Distribution of biotopes, habitats and biological values at Holmöarna and in the Kvarken Archipelago

AquaBiota Report 2013:06

Sofia A. Wikström, Carolina Enhus, Frida Fyhr,
Johan Näslund and Göran Sundblad

AquaBiota
WATER RESEARCH

STOCKHOLM, APRIL 2013

Client:
Performed by av AquaBiota Water Research for Forststyrelsen, Finland within the SUPERB project.

Authors:
Sofia A. Wikström (sofia.wikstrom@aquabiota.se), Carolina Enhus, Frida Fyhr, Johan Näslund, Göran Sundblad

Front picture:
Sofia A. Wikström, AquaBiota

Contact information:
AquaBiota Water Research AB
Adress: Löjtnantsgatan 25, SE-115 50 Stockholm, Sweden
Tel: +46 8 522 302 40
www.aquabiota.se

Quality control/assurance:
Antonia Nyström Sandman (antonia.sandman@aquabiota.se)

Internet version:
Downloadable at www.aquabiota.se

Cite as:
Wikström, S.A., Enhus, C., Fyhr, F., Näslund, J. & Sundblad, G. 2013. Distribution of biotopes, habitats and biological values at Holmöarna and in the Kvarken Archipelago. AquaBiota Report 2013:06. 52 pp.

Keywords:
Biological valuation, species distribution modelling, HELCOM underwater biotopes, Reefs, Sublittoral sandbanks

AquaBiota Report 2013:06

ISBN: 978-91-85975-26-6

ISSN: 1654-7225

CONTENTS

SUMMARY

We use bathymetric data from lidar and other sources together with field data and maps of physical variables to create maps of biotopes compatible with the HELCOM Underwater Biotopes/Habitats Classification System, Natura 2000 habitats and biological values over two study areas; Holmöarna in Sweden and the Kvarken Archipelago Natura 2000 area in Finland. We were able to model the distribution of a number of biotopes, while other was not possible to model due to a limited amount of data, an unsuitable sampling design, lack of substrate maps or other reasons. Topographic elevations that can be classified to either Sublittoral sandbanks or reefs were mapped using two different approaches, where one has been used in Sweden and the other was regarded to be more appropriate when following the definitions for these habitats in Finland. We present a method to compile different data types into maps of biological values, by primarily assigning values to biotopes and habitats and mapping those for entire regions. The results shows that the availability and quality of input data, both environmental data (e.g. maps of bathymetry and seabed substrate) and biological data from field sampling, determine how well we are able to map biological values and other factors that should be considered for marine planning and conservation.

1. INTRODUCTION

Planning, management and development of marine and coastal areas require extensive and reliable data describing the marine system, including its functions and values. In order to be included efficiently in the planning process, the data have to be collated in a way that makes them easy to use for planning authorities that may lack expertise in marine biology and geology. One of the main outcomes of the ULTRA project was that managers of marine and coastal areas in Sweden and Finland request maps showing areas of large biological value and areas of particular interest for conservation, rather than the distribution of single species or habitats.

Mapping of areas of high biological value ("högt naturvärde") are required for marine spatial planning and coastal zone management, for handling of exploitation permits and for marine conservation. There is however no standardized method for biological valuation of the marine environment. The Convention on Biological Diversity (CBD) has adopted a list of scientific criteria for identifying ecologically or biologically significant marine areas in need of protection (CBD 2008; 2009) which is regarded to be a key instrument for marine biological valuation. However, this document offers little guidance in how to practically derive a valuation according to these criteria. In the Baltic Sea a number of attempts have been done to map biological values for planning purposes, for instance at Swedish offshore banks (Naturvårdsverket 2010), in the Ålands Sea in NANNUT (www.nannut.fi), in the Swedish counties of Östergötland (Carlström et al. 2010) and Västernorrland (Florén et al. 2012) and in Swedish municipalities (e.g. Gustafsson et al. 2008). However, the methods used in these studies require further development to derive a functional and broadly accepted method for mapping of biological values.

A related and partly overlapping process is the assessment of threat status of Baltic Sea biotopes that is done within the HELCOM Red list project. During 2013 the threat status of all Baltic Sea biotopes will be assessed, which will point out endangered and vulnerable biotopes that need conservation efforts. The first step in the HELCOM Red list project has been to develop a broadly accepted biotope classification system for the Baltic Sea, HELCOM Underwater Biotopes/Habitats Classification System (Haldin et al. 2012). The biotope classification was developed by a group of experts from most countries boarding the Baltic Sea, based on field data covering a large part of the Baltic. It is an adaptation of the European habitat classification system EUNIS to the special environment of the Baltic Sea and shares with the EUNIS system a hierarchical structure where the upper levels describe the physical habitat (e.g. depth zone and seabed substratum) and the lower levels describe the biotope based on dominating vegetation and/or fauna.

Another important legal instrument for marine conservation is the EU Habitats Directive, listing a number of underwater Natura 2000 habitats that should be protected from anthropogenic impact. Mapping of these habitats has been a prioritized activity in both Sweden and Finland, but there are still many gaps in our knowledge about their distribution. Improved bathymetric mapping, for instance using lidar, opens new possibilities for large-scale mapping of certain habitats, particularly Sublittoral sandbanks and Reefs.

The aim of this study was to evaluate how data from lidar and other methods can be collated to maps of areas with high biological values and underwater Natura 2000 habitats, which can be used directly by planners and decision makers in Sweden and Finland. A second aim is to evaluate which type and quality of data that is needed to provide maps with high enough confidence to be used for planning and management.

The work was funded by the SUPERB project (Standardiserad Utveckling av Planering och Ekologiska Redskap för Bottniska viken).

2. METHODS

The project has three different parts, which are described in detail below. Firstly, we have modelled the distribution of a number of phytobenthic biotopes, compatible with the HELCOM Underwater Biotopes/Habitats Classification System (Section 2.2). Secondly, we have mapped the distribution of the Natura 2000 habitats 1110 Sublittoral sandbanks and 1170 Reefs (Section 0). Finally, we have compiled maps of biological values, based on the biotope maps and other data layers (Section 2.3).

2.1. Study areas

We worked in two different study areas within Norra Kvarken; Holmöarna in Sweden and the Kvarken Archipelago Natura 2000 area in Finland. The two areas are of different size and are chosen to illustrate the possibilities and difficulties that are encountered when working on different spatial scales.

The Holmöarna study area represents a local scale (a single protected area). It is a small area (500 km^2) around the islands of Holmöarna, a group of large islands off the Swedish coast. The largest part of the area constitutes the Holmöarna nature reserve and Natura 2000 area, but the study area includes also the parts of Holmöarna that are excluded from the protected area.

The Kvarken Archipelago area represents a regional scale, with long gradients in environmental conditions. The study area consists of a set of smaller and larger Natura 2000 areas that are spread throughout the Finnish side of Norra Kvarken, including both the south-north gradient in salinity and ice conditions from the Bothnian Sea to the Bothnian Bay and the gradient in wave exposure and Secchi depth from the coast to the open sea.

2.2. Predictive modelling of biotopes

Field data

We used already available field data, collected with different methods, as input in the predictive modelling of biotopes. Since the data was not collected specifically for modelling purpose, we first went through all data and selected a data set that was appropriate for modelling. Sampling design, i.e. how the inventoried sites are chosen, is an important aspect of predictive distribution modelling as it relates to the assumptions behind the modelling technique. Similar to most statistical studies, replicates need to be independent and in the context of predictive spatial models the environmental conditions in the area need to be well represented.

For **Holmöarna**, we had access to a number of datasets collected with different methods. Most surveys were directed towards phytobenthic habitats. This included 173 underwater video stations from a large survey of the Västerbotten County in 2011. The survey was designed to collect data for modelling of phytobenthic biotopes and the stations were placed randomly within depth strata to ensure a proper representation of all depths with benthic flora. Phytobenthic communities had further been investigated with diving transects between 2007 and 2011 (using the Swedish national standard method), snorkelling transects in shallow bays (from the base inventory of Natura 2000 habitats in 2007), underwater video transects (from the base inventory of Natura 2000 habitats in 2007) and with snorkelling in connection to surveys of juvenile fish in 2005. Beside the phytobentic data, the study area included nine grab sample stations for infauna from yearly monitoring. However, this was too little data for predictive modelling of infauna biotopes.

The stations consisting of video or dive transects were divided into sub-sections, of which 2 random sub-samples were taken. A minimum distance of 200 m between sub-samples was used in order to reduce the amount of spatial autocorrelation to acceptable levels. Furthermore, this gave sub-samples with a sample size (i.e. area coverage) similar to that used within the other surveys. The compilation of data from different sources resulted in a dataset with 310 stations where 10-25 m2 of the sea bed had been

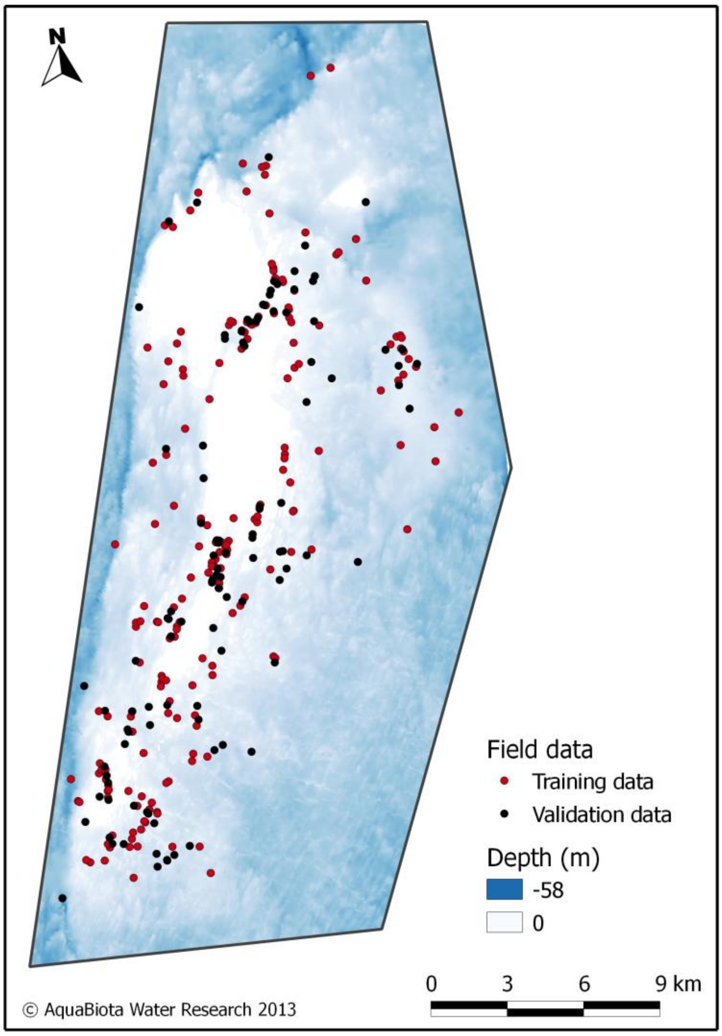

Figure 1. Overview map of field data used for modelling and validation and extent of the area that was modelled in Holmöarna.

surveyed (Figure 1). This dataset was further divided into a training dataset of 65 % and a test (i.e. validation) dataset of 35 % of the stations. The division into training and test data was done separately for each modelled biotope and was weighted by prevalence (proportion of presence samples) in order to reduce stochastic effects from the sub-sampling as the total number of samples was relatively low.

The data available for the **Kvarken Archipelago** consisted of 7251 underwater video stations, collected between 2006 and 2011 within the VELMU and FINMARINET programs. Most of the video stations were collected using a regular grid survey design, while about 5 % of the stations were randomly allocated within the study area (Figure 2). The stations from the regular grid design were not spread across the entire study area and most of the data were collected within two small areas (Norrskär and the Rönnskär islands). This design is not ideal from a spatial modelling perspective, for two reasons. Firstly, in statistical modelling most methods assume that points are independent from one another. Violations of this assumption may cause biased and unreliable models, i.e. the observed relationships are not accurate. In this study, semivariogram analysis showed that the data was spatially autocorrelated at a distance between between 1.5 and 5 km. Spatial autocorrelation means that near-lying points are more similar to each other than points further away, violating the assumption that the points are independent. The gridded design, with a dominant part of the data occurring in a grid with 100 to 1000 m between the stations, thus increases the risk for spatial autocorrelation affecting the model results.

Secondly, the majority of the stations were situated within a restricted part of the study area, predominantly in areas far away from the coast and in areas with relatively high exposure to wind and waves. Such imbalance in data collection will inevitably bias the resulting models and result in poor predictions in other parts of the study area.

In order to reduce the problems of the imperfect sampling design, we only used the randomly allocated stations, together with a random subset from the gridded data of similar size as the random data, to calibrate the models. By utilizing similar amounts of gridded and random data, i.e. giving them equal weight, we aimed to ensure that the species-environment relationships were not determined only by patterns observed in the gridded areas. This resulted in a data set including 774 stations (394 from random sampling and 380 randomly collected from the gridded data) for model calibration. One outlier with respect to depth was removed (depth 30 metres), giving a total of 773 stations for calibration of the models. For two of the biotopes (Submerged rooted plants and Pondweed, Table 1) one additional station was removed (presence locality at 11 metres depth), leaving 772 stations for those two models.

The remaining gridded data (6477 stations) was used as validation data to test the predictive ability of the models. It is important to note that since the gridded data was not spread across the study area, only part of the predicted area has been validated.

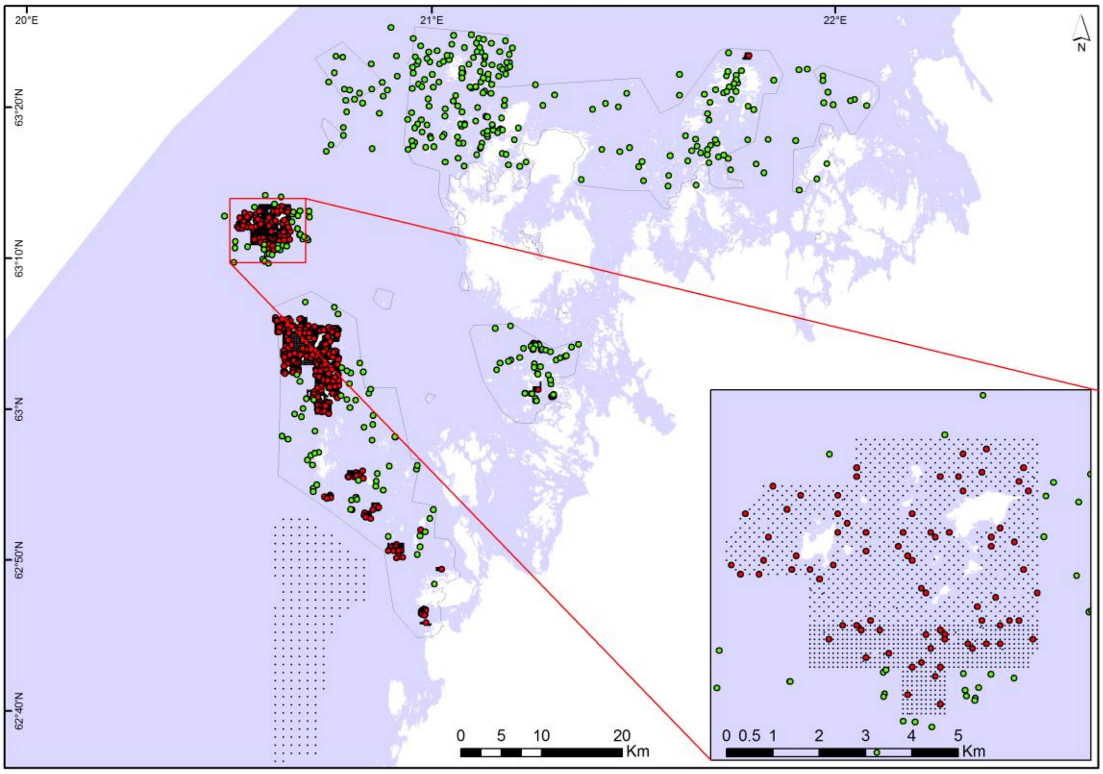

Figure 2. Data available for the modelling of the Kvarken Archipelago. Small black dots are data collected in a structured grid design and green points have been randomly allocated. Red points are a random selection from the gridded data, consisting of as many samples as there were randomly allocated samples. The green and red points together constituted the calibration data for the distribution models. The remaining small black dots were used as test data, except for the gridded data in the southwest which was completely excluded as it was outside the main area of interest. Grey lines show the extent of the Natura 2000 areas in which predictions were made. The two major concentrations of gridded data are areas where lidar measurements were available, one of which (Norrskär) is exemplified in the inset.

