



Delineating marine ecological units: a novel approach for deciding which taxonomic group to use and which taxonomic resolution to choose

Anne Fontaine^{1,2*}, Rodolphe Devillers³, Pedro R. Peres-Neto⁴ and Ladd E. Johnson¹

¹Department of Biology, Université Laval, Québec, QC, Canada, ²AgroParisTech, F-75005 Paris, France, ³Department of Geography, Memorial University of Newfoundland, St John's NL, Canada, ⁴Canada Research Chair in Spatial Modelling and Biodiversity, Department of Biological Sciences, Université du Québec à Montréal, Québec, QC, Canada

ABSTRACT

Aim Ecological maps are increasingly used to support marine management and conservation. However, the biological datasets used to produce these maps are typically limited to taxonomic groups identified to the specific taxonomic levels available. Ecological units should, however, reflect the broader marine ecosystem, independent of the datasets used. This study assessed the influence of taxonomic groups and taxonomic resolution on the process of ecological mapping.

Location Estuary and Gulf of St Lawrence (EGSL), Canada.

Methods A dataset of more than 200 taxa of benthic macrofauna was used to create a set of biological matrices corresponding to different taxonomic groups (i.e. vertebrates, invertebrates, arthropods, echinoderms, molluscs, all taxa) and different taxonomic levels from species to class. Multivariate regression trees (MRTs) were used to identify environmental drivers of taxa distribution and to create ecological maps. Similarity between maps was assessed using pairwise comparisons. First, the relationships between the two classification legends were assessed using association plots on the partitions in the corresponding trees. Then, the spatial agreement of ecological units believed to represent the same habitat types was quantified and mapped.

Results The comparison across different taxonomic groups showed a substantial level of similarity between ecological maps, indicating that ecological units defined for a specific taxonomic group can be considered to some extent as representative of the entire benthic macrofauna. Moreover, little information was lost when working at the family rather than species level, and common patterns of community distribution could still be distinguished at the class level.

Main conclusions Using a novel spatially explicit approach for comparing ecological maps, this study demonstrates that datasets limited by taxonomic breadth or resolution can perform nearly as well as more extensive datasets. These simplifications should improve our ability to manage marine ecosystems.

Keywords

benthic communities, distribution patterns, ecological mapping, Gulf of St. Lawrence, map comparison, surrogacy methods.

*Correspondence: Anne Fontaine, Department of Biology, Université Laval, Canada.
E-mail: fontaine.anne@gmail.com

INTRODUCTION

Ecological mapping has generated much interest in supporting an ecosystem-based approach in managing marine environments (Cogan *et al.*, 2009). Ecological mapping aims to spatially delineate ecologically meaningful units, often referred to as ecological units or ecoregions, that can be used to summarize broad habitat distributions and their associated biodiversity at various spatial scales. Such units are used in different contexts, including to support the design of representative marine protected areas networks, the design of sampling strategies to monitor ecological features, and the management of marine activities in an ecosystem perspective (Day & Roff, 2000; Snelder *et al.*, 2005; Costello, 2009; McBreen *et al.*, 2011). Such ecological mapping (also referred to as ecoregionalization, bioregionalization or biogeographic classification) can be defined as 'a classification process that aims to partition a large area into distinct (geographical) regions that contain groups of plants and animals and physical features that are sufficiently distinct or unique from their surroundings at the chosen scale' (UNESCO, 2009; p. 8). In practice, ecological units are identified using a great variety of approaches that differ methodologically and by the type of data used (biological and/or environmental) (Snelder *et al.*, 2007; Mackey *et al.*, 2008). Whatever approach is followed, ecological units will reflect the choice of taxa (e.g. fish) and/or taxonomic accuracy (e.g. genus).

The use of a specific taxon or a limited group of taxa as a surrogate for total biodiversity has already been successfully applied to detect pollution-induced changes (Gray *et al.*, 1988; Clarke & Warwick, 1998) and estimate total species richness (Olsgard *et al.*, 2003; Labrune *et al.*, 2008) in benthic communities. In ecological mapping studies, the selection of taxa is often limited by the availability of datasets or based on authors' preference or expertise in a particular taxonomic group. For instance, demersal fishes were used to produce a map of the Australian marine benthic bioregions (Commonwealth of Australia, 2005) whereas the national

habitat classification for Britain and Ireland was based on benthic communities of invertebrates and seaweeds (Connor *et al.*, 2004). However, research is still needed to confirm the validity of these choices, as ecological units should not only reflect the datasets used to create them, but also broader biodiversity patterns (Commonwealth of Australia, 2005; Kenchington & Hutchings, 2012). Moreover, ecological maps generated from subsets of a same database have not been directly compared to assess differences resulting from the use of different taxonomic groups.

Using levels of taxonomic resolution coarser than species (e.g. family level) is another type of surrogacy method, called 'taxonomic sufficiency' (Ellis, 1985), which has been mainly applied to monitor human impacts on marine ecosystems (Warwick, 1993; Olsgard *et al.*, 1998). This concept has only rarely been examined for distributional patterns in natural systems (Anderson *et al.*, 2005; Wlodarska-Kowalczyk & Kedra, 2007) and has not been validated for comparing marine ecological maps generated from different taxonomic levels.

To address these gaps, this study assesses the influence of the characteristics of biological datasets, the taxonomic groups considered and the taxonomic resolution chosen, for ecological mapping applications. It focuses on the Estuary and Gulf of St. Lawrence (EGSL), eastern Canada, where extensive sampling of benthic macrofauna communities provided a rich dataset for comparing ecological maps produced with (1) different taxonomic groups (i.e. vertebrates, invertebrates, arthropods, echinoderms, molluscs, all taxa) and (2) different taxonomic levels (from species to class).

METHODS

Study area and datasets

The EGSL region (Fig. 1) was selected as a study area for two reasons. First, it is a well-studied ecosystem, surveyed annually for commercial fish and shrimp biomass by

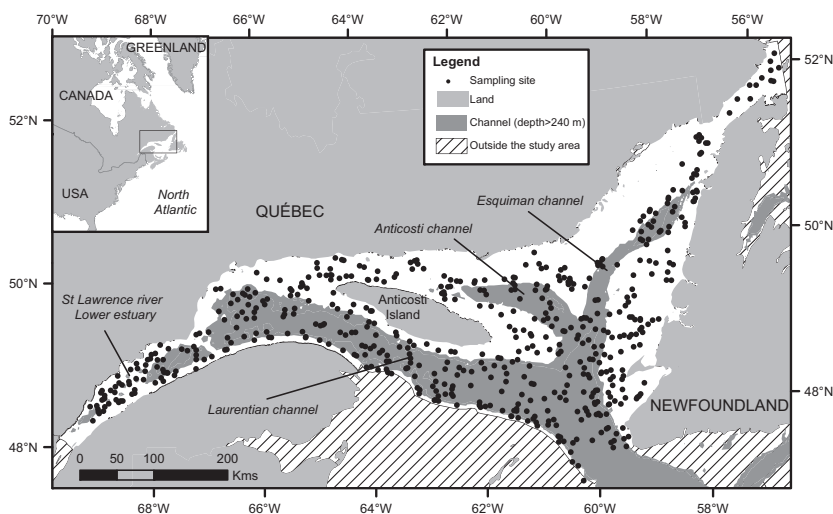


Figure 1 Map of the study area showing the distribution of sampling sites in the lower estuary and the northern part of the Gulf of St. Lawrence, Canada.

Fisheries and Oceans Canada (DFO). Between 2007 and 2009, sampling in the lower estuary and the northern Gulf of St. Lawrence was expanded to include all specimens collected in bottom trawls, generating a detailed dataset on demersal fishes and benthic macro-invertebrates over 140,000 km². Second, the EGSL is a very environmentally heterogeneous ecosystem with large spatial and temporal variations in physical conditions and oceanographic processes (Dufour & Ouellet, 2007; Dutil *et al.*, 2009).

The DFO dataset included 561 samples collected using a four-sided Campelen 1800 shrimp trawl, with a 12.7 mm mesh size for the trawl-lengthening piece and the codend (Nozères *et al.*, 2010) deployed for a standard 15-min bottom tow (an average tow distance of 1.4 km and a sampled area of 0.23 km²). Sampling sites were selected using a random-sampling design, stratified for depth from 37 to 500 m. When possible, all taxa were identified to the species level following the taxonomic classification of the Integrated Taxonomic Information System (<http://www.itis.gov/tools.html>), and the total catch weight in kg per taxon was recorded.

