EC contract no. MARE/2012/10

Knowledge base for growth and innovation in ocean economy: Assembly and dissemination of marine data for seabed mapping – Lot #3

EUSeaMap

Broad Scale Seabed Habitat Mapping Technical Guide

Based on the example of the Black Sea basin

Author: Mickaël Vasquez (orcid.org/0000-0002-7288-2394)

August 2017 - Version 1

1. Introduction

This technical guide aims to help the reader understand, with a focus on the Black Sea basin, what composes the GIS workflow models that were developed as part of EMODnet Seabed Habitats[1] phase 2 to produce a broad scale seabed habitat map, EUSeaMap.

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). We highly recommend this short document be read before going through the current one.

Further explanation on the way we employ some concepts such as fuzzy laws or Generalized Linear Models (GLM) in the context of broad scale seabed habitat mapping will be found in the EUSeaMap phase 2 technical report Populus et al (2017). We particularly recommend sections 2.4 and 2.5, as well as appendixes 9 and 12.

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 the habitat "high energy infralittoral rock" is described via three habitat descriptors: an energy regime (high energy), a biological zone (infralittoral) and a substrate type (rock). 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 Black Sea, where the EUNIS classification is not applicable, the habitat classification that was used comprises 3 habitat descriptors: the oxygen regime (oxic, suboxic, anoxic), the seabed substrate type (mud, sand, coarse sediment, mixed sediment, sandy mud, muddy sand and rock) and the biological zone (infralittoral, shallow circalittoral, deep circalittoral, bathyal and abyssal). Another one was used, not visible in the name of the habitat but of paramount importance (see section 5.1). We named it 'local biogeography' and it comprises 3 classes: 'river-influenced', 'sheltered area' and 'other'.

A habitat descriptor class typically comprises two boundaries: a 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 for the biological zone 'shallow circalittoral' the upper boundary is the one that it shares with the infralittoral, while the lower boundary is the one that it shares with the deep circalittoral. However, some habitat descriptor classes have one boundary only, e.g. the 'oxic' oxygen regime only has a lower boundary because this class is at the uppermost level in the oxygen regime habitat descriptor classification.

[1] http://www.emodnet-seabedhabitats.eu/
[2] http://eunis.eea.europa.eu/habitats.jsp

3. Modelling the presence of the habitat descriptor classes

When feasible, the presence of each individual habitat descriptor class is modelled 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 by the EMODnet Seabed Habitats project.

Basically, with this approach a GLM/fuzzy law is used to predict the presence of a habitat descriptor class given the unique environmental variable which explains that presence (e.g. light quantity available at the seabed explains the presence of the infralittoral). When fitted, a GLM/fuzzy law provides the presence probability of the habitat descriptor class via an equation which 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 environmental variable 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 predict the presence of a habitat descriptor class either one GLM/fuzzy law is required or a set of two GLMs/fuzzy laws are required: one if the habitat descriptor class has one boundary only, two if it has two boundaries.

For each GLM/fuzzy law a probability cut-off (also referred to as threshold) value is also determined 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 qualified as present, below it is qualified as absent. The approaches used to work out these threshold values is described in Populus et al (2017).

4. GIS processing

4.1. Producing layers of habitat descriptor on extensive areas

4.1.1. Scaling up GLMs/fuzzy laws

In order to produce layers of the habitat descriptor classes presence probability on extensive areas, the defined GLMs and fuzzy functions are scaled up in a GIS using full-coverage gridded layers of the environmental variables that explain the presence of the various habitat descriptor classes. Then, using the probability threshold values, for each habitat descriptor class the presence probability layers are transformed in a binary presence/absence layer. Finally for each habitat descriptor all habitat descriptor classes presence/absence layer are assembled in one single thematic layer (e.g. for the biological zones all the individual biological zones presence/absence layer are assembled in one single layer).

In the Black Sea, the infralittoral, shallow circalittoral and deep circalittoral were fully modelled in this way. Some GLM/fuzzy laws were fitted, and some continuous layers of relevant environmental variables (depth, temperature) were compiled.

4.1.2. Heads-up delineation

The approach described in section 4.1.1 requires i) GLM/fuzzy law fitting and ii) compilation of spatially/temporally extensive and continuous data of key environmental variables. These two conditions cannot not always be fulfilled: GLM/fuzzy law fitting requires some observation data which are not always available; neither are environmental variables. Therefore the boundaries of some habitat descriptor classes need to be manually delineated.

In the Black Sea for various reasons (see Populus et al, 2017) heads-up delineation had to be used for i) the bathyal and abyssal biological zones, ii) the three classes that compose the habitat descriptor "oxygen regime", and iii) the three classes of the habitat descriptor 'local biogeography'. Seabed substrate types were an input provided by the EMODnet geology in the form of classified polygon data.

4.1.3. Coding convention in the Black Sea for habitat descriptor classes

To each habitat descriptor correspond a raster layer, the classes of which are produced either by GLM/fuzzy law modelling or by heads-up delineation. Below are the codes that are used within the various habitat descriptor layers for their respective class.

Local biogeography

Code	Class
1	Other
2	River-influenced
3	Sheltered area

Biological zone

Code	Class
2	Infralittoral
3	Shallow circalittoral
4	Deep circalittoral
5	Bathyal
6	Abyssal

Seabed substrate

Code	Class
1	Mud
2	Sand
3	Coarse sediment
4	Mixed sediment
5	Sandy mud
6	Muddy sand
7	Rock

Oxygen regime

Code	Class
1	Oxic
2	Suboxic
3	Anoxic

4.2. Deriving the habitat map

4.2.1. The habitat map: a simple sum of the habitat descriptor raster layers

The habitat map is the sum of the four habitat descriptor raster layers. The sum is performed in this way:

[local biogeography] x 10000 + [biological zone] x 100 + [seabed susbstrate] x 10 + [oxygen regime]

This produces a seabed habitat raster layer whose cells have a 5-digit code (e.g. 30221). From left to right the first, third, fourth and fifth digit characterize respectively the local biogeography, the biological zone, the seabed substrate and the oxygen regime. The second digit is not used in the Black Sea but in other basins, thus always has the value 0. As a result, in the habitat raster layer each code is representative of a unique combination of a local biogeography class, a biological zone, a substrate type and an oxygen regime. For example the code 30221 correspond to the combination of the habitat descriptor classes Sheltered area + Infralittoral +Sand + Oxic (3x10000 + 2x100 + 2x10 + 1).

4.2.2. Translating 5-digit codes in habitat classes names

Each 5-digit code needs to be translated in a habitat described in the habitat classification (figure 1). Appendix 1 provides a list of all these habitats together with its matching biological zone, substrate(s), oxygen regime and local biogeography. This translation is performed via a look up table which provides for each code its corresponding habitat class in a column named 'AllDesc'. For example the code 10221, corresponding to the combination of the habitat descriptor classes Other + Infralittoral + Sand + Oxic will match the habitat class 'Infralittoral sand and muddy sand' in the column 'AllDesc' of the look up table. So will be the code 10261, as the habitat 'Infralittoral sand and muddy sand' also encompasses the seabed substrate 'Muddy sand'.

The look up table contains other columns that bring the following information.

