EUSeaMap

Technical Report No. X

Confidence

Assessment of the EUSeaMap global modelled map

1. Assessing base data layers 2
1.1 Pixel-based assessment of input layers 2
1.2 Global assessment as a weighted sum of regional scores 3
1.3 Scoring maps in the Mediterranean basin 7
2. Assessment of modelled maps versus recent habitat maps 9
3. Conclusion - Overall assessment 11
4. References 12

Introduction

There are basically two ways of considering the assessment of the EUSeaMap modelled map. As the resultant map is no more than a combination of source layers, one straightforward way is to assess the confidence of the broad-scale physical data layers that form the basis of the Eunis model and compute a weighted sum of the scores. This “internal assessment” can go from local to regional depending on the method used. Methods that could provide an assessment on each pixel would likely be deemed the best ones but their output would be a map at the same resolution as the modelled map itself, hence potentially difficult for users to read if they varied too much locally. Other methods of assessing at regional level (per blocks of data, as e.g. is the case in nautical chart showing origins of soundings) would be of easier use albeit less precisely informative.

A second aspect of internal confidence is that associated with thresholds. Even in the case where we would have highly reliable layers, the relevance of a given threshold (an amount of light or energy at the seabed) is questionable as it is at best founded on approximate field data. Therefore it is advisable to associate a fuzzy area around a sharp limit between two categories.

The other possible way (or “external assessment”) is statistical. Based on local checks carried out against external data, it would yield what is referred to in land use mapping as a “contingency matrix”, simply a global score for the whole map. Depending of the amount of ground truth data available, it could also be regionalised, i.e. computed per basin (or even sub-basin). Along with the internal assessment it could then bring an additional insight into the quality of the modelled map.

1. Assessing base data layers

This kind of assessment is either local if each pixel of each layer is assessed for the modelled area or regional when assessing base layers per regional data blocks and intersecting them.

1.1 Pixel-based assessment of input layers

The former method is the most thorough and precise, yet complex method. It is described in detail in the ABPMer- coordinated study lead for UK Defra “Assessing the confidence of broad-scale classification maps”. The uncertainty of each data layer (biological zone, substrate and energy) is assessed using a combination of three types of uncertainty: measurement error, processing error and natural variability. A probability of occurrence is computed for each parameter at each location. These probabilities are then multiplied and the result subtracted to one to yield a total probability.

The EuSeaMap could have decided to go for such an assessment, however two difficulties arose:

- the fact that data layers were generated by a number of different teams (actually at least the seven teams associated in the project, not to mention sub-contractors), hence a need for a comprehensive investigation on the methods employed and substantial uncertainty in obtaining reliable results,

- the associated costs were likely to be high, as it was necessary to hire experts of each data layer, preferably those who have computed them.

Yet the advantage of this method is its fully spatial aspect, which would lead to a reliability map at the same scale as the modelled map with assessment provided for each pixel, in line with EuSeaMap requirements. It would also enable the project to give feedback on physical layers to their providers, a condition of future improved deliveries.

However, given the above difficulties, this method was not deemed feasible within the EuSeamap frame but could be recommended for a later version.

1.2 Global assessment as a weighted sum of regional scores

1.2.1 Assessing the sediment layer

The global sediment layer is mostly historical, a result of the compilation of vintage maps made prior to the “acoustic era” with polygons drafted based on mere sample contents. Recent acoustic surveys have shown to which extent these maps gave a coarse view of the seabed reality, hence an urgent need to be quality assessed. In some cases, small size maps made from surveys are also incorporated to the final assemblage, hence The assessment of the vector substrate layer can inspire from the Mesh habitat map confidence assessment and be made per individual map. Basically the ways a substrate map and a habitat map are produced are quite similar and the method implemented in Mesh for habitat maps can easily be adapted to sediment maps.

Three types of parameters were considered (Annex 1):

- five parameters qualifying the remote sensing surveys
- six parameters qualifying the collection of field data
- another four parameters accounting for the quality of the map making process.

In Mesh these amounted to 15 individual scores which were then summarised into a final weighted sum. Most of these items hold true in the case of a sediment map except the one describing biological field data collection. A closer look could also be given to the interpretation section as to whether the four parameters are free of redundancy. The final score is a percentage. Typically low resolution vintage maps assessed with this method get scores in the range of 50 to 65%.

