Not to be cited without prior reference to the author

ICES CM 2011/G:07

Spatial predictive modelling as a tool in the Norwegian program for mapping of marine habitats

Trine Bekkby (1*), Eli Rinde (1), Heidi Olsen (2), Reidulv Bøe (2), Henning Steen (3), Nina M. Jørgensen (4), Frithjof Moy (3)

(1) Norwegian Institute for Water Research (NIVA), Gaustadalléen 21, NO-0349 Oslo, Norway. Phone: +4722185100, fax: +4722185200.
(2) Geological Survey of Norway (NGU), P.O. Box 6315 Sluppen, NO-7491 Trondheim, Norway. Phone: +4773904000
(3) Institute for Marine Research (IMR), P.O. Box 1870 Nordnes, NO-5817 Bergen, Norway. Phone: +4755238500, fax: +4755238531.
(4) Akvaplan-niva (ApN), NO-9296 Tromsø, Norway. Phone: +4777750300, fax: +4777750301.

*Contact author: Trine Bekkby, e-mail: trine.bekkby@niva.no

Abstract

There is a high pressure on coastal ecosystems from climate change and a diverse range of human activities. To develop adequate marine spatial plans and achieve knowledge based management decisions, managers and policy makers need mapped information, especially for areas of high ecological functioning and marine biodiversity. Such information has been lacking in the marine environment in Norway. In 2003, the National program for mapping and monitoring of marine biodiversity started developing methods and organisational structure for mapping selected marine habitats. As the Norwegian coast is long and complex, detailed mapping of the sea floor is costly. Consequently, it was decided to develop and apply spatial predictive modelling as a tool. The methodology includes collaboration between physical oceanographers, marine ecologists and geologists, statisticians and modellers, including GIS experts. We will here describe the methods developed and applied to the habitats “Large kelp forest” and “Carbonate sand deposits”, with Trøndelag as study area.

Keywords: GIS, Management, Marine spatial planning, Norway, Planning, Spatial predictive modelling

Background

There is a high pressure on coastal ecosystems from increasingly diverse human activities coupled with climate change impacts. According to the UN Rio convention (1992), all countries are obliged to know and protect their biological diversity. In Norway, this was followed up by a white paper (St. meld. 58 (1996-1997)), instructing all municipalities to map their biodiversity and valuate areas. This work was followed by the white paper on biodiversity (St. meld. 42 (2000-2001)) and the establishment of the Norwegian national program for mapping and monitoring of biodiversity (starting in 2003), financed by the Ministry of the Environment, the Ministry of Fishery and Coastal Affairs and the Ministry of Defence. A steering group was established, consisting of the Norwegian Directorate for Nature Management (program leader), the Directorate of Fisheries, the Norwegian Defence Estates Agency, the Climate and Pollution Agency and Tvedestrand municipality (the latter representing the users). The scientific work and field mapping is carried out by the Norwegian Institute for Water Research (NIVA scientific coordinator), Institute of Marine Research (IMR), the Geological Survey of Norway (NGU) and Akvaplan-niva (ApN).

To make appropriate management decisions, managers and policy makers need spatial information (i.e. maps) of areas of high importance in respect to biodiversity and ecosystem functioning. This information has been lacking in the marine environment. The aim of the national mapping program was therefore to develop maps for managers and planners, showing important habitats and key areas in the coastal zone (i.e. within 1 nautical mile outside the base line). The selected habitats/key areas were:

Habitats

- Eelgrass meadows and other seagrass meadows
- Soft sediments in the littoral zone
- Ice marginal deposits
- Large kelp forests
- Carbonate sand deposits

Key areas

- Oyster beds
- Large scallop populations
- Spawning areas for fish

A second aim of the mapping program was to develop criteria to distinguish between nationally, regionally and locally important instances of the different habitats/key areas. During 2003-2006, the marine part of the program focused on developing methods for each of the selected habitats/key areas and to develop a good organization structure for the forthcoming mapping. Existing data was sampled and integrated into maps, and guidelines for the mapping were developed at county level (Rinde et al., 2006). Field mapping based on the guidelines started in 2007. By 2010, mapping was finalised for the Oslofjord region, the Agder counties and Trøndelag (Bekkby et al., 2011). Due to reduced funding, some of the field mapping remained for the counties Hordaland and Troms. The program is now in the second mapping phase, which lasts from 2011-2015.

Since 2003 we have seen a rapid development in methods used for spatial predictive modelling, both with respect to methods and software. This paper presents the methods applied to the habitats “Large kelp forests” and “Carbonate sand deposits”, with Trøndelag as study area.

Methods

The coast of Norway is long (83 000 km when all islands are included) and complex, with large environmental gradients, high biodiversity and a variety of habitats. Traditional mapping of all areas is therefore too costly considering the available funding of the program. Information and knowledge that can be transferred from few samples to wider areas, as in spatial predictive modelling, are therefore of great value. Spatial predictive modelling has been applied in several studies (e.g. Lehmann, 1998; Guisan and Zimmermann, 2000; Kelly et al., 2001; Bekkby et al., 2008, 2009), and is increasingly used for mapping and nature management (e.g. within the Norwegian MAREANO programme, www.mareano.no, e.g. Dolan et al., 2009).

The study site and data

Field data were collected in Trøndelag, which consists of two counties covering large areas at the north-western coast of Norway (Figure 1), from 2007-2010. Stations were placed to cover the wide range of environmental gradients, i.e. depth, slope, terrain variability, wave exposure and current speed.

For kelp forest, a total of 1167 stations were selected randomly within stratified classes of the environmental factors. At each station, kelp coverage was recorded by NIVA and IMR using a water glass (in shallow areas, down to ~5 m) or an underwater camera (UWC, in deeper water). The UWC was supplied with light and connected to a monitor on the boat through a cable. The position of each station was recorded with a GPS (accuracy ±5 m). Coverage was classified semi-quantitatively into one of four classes: 0: no kelp, 1: single plants/scattered, 2: common/moderately dense and 3: dominating/dense). The large kelp forests as a habitat were defined as areas where *Laminaria* *hyperborea* was dense and dominating. Consequently, class 3 (dominating/dense) was defined as presence and the classes 0-2 were defined as absences. The dataset was divided into a modelling and a validation set (Table 1).

