Estuarine, Coastal and Shelf Science 176 (2016) 36–46

Contents lists available at ScienceDirect

Estuarine, Coastal and Shelf Science

journal homepage: www.elsevier.com/locate/ecss

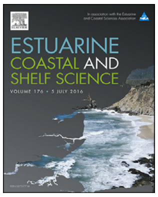


Image subsampling and point scoring approaches for large-scale marine benthic monitoring programs

Nicholas R. Perkins [a], [*], Scott D. Foster [b], Nicole A. Hill [a], Neville S. Barrett [a]

[a] Institute of Marine and Antarctic Studies, University of Tasmania, Australia
[b] Commonwealth Scientific and Industrial Research Organisation, Australia

ARTICLE INFO

Article history:
Received 11 March 2015
Received in revised form
4 March 2016
Accepted 7 April 2016
Available online 9 April 2016

Keywords:
Autonomous underwater vehicle
Image selection
Point count
Precision
Bias
Survey design

ABSTRACT

Benthic imagery is an effective tool for quantitative description of ecologically and economically important benthic habitats and biota. The recent development of autonomous underwater vehicles (AUVs) allows surveying of spatial scales that were previously unfeasible. However, an AUV collects a large number of images, the scoring of which is time and labour intensive. There is a need to optimise the way that subsamples of imagery are chosen and scored to gain meaningful inferences for ecological monitoring studies. We examine the trade-off between the number of images selected within transects and the number of random points scored within images on the percent cover of target biota, the typical output of such monitoring programs. We also investigate the efficacy of various image selection approaches, such as systematic or random, on the bias and precision of cover estimates. We use simulated biotas that have varying size, abundance and distributional patterns. We find that a relatively small sampling effort is required to minimise bias. An increased precision for groups that are likely to be the focus of monitoring programs is best gained through increasing the number of images sampled rather than the number of points scored within images. For rare species, sampling using point count approaches is unlikely to provide sufficient precision, and alternative sampling approaches may need to be employed. The approach by which images are selected (simple random sampling, regularly spaced etc.) had no discernible effect on mean and variance estimates, regardless of the distributional pattern of biota. Field validation of our findings is provided through Monte Carlo resampling analysis of a previously scored benthic survey from temperate waters. We show that point count sampling approaches are capable of providing relatively precise cover estimates for candidate groups that are not overly rare. The amount of sampling required, in terms of both the number of images and number of points, varies with the abundance, size and distributional pattern of target biota. Therefore, we advocate either the incorporation of prior knowledge or the use of baseline surveys to establish key properties of intended target biota in the initial stages of monitoring programs.

1. Introduction

Increasing human impacts on marine ecosystems heighten the need for rapid and cost effective assessment and monitoring methods (Halpern et al., 2008; Brown et al., 2011). Benthic habitats play a vital ecological role and support fisheries with high economic value (Hughes et al., 2005). Quantification of benthic habitats and biota is necessary to better understand spatial patterns, monitor changes and assess the impacts of management strategies (Molloy et al., 2013). In this context, marine imagery collected by unmanned vehicles enables surveys of large areas, access to environments that are difficult to survey such as deeper waters, and creates a permanent record that allows comparison over time (Dumas et al., 2009). Photographic approaches yield high quality quantitative information on benthic communities including species presence-absence, direct counts of individuals or colonies, areal or percent cover estimates, and estimates of size (Hill and Wilkinson, 2004; Dumas et al., 2009; Trygonis and Sini, 2012). Recent technological advances have facilitated the routine collection of increasingly larger amounts of imagery, with some survey platforms able to capture several thousand images over a few hours of deployment (Pizarro et al., 2013). There is now a pressing need to assess the statistical robustness and statistical efficiency of competing

---

[*] Corresponding author. Institute of Marine and Antarctic Studies, Private Bag 49, University of Tasmania, Hobart, 7001, Australia.
E-mail address: Nicholas.Perkins@utas.edu.au (N.R. Perkins).

http://dx.doi.org/10.1016/j.ecss.2016.04.005
0272-7714/

sampling approaches for extracting ecological data from the large volumes of imagery produced.

One common use of marine imagery in large-scale ecological surveys is the initial inventory and ongoing monitoring of key groups or communities (Brown et al., 2004; Hill and Wilkinson, 2004; Smale et al., 2012). This typically involves the estimation of the areal coverage of “visually conspicuous” (0.01–1.0 m diameter) sessile benthic organisms or assemblages indicative of a particular habitat, and measuring differences in this coverage between regions or over time (Van Rein et al., 2009). A range of platforms have been utilised for this type of work, including diver-swum transects with hand-held cameras, the use of stills from towed camera systems, remotely operated vehicles (ROVs) and autonomous underwater vehicles (AUVs). Techniques using SCUBA divers typically employ multiple short transects (<50 m) and are restricted to shallower depths (Brown et al., 2004; Leujak and Ormond, 2007). Towed systems, ROVs and AUVs are capable of covering much larger areas, can operate at greater depths and return a large amount of data in a relatively short time. For these reasons, technologies such as these are more commonly employed when surveys are to cover meso-scales (Van Rein et al., 2009). For example, an AUV has been deployed as the platform of choice for the benthic surveying and monitoring of a large number of sites around the Australian coastline (Pizarro et al., 2013). Establishing the reliability of coverage estimates in such large-scale projects is of crucial importance from both scientific and economic standpoints.

The reliability of estimates is intrinsically linked to their precision (how variable one sample is from the next) and bias (the difference between the expected value of an estimator and the true value of the parameter being estimated). Results of surveys, and estimates of variability in particular, will not only be dependent upon the natural variability in the biota but also dependent upon: (i) the method used and effort spent on scoring each image, (ii) the number of images selected from each transect and their method of selection, (iii) the number of replicate transects used in an area, and (iv) the length and design of the transects (Houk and Van Woesik, 2006). Survey approaches which provide greater precision minimise the variation attributable to the sampling design and maximise the information about the biota. Issues surrounding survey and transect design are dealt with by Foster et al. (2014). Here we focus on the issues of subsampling imagery from transects and the intensity with which individual images are scored.

When scoring individual images, percent cover is the most common metric used to quantify benthic organisms (Van Rein et al., 2011; Deter et al., 2012). Images are usually selected along a transect or within sites and arguably the most common approach is the use of point count methodology, whereby the proportion of randomly overlain points intersecting an organism or substratum is used to calculate its coverage (Pielou, 1974). This approach has been facilitated by the development of dedicated point-count software, such as Coral Point Counts (CPC) (Kohler and Gill, 2006). The optimum number of points for scoring individual images (e.g. Dumas et al., 2009; Deter et al., 2012), and the interplay between the number of points per image and the number of images has been well studied (e.g. Brown et al., 2004). However, attributes of biological organisms, such as size, abundance and spatial pattern may also affect this trade-off, but have not been considered. Further, different image selection approaches (e.g. random or systematic sampling) within a transect will result in a different spatial spread of samples, which may affect efficiency (or bias and variance), particularly when the distribution of biota is clustered, yet a formal assessment of these potential affects is lacking. We aim to take a holistic approach by simultaneously examining the effect of both attributes of the biota and sampling approaches on the accuracy and precision of estimates of the precent cover of biota. Due to the likelihood of a strong interplay between these factors, an approach is required that allows an examination of the various trade-offs simultaneously. We do this through the use of Monte Carlo simulation and resampling approaches, which offer a flexible way to vary these parameters, whilst comparing the outputs to known percent covers, a quantity typically not known in real surveys (Bros and Cowell, 1987).

We report findings on the interplay between: (i) the number of images selected, (ii) the number of points used, and (iii) image selection approach on estimates of the cover of biota of varying spatial distribution and size classes. We use AUVs as a case study as they have a number of features that make them an ideal platform for benthic surveys in cross-shelf waters although other platforms share many common issues with AUVs. In order to “ground truth” the findings of our simulation we use Monte Carlo resampling techniques on actual scored AUV imagery to analyse the implications of reducing the number of points scored per image. Through analysis of simulated and actual benthic imagery we provide general recommendations for the level of subsampling and scoring that may be appropriate for scoring of marine benthic imagery.

