

Thesis for the Degree of Doctor of Philosophy

Spatial distribution and conservation planning of seabed biological diversity

Genoveva Gonzalez Mirelis, 2011

Faculty of Science



UNIVERSITY OF GOTHENBURG

Department of Marine Ecology – Tjärnö
45296 Strömstad
Sweden

© Genoveva Gonzalez Mirelis 2011

All rights reserved. No part of this publication may be reproduced or transmitted, in any form or by any means, without written permission.

Cover illustration by Autun Purser

Printed by Ale Tryckteam AB, Bohus, Sweden 2011

ISBN 978-91-628-8391-1

LIST OF PAPERS

This thesis is a summary of the following papers ¹:

- Paper **I**. Gonzalez-Mirelis, G., Bergström, P., Lundälv, T., Jonsson, L. and Lindegarth, M. 2009. Mapping the benthos: Spatial patterns of seabed-dwelling megafauna in a Swedish Fjord, as derived from opportunistic video data. *Marine Biodiversity*, 39 (4) p. 291-302
- Paper **II**. Gonzalez-Mirelis, G., Bergström, P., and Lindegarth, M. 2011. Interaction between classification detail and prediction of community types: implications for predictive modelling of benthic biotopes. *Marine Ecology Progress Series*, 432 p. 31-44
- Paper **III**. Gonzalez-Mirelis, G., and Lindegarth, M. 2011. Predicting the distribution of out-of-reach biotopes by conditional inference: a case-study from a Swedish Fjord. *Ecological Applications*. Accepted for publication
- Paper **IV**. Gonzalez-Mirelis, G., Nilsson, P. 2011. A test of the value of modelled biotopes as surrogates for species diversity in the planning of a Swedish marine National Park. Manuscript

¹Papers **I** and **II** are reprinted with the kind permission of Elsevier and Inter-Research

Contents

1	Brief introduction	2
2	On the spatial distribution of species	2
2.1	Species distribution modelling and habitat mapping	4
2.2	Mapping the benthic landscape	6
2.3	The case-study	8
3	Considerations of the data model	10
3.1	Measuring spatial pattern	10
3.2	The mapping units	12
4	Mapping the distribution of seabed diversity through biotope modelling	12
5	Representing the distribution of seabed diversity using biotopes as surrogates	16
6	Conclusions and outlook	18
7	References	19
	Acknowledgements	24
	List of Figures	
1	Map of the study area and ROV-surveyed sites	11
2	Map of biotopes	15
	List of Tables	
1	Description of biotopes	13
2	Fishing pressure by biotope class	16

1. Brief introduction

Insight into the spatial distribution of the various components of biological diversity, not least species, paves the way to addressing fundamental ecological questions aiming to explain the abundance (or at a minimum, presence) of organisms in an ecosystem, a question of intrinsic scientific value, apart from any applications that such knowledge may have. A spatial ecologist with unlimited time and budget, and fuelled by a genuine commitment to understanding the workings of Nature in their own right, would during his or her career strive to complete at least one cycle of the following three-stage process: detect and measure spatial pattern, relate this to ecological phenomena, and formulate and test hypotheses regarding the ecological processes responsible for generating those patterns. As a spatial ecologist with the restrictions imposed on me mostly by time, funding, and my own particular interests, I have within this PhD program attempted to: detect and measure spatial pattern at the local scale across the seafloor of a Fjord ecosystem, and describe ecological phenomena associated with those patterns. However, rather than continue down the path of theoretical ecology and try to understand the processes shaping the spatial variability of my study system, I have instead put on the applied (spatial, marine) ecologist hat, and attempted to use this knowledge to solve practical problems that society faces. I propose a unified methodological framework for documenting the distribution of diversity and implementing the generated knowledge into conservation.

Throughout this thesis I will use the term '(bio)diversity' in a rather inclusive sense to refer to the full complement of species as well as all other levels of organization, of an ecosystem. It is therefore more closely related with the composition of ecosystems than with emergent properties such as taxon richness or other diversity indices.

The most important conclusions reached throughout the duration of this PhD and summarized in this Thesis are preceded by an asterisk symbol (*) so as to help the reader extract the take-home messages without isolating those from their context.

2. On the spatial distribution of species

Knowing how species are distributed over the biosphere is also a fascinating question, with a long history of attempts to disentangle its details, whether by scientists, harvesters and gatherers, nature lovers, etc. The empirical fact that species are distributed neither at random nor uniformly across space forms the very foundation of spatial ecology and hence this Thesis. Various theories have been developed to explain why some areas (e.g. ecosystems) have a particular species composition, and building from that why a species occurs and thrives in some areas and not in others. Island Biogeography theory was the first proposed (MacArthur & Wilson, 1963; 1967). This theory combined the well-known species-area relationship with the counteracting effects of rates of colonization and extinction to explain why some islands hosted many species and others few. Hubbell (2001) extended this theory and proposed a Unified Neutral hypothesis to explain the relative species abundances in a metacommunity, and all its component local communities, combining island biogeography principles (speciation and extinction) with community ecology concepts (dispersal and ecological drift) considering individuals with equivalent requirements and no competition, but it has received criticism for potentially failing to be a parsimonious model. The habitat-association approach to ecology tackles the problem from a different standpoint, contributing also a number of theories. Gradient Analysis theory (ter Braak & Prentice,

1988) is defined as the interpretation of community composition in terms of species' responses to environmental gradients and is an important part of ecological research. The niche concept has been notoriously successful in the development of ecological quantitative, predictive methods, if applied in a wide variety of meanings. The concept of 'environmental niche' (e.g. Austin 1992) is particularly useful, as it is defined from the point of view of the environmental requirements of a species. This can in turn be further divided into 'fundamental niche' and 'realized niche' (Hutchinson, 1957; Pulliam, 2000), where the former is a view of species occupying all of their suitable habitat and the latter includes the possibility of species being excluded from a part of their fundamental niche by biotic interactions. Guisan & Thuiller (2005) provide a useful overview of these concepts. Empirical data support the fundamental postulate that species respond to environmental gradients, however complex it may be to elucidate the particular details (e.g. type of ecological response) of each species-environmental variable combination, as well as to assemble those into multidimensional spaces for more comprehensive and realistic descriptions of species-environment relationships. But work is underway (e.g. Heikkinen & Mäkipää 2010; Normand et al. 2009) and slowly, theories are refined accordingly.

Besides its value in advancing fundamental science, knowledge of the spatial distribution of biota is also on high demand among conservation practitioners and resource managers. This type of information enables assessing the distribution of biological resources (e.g. fisheries), or monitoring habitat change. With complete knowledge of the distribution of biodiversity it becomes possible to identify e.g. unique ecological features (at some nominated scale) by means of systematic comparisons of ecosystem structure and composition across different localities, and hence confidently select places worthy of special management provisions on the grounds of containing 'interesting' (in this example, unique) biological features. In the context of conservation planning, the most urgent question that needs answering follows directly from the intergovernmental mandate of the Convention on Biological Diversity, revisited in the Seventh Conference of the Parties, to establish 'ecologically representative national and regional systems of protected areas' (Secretariat of the Convention on Biological Diversity, 2004). The mandate itself rests on a number of assumptions that can be summarized as follows: (1) all biodiversity occurring throughout the planet needs to be preserved; (2) not all instances of biodiversity (e.g., all localities where a species occurs) can be preserved, because other governmental objectives (economic expansion, etc.) may be incompatible with the objective of biodiversity preservation (for a given site, the two cannot be satisfied simultaneously). Precisely because biodiversity occurs in patterns, a certain amount of area can in principle be identified encompassing at least one instance of all components of biodiversity (and thus leaving only replicates off its boundaries) and assuming that its maintenance would ensure the long-term persistence of the biodiversity within it, then the job would be done.

The reality however is that species distributions are generally poorly known, placing a major impediment in addressing directly the questions posed above. The daunting task of biodiversity distribution mapping relies heavily on extensive observational data, collected across ecosystems, scales, and levels of biological organization. In addition, the cost associated with conducting such sampling programs increases dramatically in remote or inaccessible areas, leaving vast expanses of area with significant data shortages, as in the case of the seabed.