1.2.2 The Emodnet lots

It is noteworthy that the Emodnet project requires that each lot provides an accuracy assessment of its delivered layers. Two lots are of relevance to the EuSeaMap habitat map, namely bathymetry and geology (seabed substratum). However the geographic gaps between the lots are a drawback (e.g. there is no Emodnet substrate layer for the

Mediterranean as yet). Besides, the assessment of both the biological zones and the energy layers is not in their remits and thus rests fully with EuSeaMap. The geology lot is currently looking at a polygon-based assessment, whereby the map assessment methodology designed by Mesh is taken down to polygon level to express the confidence of the polygon content. This represents an intermediate way not as reliable as the full spatial pixel-based method but still yielding more detail than just using global scores as attempted here.

1.2.3 Assessing the depth zone layer

The depth zone layer is complex as each zone is made by intersecting different variables.

a) Definitions of the the infralittoral and upper circalittoral zones (the latter for the Mediterranean only) are based on the interaction between depth and respectively 1% and 0.01% of incident light. The light parameter Kpar is computed with an algorithm consistent over the whole of western Europe so no distinction can be made between locations. The accuracy with which this Kpar value is determined is bound to be fairly constant except in shallow waters where it is biased because of bottom reflection, yet with no consequence since these shallow waters are by definition infralittoral. Therefore as a simplification we can consider that the depth value is the main factor affecting confidence - at least in relative terms - on the determination of these two zones.

b) The Atlantic upper circalittoral is based on the wave base, itself the limit of wave-induced bottom disturbance computed using statistics from wave models. Similarly we have no way of simply assessing the accuracy of these statistics and are therefore reduced to consider them as consistent over the whole study area. The computation of the wave base was made using Gebco low resolution depth values, another data layer whose reliability is quite difficult to inform. Unless we use probabilities based on model computations we have no grip on the reliability of such individual layers.

c) The Mediterranean lower limit of the lower circalittoral zone is defined as the “shelf break line”, a sharper slope value located in the 120-180m range. Likewise the lower bathyal limit is the depth break line at the foot of the continental slope. In the Atlantic the lower circalittoral is more simply limited by the 200m depth line. These boundaries are directly affected by the quality of the underlying bathymetry so as a first approximation we could use a depth quality index to summarize the confidence on depth zones.

To summarise points a) to c), it appears that bathymetry is the only variable common to depth zones that could bear a criterion of quality.

1.2.4 Assessing the energy layer

We have no knowledge on the reliability of energy computed from hydrodynamical models of waves and currents, let alone on their statistical combination at the seabed. One very global way of assessing the energy would be to rank the fineness of these models, in fist approximation their resolution. When looking at this on the French part of the Channel, we actually have three types of models for currents (Rivier, 2010) going from the more offshore model with 2km resolution to coastal models with either 200m resolution at best . However let’s summarise the role of energy in our modelling. In the Atlantic wave data (wave length) are used for the low limit of the upper circalittoral zone (not the case in the Mediterranean). Currents and wave data ought to be used for rocky substrates in all basins but in the Mediterranean this had to be given up due to the coarseness of wave models close to the coast.

As a conclusion, given the relatively moderate role of energy data in the modelling, assessing the confidence in energy layers would remain marginal in an overall assessment. To lessen the contribution of this factor, we can also consider the quality of energy data as being reflected by that of the bathymetry.

1.2.5 Assessing the bathymetry

Resolution and data density

These are two different factors indeed and theoretically resolution should reflect data density. However it may happen that high resolution has been achieved even with low density data. The best way to account for this would be to have access to data density at each cell but this is beyond the scope of this assessment and left to deal with by the Emodnet hydrography lot.

To deal with quality assessment in more simple and feasible terms, we take it for granted DTM providers have appropriately adapted their cell size to the mean density of their cloud of points. Of course this assumption is likely to be false in places with lower density data but at our level we accept this fact and deem resolution does account for data density altogether. Therefore as a first approximation we are going to consider the resolution of individual DTMs as the leading factor.

