ICES CM 2004/ P:26

Modelling Fish Habitat Suitability in the Eastern English Channel. Application to community habitat level.

S. Vaz, A. Carpentier, C. Loots and P. Koubbi.

Valuable marine habitats and living resources can be found in the Eastern English Channel and in 2003, a Franco-British Interreg IIIA project, 'Eastern Channel Habitat Atlas for Marine Resource Management' (CHARM), was initiated to support decision-making for management of essential fish habitats. Fish habitat corresponds to geographic areas within which ranges of environmental factors define the presence of a particular species. Habitat Suitability index (HSI) modelling was used to relate fish geographic distribution and their relation towards environmental factors and to delineate their optimum habitat. This study was based on data obtained from 1988-2003 IFREMER's Channel Ground Fish Surveys, including both species abundance and environmental data. Suitability index (SI) functions based on generalised additive models were used to relate depth, temperature, salinity and sediment to sub-community assemblage probability of occurrence. As a result, SI values were positively related to assemblage affinity along the gradient of the environmental variables. The resulting HSI models were used to map, using GIS, the optimum habitats of communities ; sensible habitats such as spawning grounds, nurseries or areas carrying bio-diverse fish community were also defined. The information obtained will help to elaborate guidelines for the conservation and protection of natural habitats in the face of climate change and anthropogenic disturbances.

Key-words : Eastern English Channel, CHARM, Fish Optimum Habitat, GIS

S. Vaz, A. Carpentier: Ifremer, Laboratoire Ressources Halieutiques, 150 quai Gambetta, BP699, 62321, Boulogne/mer, France [tel : (+33) 3 21 99 56 00, fax : (+33) 3 21 99 56 01, e-mail :andre.carpentier@ifremer.fr]. C. Loots, P. Koubbi, LIMUL, Laboratoire d'ichtyo-écologie marine, Université du Littoral Côte d'Opale, 32 avenue Foch, 62930, Wimereux, France [tel: (+33) 3 21 99 64 08, fax: (+33)3 21 99 64 01, e-mail : koubbi@univ-littoral.fr].

Introduction

In the Eastern English Channel, an important area for small fisheries with strong hydrodynamic features, a Franco-British Interreg IIIA project, 'Eastern Channel Habitat Atlas for Marine Resource Management' (CHARM), was initiated to support decision-making for management of essential fish habitats. The principal aim of the project is to better understand the marine environment of the Dover Strait region in terms of its biological resources by simulating scenarios for impacts to marine habitats and changes in levels of exploitation that can influence marine living resources (Carpentier *et al.*, 2004)

The Eastern English Channel is a shallow sea (40-100m depth) separated from the Western English Channel by a vast area of pebbly seabed. The temperature is homogeneous across the water column as a result of strong seabed currents inducing a permanent bracing. Abundant and numerous fish species may be found and this richness is linked to the abundance of benthic animals that constitute a food source (Nival, 1991). The benthic communities distribution is linked to the tidal currents intensity occurring in the Eastern English Channel (Pingree & Maddock, 1977), which is also characterized by an unbalanced salinity pattern between France and England. Lower salinity concentrations result from the Seine and Somme river waters pouring along the coast, from the Seine estuary to the Dover strait. As a result the species and community assemblages present in this part of the Channel were suspected to be strongly affected by such heterogeneous environment.

IFREMER (French Research Institute for Exploitation of the Sea) has been carrying out since 1988 an essential ground fish survey enabling ICES to produce annual evaluations of major commercial fish stocks. Based on collected biological data on all demersal, benthic and pelagic species captured, it was possible to identify and describe the existing communities in the Eastern Channel and relate them with physical and hydrological features. Four sub-communities were defined by the TWINSPAN classification (Fig.1) and seemed to reflect an inshore to offshore gradient in its assemblages (Fig.2) (Vaz *et al.*, 2004). The community in the Eastern English Channel was found to be strongly structured spatially and clearly resulted from important community response to the environment.

It may therefore be possible to model and predict sub-community distribution from environmental parameters. Habitat Suitability Index approach, linking statistical modelling to GIS mapping, could be used to model the fish community and to verify whether the environmental descriptors available here can provide on their own (without any information of the trophic relationships or biotic interaction) an acceptable prediction of the community type.