Name	Description
ModelCode	Habitat 5-digit code
MODEL_DESC	Class name for each habitat descriptor (e.g. 'other - Shallow Circalittoral - Muddy Sand - Oxic')
Biozone	Biological zone
Substrate	Seabed substrate type
Oxygen	Oxygen regime
AllComb	Habitat code in the classification used. Not very relevant in the Black Sea. If EUNIS was used, would be e.g. 'A5.2'
AllDesc	Habitat description in the classification (e.g. 'Infralittoral sand and muddy sand')
AllcomdD	AllComb + AllDesc - Not relevant in the Black Sea
MSFD_predominant	Habitat description in the MSFD classification
Val_comm	Free text. Typically used for habitats that are predicted while they should not occur (see section 5.1 or 5.3 for examples)

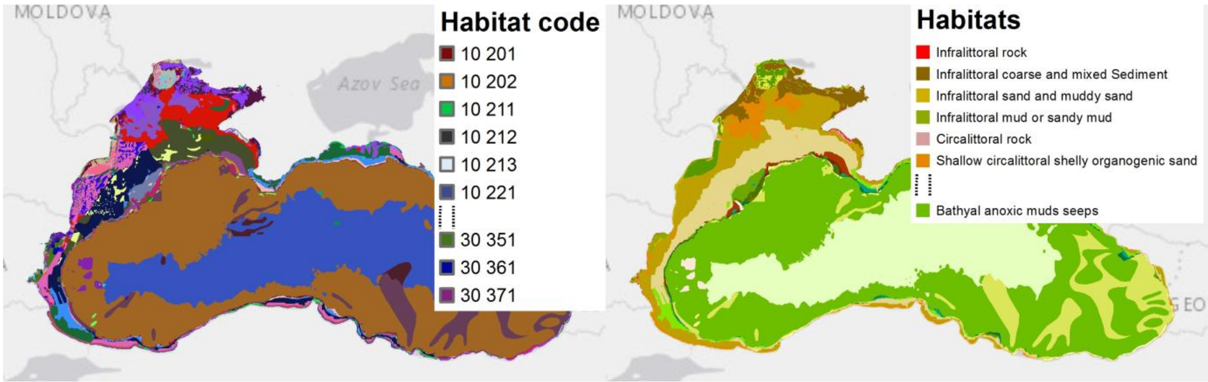

Figure 1: translation of the 5-digit code map (left) in a map with habitat classes (right)

5. Some particularities of the Black Sea basin

5.1. When mud and sandy mud seabed substrates meet the infralittoral...

Mud and sandy mud rarely occur in the infralittoral. However, the general rules used to predict the infralittoral and shallow circalittoral spatial distribution on soft bottoms, which are based on depth, produce here and there some infralittoral mud or sandy mud polygons.

These might be some actual ones: some are acknowledged in Bulgarian sheltered waters, and some others might exist in the Black Sea but have not been observed yet. This might also be a false prediction from our model due to poor input data on either sediment spatial distribution or bathymetry. Therefore, except in Bulgarian sheltered waters, we decided to clearly flag these polygons where muddy sediments intersect the infralittoral so as to inform the user of the high uncertainty of this information.

Extensive areas located at the mouth of the Danube and Dnieper-Bug rivers, hereafter referred to as river plumes, are under the influence of high fine sediment riverine input. In these areas the development of infralittoral soft bottoms communities is driven by the presence of fine superficial sands, muddy sand, coarse and mixed sediment whereas the circalittoral communities develop on sandy mud and mud substrates. Thus inside the river plumes, the boundaries of which were manually delineated, the following rules were defined:

- where sand, muddy sand, coarse or mixed sediment occur, the biological zone is infralittoral
- where mud or sandy mud occur, the biological zone is shallow circalittoral

All these particularities are the reason why we used the habitat descriptor named 'local biogeography'. This habitat descriptor, that occupies the first digit of the 5-digit habitat code, comprises 3 classes. Those (figure 2) are namely 'River-influenced' (i.e. river plume), 'Sheltered areas' (where infralittoral muddy sediments are likely to occur, i.e. in Bulgarian waters) and 'Other' (most common areas, where infralittoral muddy sediments are not expected to occur). In that way, it is the look up table that drives what habitat class name needs to be assigned whether we are in a river plume, a sheltered area or a common area, e.g. :

- 5-digit code 30211 (**Sheltered area** + Infralittoral + Mud) indicates that we are in a sheltered area where infralittoral mud/sandy mud are likely to occur, thus in the look up table the matching habitat is 'Infralittoral mud or sandy mud' with no restriction
- 5-digit code 10211 (**Other** + Infralittoral + Mud), indicates that we are in the an area where infralittoral mud/sandy mud occurrences are not expected. As a result in the look up table the matching habitat is still 'Infralittoral mud or sandy mud' but a warning message is

present in the *Val_comm* column ('Uncertain due to poor modeling of biological zone and/or substrate')

- 5-digit code 20251 (**River-influenced** + Infralittoral + Sandy mud) indicates that we are inside a plume area where mud/sandy mud occurs, thus in the look up table the matching habitat is in all cases 'Circalittoral terrigenous muds'

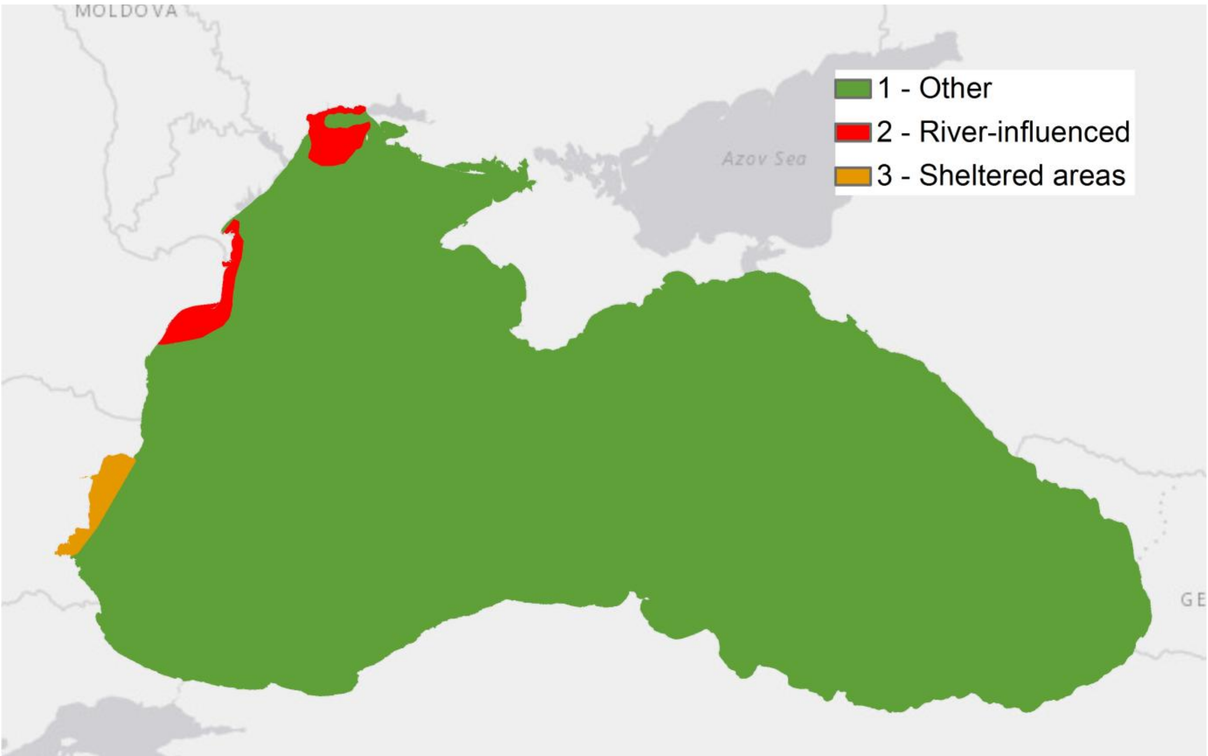

Figure 2: the 'local biogeography' habitat descriptor input layer. 'River-influenced' areas are plume areas. 'Sheltered areas' are where mud/sandy mud are known to occur in the infralittoral. 'Other' areas are common areas where mud/sandy mud are not expected to occur

5.2. Infralittoral/circalittoral modelling: soft bottoms vs. hard bottoms

The fuzzy law used to model the infralittoral and circalittoral for rocky bottoms is different from that used to model the infralittoral and shallow circalittoral on soft bottoms (Populus et al, 2017). Therefore for both of these biological zones the fuzzy equation slope and intercept input raster layers have different values whether the seabed substrate is hard or soft. This is illustrated in figures 5 and 6, section 6.2.