For carbonate sand deposits, a total of 1263 stations were selected. Stations were located along depth gradients and partly at random, but biased towards areas where field experience has shown that carbonate sand occurs and is presently produced (i.e. wave exposed and narrow sounds and sediment accumulation basins in the outer archipelago zone shallower than about 50 m water depth). Carbonate sand does generally not occur in fjords and sheltered areas or in areas of glacial sand and gravel. NGU collected grab samples at each station, and carbonate content was visually determined. The position of each station was recorded on a GPS (accuracy ±5 m). All stations with carbonate content ≥ 50 % were defined as presences, those with carbonate content <50 % as absences (Table 1). We originally had 717 absences. To get a more balanced dataset, we randomly selected 388 of these, resulting in the same number of presences and absences for the modelling.

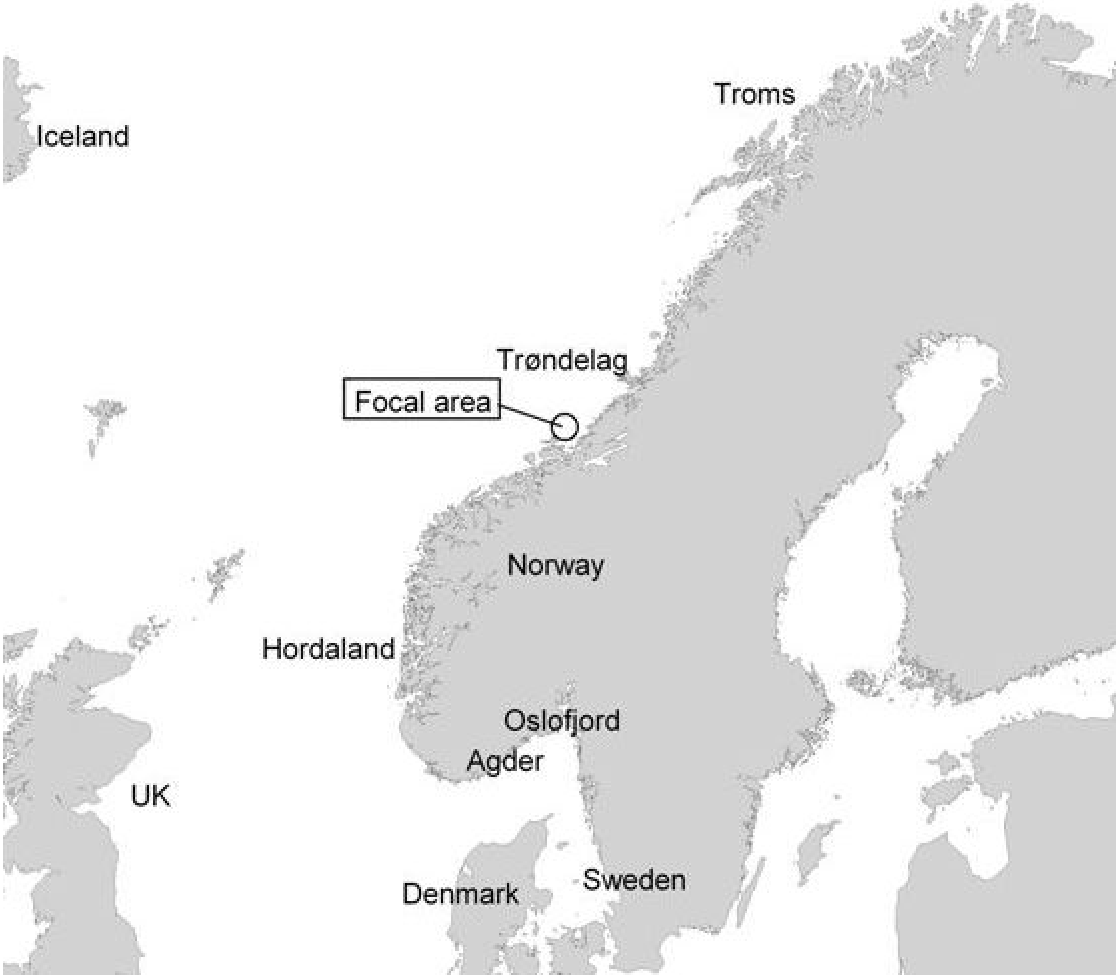

Figure 1. The study area in which details of the modelling is shown in Figures 4, 8 and 10.

Table 1. Sampled field data for kelp forest and carbonate sand modelling and validation. The datasets are divided into presences and absences. Fore kelp forest (*Laminaria hyperborea*), presences were defined as stations at which kelp was dense and dominating, absences were defined as stations with no kelp, single plants, scattered occurrences or common/moderately dense kelp forest. For carbonate sand deposits, presences were defined as stations with >50 % carbonate content, absences were defined as <50 % carbonate content.

	Large kelp forests			**Carbonate sand deposits**		
	Presences	**Absences**	**Total**	**Presences**	**Absences**	**Total**
Modelling	374	358	**732**	388	388	**776**
Validation	114	321	**435**	97	61	**158**
Total	**488**	**679**	**1167**	**485**	**449**	**934**

The predictor variables - GIS layers
The variables depth, slope, terrain curvature and wave exposure were available as GIS layers at a 25 m spatial resolution. Current speed was available at a 500 m resolution. All data and models were integrated into ArcGIS 9.2, and the current speed model was rescaled to 25 m resolution. We analysed the effect of three terrain curvature models, one with a 250-m calculation window, one with a 500-m calculation window and one with a 1-km calculation window. For current speed, depth-averaged models (averaged over the water column) of median and 90th percentile were tested.

Statistical analyses and model selection
We analysed the statistical influence of depth, slope, terrain curvature, wave exposure and current speed using generalized additive models (GAMs; Hastie and Tibshirani, 1990; occurrence as a binomial dataset, 2 d.f. for the smoothing spline function) in S-PLUS 2000. As a tool for model selection, we used the Akaike Information Criterion (AIC; Burnham and Anderson, 2001) in GRASP (an extension to S-PLUS 2000; Lehmann et al., 2003, 2004). We used the AICc calculations, as recommended by Burnham and Anderson (2004), which is the AIC adjusted to fit small sample sizes (AIC and AICc being equal at large sample sizes).

