

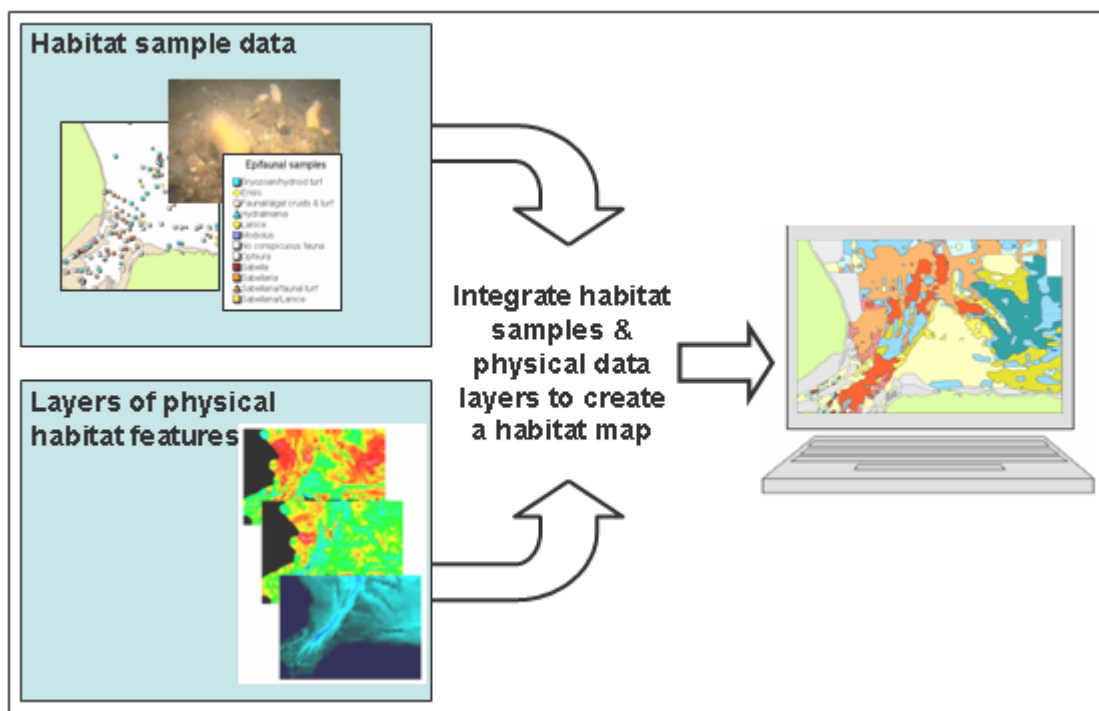
Title:	MESH Guide: How do I make a map?
Author(s):	Vera Van Lancker (UGent) and Bob Foster-Smith (Envision)
Document owner:	Vera Van Lancker (vera.vanlancker@UGent.be)
Reviewed by:	Roger Coggan (Cefas), Jacques Populus (Ifremer), Dave Long (BGS), Jon Davies (JNCC), David Connor (JNCC)
Workgroup:	
MESH action:	Action 2
Version:	Version 1
Date published:	August 2007
File name:	GMHM4 How do I make a map.pdf
Language:	English
Number of pages:	77
Summary:	<p>The <i>MESH Guide to habitat mapping</i> aims to provide a methodological framework for marine habitat mapping so that future mapping studies will produce high quality data and maps which are inter-compatible and their outputs can be assimilated into common, harmonised maps. It will help to make habitat maps more compatible by illustrating tried and tested standards and procedures in a step-by-step manner.</p> <p>This document describes how data are transformed and integrated to derive a biological habitat map that is fit for purpose. It covers the key steps in making a habitat map by the combined use of habitat and biological records and data layers of physical habitat variables. It also offers advice on map layouts and GIS.</p>
Reference/citation:	<p>Van Lancker, V. & Foster-Smith, R. 2007. How do I make a map? In: <i>MESH Guide to Habitat Mapping</i>, MESH Project, 2007, JNCC, Peterborough. Available online at: http://www.searchmesh.net/default.aspx?page=1900)</p>
Keywords:	
Bookmarks:	
Related information:	<p>This document is a printable version of the MESH Guide website: http://www.searchmesh.net/default.aspx?page=1658</p>

How do I make a map?

Vera van Lancker & Bob Foster-Smith

This section describes how data are transformed and integrated to derive a biological habitat map that is fit for purpose. The section does not cover post-survey data preparation - see [How Do I Collect My Data?](#). It is also important to note that the mapping data may themselves have gone through a process of analysis and interpretation from their original source material, especially in the case of desktop broad-scale habitat mapping.

The key steps in making a habitat map (as stressed throughout the MESH Guide) are the combined use of habitat and biological records (normally data from direct seabed sampling) and full coverage data layers of physical habitat variables, *hereafter* called physical coverages (normally sourced from remote sensing, derived from samples or predicted by models). Data integration and habitat modelling are needed to derive the relationships between the biological and physical data. These relationships can be used to predict the habitat distributions.



A summary of the seabed habitat mapping process

Habitat map production has four key steps:

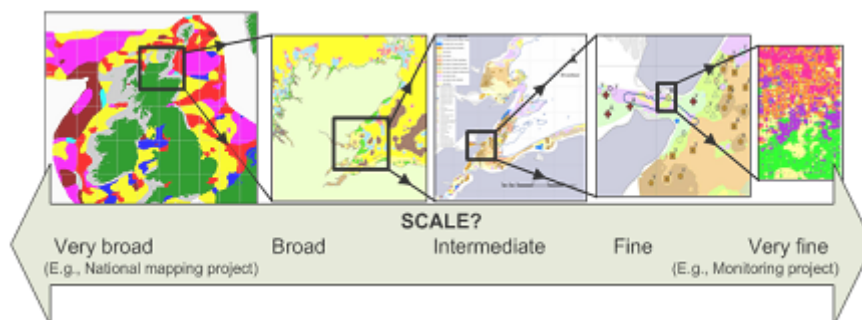
1. Optimise the ground-truth biological data analysis;
2. Selecting and deriving the best available input and most appropriate physical coverages (includes analysis of remotely-sensed data);
3. The map making using the most appropriate techniques for interpreting the data through integration and modelling, and;
4. Designing the map to create a map fit for purpose.

The success of the final habitat map can be undermined if any one of them is not done well!

The various techniques that can be applied to the four stages are explained in the following sections, after outlining suggested strategies for their use.

Determine the strategy

There are many approaches to habitat mapping, but it is convenient to illustrate the range of applications as lying along a scale spectrum: broad to fine scale. This spectrum has been adopted in many sections of the MESH Guide and should now be familiar. Broad versus fine scale best contrasts the different approaches, but become more difficult to explain for intermediate scales.



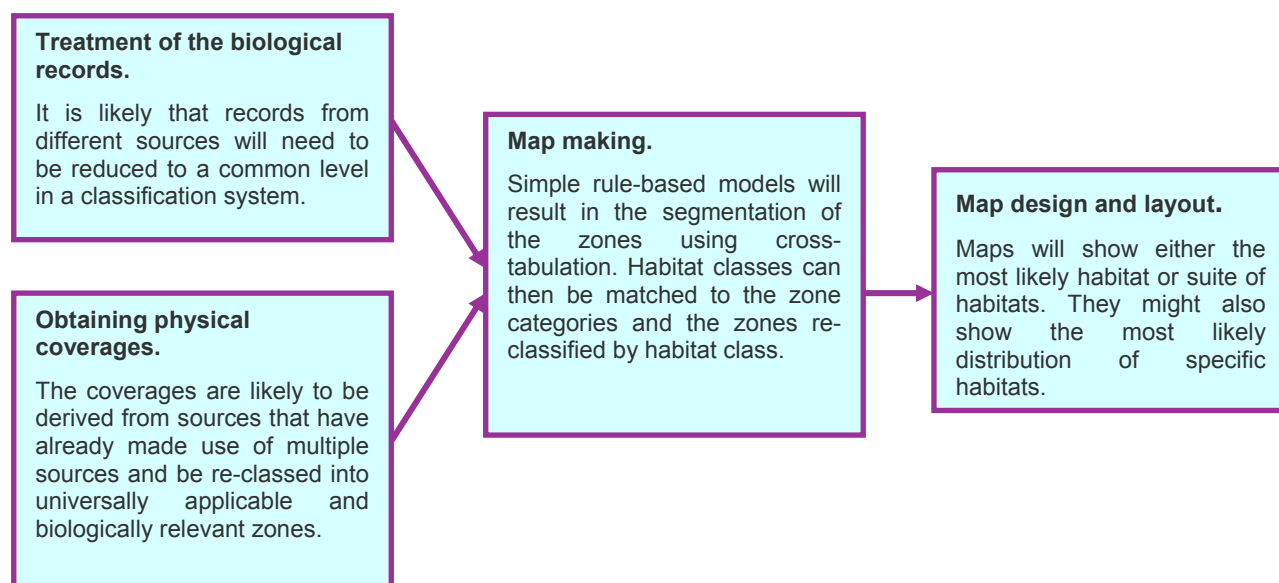
An illustration of the type of maps across the broad to fine-scale spectrum

Benthic habitat mapping will inherently involve some form of modelling. In this context, a model can be seen as any representation of the sea floor based on a systematic investigation and analysis of the evidence for habitat distribution. This is an intentionally broad definition, which can range from expert knowledge interpretation 'by eye' and rule-based models through to statistical models based on an extensive analysis of data.

Bear in mind, however, that techniques described under one heading may be applied to the other situations. For example, it is possible to use the 'broad-scale' modelling approaches for small areas and to use the 'fine-scale' detailed approach for very large areas (e.g. national seabed surveys in Ireland ([INFOMAR](http://www.marine.ie/home/services/surveys/seabed/) <http://www.marine.ie/home/services/surveys/seabed/>) and Norway ([Mareano](http://www.mareano.no/) <http://www.mareano.no/>)).

Broad-scale habitat mapping for very large areas (e.g. whole seas or national waters)

This approach most commonly uses coverages for the major physical parameters that are combined to obtain the predicted distribution of a wide range of broadly defined habitat types (e.g. EUNIS levels 3 and 4). The data used is derived from multiple sources (see *What is habitat mapping?*) and GIS's are most suitable for the modelling. The combined errors and uncertainties of the input data layers will mean that the resulting maps are necessarily rather coarse and generalised in nature.



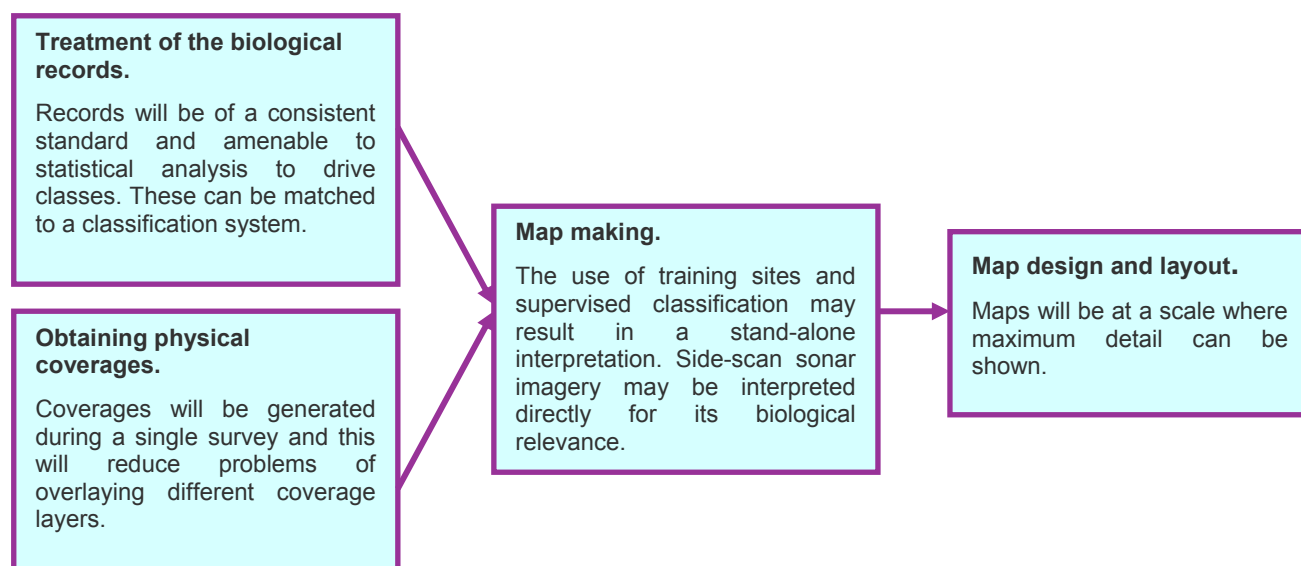
Scheme of the broad-scale mapping approach

The modelling will be based on applying general rules relying on *knowledge* of causal relationships, valid across a large geographic area. The area is divided up into habitat categories based on combinations of a limited number of zones for each of the key environmental structuring variables (e.g. depth, substratum and water energy). This may involve expert opinion to reclassify the physical parameters into a small number of biologically relevant categories without recourse to sophisticated statistical methods to justify the upper and lower boundaries of the structuring variables.

The MESH EUNIS “triplet” approach is an example where the EUNIS level 3 variables (substratum, depth and bed stress) have been combined to have a rough idea of the distribution of broad-scale habitats. The Marine Landscape approach is similar, but includes a slope layer making the product easier to understand from a physiographic viewpoint.

Fine-scale habitat mapping for small areas

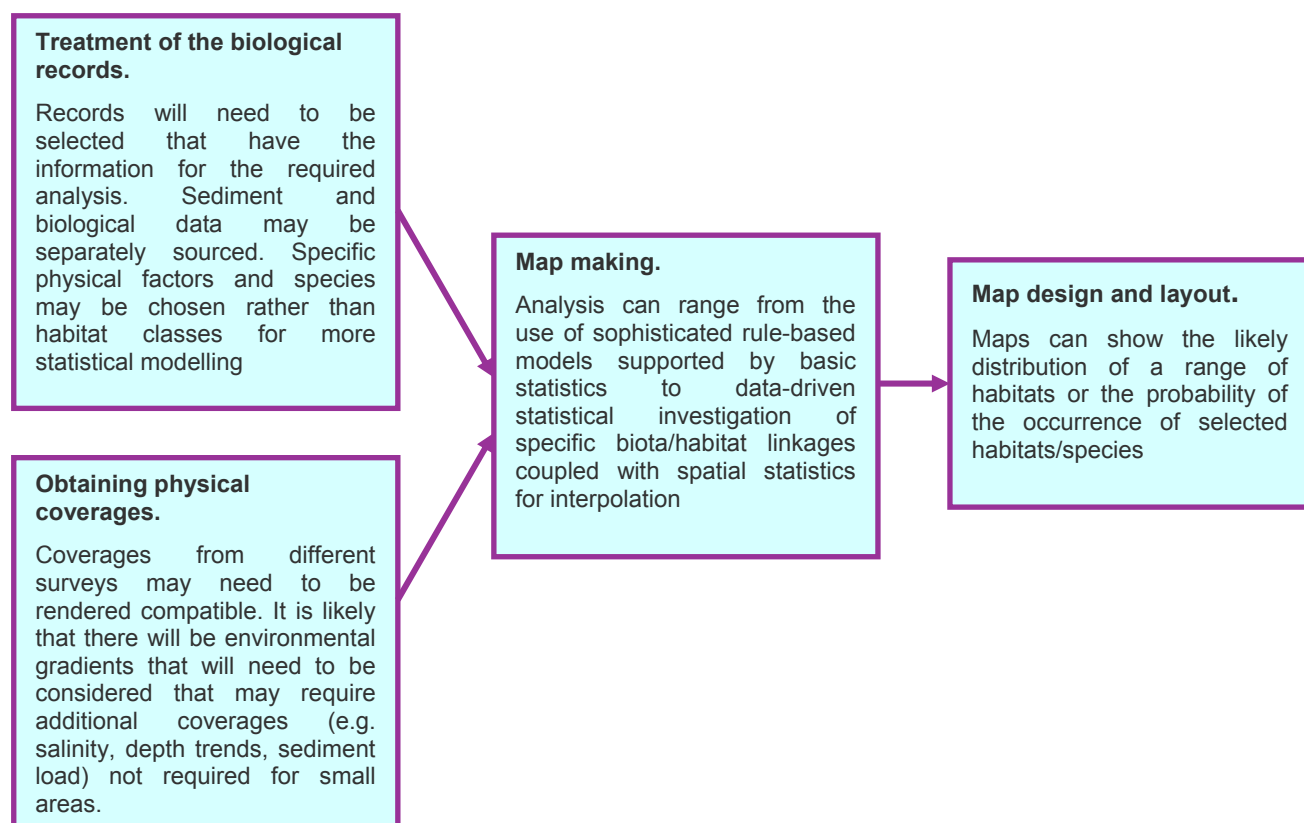
At the other end of the continuum, single surveys of small areas usually adopt a *data-driven* approach. More narrowly defined habitats (e.g. EUNIS levels 4 and 5) can now be predicted and ground-truthing is usually more detailed. Mostly, ‘signatures’ are created from the physical coverages (usually acoustic or electromagnetic spectrum (EMS) reflectance values) using the ground-truth records as training sites. Interpretation of the data from each survey is ‘self-contained’ and no attempt needs to be made to derive rules that can be exported to a second survey. This strategy is exemplified by supervised classification of remotely-sensed images and there are many examples of surveys where this approach has been used. It is likely that specific image-processing GIS software will be required. Side-scan sonar images may discriminate a much smaller set of habitats and be more straightforward to interpret than for large areas: intermediate interpretation of side-scan sonar imagery to produce physical habitat maps may be by-passed by interpreting biological habitats directly.



Scheme of the fine-scale mapping approach

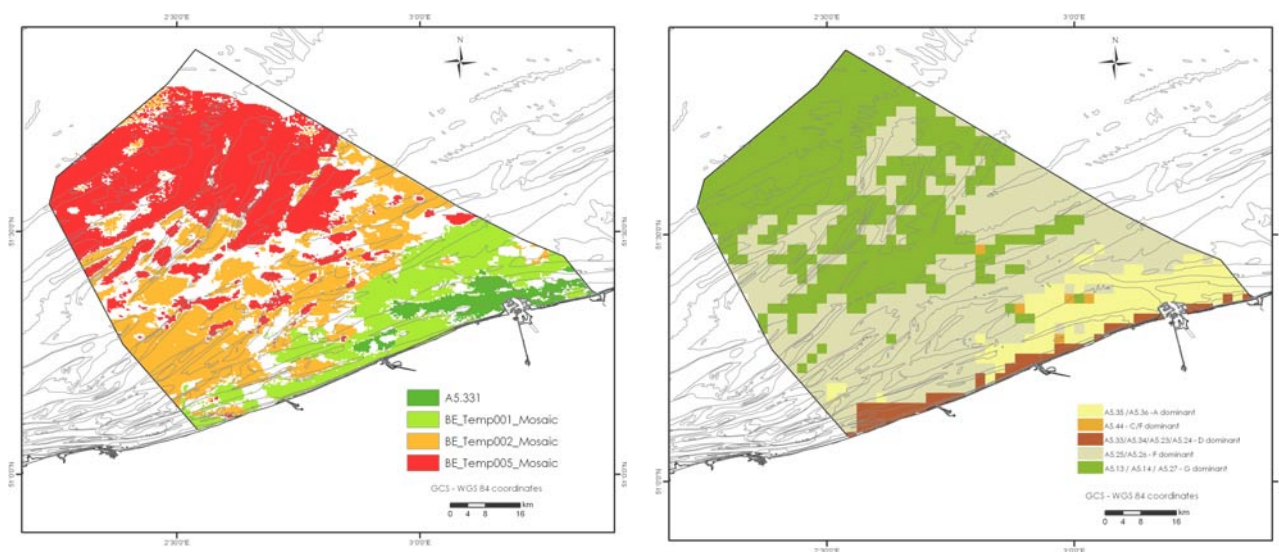
Intermediate scale mapping

It is the scale between the very broad and the fine scales that prove to be the most difficult to give guidance on. It is likely that the input data will come from many surveys over an extended campaign or different campaigns. Some of the inputs may come directly from surveys whilst others (such as bed stress) will be modelled mathematically. Strategies might range from (a) the use of statistical techniques to investigate relationships between a range of physical factors and the biology and spatial statistics to optimise interpolation, mostly data-driven and (b) rule-based or knowledge-driven models. Mixtures of the two approaches can also be used. The first statistical approach is easier to apply to a single habitat (e.g. mussel beds) than to the complete range of habitats found in an area. The second is similar to the approach adopted for very broad scale mapping and suitable for modelling the range of habitats present. The incorporation of more data-driven statistics into such a model bridges the distinction made between the two approaches. Greater emphasis on statistical models will require the use of specialist statistics and mathematics software and possibly specially written routines.



Scheme of the intermediate-scale mapping approach

An example of an intermediate scale mapping product is presented in the following figure. A probability map (250 m grid resolution) of the macrobenthic communities (EUNIS level 5) of the Belgian part of the North Sea is shown, based on data-driven statistical modelling. Related to the same area, an extract is shown of the EUNIS level 3 MESH Triplet map (1 nm grid resolution). The differences are mostly observed in the coastal zone where the occurrence of one macrobenthic community is completely missed with the broad-scale approach.



Example of an intermediate versus a broad-scale mapping product

Links to other sections:

[Modelling in habitat mapping](#)

[Examples of broad-scale mapping](#)

[Examples of fine-scale mapping](#)

[Examples of intermediate-scale mapping](#)

Links to other websites:

<http://www.marine.ie/home/services/surveys/seabed/>

<http://www.mareano.no/english/index.html>

Modelling in habitat mapping

Models are often divided into **empirical** and **deterministic**. Empirical models are based on a statistical analysis of the data and the results are expressed as probabilities. It is not necessary to understand the causal relationships between a dependent and independent variable in the analysis of the data, just as long as the relationship works well enough to predict distributions. This type of modelling is also termed **data-driven**.

Deterministic models use knowledge of causal relationships to predict distributions. These models can be expressed as mathematical formulae (e.g. the calculation of sea floor currents) or by linguistic 'rules' (e.g. stating the cut-off for light supporting algal growth). Deterministic models are often termed **rule-based** models. **Expert knowledge** is a version of rule-based modelling in that the 'rules' are in the form of expert knowledge and might not be expressed in a systematic way. If expert knowledge and direct observation are applied to coverages in order to draw habitat boundaries (e.g. shore mapping from aerial photos), this is referred to as **direct mapping**.

Mixed models combine elements of both empirical and deterministic elements of modelling.

Additionally, if the modelling involves the manipulation of maps as well as other data within a GIS, the modelling is termed **Cartographic modelling**. Much habitat mapping is based on mixed models within a GIS where various input data are used to predict the distribution of physical habitats likely to support a particular biological community or species. These models are termed **Habitat Suitability models** (HSM). These are then translated into habitat maps on the assumption that the supported habitats will be present.

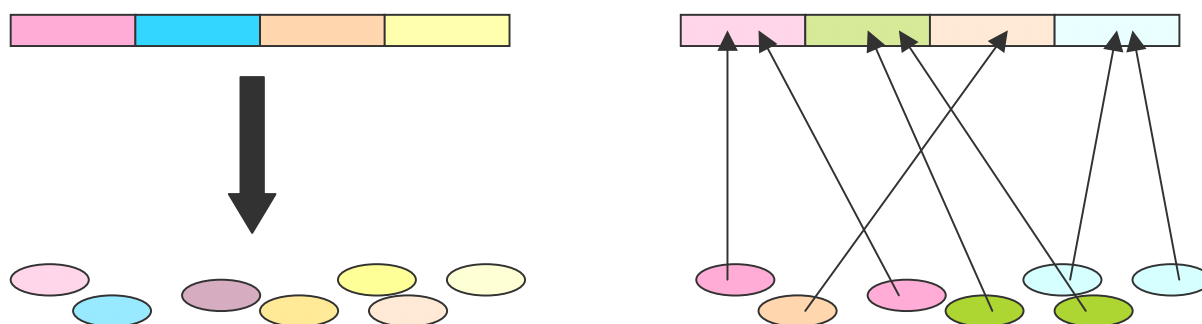
Optimise the ground-truth analysis

The first step is the treatment of the biological records, their analysis and classification to produce data relevant for further integration with physical coverages to finally obtain a biologically relevant habitat map.

For the provision of physical and biological attributes, ground-truth data are a vital element in the mapping process. They are needed to validate the remote sensing data and to assign ground types to the mapped regions (e.g. rock outcrops, sandy beaches). However, it is important to **determine the habitat classes** that will be mapped. Associations need to be established between communities and their associated physical variables as the basis for inferring the distribution of habitats where physical coverages will be available (through sampling or modelling).

There are two fundamental approaches to populate a map with habitat classes:

1. Direct application of an existing classification system ('top-down' classification) by in-situ observation; using expert judgement to match what you see in the field to the classes listed by the scheme; and
2. Determining habitat classes from ground-truth sample data; that is analysing the data to find significant associations between biological and physical parameters ('bottom-up' classification).



Schematic diagram illustrating top-down *versus* bottom-up determination of habitat classes. In top-down (left), the existing classification system (blocks) is imposed on the sample data (circles). In bottom-up (right), similarity between samples is used to determine the classes.

Both approaches require the user to have a significant knowledge of the classification scheme they are using, and how it is structured, so they can apply the scheme correctly and have confidence in the accuracy of the assigned classes.

For the identification of habitats from video imagery reference is made to White *et al.*, 2007.

Links to other sections:

[Route to classifying your records](#)

[Applying an existing classification scheme](#)

[Defining survey-specific habitat classes](#)

[Amalgamating data](#)

Route to classifying your records

The quality and quantity of data gathered depends not only on the limitations of the techniques, but on cost and recording/taxonomic skills available. Whatever format is used, the data are likely to be in the form of a site/species and site/habitat feature spreadsheet. It is assumed that the data have been prepared in this format to an acceptable standard and have been through some form of quality assurance process (taxonomy checked, for example).