5.3. Some habitat descriptor classes combinations occur but are not acknowledged

In addition to the combination muddy sediment + infralittoral discussed in section 5.1, the combination of some habitat descriptors classes is predicted by our model while their occurrence is not reported in the literature (see appendix 1, habitat names in brown colors). For these habitats a warning message is included in the look up table's *Val_comm* column (' Unexpected. Requires further investigation).

6. Workflow Models for the Black Sea

In order to automate the production of the broad scale maps, toolboxes with GIS Worflow models were developed under ArcGIS 10.2. was developed. The toolbox dedicated to the Black Sea basin comprises three workflow models, namely *Compute probability*, *Biozone processing* and *Habitats processing*.

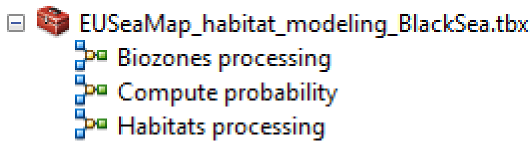

Figure 3: the three workflow models that compose the Black Sea toolbox

The three workflow models are specifically described in the following subsections.

6.1. Compute probability

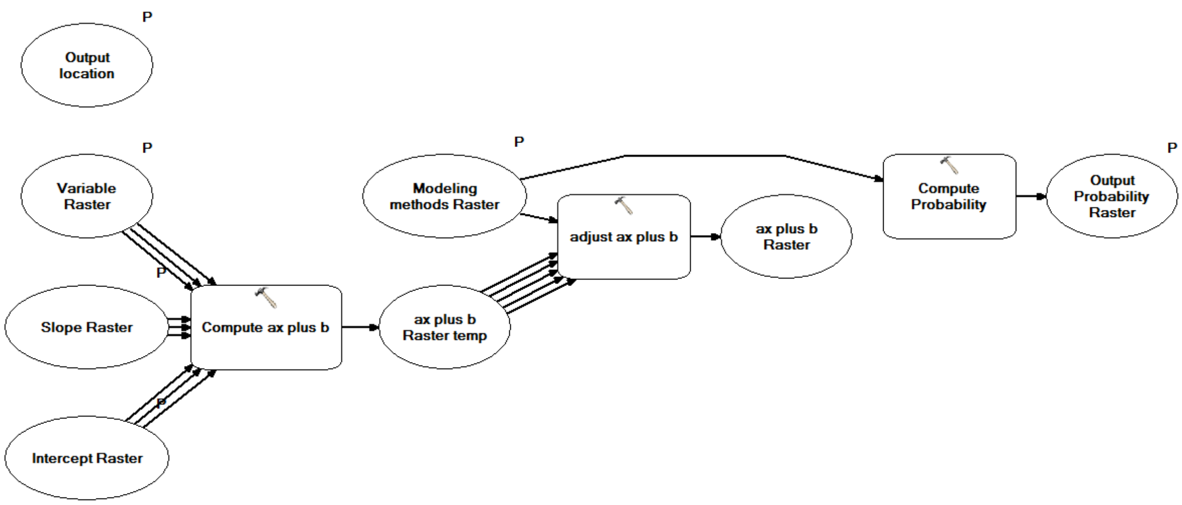

Figure 4: the '*compute probability*' workflow

This generic workflow produces a raster for the presence probability of a habitat descriptor class (e.g. shallow circalittoral) given the value of the environmental factor that explains that presence. The probability can be computed via either a GLM (Generalized Linear Model) or a fuzzy function, whose equation is in the form (see section 3)

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

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

where X is the environmental parameter 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.

6.1.1. Inputs

Name	Description	Type
Variable raster	The environmental factor values	Raster
Slope raster	Equation slope values	Raster
Intercept raster	Equation intercept values	Raster
Modeling method raster	Method type 1=fuzzy function 2=GLM	Raster
Output location	Path to the output geodatabase	Text

6.1.2. Outputs

Name	Description	Type
Output probability raster	Presence probability values	Raster

6.1.3. Processes

Name	Compute ax plus b
Inputs name	Variable raster Slope raster Intercept raster
Outputs name	*ax plus b Raster tmp*, which is a raster with ax+b values (a=slope, b=intercept)
Formula	`"%Slope Raster%"*"%Variable Raster%"+"%Intercept Raster%"`
Description	a=slope, b=intercept

Name	adjust ax plus b
Inputs name	*ax plus b Raster tmp*, the output raster of the process *Compute ax plus b* *Modeling method raster*, an input of the workflow
Outputs name	*ax plus b Raster tmp*, which is a raster with ax+b values (a=slope, b=intercept) adjusted according to below considerations
Formula	`Con("%Modeling methods Raster%"==1,Con("%ax plus b Raster temp%" > 1,1, Con("%ax plus b Raster temp%" < 0,0,"%ax plus b Raster temp%")),Con("%ax plus b Raster temp%" > 7,7,"%ax plus b Raster temp%"))`
Description	If the equation is a fuzzy function, the ax+b calculated in the previous process (Compute ax plus b) might produce out of the range [0,1]. This has to be corrected because the result of fuzzy functions are probability values, thus values in the range [0,1] • If input > 1 then output = 1 • If input < 0 then output = 0 • otherwise output=input If the equation is a GLM, the output can't be > 7 because in the subsequent process (*Compute Probability*) this output will feed an exponential function, and a technical

	limitation of ArcGIS 10.2 is that the exponential function does not work properly with values > 7 • If input > 7 then output = 7 • otherwise output=input

Name	Compute probability
Inputs name	ax plus b Raster, the output raster of the process *adjust ax plus b* that contains ax+b values Modeling method raster, an input of the workflow
Outputs name	A raster with ax+b values (a=slope, b=intercept)
Formula	`Con("%Modeling methods Raster%"==1,"%ax plus b Raster%",Exp("%ax plus b Raster%")/(1+Exp("%ax plus b Raster%")))`
Description	If the equation is a fuzzy function, output = input If the equation is a GLM, output = $e^{input} / (1+e^{input})$

6.2. Biozone processing

The workflow aims to produce both presence probability layers and spatial distribution layers for the infralittoral, the shallow circalittoral and the deep circalittoral. Bathyal and abyssal spatial distribution are provided as a single "ready for use" raster layer, because this is the way they are produced, thus can't be treated in the same manner.

The workflow 1) computes, according to GLM/fuzzy laws, the presence probability layers required for each of the biological zones infralittoral, shallow circalittoral and deep circalittoral, 2) creates a spatial distribution raster layer for each of these biological zones (i.e.; each cell where the biological zone is present has a value, each cell where it is absent has null value), 3) produces spatial distribution layers for the bathyal and the abyssal and 4) finally merges each individual presence raster layer into a single one that comprises all biological zone.

