Northwest Atlantic   Fisheries Organization

Robustness in the delineation of ecoregions on the Newfoundland and Labrador continental shelf

Pierre Pepin, Mariano Koen-Alonso, Jennifer Higdon, Neil Ollerhead

Northwest Atlantic Fisheries Centre

Fisheries and Oceans Canada

PO Box 5667

St. John's, Newfoundland & Labrador

Canada A1C 5X1

Corresponding author: P. Pepin (pierre.pepin@dfo-mpo.gc.ca); +1 709 772 2081

Abstract

Ecosystem approaches to management are essentially place-based approaches; they aim to provide management provisions and advice encompassing multiple stocks which inhabit a common and geographically-defined area. These "ecosystem management" units, and the scale at which they are defined, ideally would capture the core of a functional ecosystem. Here, we evaluated how the delineation of ecoregions on the Newfoundland Shelf could be affected [1] by changes in population abundance and ecosystem structure over time, and [2] whether relatively simple descriptions of biological variables (e.g. biomass, diversity, richness) appropriate when there is a significant spatial variation in community structure. The approach used in ecoregion delineation essentially relies on producing quantitative layers that link different features of the ecosystem, both physical and biological, through principal components analyses (PCA) to define areas with similar features based on clustering algorithms. The analyses included data from the Engels (1980-1994) and Campelen (1995-2010) surveys. To investigate whether the delineation of ecoregions was affected by averaging over extended periods, we used averaging periods of 4 (Campelen) and 5 (Engels) years to assess the effect of changes

in bottom temperature, biomass, evenness and richness. The optimal number of clusters identified from Engels surveys was variable over time, partly as a result of changes in data extent and availability over time. The period associated with collapse of NL ecosystem (1990-1994) shows a high degree of spatial fragmentation in the distribution of clusters. The optimal number of clusters from the Campelen surveys was generally stable. To investigate whether relatively simple descriptions of biological variables (e.g. biomass, diversity, richness) were appropriate when there is a significant spatial variation in community structure, we described community structure for the two survey series based on multivariate exploratory descriptors (PCA scores) using species biomass in each tow as a measure of abundance. The analyses of community structure for the Engels, Campelen and Campelen using only the species collected by the Engels surveys all revealed a strong and clear separation of taxa in relation to depth, temperature and latitude. The addition of taxonomic layers does little to the optimal number of clusters identified for each survey or the spatial structure in their delineation. The results from both analyses reported here confirm earlier conclusions that the delineation of ecoregions on the Newfoundland Shelf appears to be robust to changes in information content of the analyses.

Introduction

The delineation of biogeographic boundaries in the World's Oceans has been based largely on the identification of water masses, fronts and currents that influence the dispersal of plankton or that place physiological constraints on the habitats which organisms can occupy (Longhurst 2007). The physical characteristics of water masses has long been used to identify differences among major areas in which the primary production potential differs (e.g. Ryther 1969) which thereby serves to identify key ecosystems (e.g. shelf seas, upwelling, open ocean). Ideally, the delineation of marine ecosystems that would serve to assess the state of marine environments should be based on simple ecological criteria that include bathymetry, oceanography, productivity and trophic relationships (www.lme.noaa.gov). Delineation of spatial management units is prerequisite to establishment of an effective ecosystem approach to management of human activities in marine ecosystems. Biogeographic classification has been described as "fundamental for marine spatial planning and can serve as a framework for a number of uses from assessment and monitoring to marine protected areas network design" (CBD 2009). In recent years, there has been an increased need for the assessment of biogeographic classification schemes as many coastal states have moved their policies toward the implementation of Ecosystem Approaches to Management (EAM) in an attempt to recognize the interconnectedness of organisms with the environment in which they live rather than rely on single focus strategies, such as fisheries management. EAM also adds emphasis on the need to conserve biodiversity, unique habitats and sensitive or vulnerable marine ecosystems.

Ecosystem approaches to management are essentially place-based approaches; they aim to provide management provisions and advice encompassing multiple stocks which inhabit a common and geographically-defined area. These "ecosystem management" units, and the scale at which they are defined, ideally would capture the core of a functional ecosystem, though other considerations should also be taken into account in defining them (e.g. jurisdictional boundaries and legal issues, main fisheries and fleets, operational issues regarding surveillance and enforcement, etc.). A necessary

starting point in the process of defining "ecosystem management" units is the delineation of ecosystem boundaries and identification of major ecosystem subunits.

From a practical perspective, society, governments and international organizations (e.g. Regional Fisheries Management Organizations – RFMOs) have had to define spatial management units in the oceans often in response to the needs associated with the management, conservation and exploitation of renewable marine resources well before the need to consider EAM because of the need collected fishery statistics in order to provide scientific advice for management. Halliday and Pinhorn (1990) reviewed the history of the delimitation process of "fishing areas" in the Northwest Atlantic from the 1920s to the present circumstances following the establishment of the Northwest Atlantic Fisheries Organization (NAFO) in 1979. In their review they note:

> *"There is remarkably little documentation of the scientific knowledge which was actually utilized in decision making about the precise locations of specific statistical boundary lines. The historical record does provide accounts of the general principles used in boundary delineation. The NACFI (North American Council on Fisheries Investigations) lines were chosen 'to correspond as far as possible with natural divisions of the fish populations or with barriers to fish migrations' (Found, 1933). Barriers to migrations presumably were of topographic or oceanographic nature. Cote (MS 1953) explicitly listed topography, oceanography and stock structure (in that order) as the 'ideal' bases for subdivision of ICNAF (International Commission for the Northwest Atlantic Fisheries) subareas although, of these, ICNAF (MS 1953) emphasized stock structure only (along with uniformity of size of subdivisions, ease of use and conformity with existing divisions). Faunal composition, or at least the distribution of commercial species, (i.e. zoogeography) also influenced some boundary decisions"*

In essence, Halliday and Pinhorn's (1990) review indicated that although knowledge based, the approaches to the delineation of management units in the Northwest Atlantic had been largely qualitative and the result of a myriad of political, scientific, and arbitrary recommendations applied by a succession of decision-making entities. More recently, the Scientific Council of NAFO agreed that any ecoregion mapping of the Convention Area (NCA) must be consistent with similar mapping done by other coastal states in their respective Exclusive Economic Zones (EEZs) and tasked the Working Group on Ecosystem Approach to Fisheries Management (WGEAFM) to identify regional ecosystems across the entire jurisdictional area of the organization. To this end, WGEAFM members undertook a series of analyses based on the application of multivariate methods that served to synthesize knowledge of the bathymetry, oceanography, biological, and in some instances geological conditions to which a variety of clustering approaches could be applied to identify areas that were statistically more similar (WGEAFM 2008, 2010, 2012).

The different sources of data available to describe the ocean can represent important challenges to the application of quantitative methods for the delineation of spatial management units. Most data are obtained using some form of remote sensing, whether it is fishing nets that can bias which organisms

are being collected, satellites that only represent surface layers, or conductivity-temperature-depth (CTD) sensors that provide high degree of vertical resolution but have limited horizontal resolution owing to survey constraints of slow moving ships. Furthermore, there is considerable vertical structure in the distribution of physical, chemical and biological properties that may be difficult to capture fully across broad geographic areas. These factors can result in differences in the quality and representativeness of different data sources which may, depending on the data available in different parts of the region of interest, affect the outcome of any quantitative delineation process. This may have particularly important implications if the nature, extent and quality of information vary over time.

Previous work of WGEAFM had provided a substantive delineation of ecoregions on the US Northeast Atlantic Shelf (Fogarty and Keith, 2009; Areas 4X5YZe6ABC), the Scotian Shelf (Zwanenburg et al., 2010; Areas 4VsnWX) and the Newfoundland Shelf (Pepin et al., 2010; Areas 2J3KLNO) to identify potential management units for the EAM. The approach used in ecoregion delineation essentially relies on producing quantitative layers that link different features of the ecosystem, both physical and biological, through principal components analyses to define areas with similar features based on clustering algorithms. WGEAFM (2010) concluded that in general terms, the ecoregion analyses presented to date provided a robust basis for the discussion and identification of ecosystem-level units to be used for the initial development and implementation of the "roadmap to EAF". Currently, ecoregion delineation analyses have been based on data layers that condensate multiple years; this makes sense because there is an expectation that these regions should be relatively stable over time. However, little change does not mean "no change".