With the development of new statistical techniques and GIS tools, predictive habitat distribution models are increasingly used to relate the geographical distribution of species or communities to their environment (Guisan and Zimmermann, 2000, Eastwood *et. al*, 2003). Fish habitat corresponds to geographic areas within which ranges of environmental factors define the presence of a particular species. Habitat modelling may also be used to relate community assemblage geographic distribution and their relation to environmental factors and to delineate their optimum habitat (Franklin, 1995). However, this relationship is likely to be complex and may not be linear. Generalised additive models (GAM) represent a method of fitting a smooth relationship between two or more variables through data points. They are useful where the relationship between variables is expected to be of a complex form, not easily fitted by standard linear or non-linear models. They do not involve strong assumptions about the relationship that is implicit in standard parametric regression. Such assumptions may force the fitted relationship away from its natural shape and as a result GAM are increasingly used in ecology.

Methods

Survey design and data available

Since 1988, the CGFS survey has been taking place every year in October in the Eastern English Channel and the southern North Sea. During each fishing operation, bottom depth at each station and hydrological parameters (bottom temperature and salinity) were recorded. The abundance indices of the encountered species were standardised into density per km^2 at each station. Ninety observed species and 1326 stations were used in this study. The sediment type was obtained from the Larsonneur *et al.* (1979) map by re-sampling it at trawl haul locations using the GIS Arcmap software. The original sediment types have been simplified into 5 classes and include by decreasing order of size, pebbles, gravels, biolithoclastic sand, lithoclastic sand and muds. Survey design and environmental parameters have been described in a precedent paper (Vaz *et al.* 2004) and will not be further detailed here.

Spatial analyses and GIS

Interpolation is required for estimating the values of a property of interest at non sampled locations from sparse and irregularly spaced samples. Kriging is a geostatistical estimation different from other interpolators because it uses a model of the spatial variation – the variogram which is the central tool of geostatistics (Webster and Oliver, 2001). Geostatistical analyses were used extensively to produce environmental maps in particular temperature and salinity maps obtained during the surveys. These were averaged over all the available years. It was also applied to community type data to illustrate the assemblages geographic distribution and extent. These analyses will not be detailed in this paper but the resulting results and maps were used to predict the Community Habitat maps and to compare predicted community to observed community types.

Grids obtained through kriging were imported into ArcGIS (ESRI, ArcGIS version 8.2), plotted and projected to a common reference system (Transverse Mercator projection). Care was taken to limit the spatial extent of the interpolated maps so as to avoid keeping data resulting from extrapolation (i.e. in areas where no sampling had taken place). The created shapes were joined to resample predicted maps at station position and enable the comparison of predicted and observed value. The original sediment layer was used for model predictions. Mean sea level at mid-tide (coefficient of 70) was obtained from a hydrodynamic model of IFREMER labs in the form of a 4 square km grid, which was then combined to a classical bathymetric map (Martin *et al.* 2004). The resulting depth layer was thought more ecologically relevant and comparable to observed depth and were used for community predictions

TWINSPAN Classification

Classification techniques can describe and recognise patterns in species distribution and define communities. The method TWINSPAN (Two-Way INdicator SPecies ANalysis) (Hill *et al*., 1975) combines ordination and clustering and it is widely used in vegetation science to classify species-by-sample data. This analysis results in both a classification and an ordination of the data, from which a dendrogram can be constructed (Legendre and Legendre, 1998). The TWNSPAN procedure was used to define different sub-communities based on the species data collected during the CGFS surveys from 1988 to 2003 (Vaz *et al.*, 2004).

Habitat Modelling

The depth, bottom salinity and bottom temperature observed at each station were used for model calibration and all available data were used. This implied that the same data set was used to calibrate and evaluate the model which was not to be validated outside of its calibration range. However, the depth layer resulting from the superposition of bathymetry and mean sea level was used as the reference grid for model prediction. At each grid nods, sediment type map and kriged bottom salinity and temperature maps were resampled to produce a regular predictor data set used for model prediction.

GAM

Locally weighted approaches, such as generalised additive modelling (GAM) have already been used to predict species abundance based on maps of environmental predictors of suitable habitats (Guisan and Zimmermann, 2000). GAM models are a flexible class of models that are commonly used to implement non-parametric smoothers in regression models. The smoothers are applied independently to each predictor and additively calculate the component response. GAMs have the following general form :

$$E\left[y\right]=f\left(\beta_0+\sum_i s_i\left(x_i\right)\right) \quad (1)$$

Where x_i is the i^{th} predictor, y is the response of interest (here community type), which has some specified statistical distribution whose expectation, E(y) is a function, f( ), of the explanatory variables. The function f is the inverse of the link function (same as those available in generalised linear models). β_0 is a parameter to be estimated, and the functions s_i( ) are smoothing functions for the explanatory variables.

