Fish habitat modelling in a Baltic Sea archipelago region

Baltic Sea Region
INTERREG III B

BALANCE

Title				BALANCE Interim Report No. 11			
Fish Habitat Modelling in the Archipelago Sea							
Authors				Date May 2007			
Ulf Bergström, Alfred Sandstrøm, Göran Sundblad Swedish Board of Fisheries, Institute of Coastal Research, Box 109, SE-742 42 Öregrund, Sweden				Approved by Johnny Reker			
Revisi	Description			By	Checked	Approved	Date
1	Final report			AK	JYR	JYR	19/8-06
0	Draft report			AK	AR	JYR	19/5-06
	Front page illustration						
Key words				Classification			
BALANCE, fish habitats, Archipelago Sea, Perch, Pikeperch,				☒ Open ☐ Internal ☐ Proprietary			

Distribution	No of copies
BALANCE Secretariat	3 + pdf
BALANCE partnership	20 + pdf
BSR INTERREG IIB Joint Secretariat	1
Archive	1

CONTENTS

0   PREFACE

This report present a description of the habitat modelling activities conducted by the Swedish Board of Fisheries in the Archipelago Sea during 2006, the BALANCE Pilot Area 3. The activities are part of the BSR INTERREG IIIB project "BALANCE".

The habitat modelling has been concentrated on developing models for four of the most ecologically as well as economically important fish species in the coastal ecosystem of the Baltic Sea. These are the freshwater fishes Eurasian perch (*Perca fluviatilis*), northern pike (*Esox lucius*), pikeperch (*Sander lucioperca*) and roach (*Rutilus rutilus*). Pikeperch and northern pike can be considered as obligate piscivores, perch as a facultative piscivore and roach as a generalist and omnivore and an important prey fish for the other species. They all depend to a varying extent on shallow near-shore areas during their early life stages, mainly since such areas are heated more rapidly early in spring and high water temperature is needed for juvenile development and survival (Karås and Hudd 1993). All four species have been subject to numerous studies, and their biology and ecology is relatively well known. This information has been used in developing the habitat model for a large transnational area in the Baltic Sea.

Further information on the BALANCE project and electronic copies of this report can be obtained at www.balance-eu.org.

Ulf Bergström

The Swedish Board of Fisheries, Institute of Coastal Research

1   INTRODUCTION

Continuous maps of habitats and ecological communities are needed for an efficient management of the heavily exploited coastal zone. Detailed knowledge on the spatial extent of essential habitats for fishes and other organisms in the Baltic Sea is however sparse, as is knowledge on the effects of habitat availability and quality on population sizes and structures. Given the complex structure of the many coastal areas within the Baltic Sea Region, e.g. the fragmented archipelagos and strong environmental gradients, it is very costly to perform surveys that cover all potentially interesting areas for the species/communities of interest. Spatial predictive modelling, using key habitat characteristics to identify areas of particular interest could provide a tool to circumvent this problem. The underlying concept of these models is that certain habitat characteristics are needed to host specific species, assemblages or communities. Using these environmental variables as predictors in statistical GIS models, maps may be produced that can be used in marine spatial planning, and that may promote an enhanced understanding of the sensitivity and functioning of coastal ecosystems.

2   MODELLING AND VALIDATION OF MODELS

Modelling the correlation between the spatial distribution of species and environmental conditions has a long history (e.g. Johnston 1924), but it is since the early 1990ies, with the advancements of statistical approaches and GIS tools, major progress has been made (Lek & Guegan 1999, Guisan et al. 2002, Guisan & Thuiller 2005). Both abiotic and biotic factors (e.g. temperature, salinity, depth, substratum and vegetation) have been used in several studies to predict the distribution of species, e.g. macrophytes (Lehmann 1998), sandeel (Wright et al. 2000), spiny lobster (*Panulirius argus*) (Bello et al. 2005), spawning habitat for sole (*Solea solea*) (Eastwood et al. 2001), and small fish presence and abundance (Francis et al. 2005).

Due to their flexibility, generalized additive models (GAM) have become popular in species distribution modelling (Hastie & Tibshirani 1990, Guisan & Zimmerman 2000, Austin 2002). Basically, GAM is an extension of generalized linear models (GLM), with the advantage of better handling of highly non-linear relationships and a data rather than model-driven approach (Guisan et al. 2002).

In species distribution modelling the objective is often to link the distribution of a response variable (species/community) to environmental variables, which then can be used to map potential distribution of the species/community or habitat suitability. Modellers utilise a wide range of validation techniques, primarily to validate the model itself, while validation of spatial predictions are rarely made (Carroll et al. 1999, Vaughan & Ormerod 2005). There has been a rapid increase in available methods for species distribution modelling, and there is yet no consensus among the scientific community on how to best describe the potential and limitations of a model (Vaughan & Ormerod 2005). Many authors recommend the use of separate data sets when building and evaluating models (Chatfield 1995, Fielding and Bell 1997, McPherson et al. 2004, Vaughan & Ormerod 2005), although such an approach runs the risk of comparing sampling occasions or methods rather than of model results. Whether or not new data has been used in validation, components of accuracy and level of generalization need to be specified to aid comparisons with different models and assess model usefulness in different situations (Carroll et al. 1999, Justice et al. 1999, Pearce & Ferrier 2000). It is not enough to specify only e.g. overall accuracy or sensitivity if end-users/managers are to draw appropriate conclusions about the usefulness and limitations of a model.

3   REQUIREMENTS FOR DATA USED IN HABITAT MAPPING AND MODELLING

The requirements of datasets used in predictive spatial modelling may be relatively different from that of other traditional ecological studies. Firstly, it is important to clearly specify the theoretical concepts and the potential regulating mechanisms before conducting a survey or utilising data from older surveys. This will enable a better and more conservative selection of environmental variables concentrating initially on the potentially most important ones (Pont et al. 2005). Utilising a high number of environmental variables requires large datasets – in order to validate model predictions, the number of sampled sites should be much higher than the number of environmental variables.

Secondly, it is important that the sampling design comprises large or entire parts of the environmental gradient that governs the distribution of the target species, habitat or assemblage in question. Many of the datasets collected in the Baltic monitoring programs do not meet this criterion. They are often, instead, targeted against the core areas of the distribution of the species. When producing habitat maps over large geographical areas, it is of importance to keep in mind that there may be regional differences in the species-environment relationship. If the statistical models are based only on a smaller, restricted area, biases may occur when exporting the explanation models to larger areas, especially if the environmental variables used as predictors are not directly responsible for the distribution of the target species. A trade-off in this respect is the tendency for the strength of the model predictions to decrease with increasing prediction area, thus demanding more precise and accurate descriptions when increasing prediction area.

Thirdly, it is important that the sampling design is randomised and not too restricted to certain habitats. The latter is potentially a problem for some organism types that are only possible to sample in certain habitats. For example, many biological sampling methods are restricted either to soft or to hard bottom substrates. It may therefore be difficult to cover the potential distribution of species that are not closely associated with a certain bottom type.

Fourthly, it is, especially when considering the objectives of BALANCE and the requirements of WP 3 and 4, important that the explanation models can be used for predicting habitat distributions over large spatial scales. This condition does not only affect the requirements of the field data but also, equally important, demands high resolution maps of the environmental variables that cover the whole areas of interest.

4   *THE MODELLED SPECIES*

Within BALANCE WP2, we have concentrated on developing models for four of the most ecologically as well as economically important fish species in the coastal ecosystem of the Baltic Sea. These are the freshwater fishes Eurasian perch (*Perca fluviatilis*), northern pike (*Esox lucius*), pikeperch (*Sander lucioperca*) and roach (*Rutilus rutilus*). Pikeperch and northern pike can be considered as obligate piscivores, perch as a facultative piscivore and roach as a generalist and omnivore and an important prey fish for the other species. They all depend to a varying extent on shallow near-shore areas during their early life stages, mainly since such areas are heated more rapidly early in spring and high water temperature is needed for juvenile development and survival (Karås & Hudd 1993). All four species have been subject to numerous studies, and their biology and ecology is relatively well known.

Eurasian perch is one of the most common species in European fresh and brackish waters. It is distributed along the coasts of the Baltic Sea with the exception of the most saline south-westerly areas where its distribution is limited to estuaries. It is an important species, particularly for recreational fishery in both Finland and Sweden. The commercial catch of the species has declined in several areas along the Swedish coastline, a phenomenon believed to be partly caused by unsuccessful recruitment from the early 1990s and onwards. Perch is the dominating (biomass) species in the majority of areas where long-time trends in fish community composition has been monitored within the Swedish-Finnish archipelago region (Ådjers et al. 2006). The species is an important predator, which mainly feeds on small fishes and macro-invertebrates in both littoral and pelagic habitats.

Perch normally spawn from April to June in shallow sheltered water where a long (ca. 10-150 cm) slender, single egg strand is normally attached to a protruding substrate, e.g. submerged vegetation or fallen tree branches (Treasurer 1983, Gillet & Dubois 1995, Urho 1996, Smith et al. 2001). Eggs need about 120-200 degree days to hatch, equivalent to 8-20 days in normal spring temperatures (Thorpe 1977). Newly hatched larvae immediately move out to the pelagic zone, where they start to feed on zooplankton prey (Romare 2000). As body size increases, perch undergo an ontogenetic niche shift and move back to the littoral zone (Hjelm et al. 2000). The shift back to the littoral zone has been suggested to be an effect of size-specific predation risk and size-dependent habitat use (Byström et al. 2003). A second resource shift to larger prey, i.e. fish, takes place between 90 and 240mm (Hjelm et al. 2000). Since body size is often inversely related to mortality, fish need to optimize growth during their early life stages. In order to avoid predators, juvenile perch are often highly associated to habitats that provide shelter (Persson & Eklöv 1995, Kjellman et al. 1996, Sandström & Karås 2002b). Temperature has been shown to affect the start of the spawning season and also to have a strong influence on growth and development of later life stages in the species (Thorpe 1977, Karås & Hudd 1993, Karås 1996a, 1996b, Sandström et al. 1997).

