POLITECNICO DI TORINO
Repository ISTITUZIONALE

Integrating deep learning and spatial statistics in marine ecosystem monitoring

Original
Integrating deep learning and spatial statistics in marine ecosystem monitoring / Sangiovanni, G.M., Mastrantonio, G., Pollice, A., Ventura, D., Jona Lasinio, G.. - In: ENVIRONMENTAL AND ECOLOGICAL STATISTICS. - ISSN 1352-8505. - 33:2(2026), pp. 1161-1183. [10.1007/s10651-026-00732-7]

Availability:
This version is available at: 11583/3011164 since: 2026-11-06T14:00:09Z

Publisher:
Springer

Published
DOI:10.1007/s10651-026-00732-7

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

12 November 2026

Environmental and Ecological Statistics
https://doi.org/10.1007/s10651-026-00732-7

Integrating deep learning and spatial statistics in marine ecosystem monitoring

Gian Mario Sangiovanni[1] · Gianluca Mastrantonio[2] · Alessio Pollice[3] · Daniele Ventura[4] · Giovanna Jona Lasinio[1]

Received: 20 November 2025 / Revised: 27 April 2026 / Accepted: 30 April 2026
© The Author(s) 2026

Abstract

In ecology, photogrammetry is a crucial method for efficiently collecting non-destructive samples of natural environments. When estimating the spatial distribution of animals, detecting objects in large-scale imagery becomes essential. Deep learning-based object detection enables large-scale analysis but introduces uncertainty because detection probability depends on environmental and observational factors. To address detection bias, we model the distribution of benthic sea cucumbers in an area of the Italian Tyrrhenian coast near Giglio Island using a Thinned Log-Gaussian Cox Process (LGCP). The main objective of this study is to develop a modular framework that integrates a deep learning-based object detector with a spatial point process to correct for systematic undercounting in marine surveys. We employ the YOLOv11 architecture to automate the identification of individuals. We assume that a true underlying intensity describes the actual population, while the observed detections correspond to a subset of individuals that are not always detected, leading to a degraded intensity. The detection process is modeled through a probabilistic function that depends on environmental and observational factors, including local density, network confidence scores, and object size. Manual annotations are used as a benchmark, and we compare the thinned LGCP with an unthinned model fitted to the deep learning-based object detections. The proposed approach reduces bias in spatial intensity estimation and improves agreement with manual surveys (average Pearson residual 6.150 vs. 6.694 and average raw residual 1.979 vs. 2.092). By explicitly accounting for detection uncertainty, the framework increases the reliability of large-scale benthic monitoring, supporting habitat assessment and evidence-based marine conservation.

Keywords Detection bias · Marine biology · Object detection · Underreporting · Thinned log-Gaussian cox process

Extended author information available on the last page of the article

Published online: 15 May 2026

1 Introduction

Modern underwater imaging technology has created both an opportunity and a challenge for marine ecology. Advances in Structure from Motion (SfM) photogrammetry, when paired with Diver Propulsion Vehicles (DPVs), now enable researchers to capture high-resolution imagery across hundreds of square meters of seafloor in a single survey (Ventura et al. 2025). A typical large-scale benthic survey can generate thousands of images (once tiled in a suitable format), each containing individual organisms. This represents an unprecedented opportunity to study marine populations at spatial scales that were simply unattainable with traditional survey methods. However, this technological capability has revealed that our analytical capacity has not kept pace with our data collection capabilities. The standard approach to analyze underwater imagery remains manual annotation by trained biologists, who visually inspect each image and mark the location and identity of target organisms. While this method provides reliable, high-quality data, it is fundamentally limited by its labor-intensive nature (Beijbom et al. 2012; Mahmood et al. 2016). The bottleneck has shifted from data collection to data analysis, effectively preventing us from monitoring benthic ecosystems at the spatial and temporal scales necessary to understand ecological dynamics and support conservation planning.

Deep learning appears to offer a solution. Convolutional neural networks (CNNs) have demonstrated impressive capabilities for automated species detection and counting in ecological imagery (Weinstein 2018; De Iaco et al. 2022). These methods can process thousands of images in minutes and often achieve detection accuracy comparable to human annotators on benchmark datasets. In practice, detectors are typically trained on manually annotated images and then applied to entire surveys to produce large-scale counts. Yet most applications of deep learning in ecology treat detection as the final product rather than as the beginning of ecological inference (Goodwin et al. 2022).