Pioneer benthic ecologist Petersen (1924) expressed this problem by stating that “‘botanizing’ out at sea is a very expensive affair” (Petersen 1924, p. 688). Embedded in this accurate statement is not only the notion of the high cost associated with acquiring survey data in the benthos, but also a regret that how little data are available cannot compare in quality and quantity with those that plant ecologists have readily at their disposal. Any methods to infer spatial pattern from ecological geospatial data are therefore forced to be highly data-efficient.

The present Thesis is concerned with making the best possible use of available methods to document the spatial patterns of biological diversity in an environment with data limitations. Specifically, I ask the following two questions, how to map the distribution of benthic diversity, subject to the condition that the purpose served by the generated map should be of supporting spatial planning? and how to determine if an area (interpreted as a collection of sites) adequately represents overall biodiversity so that it can be prioritized in the face of competing demands for space? In papers **I** and **II** issues underpinning methods (i.e., the data model) are addressed. Papers **III** and **IV** deal respectively with the above two questions directly. I will argue that these questions are intrinsically related to one another.

2.1 Species distribution modelling and habitat mapping

Species distribution modelling is by and large an extension of the habitat-association approach, by which biological populations, whether marine or terrestrial, are seen to distribute themselves in space according to habitat gradients, leading to community zonation when taken as a whole. Models of environmental variables against species abundance (e.g. statistical regression) were initially used exclusively to understand species’ ecology. Once made spatially-explicit, their value to predict species presence or abundance at locations for which only the predictor environmental variables were known made species distribution models become dominant as an interpolation tool. Species distribution modelling has been noted to have become disconnected from ecological theory (e.g. Austin 2007) and is currently predominantly used as a tool for mapping the suitable habitat of a species or a community. Modelling is of course a known way of addressing the poor data coverage of observational methods or, in other words, of making efficient use of scarce data.

Species distribution models are regression and classification methods, statistical or otherwise, for the quantitative prediction of species occurrence across space. Distribution modelling is typically conducted around a grid that is overlaid on data layers (the predictor or explanatory variables), so that a prediction can be obtained for every individual grid cell, including those where no *in situ* information on the biological composition is available and only environmental data exists, and full coverage is thus achieved.

While distribution modelling got off the ground by virtue of statistical regression, parametric methods (e.g. Generalized Linear Models [GLM]), machine learning is beginning to dominate as an automated, predictive method. Machine learning (which includes among others all methods using recursive partitioning, such as Classification and Regression Trees [CART], random forests, etc.) comprises non-parametric techniques that induce rules directly from observations making use of an algorithmic framework (Guisan & Zimmermann, 2000) and is better suited to work with ecological and environmental data, particularly of a spatial nature, for several reasons. Machine

learning methods are able to model high dimensional data with hierarchical interactions and non-linear relationships. Importantly, they are free from many of the constraints that apply to statistical methods, most notably, a need to choose a shape for the expected response curves between species performance and each environmental gradient, prior to modelling. Additionally, machine learning is easily integrated both with Geographic Information Systems (GIS) and relational databases to carry out predictive mapping. Connecting a working data notification and management system to the predictive models is of crucial importance for a practical use of machine learning on ecological data (Furlanello, 2004).

The process of distribution modelling consists of the following stages:

- Conceptualization
- Data gathering and choice of data model
 - Field surveys (response data)
 - Preparation of explanatory variables (predictor data)
- Model fitting
- Model evaluation
- Spatial predictions
- Assessment of model applicability

If the sample base obtained from the field surveys were fully representative of the distribution of the species to be modelled, and if all biotic and abiotic phenomena surrounding and potentially interacting with it were adequately represented by the predictor variables, then the modelled overall ecological response obtained from the model would summarize all systematic variation in its aggregated performance, and the spatial predictions from the model would be in full accordance with its real distribution, except for stochastic variation (R. Halvorsen pers. comm.). Notably, the only two prerequisite conditions for an adequate model are related with the data model, underlining that the power and reliability of empirical models depend strongly on the data. The output from such model would be by definition the fundamental niche of a species.

Given the hierarchical nature of ecosystems, any level of biological organization from single species to broad ecosystems can be described in terms of its spatial distribution patterns and hence be depicted in map form. Species distribution modelling can easily be scaled-up to functional groups of species, or communities to obtain maps of community distribution, and even further up for maps of habitat types (Guisan & Thuiller, 2005). Such integration allows better use to be made of all available data for all surveyed species (Ferrier et al., 2002). Ferrier & Guisan (2006) reviewed approaches to modelling at the community level and classified them into three groups, depending on whether species data was assembled into communities before modelling, after modelling, or during modelling. The most common strategy has been the first, whereby site groups generated by numerical classification of survey data are modelled and extrapolated across an entire region of interest. The output from such model is a prediction of which group type each grid cell is most likely to be. But the advantages of community level modelling have not been enough, it appears, to have moved the approach into a stage of more uncritical, routine application.

In turn, habitat maps can support geospatial ecological data by providing local and regional models that capture the spatial distribution of user-defined or typology-derived classes (e.g., forest types) which can in turn be used to understand or predict species distributions. Habitat types can be mapped by any means allowing collection of data across vast areas in a cost-efficient manner (e.g. remote sensing) and does not necessarily rely on intensive data collection and modelling because the detail of the mapping units is all that much coarser.

2.2 Mapping the benthic landscape

In Petersen's times, the classification and mapping of seabed communities on sublittoral environments was limited to what little could be inferred from small numbers of scattered point samples, yielding maps with large gaps of information, or restricted to very small scales. The advent of acoustic technologies (namely sidescan sonar, and single- and multibeam echo sounders) rendered it possible to acquire high-resolution, full-coverage imagery of the seafloor over extensive areas, beyond the limit of light penetration. As a result, the geophysical attributes of the terrain that can be derived from its acoustic properties can now form the basis for a classification of e.g. depth and seafloor texture. These advances endowed benthic researchers with the ability to carry out spatially continuous, 'wall-to-wall' mapping, following in the steps of terrestrial remote sensing science. Also, it has been noted that the zonation that emerges in epibenthic megafaunal communities (or megabenthos) is most extreme (Howell, 2010). Sessile benthic organisms are also useful for habitat characterization because substrate is critical in determining their aggregated performance and because they are spatially fixed, they also become indicative of environmental conditions of the adjacent seafloor (Kostylev et al., 2001). As far as mapping is concerned, the epibenthic megafauna is the vegetation of the benthic landscape. But with ecological data being sparse at best, and remotely sensed data bearing mostly an indirect (and not fully disentangled) relationship with biological composition, the challenge of mapping the various components of benthic diversity (from genes to ecosystems) is largely a methodological one.

The integration of data from multiple surveying techniques, typically including one or more full-coverage layers depicting features visible at medium to large spatial scales (e.g., bedrock outcrops, large depth gradients, etc.), and at least one dataset from some *in situ* benthic survey technique providing insight into small-scale variation occurring mostly at the biological level, nested within the former, has proven the most rewarding technique for mapping extensive areas of the seafloor. Hereafter we refer as 'survey data' to any data collection activity delivering geo-referenced information on the value of ecological variables, and/or variables relating to the sediment or bedrock, whether quantitative (e.g., species abundance, granulometry information) or qualitative (e.g. 'presence of sessile invertebrates', 'presence of mobile sediment'), where the sampling device (e.g. grab, benthic sled, video camera) is located directly at the site from where information is being gathered.

Two general approaches can be distinguished on the basis of the role that survey data plays in the mapping process: a 'top-down approach', where survey data is used merely for ground-truthing purposes, and the process is driven by the acoustic patterns; and a 'bottom-up approach', where biotic patterns, as inferred from the survey data, drive the definition and mapping of classes. Even though the final result is equivalent, a

thematic map showing the distribution of classes that echo biotic patterns of the seafloor to a greater or lesser degree, the path followed is fundamentally different. At the most abstract level it can be described as (a) an attempt to find the attributes of polygons of (mostly) known boundaries, in the top-down case, or (b) an attempt to find the boundaries of polygons of (mostly) known attributes, in the bottom-up case. Stressing the importance of whether boundaries are known or not, the former approach is also known as 'supervised', and the latter, as 'unsupervised'.

In the top-down approach, first, a classification technique based on patterns identified from the remotely-sensed data, usually acoustic imagery, is used to derive homogeneous and distinct regions, often referred to as acoustic habitats. Techniques range from visual interpretation to highly sophisticated classification algorithms (e.g., Lamarche et al. 2011). These are essentially used as a framework within which reference sites are defined. Samples of *in situ* data are then collected from all the detected regions, or the reference sites, so as to validate the classified habitats (see Brown et al. 2002; 2004; Freitas et al. 2003; Jordan et al. 2005; Kloser et al. 2001; Kostylev et al. 2001), and classes are occasionally merged if they can be proven to have non-distinct faunas.

