Fisheries Research
September 2008, Volume 93 (1-2), Pages 29-38
http://dx.doi.org/10.1016/j.fishres.2008.02.008

Cluster analysis of linear model coefficients under contiguity constraints for identifying spatial and temporal fishing effort patterns

Stéphanie Mahévas[a, *], Lise Bellanger[b, *], Verena M. Trenkel[a]

[a] IFREMER, Département EMH, BP 21105, 44311 Nantes Cedex 03, France
[b] Université de Nantes, Laboratoire de Mathématiques Jean Leray, UMR CNRS 6629, BP 92208, 44322 Nantes Cedex 03, France

*: Corresponding author : Mahévas S., email address : Stephanie.Mahevas@ifremer.fr, Bellanger L., email address : Lise.Bellanger@univ-nantes.fr

Abstract:

For fisheries management purposes, it is essential to take into account spatial and seasonal characteristics of fishing activities to allow a reliable assessment of fishing impact on resource. This paper presents a novel technique for describing spatial and temporal patterns in fishing effort. The spatial and seasonal fishing activity patterns of the French trawler fleet in the Celtic Sea during the period 1991–1998 were analysed by modelling fishing effort (fishing time) with generalised linear models. The linear model for fishing effort included fixed effects for both spatial (statistical rectangles) and temporal units (months). In addition, spatial correlations in any given month were modelled by an exponentially decreasing function. Temporal correlations were included using the previous month's fishing effort for a given spatial unit as predictor. A method based on cluster analysis of estimated model coefficients of spatial or temporal fixed effects is proposed for identifying groups of similar spatial and temporal units. A contiguity constraint is imposed in the clustering algorithm, ensuring that only neighbouring spatial units or consecutive temporal units are grouped. The cluster analysis identified 22 spatial and 9 temporal groups. Winter and spring months stood out as being more variable than the remaining months. Spatial groups were of varying size, and generally larger offshore. The proposed method is generic and could for example be used to analyse temporal and spatial patterns in catch or catch rate data.

Résumé:

Dans un objectif de gestion des pêcheries, pour établir un diagnostic fiable de l'impact de la pêche sur la ressource, il est nécessaire d'intégrer les spécificités spatiales et temporelles de l'activité de pêche. Ce papier présente une nouvelle méthode pour décrire des structures spatiales et temporelles de l'effort de pêche. La distribution spatiale et saisonnière des chalutiers français pêchant en mer Celtique entre 1991 et 1998 est analysée en modélisant l'effort de pêche (temps de pêche) à l'aide d'un modèle linéaire généralisé. Le modèle décrivant la variabilité de l'effort de pêche incluait des effets fixes spatiaux (à l'échelle du rectangle statistique) et temporels (à l'échelle du mois). Les corrélations spatiales à un mois donné étaient modélisées par une fonction exponentielle décroissante de la distance et pour tenir compte des corrélations temporelles nous avons introduit, pour une unité spatiale donnée, l'effort de pêche du mois précédent comme variable explicative dans le modèle. Une méthode de classification des effets fixes spatiaux (respectivement temporels) du modèle statistique est alors proposée pour construire des groupes d'unités spatiales (respectivement des groupes d'unités temporelles). Des contraintes de contiguïté spatiale et temporelle sont imposées dans l'algorithme de classification pour s'assurer que seules les unités spatiales voisines et que seules les unités temporelles successives soient groupées. L'application de cette méthode de classification a permis d'identifier 22 zones et 9 saisons. Les mois d'hiver et de printemps ressortent comme étant plus hétérogènes que les autres. La taille des zones est très variable et généralement plus grande au large qu'à la côte. La méthode proposée est générique et pourrait être par exemple utilisée pour identifier des structures spatiales et temporelles des données de capture ou de taux de capture.

Keywords: Generalised linear model; Cluster analysis under contiguity constraints; Statistics for spatial data; Spatial and seasonal pattern; Allocation of Fishing effort; Fleet dynamics

1. Introduction

Fishing fleet dynamics are characterised by the choice of fishing location and the set of target species at a given time of the year (Hilborn and Ledbetter 1985). Seasonal species migrations (Biseau 1998; Vignaux 1996a), economic changes and weather conditions (Holland and Sutinen 1999; Sampson 1991) make fishing activities variable in both time and space. In order to reliably evaluate the impact of a given fishing fleet on a particular resource, it has been argued that taking account of spatial and seasonal characteristics of fishing activities is essential for reliable stock assessments and realistic forecasting models for management purposes (Booth 2000). This leads first to a decomposition of fishing effort by métier which is defined by season, location, target species and fishing gear (Biseau et al 1988), and commonly accepted as a fundamental feature of fishing activities (ICES 2004). The exploration of alternative management measures is another field of application of these spatial and temporal patterns. Babcock and Pikitch (2000) underlined the importance of spatial and seasonal knowledge of populations and fleets to design appropriate marine protected areas and successful management measures. Several simulation tools for management scenario testing have been developed that require definition of distinct spatial and temporal fishing activity units (e.g. Sparre 2003; Mahévas and Pelletier 2004; Pelletier and Mahévas 2005).

In fisheries science, hierarchical cluster analysis is commonly employed for grouping observation units, such as observation years for scientific surveys (Poulard 2001), catch composition for identifying métiers and strategies (Pelletier and Ferraris 2000) or species spatial distributions (Verdoit et al 2003). The general aim is to group sampling units that show common patterns. Here we propose a model-based cluster analysis for grouping variables, such as a temporal or spatial effects. These variables are the coefficients of a linear model, and hence assess the average features of variability of observations conditional on model formulation. Clustering estimated model effects instead of raw observations allows us to ignore local fluctuations of observations not explained by spatial or temporal factors, for instance due to autocorrelation structures in observations.

Cluster analysis is an algorithmic procedure providing partitions of the initial population. Two broad clustering families have been developed (Lebart et al 1997, Gordon 1996): mobile centroid clustering methods (Hartigan and Wong 1979) and hierarchical clustering approaches (Gordon 1987). The first family partitions directly units into disjoint groups. Allocation of units is iterative in order to minimize the distance of each unit to the estimated centroids of the clusters. Hierarchical clustering is based on an agglomerative technique grouping units two by two (or divisive technique splitting the group into two groups). Some additional constraints, usually contiguity constraints, are often required in the classification to take neighbourhoods into account. Clustering with contiguity constraints requires first the definition of a neighbourhood relationship (e.g. horizontal, spherical adjacencies) and second performing a clustering algorithm modified to take into account the neighbourhood constraints. Gordon (1996) provides a review of constrained classification methods. Different linkage methods can be used to decide whether objects are similar enough to be grouped. The most commonly used methods are complete linkage and single linkage. Complete linkage is often used in ecology when one wishes to delineate clusters with clear discontinuities (Legendre and Legendre 1998). The single linkage method has the advantage over the other methods to only use rank distance and consequently, to be rather similar to non-parametric methods. It is also the only linkage method allowing hierarchical clustering with contiguity constraints (Everitt 2001). Unfortunately, for noisy data, this linkage method is also well known to cause chaining in the dendrogam. Several studies have analysed and characterised the chaining phenomenon, see for instance Hartigan (1975) and Everitt (2001). More details of these linkage techniques can be found in Gordon (1981) and Lebart et al (1997).

In this study, we developed a clustering method with contiguity constraints based on a modified dissimilarity matrix and using a minimum linkage method independently on spatial and temporal units. Each unit is characterized by an estimated parameter value provided by a generalised linear model. The modified dissimilarity matrix is computed using 1) the 1-p-value derived from a Fisher statistical test applied to estimated values to assess the null-hypothesis of equality of pairs of parameters for temporal units (or spatial units) and 2) the neighbourhood constraint. The temporal neighbourhood relationship is assumed horizontal: sorting the units in sequential order, a temporal unit can only be grouped with the previous and following temporal unit. Spatial units are located on a regular grid and the eight neighbours of a spatial unit define the spatial neighbourhood. We apply the proposed model-based clustering algorithm for determining spatial and seasonal patterns in fishing effort for the French trawler fleet in the Celtic sea (Figure 1). In the following description we assume that the temporal unit corresponds to calendar months and the spatial units to statistical rectangles (1 degree longitude by 0.5 degree latitude).

