



EMODnet



European Marine
Observation and
Data Network

EMODnet Thematic Lot n°3 – Seabed Habitats

EASME/EMFF/2020/3.1.11/Lot3/SI2.843624

Start date of the project: 25/09/2021 - (24 months)

Centralisation Phase

Constructing EUSeaMap

User Guide

Version 2





Disclaimer

The information and views set out in this report are those of the author(s) and do not necessarily reflect the official opinion of the CINEA or of the European Commission. Neither the CINEA, nor the European Commission, guarantee the accuracy of the data included in this study. Neither the CINEA, the European Commission nor any person acting on the CINEA's or on the European Commission's behalf may be held responsible for the use which may be made of the information.

Document info

Title	Constructing EUSeaMap – User Guide
Deliverable	D1.16
Author	Mickaël Vasquez (Ifremer). Mickael.Vasquez@ifremer.fr
Dissemination level	Public
Version date	24/09/2023
Version	2
Version updates	Major update of the text

Contents

1 Introduction.....	9
2 Vocabulary and concepts.....	9
3 The EUSeaMap GIS Workflow	10
3.1 Creation of a habitat descriptor layer and its confidence layer	12
3.1.1 Workflow	12
3.1.2 Step 1: modelling the habitat descriptor classes and assessing their confidence	13
3.1.3 Step 2: merging the habitat descriptor classes in a single layer	17
3.2 Converting the habitat descriptor raster layers in polygon layers	17
3.3 Creating the habitat map	17
3.3.1 Combining the habitat descriptor layers.....	18
3.3.2 Joining the look-up table	19
4 The R scripts & ArcGIS™ model	20
4.1 Prerequisites	21
4.1.1 System requirements	21
4.1.2 Software Requirements.....	21
4.2 General principles.....	21
4.2.1 A critical requirement: raster inputs must all be spatially consistent.....	21
4.2.2 Formulation of the R script inputs via configuration files	21
4.2.3 A strict folder structure	22
4.2.4 Changing the maximum amount of memory used by terra	22
4.3 Scripts for the production of habitat descriptor layers	23
4.3.1 Modelling the habitat descriptor classes	23
4.3.2 Merging the individual habitat descriptor classes in a single layer.....	27
4.4 Creating the habitat layer and overall confidence	29
4.4.1 Combining the habitat descriptor layers.....	29
4.4.2 Joining the habitat polygon layer and the look-up table that crosswalks the modelled habitat codes and habitat classifications	31
4.4.3 Creating the habitat confidence layers	32
4.5 Additional script: combining the wave-induced and current-induced habitat descriptor layers.....	33
5 References	35
6 Appendix 1: the habitat descriptors used for EUSeaMap (v2019), their classes and codes.....	37

7 Appendix 2: calculating the slope, intercept and probability threshold for fuzzy laws.....	41
8 Appendix 3: Hands-on training	43

Executive summary

In the first two phases of EMODnet, the Seabed Habitats thematic project used a commercial software, namely ARCGIS™, to produce the pan-European broad-scale seabed habitat map, EUSeaMap. A key objective of Phase 3 (2017-2021) was to develop a new GIS workflow that is robust, repeatable and transferable across marine regions, and to implement this workflow in tools based on open source technologies. As a result, most of the workflow was implemented in the form of R scripts. For technical reasons, a small part had to be implemented in the form as an ArcGIS™ ModelBuilder model. These tools were used to build versions 2019, 2021 and 2023 of EUSeaMap.

This technical guide is intended to help the reader learn how to use these tools. The vocabulary and concepts used in the document are defined and the GIS workflow is described. The purpose of the scripts is documented in detail, as well as what they require as input and what they produce as output. Step-by-step hands-on training is provided. However, it is important to note that this document is intended for advanced users, i.e. that a good knowledge of the methods used for EUSeaMap, which are described in other documents, is an essential prerequisite for understanding the document and the associated scripts.

1 Introduction

This technical guide aims to help the reader to use the R scripts and the model created using the ArcGIS™ ModelBuilder application developed by EMODnet Seabed Habitats¹ to produce the broad-scale seabed habitat map (referred to as "EUSeamap" in the report) in a semi-automated manner.

The general methods used are largely based on a methodology that was originally developed by the MESH Project, the general principles of which are described in Coltman et al. (2008). It is strongly recommended the reader read this short document before going through this one.

The way in which some concepts such as fuzzy laws or Generalized Linear Models (GLMs) are used in the context of EUSeaMap, a thorough understanding of which is a prerequisite for the understanding of the present document and related scripts, will not be developed here. Further explanation can be found in the EUSeaMap Phase 2 Technical Report Populus et al, 2017. In particular, it is particularly recommended to read sections 2.5 to 2.7 and appendices 9 (Thresholds), 10 (Confidence) and 12.

Before describing the scripts, i.e. what they do and what their inputs and outputs are, the vocabulary and concepts used in the document are defined and the GIS workflow is described.

2 Vocabulary and concepts

A habitat is a combination of several environmental features. In habitat classifications these environmental features are formulated by what we will refer to in this document as '**habitat descriptors**', which in the seabed section of the European classification EUNIS are at least seabed substrate types and biological zones, and in some cases other seabed features such as the energy levels induced by water movements. For example, in the Atlantic EUNIS version 2007-11 describes habitats using three habitat descriptors: energy levels, biological zones and seabed substrate types. The habitat descriptors have their own classification, e.g. in the Atlantic the habitat descriptor classification for the energy levels has 3 classes: high, moderate and low. In this document we will refer to these classes as '**habitat descriptor class**'.

The following habitat descriptors are used in EUSeaMap:

- Seabed Substrate types
- Biological Zones
- Wave-induced Energy Levels
- Current-induced Energy Levels
- Oxygen Regimes
- Salinity Levels
- Masks

Seabed substrate types and biological zones are used in all marine regions. A third habitat descriptor is then usually used, depending on the region. For example, in the Atlantic combined wave and current induced energy levels are used as a third habitat descriptor, while in the Black Sea oxygen regimes are used. Further details on which habitat descriptors are used in each region can be found in Populus et al. (2017).

All the habitat descriptor classes are described in Annex 1, together with the numerical code assigned to them in the EUSeaMap GIS workflow that produces the broad-scale habitat map.

A habitat descriptor class typically has two boundaries: an upper and a lower boundary. The upper boundary is the boundary that the class shares with its upper adjacent class in the habitat descriptor classification, and the lower boundary is the boundary that the class shares with its lower adjacent class. For example, the upper

¹ <http://www.emodnet-seabedhabitats.eu/>

boundary of the Circalittoral Biological Zone is the boundary it shares with the Infralittoral, while its lower boundary is the boundary it shares with the Offshore Circalittoral.

Some habitat descriptor classes have only one boundary because they are at the top or the bottom of their classification and therefore have only one adjacent class. For example, the Infralittoral Biological Zone, the Oxidic Oxygen Regime, or the High Energy Level only have a lower boundary (i.e. no upper boundary) because these classes are at the top level of the Biological Zone, Oxygen Regime and Energy Level habitat descriptor classifications respectively. Similarly, the Abyssal Biological Zone, Anoxic Oxygen Regime or Low Energy Level have only one boundary, the upper one, because they are at the bottom of their respective classifications.

3 The EUSeaMap GIS Workflow

The general workflow consists of 4 steps:

1. The habitat descriptor layers are created individually (together with their respective confidence layer).
2. These habitat descriptor raster layers are converted into polygon layers.
3. The seabed habitat layer is created by overlaying all polygon habitat descriptor layers. In parallel, the individual habitat descriptor confidence layers are overlaid to produce an overall confidence layer.
4. The seabed habitat layer is then joined to the lookup table that crosswalks the modelled numerical habitat codes and the classes of habitat classifications such as EUNIS or the MSFD broad habitat type classification².

In EUSeaMap, some habitat descriptor layers are provided 'out-of-the-box', i.e. they don't need to be created by the EUSeaMap workflow because they are created externally. An example of this in all marine regions is the habitat descriptor 'Seabed Substrate', which is created and provided directly by EMODnet Geology.

Figure 1 illustrates the workflow with an example where four habitat descriptors are considered, three of which would be produced by the workflow (HD2, 3 and 4) and one of which (HD1) would be externally created.

Sections 3.1, 3.2, 3.3.1 and 3.3.2 describe each of the three steps above.

² COMMISSION DECISION (EU) 2017/848

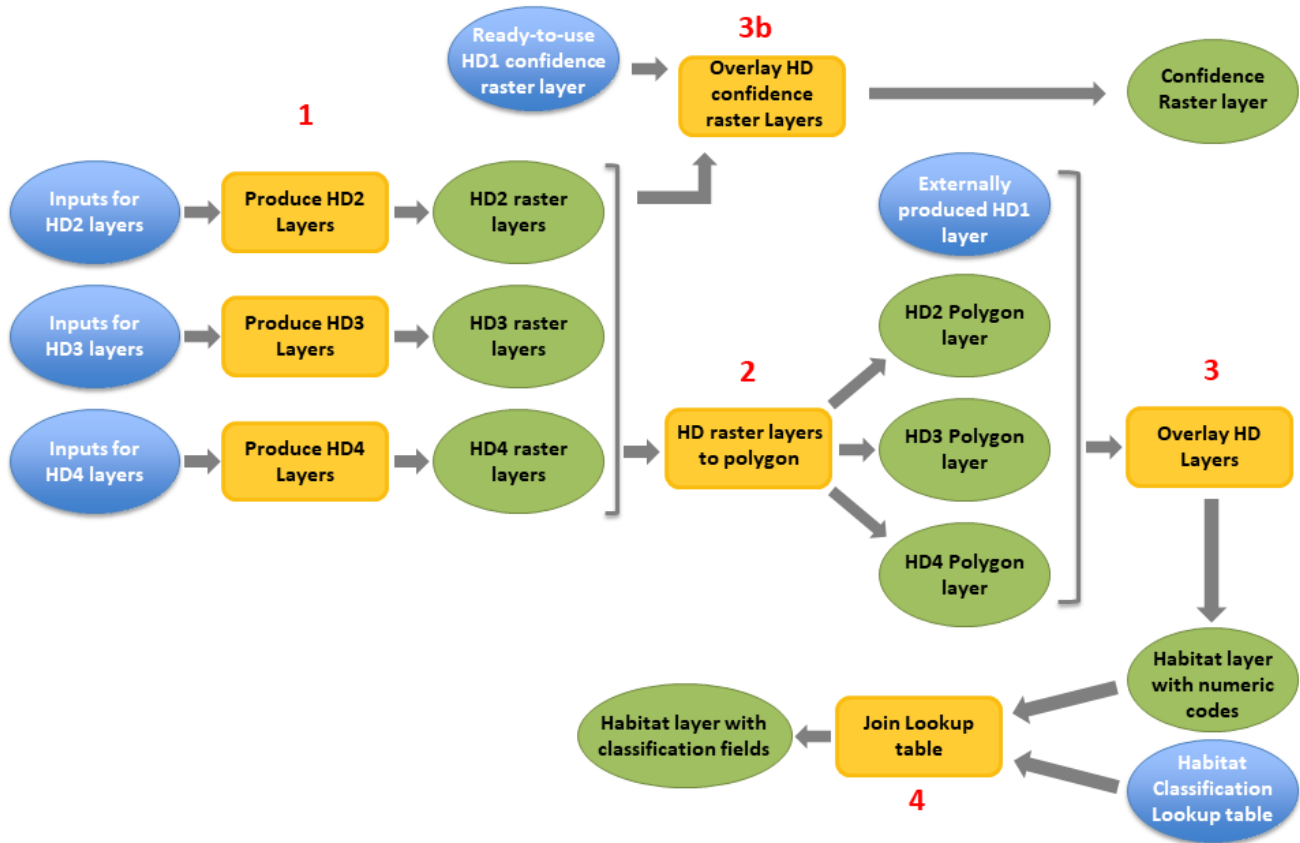


Figure 1: The general EUSeaMap GIS workflow illustrated with an example of 4 habitat descriptors (HD). Layers are produced for HD2, HD3 and HD4, while the HD1 layer is provided 'out-of-the-box' (as it is produced externally). The HD2, HD3 and HD4 layers are 1) produced as rasters (along with a confidence layer for each), 2) converted into polygons, 3) overlaid to produce the habitat layer. 3b: The HD confidence raster layers are also overlaid into a single overall confidence raster. After overlaying the HD layers, the lookup between numeric habitat codes and different classification classes is joined to the habitat layer (4).

3.1 Creation of a habitat descriptor layer and its confidence layer

3.1.1 Workflow

The production of a habitat descriptor layer and its confidence layer consists of 2 steps (Figure 2):

1. The habitat descriptor classes are modelled using spatial distribution laws. As a result, 2 raster layers are produced for each class, namely i) a spatial distribution layer and ii) a layer on the confidence in the spatial distribution.
2. All the individual class spatial distribution and confidence layers modelled in step 1 are merged into a single layer.

Sections 3.1.2 and 3.1.3 describe these two steps in more detail.

