Ecological Informatics 95 (2026) 103784

Contents lists available at ScienceDirect

Ecological Informatics

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

Assessing joint effects of sampling design and annotation quality on benthic cover estimates through Monte Carlo simulations

Joris Guerin [a,b,*], Guilherme Longo [c,d], Regina Nobre [e], Célia Blondin [a], Laure Berti-Equille [a], Daniele Ventura [f]

[a] Espace-Dev, IRD, Université de Montpellier, Montpellier, France
[b] Department of Informatics and Applied Mathematics, Universidade Federal do Rio Grande do Norte, Natal, RN, Brazil
[c] Department of Oceanography and Limnology, Universidade Federal do Rio Grande do Norte, Natal, RN, Brazil
[d] Department of Biology, University of Massachusetts Boston, Boston, MA, USA
[e] Department or Ecology, Universidade Federal do Rio Grande do Norte, Natal, RN, Brazil
[f] Department of Environmental Biology, Sapienza Università di Roma, Rome, Italy

ARTICLE INFO

Dataset link: https://github.com/jorisguerin/benthic-photoquadrat-study

Keywords:
Ecological sampling
Machine learning annotation
Uncertainty quantification
Monte Carlo simulation
Benthic monitoring

ABSTRACT

Ecological monitoring increasingly relies on image-based data collection coupled with machine learning annotation. Yet, the combined impact of sampling effort and annotation quality on indicator precision remains poorly understood, limiting evidence-based resource allocation in biodiversity monitoring programs. We investigate this question for benthic cover estimation using photo-quadrat surveys, which involve errors from spatial quadrat sampling and image annotation. While prior studies have examined these error sources independently, their joint effects remain unquantified. We developed a photo-quadrat sampling and annotation simulator and applied it within a Monte Carlo framework to quantify how different combinations of design choices affect cover estimates on known reference maps. We apply our approach to a reference map from shallow Mediterranean benthic habitat (Tyrrhenian Sea), comparing four quadrat placement strategies with varying annotation accuracy across 10,000 replicates per parameter combination. Our results challenge conventional benthic monitoring assumptions: (i) random quadrat placement significantly outperformed structured transect-based strategies, (ii) improved annotation performance did not systematically improve cover precision, and (iii) extensive sampling with imperfect annotation consistently outperformed perfect annotation of fewer samples. Our complete simulation framework, analysis code and data are made publicly available, along with a free web application for broader research and educational use.

1. Introduction

Benthic habitats form the foundation of marine ecosystems, supporting biodiversity, providing essential ecosystem services, and serving as critical indicators of environmental health (Kritzer et al., 2016; Galparsoro et al., 2014). Effective monitoring of benthic cover composition is therefore essential for marine conservation efforts, enabling researchers and managers to track ecosystem changes in space (Aued et al., 2018) and time (Biscaia Zamoner et al., 2021), assess anthropogenic impacts (Perry and Alvarez-Filip, 2019; Mello et al., 2025), define conservation priorities (Proudfoot et al., 2020), and evaluate restoration success (Jayachandran et al., 2022).

Traditional benthic survey methods have relied on direct sampling techniques (grab sampling, sediment cores) or direct visual assessments from divers to characterize seafloor communities (Eleftheriou, 2013). Recently, benthic surveys based on digital images have received increasing attention. They offer a non-invasive and cost-effective approach, enabling large-scale monitoring while preserving the integrity of these critical ecosystems. In particular, the *photo-quadrat method* is widely used for routine benthic cover monitoring programs as it offers a good balance between detailed species detection capabilities and field work efficiency (Jokiel et al., 2015).

The photo-quadrat method typically follows a standardized protocol (Fig. 1), involving three sequential steps: *(1) field sampling of photo-quadrats*, i.e., standardized top-down photos of fixed-size frames positioned on the seafloor using random or systematic sampling designs (e.g., along transects), *(2) point sampling within each image*, either randomly or on a regular grid, and *(3) annotation of sampled points* by taxonomic experts or automated algorithms (Blondin et al., 2024; Zotou

* Correspondence to: UMR Espace-Dev, Maison de la Télédétection, 500 Rue Jean François Breton, 34090 Montpellier, France.
E-mail address: joris.guerin@ird.fr (J. Guerin).

https://doi.org/10.1016/j.ecoinf.2026.103784
Received 23 September 2025; Received in revised form 16 April 2026; Accepted 17 April 2026
Available online 20 April 2026

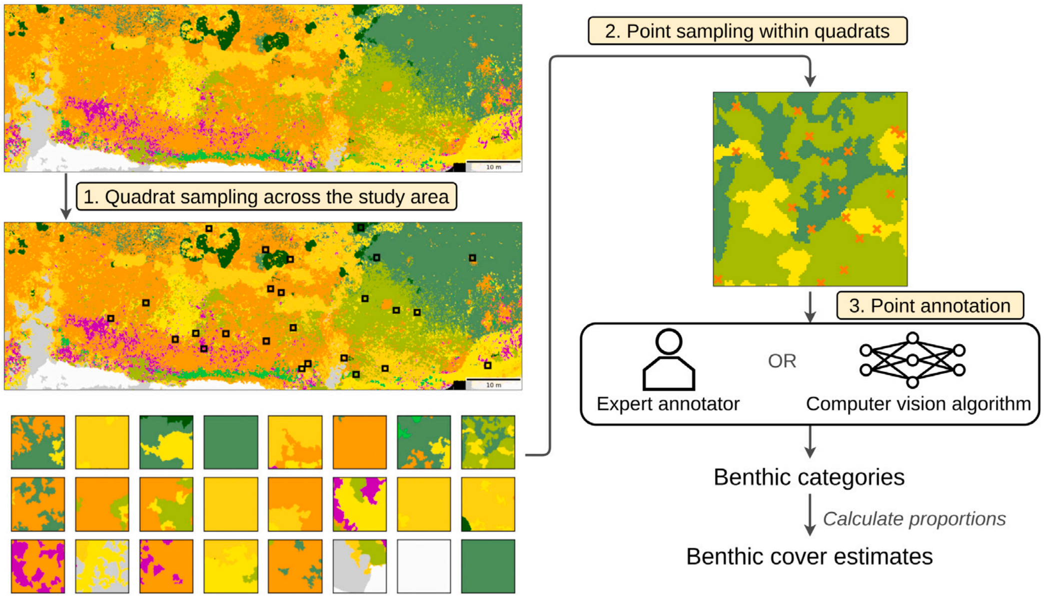


Fig. 1. The photo-quadrat methodology for benthic cover assessment involves three sequential steps: 1. field sampling of quadrat photos of the seafloor, 2. point sampling within quadrats, and 3. annotation of sampled points by experts or automated algorithms. In the first panel, colors represent different benthic elements of the study area. In the second panel, the black squares represent randomly sampled photo-quadrats, which are zoomed-in below. In the right panel, the orange marks represent random points selected for annotation.

et al., 2025; Jackett et al., 2023). The proportions of annotated points corresponding to each benthic class provide the final cover estimates.

Each of these three methodological steps introduces approximation errors that collectively influence the final benthic cover estimates. Field sampling errors arise from the limited representativeness of discrete quadrat locations. Point sampling errors arise from the finite number of points used to estimate spatially continuous cover distributions within each image. Finally, annotation errors arise from misclassifications of certain benthic organisms or substrates. This work aims to quantify how these different error sources and their individual magnitudes impact final benthic cover estimates, which is essential for optimizing survey protocols and making informed resource allocation in monitoring programs.

Previous studies have attempted to quantify the uncertainty associated with sampling design choices (number of transects, number of quadrats per transect, number of points per quadrat) to determine minimal sampling effort required for reliable benthic cover assessment. Molloy et al. (2013) conducted leave-some-out analyses on Philippine photo-quadrat data to evaluate how reduced sampling intensity affects the reliability of benthic cover estimates. Similarly, Montilla et al. (2020) used Venezuelan reef data to assess photo-quadrat sampling performance by measuring similarity between samples. Both approaches conclude that relatively modest sampling effort can achieve acceptable precision, but they share a critical methodological limitation: they evaluate sampling design performance by measuring internal consistency within their own datasets rather than against known ground truth. This approach can assess precision (how consistent are the results) but cannot evaluate accuracy (how close are the results to reality), potentially leading to biased recommendations if the original sampling failed to capture true benthic cover proportions and habitat variability. Lechene et al. (2019) addressed this limitation by using Monte Carlo simulations on high-resolution benthic habitat maps to evaluate how quadrat size and number affect sampling accuracy.

Complementarily, recent automated annotation approaches (Lowe et al., 2025; Jackett et al., 2023; Blondin et al., 2024; Zotou et al., 2025) focus on optimizing and testing annotation performance metrics in isolation. On another note, Beijbom et al. (2015) investigated the uncertainty associated with point annotation by quantifying intra- and inter-annotator variability to establish a more realistic baseline for automatic annotation methods. They showed that expert agreement varied substantially across habitat types, with high consistency for coral genera identification but lower agreement for algal functional groups. They then evaluated semi- and fully-automated annotation methods and concluded that semi-automated approaches could achieve human-level precision for cover estimation, while fully-automated methods provided rapid, unbiased estimates with increased variance. However, this approach treats annotation accuracy as an isolated component and fails to evaluate how these errors propagate through the complete monitoring pipeline. Critically, it cannot address system-level questions such as: Given a fixed sampling effort, how much annotation accuracy is needed to achieve target benthic cover precision? Or can increased sampling effort compensate for reduced annotation accuracy to maintain monitoring objectives?

