EMODnet Thematic Lot n°2 Seabed Habitats

EASME/EMFF/2018/1.3.1.8/Lot2/SI2.810241

Start date of the project: 25/09/2019 - (24 months)

EMODnet Phase III

Constructing EUSeaMap

User Guide

The European Marine Observation and Data Network (EMODnet) is financed by the European Union under Regulation (EU) No 508/2014 of the European Parliament and of the Council of 15 May 2014 on the European Maritime and Fisheries Fund.

The European Marine Observation and Data Network (EMODnet) is financed by the European Union under Regulation (EU) No 508/2014 of the European Parliament and of the Council of 15 May 2014 on the European Maritime and Fisheries Fund.

Disclaimer

Document info

Title	Constructing EUSeaMap – User Guide
Deliverable	D1.14
Author	Mickaël Vasquez (Ifremer). Mickael.Vasquez@ifremer.fr
Dissemination level	Public
Submission date	31/08/2021

Contents

Executive summary

In the first two phases of EMODnet, the Seabed Habitats thematic project used a commercial software, namely ARCGIS^{TM}, to generate the broad-scale seabed habitat map, EUSeaMap. A key objective of phase 3 (2017-2021) was to develop a new GIS workflow that would be sound, repeatable and transferrable across sea 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 bit had to be implemented in the form of a ArcGIS^{TM} ModelBuilder model. These tools were used to construct the version 2019 and 2021 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 produce as output. A 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 use the R scripts and the model made using ArcGIS™ ModelBuilder application that have been developed by EMODnet Seabed Habitats[1] to produce the broad-scale seabed habitat map (referred to as "EUSeamap" in the report) in a semi-automated way.

The general methods that are employed highly build on a methodology that was originally developed within the framework of MESH Project whose general principles are described in Coltman et al (2008). it is strongly recommended to read that short document before going through the present one.

The way some concepts such as fuzzy laws or Generalized Linear Models (GLM) are employed in the context of EUSeaMap, the in-depth understanding of which is a prerequisite for the understanding of the present document and related scripts, will not be developed here. Further explanation on this will be found in the EUSeaMap phase 2 technical report Populus et al, 2017. In that document, it is particularly recommended to read sections 2.5 to 2.7, as well as appendixes 9 (Thresholds), 10 (Confidence) and 12.

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

2 Vocabulary and concepts

A habitat is the combination of multiple environmental characteristics. In habitat classifications these environmental characteristics are formulated via what we will refer to all along this document as **'habitat descriptors'**, which in the seafloor section of the European classification EUNIS[2] is a seabed substrate type, a biological zone or any other seabed characteristic. For example in the Atlantic EUNIS version 2007-11 describes the habitats via three habitat descriptors: energy level, biological zone and substrate type. The habitat descriptors have their own classification, e.g. in the Atlantic the energy regime habitat descriptor classification comprises 3 classes: high, moderate and low. In the 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 regions. Then, usually, depending on the region, a third habitat descriptor is used. 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 individual regions are found in Populus et al, 2017.

All the habitat descriptor classes are described in annex 1, together with the numeric code that they are assigned in the EUSeaMap GIS workflow that produces the broad-scale habitat map.

---

[1] http://www.emodnet-seabedhabitats.eu/

[2] http://eunis.eea.europa.eu/habitats.jsp

A habitat descriptor class typically comprises two **boundaries**: an upper and a lower. The upper boundary is the one that the class shares with its upper neighbour class in the habitat descriptor classification, and the lower boundary is the one that the class shares with its lower neighbour. For instance, the upper boundary of the biological zone 'shallow circalittoral' is the one that it shares with the infralittoral, while its lower boundary is the one that it shares with the deep circalittoral.

Some habitat descriptor classes only have one boundary because they are the top or the bottom of their classification, hence are bordered by only one other class. For example, the 'infralittoral' biological zone, the 'oxic' oxygen regime, or the 'high energy' energy regime only have a lower boundary because these classes are at the top level of respectively the biological zone, oxygen regime and energy regime habitat descriptor classification. In the same way the 'abyssal' biological zone, the 'anoxic' oxygen regime, or the 'low energy' energy regime only have one boundary, the upper one, because they are at the bottom of their respective classification.

3 The EUSeaMap GIS Workflow

The general workflow comprises 4 steps:

1. The habitat descriptor layers are produced individually (together with their respective confidence layer)
2. These habitat descriptor raster layers are converted to polygon layers
3. The habitat layer is produced by overlaying all the polygon habitat descriptor layers. In parallel, the individual habitat descriptor confidence layers are overlaid to produce an overall confidence layer.
4. To the habitat layer is subsequently joined the lookup table that crosswalks the modelled numeric habitat codes and the classes of habitat classifications such as EUNIS or the MSFD broad habitat types classification[3].

In EUSeaMap, some habitat descriptor layers are provided "ready-to-use", i.e. they don't need to be produced by the EUSeaMap workflow because they are produced elsewhere. An example of this in all the basins is the habitat descriptor "seabed substrate", which is directly provided by EMODnet Geology.

Figure 1 illustrates the workflow with an example in which 4 habitat descriptors would be considered, 3 of which would be produced by the workflow (HD2, 3 and 4) and one of which (HD1) would be provided ready-to-use, already classified.

The sections 3.1, 3.2, 3.3.1 and 3.3.2 describe each of the three abovementioned steps.