Researchers focus primarily on maximizing detection accuracy (or related metrics), while two fundamental issues remain largely unaddressed. First, all deep learning-based detection systems imperfectly observe ecological reality. Even the most sophisticated detectors make systematic mistakes. They miss cryptic or partially obscured individuals, perform poorly under challenging lighting conditions or in turbid water, confuse target species with morphologically similar organisms or background structures, and show inconsistent detection performance across different habitats or image qualities (Zhai et al. 2022). These errors are not random noise that cancels out over large samples. They represent systematic observation bias (Guttorp et al. 2026). Secondly, beyond the detection accuracy problem lies a deeper conceptual issue about what we actually want to learn from imagery data. The spatial distribution of organisms reflects underlying ecological processes. Benthic species distribute themselves across seascapes according to environmental gradients in depth, slope, and substrate type, responding to habitat heterogeneity, interacting with other species, and reflecting historical colonization dynamics. These spatial patterns constitute the primary scientific interest in

most ecological studies. They reveal which habitats species prefer, identify the environmental factors that structure populations, and enable predictions about how distributions might shift in response to environmental change. Most detection-focused approaches collapse this rich spatial information into simple aggregate counts.

For those reasons, we need a different approach that treats deep learning-based object detections not as ground truth but as an observation process that must be modeled explicitly within a larger inferential framework. This perspective has long been central to ecological statistics. The distinction between actual ecological state and observed data forms the conceptual foundation for species distribution models, capture-recapture analyses, and occupancy models (MacKenzie et al. 2002; Warton et al. 2013). The same logic should be applied in this context. What distinguishes this work from most existing deep learning applications in ecology is the way deep learning-based object detections are treated within the inferential framework. Rather than using detector outputs as direct proxies for true abundance, we explicitly model them as a thinned observation of an underlying ecological process. Detection probability is not assumed known, nor estimated through repeated surveys or labeled absences, but inferred jointly with the spatial intensity via a thinned Log-Gaussian Cox Process (LGCP) (Møller et al. 1998; Chakraborty et al. 2011; Warton and Shepherd 2010; Panunzi et al. 2025; Lewy and Kristensen 2009). This allows us to perform ecological inference directly from presence-only detections while accounting for systematic observation bias.

We developed this approach here for sea cucumber populations assessed through large-scale underwater photographic surveys. Sea cucumbers (Holothuroidea) make an excellent model system for several reasons. Ecologically, they function as ecosystem engineers, processing organic matter in the detrital food web pathway (Ciriminna et al. 2024). They process large quantities of sediment and contribute substantially to nutrient cycling and bioturbation in benthic communities ranging from shallow seagrass beds to deep-sea environments (Purcell et al. 2016; Schneider et al. 2013; Lopez and Levinton 1987). Many sea cucumber populations face conservation concerns due to intensive fishing pressure driven by commercial markets (predominantly Asian). Improving the accuracy of spatial distribution estimates for these species has direct implications for stock assessment, habitat suitability mapping, marine protected area design, and long-term monitoring schemes. More reliable spatial information enables managers to identify priority conservation areas, detect early population declines, evaluate the effectiveness of protection measures, and anticipate distributional shifts under environmental change. From a methodological standpoint, sea cucumbers present typical challenges for object detection. They show variable appearance, exhibit cryptic colouration, resemble background substrate and occur in visually complex habitats. Methods that work for sea cucumbers should generalize reasonably well to other benthic species facing similar detection challenges.

Our approach divides the problem into two stages. First, a deep learning detector trained on manual annotated imagery (YOLOv11 (Khanam and Hussain 2024)) identifies candidate organisms and converts bounding boxes into spatial point locations using centroids as proxies for position. We do not treat these detections as

error-free observations. The second stage models the spatial distribution of detected organisms using a thinned LGCP. This formulation naturally captures complex spatial patterns and their relationships with environmental covariates. Our key innovation involves explicitly incorporating a thinning mechanism. We model the detections as arising from an underlying true point pattern (actual sea cucumber locations) that has been stochastically thinned. The thinning probability reflects detection reliability and can vary spatially, depending on factors such as image quality, habitat complexity, and others. This thinning framework provides the natural connection between our two stages. The detector provides us with presence-only data, without indicating the detection probability. The spatial model estimates this detection probability while simultaneously recovering the actual underlying spatial intensity and quantifying its relationship with environmental covariates. We can therefore correct for under detection and obtain unbiased estimates of total abundance and spatial distribution patterns. The framework also provides appropriate uncertainty quantification throughout.

