community by chance. No weighting was used in the calculation of the Cohen's kappa.

5.2.5 Habitat suitability mapping

The habitat suitability model was used to calculate the classification probabilities (i.e. community-specific habitat suitability) of each grid cell within the full coverage maps of each of the selected explanatory environmental variables (see Data availability: Full coverage maps). This habitat suitability measure was based on the grid cell's Mahalanobis distance from the different community centroids. The Mahalanobis distance (measure of distance between two points in the space defined by two or more correlated variables) is the distance between each sample and the macrobenthic community centroid in the multivariate space defined by the variables in the model. In general, the further away a grid cell is from a community centroid, the less likely it is that the habitat of the grid cell is suitable for that community. As such, a habitat suitability map (0 to 100%) for each macrobenthic community was derived. However, not all grid cells allowed a reliable habitat suitability estimate: grid cells with a Mahalanobis distance of three times the standard deviation from any macrobenthic community centroid (as calculated from the Mahalanobis distances from the model input data) were considered outliers and excluded from the map. Hence, we ascertained that no predictions were made beyond the range of the data set, used to develop the model. Using the model beyond this range could potentially lead to artefacts.

5.3 Results

5.3.1 Community analysis

Based on DCA, Cluster Analysis and TWINSPAN, 690 samples were consistently assigned to eight sample groups: 83 samples (11%) were inconsistently grouped and were excluded from further analysis. All groups consisted of 23 (sample group B) to 228 samples (sample group F), except for sample group H, which consisted of no more than five samples. Group H was therefore excluded from further analyses.
An uneven relative distribution of the samples of each sample group over the formerly defined macrobenthic communities and transitional species assemblages in the BPNS was found (Table 5.1). Because the major part of the group C samples (83%) corresponded with the *A. alba* community, defined by Van Hoey et al. (2004), group C was here defined as the *A. alba* community. Likely, groups A (max. 58%), E (max. 47%) and G (100%) were defined as the *M. balthica*, *N. cirrosa* and the *O. limacina* community, respectively. The major part of groups D and F samples (96% and 69%, respectively) were part of TSAs, each representing a link between two "parent communities". Sample group B could not be assigned to any community or TSA.

Table 5.1: Relative distribution (%) of the samples of each multivariately defined sample group over the formerly defined macrobenthic communities ([1] Van Hoey et al. 2004; [2] Degraer et al. 2003). TSA 1, transitional species assemblage (TSA) between *A. alba* and *N. cirrosa* communities; TSA 2, TSA between *N. cirrosa* and *O. limacina* communities; TSA 3, TSA between *N. cirrosa* and intertidal communities.

	Multivariately defined sample groups						
Formerly defined communities	A	B	C	D	E	F	G
Abra alba community (1)			**83**				
← TSA 1 → (1)			14	**96**	21	2	
Nephtys cirrosa community (1)					**47**	2	
← TSA 2 → (1)			2	4	25	**69**	
← TSA 3 → (1)					7	3	
Ophelia limacina community (1)			1			24	**100**
Macoma balthica community (2)	**58**	4	1	5			

5.3.2 Community habitat preferences

Clear differences in habitat preferences were found for all macrobenthic communities and for all environmental variables taken into account in this study (Figure 5.2). From the *M. balthica* community to the *O. limacina* community a preference for increasing median grain size was detected. Although less consistent, a similar positive relationship was found for depth, distance to the coast and slope. An opposite trend was detected considering sediment the mud content.

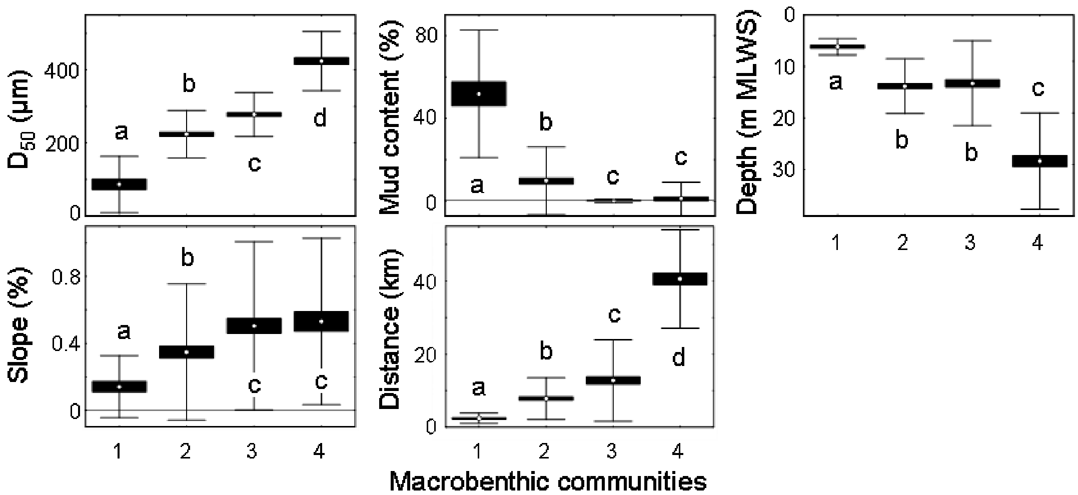


Figure 5.2: Habitat preferences of all macrobenthic communities: 1, *Macoma balthica* community; 2, *Abra alba* community; 3, *Nephtys cirrosa* community; 4, *Ophelia limacina* community. Mean, box: standard error, whiskers: standard deviation. Different letters (a, b, c, d) point to significant differences (post hoc LSD test: p < 0.05).

5.3.3 Community habitat suitability modelling

At first several combinations of environmental variables were used to develop preliminar habitat suitability models, in order to search for the optimal combination of predicting environmental variables. Prior to the analysis, the mud content, depth, distance to the coast and the slope were log transformed to obtain a homogeneous variance covariance matrix.

The slope was never selected in the preliminary models by the discriminant function analysis and was thus automatically rejected from further modelling exercises. The correlation matrix showed that the log-transformed depth and distance to the coast were correlated ($r = 0.75$). Because the standardized discriminant function coefficient of the depth (-0.167) was smaller in comparison to the distance to the coast (0.329), it was decided to exclude depth from the modeling exercise. As a result only three environmental variables were used in the preliminar models: median grain size, sediment mud content and distance to the coast. Only the first two discriminant functions (DF) were significant (DF1 Wilk's $\lambda = 0.180$, $\chi^2 = 609.5$, df = 9, $p < 0.001$ and DF2 Wilk's $\lambda = 0.593$, $\chi^2 = 185.8$, df = 4, $p < 0.001$) and explained 77.0% and 22.8% of the variance. Since the median grain size and the mud content were the most important explanatory variables for these functions (highest standardized discriminant function coefficients) only these variables were included in the final model.

Cross-validation

The performance of the habitat suitability model was tested by means of a threefold cross-validation procedure (Table 5.2). The agreement between model predictions and observations was very good and consistent between the three cross-validation model runs (e.g. Cohen's kappa: 0.70 – 0.73). This demonstrated that the modelling approach is suitable and a final model could be developed using all available samples.

Table 5.2: Model performance for a threefold cross-validation.
The data were stratified in such a way that the prevalence of a community in each fold, is proportional to the prevalence in the complete data set. CCI, % Correctly Classified Instances.

	Model run		
	1	2	3
CCI (all samples)	80.2	78.3	82.3
CCI (validation)	79.8	80.7	79.0
Cohen's kappa	0.71	0.73	0.70

Final model

Two DFs were proposed. The first DF, explaining 76.6 % of the variance, was mainly determined by the median grain size (Wilk's λ = 0.37, $p < 0.01$, standardized coefficient = -0.62 versus 0.55 for mud content). Mud content was slightly more relevant then the median grain size within the second DF (Wilk's λ = 0.36, $p < 0.01$, standardized coefficient = -1.00 versus -0.95 for median grain size), accounting for 23.4 % of the variance.
Four classification functions (i.e. one per macrobenthic community) were derived (Table 5.3).

Table 5.3: Community specific weights of all variables taken into the classification functions.
Cases are classified to the community rendering the highest score, by applying $S_i = w_{i(\text{Median grain size})}$*(Median grain size) + $w_{i(\text{Mud content})}$*($\log_{10}$(Mud content+1)) + Constant, with i = community i.

	Macoma balthica community	*Abra alba* community	*Nephtys cirrosa* community	*Ophelia limacina* community
Median grain size	0.063	0.082	0.079	0.121
$\log_{10}$(Mud content + 1)	17.685	13.421	7.541	11.457
Constant	-17.637	-15.716	-12.541	-27.323

The performance of the final model constructed with all samples, was assessed for the whole data set. Overall, 79% of the samples were assigned to the correct community. Uncorrectly classified samples were generally assigned to a neighbouring community (*M. balthica* community ↔ *A. alba* community ↔ *N. cirrosa* community ↔ *O. limacina* community) (Table 5.4). The CCI per community was between a minimum of 72% (*A. alba* community) and a maximum of 86% (*O. limacina* community), but was not related to the prevalence of each community in the original data set. The latter observation, combined with a Cohen's kappa of 0.71, indicated a very good agreement between observed and modelled macrobenthic communities (Monserud and Leemans 1992).

Table 5.4: *A posteriori* accuracy and sample classification, rows: observed classifications and columns: predicted classifications. CCI, % Correctly Classified Instances.

	Community prevalence	CCI	*M. balthica* community	*A. alba* community	*N. cirrosa* community	*O. limacina* community
M. balthica community	7.8%	82%	23	3	2	0
A. alba community	36.9%	72%	10	97	24	4
N. cirrosa community	35.3%	83%	0	5	107	17
O. limacina community	20.0%	86%	1	0	9	62
Total		79%	34	105	142	83

5.3.4 Habitat suitability maps

The habitat suitability could reliably be assessed for 53297 grid cells (resolution: 250 m; i.e. 96.3% of the BPNS): the prediction for the remaining 3.7% was considered beyond the range of the model development data (i.e. Mahalanobis distance > 3 SD from any macrobenthic community centroid, see Materials and Methods), which consequently does not allow a reliable prediction.
The habitat suitability for the four macrobenthic communities is clearly zoned throughout the BPNS (Figure 5.3). At first, a clear onshore-offshore gradient in habitat suitability can be discerned. The offshore benthic habitats are suited mainly for the *O. limacina* community (maximum modelled suitability: 99.9%), while the *A. alba* community is expected to dominate the onshore area (maximum modelled suitability: 88.8%). The habitat of the *N. cirrosa* community is taking an intermediate position (maximum modelled suitability: 92.1%). A second longshore gradient can further be found in the onshore zone. In the western part of the onshore zone a clear dominance of the habitat of the *A. alba* community is found, whereas this community is expected to co-dominate the eastern part, together with the *M. balthica* community (maximum modelled suitability: 98.9%).

5.4 Discussion

5.4.1 Habitat suitability model

From a conceptual viewpoint three different types of models exist: (1) theoretically based analytical models (cf. simplified reality), (2) process-based mechanistic models (cf. cause-effect relationships) and (3) empirical models (Levins 1966). The main purpose of the latter type being to accurately condense empirical facts, its mathematical formulation is not expected to describe realistic "cause-effect" between model input variables and predicted responses, nor to inform about underlying ecological functions and mechanisms. Because our aim was to model and predict as precisely as possible the habitat suitability our model should thus be regarded as empirical (Guisan and Zimmermann 2000).

Considering the statistical approach we selected DFA. DFA is considered a valid modelling technique, since in our case the selected response variable is a categorical entity (i.e. macrobenthic community), (Guisan and Zimmermann 2000). Such approach has already been widely applied for habitat suitability modelling of e.g. marine benthic communities (Shin 1982; Vanaverbeke et al. 2002; Caeiro et al. 2005), seagrasses (Fourqurean et al. 2003), alpine marmots (Borgo 2003), nesting griffon vultures (Xirouchakis and Mylonas 2005), freshwater benthic diatoms (Pan et al. 1999), freshwater fish (Nate et al. 2003) and black terns (Naugle et al. *2*000).

Out of a suit of five environmental variables (bathymetry, slope, median grain size, sediment mud content and distance to the coast), the forward selection procedure of the discriminant function analysis indicated median grain size and mud content to be the most important environmental variables determining the distribution of the macrobenthos. The structuring importance of both variables has already been indicated by many other studies (e.g. Wu and Shin 1997; Van Hoey et al. 2004, Willems et al. 2008). Assigning all variability in macrobenthic distribution patterns to solely the latter variables would however be an oversimplification of reality. Many other environmental variables might also contribute, as demonstrated by many other studies (e.g. hydrodynamics: Caeiro et al. 2005; turbidity: Akoumianaki and Nicolaidou 2007; primary productivity: Smith et al. 2006; organic matter: Verneaux et al. 2004). Next to the potential direct influence of median grain size and mud content on the macrobenthic distribution (e.g. burrowing capacity, de la Huz et al. 2002), both variables can however also be considered as a proxy for at least some of these other potentially structuring variables, more difficult to measure (e.g. hydrodynamics and food supply to the bottom, Herman et al. 1999). Considering median grain size and mud content as indirect gradients (*sensu* Austin et al. 1984), our model should be regarded as an empirical model (Guisan and Zimmerman 2000), and caution is thus needed when applying the model outside of the geographical range of the original model construction data. The main advantage of the selection of median grain size and mud content however is the fact that both variables were measured systematically and available at full coverage not only within our case study area, but also in many other marine areas.

Because the modelling approach generated a high and consistent predictive performance, it was considered sound to develop a final model with all data. The Cohen's kappa for the three folds of the crossvalidation was 0.70 to 0.73, which indicates a very good agreement between model and observations (Monserud and Leemans 1992). A high kappa indicates that the model is also making correct predictions for the rarer communities (e.g. *Macoma balthica* community).

The final predictive model constructed with all data showed an average CCI of 79% when applied to all samples. For each community separate, the CCI varied between 72 and 86%. The *a posteriori* (i.e. no independency of construction and test data) Cohen's kappa for the final model of 0.71 was found to be very similar to the *a priori* kappa of the models developed during the crossvalidation process (maximum Cohen's kappa: 0.70 – 0.73), indicating the good estimate of Cohen's kappa. The final model agreement should be considered very good, following Monserud and Leemans 1992).

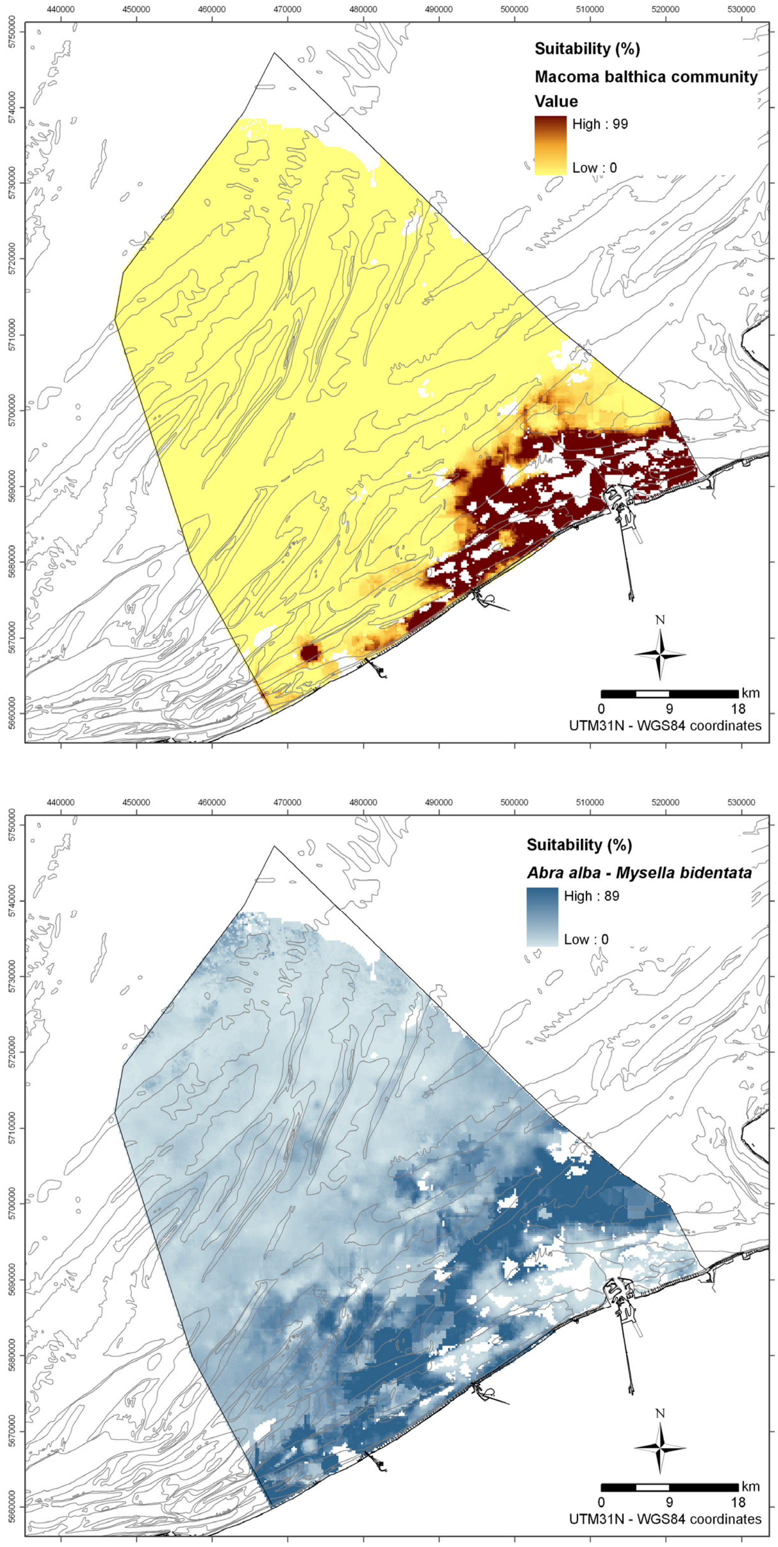

Suitability (%)
Macoma balthica community
Value
High : 99
Low : 0
N
km
0
9
18
UTM31N - WGS84 coordinates
Suitability (%)
Abra alba - Mysella bidentata
High : 89
Low : 0
N
km
0
9
18
UTM31N - WGS84 coordinates

Figure 5.3a

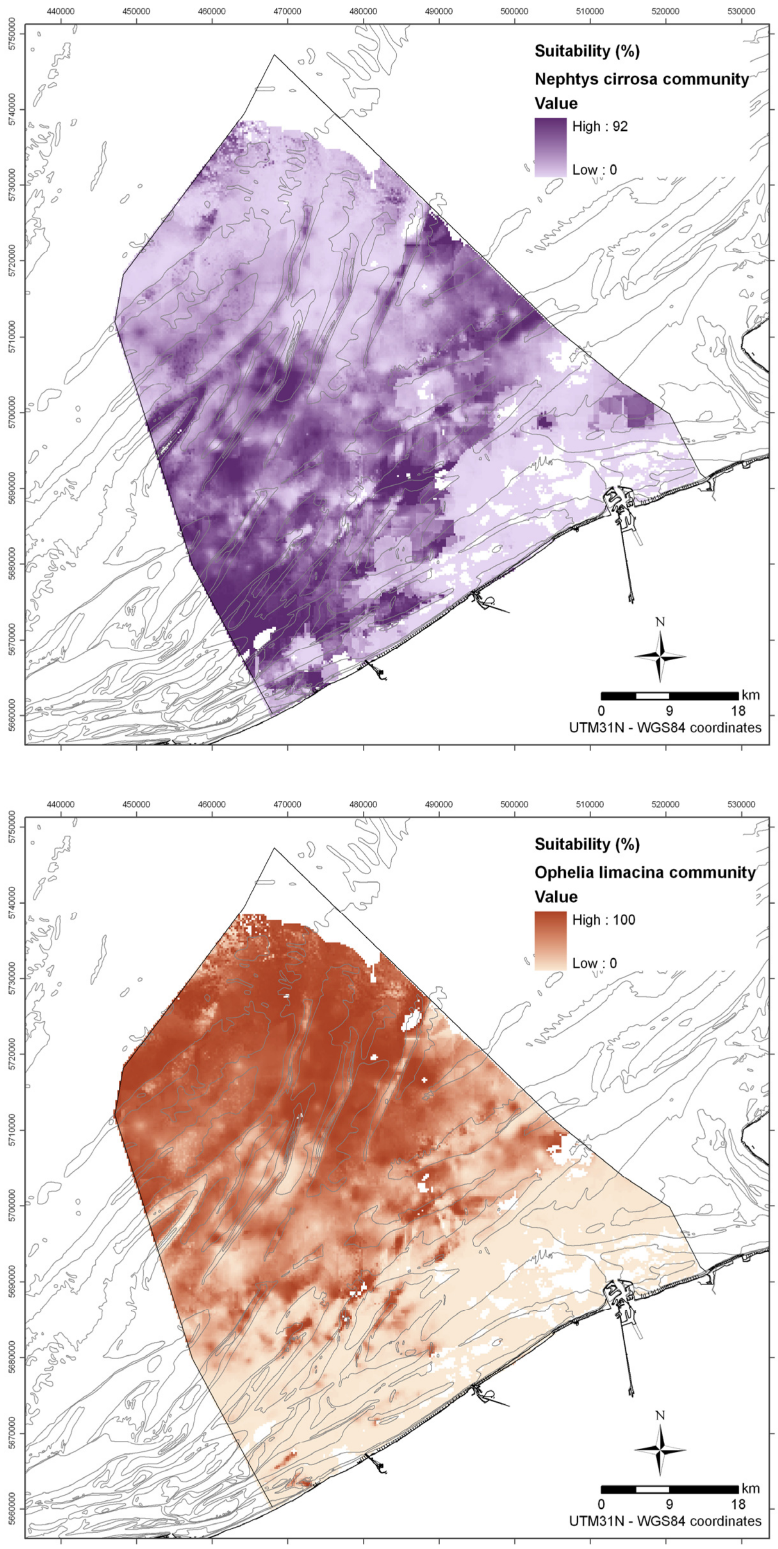


Figure 5.3b: Predicted habitat suitability maps for the *Macoma balthica* community, the *Abra alba* community, the *Nephtys cirrosa* community and the *Ophelia limacina* community in the Belgian part of the North Sea. White; no environmental data or prediction beyond the range of the model development data set.

5.4.2 Habitat suitability mapping

As demonstrated in this study the habitat suitability model can be used to predict the full coverage spatial distribution of the macrobenthic communities. Such detailed knowledge should be considered as highly relevant for marine management: a well-equilibrated marine spatial planning is particularly dependent on the data availability. However, the use of a model to increase the volume of data availability also includes some risk, of which two aspects are discussed below.

The habitat suitability model was developed based on a large data base (364 sampling stations were assigned to one of the four macrobenthic communities), but solely derived from the BPNS. Whereas the high number of sampling stations (as well as their spatial spread) included in the modelling exercise points towards a high reliability of model application within the BPNS, outside of the BPNS the model should only be used with great care. Two main types of problems may possibly be encountered. (1) Sediment types (and hence, most probably also macrobenthic communities) may be differing from the ones in the BPNS, leading to the use of the model beyond the range of the model development data set. Such error can be avoided if a threshold for maximum Mahalanobis distance from the communities centroid is set. In our mapping exercise this maximum Mahalanobis distance was set at three times the standard deviation from any macrobenthic community centroid (see Materials and Methods). (2) Although the sediment types encountered fall well within the range of sediments included in the model, the sediments host a different macrobenthic community, not present in the BPNS. Such errors are likely to occur when applying the model in other biogeographic regions, constituting a different macrobenthic species pool and/or where, next to sediment composition, other environmental variables are important in structuring the communities. It is therefore of utmost importance only to apply the model within the biogeographic region of origin. The BPNS being part of the biogeographic region of the Southern Bight of the North Sea (southern limit: 51°00'N, northern limit: 53°30'N), our model could thus be used (with care) within this region. This hypothesis is strengthened by the fact that, based on a thorough analysis of the present-day community structure in the North Sea, the macrobenthic communities from the BPNS are similar in the Southern Bight (Rachor et al. 2007). Next to its applicability to the BPNS, the model might thus be expected to perform properly in the full Southern Bight of the North Sea. Further testing is advised.

One should however always be aware that the reliability of the generated maps is depending on many aspects, of which data availability is considered extremely important. Data availability is impacting the reliability during all three stages of the habitat suitability mapping: (1) discriminating between the communities, (2) constructing the model, and (3) mapping the habitat suitability. When discriminating between the communities it is particularly important to make use of a fair amount of biological data: only when the communities can be discriminated reliably, one can go to the next step in habitat suitability modelling at the level of communities. Although there is no unambiguous method to assess this type of reliability, analytical tools such as ANOSIM (Clarke 1993) may be useful. To construct the model it is necessary to make use of enough data, linking the biological characteristics to the environmental variables: in general, the more data are available the higher the model performance. This model performance can be checked by various indicators, of which CCI and Cohen's kappa were used in this paper. Finally, the availability of environmental data becomes particularly crucial when selecting the resolution of the habitat suitability map. During this stage it is important to take care of a good balance between detail (e.g. spatial heterogeneity) and reliability. Alternatively, one could also decide not to aim at full coverage habitat suitability maps: if no spatial interpolation of environmental data is done, this last aspect of reliability can be called off.

5.4.3 Relevance for marine management

Although the spatial distribution of the macrobenthic community habitat suitability allows an easy communication with managers and policy-makers, it is important to detailedly define and comprehend its content. In this study the habitat suitability is defined as the probability to encounter a macrobenthic community in a specific habitat. Predicting the spatial distribution of a macrobenthic community however does not mean that we are able to detailedly predict its species composition at a specific site and moment. Because of short- to long-term temporal variability within temperate macrobenthic communities, the community structure should be regarded dynamic rather than static (e.g. Beukema et al. 1993; Meire et al. 1994; Turner et al. 1995; Essink et al. 1998; Herman et al. 1999). Yet, each (stable) community is expected to maintain a distinctly specific species composition and abundance respective to other communities (Turner et al. 1995), as demonstrated for the BPNS by Degraer et al. (1999b). In other words, if a habitat is found suitable for a macrobenthic community, its composing species have the possibility of colonizing the habitat, but may as well be absent because of anthropogenic impacts, such as fisheries, or natural temporal variability. Habitat suitability thus predicts the specific ecological potentials of a habitat rather than the realized ecological structure (Degraer et al 1999b).
The use of habitat suitability maps within marine management is therefore twofold: (1) a warning signal for potential anthropogenic impact and (2) a baseline map for marine spatial planning. A significant mismatch between the actual community structure and the habitat suitability map might trigger further investigation on its causes and might, as such, highlight anthropogenic impacts or eventually an ineffective marine management. Habitat suitability maps should thus be considered complementary to, rather than a substitution of, direct observations of the macrobenthic community structure. Secondly, distinguishing between areas with higher and lower macrobenthic potentials, habitat suitability maps might serve as a baseline map for marine spatial planning: taking into account the precautionary principle, high potential areas can now be avoided when spatially planning new marine activities, such as wind farms. Because, in absence of major anthropogenic impacts, the habitat suitability and thus ecological potential are far more stable through time compared to the permanently fluctuating macrobenthic communities, information on the ecological potentials of an area is of utmost importance for a scientifically-sound marine spatial planning, including MPA selection.

5.5 Conclusions

I. Median grain size and sediment mud content were selected above bathymetry, slope and distance to the coast to represent the most important environmental variables determining the macrobenthic community distribution.
II. The empirical habitat suitability model allowed to accurately predict the macrobenthic community distribution based solely information on median grain size and sediment mud content.
III. The habitat suitability could be reliably assessed for 53297 grid cells (resolution: 250 x 250 m; i.e. 96.3% of the BPNS).
IV. Next to its applicability to the BPNS, the model is anticipated to perform well in the full Southern Bight of the North Sea. Its applicability outside the Southern Bight of the North Sea should be considered at least questionable. Further testing is advised.
V. Since the habitat suitability is considered far more stable through time compared to the permanently fluctuating macrobenthic communities, information on the habitat suitability of an area is considered highly important for a scientifically-sound marine management.

5.6 Acknowledgements

The second and third author acknowledge a grant from the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). The research was conducted under the Belgian Science Policy projects *BWZee* (contract n° EV/02/37A and EV/88/37B), *HABITAT* (contract n° MN/02/89), *MAREBASSE* (contract n° EV/02/18A), and *GAUFRE* (contract n° MA/02/009 and MA/02/007), and the Agency of Maritime and Coastal Services – Coastal Division project *HABITAT* (dossier n° 99380 and 200.455). This publication contributes to the EU INTERREG III B *MESH* project (http://www.searchmesh.net), the EU Network of Excellence *MarBEF* (contract n° GOCE-CT-2003-505446; http://www.marbef.org), the Gent University *BBSea* project (contract n° GOA 01G00705; http://biology.ugent.be/bbsea), and the ICES *Study Group on the North Sea Benthos Project 2000* (http://www.ices.dk/iceswork/wgdetail.asp?wg=SGNSBP).

Chapter 6

The relevance of ecogeographical variables for marine habitat suitability modelling of *Owenia fusiformis*

SUBMITTED FOR PUBLICATION AS:

Verfaillie, E., Degraer, S., Du Four, I., Rabaut, M., Willems, W. and Van Lancker, V. (submitted to Estuarine, Coastal and Shelf Science) The relevance of ecogeographical variables for marine habitat suitability modelling of *Owenia fusiformis*.

6 The relevance of ecogeographical variables for marine habitat suitability modelling of *Owenia fusiformis*

Abstract

Predictive modelling of ecologically relevant marine habitats requires predictor or ecogeographical variables (EGVs) that govern the potential distribution of a species or community.

