

4.1 Fish habitat modelling in the Baltic Sea archipelago region (Balance Pilot Area 3)

Authors: Ulf Bergström, Alfred Sandström, Göran Sundblad

Abstract: Young fishes are often dependent on certain habitats for their survival, and protecting these habitats may be crucial for maintaining strong adult stocks. An important step in providing sufficient protection for such essential fish habitats is reliable large-scale habitat mapping. One promising approach is that of spatial predictive modelling, where statistical models relating species occurrence to environmental variables are coupled to geographic information systems (GIS). This approach was used for mapping spawning and nursery areas of a number of common coastal fishes in the large, complex, Swedish-Finnish archipelago area in the Northern Baltic Sea. Generalized additive models (GAM) were fitted to describe the relationship between fish occurrence and habitat variables. The resulting probability maps were subsequently transformed into categorical presence/absence maps. We show that using only a few environmental predictors, the main patterns of occurrence can be mapped. The maps produced in this study can be directly used in spatial planning within the area, and also will constitute the basis for analyses of the ecological connectivity of the Natura 2000-network within BALANCE.

4.1.1 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.

Background

Modelling. 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 and Guegan 1999, Guisan et al. 2002, Guisan and Thuiller 2005). Generalized additive models (GAM) have, due to their flexibility, become popular in species distribution modelling (Hastie and Tibshirani 1990, Guisan and 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 and 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 and 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 and Ormerod 2005), although such an approach runs the risk of comparing sampling occasions or methods rather than 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 and 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.

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.

The 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*), pike-perch (*Sander lucioperca*) and roach (*Rutilus rutilus*). Pike-perch 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, Sandström et al. 1997). All four species have been subject to numerous studies, and their biology and ecology is well known.

Aims

The aim of this study was to produce maps describing the distribution of nursery habitats for perch, pike, pike-perch 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).

4.1.2 Material and Methods

The study area

The 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 maps of the physical environment is needed.

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 inflow of freshwater.

Water temperature and ice cover also often varies substantially between the inner and outer parts of the archipelago. Maximum summer water temperature range is normally 15-20 °C in open water and around 25 °C 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 °C. 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 pike-perch 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.

Field sampling and data sources

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 m² (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 m² 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.

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 as a predictor layer.

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: $\text{water clarity proxy} = \log_{10}(\text{wave exposure} / \text{distance})^{0.5}$.

Response variables

Separate models were constructed for young-of-the-year of perch, pike, pike-perch and roach, as well as for spawning of perch. For each model, presence/absence of the species was used as response variable.

Model selection (methods, algorithms, software, and routines)

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 the different species/life stages. 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).

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, limited to 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 D2 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 D2, 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 and 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 and 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.

4.1.3 Results

The model for pike-perch 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 overdispersion (Df = 296 total;

Residual deviance = 378.9644), indicating that there is some unexplained spatial heterogeneity in the data. All 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 the spatial predictions 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.

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 metres.

Validation

Model evaluation statistics for the probability of presence predictions, i.e. D2, 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 pike-perch 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 pike-perch 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.

4.1.4 Discussion

We have used three 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 and 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, due to their additive structure, not allow for taking interactions between predictor variables into account. There are however ways of overcoming 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 recalculated 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 pike-perch, 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.

4.1.5 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.

4.1.6 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.