2. Material and method

2.1. Data

The data come from the French trawler fleet operating in the central part of the Celtic Sea during the period 1991 to 1998. The fleet consists of 589 trawlers between 12 and 24 meters in length. For each vessel-trip, total trawling time was available per statistical rectangle. We modelled total fleet fishing time per statistical rectangle and per month for each year (Mahévas and Trenkel 2002).

2.2. Model

The approach has three steps (Figure 2): 1) conducting an exploratory analysis of fishing time data to investigate statistical data distributions, autocorrelation structures, etc.; 2) fitting an appropriate statistical model to fishing time data to estimate spatial and temporal effects; 3) separately clustering the spatial and temporal effects estimated in the previous step to provide fishing zones and seasons.

Based on the exploratory data analysis (step 1, descriptive analysis and plots), a set of models for describing the spatial and temporal distributions of fishing time was defined (Table 1). The factors for statistical rectangles, months and years were modelled as fixed effects. A strongly right-skewed distribution was found for monthly fishing times per rectangle. We used the Box-Cox method relying on a maximum likelihood estimation to estimate the best power transformation of fishing time that would achieve normality (Draper and Smith 1998). Consequently, fishing time is normalised by a fourth-root transformation. The full model for fleet fishing time T_{ijk} in month *i,* rectangle *j* and year *k* is defined as

$$(1) \qquad T_{ijk}^{1/4} = m + \delta T_{(i-1)jk}^{1/4} + month_i + rectangle_j + year_k + \varepsilon_{ijk}$$

for $i = 1,...,12; j = 1,...,48; k = 1,...,8$, assuming $\delta T_{(0)jk}^{1/4} = \delta T_{(12)jk}^{1/4}$,

and where $\varepsilon \sim N_{n=4488}(0,\Sigma)$.

To take into account that fishing time in a rectangle might be correlated with fishing time in neighbouring rectangles, we include a spatial covariance structure with a nugget effect, and specify the covariance matrix as $\Sigma = \sigma^2 H(\varphi) + \tau^2 I$ where $(H(\varphi))_{jj'} = \rho(\varphi; d_{jj'})$, $d_{jj'}$ is the Euclidean distance between rectangle *j and j',* φ *is* the decay parameter, ρ is chosen as the classical exponential covariance function (see for example (Cressie 1993)) of fishing times in neighbouring rectangles *j'* in the same month *i* and year *k* and τ^2 is the nugget effect variance. The Euclidean distance between two rectangles is calculated using the centre of the rectangles identified by its geographical coordinates in degree (Longitude, Latitude). The model (1) includes the term $\delta T_{(i-1)jk}^{1/4}$ for describing the dependence of fishing time in a given rectangle j on the previous month's fishing time in the same rectangle (including the transition between December (i=12) and January (i=1)).
The model is parametrized using classical treatment contrasts for coding of factors. In the following, the first level is set to 0 for each factor (*January* for *month*, *25E1* for *rectangle* and *1991* for *year*) and thus each coefficient represents the difference between that level and level one.
Model comparison and selection was carried out using Akaike's information criteria (AIC) (Akaike 1974; Pinheiro and Bates 2000). Residual plots were used to check model assumptions (McCullagh and Nelder 1989). All models were fitted by maximum likelihood using R 2.5.1 (http://www.r-project.org).

2.3. Clustering algorithm with contiguity constraints

A hierarchical cluster analysis (HCA) (Lebart et al. 1997) is performed for grouping levels of spatial and temporal variables, using the set of dissimilarities for each pair of *spatial variables* (or *temporal variables*). In addition, we impose contiguity constraints on the set of allowable classification solutions: the objects in a class are required not only to be similar to one another, but also to comprise a spatial (or temporal) contiguous set of objects. For this, neighbours of a statistical rectangle are the eight adjacent rectangles and neighbours of a given month are the previous and following month. The only simple appropriate clustering method using contiguity constraints is the single linkage method. In practice, we implement a crude version of the *single linkage* clustering method with seasonal (or spatial) constraint using a classical hierarchical clustering algorithm (function `hclust` in R), setting the dissimilarities between non-adjacent months (or rectangles) to high values.

If η is the dissimilarity in the HCA , we define γ as the aggregate (joining) index in the usual HCA for *rectangle* (resp. *month*) by :

$$\gamma(rectangle_i, rectangle_{i'}) = \eta(rectangle_i, rectangle_{i'}) + \kappa(rectangle_i, rectangle_{i'})$$

where κ is the contiguity index defined by $\kappa(rectangle_i, rectangle_{i'}) = 0$ if the contiguity constraint is satisfied for $rectangle_i$ and $rectangle_{i'}$, else $\kappa(rectangle_i, rectangle_{i'}) = +\infty$. In the HC algorithm, γ is then used as set of dissimilarities. The clustering results are not sensitive to the actual value. We detail below the computation of the dissimilarities η .

2.4. Raw data Clustering

Clustering methods are classically applied to raw data. To demonstrate the necessity of using a based-model clustering approach, we first applied the hierarchical cluster method with imposed spatial contiguity constraints directly to the raw fishing effort data. For identifying fishing zones (similarly fishing seasons), averages of fishing times per ices-rectangle (similarly, per month) over the study period were calculated and the dissimilarity η between two rectangles (or two months) was calculated as the squared differences between respective averages of the fishing times.

2.5. Model-based clustering

To identify homogeneous fishing time areas and seasons, clustering of model coefficients is carried out using dissimilarities calculated using the 1-p values of statistical tests on the estimated coefficients. Let us consider a special case of the general method for constructing tests for general linear models for hypotheses involving linear functions of parameters. We denote β the vector of parameters:

$$\beta = \left[m, \delta, month_{2}, ..., month\text{©û©}, rectangle\text{©ü}, ...rectangle\text{©þ}, year\text{©ü}, ..., year\right]^T \in \Re^{67=1+1+11+47+7} .$$

We use a F-test to test the equality of pairs of coefficients for factor *month* (or *rectangle*) (Searle 1971; Rawlings et al. 2001).

We express model (1) as a "classic" general linear model $T^{*1/4} = X^*\beta + \varepsilon^*$ where $\varepsilon^* \sim N(0, \sigma^2 I)$ and maximum likelihood (ML) estimates of the model parameter vector $\beta \in \Re^{67}$ is obtained by solving an ordinary least-squares problem. For example, the single null hypothesis that two coefficients $month_i$ and $month_{i'}$ (or $rectangle_j$ and $rectangle_{j'}$) are equal is :

$$H_0 : K^T\beta = 0 \text{ against } H_1 : K^T\beta \neq 0$$

In our case, for example, to test $H_0 : month_i = month_{i'}$ against $H_1 : month_i \neq month_{i'}$, K is a row vector of length 67 with the *i*-th element $K_i = 1$, the *i'*-th element $K_{i'} = -1$ and zeros elsewhere.

The sum of squares for the hypothesis can be written $Q = (K^T\hat{\beta})^2 / (K^T (X^{*T} X^*)^{-1} K)$ and has 1 degree of freedom. Thus, using classical notions of general linear models, the *F*-ratio, equivalent to a *t*-test is :

$$\mathrm{F} = (\mathrm{Q}/1)/(\mathrm{s}^2) = (\mathrm{K}^T\hat{\beta})^2 / (K^T (X^{*T} X^*)^{-1} K)\, \mathrm{s}^2) \sim_{\mathrm{H}_0} F(1,4421)$$

where $s^2 = SC_{res}/4421$ and SC_{res} is the residual sum of squares of the model and 4421=4488-67 corresponding to the number of degrees of freedrom.

A hierarchical cluster analysis (HCA) (Lebart et al. 1997) is then performed for grouping levels of spatial and temporal variables, using the set of dissimilarities produced by the *(1-p)*-values of the previous *F*-tests for each pair of factor levels *month* (or *rectangle*) and the contiguity constraints.

As we are not interested in the complete hierarchy but only partitions, we select one of the solutions in the nested sequence of clusterings that comprise the hierarchy in cutting the dendrogram at a particular height (sometimes termed the *best cut* indicated by large changes in fusion levels in the appearance of the dendrogram based on visual inspection (Everitt et al 2001)).