This paper shows for the Belgian part of the North Sea how different combinations of EGV subsets result in different habitat suitability models of the macrobenthic species *Owenia fusiformis*. This tube-building polychaete belongs to the ecologically rich *Abra alba* community, is living in a well-defined habitat and is strongly linked to the sediment composition and topography of the seabed. Therefore, subsets of sedimentary, multi-scale topographical and other EGVs (e.g. hydrodynamics) were used to predict the distribution of the species. In earlier studies, only sedimentological and bathymetrical EGVs were considered for this species.

The habitat suitability models were derived from ecological niche factor analysis (ENFA) and the best model was selected using cross-validation. The validation showed that topographical EGVs on the smallest spatial scale were crucial for a good habitat suitability model. Surprisingly, a model based exclusively on these topographical EGVs is better than a model that also includes sedimentological and hydrodynamical EGVs.

Keywords: Habitat suitability modelling, ecogeographical variables, *Owenia fusiformis*, ecological niche factor analysis, Belgian part of the North Sea

6.1 Introduction

Owenia fusiformis is a macrobenthic species that is common on the Belgian part of the North Sea (BPNS). The species belongs to the ecologically important *Abra alba* community (Van Hoey et al. 2004; Degraer et al. 2008), characterized by high densities and diversity. *Owenia fusiformis* can be considered as an ecosystem engineer influencing the benthic community locally: there is a positive correlation between the occurrence of *O. fusiformis* and the species abundance/richness of the community (Somaschini 1993). During the last decades, this tube-building polychaete has shown a considerable increase in density. Between 1976 and 1986, the species was found in low densities of maximum 15 ind./m², whereas in the period 1994-2001, densities increased to 500 ind./m² (Degraer et al. 2006). In recent years, the species has reached average densities of 165 ind./m^2, with dense aggregations of more than 4000 tubes per m² locally (Marine Biology Section, Ugent – Belgium, 2008).
The species lives in flat, soft-sediment environments. It prefers fine-to-medium sands with a grain-size between 100 and 500 µm and is characteristic of sheltered areas with high percentages of organic matter (Fager, 1964). *O. fusiformis* forms dense aggregations or patches. This clustering behaviour is comparable to that of the tube-building, habitat-engineering polychaete *Lanice conchilega*, (Rabaut et al. 2007). The tubes of *O. fusiformis* are shorter (i.e. 12-13 cm; Fager 1964) and the organisms have a longer lifespan (i.e. on average three years; Ménard et al. 1989).
The flexible tube of cemented sand grains and shell fragments is longer than the worm itself (Hartmann-Schröder 1996) and protrudes from the surface. When the species occurs in dense aggregations, the biogenic structures of protruding tubes can be detected by acoustic methods. The patches described in Van Lancker et al. (2007) protruded 18 to 40 cm above the surrounding sediment.
Classification and modelling techniques used in habitat mapping are all based on the assumption that the suitability of an area for a certain species is related to its predictor or ecogeographical variables (EGVs), corresponding to the environmental variables relating to factors of potential relevance to the focal species (e.g. substrate type, topographical position, hydrodynamical regime or presence/absence of another species). The ecogeographical information is generally more widely available than ecological sample information of the focal species. As such, it becomes possible to predict how suitable a habitat is for a certain species, using specific combinations of these EGVs in a habitat suitability model (HSM).
For soft-substrate habitats, grain-size and silt-clay percentage are commonly thought to be the most influential EGVs for the modelling of macrobenthic species (Van Hoey et al. 2004; Degraer et al. 2008; Willems et al. 2008). However, the spatial distribution of some macrobenthic species or communities is known to be patchy or bound to topographical variation (Rabaut et al. 2007). Multi-scale topographical characteristics are therefore believed to be important EGVs too (Guisan and Thuiller 2005; Baptist et al. 2006; Wilson et al. 2007).
This paper demonstrates the influence of different combinations of EGVs on the result of the habitat suitability (HS) modeling of *O. fusiformis*. The different combinations of EGVs include or exclude specific (multi-scale) topographical, sedimentological or other EGVs.

6.2 Material and methods

6.2.1 Study area and datasets

The study area covers the entire BPNS with a surface of 3600 km², situated on the North-West European Continental Shelf (Figure 6.1). The BPNS is relatively shallow and dips gently from 0 to -50 m MLLWS (Mean Lower Low Water Springs). A highly variable topography, dominated by a series of sandbanks and swales, characterizes the seabed surface.

Thirty-seven full-coverage topographical, sedimentological and hydrodynamical EGVs were used, all available as raster maps of 250 x 250 m resolution (Table 6.1). The variables were made more symmetrical by the Box-Cox standardizing algorithm (Sokal and Rohlf 1981), a procedure that produces a distribution as close to a Gaussian as possible.

Owenia fusiformis was found in 193 stations at the BPNS (Figure 6.1) (Marine Biology Section, Ugent – Belgium, 2008). For each biological sample, the density of *O. fusiformis* was calculated. Densities were expressed as individuals/m² and were divided into four classes, giving different weights to the presence maps: 1) 1-10 specimens; 2) 11-100 specimens; 3) 101-1000 specimens, 4) 1001-6000 specimens.

6.2.2 Research strategy

The research strategy comprised four steps: (1) selection and production of EGVs; (2) Ecological Niche Factor Analysis (ENFA), identifying ecologically relevant factors; (3) production of HS models of *O. fusiformis*, using the geometric mean distance algorithm and different combinations of EGVs; (4) cross-validation of different HS models, enabling the identification of the best explanatory EGVs for each study area.

Software used was ArcGIS 9.2 for GIS analyses and mapping; Landserf 2.2.0 (Wood 2005) for multi-scale topographical analyses; Biomapper 3.2 (Hirzel et al. 2002b; and Hirzel et al. 2006) for ENFA, HS modelling and validation.

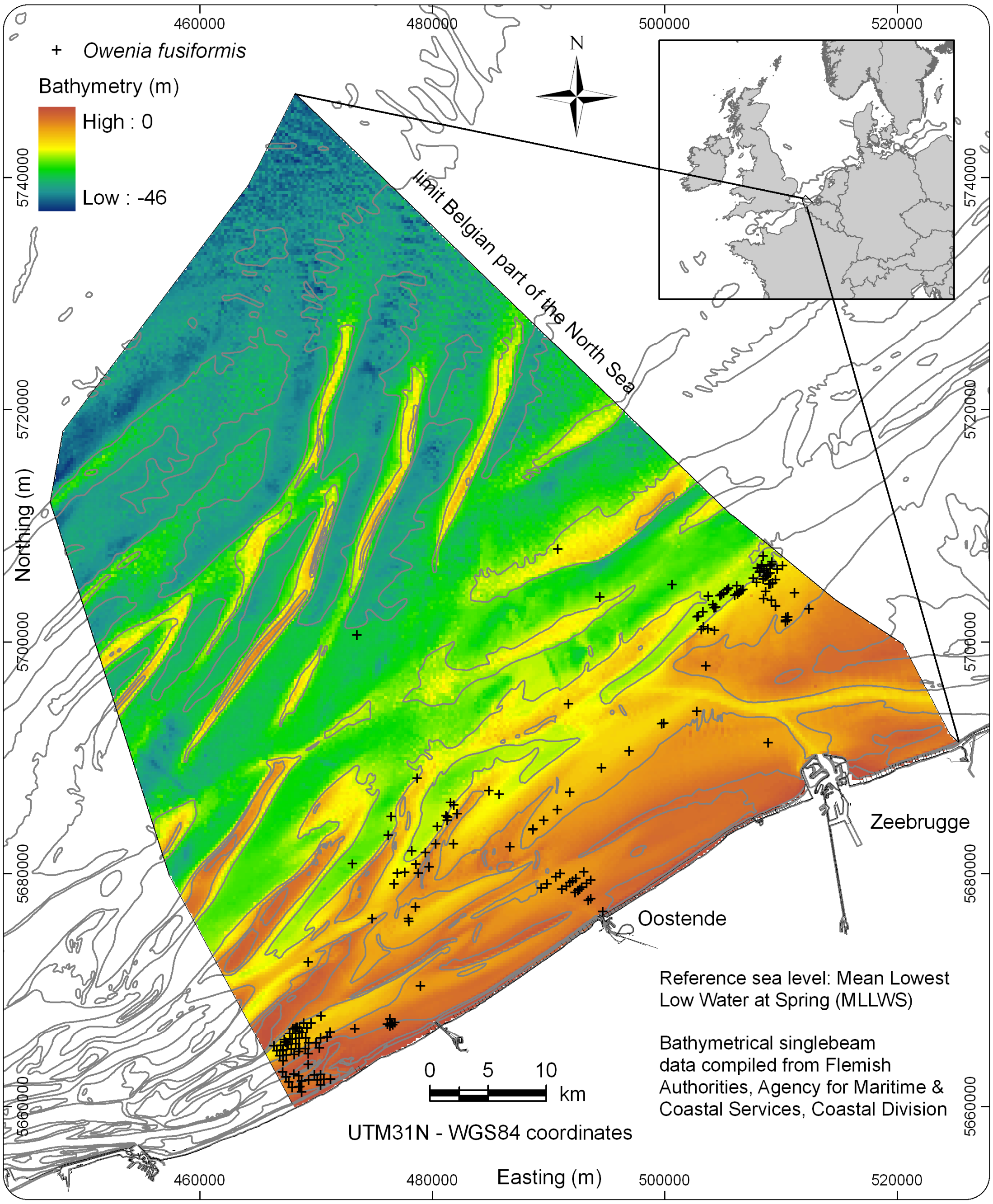


Figure 6.1: Bathymetrical map of the Belgian part of the North Sea, located in northwestern Europe, with indication of presence samples of *Owenia fusiformis*.

6.2.3 Step 1: Selection and production of EGVs

The median grain-size of the sand fraction and the silt-clay percentage were used as sedimentological EGVs (Table 6.1).

Following Wilson et al. (2007), terrain-analysis variables can be grouped into four classes, all of them being derivatives from a digital terrain model (DTM):

- slope;
- orientation (aspect);
- curvature and relative position of features;
- terrain variability.

All of these variables can be computed in GIS software or specialized terrain-analysis software (e.g. LandSerf from Wood (2005)).
Terrain analysis may be performed on a DTM represented as a raster grid, or on a continuous representation of a DTM as a double-differentiable surface (e.g. Wood 1996). The latter approach offers great flexibility in the choice of algorithms of terrain analysis and in the scales at which the analyses are performed. Following Evans (1980), a DTM is approximated by a bivariate quadratic equation:

$$Z = aX^2 + bY^2 + cXY + dX + eY + f \tag{6.1}$$

with Z = height of the DTM surface;
X and Y = horizontal coordinates;

The coefficients a, b, c, d, e and f in equation 6.1 can be solved within a window using simple combinations of neighboring cells, which is the basis for terrain analysis in most GIS software, regardless of their using grid-based methods or a mathematical representation of the DTM.
For a terrain analysis across a variety of spatial scales, Wood (1996) solves this equation for an n by n matrix with a local coordinate system (x, y, z) defined with the origin at the central pixel. The user may specify any odd number (n) for the size of the square analysis window defining the part of the raster DTM to be analyzed in relation to each central pixel in turn. To compute the terrain parameters in LandSerf (Wood 2005), an analysis window is effectively moved across the raster DTM surface such that each pixel in turn becomes the central pixel on which calculations are based.

For the BPNS, the multi-scale terrain analysis resulted in a range of EGVs such as slope and other bathymetrical derivatives, measured over four spatial scales (see Table 6.1 for an overview). Window sizes of 3, 9, 17 and 33 pixels, corresponding to 750, 2250, 4250 and 8250 m, respectively, were used for this analysis. The window sizes of 3, 9, 17 and 33 pixels, were chosen because they provide an adequate cover of different spatial scales, following Wilson et al. 2007.
Hydrodynamical EGVs were the maximum bottom shear stress and the maximum current velocity (Table 6.1). Furthermore, the MERIS satellite-derived EGVs maximum Chlorophyl a concentration and maximum Total Suspended Matter were used (Table 6.1). Finally, the distance to the coastline was computed (Table 6.1). The resolution of all of the EGVs was 250 x 250 m, except for maximum Total Suspended Matter and maximum Chlorophyl a concentration; these had original resolutions of 1 x 1 km, but were resampled to 250 x 250 m.

Table 6.1: Full-coverage ecogeographical variables (EGVs) as input for the Ecological Niche Factor Analysis and habitat suitability modelling. The multi-scale topographical EGVs have window sizes of 3, 9, 17 and 33 pixels. Fractal dimension EGVs have only 3 window sizes of 9, 17 and 33 pixels. (/ = no unit)

EGV	Unit	Reference or procedure
Sedimentology $d_s x$ = diameter for which x% of the sand fraction (63-2000 µm) in the sample has a smaller diameter • $d_s 50$ = median grain-size of sand fraction • Silt-clay percentage (silt-clay %) = 0-63 µm	 µm %	Reference: sedisurf@database hosted at Ghent University, Renard Centre of Marine Geology. Reference: Verfaillie et al. 2006 Reference: Van Lancker et al. 2007
Topography • Digital terrain model (DTM) of multibeam bathymetry	m	Reference: DTM from Flemish Authorities, Agency for Maritime and Coastal Services, Flemish Hydrography
• Slope (slp) = a first derivative of the DTM	°	Procedure: Evans (1980); Wilson et al. (2007)
Aspect = a first derivative of the DTM Indices of northness and eastness provide continuous measures (−1 to +1), describing the orientation of the slopes. • Eastness (eastn) = sin (aspect) • Northness (northn) = cos (aspect)	 / /	Procedure: Wilson et al. (2007); Hirzel et al. (2002a)
Surface curvature = a second derivative of the DTM • Profile curvature (prcurv) = rate of change of slope along a profile in the surface; useful to highlight convex and concave slopes • Plan curvature (plcurv) = rate of change of aspect in plan across the surface; useful for defining ridges, valleys and slopes • Mean curvature (mcurv) = average value obtained from maximum and minimum profile curvature, providing an indication of the relative position of features	 / / /	Procedure: Evans (1980); Wilson et al. (2007)
• Rugosity (rug) = ratio of the surface area to the planar area, across the neighbourhood of the central pixel	/	Procedure: Jenness (2002); Lundblad et al. (2006); Wilson et al. (2007)
Bathymetric Position Index (BPI) = measure of where a location, with a defined elevation, is relative to the overall landscape • BPI with a window size of 6 pixels (BPI_1500m)	 /	Procedure: Lundblad et al. (2006); Wilson et al. (2007)
• Fractal dimension (fd) = a measure of the surface complexity	/	Procedure: Mandelbrot (1983); Wilson et al. (2007)
Hydrodynamics • Maximum bottom shear stress (bstrx) = frictional force, exerted by the flow per unit area of the seabed • Maximum current velocity (mmax)	N/m^2 m/s	Reference: Management Unit of the North Sea Mathematical Models and the Scheldt estuary (MUMM)
Satellite-derived variables • Maximum Chlophyl a concentration (max Chl a) over a 2-year period (2003-2004) • Maximum Total Suspended Matter (max TSM): measure for turbidity over a 2-year period (2003-2004)	mg/m^3 mg/l	Reference: MERIS data processed by MUMM in the framework of the BELCOLOUR-2 project (ESA ENVISAT AOID3443)
• Distance to coast (distcst)	km	Computed in GIS

6.2.4 Step 2: Ecological Niche Factor Analysis

ENFA (Hirzel et al. 2002) is a statistical technique, based on Hutchinson's (1957) concept of the ecological niche, that computes suitability functions for species by comparing the EGVs of the species with those of the whole set of cells. Unlike other HS modelling techniques, ENFA needs only presence data of species. This was appropriate for the current dataset, since some samples were taken only where the presence of *O. fusiformis* was likely, as determined from highly detailed multibeam observations.

Contrary to PCA, where axes are chosen to maximize the variance of the distribution, ENFA computes ecologically relevant factors. Still, the output of ENFA is similar to that of PCA, with the results being sets of new, linearly independent variables, combining the original EGVs.

Species are generally expected to show non-random distributions with respect to EGVs, meaning that a species with e.g. an optimum depth is expected to occur within this optimal range. As such, the depth distributions of the cells in which species are observed, in comparison with the whole set of cells, may be quantified. These distributions may be different for different species, regarding their mean and standard deviations.

For one single EGV, the species' marginality (M) can be defined as the absolute difference between the global mean (m_G) and the species mean (m_S), divided by 1.96 standard deviations (σ_G) of the global distribution. A large M value close to 1, means that the species lives in a very particular habitat relative to the reference set. The operational definition of marginality as it is implemented in the Biomapper software (Hirzel et al. 2002b), is a multivariate extension of the species' marginality (i.e. the *global* marginality). This is an overall marginality M computed over all EGVs, allowing the comparison of the marginalities of different species within a given area.

Similarly, the specialization S for one single EGV can be defined as the ratio of the standard deviation of the global distribution (σ_G) to that of the focal species (σ_S). Any S value exceeding 1, indicates some form of specialization (i.e. a narrow niche breadth in comparison with the available conditions). Again, a global specialization index for all of the EGVs, can be computed. This value ranges from 1 to infinity. For ease of interpretation, the global tolerance coefficient, defined as the inverse of the specialization, is usually preferred as it ranges from 0 to 1. It is an indicator of the species' niche breadth.

The multivariate niche (i.e. for all of the EGVs) can be quantified on any of its axes by an index of marginality and specialization. ENFA chooses its first axis to account for all the marginality of the species, and the following axes to maximize specialization.

The broken-stick method (MacArthur 1960; Frontier 1976; and Legendre and Legendre 1998) was used to decide on the number of factors to retain for the HS modelling.

To test and compare the performance of the models, eight combinations of EGVs were used to compute the factors (Table 6.2): (1) all EGVs; (2) all topographical EGVs on all of the spatial scales; (3) all topographical EGVs computed with window size 3 (including DTM and BPI); (4) all topographical EGVs computed with window size 9; (5) all topographical EGVs computed with window size 17; (6) all topographical EGVs computed with window size 33; (7) sedimentological EGVs; (8) sedimentological EGVs and all topographical EGVs computed with window size 3 (including DTM and BPI).

Table 6.2: Different combinations of EGVs (/ = not used for analysis; X = used for analysis; M = used at multi-scale, with window sizes of 3, 9, 17 and 33 pixels; 3; 9; 17 and 33: used with window sizes of 3, 9, 17 and 33 pixels, respectively).

	Combinations							
	1	2	3	4	5	6	7	8
d_s50	X	/	/	/	/	/	X	X
silt-clay %	X	/	/	/	/	/	X	X
DTM	X	X	X	X	X	X	/	X
slp	M	M	3	9	17	33	/	3
eastn	M	M	3	9	17	33	/	3
northn	M	M	3	9	17	33	/	3
prcurv	M	M	3	9	17	33	/	3
plcurv	M	M	3	9	17	33	/	3
mcurv	M	M	3	9	17	33	/	3
fd	M	M	/	9	17	33	/	/
rug	3	3	3	/	/	/	/	3
BPI_1500m	X	X	X	/	/	/	/	X
bstrx	X	/	/	/	/	/	/	/
mmax	X	/	/	/	/	/	/	/
max Chl a	X	/	/	/	/	/	/	/
max TSM	X	/	/	/	/	/	/	/
distcst	X	/	/	/	/	/	/	/

6.2.5 Step 3: Habitat suitability modelling

Several algorithms are implemented in Biomapper to compute, for each grid cell, the suitability for *O. fusiformis*: median (of the species distribution on all selected niche factors), distance harmonic mean, distance geometric mean (GM) and minimal distance algorithm (Hirzel et al. 2002; Hirzel and Arlettaz 2003). The GM algorithm (Hirzel and Arlettaz 2003) was selected for the present study, as validation (Step 4) showed that the other algorithms gave systematically worse results than the GM algorithm. For this algorithm, no assumptions are made on species distribution. The suitability of any point **P** in the environmental factor space is the geometric mean H_G of N species-observation points $\mathbf{O}_i$, which is computed from the distances to all observations:

$$H_G(\mathbf{P}) = \sqrt[N]{\prod_{i=1}^{N} \delta(\mathbf{P}, \mathbf{O}_i)} \qquad (6.2)$$

For the observations of *O. fusiformis*, a frequency distribution and a GM algorithm value are computed for each ecological niche factor. The farther away a grid cell is from this value, the less suitable it is. The suitability index ranges from 0 to 1.

6.2.6 Step 4: Validation

A k-fold cross-validation splits the species data into k sets. Then, k-1 sets are used to compute a HS model and the remaining set is used to validate this model. This strategy is repeated k times. For each of the models, the cross-validation results in k different HS models. By determining how the results vary, their predictive power is assessed. For this study, use was made of the continuous Boyce index (Hirzel et al. 2006), following the approach of Boyce et al. (2002). This method partitions the habitat suitability range into *b* classes. For each class *i*, it calculates 2 frequencies: 1) P_i, the predicted frequency of evaluation points; and 2) E_i, the expected frequency of evaluation points or the frequency expected from a random distribution across the study area, given by the relative area covered by each class. For each class *i*, the predicted to expected (*P/E*) ratio is F_i. A low-suitability class should contain fewer evaluation presences than expected by chance, resulting in $F_i < 1$, whereas high-suitability classes are expected to have F_i values higher than 1. A good model is thus expected to show a monotonously increasing curve (increase of F_i and increase of habitat suitability). Boyce et al. (2002) proposed to measure this monotonous increase by the Spearman rank correlation coefficient between F_i and *i* (i.e. “Boyce index” varying between -1 and 1, corresponding to a bad and a good model, respectively). Because of the sensitivity to the number of suitability classes *b* and to their boundaries, Hirzel et al. (2006) proposed a new index (“continuous Boyce index”), determined by using a moving window of width W (e.g. 0.2), instead of fixed classes. The computation of this index starts with a first class covering the suitability range [0,W] whose *P/E* ratio is plotted against the average suitability values of the class, W/2. In the next step, the moving window is shifted a short distance upwards and *P/E* is plotted again. This is repeated until the window reaches the last possible range [1-W,1]. This results generally in a smooth *P/E* curve, from which a continuous Boyce index is computed.

For this study, the continuous Boyce index was calculated with a moving window size of 0.2. The eight HS models resulting from the eight combinations of EGVs (Table 6.2) were validated with a k value of 10.

6.3 Results

The validation using combination set 3 (all topographical EGVs computed with window size 3 (including DTM and BPI); Table 6.2) gave the best result (Figure 6.2). This model was thus selected as the final model for the BPNS. Combination sets 4, 5 and 1 (i.e. topographical EGVs with window sizes of 9 and 17; and all EGVs, respectively) also gave rather good results (indices > 0.5). The model obtained by combination set 6 (i.e. topographical EGVs with a window size of 33, i.e. computed on a large spatial scale of more than 8 km), resulted in a very bad model (with a validation index equal to zero). Combination sets 2, 7 and 8 gave intermediate results.

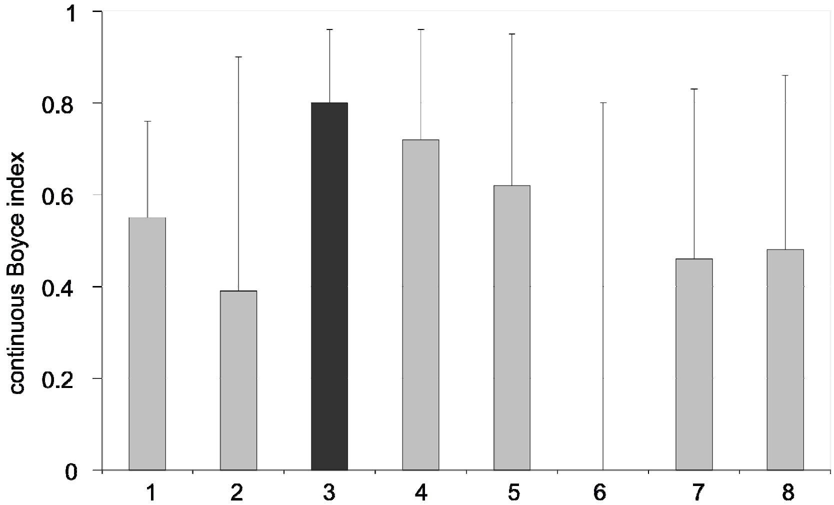


Figure 6.2: Continuous Boyce indices for the Belgian part of the North Sea for different combinations of EGVs (X-axis) (Table 6.2). The best combination is marked in black (combination 3; all topographical EGVs computed with window size 3 (including DTM and BPI) Error bars correspond to 1 standard deviation.

ENFA, based on the final model, results in a global marginality coefficient of 0.63; a global specialization coefficient of 1.92 (i.e. a global tolerance coefficient of 0.52), indicating that *O. fusiformis* lives in conditions rather uncommon for the BPNS and that its niche breadth is generally slightly narrow. After comparing the ENFA eigenvalues with the broken-stick distribution, the first 7 factors were kept as significant for the analyses (Table 6.3). They explain 97% of the information (i.e. 100% of the marginality and 95% of the specialization).

Owenia fusiformis shows the highest marginality score for the bathymetry (DTM), meaning that the species prefers higher-than-average values (i.e. a preference for shallow locations). However, the specialization for this predictor is rather low, meaning that the species is quite tolerant regarding the depth. Regarding the BPI; the species prefers flatter-and-more-depressed-than-average conditions (i.e. preference for depressions and flat areas). Regarding eastness and northness, the species has a preference for more western and more northern orientations than average, although the species is again very tolerant for these predictors.

The HS map of *O. fusiformis* in Figure 6.3 shows that high suitabilities are mainly expected in the coastal zone extending no more than 30 km offshore. Highest suitabilities are found at the N side of the Vlakte van de Raan, extending towards the Netherlands; at the S and N side of the Oostendebank; at the S and N side of the Nieuwpoortbank; and between the Middelkerkebank and Kwintebank. Too a lesser extent, high suitabilities are found at the N side of the Thorntonbank; at the S and N side of the Akkaertbank; and at the S side of the Wenduinebank.

Table 6.3: Correlation between ENFA factors and the EGVs for the final model. The percentages indicate the amount of specialization accounted for by the factor (factor 1 explains 100% of the marginality).

	Factor 1 (20%)	Factor 2 (33%)	Factor 3 (13%)	Factor 4 (10%)	Factor 5 (9%)	Factor 6 (7%)	Factor 7 (3%)
BPI_1500m	---	0	*	**	**	********	*
DTM	++++++++	0	***	0	*	***	0
eastn3	--	0	**	*	******	**	*
mcurv3	--	********	**	****	*****	0	*******
northn3	++	0	**	*	****	**	0
plcurv3	+	*	0	**	***	***	*****
prcurv3	--	******	0	*********	**	***	*****
rug	---	*	********	*	***	**	0
slp3	--	0	***	**	*	**	**

Factor 1 is the Marginality factor. Positive values indicate higher-than-average values. Negative values mean the reverse. The greater the number of symbols, the higher the correlation. 0 indicates a very weak correlation.

Factors 2 to 7 are Specialization factors. The symbol * means *O. fusiformis* is found occupying a narrower range of values than available. The greater the number of *, the narrower the range. 0 indicates a very low specialization.

Considering HS values ≥ 60% as highly suitable for the species, the following ranges of EGVs correspond to optimal environmental conditions (EGV ranges are obtained by selecting pixels with HS values higher than 60 %): (1) shallow-water environments (8 - 20 m); (2) fine sandy sediments (median grain-size between 145 and 285 µm); (3) moderately high silt-clay % (0.5 - 20 %); (4) high amounts of Total Suspended Matter (> 25 mg/l); (5) maximum Chlorophyll a concentration of 30 to 40 mg/m^3; (6) moderate maximum bottom shear stresses (1.2 - 2.3 N/m^2); (7) maximum current velocities of 0.8 to 1.0 m/s; (8) topographies with northwest orientations (northness > 0.5 and eastness < 0); and (9) flat topographies (slope around 0°). Table 6.4 gives a summary of the mean and standard deviations of the EGVs. As multi-scale EGVs give similar values for different window sizes, only values for window size 3 are given. Although only topographical EGVs with a window size of 3 are used for the final model, other EGV ranges are also given in Table 6.4.