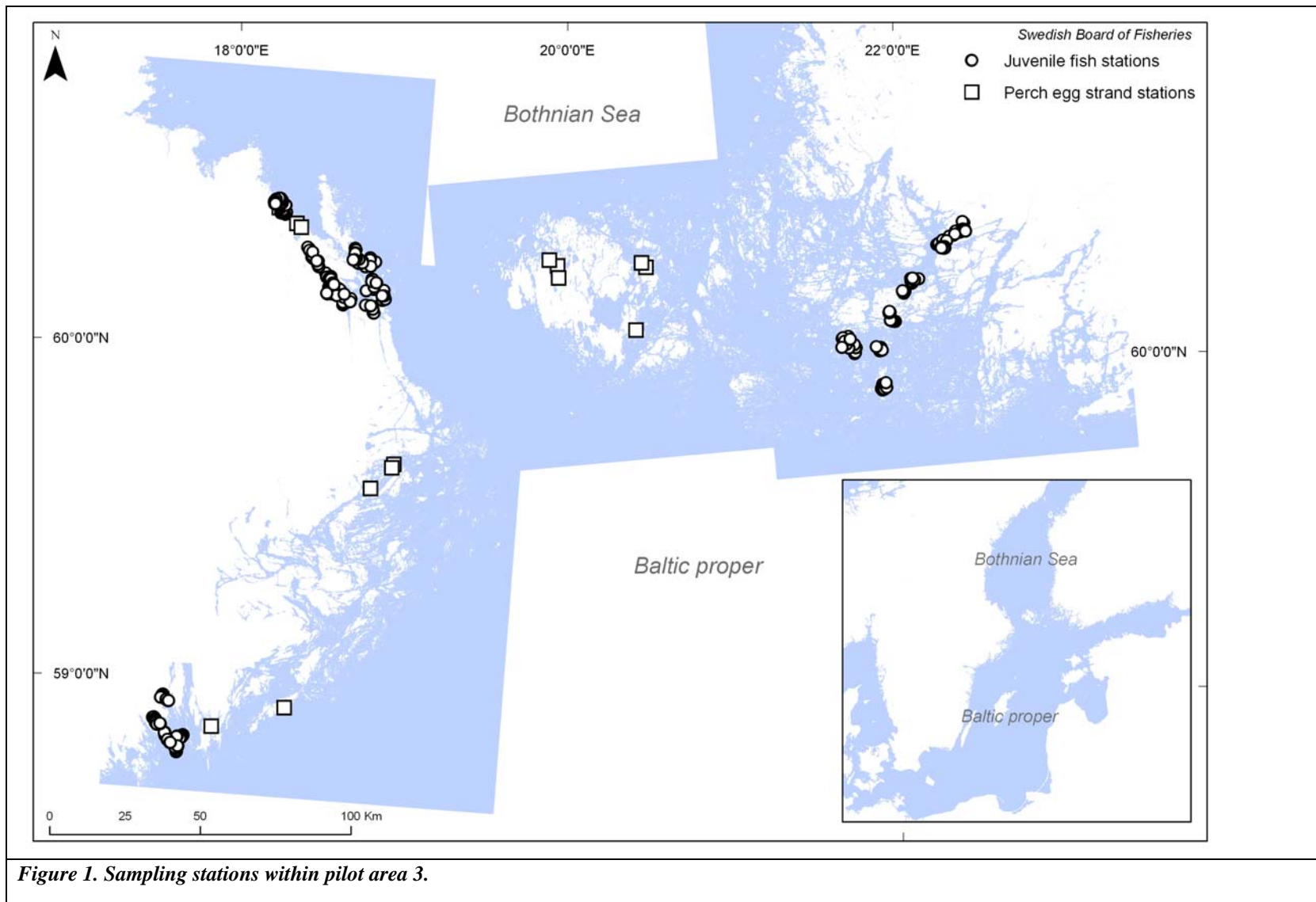
Bergström, U., Evertson, J., Karås, P., Sandström, A., Sundblad, G. and Wennberg, S. 2007. Evaluation of satellite imagery as a tool to characterise shallow habitats in the Baltic Sea. Balance interim report no. 5.