To overcome these limitations, this paper presents a novel framework to comprehensively evaluate error sources in benthic image sampling and annotation protocols. We implemented a Monte Carlo simulation tool that takes a detailed habitat map as input and systematically varies key design parameters (number of quadrats, quadrat size, spatial placement patterns, points per quadrat, and annotation error rates) to quantify how different uncertainty sources interact throughout the complete monitoring pipeline. We demonstrate this framework using a real-world detailed benthic habitat map annotated with 10 ecological classes at 1 cm per pixel resolution (Ventura et al., 2023). Our approach reveals how errors propagate, accumulate, or compensate for one another in the overall estimation process, providing insights that cannot be obtained from studying individual error sources in isolation. While demonstrated here for a Mediterranean benthic system, the simulation framework's core principles (using high-resolution habitat maps to evaluate sampling and annotation trade-offs against known ground truth) could be adapted to other ecological monitoring contexts where comparable spatial data are available.

Our method addresses fundamental questions that previous studies could not answer: (1) How do additional monitoring efforts (more quadrats, different quadrat placements, more points per quadrat, improved annotation accuracy) translate into better benthic cover estimates, enabling ecologists to design monitoring programs according

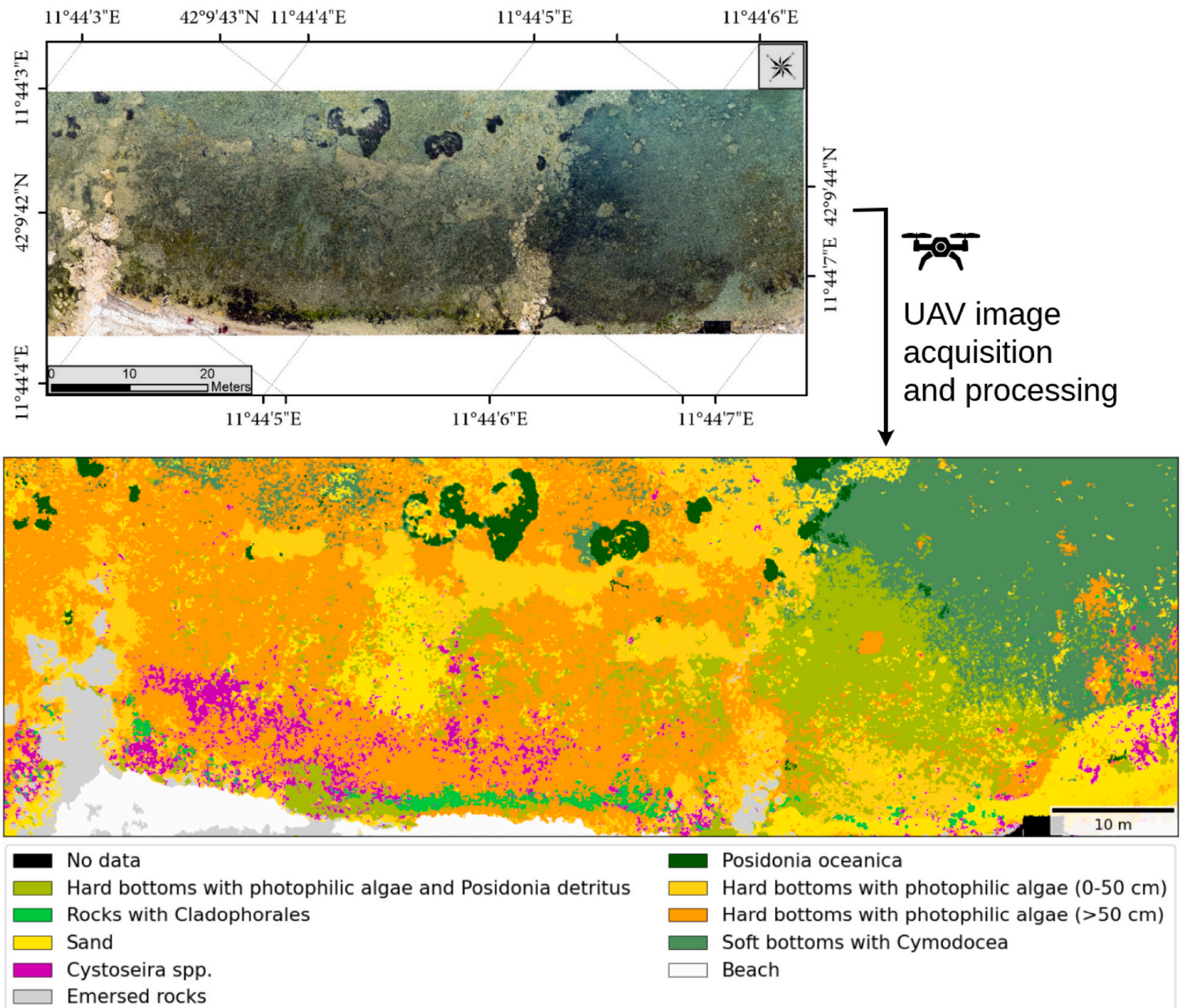


Fig. 2. Presentation of the study area located north of Civitavecchia harbor (Tyrrhenian Sea, Italy). Top: Orthophoto of the 31 m × 97 m study area. Bottom: Map of benthic cover types classified at 1cm pixel resolution.

to desired accuracy targets? (2) How much annotation accuracy is required to safely delegate annotation tasks to automated methods without sacrificing overall monitoring quality? (3) When might investing in improved sampling design be more effective than improving annotation accuracy, and vice versa? By providing a rigorous framework for joint evaluation of sampling and annotation protocols, this work enables evidence-based decisions about where to allocate limited monitoring resources for maximum impact. Our experiments reveal that sampling effort largely dominates annotation quality in determining monitoring accuracy for our study ecosystem, challenging the assumption that perfect annotation is always required. This finding suggests new priorities for resource allocation in similar monitoring contexts.

2. Materials and methods

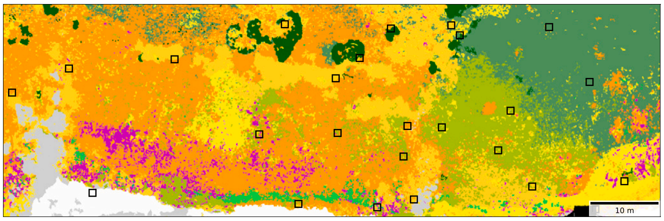
2.1. Reference data

To conduct our benthic cover sampling simulation analysis, we required a high-resolution reference map on which to simulate the photo-quadrat sampling methodology. We used the centimeter-scale benthic map produced by Ventura et al. (2023) (Fig. 2). The dataset was developed following FAIR (Findable, Accessible, Interoperable, Reusable) data principles and CARE (Collective Benefit, Authority to Control, Responsibility, Ethics) principles for collaborative data governance with local stakeholders (see Supplementary Materials S1 for metadata compliance documentation). Complete methodological details, including georeferencing accuracy, image processing, and classification validation, are provided in the original study (Ventura et al., 2023). The complete map is available upon request (see Data Availability Statement) and a lightweight version, sufficient to reproduce all analyses in this study, is included in the code repository.[1]

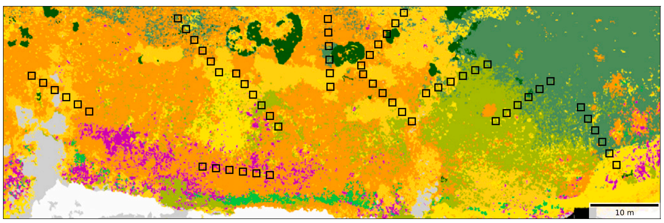
2.1.1. Study area

The research site (11.734709° E, 42.161924° N) was located north of Civitavecchia harbor (Tyrrhenian Sea, Italy), covering 0.3 hectares within a marine Site of Community Importance (SCI) called "Fondali tra Punta S. Agostino e Punta Mattonara". The seabed, located in a sheltered and shallow bay, was characterized by sandy flats, rocky formations, and vegetated habitats. Hard substrates at depths between 0.1 and 3 m supported diverse photophilous communities including seagrass patches (*Posidonia oceanica* and *Cymodocea nodosa*), brown algae (*Cystoseira* spp.), and green algae (Cladophorales). Sandy depressions between rocks frequently contained organic detritus from dead

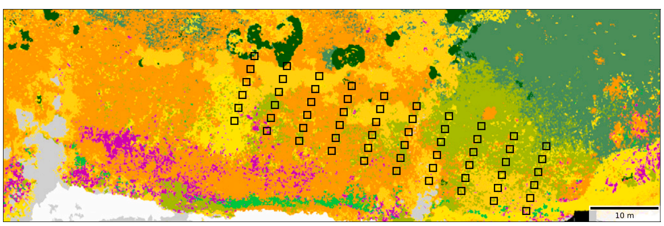
[1] https://github.com/jorisguerin/benthic-photoquadrat-study



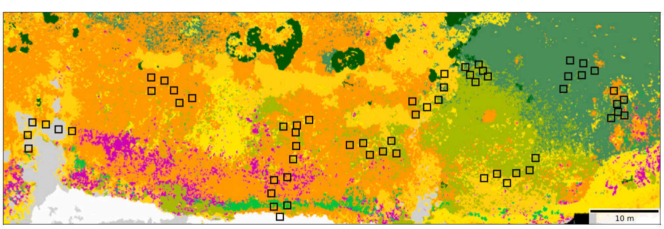
(a) Random sampling (25 quadrats, 1 m side length)



(b) Random transects (10 transects, 6 quadrats each, 2 m spacing)



(c) Parallel transects (same configuration as above)



(d) Non-directional transects (random direction between successive quadrats)

Fig. 3. Benthic cover sampling strategies. Examples of four sampling approaches generated by our simulation framework, each showing quadrat placement patterns for comparative analysis.