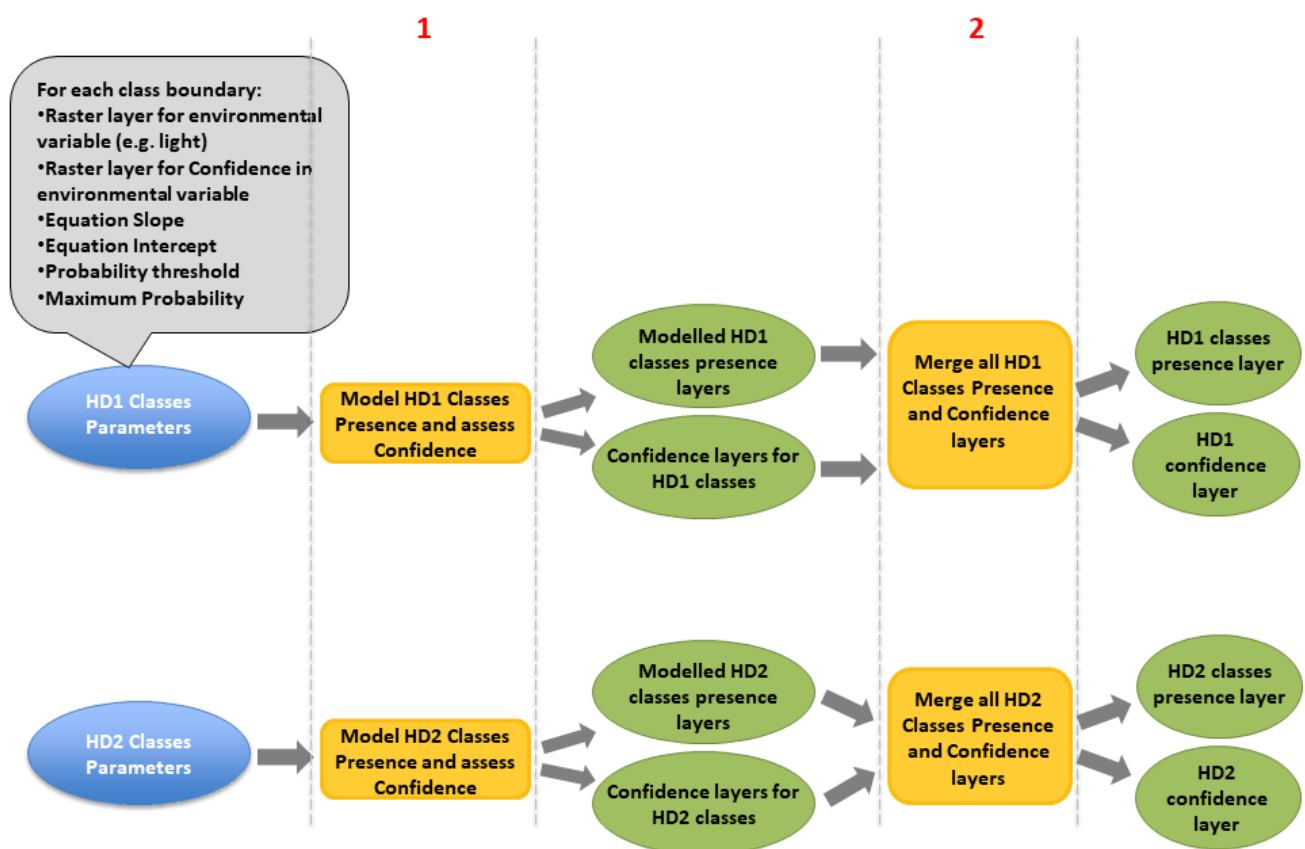


Figure 2: Production of Habitat Descriptor (HD) layers for 2 HDs, HD1 and HD2. For each HD class, presence layers are modelled and confidence is assessed (hence two output layers are produced, one for class presence and one for confidence in class presence); 2) individual HD class presence and confidence layers are merged into a single layer.

3.1.2 Step 1: modelling the habitat descriptor classes and assessing their confidence

Modelling the habitat descriptor classes: the use of spatial distribution GLM/fuzzy laws

The presence of each individual habitat descriptor class is modelled using either Generalized Linear Models (GLMs) or fuzzy laws. Details of how these GLMs and fuzzy laws are fitted in the approach developed for EUSeaMap can be found in Populus et al. (2017). In the EUSeaMap approach, GLM/fuzzy laws are used to predict the probability of the presence of a habitat descriptor class, given the value of a unique environmental predictor variable (e.g. the amount of light available at the seabed is the unique predictor variable for the presence of the infralittoral).

Here we describe the set of parameters that need to be provided to step 1 of the above workflow (see Figure 2) in order to compute a probability and presence raster using a GLM/fuzzy law.

Both GLMs (Figure 3A) and fuzzy laws (Figure 3B) allow the calculation of the probability of presence of a habitat descriptor class as a function of a predictor value.

The equation has the form:

$$P(X) = \frac{e^{ax+b}}{1+e^{ax+b}}$$
 for a GLM,

$$P(X) = ax+b$$
 for a fuzzy law

where X is the predictor value, P(X) is the probability of occurrence of the habitat descriptor class, and a and b are the slope and the intercept of the GLM or the fuzzy function, respectively.

To model the presence of a habitat descriptor class, a GLM/fuzzy law is required for each boundary of the class, i.e. the upper boundary (if any) and the lower boundary (if any).

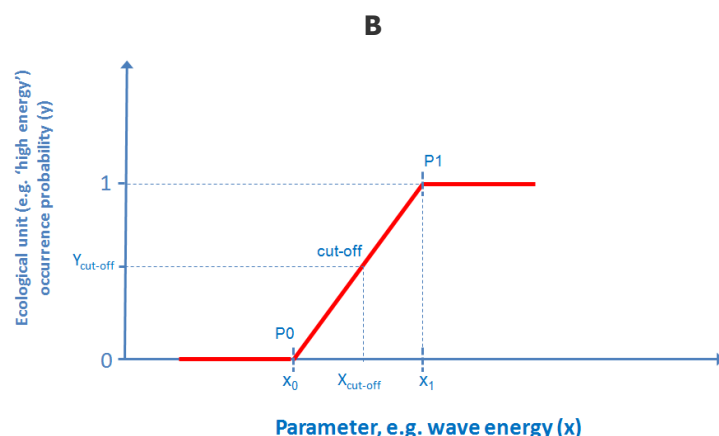
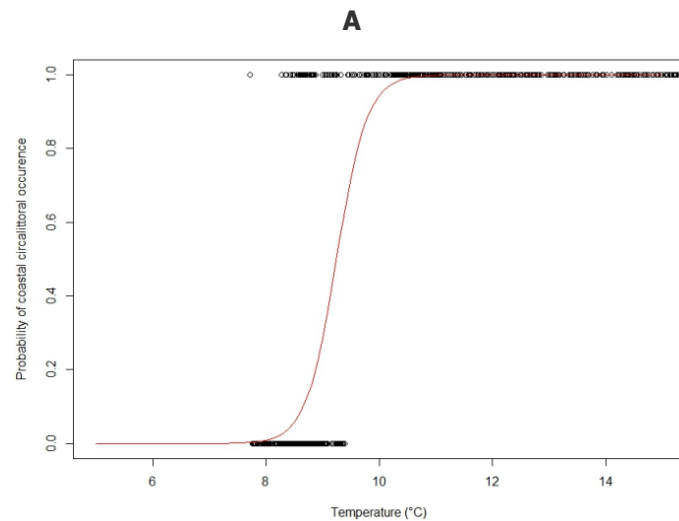


Figure 3: The 2 types of laws used to model the habitat descriptor class spatial distribution: GLM (A) and fuzzy (B) – In Populus et al, 2017

For each GLM/fuzzy law a probability cut-off (also called a threshold) is also provided so that any predicted probability value of presence can be converted to a binary value (present or absent): above the probability threshold, the habitat descriptor class is characterised as present, below it as absent. The approaches used to define these thresholds are described in Populus et al (2017).

In summary, for a GLM/fuzzy law to be used in the workflow, 3 values are required for the 2 boundaries (or the boundary if there is only one) of each habitat descriptor class: a **slope**, an **intercept** and a **probability threshold**.

Note: For background information, Annex 2 describes how to calculate the slope, intercept and probability threshold from the fuzzy law control points.

Assessing the confidence in the habitat descriptor classes

As described in Populus et al. (2017), the confidence in assigning the correct habitat descriptor class to a raster cell depends on two items:

1) The confidence **in the environmental layer values** used for each habitat descriptor class boundary: this is an input to the workflow, provided in the form of a raster layer where cell values are a categorical measure of the confidence (1=low, 2=moderate, 3=high). Some guidance on the provision of this layer can be found in Populus et al. (2017), section 2.7. Note that EMODnet Geology provides a confidence index for its polygon Seabed Substrate layer, with confidence assigned at the polygon level. Similarly, EMODnet Bathymetry provides a confidence index for its DTM, with confidence assigned at the raster cell level.

The confidence **based on the probability** predicted by the GLM/fuzzy law. For each raster cell, the workflow classifies the calculated probability (a continuous value in the range [0-1]) into one of the following categorical confidence values: 1 (for low), 2 (for moderate) or 3 (for high). The rules used to classify a probability value into a categorical confidence are described in Table 1.

Table 1: Rules used to classify a continuous probability as high, moderate or low confidence. 'probability max' is the maximum value of the probability considering the GLM/fuzzy law (typically 1, but in some cases may be less); 'range'= probability max - probability threshold

Confidence per cell	Rule
High	$\text{probability threshold} + (0.6 \times \text{range}) \leq \text{probability} \leq \text{probability max}$
Moderate	$\text{probability threshold} + (0.2 \times \text{range}) \leq \text{probability} < \text{probability threshold} + (0.6 \times \text{range})$
Low	$\text{probability threshold} \leq \text{probability} < \text{probability threshold} + (0.2 \times \text{range})$

For example, for a GLM/fuzzy law with a maximum probability of 1, and a probability threshold of 0.5, the rules would be as described in Table 2.

Table 2: Example of rules used to classify a continuous probability as high, moderate or low confidence. In this example probability max is 1 and probability threshold is 0.5.

Confidence per cell	Rule
High	$0.8 \leq \text{probability} \leq 1$
Moderate	$0.6 \leq \text{probability} < 0.8$
Low	$0.5 \leq \text{probability} < 0.6$

For each habitat descriptor class boundary, the workflow automatically creates a confidence raster layer based on probability, which is done by classifying the probability as high, moderate or low confidence. For this to be possible, the user needs to provide the maximum probability as input to the script, in addition to the 3 values mentioned above (i.e. slope, intercept and probability threshold).

In a final step, the workflow creates a single confidence raster layer for each habitat descriptor class boundary by combining the confidence raster layer for the environmental variable values and the confidence raster layer based on the probability. The logic used for the combination is described in Table 3.

Table 3: logic used to combine the two types of confidence

		Confidence in values of the environmental variable		
		H	M	L
Confidence based on the probability	H	H	H	M
	M	M	M	L
	L	L	L	L

GIS Workfkw

The workflow for this step (figure 4) requires the following input parameters **for each class boundary**:

- The raster layer of the environmental predictor variable (e.g. seabed PAR)
- The slope of the GLM/fuzzy law
- The intercept of the GLM/fuzzy law
- The probability threshold
- The maximum probability

It produces for each habitat descriptor class:

- A categorical raster layer for the occurrence of the class. In this layer all the cells have the same value, which is assigned according to the habitat descriptor class coding convention in Annex 1.
- A categorical raster layer for confidence in the occurrence of the habitat descriptor class. The values of the cell are one of the following: 1 (low confidence), 2 (moderate confidence) and 3 (high confidence)

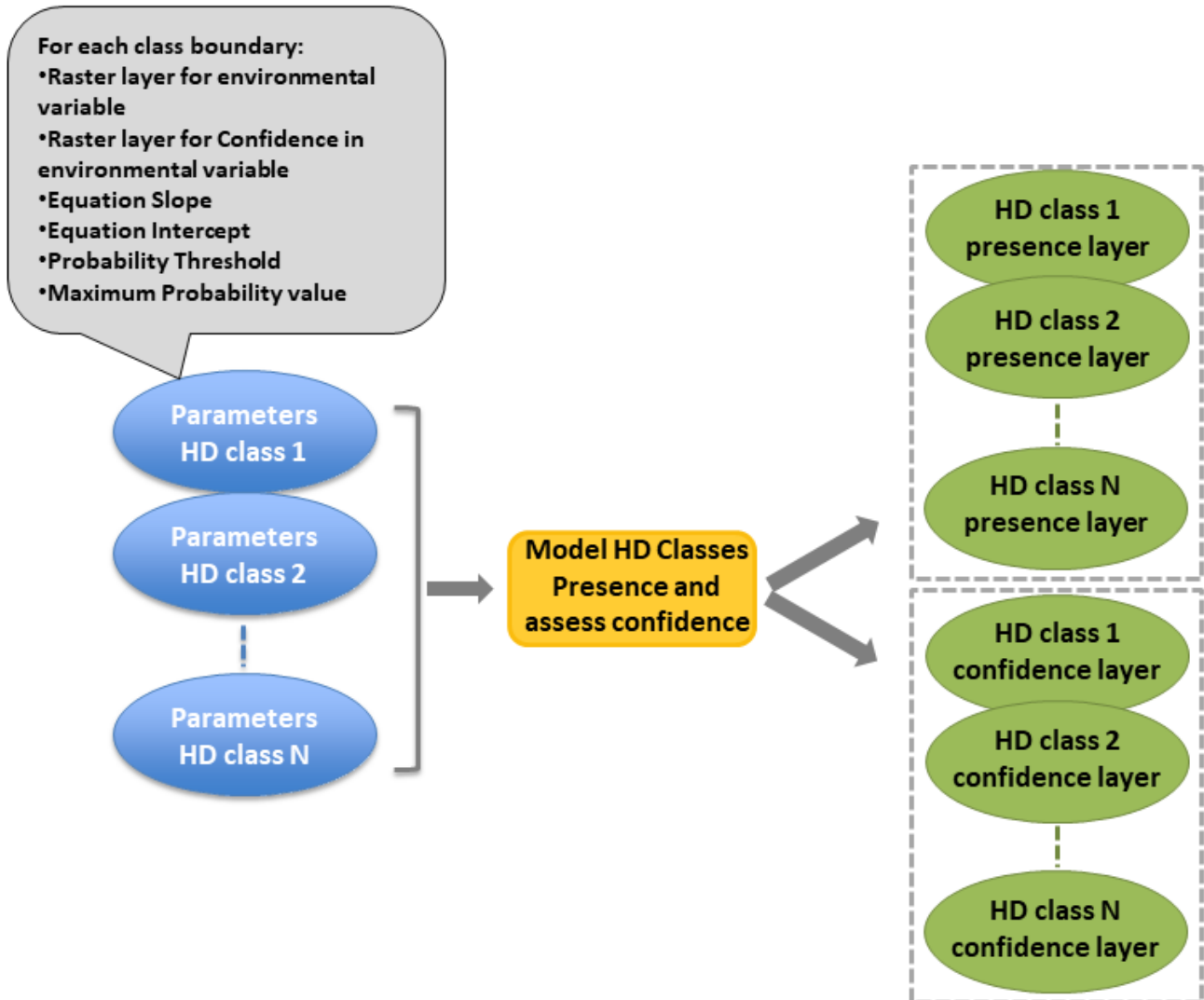


Figure 4: Workflow for the production of the individual class presence and confidence layers for a Habitat Descriptor (HD). For each class boundary, the inputs are the environmental variable raster and its confidence raster, the slope and intercept of the GLM/fuzzy law, the probability threshold and the maximum value that the probability can take considering the GLM/fuzzy law. For each class, 2 raster layers are created: a presence layer and a confidence layer.