Environmental variables were chosen on the basis of data availability and knowledge from previous studies regarding their potential influence in structuring benthic communities (Day & Roff, 2000; McArthur *et al.*, 2010). Some variables were described by different metrics (e.g. 'bottom temperature' used the annual mean, minimum, maximum, standard deviation and 90th percentile) to boost their potential contribution to the model. While some metrics appeared highly correlated, we used multivariate regression tree that handles collinearity well. Environmental data were collected from different sources, such as *in situ* surveys, historical data, satellite imagery and oceanographic models (a detailed description of the environmental metrics and sources is provided in Table S1). All together, 32 metrics (numerical and categorical) were used to characterize depth, geomorphology, surface sediment and sea-bottom oceanographic conditions (temperature, salinity, dissolved oxygen and current velocity). In addition, surface chlorophyll *a* concentration was also used in the model, being a good proxy for phytoplankton biomass, a major food source for benthic communities in deep waters (Witman & Roy, 2009). Chlorophyll *a* concentrations were derived from semi-monthly composite images captured by the MODIS ocean colour sensor. However, for the North Atlantic region, biases in these values can exceed 50% for chlorophyll concentrations measured during the winter (Gregg & Casey, 2007), mainly due to clouds and the high solar zenith angle. Therefore, the interannual chlorophyll mean for the 2007–2009 period was based only on values from April to August. All environmental metrics were available as continuous features in raster or vector GIS formats, except for bottom dissolved oxygen from the multispecies survey. This variable was interpolated using ordinary kriging in ArcGIS® v.10, using depth as a covariable, as proposed by Dutil *et al.* (2011).

Selecting a mapping strategy

Although benthic and pelagic environments are coupled (e.g. via energy transfer), the processes controlling the distribution of their species have been shown to be fundamentally different (UNESCO, 2009; Spalding *et al.*, 2012). As a consequence, most studies recommend treating them separately for the purposes of ecological mapping (Day & Roff, 2000; Costello, 2009; UNESCO, 2009). Given the available datasets, this study focused only on the benthic realm and did not include species classified as pelagic by Nozères *et al.* (2010), the World Register of Marine Species (WoRMS, <http://www.marinespecies.org>) or the Catalogue of Life (<http://www.catalogueoflife.org>).

For the EGSL, biological data came from discrete sampling sites distributed across the study area whereas numerous environmental variables (e.g. salinity, temperature) were measured or modelled over the entire spatial domain. In this case, a common strategy consists of modelling biological communities according to the suitability of environmental characteristics (Pitcher *et al.*, 2007; Degraer *et al.*, 2008). Here, we applied a multivariate regression tree (MRT) analysis (De'ath, 2002) to identify and model specific assemblages of taxa, or 'community types', in one single step. In this approach, sampling sites were recursively partitioned into two clusters, each partition being based on a split criterion (threshold value) of an explanatory variable that minimized the biological dissimilarity within each cluster. Thus, each terminal cluster (hereinafter called a 'leaf') represented one community type. The rules determined by the succession of environmental variables defining the leaf within the structure of the tree characterized its associated habitat (Fig. 2). Each cluster, corresponding to a specific community type and its associated habitat, was then mapped using environmental layers. The spatial units thus defined were called 'ecological units'. The different types of ecological units (i.e. the different habitats) formed the 'classification legend' that accompanied each ecological map.

Although not previously used to delineate ecological units, MRTs have been successfully used to explore relationships between environment and community patterns in many studies (e.g. Ruppert *et al.*, 2010; Williams *et al.*, 2010). Beyond the ease of generating ecological maps, MRTs have other advantages. The method does not require assumptions about the relationships between communities and their environments (e.g. linear versus nonlinear), accepts both numerical and categorical variables, handles missing data, is invariant to monotonic transformations of explanatory variables, is robust to collinear variables and handles interactions (De'ath, 2002). Each MRT is characterized by (1) its explanatory power as measured by the percentage of taxon variance explained, (2) its predictive power as estimated by a cross-validated relative error (CVRE) based on the division of the data into random test groups, and finally (3) its number of leaves (De'ath, 2002). For each biological dataset, the number of leaves of the MRT was selected via the '1-SE' rule

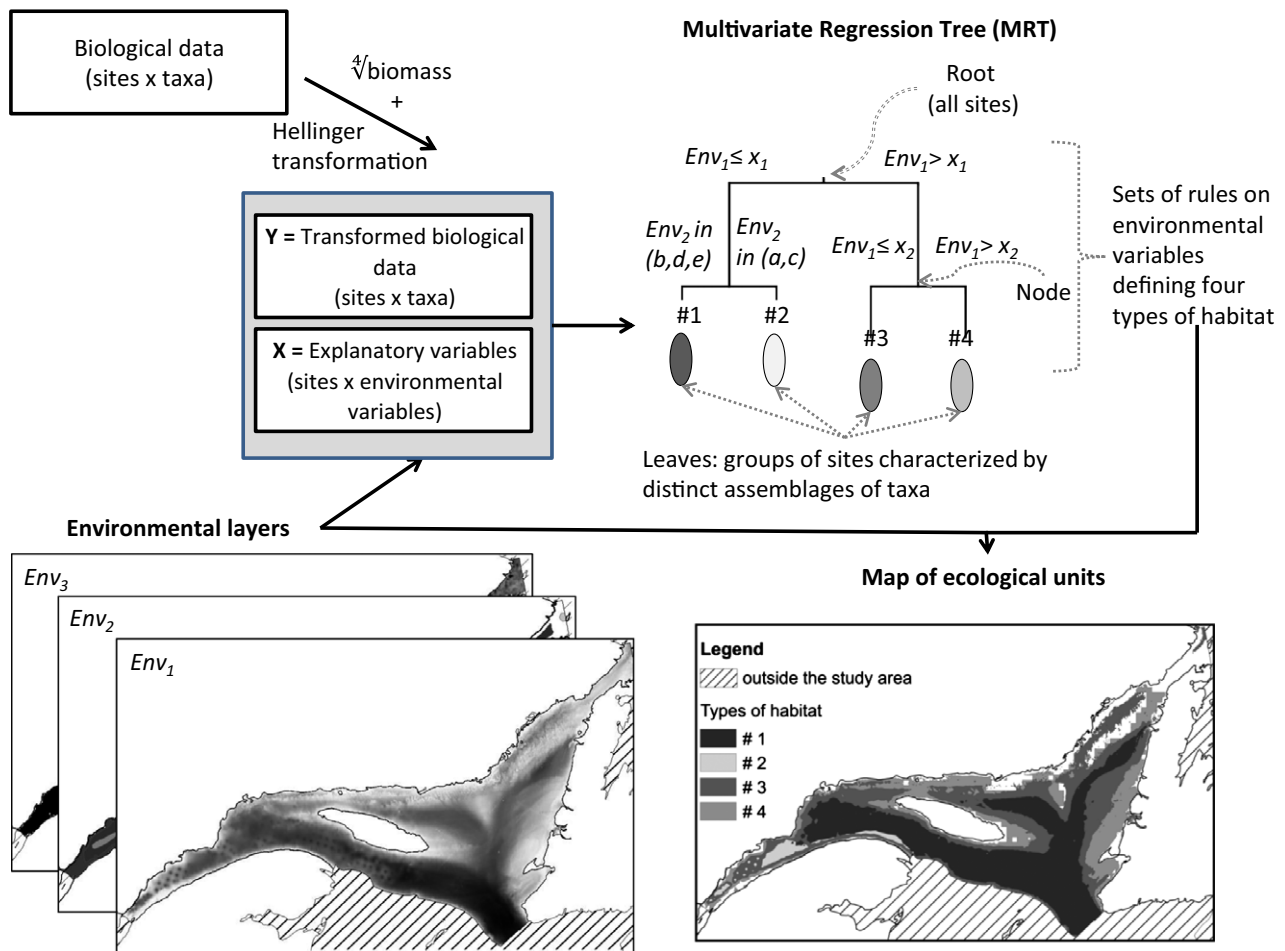


Figure 2 General framework used to delineate ecological units based on a simplified example of a multivariate regression tree (MRT) with two of three available environmental variables selected as split criteria by the algorithm. Quantitative variable Env_1 with values x_1 and x_2 acting as thresholds and qualitative variable Env_2 based on five states. Four leaves, representing four distinct assemblages of taxa, are defined by a set of rules based on environmental variables, which is then used within a geographic information system to delineate ecological units. Each ecological unit belongs to one of the four types of habitat designated by the number of the corresponding leaf in the MRT.

(Breiman *et al.*, 1984), which minimized the complexity of the tree while maintaining performance (akin to model parsimony). Cross-validation was repeated 500 times for successive and independent divisions of the sampling sites into ten test groups.

