Downloaded from orbit.dtu.dk on: Jan 23, 2019

DTU Library

Regional métier definition: A comparative investigation of statistical methods using a workflow applied to the international otter trawl fisheries in the North Sea

Deporte, Nicolas; Ulrich, Clara; Mahévas, Stephanie; Demanèche, Sebastien; Bastardie, Francois

Published in:
ICES Journal of Marine Science

Link to article, DOI:
10.1093/icesjms/fsr197

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Deporte, N., Ulrich, C., Mahévas, S., Demanèche, S., & Bastardie, F. (2012). Regional métier definition: A comparative investigation of statistical methods using a workflow applied to the international otter trawl fisheries in the North Sea. ICES Journal of Marine Science, 69(2), 331-342. DOI: 10.1093/icesjms/fsr197

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

ICES Journal of Marine Science Advance Access published January 19, 2012

ICES Journal of
Marine Science

International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

ICES Journal of Marine Science; doi:10.1093/icesjms/fsr197

Regional métier definition: a comparative investigation of statistical methods using a workflow applied to international otter trawl fisheries in the North Sea

Nicolas Deporte[1], Clara Ulrich[2,*], Stéphanie Mahévas[3], Sébastien Demanèche[1], and Francois Bastardie[2]

[1] *Ifremer, Brest, RBE/STH, BP 70, 29280 Plouzané, France*
[2] *Technical University of Denmark, National Institute for Aquatic Resources (DTU Aqua), Charlottenlund Castle, 2920 Charlottenlund, Denmark*
[3] *Ifremer, Nantes, RBE/EMH, rue de l'ile d'Yeu, BP 21105, 44311 Nantes Cedex 03, France*

**Corresponding Author: tel: +45 21 157486; fax: +45 35 883333; e-mail: clu@aqua.dtu.dk.*

Deporte, N., Ulrich, C., Mahévas, S., Demanèche, S., and Bastardie, F. Regional métier definition: a comparative investigation of statistical methods using a workflow applied to international otter trawl fisheries in the North Sea. – ICES Journal of Marine Science, doi:10.1093/icesjms/fsr197.

Received 13 July 2011; accepted 24 November 2011.

The European Common Fisheries Policy recognizes the importance of accounting for heterogeneity in fishing practices, and métier-based sampling is now at the core of the EU Data Collection Framework. The implementation of such an approach would require Member States to agree on the standard regional métier definitions and on practical rules to categorize logbook records into métiers. Several alternative approaches have been used in the past to categorize landings profiles, but no consensus has yet emerged. A generic open-source workflow is developed to test and compare a selection of methods, including principal components analysis (PCA), hierarchical agglomerative clustering (HAC), *K*-means, and Clustering LARge Applications (CLARA), and to provide simple allocation rules. This workflow is applied to a unique regional dataset consisting of bottom-trawl logbooks of five North Sea countries. No method proved to be infallible, but combining PCA with either CLARA or HAC performed best. For 2008, a hierarchical classification with 14 species assemblages is proposed. Discriminant analysis proved more robust than simple ordination methods for allocating a new logbook record into an existing métier. The whole approach is directly operational and could contribute to defining more objective and consistent métiers across European fisheries.

Keywords: cluster, Data Collection Framework (DCF), logbooks, métier, mixed fisheries, multivariate analysis, North Sea.

Introduction

The European Common Fisheries Policy (CFP) calls for the implementation of an ecosystem-based approach to fisheries management, with increasing focus on limiting the impact of fisheries on the environment. As part of the process, the CFP recognizes the importance of accounting for heterogeneity in fishing practices, and the 2011 Proposal for the Reform of the CFP shows an obvious will to move away from single-stock and towards fleet-based management (EC, 2011). Some steps have already been taken in this direction over the past decade through, for instance, differential effort reductions based on gear and mesh-size categories, and the development of mixed-fisheries scientific advice (ICES, 2010). Analysing catch and effort by fishing activity allows for more accurate estimates of partial fishing mortality induced by the various fleets. To make this approach operational, the first step is to define fishing activities. This topic is not new, and since the seminal work of Laurec *et al.* (1991), classifying fishing activities has been well investigated by the scientific community (see below). The definition of homogeneous groups of fishing operations and/or fishing vessels is intended to characterize the overall fishing activity into a few, easy-to-manage categories. However, despite many years of scientific study, no single and fully unified approach has emerged. Various criteria and scales can be used, which lead to different perceptions of the same reality (Ulrich *et al.*, 2009). Nonetheless, the establishment of the European Data Collection Framework (DCF; EC, 2008) has been an important step forwards, not least because it has led to agreement on basic concepts and terminology. The DCF has adopted the definition that we follow here: a métier is a group of fishing operations targeting a specific assemblage of species, using a

© 2012 International Council for the Exploration of the Sea. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oup.com

Downloaded from http://icesjms.oxfordjournals.org/ at DIFRES on January 23, 2012

specific gear, during a precise period of the year and/or within the specific area.

The DCF defines métiers according to a hierarchical structure using six nested levels: level 1, activity (fishing/non-fishing); level 2, gear class (e.g. trawls, dredges); level 3, gear group (e.g. bottom trawls, pelagic trawls); level 4, gear type [e.g. bottom otter trawl (OTB), bottom pair trawl]; level 5, target assemblage based on the main species type [e.g. demersal fish vs. crustaceans or cephalopods]; level 6, mesh size and other selective devices. For convenience, we also define here a further disaggregation level distinguishing targets at the true species level (e.g. cod, haddock) as level 7, as distinct from DCF level 5, which deals only with species type. This level 7 is expected to describe more accurately the actual landings profile. It must be noted that our use of the level 7 concept here may differ slightly from the current usage at the national level in some cases.

To make these definitions operational Europe-wide, it is important that all coastal EU Member States agree at the regional scale on (i) métier definition at levels 7 and 5, and (ii) practical rules to allocate their own activities to métiers. Logbook data are the main source of information on fishing activities. They detail for each trip of each vessel the quantity of the main species caught and kept on board per catch day, location, and type of gear used. Categorizing logbook data into levels 1–4 (and to a lesser extent level 6, depending on the accuracy of mesh-size reporting) is straightforward because the required information is directly available in the logbooks. Categorization into levels 5 and 7, however, is more difficult, because in the EU, fishers do not have to declare which species they are actually targeting when fishing. Therefore, the matching métier has to be inferred.

ICES (2003) provided general concepts and ideas to define these métiers, but did not provide quantitative guidelines, nor did the expert groups (EC, 2005, 2006) that led to the DCF. However, several analyses have been conducted at the national scale, and many approaches have been used. The earliest and simplest approach consists of selecting fishing trips where a certain catch proportion of selected key species is exceeded (e.g. Biseau, 1998; Ulrich *et al.*, 2001). This approach is based largely on trial and error, and often requires a qualitative, *a priori*, knowledge of the fisheries. Another approach consists of conducting multivariate analyses on species composition in catch data by trip or fishing operation (referred to as catch or landings profiles), then grouping similar profiles into métiers. This grouping can be performed by direct visual inspection (Biseau and Gondeaux, 1988; Laurec *et al.*, 1991) or statistically through cluster analysis. Within this approach, several statistical methods, settings, and software have been used, and many local applications have been published in the literature, also in response to the requirements of DCF implementation. Rogers and Pikitch (1992) used two opposite types of hierarchical clustering techniques and detrended correspondence analysis to define groundfish assemblages in Oregon and Washington waters. Lewy and Vinther (1994) used a hierarchical agglomerative clustering (HAC) analysis when identifying Danish North Sea trawl fisheries, and a similar approach was later used by Holley and Marchal (2004) and Marchal (2008) on French fisheries, by Tzanatos *et al.* (2005) in Greece, and by Jiménez *et al.* (2004) in Spain. He *et al.* (1997), Silva *et al.* (2002), and Bastardie *et al.* (2010) used *K*-means clustering approaches for fisheries in Hawaii, Spain, and Denmark, respectively. Pelletier and Ferraris (2000) combined principal components analysis (PCA) and HAC to identify métiers of both an artisanal Senegalese fishery and French Celtic Sea fisheries, a sequence of methods much used in subsequent studies (Ulrich and Andersen, 2004; Tzanatos *et al.*, 2006; Campos *et al.*, 2007; Katsanevakis *et al.*, 2010). Finally, non-hierarchical clustering methods were used to classify métiers in the Iberian Peninsula, with partitioning around medoids (PAM) used for Portuguese purse-seine fisheries (Duarte *et al.*, 2009) and its variant, Clustering LARge Applications (CLARA), used for Spanish otter trawl fisheries (Castro *et al.*, 2010, 2011; Punzón *et al.*, 2010).