Modelled biotopes

From the field data, we extracted data on HELCOM Underwater biotopes (levels 5-6; see Box 1). Since the biotope classification was still under development when the modelling of biotopes was done, we used preliminary versions of the biotope classification system. For simplicity, we did not consider the upper levels in the modelling. This means that biotopes characterised by the same species but occurring on different substrates, for instance submerged rooted plants on muddy sediment, coarse sediment and sand, are modelled as one biotope. The modelled biotopes and their definitions are given in Table 1.

Box 1. The Baltic Sea EUNIS classification of habitats and biotopes

This system is developed within the HELCOM Red list project, to be used for the threat assessment of biotopes in the Baltic Sea. It is a hierarchical classification system that follows the European habitat classification EUNIS. At the highest hierarchical level are broad habitats defined by environmental factors such as depth and seabed substrate. These habitats are then subdivided into biotopes defined by dominating species or groups of organisms.

An excerpt from BS EUNIS is shown below. At level 1, the Baltic Sea habitats and biotopes are separated from marine habitats. Level 2 splits the habitats into photic and aphotic habitats (in or below the photic zone). Level 3 separates between seabed substrate types and level 4 shows community type (characterised or not by vegetation or epibenthic fauna). Level 5 gives the biotopes characterized by different organism groups (e.g. annual algae, perennial algae or submerged rooted plants) and level 6 is a finer division of these biotopes depending on dominance of certain species or groups of species.

- A Baltic
  - AA Baltic Photic
    - AA.J Baltic Photic Muddy sediment
      - AA.J1 Baltic Photic Muddy sediment characterized by vegetation
        - AA.J12 Baltic Photic Muddy sediment characterized by submerged rooted plants
          - AA.J124 Baltic photic Muddy sediment dominated by Charales

When the **Holmöarna** area was modelled, the classification system was weakly developed at level 6 (defined by dominating species), why we focussed on modelling level 5 biotopes (defined by broad taxonomic groups). We modelled the four level 5 biotopes that we considered to be consistently inventoried in the assembled data: biotopes characterized by (1) Perennial algae and moss, (2) Annual algae, (3) Epibenthic fauna and (4) Submerged rooted plants. In addition, we modelled the level 6 biotope dominated by Charales and biotopes characterized by tall vascular plants (corresponding to biotopes dominated by Pond weed and Watermilfoil in the final biotope classification).

Furthermore, the associated substrate information in the various original data sets was highly variable and in some cases missing, why it was not included in the modelling. This is mainly an issue regarding the biotopes that occur on both sand and mud (i.e. biotopes dominated by Charales and biotopes characterized by tall vascular plants) as the other modelled biotopes only occur on hard substrates (rock, boulders and stone) in this area.

Table 1. Biotopes in Holmöarna and Kvarken Archipelago that were modelled in this study. The biotopes comes from the preliminary HELCOM Underwater Biotopes/Habitats Classification System available at the time of the modelling work (see the text for further explanation), except for Filamentous algae and High vegetation in the Kvarken Archipelago. The definition shows which taxa from the inventory data that were included in the respective biotopes.

Biotope	**Definition**
Holmöarna	
Perennial algae and moss	≥ 10 % cover of perennial algae and moss (*Aegagrophila linnaei*, *Battersia arctica*, *Fucus* sp., *Fontinalis* sp.)
Annual algae	≥ 10 % cover of annual algae (*Ceramium tenuicorne*, *Cladophora fracta*, *Cladophora glomerata*, *Ectocarpus siliculosus*, *Pylaiella littoralis*, *Rivularia*, *Ulva* sp.)
Epibenthic fauna	≥ 10 % cover of epibenthic fauna (*Balanus improvisus*, *Electra crustulenta*, hydroids and sponges)
Submerged rooted plants	≥ 10 % cover of submerged rooted plants (*Chara* spp., *Nitella* sp., *Tolypella nidifica*, *Callitriche hermafroditica*, *Isoëtes lacustris*, *Myriophyllum* spp., *Potamogeton* spp., *Ranunculus* spp., *Stuckenia* spp., *Subularia aquatica*, *Zannichellia palustris*)
Charales	≥ 10 % cover of Charales (*Chara* spp., *Nitella* sp., *Tolypella*)
Tall vascular plants	≥ 10 % cover of *Potamogeton perfoliatus*, *Stuckenia pectinata* and *Myriophyllum* spp.
Kvarken Archipelago	
Fucus sp.	Presence of *Fucus* sp.
Perennial non-filamentous corticated red algae	Presence of *Furcellaria lumbricalis*
Filamentous algae	Presence of filamentous algae (*Battersia* sp., *Ceramium* sp., *Cladophora* spp., *Dictyosiphon* sp., *Ectocarpus siliculosus*, *Pylaiella littoralis*, *Polysiphonia* sp., *Rhodochorton purpureum*, *Scytosiphon lomentaria*, *Stictyosiphon tortilis*, *Ulva* sp., *Vaucheria* sp.)
Submerged rooted plants	Presence of submerged rooted plants (*Chara* spp., *Nitella* sp., *Tolypella nidifica*, *Callitriche hermafroditica*, *Myriophyllum* sp., *Potamogeton* spp., *Ranunculus baudotii*, *Ruppia* sp., *Stuckenia* spp., *Zannichellia* sp.)
Pond weed	Presence of *Potamogeton perfoliatus* and/or *Stuckenia pectinata*
Zannichellia spp. and/or Ruppia spp.	Presence of *Zannichellia* sp. and/or *Ruppia* sp.
Watermilfoil	Presence of *Myriophyllum* sp.
Charales	Presence of charophytes (*Chara* spp., *Nitella* sp., *Tolypella nidifica*)
High vegetation	Presence of ≥ 50 cm high vegetation

When modelling the **Kvarken Archipelago** we had access to a preliminary list of level 6 biotopes, so in this area we attempted to model level 6 biotopes as far as possible. This preliminary list turned out to be similar to the final classification, but a few of the final biotopes were missing. The level 6 biotopes that were considered to be consistently inventoried in the field data were biotopes dominated by (1) *Fucus* sp., (2) perennial non-filamentous corticated red algae, (3) Pond weed, (4) *Zannichellia* spp. and/or *Ruppia* spp., (5) Watermilfoil and (6) Charales. In addition, we modelled the level 5 biotope Submerged rooted plants, in order to compare the results from modelling of different levels of detail in the biotope classification system. Since vegetations heights are registered in the Finnish survey method, we also modelled the occurrence of high vegetation (>50 cm length), since high vegetation is known to be important for many animals, including juvenile fish.

In most of the Finnish data set the filamentous algae were not identified to species and attached algae were not separated from drifting algae. This reflects that species identification can be difficult using UV video and that it is sometimes difficult to be certain whether the algae are attached. This meant that we could not characterise the filamentous algal communities based on species composition. Instead, we modelled the filamentous algae as one group. The predicted distribution of this group includes the biotopes dominated by Perennial filamentous algae (level 6), Annual filamentous algae (level 5) and *Vaucheria* sp (level 6).

The epifaunal communities were not inventoried consistently in the Finnish data. For instance, mytilids, hydroids, moss animals and sponges were only recorded in the investigations in 2011. It might be possible to model the distribution of these biotopes using a smaller, specifically selected dataset, at least in well investigated parts of the area, but we did not have time to investigate this within the current project.

In Holmöarna, we classified a biotope as present in a field station if it had at least 10 % cover, which is in line with the HELCOM definition of the level 5 biotopes (the cover of the characterizing species should be at least 10 %). In the Kvarken Archipelago, a biotope was instead classified as present as soon as one or more of the characterising species were present. This was done for two reasons, firstly because cover data was missing from a number of stations and secondly because only modelling high cover occurrences would have decreased the prevalence (i.e. amount of presences in relation to total sample size) of a number of biotopes below an acceptable level.

Environmental predictor variables

A prerequisite for successful modelling and prediction of biotopes is the use of ecologically relevant environmental variables, and that the variables are available spatially, i.e. in a GIS. The perhaps most important criterion is that the sampling design does not introduce any bias. This is done by ensuring that the sampling covers all possible combinations of the predictor variables and is normally achieved by randomly allocating survey stations (this also reduces the risk of spatial autocorrelation).

The environmental predictors used in this study are shown in Table 2. The choice of variables used for each study area was based on available data layers and whether they were judged to be potential predictors for biotope distributions in the specific area.

In **Holmöarna**, the depth layer for the predictions combines bathymetric data from the lidar survey of the southern part of the study area with data from the Swedish Maritime Administration (Sjöfartsverket) bathymetric database. First, a full cover depth layer was derived by natural neighbour interpolation with inverse distance weighting, using 2871083 depth measurement points from the Swedish Maritime Administration. For the area covered by lidar, another depth layer was produced by taking mean depth for each raster cell and expanding over gaps using the close gaps module in SAGA (the module fill gaps by interpolating from the nearest cells). Finally the two depth layers were merged, using the more detailed layer derived from lidar where it was available. The models were calibrated using field measured depth, corrected for mean water level, and the full cover depth layer was used for the predictions. Based on the depth layer, we calculated sea bed slope and an index of topographic position which show if a point on the map rise above the surrounding sea floor, represents a depression or is situated in a flat area.

Table 2. Environmental predictor layers used for modeling of Holmöarna and Kvarken Archipelago. When applicable, transformation of the predictor variables in the models is given in brackets. In the Kvarken Archipelago, we compared the results from using two different depth layers in a part of the study area.

	Holmöarna	**Kvarken Archipelago**
Depth	Interpolated from batyhymetric data from the Swedish Maritime administration and the lidar survey performed in SUPERB. Spatial resolution 10 m. [Square-root transformation]	(1) Interpolated map from FINMARINET. Spatial resolution 20 m. (2) Data from lidar surveys in SUPERB and ULTRA (covering part of the study area). Spatial resolution 10 m.
Slope	Calculated from the a DEM with 100 x 100 m grid size. Spatial resolution 10 m. [Square-root transformation]	-
Topographic position	Calculated from the a DEM with 100 x 100 m grid size with 250 m radius. Spatial resolution 10 m.	-
Wave exposure	SWM (Simplified Wave Model). Spatial resolution 25 m. [Log transformation]	SWM (Simplified Wave Model). Spatial resolution 25 m. [Log transformation]
Salinity	-	From FINMARINET: Interpolated from salinity measurments from OIVA database (summer values 1999-2008). Spatial resolution 100 m.
Secchi depth	-	From FINMARINET: Interpolated from Secchi depth measurments from OIVA database (summer values 1999-2008). Spatial resolution 10 m.

SWM (Simplified Wave Model; Isaeus 2004) was used as measure of wave exposure. The model calculates exposure to waves from the fetch in 16 directions and dominant wind conditions and has been shown to perform well as predictor of benthic communities in the Baltic Sea. Data layers for temperature, salinity, Secchi depth and ice conditions were not available at fine enough spatial resolution to be meaningful predictors at this local scale.

In the **Kvarken Archipelago**, the predictions were done using two different bathymetric data layers. Firstly, we used an interpolated depth layer from FINMARINET. The prediction from this data layer was then compared with predictions using depth layers from the lidar measurements in parts of the area (Rönnskären and Norrskär), keeping the models and all other predictor layers the same. As in Holmöarna, the models were calibrated using field measured depth corrected for mean water level and the two bathymetric layers were used for the predictions. Since we worked with a coarser bathymetry in the full study area, we did not calculate slope or topographic position. This could have been done in the areas covered by lidar, which would probably have improved the predicted maps for this area, but we prioritised to keep the same statistical model for predictions with both depth layers in order to specifically look for improvements due to a finer scale bathymetry.

Wave exposure was calculated using SWM (described above). Interpolated summer values of salinity and Secchi depth at 100 metres cell size resolution from FINMARINET were selected over the alternatives from the Finnish Meteorological Institute, partly due to difficulties with the projection but also to the apparent resolution (i.e. cell size) and previously successful use of data extracted from the OIVA database (Bergström et al. 2013, Sundblad et al. 2013). Initially, we tested to include ice conditions, measured as the number of weeks with an ice cover over 30% for the years 2009 and 2010, both independently per year and as a mean for these two years on response data from corresponding years. The effect of ice cover was expected to interact with primarily depth, and possibly wave exposure. Preliminary models showed that this was a correct assumption. However, the strength of the interaction depended on year (i.e. the amount of ice) and the response variable. Although the ice data showed strong potential as an important predictor in distribution models for this area it was removed from further analyses. The reasons were that the biological data covered a larger time period than the ice data, and that the interactive effects would require too much time to disentangle.

Unfortunately, no spatial information on substrate was available in any of the study areas. The resulting maps should therefore be interpreted with this in mind.

An important aspect to consider before attempting any type of distributional modelling is the correlation among the variables used to explain the patterns of interest. If predictor variables are correlated, i.e. when one is large so is the other and vice versa, the assumptions behind the model is violated and the conclusions drawn will be biased and not reliable. As in simple linear regression the precision of the estimates will be biased and the importance of the predictor variables will be inflated potentially leading to the wrong conclusion that a predictor is important for the distribution of the response, when in fact it is not. There are several ways to investigate if these problems exist, the simplest of which is by plotting the pairwise relationships among all predictor variables. Such a plot revealed two things that need extra consideration in the Kvarken area (Figure 3). First, the relationship between depth and wave exposure indicated that there was a lack of samples from deep and sheltered areas, but that all other combinations had been sufficiently sampled. We do not consider this as a problem since the GIS revealed that these types of environments are extremely rare in the study area (i.e. we do not risk extrapolating the model since that combination is in fact not available in the prediction area).

Secondly, and of higher importance, was the relationship between salinity and Secchi depth. There was a strong effect of insufficient sampling of many combinations between the two variables salinity and Secchi depth. For instance, areas characterised by low salinity and low visibility consisted of just a few stations. In general, the majority of

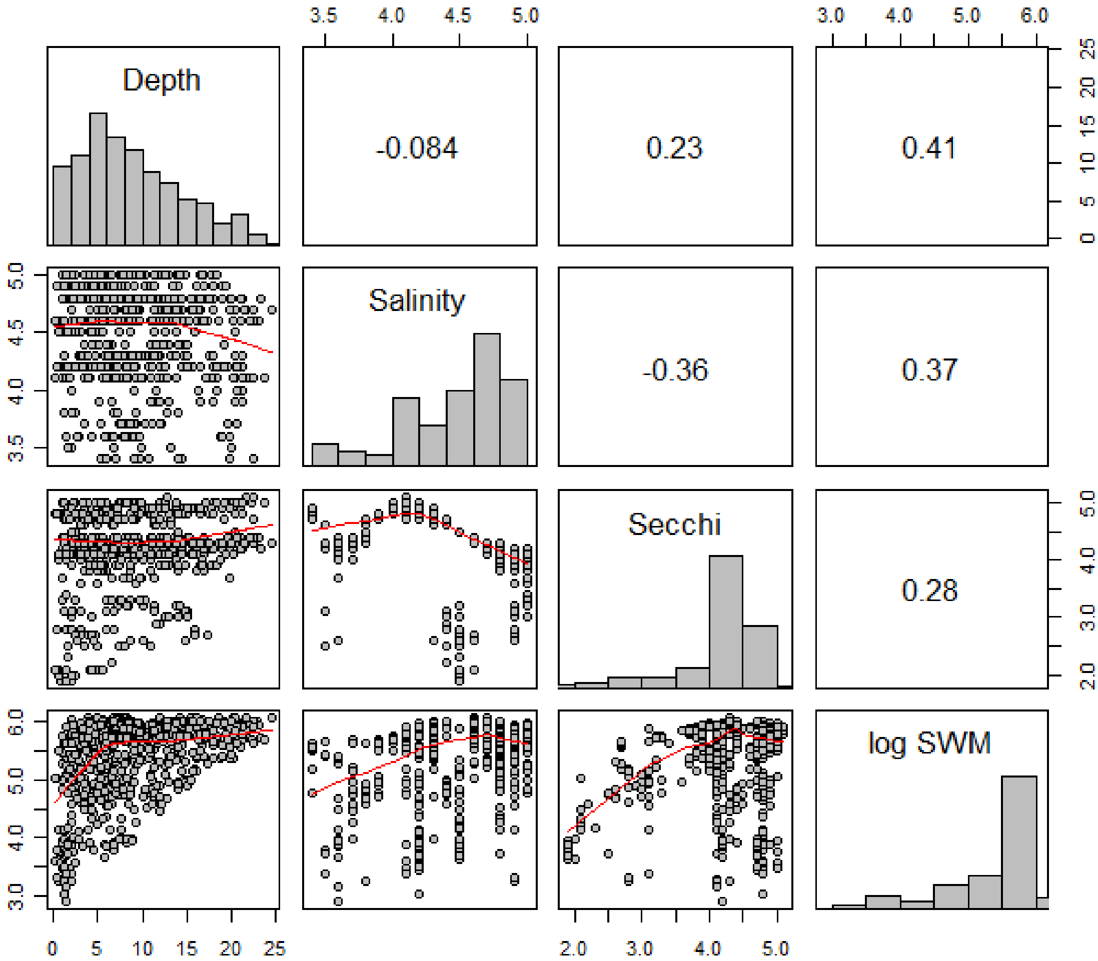

Figure 3. Correlations among predictor variables in the Kvarken Archipelago. Histograms show the distribution of the predictor variables. Upper panels show Pearson correlation coefficients and plotted in the lower panels are the pairwise correlations among predictor variables and their relationship based on a smooth line.

the stations were situated in the upper range of water visibility (Secchi grid values > 4 m), and followed a gradient in salinity from southwest to northeast, with the addition of a concentrated subset of samples around Yttre Torgrund west of Vasa with intermediate salinity and relatively poor visibility. These patterns were probably an effect of sampling being concentrated to the larger Natura 2000 areas. The consequence is that the models will be calibrated (fitted) according to the relationships observed in these areas, and when the models are projected into other areas the model will provide probability values in combinations of salinity and Secchi depth that were not included in the calibration stage. This is a form of extrapolation, i.e. extending the model outside its range, even though we are still within the calibration range of individual predictor variables. Such extrapolations may be of less concern if there is independent data to assess the credibility of the predicted map at all possible combinations of environmental predictors. Since there is a lack of such data in this project, interpretation of the maps should be made with extra care (as will be stressed throughout this report).