Transforming biological datasets

Taxa that occurred in only one sample were removed from the dataset. The resulting database included 234 taxa, representing 11 different phyla (Table 1). Invertebrates and vertebrates were represented by 162 and 72 taxa, respectively (110 and 60 identified to the species level). Biomass data were fourth-root transformed (Clarke & Green, 1988) prior to the analyses to reduce the contribution of abundant species to the assessment of site similarity. To circumvent the problem associated with the use of Euclidean distances on zero-inflated distribution matrices, data were Hellinger-transformed (Legendre & Gallagher, 2001) prior to MRT analysis.

Generating biological matrices

To compare the effect of taxonomic group on the ecological mapping, six biological matrices were constructed (Table 1): Arthropoda, Echinodermata, Mollusca, All Invertebrates, Vertebrata and All Taxa. The effect of taxonomic levels was assessed exclusively on the vertebrate dataset, which had the highest taxonomic resolution (species level for 60 of 72 taxa). From the detailed 'Species level' matrix (561 sites \times 60 species), four additional matrices were created by aggregating data at coarser taxonomic levels: genus (561 sites \times 48 genera), family (561 sites \times 20 families), order (561 sites \times 9 orders) and class (561 sites \times 3 classes).

Comparing ecological maps

Comparing ecological maps generated from different datasets required the selection of a specific dataset to be used as a reference in pairwise comparisons. In the case of comparison

Table 1 Distribution of taxa among the taxonomic groups and number of sites where the group was present.

Taxonomic group	Number of taxa	Number of phyla	Number of sites
Arthropoda	37	1	561
Echinodermata	31	1	482
Mollusca	41	1	545
All Invertebrates	162	11	561
Vertebrata	72	1	561
All Taxa	234	11	561

across taxonomic groups, the biological dataset that generated the best model was selected as the reference dataset for comparing maps. In the case of taxonomic levels, the ecological map generated with the vertebrate dataset at the finest taxonomic resolution (i.e. species level) was used as the reference to assess the influence of coarser taxonomic resolutions on ecological mapping.

However, two difficulties arose when comparing ecological maps. First, there was no *a priori* equivalence between their classification legends because ecological units of each map could represent habitats encompassing different communities (e.g. Fig. 3a, b). Consequently, a 'common classification legend' was used, showing correspondences between classes of the two initial classification legends (Fig. 3c). Second, the number of habitat types in the common classification could not exceed that of the least complex ecological map. Thus, to maximize the chances to find correspondences between classes, MRTs were reconstructed by fixing the number of leaves of the trees to be equal to that of the MRT of the reference dataset. Five hundred iterations were run for a 10-fold cross-validation.

For each pairwise comparison, the common classification legend was found by building a contingency table showing sampling sites allocated to the leaves of the two trees (Legendre & Legendre, 1998). The significance of the association was tested using Pearson's chi-square statistical test, and residuals were visualized using association plots (Fig. 4). When one leaf from each tree sharing a number of sites was significantly higher than expected by chance ($\alpha = 0.05$), it was considered to be a one-to-one relationship between the corresponding habitats of each map (Fig. 4, case A), and we used the same label to designate them in the common classification legend. When several leaves on one tree shared a significant number of sites with a single leaf of the other tree (i.e. 'one-to-many relationship'), we combined the corresponding habitats to define a common habitat in the new classification legend (Fig. 4, case B), a procedure that is commonly performed to increase the agreement among maps (Jung *et al.*, 2006). However, to remain meaningful, habitats were only combined if they shared common biological and environmental features, that is they corresponded to leaves that shared a common node. Thus, the structure of each tree was used as a decision tool for grouping of leaves (Fig. 4), and when leaves could not be grouped (e.g. found on

different nodes), only the one with the strongest relationship was used (Fig. 4, case C where Leaf 5 of 'All Invertebrates' is associated with Leaf 4 of 'Vertebrates', but Leaf 7 is not). The resulting 'correspondence table' for each pairwise comparison showed the grouping of habitats that were considered as equivalent for each map as well as the new labels used to designate these common habitats (Fig. 3c).

The ecological maps being compared (Fig. 3a, b) were then 'reconstructed' using the new common classification legend (Fig. 3c). Each ecological unit was reclassified with the label of the corresponding common habitat (Fig. 3d, e). Visualization of data and operations on data layers was performed using ArcGIS®.

Finally, the reconstructed maps were superimposed and a 'joint map' was produced to show the geographic overlap between ecological units with the same label in each map (Jung *et al.*, 2006). The resulting areas were called 'common ecological units' and identified according to the common classification legend (Fig. 3f). From an ecological point of view, these units represented geographical areas that were characterized by specific environmental features and distinct communities for both groups of taxa being compared. The overall similarity between the two ecological maps was then assessed in two ways. First, the number of common habitats was used as a measure of the 'thematic precision' of the joint map where grouping of initial habitats (e.g. leaves 12 and 13 of Vertebrata; Fig. 4) was considered to reduce resolution (i.e. information on the distribution of different communities was lost). Second, two measures were used to assess how well 'the maps agree in terms of the general location of each category'? (Pontius, 2002; p.1042). The 'overall spatial agreement' was measured using the percentage of the study area classified in the same way in the two reconstructed maps (see Foody, 2006). This overall measure corresponds to the proportion of the study area covered by common ecological units. At the individual habitat type level, a measure of the 'prevalence' (Hagen-Zanker, 2006) quantified the overlap between the reference map and the map to which it was being compared. For each common habitat, prevalence was calculated as the product of the relative degree of overlap between the two maps. Two ecological maps were thus considered to be more similar when the thematic precision was close to the number of habitats in the initial classification legends (i.e. little or no grouping of habitats was needed) and when the measures of spatial agreement, for the entire classification and for each common habitat, were high.

All data analyses were performed using the statistical software R v.2.15.3 (R Core Team, 2012), including the packages MVPART (De'ath, 2012) and MVPARTwrap (Ouellette, 2011) for MRT analysis and VCD for contingency tables analysis (Meyer *et al.*, 2012).

RESULTS

Comparing ecological maps across taxonomic groups

The nine-leaf MRT based on the 'Vertebrata' dataset emerged as the model explaining the highest variance (49%)