Posidonia oceanica leaves. This complex habitat mosaic provides critical ecosystem services including nursery grounds for endemic species and coastal protection, making accurate mapping essential for effective coastal management and monitoring (Ventura et al., 2023).

2.1.2. High-resolution map generation

High-resolution aerial images were acquired in May 2022 with a Hasselblad L1D-20c camera (20 megapixels, 1-inch CMOS sensor), mounted on a DJI Mavic 2 Pro quadcopter operating at 40 m altitude, achieving a ground sample distance of approximately 1.7 cm/px. A circular polarizer filter was used to reduce sun-glint effects.

Image geolocations were corrected using Post Processing Kinematics with RINEX files from the nearby Civitavecchia Continuously Operating Reference Station. This procedure yielded planimetric and altimetric precision of 5.8 cm and 11 cm, validated at seven ground control points.

Photogrammetric processing was performed using Agisoft Metashape v1.8.1 following standard structure-from-motion workflows: image alignment, sparse point cloud generation, dense point cloud construction using multi-view stereopsis, and orthorectified photomosaic creation through Digital Surface Model interpolation.

Benthic habitat classification was conducted in ArcMap 10.6 using a Support Vector Machine classifier trained on manually selected samples (maximum 500 pixels per class). Post-classification processing included majority filtering, boundary cleaning, and removal of isolated pixel regions to improve classification accuracy and spatial consistency.

2.1.3. Reference map validity

While the benthic habitat map used in our simulations may contain classification errors inherent to remote sensing approaches, this does not affect the validity or generality of our analysis. By treating this map as ground truth for simulation purposes, we evaluate how sampling design and annotation errors propagate through the monitoring pipeline relative to a known reference state. Any potential misclassification in the original map becomes part of our synthetic reality, allowing us to assess photo-quadrat methodology performance under realistic conditions.

2.2. Sampling and annotation simulator

This section presents the components of our simulation framework to systematically evaluate the performance of photo-quadrat sampling strategies and annotation protocols. The simulator generates synthetic sampling campaigns by extracting quadrats from our reference map according to specified design parameters, then applies controlled annotation errors to quantify their effects on benthic cover estimates.

The complete codebase for reproducibility has been openly released.[2] As a side-contribution, we also developed an openly available web application for the sampling component.[3] It can be used as a support to teach sampling design principles to ecology students.

2.2.1. Quadrat placement simulation

Our simulator implements four quadrat placement strategies commonly found in the literature (Aued et al., 2018; Biscaia Zamoner et al., 2021), allowing us to explore different aspects of spatial sampling design:

1. **Random isolated quadrats** (Fig. 3(a)): Quadrats are placed randomly across the study area. Their number is controlled by a user-defined *sample size* parameter.
2. **Random transects** (Fig. 3(b)): Quadrats are arranged along randomly oriented transects, a common field protocol where divers collect photo-quadrats along predetermined linear paths. Tunable parameters include *number of transects*, *quadrats per transect*, and *inter-quadrat spacing*.
3. **Parallel transects** (Fig. 3(c)): Transects are forced to be parallel to each other, mimicking designs often used in large-scale monitoring programs.
4. **Non-directional transects** (Fig. 3(d)): Successive quadrats are connected by random-direction movement of fixed distance. This approach maintains the field simplicity of transect-based collection while removing linearity constraints, allowing evaluation of whether different transect trajectories lead to different cover estimation errors.

All methods share a common *quadrat size* parameter and enforce non-overlapping constraints between quadrats to reproduce realistic field conditions. Fig. 3 shows examples of these sampling configurations, with corresponding benthic cover estimates presented in Table 1. Comparisons of other sampling configurations can be easily performed using our benthic simulator application at the user's convenience.

[2] https://github.com/jorisguerin/benthic-photoquadrat-study
[3] https://benthic-sampling-simulator.streamlit.app/

Table 1
Example cover estimates for different sampling strategies. True cover represents ground-truth proportions across all pixels in the study area. Sampling estimates are calculated from all pixels within the sampled quadrats. Values correspond to the realizations from Fig. 3.

Benthic cover category	True cover	Random sampling	Random transects	Parallel transects	Non-directional transects
Hard bottoms w/ photophilic algae & Posidonia detritus	12.43%	19.69%	13.72%	31.33%	23.44%
Rocks w/ Cladophorales	1.12%	0.51%	0.06%	0.00%	1.78%
Sand	8.98%	7.35%	6.36%	10.54%	4.36%
Cystoseira spp.	3.70%	1.72%	2.25%	0.98%	0.82%
Emersed rocks	3.32%	1.41%	0.05%	0.12%	3.71%
Posidonia oceanica	2.11%	5.50%	1.40%	0.67%	0.04%
Hard bottoms w/ photophilic algae (0–50 cm)	15.05%	26.93%	25.08%	22.72%	11.18%
Hard bottoms w/ photophilic algae (>50 cm)	34.05%	21.94%	34.14%	32.68%	33.54%
Soft bottoms w/ *Cymodocea*	15.82%	10.56%	16.93%	0.96%	19.52%
Beach	3.43%	4.39%	0.00%	0.00%	1.61%
Total	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**

2.2.2. Intra-quadrat point sampling

Following quadrat placement simulation, we obtain a set of images representing the collected quadrats. To assess the true representativeness of the field sampling campaign, we can use all pixels within each quadrat to estimate benthic cover proportions and compare the results against the true proportions across the entire study area. However, practitioners typically annotate only a subset of pixels per quadrat (∼25), introducing an additional error source.

Our simulator replicates this process by randomly sampling pixels within each quadrat, controlled by a single user-defined parameter: *number of points*. Varying the number of annotated pixels allows to quantify how point sampling density interacts with spatial sampling design to affect estimation accuracy. The output of step 2, in Fig. 1, shows an example of intra-quadrat pixel samples.

2.2.3. Annotation error simulation

Once points are sampled from quadrat images, they must be annotated with benthic organisms labels. Traditionally performed by marine ecology experts, annotation is increasingly delegated to automated algorithms to reduce the annotation burden (Williams et al., 2019). Both approaches can produce misclassification errors that propagate through the benthic cover estimation process.

In our simulations, we can use true labels to compute benthic cover and assess the representativeness of sampling steps without annotation error. Yet, to quantify how annotation error affect final cover estimates, we implemented a label switching mechanism to generate label annotations with controlled error rates. Our annotation error simulation requires three parameters: (1) *target class*: The benthic class for which cover proportion is estimated. (2) *precision*: The proportion of points labeled as the target class that are actually the target class. (3) *recall*: The proportion of true target class points that are correctly labeled as target class.

Then, to simulate annotation errors at desired precision and recall levels, we compute false positive and false negative rates:

$$FNR = 1 - \text{recall}, \tag{1}$$

$$FPR = \frac{\text{recall} \times n_{\text{pos}} \times (1 - \text{precision})}{\text{precision} \times n_{\text{neg}}}, \tag{2}$$

where n_{pos} is the number of true target class pixels and n_{neg} is the number of true non-target class pixels.

Finally, for each true target class pixel, we switch its label to a different class with probability FNR. For each true non-target pixel, we switch its label to the target class with probability FPR. This process reproduces the behavior of an annotation model with specified precision and recall for cover assessment of the target class.

2.2.4. Simulator stochasticity

By fixing the parameters described above and running a simulation, we generate a single cover estimate representing a specific sampling design and annotation performance combination. By design, this estimate represents what would occur when implementing a complete monitoring protocol with similar specifications in our study area.

The sampling and annotation processes in our simulator inherently involve randomness, so multiple runs under identical parameters produce different cover estimation results. This variability is actually desirable, as it reflects the stochastic nature of real monitoring campaigns.

Our simulator enables us to generate multiple iterations of the same sampling and annotation design, allowing us to precisely quantify the precision and variability of a sampling configuration against known ground truth. This capability provides the foundation for systematic parameter evaluation and performance assessment.

2.3. Comparing configurations

To evaluate how different monitoring design choices affect benthic cover estimation accuracy, we systematically vary key parameters that practitioners must decide when implementing photo-quadrat surveys. These parameters directly correspond to resource allocation decisions: quadrat number and placement strategy affect field time and costs, point density determines annotation workload, and annotation method (expert vs automated) involves trade-offs between accuracy and scalability. By quantifying estimation errors across this parameter space, we provide guidance for optimizing monitoring program design under resource constraints.

2.3.1. Individual simulation error

Once we select a parameter set θ (quadrat size, placement methodology and associated parameters, number of points per quadrat, target category, annotation performance metrics) to configure our simulator S, we can execute a simulation run to generate a stochastic sampling scenario. This produces an estimate of the cover proportion for the target class over the study area, denoted as $\bar{\rho} = S(\theta)$. By comparing this cover estimate to the known true cover proportion of the target class (denoted as ρ_{true}), we obtain an error value:

$$\epsilon = \bar{\rho} - \rho_{\text{true}}, \tag{3}$$

where $\epsilon > 0$ indicates overestimation and $\epsilon < 0$, underestimation.

2.3.2. Error distribution

To characterize the error distribution for a given configuration, we employ Monte Carlo simulation. We fix θ and repeat the simulation process N times (N = 10,000 in our experiments). The N error values collected represent different realizations of the monitoring design defined by θ.

