Received: 28 April 2022 | Accepted: 23 October 2022

DOI: 10.1111/2041-210X.14029

Methods in Ecology and Evolution

RESEARCH ARTICLE

Digitizing the coral reef: Machine learning of underwater spectral images enables dense taxonomic mapping of benthic habitats

Daniel Schürholz[1,2] | Arjun Chennu[2]

[1]Microsensor Group, Max Planck Institute for Marine Microbiology, Bremen, Germany

[2]Data Science and Technology, Leibniz Centre for Tropical Marine Research, Bremen, Germany

Correspondence
Daniel Schürholz
Email: dschuerh@mpi-bremen.de

Arjun Chennu
Email: arjun.chennu@leibniz-zmt.de

Funding information
European Union's Horizon 2020, Grant/Award Number: 813360

Handling Editor: Oscar Gaggiotti

Abstract

1. Coral reefs are the most biodiverse marine ecosystems, and host a wide range of taxonomic diversity in a complex spatial community structure. Existing coral reef survey methods struggle to accurately capture the taxonomic detail within the complex spatial structure of benthic communities.
2. We propose a workflow to leverage underwater hyperspectral image transects and two machine learning algorithms to produce dense habitat maps of 1150 m^2 of reefs across the Curaçao coastline. Our multi-method workflow labelled all 500+ million pixels with one of 43 classes at taxonomic family, genus or species level for corals, algae, sponges, or to substrate labels such as sediment, turf algae and cyanobacterial mats.
3. With low annotation effort (only 2% of pixels) and no external data, our workflow enables accurate (Fbeta of 87%) survey-scale mapping, with unprecedented thematic detail and with fine spatial resolution (2.5 cm/pixel). Our assessments of the composition and configuration of the benthic communities of 23 image transects showed high consistency.
4. Digitizing the reef habitat and community structure enables validation and novel analysis of pattern and scale in coral reef ecology. Our dense habitat maps reveal the inadequacies of point sampling methods to accurately describe reef benthic communities.

KEYWORDS
coral reefs, habitat mapping, hyperspectral imaging, machine learning, survey scale mapping, thematic detail

1 | INTRODUCTION

Under rapidly changing environmental conditions, the need for accurate and speedy ecological assessment of marine and freshwater ecosystems has greatly increased. This is particularly pressing for coral reefs, which are the most biologically diverse marine ecosystems on the planet, but have suffered significant deterioration in recent years due to a variety of stressors, such as tourism overuse, destructive fishing practices, land-based pollution and climate change (Cesar et al., 2003; Hughes et al., 2020). Continued stress

on coral reefs deteriorates their health, leading to increased coral bleaching, coral mortality, disease outbreaks, loss of coral dominance and diversity loss (Burke & Spalding, 2004). In turn, the deterioration of coral reef health world-wide will endanger the ecosystem services that these reefs provide (i.e. shoreline protection, bioprospecting, food production, etc.) to coastal populations and other associated systems (Hoegh-Guldberg et al., 2017; Moberg & Folke, 1999). Furthermore, this long-term degradation of reefs confounds an intergenerational understanding of baseline reef health that informs reef restoration and management interventions (Muldrow et al., 2020), thus highlighting the need for objective assessments of reef health through monitoring.

Modern reef monitoring efforts focus on the creation of benthic habitat maps, as they capture the spatial distributions of species and habitat features (Guisan et al., 2013; Roelfsema et al., 2020). Such information captured over a long time series forms the basis of scientific evaluation of the ecosystem's evolution, and underpins the decisions for management, conservation and restoration (Foo & Asner, 2019). The spatial, temporal and thematic scales of ecosystem mapping have a critical influence on the viability of specific scientific analyses (Lecours, 2017), such as elucidating functional drivers, detecting community phase shifts or signalling deterioration of habitats. Most reef inventories compiled from in-situ surveys lack sufficient taxonomic and spatial detail, and have been carried out in only 0.01%–0.1% of coral reef regions world-wide (Hochberg & Gierach, 2021). In addition, many surveys do not report any uncertainty information that limits the utility of the data for scaling up studies to the ecosystem-level (Reverter et al., 2022). Thus, a priority for future in-situ reef surveys should be wider biogeographic coverage, clearer uncertainty estimates and deeper taxonomic and spatial detail at the survey scale.

Satellites are increasingly used to map shallow benthic habitats and analyse regional and global phenomena affecting coral reefs (Hedley et al., 2016; Heron et al., 2016). With recent enhanced spectral and spatial resolutions of remotely sensed images (0.5–10 m per pixel), better reef monitoring products have been enabled, such as geomorphological zonation of reefs (Kennedy et al., 2021) and benthic habitat maps (Lyons et al., 2020; Roelfsema et al., 2021). However, validating the accuracy of satellite-derived maps is a difficult task, impeded by the lack of in-situ validation datasets and the lack of error estimation in existing datasets (Hochberg & Gierach, 2021; Phinn et al., 2012). While remote sensing offers a viable approach for large scale analyses of reefs, current satellite sensors lack spatial resolution to represent small organisms (<0.5 m) and the spectral resolution to potentially differentiate organisms to a deep taxonomic description (Hochberg et al., 2003; Muller-Karger et al., 2018). In contrast, in-situ surveys can provide enhanced spatial and spectral resolutions in underwater imagery, made available by advancements in instrumentation and robotic platforms (Chennu et al., 2017), both aerial (Casella et al., 2017) and underwater (Armstrong et al., 2019). Improvements in machine learning (ML), especially with the application of artificial neural networks, have contributed to better accuracy and throughput of efforts in automated classification (Beijbom et al., 2015; González-Rivero et al., 2020) and semantic segmentation (Alonso et al., 2019; Pavoni et al., 2020) of benthic images. Carefully designed 'ground-truthing', produced from images acquired via underwater/proximal sensing, and mapped through scalable and automated workflows, can provide a consistent source of validation for current and upcoming ecosystem-level studies.

Deriving validation support from in-situ surveys requires careful design conformity between the proximal and remote sensing campaigns (Roelfsema & Phinn, 2013). For example, the lack of conformity in the set of labels used between satellite and in-situ studies is a major confounding factor (Foody, 2004). The labelspace of global maps usually include broad reefgroups (coral, algae, sediment, etc.), some status indicators (dead, alive, bleached) or morphological descriptions (branching, massive, weedy; Kennedy et al., 2021). This multi-faceted and easy-to-interpret view of the reef structure is useful for coastal management (Roelfsema et al., 2020). However, describing the reef community with broad reefgroups can hide intra-group shifts and conceal key dynamics of coral reef communities (Brito-Millán et al., 2019). To enable these analyses, in-situ studies leverage thematic scales that identify organisms down to genus or species level, as well as different substrata (sand, rock, rubble) and the substrate-associated communities (cyanobacterial mats, turf algae), rendering a detailed view of the biotic and abiotic components (Althaus et al., 2015). Capturing community structure with a detailed labelspace is typically limited by the availability of expert time or by logistical constraints. For this reason, reef community structure, as assessed in a majority of reef studies, is severely undersampled—both spatially and thematically—with respect to habitat complexity, neglecting spatial distributions and locations of benthic components.

Here we demonstrate how to produce dense and detailed maps of coral reef habitats from underwater surveys (Figure 1). Dense means that all regions/pixels in each image transect are assigned a (biotic or abiotic) habitat label, resulting in full semantic segmentation of the transect without any 'background' or 'unknown' labels. Detailed refers to the thematic detail that is captured by the labels, either being taxonomic (species, genus, etc) or broad reefgroups (corals, sponges, etc). We leverage ML to automate the classification of underwater hyperspectral image transects captured over multiple weeks and locations along the coast of the Caribbean island of Curaçao (Chennu et al., 2017; Rashid & Chennu, 2020). Our workflow description (Figure 1) considers all the steps from the field survey to the classification of 500+ million pixels to the validation of aggregate habitat descriptions derived from the detailed habitat maps. We show how dense maps can be produced, with clear assessments of accuracy, into multiple thematic scales, either at a broad ('reefgroups') or taxonomic ('detailed') labelspace. By implementing two independent ML methods (neural networks and object-based ensemble classification) in parallel, we provide an assessment of the consistency between the reef community structures as described by the maps produced with each ML method. These ML methods

can be used to rapidly convert transect spectral data into habitat maps at the survey scale, without the need to augment the training data with external datasets. Finally, we exploit the dense habitat maps of island-wide transects to reveal the inadequacies of sparse point sampling methods to accurately describe reef benthic communities.

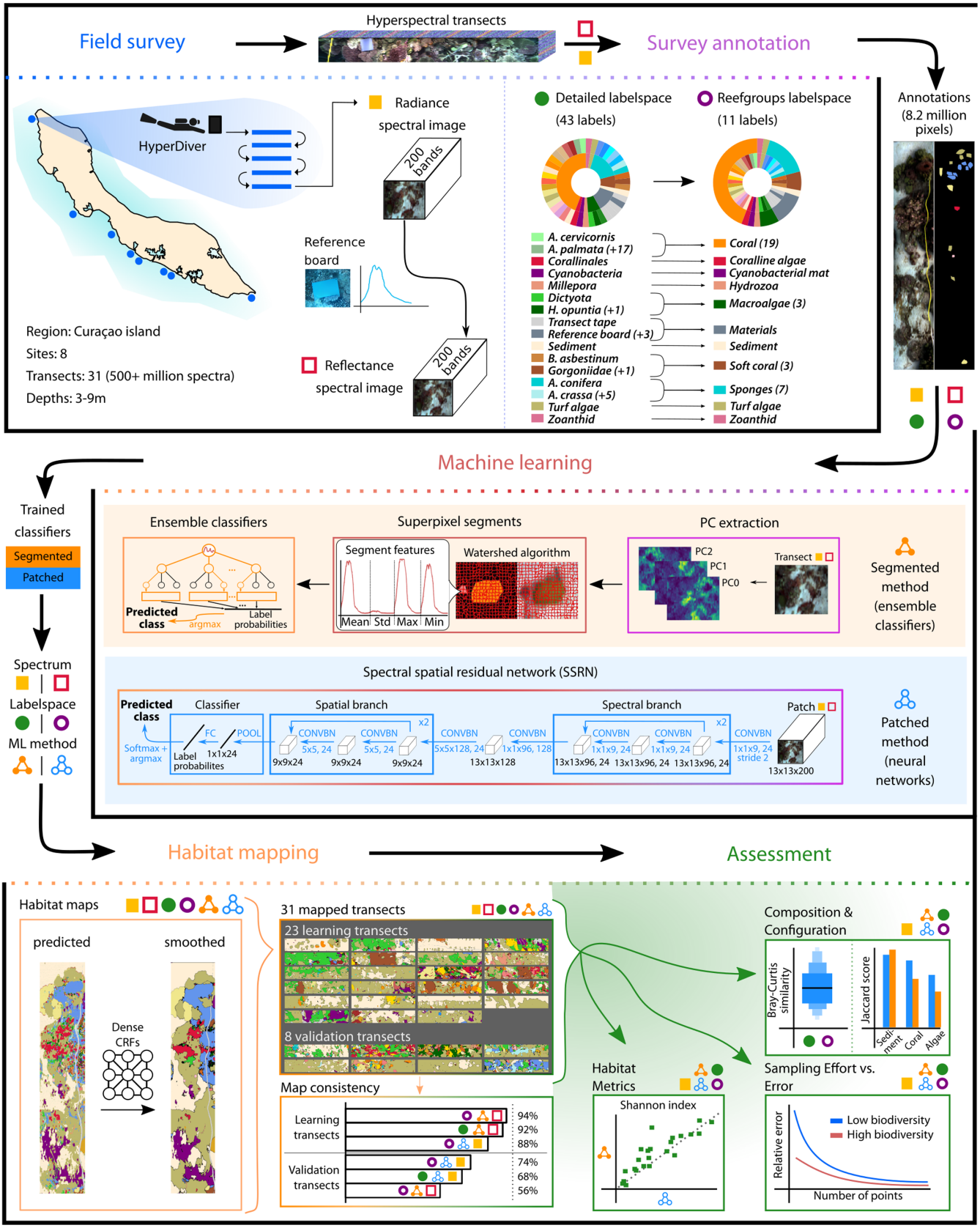


FIGURE 1 Schematic of scalable acquisition-mapping-assessment workflow for digitizing reef community structure. The underwater survey with the HyperDiver at eight study sites over 147 transects (50 m each) in Curaçao produced radiance and reflectance spectral images. A subset of 31 transect images was annotated into the detailed (43 classes) labelspace and aggregated into the reefgroups (11 classes) labelspace (see Figure S3 for complete labelspaces). Labelspaces can be adapted to encompass the underlying benthic community being studied. Annotated regions from 23 transects were used in two separate machine learning methods to classify each region of the spectral images into each labelspace independently. The segmented method used ensemble learning on image superpixel features, while the patched method used spatial-spectral neural network learning of image patches. The modular architecture of our workflow facilitates the usage of other ML models that produce probabilistic predictions. The classifiers were used to predict the label probabilities at each of 500+ million pixels in all 31 transects. The classifier-predicted probability maps were contextually smoothed and converted into densely labelled habitat maps. The habitat maps were assessed for their consistency with reference annotations as well as their ability to describe the composition and configuration of the benthic communities in the transects. The effort-versus-error relationship for point-count sampling of the reef habitats was assessed using virtual sampling of the 23 dense habitat maps.