Northern pike is, just like perch, a common species in European fresh and brackish waters. The distribution of the species in the Baltic resembles that of perch, i.e. it is distributed along most of the Baltic coast with the exception of the most saline areas where they only occur in estuaries. The species is particularly important for the recreational

fishery and it was estimated to be the single most important species for this type of fishery in Sweden. Similar to perch the commercial catch of the species has declined in several areas along the Swedish coastline, most likely due to unsuccessful recruitment from the early 1990s and onwards. As the biomass of pike cannot be appropriately measured by gill-nets, the relative importance of the species is hard to quantify. Measuring from the commercial and recreational catch statistics it is probably one of the most abundant predator fishes in littoral areas of the Baltic. The species is a highly specialised piscivore and its diet is constituted by a large range of small- and medium sized fish species.

Spawning of northern pike normally starts in the end of April or early May in shallow sheltered areas where the water is heated more rapidly in early spring (Karås & Hudd 1993, Gillet & Dubois 1995, Nilsson 2006). Pike eggs have a sticky outer layer that adhere the eggs to the spawning substrate. Eggs have been observed both on bare bottom and vegetation, with higher densities on vegetation (Wright & Shoesmith 1988). The eggs are laid over a period of approximately 2-5 days and they hatch in ca 120 degree days. Post-hatching larvae remain close to the spawning grounds where they first feed mainly on plankton and macroinvertebrates, but already at a size of 4 to 5 cm start to feed on small fish (Morrow et al. 1997). The distribution and quality of life-stage specific habitats is a central factor determining pike population size. Access to fry and juvenile-adult habitats was shown by Minns et al. 1996 to be more limiting than access to spawning habitat. This pattern was explained by the need for larger spatial areas in older life stages compared to the aggregated spawning sites.

Pikeperch has a more limited distribution in the Baltic Sea than the three other studied species. It rarely occurs in the Gulf of Bothnia and it is more limited by increasing salinity in the southern Baltic than both pike and perch. The species has a high commercial value and it is an important target species for commercial fisheries. It is also of importance for the recreational fishery. The commercial catch of pikeperch has declined in several areas along the Swedish coast, a trend mainly attributed to increased fishing pressure. Similar to pike, the species is a highly specialised piscivorous fish. It mainly occurs in the pelagic zone, where it in some areas is the main predator on pelagic prey fishes.

In the northern Baltic proper, spawning of pikeperch normally starts in late May to early June at temperatures of 8-14° C (Lehtonen et al. 1996, Saulamo & Lappalainen 2006). Spawning mainly takes place on shallow sandy grounds were the male builds a nest in which the female lays the sticky eggs in 3-4 layers, but also other substrates such as stones or roots of vegetation may be utilized (Lehtonen et al. 1996). The male then guards the nest until hatching. Post-hatching larvae stay near the bottom approximately two weeks before shifting to a pelagic habitat. After feeding on zooplankton and crustaceans (mainly mysids) they may start to feed on fish already at a size of 20-30 mm length (Lehtonen et al. 1996), although the observations by Sandström & Karås (2002a) indicated that the diet consisted almost entirely of mysids from 35 to 55-60 mm length, while fish was the dominant food item from 70 mm length. Adult pikeperch in the Baltic can tolerate higher salinities, while early larvae for a short critical period are limited to salinities <4.5-4.75‰ (Olifan 1945 in Russian, c.f. Lehtonen et al. 1996). High turbidity may be the most characteristic feature of coastal juvenile pikeperch habitats. It is well adapted to forage in low light conditions, i.e. turbid eutrophic waters with high primary and secondary production (Sandström & Karås 2002a). An increased turbidity

has also been shown to reduce predation from perch during larval stages (Pekcan-Hekim & Lappalainen 2006). Temperature has been attributed to be the most significant environmental factor affecting pikeperch population dynamics via positive effects on annual abundances, growth, year-class strengths and yields (Lappalainen 2001).

Roach is also, similar to the other three species, one of the most commonly occurring fish species in European fresh and brackish waters. The distribution of the species in the Baltic resembles that of perch and pike, although it is slightly more vulnerable to high salinity which may reduce its distribution area in the southern Baltic Sea to some extent. The species is not of any significant commercial importance. In contrast, it is instead of large ecological importance. Roach is an important prey fish for many predator fishes as well as for fish-eating birds. The species is the second most abundant fish in the areas monitored by multi-mesh gill-nets in both Swedish and Finnish coastal waters (Ådjers et al. 2006) The abundance of roach has increased the last decades, particularly in the Finnish part of BALANCE pilot area 3, which has lead to an enlargement of the main distribution areas of the species to more exposed parts of outer archipelago areas where it only occurred rarely in the past (Lappalainen 2002). This increase in biomass of roach has been shown to influence the population size and dynamics of the blue mussel (*Mytilus edulis*) in Finnish coastal waters (Lappalainen et al. 2004). Roach is an omnivore and a resource generalist that mainly feeds on zooplankton and benthic invertebrates but occasionally may incorporate plants and algae in their diets.

Roach spawn in spring, predominantly from May to June, normally in shallow sheltered areas on vegetated substrates. The larvae hatch at a size of 6-7 mm and stay in the littoral areas after hatching. Roach larvae and roach juveniles rarely move into the pelagic zone, but they may nevertheless slowly migrate to other littoral areas after hatching. It has been shown that the species gradually spreads to the outer parts of the archipelago areas as there body size increases (Urho 2002). The recruitment of the species benefits from eutrophication and also from increased temperature, albeit not as pronounced as the other three studied species.

4.1   Aims

The aim of this study was to produce maps describing the distribution of nursery habitats for perch, pike, pikeperch and roach, and of spawning habitats for perch in the BALANCE pilot area 3, in the northern Baltic proper. Maps describing the distribution of recruitment habitats of the studied species are currently lacking, why protection of these essential fish habitats through marine spatial planning is difficult. Besides being used directly in marine spatial planning by national and regional authorities in Sweden, Åland and Finland, these maps will be used for analyses of the coherence of the Natura 2000 network (WP3) and for development of GIS tools for marine spatial planning (WP4).

5   MATERIAL AND METHODS

5.1   General Characteristics of the Pilot Area

BALANCE pilot area 3 is located in the vast archipelago region that stretches from the counties of Södermanland, Stockholm and Uppsala in Sweden, over via Åland and the Finnish Archipelago Sea (Fig 1). It is a topologically and geologically very heterogeneous area that consists of numerous islands and smaller islets, and habitat patchiness is thus normally high even on smaller spatial scales. The bottom substrate normally consists of a complex mix of soft and hard substrates. Modelling this patchy distribution of habitats is a challenge, and to succeed in making accurate habitat maps detailed data on/maps of the physical environment is needed.
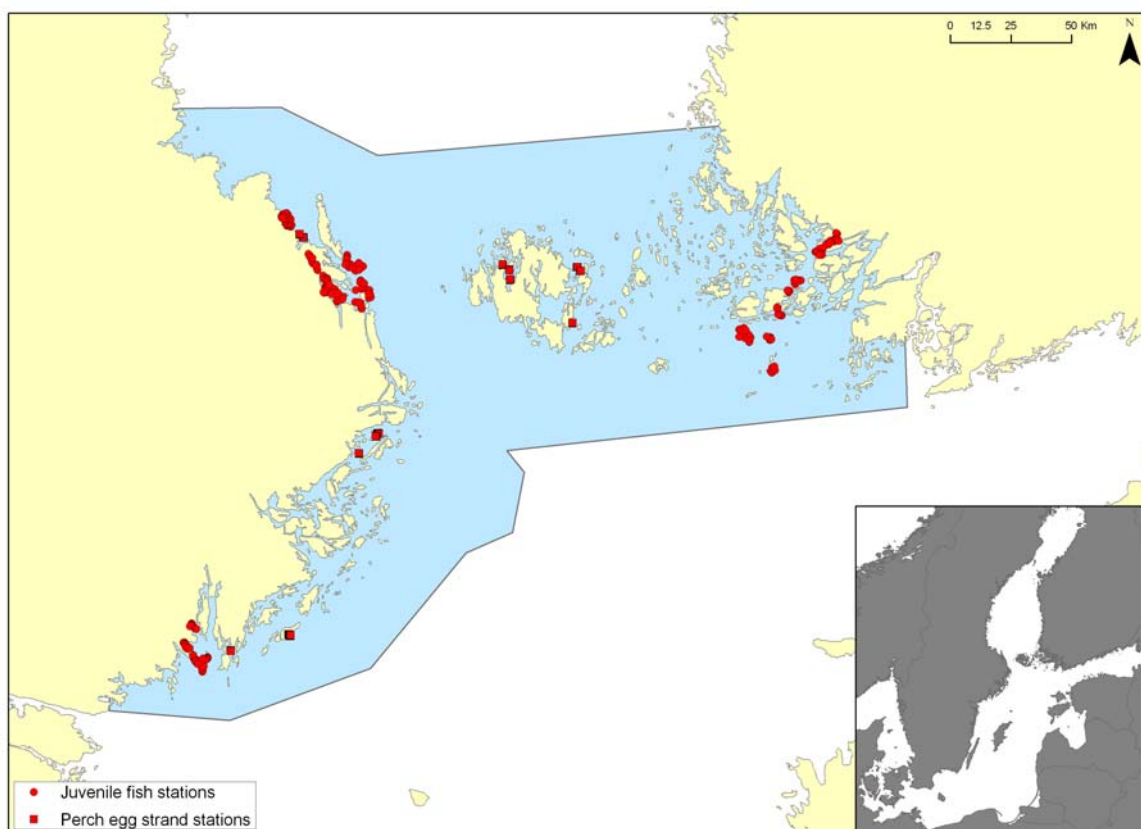

Figure 1. Sampling stations within pilot area 3.