6.2.1. Inputs

Name	**Description**	**Type**
Infra lower bound proxy raster	Environmental factor values that are used as a predictor of the infralittoral presence in the deeper part. In the Black sea they are depth values	Raster
Equation Slope Raster for Infra lower bound	Slope value of the equations that are used to compute the infralittoral presence probability in the deeper part. In the Black Sea there are 2 equations, one for sediment substrate and another one for rock, thus there are 2 values (see figure 5)	Raster
Equation Intercept Raster for Infra lower bound	Intercept value of the equations that are used to compute the infralittoral presence probability in the deeper part. In the Black Sea there are 2 equations, one for sediment	Raster

	substrate and another one for rock, thus there are 2 values (see figure 5)	
Equation Type Raster for Infra lower bound	Type of the equations that are used to compute the infralittoral presence probability in the deeper part. In the Black Sea the two equations are fuzzy ones, thus the value is 1 for all cells	Raster
Cut-off values Raster for Infra lower bound	The probability threshold values that are used to set the infralittoral lower boundary. For the Black Sea it is 0.5 everywhere	Raster
Shallow Circa upper bound proxy raster	Environmental factor values that are used as a predictor of the shallow circalittoral presence in the upper part. In the Black sea they are depth values	Raster
Equation Slope Raster for Shallow Circa upper bound	Slope value of the equations that are used to compute the shallow circalittoral presence probability in the upper part. In the Black Sea there are 2 equations, one for sediment substrate and another one for rock, thus there are 2 values (see figure 6)	Raster
Equation Intercept Raster for Shallow Circa upper bound	Intercept value of the equations that are used to compute the shallow circalittoral presence probability in the upper part. In the Black Sea there are 2 equations, one for sediment substrate and another one for rock, thus there are 2 values (see figure 6)	Raster
Equation Type Raster for Shallow Circa upper bound	Type of the equations that are used to compute the shallow circalittoral presence probability in the upper part. In the Black Sea the two equations are fuzzy ones, thus the value is 1 for all cells	Raster
Cut-off values Raster for Shallow Circa upper bound	The probability threshold values that are used to set the shallow circalittoral upper boundary. For the Black Sea it is 0.5 everywhere	Raster
Shallow Circa lower bound proxy raster	Environmental factor values that are used as a predictor of the shallow circalittoral presence in the deeper part. In the Black sea they are temperature values	Raster
Equation Slope Raster for Shallow Circa lower bound	Slope value of the equations that are used to compute the shallow circalittoral presence probability in the deeper part. In the Black Sea there is only one equation thus it is the same value everywhere (3.74)	Raster

Equation Intercept Raster for Shallow Circa lower bound	Intercept value of the equations that are used to compute the shallow circalittoral presence probability in the deeper part. In the Black Sea there is only one equation thus it is the same value everywhere (-34.59)	Raster
Equation Type Raster for Shallow Circa lower bound	Values for the type of the equations that are used to compute the shallow circalittoral presence probability in the deeper part. In the Black Sea there is only one equation and it is a GLM, thus the value is 2 for all cells	Raster
Cut-off values Raster for Shallow Circa lower bound	The probability threshold values that are used to set the shallow circalittoral lower boundary. For the Black Sea it is 0.27 everywhere	Raster
Deep Circa upper bound proxy raster	Environmental factor values that are used as a predictor of the deep circalittoral presence in the upper part. In the Black sea they are temperature values	Raster
Equation Slope Raster for Deep Circa upper bound	Slope value of the equations that are used to compute the deep circalittoral presence probability in the upper part. In the Black Sea there is only one equation thus it is the same value everywhere (-3.74)	Raster
Equation Intercept Raster for Deep Circa upper bound	Intercept value of the equations that are used to compute the deep circalittoral presence probability in the upper part. In the Black Sea there is only one equation thus it is the same value everywhere (34.59)	Raster
Equation Type Raster for Deep Circa upper bound	Type of the equations that are used to compute the deep circalittoral presence probability in the upper part. In the Black Sea there is only one equation and it is a GLM, thus the value is 2 for all cells	Raster
Cut-off values Raster for Deep Circa upper bound	The probability threshold values that are used to set the deep circalittoral upper boundary. For the Black Sea it is 0.73 everywhere	Raster
bathyal & abyssal zones	Spatial distribution of shelf, bathyal and abyssal. The following codes are used • 0 = shelf • 1 = bathyal • 2 = abyssal	Raster
Reference raster	A raster on which are snapped the outcomes of all processes	Raster

Output location	Path to the output geodatabase	Text

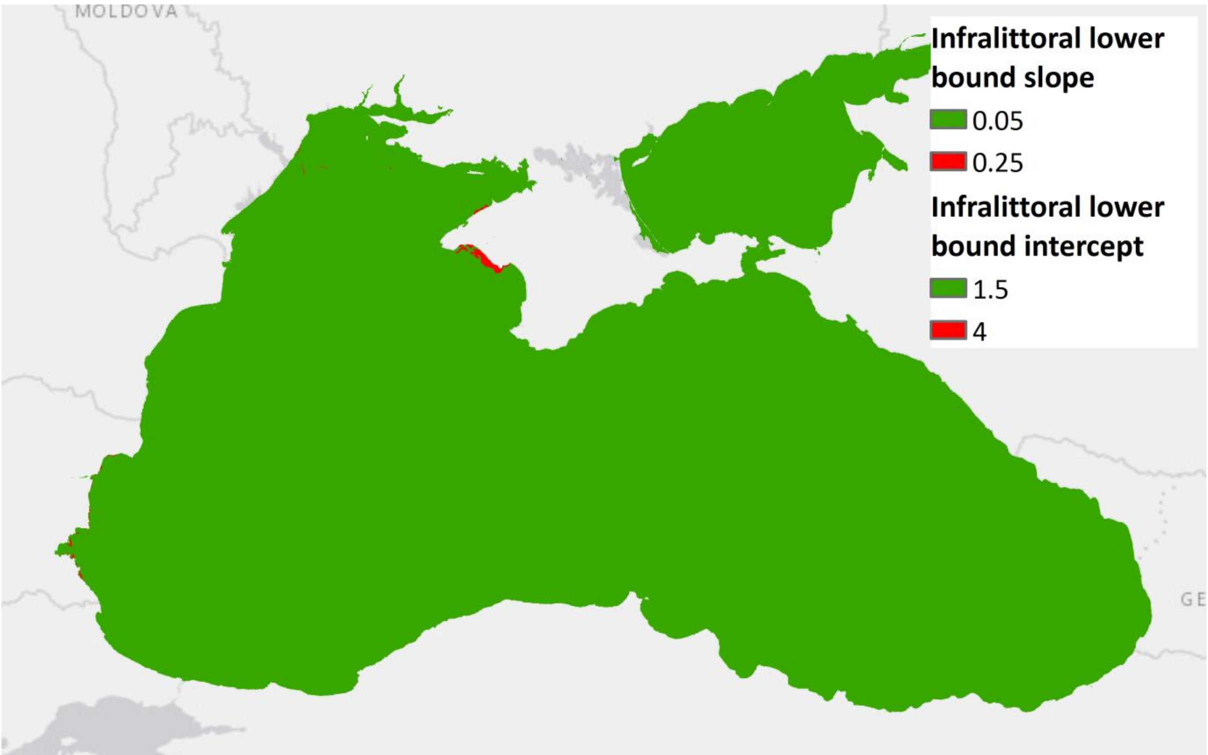

Figure 5: slope and intercept values that are used for modeling the infralittoral deeper part where the seabed substrate is sediment (green) and where it is rock (red)