Spatial predictive modelling and validation
Based on the response curves from the GAM analysis, GRASP develops a matrix of prediction. In ArcView 3.3, the matrix for the selected statistical model was used to produce an output spatial probability model (at a spatial resolution of 25 m) from GIS layers of the variables. The predicted probability distribution was validated using both a cross-validation test and an independent dataset. In both cases, we used the area under the receiver operating characteristic (ROC) curve, known as the AUC value (Swets, 1988). The cross-validation ROC test (Fielding and Bell, 1997) was made with five subsets (folds) of the same dataset used for modelling (fivefold cv-ROC).

Defining occurrences from probability models
Based on statistical analyses of the presence/absence data, the output spatial predictive model provides estimates of the probability of finding kelp forest or carbonate sand at each of the stations. We compared these probabilities (arcsin transformed) with the four kelp coverage classes observed in the field using ANOVA in StatGraphics Plus 5.1 (different classes was not established for carbonate sand deposits). Multiple range tests provide information on which means that are significantly different from others. The cut-off probability value used to determine occurrence of kelp and carbonate sand (and hence deciding the size of the polygons made available to the managers and planners) was defined as the 75th percentile of the probability values for dominating/dense kelp forest and 50 % for carbonate sand deposits.

The habitats in Trøndelag were valuated according to the following criteria (revision is ongoing):
Large kelp forests
- A (of national importance) – large kelp forests (≥500 000 m^2).
- B (of regional importance) – kelp forests 100 000-500 000 m^2. Kelp forests in sea urchin grazed areas.

Carbonate sand deposits
- A (of national importance) – large occurrences (>200 000 m^2) of carbonate sand with more than 50 % shells and shell fragments
- B (of regional importance) – occurrences (100 000-200 000 m^2) of carbonate sand with more than 50 % shells and shell fragments

Results
Spatial predictive modelling of kelp forest distribution
In Trøndelag, dense kelp forest was (according to the GAM and AIC analyses) best determined by the combined effect of depth, slope, curvature based on a 250 m calculation window, wave exposure and median current speed (AIC=718.4355, 5-fold cvROC=0.76, Figure 2). The second best model did not include current speed (AIC=722.6409). Wave exposure was the most important factor in the model

followed by depth and the terrain parameters (Figure 3). Current speed was, at the available resolution, (500 m), the least important factor in the model (Figure 3). Figure 4 shows the resulting map of the probability of finding dense/dominating kelp forests using the response curves.

The ANOVA analysis comparing the predicted probabilities (arcsin-transformed) for the four different coverage classes observed in the field (0: no kelp, 1: single plants/scattered, 2: common/moderately dense, 3: dominating/dense) shows that there is an overall significant difference in mean probability from one level of coverage to another (F=166.69, P<0.0001; Table 2). The probability of finding kelp increases with increased observed coverage (Figure 5). The multiple range tests (Table 2) show that the probability values of coverage class 1 (single plants/scattered) is not significantly different from class 2 (common/moderately dense). All other classes were significantly different from each other. The area of kelp occurrences (polygons) submitted to the managers was chosen as the areas with more than 60 % probability of dominating/dense kelp forest. This corresponds to the 75th percentile for coverage class 3 in the boxplot (Figure 5).

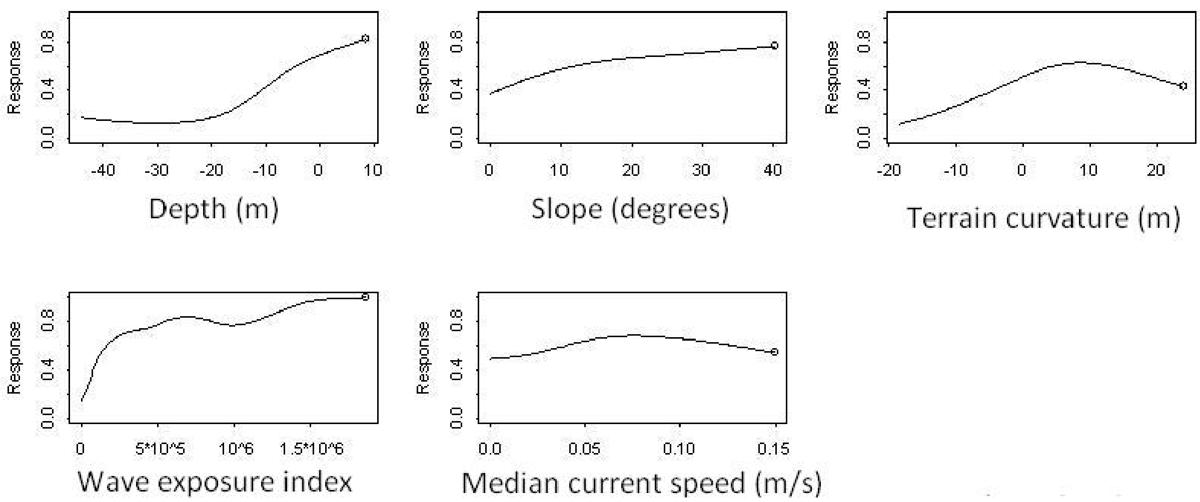

Figure 2. Response curves for kelp *Laminaria hyperborea* occurrence (presence/absence) against depth, slope, terrain curvature, wave exposure (Isæus, 2004) and median current speed. The y-axis is scaled according to the response, and is a number between 0 and 1. Terrain curvature is measured with a 250 m calculation window. A negative value indicates a basin and a positive value a shoal. The more negative the value, the deeper the basin; the more positive the value, the greater the rise in the shoal.