[3] COMMISSION DECISION (EU) 2017/848

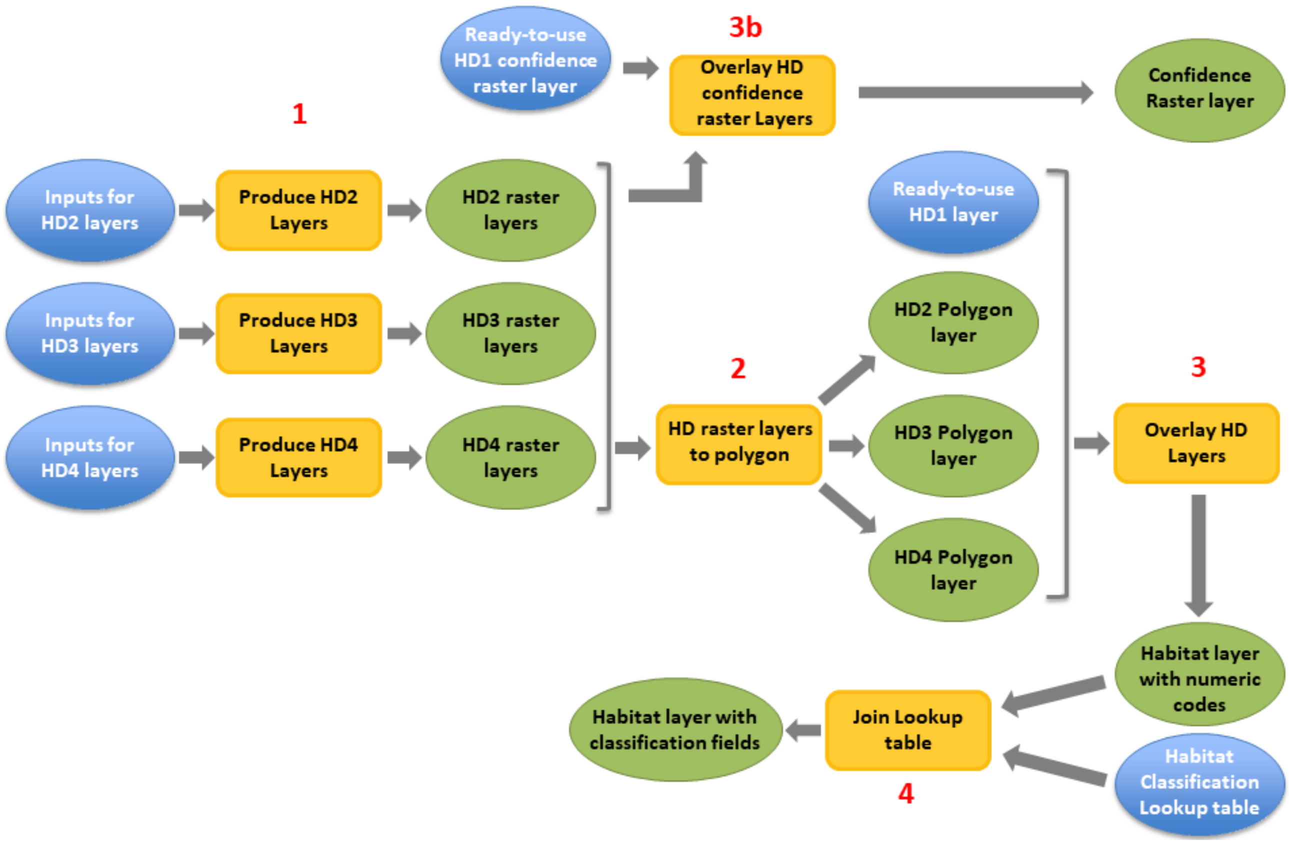


Figure 1: The EUSeaMap general GIS workflow illustrated with an example of 4 habitat descriptors (HD). Layers for HD2, HD3 and HD4 need to be produced, while HD1 layer is provided "ready-to-use" (because e.g. it is produced elsewhere). HD2, HD3 and HD4 layers are 1) produced as rasters (together with a confidence layer for each), 2) converted to polygon, 3) overlaid to create the habitat layer. 3b: The HD confidence raster layers are also overlaid in a single overall confidence raster. After the HD layer overlay, the lookup between numeric habitat codes and various classification classes is joined to the habitat layer (4).

3.1 Production of a habitat descriptor layer and its confidence layer

3.1.1 Workflow

The production of a habitat descriptor layer and its confidence layer comprises 2 steps (figure 2):

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

In sections 3.1.2 and 0 these two steps are further detailed.

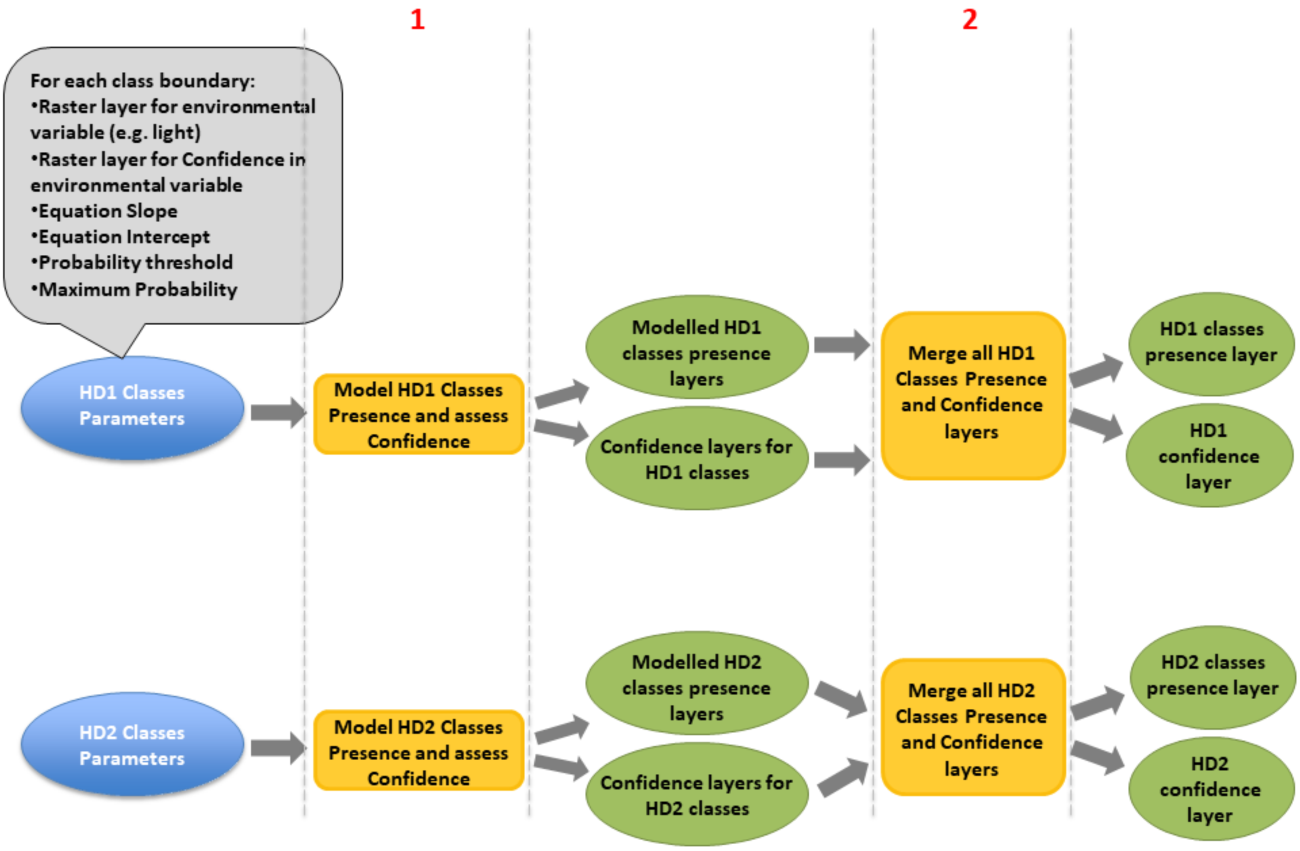


Figure 2: Habitat descriptor (HD) layers production workflow for 2 HDs, HD1 and HD2. For each HD: 1) for each HD class, presence layers are modelled and confidence is assessed (hence two output layers are produced, respectively for class presence and confidence in class presence); 2) individual HD class presence and confidence layers are merged in one 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 modeled using either Generalized Linear Models (GLM) or fuzzy laws. Details are given in Populus et al (2017) on how these GLMs and fuzzy laws are fitted in the approach developed for EUSeaMap. With the EUSeaMap approach the GLM/fuzzy law is used to predict the presence probability of a habitat descriptor class given the value of a unique environmental predictor variable (e.g. the light quantity available at the seabed is the 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 abovementioned workflow (see figure 2) for a probability and presence raster to be computed using a GLM/fuzzy law.

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

The equation is in the form:

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

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

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

In order 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 (if any) and the lower (if any) boundary.