Reviewing these studies raises a number of questions. First, although mostly statistical, clustering has generally included an element of subjective choice, and the robustness of the results to these choices is unknown. Second, although the DCF aims at unifying métier definitions at a regional scale (i.e. across nations operating in the same region), all the analyses described above were performed at the national level, involving limited datasets. The requirement for regional métiers is likely to provide different results by combining different fishing strategies across different member states and may potentially also raise computational challenges associated with larger datasets. Finally, all studies were performed on historical data aggregated over given periods of time (generally per year), but they did not usually address the requirement to assign future logbook records to a métier, as would be useful for real-time monitoring of fisheries.

To address these issues, we developed an operational framework that will allow (i) analysis of the sensitivity of métier definition (at level 7) to the classification method, (ii) linking the métier obtained at level 7 to the target assemblage at level 5, and (iii) categorizing any new logbook records into the most relevant métier class. In this work, focus is solely on level 7 as a way to enhance and operationalize level 5, so we disregard level 6. Although DCF levels 1–5 were defined at a whole European level and meant to be generically available, level 6 was defined regionally and is not available in all regions. Therefore, the methods described below deal only with the analysis of landings profiles and not with their linkages with mesh size, as was attempted by Pelletier and Ferraris (2000), Ulrich and Andersen (2004), and Marchal (2008).

The application of the whole procedure is illustrated with the example of the international OTB fishery using combined logbook data from the main countries (Denmark, England, France, Scotland, and the Netherlands) fishing in the North Sea region, i.e. ICES Subarea IV (North Sea), Division IIIaN (Skagerrak), and Division VIId (Eastern Channel). According to ICES (2010), the bottom trawling component of these nations together accounts for some half of the total landings of the main assessed species (cod, haddock, whiting, saithe, sole, plaice, and *Nephrops*) in the North Sea. This study represents the first attempt to merge national logbook data into a regional dataset, and so in addition to its generic statistical scope, it also represents a major insight into the nature of North Sea demersal trawl fisheries.

Material and methods

The workflow was developed entirely in R (R Development Core Team, 2010). The code associated with this work is included in the "vmstools" R package (http://code.google.com/p/vmstools/), which is a library of tools for fishery data-related analyses (Beare *et al.*, 2011; Hintzen *et al.*, 2011). Some R-specific extension packages were also used (FactoMineR, cluster, SOAR, amap, MASS, mda).

Downloaded from http://icesjms.oxfordjournals.org/ at DIFRES on January 23, 2012

Data

The dataset included detailed 2007/2008 logbook data for OTB fishing in ICES Subarea IV (North Sea), Division IIIa (Skagerrak–Kattegat), and Division VIId (Eastern Channel). The term "logbooks" is used here in its broad sense, implying a merging of the actual logbooks completed by fishing vessels, with cash-value information coming mainly from sales slips and information on the fishing vessel coming from the fleet register, so that complete information is available for each fishing trip (Hintzen *et al.*, 2011). In the EU, a logbook has to be completed for each fishery sequence, i.e. for each combination of fishing day, gear, mesh size, and ICES rectangle. However, in practice, national fisheries research institutes do not all have access to the same level of disaggregation. For some, data are available down to fishing operation (haul-by-haul information), whereas for others, all operations of a given fishing trip are aggregated into a single record, with only the main area fished and gear used indicated. The best information available for each country was retained, and the generic term logbook event (LE) was used to refer to each observation. This implies that for the countries whose available information was disaggregated to the fishing operation or day, a single fishing trip could be characterized eventually by one or several métiers. For 2007, 74 712 LEs were recorded, and for 2008, there were 96 758 (the initial number of LEs in 2008 was 98 017, but 1.3% of LEs had some missing value information and were removed). The exchange format used throughout the study was the standardized EFLALO format, previously used in many research projects (e.g. Marchal, 2006; Beare *et al.*, 2011). In this format, each row represents an LE, and columns include several descriptors (vessel, gear, mesh size, ICES rectangle, etc.), as well as the weight and the value of landings by species. ICES (2003) recommended that métiers be defined preferably on landings composition expressed in cash value if available, as this may reflect more accurately the actual targeting choices of the fishers. This option was retained here. The number of species recorded varied significantly between countries, from 49 (in Scotland) to 220 (in France), and when pooled together, the whole dataset included 278 species in 2007 and 296 in 2008.

In 2008, the ten main species (72% of the total value) landed by bottom trawls in the North Sea region were, in decreasing importance, *Nephrops* (NEP), sandeels (SAN), saithe (POK), Atlantic cod (COD), monkfish (MON), European plaice (PLE), haddock (HAD), herring (HER), whiting (WHG), and mackerel (MAC). However, these species were not equally spread across all LEs, because they were present in, respectively, 47, 3, 22, 42, 27, 46, 30, 4, 32, and 11% of all LEs. This means that some species are more heavily fished in dedicated métiers, often associated with specific countries, illustrating the need to collate data at a supranational level.

Methodology

The comparison of methods for the characterization and classification of landings profiles was performed through a number of sequential steps, as follows (Figure 1):

Step 1: identification of the important species out of all species recorded, and the reduction in the dataset to these key species only;

Step 2: investigation of the added value of initially running a PCA on the dataset to build preliminary groups of species;

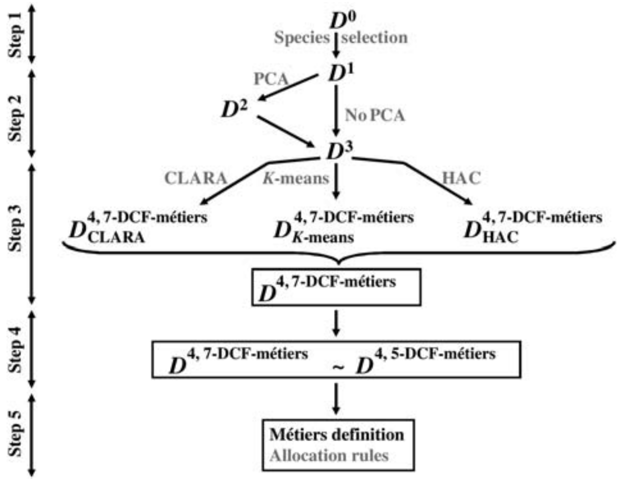
Step 1
Step 2
Step 3
Step 4
Step 5
D^0
Species selection
D^1
PCA
No PCA
D^2
D^3
CLARA
K-means
HAC
$D^{4,7\text{-DCF-métiers}}_{\text{CLARA}}$
$D^{4,7\text{-DCF-métiers}}_{K\text{-means}}$
$D^{4,7\text{-DCF-métiers}}_{\text{HAC}}$
$D^{4,7\text{-DCF-métiers}}$
$D^{4,7\text{-DCF-métiers}} \sim D^{4,5\text{-DCF-métiers}}$
Métiers definition
Allocation rules

Figure 1. Workflow of the analyses for the definition of métiers using landings profiles and the definition of the allocation rules of an LE to a métier. Symbols are explained in the text.

Step 3: running a selection of clustering methods and settings, and characterizing the species-based level 7 classifications obtained;

Step 4: relating the species-based level 7 classifications to the corresponding DCF level 5 classification;

Step 5: predicting the classification of any new LE into the defined métiers.

Each step is described below.

Step 1: identification of main species

Let us denote D^0 the initial dataset, $D^0 = (D^0_{i,j})_{i=1,\dots,n;j=1,\dots,m}$, where n refers to the number of LEs, m the number of species, and $D^0_{i,j}$ the landings in value of species j during the ith LE. Several approaches were suggested for identifying key species out of the large dataset D^0. An objective method was to use an HAC on the species observations, i.e. on the transposed dataset (all species × LE). More subjectively, main species could also be defined as the ranked list of species accumulating to a given percentage of the total catches (the perTotal method), or as the species representing at least a given percentage of at least one LE, i.e. species likely to represent a true target for part of the fishery (the perLogevent method). The three methods were thus implemented and compared.

