JNCC Report
No: 591

A three-step confidence assessment framework for classified seabed maps

Helen Lillis

July 2016

ISSN 0963-8901

For further information please contact:

Joint Nature Conservation Committee
Monkstone House
City Road
Peterborough PE1 1JY
jncc.defra.gov.uk

This report should be cited as:

Lillis, H. 2016. A three-step confidence assessment framework for classified seabed maps. *JNCC Report No. 591*. JNCC, Peterborough.

Summary

A simple and adaptable framework is presented for carrying out a qualitative confidence assessment for classified seabed maps, which is based around three criteria:

1. Remote sensing coverage.

2. Distinctness of class boundaries.

3. Amount of sampling.

It is termed a “framework” because of its ability to be adapted to the user’s needs; this includes applying it to a whole study, a single class within a study or individual polygons. The framework has been applied in several projects to date (e.g. Ellwood (2014), Ellwood and Duncan (2015) and Diesing *et al* (2015)) and examples of these applications are also presented.

The final section contains some guidance to aid the reader in developing his or her own three-step confidence assessment method based on this framework.

Contents

1 Background

The term “confidence” with regards to classified seabed maps (e.g. biotope/habitat/substrate) can have many meanings and therefore should be qualified whenever it is used. Confidence is a term sometimes applied to the accuracy[1]/uncertainty of the map based on external validation (testing the map with ground-truthing data that were not used in the mapmaking). This can be a very useful statistic but presents some challenges: data for validation are likely to be scarce, and the difference in spatial scales between a validation point (e.g. grab sample) and the map polygons means that mis-matches may be very common in spatially heterogeneous areas such as habitat mosaics.

The MESH confidence assessment (MESH Project 2008) delivers a confidence score that indicates the quality of the process used to make a biotope map and explains the relative reliability of different maps. However, because it refers to the mapping process as a whole, it does not give an indication of the probability (or likelihood) of a seabed class (e.g. biotope/habitat/substrate type) in a map being present at any location. Furthermore, although the MESH confidence assessment is straightforward, it takes some time to gain a full understanding of what it is actually showing.

As a compromise between these alternatives JNCC aimed to develop a simple, adaptable and transparent confidence assessment system that produces a qualitative score indicating the likelihood of a particular habitat being correctly mapped within a study area.

[1] Accuracy: the degree to which a measured value conforms to a true or accepted value. Accuracy is a measure of correctness.

2 The confidence assessment framework

Key features of this framework for assessing confidence in classified seabed maps are that it is:

- Simple – and therefore easy to apply and to understand.
- Adaptable – the specific application to a dataset may be altered to be relevant to the particular type of mapping method and classification used.

It is termed a "framework" because of its ability to be adapted to the user's needs; this includes applying it to a whole study, a single class within a study or individual polygons.

The framework is based around three criteria:

1. Remote sensing coverage – this is often the most important factor in accurately delineating the class boundaries. Remote sensing techniques include multi-beam or single beam echo sounder, side-scan sonar and aerial photography, among others.
2. Distinctness of class boundaries – this is a feature of the data and the particular habitats it surveyed, which are considered to have a large influence on the quality of the final map.
3. Amount of sampling – this is often the most important factor in accurately assigning the habitat type to each class. Sampling techniques include grab sampling, photos, videos, shore survey and diver observation, among others.

The confidence assessment framework can be represented by a simple three-step decision tree, in which the second and third steps can be made to depend on the answers to the previous questions, and the final score is a sum of the points awarded for each criterion (Figure 1, Table 1). The final score will range between 0 and 4 with 4 representing the 'best' type of map. Maps with equal scores are assumed to have roughly similar levels of confidence, regardless of the route through the decision tree. Note, however, that this is a qualitative assessment of the likelihood of a particular seabed class being correctly mapped within a study area; a score of 4 does not equate to a perfect or 100% accurate map.
A description of how each criterion may be assessed is given in **Table 2**. Some examples of how it has been applied are given in Section 3. Section 4 describes how to create a bespoke three-step confidence assessment based on this framework, which requires a clear explanation about how the criteria are to be assessed and a decision tree that display the possible combinations of points.