What you do with these data now depends upon the purpose(s) of the habitat mapping project. The options are:-

- Applying an existing classification system: there are three methods commonly used to assign a habitat class to records. Firstly, using an 'expert eye' approach by choosing only from those available in an existing classification system (e.g. EUNIS). Secondly, analysing the data and assigning a habitat class from an existing system, but only if the record fits adequately. If the record departs significantly from the given descriptions, then new classes can be constructed for the purposes of the local survey and put forward for consideration for inclusion in an updated version of the parent classification system. Thirdly, where all records also show significant local composition that departs from the given description, analyse the records and extend the general description. Most habitats in a classification are described generally and variability indicated. This variability can be the basis of a further lower level in the classification hierarchy.
- Multivariate analysis to define survey-specific habitat classes: the data are analysed using statistical techniques (usually multivariate) to separate records into natural groupings that show similarity. These become the basis for classes that are of significance primarily for that particular survey. However, it is expected that scientists will endeavour to relate their findings to other similar surveys and that the classes should have some wider significance.
- Measurement of specific habitat features where each record can be assigned a probability that it will support a particular species or habitat class and these probabilities can be mapped – one map per species or habitat. This means that a separate map is produced for each class, such as related to for example: multispecies data (e.g. number of species and other diversity indices); single species (presence/absence; abundance); key habitat factor (e.g. percent silt); and the likelihood of belonging to a particular habitat class

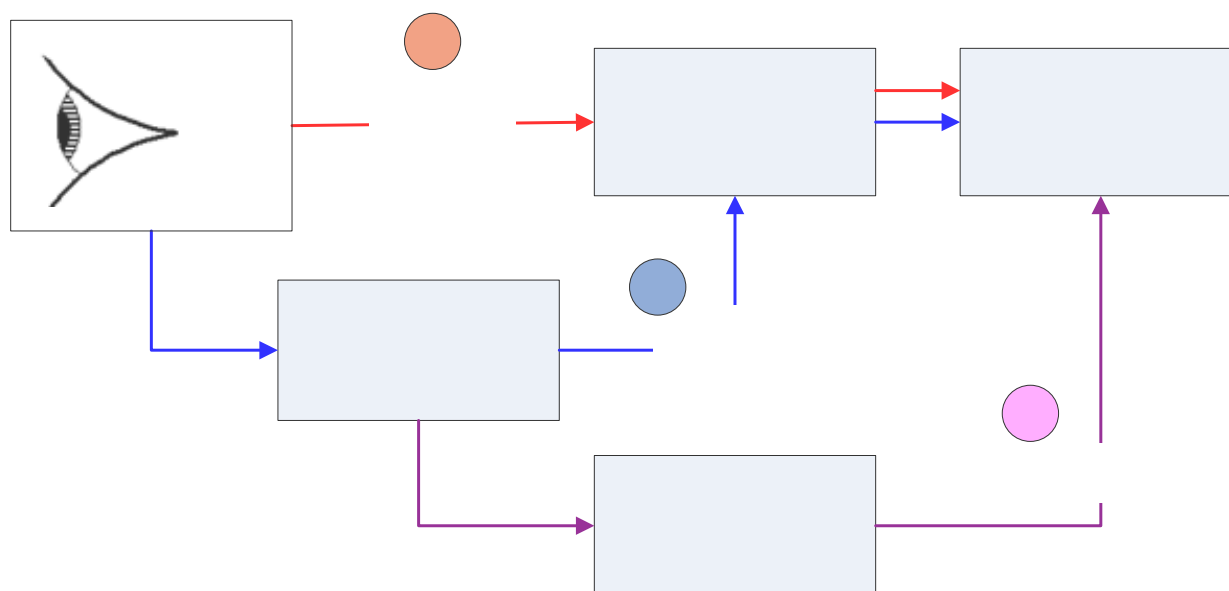
Applying an existing classification scheme

Three approaches to assigning habitat records to an existing habitat classification scheme mentioned in the section 'Classifying your records' are described below.

Expert eye analysis

This can be the most rapid way of determining habitat classes for an experience observer with a good knowledge of the classification scheme. Typically, expert judgement is used to match observations made in the field with the pre-defined habitat classes. Extensive use is made of visual ground-truthing (i.e. direct human observation) where habitats are accessible, such as on 'walk-over' surveys of the shore and diver surveys. An example is provided by an intertidal survey conducted by the Countryside Council for Wales (CCW) where shore surveys have been used

to ground-truth draft maps derived from aerial photographs of the coast. The following web-link summarises the methodology and illustrates the evolution of the final mapped output ([CCW intertidal mapping.htm](http://www.marlin.ac.uk/Conference99/Demonstrations/Countryside%20Council%20for%20Wales/MappinginWales.htm)). (<http://www.marlin.ac.uk/Conference99/Demonstrations/Countryside%20Council%20for%20Wales/MappinginWales.htm>).



Deriving habitat codes from observations

It should be pointed out that there are drawbacks to the reliance on records of location and habitat class alone: The assignment of the habitat class cannot be checked against species or habitat feature records. It is good practise to make notes of species/life form/habitat features and reference photos.

Matching composition of records to habitat class descriptions

The collection of data on species/life form/habitat feature for records permits a more rigorous and systematic approach to assigning habitat class post-survey. Habitat classifications, such as EUNIS, are described in terms of key species and physical features and are accompanied by detailed descriptions of species composition, their expected abundances and variability. The sample records can be matched to, firstly, the key descriptions and then compared to the more detailed information to check for appropriateness.

Automated habitat matching

Assigning habitats on the basis of expert judgement obviously introduces a degree of subjectivity to the interpretation, and the likelihood that different observers may reach different conclusions. Ideally, a more objective approach is required and this can be provided to some degree by an automated system that matches the available ground-truth data to the pre-defined habitat classes. This is the function of the [Habitat Matching Program](#), a software application developed by the MESH project.

The Habitat Matching Program has used an extensive body of data from existing habitat surveys to construct standard definitions of the individual habitat classes (known as *standards*) in the EUNIS classification scheme and the Marine Habitat Classification for Britain and Ireland (Connor *et al.*, 2004). Biological, physical and environmental data from ground-truth surveys are entered into the program, which

then attempts to match the sample data against the *standards*. The output lists the five best matches and provides an indication of the degree of correlation between the sample data and the *standards*. The final assignment of a habitat class requires some judgement by the user, taking into consideration their confidence in the sample data and the 'goodness-of-fit' indicated by the correlation index. Click to obtain more information on the [Habitat Matching Program](#). The program and user manual are available in the resource folder by following these links:

[Habitat Matching Program Manual](#)

[Habitat Matching Program \(EUNIS 2004 report.pdf\)](#)

[MNCR version of Habitat Matching Program \(MNCR 04 05 introduction.pdf\)](#)

Links to other sections:

[Habitat matching program](#)

Links to resources:

[Interpreting intertidal habitats Ifremer.doc](#)

[WE Imares Musselbeds.pdf](#)

[Manual for Habitat Matching Program](#)

[Habitat Matching Program EUNIS 2004 report pdf](#)

[MNCR version of Habitat Matching Program](#)

Links to other websites:

<http://www.marlin.ac.uk/>

Habitat Matching Program

The Habitat Matching Program (HMP) is a software application that automates the classification of benthic marine samples according to the habitat classes defined in the National Marine Habitat Classification for Britain and Ireland (Connor *et al.*, 2004; (<http://www.jncc.gov.uk/MarineHabitatClassification>)) and their equivalent classes in the European EUNIS habitat classification (<http://eunis.eea.europa.eu/habitats.jsp>)

The assignment or matching of sample data to habitats within these classifications at present requires a significant degree of marine biological expertise and can be time consuming. In response to this, a data matching program that aims to improve the efficiency and accuracy of this data assignment process has been developed by JNCC as part of the MESH project.

The program aims to match the user's sample data from intertidal or subtidal surveys against habitat types (known as *standards*) in these classification systems. This is achieved by the use of both comparative and statistical mechanisms. The HMP can accept either or both biological and physical data. The program is aimed at:

- Organisations that undertake benthic marine surveys whose data need to be interpreted according to EUNIS or the Marine Habitat Classification for Britain and Ireland.
- Organisations that have previously assigned survey data to earlier versions of these habitat classification systems and need it updated to the latest (2004) versions.

Main features of the HMP – is it the right tool for you?

The habitat types (standards) are based on survey data used to define the habitat classes in the classification for Britain and Ireland, ranging from intertidal habitats to those down to 200 m depth. Representation of offshore benthic habitats in the standards is more limited; the classification scheme does not currently include pelagic habitats. HMP will attempt to match any survey data undertaken in this range of habitats in the UK and surrounding European waters.

The HMP can analyse up to 250 samples at once. The program accepts biological sample data in semi-quantitative (SACFOR), quantitative (nos./m²) or presence only (1,0) formats. It also accepts physical habitat data according to nine attributes (e.g. substratum, depth).

The HMP provides the five closest habitat type matches to each sample, based on separate analyses of the biological and physical sample data. In addition it can tabulate and compare data in samples (e.g. name and number of biological taxa) against that contained in each of the five matched habitat types (standards).

HMP is a user-friendly program, requiring only a basic working knowledge of MS Excel but the user needs some marine biological knowledge to interpret the results most effectively.

Habitat 'standards'

Files for the **habitat standards** contain biological or physical profiles for each habitat type, which represent the typical characteristics of the habitat. These files are embedded within the HMP and are used to match the data in the sample files to the closest habitat type. Standard files were generated by JNCC based on the survey data (held within Marine Recorder) that was used to define each habitat type in the marine habitat classification for Britain and Ireland (Connor *et al.*, 2004).

How the matching process works

Each new sample record is compared against all the profiles in the relevant standards file, to find those standards, which match most closely to the profile of the sample, using multivariate statistical techniques. The program returns the five most closely fitting standards for both the biological data and the physical data. The results are displayed in a matrix-style format (see figure). The sample name/ code is shown on the far left on each row and the five most closely matching habitat types are listed in order from left to right. The best matches for the biological and physical data are presented together. The HMP currently supplies the five closest matches regardless of how 'distant' the available standards are to each sample. For this reason and the fact that the assigned best matches for the biological and physical data for any one sample can differ, the user can select which habitat type they consider the most appropriate for their sample. To aid this process a 'closeness-of-fit' figure is indicated by the correlation index for each of the five best-match habitat types, and various options to view and compare the results are provided.

The screenshot shows the 'Marine Biotope Matching - 1.7.0.410' window. It has a menu bar (File, Edit, Help) and a toolbar with buttons for 'Load Bio.', 'Load Phy.', 'Run', 'Report', 'Export to: Database', and 'Export to: Spreadsheet'. Below the toolbar is a 'Results - click on names on the left for information' section with an 'Abandon Run' button. This section contains two tables: 'Biological' and 'Physical Data'. Each table has columns for 'Best Match', '2nd Match', and '3rd Match'. The 'Biological' table shows three rows of sample data with their respective matches and statistical values (Max Dist, No Iter., In Com., Not In Com., Pearson). The 'Physical Data' table shows two rows of sample data with their respective matches and statistical values (Max Dist, Dist).

Results - click on names on the left for information			
	Best Match	2nd Match	3rd Match
Biological			
288.014.001_1 Max Dist = 0.5241 No Iter. = 807	CR.HCR.XFa.FluHocu Dist = 0.025111 In Com. = 0.8158 Not In Com. = 37 Pearson = 0.5974	CR.MCR.EcCr.UrtScr Dist = 0.042286 In Com. = 0.2368 Not In Com. = 41 Pearson = 0.2451	CR.HCR.FaT.CTub.Adig Dist = 0.042841 In Com. = 0.3947 Not In Com. = 55 Pearson = 0.2927
288.015.001_1 Max Dist = 0.5182 No Iter. = 425	CR.HCR.XFa.FluHocu Dist = 0.024831 In Com. = 0.7551 Not In Com. = 36 Pearson = 0.5661	CR.MCR.EcCr.UrtScr Dist = 0.036432 In Com. = 0.2449 Not In Com. = 46 Pearson = 0.3457	CR.HCR.FaT.CTub.Adig Dist = 0.037983 In Com. = 0.4082 Not In Com. = 56 Pearson = 0.3769
289.007.001_1 Max Dist = 0.7488 No Iter. = 1022	CR.HCR.XFa.FluHocu Dist = 0.035867 In Com. = 0.5854 Not In Com. = 54	CR.MCR.CSab.Sspi.ByB Dist = 0.039312 In Com. = 0.4146 Not In Com. = 47	CR.MCR.CFaVS.CuSpH.VS Dist = 0.042082 In Com. = 0.2683 Not In Com. = 46
Physical Data			
467.023.001 Max Dist = 8.061	CR.HCR Dist = 0.95254	CR.MCR Dist = 1.3005	IR.HIR Dist = 1.7977
467.025.001 Max Dist = 7.988	CR.HCR Dist = 0.99058	CR.MCR Dist = 1.3213	IR.HIR Dist = 1.6227

The **Results Matrix** showing the sample labels (left column) and three of the five corresponding best matches for both biological and physical data, as determined by the HMP analysis.

Further information

Further technical information on the Habitat Matching Program is available here: ([Technical Summary and Examples.doc](#))

The full User Manual is available here: ([Manual for Habitat Matching Program v2.pdf](#))

The Habitat Matching Program will be released on the MESH web site (<http://www.searchMESH.net>) later in 2007.

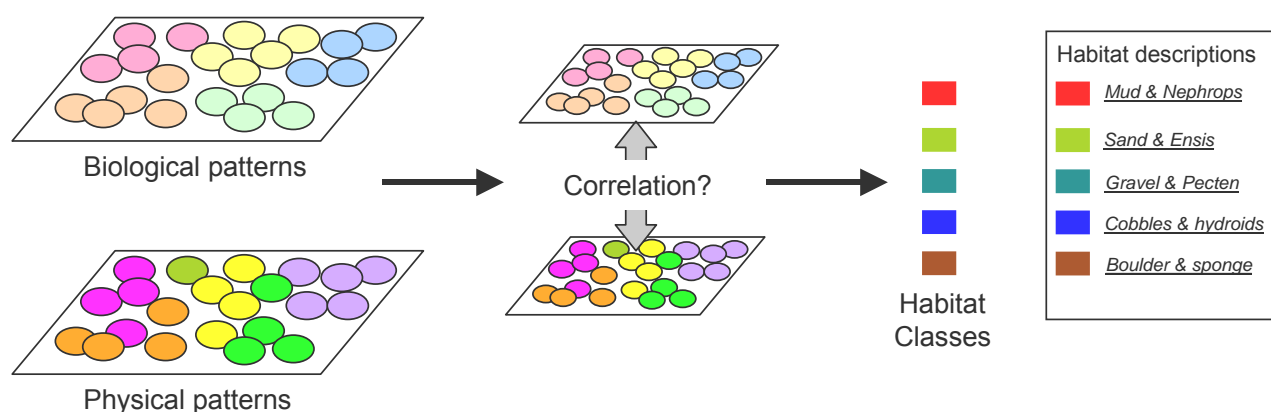
Defining survey-specific habitat classes

Sample records of species and physical features can be used to create classes based on similarity/dissimilarity. Multivariate analysis of these records is an empirical approach to determining habitat classes (often characterised by the expression 'letting the data speak for themselves') and makes maximum use of the biological, physical and environmental ground-truth data. However, the analysis is usually directed in some way to accentuate key species or rarities (see 'transformation'). In principal, it uses statistical analyses of many different samples to find significant associations between biological communities and environment conditions in which they live. These associations can then be used to define and classify different 'natural' habitats.

In concept, the process of establishing associations between biological and physical data and defining habitat characteristics is simple. Given that we have a set of

samples for which we have both biological and physical/environmental data, the process follows four steps:

1. [Look for patterns / clusters in biological data](#)
2. [Look for patterns / clusters in physical data](#)
3. [Correlate biological and physical patterns to produce 'habitat' clusters.](#)
4. [Determine the major characterising species and physical properties for each habitat](#)



Deriving biological and physical patterns and their correlation into habitat classes.

Similarity/dissimilarity is typically measured using the Bray-Curtis similarity/dissimilarity coefficient. The coefficients between each pair of samples are prepared as a matrix. Thereafter, two complementary multivariate techniques are used to help find natural groupings within the sample data: ORDINATION and CLASSIFICATION.

There are a number of analytical packages that are routinely used in benthic ecology:

- **PRIMER (Plymouth Routines In Multivariate Ecological Research)** offers ordination by non-metric multidimensional scaling (MDS) and principal components (PCA) to summarise patterns in species composition and environmental variables and hierarchical clustering into sample (or species) groups (CLUSTER and SIMPROF) (Clarke and Warwick, 2001; Clarke and Gorley, 2006). PRIMER identifies variables (species) that contribute most to sample similarities (SIMPER); analysis of similarities/differences between groups of samples (ANOSIM, an analogue of univariate ANOVA tests); testing relationships in multivariate patterns between two sets of samples (RELATE); identifying environmental variables that “best explain” the patterns in the biological data (BEST). PRIMER also has a wide range of univariate, graphical and other multivariate routines for analysing biological and physical/environmental data.
- **DECORANA (Detrended Correspondence Analysis)** is an ordination technique and is complemented by **TWINSPAN (two-way indicator species analysis)**, a tool for classifying species and samples, producing an ordered two-way table of their occurrence (Hill, 1979a,b). The process of classification

is hierarchical; samples are successively divided into categories, and species are then divided into categories on the basis of the sample classification.

- CANOCO is an extension of DECORANA (Hill, 1979b). It includes the indirect techniques of principal components analysis (PCA), (detrended) correspondence analysis and principal coordinates analysis and also the direct techniques of weighted averaging, canonical correspondence analysis, canonical variates analysis (= linear discriminant analysis) and redundancy analysis (ter Braak, 1986 and 1988). CANOCO can also test whether species are related to measured environmental variables using a Monte Carlo permutations test (ter Braak, 1988).

Data preparation

It is assumed that the data will have been prepared for analysis. In summary, this should include assurances on the following points:

- Taxonomy is correct, at least to a stated taxonomic level (or life form if uncertain) and the nomenclature is standardised to eliminate name changes (synonyms) between surveys.
- Only data from comparable techniques are amalgamated (not cores and grabs, for example).
- Where different sampling devices have been used, such as a 0.1m² grab and a 0.25 m² grab, the data should be *standardised* to remove sampling bias, by expressing abundances as number per square metre. Standardisation as percentages may also be used, but may be misleading.
- It may also be desirable to *transform* the data in order to down-weight the influence of highly abundant taxa as these can sometimes mask underlying biological patterns.
- Alternatively, it may be desirable to select key species to stress structuring or functional species and downplaying rarities, perhaps eliminating very rare species from the dataset altogether.
- In physical data sets, it is usual that a number of different variables have been measured, such as temperature (°C), salinity (‰) and depth (m). As these variables are measured in different units, the data needs to be *normalised* to put them on a unitless (and therefore comparable) scale.

Cautionary notes

The application of analytical techniques is a skilled process and requires the user to be aware of the limitations of the sample data and how the choices made during analysis can influence the results. 'Expert judgement' is always required in interpreting the results.

- The transformations employed vary as to the extent to which they weight rare species as compared to abundant species. The most severe transformation is, in fact, reducing abundance to 'presence/absence'. The SACFOR system is also a form of transformation. Counts can be transformed using square root, double square root or log. Transformation may have a very marked effect on the nature of the clusters determined.

- It is often inappropriate to include too wide a variety of habitats for multivariate analysis since the results will reflect obvious splits (gravel habitats versus mud) and hide more subtle and 'interesting' differences within these groups. It may be best to analyse very different habitats separately.
- Infaunal and epifaunal communities are rarely sampled effectively by a single technique, so it is normal to sample each component separately and to analyse them separately, leading to separate sets of community classes. Infaunal analyses will be more appropriate for unconsolidated sediments (mud, sands, gravels) and epifaunal analyses for consolidated sediments (rock, boulder). Some difficulty is to be expected for intermediate substrata such as cobbles and pebbles, where both infauna and epifauna characterise the habitat.

Biotope descriptions in the Marine Habitat Classification for Britain and Ireland (Connor *et al.*, 2004) and the European Nature Information System (EUNIS) classification tend to favour epifaunal species on harder substrata (as these are easily observable) and infaunal species on softer substrata (where they are more abundant than the epifauna). Relatively few biotopes have extensive descriptions of both infauna and epifauna. It is important to recognise these limitations of the current biotope descriptions and that your samples may provide more comprehensive data on both infauna and epifauna.

There is little current literature examining epifaunal/infaunal links. If such links could be found, biotope assessments for ground-truthing purposes would be simplified and made more cost-effective. The associations between infauna and epifauna are being addressed by a separate MESH study ([Report on infaunal / epifaunal links](#)).

Link to Resources:

[Report on infaunal / epifaunal links](#)

[Technical Summary and Examples.doc](#)

[Manual for Habitat Matching Program v2.pdf](#)

Link to other sections:

[Patterns in biological data](#)

[Patterns in physical data](#)

[Correlate biological and physical patterns](#)

[Determine the major characteristics](#)

[Match classes with classification system](#)

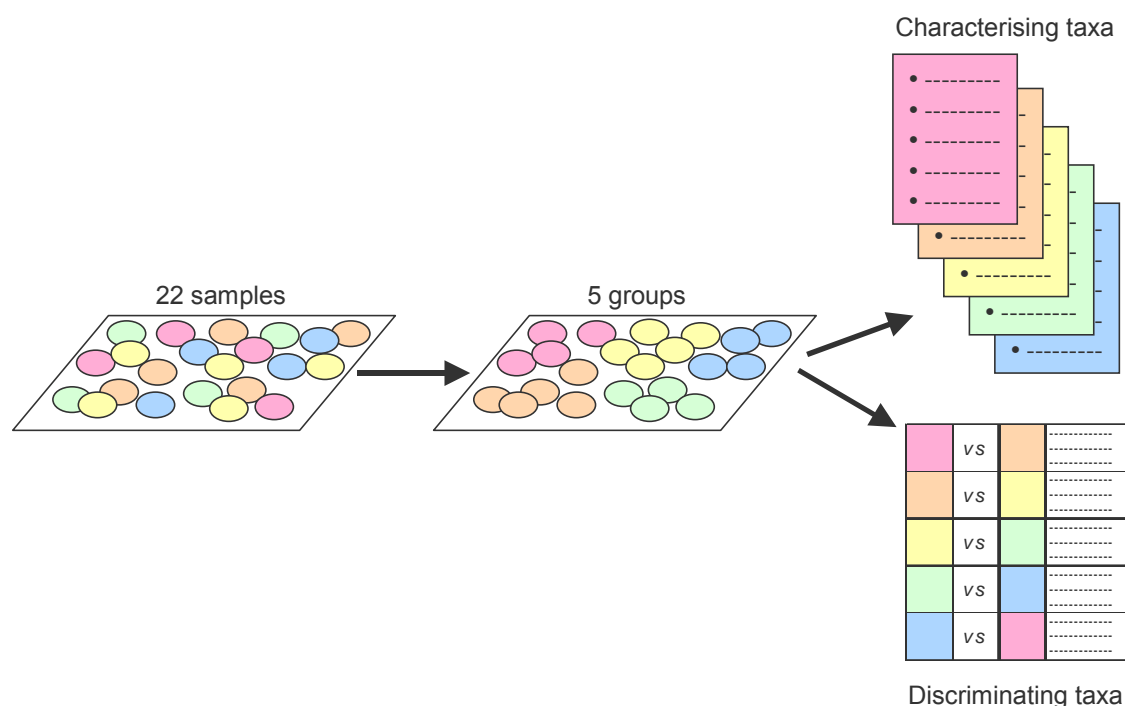
Links to websites:

<http://www.searchMESH.net>

Patterns in biological data

The goal here is to analyse the biological data from ground-truth sampling to identify and characterise distinct 'natural' communities. It requires an adequate number of samples and starts with the null-hypothesis that there is no difference in community

structure between the samples (i.e. they contain the same types of species in the same proportions). If this proves incorrect, the null hypothesis can be rejected and the samples sorted into groups (or 'clusters') that have similar communities. This establishes a pattern in the biological data, which can then be further examined to pick out species that characterise the samples within a group and those that differentiate one group from another.



Grouping sample data and deriving discriminating taxa.

Data preparation

Some form of data preparation is usually needed prior to the analysis. As well as ensuring that taxonomic identification is consistent between samples (i.e. no synonyms have been used), many data sets need to be standardised and/or transformed (e.g. method of expressing abundance). Sometimes this can be done within the analysis package.

Standardisation

Obtaining samples from the seabed can be subject to a fair degree of 'sampling error' or variability. For example, it is very difficult to ensure consistency in the collection of trawl samples as the distance covered by the trawl can vary considerably between tows, even if they are deployed for a fixed time (e.g. 10 minutes), due to the influence of tide and weather conditions on the speed of the towing vessel. Consequently, the data need to be standardised to remove this sampling variability. For trawl samples, this can be achieved by calculating the area swept by each tow (distance covered x width of trawl) and expressing the abundance (or biomass) per unit area of seabed (e.g. per 1 m², 10 m², 100 m² or 1 km², as appropriate). For grab and core samples, the accepted convention is to express abundance (or biomass) per unit area of seabed sampled, rather than per unit volume of material sampled. This is why grab and corer specifications refer to an area (e.g. 0.1m² Day grab, 0.2 m² Box corer etc). A universal method of

standardising abundance or biomass data is to express it in unit-less form; that is to convert it to a percentage.

An example table of standardising abundance data for two trawl samples covering different swept areas (200 sq m for Trawl 1, 390 sq m for Trawl 2). Although the actual abundance data appear quite different, the standardised data (Abundance per 100 sq m and % Abundance) show the samples were indeed very similar.

Species	Actual abundance		Abundance per 100 sq m		% Abundance	
	Trawl 1	Trawl 2	Trawl 1	Trawl 2	Trawl 1	Trawl 2
A	325	630	162.5	161.5	68.9	67.5
B	15	40	7.5	10.3	3.2	4.3
C	2	4	1.0	1.0	0.4	0.4
D	35	80	17.5	20.5	7.4	8.6
E	92	175	46.0	44.9	19.5	18.7
F	3	5	1.5	1.3	0.6	0.5
Sum	472	934	236.0	239.5	100	100

Transformation

In biological communities, it is common for there to be large numbers of small animals and relatively few of the larger species. This can cause a problem in community analyses based on abundance counts, because one or two highly numerous species will appear to be the only significant descriptors of the community, which may in fact be composed of a hundred or more species, some of which may be large and characteristic of the habitat (e.g. edible crab) but present in limited numbers. Hence the data need to be balanced in some way to down-weight the influence of the highly numerous species and to raise the profile of the less abundant species. This is achieved by simple mathematical transformation of the raw abundance data, as illustrated in the table. The square root, 4th root and Log(X+1) transformations have a progressively more severe effect on the original abundance data. The presence-absence (P/A) transformation is the most severe of all, removing abundance data entirely. The choice of which transformation to use is mostly down to experience and judgement, though the analyses can be repeated in a systematic way to test progressively more severe transformations.

Abundance	Transformations			
	Square Root	4th Root	Log (x+1)	P/A
0	0.00	0.00	0.00	0
1	1.00	1.00	0.30	1
10	3.16	1.78	1.04	1
100	10.00	3.16	2.00	1
1000	31.62	5.62	3.00	1
10,000	100.00	10.00	4.00	1
100,000	316.23	17.78	5.00	1
1,000,000	1000.00	31.62	6.00	1

An example table of the transformation of abundance data.

Some life forms do not lend themselves to measures/counts of abundance, such as colonial species (e.g. sponges, hydroids, bryozoans), yet these may be a major characterising fauna of certain habitats, especially cobble reefs, rock outcrops and stable gravels and sands. It is therefore quite crucial to include these fauna in the

analysis, so it may be preferable to work with biomass rather than abundance data. Presence-Absence data have a particular value in enabling an initial assessment that includes all life forms (i.e. colonial and non colonial forms), and is a useful starting point for the analysis, which may then be further refined to use abundance or biomass data. The use of abundance data normally excludes any consideration of colonial species as they have typically been scored on a non-numeric, relative abundance scale (e.g. SACFOR, as described earlier) and the analytical programs cannot cope with non-numeric data.