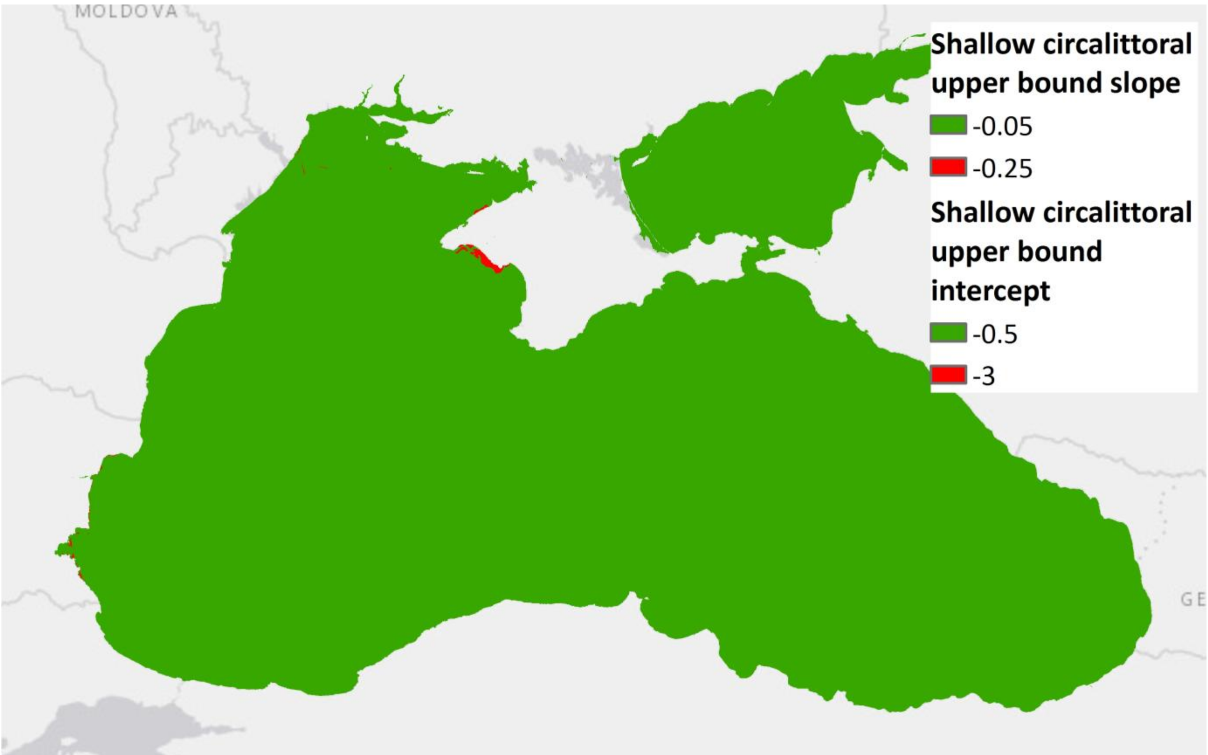

Figure 6: slope and intercept values that are used for modeling the shallow circalittoral upper part where the seabed substrate is sediment (green) and where it is rock (red)

6.2.2. Outputs

Name	Description	Type
Output Presence Probability Raster for Infra lower bound	Presence probability of the infralittoral in the deeper part	Raster
Output Presence Probability Raster for Shallow Circa upper bound	Presence probability of the shallow circalittoral in the upper part	Raster
Output Presence Probability Raster for Shallow Circa lower bound	Presence probability of the shallow circalittoral in the deeper part	Raster
Output Presence Probability Raster for Deep Circa upper bound	Presence probability of the deep circalittoral in the upper part	Raster
Output Infra presence Raster	Infralittoral spatial distribution. Where present the cell values is 2, where absent it is null	Raster
Output Shallow Circa presence Raster	Shallow circalittoral spatial distribution. Where present the cell values is 3, where absent it is null	Raster
Output Deep Circa presence Raster	Deep circalittoral spatial distribution. Where present the cell values is 4, where absent it is null	Raster
Output Biozones	All five biozones in a single layer. Values are the following • 2 = infralittoral • 3 = shallow circalittoral • 4 = deep circalittoral • 5 = bathyal • 6 = abyssal	Raster

6.2.3. Processes

1) The worflow model starts by running a series of processes, namely a, b, c and d on figure 8, that produce for each biozone and for each of its boundaries a presence probability raster according to the rules that condition the biozone's presence/absence at the boundary. By "rules" is meant an environmental factor and a GLM/fuzzy equation (both governed by a slope and an intercept).

Name	Compute Presence Probability for Infra lower bound - (a) on figure 8
Inputs name	Infra lower bound proxy raster Equation Slope Raster for Infra lower bound Equation Intercept Raster for Infra lower bound Equation Type Raster for Infra lower bound
Outputs	Output Presence Probability Raster for Infra lower bound

name	
Formula	-
Description	All inputs (proxy, equation slope, equation intercept and equation type) feed the workflow C*ompute probability* that in turn produces the presence probability raster according to the rules that govern the infralittoral presence/absence in the deeper part. A more specific description of the workflow C*ompute probability* is provided in the section dedicated to the workflow

Name	Compute Presence Probability for Shallow Circa upper bound - (b) on figure 8
Inputs name	Shallow Circa upper bound proxy raster Equation Slope Raster for Shallow Circa upper bound Equation Intercept Raster for Shallow Circa upper bound Equation Type Raster for Shallow Circa upper bound
Outputs name	Output Presence Probability Raster for Shallow Circa upper bound
Formula	-
Description	All inputs (proxy, equation slope, equation intercept and equation type) feed the workflow C*ompute probability* that in turn produces the presence probability raster according to the rules that govern the shallow circalittoral presence/absence in the upper part. A more specific description of the workflow C*ompute probability* is provided in the section dedicated to the workflow

Name	Compute Presence Probability for Shallow Circa lower bound - (c) on figure 8
Inputs name	Shallow Circa lower bound proxy raster Equation Slope Raster for Shallow Circa lower bound Equation Intercept Raster for Shallow Circa lower bound Equation Type Raster for Shallow Circa lower bound
Outputs name	Output Presence Probability Raster for Shallow Circa lower bound
Formula	-
Description	All inputs (proxy, equation slope, equation intercept and equation type) feed the workflow C*ompute probability* that in turn produces the presence probability raster according to the rules that govern the shallow circalittoral presence/absence in the deeper part. A more specific description of the workflow C*ompute probability* is provided in the section dedicated to the workflow

Name	Compute Presence Probability for Deep Circa upper bound - (d) on figure 8
Inputs name	Deep Circa upper bound proxy raster Equation Slope Raster for Deep Circa upper bound Equation Intercept Raster for Deep Circa upper bound Equation Type Raster for Deep Circa upper bound
Outputs name	Output Presence Probability Raster for Deep Circa upper bound
Formula	-
Description	All inputs (proxy, equation slope, equation intercept and equation type) feed the

	workflow C*ompute probability* that in turn produces the presence probability raster according to the rules that govern the deep circalittoral presence/absence in the upper part. A more specific description of the workflow C*ompute probability* is provided in the section dedicated to the workflow

2) At that point a presence probability raster has been produced for each biozone and for each of its boundary. Then the workflow model runs three processes, namely e, f, and g on figure 8, that produce for each biozone a binary presence raster, i.e. a raster whose cells have a value where the biozone is present, and have no value where the biozone is absent . Cell values are coded according to the coding convention that is given in table 'biological zone', section 4.1.3

Name	Compute Infra binary presence - (e) on figure 8
Inputs name	Output Presence Probability Raster for Infra lower bound Cut-off values Raster for Infra lower bound
Outputs name	Output Infra presence Raster
Formula	`SetNull("%Output Presence Probability Raster for Infra lower bound%" < "%Cut-off values Raster for Infra lower bound%",2)`
Description	The inputs (presence probability and cut-off) feed the process that in turn transform the presence probability raster in a binary infralittoral spatial distribution raster (i.e. a raster where value is either 2 if present, or null if absent) The cells that have a probability value < the cut-off value are set to null. The cells that have a probability value > the cut-off value are given the code 2 which is the code for the infralittoral The infralittoral only has one boundary, the lower one, because there is no biological zone above the infralittoral.