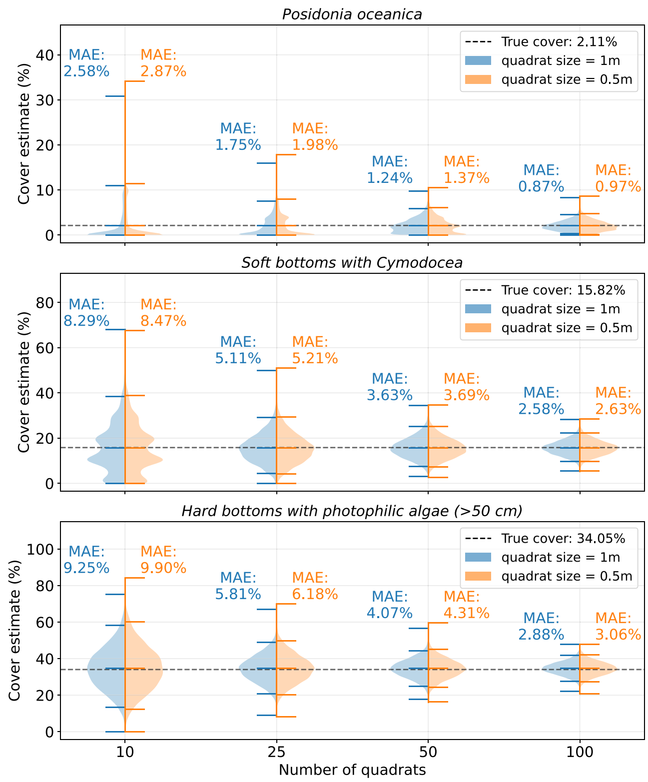


Fig. 4. Effect of quadrat number and size on cover estimates. Violin plots show distributions of benthic cover estimates from 10,000 Monte Carlo simulations for three benthic classes with different abundance levels. Each plot displays the mean estimate, 95% confidence interval, extreme values, and mean absolute error (MAE). Random quadrat placement used throughout.

The error distribution for each configuration θ is visualized using violin plots (see Fig. 4 as an example). Comparing violin plots between different monitoring configurations enables visual assessment of their relative performance.

2.3.3. Performance metrics

To quantitatively characterize the estimation performance of each configuration, we compute key metrics from the error distribution:

- **Mean Absolute Error (MAE)**: The average magnitude of estimation errors, computed as

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} |\epsilon_i|. \tag{4}$$

  MAE quantifies overall estimation precision regardless of error direction. A monitoring protocol with a higher MAE will be further away from the truth, on average.
- **Bias**: The mean error across simulations, computed as

$$Bias = \frac{1}{N}\sum_{i=1}^{N} \epsilon_i. \tag{5}$$

  Bias reveals systematic over- or underestimation tendencies in monitoring configurations. In violin plots, bias appears as the offset between the distribution center and the true cover value.

These two metrics together provide a comprehensive comparison of monitoring protocols: MAE captures precision while bias captures systematic accuracy. For practical applications, we prioritize MAE as it reflects actual estimation accuracy under resource constraints.

2.4. Assessing interactions between error sources

To quantify how sampling and annotation errors combine to produce total estimation error, we decompose the mean squared error (MSE) into components attributable to each source and their interaction. We use MSE for this decomposition rather than MAE, because it admits an exact additive decomposition into bias and variance terms, making the interaction between error sources mathematically tractable.

For a given monitoring configuration θ, we define three error quantities. The total error e_{tot} compares final cover estimates (after sampling and annotation) against ground truth over the whole area. The sampling error e_s isolates the contribution of spatial subsampling by assuming perfect annotation of sampled quadrats. The annotation error e_a isolates the contribution of imperfect annotation by comparing model-derived cover estimates against ground truth on the sampled quadrats only. As e_s and e_a are our only error sources, we have $e_{tot} = e_s + e_a$. Applying the bias–variance decomposition, we have:

$$\begin{aligned}\mathrm{MSE}(e_{tot}) &= \mathrm{Bias}(e_{tot})^2 + \mathrm{Var}(e_{tot})\\ &= \mathrm{Bias}(e_s + e_a)^2 + \mathrm{Var}(e_s + e_a).\end{aligned} \tag{6}$$

By the Bias and Variance additivity properties, we can write

$$\begin{aligned}\mathrm{MSE}(e_{tot}) &= (\mathrm{Bias}(e_s) + \mathrm{Bias}(e_a))^2\\ &\quad + \mathrm{Var}(e_s) + \mathrm{Var}(e_a) + 2\mathrm{Cov}(e_s, e_a).\end{aligned} \tag{7}$$

Expanding the quadratic term, and recomposing the individual MSE decompositions, we get

$$\begin{aligned}\mathrm{MSE}(e_{tot}) &= \mathrm{MSE}(e_s) + \mathrm{MSE}(e_a)\\ &\quad + 2\mathrm{Bias}(e_s)\mathrm{Bias}(e_a) + 2\mathrm{Cov}(e_s, e_a).\end{aligned} \tag{8}$$

This decomposition reveals how the two error sources interact. When both biases share the same sign, their contributions compound, increasing total error, e.g., an undersampled species that is also underpredicted by the annotation model (low recall). Conversely, when biases have opposite signs, errors partially cancel, and the total error may be smaller than either individual component, e.g., an undersampled species that is overpredicted (low precision). Similarly, positive covariance between e_s and e_a amplifies total error, while negative covariance has a compensating effect.

This decomposition framework enables quantitative assessment of which error source and mechanism dominates monitoring uncertainty under different resource allocation strategies.

2.5. Experimental design

To comprehensively evaluate error propagation across different scenarios, we focused our analysis on three representative classes with varying cover proportions: *Posidonia oceanica* (low abundance, ∼ 2%), *Soft bottoms w/ Cymodocea* (medium abundance, ∼ 16%), and *Hard bottoms w/ photophilic algae (> 50 cm)* (high abundance, ∼ 34%). This selection enables assessment of how estimation errors vary with target class rarity, a critical consideration for monitoring program design.

For each target category, we conducted five experiments, isolating specific design parameters while controlling others:

Experiment 1: Quadrat size and number effects. We varied quadrat size (0.5 m, 1 m) and number (10, 25, 50, 100) while fixing other parameters: random quadrat placement, all pixels annotated, and perfect annotation (no errors).

Experiment 2: Quadrat placement strategy effects. We compared four strategies (random, random transects, parallel transects, non-directional transects) using fixed parameters: 0.5 m quadrats, all pixels annotated, no annotation errors. Random placement used 50 quadrats; transect approaches used 10 transects with 6 quadrats per transect, 2 m spacing between quadrats and between parallel transects. We use more quadrats for transect designs to better reflect field effort

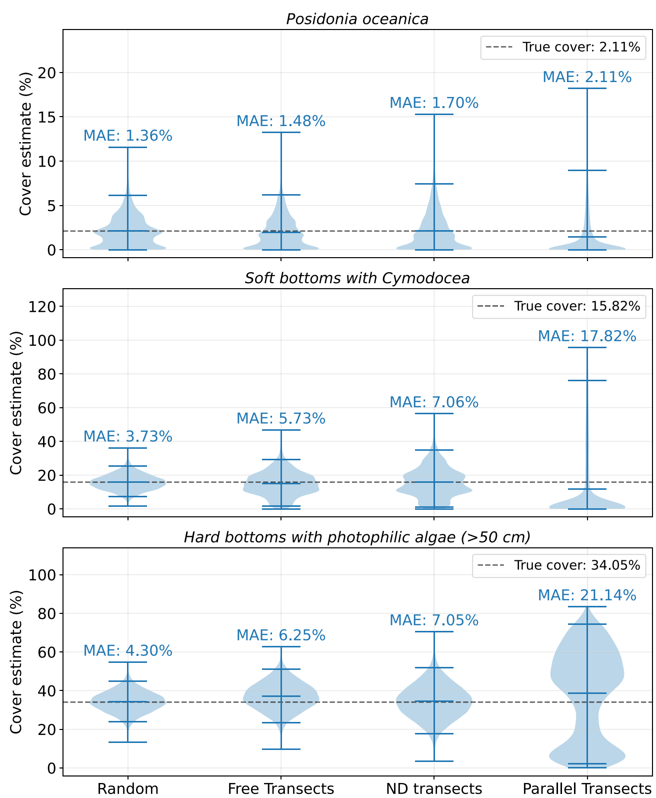


Fig. 5. Effect of quadrat placement strategy on cover estimates. Violin plots show distributions of benthic cover estimates from 10,000 Monte Carlo simulations for three benthic classes with different abundance levels. Each plot displays the mean estimate, 95% confidence interval, extreme values, and mean absolute error (MAE). Configurations: 50 random quadrats (0.5 m) vs. 60 transect quadrats (10 transects, 6 quadrats per transect, 2 m spacing).

allocation, as collecting spatially clustered quadrats along transects is more time-efficient than navigating to isolated random locations.

Experiment 3: Point sampling density effects. We varied annotation density (10, 25, 100, or all pixels per quadrat) while fixing: random placement, 50 quadrats of 0.5 m, no annotation errors.

Experiment 4: Annotation error effects. We systematically varied annotation performance using random placement (50 quadrats, 0.5 m, 10 points per quadrat). Precision and recall each varied from 0.5 to 1.0 in 0.05 increments, with all combinations evaluated.