Our approach is in contrast to recent computer vision studies that embed spatial point process models directly within detection networks (Pham et al. 2016; Mabon et al. 2022, 2023, 2024; Descombes and Zerubia 2002). Due to their flexibility and scalability, neural networks and spatial point process models have been increasingly recognized as a *convenient couple* for the statistical analysis of multivariate point patterns (Mateu and Jalilian 2022). In those settings, spatial statistics are primarily used as regularization tools to improve detection accuracy. Here, the direction of integration is reversed: detection serves ecological inference rather than the opposite. By separating detection and inference, the framework scales to large spatial domains, facilitates integration of environmental covariates, and enables robust uncertainty propagation across modeling stages.

The objective of this study is to develop and assess a unified framework that integrates deep learning object detections with a thinned spatial point process to obtain unbiased estimates of spatial distribution from presence-only data. The remainder of this paper proceeds as follows. Sect. 2 describes the study area and details both stages of our framework: the object detection phase and the thinned LGCP model, including environmental covariate preprocessing and model selection procedures. Sect. 3 presents results comparing alternative model specifications and demonstrates how our approach recovers spatial patterns and abundance estimates from detector outputs. Sect. 4 discusses the ecological implications, methodological limitations, and potential extensions of this framework to other benthic monitoring applications.

2 Materials and methods

2.1 Study Area

The research was conducted at Punta Gabbianara (N: 42.364867°; E: 10.920210°), situated along the northeastern shoreline of Giglio Island in the Tyrrhenian Sea, Italy (Fig. 1). The study area encompasses approximately 5.500 square meters of infralittoral seafloor, extending from 8 to 27 meters depth, and is characterized by

