

Report of the EMODnet WP5 workshop
Yerseke 23-24/1/14

Present:

Peter Herman, Olivier Beauchard (NIOZ)

Lennert Tyberghein, Simon Claus (VLIZ)

Jean-Marie Beckers, Mohamed Ouberdous, Sylvain Watelet (Ulg)

Carlos Pinto (ICES)

Pierre Helaouet (SAHFOS)

Invited but excused:

Francisco Hernandez, Klaas Deneudt (VLIZ)

Abigail McQuatters-Gollop, Darren Stevens (SAHFOS)

Neil Holdsworth (ICES)

Alexander Barth (ULg)

Agenda

1. Progress in constructing gridded maps of example species
2. Technical issues
3. Implementation and automation of gridding procedures
4. Priorities and planning for coming half year and beyond

1. Progress in constructing gridded maps of example species

During the first EMODnet biology workshop (Ostend, September 2013) it was decided to select a number of well-known and published cases from diverse data sources to test the methodology. The selection was based on data availability, reference to existing literature and relevance to the project. The present state of progress in the different cases was discussed and great progress was made with several cases during the workshop. This document describes the present state of affairs and the remaining tasks, including time planning where relevant.

1.1. Cod in the North Atlantic

1.1.1. Data base

Data on cod abundance in standardized bottom trawls are available from ICES, covering the last decades. The species was selected because of its relevance for fisheries policy, and its potential indicator value for shifts related to climate change (cod is a cold-water species).

During data retrieval from EUROBI/EMODnet, an inconsistency was detected with the original ICES database. This inconsistency was solved during the workshop, thanks to ICES participation. It will be checked for all fish species in EMODnet and corrected where needed.

A new and correct data file, including calculated zero observations, was prepared for gridding during the workshop.

The file covers 14°-90° N, 30°W to 70°E; period 1971-2011. All data are from bottom trawl surveys. The data base contains information on size class and age class distributions, but not all of these data have been copied from the ICES database into the EMODnet database. It will be considered to change this in the future, depending on the data products to be made. For the WP5 gridding test, data on length was already included in the database.

1.1.2. Gridded products

Abundance maps covering the distribution of the species will be prepared for all years in the time span of data availability. In addition, an animation showing decadal changes will be prepared. Details of the methodology will depend on the characteristics of the map. A sliding time window (length to be determined) averaging the spatial fields over a number of years and advancing in time, will be used as the basis for the animation.

1.1.3. Timing

The basic gridding product (maps for every year) is ready. An animation will be prepared and optimized coming weeks.

1.1.4. Discussion points remaining

An important point of discussion is what other data products, based on this data base (and including length (and maybe age) information) could be made. Possibilities are maps of harvestable biomass, maps of reproductive stocks, animations of the fate of age classes, etc. It was decided that this question be submitted to fisheries biologists before decisions will be made. EMODnet will not go into great depth, but will try and find interesting 'first views' on the data sets that are also applicable to other species.

1.1.5. Where to go from here

It was decided to use the experience gained in the gridding of cod data to extend to other fish species. This should include at least one pelagic species (herring was chosen) and one true demersal species (plaice). Automation of scripts will be the main point of attention in this extension.

1.2. *Amphiura filiformis* (macrobenthos) in the North Sea

1.2.1. Data base

Two datasets covering most of the North Sea are available for macrobenthos: the ICES North Sea Benthos Survey of 1986, and the North Sea Benthos Project of 2000. Both datasets were used separately. They are based on grab or box core samples and provide numerical density and biomass values. Retrieval of the data from the EMODnet data base posed no problem. The only complication is that the North Sea Benthos project of 2000 included some of the 1986 data. For comparison purposes it was decided to make both data sets completely independent by withdrawing the 1986 data from the 2000 data set.

1.2.2. Gridded products

Gridding with DIVA was successfully applied to both data sets.