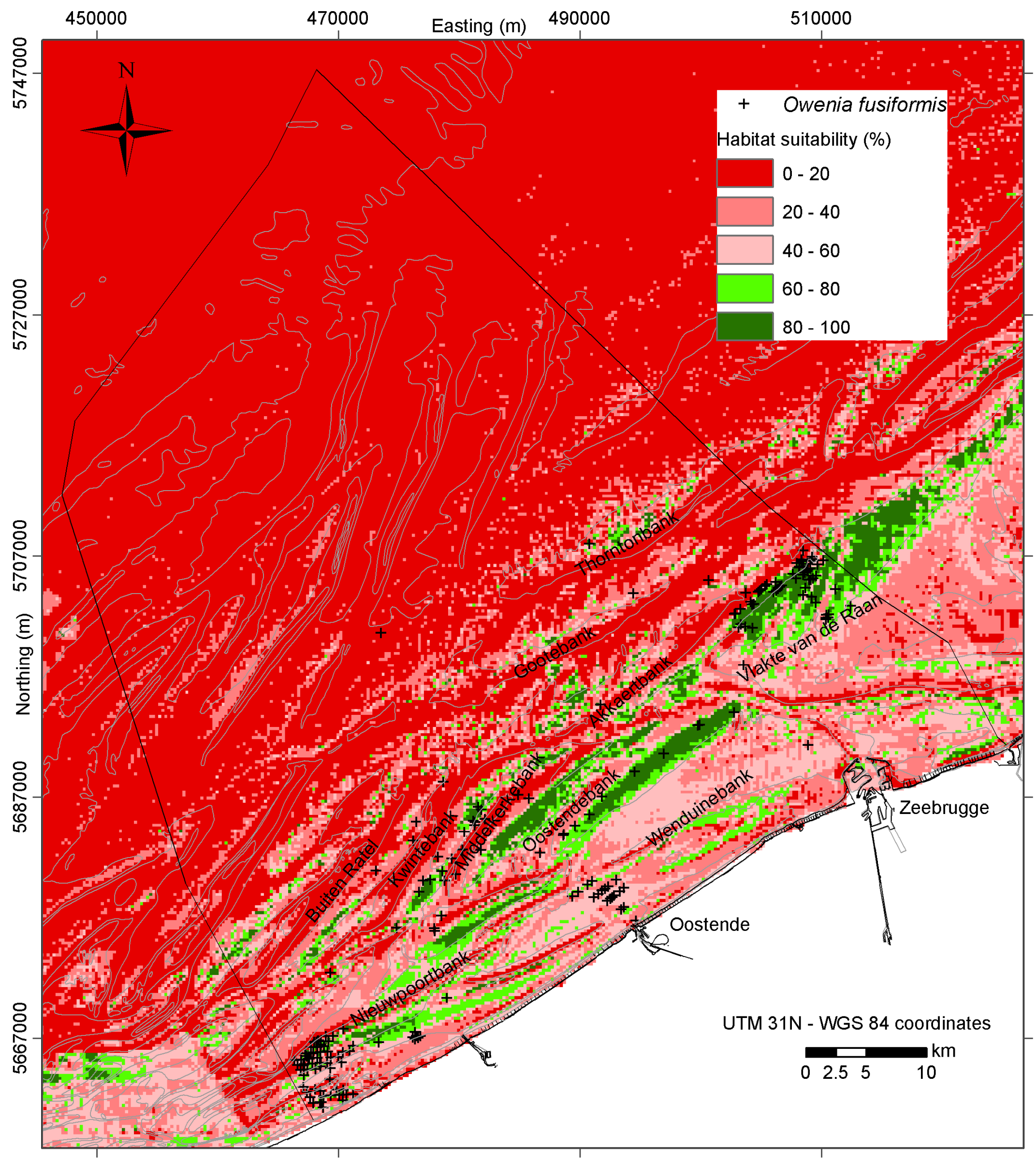


Figure 6.3: Final habitat suitability model of the Belgian part of the North Sea for *Owenia fusiformis.* The optimal niche of *Owenia fusiformis* lies in the coastal zone (maximum 30 km offshore), in between the major Coastal Banks.

Table 6.4: Mean and standard deviations (SD) for the EGVs on the BPNS, corresponding to HS values ≥ 60%. The EGVs and their units are described in Table 6.1.

EGV	Mean ± SD
d_s50	214.91 ± 68.43
silt-clay %	10.89 ± 10.73
DTM	-13.22 ± 5.49
slp	0.19 ± 0.14
eastn	-0.44 ± 0.38
northn	0.57 ± 0.58
prcurv	0.00 ± 0.01
plcurv	0.16 ± 2.47
mcurv	0.00 ± 0.01
rug	1.00 ± 0.00
BPI_1500m	0.07 ± 0.91
bstrx	1.79 ± 0.49
mmax	0.91 ± 0.11
max Chl a	35.25 ± 3.43
max TSM	34.21 ± 7.84
distcst	10.73 ± 5.99

6.4 Discussion

The aim of this paper is to examine which EGVs are the most important predictors of the species *O. fusiformis*.

6.4.1 Comparison of EGV conditions with values from the literature

For the Baie de Seine (France), *O. fusiformis* has been described as one of the ten most abundant species (ind/m²) of the *Abra alba* community (Fromentin et al. 1997; Van Hoey et al. 2004). However, d_s50 of that community varied between 80-120 µm (being much finer sediment than what has been encountered in this study), while the mean depth is around 10.5 m (somewhat shallower than the value determined in this study). In Van Hoey et al. (2005), *Owenia fusiformis* was also identified as one of the ten most abundant species of the *A. alba* community on the BPNS. According to those authors, the mean d_s50 of that community was 222 ± 45 µm, the mean silt-clay% was 14 ± 11 % and the mean depth was 10.8 m. These results are highly comparable with the values found for *O. fusiformis* in this study. This means that regarding the sedimentological and the depth conditions, the *A. alba* community and *O. fusiformis* are very similar. Dauvin et al. (2004) described that *O. fusiformis* has a preference for sand with high percentages of silt-clay, which also corresponds to the results of this study.

To our knowledge, only sedimentological and bathymetrical EGVs have ever been considered for this species. This study demonstrates that other conditions (such as bathymetric derivatives) are also important predictors. Although a good model is obtained using all of the EGVs (combination set 1 in Figure 6.2) and thus including depth and sedimentology, the best model is obtained from topographical variables alone, on a spatial scale of 750 m (including depth and BPI on a spatial scale of 1500 m; combination set 3 in Figure 6.2). This is a surprising result, as those EGVs do not contain any information other than topography (e.g. sedimentology or hydrodynamics).

6.4.2 Spatial scale and spatial structure

Figure 6.2 shows that combinations 3, 4 and 5 give similar good results for modelling the occurrence of *O. fusiformis*. This indicates that the terrain variables with window sizes of 3, 9 and 17 pixels (or spatial scales of 750, 2250 and 4250 m, respectively; similar to the spatial scale of a sandbank) contain overlapping and no complementary information. The fact that the model based on a window size of 33 pixels (or 8250 m) has a zero validation index, means that terrain variables calculated on this spatial scale are too broad or too general to be good predictors for *O. fusiformis*.

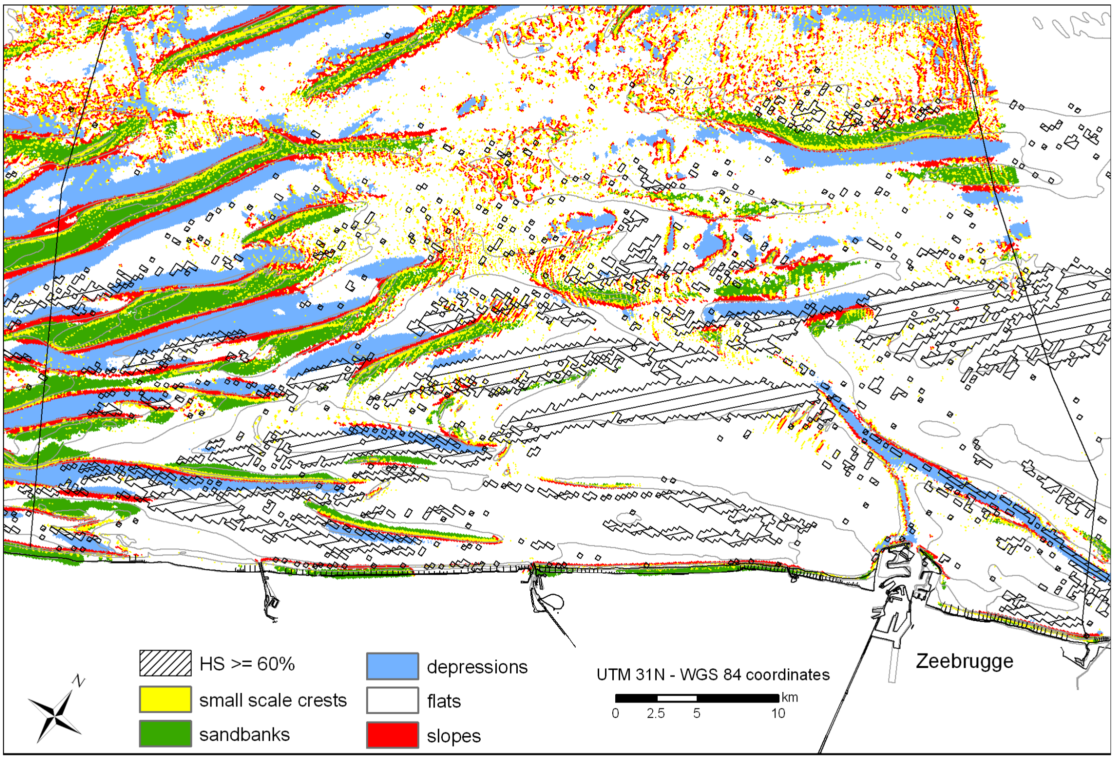


Figure 6.4: Habitat suitability (HS) zones of ≥ 60% plotted on BPI classification: small scale crests (bedforms), sandbanks, depressions (or swales), flats and slopes. The high-suitability zones exclude the large sandbanks. The area shown is the southern part of the Belgian part of the North Sea.

It is remarkable that high-suitability zones are located in areas where no sandbanks are present (Figure 6.4; representing a classification of topographical features, based on BPI and slope). The zones with higher HS are located mainly in shallow, flat zones. Still, many other shallow, flat zones have low habitat suitabilities (mainly in the southeastern coastal zone and farther away from the coastline). This is due to a combination of non-suitable EGVs, but it is difficult to determine which conditions are most constraining and which are not. At first sight, the low suitabilities around the harbour of Zeebrugge (SW part of the BPNS) are due to high silt-clay %. Still, as this predictor is not an input EGV for the final model, the low HS cannot be predicted by

this variable. Two EGVs that are input variables for the final model are rugosity and slope, and in this area of low suitability, the rugosity and the slope are generally very low compared to the rest of the BPNS (i.e. a completely flat area, without any small-scale topography that can be computed on a resolution of 250 x 250 m). This area of very low rugosity (equal to 1.00) and slope (quasi equal to 0.00), corresponds to a large extent with the zone of high silt-clay % around the harbour of Zeebrugge (Figure 6.5). Where, in this zone of high silt-clay %, slightly higher rugosity and slope values are observed and where small scale crests and the anthropogenic navigation channel are located (Figure 6.4), the model predicts higher suitabilities (although this is probably not correct in the case of the navigation channel). This suggests that *O. fusiformis* has a preference for very flat areas, that are in a very minor extent affected by some rugosity and some slope (order of magnitude between 0.1 and 0.2°). An interesting question that rises from these observations, is whether the dense aggregations of *O. fusiformis*, that protrude above the surface are reflected by these slightly higher rugosity and slope values. Further research regarding this issue, is needed.
These observations indicate as well that sedimentological and hydrodynamical EGVs, considered as crucial for predicting the species until now, can be replaced to a certain extent by topographical EGVs. Further research is necessary to examine why sedimentological and hydrodynamical EGVs are not selected by the best model. A possible cause is the fact that these EGVs are correlated to the topographical EGVs and as such contain redundant information, that is covered sufficiently by the information contained in the topographical EGVs.

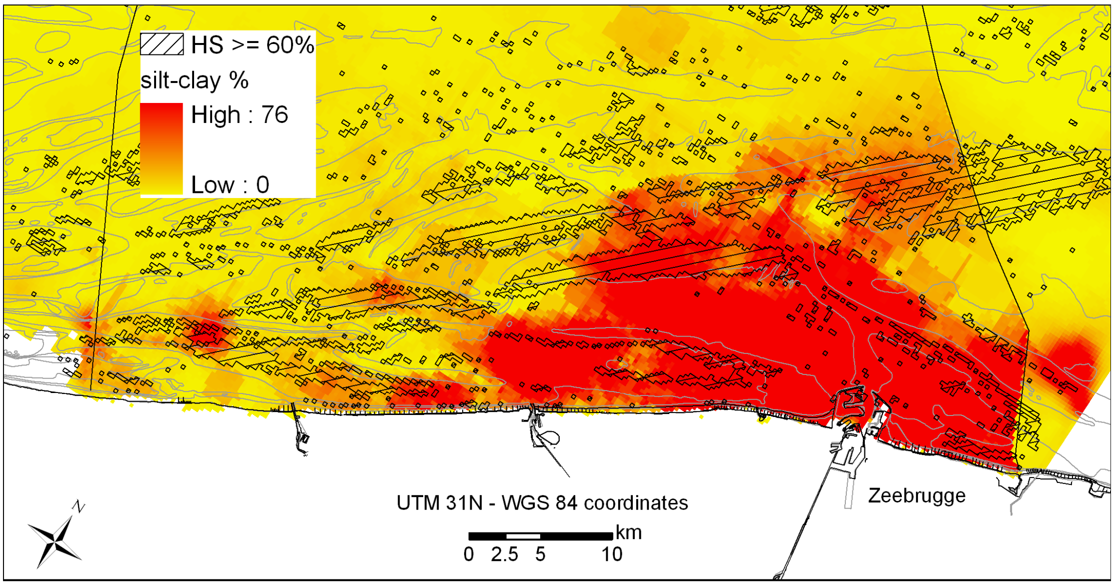


Figure 6.5: Habitat suitability (HS) zones of ≥ 60% plotted on silt-clay %. The high-suitability zones around the harbour of Zeebrugge are located mainly in the zone of higher silt-clay %, although topographical EGVs (particularly slope and rugosity) make a subtle difference between lower (extreme low slope and rugosity) and higher (slightly higher slope and rugosity) suitabilities. The area shown is the southern part of the Belgian part of the North Sea.

Of course, as explained by Legendre (1993), the occurrence of species is not only predicted by non-spatially and spatially structured environmental variance (corresponding to the variance of the EGVs). Species distribution and density can also be associated with a spatial structure that cannot be explained by environmental conditions, such as biotic processes within the population or community (e.g. competition, predation, recruitment processes, etc.). An example of such a recruitment process was demonstrated by Callaway (2003), where juveniles of the tube-building polychaete *L. conchilega* attached to adults, stimulating aggregation of the species.

6.5 Conclusion

Past studies demonstrated that the occurrence of *O. fusiformis* can be predicted by the sedimentology and the depth. The present study shows that multi-scale topographical EGVs are also important predictors. Remarkably, the selection of one single spatial scale comparable with the spatial scale of a sandbank (i.e. between 750 and 4250 m) is sufficient to obtain a successful model of the spatial distribution of *O. fusiformis*.
The main predictors selected by the best model are all terrain EGVs computed with a window size of 3 pixels or 750 m. This indicates that sedimentological EGVs, considered as crucial for predicting the species until now, contain overlapping information with the topographical EGVs.
The shallow coastal zone, away from the large sandbanks, has the highest habitat suitability for *O. fusiformis* on the BPNS. The species prefers mainly flat areas with some minimal rugosity and slope.

6.6 Acknowledgements

This paper is part of the PhD research of the first author, financed by the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). It also contributes to the EU INTERREGIII B project MESH (''Development of a framework for Mapping European Seabed Habitats''. In addition, it frames into the research objectives of the project MAREBASSE (''Management, Research and Budgeting of Aggregates in Shelf Seas related to End-users'', EV/02/18A), financed by the Belgian Science Policy. The sedisurf@database has been compiled by the Renard Centre of Marine Geology (Ghent University) and consists of data from various institutes.

Chapter 7

Discussion

7.1 Towards maximal objectivity and optimized (geo)statistical approaches in Habitat Mapping

In all of the chapters, methodologies were put forward that strive towards a **maximal objectivity in habitat mapping**. To that end, **(geo)statistical methods** were tested and applied to the marine datasets, presented in this thesis.
In **Theme 1**, the coverage data for habitat mapping were optimized with the use of geostatistical interpolation techniques (or Kriging). Kriging techniques are objective in the sense that they make use of the spatial correlation between neighbouring observations, to predict values at unsampled places (Goovaerts 1999). These techniques give an indication of the errors and uncertainties associated with the interpolated values, based on a variance surface of the estimated values (Burrough and McDonnell 1998).
If there is a linear relation between the sedimentological variable (e.g. grain-size) and a secondary variable (e.g. bathymetry), **multivariate geostatistical techniques,** such as Kriging with an external drift (KED), can be applied.
In **Chapter 2**, KED was tested for the first time on sedimentological data of the entire Belgian part of the North Sea (BPNS). The multivariate geostatistics were limited to a **single secondary dataset**, being the **bathymetry**. Because of the linear relation between the bathymetry and the median grain-size, the interpolation could be improved significantly.
In **Chapter 3**, the same technique (KED) was applied to a set of sedimentological variables (d_s10, d_s50, d_s90 and silt-clay %) for a small study area on the BPNS. Instead of one single secondary dataset, a **whole set of secondary data** was used; **multi-scale topographical datasets** such as slope, aspect and curvature, on different spatial scales, were derived from the bathymetry using spatial analyses. Those datasets were used as secondary datasets for KED, after being reduced by Principal Components Analysis (PCA); this avoided multicollinearity between the data. Compared to other Kriging techniques (e.g. the most commonly used Ordinary Kriging or OK), the KED models performed better; this was demonstrated in Chapters 2 and 3. Due to the KED technique, the model output took into account the topographical variability of the seafloor, most often a steering parameter in the prediction of soft substrata habitats. This is not possible with OK.

Examples of geostatistics in the context of marine habitat mapping are rare in literature. Jerosch et al. (2006) used indicator kriging for the mapping of the spatial distribution of benthic communities at the deep-sea submarine Håkon Mosby Mud Volcano, located on the Norwegian–Barents–Svalbard continental margin. Pesch et al. (2008) applied Ordinary Kriging to interpolate abiotic variables, as input for decision trees for the prediction of benthic communities in Germany. Pesch et al. (2008) recommended the use of multivariate geostatistical methods for the prediction of sedimentological maps, as applied in Chapter 3.
Previously, the sedimentological map of the BPNS was limited showing median grain-size with only 3 classes: very fine sand (63-125 µm), fine sand (125-250 µm) and medium to coarse sand (250-2000 µm) (Lanckneus et al. 2001). This map was made using a simple inverse distance interpolation. Other sedimentological maps are often expressed in classes as well. The most common **sediment classification** is that

of **Folk** (Folk 1974), which is based on the relative proportions of gravel, sand and mud. The advantage of this classification system is its broad applicability and use, making it easy to compare sedimentological maps of different areas. This classification was used for the marine landscape mapping of the UK (Golding et al. 2004; and Connor et al. 2006). However, median grain-size is also a commonly available sedimentological parameter and is readily available for wider areas. Ideally, the analysis would be based on raw data acquired using similar methodologies; however, this is usually not possible and various datasets have to be merged to create a sufficiently dense grid of sample data. For new initiatives, the **MESH Standards and Protocols** (Coggan et al. 2007) and the **MESH Recommended Operational Guidelines** (White and Fitzpatrick 2007) for Seabed Habitat Mapping are recommended. A major **disadvantage of a classification system** (such as the Folk classification), is that the resulting output is **not well suited for modelling purposes**. Modelling techniques require, in general, **continuous datasets** (e.g. 0-100 %), with more variation than only a few classes. The sedimentological maps of Chapters 2 and 3 do not have classes, showing a separate value for each pixel. The resulting data grids can be used easily for other modelling initiatives (e.g. the use of the sediment data grid in sediment transport modelling or as input to the prediction of sandwave dimensions).

In **Theme 2**, both **abiotic and biotic datasets** were **integrated** in view of habitat mapping. Various methodologies exist for this integration, ranging from simple cross tabulation to complex statistical habitat modelling methods.
In **Chapter 4**, emphasis was put on the integration of abiotic datasets, to produce an ecologically relevant map of **marine landscapes** on the BPNS. The methodology was based on a combination of **PCA** and **cluster analysis**. It offered a more objective strategy to cluster a diverse range of information. The methodology, worked out in Chapter 4, allowed working with **continuous datasets**, instead of with classified layers (e.g. weak, moderate or strong maximum tidal stress for the UK seas in Connor et al. (2006); photic and non-photic zone for the Baltic Sea in Al-Hamdani and Reker (2007)) and did not oblige the user to make choices about the number and the breaks of the classes or about the maximum **number of input layers**. In Al-Hamdani and Reker (2007), only three abiotic datasets (sedimentology, light zone and salinity) were included to keep the landscape classification ecologically relevant, but also to limit the number of potential combinations to a manageable number. The PCA in Chapter 4 created 6 Principal Components, being linear combinations of the original 16 abiotic datasets. Still, the most difficult issue in marine landscape mapping, is choosing the appropriate number of landscapes. The maximum number of landscapes can be very high (e.g. 5 sediment classes, 2 light zones and 6 salinity classes resulted into 60 marine landscapes in Al-Hamdani and Reker (2007)). However, it is doubtful that all landscapes have an ecological relevance. Chapter 4 proposed a **statistical criterion** (**Calinski-Harabasz index**) to obtain the **optimal number of landscapes**. Also here included was a test whether all of the landscapes were relevant ecologically. This validity was tested, using an **indicator species analysis**, with a biological dataset containing macrobenthic species data (Marine Biology Section, Ugent – Belgium, 2008). A correlation between the defined clusters and the occurrence of the macrobenthic species was shown. Still, it must be clear that it is never the absolute aim of the marine landscape mapping to predict the biology as such; for that purpose other and better predictive modelling techniques exist (Guisan and Zimmermann 2000, for an overview). Marine landscapes give an **indication about the biology**,

based on abiotic datasets only and are most interesting in areas where biological data are scarce or absent.

An interesting marine landscape in Chapter 4 is 'cluster 8' (Figure 4.2), characterized by its **gravel occurrence**. Only 6 biological samples were available in this landscape to validate the ecological relevance. Moreover, the sampling technique (Van Veen grab sample) is not suitable to sample gravel, resulting in a biased ecological validation of this landscape. However, the marine landscape map does provide insight in **possible ecological valuable areas**; areas that may be unknown (or forgotten) and biologically under-sampled today. Van Beneden (1883) and unpublished data from Gilson (Royal Belgian Institute of Natural Sciences, 1899-1910) examined the former co-existence of several important ecosystem functions, associated to forgotten offshore "boulder fields" in the area of the Westhinder sandbank (Houziaux et al. 2007a). For this area, Houziaux et al. (2007a) inferred that high levels of epibenthic species richness and taxonomic breadth, wild beds of the European flat oyster (*Ostrea edulis*), and a spawning ground for the North Sea herring (*Clupea harengus*) seemed to occur in the 19th and at the beginning of the 20th century. It is as well in this area (at the SE side of the Westhinderbank) that the marine landscapes map indicates the presence of gravel fields. Houziaux et al. (2007a) showed that even today a high biodiversity of epibenthos exists in this area (e.g. 70 species of bryozoans at one single sampling station). Although not ground-truthed by the available biological samples or the macrobenthic communities on the BPNS, the marine landscapes map of Chapter 4 shows the presence of a gravel landscape in this area, with a possibly high biodiversity. This is not only the case for the Westhinder area, where Houziaux et al. (2007a) already demonstrated the high biodiversity. Other offshore areas might be valuable as well, as are the swales of some parts of the Hinderbanken, the swale at the SE side of the Oosthinderbank, and the swale between the Goote- and Thorntonbank. In the northernmost part of the BPNS, the patch of 'gravel' consists of shell-rich sand.

The strength of two **statistically-based habitat suitability modelling** (HSM) **techniques** was demonstrated in Chapters 5 and 6.

In **Chapter 5**, HSMs were produced for the 4 macrobenthic communities on the BPNS. The modelling technique used **discriminant function analysis** (DFA), to determine which variables discriminate between two or more naturally occurring groups. A limited number of abiotic datasets was used as input for the model (median grain-size, silt-clay %, depth, slope and distance to coast). The model selected median grain-size and silt-clay % as significant explanatory variables for the modelling of the macrobenthic communities. A three-fold **cross-validation** showed that the agreement between the model predictions and observations was very good and consistent. Moreover, the percentage correctly classified instances (CCI) and the Cohen's kappa index confirmed this agreement.

Important for the resulting models is that a suitable habitat, for a species or community, means that its composing species have the possibility of colonizing the habitat, but may as well be absent because of anthropogenic impacts, such as fisheries, or natural temporal variability. Habitat suitability thus predicts the specific ecological potentials of a habitat rather than the realized ecological structure (Degraer et al. 1999b). The value of DFA for HSMs of marine benthic communities has been demonstrated as well in Shin (1982); Vanaverbeke et al. (2002); and Caeiro et al. (2005).

In **Chapter 6**, another method for habitat suitability modelling was demonstrated. Using **ecological niche factor analysis** or ENFA, the occurrence of the macrobenthic species *Owenia fusiformis* was predicted for the BPNS. This modelling technique had the advantage of its capability to handle an unlimited number of abiotic datasets or ecogeographical variables (EGVs). This number was again reduced by **factor analysis** (ENFA), selecting the ecologically most relevant factors for *O. fusiformis*. Different algorithms and subsets of EGVs were applied, resulting in different HSMs. The best model was retained by cross-validation. Earlier, ENFA had been applied successfully for marine HSMs in the following examples: for deep-water gorgonian corals on the Atlantic and Pacific Continental Margins of North America (Bryan and Metaxas 2007); for squat lobsters in the Macnas Mounds in Porcupine Seabight, SW Ireland (Wilson et al. 2007) and for the northern gannet in Bass Rock, western North Sea in Scotland, UK (Skov et al. 2008). In addition, to showing the value of ENFA, Chapter 6 demonstrated the value of **multi-scale EGVs** as input for the HSMs. The importance of considering variations in spatial scale to predict the occurrence of marine benthic species was demonstrated also in Murray et al. (2002); Ysebaert and Herman (2002); Baptist et al. (2006); and Wilson et al. (2007).

The map of the median grain-size of the BPNS, constructed using KED (Chapter 2), was extended to the **Southern North Sea** (Figure 7.1a), including the southern part of the Dutch continental shelf and the southeastern part of the English continental shelf. This was justified on the basis of a linear relation between the median grain-size and the bathymetry for the whole study area. A map of the silt-clay% using OK, was completed for this area as well (Figure 7.1b). From this, the HSMs of the 4 macrobenthic communities (Chapter 5) could be applied to a wider North Sea region (Figure 7.1c-d-e-f). Although these maps are only preliminary versions, some first remarks can be formulated: 1) the Southern North Sea belongs to the same biogeographical area as the BPNS with comparable sediment type and communities; 2) the shading areas on the 4 HSMs are zones outside the limits of the model; this can be due to the presence of gravel (and as an abiotic variable representing gravel is no input for the model, typically associated fauna can not be predicted); 3) the high suitabilities of the *A. alba* community around the shading areas are modelling artefacts; 4) transborder modelling is usually not straightforward because of different resolutions of input data and different sampling and processing methods; and 5) the absence of the *M. balthica* community around Western Scheldt estuary can be due to possible lower estimations of Dutch silt-clay percentages.
The proposed methodology of Chapter 4 was not tested yet outside of the study area. This is due mainly to the unavailability of standardized abiotic datasets over larger areas. For initiatives in this direction, reference is made to Connor et al. (2006) and Al-Hamdani et al. (2007) who compiled marine landscape maps for the UK seas and the entire Baltic Sea (comprising 7 countries), respectively.

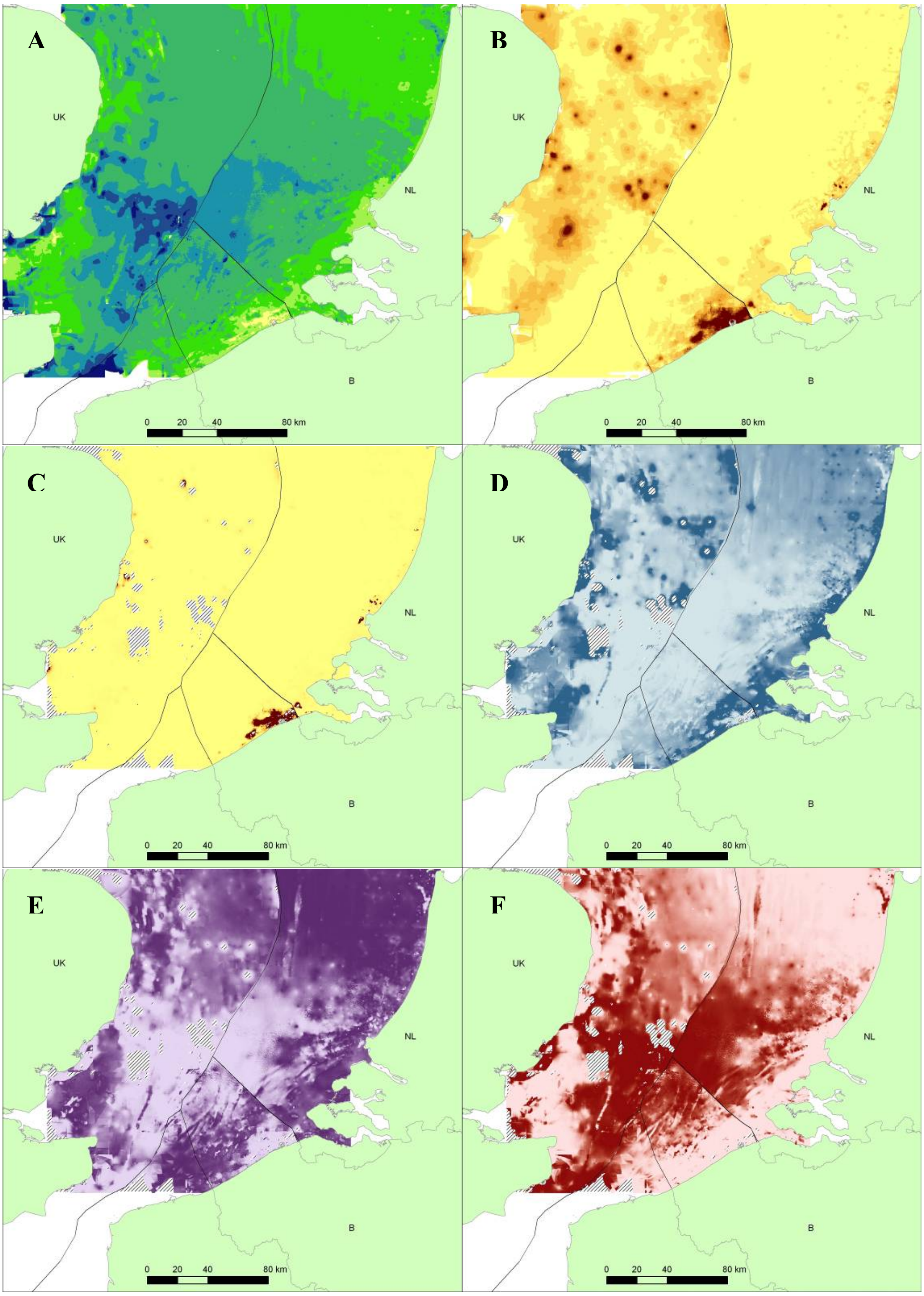


Figure 7.1: Methodologies of Chapters 2-5 applied to the Southern North Sea. A) Median grain-size of the sand fraction; B) Silt-clay %; C) HSM of *Macoma balthica* community; D) HSM of *Abra alba* community; E) HSM of *Nephtys cirrosa* community; and F) HSM of *Ophelia limacina* community.
The maps are orientated to the North. Colour scales range from pale to dark and represent gradients of low to high values (63-2000 µm for A and 0-100% for B, C, D, E and F).