2 | MATERIALS AND METHODS

2.1 | Underwater hyperspectral surveying

Underwater hyperspectral transects were acquired with the HyperDiver surveying system (Chennu et al., 2017), and a detailed description of the acquisition and processing is available in a data descriptor (Rashid & Chennu, 2020). A brief overview is provided here.

Hyperspectral transects were acquired in a survey conducted along the leeward coastline of the Caribbean island of Curaçao (Figure 1 'Field Survey'). At each of the eight survey sites, 10 to 20 transect images of 50 × 1 m area were acquired by divers at varying depths (3 to 9 m range). The resulting dataset contains 147 hyperspectral transects, from which 31 transects were selected for testing the proposed workflow. The 31 transects were comprised by 23 transects selected from the 3, 6 and 9 m depth for each site (except for one site where the 6 m transect was missing), and 8 additional transects randomly selected across the depth gradient. The hyperspectral push-broom imager captures lines of 640 pixels at a time. Each pixel contained 12-bits of radiometric information for each of the 480 wavelength bands in the 400–800 nm range. The spectral images were interpolated and reduced to 200 bands of 8-bit radiometric precision. Although the spectral transect images contain 60 times more colour information per pixel, the overall data size is smaller or comparable with high-resolution colour photography used in reef surveys.

The hyperspectral transect images were captured as radiance data under natural and varying light conditions due to depth, cloud cover, water surface conditions, etc. To be independent of these lighting conditions, each transect's radiance image was converted to pseudo-reflectance images (for brevity we refer to pseudo-reflectance as reflectance for the rest of this manuscript) by dividing out the average radiance signal of a grey reference board present at the ends of each transect plot.

2.2 | Benthic annotation and thematic flexibility

Annotations were created by human experts to support automated classification of the transects by ML methods. The annotations consisted of 2089 small polygons covering 8.2 million pixels with a corresponding habitat label across the 31 transects (Figures S1 and S2; Figure 1 'Survey annotation'). Biotic classes were annotated to the deepest taxonomic level possible, such as family, genus or species. Substrate classes are represented as sediment, cyanobacterial mat or turf algae. Survey materials were also included to give semantic labels to any object found in the transects, that is, transect tape or reference board. Three classes were removed given their very low representation in the selected transects (<2 annotated regions). The resulting 'detailed' labelspace had 43 final labels (Figure S3). Loosely speaking, the detailed labelspace describes the habitat in the perspective of a reef ecologist, aiming for full taxonomic resolution of the studied reef community.

From the perspective of a reef manager, typically interested in the broad demographic description of a reef, taxonomic detail is not useful or a detriment to management analysis. We created a labelspace to serve the perspective of a reef manager, by abstracting each label in the detailed labelspace to a corresponding broad reefgroup class (Figure S1). For example, the 19 coral species and genera were abstracted to a class called 'Coral'. The 11 resulting classes formed the 'reefgroup' labelspace. This thematic flexibility allowed us to run the ML setup with annotations in either labelspace, to measure the workflow's ability to classify into both labelspaces correctly. To compare classifications across labelspaces, we created an abstracted 'detailed-to-reefgroup' version of the detailed labelspace maps, that is, assigning all labels to their corresponding reefgroup label. Then the reef community composition was compared between the detailed-to-reefgroup maps and the reefgroup maps.

The reference annotations were created with a bias towards reducing human effort rather than providing uniform coverage of samples across the survey data (Rashid & Chennu, 2020). This resulted in a ML dataset with a relatively high degree of label imbalance, either when considered as a set of polygons or as a set of pixels across the annotated transects (Figures S1 and S2). The degree of imbalance meant that, for example, the 5 most abundant classes (*Sediment*, *Turf algae*, *Diploria strigosa*, *Dictyota*, *Siderastrea siderea*) had 789 polygons and 3.16 million pixels, while the 5 rarest classes (*Aplysina cauliformis*, *Briareum asbestinum*, *Dichocoenia stokesii*, *Zoanthid*, *Lobophora variegata*) had only 14 polygons and 18,464 pixels.

The annotated data consisted of 23 'learning transects' (373+ million pixels), that were used to train and test classifiers in the ML steps of our workflow, and 8 separate 'validation transects' (150+ million pixels) that were used to assess the performance of our workflow on unseen data outside the learning transects. Overall, it was possible to represent each annotated transect with two types of signal (radiance or reflectance) and labelspaces (detailed or reefgroups) for ML towards automated classification. The automation of mapping steps in our workflow (i.e. ML classifier training, map creation and assessment) was implemented with the snakemake workflow management system (Mölder et al., 2021).

2.3 | Machine learning for benthic mapping

ML classifiers were then created to predict the identity of each image region based on its spectral-spatial features (Figure 1). Two separate ML methods—'patched' and 'segmented'—were independently implemented for each combination of signal type (radiance, reflectance) and labelspace (detailed, reefgroups). The predictions of both methods for each image region were a probability value for each label/class in a labelspace.

For the patched method, a deep learning network called spectral-spatial residual network (Zhong et al., 2018) was used to train a classifier. This network identifies spectral and spatial features by first convolving 1D filters in a spectral branch and then convolving 2D filters in a spatial branch over square patches from the hyperspectral image. Our hyperparameter tuning experiments indicated good performance for parameter values close to original study (see Supplement). For each pixel in the image, the probability of labels is predicted for the central pixel based on the neighbouring pixels in a regular image patch (hence the name 'patched'). The image was padded with reflection of border pixels to enable selection of patches at the image edges. After training, each transect was mapped by passing every image patch through the trained network to obtain the predicted label probabilities for the central pixel.

The segmented method consisted of three sequential processes to obtain the label probabilities for each image region (Figure 1). The first step was to reduce each transect image to six principal components and calculate the mean at each pixel. The second step was superpixel over-segmentation of the transect using the mean image as input to the watershed algorithm. The parameters for the watershed algorithm were a batch size of 2000×640 pixels, 12,000 markers per batch and a compactness of 1000. This reduced the transect image into a set of irregularly shaped superpixels, which were contiguous image segments of similar pixels (hence the name 'segmented'). Descriptors for each hyperspectral image segment were calculated for each spectral band: mean, standard deviation, minimum and maximum values. There were concatenated into a feature vector of 800 features for each image segment. These vectors were then used as input samples for the random forest ensemble classifier in the scikit-learn library. The parameters for the classifier used for transect mapping were 300 base estimators, 2 minimum leaves per tree, 25 as the maximum tree depth and a minimum of 3 samples per tree split. The function used to measure the quality of a split was the Gini inequality function and the class weights were adjusted inversely to the class abundance in the samples. The classifier output was the label probabilities for an image segment, which were assigned to all the pixels within the segment to generate the class probability map of each transect.

Both methods were set up to take as input an image segment/patch from either of the signal types and produce the same output: an array of probability values for each label in the labelspace linked to all pixels inside a segment or the central pixel in a patch. The predictive performance of the trained classifiers was tested on a set of image annotations, which was spatially disjoint (no shared pixels) from the annotations used for training, as recommended in recent reviews (Paoletti et al., 2019). This testing set comprised of 15,496 image segments for the segmented method and 50,000 pixels for the patched method. The ML setup allowed us to use the segmented method (ensemble classifiers) and the patched method (deep learning) as interchangeable ML components in the workflow for scalable reef mapping.

The performance metrics used to evaluate the classifiers on the testing dataset were: overall accuracy (OA), recall (or producer accuracy), precision (or user accuracy), Fbeta (or F1-score) and Cohen's kappa (Figure 2; Table S1). OA was calculated by dividing the number of correctly predicted by total predicted segments/pixels. Recall, precision and Fbeta values were calculated for each class separately, and then aggregated using weighted averaging, with weights corresponding to the inverse of the class proportion in the testing dataset. Recall was calculated as the fraction of segments/pixels of a given class that were correctly classified. Precision for a class was the fraction of predicted segments/pixels that were annotated as that class. Fbeta is the harmonic mean of recall and precision. Cohen's kappa measures the performance of a classifier as a distance to an uninformed classifier (value of 0) and to a perfect classifier (value of 1), considering the dataset class imbalance.

We studied how both classifiers performed depending on the quantity and quality of annotated spectral pixels (Figure 3a,b). The quantity of annotated data was measured as unique pixels (in each patch or segment) used during training. Classifier performance was measured by training on various quantities of data but keeping the amount of computing effort constant (see *Supplement*). Furthermore, to measure the effect of the quality of the spectral information on the classifiers' performance, a subset ($N = [10, 25, 50, 100, 150]$) of equally spaced spectral bands were selected (out of the original 200 bands) from the transect images (Figure 3c). Each method was separately trained and tested on the transect images with the subsampled spectral bands.