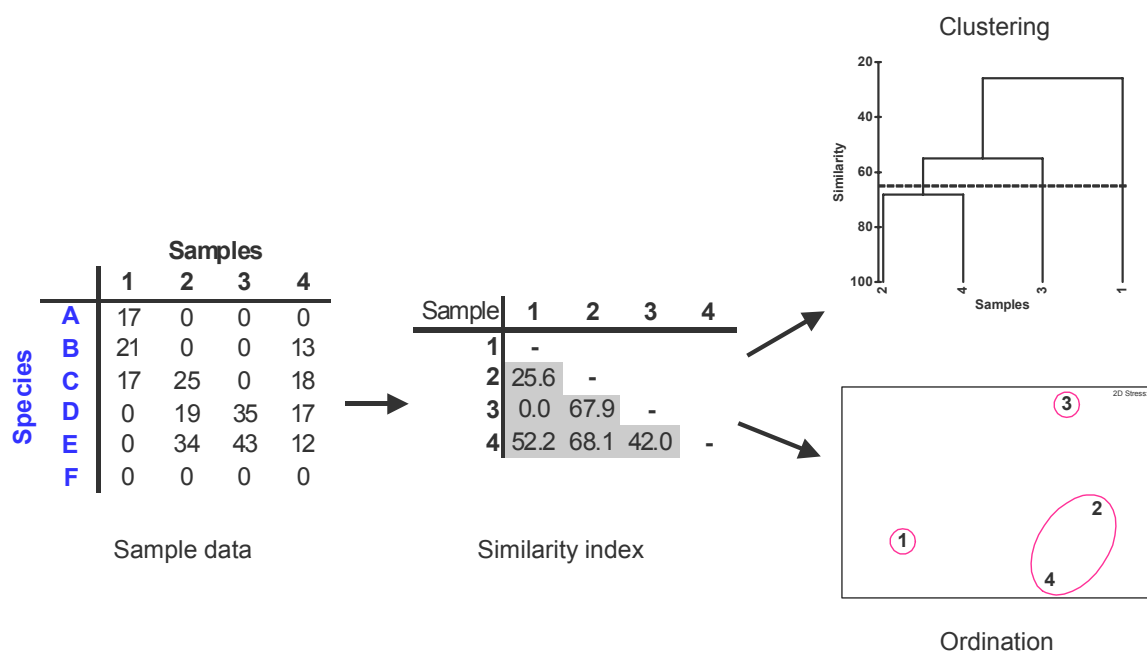
In this case, the relative abundance or biomass records can be converted to “numerical equivalents” by assigning what would appear to be an appropriate abundance. The methods for doing this consistently are at an early stage of development, but it should not be a conceptual barrier if you are aware of the effect that any likely data transformations will have on the assigned values. It is more important to give an assigned value the correct order of magnitude than a precise number, as can be seen in the table:

Assigned value	Sq root	4th root	Log (x+1)
10	3.16	1.78	1.04
20	4.47	2.11	1.32
100	10.00	3.16	2.00
200	14.14	3.76	2.30
1000	31.62	5.62	3.00
2000	44.72	6.69	3.30

Methodology

The most commonly used method for determining patterns in biological data is cluster analysis. The similarities between samples can be visualised by plotting them in a dendrogram (cluster analysis) and/or an ‘ordination’ plot (see further). There are a number of computer programs available for carrying these analyses: e.g. PRIMER (<http://www.primer-e.com>), MVSP (Multi Variate Statistical Package; <http://www.kovcomp.com>) and CLUSTAN (Cluster Analysis; <http://www.clustan.com>)

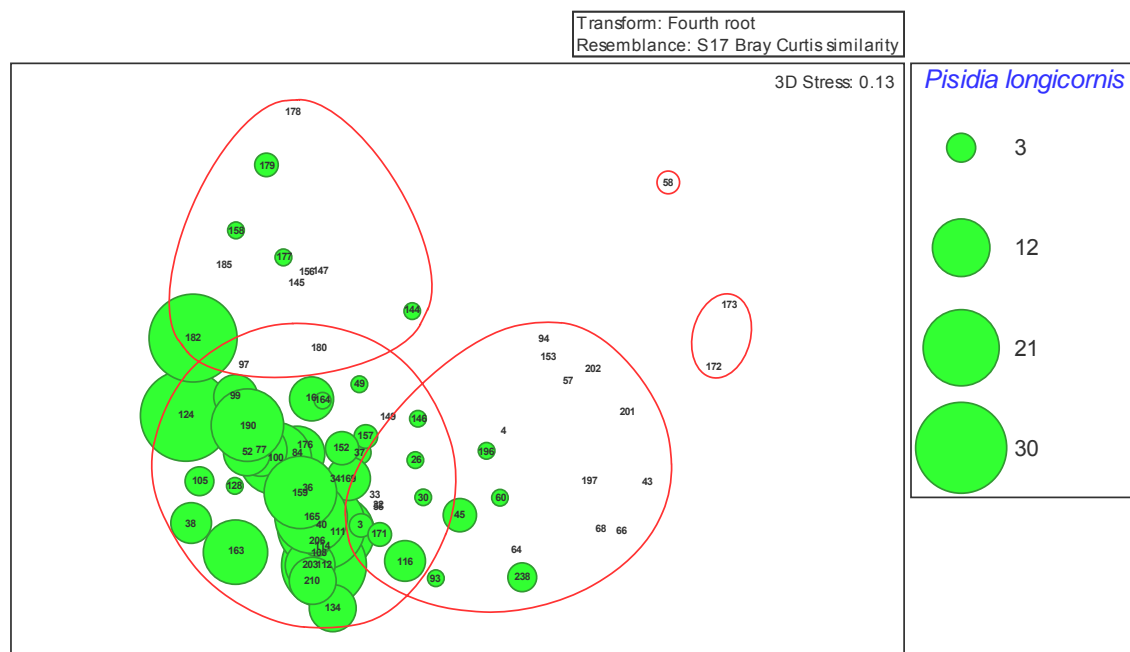
Cluster analysis is a technique whereby the species composition of each sample is compared with every other sample (pair-wise comparison) and is suitable for presence-absence, abundance or biomass data. In Europe, the most commonly used comparative index is the Bray-Curtis similarity/dissimilarity coefficient. A triangular matrix of inter-sample similarities is produced and used in an agglomerative clustering procedure where samples that are most similar to each other are grouped together. These groupings are represented in a cluster diagram (dendrogram), depicting the hierarchical similarity between samples and groups of samples (see diagram).



Schematic diagram of cluster analysis. Starting with a species-by-samples data matrix, a similarity index is calculated (centre) and represented diagrammatically as a dendrogram (cluster plot) and/or an ordination with similarity contours (after Clarke & Warwick, 2001).

The ordination plot is another way of representing these similarities/dissimilarities. Sample points are plotted in a multidimensional space such that the distance between points represents their relative similarity/dissimilarity (i.e. the closer the points, the greater their similarity). A popular method, known as 'non-metric Multi Dimensional Scaling (or MDS for short), uses ranked similarity values, so the two closest points are the most similar and the two most distant points are the least similar. In reality, the plot has $n-1$ dimensions, where n = the number of sample points, but to aid interpretation this is 'compressed' into a 2-dimensional or 3-dimensional plot, accompanied by a 'stress' value which indicates how well the 2D (or 3D) plot depicts full multi-dimensional ordination.

The interpretation of biological MDS ordinations can be assisted by overlaying similarity contours (derived from the cluster analysis) to highlight the cluster groups. For example, in this MDS plot for epifaunal data, three major cluster groups are delineated (red contour lines). The sample points can also be overlain with scaled abundance symbols (a 'bubble plot') where the size of the bubbles are indicative of the relative abundance of a chosen species at each sample station. Here, the long-clawed porcelain crab *Pisidia longicornis* is clearly a characterizing species for one of the clusters.



Example of a 2-dimensional representation of an MDS ordination for epifaunal data. Sampling stations are labelled with their sequential Station Number. The red lines represent a 42% similarity contour, so all samples inside the contour have a similarity > 42%. At this level of similarity, three major clusters are evident. A bubble plot showing the relative abundance of the Porcelain crab, *Pisidia longicornis*, overlies each sample point. A scale for the bubble plot is shown on the right.

Patterns in physical data

The goal here is to identify patterns in the physical and environmental parameters that determine habitat type. Substrata (rock, gravel, sand mud etc) is usually the principal factor as each different type provides different living conditions and so is usually inhabited by recognisably different communities. Those conditions are further modified by environmental gradients such as depth, exposure (to wave and tidal energy), salinity etc, forming a multidimensional matrix of physical habitats. Analysing the physical and environmental data indicates which parts of this matrix your sample represents and which are the principal components that characterise sample groups and differentiate between them.

Data preparation

Only true variables such as the % gravel, sand or mud in sediments, or actual temperature and salinity measurements can be used on numerical analyses. Sometimes it may be desirable to simplify the analysis by converting variable data into categorical classes, which are then used as factors. This is a common approach with sediment data, using the Folk triangle to classify sediment samples into categories such as 'gravelly-sand', 'sandy-gravel', 'sandy-mud' etc, as seen on many seabed sediment maps.

Many environmental factors or descriptors are recorded as categorical data, such as wave exposure (exposed, sheltered, extremely sheltered) or biological zones (eulittoral, infralittoral, circalittoral etc). Often these are given shorter codes, which are easier to handle in spreadsheets and read on graphical outputs. If numerals are

used as codes it is important to remember that the data is still categorical and cannot be used in numerical analyses.

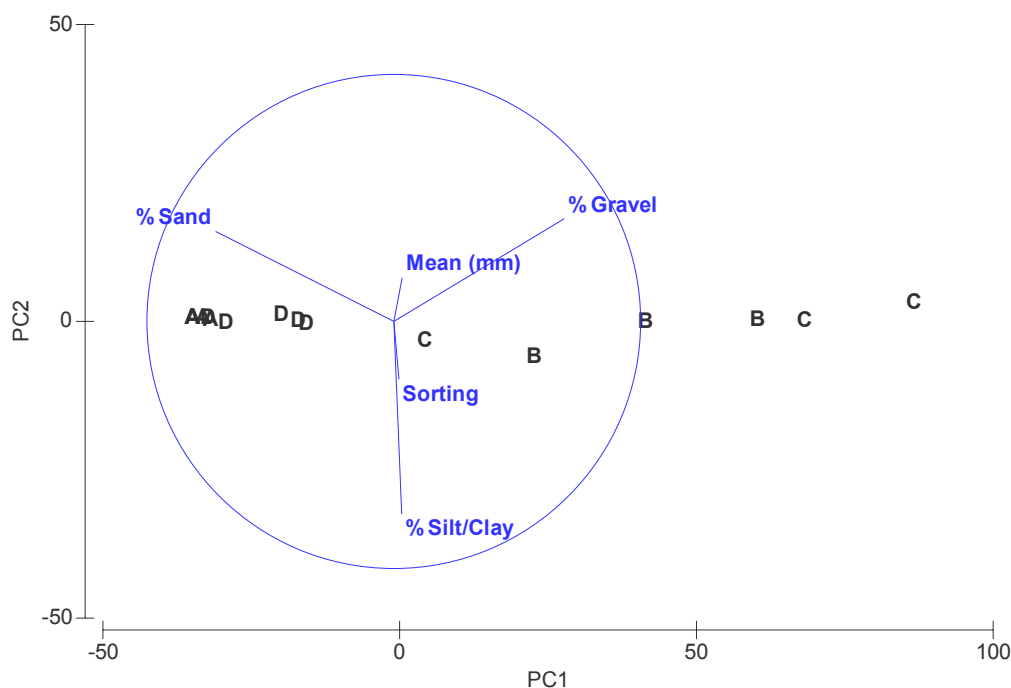
Where a suite of physical or environmental variables has been measured, the data will be expressed in a variety of different units (e.g. temperature in °C, salinity in ‰, current speed in knots or metres per second). It is futile to attempt numerical analysis using the absolute data values, as the data set with the largest range will always appear to be the most influential. To overcome this, the data need to be expressed as relative, 'unitless' values. This is achieved by a simple mathematical operation called '*normalisation*', which uses the mean and standard deviation of the data set. The mean is subtracted from each data value and the result divided by the standard deviation.

Temperature (deg C)			Salinity (parts per thousand)		
A	B	C	A	B	C
(data value)	(A - mean)	(B/StDev)	(data value)	(A - mean)	(B/StDev)
5	-4.5	-1.49	33.1	-0.45	-1.49
6	-3.5	-1.16	33.2	-0.35	-1.16
7	-2.5	-0.83	33.3	-0.25	-0.83
8	-1.5	-0.50	33.4	-0.15	-0.50
9	-0.5	-0.17	33.5	-0.05	-0.17
10	0.5	0.17	33.6	0.05	0.17
11	1.5	0.50	33.7	0.15	0.50
12	2.5	0.83	33.8	0.25	0.83
13	3.5	1.16	33.9	0.35	1.16
14	4.5	1.49	34.0	0.45	1.49
range	9.00	2.97	range	0.90	2.97
mean	9.50	0.00	mean	33.55	0.00
StDev	3.03	1.00	StDev	0.30	1.00

In the example, the temperature values range from 5 to 14°C, and salinity from 33.1 to 34.0‰. There appears to be very little similarity between these data sets when the absolute values are inspected (column A), they even have different ranges, means and standard deviation. Normalising the data removes their dependence on the units of measurement. Comparing the normalised data (column C) shows that temperature and salinity showed identical patterns.

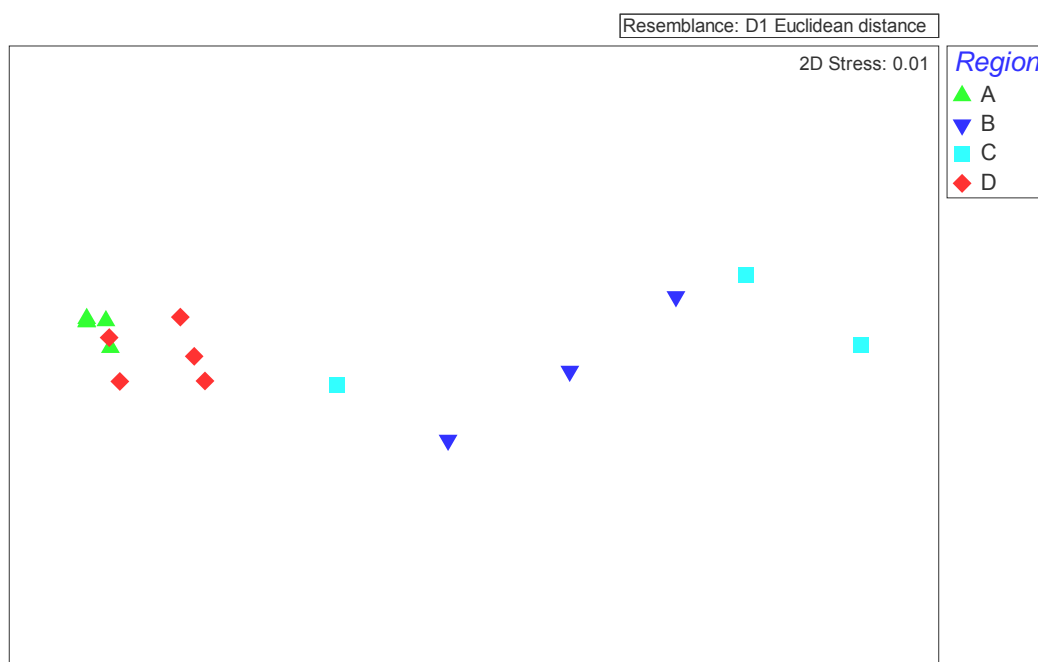
Methodology

There is a variety of ways that patterns in physical data can be determined. Perhaps the most widely known is Principal Components Analysis (PCA). This ordination method simplifies the physical dataset by transforming the data to a new coordinate system such that the greatest variance lies in the first coordinate (First Principal Component), the second greatest variance then forms the Second Principal Component and so on. In this method of indirect gradient analysis, samples are spread out relative to the PCA axes. The principal components represent linear combinations of the variables.



Graphical output from a Principal Components Analysis (PCA) on sediment samples from four areas (A to D) in the Hastings Shingle Bank in the English Channel (Brown *et al.*, 2001). The variables were mean particle size (mm), sorting coefficient, % gravel, % sand and % silt/clay content. The greatest variance (along PC Axis 1) is clearly driven by the sand:gravel content of the samples.

Nevertheless, the same techniques used in analysing biological data can also be applied to physical data. Thus cluster analysis and MDS plots offer simple methods for pattern determination. In contrast to the biological data, however, the Bray-Curtis similarity coefficient is not appropriate, because in physical data zero has no special meaning; it is simply one point on a scale (in biological data zero indicates the absence of a species). As the variables will likely be on different scales, normalising procedures will produce negative and positive values. Distance coefficients such as Euclidean Distance are therefore the preferred measure of similarity for physical/environmental data sets.



MDS plot of the same set of samples displayed in the PCA image (Brown *et al.*, 2001). The tighter clustering of samples from regions A and D indicates they were similar and more consistent in their composition than samples from regions B and C.

Correlate biological and physical patterns

A number of methods are available to examine the correlation between biological and physical patterns. Two commonly used computer packages with a variety of multivariate methods are PRIMER and CANOCO.

PRIMER

PRIMER comprises a wide range of univariate, graphical and multivariate routines for analysing biological and physical/environmental data (Clarke & Warwick, 2001; Clarke & Gorley, 2006). 'BEST' and 'LINKTREE' are two routines targeted at linking multivariate biological patterns with single or multiple environmental variables.

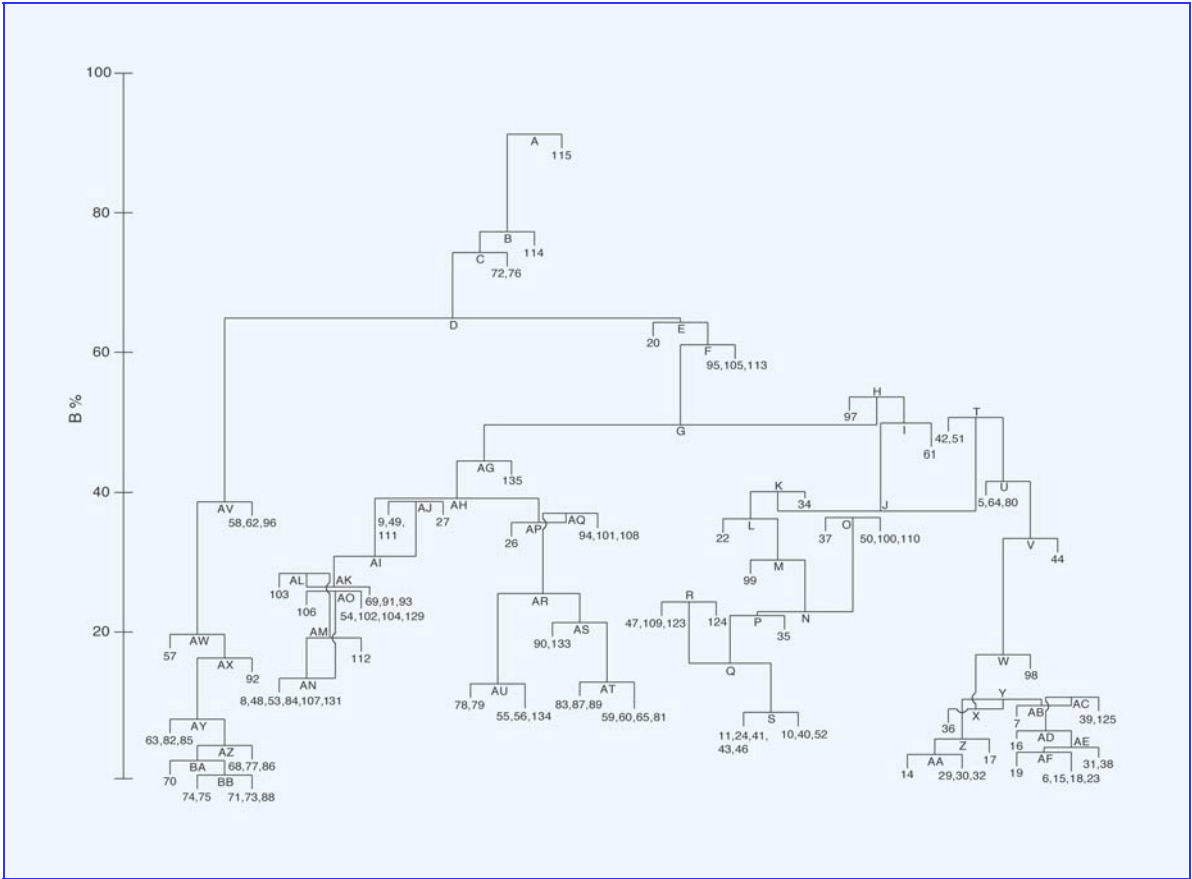
The BEST routine available in PRIMER v6 combines the BIO-ENV and BV STEP procedures found in PRIMER v5. BIO-ENV uses all the available environmental variables to find the combination that 'best explains' the patterns in the biological data. However, when large numbers (>15 or 16) of environmental variables are used the procedure can become impractical, as computation time may be excessive. In such cases the BVSTEP option can be employed to carry out a stepwise search of the variables, employing both *forward selection* and *backward elimination*. Starting with the variable showing the maximum matching coefficient, variables are successively added, the combinations tested and (at each stage) the variable contributing least eliminated. Several iterations of the procedure are carried out from a random selection of (e.g. ≤6) variables to ensure that the 'best' match is found.

An example table of a BIO-ENV analysis in relation to benthic polychaete distributions in the Irish Sea (from Mackie *et al.*, 1997). A combination of three environmental variables (%gravel, %silt and depth) provides the best match to the patterns observed in the biological data.

Number of variables	Best variable combinations	Correlation (p_w)
3	Gravel-Silt-Depth	0.70
	Gravel-Silt/Clay-Depth	0.69
	Gravel-Carbon-Depth	0.68
4	Gravel-Silt-Depth-Latitude	0.69
	Gravel-Silt-Carbon-Depth	0.69
	Gravel-Silt/Clay-Depth-Latitude	0.68
	Gravel-Silt/Clay-Carbon-Depth	0.68
5	Gravel-Silt-Carbon-Depth-Latitude	0.68
	Gravel-Silt/Clay-Carbon-Depth-Latitude	0.68

The LINKTREE routine takes the combination of variables that were identified as 'best' in BIO-ENV together with the faunal inter-station similarities to find the most effective way of describing the biological-environment relationships relative to the successive use of single variables. Starting with the group of all samples, it divides them into two groups (a binary split), determined by the most influential environmental variable(s). So, the first split could be on the grounds that the two resulting groups are most dissimilar in terms of their salinity. By iteratively repeating this procedure on the resulting groups, the samples are divided into a number of groups, within which all the samples have similar biological and physical characteristics. Expressed more technically, the group of samples is successively divided according to the environmental variable(s) that maximise the separation between the groups in multidimensional space. Sometimes more than one variable is determined at a split (if variable each gives the same result). A statistical test is used to examine the significance (5% level) before each split, with division stopped when non-significant. An output value (B%, see table) provides an absolute measure of group differences, and low values occur when samples are most similar.

This is divisive clustering, as opposed to agglomerative in cluster analysis, and inversions can sometimes occur in the clustering pattern. Unlike BIO-ENV the environmental variables are non-additive and one advantage is that a variable can be identified as important in part of the overall faunal distribution, yet not so in other parts (conversely, BIO-ENV examines the overall wider situation). The LINKTREE procedure also has potential for prediction: if the environmental conditions are known for a new sample station, then the LINKTREE results may allow it to be assigned to a particular assemblage or group of sites.



An example of a LINKTREE analysis in PRIMER, repeated binary splits dividing the samples into groups with similar biological and physical properties. The following table shows that at node G, the samples are split into two further groups, AG and H, determined by their percentage sand content (see first line of following table).

An example table of the part of the descriptive information for a LINKTREE analysis of benthic macrofaunal distributions in the Outer Bristol Channel (from Mackie *et al.*, 2006)

Node/Stn split	Variable	LHS (RHS) Split	π	Sign.	(p) R	B%
G→AG, H	Sand	>90.4 (<87.9)	7.09	0.1	0.39	49.6
AG→AH, (135)	Mean phi, Depth	<1.99 (>2.08) <52.1 (>52.1)	5.30	0.1	0.41	44.5
AH→AI, AP	Mud	<0.03 (>0.04)	5.36	0.1	0.39	39.2
AI→AK, AJ	Mean phi	>1.11 (<1.1)	4.00	0.1	0.40	30.8
AK→AL, (69, 91, 93)	Depth	>25 (<21)	3.66	0.1	0.48	25.9
AL→103, (AM)	Depth, Mud	>49 (<46) 0 (>0)	3.79	0.1	0.56	28.4

CANOCO

CANOCO is a computer program for **CAN**onical **C**ommunity **O**rdination by (partial/detrended/canonical) correspondence analysis, principal components analysis and redundancy analysis (ter Braak, 1986 and 1988), that originated as an extension of DECORANA (Hill, 1979b). Over the last 20 years it has evolved to include a variety of multivariate ordination methods and the current version (4.5) is available with a Microsoft Windows interface (ter Braak and Smilauer, 2002).

Jongman *et al.* (1995) provide a detailed account of the theory and implementation of the various techniques.

Ordinations, like cluster analysis, are 'indirect' methods of analysing species-environment relationships since additional procedures are necessary to correlate the biological patterns to the environmental variables. Canonical (or constrained) analyses overcome this by integrating ordination with regression.

The methods available fall into four categories:

- **Unconstrained ordinations** describe the structure in a single data set
- **Canonical ordinations** explain one data set by another data set (ordinations are constrained by explanatory variables)
- **Partial ordinations** describe the structure in a data set after accounting for variation explained by a second data set (co-variable data)
- **Partial canonical ordinations** explain one data set by another data set after accounting for variation by a third data set (co-variable data)

ter Braak & Verdonschot (1995) examine the use of Canonical Correspondence Analysis (CCA) in aquatic ecology and this technique is the most commonly used direct gradient analysis method. It has been widely used in marine benthic situations, from the intertidal to deep water (Ysebaert & Herman, 2002; Narayanaswamy *et al.*, 2003; Bergquist *et al.*, 2005). In CCA the ordination axes are derived from linear combinations of the environmental variables such that the dispersion of the species (and sample) scores are maximised. Environmental variables are shown on the ordinations as arrows directed from the origin of the plot where the origin represents the grand mean for each variable. Longer arrows are more strongly correlated with the ordination axes than short ones.

In the following example, CCA was employed to investigate the species-environment relationships of benthic polychaetes in the Irish Sea (Mackie *et al.*, 1997). Forward selection of the variables revealed seven that 'best' explained the data. At each step, a Monte Carlo permutation test was used to determine the significance of each variable. The first five variables were highly significant ($P < 0.0001$), the others less so ($P < 0.05$). The seven variables collectively explained 34.75% of the total inertia.

