A pattern recognition approach to identify biological clusters acquired by acoustic multi-beam in Kongsfjorden

Giovanni Giacalone, Marco Barra, Angelo Bonanno, Gualtiero Basilone, Ignazio Fontana, Monica Calabrò, Simona Genovese, Rosalia Ferreri, Giuseppa Buscaino, Salvatore Mazzola, Riko Noormets, Christopher Nuth, Giosuè Lo Bosco, Riccardo Rizzo and Salvatore Aronica

ARTICLE INFO

Keywords:
Fish school
multi-beam
k-means
3D pattern
cluster

ABSTRACT

The Svalbard archipelago is one of the most intensively studied marine regions in the Arctic area; here the composition and distribution of marine assemblages are changing under the effect of global change, and marine communities are monitored in order to understand the long-term effects on marine biodiversity. In the present work, acoustic data collected in the Kongsfjorden (Svalbard) using multi-beam technology was analyzed to develop a methodology for identifying and classifying 3D acoustic patterns related to fish aggregations (schools). For each identified aggregations, different features were taken into account thus allowing to develop a multi-variate classification procedure considering morphological, energetic and bathymetric characteristics associated to such aggregations. The results obtained from clustering suggest that from a mathematical point of view three distinct groups could be identified. The proposed approach, that allows to discriminate the acoustic patterns identified in the water column, seems promising for improving the monitoring programs of the marine resources, also in view of the ongoing climate changes.

1. Introduction

Acoustic methods, employing single-beam echo-sounder technology in the last decades represented the most effective tool in monitoring the abundance and spatial distribution of marine organisms inhabiting the water column [57]. Recently, the multi-beam technology, historically adopted in morpho-bathymetric surveys, was adapted to identify targets in the water column, thus allowing fishery scientists to obtain a more detailed view of fish aggregations [35]. In this framework, a key aspect to promptly develop specific sustainable management strategies is the design of efficient monitoring programs able to provide early warnings highlighting changes in marine community. Monitoring programs are even more important also under the climate-change scenario. In particular, Arctic areas are experiencing the most evident climate-changes related modifications; for example, in some arctic regions, such as Greenland, in 2019 the highest ice mass loss was observed, reaching in July 2019 about 5 million tons per minute [30], [31], [56]. The marine ecosystem strongly depends on sea ice, which serves as a platform for marine life and plays a crucial role in controlling the physical and chemical processes that affect biological activity [62], [42]. The invasion of alien species in the areas surrounding the Arctic is also an expanding phenomenon negatively impacting on endemic species through complex trophic interactions. The case of Atlantic cod (Gadus morhua) competing with the Arctic cod (Polar cod, Boreogadus saida) [23] is one such example. In the present work, multi-beam data collected in the framework of the CalvingSEIS project (Glacier dynamic ice loss quantified through seismic eyes) were analyzed. Under the umbrella of CalvingSEIS project, a group of researchers belonging to different European institutions (University of Oslo, Norway; National Research Council, Italy; University in Kiel, Germany; GAMMA Remote Sensing Ag, Switzerland; NORSAR, Norway) evaluated the influence of glaciers melting dynamics on terrestrial and marine Arctic ecosystems. For the first time, a multi-beam scientific echo-sounder was used for a bathymet-
A pattern recognition approach

ric survey in Kongsfjorden (Norway), allowing to consistently evaluate the morphological and energetic features of fish aggregations inhabiting the fjord. The collected acoustic data was analyzed to develop a methodology for identifying and classifying 3D acoustic patterns related to fish aggregations (schools).