Statistical modelling of biotope distribution

Generalized additive models (GAM) were used to relate the distribution of the biotopes to the environmental predictor variables by using the statistical software R and the library 'mgcv' (Wood 2006). GAM is a type of statistical technique that is highly flexible as it follows the patterns observed in the data using smooth lines (functions). The high flexibility allows fitting non-linear responses, but similar to many other techniques it is also sensitive to the quality of the data that is used to calibrate the models.

We initially fitted a full model, i.e. using all environmental predictors, and model selection then consisted of trying to reduce the full model by removing the least important predictor until no further improvement could be made. Included also was an automated procedure that involves a penalizing function trying to limit each relationship to a horizontal line (i.e. reducing the complexity and potentially removing the variable completely; Wood & Augustin 2002). For all cases a binomial response has been used, i.e. the presence or absence of the response.

For the Kvarken Archipelago, the final model for each response was used in a cross-validation procedure to obtain an estimate of the stability of the relationships, i.e. to assess whether only a subset of the data was responsible for the observed patterns or if the selected model was supported by all parts of the data. Depending on the amount of presence data available, ten or five folds were used in the cross validation procedure. Cross validation means that the dataset is partitioned in a number of folds (here five or ten), where each fold has the same prevalence as the full dataset. The model is then calibrated using all but one fold, which is instead used as a test dataset. This procedure is iteratively done so that all folds are tested once, and included in model calibration for the remainder of the runs. The advantage is that all data is used in model calibration and validation, and that the uncertainty of model performance can be estimated. The cross validation procedure was not appropriate to use in Holmöarna because of the low number of stations.

Model performance was evaluated based on deviance explained (how well the model fits the data) and the area under curve value (AUC) of receiver operating characteristic plots (Fielding and Bell, 1997). AUC values range between 0.5 and 1 and is independent of any cut off for the probabilities. AUC is a measure of the discriminatory ability of the model, i.e. how well presences are separated from the absences. Models that performs no better than guessing has an AUC value of 0.5, while a model that perfectly discriminates between presence and absence has a value of 1. Values above 0.8 can be considered excellent (Hosmer and Lemeshow, 2000). In other words, a value of 0.8 means that a randomly selected presence will have a higher predicted probability than a randomly selected absence in 80 % of the cases. Variable importance (how important each predictor is in relation to the other included predictors) was measured as the associated chi2 value from the model.

Prediction to maps

The final models were used together with the predictor layers to predict the distribution of the modelled biotopes across the study areas. The predictions were made with a cell size matching the resolution in the depth data (i.e. 10 m cell size in Holmöarna and the lidar areas in the Kvarken Archipelago and 20 m cell size for the entire Kvarken Archipelago). The predicted maps show the 'probability of presence' of the modelled biotopes. All maps only display the probabilities above a cut off value and therefore only include the areas where there is a relatively high probability of finding the biotope. The cut off value was calculated to maximise the sensitivity and specificity, i.e. the ability to correctly predict both presences and absences (Jiménez-Valverde and Lobo, 2007). Thus, the maps show probability of occurrence in potentially suitable habitats, as defined by the environmental predictor variables used.

The quality of the output maps was evaluated with the independent validation data, i.e. field stations that were not used to train the models, using the area under curve value (AUC).

Mapping of Sublittoral sandbanks and Reefs

Identification of elevations

We tested two different approaches to map the Natura 2000 habitats *1110 Sublittoral sandbanks* and *1170 Reefs*. The first approach has been used for large-scale mapping of these habitat types in Sweden (Fyhr 2012) and was used for mapping of elevations in the entire Kvarken Archipelago. A vector file containing 1 m depth curves was created based on the interpolated depth layer from FINMARINET with 20 m resolution (Figure 4). The depth curves were then used to identify and select elevations from the sea floor, choosing *the deepest curve* that surrounds an elevation to represent the edge of the potential reef or sandbank (Figure 5, left panel). This means that an elevation could contain a number of smaller elevations. A two meter difference in altitude was set as a minimum of what could be considered an elevation.

All elevations within the Kvarken Archipelago Natura 2000 area were identified. Some elevations were located on the border of the Natura 2000 area, and these were chosen as well (Figure 5, left panel).

The second approach was used in a smaller area (Rönnskär), where lidar measurements were available. From the lidar data we created a finer scale depth grid with 2 m resolution. Gaps in the lidar-data at this fine spatial scale was filled in by creating depth layers of 10 and 20 m resolution from the lidar data and always using the value from the finest available resolution. The grid was converted to a vector file with 1 m depth curves, as in the first approach (Figure 6) and elevations with an approximate width of 10-20 m were selected from the depth curves (Figure 5, right panel). This method was more subjective than the first approach, but allowed more flexibility to chose a certain size of the elevations.

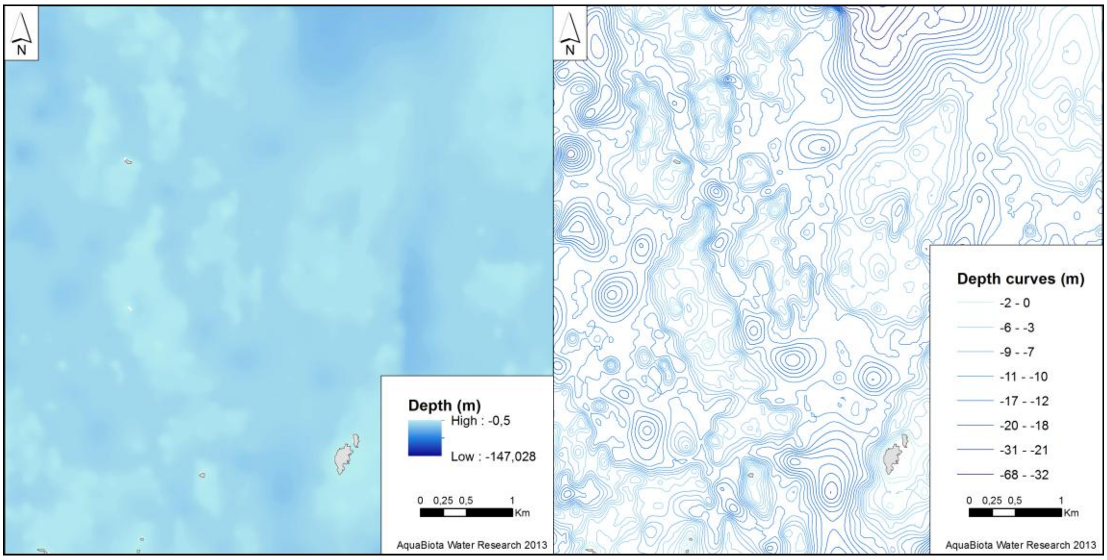

Figure 4. A section of the FINMARINET depth grid from the Kvarken Archipelago (left panel) and the depth curves created from it (right panel).

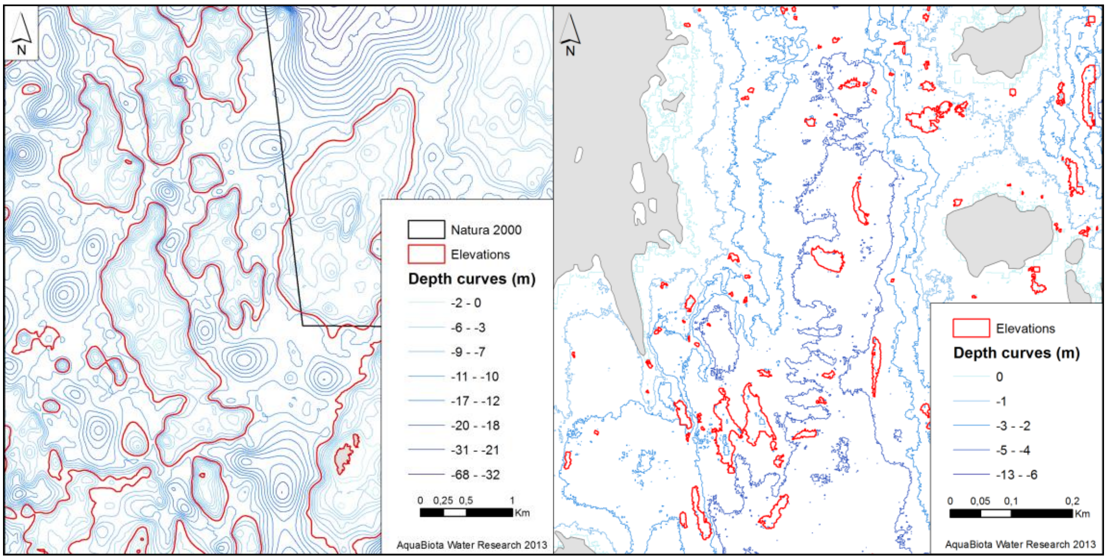

Figure 5. Elevations identified by the two different approaches. The deepest curve was used to represent the edge of the elevations using the FINMARINET depth grid of the entire Kvarken Archipelago (left panel). In the Rönnskär area, elevations with an approximate width of 10-20 m were selected (right panel). Note the difference in scale between the two maps.

Identification of substrate within elevations

In the Habitate Directive reefs are described as: *[...] hard compact substrata on solid and soft bottoms, which arise from the sea floor in the sublittoral and littoral zone. "Hard compact substrata" are: rocks (including soft rock, e.g. chalk), boulders and cobbles (generally >64 mm in diameter).* Sublittoral sandbanks are described as: *consist[ing] mainly of sandy sediments, but larger grain sizes, including boulders and cobbles, or smaller grain sizes including mud may also be present on a sandbank.*

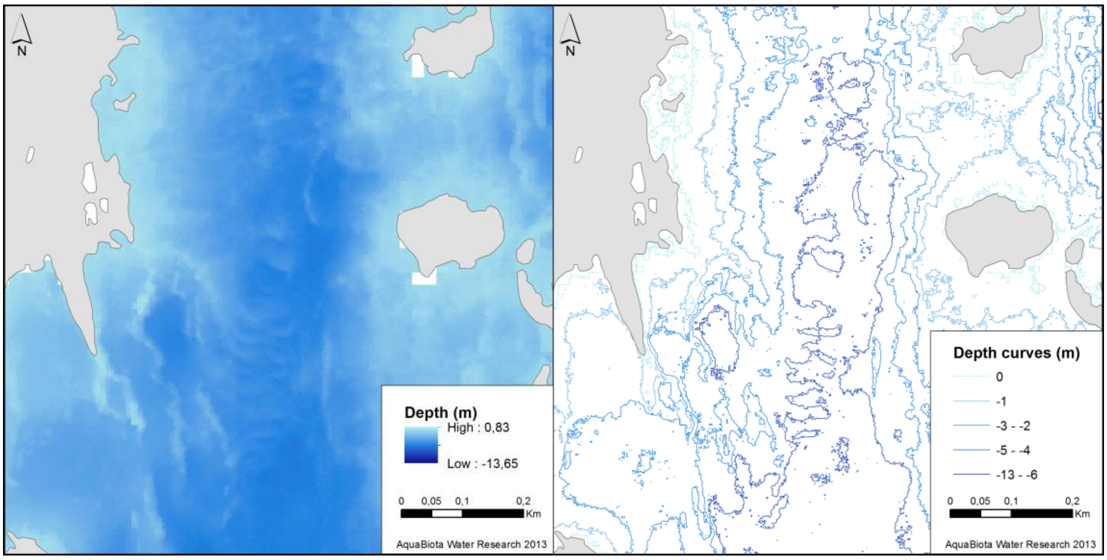

Figure 6. A section of the fine scale depth grid from lidar data from the Rönnskär area in the Kvarken Archipelago (left panel) and the depth curves created from it (right panel).

Table 3. The three substrate groups used for identifying dominating substrate on the elevations; hard substrate (BR_BL_ST), sand and coarse sediment (GR_SA_SI) and fine sediment (CL_SB).

BR_BL_ST	GR_SA_SI	CL_SB
Bedrock	Gravel 6-0.2 cm	Clay
Boulder >300 cm Boulder 300-120 cm	Sand 0,2-0,006 cm Silt 0,006-0,0002 cm	Soft bottom
Boulder 120-60 cm		
Stone 60-10 cm		
Stone 10-6 cm		

To be able to determine the dominating substrate within an elevation, it is necessary to use a map of substrate types in the area. In this case no such data was available. Instead the data from drop video stations in the area, collected between 2006 and 2011 within the VELMU and FINMARINET programs, was used to estimate which substrates that occur on the elevations.

The substrate in the drop video data was divided into three groups based on type and size (Table 3) and the dominating substrate group was determined for each drop video station as the group with a total cover of >50%. If none of the substrate groups reached >50 %, no substrate was considered dominating. The shapefile containing drop video data was then cut to fit the selected elevations.

2.3. Mapping of biological values

A key instrument for marine biological valuation is the list of criteria for identifying ecologically or biologically significant marine areas adopted by the Convention on Biological Diversity (CBD 2008; 2009). It states that the valuation should build on a number of scientific criteria such as uniqueness, biological diversity and importance for threatened species. Mapping of biological values across a larger area or a region pose a special challenge, since we often lack covering maps that can be used to assess these criteria. One approach that has been used for regional mapping of biological values is to use spatial prediction of the distribution of specific habitats or biotopes with high biological value (Carlström et al. 2010; Florén et al. 2012). Here, we develop this approach further by connecting it to the newly developed HELCOM Underwater Biotopes and using the CBD criteria for valuation of the different biotopes. We also include other data to account for values that are not properly described by the benthic biotopes or habitats.

The process to produce maps of biological values included three steps, which are described in detail below. The first step was to identify which criteria are important to describe biological values in the study area and what data can be used to measure these values. Based on this, we listed a number of elements (biotopes, habitats and other elements) for which we could derive spatially covering maps and a list of criteria to determine their value. Secondly, we used available data and specialist judgment to assign a value to each element. Finally, the values from different elements were added to a combined map of biological values.

Criteria and measures of biological value

The identification of criteria and measures of biological diversity was initiated during the SUPERB workshop on biological valuation of the underwater environment in the Gulf of Bothnia, held in Umeå 1-2 December 2011. The workshop assembled specialists and managers from the Gulf of Bothnia regions in both Sweden and Finland to discuss which indicators and measures that should be used to identify high biological values in the region. The identified indicators and methods are presented in Dahlgren et al. (2013).

Based on this workshop, we identified a list of elements for which it was possible to derive spatially covering maps for the study areas. This included the distribution phytobenthic biotopes compatible with HELCOM Classification System and a few specific habitats (roughly corresponding to Natura 2000 habitat types) that were identified as important during the workshop. The distribution of biotopes were modelled in this study (see Section 3.1). Modelled distribution maps of Natura 2000 habitat types *1130 Estuaries*, *1160 Large shallow inlets and bays*, and *1620 Boreal baltic islets and small islands* were taken from the EU INTERREG IIIB project BALANCE (Wennberg et al 2008). Additional maps of *1620 Boreal baltic islets and small islands*, *1150 Coastal lagoons* and *1170 Reefs* were taken from the base inventory of Natura 2000 habitats and protected areas which the Swedish Environmental Protection Agency carried out in 2004-2008 (Naturvårdsverket 2009). The habitats are shown in Figure 7.