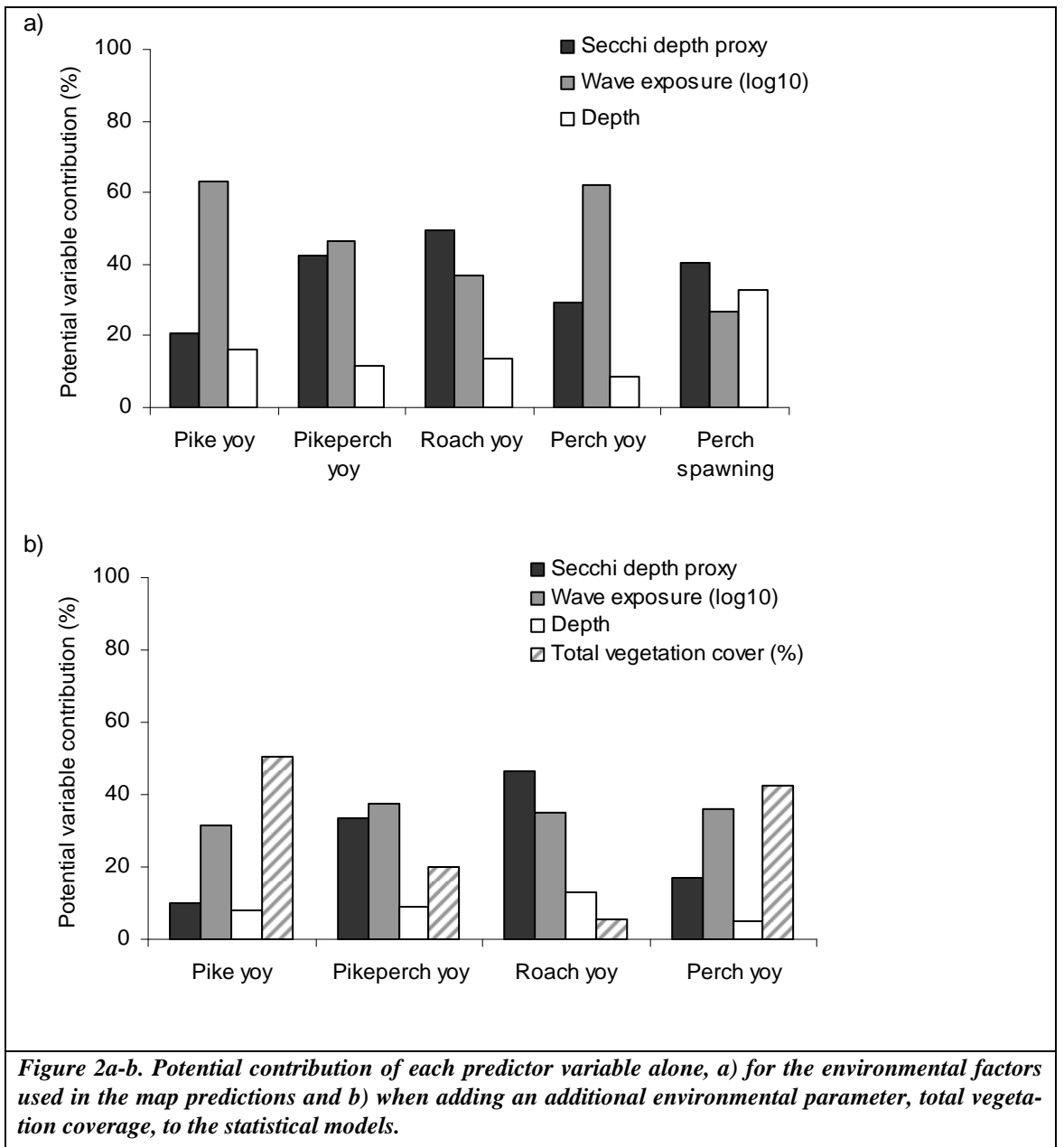
Carroll, C., W. J. Zielinski, and R. F. Noss. 1999. Using presence-absence data to build and test spatial habitat models for the fishes in the Klamath region, U.S.A. *Conservation Biology* 13:1344-1359.

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.
- 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.
- 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.
- 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 generalizability of prognostic information. *Annals of Internal Medicine* 130:515-524.
- Karås, P., and R. Hudd. 1993. Reproduction areas of fresh-water fish in the Northern Quark (Gulf of Bothnia). *Aqua Fennica* 23:39-49.
- 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.
- 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.