A technical issue arising from the test application was the determination of the correlation length (a parameter in DIVA). The automatic method to determine an optimal value of this parameter failed, due to the regular spacing of samples. We discussed this further under 'technical issues'.

This test case was used to investigate the feasibility of using DIVA on presence/absence data. Many datasets in EurOBIS only contain presence/absence information rather than numerical information on abundance or biomass. The 1986 data of *A. filiformis* were reduced to presence/absence (1/0), and different versions of DIVA were applied to this reduced data set. In all trials the probability of occurrence (0 to 1) was calculated in the gridding. By extrapolation, these probabilities sometimes slightly exceed the limits (0 and 1), which will be solved by a logit-like transformation in the software.

Fig. 1 compares the results of a version of the algorithm that does not bring down the interpolated values to the background value in regions where observations are scarce, with the standard version of the algorithm. Differences are relatively small, and where they exist, the standard DIVA version seems to produce the most reliable estimates in this case (where there are good reasons to assume that the species is absent in the deeper waters surrounding the North Sea). In other cases different conclusions might be reached, however.

The *A. filiformis* case has also been used to investigate how to bring habitat information into the gridded products. See 'technical issues' for this discussion and the results.

1.2.3. Timing

The gridded products are ready, including presence/absence gridding and gridding with account of depth as an environmental factor.

1.2.4. Discussion points remaining

None for this case

1.2.5. Where to go from here?

It was decided to select 10 other species from the macrobenthos databases, ranging from very common to rather rare species, and perform the gridding on their abundance data as well as on their presence/absence data. The species selected will also reflect different types of macrobenthos, according to their biological traits.

Further, the macrobenthos dataset of 1986, which was extensively analyzed and published, will be used to develop gridding products of community indicators. The idea is to first perform a multivariate analysis and then spatially grid the major axes of the multivariate analysis. The result can be compared to published results of the analyses done before.

1.3. *Calanus finmarchicus* and *Calanus helgolandicus* (zooplankton) in the N. Atlantic

1.3.1. Data base

The analysis is based on data provided by SAHFOS. Abundance data of both species have been made available for the period January 1958 till December 2012 covering the entire N. Atlantic ocean. The data used have been checked during the workshop, and the resulting maps compared with data products produced by SAHFOS. Good correspondence was found and it is believed that the data base is correct.

The data shows strong spatial gradients and patterns, as well as strong seasonal variation. Due to the large domain used, the seasons are not synchronous over the entire domain. A compromise will have to be used, unless one can accurately model the effect of season on the species in different spatial areas. This, however, is beyond the level of analysis that can be done in EMODnet.

1.3.2. Gridded products

Due to the size of the database and the large year-to-year variability, an appropriate presentation method must be found that is able to show the large-scale decadal changes in the distribution of these species. This presentation, when focusing on decadal changes, should also not be too much disturbed visually by seasonal dynamics or by year-to-year variability. We chose to calculate spatial fields for 10-year periods, and to use a sliding time window shifting one year at a time, to include into an animation. We will make separate long-term animations for the four seasons, but consider also other forms of animations, when appropriate, that focus on yearly averages or that do show seasonality as it has evolved over the past decades. During the workshop, we have prepared a DIVA4D script to perform the analysis of sliding time windows per season.

In order to keep the different gridded maps as comparable as possible, we have determined an optimal correlation length on the totality of the data and kept it fixed for all gridded products. All products have used the average density of the data used as background value.

We had to compromise with respect to the definition of the seasons. We have used Dec-Feb, Mar-May, Jun-Aug, Sep-Nov as seasons throughout the domain.

1.3.3. Timing

First gridded data product ready during workshop. Other animations will be prepared during the coming months.

1.3.4. Discussion points remaining

None for this case

1.3.5. Where to go from here?

We will use our experience with these two species to extend in two different directions. First, we will repeat the same gridding for the same two species, but with presence/absence rather than numerical abundance as input data. Most SAHFOS data are available in EUROBIS as presence/absence data, and it would be most worthwhile to know how useful these are in comparison with the full numerical data. Second, we will extend to new variables in the SAHFOS data set.