Community data are composed of binary data for each sub-community type. The logistic regression for the Bernouilli component of the model was chosen and the logit link function was used:

$$G\left(x\right)=Log\left(\frac{1}{1-x}\right) \quad (2)$$

The resulting estimation is a probability of occurrence (likelihood) ranging from 0 to 1. Only smoothing splines for si were used although other functions were available. The flexibility of the smoothing splines is summarised in their associated degree of freedom. A degree of freedom approximatively equal to 4 was used and compared to linear or quadratic predictors (df = 1 or 2). Spline functions could only be applied to temperature, salinity and depth predictors as the sediment predictors were categorical variables. Only the linear interaction between temperature and salinity was considered to reduce model selection to manageable proportion. The model selection was performed by comparing the full model (equation 3) to every possible reduced model

E(x) = s(temperature) + s(salinity) + s(depth) + gravels + pebbles + biolithoclastic sands + lithoclastic sands + muds + (temperature:salinity) (3)

Each predictor were removed in turn and the resulting model was compared to the full model following a process of backward elimination. The interaction term was first excluded, then each sediment types in turn, then the smooth function for each continuous variables, then the continuous variables themselves. Model comparison and testing against the full model was used at each step and made possible by ANOVA (indicated by the t statistics). This enable the choice which of several statistical models most appropriately reflects the relationship between the response and the predictor variables. Such test is appropriate for determining whether to select a more complex model with many predictors or a simpler model with a subset of the predictors. The models in this paper have been implemented within the S-PLUS software using the existing GAM technology and algorithms.

Model validation

Spearman's Rank Correlation Coefficient

Spearman's Rank Correlation Coefficient is a measure of association between the rankings of two variables measured on n individuals (i.e. two vectors of length *n*). The correlation coefficient is calculated from the two vectors of ranks for the samples: let $\{x_i ; i=1...n\}$ and $\{y_i ; i=1...n\}$ be the vectors of ranks for sample 1 and sample 2 respectively, then the coefficient r is based on the vector of differences between ranks: $\{d_i = x_i - y_i\ ;\ i=1...n\}$ and is calculated by :

$$r = 1 - 6\left(\sum_{i=1\dots n} \frac{d_i^2}{n(n^2-1)} \right)$$

If ties are present, then the statistic will be biased, and must be recalculated taking account of ties by :

$$r = \frac{\sum x_i^2 + \sum y_i^2 + \sum d_i^2}{2\sqrt{\sum x_i^2 \times \sum y_i^2}} \qquad (4)$$

where $\sum x_i^2 = \frac{n^3-n}{12} - t_x$, $\quad \sum y_i^2 = \frac{n^3-n}{12} - t_y$, $\quad t_k = \frac{\sum t_j^3 - t_j}{12}$

and t_j is the number of observations in the group with rank j.
The t-approximation for this statistic, T, is valid for samples of size 8 upwards, and is calculated by :

$$T = r\sqrt{\frac{n-2}{1-r^2}} \qquad (5)$$

It has approximately a t-distribution on $n-2$ degrees of freedom, and can be used for a test of the null hypothesis of independence between samples. This test was used to compare and validate estimated community type from GAM and observed values.

Results

Model selection

Each community type was modelled in the way described earlier and the resulting models were summarised in Table 1. Each reduced models were tested against the full model (eq. 4) and were found to be statistically similar to the full model thus validating the exclusion of some irrelevant predictors (Table 2). The number and complexity of predictors increased from group 1 to 4 revealing the increasing complexity of the community relationship to the environment.

Although temperature and depth descriptors were significant for all community types, salinity was found to be of lesser importance for the first two groups. It was used as linear predictor in group one and was excluded from the group 2 model. Similarly, the first two groups were significantly related to coarse sediment types (gravel, pebbles and bio-lithoclastic sands) while muds was relevant only to the group 3 and 4 predictive models. Lithoclastic sands were never found useful to model building. Finally, for the group 4 the interaction between salinity and temperature was found to act significantly on the sub-community repartition.