An option: providing the slope, intercept, and probability threshold as a raster instead of a constant value

Within the same region, the GLM/fuzzy law slope and intercept, and the probability threshold defining a habitat descriptor class boundary may vary spatially.

An example of this is shown in Figure 5. In the Black Sea, the slope and intercept values that define the infralittoral lower boundary are not the same everywhere. In this particular case, where the GLM/fuzzy law parameters vary spatially, instead of a constant value for slope and intercept, the workflow requires a raster layer as input for both.

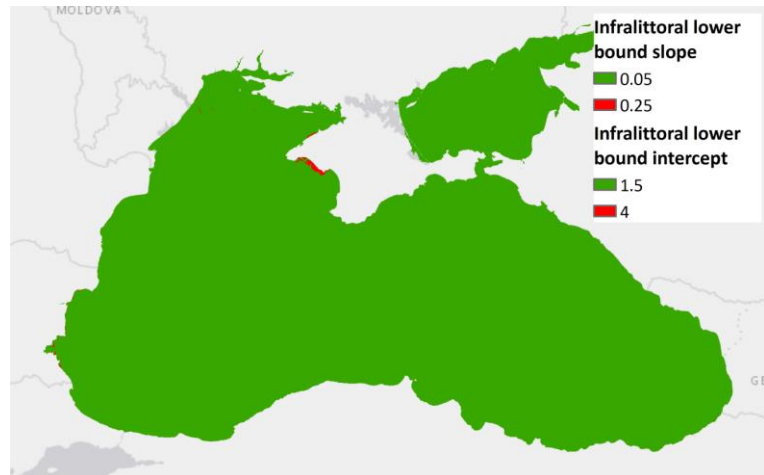


Figure 5: Example of spatially varying GLM/fuzzy slope and intercept for a habitat descriptor class boundary: the infralittoral lower boundary in the Black Sea.

3.1.3 Step 2: merging the habitat descriptor classes in a single layer

As shown in Figure 2, step 2 of the workflow merges all the habitat descriptor class layers and the confidence layers produced in step 1. The output is a single habitat descriptor raster layer containing all the classes (e.g. all biological zones) and a single habitat descriptor confidence raster layer.

3.2 Converting the habitat descriptor raster layers in polygon layers

The outputs of step 1 of the general workflow (see Figure 1) are habitat descriptor raster layers. As the creation of the habitat layer (Figure 1, step 3) involves the overlay of the habitat descriptors in vector mode, all the habitat descriptor layers need to be converted to polygon layers (Figure 1, step 2).

3.3 Creating the habitat map

The final step in the general workflow is the creation of the habitat map. In a first step the habitat descriptor layers are all combined by a geometric intersection to create the habitat layer (see Figure 1, step 3). In a second step, to the attribute table of this habitat layer is joined the lookup table that crosswalks the habitat codes resulting from the union and the classes of habitat classifications such as EUNIS or the MSFD Broad Benthic Habitat Types (see Figure 1, number 4). These two steps are described below.

3.3.1 Combining the habitat descriptor layers

General principle

All the habitat descriptor layers are combined by a geometric intersection. This results in a habitat polygon layer, the attribute table of which contains one column per habitat descriptor layer. For example, in the Mediterranean, where 3 habitat descriptors are considered, namely seabed substrate types, biological zones, and mask values, the habitat layer will contain 3 columns.

For each polygon, the values contained in these columns are summed in a new column. This results in a code that is representative of a unique combination of the habitat descriptor classes considered in the region (e.g., in the Mediterranean, representative of a seabed substrate type, a biological zone and a mask value).

Coding convention

For consistency across regions, it was agreed that all the habitat descriptor classes would be coded to two digits except, with the exception of the mask habitat descriptor classes, which are coded to one digit due to technical limitations. It was also agreed that, as shown in Figure 6, the two digits from the right would be devoted to the seabed substrate and then, from right to left, the biological zone 2 digits, the energy 2 digits, the (temperature, oxygen or salinity) 2 digits, and the mask digit.

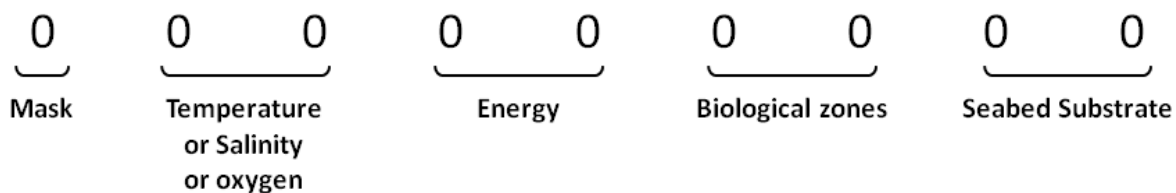


Figure 6: Coding convention for modelled habitat codes

Thus, in order for the above sum of the habitat descriptor codes to produce such coded values, each habitat descriptor code must be multiplied by the factors given in Table 4.

Table 4: Multiplication factors used for each habitat descriptor

Habitat descriptor name	Multiplication factor
Seabed substrate	1
Biological zone zones	100
Energy	10000
Oxygen, Salinity, Temperature	1000000
Mask	100000000

A few examples of codes

In the Mediterranean Sea, 200001030 means seabed substrate=30 (coarse sediment), biological zone = 10 (i.e. infralittoral) and mask = 2 (i.e. mask 1).

In the Black Sea, 100002020 means seabed substrate=20 (sand), biological zone = 20 (i.e. circalittoral) and mask = 1 (i.e. no mask).

In the Atlantic 103020 would mean seabed substrate=20 (sand), biological zone = 30 (i.e. deep circalittoral) and energy = 10 (i.e. low energy)

3.3.2 Joining the look-up table

The final step is to join the habitat layer and the look-up table, i.e. the table that crosswalks the digital habitat codes mentioned above and the habitat classes described in several habitat classifications (e.g. EUNIS).

modelCod	Biozone	Substrat	EUNISc	EUNISd
100001000	Infralittoral	Unknown	Na	Na
100001010	Infralittoral	Mud	A5.34	Infralittoral fine mud
100001020	Infralittoral	Sand	A5.23	Infralittoral fine sands
100001030	Infralittoral	Coarse & mixed sediment	A5.13	Infralittoral coarse sediment
100001040	Infralittoral	Coarse & mixed sediment	A5.13	Infralittoral coarse sediment
100001050	Infralittoral	Sandy Mud	A5.33	Infralittoral sandy mud
100001060	Infralittoral	Muddy Sand	A5.23	Infralittoral fine sands
100001070	Infralittoral	Rock or other hard substrata	A3	Infralittoral rock and other hard substrata
100001071	Infralittoral	Posidonia oceanica	A5.535	[Posidonia] beds
100001072	Infralittoral	Cymodocea nodosa	A5.531	[Cymodocea] beds
100002000	Circalittoral	Unknown	Na	Na
100002010	Circalittoral	Mud	A5.39	Mediterranean biocoenosis of coastal terrigenous muds
100002020	Circalittoral	Sand	A5.46	Mediterranean biocoenosis of coastal detritic bottoms
100002030	Circalittoral	Coarse & mixed sediment	A5.46	Mediterranean biocoenosis of coastal detritic bottoms
100002040	Circalittoral	Coarse & mixed sediment	A5.46	Mediterranean biocoenosis of coastal detritic bottoms
100002050	Circalittoral	Sandy Mud	A5.38	Mediterranean biocoenosis of muddy detritic bottoms
100002060	Circalittoral	Muddy Sand	A5.46	Mediterranean biocoenosis of coastal detritic bottoms
100002070	Circalittoral	Rock or other hard substrata	A4.26 or A4.32	Mediterranean coralligenous communities moderately exposed to or sheltered from hydrodynamic action
100002071	Infralittoral	Posidonia oceanica	A5.535	[Posidonia] beds
100002072	Infralittoral	Cymodocea nodosa	A5.531	[Cymodocea] beds

Figure 7: Example of look-up table. In the column "modelCod" is the modelled habitat code; the columns EUNISc and EUNISd are the EUNIS code and description that match the code

4 The R scripts & ArcGIS™ model

The R scripts and the ArcGIS™ toolbox are available on Github:

https://github.com/emodnetseabedhabitats/EUSeaMap_creation

Two R scripts are proposed for the two steps that allow the production of a habitat descriptor layer and its confidence layer (section 3.1). For the other tasks in the general workflow, i.e. converting the habitat descriptor raster layers into polygon layers (section 3.2) and combining them to produce the habitat layer (section 3.3.1), and joining the habitat layer and the look-up table (section 3.3.2), we recommend the use of ARCGIS™. The various R package functions tested for raster-to-polygon conversion did not prove sufficiently efficient for large numbers of polygons (which is the case for EUSeaMap). Other tests of ad hoc R packages for intersecting polygon layers also showed some limitations (for more details see Vasquez et al. 2020, annex 4). A ARCGIS™ ModelBuilder model is proposed that allows raster-to-polygon conversion and polygon intersection to be performed together. An ARCGIS™ ModelBuilder model is also proposed for joining the habitat layer and the look-up table. Due to the large number of polygons, it is recommended that the outputs of the ARCGISTM ModelBuilder models be produced in geodatabase format rather than shapefile format.

The scripts and the ARCGIS™ ModelBuilder models are documented in section 4.3 (Script for the production of habitat descriptor layers) and 4.4 (Creating the habitat layer and overall confidence). Figure 8 shows the general workflow with references in brackets to the subsection describing the different tasks.



Figure 8: General workflow and subsection (in brackets) that documents the script or ModelBuilder model corresponding to the various tasks

4.1 Prerequisites

4.1.1 System requirements

RAM/Memory: the R package *terra* is memory hungry. Minimum requirement is 16 GB, ideal 32 GB.

4.1.2 Software Requirements

R

The scripts are written in the R language. Therefore, R has to be installed.

R packages

The R package *terra* is required

ArcGIS™

ARCGIS™ ModelBuilder models require ArcGIS™ 10.0 or higher. No extension to native ArcGIS™ is required.

4.2 General principles

4.2.1 A critical requirement: raster inputs must all be spatially consistent

Since most of the scripts perform raster overlay operations, a key requirement is that all input rasters must be fully spatially consistent, i.e. strictly the same spatial extent and resolution. If this requirement is not met, the script will fail and return an error message such as "different extent" or "different resolution".

4.2.2 Formulation of the R script inputs via configuration files

The R scripts were designed to be run by non-R experts, i.e. as a "black box". Most of the inputs passed to the scripts are formulated via configuration files in csv (comma separated values) format (Figure 9). In general, very few parameters, such as the working directory, need to be formulated directly in the script (Figure 10).

shortName	longName	code	upper_boundary_equation_type	upper_boundary_variable	upper_boundary_slope	upper_boundary_intercept
infra	Infralittoral	10				
shallowCirca	Shallow circalittoral	20	1	SeabedLightEnergy.tif	-0.926	2.102
deepCirca	Deep circalittoral	30	1	SeabedLightPercentage.tif	-2000	1.5

Figure 9: Screenshot of a csv configuration file in which script inputs are formulated

```
#----- script parametrisation -----
#chunk size and max memory to load for processings
chunksize<-1e+07
maxmemory<-1e+09

#working directory
workingDirectory<-"d:/travail/euseamap_phase3/wp2_modelling_thresholds_confidence/Modeling/r_draft/medsea_small_with_new_codif"

#the config csv file that describes all the bits that have to be merged
config_csvFileName<-"medsea_habitat_descriptor_merging_inputs.csv"

#the name of the habitat descriptor as it appears in the config csv file, column "habitat_descriptor_shortName"
habitat_descriptor_shortName<-"biozones"
#-----
```

Figure 10: Example of parameters that are formulated in a script. In this example, there are 5 parameters (*workingDirectory*, *config_csvFileName*, *habitat_descriptor_shortName*)

In the script description sheets the inputs passed via the configuration file are referred to as 'Inputs' and the parameters formulated in the script are referred to as 'script parameters'.

4.2.3 A strict folder structure

For a given study area (e.g. Mediterranean Sea), the user must follow a standard folder structure (Figure 11) in order to run the script.

The root folder is the **working directory**. Below this are the following folders:

- **config_files:** folder where the script expects the csv configuration files to be stored.
- **input:** folder where the script expects the input files to be stored
- **output:** folder where the script writes the output files

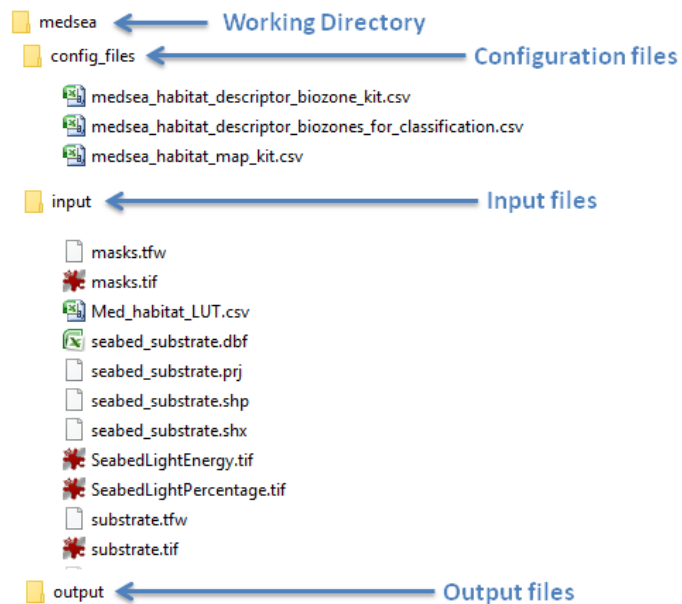


Figure 11: The folder structure

4.2.4 Changing the maximum amount of memory used by terra

The "terra" R package uses a lot of memory as it performs all calculations in memory. All the scripts that use the terra package have a script parameter, 'maxmemory', which specifies the maximum amount of RAM (in GB) that terra is allowed to use when processing a raster dataset. Depending on the amount of RAM available on your computer, you may wish to increase this value.

