Progress in Oceanography 222 (2024) 103214

Contents lists available at ScienceDirect

Progress in Oceanography

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

Improving coral monitoring by reducing variability and bias in cover estimates from seabed images

Emma J. Curtis [a,b,*], Jennifer M. Durden [c], Brian J. Bett [c], Veerle A.I. Huvenne [c], Nils Piechaud [d], Jenny Walker [b], James Albrecht [e], Miquel Massot-Campos [b], Takaki Yamada [b], Adrian Bodenmann [b], Jose Cappelletto [b], James A. Strong [c], Blair Thornton [b,f]

[a] School of Ocean and Earth Science, National Oceanography Centre, University of Southampton Waterfront Campus, Southampton, SO14 3ZH, UK
[b] Centre for In Situ and Remote Intelligent Sensing, University of Southampton, Southampton, SO16 7QF, UK
[c] Ocean BioGeosciences, National Oceanography Centre, European Way, Southampton, SO14 3ZH, UK
[d] Institute of Marine Research, P.O. Box 1870, Nordnes, 5817 Bergen, Norway
[e] Joint Nature Conservation Committee, 2 East Station Road, Fletton Quays, Peterborough, PE2 8YY, UK
[f] Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba Meguro-ku, Tokyo 153-8505, Japan

ARTICLE INFO

Keywords:
Underwater photography
Image annotation
Data imputation
Environmental monitoring
Conservation
Ocean floor

ABSTRACT

Seabed cover of organisms is an established metric for assessing the status of many vulnerable marine ecosystems. When deriving cover estimates from seafloor imagery, a source of uncertainty in capturing the true distribution of organisms is introduced by the inherent variability and bias of the annotation method used to extract ecological data. We investigated variability and bias in two common annotation methods for estimating organism cover, and the role of size selectivity in this variability. Eleven annotators estimated sparse cold-water coral cover in the same 96 images with both grid-based and manual segmentation annotation methods. The standard deviation between annotators was three times greater in the grid-based method compared to segmentation, and grid-based estimates from annotators tended to overestimate coral cover. Size selectivity biased the manual segmentation; the minimum size of colonies segmented varied between annotators fivefold. Two modelling techniques (based on Richard's selection curves and Gaussian processes) were used to impute areas where annotators identified colonies too small for segmentation. By imputing small coral sizes in segmentation estimates, the coefficient of variation between annotators was reduced by approximately 10%, and method bias (compared to a reference dataset) was reduced by up to 23%. Therefore, for sparse, low cover organisms, manual segmentation of images is recommended to minimise annotator variability and bias. Uncertainty in cover estimates may be further reduced by addressing size selectivity bias when annotating small organisms in images using a data-driven modelling technique.

1. Introduction

Ecological monitoring is critical to the conservation of habitats that are susceptible to anthropogenic impacts (Lindenmayer and Likens, 2010), and monitoring of ocean health is recognised as internationally important for environmental sustainability strategy (Danovaro et al., 2020; IOC-UNESCO, 2020). Monitoring has revealed significant changes to important seabed habitats: a three-decade decline in coral cover on the Great Barrier Reef (De'ath et al., 2012); a coral bleaching and disease event resulting in a 13% decline in coral cover in the Caribbean (Miller et al., 2009); ocean acidification weakening coral framework (Hennige et al., 2015); the destruction of sponge aggregations by bottom trawling (Vieira et al., 2020; Kazanidis et al., 2019); the slow recovery of cold-water corals from trawling (Clark et al., 2019; Huvenne et al., 2016); and the ingestion of plastics across the marine food web, even in remote areas such as the deep sea (Taylor et al., 2016). Such monitoring outcomes are important to conservation decision-making and environmental management, including the evaluation of Vulnerable Marine Ecosystems (VMEs) (Obura et al., 2019), the designation of Marine Protected Areas (MPAs) (Almany et al., 2009), and the management of industrial activities (Levin et al., 2020).

To be effective, monitoring data must be collected over ecologically-relevant spatial and temporal scales, and to commonly-agreed metrics. The spatial repeatability, larger areal coverage and greater flexibility

* Corresponding author at: School of Ocean and Earth Science, National Oceanography Centre, University of Southampton Waterfront Campus, Southampton, SO14 3ZH, UK.

E-mail address: ec9g15@soton.ac.uk (E.J. Curtis).

https://doi.org/10.1016/j.pocean.2024.103214
Received 7 April 2023; Received in revised form 7 November 2023; Accepted 31 January 2024
Available online 6 February 2024

of underwater imaging compared to traditional sample-return based survey methods has lead to the increased use of photographic and video surveys in seabed monitoring programmes (Morris et al., 2014; Brown et al., 2004; Williams et al., 2016). Image based surveys also provide more opportunity for better understanding deep-sea (>200 m water depth) VMEs, which often need to be monitored over large spatial and temporal scales due to their composite organisms having sparse distributions and long life histories (Clark et al., 2019; Vieira et al., 2020). Essential Ocean Variables have been developed as an international standard to establish baseline conditions and assess change (Constable et al., 2016; Miloslavich et al., 2018; Levin et al., 2019), with percentage cover as an established metric for assessing the state of habitats with colonial or seafloor-covering biota, including hard coral (Brown et al., 2004; Hill and Wilkinson, 2004; Obura, 2018). Methods for manually estimating percentage cover from images include grid-based and manual segmentation-based methods (Leujak and Ormond, 2007), in which researchers annotate, that is, locate, identify and size the target organisms. These methods are often contrasted by their efficiency, with manual segmentation considered to require more time than grid-based cover estimation, but with the accuracy and time efficiency of grid-based estimates being strongly linked to the grid size selected (Trygonis and Sini, 2012; Dethier et al., 1993).

The reliability of cover estimates is in part determined by the bias and variability associated with the image capture and annotation method (Sayer and Poonian, 2007). Variability in image-derived ecological data can result from the chosen image acquisition and processing techniques (Jones et al., 2007; Schoening et al., 2020), and from the annotation process. Annotation to produce robust cover estimates requires the correct detection and identification of target organisms, and the correct evaluation of size, each of which is subject to variation and bias. Annotation method choice may impact this variation; for example, Leujak and Ormond (2007) compared six different in situ and visual imaging survey methods resulting in coral cover estimates ranging from 34%–46% in a high cover coral reef in the Red Sea. Purser et al. (2009) estimated cold-water coral and sponge cover using multiple annotation techniques including an automated analysis technique and a manual point assessment technique at two different resolutions (15 points versus 100 points). They found the lower resolution technique (15 points) struggled to accurately detect sparsely distributed coral colonies and the autoanalysis technique underestimated abundant coral cover. Sponge cover estimates varied between methods, with both the 15 points technique and the autoanalysis producing inaccurate cover estimates. Perkins et al. (2022) also found that image annotation technique impacted the statistical power to monitor change in target organisms (such as sponge bleaching). Annotators vary in both detecting and identifying organisms. A comparison of eight annotators identifying dinoflagellate species found agreement from 43%–95%, with self-consistency levels on repeat annotations as low as 67% (Culverhouse et al., 2003). Beijbom et al. (2015) found both intra- and inter-annotator variability in identifying different substrata found in images captured at four Pacific coral reefs, and that annotator experience with the study site impacted annotator error. A study of fauna in seafloor images found that the majority of variability in the three annotators' faunal composition estimates derived from varying detection rather than disagreements of identity (Durden et al., 2016). In a dive survey, seabed cover estimates were found to vary between observers, with precision in estimates of seabed cover found to be related to quadrat size (Benedetti-Cecchi et al., 1996). The magnitude of such variability in cover estimates may be significant to the ecological conclusions drawn, particularly for sparsely-distributed organisms. Annotator variability may also obscure real environmental change, with implications for the designation of protected areas or implementation of policy measures.

Efforts have been made to provide guidance on and establish thresholds for the identification of VMEs from images. High density VMEs can at times be identified from a single image, however, when VME forming taxa are present at lower densities, their distribution at larger spatial scales is required to identify the VME (Baco et al., 2023). Cold-water coral habitat density is highly dependent on the composite taxa and the abiotic conditions found at sites, making it challenging to define a threshold for reefs and gardens (OSPAR Commission, 2008; Bullimore et al., 2013). Rogers et al. (2013) provide a coral colony density threshold of greater than 10 times the background densities (usually 0.1 m^{-2} but scale dependent) as their definition of coral gardens, however, Bullimore et al. (2013) found this criteria to exclude too many coral garden habitats due to the broad range of comprising biotopes. Price et al. (2019) and Rowden et al. (2020) both suggested a threshold of approximately 30% living or dead coral framework cover to establish a reef habitat, over spatial scales of 2 m^2 to 50 m^2. The designation of MPAs and other policy measures can also occur when VME forming taxa (such as cold-water corals) are present in unique environmental conditions (Huvenne et al., 2016). When density or cover estimates are used to define VMEs and influence management actions, uncertainty in estimates derived from seabed images due to annotator error and variability can impact the protection of these ecosystems.

We investigate uncertainty in cover estimates derived from seafloor photographs taken at a site with sparse coral cover, by comparing the annotation results of 11 annotators using two methods of cover estimation and additionally proposing two techniques to reduce annotator variability based on data-driven annotator modelling. We compare grid-based and manual segmentation image annotation methods, and assess the inter-annotator and method-based variability and bias in estimating sparse cold-water coral cover. We further explore inter-annotator variability in the manual segmentation method by assessing annotator bias in relation to coral colony size. To address the revealed inter-annotator size bias, we apply and evaluate two different modelling techniques for imputing missing coral size data. We discuss the ecological implications of the observed annotator variability, and the generalisable implications for routine monitoring of seafloor covering biota.

2. Methods

2.1. Study site and image collection

This study investigates annotator variability using seabed images gathered in the Western Darwin Mounds area in 2019 (Table 1; Huvenne and Thornton (2020)). The Darwin Mounds are a field of sparsely-distributed cold-water coral mounds, each up to approximately 75 m in diameter and 5 m in height. The mounds are unusual in hosting scleractinian corals, *Desmophyllum pertusum* and *Madrepora oculata*, in a sedimentary rather than rocky environment (similar to the Moira Mounds in the Porcupine Seabight; Wheeler et al. (2011)). The mounds are thought to have initially formed in the early Holocene, as a result of cold-water corals baffling sandy sediments from the prevailing contour current (Victorero et al., 2016). Although being sparse in cover, the coral framework on the mounds still provides habitats for commercially important fish species and various invertebrates (Bett, 2001; Costello et al., 2005). The Darwin Mounds effectively became the UK's first offshore MPA in August 2003 when the area was closed to bottom fisheries under the EU Common Fisheries Policy (Council of the European Union, 2004; De Santo and Jones, 2007). Since 2015, it has been designated as a Special Area of Conservation and the area remains a target of continued monitoring (Huvenne et al., 2016; Chaniotis et al., 2020).

Vertical still images were collected with a combined stereo camera and double laser line scanner system known as BioCam (West et al., 2020), mounted to the *Autosub6000* autonomous underwater vehicle. See Thornton et al. (2021) for details of the acquisition system. The survey over the Western Darwin Mounds captured 21,279 images in an area of approximately 29 hectares (Table 1). The images were processed to correct for colour, illumination beam pattern and lens distortion (Bodenmann et al., 2017), and were clustered into 11

Table 1
Western Darwin Mounds cold-water coral cover study image collection and annotation dataset preparation information. SD = standard deviation.