Experiment 5: Comparing resource allocation strategies. To directly evaluate the trade-off between sampling effort and annotation quality that practitioners face under resource constraints, we compared two realistic monitoring scenarios: Scenario 1 employed extensive sampling (200 quadrats) with imperfect automated annotation (precision/recall = 0.75), representing extensive resource allocation to field campaigns with minimal efforts for in-lab annotation. Conversely, scenario 2 employed limited sampling (50 quadrats) with high-quality expert annotation (precision/recall = 0.99). Both scenarios used random quadrat placement, 0.5 m quadrats, and 10 points per quadrat. To quantify the magnitude of errors due to sampling and annotation quality, we applied the error decomposition framework (Section 2.4) to each scenario, to partition MSE into sampling, annotation, and interaction components.

3. Results

We systematically evaluated different aspects of photo-quadrat monitoring design through Monte Carlo simulations across five experiments examining quadrat number and size, placement strategies, point sampling density, annotation errors, and resource allocation trade-offs. Our analysis reveals distinct patterns in how different error sources affect cover estimation, with findings that challenge conventional assumptions about sampling optimization and annotation quality requirements.

3.1. Field sampling design effects

We first examined how field sampling decisions affect cover estimation accuracy by comparing quadrat size, number, and placement strategies while maintaining perfect annotation of all pixels.

3.1.1. Size and number of quadrats

Fig. 4 reveals contrasting effects of quadrat size and number on estimation accuracy. Surprisingly, doubling quadrat size (0.5 m vs 1 m) has minimal impact on benthic cover error despite quadrupling the sampled area. In contrast, increasing quadrat number consistently reduces estimation errors across all abundance classes. MAE decreases by over 6% as the number of quadrats increases from 10 to 100.

3.1.2. Quadrat placement strategy

Fig. 5 demonstrates that random quadrat placement achieved lower estimation errors (MAE) than all structured transect approaches. Random placement also exhibits minimal systematic bias across all abundance classes.

Both free transects (linear transects of random orientation and position) and non-directional transects are also unbiased estimators of the true cover area. However, free transects achieved lower MAE than their non-directional counterparts.

Parallel transects, despite being very common in marine ecology, performed poorly across all investigated classes. This strategy not only exhibited significantly higher mean absolute errors but also introduced systematic bias in cover estimates.

3.2. Point sampling density effects

Fig. 6 demonstrates that annotation density has minimal impact on benthic cover estimation accuracy. Increasing the number of annotated points per quadrat from 10 to 100, or even to all pixels, produces negligible improvements across all classes of varying abundance.

For *Posidonia oceanica*, MAE decreased only marginally from 1.38% (10 points) to 1.35% (all pixels). Similarly modest improvements were observed for *Soft bottoms w/ Cymodocea* (3.80% to 3.76%) and Hard bottoms w/ photophilic algae (4.41% to 4.29%).

3.3. Annotation errors effects

In Fig. 7, minimum MAE consistently occurs at intermediate annotation performance levels rather than perfect classification (precision/recall = 1.0), across all three abundance classes. This challenges the fundamental assumption that improving annotation quality always improves monitoring accuracy.

For *Posidonia oceanica*, estimation accuracy remains relatively stable across annotation performance levels, with subtle but systematic patterns. Decreasing recall tends to slightly increase estimation errors, while decreasing precision actually improves accuracy (lower MAE).

For *Soft bottoms w/ Cymodocea* and *Hard bottoms w/ photophilic algae*, the opposite pattern emerges: minimum estimation errors occur at low recall (~ 0.5) combined with moderate precision (~ 0.75).

Examining bias patterns across all three classes reveals diagonal bands of low systematic error, including the theoretically optimal configuration (precision/recall = 1.0). This pattern confirms that random spatial sampling introduces no systematic bias, consistent with sampling theory. However, the existence of multiple low-bias configurations at suboptimal annotation performance levels demonstrates that perfect annotation is not necessary for unbiased estimation.

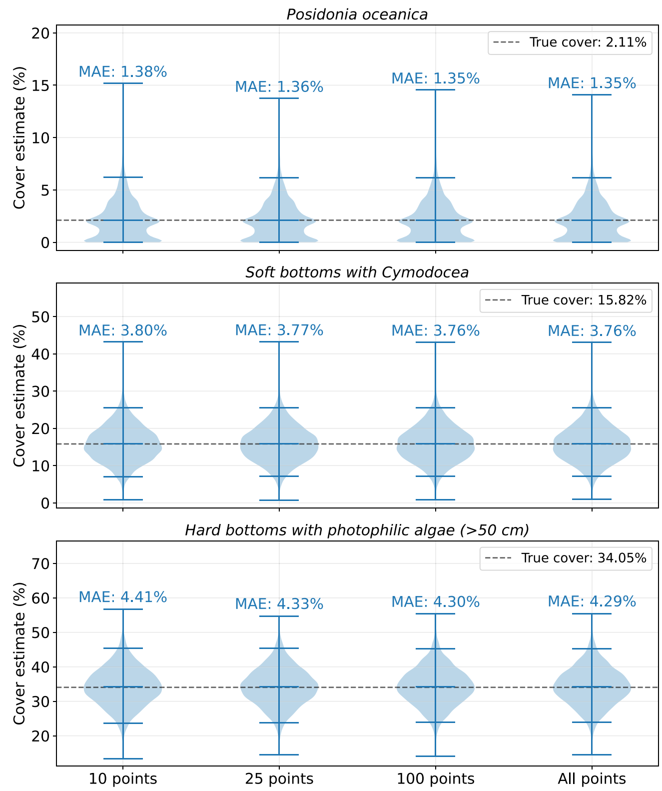


Fig. 6. Effect of annotation density on cover estimates. Violin plots show distributions of benthic cover estimates from 10,000 Monte Carlo simulations for three benthic classes with different abundance levels. Each plot displays the mean estimate, 95% confidence interval, extreme values, and mean absolute error (MAE). Base configuration: 50 randomly placed quadrats (0.5 m).

3.4. Resource allocation trade-offs

Fig. 8 directly compares the practical trade-off between sampling effort and annotation quality that practitioners face under resource constraints. The extensive sampling scenario (200 quadrats with lower-quality annotation) consistently outperforms the limited sampling scenario (50 quadrats with expert annotation) across all abundance classes.

For all classes, the extensive sampling strategy error distributions are substantially tighter around true cover values, with significantly lower MAE despite reduced annotation quality. To understand the relative contribution of sampling versus annotation to these outcomes, we next decompose total error into its constituent sources.

3.5. Error source decomposition

Table 2 partitions total estimation error into sampling, annotation, and interaction components for both resource allocation scenarios. Several patterns emerge from the bias–variance decomposition.

Sampling error exhibits high variance but is largely unbiased, which is a known properties of random sampling. Annotation error, by contrast, contributes little variance but introduces a small systematic bias, reflecting the consistent misrepresentation inherent to automated classifiers. As a result, sampling error accounts for nearly all observed MSE ($\mathrm{MSE}(e_{tot}) \approx \mathrm{MSE}(e_s)$), with annotation playing a negligible role.

This asymmetry has a direct implication for monitoring design: improving annotation quality reduces bias, which was already small, while increasing sample size reduces sampling variance, which dominates total error. The performance advantage of extensive sampling observed in Section 3.4 therefore reflects reduced variance for sampling error.

Finally, interaction terms are small relative to sampling variance. Additionally, several interactions are negative, confirming that improvements in annotation quality sometimes slightly degrade final estimates by reducing the compensating effect.

4. Discussion

Our Monte Carlo simulation framework reveals fundamental insights into the trade-offs between sampling design and annotation quality in benthic monitoring programs. By systematically evaluating how different error sources propagate through the complete photo-quadrat pipeline, our results challenge conventional assumptions about resource allocation in marine ecological surveys. The following discussion examines the implications of these findings for monitoring protocol design and highlights opportunities to optimize survey efficiency.

4.1. Interpretation and practical implications of key findings

This section examines the practical implications of our key results for benthic monitoring program design. We interpret the relative performance of different sampling strategies, evaluate the trade-offs between spatial coverage and annotation quality, and discuss how these results can inform resource allocation decisions in real-world survey contexts.

4.1.1. Quadrat number and size

Our experiments revealed that reducing quadrat size from 1 m to 0.5 m has minimal impact on benthic cover estimates, while increasing the number of quadrats is the most critical parameter for improving estimation accuracy. The minimal effect of quadrat size likely reflects the strong spatial clustering of benthic classes within our study area. Once a quadrat captures a particular habitat patch, increasing its size provides limited additional information about class distributions. This suggests that spatial replication is more valuable than individual quadrat coverage for improving estimation precision. Smaller quadrats have already been shown to be more precise and time-efficient in the literature (Pringle, 1984). From a practical perspective, smaller quadrats offer operational advantages through improved maneuverability during field missions, though the requirement for larger sample sizes increases survey duration and costs.

4.1.2. Quadrat placement strategies

Benthic monitoring programs often rely on transect-based sampling (Mantelatto et al., 2018; Leujak and Ormond, 2007). Our results challenge the widespread use of this method in marine monitoring programs, with random quadrat placement achieving lower estimation errors than all structured approaches. The superior performance of random sampling likely reflects reduced spatial autocorrelation effects, as random placement naturally maximizes spatial dispersion and enables coverage of more diverse habitat patches.

Among transect methods, free transects (linear transects of random orientation and position) performed better than their non-directional counterparts, likely due to broader spatial dispersion of quadrats. Indeed, linear placement maximizes the average distance between quadrats, enabling coverage of more diverse habitat areas and reducing spatial autocorrelation effects. In contrast, parallel transects showed particularly poor performance across all investigated classes. This poor performance results from increased spatial autocorrelation, as parallel transects tend to sample similar habitat patches, reducing effective sample independence and limiting habitat diversity captured. The systematic bias observed with parallel transects further compounds these issues.