4.3 Scripts for the production of habitat descriptor layers

4.3.1 Modelling the habitat descriptor classes

Script name	habitat_descriptor_modelling.R									
Author	Mickaël Vasquez (Ifremer)									
Date	26/05/2023									
Descr.	<p>Produces rasters of the spatial distribution of the classes of a habitat descriptor according to the modelling law parameters that characterise the class boundaries. The script produces a spatial distribution raster for each class whose input parameters are described in the csv configuration file. This configuration file specifies for each class boundary the predictor (i.e. the environmental variable) raster, the equation type (GLM or fuzzy), the slope and intercept of the equation, the probability threshold and the maximum probability of the law. See section 3.1 for more details on these parameters.</p> <p>For each class, the script will produce a spatial distribution raster, a confidence raster, a confidence raster based on probability and, optionally, a probability raster.</p> <p>If it's a GLM, the equation used for the probability raster calculation is</p> $P(\text{predictor}) = \frac{e^{(\text{slope} * \text{predictor} + \text{intercept})}}{1 + e^{(\text{slope} * \text{predictor} + \text{intercept})}}$ <p>If it's a fuzzy law, it is</p> $P(\text{predictor}) = \text{slope} * \text{predictor} + \text{intercept}$									
Workflow	Figure 2, (see number 1) and Figure 4									
Inputs	<p>All inputs are described in a csv configuration file. The csv file contains one row for each habitat descriptor class. If the habitat descriptor class has only one boundary (see section 2 for further explanation of boundaries), the csv file cells for the missing boundary parameters are left blank.</p> <p>Important: all input raster files must be located in the input folder</p> <table border="1"> <thead> <tr> <th>Column Name</th> <th>Type*</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>shortName</td> <td>C</td> <td>The habitat descriptor class short name. Crucial. it will be used in the name of the output rasters. Must not contain special characters</td> </tr> <tr> <td>longName</td> <td>C</td> <td>The habitat descriptor class short name. Only there for the sake of clarity in the messages displayed by the script</td> </tr> </tbody> </table>	Column Name	Type*	Description	shortName	C	The habitat descriptor class short name. Crucial. it will be used in the name of the output rasters. Must not contain special characters	longName	C	The habitat descriptor class short name. Only there for the sake of clarity in the messages displayed by the script
Column Name	Type*	Description								
shortName	C	The habitat descriptor class short name. Crucial. it will be used in the name of the output rasters. Must not contain special characters								
longName	C	The habitat descriptor class short name. Only there for the sake of clarity in the messages displayed by the script								

code	N	The habitat descriptor class code that will be given to the output class raster. Must follow the conventions (see annex 1)
upper_boundary_equation_type	N Or C	For the class upper boundary, Integer value that indicates the type of modelling law 1 = fuzzy 2 = GLM If the type of modelling law varies spatially across the study area, a raster name must be provided instead of an integer value. The raster cells will indicate, via the same values as above, if the type is fuzzy or GLM
upper_boundary_variable	C	For the class upper boundary, the raster file name for the environmental variable
upper_boundary_confidence_in_variable	C	For the class upper boundary, the raster file name for the confidence in the environmental variable
upper_boundary_slope	N Or C	For the class upper boundary, the slope of the GLM/fuzzy law. If the slope varies spatially across the study area, a raster name must be provided instead of a numeric value. The raster cells will indicate the value of the slope
upper_boundary_intercept	N Or C	For the class upper boundary, the intercept of the GLM/fuzzy law. If the intercept varies spatially across the study area, a raster name must be provided instead of a numeric value. The raster cells will indicate the value of the intercept
upper_boundary_threshold	N Or C	For the class upper boundary, the probability threshold. If the probability threshold varies spatially across the study area, a raster name must be provided instead of a numeric value. The raster cells will indicate the value of the threshold

	upper_boundary_max_probability	N Or C	For the class upper boundary, the maximum value that the probability can have (in some cases may be less than 1) If the maximum probability varies spatially across the study area, a raster name must be provided instead of a numeric value. The raster cells will indicate the value of the maximum probability
	lower_boundary_equation_type	N Or C	For the class lower boundary, Integer value that indicates the type of modelling law 1 = fuzzy 2 = GLM If the type of modelling law varies spatially across the study area, a raster name must be provided instead of an integer value. The raster cells will indicate, via the same values as above, if the type is fuzzy or GLM
	lower_boundary_variable	C	For the class lower boundary, driver raster name
	lower_boundary_confidence_in_variable	C	For the class lower boundary, the raster file name for the confidence in the environmental variable
	lower_boundary_slope	N Or C	For the class lower boundary, the slope of the GLM/fuzzy law. If the slope varies spatially across the study area, a raster name must be provided instead of a numeric value. The raster cells will indicate the value of the slope
	lower_boundary_intercept	N Or C	For the class lower boundary, the intercept of the GLM/fuzzy law. If the intercept varies spatially across the study area, a raster name must be provided instead of a numeric value. The raster cells will indicate the value of the intercept
	lower_boundary_threshold	N Or C	For the class lower boundary, the probability threshold. If the probability threshold varies spatially across the study area, a raster name must be provided instead of a

	<p>numeric value. The raster cells will indicate the value of the threshold</p> <p>lower_boundary_max_probability N For the class lower boundary, the maximum value that the probability can have (in some cases may be less than 1)</p> <p>Or</p> <p>C If the maximum probability varies spatially across the study area, a raster name must be provided instead of a numeric value. The raster cells will indicate the value of the maximum probability</p> <p>(*) C=character, N=numerical</p>															
<p>Outputs</p>	<p>For each habitat descriptor class described in the csv file (i.e. each row of the configuration file), the outputs are as follows:</p> <ul style="list-style-type: none"> • A spatial distribution raster, the cells of which have the value indicated in the csv file column "code". The name of the tif file is the class short name (indicated in the csv file column "shortName") • A confidence raster, obtained by combining the raster layer of the confidence in values of the environmental variable and the raster layer of the confidence based on the probability. The name of the tif file is the class short name (indicated in the csv file column "shortName") + "_confidence_overall" • A raster for the confidence based on the probability. The name of the tif file is the class short name (indicated in the csv file column "shortName") + "_confidence_based_on_proba.tif" • Optionally, A probability raster. The name of the tif file is the class short name (indicated in the csv file column "shortName") + "_proba" <p>All the outputs will be created in the folder 'output'.</p>															
<p>Scripts parameters</p>	<table border="1"> <thead> <tr> <th>Name</th> <th>Type*</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>maxmemory</td> <td>N</td> <td>See section Erreur ! Source du renvoi introuvable.</td> </tr> <tr> <td>workingDirectory</td> <td>C</td> <td>Full path to the working directory</td> </tr> <tr> <td>config_csvFileName</td> <td>C</td> <td>Name of the configuration name</td> </tr> <tr> <td>habitat_descriptor_probability_rasters_as_output</td> <td>B</td> <td>If TRUE for each habitat descriptor class a raster file will be created for the probability rasters. If FALSE, no file will be created</td> </tr> </tbody> </table> <p>(*) C=character, N=numerical , B=Boolean</p>	Name	Type*	Description	maxmemory	N	See section Erreur ! Source du renvoi introuvable.	workingDirectory	C	Full path to the working directory	config_csvFileName	C	Name of the configuration name	habitat_descriptor_probability_rasters_as_output	B	If TRUE for each habitat descriptor class a raster file will be created for the probability rasters. If FALSE, no file will be created
Name	Type*	Description														
maxmemory	N	See section Erreur ! Source du renvoi introuvable.														
workingDirectory	C	Full path to the working directory														
config_csvFileName	C	Name of the configuration name														
habitat_descriptor_probability_rasters_as_output	B	If TRUE for each habitat descriptor class a raster file will be created for the probability rasters. If FALSE, no file will be created														
<p>How to run it?</p>	<ul style="list-style-type: none"> • Create a configuration file and put it in the Config_files folder • Open the script in RStudio • Edit the script parameters described above 															

- Click Button 'Source' in RStudio

4.3.2 Merging the individual habitat descriptor classes in a single layer

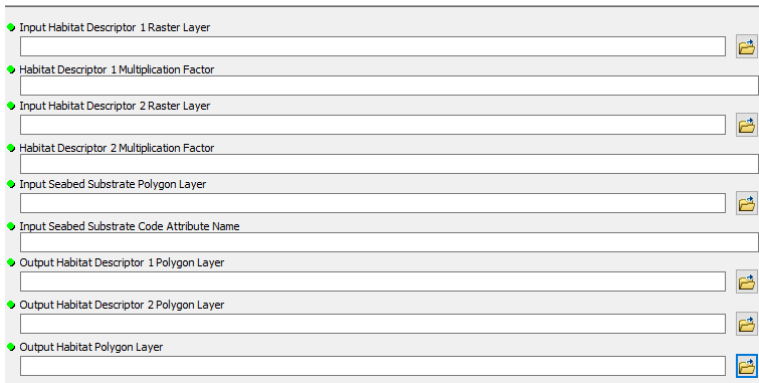
Script name	habitat_descriptor_merging_classes.R																		
Author	Mickaël Vasquez																		
Date	04/2021																		
Descr.	<p>For one or more habitat descriptors</p> <ul style="list-style-type: none"> • All the spatial distribution rasters for each individual class are merged into a single raster layer • All the confidence rasters for each individual class are merged into a single raster layer • All the confidence based on probability rasters for each individual class are merged into a single raster layer • All the habitat presence probability rasters for each individual class are merged into a single raster layer 																		
Workflow	Figure 2, (see number 2)																		
Inputs	<p>All inputs are described via a csv configuration file. The csv file may describe the inputs for several habitat descriptors (i.e. no need to prepare a csv file for each habitat descriptor). For each habitat descriptor, the csv file contains one row for each raster layer that that will be merged in the output layer (so one row per habitat descriptor class to be merged).</p> <table border="1"> <thead> <tr> <th>Column Name</th> <th>Type*</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>habitat_descriptor_shortName</td> <td>C</td> <td>The name of the habitat descriptor to which the class is attached. Crucial. it will be used in the name of the output rasters. Must not contain special characters</td> </tr> <tr> <td>class_fileName</td> <td>C</td> <td>The name of the input raster file that characterises the habitat descriptor class spatial distribution</td> </tr> <tr> <td>proba_fileName</td> <td>C</td> <td>The name of the input raster file that characterises the habitat descriptor class presence probability</td> </tr> <tr> <td>based_on_proba_confidence_fileName</td> <td>C</td> <td>The name of the input raster file that characterises the confidence in the habitat descriptor class occurrence based on the probability</td> </tr> <tr> <td>overall_confidence_fileName</td> <td>C</td> <td>The name of the input raster file that characterises the overall confidence in the habitat descriptor class occurrence</td> </tr> </tbody> </table>	Column Name	Type*	Description	habitat_descriptor_shortName	C	The name of the habitat descriptor to which the class is attached. Crucial. it will be used in the name of the output rasters. Must not contain special characters	class_fileName	C	The name of the input raster file that characterises the habitat descriptor class spatial distribution	proba_fileName	C	The name of the input raster file that characterises the habitat descriptor class presence probability	based_on_proba_confidence_fileName	C	The name of the input raster file that characterises the confidence in the habitat descriptor class occurrence based on the probability	overall_confidence_fileName	C	The name of the input raster file that characterises the overall confidence in the habitat descriptor class occurrence
Column Name	Type*	Description																	
habitat_descriptor_shortName	C	The name of the habitat descriptor to which the class is attached. Crucial. it will be used in the name of the output rasters. Must not contain special characters																	
class_fileName	C	The name of the input raster file that characterises the habitat descriptor class spatial distribution																	
proba_fileName	C	The name of the input raster file that characterises the habitat descriptor class presence probability																	
based_on_proba_confidence_fileName	C	The name of the input raster file that characterises the confidence in the habitat descriptor class occurrence based on the probability																	
overall_confidence_fileName	C	The name of the input raster file that characterises the overall confidence in the habitat descriptor class occurrence																	

	<p>folder C The name of the folder that contains the input raster file. Classically the output folder</p> <p>(*) C=character, N=numerical</p>																								
Outputs	<p>For each habitat descriptor described in the csv file, creates (all the outputs will be created in the folder 'output'):</p> <ul style="list-style-type: none"> • A spatial distribution raster, the name of which is what is indicated in the csv file in the column "habitat_descriptor_shortName" • A probability raster, result of all the class probability rasters merging, the name of which is what is indicated in the csv file in the column "habitat_descriptor_shortName" + "_proba" • A confidence based on probability raster, result of all the class confidence based on probability rasters merging, the name of which is what is indicated in the csv file in the column "habitat_descriptor_shortName" + "_confidence_based_on_proba" • An overall confidence raster, result of all the class overall confidence raster rasters merging, the name of which is what is indicated in the csv file in the column "habitat_descriptor_shortName" + "_confidence_overall" 																								
Scripts parameters	<table border="1"> <thead> <tr> <th>Name</th> <th>Type*</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>maxmemory</td> <td>N</td> <td>See section Erreur ! Source du renvoi introuvable.</td> </tr> <tr> <td>workingDirectory</td> <td>C</td> <td>Full path to the working directory</td> </tr> <tr> <td>config_csvFileName</td> <td>C</td> <td>Name of the configuration name</td> </tr> <tr> <td>output_habitat_descriptor_raster</td> <td>B</td> <td>If TRUE a raster file will be created for the spatial distribution. If FALSE no file will be created</td> </tr> <tr> <td>output_overall_confidence</td> <td>B</td> <td>If TRUE a raster file will be created for the overall confidence. If FALSE no file will be created</td> </tr> <tr> <td>output_confidence_based_on_proba</td> <td>B</td> <td>If TRUE a raster file will be created for the confidence based on probability. If FALSE no file will be created</td> </tr> <tr> <td>output_probability_raster</td> <td>B</td> <td>If TRUE a raster file will be created for the probability. If FALSE no file will be created</td> </tr> </tbody> </table> <p>(*) C=character, N=numerical, B=boolean</p>	Name	Type*	Description	maxmemory	N	See section Erreur ! Source du renvoi introuvable.	workingDirectory	C	Full path to the working directory	config_csvFileName	C	Name of the configuration name	output_habitat_descriptor_raster	B	If TRUE a raster file will be created for the spatial distribution. If FALSE no file will be created	output_overall_confidence	B	If TRUE a raster file will be created for the overall confidence. If FALSE no file will be created	output_confidence_based_on_proba	B	If TRUE a raster file will be created for the confidence based on probability. If FALSE no file will be created	output_probability_raster	B	If TRUE a raster file will be created for the probability. If FALSE no file will be created
Name	Type*	Description																							
maxmemory	N	See section Erreur ! Source du renvoi introuvable.																							
workingDirectory	C	Full path to the working directory																							
config_csvFileName	C	Name of the configuration name																							
output_habitat_descriptor_raster	B	If TRUE a raster file will be created for the spatial distribution. If FALSE no file will be created																							
output_overall_confidence	B	If TRUE a raster file will be created for the overall confidence. If FALSE no file will be created																							
output_confidence_based_on_proba	B	If TRUE a raster file will be created for the confidence based on probability. If FALSE no file will be created																							
output_probability_raster	B	If TRUE a raster file will be created for the probability. If FALSE no file will be created																							
How to run it?	<ul style="list-style-type: none"> • Create a configuration file and put it in the Config_files folder • Open the script in RStudio • Edit the script parameters described above • Click Button 'Source' in RStudio 																								