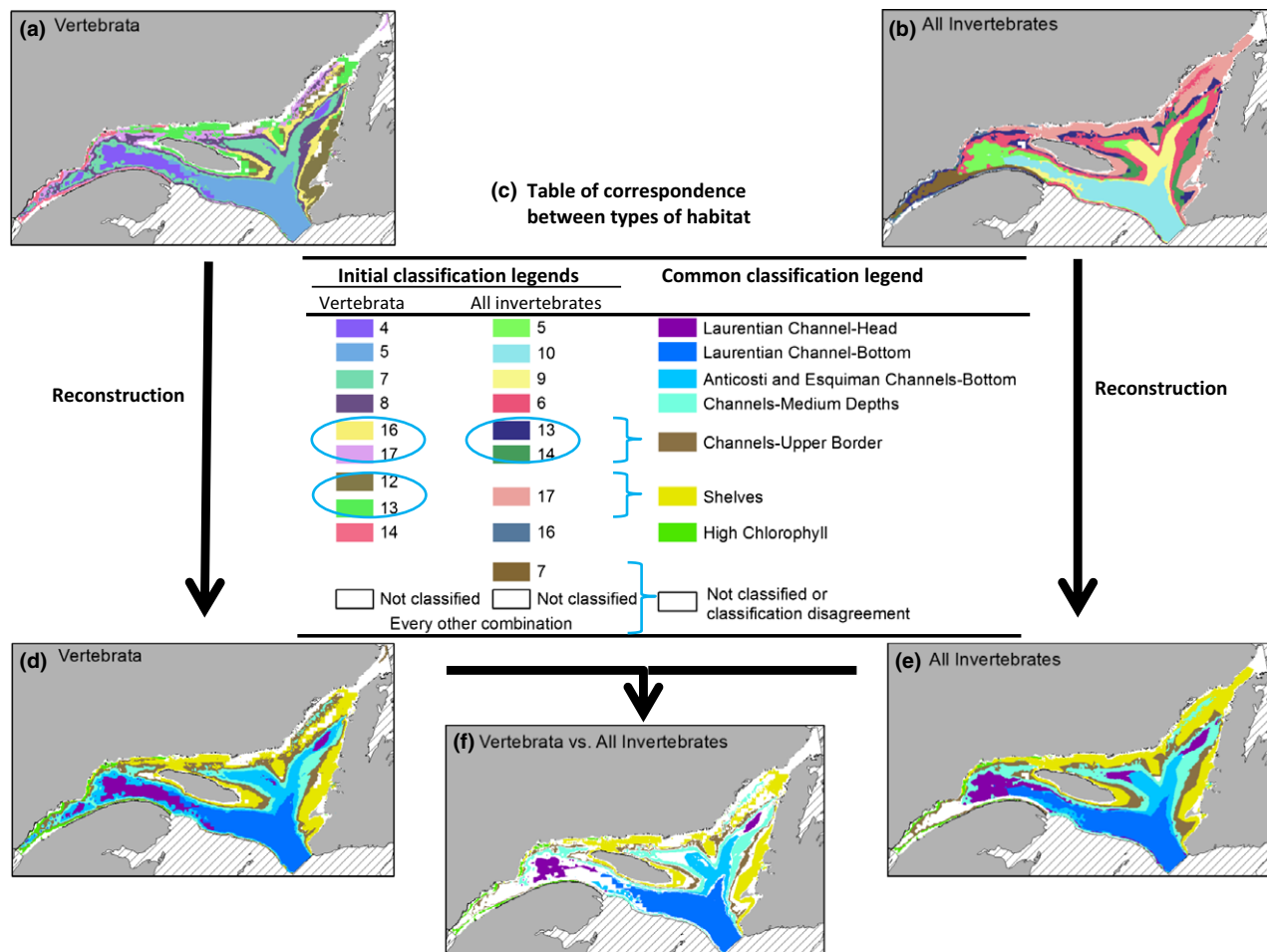


Figure 3 The three steps of the general framework for comparing ecological maps using those established from nine-leaf MRTs based on 'Vertebrata' and 'All Invertebrates' datasets (a,b) as an example. Types of habitat for each taxonomic group are identified in the initial classification legends in the table of correspondence (c), which is then used to reconstruct the two maps (d, e) following the common classification legend. Blue circles in the table indicate habitats that are grouped in the common classification legend. Finally, a joint map 'Vertebrata versus All Invertebrates' (f) shows the common ecological units according to the common classification legend.

and with the lowest CVRE (0.58), followed closely by 'Arthropoda' (45%, 0.61), 'All Taxa' (43%, 0.63), 'All Invertebrates' (34%, 0.72), 'Echinodermata' (27%, 0.75) and finally 'Mollusca' MRTs (7%, 0.93) (Table 2). After adjusting all trees to a common number of nine leaves, the 'Vertebrata' MRT remained the best model, followed by the 'Arthropoda' MRT. The 'Mollusca' MRT remained as having the worst overall performance with only 25% of the variance explained and a CVRE of 0.90 (Table 2). Eleven of the 33 available environmental metrics were retained and used as split criteria in the trees. These metrics (Tables 2 and S1) were related to bottom oxygen, bottom salinity, bottom temperature, chlorophyll *a*, depth and surficial sediment.

A visual comparison of ecological maps based on different MRTs was performed (see Fig. 3a, b for the 'Vertebrata', 'All Invertebrates' and Fig. 5a–c for 'Arthropoda', 'Mollusca' and 'All Taxa' ecological maps; see Fig. S1a for the 'Echinodermata' map). Channels deeper than 240 m (Fig. 1) appeared as persistent features in all ecological maps. Other spatial patterns were also visually identified, most notably the suc-

cession of ecological units from the deepest part of the Esquiman and Anticosti channels (Fig. 1) to the coastal shelf. The lower St. Lawrence Estuary also appeared as a distinct feature.

Pairwise comparisons against the reference 'Vertebrata' MRT revealed varying correspondence between the leaves of the two MRTs. For example, for the 'Vertebrata versus All Invertebrates' comparison (Fig. 3 and Fig. 4), there was a clear one-to-one correspondence between Leaf 5 of the 'Vertebrata' partition and Leaf 10 of the 'All Invertebrates' partition (Fig. 4, case A), which allowed the corresponding habitats to be designated as similar, that is 'Laurentian Channel-Bottom' (Fig. 3c). In contrast, leaves 12 and 13 of the 'Vertebrata' partition both showed a significant relationship with Leaf 17 of the 'All Invertebrates' partition (Fig. 4, case B – 'one-to-many'). The habitats corresponding to these leaves were combined and identified with the label 'Shelves' in the new classification legend (Fig. 3c). Another one-to-many relationship involved Leaf 4 of the 'Vertebrata' partition and leaves 5 and 7 of the 'All Invertebrates' partition.

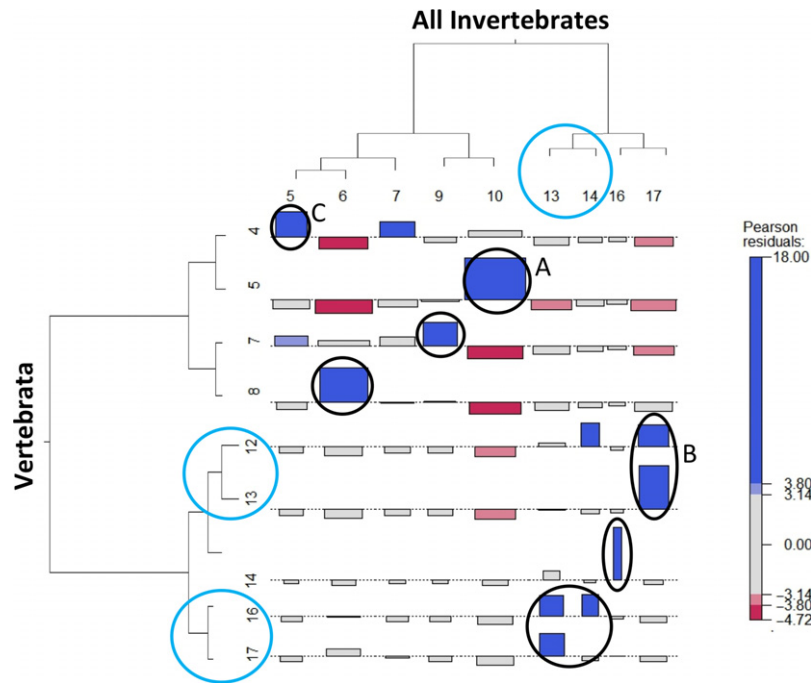


Figure 4 Correspondences between leaves of the 'Vertebrata' and 'All Invertebrates' MRT. For each MRT, the tree structure shows the root, branches and leaves, the latter being identified by numbers assigned automatically by the algorithm. In each cell of the association plot, Pearson's residual is indicated by a rectangle whose size is proportional to the deviation from an independent relationship and whose colour is determined by the probability level associated with a χ^2 test, that is full and light blue for positive relationships significant at $\alpha = 0.05$ and $\alpha = 0.1$, respectively, and negative or non-significant relationships in red and grey. Black circles indicate significant relationships used to give the equivalence of habitats in the corresponding maps. Letters refer to examples of the two different types of relationship: one-to-one (case A), one-to-many (cases B, C). Blue circles indicate habitats that are combined in the common classification legend.

However, because the latter two leaves were separated by more than one node (Fig. 4, case C), only the most significant relationship, that is between Leaf 4 and Leaf 5, was used to create the class 'Laurentian Channel-Head' in the common classification legend (Fig. 3c). Class 7 of the 'All Invertebrates' map was combined with other classification disagreements in the new classification legend (Fig. 3c). Finally, the 'Vertebrata' and 'All Invertebrates' maps were reclassified using the new classification legend (Fig. 5d–e). Common ecological units, that is the intersection between units in both maps designated with the same label, covered 61% of the study area. This overlap represented a substantial overall degree of agreement between the two maps, especially considering that the thematic precision of the new classification legend, that is seven common habitats, was close to the initial nine-leaf classification (Fig. 3f, Table 3). However, the degree of overlap varied greatly depending on specific habitats, with prevalence ranging from 21% for the 'Laurentian Channel – Head' habitat to 79% for 'Laurentian Channel – Bottom' habitat.

Common ecological units resulting from pairwise comparisons between the reference 'Vertebrata' map and the 'All Taxa' and 'Mollusca' maps showed the existence of strong common patterns in the distribution of the three groups of taxa (Fig. 6a, b), although the performance of the corre-

sponding models differed notably (Table 2). The highest similarity was found for the 'Vertebrata versus All Taxa' comparison, both in terms of thematic precision of the common classification legend (i.e. nine classes) and the overall spatial agreement (i.e. 75%) (Fig. 6a, Table 3). The pairwise comparison with the 'Mollusca' map gave exactly the same quantitative results as in the 'Vertebrata versus All Invertebrates' case, that is seven classes of common ecological units and an overall spatial agreement of 61% (Table 3). Visually, the corresponding joint maps were very similar (Fig. 3f and Fig. 6b). The worst level of similarity between ecological maps was seen in the 'Vertebrata versus Arthropoda' case, which still showed a thematic precision of six classes and an overall spatial agreement of 55% (Table 3, Fig. S2a). In all comparisons, the range of prevalence for the different common habitats was highly variable (Table 3), the highest score being obtained, in four comparisons of five, by the habitat corresponding broadly to the 'Laurentian Channel – Bottom' in the 'Vertebrata versus All Invertebrates' map.