These findings suggest that random quadrat placement should be the preferred method for benthic monitoring programs. However, random placement presents practical field challenges, including extensive underwater navigation and difficulties achieving true randomness during diving operations. Free transects offer a practical compromise, providing better performance than parallel transects while

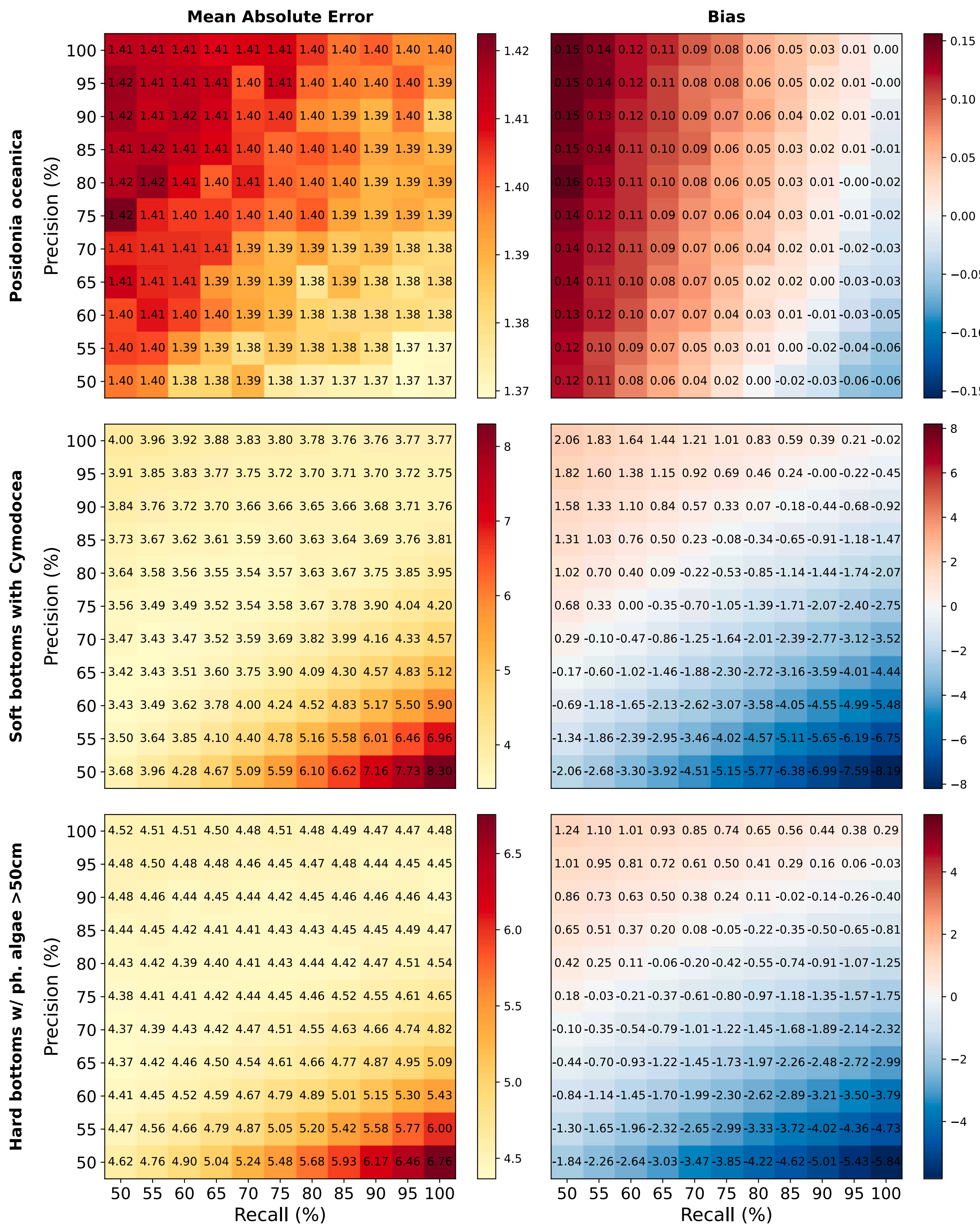


Fig. 7. Effect of annotation performance on cover estimates. Heatmaps show mean absolute error (MAE) and bias as functions of precision and recall for three benthic classes with different abundance levels, computed over 10,000 Monte Carlo simulations. Configuration: 50 random quadrats (0.5 m), 10 points per quadrat.

reducing field effort, though they still require generating random transect positions and orientations. An approach proposed by Foster et al. (2020) allows to optimize transect placement given constraints provided by the end user. Given these results, we strongly discourage the use of parallel transects and recommend prioritizing sampling independence, whether through random quadrat placement or carefully designed transect strategies that maximize spatial dispersion.

It is important to underline that our finding that simple random quadrat placement outperforms transect-based designs contrasts with other marine benthic surveys. Wang et al. (2020) found cluster sampling (similar to our non-directional transect approach) outperformed simple random sampling for tidal flat macrobenthos, while Delargy et al. (2025) showed spatially balanced designs outperformed simple random for sea scallops. However, as Delargy et al. (2025) emphasized,

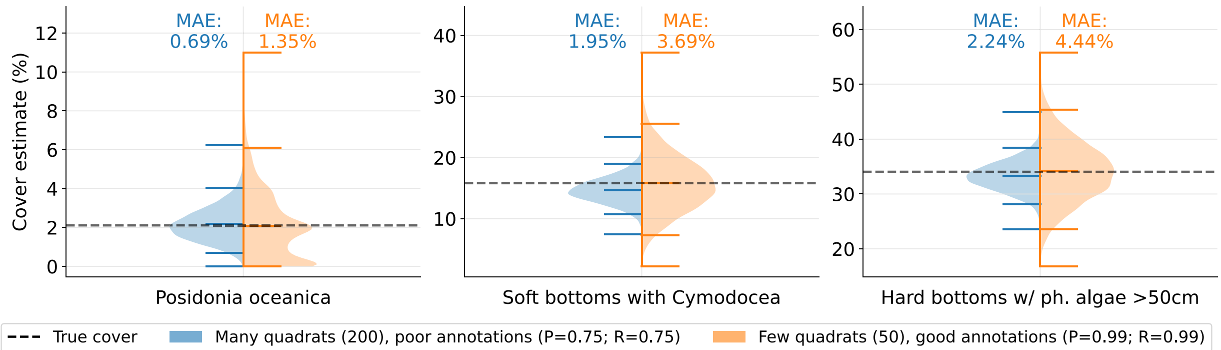


Fig. 8. Comparison of two opposite resource allocation strategies. Violin plots show distributions of benthic cover estimates from 10,000 Monte Carlo simulations for three benthic classes with different abundance levels. Each side represents a different resource allocation scenario: extensive sampling with automated annotation (200 quadrats, precision/recall = 0.75) versus limited sampling with expert annotation (50 quadrats, precision/recall = 0.99). Configuration: random quadrat placement, 0.5 m quadrats, 10 points per quadrat.

Table 2
Error decomposition for resource allocation scenarios. MSE(e_{tot}) is the observed estimation mean squared error, MSE(e_s) isolates sampling error assuming perfect annotation, and MSE(e_a) isolates annotation error on sampled quadrats. The interaction terms, Bias(e_s)Bias(e_a) and Cov(e_s, e_a), indicate whether the two error sources amplify each other (positive values) or partially cancel (negative values).

Extensive sampling, automated annotation
(200 quadrats, precision/recall = 0.75)

	MSE			Bias			Variance			Interaction terms	
Benthic cover category	e_{tot}	e_s	e_a	e_{tot}	e_s	e_a	e_{tot}	e_s	e_a	Bias(e_s)Bias(e_a)	Cov(e_s, e_a)
Posidonia oceanica	0.755	0.759	0.010	0.073	0.006	0.067	0.750	0.759	0.006	≃ 0	−0.007
Soft bot. w/ *Cymodocea*	5.770	5.195	1.390	−1.201	−0.118	−1.083	4.328	5.181	0.217	0.128	−0.535
Hard bot w/ phot. algae	7.684	7.346	1.311	−0.793	0.293	−1.086	7.056	7.260	0.132	−0.318	−0.168

Limited sampling, expert annotation
(50 quadrats, precision/recall = 0.99)

	MSE			Bias			Variance			Interaction terms	
Benthic cover category	e_{tot}	e_s	e_a	e_{tot}	e_s	e_a	e_{tot}	e_s	e_a	Bias(e_s)Bias(e_a)	Cov(e_s, e_a)
Posidonia oceanica	3.037	3.034	0.001	0.007	0.004	0.002	3.037	3.034	0.001	≃ 0	0.001
Soft bot. w/ *Cymodocea*	21.87	22.00	0.027	−0.129	−0.088	−0.042	21.85	22.00	0.025	0.004	−0.085
Hard bot. w/ phot. algae	30.34	30.39	0.018	0.222	0.265	−0.043	30.29	30.32	0.016	−0.011	−0.020

optimal statistical survey design is highly specific to the spatial nature of the species and habitats being targeted.

4.1.3. Number of annotated point per quadrat

Our simulations demonstrate that annotation density has minimal impact on benthic cover estimation accuracy, which suggests that current field practices of annotating over 25 points per quadrat Zotou et al. (2025), Blondin et al. (2024) may represent inefficient resource allocation. This finding aligns with Raitsos (2024), who evaluated point intercept methods for macroalgal cover estimation in Greek seas. Their direct comparison of 50, 100, and 150 points per photograph against reference grid measurements found that all three densities produced similar coverage estimates, with the 50-point method requiring significantly less analysis time.