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





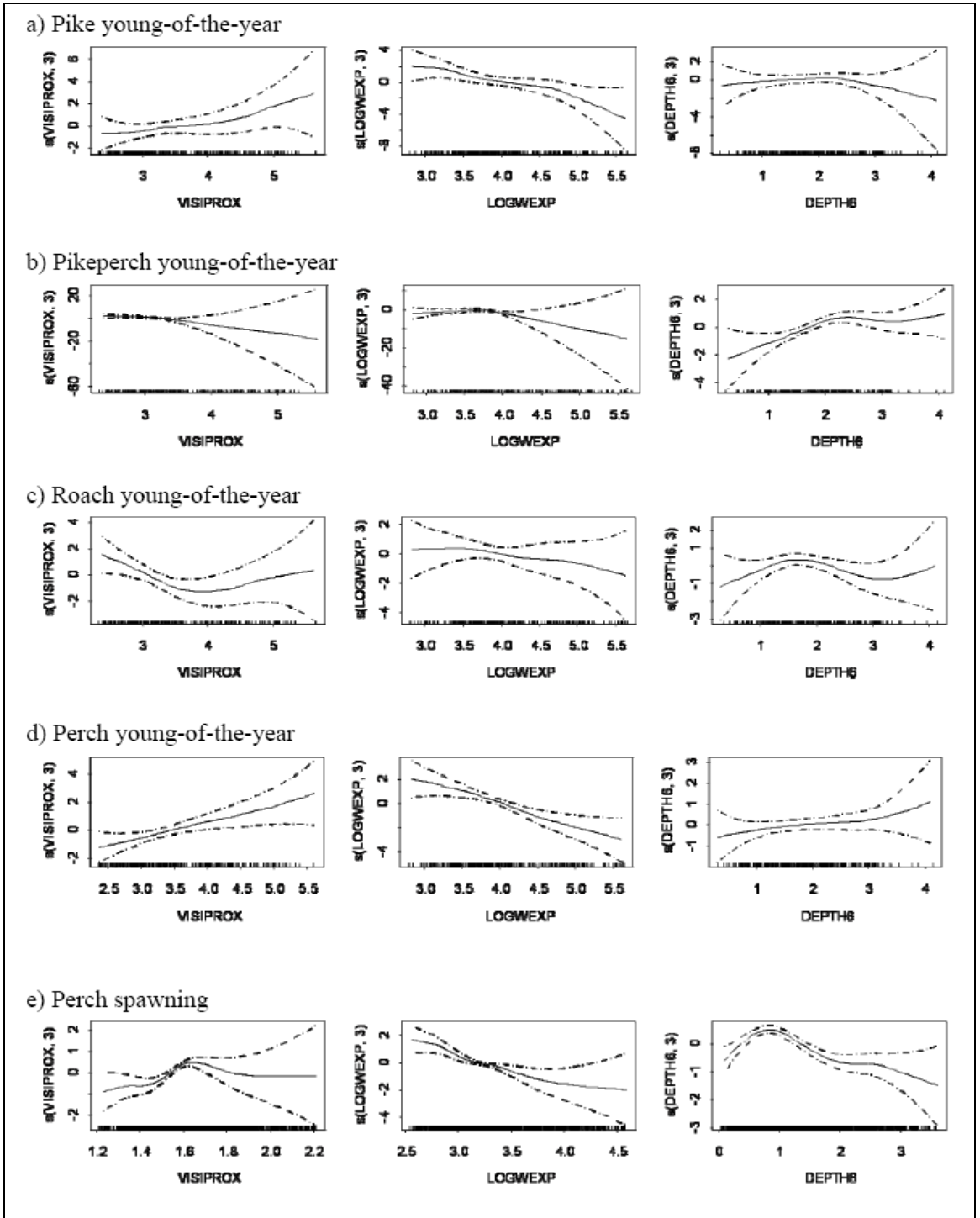


Figure 3 a-e. Partial response curves for the GAMs of the species and life stages modelled. The y-axis represents the response variable in the linear predictor scale. Visiprox denotes the Secchi depth proxy, logwexp the logarithm of the wave exposure index, and depth6 water depth. Dotted lines indicate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