The northern Baltic proper is a non-tidal, brackish water body. There are strong salinity gradients in the area, both in a north-south direction as well as from the inner parts of the archipelago to the outer parts. The salinity in the outer archipelago varies from around 5 psu in the northern parts to around 7 in the south. In the innermost bays and fladas the salinity may be as low as 3-4 psu, and occasionally even lower, depending on the inflow of freshwater.

A substantial influence on the landscape of the region is caused by the land-uplift of 4-6 mm per year, which continuously changes the landscape. Due to this slow geological process, shallow coastal areas either turns into land or into wetlands or small lakes, eventually entirely separated from the sea. Land-uplift thus slowly alters the prevailing geomorphology of a specific site with subsequent changes in wave-exposure, water depth and water exchange with the open sea. This, in turn, influences a number of parameters that may structure aquatic communities, such as salinity, temperature, nutrient levels, sediment characteristics and ice conditions (Munsterhjelm 2005).

Water temperature and ice cover also vary between the inner and outer parts of the archipelago. Maximum summer water temperature range is normally 15-20 ºC in open water and around 25 ºC in sheltered shallow areas. The most dramatic gradients in temperature regime normally occur in spring - early summer when temperature differences between sheltered shallow areas and the open sea may exceed 10 ºC. Shallow and sheltered areas on the other hand cool off more rapidly in autumn than the open sea. Since fish are very sensitive to temperature changes, temperature variations may have a large influence on fish migrations within the area. Many coastal fish species, such as pike, perch and pikeperch need warm water for reproduction, and mainly utilise shallow, sheltered areas in the archipelago as nursery areas. Also fish species like turbot, flounder, whitefish and herring utilize the productive coastal areas for spawning. The adults of these species prefer cold water, and therefore migrate to the outer parts of the archipelago and/or to deeper water layers when the water in the inner parts gets too warm in summer.

Shallow coastal areas are normally characterised by soft sediment bottoms in sheltered areas with a macrophyte community dominated by habitat structuring species such as pondweeds (*Potamogeton* spp.), stoneworts (*Chara* spp.), milfoils (*Myriophyllum* spp.) and emergent reed, (*Phragmites australis* L.). When wave exposure increases, substrates go from soft to a mix of soft and hard. In exposed parts of the archipelago hard substrates dominate at shallow sites. These areas are often dominated by bladderwrack (*Fucus vesiculosus* L.), and in the northern parts of the area by the closely related endemic *Fucus radicans*, as the main habitat forming species. The maximum depth for vegetation differs depending on light attenuation, ranging from around 2 to 10 m, with the shallowest maximum depth in turbid inner bays. There are large variations in nutrient runoff from land, with the highest loadings in the densely populated parts of the archipelago, e.g. the Stockholm and Turku areas. The variability in nutrient loadings in combination with large differences in water turnover time result in strong turbidity gradients on both large (km) and small (m) scales.

5.2   *Field sampling*

Sampling of juvenile fishes was conducted in late July-August 2005-2006. Juvenile fish were monitored by point abundance sampling with small detonations that stun small fish within an area of ca. 60 m^2 (evaluated in Snickars et al. 2007). This method allows quantitative sampling of fish (15-150 mm) with well-developed swim bladders in all shallow habitats, including dense vegetation. All stunned individuals (floating and sinking) were collected via snorkelling for later counting of the number

of individuals, determination of species and length measurements. The distribution of sampling sites was stratified along wave exposure and archipelago zonation gradients, in order to cover the whole ranges of distribution of the studied species (Fig. 1).

Surveys of perch egg strands were conducted three times during a period from late April to mid June, with intervals of 14-20 days in 2003. The survey was conducted by snorkelling along parallel transect lines (length 20-480 m, 4-8 lines per site) drawn perpendicular to the length axis of each site from one shore to the opposite shore until the entire site was covered. All visible egg strands within one metre on both sides of the transect lines were registered. Totally 22 000 m^2 were surveyed for eggs covering 22 studied sites. The majority of the sites were shallow inlets and selected from a previous large-scale survey of inlets. Selection criteria included a minimum level of anthropogenic disturbance and a considerate but not extreme variation in geomorphometry and wave-exposure. Sites were spread over a relatively large geographical area in order to study general trends within the Baltic Sea (Fig. 1)

In both the surveys of juvenile fish and perch egg strands, representative information on environmental variables such as depth, temperature, turbidity, substrate and vegetation were collected for all sampling points.

The surveys were conducted by the Swedish Board of Fisheries in collaboration with the Foundation for Uppland, the Finnish Game and Fisheries Research Institute, Metsähallitus in Finland and Åbo Akademi University.

5.3   *Predictor layers*

Few environmental predictor variables were available as continuous maps for the whole BALANCE pilot area 3. Only wave exposure (Isæus 2004) and depth from sea charts were considered to have a spatial coverage, resolution and accuracy suitable for the fine-scale fish habitat modelling. In some areas the depth information is very coarse, mainly due to military restrictions affecting the access to the existing data, which in these areas substantially limits the usefulness of this variable for these fine-scale modelling purposes.

Wave exposure and depth are important predictors for fish distribution, but further variables are needed to obtain high quality habitat maps. One important predictor variable for fish distribution that was not available at a sufficient resolution was water clarity. To cover this gap, a GIS model of Secchi depth was developed. The index was based on distance from the base line and wave exposure. Evaluation of the index was made against 293 in-situ measurements of Secchi depth (range 0.3-10 m), from different years but the same season. There was a significant positive correlation between the index and field measurements of Secchi depth, but the index only explained a minor part of the variation in the level of Secchi depth (r^2=0.2911). The input data was from different years, thus the model cannot be expected to be very accurate, since water turbidity is highly variable even at short time scales. The proxy, however, still captured the gradients in water clarity that are found at small spatial scales from small sheltered bays to open areas, which are important for the distribution of juvenile fish, very well, and was therefore considered to be valuable in the modelling.

All layers were in ESRI raster format with 25 meter cell size and in UTM34N projection. Depth was limited to maximum 6 metres depth, as the data from the fish surveys was available down to this depth. The Wave exposure index used was a log10-transformation of data produced using WaveImpact (Isæus 2004). The proxy on water clarity was calculated using water distance from the base line (connecting the outmost islands and thereby defining the archipelago zone) and the wave exposure index, using the equation:

water clarity proxy = logwaveexposure/logdistance ^ 0.5.

5.4 Modelling methods

Using Hawth's tool in ArcGIS, data from each predictor layer was extracted for the fish sampling positions. GAMs were used to model the probability of occurrence for young-of-the-year of perch, pike, pikeperch and roach, as well as for perch egg strands (perch spawning). Subsequently, presence/absence of fish juvenile and perch egg strands were used as input in the models. Modelling was conducted in S-PLUS using the GRASP work package. All predictor variables were forced into binomial models with 3 degrees of freedom to fit the spline function. No weights on prevalence were applied since this has been shown to overestimate the probability of presence when making spatial predictions (Maggini et al. 2006). All species were run simultaneously when modelling the juvenile fish, while the perch egg strand model was run separately.

The models were used for producing spatial predictions in GIS using script provided from the GRASP work package. The extent of the predictions was the whole pilot area 3, in areas shallower than 6 m depth. All predictor rasters were in 25 metre resolution, as was the resulting maps, showing probability of presence. ESRI Spatial Analyst was used to reclassify the continuous probability of presence predictions into dichotomized maps of suitable and unsuitable habitat for each response variable. The threshold for habitat suitability was determined using the true-skill statistic TSS, (Allouche et al. 2006), where the sum of specificity and sensitivity is maximised. There are many methods for determining thresholds of presence, and TSS has the advantage of not being affected by the prevalence or the size of the validation set (Allouche et al. 2006). Thus, two different sets of grids were made. The first showing continuous probability of presence for each species/life stage, and the second set showing categorized habitat in suitable/unsuitable habitat.

Additional explanatory Y-O-Y models were built using the same data set with the addition of total vegetation cover (%) in order to show differences in modelling potential with alternative predictor variables.

Model performance was evaluated by an analysis of deviance, which is equivalent to variance analysis in general linear models. The overall test statistic is called D^2 and is a measure of goodness-of-fit for the overall model. The increased flexibility of these models can however lead to overdispersed errors. Over dispersion is characterised by largely inflated residual deviance, which was examined in all models. A general recommendation is that one should rely more on empirical evaluation rather than D^2, due to a tendency of overfitting the calibration data (Guisan et al. 1999). Model evaluation was therefore also based on receiver-operating-characteristic (ROC) plots.

ROC-plots are obtained by plotting all sensitivity values on the Y axis and (1-specificity) on the X axis for all available thresholds on the X axis (DeLeo 1993, Fielding & Bell 1997). ROC-plots give an area-under-curve (AUC) value that range between 0.5 and 1. An AUC-value of 1 indicates no overlap between the two group distributions, i.e. true positives and false positives. A value of 0.75 shows that 75% of the time a random selection from the positive group will have a score greater than a random selection from the negative group (DeLeo 1993, Fielding & Bell 1997). ROC (AUC) has been recommended as a measure of accuracy since it is insensitive to the response variables prevalence (McPherson et al. 2004). To aid model evaluation in comparative studies it has been recommended that ROC and sampling prevalence should be reported (McPherson et al. 2004).

The built-in validation procedure with ROC-plots in GRASP produces both a ROC-plot for the entire data set as well as a cross-validation with subsets of the data to compare predicted versus observed values. All models were validated using 5 groups in the cross-validation.

5.5 Results

The model for pikeperch was the most accurate (Fig. 4; ROC = 0.90), and the models for perch spawning, pike Y-O-Y and roach Y-O-Y also performed reasonably well (Fig. 5, 6 and 8; ROC = 0.75-0.81). The model for perch Y-O-Y proved less accurate (Fig. 7; ROC = 0.66). This model also showed a tendency for over dispersion (Df = 296 total; Residual deviance = 378.9644), indicating that there is some unexplained spatial heterogeneity in the data. All the other models were well within limits.