2. Materials and Methods

2.1. Study area

The Svalbard archipelago is located between the parallels 74° N and 81° N, is heavily influenced by Atlantic circulation [33], and extends over an area of 63,000 km². Here the ice cover 57% of land with a mix of land and sea glaciers and ice caps [46]. The Kongsfjorden is an area of highly scientific interest due to its biodiversity [33]. The innermost part of the fjord is characterized by one of the largest glaciers of Svalbard, called Kronebreen. Its front ends directly on part of the shelf, but can assume a semi-pelagic lifestyle in specific habitat areas, with gravel, rocks or boulders, which provide protection from predators [10]. Adults are usually found in deeper and colder waters. During the day, they form schools that move about 30-80 m from the bottom, while during the night they disperse in the water column to feed. This is an eurythermal and euryhaline species tolerating wide temperatures and salinity ranges [18]. Finally, the haddock is a saltwater demersal fish that is generally found between 10 - 200 m [45]. Adults are most commonly found between 80 and 200 m depth; females predominate in shallow water and males in offshore areas. The temperature range in which they can be found goes from 4 to 10 °C. Among these three species, typically arctic and sub-arctic, the one characterized by a wider distribution is the Polar cod (Boreogadus saida) [33]. The Atlantic cod prevails in the shallow waters [7]. Hence, Atlantic species have gradually established themselves in Arctic regions with a significant impact on ecosystem structure and ecological functioning [54]. Although there appears to be an overlap in the distribution between Atlantic cod and haddock, the former, tolerating lower temperatures, typically expands further north than haddock [16], [4]. There seems to be no trace of the haddock both in the innermost part of the fjord [24] and in the proximity of the Kronebreen glacier [8].

2.2. Multi-beam data processing

Marine species characterized by aggregative behaviour and inhabiting the water column are generally monitored by using acoustic techniques [22], [6], [5]. These techniques allow, with a non-invasive methodology, to investigate large portions of the sea in a relatively short time. Even if the main aim of acoustic survey is the evaluation of the abundance and spatial distribution of target species, acoustic methods allow also to identify the morphological, energetic [25], and bathymetric characteristics of these patterns [1]. (Fig. 2). Multi-beam data were acquired in the period 23th to 26th August 2016 on board the R/V “Viking Explorer”. The data were acquired using a Kongsberg-Maritime EM2040 MBES characterized by an operating frequency of 300 kHz and pole-mounted on one side of the boat. The sampling design focused on the interface with the glacial front inside the fjord and the path that goes from the base of NY Ålesund up to it (Fig. 1). All the acquired data were stored in binary format on dedicated server. Subsequently, the collected files were converted in a format suitable for analysis and ad-hoc algorithms were applied to detect fish schools [25]. For each identified fish school, the morphological and energetic characteristics were extracted (Table 1). In addition, in order to take into account for the position of the schools in the water column, the school depth was also considered (Table 1).

Figure 1: The study area of the Kongsfjorden in the context of Svalbard. Source: http://cruise-handbook.npolar.no/site-images-new/kongsfjorden/Kongsfjorden_2010_3-5-10.jpg

Due to the ice presence, most of the studies carried out in the Kongsfjorden were conducted during the summer [33], [32], [60], when this area is accessible for navigation, although different life processes also take place during polar winter [24]. In the area, besides several species of marine mammals, three species belonging to Gadidae family dominate the water column. One that has been found to have the greatest distribution is the Polar cod (Boreogadus saida) [33]. The other two species are the Atlantic cod (Gadus morhua) and haddock (Melanogrammus aeglefinus) [53], that have been favored by the increase in the inflow of warmer water from the Atlantic to the western coast of Svalbard. The Polar cod is a brackish water bony fish inhabiting the water column [37], [59], [14]; it is a circum-arctic species common in open and ice-covered waters of the seas of the Arctic shelf [49], [51], [15] that performs offshore-inshore migrations associated with spawning and ice movement [43]. The reproduction period is between November and March [32], [51]. The Atlantic cod, widely present in the area, is a ma-
A pattern recognition approach

Figure 2: Diagram showing multi-beam instrument acquisition.

Figure 3: Frequency distribution of the 13 characteristic parameters.

Figure 4: Histograms of variables following logarithmic transformation.

