Trabajo Fin de Grado

Diatoms, filamentous algae and macrovegetation distribution modeling in Gulf of Bothnia

Autor

Ander Lopez Conde

Director

Jarno Vanhatalo

Ponente

Alfredo Serreta Oliván

Faculty of Biological and Environmental Sciences
Escuela Politécnica Superior de Huesca
2015

ABSTRACT

We have evaluated the distribution and extent of sea bottom vegetation divided in three groups: Diatoms, macrovegetation and filamentous algae in the Gulf of Bothnia, northernmost area of the Baltic Sea, and relate the increment in the distribution of the filamentous algae with the increasing problem of the eutrophication. The distribution modeling of these groups of species has been done by combining data from species abundance (distribution data) with GIS environmental raster variables based of environmental information in a binomial model to predict the spatial probability of each group of species using MatLab and the GPstuff toolbox. From all the variables used the most important ones were the bottom type and variables related to the exposure of an area (weighted fetch, number of islands and distance to shallow waters) to explain the predicted distribution of the group of the species. It is shown that the main group of species in the Gulf of Bothnia is the filamentous algae, with and elevated predicted probability in almost all the Gulf of Bothnia. Preferring hard bottoms like rock or stones and exposed areas, the number of filamentous algae is increasing every year, reducing macrovegetation populations into more protected areas. The number of nutrients and filamentous algae has increased in the last decades. We discuss a relation between evolution of eutrophication and the increase of filamentous algae, which follows the same south to north and west to east gradients, been the south and west more eutrophied. This work aims to be a tool to assess the environmental protection and coastal management of eutrophication by predicting the probability of presence of the different vegetation groups and analysing the relation of these groups with the eutrophication.

KEY WORDS: Diatoms · Macrovegetatation · Filamentous algae · Eutrophication · Modeling · Distribution · Gulf of Bothnia

INTRODUCTION

Eutrophication

Eutrophication of coastal waters is happening worldwide, creating more areas suffering hypoxia, and consequently creating more “dead zones” (Diaz & Rosenberg 2008), and therefore causing a loss of habitat and spawning areas, elimination of benthic animals and alteration of the food chains. In the Baltic Sea, eutrophication is a known actual problem.

The human activities surrounding the Baltic Sea are numerous because of the high population in it. Activities such as agriculture, municipal sewage or industries are common, and atmospheric deposition and nitrogen fixation have created an excess in the nitrogen and phosphorus in the waters.

During the last fifty years, the nutrient amount in the Baltic Sea basins has increased approximately the double even the triple, and the rates of the biogeochemical processes have increased even more (Savchuk et al. 2008). Actually, some reconstructions shown in *Savchuk et al. (2012)* says that in the last half century the Baltic Sea has received approximately 50 million tonnes of nitrogen and 2.25 million tonnes of phosphorus from the land and atmosphere.

In the Gulf of Bothnia the surrounding activities are lower due to its northernmost position, and therefore the eutrophication in it is lower than in the rest of the Baltic Sea. However, eutrophication is spreading, especially because of inflow of organic material from rivers.

As a result of the excessive nutrients loads by human activities, and amplified by the factors that the Baltic Sea has a slow water renewal and strong stratification, eutrophication is now a serious problem.

Current status

The increase of eutrophication and with it the increase of the filamentous algae is a known actual problem in the Baltic Sea (Raffaelli et al. 1998). In the publication *HELCOM (2010)* they state that with exception of the Bothnian Bay all the open waters of the Baltic Sea are affected by eutrophication. The same way, all the coastal areas with the exclusion the Gulf of Bothnia are also affected by eutrophication.

This has modified all the trophic levels in the Baltic Sea ecosystems, changing the structure and the species composition of different communities (HELCOM 2002). During the last decades algal vegetation has increased (Eriksson et al. 1998), and an increase in the concentration of nutrients in the Gulf of Bothnia has been noticed during the last 30 years, doubling its quantity (HELCOM 1996, Karjalainen 1999).

The eutrophication in the Gulf of Bothnia is spreading, following a south-north gradient, which is a reflection of how eutrophication in large scale is evolving (Lundberg et al. 2009).

Objectives and justification

In management and conservation planning, estimates of species distribution are widely used. In the assessment and evaluation of the eutrophication status of the Baltic Sea and also in the Gulf of Bothnia, useful tools are indicators. Some well know indicators for aquatic vegetation are the distribution of Bladderwrack (*Fucus vesiculosus*) and distribution of Eelgrass (*Zostera marina*), as well as the proportion of opportunistic species in the vegetation communities (HELCOM 2009). Both species mentioned above are considered as macrovegetation, while lot of the opportunistic species are considered filamentous algae.

Therefore, knowing the distribution of these two groups of species can help in the assessment and the evaluation of the eutrophication in the Gulf of Bothnia, which is the less eutrophied zone of the Baltic Sea. With this objective a useful tool is the species distribution modelling, which can give predictions at unsurveyed locations and explain the environmental variables effect in that prediction.

Species distribution modelling

Species distribution modelling are numerical tools that combine observations of species occurrence or abundance with environmental estimates (Elith & Leathwick 2009). Distribution modeling is used to predict distributions in a concrete study area, from occurrence or abundance data from the same area or being extrapolated from a different area. In the same way it is possible to extrapolate the data along the time to create a prediction in the future.

Nowadays distribution modeling is used for all kind of landscapes (terrestrial or aquatic) for describing patterns and making predictions in conservation or management processes. Surveying big areas is expensive and time consuming, and modeling can give us predictions of unsurveyed areas and show which environmental variables are more for important for each species.

The reliability and robustness of a model falls in the relevant predictors and modelling method, consideration of scale, the extent of extrapolation and how they are considered the interplay between environmental and geographic factors (Elith & Leathwick 2009). The measure of realism of the model is based in these factors.

One of the weak points of the species distribution modeling is the links between the ecological knowledge and the modeling practice, especially when it is referred to the biotic interactions. Improving this and reducing models uncertainty are the big challenges in the ecological modeling.

A wide variety of techniques allows modeling to be used in diverse applications. The degrees of success is different for different models and techniques, but species distribution models have a good performance predicting natural distributions of species, especially when the data has been correctly surveyed and the selection of the model and relevant predictors have been appropriately selected, showing a good predictive capability. On the other hand, extrapolation in time or space could be much more challenging, and using incomplete or inadequate data could bring us to wrong results and false predictions (Elith & Leathwick 2009).

Species distribution modeling combines data from species occurrence or abundance (distribution data) with environmental variables based on environmental information as well as a spatial factor of study area. The results can shows us the importance of each variable and make us understand better the species we are studying, as well as creating a prediction of distribution. But for getting the correct results and a reliable prediction, a good model has some key steps that we have to follow in the modelling practice.

First of all the base of the analysis will be to have a relevant and as complete data as we can, as accurate as it is possible. We also will have to take into account the correlated predictors variables, and select the proper algorithm for our case. Some algorithms could be more suitable to some analysis (for example terrestrial landscape or aquatic landscape). Then we have to adequate the model to the training data, and evaluate the model to see if it is reliable and realist, and see if the response functions fit with our model. In that evaluation we have also to test the predictive performance. Finally we have to map the predictions to a geographic space (Elith & Leathwick 2009).