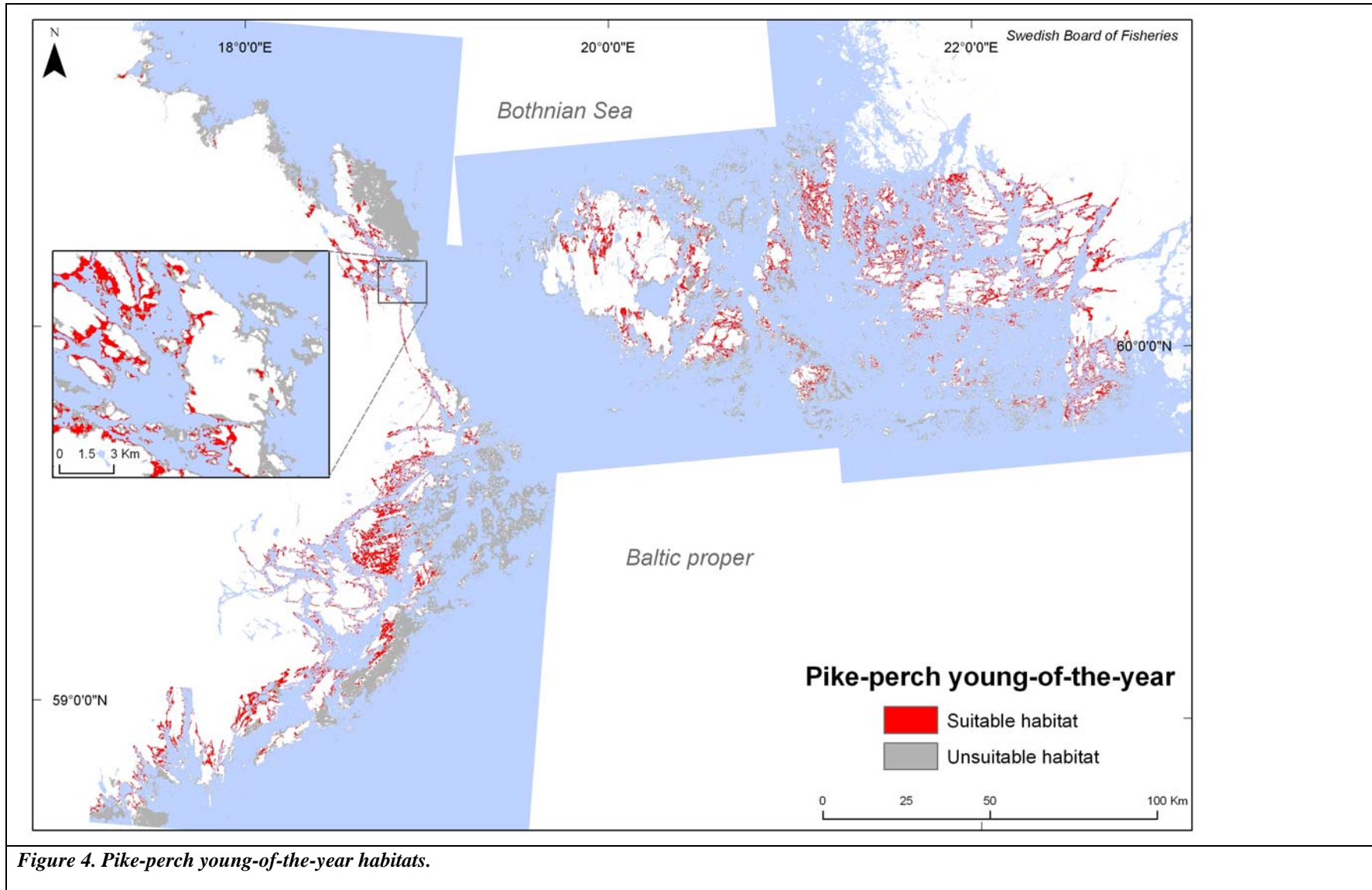


Figure 4. Pike-perch young-of-the-year habitats.