<table>
<thead>
<tr>
<th>CHARACTERISTICS OF THE EXTRACTED PATTERNS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energetic</td>
</tr>
<tr>
<td>Sv Mean</td>
</tr>
<tr>
<td>Std Sv</td>
</tr>
<tr>
<td>Sv Max</td>
</tr>
<tr>
<td>Sv Min</td>
</tr>
<tr>
<td>Bathymetric</td>
</tr>
<tr>
<td>Depth mean</td>
</tr>
<tr>
<td>Depth max</td>
</tr>
<tr>
<td>Depth min</td>
</tr>
</tbody>
</table>

Table 1
Energetic, bathymetric and morphological characteristics extracted.

2.3. Exploratory data analysis
Clustering methods are known to be influenced by the presence of outliers, and acoustic data are often characterized by the presence of skewed distribution and extreme values. Thus, in order to avoid possible biases, the frequency distribution of considered variables and the skewness index [26] were computed. The obtained histograms evidenced in some cases the presence of skewed distribution (see Fig. 3). The combined analysis of the frequency histograms and the asymmetry index allowed to identify the most problematic variables (Table 2; the variables subjected to log transformation are highlighted in bold). To reduce the influence of the extreme values, linked to long-tailed frequency distributions, a log transformation was performed. The frequency distribution of log-transformed variables are shown in Fig. 4. Finally, in order to avoid scale-related effects during the clustering procedure, variables were standardized.

2.4. Clustering method
In this paper the k-means algorithm was used to identify different kind of patterns along the water column. The k-means algorithm requires as input the number of clusters (k) to be identified. Thus, as the number of clusters was not known a priori, a preliminary analysis was performed to evaluate the partitioning of observations among the different clusters, considering a number of clusters between 2 and 10. For each k value, 10 iterations were performed, taking into consideration the relative average values obtained. In order to further evaluate how many clusters to take into consideration to determine an optimal solution, the internal validity indices were applied. Only the indices suitable for k-means (Table 3) were considered, using the NBClust package [11] in R statistical environment program [50]. Once the number of clusters was identified, boxplots were used to evidence for each variable the differences among the identified groups.

3. Results
The exploratory analysis suggested that the optimal number of clusters is 2 or 3 (Fig. 5). In particular, for both k = 2 and k = 3 the observations were almost equally distributed among the clusters. On the contrary, increasing the number of clusters, some of the clusters were consistently characterized by a low number of observations. Validation indices provided an equal probability for 2, 3 and 4 clusters, as shown in Table 3. In order to evaluate the differences among the groups, the box-plots for each of the variables obtained with k between 2 and 4 were examined. From the analysis of the box plots obtained with k = 4 in Fig. 6, it is not possible to highlight a clear separation between the variables in the 4 clusters. For k = 4, a pairwise similarity among the vari-
A pattern recognition approach

### Table 2

Basic statistical parameters used to pre-process the data. The bold rows indicate the parameters that strongly deviates from the symmetry.