Pepin et al. (2010) undertook an assessment of the effect of using different subsets of data to carry out ecoregion delineation and concluded that although the environmental features (e.g. bathymetry, temperature, etc.) dominate the identification of ecoregions, biological variables (e.g. demersal biomass, richness, evenness) demonstrated a close association with those descriptors as a result of the substantial environmental along- and cross-shelf gradients in the region. Because of the need to provide robust advice concerning the delineation of ecosystem units and subareas for application of the ecosystem approach to management, we evaluated how the delineation of ecoregions could be affected [1] by changes in population abundance and ecosystem structure over time, and [2] whether relatively simple descriptions of biological variables (e.g. biomass, diversity, richness) are appropriate when there is a significant spatial variation in community structure. Here, we undertook an analysis aimed at addressing these issues based on available information for the Newfoundland and Labrador Shelves (42-58°N; Figure 1) which have been identified as one of the twelve major marine biogeographic units in the Canadian EEZ by Fisheries and Oceans Canada (DFO 2009). The region, which includes the Grand Banks, is located at the confluence of four of Longhurst's (2007) major biogeographical provinces: Polar, Arctic, Gulf Stream and Northwest Atlantic Shelf, the latter of which extends all the way along the eastern coast of North America. The area is influenced by a strong equatorward current carrying cold Arctic water (Labrador Current), as well as the circulation of the subpolar gyre (Labrador and Irminger Seas) to the East and the Gulf Stream and its meanders in the South (Loder et al., 1998; Han et al., 2008). In addition, regional sea-ice cycles and transport play an important role in the freshwater balance of the region, with pack-ice extending as far South as the Grand Banks during climatic extremes (Colbourne et

al., 1997). The region of interest is relatively data poor when contrasted with other parts of the NAFO Convention Area, such as the Scotian Shelf or US Northeast Continental Shelf. The work represented here builds on analyses presented by Pepin et al. (2010).

Methods

The first step was to acquire all the datasets to be used in the analysis. The variables selected for inclusion in the analyses presented here were based on the reduced set used in the final recommendations by Pepin et al. (2010) and include: bathymetry, surface and bottom temperatures, chlorphyll *a* concentrations, primary productivity, and biomass, richness and evenness derived from multispecies bottom trawl surveys. This is considerably fewer variables than the information content used on the analyses performed for the US continental shelf areas (Fogarty and Keith 2009) but the analyses presented by Pepin et al. (2010) provided clear evidence that the overall biogeographic structure of the Newfoundland Shelf and Grand Banks could be defined reliably using the subset of data chosen for this analysis. Most of the data were not continuous surfaces, but rather vector point database format, and therefore had to be interpolated to a common gridded surface (Figure 2). Once all the data were represented as continuous surfaces they had to be made spatially comparable (i.e. perfectly overlapping cells of the same size); therefore, all raster datasets were aligned, resampled, and/or aggregated to a standard 20 km grid. Before the datasets were used in multivariate analyses, all were standardized to a common scale (mean = 0; s.d. = 1). Following these steps, the data were analyzed using principal component analysis (PCA) and then classified and mapped using k-means clustering.

Any variables that were originally vector datasets were converted into continuous raster surfaces using appropriate interpolation methods (Goulet et al., 2010). These vector datasets were all provided from various Fisheries and Oceans Canada (DFO) surveys (Goulet et al., 2010). The cell size of the interpolated surfaces was dependent on the spatial distribution of samples from the original datasets, ranging from 2 km to 20 km. The demersal and bottom temperature datasets were dense and interpolated to grids of 2 km or 2.5 km depending on the dataset (Goulet et al., 2010). To maintain consistency between the raster and vector datasets, all surfaces were converted to a 20 km grid using the aggregate tool in ArcGIS (ESRI, 2008). The value assigned to each 20 km cell was calculated by taking the mean values of all the original 2 km, or 2.5 km, cells within the larger 20 km cell. Certain datasets (sea surface temperature, primary production, Chl-*a*, and bathymetry) could not be brought to a 20 km grid using the aggregate tool as they were not an integer factor of 20, which is a requirement of the aggregate tool (ESRI, 2008). Therefore, to calculate the mean values of these variables within each 20 km cell the datasets were resampled, using the nearest neighbour algorithm, to create a raster that was an integer factor of 20. The factor value of 20 that was closest to the original cell size was used, for example, the sea surface temperature raster had a cell size of ~4.8 km so it was resampled to a 5 km cell size. Once this was completed, the resampled raster was aggregated into 20 km cells using the aggregate tool.

The GEBCO (General Bathymetric Chart of the Oceans) bathymetry dataset was downloaded from the GEBCO website (www.gebco.net). GEBCO is composed of an international group of experts who work on

the development of bathymetric datasets and operates under the auspices of the International Hydrographic Organization (IHO) and the Intergovernmental Oceanographic Commission (IOC) of UNESCO.

Sea surface temperature dataset was provided by NOAA. Sea surface temperatures were measured daily using AVHRR Satellite starting in 1985 (for more information see www.nodc.noaa.gov/SatelliteData/pathfinder4km/).

The Bedford Institute of Oceanography (BIO) provided the chlorophyll *a* (Chl-*a*) and primary production (PP) datasets. The original Chl-*a* datasets were acquired from the SeaWiFS (Sea Viewing Wide Field of View Sensor) satellite sensor mounted on the Orbview-2 satellite operated by GeoEye (http://www.geoeye.com/CorpSite/). The Chl-*a* estimates are derived using the OC4.v4 algorithm (O'Reilly et al., 2000). PP estimates were derived from the Chl-*a* datasets (Platt et al. 2009). The PP image is the average over all years (98-04) and the four Chl-*a* datasets are seasonal averages (spring, summer, fall, winter) over all years (97-07). To maintain consistency, the four Chl-*a* seasonal averages were averaged to produce a single Chl-*a* dataset, matching the PP dataset.

Demersal and bottom temperature raster datasets were all interpolated using ordinary kriging (Goulet et al., 2010). Richness was estimated as the number of taxa per demersal or pelagic tow. Diversity (evenness) was represented using Shannon's Evenness Index ($\left[-\sum_{i=1}^{S} p_i \ln p_i \right] / \ln S$; where *pi* is the proportion of species *i*, and *S* is the number of species).

All the data were standardized to a mean of 0 and standard deviation of 1 ([x – mean (x)] / s.d.). The Raster Calculator inside ArcGIS was used to perform this calculation on the raster datasets.

Analysis

Principal components analysis (PCA) is often performed on high-dimensional data to eliminate redundancy, find patterns, emphasize variance within the variables, and improve interpretation (ESRI, 2010). Essentially, PCA transforms the data in multivariate space to a new multivariate space whose axis are rotated so that the greatest variance is explained by the first principal component, the second principal component (orthogonal to the first) explains the second greatest variance, and so on. The first three or four principal components typically explain the most variance and by analyzing only these components, one reduces the number of dimensions without much loss of information.

The PCA results were used in a k-means clustering procedure to classify the data. K-means clustering is an unsupervised classification technique, meaning there is no prior knowledge on what information classes exist in the data. An information class is a similar grouping of values that are known to belong to a specific class, for example, in satellite imagery classification a class may be defined as a meaningful grouping of locations representing real world objects such as water or forest. A cluster, on the other hand, is simply a statistical grouping in the data with similar attribute values in multivariate space with

no knowledge on what that cluster represents in the real world. These clusters must be interpreted into meaningful classes by the user.

The raster outputs of the first four principal components from the ArcGIS PCA analysis were used as input into k-means clustering using the algorithm of Legendre (2001). The number of clusters we investigated ranged from 2 to 10. Legendre's (2001) algorithm provides the optimal number of clusters as determined by the Calinksi-Harabasz (C-H) statistic, and the count of observations within each cluster. The C-H statistic is calculated for different number of clusters using the following equation (Legendre, 2001):

$$C\text{-}H = [R^2/(K-1)]/[(1-R^2)/(n-K)]$$

where R^2 = (SST − SSE)/SST, SST is the total sum of squared distances, SSE is the sum of squared distances of the objects to their group's own centroids, and K is the number of groups. The number of clusters that yields the highest C-H criterion corresponds to the most compact set of clusters, or optimal number of groups (Legendre, 2001). The output from the k-means clustering was then mapped to visualize the distribution of clusters from each run.

Results

Time-averaging

In this section, we aimed to assess whether changes in population abundance and ecosystem structure over time influence the potential delineation of ecoregions. The analyses reported by Pepin et al. (2010 were based on the average of multispecies surveys performed with a Campelen trawl over the years 1995-2007. Here we extend the analyses to include an earlier survey period when an Engels trawl served as the basis for stock assessments 1980-1994, and the Campelen survey period was extended to include 2008-2010. Coverage of northern (2J3KL) started in 1980 and extended to include 3NO in 1990. Survey methods also differed between the two period with tows routinely lasting 30 minutes for Engels (1.8 nm), and 15 minutes for Campelen (0.8 nm) surveys. However, because the surveys are primarily designed to provide an index of numbers and biomass for a diversity of species, we considered that each survey series was internally consistent and so we did not combine the two series. To determine if the delineation of ecoregions was affected by averaging over extended periods, we used short averaging periods of 4 (Campelen) and 5 (Engels) years to assess the effect of changes in bottom temperature, biomass, evenness and richness. Bathymetry, surface temperature, phytoplankton abundance and production were kept constant.