Besides, we had some spatial data on distribution and important habitats for plants, fish and seabirds for Holmöarna, which could be used to complement the map of values based on only biotopes and habitats. For plants, we had data on the occurrence of two rare species; the brown alga *Fucus radicans* and the charophyte *Chara tomentosa*. Both species occurred at one single site in the Holmöarna study area (Figure 8. Occurrence of rare plant species and important habitats for fish in the Holmöarna study area. The fish habitats comes from the Digital Environmental Atlas.Figure 8). For fish, we had areas delineated to be important for feeding and reproduction of certain coastal fish species (grayling, perch, pike and whitefish) from "The Digital Environmental Atlas" (www.gis.lst.se/miljoatlas), containing data on areas vulnerable to oil spill (Figure 8). For seabirds, we had access to a valuation of bird islands in a restricted part of the Holmöarna area, based on inventory data.

The workshop identified five criteria important for valuation of habitats and biotopes; (i) uniqueness or rarity, (ii) importance for threatened, endangered or declining species and/or habitats, (iii) biological diversity, (iv) naturalness and (v) ecological function. The last criterion was not taken from CBD but was added because it was regarded as important. We had to exclude the Naturalness criterion since it was not possible to assess the general naturalness for habitats or biotopes – a certain biotope can have a high naturalness in some areas but be strongly impacted in other. Naturalness can be used in a second step in the valuation by combining information on impacted areas with the map of biological values.

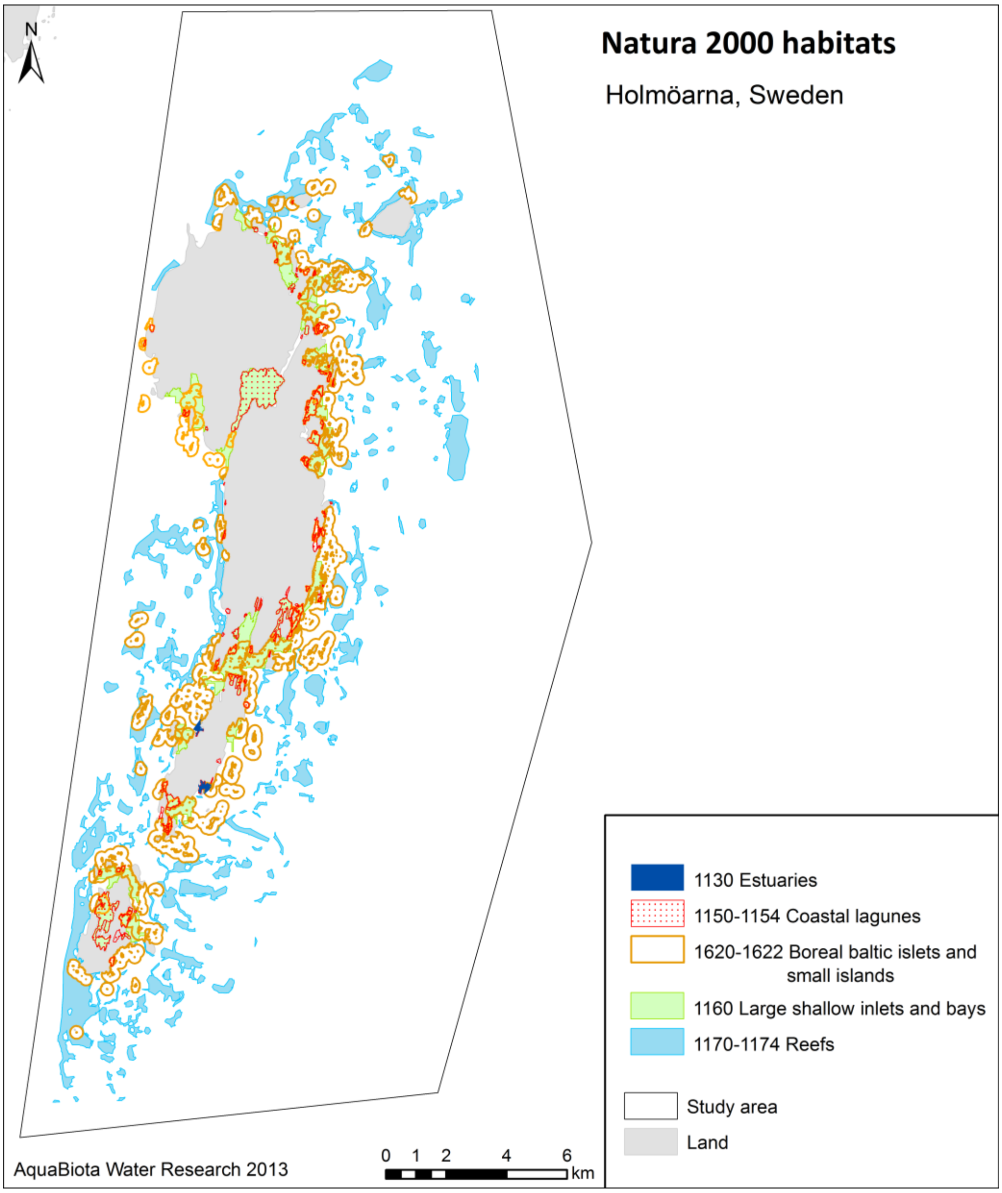

Figure 7. Natura 2000 habitats in the Holmöarna study area, from BALANCE and the base inventory of Natura 2000.

The rare plants are given a value according to the uniqueness criterion and the feeding and reproduction areas are valuated according to the CBD criterion “special importance for life-history stages of species”. The valuation of the bird islands was based on the criteria (i) uniqueness or rarity, (ii) importance for threatened, endangered or declining species and/or habitats, (iii) biological diversity, and (iv) special importance for life-history stages of species.

For the Kvarken Archipelago, only the modelled biotopes were available and the valuation was based only on these.

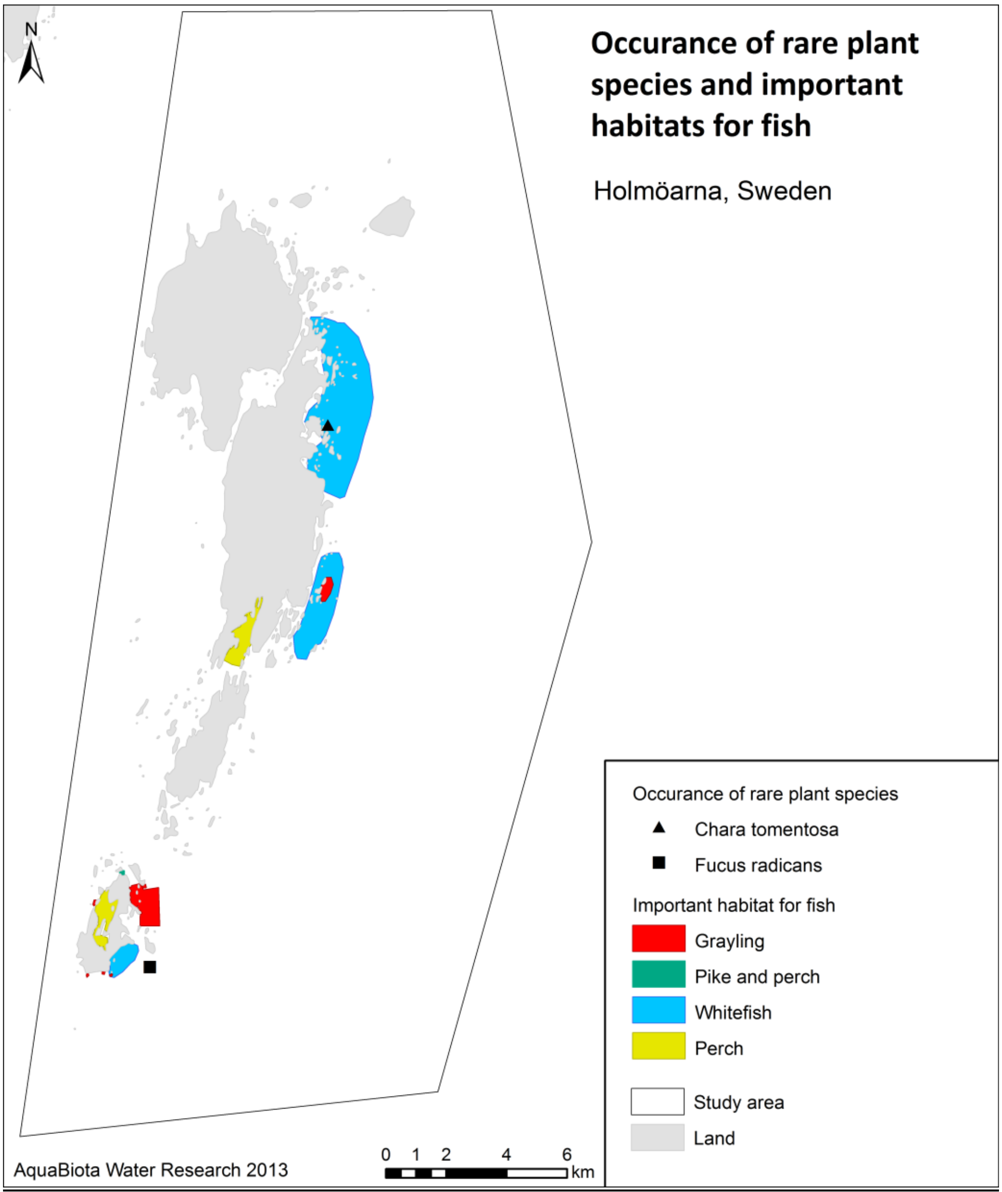

Figure 8. Occurrence of rare plant species and important habitats for fish in the Holmöarna study area. The fish habitats comes from the Digital Environmental Atlas.

Valuation

Each **biotope and habitat** was assigned a value according to the four valuation criteria (uniqueness, threat, diversity and ecological function; Table 4). As far as possible, the values were derived using empirical data. We used a three-graded scale; high biological value (10), low biological value (1) and no value (0). The aim was to set the threshold between high and low value so that high values would be reserved for biotopes and habitats that have a high priority for conservation, but we could not find an objective method to define this threshold across the valuation criteria. This means that the thresholds between high and low value used in this example are very subjective (described for each criterion below) and could be discussed.

*Table 4. Valuation of biotopes compatible with the HELCOM classification system and Natura 2000 habitats. All biotopes that were recorded reliably in the underwater video survey were listed and assessed, although we only had covering maps for a few of them. Level 6 biotopes are listed with an indent below the corresponding level 5 biotope. Biotopes used in the mapping of biological values in either Holmöarna or the Kvarken Archipelago are marked with *.*

	Unique-ness	Threat	Diversity	Function	Total value
Biotopes					
Annual algae*	0	0	0	1	1
Vaucheria sp.	1	0	0	1	2
Perennial algae and moss*	0	0	1	1	2
Fucus sp.*	10	0	1	10	10
Perennial filamentous algae	0	0	1	1	2
Aquatic mosses (Bryophyta)	1	0	0	1	2
Submerged rooted plants*	1	5	1	10	10
Pond weed (*Potamogeton perfoliatus*/*Stuckenia pectinata*)	1	5	1	10	10
Zannichellia spp./*Ruppia* spp.	1	5	1	1	8
Watermilfoil (*Myriophyllum* sp.)	1	5	1	10	10
Charales	1	5	1	10	10
Epibenthic fauna = Sponges (Porifera)	1	0	1	1	3
Habitats					
Coastal lagoons	10	10		10	10
Estuaries	0	1		1	2
Large shallow inlets and bays	0	1		1	2
Reefs	1	0		1	2
Boreal Baltic islets and small islands	0	0		1	1

Both for biotopes and habitats, the **uniqueness** was evaluated on a regional scale (using data/measurements from the Västerbotten County). Biotopes and habitats can also be unique on a larger (global, Baltic Sea) or smaller (local) scale. We chose to take a regional perspective in this analysis, but this could be complemented in the future when data on other spatial scales are available. Ideally, we would have used data from both Sweden and Finland in this analysis, but the sampling design used in the Finnish monitoring programs (most sampling in outer parts of the archipelago) meant that there was a risk for a systematic bias in the frequency estimations.

For biotopes, the value for uniqueness was derived from statistical analyses of a dataset of 870 drop video stations from the entire Västerbotten County. The data were collected with a stratified random sampling design (randomised sampling within defined depth strata), which made them suitable for determination of the relative occurrence frequency of different biotopes in this coastal area. Each station was classified to a HELCOM underwater biotope based on the cover of substrate and of all species recorded at the site. The classification was done both to level 5 (characterised by broad taxonomic groups) and level 6 (characterised by one or a few species). For each biotope, we then calculated the relative occurrence frequency, accounting for that the shallow depth strata were oversampled in relation to their total cover of the marine area in the Västerbotten County. Of the biotopes that are well inventoried by drop video (phytobentic biotopes except for those dominated by emergent reed and sedges and crustose algae), the *Fucus* biotope stand out as very rare and was given the value 10 (Table 4). Among the other level 6 biotopes, the perennial filamentous algae biotope was by far the most frequent, while all the other had relatively low occurrence frequency and were given the value 1. Among the level 5 biotopes, only the rooted plants biotope was given a value based on this criterion.

For habitats, the value for uniqueness was derived by calculating the area covered by these habitats as a fraction of the total marine area of the Västerbotten County. Lagoons was the rarest of these habitats, covering <0.01‰ of the marine area, and was given the value 10 for this criterion. Reefs covered 0.7‰ and was given the value 1. The other habitats covered around 2‰, which is still little in comparison with the entire marine area but means that they are relatively common habitats in the coastal zone.

Values for **threat** for biotopes were taken from the ongoing Red List assessment of Baltic Sea biotopes. The assessment was not finalized when the valuation was done, but we have taken part in the discussions in the HELCOM work group and used a preliminary version of the habitat red list. The HELCOM red list assess threat status on Baltic Sea scale, so we used a regional expert (Johnny Berglund, County Administrative Board of Västerbotten) to complement with the regional perspective. The most threatened biotopes in the region are those confined to sheltered areas with fine sediment substrate. This includes the rooted plant biotopes, but since these also occur in less sheltered areas they were given the value 5 instead of 10.

For habitats, the value for threat was taken from the Swedish assessment of habitats according to the Habitats Directive 2007 (Sohlman 2008). Coastal lagoons are regarded as the most threatened habitat and were given the value 10, while other sheltered habitats (Estuaries and Large shallow inlets and bays) are also threatened and were given the value 1 for this criterion.

Diversity of biotopes was assessed as mean species richness of macrophytes and sessile invertebrates, using the same data set as used for assessment of uniqueness. This gives a coarse measure of species richness, since the taxonomic resolution of video inventory is relatively low. Comparisons of species richness measured with diving and underwater video have however shown a correlation between species the two measures (Gullström et al 2013), indicating that the richness measured with underwater video at least gives the relative differences in richness between stations.

The number of macrophyte and sessile invertebrate species was counted for each video station. In stations with a mix of hard and sediment substrate, we only counted the number of species associated to the dominating biotope (to which the station was classified). If, for instance, the station was classified as a perennial algae biotope we only counted the number of hard substrate species as a measure of species richness. The reason for this was that the video inventory was not directed towards mapping distinct biotopes, meaning that a certain station often includes a mix of biotopes. The HELCOM

Underwater Biotopes/Habitats Classification System do include biotopes that contain a mix of hard and sediment substrate, but these could not be separated in this data set.

Three biotopes stood out with a low number of species (moss, *Vaucheria* and annual algae biotopes, all with a mean around 1). Among the other, the *Fucus* biotope had a mean of 5 species and the rest of the biotopes had means between 2 and 3 species per station. The higher number for *Fucus* was based on only two samples, so we chose to give all biotopes with more than 2 species the value 1 for the diversity criterion.

Diversity could not be assessed for the habitats since we had no comparable data set from all habitats.

Ecological function was identified as an important criterion at the workshop, but it was difficult to find data to give the biotopes or habitats a value for this criterion. In this example, we therefore use expert opinion to assign a value to each biotope and habitat. The expert judgement was based on three functions: food or habitat for other species, primary production and filtration. We regard all biotopes to perform at least one of these functions and all are thus given a value of at least 1. Biotopes dominated by large macrophytes (*Fucus* and the large rooted plants) are given the value 10 since they both have a high primary production and provide an important habitat for many other species, including fish.