2. Methods

2.1. Simulation methodology

As the basis for our simulation study, we used spatial scales of deployment and transect designs that have been used in AUV deployments under the government funded Integrated Marine Observing System (IMOS) monitoring program in Australian shelf waters (Barrett et al., 2010; Pizarro et al., 2013). All surveys for the IMOS program were conducted using the AUV Sirius, an AUV sampling platform designed by the Australian Centre for Field Robotics (ACFR) at the University of Sydney (http://www.acfr.usyd.edu.au/research/projects/subsea/auvSIRIUS.shtml).

Our simulation involved two major components: (i) the creation of biological distributions within a virtual seascape, and (ii) the sampling and subsampling of this distribution with various image and point scoring approaches along an AUV transect.

2.2. Simulating the distribution of biota

To create the biological distributions, a sample frame of 400 × 800 m, that encompasses the approximate survey area that has been used for AUV transects in the IMOS program, was established. Within the sample frame a biological distribution was created using either random or clustered spatial point patterns (see Diggle, 1983), with points assigned one of three size classes and abundances (Table 1).

Random biological distributions were simulated using a Poisson process. Clustered distributions of biota are commonly observed in ecological studies due to fine-scale biological and environmental factors such as dispersal and habitat structure (Dormann et al., 2007). Clustered distributions were simulated using a Neyman-Scott clustering process (see Diggle, 1983). The intensity of the processes were given values that resulted in percent covers that were similar to those of organisms of interest in previously scored data (see below). For the Neyman-Scott clustering processes, a homogenous Poisson process was used to establish the ‘parent points’. Each parent point spawned daughter points whose number follows a Poisson distribution. These daughter points were spatially distributed at a random angle and distance from the parent point. We examined two different clustered processes: (i) Mildly clustered: 10 daughter points, with the distance daughter points fell from the parent being drawn from a random normal distribution with a mean of zero and standard deviation of 5 and maximum

Table 1
The biological distributions used in the simulation component of the study. The clustered process is a Neyman-Scott process where: Clustered[1] has an offset distance of mean zero, sd 5 m and a mean of 10 daughter points, and Clustered[2] has an offset distance of mean zero, sd 3 m and a mean of 20 daughter points. For more details see Methods.

Biological distribution	Size range	Abundance	Point pattern distribution
Large	25–80 cm diameter	~40% coverage	Random
Large (mildly clustered)	25–80 cm diameter	~40% coverage	Clustered [1]
Large (clustered)	25–80 cm diameter	~40% coverage	Clustered [2]
Medium	10–25 cm diameter	~5% coverage	Random
Medium (mildly clustered)	10–25 cm diameter	~5% coverage	Clustered [1]
Medium (clustered)	10–25 cm diameter	~5% coverage	Clustered [1]
Small	1–10 cm diameter	~0.1% coverage	Random
Small (mildly clustered)	1–10 cm diameter	~0.1% coverage	Clustered [1]
Small (clustered)	1–10 cm diameter	~0.1% coverage	Clustered [2]

radius of 20 m (ii) Clustered: 20 daughter points, with the distance daughter points fell from the parent being drawn from a random normal distribution with a mean zero, standard deviation of 3 and maximum radius of 5 m. The number of parent points was varied to allow the expectation of the percent cover created for equivalent random and clustered processes to be kept equal for comparative purposes.

A size class was assigned to each point pattern created: either small (1 cm–10 cm diameter), medium (10 cm–25 cm diameter) or large (25 cm–80 cm diameter) to represent different sized biota. A random uniform distribution of sizes was assigned to the points within these specified limits. These size classes were chosen as they were roughly analogous to the sizes of the organisms chosen for the analysis of previously scored data (see below). Also, these sizes are typical of benthic organisms that may be scored in AUV imagery.

2.3. Sampling of the biological distribution

The sampling component of the simulation involved generating an 'S' shaped transect with a random start point within the sample frame, applying one of three image selection approaches to images along the transect, and then overlaying random points within the images. The 'S' shaped transect (see Fig. 1), consisted of three longer sections of 500 m and two shorter sections of 150 m, giving a total transect length of 1.8 km. This design was chosen as it is a stylisation of a design that is currently widely employed for AUV surveys around Australia (Williams et al., 2012). Random starting points for the transects were included to simulate the difficulties in precisely locating transects in the marine environment. For each simulation a set of non-overlapping images (1.6 × 1.2 m -the average size of an AUV image) was defined along the length of the transect. Images were selected using a specified image sampling methodology and number of images, and a specified number of random points were overlaid within the image boundaries.

The 'sampled' percent cover for each image was determined as the number of points that fell inside the biological pattern divided by the total number of points used within the image. The mean percent cover for the transect and the standard deviation of this estimate was then determined by taking the average (and standard deviation) of the image percent covers. The 'sampled' percent cover could then be compared with the 'true' simulated cover. An illustration of two potential distributions and image sampling approaches, as well as the within image point sampling is provided in Fig. 1.

Calculation of the 'true' coverage of the biological distribution over the survey area, which the point count was estimating, was complicated by the fact that some members of the biological distribution had a degree of overlap. So, whilst the total area of all

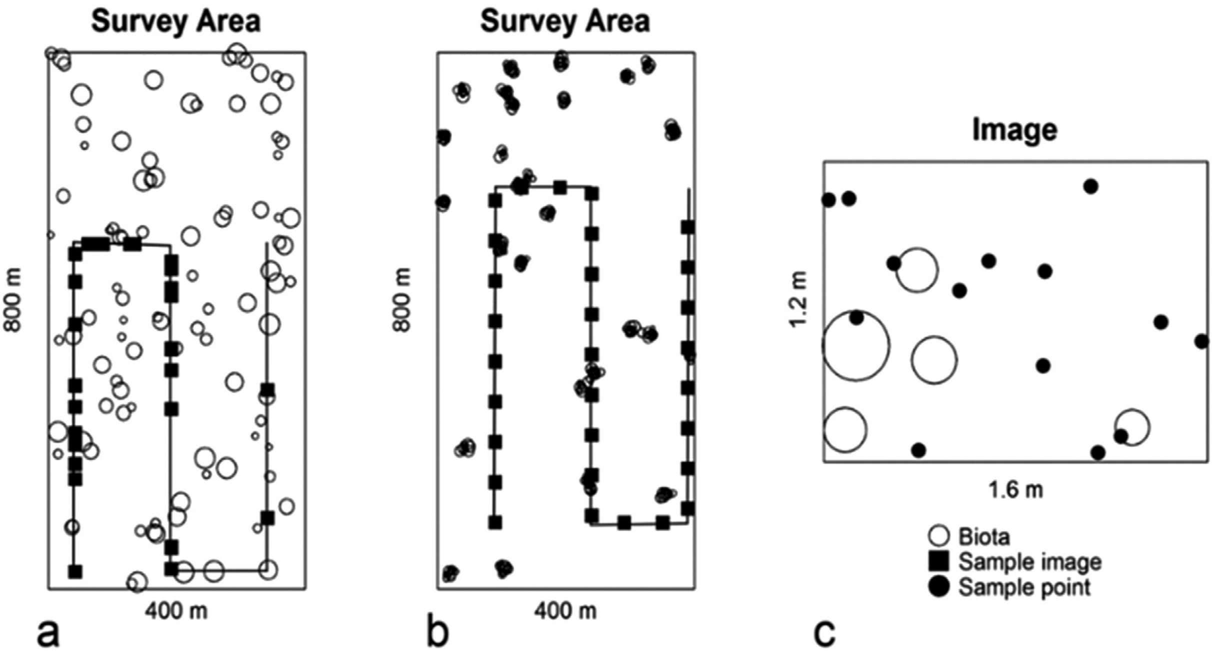


Fig. 1. Two realisations of biological distributions and image sampling strategies: (a) a random biological distribution with a random image sampling strategy and 30 images; and (b) a clustered biological distribution with a systematic image sampling strategy and 30 images. Panel (c) illustrates within image subsampling of medium-sized biota with 14 random points, with 3 points falling on the biota. Biota and image sizes are not shown to scale, and are used for illustrative purposes only.