Relation between modelled species and eutrophication

The structure and species composition of aquatic bottom vegetation communities are affected by eutrophication, favouring filamentous annual species due to increase in the amount of nutrients and the diminishing capacity of penetration of the light due to decrease of the water transparency, changing the distribution of the species.

The hypoxia created by the eutrophication also creates a more attractive environment for filamentous algae. While larger vegetation has a higher request of oxygen no longer survive, they are replaced by smaller and fast-growing species like some species of filamentous algae.

Consequently, the eutrophication has incremented the growth of annual filamentous algae in the whole Baltic Sea and also in the Gulf of Bothnia, in detriment of the macrovegetation biomass, depth and geographic distribution (Nielsen et al. 2002). In this sense, it is possible to create a relation between the eutrophication and the distribution of different groups of vegetation.

MATERIAL AND METHODS

Study area

The Gulf of Bothnia is located in the northernmost extension of the Baltic Sea. Surrounded by Finland's west coast and Sweden's east coast, the gulf is 80–240 km wide and 725 km long but the study area only covers the northernmost 600 km. The average depth is 60 m and the maximum depth is 295 m. In the south it is almost closed off by the Åland Islands. It is composed by the Bothnian Sea and Bothnian Bay.

The coast contains plenty of archipelagos and estuaries and numerous rivers for both coast sides. The most common coast type is exposed open shores, vulnerable to the strong wind even though there are no powerful tides in the study area.

The freshwater entering via rivers influences the Gulf of Bothnia decreasing the salinity from the north to the south (Håkansson et al. 1996). This influence creates a gradient of salinity which in the north can result in a very low salinity, less than 0,5%, while in the south the water salinity is similar to the rest in the Baltic sea. The water residence time is ca. 7 years (Algesten et al. 2006).

The Gulf of Bothnia is usually covered by ice in normal winters, having an ice covering time duration between 60 and 194 days, from October-November to April-May (Veneranta et al. 2013)

Sampling design and method

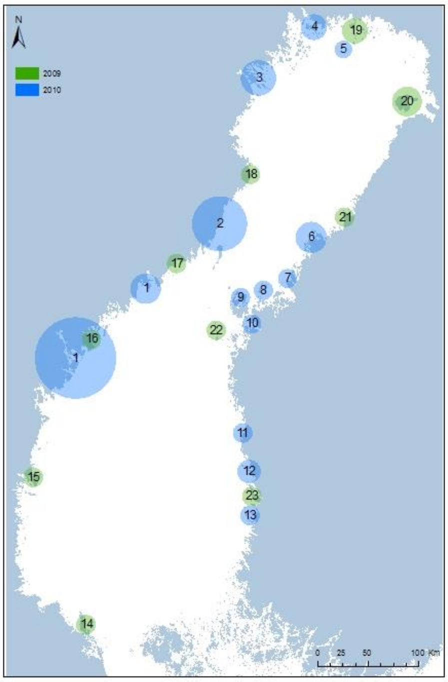

Figure 1: Map of Gulf of Bothnia showing the sampling points in 2009 and 2010 (Finnish game and Fisheries Research Institute).

The sampling and data collection was realized by the Finnish game and Fisheries Research Institute.

The classification of the vegetation types has been done along the coasts of the Gulf of Bothnia in a total of 225 sample points, divided into 13 areas. The sampling sites were placed with maximum based on wind exposure, ice winter length and shoreline length. For stratification, these variables were combined to a new, eight class layer. The minimum distance between sampling sites was set to 1.5 km. The sampling timing was set to 1-2 weeks after ice break-up in winter independent of latitude.

The location of each sample was recorded both in a manual GPS and in GPS included in the aquascope.

In each sampling site pictures were taken with a 12 MP digital camera which was attached to an aquascope. At each sampling site 5-13 photos were taken with a minimum distance of 1 m. The legs of water scope had white plates to correct the tone of light. The depth of each sampling point was 0.3 m, the length of the aquascope leg, which represents the shallowest littoral area. Each photograph was determined by 16 points and the total number of photos in data is 2427.

For the interpretation of the photograph, a CPCe software (*http://www.nova.edu/ocean/cpce/*) was used, where it was added the species list of Bothnian Sea/Bay and suitable classifications. For the classification of the vegetation 16 points were used in each image, marking each point with a figure if the vegetation is present and a cross if not. During the interpretation, if there were several different plants in a field, the sampling point shows the plant with the greatest surface in the area. An exception applies diatoms, which are only indicated if the area is no other vegetation

From the classification made in the sampling (species, size, condition and bottom type), the species were grouped into diatoms, filaments (filamentous algae) or macrovegetation (eg. actual plants) for the statistical analysis as it is shown in the Table 1.

Table 1: Species grouped into diatoms, filamentous algae or macrovegetation.

Diatoms	Filamentous algae	Macrovegetation
Diatoms are not possible do define on species from this kind of imagery	– Cladophora glomerata – Ulva intestinalis – Furcellaria lumbricalis – Pilayella littoralis – Ectocarpus siliculosus – Dictyosiphon/Stictyosiphon sp.	– Potamogeton sp. – Potamogeton pectinatus – Zannichellia palustris – Phragmites australis – Ranunculus circinatus – Ceratophyllum sp. – Myriophyllum sp. – Nitella sp.

Each data entry was based on the vegetation type and the location of the sampling position. Also was determined the known environmental variables of the sampling for each sampling station (fetch, the amount of shoreline per surface area, distance to the depth of the zone surrounding the low surface area, the length of the ice season, nutrients, distance from rivers, etc.)

Due to the wide area covered by the study area, there are potential differences between the coastal areas which can be appreciated in the analysis of the data.

Environmental GIS variables

Variables measured in the field

Some of the variables were measured in the field. The ones used in the statistical analysis are summarized in the Table 2.

Table 2: Environmental variables measured in the field used to analyse the probability distribution of the diatioms, filaments and macrovegetation (Vanhatalo et al. 2012; Veneranta et al. 2013).

Variable	Description	Unit	Value range
Shoreprofile	Shoreline profile	Classification	I = open water, II = steep, III = gently sloping with steep edge, IV = gently sloping, V = shallow, VI = Shallow with sand bar
Bottom	Bottom type	Classification	Soft, silt, sand, sand/stone, stones, rocks (>30cm *D*)
Bottomcov	Bottom coverage	Classification	No coverage, <10%, 10-25%, 25-50%, >50%

GIS variables

All the variables used for this analysis were the same used in the *Veneranta et al. (2013)*. The data proceed from the Finnish Meteorological Institute (FMI), Finnish Environmental Institute (FEI), Swedish Environmental Protection Agency (SEPA), HELCOM or Finnish game and Fisheries Research Institute (FGFRI).

All GIS analyses were performed using ArcGIS (ArcMAP 10.1). All the variables were converted to 300 m.

Table 3: Environmental GIS variables used to analyse the probability distribution of the diatoms, filaments and macrovegetation (Vanhatalo et al. 2012, Veneranta et al. 2013).