7.2 Reliability of abiotic data coverages and habitat maps

In the context of marine habitat mapping, the **importance of high quality abiotic data** is often overlooked. A high quality abiotic dataset or coverage implies that e.g. the spatial and temporal scale, ground-truth data, interpolation and model techniques are appropriate for the dataset and that they have an ecological relevance for the species under consideration. Habitat modellers tend to have a high confidence in the abiotic data they use, and one step further, managers rely particularly on the habitat maps to make decisions on the allocation of marine protected areas or for spatial planning purposes. However, if the topography, the substrate of the seafloor or the hydrodynamic regime is reconstructed from limited coverage, or if inappropriate scales are being used, the HSM application will likely poorly represent the species' or the community's spatial distribution.
Coverages are usually based on **models** (e.g. bottom current velocities or maximum bottom shear stress), or on **interpolations** of ground-truth samples (e.g. median grain-size, Chapters 2 and 3). Even if complex techniques are used to create high quality EGVs (e.g. multivariate geostatistics, Chapter 2), one has to be aware of the **limitations of the datasets**. In the case of models or interpolations, the reliability of the data is decreasing rapidly, at an increasing distance from the location of the ground-truth samples. This is demonstrated in Chapter 2 (Figure 2.12), where a map of the estimation variance of the kriging analysis gives an indication of the overall reliability of the interpolation. However, it gives more an indication of the sampling density than an absolute measure of reliability of the kriging estimate (Journel 1993; Armstrong 1994; and Goovaerts 1997). Still, this map shows where the data density is sufficient or not to obtain a reliable map of the grain-size. Furthermore, **error propagation**, resulting from the combination of erratic abiotic datasets, can further deteriorate the model. Datasets of **full coverage imagery** (e.g. multibeam, satellite imagery, lidar altimetry), are the **most reliable**, as they contain mostly directly measured data for each pixel of the dataset. When the data are available directly from the imagery (e.g. depth data or mathematically derived topographical coverages from multibeam), the reliability can be considered as high. Still, interpreted data from full coverage imagery (e.g. acoustic seabed classification, derived from multibeam backscatter in Van Lancker et al. 2007), will always include a measure of uncertainty. Finally, if relationships exist between coverages and species/communities, but if the coverages are unavailable (e.g. because of high costs) or the relationships are absolutely unknown, any habitat maps will be a flawed reflection of the actual situation.

Foster-Smith et al. (2007b) describe how the accuracy and confidence of marine habitat maps can be assessed. The **accuracy** of a habitat map is a mathematical measure of the **predictive power of a map** to predict correctly the habitat for a particular point (or pixel). By overlaying ground-truth data on the predicted habitat map, an error matrix can be made, by counting 'hits' and 'misses' and calculating accuracy indices (Foster-Smith et al. 2007b, for an overview). **Confidence** is more subjective than accuracy; it is an assessment of the **reliability of a map given its purpose**. Foster-Smith et al. (2007b) proposed a **multi-criteria approach** to determine the confidence one can have in habitat maps. This approach was implemented in the **MESH Confidence Tool**. Three main questions are posed:

- How good is the **ground-truthing**?
- How good is the **remote sensing**?
- How good is the **interpretation**?

The 3 groups correspond to the 3 first steps of the habitat mapping scheme (Figure 1.2), where remote sensing data correspond to coverage datasets. For each group, a number of questions are posed, resulting in a group score, expressed as a percentage. Questions are related to the techniques (appropriate or not for the habitat), coverage (e.g. tracklines spread widely or full-coverage), positioning (e.g. differential GPS or chart based navigation), standards applied (internal or internationally agreed), age of datasets, interpretation (e.g. is the interpretation method documented or not), level of detail of the resulting map and the map accuracy.
The tool is intended to evaluate habitat maps, and preferably for habitat classifications based on acoustical data and validated with biological samples. Still, for the marine landscape map and the HSMs of Chapters 4, 5 and 6, respectively, the tool can be applied as well (Table 7.1).

Table 7.1: Confidence assessment of the habitat maps of Chapters 4, 5 and 6 using the MESH Confidence Tool (Foster-Smith et al. 2007b). Scores are expressed as percentages (GT = ground-truthing; RS = remote sensing; IN = interpretation; BPNS = Belgian part of the North Sea; HSM = habitat suitability model; *O. fusiformis* = *Owenia fusiformis*).

Chapter	Description	GT score	RS score	IN score	Overall score
4	Marine landscapes (BPNS)	65	73	58	**66**
5	HSMs macrobenthic communities (BPNS)	65	73	92	**77**
6	HSM *O. fusiformis* (BPNS)	65	73	92	**77**

The marine landscape map has, relatively to the other habitat maps, the worst score of confidence (66 %). This is due mainly to the limited availability of samples to validate the marine landscapes (called ‘interpretation’ by the Confidence Tool). The HSMs of both the macrobenthic communities and *O. fusiformis* on the BPNS have the same score (77 %), because a comparable methodology on the same spatial scale was applied, integrating biological and abiotic data from the beginning. Still, the HSM of *O. fusiformis* is based on more abiotic variables than the HSM of the *A. alba* community (respectively 9 abiotic variables, representing the topographical nature; and 2 abiotic variables, representing the sedimentology only). As such, the first model can be considered as more reliable, but this is not accounted for in the Confidence Tool.
As demonstrated for the HSMs of *O. fusiformis* and the *A. alba* community, these scores have to be interpreted carefully, because of possible subjective assignments of the scores or of criteria that were not considered.

7.3 Comparison of habitat maps on the BPNS

7.3.1 Comparison of marine landscapes with habitat suitability models on the BPNS

The sedimentological maps of Chapters 2 and 3 and the habitat maps of Chapters 4, 5 and 6 are not fully comparable in the sense that the former were used as input for the latter. Still, the marine landscape map and the HSMs of both the macrobenthic communities and of *O. fusiformis* on the BPNS, can be overlain and this result can be evaluated.

Boxplots comparing the marine landscapes (Chapter 4, Figure 4.2) and the HSM of *O. fusiformis* (Chapter 6, Figure 6.3), show a high suitability for this species of landscapes 1 and 2, whereas landscapes 3, 4 and 8 are less suitable (Figure 7.2). The remaining landscapes are unsuitable.

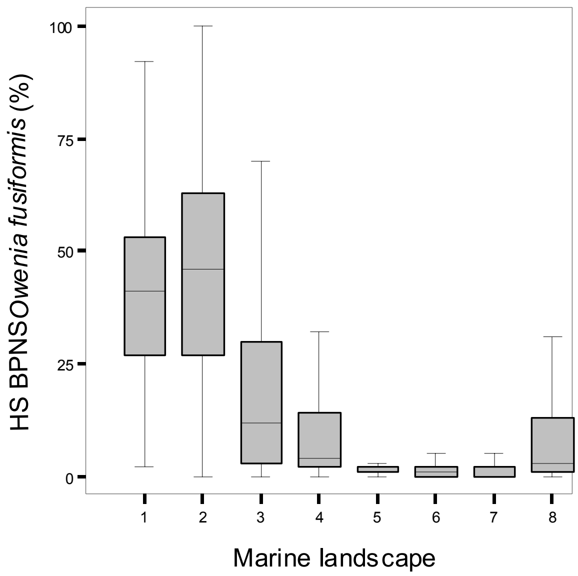


Figure 7.2: Boxplots of HSM of *Owenia fusiformis* on the BPNS, overlain on the marine landscape map, showing that landscapes 1 and 2 have the highest suitability for *O. fusiformis*. The middle line in the box is the median, the lower and upper box boundaries mark the first and third quantile. The whiskers are the vertical lines ending in horizontal lines at the largest and smallest observed values that are not statistical outliers (values more than 1.5 interquantile range).

Landscapes 1 and 2 correspond to "shallow, high silt-clay percentage, high current velocity, high bottom shear stress, turbid, high Chl a concentration" and "shallow NW orientated flats and depressions, fine sand, slightly turbid, high Chl a concentration", respectively. This corroborates the habitat preferences of *O. fusiformis* as presented in Chapter 6, although a moderately high silt-clay %, a moderate maximum bottom shear stress and moderate maximum current velocities were now added as well. Landscapes 3, 4 and 8, respectively, correspond with "shallow SE orientated sandbanks, fine to medium sand, slightly turbid, high Chl a concentration", "deep NW orientated flats and depressions, medium sand" and "high percentage of gravel – shell fragments".

When the same exercise is done for the habitat suitabilities of the macrobenthic communities (Chapter 5, Figure 5.3), overlain on the marine landscapes (Figure 7.3), the following observations can be made:
(1) The *Macoma balthica* community has a pronounced preference for landscape 1; this is logical, as this species was known already to occur mainly in fine sediments with a high silt-clay %; all other communities show much more variation;
(2) The *A. alba* community shows highest suitabilities in landscapes 2 and 3, corresponding more or less with the preference of *O. fusiformis* (Figure 7.2); still, the *A. alba* community shows more variation in the other landscapes;
(3) The *Nephtys cirrosa* community has similar preferences for all of the landscapes, except for landscape 1, meaning that this community has a very broad niche and that its habitat is not at all well defined; and
(4) The *Ophelia limacina* community has a preference for the combination of landscape 4 ("Deep NW orientated flats and depressions, medium sand"), 5 ("Deep SE orientated flats and depressions, medium sand"), 6 ("Crests of sandbanks, medium sand"), 7 ("Slopes of sandbanks, medium sand") and 8 ("High percentage of gravel – shell fragments"). Still, the preferences are very well pronounced.

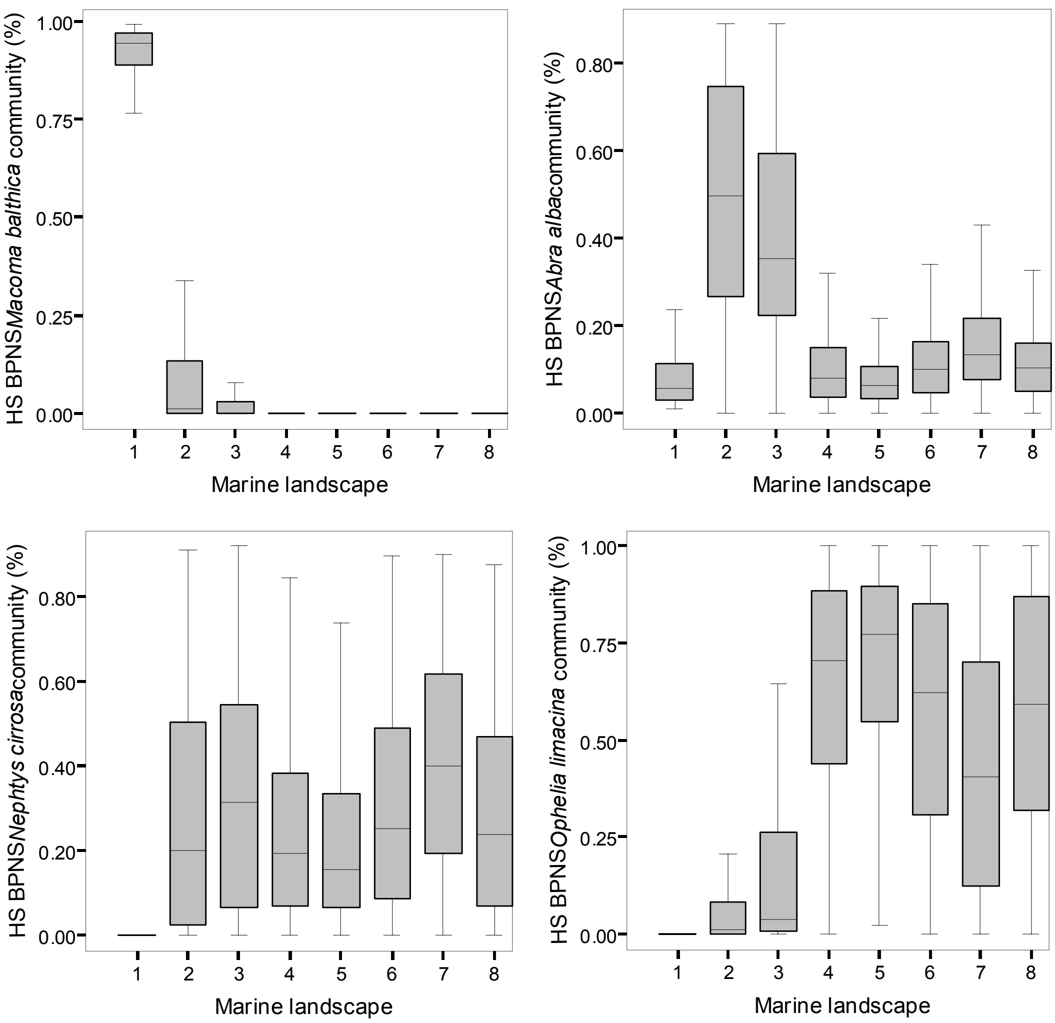


Figure 7.3: Boxplots of HSM of the 4 macrobenthic communities on the BPNS, overlain on the marine landscape map.

As such, there is a discrepancy between the number of macrobenthic communities (4) and the number of marine landscapes (8). However, as only 4 communities have been defined on the BPNS, it is possible that more communities or gradations between communities are present. As discussed in Section 7.1, marine landscape mapping can reveal other ecologically interesting areas that are still unknown today or that are difficult to sample (e.g. landscape 8, representing a 'High percentage of gravel – shell fragments').

7.4 Habitat maps to support a sustainable management of the marine environment

Generally, habitat maps are produced by scientists. Their translation towards simple maps, of use for all stakeholders, is often a difficult task. Still, habitat maps are a crucial tool in the context of a **sustainable management of the marine environment**. In particular, they are useful for:

- the designation of **Marine Protected Areas** (MPAs) (e.g. improved knowledge on valuable habitats with a high biodiversity);
- setting-up **baselines** for future impacts of new anthropogenic activities (e.g. windmill parks);
- **marine spatial planning** (e.g. where to avoid certain activities);
- setting-up new **monitoring studies** (e.g. indication of possible interesting habitats);
- **informing the stakeholders** (e.g. fishermen, aggregate extraction industry, coastal managers, inhabitants of coastal villages…);
- the implementation of **European Directives** (e.g. new proposals for Habitats Directive areas).

For stakeholders, it can be difficult to interpret habitat maps correctly. Generally, they have a preference for classified habitat maps with clear boundaries between valuable and less valuable areas, although this does not correspond with the real situation. It should be the task of scientists to educate stakeholders how to translate or interpret the results, whether or not classified habitat maps or maps with gradual scales are available.
A first attempt was made to translate the 4 HSMs of the macrobenthic communities (Chapter 5) to a EUNIS classification (Schelfaut et al. 2007), being a pan-European classification system of habitats (EUNIS 2002). However, only 1 of the classes corresponded with an existing EUNIS class, while 3 classes could not be correlated. This is the case for many countries (e.g. France) and pleads for an overall updating of the EUNIS classification system (Foster-Smith et al. 2007a).

As described in the legal characterization of the BPNS (Section 1.2.6), there are **5** legally designated Belgian **MPAs**: 3 Special Protection Areas (SPAs), in the framework of the Birds Directive; 1 Special Area of Conservation (SAC), in the framework of the Habitats Directive; and 1 specific marine nature reserve (Figure 1.12). Initially a second SAC was designated on the Vlakte van de Raan (SAC2), but this has been cancelled in February 2008, because of a complaint of the energy company Electrabel.
The designation of the SPAs on the BPNS, is based on selected important bird areas (Haelters et al. 2004). The designation of SAC1 and the former SAC2 is based on the

consultation of experts on different ecosystem components (macro- and epibenthos, fish, seabirds and sea mammals) and the investigation of habitat data (Derous et al. subm. b). The main issue remains the estimation of the biological value of certain areas which is based ideally on all marine ecosystem components and taking into account a number of criteria for their valuation (Derous et al. 2007). For the BPNS, Derous et al. (subm. a) proposed a **biological valuation map** (BVM), based on datasets of macro-, hyper- and epibenthos, demersal fish and avifauna. The HSMs of the 4 macrobenthic communities (Chapter 5; Figure 5.3), served as a major input for the macrobenthos of the BVM, next to point data on densities and species richness. Derous et al. (subm. b) compared the high valuable zones of the Belgian BVM (Derous et al. subm. a) with the Belgian SACs and SPAs of the European Directives and found a good overlap between both, although some valuable areas of the BVM, were not covered by the SACs and SPAs. Derous et al. (subm. b) recommends proposing other priority areas under the Habitats Directive: e.g. the area at the northern side of the former SAC2 (Vlakte van de Raan) (Introduction; Figure 1.12) or the sandbank complex of the Thorntonbank (Introduction; Figure 1.3).

The HSM of *O. fusiformis* on the BPNS (Chapter 6; Figure 6.3), being an important indicator species of the ecologically important *A. alba* community (Van Hoey 2004), confirms the importance of the area at the **northern side of the Vlakte van de Raan** as a **possible priority area**.

The marine landscapes map (Chapter 4; Figure 4.2) shows a clear distinction between the coastal and the more offshore area (with landscape 2 and 3 as most interesting for the *A. alba* community). Moreover, sandbanks, swales, slopes and gravel areas are distinguished, indicating possible interesting habitats that should be examined to discover their ecological relevance. Especially the potential gravel habitat should be explored further (see Section 7.1). **Possible priority areas** could include the area in the swale at the **SE side of the Westhinder sandbank** or the coarser sediment patch in the swale at the **SE side of the Noordhinder sandbank** (Chapter 4; Figure 4.2; cluster 8). Another potential area is located in the swale between the Goote- and Thorntonbank. In Van Lancker et al. (2007), the distribution of possible gravel areas (Introduction; Figure 1.8), was even more extensive than on the marine landscapes map. This is due to the fact that the gravel input layer of the marine landscapes map was based purely on quantitative sample data, whilst the potential gravel area of Van Lancker et al. (2007) is based additionally on geological, multibeam backscatter and diving information. Based on Van Lancker et al. (2007), gravel is expected in the **entire swale system of the southern part of the Hinder Banks**. However, it must be stressed that the gravel on the BPNS is not a continuous layer; it occurs in patches mainly and in many cases the gravel is covered with a sand veneer (Van Lancker et al. 2007). Especially in areas where sand dynamics are high, species richness and diversity decreases and mostly only the *O. limacina* community might be present (Van Lancker et al. 2007). This community is indeed typical for medium to coarse sediments and is often associated with gravel and shell fragments. In Figure 7.4, the **HSM of the *A. alba* community** is combined with the potential gravel areas of the marine landscapes map and of Van Lancker et al. (2007). As the HSM of the *A. alba* community is similar to that of *O. fusiformis*, only the former is presented on the map to avoid confusion. Moreover, *O. fusiformis* is an important indicator species of the *A. alba* community (Van Hoey et al. 2004). The whole zone of high suitability of the *A. alba* community (being the dark orange zone), combined with the potential gravel areas can be considered as potentially valuable seabed habitats.

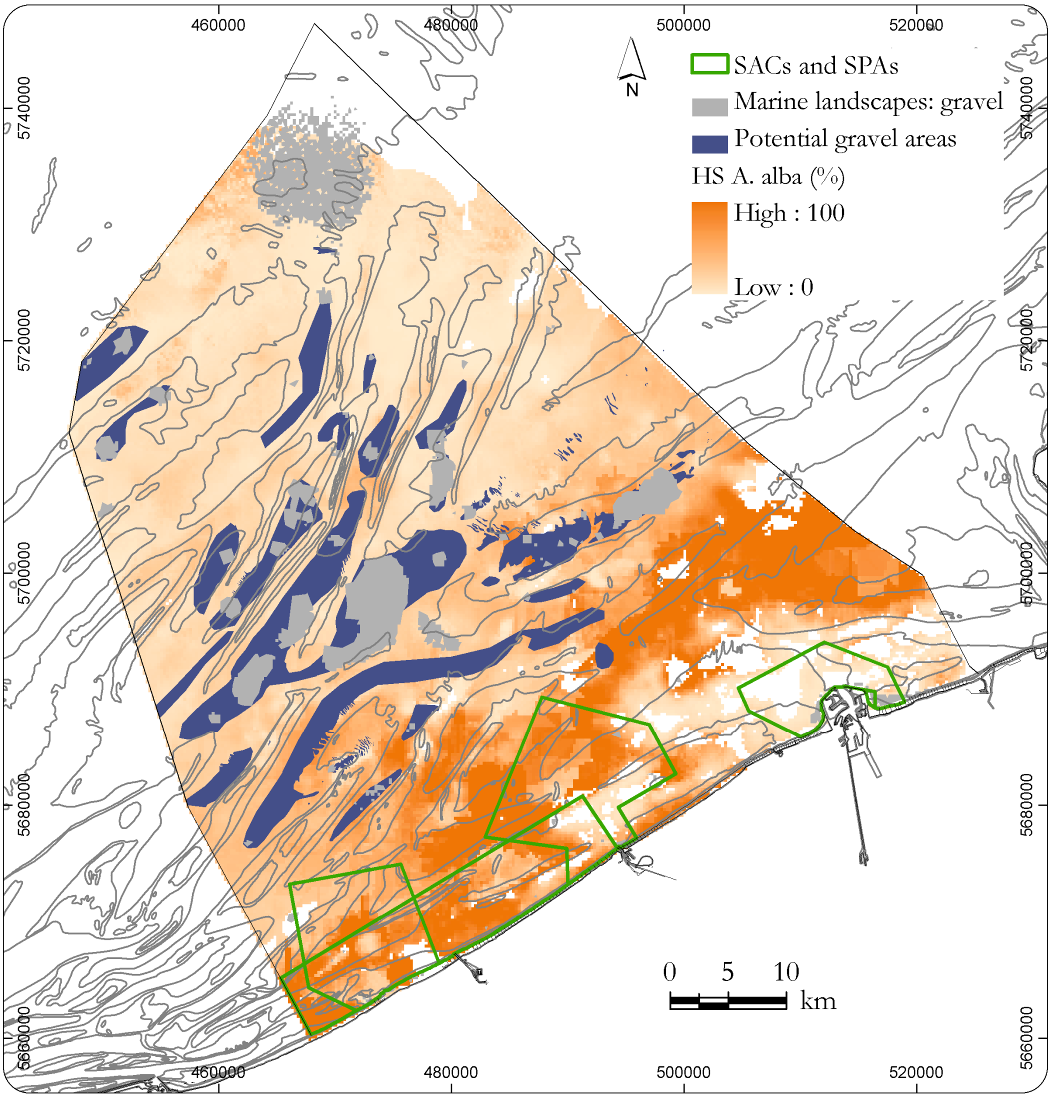


Figure 7.4: Combination of the habitat suitability model of the *A. alba* community, gravel patches from the marine landscapes map and potential gravel areas from Van Lancker et al. (2007).
Based on this map, the whole dark orange area, with the possible gravel areas are potentially valuable seabed habitats. SACs = Special Areas of Conservation; SPAs = Special Protection Areas; HS = habitat suitability; *A. alba* = *Abra alba*.

However, nature conservation is not the only 'user' of the BPNS; as such the **anthropogenic activities** on the BPNS are overlain on the same datasets (Figure 7.5). If all anthropogenic activities are left unchanged, **little space** is left **for nature conservation**. Still, it must be clear, that this space is not constantly occupied (e.g. the use intensity of military exercises is for most of the designated zones less than 11 exercise days/year/km² (Maes et al. 2005)) and multiple occupation of already designated zones could be considered.

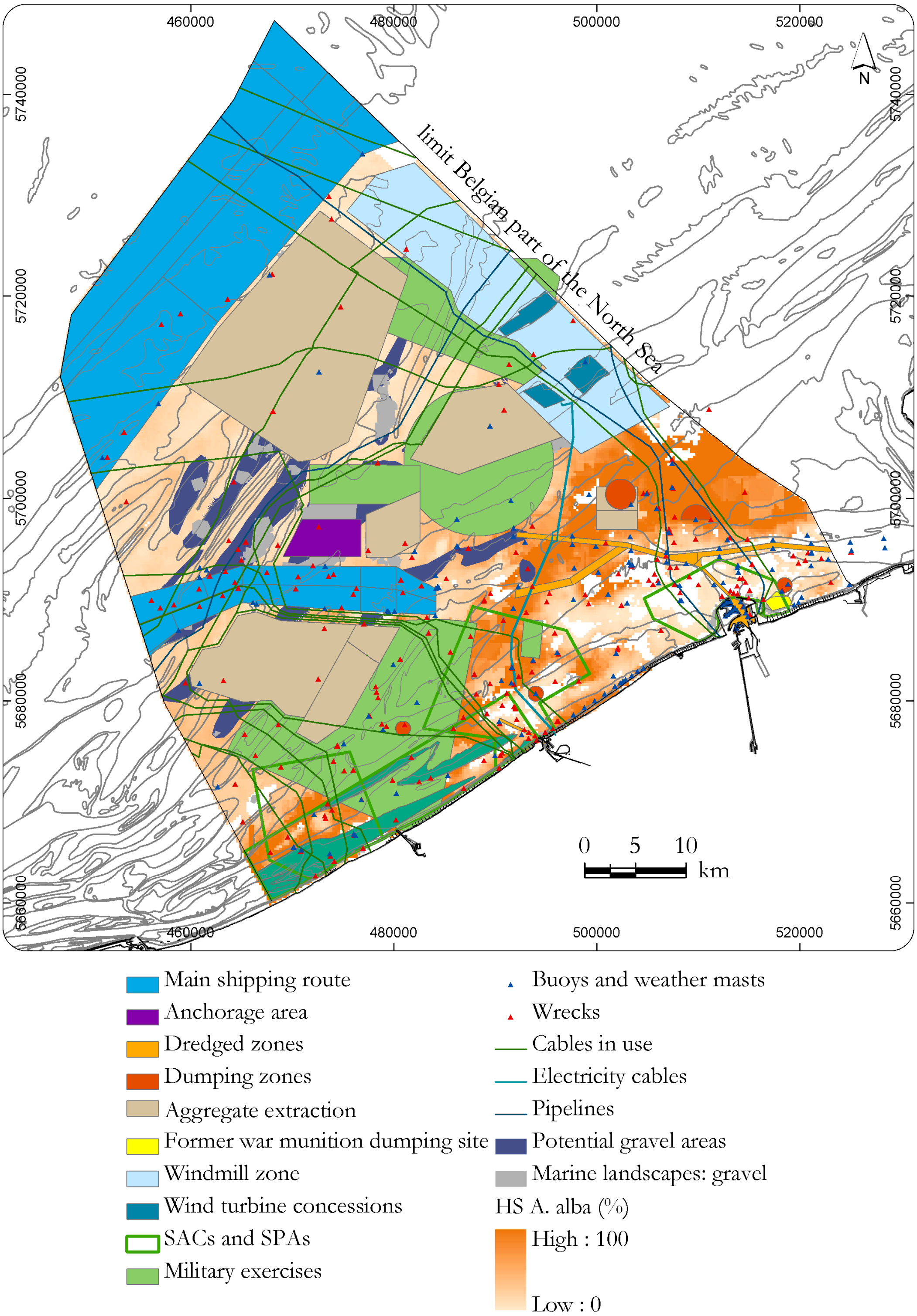


Figure 7.5: Overlay of anthropogenic activities on the habitat suitability model of the *A. alba* community, gravel patches from the marine landscapes map and potential gravel areas from Van Lancker et al. (2007). SACs = Special Areas of Conservation; SPAs = Special Protection Areas; HS = habitat suitability; *A. alba* = *Abra alba*.

In Figure 7.6, all areas of highest suitabilities of the *A. alba* community and possible gravel areas, that are not overlapping with anthropogenic activities (except for cables, pipelines, wrecks, buoys and weather masts), are marked as potentially valuable seabed habitats. For zones 1 until 3, the ecological relevance has been shown in Chapter 5 for the *A. alba* community and in Chapter 6 for *O. fusiformis* (although the highest suitability zones are somewhat less extensive for *O. fusiformis* than for the *A. alba* community). The areas 4 until 11 are based purely on abiotic knowledge from the marine landscapes (Chapter 4) and the potential gravel patches from Van Lancker et al. (2007). Still, comparing these areas with the species richness data from Houziaux et al. (2007b), it is observed that areas 5, 9, 10 and 11 coincide well with their zones of largest taxonomic breadth (Figure 7.7).