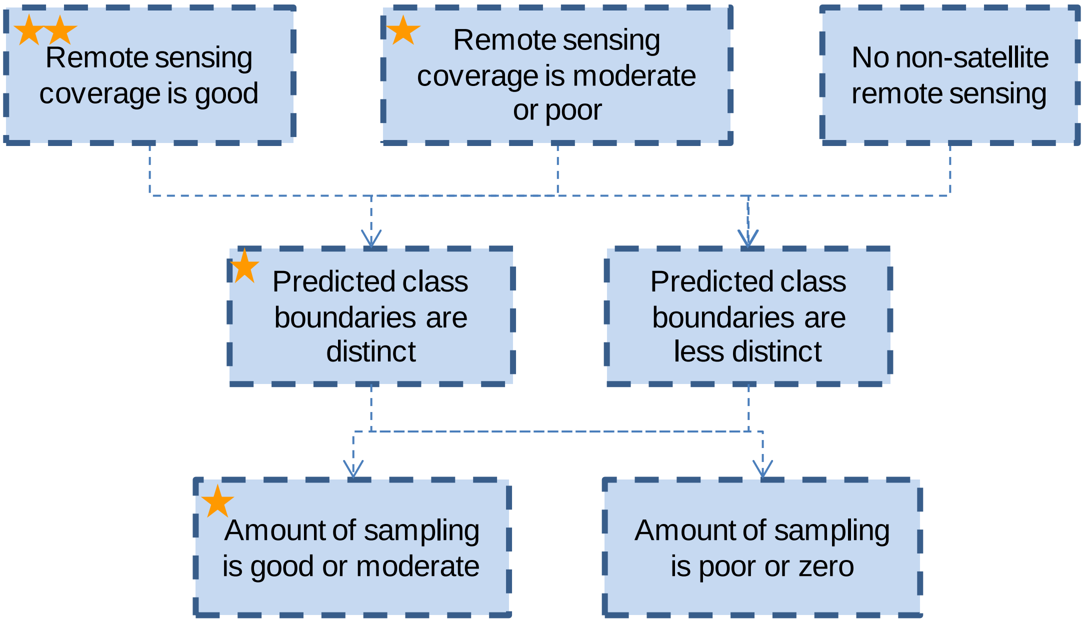

Figure 1: three-step decision tree; the assessor starts at the top and follows the arrows. Routes through the decision tree are displayed as dashed lines to indicate that these are potential routes, which may be edited or removed for certain situations. Stars/points are awarded according to the answers given and the final score is the sum of the stars/points.

Table 1: All combinations of scores that are possible under the three-step confidence assessment framework. Some combinations may not be possible under certain applications.

Remote sensing coverage	Distinctness of class boundaries	Amount of sampling	Total score
★★	★	★	4
★★	★		3
★★		★	
★	★	★	
★★			2
★	★		
★		★	
	★	★	
★			1
	★		
		★	
			0

Table 2: Further description of the generic criteria in a three-step confidence assessment.

Name of criterion	Application
Remote sensing coverage Possible scores: 0, 1 or 2	There is ambiguity about the amount of remote sensing that classes as "good" (2 points) and "moderate or poor" (1 point). This is to allow you to set appropriate thresholds for your own purposes, based on the seabed type of interest, the remote sensing technique used, homogeneity of the seabed, whether the survey was inter-tidal or sub-tidal and any other factors considered relevant. If you are unsure, a suggested rule of thumb is that over around 90% coverage may be classed as "good".
Distinctness of class boundaries Possible scores: 0 or 1	An example of when predicted class boundaries are distinct is when a map only contains rock and mud and the area has been surveyed with a multi-beam echo sounder. An example of when predicted class boundaries are less distinct is when the map contains subtidal coarse sediment, mixed sediment, and perhaps several different biotopes. Coarse sediment and mixed sediment are difficult to distinguish in acoustic remote sensing data, and biotopes are often impossible to distinguish. You may also vary how this criterion is assessed depending on the points scored for Remote sensing coverage, e.g. bypass this criterion if there is no remote sensing (a score of 0) (see EUNIS and OSPAR examples in Sections 3.1 and 3.2).
Amount of sampling Possible scores: 0 or 1	There is ambiguity about the amount of sampling that classes as "good or moderate" (1 point). This is allow you to set appropriate thresholds for your own purposes, based on the seabed type of interest, the sampling technique used, homogeneity of the seabed, whether the survey was inter-tidal or sub-tidal and any other factors considered relevant. If you are unsure, a suggested rule of thumb is that one sample per class or per polygon may be classed as "good or moderate". You may also vary how this criterion is assessed depending on the points scored for Remote sensing coverage, e.g. per class if there is remote sensing or per polygon if not (see EUNIS example in Section 3.1). It may be further varied depending on the points scored for Distinctness of class boundaries, e.g. per class if class boundaries are distinct or per polygon if not (see OSPAR example in Section 3.2).

