

Use of EMODNET Biology Data for invasive species policies. What can we learn?

Peter M.J. Herman

20 juni 2018

Introduction

Invasive species form an increasing part of marine communities worldwide. One of the responsible vectors for the spreading of invasive species is ship ballast water. Organisms can survive in ballast water taken in at some point in the world, and when not properly treated be released elsewhere. For this reason the use of ballast water is increasingly subjected to regulations. See <https://www.ospar.org/work-areas/eiha/shipping> and links therein for more information.

One of the key elements of the OSPAR “General Guidance on the Voluntary Interim application of the D1 Ballast Water Exchange Standard” are extensive checks of the ballast waters when ships enter harbours. This is a time-consuming process, that would probably not be justified when ships travel short distances between harbours entirely sharing their set of invasive species. One cannot introduce what is already there, after all. OSPAR and HELCOM have been working on a decision support tool that can help defining exemptions from the rules, where the danger of new introductions is deemed negligible.

In this analysis we explore what EMODNET could offer to the efforts devoted to this decision support tool. The EMODNET Biology product in this case is a workflow, rather than a fixed end product. We hope that the workflow, documented at code level in this product, can be taken up and further extended by the authorities responsible for invasive species management.

Pre-emption of the extensive checks is possible when the harbours have very similar fauna and essentially the same invasive species, because under this condition transport of water from one harbour to the other would change nothing in the distribution pattern of the invasive species. This is translated in the decision support tool under development by HELCOM and OSPAR (http://jointbwmexemptions.org/ballast_water_RA/apex/f?p=104:13:.....)

The decision support tool is based on extensive sampling of the fauna in different European harbours. This dataset is part of the website of the decision support tool, referred to earlier. It contains, in total, appr. 10000 records. In as far as it correctly documents which species do and do not occur in each of the harbours, it can be used to determine whether control for traffic between any pair of harbours is needed. However, these decisions could be flawed by false negatives. Suppose a species does occur in the harbour of departure, but was not detected in this dataset. This is a ‘false negative’. Suppose moreover that the species is absent in the harbour of destination. Then, based on the false negative, a clearance would be given whereas precaution would have been needed. In reverse, if a species is present in the harbour of departure and in the harbour of arrival, no special precaution is needed. However, unnecessary precaution would be applied in the case of a false negative in the harbour of arrival. There is, in short, a need for control for false negatives as they may lead to either unnecessary costs or unnecessary environmental risks.

In this product we use the EMODNET data base as a tool to check for false negatives in the harbour dataset. We assume that this dataset has been correctly assembled, so that any positives reflect the real occurrence of the species in the harbour. Therefore we do not check the positives. We discuss this choice in the discussion section.

Here we document the technical approach to the problem, given in the exact procedure followed for the data analysis. This is included not only for transparency, but also to facilitate adoption of this approach by others. Next we shortly discuss some examples of results, and their possible implications for the decision support tool.

We have not invested much effort in the quality of the graphics in this analysis. Although in general we realise that a good representation of results helps in conveying a message, we consider the workflow in this product as a technical contribution to a more general workflow that will also take care of a proper representation of results.

Approach

We have downloaded the list of target species and candidate target species from the website of the tool ([http://jointbwmexemptions.org/ballast_water_RA/apex/f?p=104:12:::~::~](http://jointbwmexemptions.org/ballast_water_RA/apex/f?p=104:12:::)). This list was manually checked in WoRMS (<http://www.marinespecies.org>), and accepted scientific names and AphiaIDs were looked up. AphiaIDs are unique numerical identifiers of species - see WoRMS website for details. In EMODNET biology records, all species are identified by their AphiaID. Subsequently, a list of unique species was drawn from the harbour data base, and this list was also manually submitted to WoRMS. We subsequently produced maps of the distribution, both in the EMODNET database and in the harbour database. The maps are a basis for further reflection on the use of the harbour database, and can be used by experts to improve the decision support tool. The maps can also show whether the EMODNET database can bring useful additional information for this further improvement.