3. Results

Exploratory data analysis showed that average monthly fleet fishing times varied considerably in space suggesting a strong rectangle effect (Figure 1). In contrast, monthly averages (respectively annual monthly averages) did not show any specific month (respectively year) patterns. For each statistical rectangle, temporal autocorrelation was analysed using the Durbin Watson statistic and we concluded that there existed a temporal autocorrelation of order one for all but two rectangles, 26E3 and 29E5 (Figure 3a). These two rectangles were characterized by random fishing times. We explored the spatial autocorrelation structure plotting the semi-variogram. Figure 3b) shows an exponential spatial autocorrelation structure. The neighbourhood retained was 3.162278 degrees (190 nautial miles), that is the eight adjacent rectangles.

Several models nested within the full model (eq. 1) were fitted to the fishing time data. Residual plots did not show any strong trend, indicating that all the models had reasonable fits (Figure 4). The comparison of AIC values for the different models showed that overall temporal variations were more important than spatial variations (Table 1). The best model (smallest AIC) was the full model, which included both temporal and spatial correlations. The correlations between explanatory variables were examined and found to be less than 0.01. The temporal relationship with fishing time in the previous month was highly significant and estimated as $\delta = 0.47$ (95% Confidence interval [0.45; 0.5]) (Table 2). The estimated spatial correlation coefficient was $\varphi = 0.78$ (95% Confidence interval [0.71; 0.86]) demonstrating a strong spatial pattern in fishing time allocation. Year was not a significant factor (Table 2), which indicates that spatial and temporal exploitation patterns were stable over the study period.

We applied the clustering algorithm with contiguity constraints to raw fishing time. We only present the results for the spatial study but a similar conclusion was obtained for the temporal study. The dendrogram for the single linkage clustering with spatial constraints of the mean fishing effort showed a strong chaining effect (Figure 5). Thus it did not provide interpretable results and the choice of a particular partition is not obvious.

The application of the clustering algorithm to the estimated month and rectangle coefficients of the full model resulted in 22 fishing areas by grouping statistical rectangles (Figure 6) and 9 fishing periods by grouping months (Figure 7). The number of clusters was determined by the best cut on the dendrogram. These fishing areas and periods exhibit similar fishing exploitation patterns.

The identified seasons were : 1) January, 2) February, 3) March, 4) April, 5) May, 6) June, 7) July and August, 8) September, October and November and finally 9) December. January is the month during which the fishing activity is the most intense, closely followed by March and then July, August and April (Table 3). Given that January and March are not consecutive months, the clustering analysis did not group them in the same cluster. The fishing activity of the studied fleet was the least in December followed by February, November, May, October, September and June. Differences in month effects were larger in winter and spring than in summer (Table 3). All winter and spring months therefore appeared in separate clusters leading to 7 distinct seasons in winter and spring and two in summer. December, January and February stood out as the months that were grouped last in the clustering process. With respect to the estimated month effects and the number of seasons, we concluded that fishing activity was less stable in winter than in summer. The longest season was obtained in autumn (September to November) and this season was characterized by an intermediate fishing activity.

The spatial clustering partitioned the rectangles into 22 areas of similar fishing effort, shown with the same colour with respect to their fishing effort level (Figure 8). The number of rectangles per fishing area varied from one to six. The coastal clusters were smaller (for example, areas 1, 2, 4, 5) than the off-shore clusters (for instance, areas 12, 13, 22). Five off-shore fishing areas (areas 3, 6, 8, 24, 21) were however defined by a single rectangle. Area 3 (31E3 rectangle) and area 6 (25E3 rectangle) were the most visited fishing zones, whereas area 8 (27E1 rectangle) was among the least visited fishing zones. Regarding their estimated effects, fishing times in these Ices-rectangles contrast with fishing times in the neighbouring rectangles (Table 4). Areas 19 and 21 are located on the shelf break and probably are subject to both shelf and deep water fishing activities. This might explain their position within single-rectangle clusters. The fishing zones most visited by the French trawler fleet were located South of Ireland (areas 3, 12, 13), off Cornwall (area 9) and off the West of France (areas 6, 7). The waters close to the Irish and English cost were the least visited areas by the French fleet (areas 1,2,4,5, 16). The largest homogeneous fishing area (6 rectangles, area 22) was located in the South-Western part of the Celtic sea.

4. Discussion

4.1. Model-based clustering

We propose a method for characterising spatial and temporal patterns in fishing effort based on a hierarchical cluster analysis of coefficients from a linear model with imposed constraints of spatial and temporal contiguity.

Clustering methods are classically applied to raw data. However, this could be not appropriate when contiguity constraints are used. The single linkage criterion applied to the raw (noisy) data induced a strong chaining structure and the derived dendrogram did not allow valid clustering. In contrast, applying the proposed method to model coefficients, no chaining effect occurred in the dendrogram (Figure 6) and a valid and interpretable partition could easily be performed (best cut at 22 clusters). Indeed, fitting a model accounting for month, rectangle and year explanatory variables allowed to estimate jointly seasonal and spatial effects while filtering the data from inter-annual variations. As the exploratory analysis showed the presence of auto-correlation structures in the fishing time data, the statistical model explicitly took them into account. Consequently, our model-based approach overcame the problem of chaining which is characteristic of single linkage clustering.

The method was illustrated by an application to fishing time data for the French trawler fleet fishing in the central Celtic Sea during the period 1991-1998. In the example, spatial correlations were modelled by an exponential function and temporal correlations by introducing previous month's fishing time in a given rectangle as predictor in the model. In this study, due to the absence of fishing in some rectangles during some months there was an unequal number of observations in each cell. The total number of observations is 4488 instead of 4608(=12*48*8) in the case of a balanced design, representing less than 3% of missing values. As our factorial design is unbalanced, the orthogonality property of main effects (and interactions also) present in balanced data is no longer valid (Montgomery 2005). This means that changing the order of the factors in the model could lead to differences in estimated effects. Fortunately, in our case the imbalance is too small to impact the results of the model. Moreover, as the total number of observations compared to the number of missing values is large, the results of the hierarchical cluster analysis performed on 1-p values from the F tests are not affected by this problem of missing observations (the critical region of the F test with 5% significance level is nearly the same $f_{95\%;1,4421} \approx f_{95\%;1,4541} \approx 3.84$). But, it is important to stress that our method is suitable only for balanced design or nearly balanced design for a large data set. In other cases, users can encounter problems with fitting the model or /and performing the cluster analysis.

In this paper, we selected as the full model a model with only main effects rather than a model with interactions for which model parameters would have been more difficult to interpret. In the Celtic sea case study no clear interaction between month and rectangle was detected by exploratory analyses. Nevertheless, the introduction of a month*rectangle interaction into the model could improve its goodness of fit but would above all increase the difficulty of interpretation. If interactions were significant, the method might be adjusted to provide season-areas with similar patterns. Thus instead of clustering the coefficients of rectangles and months separately, the resulting estimated coefficients *month*rectangle* would be clustered in two steps. First fixing the month, it would lead to maps of fishing areas per month. Second, fixing the rectangle, it would lead to one seasonal year-split per rectangle. This approach will be considered in a future analysis using indicators to quantify temporal stability and spatial heterogeneity of fishing areas. Other recently proposed spatio-temporal models (Banerjee et al. 2004) such as STAR (spatio-temporal autoregressive) models or Bayesian spatial models might be another approach to deal with interactions.

In the context of mixed fisheries, management advice needs to be based on integrated approach accounting for fleets and species, spatial and temporal features of the fishery. Fishing time data inform on where, when and how long fishermen fish and hence this data is suitable for describing fishing zones and seasons. However, spatial management requires additional information to these raw definitions of zones and seasons to regulate fishing access. For instance it would be necessary to characterize each zone and season by the set of targeted species, using catch data available in commercial log-books as additional explanatory variables in the model. In this study, we have not used catch data given that reported catches often have a poor spatial resolution and might bias the perception of the catch process. Indeed, the onboard sorting process leads to discards which are not reported and still not-sufficient understood to provide an accurate estimate of real catch (Rochet et al 2002). Considering the fishery as a group of vessels sharing a similar fishing activity (ICES 2004), an alternative approach integrating target species information might be to split the studied fleet into fisheries and then to apply the proposed modeling scheme to each fishery separately.
Finally, the approach used in this study is generic, and can be generalised to any response variable. For instance, the method could be applied for analysing spatial and temporal patterns in monthly catches.