Remote sensing by hydroacoustics is highly effective in classifying habitats over large areas of seabed. The approach effectively reveals boundaries created by discontinuities in substrate types, which in turn give rise to sharp changes in community composition. But the more gradual and subtle changes which may emerge in response to factors other than substrate, are wholly overlooked. A more serious issue is that assemblages may be identified from the survey data which have no corresponding acoustic class (Brown et al., 2005; Freitas et al., 2003). The problem is that the spatial detail of the map is limited by the scale at which the acoustic regions are defined, and although boundaries can be modified to a limited extent on the basis of the biotic patterns (which is possible only at a scale defined by the distance between reference sites, i.e., the same scale at which acoustic habitats are defined), the resolution can only decrease as a result of classes being dropped, but never increase.

Indeed, the question is raised more and more whether acoustically derived habitats are a good representation of the patterns of variability of epibenthic communities (Eastwood et al., 2006; Hewitt et al., 2004). Stevens & Connolly (2004) concluded that the ability of abiotic surrogates to predict patterns of biological similarity was indeed poor. Parry et al. (2003) detected a nested hierarchy of spatial structure within the megafaunal assemblage of a large, apparently homogeneous, soft-bottom habitat unit. Because the subset of biota of interest is precisely the epibenthic megafauna, dissatisfaction prompted the development of improved mapping methods.

The bottom-up approach emerged in response to this call, and Field et al. (1982) summarize it as 'letting the species tell their story' (p. 38, citing from a previous paper co-authored by J. G. Field). In this approach, the mapping units are defined on the basis of multivariate species patterns (e.g. peaks of homogeneity within the continuous gradient of faunal composition, Brown et al. 2002), which are in turn assumed to define sets of distinct environmental factors (Kostylev et al., 2001). Eastwood et al. (2006) compared top-down versus bottom-up approaches to classifying, and mapping seabed assemblages and found that, when "the seabed comprises relatively homogeneous, unconsolidated sediments and the main driver is the development of the best possible biological assemblage map, then a bottom-up, unsupervised approach is likely to arrive at a set of assemblages that are defined equally well or slightly better compared with a

top-down approach" (Eastwood et al. 2006 p.1544).

The evolution of bottom-up methods has largely tracked that of species (and communities) distribution modelling, but this has often failed to be explicitly recognized. Benthic ecologists have been able to capitalize on the expansion of the field of distribution modelling due to the mechanistic similarities between conspicuous epibenthic fauna, and vegetation. But lack of collaboration and knowledge of the literature has so far kept benthic modelling and mapping lagging behind vegetation modelling, and new methods are adopted by benthic ecologists only years after they have been developed. Within benthic mapping studies using a bottom-up approach, two further varieties can be distinguished: one can be described exactly as the type of community-level modelling described by Ferrier & Guisan (2006) as classifying sites prior to modelling, thus making use of all species data available (Buhl-Mortensen et al., 2009); this can be said to have been introduced formally by the study of Field et al. (1982) who only fail to complete the 'prediction' stage. The other and more common approach skips over the species-level (raw) data altogether, classifying community types (e.g. 'mixed brown algae') in the field as the data are collected, thus incurring an unfortunate loss of information (Holmes et al., 2008; Ierodiaconou et al., 2011).

The so-called 'bottom-up' and 'top-down' approaches to mapping the benthic landscape can also be paralleled with the concepts of 'biotopes' (Connor et al., 2004; CORINE, 1991; EUNIS, 2005; HELCOM, 1998) and 'habitats' (Allee et al., 2001; Greene et al., 1999; Valentine et al., 2005) respectively. Especially in the applied environmental literature, researchers have gravitated towards these two types of mapping theme presumably because they meet two crucial conditions: they are easy to map, and they are biologically meaningful. The current meaning of 'biotope', which in fact became popular in the marine realm before the terrestrial, combines the "physical environment [...] and its distinctive assemblage of conspicuous species" (Olenin & Ducrotoy 2006, p. 22) where, crucially the concept incorporates geographic location, thus rendering it scale-dependent. In the sections that follow I describe the 'bottom-up' approach in detail and relate it explicitly to distribution modelling at the community level. By means of a case study taken from a Swedish Fjord recently designated as a multiple-use marine protected area I make the case that benthic biotope modelling provides the most effective means to map and spatially manage seabed diversity.

2.3 The case-study

The Koster Fjord, a part of the Koster-Väderö Trough, is a fault fissure running parallel to the northern part of the Swedish West coast, and connected to Atlantic deep water through the Norwegian Trough, over a sill approximately 100 m deep. The Koster archipelago lies between the fjord and the open Skagerrak. The maximum depth of the fjord is 262 m, and rocky bottoms occur down to more than 200 m depth. The width of the trough varies between 2 km and 300 m. The Koster-Väderö Trough is the most oceanic and diverse marine environment found in Sweden, and features a variety of species which elsewhere occur only in deeper water, due to the inflow of Atlantic deep water, a condition known as "fjord effect". At a Swedish level, approximately 200 marine species are unique to this area, including the reef-forming deep water coral *Lophelia pertusa* with a rich associated fauna, gorgonian corals, the upper bathyal sea-pen *Kophobelemnon stelliferum*, the large hard-bottom dwelling lamellibranch *Acesta excavata*, and the sponge *Geodia barretti* (Lundälv & Jonsson, 2000).

The *Kosterhavet* National Park was designated in 2009, conferring the fjord and archipelago of this unique site a new status not without its responsibilities. Among those stands out the conservation-related objective of long-term preservation of the marine ecosystems, habitats, and species occurring naturally in the region, while ensuring the sustainable use of local biological resources.

Notably, the Park is one where multiple uses are allowed. Various kinds of commercial and recreational fisheries have a stake in the area, alongside with tourism, and to a lesser extent, shipping. There is a well established trawl fishery for shrimp (*Pandalus borealis*), and Norway lobster (*Nephrops norvegicus*) comprising around 30 boats, mostly under 12 m. An additional small fleet of local fishermen target lobster (*Homarus gammarus*) by means of static gear. The area is also important for recreation, with many hundreds of yachts and motor boats staying in the area over the summer. Tourism has increased 50% over the last decade with around 80,000 people visiting, mainly in July and August. In light of this, a pressing need emerges to lay out, and ensure mechanisms of enforcement of, a management system within which competing demands inside the multiple-use park are adequately accommodated, while not compromising the conservation goals of the park.

This can only be achieved by means of adequate spatial planning, a tool now widely recognized to be suited for implementing an ecosystem-based approach to the management of ecosystems. Marine spatial planning involves the practice of zoning (to spatially and temporally designate areas for specific purposes), with the aim of reducing conflict both among different users competing for the same space, and more importantly, between users and the environment (Douvere, 2008), ensuring that the capacity of the ocean to provide goods and services remains undiminished. Far from a straightforward task, the main challenge that conservation spatial planning faces today is undoubtedly that goals be achieved in a way that minimizes forgone opportunities for production (Margules & Pressey, 2000).

The zoning system currently in place involves three co-occurring management regimes (see Figure 5 in paper III). Three sites are afforded the highest degree of protection, and are designated as Seabed Protection Areas (SPAs), with a full trawl ban in place and where anchoring and use of other equipment that can damage seafloor are prohibited. The remainder of the area of the National Park is divided into two zones on the basis of depth, whereby areas above 60 m benefit from partial restrictions (partial protection zone) with no commercial fisheries allowed, and areas deeper than 60 m are open for a specially-regulated fishery.