An example table of the forward selection of variables in a study of the distribution of benthic polychaetes in the Irish Sea

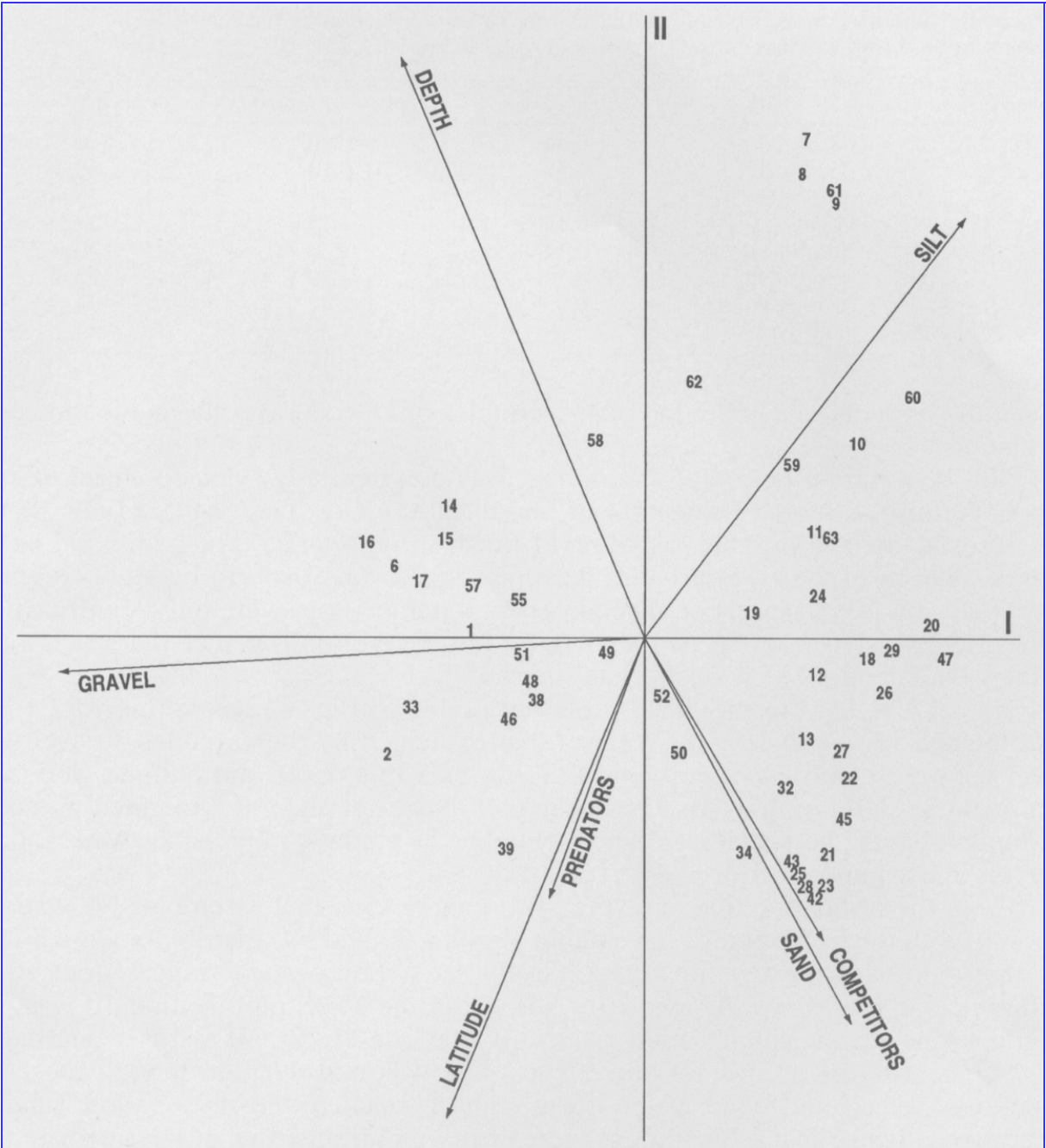
Selection order	Variable	Variance explained	Cumulative variance	Cumulative % explained
1	Gravel	0.46	0.46	10.87
2	Depth	0.32	0.78	18.44
3	Silt	0.21	0.99	23.40
4	Sand	0.15	1.15	27.19
5	Latitude	0.14	1.29	30.50
6	Predators	0.10	1.39	32.86
7	Competitors	0.09	1.47	34.75
All 12 variables			1.90	44.92

In the ordination, Axes I and II were the most important accounting for 21.3% of the species variance and 61.2% of that explained by the variables.

An example table of CCA ordination summary for the polychaete-environment relationship

Axes	I	II	III	IV	I-IV
Eigenvalues	0.520	0.382	0.178	0.147	1.227
Species-environment correlations	0.955	0.932	0.875	0.917	
Cumulative percent variance					
of species data	12.3	21.3	25.5	29.0	
of species environment relation	35.3	61.2	73.3	83.3	
Sum of all unconstrained eigenvalues				4.231	
Sum of all canonical eigenvalues				1.473	

As can be seen by the ordination plot and the correlation table, sediment gravel content was most influential for axis I. Depth and latitude were most important in defining axis II. Variables such as depth (and latitude) may however be proxies for other co-varying factors (e.g. temperature, pressure, currents) rather than the variable itself.



An example table of ontraset correlations of environmental variables for axes I-IV

Axis variable	Axis			
	I	II	III	IV
Gravel	-0.922	-0.049	-0.153	0.168
Depth	-0.388	0.789	0.241	-0.391
Silt	0.503	0.570	-0.554	0.037
Sand	-0.326	-0.521	0.348	-0.643
Latitude	-0.310	-0.649	-0.460	0.116
Predators	-0.147	-0.348	-0.634	0.047
Competitors	0.281	-0.406	-0.307	-0.504

Although omitted from the CCA plot displayed here, species can also be displayed. This can be on the same plot alongside the sample stations, or (for clarity) separately. The species displayed can be selected to those showing the best relationships with the environmental factors. Likewise, the species-environment

relationships could be investigated further through partial CCA. Oug (1998) demonstrated this in a study of the benthic macrofauna near Tromsø, Norway.

Determine major characteristics

Correlating patterns in the biological and physical/environmental data will produce a number of different groups of clustered samples which are likely to be representative of different habitats. Within each group, the samples will have similar properties, and the task now is to determine which characteristics can be used to define the habitat and its relevant scale.

Biological characteristics

Ranked abundance lists are useful in giving a simple overview of the relative dominance of the various taxa within each group, but they would not usually include colonial species or plant species as these are rarely given an abundance value). It will be important to include such species if these are characteristic of the habitat, particularly if they provide some degree of structure (e.g. erect forms such as seaweeds, sponges etc) that may be utilised by other species.

The real issue is to identify those species that have been instrumental in determining the observed biological patterns (clustering). A list of characterising taxa would include those that contribute most to the similarity within a cluster, while discriminating taxa are those that contribute most to dissimilarity between clusters. The PRIMER package has a single routine (SIMPER: 'similarity percentages') giving both outputs, as per the following example.

An example table of the results of SIMPER analysis for beam trawl samples from a study site in the Eastern Channel, where two major clusters were identified. This first part of the output lists the top ten taxa in each cluster, giving their average abundance and contribution (% and cumulative %) to the average similarity of samples in the clusters.

Taxon	Average Abundance	% Contribution	Cumulative %
Cluster 1			
<i>Hinia sp.</i>	1071.00	14.57	14.57
<i>Pagurus bernhardus</i>	256.25	11.11	25.68
<i>Ophiura ophiura</i>	214.25	7.63	33.31
<i>Ophiura albida</i>	3137.50	5.97	39.28
<i>Callionymus sp.</i>	9.00	4.86	44.13
<i>Macropodia sp.</i>	16.75	4.63	48.76
<i>Pagurus prideauxi</i>	5.50	4.30	53.06
<i>Aphrodita aculeata</i>	8.50	4.13	57.19
<i>Corbula gibba</i>	71.25	4.10	61.29
<i>Gobiidae</i>	25.25	3.96	65.24

Taxon	Average Abundance	% Contribution	Cumulative %
Cluster 2			
<i>Pagrus bernhardus</i>	92.54	12.50	12.50
<i>Anapagrus laevis</i>	31.85	8.90	21.40
<i>Ophiura albida</i>	123.92	7.74	29.14
<i>Hinia sp.</i>	32.00	6.77	35.91
<i>Callionymus sp.</i>	5.92	5.47	41.38
<i>Macropodia sp.</i>	10.69	4.76	46.14
<i>Aequipecten opercularis</i>	7.38	4.60	50.74
<i>Pagrus prideauxi</i>	6.69	4.49	55.23
<i>Liocarcinus holsatus</i>	3.92	4.25	59.48
<i>Gobiidae</i>	4.00	3.99	63.47

An example table of the results of SIMPER analysis for beam trawl samples from a study site in the Eastern Channel, where two major clusters were identified. This second part of the output lists taxa contributing to the first 40% of the dissimilarity between clusters, giving their average abundance in each cluster and their contribution (% and cumulative %) to that dissimilarity.

Taxon	Average Abundance		%	Cumulative
	Cluster 1	Cluster 2	Contribution	%
<i>Aphrodita aculeata</i>	8.50	0.31	4.19	4.19
<i>Phaxas pellucidus</i>	4.00	0.00	3.61	7.80
<i>Acanthocardia sp.</i>	7.00	0.00	3.61	11.41
<i>Limanda limanda</i>	7.50	0.08	3.53	14.94
<i>Abra sp.</i>	113.25	0.08	3.41	18.35
<i>Echinocardium cordatum</i>	9.75	0.15	3.22	21.57
<i>Liocarcinus pusillus</i>	0.75	2.08	3.10	24.68
<i>Buglossidium luteum</i>	5.75	0.92	2.78	27.46
<i>Echiichthys vipera</i>	0.25	1.15	2.72	30.18
<i>Ophiura ophiura</i>	214.25	2.31	2.68	32.86
<i>Aequipecten opercularis</i>	1.50	7.38	2.53	35.39
<i>Crangon allmanni</i>	0.50	3.46	2.51	37.91
<i>Psammechinus miliaris</i>	0.00	5.15	2.45	40.36

Consideration also needs to be given to the fidelity and specificity of species, high fidelity meaning a species is *always* present in a particular habitat and high

specificity meaning it is *only* found there. Clearly these are important attributes of characterising taxa. In the example above several taxa can be seen to exist in one cluster, but not the other (e.g. *Phaxus pellucidus*, *Psammechinus miliaris*).

Finally, it will be important to record the full species list or, at least, the most frequently occurring species for the habitat (with relative abundance/biomass data, pooled for all samples in the cluster). Such records will be needed to describe the habitat class being mapped and when erecting a new habitat class within an existing classification system, and to help define the '*standard*' for the new habitat within the Habitat Matching Program.

TWINSpan provides a two-way indicator species analysis for classifying species and samples, producing an ordered two-way table of their occurrence (Hill, 1979a, b). Again, the output needs an 'ecological' interpretation. The classification process is hierarchical. Samples are successively divided into categories, and species are then divided into categories on the basis of the sample classification. This technique has been used extensively in botanical studies, but there are relatively few marine benthic applications (e.g. Heip and Craeymeersch, 1995; Hensley, 1996; Rendall, 1997). Mettam *et al.* (1994) employed both cluster analysis and TWINSpan in their study of the benthic macrofauna of the Severn Estuary. Ten faunal groups were identified using cluster analysis and nine were delimited using TWINSpan. After considering both sets of results together, the authors recognised eight faunal groups — each are occupying different sediment types and depths.

Physical characteristics

The results of BIO-ENV (PRIMER) or CANOCO analyses will have indicated the physical variables that are most influential in determining the biological patterns. Of principal interest among these will be the substratum type, which should be described as fully as possible. Other physical variables should be included in the description, such as salinity, depth and temperature ranges, as should details of hydrodynamic and environmental/ecological descriptors such as wave exposure, current speed, biological zone etc.

Match classes with classification system

The final stage of the process is to see if any of the habitat descriptions derived from the analysis and interpretation of ground-truth data match, fully or partially, any of the habitats identified in the existing classification system. Although many studies to date have not done this and rely solely on defining the habitat classes within the study rather than trying to relate these to an external '*standard*' scheme, this practice is not of great benefit. Comparison with external schemes and preparation of maps which have been correlated fully or partially with such schemes is important in helping to standardise the maps and make the data interoperable (see *How can I use my map?*).

If a satisfactory match is not found, the information should be used to erect a new habitat class, to enhance the utility of the classification system. There is a procedure by which new submissions can be made to the EUNIS system using the following documents that can be found in the resources folder.

'EUNIS marine habitat classification: Application, testing and improvement' provides guidance on defining new EUNIS habitat types ([EUNIS application v3.doc](#)).

'EUNIS marine proposal pro-forma' is an Excel spreadsheet providing a worked example for a new submission ([EUNIS marine proposal proforma v3.xls](#)).

Proposed new EUNIS habitat types should be entered on the pro-forma and sent to the JNCC (mailto: david.connor@jncc.gov.uk)

Resources:

[EUNIS application v3.doc](#)

[EUNIS marine proposal proforma v3.xls](#)

Amalgamating data

Much of the analysis explained above presupposes that the data have been obtained using standardised techniques either from a single survey (or series of coordinated surveys), or from different surveys that are similar enough for the data to be combined into a single data set. However, where there are substantial differences in techniques, there will clearly be an effect on the composition of the combined dataset.

It is a common experience that datasets reflect the skills and biases of the recorders, even when the techniques are nominally equivalent between surveys. It may require the editing out of dubious identifications from one or more datasets, or the use of higher taxonomic units than those supplied.

Transformation of counts or even reduction to SACFOR or presence/absence may be needed to find a common denominator between datasets, especially where it is suspected that counts show a bias. It may be important to remove rarities from the datasets so that the transformed data reflect the species occurrences that can be regarded with confidence.

Data might be in the form of habitat classes and these may need to be translated into the latest classification nomenclature. Where there is no clear translation, the habitats may need to be amalgamated using a higher common level in the classification system. This may lead to a loss of information and an alternative is to use a life form or biotope complex that captures the essence of the biota of the habitats.

Optimise the physical coverages

In a second step, the most appropriate physical coverages will need to be selected and this may require some preparation from the original data (e.g. if new remote survey data have been produced).

At this stage, it is assumed that data layers from single surveys (or series of coordinated surveys), such as acoustic maps, have been prepared to an acceptable standard (see *How do I collect my data?*) and the emphasis is on preparing these and other sources of input data for integration with the ground-truth data to create biologically relevant habitat maps. 'Rough and ready' to more sophisticated approaches can be used; the latter extensively investigating the role of many variables. The experience gained from the ground-truth data analysis will drive the selection of the main variables.

Whether or not a particular location on the sea floor is **suitable** for a particular species (and by extension, assemblages of species characteristic of habitat types) is largely determined by the physical and environmental factors. **However, the presence of favourable habitat factors does not mean that a particular species will be present since many other factors may result in the species being absent (such as competition, recruitment history, predation, anthropogenic activities and so on).**

A few physical and environmental factors are universally important for determining the distribution of all marine species and habitat types. Three are considered to be fundamental in the EUNIS scheme and form the basic input data for the predictive mapping of the EUNIS Level 3/4 map of the whole MESH area (the so-called 'Triplet Map'). A general discussion on the kind of data needed is provided in *What do I want to map?*; here the importance of the necessary physical coverages is stressed.

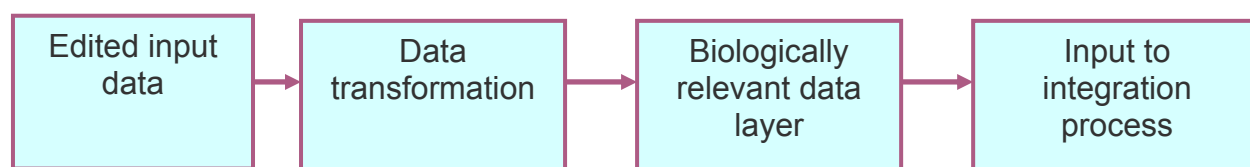
- **Substratum** (e.g. rock, sand, mud, boulders), to be mapped as classes along a scale based on the grain size (e.g. the Wentworth scale) or on the relative proportions of silt, sand and gravel (the Folk triangle). However, the critical substratum characteristic will vary considerably between species and habitats and specific continuous variables, such as percentage silt or median grain size may be more biologically meaningful and are thus more suitable for modelling purposes. Some of these variables may be derived from remote sensing data through expert interpretation or through automated classifications;
- **Bathymetry**: many important environmental constraints on species distribution are derived from depth data. Light penetration is often quoted in terms of depth, but this depends on local and regional turbidity; and
- **Water energy**: energy on the sea floor can be measured; however, it is often modelled from other factors and there are many ways of expressing and calculating water energy.

One should realise that there are other variables that may prove important for constraining the distribution of particular species and habitats such as water temperature, salinity, sediment transport and bedform features. In addition, surveys may result in the measurement of properties that are not immediately interpretable in

terms of any biologically relevant factors, such as acoustic reflectance and backscatter.

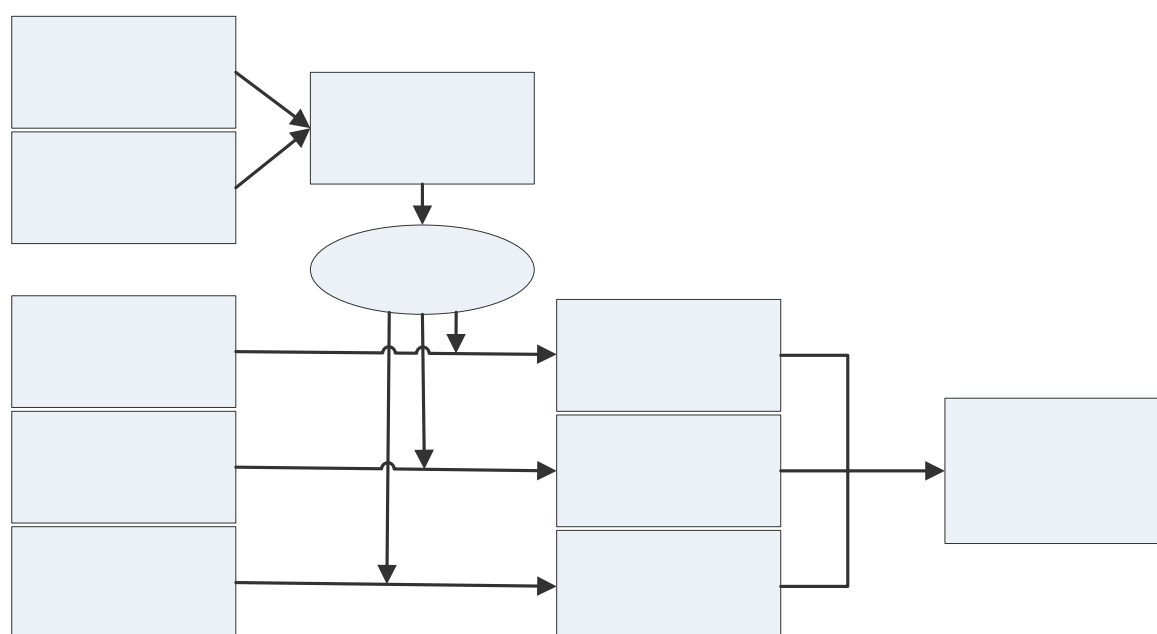
Which variables are needed for habitat mapping depends upon the area and the habitats to be mapped and its purpose: large areas embracing a wide range of habitat types may require a wide range of variables, whilst variation in light penetration and salinity may be considered to be constant within a small area and thus not needed.

As already indicated, the selected variables and coverages might need a **transformation** before integration with the ground-truth data. A range of transformation techniques can be applied to the input data in order to derive the required coverages for integration with the habitat records.



The transformation process can be quite simple, such as changing formats (e.g. vector data may need to be transformed into a raster format or vice versa), or require more complicated processes such as raster calculations and classification (e.g. slope from bathymetry), reclassification or even sophisticated modelling (e.g. advanced geostatistical techniques to optimally interpolate sediment distributions).

Finally, the obtained coverages need to be **combined** following a strategy that is most meaningful towards its biological relevance.



Example of a scheme that uses rules for habitat suitability to transform/reclassify original datasets into biologically relevant zones that are combined into habitat suitability classes.

Links to other sections:[Where do coverages come from?](#)[Transforming data](#)[Combining data](#)**Links to resources:**[EUNIS application v3.doc](#)[EUNIS marine proposal proforma v3.xls](#)**Links to websites:**<http://www.primer-e.com><http://www.kovcomp.com><http://www.clustan.com>**Where do coverages come from?**

As a reminder, coverages can be remotely-sensed data (acoustic, satellite images, aerial photography, LIDAR etc), indirectly linked through an interpretation of the original coverage data (as intermediate, proxy maps of some physical factor), or created using deterministic models (sediments, exposure, sea floor currents etc.).

A list of parameters that influence the occurrence and distribution of seabed habitats is given in the table. A more extensive discussion on the datasets can be found in Connor, 2007.

An example table of datasets that can be used for habitat mapping (ICES, 2006)

Variable	Unit	Application
Shoreline/coastline (HAT - Highest Astronomical Tide)	m	Defines the land/sea boundary Complexity of coast/aspect/ruggedness
Bathymetry (including topography)	m/gradient	Topography, 3D modelling, slope, ruggedness, bedforms. Relation to biological zonation
Surficial substrata	Top 1 meter of sediment. Sediment structure (phi, mm) Lithology. Redox discontinuity (mv)	Identification of seabed sediments, potential habitats and range of biological communities. Contaminant sinks/anoxic zones
Maximum wave base	Meters below sea surface (mean value – at least over 1 year, preferably over last 10 years)	To assess the degree of seabed disturbance which may affect biological communities
Wave exposure/fetch	Exposure coefficient/shear Orbital velocity (e.g. for relevant storm conditions);	Identification of potential habitats, range of organisms, seabed disturbance.

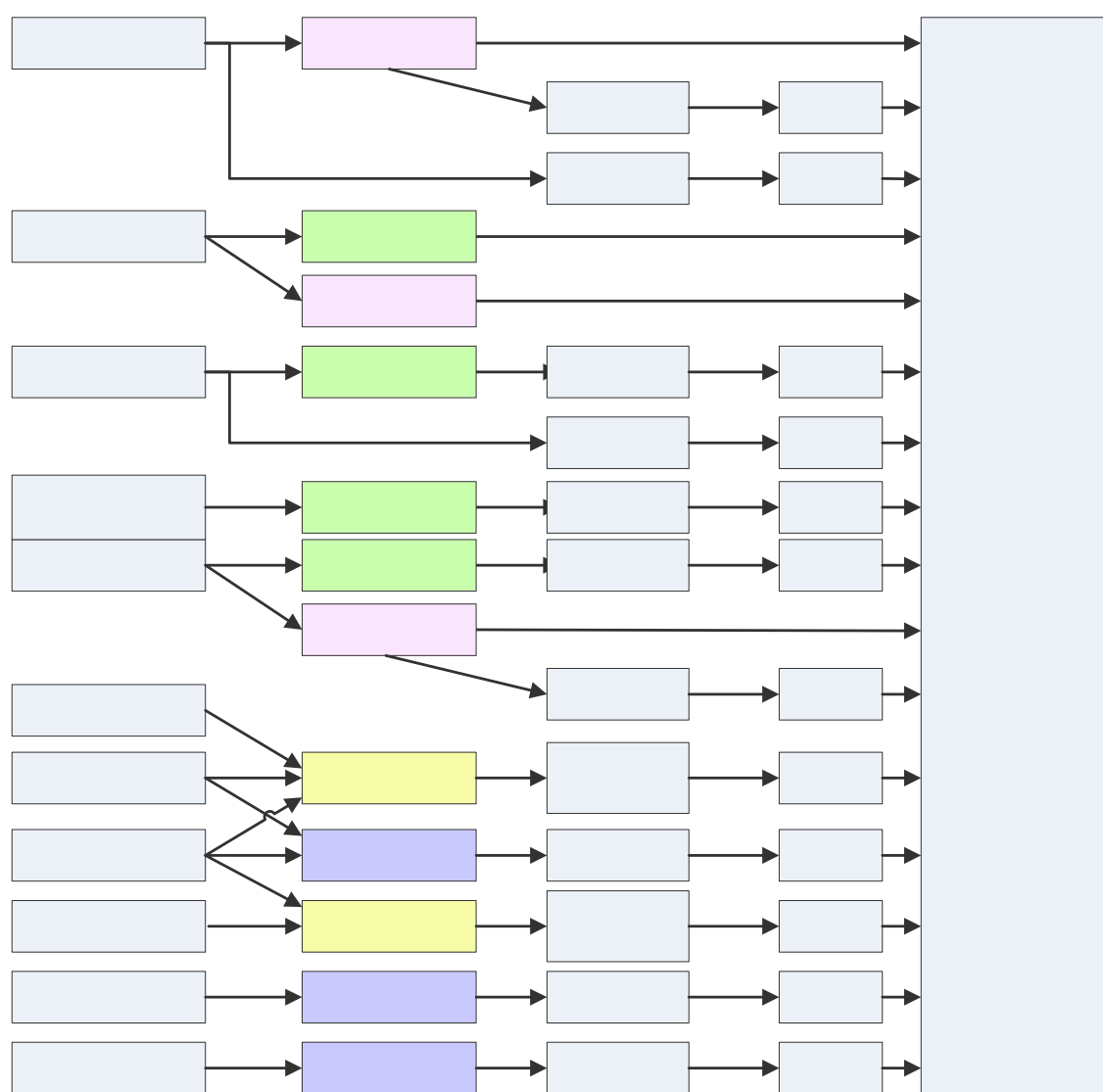
Variable	Unit	Application
	may be dependant of life span of the relevant organisms	
Temperature (surface, bottom, profile)	°C	Biogeographic zones
Thermocline	Annual average	Special communities Stratification
Salinity (surface, bottom, profile)	‰	Potential habitats, range of organisms
Halocline		
Current speed (residual/max near-bed stress)	U .cms-1 (see below)	Identification of potential habitats, sediment distribution
Stratification	Potential energy anomaly Jm3	Water column stability
Mixing	Spatial and temporal extent (see below)	Retention of juvenile Development of anoxia
Tidal range/sea level changes	cm/meter	Identification of potential habitats, zonation, exposure time
Transparency/light attenuation	Secchi depth (m)	Depth of photic zone
Turbidity	FTU (turbidity)	Potential habitats (macroalgae/maerl etc)
Ice Cover, (seasonal surface cover not anchor ice)	Number of days with ice coverage and area covered.	Range of sessile organisms
	Thickness (m)	Tendency for anoxia in shallow basins
Ph sediment/water column		acidification
Dissolved gases Oxygen/methane	.mg/l percent saturation	Anoxic area or time period of oxygen sag Special communities
Water quality nutrients	DIN/DIP/silicate uml-1	Anthropogenic enrichment
Anthropogenic activities	Multiple	Habitat modifiers
Occurrence frequency/ intensity of algal blooms	Chlorophyll a Presence of toxic species	Eutrophication/may lead to anoxia/ toxic species may selectively impact species
Benthic species	Benthic community metrics(abundance/diversity etc)	Range of organisms, benthic diversity. May form habitat (biogenic reefs or modify habitat). Not needed for prediction but for validation
Pelagic species	Pelagic community	Range of organisms pelagic diversity - for model validation

Some of the parameters are of lesser relevance and may be used on a case-by-case basis. The primary variables remain sediment type, bathymetry and water energy; secondary variables mostly result from a more or less complex computation or association of the former.

Depending on the environment, different variables become more important in determining the habitat. In particular, rocky substrata differ markedly from sandy substrata and their characterisation requires a different variable selection. Variables explaining rocky substrata are very important in France, the UK and Ireland, while the sandy substratum is predominant in Belgium and in the Netherlands

At the risk of jumping ahead to the next Section on integration of remotely-sensed data and ground-truth records, the use of these coverages for the derivation of a habitat map can be through direct interpretation, statistical analysis and habitat suitability modelling. For completeness, these links are shown in the following diagram but are described more fully in [Optimise the map making](#).

The flow diagram illustrates many of the sources of coverages that could be used for habitat mapping. This is not intended to be a complete list of coverages, but does show the variety of possibilities that might be considered.



Ways of deriving habitat maps from original datasets

Transforming data

Coverages may not be available in a suitable format for modelling (e.g. vector data may need to be transformed into a raster format or vice versa). Thus, it is likely that

data will need to be **transformed** to create the input layers required for mapping the habitat distribution.

This transformation process can be quite simple, such as changing formats, or require more complicated processes such as reclassification and even sophisticated modelling e.g. to derive sediment distributions. The latter may require its own integration of sediment sample data with full coverage data. A range of transformation techniques can be applied to the input data in order to derive the required coverages for integration with the habitat records.

Common transformation processes are:

- **Converting data to a common raster format:** The raster format (see definition) is particularly suitable for many forms of mathematical manipulation and modelling with multiple layers.
- **Reclassification and cross tabulation of data:** One of the most common ways of manipulating raster data is to reclassify continuous variable data into discrete classes. Two or more reclassified datasets can be combined through cross tabulation (a process that compares two maps of classed data and returns a unique value for every combination of input values).

A	B	B		A	C	A		AA	BC	BA
C	C	A	+	A	C	A	=	CA	CC	AA
C	A	B		B	C	B		CB	AC	BB

Illustration of a cross tabulation of 2 raster datasets.