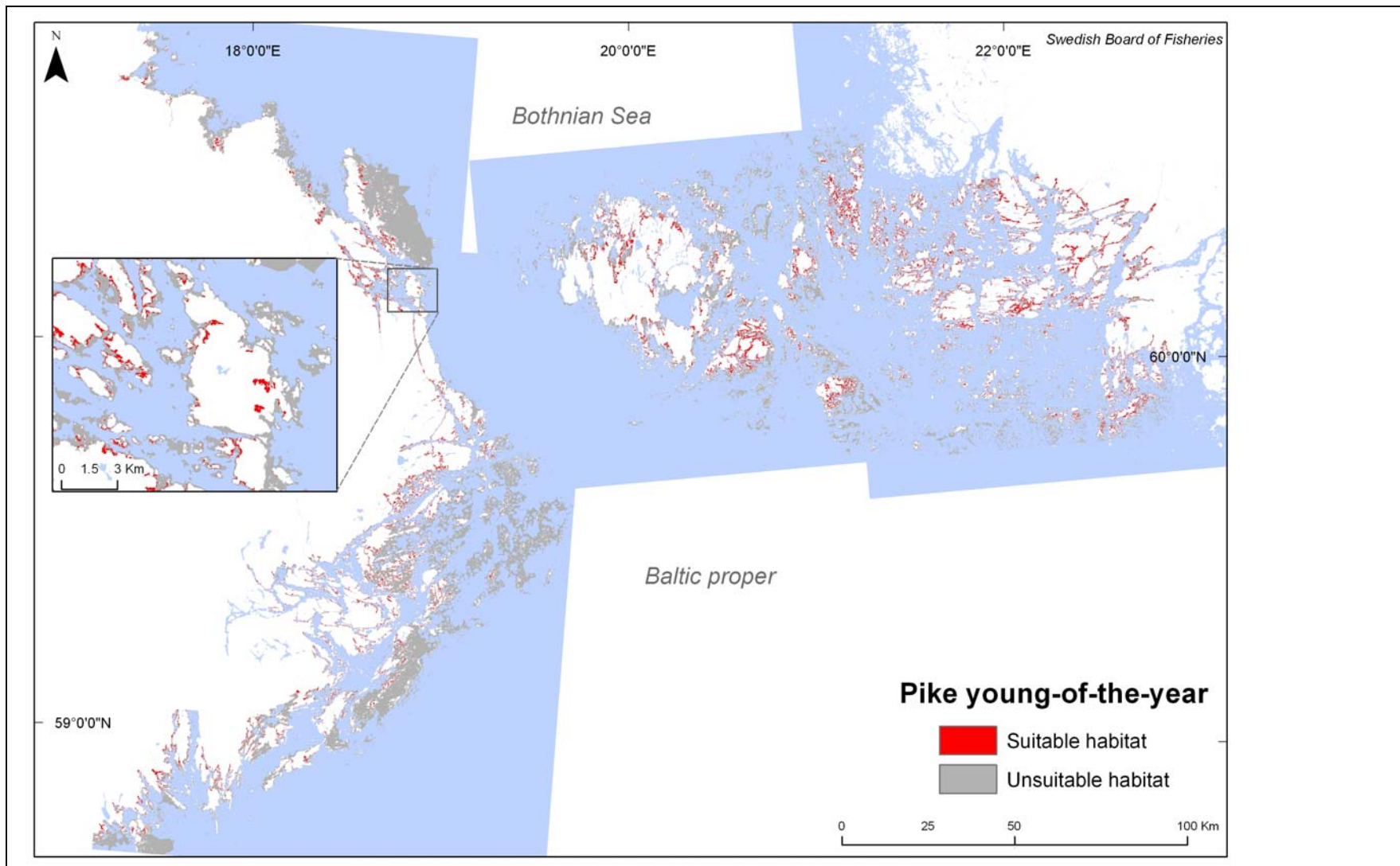


Figure 5. Pike young-of-the-year habitats.