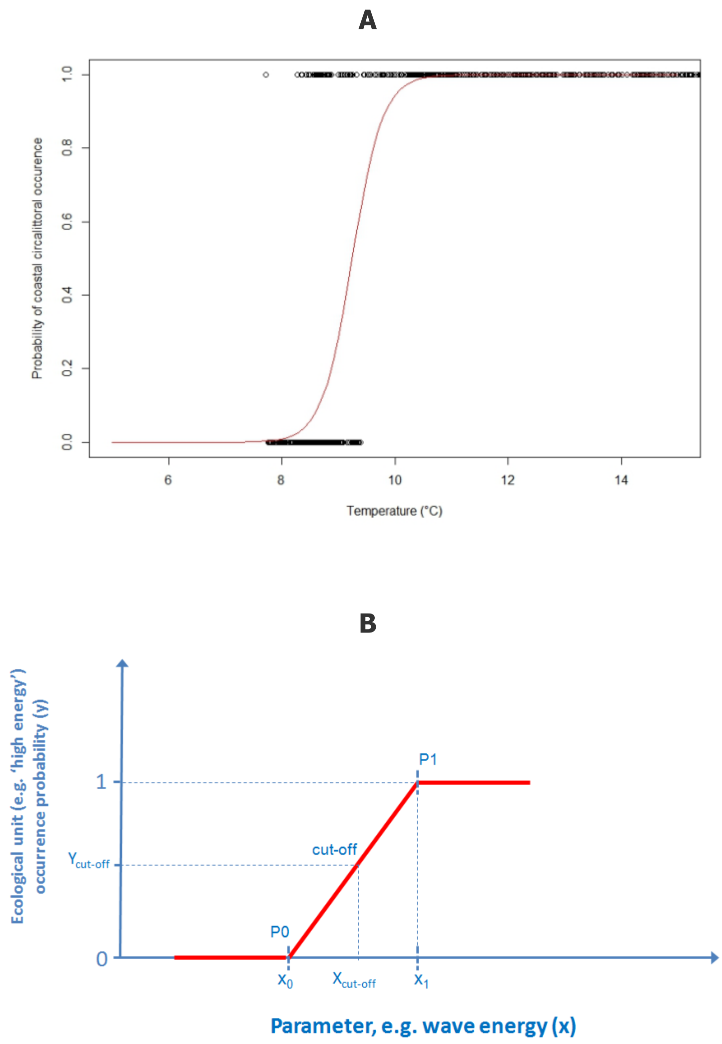


Figure 3: The 2 types of laws used to model the habitat descriptor class spatial distribution: GLM (A) and fuzzy (B) – In Populus et al, 2017

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

To summarize, for a GLM/fuzzy lay to be used in the workflow, 3 values are required 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 the way the slope, intercept and probability threshold can be calculated from the fuzzy law control points.

Assessing the confidence in the habitat descriptor classes

As described in Populus et al (2017), the confidence that a habitat descriptor class is correct depends on two items:

1) the confidence **in values of the environmental layers** that are 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 for the provision of that layer is provided in Populus et al (2017), section 2.7. Note that for its polygon seabed substrate layer, EMODnet Geology provides a confidence index (the confidence is provided at polygon level). So does EMODnet Bathymetry for its DTM (the confidence is provided at cell level).

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

Table 1: rules used to classify a continuous probability in 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); probability threshold is the probability cut-off value; range= probability max - probability threshold

Confidence per cell	Rule
High	1.6 x range ≤ probability ≤ probability max
Moderate	1.2 x range ≤ probability < 1.6 x range
Low	probability threshold ≤ probability < 1.2 x range

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

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

Confidence per cell	Rule
High	0.8 ≤ probability ≤ 1
Moderate	0.6 ≤ probability < 0.8
Low	0.5 ≤ probability < 0.6

For each habitat descriptor class boundary, the workflow automatically creates a raster layer of the confidence based on the probability by performing the classification of the probability in high, moderate or low confidence. For this to be possible, additionally to the 3 abovementioned values (i.e. slope, intercept and probability threshold) the user will need to provide the **maximum probability** as input to the script.

In a last step, for each habitat descriptor class boundary the workflow creates one single raster layer of confidence 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 logic used for the combination is described in table 3.

Table 3: logic used for combining 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 Workfkow

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

- The environmental predictor variable raster layer (e.g. light)
- 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 that layer all the cells have the same value, which is assigned according to the habitat descriptor class coding convention that is in annex 1.
- a categorical raster layer for confidence in the habitat descriptor class occurrence. The values of the cell are one of the following: 1 (low confidence), 2 (moderate confidence) and 3 (high confidence)

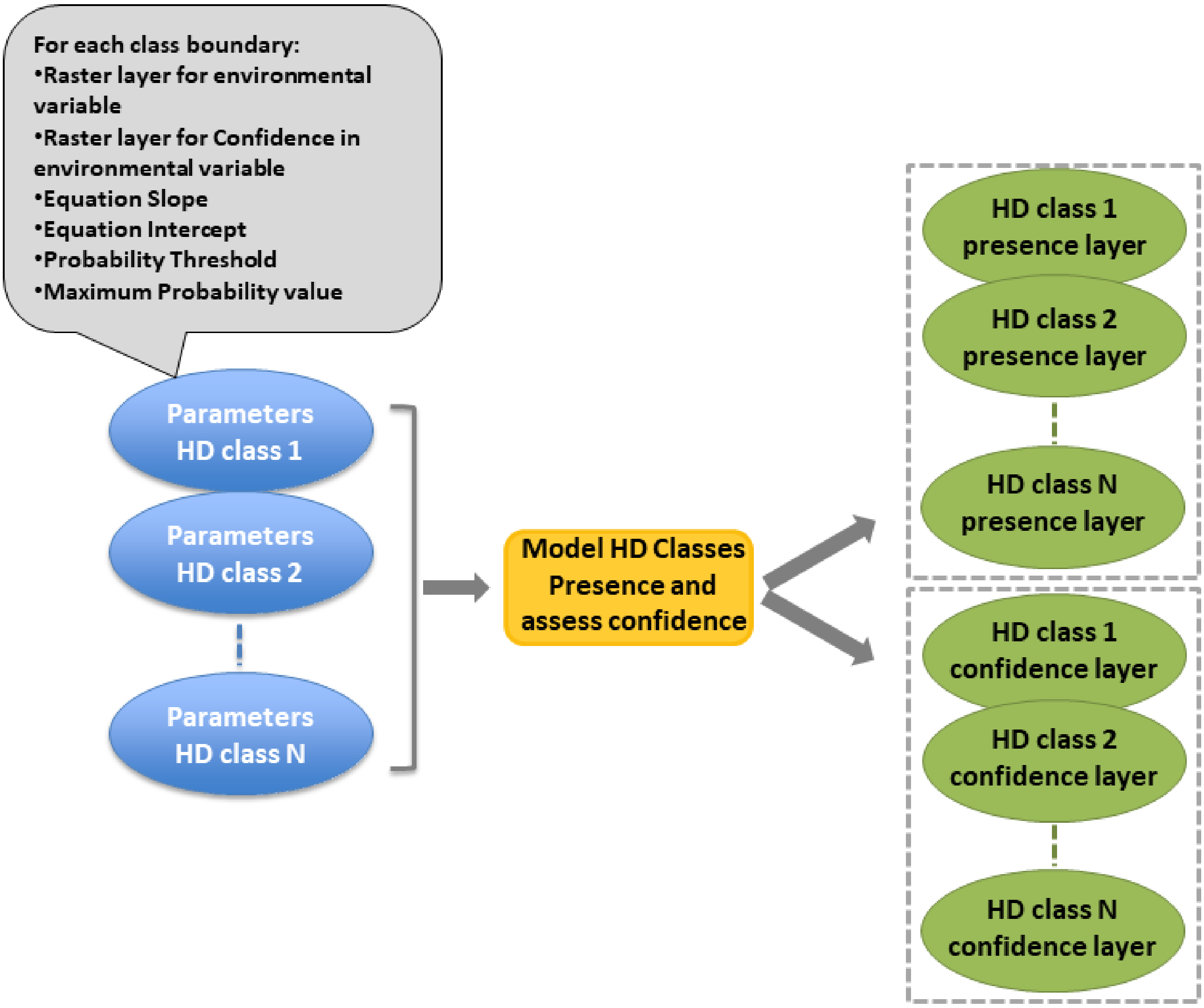


Figure 4: Workflow for the production of the individual class presence and confidence layers for a Habitat descriptor (HD). For each boundary of the class, the inputs are the environmental variable raster and its confidence raster, the slope and intercept of the GLM/fuzzy law, the probability threshold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