Study dates	16–17/09/2019
Study area (centre)	59.81°N 7.36°W
Water depth (m)	957–970
AUV mean altitude ± SD (m)	5.05 ± 0.67
Seabed area covered (m^2)	288 000
Mean image seabed area ± SD (m^2)	42.1 ± 12.0
Image mean spatial resolution ± SD (mm/pixel)	2.7 ± 0.4
Number of images collected	21 279
Number of images annotated	96
Total seabed area of annotated images (m^2)	3950
Number of annotators	11

visually similar groupings using unsupervised features extracted by a deep-learning location guided autoencoder (Yamada et al., 2020). Eight non-overlapping images were randomly selected from within each cluster to generate a cluster-stratified selection of seafloor images to function as the annotation dataset. To increase the number of corals present in our dataset, 16 rather than eight non-overlapping images were selected from the image cluster known to contain coral, creating an annotation dataset of 96 images in total.

2.2. Image annotation and cover estimation

Eleven annotators annotated the images in random order using the online annotation platform Squidle+ (Bewley et al., 2015). Manual annotation was repeated using two different methods: grid-based percentage estimation and manual segmentation. Each method was used to estimate the percentage cover of living scleractinian coral (defined as coral framework with living polyps, hereafter referred to as coral colonies) in each image. Prior to annotation, annotators received instructions on the chosen annotation methods and corals of interest, with practice on a smaller tutorial dataset containing 15 images captured in the same deployment but not included in the annotation dataset. Guidance included a minimum coral colony size to annotate, example annotations on the tutorial dataset and opportunities for feedback and to confer with other annotators. This work was carried out using annotators' own workspace setups.

In grid-based annotation, images were divided into 56 cells with an eight by seven grid to produce cell sizes of approximately 1 m^2 seabed area (320 × 308.6 pixel cell size). Annotators estimated the summed percentage cover of living and dead coral colonies found within each cell (Fig. 1a; 1b). Percentage cover estimates were given in 10% ranges (Jokiel et al., 2015). For example, if the annotator estimated around 10% of the annotated cell to be covered by living coral, the cell would be annotated with the following tags: live, 5 < x ≤ 15%. The lowest and highest cover ranges that could be reported were $0 < x \leq 5$ and $95 < x \leq 100$% respectively. If no scleractinian coral was present in a cell, the cell would be labelled as sediment only. Coral cover percentages for each cell were made proportional to the image by dividing the cover values by 56 and summed to produce a living coral cover estimate for the whole image. For manual segmentation, annotators drew polygons around all living coral colonies detected in each image (Fig. 1c; 1d). The areas of the drawn segments were summed to produce image cover estimates. Where colonies were too small for manual segmentation, annotators assigned a point indicating the presence of live coral, but did not provide a drawn segment from which colony size could be calculated from ("point label"; Fig. 1d). For both methods, annotators could add an additional "uncertain" label when they were not fully confident in their annotations. Uncertain annotations without a live or dead label were removed before analyses. For an example of the Squidle+ interface used for both annotation methods, see the supplementary material (Supplement 1).

Annotators were instructed to complete their annotations within one month, in as many sessions as required. The time taken for each annotator to fully annotate the 96 images with each method was recorded.

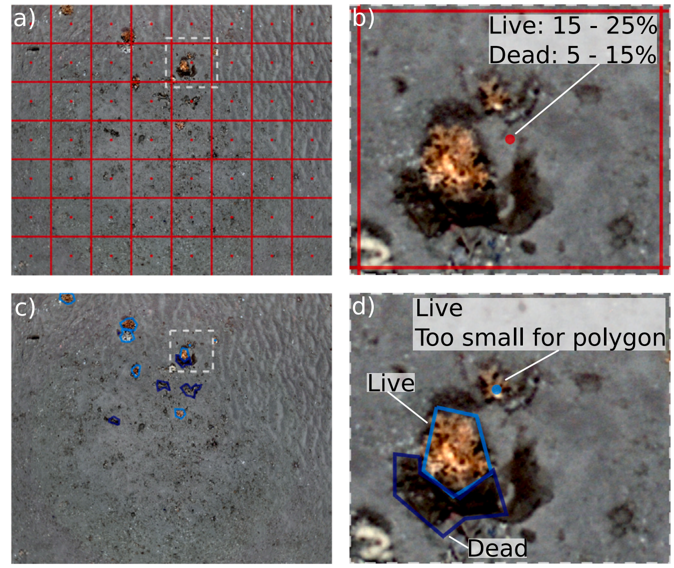

Fig. 1. Illustration of the image annotation methods used to estimate coral cover: grid-based (a; b) and manual segmentation (c; d). A whole image is presented in (a) and (c), and a single grid cell (b) and (d). Coral colonies are shown approximately to scale.

The pixel dimensions and seabed area of each image were calculated using *Autosub6000*'s altitude and the camera acceptance angles (69.5° and 60.7°), following Jones et al.'s (2007) approach for vertical seabed images scaling calculation (Jamieson et al., 2013; Durden et al., 2015; Piechaud and Howell, 2022). Image coral cover estimates were converted into real-world spatial cover estimates assuming a flat-world scenario. Any protrusion of coral and other three dimensional structures from the surface of the seabed was assumed to be negligible, as the AUV's mean flying altitude during image acquisition was large (approximately 5 m) in comparison to the relatively small height of the coral structures (estimated to be < 1 m and likely < 0.5 m for the coral colonies included in the present study) in the Darwin Mounds area (Huvenne and Thornton, 2020).

2.3. Analyses of variability in annotator cover estimation

Annotator variability was assessed between annotation methods by comparing the cover estimates per image. Differences between annotators in coral cover estimates for each annotation method were assessed using the Friedman rank-sum tests adjusted for ties, where the annotator identity (1–11) was the treatment condition and the image identifier (1–96) was the blocking condition, (Hollander et al., 2013; Vallat, 2018). Differences in coral cover estimates per image between annotators and between the grid-based annotation method and segmentation were also assessed by a Friedman test (identifier combining annotator identity and annotation method used as 1–22 treatment condition, image identifier 1–96 blocking condition). Post-hoc pairwise Conover tests on the method-annotator cover estimates were performed using a false discovery rate p-value adjustment method (Conover, 1999; Benjamini and Hochberg, 1995; Terpilowski, 2019).

Coral percentage cover was estimated across all the images used in this study for each annotator and annotation method. To generate confidence intervals for the annotator cover estimates, image cover estimates made by each annotator with each annotation method were pooled separately and randomly resampled with replacement to create image sets with total imaged seabed areas near equal to the original total image seabed area photographed (3950.44 m^2). The measured areas of coral colonies present in the sampled image sets were summed to give a total living coral area, which was divided by the total imaged area to create standardised living coral percentage cover estimates. Resampling was repeated 1000 times to determine the area-weighted mean of coral percentage cover and 95% confidence intervals. Confidence intervals were calculated using the bias corrected and accelerated method to account for bias and skew present in the resampled coral cover distributions (Roff, 2001). Percentage cover was estimated both for each annotator and across all annotators. For comparisons across all annotators, image estimates made by all annotators were pooled together for each annotation method and resampled with replacement. Annotator bias within each method were evaluated by calculating the percentage difference between the area-weighted mean cover estimate for each annotator and the mean cover estimate across all annotators.

Coral percentage cover estimates and annotation speed were compared between methods with Wilcoxon signed-rank tests. Annotation time, number of detected colonies and number of drawn segments in the manual segmentation method was tested for correlation with Kendall rank correlations.

2.4. Analyses of method bias

Method bias was evaluated with the area-weighted mean coral cover estimated across all annotators for each annotation method. A reference dataset combining all 11 annotators' drawn segments was used to compare each annotation method against. To create this reference dataset, the mean area of drawn segments around any coral colony segmented by at least one annotator was used (note, this is likely a slight overestimate of coral cover; Chalana and Kim (1997), Tong et al. (1998)). The difference between the area-weighted mean coral cover estimated by the reference dataset and by each annotation method were then standardised to give the percentage differences from the reference dataset. One sample Wilcoxon signed rank tests were performed to evaluate whether annotators' area weighted mean coral cover estimates were significantly different to the coral cover estimated by the reference dataset. Combining all annotators also produced the number of coral detections made by at least one annotator, which was then used to estimate apparent detection success in segmentation (Durden et al., 2016).

2.5. Reducing uncertainty in manual segmentation image annotation

To create more complete coral cover estimates and reduce variability between annotators in the segmentation method, we employed two techniques to model the colony size distributions of each annotator and impute the missing coral area data. The first is based on established sample selection relationships used in fishery studies (Richard's selection curves), with bootstrapping to characterise uncertainty within each annotator's model. The second employed constrained Gaussian processes regressions (GP; Agrell (2019)), which make model predictions together with uncertainty estimates.

Each annotators' segmented coral sizes were plotted against their relative cumulative frequencies (see Supplement 2 for examples of these plots) and both modelling techniques were applied separately. The techniques were implemented with a defined minimum size for the modelled area, set to a circular area with the diameter of a single corallite (10 mm; Gass and Roberts (2011); three times the resolution of the acquisition system). To remove potential outliers from the modelled data, in cases where the smallest colony area drawn by an annotator was inconsistent with the general distribution trend of the larger colonies drawn, those values were reset to point labels and the corresponding area estimated by the imputation process. For more information on the removal process, refer to the supplementary material (Supplement 2).

2.5.1. Imputing coral sizes - modelling using a known relationship

Selection curves are often used in fishery studies to provide the probability that a fish of a certain length is retained if they make contact with trawl gear (Wileman et al., 1996; Millar and Fryer, 1999). Following the logic that larger organisms have a greater probability of being retained (or here, detected) in a trawl (in an image), the selection curve tends to resemble an S-shape (Millar and Fryer, 1999; Stepputtis et al., 2016). To impute the point labelled coral sizes, the relative cumulative frequencies of coral colonies and colony areas were modelled using the inverse of a selection curve. Inverse Richard's selection curves (IRC) were chosen to describe the coral areas drawn by each annotator, where $r(l)$ was the relative cumulative frequency of a coral colony with a natural logarithm of real-world area l being detected by an annotator:

$$l = \frac{ln(\frac{r(l)^{\delta}}{1-r(l)^{\delta}}) - a}{b} \tag{1}$$

The equation is specified by the parameters a and b and includes an additional parameter, δ, to account for varying degrees of asymmetry in the distribution (Millar and Fryer, 1999). The natural logarithm of colony areas were used to ensure only positive values of colony sizes were predicted. The natural logarithm of colony areas from each annotator and the minimum corallite area with a relative cumulative frequency of zero were combined and bootstrapped 1000 times to generate parameter estimates using the curve_fit function in the Scipy Python module (Virtanen et al. (2020); Supplement 3). The resulting bootstrapped parameter estimates for the IRCs (a, b and δ) were used to estimate point label sizes by providing randomly selected relative cumulative proportion $r(l)$ values between the inverse of the total number of labels an annotator made, and the proportion of point labels for each annotator. These values were used along with the bootstrapped parameters to generate a range of possible coral colony areas.