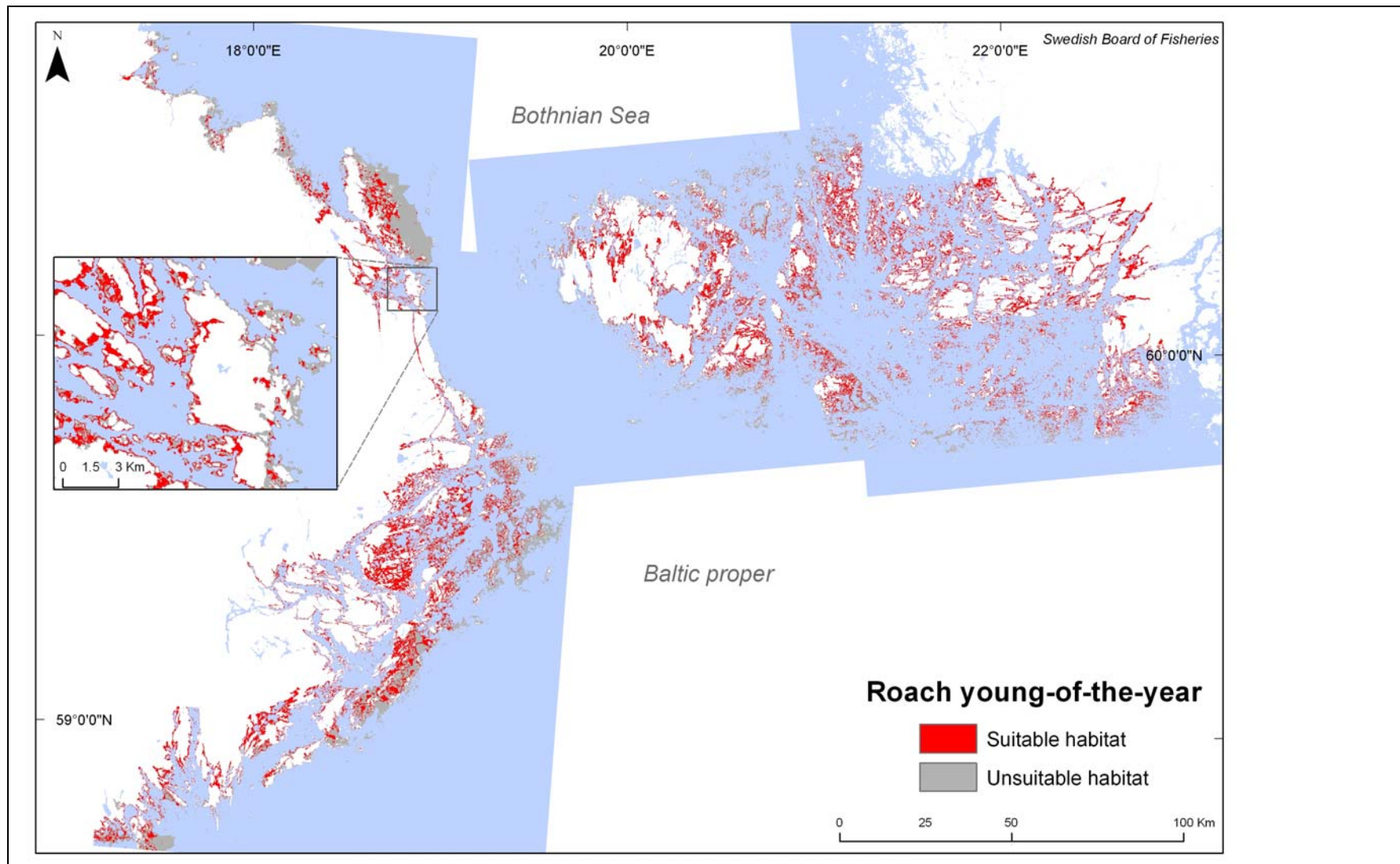


Figure 6. Roach young-of-the-year habitats.

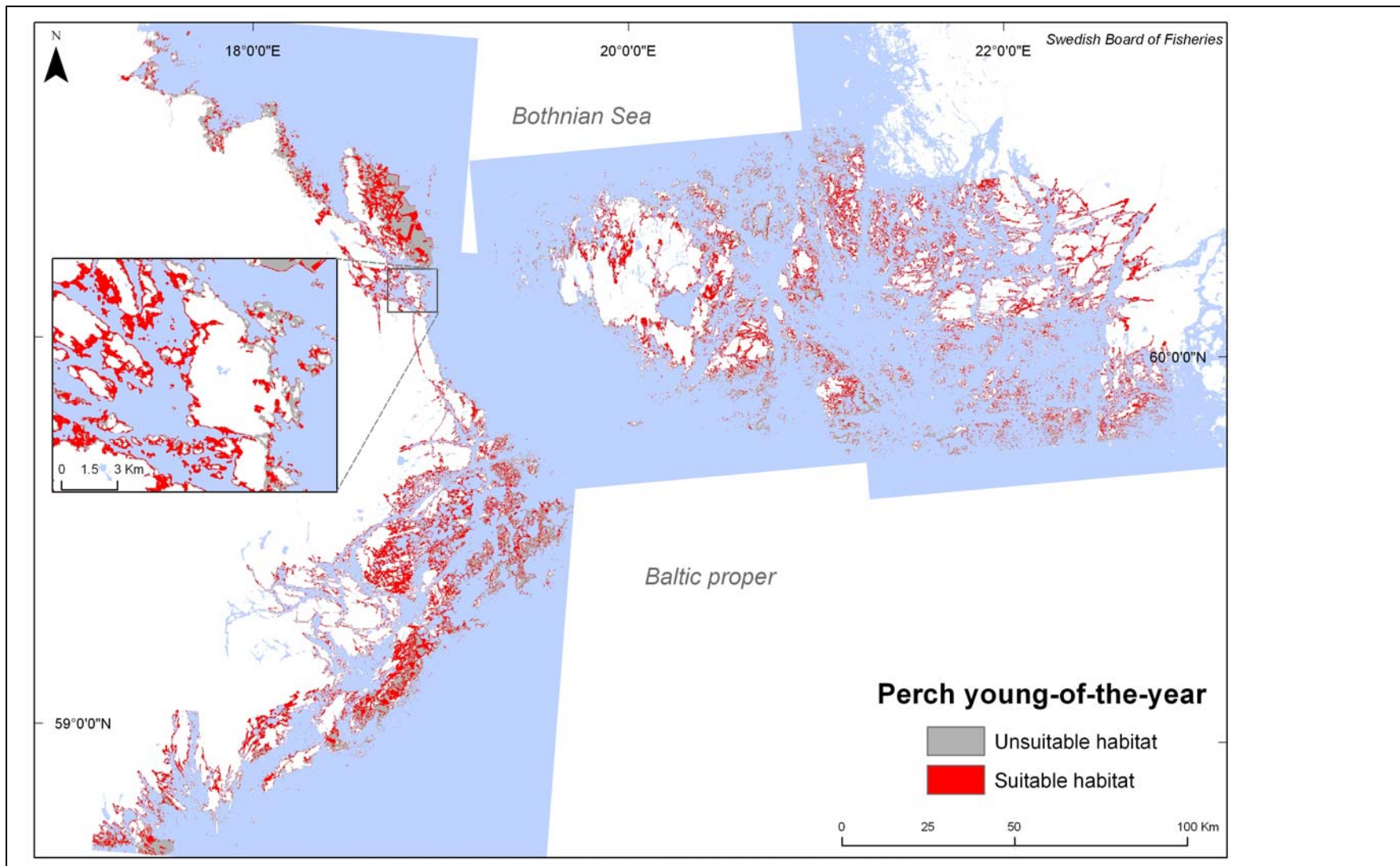


Figure 7. Perch young-of-the-year habitats.