Comparing ecological maps across taxonomic levels

The 'Species level' MRT showed characteristics similar to the general 'Vertebrata' MRT, that is 47% of the total variance explained and a CVRE of 0.58 (Table 4). A similar level of

Table 2 Multivariate regression trees (MRTs) based on different taxonomic groups for an optimal size (in bold) and for an imposed size of nine leaves (in *italics*). CVRE represents the cross-validated relative error. Environmental variables and metrics used as split criteria are indicated by a cross (x). See Supporting Information (Table S1) for a full description of environmental variables and metrics.

Taxonomic group used in the MRT	MRT characteristics		Environmental variables and metrics used as split criteria by the MRT algorithm										
	Number of leaves	Variance explained (%)	Bottom oxygen			Bottom salinity			Bottom temperature			Chlorophyll <i>a</i>	Sediment
			<i>O₂_mn_Stats</i>	<i>O₂_Survey</i>	<i>S_max_M</i>	<i>S_mn_M</i>	<i>S_n90_M</i>	<i>T_min_M</i>	<i>T_mn_M</i>	<i>Depth</i>	<i>Chl_mn</i>	<i>Depth</i>	<i>Sed_69_classes</i>
Arthropoda	6	45					x			x	x	x	x
	<i>9</i>	<i>49</i>					x			x	x	x	x
Echinodermata	4	27		x							x		
	<i>9</i>	<i>34</i>	x					x		x	x	x	x
Mollusca	2	7			x								
	<i>9</i>	<i>25</i>	x		x					x	x	x	x
All invertebrates	9	34									x	x	x
Vertebrata	9	49	x			x			x		x	x	x
All taxa	10	43									x	x	x
	<i>9</i>	<i>42</i>					x			x	x	x	x

performance for both criteria (i.e. 46% and 0.60, respectively) was obtained at the family level. At the genus level, the percentage of variance explained dropped to 42% while the CVRE rose to 0.64, the worst performance seen in this set of trees. The best models were the 'Order level' and 'Class level' MRTs, the last one reaching a percentage of 67% for the variance explained and a CVRE of 0.50. However, these relatively good performances must be balanced by the fact that the biological matrices modelled were far simpler at those taxonomic levels, with only nine orders and three classes, compared to the 60 taxa of the 'Species level' matrix. Bottom oxygen, bottom salinity, bottom temperature, depth, chlorophyll *a* and sediment were commonly selected as explanatory variables (Table 4). Ecological maps corresponding to the five taxonomic levels were quite similar visually and shared the same broad features as already mentioned in the comparison between different taxonomic groups (see Fig. 5d–f for the species, family and class levels and Fig. S1b–c in Supporting Information for genus and order levels). Spatial agreements of the ecological maps at the family, order and class levels with the 'Species level' ecological map were high, with an overall spatial agreement of 70 or 71% and a range of prevalence varying from 60% to 93% for the different habitats (Table 5; Fig. 6c, Fig. S2c, Fig. 6d for the respective joint maps). The habitat corresponding to the 'Laurentian Channel – Bottom' appeared again as the most spatially congruent among all comparisons. However, the thematic precision of the common classification legend was better at the family level (seven classes) than at order and class levels (five classes). The 'Family level' ecological map can then be considered as giving information closer to the reference map than the two other maps. Eight common classes were identified in the 'Species versus Genus level' comparison, but the overall spatial agreement between the two maps (63%) was lower compared to the other cases and the prevalence was also fairly low (34%) for one specific habitat (Table 5; Fig. S2b in Supporting Information).

DISCUSSION AND CONCLUSIONS

The goal of this paper was to assess the influence of selecting specific taxonomic groups and taxonomic resolution on the creation of ecological maps. Our results indicate that datasets of limited taxonomic breadth or limited taxonomic resolution can generate ecological maps that are very similar to those created by more extensive biological datasets.

Influence of taxonomic groups in ecological mapping

Among all pairwise comparisons, the 'Vertebrata versus All Taxa' comparison should be considered as a special case as the first dataset was a subset of the second. The high level of similarity found between the two maps and the strong performance of the 'Vertebrata' MRT suggest that this taxonomic group could be a good surrogate to represent benthic

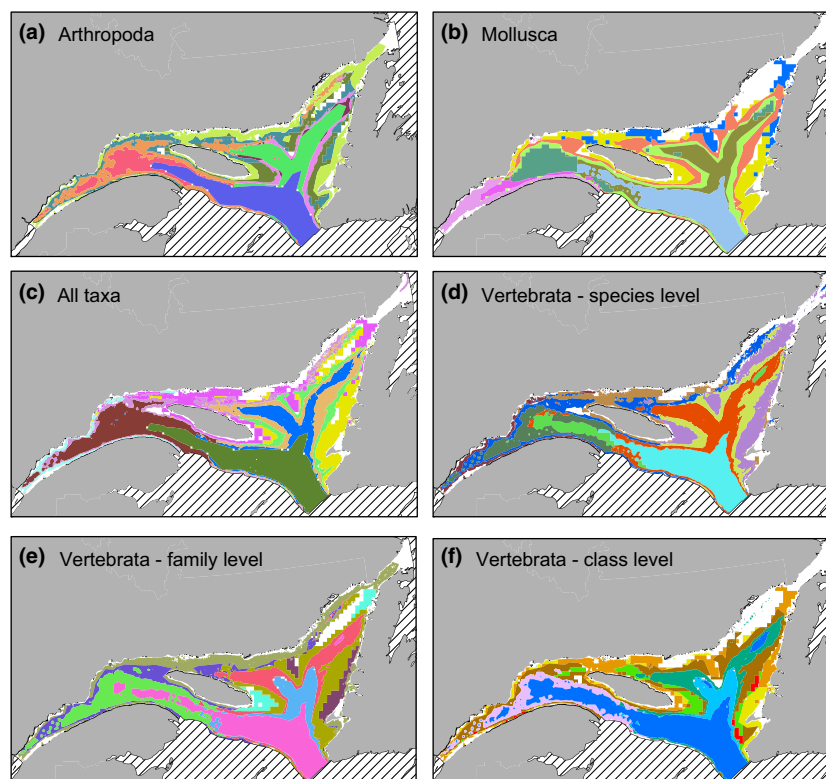


Figure 5 Ecological maps established from nine-leaf MRTs based on different biological datasets. Distinct habitats are identified by different colours. Areas not classified by the model appear in white. Maps (a) Arthropoda, (b) Mollusca and (c) All taxa are used for comparison across taxonomic groups. Maps (d) to (f) are used for comparison across taxonomic levels and are based on the 'Vertebrata' dataset reduced to taxa identified at the species level. Map (d) is based on records at the taxonomic level of species whereas maps (e) and (f) are based on datasets aggregated at family and class levels, respectively.

biodiversity in ecological mapping applications. However, some caution is needed as this high level of similarity and a prevalence of 100% found for five common habitats out of nine could also indicate that the vertebrates' group has a large influence on the dissimilarity measure calculated on the 'All Taxa' dataset, at least in our study. In contrast, for the other pairwise comparisons, biological datasets were mutually exclusive so that the high degree of correspondence found between the maps could be truly considered as an indication of redundancy of the biogeographic information contained in the two datasets. In comparison, previous studies that analysed biogeographic patterns for different taxonomic groups showed inconsistent results. On one hand, the similarity between the biodiversity patterns of invertebrate megabenthos and the limits of ecoregions derived from the distribution of fishes for the Australian margin (Williams *et al.*, 2010) is consistent with the present study ('Vertebrate' versus 'All Invertebrates' comparison). On the other hand, in contrast to the present study, those patterns were not necessarily preserved when each major taxonomic group was considered separately (e.g. Molluscs, Demospongiae). Likewise, dissimilarity of distribution patterns was seen among different invertebrate groups (i.e. Arthropoda, Mollusca, Annelida, Bryozoa) and the entire invertebrate assemblage (Anderson *et al.*, 2005). Even when a specific taxonomic group (e.g. polychaetes) was determined to be a good surrogate for assessing natural benthic distributional patterns (Włodarska-Kowalczyk & Kedra, 2007), the taxonomic dominance of the group was considered to have strongly influenced the pat-

terns observed for the entire benthic community. Our findings contrast with most previous work as they clearly indicate that different taxonomic groups can lead to spatially similar ecological units. Thus, these groups can be used individually in ecological mapping applications to represent benthic marine biodiversity when more comprehensive biological datasets are not available. This conclusion, thought to be true for demersal fishes (Commonwealth of Australia, 2005; Briggs & Bowen, 2012), has not, however, been well documented until now.