Variable	**Description**	**Unit**	**Value range**
Shoprofile	Shore profile	NA	1-6
depth	Depth	m	0.1-47
FE300W	Fetch weighted	m	90-260 084
FE300ME	Fetch mean	m	9-124 247
D20M	Distance to 20 m depth curve	m	0-27 473
LINED	Shore line density in a circle of 3 km	km/km^2	0-323
ISLANDN	Number of islands in a circle of 10 km	I	0-549
DSAND	Area weighted distance to sand	I	0-116
DSHALLO	Area weighted distance to shallow	I	0-6168
PE900	Water area per shoreline length	I	0-1046
SALSPR	Spring salinity	psu	0-6.2
SALWIN	Winter salinity	psu	0-6.0
ICEWIN	Length of ice winter (2009)	m	0-24
ICELAST	Last ice cover (concentration <30%)	wk	0-21
EKOSTAT	Ecological status of coastal waters	type	0-4
PHOSP	Dissolved inorganic phosphorus	I	14-49
NITROG	Dissolved inorganic nitrogen	I	10-49
CHLA	Chlorophyll a phytoplankton concentration	I	11-45
SECCHI	Secchi depth	I	10-43
RIVERS	Distance to the nearest river mouth	m	0-56 700
SHAREA	Shallow area	index	0-1 122
BOTCLS	Bottom class in shallow areas	NA	1 (sand), 2 (sand and mud), 3 (sand and rocks), 4 (other classes)

Species distribution model

The first predictions were modelled using a presence-absence Bernoulli observation model for presence absence observations y with an occurrence probability π, following *Vanhatalo et al. (2012).* It is based in the relationship of occurrence probability π to the environmental variables *x* and the spatial parameters *s*:

$$Y(x,s) \sim Bernoulli\ [\pi(x,s)]$$

$$\pi(x,s) = g\ [f(x) + \rho(s) + \alpha]$$

Where, *f(x)* is the predictive function dependent of the environmental variables, *ρ(s)* is the spatial component, α is the intercept and g, the link function. Link function g is logistic in case of Bernoulli observations.

The model

The binomial model states that:

$$y_{i\,|z_i} \sim Binomial\ (z_i, \pi(x_i, s_i))$$

In a simple binomial model, the objective is to estimate the distribution of the unknown population proportion from the results of a sequence of “Bernoulli trials”; which is the data $y_{i,1}, \dots, y_{i,z}$ which are either a 0 or 1. The binomial distribution creates a model for the data from a sequence of *z* exchangeable trials from a population were each trial gives rise to two possible results, 0 for failure and 1 for success.

Because of that, data can be summarized in the number of success in trials, which is the result variable, *y*. The parameter z represents the proportion of successes or, in other words, the probability of success in each trial (Gelman et al. 2003).

$$y_i = \sum_{j=1}^{z_i} y_{i,j}$$

The main difference with the previous model is the addition of discrete and continuous covariates. In this case, instead of having an absence-presence model we can measure the abundance using the binomial model.

As different as before, in this time y_i is the number of count points where species is present in site i, and z_i is the total number of count points in site i. This allows us to take into account the abundance for the modelling, because it is important to differentiate the occasional presence with a numerous abundant presence both to select the variables effect and for the prediction created.

The selection of the variables with the higher effect in the prediction was made according to the average predictive comparison test (Vanhatalo et al. 2012) which is summarized in figures A, B and C, in which is shown the effect of each variable, and the posterior observation of the response curves of each individual variable.

In the generalized linear model, which our model extends, the importance or the strength of each variable is defined by β_x:

$$\pi(x) = g\ [\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots]$$

Once we add the location (spatial correlation), we have to use the Gaussian Process model described until now, changing $\beta_1 x_1$ with a non-linear function *f(x)* and giving prior to it. As we mentioned before, in a generalized linear model the strength of each variable is defined by β_x. In a non-linear model with interactions between variables, the difference in the strength of the effect is based in the values of the variables. The predictive effect of each variable vary due to the interaction between variables, because the predictive effect of one variable may depend on the value of other variables.

For this non-linear model, the strength of the effect of a variable in the prediction can be calculated with the average predictive comparison (Gelman & Pardoe 2007). The APC values in Figures 2 to 4 are analogous to absolute values of beta in the generalized linear model.

Covariance and hyperparameters

GP model defines the probability distribution over functions and it is defined by mean and covariance function. Different kind of covariance exists, and each one has a number of free hyperparameters, whose values also need to be determined.

The specification of the covariance function determines what type of latent functions *f(x)* are possible, and consequently, the selection of the properties of the latent functions, such as the variability or smoothness (Vanhatalo et al. 2012).

Therefore, choosing a covariance function can be understood as a model selection, and the selection of the hyperparameters and its value as the training of a Gaussian process model (Rasmussen & Williams 2006).

The predictive and spatial functions are given a Gaussian process prior with neural network and exponential covariance functions, expressed in the next way:

$$f(x) \sim GP\left(0, k_f(x,x')\right) \quad \rightarrow \quad k_f(x,x') = neural\ network\ covariance\ function$$

$$\rho(s) \sim GP\left(0, k_\rho(s,s')\right) \quad \rightarrow \quad k_\rho(s,s') = exponential\ covariance\ function$$

$$k_f(x_i, x_j) = \frac{2}{\pi} sin^{-1}\left(\frac{2x_i^T \sum_{nn} x_j}{\left(1 + 2x_i^T \sum_{nn} x_i\right)\left(1 + 2x_j^T \sum_{nn} x_j\right)}\right)$$

$$k_\rho(s,s') = e^{-\sqrt{\sum_{d=1}^{2}(s_d s'_d)^2/l_d}}$$

$$\alpha \sim N(0,10)$$

$$f(x) + \rho(s) + \alpha \sim N(0,10 + k_f + k_\rho)$$

We use the neural network covariance function for *f(x)* which is a good choice for it because it allows interaction between environmental variables and has good extrapolation power. For the spatial random effect we use an exponential covariance function, which is a usual choice for modelling spatial random fields (Vanhatalo et al. 2012).

Spatial autocorrelation is an important concept when our model relates environmental and geographic variables. Values in a spatial random field are spatially correlated. In its most basic form this means that adjacent values do not differ as much as values that are further apart. Usually, the values are defined over a continuous domain and the spatial random field is defined by function valued random variable. In our model the spatial random field is defined by the exponential covariance function.

Different hyperparameters give different explanations of the data, so they are of great importance when we are trying to understand the data. To estimate the parameters of the covariance function we searched their maximum a posterior estimate with gradient based optimization where we approximated the marginal likelihood of the model with expectation propagation algorithm (Rasmussen & Williams 2006, Vanhatalo et al. 2012).

As we mentioned before, the covariance functions usually have some free parameters. In the covariance functions the parameters are *l*, the length-scale parameter and σ^2, the signal variance parameter, which can be varied, are call hyperparameters in the Bayesian hierarchical model (Rasmussen & Williams 2006). The length-scale parameter defines the declination of the correlation with the distance and the signal variance parameter shows the variability of the spatial field (Vanhatalo et al. 2012).