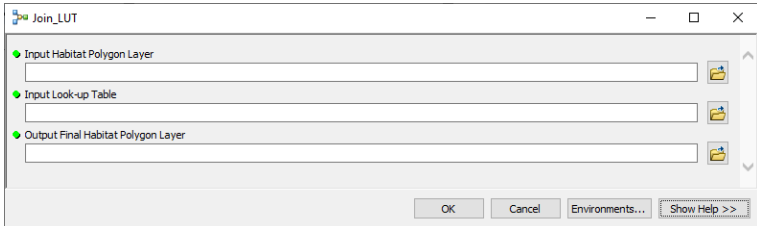
4.4 Creating the habitat layer and overall confidence

4.4.1 Combining the habitat descriptor layers

ArcGIS™ Model Builder Model name	Combine_habitat_descriptors		
Author	Mickaël Vasquez		
Date	04/2021		
Descr.	<p>Requires ArcGIS™ 10.0 (or higher). Creates the habitat layer, which includes:</p> <ol style="list-style-type: none"> 1) Converting the modelled habitat descriptor rasters into polygon layers 2) Overlaying (by intersection) all the habitat descriptor polygon layers, including the seabed substrate (which is expected to be already in the form of a polygon layer) 3) Calculating the habitat code, which is the sum of: seabed substrate code + (habitat descriptor 1 code x habitat descriptor 1 multiplication factor) + (habitat descriptor 2 code x habitat descriptor 2 multiplication factor). For the multiplication factor, the user may use the convention elaborated for EUSeaMap (see table 4) <p>Note: the model assumes that there are 2 modelled habitat descriptors, as this is typically the case in EUSeaMap. If there are more habitat descriptors, the model will need to be edited.</p>		
Workflow reference	Figure 1, (numbers 2 and 3)		
Inputs	Model Variable name	Type	Description
	Input Seabed Substrate Polygon Layer	Feature class	The input Seabed Substrate polygon layer
	Input Seabed Substrate Code Attribute Name	String	The name of the seabed substrate polygon layer attribute that describes the seabed substrate code
	Input Habitat Descriptor 1 Raster Layer	Raster dataset	The input habitat descriptor 1 raster layer
	Habitat Descriptor 1 Multiplication Factor	Variant	The multiplication factor for the Habitat Descriptor 1. The user may use the convention elaborated for EUSeaMap (see table 4)
	Input Habitat Descriptor 2 Raster Layer	Raster dataset	The input habitat descriptor 2 raster layer

	<p>Habitat Descriptor 2 Multiplication Factor Variant The multiplication factor for the habitat descriptor 2. The user may use the convention elaborated for EUSeaMap (see table 4)</p>												
<p>Outputs</p>	<table border="1"> <thead> <tr> <th>Model Variable name</th> <th>Type</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>Output Habitat Descriptor 1 Polygon Layer</td> <td>Feature class</td> <td>The habitat descriptor 1 polygon obtained by the raster-to-polygon conversion. is recommended to create a geodatabase feature class rather than a shapefile</td> </tr> <tr> <td>Output Habitat Descriptor 2 Polygon Layer</td> <td>Feature class</td> <td>The habitat descriptor 2 polygon obtained by the raster-to-polygon conversion. It is recommended to create a geodatabase feature class rather than a shapefile</td> </tr> <tr> <td>Output Habitat Polygon Layer</td> <td>Feature class</td> <td>The habitat polygon obtained by intersection of habitat descriptor 1 polygon layer, the habitat descriptor 2 polygon layer, and the seabed substrate. It contains an attribute, “hab_code”, the values of which are the sum of: seabed substrate code + (habitat descriptor 1 code x habitat descriptor 1 Multiplication Factor) + (habitat descriptor 2 code x habitat descriptor 2 Multiplication Factor). It is recommended to create a geodatabase feature class rather than a shapefile</td> </tr> </tbody> </table>	Model Variable name	Type	Description	Output Habitat Descriptor 1 Polygon Layer	Feature class	The habitat descriptor 1 polygon obtained by the raster-to-polygon conversion. is recommended to create a geodatabase feature class rather than a shapefile	Output Habitat Descriptor 2 Polygon Layer	Feature class	The habitat descriptor 2 polygon obtained by the raster-to-polygon conversion. It is recommended to create a geodatabase feature class rather than a shapefile	Output Habitat Polygon Layer	Feature class	The habitat polygon obtained by intersection of habitat descriptor 1 polygon layer, the habitat descriptor 2 polygon layer, and the seabed substrate. It contains an attribute, “hab_code”, the values of which are the sum of: seabed substrate code + (habitat descriptor 1 code x habitat descriptor 1 Multiplication Factor) + (habitat descriptor 2 code x habitat descriptor 2 Multiplication Factor). It is recommended to create a geodatabase feature class rather than a shapefile
Model Variable name	Type	Description											
Output Habitat Descriptor 1 Polygon Layer	Feature class	The habitat descriptor 1 polygon obtained by the raster-to-polygon conversion. is recommended to create a geodatabase feature class rather than a shapefile											
Output Habitat Descriptor 2 Polygon Layer	Feature class	The habitat descriptor 2 polygon obtained by the raster-to-polygon conversion. It is recommended to create a geodatabase feature class rather than a shapefile											
Output Habitat Polygon Layer	Feature class	The habitat polygon obtained by intersection of habitat descriptor 1 polygon layer, the habitat descriptor 2 polygon layer, and the seabed substrate. It contains an attribute, “hab_code”, the values of which are the sum of: seabed substrate code + (habitat descriptor 1 code x habitat descriptor 1 Multiplication Factor) + (habitat descriptor 2 code x habitat descriptor 2 Multiplication Factor). It is recommended to create a geodatabase feature class rather than a shapefile											
<p>How to run it?</p>	<ul style="list-style-type: none"> • Open the model in the ActoolBox. The following dialog box will appear  <p>If an update of the tool is required, (e.g. add an habitat descriptor), edit it</p>												
<p>Main ArcGIS™ tools used</p>	<ul style="list-style-type: none"> • Raster to Polygon (Conversion) • Intersect (Analysis) 												

4.4.2 Joining the habitat polygon layer and the look-up table that crosswalks the modelled habitat codes and habitat classifications

ArcGIS™ Model Builder Model name	Join_LUT												
Author	Mickaël Vasquez												
Date	09/2023												
Descr.	Joins the habitat shapefile produced by the model "Combine_habitat_descriptors" (see section 4.4), in which habitats are described via a code, to a look-up table (in dbase or geodatabase table format) that crosswalks each habitat code with various habitat classification (e.g. EUNIS).												
Workflow reference	Figure 1, (see number 4)												
Outputs	A shapefile containing the attributes of the look-up table												
Scripts parameters	<table border="1"> <thead> <tr> <th>Model Variable name</th> <th>Type</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>Input Habitat Polygon Layer</td> <td>C</td> <td>The Habitat polygon layer. It is recommended to use a geodatabase feature class rather than a shapefile</td> </tr> <tr> <td>Input Look-up Table</td> <td>C</td> <td>Look-up table. In dbase or geodatabase table format</td> </tr> <tr> <td>Output Final Habitat Polygon Layer</td> <td>C</td> <td>The output Habitat polygon layer</td> </tr> </tbody> </table>	Model Variable name	Type	Description	Input Habitat Polygon Layer	C	The Habitat polygon layer. It is recommended to use a geodatabase feature class rather than a shapefile	Input Look-up Table	C	Look-up table. In dbase or geodatabase table format	Output Final Habitat Polygon Layer	C	The output Habitat polygon layer
Model Variable name	Type	Description											
Input Habitat Polygon Layer	C	The Habitat polygon layer. It is recommended to use a geodatabase feature class rather than a shapefile											
Input Look-up Table	C	Look-up table. In dbase or geodatabase table format											
Output Final Habitat Polygon Layer	C	The output Habitat polygon layer											
How to run it?	<ul style="list-style-type: none"> Open the model in the ActoolBox. The following dialog box will appear 												
Main ArcGIS™ tools used	Multipart To Singlepart, Join Field, Repair Geometry												

4.4.3 Creating the habitat confidence layers

Script name	habitat_map_calculating_confidence.R												
Author	Mickaël Vasquez												
Date	05/04/2023												
Descr.	For one or more habitat classifications, create the habitat confidence layer, which is the combination of the individual confidence layers of the habitat descriptors that are involved in the habitat classification. For example in the EUNIS classification the habitat descriptors that are involved (seabed substrate, biological zone and energy level) are not the same as the ones involved in the MSFD broad habitat types classification (seabed substrate and biological zone only). For each cell, the result value is the minimum confidence value of the input habitat descriptor confidence layers.												
Workflow	Figure 1, (see number 3b). In figure 1, in the interest of readability only one output is represented. But there can be more than one depending on the number of classifications.												
Inputs	<p>All inputs are described via a csv configuration file. The csv file may describe the inputs for several habitat classifications (i.e. no need to prepare a csv file for each classification). For each habitat classification, the csv file contains one row for each habitat descriptor confidence raster layer that that will be combined in the output layer (so one row per habitat descriptor confidence to be merged).</p> <table border="1"> <thead> <tr> <th>Column Name</th> <th>Type*</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>habitat_classification_name</td> <td>C</td> <td>The name of the habitat classification for which the habitat descriptor is used. Crucial. it will be used in the name of the output raster. Must not contain special characters</td> </tr> <tr> <td>confidence_fileName</td> <td>C</td> <td>The name of the habitat descriptor confidence raster file</td> </tr> <tr> <td>folder</td> <td>C</td> <td>The path to the folder that contains the file, relatively to the working directory</td> </tr> </tbody> </table> <p>(*) C=character, N=numerical</p>	Column Name	Type*	Description	habitat_classification_name	C	The name of the habitat classification for which the habitat descriptor is used. Crucial. it will be used in the name of the output raster. Must not contain special characters	confidence_fileName	C	The name of the habitat descriptor confidence raster file	folder	C	The path to the folder that contains the file, relatively to the working directory
Column Name	Type*	Description											
habitat_classification_name	C	The name of the habitat classification for which the habitat descriptor is used. Crucial. it will be used in the name of the output raster. Must not contain special characters											
confidence_fileName	C	The name of the habitat descriptor confidence raster file											
folder	C	The path to the folder that contains the file, relatively to the working directory											
Outputs	For each habitat classification described in the csv file, creates (in the folder 'output') the overall confidence raster, the name of which is what is indicated in the csv file in the column "habitat_classification_name"+"_confidence_overall"												
Scripts parameters	<table border="1"> <thead> <tr> <th>Name</th> <th>Type*</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>maxmemory</td> <td>N</td> <td>See section Erreur ! Source du renvoi introuvable.</td> </tr> </tbody> </table>	Name	Type*	Description	maxmemory	N	See section Erreur ! Source du renvoi introuvable.						
Name	Type*	Description											
maxmemory	N	See section Erreur ! Source du renvoi introuvable.											

	<p>workingDirectory C Full path to the working directory</p> <p>config_csvFileName C Name of the configuration name</p> <p>(* C=character, N=numerical)</p>
How to run it?	<ul style="list-style-type: none"> • Create a configuration file and put it in the Config_files folder • Open the script in RStudio • Edit the script parameters described above • Click Button 'Source' in RStudio

4.5 Additional script: combining the wave-induced and current-induced habitat descriptor layers

Not shown in figure 1 (because specific to the energy habitat descriptor, which itself is specific to the EUNIS 2007 classification and to the Atlantic region) is the step that comprises combining the wave-induced and current-induced habitat descriptor layers and their respective confidence layers. The script performs this intermediate task.