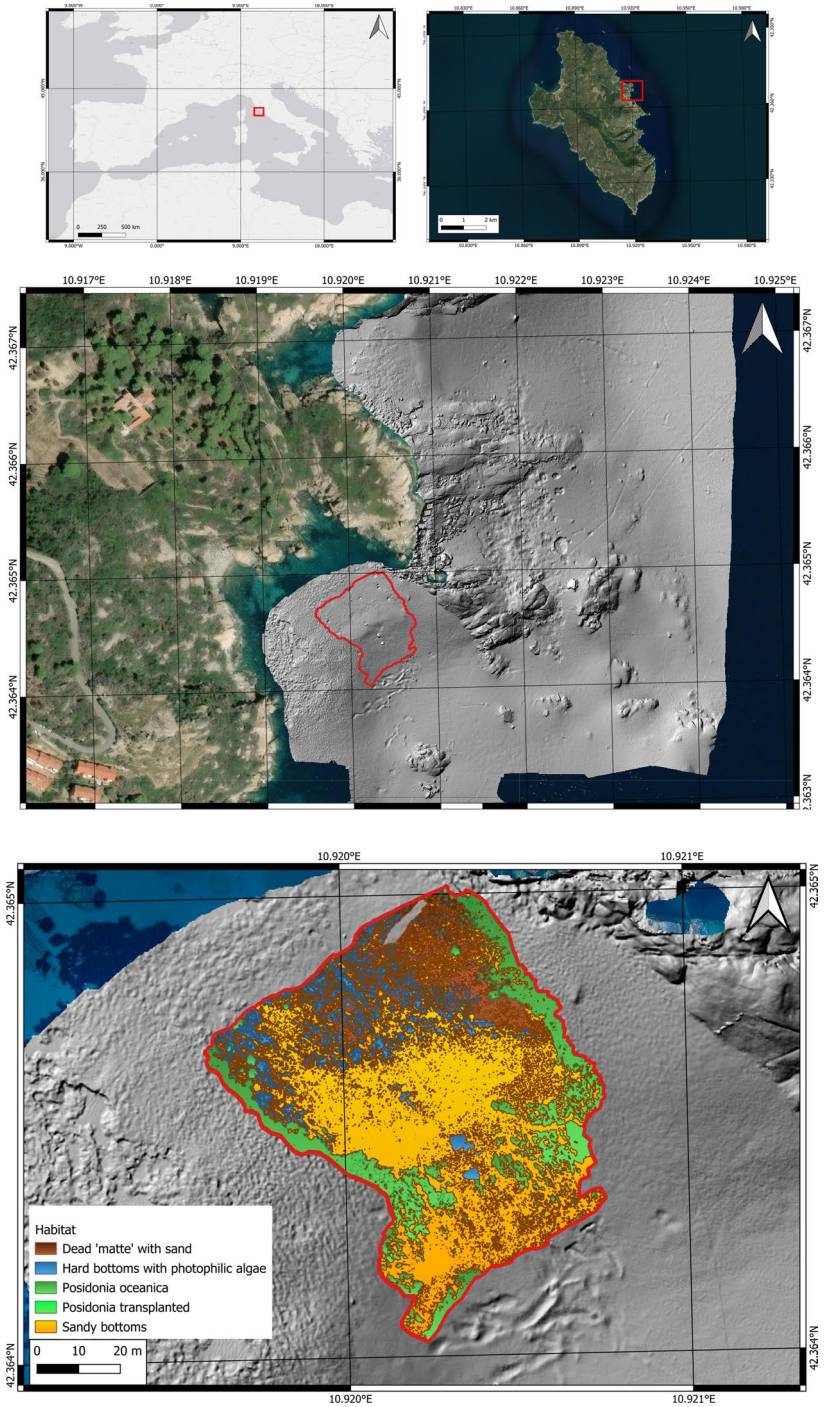

Fig. 1 Punta Gabbianara study site

heterogeneous benthic habitats that have experienced substantial anthropogenic impacts. The site has undergone ecological disturbance and subsequent restoration following the Costa Concordia shipwreck in 2012 (Casoli et al. 2017), with large-scale *Posidonia oceanica (L.) Delille* (P. Oceanica) transplantation initiated in 2019 (Mancini et al. 2022). The current benthic landscape represents a complex mosaic of Mediterranean infralittoral communities, defined by sandy sedimentary areas interspersed with granitic geological formations, including boulders and cobbles colonized by photophilic macroalgal assemblages. The site exhibits both remnant natural *P. Oceanica* patches and newly established transplanted meadows, which are progressively colonizing fragmented areas of dead rhizome matrices (locally termed *matte*) and adjacent sandy substrates. This heterogeneous habitat configuration provides an ideal setting for investigating benthic invertebrate distributions across varying substrate types and restoration stages, making it particularly suitable for automated detection and spatial modeling approaches.

2.2 Data collection

In 2022, we conducted five photogrammetric mapping missions spanning January through November, applying methodologies previously validated for tracking seagrass transplant distribution (Ventura et al. 2022) and characterizing sea cucumber microhabitat selection (Ventura et al. 2025). Underwater image acquisition employed a GoPro Hero 10 action camera (23 MP,5568×4176 pixel resolution) affixed to a DPV, enabling extensive spatial coverage while constraining dive duration. The camera was positioned approximately 5 meters above the seabed, maintaining 75% image overlap between successive frames, yielding 3 mm pixel ground sample distance which was sufficient for reliable organism detection and substrate characterization.

We incorporated the same suite of spatial variables utilized in previous analyses (Mastrantonio et al. 2024; Ventura et al. 2025), maintaining a resolution of $0.21 \times 0.24\ m$. Continuous environmental variables, including seafloor depth, slope angle and terrain roughness, which were all derived from the high-resolution digital surface models. Environmental predictors were selected based on ecological knowledge of holothurian habitat use, which is primarily influenced by substrate type, structural complexity, and organic matter availability. Collinearity was assessed using pairwise Pearson correlations, and terrain roughness was not included to avoid identifiability issues given the strong correlation ($\rho \approx 0.7$) with slope (Fig. 2). In addition to topographic predictors, we included texture metrics derived from orthophotomosaics to capture habitat heterogeneity not represented by elevation-based variables. Specifically, we applied Laplacian and Sobel operators (Robinson 1977) to capture complementary aspects of seafloor texture. The Laplacian (second-derivative filter) highlights fine-scale textural discontinuities and blob-like features, while the Sobel operator (first-derivative filter) emphasizes directional edges and linear boundaries. For the Sobel filter, we computed horizontal and vertical gradients separately, then calculated edge magnitude as the Euclidean norm of the gradients. Both Laplacian and Sobel magnitude values were averaged across red-green-blue