Our findings could simplify substantially the logistics in ecological mapping studies by focusing on specific taxonomic groups. Given the relatively good knowledge of fish distributions via commercial catches and scientific surveys, this group should be a good candidate in many ecological mapping applications. Further, it would be interesting to determine whether a fish dataset limited to a few species could still be robust in delineating meaningful ecological units. The work completed by Reygondeau *et al.* (2012) using catch rates on a limited number of large fish species suggests that this is a possibility, at least for the pelagic realm. However, it should be noted that the Gulf of St Lawrence is a highly stratified ecosystem (Koutitonsky & Bugden, 1991), which may have accentuated the differences among habitats and thus favoured the identification of common patterns between different taxonomic groups. Thus, caution should be used before applying these results indiscriminately to other environments, and other studies are needed to confirm or nuance these conclusions.

Influence of taxonomic resolution in ecological mapping

The ecological rationale for taxonomic sufficiency resides in the idea that phylogeny can reflect morphological similarities between taxa that share functional traits and thus can be influenced in the same way by environmental drivers (Roy *et al.*, 1996; Anderson *et al.*, 2005; Włodarska-Kowalczyk & Kedra, 2007). The use of coarser taxonomic resolution

Table 3 Measures of the degree of agreement between ecological maps based on different taxonomic groups. The 'Vertebrata' ecological map was used as a reference for each pairwise comparison. Thematic precision corresponds to the number of habitats in the common classification legend. The overall spatial agreement represents the intersection between units in both maps identified with the same label in the common classification legend and is measured by the proportion of the study area covered by common ecological units. For each common habitat, the prevalence expresses the degree to which the habitat in one map overlaps with the same habitat in the other map.

Taxonomic groups used in the comparison of ecological maps	Thematic precision	Spatial agreement (in %)	
		Overall	Range of prevalence for the different common habitats
Vertebrata versus Arthropoda	6	55	17–79
Vertebrata versus Echinodermata	6	66	32–92
Vertebrata versus Mollusca	7	61	51–89
Vertebrata versus All Invertebrates	7	61	21–79
Vertebrata versus All Taxa	9	75	40–100

presents obvious advantages such as reducing costs related to species identification. Identification at coarser levels provides savings that can be estimated by taking the complement of the ratio of the number of coarser taxa to the number of species (Ferraro & Cole, 1995). In the case of vertebrates, identification at the genus, family, order and class level could reduce costs on the order of 20%, 67%, 85% and 95%, respectively. Moreover, the use of coarser taxonomic resolution can also reduce the risk of potential identification errors (Włodarska-Kowalczyk & Kedra, 2007).

The similarities found between maps based on data aggregated at different taxonomic levels confirm previous work showing the concept of taxonomic sufficiency could be applied to both studies of natural benthic variability and pollution studies. In most earlier studies, a taxonomic resolution at the family level was sufficient to describe natural distribution patterns for different taxonomic groups (Roy *et al.*, 1996; Pitcher *et al.*, 2002; De Biasi *et al.*, 2003; Linse *et al.*, 2006; Joydas & Damodaran, 2013). This finding is consistent with the high degree of agreement found here between ecological maps at species and family levels. However, our study also showed a good correspondence at coarser taxonomic levels, that is order and class, although with some loss of thematic precision in the common classification legend. Whereas the use of coarse taxonomic levels has been relatively well studied in the context of perturbation detection (e.g. Olsfard *et al.*, 1998), very few studies have addressed this issue with regard to natural distribution patterns. At least two other studies based on different ecosystems (kelp holdfast communities (Anderson *et al.*, 2005) and benthic macrofauna of an Arctic fjord (Włodarska-Kowalczyk & Kedra, 2007), established that different taxonomic levels up to the order level provided similar levels of distinctiveness in their distributional patterns, becoming less distinct at the class and phylum levels. Our study showed that even at the class level, for which our vertebrate dataset contained only three classes, we were still able to detect the main ecological units in a relatively diverse ecosystem. To our knowledge,

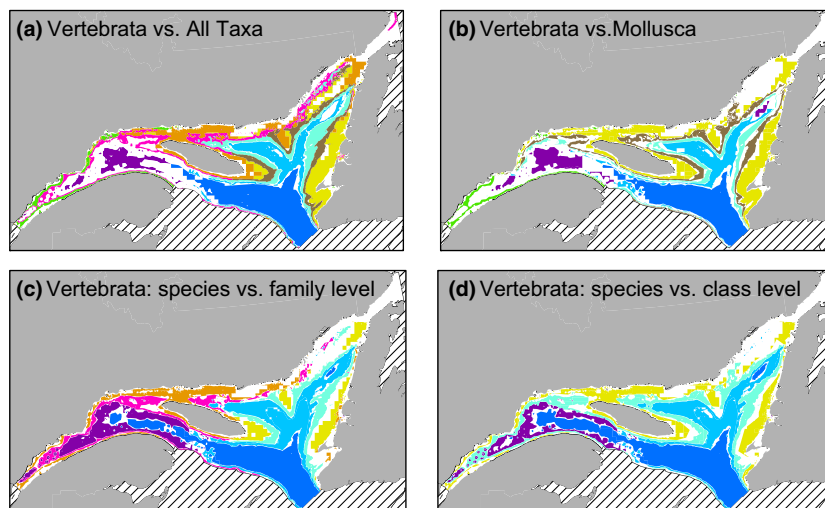


Figure 6 Joint maps showing common ecological units for the comparison of four pairs of ecological maps. Distinct common habitats are identified by different colours. Areas not classified or showing classification disagreement appear in white. For comparison across taxonomic groups, 'Vertebrata' ecological map is compared to (a) 'All Taxa' and (b) 'Mollusca' ecological maps. For comparison across taxonomic levels, the ecological map based on the 'Vertebrata' dataset at the species level is compared to the map created from the same dataset aggregated at (c) family and (d) class levels.

Table 4 Multivariate regression trees (MRTs) based on taxa of the 'Vertebrata' dataset aggregated at different taxonomic levels for a fixed size of nine leaves. CVRE represents the cross-validated relative error. Environmental variables and metrics used as split criteria are indicated by a cross (x). See Supporting Information (Table S1) for a full description of environmental variables and metrics.

MRT characteristics		Environmental variables and metrics used as split criteria									
Taxonomic level used in the MRT	Variance explained (%)	CVRE	Bottom oxygen <i>O₂_mn_Stats</i>	Bottom salinity			Bottom temperature		Chlorophyll <i>a</i> <i>Chl_mn</i>	Depth <i>Depth</i>	Sediment <i>Sed_69_classes</i>
				<i>S_min_M</i>	<i>S_mn_M</i>	<i>S_n90_M</i>	<i>T_min_Stats</i>	<i>T_n90_M</i>			
Species	47	0.58	x						x	x	
Genus	42	0.64	x			x			x	x	x
Family	46	0.60		x			x	x	x	x	x
Order	55	0.54			x			x	x	x	x
Class	67	0.50			x			x	x	x	x

Table 5 Measures of the degrees of agreement between ecological maps based on species of the 'Vertebrata' dataset aggregated at different taxonomic levels. The 'Species level' ecological map is used as a reference for each pairwise comparison. Thematic precision corresponds to the number of habitats in the common classification legend. The overall spatial agreement represents the intersection between units in both maps identified with the same label in the common classification legend and is measured by the proportion of the study area covered by common ecological units. For each common habitat, the prevalence expresses the degree to which the habitat in one map overlaps with the same habitat in the other map.

Taxonomic levels used in the comparison of ecological maps	Thematic precision	Spatial agreement (in %)	
		Overall	Range of prevalence for the different common habitats
Species level versus Genus level	8	63	34–89
Species level versus Family level	7	71	60–93
Species level versus Order level	5	71	68–93
Species level versus Class level	5	70	68–90

this is the first successful use of such coarse taxonomic levels in marine ecological mapping applications.