In the HAC method, groups of species were identified by iterative pairwise agglomeration of elements based on the Ward minimum variance criterion (Ward, 1963). The analysis was carried out on landings proportion by LE (landings profile) rather than on absolute values, to be independent of the LE total value. We used a Scree test (Cattell, 1966) to select the cut height of the dendrogram. The Scree test cuts the dendrogram at the successive largest gain in the clustering variance ratio (variance between clusters/total variance of the dataset). This first run isolated a number of principal species and pooled the remaining species of lesser importance in a group of residual species. As this step generally isolated only few main species, similar HACs were subsequently run through a loop on the residual group, and each new species isolated by a monospecific cluster was

Downloaded from http://icesjms.oxfordjournals.org/ at DIFRES on January 23, 2012

added to the list. The loop stopped when all clusters contained more than one species. In the perTotal method, the percentage threshold was increased in steps of 5% from 5 to 100%, and the ranked species summing up to this value was recorded. In the perLogevent method (working on landings profiles), the percentage threshold was decremented in steps of 5% from 100 to 5%, and all species representing at least this value in at least one LE were recorded.

This approach allowed the exploration of the variability and the sensitivity of the definition of key species to differences in concepts and thresholds, and the derivation of a robust list of species using a combination of outcomes from the three methods. Subsequently, the initial dataset (landings by LE × all species) was transformed into landings profiles and reduced to the principal species only (LE × principal species), defined by $D^1 = (D^1_{i,j})_{i=1,\ldots,n;j=1,\ldots,p}$, where p is the number of principal species ($p \leq m$) and $D^1_{i,j} = D^0_{i,j}/\sum_{j=1}^{m} D^0_{i,j}$.

After that step and based on this conveniently reduced number of species, national differences observed in the coding of unsorted mixed groups of cuttlefish, squid, monkfish, and skates and rays were made consistent to avoid the potential bias in the results.

Step 2: PCA transformation

Many of the studies cited above made use of a PCA before the actual clustering, and it was therefore decided to investigate the relevance of this choice. There are two reasons for applying a PCA. First, it helps to reduce the multidimensional catch matrix to a smaller number of informative components represented by the first n axes of the PCA transformation. Second, it is informative about the interactions among species across LEs.

When running a PCA, it is necessary to specify the number of axes to be retained. Two possible criteria were implemented (Hartigan, 1975): (i) using a second-order Cattell Scree test looking for the significant marginal increases of explained inertia, and (ii) selecting all axes accumulating 70% of the explained inertia. The resulting dataset is the matrix of new LEs coordinates using the retained axes of the PCA, $D^2 = (D^2_{i,j})_{i=1,\ldots,n;j=1,\ldots,k}$, where k is the number of retained axes $(k \leq p)$ and $D^2_{i,j}$ is the coordinate of the ith LE on the jth axis of the PCA.

Step 3: clustering

Let us denote D^3 the input dataset of the clustering step of the analysis. $D^3 = D^2$ if a PCA was previously used, otherwise $D^3 = D^1$. Three clustering methods were selected and implemented to be applied on D^3: HAC (Hartigan, 1975). K-means (Hartigan and Wong, 1979), and CLARA (Kaufman and Rousseeuw, 1990). All methods lead to a classification of all individual LEs, but they are based on different approaches and algorithms.

The HAC method initially assigns each LE to its own cluster, then the algorithm proceeds iteratively, at each stage joining the two most similar clusters, until there is just a single cluster. The similarity is computed using a distance function (default set to the Euclidean distance) and a clustering method (default set to the Ward criterion). The number of clusters is determined thereafter, once all combinations have been calculated, using a Scree test, which can return successive thresholds for increasing number of clusters.

The two other methods assume that the final number of clusters is known *a priori* and set at K. In contrast to the HAC, these methods select first K LEs, then assign them to their own clusters (kernels), and subsequently assign each LE of the dataset to the closest cluster, according to a similarity criteria (default set to the Euclidean distance).

With the K-means method, each cluster is characterized by its gravity centre (i.e. an average virtual LE) and the method is applied to the whole LE dataset (no sampling). For each new observation, the gravity centre of the cluster and the distances to the next LE are recalculated. The procedure is repeated with increasing values of K, and the most appropriate number of clusters can be selected by detecting the largest marginal losses of inertia between two consecutive numbers of clusters, similar to the HAC Scree test.

The CLARA method was developed to cope with large datasets. Subsets of the data, of user-defined number and size, are sampled and clustered using the PAM (Kaufman and Rousseeuw, 1990) algorithm. With this approach, the centre of the cluster is defined as the medoid, i.e. the most central LE that shows the smallest dissimilarity with other LEs within the cluster. Identifying the medoid requires computing all dissimilarities within the cluster and comparing it with the sum of dissimilarities if any other LE in the cluster had been the medoid. The medoid is then adjusted accordingly, until convergence of the procedure. As in K-means, the procedure is run with increasing values of K, and the most appropriate number of clusters is detected using the local maximum values of the silhouette of the classification. The silhouette provides an average comparison of the distance between an LE and the other LEs from its cluster, and between the same LE and the other LEs of the nearest cluster. Subsequently, all remaining LEs are assigned to their nearest medoid using a user-defined method for calculating distances (default set to the Euclidean distance).

Regardless of the clustering method used, step 3 leads to a new matrix $D^{4,\text{métiers}}_{\text{ClusteringMethod}} = (D^{4,\text{métiers}}_{\text{ClusteringMethod}_{i,j}})_{i=1,\ldots,n;j=1,\ldots,m+1}$ equal to D^0, with an additional column containing the result of the classification as an identifier of the cluster associated with each LE.

To perform a rigorous comparison of the performance and outcomes of the three clustering methods, an objective analysis plan was established. First, because the HAC method initially proved to be fairly computer-demanding and not straightforward to implement in R for the whole dataset, the approach was modified along the principles of the CLARA procedure, i.e. the HAC was performed on a sample of randomly selected LEs representing a given percentage of all initial LEs. The remaining LEs were subsequently allocated to the defined clusters using linear discriminant analysis. As for CLARA, to avoid issues linked to the random sampling of LEs, the procedure was set to be repeated a number of times, and the best classification with regard to maximizing the clustering variance ratio was retained for the final outputs.

The comparison between HAC and CLARA was made based on five samples within each method and with sample size set at 1, 5, 10, or 30% of the whole dataset. As the sampling algorithm is performed internally within each method, the sampled datasets may differ in both methods, but because both methods proceed with multiple samples, they should ideally be robust to the sampling bias. Monitoring this robustness is exactly the purpose of the second test we performed as described below.

The second test follows from the observation that combining national logbook data resulted in a stratified dataset with rows intrinsically grouped by country and season, which affected the row sampling of HAC and CLARA. For this reason, the robustness and the stability of the three methods were tested by applying them to three permutations of the whole dataset, where the LE lines were randomly shuffled before the analyses.

Downloaded from http://icesjms.oxfordjournals.org/ at DIFRES on January 23, 2012

Third, the three methods build on finding best performance of given criteria across the consecutive number of clusters, but several thresholds can be observed when increasing the number of clusters. A first threshold is expected to distinguish the most well-defined fisheries (e.g. sandeel, northern prawn) from the bulk of more mixed operations, and a second threshold is expected to return a more accurate description of these mixed fisheries. Therefore, two classifications were investigated for each method, returning both the first and the second threshold encountered with increasing numbers of clusters (an initial threshold could sometimes be observed for small numbers of clusters but was disregarded), corresponding to the second and third Scree test minima for HAC, the first and second largest inertia losses for *K*-means, and the first and second silhouette maxima for CLARA, respectively.

Ultimately, the most appropriate approach was selected based on multiple considerations, i.e. (i) the computational cost and data limitations (on a Linux server), (ii) the clustering variance ratio, (iii) the stability of results across different permutations of the initial dataset (shuffles), (iv) the projections of LEs by cluster on first factorial axes, with clusters expected to be well distinguished by axes and LEs to be well grouped together, (v) some characteristics of outcomes (a) the final number of clusters, (b) the number of LEs by cluster, and (c) the list of characteristic species by cluster, representing both the main species in value and the most representative species caught in the cluster—empirical rules were defined, so characteristic species should appear in the ranked list accumulating 75% of the average value of the cluster and be characterized by a test value (a statistical metric comparing the proportion of that species in the landings of the cluster with the average proportion of the same species in the whole dataset; Lebart *et al.*, 1995) >50% and should appear in at least 30% of LEs in the cluster, and (vi) the meaning and the relevance of clusters. The last criterion is based mainly on expert knowledge. Experts played an important role in judging the validity of some métier definitions, e.g. if a cluster gathered two important species known for not being caught together, if small but very specific fisheries could be identified, or if two clusters seemed redundant.