We suggest to adopt:

- for hectometric resolution : score 1
- for 500m range resolution : score 2
- for kilometric resolution : score 3

Vintage

Usually individual DTMs are made using a variety of data sources and even though the adopted resolution is assumed to be representative of them all, it is important to take account of data vintage (as is shown fro example in the cartouche of nautical charts). This criterion is most important for soft bottom that change over time and probably of less relevance on hard bottom. It is common to see historical soundings collected over a century ago be collated along with more recent data to produce a DTM. According to practise it seems reasonable to split data age in three categories.

We suggest to adopt:

- before 1945 : score 1
- 1945 to 1980 : score 2
- 1980 to present day : score 3

Origin of data

What is relevant here is not the primary origin of data but that of actual data used to build the DTMs.

The most common source of data is from individual soundings and many local DTM have used certified digital bathymetry fair sheets. These soundings were collected in the old days with a lead line until the advent of single beam sounder. It would be very difficult to assess the quality of such data in great detail, as even very old lead data at the coast can be quite accurate. It is sufficient to say these data have been used by hydrographic surveys to produce their nautical charts and as such are guaranteed. Swath systems (e.g. multibeam) used in modern hydrography are recognised of highest quality. Associated with recent positioning techniques they provide highly relevant DTMs, even away from the coast.

In cases where the bathymetry fair sheets are not available an alternative solution is to pick out soundings from map values (points or isolines). In this case depth data are biased and of lesser quality.

To summarise we suggest to adopt:

- multibeam : score 3
- survey data (fair sheets) : score 2
- map data : score 1

The three factors above are reasonably decorrelated and even though they do not account for the full variety of depth data types and DTM processing modes, they can be taken as a first approach to measuring confidence.

Table 1: Example of scoring for French bathymetry maps

Depth data outline	**Data set name**	**DTM resolution** KM : 1 500m : 2 100m : 3	**Vintage** Y<1945: 1 1945<Y<1980 : 2 Y>1980 : 3	**Data origin** Chart values : 1 Fair sheets : 2 Multibeam : 3	**Total depth score (DS)**
Frame 1	1/10000 digital minute (Cap d'Agde)	3	2	2	**7**
Frame 2	Multibeam French EEZ	2	3	3	**8**
……					
Frame n					

Further considerations

A deeper insight into confidence would imply looking at how some of the DTMs were produced. Basically depth data come two from types of sounding tools. In the case of full coverage multibeam data, DTM reliability is very high at all locations and can be considered homogeneous. In the case of single beam sounders, as data density varies dramatically in the two directions within a single fair sheet, the quality of depth determination is commensurate to the interpolation quality (i.e. the interpolation error, easily obtainable for each pixel from advanced algorithms such as kriging). When such an error is available from the map producer, we can obtain a true spatial confidence assessment with a high spatial resolution.

1.3 Scoring maps in the Mediterranean basin

1.3.1 Sediment maps

Sediment maps come from many sources, as was illustrated by the report section on source data layers. For example on the French Mediterranean coasts four types of maps are concerned a) the Lima project sediment maps for Corsica at scale 1/100000, the "G series" maps from SHOM at scale 1/50000, the BRGM series for the PACA region at scale 1/50000 and the sediment map of the Golfe du Lion from the University of Perpignan (scale 1/250000). These maps originally all featured different classifications as adopted by their authors.