1.4. *Mnemiopsis* in the Black Sea

1.4.1. Data base

All field data on numerical abundance of this species in the Black Sea have been collected into a single data base. The IBSS in Sevastopol has made these data available. However, upon inspection of the data, strange order-of-magnitude differences between the different data sets were found, even in the same region and period. Inspection revealed that there were errors in the units used in the data base. A number of subsets of the data has been corrected, but contacting the data originators has not been successful for all data sets. Some serious doubts remain as the consistency of the entire dataset and this is currently under further scrutiny.

1.4.2. Gridded products

No products are available at the moment. A script for preparing the data files, executing the DIVA analyses and post-processing the results is ready, but it cannot yet be used because of the problem with the data set. As soon as reliable data are available, a gridded product can be made within hours.

1.4.3. Timing

Depends on operations in IBSS.

1.4.4. Discussion points remaining

Unclear, since no clear picture of reliable results is available. However, it can be expected that seasonal variation could be hampering the production of maps considerably. This point will have to be looked at once the data are available.

1.4.5. Where to go from here

There are two possible directions for an extension of the present case: extending the database of *Mnemiopsis* to other European seas, or extending the data presented on the Black Sea. The first direction is hampered by the fact that most data in NW Europe will not be freely available by the end of the EMODnet project. The second direction needs to be further explored, together with partners from the Black Sea area.

1.5. *Marezzelleria* in the Baltic

1.5.1. Data base

Two data bases have been used: one from Denmark (DHI – Alf Josefson) and one from Sweden (SMHI – Lotta Fryberg & Patrik Strömberg). The Danish data base has been completed with all zero observations and can be used as a basis for gridded mapping. The request for this data to SMHI was positively replied and as such absences from this dataset were generated as well (not yet included in products developed during the workshop).

A problem with the data set is that it is now clear that the genus has several species in the Baltic, and that there may have been several invasion events with different species involved. It is not possible to reconstruct this from the data that we received, since the majority of the data records does not have species identification. This hampers the interpretation and reduces the relevance of the case.

1.5.2. Gridded products

We were limited to the Danish part of the data base, because no zero observations were available for the Swedish part. *Marezzelleria* is a species mostly restricted to the near-coast area. It is not entirely clear why almost all observations are coastal. Either the species is limited to very shallow depths or to habitats that only occur in shallow water. Alternatively, the sampling effort may have concentrated on this area and thus have been unable to detect the species elsewhere by a lack of sampling. This will have to be clarified through consultation of the data originator.

Since all data were restricted to shallow depth in practice, we have restricted the gridding to this zone by adapting the correlation length based on a mask for shallow water (see below for technical details). The current maps are made for decadal periods. Data availability changes greatly from year to year, it is not likely that an animation based on yearly data sets will be very meaningful. An approach with sliding time windows will be tried for this case.

1.5.3. Timing

Gridded products can be made once data are available from Sweden. For Denmark a practical problem with the complicated coastlines needs to be solved. Once these points are solved, gridded products can be completed rapidly.

1.5.4. Discussion points remaining

Coastal restriction of the species. Species identification problem. These points will be raised with the data originators.

1.5.5. Where to go from here?

We might extend this data products with monitoring data from the Finnish coast to further complete the data products. We should find relevant variables with good data coverage for the Baltic Sea. A choice of these will be based on priorities within DEVOTES. We could also extend to other invasive species, if good data sets can be provided (e.g. *Ensis*?)

2. Technical issues

2.1. Completion of data bases with zero observations

The availability of zeroes is essential for the gridding procedure using DIVA. Therefore, a list of datasets for which zeroes can be reconstructed in a consistent way must be made before gridding can be automated with EMODnet biological data. VLIZ is constructing such a list. It will be used to determine further goals for the (semi-)automatic data gridding to be applied to many biological data.