At the end of this step, a single clustering approach was selected, with $D^{4,\mathrm{level\,7-métiers}}$ denoting the resulting outcome. Each métier at level 7 was named using the combination of gear used and the list of characteristic species.

Step 4: linkage with DCF level 5

The DCF (EC, 2008) currently uses the métier definition at level 5, based on species assemblage. For the OTB gear in the North Sea region, the DCF recognizes in principle the existence of the mixed-assemblage crustaceans–DEF (MCD) and cephalopods–DEF (MCF). However, it is also specified that "The target assemblage that comes up at the first position should be considered as the target assemblage to be reported in the matrix" (EC, 2008), which in practice would imply that only one type is defined, not a mix. The notion of first position assemblage can be interpreted in different ways, and in the absence of clear quantitative guidelines, different national fisheries laboratories may have implemented different sets of simple empirical rules (referred to as ordination rules), e.g. a first species method, identifying the single most abundant species in the LE and allocating the LE to the corresponding level 5 assemblage of this species, and a first group method, summing the catch of the LE within the respective species types and allocating the LE to the most important level 5 assemblage of the LE.

Assuming that level 7 is a more accurate representation of fishing activities, we used this as a baseline to assess the validity of these ordination rules, by calculating the extent of correspondence between both. The two ordination rules were first applied directly to each LE of the initial dataset D^0 on its own catch profile, then the list of characteristic species, as defined above, of each LE at its level 7 cluster was considered using $D^{4,\mathrm{level\,7-métiers}}$ and related to their corresponding assemblage, allowing the appearance of mixed assemblages at level 5. Finally, the contingency of both classifications was measured in terms of the number of LEs in each level 5. Obviously, for the clusters dominated by a single species, both classifications would be in broad agreement, but this may be less clear for clusters defined by more than one main species.

Step 5: prediction of the métier of a new LE

Métiers are not only used to characterize a fishery from historical data, but also to improve sampling monitoring and fishing mortality estimates. Some quantitative rules are therefore necessary to identify the most representative métier of any new LE (e.g. during the current year, before all annual data have been collected and new analyses run). Discriminant analysis is an efficient technique for deriving such quantitative allocation rules. It allows the classification of a set of observations into predefined clusters, fitting a multichoice model using one function for each cluster. This model is expected to predict the cluster of an LE based on a set of predictors, here the landings per species: given a new LE, all K discriminant functions (K being the number of clusters) are evaluated and the observation assigned to class i if the ith discriminant function has the highest value. This analysis is performed using the matrix $D^{4,\mathrm{level\,7-métiers}}$ corresponding to a set of observations for which the clusters are known.

Results

The detailed comparison of methods was performed on the largest and most recent dataset, i.e. 2008. Subsequently, the selected method was applied to the 2007 dataset on which the discriminant analysis was also fitted. Finally, the results for 2008 were compared with the predicted métiers using such allocation rules.

Step 1

The HAC method retained 31 main species of the 296 of the initial dataset (Figure 2). The perTotal method was more selective: given the strong dominance of few species in the total value of the dataset, the incremental slope was very low, and 26 species represented 95% of the total value. The perLogevent method returned the largest range of species, with 60 species representing 100% of the value of at least one LE. Combining these three sets of species and harmonizing a few national codes led to a reduced dataset of 58 main species, 99.1% of the total value (Table 1).

Step 2

The PCA showed some species clustering over the first axes (Figure 3), but little information was conveyed by each axis, and it was necessary to retain 36 axes to reach the threshold of 70% inertia. The Scree test was considered to retain too little information, because it retained just nine axes cumulating 24.4% of the inertia. Therefore, subsequent analyses were performed on the

Downloaded from http://icesjms.oxfordjournals.org/ at DIFRES on January 23, 2012

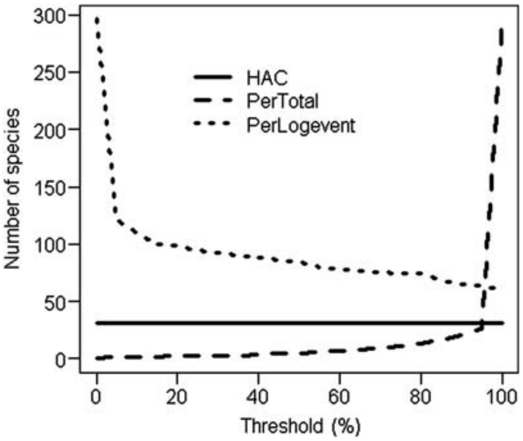

Figure 2. The number of main species selected depending on the method and threshold chosen.

Table 1. Species name, FAO code, and the corresponding DCF level 5 for the main species retained in 2008 (DEF, demersal fish; CRU, crustacean; CEP, cephalopod; SPF, small pelagic fish).

FAO code	Scientific name	English name	DCF level 5
BIB	*Trisopterus luscus*	Bib	DEF
BRB	*Spondyliosoma cantharus*	Black sea bream	DEF
BSS	*Dicentrarchus labrax*	European sea bass	DEF
COD	*Gadus morhua*	Atlantic cod	DEF
CSH	*Crangon crangon*	Common shrimp	CRU
CTC	*Sepia officinalis*	Common cuttlefish	CEP
GUU	*Chelidonichthys lucerna*	Tub gurnard	DEF
HAD	*Melanogrammus aeglefinus*	Haddock	DEF
HER	*Clupea harengus*	Atlantic herring	SPF
HKE	*Merluccius merluccius*	European hake	DEF
LEM	*Microstomus kitt*	Lemon sole	DEF
MAC	*Scomber scombrus*	Atlantic mackerel	SPF
MEG	*Lepidorhombus whiffiagonis*	Megrim	DEF
MON	*Lophius piscatorius*	Monkfish	DEF
MUR	*Mullus surmuletus*	Surmullet	DEF
NEP	*Nephrops norvegicus*	Norway lobster	CRU
NOP	*Trisopterus esmarkii*	Norway pout	DEF
PLE	*Pleuronectes platessa*	European plaice	DEF
POK	*Pollachius virens*	Saithe	DEF
POL	*Pollachius pollachius*	Pollack	DEF
PRA	*Pandalus borealis*	Northern prawn	CRU
RAJ	Rajidae	Rays and skates nei	DEF
SAN	*Ammodytes* spp.	Sandeels nei	DEF
SDV	*Mustelus* spp.	Smooth-hounds nei	DEF
SOL	*Solea solea*	Common sole	DEF
SPR	*Sprattus sprattus*	European sprat	SPF
SQU	Loliginidae, Ommastrephidae	Various squids nei	CEP
SYC	*Scyliorhinus canicula*	Small-spotted catshark	DEF
TUR	*Psetta maxima*	Turbot	DEF
WHG	*Merlangius merlangus*	Whiting	DEF
WIT	*Glyptocephalus cynoglossus*	Witch flounder	DEF

Only 31 of the 58 retained species named in text are displayed.

36 principal components, decreasing by 40% the number of columns used for clustering.

Step 3

The comparative analysis underlined large differences in the performance of the various methods and settings used (Table 2).

The *K*-means method was the fastest to compute, did not return any dataset size limitations, and had greater variance ratios. However, its outcomes seemed largely unreliable and difficult to interpret. The number and characterization of clusters was unstable across the three shuffles, and the method seemed to emphasize minor clusters (e.g. edible crab, *Sebastes* sp.) while pooling together most of the important species. On this basis, the *K*-means method was not considered appropriate and was not analysed further.