The optimal number of clusters identified from Engels surveys was variable over time (Figure 3). This is partly a result of changes in data extent and availability over time (Figure 4). However, the general distinction between the Grand Banks and the Newfoundland Shelf, identified by the analyses of Pepin et al. (2010), remain clearly apparent. We note that the period associated with collapse of NL ecosystem (1990-1994) shows a high degree of spatial fragmentation in the distribution of clusters. Changes in association of biomass, diversity and richness relative to physical structure of environment, resulting

from declines in the abundance of demersal stocks, may also be leading to variations in delineation of clusters (Table 1). The optimal number of clusters from the Campelen surveys was generally stable but the three additional years (2008-2010), relative to the work of Pepin et al. (2010), suggest that a lower number of clusters may be optimal (Figure 5). This may have been caused by the relatively low number of observations during this period (1678 versus 2476-3183 during other Campelen averaging periods). However, this may also be indicative that changes in community or ecosystem structure associated with recovery of stocks can be a source of uncertainty and variability in the stability of ecoregion delineation. This interpretation is supported somewhat by a comparison of the final PCA run described by Pepin et al. (2010) (i.e., with the same variables that have been applied in this analysis) with the results obtained by the addition of data from 2008-2010 (Figure 6). The re-analysis reveals limited difference in the Calinski-Harabasz statistic for an optimal number of clusters ranging from 4 to 6, but in all instances the spatial distribution of clusters maintains the overall ecoregion structure identified by Pepin et al. (2010).

Taxonomic information

In this section, we address whether relatively simple descriptions of biological variables (e.g. biomass, diversity, richness) are appropriate when there is a significant spatial variation in community structure. Here, we describe community structure for the Engels (1980-1994) and Campelen (1995-2010) surveys based on multivariate exploratory descriptors (PCA scores) using the biomass of taxa (identified to the species level or grouped to genus in instances where species identification was uncertain or inconsistent) in each tow as a measure of abundance. Only taxa that had an overall occurrence frequency greater than or equal to 3% of all trawl sets were included in the analyses. The scores from first three principal components were included as additional variables to the ecoregion PCA and k-means clustering steps described in Figure 2. Because the Campelen surveys use a finer mesh, they yield a greater number of species than the Engels surveys, we performed two analyses of Campelen surveys: [1] using only the taxa collected in the Engels surveys; [2] using all taxa with > 3% occurrence.

The analyses of community structure for the Engels, Campelen and Campelen using only the taxa collected by the Engels surveys all revealed a strong and clear separation of taxa along PC1 – PC3. The high number of observations ($N_{Engels} = 7746$; $N_{Campelen} = 9983$) and high sampling variability contribute to low explanatory power of single components (PCs) of the multivariate analysis (Table 3). PCs 1-3 explain between 26% and 33% of variance in taxonomic composition, depending on the choice of survey and taxa included in the analyses. The overall results were largely unaffected by analytical method or transformation of data. Eigenvalues 1 to 10 were statistically significant for the 34 taxa analyses; eigenvalues 1-17 were significant for the 64 taxa analysis. The general pattern of separation of taxa was consistent in all the analyses of the two survey series (Engels versus Campelen) whether the taxa of the Campelen were a subset based on the taxa composition of the Engels survey or not. Depth was strongly positively correlated with PC1. This was reflected in the illustrative example presented in Figure 8 where there is clear separation of shallow and deep water taxa, cold water and warm water groups. The association of temperature with PC1 to PC3 was variable among the different analyses. Latitude and longitude were associated with PC3 with Engels surveys but most strongly associated with PC2 and PC3 for the Campelen surveys.

Adding the principal component scores of the analyses of the taxonomic structure of the multispecies surveys to the ecoregion PCA resulted in similar outcomes irrespective of the survey series used in the analysis (Tables 4-6). The addition of taxonomic layers does not change the optimal number of clusters from Engels surveys (Figure 9) or the spatial structure in their delineation (Figure 10). Similar results were obtained for the Campelen surveys although optimal number of cluster appears to be greater than in simpler analyses (Figure 9). The differences between the analyses of the Campelen surveys, with taxonomic information relative to those without, are most notable in southern portions of region where there is strong gradient in community structure (Figures 11 and 12).

Discussion

The delineation of subareas (clusters) of ecoregions appears to be dependent on representativeness of data. The Campelen surveys provided more comprehensive sampling of ecosystem and therefore better representation than the Engels surveys. Furthermore, the limited coverage by Engels surveys affected value of these data in delineation of subareas. It is clear that changes in community or ecosystem status will affect subarea delineation process based on relatively coarse (biomass, diversity, richness) metrics of biological variables. However, overall definition of ecoregions in the NAFO zones of interest to this study would only modestly be affected by changes in subarea delineation – Grand Banks and NL Shelf still appear as distinct. How to base the final delimitation of ecoregions may ultimately depend on the significance of subareas to ecosystem function, which may represent critical unknowns in the identification of operational ecosystem elements (management units).

Analyses of community structure demonstrated that the distribution of the major taxa on the Newfoundland Shelf is strongly tied to bathymetric and latitudinal gradients in the region. The addition of taxonomic information did not enhance the inferences based on the Engels surveys, probably because of limited spatial scope of the surveys prior to 1990 rather than because of the value of the taxonomic information. Analyses of the Campelen surveys with the same taxa collected in the Engels trawl revealed greater detail in the delineation of subareas within the ecoregions. In the case of the Campelen surveys with all taxa with >3% occurrence, the addition of three layers aimed to represent the taxonomic diversity across the region without overwhelming multivariate analysis had limited impact on the definition of ecoregions although they provided greater definition of the subareas within each ecoregion. The greatest influence of adding the taxonomic information is on delineation of subareas located in the southern portion of the Newfoundland Shelf. In particular, the analysis reveals that the deep water community on southern edge of Grand Banks is different from the one on the eastern slope of the Newfoundland Shelf.

The results from both analyses reported here confirm earlier conclusions by Pepin et al. (2010) that the delineation of ecoregions on the Newfoundland Shelf appears to be robust to changes in information

References

CBD 2009. Secretariat of the Convention on Biological Diversity (2010). Year in Review 2009. Montreal, 42 pages.

Colbourne, E., deYoung, B., Narayanan, S., Helbig, J., 1997. Comparison of hydrography and circulation on the Newfoundland Shelf during 1990-1993 with comparison to the long-term mean. Canadian Journal of Fisheries and Aquatic Sciences 54, 68-80.

Côté J. MS 1953. Subdivision of Subareas. ICNAF Meeting Document., No. 18, Serial No. 90, 2 p. DFO 2009. Development of a Framework and Principles for the Biogeographic Classification of Canadian Marine Areas. Canadian Scientific Advisory Secretariat, Scientific Advisory Report 2009/056, 17p.

ESRI 2010. ArcGIS Version 10.0. Environmental Systems Research Institute (ESRI), Redlands, CA, USA.

Fogarty, M.J. and Keith, C. (2009 – Unpublished). Delineation of regional ecosystem units on the U.S. Northeast Continental Shelf. NEFSC Discussion Paper. Found, W. A. 1933. North American Council on Fishery Investigations. Third annual report of the Department of Fisheries for the year 1932-33. Ottawa, p. 25-27.

Halliday, R.G., and Pinhorn, A.T. 1990. The delineation of fishing areas in the Northwest Atlantic. J. North. Atl. Fish. Sci., 10: 1-51.

Han, G.Q., Lu, Z.S., Wang, Z.L., Helbig, J., Chen, N., de Young, B., 2008. Seasonal variability of the Labrador Current and shelf circulation off Newfoundland. Journal of Geophysical Research-Oceans 113, C10013 10010.11029/12007jc004376.

Goulet, P., Walsh, M., Cuff, A., Devillers, R., Edinger, E. 2010. Data analysis towards GIS-based ecoregion delineation of the Newfoundland Shelf, 1995-2007. Data Report, Department of Geography, Memorial University of Newfoundland, 43p.

ICNAF 1953. Third annual meeting, fourth and final plenary session. ICNAF Meeting Proceedings, No. 18, Serial No. 131, 4p. + Appendix.

Legendre, P. 2001. Program K-means user's guide. Département de sciences biologiques, Université de Montréal. 11 p.

Loder, J.W., Petrie, B.D., Gawarkiewicz, G., 1998. The coastal ocean off northeastern North America: A large-scale view. In K.H. Brink, A.R. Robinson (Eds.), The Global Coastal Ocean: Regional Studies and Synthesis, Vol. 11 (pp. 105-133). Hoboken, N. J.: John Wiley.