circles was known, the extent to which any adjacent circles overlapped reduced the total areal coverage in any particular simulated distribution. Determining the exact area in each case of overlap is empirically complicated, particularly in cases of multiple overlap (for a discussion of the mathematical complications see Librino et al., 2014). Therefore, it was necessary to calculate the 'true' cover by dividing up the sample frame into 80 rectangular slices, each 10 m × 400 m. This was done due to the computational load of calculating the whole area at once, although with higher computing power this step may be unnecessary. Within each slice, point positions based on a set of 10,000 quasi-random numbers (Press, 2007) were generated and it was subsequently determined which of these points fell within the simulated biota. True percent cover could then be calculated by dividing the number of points that fell inside the biota by the total number of points in the rectangular slice. The proportions for each segment were totalled to give the true percent coverage over the entire area. In test runs 10,000 points in each slice provided low levels of variance (coefficients of variation <0.05) in true cover estimates.

2.4. Examining the trade-off between the number of images and number of points

For each of the biological distributions outlined in Table 1, the trade-off between the number of images (between 2 and 100 in increments of 2) selected within a transect and the number of points (between 1 and 50 in increments of 1) within an image on the bias and precision of percent cover estimates was explored. In order to keep results brief, for this component of the study we chose to examine only systematic image sampling. Systematic image sampling was used as this is the approach that has been used in scoring real AUV transects in the IMOS program. Monte Carlo simulations were conducted where a biological distribution was created for each combination of pattern/abundance/size (Table 1), and all sampling combinations of number of images and number of points were tested over 200 runs, with random transect starting points for each run. Therefore, there were a total of 2500 combinations of number of images/number of points, tested for each of the six biological distributions over 200 runs. This gave a total of 3,000,000 simulated data sets.

The trade-off between points per image and images was assessed using the mean bias and the coefficient of variation (CV = standard deviation/mean), which were calculated for each sampling regime over all simulations. Note that we use the means and standard deviations of our proportion data, and do not transform our data in any way. We also note that the point estimate from taking the first moment (mean), as done here will coincide with that obtained from maximum likelihood methods that are suggested in Warton and Hui (2011). Thus, we avoid any potential inference complications that are a consequence of transformations (Warton and Hui, 2011). The mean bias was calculated as the difference between the 'true cover' and the 'sampled cover', the latter from the point count methodology. As our biological distributions had markedly differing mean coverages (Table 1), using CV provided a normalised measure of dispersion that allowed comparison of the precision of different sampling approaches for the various biological distributions. Lower CV values indicate more precise estimates, and hence more effective sampling. The resultant CVs for all image/point combinations were plotted separately for each biological distribution outlined in Table 1. These plots were used in order to highlight how the number of images and the number of points act in tandem in determining the precision of estimates. Contour lines were added to these plots for CVs of 0.1 and 0.2 for the large and medium categories. These CVs were chosen for illustrative purposes as they represent relatively precise estimates (equating to standard deviations within 10% and 20% of the mean respectively). Actual levels of precision required for a monitoring would be dependent on the question being addressed, such as the level of power required to detect a change or the size of the change to be detected. CVs of 0.1 and 0.2 fell outside the sampling intensity used for the small-low category and a CV of 0.5 was used. Optimum sampling regimes for the given AUV transect were determined by summing the total number of points (i.e. number of images by number of points) required to reach the 0.1, 0.2 and 0.5 CV thresholds and calculating which combination gave the minimum number of scored points.

2.5. Image selection approaches

When selecting images, consideration may need to be given to issues such as the patchiness of the habitat and/or the spatial distribution of organisms of interest. In order to examine the effect of image selection approach, we conducted a separate simulation that tested three differing image selection approaches: random, systematic and a spatially balanced approach 'Generalized Random Tessellation Stratified '(GRTS, Stevens and Olsen, 2004) that has been used in wildlife monitoring in the USA, with recent application to marine ecosystems (Hill et al., 2014). We tested the effect of these three approaches at varying image and point sampling intensities for the medium (clustered) distribution (Table 1). We chose this clustered pattern as it was expected that GRTS may perform better for a distribution that had a spatial pattern as it has been shown to provide a more efficient sampling design when a spatial signal exists.

For this simulation, 500 Monte Carlo simulations were conducted. Image selection intensities ranged from 20 to 100 by 20, and point count intensities from 10 to 50 by 10. Both the mean bias and standard error of the mean estimate were compared across the image sampling approaches at all intensities. Mean bias was calculated as the average of the difference between the calculated and true covers for the particular sampling strategy across all runs. Standard errors were calculated from the standard deviation of the mean estimate over all runs. Standard errors rather than CVs were used in this simulation, as only one biological pattern was used with the same mean percentage cover over all simulations.

2.6. Monte Carlo resampling of previously scored data

Fourteen previously scored AUV transects from the East Tasman Peninsula in Tasmania, Australia were used to analyse the effect of altering the number of points per image used to estimate the cover of several benthic taxa. This allowed an assessment of whether a less intense sampling effort would still provide robust estimates, as well as providing some ground-truthing for the simulation. These transects were scored by selecting every 100th image and scoring with 50 random points overlaid on the image using CPC software (Kohler and Gill, 2006). Every 100th image equated to an average of approximately 60 images per transect with a spacing of ~35 m between images. Transects were originally scored to a detailed "morphospecies" level, with each distinct morphotype being given a unique classifier. These classifications were subsequently amalgamated into broader morphological groups under the CATAMI classification system (Althaus et al., 2013), hereafter referred to as "morphotypes". The CATAMI system is a heirarchical classification scheme designed specifically for marine imagery, which is based on a combination of taxonomy and morphology.

Monte Carlo simulations were conducted in order to test the effect of reducing the number of points per image scored on the mean and standard deviation estimates of the percent cover of selected morphotypes over each transect. The morphotypes selected for the analysis were: canopy forming brown algae, erect branching sponges, simple massive sponges, cup sponges and bramble coral and represent organisms with high to low percent cover, respectively, over transects. Five representative transects were selected to compare the results of our simulation study where coverage of canopy algae was analogous to our large class, the three sponge morphotypes to our medium class and bramble coral to our small-low class.

1000 resampling simulations were run, calculating the percent cover for each of these morphotypes when using all numbers of points between 50 and 1. For each simulation the following steps were taken: (i) a subsample of all potential number of points from 1 to 50 was made by taking a random sample with replacement from the total of 50 points within each image; (ii) for each number of points used, the percent cover of each of the selected morphotypes was calculated for each image within the transect; and (iii) the within image percent cover using each number of points was averaged to give a percent cover for each selected morphotype over the transect. Finally, over all simulations, the CV of the cover estimates was calculated for each number of points used. We once again use CV in order to allow us to compare the precision of estimates with vastly different means, and also to allow comparison with our simulations results.

3. Results

3.1. Simulation results

For all distributions, once sampling exceeded approximately 30 images with 10 points, no systematic bias in mean estimates was evident (see supplementary materials S1). This threshold was higher for the clustered compared to the random distributions and for organisms with lower abundance. At higher sampling intensities, (e.g. those that gave CVs <0.2 for the large and medium distributions and <0.5 for the small distributions – see below), bias was typically <2% of the mean values.