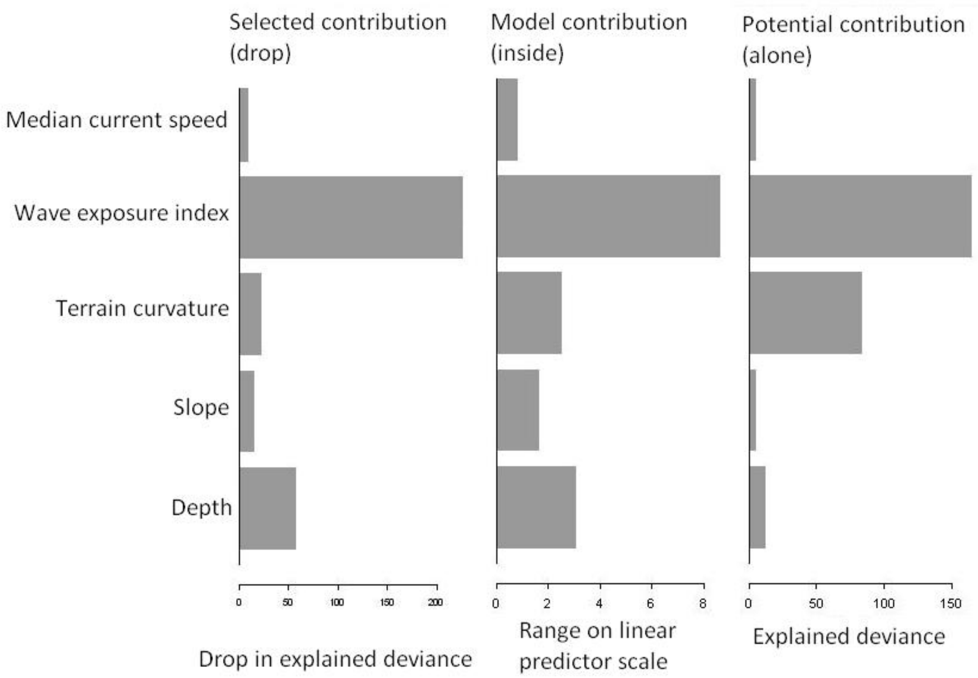

Figure 3. Deviance for kelp forest (*Laminaria hyperborea*) explained by each of the predictors in the best model (defined by the AIC analyses), shown by drop in explained deviance for each parameter ("drop", to the left), the parameters relative contribution within the model ("inside", in the middle), and the potential contribution of each parameter as a single factor ("alone", to the right).

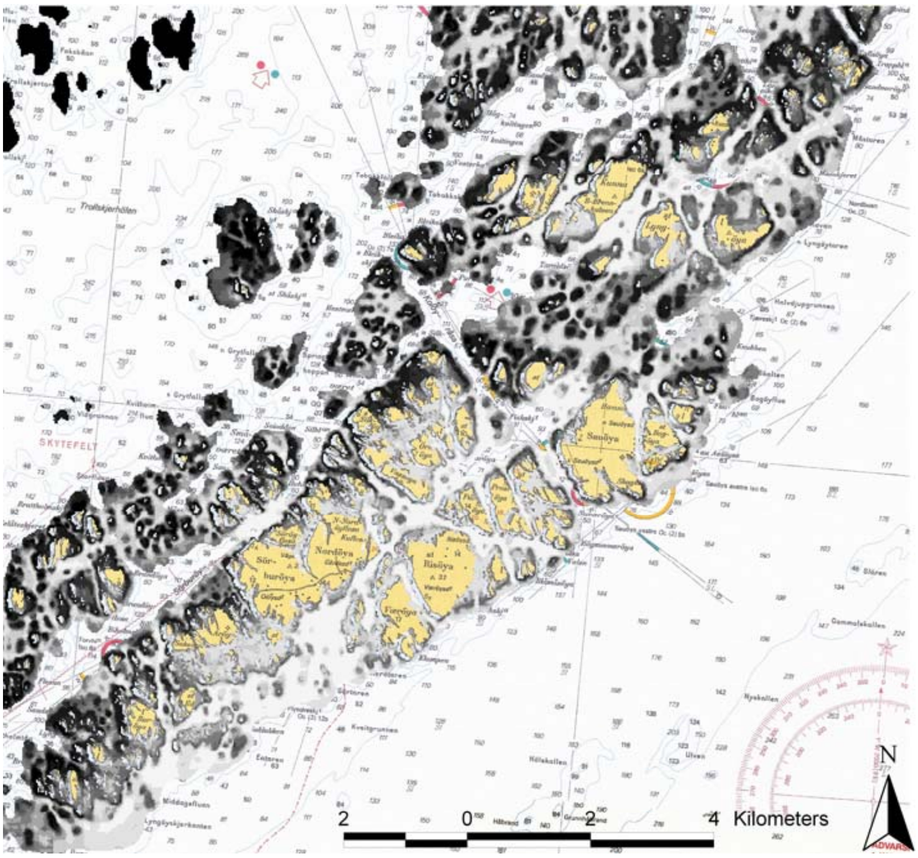

Figure 4. Predicted probability of finding dense/dominating kelp forests (*Laminaria hyperborea*) in Trøndelag, exemplified in a smaller area (see Figure 1 for overview), based on the response curves for depth, slope, terrain curvature with a 250 m calculation window, wave exposure and median current speed. The darker the colour, the higher the probability of finding dense/dominating kelp forest. The spatial resolution of the model is 25 m. The model is shown on top of a standard nautical chart.

Table 2. Results from the Duncan's multiple range tests, providing information on which kelp (*Laminaria hyperborea*) coverage class had predicted probability means that were significantly different from the other coverage classes (probability values arcsin-transformed). A minus sign indicates no significant different, a plus sign a significant difference (p<0.05).