2.5.2. Imputing coral sizes - non-parametric modelling

The relative cumulative frequencies of each annotator's segmented coral colonies were modelled with respect to their real-world sizes using non-parametric Bayesian inference known as GPs (Rasmussen and Williams, 2006). Whilst GPs do not require a known relationship prior to modelling, some initial hyper-parameters are needed which influence the resulting modelled data. The scleractinian coral taxa in this study show linear colony growth (Orejas et al. (2011); less than 4 cm yr^{-1}; Buscher et al. 2019) and due to there being little to no size selective pressure in the study area (the Western Darwin Mounds did not have major trawling damage; Huvenne et al. (2016)), the size population of imaged coral colonies was expected to show a smooth distribution. Function estimation was achieved using GPs with the radial basis function kernel, a common choice in ecological and growth studies to model smooth and stationary functions and easy to optimise (Rasmussen and Williams, 2006; Schulz et al., 2018). The radial basis function kernel has two hyper-parameters; the length-scale λ and the signal variance σ_f^2. Length-scales and signal variances values were chosen based on each annotator's data:

$$\text{Length-scale } \lambda = 0.25 \times prop_{point} \tag{2}$$

$$\text{Signal variance } \sigma_f^2 = (\frac{max_c - min_c}{4})^2 \tag{3}$$

Where $prop_{point}$ was the proportion of identified coral colonies with a point label, max_c was the maximum size of drawn coral colonies by an annotator and min_c was the minimum possible size of a coral colony. An additional parameter, noise variance σ_n^2 was included in the GP to represent the variance in coral area due to the difference in resolution when images were captured at different altitudes above the seabed, and was the same for all annotators (Supplement 4). Boundary constraints were used to improve GP predictions and produce more realistic estimates of model uncertainty (Agrell, 2019). In addition

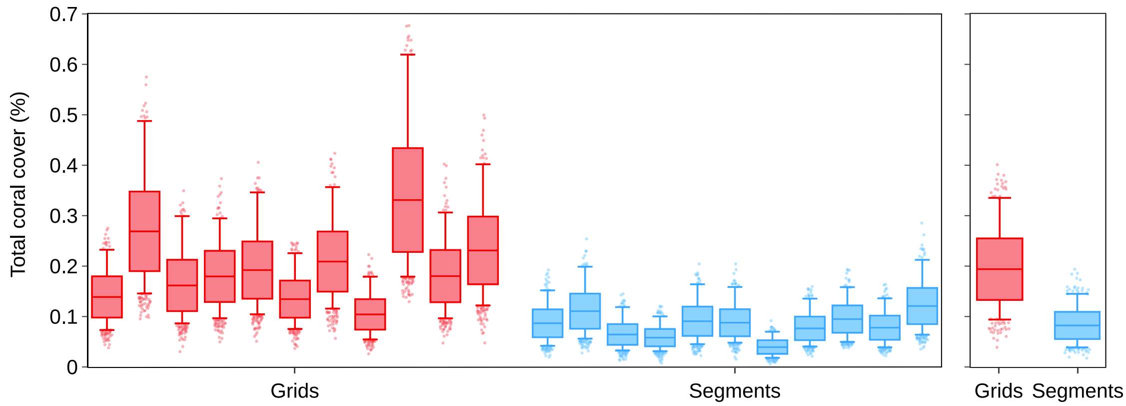

Fig. 2. Area-weighted mean coral cover over total study area estimated by each annotator (left) and across all annotators (right), using the grid-based and manual segmentation methods. Box plots show means and standard deviations (solid lines and boxes) with 95% confidence interval whiskers and points representing outliers.

to the minimum boundary for coral colony size of a single corallite, the maximum size of a coral colony segmented by each annotator was used as the corresponding upper bound for predicted values. The model function was also forced to monotonically increase so that the coral colony size increased with the relative cumulative frequency of annotated corals. The constrained GPs were implemented using the GPConstr Python package (Agrell, 2019), and to limit computational resource use, λ was constrained to be greater than or equal to 0.10.

2.5.3. Model validation and cover analyses

Model validation was performed by removing the smallest segmented colonies from each annotator's dataset in 5% increments in an iterative process until only 25% of each annotator's colony area data remained in the training subset. The removed size data was used to validate model predictions at each iteration. Both modelling techniques were evaluated based on the amount and proportion of data they required for training to consistently predict the size of each annotator's removed drawn coral colonies.

Analyses of annotator cover estimation variability were repeated as described in sections 2.3 and 2.4 including the imputed areas for each technique, and were compared to the manual segmentation results before the imputation of point labelled coral colonies.

3. Results

3.1. Darwin mounds coral cover

Total coral cover was estimated from the reference dataset as 0.13% (95% CI, 0.07 – 0.23%), made up by a total of 313 detected coral colonies distributed over the 3950.44 m^2 photographed seabed area. Living coral was detected in 22 of the 96 seabed images of the Western Darwin Mounds, with five of the images containing more than 1% coral and the maximum recorded image coral cover being 2.68%. Where live coral was detected in an image, the number of detected coral colonies ranged from 1–55, the largest colony recorded covering approximately 821 cm^2 seabed area.

3.2. Grid-based annotation and manual segmentation cover estimates

3.2.1. Annotator variability and bias within grid-based estimates

Annotator area-weighted mean cover estimates had a coefficient of variation of 33%, with the largest annotator estimate being three times the smallest (Fig. 2). Area-weighted mean coral cover across all annotators was 0.19% (95% CI, 0.09–0.33%). The maximum individual annotator bias for the grid method overestimated coral cover by 74% relative to the grid cover estimate across all annotators (Table 2). Image coral cover estimates were significantly different between annotators using the grid method, with annotators' maximum recorded image coral covers ranging from 2.14% to 5.85% (Friedman Q = 106.76, p <0.001; Table 3).

Where annotators detected live coral in image cells, 10 out of the 11 most commonly estimated $0 < x \leq 5\%$ coral cover, the last annotator estimating $5 < x \leq 15\%$ most frequently. Four annotators estimated 25% or more coral cover with a grid cell and no annotators estimated more than 45% coral cover being present within a grid cell.

3.2.2. Annotator variability and bias within manual segmentation estimates

Annotator area-weighted mean cover estimates using manual segmentation ranged between 0.04% and 0.12%, and the coefficient of variation between annotator cover estimates was 28% (Fig. 2). Area-weighted mean coral cover across all annotators was 0.08% (95% CI, 0.04–0.14%). The annotator with the greatest bias in the segmentation method underestimated coral cover by 52% relative to the segmentation estimate across all annotators (Table 2). Image coral cover estimates were significantly different between annotators using manual segmentation and maximum recorded image coral cover ranged from 1.38% to 2.61% across annotators (Friedman Q = 79.21, p <0.001; Table 3).

Compared to the reference dataset, the number of colonies detected by a single annotator ranged from 172 to 273, and 113 colonies were detected by all annotators. The mean apparent detection success across all annotators was 72%. The proportion of point labelled coral colonies ranged from 8%–82%, with seven annotators not providing an associated area for more than 50% of the colonies they detected in images.

3.2.3. Annotator variability and bias between annotation methods

For image examples of annotator variability between annotation methods, refer to the supplementary material (Supplement 5). There was significant disagreement between annotators and annotation method in coral cover estimates per image (Friedman Q = 328.85, p <0.001; Table 3). Ten of the 11 annotators estimated significantly higher image cover estimates with the grid-based method compared to segmentation, with seven annotators producing grid image cover estimates at least two times greater than their estimates made with segmentation (Supplement 6). Image cover estimates by the grid-based method showed more within method annotator disagreement than segmentation (Table 3; Supplement 6).

Greater inter-annotator variability in cover estimates occurred with the grid method compared to the segmentation of the same colony, and grids tended to overestimate colony size. Overestimation was more significant for small colonies (Fig. 3; Supplement 7). This systematic bias resulted in significant differences in whole dataset cover estimates and annotator variability between annotation methods (Fig. 2).

Table 2
Bias in annotator coral cover estimates, as the difference between individual annotators' coral estimates and the average across all annotators for each method. Presented as a percentage of the average cover estimate. Negative values indicate the annotator underestimated compared to the annotator average. IRC = Inverse Richard's selection curves, GP = Gaussian processes).

Annotator	Grids (%)	Segments without point labels (%)	Segments IRC modelled (%)	Segments GP modelled (%)
1	−30	5.0	19	22
2	39	34	14	7.2
3	−17	−22	−2.6	2.7
4	−4.2	−30	−23	−17
5	−0.2	9.8	−13	−19
6	−30	6.3	−2.7	−2.5
7	5.1	−52	−27	−29
8	−47	−7.3	4.6	12
9	74	15	18	18
10	−7.9	−5.7	−7.3	−6.9
11	19	46	21	11
Range	−30	−52	−27	−29
	74	46	21	22

Table 3
Variability between annotators and methods in average coral cover estimates per image. IRC = Inverse Richard's curves, GP = Gaussian processes.

Annotation method	Coral cover per image (%)	Friedman Q
Grids	0.14–0.41	106.76***
Segments:		
Without point labels	0.06–0.15	79.21***
IRC modelled	0.10–0.16	58.72***
GP modelled	0.10–0.18	71.76***
Grids and segments without point labels	0.04–0.33	328.85***

*** p-values <0.001; see also Supplement 6.

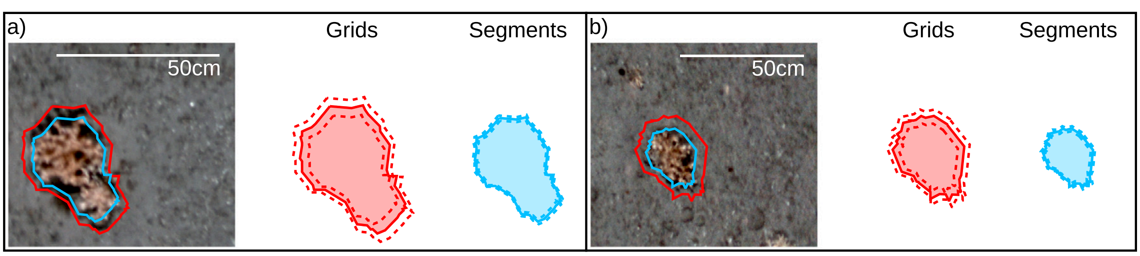

Fig. 3. Examples of method and annotator variability in cold-water coral image annotation on individual colony (a; b) scales. Images cropped to represent one grid cell. Coral colony sizes estimated by 11 annotators using both grid (red) and segment (blue) annotation methods. For the individual colonies — the union of the drawn colonies during manual segmentation was scaled to the mean size estimated by each method and overlaid to visualise trends. Adjacent are the mean sizes (solid line with shading) plus or minus the standard deviation (dotted lines) for each annotation method and the estimated areas.

For the study area, the standard deviation in grid-based coral cover estimates across annotators was three times the equivalent standard deviation using segmentation (0.06% compared to 0.02%, respectively). The area-weighted mean coral percentage cover across all annotators using the grid-based method was more than double that obtained using the segmentation method (Wilcoxon $T = 0.0$, p <0.001).

3.2.4. Annotation method bias and effort

For both annotation methods, annotators' area weighted mean coral cover estimates were significantly different to the coral cover estimated by the reference dataset (Wilcoxon, grids $T = 4.0$, p <0.01; segments $T = 0.0$, p <0.001). The grid-based annotation method generated the greatest method bias, overestimating coral cover by 45% (95% CI, -34%–144%; Fig. 4). Manual segmentation under-estimated coral cover by 38% (95% CI, −72 – 4.03%) compared to the reference dataset.

Annotators varied in mean time spent annotating per image both within and between annotation methods, with some annotators being twice as fast as other annotators using the same method (Supplement 8). Total annotation time was not significantly different between annotation methods (mean ± SE, grids = 179.39 ± 15.13 min; segments = 217.64 ± 21.97 min, Wilcoxon $T = 13.0$, $p = 0.08$).