Four probability maps were created from the predicted values from each model (Fig.4). These likelihood maps reflected the differences of spatial distribution between each classes. The occurrence probability of the group 1 (Fig. 4a) was higher offshore and in the Dover Strait area. The predicted group 2 was likely to occur closer to the coast in area where the group 1 has less than 50% chance of occurring (Fig. 4b). The group 3 was likely to occur in coastal areas (Fig. 4c). Finally the group 4 occurrence probability was higher on very coastal areas, in particular the Seine bay and the English coast of the Dover Strait. This result was particularly interesting since group 3 and 4 were found to be the most diverse assemblage types and may reflect the occurrence of many species juveniles (Vaz *et al.*, 2004).

These probability maps were produced on the same prediction grid and were merged by selecting for each grid cell the most probable predicted group (the selected group probability of occurrence was never less than 0.3). The resulting predicted community distribution (Fig. 5) illustrated the most likely community habitats distribution for all 4 types. The distribution patterns displayed on this map closely resembled those of the original community classification (Fig. 3a) and seemed more realistic than the produced interpolated map (Fig. 3b)

Model Validation

For each sub-community type a Spearman correlation coefficient was computed between the predicted most probable group and the observed group membership. This was done by joining observation locations to the closest predicted values in ArcGis. The predicted data grid had a much lower resolution than the discrete and irregular stations position and joined values distant from over 0.05° were removed. Spearman's Rank correlation coefficient between observed and predicted group membership was computed for each group and was found to be significant (Table 3). Within each group, observed and predicted values are significantly correlated.

The study of the occurrence frequency of misclassified community types revealed that the right community type was the most likely to be predicted and that misclassification is more likely to occur in immediate neighbouring groups than other groups (Table 4). For example group 1 is more likely to be misclassified in group 2 than in group 3 and is never misclassified in group 4. This result illustrated the ecological continuum already observed between the four community types (Vaz *et al.*, 2004).

Discussion

For each community type a different model, including a different set of predictors, was selected. Each assemblage reflected a different set of environmental condition which determined its habitat distribution. Depth and temperature were found of paramount importance to predict any community type and spline smoothers were often required to take into account the complex relationship among them. Group 1 and 2 were found in the deepest and warmest areas (corresponding to offshore oceanic waters) while group 3 and 4 were restricted to shallower coastal areas, for which the temperature is cooler at that time of the year (October). Salinity was found to be more relevant to coastal sub-communities (group 3 and 4) and to exhibit in these instances complex relationships to the community requiring spline smoothers. The occurrence of the first two groups were linked to the distribution of coarse sediment types (gravels, pebbles and biolithoclastic sand), which increased their likelihood. This was particularly true for group one, for which high likelihood was predicted in the Dover Strait: this was certainly linked to its high affinity to the coarse sediments present there. On the contrary, groups 3 and 4 responded positively to the occurrence of muddy sediments and negatively to coarser sediments. Lithoclastic sand were representing a broad range of sands including coarser and finer types (Larsonneur *et al.*, 1979) and were not found to be indicative of the community type.

These predicted patterns were very similar to those observed from the survey data (Vaz *et al.*, 2004) and the model predictions were remarkably consistent with the original data although they were obtained using a different set of environmental predictors (depth, temperature and salinity). The distribution patterns of the community types predicted by the combination of the four models is remarkably close to the observed community geographic distribution and gave more realistic patterns than those from interpolated maps. Not only did this approach constituted a good way of predicting community distribution but it also had a clear explicative power as too what parameter is responsible for the delineation of the different sub-community habitats. The predicted coastal distribution of group 3 and 4, linked not only to depth but also to complex temperature and salinity interaction and soft sediment occurrence may constitute a useful information to delineate the probable suitable habitat for many species nursery grounds. These groups were also found to be the most diverse and even ones (Vaz *et al.* 2004) and the knowledge of their potential area of distribution is of premium importance to aid decision-making and planning in the marine environment of the Dover Strait and adjacent waters.

The strong effect of the environment of the community structure implied that there was little biotic interaction occurring at this level of community division. This certainly resulted from the strong environmental heterogeneity of the area as well as important sources of habitat disturbances (currents, tides, fishery pressure). The species constituting the sub-communities described here seemed to coexist on the same habitat but not to interact strongly as they appeared to respond almost exclusively to their environment. This certainly resulted from the fact that these sub-communities included species that seldom met in their natural environment (eg. Crustacean and small pelagic species) and were far apart on trophic network. This community definition, however, may reveal itself a useful indicator of climatic change (implying a change on its structuring environmental factors) and would probably not reflect so clearly other anthropic impacts in the area.