Name	Compute Shallow Circa binary presence - (f) on figure 8
Inputs name	Output Presence Probability Raster for Shallow Circa upper bound Cut-off values Raster for Shallow Circa upper bound Output Presence Probability Raster for Shallow Circa lower bound Cut-off values Raster for Shallow Circa lower bound
Outputs name	Output Shallow Circa presence Raster
Formula	`SetNull(("%Output Presence Probability Raster for Shallow Circa upper bound%" < "%Cut-off values Raster for Shallow Circa upper bound%") | ("%Output Presence Probability Raster for Shallow Circa lower bound%" < "%Cut-off values Raster for Shallow Circa lower bound%"),3)`
Description	The inputs (presence probability and cut-off) feed the process that in turn transform the presence probability raster in a binary shallow circalittoral spatial distribution raster (i.e. a raster where value is either 3 if present, or null if absent) For both boundaries, the cells that have a probability value < the cut-off value are set to null. The cells that have a probability value > the cut-off value are given the code 3 which is the code for the shallow circalittoral

Name	Compute Deep Circa binary presence - (g) on figure 8
Inputs name	Output Presence Probability Raster for Deep Circa upper bound Cut-off values Raster for Deep Circa upper bound

Outputs name	Output Deep Circa presence Raster
Formula	`SetNull("%Output Presence Probability Raster for Deep Circa upper bound%" < "%Cut-off values Raster for Deep Circa upper bound%",4)`
Description	The inputs (presence probability and cut-off) feed the process that in turn transform the presence probability raster in a binary deep circalittoral spatial distribution raster (i.e. a raster where value is either 4 if present, or null if absent) The cells that have a probability value < the cut-off value are set to null. The cells that have a probability value > the cut-off value are given the code 4 which is the code for the deep circalittoral The deep circalittoral only has one boundary, the upper one, because below the deep circalittoral is the bathyal and the boundary between deep circalittoral and bathyal is defined as a slope change which is manually drawn. Bathyal spatial distribution is provided "ready-for-use" to the model workflow, and *de facto* so is the boundary between the deep circalittoral and the bathyal.

3) At that point a binary spatial distribution raster has been produced for the infralittoral, the shallow circalittoral and the deep circalittoral. The bathyal and the abyssal are treated in a different way because their boundaries are produced separately by manual delineation. Their spatial distribution is provided to the workflow in the form of a single raster layer, the cells of which have the value 1 where it is bathyal and the value 2 where it is abyssal. Two processes, namely h and i on figure 8, are in charge of making 2 layers from that single one, i.e. one for the bathyal spatial distribution and another one for the abyssal spatial distribution.

Name	bathyal calculation - (h) on figure 8
Inputs name	bathyal & abyssal zones
Outputs name	bathyal
Formula	`SetNull("%bathyal & abyssal zones%" != 1,5)`
Description	The cells that in the input have another value than that of the bathyal (i.e. 1) are set to null. The cells that in the input have the value 1 are set to 5 which is the code for the bathyal

Name	abyssal calculation - (i) on figure 8
Inputs name	bathyal & abyssal zones
Outputs name	abyssal
Formula	`SetNull("%bathyal & abyssal zones%" != 2,6)`
Description	The cells that in the input have another value than that of the abyssal (i.e. 2) are set to null. The cells that in the input have the value 2 are set to 6 which is the code for the abyssal

A last process (j on figure 8) assembles each individual spatial distribution layer in order to produce a single biological zone raster layer (figure 7).

Name	Assemble - (j) on figure 8
Inputs name	Output Infra presence Raster Output Shallow Circa presence Raster Output Deep Circa presence Raster bathyal abyssal
Outputs name	Output Biozones
Formula	-
Description	The *Mosaic To New Raster* ArcGIS native tool is used.