2.2. Maps using presence/absence data only

The methodology of producing maps of probability of occurrence using p/a data with DIVA has been tried with *Amphiura filiformis* data in the North Sea, and with *Marenzelleria* data in Danish waters. Initially the method worked, however it needed to be finished by a pseudo logit transformation in order to avoid extrapolated values outside the [0,1] interval. This pseudo logit transformation is currently included in version 4.6.4 (to be made available officially in March 2014). The new version also allows easy use of the semi-normed diva version (indicator kriging where analysis is not forced to background values in regions void of data). These options only work in DIVA 3D or 4D. So for 2D, it is necessary to work with a single layer.

SAHFOS data are made publicly available only as presence/absence data, not numerical abundances. Testing how much can be shown based on this public subset of the database, in comparison to what is shown by the full data, is therefore important. This will be done using the *Calanus* data set.

2.3. Comparison with standard geostatistical models

In one case (*Amphiura filiformis*) we compared the DIVA output with standard geostatistical methods, in this case isotropic kriging based on a standard (spherical) variogram. The output of the two methods is compared in Fig. 1. In general, the outputs are very well comparable. The correlation length used in DIVA seems to be a bit smaller than that chosen in the kriging algorithm.

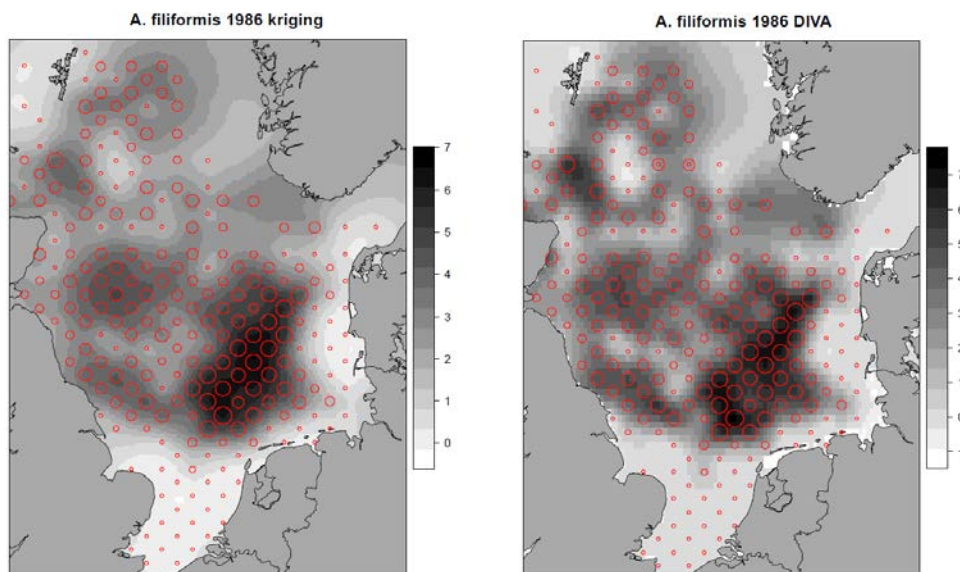


Fig. 1: Density (natural log scale) of *A. filiformis* in the 1986 North Sea Benthos Survey (ref). Size of marks proportional to observed density. Shades of gray produced by (left) DIVA interpolation and (right) ordinary kriging. (R script: plot_diva_A_f_86.R)

As a consequence, the DIVA interpolation follows the data somewhat more closely. Both methods also differ slightly in their extrapolation outside the domain of sampling. In general, however, the similarity between both dominates the picture. It is beyond the scope of the present project to do an in-depth comparison between both methods, but it is reassuring to see that for a relatively straightforward case as the present one, both methods do not differ much in their results.