<table>
<thead>
<tr>
<th>Surface area</th>
<th>min</th>
<th>Q2.5</th>
<th>Q25</th>
<th>Median</th>
<th>Q75</th>
<th>Q97.5</th>
<th>max</th>
<th>mean</th>
<th>Sk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>1.59</td>
<td>2.53</td>
<td>5.76</td>
<td>9.37</td>
<td>16.22</td>
<td>45.92</td>
<td>83.85</td>
<td>13.28</td>
<td>3.52</td>
</tr>
<tr>
<td>Width</td>
<td>1.91</td>
<td>2.85</td>
<td>6.56</td>
<td>10.50</td>
<td>18.87</td>
<td>41.77</td>
<td>90.26</td>
<td>14.47</td>
<td>3.19</td>
</tr>
<tr>
<td>Depth min</td>
<td>3.43</td>
<td>4.17</td>
<td>11.15</td>
<td>17.15</td>
<td>66.32</td>
<td>105.03</td>
<td>137.20</td>
<td>37.27</td>
<td>0.90</td>
</tr>
<tr>
<td>Depth max</td>
<td>4.91</td>
<td>6.79</td>
<td>15.64</td>
<td>24.28</td>
<td>72.55</td>
<td>112.94</td>
<td>144.29</td>
<td>43.46</td>
<td>0.88</td>
</tr>
<tr>
<td>Height</td>
<td>0.81</td>
<td>1.01</td>
<td>2.68</td>
<td>4.55</td>
<td>7.09</td>
<td>20.08</td>
<td>36.44</td>
<td>6.20</td>
<td>2.47</td>
</tr>
<tr>
<td>Volume</td>
<td>0.60</td>
<td>1.19</td>
<td>8.10</td>
<td>39.83</td>
<td>180.62</td>
<td>2764.08</td>
<td>9227.65</td>
<td>338.57</td>
<td>6.32</td>
</tr>
<tr>
<td>Depth mean</td>
<td>4.07</td>
<td>5.48</td>
<td>13.45</td>
<td>21.60</td>
<td>70.03</td>
<td>108.34</td>
<td>140.67</td>
<td>40.39</td>
<td>0.88</td>
</tr>
<tr>
<td>Roughness</td>
<td>3.02</td>
<td>3.35</td>
<td>4.86</td>
<td>12.12</td>
<td>17.01</td>
<td>28.89</td>
<td>32.28</td>
<td>12.44</td>
<td>0.59</td>
</tr>
<tr>
<td>Sv mean</td>
<td>-65.91</td>
<td>-64.29</td>
<td>-41.10</td>
<td>-33.26</td>
<td>-25.68</td>
<td>-17.06</td>
<td>-3.99</td>
<td>-34.79</td>
<td>-0.52</td>
</tr>
<tr>
<td>Std Sv</td>
<td>-66.02</td>
<td>-64.53</td>
<td>-34.64</td>
<td>-26.66</td>
<td>-18.26</td>
<td>-8.44</td>
<td>13.60</td>
<td>-28.18</td>
<td>-0.54</td>
</tr>
<tr>
<td>Sv max</td>
<td>-59.27</td>
<td>-56.08</td>
<td>-23.94</td>
<td>-15.56</td>
<td>-6.14</td>
<td>7.68</td>
<td>33.44</td>
<td>-16.25</td>
<td>-0.41</td>
</tr>
<tr>
<td>Sv min</td>
<td>-117.12</td>
<td>-114.92</td>
<td>-97.08</td>
<td>-83.92</td>
<td>-75.33</td>
<td>-67.12</td>
<td>-65.56</td>
<td>-86.34</td>
<td>-0.46</td>
</tr>
</tbody>
</table>

### Table 3

Indexes used to evaluate the clusters number [12], with the corresponding condition and the number of cluster obtained from the condition.

<table>
<thead>
<tr>
<th>Index Name</th>
<th>Number of clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;ch&quot; [9]</td>
<td>2</td>
</tr>
<tr>
<td>&quot;gap&quot; [61]</td>
<td>2</td>
</tr>
<tr>
<td>&quot;mcclain&quot; [38]</td>
<td>2</td>
</tr>
<tr>
<td>&quot;silhouette&quot; [55]</td>
<td>2</td>
</tr>
<tr>
<td>&quot;cindex&quot; [34]</td>
<td>3</td>
</tr>
<tr>
<td>&quot;db&quot; [17]</td>
<td>3</td>
</tr>
<tr>
<td>&quot;ratkowsky&quot; [52]</td>
<td>3</td>
</tr>
<tr>
<td>&quot;ball&quot; [2]</td>
<td>3</td>
</tr>
<tr>
<td>&quot;duda&quot; [19]</td>
<td>4</td>
</tr>
<tr>
<td>&quot;beale&quot; [3]</td>
<td>4</td>
</tr>
<tr>
<td>&quot;ptbiserial&quot; [40][39]</td>
<td>4</td>
</tr>
<tr>
<td>&quot;sdindex&quot; [28]</td>
<td>4</td>
</tr>
<tr>
<td>&quot;kl&quot; [36]</td>
<td>7</td>
</tr>
<tr>
<td>&quot;hartigan&quot; [29]</td>
<td>8</td>
</tr>
<tr>
<td>&quot;sdbw&quot; [27]</td>
<td>10</td>
</tr>
</tbody>
</table>