Posterior inference

The EP algorithm was used to approximate the posterior distribution of the covariate functions. Basing in the Bayes theorem, we calculate the posterior distribution of *f(x)* and $\rho(s)$, and then use both to calculate the probability of occurrence. The model then, concretely the predictive function extrapolates the occurrence probability of the sampled places to unsampled areas, and the spatial component models the spatial structure when the environmental variables cannot.

Even if the distribution of the group of species is mostly defined by the environmental factors, a properly specified model with a correct number of predictors will display minimal spatial autocorrelation. The spatial random field is more important than the environmental variables when some of the key environmental variables are missing, the predictive model is mis-specified or the geographic factors are much more influential (Elith & Leathwick 2009).

All the results were computed using MatLab and the GPstuff toolbox developed by *Vanhatalo et al. (2012).*

RESULTS

Selection of variables

Based on the APC test we can observe the influence of variable in the prediction in the next Figures 2 to 4 for each type of vegetation. For the diatoms the most influential variable is the bottom type, the percentage of covered bottom, the distances to the sand and the distance to shallow water as well as the number of island surrounding the sampling area.

In the other side, the most influential variables for the filamentous algae are the bottom type once again and the fetch weighted. The percentage of bottom covered, the number of surrounding islands and the water area per shoreline length have also a relevant effect in the prediction. Finally the depth and the distance to the 20 metres curve affect also the prediction.

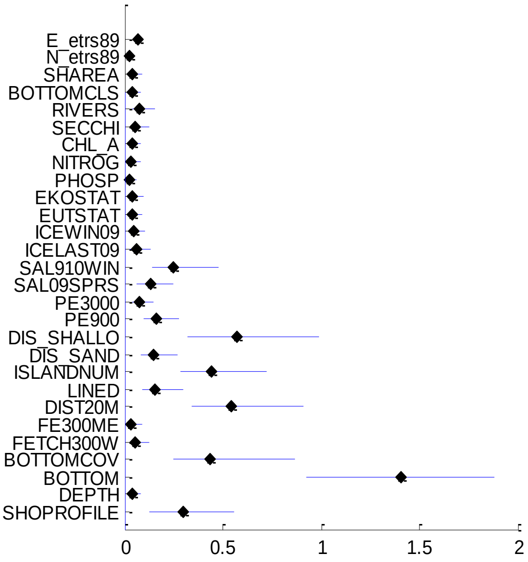

Figure 2: Strength of the effect of the variables in the predictive distribution for the diatoms.

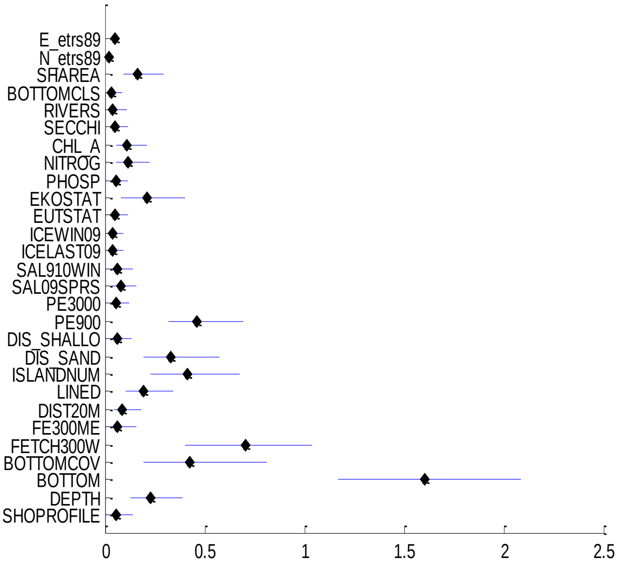

Figure 3: Strength of the effect of the variables in the predictive distribution for the filaments.

The prediction for the macrovegetation is extensively affected by the type of bottom, the distance to the 20 metres curve, the fetch weighted, the distance to the sand and bottom classification. The percentage of covered bottom has less effect than the others, but still affects the prediction.

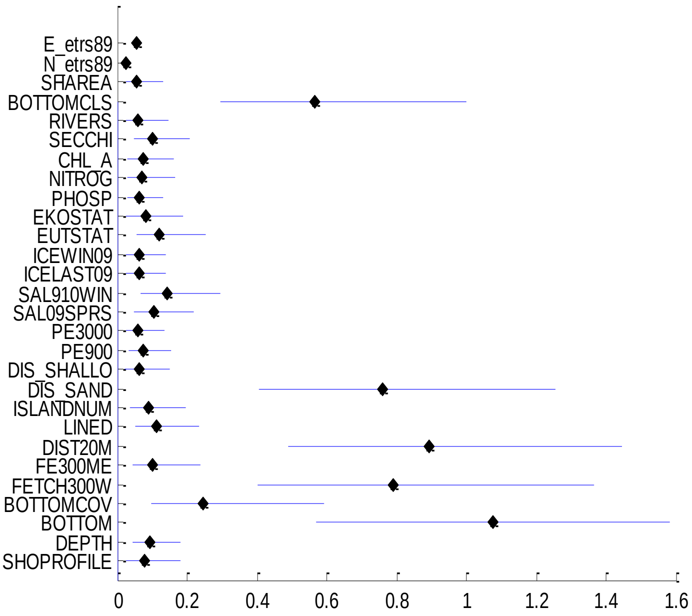

Figure 4: Strength of the effect of the variables in the predictive distribution for the macrovegetation.

Response curves

In the appendix I there are the response curves of all the variables, but in the following figures we have summarize the response curves of the most important variables. We can observe a strong relationship between the bottom type and the presence of diatoms, as well as with the distance to shallow waters. With stone/rock bottom the probability decreases, and also when the distance to shallow water increases. On the other hand, the probability slightly increases when the percentage of covered bottom increases, and the same happens when number of surrounding island increases and the distance to the 20 metres curve gets stronger.

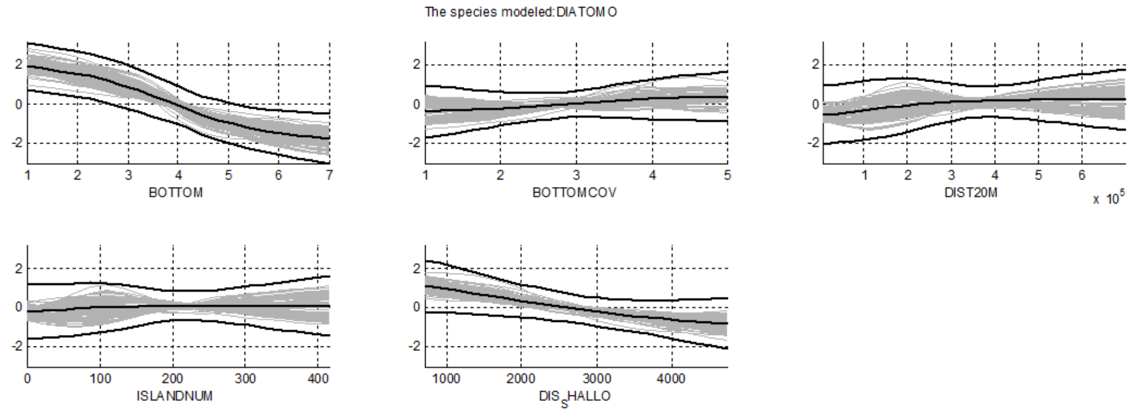

Figure 5: Response curves of binomial modelling of the DIATOM using only the covariates which showed a high influence in the prediction in the binomial modelling with all the covariates.