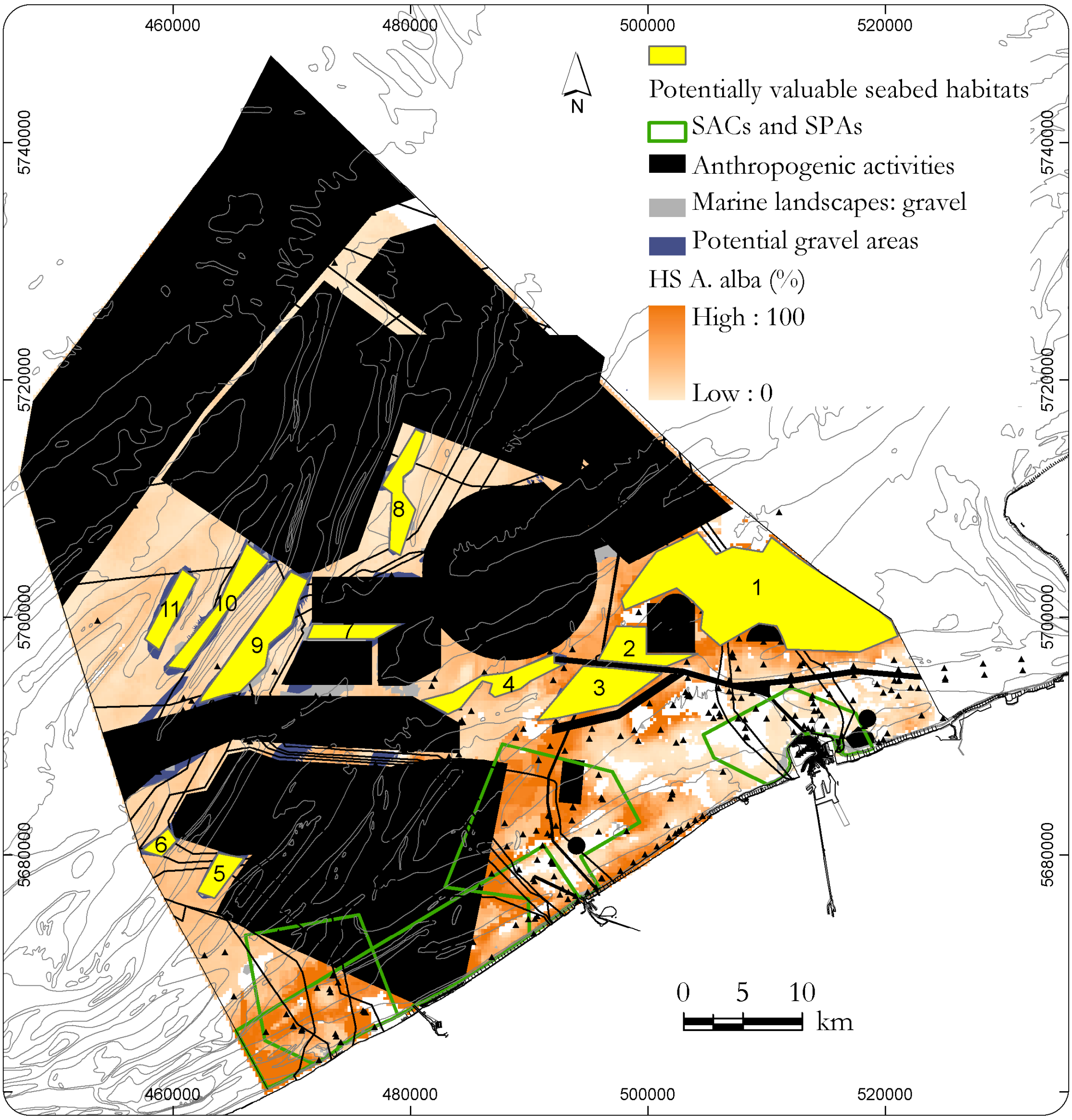


Figure 7.6: Potentially valuable seabed habitats, not overlapping with anthropogenic activities and in addition to the already designated SACs and SPAs. The zones correspond to the highest suitabilities of the *A. alba* community (zone 1, 2 and 3) and the gravel patches from the marine landscapes map and from Van Lancker et al. (2007) (zone 4, 5, 6, 7, 8, 9, 10 and 11). SACs = Special Areas of Conservation; SPAs = Special Protection Areas; HS = habitat suitability; *A. alba* = *Abra alba*.

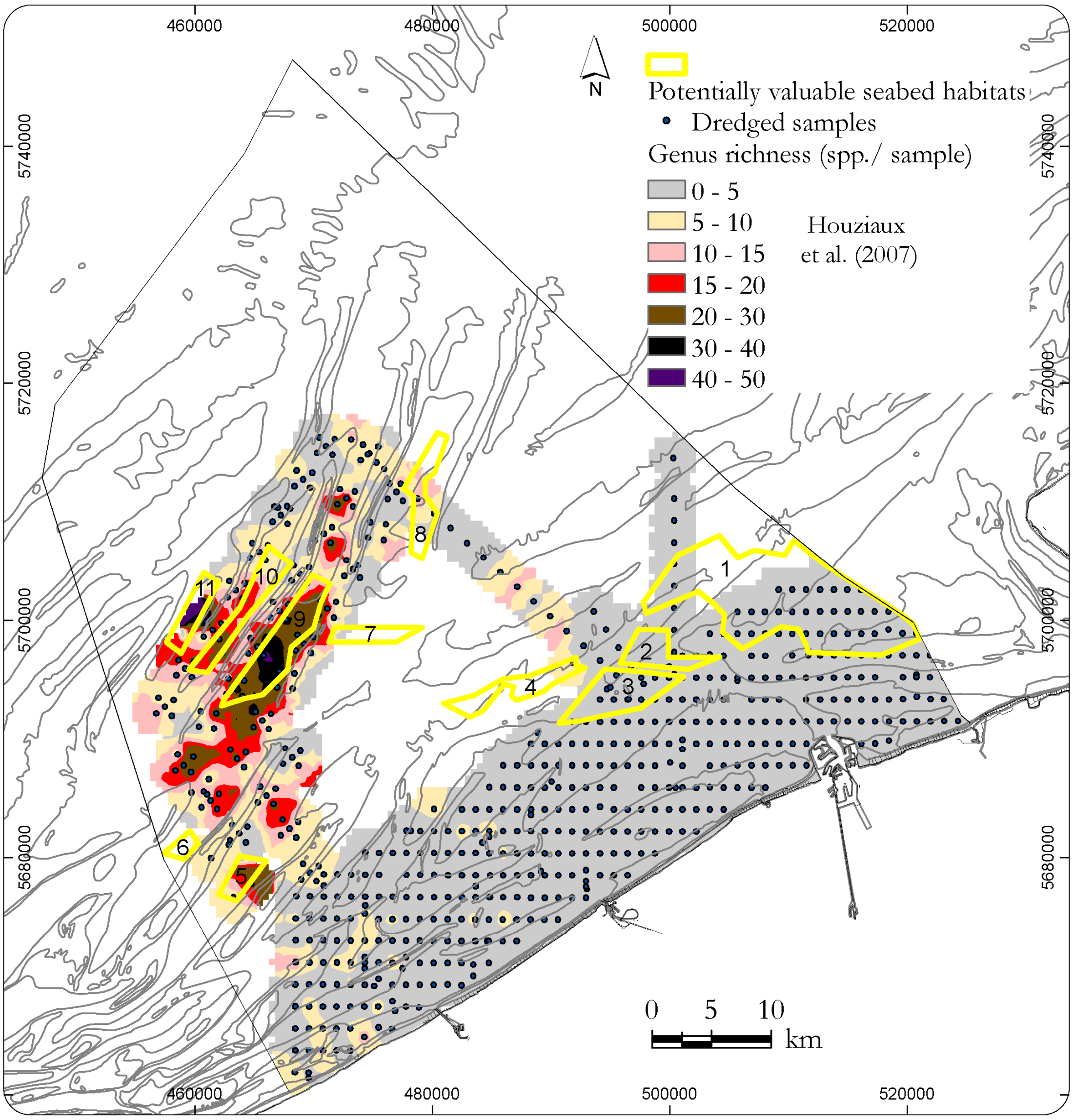


Figure 7.7: Potentially valuable seabed habitats overlain on the species richness map on genus level of epibenthos (excluding infauna), based on historical datasets of Gilson from the beginning of the 20th century (Houziaux et al. 2007b).

Table 7.2 is a summary of the potentially valuable seabed habitats, with their location on the BPNS, their ecological relevance (based on macrobenthos HSMs from Chapter 5 and 6; and epibenthos data excluding infauna, from Houziaux et al. (2007b)) together with possible conflicts with anthropogenic activities. The ecological relevance of avifauna or sea mammals is not taken into account.

Haelters et al. (2007) proposed an OSPAR[19] MPA, called the 'Westhinder MPA'. For this delimitation, historical information about the benthic biodiversity and the extent of former oyster beds in the 19th century was used (Lanszweert 1868; and Houziaux et al. 2007b), together with recent information on biodiversity (Houziaux et al. 2007b) and the presence of gravel (Van Lancker et al. 2007). The area follows the southern slope of the Westhinder sandbank and the northern slope of the Oostdijck sandbank. For the northeastern and western delimitation, the position of the Westhinder anchorage area and the position of some buoys has been used (Haelters et al. 2007).

Figure 7.8 represents the valuable seabed habitats as a result from this research together with their ecological relevance. From this, there are 5 zones with a high ecological relevance: zone 1, 2, 3, 9 and 11.
As both zones 2 and 3 are situated in a highly important area for navigation, they are less suitable. For zone 1, the biodiversity of the **macrobenthos** is high to very high, as also for birds (Derous et al. subm. a). The delineation of zone 2 and 3 is based on the results of a **predictive model** of habitat suitability of the *A. alba* community. With any model (also demonstrated in 7.2), it is clear that the predictions need sound validation in terms of ground-truthing.
Regarding the **species richness of epibenthos**, **zones 9 and 11** seem most valuable. Zone 9 corresponds mostly with the proposed 'Westhinder MPA' of Haelters et al. (2007).

Table 7.2: Potentially valuable seabed habitats, with their location; ecological relevance (limited to macrobenthos), with a scale ranging from low to high (+ to +++) with reference; and possible conflicts with anthropogenic activities (anthropogenic activities based on Maes et al. 2005).

Zone	Location	Ecological relevance		Possible conflicts
1	N & central part VVR, E end AB	+++	Chapter 5, 6	Fisheries
2	NW part VVR, SE part AB	+++	Chapter 5, 6	Fisheries
3	S of central AB	+++	Chapter 5, 6	Shipping, fisheries
4	W & central part AB	?	/	Shipping, fisheries
5	Swale OD - BR	++	Houziaux et al. 2007b	/
6	Swale OD - BeB	?	/	/
7	Swale OH - GB	?	/	Fisheries (minor part)
8	Swale BB - OH	+	Houziaux et al. 2007b	Fisheries (partly)
9	Swale OH - WH	+++	Houziaux et al. 2007b	Fisheries (partly)
10	Swale WH - NH	++	Houziaux et al. 2007b	Fisheries (partly)
11	Swale NH - FB	+++	Houziaux et al. 2007b	Fisheries (partly)

VVR = Vlakte van de Raan; AB = Akkaertbank; OD = Oostdijck; BR = Buiten Ratel; BeB = Berguesbank; OH = Oosthinder; GB = Gootebank; BB = Bligh Bank; WH = Westhinder; NH = Noordhinder; FB = Fairybank; ? = unknown ecological relevance.

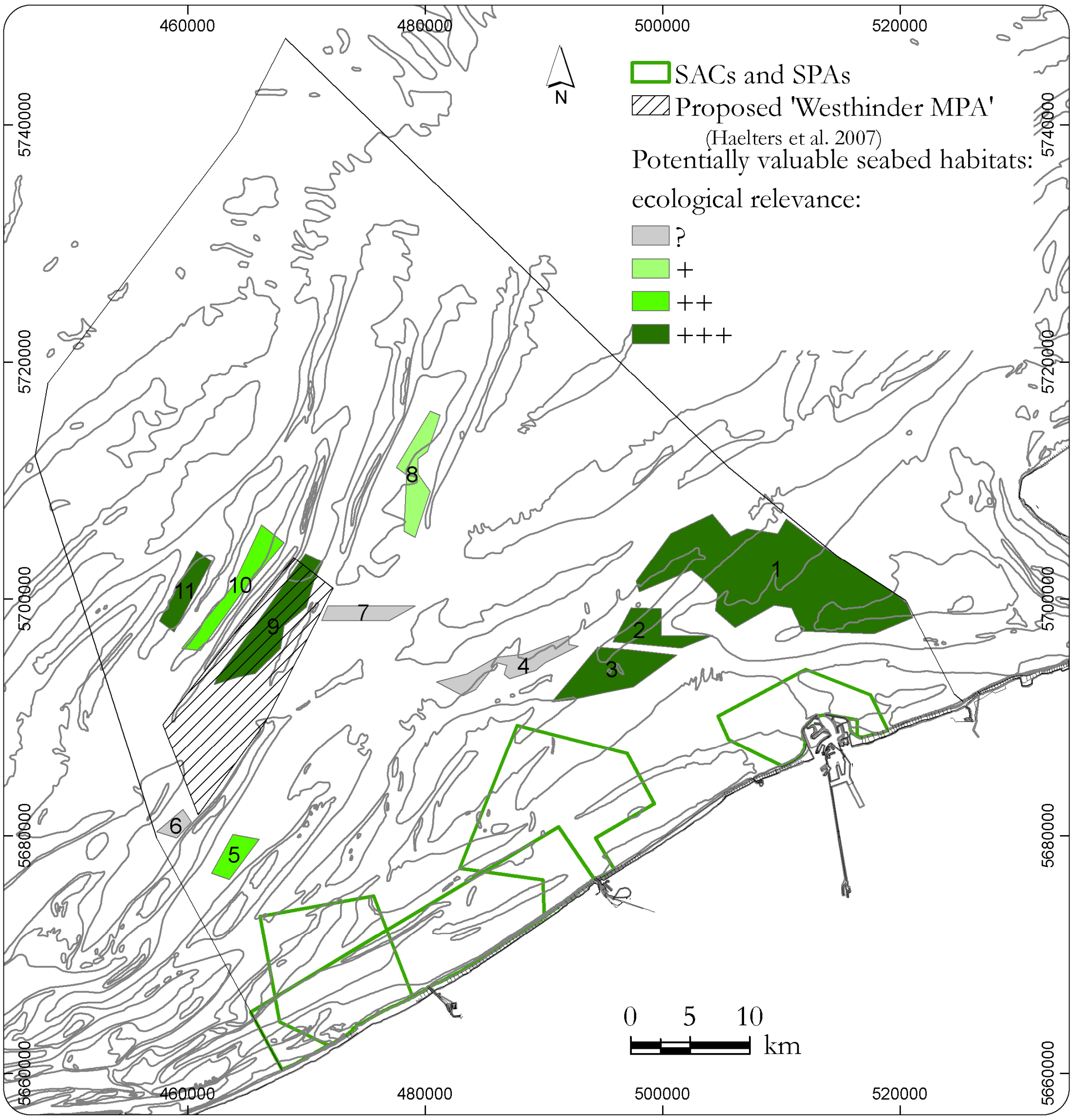


Figure 7.8: Potentially valuable seabed habitats, with their ecological relevance (scale ranges from unknown (?) to high (+++)) (Table 7.2); existing Special Areas of Conservation (SACs) and Special Protection Areas (SPAs); and proposed 'Westhinder MPA' of Haelters et al. (2007).

Finally, it is hoped that this research will contribute to the future implementation of the **EU Marine Strategy Directive**. This Directive aims at both the implementation of an ecosystem-based approach in marine waters and a sustainable use of marine goods and services (CEC 2005). According to Annex II of this Directive, as many biological, physical and chemical characteristics should be included in the assessment of the environmental status of the sea. Habitat maps, at various spatial and temporal scales, and the integration of various datasets will likely be needed to improve management practices. In addition, the techniques and approaches, presented in this thesis, can assist in the prediction of ecologically valuable areas in the marine environment.

Chapter 8

Conclusion

8 Conclusion

As formulated in the research objectives, the general objective of this thesis was to develop **spatial distribution models** as **input for habitat mapping**. Results focussed on the creation of: **abiotic coverages** and **habitat maps**, the latter being the result of the **integration of ground-truth data and the abiotic coverages**. The methodologies were set-up, following an **objective** and **(geo)statistically sound** approach, with emphasis on **intercomparison and validation.**

More specific research objectives were the following:

ii. Proposing a **new approach for marine landscape mapping**, being simple, statistically sound and easy applicable to other regions. The proposed methodology should be a step forward in standardising marine landscape mapping throughout Europe.

iii. Developing methodologies that are straightforward, **objective** and **statistically sound**.

iv. Using a **maximum input of ecogeographical variables** (EGVs), for the modelling of the marine landscapes and habitat suitability modelling (HSM) and as secondary variables for multivariate geostatistics.

v. **Optimising the reliability of EGVs** by applying multivariate geostatistics for the sedimentological maps and **optimising the reliability of habitat maps** by applying statistically sound methodologies; **validating** the results obtained by the (geo)statistical methodologies.

vi. Deriving **multi-scale topographical EGVs** from the bathymetry, for both the modelling of sedimentological maps and for the modelling of the HSMs of the species *Owenia fusiformis*.

As explained in the Introduction, the habitat mapping process can be subdivided into 4 steps:

1) Getting the best out of ground-truth data;
2) Selecting and getting the best out of coverage data;
3) Integration of ground-truth and coverage data;
4) Habitat map design and layout.

Main emphasis of the research was on Step 2 and Step 3 of the habitat mapping process; as such the thesis was subdivided into 2 themes, corresponding with:

1) Best coverages for habitat mapping;
2) Integration of datasets in the view of habitat mapping.

<u>Theme 1: Best coverage data for habitat mapping</u>

In the first theme, emphasis was put upon the **application of multivariate geostatistical techniques**; these are used increasingly in certain research domains (e.g. soil science and climatology), but are still unexplored in marine science. To obtain reliable sedimentological maps of scattered ground-truth data, it is crucial to apply the most suitable interpolation techniques. If there is a **linear relation** between the **sedimentological variable** and one or more **secondary variables**, multivariate geostatistics can be applied. **Kriging with an external drift** or KED (Goovaerts 1997) is an example of such a technique.

KED was applied successfully on sedimentological data of 2 study areas: 1) the Belgian part of the North Sea (BPNS) in Chapter 2; and 2) a small study area at the southeastern part of the BPNS in Chapter 3. For the first study area, the median grain-size of the sand fraction (d_s50) was interpolated with only one secondary variable, the bathymetry. For the second study area, a whole set of secondary variables was used for the interpolation of the d_s10 (10^{th} percentile of the sand fraction), d_s50, d_s90 (90^{th} percentile of the sand fraction) and silt-clay% (fraction below 63 µm). These secondary variables comprised the bathymetry and its multi-scale derivatives (e.g. slope, curvature, fractal dimension on different spatial scales).

Both results were **compared** and **validated** against results obtained by linear regression (Chapter 2) and Ordinary Kriging (OK) (Chapter 2 and 3). Validation indices revealed that all results, based on KED, gave significantly better results (except for the map of the silt-clay%, which correlated very weakly with the secondary variables).

The result for both study areas were **highly reliable sedimentological maps**, reflecting well the natural morphology of the seabed. Moreover, the followed methodology was **objective**, **straightforward** and **scientifically sound**.

Theme 2: Integration of datasets in the view of habitat mapping

The second part of the thesis comprised the **integration of abiotic datasets with biological data.** In this theme, the main aim was to apply **statistical methods** to obtain **highly reliable habitat maps**.

In Chapter 4, marine landscapes were modelled, by integrating a whole set of abiotic variables. Therefore, **principal components analysis** (PCA) was used to reduce the number of abiotic variables to a set of non-correlating components. Moreover, a **cluster analysis** was used to cluster these components to meaningful marine landscapes. To determine the number of landscapes, the Calinski and Harabasz (1974) index was used, designated by Milligan and Cooper (1985) as giving the best results out of a number of indices. Only at the end of the process, biological data were used to test the ecological meaning of the marine landscapes. This was confirmed for the marine landscapes of the BPNS.

HSMs were produced in Chapter 5 and 6. Therefore, two different statistical modelling techniques were applied: **Discriminant Function Analysis** (DFA) and **Ecological Niche Factor Analysis** (ENFA).

DFA was applied in Chapter 5 for the modelling of the 4 macrobenthic communities on the BPNS (Van Hoey et al. 2004). Moreover, a three-fold **cross-validation** showed that the agreement between the model predictions and observations was very good and consistent.

In Chapter 6, ENFA was applied for the modelling of the species *Owenia fusiformis*, on the BPNS. As this species is an important indicator species of the ecologically important *A. alba* community (Van Hoey et al. 2004), the HSM of the entire BPNS corresponded, in essence, with the HSM of the *A. alba* community, although subtle differences were observed. A cross-validation was used to select the best model out of a set of models, based on different modelling algorithms and different subsets of EGVs.

In the Discussion (Chapter 7), results of this research were compared, overlain and integrated.
The **reliability of the habitat maps** was calculated following a multi-criteria approach (Foster-Smith et al. 2007b). The reliability of all of the habitat maps (Chapter 4, 5 and 6) ranged between 66 and 77 %, with the lowest score for the marine landscapes map and the higher scores for the HSMs on the scale of the BPNS.
The methodologies of Chapter 2 (multivariate geostatistics) and Chapter 5 (Discriminant Function Analysis) were tested successfully on the **Southern North Sea** to obtain maps of the median grain-size, silt-clay% and HSMs of the 4 macrobenthic communities.

Based on the HSM of the ecologically important *A. alba* community (Chapter 5), combined with the potential gravel areas from the marine landscapes (Chapter 4) and from Van Lancker et al. (2007), propositions were made for the **designation of potentially valuable seabed habitats** on the BPNS.

Appendix

9 Appendix

In the Appendix, a short overview is given of the theoretical background of the statistical techniques used throughout Chapter 3, 4, 5 and 6. Geostatistical techniques have been explained in the Material and Methods section of Chapter 2 and 3 and are thus omitted in this Appendix.
Following techniques are summarized:

- Ordination techniques:
  - Principal Components Analysis (Chapter 3 and 4);
  - Discriminant Function Analysis (Chapter 5);
  - Ecological Niche Factor Analysis (Chapter 6);
- Clustering techniques:
  - Ward's minimum variance method (Chapter 4);
  - *K*-means partitioning (Chapter 4);
  - Indicator species analysis (Chapter 4);
- Multi-scale terrain analysis (Chapter 3 and 6).

As the book 'Numerical Ecology' of Legendre and Legendre (1998) provides a very good overview of mathematical methods in the context of ecological applications, the theory in this Appendix is mainly based on this reference (except Ecological Niche Factor Analysis, based on Hirzel et al. 2002a and multi-scale terrain analysis, based on Wilson et al. 2007). Other specific references are cited in the text.

9.1 Ordination techniques

In the context of multivariate statistics, ordination comes from ecology where it refers to the representation of objects (e.g. sites, stations, variables) as points along one or several axes of reference (Gower 1984). Ordination in reduced space is also called *factor analysis*, since it is based on the extraction of eigenvectors or factors of the association matrix.

9.1.1 Principal Components Analysis

Principal Components Analysis (PCA) can be performed in multivariate analysis for data reduction and to avoid multicollinearity (i.e. high degree of linear correlation) of a set of variables. It computes a reduced set of new, linearly independent variables, called principal components that account for most of the variance of the original variables. The principal components are a linear combination of the original variables. A maximum of p principal axes may be derived from a data table containing p variables. The principal axes of a dispersion matrix **S** are found by solving:

$$(\mathbf{S} - \lambda_k \mathbf{I})\, \mathbf{u}_k = 0 \qquad (9.1)$$

with λ_k = eigenvalues;
$\mathbf{u}_k$ = eigenvectors;
I = identity matrix (i.e. diagonal matrix where all diagonal elements are equal to unity).

whose characteristic equation

$$|\, \mathbf{S} - \lambda_k \mathbf{I} \,| = 0 \qquad (9.2)$$

is used to compute the eigenvalues λ_k. The eigenvectors $\mathbf{u}_k$ associated with the λ_k are found by putting the different λ_k values in turn into equation 9.1.
PCA can be defined for a dispersion (or covariance) matrix **S** or for a correlation matrix **R**. A dispersion matrix contains the variances and covariances of p variables (covariance is the extension, to two variables, of the concept of variance; being a measure of dispersion of a random variable around its mean). The covariance measures the joint dispersion of two random variables around their means, while the correlation is defined as a measure of the dependence between two random variables $\mathbf{y}_j$ and $\mathbf{y}_k$. The correlation matrix is the dispersion matrix of the standardized variables (i.e. dimensionless variables with mean equal to zero; and a variance and thus standard deviation equal to one). Standardization of a variable y_i is achieved by subtracting the mean $\bar{y}$ and dividing by the standard deviation s_y of the variable.

$$z_i = \frac{y_i - \bar{y}}{s_y} \qquad (9.3)$$

In an **R** matrix, all diagonal elements are equal to one. The sum of eigenvalues of **S** is equal to the sum of variances s^2, while the sum of eigenvalues of **R** is equal to p (i.e. the number of variables).
The PCs of a correlation matrix are computed from matrix **U** of the eigenvectors of **R** and the matrix of standardized observations:

$$\mathbf{F} = \left[\frac{y - \bar{y}}{s_y} \right] \mathbf{U} \qquad (9.4)$$

F is the matrix of component scores.

For this study, where it was aimed to use equally contributing variables to the clustering of objects, PCA was based on a correlation matrix. Successive principal components correspond to progressively smaller fractions of the total variance. As such, a main problem regarding PCA, is to determine how many components are meaningful for the specific purpose of the dataset. There is an empirical rule that suggests to interpret a principal component if the corresponding eigenvalue λ is larger than the mean of the λ's. In the specific case of standardized data (i.e. in case of a correlation matrix), the mean of the λ's is 1, implying that only the components whose λ's are larger than 1 should be preserved as meaningful components for the analysis (i.e. the Kaiser-Guttman criterion).

9.1.2 Discriminant Function Analysis

Discriminant Function Analysis or Discriminant Analysis (DA) is a specific case of canonical analysis, which is in turn a kind of ordination. For canonical analysis, two or eventually more data matrices are simultaneously compared. An example of this kind of analysis is the relationship between a first table describing macrobenthic communities and a second table of abiotic variables, observed at the same locations.
DA is a usual step in ecological analysis, where an already known group of the objects (i.e. a qualitative response variable **y**) is explained by a set of quantitative descriptors (i.e. the explanatory variables **X**). The already known group at the start of the analysis may be the result of a cluster analysis computed from a different dataset. DA allows interpreting the groups or clusters.
It is a method of linear modelling and proceeds in two steps: 1) testing for differences in the explanatory variables **X** and 2) if the test shows that significant differences between groups are found in the **X** variables, the analysis proceeds to find linear combinations (i.e. discriminant functions) of the **X** variables that best discriminate between the groups.
DA is based upon an explanatory data matrix **X** of size (n x m), where n objects are described by m descriptors. **X** is meant to discriminate between the groups by a distinct classification criterion vector **y**.
For DA, linear combinations of the discriminant descriptors in matrix **X** should be found that maximize the differences among groups while minimizing the variation within the groups. The solution to this problem calls for the eigenvalues and eigenvectors of a matrix corresponding to the ratio of the among-group dispersion **A** to the pooled within-group dispersion **V**. The following matrix equation states the maximization problem:

$$(\mathbf{V}^{-1}\mathbf{A} - \lambda_k\mathbf{I})\ \mathbf{u}_k = 0 \qquad (9.5)$$

with λ_k = eigenvalues;
$\mathbf{u}_k$ = eigenvectors;
$\mathbf{I}$ = identity matrix.

This equation is analogous as the basic equation for PCA (9.1).
To test whether the hypothesis of homogeneity of the within-group dispersion matrices is met, a useful test statistic is Wilks' Λ (1932), which measures to what extent the groups differ in the positions of their centroids.

9.1.3 Ecological Niche Factor Analysis

Factor Analysis

Ecological Niche Factor Analysis (ENFA) (Hirzel et al. 2002a) is an ordination technique, based on Hutchinson's (1957) concept of the ecological niche that computes suitability functions for species by comparing the ecogeographical variables (EGVs) (e.g. depth, grain-size) of the species with that of the whole set of cells.

Unlike other habitat suitability modelling techniques, ENFA only needs presence data of species.
In contrary to PCA, where axes are chosen as to maximize the variance of the distribution, ENFA computes ecologically relevant factors. Still, the output of ENFA is similar to a PCA, with the results being a set of new, linearly independent variables, combining the original EGVs.
Species are generally expected to show non-random distributions regarding EGVs, meaning that a species with e.g. an optimum depth is expected to occur within this optimal range. As such, the depth distribution of the cells in which the species was observed, in comparison with the whole set of cells, may be quantified. These distributions may be different regarding their mean and standard deviations (Figure 9.1).

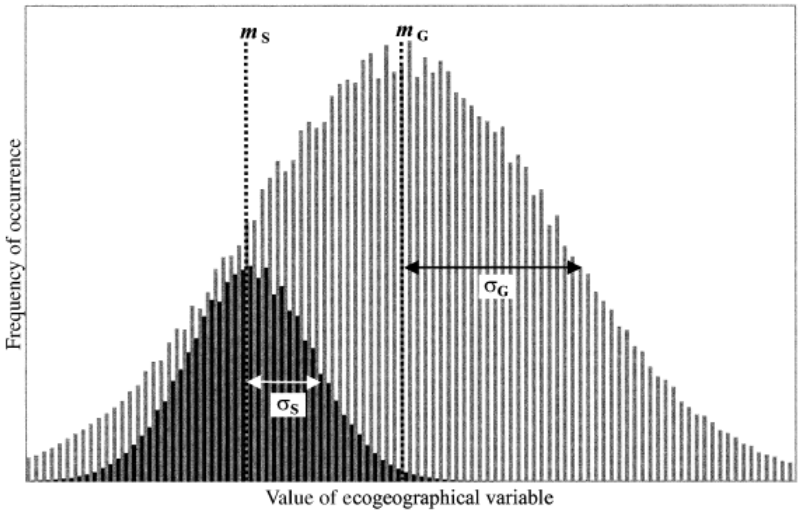


Figure 9.1: Global and species distribution, as defined for one ecogeographical variable (EGV). The distribution of the focal species on any EGV (black bars) is in general different from that of the whole set of cells (grey bars) with respect to its mean ($m_S \neq m_G$) or its standard deviation ($\sigma_S \neq \sigma_G$), allowing respectively marginality and specialization to be defined (from Hirzel et al. 2002a).