The values for the four criteria were then summed to a total value for each biotope or habitat (Table 4). The sum was not allowed to be higher than 10, which means that a high value for one of the criteria will automatically give a biotope/habitat a high value.

Occurrence of the **rare plants** was given the value 10. For **fish**, important habitats for perch, pike and whitefish were given the value 1, while habitats for grayling were given the value 10 since coastal breeding in this species is a unique feature.

The valuation of **bird islands** was done by experts at the County Administrative Board of Västerbotten, using a partly different approach. Each inventoried island was given a separate value based on the inventory results and the chosen valuation criteria (uniqueness, threat, diversity and importance for life-history; data not shown in Table 4). Species regarded as **unique** were given a value of 1 (locally unique), 2 (unique for the Baltic Sea) or 3 (globally unique). The data included one globally unique species (Baltic Sea Herring Gull) and two unique for the Baltic Sea (Velvet Scoter, Black Guillemot). If more than one species were present their values were added together. For **threat**, the presence of each species on the HELCOM red list was given a value of 1. For the **diversity** criterion, presence of > 30% of the approximately 40 sea bird species breeding in the Västerbotten County was given the value 10 and presence of >15% of the species the value 1. Finally, the total abundance of birds recorded at one occasion was used as a measure for the criterion **importance for life-history**. A total number of >200 birds was given the value 10 and a total of >50 bird the value 1. The values for the four criteria were then summed to a total value for each island. As for the biotopes and habitats, the sum was not allowed to be higher than 10.

Mapping

The total biological value from Table 4 was used together with the available map layers to compile maps of biological values for the Holmöarna and Kvarken Archipelago study areas.

For **Holmöarna**, this was done in two steps. First we made four maps of biological values for (i) biotopes, (ii) important habitats, (iii) rare plant species and (iv) fish habitats, respectively. Each map was made by taking the highest biological value for an element occurring in each grid cell of the map. This means that if more than one biotope or habitat occurs in one grid cell (for instance both annual and perennial algae) their values are not summed. In the second step, we added these four maps together to an integrated map of biological values for phytobenthic biotopes and species, habitats and fish. Thus, the value of a habitat was higher if it contained biotopes or species with a biological value or if they were mapped as important habitats for fish.

Since the bird data only covered a small part of the Holmöarna study area, these values were not added to the map of total biological value. Instead, these data are presented in a separate map.

For the **Kvarken Archipelago** the map of biological values was only based on the biotopes. The map was done by taking the highest biological value for a biotope occurring in each grid cell of the map (as step 1 in the method used for Holmöarna). The filamentous algae biotope included biotopes with different biological value (biotopes dominated by annual algae, perennial filamentous algae and *Vaucheria* sp.; Table 4) and was given the value 1. The biotope maps over this area show the probability for occurrence of the species characterizing the biotope rather than 10% cover of these species (as used in the HELCOM definition). For the *Fucus* and rooted plants biotopes, 50-60% of the stations with occurrence of the characterizing species had a cover of 10% or more and would thus have been classified as this biotope according to the HELCOM definition. We therefore modified the occurrence maps for the biotopes based on the cover information from the modelling data set. We included the 50% of the area with predicted occurrence of the biotope in the map of biological values, removing the 50% with the lowest probability of occurrence of the species. For the filamentous algae biotope more than 90% of the stations with occurrence had at least 10% cover, so we used the entire predicted distribution of occurrence of the characterizing species as predicted distribution for this biotope.

3. RESULTS AND DISCUSSION

3.1. Modelling of biotopes in Holmöarna

Depth and SWM were the most important predictor variables for all modelled biotopes, while slope and topographic position contributed relatively little to the models (Table 5). For four of the biotopes (Perennial algae and moss, Annual algae, Submerged rooted plants and Tall vascular plants), the models got good evaluation results. Deviance explained was relatively high (around 40 %), considering that substrate was not included in the models, and the cross validated discriminatory ability was good (>0.85 AUC). External validation of the prediction maps also indicated that the models were able to produce reliable prediction across the study area.

The two remaining biotopes (Epibenthic fauna and Charales) were more difficult to model reliably. For these biotopes, less than 30 % of the deviance was explained by the model (Table 5). The cross validated AUC was lower although still acceptable, while external validation indicated a poor result for Charales. Based on the low explained deviance and poor external validation, the prediction maps were judged to be too unreliable to be used for management and are not shown in the report or used for further analyses.

The low modelling success for epibenthic fauna is likely due to the fact that important predictors for the distribution of this group was not included in the models. Availability of hard substrate is likely a key factor for this group, so inclusion of substrate would probably improve the model. For Charales the reason is less obvious, since this group can be expected to respond to the same predictors as the other plant groups. We suspect that the group includes species with different ecological requirements and that it would be useful to model Charales biotopes on muddy sediment and sand/coarse sediment separately, but we did not have time to test this within the present project.

The predictions that got a good validation result are shown in Figure 9. As expected, the macroalgal biotopes occur in shallow areas exposed to wind and waves, with the Annual algae biotope typically occurring in the shallowest areas and the Perennial algae biotope extending deeper. The Submerged rooted plants biotope replace the macroalgal biotopes in sheltered areas and the distribution of Tall vascular plants (equivalent to the biotopes dominated by Pondweed and Watermilfoil) had a similar but more restricted distribution.

Table 5. Model performance (measured as cross validated AUC and amount of deviance explained) and predictive performance (showing the quality of the output prediction maps according to an external validation data). AUC is a measure of the discriminatory ability and shows in how many cases a randomly chosen presence will have a higher predicted probability of occurrence than a randomly selected absence. Also shown is relative variable importance for the variables in the final model and prevalence (Prev.) in the calibration data set (calculated as number of presences divided by the total sample size). TPI = Topographic position index. Models that were judged to perform well and produce reliable predictions are indicated with an asterisk.

	Model performance		Predictive performance	Variable importance (%)				
Response	**cvAUC**	**Deviance**	**extAUC**	**Depth**	**Slope**	**TPI**	**SWM**	**Prev. (%)**
Perennial algae and moss*	0.90	0.44	0.89	30		9	61	30
Annual algae*	0.86	0.45	0.84	40	6		54	39
Epibenthic fauna	0.84	0.28	0.84	20			80	16
Submerged rooted plants*	0.88	0.42	0.89	30	3		67	32
Charales	0.80	0.24	0.79	57		3	41	29
Tall vascular plants*	0.89	0.39	0.90	12		7	81	21

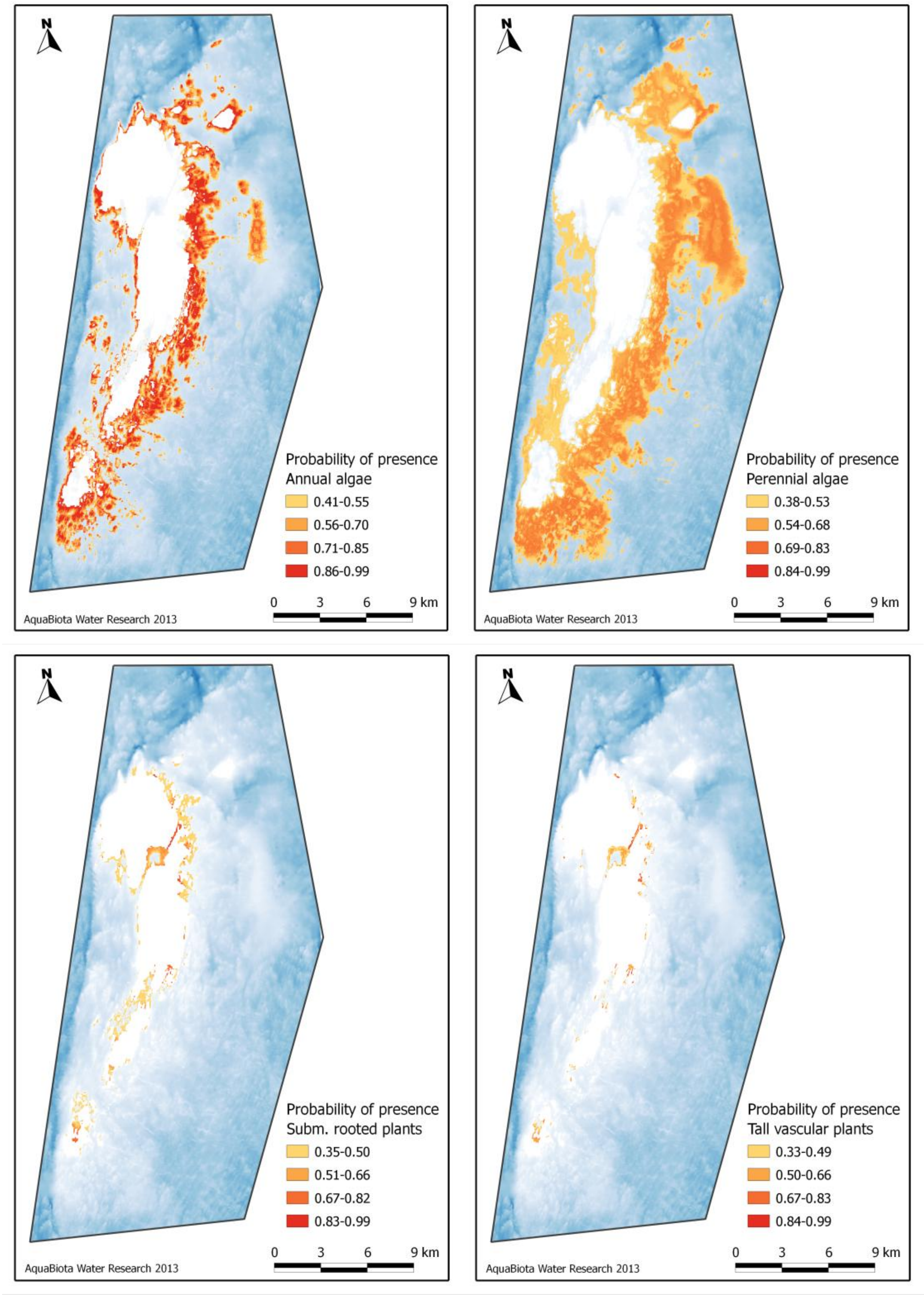

Figure 9. Predicted distribution of the BS EUNIS biotopes characterized by annual algae, perennial algae and moss and submerged rooted plants (at least 10% cover of the biotope-forming species), and predicted distribution of at least 10% cover of tall vascular plants in Holmöarna.

Modelling of biotopes in the Kvarken Archipelago

Model performance

The biotope "*Zannichellia* spp. and/or *Ruppia* spp." only occurred in 0.8 % of the calibration data which was not enough to produce reliable results and hence no results are shown for this biotope. The other biotopes varied between common to rare. "Filamentous algae" was the most commonly occurring biotope, followed by "*Fucus* sp." and "Submerged rooted plants" (Table 6).

Depth was the only predictor variable included in all models, although it was not always the most important predictor variable. The second most used predictor was wave exposure, which was included in all but one model ("Charales"). Salinity and Secchi depth were included in three models each (Table 6). The general importance of depth

Table 6. Model performance measured as cross validated AUC and amount of deviance explained (± SD). AUC is a measure of the discriminatory ability and shows in how many cases a randomly chosen presence will have a higher predicted probability of occurrence than a randomly selected absence. Also shown is relative variable importance (± SD) for the variables in the final model and prevalence in the calibration data set (calculated as number of presences divided by the total sample size). Responses with a prevalence <10 % were evaluated using five-fold cross validation and >10 % using ten-fold cross validation). Models that were judged to perform well and produce reliable predictions are indicated with an asterisk.

	Model performance		**Variable importance (%)**				
Response	**cvAUC**	**Deviance**	**Depth**	**SWM**	**Secchi**	**Salinity**	**Prev. (%)**
Fucus sp.*	0.94 (0.04)	0.53 (0.01)	62 (5)	16 (3)	4 (2)	18 (5)	23
Perennial non-fil. corticated red algae	0.91 (0.04)	0.40 (0.05)	38 (15)	11 (5)	51 (15)		4.0
Submerged rooted plants*	0.95 (0.03)	0.59 (0.01)	35 (15)	47 (13)	18 (7)		15
Pondweed*	0.94 (0.05)	0.50 (0.01)	24 (13)	76 (13)			12
Watermilfoil	0.91 (0.02)	0.40 (0.02)	81 (5)	11 (11)		8 (12)	3.7
Charales	0.92 (0.04)	0.41 (0.03)	63 (10)			37 (10)	4.1
Filamentous algae*	0.90 (0.04)	0.43 (0.01)	61 (2)	39 (2)			69
High vegetation > 50 cm*	0.92 (0.04)	0.40 (0.03)	32 (15)	68 (15)			3.8

and wave exposure both here and in Holmöarna shows that they are important predictors for phytobenthos both at local and regional scale in the Northern Quark. This means that a detailed and reliable depth map is essential for successful prediction of the distribution of phytobenthic biotopes.

Deviance explained was in general relatively high (40-59 %), considering that substrate was not included in the models. Cross validated discriminatory ability was excellent for all models (>0.90 AUC) and the models were generally stable as indicated by low AUC standard deviation (Table 6).

Predicted distribution

The distribution of the biotopes across the entire Kvarken Archipelago areas was predicted using the FINMARINET bathymetric model together with the other predictor variables. External validation of the maps indicated a good performance of all models (Table 7). However, examination of the models and the prediction maps indicated that the quality of the predictions was poor for three of the biotopes; Perennial non-filamentous red algae (i.e. *Furcellaria lumbricalis*), Watermilfoil (*Myriophyllum* sp.) and Charales. For *F. lumbricalis*, this may be due to the fact that the species has its distributional limit in the northern Quark. The distribution of marginal populations can be expected to be less predictable due to chance extinctions and colonisations. The Charales group gave a poor model also for Holmöarna, which as mentioned before could be due to the fact that the group includes species with different ecology.

Table 7. Predictive performance (external AUC) in the entire area and in the subset based where lidar measurements were available. In the latter case, predictions using field measured depth, lidar measurements and the FINMARINET bathymetry model were compared, keeping the remainder of the data static. Note that the AUC values from the entire area are not directly comparable to those of the subset since the prediction in the subset area was done in a narrower environmental range (see the text for explanation). Models that were judged to perform well and produce reliable predictions are indicated with an asterisk.

	Entire area	**Subset (Norrskär and Rönnskär)**		
Response		**Field depth**	**Lidar**	**Finmarinet**
Fucus sp.*	0.93	0.87	0.86	0.85
Perennial non-fil. corticated red algae	0.85	0.70	0.69	0.68
Submerged rooted plants*	0.93	0.90	0.91	0.89
Pondweed*	0.93	0.88	0.88	0.87
Watermilfoil	0.93	0.88	0.88	0.85
Charales	0.88	0.73	0.74	0.74
Filamentous algae*	0.83	0.80	0.79	0.79
High vegetation > 50 cm*	0.92	0.89	0.89	0.85

An alternative, but not exclusive, explanation to the poor result for these three biotopes is that there were too few occurrences to allow modelling of the distribution of these species over such a large area with several strong gradients. All these biotopes had a low prevalence in the field data (around 4%, see Table 6). This raises an important issue related to the modelling of HELCOM Underwater Biotopes. Modelling of level 6 biotopes (e.g. *Fucus*) will always give lower prevalence than modelling of level 5 biotopes (e.g. perennial algae and moss). On the other hand, we may be more interested in the level 6 biotopes for management. The trade-off between taxonomic detail and suitability for distribution modelling has to be considered in future work and a general advice is that it is probably most effective to model the highest level biotope that is relevant for the purpose of the study. In the Kvarken Archipelago, we excluded a large part of the available data in order to get a data set that was representative for the entire study area. It would probably be possible to get better models for level 6 biotopes if only modelling the best surveyed areas.

The prediction maps that passed the quality check are shown in Figure 10 and 9. The map of pondweed was very similar to that of submerged rooted plants and is not shown here. Since the maps show the probability for presence of one or more of the characterising species, the maps are likely overestimating the actual distribution of the biotopes, which should have a cover of at least 10% of the characterising species. The map of filamentous algae includes the biotopes dominated by Perennial filamentous algae, Annual filamentous algae and *Vaucheria* sp.

When looking at the external validation of these maps (Table 7), it is important to note that due to the sampling method the evaluation data are not spread across the entire region (see Figure 2). This means that it has not been possible to validate the model in most of the northern part of the study area. This could be done at a later stage by collecting additional field data in these areas.