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

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

An example of this is illustrated in figure 5. In the Black sea, the slope and intercept values that define the Infralittoral lower boundary are not the same everywhere. In that specific case where GLM/fuzzy law parameters vary spatially, in lieu of a constant value for the slope and intercept the workflow will require **a raster layer** as input for both.

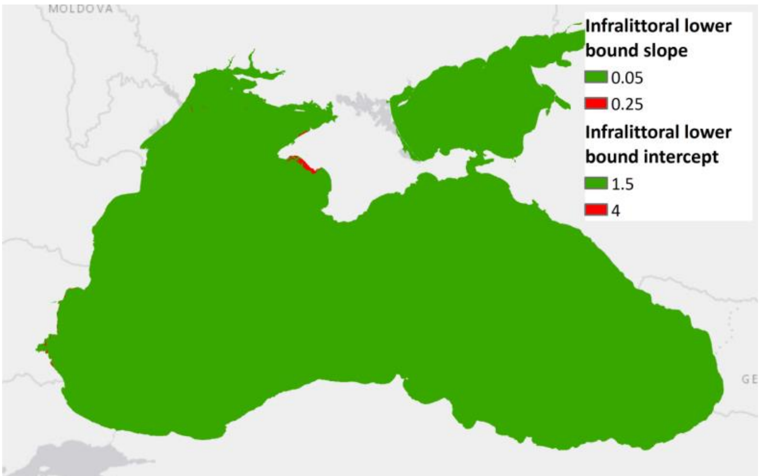


Figure 5: Example of GLM/fuzzy slope and intercept varying spatially: the infralittoral lower boundary in the Black sea

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

As shown in figure 2, in step 2 of the workflow all the habitat descriptor class layers and the confidence layers produced in step 1 are merged.

The output is a single habitat descriptor raster layer that comprises all the classes (e.g. all the biological zones) and a single confidence raster layer for the habitat descriptor.

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) comprises the overlay of the habitat descriptors in vector mode, all the habitat descriptor layers have to be converted in polygon layers (figure 1, step 2).

3.3 Creating the habitat map

The final step of the general workflow is the creation of the habitat map. In a first step the habitat descriptor layers are all combined via a geometric intersection to produce the habitat layer (see figure 1, step 3). In a second step, to this habitat layer's attribute table is joined the lookup table that correlates the habitat codes that result from the union and the classes of habitat classifications such as EUNIS or the MSFD predominant habitat classification (see figure 1, number 4). The following describes these 2 steps.

3.3.1 Combining the habitat descriptor layers

General principle

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

For each polygon the values contained by 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 coherence across regions, it has been agreed that all the habitat descriptor classes would be coded on 2 digits except, due to technical limitations, the classes of the mask habitat descriptor that are coded on one digit. It has also been agreed that, as illustrated in figure 6, the 2 first digits from the right would be devoted to the seabed substrate, and then, from right to left, would follow the biological zone 2 digits, the energy 2 digits, the (temperature, oxygen or salinity) 2 digits, and the mask digit.

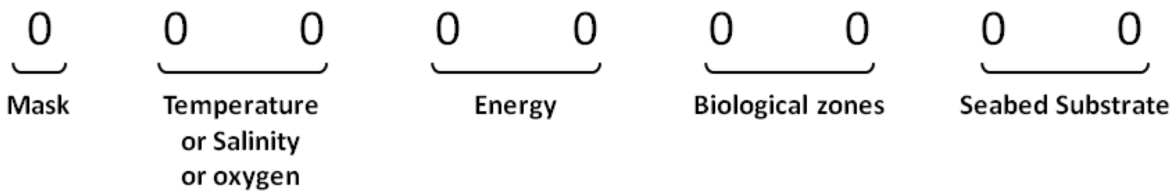


Figure 6: The coding convention for modelled habitat codes

Thus, in order for the abovementioned sum of the habitat descriptor codes to output values coded in such a way, each habitat descriptor code have to 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 last step is joining the habitat layer and the the look-up table, which is that table that crosswalks the abovementioned digital habitat codes and the habitat classes described in 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: Screenshot of an 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 are all available on Github:

https://github.com/emodnetseabedhabitats/EUSeaMap_creation

2 R scripts are proposed for the 2 steps that allow for the production of a habitat descriptor layer and its confidence layer (section 3.1). For the other tasks of the general workflow, i.e. converting the habitat descriptor raster layers in polygon layers (section 3.2) and combining them to create the habitat layer (section 3.3.1), we recommend the use of ARCGIS™. The various R package functions tested for raster conversion into polygons did not prove sufficiently efficient when the number of polygons is high (which is the case for EUseaMap). Other tests of *ad hoc* R packages for intersecting polygon layers also demonstrated some limitations (for further details see Vasquez et al 2020, annex 4). A modelBuilder model that enables to perform these 2 tasks together is proposed. For the last step, joining the habitat layer and the look-up table (section 3.3.2), a R script is proposed.

The scripts and the modelBuilder model are documented in section 4.3 (Script for the production of habitat descriptor layers) and 4.4 (Creating the habitat layer and overall confidence). In figure 8 is represented the general workflow. In brackets are references to the sub-section that encompasses the different tasks.