For manual segmentation, there were no significant correlations between the coral detection success of annotators and the total time spent annotating (Kendall $\tau = 0.07$, $p = 0.75$), the proportion of point labels annotators had and the total time spent annotating ($\tau = -0.13$, $p = 0.65$), nor the number of colonies annotators detected and the number of colonies annotators segmented ($\tau = 0.18$, $p = 0.43$; Supplement 9).

3.3. Imputing missing cover data in manual segmentation

3.3.1. Model validation and performance

Model validation showed that, for both techniques, prediction performance of coral sizes improved as larger proportions of colony size observations were included in the training data. Most model iterations needed at least 20% of the annotator's detected coral colonies (approximately 50 colonies) to have a drawn segment to train with to predict realistic trends in colony size. The IRCs under-predicted known coral sizes whereas the GPs over-predicted them (Fig. 5). In model validation, the extrapolation performance of both models varied between annotators, achieving more accurate predictions of known

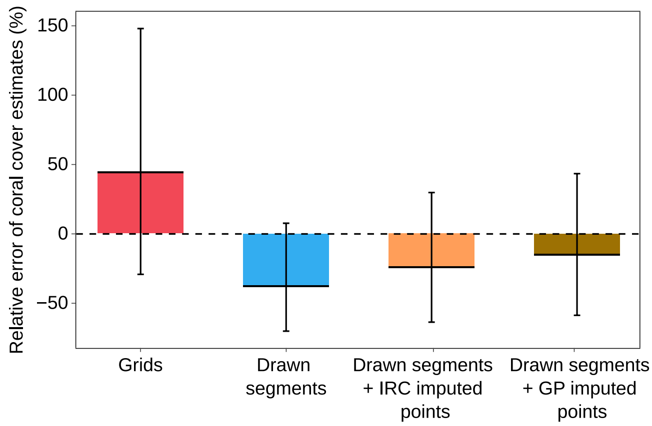

Fig. 4. Relative error of annotation methods and imputation techniques for annotating cold-water coral cover. Errors (coloured bars and solid black lines) are presented as the percentage differences between each annotation approach's respective area-weighted mean cover estimate across all annotators and the area-weighted mean cover estimated from a reference dataset generated from all 11 annotators' manual segmentation results. Error bars are annotation approaches' 95% confidence intervals. IRC = Inverse Richard's curves, GP = Gaussian processes.

colony size with annotators that had greater numbers of segmented colonies (Fig. 6).

Imputing the test coral colony sizes with either modelling technique significantly improved the accuracy of coral cover estimates compared to ignoring the removed colony sizes (Fig. 6; Wilcoxon signed rank test for significant coral cover difference when predicting the known sizes of at least 25 coral colonies compared to ignoring them $T = 0.0$, $p < 0.001$).

Segmented colony sizes were directly used for cover estimation and so only model extrapolation of the small point labelled coral colony sizes are of interest in this study (Fig. 7). The IRC models estimated smaller colony sizes than the GPs, with colony sizes gradually declining with decreasing cumulative frequency (Fig. 7). In the GP models, the decline in colony size with decreasing cumulative frequency was initially less than in the IRC models, increasing just before reaching zero cumulative frequency (Fig. 7). The GP modelled colony sizes had larger associated uncertainties, which increased when fewer segmented coral colony sizes were provided. The uncertainty in the coral sizes predicted with IRCs depended on both the proportion of segment annotations used in model training and the distribution of colony sizes provided by the annotator (Fig. 7a; 7b). The uncertainty in IRC modelled colony sizes first increased and then decreased again as imputed colony size reduced, tapering near zero. Both models agreed best with observed coral colony sizes for annotators with higher proportions of drawn segments (Fig. 7).

3.3.2. *Annotator variability and bias after cover imputation*

Coral cover estimates per image were increased and showed less inter-annotator variability after point labelled colony sizes were imputed and contributing to cover estimates (Table 3).

Imputing colony sizes increased segmentation coral cover estimates (Fig. 8). Area weighted mean cover estimates were 0.10% (95% CI, 0.05 – 0.17%) and 0.11% (95% CI, 0.05 – 0.19%), with colony sizes imputed by IRC models and GPs respectively . The coefficient of variation between annotators was reduced from 28% before imputation to 17% and 16% after imputing with IRC models and GPs respectively. The maximum annotator bias was reduced from an underestimate of 52% compared to the average cover estimate across all annotators, to underestimates of 27% and 29% (with IRC and GP imputed colony sizes, respectively; Table 2).

3.3.3. *Annotation method bias after cover imputation*

Annotator's area weighted mean coral cover estimates were still significantly different to the coral cover estimated by the reference dataset (Wilcoxon, IRC imputed $T = 0.0$, $p < 0.001$; GP imputed $T = 2.0$, $p < 0.01$). However, imputing point-labelled colony sizes with IRCs and GPs reduced segmentation underestimation from 38% compared to the reference dataset to 23% (95% CI, −65%–29%) and 15% (95% CI, −59%–40%) respectively (Fig. 4).

4. Discussion

We found that coral cover estimates made using both manual segmentation and grid-based annotation methods varied between annotators, with some annotator cover estimates being three times that of others using the same annotation method. Manual segmentation produced lower coral cover estimates with lower annotator variability and bias compared to the grid-based estimation method in our sparse coral cover dataset. The results also show that the use of modelling techniques to impute unmeasured coral sizes in the manual segmentation method can reduce size selectivity bias, annotator variability and coral cover underestimation. Based on these findings, we provide generalised recommendations for planning image annotation and for accounting for annotator variability in datasets when evaluating percentage cover from seabed images. These recommendations can be applied both to multi-annotator and single annotator studies.

4.1. *The impact of variability and bias in cover estimates for monitoring programmes*

The observed variability in cover estimates within and between annotation methods may raise concerns for the effectiveness of monitoring. For example, where low seabed cover values are employed to both define and monitor the status of VMEs, even modest variations could substantially influence interpretations and outcomes.

The almost unique case of scleractinian coral growing on sandy substrata at the Darwin Mounds makes the on-mounds cold-water coral habitat an important deep-sea VME for conservation (De Santo and Jones, 2007). Considering on-mound images only (38 m^2 average seabed area), live coral cover was 0.57% (95% CI, 0.00–2.40%). Rowden et al. (2020) provide data on live coral colony numerical density and associated species richness that they use to define objective cold-water coral VME thresholds in studies based on 25 and 50 m^2 video transects. We have used those data to establish coral colony density thresholds at the points where 90% of the asymptotic associated species richness were achieved for the two sample unit sizes, both approximately 0.1 live colonies m^{-2}. In the Darwin Mounds case, that density threshold corresponds to 0.22% live coral cover. Our on-mound live coral cover estimate made from the reference dataset meets that VME threshold, as do the estimates made with each annotation and imputation technique when estimated across all annotators. However, when considering individual annotator estimates, one segmentation without imputation value fails to reach the threshold (Supplement 10).

Semi- and fully automated image analysis techniques are increasingly being used in seabed monitoring studies to support the evaluation of marine imaging data at effective timescales. Some examples include automated point count annotation methods for estimating coral reef substrate cover (Beijbom et al., 2015); automated semantic segmentation of coral, sponges (Purser et al., 2009) and poly-metallic nodules (Schoening et al., 2016) from seafloor images; automated cold-water coral reef classification from photogrammetric reconstructions (de Oliveira et al., 2021); real-time detection and segmentation of litter from seafloor images (Corrigan et al., 2023); artificial intelligence assisted semantic segmentation of corals from orthophotos (Pavoni et al., 2021) and machine learning assisted detection and segmentation of seafloor organisms within online annotation platforms (Zurowietz

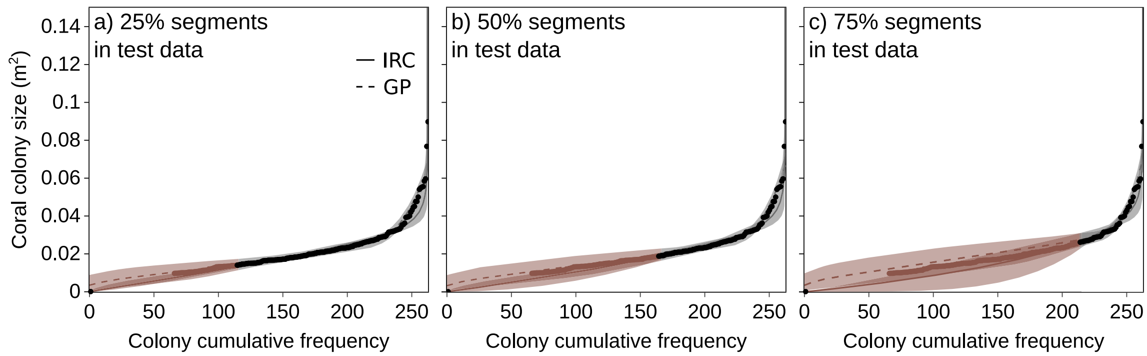

Fig. 5. Model validation for imputing coral colony sizes with inverse Richard's selection curves (IRC; solid lines) and constrained Gaussian processes (GP; dashed lines), using data from the annotator with the greatest number of drawn segments as an illustrative example. Models extrapolated mean and 95% confidence intervals (coloured lines and shaded bands) of small coral colony sizes, including colonies moved from training (black markers) to test datasets (coloured markers).

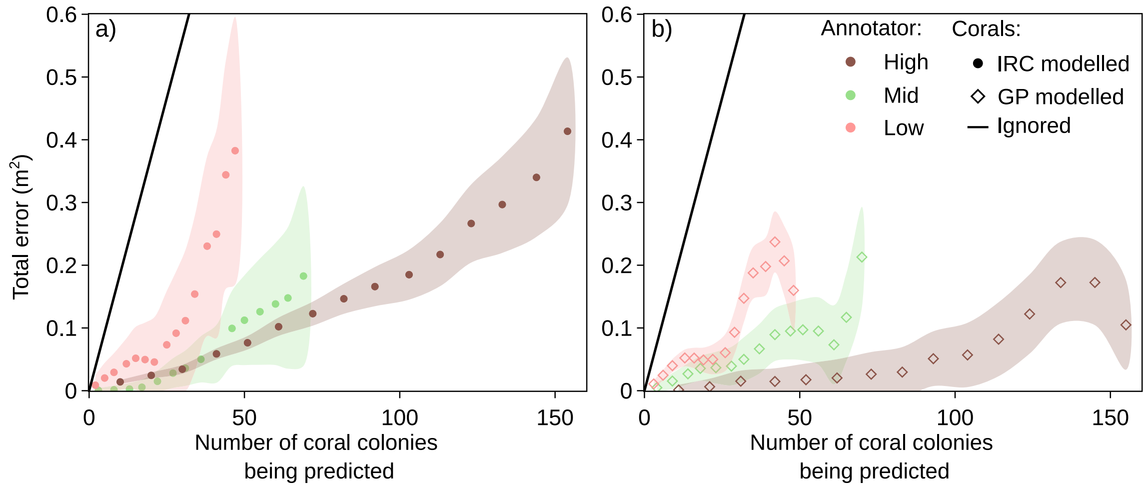

Fig. 6. Total error for imputation model validations, measured as absolute differences between modelled small coral colony sizes and drawn colony sizes, shown for three annotators. The three annotators selected had either a low ('Low', pink), medium ('Mid', green), or high ('High', brown) proportion of drawn segments compared to the total number of coral colonies each detected. (a) Inverse Richard's selection curve modelled (IRC); (b) Gaussian process modelled (GP).