The class probability maps of each transect from either ML method were smoothed by refining the probabilities with dense conditional random fields (DCRF; Krähenbühl & Koltun, 2011). DCRFs were used to update the probability of each pixel based on the

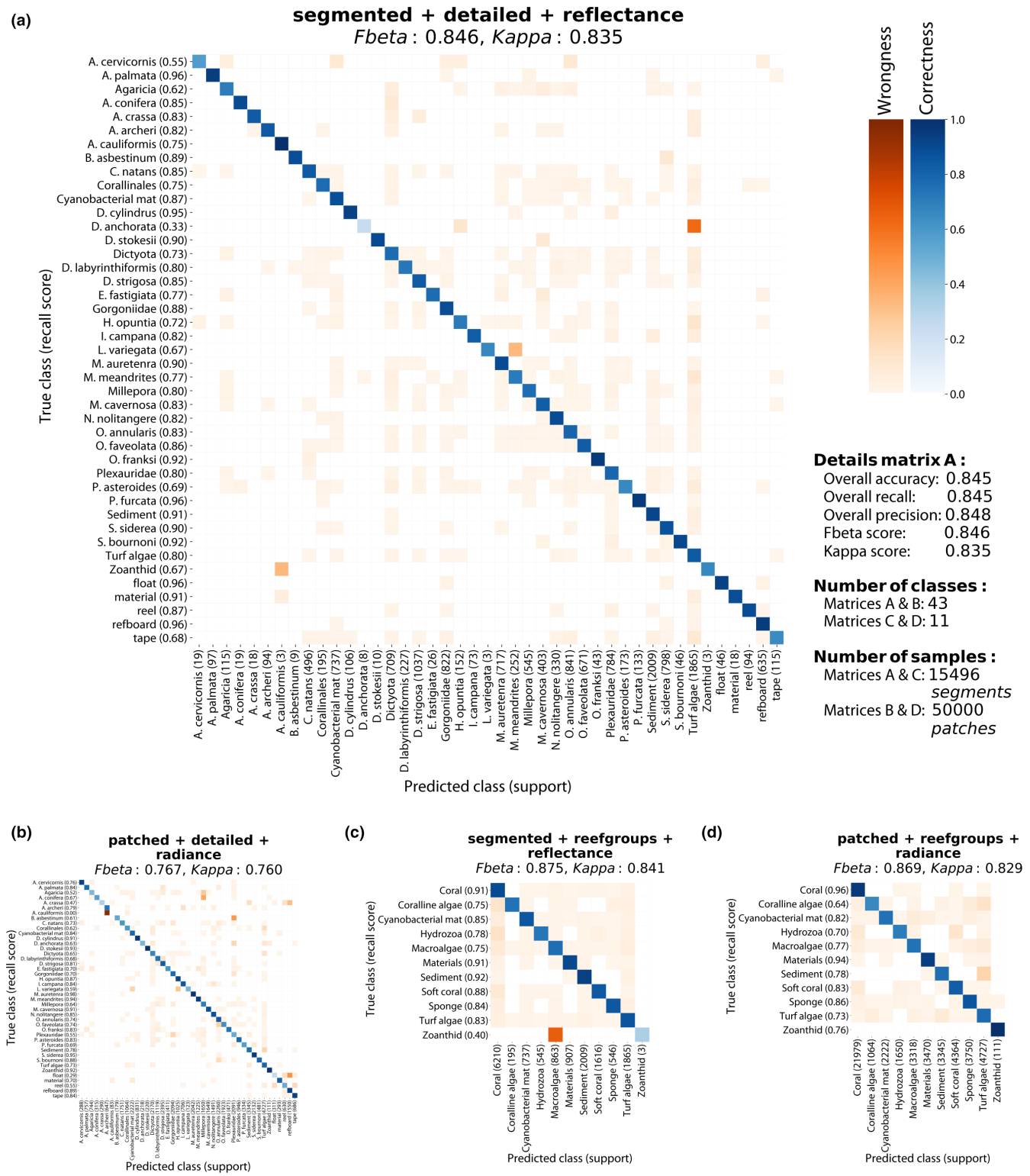


FIGURE 2 Performance evaluation of classifiers in detailed and reefgroups labelspaces. Recall confusion matrices of classifiers from a subset of ML method+labelspace+signal-type combinations were used to asses performance on a held-out testing set. The segmented method (a) had an excellent overall performance (84.6% Fbeta score; other metrics shown in the side notes) on the detailed labelspace. It presented little (<3%) to minor (<20%) confusions for the rare classes such as *Zoanthid* and *D. anchorata*. The patched method (b) showed a lower overall performance (Fbeta of 76.7%) in the detailed labelspace; with similar confusion for rare classes, such as *A. cauliformis* and *refboard*. On the reefgroup labelspace, both the segmented (c) and patched (d) methods showed excellent overall performance (87.5% and 86.9% Fbeta, respectively), with high recall (92%) shown by the segmented method for *Sediment* and similarly by the patched method for *Corals* (96%). The segmented method showed some relevant confusion between the *Zoanthid* and *Macroalgae* classes, due to the rarity (only three segments) of *Zoanthids* in the dataset.

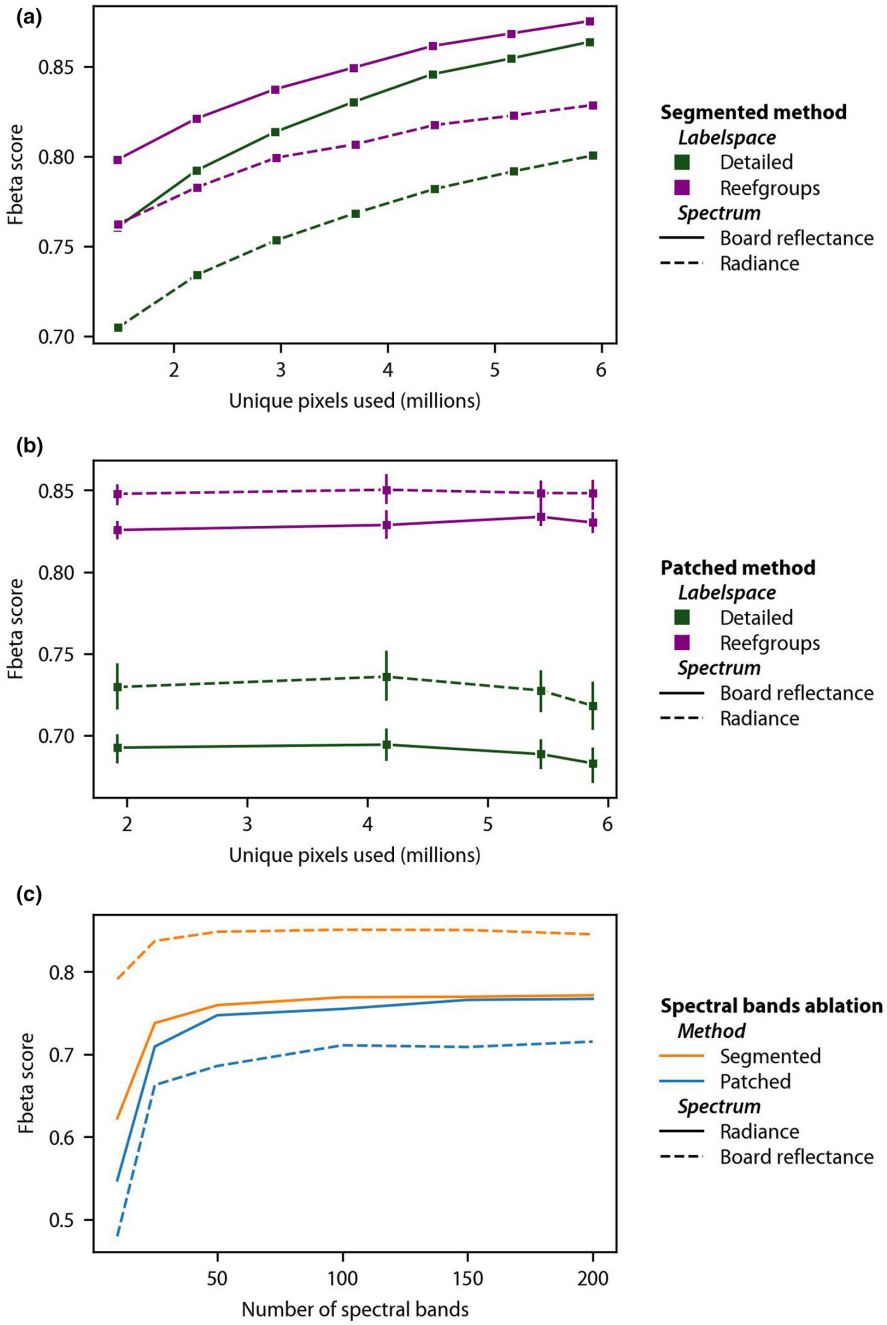


FIGURE 3 Effect of quantity and quality of spectral data on classifier performance. The effects of increasing the unique number of pixels in the data, but training the ML models with the same computational effort, were measured (Fbeta score) for both methods and labelspaces. The segmented method (a) showed improved performance whereas the patched method (b) showed little change in performance. Both methods performed better at predicting into the reefgroup labelspace (with 11 classes) than the detailed labelspace (with 43 classes) irrespective of the spectral signal type. (c) Both methods performed better when using an increasing number of (uniformly sampled) spectral bands for training, with limited improvement beyond 50 spectral bands.

surrounding context, that is, label probabilities. DCRFs interconnect every pixel in an image through a graph model, thus allowing fusing of long-range and short-range context within the image. The class probability maps were used as unary potential inputs to the DCRF, to obtain the smoothed probability maps (Figure 1 'Habitat mapping'; Figures 4a,b and 5).

Each class probability map—from either ML method and with or without smoothing—was converted to a class map by assigning the identity of the class with the highest probability at each transect image location. The result was a categorical habitat map where every pixel was assigned to one label in the labelspace of the trained classifier.

2.4 | Comparison of community structure

We assessed and compared the compositional and configurational structures of the reef benthos from the 23 (learning) transects distributed across Curaçao island (Figure 1 'Assessment'; Figure 6). Since both ML methods independently generated habitat maps in each labelspace of each transect, the habitat maps derived from each method were used for pairwise comparisons of the transect's community structures. For each transect the percentage cover (P_i) was calculated as:

$$P_i = 100 \times (C_i / N),$$

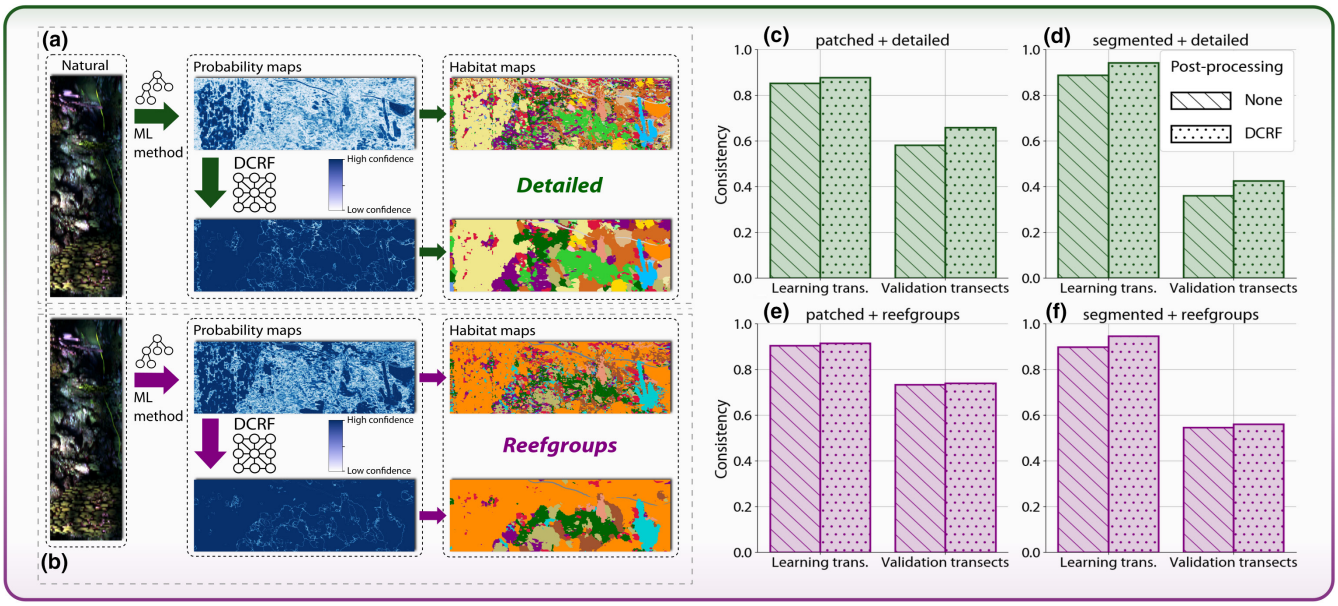


FIGURE 4 Smooth and consistent habitat mapping. The process of producing consistent habitat maps in the detailed (a) and reefgroups (b) labelspaces. Habitat maps produced from the classifiers' probability maps exhibit small-scale spatial noise and regions of low confidence. Smoothing the label probabilities with DCRFs renders spatially cohesive maps. (see Figure S3 for colour legend.) (c–f) The use of DCRFs improved the map consistency in all cases. The patched method was better than the segmented method at generalization and was more accurate at mapping the validation transects, which were completely disjoint from the learning transects used for training the ML classifiers.

where C_i is the count of pixels of class i in the transect and N is the total pixel count in the transect.