4.2. Temporal and spatial fishing patterns in Celtic Sea

French trawlers had a tendency to return to the same statistical rectangles in subsequent months. This was shown by a positive temporal autocorrelation. Given that total fleet fishing times per rectangle were analysed, this pattern could have been produced by the same or by different vessels. Similar effects have been reported for the New England trawl fishery (Holland and Sutinen 1999) and the New Zealand hoki fishery (Vignaux 1996b). In Holland and Sutinen (1999), variables describing behaviour types, i.e., fishing areas visited during the previous ten days, were the most important explanatory variables when modelling fishing revenues. Vignaux (1996b) found that the previous day's fishing location significantly explained the choice of the fishing area on a given day. In this study, spatial correlations of fishing times between neighbouring rectangles in a given month were found. This may be explained by assuming similar fish habitats and consequently similar population abundances and community structures. It must be noted that statistical rectangles are arbitrary spatial units. If independent information concerning the relevant ecological factors were available, these could be used in the model.

The cluster analysis performed on the model coefficients for months and rectangles resulted in 22 fishing areas and nine fishing periods. Seasonal and spatial patterns, either in CPUE data (Vignaux, 1996b; Silvano et al 2001) or in fishing times (Greenstreet et al. 1999; Jenning et al. 1999; Béné and Tewfik 2000), have been identified for many fisheries.

The activity of the French fleet was found to be stable during summer and autumn months (two seasons over five months). The same seasonal pattern was also pointed out by Greenstreet et al (1999) for the UK trawler fleet in North Sea. In contrast, during winter and spring the fishing activity appeared to be much more variable from one month to another as each month formed a separate cluster (seven seasons over seven months). In this analysis, December was not surprisingly identified as the month characterized by the lowest fishing time : the second part of December is known to be a period off, specially for French vessels going out for one or two weeks-trips.
The size of the estimated fishing areas was rather variable: one area consisted of more than six statistical rectangles (area 22) whereas others were made up of only one rectangle (area 1 to area 8, area 10, area 16, area 17, area 19, area 21). This suggests that the latter set of rectangles had their own particular fishing dynamics despite significant overall spatial correlations of fishing times. The South-Western Celtic sea seems to be more homogeneous than the borders, which often consisted of single rectangle clusters such as cluster 1 close to the Irish coast and cluster 2 close to Cornwall. These results confirm earlier observations and might be explained by more pronounced heterogeneity in coastal fishing activities compared to off-shore fishing (Biseau et al. 1999; Greenstreet et al. 1999). However, three off-shore areas (8, 19 and 21) are clusters constituted by a single rectangle. Areas 8 and 19 are rectangles characterized by average monthly fishing efforts smaller than their six neighbouring rectangles, but no other obvious factors can explain this differences. By contrast, area

21 is a rather atypical fishing area. This rectangle is situated on the continental slope with large variations in bathymetry. As also indicated by the exploratory analysis (Figure 1), the statistical rectangle 31E3 formed a cluster on its own characterised by consistently the largest fishing times (Table 4). This large discrepancy between this fishing area and the others has already been observed by Pelletier and Ferraris (2000) and is still observed in 2005 (SIH-Ifremer 2007). This area is an attractive fishing zone because of the availability of valuable species. It is known to be visited by vessels targeting Nephrops (*Nephrops norvegicus*) (Coull et al 1998) and has also been characterized by a large abundance of whiting (*Merlangius merlangus*) (Verdoit et al 2003). The large fishing times in areas 13 and 12 might be similarly explained (Biseau et 1999). Regarding area 6 and area 7, large CPUEs of megrim (*Lepidorhombus wiffiagonis*) and monkfish (*Lophius piscatorius and L. budegassa*) (Petitgas et al 20003 and Biseau et al 1999) but also the proximity to fishing ports might explained their high fishing times. These two areas are crossed on the way back, and consequently might be used for last fishing operations.
The fishing areas and periods obtained by our method can be used in several ways, for example as the basic units in a spatial management model (Pelletier et al 2001; Mahévas and Pelletier 2004; Pelletier and Mahévas 2005) or in a spatial stock assessment model (Stefansson and Palsson 1997) or to apportion global fishing mortality in space and time.

Acknowledgment

We would like to thank Ludger Evers and Céline Metaireau for carrying out preliminary analyses, Steven Juggins for suggestions for clustering algorithms, Richard Tomassone and two anonymous referes for comments on the manuscript. This study was performed using logbook data registered by the French Fishery ministry (DPMA) and extracted from Harmonie, the database containing the French Fisheries Information System managed by Ifremer. This work was funded by the European Commission, project contract TECTAC (Q5RS-2002-01291).

References

Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19: 716-725.
Anderson, T.W. 1994. The statistical Analysis of Time series. I. John Wiley & Sons. New-York, USA.
Babcock, E.A. and Pikitch E.K. 2000. A dynamic model of fishing strategy choice in a multispecies trawl fishery with trip limits. Canadian Journal of Fisheries and Aquatic Sciences. 57: 357-370.
Banerjee, S., Carlin, B. P., Gelfond, A. E. 2004. Hierarchical Modeling and Analysis for Spatial Data, Chapman & Hall/CRC, New-York, USA.
Béné, C. and Tewfik, A. 2000. Analysis of fishing effort allocation and fishermen behaviour through a system approach. CEMARE No. 155.
Biseau, A. and Gondeau, O. 1988. Apport des méthodes d'ordination en typologie des flottilles. Journal du Conseil International pour l'Exploration de la Mer 44: 286-296.
Biseau, A. 1998. Definition of a directed fishing effort in a mixed-species trawl fishery and its impact on stock assessments. Aquatic Living Resources 11 (3): 119-136.
Biseau, A., Maguer, C. and Sanz-Aparicio, C. 1999. Pêcheries bigoudènes. Bilan des connaissances. Contrat CE (DG XIV) N°97/0028 No.
Booth, A.J. 2000. Incorporating the spatial component of fisheries data into stock assessment models. ICES Journal of Marine Science 57: 858-865.
Cressie, N.A.C; 1993, *Statistics for spatial data*, Wiley series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley and Sons Inc. New York.
Coull, K.A., Johnstone,R. and Rogers, S.I.. 1998. Fisheries Sensitive Maps in British Waters. UKOOALtd. 58pp.
Davidian, M. and Giltinan, D.M. 1995. Non linear models for repeated measurement data. Chapman & Hall/CRC, New-York, USA.
Draper, N.R. and Smith, H. 1998. Applied Regression Analysis. John Wiley and Sons. New-York, USA.
Everitt, B.S., Landau, S., Leese, M. 2001. Cluster Analysis, Arnold, London, UK.