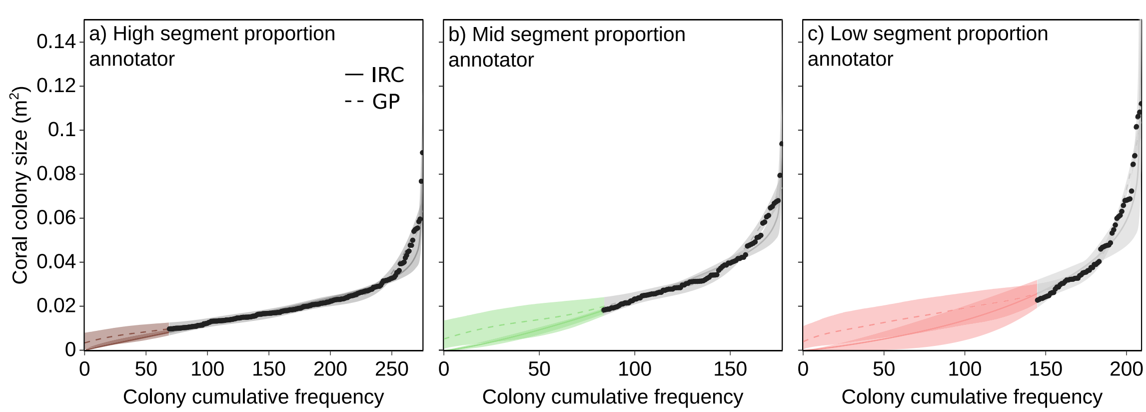

Fig. 7. Coral colony sizes modelled with inverse Richard's selection curves (IRC; solid lines) and constrained Gaussian processes (GP; dashed lines) to extrapolate small coral colony sizes for three example annotators.

et al., 2018) . These non-manual image annotation methods can compensate for some of the limitations in manual annotations highlighted in this study, such as the minimum mask size human annotators are reliably able to draw, and error attributable with fatigue or other effects not experienced by machines (Culverhouse, 2007). Manual annotation techniques are, however, still being used to assess the state of ecosystems and automated techniques either rely on manually generated training data or use manually annotated data as ground truthing to evaluate model performance against. Therefore, accurate manual annotations are still required as we move towards more automated image analysis techniques.

The variability in cover estimates between annotators, both on-mound and across the whole study, suggests that studies monitoring low seabed cover VMEs could potentially encounter times three variation in estimated cover between monitoring timepoints attributable to annotation method or annotator variability alone. The impact of variability on cover estimates will not be of the same magnitude in high

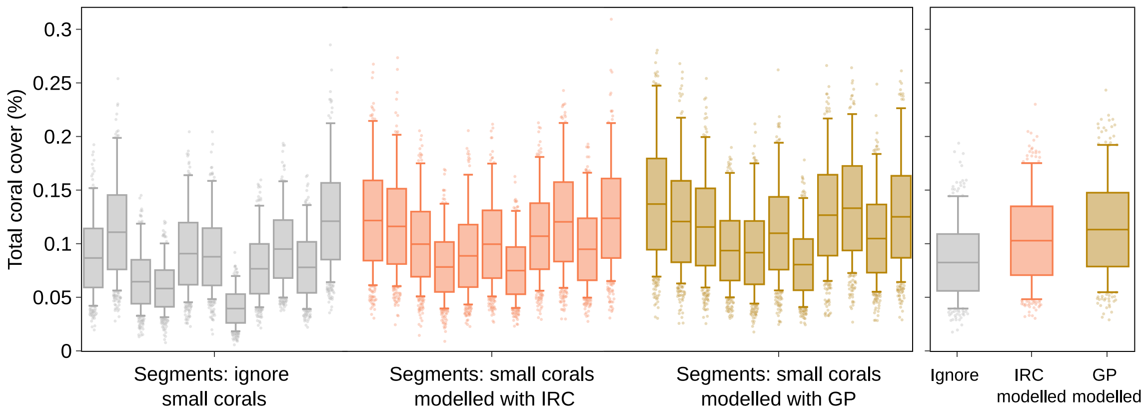

Fig. 8. Area-weighted mean coral cover over total study area estimated by each annotator (left) and across all annotators (right), using the manual segmentation annotation method and considering identified corals with no associated area in three ways. Box plots show mean and standard deviations (solid line and box) with whiskers as 95% confidence intervals and outliers as points. IRC = Inverse Richard's curves, GP = Gaussian processes.

seabed cover habitats (e.g., tropical coral reefs) but likely still exists and should also be considered in corresponding monitoring studies, particularly in cases monitoring gradual decline. With increasing reuse of image annotation data, reporting potential annotator bias and variability is critical to enabling meaningful comparisons between studies and timepoints. Therefore, whether in a one-off survey or in repeat surveys, formal assessments of potential bias, inter-annotator and inter-survey sources of variability should be carried out as a matter of best practise.

4.2. Annotation method choice influence on cover estimates

Estimates of percentage coral cover from grid-based seabed image annotations had higher annotator variability and bias than manual segmentation. Annotator variability and bias in the grid-based method was largely a result of annotators differing in their evaluations of colony sizes rather than in their detection of the colonies.

Grid estimates are more subjective than segmentation, as it requires annotators to judge the percentage of cover present in each grid cell. Human perception of percentage cover can be prone to overestimation and has a higher likelihood of varying between annotators (Olmstead et al., 2004; Finn et al., 2010). This bias and variability is especially evident when the subjects of interest are much smaller than the cell size (Benedetti-Cecchi et al., 1996; Trygonis and Sini, 2012); the coral colonies in this study were all smaller than 0.1 m^2 in size, in comparison to a mean grid cell size of 0.73 m^2. To aid in reducing bias associated with the chosen cell size, Dethier et al. (1993) recommended using many small subdivisions in quadrats to aid censusing scattered and low-coverage biota. However, while a reduction in cell size may reduce bias, it has previously been found to reduce the time efficiency of generating cover estimates (Benedetti-Cecchi et al., 1996; Deter et al., 2012; Trygonis and Sini, 2012). Annotators were instructed to estimate cell cover in ten percent increments. As most annotators modal cell cover category was $0 < x \leq 5\%$, cover estimate precision may have been improved by annotators using smaller cover increments or providing a continuous estimate instead, so as to remove decision rules as a source of annotator variability (Dethier et al., 1993).

The sparse distribution and small size of many of the coral colonies recorded in this study meant the grid-based estimations overestimated the size of coral colonies and the annotation method was not significantly faster than manual segmentation, which produced estimates with lower annotator bias and uncertainty. Unlike grid-based estimates, manual segmentation can also provide additional information on colony shape, orientation and size distributions, giving insight into coral condition and recovery (De Clippele et al., 2018). Furthermore, drawn segments can be used to train machine learning algorithms for automated image analyses (Buškus et al., 2021; Piechaud and Howell, 2022). However, manual segmentation showed a similar magnitude of method bias to grid-based estimates (45% with grids, 38% with segmentation) but underestimated coral cover compared to the reference annotation set. This underestimation was in part caused by size selectivity bias.

4.3. Size selectivity bias in manual segmentation

High inter-annotator variability in cover by the segmentation method occurred as a result of differences in the smallest segmented colony and the proportion of identified coral assigned only point labels. Annotator coral colony detection success was not related to annotator segmented colony proportions, so differences in cover estimates were related to the lack of area assignment rather than coral detection. The proportion of point labels did not correlate with total time spent annotating, suggesting that differences were not a function of annotator effort. Personal differences may have contributed to annotator variability in coral colony segmentation, including personal workspace setups, screen resolutions and equipment, dexterity and visual perception (Van Coillie et al., 2014). Trends towards big-data studies and the use of shared image repositories and multi-user annotation platforms (Bewley et al., 2015; Langenkämper et al., 2017), indicate that these differences should be expected, emphasising the need for studies to set realistic bounds on their expected impact.

For the purposes of coral monitoring, the detection of coral recruitment and new colony formations can provide evidence for potential recovery of a site after disturbance events such as trawling or bleaching (Connell et al., 1997; Beazley et al., 2021). These new colonies are likely to be small in size, depending on coral growth rates and time after disturbance. Failure to detect or provide a size estimate for small colonies due to size selection bias can restrict our monitoring outputs and insight into habitat recovery (Perkins et al., 2022), and so annotators should strive to provide size estimates for as many detected colonies as possible.

Size selectivity bias can occur both in visual surveys and in physical sampling techniques such as trawling and is likely a source of uncertainty for studies of small target species, or those examining a broad spectrum of organism sizes. This has been reported in both marine and terrestrial ecological studies, where imperfect detection leads to small, difficult to detect species being falsely absent or under-detected (Issaris et al., 2012; Kellner and Swihart, 2014; Breton et al., 2013; Warwick and Clarke, 1996). Size selectivity bias has been reported in a study of individual megafauna in seabed images (Schoening et al., 2020), where apparent faunal density increased and the median dimensions of annotated fauna decreased with increasing image resolution. This effect needs to be considered for long-term monitoring where image acquisition equipment and deployment strategies can be expected to vary over time (Clark et al., 2019).

4.4. Modelling colony size to impute missing data

The models used to impute missing coral colony sizes successfully reduced inter-annotator variability by 10% and increased coral cover estimates (Fig. 8), but differed in some practical aspects and some details of results that may be important to consider. The IRC models were entirely data-driven, requiring only a single end-point for the minimum colony size to be defined for all annotators. By contrast, the constrained GPs required a larger number of parameters to define the model, related to coral morphology, image acquisition properties and attributes of each annotator's data. Care should be taken to choose appropriate initial parameters for GP's that are representative of the modelled data's underlying characteristics or there is a risk of the model over- or under-fitting (Rasmussen and Williams, 2006). Both models performed well in validation tests when extrapolated sizes were compared to drawn segment area estimates (Fig. 5).

Both models followed a smooth size distribution of coral colonies and would not account for large changes within a single annotation set, created from a change in annotation setup for example. Since model performance related to both the proportion of observed colony sizes and the total number of area assigned annotations, annotators should aim to annotate small specimens or colonies in images as much as possible to improve model performance, where noise related to error in size assignment of small targets can be identified and rejected prior to training. Most notably, model validation showed that both models represented the test coral colony sizes more accurately than if these colonies were ignored. Therefore, use of either technique is recommended when segmented annotations have known size selectivity bias.

Reducing size selection bias and variability (through imputation or annotating smaller specimens) improves the ability (statistical power) to detect change in coral cover. Additionally, the modelling techniques addressed some of the method bias present in the manual segmentation method, reducing the underestimation of coral cover caused by point labelled colonies not being accounted for in cover estimates. The reference annotation set created by combining all annotators manual segmentation outputs represents the upper-bound number of coral colonies present in the annotation set. Imperfect detection was still present after imputing small coral colony sizes, and further error likely still existed, in the form of misclassified colonies. Despite these remaining issues, imputing missing coral colony sizes increased annotator averaged cover estimates by a relative 25%–37%, suggesting that it is a worthwhile endeavour.