3 Examples of applications

3.1 EUNIS level 3 habitat maps

JNCC periodically compiles a UK dataset for EUNIS level 3 habitats by combining polygon maps from individual surveys into a single layer. As part of this process, where two maps overlap with conflicting information, the three-step confidence score is used to determine which should be used (see Ellwood (2014) for full method).

The three-step confidence assessment is performed on a **per-survey basis**, with the second and third criteria depending on the answer to the first criterion (Figure 2, Table 3). The specific application of the three criteria is summarised in Table 4.

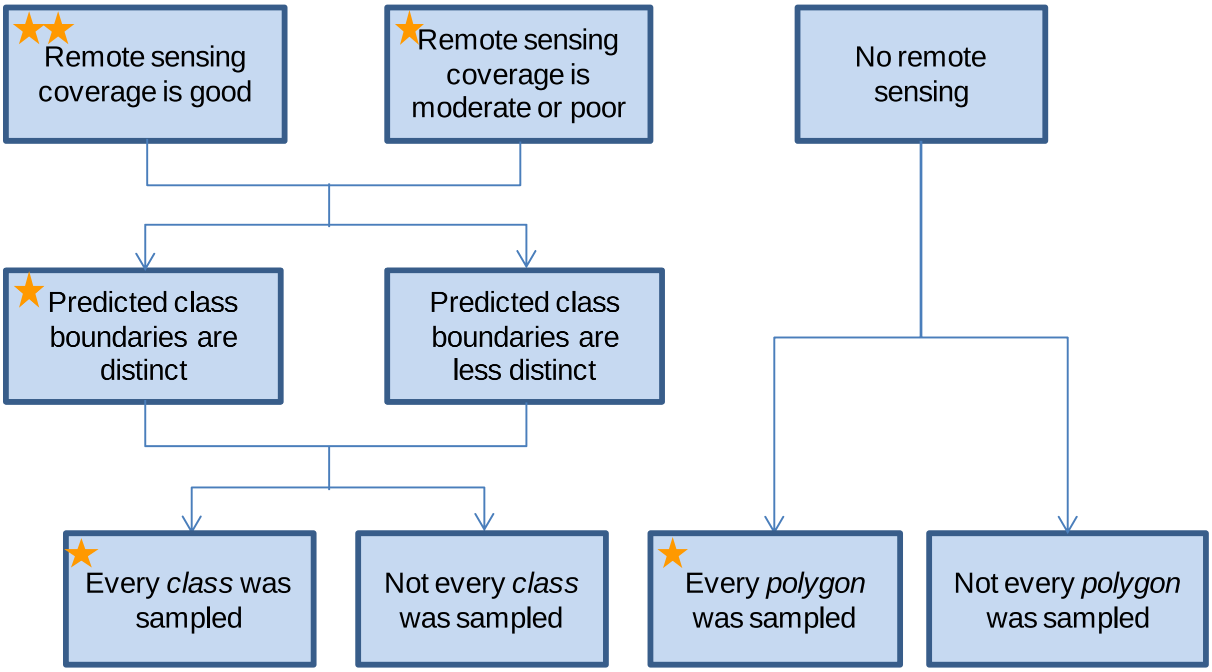

Figure 2: Confidence decision tree for EUNIS habitat mapping; the assessor starts at the top and follows the arrows.

Table 3: All combinations of scores that are possible under the three-step confidence assessment method for EUNIS level 3 habitat maps.

Remote sensing coverage	Distinctness of class boundaries	Amount of sampling	Total score
★★	★	★	4
★★	★		
★★		★	3
★	★	★	
★★			
★	★		2
★		★	
★			1
		★	
			0

Table 4: Application of the three-step confidence assessment method in scoring EUNIS level habitat maps.

Name of criterion	Description
Remote sensing coverage	How much of the study area is surveyed by remote sensing? • 2 points: coverage is good – a rule of thumb may be around 90 %; but expert judgement may be used • 1 point: coverage is moderate or poor • 0 points: no remote sensing used
Distinctness of class boundaries	This question is only answered if there is remote sensing data (i.e. if Remote sensing coverage scores 1 or 2). How easy is it to distinguish the classes in the remote sensing data and the boundaries between the classes? • 1 point: most classes are distinct in the remote sensing data • 0 points: some of the classes are difficult to distinguish in the remote sensing data.
Amount of sampling	Was there an adequate amount of sampling to identify every polygon? If there is any remote sensing data (i.e. if Remote sensing coverage scores 1 or 2): • 1 point: every/almost every class in the map was sampled. • 0 points: not every class in the map was sampled. If there was no remote sensing (i.e. if question one scores 0): • 1 point: every/almost every polygon in the map was sampled. • 0 points: not every polygon in the map was sampled.