Gaertner, D., Pagavino, M. and Marcano, J. 1996. Utilisation de modèles linéaires généralisés pour évaluer les stratégie de pêche thonière à la senne en présence d'espèces associées dans l'Atlantique ouest. Aquatic Living Resources **9**: 305-323.
Gordon, A. D. 1981. Classification. Chapman & Hall/CRC, New-York, USA.
Gordon, A. D. 1987. A review of Hierarchical Classification. Journal of the Royal Statistical Society 150(2): 119-137.
Gordon, A. D. 1996. A survey of constrained classification. Computational Statistics and Data Analysis 21: 17-29.
Greenstreet, S.P.R., Spence, F.B., Shanks, A.M. and McMillan, J.A. 1999. Fishing effects in north-east Atlantic shelf seas: patterns in fishing effort, diversity and community structure. II. Trends in fishing effort in the North Sea by UK registered vessels landing in Scotland. Fisheries Research 40: 107-124.
Hartigan, J.A.1975. Clustering Algorithms. Wiley, New York.
Hartigan, J. A. and Wong, M. A. 1979. A K-Means Clustering Algorithm. Applied Statistics 28: 100-108.
Hilborn, R. and Ledbetter, M. 1985. Determinants of catching power in the British Columbia salmon purse seine fleet. Canadian Journal of Fisheries and Aquatic Sciences 42: 51-56.
Holland, D.S. and Sutinen, J.G. 1999. An empirical model of fleet dynamics in New England trawl fisheries. Canadian Journal of Fisheries and Aquatic Sciences 56: 253-264.
ICES 2004. Report of the Study Group on the Development of Fishery-based Forecasts. CM/ACFM 11.
Jenning, S., Alvsvag, J., Cotter, A.J.R., Ehrich, S., Greenstreet, S.P.R., Jarre-Teichmann, A., Mergardt, N., Rijnsdorp, A.D. and Smedstad, O. 1999. Fishing effects in north-east Atlantic shelf seas: patterns in fishing effort, diversity and community structure.III. International trawling effort in the North Sea: an analysing of spatial and temporal trends. Fisheries Research 40: 125-134.
Kendall, M. and A. Stuart (1979). The Advanced Theory of Statistics. Vol 3. Charles Griffin & Compagny Limited. London.
Lebart, L., Morineau, A. and Piron, M. 1997. Statistique Exploratoire Mutidimensionnelle. Dunod (2nd Edition). Paris, France.
Legendre, P. and Legendre, L. 1998. Numerical Ecology. Developments in Environmental Modelling 20. Elsevier.
Mahévas, S. and Trenkel, V. 2002. Utilisation de modèles mixtes pour décrire la distribution spatio-temporelle du temps de pêche de la flottille française en mer Celtique. Journal de la Société Française de Statistiques - Modèles Mixtes et Biométrie 143: 177-186.
Mahévas, S., Sandon, Y. and Biseau, A. 2004. Quantification of annual variations in fishing power and explanation of differences in efficiency by technical characteristics: an application to the bottom-trawlers of South-Brittany targeting anglerfish from 1983 to 1998. ICES Journal of Marine Science 61: 71-83.
Mahévas, S. and Pelletier, D. 2004. ISIS-FISH, a generic and spatially explicit simulation tool for evaluating the impact of management measures on fisheries dynamics. Ecological Modelling 171: 65-84.
Marchal, P. 1997. Managing growth overfishing with multi annual compromise strategies. Canadian Journal of Fisheries and Aquatic Sciences 54: 2255-2276.
McCullagh, P. and Nelder, J.A. 1989. Generalized Linear Models. Chapman and Hall. New-York, USA.
Montgomery, D.C.; 2005, *Design and Analysis of Experiments*, 6th Ed., John Wiley and Sons Inc. New York.
Pelletier, D. and Ferraris, J. 2000. A multivariate approach for defining fishing tactics from commercial catch and effort data. Canadian Journal of Fisheries and Aquatic Sciences **57**: 51-65.
Pelletier, D., Mahévas, S., Poussin, B., Bayon, J., Andre, P. and Royer, J.T. 2001. A conceptual Model for Evaluating the Impact of Spatial Management Measures on the Dynamics of a Mixed Fishery. . Spatial Processes and Management of Marine Populations Alaska Sea Grant College Program. AK-SG-01-02: 54-66.
Pelletier, D. and Mahévas, S. 2005. Fisheries simulation models for evaluating the impact of management policies, with emphasis on marine protected areas. Fish and Fisheries **6**: 307-349.
Petitgas, P., Poulard, J.-C., and Biseau, A. 2003. Comparing commercial and research survey catch per unit of effort: megrimin the Celtic sea. ICES Journal of Marine Science, 60: 66-76.
Pinheiro, J.C. and Bates, D.M. 2000. Mixed-Effects Models in S and S-Plus. Statistics and Computing Springer and Verlag. New-York, USA.
Poulard, J. C. 2001. Distribution of hake (Merluccius merluccius, Linnaeus, 1758) in the Bay of Biscay and the Celtic Sea from the analysis of French commercial data. Fisheries Research 50: 173-187.

Rawlings, J. O., Pantula, S. G., Dickey, D. A. 2001. Applied Regression Analysis: A Research Tool, 2nd ed., Springer and Verlag. New-York, USA.
Sampson, D.B. 1991. Fishing tactics and fish abundance, and their influence on catch rates. ICES Journal of Marine Science 48: 291-301.
Searle S. R. 1997. Linear Models, Wiley Classics Library, New York, USA.
SIH-Ifremer, 2007. Synthèse des flottilles de pêche 2005. Flottle mer du Nord-Manche-Altantique. Système d'Informations Halieutiques de l'Ifremer.55p.
Silvano, R.A.M. and Begossi, A. 2001. Seasonal dynamics of fishery at the Piracicaba River. Fisheries Research 51: 69-86.
Sparre, P. 2003. An EXCEL-based software toolbox for stochastic fleet-based forecast. ICES Annual conference. CM 2003/V:07.
Stefansson, G. and Palsson, O.K. 1997. BORMICON : A Boreal Migration and Consumption model. Marine Research Institute No. 59.Reykjavik, Iceland.
Venables, W.N. and Dichmont, C.M. 2004. GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. Fisheries Research 70: 319-337.
Verdoit, M., Pelletier, D. and Bellail R. 2003. Are commercial logbook and scientic data useful for characterizing the spatial and seasonal distribution of exploited populations? The case of the Celtic Sea whiting. Aquatic Living Resources. 16: 467-485.
Vignaux, M. 1996a. Analysis of spatial structure in fish distribution using commercial catch and effort data from New Zealand hoki fishery. Canadian Journal of Fisheries and Aquatic Sciences **53**: 963-973.
Vignaux, M. 1996b. Analysis of vessel movements and strategies using commercial catch and effort data from the New Zealand hoki fishery. Canadian Journal of Fisheries and Aquatic Sciences **53**: 2126-2136.

Tables

Table 1. Comparison of different model fits (AIC) for total fishing time ($T^{1/4}$) allocated by month and rectangle in the Celtic Sea by the French trawler fleet during 1991-1998; df degrees of freedom.

Model	Explanatory variables	**df**	**AIC**
Basic	Month+rectangle+ year	67	13547.95
AR(month-1)	(Previous month's fishing time)$^{1/4}$+Basic	68	12051.66
Spatial correlation	Basic + exp(neighbouring rectangle time)	69	12892.28
Full	Basic + (Previous month's fishing time)$^{1/4}$ + exp(neighbouring rectangle time)	70	**11582.09**

Table 2. Analysis of variance table for full model for fishing time per rectangle and month by the French bottom trawlers operating in the Celtic sea during 1991-98 (see Table 1 for model definition).

Effect	**df**	**F-value**	**p- value**
Intercept	1	61682.91	<0.0001
(Previous month's fishing time)$^{1/4}$	1	8246.55	<.0001
Month	11	8.30	<.0001
Rectangle	47	18.53	<0.0001
Year	7	1.56	0.1415
Corr Struct	**lower**	**est.**	**upper**
Spatial range	0.718	0.784	0.855
nugget	5.967365e-43	2.872111e-08	1.0000000

Table 3 : Fishing seasons and estimated month effects with 95% confidence intervals using model (1).

Season number	Month	Lower bound	Month effect	Upper bound
1	January	0	0	0
2	February	-0.41	-0.64	-0.18
3	March	-0.01	-0.24	0.22
4	April	-0.15	-0.38	0.08
5	May	-0.39	-0.62	-0.16
6	June	-0.23	-0.45	0.004
7	July	-0.11	-0.34	0.11
	August	-0.13	-0.36	0.10
8	September	-0.29	-0.52	-0.06
	October	-0.33	-0.56	-0.10
	November	-0.40	-0.63	-0.17
9	December	-0.78	-1.01	-0.55

Table 4 Fishing zones and estimated rectangle effects with a 95% confidence intervals using model (1).