For the large and medium distributions, increased precision was achieved more rapidly by including more images in the sample rather than increasing the number of points per image (Fig. 2). For these distributions, Fig. 2 shows CVs decreasing more rapidly along the images axis compared with the points axis. Overall, increased sampling intensities (either more images or more points per image) were required for distributions that have lower overall cover, and for those with clustered distributions when compared to random distributions with equivalent cover. For the large distributions, a low number of points per image (<10) were required to gain precise estimates (CV = 0.1) once more than approximately 30 images were selected within the transects. For the medium distributions, in order to achieve CV = 0.1 a sufficiently high (>25–46) number of points were required along with a large number of images (>90). A CV of 0.2 could be achieved with a low number of points (<10) provided a large number of images (>60–80) were included in the sample. For the small-low distributions, both a high number of images (>80–90) and points per image (>40) needed to be sampled in order to reach a relatively low level of precision (CV = 0.5). The sampling intensity required to reach CV levels of 0.1 or 0.2 fell outside of the bounds that we tested for the small-low distributions.

Increased sampling was required (both the number of images and number of points) with increased clustering in the distributions (Fig. 2). This effect was particularly noticeable for the medium distribution, where the level of sampling required to reach a CV of 0.1 increased noticeably as the distribution became more clustered.

If it is assumed that an equal amount of time is required to score each point then scoring the minimum number of points required to achieve the desired CV also optimises the cost-benefit of scoring. The 'optimal' combination of the number of images and points per image required to reach CVs of 0.1 and 0.2 for the large and medium distributions are summarized in Table 2. For the small-low distributions, sampling intensities close to the limit of those tested needed to be employed to reach a CV = 0.5, and hence results are not displayed in Table 2.

The various image selection approaches performed equally well in terms of the estimation of the standard error of the mean, and levels of bias in mean estimates (Fig. 3). There was no evidence of any image sampling approach providing better estimates at any combination of number of images or number of points tested.

3.2. Monte Carlo resampling of previously scored data

Reducing the number of points used to score the cover of five morphotypes across five representative transects gave remarkably similar results to those of our simulation (Fig. 4). On average there were approximately 60 images scored in each transect. To achieve a CV of 0.1 required approximately 3 points per image for the canopy algae (analogous to our large group). 15–20 points per image were required for a CV of 0.2 for the sponge morphotypes (medium), and achieving a CV of 0.1 was close to or beyond the limit of sampling. For bramble coral (small), a CV of 0.5 was not achievable using all images with 50 points. These results compare well with using 60 images in the Monte Carlo resampling to the plots in Fig. 2. Results for all 14 transects studied are provided in the supplementary materials S2 and show the number of points at which the standard deviation begins to asymptote in the representative transects.

4. Discussion

A rigorous evaluation of various survey and sampling methodologies is an important precursor to initiating monitoring programs (Brown et al., 2004). Whilst such programs may have a variety of end goals, if the sampling design utilised does not provide sufficient accuracy (i.e. low bias and good precision) then inferences are ambiguous at best (Miller and Ambrose, 2000). Here we provide practical advice for those who intend to utilise benthic imagry for broad-scale surveys and monitoring. Through the use of simulated biological distributions that incorporate a range of size classes, abundances and patterns we provide results that are generalizable and can be applied in a wide range of settings. Furthermore, results of our simulation are supported through Monte Carlo resampling analysis of previously scored imagery, adding further evidence that our results are applicable to real world benthic surveys. Our results highlight the importance of taking into account the abundance, size and distribution of target biota to be used in monitoring programs, and hence the need for either incorporating prior knowledge or conducting baseline surveys to establish these key properties prior to determining sampling protocols.

This work builds on a growing body of literature that has explored the issues surrounding sampling of marine benthic imagery (e.g. Brown et al., 2004; Leujak and Ormond, 2007; Molloy et al., 2013). In particular, it adds weight to the evidence that for broad-scale monitoring work it is more important to replicate sampling at gross rather than finer scales by increasing the number of images rather than points within images (Brown et al., 2004; Houk and Van Woesik, 2006; Molloy et al., 2013). This is echoed in Foster et al. (2014) who found that better survey designs were those that evenly dispersed sampling effort (images and transects) over environmental gradients in the survey area. We also found that for our simulated distributions, the image sampling strategy used did not confer any advantage in terms of the mean or standard error of cover estimates. This result was unexpected as it was

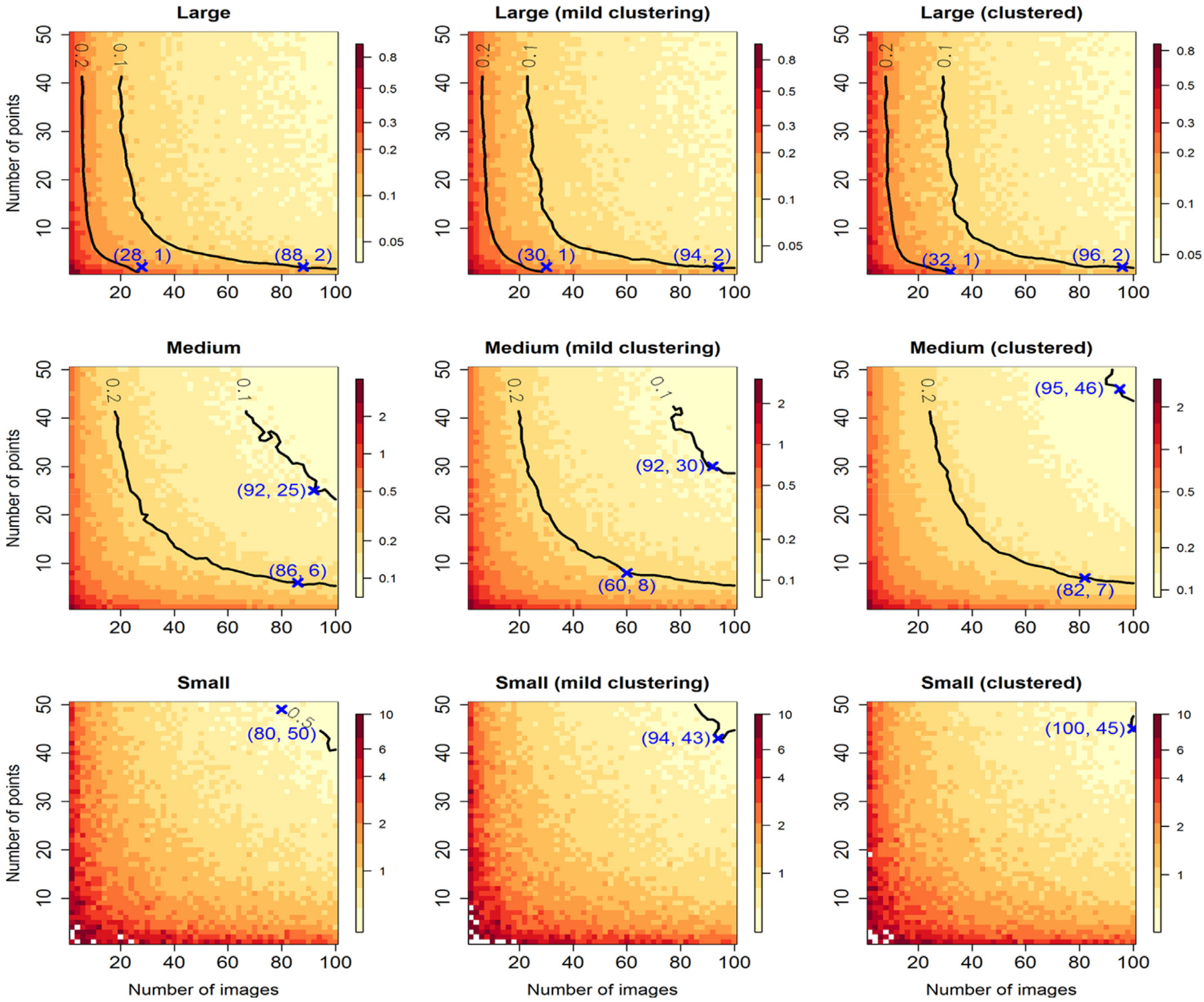


Fig. 2. The relationship between the number of images, number of points and Coefficient of Variation (CV) for the nine simulated biological distributions. The colour scale represents the CV, with lower CVs representing more precise estimates. Reference contour lines are included, with blue crosses and corresponding numbers (images, points) indicate the optimum sampling scenarios in terms of the minimum number of points needed to achieve the respective CVs. Images were selected systematically along the transect. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Optimum scoring in terms of the number of images and points per image used to obtain CV's of 0.1 and 0.2 for the large and medium distributions. Optimum scoring was determined by calculating the minimum total number of points required to achieve the relevant CVs.