Analysis

We begin the analysis by reading the harbour database (which was manually downloaded from the tool website) and extracting the list of species in that database. This list of species is written to a text file, in order to be submitted manually to WoRMS. However, before doing so we had to make a number of changes to the original species names. Some were not correctly spelled, others were not marine, not proper scientific names, or named forms or varieties that were not included in WoRMS. Based on this list of problems, we collected a number of corrections to be performed in the basic data set `hd`, before a species list can be extracted and submitted to WoRMS. This correction is done in the following code.

```
#####
##  read and species-check harbour data base #####
#####
hd<-read.csv("occddata.csv",stringsAsFactors = F,header = T)
#
##### this section treats generic problems #####
# remove all sp, sp., spp, spp. substrings
ll<-which(substr(hd$Species.Name,(nchar(hd$Species.Name)+1)-5,nchar(hd$Species.Name)))==" spp.")
hd$Species.Name[ll]<-substr(hd$Species.Name[ll],1,nchar(hd$Species.Name[ll])-5)
ll<-which(substr(hd$Species.Name,(nchar(hd$Species.Name)+1)-4,nchar(hd$Species.Name)))==" sp.")
hd$Species.Name[ll]<-substr(hd$Species.Name[ll],1,nchar(hd$Species.Name[ll])-4)
ll<-which(substr(hd$Species.Name,(nchar(hd$Species.Name)+1)-4,nchar(hd$Species.Name)))==" spp")
hd$Species.Name[ll]<-substr(hd$Species.Name[ll],1,nchar(hd$Species.Name[ll])-4)
ll<-which(substr(hd$Species.Name,(nchar(hd$Species.Name)+1)-3,nchar(hd$Species.Name)))==" sp")
hd$Species.Name[ll]<-substr(hd$Species.Name[ll],1,nchar(hd$Species.Name[ll])-3)
# replace "v." or "var" by "var."
hd$Species.Name<-gsub(" v. ", " var. ",hd$Species.Name)
hd$Species.Name<-gsub(" v ", " var. ",hd$Species.Name)
hd$Species.Name<-gsub(" var ", " var. ",hd$Species.Name)
##### this section treats specific problems with the names #####
##
##### first we define a function to replace a name in the dataset ###
corr_name<-function(old_name,new_name,hd){
  if(new_name=="delete"){
```

```

    ll<-which(hd$Species.Name==old_name)
    hd<-hd[~ll,]
  }else{
    hd$Species.Name[hd$Species.Name==old_name]<-new_name
  }
  return(hd)
}
##### then we perform all corrections #####

hd<-corr_name("Amphidinium af. simplex","Amphidinium simplex",hd)
hd<-corr_name("Clausocalanus styliremis","Clausocalanus styliremis",hd)
hd<-corr_name("Dolichospermum cf. affine","Dolichospermum affine",hd)
hd<-corr_name("Dolichospermum cf. circinale","Dolichospermum circinale",hd)
hd<-corr_name("Dolichospermum cf. lemmermannii","Dolichospermum lemmermannii",hd)
hd<-corr_name("Dolichospermum cf. spiroides","Dolichospermum spiroides",hd)
hd<-corr_name("Escherichia coli bglucuronidasa +","Escherichia coli",hd)
hd<-corr_name("Koliella longiseta f. longiseta","Koliella longiseta",hd)
hd<-corr_name("Microdeutopus gryllotalpa/propinquus","Microdeutopus gryllotalpa",hd)
hd<-corr_name("Nitzschia af. fusiformis","Nitzschia fusiformis",hd)
hd<-corr_name("Oocystis cf. solitaria","Oocystis solitaria",hd)
hd<-corr_name("Phaeophyta/Rhodophyta","Rhodophyta",hd)
hd<-corr_name("Rhizosolenia hebetata f. Semispina","Rhizosolenia hebetata f. semispina",hd)
hd<-corr_name("Stauroneis cf. gracillima","Stauroneis gracillima",hd)
hd<-corr_name("Unicell","delete",hd)
hd<-corr_name("Unknown","delete",hd)
hd<-corr_name("biofilm","delete",hd)
hd<-corr_name("Chaetoceros spore","Chaetoceros",hd)
hd<-corr_name("Aeromonas veronii","Aeromonas",hd)
hd<-corr_name("Agraylea","delete",hd)
hd<-corr_name("Amphidinium simplex","Amphidinium",hd)
hd<-corr_name("Aulacoseira islandica helvetica","Aulacoseira islandica",hd)
hd<-corr_name("Aulacoseira islandica islandica","Aulacoseira islandica",hd)
hd<-corr_name("Calanoidea","Calanoida",hd)
hd<-corr_name("Chaetoceros ceratosporus var. ceratosporus","Chaetoceros ceratosporus",hd)
hd<-corr_name("Chaetoceros thronsenii var. thronsenii","Chaetoceros thronsenii",hd)
hd<-corr_name("Chrysophyte","Chrysophyceae",hd)
hd<-corr_name("Clausocalanus styliremis","Calocalanus styliremis",hd)
hd<-corr_name("Cocconeis didyma","Diploneis didyma",hd)
hd<-corr_name("Desmodesmus armatus var. armatus","Desmodesmus armatus",hd)
hd<-corr_name("Desmodesmus opoliensis var. opoliensis","Desmodesmus opoliensis",hd)
hd<-corr_name("Diaphanosoma mongolianum","Diaphanosoma",hd)
hd<-corr_name("Eubosmina maritima","Bosmina (Eubosmina) coregoni",hd)
hd<-corr_name("Filamentous phaeophyceae","Phaeophyceae",hd)
hd<-corr_name("Fish","Pisces",hd)
hd<-corr_name("Intestinal enterococci","Enterococcus",hd)
hd<-corr_name("Lagerheimia longiseta var. longiseta","Lagerheimia longiseta",hd)
hd<-corr_name("Mytilus edulis trossulus","Mytilus trossulus",hd)
hd<-corr_name("Nauplii","Copepoda",hd)
hd<-corr_name("Navicula viminoides var. cosmomarina","Navicula viminoides",hd)
hd<-corr_name("Pediastrum boryanum var. boryanum","Pediastrum",hd)
hd<-corr_name("Pediastrum duplex var. duplex","Pediastrum duplex",hd)
hd<-corr_name("Pennaes","Bacillariophyceae",hd)
hd<-corr_name("Pseudokirchneriella contorta","delete",hd)