For the filamentous algae, the bottom type has a strong effect to the prediction. When bottom is more similar to a stone/rock bottom type, the probability of presence gets much higher. The increase in the fetch weighted also increases the probability even in a not so strong way. The percentage of covered bottom and the number of surrounding islands increase slightly the probability when they are higher, and the water area per shoreline length remains neutral but shows a high variability, affecting the prediction in a different ways. The distance to sand decreases the probability also in very soft way.

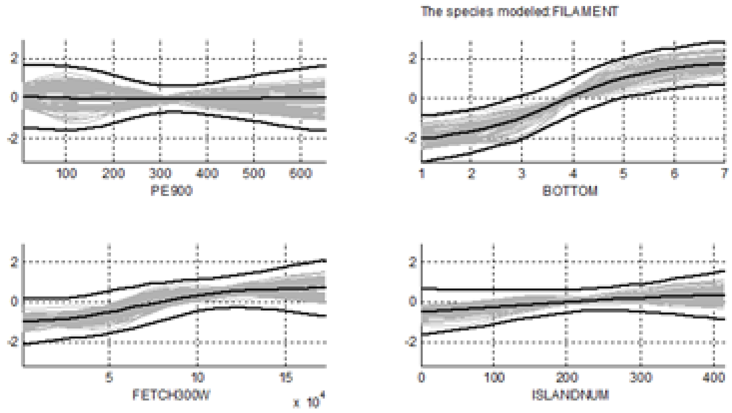

Figure 6: Response curves of binomial modelling of FILAMENT using only the covariates which showed a high influence in the prediction in the binomial modelling with all the covariates.

The macrovegetation is clearly affected by the bottom type, the fetch weighted, the distance to the 20 meter curve and the distance to the sand. In the first two variables, the probability decreases in stone/rock bottoms, and also when the weighted fetch is higher. In the case of the distance to sand and the distance to the 20 meter curve is in the just the contrary. When the distance is higher the probability is increased. The same happens with percentage of covered bottom, but the strength of this variable in the prediction is clearly much lower. The bottom class in shallow waters also affects the prediction slightly.

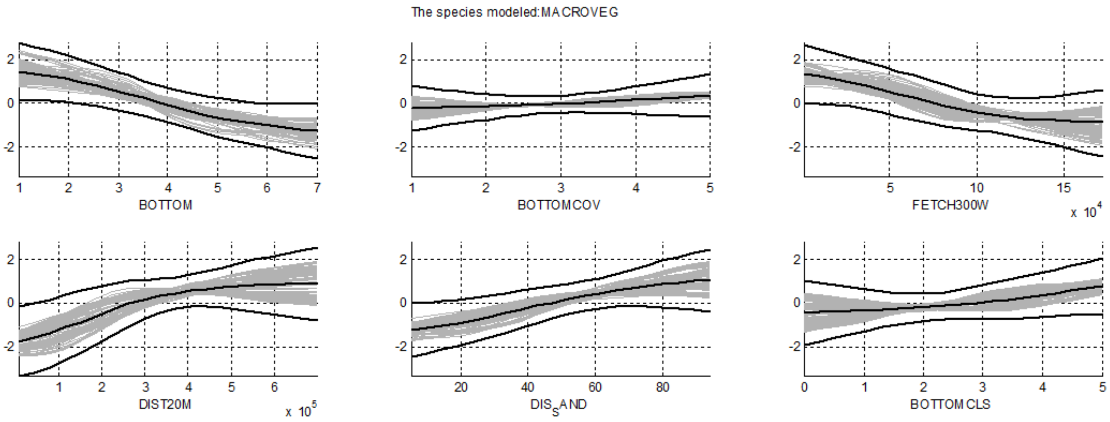

Figure 7: Response curves of binomial modelling of MACROVEG using only the covariates which showed a high influence in the prediction in the binomial modelling with all the covariates.

Predicted distribution of the species

The prediction probability of the three groups of species in the Gulf of Bothnia is shown in the following maps, in a scale from 0 to 1 with the same scale of colours. It 's obvious that the most extended group of bottom vegetation is the group of filamentous algae, which has a probability higher than 0.8 in almost the whole gulf. For the prediction of the diatoms, on the other hand, exists a high variability in function of the area. Evidently the macrovegetation has the lower probability in the whole gulf, having a high probability only the east coast of the Bothnian Sea, especially in the north part and close to the sea shore.

We can observe a big difference between different areas. The diatoms have a medium-high probability in the east coast of the Bothnian Sea, while in the west coast the prediction is generally low. In the Bothnian Bay, the probability of the diatoms is around 0.5. The prediction of the filamentous algae does not vary so much, there are only three observable areas where the probability descends from high to medium: in both sides in the north of the Bothnian Sea and in the north of the Bothnian Bay.

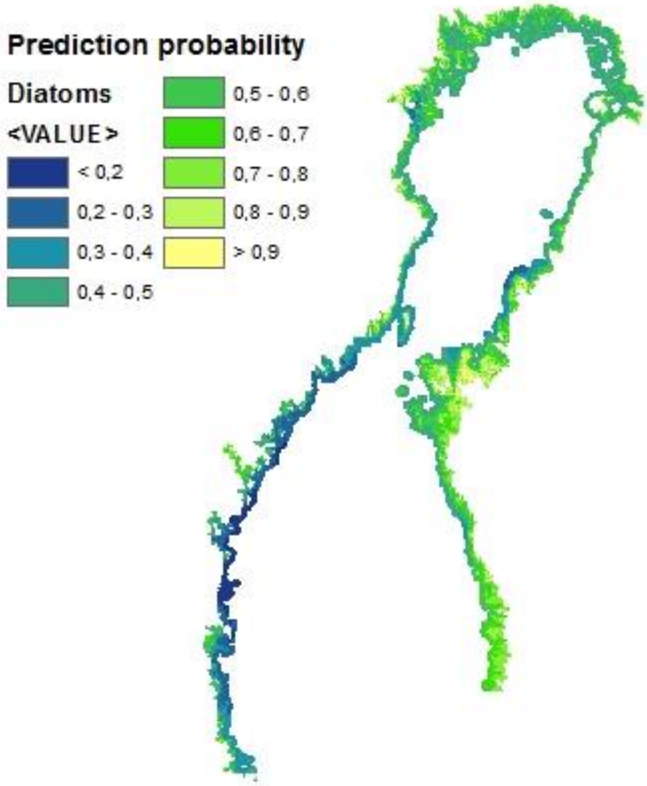

Figure 8: Distribution of the Diatom with prediction probability, using only the variables which showed a high influence in the prediction in the binomial modelling with all the covariates.