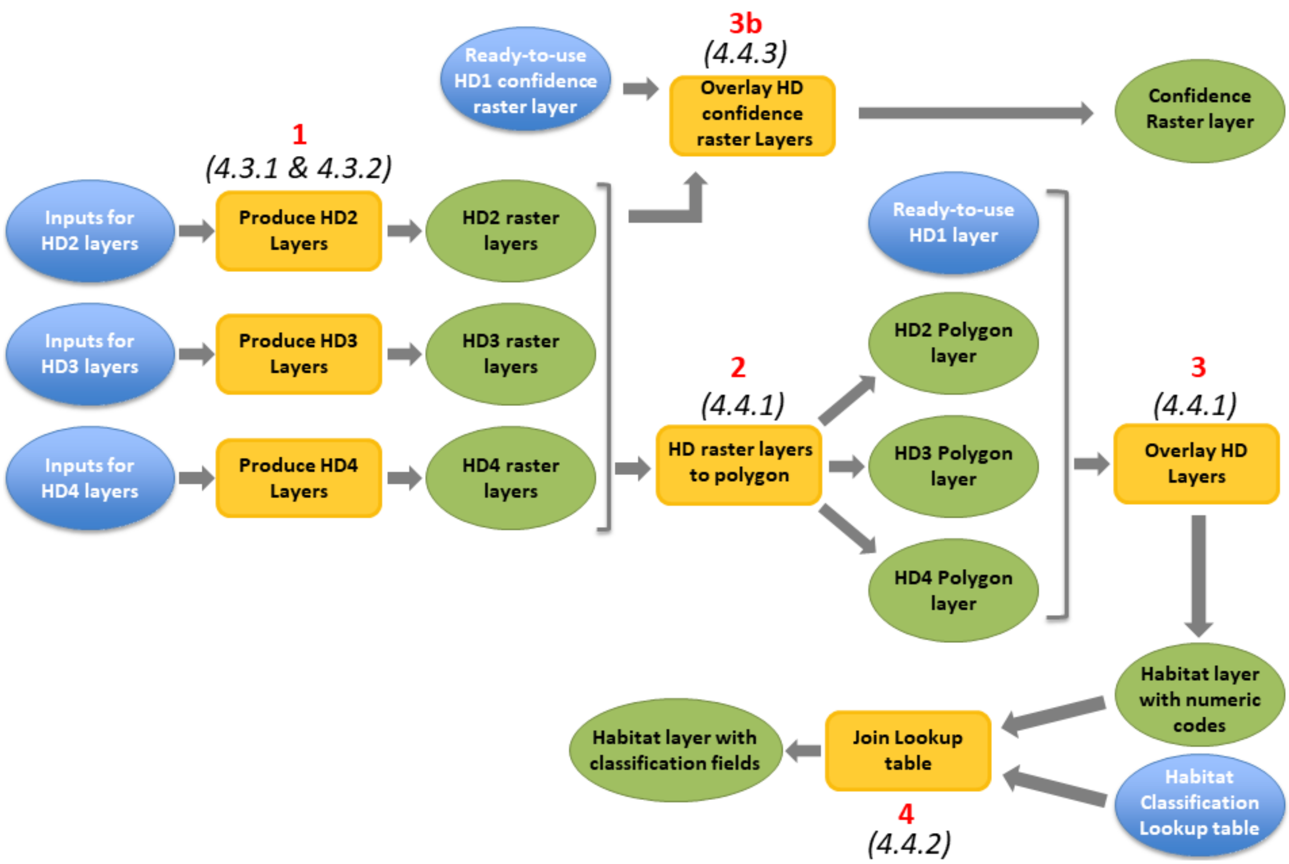


Figure 8: general workflow and sub-section (in brackets) that documents the script or modelbuilder model corresponding to the various tasks

4.1 Prerequisites

4.1.1 System requirements

RAM/Memory: the R package *raster* is fond of memory. Minimum requirement is 8 GB, recommended 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. It is also recommended to use the software Rstudio, which is more user-friendly than the native R editor.

R : https://cran.r-project.org/mirrors.html

R packages

The following R packages are required:

- raster
- rgdal
- readxl

ArcGIS™

The model builder model presented in section 4.4 requires 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 spatially match

A crucial requirement of the "raster" R package is that all rasters that are overlaid must fully spatially match, i.e. must strictly have the same spatial extent and resolution. As a result, as most of the scripts perform raster overlay, all the input rasters to the scripts will have to spatially match. If this requirement is not fulfilled, 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 have been developed with the objective that they can be run by non-R experts, i.e. as a "black box". The majority of the inputs passed to the scripts are formulated via configuration files in csv (comma-separated values) format (figure 9). Very few parameters, such as the working directory, are 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 fine, 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 (chunksize, maxmemory, workingDirectory, config_csvFileName, habitat_descriptor_shortName)

In the script description sheets (sections 4.3, 4.4.2 and 4.4), inputs passed via the configuration file will be referred to as "Inputs" and the parameters formulated in the script will be referred to as "script parameters".

4.2.3 Folder structure

For a given study area (e.g. Mediterranean Sea) for the script to run the user will need to follow a standard folder structure (figure 11).

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

- **config_files**: folder in which the script will expect the csv configuration files to be stored
- **input**: folder in which the script will expect the input files to be stored
- **output**: folder in which the script will write the outputs

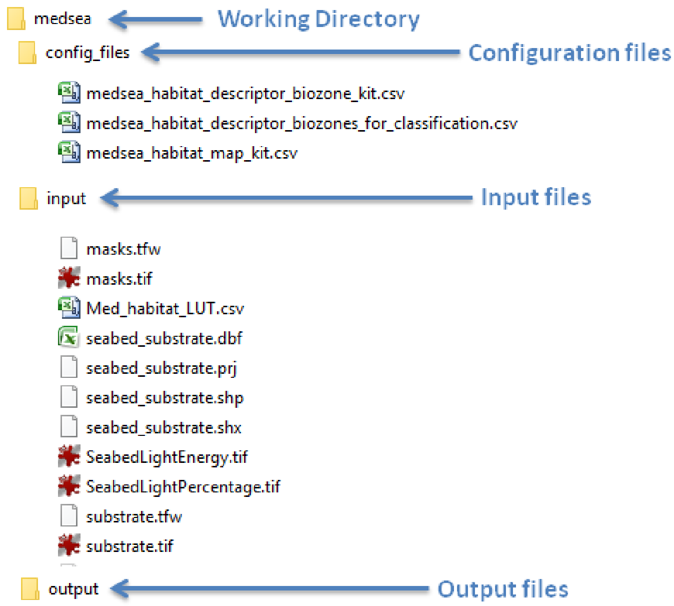


Figure 11: The folder structure

4.2.4 Tuning the load memory size in R

The "raster" R package uses a lot of memory, as by default it performs all calculations in memory. With the default options, depending on the RAM available your computer might not be able to run the scripts with the default options. For that reason all the scripts that use the Raster package have 2 script parameters, namely 'chunksize' and 'maxmemory', that enable to tune the memory used by the package while calculating. 'maxmemory' is the maximum number of pixels that is loaded in memory during a raster calculation. By default, both values are set to 1e+09. Depending on the capacity of your computer, these values might be too high and, as a result, the script might not be able to run (memory error message). In that case, decrease the values by a factor 10 (1e+08, then 1e+07, etc) until the script starts running.

4.3 Script 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	04/2021
Descr.	Models the classes' spatial distribution of <u>one</u> habitat descriptor according to the ecological law parameters that characterise the class boundaries. The script models the spatial distribution for each class whose input parameters are described in the csv configuration file. In that configuration file are indicated 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. For further details on these parameters, see section 3.1. 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. If it's a GLM, the equation used for the probability raster calculation will be $P(\text{predictor}) = e^{(\text{slope}*\text{predictor}+\text{intercept})} / (1+e^{\text{slope}*\text{predictor}+\text{intercept}})$ If it's a fuzzy law, it will be $P(\text{predictor}) = \text{slope}*\text{predictor}+\text{intercept}$
Workflow	Figure 2, (see number 1) and Figure 4
Inputs	All inputs are described via a csv configuration file. The csv file will contain a row for each habitat descriptor class that has to be modelled. If the habitat descriptor class has one boundary only, the csv file cells for the boundary parameters (equation type) have to be left blank Crucial: all input raster files have to be in the input folder