5. Conclusions and recommendations

This study highlights the potential bias and variability in annotator cover estimates from seabed images. These have a significant influence when deriving cover estimates of organisms with sparse, patchy distributions, such as the Darwin Mounds cold-water corals considered in this study. To reduce inter-annotator variability and bias when estimating biota cover from seafloor images, we provide the following recommendations:

1. Before starting image annotation, the method should be chosen based on the distribution of biota: for small/sparsely distributed biota, manual segmentation may reduce inter-annotator variability and bias.
2. Impute missing data: where annotators can detect but not measure any fraction of the target population, post-process these data via imputation for improved cover estimates.
3. Annotators should provide measurement estimates for as much of the target population as feasibly possible, not only to improve imputation model robustness, but also to provide population size structure data which can be used to evaluate habitat recovery.
4. Studies should report potential annotator bias and variability for quality control purposes and to allow for meaningful comparisons between annotators, annotations methods, and time points. A common subset of images should randomly be repeatedly annotated and from it, annotator variability, bias, and detection success can be estimated and if possible, accounted for.

These recommendations should reduce both annotator bias and variability and consequently improve the monitoring of seabed cover by VMEs and other targets of interest.

CRediT authorship contribution statement

Emma J. Curtis: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Jennifer M. Durden:** Conceptualization, Investigation, Resources, Writing – review & editing, Supervision. **Brian J. Bett:** Conceptualization, Formal analysis, Resources, Writing – review & editing, Supervision. **Veerle A.I. Huvenne:** Investigation, Resources, Writing – review & editing. **Nils Piechaud:** Methodology, Investigation, Resources. **Jenny Walker:** Investigation, Data curation. **James Albrecht:** Resources, Writing – review & editing. **Miquel Massot-Campos:** Investigation. **Takaki Yamada:** Investigation. **Adrian Bodenmann:** Investigation. **Jose Cappelletto:** Investigation, Writing – review & editing. **James A. Strong:** Investigation. **Blair Thornton:** Conceptualization, Methodology, Investigation, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Readers can access the annotation data used in this study via the Squidle+ link provided after submitting a request to the authors.

Acknowledgements

Many thanks to the captain and crew of the RRS *Discovery* DY108/109 and the *Autosub6000* team. We also thank the members of the University of Southampton and the National Oceanography Centre who provided the image annotations for this study. This research was funded by the UK Natural Environment Research Council's Oceanids programme (NE/P020887/1 and NE/P020739/1). Curtis was funded by the UK Natural Environment Research Council's INSPIRE programme (NE/S007210/1). Durden, Bett and Huvenne were also funded by the Climate Linked Atlantic Sector Science (CLASS) project supported by the UK Natural Environment Research Council's National Capability funding (NE/R015953/1). Huvenne enjoyed a Fellowship from the Hanse-Wissenschaftskolleg Institute for Advanced Study during the final preparation stages of this manuscript. Data used in this article are available on request on the online benthic imaging repository Squidle+ (www.soi.squidle.org).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.pocean.2024.103214.

References

Agrell, C., 2019. Gaussian processes with linear operator inequality constraints. 20, pp. 1–36. http://dx.doi.org/10.48550/arXiv.1901.03134, arXiv.

Almany, G.R., Connolly, S.R., Heath, D.D., Hogan, J.D., Jones, G.P., McCook, L.J., Mills, M., Pressey, R.L., Williamson, D.H., 2009. Connectivity, biodiversity conservation and the design of marine reserve networks for coral reefs. Coral Reefs 28 (2), 339–351. http://dx.doi.org/10.1007/s00338-009-0484-x.

Baco, A.R., Ross, R., Althaus, F., Amon, D., Bridges, A.E.H., Brix, S., Buhl-Mortensen, P.a., Colaco, A., Carreiro-Silva, M., Clark, M.R., Du Preez, C., Franken, M.-l., Gianni, M., Gonzalez-Mirelis, G., Hourigan, T., Howell, K., Levin, L.A., Lindsay, D.J., Molodtsova, T.N., Morgan, N., Morato, T., Mejia-Mercado, B.E., O'Sullivan, D., Pearman, T., Price, D., Robert, K., Robson, L., Rowden, A.A., Taylor, J., Taylor, M., Victorero, L., Watling, L., Williams, A., Xavier, J.R., Yesson, C., 2023. Towards a scientific community consensus on designating Vulnerable Marine Ecosystems from imagery. PeerJ 11, e16024. http://dx.doi.org/10.7717/peerj.16024, URL https://peerj.com/articles/16024.

Beazley, L., Kenchington, E., Korabik, M., Fenton, D., King, M., 2021. Other Effective Area-Based conservation measure promotes recovery in a cold-water coral reef. Global Ecol. Conservat. 26, e01485. http://dx.doi.org/10.1016/j.gecco.2021.e01485.

Beijbom, O., Edmunds, P.J., Roelfsema, C., Smith, J., Kline, D.I., Neal, B.P., Dunlap, M.J., Moriarty, V., Fan, T.-Y., Tan, C.-J., Chan, S., Treibitz, T., Gamst, A., Mitchell, B.G., Kriegman, D., 2015. Towards automated annotation of benthic survey images: Variability of human experts and operational modes of automation. In: Chen, C.A. (Ed.), PLoS One 10 (7), e0130312. http://dx.doi.org/10.1371/journal.pone.0130312, URL https://dx.plos.org/10.1371/journal.pone.0130312.

Benedetti-Cecchi, L., Airoldi, L., Abbiati, M., Cinelli, F., 1996. Estimating the abundance of benthic invertebrates: A comparison of procedures and variability between observers. Mar. Ecol. Prog. Ser. 138, 93–101. http://dx.doi.org/10.3354/meps138093.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Stat. Methodol. 57 (1), 289–300, URL http://www.jstor.org/stable/2346101.

Bett, B.J., 2001. UK Atlantic Margin Environmental Survey: Introduction and overview of bathyal benthic ecology. Cont. Shelf Res. 21, 917–956. http://dx.doi.org/10.1016/S0278-4343(00)00119-9, URL https://linkinghub.elsevier.com/retrieve/pii/S0278434300001199.

Bewley, M., Friedman, A., Ferrari, R., Hill, N., Hovey, R., Barrett, N., Marzinelli, E.M., Pizarro, O., Figueira, W., Meyer, L., Babcock, R., Bellchambers, L., Byrne, M., Williams, S.B., 2015. Australian sea-floor survey data, with images and expert annotations. Sci. Data 2 (150057), http://dx.doi.org/10.1038/sdata.2015.57.

Bodenmann, A., Thornton, B., Ura, T., 2017. Generation of high-resolution three-dimensional reconstructions of the seafloor in color using a single camera and structured light. J. Field Robotics 34 (5), 833–851. http://dx.doi.org/10.1002/rob.21682.

Breton, A.R., Hawkins, J.A., Winkelman, D.L., 2013. Correcting length-frequency distributions for imperfect detection. North Am. J. Fish. Manag. 33 (6), 1156–1165. http://dx.doi.org/10.1080/02755947.2013.829141.

Brown, E.K., Cox, E., Jokiel, P.L.P.L., Rodgers, K., Smith, W.R., Tissot, B.N., Coles, S.L.S.L., Hultquist, J., Rodgers, S.K., Smith, W.R., Tissot, B.N., Coles, S.L.S.L., Hultquist, J., 2004. Development of benthic sampling methods for the Coral Reef Assessment and Monitoring Program (CRAMP) in Hawai'i. Pacific Sci. 58 (2), 145–158. http://dx.doi.org/10.1353/psc.2004.0013.

Bullimore, R.D., Foster, N.L., Howell, K.L., 2013. Coral-characterized benthic assemblages of the deep Northeast Atlantic: Defining "Coral Gardens" to support future habitat mapping efforts. ICES J. Mar. Sci. 70 (3), 511–522. http://dx.doi.org/10.1093/icesjms/fss195, URL https://academic.oup.com/icesjms/article/70/3/511/914657.

Buškus, K., Vaičiukynas, E., Verikas, A., Medelytė, S., Šiaulys, A., Šaškov, A., 2021. Automated quantification of brittle stars in seabed imagery using computer vision techniques. Sensors 21 (22), 1–15. http://dx.doi.org/10.3390/s21227598.

Chalana, V., Kim, Y., 1997. A methodology for evaluation of boundary detection algorithms on medical images. IEEE Trans. Med. Imaging 16 (5), 642–652. http://dx.doi.org/10.1109/42.640755, URL http://ieeexplore.ieee.org/document/640755/.

Chaniotis, P.D., Robson, L.M., Lemasson, A.J., Cornthwaite, A.L., Howell, K.L., 2020. UK deep-sea conservation: Progress, lessons learned, and actions for the future. Aquat. Conserv. Mar. Freshw. Ecosyst. 30 (2), 375–393. http://dx.doi.org/10.1002/aqc.3243.

Clark, M.R., Bowden, D.A., Rowden, A.A., Stewart, R., 2019. Little evidence of benthic community resilience to bottom trawling on seamounts after 15 years. Front. Mar. Sci. 6 (63), 1–16. http://dx.doi.org/10.3389/fmars.2019.00063.

Connell, J.H., Hughes, T.P., Wallace, C.C., 1997. A 30-year study of coral abundance, recruitment, and disturbance at several scales in space and time. Ecol. Monograph 67 (4), 461–488. http://dx.doi.org/10.1890/0012-9615(1997)067[0461:AYSOCA]2.0.CO;2.

Conover, W.J., 1999. Practical Nonparametric Statistics, third ed. vol. 350, John Wiley & Sons, p. 608. http://dx.doi.org/10.2307/2284744, URL https://www.jstor.org/stable/2284744?origin=crossref.

Constable, A.J., Costa, D.P., Schofield, O., Newman, L., Urban, E.R., Fulton, E.A., Melbourne-Thomas, J., Ballerini, T., Boyd, P.W., Brandt, A., de la Mare, W.K., Edwards, M., Eléaume, M., Emmerson, L., Fennel, K., Fielding, S., Griffiths, H., Gutt, J., Hindell, M.A., Hofmann, E.E., Jennings, S., La, H.S., McCurdy, A., Mitchell, B.G., Moltmann, T., Muelbert, M., Murphy, E., Press, A.J., Raymond, B., Reid, K., Reiss, C., Rice, J., Salter, I., Smith, D.C., Song, S., Southwell, C., Swadling, K.M., Van de Putte, A., Willis, Z., 2016. Developing priority variables ("ecosystem Essential Ocean Variables" - eEOVs) for observing dynamics and change in Southern Ocean ecosystems. J. Mar. Syst. 161, 26–41. http://dx.doi.org/10.1016/j.jmarsys.2016.05.003.

Corrigan, B.C., Tay, Z.Y., Konovessis, D., 2023. Real-time instance segmentation for detection of underwater litter as a plastic source. J. Mar. Sci. Eng. 11 (8), http://dx.doi.org/10.3390/jmse11081532.

Costello, M.J., McCrea, M., Freiwald, A., Lundälv, T., Jonsson, L., Bett, B.J., van Weering, T.C.E., de Haas, H., Roberts, J.M., Allen, D., 2005. Role of cold-water Lophelia pertusa coral reefs as fish habitat in the NE Atlantic. Cold-Water Corals Ecosyst. (1), 771–805. http://dx.doi.org/10.1007/3-540-27673-4_41.

Council of the European Union, 2004. Council REgulation (EC) No 602/2004 of 22 2004 amending Regulation (EC) No 850/98 as regards the protection of deepwater coral reefs from the effects of trawling in an area North West of Scotland. Off. J. Eur. Union L 97 (1.4.2004), 30–31.

