ICES Journal of
Marine Science

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

ICES Journal of Marine Science (2020), 77(2), 835–845. doi:10.1093/icesjms/fsz249

Original Article

Optimizing image-based protocol to monitor macroepibenthic communities colonizing artificial structures

Bastien Taormina [1,2]*, Martin P. Marzloff[2], Nicolas Desroy[3], Xavier Caisey[2], Olivier Dugornay[4], Emmanuelle Metral Thiesse[2], Aurélien Tancray[5], and Antoine Carlier[2]

[1]*France Energies Marines, 525 Avenue Alexis de Rochon, 29280 Plouzané, France*
[2]*Ifremer, Centre de Bretagne, DYNECO—Laboratoire d'écologie benthique, ZI de la Pointe du Diable, CS 10070, 29280 Plouzané, France*
[3]*Ifremer, Laboratoire Environnement Ressources Bretagne Nord, 38 rue du Port Blanc, 35801 Dinard, France*
[4]*Ifremer, Centre de Bretagne, Direction de la Communication—Pôle audiovisuel, ZI de la Pointe du Diable, CS 10070, 29280 Plouzané, France*
[5]*Ifremer, Centre de Bretagne, Laboratoire Comportement des Structures en Mer, ZI de la Pointe du Diable, CS 10070, 29280 Plouzané, France*

**Corresponding author: tel: (+33)2 98 49 98 55; e-mail: Bastien.Taormina@france-energies-marines.org.*

Taormina, B., Marzloff, M. P., Desroy, N., Caisey, X., Dugornay, O., Metral Thiesse, E., Tancray, A., and Carlier, A. Optimizing image-based protocol to monitor macroepibenthic communities colonizing artificial structures. – ICES Journal of Marine Science, 77: 835–845.

Received 18 June 2019; revised 21 November 2019; accepted 26 November 2019; advance access publication 3 January 2020.

Underwater imagery is increasingly used as an effective and repeatable method to monitor benthic ecosystems. Nevertheless, extracting ecologically relevant information from a large amount of raw images remains a time-consuming and somewhat laborious challenge. Thus, underwater imagery processing needs to strike a compromise between time-efficient image annotation and accuracy in quantifying benthic community composition. Designing and implementing robust image sampling and image annotation protocols are therefore critical to rationally address these trade-offs between ecological accuracy and processing time. The aim of this study was to develop and to optimize a reliable image scoring strategy based on the point count method using imagery data acquired on tide-swept macroepibenthic communities. Using a stepwise approach, we define an underwater imagery processing protocol that is effective in terms of (i) time allocated to overall image, (ii) reaching a satisfactory accuracy to estimate the occurrence of dominant benthic taxa, and (iii) adopting a sufficient taxonomic resolution to describe changes in community composition. We believe that our method is well adapted to investigate the composition of epibenthic communities on artificial reefs and can be useful in surveying colonization of other human structures (wind turbine foundations, pipelines, etc.) in coastal areas. Our strategy meets the increasing demand for inexpensive and time-effective tools for monitoring changes in benthic communities in a context of increasing coastal artificialization pressures.

Keywords: benthic monitoring, fouling community, sampling design, taxonomic resolution, underwater imagery.

Introduction

Coastal benthic ecosystems are increasingly impacted by a cocktail of anthropogenic pressures, including sea bottom fishing (trawling/dredging in particular), harbour development, tourism, industry, energy production, and urban coastal development (Halpern *et al.*, 2008). As a direct consequence, both quality and extent of vulnerable coastal habitats have declined worldwide (Jackson *et al.*, 2001; Lotze and Milewski, 2004; Lotze *et al.*, 2006; Le Pape *et al.*, 2007). In this context, there is an increasing demand for regular cost-effective monitoring of ecosystems ecological status. Underwater imagery has been increasingly used as an effective and repeatable method to monitor benthic ecosystems, for several reasons. First, the collection of large amounts of high-resolution information on benthic biodiversity is rapid; second, the method is non-invasive, which is a key for long-term monitoring of selected sites (no or limited perturbation of ecological communities); and third, cameras operated by scuba divers or underwater vehicles provide access to remote sites (for instance due to depth or seafloor topography) that are difficult to sample with classic methods. Consequently, underwater imagery is widely used to describe a diverse range of coastal benthic habitats, such as tropical coral reefs (Brown *et al.*, 2004; Lam *et al.*, 2006; Dumas *et al.*, 2009; Molloy *et al.*, 2013; Jokiel *et al.*, 2015), algal assemblages (Preskitt *et al.*, 2004; Vroom and Timmers, 2009; Deter *et al.*, 2012; Berov *et al.*, 2016), rocky substrates (Macedo *et al.*, 2006; Van Rein *et al.*, 2011), artificial reefs (Page *et al.*, 2006; Walker *et al.*, 2007; Jerabek *et al.*, 2016; Jimenez *et al.*, 2017), highly hydrodynamic sites (Foveau *et al.*, 2017; O'Carroll *et al.*, 2017), and mesophotic or deep-sea ecosystems (Sen *et al.*, 2016; Domke *et al.*, 2017; Marzloff *et al.*, 2018).

While underwater imagery produces large amounts of raw data of seafloor communities, the extraction of ecologically relevant information through taxonomic identification to species level is often challenging, sometimes impossible without collected specimens, expert knowledge, or extensive taxonomy literature (Althaus *et al.*, 2015). So, benthic ecologists have developed classification methods adapted to assess benthic biodiversity solely from imagery. Such classifications are often region specific and inconsistent as they may use different terminologies to label a given category of organism (Schlacher *et al.*, 2010; Harrison and Smith, 2012; Oh *et al.*, 2015). In response to these inconsistencies across worldwide image-based benthic surveys, Althaus *et al.* (2015) developed a standardized classification for identifying benthic categories from underwater imagery called Collaborative and Automated Tools for Analysis of Marine Imagery (CATAMI), which aims to facilitate image annotation, data management, and data sharing.

However, even with the appropriate classification, the extraction of relevant information (taxon occurrence, count of individuals or colonies, size or cover estimation, etc.) from the entire raw images relies on laborious and time-consuming analysis (Pech *et al.*, 2004; Preskitt *et al.*, 2004; Nakajima *et al.*, 2010). For instance, concerning benthic sessile communities on hard substrates, the challenge lies in quantifying the occurrence or percentage cover of each taxon on each image to describe the community composition. Image scoring can rely either on labelling each conspicuous organisms on the picture (to estimate presence or abundance) or on exhaustively delineating their shape (to estimate percentage cover). However, this method is highly time-consuming so it cannot be applied to a large set of images or to diverse encrusting communities. The "point count" approach provides a reliable time-effective alternative to this comprehensive image analysis (Pielou, 1974). It consists in distributing a certain number of points on an image and then visually labelling the benthic category (taxa or substratum type) lying under each point. Then, the community composition can be assessed by calculating the percentage cover of each category as the ratio between the number of points attributed to a target category and the total number of points, on a given sampled surface. This method was facilitated by the development of dedicated software, such as CPCe (Coral Point Count estimation; Kohler and Gill, 2006), PhotoQuad (Trygonis and Sini, 2012), or more recently BIIGLE (Langenkämper *et al.*, 2017). However, the accuracy of the percentage covers estimated with this method increases with the density of points scored and depends also on the method used to project points on the image. So, the optimal point density strikes a compromise between the desired accuracy level and the time needed for image processing. It also depends on the seafloor area sampled per image, as well as the size, relative occurrence, and distribution patterns of the targeted taxa (Pante and Dustan, 2012; Perkins *et al.*, 2016).