3.2 OSPAR threatened and/or declining habitat maps

JNCC annually compiles a UK dataset for OSPAR habitats by combining polygon maps from individual surveys into a single layer. As part of this process, where two maps overlap with conflicting information, the three-step confidence score is used to determine which should be used (see Ellwood and Duncan (2015) for full method).

The three-step confidence assessment is performed on a **per-habitat**, **per-survey basis**, with the second and third criteria depending on the answer to the first criterion (Figure 3). The possible combinations of scores under this method is the same as for EUNIS level 3 habitat maps (Table 3). The specific application of the three criteria is summarised in Table 5.

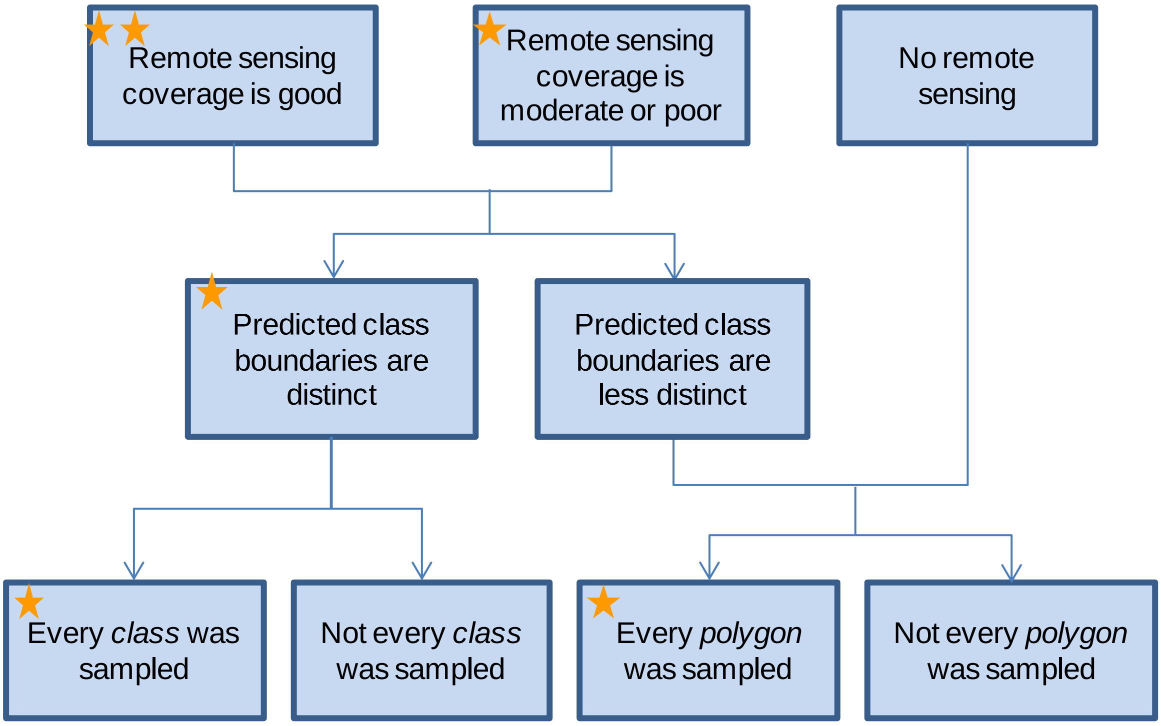

Figure 3: Confidence decision tree for OSPAR habitat mapping; the assessor starts at the top and follows the arrows.

Table 5: Application of the three-step confidence assessment method in scoring OSPAR habitat maps.

Name of criterion	Description
Remote sensing coverage	How much of the mapped habitat within the study area is surveyed by remote sensing? • 2 points: coverage is good – a rule of thumb may be around 90 %; but expert judgement may be used. • 1 point: coverage is moderate or poor • 0 points: no remote sensing used

Distinctness of class boundaries	This question is specific to the habitat and is only answered if there is remote sensing data (i.e. if Remote sensing coverage scores 1 or 2).
	