Culverhouse, P.F., 2007. Human and machine factors in algae monitoring performance. Ecol. Inform. 2 (4), 361–366. http://dx.doi.org/10.1016/j.ecoinf.2007.07.001.

Culverhouse, P.F., Williams, R., Reguera, B., Herry, V., González-Gil, S., 2003. Do experts make mistakes? A comparison of human and machine identification of dinoflagellates. Mar. Ecol. Prog. Ser. 247, 17–25. http://dx.doi.org/10.3354/meps247017.

Danovaro, R., Fanelli, E., Aguzzi, J., Billett, D., Carugati, L., Corinaldesi, C., Dell'Anno, A., Gjerde, K., Jamieson, A.J., Kark, S., McClain, C., Levin, L., Levin, N., Ramirez-Llodra, E., Ruhl, H., Smith, C.R., Snelgrove, P.V., Thomsen, L., Van Dover, C.L., Yasuhara, M., 2020. Ecological variables for developing a global deep-ocean monitoring and conservation strategy. Nat. Ecol. Evol. 4 (2), 181–192. http://dx.doi.org/10.1038/s41559-019-1091-z.

De Clippele, L.H., Huvenne, V.A., Orejas, C., Lundälv, T., Fox, A., Hennige, S.J., Roberts, J.M., 2018. The effect of local hydrodynamics on the spatial extent and morphology of cold-water coral habitats at Tisler Reef, Norway. Coral Reefs 37 (1), 253–266. http://dx.doi.org/10.1007/s00338-017-1653-y.

de Oliveira, L.M.C., Lim, A., Conti, L.A., Wheeler, A.J., 2021. 3D classification of cold-water coral reefs: A comparison of classification techniques for 3D reconstructions of cold-water coral reefs and seabed. Front. Mar. Sci. 8 (March), 1–19. http://dx.doi.org/10.3389/fmars.2021.640713.

De Santo, E., Jones, P., 2007. The Darwin Mounds: From undiscovered coral to the development of an offshore marine protected area regime. Bull. Mar. Sci. 81 (Supplement 1), 147–156, URL http://discovery.ucl.ac.uk/149542/.

De'ath, G., Fabricius, K.E., Sweatman, H., Puotinen, M., 2012. The 27-year decline of coral cover on the Great Barrier Reef and its causes. Proc. Natl. Acad. Sci. 109 (44), 17995–17999. http://dx.doi.org/10.1073/pnas.1208909109, URL http://www.pnas.org/cgi/doi/10.1073/pnas.1208909109.

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. 418–419, 75–82. http://dx.doi.org/10.1016/j.jembe.2012.03.006.

Dethier, M.N., Graham, E.S., Cohen, S., Tear, L.M., 1993. Visual versus random-point percent cover estimations: "objective' is not always better. Mar. Ecol. Prog. Ser. 96 (1), 93–100. http://dx.doi.org/10.3354/meps096093.

Durden, J.M., Bett, B.J., Jones, D.O.B., Huvenne, V.A.I., Ruhl, H.A., 2015. Abyssal hills – hidden source of increased habitat heterogeneity, benthic megafaunal biomass and diversity in the deep sea. Prog. Oceanogr. 137, 209–218. http://dx.doi.org/10.1016/j.pocean.2015.06.006.

Durden, J.M., Bett, B., Schoening, T., Morris, K., Nattkemper, T., Ruhl, H., 2016. Comparison of image annotation data generated by multiple investigators for benthic ecology. Mar. Ecol. Prog. Ser. 552, 61–70. http://dx.doi.org/10.3354/meps11775, URL http://www.int-res.com/abstracts/meps/v552/p61-70/.

Finn, P.G., Udy, N.S., Baltais, S.J., Price, K., Coles, L., 2010. Assessing the quality of seagrass data collected by community volunteers in Moreton Bay Marine Park, Australia. Environ. Conserv. 37 (1), 83–89. http://dx.doi.org/10.1017/S0376892910000251.

Gass, S.E., Roberts, J.M., 2011. Growth and branching patterns of Lophelia pertusa (Scleractinia) from the North Sea. J. Mar. Biol. Assoc. United Kingdom 91 (4), 831–835. http://dx.doi.org/10.1017/S002531541000055X.

Hennige, S.J., Wicks, L.C., Kamenos, N.A., Perna, G., Findlay, H.S., Roberts, J.M., 2015. Hidden impacts of ocean acidification to live and dead coral framework. Proc. R. Soc. B: Biol. Sci. 282 (1813), http://dx.doi.org/10.1098/rspb.2015.0990.

Hill, J., Wilkinson, C., 2004. Methods for Ecological Monitoring of coral reefs, 1 (9). Australian Institute of Marine Science, pp. 1–10.

Hollander, M., Wolfe, D.A., Chicken, E., 2013. Nonparametric Statistical Methods. John Wiley & Sons, Incorporated, Somerset, UNITED STATES.

Huvenne, V.A., Bett, B.J., Masson, D.G., Le Bas, T.P., Wheeler, A.J., 2016. Effectiveness of a deep-sea cold-water coral Marine Protected Area, following eight years of fisheries closure. Biol. Cons. 200, 60–69. http://dx.doi.org/10.1016/j.biocon.2016.05.030.

Huvenne, V.A., Thornton, B., 2020. Discovery Cruise DY108-109, 6 Sept - 2 Oct 2019. CLASS – Climate-linked Atlantic System Science Darwin Mounds Marine Protected Area habitat monitoring, BioCAM - first equipment trials. BLT- Recipes: Pilot study. In: National Oceanography Centre Cruise Report, (no. 66), National Oceanography Centre, Southampton, p. 224, URL http://nora.nerc.ac.uk/id/eprint/526682/.

IOC-UNESCO, 2020. Global Ocean Science Report 2020 - Charting Capacity for Ocean Sustainability. UNESCO Publishing, Paris, p. 250, URL https://unesdoc.unesco.org/ark:/48223/pf0000375147,

Issaris, Y., Katsanevakis, S., Salomidi, M., Tsiamis, K., Katsiaras, N., Verriopoulos, G., 2012. Occupancy estimation of marine species: Dealing with imperfect detectability. Mar. Ecol. Prog. Ser. 453, 95–106. http://dx.doi.org/10.3354/meps09668.

Jamieson, A.J., Boorman, B., Jones, D.O., 2013. Deep-sea benthic sampling. In: Eleftheriou, A. (Ed.), In: Methods for the Study of Marine Benthos, (no. i), John Wiley & Sons, Ltd, Oxford, UK, pp. 285–347. http://dx.doi.org/10.1002/9781118542392.ch7, URL http://doi.wiley.com/10.1002/9781118542392.ch7.

Jokiel, P.L., Rodgers, K.S., Brown, E.K., Kenyon, J.C., Aeby, G., Smith, W.R., Farrell, F., 2015. Comparison of methods used to estimate coral cover in the Hawaiian Islands. PeerJ 2015 (5), http://dx.doi.org/10.7717/peerj.954.

Jones, D.O.B., Bett, B.J., Tyler, P.A., 2007. Megabenthic ecology of the deep Faroe–Shetland channel: A photographic study. Deep Sea Res. I: Oceanogr. Res. Pap. 54 (7), 1111–1128. http://dx.doi.org/10.1016/j.dsr.2007.04.001.

Kazanidis, G., Vad, J., Henry, L.A., Neat, F., Berx, B., Georgoulas, K., Roberts, J.M., 2019. Distribution of deep-sea sponge aggregations in an area of multisectoral activities and changing oceanic conditions. Front. Mar. Sci. 6 (163), 1–15. http://dx.doi.org/10.3389/fmars.2019.00163.

Kellner, K.F., Swihart, R.K., 2014. Accounting for imperfect detection in ecology: A quantitative review. PLoS One 9 (10), http://dx.doi.org/10.1371/journal.pone.0111436.

Langenkämper, D., Zurowietz, M., Schoening, T., Nattkemper, T.W., 2017. BIIGLE 2.0 - Browsing and annotating large marine image collections. Front. Mar. Sci. 4 (83), 1–10. http://dx.doi.org/10.3389/fmars.2017.00083.

Leujak, W., Ormond, R.F., 2007. Comparative accuracy and efficiency of six coral community survey methods. J. Exp. Mar. Biol. Ecol. 351 (1–2), 168–187. http://dx.doi.org/10.1016/j.jembe.2007.06.028.

Levin, L.A., Amon, D.J., Lily, H., 2020. Challenges to the sustainability of deep-seabed mining. Nat. Sustain. 3 (10), 784–794. http://dx.doi.org/10.1038/s41893-020-0558-x.

Levin, L.A., Bett, B.J., Gates, A.R., Heimbach, P., Howe, B.M., Janssen, F., McCurdy, A., Ruhl, H.A., Snelgrove, P., Stocks, K.I., Bailey, D., Baumann-Pickering, S., Beaverson, C., Benfield, M.C., Booth, D.J., Carreiro-Silva, M., Colaço, A., Eblé, M.C., Fowler, A.M., Gjerde, K.M., Jones, D.O.B., Katsumata, K., Kelley, D., Le Bris, N., Leonardi, A.P., Lejzerowicz, F., Macreadie, P.I., McLean, D., Meitz, F., Morato, T., Netburn, A., Pawlowski, J., Smith, C.R., Sun, S., Uchida, H., Vardaro, M.F., Venkatesan, R., Weller, R.A., 2019. Global observing needs in the deep ocean. Front. Mar. Sci. 6 (241), 1–32. http://dx.doi.org/10.3389/fmars.2019.00241.

Lindenmayer, D.B., Likens, G.E., 2010. The science and application of ecological monitoring. Biol. Cons. 143 (6), 1317–1328. http://dx.doi.org/10.1016/j.biocon.2010.02.013.

Millar, R.B., Fryer, R.J., 1999. Estimating the size-selection curves of towed gears, traps, nets and hooks. Rev. Fish Biol. Fish. 9, 89–116. http://dx.doi.org/10.1023/A:1008838220001.

Miller, J., Muller, E., Rogers, C., Waara, R., Atkinson, A., Whelan, K.R., Patterson, M., Witcher, B., 2009. Coral disease following massive bleaching in 2005 causes 60% decline in coral cover on reefs in the US Virgin Islands. Coral Reefs 28 (4), 925–937. http://dx.doi.org/10.1007/s00338-009-0531-7.

Miloslavich, P., Bax, N.J., Simmons, S.E., Klein, E., Appeltans, W., Aburto-Oropeza, O., Garcia, M.A., Batten, S.D., Benedetti-Cecchi, L., Checkley, D.M., Chiba, S., Duffy, J.E., Dunn, D.C., Fischer, A., Gunn, J., Kudela, R., Marsac, F., Muller-Karger, F.E., Obura, D., Shin, Y.J., 2018. Essential ocean variables for global sustained observations of biodiversity and ecosystem changes. Global Change Biol. 24 (6), 2416–2433. http://dx.doi.org/10.1111/gcb.14108.

Morris, K.J., Bett, B.J., Durden, J.M., Huvenne, V.A.I., Milligan, R., Jones, D.O.B., McPhail, S., Robert, K., Bailey, D.M., Ruhl, H.A., 2014. A new method for ecological surveying of the abyss using autonomous underwater vehicle photography. Limnol. Oceanogr.-Methods 12, 795–809. http://dx.doi.org/10.4319/lom.2014.12.795.