Strategies to process underwater imagery are heterogeneous across published studies and rarely justified. However, a number of methodological choices determines the information extracted from image scoring (e.g. quadrat size, number of images scored, number of subsampling points per image). In most cases, these choices are based on empirical rules that depend on a number of study-specific factors (e.g. size of taxa, scope of the study, budget). For instance, the first choice to make, which is the definition of the quadrat size, is empirically determined to appropriately scale-match with targeted taxa: intuitively, studies on coral reef will use wider quadrat than those on biofouling. Nevertheless, a limited number of methodological works clearly rationalize image subsampling strategies (Dumas *et al.*, 2009; Deter *et al.*, 2012; Pante and Dustan, 2012; Berov *et al.*, 2016; Perkins *et al.*, 2016). Nevertheless, all of these mentioned studies focus on benthic organisms of sizes superior to 10 cm (i.e. megafauna/flora). Thus, to our knowledge, very few information (Sartoretto *et al.*, 2017) are readily available concerning optimal method when targeting low sized and/or encrusting macroepibenthic communities.

The aim of this study was to develop and optimize a protocol of underwater image analysis suitable for describing macroepibenthic communities colonizing natural and artificial hard substrates. Using a stepwise approach, we defined a reliable image scoring strategy using imagery data acquired on subtidal tide-swept encrusting benthic communities by optimizing (i) density of points, (ii) way of point projection, (iii) total sampling area, and (iv) taxonomic resolution (by testing the CATAMI classification).

Methods

Study site

The study site is a 15-km-long submarine power cable (8 MVA, 10 kVDC) set-up in 2012 to connect the tidal test site of Paimpol–Bréhat (Brittany, France; Figure 1) developed by Electricité de France Energies Nouvelles. Because of the seafloor characteristics (dominance of pebbles and rocks), 11 km of cable is unburied but fully protected with nested iron half-shells (50-cm long, 15-cm diameter). The cable is also stabilized by 120 concrete mattresses (6-m long, 3-m wide) installed in 2013 (Figure 2), which prevent its displacement due to high hydrodynamic site conditions (current speed up to 5 knots during spring tides). Due to several setbacks in the commissioning progress of the project, no electric current has transited through the cable so far and associated protection structures have actually acted as a simple artificial reef.

Image acquisitions

An underwater imagery benthic survey undertaken by divers was performed at three sites along the cable route: A, B, and C (Figure 1). The three sites present similar depths (between 18 and 20 m). At each site, high-definition photographs of benthic communities were taken by divers both on natural bottom and artificial habitats that protect the cable (iron half-shells for sites A, B, and C and concrete mattresses for sites B and C) with the following strategy:

(i) each side of each 50-cm-long iron half-shell on a 10-m transect using overlapping still imagery;

(ii) 16 regularly spaced concrete units (whether 47 cm × 38 cm or 47 cm × 20 cm) of the mattress; and

(iii) quadrat of 25 cm × 25 cm randomly placed on the natural habitat 10 m apart from the cable route.

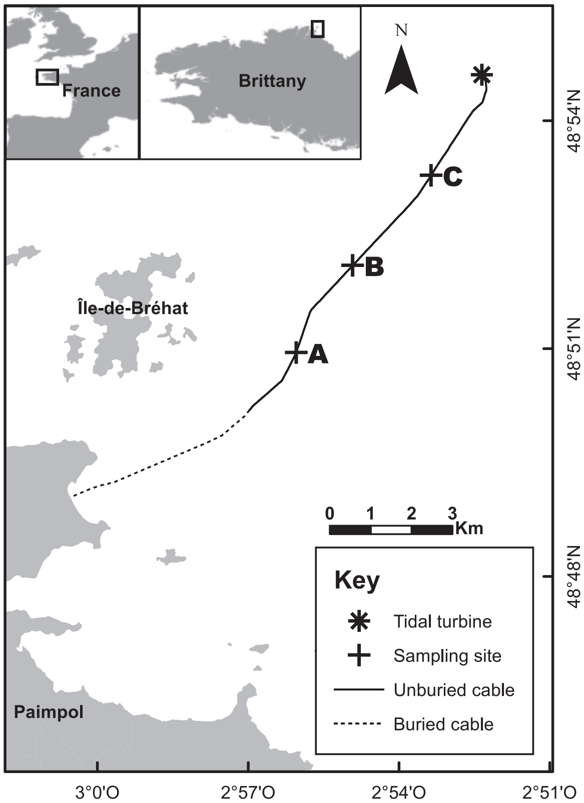

Figure 1. Map of the study area of the north coast of Brittany in western France (top-left and top-centre panels), which shows the location of the Paimpol–Bréhat tidal turbine test site where A, B, and C indicate the three study sites surveyed along the cable route (bottom).

Photographs were taken at a resolution of 37 million pixels per image using a Nikon D810 inside an Ikelite underwater housing, with a 20-mm lens and 2 Keldan LED lights (105 W, 9000 lumens). All images were calibrated with a scale bar.

The image-processing protocol optimization occurred at three different levels as illustrated in Figure 3: we first defined (i) the optimal point count method at the image level, then (ii) the sampling effort, i.e. the number of images, required at the site level, and finally (iii) the relevant taxonomic resolution.

Point count strategy at the image level

Briefly, we followed a three-step approach (detailed in the following sections) to define the optimal image scoring strategy, in terms of number of points and point projection method, by:

(i) describing exhaustively the benthic biodiversity on nine "reference" images (three for each type of habitat);

(ii) using these nine "reference" images, assessing how the point sampling designs (point density combined with projection method) impact the estimation of benthic biodiversity;

(iii) based on the obtained relationships, identifying the optimal density of point and projection method.

Exhaustive analysis of "reference" images

We selected three images representative of the complexity of the benthic community (in terms of diversity and spatial heterogeneity) for each habitat (half-shell, mattress, and natural bottom). On these nine "reference" pictures, an area equivalent to 625 cm^2 was cropped for analysis. Using ArcGIS, all benthic categories (being either taxa or substrates) visible in this area were manually cut out and annotated after visual identification (at the lowest possible taxonomic level for biological categories). The comprehensive scoring of each reference image took between 14 and 21 h. This first step resulted in nine raster files that provided a comprehensive description of benthic biodiversity and for which each pixel was assigned to a benthic category (Figure 4b).

Point count simulations

Then, we tested how a range of point count image-scoring strategies effectively reflects the true benthic community composition. These point sampling strategies were generated by combining 100 different point densities (from 5 to 500 points per 625 cm^2 image area, by the increments of 5 points) and two different projection methods (random and stratified random; Figure 4c). For each of the nine "reference" images, 1000 random simulations were performed for each combination, giving a total of 200000 simulations. For each simulation, we computed the percentage cover of each benthic category. All the simulations were performed with RStudio (v 1.0.0143) using the SpCosa package to implement stratified-random sampling (Walvoort *et al.*, 2010).

Selection of the optimal method

Our aim was to achieve an optimal scoring method that would enable us to estimate the occurrence of benthic categories with a percentage cover superior or equal to 5% and an accuracy corresponding to a coefficient of variation (*C.V.*) of the estimated occurrence $\leq$0.25. This threshold was chosen because it has been shown that the point count method is generally not suitable to accurately characterize benthic categories with a percentage cover inferior to 5% (Dumas *et al.*, 2009; Deter *et al.*, 2012; Perkins *et al.*, 2016).

To assess the accuracy of alternative point sampling strategies, we computed the *C.V.* of the estimation of percentage cover computed for each category across 1000 random simulations:

$$C.V.\ (i, n, m) = \frac{\sigma(i, n, m)}{\bar{X}(i, n, m)}, \tag{1}$$

where i is the ith benthic category, n is the number of points scored ($5 \leq n \leq 500$ by interval of 5), m is the the projection method (random or stratified random), $\bar{X}$ (i, n, m) is the mean percentage cover of category i across 1000 simulations under a given method; σ (i, n, m) is the standard deviation of the percentage cover of category i across 1000 simulations under a given method. The *C.V.* constitutes a good proxy of accuracy in percentage estimates across repeated measures (the higher the *C.V.*, the lower the accuracy).