Script name	habitat_descriptor_calculating_combined_energy.R
Author	Mickaël Vasquez
Date	03/05/2023
Descr.	Calculates the combined wave-induced and current-induced habitat descriptor layer and its confidence layer. The 2 habitat descriptor layers are combined by keeping, for each cell, the highest value of the 2 layers, e.g. if for a cell the current is moderate and the wave is low, then the combined energy is moderate. When it comes to the confidence, for each cell the confidence is that of the habitat descriptor that contributes to the calculation of the overall energy, ie if for a cell the combined energy is that of wave (because wave > current), then the combined confidence is that of wave. If for a cell the combined energy is that of wave and current (because wave = current), the confidence is the averaged confidence of wave and current, rounded up.
Worklow reference	Not shown in figure 1, but the outputs of figure 1/step 1 are individual habitat descriptor layers for current and wave. These 2 habitat descriptor layers subsequently have to be combined in order to produce a wave/current-induced energy layer, which together with the other habitat descriptor layers (biological zones, seabed substrate) is then provided as input to figure 1/step 2, Similarly, figure 1/step 3b requires as input a combined wave/current confidence layer.
Outputs	2 raster files: <ul style="list-style-type: none"> • One for the combined wave/current energy • One for the confidence in the combined wave/current energy

Scripts parameters	Name	Type*	Description
	workingDirectory	C	<i>Full link to the working directory (i.e. the directory that contains the habitat shapefile)</i>
	current_fileName	C	path to the input raster file name corresponding to the current-induced energy
	current_confidence_fileName	C	path to the input confidence raster file name corresponding to the current-induced energy
	wave_fileName	C	path to the input raster file name corresponding to the current-induced energy
	wave_confidence_fileName	C	path to the input confidence raster file name corresponding to the current-induced energy
	ouput_fileName	C	path to the output energy file name
	ouput_confidence_fileName	C	path to the output confidence file name
	calculateConfidence	B	Is it required to calculate confidence? If FALSE, the confidence layer is not calculated
	(*) C=character, B=Boolean		
How to run it?	<ul style="list-style-type: none"> • Open the script in RStudio • Edit the script parameters described above • Click Button 'Source' in RStudio 		

5 References

Coltman, N., Golding, N., Verling, E., 2008. Developing a broadscale predictive EUNIS habitat map for the MESH study area. 16 pp. Available online at www.emodnet-seabedhabitats.eu/pdf/MESH%20EUNIS%20model.pdf

Populus J., Vasquez M., Albrecht J., Manca E., Agnesi S., Al Hamdani Z., Andersen J., Annunziatellis A., Bekkby T., Bruschi A., Doncheva V., Drakopoulou V., Duncan G., Inghilesi R., Kyriakidou C., Lalli F., Lillis H., Mo G., Muresan M., Salomidi M., Sakellariou D., Simboura M., Teaca A., Tezcan D., Todorova V. and Tunesi L., 2017. EUSeaMap, a European broad-scale seabed habitat map. 174p. <http://doi.org/10.13155/49975>

Vasquez M., Manca E., Inghilesi R., Martin S., Agnesi S., Al Hamdani Z., Annunziatellis A., Bekkby T., Pesch R., Askew N., Bentes L., Castle L., Doncheva V., Drakopoulou V., Gonçalves J., Laamanen L., Lillis H., Loukaidi V., McGrath F., Mo G., Monteiro P., Muresan M., O'Keeffe E., Populus J., Pinder J., Ridgeway A., Sakellariou D., Simboura M., Teaca A., Tempera F., Todorova V., Tunesi L., Virtanen E. (2020). EUSeaMap 2019, A European broad-scale seabed habitat map, technical report . EASME/EMFF/2018/1.3.1.8/Lot2/SI2.810241– EMODnet Thematic Lot n° 2 – Seabed Habitats. <https://doi.org/10.13155/74782>

6 Appendix 1: the habitat descriptors used for EUSeaMap (v2019), their classes and codes

The following codes were used for the development of EUSeaMap 2019.

Seabed substrate

2-digit code	Name
00	No substrate data
10	Mud
20	Sand
30	Coarse sediment
40	Mixed sediment
50	Sandy mud
60	Muddy sand
70	Rock
71	Posidonia
73	Dead mattes of posidonia

Energy

2-digit code	Name
0	No Energy data
10	Low energy
20	Moderate energy
30	High energy

Biological zones

2-digit code	Name	2-digit code	Name
0	No information	74	Atlanto-Mediterranean lower bathyal
10	Infralittoral	75	Upper Abyssal
20	Circalittoral	76	Arctic upper abyssal
30	Deep circalittoral	77	Atlanto-arctic upper abyssal
40	Bathyal	78	Atlantic upper abyssal
50	Abyssal	79	Atlanto-Mediterranean upper abyssal
60	Upper bathyal	80	Mid abyssal
61	Atlantic upper bathyal NORTH	81	Arctic mid abyssal
62	Atlanto-Arctic upper bathyal	82	Atlanto-arctic mid abyssal
63	Atlantic upper bathyal	83	Atlantic mid abyssal
64	Atlanto-Mediterranean upper bathyal	84	Atlanto-Mediterranean mid abyssal
65	Mid Bathyal	85	Lower abyssal
66	Atlanto-Arctic upper bathyal in the NORTH Arctic	86	Arctic lower abyssal
67	Arctic mid bathyal	87	Atlanto-arctic lower abyssal
68	Atlantic mid bathyal	88	Atlantic lower abyssal
69	Atlanto-Mediterranean mid bathyal	89	Atlanto-Mediterranean lower abyssal
70	Lower Bathyal	91	Atlantic upper bathyal in South Arctic
71	Arctic lower bathyal	96	Atlanto-Arctic upper bathyal in South Arctic
72	Atlanto-Arctic lower bathyal		
73	Atlantic lower bathyal		

Oxygen

2-digit code	Name
0	No oxygen data
10	oxic
20	suboxic
30	anoxic

Salinity

2-digit code	Name
0	No salinity data
10	Oligohaline
20	Mesohaline
30	Polyhaline
60	Euhaline

Mask

1-digit code	Name
1	No mask
2	Mask 1
3	Mask 2

7 Appendix 2: calculating the slope, intercept and probability threshold for fuzzy laws

Figure 2.1 illustrates a fuzzy law for one of the two boundaries of a habitat descriptor class (e.g. lower boundary of the class 'shallow circalittoral'). In abscissa are the driver values (e.g. temperature). In ordinate is the probability of occurrence for the habitat descriptor class. The shape of the fuzzy function is governed by two control points, P0 and P1. P0 (x0,0) indicates where the probability begins to increase above 0. P1(x1,1) is the point where the probability starts to be 1.

In-between is a simple straight line, whose slope a and intercept b are defined as:

$$a = 1 / (x_1 - x_0)$$

$$b = -x_0 / (x_1 - x_0)$$

As with a GLM approach, a probability threshold (or cut-off) value has to be worked out. It is the probability value above which the habitat descriptor class will be classified as present and below which it will be classified as absent.

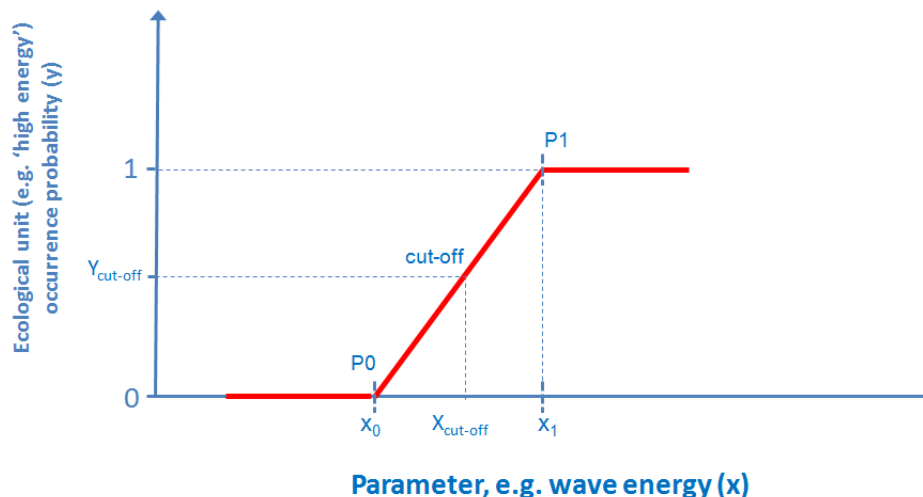


Figure 2.1. The fuzzy classifier shape is governed by 2 control points P0 and P1. The cut-off is the point whose y coordinate is the probability value above which the habitat descriptor class will be classified as present and below which it will be classified as absent.

Usually, in a threshold analysis, the values that are worked out are x_0 , x_1 and $x_{cut-off}$, while the GIS workflow requires as input the slope, the intercept and the probability threshold ($y_{cut-off}$ in figure 2.1). An excel file that is provided in the supplemental material will facilitate the calculation of these items from x_0 , x_1 and $x_{cut-off}$. The name of the file is "slope_intercept_fuzzy_calculation.xlsx".

	X1	X0	threshold	slope	intercept	probability threshold	
Infra lower boundary		2.27	1.19	1.82	0.93	-1.10	0.58
shallow circa upper boundary		1.19	2.27	1.82	-0.93	2.10	0.42
shallow circa lower boundary		0.00075	0.00025	0.0005	2000.00	-0.50	0.50
deep circa upper boundary		0.00025	0.00075	0.0005	-2000.00	1.50	0.50

\downarrow $1 / (x_1 - x_0)$ \downarrow $-x_0 / (x_1 - x_0)$ \downarrow threshold x slope + intercept

Figure 2.2. Screenshot of the excel file provided for the calculation of the slope, intercept and probability threshold from x_0 , x_1 and $x_{cut-off}$

8 Appendix 3: Hands-on training

In exercises 1 to 3 we will model the spatial distribution of the following habitat descriptor classes:

- Biological zones
- Wave-induced energy levels
- Current-induced energy levels

Their respective confidence raster layers will also be created.

In exercise 4, the class spatial distribution layers will be merged so that all classes of a habitat descriptor are combined into one layer.

In exercise 5, the habitat descriptor layers for current-induced energy levels and wave-induced energy levels will be combined in order to create a combined wave/current-induced energy level layer.

In exercise 6, the habitat layer will be created by combining the 3 habitat descriptor layers 'seabed substrates', 'energy levels' and 'biological zones'. In exercise 7 this layer will be joined to a table that crosswalks the modelled habitats with two standard classifications (EUNIS 2007-11 and MSFD broad habitat types).

The R scripts and the ArcGIS™ toolbox are available are all available on Github:

https://github.com/emodnetseabedhabitats/EUSeaMap_creation

The input data used for the hands-on training is available here:

<https://cloud.ifremer.fr/index.php/s/waZJVfyd8PfkIy7>

1 - Modelling the biological zone classes

1.1 – Becoming familiar with the csv configuration file

The csv configuration file describes all the elements required by the script to model a series of habitat descriptor classes. It is therefore essential to become familiar with the organisation of the file. The meaning of the file columns is documented in the report, section **Erreur ! Source du renvoi introuvable.**

- Open the csv configuration file (in the 'config_files' folder, file habitat_descriptor_modelling_biozones.csv). This file describes all the elements required by the script to model the biological zones.

Questions

- For the Infralittoral lower boundary, what is the name of the raster file used as environmental variable?

Answer: variable_seabed_par.tif

- For the Infralittoral lower boundary, what is the name of the raster file used as confidence in the environmental variable?

Answer: confidence_seabed_par.tif

- For the Shallow Circalittoral upper boundary, what is the name of the raster file used as environmental variable?

Answer: variable_seabed_par.tif. Remark: it is the same as for the infralittoral lower boundary, which is normal because the infralittoral lower boundary and the shallow circalittoral upper boundary are ... the same boundary, i.e. the boundary between the infralittoral and the shallow circalittoral

- For the Shallow Circalittoral lower boundary, what is the name of the raster file used as environmental variable?

Answer: variable_wavebase.tif. Remark: this variable is the ratio λ/h , where λ =wave wavelength and h =depth

- For the Shallow Circalittoral lower boundary, what is the name of the raster file used as confidence in the environmental variable?

Answer: confidence_wavebase.tif

- What is the type of equation used for the infralittoral lower boundary?

Answer: the value is 2, which means that the equation type is GLM

- What is the type of equation used for the abyssal upper boundary?

Answer: the value is 1, which means that the equation type is fuzzy

- why is the max probability for the shallow circalittoral 0.69 upper boundary and not 1?

Answer: the max probability will be for PAR at seabed=0 mol.pho.m⁻².d⁻²

As slope=-1.076 and intercept=0.777

$P(X=0) = \exp(-1.076*0+0.777) / (1+\exp(-1.076*0+0.777)) = 0.69$

- Check that in the csv file all the input parameters (class code, equation type, slope, intercept, probability thresholds, max probability) defined for each biological zone boundary are the same as those listed in table 3.1.

Was there any error?

1.2 – Looking at the input data

- In a GIS software, open all the input raster files (i.e. *variable* rasters and *confidence in variable* rasters) listed in the csv configuration file and take a look at them individually (remember, they are all in the 'input' folder).

- In the confidence in variable rasters, what are the possible values? What do they mean?

Answer: for a raster cell, the possible values of confidence in the value of the environmental variable are 1, 2 and 3, meaning Low, Moderate and High confidence, respectively.

1.3 – Running the script

- Open habitat_descriptor_modelling.R in Rstudio.

- Assign to the script variable `workingDirectory` the path to the working directory (which is the folder that contains the 'config_files', 'input', and 'ouput').

- Assign to the script variable `config_csvFileName` the name of the configuration file, i.e. 'habitat_descriptor_modelling_biozones.csv'

```
config_csvFileName<-"habitat_descriptor_modelling_biozones.csv"
```

- In Rstudio, click the 'Source' button. The script starts to run. Depending on the capacity of your computer, it may take some time to complete. When it has finished, the Rstudio console will display a message mentioning "Completed in..." followed by the duration.