The potential contribution from each predictor variable, calculated by creating new models with only one predictor, showed that wave exposure and Secchi depth contributed most to the performance of the models while depth did not contribute as much to model strength (Fig. 2a). Adding vegetation coverage would probably improve model performance substantially, especially when modelling the distribution of perch and pike Y-O-Y (Fig. 2b). However, as vegetation was not available as a continuous map, no fish habitat maps using vegetation coverage as a predictor could be produced.
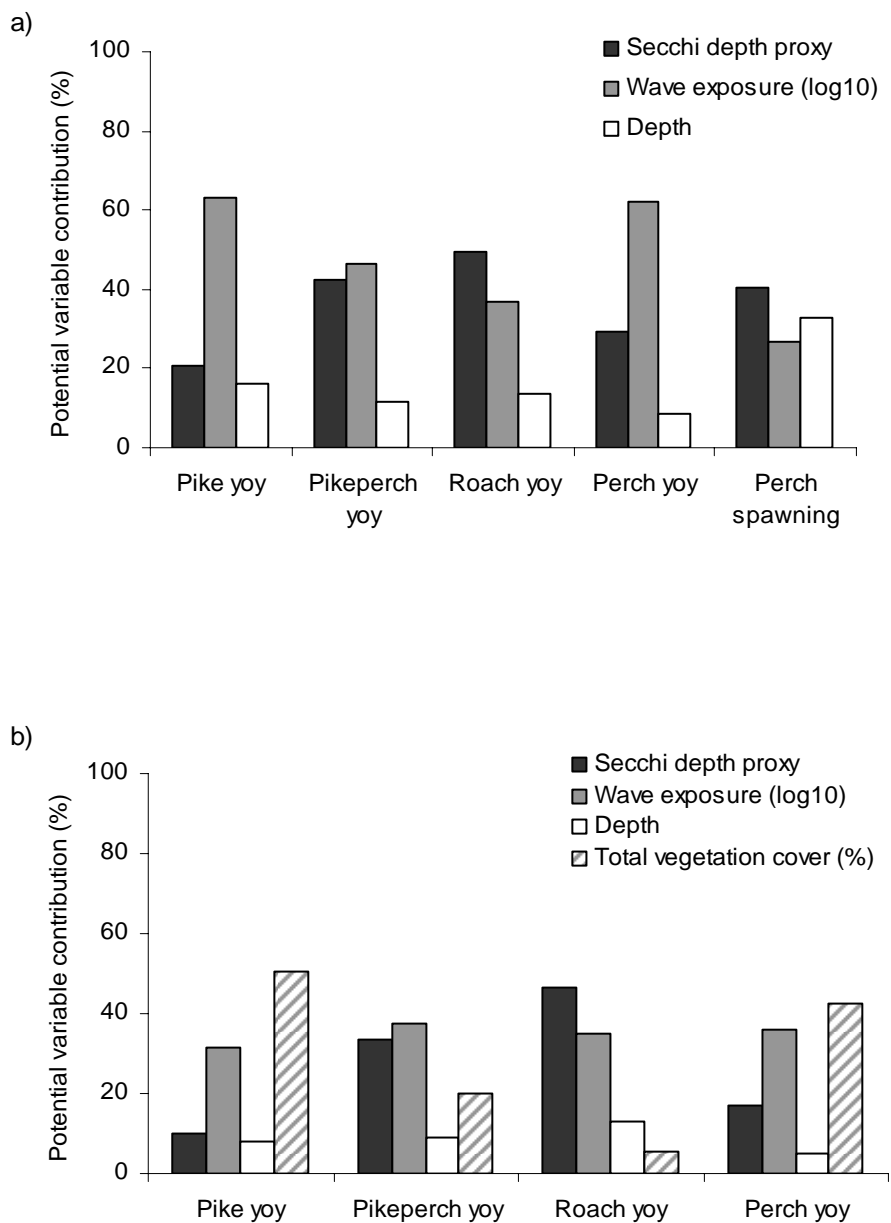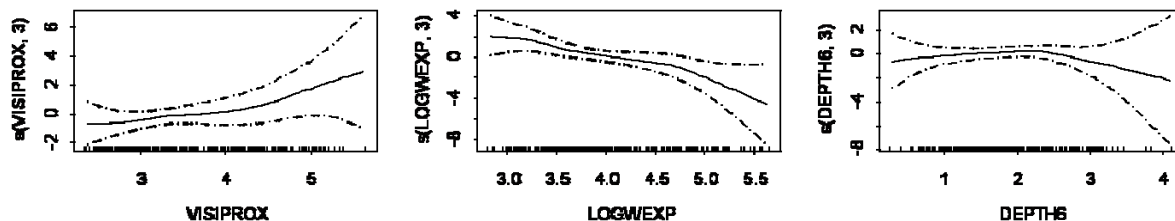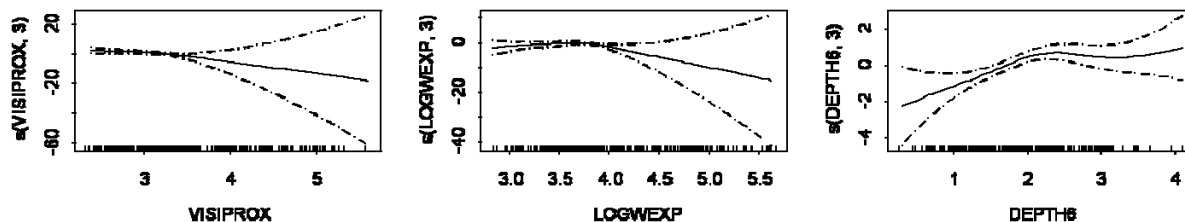
a)

b)

Figure 2a-b. Potential contribution of each predictor variable alone, a) for the environmental factors used in the map predictions and b) when adding an additional environmental parameter, total vegetation coverage, to the statistical models.
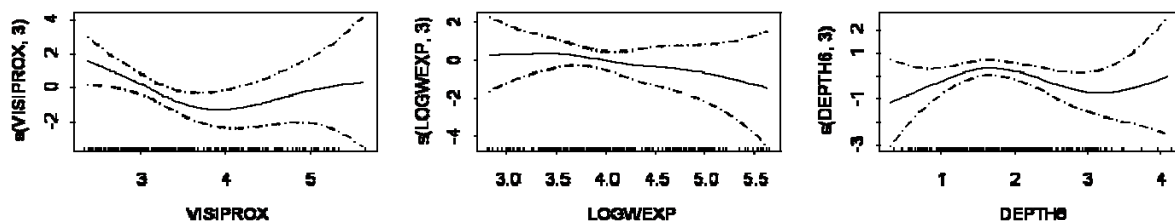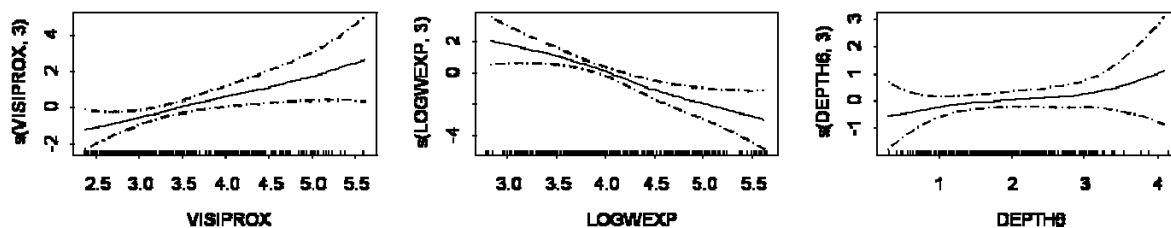
a) Pike young-of-the-year

b) Pikeperch young-of-the-year

c) Roach young-of-the-year
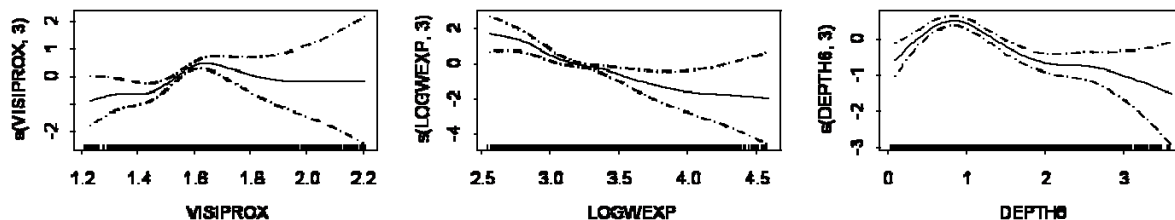
d) Perch young-of-the-year

e) Perch spawning

Figure 3 a-e. Partial response curves for the GAMs of the species and life stages modelled. The y-axis represents the response variable in the linear predictor scale. Visiprox denotes the Secchi depth proxy, logwexp the logarithm of the wave exposure index, and depth6 water depth. Dotted lines indicate twice point standard errors and the dots on the x-axis represent the samples along each predictor variable gradient.

The partial response curves for the GAMs illustrate how each explanatory variable affects the distributions of the species and life stages modelled. For pike Y-O-Y,

there was a positive effect of increased water clarity, a negative effect of increased wave exposure and a slight negative effect of depths >3 metres (Fig. 3a). For pike-perch Y-O-Y, there was a negative effect from increased water clarity and wave exposure, while depth had little effect (Fig. 3b). In the roach Y-O-Y model, water clarity had the highest contribution, with a positive effect of low clarity. There was a negative effect of increased wave exposure and a tendency of preferred depth around 2 metres depth. For the perch Y-O-Y model, there was a clear positive effect of increased water clarity and a negative effect of increased wave exposure, while the effect of depth was not as evident. In the perch spawning model, water clarity had a positive effect, and wave exposure a negative effect. Depth had a negative impact from an optimum depth around 1 metre.

5.6   Validation

Model evaluation statistics for the probability of presence predictions, i.e. D^2, AUC for both cvROC and ROC as well as prevalence, are given in table 1. These statistics show that both accuracy and level of generalisation was highest for the pikeperch model, intermediate for the perch spawning and roach models and lowest for the pike and perch Y-O-Y models.