Ous clusters for the morphological characteristics was identified, while clear pattern were not identified for the energetic and bathymetric characteristics. Therefore, taking into account these considerations and the results obtained by the exploratory analysis (Fig. 5), which highlighted the presence of two or three main clusters, the existence of 4 clusters may likely be excluded. In the case of k = 3, a better separation emerged (Fig. 7) even if, it can be noted that in terms of morphological variables, classes 2 and 3 show some similarities, while only a slight separation is evident for the same classes in the energetic characteristics. In the case with k = 2 (Fig. 8), there is a clear separation between the classes in all morphological variables considered and in the minimum of the Sv variable. Then, among the possible solutions obtained, the ones that better represent the dataset are the division into two or three distinct clusters. To choose between these two options, it was decided to check the labeling of the patterns for both clustering cases (k = 2 and k = 3). In case of k-means algorithm application with k = 3, it emerges that observations belonging to clusters 2 and 3 (respectively blue and light-blue in Fig. 7) are grouped into a single cluster (cluster 1 in red color in Fig. 8) when k-means is applied with k = 2, with the exception of only 4 observations out of a total of 77. The remaining cluster 1, obtained with k-means for k = 3, groups the same observations as cluster 2 obtained with k-means for k = 2; this cluster also contains the four previously excluded observations. A more in-depth view of the results allows us to state that three is the number of clusters that better represents the data set from a numerical point of view. To further investigate the relationships among the identified clusters, the spatial pattern was explored (Fig. 9). It is important to highlight that the positions (in terms of longitude and latitude) of the analyzed fish schools were not used as input variable for the clustering algorithm. From visual observation, we can detect a clear spatial separation of the three clusters. In detail, the class identified with the blue color is located in the area.
A pattern recognition approach

Farthest from the glacier and, considering the average depth, such fish schools are positioned at the greatest depths. On the contrary, the fish schools belonging to the class identified with the light blue color are mainly found in the central part of the study area and are characterized by a lower depth than the other clusters (about 30 m deep). Finally, the red class includes all fish schools that are positioned in the area closest to the glacier, also occupying shallow depths. However, the morphological characteristics analyzed show strong similarities between the fish schools labeled with the colors blue and light blue respectively (clusters 2 and 3 in Fig. 7), while the morphological characteristics differ for the red fish schools (cluster 1 in Fig. 7). In the same figure, a slight difference can be seen between the energy characteristics of clusters 2 and 3 while greater similarities are found in some of the energy characteristics belonging to clusters 1 and 3.