For one single EGV, the species' marginality (*M*) can be defined as the absolute difference between the global mean (m_G) and the species mean (m_S), divided by 1.96 standard deviations (σ_G) of the global distribution:

$$M = \frac{|m_G - m_S|}{1.96\sigma_G} \qquad (9.6)$$

A large *M* value close to 1, means that the species lives in a very particular habitat relative to the reference set.
The operational definition of marginality as it is implemented in the Biomapper software (Hirzel et al. 2002b), is a multivariate extension of equation 9.6 (i.e. the *global* marginality):

$$M = \frac{\sqrt{\sum_{i=1}^{V} m_i^2}}{1.96} \qquad (9.7)$$

with V = number of EGVs for each cell i

This is an overall marginality M computed over all EGVs, enabling to compare the marginalities of different species within a given area.
Similarly, the specialization S for one single EGV can be defined as the ratio of the standard deviation of the global distribution (σ_G) to that of the focal species (σ_S):

$$S = \frac{\sigma_G}{\sigma_S} \qquad (9.8)$$

Any S value exceeding 1, indicates some form of specialization.
Again, a global specialization index for all of the EGVs, can be computed as:

$$S = \frac{\sqrt{\sum_{i=1}^{V} \lambda_i}}{V} \qquad (9.9)$$

The multivariate niche (i.e. for all of the EGVs) can be quantified on any of its axes by an index of marginality and specialization. ENFA chooses its first axis as to account for all the marginality of the species, and the following axes as to maximize specialization.
To determine how many components are meaningful for the specific purpose of the dataset, the Biomapper software has incorporated the broken stick model (MacArthur 1960; Frontier 1976; Legendre and Legendre 1998). This model compares the list of decreasing eigenvalues to the decreasing values of the broken stick model. A stick of unit length may be broken at random into p pieces by placing on the stick (p – 1) random break points selected using a uniform [0, 1] random number generator. Frontier (1976) has computed the percentage of variance associated with successive eigenvalues, under the broken stick null model, for 2 to 20 eigenvalues. Only those principal axes are retained that explain a fraction of the variance as small as or smaller than that predicted by the broken stick null model.

Habitat suitability maps

On the basis of the retained factors, the next step of ENFA consists of calculating habitat suitability maps. Habitat suitability maps predict the specific ecological potentials of a habitat rather than the realized ecological structure (Degraer et al 1999b). In other words, if a habitat is found suitable for a species or community, the species have the possibility of colonizing the habitat, but may as well be absent because of anthropogenic impacts, such as fisheries, or natural temporal variability (Degraer et al. 2008).
To calculate habitat suitability maps, different algorithms are incorporated into the Biomapper software (Hirzel et al. 2002b):

- *Median algorithm* (Hirzel et al. 2002a); this algorithm is based on the assumption that the best habitat occurs at the median of the species distribution on each factor (Figure 9.2). This algorithm is limited by the requirement of a unimodal and symmetrical species distribution on each factor.

- *Distance geometric mean algorithm* (Hirzel and Arlettaz 2003). For this algorithm, no assumptions are made on species distribution. The suitability of any point **P** in the environmental factor space is the geometric mean H_G of N species observation points $\mathbf{O}_i$, which is computed from the distances to all observations:

$$H_G(\mathbf{P}) = \sqrt[N]{\prod_{i=1}^{N} \delta(\mathbf{P}, \mathbf{O}_i)} \tag{9.10}$$

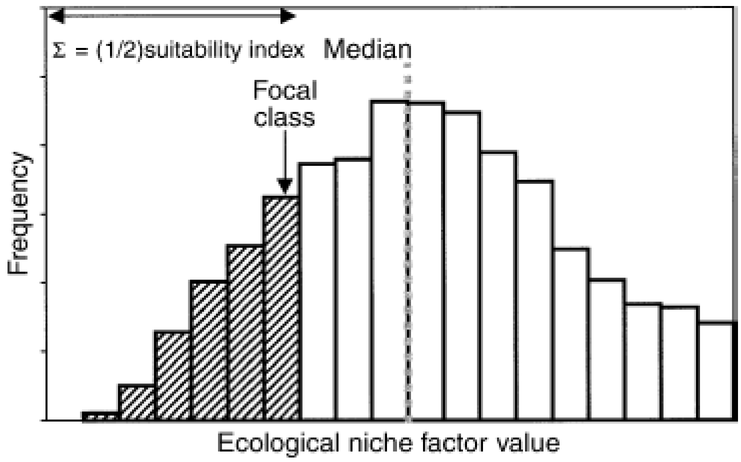


Figure 9.2: The median algorithm for habitat suitability maps. The suitability of any cell from the global distribution is calculated from its situation (arrow) relative to the species distribution (histogram) on all selected niche factors. Specifically, it is calculated as twice the dashed area (sum of all cells from the species distribution that lie as far or farther from the median dashed vertical line) divided by the total number of cells from the species distribution (surface of the histogram) (from Hirzel et al. 2002a).

- *Distance harmonic mean algorithm* (Hirzel and Arlettaz 2003). No assumptions are made on the species distribution. The suitability of any point **P** in the environmental factor space is the harmonic mean H_H of N species observation points $\mathbf{O}_i$, which is computed from the distances to all observations:

$$H_H(\mathbf{P}) = \frac{1}{\frac{1}{N} \sum_{\substack{i=1 \\ \mathbf{P} \neq \mathbf{O}_i}}^{N} \frac{1}{\delta(\mathbf{P}, \mathbf{O}_i)}} \tag{9.11}$$

- *Minimum distance algorithm* (Hirzel and Arlettaz 2003). No assumption is made on the species distribution and the computations are based on:

$$H_M(\mathbf{P}) = \underset{i=1}{\overset{N}{Min}}\{\delta(\mathbf{P}, \mathbf{O}_i)\} \tag{9.12}$$

Evaluation of habitat suitability maps

To compare habitat suitability maps based on different algorithms, a whole set of evaluation indices exist (see Hirzel et al. 2006 for an overview). For this study, use was made of the continuous Boyce index (Hirzel et al. 2006), based on the approach of Boyce et al. (2002). This method consists in partitioning the habitat suitability range into b classes. For each class i, it calculates 2 frequencies:

1) P_i, the predicted frequency of evaluation points:

$$P_i = \frac{p_i}{\sum_{j=1}^{b} p_j} \tag{9.13}$$

with p_i = number of evaluation points predicted by the model to fall in the habitat suitability class i;
$\sum p_j$ = total number of evaluation points.

2) E_i, the expected frequency of evaluation points or the frequency expected from a random distribution across the study area, given by the relative area covered by each class:

$$E_i = \frac{a_i}{\sum_{j=1}^{b} a_j} \tag{9.14}$$

with a_i = number of grid cells belonging to habitat suitability class i (or area covered by class i);
$\sum a_j$ = overall number of cells in the whole study area.

For each class i, the predicted to expected (P/E) ratio F_i is given by:

$$F_i = \frac{P_i}{E_i} \tag{9.15}$$

A low suitability class should contain fewer evaluation presences than expected by chance, resulting in $F_i < 1$, whereas high suitability classes are expected to have F_i values increasingly higher than 1. A good model is thus expected to show a monotonically increasing curve (increase of F_i and increase of habitat suitability). Boyce et al. (2002) proposed to measure this monotonic increase by the Spearman

rank correlation coefficient between F_i and i (i.e. "Boyce index" varying between -1 and 1, respectively corresponding with a bad and a good model).
Because of the sensitivity to the number of suitability classes b and to their boundaries, Hirzel et al. (2006) proposed a new index ("continuous Boyce index"), based on a moving window of width W (e.g. 0.2), instead of fixed classes. The computation of this index starts with a first class covering the suitability range [0,W] whose *P/E* ratio is plotted against the average suitability values of the class, W/2. In the next step, the moving window is shifted from a small amount upwards and *P/E* is plotted again. This is repeated until the window reaches the last possible range [1-W,1]. This results generally in a smooth *P/E* curve, on which a continuous Boyce index is computed.

9.2 Cluster Analysis

Cluster analysis is classifying objects into collective categories. Clustering requires the recognition of discontinuities in an environment which is most often continuous; objects are recognized to be sufficiently similar to be put in the same group and to also identify distinctions or separations between groups. Clustering is an operation of multidimensional analysis which consists in partitioning the collection of objects (or descriptors) in the study.
Most clustering methods proceed from association matrices. Association is a general term to describe any measure or coefficient used to quantify the resemblance or difference between objects or descriptors. Similarity coefficients are distinguished from distance (or dissimilarity) and dependence coefficients. Similarity coefficients represent the largest group in the literature. They are maximum when two objects are identical and minimum when the objects are completely different. Distance coefficients follow the opposite rule. For dependence coefficients, zero corresponds to no association. A typical association measure to be used with abiotic descriptors is the Euclidean distance, computed using Pythagoras' formula, from site-points positioned in a p-dimensional space (i.e. a metric or Euclidean space). For the various descriptors y_j (j = 1…p), the Euclidean distance between objects $\mathbf{x}_1$ and $\mathbf{x}_2$ is computed as follows:

$$D_1\ (\mathbf{x}_1,\mathbf{x}_2) = \sqrt{\sum_{j=1}^{p} (y_{1j} - y_{2j})^2} \qquad (9.16)$$

The case of two descriptors is represented in Figure 9.3.

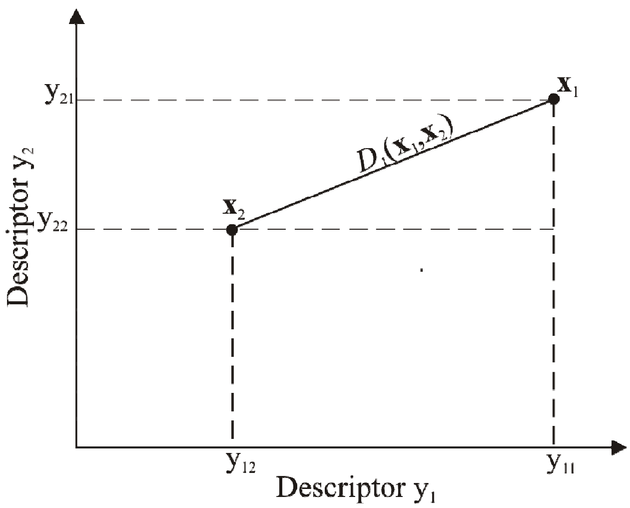


Figure 9.3: Euclidean distance (D_1) between objects x_1 and x_2 in the case of 2 descriptors (i.e. a 2-dimensional space) (after Legendre and Legendre 1998).

Sneath and Sokal (1973) propose a classification of clustering procedures. For this study, only two dichotomies are briefly described, as only those dichotomies are relevant for the techniques used in Chapter 4.

- **Agglomeration versus division**

Agglomerative methods start with the discontinuous partition of all objects. They are successively grouped into larger and larger clusters until a single cluster is obtained. If a single group containing all objects is used as starting point, *divisive* algorithms subdivide the group into sub-clusters, until the discontinuous partition is reached.

- **Hierarchical versus non-hierarchical**

Hierarchical methods put the members of inferior-ranking clusters into larger, higher-ranking clusters. A typical product of hierarchical clustering is a dendrogram. *Non-hierarchical* methods are very useful in ecology; they produce a single partition which optimizes within-group homogeneity. The latter techniques should be used where a direct representation of the relationships among objects should be obtained instead of a summary of their hierarchy.

9.2.1 Ward's minimum variance method

Ward's (1963) or Orlóci 's (1967) minimum variance method is a hierarchical, agglomerative clustering algorithm that minimizes an objective function which is the same "squared error" criterion that is used in multivariate analysis of variance and results into clusters with a minimal variance between each cluster. At each clustering step, this method finds the pair of objects or clusters whose fusion increases as little as possible the sum, over all objects of the squared Euclidean distances between objects and cluster centroids (i.e. the type-objects of the cluster, whether these objects actually exist or are only a mathematical construct). At the beginning of the analysis, all samples are in a separate cluster, with the sum of squared distances equal to zero, since each sample coincides with the centroid of its cluster. The distance of object $\mathbf{x}_i$ to the centroid $\mathbf{m}$ of its cluster is computed using the Euclidean distance formula (eq. 9.16) over the various descriptors $\mathbf{y}_j$ (j = 1…p):

$$\sum_{j=1}^{p}\left[y_{ij} - m_j\right]^2 \qquad (9.17)$$

The sum of squared distances of all objects in cluster k to their common centroid (i.e. 'error' in analysis of variance or ANOVA) is defined as follows:

$$e_k^2 = \sum_{i=1}^{n_k}\sum_{j=1}^{p}\left[y_{ij}^{(k)} - m_j^{(k)}\right]^2 \qquad (9.18)$$

where $y_{ij}^{(k)}$ is the value of descriptor $\mathbf{y}_j$ for an object i member of group (k) and $m_j^{(k)}$ is the mean value of descriptor j over all members of group k.
At each clustering step, the criterion to be minimized, is the *sum of squared errors* E_K^2 (or the within-cluster squared errors) for all clusters corresponding to a given partition:

$$E_K^2 = \sum_{k=1}^{K} e_k^2 \qquad (9.19)$$

9.2.2 K-means partitioning

K-means partitioning is a non-hierarchical, divisive clustering technique. Partitioning clustering methods produce clusters in a predefined number of groups (K). Objects are allocated to a cluster to which the distance to its centre is minimal. The procedure stops if all objects have been allocated.
The objective function that the new partition should minimize is the same as in Ward's minimum variance method: E_K^2. Hereby, the major problem is the initial position of the cluster centroids (a non-existing problem in the case of Ward's method that proceeds iteratively by hierarchical agglomeration). This problem is known as the "local mimimum" problem in algorithms. A commonly used approach for solving this problem is to provide an initial configuration of cluster centroids corresponding to the result of a hierarchical clustering (e.g. Ward's method). The *K*-means algorithm is then used to rearrange the group membership and to look for a better overall solution (i.e. a lower E_K^2 value).

Another problem related to the *K*-means algorithm is the most appropriate number of clusters (K). From a simulation study comparing 30 indices as 'stopping rules', Milligan and Cooper (1985) proposed the Calinski-Harabasz (1974) or C-H criterion as giving the best results. C-H is the F-statistic of multivariate analysis of variance and canonical analysis. F is the ratio of the mean square for the given partition, divided by the mean square for the residuals. The number of clusters corresponding with the highest C-H value is the optimal solution in the least squares sense. It is defined by the ratio of the mean square for the given grouping divided by the mean square of the residuals (formulas from Orpin and Kostylev 2006):

$$\text{C-H} = \frac{\left[\frac{R^2}{(K-1)}\right]}{\left[\frac{(1-R^2)}{(n-K)}\right]} \qquad (9.20)$$

with n = number of objects;
K = number of clusters; and where:

$$R^2 = \frac{(SST - SSE)}{SST} \qquad (9.21)$$

with SST = total sum of the squared distances to the overall centroid (similar to the between groups sum of the squares);
SSE = sum of the squared distances of the objects to the groups own centroids (or the E_K^2 or within group sum of the squares).

The number of clusters with a maximum C-H value, is the optimal grouping solution in terms of a least-squares solution (similar to a pseudo F-test).

9.2.3 Species indicator analysis

Species indicator analysis or INDVAL (Dufrêne and Legendre 1997) can be used to search for indicator species characterizing clusters. The index is maximum when all individuals of a species are found in a single group of sites and when the species occurs in all sites of that group. The INDVAL index is defined as follows:

$$\text{INDVAL}_{ij} = A_{ij} \times B_{ij} \times 100 \qquad (9.22)$$

with A_{ij} = $\text{Nindividuals}_{ij}/\text{Nindividuals}_i$ or the mean abundance of species *i* in the sites of group *j* compared to all groups in the study. A_{ij} is a measure of specificity and is maximum when species *i* is only present in cluster *j*;
B_{ij} = $\text{Nsites}_{ij}/\text{Nsites}_j$ or the relative frequency of occurrence of species *i* in the sites of group *j*. B_{ij} is a measure of fidelity and is maximum when species *i* is present in all sites of cluster *j*.

9.3 Multi-scale terrain analysis

Following Wilson et al. (2007), terrain analysis variables can be grouped into four groups, all of them being derivatives from a digital terrain model (DTM):

- slope;
- orientation (aspect);
- curvature and relative position of features;

- terrain variability.

All of these variables can be computed in GIS software or a specialized terrain analysis software (e.g. LandSerf from Wood (2005)).

Terrain analysis on a DTM may be performed on the basis of the DTM represented as a raster grid, or based on a continuous representation of a DTM as a double-differentiable surface (e.g. Wood 1996). The latter approach offers a great flexibility in the choice of algorithms of terrain analysis and the scales at which the analyses are performed. Following Evans (1980), a DTM is approximated by a bivariate quadratic equation:

$$Z = aX^2 + bY^2 + cXY + dX + eY + f \qquad (9.23)$$

with Z = height of the DTM surface;
X and Y = horizontal coordinates;

The coefficients a, b, c, d, e and f in equation 9.22 can be solved within a window using simple combinations of neighboring cells, which is the basis for terrain analysis in most GIS softwares, whether they use grid-based methods or a mathematical representation of the DTM.
For a terrain analysis across a variety of spatial scales, Wood (1996) solves this equation for an n by n matrix with a local coordinate system (x, y, z) defined with the origin at the central pixel where the user may specify any odd number (n) for the size of the square analysis window defining the portion of the raster DTM to be analyzed in relation to each central pixel in turn. To compute the terrain parameters in LandSerf (Wood 2005), an analysis window is effectively moved across the raster DTM surface such that each pixel in turn becomes the central pixel on which calculations are based.

- **Slope**

Slope is a first spatial derivative of the DTM and is calculated as follows:

$$\frac{\partial Z}{\partial X} = S_x = 2aX + cY + d \quad \text{and} \quad \frac{\partial Z}{\partial Y} = S_y = 2bY + cX + e \qquad (9.24)$$

By adopting a local coordinate system with the origin at the central point of the analysis window (x, $y = 0$), the slope at the centre of the (moving) analysis window is:

$$\text{slope} = \arctan\left(\sqrt{d^2 + e^2}\right) \qquad (9.25)$$

- **Orientation**

Aspect is as well a first spatial derivative of the DTM and represents the orientation of the slope. It is calculated as follows:

$$\text{aspect} = \arctan\left(\frac{e}{d}\right) \qquad (9.26)$$

As this parameter is typically measured in degrees, completely different values, may be oriented in the same direction (e.g. 1° and 359°). As such, aspect can be splitted into two components (*eastness* and *northness*), following a conversion from degrees to radians (Hirzel et al. 2002a):

$$\text{eastness} = \sin(\text{aspect}) \tag{9.27}$$

$$\text{northness} = \cos(\text{aspect}) \tag{9.28}$$

These indices provide continuous measures (−1 to +1) describing orientation.

- **Curvature and relative position of features**

Surface curvature is a second spatial derivative of the DTM. Difference is made between profile, plan and mean curvature, corresponding respectively with:

Profile curvature is the rate of change of slope along a profile in the surface; useful to highlight convex and concave slopes.

$$\text{Profile curvature} = \frac{-200(ad^2 + be^2 + cde)}{(e^2 + d^2)(1 + e^2 + d^2)^{1.5}} \tag{9.29}$$

Plan curvature is the rate of change of aspect in plan across the surface; useful for defining ridges, valleys and slopes.

$$\text{Plan curvature} = \frac{200(bd^2 + ae^2 - cde)}{(e^2 + d^2)^{1.5}} \tag{9.30}$$

Mean curvature is the average value obtained from maximum and minimum profile curvature, providing an indication of the relative position of features.

$$\text{Mean curvature} = -a - b \tag{9.31}$$

The *Bathymetric Position Index* (BPI) is a measure of where a location, with a defined elevation, is relative to the overall landscape. The calculation is a raster-grid based method rather than one based on quadratic representation of the DTM surface. The definition is as follows (Lundblad et al. 2006):

$$\text{BPI}\langle \textit{scalefactor} \rangle = \text{int}(Z_{grid} - \text{focalmean}(Z_{grid}, \text{circle}, r)) + 0.5) \tag{9.32}$$

with scalefactor = radius in map units multiplied by data resolution of DTM;
int = integer;
Z_{grid} = raster grid of DTM;
focalmean = raster calculation of the mean of the raster within the circle of radius r

- **Terrain variability**

Rugosity (Jenness 2002) is the ratio of the surface area to the planar area, across the neighbourhood of the central pixel and is computed as follows:

Rugosity = surface area of 3x3 neighbourhood / planar area of 3x3 neighbourhood (9.33)

By this method flat areas will have a rugosity value near to 1, while high relief areas will exhibit higher values of rugosity.

Fractal dimension (Mandelbrot 1983) is a measure of the surface complexity. There are a variety of methods available for the calculation of the fractal dimension (D) of topography that will give a D value between 2 (flat surface) and 3 (a space filling rough surface). Landserf (Wood 2005) implements the variogram method which calculates fractal dimension through a plot of the log of variance against the log of lag, typically referred to as the log-log variogram. The variogram is calculated as follows:

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{n} \sum_{j=1}^{n} (z_i - z_j)^2 \qquad (9.34)$$

with h = lag between measured cells;
n = number of pairs.

References

Akoumianaki, I. and Nicolaidou, A., 2007. Spatial variability and dynamics of macrobenthos in a Mediterranean delta front area: The role of physical processes. Journal of Sea Research, 57: 47-64.

Al-Hamdani, Z. and Reker, J. (eds.), 2007. Towards marine landscapes in the Baltic Sea. BALANCE interim report #10. 117 pp. Available online at (http://balance-eu.org/).

Allee, R.J., Dethier, M., Brown, D., Deegan, L., Ford, R.G., Hourigan, T.F., Maragos, J., Schoch, C., Sealey, K., Twilley, R., Weinstein, M.P., and Yoklavich, M.M., 2000. Marine and Estuarine Ecosystem and Habitat Classification. NOAA Technical Memorandum NMFS-F/SPO-43, National Oceanic and Atmospheric Administration, National Marine Fisheries Service.

Anderson, J.T., Gregory, R.S. and Collins, W.T., 2002. Acoustic classification of marine habitats in coastal Newfoundland. ICES Journal of Marine Science, 59(1): 156-167.

Armstrong, M., 1994. Is research in mining geostats as dead as a dodo? In: Dimitrakopoulos, R. (Ed.), Geostatistics for The Next Century. Kluwer Academic Publishers, Dordrecht, pp. 303–312.

Ashley, G.M., 1990. Classification of large-scale subaqueous bedforms: a new look at an old problem. Journal of Sedimentary Petrology, 60(1): 160-172.

Austin, M.P., Cunningham, R.B. and Fleming, P.M., 1984. New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. Vegetatio, 55: 11–27.

Azovsky, A.I., Chertoprood, M.V., Kucheruk, N.V., Rybnikov, P.V. and Sapozhnikov, F.V., 2000. Fractal properties of spatial distribution of intertidal benthic communities. Marine Biology, 136(3): 581-590.

Banks, S.A. and Skilleter, G.A., 2002. Mapping intertidal habitats and an evaluation of their conservation status in Queensland, Australia. Ocean & Coastal Management, 45(8): 485-509.

Baptist, M.J., Van Dalfsen, J., Weber, A., Passchier, S. and Van Heteren, S., 2006. The distribution of macrozoobenthos in the southern North Sea in relation to meso-scale bedforms. Estuarine, Coastal and Shelf Science, 68 (3-4): 538-546

Bates, C.R. and Oakley, D.J., 2004. Bathymetric sidescan investigation of sedimentary features in the Tay Estuary, Scotland. International Journal of Remote Sensing, 25(22): 5089-5104.

Bellec, V., Van Lancker, V., Degrendele, K., Roche, M., Schotte, P. and Le Bot, S., in press. Geo-environmental characterization of the Kwinte Bank. Journal of Coastal Research.

Berné, S., Trentesaux, A., Stolk, A., Missiaen, T. and De Batist, M., 1994. Architecture and long term evolution of a tidal sandbank: The Middelkerke Bank (southern North Sea). Marine Geology, 121(1-2): 57-72.

Beukema, J.J., Essink, K., Michaelis, H.and Zwarts, L., 1993. Year-to-year variability in the biomass of macrobenthic animals on tidal flats of the Wadden Sea: how predictable is this food source for birds? Netherlands Journal of Sea Research, 31: 319–330.

Borgo, A., 2003. Habitat requirements of the Alpine marmot *Marmota marmota* in re-introduction areas of the Eastern Italian Alps. Formulation and validation of habitat suitability models. Acta Theriologica, 48: 557–569.

Borja, A., Franco, J. and Pérez, V., 2000. A marine biotic index to establish the ecological quality of soft-bottom benthos within European estuarine and coastal environments. Marine Pollution Bulletin, 12: 1100–1114.

Bourennane, H., King, D. and Couturier, A., 2000. Comparison of kriging with external drift and simple linear regression for predicting soil horizon thickness with different sample densities. Geoderma, 97: 255–271.

Bourennane, H. and King, D., 2003. Using multiple external drifts to estimate a soil variable. Geoderma, 114: 1–18.

Boyce, M.S., Vernier, P.R., Nielsen, S.E. and Schmiegelow, F.K.A., 2002. Evaluating resource selection functions. Ecological Modelling, 157: 281-300

Bradshaw, C.J.A., Davis, L.S., Purvis, M., Zhou, Q. and Benwell, G.L., 2002. Using artificial neural networks to model the suitability of coastline for breeding by New Zealand fur seals (*Arctocephalus forsteri*). Ecological Modelling, 148(2): 111-131.

Brinkman, B., Dankers, N. and van Stralen, M., 2002. An analysis of mussel bed habitats in the Dutch Wadden Sea. Helgoland Marine Research, 56(1): 59-75.

Brown, C.J., Cooper, K.M., Meadows, W.J., Limpenny, D.S. and Rees, H.L., 2002. Small-scale Mapping of Sea-bed Assemblages in the Eastern English Channel Using Sidescan Sonar and Remote Sampling Techniques. Estuarine, Coastal and Shelf Science, 54(2): 263-278.

Brown, C.J. and Collier, J.S., 2008. Mapping benthic habitat in regions of gradational substrata: An automated approach utilising geophysical, geological, and biological relationships. Estuarine, Coastal and Shelf Science, 78(1): 203-214.

Bryan, T.L. and Metaxas, A., 2007. Predicting suitable habitat for deep-water gorgonian corals on the Atlantic and Pacific Continental Margins of North America. Marine Ecology-Progress Series, 330: 113-126.

Burrough, P.A. and McDonnell, R.A., 1998. Principles of Geographic Information Systems. Oxford University Press, Oxford. 333 pp.

Caeiro, S., Costa, M.H., Goovaerts, P. and Martins, F., 2005. Benthic biotope index for classifying habitats in the sado estuary: Portugal. Marine Environmental Research, 60: 570-593.

Calinski, T., and Harabasz, J., 1974. A dendrite method for cluster analysis. Communications in Statistics, 3: 1–27.

Cao, Y., Bark, A.W. and Williams, W.P., 1997. A comparison of clustering methods for river benthic community analysis. Hydrobiologia, 347(1): 24-40.

Cardillo, M., Macdonald, D.W. and Rushton, S.P., 1999. Predicting mammal species richness and distributions: testing the effectiveness of satellite-derived land cover data. Landscape Ecology, 14(5): 423–435.

Cattrijsse, A. and Vincx, M., 2001. Biodiversity of the benthos and the avifauna of the Belgian Coastal Waters. Summary of data collected between 1970-1998. Belgian Science Policy, PODO I. Report. Brussels. 48 pp.

Clarke, K.R., 1993. Non-parametric analyses of changes in community structure. Australian Journal of Ecology, 18: 117–143.

Clifford, H.T. and Stephenson, W., 1975. An Introduction to Numerical Classification. Academic Press, New York, USA. 229 pp.

Cliquet, A., Bogaert, D., De Waen, D. and Maes, F., 2007. The designation of Marine Protected Areas in Belgium: From government to governance? Proceedings MARE Conferentie People and the Sea IV. Who ows the Coast, Amsterdam, 5-7 juli 2007. 10 pp.

Cochrane, G.R. and Lafferty, K.D., 2002. Use of acoustic classification of sidescan sonar data for mapping benthic habitat in the Northern Channel Islands, California. Continental Shelf Research, 22(5): 683-690.

Coggan, R., Populus, J., White, J., Sheehan, K., Fitzpatrick, F. and Piel, S. (eds.), 2007. Review of Standards and Protocols for Seabed Habitat Mapping. In: MESH Guide to Habitat Mapping, MESH Project, 2007, JNCC, Peterborough. 210 pp. Available online at: (http://www.searchmesh.net/).

Commission of the European Communities, 2005. Proposal for a Directive of the European Parliament and the Council establishing a framework for Community action in the field of marine environmental policy (Marine Strategy Directive). [SEC(2005)1290], Brussels, 26 October 2005. 35 pp.

Commission of the European Communities, 2006. Green Paper: Towards a future Maritime Policy for the Union: A European vision for the oceans and seas. COM(2006) 275 final, Brussels, 7 June 2006. 49 pp.

Connor, D.W., Allen, J.H., Golding, N., Howell, K.L., Lieberknecht, L.M., Northen, K.O. and Reker, J.B., 2004. The National Marine Habitat Classification for Britain and Ireland. Version 04.05. Joint Nature Conservation Committee, Peterborough, UK, Available online at: (www.jncc.gov.uk/MarineHabitatClassification).

Connor, D.W., Gilliland, P.M., Golding, N, Robinson, P., Todd, D., and Verling, E., 2006. UKSeaMap: the mapping of seabed and water column features of UK seas. Joint Nature Conservation Committee, Peterborough. 120 pp.

Dartnell, P. and Gardner, J.V., 2004. Predicting seafloor facies from multibeam bathymetry and backscatter data. Photogrammetric Engineering & Remote Sensing, 70(9): 1081-1091.