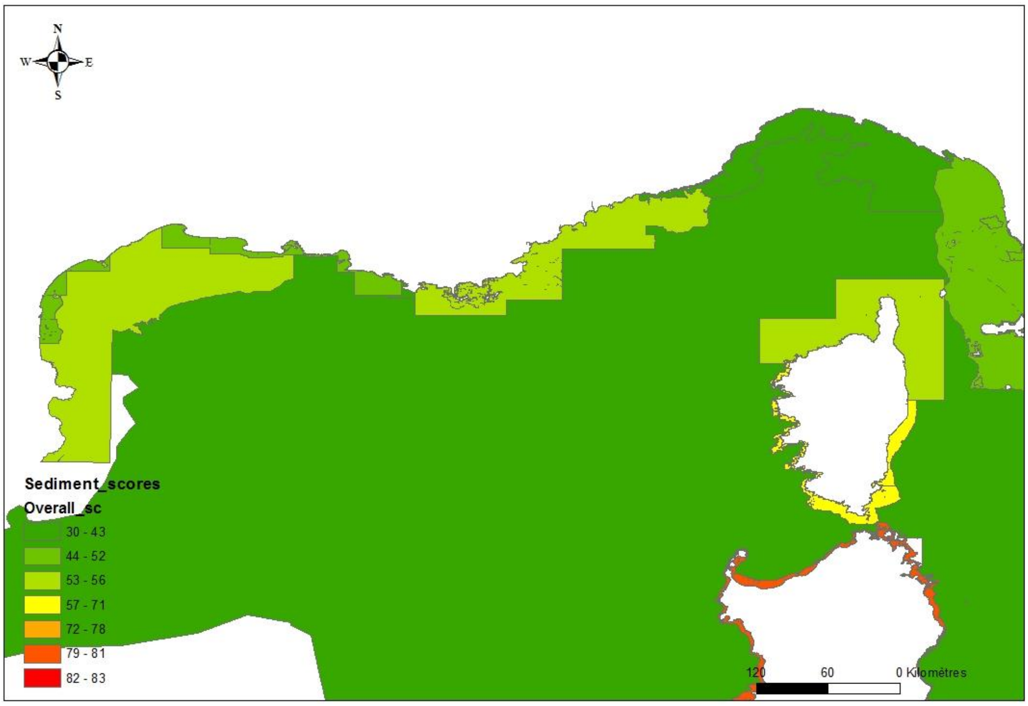

Figure 1: Scores of sediment maps for northern part of western Mediterranean.

1.3.2 Bathymetry maps

The EuSeaMap depth data in the Mediteranean come basically from three sources.

a) Participating countries have their own bathymetric DTMs built from local data sets. These are usually limited to continental platforms and in the case of steep shelves have a limited extension seaward. The original resolution of these data was usually higher than EuSeaMap's current resolution of 250m and therefore several pixels or soundings may have been averaged in 250m cells.

b) Further offshore a compilation of Mediterranean depth data is available from CIESM, with a current resolution of 500m. All of these were obtained from recent multibeam surveys, so their quality is homogeneous throughout the map.

c) In the centre of the basin and Tyrrhenian Sea, the default bathymetry is from Gebco (1 km resolution).

The simplest way to go about confidence would be to consider resolution as the key quality flag for each block of data, however this could be over-simplistic.

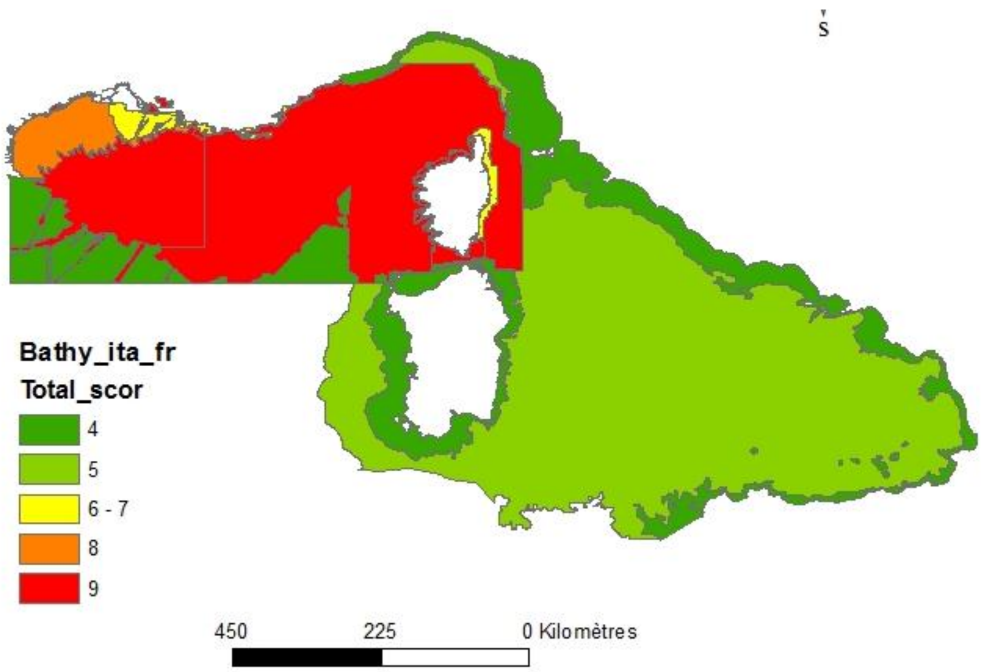

Figure 2: Scores of bathymetry maps for Italy and France.