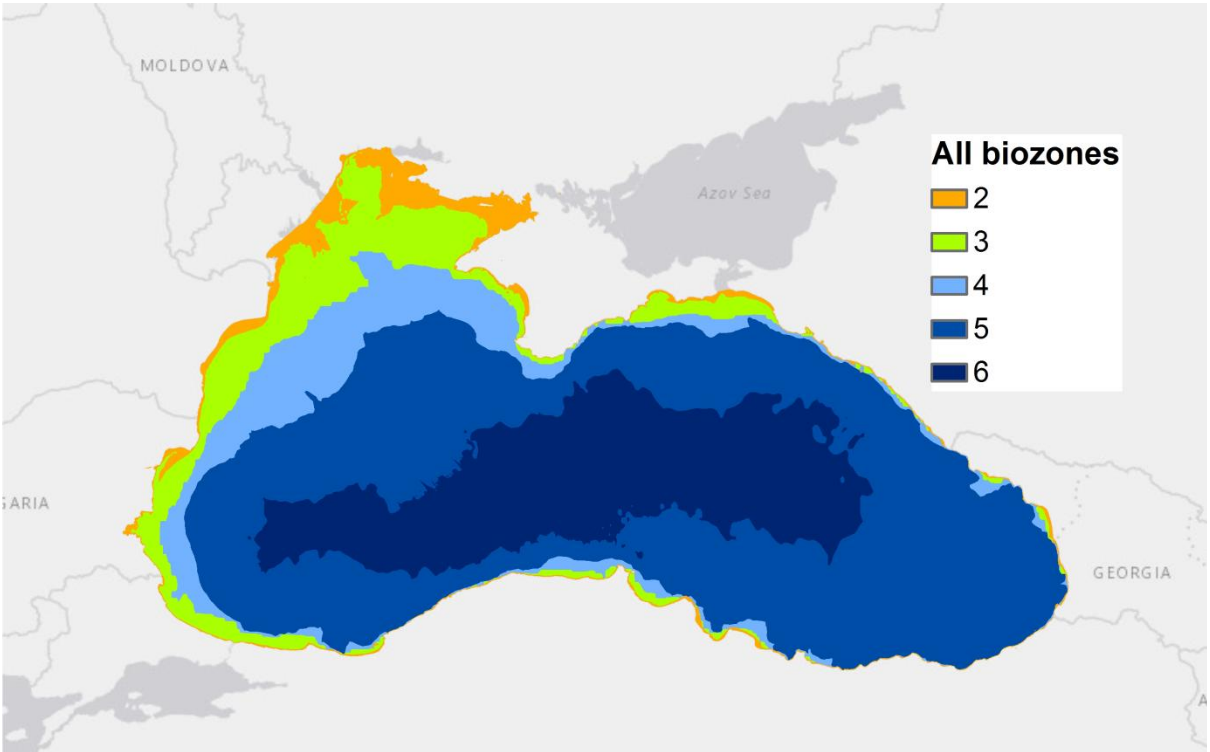

Figure 7: final biological zone layer

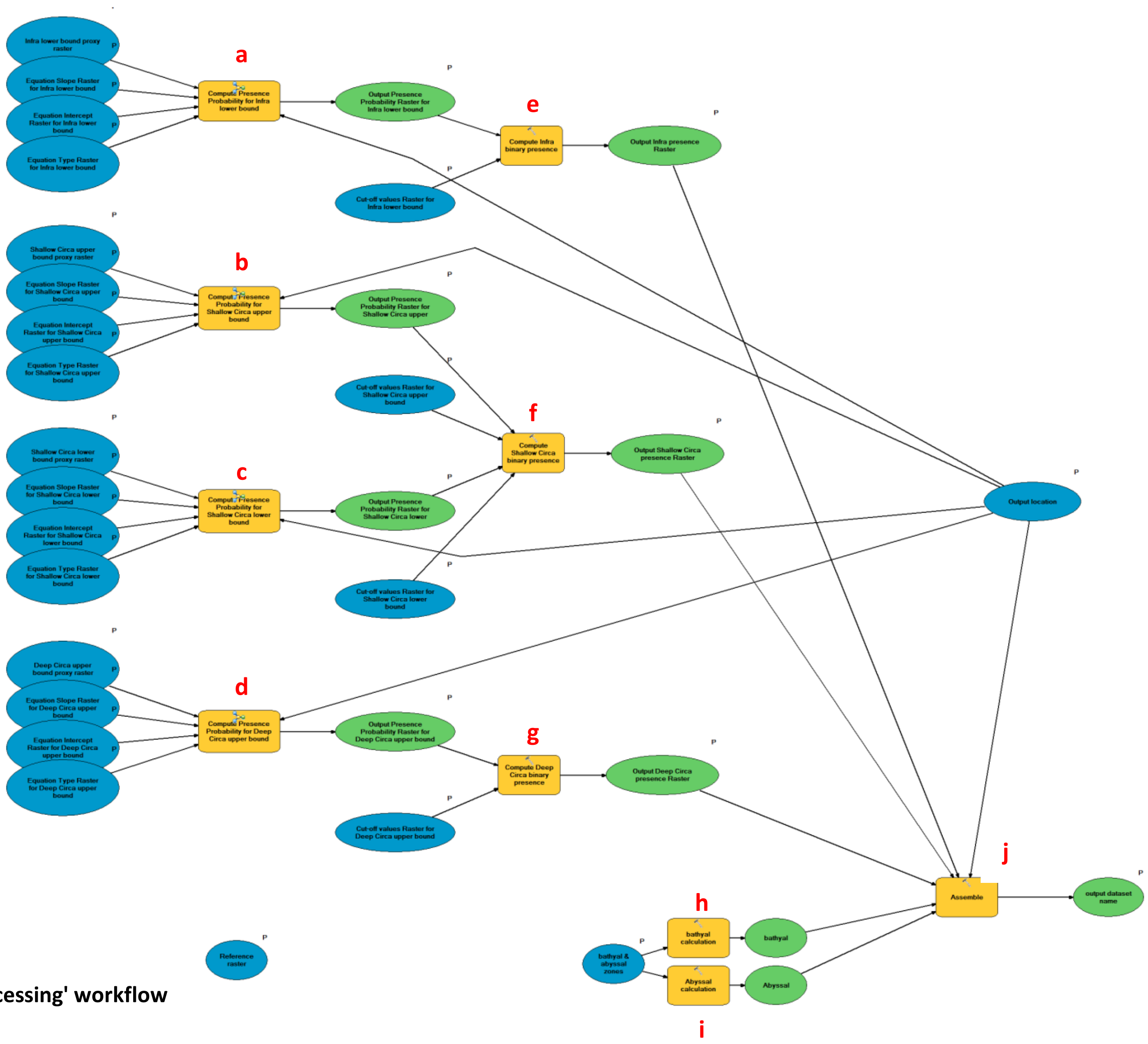

Figure 8: the 'biozone processing' workflow

6.3. Habitats processing

The workflow (figure 10) aims to produce the broad scale habitat map by summing the four habitat descriptor layers that are the local biogeography, the biological zones, the seabed substrate types and the oxygen regime (figure 9). The 5-digit code raster layer that results from that sum is subsequently converted in a polygon layer, to which is joined the look up table (see section 4.2.2) which contains the seabed habitats classes that match each 5-digit code.

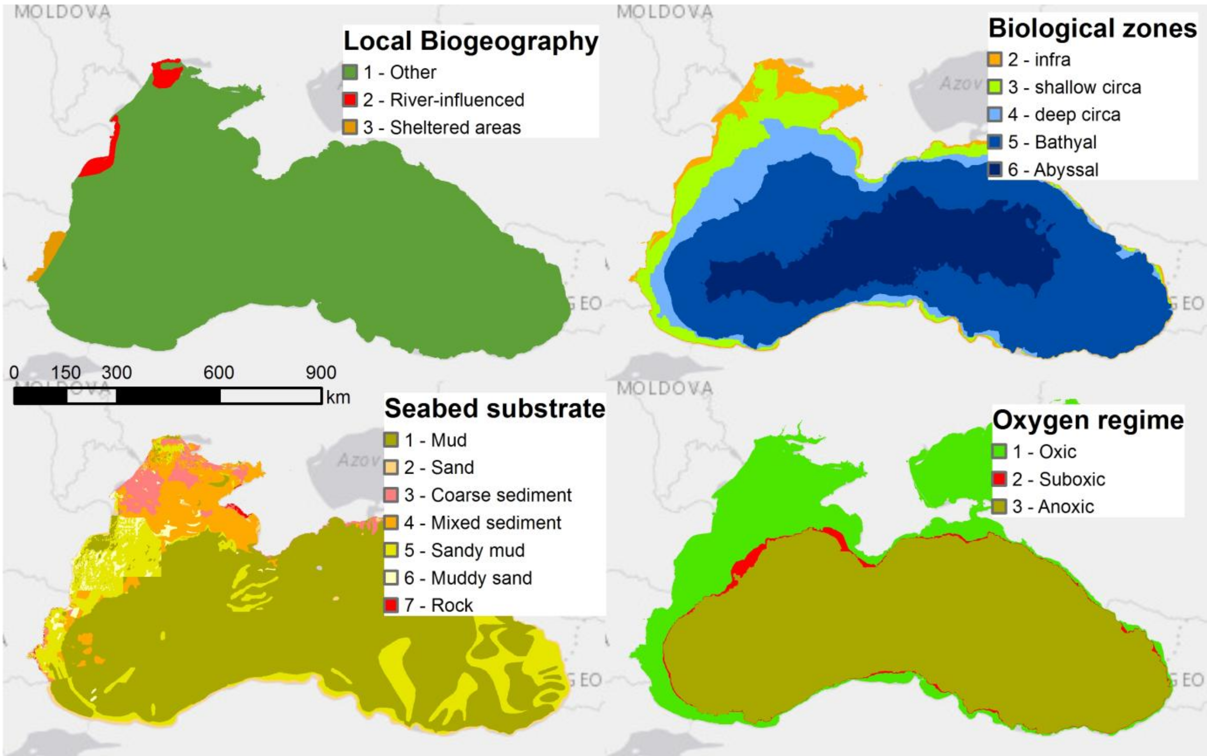

Figure 9: the 4 habitat descriptor raster layers that are summed to produce the habitat map

6.3.1. Inputs

Name	Description	Type
Input seabed substrate	Seabed substrate types	Raster
Input oxygen conditions	Oxygen regimes	Raster
Input local biogeography	Plume, sheltered area or other	Raster
Input biological zones	Biological zones	Raster
Habitats LUT	Look up table	Table
Output location	Path to the output geodatabase	Text

6.3.2. Outputs

Name	Description	Type
Output habitat Raster	Output 5-digit code habitat raster layer	Raster
Output habitat polygons	Habitat layer in vector form	Polygons

6.3.3. Processes

Name	Compute Habitats
Inputs name	Input seabed substrate Input oxygen conditions Input local biogeography Input biological zones
Outputs name	Output habitat Raster
Formula	"%Input local biogeography%" * 10000 + "%Input biological zones%" * 100 + "%Input seabed substrate%" * 10 + "%Input oxygen conditions%"
Description	Performs the sum of the 4 habitat descriptor layers in a way the local biogeography, the biozones, the seabed substrate and the oxygen regime are coded in the 5-digit code output habitat raster layer respectively in the 10000s, 100s, 10s and 1s

Name	Join Fields To Raster Table
Inputs name	Output habitat Raster
Outputs name	Output habitat Raster with all Fields
Formula	-
Description	Joins the look up table to the the 5-digit code habitat raster layer

Name	Raster to Polygon
Inputs name	Output habitat Raster with all Fields
Outputs name	Output habitat polygons
Formula	-
Description	Converts the habitat raster layer in polygon layer

Name	Join Fields To PolygonTable
Inputs name	Output habitat polygons
Outputs name	Output habitat polygons with all fields
Formula	-
Description	Joins the look up table to the habitat polygon layer