2.4. Incorporation of environmental information into the gridding procedure

We investigated ways to incorporate environmental information into the gridding procedure. It is hoped that using independent physical, geological or chemical information could improve the interpolation, by restricting extrapolation of observations from favourable into unfavourable regions. Note that this is not a modeling approach as in regression models. We keep a strict interpolation point of view, but try to restrict information flow from one type of area into another type of area. This can be done in DIVA by spatially varying the correlation length. Where it is small, information does not flow from one point in space to another.

During the workshop, we worked out two examples. The first concerns *A. filiformis* in the North Sea. From the 2000 data set we know that the species is restricted to a depth range of approximately 30-150 m (Fig. 2)

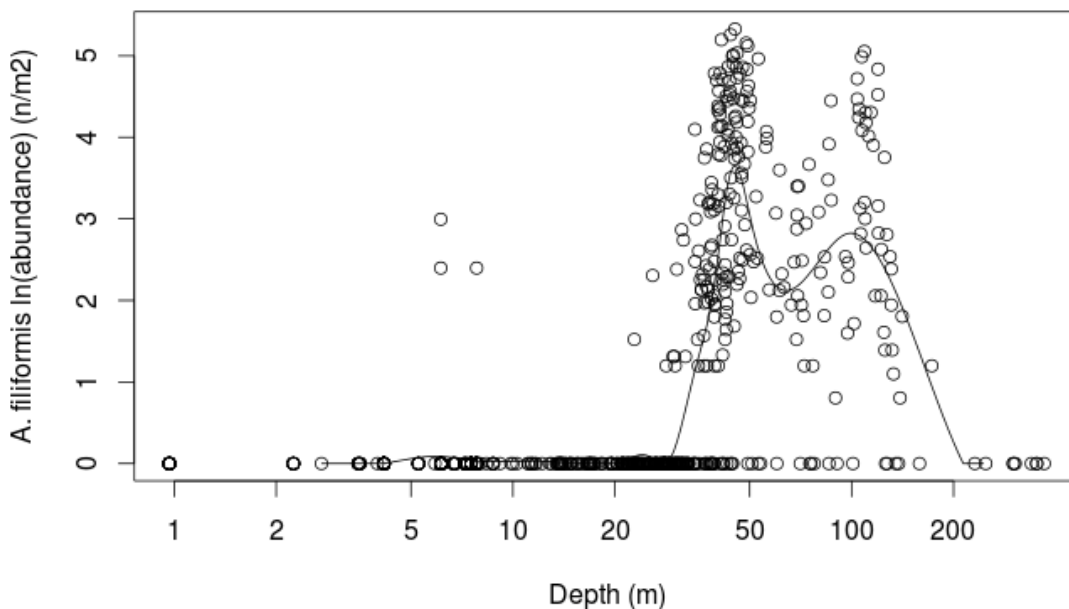


Fig. 2 (R script: af_depth.R)

From this information we created a mask, where favourable areas are white (value 1) and unfavourable areas are black (value 0), as shown in Fig. 3

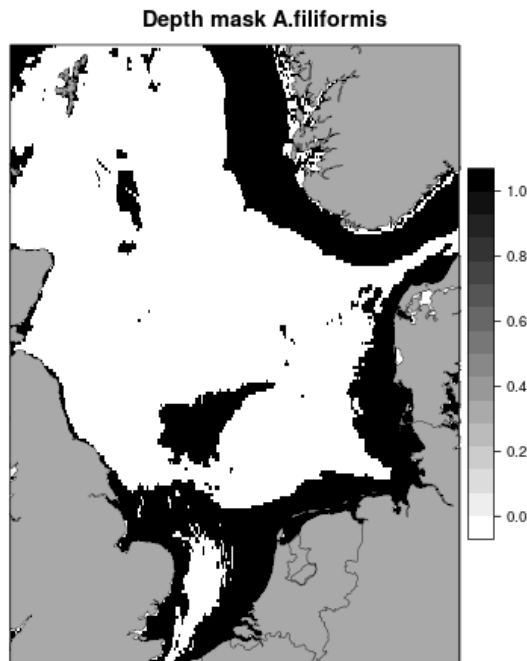


Fig. 3. Mask file based on bathymetry. (R script: DepthmaskAf.R)