Since the approval of the denomination in the early 2000's Remotely Operated Vehicle (ROV) surveys have been conducted across the area (see Figure 1) and underwater video footage amassed resulting in a number of reports and comprehensive inventory of epibenthic fauna. Underwater filming is widely praised for yielding a source of material that can be revisited and reanalyzed. In the spirit of taking full advantage of the existing library of video material, and in view of a pressing need for further documenting the distribution of biological diversity of (the benthic portion of) the National Park, a project was commissioned that would use a predictive mapping approach, using already-available ROV data as a basis, to produce a map that would serve to facilitate the description and quantification of regional benthic diversity (i.e., a biotope map). Two main conditions applied, namely, that the output be as close to the 'truth' as possible (i.e. the realized distribution of assemblages), and the method as

objective, automated, and repeatable as possible. The text box below summarizes our predictive mapping process:

Conceptualization

Approach: first classify, then predict

Species data: epibenthic megafauna including species and taxon complexes; sessile, or free-living but in contact with bottom

Response data

Split ROV tracks at a chosen scale of analysis and analyze video footage

Classify video sequences into biotope classes ($n=417$)

Predictor data

Create layers of environmental predictors and calculate values for grid cells

Model building

Model training ($n=313$)

Model testing ($n=104$)

Spatial predictions

Predict for $n=2034345$ cells

Assessment of model applicability

Spatial overlay, systematic conservation planning

Text Box 1

3. Considerations of the data model

3.1 Measuring spatial pattern

In distribution modelling the size of the sampling unit, referred to as 'grain size', dictates the level of spatial detail that will be depicted by the output map (and by the same token, the amount of detail that will be ignored). Only if grain size is made to be equal to, or smaller than the smallest areal unit with homogeneous community composition (e.g., a patch), will we obtain the greatest spatial resolution possible. It is therefore one of the most consequential choices to be made. It is however, very rarely addressed, and grain size is often pragmatically determined e.g., by the scale at which predictor data are available.

When the aim of ecological research is the identification of patterns (e.g., species performance relative to an environmental variable) it has been suggested that the size of the sampling unit should be smaller than the structures resulting from the unit process (Dungan et al., 2002; Kenkel et al., 1989). This can only be elucidated by conducting a multiple scale analysis prior to setting out to investigate the phenomenon of choice. Sample size must then strike a balance between the right spatial scale to detect patterns without introducing noise, and variability among samples, which needs to be retained. Accounting for this enables attaining the most parsimonious sampling scheme to explore an ecological phenomenon. Moreover, conducting a study using an inadequate size of sampling unit may result in one of two consequences: the need to sample an area too intensively (i.e., inefficiently), or the risk of overlooking existing patterns (Wiens, 1989).

In the present study virtually no restrictions applied to the choice for grain size. I therefore carried out analyses to unravel the scale at which heterogeneity in faunal composition could be detected so as to ensure that the resolution of the generated map

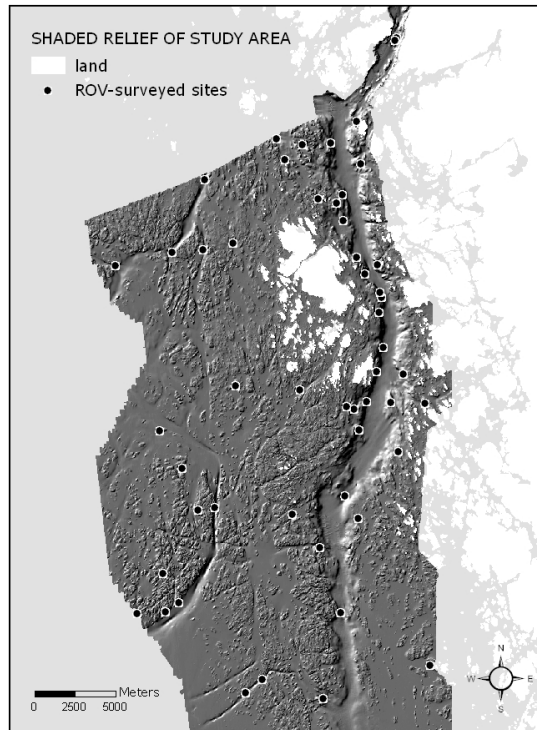


Fig. 1. Map of the study area showing shaded relief and ROV-surveyed sites (filled circles)

matched the scale of the mosaic of faunal patches, to maximize the chances that any existing patterns would be readily picked up by the model, and to make the most efficient use of the data. In Paper I I report on the findings from these analyses.

It was determined that *data aggregated into samples smaller than $\approx 35\text{-}40\text{ m}$ (linear scale) have a megafaunal composition that varies mostly due to stochastic processes, whereas if data were aggregated into larger samples, megafaunal assemblages would be distinguishable, in the form of patches, within. These findings elicit support for grain of any size $<35\text{ m}$, and more conservatively $<20\text{ m}$, which was the smallest of all intercepts (see paper I), and furthermore, for the use of a regular grid covering the whole of the modelling area without a need for partitioning it into regions (Osborne & Suárez-Seoane, 2002), on the basis of nearly identical local scale patterns throughout (note that at one location, site number 6 located at the bottom of the Fjord, no heterogeneity was detected; this means that patches should be even larger than the scales studied, and therefore the chosen grain size, while still adequate, could have been larger there).

Based on these results the available ROV video observations were binned into 15 m by 15 m grid cells for subsequent building of a species-by-site matrix, where 'sites' are linearly-georeferenced, grid cell-bounded video sequences. This procedure defines the spatial resolution of the map and determines that only small scale patterns will be depicted, leaving large scale trends to remain obscure.

3.2 The mapping units

A bottleneck in the process of predictive mapping of the benthos is the lack of a universally-applicable classification system of biodiversity (e.g. Howell et al. 2010), which captures true discontinuities in the megabenthos so as to be readily incorporated into a model as the target pattern. This is why a classification step needs to be implemented in the process. But when user-defined numerical classifications are used arbitrary choices are almost invariably involved and resulting classifications may exhibit suboptimal performance (Snelder et al., 2010).

In order for the classification to be useful its level of detail, in a way mirroring spatial detail, must be able to split up the data into classes that while responding to environmental gradients, they do not introduce excess noise. This is achieved when the species making up the classes (equivalent to assemblages) show a collective pattern of response (Zimmermann & Kienast, 1999). This is in turn contingent upon the degree of internal structure within classes. By representing the readiness with which classes are discriminated by a set of environmental variables ('discriminability') as a function of classification detail (measured as a cut off level of biological similarity) I showed that *how coarsely or finely assemblages are defined can have an effect on the detection of distribution patterns (see paper II). Intermediate detail classifications emerged as most prone to show patterns. This is a largely unresearched question, and it was challenged when presented at an international conference, though no evidence was put forward, in that it was suggested that the more detailed the classification the easier it should be to detect patterns. Further research is therefore needed.

The levels of the response variable, i.e. biotopes, were defined taking into consideration the different predictability of classes determined by different levels of within-class similarity. Other factors were taken into account as well, such as readiness to distinguish between classes solely by visual inspection (as it would be needed if the classification were to be used in future investigations), and closeness to classes in existing biotope classifications (particularly that of Connor et al. 2004). This procedure could not be automated and a different analyst could have arrived at a different classification.

This, then, defines the biological resolution of the study. The classification developed collapsed the most finely-defined communities into their containing classes, e.g. communities dominated by *Swiftia rosea* and/or *S. pallida*, *Acesta excavata*, and others. (Note that the nomenclature was changed in paper IV and this Thesis to not create confusion by having three-character labels and two-character labels). The final classification is summarized in Table 1.

4. Mapping the distribution of seabed diversity through biotope modelling

In the benthos, community composition and diversity have been found to be structured by water column processes (suspended particulate matter loading, food availability, and

Class label	Physical habitat description	Characterizing taxa
AA	Offshore circalittoral rock	Axinellid sponges, <i>Terebratulina retusa</i> , Anomids, <i>Placostegus tridentatus</i> , Spirorbinids
AB	Lower circalittoral rock	Flustrids
AC	Offshore circalittoral biogenic reef	<i>Hathrometra sarsii</i> , <i>Lophelia pertusa</i> , <i>Mycale lingua</i> , <i>Ascidia obliqua</i> , <i>Filograna implexa</i>
BA	Lower circalittoral mixed sediment and rock	<i>Pandalus borealis</i> , <i>Liocarcinus</i> sp., <i>Sabella pavonina</i> , <i>Munida rugosa</i>
BB	Offshore circalittoral fine sediment	<i>Nephrops norvegicus</i> , Cerianthids
BC	Offshore circalittoral mixed sediment and rock	<i>Spirontocaris lilljeborgii</i> , <i>Lithodes maja</i>
BD	Lower circalittoral sediment	<i>Kophobelemnon stelliferum</i> , <i>Pachycerianthus multiplicatus</i>
C	Near-shore lower circalittoral sediment	<i>Pennatula phosphorea</i>
D	Near-shore lower circalittoral coarse sediment including shell hash	Gobids, <i>Pecten maximus</i>

Table 1. Description of biotopes

hydrodynamic stress) as well as substrate properties, temperature, and salinity. The set of predictor, abiotic variables comprised: depth, substrate type, seafloor topographic complexity (measured as total surface area), aspect, and geomorphological landform. These variables are assumed to be proxies for one or more of the functionally relevant variables.