An issue that should also be kept in mind when interpreting the prediction maps is that they are produced without data on seabed substrate. This means that in many shallow areas, both hard substrate and sediment biotopes are predicted to occur in the same grid cells of the map. This might reflect the true situation – in this region as in many parts of the Baltic Sea the seabed in shallow areas often consists of a mosaic of hard and soft substrate when looking at a spatial resolution of 20 x 20 m. If, however, data on seabed substrate would become available, it could be used to refine the predictions by excluding biotopes from grid cells with wrong substrate.

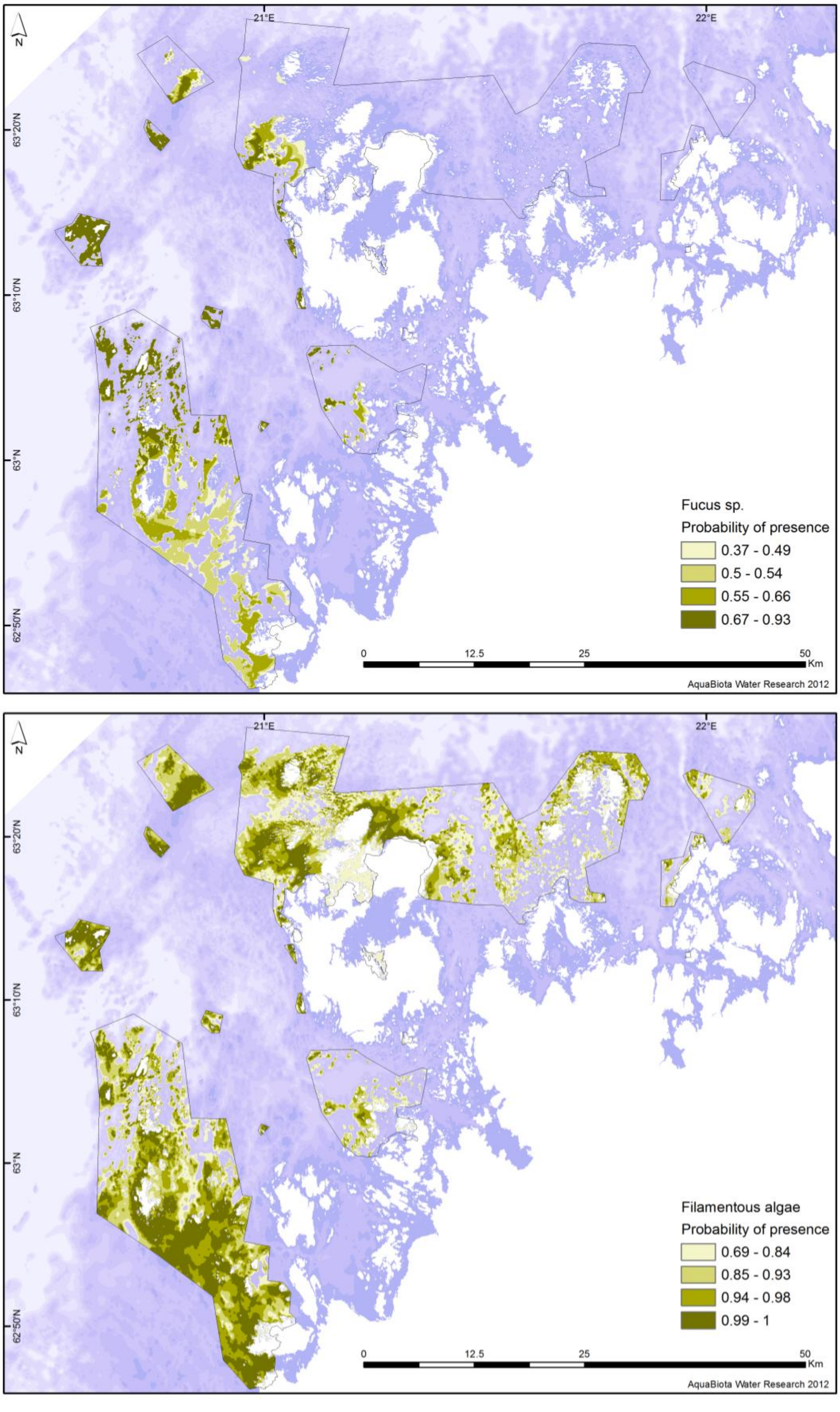

Figure 10. Predicted distribution of Fucus sp. and filamentous algae in the Kvarken Archipelago.

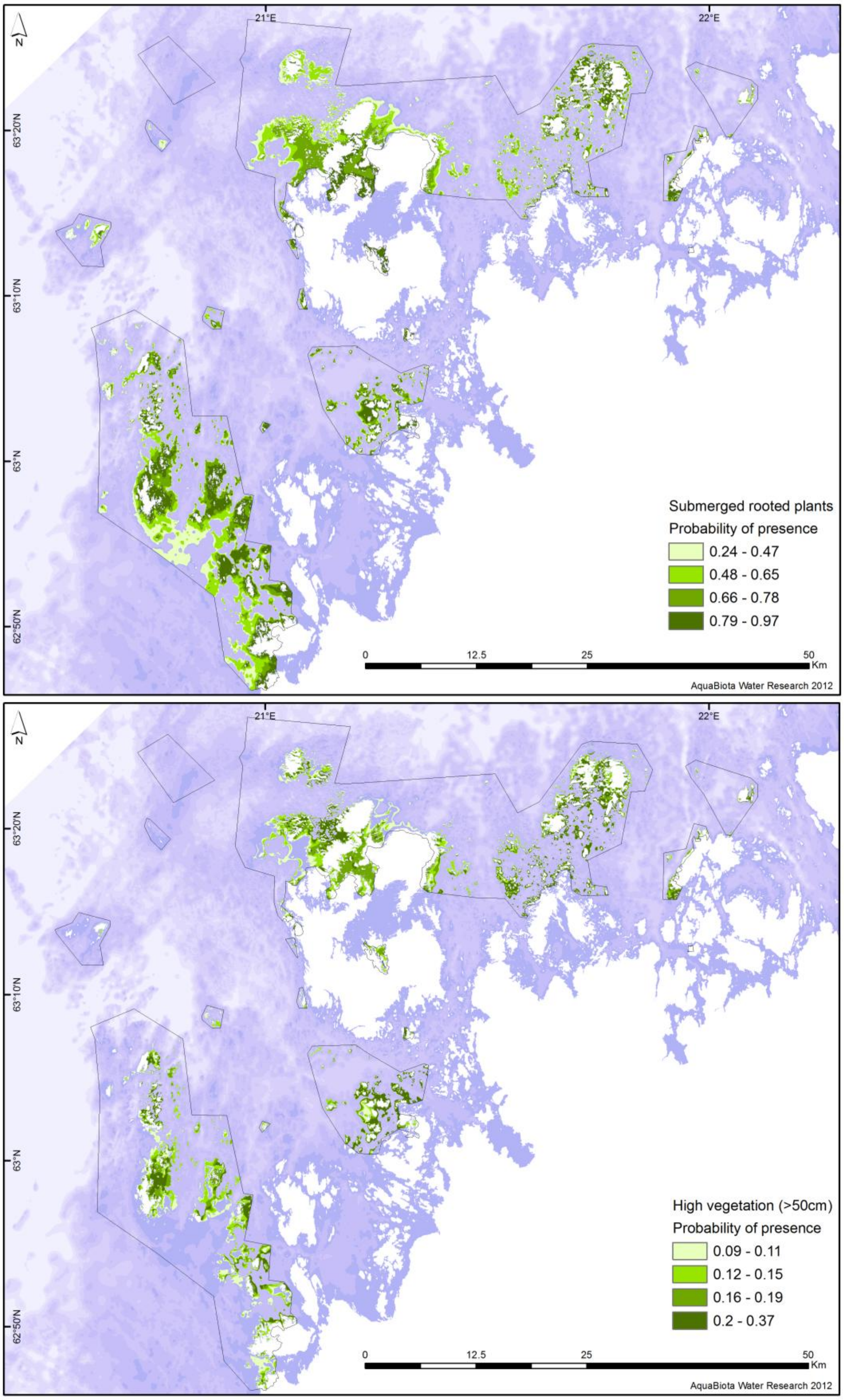

Figure 11. Predicted distribution of submerged rooted plants and high vegetation in the Kvarken Archipelago.

Comparison between different bathymetric sources in a subset of the area

Keeping everything else similar we evaluated the effect of three different sources of depth, field measured, the FINMARINET model and more detailed lidar bathymetry. The comparison showed that there was a statistically significant loss in predictive performance when comparing predictions using the FINMARINET depth model compared to using field measured depth (pairwise t-test: mean AUC field measured depth 0.83, mean AUC FINMARINET depth 0.82, t=3.05, df=7, p=0.019). However, there was no loss in predictive performance when using lidar depth compared to field measured depth (pairwise t-test: mean AUC field measured depth 0.83, mean AUC lidar depth 0.83, t=0.42, df=7, p=0.68). This result indicates that in the context of predictive species distribution modelling, lidar bathymetry is of the same quality as field measured depth.

It should be noted that despite the statistical significance also the FINMARINET depth model resulted in relatively high AUC (Table 7). The difference between the two depth layers becomes most apparent when comparing the output maps (Figure 12). The map produced with lidar bathymetry picks up much more detailed distribution patterns compared to the FINMARINET depth model. The broad distribution patterns are however the same, which means that this data layer can be useful when lidar data is unavailable.

The comparison was restricted to the area where lidar bathymetry was available, and the environmental ranges therefore differed in comparison to the whole study area. Since the evaluation in the lidar area had a narrower environmental range, the associated evaluation values (AUC) are not comparable to those obtained for the whole study area. This difference is in many cases due to including a larger amount of (ddep) unsuitable habitat that increases the AUC value. When the deepest presence is 10 m, evaluating the model down to 15 meters will give a lower AUC (worse performance) than evaluating the model down to 100 m since there will then be many more cases were the model is correct.

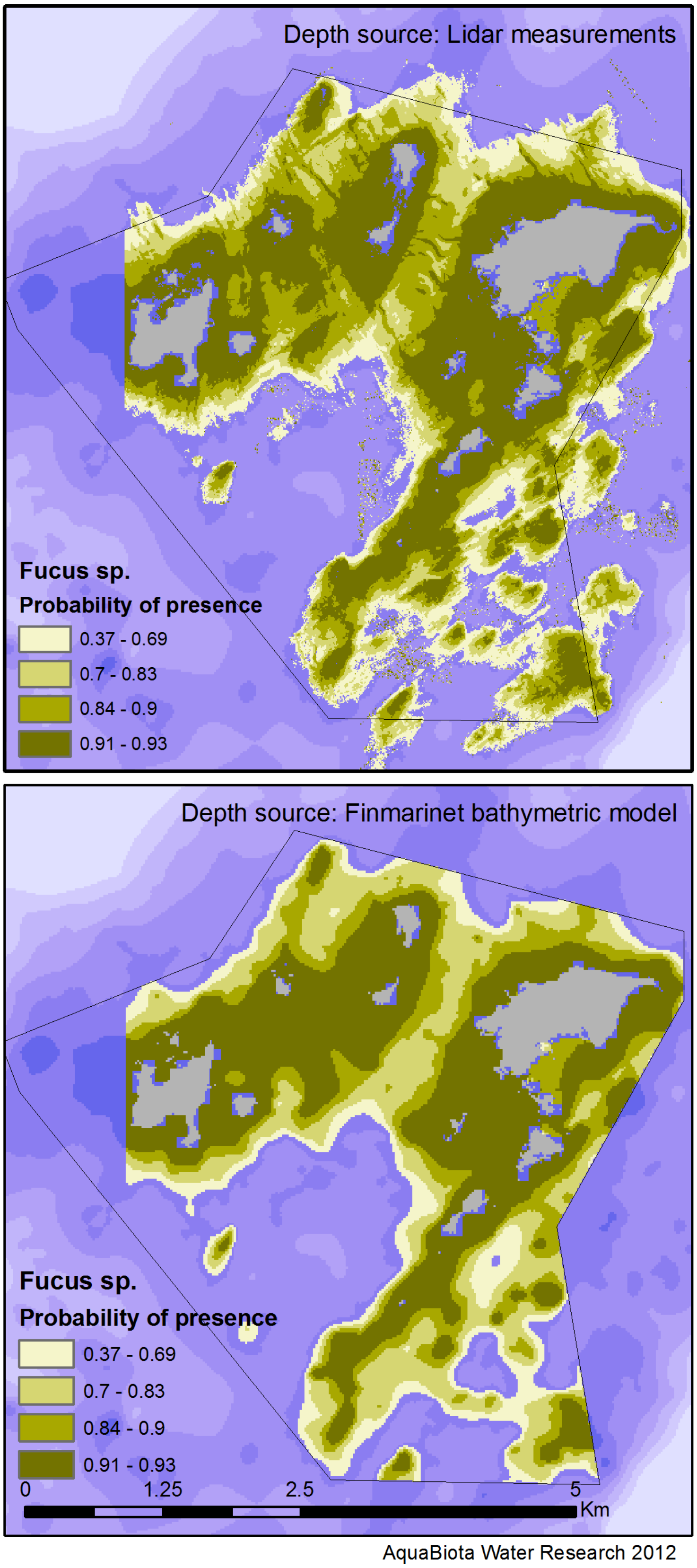

Figure 12. Comparison of prediction maps of the distribution of Fucus sp. using bathymetric data from lidar measurements and from the FINMARINET bathymetric model. The figure shows a subset of the Kvarken Archipelago study area (Norrskär).

3.3. Mapping of Sublittoral sandbanks and Reefs

In the entire Kvarken Archipelago, 625 elevations were identified based on the FINMARINET depth grid and the first mapping approach (Figure 13). The total area of the elevations was 525 km^2. A few of these elevations were located on the boarder of the Kvarken Archipelago Natura 2000 area, the total area of the elevations within the Natura 2000 area was 400 km^2. Since we used the deepest depth curve to delineate the elevation, several of the elevations cover very large areas. For instance, the entire Norrskär ends up in one elevation. This approach has been used for large-scale mapping of Sublittoral sandbanks and Reefs in Sweden (Fyhr 2012), but according to our project partners Anette Bäck and Michael Haldin (Forsstyrelsen) it does not comply with the Finnish definitions of these habitats.

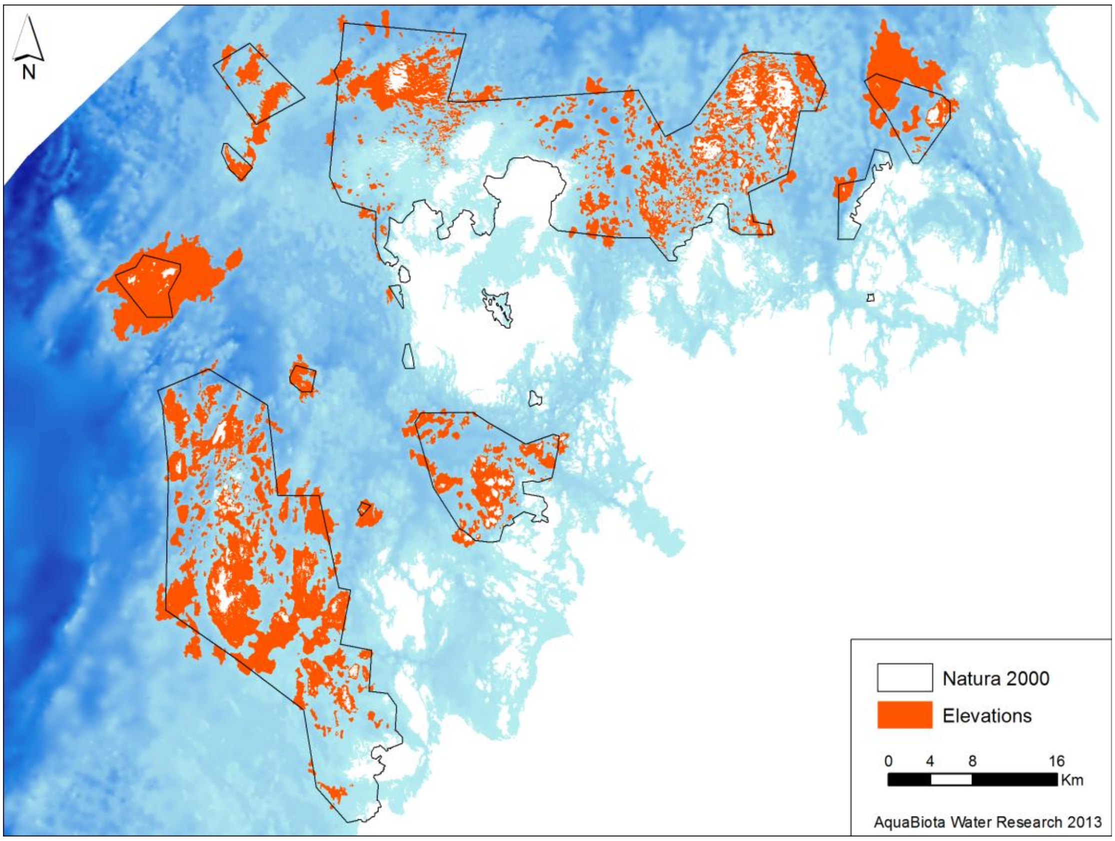

Figure 13. Elevations in the Kvarken Archipelago Natura2000 area, identified based on the FINMARINET depth grid.