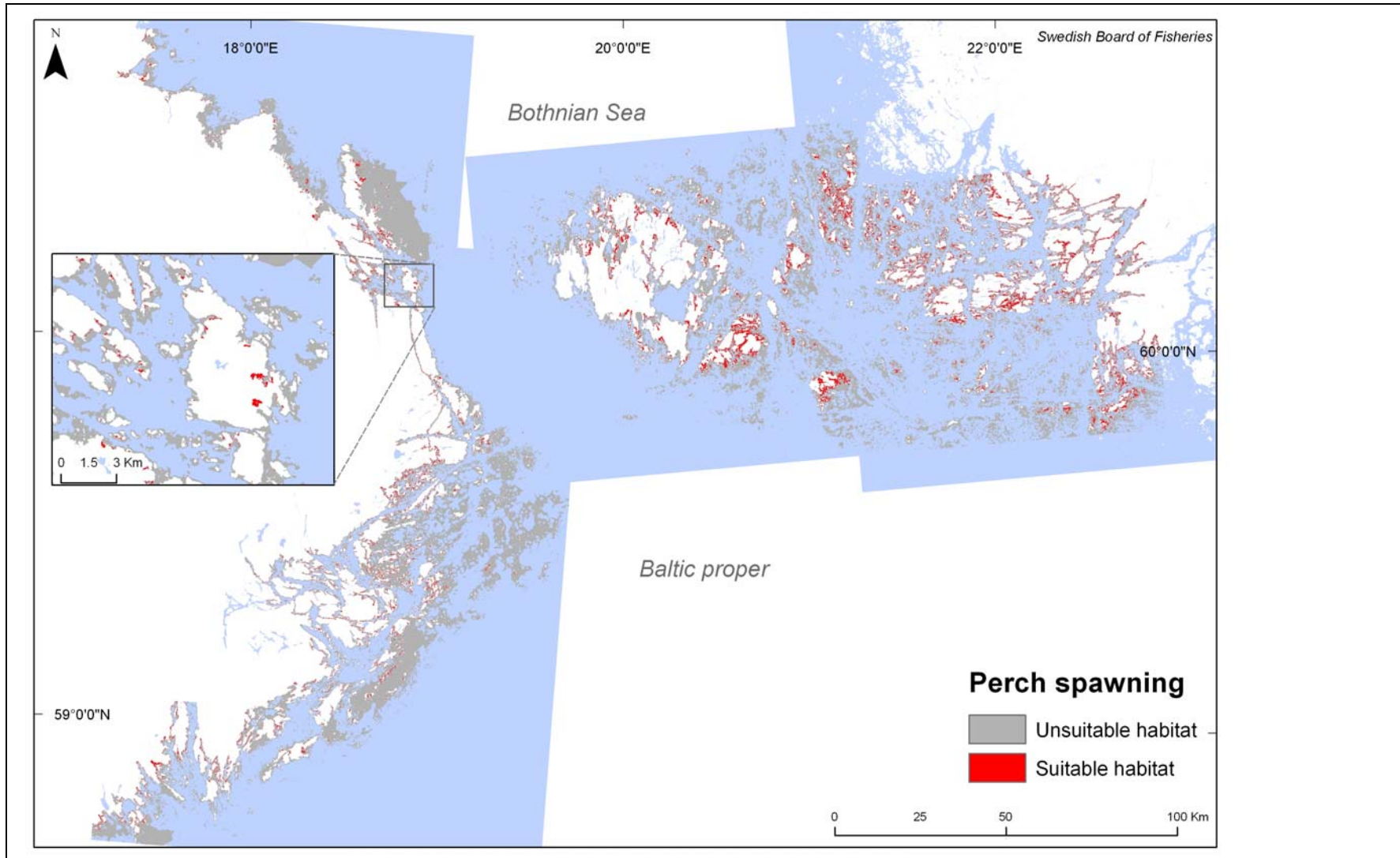


Figure 8. Perch spawning habitats.

| Table 1. Model evaluation statistics for the five recruitment habitat models. | | | | | |
|--|----------------|-------|------|----------------------|----------------------------|
| Species / Statistic | D ² | cvROC | ROC | Prevalence indata | TSS, catego- rised maps |
| Pike Y-O-Y | 0.11 | 0.64 | 0.75 | 0.08 | 28.6 |
| Pike-perch 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 |