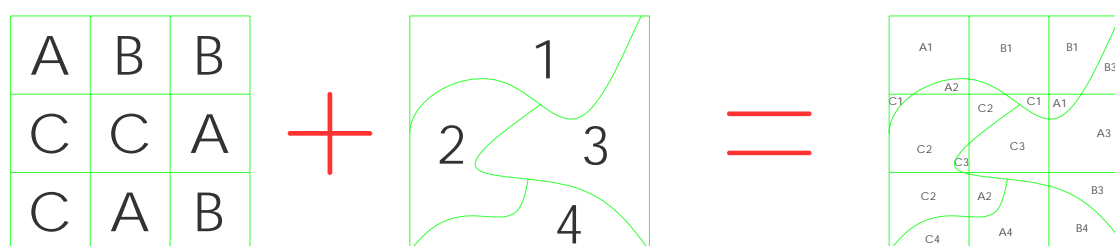


Illustration of a cross tabulation of a raster and a vector dataset

- **Interpolation of point data to form modelled full coverages:** Many data sets require interpolation to create a complete coverage since the raw data are in the form of points. For example, a hydrographic survey echo sounder will create a series of point data along the lines run by a vessel.

Also for sediment data, the interpolation method is crucial to obtain reliable coverages. Spatial statistics and optimal interpolation are discussed under [deriving optimal data coverages](#). 4.3.2.3

- **Applying moving-window algorithms:** Moving window processes involve passing a window (e.g. a square 9 pixels high and wide, although other shapes can be used) over the image pixel-by-pixel and returning a value for the central pixel according to the particular mathematical calculation of all the

pixel-values covered by the window. The simplest algorithms return statistics such as the average or mode value/class and are used to simplify images. More complex algorithms return values for variance and heterogeneity and (from bathymetric data) slope or aspect. These derived layers can often reveal much useful information from the base layer and are the starting point for detection of sea bed features and landscape evaluation.

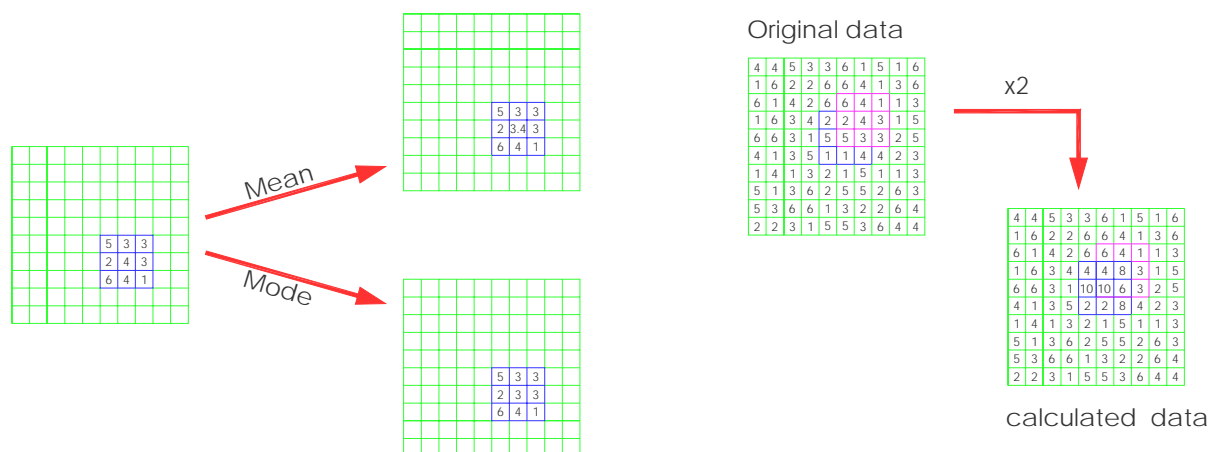


Illustration of a moving window algorithm.

- **Expert analysis 'by eye':** Used systematically by experienced workers, 'by eye' examination of an image can result in a very sophisticated interpretation of a complex image.
- **Unsupervised classification and automated pattern recognition:** Data may be subjected to computer-aided techniques for the detection of 'natural' groupings of values within data. These are unsupervised if the software are permitted to partition or segment the data sets without guidance from ground-truth records. These techniques are often used to guide sampling and also to match ground-truth records to 'natural' divisions within remotely-sensed data coverages.
- **Supervised classification:** This is discussed more fully in 4.4 as a major tool for integrating ground-truth samples and remotely-sensed images. However, it is included here since the same tool can be used to produce coverages of sediment type as an input into habitat suitability modelling (HSM). Signatures are created from the ground-truth data which are then applied back to the coverages to predict sediment class.

Links to other sections

[Expert-eye analysis of remote-sensing data](#)

[Transformation of acoustic backscatter images](#)

[Deriving optimal data coverages](#)

[Optimise the map making](#)

Expert-eye analysis of remote-sensing data

When remote sensing has been used to draw up a draft physical map of an area, the interpretation of that map in terms of the substratum classification and bedforms should undergo some form of validation procedure using observational and/or sampling techniques. Failure to do so will reduce the confidence in the final mapped output.

Ideally, each of the mapped polygons should have been sampled during the ground-truth survey, but this is often impractical. It is more likely that each substratum class has been sampled using a variety of techniques. The data and/or information from these samples should now be used to validate the draft physical map. If the provisional classifications prove to be incorrect, they should be modified accordingly.

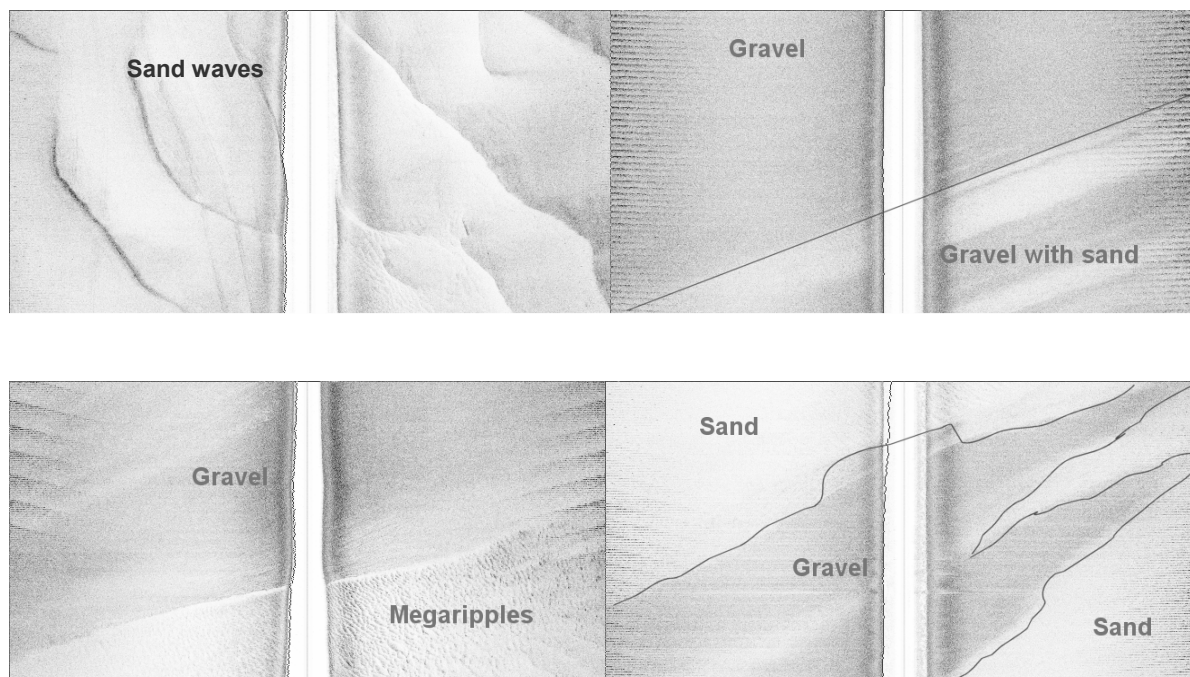
Where soft sediments predominate, a full geophysical analysis of samples will provide the highest confidence in validating the sediment classes. This may involve a particle size analysis and/or testing the geotechnical properties of the sediments (penetrometry, shear strength etc). If this is not available, moderate confidence can be gained from sediment descriptions resulting from direct inspection of samples, such as in a walk-over survey of a beach or examining sediments from grabs or core samples. Failing this, the validation may have to rely on the results of remote observation, where video and/or photographs have been used to image the seabed.

Where hard substrata predominate (i.e. rock outcrops, boulders, cobbles), validation relies heavily on observational techniques. In-situ, direct human observation provides the greatest confidence, using results from walk over surveys of the shore or diver surveys in shallow waters. For deeper waters, observation may be restricted to the use of video or photographic techniques, and confidence can be less on account of their limited field of view. If observational techniques have not been used, some degree of validation can be gained from samples collected by rock dredges or very large grabs.

The draft physical maps may include some interpretation of bedforms, such as 'sand waves' or 'mega-ripples'. In many cases, the remote survey techniques are the best method of identifying these bedforms, and opportunities for validation are extremely limited. Here, cross-validation may be appropriate, comparing the outputs of two remote sensing techniques to see if they both detect the same bedform (e.g. mega-ripples detected by both multibeam and side-scan sonar). Sampling techniques can rarely be used to validate bedform interpretations, as the scale at which they sample is usually far smaller than the scale at which bedforms exist (grabs may only sample an area of 0.1 m², but sandwaves may have a wavelength of several metres). Observational techniques (video and diver surveys) are probably the only affective ground-truth method for verifying sub-tidal bedforms, though they may only be applicable to bedforms on the metre rather than decimetre scale (due to restricted visibility).

The validation process should also consider verifying the boundaries between different sediment types and/or bedforms detected by remote survey techniques. The placement of borders may be the most critical part of a habitat map, so their location and characteristics are important to the overall confidence of the map. Are the borders in the right place, and do they represent distinct or gradual changes in sediment type? Here again, cross validation between the outputs of two or more remote survey techniques can be informative. Direct observation or video transects

can also be used. In shore surveys, a portable GPS system can be used to track a detailed outline of habitat boundaries.



Examples of interpretation of sediments, bedforms and borders from side-scan sonar surveys. The interpretations should be validated using data from ground-truth surveys

Transformation of acoustic backscatter images

Transforming acoustic backscatter images into usable maps for habitat mapping purposes can range from a simple to a quite complex process. Mostly, images are interpreted in terms of their reflectivity, texture and patterns, which can be done in a more or less unbiased way.

However, one should be aware that the backscattering is in decreasing order determined by (Blondel & Murton 1997):

1. the geometry of the sensor-target system (angle of incidence of each beam, local slope, etc.);
2. the physical characteristics of the surface (micro-scale roughness, etc.); and,
3. the intrinsic nature of the surface (composition, density, relative importance of volume versus surface diffusion/scattering for the selected frequency).

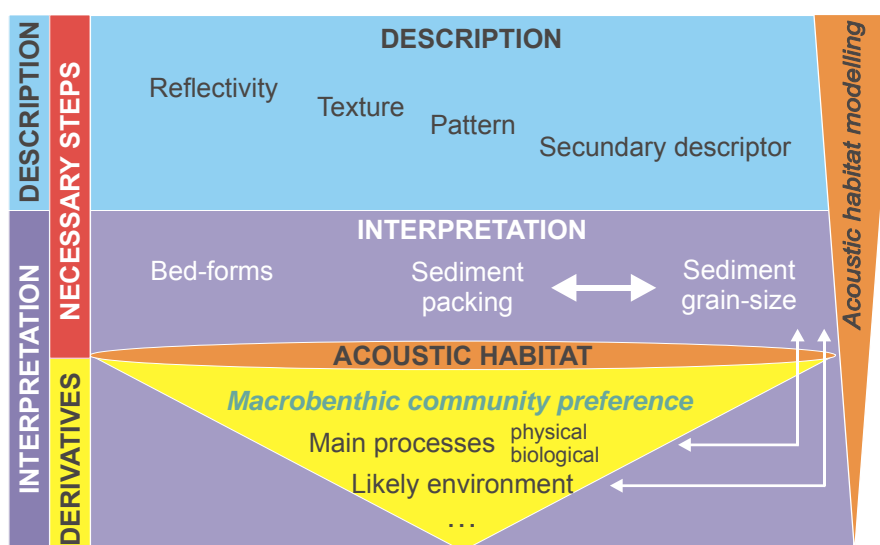
Moreover, the backscatter quality will vary with different acoustic frequencies and their resolution. Imagery may vary from place to place and with bio-geographic conditions

Interpretation in terms of sediment nature and distribution remains difficult and requires good quality imagery. The most important parameters controlling the acoustic response of marine sediments, ranked in order of importance are:

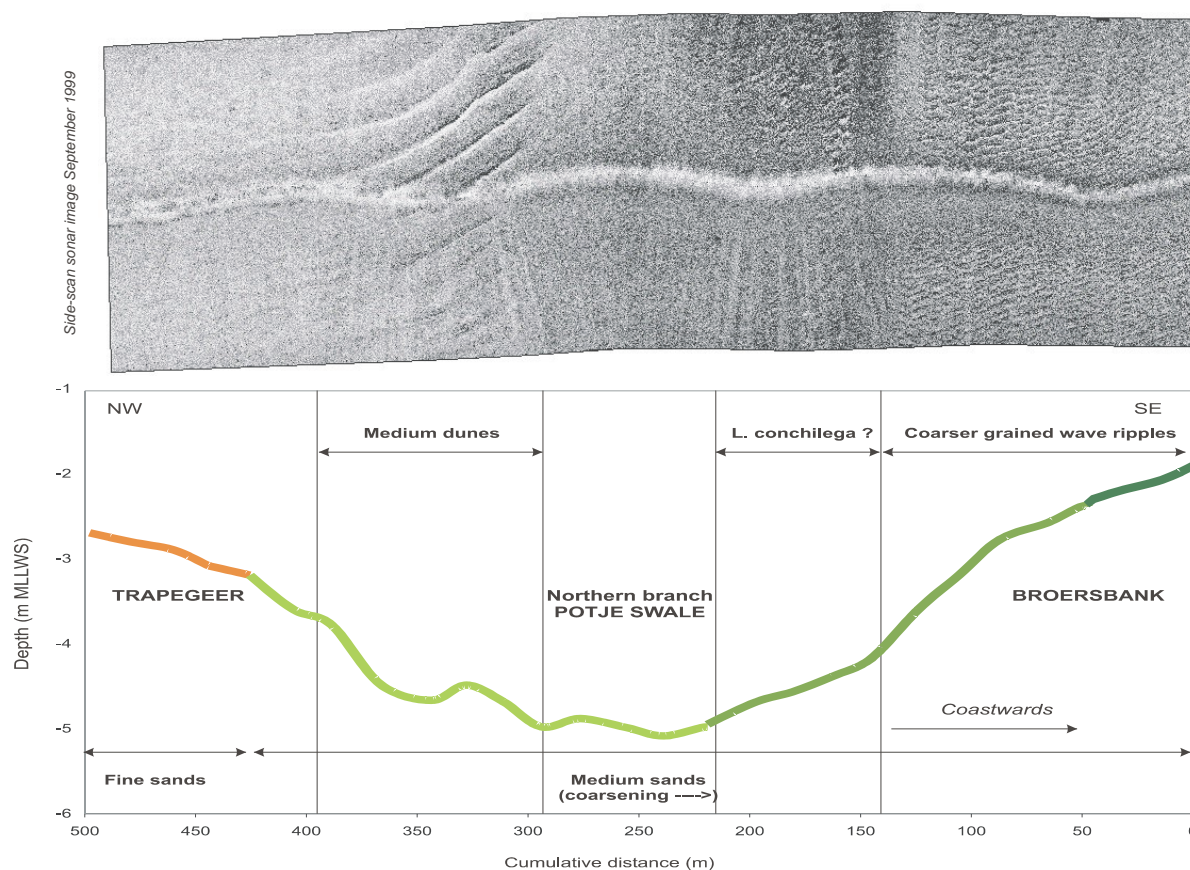
1. porosity;
2. density;
3. the overburden stress;

4. the degree and type of lithification; and,
5. the grain size and distribution.

Most probable, one can distinguish between soft and hard grounds, but going beyond that might require extensive ground-truthing. When the occurrence of macrobenthic species or communities correlates highly with sediment nature, it might be attempted to interpret the imagery in terms of sediment nature first and then to link up with the biology (*indirect interpretation*). In this case, it will be most crucial to estimate variations in porosity/density first. As such, highly reworked areas, such as sandbank tops, might have a low backscatter, although the grain size is often coarser than its surrounding environments. Additionally, sediments with even minor silt-clay enrichment are more compacted and the area will have a higher backscatter. This is illustrated in the following figure.



Interpretation scheme for the interpretation of soft substrata habitats (Van Lancker et al., 2001).



Side-scan sonar image and associated bathymetry of a shallow sandbank-swale system near the Belgian coast. A relatively low acoustic backscatter is seen along the Trapegeer sandbank corresponding with medium-grained sands. In the swale, the high reflectivity facies corresponds with fine sands with a silt-clay component. Along the foot of the Broersbank the patchy facies correlates with the occurrence of the tube building polychaete *Lanice conchilega*. The high reflectivity facies of the Broersbank is due to the presence of coarse-grained wave ripples.

Nowadays multibeam datasets allow for a full-coverage predictive modelling of sediment nature based on variations in acoustic backscatter. As an example Ferrini and Flood (2006) investigated the relationship between backscatter intensity (using a 300kHz multibeam system), grain size distribution and seabed roughness of a sandy site using principal component analysis and multiple regression analysis. It was possible to draw a correlation between roughness, grain size and backscatter intensity, but the relative importance of the factors varied, even though the survey was restricted to sandy sites. The main variables contributing to backscatter variation were median grain size, standard deviation of grain size and range of heights, derived from a small-scale roughness map. The latter is crucial and requires high standards of the multibeam survey and the ground-truthing. For more gravely sites, the percentage gravel and also the percentage shells will highly influence the backscatter intensities.

An overview of the strategies and processes involved in acoustic seabed mapping can be found in the worked example on the mapping of [shallow coastal benthic habitats](#).

Deriving optimal data coverages

Interpolation of point data

When quantitative data is to be used, the density of the sampling points should be sufficient to obtain an adequate data coverage. Various interpolation techniques exist and should be intercompared. A good data exploration with a histogram, a frequency distribution, and some descriptive statistics is necessary to get insight into the quality of the data and to become familiar with the data.

Geostatistics, generally known as Kriging techniques are interpolation techniques that take into account 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).

Where the density of sampling points is low, geostatistics can significantly improve the interpolation. Validation with independent data shows where extra sampling is necessary and where the quality of the map is less good.

In some cases, where the sampling data are correlated with other coverage data, it is possible to use multivariate geostatistics. An example is the correlation between grain size and bathymetry. In soft substrata and higher dynamic areas, the bathymetry will influence the sediment distribution and especially in the case of sandbanks and swales, the sediment distribution differs considerably according to the morphological entity. Verfaillie *et al.*, (2006) demonstrated the strength of this technique and applied it to obtain a full coverage map of the median grain size of the sand fraction over the entire Belgian part of the North Sea ([UGent Multivariate geostatistics.pdf](#)). The application has been extended to produce maps for the southern North Sea, covering the Belgian Continental Shelf, the southern part of the Dutch Continental Shelf and a small part of the south-eastern UK waters. The same technique has been applied to interpolate individual sediment fractions over the entire Dutch Continental Shelf ([Dummy file WE TNO interpolation fractions.doc](#))

Deriving biologically relevant bathymetry related parameters

A good bathymetry dataset is crucial in any habitat mapping study and should be invested in. Most often, the depth itself will not be the determining parameter, but derived parameters such as slope, aspect, topographic features are often more important in explaining habitat variability. When bathymetric point data is interpolated into digital elevation models (DEM), a suite of mathematical derivatives can be calculated at any location.

Slope and Aspect

Slope and aspect are two first order derivatives of the bathymetric surface. Slope is defined by a plane tangent to the surface as modelled by the DEM at any given point and comprises two components namely gradient, the maximum rate of change of altitude and aspect, the compass direction of this maximum rate of change (definitions by Evans, 1980 and Burrough *et al.*, 1999). Mainly slope is used in the context of habitat mapping e.g. higher amount of species associated with slopes, potentially because of increased food availability.

Bathymetric position index

The [Bathymetric Position Index](#) (BPI) is a second order derivative of the surface. It is a measure that allows calculating where a certain location with a defined elevation is relative to the overall landscape. The result is a map with geomorphological features like slopes, depressions, crest lines and flat areas.

In addition, sandwave heights may be estimated ([Dummy file WE TNO Sand wave height map DCS.doc](#)) in an automated way. The method works well for individual sandwave areas, but becomes inaccurate where sandwaves are superimposed on sandbanks.

An alternative for the BPI is a topographic feature recognition from bathymetric images.

Moreover, in GIS the following terrain based tools may help in getting more out of your data: terrain shading; hill shading (sun angle shading); contour lines; height above threshold (HAT) and combinations of all of the above.

All of these techniques can be applied on coverage depth data of acoustic or optical remote sensing techniques. Applications using hydrographic LIDAR (Light Detection And Ranging) data can also be found ([WE Ifremer lidar.pdf](#)).

Combining data

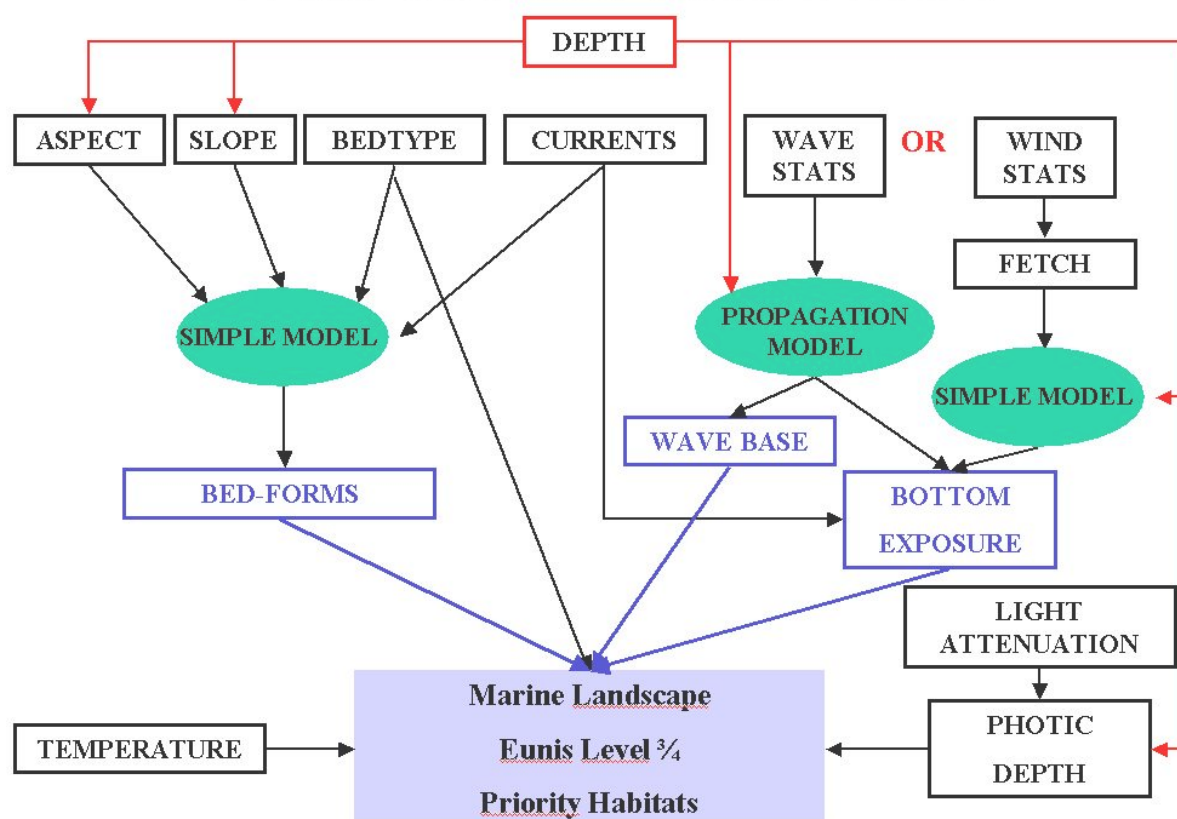
One of the most inspiring approaches to combine data coverages for habitat mapping is described by Roff and Taylor (2000). The latter proposed a marine landscape approach, which enables the mapping of habitats based on geophysical features alone, but in the view that these are important in determining the nature of biological communities. In this approach, the biology is only used passively to verify the final results. The concept also anticipates on the growing realization that conservation at the scale of spaces or landscapes might be more reasonable than conserving individual species. The spaces concept requires a top-down manner of working (Laffoley *et al.*, 2000), which is exactly what is proposed in the paper of and Taylor (2000).

The data combination can be done through simple mapping and GIS techniques, preferably based on the transformed data, described previously. The final result may however largely differ dependent on:

1. the quality of the underlying data sets;
2. the completeness of the data sets;
3. the data layers that have been used;
4. the resolution of the data layers;
5. usage of vector polygon data, raster data or vector grids;
6. the break values used for the classification of each data layer;
7. usage of supervised or unsupervised classifications;
8. the types of queries performed (different combinations of data sets, changes in practical criteria such as varying class breaks...).

Also, the extent to which biological data is used in the process might influence largely the final outcome.

The diagram is an attempt to show how the main parameters can be combined to produce meaningful results for habitat mapping. The top level shows primary variables, which combine in a way to produce secondary ones. The red colour enhances the role played by depth in many ways. Depth can interfere indirectly in computations such as wave propagation, current computation and also to determine the near-bed stress (or bottom exposure) as a result of surface exposure. It is otherwise mostly expressed by its derivatives (slope, aspect, topography). In the marine landscape modelling procedure, depth can also be combined with wave penetration and light penetration to produce biological “depth zones”, namely the infra and circa-littoral zones. Note that the diagram is only an example of data layers that can be combined. The combinations will differ according to the purpose, the geographical area and the kind of substratum.



Ways of combining environmental data for habitat modelling. Here an example relevant for the modelling of rocky substrata. Depending on the resolution of the data layers, the final product may be a 'Marine Landscape', a EUNIS level 3 or 4, or focussed on a priority habitat.

Links to other sections

[Marine Landscape mapping approach](#)

Marine landscape mapping approach

The methods to come to a marine landscape map are well defined (Roff & Taylor, 2000; Golding *et al.*, 2004; Connor *et al.*, 2006). The techniques are mostly GIS based and consist of: defining a series of environmental data layers characterizing the seabed; processing the data layers for further analysis; identifying meaningful thresholds by means of classification; production of the marine landscape units by

means of summarizing and querying the different datasets; and, ecologically validating the results.

1. The environmental parameters should function as indicators for habitat types across ranges of scales and are preferably stable in nature. There are many abiotic features that can be included and which are relevant towards habitats, but not all of them have the same importance. The selected data layers may vary along different regions as some datasets might show too little variation or prevent discriminating among habitat types.
2. In processing the selected data, it must be decided which data type will be used throughout the process. This may be vector, raster or vector grid format, and dependant on the user's choice.
3. Some datasets will be available as unclassified continuous datasets, whilst others might already be in a classified form. The ways in which they are classified have a strong influence on the final classification/map. However, a huge amount of classes might reflect the complexity of the environment, usage of too many classes, within a classification, complicates the further handling of the data. Moreover, it will not necessarily lead to a map with well-defined marine landscape types. A small number of relevant classes have the greatest possibility to lead to ecological meaningful habitat types. So far, no commonly accepted classification technique exists and thus, a lot depends on the expert knowledge of the involved scientists.
4. Once each data set is classified and annotated with their necessary metadata and attributes, they are combined. The technique applied depends on the chosen data format. The query process can be performed in GIS or within MS Access. From the resulting layer or from the composite database, practical criteria can be developed to assist in the demarcation of marine landscapes into distinct types. The key criteria for the queries are dependant of the input data sets (e.g. bathymetry, slope, median grain size, photic depth...). Analysis of the combined data set/table can lead to many different combinations for each location, eventually identifying different marine landscape types.
5. As a final step, the marine landscapes should be evaluated against their ecological value. This can be a simple cross tabulation of the landscapes against biological databases. However, some habitats may be well under- or oversampled and an unbiased evaluation may be hard to procure. Ideally, the results are verified against detailed survey data.