Longhurst, A.R., 2007. Ecological geography of the sea. Burlington: Academic Press.

O`Reilly, J.E., et al. 2000. SeaWiFS postlaunch calibration and validation analysis, Part 3. In Hooker, S.B. and Firestone, E.R. (Eds.), SeaWiFS Postlaunch Technical Report Series, Volume 11. NASA Goddard Space Flight Center.

Pepin, P., Cuff, A., Koen-Alonso, M. and Ollerhead, N. (2010). Preliminary analysis for the delineation of marine ecoregions on the Newfoundland and Labrador Shelves. NAFO SCR Doc. 10/72, 24p.

Platt, T., Sathyendranath, S., Forget, M-H., White III, G.N., Caverhill, C., Bouman, H., Devred, E., Son, S.H. 2008. Operational estimation of primary production at large geographical scales. Remote Sensing of Environment, 112, 3437 – 3448.

Ryther, J.H., 1969. Photosynthesis and fish production in the Sea. Science 166, 72-76.

WGEAFM 2008. Report of the NAFO Scientific Council Working Group on Ecosystem Approaches to Fisheries Management. NAFO Science Summary Report Doc. 08/10, 70p.

WGEAFM 2010. Report of the 3rd Meeting of the NAFO Scientific Council Working Group on Ecosystem Approaches to Fisheries Management. NAFO Science Summary Report Doc. 10/24, 75p.

WGEAFM 2012. Report of the 5th Meeting of the NAFO Scientific Council Working Group on Ecosystem Approaches to Fisheries Management. NAFO Science Summary Report Doc. 12/26, 130p.

Zwanenburg, K., Horsman, T., and Kenchington, E. 2010. Preliminary analysis of biogeographic units on the Scotian Shelf. NAFO Scientific Report Doc. 10/06, 30p.

Period	Layer	PC1	PC2	PC3
Engels 1980-1984	**Percent variance**	36%	22%	15%
	Sea Surface Temperature	0.655	0.188	0.194
	Primary Production	0.453	0.417	0.257
	Chlorophyll *a*	-0.117	-0.031	0.031
	Bathymetry	-0.572	0.694	0.313
	Biomass	-0.088	-0.279	0.093
	Evenness	0.020	-0.017	0.496
	Richness	-0.132	-0.392	0.391
	Bottom Temperature	-0.017	-0.276	0.625
Engels 1985-1989	**Percent variance**	30%	29%	15%
	Sea Surface Temperature	-0.359	-0.079	-0.128
	Primary Production	-0.251	-0.218	-0.352
	Chlorophyll *a*	0.052	0.050	-0.005
	Bathymetry	0.884	-0.256	-0.159
	Biomass	0.055	0.459	0.136
	Evenness	-0.040	-0.018	-0.867
	Richness	0.101	0.571	-0.175
	Bottom Temperature	0.097	0.584	-0.182
Engels 1990-1994	**Percent variance**	37%	20%	18%
	Sea Surface Temperature	0.673	0.169	0.139
	Primary Production	0.443	0.373	0.221
	Chlorophyll *a*	-0.115	-0.050	0.034
	Bathymetry	-0.546	0.601	0.482
	Biomass	0.155	-0.047	0.574
	Evenness	-0.016	0.108	-0.100
	Richness	-0.119	-0.536	0.248
	Bottom Temperature	0.027	-0.409	0.546

Table 1. Principal component loadings of variables used in analyses for different periods based on the surveys performed using the Engels trawl (1980-1994)**.**

Period	Layer	PC1	PC2	PC3	PC4
Campelen 1995-1999	**Percent variance**	35%	25%	17%	12%
	Sea Surface Temperature	0.663	-0.076	0.205	0.240
	Primary Production	0.459	0.145	0.304	0.271
	Chlorophyll *a*	-0.121	-0.006	-0.005	-0.037
	Bathymetry	-0.511	0.432	0.504	0.483
	Biomass	-0.060	-0.511	0.477	0.077
	Evenness	0.115	0.189	-0.475	0.642
	Richness	-0.214	-0.435	-0.374	0.382
	Bottom Temperature	-0.108	-0.549	0.137	0.266
Campelen 2000-2003	**Percent variance**	35%	24%	17%	11%
	Sea Surface Temperature	0.664	-0.067	0.192	0.234
	Primary Production	0.457	0.148	0.298	0.269
	Chlorophyll *a*	-0.122	-0.014	0.001	-0.036
	Bathymetry	-0.514	0.416	0.519	0.473
	Biomass	-0.029	-0.504	0.488	0.075
	Evenness	0.171	0.361	-0.367	0.511
	Richness	-0.179	-0.339	-0.479	0.501
	Bottom Temperature	-0.096	-0.548	0.057	0.361
Campelen 2004-2007	**Percent variance**	39%	22%	18%	10%
	Sea Surface Temperature	0.666	0.046	0.184	-0.262
	Primary Production	0.458	0.299	0.225	-0.208
	Chlorophyll *a*	-0.121	-0.021	0.017	0.017
	Bathymetry	-0.504	0.678	0.322	-0.341
	Biomass	-0.065	-0.392	0.661	-0.325
	Evenness	0.183	0.346	-0.491	-0.366
	Richness	-0.200	-0.392	-0.362	-0.692
	Bottom Temperature	-0.030	-0.148	0.048	-0.232
Campelen 2008-2010	**Percent variance**	39%	22%	19%	8%
	Sea Surface Temperature	0.673	0.106	0.199	0.269
	Primary Production	0.450	0.361	0.261	0.143
	Chlorophyll *a*	-0.118	-0.043	0.020	-0.107
	Bathymetry	-0.547	0.546	0.532	0.270
	Biomass	0.086	-0.331	0.593	-0.040
	Evenness	0.045	0.438	-0.467	0.364
	Richness	-0.150	-0.459	-0.077	0.819
	Bottom Temperature	0.002	-0.215	0.182	0.135

Table 2. Principal component loadings of variables used in analyses for different periods based on the surveys performed using the Campelen trawl (1995-2010).

Engels					N=7746	34 Taxa
	Axis	Variance	Latitude	Longitude	Depth	Temperature
	PC1	14.1%	0.275	0.062	0.721	0.642
	PC2	10.5%	0.029	-0.034	-0.307	0.061
	PC3	5.7%	0.470	-0.454	0.182	0.063

Campelen					N=9983	64 Taxa
	Axis	Variance	Latitude	Longitude	Depth	Temperature
	PC1	13.9%	-0.055	0.344	0.892	0.486
	PC2	7.0%	0.547	-0.347	0.256	-0.306
	PC3	5.6%	0.333	-0.105	0.110	0.390

Campelen with Engels Taxa					N=9983	34 Taxa
	Axis	Variance	Latitude	Longitude	Depth	Temperature
	PC1	16.9%	0.019	0.308	0.873	0.550
	PC2	9.5%	-0.390	0.313	-0.239	0.346
	PC3	6.6%	0.491	-0.208	0.027	0.158

Table 3. Summary results of the principal component analysis of community structure from the Engels and Campelen surveys. The axis column provides the percentage of variation explained by each principal axis followed by the correlation of PC scores with latitude, longitude, depth and bottom temperature for each observation.

Layer	PC1	PC2	PC3
Sea Surface Temperature	0.599	0.260	0.122
Primary Production	0.467	0.014	0.306
Chlorophyll *a*	-0.107	-0.036	0.007
Bathymetry	-0.354	-0.556	0.588
Bottom Temperature	-0.162	0.412	0.339
Biomass	-0.053	0.300	0.444
Richness	-0.315	0.400	0.027
Evenness	0.056	-0.048	0.132
Taxonomy PC1	-0.270	0.418	0.068
Taxonomy PC2	-0.091	0.123	0.210
Taxonomy PC3	-0.271	0.100	-0.408

Eigenvalues	1.483	1.128	0.733
Percent Variance	30%	23%	15%
Cumulative Variance	30%	53%	68%

Table 4. Principal component loadings of variables used in analyses that included the principal component scores of the principal component of community structure based on the surveys performed using the Engels trawl (1980-1994)**.**

Layer	PC1	PC2	PC3
Sea Surface Temperature	0.619	-0.192	0.075
Primary Production	0.467	0.073	0.166
Chlorophyll *a*	-0.107	0.020	0.042
Bathymetry	-0.370	0.551	0.438
Bottom Temperature	-0.111	-0.457	0.326
Biomass	-0.050	-0.313	0.550
Richness	0.132	0.112	-0.390
Evenness	-0.253	-0.327	-0.296
Taxonomy PC1	0.058	-0.241	-0.021
Taxonomy PC2	-0.269	-0.107	-0.351
Taxonomy PC3	-0.281	-0.398	0.036