After implementation of the sampling procedure, the HAC method computed smoothly, even with a large sample size. At 30%, the method consistently returned eight clusters, with a variance ratio ∼20%, at first threshold classification. However, these were not completely consistent across the three shuffled datasets, with small differences observed in the characteristic species. More problematically, the method did not reach consistency at

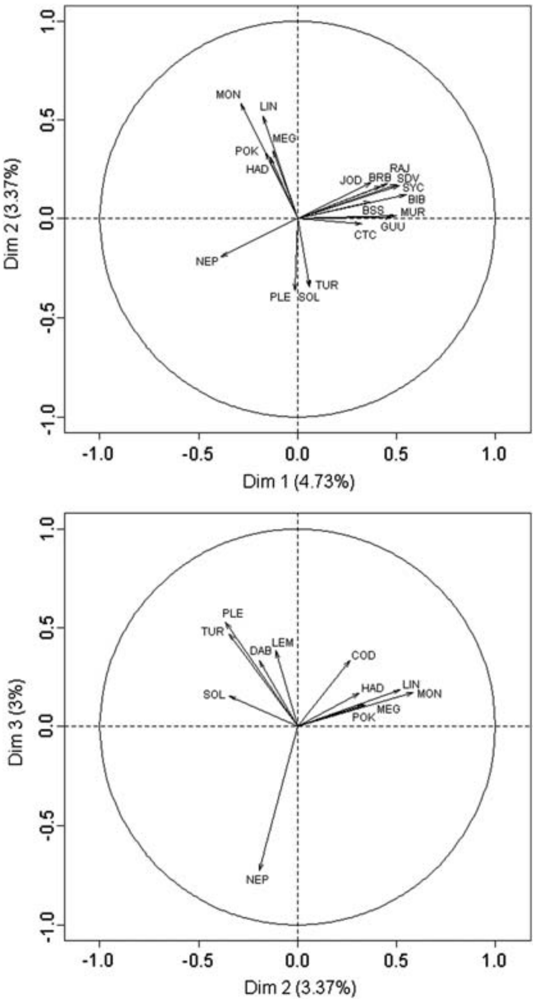

Figure 3. PCA projection of the species on axes 1 and 2 (top) and 2 and 3 (bottom). The percentage numbers on the axes are the percentages of inertia explained by that axis.

Downloaded from http://icesjms.oxfordjournals.org/ at DIFRES on January 23, 2012

Table 2. Performance and outcomes of the three clustering methods for two levels of classification, three shuffled datasets, and four sample sizes (HAC and CLARA methods only), with the average computing time (min), the resulting number of clusters, and the clustering variance ratio (%).

			Sample size			
Criterion	**Method**		**1%**	**5%**	**10%**	**30%**
First threshold	HAC	Computing time	1.2	1.6	2.7	12.5
		Shuffle 1	7 (18.9)	8 (20.6)	8 (21.2)	8 (20.6)
		Shuffle 2	7 (18.8)	8 (21.0)	7 (18.7)	8 (20.8)
		Shuffle 3	6 (15.4)	9 (23.1)	8 (21.4)	8 (20.6)
	CLARA	Computing time	1.22	11.74	31.62	
		Shuffle 1	7 (19.2)	8 (21.5)	6 (16.7)	
		Shuffle 2	7 (19.0)	7 (19.0)	6 (16.6)	
		Shuffle 3	6 (16.6)	6 (16.1)	6 (16.1)	
	K-means	Computing time	4.73			
		Shuffle 1	9 (23.9)			
		Shuffle 2	12 (28.5)			
		Shuffle 3	10 (25.9)			
Second threshold	HAC	Computing time	1.3	1.7	2.7	11.9
		Shuffle 1	10 (25.5)	9 (23.0)	10 (24)	12 (28.3)
		Shuffle 2	9 (22.4)	9 (23.6)	9 (23.6)	11 (27.1)
		Shuffle 3	8 (19.5)	11 (25.6)	12 (28.0)	9 (23.3)
	CLARA	Computing time	1.79	35.35	194	
		Shuffle 1	7 (18.6)	14 (32.5)	13 (30.2)	
		Shuffle 2	15 (33.7)	14 (32.5)	14 (32.6)	
		Shuffle 3	10 (24.4)	13 (30.2)	16 (36.1)	
	K-means	Computing time	4.8			
		Shuffle 1	11 (27.2)			
		Shuffle 2	14 (32.3)			
		Shuffle 3	12 (29.3)			

the second threshold, returning 9–12 clusters for the different shuffles.

The CLARA method turned out to be slowest to compute even with a smaller sample size and could not run with a 30% sample size. However, six consistent clusters were obtained across shuffles at a 10% sample size at the first threshold. At the second threshold, different numbers of clusters were also returned, explaining ~30% of the inertia. These clusters were meaningful and broadly consistent in terms of characteristic species, the additional clusters obtained being characterized by fisheries with few LEs such as Norway pout or herring (usually a little caught with OTB), and which could therefore have been missed by the sampling.

A closer comparison of the HAC and CLARA results obtained with the second shuffle indicated some degree of consistency (Table 3, Figure 4). At the first threshold, six meaningful clusters were identified in common by both methods, three largely dominated by a single species (sandeel, *Nephrops*, sole) and three of mixed nature (gadoid–monkfish, squid–mullet–cuttlefish–whiting, and plaice–lemon sole). In addition, HAC, unlike CLARA, identified two additional single-species fisheries, common shrimp and northern prawn. Both classifications appeared relevant, although the additional HAC clusters were preferred. At the second threshold, some large clusters were cut into smaller ones, providing a finer description of the mixed groups. There was again a high degree of similarity between outcomes of both methods, but whereas most clusters appeared meaningful, HAC also returned some inconsistent results, pooling Norway pout and herring as characteristic of the same cluster when these two species are not caught routinely together.

All considered, it did not appear obvious that one method performed significantly better than the other. Results were robust at the first threshold, but neither of the two methods proved infallible and fully robust to sampling size and row-shuffling when investigating more precise definitions at the second threshold. At that threshold, however, the results provided by HAC seemed slightly less consistent; the variance ratio was lower and the clusters more overlapping on the first factorial axes (Figure 4). Therefore, we chose to pursue the analyses with the results obtained with CLARA, with a 10% sample size and at the second threshold, as the baseline classification for illustration (Figure 5).

Step 4

The results above were used as the basis for evaluating the accuracy of simple ordination rules allowing the allocation of an LE to DCF level 5 (Table 4).

The simple métiers (DEF, CRU) were generally well captured by the rules (~95% of overlap), but these simple approaches obviously could not capture clusters with mixed target type. For example, our clusters 3 and 7 were both dominated by a different combination of cephalopods and DEF (MCF), with species being very characteristic of the cluster (high test value) and a large degree of co-occurrence (observed in >80% of the LE of the cluster). However, LEs in these clusters were split into two distinct level 5 categories with the ordination rules depending on the dominating species; 43% of them being considered as DEF and 52% as CEP with the first species method, and 59 and 38%, respectively, with the first group method.

Both ordination methods were generally in broad agreement with each other. However, the first group method tended to favour the DEF métier, because of the large number of species belonging to this group. As a consequence, some LEs would be

Downloaded from http://icesjms.oxfordjournals.org/ at DIFRES on January 23, 2012

Table 3. Identification of the métiers at level 7 and the corresponding level 5 and comparison of the corresponding clusters obtained with HAC and CLARA at both first and second thresholds, indicating for each cluster which are the characteristic species; for each of these, the average proportion of the species in the total value of each LE in the cluster, as well as the proportion of LE in the cluster where the species is recorded, are indicated in brackets (expressed as a percentage).