The limited benefit of increased annotation density reflects the spatially clustered nature of benthic communities. Individual quadrats are homogeneous, typically containing only a few classes. Under these conditions, sparse point sampling adequately captures proportional composition, while additional points provide redundant information rather than improved precision. The diminishing returns observed by Raitsos (2024), where tripling annotation density from 50 to 150 points yielded negligible accuracy gains, directly supports this interpretation.

From a practical standpoint, reducing annotation density has substantial implications for manual annotation workflows. Decreasing from 25 to 10 points per quadrat reduces the annotation workload by more than half, significantly decreasing the time human experts spend on taxonomic identification and data entry. These findings suggest that practitioners could redirect resources from intensive within-quadrat annotation toward increased spatial sampling, potentially achieving greater overall monitoring accuracy through enhanced spatial coverage rather than annotation density.

This resource reallocation strategy aligns with our broader finding that spatial sampling represents the primary bottleneck in monitoring accuracy, rather than annotation precision within individual quadrats.

4.1.4. Annotation performance

A fundamental assumption in automated monitoring is that improved annotation performance directly translates to better system-level accuracy. Fig. 7 challenges the generality of this assumption as in our experiments, optimal benthic cover estimation accuracy does not occur at maximum annotation precision and recall.

Across all three abundance classes, the configuration yielding minimum estimation error occurs at intermediate annotation performance levels rather than perfect classification. This finding demonstrates that annotation quality optimization in isolation does not guarantee improved overall monitoring performance and may actually degrade it in certain scenarios. Many recent works propose approaches to automate benthic image annotation (Beijbom et al., 2015; Lowe et al., 2025; Jackett et al., 2023; Blondin et al., 2024; Zotou et al., 2025), focusing on optimizing and testing annotation performance metrics in isolation. However, they focus on optimizing and testing against annotation performance metrics alone. Our results demonstrate that evaluating annotation quality against known ground truth within the complete monitoring pipeline reveals system-level behaviors that cannot be predicted from annotation performance metrics alone.

For low-abundance classes like *Posidonia oceanica* (∼ 2%), estimation accuracy remains relatively stable across different annotation performance levels, with decreasing precision actually improving estimation accuracy. This behavior reflects the systematic underrepresentation of rare classes in limited samples, where the target class may be entirely absent from sampled quadrats. Under these conditions, reduced precision increases false positive predictions, compensating for systematic undersampling and improving overall cover estimation accuracy. While decreasing precision appears to increase systematic bias, in resource-constrained settings this bias precisely compensates for limited sample representativity and decreases mean absolute error.

For higher abundance classes (Soft bottoms w/ Cymodocea and Hard bottoms w/ photophilic algae), minimum errors occur at low recall values (∼ 0.5) combined with moderate precision (∼ 0.75). This contrasting behavior reflects their substantially higher abundance levels, where systematic overrepresentation in limited samples becomes problematic rather than beneficial. A problematic region exists at the intersection of low precision and high recall, but other precision–recall combinations yield acceptable estimation performance.

These findings reveal that the relationship between annotation quality and monitoring accuracy depends critically on target class abundance and sampling intensity.

It is also important to emphasize that the above conclusions are drawn from simulations based on a single high-resolution benthic habitat map in the Tyrrhenian Sea. The observed patterns between annotation quality and estimation accuracy likely depend on habitat spatial structure, taxonomic composition, and the specific abundance distributions present in this system. For example, systems with higher taxonomic diversity would likely require larger sample sizes to characterize all classes, as each additional class mechanistically reduces individual class abundance. For rare species in particular, ensuring the classifier performs better than chance becomes non-trivial, and the regime where precision falls below 0.5 was not explored in our study. More spatially uniform habitats may also influence these trade-offs, potentially increasing the relative importance of annotation quality as spatial variability no longer drives sampling error.

Nevertheless, the existence of scenarios where improving annotation quality does not improve monitoring accuracy has important practical implications. It demonstrates that annotation performance optimization in isolation can lead to suboptimal resource allocation. Before investing substantial resources in annotation refinement or model development, practitioners should conduct preliminary assessments to determine whether annotation quality represents a bottleneck in their specific monitoring system.

4.1.5. Comparing resource allocation strategies

In our direct comparison of resource allocation strategies, extensive sampling with automated annotation (200 quadrats, precision/recall = 0.75) consistently outperformed limited sampling with expert annotation (50 quadrats, precision/recall = 0.99) across all three categories. This finding suggests that prioritizing spatial sampling over annotation perfection can improve monitoring performance in resource-constrained monitoring programs.

These results indicate that automated annotation systems, even with moderate performance levels, enable more effective monitoring strategies than traditional approaches emphasizing annotation accuracy. The superior performance of extensive sampling with imperfect automation suggests that marine ecologists should prioritize field data collection over laboratory annotation precision. Reallocating resources from intensive manual annotation toward increased spatial sampling can substantially improve benthic cover estimation accuracy and represents a fundamental shift in optimal monitoring program design.

Our error source decomposition supports this resource allocation recommendation. In our experiment, sampling-related errors strongly dominated annotation-related errors across all three benthic classes, and costly annotation improvements yielded minimal gains in cover estimation accuracy compared to collecting additional photo-quadrat samples.

To the best of our knowledge, this work is the first to quantitatively compare the relative contributions of sampling effort and annotation quality to overall monitoring error using controlled ground truth validation in a benthic monitoring context.

4.2. Perspectives from systems engineering: beyond component-level optimization

Our findings exemplify a fundamental principle in Systems Engineering: optimizing individual system components in isolation often leads to suboptimal overall system performance (Sage and Rouse, 2011). Traditional approaches to benthic monitoring have treated data collection (Molloy et al., 2013) and annotation (Beijbom et al., 2015; Zotou et al., 2025) as separate optimization problems, with each stakeholder group focusing on their own constraints and objectives rather than system-level performance.

Marine ecologists have historically optimized their data collection protocols under strict resource constraints. Field time is expensive and limited, requiring careful balance between sample quantity and logistical feasibility. Similarly, laboratory processing time constrains annotation effort — more samples mean longer processing times for expert taxonomists. This leads ecologists to collect moderate sample sizes with intensive annotation protocols to ensure high-quality labels.

Simultaneously, machine learning engineers optimize annotation algorithms toward maximum precision and recall, treating prediction accuracy as the primary objective. Substantial resources are invested in developing sophisticated models, collecting extensive training datasets, and fine-tuning architectures to minimize classification errors. This component-level optimization assumes that better annotation models automatically improve monitoring system performance.

However, our results reveal that this dual optimization strategy can produce a suboptimal equilibrium. Our case study demonstrates that accepting lower annotation accuracy in favor of increased sampling effort can drastically improve cover estimates. This suggests a fundamentally different resource allocation strategy: many samples with fast, imperfect annotation outperform fewer samples with perfect annotation.

From an ecological perspective, this implies shifting toward extensive sampling protocols with minimal concern for laboratory processing time, since automated models can handle annotation workload regardless of sample volume. Field efforts should prioritize spatial coverage over annotation quality.

From a machine learning perspective, this finding shifts the objective from developing perfect annotation models to creating fast, deployable systems that enable rapid data processing. The goal becomes minimizing the time-to-deployment for working models rather than maximizing final model performance. At the system level, where the objective is optimizing final benthic cover estimate accuracy, our results suggest that many samples with automated annotation, even containing small systematic errors, outperform traditional approaches emphasizing annotation perfection. This systems perspective reveals that the bottleneck in monitoring accuracy lies in spatial sampling, not annotation quality, fundamentally challenging current resource allocation practices in marine monitoring programs.

4.3. Study limitations and generalizability

While our study provides valuable insights, we must acknowledge that our Monte Carlo simulations were conducted on a single high-resolution benthic map from the Tyrrhenian Sea, Italy. While this approach enables rigorous quantification of error propagation under controlled conditions, the generalizability of our conclusions to different environments, habitat types, and spatial scales requires careful consideration.

The specific habitat composition and spatial structure of our study site influence how sampling and annotation errors interact. For instance, highly fragmented habitats or regions dominated by morphologically similar species could alter the relationship between sampling effort and annotation quality. Additionally, the taxonomic resolution of classification schemes (e.g., species-level versus functional group classification) may influence annotation-sampling error interactions, as finer taxonomic distinctions typically increase annotation difficulty by increasing the number of output classes for the model.

Conducting similar analyses across diverse marine ecosystems would strengthen our conclusions, yet such studies remain challenging due to the scarcity of centimeter-resolution benthic habitat maps. The intensive field work, advanced remote sensing techniques, and expert annotation required to generate reference maps of comparable quality represent significant barriers to replication. Technological limitations also limit replicability in deeper or more turbid waters, where generating high-resolution maps is still difficult. Nevertheless, our study region exhibits habitat characteristics common to many coastal environments, suggesting that our broader conclusions regarding optimal sampling design and error interactions are likely to be applicable to other Mediterranean coastal marine ecosystems, even if some variability in the estimates is expected.

Importantly, the demonstrated existence of annotation-sampling error compensation in at least one realistic case establishes this phenomenon as a critical consideration for monitoring program design. Given the potential for suboptimal resource allocation, we encourage practitioners to evaluate these trade-offs systematically (e.g., using simulation approaches like ours) rather than relying on assumptions about error dominance.