Obura, D., 2018. Essential Ocean Variables (EOV) for Biology and Ecosystems: Hard coral cover and composition. URL https://oceanexpert.org/downloadFile/41062.

Obura, D.O., Aeby, G., Amornthammarong, N., Appeltans, W., Bax, N., Bishop, J., Brainard, R.E., Chan, S., Fletcher, P., Gordon, T.A., Gramer, L., Gudka, M., Halas, J., Hendee, J., Hodgson, G., Huang, D., Jankulak, M., Jones, A., Kimura, T., Levy, J., Miloslavich, P., Chou, L.M., Muller-Karger, F., Osuka, K., Samoilys, M., Simpson, S.D., Tun, K., Wongbusarakum, S., 2019. Coral reef monitoring, reef assessment technologies, and ecosystem-based management. Front. Mar. Sci. 6 (SEP), 1–21. http://dx.doi.org/10.3389/fmars.2019.00580.

Olmstead, M.A., Wample, R., Greene, S., Tarara, J., 2004. Nondestructive measurement of vegetative cover using digital image analysis. HortScience 39 (1), 55–59. http://dx.doi.org/10.21273/hortsci.39.1.55.

Orejas, C., Ferrier-Pagès, C., Reynaud, S., Gori, A., Beraud, E., Tsounis, G., Allemand, D., Gili, J.M., 2011. Long-term growth rates of four Mediterranean cold-water coral species maintained in aquaria. Mar. Ecol. Prog. Ser. 429, 57–65. http://dx.doi.org/10.3354/meps09104.

OSPAR Commission, 2008. Case Reports for the OSPAR List of threatened and/or declining species and habitats. In: Biodiversity Series, URL https://qsr2010.ospar.org/media/assessments/p00358_case_reports_species_and_habitats_2008.pdf.

Pavoni, G., Corsini, M., Ponchio, F., Muntoni, A., Edwards, C., Pedersen, N., Sandin, S., Cignoni, P., 2021. TagLab: AI-assisted annotation for the fast and accurate semantic segmentation of coral reef orthoimages. J. Field Robotics (August), http://dx.doi.org/10.1002/rob.22049.

Perkins, N., Zhang, Z., Monk, J., Barrett, N., 2022. The annotation approach used for marine imagery impacts the detection of temporal trends in seafloor biota. Ecol. Indic. 140 (June), 109029. http://dx.doi.org/10.1016/j.ecolind.2022.109029.

Piechaud, N., Howell, K.L., 2022. Fast and accurate mapping of fine scale abundance of a VME in the deep sea with computer vision. Ecol. Inform. 71 (August), 101786. http://dx.doi.org/10.1016/j.ecoinf.2022.101786.

Price, D.M., Robert, K., Callaway, A., Lo lacono, C., Hall, R.A., Huvenne, V.A., 2019. Using 3D photogrammetry from ROV video to quantify cold-water coral reef structural complexity and investigate its influence on biodiversity and community assemblage. Coral Reefs 38 (5), 1007–1021. http://dx.doi.org/10.1007/s00338-019-01827-3.

Purser, A., Bergmann, M., Lundälv, T., Ontrup, J., Nattkemper, T.W., 2009. Use of machine-learning algorithms for the automated detection of cold-water coral habitats: A pilot study. Mar. Ecol. Prog. Ser. 397, 241–251. http://dx.doi.org/10.3354/meps08154.

Rasmussen, C.E., Williams, C.K.I., 2006. Gaussian Processes for Machine Learning, vol. 7, (no. 5), The MIT Press, Massachusetts Institute of Technology, p. 266, URL http://www.gaussianprocess.org/gpml.

Roff, D.A., 2001. The bootstrap. In: Introduction to Computer-Intensive Methods of Data Analysis in Biology. Cambridge University Press, Cambridge, pp. 66–101. http://dx.doi.org/10.1017/CBO9780511616785.005, URL https://www.cambridge.org/core/product/identifier/CBO9780511616785A009/type/book_part.

Rogers, A.D., Laffoley, D., Polunin, N., Tittensor, D.P., 2013. Ocean conservation: Current challenges and future opportunities. In: Key Topics in Conservation Biology, vol. 2, pp. 161–183. http://dx.doi.org/10.1002/9781118520178.ch10.

Rowden, A.A., Pearman, T.R., Bowden, D.A., Anderson, O.F., Clark, M.R., 2020. Determining coral density thresholds for identifying structurally complex vulnerable marine ecosystems in the deep sea. Front. Mar. Sci. 7 (February), 1–11. http://dx.doi.org/10.3389/fmars.2020.00095.

Sayer, M.D., Poonian, C., 2007. The influences of census technique on estimating indices of macrofaunal population density in the temperate rocky subtidal zone. Underw. Technol. 27 (3), 119–139. http://dx.doi.org/10.3723/175605407783360053.

Schoening, T., Kuhn, T., Jones, D.O., Simon-Lledo, E., Nattkemper, T.W., 2016. Fully automated image segmentation for benthic resource assessment of poly-metallic nodules. Methods Oceanogr. 15–16, 78–89. http://dx.doi.org/10.1016/j.mio.2016.04.002.

Schoening, T., Purser, A., Langenkämper, D., Suck, I., Taylor, J., Cuvelier, D., Lins, L., Simon-Lledó, E., Marcon, Y., Jones, D.O.B., Nattkemper, T., Köser, K., Zurowietz, M., Greinert, J., Gomes-Pereira, J., 2020. Megafauna community assessment of polymetallic-nodule fields with cameras: Platform and methodology comparison. Biogeosciences 17 (12), 3115–3133. http://dx.doi.org/10.5194/bg-17-3115-2020.

Schulz, E., Speekenbrink, M., Krause, A., 2018. A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. J. Math. Psych. 85, 1–16. http://dx.doi.org/10.1016/j.jmp.2018.03.001.

Stepputtis, D., Santos, J., Herrmann, B., Mieske, B., 2016. Broadening the horizon of size selectivity in trawl gears. Fish. Res. 184, 18–25. http://dx.doi.org/10.1016/j.fishres.2015.08.030.

Taylor, M.L., Gwinnett, C., Robinson, L.F., Woodall, L.C., 2016. Plastic microfibre ingestion by deep-sea organisms. Sci. Rep. 6 (May), 1–9. http://dx.doi.org/10.1038/srep33997.

Terpilowski, M.A., 2019. scikit-posthocs: Pairwise multiple comparison tests in Python. J. Open Source Softw. 4 (36), 1169. http://dx.doi.org/10.21105/joss.01169.

Thornton, B., Bodenmann, A., Yamada, T., Stanley, D., Massot-Campos, M., Huvenne, V., Durden, J., Bett, B., Ruhl, H., Newborough, D., 2021. Visualizing multi-hectare seafloor habitats with BioCam. Oceanography 34 (4), 92–93. http://dx.doi.org/10.5670/oceanog.2021.supplement.02-34.

Tong, S., Cardinal, H.N., McLoughlin, R.F., Downey, D.B., Fenster, A., 1998. Intra- and inter-observer variability and reliability of prostate volume measurement via two-dimensional and three-dimensional ultrasound imaging. Ultrasound Med. Biol. 24 (5), 673–681. http://dx.doi.org/10.1016/S0301-5629(98)00039-8.

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. http://dx.doi.org/10.1016/j.jembe.2012.04.018.

Vallat, R., 2018. Pingouin: Statistics in Python. J. Open Source Softw. 3 (31), 1026. http://dx.doi.org/10.21105/joss.01026.

Van Coillie, F.M., Gardin, S., Anseel, F., Duyck, W., Verbeke, L.P., De Wulf, R.R., 2014. Variability of operator performance in remote-sensing image interpretation: The importance of human and external factors. Int. J. Remote Sens. 35 (2), 754–778. http://dx.doi.org/10.1080/01431161.2013.873152.

Victorero, L., Blamart, D., Pons-Branchu, E., Mavrogordato, M.N., Huvenne, V.A., 2016. Reconstruction of the formation history of the Darwin Mounds, N Rockall Trough: How the dynamics of a sandy contourite affected cold-water coral growth. Mar. Geol. 378, 186–195. http://dx.doi.org/10.1016/j.margeo.2015.12.001.

Vieira, R.P., Bett, B.J., Jones, D.O., Durden, J.M., Morris, K.J., Cunha, M.R., Trueman, C.N., Ruhl, H.A., 2020. Deep-sea sponge aggregations (Pheronema carpenteri) in the Porcupine Seabight (NE Atlantic) potentially degraded by demersal fishing. Prog. Oceanogr. 183, 102189. http://dx.doi.org/10.1016/j.pocean.2019.102189.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., Vijaykumar, A., Bardelli, A.P., Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C.N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D.A., Hagen, D.R., Pasechnik, D.V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G.A., Ingold, G.L., Allen, G.E., Lee, G.R., Audren, H., Probst, I., Dietrich, J.P., Silterra, J., Webber, J.T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J.L., de Miranda Cardoso, J.V., Reimer, J., Harrington, J., Rodríguez, J.L.C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N.J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P.A., Lee, P., McGibbon, R.T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T.J., Robitaille, T.P., Spura, T., Jones, T.R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y.O., Vázquez-Baeza, Y., 2020. SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nature Methods 17 (3), 261–272. http://dx.doi.org/10.1038/s41592-019-0686-2.

Warwick, R.M., Clarke, K.R., 1996. Relationships between body-size, species abundance and diversity in marine benthic assemblages: Facts or artefacts? J. Exp. Mar. Biol. Ecol. 202 (1), 63–71. http://dx.doi.org/10.1016/0022-0981(96)00031-7.

West, G., Bodenmann, A., Newborough, D., Thornton, B., 2020. Resolution and coverage - The best of both worlds in the BioCam 3D visual mapping project. J. Ocean Technol. 15 (3), 67–76, URL https://www.thejot.net/article-preview/?show_article_preview=1186.

Wheeler, A.J., Kozachenko, M., Henry, L.A., Foubert, A., de Haas, H., Huvenne, V.A., Masson, D.G., Olu, K., 2011. The Moira Mounds, small cold-water coral banks in the Porcupine Seabight, NE Atlantic: Part A-an early stage growth phase for future coral carbonate mounds? Mar. Geol. 282 (1–2), 53–64. http://dx.doi.org/10.1016/j.margeo.2010.08.006.

Wileman, D.A., Ferro, R.S.T., Fonteyne, R., Millar, R.B., 1996. Manual of Methods of Measuring the Selectivity of Towed Fishing Gears. In: ICES Cooperative Research, (no. 215), ICES, p. 126. http://dx.doi.org/10.17895/ices.pub.4628.

Williams, S.B., Pizarro, O., Steinberg, D.M., Friedman, A., Bryson, M., 2016. Reflections on a decade of autonomous underwater vehicles operations for marine survey at the Australian Centre for Field Robotics. Annu. Rev. Control 42, 158–165. http://dx.doi.org/10.1016/j.arcontrol.2016.09.010.

Yamada, T., Prügel-Bennett, A., Thornton, B., 2020. Learning features from georeferenced seafloor imagery with location guided autoencoders. J. Field Robotics 38 (1), 52–67. http://dx.doi.org/10.1002/rob.21961.

Zurowietz, M., Langenkämper, D., Hosking, B., Ruhl, H.A., Nattkemper, T.W., 2018. MAIA—A machine learning assisted image annotation method for environmental monitoring and exploration. PLoS One 13 (11), http://dx.doi.org/10.1371/journal.pone.0207498.