Coverage	**0 (no kelp)**	**1 (single plants/scattered)**	**2 (common/ moderately dense)**	**3 (dominating/ dense)**
0 (no kelp)		+	+	+
1 (single plants/scattered)	+		-	+
2 (common/moderately dense)	+	-		+
3 (dominating/dense)	+	+	+	

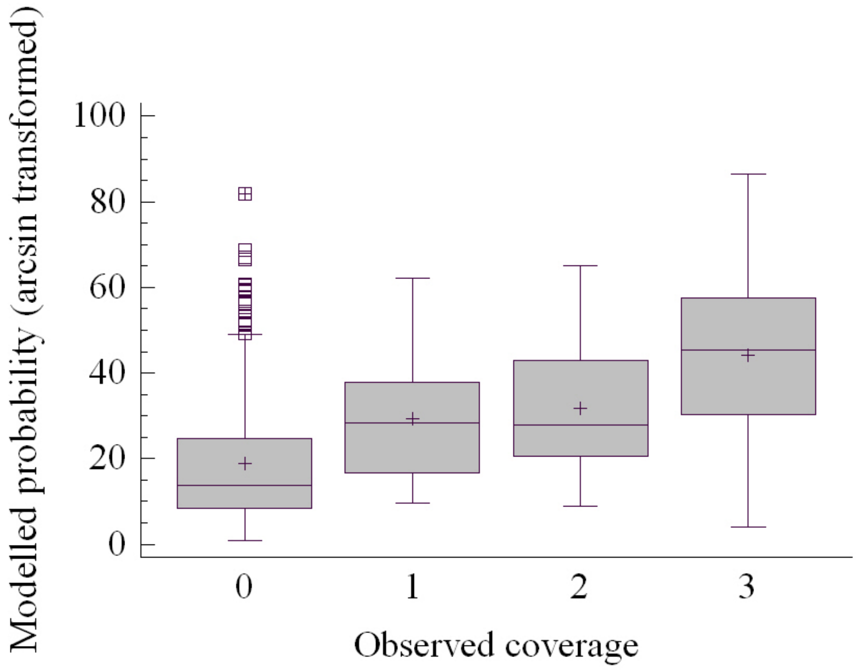

Figure 5. Box-and-whisker plot showing the variability in the predictive probability values within each coverage class (0: no kelp, 1: single plants/scattered, 2: common/moderately dense, 3: dominating/dense). The Duncan's multiple range tests (Table 2) showed that the probability values (arcsin-transformed) of coverage class 1 were not significantly different from class 2, but that all other classes were significantly different from each other. The data in the plot are divided into four areas of equal frequency (quartiles). The grey box in the plot encloses the middle 50 % values, the median is drawn as a horizontal line inside the box and the mean as a +. The bottom and top of the box is the 25th and 75th percentile of the data. The open squares are the outliers.

Spatial predictive modelling of carbonate sand deposits

Carbonate sand deposits with more that 50 % shells and shell fragments were best determined by the combined effect of depth, wave exposure and the maximum current speed (90th percentile) (AIC=990.1635, 5-fold cvROC=0.66, Figure 6). The second best model also included slope and terrain curvature with a 500 m calculation window (AIC=991.6656). Depth is the single most important factor in the model (Figure 7). Figure 8 shows the resulting map of the probability of finding carbonate sand deposits using the response curves. Only carbonate sand occurrences deposited recently and within the marine part of the coastal zone are covered. Water depths have changed by up to 180 m since carbonate sand started to form about 10 000 years ago, hence many occurrences are above sea level today.

The ANOVA analysis comparing the predicted probabilities (arcsin-transformed) for the presences and the absences shows that there is an overall significant difference in mean probability (F=115.20, P<0.0001, Figure 8). The areas of carbonate sand deposit (i.e. the polygons) submitted to the managers were areas with more than 40 % probability of carbonate sand, which corresponds to the 75th percentile (Figure 8). As the model overestimate the occurrence of carbonate sand deposits, the identified areas were filtered with respect to modelled terrain curvature. As carbonate sand deposits are most commonly found in basins and not on shoals, shoals (identified as curvature values >1 using a 250 calculation window) were excluded. As we had little data in deep areas, predicted areas deeper than 50 m depth were also excluded. Twenty years of experience has shown that carbonate sand is presently not formed in fjord areas. Fjords were thus not sampled and all modelled carbonate sand deposits in inner fjord areas were manually excluded. The filtered occurrences fulfilling the size criteria described earlier are shown together with the resulting predicted kelp forest areas in Figure 10.

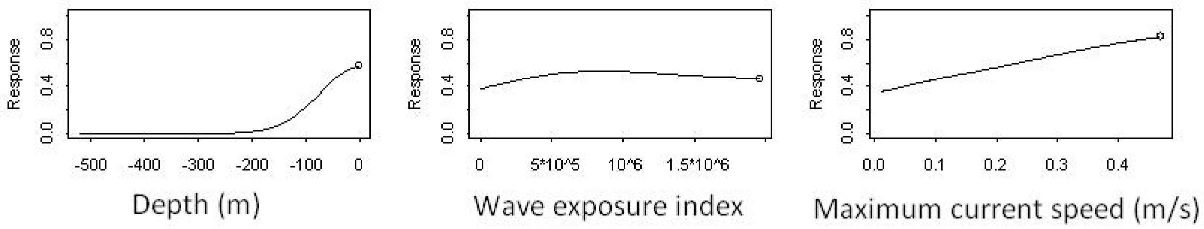

Figure 6. Response curves for carbonate sand deposit occurrence (presence/absence) against depth, wave exposure (Isæus, 2004) and maximum current speed. The y-axis is scaled according to the response, and is a number between 0 and 1.