			First threshold		Second threshold	
Cluster number	**Level 7**	**Level 5**	**HAC**	**CLARA**	**HAC**	**CLARA**
1	OTB *Nephrops*	CRU	NEP (85, 100)	NEP (79, 91)	NEP (84, 100)	NEP (94, 100)
2	OTB *Nephrops* – monkfish	MCD				NEP (61, 94) MON (12, 77)
14	OTB Northern prawn	CRU	PRA (50, 59)		PRA (88, 100)	PRA (89, 100)
8	OTB shrimp	CRU	CSH (100, 100)		CSH (100, 100)	CSH (100, 100)
5	OTB cod – haddock	DEF	COD (21, 79) MON (21, 66) POK (16, 67) HAD (12, 70)	MON (20, 69) COD (19, 77) POK (13, 65) HAD (11, 69)	COD (25, 82) POK (19, 69) HAD (14, 67) MON (13, 57)	COD (39, 88) HAD (24, 71)
9	OTB saithe – hake	DEF				POK (39, 90) HKE (8, 63)
6	OTB monkfish – megrim	DEF			MON (53, 98) MEG (13, 73)	MON (48, 98) MEG (9, 60)
3	OTB squid – whiting	MCF	SQU (22, 69) WHG (13, 66) MUR (9, 62) BSS (8, 52) CTC (7, 47) BIB (2, 60)	SQU (24, 74) WHG (15, 69) MUR (10, 67) BSS (9, 55) CTC (7, 50) BIB (2, 64) GUU (2, 60)	SQU (35, 71) WHG (22, 73) MAC (8, 52)	SQU (39, 83) WHG (23, 79) MUR (6, 60) MAC (6, 55)
13	OTB sea bass	DEF			BSS (15, 65) MUR (14, 74) CTC (13, 65) BRB (5, 57) BIB (3, 70) GUU (2, 68)	BSS (24, 78) BRB (10, 75) RAJ (9, 75) SDV (5, 67) SYC (3, 78)
7	OTB cuttlefish – surmullet	MCF				CTC (31, 86) MUR (27, 90) GUU (4, 73)
4	OTB plaice – lemon sole	DEF	PLE (49, 95) LEM (13, 51) TUR (10, 55)	PLE (53, 96) LEM (14, 52) TUR (9, 51)	PLE (49, 95) LEM (13, 51) TUR (10, 55)	PLE (54, 97) LEM (14, 52) TUR (9, 52)
11	OTB sole	DEF	SOL (75, 99)	SOL (66, 94)	SOL (75, 99)	SOL (68, 95)
10	OTB sandeel	DEF	SAN (99, 100)	SAN (99, 100)	SAN (99, 100)	SAN (99, 100)
12	OTB Norway pout	DEF			NOP (29, 32) HER (17, 58)	NOP (95, 100)

Results were obtained with the shuffle 2 dataset, using a 10% sample size for CLARA and a 30% sample size for HAC. The horizontal lines indicate how the clusters overlap with each other across the four classifications. Cluster numbers are the arbitrary ranking of classification outputs and are only retained for the description of Figures 4 and 5.

Downloaded from http://icesjms.oxfordjournals.org/ at DIFRES on January 23, 2012

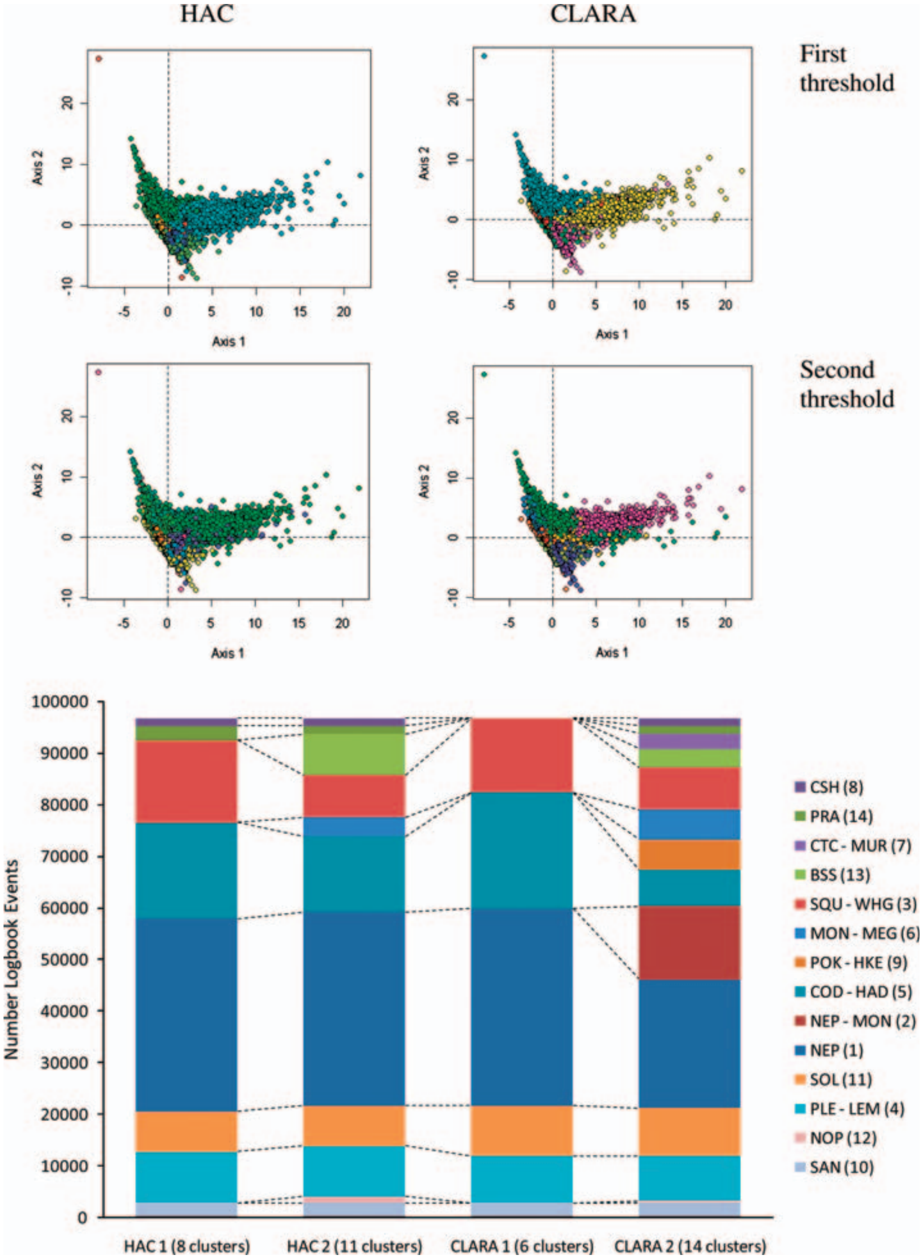

Figure 4. (Top) Projection of LEs in the subspace of the two first axes for the clustering methods HAC and CLARA and at two threshold levels. Each identified cluster is coloured. (Bottom) The number of LEs by level 7 obtained with HAC and CLARA at both first and second thresholds. The horizontal lines indicate how the clusters overlap with each other across the four classifications. The results were obtained with the shuffle 2 dataset, using a 10% sample size for CLARA and a 30% sample size for HAC. In parenthesis are the cluster numbers in the order they appear in the output (see also Table 3, Figure 5).

classified as DEF, despite their primary species in value being from a type other than DEF.

Step 5

The same series of steps was applied to the LE of 2007, on a single shuffle. In all, 12 clusters were identified, which were mostly similar to those of 2008 (Table 5), except the Norway pout cluster, which did not appear because the fishery for that species was closed in 2007. The discriminant analysis was conditioned on the 2007 results and applied to the 2008 LEs. The correspondence between the predicted and estimated level 7 categorizations for 2008 was generally high, with systematically >80% of LEs being correctly predicted by the allocation rules.

Discussion

In summary, we suggest a full sequence to analyse historical catch data and assign them to métier category, synthesized below for clearer understanding.

Downloaded from http://icesjms.oxfordjournals.org/ at DIFRES on January 23, 2012

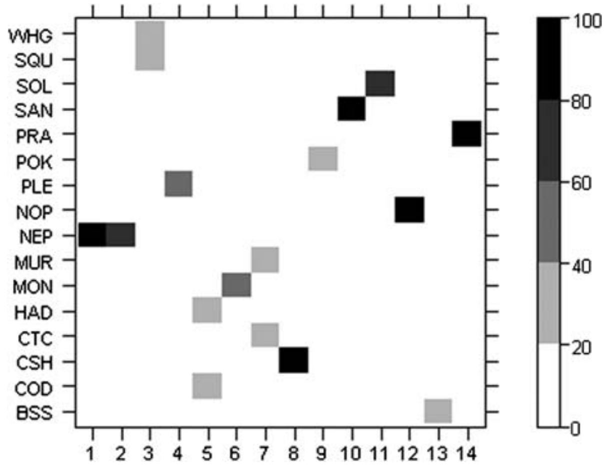

Figure 5. Percentage of the cash value per species per cluster. Only species with values >20% within at least one cluster are displayed. The results are obtained with the shuffle 2 dataset, using a 10% sampling size and the CLARA method. Cluster numbers on the horizontal axis correspond to the numbers in Table 3.