	PC1	PC2	PC3
Eigenvalues	1.532	1.099	0.775
Percent Variance	30%	21%	15%
Cumulative Variance	30%	51%	66%

Table 5. Principal component loadings of variables used in analyses that included the principal component scores of the principal component of community structure based on the surveys performed using the Campelen trawl (1995-2010) using all taxa with >3% occurrence**.**

Layer	PC1	PC2	PC3
Sea Surface Temperature	0.632	-0.140	0.005
Primary Production	0.471	0.117	0.100
Chlorophyll *a*	-0.107	0.011	0.055
Bathymetry	-0.382	0.542	0.436
Bottom Temperature	-0.077	-0.470	0.371
Biomass	-0.005	-0.287	0.537
Richness	0.098	0.074	-0.333
Evenness	-0.251	-0.367	-0.225
Taxonomy PC1	-0.011	-0.348	0.033
Taxonomy PC2	0.198	-0.023	0.435
Taxonomy PC3	-0.321	-0.328	-0.135

	PC1	PC2	PC3
Eigenvalues	1.524	1.07071	0.83402
Percent Variance	30%	21%	16%
Cumulative Variance	30%	50%	67%

Table 6. Principal component loadings of variables used in analyses that included the principal component scores of the principal component of community structure based on the surveys performed using the Campelen trawl (1995-2010) using only the taxa included in the analysis of the Engels surveys**.**
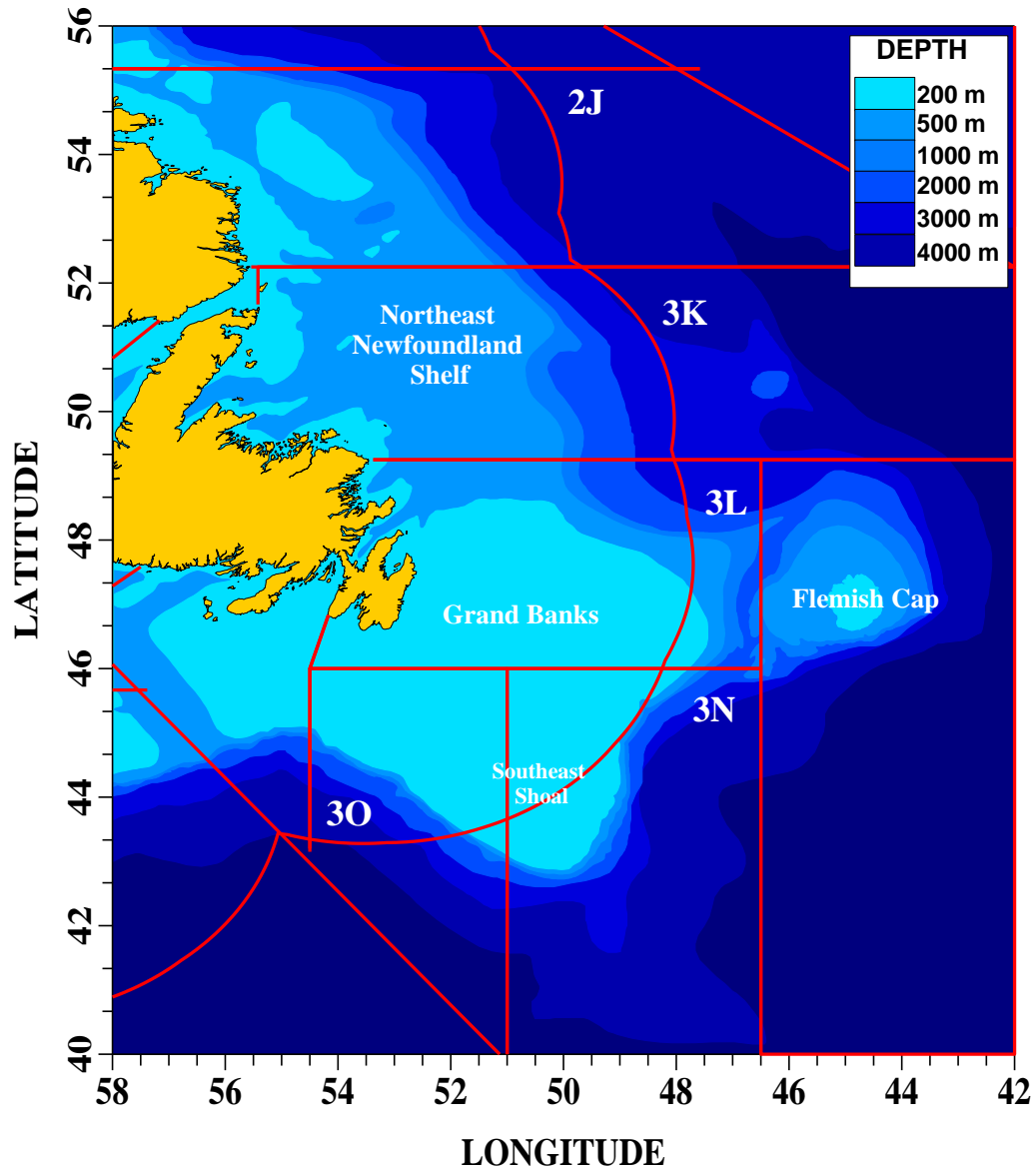
Figure 1. NAFO areas of interest for the assessment of Newfoundland Ecoregions (2J3KLNO). The major bathymetric zones showing the Northeast Newfoundland Shelf, Grand Banks, Flemish Cap as well as the Southeast Shoal region at the tail of the Grand Banks.
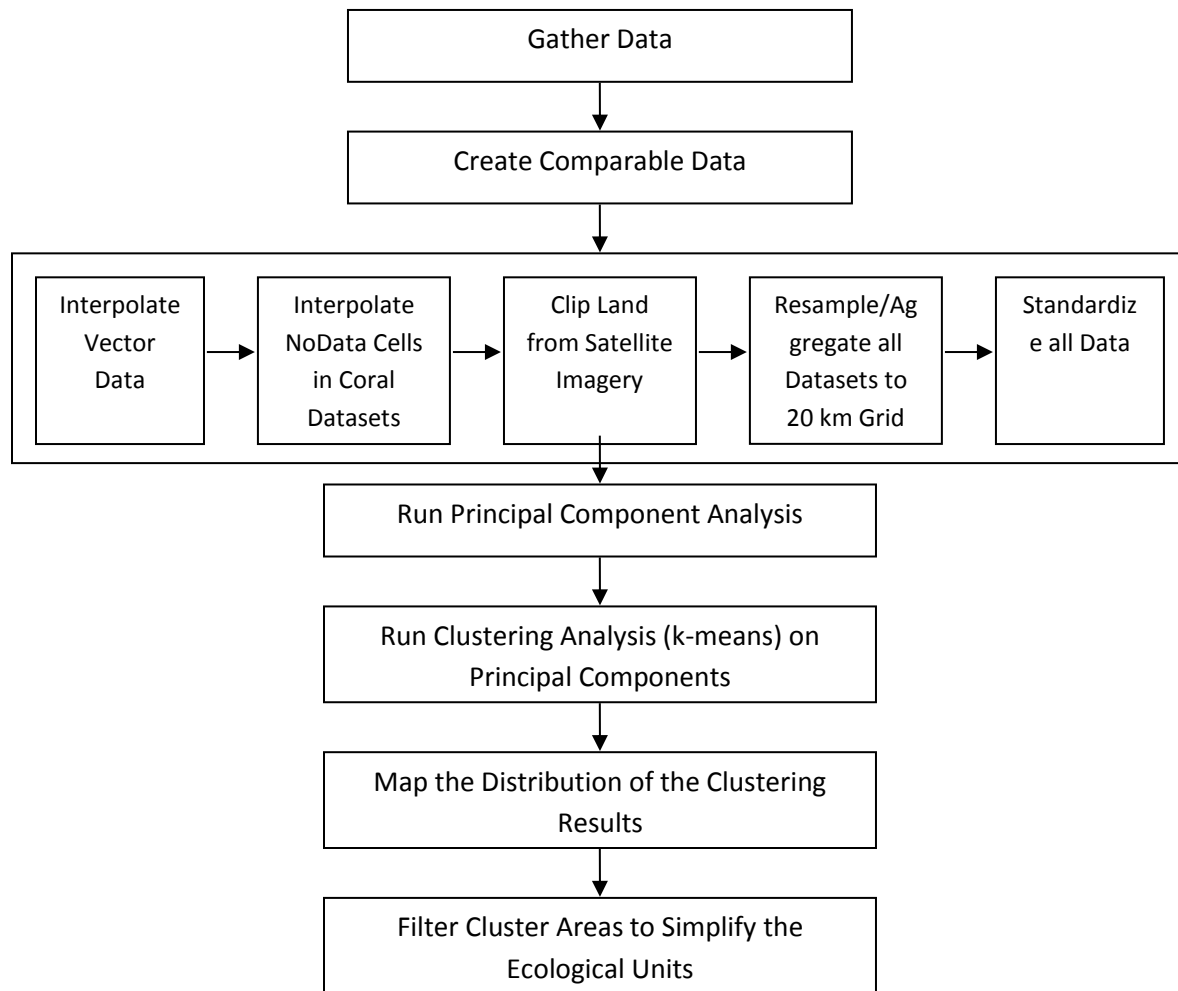
Figure 2: Flowchart of the analytical approach.
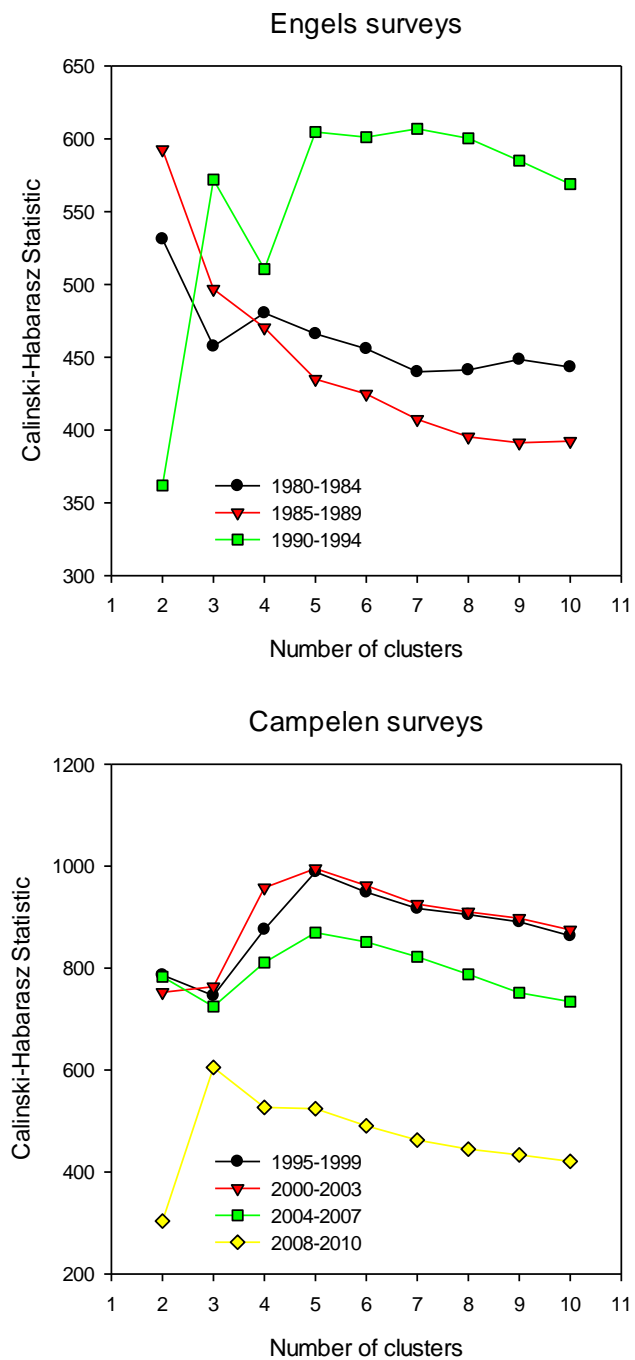
Figure 3. Period specific profile of Calinski-Harabasz Statistic for the Engels (top panel) and Campelen (bottom panel) surveys.
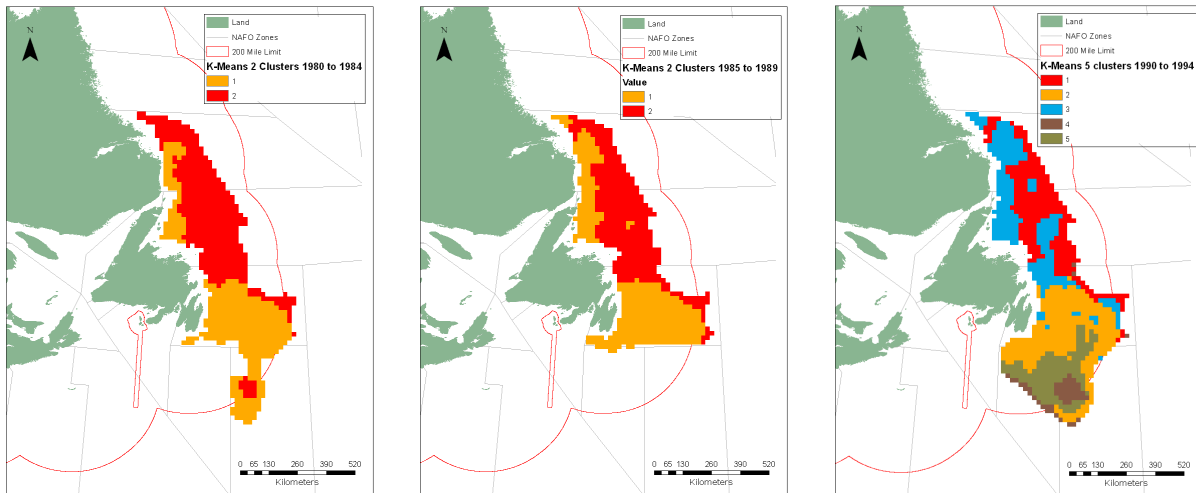
Figure 4. Spatial distribution of the optimal number of clusters for different time intervals of the Engels surveys (left panel: 1980-1984; centre panel: 1985-1989; right panel: 1990-1994).
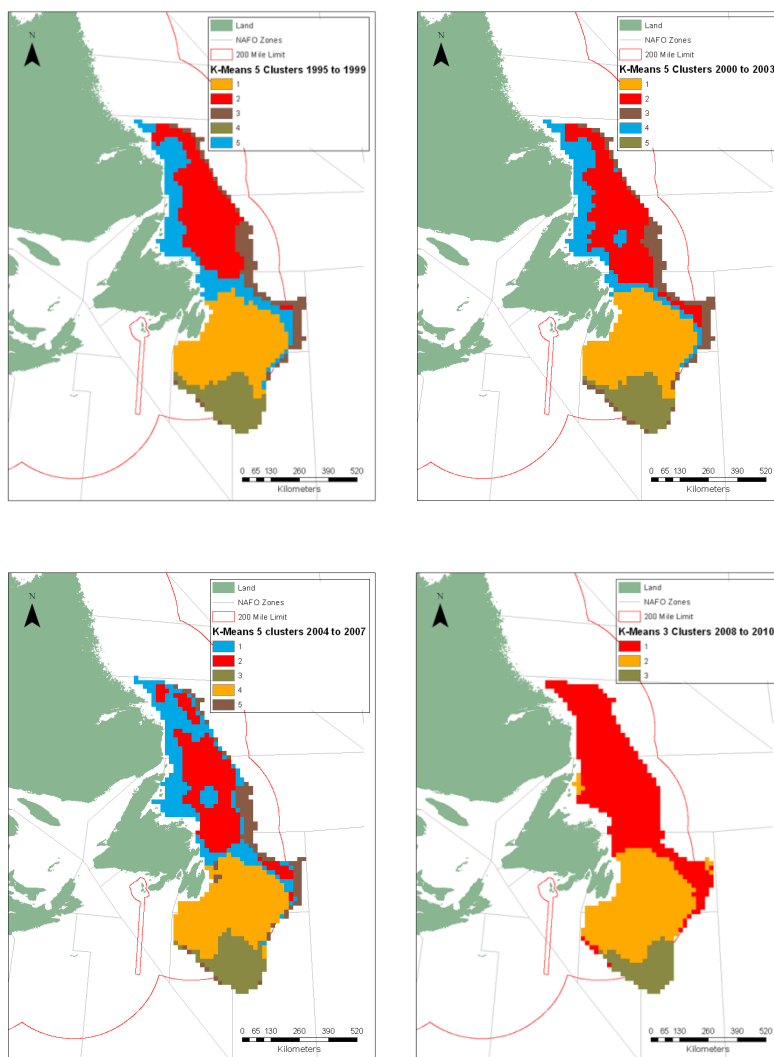