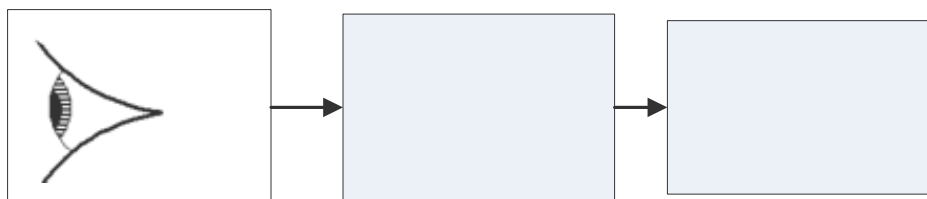
A detailed description of the processes involved can be found in Connor *et al.*, (2006).

Optimise the map making

The third step focuses on data integration and modelling and may require a range of strategies to integrate the ground-truth records with the physical coverages. The aim is to prepare a coverage of the distribution of habitat classes. As with the other stages, the methods for integrating the biological ground-truth data with the physical coverages range from the simple to the sophisticated ones:

- Biological habitat classes can be overlain on physical coverages and their **match** verified. Similar matches can then be grouped together. Cross-tabulation is the most common technique to be used. A number of problems can arise, mostly related to the distribution of sampling points which is most often largely biased to particular habitats.
- Images from side-scan sonar, swath backscatter, swath sun-illuminated bathymetric models, aerial photography and satellite can be interpreted in terms of the habitat, directly, using expert knowledge and judgement. On the shore, this might be a **direct interpretation** of aerial photos in the field; e.g. walking along the shoreline or using underwater cameras). The MESH Habitat Signature Catalogue gives an overview of known signatures of some habitats.

Using signatures may be relatively successful for small, simple areas, but as the survey area increases, confusion over the distribution of biota over similar patterns is more likely to occur. After integration with the ground-truthing, some images might need a **reclassification**



Scheme of a direct mapping approach.

- A suite of **statistical techniques** can be used to model habitats. Most of the techniques can be grouped into seven categories (Guisan and Zimmermann, 2000): multiple regression and its generalized forms, classification techniques, environmental envelopes, ordination techniques, bayesian approaches, neural networks and other approaches including mixed approaches. The extent and quality of the data will mainly drive the choice of the approach. Specialist software and advanced technical skills might be required.

Links to other sections

[Marine Landscape mapping approach](#)

[Matching biological habitat classes](#)

[Direct mapping](#)

[Statistical modelling](#)

[Examples of map making](#)

Links to resources:

[Shallow coastal benthic habitats](#)

[UGent Multivariate geostatistics.pdf](#)

[Dummy file WE TNO interpolation fractions.doc](#)

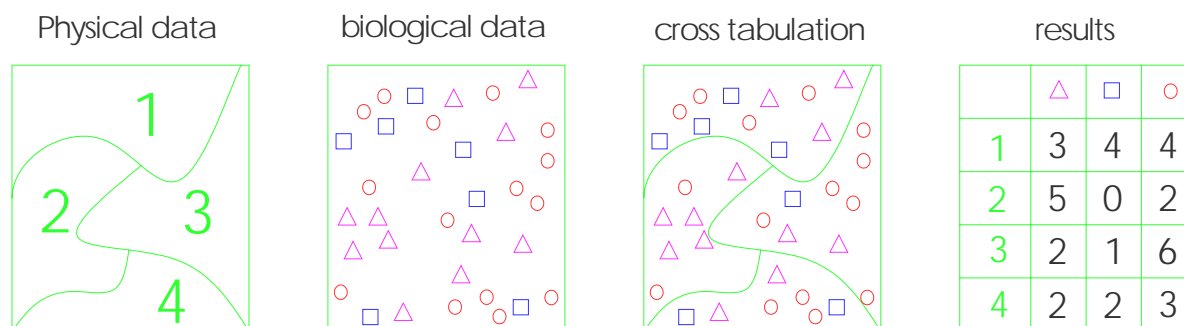
[Bathymetric Position Index](#)

[Dummy file WE TNO Sand wave height map DCS.doc](#)

[WE Ifremer lidar.pdf](#)

Matching biological habitat classes

Matching biological habitat classes to physical coverages resulting from cross-tabulation is simple in principle.



Cross tabulation of physical and biological data

Records are binned into categories according to where they occur on the cross tabulation map. The records are then ranked by frequency and the physical coverages category reclassified to the dominant biological habitat. However, there are a number of problems with this technique:

- Certain habitats are often sampled much more frequently than others (e.g. divers sample rocky habitats more than sand) and the frequencies might need to be corrected for this bias.
- As the number of classes of the physical coverages increases with cross tabulation, the number of records in each category falls – eventually to the point where the frequencies have little value. This is especially true for the more remote (deep, offshore) categories.
- The map will always show the dominant habitat and this means that rare but important habitats are never shown. This is especially the case with coarse resolution maps (e.g. 250 m pixel size) where the habitats are spatially smoothed.

For these reasons, the technique is best suited to very broad scale indicative maps where the 'rough and ready' approach is emphasised e.g. for the marine landscape maps.

Direct mapping

Direct interpretation of images / reclassification of images

When dealing with intertidal habitats, the interpreter can go to the field and obtain direct observation of ground facies against how they appear on the imagery. He then quite easily forges his own idea of a "signature". (Note that this use of the term 'signature' differs from that used in supervised classification. In the present context

signatures indicate a relationship between habitat characteristic as observed directly and patterns as observed in the coverage image. (For supervised classification, signatures have a more statistically rigorous meaning.) Generally such signature data are rather abundant as ground-truthing the intertidal zone is not so heavily constrained, and signatures can be refined to a high level of quality. In the case of subtidal habitats, direct observation is more difficult to achieve. Therefore getting reliable signatures may be more difficult in subtidal mapping.

Sidescan Images may be directly interpreted in terms of physical habitat features (sediment, bedform, topography) with a minimum of direct observation and then re-interpreted knowing which biological communities they support. For example, side-scan images may have been interpreted in terms of bedform and sediment types prior to integration and might need a reclassification. Ground-truth biological records can be overlain and matched by eye. It is then assumed that similar patterns support the same biota. If other layers, such as depth contours and slope, are also incorporated, then the substrata can be divided by depth zone etc and the extrapolation of the pattern/biota links will be further constrained by these additional factors, improving the final map. This may be relatively successful for small, simple areas but as the survey area increases, confusion over the distribution of biota over similar patterns in the backscatter is more likely to occur.

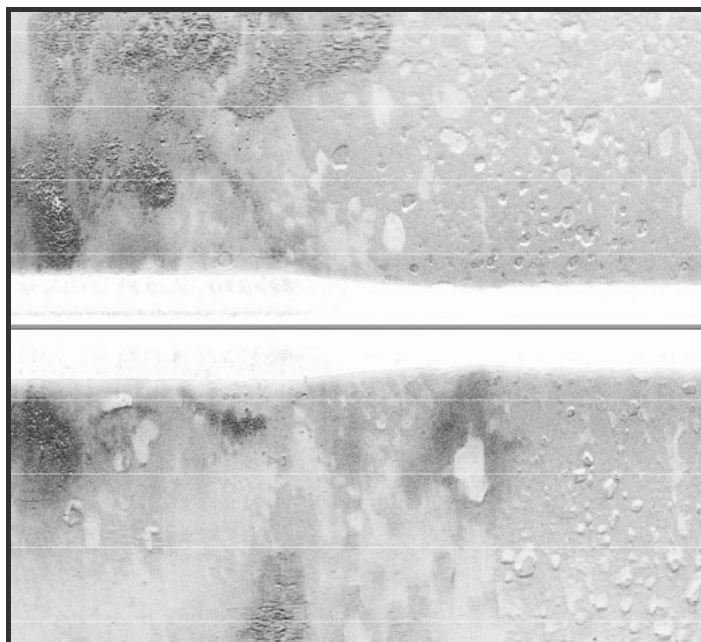
When interpreting remote sensing imagery of seabed habitats, the interpreter needs to relate the expression of ground units given by the imagery to their reality in the field, in other words he looks at signatures of his facies or habitats.

Signatures can vary for a given facies or ground unit. They depend on natural factors, which will make them different to the observer even though they may be clearly assigned to a given class. Slight colour nuances in the composition of the facies itself can be of considerable influence on the way they are remotely-sensed. For example muddy sand can be covered with some microflora, which modifies its aspect on aerial photographs. The mappers use their knowledge and expertise to allow for this variability.

An example is given in the Bay of Concarneau (France) with a “pockmarked texture” (figures), a spectacular facies never highlighted in previous studies in this bay. These bedforms are usually seen in deeper water. The pockmarks form a highly unusual chain pattern here. The numerous samples taken from these bottoms were made up of a compact mud with highly concentrated *Haploids* communities on their surface.



Still camera view showing a small-sized pockmark.



Side-scan sonar imagery of pockmarks in the Bay of Concarneau. The EUNIS class is “Sandy mud”. The facies is described as “clear grey facies with craters”.

These signatures can then either be used in automated classification or in direct manual interpretation, depending on the methodological choice of the interpreter. The quality of the interpretation will depend directly on the quality of these signatures.

Links to other sections:

[Habitat signature catalogue](#)

Habitat signature catalogue

The lack of signatures directly obtained by surveying can be compensated to a certain extent by using “historic signatures” collected by other surveyors. The MESH project recognised that the many examples collected by the partnership had a value beyond the project and that they should be collated and made available to the wider community. The value of such a catalogue is largely enhanced by its diversity in terms of types of techniques, conditions of acquisition and geographic locations of the targets. Only by looking at a large number of cases is the interpreter likely to find a signature identical to the one he is investigating and to really be aided in his interpretation. When designing a catalogue it is therefore recommended not to limit the number of occurrences. Besides, the signature metadata have to be carefully collected for the user to be fully aware of the conditions in which the signatures were generated.

The MESH catalogue of signatures

A number of habitats (see table) were initially identified by the MESH partnership. Only habitats that showed conspicuous signatures when viewed by remote sensing tools were selected. These habitats may belong to various EUNIS levels: generally higher levels are concerned, where abiotic drivers shape the signature. An example of this is “fine sand”, where many different lower type habitats can be encountered.

However, some belong to lower levels where the infauna on sediment substratum or the flora on rocky substratum has a strong influence on the habitat outlook.

At first there was a need to formally describe all the habitats by a summary and keywords. An example table is given for mussel beds. Then comes a list of “blocks”, a block defined as occurrences of this habitat at a particular location. Each block contains as many signatures as were collected by the various remote sensing techniques. An example from les Abers, Brittany is given where seagrass is viewed by aerial photography, side-scan sonar, hydrographic LIDAR and field pictures. Each signature also contains metadata describing in which conditions it was collected.

The geographic location of each block is shown in an insert, and the catalogue is also dynamically linked to the MESH webGIS in such a way that at any time a user looking at a habitat map can instantly view a signature located in the same area. The MESH catalogue is online at (<http://www.rebent.org/mesh/signatures/>).

The MESH catalogue is expected to be a live tool and any user a potential contributor. More signatures are welcome, provided they illustrate different aspects of habitats. An instruction manual as how to populate the catalogue or consult can be found in ([manual catalogue signatures.pdf](#)).

[Habitats usual names](#)
[EUNIS Code search](#)
[Keyword search](#)
[Multi-Criteria search](#)

Here is the comprehensive list of habitat sheets available in the catalogue of seabed signatures. To access one habitat sheet, click on the habitat name.

Search results : 31 habitats usual names

Barnacles	Dredged gravel beds (circalittoral)	Mussel Beds
Bedrock outcrops	Dump spots of dredged materials	Oyster beds
Boulder fields	Fine sand	Pockmark
Boulder reef (circalittoral)	Fucoid beds	Sabellaria alveolata beds
Carbonate mounds	Gravel and Sand ribbons (circalittoral)	Sabellaria spinulosa reef
Circalittoral sand waves	Kelp beds	Saltmarshes
Coarse sand	Lanice conchilega reef	Sand bank
Cobble reef (circalittoral)	Maerl beds	Undredged gravel beds (circalittoral)
Cobble reef (intertidal)	Medium sand	Zostera marina beds
Collapsed mine shafts (circalittoral)	Megaripple field (circalittoral)	
Crepidula Beds	Mixed sediments (circalittoral)	

Habitats as found in the catalogue of signatures

[Multi-Criteria search](#)
[Habitats usual names](#)
[EUNIS Code search](#)
[Keyword search](#)

Mussel Beds

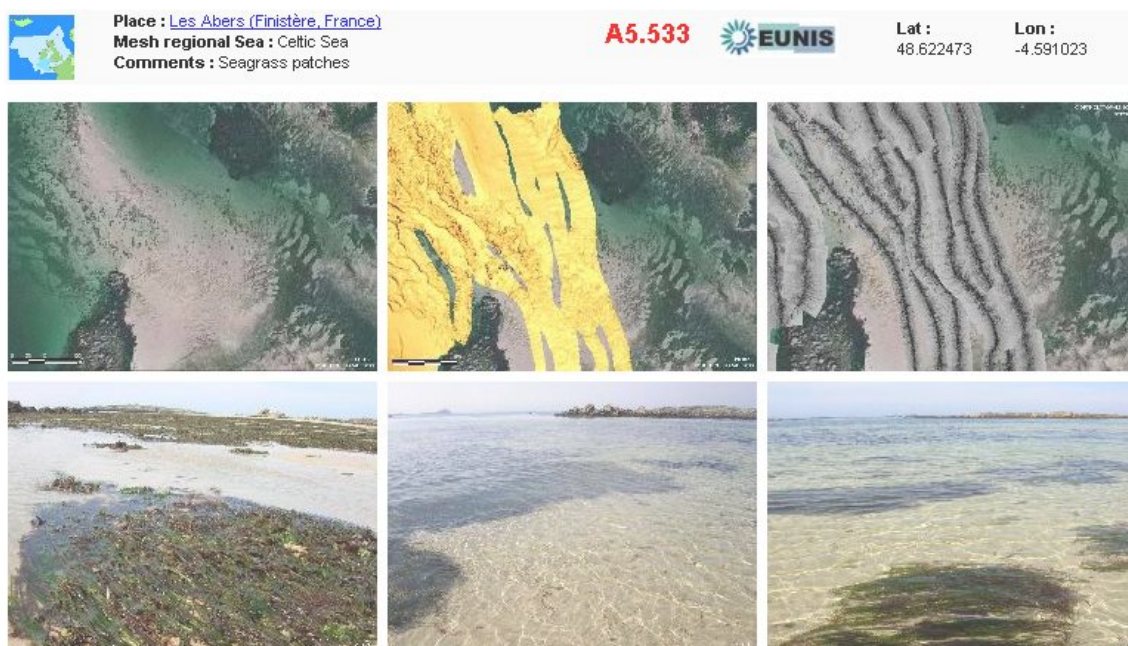
Biotope description :
'A mussel bed is a benthic community structured by blue mussels. It may consist of a spatially well defined irregular collection of more or less protruding smaller beds, which may be called patches, separated by open spaces. This description entails also young beds with a high abundance of small mussels. The described structure may not be so distinct in young beds or just settled beds (spatfall). Within these mussels many species may occur because the bed provides shelter and hard substrate. Typical species are barnacles, bryozoans and macro algae. At low tide many birds are present.

Weblinks :

Key-words : [mussel](#), [mussel bed](#), [biogenic reef](#), [Mytilus edulis](#)

Habitat description :
Intertidal beds occur in the lower reaches of the intertidal between the low water level and mean tide. Above that seedbeds may occur but won't survive because of long emergence and lack of feeding time. Beds can occur on substrates ranging from sand to mud including shell deposits. On hard substrates like rocky shores, piers and groynes mussels may occur in high densities, but these are not considered to be mussel beds.

Example of habitat description (Mussel beds)



Three signatures of seagrass in shallow water, aerial photograph, Lidar, side-scan sonar and three field pictures (les Abers, Brittany).

Link to resources:

[Manual catalogue signatures.pdf](#)

Links to websites:

<http://www.rebent.org/mesh/signatures/>

Statistical modelling

A suite of statistical techniques exist that can be used to model habitats. An excellent overview of statistical modelling techniques to predict habitat distributions can be found in Guisan and Zimmermann (2000). Most of the statistical modelling techniques can be grouped into seven categories: Multiple regression and its generalized forms, Classification techniques, Environmental envelopes, Ordination techniques, Bayesian approaches, Neural networks and other approaches including mixed approaches (Guisan and Zimmermann (2000)). Also, the following link provides a good overview of statistical techniques out of the context of habitat modelling (<http://www.statsoft.com>)

In Guisan and Zimmermann (2000), an overview is given of different statistical methods that can help in the prediction of habitats. In some cases, habitat suitability models will be developed with the most important goal defining quantitatively the niche of a species. Habitat suitability models can assign a probability of occurrence for each location, based on the local environmental variables. Additionally the models can be applied to construct full cover species distributions maps. These maps are generated by feeding full coverage maps for each environmental variable into the model. The model will predict the probability of occurrence for the species for

each pixel of the raster. The modelling also helps to assess the ecological niche of the species in a more quantitative way and gives a probability of occurrence for each variable combination.

The main statistical approaches to modelling can be grouped into seven categories:

1. Generalized regressions

Regression techniques relate a response variable to a single or a combination of environmental predictors or explanatory variables. The classical least square regression (LS) is only valid when the response variable has a normal distribution. Generalized linear models (GLM) can handle distributions such as the Gaussian, Poisson, Binomial or Gamma. LS regressions could predict unfeasible values such as negative values or probabilities higher than 100%, while GLM can only yield predictions within the limits of observed values (e.g. probabilities of species between 0 and 1). Examples of the use of regression techniques for habitat modelling can be found in:

Links to resources:

[WE Ifremer predictive modelling seaweeds.pdf](#)

[Habitat suitability modelling MESH.pdf](#)

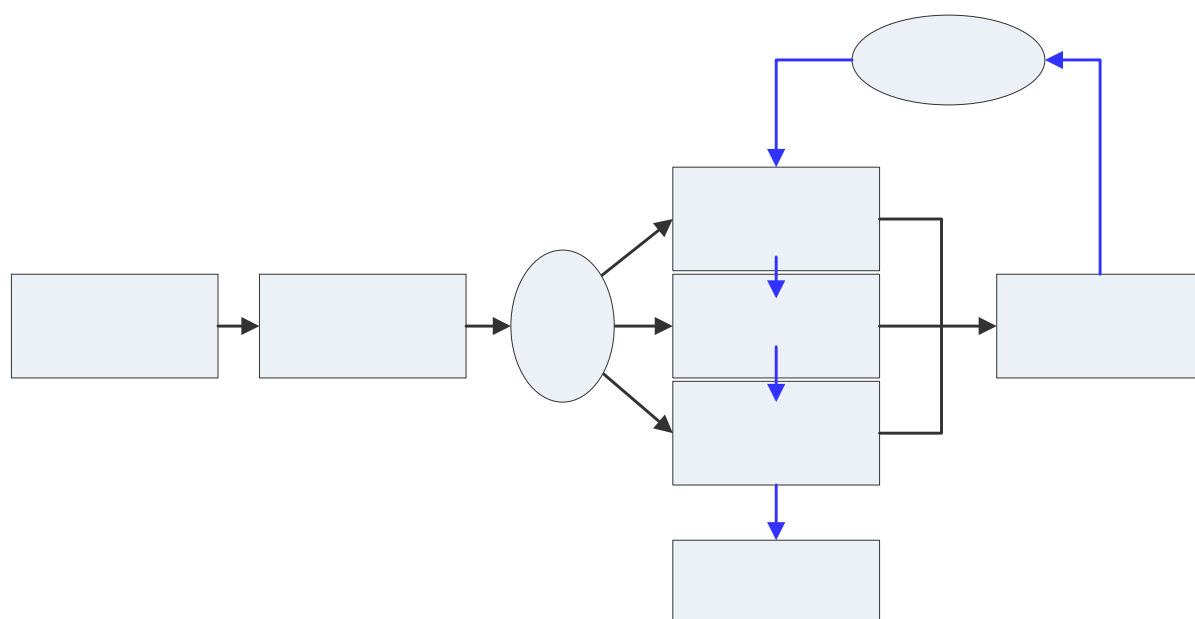
[maerl regression.pdf](#)

Links to websites:

<http://www.statsoft.com>

2. Classification techniques

A broad range of techniques falls under the category of classification. Those techniques assign a certain class of the response variable to each combination of the environmental predictive variables. Some examples of techniques are classification and regression trees, rule-based classification and maximum likelihood classification.



Deriving statistical signatures from ground-truth samples.

In the context of image classification (e.g. satellite imagery), supervised and unsupervised classification are common approaches. In the case of supervised classification, the software system delineates certain classes based on statistical characterization data drawn from known examples in the image ('training sites') (Eastman, 1999). The most common form of supervised classification is maximum likelihood classification (e.g. ([ENV CS05 Maximum Likelihood Classification of AGDS Data.pdf](#))). Supervised classification can be applied directly using the biological ground-truth data without going through the intermediary of an interpretation of sediment type. In a sense the integration of physical coverages and biota is done in one step. This works well for small areas, but the opportunity for confusion of signatures between different habitats increases with the survey area. However, supervised classification can be refined by using prior probability images based on what is known about broad trends in habitat distribution to constrain the predicted habitat distribution (e.g. [ENV CS09 Prior Probability Classification.pdf](#)). Unsupervised classification uses clustering software to classify the image, without the use of training sites. Segmentation is a form of unsupervised classification where adjacent pixels with similar pixel values are grouped to segments. This might be used during surveys for a quick interpretation of remotely-sensed data, but is not often used for the final analysis of the data. An example of unsupervised classification can be found in:

Links to resources:

[WE Ifremer predictive modelling seaweeds.pdf](#))

[ENV CS05 Maximum Likelihood Classification of AGDS Data.pdf](#)

[ENV CS09 Prior Probability Classification.pdf](#)

3. Environmental envelopes

The environmental envelope of a species is defined as the set of environments within which it is believed that the species can persist: that is where its environmental requirements can be satisfied (Walker and Cocks, 1991). Many large-scale vegetation or species models are based on environmental envelope techniques.

Links to resources:

[Maerl regression.pdf](#)

4. Ordination techniques

The term 'ordination' was introduced by Goodall (1954) for methods that arrange samples or species in relation to 'a multidimensional series'.

A well-known example of ordination is Principal Components Analysis, being a linear dimensionality reduction technique, which identifies orthogonal directions of maximum variance in the original data, and projects the data into a lower-dimensionality space formed of a sub-set of the highest-variance components (Bishop, 1995).

Most habitat models that predict the distribution of species or communities that use ordination techniques are based on canonical correspondence analysis (CCA). In

this direct gradient analysis the principal ordination axes are constrained to be a linear combination of environmental descriptors (ter Braak, 1988).

5. Bayesian approach

Models based on Bayesian statistics combine 'a priori' probabilities of observing species or communities with their probabilities of occurrence conditional to the value of each environmental predictor (Guisan and Zimmermann, 2000).

6. Neural networks

(Artificial) neural Networks (ANN) are analytic techniques modelled after the (hypothesized) processes of learning in the cognitive system and the neurological functions of the brain and capable of predicting new observations (on specific variables) from other observations (on the same or other variables) after executing a process of so-called learning from existing data (definition (<http://www.statsoft.com/textbook/stathome.html>)). ANN is not often used for habitat distribution models. An example of the use of ANN can be found in:

Links to resources:

[Habitat suitability modelling MESH.pdf](#)

7. Other approaches

Examples of other approaches are simple models in GIS, e.g. using overlays of environmental variables and absence/presence of species.

Discriminant function analysis is another example of another approach. Discriminant function analysis is used to determine which variables discriminate between two or more naturally occurring groups (it is used as either a hypothesis testing or [exploratory](#) method) (definition (<http://www.statsoft.com/textbook/stathome.html>)). An example of Discrimination function analysis can be found in: [WE UGent Habitatsuitability EUNIS.pdf](#)

Links to other sections:

[Supervised classification using image processing tools](#)

Links to resources:

[WE UGent Habitatsuitability EUNIS.pdf](#)

Links to websites:

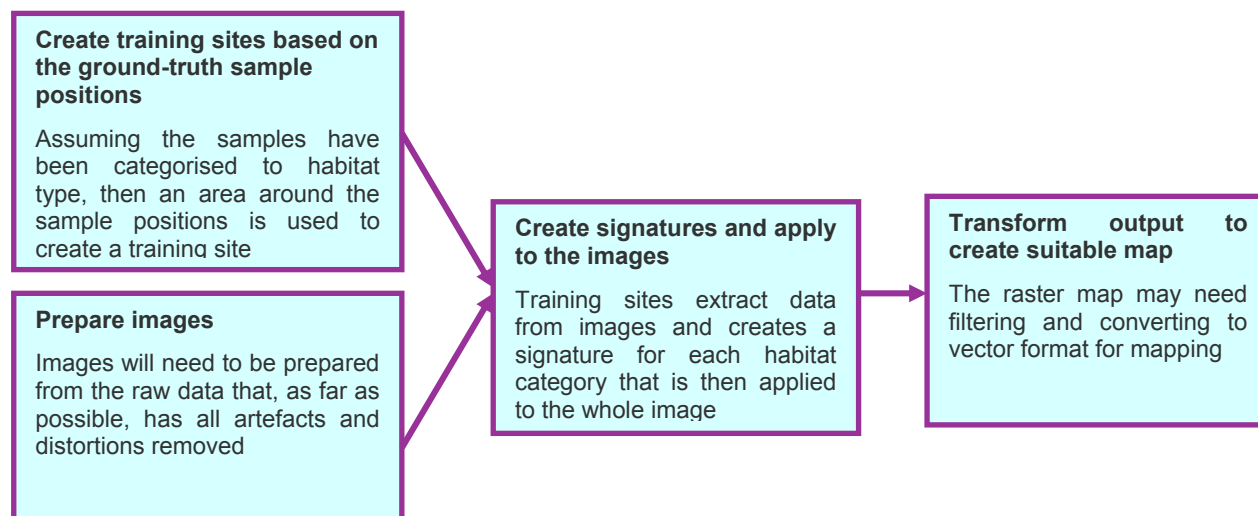
<http://www.statsoft.com/textbook/stdatmin.html#eda>

Supervised classification using image-processing tools

Principles

Classification using maximum likelihood classifiers is well established as a technique for the interpretation of satellite and airborne remotely-sensed images and there are

very well developed software packages that provide a range of sophisticated tools for the whole process of image-processing. The principals are quite straightforward:



Scheme of a supervised classification using image-processing tools

Supervised classification is a **data-driven (empirical)** modelling tool in that the process derives statistical relationships between the input variables and the ground-truth habitats.

The traditional image is the result of the detection of reflected ambient daylight (visible and near-visible spectra). The spectral response depends on the nature of the objects the light is reflected off and this is used to create a characteristic signature for each habitat type. The training site is like a 'cookie-cutter' in that it cuts through all the spectral image layers and extracts the values for each spectrum. These spectral values are then used to create the habitat signature. The 'signature' is in the form of a statistical probability distribution in as many dimensions as there are input images ('n' dimensions). The probability distribution is calculated using the maximum likelihood estimator (MLE). Each habitat will have its own signature and together they form a signature catalogue.