As a diversity measure for each transect we used Shannon diversity index (H'), defined as

$$H' = -\sum_{i=1}^{R} p_i ln(p_i),$$

where p_i is the proportion of elements of a class i and R is the total number of classes in the labelspace.

To identify biases in the ML methods, we applied a Bland–Altman analysis on the habitat metrics derived for the transects from each ML method (Figure 6a–d). This analysis consists of two plots that help identify agreement between two quantitative methods of measurement. The first plots the values of both methods for the specific variable—percentage cover of a class or Shannon diversity index—against each other, to identify values deviating from the one-to-one correlation line. In the second plot, the differences of the paired measurements are plotted against their averages, to identify the mean of the difference and its ±1.96 standard deviation lines. The bias is read as the gap between the mean of the difference and the 0 difference line. The two methods are considered to be in agreement if 95% of the values lie within the standard deviation lines in the second plot.

To measure the compositional similarity between two classified maps we used the Bray–Curtis similarity (BCS) index defined as:

$$BCS = 1 - \frac{\sum_{i=1}^{R} |A_i - B_i|}{\sum_{i=1}^{R} |A_i + B_i|},$$

where A_i is the count of pixels of class i in map A and B_i is the count of pixels of class i in map B. R is the total number of classes in the labelspace. We used the 'braycurtis' function from scipy python library (Virtanen et al., 2020). The closer the BCS value is to 1 the more similar the composition of the two compared communities.

We measured the similarity in the configuration of the communities between each reefgroup habitat map and their corresponding detailed-to-reefgroup map provided by each method, by calculating the Jaccard score for each reefgroup class. The Jaccard score (J) is defined as

$$J_i = \frac{A_i \cap B_i}{A_i \cup B_i},$$

When the Jaccard score is 1, then the two maps are identical in configuration and when the score is 0 then the maps are entirely dissimilar. A high Jaccard score requires that across the habitat maps from both ML methods, both the identity and the location of the pixels are a match. Thus, it is a stringent measure of the similarity between two maps or sets.

2.5 | Effort-versus-error of point-count sampling

We conducted simulations to estimate the error associated with a certain level of sampling effort in assessing the diversity and coverage of key groups through sparse point sampling of transects.

For this simulated experiment, we selected 4 transects with Shannon diversity index from low to high (H' = {0.61, 1.26, 1.6, 2.64}). Each transect was divided into 50 non-overlapping quadrats of size 640 pixels × 406–705 pixels. Each of these quadrats was sparsely sampled with N = {5, 10, 20, 40, 80,160, 240, 320, 480, 640, 960} randomly selected points. For example, this means that when N = 5 points, 250 points (50 quadrats x 5 points) were randomly selected from the habitat map of the transect. The

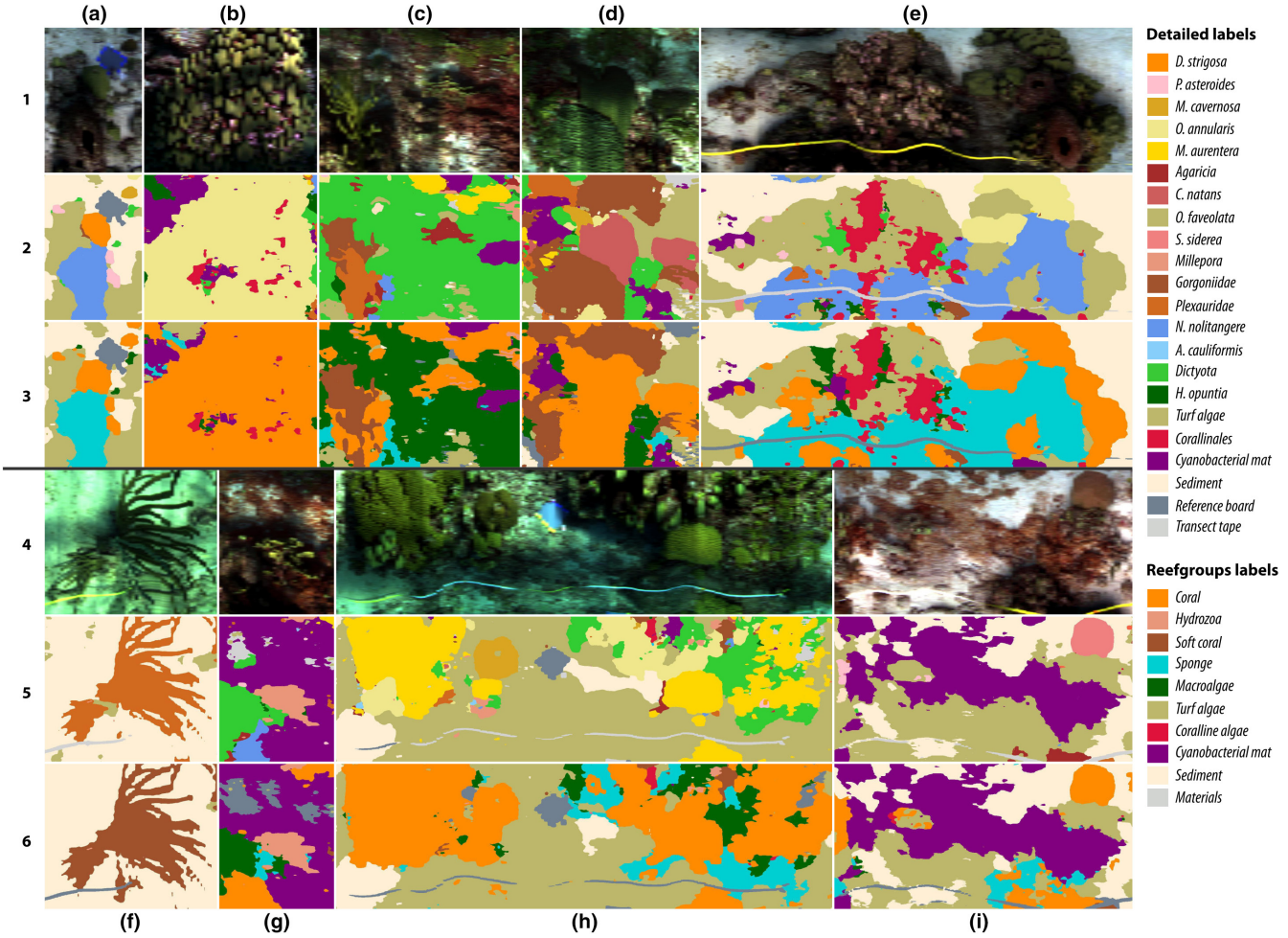
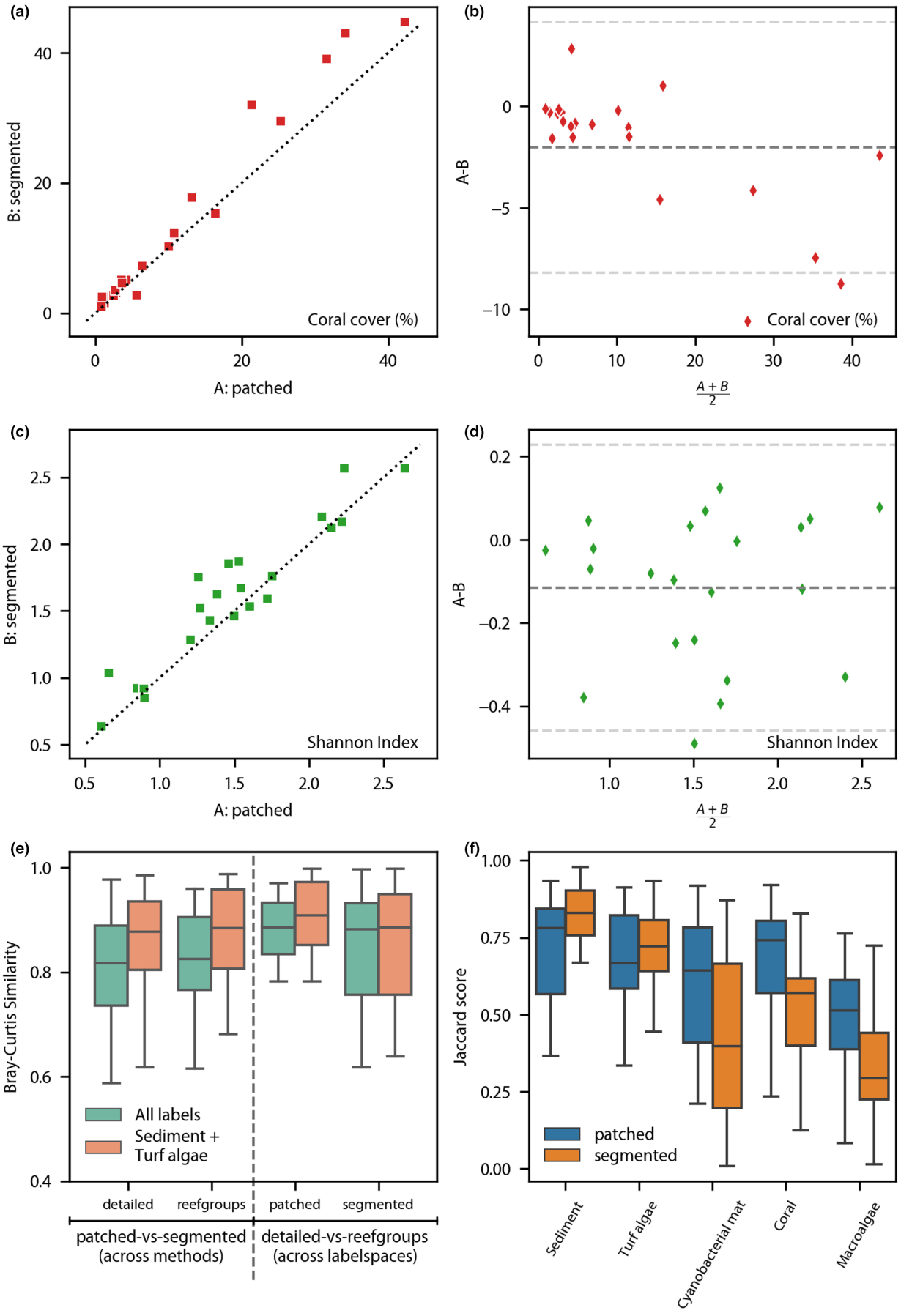


FIGURE 5 The rich structure of a digitized reef community. Sections of habitat maps produced with the patched ML method, with colours corresponding to classes from each labelspace (see legend and Figure S3). Rows 1 and 4 are the natural view, as would be seen by a human observer. Rows 2 and 5 show the sections in the detailed labelspace and rows 3 and 6 in the reefgroup labelspace. Our proposed workflow accurately discerns among a large labelspace and delineates complicated shapes of reef biota. In images a:2 and a:3, the habitat maps show correctly classified and well-delineated instances of three coral species (*D. strigosa*, *P. asteroides* and *M. cavernosa*), a sponge (*N. nolitangere*) as well as regions of *Sediment*, *Turf algae* and *Dictyota* macroalgae. The maps in f:5 and f:6 show another example of fine-grained segmentation of the branches of a specimen of the *Plexauridae* soft coral family. Comparing the maps in d:2 and d:3, or in h:5 and h:6, shows the number of different coral species that can be identified under the broad *Coral* group (in orange). Small encrusting taxa such as *Coralline algae* are visible in b:2 and e:2. The delineation of *Cyanobacterial mat* in g:2 and i:2, along with the many regions of *Sediment* and *Turf algae* represent substrate and microbial components of coral reef benthos.