We used a nonlinear model (function *nls* of the R package *stats*) using Rstudio (RStudio Team, 2015; v 1.0.0143) to characterize the number of points required to reach a *C.V.* of 0.25 for taxa that exhibit a range of percentage cover (represented in bold white line in Figure 5). The black dotted line highlights the specific case of benthic categories associated with a 5% cover. For each habitat (natural bottom, iron half-shell, concrete mattress) and projection method, we identified the minimum number of points required to achieve a *C.V.* of $\leq$0.25 for benthic categories

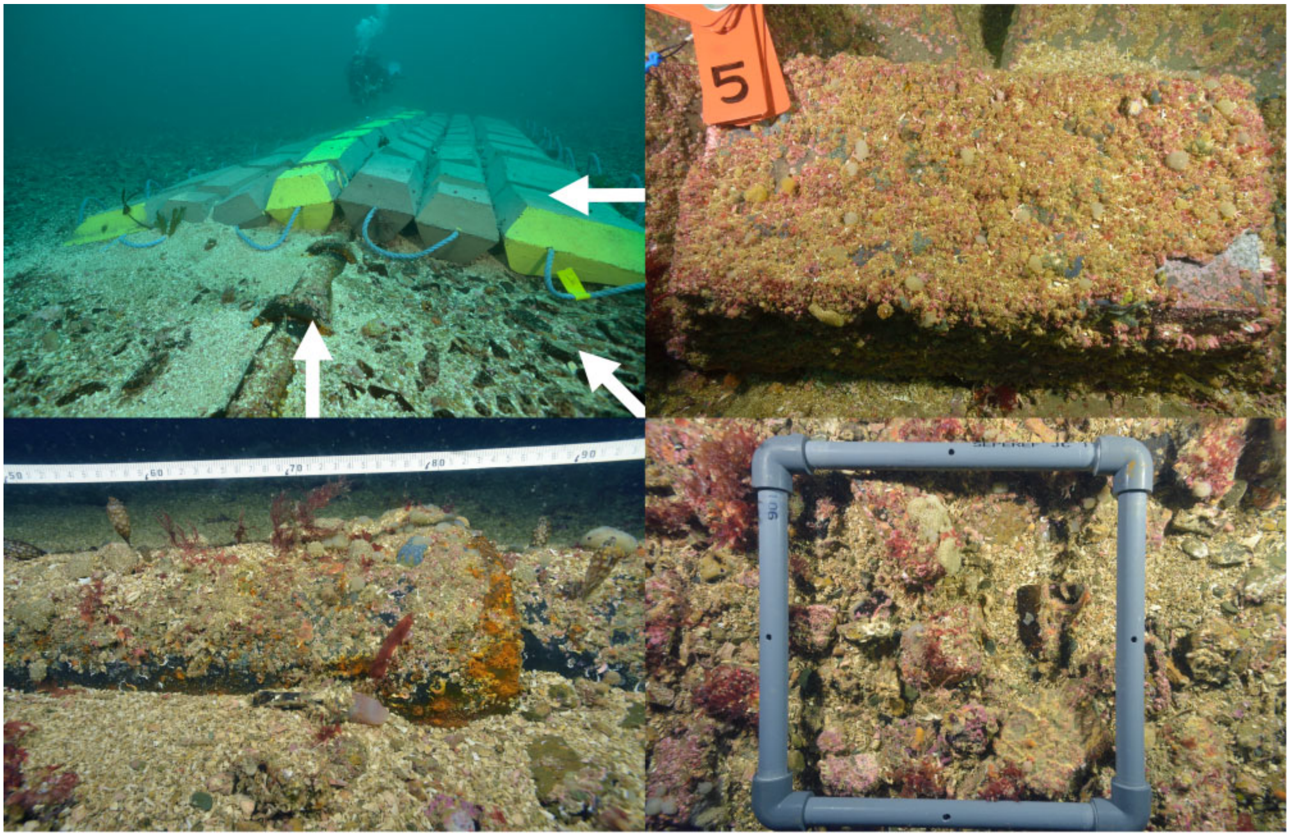

Figure 2. Overall view of one of the survey sites including cast-iron half-shells, a concrete mattress (freshly installed). and natural habitat (top left). Close-up views of one of the mattresses concrete units (top right), one cast-iron half-shell (bottom left), and one of the quadrats placed on the natural habitat (bottom right) (courtesy: Olivier Dugornay).

with a 5% cover (which corresponds to our accuracy threshold). Based on these *C.V.* estimates, we identified an optimal strategy across all habitats, in terms of minimum number of points and projection method.

Sampling effort at the site level

Once the optimal point count strategy is adopted to efficiently capture benthic community composition within an image (which could be considered as a replicate), the second step was to determine the most relevant sampling area, i.e. the total area observed at the site level for a given habitat (defined as number of images $\times$ quadrat size).

To assess this optimum sampling area, we first applied the optimal point count method (defined in the previous part) to all the analysable images from a large dataset (110 images) obtained at a single site and date. These image analyses were performed using the free software PhotoQuad (Trygonis and Sini, 2012). A benthic category was assigned to each projected point, and the percentage cover was estimated for each encountered category. The biological categories were determined at the lowest possible taxonomic level (i.e. species when possible). For natural bottom and concrete mattresses, 55 and 21 photos of 625 cm^2 were analysed, respectively, and 34 photos of 400 cm^2 were analysed for iron half-shells. For the rest of the procedure, only the biological categories were considered to focus on the composition of the benthic communities.

Then, we used Monte-Carlo simulations to construct curves of taxonomic similarity–area for each type of habitats, a straightforward approach to determine adequate sampling size (Weinberg, 1978; Kronberg, 1987; Schmera and Eros, 2006). For a given sampling area (n images), two independent sets of n images were randomly chosen from the total data set. Bray–Curtis similarity indices were calculated to compare the diversity sampled in each of these two sets. This process was repeated 1000 times for each level of sampling area. We then produced habitat-specific (i.e. natural bottom, mattress, iron half-shell) similarity–sampling area curves using the package CommEcol (Schneck and Melo, 2010) in RStudio (v 1.0.0143) by plotting mean estimates of Bray–Curtis similarity for each level of sampling effort. The non-linear relationship between similarity and the sampling area was modelled using the function *nls* of the R package *stats*. We defined the optimum sampling area as the number of survey images associated with the asymptotic point of the similarity–sampling area curve, i.e. when increasing sample number only marginally increases between-sample similarity (by <0.1%).

Taxonomic resolution

The CATAMI classification developed for underwater image analysis, combines a coarse-level taxonomy and the integration of organism morphology for the identification of benthic taxa (Althaus *et al*., 2015). We tested this classification frame by examining how it affects diversity patterns obtained with the finest taxonomic frame that we could provide.

We used the same data set (110 images Site B, September 2015) that served to determine the optimum sampling area at the site level. All the taxa identified at the lowest taxonomic level (LTL) were also labelled using the CATAMI classification at its highest resolution. Thus, we produced two alternative community datasets, corresponding to these two different taxonomic resolutions: the LTL and the CATAMI resolution. As an example, the ascidian species *Styela clava*, which is easily recognizable from imagery, can be identified by experts as (i) "Styela clava" using the LTL classification while it will be scored as (ii) "Solitary stalked Ascidian" (which encompasses a number of species) using the CATAMI classification. Resemblance matrices were computed for both taxonomic resolutions by calculating Bray–Curtis similarities between samples. The two similarity matrices were visually compared by computing two non-metric multi-dimensional scaling (nMDS) ordinations with Rstudio (v 1.0.0143). Potential correlation between the LTL and the CATAMI matrices was examined using Spearman's rank correlation coefficient, and the significance of the relationship was determined with the Monte-Carlo permutation routine RELATE of the PRIMER programme (Clarke and Warwick, 2001).