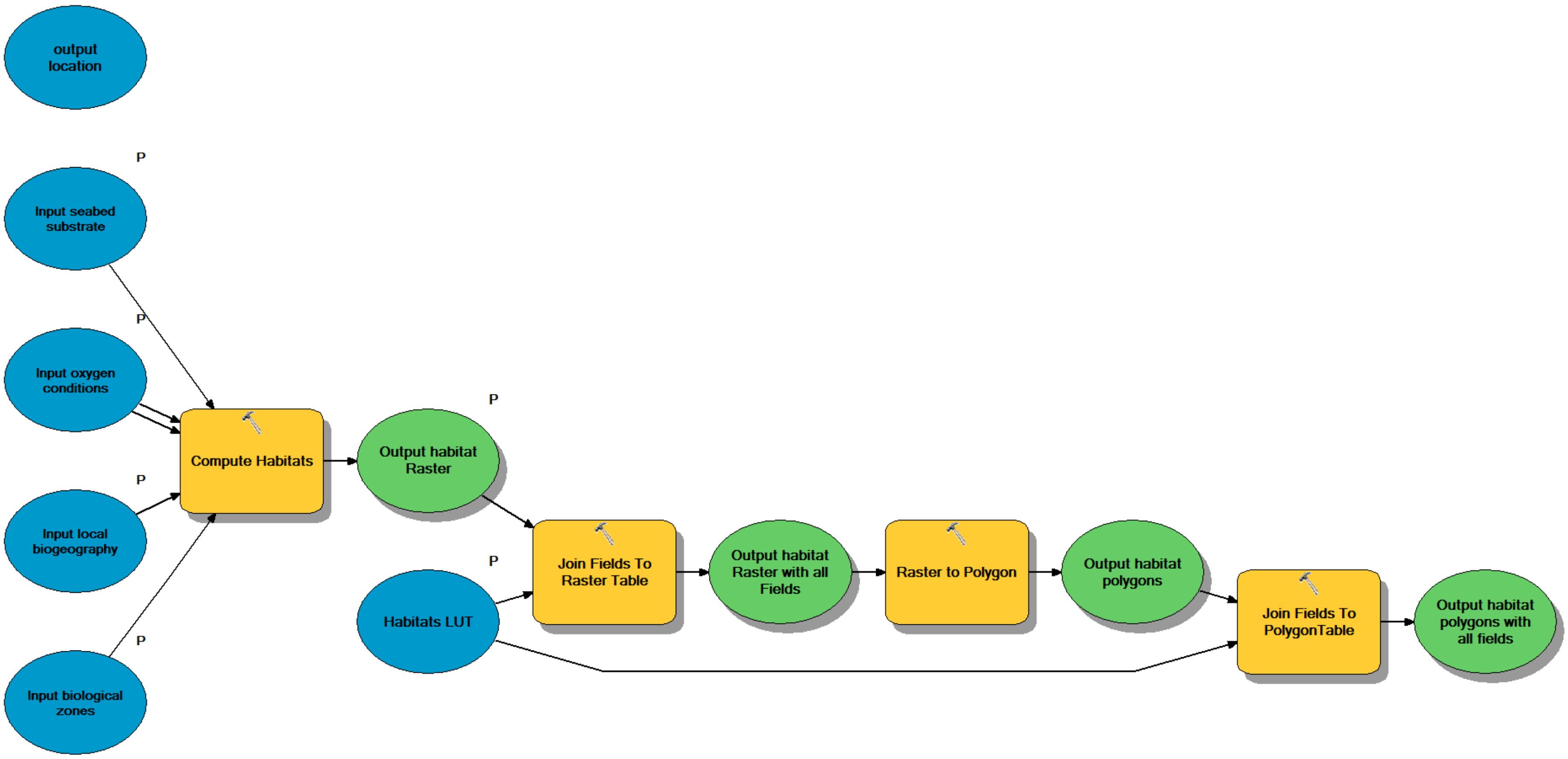

Figure 10: the 'Habitat processing' workflow

7. Supplemental material

All supplemental material is available in zip file (85 MB) at the following address. It contains the toolbox, input/output data and some ArcGIS documents. ArcGIS 10.2 version is required.

https://cloud.ifremer.fr/index.php/s/orQJsWyEDJ3L47M

8. 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

Appendix 1: Habitat names that match each individual combination biozone + substrate + oxygen regime + local biogeography. Red text is for those combinations that are considered as uncertain (i.e. the habitat is not acknowledged but is predicted in some places), and brown text is for those that are unexpected (i.e. the combination requires further investigation).

Classification habitat name	Biological Zone	Substrate	Oxygen regim	Local biogeography
Infralittoral sand (Plume)	INFRA	SAND	OXIC	RIVER-INFLUENCED
Infralittoral muddy sand (Plume)	INFRA	MUDDY SAND	OXIC	RIVER-INFLUENCED
Shallow circalittoral coarse and mixed sediment (Plume)	CIRCA	MIXED/COARSE	OXIC	RIVER-INFLUENCED
Circalittoral terrigenous muds (Plume)	CIRCA	MUD/ SANDY MUD	OXIC	RIVER-INFLUENCED
Infralittoral rocks with photophilic algae	INFRA	ROCK	OXIC	OTHER
Infralittoral mud and sandy mud	INFRA	MUD; SANDY MUD;	OXIC	OTHER
Infralittoral mud and sandy mud	INFRA	MUD; SANDY MUD;	OXIC	SHELTERED AREAS
Infralittoral sand and muddy sand	INFRA	SAND; MUDDY SAND	OXIC	OTHER
Infralittoral Coarse and Mixed Sediment	INFRA	COARSE; MIXED	OXIC	OTHER
Circalittoral rock	CIRCA	ROCK	OXIC	OTHER
Shallow circalittoral mud and organogenic sandy mud/muddy sand	CIRCA	MUD; SANDY MUD; MUDDY SAND; MIXED	OXIC	OTHER
Shallow circalittoral shelly organogenic sand (clean shelly debris without mud)	CIRCA	SAND; COARSE (shelly with no mud)	OXIC	OTHER
Deep circalittoral mixed sediments	DEEP CIRCA	COARSE; MIXED; MUDDY SAND, SANDY MUD	OXIC	OTHER
Deep circalittoral sand (unexpected)	DEEP CIRCA	SAND	OXIC	OTHER
Deep circalittoral mud	DEEP CIRCA	MUD	OXIC	OTHER
Deep circalittoral suboxic calcareous muds	DEEP CIRCA	MUD	SUBOXIC	OTHER

Classification habitat name	Biological Zone	Substrate	Oxygen regim	Local biogeography
Deep circalittoral suboxic muddy sand (unexpected)	DEEP CIRCA	MUDDY SAND	SUBOXIC	OTHER
Deep circalittoral suboxic sand (unexpected)	DEEP CIRCA	SAND	SUBOXIC	OTHER
Deep circalittoral suboxic sandy mud (unexpected)	DEEP CIRCA	SANDY MUD	SUBOXIC	OTHER
Deep circalittoral anoxic muds	DEEP CIRCA	MUD	ANOXIC	OTHER
Deep circalittoral anoxic muddy sand (unexpected)	DEEP CIRCA	MUDDY SAND	ANOXIC	OTHER
Deep circalittoral anoxic sand (unexpected)	DEEP CIRCA	SAND	ANOXIC	OTHER
Deep circalittoral anoxic sandy mud (unexpected)	DEEP CIRCA	SANDY MUD	ANOXIC	OTHER
Bathyal anoxic muds	BATHYAL	MUD	ANOXIC	OTHER
Bathyal coarse sediment (unexpected)	BATHYAL	COARSE	ANOXIC	OTHER
Bathyal mixed sediment (unexpected)	BATHYAL	MIXED	ANOXIC	OTHER
Bathyal muddy sand (unexpected)	BATHYAL	MUDDY SAND	ANOXIC	OTHER
Bathyal sand (unexpected)	BATHYAL	SAND	ANOXIC	OTHER
Bathyal sandy mud (unexpected)	BATHYAL	SANDY MUD	ANOXIC	OTHER
Abyssal seabed	ABYSSAL	ANY	ANOXIC	OTHER