4. Discussion

To the best of our knowledge, this work represents the first attempt to use the multi-beam data acquired in the water column in the Svalbard Islands to evaluate the presence of different fish species. For this area there is no previous knowledge relating to the acoustic patterns identified. The use of the morphological, energetic and bathymetric characteristics of the aggregations, along with clustering results, allowed to hypothesize the presence of three distinct clusters of fish schools. Two of them are quite different in terms of morphological and energetic variables. The presence of a third cluster, which differs from the other two only for some characteristics, suggests that this may refer to one of the species associated with one of the other two clusters, and that therefore it is the same species. Specifically, it is possible to hypothesize that in the three-cluster classification (Fig. 7), cluster 2 and cluster 3 correspond to the same species, showing the same morphological features and differing in terms of bathymetric and energetic ones (likely correlated to the specimens size). In Fig. 7 it is possible to see that cluster 3 prefers lower depths than cluster 2. This result, associated with the high variability of the same cluster in the energy characteristics and the period of the survey, makes us hypothesize that the species is present in a juvenile stage of
life. This can be justified by the fact that some fish species show different behaviors in relation to the life stage. From the analysis of the box plots, in cluster 3 a greater range between the values of $S_{c_{\text{max}}}$ and $S_{c_{\text{min}}}$ is visible more than in the other clusters. This characteristic is in agreement with literature studies that observed that juvenile individuals form very compact schools in the central part and more dispersed in the peripheral/external part [44]. Therefore, the similarities of the morphological characteristics between clusters 2 and 3 lead to the hypothesis that the species identified by both is only one, the Atlantic cod. In fact, it is known from literature that the biological behavior of this species is different in its main life stages, as juveniles and as adults. Furthermore, this species appears to be among the most present species [24] in the central area of the fjord as a species capable of tolerating a high range of variability in both salinity and temperature [18]. From the study of the biological characteristics and behavior of this species it can be highlighted how the morphological and energetic characteristics as well as the bathymetric preferences support this hypothesis. This species reproduces in a period between December and June [14], [21], this would justify the presence of juveniles in our sampling period (August). Fisher et al. (2017) [24] stated that in the period of August 2014 a great abundance of cod was recorded in shallow waters with an average standard length between 40 and 80 mm, while adult individuals up to 300 mm have been observed sporadically in shallow waters. These individuals, once having reached an average length of 100-125 mm, appear to abandon shallow water habitats by migrating to deeper habitats or offshore [24]. Our results are in agreement with literature data as regards the greater presence of Atlantic cod in the area compared to the others [33], [53], [7]. Furthermore, to corroborate this it must be considered that this species, even if tolerant to low temperatures, is not able to withstand the low salinity values present in the proximity of the glacier. In fact, it can easily be assumed that cluster 1, located in the innermost part of the fjord, is the other species, endemic to the area, and namely the Polar cod. Among the species considered so far, the latter appears to be the only one able to tolerate low salinity values typical of a marine area in which there is a strong dilution of fresh water coming from the melting of the Kronebreen glacier. The depth preference is also in line with what reported in literature. Marine fish population can react in different ways to environmental changes; unfavorable environmental conditions could impact on the population dynamics leading to changes in abundance [58] or in the depth and/or geographical distribution [13], [47], [20]. The intensity and extent of such responses depends on a number of factors such as the thermal tolerance, the migratory capacity and the ability to adapt to different environmental conditions. Nonetheless, the adaptation rate is population and species-specific [30], [48]. Furthermore, new species (often referred as alien species) entering in a specific ecosystem could lead to deep modification in community structure, even causing local extinctions. The case of Atlantic cod (Gadus morhua) competing with the Polar cod (Boreogadus saida) is one such example. The increasing abundance of Atlantic cod, observed in the Kongsfjorden area, is considered as one of the effects of the ongoing environmental changes.

5. Conclusions
The results of this work show that the dataset is composed, from a numerical point of view, of three clusters while, from a biological point of view, the species associated with the identified patterns are likely two, and one of them with different life stages and different energetic characteristics. In the light of the results obtained, we can sustain that it is possible, starting from raw acoustic data, to distinguish fish schools in relation to some morphological, energetic and bathymetric positional parameters of the aggregations, in particular study areas such as the one analyzed in this context. The results obtained are in agreement with other studies based on biological data published for the same study area. Therefore, this method, based on the analysis of acoustic data, has proved to be very useful for the recognition of species and can be used to support biological research on target species.
The main advantage lies in the fact that the analyzed characteristics may allow to obtain a classification of the species on the basis of their biological behavior, without using invasive sampling methods, thus reducing the fishing effort. This work opens up different scenarios on the possibility of using this data for identifying fish species and therefore represents a starting point for possible future works in this area and in other restricted areas.

Acknowledgement

This research was carried with the support of the CALVINGSEIS project (Grant No. 244196/E10) funded by The Norwegian Research Council. The authors thank the Department of Earth Systems Science and Environmental Technologies of the Italian National Research Council for the logistic support in Ny-Ålesund.

References


A pattern recognition approach

higher trophic levels. In: Duarte CM, Agusti S (eds) Impacts of warming on polar ecosystems FBBVA Press, Madrid, 139–175.