Table 4. Percentage of correspondence between métiers defined by the CLARA algorithm aggregated at level 5 (in columns) with the métiers defined by the "First Species" and "First Group" ordination methods (in rows).

	CLARA			
Criterion	**CRU**	**MCD**	**MCF**	**DEF**
First species				
CEP	–	–	52	1
CRU	100	93	–	5
DEF	–	5	43	93
SPF	–	2	5	1
First group				
CEP	–	–	38	–
CRU	100	82	–	2
DEF	–	16	59	97
SPF	–	2	3	1

(i) Select the main species to be retained (step 1) and run a PCA (step 2).

(ii) Perform the clustering (step 3). Depending on the precision scale required and the size of the dataset, various classifications should be explored, as done here, to identify the most stable and meaningful patterns. The whole analysis (steps 1–3) can either be performed yearly if the time-trends in métier composition are to be monitored, or by pooling, for instance, the most recent 3 years of data if a broader overall picture is required.

(iii) Define the most characteristic species of each level 7 métier obtained, using for example a combination of importance in value, specificity, and broad representation across LEs. This step includes an element of subjectivity, so the resulting characterization of métiers should be inspected carefully.

(iv) Characterize each métier to its level 5 simply by relating each of its characteristic species to its assemblage type (step 4). The various level 7 métiers of similar assemblages would be aggregated within the corresponding level 5 métiers, and this would form the basis of the planning of the sampling programme for the subsequent year.

(v) Perform a discriminant analysis on level 7 métiers to obtain automatic allocation rules (step 5). Using these rules, any new LE entered into the database will be allocated immediately to a métier (at levels 7 and 5). Using this approach, the sampling programme over the year could be monitored in real time.

(vi) At the end of the year, perform a complete new statistical analysis of all new logbooks events collected to detect any changes in métier distribution that may have arisen, and if necessary update the sampling programme for the forthcoming year.

The definition of clusters of fishing activities is the primary step without which no further analysis can be conducted and no progress in fleet-based or mixed-fisheries management achieved. The many previous studies addressing métier definition (cited in the Introduction) have used different methods on different

Table 5. Correspondence between the 2008 clusters defined by cluster analysis (in rows) with the 2008 clusters predicted by discriminant analysis applied on the 2007 clusters (in columns), expressed as a percentage of the number of LEs by row, and the total number of LEs in each cluster.

Level 7	NEP	PRA	CSH	COD – HAD	MON – POK	SQU – WHG	BSS – BRB	MUR – CTC	PLE – LEM	SOL	SAN	SPR – HER	Non alloc.	Total
NEP	99.8									0.1				24 865
NEP – MON	82.3			3.9	9.5	0.3	0.3		0.6	0.4		2.4	0.2	14 252
PRA		100												1 527
CSH			100											1 513
COD – HAD	0.2			89.5	5.0	1.0	3.5		0.5	0.2			0.1	7 040
POK – HKE	4.1	0.3		11.6	83.5				0.3					5 757
MON – MEG	2.5			4.2	92.9	0.2			0.2					5 919
SQU – WHG				2.3		88.2	3.2	3.4		2.5		0.2		8 070
BSS – BRB				1.4	0.2	3.9	90.4	2.7	0.3	1.0				3 680
CTC – MUR						11.5	0.5	86.6	0.3	1.0				2 903
PLE – LEM	1.6			2.7		1.7	0.1	0.2	92.8	0.9				8 702
SOL	1.6			1.6		0.4	0.3	0.2	0.8	95.0				9 289
SAN											99.5	0.5		2 877
NOP				8.2	6.3	1.9						83.5		364
Total	37 248	1 548	1 513	8 418	12 037	7 922	3 930	2 935	8 349	9 264	2 865	685	44	96 758

Downloaded from http://icesjms.oxfordjournals.org/ at DIFRES on January 23, 2012

fisheries. However, none of those studies addressed quantitatively the importance of the choice of the method itself, nor did they implement prediction methods to allocate new LEs of future years into predefined métiers. Although some studies were conducted within the framework of the DCF, they typically failed to link explicitly their outcomes with the actual needs and levels defined by the framework. In addition, the previous analyses did not address the issues of national differences at a regional scale. In this regard, we consider that the current analysis is a significant step forwards, in that it has succeeded in (i) addressing the persistent issue of métier identification in the most objective way, and (ii) providing operational and generic open-source tools directly applicable to any regional fisheries where logbook data are routinely available, thus contributing to the practical implementation of the DCF in Europe. In addition, key results were obtained at a regional North Sea level by compiling, for the first time, a comprehensive international dataset, and the results obtained were clearly different from those obtained applying the same workflow to national data (results not shown), although some clusters here reflect national rather than international fisheries.

With regard to the importance of methodological choice, it has become clear that different methods returned different classifications. Based on the accuracy and robustness of clustering, and basic knowledge of the fisheries, there was no entirely clear basis for choosing between the CLARA and HAC methods. The results showed many similarities. In both cases, the clusters obtained were mostly meaningful and well balanced in size, especially at the first threshold classification, which returned a broad picture distinguishing major single-species fisheries from the main mixed fisheries. The second threshold provided finer description of the mixed groups, but the results were less robust and some inconsistencies appeared. Both methods required subsets of the dataset, and sample size appeared to be a determinant factor for improving the robustness of results. However, computation time increased with sampling size. In particular, CLARA computed fairly slowly with a large sample size, because the method runs through a loop of increasing kernel number, which slows the process. In comparison, the *K*-means method performed poorly. Although efficient in computation, the results obtained were largely unstable and often meaningless.

Clustering results were also compared with and without running a PCA in step 2 (results not shown). Overall, the clusters obtained were fairly similar in average catch composition, but there were differences in the allocation of some of the LE located at the edges between several clusters in terms of landings profile. Not using a PCA resulted in comparably smaller and more accurately defined clusters for some of the main species, but in a bigger pool of less-defined groups for other species. Running a PCA decreased significantly the computing time, by reducing the number of explanatory variables, and results were potentially more independent of the choice of the number of species retained in step 1. It appears therefore that the PCA is preferable for describing global trends, but that the exact allocation of some least characteristic individual LEs could be challenged.

It is necessary to characterize each cluster in terms of its main species. This step is in itself not as trivial as it sounds, because it requires defining some criteria for selecting the characteristic species. This represented, therefore, the only part of the whole workflow where subjective choices had to be made. The average value has often been retained as the only criterion, but we propose more detailed empirical rules, based on exploratory analyses of the outcomes and common sense. The list of characteristic species of a cluster included species not only important in average value, but also those which were significantly more abundant in the cluster than in the overall dataset, as well as those that were much represented across LEs. These criteria and the corresponding thresholds are, however, not universal, and any application of this method needs to pay particular attention to the definition of relevant criteria.

The results support, but also question, some aspects of the DCF of the EU. A number of pan-European workshops (ICES, 2003; EC, 2005, 2006) led to this hierarchical métier definition, so the DCF design emerged from intensive (still ongoing) scientific discussions, where compromises had to be found to reach a one-size-fits-all model covering all EU fisheries. In contrast to DCF levels 1–4, level 5 on target assemblage remains controversial, owing to the difficulty in defining and quantifying it accurately. The concept of target is by its essence vague when fishers catch a number of species in varying proportions and no information on fisher intention before the fishing operation is collected routinely. With hindsight, the choice of aggregating species by type does not necessarily simplify the establishment of sampling schemes. Fishing does not operate towards a given type, but rather towards a given valuable mix of co-occurring species, possibly of different types. The approach leads to the pooling of widely different species caught by distinct métiers together within the same level 5 assemblage, e.g. common shrimp and *Nephrops*, and only level 6 categorization can help to distinguish them, provided mesh-size information is reported accurately in logbooks. Ultimately, level 7 may actually be preferred to level 5, by, for example, selecting the first threshold criterion that may not necessarily return a larger number of categories than the current level 5, but which returns categories that are more appropriate.

Conclusions

A robust and operational workflow for planning and monitoring DCF sampling programmes at level 5 has been suggested here, rather than empirical ordination rules, and métiers have been considered at a regional scale by analysing catch declarations from different Member States simultaneously. We believe that the open-source programs developed for conducting these analyses are extremely powerful in terms of flexibility, general applicability, and computing speed. Despite a fair level of complexity embedded in the computing functions, these programs remain straightforward enough to operate (see scripts and examples of use at http://code.google.com/p/vmstools/wiki/MetiersLogbook) and can be used with any logbook database standardized in the relevant input format. They can also be combined with other similar tools for the analysis of logbook and VMS data (Hintzen *et al*., 2011). These tools are directly operational and useful for standardizing métier definitions across Europe, so in our opinion can contribute to improved fleet-based and ecosystem-based fisheries management.