The second approach identified 1582 elevations based on the fine scale depth grid in the Rönnskären area (Figure 14). These elevations were selected using a specific size criteria (10-20 m width), to better fit the Finnish definition of Sublittoral sandbanks and Reefs, and are thereby much smaller. The mean area of the elevations identified by this approach was 193 m^2 and the total area of the elevations in the Rönnskären area was 0.31 km^2.

In these examples, we used different approaches with different data sets, but it would of course be possible to delineate smaller objects in the coarse-scale depth grid. The

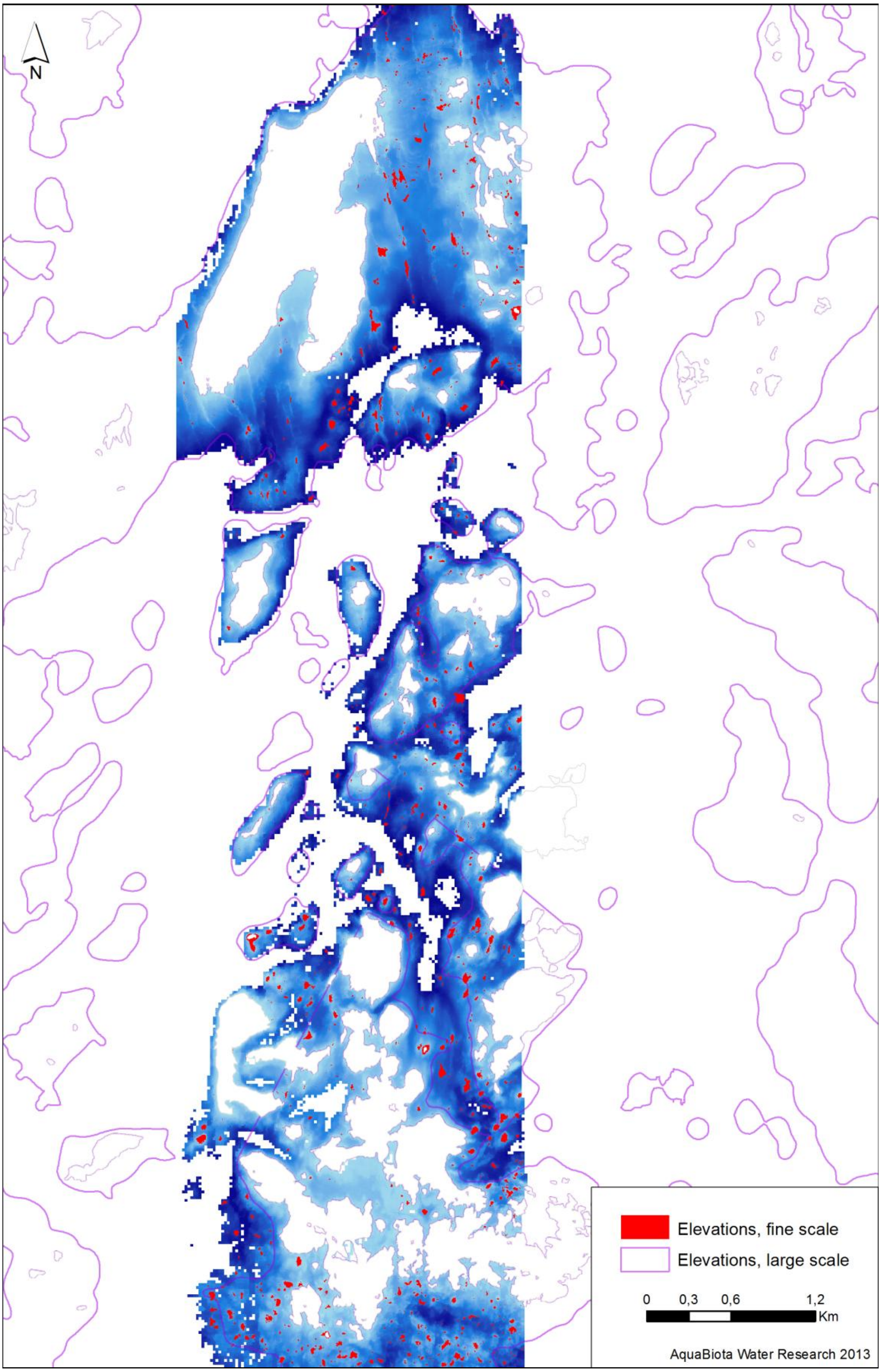

Figure 14. Elevations in the Rönnskären area identified based on the fine scale depth grid from lidar (elevations based on the FINMARINET depth grid are shown for comparison).

resolution of the depth layer will however determine how small elevations that can be mapped. If the Finnish definition of Sandbanks and Reefs only includes elevations with a size of a few tenths of meters, the FINMARINET grid is likely too coarse to be used for this purpose.

In order to assign the mapped elevations to either Sublittoral sandbanks or Reefs, information on dominating substrate is necessary. Only 73 out of 625 elevations derived from the coarse-scale depth layer contained at least one drop video station. Half of these had only one single station, only 15 had 10 stations or more and only five contained more than 100 stations. Similarly, only 24 out of 1582 elevations derived from the fine scale depth layer contained in a drop video inventory station. Figure 15 show two examples from the Rönnskären area, where the elevations derived by the two different depth layers are overlaid by drop video data with substrate information. In cases where drop video stations were available we added information on dominating substrate to the shape file with the elevations, but it should be noted that in most cases this information is based on very few observations.

Since we could not classify most of the elevations to Sublittoral sandbanks or Reefs, we did not eliminate elevations deeper than 30 m, which should be done for sandbanks according to the definition for this habitat.

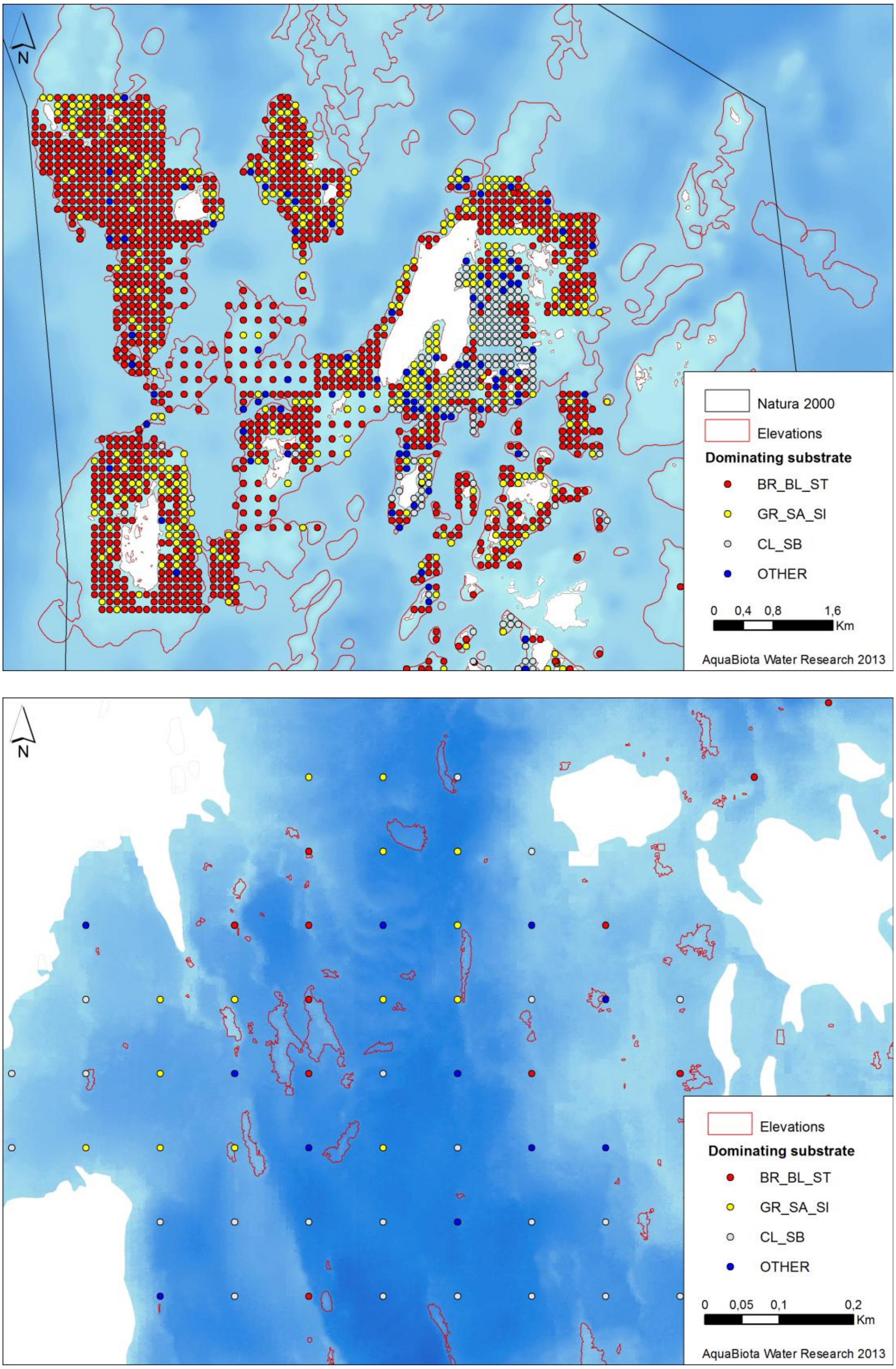

Figure 15. Identified elevations in the Rönnskären area in the Kvarken Archipelago, based on the coarse-scale (upper panel) and fine-scale (lower panel) depth layers, overlaid by drop video stations with substrate information; hard substrate (BR_BL_ST), sand and coarse sediment (GR_SA_SI), fine sediment (CL_SB) and sites where none of these groups dominated (OTHER). Note the difference in map scale between the two panels.

3.4. Biological valuation

Holmöarna

The map of biological values for Holmöarna, when considering phytobenthic biotopes and habitats and fish, is shown in Figure 16. Values of 10-14 show the presence of one or more elements (biotopes, habitats or important fish areas) with a high biological value and areas with values of 20-22 have two such elements, combined or not with elements of lower value. The map shows that high biological values for these groups are largely confined to the shallowest areas and to enclosed bays and sounds. These areas get high values from rooted plant biotopes and/or the habitat Coastal lagoons, which represent threatened biotopes and habitats with an important function as habitat for other species.

It is important to note that the map only shows values from phytobenthic communities and fish. For instance, both hard substrate biotopes characterised by epifauna and sediment biotopes characterised by infauna are excluded since we were not able to model their distributions. Epifauna biotopes were given a low biological value, comparable to the value of the filamentous algae biotopes. Infauna biotopes could not be assessed due to lack of data. The infauna is not seen in the underwater video and an appropriate mapping and valuation of these biotopes would require collection of sediment grab data with a sufficient spatial dispersion. Inclusion of epi- and infaunal biotopes would probably change the map of biological values. Specifically; the absence of biological values in deeper areas of Holmöarna is due to absence of distributional data on faunal biotopes and should not be interpreted as a lack of values here.

It is also important to note that the data layer on important areas for fish, taken from the Digital Environmental Atlas, is not necessarily complete. For instance, sheltered bays and sounds that are known to be important reproduction areas for several coastal fish populations are not pointed out. This could be due to lack of data or to the fact that the data layer was developed for oil spill emergency situations, which are mainly a threat to more open coastal areas. In the present map, the more sheltered reproduction areas are probably covered fairly well by the distribution of rooted plant biotopes and coastal lagoons. Nevertheless, in order to map all biological values for fish the data used here should be complemented, for instance with the distribution of important reproduction areas for different fish species that could be derived by species distribution modelling.

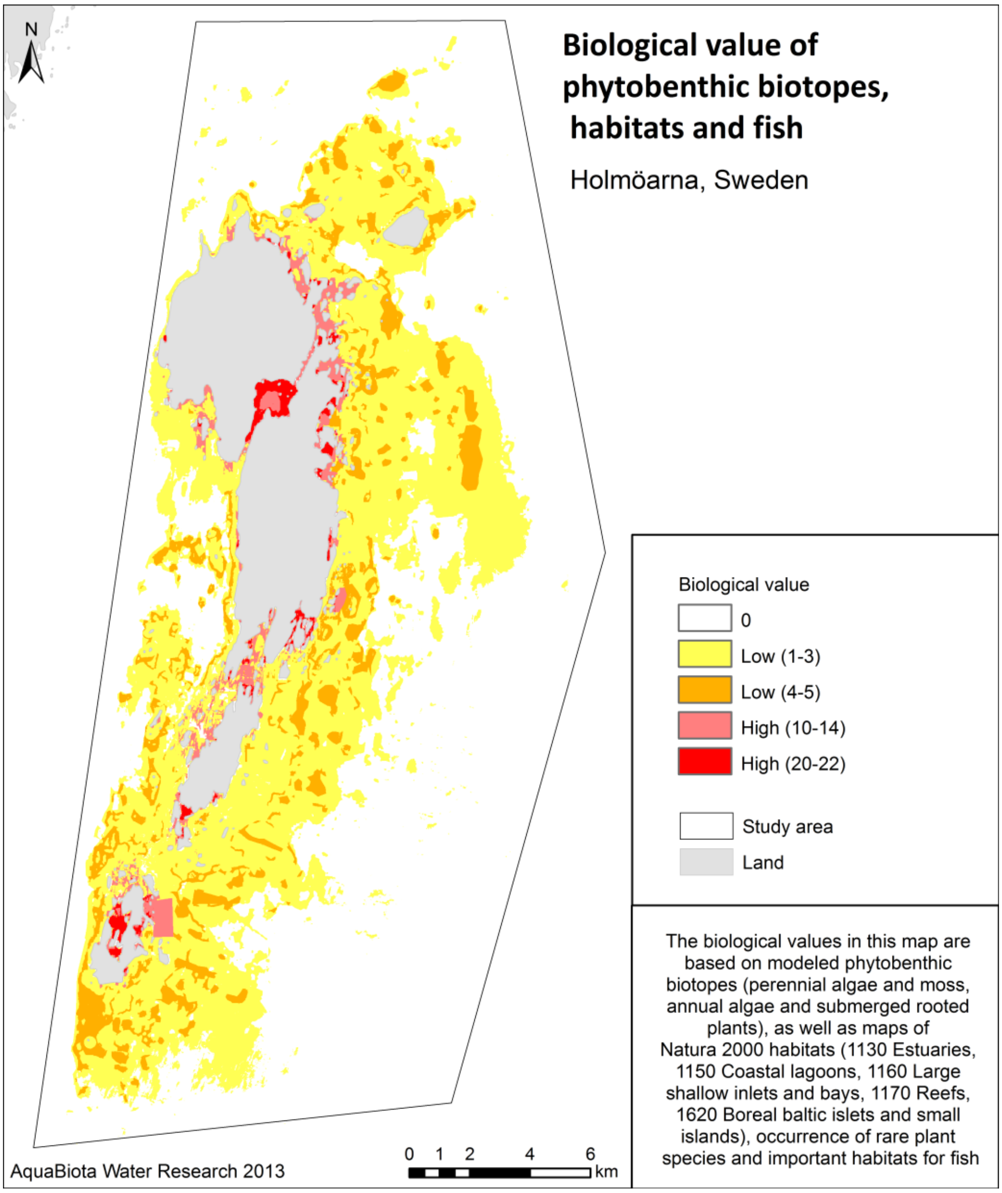

Figure 16. Map of biological values in Holmöarna, based on the modelled distribution of phytobenthic biotopes, Natura 2000 habitats, occurrence of rare plants species and important habitats for fish from the Digital Environmental Atlas.