These signatures are then applied to the whole image (actually, a stack of images, one for each spectrum). The spectral values for each pixel (one value per spectrum) are matched to the signature catalogue and each pixel is given a probability value of belonging to each habitat category depending upon where it lies in the n-dimensional probability distribution.

Usually, the corresponding pixel of the habitat image is assigned to the habitat that has the highest probability. This decision-rule is termed a 'hard' classifier since it does not take account of uncertainty. However, the probabilities can be used in other ways to create maps that reflect uncertainty of the classification (e.g. through fuzzy classification).

Application

Supervised classification has been developed for satellite image-processing where it has been applied to the classification of spectral layers. However, it can also be applied to other forms of remote sensing and has been used for the classification of interpolated acoustic reflectance data, for example AGDS values. It can also be used to combine different forms of data, not just 'spectral' values. For example, classification can combine reflectance, depth (height), variability etc. It does this by

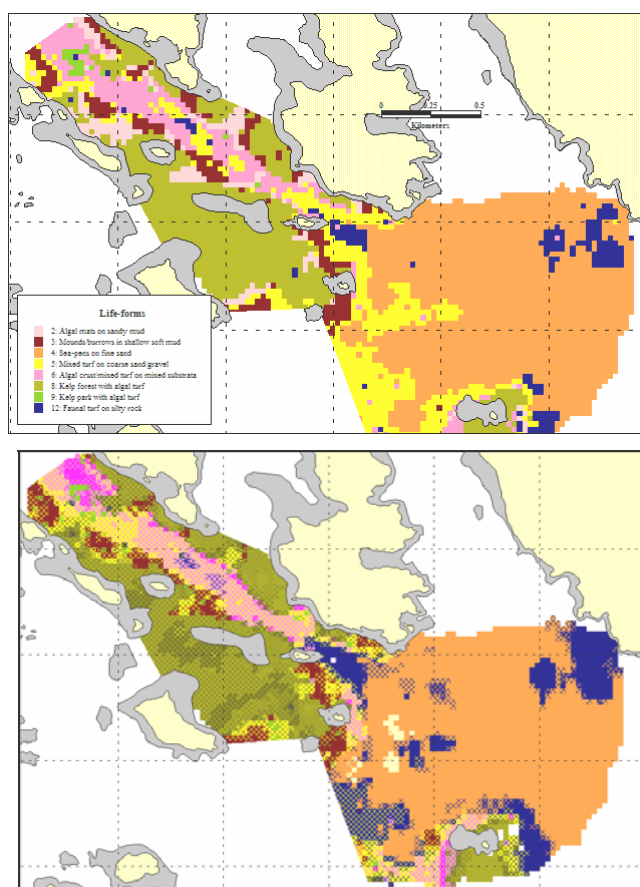
stretching values of every dimension to lie between 0-255. Although many of the variables may be correlated, supervised classification is a very versatile and statistically robust tool.

However, to work well the images should be distortion-free. This is often difficult to achieve with different degrees of distortion across an image. Distortion away from nadir (the point on the sea floor directly under the remote sensing instrument) is a particular problem for many acoustic swath systems and is particularly noticeable where image strips have been mosaiced together.

One of the disadvantages of this data-driven approach is that the signatures, representing the links between the variables and the habitat categories, are largely hidden from the user and not easily exportable to other images. Whilst it is possible to export signatures, more usually each new survey needs to be interpreted independently of previous surveys. Combining surveys may often rely simply on overlapping maps and seeing where there is or is not concordance and editing 'by eye'.

However, more sophisticated use can be made of the underlying probabilities and many of the apparent limitations of supervised classification can be overcome.

The following is an example of hard supervised classification and a 'fuzzy' classification of the same area using the same AGDS and ground-truth data of Loch Maddy, Scotland. The predictive capability of the map was greatly improved using the fuzzy approach, but the map may be considered to be more difficult to read and use?



Supervised classification based on AGDS data. The track data were interpolated in Surfer™ and then classified using Idrisi™

Examples of map making

The mapping strategy and the concepts of broad-, fine- and intermediate-scale mapping have been outlined in Section 4.1. Here, a few examples will be presented for each type of mapping. The differences are reflected on all levels of the mapping process; from the broad to the small scale:

- the treatment of biological records differs from a need to significantly reduce a diverse set of data or to have a consistent data set to match directly to a classification system;
- the coverages are derived from multiple sources or are gathered during one survey;
- the integration is rather simple or makes use of sophisticated statistical analysis; and,
- the final map is able to show a range of habitats or probabilities or has a maximum detail on a species or community level.

Links to other sections:

[Examples of broad-scale maps for large areas](#)

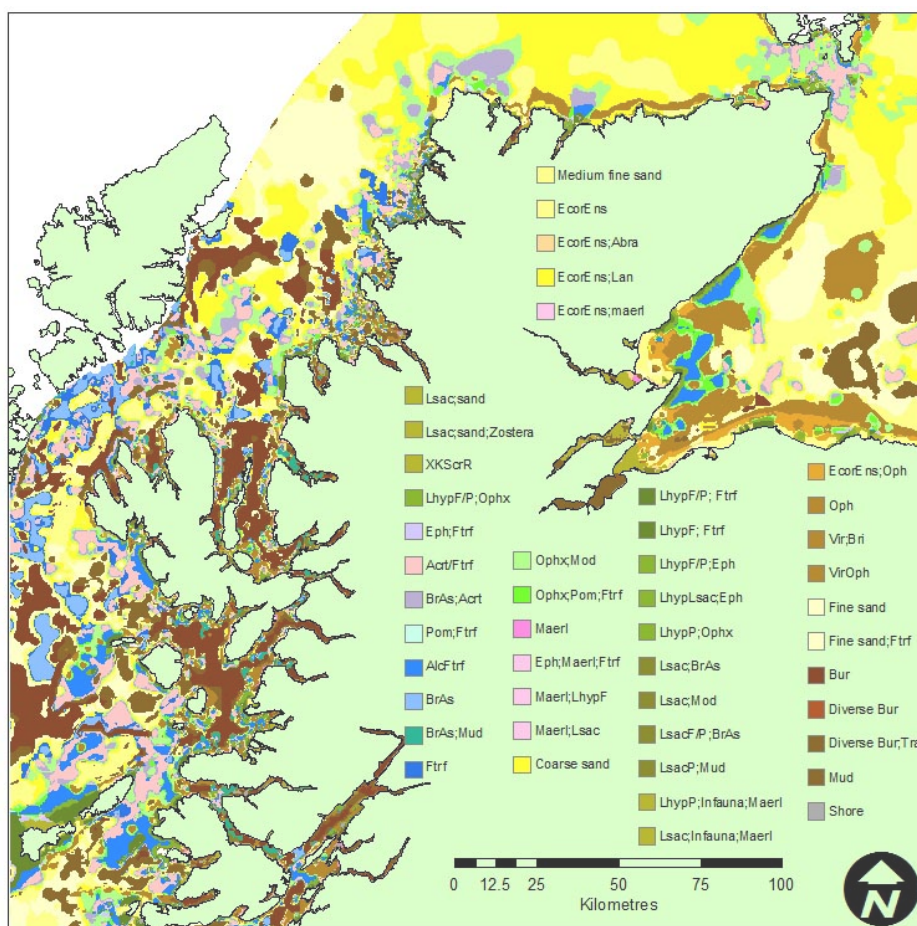
[Examples of fine-scale maps for small areas](#)

[Examples of intermediate-scale maps](#)

Examples of broad-scale maps for large areas

Highland Shellfish Management Organisation study for Scottish coastal waters

The HSMO case study ([Highland Shellfish Management.pdf](#)) presents the results of a 'top-down' GIS project undertaken to offer support to the Highland Shellfish Management Organisation (HSMO) in its review of inshore fisheries management. A GIS was built with data on physical geography, environment, habitats and biotopes, shellfish resources and fishing effort, infrastructure and conservation interests. The GIS was intended to assist HSMO in its management functions by providing a broad environmental overview of the area, its natural heritage features and the interactions that exist between the relevant fisheries and these features. A biotope map was derived by modelling, using the many source data sets. Problems were encountered with the biotope records that lacked spatial precision and there was a poor coverage for the Highland Region. Modelling exposure was not a sophisticated process and the exposure classes were not exactly defined. However, the case study does provide meaningful analysis at the broadest scale, e.g. areas of fisheries conflicts can be successfully predicted. From a management perspective, these GI systems can help set priorities and focus resources, at a regional scale. The main strength of the system is the cost effective use that is made of existing data. It also has the potential to support SEA and EIA and can help facilitate management integration.



Biotope map of Scottish coastal waters: the map is a derivative of substratum, bathymetry, exposure and physiographic data layers that have been cross tabulated and combined with biological point records (Envision Ltd).

EUNIS 'triplet' modelling of the MESH area

MESH adopted the EUNIS habitat classification as its standard scheme for presentation of habitat maps. Because the coverage of detailed maps is somewhat limited, a broader modelling approach was also adopted to predict the distribution of EUNIS habitats across the MESH area (north-west Europe).

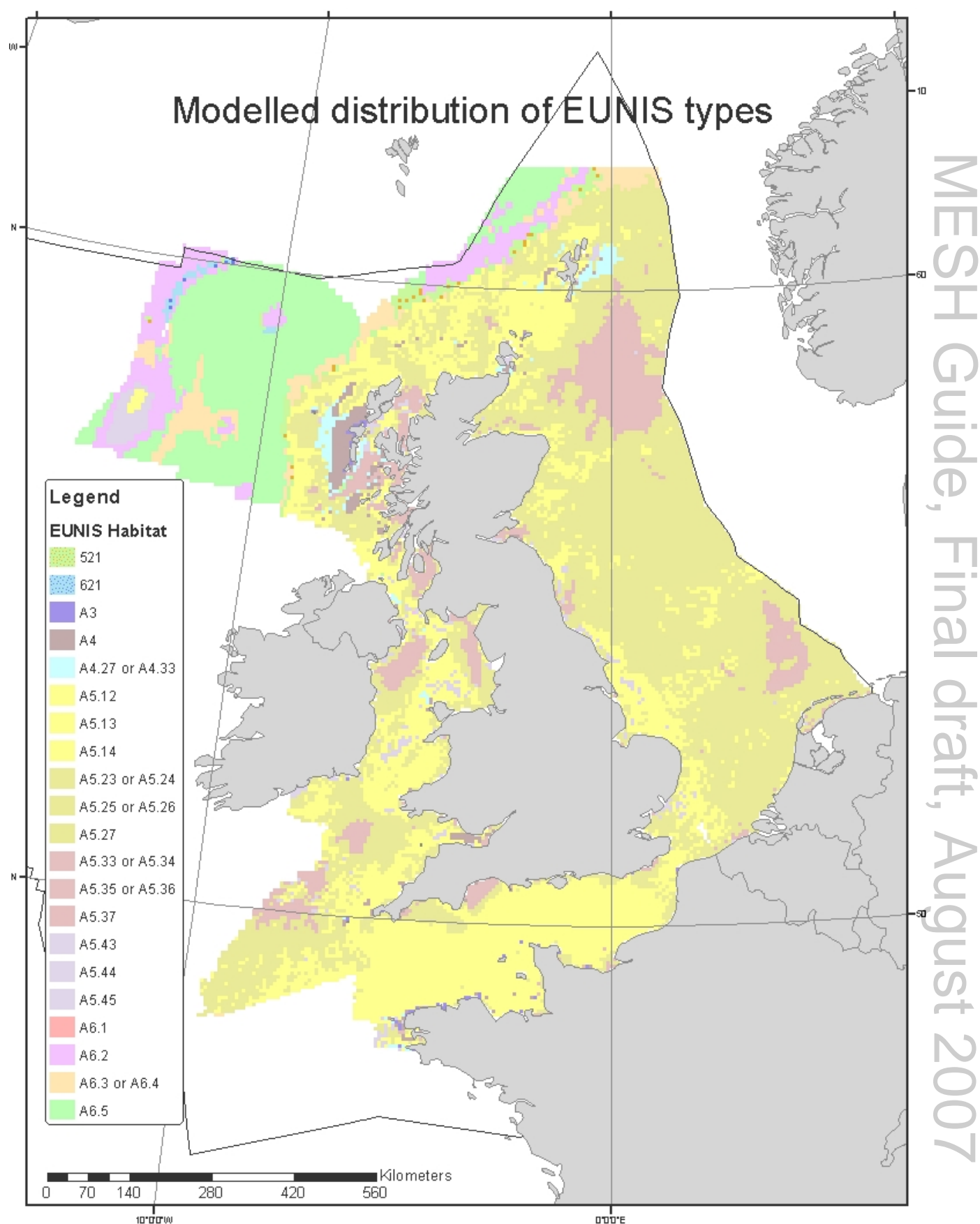
Given the size of the MESH area (five countries) and the need for full coverage data layers to use as input data for the modelling, it was necessary to restrict the modelling to a coarse level in the EUNIS classification (levels 3/4, no biological data involved) and to use the following three data layers.

- Substratum
- Depth zone (based on bathymetry, light penetration and wave base)
- Bed stress

The necessary data layers were sourced either from single data sets or models covering the MESH area (bathymetry, light penetration, wave base, bed stress), or by integrating data from a number of national sources (substratum). Each data layer was categorised according to the classes used in the EUNIS scheme (e.g. bed stress as high, moderate or low) and converted to a vector grid of about 1 nm

square. Each combination of the three data layers was equated to a particular EUNIS habitat class (e.g. Rock substratum, photic depth zone and high bed-stress = EUNIS class A3.1). By analysing the data for each cell in MS Access, it was possible to produce a map showing the predicted distribution of each EUNIS class. More details about the EUNIS model developed by MESH can be found here [\(Worked Example - MESH EUNIS Model.pdf\)](#)

While these predicted EUNIS habitat maps do not replace mapping studies completed using more traditional techniques, they are useful in that they provide a degree of information for areas that remain unmapped, particularly areas away from the coast. They may also highlight possible areas of habitat heterogeneity or potential areas of rare habitat where future mapping studies should focus. The main limit to the resolution of this modelling is the resolution of the individual datasets used. A related problem is the difficulty in obtaining full coverage physical datasets; for example a wave exposure layer.



Map showing the predicted distribution of EUNIS habitats for the MESH area.

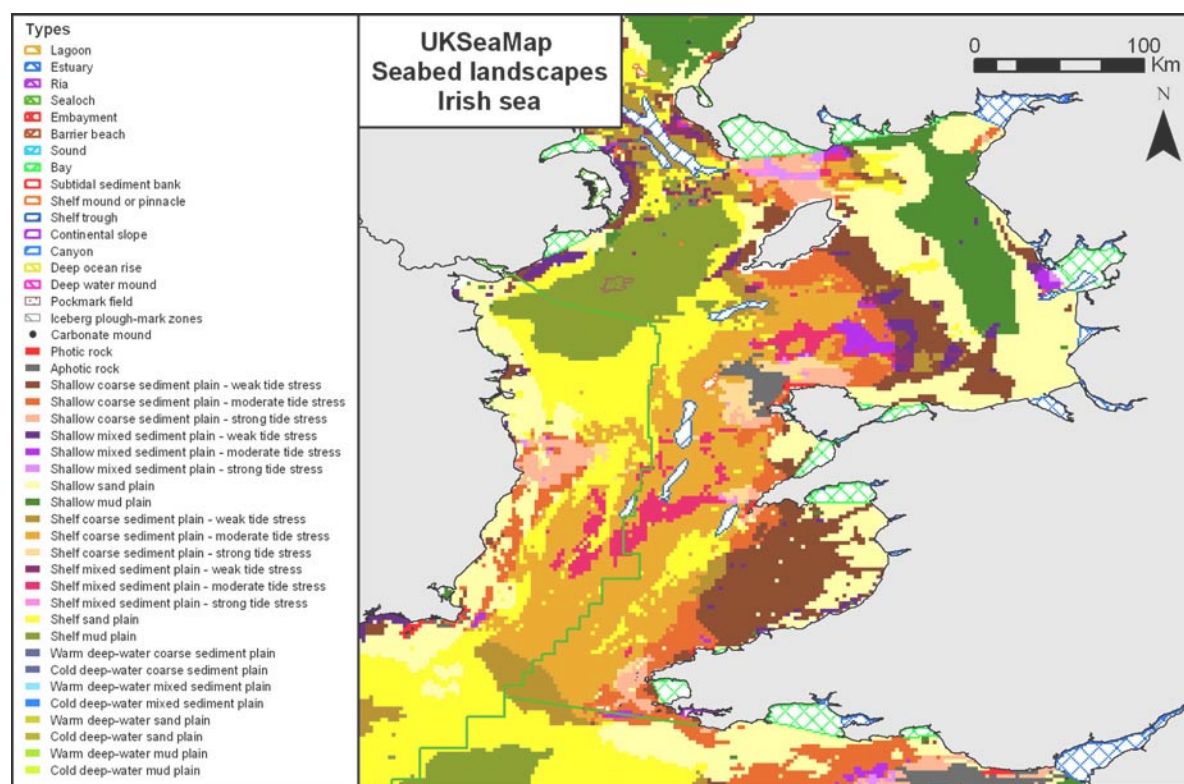
UKSeaMap

The UKSeaMap project (Connor *et al.*, 2006; (http://www.jncc.gov.uk/UKSeaMap/UKSeaMap_FinalReportJan2007.pdf)) adopted a broadly similar approach to the EUNIS triplet study, in using a similar set (the same in most cases) of physical data, summarized to a vector grid, to model the distribution of broad scale habitats for UK waters. However, because EUNIS lacks a topographic element to the way it classifies the seabed, UKSeaMap sought to produce a more topographically based landscape map that would be better suited to regional and national planning and management needs.

UKSeaMap developed this broad 'marine landscape' map by combining three elements:

- A set of topographic and bedform features (e.g. pinnacles, banks, troughs, mounds), identified by shape and slope using bathymetric data,
- A set of coastal physiographic features (e.g. bays, estuaries, lagoons, sealochs), identified by coastal shape, topography and salinity regime, and
- A set of modelled broad habitat features (e.g. shelf mud plain), identified by modelling habitat parameters to determine broad habitat types equivalent to EUNIS classes.

The development of the broad-scale modelling work within UKSeaMap has yielded several advantages over the EUNIS triplet approach. Firstly, end-users readily understand the topographic and coastal features mapped, particularly as many equate to features listed in the Habitats Directive and by the OSPAR Commission as requiring protection, and are thus important in management and policy terms. Secondly, the modelling of habitat features has followed a more flexible approach than EUNIS, to map at both a coarser level where appropriate (inshore rocky features) and at a finer level where appropriate (offshore sediment features) to better suite both the scale of the mapped area and end-user needs.



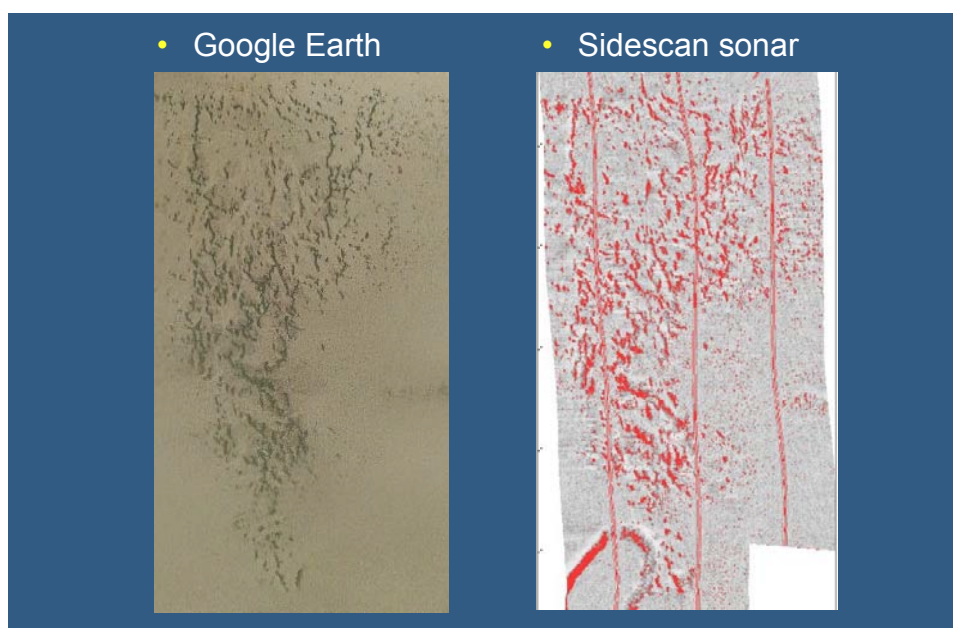
'Marine landscape' map for the Irish Sea, extract from UKSeaMap (Connor *et al.*, 2006).

In the MESH context, additional examples of Marine Landscape maps can be found for Belgian [UGent Marine Landscapes BCS.pdf](#) and Dutch [WE TNO Marine landscape Map DCS.doc.doc](#) waters. The methodology has been adopted in various ways dependent on the purpose of the maps and the available datasets. Although, marine landscape mapping is generally a broad-scale approach to habitat mapping, the resolution of the datasets determined the scale of the final product.

Examples of fine-scale maps for small areas

Detection of (sub)littoral oysters and mussels along Dutch estuaries and tidal flats

Side-scan-sonar has been applied for the mapping and monitoring of shellfish in very shallow waters of the Dutch estuaries and tidal flats [WE TNO SSS oysters mussels.doc](#). Filtering and automatic contouring has been applied as a first step towards quantitatively estimating shellfish densities. The estimation is based on counting the number of peaks in backscattering. This can be done through binning, which counts numbers within unit areas. Binning maps are suitable for automated contouring, which can be used for delineating shellfish banks and accentuating their patterns. The results have been compared with optical images displayed on Google Earth.



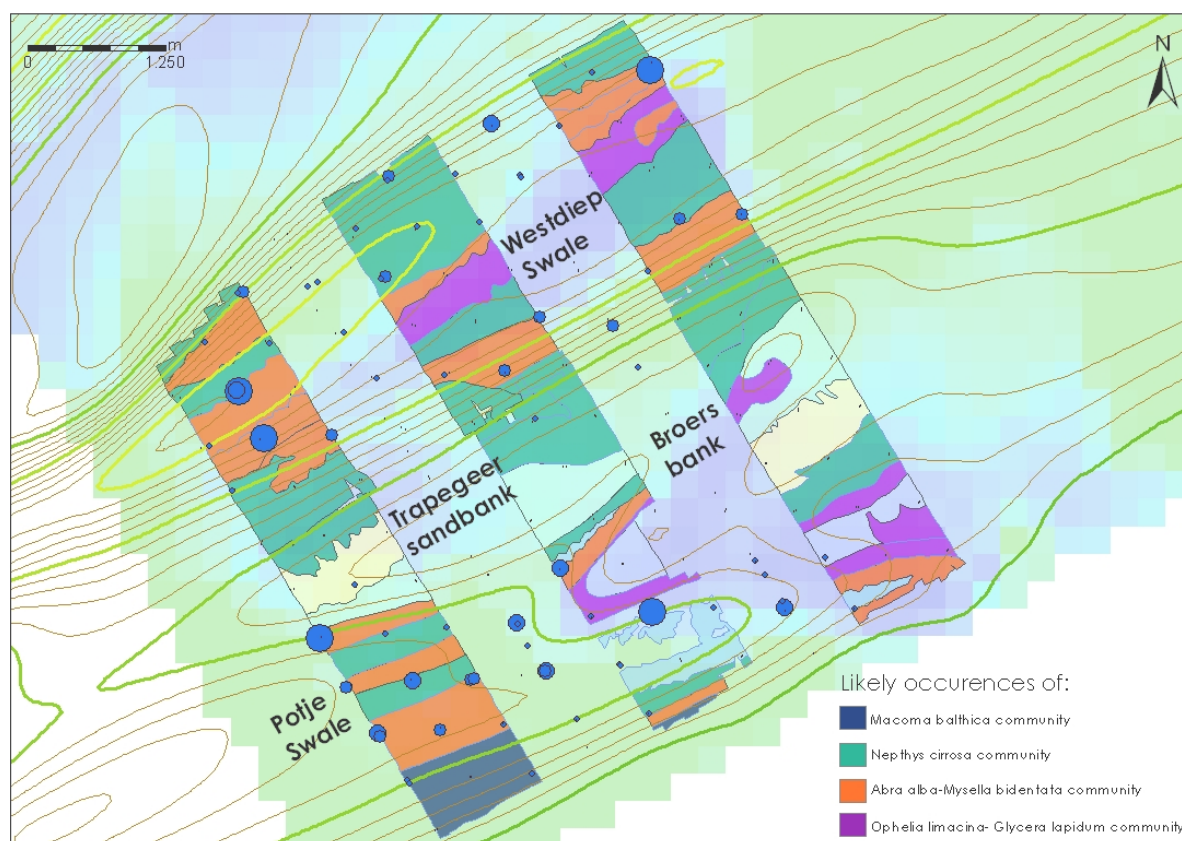
Acoustic and optical images of a dense mussel bank with open spaces (TNO).

Macrobenthic community and species maps of the Belgian part of the North Sea

To obtain small-scale, full-coverage maps of macrobenthic communities and species, very-high resolution side-scan sonar imagery has been used on selected areas of the Belgian part of the North Sea. An example is shown from the Western Coastal Banks (5x5 km²), a geomorphologically and sedimentologically highly diverse area with water depths of 0 to –15 m (MLLWS) (Degraer *et al.*, 2002). Extensive sedimentological and biological samples were taken and additional physico-chemical parameters were measured. A detailed interpretation of the side-scan sonar imagery is shown in the figure. Acoustic facies have been delineated on the basis of their reflectivity, texture and patterns (Van Lancker *et al.*, 2001). Subsequently, these facies were interpreted in terms of sediment packing and distribution. This enabled to link with the habitat preferences of the 4 main macrobenthic communities, found on the Belgian part of the North Sea:

1. *Macoma balthica*;
2. *Abra alba* – *Mysella bidentata*;
3. *Nephtys cirrosa*; and,
4. *Ophelia limacina*.

They occur along the gradient muddy sands, fine to medium sands with mud, well-sorted fine to medium sands and medium to coarse sands. The background of the figure is a probability map of the presence/absence of the tube building polychaete *Lanice conchilega*. The predictive modelling results, at a grid of 250 m, were obtained using artificial neural networks (Willems *et al.*, in press). In the orange facies, often patchy patterns with a slightly higher backscatter can be distinguished; they correlate well with the presence of dense colonies of *L. conchilega*.