```

```

hd<-corr_name("Rhizosolenia alata indica","Rhizosolenia indica",hd)
hd<-corr_name("Sygalium squamatum","delete",hd)
hd<-corr_name("Tabellaria flocculosa var. asterionel","Tabellaria flocculosa var. asterionelloides",hd)
hd<-corr_name("Thalassiosira oestrupii var. venride","Thalassiosira oestrupii var. venrickae",hd)
hd<-corr_name("Vibrio mimicus","Vibrio",hd)
hd<-corr_name("Acaria","Acari",hd)
hd<-corr_name("Acarina","Acari",hd)
hd<-corr_name("Heterocapsa arctica frigida","Heterocapsa arctica subsp. frigida",hd)
hd<-corr_name("Kellicottia longiseta","Kellicottia longispina",hd)
hd<-corr_name("Pseudopediniella","Pseudopedinella",hd)
hd<-corr_name("Thalassiosira levanderii","Thalassiosira levanderi",hd)
hd<-corr_name("Thalassiosira eccentric","Thalassiosira eccentrica",hd)
hd<-corr_name("Cercopagis pengoi","Cercopagis (Cercopagis) pengoi",hd)
hd<-corr_name("Bosmina longirostris","Bosmina (Bosmina) longirostris",hd)
## we extract the list of species names, order it and write to file for submission to WoRMS
splist<-unique(as.character(hd$Species.Name))
splist<-splist[order(splist)]
write.csv(splist,file="hs2.txt",row.names = F)

```

The species list that was written in the last step was then submitted to WoRMS, and enriched with the AphiaID and taxonomy. The result of this offline operation is stored in “hs2_matched.csv”. It will be merged with the dataframe containing the harbour data.

We also read the list of target species, that has also been matched with WoRMS, and is stored as “tsm_matched.csv”, and order them alphabetically.

```

# corrected species names and aphiaIDs from WoRMS
sptax<-read.csv("hs2_matched.csv",stringsAsFactors = F,header = T)
hd<-merge(hd,sptax,by="Species.Name")
# target species names and aphiaIDs, as derived from the list of OSPAR, WoRMS corrected
tsp<-read.csv("tsm_matched.csv",stringsAsFactors = F,header = T)
tsp<-tsp[order(tsp$ScientificName_accepted),]
# a file with (approximate) geographic coordinates of all harbours concerned
harbpos<-read.csv("harbpos.csv",stringsAsFactors = F,header = T)
hd<-merge(hd,harbpos,by="HarbourName")

```

With all the taxonomic controls done, we can now proceed to producing maps. For all target species and all harbours, we map whether the species has been found in the harbour, whether (in absence of the species) a congeneric species has been found, or whether neither species nor genus have been found in the harbour. We use a colour code for these cases. On top of that, we also look up the locations where the species (not the genus) has been recorded in the EUROBIS data base. We do that online, using the EMODNET Biology WFS service. We plot all occurrences of the species on the same map with the harbour data, but use a different symbol.