The evaluation of the categorised presence-absence maps, i.e. the TSS scoring based on an error matrix, gives a similar picture. The highest scores were found for the pikeperch and roach Y-O-Y and perch spawning habitat maps, while pike and perch Y-O-Y maps were less accurate. The misclassification rates were 30-38% in all models.

Table 1. Model evaluation statistics for the five recruitment habitat models.

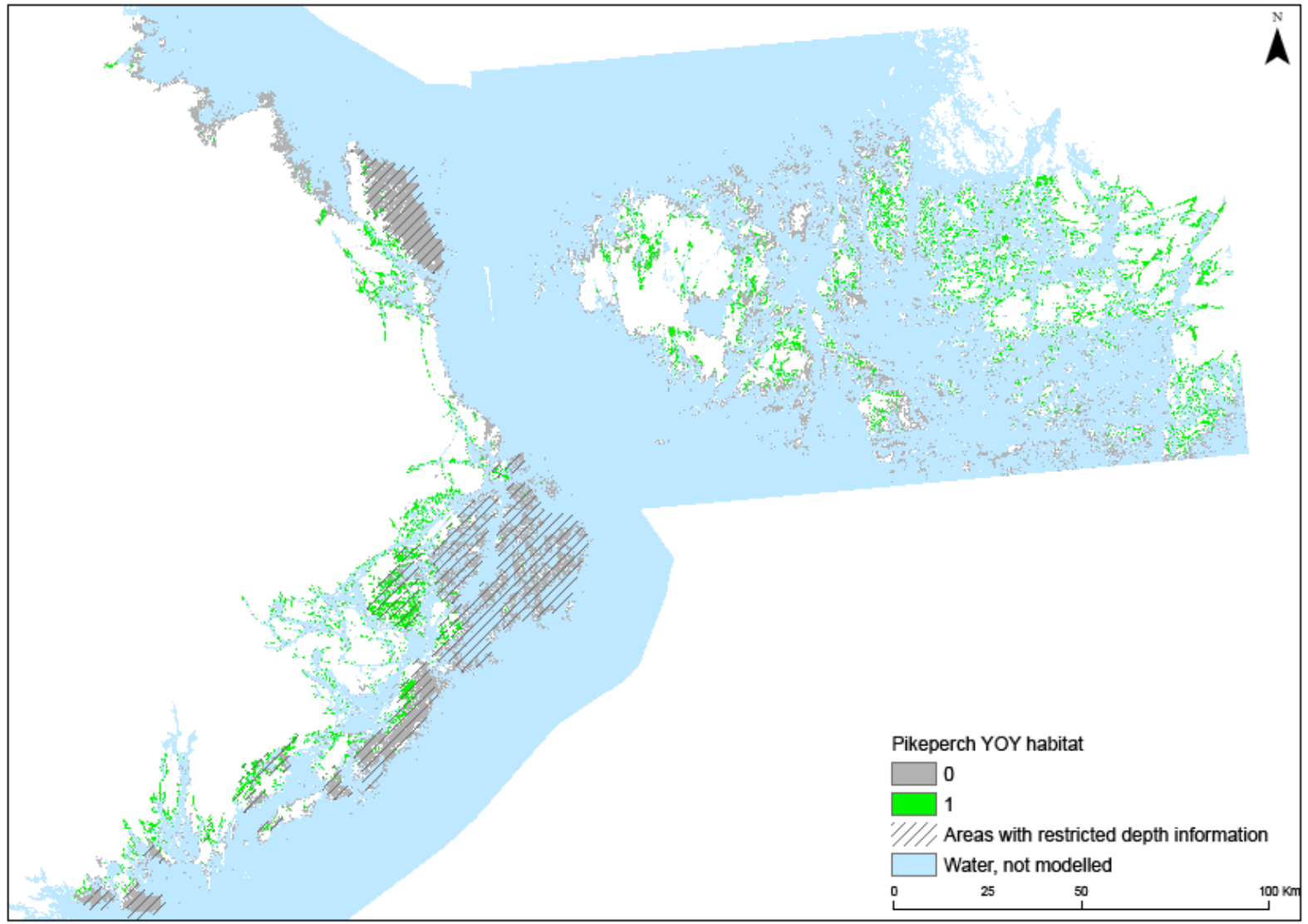| Species / Statistics	D^2	cvROC	ROC	Prevalence indata	TSS, catego-rised maps
Pike Y-O-Y	0.11	0.64	0.75	0.08	28.6
Pikeperch Y-O-Y	0.34	0.85	0.90	0.09	34.6
Roach Y-O-Y	0.19	0.74	0.81	0.13	39.6
Perch Y-O-Y	0.06	0.61	0.66	0.43	23.9
Perch spawning	0.12	0.77	0.78	0.08	35.0

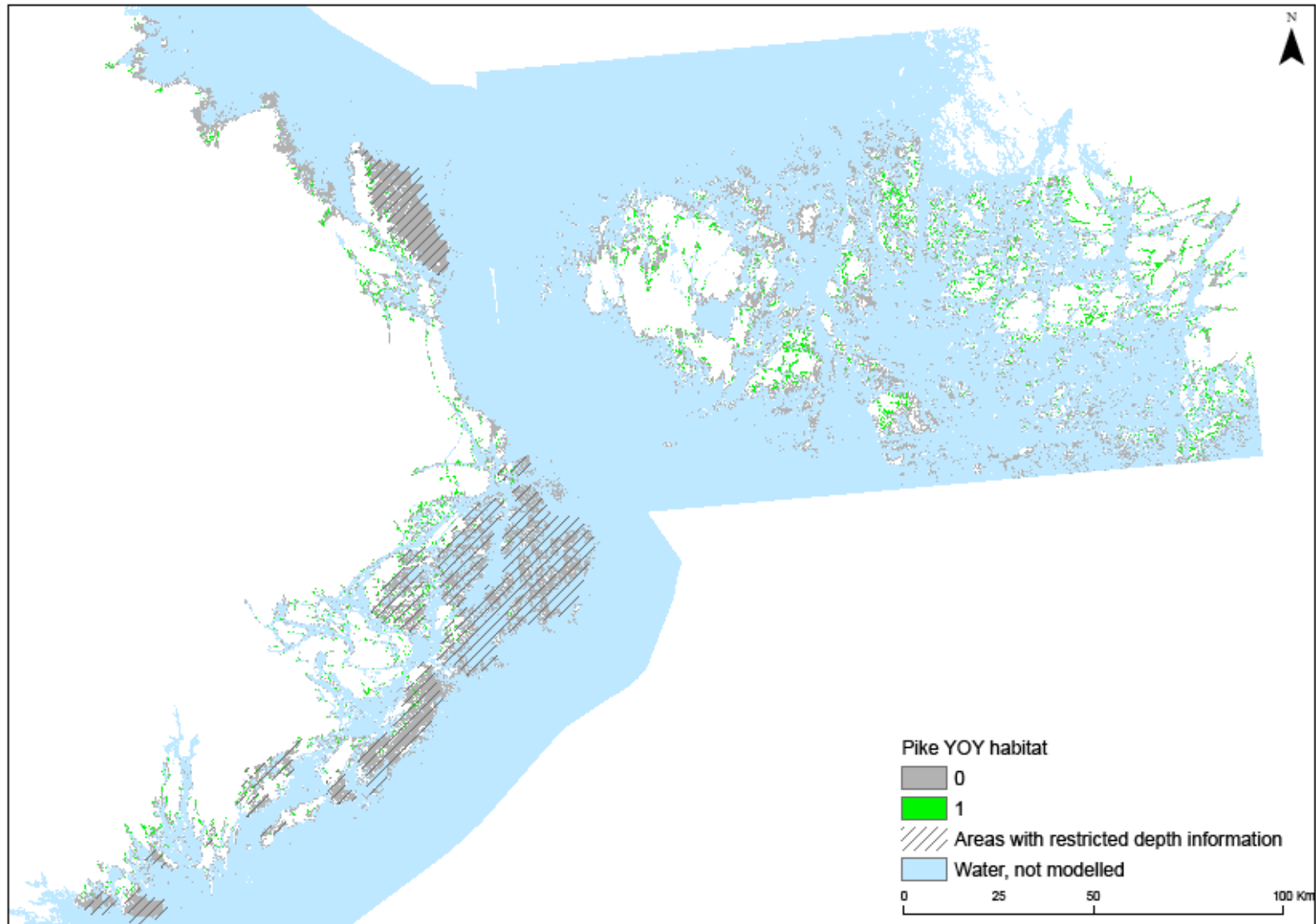*Figure 4. Pikeperch young-of-the-year habitats.*