FIGURE 6 Assessing the benthic community structure from dense maps. (a) Coral cover of 22 learning transects was compared from the habitat maps for each ML method. The segmented method predicted slightly more coral cover than the patched method in the transects. (b) No clear bias was noted for either method in the Bland–Altman plot. (c) The Shannon index values for 22 compared transects were highly correlated between the maps from the segmented and patched methods. The segmented method produced habitat maps with slightly higher Shannon diversity. (d) No significant bias was found in either ML method for the Shannon index comparison. (e) The median Bray–Curtis similarity between the mapped communities was ~80% across ML methods and ~88% across labelspaces. Note that the orange bars refer to a labelspace where *Sediment* and *Turf algae* were combined into a single class, resulting in a higher compositional similarity. (f) Configurational similarity assessed using the Jaccard index between the reefgroups and detailed-to-reefgroup habitat maps for the top-five dominant labels is shown. Given the 500+ millions of pixels in this assessment, the maps showed very good consistency in configuration.



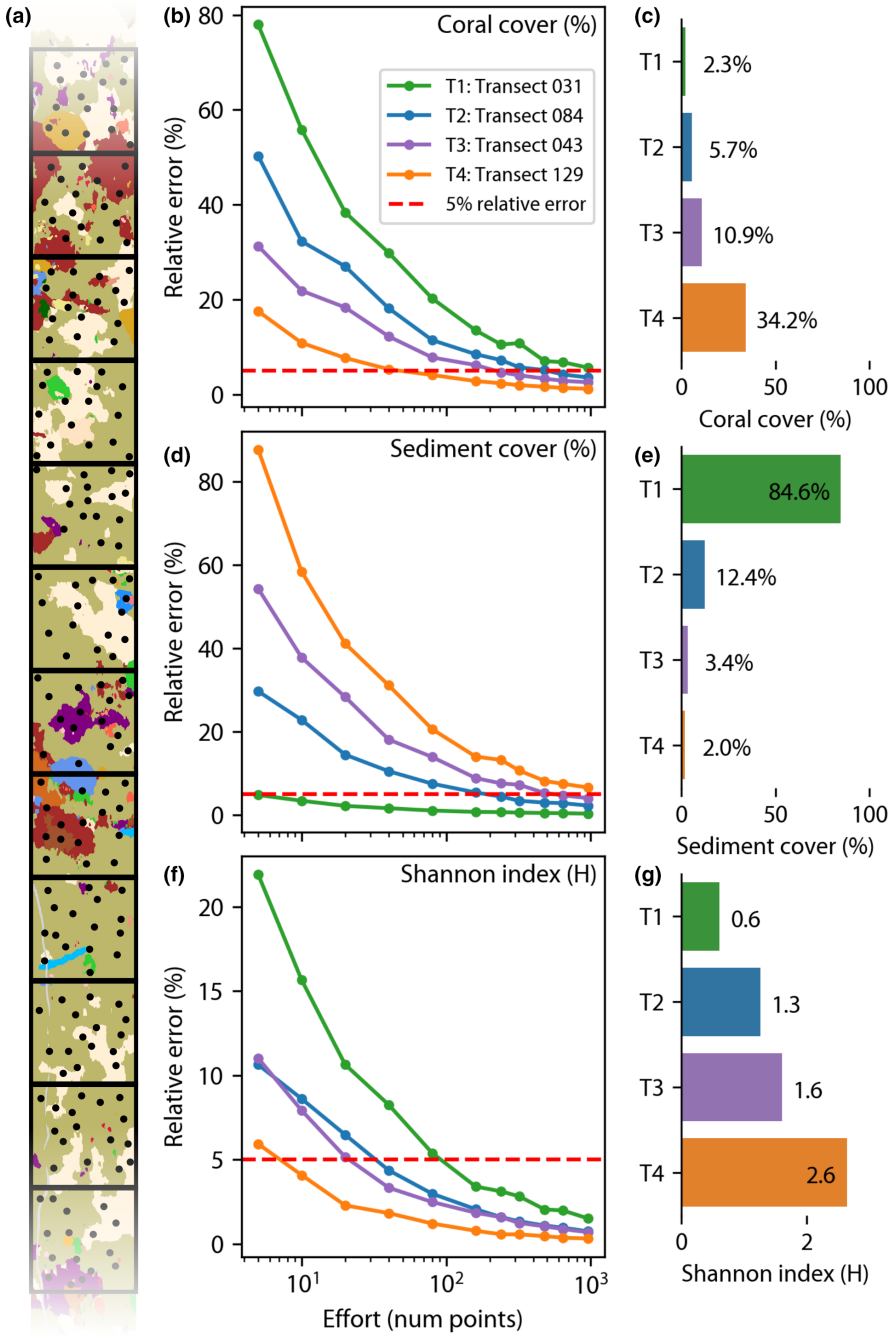


FIGURE 7 Effort-versus-error analysis for point-count sampling of reef community structure. (a) Schematic of the simulations of quadrat-wise random sparse sampling of dense habitat maps to estimate metrics (coverage or Shannon index) through point-count estimates of four transects with different biodiversity values (H ={0.61, 1.26, 1.6, 2.64}). The error in the habitat metric from sparse random points relative to the metric of the full map was calculated from repeated trials. The number of point samples per quadrat required to achieve a relative error lesser than 5% (dashed red line) was assessed. (b, c) The transect with the least coral coverage (T1) required more than 960 points per quadrat to estimate it within the 5% error limit, whereas the transect with the highest coral coverage (T4) required 80 or less points per quadrat. (d, e) The transect with the least sediment coverage (T4) required more than 960 points per quadrat to estimate sediment coverage within the 5% error limit, whereas the transect with the highest sediment coverage (T1) required only 5 points per quadrat. (f, g) The least diverse transect (T1) required 160 points, compared with 40 points per quadrat for the more diverse transect (T4), to estimate the Shannon diversity index within the 5% error limit. These results suggest that rarer species require more sampling effort and that over 80 points per quadrat should be used to estimate habitat metrics in reef transects where the expected diversity or coverage is not previously known.

random sampling of the quadrats was conducted 250 times for each effort level. From each set of subsampled points in each transect, the coral coverage, sediment coverage and Shannon diversity index were calculated. For each metric, the relative deviation of the value obtained from the subsampled points from the value obtained from all the points in each transect was calculated. We selected 5% relative error as the limit for acceptable error, similar to a previous simulation study (Pante & Dustan, 2012). The resulting error from changing sampling effort was compared for each of the transects containing a significantly different species diversity and coverage distribution (Figure 7).

3 | RESULTS

3.1 | Automated workflow for scalable benthic mapping

To measure the performance of the ML methods on the expert annotations several experiments with combinations of signal type and labelspace were executed (see Section 2). For the detailed labelspace, the segmented+reflectance combination had the best predictive performance with an Fbeta score of 84.5% (Figure 2a; Table S1). The classifier had 80% to 96% recall for a majority of the

43 labels with sufficient data support (see diagonal of Figure 2a). Some labels with low data support showed excellent recall (*Aplysina archeri*, *Aiolochroia crassa*, *B. asbestinum* and *D. stokesii*), while others showed significant errors (*Zoanthid*, *L. variegata* and *Desmapsamma anchorata*). Despite having to distinguish between 43 labels, the segmented+reflectance classifier had a high Cohen's kappa coefficient indicating performance which is 83.5% of a perfect classifier (Figure 2a; Table S1). For the same labelspace, the patched+radiance combination (Figure 2b; larger version in Figure S4) performed with a 9% lesser Fbeta score (76.7%). Most classes had a recall value between 70% and 90%, with some rare classes, such as *A. cauliformis* and *B. asbestinum*, having significant errors.

For the reefgroup labelspace, the best predictive performance was 87.5% in the segmented+reflectance combination (Figure 2c) followed closely by the patched+radiance combination with 86.9% (Figure 2d). The latter showed high recall values for all 11 classes, even reaching 96% correctness for the *Coral* class—which consists of 19 different coral genera and species (Figure S3). The highest confusion occurred between *Turf algae* and *Sediment* classes. For both ML methods classifiers had a Cohen's kappa score of 83% towards a perfect classifier (Figure 2; Table S1). Overall, both ML methods showed Fbeta scores between 72% and 87% with better performances on the reefgroup labelspace than on the detailed labelspace (Figure 2; Figures S4–S8; Table S1).

Both classifiers had different responses to the amount of unique input data seen during training. The classifier performance in the segmented method improved significantly with greater quantities of training data (Figure 3a). The greatest improvement was an Fbeta from 76% with 1.47 million pixels to 86% with 5.89 million pixels seen in the detailed labelspace using reflectance spectra. The performance also improved with greater quantities of radiance spectra, but overall the segmented method performed better on reflectance rather than radiance spectra. In contrast, the patched method showed no improvement with larger quantities of training data (Figure 3b). Instead, the performance in the detailed labelspace showed a 1% deterioration when the same computing effort was distributed over all the available data (5.93 million pixels). This seemingly unexpected result of poorer performance with more data can be understood by considering the lower number of training iterations over the larger dataset to maintain the same computing effort.

The impact of data quality, or spectral resolution, on predictive performance was assessed by using a subset of 10–100 spectral bands in the training data. In both the segmented and patched methods (Figure 3c), the predictive performance showed a strong 5% to 15% improvement when the number of spectral bands was increased from 10 to 25 with diminishing improvements when using 50+ spectral bands. Overall, the availability of greater spectral resolution, even when the bands were chosen without special consideration, had a large effect on the performance of the ML methods.

To produce habitat maps for further analysis, we selected one patched classifier that was trained on all 200 bands and with 62,500 patches (~4.1 million unique pixels) and one segmented classifier that was trained on all 200 bands and with 62,332 segments (~5.8 million unique pixels).

3.2 | Smooth habitat maps to digitize reef community structure

The label probability map obtained directly from the classifier's prediction showed generally noisy spatial distribution, with many areas of low confidence (Figure 4). This effect of low and noisy confidence was larger in the detailed than the reefgroup labelspace. Processing the predicted probability map with DCRFs produced a more uniform map of probabilities with high confidence except for the border pixels between adjacent targets (Figure 4c,e). The habitat maps from the DCRF-processed probability maps better delineated the benthic scene with smooth and contiguous regions (Figure 4a,b).

Beyond the visual cohesiveness, the smoothed habitat maps were quantitatively more consistent than the raw habitat maps in all combinations of ML methods and labelspaces for transects (Figure 4c,f; Figure S9A,B). The map consistency was measured as the average of the label accuracy in each of the annotation regions in all of the annotated transects, that is, regions used for both training and testing the ML methods. The consistency of the habitat maps in the regions of the validation transects were lower than in the learning transects (Figure 4c–f): consistency for the segmented method dropped from 94% to 43% and from 95% to 56% with the detailed and reefgroup labelspaces (Figure 4d,f), respectively, and from 88% to 66% and from 92% to 74% for the patched method (Figure 4c,e). This indicated that the patched method (convolutional neural networks) was better than the segmented method (ensemble object classifiers) at generalizing to unseen data. Classification of transects with the reflectance signals resulted in a large drop of consistency, with a worst case change from 92% for the learning transects to 18% for the validation transects (Figure S9). Overall, the best predictive performance on data from the validation transects, which was not used in any ML step, was from the patched method.

Smooth habitat maps in both labelspaces were produced for all 31 transects. With each transect approximately 50×1 m in size, this task involved assigning each of the 500+ million spectral pixels to one of 43 labels (detailed) or one of 11 labels (reefgroups) independently. Montages of the habitat maps for all transects were visualized (Figures S10–S13). A selection of interesting sections of these habitat maps were visualized (Figure 5) with the natural view (rows 1 and 4), the detailed map (rows 2 and 5) and the reefgroup map (rows 3 and 6) shown together. Despite overall conformity, the habitat map sections also display some confusions: different sections of the same colony in C:2 are assigned to *Plexauridae* and *Gorgoniidae*, which are both soft coral families with similar digitate morphologies, while in C:3 this same colony is assigned between *Coral* and *Soft coral*. The incorrectly labelled regions of *Neofibularia nolitangere* sponge in E:2 (and *Sponge* in E:3) along the image edge are also errors. These errors likely occur due to the poorly illuminated shadow regions that received predictions with low confidence, and then got reassigned to the *Sponge* class by the DCRF process due to nearby high-confidence regions. Note however that this did not occur for the *Transect tape* in the shadow regions which was predicted with high confidence.