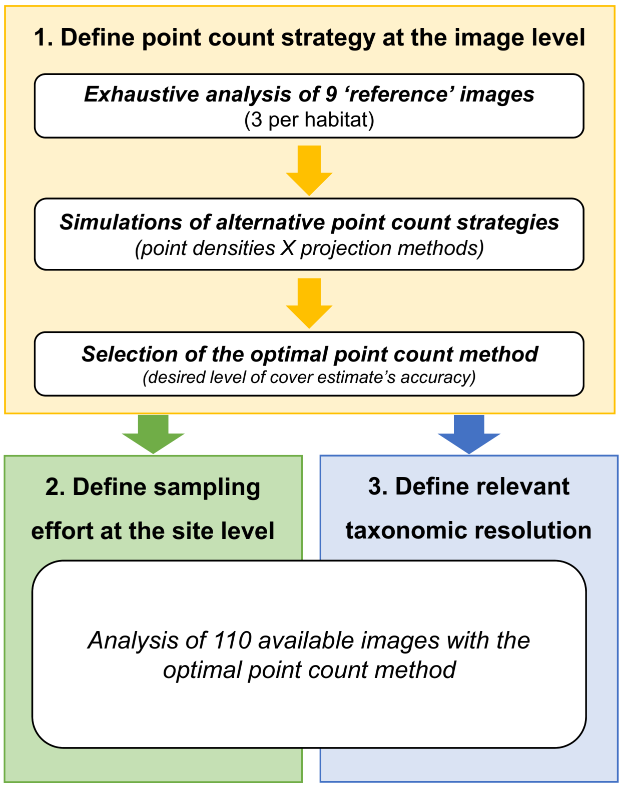

Figure 3. Flowchart describing the stepwise approach used to optimize the method of underwater imagery processing for accurately monitoring changes in epibenthic biodiversity on coastal artificial structures.

Results

Point count optimization at the image level

Figure 5 presents the aggregated results across all the point densities simulated (from 5 to 500 points per image) to determine the scoring effort required per image to reach a satisfactory accuracy for each habitat type (i.e. natural bottom, mattress, half-shell) and each type of point projection (random or stratified random). Across all simulations, the *C.V.* of the estimated percentage cover of taxa decreases rapidly as the number of points and/or the occurrence of the benthic categories increase. This reflects that percentage cover estimates are more accurate for a high density of point and/or for more abundant benthic categories (common taxon). For instance, across all investigated habitat and projection methods, ~50 point scores per image are sufficient to achieve a *C.V.* of ≤0.25 for abundant taxa with percentage cover >20%. For a given point score strategy (point density and projection method), the accuracy of percentage cover estimate varies according to the habitat considered, in particular for taxa with percentage cover <10%. To reach a *C.V.* value of 0.25 for categories with percentage cover ~5%, 322, 345, and 342 randomly projected points per image are needed, for half-shell, mattress, and natural bottom, respectively (Table 1). When using stratified-random projection, the number of points needed dropped to 199, 248, and 211 per image. Beyond that, improving the accuracy of percentage cover estimates of 5% cover categories is costly in terms of scoring effort since ~50 and 300% extra points are required to attain *C.V.* 0.2 and 0.1, respectively, compared to the number of

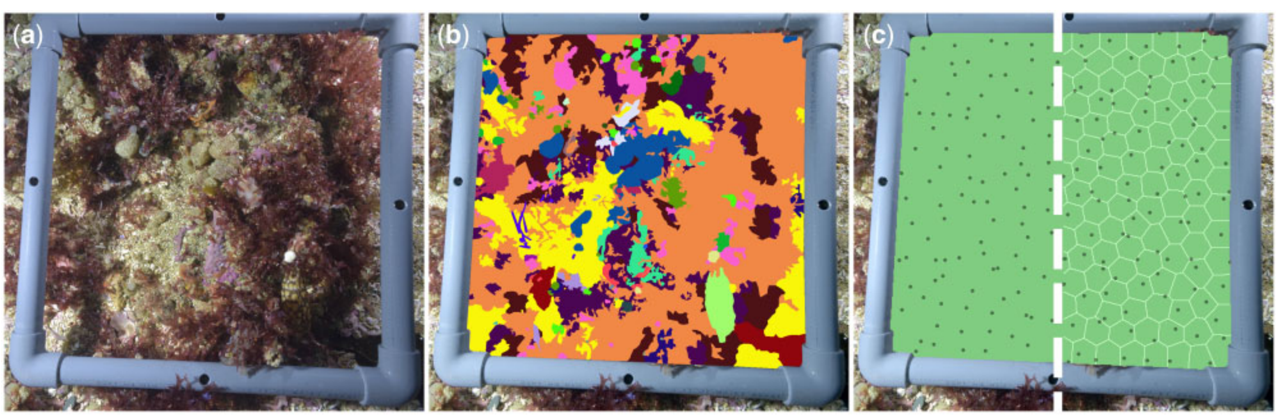

Figure 4. Illustration of image processing. (a) An example of 25 cm ×25 cm quadrat image of the natural bottom (site B, September 2017—courtesy: Olivier Dugornay). (b) Result of the exhaustive picture taxonomic analysis performed with ArcGIS, each colour corresponding to a different benthic category (i.e. substratum type or taxon). (c) Example of point count simulation with 200 points (i.e. 0.32 points per cm^2), using the random (left) or stratified-random (right) projection methods.