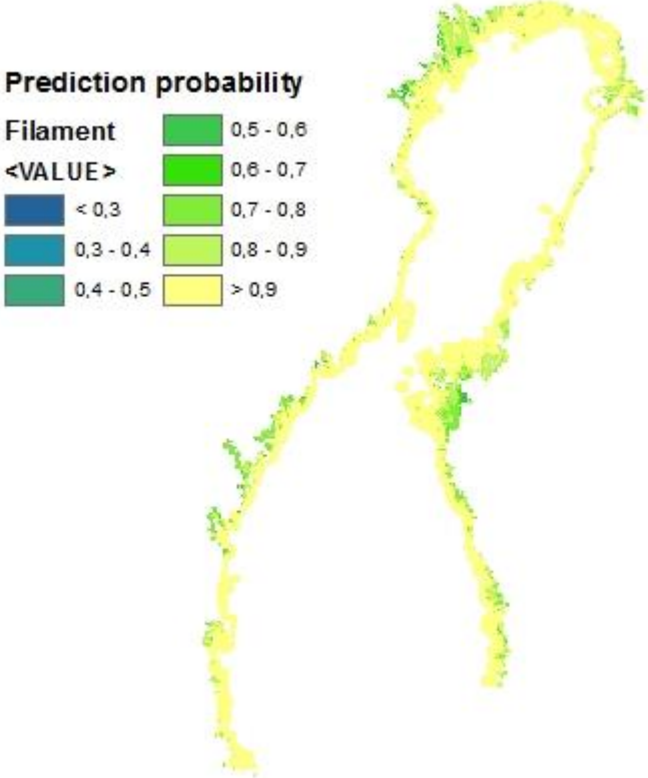

Figure 9: Distribution of the Filament with prediction probability, using only the variables which showed a high influence in the prediction in the binomial modelling with all the covariates.

The prediction of macrovegetation in the figure 12 also has a strong difference between the different areas. The probability is generally low, but we can observe some medium even high probabilities in the east coast of the Bothnian Sea.

We have noticed also a relation between the distance to the sea shore and the prediction probability.

The macrovegetation increases its probability in areas close to the sea shore. Not only for the macrovegetation, we can observe that the probability for the diatoms also increases in areas closer to the coast, while the filamentous algae increases its probability in areas further to the coast.

The results additionally show an inverse relation between the filamentous algae and the macrovegetation, in which clearly we can see that the only areas with high probability for the macrovegetation coincide with the areas in where the filamentous algae reduce its probability.

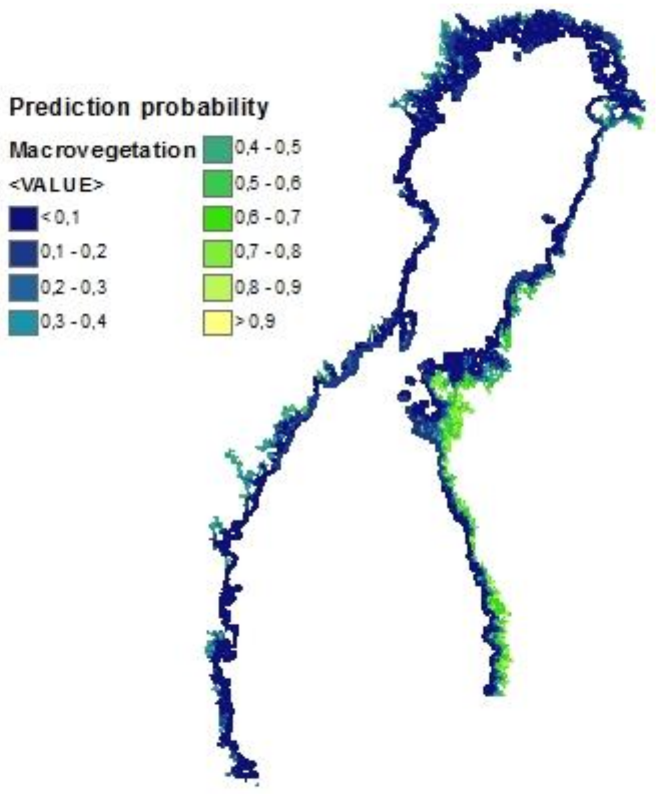

Figure 10: Distribution of the Macroveg with prediction probability, using only the variables which showed a high influence in the prediction in the binomial modelling with all the covariates.

DISCUSSION

Table 4 summarizes the most influential variables in the prediction for each group of species. The bottom type is the most influential one for the three of them, and the percentage of covered bottom is also present in the three groups. As we can see, macrovegetation and filamentous algae share four variables, but if we look in the curve responses we can see that three of the four variables response are opposite for macrovegetation and filamentous algae. The only variable with a similar response is the percentage of covered bottom, which is the one with less effect on the prediction and the slope is almost zero.

Table 4: Summarizing table of the most influential variables for each group of species using a binomial model.

VARIABLES	DIATOM	FILAMENT	MACROVEGETATION
BOTTOM	X	X	X
BOTTOMCOV	X	X	X
BOTTOMCLS			X
CHL			
DEPTH			
DISAND		X	X
DIST20M	X		X
DSHALLO	X		
FE300ME			
FETCH300W		X	X
ICELAST09			
ISLANDIUM	X	X	
PE3000			
PE900		X	
SAL910WIN			
SECCHI			
SHAREA			

In the diatoms, the variable with the higher effect is the bottom type, and the response curve indicates a preference for soft and sandy bottoms. However, other publications (Snoeijs 1994, Busse & Snoeijs 2006) show that diatoms prefer hard bottoms of rock or stones. This difference in the response of the variable maybe causes by a mistake in the interpretation of bottom type and diatom appearance in the original classification.

The distance to shallow water is the second variable with higher effect in the prediction. The response shows a strong decrease in the probability when the distance to the shallow water gets higher. The type of shore profile preferred by the diatoms is the type five (shallow waters), and the response of both variables are according to the generated distribution map of the diatoms, where we can see the higher probability in the places closest to the shore, in shallow waters were the large numbers of living diatoms overwinter beneath the ice on benthic substrates (Kingston et al. 1983).

The distance to 20 m curve describes actually the surface area of relatively shallow coast. The higher distance, the more we have shallow, productive area. In similar way, the higher the island count per area is, the more shallow the area might be. The shallow and structurally complex area warms up early - thus being an productive area in early spring. The dense and shallow archipelago areas work also as a kind of filters, thus enhancing the productivity.

The filamentous algae shows a logic response according to the biological knowledge. This group shows a preference for rock and stone bottoms according the bottom type variable. The increase in the weighted fetch variable also shows an increase in the probability, which is in concordance with the bottom type variable, because both variables are related with the exposition of areas to the wind or waves. Taking into account the responses of the two variables we can say that filamentous algae shows a preference for exposed areas. This is explained because hard bottoms are usually in very exposed areas, like older parts of the archipelagos, and therefore explains the increase of probability for filamentous algae in wind exposed areas. This was confirmed by a study made by *Einav et al. (1996*) which showed that more exposed areas in islands had a larger stock and species richness of algae than areas with less exposure.