Conclusion

This study confirmed that the fish communities distribution found in the Eastern English Channel resulted from the heterogeneous environmental patterns in the area. The habitat distribution of the observed community may therefore be deduced from specific environmental predictors depending on the community type investigated.

This modelling methodology enable both the identification the environmental condition suitable for each community type but also the prediction of their likely geographic limits in unexplored areas. The result of this study may therefore be useful for the purpose of management decision-making when knowledge of the community type is required in a particular area of interest.

Such community habitat modelling approach could enable the simulation of environmental change scenarios and the study of their likely effect on the distribution of species assemblages in the Eastern English Channel.

References

Carpentier A. , C. Martin, S. Vaz, F. Coppin, P. Koubbi, A. Lefebvre, G. J. Meaden, M. Walkey, S. Harrop,Z. Kemp, N. Desroy, J.C. Dauvin & J.M. Dewarumez, P. D. Eastwood. 2004. Eastern Channel Habitat Atlas for Marine Resource Management (CHARM). ICES CM 2004/ P:41.

Eastwood P. D., Meaden G. J., Carpentier A. and Rogers S. I., 2003. Estimating limits to the spatial extent and suitability of sole (*Solea solea*) nursery grounds in the Dover Strait. Journal of Sea Research, 50(151-165).

Franklin J., 1995. Predictive vegetation mapping: geographical modelling of biospatial patterns in relation to environmental gradients. Prog. Phys. Geogr. 19, 474-499.

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

Larsonneur C., Vaslet D. and Auffret J. –P., 1979. Les Sédiments Superficiels de la Manche, Carte Géologique de la Marge Continentale Française, Bureau des Recherches Géologiques et Minières, Ministère de l'Industrie, Service Géologique National, Orléans, France.

Legendre P & Legendre L, 1998. *Numerical Ecology*. Elsevier, Amsterdam.

Martin C. S., Carpentier A., Coppin F., Eastwood P. D., Meaden G. J., Walkey M., Harrop S., Kemp Z. and Vaz S., 2004. A Digital Atlas for the Physical Environment of Eastern English Channel (CHARM Project). ICES CM 2004/P42

Nival P. 1991. Manche et mer du Nord : riches et fragiles. Science et Vie n°176. Sept. 91. Hors-série. La vie des océans : 106-113.

Pingree R.D. & L. Maddock. 1977. Tide residuals in the English Channel. J. Mar. Biol. Ass. U.K., 57, 339-354.

Vaz S., Carpentier A. and Coppin F., 2004. Eastern English Channel Fish Community from 1988 to 2003 and its Relation to the Environment. ICES CM 2004/K:40

Webster R & Oliver MA, 2001. *Geostatistics for Environmental Scientists*. Wiley, Chichester.

Tables

Table 1 : Summary of selected model predictors for each sub-community types

Predictors	Full model	Response Group 1	Group 2	Group 3	Group 4
Btemp	S	S	+	S	S
Bsal	S	+	-	S	S
Depth	S	S	S	S	S
P	+	+	+	+	-
G	+	+	+	+	+
BS	+	-	+	+	+
LS	+	-	-	-	-
M	+	-	-	+	+
Btemp:Bsal	+	-	-	-	+

S: spline smoother, +: linear, -: not included. Btemp : Bottom temperature, Bsal : bottom salinity, Depth : measured bottom depth, P : pebbles, G : gravels, BS : bio-lithoclastic sand, LS : lithoclastic sand, M : muds, Btemp:Bsal : interaction term between temperature and salinity.

Table 2 : Summary ANOVA table comparing and testing the selected reduced model against the full model

	Full Model		Reduced Model		Comparison Test	
	Resid. Df	Resid. Dev	Resid. Df	Resid. Dev	F Value	Pr(F)
Group 1	630.1002	427.4204	637.0160	436.7603	1.537535	0.1525356
Group 2	630.5014	504.8808	640.1597	516.0072	0.9159484	0.515439
Group 3	630.1339	533.5285	632.1417	534.6039	0.6432443	0.5265149
Group 4	630.1247	358.4993	632.2952	361.3442	1.584135	0.2038579

Table 3 Spearman's Rank Correlation Coefficient testing the independence between observed and predicted community classes

	Group 1	Group 2	Group 3	Group 4
Sample size	1268	1268	1268	1268
Correlation	0.7659	0.6906	0.6990	0.8440
Adjusted for ties	0.6742	0.4293	0.4844	0.4384
t Approximation	32.48	16.91	19.70	17.35
Degrees of freedom	1266	1266	1266	1266
P-value	0.000	0.000	0.000	0.000