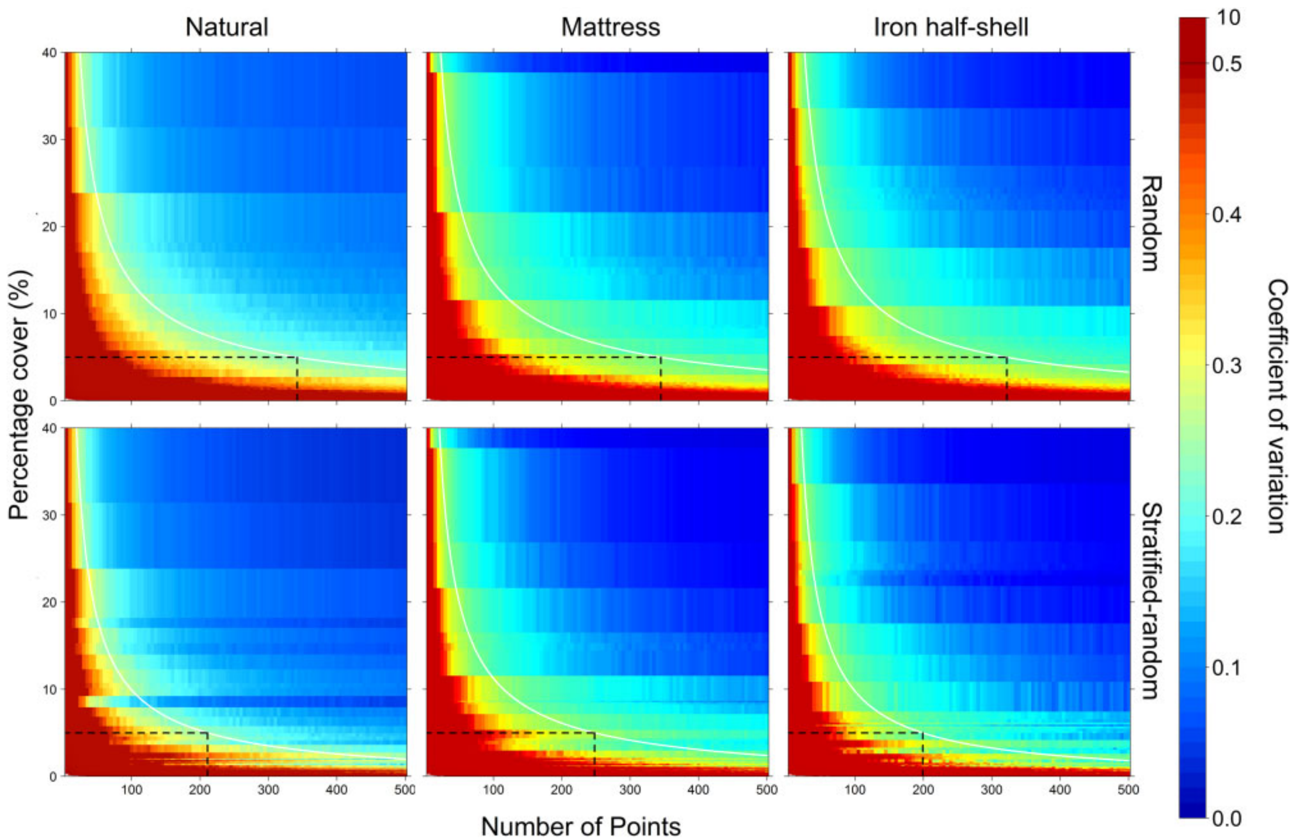

Figure 5. Change in *C.V.* of percentage cover estimates as a function of number of points scored per image (*x*-axis) and actual percentage cover of benthic categories (*y*-axis). The six panels correspond to the two different projection methods (i.e. random and stratified random) and the three different habitats (i.e. natural, mattress and half-shell). *C.V.*, represented by the colour scale, indicates the proportion of variation around mean cover estimates (the smaller the *C.V.*, the more accurate the estimate). The white thick line delineates the *C.V.* values of 0.25. The black dotted lines represent the intersection between benthic categories with a percentage cover of 5% and the number of points to obtain a *C.V.* value of 0.25. We defined the optimal number of points in each scenario as the intersect between these two lines.

points required to obtain a *C.V.* of 0.25 (Table 1). Consequently, the optimal method that fulfils our criteria (i.e. the *C.V.* of 0.25 for rare categories of 5% cover) requires 248 points per picture of 625 cm^2 (rounded to 250 points, i.e. 0.4 points per cm^2) using a stratified-random projection.

Sampling area at the site level

For the three investigated habitats, relationships between the taxonomic similarity between samples and the sampling effort (number of image scored) result in similar typical accumulation curves (Figure 6). The asymptote was reached slightly faster for half-shell than for mattress and natural bottom. According to our criteria (scoring an additional image represents a benefit as long as the similarity index is improved by >1%), the required sampling areas are 0.36 m^2 (corresponding to 9.05 pictures) for the half-shell, 0.55 m^2 (8.85 pictures) for the mattress and 0.52 m^2 (8.35 pictures) for the natural bottom (Table 2).

Fitting taxonomic resolution

The analysis of pictures taken at Site B in September 2015 using the LTL underlines 44 distinct biological categories across communities of natural bottom, mattress, and iron half-shell, mainly dominated by red algae (encrusting and foliose) and ascidians (solitary and colonial). nMDS analysis shows a clear taxonomic difference between the community settled on natural bottom and those developing on artificial (mattress and half-shell) habitats (Figure 7a). When using the CATAMI classification, the number of biological categories drops from 44 to 27 (a decrease of 39%). Despite this coarser taxonomic resolution, the corresponding nMDS (Figure 7b) shows a very similar pattern to the one obtained with the LTL classification. However, the visual comparison needs to be treated carefully considering the moderate stress values of the different nMDS representations. Spearman's correlation coefficient between the two patterns of taxonomic similarity is high ($\rho = 0.986$), and the permutation routine confirms this correlation as statistically significant ($P < 0.05$).

Discussion

Studies of benthic biodiversity based on underwater imagery are faced with a similar challenge: the need to strike a compromise between time-efficient imagery processing and extraction of ecologically relevant information (Van Rein *et al.*, 2009; Molloy *et al.*, 2013). Our stepwise optimization protocol provides an effective means to rationalize image-processing trade-offs in terms of (i) time allocated to images annotation, (ii) accuracy reached in percentage cover estimates of taxa, and (iii) taxonomic resolution. This method can easily be adapted to survey natural reefs or man-made structures in coastal areas by accounting for study-

Table 1. Number of points required to reach a *C.V.* of 0.1, 0.2, and 0.25 for 5% cover benthic categories, the two different projection methods and the three different habitats.

		Natural		Mattress		Iron half-shell	
Percentage cover	***C.V.***	**Stratified random**	**Random**	**Stratified random**	**Random**	**Stratified random**	**Random**
5	0.1	727	1 733	873	1 526	783	1 502
	0.2	290	529	351	517	288	490
	0.25	211	342	**248**	345	199	322

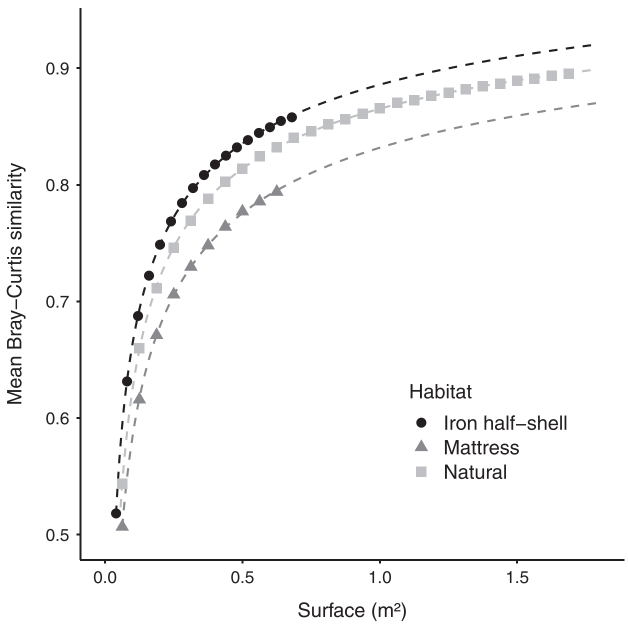

Figure 6. Evolution of the mean Bray–Curtis similarity between two equal subsamples (see "Methods" section) in function of the sampling area (m^2) for three different habitats.

Table 2. Number of pictures and corresponding sampling area required to reach the asymptotic point of the similarity–area curve for each habitat.

Habitat	Number of pictures	Area (m^2)
Natural	9.35	0.52
Mattress	8.85	0.55
Iron half-shell	9.05	0.36

specific features related to targeted communities. Our approach can be suitable for low subtidal and circalittoral hard bottom areas where benthic communities have no or low stratified physical structure without dense macroalgae development. In this sense, it may be particularly useful to monitor benthic colonization of offshore artificial structures (wind turbine foundations, pipelines, etc.).

An optimized imagery processing protocol to study macroepibenthic communities

The focus of the present study on the detection of fine-scale spatiotemporal changes in macroepibenthic communities colonizing artificial structures led us to define our optimum method following a stepwise approach. Indeed, given that most of the targeted organisms have small mean size (~10 mm), we first designed the way that images should be described to accurately estimate the cover of taxa at the finest spatial scale (i.e. image) and then we assessed the sampling effort required at the site scale to encompass the larger spatial heterogeneity of local benthic diversity. This stepwise approach can serve as a general guideline for other image-based benthic studies, even though other approaches can be considered. For instance, Perkins *et al.* (2016) simultaneously optimized the number of pictures per site and the density of points per image along transects, albeit *in silico* using a computer-generated data set. They showed that increasing the number of images more effectively increased precision than increasing the number of points. While it seems a key to apply on field images of benthic communities a similar optimization procedure that considers both parameters at the same time, it remains difficult to achieve in practice. Indeed, this would require to exhaustively describe a large set of images, which is time-consuming (~15 h per image in the present study). In our case, we privileged a high point density per image rather than the number of images per site because macroepibenthic sessile communities are typically characterized by a high number of rare taxa with a low patchiness (i.e. homogeneous repartition) and the dominance of small and encrusting individuals. Furthermore, studies on benthic colonization of artificial habitats, such as ours, generally examine differences in community composition between natural and artificial habitats, or between different artificial habitats, at local scale. Thus, they require detection of quantitative differences in taxonomic composition at fine scales (e.g. across locally heterogeneous substrates). To detect subtle changes in the occurrence of particular species within relatively homogeneous epibenthic communities, a high scoring effort is required at the image level by increasing point density, rather than at the site level.