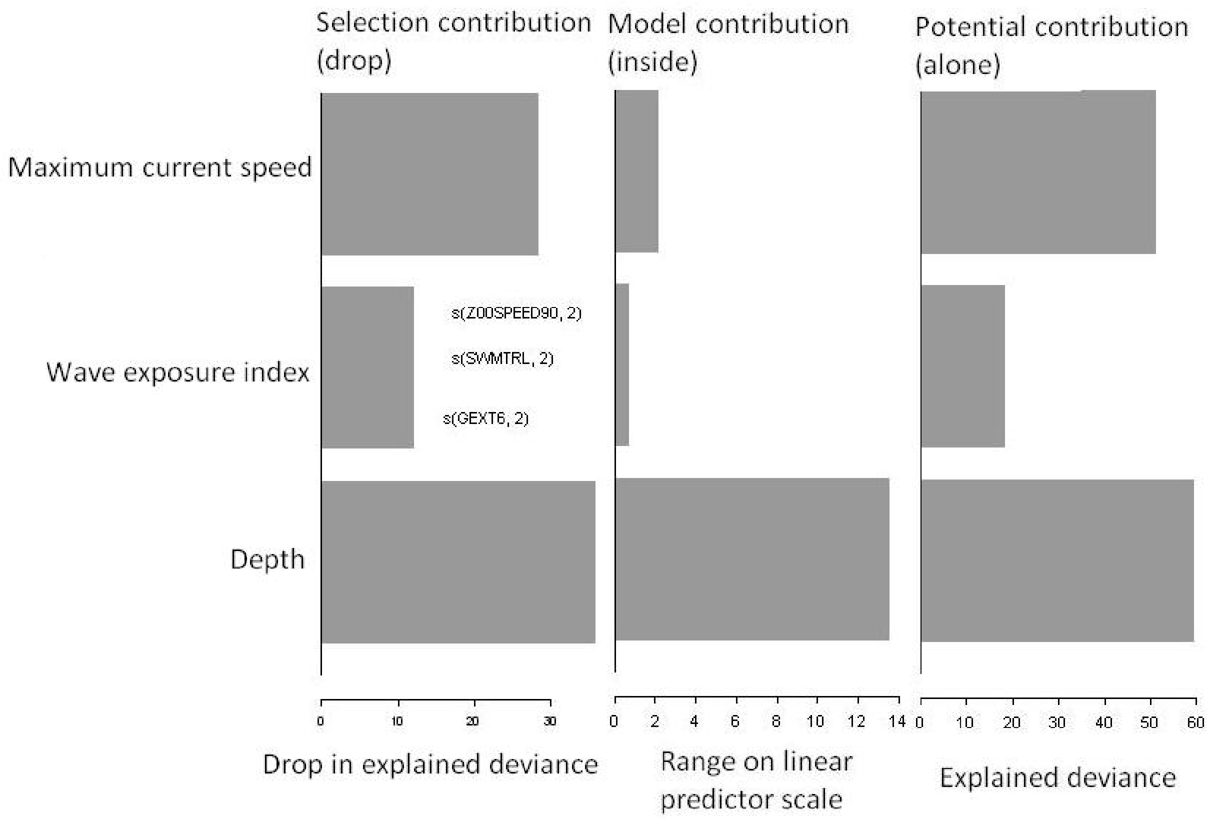

Figure 7. Deviance explained by each of the predictors in the best carbonate sand model (defined by the AIC analyses), shown by drop in explained deviance for each parameter (“drop”, to the left), the parameters relative contribution within the model (“inside”, in the middle), and the potential contribution of each parameter as a single factor (“alone”, to the right).

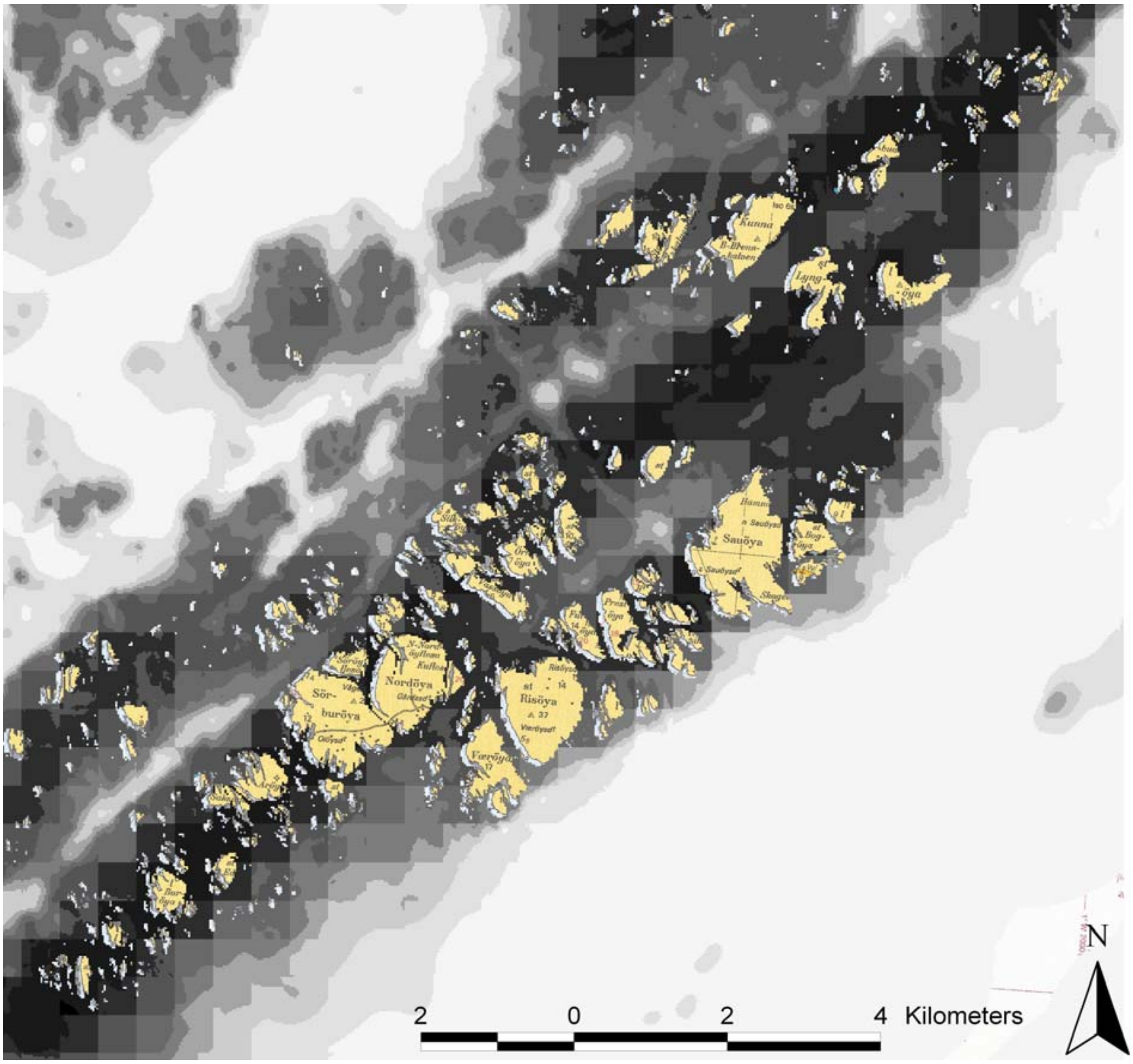

Figure 8. Predicted probability of finding carbonate sand deposits with more that 50 % shells and shell fragments in Trøndelag, exemplified in a smaller area (see Figure 1 for overview), based on the response curves for depth, wave exposure and maximum current speed. The darker the colour, the higher the probability of finding carbonate sand deposits. The spatial resolution of the model is 25 m. The coarse pixels are due to the input current speed model having a 500 m spatial resolution.