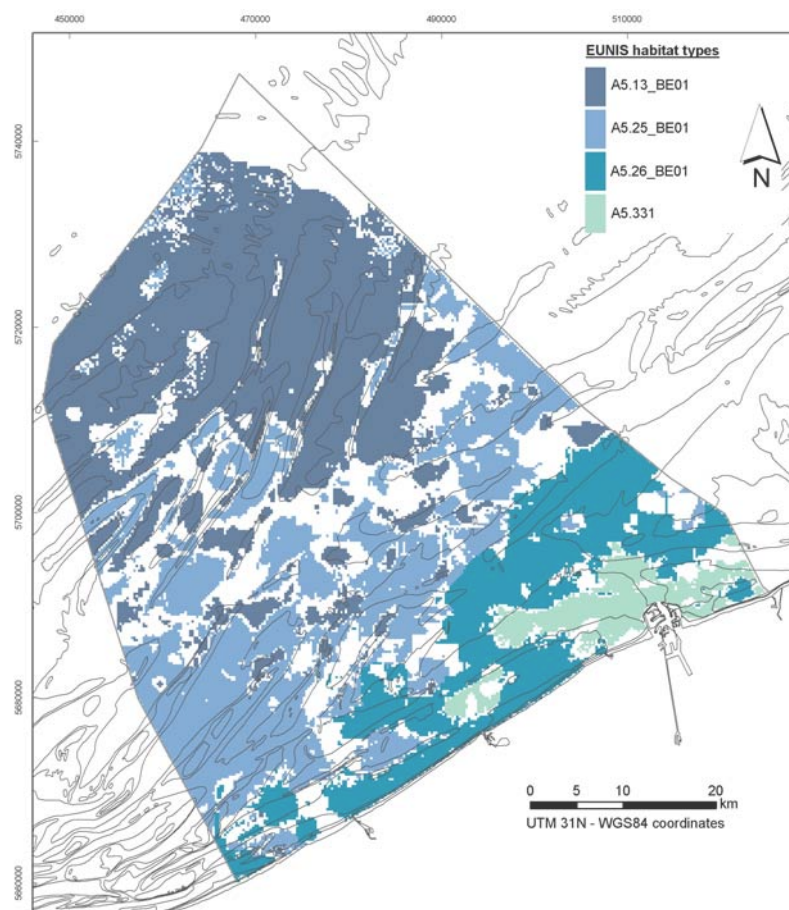


Macro-benthic community map of the Western Coastal Banks based on acoustic facies delineation (March 2000). The superimposed sampling points indicate densities of the tube building polychaete *Lanice conchilega*. The background is a probability map of the presence (green) / absence (blue) of this species. The orange coloured facies corresponds with the likely occurrence of the *Abra alba* - *Mysella bidentata* community, of which *L. conchilega* is the key species (Ghent University).

Examples of intermediate scale maps

Probability distribution maps of macrobenthic communities covering the Belgian part of the North Sea

Probability distribution maps of macrobenthic communities have been produced covering the entire Belgian part of the North Sea [WE UGent Habitatsuitability EUNIS.pdf](#) from biological and sedimentological databases. A habitat model, based on discriminant function analyses was developed based on the biological dataset, comprising both species and environmental data. Median grain-size and silt-clay percentage were the most discriminating parameters. These variables were modelled on a grid of 250 m using advanced geostatistical tools. In GIS, the habitat model was applied on these datasets. Probability maps of the 4 macrobenthic communities were constructed. These were then translated into a EUNIS map. At present, not all of the macrobenthic communities can be attributed to a EUNIS class. Ecological validation proved highly successful.



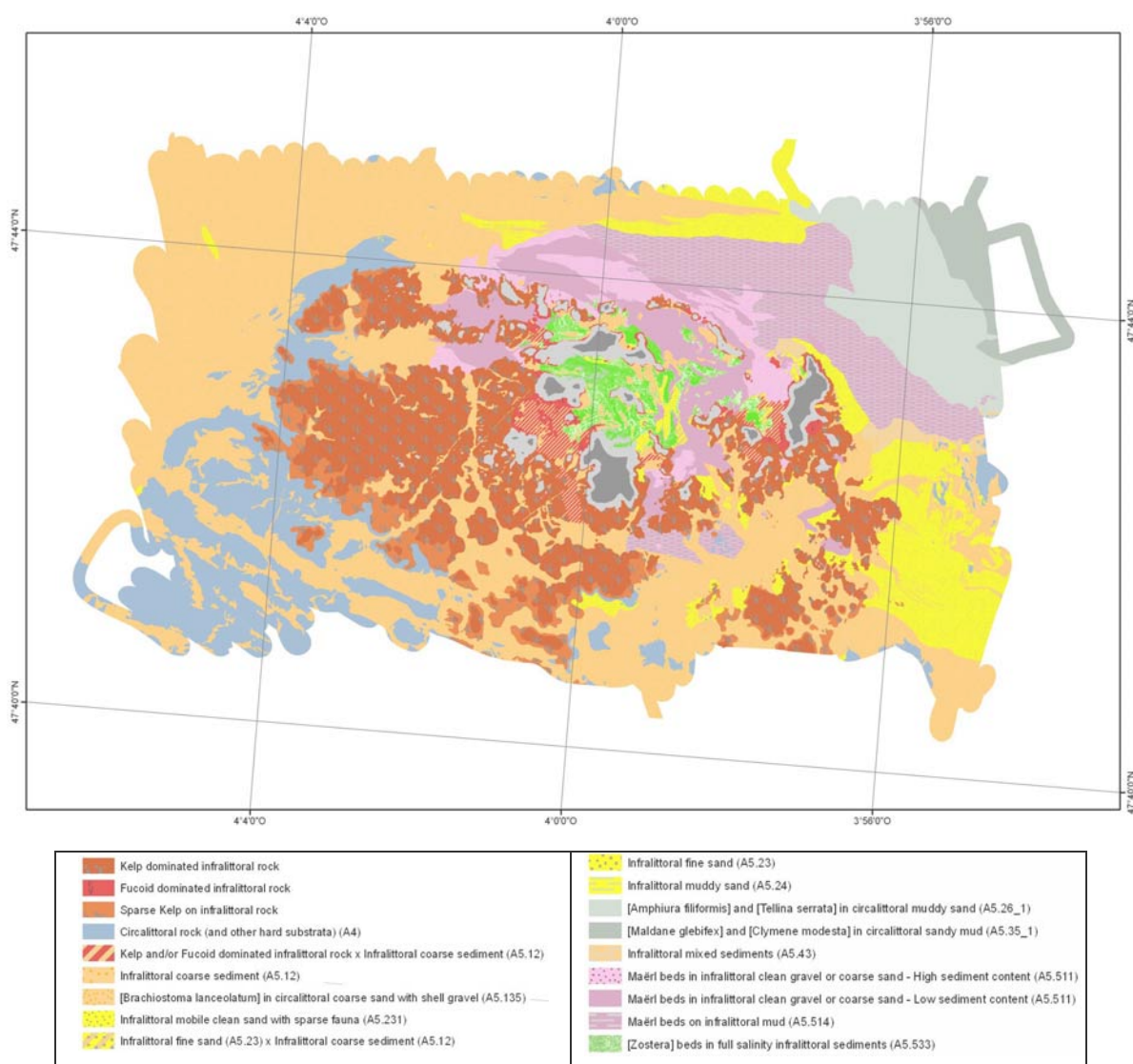
Provisional EUNIS habitat types as defined on the Belgian Continental Shelf (Ghent University).

Mapping EUNIS habitats along the Glénan Archipelago

The Glénan archipelago is known for both its outstanding geomorphological structure and its highly diverse benthic habitats. Numerous studies have been made on this sector, many of them specific and independent. The archipelago provides an optimal study area for general mapping of benthic habitats made by collating data from several sources ([Glenan Archipelago Case Study.pdf](#)). A large number of remarkable habitats are located in the Glénan archipelago. There are extensive maerl beds at the centre and to the North-east of the archipelago. *Zostera marina* eelgrass beds are the main habitat in the centre of the sector. To the southwest of the principal islands, there is a large rocky flat, mostly colonised by macroalgae like kelp. Finally, although their surface area is limited, in the intertidal sector there are large belts of fucoids and large areas covered by fields of boulders that are particularly rich in terms of species present.

The baseline data were coastal orthophotography for the intertidal zone and shallows (depths less than 8-10 m), the SHOM's sedimentological G map, modified by photo-interpretation for shallow bottoms, along with observations made on hyperbaric dives at greater depths. Samples were taken for granulometric analysis or sometimes the grain size class was visually estimated. Coastal orthophotographs (resolution 1 m) were processed by enhancement before being put into unsupervised categories. Ground-truthing was then used to classify each pixel of the image as a

habitat type with respect to its spectral signature. Additional survey work included dual hydrographic LIDAR and very-high resolution side-scan sonar and multibeam, validated with grab sampling and video. SPOT satellite imagery was also available and many photographs. The habitats were described in accordance with the Natura 2000 typology given in detail in the Natura 2000 habitat manuals (2004).



Benthic habitats in the Glénan archipelago. The map is a result of many different data sources. The final interpretation is translated into EUNIS classes (Ifremer).

Biotope characterisation, Bristol Channel

The Outer Bristol Channel Marine Habitat Study (Mackie *et al.*, 2006) examined the biology and geology of the seabed in an area with potential as a marine aggregate resource. The infaunal distributions were most strongly correlated to depth and five sediment parameters. The biotope map was derived from a semi-quantitative cluster analysis of both infauna and epifauna, adjusted with reference to the seabed character and bedform map. The three main EUNIS infaunal biotopes were A5.242 (SS.SSa.lmuSa.FfabMag) in the predominantly fine sands of Carmarthen Bay, A5.124 (SS.SCS.ICS.HeloMsim) in the extensive NOBel Sands sandwave field, and A5.132 (SS.SCS.CCS.MedLumVen) in the predominantly gravely sediments of the

SOBel Sands to the south. In many areas multiple biotopes co-occurred, and epifaunal biotopes A5.232 and A5.444 (hydroids) and A5.611 (*Sabellaria spinulosa*) were present as overlays, particularly on the more stable gravely sediments. The biotopes were defined as EUNIS or taken from (<http://www.jncc.gov.uk/marineclassification>).

The biotope map is primarily based on the biology at each sample position, adjusted with respect to a full-coverage seabed character map. The latter map has been constructed on the basis of side-scan sonar and multibeam imagery, validated with sampling and video imagery.

The Bristol Channel mapping is extensively described and illustrated in Mackie *et al.*, (2006).

Links to resources:

[ENV CS07 Highland Shellfish Management.pdf](#)

[Worked Example - MESH EUNIS Model.pdf](#)

[UKSeaMap_FinalReportJan2007.pdf](#)

[WE_UGent_MarineLandscapesBCS.pdf](#)

[Dummy file WE TNO Marine Landscapes Wadden.doc](#)

[WE TNO Marine landscape Map DCS.doc\).doc](#)

[WE TNO SSS oysters mussels.doc](#)

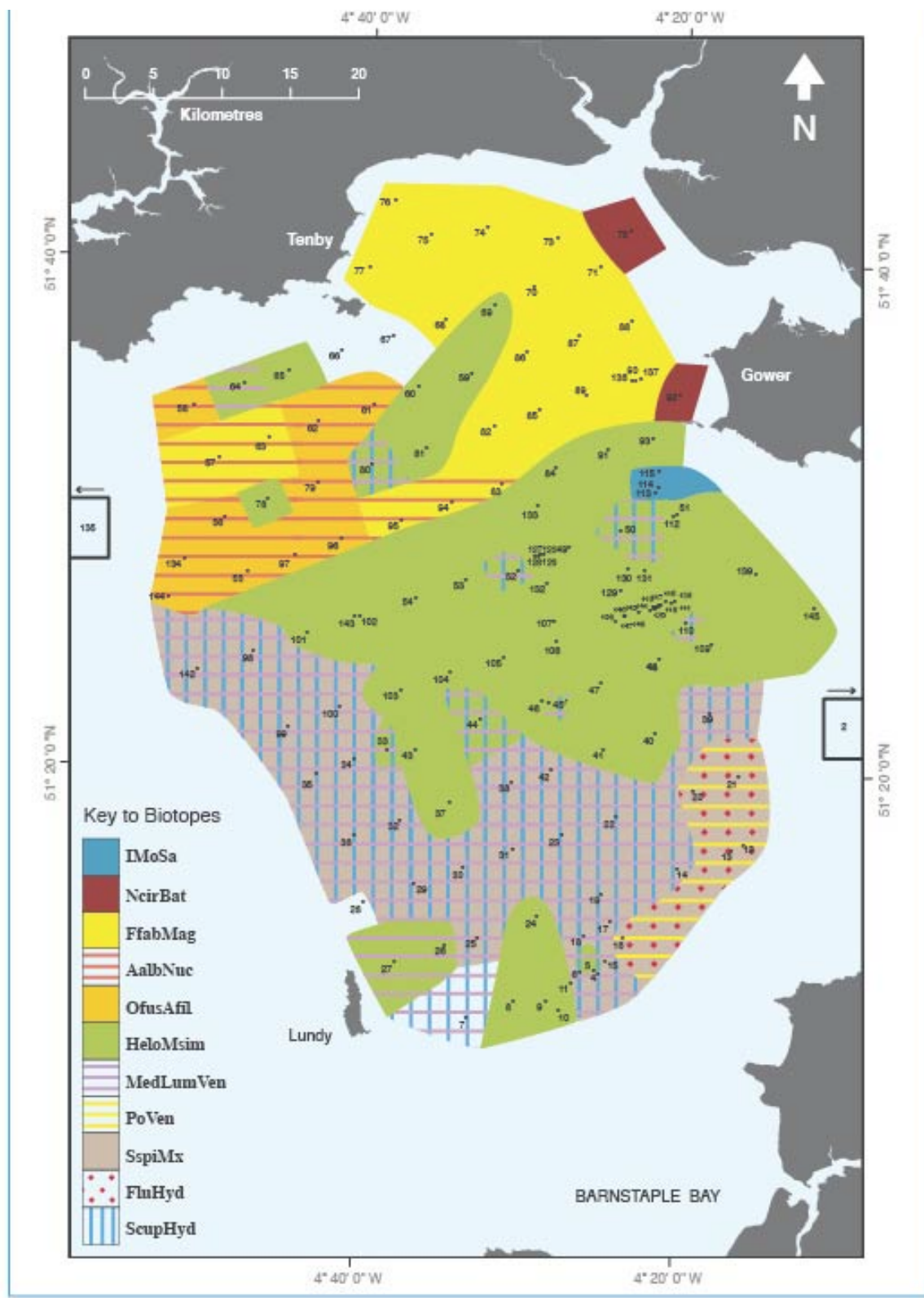
[WE_UGent_Habitatsuitability_EUNIS.pdf\)](#)

[Glenan Archipelago Case Study.pdf\)](#)

Links to other websites:

<http://www.jncc.gov.uk/UKSeaMap>

<http://www.jncc.gov.uk/marineclassification>



Biotope map of the Bristol Channel integrating biological and environmental parameters. It is derived from cluster analysis and a seabed character and bedform map (Mackie *et al.*, 2006).

Design of the habitat map

Finally, when a map is designed, it is important to consider whether the final product needs to be a printed map or an electronic map. With electronic maps, users can zoom in and out freely to the desired resolution, whereas with printed maps, users are obliged to interpret the map at a specified scale. Still, with electronic mapping, users should be made aware to which scale the data still supports an adequate or meaningful level of zooming.

A quick guide will be offered to some of the common problems associated with *electronic cartography*, specifically with the creation of suitable habitat maps of the seabed and visualising them in Geographic Information Systems (GIS). Some practical solutions are suggested to improve map delivery.

Some important items for designing maps are the correct use of coordinate systems of data, metadata (e.g. data owner), attribute information being information stored in a table such as unique codes, names of features etc.

Links to other sections:

[Making electronic maps using GIS](#)

[Map layout](#)

Making electronic maps using GIS

The two most commonly used GIS software for creating seabed maps in Europe are the suite of GIS software products produced by ESRI™ (<http://www.esri.com>), commonly referred to as ArcGIS™, and MapInfo Corporation's MapInfo Professional™ (<http://www.mapinfo.com>). The advice described in the following sections is based on the experience using ESRI™ products, but the same basic principles can be applied to MapInfo Professional.

Coordinate systems

It is possible to create polygons in a data resource (essentially an electronic map) without defining a coordinate system associated with your features. For future use however, it is essential that the coordinate system is clearly defined. You can verify the coordinate system used by a data resource in GIS. It is not possible to view files lacking a coordinate system together with those that do have a defined coordinate system in a GIS. There are tools available in GIS to define (or change) coordinate systems. For example, ArcToolbox™ contains a "Projections Toolset" which can:

- Define a dataset's coordinate system information if it is missing;
- Modify a dataset's existing coordinate system information, and;
- Convert a dataset from one coordinate system into a different coordinate system.

Metadata

Adequate documentation of data resources with metadata can save time in the future. Metadata are described in detail in the section *What can I do with my map?*. Metadata can be stored with the resource itself (internal) or held in a searchable

metadata catalogue (external). Internal metadata are transferred with the data and are easily accessible for people using the data to reference. For example, the ArcCatalog™ application allows users to create internal metadata. It is possible to export internal metadata to databases, typically using XML (Extensible Markup Language). However, external metadata allow more efficient searching across multiple files, particularly when they are stored in a database. Where external metadata are used, it is crucial to clearly link the external metadata to the data resources, ideally using unique identifiers in the file name of the resource. For example FR000001 refers to the first data resource collated by the MESH partner in France, Ifremer. When you are considering how to create metadata for your data resources, the first question to ask yourself is whether your needs will be best served by using internal metadata, external metadata or a combination of the two.

Attributes

Features in a vector data resource (polygons, points, lines) will normally be associated with some other information, most commonly data or information describing the feature. Such information is known as an attribute and is stored in a data field of a GIS data file. Features may have many associated attributes (for example, identifier, feature name, date, time, analyst name) that will each be stored as separate data fields. These data files are commonly known as attribute tables and are normally used for data analysis and map production.

It is essential that the features in your map are correctly attributed so that other users in the future can derive maximum value from the data resource. As a user of an otherwise excellent electronic map, one of the most frustrating situations occurs when the information about what is actually on the seabed at a particular location is not held as part of the data resource. For example, when a cartographer uses a linked table or legend file to symbolise the map, these crucial look-up tables are often lost if they are not stored with the data resource. Another common problem is the use of an alphanumeric code to represent the categories in the data (such as seabed habitats) where the code is used to thematically colour the polygons to create the map, but the legend only includes a text description of the categories. It can prove difficult to 'crack' the code to relate the text descriptions to their respective codes. To avoid such problems, always record any relevant descriptive information associated with features as attributes in the attribute table of the data resource.

Names of attributes in ArcGIS™ applications are limited to ten characters. Do not choose names for attributes, which are not unique or recognisable when they are truncated to ten characters, otherwise you may struggle to identify attributes. In addition to a name, an attribute also has a data type and field length. For example, an attribute of a habitat code may be a data type of 'Text' with a field length of '20'. Define attribute types and lengths correctly: often information is lost if fields are not sufficiently long to contain the information required in the record.

Although most modern database software (including ArcGIS™ and MS Access) is not case-sensitive to the characters you choose for the names of attributes, it is good practice to add attributes to the features in your data file in a standard way, for example always using upper case letters. Remember that other software or computer languages (particularly older software or software based on older principles) may be sensitive to case when they search your attribute table for data. For example, the HTML templates used by the [MESH webGIS](http://www.searchmesh.net/webGIS) (<http://www.searchmesh.net/webGIS>) to query the attribute tables of the map

shapefile are case-sensitive and an attribute name in the incorrect case will cause the online query to fail.

Topology

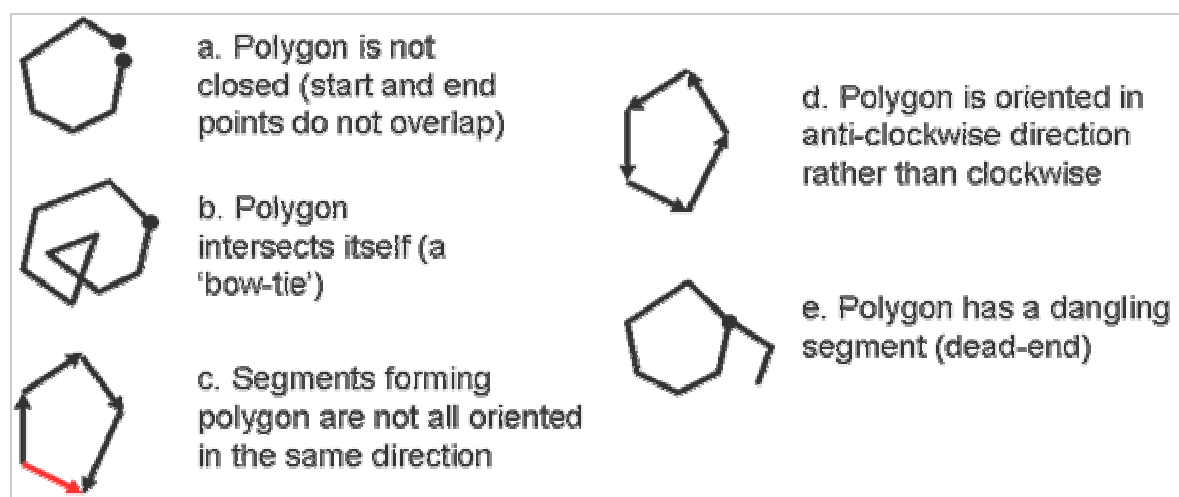
In addition to defining a coordinate system, recording metadata and adding attributes, there is one final aspect which needs to be considered when creating a vector data resource - topology. In GIS today, topology refers to the relationship between adjacent features. Whereas in vector data structures the topology between different units is explicitly recorded, in raster databases, this is only implicitly coded through the attribute values in the pixels. Topology may seem remote from seabed habitat mapping but it is essential to take topology into account when creating map data. Topological rules assume that geographic features occur on a two-dimensional plane. Spatial features are then denoted by nodes (0-dimensional cells), edges (1-dimensional cells), or polygons (2-dimensional cells).

Links to other sections:

[Tools to check and edit map data](#)

Tools to check and edit map data

The topological rules used by GIS packages mean that certain functions require vector data resources to contain topologically correct (simple) features for these functions to operate successfully. For example, the tools in the Geoprocessing Wizard of ArcGIS™ (dissolve, merge, union, intersect) may fail if the input shapefiles have features with topological errors (often known as non-simple features). In shapefiles generated from seabed habitat mapping data, common topological problems are: features oriented in an anti-clockwise direction rather than clockwise; 'bow-ties' caused by self-intersecting features; and, dangling segments within features. A set of simple diagrams illustrating non-simple features is shown.



Illustrations of non-simple features according to ESRI™ topological rules for features.

Therefore, if you are generating shapefiles that you plan to analyse or edit using spatially analytical tools, it is important to check their topology and simplify any non-simple feature where possible.

Time will obviously be saved if shapefiles are checked in batches. For this purpose, MESH has developed a suite of tools for use in ArcGIS™ that can process shapefiles in batches:

- *Document shapefiles*: identifies shapefiles containing non-simple features;
- *Simplify shapefiles*: simplifies shapefiles containing non-simple features, and;
- *Dissolve shapefiles*: dissolves a batch of shapefiles based on an attribute specified by the user (useful for making boundary polygons).

Sometimes the *Simplify shapefiles* tool is unable to simplify all the features in a shapefile. In this case there is an additional set of tools that process the shapefiles one at a time to help trace the error:

- *Find non-simple features*: finds and simplifies non-simple features in a shapefile and deletes features with an empty geometry;
- *Split multipart features*: splits multipart features into individual features which retain the attributes of the original multipart polygon;
- *Remove interior rings*: removes interior rings from features which is useful for removing artefact slivers resulting from union operations, and;
- *Re-order shapefile*: a useful tool for visualising data which helps prevent larger features obscuring smaller ones by drawing the largest polygon first.

The ArcGIS MESH Tools are available for anyone to use, but please note that neither the MESH Project nor JNCC offer any software support for these tools; they are used at your own risk. To date they have been used only in ArcGIS™ 8.2 and 8.3. Please keep back-up copies of all data before using the tools.

Map layout

Visualising data

Thematic maps that differentiate between features on the basis of their attribute information rely on colour and/or shading to emphasise the range of attribute values associated with the map features. Selection of colours and shades has an important bearing on the visual impact of the map, and how easily a user can interpret its message. Inappropriate use of colour and shading can ruin an otherwise excellent map! The following advice is based on our experience reviewing seabed habitat maps during the MESH Project. It is not intended to be a prescriptive or exhaustive list, rather just some simple 'dos' and 'do nots' to remember when deciding on a colour scheme to apply to a map of the seabed:

Remember to:

- Take account of how the colours of adjacent polygons will appear when they are shown on the map: are they possible to distinguish at the 'normal' viewing scale?
- If possible, check the appearance of colours on other computer screens or a variety of printers since devices often render the same colour slightly differently.

- Reflect on the relative sizes of polygons in your map. For example, in a map that incorporates mapping of both the intertidal and subtidal zones, the subtidal polygons are usually larger than the intertidal polygons. Larger polygons are better illustrated with pale colours to prevent them to visually dominate the map. Use bold colours to highlight the smaller intertidal areas of the map.
- Remember that a user will often view the map at different scales either broad scale (zoomed out) or fine scale (zoomed in). Electronic mapping systems can allow the user to zoom in and out freely and at very broad scales a colour scheme applied to a very detailed level of a seabed classification scheme might become impractical.
- Consider users with colour vision deficiency (colour 'blindness'), especially those who are unable to distinguish between red and green.
- Investigate how an existing colour scheme will look if it is applied to your map. This may add value to the map, as users accustomed to the current colour schemes can easily interpret it. MESH has designed a colour scheme for EUNIS (2004 version), which can be used to symbolise polygons attributed with EUNIS habitat types. Below level 3 the EUNIS habitat types are grouped so that habitat types can only be distinguished at level 3 or above; for example, A1.1.
- Think about how the majority of users will view the map – printed or on screen? If the main usage will be electronic in GIS packages, the colour scheme can be kept relatively simple because the standard query tools allow users to get more detail if required. If the map is to be printed, ensure that your chosen colour scheme will be adequate at the planned printing scale.
- Record your colour scheme and store it together with the file. For example, you could create an ArcGIS™ style if you have more than one map to which you wish to apply the colour scheme. Alternatively, for a colour scheme unique to a particular shapefile, you could create a layer file (.lyr) but ensure that this layer file always stays with the shapefile, particularly when passed to a third party! The component RGB values (Red, Green, Blue) of each map colour could even be noted in a text file if other options are not available, ensuring that other users can re-create the colours in another software environment (a different GIS or perhaps a graphics package for publication). It seems a waste to design a colour scheme for a publication and then lose the ability to reproduce it at a later date.

Remember not to:

- Create a colour scheme that is too complex to be interpreted. Although it is possible to use an almost infinite range of colours and patterns in GIS software, this does not necessarily make a clear map! Remember that it is not compulsory to apply a different colour or pattern to every feature having a different attribute. Grouping seabed habitats into similar types is a practical alternative. An often-used 'rule of thumb' restricts the number of different colour categories to approximately 15 because it is thought that is the maximum that can be easily distinguished by a user.

- Symbolise a map using attribute codes that lack a common meaning (e.g. integer values to represent habitat descriptions), and then write text to describe the habitats as part of a legend. Instead, add descriptive attributes to the vector data resource that can be interpreted by other users (e.g. habitat descriptions or codes from an accepted scheme). Then the colour scheme can be applied to any attribute, while the important information is stored as part of the data resource.
- Write lengthy text as part of a legend; it is extremely difficult to interpret the map if the legend is too complex, and the visual impact of the map may be compromised.
- Use boundaries for features which are too dark and consequently obscure the internal colour of the features; this is often a problem for small features.
- Design a colour scheme for a printed map without testing how the colours appear when printed; two colours appearing distinct on screen may look similar when printed (and vice versa).

Layout

Traditional elements of map layout (titles, text, legend, scale bar, logos) are sometimes omitted when creating electronic maps. However, they are crucial for a good map that is easy for users to interpret. Important things to consider are:

- A *north arrow* indicating the orientation of the map;
- A *scale bar* providing a visual indication of the size of features and distance between features on the map. The bar is usually divided into several parts and labelled with a map unit (e.g. km, metres);
- *Scale text* to represent the scale of your map rather than a scale bar. Scale text indicates the scale of the map and of features on the map (e.g. 1/10.000, saying that 1 cm on the map equals 10 kilometres on the ground).
- *Resolution* of the acoustic images
- *Coordinate system*

The table of contents often takes the place of a map legend in GIS. If it is present in a map, it should have clear, intuitive names for data frames, layer names, headings, and class names. The table of contents is also a good place to provide map readers with additional information.

Links to other sections:

[Tools to check and edit map data](#)

Links to websites:

<http://www.esri.com>

<http://www.mapinfo.com>

<http://www.searchmesh.net>