The following code defines a function that, for a single species, looks up all occurrences in the EMODNET data base and plots these together with the observed occurrences of the species and the genus in the harbours database.

```

require(sp)
require(maps)
require(mapdata)
# make a SpatialPointsDataFrame object based on the file of harbour positions
coordinates(harbpos)<- ~lon+lat
# function to make the map with harbour and EMODNET occurrences of a single species

```

```

plotsp<-function(name,genus,aphia){
  # make a selection of data base containing all records of the species in the harbours
  thd<-hd[hd$ScientificName_accepted == name,]
  # same for the genus
  thdc<-hd[hd$Genus == genus,]
  # make a vector of harbours where the species was found
  thdpres<-unique(thd$HarbourName)
  # make a vector of harbours where the genus was found
  thdgenpr<-unique(thdc$HarbourName)
  # make a column in the SpatialPointsDataFrame harbpos with coded values for presence
  # of the species concerned: 1 if species present, 2 if genus present, 3 if both absent
  harbpos$p<-ifelse(harbpos$HarbourName %in% thdpres,1,
                    ifelse(harbpos$HarbourName %in% thdgenpr,2,3))
  # define plotting colours for the three categories
  cols<-c("red","black","green")
  # now retrieve all occurrences of the species in the EMODNET data base
  # first define the url for the WFS
  spURL<-paste("http://geo.vliz.be/geoserver/Dataportal/ows?service=WFS",
               "&version=1.0.0&request=GetFeature",
               "&typeName=Dataportal:eurobis&viewParams=where:aphiaidaccepted=",
               aphia,"&outputformat=csv",sep="")
  # then download the data
  eod<-read.csv(spURL,stringsAsFactors = F)
  # make plot of occurrences of species/genus in the harbours
  plot(harbpos,pch=19,cex=2,col=cols[harbpos$p],main=name,asp=1,xlim=c(-20,30),ylim=c(25,65))
  # add a world map
  map('worldHires',add=T)
  # add all occurrences of the species in the EMODNET data base
  points(harbpos,pch=19,cex=2,col=cols[harbpos$p])
  points(eod$longitude,eod$latitude,pch=15,cex=.8,col="red")
  legend("bottomright",legend=c("Species present in harbour","Species absent, genus present in harbour",
                                "Species and genus absent in harbour","Species present in Eurobis"),
        pch=c(19,19,19,15),col=c("red","black","green","red"),cex=.5,pt.cex=c(1,1,1,.6),
        bg="white")
}

```

In the subsequent code chunk, we produce the maps of all target species. These maps are stored in a separate pdf file (maps.pdf).

```

pdf("maps.pdf")
for(i in 1:nrow(tsp)){ # loop over all target species
  plotsp(name=tsp$ScientificName_accepted[i],genus=tsp$Genus[i],aphia=tsp$AphiaID_accepted[i])
}
dev.off()