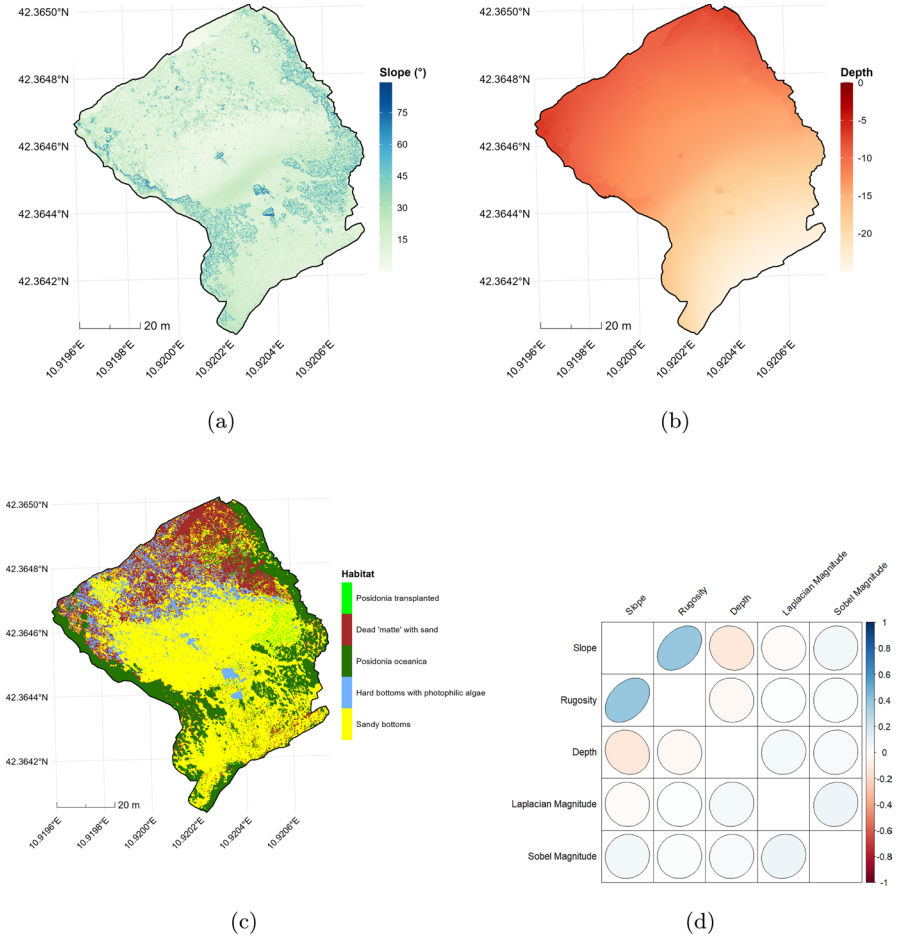

Fig. 2 Spatial distribution of **a** slope, **b** depth, and **c** habitat types within the study area. Panel **d** shows the correlation matrix among environmental covariates

channels to produce single-band texture metrics. These textural variables may capture ecologically relevant habitat features such as rubble-sand interfaces or algal canopy boundaries that are not adequately represented by topographic metrics alone.

To classify substrate types from the exceptionally detailed orthophoto mosaics, we applied an object-based image analysis framework (Ventura et al. 2025; Fallati et al. 2024). This approach first segments imagery into discrete units sharing similar spectral signatures and spatial characteristics, then applies supervised machine learning to assign each unit to a habitat category. Five benthic classes were identified: sandy bottoms, hard bottoms with photophilic algae, dead matte mixed with sand, natural *P. oceanica* meadow, and transplanted *P. oceanica*. All environmental variables were aggregated to a regular 1×1 m grid using zonal averaging. For each grid cell, we derived two representations of habitat information: the fractional

coverage of each habitat type and the dominant habitat class. This dual representation enables subsequent models to capture both fine-scale habitat heterogeneity and the primary substrate composition at the analysis resolution.