The modelling method of choice was a Conditional Inference (CI) Tree-based forest, which was found to outperform other decision tree-based models both in terms of classification accuracy and the ability to discriminate between classes. CI trees have been developed recently by Hothorn et al. (2006). The framework uses a combination of machine learning principles and hypothesis testing that renders it robust and powerful, albeit computationally demanding. Note that the statistical framework is nested within an overall machine learning system: hypothesis testing is employed to measure the strength of the relationships between predictors and response, and predictions are generated by applying a set of decision rules.

Because I implemented a forest (i.e., an ensemble) of CI trees it is not possible to extract decision rules, which could shed some light on environment-biota relationships (e.g. which habitat factors are the most important in organizing these associations? Which thresholds along the defined gradients represent natural breaks?). The method produces very robust predictions through an eminently inductive process, but it is a black box. Overall the model can be considered (to a certain extent, see below) a test of the implicit theory that species distributions and by extension community composition are at least in part determined by environmental variables, because it predicted better than the least-possible informed prediction system, i.e. to predict every case as the

most common class. However, no contributions can be made to existing explicit theory regarding species response to environmental gradients and resources (Austin, 2007) (see section 2). A lack of ability to explain patterns, however, was offset by a high predictive accuracy, which was prioritized in accordance with the overall aim of the study.

Spatial autocorrelation is known to occur at all scales. While spatial autocorrelation of videoframe-sized samples and distances up to 100's of meters was accounted for in the size of the grid, it is likely that the data points of the predictive model (cell-based samples) were not completely independent from one another but were also spatially autocorrelated as a result of being tightly clustered. It is difficult to disentangle the effect that this may have had on predictive accuracy. This will have had an effect on the choice of variables and splitting points implemented at every node of every tree in the forest, in that effects of spatial autocorrelation may have been taken for true effects of predictors on response, and therefore strength of relationships may have been inflated. I have argued, however, that the overall effect on predictive accuracy may have been positive and should indeed be sought after in mapping studies.

Yet another compromise had to be made when setting up the model entailing that only typical assemblages were adequately modelled. While the model can be useful to select representative areas, nothing can be said about the location of sites where processes may give rise to less typical (and hence more unique) communities. For instance, in one site (number 46) a very large patch of a species of *Corallimorpharia* (tentatively identified as *Sideractis glacialis*, but unfortunately unconfirmed) was found on a rocky outcrop; vast fields of basket stars (both *Ophiothrix fragilis* and *Ophiocomina nigra*, in fully segregated aggregations) were found in another site (number 73). These types of occurrences can neither be explained nor predicted by the model.

*The outcome from the mapping process followed is not only a map of conservation features and potential biodiversity surrogates (see below), but actually a formal definition of biotopes: Firstly, the fact that classes are defined according to species association patterns accounts for the biotic aspect of the concept. Second, predicting the presence of classes on the basis of environmental data accounts for the abiotic aspect. And third, the spatially-explicit nature of the model resolves the definition of geographic boundaries.

A map of benthic biotope classes was produced by assigning the most likely class, as determined by the model, to every single grid cell across the modelling area. Here (see Figure 2) I reproduce the biotope map as clipped to the National Park area, where the results have been applied and where the confidence on the predictions was (on subjective grounds) greater. While the accuracy resulting from the confusion matrix was 72%, it is worth drawing attention to the observation that the model appears to have performed well at identifying boundaries between classes (G. Quinn pers. comm.) To illustrate the applicability of the model to management questions I describe below two examples, only one of which is included briefly in paper III.

By means of the species-by-site matrix and the site classification it is easy to calculate the probability of association of each single species to a given biotope using the frequency of occurrence. This was done for all species and all biotopes, and the resulting data product is available from this author on request. One particular species provides an interesting case. *Kophobelemnion stelliferum* is a species of sea-pen of conservation importance within Sweden (P. Nilsson pers. comm.) and it was highly associated with biotope BD to the point of being one if its characterizing taxa. Through the biotope

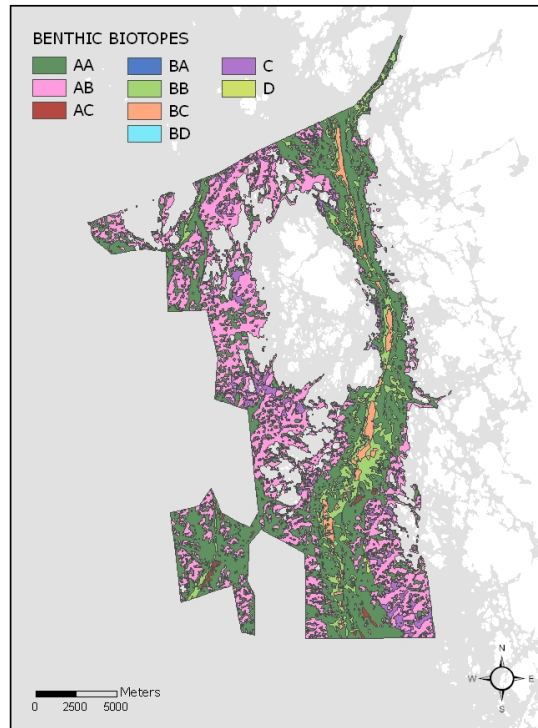


Fig. 2. Map of biotopes, as predicted for each 15 x 15 m cell. AA: Offshore circalittoral hard substrate (Axinellid sponges, *T. retusa*, *P. tridentatus*), AB: Lower circalittoral hard substrate (flustrids turf), AC: Biogenic reef (*L. pertusa*, *H. sarsii*, *M. lingua*, *F. implexa*), BA: Circalittoral mixed rock and sediment (*P. borealis*, *S. pavonina*), BB: Circalittoral fine sand (*N. norvegicus*, cerianthids), BC: Circalittoral mixed rock and sediment (*B. tuediae*, *S. lilljeborgii*, *L. maja*), BD: Circalittoral fine sand (*K. stelliferum*, *F. quadrangularis*, *P. multiplicatus*), C: Lower circalittoral mud (*P. phosphorea*, *F. quadrangularis*), D: Lower circalittoral coarse sediments (*P. maximus*, gobid fishes).

map produced, not only is it possible to plan a targeted survey to further quantify the occurrence of the species in the National Park (which indeed is predicted to occur across a very small area), but also it has emerged that this particular biotope is fully unrepresented in any of the SPAs in place.

For additional analyses fishing data was acquired from the Swedish Board of Fisheries in the form of Vessel Monitoring System records for years 2007-2010. Fishing positions were gridded to 1 ha cells, and the maximum number of pings out of the four-year period was used as an estimate of fishing effort for the location. Spatial overlay of highly fished areas (defined as cells with more than 3 pings; note that the maximum

number was 10) and biotope distribution enabled a very simple quantitative analysis of the level of pressure each biotope endures (see table 2). While the high amount of biotope AA that appears to be heavily fished may be misleading, as this biotope is very often contiguous to BB polygons, it becomes clear that BB areas are the most heavily impacted by fisheries (this additionally served as proof that the biotope model returned a very realistic output). The AC areas from table 2 are sites that no longer have *Lophelia pertusa* colonies. When considering all area fished (not shown) it emerged that 10% of biotope BD was under fishing pressure, posing an even more serious risk to the sea-pen.

Biotope class	Area in high fishing effort zone (ha)	Percent of total in NP (%)
AA	49.3	0.4
AC	1.4	0.7
BB	34.7	1.5
BC	1.5	0.2
BD	0.01	0.3

Table 2. Fishing pressure by biotope class

This is one of the first instances where CI trees or their forest implementation are used in predictive mapping (see Czúcz et al. 2011), and the first in a benthic environment. In paper III I elaborate on how better collaboration with vegetation scientists would boost benthic mapping.