```

```

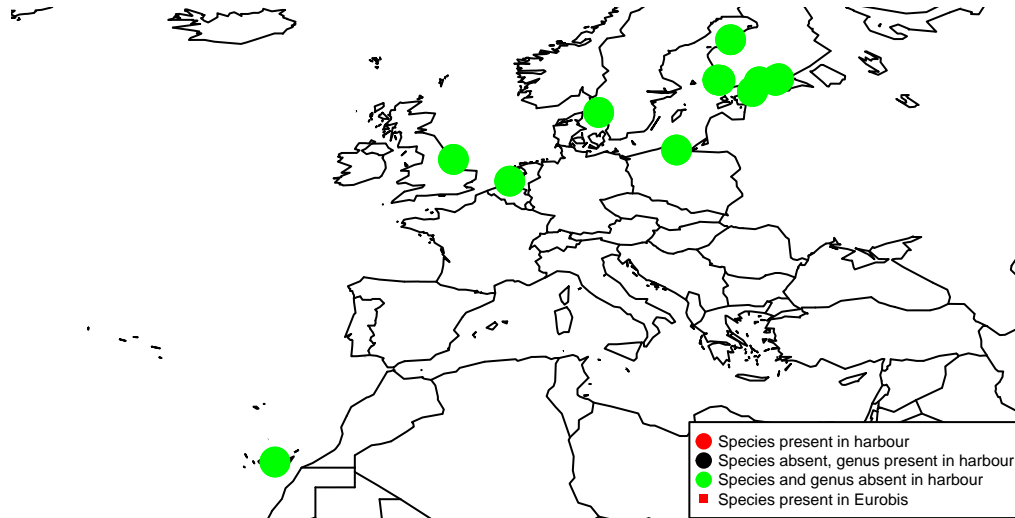
## pdf
## 2

```

Results

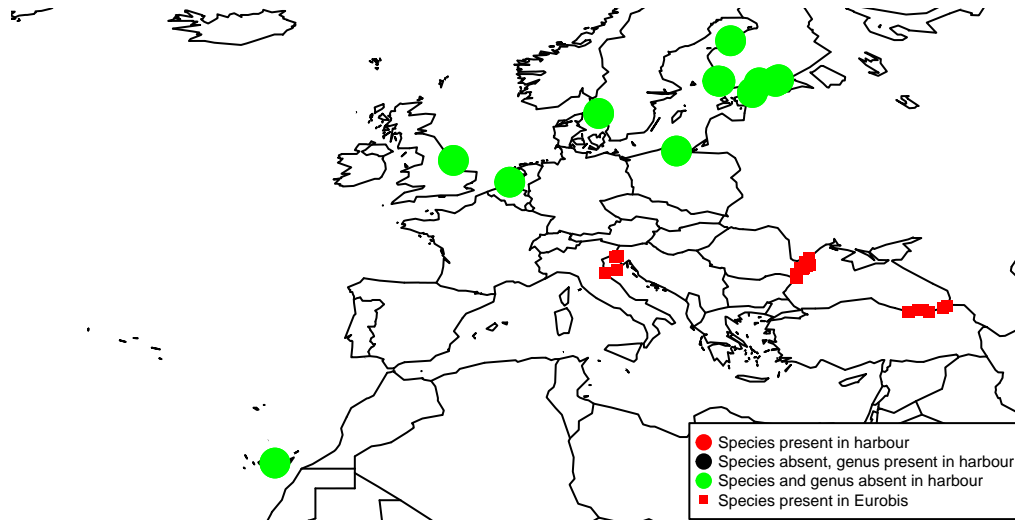
We divided the species in a number of categories. The first category (Group I) consists of species that have been found in none of the harbours, and whose absence is confirmed by the complete absence of EUROBIS records in the mapped part of Europe (including the Atlantic coasts, Baltic, and NW Mediterranean). An example of this type of species is *Ponticola kessleri*.

Potamocorbula amurensis



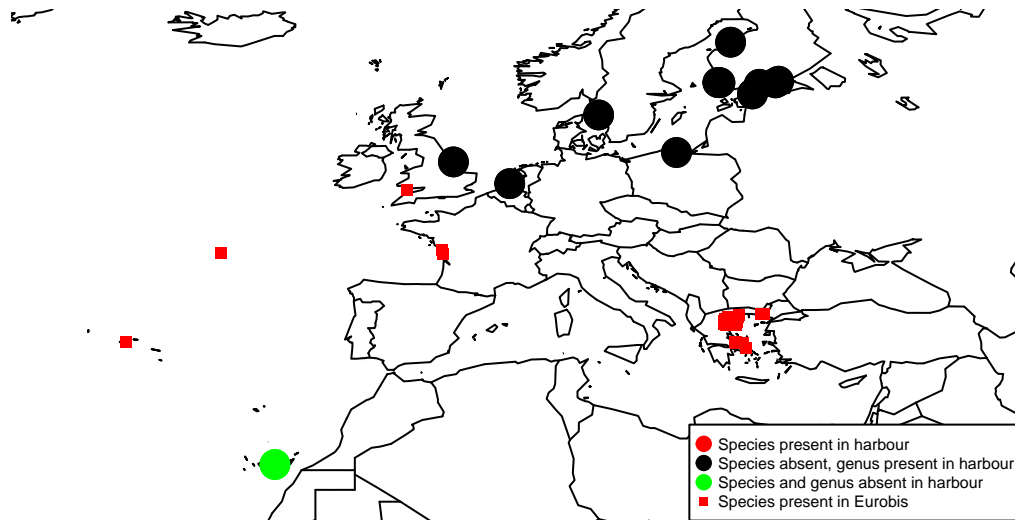
Species in Group II were also not found in any of the harbours, were found in the NW Mediterranean, but not in the Atlantic and Baltic parts of Europe. An example of this group is *Anadara inaequalis*. Absence of these species from the considered harbours is also very likely, although invasion may have a higher probability than for the first list of species.

Anadara inaequalis



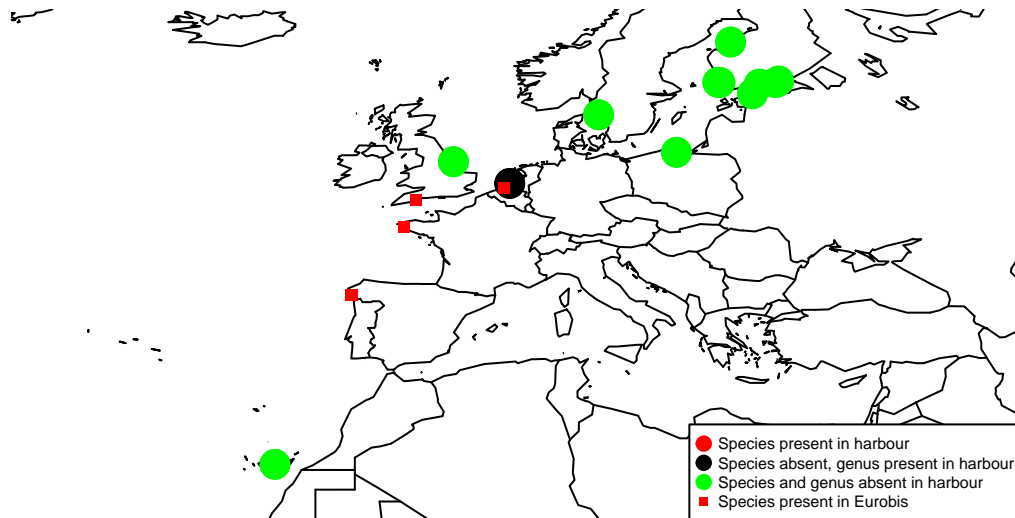
Group III species have also not been recorded in the harbours, but are known from Western Europe at relatively long distances from the harbours considered. An example is *Amphibalanus eburneus*. In all of these cases it is up to judgement what is ‘far’ from the harbours, but this list seems to contain mostly species that can reasonably be thought to be (still) absent in the harbours considered.