2.3 Detection task

To automate the identification of sea cucumbers within orthophotomosaics, we developed an object detection workflow based on the You Only Look Once version 11 (YOLOv11) architecture (Khanam and Hussain 2024). While the original YOLO framework established the foundation for real-time detection (Redmon et al. 2016), this study utilizes the more advanced YOLOv11 implemented using the Ultralytics repository. Compared to earlier YOLO versions, YOLOv11 incorporates updated backbone and feature-aggregation modules, anchor-free detection strategies, and improved training procedures, resulting in enhanced performance for small objects and visually complex scenes such as underwater habitats (Ünel et al. 2019).

The training dataset was derived from twelve high-resolution orthophotomosaics acquired in 2021 at ecologically similar sites outside the main study area, including both transplantation sites and natural habitats such as *Cala Cupa* and *Cala di Mezzo* along the northeastern coast of the island. By training on spatially independent sites, we ensured that the model learned generalized morphological features of sea cucumbers rather than site-specific background correlations, thereby enhancing its ability to generalize to the Punta Gabbianara study area. To match the scale of the target organisms, each mosaic was tiled into 640×640 pixel patches at a spatial resolution of 0.5 cm/pixel using the QGIS plugin Deepness (Aszkowski et al. 2023). From an initial pool of 3609 tiles, we curated a dataset of 2492 manually annotated images, including 2292 positive tiles and 200 negative samples depicting only the seabed. The inclusion of these negative examples, spanning various habitats, was a strategic choice to enhance the model's generalization capacity and reduce false positives. A rigorous, dual-reviewer annotation process on the Roboflow platform (Alexandrova et al. 2015) ensured high-quality labels. The dataset was partitioned into training (70%), validation (20%), and test (10%) sets.

To further enhance robustness and generalization capabilities, a comprehensive suite of data augmentation techniques was applied. These included geometric transformations (e.g., rotations, flips, scaling and cropping) and photometric adjustments (e.g., variations in brightness, contrast, hue, and saturation) designed to simulate realistic changes in underwater imaging conditions. Additional effects such as blur and noise were incorporated to reproduce the impact of water turbidity and camera motion. This combination of transformations aimed to improve the model's resilience to variations in seabed texture, lighting conditions and organism orientation (Zoph et al. 2020). For the detection task, we employed the medium-sized YOLOv11 model (YOLOv11m), which offers a favorable trade-off between detection performance and speed. The model was initialized via transfer learning using weights pre-trained on the Common Objects in Context (COCO) dataset (Lin et al. 2014), a large-scale benchmark comprising natural images of 80 object categories that facilitates feature generalization across visual domains. The confidence

threshold for predictions was selected by maximizing the F_1 score on the validation set. Preliminary evaluation on a subset of the 2021 training data devoted for testing demonstrated the model's high sensitivity, yielding a recall of 92%, a precision of 75%, and a balanced F_1 score of 82%. Further methodological details, including training procedures, and hyperparameter optimisation are reported in Sangiovanni et al. (2025).

The trained model was applied to five orthophoto campaigns from the main study area. Detection performance was evaluated by comparing predicted individuals with manual annotations using precision, recall, and the F_1 score. The results are summarized in Table 1, while Fig. 3 illustrates the spatial distribution of manually annotated and detected individuals. Overall precision remained consistently high ($0.83 - 0.92$), indicating a low rate of false positives across campaigns. Recall showed greater variability, ranging from 0.70 in spring surveys to 0.39 in the late-autumn campaign, leading to F_1 scores between 0.80 and 0.54. This pattern suggests that the detector reliably identifies individuals when present but may miss a proportion of organisms under more challenging visual conditions. Detection performance was strongest in April ($F_1 \approx 0.80$), when water clarity and seabed visibility were optimal, and declined progressively toward autumn. The marked reduction in recall