Table 3.1: variable, classification method and parameters (slope, intercept, probability threshold and max probability) used as input to the script for each biological zone boundary. Other parameters (variable threshold, fuzzy X1 and fuzzy X0) are provided for background information

Habitat descriptor class boundary	Class code	Equation type	Variable	Variable Threshold	Fuzzy X1	Fuzzy X0	Slope	Intercept	Probability threshold	Max probability
Infralittoral Lower Boundary	10	GLM	PAR at seabed (mol.pho.m ⁻² .d ⁻¹)	0.7	-	-	1.076	-0.777	0.49	1
Shallow Circalittoral Upper Boundary	20	GLM	PAR at seabed (mol.pho.m ⁻² .d ⁻¹)	0.7	-	-	-1.076	0.777	0.51	0.69
Shallow Circalittoral Lower Boundary	20	GLM	Wave base ratio	1.5	-	-	19.2	-28.7	0.41	1
Deep Circalittoral Upper Boundary	30	GLM	Wave base ratio	1.5	-	-	-19.2	28.7	0.59	1
Deep Circalittoral Lower Boundary	30	Fuzzy	Depth (m)	-200	-180	-220	0.025	5.5	0.5	1
Upper Bathyal Upper Boundary	60	Fuzzy	Depth (m)	-200	-220	-180	-0.025	-4.5	0.5	1
Upper Bathyal Lower Boundary	60	Fuzzy	Depth (m)	-1300	-1017	-1583	0.0018	2.797	0.5	1
Lower Bathyal Upper Boundary	70	Fuzzy	Depth (m)	-1300	-1583	-1017	-0.0018	-1.797	0.5	1
Lower Bathyal Lower Boundary	70	Fuzzy	Depth (m)	-2200	-1912	-2488	0.0017	4.319	0.5	1
Abyssal Upper Boundary	75	Fuzzy	Depth (m)	-2200	-2488	-1912	-0.0017	-3.319	0.5	1

1.4 – Looking at the results

a) Check that as many spatial distribution rasters have been created as there are biological zones listed in the file 'habitat_descriptor_modelling_biozones.csv'. The tif files have been named according to what is written in the 'shortName' column of the csv file. In addition, 2 rasters have been created for each biological zone, i.e. one for the confidence based on the probability (the name of the tif is in the form 'shortName'_confidence_based_on_proba) and another one for the overall confidence (the name of the tif is in the form 'shortName'_confidence_overall), the latter being the combination of i) the former and ii) the raster for confidence in the environmental variable provided as input in the csv file (columns 'upper_boundary_confidence_in_variable' and 'lower_boundary_confidence_in_variable'). The output folder should contain the following files:

Spatial distribution files	Confidence based on probability	Overall confidence
Infra.tif	Infra_confidence_based_on_proba.tif	Infra_confidence_overall.tif
Shallow_Circa.tif	Shallow_Circa_confidence_based_on_proba.tif	Shallow_Circa_confidence_overall.tif
Deep_Circa.tif	Deep_Circa_confidence_based_on_proba.tif	Deep_Circa_confidence_overall.tif
Upper_Bathyal.tif	Upper_Bathyal_confidence_based_on_proba.tif	Upper_Bathyal_confidence_overall.tif
Lower_Bathyal.tif	Lower_Bathyal_confidence_based_on_proba.tif	Lower_Bathyal_confidence_overall.tif
Abyssal.tif	Abyssal_confidence_based_on_proba.tif	Abyssal_confidence_overall.tif

b) Take some time to look at these rasters in a GIS software. Check that the code assigned to each class spatial distribution raster cell matches the code given in the 'code' column of the csv configuration file.

2 - Modelling the current-induced energy classes

2.1 – Looking at the csv configuration file

- Open the csv configuration file (in the folder 'config_files', file habitat_descriptor_modelling_currents.csv). This file describes all the elements required by the script to model the current energy levels. The meaning of the file columns is documented in the master report, section **Erreur ! Source du renvoi introuvable.**

- Check that all the input parameters (class code, equation type, slope, intercept, probability thresholds, max probability) defined for each energy level boundary are the same as those listed in table 3.2.

Was there any error ?

Table 3.2: variable, classification method and parameters (slope, intercept, probability threshold and max probability) used as input to the script for each current level boundary. Other parameters (variable threshold, fuzzy X1 and fuzzy X0) are provided for background information

Habitat descriptor class boundary	Class code	Equation type	Variable	Variable Threshold	Fuzzy X1	Fuzzy X0	Slope	Intercept	Probability threshold	Max probability
Current High Lower Boundary	10	Fuzzy	Current-induced kinetic energy (N.m ⁻²)	400	450	350	0.01	-3.5	0.5	1
Current Moderate Upper Boundary	20	Fuzzy	Current-induced kinetic energy (N.m ⁻²)	400	350	450	-0.01	4.5	0.5	1
Current Moderate Lower Boundary	20	Fuzzy	Current-induced kinetic energy (N.m ⁻²)	80	100	60	0.025	-1.5	0.5	1
Current Low Upper Boundary	30	Fuzzy	Current-induced kinetic energy (N.m ⁻²)	80	60	100	-0.025	2.5	0.5	1

2.2 – Looking at the input data

Open in a GIS software all the input raster files (i.e. *variable* rasters and *confidence in variable* rasters) listed in the csv configuration file and take a look at them individually (remember, they are all in the 'input' folder).

2.3 – Running the script

- Open `habitat_descriptor_modelling.R` in Rstudio.

- Assign to the script variable `workingDirectory` the path to the working directory (which is the folder that contains the 'config_files', 'input', and 'output').

- Assign to the script variable `config_csvFileName` the name of the configuration file, i.e. 'habitat_descriptor_modelling_currents.csv'

```
config_csvFileName<-"habitat_descriptor_modelling_currents.csv"
```

- In Rstudio, click the 'Source' button. The script starts to run. Depending on the capacity of your computer, it may take some time to complete. When it is finished, the Rstudio console will display a message saying "Completed in..." followed by the duration.

2.4 – Looking at the results

a) Check that all the raster have been created. The 'output' folder should contain the following files:

Spatial distribution files	Confidence based on probability	Overall confidence
Currents_High.tif	Currents_High_confidence_based_on_proba.tif	Currents_High_confidence_overall.tif
Currents_Moderate.tif	Currents_Moderate_confidence_based_on_proba.tif	Currents_Moderate_confidence_overall.tif
Currents_low.tif	Currents_Low_confidence_based_on_proba.tif	Currents_Low_confidence_overall.tif

b) Take some time to look at these rasters in a GIS software. Check that the code assigned to each class spatial distribution raster cell matches the code given in the 'code' column of the csv configuration file.

3 - Modelling the wave-induced energy classes

3.1 – Looking at the csv configuration file

- Open the csv configuration file (in the folder 'config_files', file `habitat_descriptor_modelling_wave.csv`). This file describes all the items required by the script to model the wave energy levels. The meaning of the file columns is documented in the master report, section **Erreur ! Source du renvoi introuvable.**

- Check that all the input parameters (class code, equation type, slope, intercept, probability thresholds, max probability) defined for each energy level boundary are the same as those listed in table 3.3.

Was there any error ?

Table 3.3: variable, classification method and parameters (slope, intercept, probability threshold and max probability) used as input to the script for each wave level boundary. Other parameters (variable threshold, fuzzy X1 and fuzzy X0) are provided for background information

Habitat descriptor class boundary	Class code	Equation type	Variable	Variable Threshold	Fuzzy X1	Fuzzy X0	Slope	Intercept	Probability threshold	Max probability
Wave High Lower Boundary	10	Fuzzy	Wave-induced kinetic energy (N.m ⁻²)	22	27	17	0.1	-1.7	0.5	1
Wave Moderate Upper Boundary	20	Fuzzy	Wave-induced kinetic energy (N.m ⁻²)	22	17	27	-0.1	2.7	0.5	1
Wave Moderate Lower Boundary	20	Fuzzy	Wave-induced kinetic energy (N.m ⁻²)	7.6	10.6	4.6	0.16666667	-0.76666667	0.5	1
Wave Low Upper Boundary	30	Fuzzy	Wave-induced kinetic energy (N.m ⁻²)	7.6	4.6	10.6	-0.16666667	1.76666667	0.5	1

3.2 – Looking at the input data

Open in a GIS software all the input raster files (i.e. *variable* rasters and *confidence in variable* rasters) listed in the csv configuration file and take a look at them individually (reminder: they are all in the 'input' folder).

3.3 – Running the script

- Open `habitat_descriptor_modelling.R` in Rstudio.
- Assign to the script variable `workingDirectory` the path to the working directory (which is the folder that contains the 'config_files', 'input', and 'output').
- Assign to the script variable `config_csvFileName` the name of the configuration file, i.e. 'habitat_descriptor_modelling_wave.csv'

```
config_csvFileName<-"habitat_descriptor_modelling_wave.csv"
```

- In Rstudio, click the 'Source' button. The script starts to run. Depending on the capacity of your computer, it may take some time to complete. When it is finished, the Rstudio console will display a message saying "Completed in..." followed by the duration.

3.4 – Looking at the results

a) Check that all the raster have been created. The 'output' folder should contain the following files:

Spatial distribution files	Confidence based on probability	Overall confidence
Wave_High.tif	Wave_High_confidence_based_on_proba.tif	Wave_High_confidence_overall.tif
Wave_Moderate.tif	Wave_Moderate_confidence_based_on_proba.tif	Wave_Moderate_confidence_overall.tif
Wave_Low.tif	Wave_Low_confidence_based_on_proba.tif	Wave_Low_confidence_overall.tif

b) Take some time to look at these rasters in a GIS software. Check that the code assigned to each class spatial distribution raster cell matches the code given in the 'code' column of the csv configuration file.

4 – Merging all the habitat descriptor classes to create one single raster per habitat descriptor

The outputs of exercises 1 to 3 are individual spatial distribution raster layers and confidence raster layers for each class of the 3 habitat descriptors biological zones, current-induced energy levels and wave-induced energy levels. Here for the 3 habitat descriptors we will merge all the class raster layers into a single raster layer.

The script that does this is 'habitat_descriptor_merging_classes.R', documented in the master report, section 4.3.2.

4.1 – Looking at the csv configuration file

- Open the csv configuration file (in the 'config_files' folder, file habitat_descriptor_merging_inputs.csv). This file describes all the elements required by the script to create the habitat descriptor rasters by merging their corresponding class rasters. The meaning of the columns in the file is documented in the master report, section 4.3.2.

We want to create 3 rasters for each habitat descriptor, i.e for

- The spatial distribution
- The confidence based on probability
- The overall confidence

We do not want to merge the continuous probability rasters. That's why the column 'proba_fileName' of the csv file is empty.

We need to give the script the name of the raster files created in exercise 1, 2 and 3 as input so that it can merge them. That's what the columns 'class_fileName', 'based_on_proba_confidence_fileName' and 'overall_confidence_fileName' are for.

Indeed, for each habitat descriptor (whose name is in the column 'habitat_descriptor_shortName'), the rasters created in exercises 1, 2 and 3 are listed in the columns 'class_fileName' (class spatial distribution rasters), 'based_on_proba_confidence_fileName' (rasters for confidence based on probability), and 'overall_confidence_fileName' (rasters for overall confidence). The 'folder' column contains the name of the folder containing all these files. As they have all been created by us using the habitat_descriptor_modelling.R script, they are all in the 'output' folder.

4.2 – Running the script

- Open habitat_descriptor_merging_classes.R in Rstudio.

- Assign to the script variable `workingDirectory` the path to the working directory (which is the folder that contains the 'config_files', 'input', and 'output').

- Assign to the script variable `config_csvFileName` the name of the configuration file, i.e. 'habitat_descriptor_merging_inputs.csv'

```
config_csvFileName<-"habitat_descriptor_merging_inputs.csv"
```

- Set the 4 boolean variables as follows:

```
output_habitat_descriptor_raster<-TRUE
```

```
output_overall_confidence<-TRUE
```

```
output_confidence_based_on_proba<-TRUE
output_probability_raster<-FALSE
```

- In Rstudio, click the 'Source' button. The script starts to run. Depending on the capacity of your computer, it may take some time to complete. When it is finished, the Rstudio console will display a message saying "Completed in..." followed by the duration.

4.3 – Looking at the results

a) Check that all the rasters have been created.

3 rasters have been created for each habitat descriptor listed in the csv file, one for the spatial distribution, one for the confidence based on probability, and a last one for the overall confidence that the modelled class is correct.

As a result, the 'output' folder should contain the following files:

Spatial distribution files	Confidence based on probability	Overall confidence
Biozones.tif	Biozones_confidence_based_on_proba.tif	Biozones_confidence_overall.tif
Currents.tif	Currents_confidence_based_on_proba.tif	Currents_confidence_overall.tif
Wave.tif	Wave_confidence_based_on_proba.tif	Wave_confidence_overall.tif

b) Take some time to look at these rasters in a GIS software.

5 - combining the wave-induced and current-induced habitat descriptor layers

In exercises 1 to 4 we have created the raster layers for the habitat descriptors biological zones, wave-induced and current-induced energy levels. Before creating the habitat layer, we need to combine the latter two to create a single layer for the combined current/wave energy levels. The confidence raster layers are also combined. The script 'habitat_descriptor_calculating_combined_energy.R' (see master report, section 4.5) performs these tasks.

5.1 – Running the script

- Open `habitat_descriptor_calculating_combined_energy.R` in Rstudio.
- Assign to the script variable `workingDirectory` the path to the working directory (which is the folder that contains the 'config_files', 'input', and 'ouput').

- Assign as follows the R variables that indicate the path to the input rasters:

```
#path to the input raster file name corresponding to the current-induced energy
current_fileName<-"output/Currents.tif"

#path to the input confidence raster file name corresponding to the current-induced
energy
current_confidence_fileName<-"output/Currents_confidence_overall.tif"

#path to the input raster file name corresponding to the current-induced energy
wave_fileName<-"output/Wave.tif"

#path to the input confidence raster file name corresponding to the current-induced
energy
wave_confidence_fileName<-"output/Wave_confidence_overall.tif"
```

- Assign as follows the R variables that indicate the path to the output rasters:

```
#path to the output energy file name
ouput_fileName<-"output/energy.tif"

#path to the output confidence file name
ouput_confidence_fileName<-"output/energy_confidence_overall.tif"
```

- Set the following R boolean variable to TRUE for the script to output the combined confidence raster layer as well.

```
calculateConfidence<-TRUE
```

- In Rstudio, click the 'Source' button. The script starts to run. Depending on the capacity of your computer, it may take some time to complete. When it is finished, the Rstudio console will display a message saying "Completed in..." followed by the duration.

5.2 – Looking at the results

- a) Check that all the rasters have been created. There should be 2 files in the 'output' folder: `energy.tif` and `energy_confidence_overall.tif`.
- b) Take some time to look at these rasters in a GIS software. You can also look at the input rasters and see what the output rasters look like compared to the input rasters.