Amphibalanus eburneus



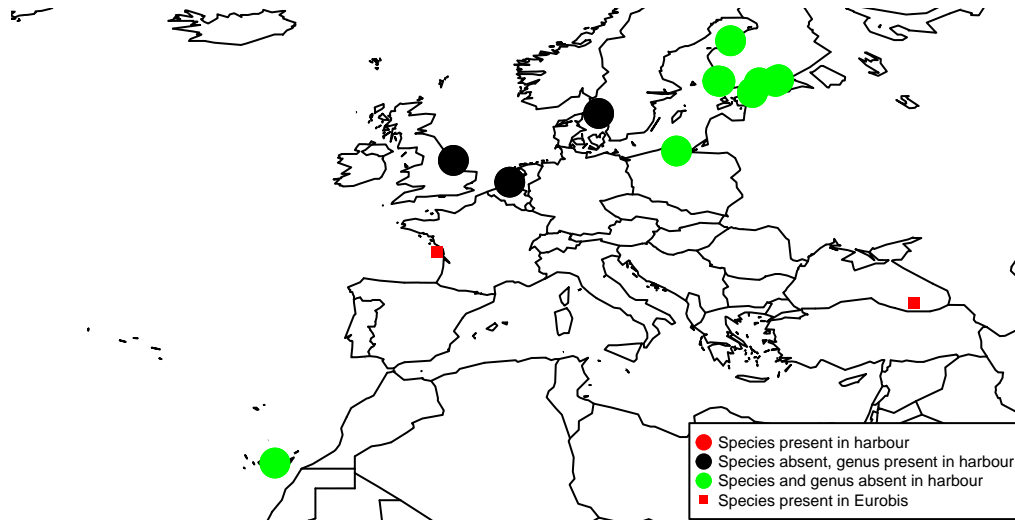
Group IV consists of species that have not been recorded from any of the harbours, but have been found in EUROBIS at locations in the proximity of the harbours. These findings suggest that the negatives in some harbours run the risk of being false negatives, even though there might be some ecological differences between the places of occurrence and the harbours. For all Group IV species, the number of occurrences in EUROBIS was relatively small. This is the case for the following species: *Asterias amurensis* (close to Rotterdam), *Callinectes sapidus* (close to Rotterdam), *Caulerpa cylindracea* (close to Las Palmas), *Caulerpa taxifolia* (close to Las Palmas), *Corbicula fluminea* (close to Rotterdam), *Dikerogammarus villosus* (close to Rotterdam), *Dinophysis sacculus* (close to Rotterdam, Hull and Las Palmas), *Fibrocapsa japonica* (close to Gothenburg), *Gracilaria vermiculophylla* (close to Gothenburg and Rotterdam), *Grateloupia turuturu* (close to Rotterdam), *Hemimysis anomala* (close to several harbours, especially in Finland but also elsewhere), *Hydroides dianthus* (close to Rotterdam), *Hydroides elegans* (close to Hull), *Maeotias marginata* (close to eastern Baltic harbours), *Mytilus galloprovincialis* (close to Rotterdam), *Palaemon macrodactylus* (close to Rotterdam), *Pseudochattonella verruculosa* (close to Gothenburg and Gdynia), *Rapana venosa* (close to Rotterdam), *Undaria pinnatifida* (close to Rotterdam and Hull).