3.3 | Assessing the community structure

A Bland–Altman analysis of the coral coverage (Figure 6a,b) and the Shannon diversity index (Figure 6c,d) from the reefgroup labelspace maps showed a high degree of correlation and low bias between the segmented and patched methods. The coverage of all other reefgroup classes, except for *Sponges*, derived from either mapping method, were comparable across the range of values in all the transects (Figure S14). The Bray–Curtis similarity, which compares the compositional structure between the communities in the maps of both methods, had a median value of 82% in the detailed and reefgroup labelspaces with a quartile range of 72%–89% and 75%–90%, respectively (Figure 6e). With the *Sediment* and *Turf algae* substrate classes merged together, the median value for the similarity rose to 89% for the detailed and reefgroup labelspaces, with a quartile range of 78%–92% and 79%–94%, respectively. This improvement in the similarity index indicates a large effect of the inherent definition problem of *Turf algae* on reef habitat mapping.

The similarity assessment across the learning transects for the patched method showed an 88% similarity median value with a quartile range of 84%–92%, while the segmented method showed an 87% median value with a quartile range of 77%–91%. With the *Sediment* and *Turf algae* classes merged, the patched method showed a similarity median of 90% (quartile range 88%–97%), whereas the segmented method showed a barely improved median similarity of 88% (quartile range 77%–93%). Our proposed workflow recovered a reef community composition that was highly consistent between the taxonomic and broad reefgroup descriptions of the reef benthos.

The spatial configuration analysis between the detailed-to-reefgroup and reefgroup maps showed that three classes had higher configurational similarity for the patched method than for the segmented method: *Cyanobacterial mat* (64% vs. 40%), *Coral* (74% vs. 57%) and *Macroalgae* (51% vs. 29%; Figure 6f). Two classes showed higher configurational similarity for the segmented method than the patched method: *Sediment* (83% vs. 78%) and *Turf algae* (72% vs. 67%). The Jaccard score was lower for both methods on rarer classes (Figure S15).

3.4 | Evaluating the effort-versus-error compromise in reef sampling

We exploited our dense and accurate habitat maps to revisit the effort-versus-error relationship of sparser reef sampling techniques (see *Methods* for point selection). The number of random point samples required to achieve a relative error lesser than 5% was assessed (Figure 7). To recover the hard coral coverage in transect T1—which had low biodiversity ($H' = 0.6$) and low coral coverage (2.3%)—960 random points per quadrat were required (Figure 7b,c). For the transect T4, with $H' = 2.6$ and coral coverage of 34.2%, 80 random points per quadrat were sufficient. Similarly, 960 points per quadrat were needed to recover the sediment coverage in transect T4, where sediment covers only 2% of the benthos (Figure 7d,e). In contrast, only 5 points per quadrat were required to capture the sediment coverage in transect T1, which has the highest sediment coverage (84.6%). The Shannon index was recovered with 10 sampling points per quadrat for transect T4 ($H' = 2.6$), with 40 points per quadrat for the transects T2 ($H' = 1.6$) and T3 ($H' = 1.3$), and with 160 points per quadrat for transect T1 ($H' = 0.6$; Figure 7f,g). Overall, higher sampling effort was required to accurately recover the coverage of rare species or to capture the Shannon biodiversity index of scenes with low biotic coverage.

4 | DISCUSSION

4.1 | Dense and detailed mapping of benthic communities

The presented reef mapping workflow was able to produce dense habitat maps with an unprecedented degree of thematic (43 labels) detail and high spatial (~2.5 cm/pixel) resolution. The rich spectral detail in the HyperDiver data was leveraged by ML classifiers to produce highly accurate habitat maps (87% Fbeta) with little annotation effort (2% pixels in ~20 hours). The two labelspaces used in the maps describe the reef benthic biodiversity down to genus and species level as well as abiotic and microbial components, such as sediment, turf algae and cyanobacterial mats. Our habitat maps provide a no-pixel-left-behind dense view of entire 50 m long transects, which allowed us to identify, localize and delineate the components of the surficial reef benthic community. Two ML methods with different complexities were used independently to produce dense and detailed habitat maps, thus facilitating objective comparison of reef descriptions at big data scale. Our workflow provides a deep description of community structure (diversity, coverage, composition and configuration), which demonstrated high convergence between both ML methods. Nonetheless, our detailed assessment indicates that deep learning classifiers (i.e. the patched method) are better at generalizing towards new and unseen datasets under comparable annotation and computational effort.

We designed the thematic detail in our workflow to target multiple user groups. Our workflow uses a detailed thematic scale in the form of 43 benthic categories that describe biotic and abiotic components of the reef habitat. Subsequently, a reefgroup labelspace, comprised of broad groups of reef biodiversity, was abstracted from the taxonomically detailed labelspace through interconnected ontologies, similar to hierarchical geomorphic zones developed previously (Kennedy et al., 2021). By independently mapping into the reefgroup labelspace we showed that the workflow consistently retrieved the composition and configuration of the reef transects across thematic resolutions (Figure 6e,f). This enables comparisons between our maps and previous/historical datasets that may have different thematic resolutions, as well as allowing for the workflow to 'translate' between the needs of different expert groups, such as reef ecologists or managers (Lecours et al., 2015; Roelfsema et al., 2021). We incorporated this thematic

flexibility in our workflow, so that it can be reused in other benthic mapping scenarios (i.e. other coral reef sites, seagrass meadows, rocky reef sites).

4.2 | Comparison of machine learning methods

To automate the digitization of reef communities and community structure, careful consideration of workflow parameters is recommended. In our workflow, we independently used two ML approaches: one based on object-based image analysis (i.e. segmented method) and the other on deep neural networks (i.e. patched method). Continual increase in complexity and specificity of ML tasks for automation impedes a clear judgement of an ML method's ability to generalize (Paoletti et al., 2019). We tested two ML methods with different operational paradigms on a non-overlapping dataset to explore the trade-off in performance versus complexity of automation. Both segmented and patched methods showed 80% ± 5% Fbeta scores in both labelspaces (Figure 2), the segmented approach performing better with reflectance data and the patched method with radiance data. We do not consider this to be generally indicative for future efforts, because both our ML models have no consideration of the optical physics between the two signal types. Another work targeting a similar sized labelspace (35 labels) achieves a mean pixel accuracy of 49.9% with a deep semantic segmentation network (DeepLabV3+) on sparse samples in 729 test images of coral reef orthophoto mosaics (Alonso et al., 2019). Given only 2% of annotated pixels, our workflow mapped, with higher accuracy on 43 labels, underwater transects with high natural variability. We show that both ML methods can produce accurate mapping of reef transects, apparently due to the spectral detail (Figure 3c). Nonetheless, we found significant differences in the data requirements of both ML methods, and in their generalization abilities.

To determine the data requirements for the ML algorithms, we assessed the performance of both methods—under a constant computational effort—based on the number of unique pixels (in segments or patches) used for training (Figure 3a,b). The patched method needed less data to achieve its peak performance under the same amount of computing power and annotation effort. Despite the variable lighting conditions and methodological artefacts between training and validation transects, the patched method classified into both labelspaces more consistently (Figure 4c–f). Although classifier performance metrics on the training transects are better for the segmented method (Figure 2), the patched method was 23% better at classifying out-of-distribution data (i.e. validation transects) in the detailed labelspace and 18% better in the reefgroup labelspace than the segmented method (Figure 4c,d). Given that expert annotation is the biggest bottleneck for reef survey analysis (Beijbom et al., 2015; Roelfsema & Phinn, 2013), the patched method, with its better generalization capability, provides better performance-per-human-effort compared with the segmented method.

4.3 | How well do the habitat maps capture the community structure?

The smoothed habitat maps from our workflow show spatial and thematic detail of the structure of the coral reef benthic community (Figure 5). Benthic targets are clearly separated into meaningful regions, which represent different substrata, different organisms of various sizes. Small coral colonies and intricate shapes of branching corals, soft corals and sponges, and even transect tapes are correctly delineated and classified. Our workflow is able to accurately map bare sediment, turf algae and cyanobacterial mats achieving a previously missing capability in reef habitat mapping: dense mapping of the microbial components of reef substrata, while integrating them into a benthic community labelspace. This can be used to quantify changes in abundance of cyanobacterial mats or turf algae, usually an indicator of reef deterioration (de Bakker et al., 2017).

The primary focus of our workflow development was to convert underwater spectral images into dense habitat maps. The target of our assessment was to go beyond comparing classifier-level metrics, and assess the final habitat maps for consistency and accuracy. We considered the goal of comparing the densely classified maps to photo-quadrats with point-count estimates (Rashid & Chennu, 2020), but the large difference in sample sizes—~2000 quadrat points versus millions of classified points—made the experiment statistically unsound. Given that 98% pixels (out of 500+ million) do not have reference label annotations and the complex spatial structure of the transects (Figures S10–S13), it is difficult to assess the maps accuracy on a pixel level. To overcome this limitation, we exploited the fact that the two ML methods produced the same type of output, but worked fully independent of each other in terms of method, input signal and parametrization. We compared the community structure of the dense maps in describing the same physical transect of the seafloor. To achieve this we used coverage, as well as composition and configuration metrics, which are key descriptors of community structure (Nowosad & Stepinski, 2019; Riitters, 2019). We consider that if the statistical properties of the habitat maps from independent methods are similar, our workflow will have succeeded in representing the true composition and configuration of the coral reef transects captured by the underwater surveys.

The Bland–Altman analysis showed that the habitat maps agree on coverage and diversity metrics, except for some small discrepancies (Figure 6a,b; Figure S14). These discrepancies are noticeable in coverage of corals, sponges and cyanobacterial mats which were overestimated by the segmented method in a few transects (Figure S14I,J). The demographic composition of the communities between pairs of dense habitat maps (considering all pixels) was highly similar as shown by the Bray–Curtis similarity index (Figure 6e). This shows that both ML methods independently ascribe similar classes and a similar number of those classes to the same transect. Furthermore, the assessment of the communities between the two labelspaces was also very similar (Figure 6e), providing confidence in the thematic flexibility of our workflow. We infer that automated mapping with ML methods of underwater hyperspectral transects

can handle intra-class variability (detailed labelspace) as well as inter-group variability (reefgroup labelspace) with high accuracy.

Going beyond the composition, we also assessed the spatial configuration similarity of the benthic community in the transects described by both ML methods using the Jaccard index. The regions from the abstracted detailed-to-reefgroup maps and direct reefgroup maps yielded Jaccard scores over 60% for the dominant labels (Figure 6f). Given that these comparisons are across hundreds of millions of pixels and over 43 different labels, these results indicate high configurational similarity between the maps. Therefore, our workflow is able to correctly localize and delineate important targets in benthic habitat maps, despite the degree of thematic detail. Nonetheless, these assessments are inter-comparisons within our workflow and a correct assessment of the configuration of the community structure would require dense manual annotations of the transects. The high degree of convergence in the community structure mapped independently by the two ML methods using two different hyperspectral signal types gives confidence in the ability of the workflow to produce accurate in-situ descriptions of coral reef habitats with high detail and analytical throughput.