Zone number	**Ices rectangle**	**Lower bound**	**Rectangle effect**	Upper bound
1	32E1	-1.61	-1.34	-1.07
2	31E5	-1.28	-1.01	-0.73
3	31E3	1.46	1.73	2.01
4	29E5	-1.62	-1.34	-1.06
5	29E6	-1.23	-0.96	-0.69
6	25E3	0.49	0.74	0.99
7	25E4	0.28	0.54	0.79
8	27E1	-0.94	-0.69	-0.44
9	28E5	0.10	0.35	0.61
	28E6	0.19	0.45	0.71
10	26E8	-0.03	0.23	0.49
11	27E7	-0.30	-0.04	0.21
	28E7	-0.27	-0.01	0.25
	29E7	-0.22	0.03	0.29
12	30E2	0.51	0.77	1.03
	31E2	0.42	0.68	0.94
13	28E2	0.39	0.64	0.90
	29E1	0.06	0.31	0.57
	29E2	0.27	0.52	0.78
	30E1	0.17	0.43	0.69
	31E1	0.089	0.35	0.60
14	32E2	-0.51	-0.25	0.01
	32E3	-0.44	-0.19	0.07
15	30E4	-0.03	0.22	0.48
	31E4	-0.14	0.11	0.37
	32E4	-0.20	0.06	0.32
16	26E6	-1.20	-0.93	-0.66
	26E7	-1.12	-0.86	-0.59
	27E8	-1.04	-0.77	-0.51
17	25E5	-0.58	-0.33	-0.07

18	26E5	-0.80	-0.55	-0.29
	27E4	-0.79	-0.54	-0.28
	27E4	-0.85	-0.59	-0.33
	27E6	-0.70	-0.44	-0.18
19	26E1	-0.45	-0.23	-0.01
20	27E5	-0.43	-0.17	0.08
	28E3	-0.56	-0.30	-0.05
	28E4	-0.31	-0.05	0.20
	29E3	-0.59	-0.33	-0.07
	29E4	-0.43	-0.17	0.09
	30E3	-0.60	-0.34	-0.08
21	25E1	0	0	0
22	25E2	-0.002	0.21	0.44
	26E2	-0.06	0.17	0.41
	26E3	-0.21	0.03	0.29
	26E4	-0.20	0.06	0.31
	27E2	-0.20	0.04	0.30
	28E1	-0.34	-0.08	0.17

Figures

Figure 1. Monthly fishing times of the French bottom trawlers in the Celtic Sea averaged over the years 1991-1998. For each statistical rectangle, the grey level is proportional to the monthly average. The rectangle name is the combination of a number stated on the left and a letter-number stated at the bottom of the graphic.

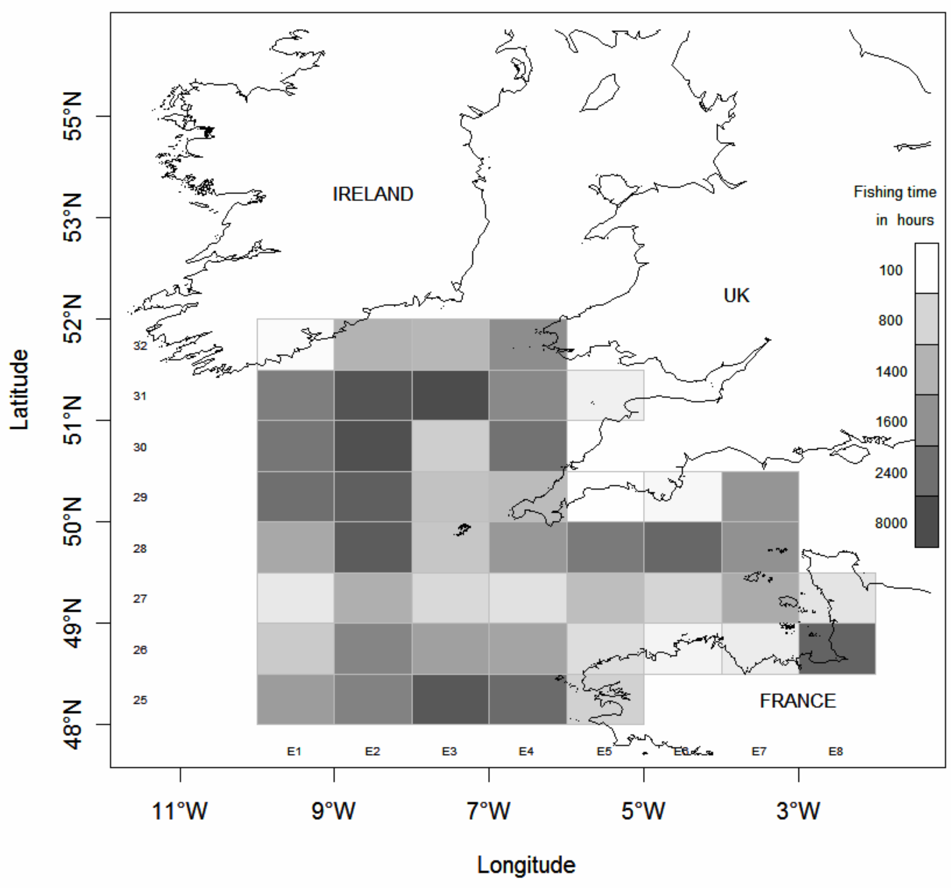

Figure 2 : Flowchart of the proposed modelling approach :1) exploratory analysis of fishing time data, 2) model fitting to estimate month and ices-rectangle effects and 3) cluster analysis to provide fishing seasons and zones.

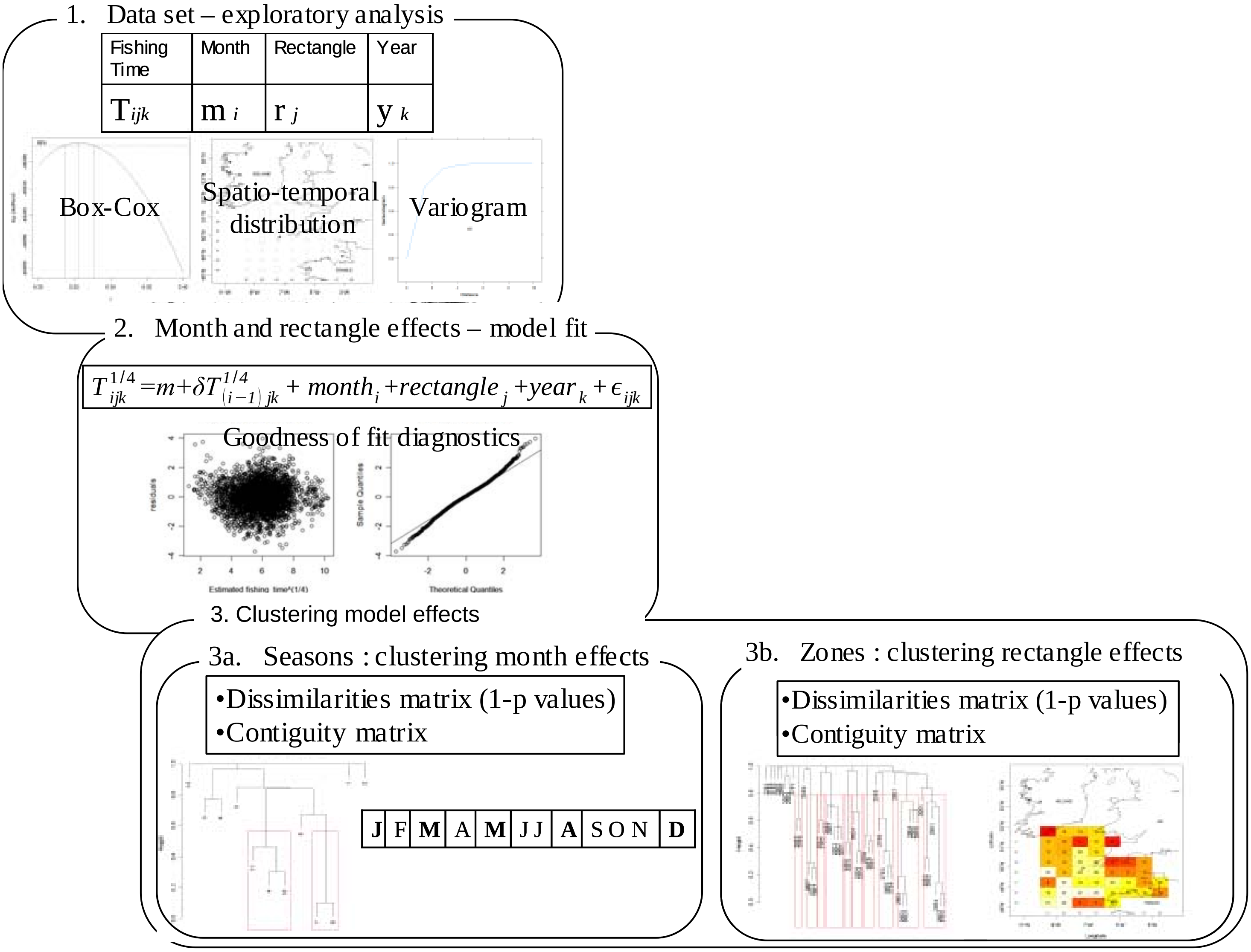

Figure 3 : Exploratory analysis of autocorrelation structures. a) For each rectangle, the coefficient estimates of temporal autocorrelation of order 1 calculated using monthly fishing time series. b) Semi-variogram values corresponding to the exponential correlation model calculated using monthly fishing times per rectangle and the Euclidean distance between rectangles. The selected neighbourhood distance is equal to 3.16 degrees.