Darwin, C., 1859. On the Origin of Species by Means of Natural Selection (6th Edition). John Murray, London. 502 pp.

Dauvin, J.C., Thiébaut, E., Gomez Gesteira, J.L., Ghertsos, K., Gentil, F., Ropert, M., and Sylvand, B., 2004. Spatial structure of a subtidal macrobenthic community in the Bay of Veys (western Bay of Seine, English Channel): Journal of Experimental Marine Biology and Ecology, 307(2): 217-235.

De Batist, M., 1989. Seismo-stratigrafie en Struktuur van het Paleogeen in de Zuidelijke Noordzee. PhD Thesis Thesis, Rijksuniversiteit Gent, 107 pp.

De Batist, M. and Henriet, J.-P., 1995. Seismic sequence stratigraphy of the Palaeogene offshore of Belgium, southern North Sea. Journal of the Geological Society of London, 152: 27-40.

DEFRA, 2006. A Marine Bill: A consultation document. Department for Environment, Food and Rural Affairs, London. 309 pp.

Degraer, S., Vincx, M., Meire, P. and Offringa, H., 1999a. The macrozoobenthos of an important wintering area of the Common scoter (*Melanitta nigra*). Journal of the Marine Biological Association of the United Kingdom, 79: 243–251.

Degraer, S., Vincx, M. and Meire, P., 1999b. Macrobenthic community changes: dynamics versus stability and their importance in coastal zone management. In: Degraer, S. (Ed.), Macrobenthos of Shallow Marine Habitats (Belgian coast) and its Use in Coastal Zone Management. Ph.D. Thesis, Ghent University, Belgium, unpublished. 158 pp.

Degraer, S., Van Lancker, V., Moerkerke, G., Van Hoey, G., Vincx, M., Jacobs, P. and Henriet, J.-P., 2002. Intensive evaluation of the evolution of a protected benthic habitat: HABITAT. Report, Federal Office for Scientific, Technical and Cultural Affairs (OSTC), Brussels, Belgium. 124 pp.

Degraer, S., Van Lancker, V., Moerkerke, G., Van Hoey, G., Vanstaen, K., Vincx, M. and Henriet, J.-P., 2003. Evaluation of the ecological value of the foreshore: habitat-model and macrobenthic side-scan sonar interpretation: extension along the Belgian Coastal Zone. Final report. Ministry of the Flemish Community, Environment and Infrastructure Department. Waterways and Marine Affairs Administration, Coastal Waterways, 63 pp.

Degraer S., Wittoeck, J., Appeltans, W., Cooreman, K., Deprez, T., Hillewaert, H., Hostens, K., Mees, J., Vanden Berghe, E. and Vincx, M., 2006. The macrobenthos atlas of the Belgian part of the North Sea. Belgian Science Policy. D/2005/1191/3. ISBN 90-810081-6-1. 164 pp.

Degraer, S., Verfaillie, E., Willems, W., Adriaens, E., Vincx, M. and Van Lancker, V., 2008. Habitat suitability modelling as a mapping tool for macrobenthic communities: An example from the Belgian part of the North Sea. Continental Shelf Research, 28: 369-379.

de la Huz, R., Lastra, M. and Lopez, J., 2002. The influence of sediment grain size on burrowing, growth and metabolism of *Donax trunculus L.* (Bivalvia: Donacidae). Journal of Sea Research, 47: 85–95.

Deleu, S.,Van Lancker, V., Van den Eynde, D. and Moerkerke, G., 2004. Morphodynamic evolution of the kink of an offshore tidal sandbank: the Westhinder Bank (Southern North Sea). Continental Shelf Research, 24: 1587-1610.

Deleu, S. and Van Lancker, V., 2007. Geological setting of gravel occurrences on the Belgian part of the North Sea. In: Van Lancker, V., Du Four, I., Verfaillie, E., Deleu, S., Schelfaut, K., Fettweis, M., Van den Eynde, D., Francken, F., Monbaliu, J., Giardino, A., Portilla, J., Lanckneus, J., Moerkerke, G. and Degraer, S., 2007. Management, research and budgetting of aggregates in shelf seas related to end-users (Marebasse). Brussel (B), Belgian Science Policy (D/2007/1191/49), 139 pp. + DVD GIS@SEA + Habitat Signature Catalogue.

Derous, S., Verfaillie, E., Van Lancker, V., Courtens, W., Stienen, E.W.M., Hostens, K., Moulaert, I., Hillewaert, H., Mees, J., Deneudt, K., Deckers, P., Cuvelier, D., Vincx, M. and Degraer, S., 2007. A biological valuation map for the Belgian part of the North Sea: BWZee, Final report, Research in the framework of the BELSPO programme "Global change, ecosystems and biodiversity" – SPSD II, March 2007, 99 pp. (+ Annexes).

Derous, S., Courtens, W., Deckers, P., Deneudt, K., Forero, C.E., Hostens, K., Moulaert, I., Stienen, E.W.M., Vanden Eede, S., Van Lancker, V., Verfaillie, E., Vincx, M. and Degraer, S. (submitted (a) to Ecological Applications) Application of the protocol for marine biological valuation to selected case study areas.

Derous, S., Van Hoey, G., Van Damme, S., Vincx, M. and Degraer, S. (submitted (b) to Marine Policy). Relevance of the marine biological valuation protocol for the implementation of present and future marine European Directives: Comparison with other implementation methods.

Deutsch, C.V. and Journel, A.G., 1998. GSLIB: Geostatistical Software Library and User's Guide, second ed. Oxford Univ. Press, New York. 369 pp.

De Waen, D., 2004. Het macrobenthos van slibrijkse substraten. Een ongekende gemeenschap van het Belgisch Continentaal Plat. Unpublished Msc Thesis, Universiteit Gent. 75 pp.

Douvere, F., Maes, F., Vanhulle, A. and Schrijvers, J., 2007. The role of marine spatial planning in sea use management: The Belgian case. Marine Policy, 31(2): 182-191.

Dufrene, M. and Legendre, P., 1997. Species assemblages and indicator species: The need for a flexible asymmetrical approach. Ecological Monographs, 67(3): 345-366.

Eastwood, P.D., Meaden, G.J. and Grioche, A., 2001. Modelling spatial variations in spawning habitat suitability for the sole *Solea solea* using regression quantiles and GIS procedures. Marine Ecology-Progress Series, 224: 251-266.

Edwards, B.D., Dartnell, P. and Chezar, H., 2003. Characterizing benthic substrates of Santa Monica Bay with seafloor photography and multibeam sonar imagery. Marine Environmental Research, 56(1-2): 47-66.

Essink, K., Beukema, J.J., Madsen, P.B., Michaelis, H. and Vedel, G.R., 1998. Long-term development of biomass of intertidal macrozoobenthos in different parts of the Wadden Sea. Governed by nutrient loads? Senckenbergiana maritime, 29: 25–35.

EUNIS, 2002. EUNIS Habitat Classification. European Environment Agency. Available online at: (http://eunis.eea.eu.int).

European Commission, 2006. EU Marine Strategy. The story behind the strategy. Luxembourg: Office for Official Publications of the European Communities. 30 pp. Available online at: (http://europa.eu.int/comm/environment/water/marine.htm).

Evans, I.S., 1980. An integrated system of terrain analysis and slope mapping. Zeitschrift für Geomorphologic Suppl-Bd, 36:274–295.

Fager E.W., 1964. Marine Sediments: Effects of a Tube-Building Polychaete. Science, 143 (3604): 356-358.

Fairbanks, D.H.K., 2000. Physio-climatic classification of South Africa’s woodland biome. Plant Ecology, 149: 71-89.

Fielding, A.L. and Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation, 24: 38-49.

Folk, R.L., 1974. The Petrology of Sedimentary Rocks. Hemphill Publishing Co, Austin. 182 pp.

Foster-Smith, R., Connor, D. and Davies, J., 2007a. What is habitat mapping? In: MESH Guide to Habitat Mapping, MESH Project, 2007, JNCC, Peterborough. 66 pp. Available online at: (http://www.searchmesh.net/).

Foster-Smith, R., Coltman, N. and Fitzpatrick, F. 2007b. How good is my map? In: MESH Guide to Habitat Mapping, MESH Project, 2007, JNCC, Peterborough. 36 pp. Available online at: (http://www.searchmesh.net/).

Fourqurean, J.W., Boyer, J.N., Durako, M.J., Hefty, L.N. and Peterson, B.J., 2003. Forecasting responses of seagrass distributions to changing water quality monitoring data. Ecological Applications, 13: 474–489.

Franklin, E.C., Ault, J.S., Slutg, S.G., Luo, J., Meester, G.A., Diaz, G.A., Chiappone, M., Swanson, D.W., Miller, S.L. and Bohnsack, J.A., 2003. Benthic habitat mapping in the Tortugas region, Florida. Marine Geodesy, 26(1-2): 19-34.

Freitas, R., Rodrigues, A.M. and Quintino, V., 2003a. Benthic biotopes remote sensing using acoustics. Journal of experimental marine biology and ecology, 285-286: 339-353.

Freitas, R., Silva, S., Quintino, V., Rodrigues, A.M., Rhynas, K. and Collins, W.T., 2003b. Acoustic seabed classification of marine habitats: studies in the western coastal-shelf area of Portugal. ICES Journal of Marine Science, 60(3): 599-608.

Fromentin, J.M., Ibanez, F., Dauvin, J.C., Dewarumez, J.M. and Elkaim, B., 1997. Long-term changes of four macrobenthic assemblages from 1978 to 1992. Journal of the Marine Biological Association of the United Kingdom, 77(2): 287-310.

Frontalini, F. and Coccioni, R., 2008. Benthic foraminifera for heavy metal pollution monitoring: A case study from the central Adriatic Sea coast of Italy. Estuarine, Coastal and Shelf Science, 76(2): 404-417.

Frontier, S., 1976. Etude de la décroissance des valeurs propres dans une analyse en composantes principales: comparison avec le modèle du bâton brisé. Journal of Experimental Marine Biology and Ecology, 25: 67-75.

Gauch, H.G. and Whittaker, R.H., 1981. Hierarchical classification of community data. Journal of Ecology, 69: 537–557.

Golding, N., Vincent, M.A. and Connor, D.W., 2004. Irish Sea Pilot—Report on the development of a Marine Landscape classification for the Irish Sea, Joint Nature Conservation Committee, unpublished. 172 pp. Available online at: (www.jncc.gov.uk/irishseapilot).

Goovaerts, P., 1997. Geostatistics for Natural Resources Evalation. Oxford University Press, New York. 483 pp.

Goovaerts, P., 1999. Using elevation to aid the geostatistical mapping of rainfall erosivity. Catena, 34: 227–242.

Gower, J.C., 1984. Ordination, multidimensional scaling and allied topics. 727-781. In: Lederman, W. (ed.) Handbook of Applicable Mathematics Vol. VI: Lloyd, E. (ed.) Statistics. Wiley, Chichester.

Greene, H.G., Yoklavich, M.M., Starr, R.M.O., Connell V.M., Wakefield, W.W., Sullivan, D.E., McRea, J.E. and Cailliet, G.M., 1999. A classification scheme for deep seafloor habitats. Oceanologica Acta, 22(6): 663-678.

Gregr, E.J. and Bodtker, K.M., 2007. Adaptive classification of marine ecosystems: Identifying biologically meaningful regions in the marine environment. Deep Sea Research Part I: Oceanographic Research Papers, 54(3): 385-402.

Greulich, S., Bornette, G., Amoros, C. and Roelofs, J.G.M., 2000. Investigation on the fundamental niche of a rare species: an experiment on establishment of *Luronium natans*. Aquatic Botany, 66(3): 209-224.

Guisan, A. and Zimmerman, N.E., 2000. Predictive habitat distribution models in ecology. Ecological Modelling, 135(2-3): 147-186.

Guisan, A., and Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters, 8(9): 993–1009.

Haelters, J., Vigin, L., Stienen, E.W.M., Scory, S., Kuijken, E. and Jacques, T.G., 2004. Ornithologisch belang van de Belgische zeegebieden: identificatie van mariene gebieden die in aanmerking komen als Speciale Beschermingszones in uitvoering van de Europese Vogelrichtlijn. Bulletin van het Koninklijk Belgisch Instituut voor Natuurwetenschappen, Biologie 74 (supplement). 91 pp.

Haelters, J., Kerckhof, F. and Houziaux, J.-S., 2007. De aanduiding van mariene beschermde gebieden in de Belgische Noordzee: een mogelijke uitvoering van OSPAR Aanbeveling 2003/3 door België. Federaal Wetenschapsbeleid, Koninklijk Belgisch Instituut voor Natuurwetenschappen, Beheerseenheid Mathematisch Model Noordzee (BMM), Brussel. 46 pp.

Hartmann-Schröder, G., 1996. Annelida, Borstenwürmer, Polychaeta. 2nd revised ed. The fauna of Germany and adjacent seas with their characteristics and ecology, 58. Gustav Fisher: Jena, Germany. 648 pp.

Heino, J. and Mykrä, H., 2006. Assessing physical surrogates for biodiversity: Do tributary and stream type classifications reflect macroinvertebrate assemblage diversity in running waters? Biological Conservation, 129(3): 418-426.

Hengl, T., Heuvelink, G.B.M.and Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma, 120: 75–93.

Hengl, T., Heuvelink, G. B. M. and Rossiter, D. G., 2007a. About regression-kriging: from theory to interpretation of results. Computers & Geosciences, 33: 1301–1315.

Hengl, T., 2007b. A Practical Guide to Geostatistical Mapping of Environmental Variables. EUR 22904 EN Scientific and Technical Research series, Office for Official Publications of the European Communities, Luxemburg. 143 pp.

Hengl, T., Toomanian, N., Reuter, H. I. and Malakouti, M. J., 2007c. Methods to interpolate soil categorical variables from profile observations: lessons from Iran. Geoderma, 140 (4): 417–427.

Herman, P.M.J., Middelburg, J.J., Van De Koppel, J. and Heip, C., 1999. Ecology of estuarine macrobenthos. Advances in Ecological Research, 29: 195–240.

Hewitt, J.E., Thrush, S.F., Legendre, P., Funnell, G.A., Ellis, J. and Morrison, M., 2004. Mapping of marine soft-sediment communities: integrated sampling for ecological interpretation. Ecological Applications, 14(4): 1203–1216.

Hill, M.O., 1979. TWINSPAN—a FORTRAN program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes. Ecology and Systematics, Cornell University, Ithaca, New York. 31 pp.

Hill, M.O. and Gauch, H.G., 1980. Detrended correspondence analysis: an improved ordination technique. Plant Ecology, 42: 47–58.

Hirzel, A.H., Hausser, J., Chessel, D. and Perrin, N., 2002a. Ecological-Niche Factor Analysis: How to compute Habitat-Suitability Maps without absence data? Ecology, 83(7): 2027–2036.

Hirzel, A., Hausser, J., Perrin, N., 2002b. Biomapper 3.1. Lausanne, Lab. for Conservation Biology. Available online at: (http://www.unil.ch/biomapper).

Hirzel, A.H. and Arlettaz, R., 2003. Environmental-envelope based habitat-suitability models. In 1st Conference on Resource Selection by Animals (ed. B.F.J. Manly). Omnipress, Laramie, USA. p. 67-76.

Hirzel, A. H., Lay, G.L., Helfer, V., Randin, C. and Guisan, A., 2006. Evaluating the ability of habitat suitability models to predict species presences. Ecological Modelling, 199: 142–152.

Holtmann, S.E., Groenewold, A., Schrader, K.H.M., Asjes, J., Craeymeersch, J.A., Duineveld, G.C.A., van Bostelen, A.J. and van der Meer, J., 1996. Atlas of the Zoobenthos of the Dutch Continental Shelf. Ministry of Transport, Public Works and Water Management, Rijswijk, the Netherlands. 243 pp.

Houziaux, J.-S., 2005. The Hinder Banks: yet an important region for the Belgian marine iodiversity? SPSDII project EV/36/45A, mid-term scientific report, Belgian Science olicy Office, 89 pp.

Houziaux, J-S., Haelters, J.and Kerckhof, F., 2007a. Facts from history – The former ecological value of gravel grounds in Belgian marine waters: their importance for biodiversity and relationship with fisheries. ICES SGBIODIV Report 2007: 18-25.

Houziaux, J.-S., Kerckhof, F., Degrendele, K., Roche, M. and Norro, A., 2007b. The Hinder banks: yet an important area for the Belgian marine biodiversity? Brussel (B), Belgian Science Policy (EV/36/45). 123 pp. + Annexes.

Hubert, L.J. and Levin, J.R., 1976. A general statistical framework for assessing categorical clustering in free recall. Psychological Bulletin, 83: 1072-1080.

Hudson, G. and Wackernagel, H., 1994. Mapping temperature using kriging with external drift: theory and an example from Scotland. International Journal of Climatology, 14: 77–91.

Hutchinson, G.E., 1957: Concluding remarks. – Cold Spring Harbour Symposium on Quantitative Biology, 22: 415-427.

Iampietro, P. and Kvitek, R., 2002. Quantitative seafloor habitat classification using GIS terrain analysis: effects of data density, resolution, and scale. In Proceedings of the 22nd Annual ESRI User Conference. San Diego, CA, 8–12 July, unpublished. Available online at: (http://gis.esri.com/library/ userconf/proc02).

Iampietro, P.J. and Kvitek, R.G.M., E., 2005. Recent Advances in Automated Genus-specific Marine Habitat Mapping Enabled by High-resolution Multibeam Bathymetry. Marine Technology Society Journal, 39(9): 83-93.

ICES, 2005. Report of the Working Group on Habitat Mapping (WGMHM), Bremerhaven, Germany, 5–8 April, ICES CM 2005/E:05. 85 pp.

ICES, 2006. Report of the Working Group on Marine Habitat Mapping (WGMHM), 4-7 April, 2006, Galway, Ireland. ICES CM 2006/MHC:05 Ref. FTC, ACE. 132 pp.

ICES, 2007. Report of the Working Group on Marine Habitat Mapping (WGMHM), 17–20 April 2007, Woods Hole, USA. ICES CM 2007/MHC:07. 129 pp.

Jacobs, P., De Ceukelaire, M., Moerkerke, G. and Polfliet, T., 2002. Toelichtingen bij de geologische kaart. 1:50.000. Kaartblad Blankenberge - Westkapelle - Oostduinkerke - Oostende (4-5-11-12). Ministerie van Economische Zaken – Belgische Geologische Dienst, Ministerie van de Vlaamse Gemeenschap - Afdeling Natuurlijke Rijkdommen en Energie, Brussel.

Jenness, J., 2002. Surface Areas and Ratios from Elevation Grid (surfgrids.avx) extension for ArcView 3.x—version 1.2. Jenness Enterprises. Available online at: (http://www.jennessent.com/arcview/grid tools.htm).

Jerosch, K., Schlüter, M. and Pesch, R., 2006. Spatial analysis of marine categorical information using indicator kriging applied to georeferenced video mosaics of the deep-sea Håkon Mosby Mud Volcano. Ecological Informatics, 1(4): 391-406.

Jerosch, K., Lüdtke, A., Schlüter, M. and Ioannidis, G.T., 2007. Automatic content-based analysis of georeferenced image data: Detection of Beggiatoa mats in seafloor video mosaics from the Håkon Mosby Mud Volcano. Computers & Geosciences, 33(2): 202-218.

Jobin, B., Beaulieu, J., Grenier, M., Bélanger, L., Maisonneuve, C., Bordage, D. and Filion, B., 2003. Landscape changes and ecological studies in agricultural regions, Québec, Canada. Landscape Ecology, 18(6): 575-590.

Jongman, R H G., ter Braak, C.J.E. and van Tongeren, O.E.R., 1987. Data analysis in community and landscape ecology. Pudoc, Wageningen. 299 pp.

Journel, A.G., 1993. Geostatistics: roadblocks and challenges. In: Soares, A. (Ed.), Geostatistics Troia ’92. Kluwer Academic Publishers, Dordrecht. pp. 213–224.

Journel, A.G. and Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press Inc, London, UK. 600 pp.

Keytsman, E. and Jones, P.T., 2007. Het klimaatboek - Pleidooi voor een ecologische omslag. EPO, Berchem-Antwerpen. 184 pp.

Kobler, A., Džeroski, S. and Keramitsoglou, I., 2006. Habitat mapping using machine learning-extended kernel-based reclassification of an Ikonos satellite image. Ecological Modelling, 191(1): 83-95.

Kostylev, V.E., Todd, B.J., Fader, G.B.J., Courtney, R.C., Cameron, G.D.M., and Pickrill, R.A. 2001. Benthic habitat mapping on the Scotian Shelf based on multibeam bathymetry, surficial geology and sea floor photographs. Marine Ecology Progress Series, 219: 121–137.

Kostylev, V.E., Courtney, R.C., Robert, G. and Todd, B.J., 2003. Stock evaluation of giant scallop (*Placopecten magellanicus*) using high-resolution acoustics for seabed mapping. Fisheries Research, 60(2-3): 479-492.

Kostylev, V.E., Erlandsson, J., Ming, M.Y. and Williams, G.A., 2005. The relative importance of habitat complexity and surface area in assessing biodiversity: Fractal application on rocky shores. Ecological Complexity, 2(3):272–286.

Kyriakidis, P.C., Kim, J., and Miller, N.L., 2001. Geostatistical mapping of precipitation from rain gauge data using atmospheric and terrain characteristics, Journal of Applied Meteorology, 40(11): 1855-1877.

Lanckneus, J., Van Lancker, V., Moerkerke, G., Van den Eynde, D., Fettweis, M., De Batist, M. and Jacobs, P., 2001. Investigation of the natural sand transport on the Belgian Continental Shelf (BUDGET). Final report. Federal Office for Scientific, Technical and Cultural Affairs OSTC, Brussels. 104 pp. + 87 pp. Annex

Lanszweert, E., 1868. Les bancs d'huitres devant Ostende, Bulletin des Seances de la Societe Royale Malacologique de Belgique III: XVII-XVIII.

Lathrop, R.G., Cole, M., Senyk, N. and Butman, B., 2006. Seafloor habitat mapping of the New York Bight incorporating sidescan sonar data. Estuarine, Coastal and Shelf Science, 68(1-2): 221-230.

Le Bot, S., Van Lancker, V., Deleu, S., De Batist, M. and Henriet, J.P., 2003. Tertiary and Quaternary geology of the Belgian continental shelf. Scientific Support Plan for a Sustainable Development Policy. SPSD II North Sea. Brussels, PPS Science policy publication D/2003/1191/12. 75 pp.

Leecaster, M., 2003. Spatial analysis of grain-size in Santa Monica Bay. Marine Environmental Research, 56: 67–78.

Legendre, P., 1993. Spatial Autocorrelation - Trouble or New Paradigm? Ecology, 74(6): 1659-1673.

Legendre, P. and Legendre, L., 1998. Numerical ecology. 2nd English edition. Elsevier Science BV, Amsterdam. 853 pp.

Legendre, P., Ellingsen, K.E., Bjørnbom, E. and Casgrain, P., 2002. Acoustic seabed classification: improved statistical method. Canadian Journal of Fisheries and Aquatic Sciences, 59: 1085-1089.

Legendre, P., 2003. Reply to the comment by Preston and Kirlin on "Acoustic seabed classification: improved statistical method" Canadian Journal of Fisheries and Aquatic Sciences, 60: 1301-1305.

Le Pape, O., Chauvet, F., Mahévas, S., Lazure, P., Guérault, D. and Désaunay, Y., 2003. Quantitative description of habitat suitability for the juvenile common sole (*Solea solea*, L.) in the Bay of Biscay (France) and the contribution of different habitats to the adult population. Journal of Sea Research, 50(2-3): 139-149.

Levins, R., 1966. The strategy of model building in population biology. American Scientist, 54: 421–431.

Liu, A.C., De Batist, M., Henriet, J.P. and Missiaen, T., 1993. Plio-Pleistocene scour hollows in the Southern Bight of the North Sea. Geologie & Mijnbouw, 71: 195-204.

Lloyd, C.D., 2005. Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain. Journal of Hydrology, 308(1-4): 128-150.

Lu, L., 2005. The relationship between soft-bottom macrobenthic communities and environmental variables in Singaporean waters. Marine Pollution Bulletin, 51: 1034–1040.

Lundblad, E., Wright, D.J., Miller, J., Larkin, E.M., Rinehart, R., Naar, D.F., Donahue, B.T., Anderson, S.M. and Battista, T., 2006. A Benthic Terrain Classification Scheme for American Samoa. Marine Geodesy, 29(2): 89–111.

MacArthur, R. H., 1960. On the Relative Abundance of Species. American Naturalist, 94: 25–36.

Maes, F., Schrijvers, J., Van Lancker, V., Verfaillie, E., Degraer, S., Derous, S., De Wachter, B., Volckaert, A., Vanhulle, A., Vandenabeele, P., Cliquet, A., Douvere, F., Lambrecht, J. and Makgill, R., 2005. Towards a spatial structure plan for sustainable management of the sea. Research in the framework of the BSP programme "Sustainable Management of the Sea" – PODO II, June 2005. 539 pp.

Malvern Instruments. "Mastersizer 2000". 01 April 2008. Available online at: (http://www.malvern.com/LabEng/products/Mastersizer/MS2000/mastersizer2000.htm).

Mandelbrot, 1983. The Fractal Geometry of Nature. New York: W. H. Freeman and Company. 468 pp.

Mantelatto, F.L.M. and Fransozo, A., 1999. Characterization of the physical and chemical parameters of Ubatuba Bay, Northern coast of Sio Paulo State, Brazil. Review of Brasilian Biology, 59(1): 23-31.

Maréchal, R., Henriet, J.-P., Mostaert, F., De Batist, M., Moons, A. and Verschuren, M., 1986. Studie oppervlaktelaag van het Belgisch Continentaal Plat. Seismisch prospectie sector B. Onuitgegeven Rapport, Ministerie voor Economische Zaken. 52 pp.

Marine Biology Section, Ugent – Belgium, 2008. Macrodat. Macrobenthos of the Belgian Part of the North Sea and associated beaches since 1971. Available online at (http://www.vliz.be/imis/imis.php?module=dataset&dasid=1615, accessed on 02/02/2005).

Martinez-Cob, A. and Cuenca, R.H., 1992. Influence of elevation on regional evapotranspiration using multivariate geostatistics for various climatic regimes in Oregon. Journal of Hydrology, 136: 353–380.

McCune, B. and Mefford, M.J., 1999. PC-ORD. Multivariate Analysis of Ecological Data. Version 4.41 MjM Software, Gleneden Beach, Oregon, U.S.A.

McLachlan, A., 1996. Physical factors in benthic ecology: effects of changing sand particle size on beach fauna. Marine Ecology-Progress Series, 131: 205-217.

Meire, P.M., Seys, J., Buijs, J. and Coosen, J., 1994. Spatial and temporal patterns of intertidal macrobenthic populations in the Oosterschelde: are they influenced by the construction of the storm-surge barrier? Hydrobiologia, 282/283: 157–182.

Ménard, F., Gentil, F. and Dauvin, J.C., 1989. Population Dynamics and Secondary Production of *Owenia fusiformis* Delle Chiaje (Polychaeta) from the Bay of Seine (Eastern English Channel). Journal of Experimental Marine Biology and Ecology, 133: 151-167.

Meul, M. and Van Meirvenne, M., 2003. Kriging soil texture under different types of non-stationarity. Geoderma, 112: 217–233.

Milligan, G.W., 1980. An Examination of the Effect of Six Types of Error Perturbation on Fifteen Clustering Algorithms. Psychometrika, 45: 325 -342.

Milligan, G.W. and Cooper, M.C., 1985. An Examination of Procedures for Determining the Number of Clusters in a Data Set. Psychometrika, 50: 159 -179.

Monserud, A.R. and Leemans, R., 1992. Comparing global vegetation maps with the kappa statistic. Ecological Modelling, 62: 275–293.

Moreda-Piñeiro, A., Seco-Gesto, E.M., Bermejo-Barrera, A. and Bermejo-Barrera, P., 2006. Characterization of surface marine sediments from Ría de Arousa estuary according to extractable humic matter content. Chemosphere, 64(5): 866-873.

Mouillot, D., Culioli, J.M. and Chi, T.D., 2002. Indicator species analysis as a test of non-random distribution of species in the context of marine protected areas. Environmental Conservation, 29(3): 385-390.

Miras-Avalos, J.M., Paz-Gonzalez, A., Vidal-Vazquez, E. and Sande-Fouz, P., 2007. Mapping monthly rainfall data in Galicia (NW Spain) using inverse distances and geostatistical methods. Advances in Geosciences, 10: 51-57.

Mishra, D., Narumalani, S., Rundquist, D. and Lawson, M., 2006. Benthic Habitat Mapping in Tropical Marine Environments Using QuickBird Multispectral Data. Photogrammetric Engineering & Remote Sensing, 72(9): 1037-1048.

Mumby, P.J. and Harborne, A.R., 1999. Development of a systematic classification scheme of marine habitats to facilitate regional management and mapping of Caribbean coral reefs. Biological Conservation, 88(2): 155-163.

Murray, J.M.H., Meadows, A. and Meadows, P.S., 2002. Biogeomorphological implications of microscale interactions between sediment geotechnics and marine benthos: a review. Geomorphology, 47: 15–30.

Nate, N.A., Bozek, M.A., Hansen, M.J., Ramm, C.W., Bremigan, M.T. and Hewett, S.W., 2003. Predicting the occurrence and success of walleye populations from physical and biological features of northern Wisconsin lakes. North American Journal of Fisheries Management, 23: 1207–1214.

Naugle, D.E., Higgins, K.F., Estey, M.E., Johnson, R.R. and Nusser, S.M., 2000. Local and landscape-level factors influence black tern habitat suitability. Journal of Wildlife Management, 64: 253–260.