Based on this mask, a map of correlation lengths with high value (0.5 degree) in the white area, and low (0.1 degree) in the black area, was created. This map of correlation lengths was used with data on presence/absence, to produce the interpolation map in Fig. 4. Note that within the white area, this interpolation field is very similar to the standard run, but that it has sharper boundaries and does not extrapolate the species' presence to the shallow Dogger Bank or the deep channel off Norway.

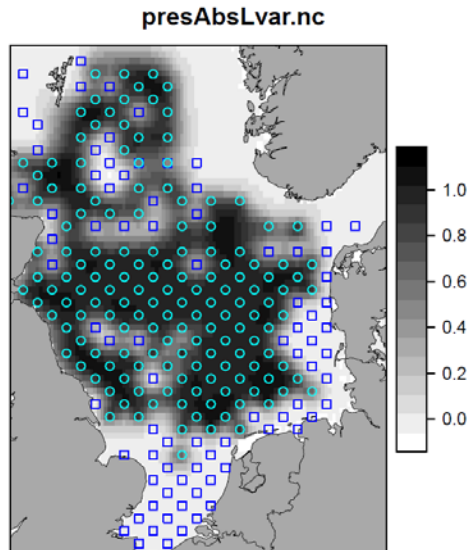


Fig. 4 Diva interpolation of *A. filiformis* in the North Sea, based on presence/absence data and incorporating environmental information by prohibiting extrapolation of presence into unfavourable habitats (R script plot_diva_A_f_pres_abs.R)

2.5. Gridding coastal species abundance

The *Marenzelleria* case posed a particular problem: how to grid abundance data of coastal species? When we can suspect that the species does not occur in deeper water, there is no point in extrapolating from the coast into these deeper waters. We approached this problem in the same way

as used for incorporation of environmental information in general. A mask was made from the coast to a depth of 10 m. Within this zone, correlation length was set much larger than outside. Thus, presence of the species in coastal waters did not lead to extrapolation far offshore, because in offshore waters information does not travel very far. This has led to the result in Fig. 5, for *Marenzelleria* sp. in Denmark during the years 2001-2012.

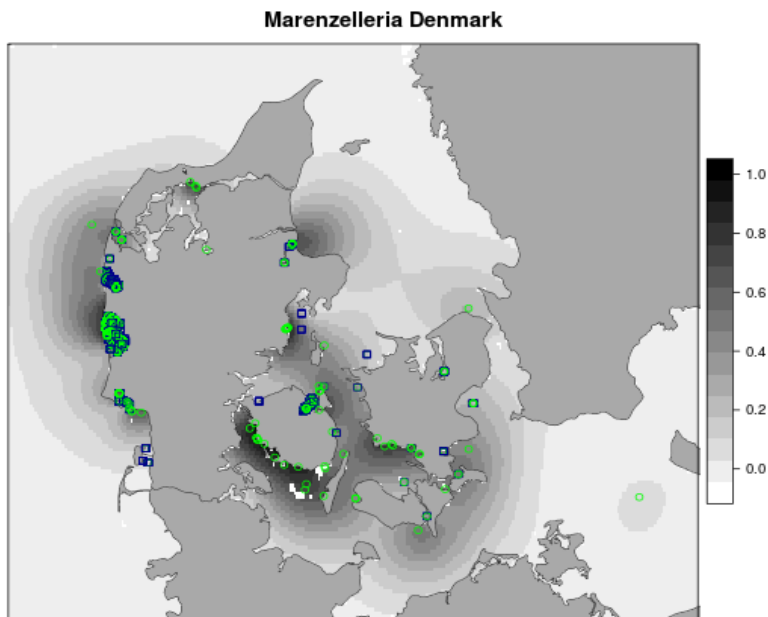


Fig. 5: Probability of occurrence of *Marenzelleria* sp. in Denmark for the years 2001-2012. Green circles are presences, blue squares absences in the observed data. (R Script: plot_diva_mar_DK.R)

In order to facilitate this particular problem, Ulg prepared a small program which from a simple mask creates the necessary files for using a smaller L in selected regions. Since this will be different for each species, not further automation was yet included and one needs to create the RL files individually and based on the mask info.