As a companion to this thesis an interactive map server is provided as a html file, also available as a CD on request, where the user can visualize, navigate and query the layers used as predictor data, as well as zoom in and out so as to gain an appreciation of the spatial detail of the biotopes not visible on the printed figures. ROV tracks and data grid cells have also been included, as well as some geo-referenced still images to give an impression of the quality of the raw, species data.

5. Representing the distribution of seabed diversity using biotopes as surrogates

A common approach to selecting candidate areas for protection, given incomplete knowledge of the distribution of the entities that are ultimately the target of those protection measures, is by use of surrogates which are correlated with species patterns. This in itself is insufficient because selecting all areas of high diversity (that is, assumed to have high diversity based on surrogate coverage) would require much more area to represent the full complement of species than if complementarity in the species composition of each individual area were taken into account. A combination of an effective surrogate scheme and systematic conservation planning (by way of site-selection algorithms) is a valid approach for implementing networks of conservation areas (or for prioritizing areas for conservation in a given planning region) that would ensure efficiency and therefore political and social support. The weak point of this reasoning lies on the degree of effectiveness of the surrogates chosen in representing (species) diversity. Surrogacy can be achieved in two ways: using features at high levels of biological organization whose spatial distribution is known (e.g. vegetation classes), or selecting areas from environmentally different regions. Because surrogates are developed with the intention to be used in managerial applications, conditions apply, namely, that they truly are more readily collected than biological data

(as is the case with remote sensing data), and that the strength of the relationship with biodiversity patterns is indisputable.

One way of determining whether a given surrogate is effective at substituting biodiversity (species) data is by selecting a conservation network aiming to represent all surrogates (i.e. ignoring species level information) and then measuring how many species are represented inside the network by incidental representation. However, if it could be demonstrated that the spatial patterns of the surrogate are fully coincident with the spatial patterns of species, that would be proof that the surrogate is effective, irrespective of any decisions made regarding area selection. This has been criticized as too stringent a test of surrogacy (Sarkar et al., 2005), but if there is a relationship between the target and the surrogate (i.e., if there is a biological model linking the two) this is a rather trivial fact.

I have demonstrated, in paper **IV**, that using the predicted biotopes as surrogates in a systematic conservation planning framework based on simulated annealing for site selection effectively captures the diversity of epibenthic megafauna. I used a new test that proved that *at local scales, biotope heterogeneity is correlated with species diversity, exactly as dictated by the idea of biodiversity surrogacy. Moreover, sites of low taxon richness and low conservation utility, such as the evaluation site located at the bottom of the Fjord, were observed in areas where homogeneity had been demonstrated (see Section 3.1, as well as paper **IV** for a definition of conservation utility). Further, this was only in part due to the fact that biotopes were defined on the basis of epibenthic megafauna. The other side of the coin is of course that additional surrogates would need to be defined for the representation of co-occurring species with different environmental requirements, e.g. infauna, and likely, pelagic diversity. The evidence suggests that biotopes obtained via community-level distribution modelling are good surrogates for the patterns of the fraction of biodiversity that was used to define them.

To the best of my knowledge, the postulate on which biodiversity surrogacy is based has not been set into ecological theory, which may be part of the reason why so much debate surrounds the question of surrogate efficacy (Grantham et al., 2010; Rodrigues & Brooks, 2007). One noteworthy exception is Araújo et al. (2004; 2001), who has incidentally recurrently found evidence against environmental surrogacy. Below I argue that issues of surrogacy are intrinsically related to the theories underpinning the concepts of this thesis since the problem of species diversity can be solved if one solves the problem of just what drives species distribution. My purpose is to put forward evidence that the issue should be further looked into.

The idea that environmental variables can predict the distribution patterns of species is equivalent to the idea that environmental heterogeneity can be a surrogate for species diversity patterns, only the two types of relationships occur at different levels of abstraction. Indeed both species distribution modelling and environmental surrogacy rely on the same assumption: if species distributions are at equilibrium with governing environmental factors, as predicted by Niche theory (Hutchinson, 1957) then (1) it becomes legitimate to make spatial predictions from species response curves (e.g. Guisan & Thuiller 2005) and (2) adequately representing environmental space would maximize species representation (Araújo et al., 2001).

Spatial modelling has however been much more successful (i.e. less controversial) than finding appropriate surrogates because modellers look intentionally to use variables as predictors that are functionally related (whether directly or indirectly)

with the response species, whereas finding good surrogates has largely relied on chance (available data, etc.). Araújo et al. (2001) found that the only target group adequately represented by environmental diversity (itself consisting of temperature and precipitation patterns) was plants, and the authors argue that this follows logically from a well-established relationship between plant distribution and environmental limiting factors.

The approach of Araújo et al. (2004; 2001), Arponen et al. (2008) and others is based on the use of continuous environmental data or assemblage data (i.e. compositional turnover) in combination with site-selection algorithms drawing from sampling theory (e.g. the p-median location allocation algorithm) as a method for selecting networks of protected areas (Faith & Walker, 1996). Other authors (e.g., Bonn & Gaston 2005; Lombard et al. 2003; Meynard et al. 2009; Sarkar et al. 2005) have used categorical data combined with algorithms that solve the minimum set problem. I have used an intermediate approach that effectively modified continuous environmental data on the basis of assemblage-level data at the modelling step, and then generated categorical conservation features from the model outputs to be fed into a site selection algorithm (of the type solving the minimum set problem), whose performance was highly satisfactory. Therefore, modelling the distribution of communities not only provides a means to gain insight into ecological pattern, but also an effective foundation for defining surrogates that can speed up, and decrease the cost of, conservation action.

A unified framework is hereby proposed to document and spatially manage diversity which building from environment-biota relationships improves the data-efficiency and robustness of methods by: (1) collecting species-level ecological data at the right spatial scale, and throughout environmental gradients, the latter of which assisted by use of sampling theory (which is incidentally what the theory was designed for), and (2) integrating biological data, whether as species data or if classified, then at the right biological resolution, into continuous environmental data (e.g. remotely sensed) by means of distribution modelling and habitat suitability mapping. Ecological theory can be invoked depending on the goal of each particular project to either further document community distribution or extract adequate surrogates for biodiversity.

6. Conclusions and outlook

If planning for conservation in the sense of designing networks of protected areas is usually a practice of hope, in the marine realm, where areas are prioritized at best on the basis of scant information and (more or less questionable) estimates of species occurrence to fill in the blanks, it is almost a practice of faith. A lack of robustness in the process of marine conservation and spatial planning not only jeopardizes the efficacy of conservation areas, but also reduces vital societal support. And while it will never be possible to make conservation planning decisions based on complete knowledge of the distribution of the full complement of biodiversity, from genes to ecosystems, a degree of robustness can be achieved, as I have shown, by means of mapping the distribution of biotopes and applying spatial planning principles. I have argued that in cases where distribution models can be developed the knowledge generated should be exploited not only to make predictions of species distributions, but also to define appropriate biodiversity surrogates. An opportunity thus emerges to bridge the gap between community-level modelling and conservation planning.

The issue at the core however, is unravelling the factors that explain the spatial

distribution of species, whether benthic or otherwise. I have focused on methodological issues, but the role that ecological theory plays into this process cannot be understated. A stronger footing on ecological theory will help develop better distribution models that produce more reliable spatial predictions. This can be possible upon elucidation of spatial patterns. In the area studied, stochastic variation characterized the structure of megabenthos at microscales (<15 m), while a mosaic of faunal patches was detected (and predicted) at intermediate scales (between 15 m and 100's of meters). Further, at even larger scales (>100's of meters) patterns of emergent properties of megabenthic community structure began to unravel.

I would like to highlight one area whose obscurity is particularly troubling: that of the distribution of *atypical* assemblages. Arguably, areas of unique processes sustaining atypical diversity, like 'beautiful freaks', are the most interesting of all. However, these are unpredictable by definition and can be revealed only by means of thorough and systematic surveying. A plea is made here to continue and improve surveying efforts to collect more observational data, which would not only enable the discovery of atypical areas but also improve existing models and theory.

Besides variation in the physical environment and biological interactions (not discussed here) the heterogeneity of benthic communities is known to be strongly influenced by the 'vagaries of larval recruitment' (Johnson, 1970), setting the marine realm quite apart from the terrestrial. Indeed, the megabenthos cannot be ecologically equated to vegetation. New methods are needed for mapping dynamic processes of the water column, so that areas that fulfil an important role in the functioning of marine ecosystems (e.g. areas of upwelling, corridors for larval transport, etc. acting as sort of 'keystone spaces') can be incorporated in the process of marine spatial planning. Surrogates for pelagic diversity are also urgently needed. Only by marrying the two realms, benthic and pelagic, will we gain the ability to fully understand, predict, and protect the distribution of marine organisms.