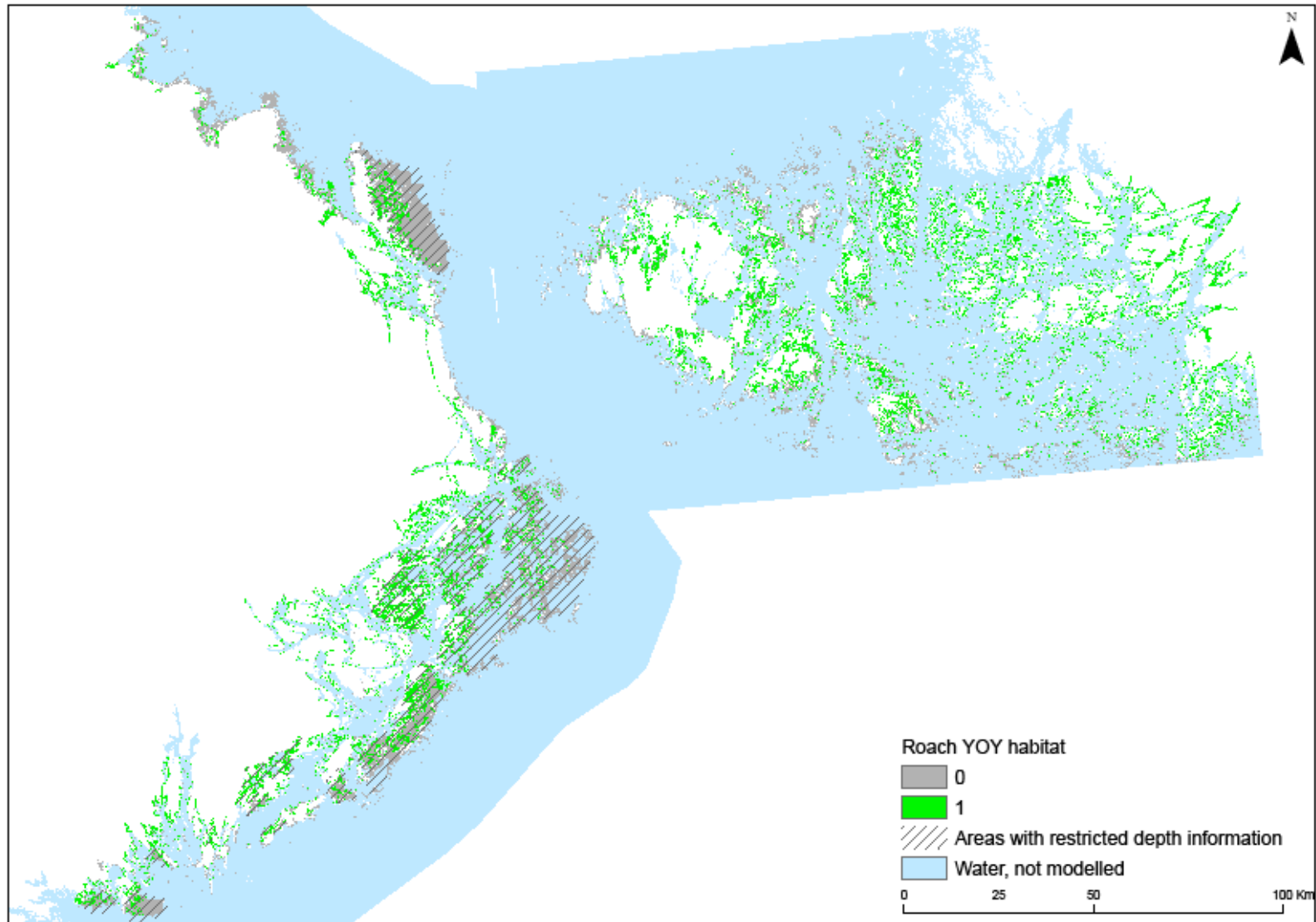*Figure 5. Pike young-of-the-year habitats.*

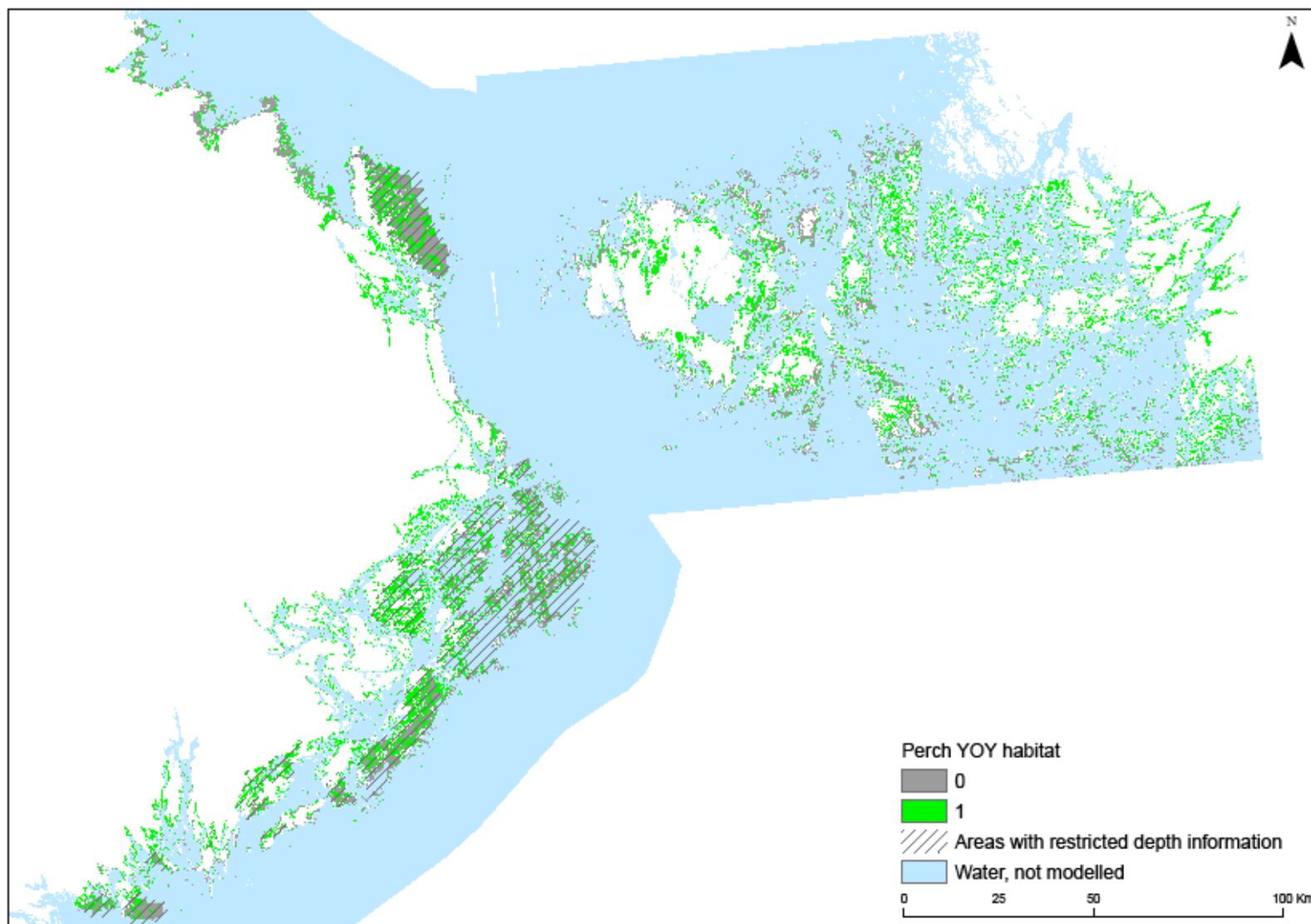*Figure 6. Roach young-of-the-year habitats.*

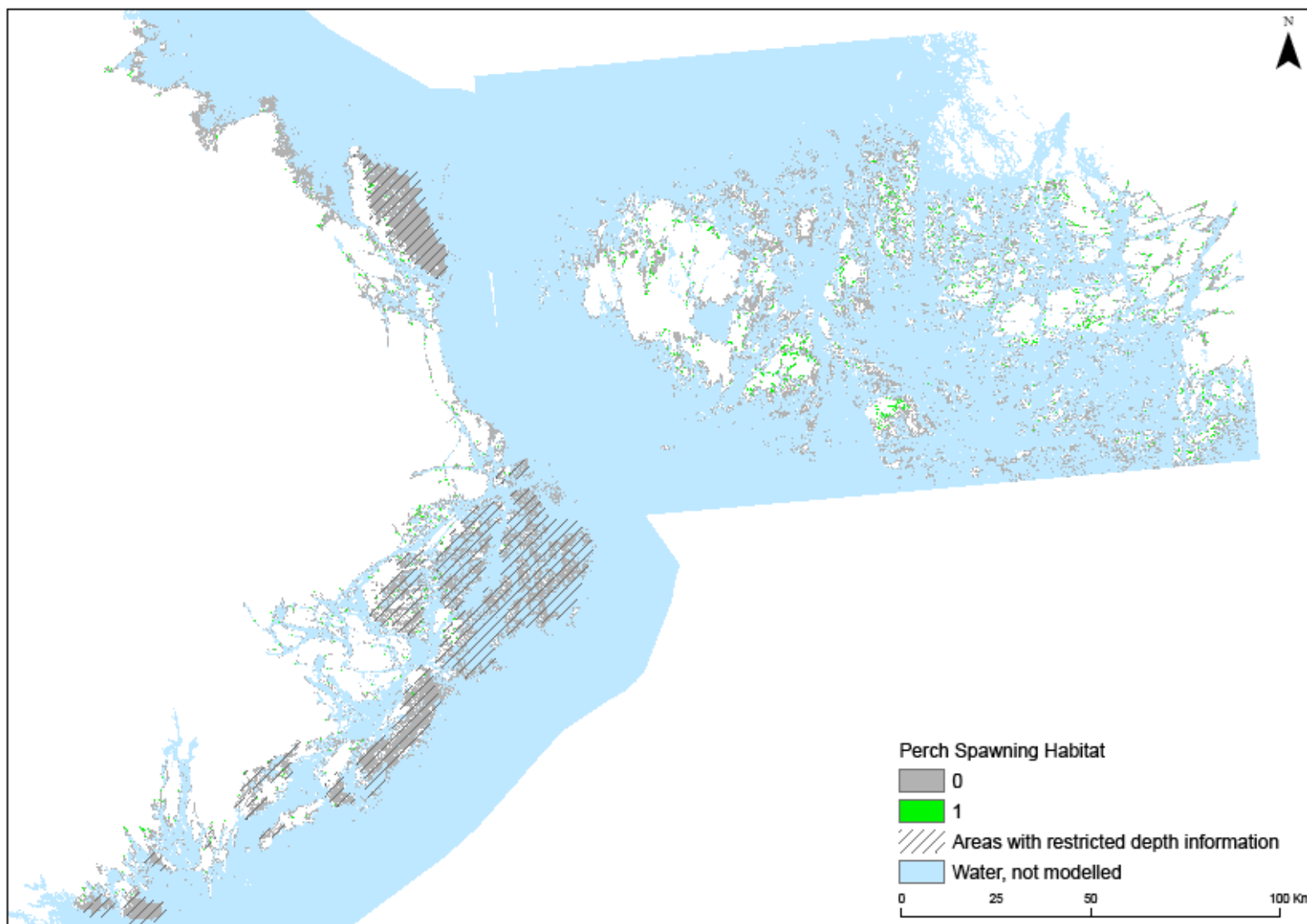*Figure 7. Perch young-of-the-year habitats.*

Figure 8. Perch spawning habitats.