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 numeric value. The raster cells will indicate the value of the threshold **lower_boundary_max_probability** — N Or C — For the class lower 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 **(*) C=character, N=numerical**
Outputs	**For each habitat descriptor class** described in the csv file (i.e. each row of the configuration file), the outputs are as follows: • 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" All the outputs will be created in the folder 'output'.
Scripts parameters	Name — Type* — Description **chunksize** — N — See section 4.2.4 **maxmemory** — N — See section 4.2.4 **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 **(*) C=character, N=numerical , B=Boolean**
How to run it?	• 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
R functions used	• calc (raster package) for the calculation of the probability when the input for slope, intercept is provided via a constant value (no spatial variation of these parameters) • overlay (raster package) for the calculation of the probability when the input for slope, intercept is provided via a raster (due to spatial variation of these parameters)

4.3.2 Merging the individual habitat descriptor classes in a single layer

<table>
<tr><td>Script name</td><td>habitat_descriptor_merging_classes.R</td></tr>
<tr><td>Author</td><td>Mickaël Vasquez</td></tr>
<tr><td>Date</td><td>04/2021</td></tr>
<tr><td>Descr.</td><td>For one or several habitat descriptor(s)
• All the spatial distribution rasters for each individual class are merged in one single raster layer
• All the confidence rasters for each individual class are merged in one single raster layer
• All the confidence based on bathymetry rasters for each individual class are merged in one single raster layer
• All the habitat presence probability rasters for each individual class are merged in one single raster layer</td></tr>
<tr><td>Workflow</td><td>Figure 2, (see number 2)</td></tr>
<tr><td>Inputs</td><td>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).
<table>
<tr><th>Column Name</th><th>Type*</th><th>Description</th></tr>
<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>
</table></td></tr>
</table>

	overall_confidence_fileName	C	The name of the input raster file that characterises the overall confidence in the habitat descriptor class occurrence
	folder	C	The name of the folder that contains the input raster file. Classically the output folder
	(*) C=character, N=numerical		
Outputs	For each habitat descriptor described in the csv file, creates (all the outputs will be created in the folder 'output'): • 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	**Name**	**Type***	**Description**
	chunksize	N	See section 4.2.4
	maxmemory	N	See section 4.2.4
	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
	(*) C=character, N=numerical, B=boolean		

How to run it?	• 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
R functions used	• merge (raster package) for merging the spatial distribution rasters • overlay (raster package) for merging the probability rasters

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.	Requires ArcGIS™ 10.0 (or higher). Creates the habitat layer, which comprises: 1) Converting in polygon layers the modelled habitat descriptors 2) Overlaying (by intersection) all the habitat descriptors 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 following sum: 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) 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 require changes. Note: ArcGIS™ is used here because with layers having lots of detail tested R packages (raster, sf) proved efficient neither for Raster To Polygon conversion nor for intersection of polygon layers
Worklow reference	Figure 1, (numbers 2 and 3)
Inputs	**Model Variable name** / **Type** / **Description** **Input Seabed Substrate Polygon Layer** / Shapefile / The input Seabed Substrate polygon layer

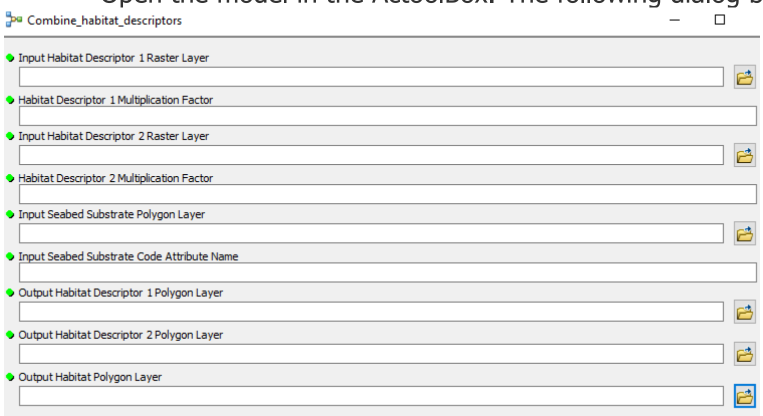
Input Seabed Substrate Code Attribute Name	String	The name of the seabed substrate polygon layer attribute that describes the seabed substrate code
Input Habitat Descriptor 1 Raster Layer	Raster dataset	The input habitat descriptor 1 raster layer
Habitat Descriptor 1 Multiplication Factor	Variant	The 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
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)

Outputs

Model Variable name	Type	Description
Output Habitat Descriptor 1 Polygon Layer	Feature class	The habitat descriptor 1 polygon obtained by conversion of raster to polygon
Output Habitat Descriptor 2 Polygon Layer	Feature class	The habitat descriptor 2 polygon obtained by conversion of raster to polygon
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 following sum: seabed substrate code + (habitat descriptor 1 code x habitat descriptor 1 Multiplication Factor) + (habitat descriptor 1 code x habitat descriptor 1 Multiplication Factor)

How to run it?

- Open the model in the ActoolBox. The following dialog box will appear



If an update of the tool is required, (e.g. add an habitat descriptor), edit it

Main ArcGIS™ tools used	• 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

Script name	habitat_map_joining_LUT.R
Author	Mickaël Vasquez
Date	04/2021
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 excel format) that crosswalks each habitat code with various habitat classification (e.g. EUNIS). The input habitat shapefile will have to be in the folder 'output' of the working directory
Worklow reference	Figure 1, (see number 4)
Outputs	A shapefile containing the attributes of the look-up table
Scripts parameters	See table below
How to run it?	• Open the script in RStudio • Edit the script parameters described above • Click Button 'Source' in RStudio

Name	Type*	Description
workingDirectory	C	Full path to the working directory
habitat_shapefile_name	C	Name of the input habitat shapefile (with no ".shp" file extension)
modelled_habitat_code_column	C	Name of the habitat shapefile column that contains the habitat codes
LUT_table_excel_file	C	Full path to the excel file that contains the look-up table
LUT_habitat_code	C	Name of the look-up table column that contains the habitat codes
output_habitat_shapefile_name	C	Name of the shapefile that will be created

(*) C=character, N=numerical

R function used	merge (native R function)

4.4.3 Creating the habitat confidence layers

<table>
<tr><td>Script name</td><td>habitat_map_calculating_confidence.R</td></tr>
<tr><td>Author</td><td>Mickaël Vasquez</td></tr>
<tr><td>Date</td><td>04/2021</td></tr>
<tr><td>Descr.</td><td>For one or several habitat classifications, creation of 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 each cell, the minimum confidence value of the input habitat descriptor confidence layers is kept. Why is this done for each classification? Because 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).</td></tr>
<tr><td>Workflow</td><td>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 whether the confidence has to be calculated for one or several map(s).</td></tr>
<tr><td>Inputs</td><td>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).
<table>
<tr><th>Column Name</th><th>Type*</th><th>Description</th></tr>
<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>
<tr><td colspan="3">(*) C=character, N=numerical</td></tr>
</table></td></tr>
<tr><td>Outputs</td><td>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”</td></tr>
</table>

<table>
<tr><td>Scripts parameters</td><td>
<table>
<tr><th>Name</th><th>Type*</th><th>Description</th></tr>
<tr><td>chunksize</td><td>N</td><td>See section 4.2.4</td></tr>
<tr><td>maxmemory</td><td>N</td><td>See section 4.2.4</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 colspan="3">(*) C=character, N=numerical</td></tr>
</table>
</td></tr>
<tr><td>How to run it?</td><td>• 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</td></tr>
<tr><td>R functions used</td><td>• stack (raster package)
• min (raster package)</td></tr>
</table>