Methodological considerations

The approach used in this study for assessing the congruence of distributional patterns between datasets was innovative compared to earlier works. Previous studies examined the degree of agreement of different biological datasets with previous biogeographic classification using ordination methods, for example non-metric multivariate dimensional scaling

(nMDS) or statistical tests such as analyses of similarities (ANOSIM). By contrast, the use of MRT analysis allowed us to spatially delineate ecological units by taking into account only the portion of biological dissimilarities that could be explained by environmental drivers. The comparisons between different biological datasets were thus less influenced by compositional variability. This characteristic can also help explain how a rather poor model like the one based on 'Mollusca' allowed delineating ecological units fairly similar to those of the 'Vertebrata'. Using a 10-fold cross-validation based on 500 bootstrap resamplings, the predicted values used to produce the mapping of ecological units closely followed the random forest framework (Thomson *et al.*, 2014). This framework is robust against the weakness of the standard MRT in which the selection of variables used as split criteria may influence and may be influenced by the entrance order of the other variables (Ellis *et al.*, 2012; Loh & Zheng, 2013). Therefore, maps based on the cross-validated predicted values are quite robust against potential biases in variable selection.

Another novel aspect in our approach is the ability to create joint maps that allow patterns common to two groups of taxa to be visualized. Because common ecological units are based on separate determination for each group, they provide a double 'biological legitimacy' in terms of being supported by independent datasets. A high percentage of prevalence in most of the comparisons for a given habitat, for example the 'Laurentian Channel – bottom', is an indication that the delineation of that habitat is meaningful for the taxonomic groups being compared. However, a low percentage of prevalence is an indication that the limits of this habitat are not clear and may be sensitive to specific taxonomic group used in the analysis. Zones of disagreement between the two maps, indicated in white on a joint map, should thus be viewed as transition zones, implying that boundaries between ecological units are likely gradual as suggested previously (Day & Roff, 2000; Cameron & Askew, 2011). For each comparison, common habitats with the highest degree of

prevalence indicate that the corresponding ecological units are the most spatially congruent. Finally, we hope that spatial-explicit approaches such as the ones presented in this paper can help better understand how biological datasets characteristics such as taxonomic breadth or resolution influence the creation of ecological maps. Such an understanding should lead to a more rigorous use of ecological maps that are increasingly used to support diverse ocean management decisions.

ACKNOWLEDGEMENTS

Thanks to D. Bernier (DFO-Quebec) for providing the annual bottom trawl surveys datasets and to M. Lévesque and her colleagues (ISMER) for having improved the identification of benthic invertebrates from the DFO surveys. Thanks to S. Senneville and S. St-Onge Drouin (ISMER) for the calculations of oceanographic parameters using the 3D physical model of the Gulf of St Lawrence (Saucier *et al.*, 2003), J.-D. Dutil, S. Proulx and their colleagues (DFO-Quebec) for the data compilation provided in Dutil *et al.* (2011) and C. Caverhill (BIO/DFO-Atlantic) for providing composite images of chlorophyll *a* concentrations. Finally, thanks to Kyle Chapman for proofing the manuscript. This project is a contribution to the research programme of Québec-Océan.

REFERENCES

- Anderson, M.J., Diebel, C.E., Blom, W.M. & Landers, T.J. (2005) Consistency and variation in kelp holdfast assemblages: Spatial patterns of biodiversity for the major phyla at different taxonomic resolutions. *Journal of Experimental Marine Biology and Ecology*, **320**, 35–56.
- Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.G. (1984) *Classification and regression trees*. Wadsworth International Group, Belmont, California.
- Briggs, J.C. & Bowen, B.W. (2012) A realignment of marine biogeographic provinces with particular reference to fish distributions. *Journal of Biogeography*, **39**, 12–30.
- Cameron, A. & Askew, N. (eds.) (2011) EUSeaMap - Preparatory Action for development and assessment of a European broad-scale seabed habitat map final report. Available at: <http://jncc.gov.uk/euseamap>. (accessed 13 August 2015).
- Clarke, K.R. & Green, R.H. (1988) Statistical design and analysis for a biological effects study. *Marine Ecology Progress Series*, **46**, 213–226.
- Clarke, K.R. & Warwick, R.M. (1998) Quantifying structural redundancy in ecological communities. *Oecologia*, **113**, 278–289.
- Cogan, C.B., Todd, B.J., Lawton, P. & Noji, T.T. (2009) The role of marine habitat mapping in ecosystem-based management. *ICES Journal of Marine Science*, **66**, 2033–2042.
- Commonwealth of Australia (2005) Benthic Marine Bioregionalisation of Australia's Exclusive Economic Zone. A Report to the National Oceans Office on the Development of a National Benthic Marine Bioregionalisation in support of Regional Marine Planning. Geoscience Australia, Canberra. Available at: <http://www.environment.gov.au/node/18076#benthic>. (accessed 13 August 2015).
- Connor, D.W., Allen, J.H., Golding, N., Howell, K.L., Lieberknecht, L.M., Northen, K.O. & Reker, J.B. (2004) The national marine habitat classification for Britain and Ireland. Version 04.05. JNCC, Peterborough ISBN 1 861 07561 8 (internet version). Available at: <http://jncc.defra.gov.uk/MarineHabitatClassification>. (accessed 13 August 2015).
- Costello, M.J. (2009) Distinguishing marine habitat classification concepts for ecological data management. *Marine Ecology Progress Series*, **397**, 253–268.
- Day, J.C. & Roff, J.C. (2000) *Planning for Representative Marine Protected Areas*. WWF-Canada, Toronto.
- De Biasi, A.M., Bianchi, C.N. & Morri, C. (2003) Analysis of macrobenthic communities at different taxonomic levels: an example from an estuarine environment in the Ligurian Sea (NW Mediterranean). *Estuarine, Coastal and Shelf Science*, **58**, 99–106.
- De'ath, G. (2002) Multivariate regression trees: a new technique for modeling species-environment relationships. *Ecology*, **83**, 1105–1117.
- De'ath, G. (2012). mvpart: Multivariate partitioning. R package version 1.6-0. Available at: <https://cran.r-project.org/src/contrib/Archive/mvpart/>. (accessed 13 August 2015).
- Degraer, S., Verfaillie, E., Willems, W., Adriaens, E., Vincx, M. & Van Lancker, V. (2008) Habitat suitability modelling as a mapping tool for macrobenthic communities: an example from the Belgian part of the North Sea. *Continental Shelf Research*, **28**, 369–379.
- Dufour, R. & Ouellet, P. (2007) Estuary and Gulf of St Lawrence marine ecosystem overview and assessment report. *Canadian Technical Report of Fisheries and Aquatic Sciences* 2744E.
- Dutil, J.-D., Nozères, C., Scallion-Chouinard, P.-M., Van Guelpen, L., Bernier, D., Proulx, S., Miller, R. & Savenkoff, C. (2009) Poissons connus et méconnus des fonds marins du Saint-Laurent. *Naturaliste Canadien*, **133**, 70–82.
- Dutil, J.-D., Proulx, S., Chouinard, P.-M. & Borcard, D. (2011) A Hierarchical Classification of the Seabed Based on Physiographic and Oceanographic Features in the St Lawrence. *Canadian Technical Report of Fisheries and Aquatic Sciences*, 2916.
- Ellis, D. (1985) Taxonomic sufficiency in pollution assessment. *Marine Pollution Bulletin*, **16**, 459.
- Ellis, N., Smith, S.J. & Pitcher, C.R. (2012) Gradient forests: calculating importance gradients on physical predictors. *Ecology*, **93**, 156–168.
- Ferraro, S.P. & Cole, F.A. (1995) Taxonomic level sufficient for assessing pollution impacts on the Southern California Bight macrobenthos- Revisited. *Environmental Toxicology and Chemistry*, **14**, 1031–1040.
- Foody, G.M. (2006) What is the difference between two maps? A remote sensor's view. *Journal of Geographical Systems*, **8**, 119–130.