6   DISCUSSION

We have used three relatively easily obtainable GIS-layers to produce large scale habitat maps for four species and two life stages in BALANCE pilot area 3. For all the investigated species/life stages, except for perch Y-O-Y, the resulting statistical models were relatively strong, especially considering the high resolution of the predictions in relation to the large extent of the study. Our results show that the predictor variables used are important components when determining the habitats of these species, and that GIS-modelling could develop into an indispensable tool in large-scale mapping of essential fish habitats.

Two of the predictor variables, wave exposure and the visibility proxy, are completely GIS-derived and can be considered as indirect variables (Austin 2002). Still, in all Y-O-Y-models they contribute the most in explaining species distribution of the variables tested. Depth can also mainly be characterised as an indirect variable encapsulating other more direct variables such as light (attenuated with increased depth), temperature, water movement (decreasing with depth) and vegetation. Depth has the least influence on the models, not because it is unimportant for determining species distributions, but rather because no predictions were made deeper than 6 m. The low impact of depth in the models only shows that there are no large differences in habitat quality between 0 and 6 m depth – deeper areas are certainly less suitable to these young life stages, but the field data covered areas only down to 6 m.

Information on water turbidity for the archipelago area between Sweden and Finland is also available only at a resolution too coarse for the kind of habitat modelling undertaken in this study. The GIS-derived visibility proxy was reasonably successful in detecting small-scale gradients in Secchi depth, and proved to be an important predictor variable in all models. For future modelling work, an alternative approach to attaining large-scale maps of turbidity could be to use satellite imagery. In another study within BALANCE, covering two smaller areas of pilot area 3, it has been shown that turbidity can be accurately interpreted from SPOT 5 images, at a resolution of only 10 m (Bergström et al. 2007).

Comparing explanatory models based on the three variables used to produce the predictions with additional data on total vegetation coverage, a more direct predictor variable, showed that both Y-O-Y pike and perch distributions also are strongly governed by vegetation coverage (Fig. 2b). Vegetation adds habitat complexity and can be important both as a refuge against predators as well as a host to many prey animals (Persson & Eklöv 1995). The selection of spawning sites by perch is also known to depend largely on vegetation type (Thorpe, 1977; Treasurer, 1983). Thus, including vegetation coverage as a predictor variable would most likely increase the predictive power of the models. Producing high-resolution, large scale GIS-layers of vegetation coverage is therefore an important step towards increasing the precision of many fish habitat models.

Statistical modelling using GAM and the GRASP work package proved to be a flexible and accessible technique for describing species-habitat relationships. A potential disadvantage of GAMs is that they do not allow for taking interactions between predictor variables into account, due to their additive structure. There is however ways of over-

coming this limitation, which should be further, explored. For example, interaction terms may be added manually to a model, by simply creating interaction terms e.g. by multiplying two predictors and adding the result as a separate predictor. Another approach for adding interaction terms can be found in Maggini et al. (2006). A regression tree can be fitted on the residuals of a first model, where after the branches/leaves of the resulting regression tree is used to classify each sample. Then a new model is built using the resulting classification as an additional predictor. This procedure is a way of incorporating interactions between all variables in a single term.

The map predictions are based on life-stage specific relationships from a limited number of study sites and a limited set of environmental variables, which in a GIS have been re-calculated to show probability of occurrence. Conceptually, the maps therefore show the potential distribution of the modelled life-stages based on the environmental maps, rather than the true distribution. For juvenile pikeperch, the habitat map appears to overestimate the distribution of the species. This pattern may be an effect of limitations in earlier life stages, for example in access to suitable spawning habitats. Thus, areas lacking spawning sites will naturally also lack Y-O-Y fishes even though suitable habitats for juveniles are abundant. An interesting application of this kind of modelling work may thus be to identify habitat bottlenecks, as well as regions where habitat degradation has had negative effects on fish stocks.

7   PERSPECTIVES

There is a high demand for detailed maps of essential fish habitats for a range of physical planning activities. The habitat maps presented in this report are already used by several regional authorities, for example in fisheries restoration and management plans and in the design and zonation of forthcoming MPAs.

So far, few studies exist that use statistical modelling coupled with GIS for large-scale mapping of essential fish habitats. Based on our experiences so far we believe that this approach will become widely used in the future, and that these habitat maps will probably become a central constituent in marine spatial planning. Both techniques and data basis are in a phase of rapid development, and both the spatial coverage and the accuracy of the maps can therefore be expected to increase steadily.

The success of fish habitat modelling initiatives will, however, not only depend on the understanding of the dynamics of fish populations and their reaction to environmental variables, but also on the accuracy of the maps of the environmental variables that the predictions are based upon. Currently, a lack of high-resolution maps of for example bathymetry, surface sediments, hydrography, and in the case of young fishes, vegetation coverage, is limiting the production of accurate habitat maps. For bathymetry, this deficiency may be alleviated for example by opening access to classified maps, and by easing restrictions on collection and usage of bathymetric data. For other environmental variables, and for bathymetry in some areas, additional high-resolution mapping is needed. Development of new techniques, such as remote sensing for identification of coastal habitat characteristics (Bergström et al. 2007), as well as GIS-modelling techniques similar to those used within this BALANCE report, may provide efficient tools for producing high-resolution maps at reasonable costs.

8   REFERENCES

Allouche, O., A. Tsoar, and R. Kadmon. (2006). *Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS).* Journal of Applied Ecology 43:1223-1232.

Austin, M. P. (2002). *Spatial prediction of species distribution: an interface between ecological theory and statistical modelling.* Ecological Modelling 157:101-118.

Bello, P. J., L. V. Rios, C. M. A. Liceaga, M. C. Zetina, C. K. Cervera, B. P. Arceo, and N. H. Hernandez. (2005). *Incorporating spatial analysis of habitat into spiny lobster stock assessment at Alacranes reff, Yucatan, Mexico.* Fisheries research 73:37-47.

Bergström, U., Evertson, J., Karås, P., Sandström, A., Sundblad, G. and Wennberg, S. (2007). *Evaluation of satellite imagery as a tool to characterise shallow habitats in the Baltic Sea.* Balance interim report no. 5.

Byström, P., L. Persson, E. Wahlström, and E. Westman. (2003). *Size- and density-dependent habitat use in predators: consequences for habitat shifts in young fish.* Journal of Animal Ecology 72:156-168.

Carroll, C., W. J. Zielinski, and R. F. Noss. (1999). *Using presence-absence data to build and test spatial habitat models for the fishes in the Klamath region, U.S.A.* Conservation Biology 13:1344-1359.

Chatfield, C. (1995). *Model uncertainty, data mining and statistical inference.* Journal of the Royal Statistical Society, Series A 158:419-466.

DeLeo, J. M. (1993). *Receiver operating characteristic laboratory (ROCLAB): Software for developing decision strategies that account for uncertainty.* Pages 318-325 in Uncertainty Modeling and Analysis, 1993. Proceedings., Second International Symposium on, College Park, MD, USA.

Eastwood, P. D., G. J. Meaden, and A. Grioche. (2001). *Modelling spatial variations in spawning habitat suitability for the sole (Solea solea) using regression quantiles and GIS procedures.* Marine Ecology Progress Series 224:251-266.

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

Francis, M. P., M. A. Morrison, J. Leathwick, C. Walsh, and C. Middleton. (2005). *Predictive models of small fish presence and abundance in northern New Zealand harbours.* Estuarine, Coastal and Shelf Science 64:419-435.

Gillet, C., and J.-P. Dubois. (1995). *A survey of the spawning of perch (Perca fluviatilis), pike (Esox lucius), and roach (Rutilus rutilus), using artificial spawning substrates in lakes.* Hydrobiologia 300/301:409-415.

Guisan, A., T. C. Edwards Jr., and T. Hastie. (2002). *Generalized linear and generalized additive models in studies of species distributions: setting the scene.* Ecological Modelling 157:89-100.

Guisan, A., and W. Thuiller. (2005). *Predicting species distribution: offering more than simple habitat models.* Ecology Letters 8:993-1009.

Guisan, A., S. B. Weiss, and A. D. Weiss. (1999). *GLM versus CCA spatial modeling of plant species distribution.* Plant Ecology 143:107-122.

Guisan, A., and N. E. Zimmerman. (2000). *Predictive habitat distribution models in ecology.* Ecological Modelling 135:147-186.

Hastie, T. J., and R. J. Tibshirani. (1990). *Generalized Additive Models.* Chapman & Hall.

Hjelm, J., L. Persson, and B. Christensen. (2000). *Growth, morphological variation and ontogenetic niche shifts in perch (Perca fluviatilis) in relation to resource availability.* Oecologia 122:190-199.

Isæus, M. (2004). *Factors structuring Fucus communities at open and complex coastlines in the Baltic Sea.* Stockholm University, Stockholm.

Johnston, T. H. 1924. *The relation of climate to the spread of prickly pear.* Transactions of the Royal Society of South Australia 48:269-295.

Justice, A. C., K. E. Covinsky, and J. A. Berlin. (1999). *Assessing the generalizabilityof prognostic information.* Annals if internal medicine 130:515-524.

Karås, P. (1996a). *Basic abiotic conditions for production of perch (Perca fluviatilis L.) young-of-the-year in the Gulf of Bothnia.* Annales Zologici Fennici 33:371-381.

Karås, P. (1996b). *Recruitment of perch (Perca fluviatilis L.) from Baltic coastal waters.* Archiv für hydrobiologie 138:99-121.