Note that the level of accuracy required to tackle an ecological question impacts the design of the imagery processing protocol. In our case study, we optimized image scoring so as to reach a desired accuracy arbitrarily set as a cover estimate's *C.V.* of <0.25 for benthic categories with a percentage cover ≥5%. This threshold ought to be adjusted depending on the scope of the study. When the objective is only to detect large variations in benthic community composition (for instance over a large spatial scale), such a high accuracy in percentage cover estimates might only be required for most common benthic categories (e.g. with percentage covers ≥10 or 20%) and a lower point density than in the present study might then be optimal. Thus, we judge essential to explicitly define *a priori* (i.e. before designing and implementing the image scoring protocol) the degree of accuracy required to tackle the ecological question(s) at stake. When image scoring accuracy is not explicitly set at the onset of the study, it is critical

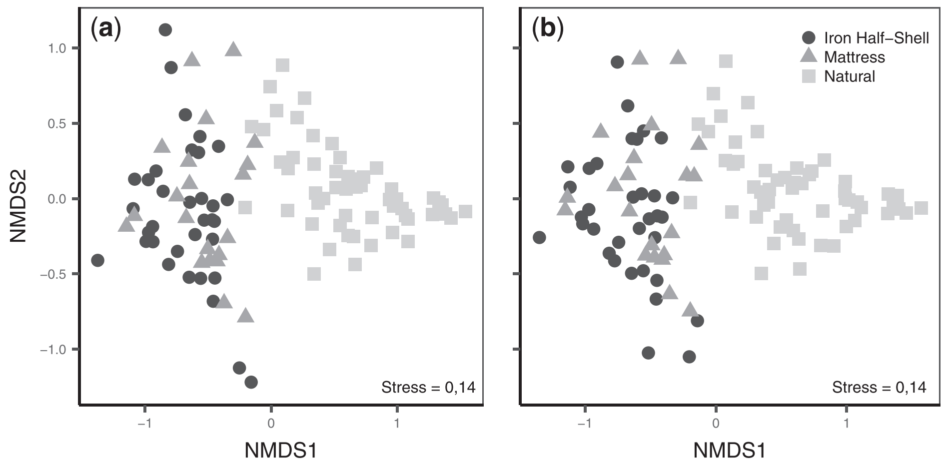

Figure 7. nMDS of Bray–Curtis similarities of benthic community composition from underwater images of site B in September 2015. Benthic organisms were described (a) at the lowest possible taxonomic level or, (b) using the coarser CATAMI classification. Each point represents a single picture.

to assess the quality and robustness of the biological information extracted from underwater imagery to avoid flawed ecological interpretations.

Accounting for spatial distribution of benthic taxa

Taxa are rarely uniformly distributed in nature (i.e. homogeneous distribution) and rather exhibit different degrees of clustering (i.e. heterogeneous distribution of the individuals). This can impact the effectiveness of spatially structured sampling methods (Cochran, 1946; Dutilleul, 1993; Legendre *et al.*, 2002; McGarvey *et al.*, 2016), such as the way count points are projected on the images. The literature shows that stratified-random sampling design performs better than random sampling design to estimate relative abundance of clustered taxa (i.e. higher accuracy in cover estimates in our case; Cochran, 1946; McGarvey *et al.*, 2016). When a community tends towards a homogeneous spatial distribution pattern, the different methods tend to perform equally. Consequently, whatever the spatial pattern of the community, stratified-random designs are always at least as accurate as the random sampling designs (Cochran, 1946). This point explains why in our study cases, the sampling effort (i.e. number of points) required with random projection was always higher than with the stratified-random projection to reach a similar precision. Nevertheless, although random designs give wider confidence interval of the percentage cover, these are unbiased, in the sense that they will not be impacted by spatial pattern of the taxa (McGarvey *et al.*, 2016). Thus, the absence of regularity in spatial distribution patterns of organisms has incited some authors to generalize the use of the random design at the expense of stratified random (Dethier *et al.*, 1993; McGarvey *et al.*, 2016).

In our study, we identified that spatial clustering of the surveyed taxa influenced the accuracy of our estimates at two different spatial scales, namely within images and across images at the site level. At the image scale, the stratified-random projection significantly reduces image-processing time as the number of points required to accurately estimate percentage cover is up to 38% smaller than with the random projection. Nevertheless, the optimal point density showed between-habitat differences that are more pronounced with the stratified-random projection than with the random projection. Since we determined the optimal number of points in a consistent way across habitats (i.e. to reach a satisfactory accuracy for "≥5% cover" categories), the fact that a given accuracy is reached with a higher point density on mattresses with respect to natural or half-shell habitats can only be explained by a difference in spatial patterns of these categories. Indeed, our exhaustive picture analyses (dedicated to image sampling strategy) showed that benthic categories with a cover between 5 and 10% occurred in more numerous and smaller patches on mattress habitat (17.9 ± 7.0 patches of 0.58 ± 0.30 cm^2; results not shown) than on the two other habitats ($9/0 \pm 2.0$ patches of $1.6 \pm 1/0$ cm^2 for natural habitat and 9.4 ± 7 patches of 1.4 ± 0.7 cm^2 for the iron half-shell; results not shown). This suggests a more homogeneous spatial repartition of categories (i.e. a lower level of clustering) on the mattress habitat, which is consistent with the homogeneous nature and flatness of each single concrete unit. Consistent with the statements exposed above (Cochran, 1946; McGarvey *et al.*, 2016), accurate estimating of percentage cover of "≥5% cover" taxa on mattress habitat requires the highest number of points with stratified-random projection.

At the site scale, we found that the minimum sampling areas required to accurately describe benthic communities are habitat specific, which reflects different levels of heterogeneity in community structure across images in each habitat. Specifically, a larger sampling area is required to reach accurate estimating of community composition on mattress and natural habitats relative to half-shell habitat. Since our optimization approach is based on the taxonomic similarity between images within a site, a larger optimum sampling area means that the photographs are more different from each other, or in other words, that the spatial distribution of communities is more heterogeneous (i.e. more clustered repartition at the scale of sites). Such observations are in agreement with recent simulations showing that a larger sampling area was required to achieve an equivalent level of precision for clustered distributions relative to homogeneously distributed communities (Perkins *et al.*, 2016).

To summarize, accurate estimating of macroepibenthic community composition requires a higher point density and a larger sampling area on mattresses relative to natural and half-shelf habitats. These are the consequences of a more homogenous spatial distribution of taxa within images, while the community appears more variable across images at the site scale (which is consistent with the fact that the exposition of concrete units to the current is variable).