3. Implementation and automation of gridding procedures

In order to automate the gridding procedures a number of steps have to be taken. First, a list is needed of all datasets in EMODNET for which zero observations can be reconstructed. This list is currently being made by VLIZ.

Second, more experience needs to be gained with new datasets, in order to know how to optimise the gridding procedure. We will take two different actions. The first is to start from the present examples and work from there, adding more species from the same dataset. The second is to open new datasets and data domains, approaching it as we have done for the present cases: we start by selecting a single representative species and work from there to broaden the approach.

An ftp site is available for exchange of data and programs. All scripts used for pre- and post-processing will also be stored on this ftp server (ftp://emodnet_bio@ftp.vliz.be/).

The current configuration and data content of the ftp server looks like this:

Species_data:

- NSBS: *Scoloplos armiger*, *Goniada maculata*, *Amphiura filiformis*, *Nephtys hombergii*, *Chaetozone setosa*, *Owenia fusiformis*, , *Mysella bidentata*, *Ophelia borealis*, *Bathyporeia elegans*, *Lunatia poliana*, *Nephtys longosetosa*, *Echinocardium cordatum*, *Abra prismatica*, *Spiophanes kroyeri*, *Ophiura albida*, *Harpinia antennaria*, *Spio filicornis*
- Mnemiopsis: no products yet
- Marenzelleria: *Marenzelleria sp.*
- Gadus: *Gadus morhua*
- Calanus: *Calanus finmarchicus*, *Calanus helgolandicus*
- Amphiura: *Amphiura filiformis* 1986, 2000

Rscripts:

Meetings_Reports

Environmental_data

- Mask script Ulg
- Seabed habitat layer

4. Priorities and planning for coming half year and beyond

The following tasks were identified and will be performed in the coming months:

- setting up a specific timeframe for following tasks: Simon, Peter
- gridding a representative sample of North Sea benthic species: Olivier
- gridding results of multivariate analysis North Sea benthos: Olivier
- gridding North Sea birds and mammals, based on ESAS and SCANS data files: Olivier
- contact IBSS (Savastopol); completing the *Mnemiopsis* case: Peter
- Check for *Mnemiopsis* data outside Black Sea: Simon
- Clarifying sampling effort and completing the *Marenzelleria* case: Peter
- gridding a suitable additional variable from the SAHFOS data set: Pierre, Olivier
- SAHFOS will deliver new data on Total small copepods, Total diatoms, Total dinoflagellates, Phytoplankton Colour Index: Pierre, Lennert
- gridding presence/absence data of *Calanus* species for comparison: Olivier, Sylvain
- gridding additional fish species from ICES data base: Mohamed, Lennert, Olivier
- collecting phytoplankton data (e.g. French coastal monitoring program?): Lennert
- collecting data on seagrass (*Posidonia*) and algae (*Caulerpa*), both in Med Sea: Simon (contacting Ifremer), Lennert
- collecting data on other groups to be analysed: Lennert
- make up a list of all datasets in EMODNET for which zero observations can be reconstructed: Lennert
- Exploration of possibilities for maps of harvestable biomass, maps of reproductive stocks, animations of the fate of age classes, etc.: Peter
- Automation of the sub processes (data retrieval from biodiversity data bases, DIVA interpolation, post-processing) and exploration of automation of the complete process: Lennert, Peter
- Writing down guidelines and criteria for the creation of products:
- Publication of the results (importance of having a good official reference for the methodology of our products):