Karås, P., and R. Hudd. (1993). *Reproduction areas of fresh-water fish in the Northern Quark (Gulf of Bothnia).* Aqua Fennica 23:39-49.

Kjellman, J., R. Hudd, and L. Urho. (1996). *Monitoring 0+ perch (Perca fluviatilis) abundance in respect to time and habitat.* Annales Zologici Fennici 33:363-370.

Lappalainen, J. (2001). *Effects of environmental factors, especially temperature, on the population dynamics of pikeperch (Stizostedion lucioperca (L.)).* Doctorate thesis, University of Helsinki, Helsinki.

Lappalainen, A. (2002). *The effects of recent eutrophication on freshwater fish communities and fishery on the northern coast of of Finland, Baltic Sea.* PhD-thesis, University of Helsinki, 74 pp.

Lappalainen, A., Westerbom, M. and Vesala, S. (2004). *Blue mussels (Mytilus edulis) in the diet of roach (Rutilus rutilus) in outer archipelago areas of the western Gulf of*

Finland, Baltic Sea. Hydrobiologia 514:87-92. H. Kautsky & Snoeijs (eds), Biology of the Baltic Sea.

Lehmann, A. (1998). *GIS modeling of submerged macrophyte distribution using Generalized Additive Models*. Plant Ecology 139:113-124.

Lehtonen, H., S. Hansson, and H. Winkler. (1996). *Biology and exploitation of pikeperch, Stizostedion lucioperca (L.), in the Baltic Sea area.* Annales Zologici Fennici 33:525-535.

Lek, S., and J.-F. Guegan. (1999). *Artificial neural networks as a tool in ecological modelling, an introduction.* Ecological Modelling 120:65-73.

Maggini, R., A. Lehmann, N. E. Zimmerman, and A. Guisan. (2006). *Improving generalized regression analysis for the spatial prediction of forest communities.* Journal of Biogeography 0:1-21.

Minns, C. K., R. G. Randall, J. E. Moore, and V. W. Cairns. (1996). *A model simulating the impact of habitat supply limits on northern pike, Esox lucius, in Hamilton Harbour, Lake Ontario.* Canadian Journal of Fisheries and Aquatic Sciences 53:20-34.

Morrow, J.V., Miller G.L., Killgore, K.J. (1997). D*ensity, Size, and Foods of Larval Northern Pike in Natural and Artificial Wetlands.* North American Journal of Fisheries Management 1997;17:210–214.

McPherson, J. M., W. Jetz, and D. J. Rogers. (2004). *The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact?* Journal of Applied Ecology 41:811-823.

Munsterhjelm, R. (2005). *Natural succession and human induced changes in the soft-bottom macrovegetation of shallow brackish bays on the southern coast of Finland. –* PhD thesis, Walter and Andree de Nottbeck Foundation. Sci. Rep. 26: 1–53 + 4 reprints.

Nilsson, J. (2006). *Predation of northern pike (Esox lucius L.) eggs: a possible cause of regionally poor recruitment in the Baltic Sea.* Hydrobiologia 553:161-169.

Pearce, J., and S. Ferrier. (2000). *Evaluating the predictive performance of habitat models developed using logistic regression.* Ecological Modelling 133:225-245.

Pekcan-Hekim, Z., and J. Lappalainen. (2006). *Effects of clay turbidity and density of pikeperch (Sander lucioperca) larvae on predation by perch (Perca fluviatilis).* Naturwissenschaften 93:356-359.

Persson, L., and P. Eklöv. (1995). *Prey refuges affecting interactions between piscivorous perch and juvenile perch and roach.* Ecology 76:70-81.

Pont, D., B. Hugueny, and T. Oberdorff. (2005). *Modelling habitat requirement of European fishes: do species have similar responses to local and regional environmental constraints?* Canadian Journal of Fisheries and Aquatic Sciences 62:163-173.

Romare, P. (2000). *Growth of larval and juvenile perch: the importance of diet and fish density.* Journal of Fish Biology 56:876-889.

Sandström, A., and P. Karås. (2002a). *Effects of eutrophication on young-of-the-year freshwater fish communities in coastal areas of the Baltic.* Environmental Biology of Fishes 63:89-101.

Sandström, A., and P. Karås. (2002b). *Tests of artificial substrata as nursery habitat for young fish.* Journal of Applied Ichtyology 18:102-105.

Sandström, O., I. Abrahamsson, J. Andersson, and M. Vetemaa. (1997). *Temperature effects on spawning and egg development in Eurasian perch.* Journal of Fish Biology 51:1015-1024.

Saulamo, K., and J. Lappalainen. (2006). *Effects of abiotic factors on movements of pikeperch during pre-spawning and spawning season in a Baltic archipelago.* Hydrobiologia 579:271-277.

Smith, C., A. Douglas, and P. Jurajda. (2001). *Oviposition site selection and embryo mortality in perch.* Journal of Fish Biology 58:880-882.

Snickars, M., A. Sandström, A. Lappalainen, and J. Mattila. (2007). *Evaluation of low impact pressure waves as a quantitative sampling method for small fish in shallow water.* Journal of Experimental Marine Biology and Ecology 343:138-147.

Thorpe, J. (1977). *Synopsis of biological data on the perch Perca fluviatilis Linnaeus, 1758 and Perca flavescens Mitchill, 1814.* FAO, Rome.

Treasurer, J. W. (1983). *Estimates of egg and viable embryo in a lacustrine perch, Perca fluviatilis.* Environmental Biology of Fishes 8, 3-16.

Vaughan, I. P., and S. J. Ormerod. (2005). *The continuing challenges of testing species distribution models.* Journal of Applied Ecology 42:720-730.

Urho, L. (1996). *Habitat shift of perch larvae as survival strategy.* Annales Zologici Fennici 33:329-340

Urho, L. (2002). *The importance of larvae and nursery areas for fish production.* PhD-thesis, University of Helsinki, 215 pp.

Wright, P. J., H. Jensen, and I. Tuck. (2000). *The influence of sediment type on the distribution of the lesser sandeel, Ammodytes marinus.* Journal of Sea Research 44:243-256.

Wright, R. M., and E. A. Shoesmith. (1988). *The reproductive success of pike, Esox lucius: aspects of fecundity, egg density and survival.* Journal of Fish Biology 33:623-636.

Ådjers, K., Appelberg, M., Eschbaum, R., Lappalainen, A., Minde, A., Rpecka, R. and Thoresson, G. (2006). *Trends in coastal fish stocks of the Baltic Sea.* Boreal Environment Research 11:13-25.

About the BALANCE project:

This report is a product of the BSR INTERREG IIIB project "BALANCE".

The BALANCE project aims to provide a transnational marine management template based on zoning, which can assist stakeholders in planning and implementing effective management solutions for sustainable use and protection of our valuable marine landscapes and unique natural heritage. The template will be based on data sharing, mapping of marine landscapes and habitats, development of the blue corridor concept, information on key stakeholder interests and development of a cross-sectoral and transnational Baltic zoning approach. BALANCE thus provides a transnational solution to a transnational problem.

The BALANCE partnership is composed of the following institutions based in 10 countries: The Danish Forest and Nature Agency (Lead), The Geological Survey of Denmark and Greenland, The National Environmental Research Institute/University of Aarhus, The Danish Institute for Fisheries Research, WWF Denmark, WWF Germany, Institute of Aquatic Ecology at University of Latvia, Estonian Marine Institute at University of Tartu, Coastal Research and Planning Institute at Klaipeda University, Metsähallitus Natural Heritage Service, The Finnish Environment Institute, The Geological Survey of Finland, WWF Finland, The Swedish Environmental Protection Agency, The National Board of Fisheries – Department of Research and Development, The Geological Survey of Sweden, County Administrative Board of Stockholm, Department of Marine Ecology at Gothenburg University and WWF Sweden. The following institutes contribute as consultants to the partnership: The Geological Survey of Norway, Norwegian Institute for Water Research, DHI Water & Environment, The Leibniz Institute of Marine Sciences, The Sea Fisheries Institute, The Finnish Game and Fisheries Research Institute, Metria Miljöanalys and The Nature Conservancy.

The **BALANCE Report Series** included at the 30th of March 2007:

BALANCE Interim Report No. 1 "Delineation of the BALANCE Pilot Areas".

BALANCE Interim Report No. 2 "Development of a methodology for selection and assessment of a representative MPA network in the Baltic Sea – An interim strategy".

BALANCE Interim Report No. 3 "Feasibility of hyperspectral remote sensing for mapping benthic macroalgal cover in turbid coastal waters of the Baltic Sea".

BALANCE Interim Report No. 4 "Literature review of the "Blue Corridors" concept and its applicability to the Baltic Sea".

BALANCE Interim Report No. 5 "Evaluation of remote sensing methods as a tool to characterise shallow marine habitats".

BALANCE Interim Report No. 6 "BALANCE Cruise Report – The Archipelago Sea".

BALANCE Interim Report No. 7 "BALANCE Cruise Report – The Kattegat".

BALANCE Interim Report No. 8 "BALANCE Stakeholder Communication Guide".

BALANCE Interim Report No. 9 "Model simulations of blue corridors in the Baltic Sea".

BALANCE Interim Report No. 10 "Towards marine landscapes of the Baltic Sea (June 2007)"

BALANCE Interim Report No. 11 "Fish habitat modelling in the Archipelago Sea"

BALANCE Interim Report No. 12 "Evaluation of remote sensing methods as a tool to characterise shallow marine habitats II"

BALANCE Interim Report No. 13 "Harmonizing marine geological data with the EUNIS habitat classification"

BALANCE Interim Report No. 14 "Intercalibration of sediment data from the Archipelago Sea"

BALANCE Interim Report No. 15 "Marine spatial planning in the Baltic Sea – an interim report"

BALANCE Interim Report No. 16 "The stakeholder – nature conservation's best friend or its worst enemy?"

For more information please see www.balance-eu.org and http://maps.sgu.se/Portal