Biological distribution	Optimum number of images/points (total) for CV = 0.2	Optimum number of images/points (total) for CV = 0.1
Large	28/1 (28)	88/2 (176)
Large (mildly clustered)	30/1 (30)	94/2 (188)
Large (clustered)	32/1 (32)	96/2 (192)
Medium	86/6 (516)	92/25 (2300)
Medium (mildly clustered)	60/8 (540)	92/30 (2760)
Medium (clustered)	82/7 (574)	95/46 (4370)

anticipated that a spatially balanced design may outperform a random design by more evenly sampling the area, particularly for the clustered distributions.

4.1. Points versus images

Relatively few points per image were required for sampling designs to exhibit little to no bias in estimates of biological groups that have greater than 5% coverage across the survey area (i.e. our medium and large groups). Similarly, Dumas et al. (2009) found that varying the number of points per image between 9 and 99 had little effect on bias in 20 m long transects where scored images were contiguous along the entire length of the transect. We found that this result also holds true for much longer transects and for a

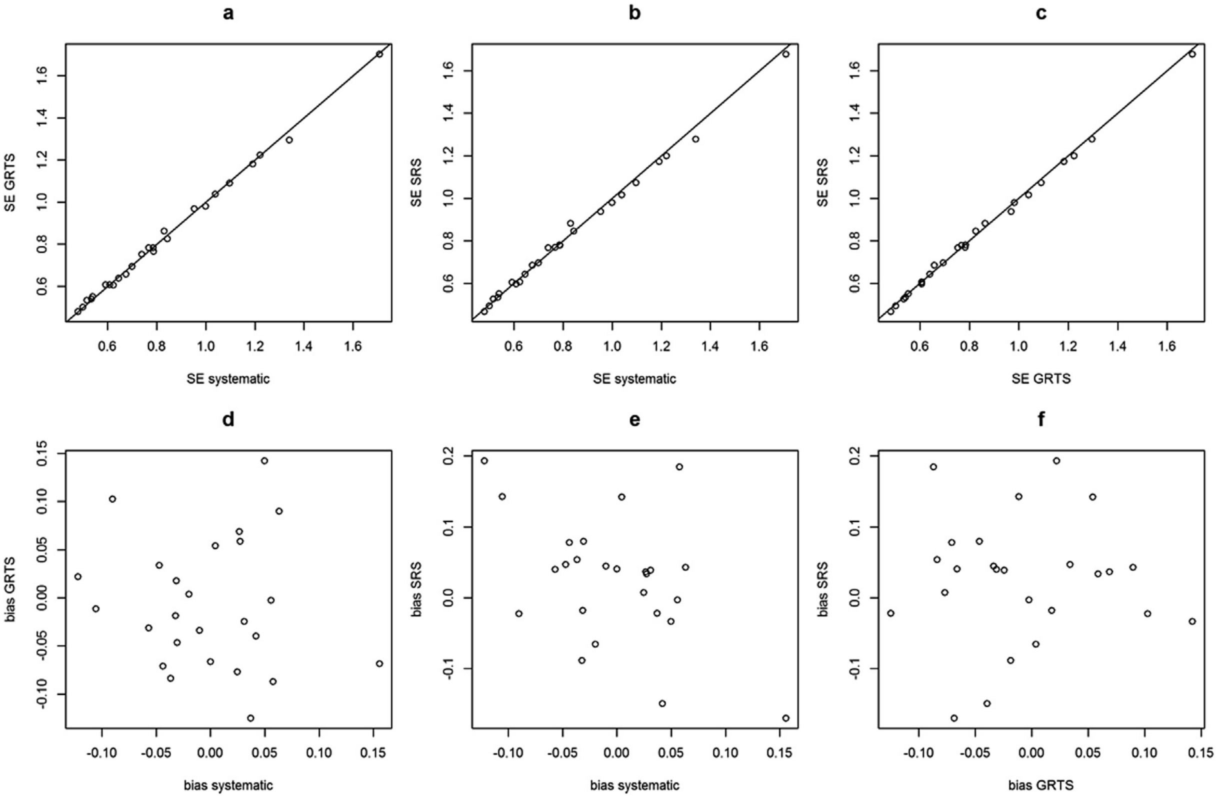


Fig. 3. Comparison of (i) standard errors (SE) using different image sampling protocols: (a) systematic vs GRTS, (b) systematic vs SRS, and (c) GRTS vs SRS; and (ii) bias using (d) systematic vs GRTS, (e) systematic vs SRS and (f) GRTS vs SRS. The biological distribution used was the medium (clustered) distribution. The number of images used range from 20 to 100 by 20, and the number of points per image from 10 to 50 by 10. Each point in the plots represents a unique combination of images and points. The lines in the SE plots represent a one-to-one relationship.

variety of spatial distributions. This shows that given adequate sampling intensity for an organism of interest, the point count approach is capable of providing unbiased estimates of percent cover for a surveyed area.

We also found a general pattern that for the distributions that could be sampled with a reasonably high precision (i.e. our large and medium distributions), the optimum sampling strategy, in terms of reducing bias and increasing precision, involved an increased number of images rather than an increased number of points per image. There is also support for our findings from a number of other studies using marine benthic imagery, which have found that spacing out and increasing sampling units, either images

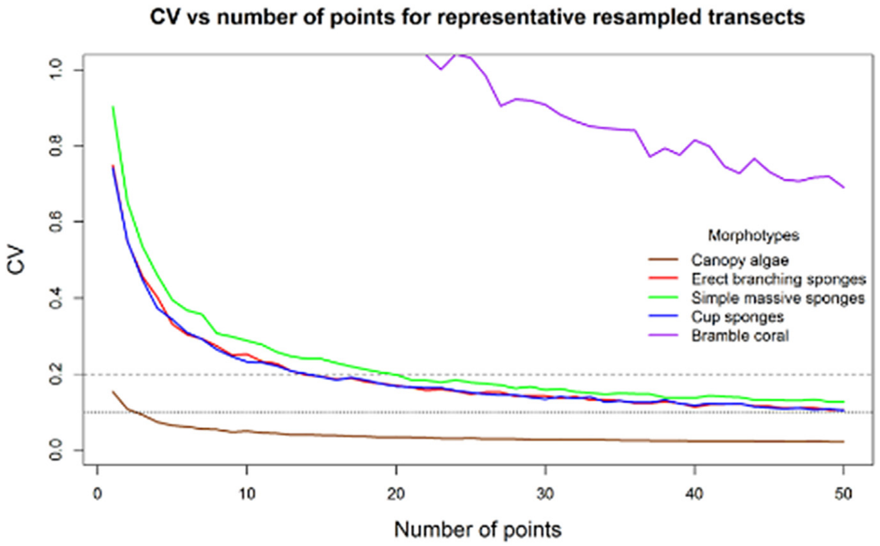


Fig. 4. Relationship between the coefficient of variation (CV) and the number of points per image used to score 5 "morphotypes" from 5 representative transects from the East Tasman Peninsula, Tasmania. Transects were scored by selecting every 100th image and scoring with 50 random overlaid points. The heavily dashed line represents a CV of 0.2. The lightly dashed line represents a CV of 0.1. The crossing points of the CV lines with these dashed lines represents the number of points per image required to achieve the relevant CV for that morphotypes within the representative transect.

within transects, or an increased number of transects over the study region, shows improvements in the precision of estimates (Brown et al., 2004; Ryan, 2004; Houk and Van Woesik, 2006; Molloy et al., 2013). We expand on the findings of previous studies by considering a range of biological characteristics of potential organisms of interest. Our results hold for more abundant groups (medium and large) regardless of whether their distribution is clustered or not. Whilst Fig. 2 shows that a reasonably high number of points (>25–46) were required to achieve a CV = 0.1 for the medium distributions, optimum samples were also at the high end of the number of images we tested (>90 in all cases). The slopes of the contour lines suggest that if an even larger number of images was tested that optimum sampling in terms of the total number of points scored would have favoured a larger number of images with less points. Reliably precise estimates fell outside of the scope of sampling that we tested for our small-low groups, and we were therefore unable to test whether this pattern held true for rarer organisms. Point count sampling approaches are unlikely to be adequate for rare organisms, as even with a large number of images or points within images, they will often be missed (Thompson, 2004).