Table 4 Misclassification frequency of predicted vs observed community types

predicted classes	observed classes 1	2	3	4
1	0.76	0.16	0.04	0.04
2	0.21	0.61	0.15	0.03
3	0.04	0.17	0.52	0.26
4	0.00	0.02	0.17	0.81

Figures

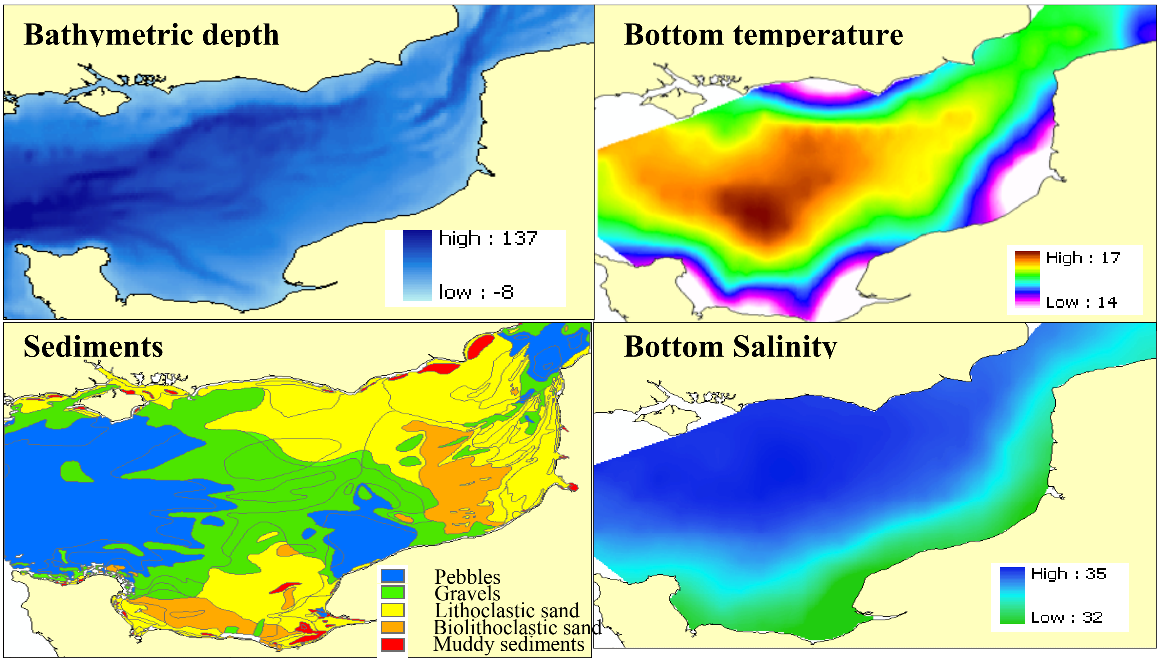

Figure 1 Eastern English Channel physical and hydrological features: Bathymetric depth and simplified sediment types representation. Survey bottom temperature and bottom salinity (averaged for 1997 to 2003) obtained by kriging.