Nicolas, D., Le Loc'h, F., Desaunay, Y., Hamon, D., Blanchet, A. and Le Pape, O., 2007. Relationships between benthic macrofauna and habitat suitability for juvenile common sole (*Solea solea*, L.) in the Vilaine estuary (Bay of Biscay, France) nursery ground. Estuarine Coastal and Shelf Science, 73(3-4): 639-650.

Ojeda, G.Y., Gayes, P.T., Van Dolah, R.F. and Schwab, W.C., 2004. Spatially quantitative seafloor habitat mapping: example from the northern South Carolina inner continental shelf. Estuarine, Coastal and Shelf Science, 59(3): 399-416.

Olsson, J.A. and Andersson, L., 2007. Possibilities and problems with the use of models as a communication tool in water resource management. Water Resources Management, 21: 97–110.

Orlóci, L., 1967. An agglomerative method for the classification of plant communities. Journal of Ecology, 55: 193-206.

Orpin, A.R. and Kostylev, V.E., 2006. Towards a statistically valid method of textural sea floor characterization of benthic habitats. Marine Geology, 225 (1–4): 209–222.

Pan, Y.D., Stevenson, R.J., Hill, B.H., Kaufmann, P.R. and Herlihy, A.T., 1999. Spatial patterns and ecological determinants of benthic algal assemblages in Mid-Atlantic streams, USA. Journal of Phycology, 35: 460–468.

Pannatier, Y., 1996. Variowin-Software for Spatial Data Analysis in 2D. Springer, New York. 91 pp.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Computers & Geosciences, 30: 683-691.

Perrin, P.M., Martin, J.R., Barron, S.J. and Roche, J.R., 2006. A cluster analysis approach to classifying Irish native woodlands. Biology and Environment: Proceedings of the Royal Irish Academy, 106b (3): 261-275.

Pesch, R., Pehlke, H., Jerosch, K., Schröder, W. and Schlüter, M., 2008. Using decision trees to predict benthic communities within and near the German Exclusive Economic Zone (EEZ) of the North Sea. Environmental Monitoring Assessment, 136(1-3): 313-325.

Porter-Smith, R., Harris, P.T., Andersen, O.B., Coleman, R., Greenslade, D. and Jenkins, C.J., 2004. Classification of the Australian continental shelf based on predicted sediment threshold exceedance from tidal currents and swell waves. Marine Geology, 211(1-2): 1-20.

Preston, J. M. and Kirlin, T.L., 2003. Comment on “Acoustic seabed classification: improved statistical method” Canadian Journal of Fisheries and Aquatic Sciences, 60: 1299-1300.

Punj, G. and Stewart, D.W., 1983. Cluster Analysis in Marketing Research: Review and Suggestions for Application. Journal of Marketing Research, 20 (2): 134-148.

R version 2.6.1 (2007-11-26) Copyright (C) 2007 The R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rabaut, M., Guilini, K., Van Hoey, G., Vincx, M. and Degraer, S., 2007. A bio-engineered soft-bottom environment: The impact of *Lanice conchilega* on the benthic species-specific densities and community structure. Estuarine, Coastal and Shelf Science, 75: 525-536.

Rachor, E., Reiss, H., Degraer, S., Duineveld, G.C.A., Van Hoey, G., Lavaleye, M., Willems, W. and Rees, H.L., 2007. Structure, distribution, and characterizing species of the North Sea macro-zoobenthos communities in 2000. p. 46-59. In: Rees, H.L., Eggleton, J.D., Rachor, E., Vanden Berghe, E. (Eds.), Structure and dynamics of the North Sea benthos. ICES Cooperative Research Report No. 288, Copenhagen, Denmark. 258 pp.

Rees, H.L., Eggleton, J.D., Rachor, E. and Vanden Berghe, E. (Eds.), 2007. Structure and dynamics of the Northern Sea benthos. ICES Cooperative Research Report No. 288, Copenhagen, Denmark. 258 pp.

Reinstorf, F., Binder, M., Schirmer, M., Grimm-Strele, J. and Walther, W., 2005. Comparative assessment of regionalisation methods of monitored atmospheric deposition loads. Atmospheric Environment, 39(20): 3661-3674.

Revel, M., Cremer, M., Grousset, F.E. and Labeyrie, L., 1996. Grain-size and Sr-Nd isotopes as tracer of paleo-bottom current strength, Northeast Atlantic Ocean. Marine Geology, 131(3-4): 233-249.

Roff, J.C. and Taylor, M.E., 2000. National frameworks for marine conservation - a hierarchical geophysical approach. Aquatic conservation: marine and freshwater ecosystems, 10: 209-223.

Roff, J.C., Taylor, M.E. and Laughren, J., 2003. Geophysical approaches to the classification, delineation and monitoring of marine habitats and their communities. Aquatic conservation: marine and freshwater ecosystems, 13: 77-90.

Rooper, C.N. and Zimmermann, M., 2007. A bottom-up methodology for integrating underwater video and acoustic mapping for seafloor substrate classification. Continental Shelf Research, 27(7): 947-957.

Rosa-Freitas, M.G., Tsouris, P., Peterson, A.T., Honorio, N.A., Monteiro de Barros, .S., Barros de Aguiar, D., da Costa Gurgel, H., de Arruda, M.E., Vasconcelos, S.D. and Luitgards-Moura, J.F., 2007. An ecoregional classification for the state of Roraima, Brazil. The importance of landscape in malaria biology. Memorias do Instituto Oswaldo Cruz, 102(3): 349-357.

Schelfaut, K., 2005. Defining Marine Landscapes on the Belgian continental shelf as an approach to holistic habitat mapping. MSc Thesis, Universiteit Gent, Belgium, unpublished. 38 pp.

Schelfaut, K., Verfaillie, E., Degraer, S. and Van Lancker, V., 2007. Translation of biological and sedimentological point data towards habitat suitability maps of biological communities and EUNIS level 5 maps. Part 2: From habitat suitability maps of biological communities towards EUNIS level 5 maps, p. 11-16.

Shin, P.K.S., 1982. Multiple discriminant analysis of macrobenthic infaunal assemblages. Journal of Experimental Marine Biology and Ecology, 59: 39–50.

Service, M., 1998. Monitoring benthic habitats in a marine nature reserve. Journal of Shellfish Research, 17(5): 1487-1489.

Skov, H., Humphreys, E., Garthe, S., Geitner, K., Gremillet, D., Hamer, K.C., Hennicke, J., Parner, H. and Wanless, S., 2008. Application of habitat suitability modelling to tracking data of marine animals as a means of analyzing their feeding habitats. Ecological Modelling, 212(3-4): 504-512.

Smith, C.R., Mincks, S. and De Master, D.J., 2006. A synthesis of bentho-pelagic coupling on the Antarctic shelf: Food banks, ecosystem inertia and global climate change. Deep-Sea Research Part II-Topical Studies in Oceanography, 53: 875-894.

Sneath, P.H.A. and Sokal, R.R., 1973. Numerical taxonomy – The principles and practice of numerical classification. W.H. Freeman, San Francisco, USA. 573 pp.

Sokal, R.R., and Rohlf, F.J., 1981. Biometry: the principles and practice of statistics in biological research. W.H. Freeman, New York, New York, USA. 859 pp.

Somaschini, A., 1993. A Mediterranean Fine-sand Polychaete Community and the Effect of the Tube-dwelling Owenia fusiformis on Community Structure. Internationale Revue der gesamten Hydrobiologie und Hydrographie, 78 (2): 219-233.

Sotheran, I.S., Foster-Smith, R.L. and Davies, J., 1997. Mapping of Marine Benthic Habitats Using Image Processing Techniques Within a Raster Based Geographic Information System. Estuarine, Coastal and Shelf Science, 44: 25-31.

Stolk, A., 1996. Chapter 16. Task 6: Analysis and control of internal structures by coring. Subtasks 6.1-6.2-6.3-6.4: Middelkerke Bank area: Sedimentological investigations. In: I. Heyse and G. De Moor (Editors), STARFISH final report. University of Gent (B), Department of Geography, Gent, pp. 16.1-16.33.

Stride, A.H., 1982. Offshore tidal sands, processes and deposits. Chapman and Hall. 213 pp.

Svoray, T., Mazor, S. and Bar (Kutiel), P., 2007. How is shrub cover related to soil moisture and patch geometry in the fragmented landscape of the Northern Negev desert? Landscape Ecology, 22: 105-116.

Trentesaux, A., 1993. Structure et dynamique sédimentaire du Middelkerke Bank, Mer du Nord méridionale. Unpublished Ph.D. Thesis, Université des Sciences et Technologies de Lille, Lille. 229 pp.

Trentesaux, A., Stolk, A. and Berné, S., 1999. Sedimentology and stratigraphy of a tidal sand bank in the southern North Sea. Marine Geology, 159: 253-272.

Turner, S.J., Thrush, S.F., Pridmore, R.D., Hewitt, J.E., Cummings, V.J. and Maskery, M., 1995. Are soft-sediment communities stable? An example from a windy harbour. Marine Ecology-Progress Series, 120: 219–230.

Turney, P., 1995. Data engineering for the analysis of semiconductor manufacturing data. In: Proceedings of the IJCAI-95 Workshop on Data Engineering for Inductive Learning. 10 pp.

Valentine, P.C., Todd, B.J., and Kostylev, V.E. 2005. Classification of marine sublittoral habitats with application to the northeastern North America region. Ed. by P.W. Barnes and J.P. Thomas. In Benthic Habitats and the Effects of Fishing. American Fisheries Society Symposium, 41. American Fisheries Society, Bethesda, Maryland, p. 183–200.

Vanaverbeke, J., Gheskiere, T., Steyaert, M. and Vincx, M., 2002. Nematode assemblages from subtidal sandbanks in the Southern Bight of the North Sea: effect of small sedimentological differences. Journal of Sea Research, 48: 197–207.

Van Beneden, E. 1883. Compte rendu sommaire des recherches entreprises à la Station biologique d'Ostende pendant les mois d'été 1883. Bulletin de l'Académie Royale des Sciences, Litérature et Beaux-Arts de Belgique, 3me Série, T6, no 11, II: 458–483.

Van den Eynde, D., 2004. Interpretation of tracer experiments with fine-grained dredging material at the Belgian Continental Shelf by the use of numerical models. Journal of Marine Systems, 48: 171-189.

Van Hoey, G., Degraer, S. and Vincx, M., 2004. Macrobenthic community structure of soft-bottom sediments at the Belgian Continental Shelf. Estuarine, Coastal and Shelf Science, 59: 599–613.

Van Hoey, G., Vincx, M. and Degraer, S., 2005. Small- to large-scale geographical patterns within the macrobenthic *Abra alba* community. Estuarine, Coastal and Shelf Science, 64(4): 751-763.

Van Hoey, G., Vincx, M. and Degraer, S., 2007. Temporal variability in the *Abra alba* community determined by global and local events. Journal of Sea Research, 58: 144–155.

Van Lancker, V. and Foster-Smith, R., 2007. How do I make a map? In: MESH Guide to Habitat Mapping, MESH Project, 2007, JNCC, Peterborough. 78 pp. Available online at: (http://www.searchmesh.net/)

Van Lancker, V., Du Four, I., Verfaillie, E., Deleu, S., Schelfaut, K., Fettweis, M., Van den Eynde, D., Francken, F., Monbaliu, J., Giardino, A., Portilla, J., Lanckneus, J., Moerkerke, G. and Degraer, S., 2007. Management, research and budgetting of aggregates in shelf seas related to end-users (Marebasse). Brussel (B), Belgian Science Policy (D/2007/1191/49). 139 pp. + DVD GIS@SEA + Habitat Signature Catalogue.

Van Lancker, V., Bonne, W., Velegrakis, A. and Collins, M.B., in press. Aggregate extraction from tidal sandbanks: is dredging with nature an option? Introduction. Journal of Coastal Research.

Verfaillie, E., Van Lancker, V. and Van Meirvenne, M., 2006. Multivariate geostatistics for the predictive modelling of the surficial sand distribution in shelf seas. Continental Shelf Research, 26: 2454-2468.

Verneaux, V., Verneaux, J., Schmitt, A. and Lambert, J.C., 2004. Relationships of macrobenthos with dissolved oxygen and organic matter at the sediment-water interface in ten French lakes. Archiv für Hydrobiologie, 160: 247-259.

Wackernagel, H., 1998. Multivariate Geostatistics. Springer, Berlin. 291 pp.

Ward, J.H., 1963. Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, 58: 236-244.

Weiss, A.D., 2001. Topographic positions and landforms analysis (Conference Poster). ESRI International User Conference. San Diego, CA, 9–13July, unpublished.

White, J. and Fitzpatrick, F., 2007. How do I collect my data? In: MESH Guide to Habitat Mapping, MESH Project, 2007, JNCC, Peterborough. 36 pp. Available online at: (http://www.searchmesh.net/)

Willems, W., Goethals, P., Van den Eynde, D., Van Hoey, G., Van Lancker, V., Verfaillie, E., Vincx, M. and Degraer, S., 2008. Where is the worm Predictive modelling of the habitat preferences of the tube-building polychaete *Lanice conchilega*. Ecological Modelling, 212(1-2): 74-79.

Wilks, S.S., 1932. Certain generalizations in the analysis of variance. Biometrika, 24: 471-494.

Willems, W., Degraer, S., Vincx, M. and Goethals, P., in prep. Habitat Suitability Models as tools in Marine Management: challenges and pitfalls.

Wilson, M.F.J., O'Conell, B., Brown, C., Guinan, J. and Grehan, A., 2007. Multiscale Terrain Analysis of Multibeam Bathymetry Data for Habitat Mapping on the Continental Slope. Marine Geodesy, 30: 3-35.

Wishart, D., 1987. Clustan User Manual, University of St Andrews, Scotland. 250 pp.

Witten, I.H. and Frank, E., 2000. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann Publishers, San Francisco, USA. 416 pp.

Wood, J., 1996. The Geomorphological Characterisation of Digital Elevation Models. PhD Thesis, University of Leicester. 185 pp.

Wood, J., 2005. Landserf Version 2.2. Available online at: (www.landserf.org).

Worth, A.P. and Cronin, M.T.D., 2003. The use of discriminant analysis, logistic regression and classification tree analysis in the development of classification models for human health effects. Journal of Molecular Structure—Theochem, 622: 97–111.

Wu, R.S.S. and Shin, P.K.S., 1997. Sediment characteristics and colonization of soft-bottom benthos: a field manipulation experiment. Marine Biology, 128: 475–487.

Xirouchakis, S.M. and Mylonas, M., 2005. Selection of breeding cliffs by Griffon Vultures Gyps fulvus in Crete (Greece). Acta Ornithologica, 40: 155–161.

Young, A., 2007. MESH maps come online. A framework to support seabed habitat mapping. Hydro International, 11: 6–9.

Ysebaert, T. and Herman, P.M.J., 2002. Spatial and temporal variation in benthic macrofauna and relationships with environmental variables in an estuarine, intertidal soft-sediment environment. Marine Ecology-Progress Series, 244: 105-124.

Zacharias, M.A., Morris, M.C. and Howes, D.E., 1999. Large scale characterization of intertidal communities using a predictive model. Journal of experimental marine biology and ecology, 239(2): 223-242.

Zacharias, M.A. and Roff, J.C., 2000. A hierarchical ecological approach to conserving marine biodiversity. Conservation Biology, 14 (5): 1327-1334.

Zajac, R.N., Lewis, R.S., Poppe, L.J., Twichell, D.C., Vozarik, J. and DiGiacomo-Cohen, M.L., 2000. Relationships among sea-floor structure and benthic communities in Long Island Sound at regional and benthoscape scales. Journal of Coastal Research, 16(3): 627-640.

Zajac, R.N., Lewis, R.S., Poppe, L.J., Twichell, D.C., Vozarik, J. and DiGiacomo-Cohen, M.L., 2003. Responses of infaunal populations to benthoscape structure and potential importance of transition zones. Limnology and Oceanography, 48(2): 829-842.

Zharikov, Y., Skilleter, G.A., Loneragan, N.R., Taranto, T. and Cameron, B.E., 2005. Mapping and characterising subtropical estuarine landscapes using aerial photography and GIS for potential application in wildlife conservation and management. Biological Conservation, 125(1): 87-100.

List of publications

Peer-reviewed publications

Cherlet, J., Besio, G., Blondeaux, P., Van Lancker, V., Verfaillie, E. and Vittori, G., 2007. Modeling sand wave characteristics on the Belgian Continental Shelf and in the Calais-Dover Strait. Journal of Geophysical Research, 112: C06002.

Degraer, S., Verfaillie, E., Willems, W., Adriaens, E., Van Lancker, V. and Vincx, M., 2008. Habitat suitability as a mapping tool for macrobenthic communities: An example from the Belgian part of the North Sea. Continental Shelf Research, 28(3): 369-379.

Derous, S., Agardy, T., Hillewaert, H., Hostens, K., Jamieson, G., Lieberknecht, L., Mees, J., Moulaert, I., Olenin, S., Paelinckx, D., Rabaut, M., Rachor, E., Roff, J., Stienen, E.W.M., Van der Wal, J.T., Van Lancker, V., Verfaillie, E., Vincx, M., Weslawski, J.M. and Degraer, S., 2007. A concept for biological valuation in the marine environment. Oceanologia, 49(1): 99-128.

Derous, S., Courtens, W., Deckers, P., Deneudt, K., Hostens, K., Moulaert, I., Paelinckx, D., Rabaut, M., Roff, J.C., Stienen, E.W.M., Van Lancker, V., Verfaillie, E., Vincx, M. and Degraer, S. (submitted to Aquatic Conservation: Marine and Freshwater Ecosystems) Biological valuation: Guidelines for a transparent and generally applicable protocol for the marine environment.

Derous, S., Courtens, W., Deckers, P., Deneudt, K., Forero, C.E., Hostens, K., Moulaert, I., Stienen, E.W.M., Vanden Eede, S., Van Lancker, V., Verfaillie, E., Vincx, M. and Degraer, S. (submitted to Ecological Applications) Application of the protocol for marine biological valuation to selected case study areas.

Verfaillie, E., Van Meirvenne, M. and Van Lancker, V., 2006. Multivariate geostatistics for the predictive modelling of the surficial sand distribution in shelf seas. Continental Shelf Research, 26(19): 2454-2468.

Verfaillie, E., Du Four, I., Van Meirvenne, M. and Van Lancker, V., accepted. Geostatistical modelling of sedimentological parameters using multiscale terrain variables: application along the Belgian Part of the North Sea. International Journal of Geographical Information Science.

Verfaillie, E., Degraer, S., Schelfaut, K., Willems, W. and Van Lancker, V. (submitted to Estuarine, Coastal and Shelf Science) A protocol for classifying ecologically relevant marine landscapes, a statistical approach.

Verfaillie, E., Degraer, S., Du Four, I., Rabaut, M., Willems, W. and Van Lancker, V. (submitted to Estuarine, Coastal and Shelf Science) The relevance of ecogeographical variables for marine habitat suitability modelling of *Owenia fusiformis*.

Willems, W., Goethals, P., Van den Eynde, D., Van Hoey, G., Van Lancker, V., Verfaillie, E., Vincx, M. and Degraer, S., 2008. Where is the worm? Predictive modelling of the habitat preferences of the tube-building polychaete *Lanice conchilega*. Ecological Modelling, 212(1-2): 74-79.

Books

Maes, F., Calewaert, J.-B., Schrijvers, J., Van Lancker, V., Verfaillie, E., Degraer, S., Derous, S., De Wachter, B., Volckaert, A., Vanhulle, A. and Vandenabeele, P., 2007. Chapter 6: GAUFRE: towards a spatial structure plan for the sustainable management of the Belgian part of the North Sea, pp 186-258. In: F. Maes and J.-B. Calewaert (Editors), Science and Sustainable Management of the North Sea: Belgian case studies. Gent, Academia Press Wetenschappelijke Uitgeverij. 314 pp.

Van Lancker, V., Deleu, S., Bellec, V., Du Four, I., Schelfaut, K., Verfaillie, E., Fettweis, M., Van Den Eynde, D., Francken, F., Monbaliu, J., Giardino, A., Portilla, J., Lanckneus, J., Moerkerke, G. and Degraer, S., 2007. Chapter 4. Marebasse: bridging gaps towards other disciplines and endusers, pp. 88-141. In: F. Maes and J.-B. Calewaert (Editors), Science and Sustainable Management of the North Sea: Belgian case studies. Gent, Academia Press Wetenschappelijke Uitgeverij. 314 pp.

Scientific reports

Derous, S., Verfaillie, E., Van Lancker, V., Courtens, W., Stienen, E.W.M., Hostens, K., Moulaert, I., Hillewaert, H., Mees, J., Deneudt, K., Deckers, P., Cuvelier, D., Vincx, M. and Degraer, S., 2007. A biological valuation map for the Belgian part of the North Sea: BWZee, Final report, Research in the framework of the BELSPO programme “Global change, ecosystems and biodiversity” – SPSD II, March 2007. 99 pp. (+ Annexes).

Maes, F., Schrijvers, J., Van Lancker, V., Verfaillie, E., Degraer, S., Derous, S., De Wachter, B., Volckaert, A., Vanhulle, A., Vandenabeele, P., Cliquet, A., Douvere, F., Lambrecht, J. And Makgill, R., 2005. Towards a spatial structure plan for sustainable management of the sea. Research in the framework of the BSP programme "Sustainable Management of the Sea" – PODO II, June 2005. 539 pp.

Van Lancker, V., Deleu, S., Bellec, V., Le Bot, S., Verfaillie, E., Fettweis, M., Van den Eynde, D., Francken, F., Pison, V., Wartel, S., Monbaliu, J., Portilla, J., Lanckneus, J., Moerkerke, G. and Degraer, S., 2004. Management, research and budgeting of aggregates in shelf seas related to end-users (Marebasse). Scientific Report Year 2. Belgian Science Policy, SPSDII North Sea. 144 pp.

Van Lancker, V., Deleu, S., Bellec, V., Du Four, I., Verfaillie, E., Fettweis, M., Van den Eynde, D., Francken, F., Monbaliu, J., Giardino, A., Portilla, J., Lanckneus, J., Moerkerke, G. and Degraer, S., 2005. Management, research and budgeting of aggregates in shelf seas related to end-users (Marebasse). Scientific Report Year 3. Belgian Science Policy. 103 pp.

Van Lancker, V. and Verfaillie, E., 2005. Geophysical zonation. In: Maes, F., Schrijvers, J., Van Lancker, V., Verfaillie, E., Degraer, S., Derous, S., De Wachter, B., Volckaert, A., Vanhulle, A., Vandenabeele, P., Cliquet, A., Douvere, F., Lambrecht, J. and Makgill, R., 2005. Towards a spatial structure plan for sustainable management of the sea. Research in the framework of the BSP programme "Sustainable Management of the Sea" – PODO II, June 2005. 539 pp.

Van Lancker, V. and Verfaillie, E., 2005. Homogeneous zones. In: Maes, F., Schrijvers, J., Van Lancker, V., Verfaillie, E., Degraer, S., Derous, S., De Wachter, B., Volckaert, A., Vanhulle, A., Vandenabeele, P., Cliquet, A., Douvere, F., Lambrecht, J. and Makgill, R., 2005. Towards a spatial structure plan for sustainable management of the sea. Research in the framework of the BSP programme "Sustainable Management of the Sea" – PODO II, June 2005. 539 pp.

Van Lancker, V., Du Four, I., Verfaillie, E., Deleu, S., Schelfaut, K., Fettweis, M., Van den Eynde, D., Francken, F., Monbaliu, J., Giardino, A., Portilla, J., Lanckneus, J., Moerkerke, G. and Degraer, S., 2007. Management, research and budgetting of aggregates in shelf seas related to end-users (Marebasse). Brussel (B), Belgian Science Policy (D/2007/1191/49). 139 pp. + DVD GIS@SEA + Habitat Signature Catalogue.

Verfaillie, E. and Van Lancker, V., 2005. Cables and pipelines. In: Maes, F., Schrijvers, J., Van Lancker, V., Verfaillie, E., Degraer, S., Derous, S., De Wachter, B., Volckaert, A., Vanhulle, A., Vandenabeele, P., Cliquet, A., Douvere, F., Lambrecht, J. and Makgill, R., 2005. Towards a spatial structure plan for sustainable management of the sea. Research in the framework of the BSP programme "Sustainable Management of the Sea" – PODO II, June 2005. 539 pp.

Verfaillie, E., Deleu, S. and Van Lancker, V., 2005. Sand and gravel extraction. In: Maes, F., Schrijvers, J., Van Lancker, V., Verfaillie, E., Degraer, S., Derous, S., De Wachter, B., Volckaert, A., Vanhulle, A., Vandenabeele, P., Cliquet, A., Douvere, F., Lambrecht, J. and Makgill, R., 2005. Towards a spatial structure plan for sustainable management of the sea. Research in the framework of the BSP programme "Sustainable Management of the Sea" – PODO II, June 2005. 539 pp.

Verfaillie, E., Du Four, I. and Van Lancker, V., 2005. Dredging and dumping operations. In: Maes, F., Schrijvers, J., Van Lancker, V., Verfaillie, E., Degraer, S., Derous, S., De Wachter, B., Volckaert, A., Vanhulle, A., Vandenabeele, P., Cliquet, A., Douvere, F., Lambrecht, J. and Makgill, R., 2005. Towards a spatial structure plan for sustainable management of the sea. Research in the framework of the BSP programme "Sustainable Management of the Sea" – PODO II, June 2005. 539 pp.

Verfaillie, E., Du Four, I., Degraer, S. and Van Lancker, V., 2007. Acoustic habitat modelling for the mapping of biological communities. In: Van Lancker, V., Du Four, I., Verfaillie, E., Deleu, S., Schelfaut, K., Fettweis, M., Van den Eynde, D., Francken, F., Monbaliu, J., Giardino, A., Portilla, J., Lanckneus, J., Moerkerke, G. and Degraer, S., 2007. Management, research and budgetting of aggregates in shelf seas related to end-users (Marebasse). Brussel (B), Belgian Science Policy (D/2007/1191/49). 139 pp. + DVD GIS@SEA + Habitat Signature Catalogue.

Other

Henriet, J.-P. and Verfaillie, E., 2007. Beelden van de diepzee. De Grote Rede, 19: 2-6.

[1] Convention on Wetlands of International Importance, Ramsar, 2 February 1971. Available online at: (www.ramsar.org).

[2] Convention for the protection of the marine environment of the North-East Atlantic, Paris, 22 September 1992. Available online at: (www.ospar.org).

[3] Convention on Biological Diversity, Rio de Janeiro, 5 June 1992. Available online at: (www.cbd.int).

[4] United Nations. Report of the World Summit on Sustainable Development. Johannesburg, South Africa, 26 August- 4 September 2002. A/Conf.199/20. Available online at: (www.unmillenniumproject.org/documents/131302_wssd_report_reissued.pdf).

[5] CBD-COP7 Decision VII/5 Marine and coastal biological diversity. Available online at: (www.cbd.int/doc/decisions/COP-07-dec-en.pdf).

[6] Council Directive 79/409/EEC of 2 April 1979 on the conservation of wild birds, Official Journal L 103 , 25/04/1979. Available online at: (http://ec.europa.eu/environment/nature/home.htm).

[7] Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora, Official Journal L 206 , 22/07/1992. Available online at: (http://ec.europa.eu/environment/nature/home.htm).

[8] Communication from the Commission. Halting the Loss of Biodiversity by 2010 - and beyond. Sustaining ecosystem services for human well-being, COM (2006)216 final. Available online at: (http://eurlex.europa.eu/LexUriServ/site/en/com/2006/com2006_0216en01.pdf).

[9] Law of 20 January 1999 on the protection of the marine environment (Wet ter bescherming van het mariene milieu in de zeegebieden onder de rechtsbevoegdheid van België), published in the Belgian Official Journal 12 March 1999. Amended by Law of 17 September 2005, Belgian Official Journal 13 October 2005.

[10] Royal Decree of 14 October 2005 on the conditions, conclusion, implementation and termination of user agreements and the drawing up of policy plans for the marine protected areas in marine areas under Belgian jurisdiction (KB van 14 oktober 2005 betreffende de voorwaarden, sluiting, uitvoering en beëindiging van gebruikersovereenkomsten en het opstellen van beleidsplannen voor de beschermde mariene gebieden in de zeegebieden onder de rechtsbevoegdheid van België), Belgian Official Journal 31 October 2005.

[11] Same reference as endnote 9.

[12] Same reference as endnote 10.

[13] Same reference as endnote 6.

[14] Same reference as endnote 7.

[15] Raad van State, Afdeling Bestuursrechtspraak. Arrest nr. 179.254 van 1 februari 2008 in de zaak A. 169.087/VII-35.198. In zake : 1. de n.v. Electrabel, 2. de n.v. Ondernemingen Jan De Nul, samen vormend de tijdelijke vereniging Electrabel Ondernemingen Jan De Nul, 3. de n.v. Electrabel Seanergy, die woonplaats kiezen bij advocaten T. Vandenput en P. De Maeyer tegen : de Belgische Staat, vertegenwoordigd door de minister van Mobiliteit, die woonplaats kiest bij advocaat L. Denys.

[16] Royal Decree of 5 March 2006 for the establishment of a specific marine reserve in the marine areas under Belgian jurisdiction and for the amendment of Royal Decree of 14 October 2005 (KB van 5 maart 2006 tot instelling van een gericht marien reservaat in de zeegebieden onder de rechtsbevoegdheid van België en tot wijziging van het koninklijk besluit van 14 oktober 2005 tot

instelling van speciale beschermingszones en speciale zones voor natuurbehoud in de zeegebieden onder de rechtsbevoegdheid van België), Belgian Official Journal 27 March 2006.

[17] Same reference as endnote 10.

[18] Same reference as endnote 16.

[19] Same reference as endnote 2.