Asterias amurensis



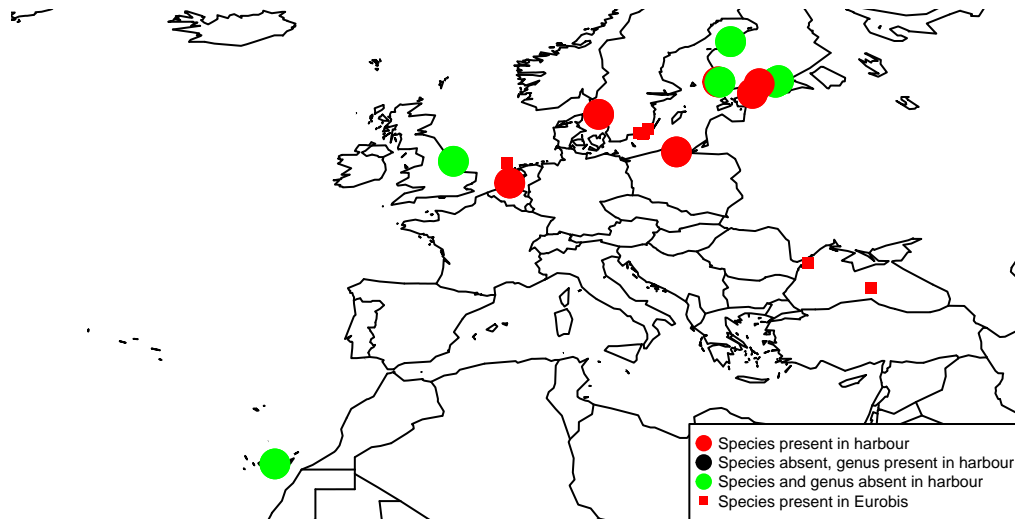
Species in Group V show many occurrences in EUROBIS, often over a large area, but have, remarkably, never been recorded in any of the harbours. An example of these species is *Phaeocystis pouchetii*. It is likely that seasonality of the sampling has played an important role in this absence. For other species, such as *Ensis leyi*, taxonomic confusion may have played a role.

Polysiphonia morrowii



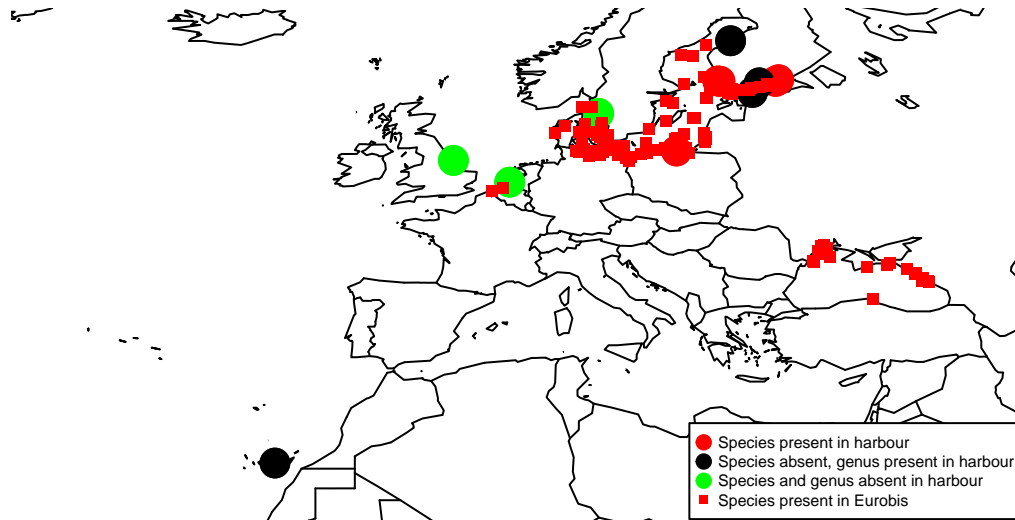
Among the species that have been recorded in at least one harbour, Group VI consists of a single species. This is *Neogobius fluviatilis*, a species that was found in one of the harbours (Rotterdam) but had not yet been recorded in the EUROBIS database for NW Europe. Note however that a congeneric species (*Neogobius melanostomus*) has been recorded several times. *Neogobius fluviatilis* was known in the data base from the Black Sea and may have invaded recently.

Neogobius melanostomus



All the other species (Group VII) have been found in at least one of the harbours, but very often there seems to be a high chance that some of the harbours report a false negative. An example is *Acartia* (*Acanthacartia*) *tonsa*, a species that is widely distributed in the Baltic and down to the Southern Bight of the North Sea, but has not been reported from several harbours in its region of distribution.

Acartia* (*Acanthacartia*) *tonsa



A summary of the groups to which the different species belong, is provided in Figure 8.