Relevant taxonomic sufficiency

Identification of benthic taxa from underwater imagery is difficult and often cannot be performed at a high level of taxonomic resolution, even by specialists. Consequently, using a suitable taxonomic classification is critical to annotate benthic taxa from underwater imagery. In our case, we showed consistent differences in community composition between the artificial (half-shell and mattress) and natural habitats regardless of the taxonomic resolution used. While the CATAMI classification used at its most precise level is coarser than the LTL, with 39% less taxa (27 and 44 taxa, respectively), it provides sufficient taxonomic resolution to detect community-level changes. For instance, a clear difference in taxonomic composition was highlighted between artificial and natural habitats' epibenthic communities (with both classifications) and a decrease in taxonomic resolution does not significantly impact the output of our community analysis. Similarly, James *et al.* (2017) showed that CATAMI performed as well as well-resolved classifications when assessing latitudinal gradient in benthic community structure. Nevertheless, these researchers did not demonstrate the robustness of CATAMI to characterize fine-scale changes in community structure. In the present study, we consolidate these conclusions by showing that the CATAMI image annotation scheme is also effective in characterizing local-scale changes in community composition across different hard habitats.

Our results corroborated by several studies on taxonomic sufficiency showing that identification at high taxonomic level (i.e. family level) allows reliable spatiotemporal analysis of benthic communities (Warwick, 1988; Urkiaga-Alberdi *et al.*, 1999; De Biasi *et al.*, 2003; Doerries and Van Dover, 2003). Warwick (1993) explains these results by the fact that the family level often brings together organisms showing similar major functional traits, which are supposed to react similarly to environmental fluctuations. In this study, we consider a resolution even coarser than family taxonomic rank, but a similar hypothesis can be applied to the different morphotype groups we used in the CATAMI classification. In our case, it should be noted that the differences in taxonomic resolution between the two classifications are sometimes marginal. Indeed, for 45% of the taxa, the lowest possible taxonomic level identified from imagery corresponds actually to the morphotype level used with the CATAMI typology at its more precise level. In this sense, CATAMI classification is well adapted for image-based descriptions of macroepibenthic communities.

In addition to providing consistent results relative to a study-specific taxonomic classification, the standardized classification CATAMI can make image analysis not only faster but also more reliable. Indeed, identification at a lower taxonomic resolution decreases misidentification risks and allows non-specialists to analyse large sets of images. These advantages make CATAMI a well-suited classification scheme in our case, and we recommend its broader application for underwater imagery annotation to facilitate the comparisons of ecological patterns across studies.

Conclusions

While our optimal image-processing protocol remains specific to our case study, we believe that our stepwise strategy provides transposable guidelines to rationally tackle the challenges inherent to underwater image annotation. Specifically, we described how to balance out the different imagery annotation choices (i.e. point score density, sampling effort per site, and taxonomic resolution) to reach a set level of accuracy in percentage cover estimates in a time-effective manner.

Acknowledgements

This study is sponsored by the Région Bretagne, France Energies Marines, and the National Research Agency within the framework of Investments for the Future programme under reference ANR-10-IED-0006-17. The authors thank Nolwenn Quillien, Morgane Lejart, Olivier Gauthier, and Nicolas Job for their kind assistance.

References

Althaus, F., Hill, N., Ferrari, R., Edwards, L., Przeslawski, R., Schönberg, C. H. L., Stuart-Smith, R., *et al.* 2015. A standardised vocabulary for identifying benthic biota and substrata from underwater imagery: the CATAMI classification scheme. PLoS One, 10: e0141039–18.

Berov, D., Hiebaum, G., Vasilev, V., and Karamfilov, V. 2016. An optimised method for scuba digital photography surveys of infralittoral benthic habitats: a case study from the SW Black Sea Cystoseira-dominated macroalgal communities. Underwater Technology, 34: 11–20.

Brown, E. K., Cox, E., Jokiel, P. L., (Paul, L., Rodgers, S. K., Smith, W. R., Tissot, B. N., *et al.* 2004. Development of benthic sampling methods for the Coral Reef Assessment and Monitoring Program (CRAMP) in Hawai'i. Pacific Science, 58: 145–158.

Clarke, K. R., and Warwick, R. M. 2001. Change in marine communities. An approach to statistical analysis and interpretation, 2nd edition. PRIMER-E, Plymouth-UK. 1–172 pp.

Cochran, W. G. 1946. Relative accuracy of systematic and stratified random samples for a certain class of populations. The Annals of Mathematical Statistics, 17: 164–177.

De Biasi, A. M., Bianchi, C. N., and 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.

Deter, J., Descamp, P., Boissery, P., Ballesta, L., and Holon, F. 2012. A rapid photographic method detects depth gradient in coralligenous assemblages. Journal of Experimental Marine Biology and Ecology, 418–419: 75–82.

Dethier, M. N., Graham, E. S., Cohen, S., and Tear, L. M. 1993. Visual versus random-point percent cover estimations: "objective" is not always better. Marine Ecology Progress Series, 96: 93–100.

Doerries, M. B., and Van Dover, C. L. 2003. Higher-taxon richness as a surrogate for species richness in chemosynthetic communities. Deep Sea Research Part I, 50: 749–755.

Domke, L., Lacharité, M., Metaxas, A., and Matabos, M. 2017. Influence of an oxygen minimum zone and macroalgal enrichment on benthic megafaunal community composition in a NE Pacific submarine canyon. Marine Ecology, 38: e12481–19.

Dumas, P., Bertaud, A., Peignon, C., Léopold, M., and Pelletier, D. 2009. A 'quick and clean' photographic method for the description of coral reef habitats. Journal of Experimental Marine Biology and Ecology, 368: 161–168.

Dutilleul, P. 1993. Spatial heterogeneity and the design of ecological field experiments. Ecology, 74: 1646–1658.

Foveau, A., Haquin, S., and Dauvin, J. C. 2017. Using underwater imagery as a complementary tool for benthos sampling in an area with high-energy hydrodynamic conditions. Journal of Marine Biology & Oceanography, 6: 1–7.

Halpern, B. S., Walbridge, S., Selkoe, K. A., Kappel, C. V., Micheli, F., D'Agrosa, C., Bruno, J. F., *et al.* 2008. A global map of human impact on marine ecosystems. Science, 319: 948.

Harrison, M. A., and Smith, S. D. A. 2012. Cross-shelf variation in the structure of molluscan assemblages on shallow, rocky reefs in subtropical, eastern Australia. Marine Biodiversity, 42: 203–216.

Jackson, J. B. C., Kirby, M. X., Berger, W. H., Bjorndal, K. A., Botsford, L. W., Bourque, B. J., Bradbury, R. H., *et al.* 2001. Historical overfishing and the recent collapse of coastal ecosystems. Science, 293: 629–637.

James, L., Marzloff, M., Barrett, N., Friedman, A., and Johnson, C. 2017. Changes in deep reef benthic community composition across a latitudinal and environmental gradient in temperate Eastern Australia. Marine Ecology Progress Series, 565: 35–52.

Jerabek, A. S., Wall, K. R., and Stallings, C. D. 2016. – practical application of reduced-copper antifouling paint in marine biological research. PeerJ, 4: e2213.

Jimenez, C., Hadjioannou, L., Petrou, A., Andreou, V., and Georgiou, A. 2017. Fouling communities of two accidental artificial reefs (modern shipwrecks) in Cyprus (levantine sea). Water, 9: 11.

Jokiel, P. L., Rodgers, K. S., Brown, E. K., Kenyon, J. C., Aeby, G., Smith, W. R., and Farrell, F. 2015. Comparison of methods used to estimate coral cover in the Hawaiian Islands. PeerJ, 3: e954.

Kohler, K. E., and Gill, S. M. 2006. Coral Point Count with Excel extensions (CPCe): a Visual Basic program for the determination of coral and substrate coverage using random point count methodology. Computers & Geosciences, 32: 1259–1269.

Kronberg, I. 1987. Accuracy of species and abundance minimal areas determined by similarity area curves. Marine Biology, 96: 555–561.

Lam, K., Shin, P. K. S., Bradbeer, R., Randall, D., Ku, K. K. K., Hodgson, P., and Cheung, S. G. 2006. A comparison of video and point intercept transect methods for monitoring subtropical coral communities. Journal of Experimental Marine Biology and Ecology, 333: 115–128.