- Gray, J.S., Aschan, M., Carr, M.R., Clarke, K.R., Green, R.H., Pearson, T.H., Rosenberg, R. & Warwick, R.M. (1988) Analysis of community attributes of the benthic macrofauna of Frierfjord-Langesundfjord and in a mesocosm experiment. *Marine Ecology Progress Series*, **46**, 151–165.
- Gregg, W.W. & Casey, N.W. (2007) Sampling biases in MODIS and SeaWiFS ocean chlorophyll data. *Remote Sensing of Environment*, **111**, 25–35.
- Hagen-Zanker, A. (2006) Map comparison methods that simultaneously address overlap and structure. *Journal of Geographical Systems*, **8**, 165–185.
- Joydas, T.V. & Damodaran, R. (2013) Testing depth-related multivariate patterns of macrofauna on the Indian continental shelf using reduced taxonomic resolution and data transformation. *Journal of the Marine Biological Association of the United Kingdom*, **93**, 37–45.
- Jung, M., Henkel, K., Herold, M. & Churkina, G. (2006) Exploiting synergies of global land cover products for carbon cycle modeling. *Remote Sensing of Environment*, **101**, 534–553.
- Kenchington, R. & Hutchings, P. (2012) Science, biodiversity and Australian management of marine ecosystems. *Ocean & Coastal Management*, **69**, 194–199.
- Koutitonsky, V.G. & Bugden, G.L. (1991) The physical oceanography of the Gulf of St. Lawrence: a review with emphasis on the synoptic variability of the motion. *Le Golfe du Saint-Laurent: petit océan ou grand estuaire?* (ed. by J.-C. Theriault), pp. 57–90. Canadian Special Publication of Fisheries and Aquatic Sciences, 113, Mont-Joli.
- Labruno, C., Gremare, A., Amouroux, J.M., Sarda, R., Gil, J. & Taboada, S. (2008) Structure and diversity of shallow soft-bottom benthic macrofauna in the Gulf of Lions (NW Mediterranean). *Helgoland Marine Research*, **62**, 201–214.
- Legendre, P. & Gallagher, E.D. (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia*, **129**, 271–280.
- Legendre, P. & Legendre, L. (1998) *Numerical Ecology*, 2nd, English edn. Elsevier, Amsterdam, New York.
- Linse, K., Griffiths, H.J., Barnes, D.K.A. & Clarke, A. (2006) Biodiversity and biogeography of Antarctic and sub-Antarctic mollusca. *Deep-Sea Research Part II-Topical Studies in Oceanography*, **53**, 985–1008.
- Loh, W.Y. & Zheng, W. (2013) Regression trees for longitudinal and multiresponse data. *Annals of Applied Statistics*, **7**, 495–522.
- Mackey, B.G., Berry, S.L. & Brown, T. (2008) Reconciling approaches to biogeographical regionalization: a systematic and generic framework examined with a case study of the Australian continent. *Journal of Biogeography*, **35**, 213–229.
- McArthur, M.A., Brooke, B.P., Przeslawski, R., Ryan, D.A., Lucieer, V.L., Nichol, S., McCallum, A.W., Mellin, C., Cresswell, I.D. & Radke, L.C. (2010) On the use of abiotic surrogates to describe marine benthic biodiversity. *Estuarine Coastal and Shelf Science*, **88**, 21–32.
- McBreen, F., Askew, N., Cameron, A., Connor, D., Ellwood, H. & Carter, A. (2011) UK SeaMap 2010. Predictive mapping of seabed habitats in UK waters. JNCC Report 446. Available at: <http://jncc.defra.gov.uk/page-5955>. (accessed 13 August 2015).
- Meyer, D., Zeileis, A., Hornik, K. & Friendly, M. (2012) Package 'vcd'. R package version 1.2-13. Available at <http://CRAN.R-project.org/package=vcd>. (accessed 13 August 2015).
- Nozères, C., Archambault, D., Chouinard, P.-M., Gauthier, J., Miller, R., Parent, E., Schwab, P., Savard, L. & Dutil, J.-D. (2010) Identification guide for marine fishes of the estuary and northern Gulf of St. Lawrence and sampling protocols used during trawl surveys between 2004 and 2008. Canadian Technical Report of Fisheries and Aquatic Sciences, 2866.
- Olsgard, F., Somerfield, P.J. & Carr, M.R. (1998) Relationships between taxonomic resolution, macrobenthic community patterns and disturbance. *Marine Ecology Progress Series*, **172**, 25–36.
- Olsgard, F., Brattegard, T. & Holthe, T. (2003) Polychaetes as surrogates for marine biodiversity: lower taxonomic resolution and indicator groups. *Biodiversity and Conservation*, **12**, 1033–1049.
- Ouellette M.H. with contributions from P. Legendre (2011) MVPARTwrap: Additional functionalities for package mvpart. R package version 0.1-9. <https://cran.r-project.org/src/contrib/Archive/MVPARTwrap/>. (accessed 13 August 2015).
- Pitcher, R., Venables, W., Ellis, N., McLeod, I., Pantus, F., Austin, M., Cappo, M., Doherty, P. & Gribble, N. (2002) Great Barrier Reef seabed biodiversity mapping project: Phase 1. CSIRO/AIMS/QDPI. Report to CRC-Reef. CSIRO Marine Research, Brisbane.
- Pitcher, C.R., Doherty, P., Arnold, P. *et al.* (2007) Seabed Biodiversity on the Continental Shelf of the Great Barrier Reef World Heritage Area. AIMS/CSIRO/QM/QDPI CRC Reef Research Task Final Report, Cleveland. Available at: <http://epubs.aims.gov.au/handle/11068/7785?show=full>. (accessed 13 August 2015).
- Pontius, R.G. (2002) Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. *Photogrammetric Engineering and Remote Sensing*, **68**, 1041–1049.
- R Core Team (2012). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>
- Reygondeau, G., Maury, O., Beaugrand, G., Fromentin, J.M., Fonteneau, A. & Cury, P. (2012) Biogeography of tuna and billfish communities. *Journal of Biogeography*, **39**, 114–129.
- Roy, K., Jablonski, D. & Valentine, J.W. (1996) Higher taxa in biodiversity studies: patterns from eastern Pacific marine molluscs. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **351**, 1605–1613.
- Ruppert, J.L.W., Fortin, M.J., Rose, G.A. & Devillers, R. (2010) Environmental mediation of Atlantic cod on fish community composition: an application of multivariate regression tree analysis to exploited marine ecosystems. *Marine Ecology-Progress Series*, **411**, 189–201.
- Saucier, F.J., Roy, F., Gilbert, D., Pellerin, P. & Ritchie, H. (2003) Modeling the formation and circulation processes

- of water masses and sea ice in the Gulf of St Lawrence, Canada. *Journal of Geophysical Research-Oceans*, **108**, 1–20.
- Snelder, T., Leathwick, J., Dey, K., Weatherhead, M., Fenwick, G., Francis, M., Gorman, R., Grieve, J., Hadfield, M., Hewitt, J., Hume, T., Richardson, K., Rowden, A., Uddstrom, M., Wild, M. & Zeldis, J. (2005) The New Zealand Marine Environment Classification. Ministry for the Environment, Wellington. Available at: <https://www.mfe.govt.nz/sites/default/files/marine-environment-classification-jun05.pdf>. (accessed 13 August 2015).
- Snelder, T., Leathwick, J., Dey, K., Rowden, A., Weatherhead, M., Fenwick, G., Francis, M., Gorman, R., Grieve, J., Hadfield, M., Hewitt, J., Richardson, K., Uddstrom, M. & Zeldis, J. (2007) Development of an ecologic marine classification in the New Zealand region. *Environmental Management*, **39**, 12–29.
- Spalding, M.D., Agostini, V.N., Rice, J. & Grant, S.M. (2012) Pelagic provinces of the world: a biogeographic classification of the world's surface pelagic waters. *Ocean & Coastal Management*, **60**, 19–30.
- Thomson, R.J., Hill, N.A., Leaper, R., Ellis, N., Pitcher, C.R., Barrett, N.S. & Edgar, G.J. (2014) Congruence in demersal fish, macroinvertebrate, and macroalgal community turnover on shallow temperate reefs. *Ecological Applications*, **24**, 287–299.
- UNESCO (2009) Global Open Oceans and Deep Seabed (GOODS) biogeographic classification. United Nations Educational, Scientific and Cultural Organization Intergovernmental Oceanographic Commission. IOC Technical Series, 84, Paris.
- Warwick, R.M. (1993) Environmental-impact studies on marine communities - Pragmatical considerations. *Australian Journal of Ecology*, **18**, 63–80.
- Williams, A., Althaus, F., Dunstan, P.K., Poore, G.C.B., Bax, N.J., Kloser, R.J. & McEnnulty, F.R. (2010) Scales of habitat heterogeneity and megabenthos biodiversity on an extensive Australian continental margin (100–1100 m depths). *Marine Ecology- an Evolutionary Perspective*, **31**, 222–236.
- Witman, J.D. & Roy, K. (2009) *Marine macroecology*. University of Chicago Press, Chicago, Illinois.
- Włodarska-Kowalczyk, M. & Kedra, M. (2007) Surrogacy in natural patterns of benthic distribution and diversity: selected taxa versus lower taxonomic resolution. *Marine Ecology Progress Series*, **351**, 53–63.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1 Environmental variables and metrics.

Figure S1 Ecological maps (complement to Fig. 5).

Figure S2 Joint maps (complement to Fig. 6).

BIOSKETCH

Anne Fontaine works in environmental management for the French Ministry of Ecology. Her research interest lies in developing tools, especially maps, which help policymakers in various aspects of marine biodiversity assessment and management.

Editor: Enrico Di Minin