Since the bird inventory data only covered a small part of the Holmöarna area, it was not included in the overall map of biological values but is presented in a separate map (Figure 17). This data differs from the other background data in that is consists of an inventory of single spatial units (bird islands) which have a position and extent. A comparable dataset for habitats would for instance be detailed inventories of single coastal lagoons, which could then be compared in terms of biological values, or the valuation of the Swedish offshore banks (Naturvårdsverket 2010). This method allows a more detailed mapping of values where different occurrences of a habitat or biotope can be given different values and is clearly preferable for some purposes. However, it requires more resources and time compared to the large-scale mapping that is possible to achieve with spatial modelling approaches, which is also reflected by the restricted cover of the map of values for birds compared to the entire Holmöarna area. Another weakness is that it only can show a value for sites that have been inventoried, meaning that areas of large value can be missing from the map. We see this as two different approaches to mapping of biological values, where the first (assigning values to biotopes/habitats and mapping those) is useful for large-scale mapping of biological values across areas or regions, while the second (inventory of single objects) provides a detailed valuation of smaller areas. One way of combining the approaches could be to start with large-scale mapping of biotopes and habitats of high value and then perform more detailed investigations in areas identified to be important.

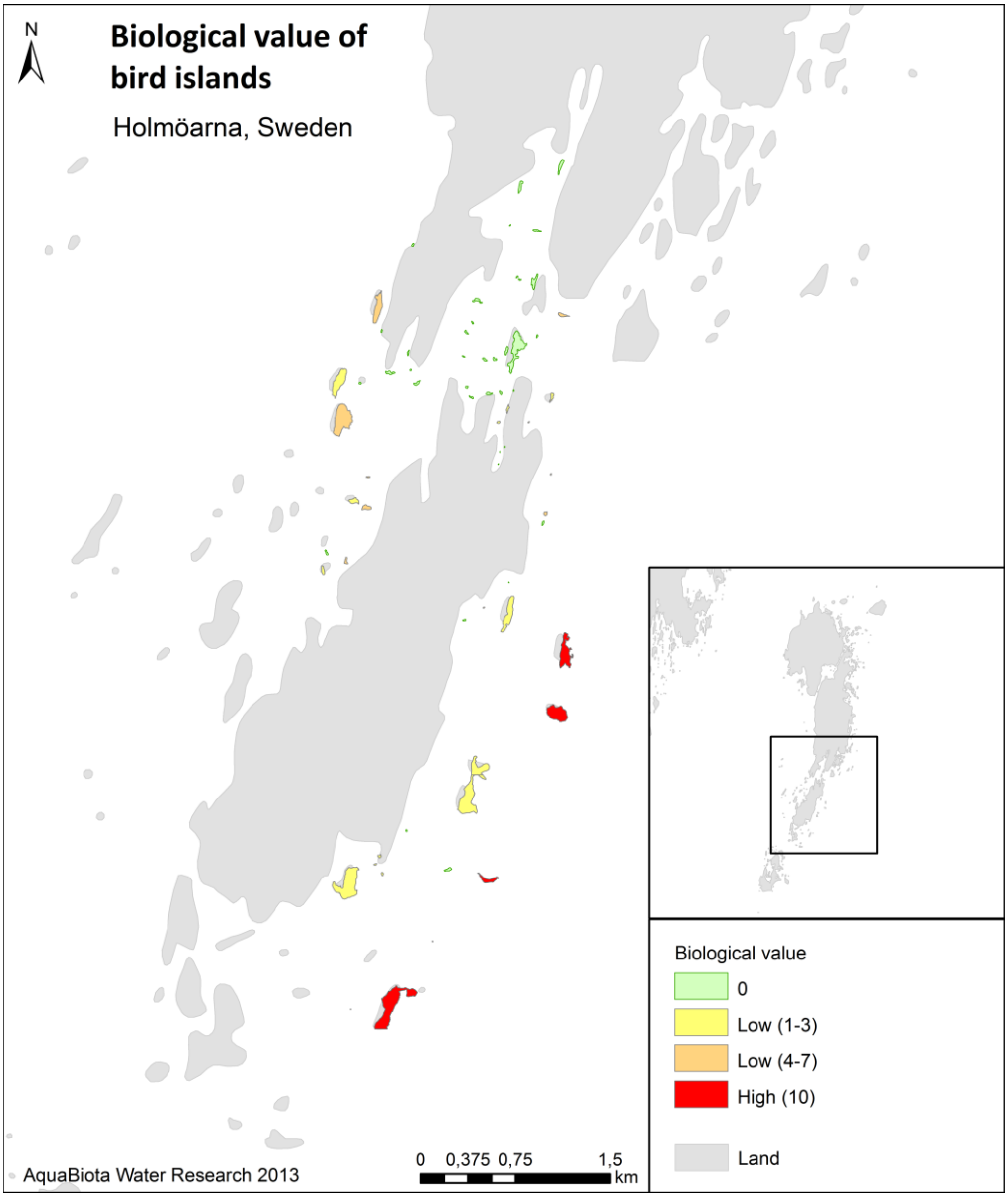

Figure 17. Map of biological values for birds in Holmöarna.

The Kvarken Archipelago

For the Kvarken Archipelago we had no spatially covering data of the Natura 2000 habitats. A map of the spatial distribution of whitefish larvae is available from the project INTERSIK (Vanhatalo et al 2012), but was not included since this was the only available data on important fish areas. Instead, the map of biological values is based only on the modeled distribution of phytobenthic biotopes (Figure 18). The map indicates large areas of high values in sheltered parts across the study area. These areas get high values from the presence of biotopes dominated by rooted plants and/or *Fucus* sp.

As for the map of biological values for Holmöarna, it should be pointed out that values connected to epifaunal and infaunal biotopes are not included in the map.

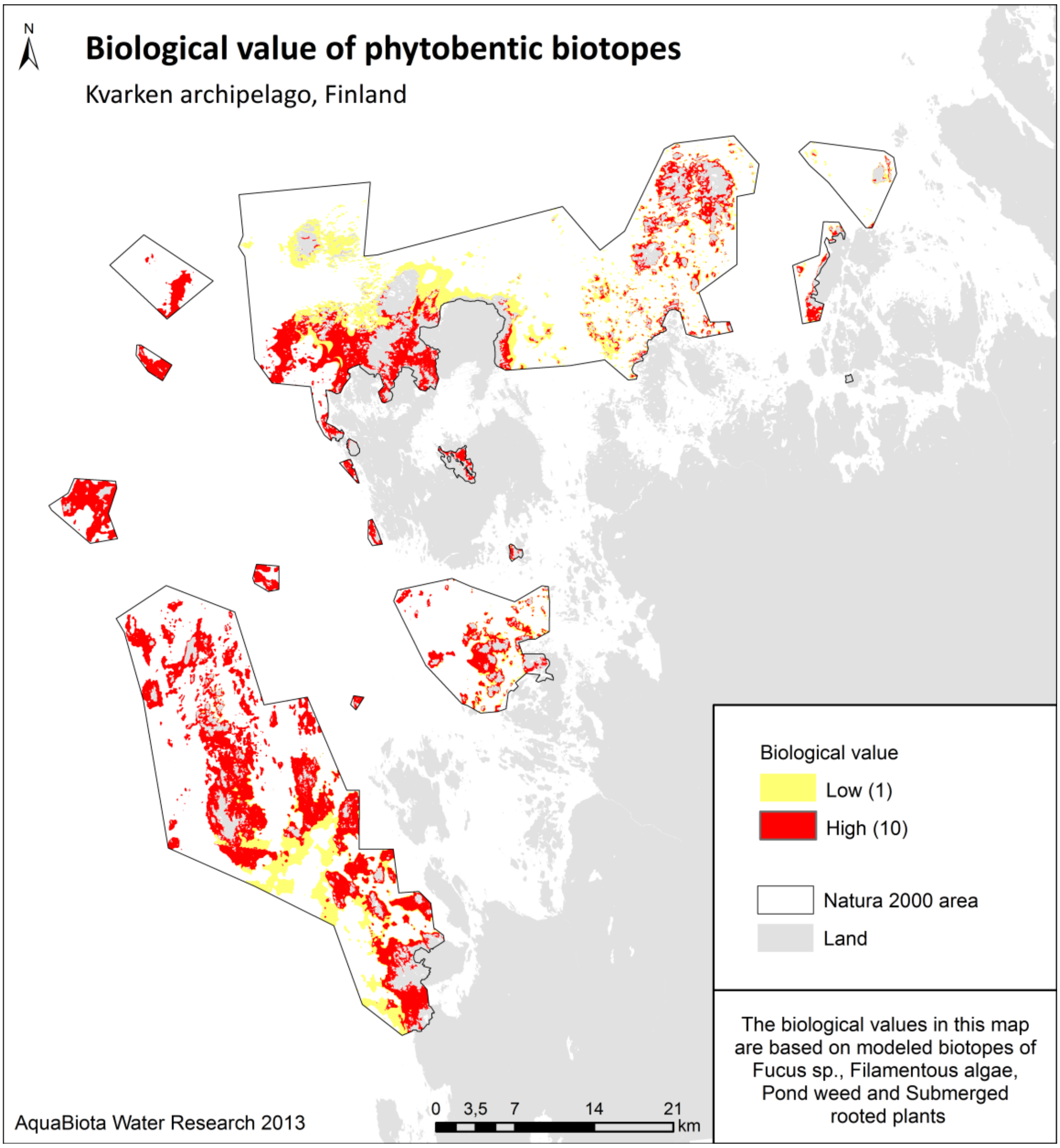

Figure 18. Map of biological values in the Kvarken Archipelago, based on the modelled distribution of phytobenthic biotopes.

4. CONCLUSIONS

- Depth was an important predictor for all the modelled biotopes, which shows that a good bathymetric map is a prerequisite for modelling of many benthic biotopes. The comparison of the bathymetric map derived from lidar with the coarser depth layer from FINMARINET shows that the accuracy and level of detail of the prediction maps increased considerably. The lidar bathymetry also allowed mapping of fine scale topographic elevations that can be calssified as Sublittoral sandbanks or Reefs. However, it should be noted that the coarse FINMARINET bathymetry still gave a good regional picture of the distribution of biotopes.

- The lack of substrate maps was clearly a limitation for the mapping of biotopes, Natura 2000 habitats and biological values.

- The sampling design for field data has a large effect on how the data can be used. Random or stratified random sampling is preferable for distribution modelling since it reduces the risk for prediction biases. Likewise, statistical determination of occurrence frequency of species, biotopes or habitats (as we did for the valuation) requires random sampling. In the case of the Kvarken Archipelago, we had to exclude a large part of the existing data to get a balanced data set for the entire study area. It is probably possible to get better results for the areas with a lot of field data if these would be modelled separately, as the amount of available data is important for what can be modelled.

- Mapping of the Natura 2000 habitats 1110 Sublittoral sandbanks and 1170 Reefs requires a detailed and reliable bathymetric map and a map of seabed substrate. The choice of mapping method depends on how the habitats are defined.

- We present a method for regional mapping of biological values, by valuating habitats and biotopes that can be mapped on a regional scale. The success of this approach will depend on that a sufficient number of biotope and habitat maps can be produced for a study area and that data is available to objectively assess the value for each biotope and habitat along the criteria that are chosen. Ideally, the choice of criteria and the valuation is done in collaboration between specialists and managers with good knowledge about local conditions.

- For birds we used a potentially complementing approach to mapping of biological values, where the value of single objects (in this case bird islands) are assessed. This approach allows a more detailed mapping of values, where the value can vary within a single biotope or habitat (e.g. separating between coastal lagoons with high or low value). However, it requires much more extensive field surveys and are only possible to carry out in limited areas or a limited number of biotopes/habitats.

ACKNOWLEDGEMENTS

We thank Anette Bäck and Michael Haldin (Forststyrelsen) and Johnny Berglund and Kristin Dahlgren (County Administrative board of Västerbotten) for valuable discussions and input throughout the project. The development of the method for biological valuation greatly benefitted from input by Johnny Berglund and Kristin Dahlgren (County Administrative board of Västerbotten) and Martin Isaeus, Karl Florén, Joakim Hansen and Antonia Nyström Sandman (AquaBiota). We are also grateful to all experts that participated in the workshop on biological valuation of the underwater environment in the Gulf of Bothnia (Umeå, December 2011). Anna Nikolopoulos (AquaBiota) prepared the bathymetric data layers for the species distribution modelling.

The work was financed by EU Interreg IV A through the project SUPERB (Standardiserad Utveckling av Planering och Ekologiska Redskap för Bottniska viken).

REFERENCES

Bergström, U., Sundblad, G., Downie, A.-L., Snickars, M., Boström, C. & Lindegarth, M. 2013. Evaluating eutrophication management scenarios in the Baltic Sea using species distribution modelling. Journal of Applied Ecology (published online).

Carlström J., Florén, K., Isaeus M., Nikolopoulos A., Carlén I., Hallberg O., Gezelius L., Siljeholm E., Edlund J., Notini S., Hammersland J., Lindblad C., Wiberg P. & Årnfelt E. 2010. Modellering av Östergötlands marina habitat och naturvärden. Länsstyrelsen Östergötland, rapport 2010:9.

CBD 2008. Decition adopted by the conference of parties to the convention on biological diversity at its ninth meeting. IX/20 Marine and coastal biodiversity. UNEP/CBD/COP/DEC/IC/20

CDB 2009. Azores scientific criteria and guidance for identifying ecologically or biologically significant marine areas and designing representative networks of marine protected areas in open ocean waters and deep sea habitats.

Dahlgren K., Berglund J., Wikström S.A. & Stenroth P. 2013: Naturvärdesbedömningar i Bottniska viken.

Fielding, A.H. & Bell, J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation, 24: 38-49.

Florén K., Nikolopoulos A., Fyhr F., Nygård L., Hammersland J., Lindblad C., Wiberg P., Näslund J., & Isæus M. 2012: Modellering av Västernorrlands marina habitat och naturvärden. Länsstyrelsen Västernorrland Rapport 2012:03.

Fyhr F. 2012: GIS-utsökning av potentiella Natura 2000 naturtyper 1170 rev och 1110 sandbankar. Med särskilda fokusområden i Hanöbukten, Västernorrland och Skagerrak. AquaBiota Notes 2012:01

Gullström M., Sundblad G. Mörk E., Lilliesköld Sjöö G., Johansson M., Halling C. & Lindegarth M. 2013. Utvärdering av visuella undervattensmetoder för uppföljning av marina naturtyper och typiska arter: variation, precision och kostnader. Project report, Naturvårdsverket.

Gustafsson, A., Hamrén, U. & Rydin, E. 2008. Kustnära marina naturvärden i Österåker kommun. Rapport från Naturvatten och Ekologigruppen.

Haldin J., Leinikki, J., Näslund, J. & Laamanen, M. 2012. EUNIS Compatible Classification of Baltic Sea Biotopes. Conference proceeding in: Galparsoro, I. (Ed.). Using EUNIS Habitat Classification for Benthic Mapping in European Seas. Revista de Investigación Marina, AZTI-Tecnalia, 19: 21-70.

Hosmer, D.W. & Lemeshow, S. 2000. Applied logistic regression, Wiley, New York

Isæus, M. 2004. Factors structuring Fucus communities at open and complex coastlines in the Baltic Sea. In Department of Botany, p. 35. Stockholm University, Stockholm.

Jiménez-Valverde, A. & Lobo, J.M. 2007. Threshold criteria for conversion of probability of species presence to either-or presence-absence. Acta Oecologia, 31: 361-369.

Naturvårdsverket. 2009. Basinventering av Natura 2000 och skyddade områden 2004-2008. Report 5990.

Naturvårdsverket 2010. Undersökning av utsjöbankar. Inventering, modellering och naturvärdesbedömning. Naturvårdsverkets rapport 6385.

Sohlman, A. (red.) 2008. Arter och naturtyper i habitatdirektivet – tillståndet i Sverige 2007. ArtDatabanken SLU, Uppsala

Sundblad, G., Bergström, U., Sandström, A. & Eklöv, P. 2013. Nursery habitat availability limits adult stock sizes of predatory coastal fish. ICES Journal of Marine Science (published online).

Vanhatalo, J., Veneranta, L. & Hudd, R. 2012: Species distribution modeling with Gaussian processes: A case study with the youngest stages of sea spawning whitefish (*Coregonus lavaretus* L. s.l.) larvae. Ecological Modelling 228: 49– 58

Wennberg, S., Nöjd, A. & Lindblad, C. 2008. Mapping of the NATURA 2000 Annex 1 habitats in Finnish and Swedish waters using GIS analyses. In: Dinesen (et al). 2008. Mapping and modelling of marine habitats in the Baltic Sea region. Draft. BALANCE interim report No. 27. Available at http://balance-eu.org/

Wood, S.N. 2006. Generalized Additive Models: an introduction with R, CRC.

Wood, S.N. & Augustin, N.H. 2002. GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. Ecological Modelling, 157: 157-177.

www.aquabiota.se