1.3.3 Overall assessment: a weighted sum

Let DS and SS be respectively the bathymetry and substrate scores. The DS range is from 0 to 9 while the SS range is theoretically 0-100%. However even the highest quality substrate maps will never reach 100% (more currently 80%), so there may be a need to normalize these two values.

We suggest to apply the following formula to reach the final score:

$$FS (\%) = 0.5 [(8.3DS+5.1) + 100 SS] /100$$

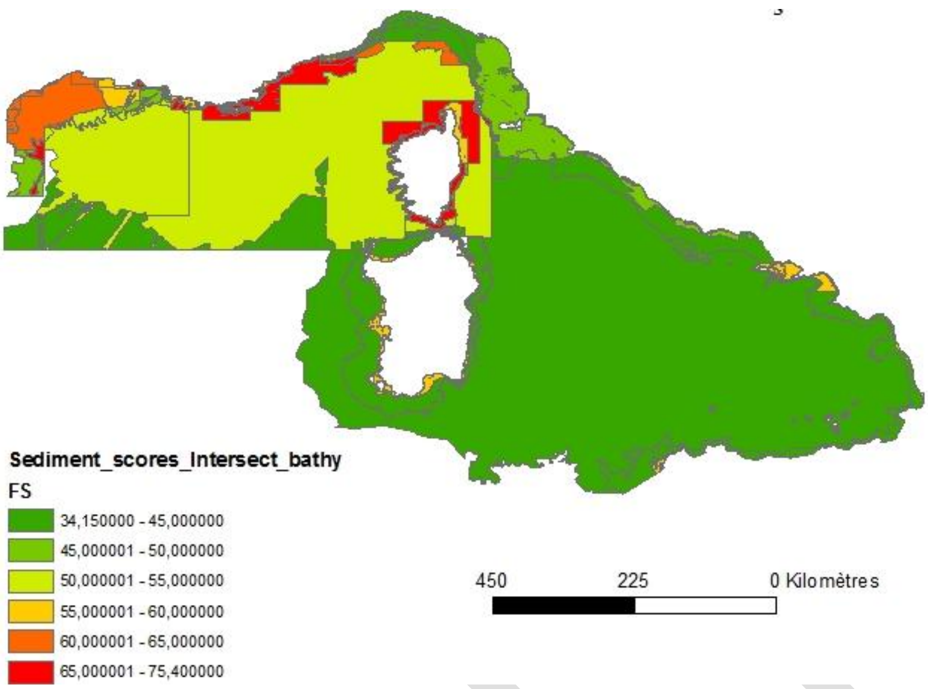

Figure 3: Global score for Italy and France.

2. Assessment of modelled maps versus recent habitat maps

2.1 The contingency matrix

The second way of assessing the EUSeaMap is referred to as an "external check" and is carried out by building a contingency matrix, i.e. a matrix with modelled data in one dimension and surveyed data from recent local habitat maps in the other.

Référence historique à la TD.

	Nephthys	Ophiura	Kelp	Faunal turf	Ophithrix
Nephthys	78	8	0	0	0
Ophiura	5	54	0	0	0
Kelp	0	0	8	0	0
Faunal turf	0	2	0	16	2
Ophithrix	0	0	0	1	18
% accuracy			90.6		
Kappa			0.89		

Figure 4: Example of contingency matrix (Mesh, 2008).

In our case, we have to make sure the assessment is significant, in other words that we are comparing comparable objects. In land use mapping polygons resulted from a classification of imagery, whereas here we are considering modelled polygons.

Dispite being of somewhat different nature, ground truth points and modelled polygons are both classified according to Eunis, however they are a number of reasons why they are likely to depart which need to be detailed before proceeding to a comparison.