Figure 5. Spatial distribution of the optimal number of clusters for different time intervals of the Campelen surveys (top left: 1995-1999; top right: 2000-2003; bottom left: 2004-2007; bottom right: 2008-2010).
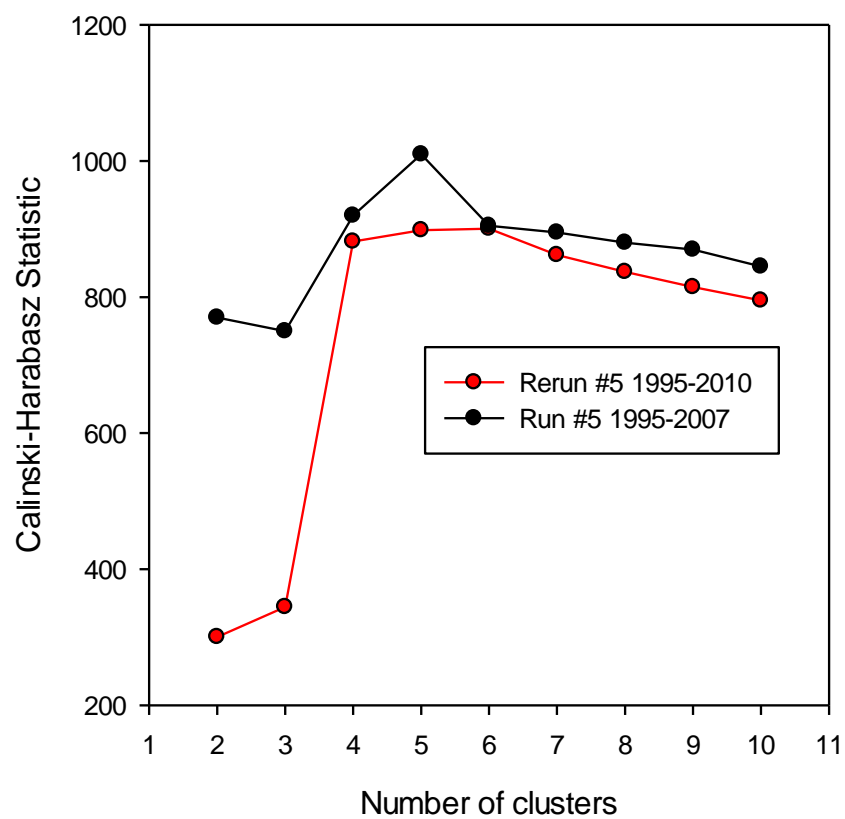
Figure 6. Profile of Calinski-Harabasz Statistic for the analyses reported by Pepin et al. (2010) (Run #5 1995-2007) with a re-analysis with three additional years of information (Rerun #5 1995-2010) from the Campelen surveys.
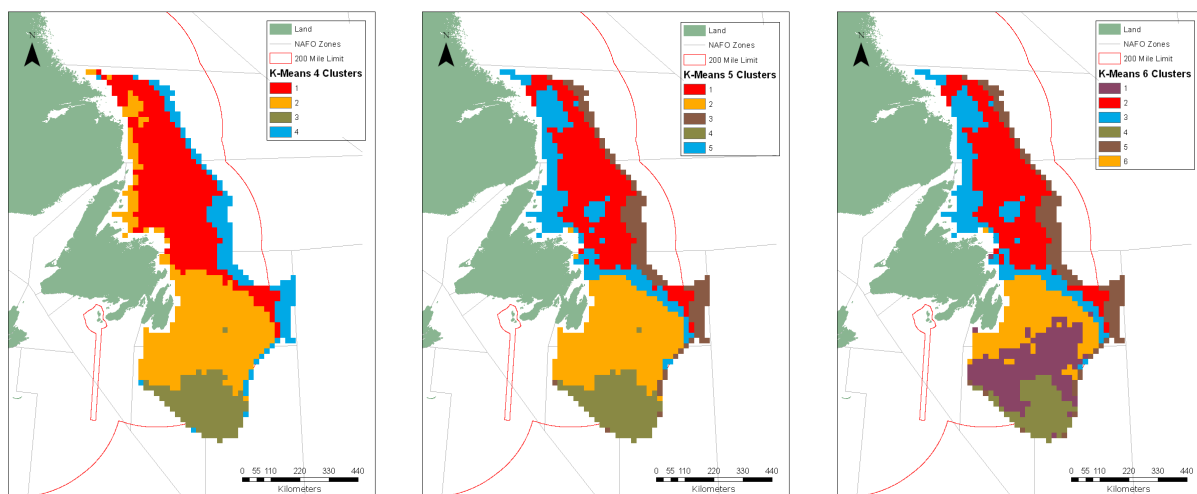
Figure 7. Spatial distribution of clusters based on different quasi-equal optimal solutions of the k-means clustering of the analyses based on the Campelen survey data from 1995-2010.
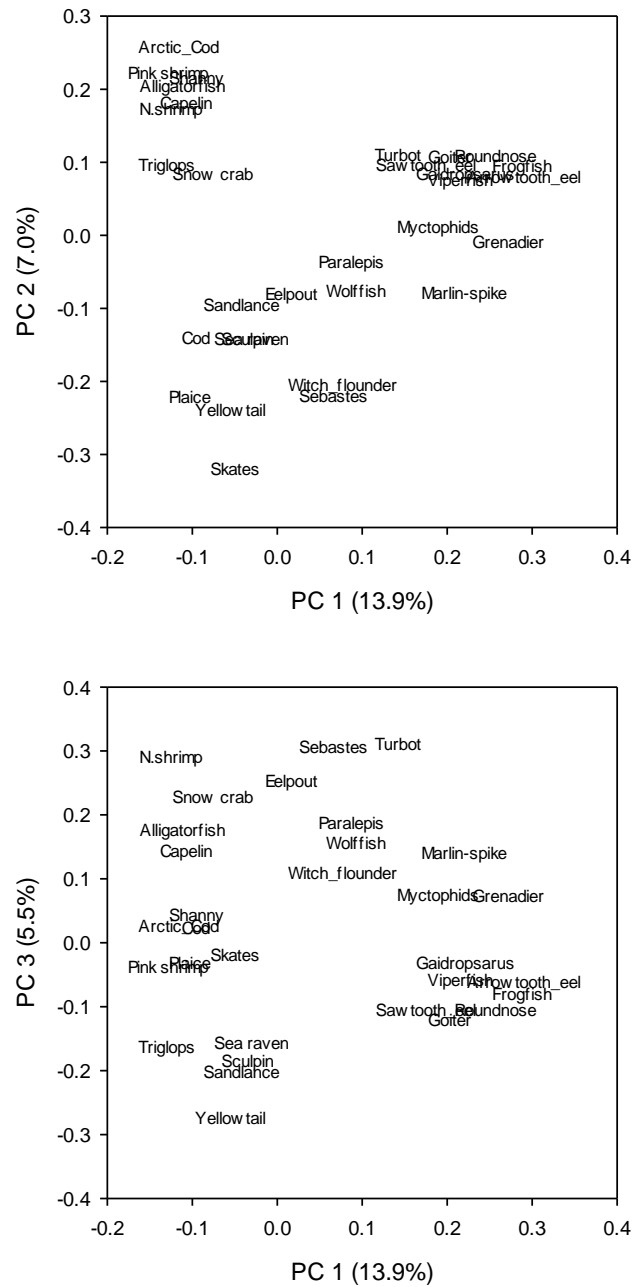
Figure 8. Principal component loadings along the first three principal axes of individual taxa for the analysis of community structure based on all taxa with an occurrence >3% of Campelen sets 1995-2010. Values in brackets represent the percentage of overall variance in community composition based on 9983 trawl sets.