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	04/2021
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: • One for the combined wave/current energy • One for the confidence in the combined wave/current energy
Scripts parameters	(see below)

Scripts parameters

Name	Type*	Description
workingDirectory	C	*Full link to the working directory (i.e. the directory that contains the habitat shapefile)*
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?	• Open the script in RStudio • Edit the script parameters described above • Click Button 'Source' in RStudio		
R function used	• stack (raster package) • max (raster package) • overlay (raster package)		

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
0	No information
10	Infralittoral
20	Circalittoral
30	Deep circalittoral
40	Bathyal
50	Abyssal
60	Upper bathyal
61	Atlantic upper bathyal NORTH
62	Atlanto-Arctic upper bathyal
63	Atlantic upper bathyal
64	Atlanto-Mediterranean upper bathyal
65	Mid Bathyal
66	Atlanto-Arctic upper bathyal in the NORTH Arctic
67	Arctic mid bathyal
68	Atlantic mid bathyal
69	Atlanto-Mediterranean mid bathyal
70	Lower Bathyal
71	Arctic lower bathyal
72	Atlanto-Arctic lower bathyal
73	Atlantic lower bathyal

2-digit code	Name
74	Atlanto-Mediterranean lower bathyal
75	Upper Abyssal
76	Arctic upper abyssal
77	Atlanto-arctic upper abyssal
78	Atlantic upper abyssal
79	Atlanto-Mediterranean upper abyssal
80	Mid abyssal
81	Arctic mid abyssal
82	Atlanto-arctic mid abyssal
83	Atlantic mid abyssal
84	Atlanto-Mediterranean mid abyssal
85	Lower abyssal
86	Arctic lower abyssal
87	Atlanto-arctic lower abyssal
88	Atlantic lower abyssal
89	Atlanto-Mediterranean lower abyssal
91	Atlantic upper bathyal in South Arctic
96	Atlanto-Arctic upper bathyal in South Arctic

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 / (x1-x0)

b = -x0 / (x1-x0)

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.

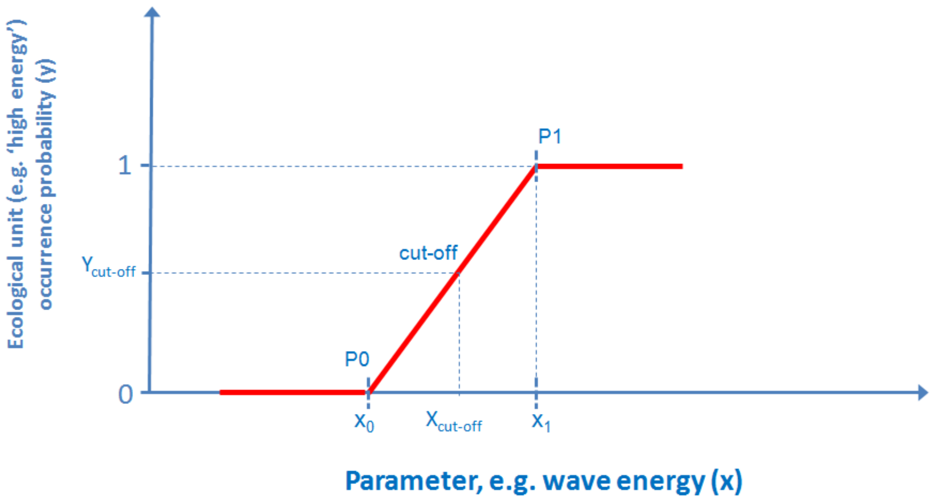


Figure 2.1. The fuzzy classifier shape is governed by 2 control points P0 and P1. The cut-off is the point whose y coordinate is the probability value above which the habitat descriptor class will be classified as present and below which it will be classified as absent.

Usually, in a threshold analysis, the values that are worked out are X_0, X_1 and $x_{cut\text{-}off}$, while the GIS workflow requires as input the slope, the intercept and the probability threshold ($Y_{cut\text{-}off}$ 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 $X_{cut\text{-}off}$. 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

1 / (X1-X0)   -X0 / (X1-X0)   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 $X_{cut\text{-}off}$

8 Appendix 3: Hands-on training

In exercises 1 to 3 we are going to 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 will be gathered in 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 that 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 how things are organised in the file. The meaning of the file columns is documented in the master report, section 4.3.1.

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

<u>Questions</u>

- 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 ʎ/h, where ʎ=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 circalitttoral 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

- 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).

- In the confidence in variable rasters, what are the possible values? What is their meaning?

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

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 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.

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 as biological zones listed in the file 'habitat_descriptor_modelling_biozones.csv' have been created. The tif files have been named according to what is written in the column 'shortName' of the csv file. Alongside, for each biological zone 2 rasters have been created, i.e. a raster for the confidence based on the probability (name of the tif is in the form 'shortName'_confidence_based_on_proba) and another one for the overall confidence (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'). So there should be the following files in the folder 'output':

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 coded assigned to each class spatial distribution raster pixels are those indicated in the column 'code' 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 items required by the script to model the current energy levels. The meaning of the file columns is documented in the master report, section 4.3.1.

- Check that all the input parameters (class code, equation type, slope, intercept, probability thresholds, max probability) defined for each energy levels 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 (reminder: 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 'ouput ').

- 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 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.

2.4 – Looking at the results

a) Check that all the raster have been created. There should be the following files in the folder 'output':

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 coded assigned to each class spatial distribution raster pixels are those indicated in the column 'code' 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 4.3.1.

- 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 'ouput '.

- 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 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.

3.4 – Looking at the results

a) Check that all the raster have been created. There should be the following files in the folder 'output':

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 coded assigned to each class spatial distribution raster pixels are those indicated in the column 'code' 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 are going to merge all the class raster layers in a single raster layer.

The script that performs 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 folder 'config_files', file habitat_descriptor_merging_inputs.csv). This file describes all the items required by the script to create the habitat descriptor rasters by merging their respective class rasters. The meaning of the file columns is documented in the master report, section 4.3.24.3.1.

We are willing to create for each habitat descriptor 3 rasters, i.e fors.

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

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

We need to provide as input to the script the name of the rasters files that we have created in exercise 1, 2 and 3 for it to merge these files. That's what the columns 'class_fileName', 'based_on_proba_confidence_fileName' and 'overall_confidence_fileName' are here for.

Indeed, for each habitat descriptor (the name of which is in column 'habitat_descriptor_shortName'), the rasters created in exercises 1, 2 and 3 are listed in 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 column 'folder' contains the name of the folder that contains all these files. As they were all created by us using the script habitat_descriptor_modelling.R, 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 'ouput ').

- 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"
```

- Initialise 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 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.

4.3 – Looking at the results

a) Check that all the raster have been created.

For each habitat descriptor listed in the csf file, 3 rasters have been created, 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, there should be the following files in the folder 'output':

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 in order to have a unique current/wave combined energy level layer. 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"
```

- Assign 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 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.

5.2 – Looking at the results

a) Check that all the raster have been created. 2 files should have been created in the folder 'output': 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

It is time now to create the habitat map. The first step is to combine the habitat descriptor layer 'seabed substrate', which is a polygon layer provided 'ready-to-use' 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 keep all the spatial detail of the seabed substrate layer, the combination is done in polygon mode. As a result, prior to combining, the raster layers need to be converted to polygon. The combination is then performed by intersecting the layers.