2.2 Main features of benthic ground truth data

Ground truth data is a description of the habitats expressed at the locations of the seabed samples or observations used to interpret the acoustic imagery of the seabed. These can be of several types, either observations (still camera, video) or samples (grab, trawl).

On rocky substrates, only the former can be used but this type of seabed is seldom investigated which results in little knowledge of them. When observations were indeed made and could provide a Eunis level 4 description (biological communities present on the substrate), it would be useless for us as what we need is level 3 with the description of energy at the seabed. In this case the comparison would not possibly go beyond level 2.

The latter type of ground truthing by sampling the sediment lends itself to the analysis of both the sediment grain size and the endofauna. From there authors describe the seabed in categories which at best are Eunis-labelled or otherwise remain in the author’s classification. When translated these descriptions reach Eunis level 5 or possibly 6. However it may happen biologists do not describe the depth zone because they have no knowledge of it other than their measure of the depth and because the biology does not express it at all times, some sorts of endofauna being found irrespective of depth zones. This prevents the subsequent translation to Eunis and these types of sample data cannot be used.

2.3 Main features of the EUSeaMap modelled map

Likewise we need to summarise the main Eunis features of the EUSeaMap.

Substrate types come from historic sediment maps made by interpolating polygons from sample data. Therefore while being fully Eunis-compliant, these polygons have limited reliability away from the samples having defined them, leading to a potentially high discrepancy between the modelled map and the surveyed map.

Depth zones are defined from physical data layers (depth gradient, light, energy at the seabed) which do have a relation to the biology by virtue of proper thresholding burt are not a direct expression of a biological content.

Energy at the seabed is a result of combining currents and waves statistical data which are alos thresholded according to biological knowledge to ensure maximum relevance.

These three features when combined provide a Eunis level 3 map of the seabed. Note that in the relevance of seabed energy on hard substrate in the Mediterranean could not be established at the project’s resolution.

2.4 Choice of ground truth samples

Ground truth samples are best chosen from recent surveys where their authors have translated them to a Eunis code. This is the most reliable case and the comparison will be as far as possible based on such points. In UK a large number of widespread Eunis benthic samples are available which makes the comparison fully relevant for the Celtic and North Seas. In the Mediterranean only a limited number of Italian points are currently available. The Baltic is more of a problem since benthos samples there are usually not Eunis-labelled.

Another way of running the comparison when actual survey samples are not available is to select polygons centroids from recent biocenosis maps. However this degrades the quality of the reference data and is likely to make the comparison more hazardous. For these points the Eunis class was summarised to level 2 for rocky substrate and level 3 for soft bottom using a point to polygon spatial link.

2.5 Results and discussion

The comparison has been run over each basin in a stepwise way. The first step was to select only survey points and check they were in sufficient numbers to allow reliable statistics. Otherwise polygons centroids were added to the point file to make it larger and hence significant.

The second step was to carry out comparisons separately for rocky and soft substrate points. As has been explained above, on the former the comparison can only be run to Eunis level 2, whereas on the latter it is possible to level 3. Again, the cardinal number of the comparison points should be checked to provide reliable statistics.

Two scores were therefore produced that were subsequently added together by a weighted sum.

MORE

3. Conclusion - Overall assessment

We do not provide here an off-the-shelf method to assess the confidence of our modelled map but rather a suite of tools to advise the user of the map of its potential limitations. Quite often users take the content of a map for granted. While it can be done with high reliability for surveyed maps, it is not the case here mainly owing to the fact we used available historic data sources whose quality is very hard to appraise. The so called “external validation” also has obvious limitations because of the largely different nature of the modelled map on one hand and the individual samples on the other.

As was explained above, further to an insight into the quality of the individual layers there is a need to consider all boundaries between categories with caution since sharp boundaries are quite rare in nature. This is the reason why a third method for quality assessment was developed by building fuzzy limits around sharp boundaries to qualifying the uncertainty of their positions.

These views are tricky to reconcile. Rather, we advise the users to consult each of these “confidence layers” for its own right in order to make up their minds as best they can. On paper this would certainly be irksome, however today's webGIS capabilities offer this capability at no extra cost. It is planned in this contract to make available for each pixel of the modelled map the quality values found in the underlying layers.

4. References

ABP Mer report
Mesh confidence assessment
Emodnet geology document
Rivier, 2010