7. References

- Allee, R., Dethier, M., Brown, D., Deegan, L., Ford, G., Hourigan, T., Maragos, J., Schoch, C., Sealy, K., Twilley, R., Weinstein, M. & Yoklavich, M. (2001). Marine and estuarine ecosystem and habitat classification, *Technical Report NMFS-F/SPO-43*, NOAA Technical Memorandum, Silver Spring, Maryland.
- Araújo, M. B., Densham, P. J. & Williams, P. H. (2004). Representing species in reserves from patterns of assemblage diversity, *Journal of Biogeography* **31**: 1037–1050.
- Araújo, M. B., Humphries, C. J., Densham, P. J., Lampinen, R., Hagemeyer, W. J. M., Mitchell-Jones, A. J. & Gasc, J. P. (2001). Would environmental diversity be a good surrogate for species diversity?, *Ecography* **24**: 103–110.
- Arponen, A., Moilanen, A. & Ferrier, S. (2008). A successful community-level strategy for conservation prioritization, *Journal of Applied Ecology* **45**: 1436–1445.
- Austin, M. (2007). Species distribution models and ecological theory: A critical assessment and some possible new approaches, *Ecological Modelling* **200**(1-2): 1–19.
- Austin, M. P. (1992). Modeling the environmental niche of plants - implications for plant community response to elevated CO₂ levels, *Australian Journal Of Botany* **40**: 615–630.

- Bonn, A. & Gaston, K. J. (2005). Capturing biodiversity: selecting priority areas for conservation using different criteria, *Biodiversity and Conservation* **14**: 1083–1100.
- Brown, C., Cooper, K., Meadows, W., Limpenny, D. & Rees, H. (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**: 263–278.
- Brown, C., Hewer, A., Meadows, W., Limpenny, D., Cooper, K. & Rees, H. (2004). Mapping seabed biotopes at Hastings Shingle Bank, eastern English Channel. Part 1. Assessment using sidescan sonar, *Journal of the Marine Biological Association of the UK* **84**: 481–488.
- Brown, C. J., Mitchell, A., Limpenny, D. S., Robertson, M. R., Service, M. & Golding, N. (2005). Mapping seabed habitats in the Firth of Lorn off the west coast of Scotland: evaluation and comparison of habitat maps produced using the acoustic ground-discrimination system, RoxAnn, and sidescan sonar, *ICES Journal of Marine Science: Journal du Conseil* **62**(4): 790–802.
- Buhl-Mortensen, P., Buhl-Mortensen, L., Dolan, M., Dannheim, J. & Kröger, K. (2009). Megafaunal diversity associated with marine landscapes of northern Norway: a preliminary assessment, *Norwegian Journal of Geology* **89**: 163–171.
- Connor, D. W., Allen, J., Golding, N., Howell, K., Lieberknecht, L., Northen, K. O. & Reker, J. (2004). The marine habitat classification for Britain and Ireland, Version 04.05, *Technical Report ISBN 1 861 07561 8 (internet version)*, JNCC, Peterborough. <http://www.jncc.gov.uk/MarineHabitatClassification/>.
- CORINE (1991). Biotopes manual. a method to identify and describe consistently sites of major importance for nature conservation data specifications, *Technical report*, European Communities Commission EUR 12587, 126 pp.
- Czúcz, B., Gálhidy, L. & Mátyás, C. (2011). Present and forecasted xeric climatic limits of beech and sessile oak distribution at low altitudes in Central Europe, *Annals of Forest Science* **68**: 99–108.
- Douve, F. (2008). The importance of marine spatial planning in advancing ecosystem-based sea use management, *Marine Policy* **32**: 762–771.
- Dungan, J. L., Perry, J. N., Dale, M. R. T., Legendre, P., Citron-Pousty, S., Fortin, M.-J., Jakomulska, A., Miriti, M. & Rosenberg, M. S. (2002). A balanced view of scale in spatial statistical analysis, *Ecography* **25**: 626–640.
- Eastwood, P., Souissi, S., Rogers, S., Coggan, R. & Brown, C. (2006). Mapping seabed assemblages using comparative top-down and bottom-up classification approaches, *Canadian Journal of Fisheries and Aquatic Sciences* **63**: 1536–1548.
- EUNIS (2005). European Nature Information System, *Technical report*.
- Faith, D. P. & Walker, P. A. (1996). Environmental diversity: on the best-possible use of surrogate data for assessing the relative biodiversity of sets of areas, *Biodiversity and Conservation* **5**: 399–415.
- Ferrier, S., Drielsma, M., Manion, G. & Watson, G. (2002). Extended statistical approaches to modelling spatial pattern in biodiversity in north-east New South Wales. II. Community-level modelling, *Biodiversity and Conservation* **11**: 2309–2338.
- Ferrier, S. & Guisan, A. (2006). Spatial modelling of biodiversity at the community level, *Journal of Applied Ecology* **43**: 393–404.

- Field, J. G., Clarke, K. R. & Warwick, R. M. (1982). A practical strategy for analysing multispecies distribution patterns, *Marine Ecology Progress Series* **8**: 37–52.
- Freitas, R., Silva, S., Quintino, V., Rodrigues, A. M., Rhynas, K. & Collins, W. T. (2003). Acoustic seabed classification of marine habitats: studies in the western coastal-shelf area of Portugal, *ICES Journal of Marine Science* **60**(3): 599–608.
- Furlanello, C. (2004). GIS-based predictive models for ecology, in S. Džeroski, B. Ženko & M. Debeljak (eds), *The Fourth International Workshop on Environmental Applications of Machine Learning (EAML 2004) Bled, Slovenia, September 27-October 1, 2004*, Jožef Stefan Institute, Ljubljana, Slovenia.
- Grantham, H. S., Pressey, R. L., Wells, J. A. & Beattie, A. J. (2010). Effectiveness of biodiversity surrogates for conservation planning: Different measures of effectiveness generate a kaleidoscope of variation, *PLoS ONE* **5**(7): e11430.
- Greene, H., Yoklavich, M. M., Starr, R. M., O'Connell, V. M., Wakefield, W., Sullivan, D. E., Jr., J. E. M. & Cailliet, G. M. (1999). A classification scheme for deep seafloor habitats, *Oceanologica Acta* **22**(6): 663–678.
- Guisan, A. & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models, *Ecology Letters* **8**: 993–1009.
- Guisan, A. & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology, *Ecological Modelling* **135**(2-3): 147–186.
- Heikkinen, J. & Mäkipää, R. (2010). Testing hypotheses on shape and distribution of ecological response curves, *Ecological Modelling* **221**(3): 388 – 399.
- HELCOM (1998). Red list of marine and coastal biotopes and biotope complexes of the Baltic Sea, in H. Nordheim & D. Boedeker (eds), *HELCOM - Baltic Sea Environment Proceedings*, Vol. 75, p. 115.
- Hewitt, J. E., Thrush, S., Legendre, P., Funnell, G., Ellis, J. & Morrison, M. (2004). Mapping of marine soft-sediment communities: Integrated sampling for ecological interpretation, *Ecological Applications* **14**: 1203–1216.
- Holmes, K., Niel, K. V., Radford, B., Kendrick, G. & Grove, S. (2008). Modelling distribution of marine benthos from hydroacoustics and underwater video, *Continental Shelf Research* **28**: 1800–1810.
- Hothorn, T., Hornik, K. & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework, *Journal of Computational and Graphical Statistics* **15**: 651–674.
- Howell, K., Davies, J. & Narayanaswamy, B. (2010). Identifying deep-sea megafaunal epibenthic assemblages for use in habitat mapping and marine protected area network design, *Journal of the Marine Biological Association of the United Kingdom* **90**(1): 33–68.
- Howell, K. L. (2010). A benthic classification system to aid in the implementation of marine protected area networks in the deep/high seas of the NE Atlantic, *Biological Conservation* **143**(5): 1041–1056.
- Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography*, Princeton University Press.
- Hutchinson, G. (1957). Concluding remarks, *Cold Spring Harbor Symposia on Quantitative Biology* **22**: 415–427.
- Ierodiaconou, D., Monk, J., Rattray, A., Laurenson, L. & Versace, V. (2011). Comparison of automated classification techniques for predicting benthic biological communities using hydroacoustics and video observations, *Continental Shelf*