Related to the previous point, the higher the number of islands is, the higher the probability becomes for filamentous algae (Bonsdorff et al. 1997). As the analysis has been made in a large scale, we cannot focus in the regional scale to compare the number of local islands with the prediction. However, the southern part of the study area has more islands than the northern part. This means that the Bothnian Bay is mostly composed by open areas with not so many islands, so in the Bothnian Sea there are more islands, which increases the prediction in this area the for filamentous algae.

The other variables affecting the prediction probability for the filamentous algae like distance to sand and perimeter don't have a clear significance. In general, we can observe a high probability of the filamentous algae in the whole study area, with the exception of none exposed areas like small gulfs or protected bays.

As we have mentioned before, the prediction for the filamentous algae is opposite to prediction of the macrovegetation. This is something that we can observe in the maps very clearly. As the scale is coarse and the resolution is not enough, actually we have a higher prediction of the macrovegetation in the inner parts of the archipelago areas or small bays, areas with low exposition and well protected from the wind and the waves, unlike with the filamentous algae.

The bottom type preferred by the macrovegetation is soft or sandy bottoms, as well as places where the weighted fetch is low. Once again both variables are in concordance, because both variables are related with the exposition of areas to the wind or waves. Contrary to what I mentioned earlier, in this case we observe that soft or sandy bottoms are in non-exposed areas, zones were the probability for macrovegetation will be higher. Usually macrovegetation is mostly in eutrophic areas were we have lot of nutrients and were the production is high like in the littoral areas.

For macrovegetation, the higher the distance from sand is the higher the influence on prediction is. This is probably caused by the properties of the Gulf of Bothnia. In Bothnian Bay and Swedish coast of Bothnian Sea, the relative area of sandy shores is high. These are also exposed coasts, and thus there are no suitable places for macrovegetation to grow, as in archipelago or sheltered estuary areas.

The dominance of filamentous algae is a sign of eutrophication in exposed areas, where the water is rich in nutrients and the production is high. The increase of eutrophication and with it the increase of the filamentous algae is a known actual problem in the Baltic Sea (Raffaelli et al. 1998). Big changes have been originated along all trophic levels in the Baltic Sea ecosystems by the eutrophication during the last years (HELCOM 2002), and the annual algal vegetation had increased during the last decades (Eriksson et al. 1998). The same way, during the last 30 years, a small increase in the concentration of nutrients in the Gulf of Bothnia has taken place (HELCOM 1996, Karjalainen 1999).

Eutrophication of the Baltic Sea has increased both the growth of annual filamentous algae and the rate of sedimentation. Together these factors may have a detrimental effect on the macrovegetation populations, because as we have seen, they share various variables which affect the distribution of both groups, but with an inverted response curve.

The environmental main reasons for this are that the increased competition from annual, fast-growing filamentous algae, which gets advantage from the increase of nutrients and the reduced light conditions caused by eutrophication and the decrease of oxygen caused by the hypoxia.

The oxygen deficit leads to changes in benthic communities; were larger vegetation, with higher request of oxygen no longer survive, and they are replaced by smaller and fast-growing species that live on the sediment surface and can tolerate low concentrations of oxygen. Even if the oxygen deficit is not so high in the Gulf of Bothnia, the filamentous algae are favoured because they can survive while macrovegetation has problems to get enough oxygen.

With this increase in filamentous algae, the consequences a lead to light deprivation for aquatic macrovegetation, reducing their biomass, depth and geographic distribution. This means that the loss of species like eelgrass and bladder wrack which provide substrate for feed, reproduction, and shelter for associated fauna, will influence and change the coastal Baltic ecosystem, and thereby coastal fish catches. More precisely, *Borg et al. (1997)* emphasised that eutrophication-induced changes in habitat structure, such as an increased dominance by filamentous algae, could alter the availability of predation, refuges and foraging habitats for other species.

Regarding the Baltic fish communities, eutrophication is one of the major factors affecting the composition and development of the communities and causing changes in fish community structure and function (Lappalainen 2002). Depending of the fish species the eutrophic areas may favour the abundance of some fish species, for example these areas are important for the reproduction of coastal fish species like pike, roach or berch. However, other species prefer more oligotrophic waters. This is the case of several species which depend on seagrass or higher algae, which may disappear owing to the effects of eutrophication (HELCOM 2006). Sedimentation also affects the reproduction of some fish species, as coregonids. Exposition affects sedimentation, and in eutrophic and exposed areas the high sedimentation can affect the spawning and incubation time of the eggs.

The eutrophication in the Baltic Sea follows a south-north gradient, which is a reflection of how eutrophication in large scale is evolving. Therefore, as *Lundberg et al.* says about the spreading of eutrophication in the Gulf of Bothnia, the changes in the environmental conditions follow a southward gradual change, which is partly caused by the inflow of organic material from rivers. Also, there is a difference between the inner archipelago areas have been more affected by eutrophication than outer archipelago and exposed areas, which have reminded more stable over the time (Lundberg et al. 2009).

The same way, we can observe a higher eutrophication in the east coast of the Gulf of Bothnia than in the west coast. This is explained because due to the shallower water and higher nutrient load the east coast is more predisposed to eutrophication (HELCOM 2009, Andersen et al. 2011).

We can observe the three spatial ways of spreading in the filamentous prediction probability map. In the south and east the probability is usually higher, with the exception of areas which are less exposed (inner archipelago areas). However, the ecological status of coastal waters and eutrophication does not strictly follow the same gradient, because there are more factors taken into account in when calculating the ecological status (Venaranata et al. 2013).

CONCLUSIONS

Compared to the rest of the Finnish coastal waters and the Baltic Sea, the Gulf of Bothnia is in good environmental condition (Lundberg et al. 2009). It has been shown that the main group of species in the Gulf of Bothnia is the filamentous algae, with and elevated predicted probability in almost all the Gulf of Bothnia. This group of species follow the eutrophication spreading direction, responding logically to the biological knowledge as expected. Preferring hard bottoms like rock or stones and exposed areas, the number of filamentous algae is increasing every year, reducing macrovegetation populations into more protected areas.

The predicted distribution for the macrovegetation is therefore opposite to the filamentous algae. In three groups of species the most important variable is the bottom type. But in the case of filamentous algae and macrovegetation, the rest are more related to exposition, having a contrary response for each group of species. As a result we can observe a generally low probability in the whole area for the macrovegetation, and having just high probability in few localized areas were the filamentous algae is not dominant like inner parts of the archipelagos or small bays, areas with low exposition and well protected from the wind and the waves.

Another important factor in the predicted probability is a spatial dependence of the factors. As we have mention before there is south-north gradient in the spread of filamentous algae, but in the same way there is an east-south gradient, were the east coast has higher probability for filamentous algae, and consequently eutrophication. These gradients are caused by the higher number of islands in the south, the shallower waters in the east and the higher nutrients load and inflow of nutrients from the rivers in the more eutrophied areas.

Undoubtedly there is a relation between evolution of eutrophication and the increase of filamentous algae. The increased competition from filamentous algae, which gets advantage from the increase of nutrients and the reduced light conditions and the decrease of oxygen caused by the hypoxia as an effect of eutrophication, produce a change in the vegetation communities were smaller and fast-growing species of filamentous algae which can tolerate low concentrations of oxygen replaces other macrovegetation like Eelgrass or Bladder wrack.