Acknowledgements

The work was funded through the Project “Development of tools for logbook and VMS analysis”, studies for carrying out the Common Fisheries Policy No MARE/2008/10 Lot 2. This support is gratefully acknowledged. We also thank our project partners and colleagues from Ifremer, DTU-Aqua, IMARES, Cefas, and Marine Scotland Science for providing the data

Downloaded from http://icesjms.oxfordjournals.org/ at DIFRES on January 23, 2012

necessary for carrying out the analyses. Finally, we thank Stuart Reeves for his final reading and language improvement.

References

Bastardie, F., Nielsen, J. R., Ulrich, C., Egekvist, J., and Degel, H. 2010. Detailed mapping of fishing effort and landings by coupling fishing logbooks with satellite-recorded vessel geo-location. Fisheries Research, 106: 41–53.

Beare, D. J. B., Hintzen, N. T., Piet, G. J., Nielsen, J. R., Manco, F., South, A., Bastardie, F., *et al.* 2011. Development of tools for logbook and VMS data analysis. Studies for carrying out the common fisheries policy. EU Open Call for Tenders, MARE/2008/10 Lot 2.

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: 119–136.

Biseau, A., and Gondeaux, E. 1988. Apport des méthodes d'ordination en typologie des flottilles. ICES Journal of Marine Science, 44: 286–296.

Campos, A., Fonseca, P., Fonseca, T., and Parente, J. 2007. Definition of fleet components in the Portuguese bottom trawl fishery. Fisheries Research, 83: 185–191.

Castro, J., Marín, M., Pierce, G. J., and Punzón, A. 2011. Identification of métiers of the Spanish set-longline fleet operating in non-Spanish European waters. Fisheries Research, 107: 100–111.

Castro, J., Punzón, A., Pierce, G. J., Marín, M., and Abad, E. 2010. Identification of métiers of the Northern Spanish coastal bottom pair trawl fleet by using the partitioning method CLARA. Fisheries Research, 102: 184–190.

Cattell, R. B. 1966. The Scree test for the number of factors. Multivariate Behavioral Research, 1: 245–276.

Duarte, R., Azevedo, M., and Afonso-Dias, M. 2009. Segmentation and fishery characteristics of the mixed-species multi-gear Portuguese fleet. ICES Journal of Marine Science, 66: 594–606.

EC. 2005. Commission Staff Working Paper: Report of the *Ad Hoc* Meeting of Independent Experts on Fleet-Fishery based Sampling, 23–27 May 2005, Nantes, France. 34 pp.

EC. 2006. Commission Staff Working Paper: Report of the *Ad Hoc* Meeting of Independent Experts on Fleet-Fishery based Sampling, 12–16 June 2006, Nantes, France. 98 pp.

EC. 2008. Commission Decision (2008/949/EC) of 6 November 2008 adopting a multiannual Community programme pursuant to Council Regulation (EC) No 199/2008 establishing a Community framework for the collection, management and use of data in the fisheries sector and support for scientific advice regarding the Common Fisheries Policy. Official Journal of the European Union, L 346/37.

EC. 2011. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Reform of the Common Fishery Policy. EU COM(2011) 417. 12 pp.

Hartigan, J. A. 1975. Clustering Algorithms. John Wiley, New York.

Hartigan, J. A., and Wong, M. A. 1979. A K-means clustering algorithm. Applied Statistics, 28: 100–108.

He, X., Bigelow, K. A., and Boggs, C. H. 1997. Cluster analysis of longline sets and fishing strategies within the Hawaii-based fishery. Fisheries Research, 31: 147–158.

Hintzen, N. T., Bastardie, F., Beare, D. J. B., Piet, G., Ulrich, C., Deporte, N., Egekvist, J., *et al.* 2011. VMStools: open-source software for the processing, analysis and visualization of fisheries logbook and VMS data. Fisheries Research, doi:10.1016/j.fishres.2011.11.007.

Holley, J-F., and Marchal, P. 2004. Fishing strategy development under changing conditions: examples from the French offshore fleet fishing in the North Atlantic. ICES Journal of Marine Science, 61: 1410–1431.

ICES. 2003. Report of the Study Group on the Development of Fishery-based Forecasts (SGDFF), 18–21 February 2003, Boulogne, France. ICES Document CM 2003/ACFM: 08. 37 pp.

ICES. 2010. Report of the Working Group on Mixed Fisheries Advice for the North Sea (WGMIXFISH), 31 August–3 September 2010, Copenhagen. ICES Document CM 2010/ACOM: 35. 93 pp.

Jiménez, M. P., Sobrino, I., and Ramos, F. 2004. Objective methods for defining mixed-species trawl fisheries in Spanish waters of the Gulf of Cádiz. Fisheries Research, 67: 195–206.

Katsanevakis, S., Maravelias, C. D., and Kell, L. T. 2010. Landings profiles and potential métiers in Greek set longliners. ICES Journal of Marine Science, 67: 646–656.

Kaufman, L., and Rousseeuw, P. J. 1990. Finding Groups in Data: an Introduction to Cluster Analysis. John Wiley, New York.

Laurec, A., Biseau, A., and Charuau, A. 1991. Modelling technical interactions. ICES Marine Science Symposia, 193: 225–236.

Lebart, L., Morineau, A., and Piron, M. 1995. Statistique Exploratoire Multidimensionnelle. Dunod, Paris.

Lewy, P., and Vinther, M. 1994. Identification of Danish North Sea trawl fisheries. ICES Journal of Marine Science, 51: 263–272.

Marchal, P. (Ed.). 2006. Technological developments and tactical adaptations of important EU fleets. Final Report, TECTAC QLK5-2001-01291. 651 pp. (main report), 369 pp. (annexes).

Marchal, P. 2008. A comparative analysis of métiers and catch profiles for some French demersal and pelagic fleets. ICES Journal of Marine Science, 65: 674–686.

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.

Punzón, A., Hernández, C., Abad, E., Castro, J., Pérez, N., and Trujillo, V. 2010. Spanish otter trawl fisheries in the Cantabrian Sea. ICES Journal of Marine Science, 67: 1604–1616.

R Development Core Team. 2010. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org.

Rogers, J. B., and Pikitch, E. K. 1992. Numerical definition of groundfish assemblages caught off the coasts of Oregon and Washington using commercial fishing strategies. Canadian Journal of Fisheries and Aquatic Sciences, 49: 2648–2656.

Silva, L., Gil, J., and Sobrino, I. 2002. Definition of fleet components in the Spanish artisanal fishery of the Gulf of Cádiz (SW Spain ICES division IXa). Fisheries Research, 59: 117–128.

Tzanatos, E., Dimitriou, E., Katselis, G., Georgiadis, M., and Koutsikopoulos, C. 2005. Composition, temporal dynamics and regional characteristics of small-scale fisheries in Greece. Fisheries Research, 73: 147–158.

Tzanatos, E., Somarakis, S., Tserpes, G., and Koutsikopoulos, C. 2006. Identifying and classifying small-scale fisheries métiers in the Mediterranean: a case study in the Patraikos Gulf, Greece. Fisheries Research, 81: 158–168.

Ulrich, C., and Andersen, B. S. 2004. Dynamics of fisheries, and the flexibility of vessel activity in Denmark between 1989 and 2001. ICES Journal of Marine Science, 61: 308–322.

Ulrich, C., Gascuel, D., Dunn, M., Le Gallic, B., and Dintheer, C. 2001. Estimation of technical interactions due to the competition for resource in a mixed-species fishery, and the typology of fleets and métiers in the English Channel. Aquatic Living Resources, 14: 267–281.

Ulrich, C., Wilson, D. C., Nielsen, J. R., and Reeves, S. A. 2009. Potentials and challenges in fleet- and métier-based approaches for fisheries management in the CFP. ICES Document CM 2009/R: 06.

Ward, J. H. 1963. Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, 58: 236–244.

Handling editor: Sarah Kraak

Downloaded from http://icesjms.oxfordjournals.org/ at DIFRES on January 23, 2012