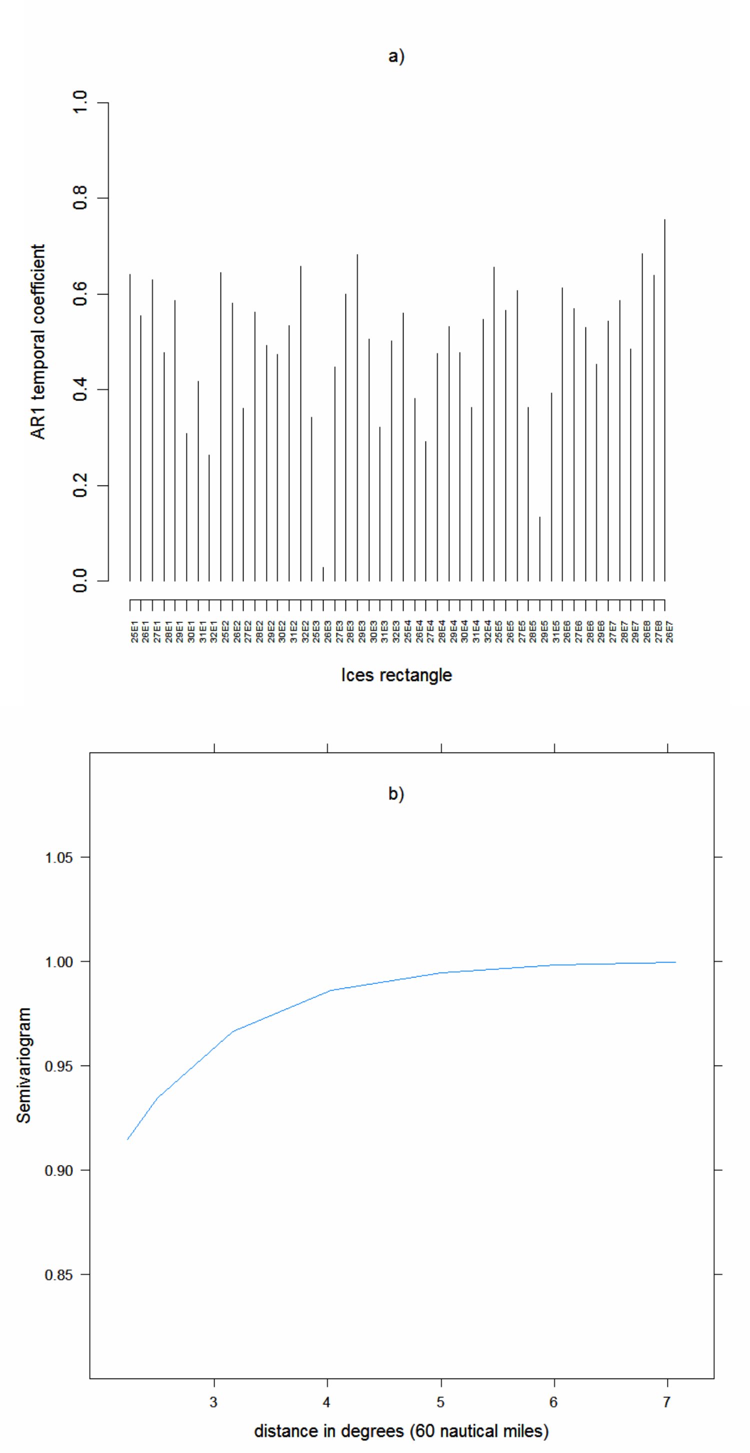

Figure 4. Diagnostic plots : a) Log-likelihood plot for the Box-Cox transformation of fishing time. 95% Confidence interval for the transformation indicates that ¼ is a reasonable choice for the sake of interpretability, b) fitted transformed fishing time versus observed transformed fishing time, c) deviance residuals versus fitted transformed fishing time and d) qq-plots of the deviance residuals. b), c) and d) were generated using model (1).

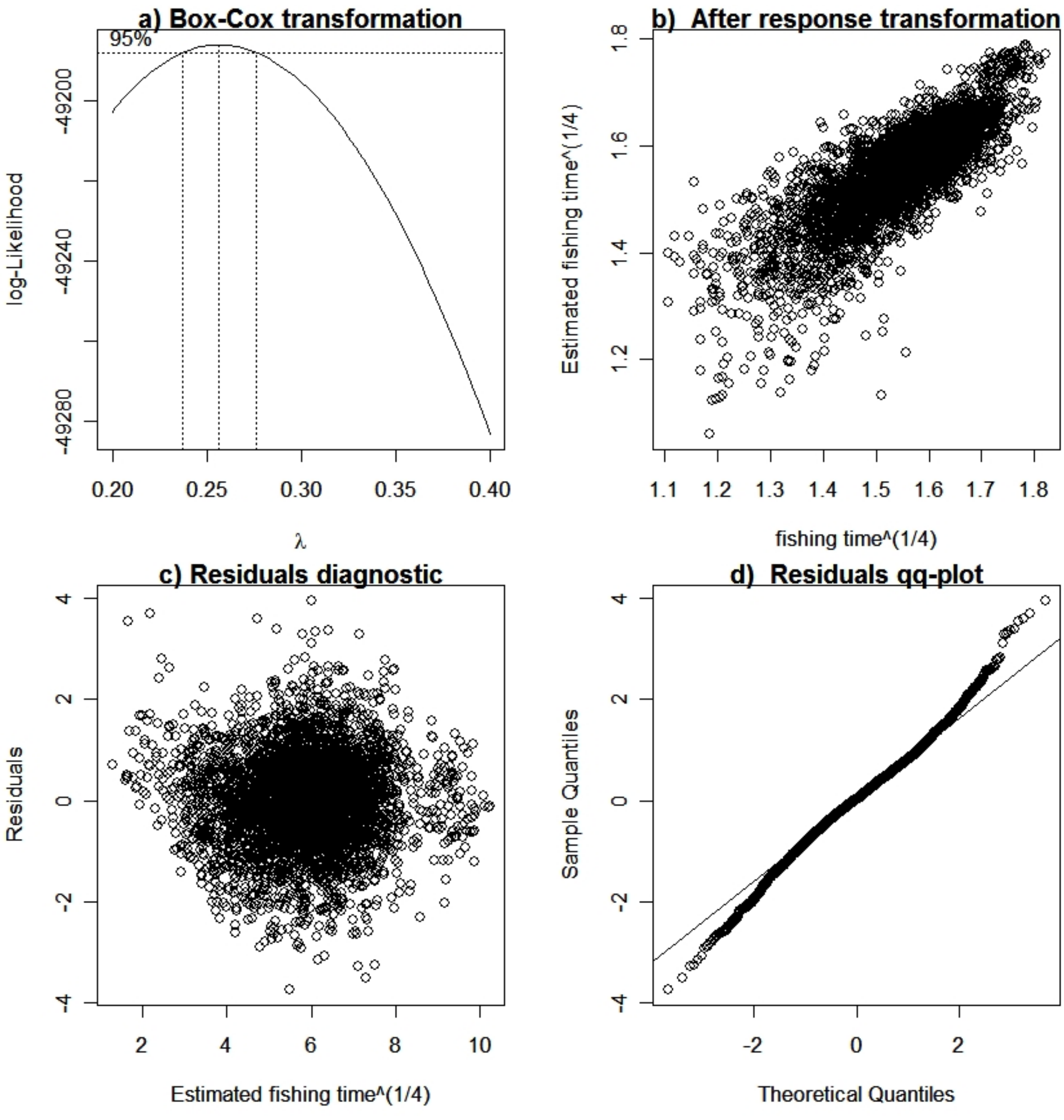

Figure 5. Dendrogram showing single linkage clustering of the Euclidian distance between raw fishing times per rectangle (average fishing time per ICES-rectangle over the study period).