	• 1 point: it is possible to distinguish the habitat in remote sensing data. As a rule-of-thumb, this can be applied to:
	    o Littoral chalk communities
	    o Intertidal Mytilus edulis beds on mixed and sandy sediments
	    o Intertidal mudflats
	    o Zostera beds
	    o Ostrea edulis beds
	    o Modiolus modiolus horse mussel beds
	    o Sabellaria spinulosa reefs
	    o Maerl beds
	    o Lophelia pertusa reefs
	    o Carbonate mounds
	    o Seamounts
	
	• 0 points: the habitat is not usually possible to detect in remote sensing data. As a rule-of-thumb, this can be applied to:
	    o Sea-pen and burrowing megafauna communities
	    o Coral gardens
	    o Deep-sea sponge aggregations

Amount of sampling	Was there an adequate amount of sampling to identify every polygon of the habitat in the survey? If the habitat is distinguishable in remote sensing data (i.e. if Distinctness of class boundaries scores 1) and if there is any remote sensing data (i.e. if Remote sensing coverage scores 1 or 2): • 1 point: every/almost every habitat type in the map was sampled. • 0 points: not every habitat type in the map was sampled. If the habitat is not distinguishable in remote sensing data (i.e. if Distinctness of class boundaries scores 0) and/or if there was no remote sensing (i.e. if Remote sensing coverage scores 0): • 1 point: every/almost every polygon in the map was sampled. • 0 points: not every polygon in the map was sampled.

3.3 Broad-scale hard substrate mapping

This project produced a map of hard substrate for the Channel and Celtic Sea based on a one arc second digital elevation model, derived layers, other environmental models and sample points. The mapping was carried out in two stages (see Diesing *et al* (2015) for full method):

1. automated spatial prediction of the presence and absence of rock at the seabed using a random forest ensemble model, and
2. manual editing of the model outputs based on ancillary geological data and expert knowledge.

The three-step confidence assessment was performed on a **per-polygon basis** due to the fact that the data used were not separated by survey and came from various sources (Figure 4). All combinations of scores are possible under this method, as shown in Table 1. The specific application of the three criteria is explained in Table 6.

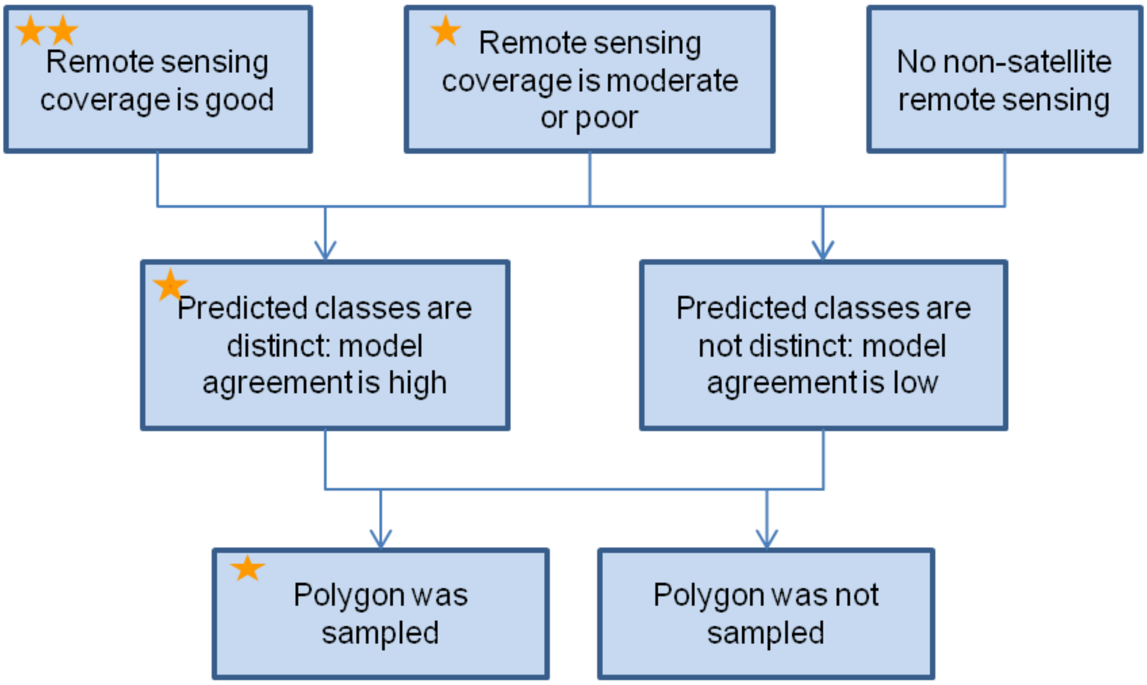

Figure 4: Confidence decision tree for broad-scale hard substrate mapping; the assessor starts at the top and follows the arrows.