	Group	Absent in all harbours					Present in at least one harbour	
		absent in database	Med Sea in database	in database few and far	in database few but close	in database abundant	not in database	in database close to negatives
Alexandrium monilatum	I							
Anadara transversa	II							
Arcuatula senhousia	III							
Babka gymnotrachelus	I							
Blackfordia virginica	I							
Brachidontes pharaonis	I							
Callinectes sapidus	IV							
Caulerpa cylindracea	IV							
Caulerpa taxifolia	IV							
Chama pacifica	I							
Chionoecetes opilio	I							
Corbicula fluminea	IV							
Coscinodiscus wailesii	VII							
Crepidula fornicata	VII							
Didemnum vexillum	I							
Dikerogammarus villosus	IV							
Dinophysis sacculus	IV							
Dreissena polymorpha	VII							
Ensis leei	V							
Fibrocapsa japonica	IV							
Gammarus tigrinus	VII							
Gracilaria vermiculophylla	IV							
Grandidierella japonica	III							
Grateloupia doryphora	III							
Grateloupia turuturu	IV							
Halophila stipulacea	II							
Hemimysis anomala	IV							
Hydroides dianthus	IV							
Hydroides elegans	IV							
Karenia mikimotoi	V							
Lophocladia lallemantii	II							
Maeotias marginata	IV							
Magallana gigas	VII							
Marenzelleria neglecta	VII							
Marenzelleria viridis	VII							
Melanothamnus harveyi	III							
Microcosmus squamiger	I							
Mnemiopsis leidyi	VII							
Mytilopsis leucophaea	VII							
Mytilus galloprovincialis	IV							
Palaemon elegans	VII							
Palaemon macrodactylus	IV							
Paralithodes camtschaticus	I							
Pfiesteria piscicida	I							
Phaeocystis pouchetii	V							
Polysiphonia morrowii	III							
Ponticola kessleri	I							
Potamocorbula amurensis	I							
Potamothenix vejovskyi	I							
Pseudochattonella verruculosa	IV							
Rangia cuneata	VII							
Rapana venosa	IV							
Spondylus spinosus	I							
Styela clava	VII							
Undaria pinnatifida	IV							
Alexandrium acatenella	I							
Alexandrium ostenfeldii	V							
Amphibalanus eburneus	III							
Anadara inaequalis	II							
Asterias amurensis	IV							
Cercopagis (Cercopagis) pengoi	VII							
Eriocheir sinensis	V							
Neogobius fluviatilis	VI							
Styopodium schimperi	I							

Figure 1: Overview of the groups of species, according to occurrence in harbour data and in EUROBIS database

Discussion

The examples discussed and shown here demonstrate that the information from the EUROBIS database can help in the interpretation of the data set collected in the harbours. As indicated, there is a high probability that many of the observations contain false negatives. Of course, one cannot prove this beyond doubt, but if a species is widely distributed in the neighbourhood of a harbour, there is at least a reasonable chance that the species or its propagules occur in this harbour. As indicated in the introduction, the detection and correction of false negatives can affect the decisions as to whether control of ballast water is needed between pairs of harbours. False negatives can increase costs, but also environmental risk, depending on whether the false negative occurs in the harbour of depart or arrival.

Harbours that are very far apart will always require control. Harbours that are very close together have a high probability of sharing the same invasive species. Therefore, judgment is most important for harbour pairs at intermediate distances. Should Gothenburg be separated from the other harbours in the Baltic (and in the North Sea)? The completion of the databases, using EUROBIS as an additional data source, can make these decisions less dependent on the vagaries of sampling at one or few occasions.

We do not think that EUROBIS can help in judging the positive records, i.e. the recorded presence in the harbours. There are a few cases where presence in a harbour is recorded while the species is not yet found in the EUROBIS database for W and N Europe. Considering these species are invasive and that collection of data in EUROBIS lags behind the actual sampling, sometimes by several years, this can be the consequence of the fact that the species has only recently arrived in the region of interest. In some cases, the species were riverine rather than marine, which can also explain their absence from EUROBIS. Both reasons decrease the effectiveness of controlling the positives using the database. However, control of the negatives can be a useful contribution from EMODNET to this decision support tool.

Further developments

The graphical quality of the present product is very low. This document is only meant as a technical contribution to a more extensive decision support tool, therefore not much attention has been paid to the graphical aspect. An obvious further development is to produce the maps in an interactive web-based application. That should not require a major investment and will make the product easier to use and to integrate into existing infrastructure.

When making the product interactive, it will also be relatively easy to include a time dimension to the EMODNET data. A time slicer can be added to an interactive application, allowing to view occurrences in the database for any time periods chosen by the user. In this way, it will be possible to investigate whether the positive findings are recent (and therefore highly relevant) or accidental observations that have never been repeated. For well-established species, it will be possible to investigate whether they have still been expanding in recent years or not. The EMODNET data can also be made available for download in tabular form. This will allow users to make their own calculations on the basis of the available data. However, the actual use and incorporation into the decision support tool is not a task for EMODNET, and should be left to the developers and users of the tool.