By using the total number of points scored as our measure of optimum sampling we have not taken into account the fact that scoring more images may result in increases in scoring time as opposed to scoring more points within an image. We anticipate that this time would be minimal, particularly as software becomes more powerful and user-friendly, although these assumptions require further testing.

We note that, using simplifying assumptions, sampling theory provides an algebraic solution for calculating the standard error of the binomial proportion within an image (see Appendix). This suggests that the number of points required to minimise error within an image could be found algebraically, and would be dependent on the proportion of space occupied by an organism within an image. When generalising to multiple images however, the simple form of this approach requires further assumptions, such as independence (and identical distribution) of the percent cover in an image. This assumption will not match reality for systematic sampling and GRTS methods, and also for biota with clustered spatial distributions. We note that over-dispersion can be incorporated into these algebraic expressions, through a hierarchical representation of the data generating mechanism (giving a beta-binomial distribution), but this is a representation of mathematical convenience and not of biological realism. While these algebraic results may provide a workable solution for some researchers in some circumstances, we believe that our results based upon simulation of sampling outcomes for different distributions provide more readily interpretable results, which are more biologically believable and robust to unnecessary assumptions.

Whilst some studies have shown that a large number of points per image (up to 100) may be needed to capture the diversity of biota within an image (Van Rein et al., 2011; Deter et al., 2012), the effects of increasing the number of images when estimating cover of biota over an area is often not considered. For example, Deter et al. (2012) found that a minimum of 64 points per image were necessary to capture all recognisable features within an image. However, such results are unlikely to be generalizable as they will depend on factors such as the size of the organisms present and the size of the image. While Van Rein et al. (2011) found that a higher number of points per image (100) detected more taxa within the image, they also showed that an increased number of images with less points (25) was likely to increase the number of taxa recorded per unit effort. Our simulation showed that at a reasonable level of sampling intensity (greater than approximately 20 images with 5 points per image) even our small-low groups, which had coverage of only 0.1%, were regularly detected within transects, albeit with relatively high variance in estimates. When planning monitoring programs it is therefore necessary to carefully consider the effects of size, abundance and patterns of distribution on the precision of estimates. We therefore advocate the use of baseline surveys to establish these key properties of potential target biota, particularly in previously unsurveyed areas.

4.2. Size, abundance and detecting change

With chronic disturbances such as climate change, long-term fishing pressure, pollution and introduced species potentially affecting many benthic communities, detecting the impacts these disturbances have on the cover of organisms over time is often a primary aim of monitoring programs (Bernstein and Zalinski, 1983; Smale et al., 2012). Estimates of the cover of organisms with lower abundances often have higher variances (Houk and Van Woesik, 2006; Smale et al., 2012). This will result in lower statistical power to detect change in the cover of such organisms unless surveys include increased spatial and/or temporal replication. Therefore, when considering candidate biota for long-term monitoring projects, the cover of that organism across the region of interest should be considered in addition to their potential to respond to pressures. For example Stoddart et al. (2005) found that the power to detect a 10% change in cover was heavily dependent on the absolute cover of coral at a site, with coverage greater than 60% at a site requiring only 5 points per image, whereas when cover was less than 30% up to 50 points per image may be required. Based on our simulation and resampling analysis we would suggest that organisms that fall in the range of sizes and abundances between our large (i.e. 25–80 cm diameter with 40% cover) and medium (i.e. 10–25 cm diameter with 5% cover) classes are able to be sampled with point count methods at a high enough precision for long-term monitoring projects. Organisms that fall into this category may be a relatively small subset of the total pool of candidate morphotypes, and pilot surveys made need to be conducted in order to determine likely candidate species. For example, using the levels of sampling employed in this study, the cover of only 4 morphotypes (the algal and sponge morphotypes) out of 30 in our real world data set could be estimated with a CV of 0.1.

In contrast, organisms that are small and have low abundance/coverage across the study region (i.e. are rare) are unlikely to be good candidates for long-term monitoring projects employing point count methods. Where such organisms are the focus of surveys, an alternative sampling strategy should be employed. For example, image-level observations (e.g. presence-absence), which do not require percentage estimates, or more intensive image and point sampling protocols, stratified sampling based on known covariates or some form of adaptive sampling (Thompson, 1996). Conversely, our findings show that where large conspicuous organisms are the focus of monitoring effort, that low within image sampling (2 points) will give reliable estimates provided a sufficient number of images are used. This indicates that if such groups were the sole focus of a survey that lower resolution methods may allow much larger areas to be surveyed. For example, visibility permitting, AUVs, ROVs or towed systems could be deployed at a higher altitude off the bottom in order to characterise dominant macroalgal groups in temperate waters or corals in tropical waters.

4.3. Patterns of distribution and sampling

Species are rarely dispersed uniformly in nature, and monitoring programs need to take this spatial heterogeneity into account (Pielou, 1974; Miller and Ambrose, 2000). Clumping of the distribution of an organism within an area can result in an overall reduction in sampling precision (Andrew and Mapstone, 1987). In some cases the sampling design may induce bias into the study's results. For example, using systematic sampling (e.g. every nth image) there is the possibility of confounding the sampling pattern with spatial oscillations in the substrate or scales of distribution of organisms (Ryan, 2004). For our simulation, we found a larger sample size was required to achieve an equivalent level of accuracy (i.e. bias – see supplementary materials S1) and precision (Fig. 2) for all clustered distributions when compared to random distributions with the same size and abundance. Here we have simulated three potential levels of clustering, and note that spatial patterns in target organisms can have a strong influence on the level of sampling required to reach high levels of precision. Our clusters are distributed across our site, whereas in real benthic environments clusters may be further restricted by environmental gradients or habitat availability. Other studies of the marine benthos have shown that there is often large variability in individual species abundances over small scales (from centimetres to metres) (Underwood et al., 2000; Ysebaert and Herman, 2002; Fraschetti et al., 2005). Where possible, future studies and simulations should incorporate distributional properties of their organisms of interest based on realistic data gained from pilot studies.

For our simulated distributions none of the image sampling protocols offered any relative advantage (Fig. 3). When systematic trends exist in the distribution of biota then spatially balanced image selection approaches such as GRTS may provide an advantage. When prior knowledge exists of distinctly different sections within a study site, stratification may provide another approach to reduce spatial variability among sampling units (Andrew and Mapstone, 1987). For example, this could involve the stratification of sites into areas that contain dominant habitat forming groups with high cover, and other areas and allocating sampling intensities based on this stratification.