These eutrophication-induced changes in the vegetation communities have consequences in the habitat structure, affecting for example the availability of predation, refuges and foraging habitats for other species. The same way, big changes are produced in the composition and development of the communities and causing changes in fish community structure and function. Therefore, we can say that eutrophication has modified all the trophic levels in the Baltic Sea and Gulf of Bothnia ecosystems, changing the structure and the species composition of different communities

Changes in vegetation can occur very fast and not as a gradual change (Dahlgren & Kautsky, 2004), so it should be particularly important to focus in the regions with less eutrophication and establish a follow up process to detect potential changes in key zones of the Gulf of Bothnia. As the scale is coarse and the resolution is not enough, more precise studies have to focus in these shallow coastal waters to start managing the eutrophication process in the Gulf of Bothnia.

MAIN BIBLIOGRAPHY

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

- Veneranta, L., Hudd, R. & Vanhatalo, J. (2013): *Reproduction areas of sea-spawning coregonids reflect the environment in shallow coastal waters.* Marine Ecology Progress Series 477:231-250

CITED LITERATURE

- Algesten, G., Brydsten, L., Jonsson, P., Kortelainen, P., Lofgren, ¨ S., Rahm, L., Raike, A., Sobek, S., Tranvik, L., Wikner, J., and ¨ Jansson, M. (2006): *Organic carbon budget for the Gulf of Bothnia.* J. Marine Syst., 63, 155–161.

- Andersen, J. P., Axe, H., Backer, J., Carstensen, U., Claussen, V., Fleming-Lehtinen, M., Järvinen, H., Kaartokallio, S., Knuuttila, S., Korpinen, A., Kubiliute, M., Laamanen, E., Lysiak-Pastuszak, G., Martin, C., Murray, F., Møhlenberg, G., Nausch, A., Norkko, and A., Villnäs. (2011): *Getting the measure of eutrophication in the Baltic Sea: towards improved assessment principles and methods*. Biogeochemistry 106: 137–156.

- Bonsdorff, E., Blomqvist, E.M., Mattila, J., Norkko, A. (1997): *Coastal eutrophication: causes, consequences and perspectives in the archipelago areas of the Northern Baltic Sea*. Estuar Coast Shelf Sci 44: 63-72

- Borg A., Pihl L., & H. Wennhage, H. (1997): *Habitat choice by juvenile cod (Gadus morhua L.) on sandy soft bottoms with different vegetation types*. Helgol. Meeresunters, 51: 2, pp.197-212.

- Busse, S. & P. Snoeijs (2002): *Gradient responses of diatom communities in the Bothnian Bay, northern Baltic Sea*. - Nova Hedwigia 74: 501-525.

- Diaz, R.J. & R. Rosenberg (2008): *Spreading dead zones and consequences for marine ecosystems.* Science 321:926–929.

- Einav, R., Sharon, Y., & Zahavi, A. (1996*): The relationship between wave energy and botanical population (macroalgae and terrestrials plants) on the Pigeon Island, The Mediterranean sea, Israel*. Six international Conference of the Israeli Society for Ecology & Environmental Quality Sciences, June - July, B:532-537.

- Elith, J. and J.R. Leathwick, 2009. *Species Distribution Models: Ecological Explanation and Prediction Across Space and Time*. Annual Review of Ecology, Evolution, and Systematics, 40: 677-697.

- Eriksson, B.K., Johansson, G. & snoeijs, P. (1998): *Long-term changes in the sublittoral zonation of brown algae in the southern Bothnian Sea.* Eur. J. Phycol., 33: 241 – 249.

- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. 2003. *Bayesian Data Analysis*. Chapman & Hall/CRC, Boca Raton, FL.

- Gelman, A., Pardoe, I., (2007): *Average predictive comparisons for models with nonlinearity, interactions, and variance components* .Sociol Methodol 37:23-51.

- Håkansson, B., Alenius, P., Brydsten, L., (1996): *Physical environment in the Gulf of Bothnia.* Ambio 8, 5–12. Special Report.

- HELCOM, (1996): *Third Periodic Assessment of the State of the Marine Environment of the Baltic Sea area*, 1989-1993. Balt. Sea Environ. Proc. No. 64A. 25 pp.

- HELCOM, (2002): *Fourth Periodic Assessment on the State of the Marine Environment of the Baltic Sea area* 1994–1998. Baltic Sea Environ. Proc., No. 82B. 215 pp.

- HELCOM, (2009): *Eutrophication in the Baltic Sea – An integrated thematic assessment of the effects of nutrient enrichment and eutrophication in the Baltic Sea region*. Balt. Sea Environ. Proc. No. 115B.

- HELCOM, (2010): *Ecosystem Health of the Baltic Sea 2003–2007: HELCOM Initial Holistic Assessment*. Balt. Sea Environ. Proc. No. 122.

- Karjalainen, M. (1999*): Effect of nutrient loading on the development of the state of the Baltic Sea – an overview.* Walter and Andrée de Nottbeck Foundation Scientific Reports 17, 35.

- Kingston, J.C., Lowe R.L., Stoermer, E.F., & Ladewski, T.B.(1983): *Spatial and temporal distribution of benthic diatoms in Northern Lake Michigan*. Ecology, 64, 1566-1580.

- Lappalainen, A. (2002): *The effects of recent eutrophication on freshwater fish communities and fishery on the northern coast of the Gulf of Finland, Baltic Sea*. Ph.D. thesis. Finnish Game and Fisheries Institute, Helsinki. 24 pp.

- Lundberg C., Jakobsson B.M. & Bonsdorff E. (2009): *The spreading of eutrophication in the eastern coast of the Gulf of Bothnia, northern Baltic Sea — an analysis in space and time*. Est. Coast. Shelf Sci. 82: 152–160.

- Raffaelli, D.G., Raven, J.A., Poole, L.J. (1998): *Ecological impact of green macroalgal blooms.* Oceanogr Mar Biol Annu Rev 36:97–125

- Rasmussen, C. E., & Williams, C. K. I., (2006): *Gaussian Processes for Machine Learning*. Massachusetts Institute of Technology. MIT Press, 2006. ISBN 026218253X.

- Savchuk, O.P., F. Wulff, S. Hille, C. Humborg and F. Pollehne. (2008): *The Baltic Sea a century ago – a reconstruction from model simulations, verified by observations*. J. Mar. Syst., 74: 485–494.

- Savchuk, O.P., Eilola, K., Gustafsson, B.G., Rodríguez Medina, M., Ruoho-Airola, T. (2012): *Long-term reconstruction of nutrient loads to the Baltic* Sea, 1850–2006. Baltic Nest Institute Techn. Rep. Ser. 6, 9 pp. ISBN: 978-91-86655-05-1

- Snoeijs, P. (1994*): Distribution of epiphytic diatom species composition, diversity and biomass on different macroalgal hosts seasonal and salinity gradients in the Baltic Sea.* Diatom Res. 9 (1), 189–211.