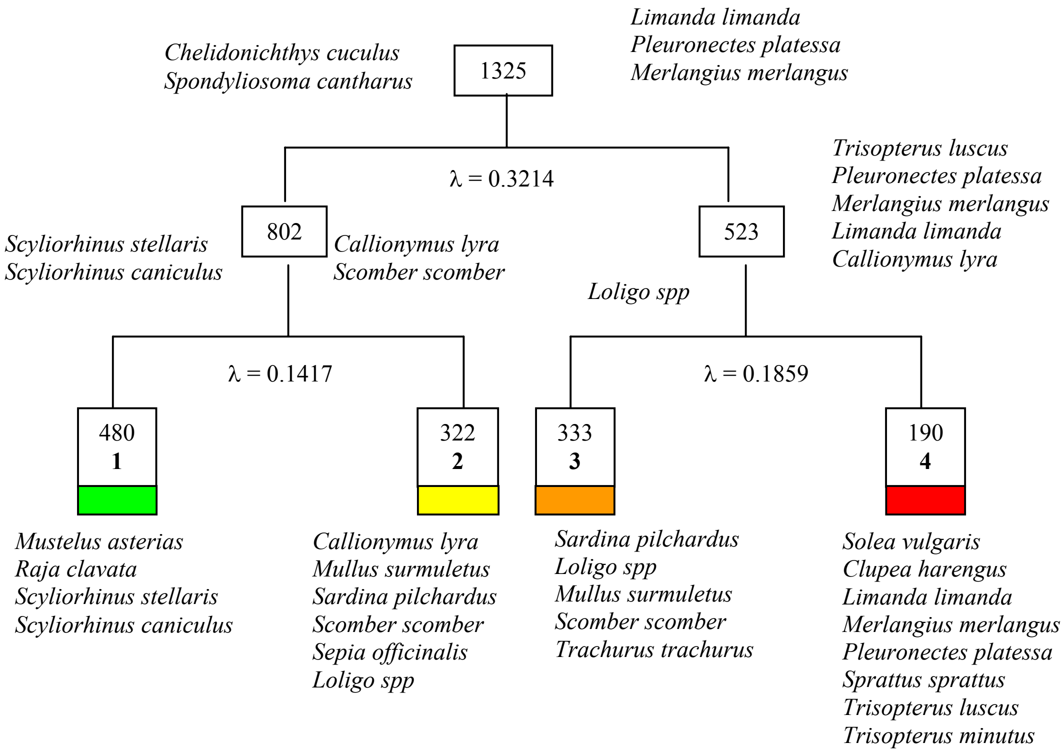

Figure 2 TWINSPAN Classification of Fish Community : the dendrogram represents the first two levels of division. DCA first axis eigen-values are represented for each division and for each group, the corresponding indicator species are given. The number of samples in each sub-group is indicated in the boxes. The preferential species of the four sub-communities are listed at the bottom of the dendrogram.

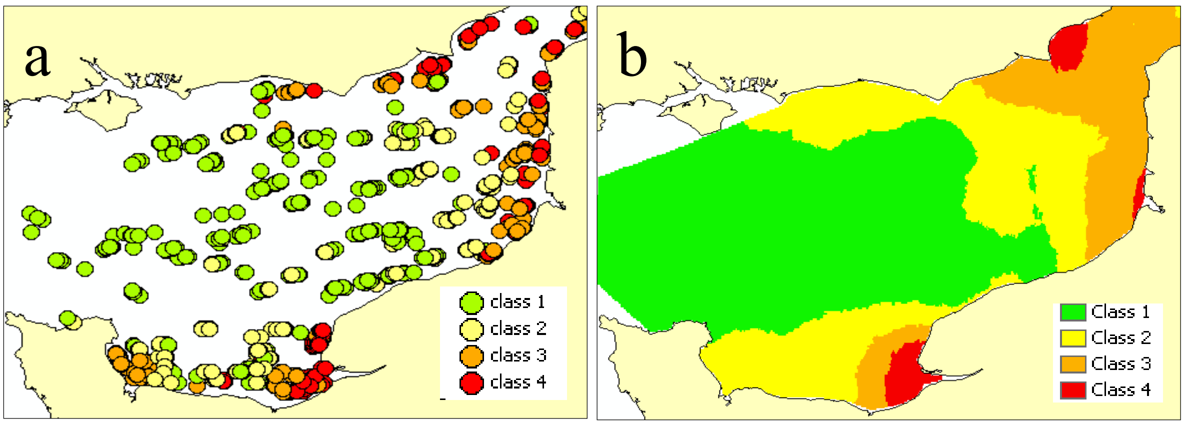

Figure 3 Spatial distribution of Fish Subcommunities in the Eastern Channel from 1988 to 2003. (a) observed assemblage type at each station, (b) kriged interpolation of assemblage type in the prospected area. These illustrate the gradation from open sea community to coastal and estuarine communities.