Unfortunately, this is not done via a R script. The reason is that R packages have proved efficient neither for raster to polygon transformation nor for polygon layer intersection when the number of polygons is high, which is usually the case for EUSeaMap. Therefore, we use here ArcGIS™. 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.

- Take a look at the attribute table. It contains a lot of attributes, one of which is the code of the substrate types according the the EUSeaMap standard (see master report, appendix 1).

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

6.2 – Run the ArcGIS™ modelbuider model

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

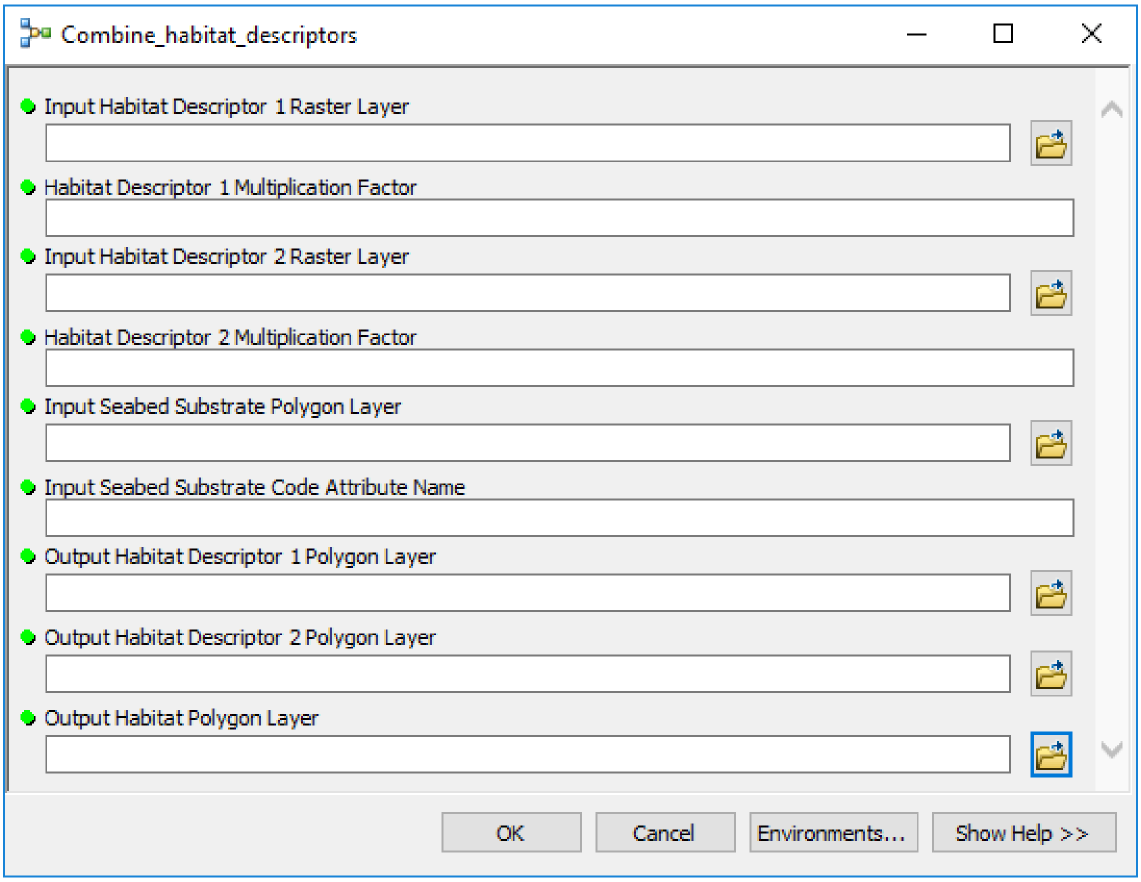
The ArcGIS™ modelbuider model will enable to

1) Convert the biological zone and energy raster layers to polygon layers

2) perform an intersection of the polygon layers biological zone, energy raster and seabed substrate

3) In the polygon layer resulting from the intersection, a habitat code will be calculated in a new column. The calculation will be a sum of the biological zone, energy raster and seabed substrate codes. For the habitat code to follow the the EUSeaMap coding convention (see master report, section 3.3.1), prior to summing, the biological zone code will be multiplied by 100, the energy level code by 10000, and the seabed substrate code will be kept as it is.

- In ArcGIS™, from the toolbox 'Toolbox_EUSeaMap_ArcGIS10.tbx' open (i.e. double-click) the tool 'Combine_habitat_descriptors'. The window illustrated below opens. 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 file *biozone_polygon.shp* in the 'output' folder

Output Habitat Descriptor 2 Polygon Layer: we will create the file *energy_polygon.shp* in the 'output' folder

Output Habitat Polygon Layer: we will create the file *habitat_polygon.shp* in the 'output' folder

Click OK. The creation of the output may take some time depending on your computer capacity.

6.3 – Looking at the results

- In a GIS sotfware, open the shapefile 'habitat_polygon.shp' that we have just created the output folder. 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:

$$\underbrace{0}_{\text{Mask}}\quad \underbrace{0\quad 0}_{\substack{\text{Temperature}\\ \text{or Salinity}\\ \text{or oxygen}}}\quad \underbrace{0\quad 0}_{\text{Energy}}\quad \underbrace{0\quad 0}_{\text{Biological zones}}\quad \underbrace{0\quad 0}_{\text{Seabed Substrate}}$$

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.

6.4 – 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 seabeb 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 we would have lost some spatial detail.

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 classications: 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 shapefile 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 shapefile, 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 script

- Open habitat_map_joining_LUT.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 R variable `habitat_shapefile_name` the name of input habitat shapefile without the extention ".shp" (beware: that file has to be in the folder 'ouptut' of the working directory, which in our case is ok because we created it there):

```
habitat_shapefile_name<-"habitat_polygon"
```

- Assign to the R variable `modelled_habitat_code_column` the name of the habitat shapefile column that describes the habitat code: the modelBuilder model named it "hab_code"

```
modelled_habitat_code_column<-"hab_code"
```

- Assign to the R variable `LUT_table_excel_file` the path for the look-up table (relative to the working directory).

```
LUT_table_excel_file<-"input/LUT_ATLPOL20190426.xlsx"
```

- Assign to the R variable `LUT_habitat_code` the name of the look-up table column that describes the habitat codes

```
LUT_habitat_code<-"ModelCode"
```

- Assign to the R variable `output_habitat_shapefile_name` the name of output habitat shapefile

```
output_habitat_shapefile_name<-"final_habitat_polygon"
```

- 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, particularly for opening the shapefile. When it is finished, a message in the Rstudio console will say "Completed in…" followed by the duration.

7.2 – Looking at the results

- In a GIS software, open the shapefile 'final_habitat_polygon.shp' that is in the output folder. 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 last 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. Here we will make two habitat confidence layers because each habitat is proposed in two classifications, namely EUNIS 2007-11 and the MSFD broad habitat types, and these two classifications do not involve the same habitat descriptors: EUNIS 2007-11 involves the seabed substrates, the biological zones and the energy levels; the MSFD broad habitat types involve the seabed substrates and the biological zones only. Hence, 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 items required by the script to create confidence rasters by merging their respective habitat descriptor confidence. The meaning of the file columns is documented in the master report, section 4.4.34.3.1. Basically, the csv file describes for each classification the habitat descriptor confidence rasters that have to me 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 ‘ouput ‘).

- 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.