- Research 31(2), Supplement 1*: S28–S38.
- Johnson, R. G. (1970). Variations in diversity within benthic marine communities, *The American Naturalist* **104**(937): 285–300.
- Jordan, A., Lawler, M., Halley, V. & Barrett, N. (2005). Seabed habitat mapping in the Kent Group of islands and its role in Marine Protected Area planning, *Aquatic Conservation: Marine and Freshwater Ecosystems* **15**(1): 51–70.
- Kenkel, N., Juhász-Nagy, P. & Podani, J. (1989). On sampling procedures in population and community ecology, *Vegetatio* **83**: 195–207.
- Kloser, R. J., Bax, N. J., Ryan, T., Williams, A. & Barker, B. A. (2001). Remote sensing of seabed types in the Australian South East Fishery; development and application of normal incident acoustic techniques and associated 'ground truthing', *Marine and Freshwater Research* **52**: 475–489.
- Kostylev, V., Todd, B., Fader, G., Courtney, R., Cameron, G. & Pickrill, R. (2001). Benthic habitat mapping on the Scotian Shelf based on multibeam bathymetry, surficial geology and seafloor photographs, *Marine Ecology Progress Series* **219**: 121–137.
- Lamarque, G., Lurton, X., Verdier, A.-L. & Augustin, J.-M. (2011). Quantitative characterisation of seafloor substrate and bedforms using advanced processing of multibeam backscatter: Application to Cook Strait, New Zealand, *Continental Shelf Research* **31**(2, Supplement): S93–S109.
- Lombard, A. T., Cowling, R. M., Pressey, R. L. & Rebelo, A. G. (2003). Effectiveness of land classes as surrogates for species in conservation planning for the Cape Floristic Region, *Biological Conservation* **112**(1-2): 45–62.
- Lundälv, T. & Jonsson, L. G. (2000). *Inventering av Koster-Väderömrådet med ROV-teknik. En pilotstudie*, naturvårdsverket förlag.
- MacArthur, R. H. & Wilson, E. O. (1963). An equilibrium theory of insular zoogeography, *Evolution* **17**(4): 373–387.
- MacArthur, R. H. & Wilson, E. O. (1967). *The theory of island biogeography*, Princeton, N.J., Princeton University Press.
- Margules, C. R. & Pressey, R. L. (2000). Systematic conservation planning, *Nature* **405**(6783): 243–253.
- Meynard, C. N., Howell, C. A. & Quinn, J. F. (2009). Comparing alternative systematic conservation planning strategies against a politically driven conservation plan., *Biodiversity and Conservation* **18**: 3061–3083.
- Normand, S., Treier, U. A., Randin, C., Vittoz, P., Guisan, A. & Svenning, J.-C. (2009). Importance of abiotic stress as a range-limit determinant for European plants: insights from species responses to climatic gradients, *Global Ecology and Biogeography* **18**(4): 437–449.
- Olenin, S. & Ducrotoy, J.-P. (2006). The concept of biotope in marine ecology and coastal management, *Marine Pollution Bulletin* **53**: 20–29.
- Osborne, P. E. & Suárez-Seoane, S. (2002). Should data be partitioned spatially before building large-scale distribution models?, *Ecological Modelling* **157**(2-3): 249–259.
- Parry, D., Kendall, M., Pilgrim, D. & Jones, M. (2003). Identification of patch structure within marine benthic landscapes using a remotely operated vehicle, *Journal of Experimental Marine Biology and Ecology* **285-286**: 497–511.
- Petersen, C. (1924). The necessity for quantitative methods in the investigation of the animal life on the sea-bottom, *Proceedings of the Zoological Society of London*

94: 687–694.

- Pulliam, H. (2000). On the relationship between niche and distribution, *Ecology Letters* 3: 349–361.
- Rodrigues, A. S. & Brooks, T. M. (2007). Shortcuts for biodiversity conservation planning: The effectiveness of surrogates, *Annual Review of Ecology, Evolution, and Systematics* 38: 713–737.
- Sarkar, S., Justus, J., Fuller, T., Kelley, C., Garson, J. & Mayfield, M. (2005). Effectiveness of environmental surrogates for the selection of conservation area networks, *Conservation Biology* 19(3): 815–825.
- Secretariat of the Convention on Biological Diversity (2004). Programme of Work on Protected Areas (CBD Programmes of Work), *Technical report*, Montreal: Secretariat of the Convention on Biological Diversity.
- Snelder, T., Lehmann, A., Lamouroux, N., Leathwick, J. & Allenbach, K. (2010). Effect of classification procedure on the performance of numerically defined ecological regions, *Environmental Management* 45: 939–952.
- Stevens, T. & Connolly, R. M. (2004). Testing the utility of abiotic surrogates for marine habitat mapping at scales relevant to management, *Biological Conservation* 119(3): 351–362.
- ter Braak, C. J. & Prentice, I. C. (1988). *A theory of gradient analysis*, Advances in Ecological Research Vol. 34, Elsevier.
- Valentine, P., Todd, B. & Kostylev, V. (2005). Classification of marine sublittoral habitats, with application to the Northeastern North America region, *American Fisheries Society Symposium* 41: 183–200.
- Wiens, J. A. (1989). Spatial scaling in ecology, *Functional Ecology* 3(4): 385–397.
- Zimmermann, N. & Kienast, F. (1999). Predictive mapping of alpine grasslands in Switzerland: Species versus community approach, *Journal of Vegetation Science* 10: 469–482.



Acknowledgements

Completing this PhD has taken five years, and what has been summarized here is only the tip of the iceberg of a monumental learning process expanding far beyond the academic. Since I moved to Sweden I have learnt the rudiments of a language I never before thought I would use (although my Swedish is only good enough to conduct such mundane activities as are grocery shopping, making hair appointments, and the like); I have learnt a whole new Budo sport I did not know I had a skill for (I am now a proud brown belt holder in the art of Taïdo) complete with a few Japanese words which I will surely pronounce with a Swedish accent; and I have learnt to master the skills required by the Swedish way of living, including such challenges as driving on several layers of ice and snow, finding my way in the woods in the dark, and the art of sailing avoiding the Norwegian cruisers, among others.

But it is the people around me that have made all this learning possible. All my peers and staff at Tjärnö have been wonderful, but the help and support I got from Per Bergström has been invaluable, and I would like to wish him the very best of luck in his up-coming endeavours, including gaining his own PhD. My special thanks go also to Elin Renborg, who was always prompt to remind me how insignificant some things are. Mikael Dahl, who was my downstairs neighbour for the first few years, is also due to wrap up his PhD in the near future: hang in there, Micke!

My professors have of course taught me a great deal. My supervisor, Mats Lindegarh, has made great efforts to be there when I needed him, as well as has been very skilled at telling when he was not needed, letting me make my own path as I advanced through my PhD but never too far behind me to make sure I would not make a wrong turn. Tomas Lundälv has taught me that a passion for the sea is something to be treasured.

The Strömstad taïdo club Sensu Dojo has given me much more than a brown belt. Not only have I made more than a few great friends there but also, learning "the way for the body and the mind" has had an enormous impact in all aspects of my life. I owe sensei Björn my deepest gratitude for helping keep my life in balance.

My thanks go also to my family (Mamá, Papá, Vicente, Jesús, and Rosario), my biggest fans and most unconditional supporters, who I take for granted all too often. *Agradezco también el apoyo de mi familia (Mamá, Papá, Vicente, Jesús, y Rosario), mis mayores y más incondicionales forofos, a los que me consta no les nuestro mi agradecimiento lo suficiente.*

The final sprint of thesis write-up was surprisingly enjoyable and smooth, and Massimo Di Stefano deserves all the credit for that. In the short time that he has been around he has already managed to change my life. Three solid pillars, both in Sweden and my home country Spain, continue to keep me standing. Tobias Carlsson and Anders Hedin have been companion, confidants, shoulders to cry on, and a constant boost for my not always reliable self-esteem, for the duration of my stay in Sweden. Back home, Ana Steinberg's support, understanding and infinite patience have been more important to me than words can convey. Thank you; *tack; gracias.*

Finally, if you are reading this and your name is not here, I apologize, because it should have been.