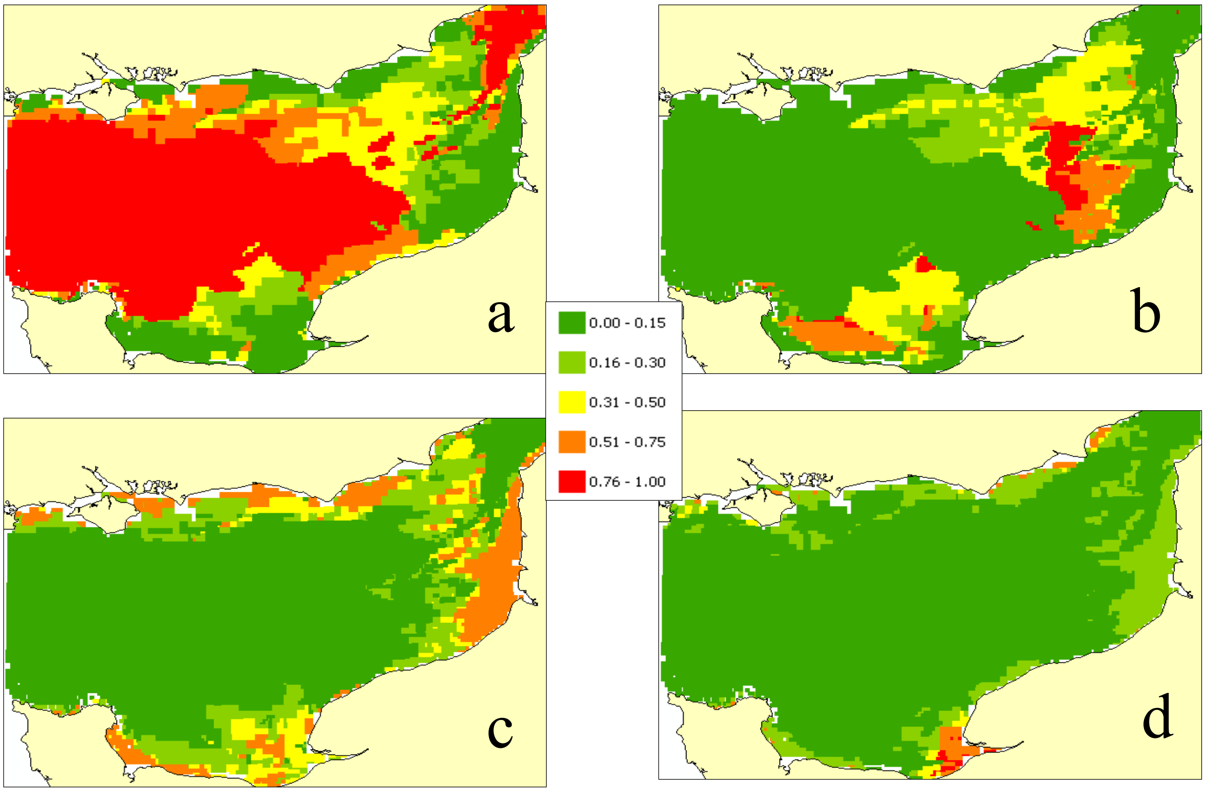

Figure 4 Predicted probability of occurrence of each community type based on habitat modelling (a) Class 1 (b) Class 2 (c) Class 3 (d) Class 4. The suitable habitats of each community type were clearly separated in space and comfirmed the inshore – offshore structure gradient of the fish community.

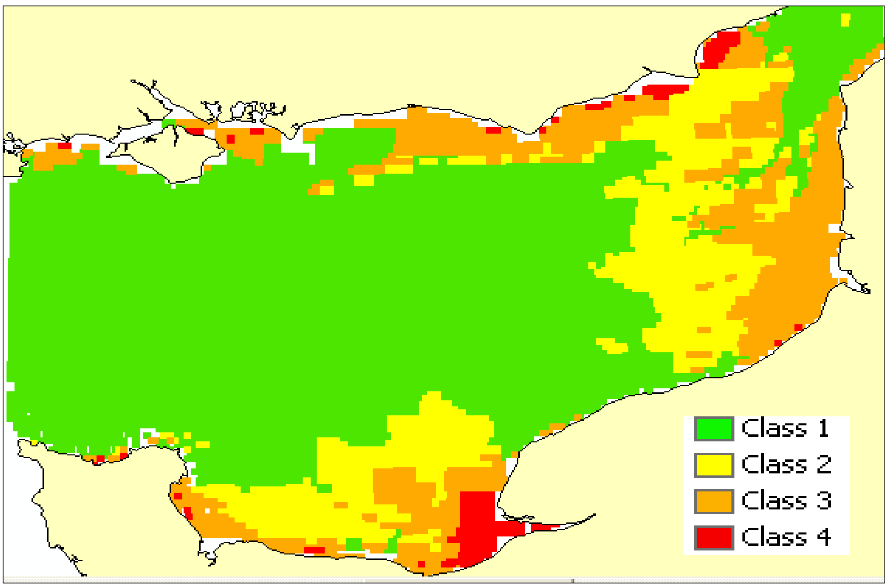

Figure 5 Predicted community distribution. In each prediction cell, the community type with the highest occurrence probability was chosen. This map displayed the same distribution patterns as those observed in Fig. 3.