5. Conclusion

The use of imagery as a benthic survey tool is set to expand in the future as novel technologies provide increasingly powerful tools capable of conducting surveys over large areas in a relatively short amount of time. This makes survey platforms such as AUVs ideal candidates for broad-scale monitoring. Here we provide practical advice to scientists and managers conducting such surveys in terms of both sampling protocols necessary to give precise estimates and also likely candidate species or groups for ongoing monitoring. Our findings indicate that increases in precision over transects can best be gained through a more even and comprehensive sampling of the area by including a larger number of images rather than sampling more points within images. Where no prior information is available to stratify the sampling of images, whether images are selected in a random, systematic or spatially balanced manner is unlikely to affect estimates. These outcomes are generalizable across a number of size classes, abundances and distributions and are likely to hold true for broad-scale monitoring surveys that focus on conspicuous functional groups which are not too small and typically have cover greater than 5% over a site. We suggest such groups should be focussed upon as candidates for monitoring programs, as the sampling effort required to gain reliable estimates of their coverage is not likely to be overly onerous. Target indicators that are dominant space occupiers in a system may be effectively monitored at lower resolutions, by deploying image capturing devices such as AUVs at higher altitudes allowing greater spatial coverage. Patterns of distribution, in particular the level of clustering of organisms of interest, will also be important in determining the sampling intensity required. When establishing long term monitoring programs, we would suggest that pilot surveys should be conducted or expert knowledge incorporated in order to determine the size, abundance and distribution of potential candidate groups. Our findings can then be utilised as a basis for determining the likely range of image and point scoring protocols for the ongoing monitoring of these groups.

Acknowledgements

We would like to thank Justin Hulls, Lisa Meyer, David Peel and Geoff Hosack. This work was supported by the Marine Biodiversity Hub, a collaborative partnership supported through funding from the Australian Government's National Environmental Research Program (NERP). NERP Marine Biodiversity Hub partners include the Institute for Marine and Antarctic Studies, University of Tasmania; CSIRO Wealth from Oceans National Flagship, Geoscience Australia, Australian Institute of Marine Science, Museum Victoria, Charles Darwin University and the University of Western Australia. The IMOS program supported the AUV deployments via the Australian Centre for Field Robotics at the University of Sydney.

Appendix

An algebraic solution to binomial sampling of points within images

Let X_i be the number of points that fall on the organism of interest in image *i*. Let k be the number of points on an image. Then assume that $X_i \sim Binomial(p_i, k)$, which implies that $\mathbb{E}(X_i) = kp_i$ and $\mathbb{V}ar(X_i) = kp_i(1 - p_i)$. Now, let $\widehat{p}_i = \frac{X_i}{k}$, so that $\mathbb{E}(\widehat{p}_i) = p_i$ and $\mathbb{V}ar(\widehat{p}_i) = \frac{1}{k}p_i(1 - p_i)$. This gives the standard error for the proportion within an image to be $(\widehat{p}_i) = \sqrt{\frac{p_i(1-p_i)}{k}}$.

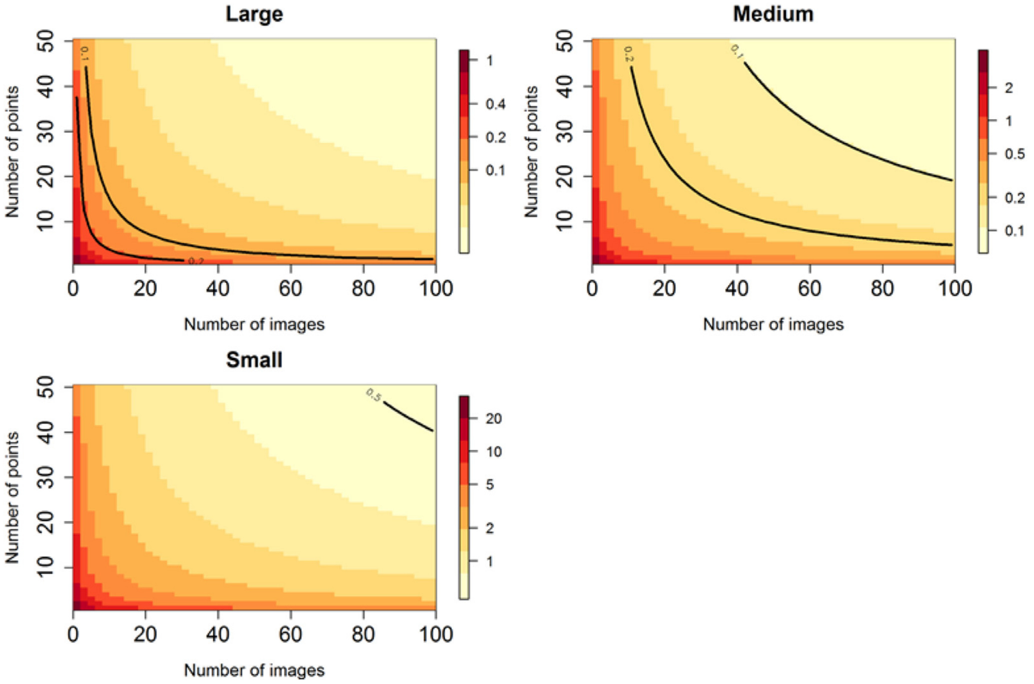
An algebraic solution to binomial sampling of multiple images

Now, let X_i, $i = 1...N$, be the number of points where there is a ‘success’ (i.e. a point falls on an organism of interest) in each of N images. Each image has k points in total. Define the mean number of ‘success’ points to be $\overline{X} = \frac{1}{N}\sum X_i$ and define the mean percentage cover to be $\overline{\widehat{p}} = \frac{1}{N}\sum\frac{X_i}{k}$. For this work, we require an estimate of the standard error of the mean percentage $\overline{\widehat{p}}$. Now,

$$\begin{aligned}\mathbb{V}ar(\overline{\widehat{p}}) &= \frac{1}{N^2k^2}Var\left(\sum X_i\right)\\ &= \frac{1}{N^2k^2}\,N\,\mathbb{V}ar(X_i),\\ &= \frac{1}{Nk}p(1-p),\ \text{giving}\end{aligned}$$

$$SE(\overline{\widehat{p}}) = \sqrt{\frac{1}{Nk}p(1-p)}$$

Coefficient of variation plots predicted by the algebraic solution are shown in the figure below. The algebraic solution can be seen to provide a similar outcome to the simulated random distribution. This is expected, as the algebraic solution makes the assumption that observations across images are independent and identically distributed, such as would occur under a random distribution. The simple algebraic solution is likely to underestimate the sampling required to reach a certain level of error where this assumption is violated, such as when clustering occurs. In such cases, the additional variation not accounted for by the model results in over-dispersion.



Incorporating Over-dispersion

If there is over-dispersion, possibly due to clustering, and if that over-dispersion can be encapsulated as a beta-binomial distribution, then we can assume that $X_i \sim \mathit{Beta\ Binomial}(p,\ k,\ \phi)$ where ϕ is an over-dispersion parameter so that $\mathbb{V}ar\left(\frac{X_i}{k}\right) = \frac{\mu\ (1-\ \mu)}{k}\phi$, which leads to (using the same argument as before) SE $(\overline{\hat{p}}) = \sqrt{\frac{\phi p(1-p)}{Nk}}$.

This approach requires the specification of ϕ, a quantity which will vary between organisms and systems. The interpretation of this parameter in relation to the observed biological distribution may not be readily apparent to ecologists. For this reason we have taken the simulation approach which relates sampling outcomes directly to potential distributions in terms of size, abundance and clustering.

Where simulation is not possible, or not immediately possible, the algebraic approach may offer a useful minimum starting point for survey design.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.ecss.2016.04.005.

References

Althaus, F., Hill, N.A., Rees, R., Jordan, A., Colquhoun, J., 2013. CATAMI standardised classification scheme visual look-up guide. In: Edwards, L. (Ed.), Collaborative and Automative Tools for Marine Imagery (CATAMI).

Andrew, N.L., Mapstone, B.D., 1987. Sampling and the description of spatial pattern in marine ecology. Oceanogr. Mar. Biol. 25, 39–90.

Barrett, N., Seiler, J., Anderson, T., Williams, S.B., Nichol, S., Hill, N.A., 2010. Using an autonomous underwater vehicle to inform management of biodiversity in shelf waters. In: Proceedings of IEEE Oceans 2010. Sydney, Australia.