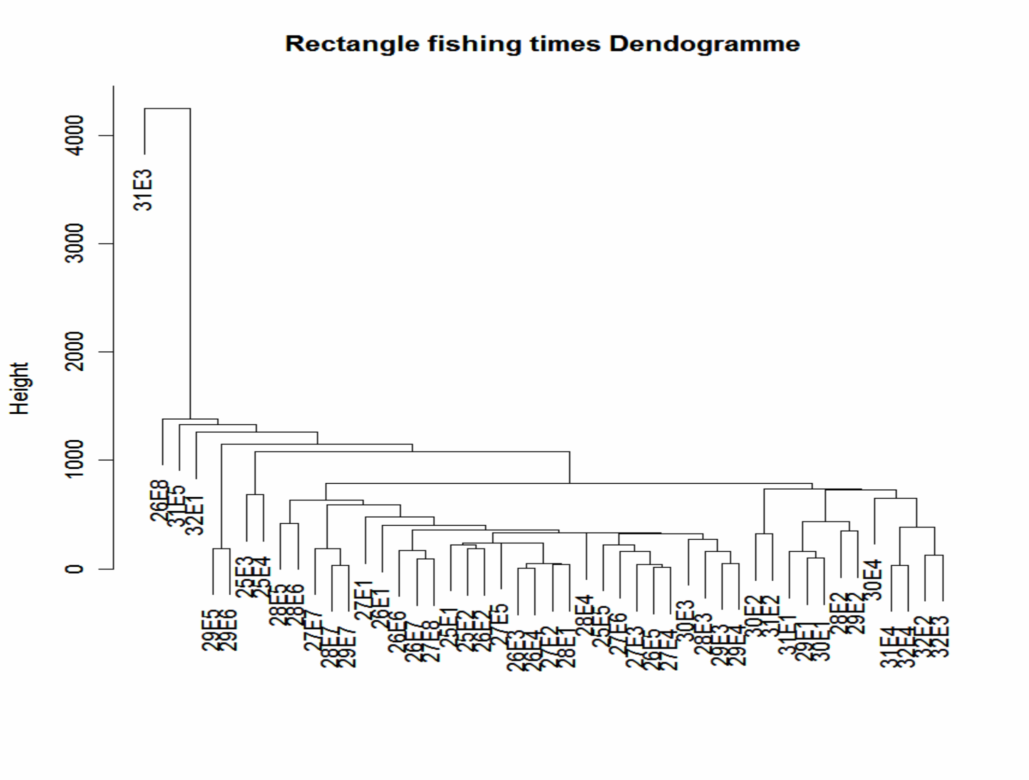

Figure 6. Dendrogram showing single linkage clustering of 1-p values of tests on estimated coefficients of rectangles for which levels are different. The boxes characterize the grouped rectangles by cutting the dendrogram at height 1-p equals 0.75.

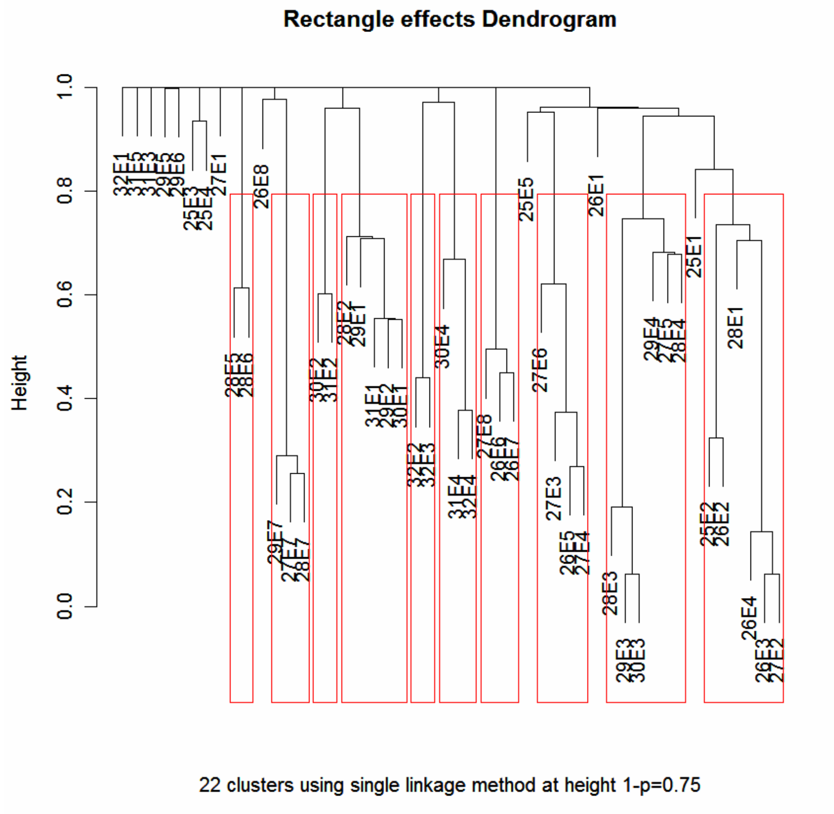

Figure 7. Dendrogram showing single linkage clustering of 1-p values of tests on difference between levels of estimated model coefficients for *month*. The boxes characterize the grouped months by cutting the dendrogram at height p equals 0.5.

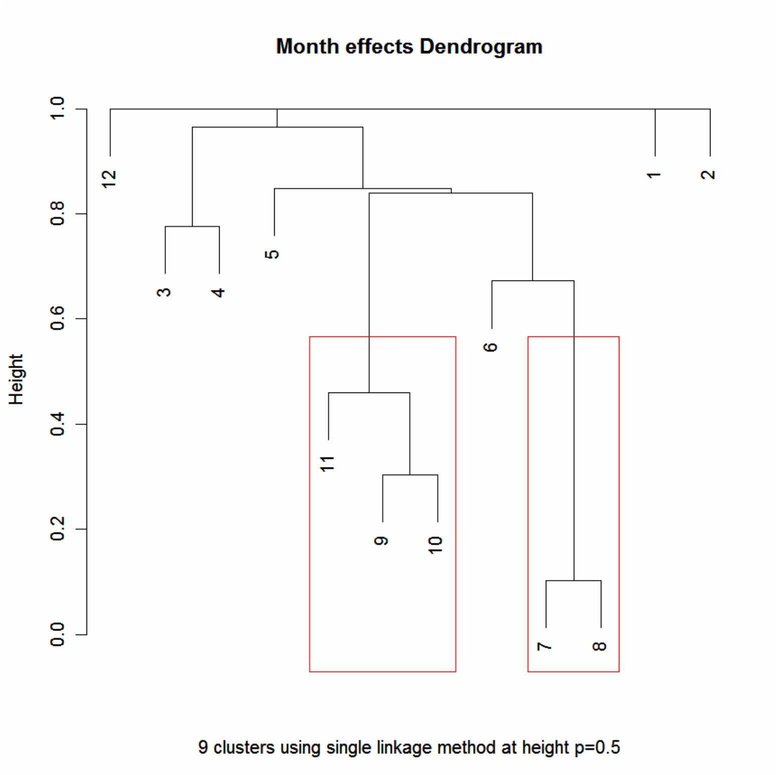

Figure 8. Map of fishing areas resulting from the cluster analysis on estimated model coefficients for model of total fishing time by French trawler fleet in the Celtic Sea. All statistical rectangles belonging to the same fishing area have the same shading and the same number. The grey level is proportional to the estimated spatial fishing effort (average rectangle effect).