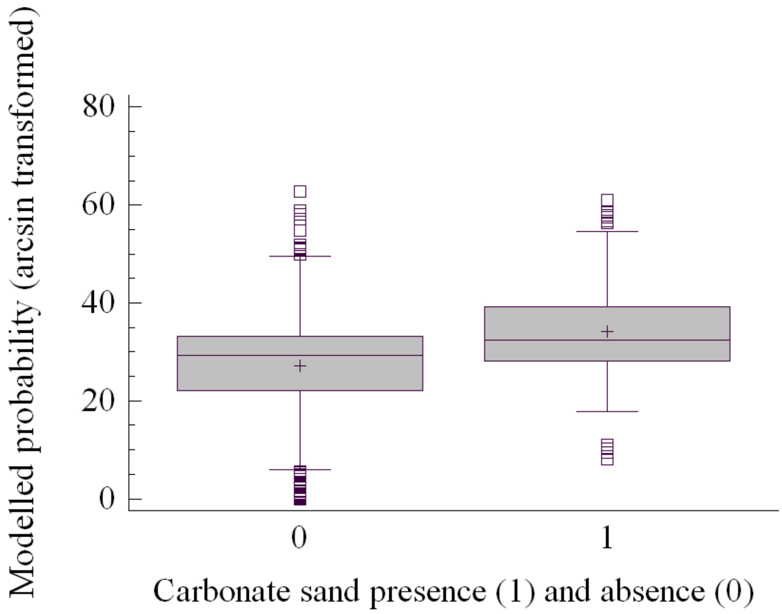

Figure 9. Box-and-whisker plot showing the variability in the predictive probability values for presences and absence, presences being carbonate sand deposits with more than 50 % shells and shell fragments. The data in the plot are divided into four areas of equal frequency (quartiles). The grey box in the plot encloses the middle 50 %, where the median is drawn as a horizontal line inside the box and the mean as a +. The bottom and top of the box is the 25th and 75th percentile of the data. The open squares are the outliers.

Discussion – relevance to management and planning

All habitats mapped within the national mapping program are visualised and made available through www.naturbase.no, a portal developed by the Norwegian Directorate for Nature Management. Figure 10 shows the kelp forest areas and carbonate sand deposits as they occur after filtering and before they are submitted to the database. All information from the portal can be downloaded as vector (shape) or SOSI files. The portal is used by managers and planners that make decision on where to allow activities in the costal zone. Consequently, in addition to field mapping and developing maps of the likely distribution of the different habitats, a lot of work has been dedicated to provide a description of different areas and occurrences. .

At the annual meetings arranged by the programme management and researchers, planners and researchers have discussed different aspects of the programmes activities, such as methodology, testing/developing criteria for valuation and the application and usefulness of the maps. The presentations and discussions at these meetings have clearly pointed out that the produced maps are highly sought after by managers and planners. Without such maps, planning and management for the coastal zone are made without any knowledge of the marine life.

One thing is to map selected habitats and key areas. Another is providing managers and planners with a tool for giving preference to different habitats and areas when planning activities in the coastal zone. During the pilot period (2003-2007), criteria for valuating habitats and key areas were developed. However, years of field mapping experience have resulted in a need to change and refine these criteria. At present, new criteria has been suggested and work with testing these on data is ongoing.