Table 6: Application of the three-step confidence assessment method in scoring semi-automated rock mapping outputs.

Name of criterion	Details of application
Remote sensing coverage	The Defra digital elevation model (DEM) has an accompanying confidence layer, which scores each pixel from 0 to 9 depending on the types and provenance of the data input to form the model (described in Astrium Oceanwise (2012)). These scores were averaged within each polygons and the following correspondence was applied: • 2 points: Defra DEM score = 7, 8 or 9 (implying multibeam-derived DEM values). • 1 point: Defra DEM score = 4, 5 or 6 (implying singlebeam-derived DEM values). • 0 points: Defra DEM score = 0, 1, 2 or 3 (implying electronic navigational charts or GEBCO-derived DEM values).
Distinctness of class boundaries	This was primarily assessed using the levels of agreement between model iterations in the automated modelling phase. • 1 point: >75% of model iterations predicting presence or absence of rock and/or if polygon was manually modified or added based on expert judgement and/or ancillary data. • 0 points: 25-75% of model iterations predicting presence or absence of rock, with no manual modification.
Amount of sampling	• 1 point: the polygon was sampled and the majority of samples agree with the classification. • 0 points: the polygon was not sampled or the majority of samples within the polygon disagree with the classification.

4 How to tailor the framework for your needs

In order to use a three-step confidence assessment system based upon this framework you will need to consider the following questions:

1. What types of seabed classes are being mapped? And therefore:
    a. What amount of remote sensing coverage would be considered good or moderate?
    b. What amount of sampling would be considered good or moderate?
    c. Which of these classes could be categorised as ‘distinct’ or ‘less distinct’?

2. Does the way certain criteria are assessed depend on the answers to other criteria?

3. How is the classified seabed map being produced, and therefore what is a feasible, practical way of assessing the three criteria?

Before carrying out the assessment it is recommended that you:

1. Create a bespoke decision tree, with the arrows positioned as appropriate to link the questions in suitable way (e.g. Figure 2, Figure 3, Figure 4).

2. Write down exactly how each criterion will be assessed, with some justification (e.g. Table 4, Table 5, Table 6).

5 References

ASTRIUM OCEANWISE. 2012. Creation of a high resolution digital elevation model (DEM) of the British Isles continental shelf v2.0. *Ref: contract 13820. Date: 05/01/2012.Commercial in confidence report*.

DIESING, M., GREEN, S.L., STEPHENS, D., COOPER, R. & MELLETT, C.L. 2015. Semi-automated mapping of rock in the English Channel and Celtic Sea. May 2015. *JNCC Report No. 569*. JNCC, Peterborough. Available online at http://jncc.defra.gov.uk/page-7074

ELLWOOD, H. 2014. Creating a EUNIS level 3 seabed habitat map integrating data originating from maps from field surveys and the EUSeaMap model v1.0. *Date: 11/03/2014.* Available online at http://jncc.defra.gov.uk/page-6655#EUNIScombined.

ELLWOOD, H. & DUNCAN, G. 2015. Creating a composite OSPAR threatened and/or declining habitat map for the UK v1.0. *Date: 10/06/2015.* Available online at http://jncc.defra.gov.uk/page-1583.

MESH PROJECT. 2008. “MESH Confidence Assessment”. Online resource available at http://www.emodnet-seabedhabitats.eu/default.aspx?page=1635

Appendix 1: development of the three-step confidence assessment framework

The three-step confidence assessment was originally developed to aid the production of a full-coverage map EUNIS level 3 layer integrating data from surveys and broad-scale models. Ellwood (2014) describes the approach and justification in full.

The three criteria were selected by considering each of the 15 MESH confidence criteria (see MESH Project 2008) together with other factors affecting map quality and choosing those likely to have the greatest effect on the overall accuracy of the habitat assignments. *Remote sensing coverage* and *amount of sampling* are similar to the MESH criteria *remote sensing coverage* and *ground truthing density*. The *distinctness of class boundaries* criterion is not solely based on the techniques used to make the map and therefore does not have an equivalent in the MESH confidence assessment.