Langenkämper, D., Zurowietz, M., Schoening, T., and Nattkemper, T. W. 2017. BIIGLE 2.0—browsing and annotating large marine image collections. Frontiers in Marine Science, 4: 1–10.

Le Pape, O., Gilliers, C., Riou, P., Morin, J., Amara, R., and Désaunay, Y. 2007. Convergent signs of degradation in both the capacity and the quality of an essential fish habitat: state of the Seine estuary (France) flatfish nurseries. Hydrobiologia, 588: 225–229.

Legendre, P., Dale, M. R. T., Fortin, M.-J., Gurevitch, J., Hohn, M., and Myers, D. 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. Ecography, 25: 601–615.

Lotze, H. K., Lenihan, H. S., Bourque, B. J., Bradbury, R. H., Cooke, R. G., Kay, M. C., Kidwell, S. M., *et al.* 2006. Depletion, degradation, and recovery potential of estuaries and coastal seas. Science, 312: 1806–1809.

Lotze, H. K., and Milewski, I. 2004. Two centuries of multiple human impacts and successive changes in a North Atlantic Food Web. Journal of Ecological Society of America, 14: 1428–1447.

Macedo, I. M., Pereira Masi, B., and Zalmon, I. R. 2006. Comparison of rocky intertidal community sampling methods at the Northern coast of Rio de Janeiro state, Brazil. Brazilian Journal of Oceanography, 54: 147–154.

Marzloff, M. P., Oliver, E. C. J., Barrett, N. S., Holbrook, N. J., James, L., Wotherspoon, S. J., and Johnson, C. R. 2018. Differential vulnerability to climate change yields novel deep-reef communities. Nature Climate Change, 8: 873–878.

McGarvey, R., Burch, P., and Matthews, J. M. 2016. Precision of systematic and random sampling in clustered populations: habitat patches and aggregating organisms. Ecological Applications, 26: 233–248.

Molloy, P. P., Evanson, M., Nellas, A. C., Rist, J. L., Marcus, J. E., Koldewey, H. J., and Vincent, A. C. J. 2013. How much sampling does it take to detect trends in coral-reef habitat using photoquadrat surveys? Aquatic Conservation: Marine and Freshwater Ecosystems, 23: 820–837.

Nakajima, R., Nakayama, A., Yoshida, T., Kushairi, M. R. M., Othman, B. H. R., and Toda, T. 2010. An evaluation of photo line-intercept transect (PLIT) method for coral reef monitoring. Galaxea. Journal of Coral Reef Studies, 12: 89–90.

O'Carroll, J. P. J., Kennedy, R. M., and Savidge, G. 2017. Identifying relevant scales of variability for monitoring epifaunal reef communities at a tidal energy extraction site. Ecological Indicators, 73: 388–397.

Oh, E. S., Edgar, G. J., Kirkpatrick, J. B., Stuart-Smith, R. D., and Barrett, N. S. 2015. Broad-scale impacts of salmon farms on temperate macroalgal assemblages on rocky reefs. Marine Pollution Bulletin, 98: 201–209.

Page, H. M., Dugan, J. E., Culver, C. S., and Hoesterey, J. C. 2006. Exotic invertebrate species on offshore oil platforms. Marine Ecology Progress Series, 325: 101–107.

Pante, E., and Dustan, P. 2012. Getting to the Point: accuracy of Point Count in Monitoring Ecosystem Change. Journal of Marine Biology, 2012: 1–7.

Pech, D., Condal, A. R., Bourget, E., and Ardisson, P. L. 2004. Abundance estimation of rocky shore invertebrates at small spatial scale by high-resolution digital photography and digital image analysis. Journal of Experimental Marine Biology and Ecology, 299: 185–199.

Perkins, N. R., Foster, S. D., Hill, N. A., and Barrett, N. S. 2016. Image subsampling and point scoring approaches for large-scale marine benthic monitoring programs. Estuarine, Coastal and Shelf Science, 176: 36–46.

Pielou, E. C. 1974. Population and Community Ecology: Principles and Methods. Gordon and Breach Science Publishers, New York/Paris/London. 432 pp.

Preskitt, L. B., Vroom, P. S., and Smith, C. M. 2004. A Rapid Ecological Assessment (REA) quantitative survey method for benthic algae using photoquadrats with scuba. Pacific Science, 58: 201–209.

RStudio Team. 2015. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA.

Sartoretto, S., Schohn, T., Bianchi, C. N., Morri, C., Garrabou, J., Ballesteros, E., Ruitton, S., *et al.* 2017. An integrated method to evaluate and monitor the conservation state of coralligenous habitats: the INDEX-COR approach. Marine Pollution Bulletin, 120: 222–231.

Schlacher, T. A., Williams, A., Althaus, F., and Schlacher-Hoenlinger, M. A. 2010. High-resolution seabed imagery as a tool for biodiversity conservation planning on continental margins. Marine Ecology, 31: 200–221.

Schmera, D., and Eros, T. 2006. Estimating sample representativeness in a survey of stream caddisfly fauna. Annales de Limnologie—International Journal of Limnology, 42: 181–187.

Schneck, F., and Melo, A. S. 2010. Reliable sample sizes for estimating similarity among macroinvertebrate assemblages in tropical streams. Annales de Limnologie—International Journal of Limnology, 46: 93–100.

Sen, A., Kim, S., Miller, A. J., Hovey, K. J., Hourdez, S., Luther, G. W., and Fisher, C. R. 2016. Peripheral communities of the Eastern Lau Spreading Center and Valu Fa Ridge: community composition, temporal change and comparison to near-vent communities. Marine Ecology, 37: 599–617.

Trygonis, V., and Sini, M. 2012. PhotoQuad: a dedicated seabed image processing software, and a comparative error analysis of four photoquadrat methods. Journal of Experimental Marine Biology and Ecology, 424–425: 99–108.

Urkiaga-Alberdi, J., Pagola-Carte, S., and Saiz-Salinas, J. I. 1999. Reducing effort in the use of benthic bioindicators. Acta Oecologica, 20: 489–497.

Van Rein, H. B., Brown, C. J., Quinn, R., and Breen, J. 2009. A review of sublittoral monitoring methods in temperate waters: a focus on scale. Underwater Technology, 28: 99–113.

Van Rein, H. B., Schoeman, D. S., Brown, C. J., Quinn, R., and Breen, J. 2011. Development of benthic monitoring methods using photoquadrats and scuba on heterogeneous hard-substrata: a boulder-slope community case study. Aquatic Conservation: Marine and Freshwater Ecosystems, 21: 676–689.

Vroom, P. S., and Timmers, M. A. V. 2009. Spatial and temporal comparison of algal biodiversity and benthic cover at gardner pinnacles, Northwestern Hawaiian Islands. Journal of Phycology, 45: 337–347.

Walker, S. J., Schlacher, T. A., and Schlacher-Hoenlinger, M. A. 2007. Spatial heterogeneity of epibenthos on artificial reefs: fouling communities in the early stages of colonization on an East Australian shipwreck. Marine Ecology, 28: 435–445.

Walvoort, D. J. J., Brus, D. J., and de Gruijter, J. J. 2010. An R package for spatial coverage sampling and random sampling from compact geographical strata by k-means. Computers and Geosciences, 36: 1261–1267.

Warwick, R. M. 1988. Analysis of community attributes of the macrobenthos of Frierfjord-Langesundfjord at taxonomic levels higher than species. Marine Ecology-Progress Series, 46: 167–170.

Warwick, R. M. 1993. Environmental impact studies on marine communities: pragmatical considerations. Australian Journal of Ecology, 18: 63–80.

Weinberg, S. 1978. The minimal area problem in invertebrate communities of Mediterranean rocky substrata. Marine Biology, 49: 33–40.

Handling editor: Joanna Norkko