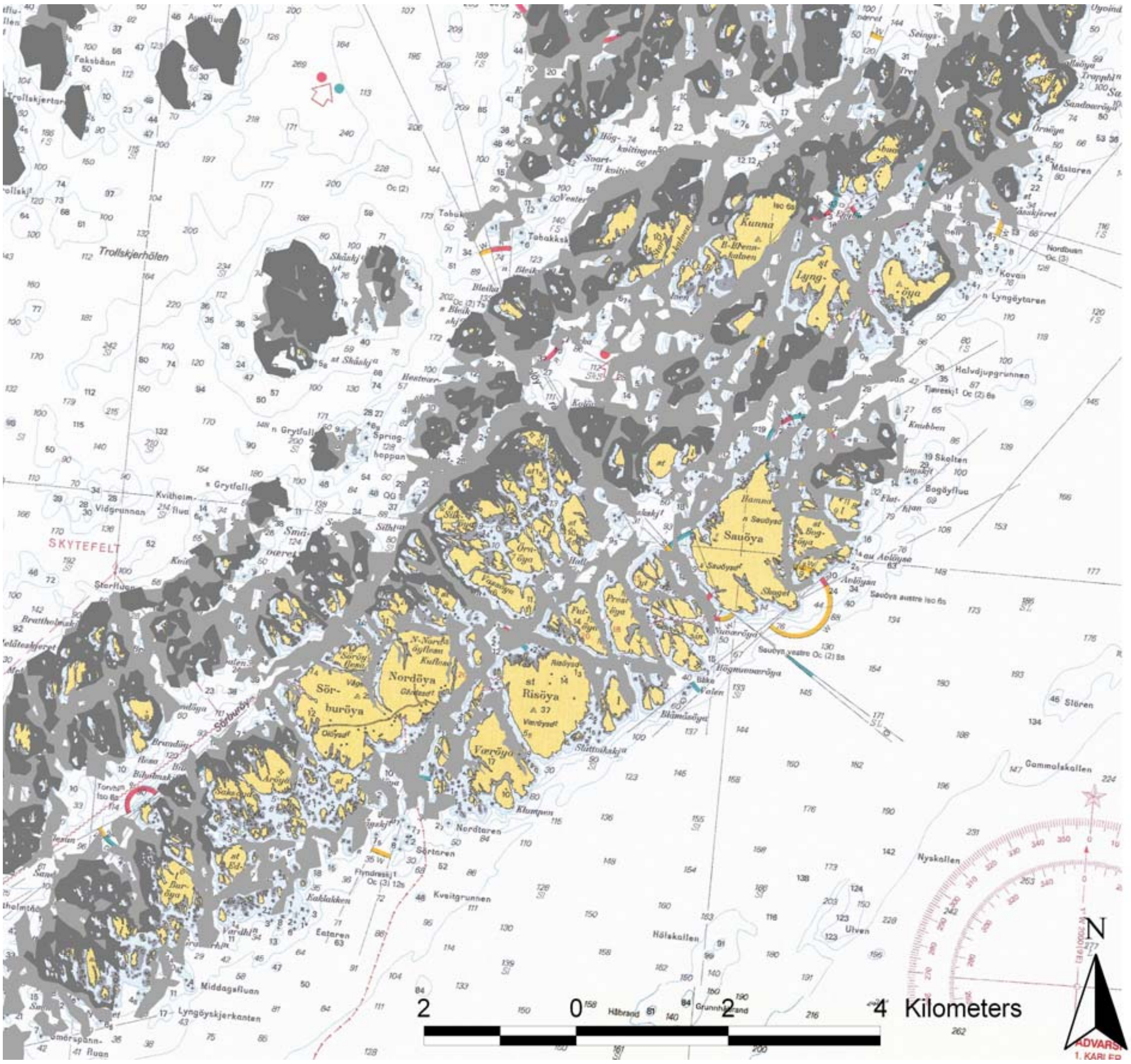

Figure 10. Dominating/dense kelp forests (dark grey) and carbonate sand deposits (light grey) after filtering. Large kelp forests were selected from the model using 60 % probability as the cutoff-value. Only kelp forest occurrences larger than 100 000 m2 are shown, as these are the ones submitted to www.naturbase.no. Carbonate sand deposits (containing more than 50 % shells and shell fragments) were selected from the model using 40 % probability as the cutoff-value and filtered against curvature and depth. Deposits in fjords, areas in which we have little field data, were excluded manually. The occurrences are shown on top of a standard nautical chart.

Acknowledgement

The Norwegian Ministry of the Environment, Ministry of Fisheries and Coastal Affairs and Ministry of Defence have for the period 2007-2010 financed the program. All research institutes involved have contributed with own financing through synergy with other projects. Thanks to researchers, research assistants and technical staff at the Norwegian Institute for Water Research (NIVA), Institute for Marine Research (IMR), the Geological Survey of Norway (NGU) and Akvaplan-niva for all you effort.

References

Bekkby, T., Bodvin, T., Bøe, R., Moy, F.E., Olsen, H., Rinde, E. 2011. Nasjonalt program for kartlegging og overvåking av biologisk mangfold - marint. Sluttrapport for perioden 2007-2010. (National program for mapping and monitoring of marine biodiversity in Norway. Final report for the period 2007-2010). NIVA Report LNR 6105, 31 pp. ISBN 978-82-577-5840-0. (Norwegian, English abstract).

Bekkby, T., Rosenberg, R. 2006. Marine habitaters utbredelse - terrengmodellering i Gullmarsfjorden. Report to the Västra Götaland administrary countu, Vattenvårdsenheten, Report 2006:07. ISSN 1403-168X. 33 pp. (in Norwegian/Swedish with English abstract).

Bekkby, T., Rinde, E., Erikstad, L.; Bakkestuen, V. 2009. Spatial predictive distribution modelling of the kelp species *Laminaria hyperborea*. ICES Journal of Marine Science 66(10): 2106-2115.

Bekkby, T., Rinde, E., Erikstad, L., Bakkestuen, V., Longva, O., Christensen, O., Isæus, M., Isachsen, P.E.. 2008. Spatial probability modelling of eelgrass Zostera marina L. distribution on the West coast of Norway. ICES Journal of Marine Science 65: 1093-1101.

Burnham, K. P., Anderson, D. R. 2001. Kullback–Leibler information as a basis for strong inference in ecological studies. Wildlife Research 28: 111-119.

Burnham, K. P., Anderson, D. R. 2004. Multimodel inference: understanding AIC and BIC in model selection. Sociological Methods and Research 33: 261-304.

Dolan M.F.J., Mortensen P.B., Thorsnes T., Buhl-Mortensen L., Bellec V., Bøe, R. 2009. Developing seabed nature-type maps offshore Norway: initial results from the MAREANO programme. Norwegian Journal of Geology 89: 17-28.

Elith, J., Graham, C. H., Anderson, R. P., Dudı´k, M., Ferrier, S., Guisan, A., Hijmans, R. J., et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29: 129-151.

Fielding, A. H., Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24: 38-46.

Guisan, A., Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135: 147-186.

Hastie, T. J., Tibshirani, R. J. 1990. Generalized Additive Models. Chapman and Hall, London.

Isæus, M. 2004. Factors structuring Fucus communities at open and complex coastlines in the Baltic Sea. Doctoral thesis, Department of Botany, Stockholm University, Sweden. 165 pp. http://www.aquabiota.se/publications/pdf/Avhandling_Isaeus.pdf.

Kelly, N. M., Fonseca, M., Whitfield, P. 2001. Predictive mapping for management and conservation of seagrass beds in North Carolina. Aquatic Conservation: Marine and Freshwater Ecosystems 11: 437-451.

Lehmann, A. 1998. GIS modeling of submerged macrophyte distribution using generalized additive models. Plant Ecology 139: 113-124.

Lehmann, A., Leathwick, J. R., Overton, J. M. 2004. GRASP v.3.1. User's Manual. Swiss Centre for Faunal Cartography, Switzerland.

Lehmann, A., Overton, J. M., and Leathwick, J. R. 2003. GRASP: generalized regression analysis and spatial prediction. Ecological Modelling 160: 165-183.

Rinde, E., Rygg, B., Norderhaug, K.M., Nygaard, K., Longva, O., Olsen, H., Bordvin, T., Steen, H. 2007. Veileder til startpakkene for kartlegging av marint biologisk mangfold (Instruction manual for mapping of marine biological diversity), NIVA reports for all counties included in the project.

Swets, J. A. 1988. Measuring the accuracy of diagnostic systems. Science 240: 1285-1293.

Wilson, M. F. J., O'Connell, B., Brown, C., Guinan, C. J., Grehan, A. J. 2007. Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. Marine Geodesy 30: 3-35.