4.4 | The effect of sampling effort on community structure

There has been some debate about the amount of sampling effort necessary to accurately measure reef community structure. In benthic surveys with photo-quadrats, the 'adequate' number of samples (points-per-quadrat or quadrat-per-area) to accurately describe the community has been a topic of debate (Dumas et al., 2009; Pante & Dustan, 2012; Perkins et al., 2016). The goal is to find a good balance between expert effort (labelling the sampling units in images) and reliability in the derived scene description. At the reef area scale, the number of quadrats per area (i.e. sampling density) is an important determinant of the precision of coverage estimations (Lechene et al., 2019; Perkins et al., 2016). At the quadrat scale, various studies have used different numbers of points per quadrat (from 5 to 99) within indeterminate quadrat distributions, and hence, a similar analysis at the point scale is valuable. Simulations of sampling synthetic habitat maps, based on normal distribution of class abundances, to recover the coral and sponge coverage estimates with less than 5% relative error, revealed that the optimal number of points per quadrat ranged between 13 points for a heterogeneous area with high coral coverage and over 600 points for a very homogeneous region with low coral coverage (Pante & Dustan, 2012). The recommended number of points per quadrat was 80 for transects of unknown community structure, but generally depended on the true underlying diversity and dispersion of the community configuration.

We contribute to the debate with a reef-scale analysis based on empirical community structure derived from our benthic habitat maps (Figure 7). We simulated quadrat sampling with various degrees of sampling effort, and found that higher densities of sampling points reduced the estimation error, similar to the results at the quadrat scale (Lechene et al., 2019). The number of point per quadrat to accurately recover the coverage of rare classes exceeded the recommended values in the literature. Even with 1000 sampling points per quadrat, rare classes at the transect level could not be detected within the 5% relative error limit (Figure 7b,d). In contrast to the coverage of individual classes, the Shannon diversity index was captured within 5% relative error with 160 sampling points per quadrat (Figure 7e). We demonstrate that the rarity and skewness of occurrence significantly impacts the error associated with a constant sampling effort. Ultimately, the precision of survey estimations is determined by the tension between distribution and density of sampling units, whether points or quadrats. When dense mapping at new sites is not possible and logistics constrain the number of quadrats, we recommend to sample over 160 points per quadrat during generation of baseline data, as well as to communicate the uncertainty generated from the sampling design (Hochberg & Gierach, 2021).

4.5 | Limitations and outlook

Although the results of our workflow are encouraging enough to recommend the different methods applied in this work, some limitations are noteworthy. For example, the nature of the images gathered by the push-broom hyperspectral camera (without geo-referencing) meant that the resulting transects are not ideal for photogrammetric techniques. This hinders the generation of 3D models and orthophoto mosaics, which provide a more comprehensive view of reef sites and facilitate temporal studies through geo-referencing. Recent studies are investigating novel techniques to overcome this limitation of hyperspectral push-broom sensors and have succeeded in producing rectified orthophoto mosaics of hyperspectral transects (Jurado et al., 2021; Moroni et al., 2012). Similarly, new robotic platforms, survey methods and data sources are being developed to improve benthic habitat mapping. We believe that the future direction of reef mapping is to develop end-to-end workflows that can handle mapping at the reefscape scale, with thematic and technical flexibility. Our workflow lays the groundwork for such end-to-end frameworks, where spectrally rich data flows are leveraged for mapping coral reefs.

It is also important to note, that the ML methods presented in this study are tailored for hyperspectral data and would not produce similarly high performance on RGB images. Another limitation of our ML methods is that they show low classification confidence in darkened areas (i.e. shadows). Thus, caution is advised when measuring small changes in temporal studies or reporting size of benthic organisms. A simple solution would be to group such problematic regions in a '*Shadow*' class of the labelspace. It must also be mentioned, that deep neural networks require high computational power to train and predict on hyperspectral data within reasonable time frames.

Regardless of the imaging techniques and ML methods used in reef surveys, we consider it important to evaluate the dense habitat maps (and not the classifiers) in terms of accuracy and completeness,

as the measure of progress. To disentangle the effects of changes in ML methods and data, we urge that the original images and annotations be made publicly available so that they can be re-evaluated independently. We have made the complete datasets (Chennu et al., 2020; Schürholz & Chennu, 2022a) and source code available to reproduce the results presented in this work (Schürholz & Chennu, 2022b).

Even though our workflow produces dense habitat maps with species-level resolution in several reefgroups, it is not a replacement for biodiversity assessments, where every species is recorded. New ML paradigms might be necessary to resolve the taxonomical hierarchy within reef organisms. Habitat descriptions that are derived from purely surficial surveys neglect cryptic biota, which can account for as much as half of the reef community and are, hence, critical to biodiversity assessments (Kornder et al., 2021). Further development of interdisciplinary efforts intersecting ML, computer vision, robotics, environmental DNA analysis and reef ecology will be required to automate survey outputs that directly enable biodiversity assessments and detailed reef inventories.

5 | CONCLUSIONS

Our proposed workflow showcases a way to generate dense habitat maps of coral reefs with flexible thematic detail. This thematic and spatial detail in the maps enables fine-grained analyses of coral reef functions and community dynamics by coral reef ecologists. We seek to contribute to unifying the perspective of ecologists, environmental managers, remote sensing and ML communities involved in the study of coral reefs. Particularly for ecologists and managers, our approach provides a consistent habitat description with adaptive thematic detail. Between remote sensing and ML experts, it offers a perspective on bridging the 'measurement gap' between ML classifiers and the ultimate data products, that is, habitat maps. The consistency achieved by our mapping workflow, and the patched method in particular, is related to the richness in spectral detail and the spatial acuity of our proximal sensing vantage point of underwater surveys (Chennu et al., 2017; Rashid & Chennu, 2020). When certain limitations are overcome and with improvements in cost and performance of underwater spectral surveying technology, it will become feasible to integrate it as a standard in-situ reef monitoring technique. The widespread use of underwater spectral surveying and automated benthic habitat mapping promises to provide the best validation data for aerial Earth observation efforts to map coral reefs globally. The integration of thematic detail into global habitat mapping promises to enable novel analyses of pattern and scale in coral reef ecology.

AUTHOR CONTRIBUTIONS

Daniel Schürholz—Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing—Original Draft, Writing—Review & Editing, Visualization, Project administration. Arjun Chennu—Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing—Original Draft, Writing—Review & Editing, Visualization, Supervision, Project administration, Funding acquisition.

ACKNOWLEDGEMENTS

We would like to thank Carsten John and Oliver Artmann at the Max Planck Institute for Marine Microbiology for their IT support. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement number 813360 '4D-REEF'. Open Access funding enabled and organized by Projekt DEAL.

CONFLICTS OF INTEREST

All authors declare that they have no conflict of interest with the work presented in this publication.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/2041-210X.14029.

DATA AVAILABILITY STATEMENT

All data needed to evaluate the conclusions in this paper are present in the manuscript, the Supplementary Materials and also in the following public data repositories: (i) the raw hyperspectral transects and annotations: https://doi.org/10.1594/PANGAEA.911300 (Chennu et al., 2020) (ii) the produced habitat maps for 31 transects in all combination of signal type, ML model and labelspace, as well as the data necessary to reproduce the figures presented in this work: https://doi.org/10.1594/PANGAEA.946315 (Schürholz & Chennu, 2022a) (iii) the source code as a reproducible project: (archive) https://doi.org/10.5281/zenodo.7185108 (Schürholz & Chennu, 2022b), and (live repo) https://bitbucket.org/hyperdream/digitizing-the-coral-reef/src/master/.

ORCID

Daniel Schürholz https://orcid.org/0000-0003-0213-9324
Arjun Chennu https://orcid.org/0000-0002-0389-5589

REFERENCES

Alonso, I., Yuval, M., Eyal, G., Treibitz, T., & Murillo, A. C. (2019). CoralSeg: Learning coral segmentation from sparse annotations. *Journal of Field Robotics*, *36*(8), 1456–1477. https://doi.org/10.1002/rob.21915

Althaus, F., Hill, N., Ferrari, R., Edwards, L., Przeslawski, R., Schönberg, C. H. L., Stuart-Smith, R., Barrett, N., Edgar, G., Colquhoun, J., Tran, M., Jordan, A., Rees, T., & Gowlett-Holmes, K. (2015). A standardised vocabulary for identifying benthic biota and substrata from underwater imagery: The CATAMI classification scheme. *PLoS ONE*, *10*(10), e0141039. https://doi.org/10.1371/journal.pone.0141039

Armstrong, R. A., Pizarro, O., & Roman, C. (2019). Underwater robotic Technology for Imaging Mesophotic Coral Ecosystems. In Y. Loya, K. A. Puglise, & T. C. L. Bridge (Eds.), *Mesophotic coral ecosystems* (pp. 973–988). Springer International Publishing. https://doi.org/10.1007/978-3-319-92735-0_51

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. *PLoS ONE*, *10*(7), e0130312. https://doi.org/10.1371/journal.pone.0130312

Brito-Millán, M., Vermeij, M. J. A., Alcantar, E. A., & Sandin, S. A. (2019). Coral reef assessments based on cover alone mask active dynamics of coral communities. *Marine Ecology Progress Series*, *630*, 55–68. https://doi.org/10.3354/meps13128

Burke, L., & Spalding, J. M., & contributing authors: M. Spalding. (2004). Reefs at Risk in the Caribbean. https://www.wri.org/research/reefs-risk-caribbean

Casella, E., Collin, A., Harris, D., Ferse, S., Bejarano, S., Parravicini, V., Hench, J. L., & Rovere, A. (2017). Mapping coral reefs using consumer-grade drones and structure from motion photogrammetry techniques. *Coral Reefs*, *36*(1), 269–275. https://doi.org/10.1007/s00338-016-1522-0

Cesar, H. S. J., Burke, L., & Pet-Soede, L. (2003). *The economics of worldwide coral reef degradation*. Cesar Environmental Economics Consulting (CEEC). https://books.google.de/books?id=WicVAQAAIAAJ

Chennu, A., Färber, P., De'ath, G., de Beer, D., & Fabricius, K. E. (2017). A diver-operated hyperspectral imaging and topographic surveying system for automated mapping of benthic habitats. *Scientific Reports*, *7*(1), 1–12. https://doi.org/10.1038/s41598-017-07337-y

Chennu, A., Rashid, A. R., den Haan, J., & de Beer, D. (2020). *Taxonomically annotated underwater hyperspectral and color images of coral reef transects from Curaçao [data set]*. PANGAEA. https://doi.org/10.1594/PANGAEA.911300

de Bakker, D. M., van Duyl, F. C., Bak, R. P. M., Nugues, M. M., Nieuwland, G., & Meesters, E. H. (2017). 40 years of benthic community change on the Caribbean reefs of Curaçao and Bonaire: The rise of slimy cyanobacterial mats. *Coral Reefs*, *36*(2), 355–367. https://doi.org/10.1007/s00338-016-1534-9

Dumas, P., Bertaud, A., Peignon, C., Leopold, M., & Dominique, P. (2009). A "quick and clean"photographic method for the description of coral reef habitats. *Journal of Experimental Biology and Ecology*, *368*, 161–168. https://doi.org/10.1016/j.jembe.2008.10.002

Foo, S. A., & Asner, G. P. (2019). Scaling up coral reef restoration using remote sensing technology. *Frontiers in Marine Science*, *6*, 79. https://doi.org/10.3389/fmars.2019.00079

Foody, G. M. (2004). Thematic map comparison. *Photogrammetric Engineering & Remote Sensing*, *70*(5), 627–633. https://doi.org/10.14358/PERS.70.5.627

González-Rivero, M., Beijbom, O., Rodriguez-Ramirez, A., Bryant, D. E. P., Ganase, A., Gonzalez-Marrero, Y., Herrera-Reveles, A., Kennedy, E. V., Kim, C. J. S., Lopez-Marcano, S., Markey, K., Neal, B. P., Osborne, K., Reyes-Nivia, C., Sampayo, E. M., Stolberg, K., Taylor, A., Vercelloni, J., Wyatt, M., & Hoegh-Guldberg, O. (2020). Monitoring of coral reefs using artificial intelligence: A feasible and cost-effective approach. *Remote Sensing*, *12*(3), 489. https://doi.org/10.3390/rs12030489