In summary, our study demonstrates how integrating computational and ecological approaches can yield practical knowledge for biodiversity monitoring. Beyond addressing challenges in marine ecosystems, the core idea of this paper (i.e., jointly evaluating sampling and annotation error impacts on ecological indicators through simulation) can be adapted to optimize monitoring designs across diverse contexts in ecology and evolution.

CRediT authorship contribution statement

Joris Guerin: Writing – original draft, Validation, Software, Methodology, Formal analysis, Conceptualization. **Guilherme Longo:** Writing – review & editing, Conceptualization. **Regina Nobre:** Writing – review & editing, Conceptualization. **Célia Blondin:** Writing – review & editing, Software. **Laure Berti-Equille:** Writing – review & editing, Conceptualization. **Daniele Ventura:** Writing – review & editing, Visualization, Data curation.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used CLAUDE SONNET 4 model in order to improve writing clarity and conciseness. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Joris GUERIN reports financial support and article publishing charges were provided by IRD. Guilherme Longo reports a relationship with National Council for Scientific and Technological Development that includes: funding grants. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the French Institute for Sustainable Development (IRD). Guilherme Longo received a research productivity scholarship provided by the National Council for Scientific and Technological Development – CNPq (308072/2022-7).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ecoinf.2026.103784.

Data availability

The complete reference benthic habitat classification map and associated RGB orthophoto used in this study are available upon reasonable request to the author (*daniele.ventura@uniroma1.it*). Full dataset documentation following FAIR (Findable, Accessible, Interoperable, Reusable) and CARE (Collective Benefit, Authority to Control, Responsibility, Ethics) principles is provided in Supplementary Material S1. A lightweight version of the reference map sufficient to reproduce all simulations and analyses presented in this paper is freely available in the code repository at https://github.com/jorisguerin/benthic-photoquadrat-study.

References

Aued, A.W., Smith, F., Quimbayo, J.P., Cândido, D.V., Longo, G.O., Ferreira, C.E.L., Witman, J.D., Floeter, S.R., Segal, B., 2018. Large-scale patterns of benthic marine communities in the Brazilian province. PLoS One 13 (6), 1–15. http://dx.doi.org/10.1371/journal.pone.0198452.

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), 1–22. http://dx.doi.org/10.1371/journal.pone.0130312.

Biscaia Zamoner, J., Aued, A.W., Macedo-Soares, L.C.P., Picolotto, V.A.P., Eiras Garcia, C.A., Segal, B., 2021. Integrating oceanographic data and benthic community structure temporal series to assess the dynamics of a marginal reef. Front. Mar. Sci. Volume 8 - 2021, http://dx.doi.org/10.3389/fmars.2021.762453.

Blondin, C., Guerin, J., Berti-Equille, L., Ortigara Longo, G., Inagaki, K., 2024. Hierarchical classification for automated image annotation of coral reef benthic structures. In: NeurIPS 2024 Workshop on Tackling Climate Change with Machine Learning.

Delargy, A.J., Cassidy, K.S., Lisi, A.D., Stokesbury, K.D., 2025. A comparison of survey designs for marine benthic invertebrate sampling. Fish. Res. 285, 107384. http://dx.doi.org/10.1016/j.fishres.2025.107384.

Eleftheriou, A., 2013. Methods for the study of marine benthos. John Wiley & Sons, Ltd, http://dx.doi.org/10.1002/9781118542392.

Foster, S.D., Hosack, G.R., Monk, J., Lawrence, E., Barrett, N.S., Williams, A., Przeslawski, R., 2020. Spatially balanced designs for transect-based surveys. Methods Ecol. Evol. 11 (1), 95–105. http://dx.doi.org/10.1111/2041-210X.13321.

Galparsoro, I., Borja, A., Uyarra, M.C., 2014. Mapping ecosystem services provided by benthic habitats in the European north atlantic ocean. Front. Mar. Sci. Volume 1 - 2014, http://dx.doi.org/10.3389/fmars.2014.00023.

Jackett, C., Althaus, F., Maguire, K., Farazi, M., Scoulding, B., Untiedt, C., Ryan, T., Shanks, P., Brodie, P., Williams, A., 2023. A benthic substrate classification method for seabed images using deep learning: Application to management of deep-sea coral reefs. J. Appl. Ecol. 60 (7), 1254–1273. http://dx.doi.org/10.1111/1365-2664.14408.

Jayachandran, P., Bijoy Nandan, S., Jima, M., Philomina, J., Vishnudattan, N., 2022. Chapter 10 - benthic organisms as an ecological tool for monitoring coastal and marine ecosystem health. In: Godson, P.S., Vincent, S.G.T., Krishnakumar, S. (Eds.), Ecology and Biodiversity of Benthos. Elsevier, pp. 337–362. http://dx.doi.org/10.1016/B978-0-12-821161-8.00004-0.

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 3, e954. http://dx.doi.org/10.7717/peerj.954.

Kritzer, J.P., DeLucia, M.-B., Greene, E., Shumway, C., Topolski, M.F., Thomas-Blate, J., Chiarella, L.A., Davy, K.B., Smith, K., 2016. The importance of benthic habitats for coastal Fisheries. BioScience 66 (4), 274–284. http://dx.doi.org/10.1093/biosci/biw014.

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. Sens. 11 (24), http://dx.doi.org/10.3390/rs11242907.

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

Lowe, S.C., Misiuk, B., Xu, I., Abdulazizov, S., Baroi, A.R., Bastos, A.C., Best, M., Ferrini, V., Friedman, A., Hart, D., et al., 2025. BenthicNet: A global compilation of seafloor images for deep learning applications. Sci. Data 12 (1), 230. http://dx.doi.org/10.1038/s41597-025-04491-1.

Mantelatto, M.C., Cruz, I.C.S., Creed, J.C., 2018. A review of the knowledge of shallow subtidal benthic communities in Brazil. Mar. Environ. Res. 140, 69–77. http://dx.doi.org/10.1016/j.marenvres.2018.05.022.

Mello, T.J., Longhini, C.M., Wanderley, B.M.S., da Silva, C.A., Lehrback, B.D., Bom, F.C., Neto, R.R., Sá, F., Vieira, E.A., Costa, V.E., Longo, G.O., 2025. Pollution affects even oceanic marine protected areas in southwestern atlantic. Environ. Pollut. 366, 125485. http://dx.doi.org/10.1016/j.envpol.2024.125485.

Molloy, P., Evanson, M., Nellas, A., Rist, J., Marcus, J., Koldewey, H., Vincent, A., 2013. How much sampling does it take to detect trends in coral-reef habitat using photoquadrat surveys? Aquat. Conserv.: Mar. Freshw. Ecosyst. 23 (6), 820–837. http://dx.doi.org/10.1002/aqc.2372.

Montilla, L.M., Miyazawa, E., Ascanio, A., López-Hernández, M., Mariño-Briceño, G., Rebolledo-Sánchez, Z., Rivera, A., Mancilla, D.S., Verde, A., Cróquer, A., 2020. The use of pseudo-multivariate standard error to improve the sampling design of coral monitoring programs. PeerJ 8, e8429. http://dx.doi.org/10.7717/peerj.8429.

Perry, C.T., Alvarez-Filip, L., 2019. Changing geo-ecological functions of coral reefs in the anthropocene. Funct. Ecol. 33 (6), 976–988. http://dx.doi.org/10.1111/1365-2435.13247.

Pringle, J.D., 1984. Efficiency estimates for various quadrat sizes used in benthic sampling. Can. J. Fish. Aquat. Sci. 41 (10), 1485–1489. http://dx.doi.org/10.1139/f84-182.

Proudfoot, B., Devillers, R., Brown, C.J., 2020. Integrating fine-scale seafloor mapping and spatial pattern metrics into marine conservation prioritization. Aquat. Conserv.: Mar. Freshw. Ecosyst. 30 (8), 1613–1625. http://dx.doi.org/10.1002/aqc.3360.

Raitsos, D.E., 2024. Comparative Study of Benthic Cover Estimation Methods from Macroalgal Photographic Samples (Ph.D. thesis). Hellenic Centre for Marine Research.

Sage, A.P., Rouse, W.B., 2011. Handbook of systems engineering and management. John Wiley & Sons.

Ventura, D., Grosso, L., Pensa, D., Casoli, E., Mancini, G., Valente, T., Scardi, M., Rakaj, A., 2023. Coastal benthic habitat mapping and monitoring by integrating aerial and water surface low-cost drones. Front. Mar. Sci. Volume 9 - 2022, http://dx.doi.org/10.3389/fmars.2022.1096594.

Wang, J., Zhang, C., Xue, Y., Xu, B., Ren, Y., Chen, Y., 2020. Comparison of different sampling designs for macrozoobenthos survey in a tidal flat. Reg. Stud. Mar. Sci. 35, 101113. http://dx.doi.org/10.1016/j.rsma.2020.101113.

Williams, I.D., Couch, C.S., Beijbom, O., Oliver, T.A., Vargas-Angel, B., Schumacher, B.D., Brainard, R.E., 2019. Leveraging automated image analysis tools to transform our capacity to assess status and trends of coral reefs. Front. Mar. Sci. Volume 6 - 2019, http://dx.doi.org/10.3389/fmars.2019.00222.

Zotou, M., Sini, M., Trygonis, V., Greggio, N., Mazaris, A.D., Katsanevakis, S., 2025. Towards scalable ecological monitoring: Assessing AI-based annotation of benthic images. J. Mar. Sci. Eng. 13 (9), http://dx.doi.org/10.3390/jmse13091721.