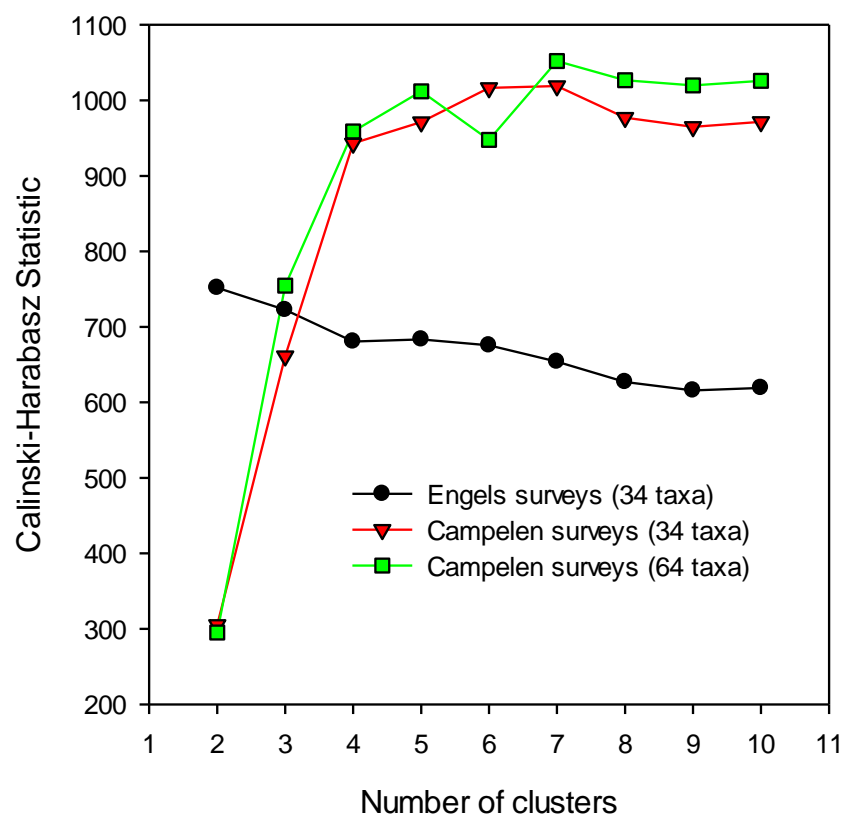
Figure 9. Profile of Calinski-Harabasz statistic for the principal component analyses that included information on taxonomic community structure based on the Engels and Campelen surveys.
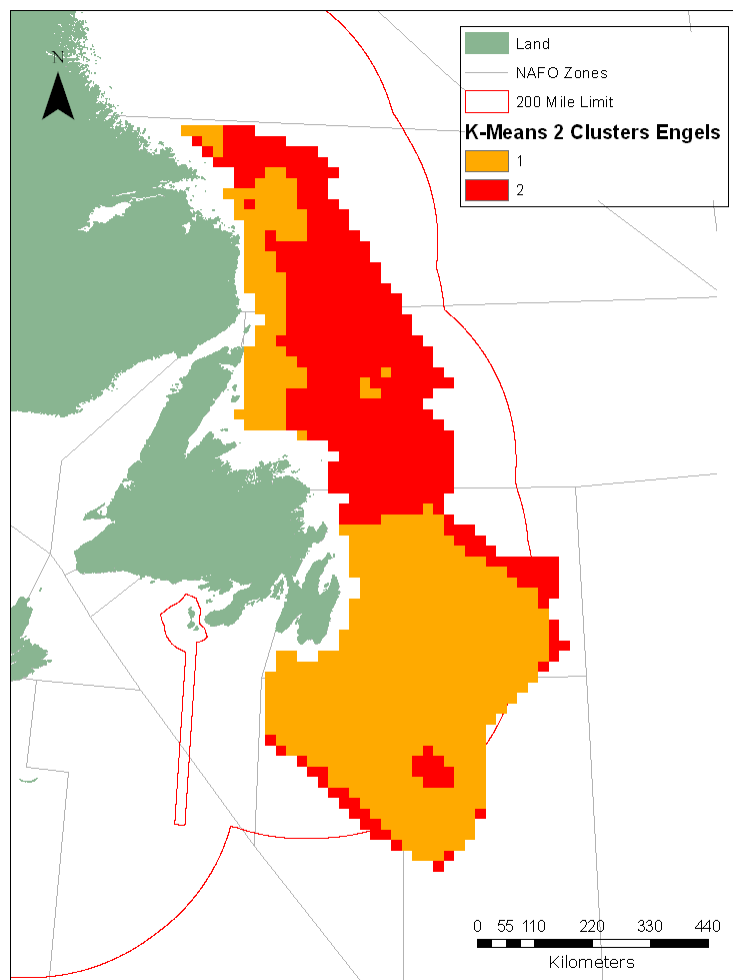
Figure 10. Spatial distribution of the optimal number of clusters for the Engels surveys (1980-1994) based on a principal component analysis that included information on taxonomic community structure with 34 taxa.
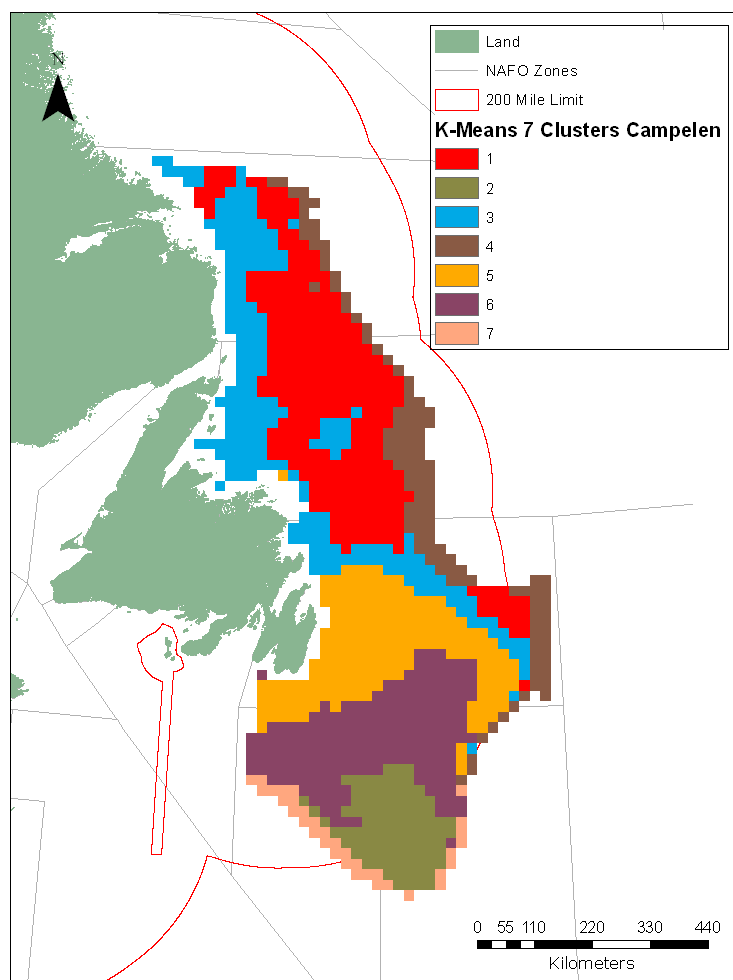
Figure 11. Spatial distribution of the optimal number of clusters for the Campelen surveys (1995-2010) based on a principal component analysis that included information on taxonomic community structure with 64 taxa.
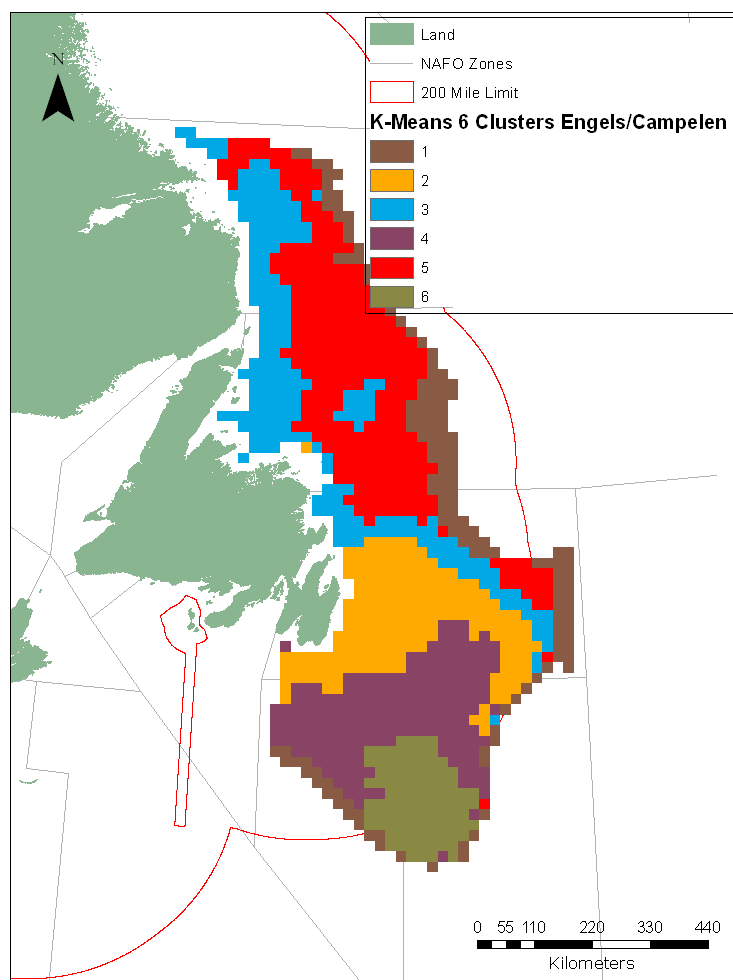
Figure 12. Spatial distribution of the optimal number of clusters for the Campelen surveys (1995-2010) based on a principal component analysis that included information on taxonomic community structure with the 34 taxa captured in the Engels surveys.