6 - Combining the habitat descriptor layers

Now it is time to create the habitat map. The first step is to combine the habitat descriptor layer 'seabed substrate', which is a polygon layer provided by EMODnet Geology, with the 2 habitat descriptor layers that we have created in exercises 1 to 5, i.e. the biological zones and the combined wave/current-induced energy levels. In order to preserve all the spatial detail of the seabed substrate layer, the combination is done in polygon mode. Therefore, the raster layers must be converted to polygons prior to combination. The combination is then performed by intersecting the layers.

Unfortunately, this is not done using a R script. The reason for this is that R packages have not proven to be efficient for either raster-to-polygon transformation or intersection of polygon layers when the number of polygons is large, which is usually the case for EUSeaMap. Therefore, ArcGIS™ is used here. A ModelBuilder model has been developed. It is documented in the master report, section 4.4.1.

6.1 – Looking at the seabed substrate layer

In a GIS software, open the shapefile 'seabed_substrate.shp' that is in the input folder.

- Look at the attribute table. It contains many attributes, one of which is the code for the substrate types according to the EUSeaMap standard (see master report, appendix 1).
- Create a symbology according to the attribute named 'EUSeaMap' to display the different substrate types in different colours.

6.2 – Creating a geodatabase

The performance of ArcGIS is significantly better when geodatabase feature classes are used instead of shapefiles. Create a File Geodatabase in the output folder and name it *habitat_output.gdb*.

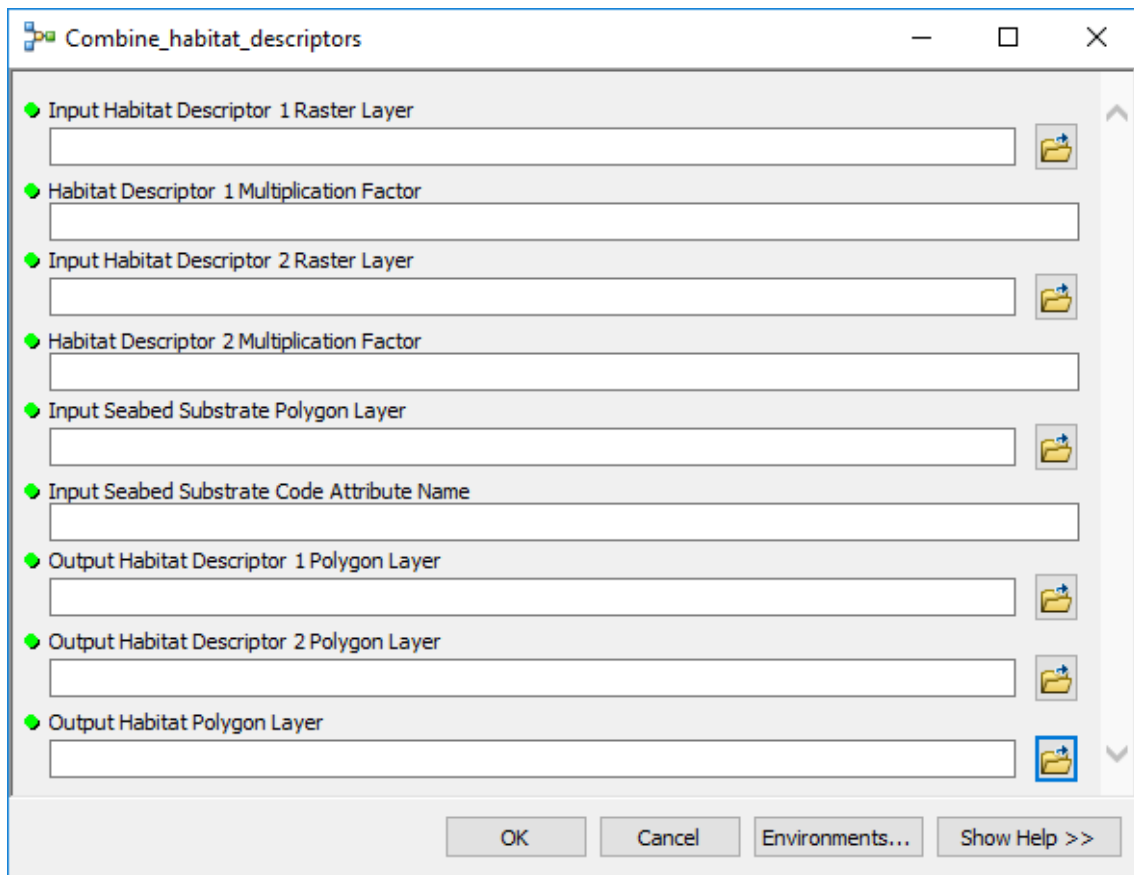
6.3 – Running the ArcGIS™ modelbuilder model

The ModelBuilder model will run under ArcGIS™ 10.0 (or higher).

The ArcGIS™ modelbuilder model will

- 1) convert the biological zone and energy raster layers into polygon layers
- 2) perform an intersection of the biological zone, energy raster and seabed substrate polygon layers
- 3) calculate a habitat code in a new column in the polygon layer resulting from the intersection. The calculation will be a sum of the biological zone, energy raster and seabed substrate codes. In order for the habitat code to follow the EUSeaMap coding convention (see master report, section 3.3.1), the biological zone code is multiplied by 100, the energy level code is multiplied by 10000, and the seabed substrate code is left as it is before summing.

- In ArcGIS™, open (i.e. double-click) the tool 'Combine_habitat_descriptors' from the toolbox 'Toolbox_EUSeaMap_ArcGIS10.tbx'. The window shown below will open. The window entries are documented in the master report, section 4.4.1.



Input Habitat Descriptor 1 Raster Layer: select the file *Biozones.tif* (folder 'output')

Habitat Descriptor 1 Multiplication Factor: type in *100*

Input Habitat Descriptor 2 Raster Layer: select the file *energy.tif* (folder 'output')

Habitat Descriptor 2 Multiplication Factor: type in *10000*

Input Seabed Substrate Polygon Layer: select the file *seabed_substrate.shp* (folder 'input')

Input Seabed Substrate Code Attribute Name: type in *EUSeaMap*

Output Habitat Descriptor 1 Polygon Layer: we will create the feature class *biozones* in the *habitat_output.gdb* geodatabase

Output Habitat Descriptor 2 Polygon Layer: we will create the feature class *energy_levels* in the *habitat_output.gdb* geodatabase

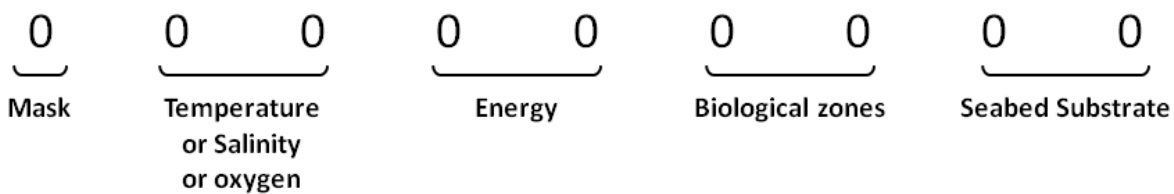
Output Habitat Polygon Layer: we will create the feature class *habitats* in the *habitat_output.gdb* geodatabase

Click OK. It may take some time to generate the output, depending on the capacity of your computer.

6.4 – Looking at the results

- In ArcGIS, open the feature class *habitats* that we have just created in the geodatabase. Each polygon corresponds to an intersection of a biological zone, an energy level and a seabed substrate.
- Open the attribute table. Look at the values in the attribute 'hab_code'
- What does the code 203070 mean in terms of biological zone, energy level and seabed substrate?

Reminder: the coding convention is as follows:



Answer: 203070 means 20 Energy=20 (i.e. 'Moderate'), Biological zone=30 (i.e. 'Deep circalittoral') and seabed substrate=70 (i.e. 'Rock')

- Make a symbology according to the attribute named 'hab_code' to display the various habitat types in different colors.

6.5 – Subsidiary question

The input biological zone and energy level layers were in raster form. Only the seabed substrate layer was in polygon form. Why didn't we just rasterise the seabed substrate layer and perform the combination in raster mode rather than in polygon mode? Doing so would have been much easier and faster!

Answer: performing the combination in raster mode would have indeed been much easier and faster, but would required converting the seabed substrate polygon layer to raster at a fixed resolution (the same resolution as the energy and the biological zone layers, i.e. approx. 100m), hence some spatial detail would have been lost.

7 - Joining the habitat polygon layer and the look-up table that crosswalks the modelled habitat codes and habitat classifications

The last step for creating the habitat map is to crosswalk the modeled habitat code with habitat types described in two habitat classifications: EUNIS 2007-11 and the MSFD broad habitat types. This is performed by the script 'habitat_map_joining_LUT.R', documented in the master report, section 4.4.2.

It will require as input the *habitat* feature class created in the previous exercise and a look-up table, the role of which is to crosswalk the individual modeled codes with a EUNIS class and a MSFD broad habitat type. The output will be a new feature class, the columns of which will be the ones that are in the look-up table.

7.1 – Looking at the look-up table

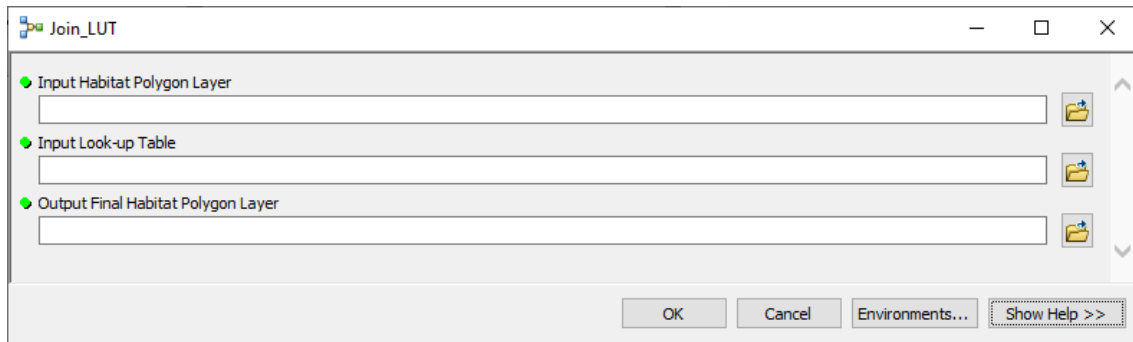
Open the file LUT_ATLPOL20190426.xlsx that is in the folder 'input'. The column contains all possible codes in our study area. Then there are several other columns that crosswalk the code with several items, including the EUNIS 2007-11 class (column 'EUNISCombD') and the MSFD broad habitat types (column 'MSFD_BBHT').

7.2 – Running the ArcGIS™ modelbuilder model

The ModelBuilder model will run under ArcGIS™ 10.0 (or higher).

The ArcGIS™ modelbuilder model will join the look-up table that crosswalks the habitat layer created in the previous exercise and the look-up table. Additionally it will run a Multipart To Singlepart and a Repair Geometry.

- In ArcGIS™, open (i.e. double-click) the tool 'Join_LUT' from the toolbox 'Toolbox_EUSeaMap_ArcGIS10.tbx'. The window shown below will open. The window entries are documented in the master report, section 4.4.2.



Input Habitat Polygon Layer: select the feature class *habitats* (geodatabase *habitat_output.gdb* geodatabase)

Input Look-up Table: select the file *LUT.dbf* (folder 'input')

Output Final Habitat Polygon Layer: we will create the feature class *habitats_final* in the *habitat_output.gdb* geodatabase

Click OK. It may take some time to generate the output, depending on the capacity of your computer.

7.2 – Looking at the results

- In a GIS software, open the feature class *habitats_final*. It has the same polygons as the habitat layer created in the previous exercise. The attribute table, though, is quite different. Take a look at it.

- Make a symbology according to the attribute named "EUNIScomBD" to display the various EUNIS habitat types in different colors.

- duplicate the layer in your software and with this layer make a symbology according to the attribute named "MSFD_BBHT" to display the various MSFD broad habitat types in different colors.

8 - Creating the habitat confidence layers

The final step is to create the habitat confidence raster layers. As the habitat layer is the result of the combination of habitat descriptor layers, the habitat confidence layer is the combination of the confidence in the habitat descriptor layers. Two habitat confidence layers are produced here because each habitat is proposed in two classifications, namely EUNIS 2007-11 and the MSFD broad habitat types, and these two classifications do not include the same habitat descriptors: EUNIS 2007-11 includes seabed substrates, biological zones and energy levels; the MSFD broad habitat types include only seabed substrates and biological zones. Therefore, the confidence raster layer for EUNIS 2007-11 will be the combination of the confidence layers for seabed substrates, biological zones and energy levels, while the confidence raster layer for the MSFD broad habitat types will be the combination of the confidence layers for seabed substrates and biological zones.

This is performed by the script 'habitat_map_calculating_confidence.R', documented in the master report, section 4.4.3.

8.1 – Looking at the csv configuration file

- Open the csv configuration file (in the folder 'config_files', file `habitat_calculating_habitat_confidence.csv`). This file describes all the elements required by the script to create confidence rasters by merging their corresponding habitat descriptor confidence. The meaning of the file columns is documented in the master report, section 4.4.3 **Erreur ! Source du renvoi introuvable.** Basically, the csv file describes for each classification the habitat descriptor confidence rasters that have to be combined.

- Check that the names listed in the column 'confidence_fileName' are correct. They should be the names of the raster files that were created for the individual habitat descriptor overall confidence in exercise 4.

8.2 – Running the script

- Open `habitat_map_calculating_confidence.R` in Rstudio.

- Assign to the script variable `workingDirectory` the path to the working directory (which is the folder that contains the 'config_files', 'input', and 'output').

- Assign to the script variable `config_csvFileName` the name of the configuration file, i.e. 'habitat_calculating_habitat_confidence.csv'

- In Rstudio, click the 'Source' button. The script starts to run. Depending on your computer capacity, it may take some time before it is finished. When it is finished, a message in the Rstudio console will say "Completed in..." followed by the duration.

8.3 – Looking at the results

- Check that all the raster have been created in the folder 'output'. There should be two of them, namely

- `habitat_eunis2007_confidence_overall.tif` for the EUNIS 2007-11 classification
- `habitat_MSFD_BBHT_confidence_overall.tif` for the MSFD benthic broad habitat types

- Take some time to look at these rasters in a GIS software.