Bernstein, B.B., Zalinski, J., 1983. Optimum sampling design and power tests for environmental biologists. J. Environ. Manag. 16, 35–43.

Bros, W.E., Cowell, B.C., 1987. A technique for optimizing sample-size (replication). J. Exp. Mar. Biol. Ecol. 114, 63–71.

Brown, C.J., Smith, S.J., Lawton, P., Anderson, J.T., 2011. Benthic habitat mapping: a review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. Estuar. Coast. Shelf Sci. 92, 502–520.

Brown, E., Cox, E., Jokiel, P., Rodgers, K., Smith, W., Tissot, B., Coles, S.L., Hultquist, J., 2004. Development of benthic sampling methods for the coral reef assessment and monitoring program (CRAMP) in Hawai'i. Pac. Sci. 58, 145–158.

Deter, J., Descamp, P., Boissery, P., Ballesta, L., Holon, F., 2012. A rapid photographic method detects depth gradient in coralligenous assemblages. J. Exp. Mar. Biol. Ecol. 418, 75–82.

Diggle, P.J., 1983. Statistical Analysis of Spatial Point Patterns. Academic Press, London.

Dormann, C.F., McPherson, J.M., Araujo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kuhn, I., Ohlemuller, R., Peres-Neto, P.R., Reineking, B., Schroder, B., Schurr, F.M., Wilson, R., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30, 609–628.

Dumas, P., Bertaud, A., Peignon, C., Leopold, M., Pelletier, D., 2009. A “quick and clean” photographic method for the description of coral reef habitats. J. Exp. Mar. Biol. Ecol. 368, 161–168.

Foster, S.D., Hosack, G.R., Hill, N.A., Barrett, N.S., Lucieer, V.L., Spencer, M., 2014. Choosing between strategies for designing surveys: autonomous underwater vehicles. Methods Ecol. Evol. 5, 287–297.

Fraschetti, S., Terlizzi, A., Benedetti-Cecchi, L., 2005. Patterns of distribution of marine assemblages from rocky shores: evidence of relevant scales of variation. Mar. Ecol. Prog. Ser. 296, 13–29.

Halpern, B.S., Walbridge, S., Selkoe, K.A., Kappel, C.V., Micheli, F., D'Agrosa, C., Bruno, J.F., Casey, K.S., Ebert, C., Fox, H.E., Fujita, R., Heinemann, D., Lenihan, H.S., Madin, E.M.P., Perry, M.T., Selig, E.R., Spalding, M., Steneck, R., Watson, R., 2008. A global map of human impact on marine ecosystems. Science 319, 948–952.

Hill, J., Wilkinson, C., 2004. Methods for Ecological Monitoring of Coral Reefs. AIMS, Townsville, QLD, p. 123.

Hill, N.A., Barrett, N., Lawrence, E., Hulls, J., Dambacher, J.M., Nichol, S., Williams, A., Hayes, K.R., 2014. Quantifying fish assemblages in large, offshore marine protected areas: an Australian case study. Plos One 9, e110831.

Houk, P., Van Woesik, R., 2006. Coral reef benthic video surveys facilitate long-term monitoring in the commonwealth of the Northern Mariana Islands: toward an optimal sampling strategy. Pac. Sci. 60, 177–189.

Hughes, T.P., Bellwood, D.R., Folke, C., Steneck, R.S., Wilson, J., 2005. New paradigms for supporting the resilience of marine ecosystems. Trends Ecol. Evol. 20, 380–386.

Kohler, K.E., 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. Comput. Geosci. 32, 1259–1269.

Leujak, W., Ormond, R.F.G., 2007. Comparative accuracy and efficiency of six coral community survey methods. J. Exp. Mar. Biol. Ecol. 351, 168–187.

Librino, F., Levorato, M., Zorzi, M., 2014. An algorithmic solution for computing circle intersection areas and its applications to wireless communications. Wirel. Commun. Mob. Comput. 14, 1672–1690.

Miller, A.W., Ambrose, R.F., 2000. Sampling patchy distributions: comparison of sampling designs in rocky intertidal habitats. Mar. Ecol. Prog. Ser. 196, 1–14.

Molloy, P.P., Evanson, M., Nellas, A.C., Rist, J.L., Marcus, J.E., Koldewey, H.J., Vincent, A.C.J., 2013. How much sampling does it take to detect trends in coral-reef habitat using photoquadrat surveys? Aquat. Conserv. Mar. Freshw. Ecosyst. 23, 820–837.

Pielou, E.C., 1974. Population and Community Ecology: Principles and Methods. Gordon and Breach Science Publishers, London.

Pizarro, O., Williams, S.B., Jakuba, M.V., Johnson-Roberson, M., Mahon, I., Bryson, M., Steinberg, D., Friedman, A., Dansereau, D., Nourani-Vatani, N., Bongiorno, D., Bewley, M., Bender, A., Ashan, N., Douillard, B., 2013. Benthic Monitoring with Robotic Platforms-The Experience of Australia (Tokyo).

Press, W.H., 2007. Numerical recipes : the art of scientific computing/William H. Press... [et al.], third ed. Cambridge University Press, New York.

Ryan, D.A.J., 2004. Point sampling strategies for estimating coverage from benthic video transects. Environmetrics 15, 193–207.

Smale, D.A., Kendrick, G.A., Harvey, E.S., Langlois, T.J., Hovey, R.K., Van Niel, K.P., Waddington, K.I., Bellchambers, L.M., Pember, M.B., Babcock, R.C., Vanderklift, M.A., Thomson, D.P., Jakuba, M.V., Pizarro, O., Williams, S.B., 2012. Regional-scale benthic monitoring for ecosystem-based fisheries management (EBFM) using an autonomous underwater vehicle (AUV). ICES J. Mar. Sci. 69 (6).

Stevens, D.L., Olsen, A.R., 2004. Spatially balanced sampling of natural resources. J. Amer. Stat. Assoc. 99 (465), 262–278.

Stoddart, J.A., Blakeway, D.R., Grey, K.A., Stoddart, S.E., 2005. Corals of the Dampier Harbour: Their Survival and Reproduction during the Dredging Programs of 2004. MScience Pty Ltd, Perth WA, p. 78.

Thompson, S.K., 1996. Adaptive Sampling. Wiley.

Thompson, W.E., 2004. Sampling Rare or Elusive Species. Island Press.

Trygonis, V., Sini, M., 2012. photoQuad: a dedicated seabed image processing software, and a comparative error analysis of four photoquadrat methods. J. Exp. Mar. Biol. Ecol. 424–425, 99–108.

Underwood, A.J., Chapman, M.G., Connell, S.D., 2000. Observations in ecology: you can't make progress on processes without understanding the patterns. J. Exp. Mar. Biol. Ecol. 250.

Van Rein, H., Schoeman, D.S., Brown, C.J., Quinn, R., Breen, J., 2011. Development of benthic monitoring methods using photoquadrats and scuba on heterogeneous hard-substrata: a boulder-slope community case study. Aquat. Conserv. Mar. Freshw. Ecosyst. 21, 676–689.

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

Warton, D.I., Hui, F.K.C., 2011. The arcsine is asinine: the analysis of proportions in ecology. Ecology 92, 3–10.

Williams, S.B., Pizarro, O.R., Jakuba, M.V., Johnson, C.R., Barrett, N.S., Babcock, R.C., Kendrick, G.A., Steinberg, P.D., Heyward, A.J., Doherty, P.J., Mahon, I., Johnson-Roberson, M., Steinberg, D., Friedman, A., 2012. Monitoring of benthic reference sites using an autonomous underwater vehicle. IEEE Robot. Autom. Mag. 19, 73–84.

Ysebaert, T., Herman, P.M.J., 2002. Spatial and temporal variation in benthic macrofauna and relationships with environmental variables in an estuarine, intertidal soft-sediment environment. Mar. Ecol. Prog. Ser. 244, 105–124.