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., Regan, T. J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T. G., Rhodes, J. R., Maggini, R., Setterfield, S. A., Elith, J., Schwartz, M. W., Wintle, B. A., Broennimann, O., Austin, M., … Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, *16*(12), 1424–1435. https://doi.org/10.1111/ele.12189

Hedley, J. D., Roelfsema, C. M., Chollett, I., Harborne, A. R., Heron, S. F., Weeks, S., Skirving, W. J., Strong, A. E., Eakin, C. M., Christensen, T. R. L., Ticzon, V., Bejarano, S., & Mumby, P. J. (2016). Remote sensing of coral reefs for monitoring and management: A review. *Remote Sensing*, *8*(2), 118. https://doi.org/10.3390/rs8020118

Heron, S. F., Johnston, L., Liu, G., Geiger, E. F., Maynard, J. A., De La Cour, J. L., Johnson, S., Okano, R., Benavente, D., Burgess, T. F. R., Iguel, J., Perez, D. I., Skirving, W. J., Strong, A. E., Tirak, K., & Eakin, C. M. (2016). Validation of reef-scale thermal stress satellite products for coral bleaching monitoring. *Remote Sensing*, *8*(1), Article 1. https://doi.org/10.3390/rs8010059

Hochberg, E. J., Atkinson, M. J., & Andréfouët, S. (2003). Spectral reflectance of coral reef bottom-types worldwide and implications for coral reef remote sensing. *Remote Sensing of Environment*, *85*(2), 159–173. https://doi.org/10.1016/S0034-4257(02)00201-8

Hochberg, E. J., & Gierach, M. M. (2021). Missing the reef for the corals: Unexpected trends between coral reef condition and the environment at the ecosystem scale. *Frontiers in Marine Science*, *8*, 1191. https://doi.org/10.3389/fmars.2021.727038

Hoegh-Guldberg, O., Poloczanska, E. S., Skirving, W., & Dove, S. (2017). Coral reef ecosystems under climate change and ocean acidification. *Frontiers in Marine Science*, *4*, 158. https://doi.org/10.3389/fmars.2017.00158

Hughes, D. J., Alderdice, R., Cooney, C., Kühl, M., Pernice, M., Voolstra, C. R., & Suggett, D. J. (2020). Coral reef survival under accelerating ocean deoxygenation. *Nature Climate Change*, *10*(4), 296–307. https://doi.org/10.1038/s41558-020-0737-9

Jurado, J. M., Pádua, L., Hruška, J., Feito, F. R., & Sousa, J. J. (2021). An efficient method for generating UAV-based hyperspectral mosaics using push-broom sensors. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 6515–6531. https://doi.org/10.1109/JSTARS.2021.3088945

Kennedy, E. V., Roelfsema, C. M., Lyons, M. B., Kovacs, E. M., Borrego-Acevedo, R., Roe, M., Phinn, S. R., Larsen, K., Murray, N. J., Yuwono, D., Wolff, J., & Tudman, P. (2021). Reef cover, a coral reef classification for global habitat mapping from remote sensing. *Scientific Data*, *8*(1), Article 1. https://doi.org/10.1038/s41597-021-00958-z

Kornder, N. A., Cappelletto, J., Mueller, B., Zalm, M. J. L., Martinez, S. J., Vermeij, M. J. A., Huisman, J., & de Goeij, J. M. (2021). Implications of 2D versus 3D surveys to measure the abundance and composition of benthic coral reef communities. *Coral Reefs*, *40*(4), 1137–1153. https://doi.org/10.1007/s00338-021-02118-6

Krähenbühl, P., & Koltun, V. (2011). Efficient inference in fully connected CRFs with gaussian edge potentials. In *Proceedings of the 24th international conference on neural information processing systems (NeurIPS)* (pp. 109–117). Curran Associates Inc. https://proceedings.neurips.cc/paper/2011/hash/beda24c1e1b46055dff2c39c98fd6fc1-Abstract.html

Lechene, M. A. A., Haberstroh, A. J., Byrne, M., Figueira, W., & Ferrari, R. (2019). Optimising sampling strategies in coral reefs using large-area mosaics. *Remote Sensing*, *11*(24), Article 24. https://doi.org/10.3390/rs11242907

Lecours, V. (2017). On the use of maps and models in conservation and resource management (warning: Results may vary). *Frontiers in Marine Science*, *4*, 288. https://doi.org/10.3389/fmars.2017.00288

Lecours, V., Devillers, R., Schneider, D. C., Lucieer, V. L., Brown, C. J., & Edinger, E. N. (2015). Spatial scale and geographic context in benthic habitat mapping: Review and future directions. *Marine Ecology Progress Series*, *535*, 259–284. https://doi.org/10.3354/meps11378

Lyons, B., Roelfsema, M. M., Kennedy, C. V., Kovacs, E. M., Borrego-Acevedo, E., Markey, R., Roe, K., Yuwono, M. M., Harris, D. L., Phinn, D. R., Asner, S., Li, G. P., Knapp, J. E., Fabina, D. S., Larsen, N., Traganos, K. D., & Murray, N. J. (2020). Mapping the world's coral reefs using a global multiscale earth observation framework. *Remote Sensing in Ecology and Conservation*, *6*(4), 557–568. https://doi.org/10.1002/rse2.157

Moberg, F., & Folke, C. (1999). Ecological goods and services of coral reef ecosystems. *Ecological Economics*, *29*(2), 215–233. https://doi.org/10.1016/S0921-8009(99)00009-9

Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Tomkins-Tinch, C. H., Sochat, V., Forster, J., Lee, S., Twardziok, S. O., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., & Köster, J. (2021). Sustainable data analysis with Snakemake. *F1000Research*, *10*, 33. https://doi.org/10.12688/f1000research.29032.2

Moroni, M., Dacquino, C., & Cenedese, A. (2012). Mosaicing of hyperspectral images: The application of a spectrograph imaging device. *Sensors (Basel, Switzerland)*, *12*(8), 10228–10247. https://doi.org/10.3390/s120810228

Muldrow, M., Parsons, E. C. M., & Jonas, R. (2020). Shifting baseline syndrome among coral reef scientists. *Humanities and Social Sciences Communications*, *7*(1), 1–8. https://doi.org/10.1057/s41599-020-0526-0

Muller-Karger, F. E., Hestir, E., Ade, C., Turpie, K., Roberts, D. A., Siegel, D., Miller, R. J., Humm, D., Izenberg, N., Keller, M., Morgan, F., Frouin, R., Dekker, A. G., Gardner, R., Goodman, J., Schaeffer, B., Franz, B. A., Pahlevan, N., Mannino, A. G., … Jetz, W. (2018). Satellite sensor requirements for monitoring essential biodiversity variables of coastal ecosystems. *Ecological Applications*, *28*(3), 749–760. https://doi.org/10.1002/eap.1682

Nowosad, J., & Stepinski, T. F. (2019). Information theory as a consistent framework for quantification and classification of landscape patterns. *Landscape Ecology*, *34*(9), 2091–2101. https://doi.org/10.1007/s10980-019-00830-x

Pante, E., & Dustan, P. (2012). Getting to the point: Accuracy of point count in monitoring ecosystem change. *Journal of Marine Biology*, *2012*, e802875. https://doi.org/10.1155/2012/802875

Paoletti, M. E., Haut, J. M., Plaza, J., & Plaza, A. (2019). Deep learning classifiers for hyperspectral imaging: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *158*, 279–317. https://doi.org/10.1016/j.isprsjprs.2019.09.006

Pavoni, G., Corsini, M., Callieri, M., Fiameni, G., Edwards, C., & Cignoni, P. (2020). On improving the training of models for the semantic segmentation of benthic communities from orthographic imagery. *Remote Sensing*, *12*(18), Article 18. https://doi.org/10.3390/rs12183106

Perkins, N. R., Foster, S. D., Hill, N. A., & Barrett, N. S. (2016). Image subsampling and point scoring approaches for large-scale marine benthic monitoring programs. *Estuarine, Coastal and Shelf Science*, *176*, 36–46. https://doi.org/10.1016/j.ecss.2016.04.005

Phinn, S. R., Roelfsema, C. M., & Mumby, P. J. (2012). Multi-scale, object-based image analysis for mapping geomorphic and ecological zones on coral reefs. *International Journal of Remote Sensing*, *33*(12), 3768–3797. https://doi.org/10.1080/01431161.2011.633122

Rashid, A. R., & Chennu, A. (2020). A trillion coral reef colors: Deeply annotated underwater hyperspectral images for automated classification and habitat mapping. *Data*, *5*(1), Article 1. https://doi.org/10.3390/data5010019

Reverter, M., Helber, S. B., Rohde, S., de Goeij, J. M., & Schupp, P. J. (2022). Coral reef benthic community changes in the Anthropocene: Biogeographic heterogeneity, overlooked configurations, and methodology. *Global Change Biology*, *28*(6), 1956–1971. https://doi.org/10.1111/gcb.16034

Riitters, K. (2019). Pattern metrics for a transdisciplinary landscape ecology. *Landscape Ecology*, *34*(9), 2057–2063. https://doi.org/10.1007/s10980-018-0755-4

Roelfsema, C. M., Kovacs, E. M., Ortiz, J. C., Callaghan, D. P., Hock, K., Mongin, M., Johansen, K., Mumby, P. J., Wettle, M., Ronan, M., Lundgren, P., Kennedy, E. V., & Phinn, S. R. (2020). Habitat maps to enhance monitoring and management of the great barrier reef. *Coral Reefs*, *39*(4), 1039–1054. https://doi.org/10.1007/s00338-020-01929-3

Roelfsema, C. M., Lyons, M., Murray, N., Kovacs, E. M., Kennedy, E., Markey, K., Borrego-Acevedo, R., Ordoñez Alvarez, A., Say, C., Tudman, P., Roe, M., Wolff, J., Traganos, D., Asner, G. P., Bambic, B., Free, B., Fox, H. E., Lieb, Z., & Phinn, S. R. (2021). Workflow for the generation of expert-derived training and validation data: A view to global scale habitat mapping. *Frontiers in Marine Science*, *8*, 228. https://doi.org/10.3389/fmars.2021.643381

Roelfsema, C. M., & Phinn, S. R. (2013). Validation. In J. A. Goodman, S. J. Purkis, & S. R. Phinn (Eds.), *Coral reef remote sensing: A guide for mapping, monitoring and management* (pp. 375–401). Springer Netherlands. https://doi.org/10.1007/978-90-481-9292-2_14

Schürholz, D., & Chennu, A. (2022a). *Dense and taxonomically detailed habitat maps of coral reef benthos machine-generated from underwater hyperspectral transects in Curaçao* [Data set]. PANGAEA. https://doi.org/10.1594/PANGAEA.946315

Schürholz, D., & Chennu, A. (2022b). Digitizing the coral reef: A complete workflow for dense taxonomic mapping of benthic habitats with machine learning of underwater hyperspectral images (1.0.0). *Zenodo*. https://doi.org/10.5281/zenodo.7185108

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. J., Jones, E., Kern, R., Larson, E., … SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Zhong, Z., Li, J., Luo, Z., & Chapman, M. (2018). Spectral–spatial residual network for hyperspectral image classification: A 3-D deep learning framework. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(2), 847–858. https://doi.org/10.1109/TGRS.2017.2755542

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Schürholz, D., & Chennu, A. (2023). Digitizing the coral reef: Machine learning of underwater spectral images enables dense taxonomic mapping of benthic habitats. *Methods in Ecology and Evolution*, *14*, 596–613. https://doi.org/10.1111/2041-210X.14029