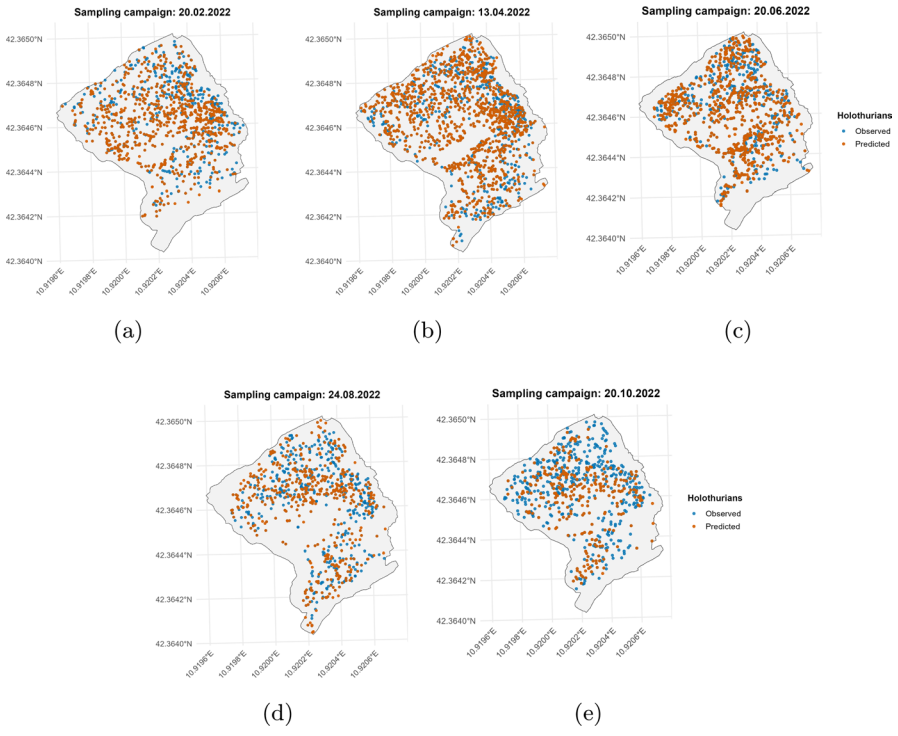

Fig. 3 Distribution of manually annotated and detected points by the object detector in the study area for the five different campaigns

Table 1 Detection performance across five monitoring campaigns

Campaign	Manually annotated/predicted counts by habitat					Performance metrics			
	P. Transpl.	Dead Matte	P. Oceanica	Hard	Sandy	Manual	Auto	P	R/F1
20/02/2022	35/23	189/108	54/35	90/56	616/488	985	721	0.900	0.659/0.761
13/04/2022	302/174	282/210	250/157	433/299	235/301	1479	1137	0.916	0.704/0.796
20/06/2022	20/8	227/163	39/30	101/55	562/440	938	672	0.880	0.638/0.743
24/08/2022	30/9	171/81	13/10	103/48	468/328	770	472	0.826	0.506/0.628
20/10/2022	15/2	151/52	21/10	96/23	388/197	668	284	0.908	0.386/0.542

Habitat-specific counts show manually annotated and predicted individuals. Performance metrics include precision, recall, and F_1-score

observed in October is likely linked to increased accumulation of detrital material, particularly dead *Posidonia* leaves deposited after late-summer storms, which reduced contrast between organisms and the surrounding substrate and made visual identification more difficult.

2.4 Data modeling

To connect the deep learning-based object detection with the statistical procedure, we treat the detections produced by YOLOv11m as a spatially biased sample of the true underlying sea cucumber distribution. Although the detector performs well when individuals are clearly visible, some organisms are systematically missed, while others may be falsely detected. In the point process literature, this is commonly referred to as point contamination, with the latter type known as ghost points (Guttorp et al. 2026). The statistical model therefore estimates the latent ecological intensity of sea cucumbers, correcting for detection bias driven by enviromental covariates that is only partially captured by the automated detector.

Define $\mathcal{D} \subset \mathbb{R}^2$ as the study area encompassing all sampled habitats. For each monitoring campaign ($t = 1, \dots, 5$), the detector produced $\mathbf{U}_t^* = \{\mathbf{u}_{1t}^*, \dots, \mathbf{u}_{n_t t}^*\} \subset \mathcal{D}$, corresponding to the centroids of predicted bounding boxes, where n_t is the number of detected individuals. In parallel, manually annotated positions $\mathbf{U}_t = \{\mathbf{u}_{1t}, \dots, \mathbf{u}_{m_t t}\} \subset \mathcal{D}$ were available, with m_t denoting the number of verified individuals. As shown in Table 1, the automated system exhibits high precision but moderate recall ($n_t < m_t$ for all t), reflecting systematic undercounting. This conservative detection strategy prevents ghost points from introducing spurious spatial patterns, which cannot be easily distinguished from ecological signals.

To correct for this bias and recover the latent spatial distribution of sea cucumbers, we conceptualize the detected pattern $\mathbf{U}_t^*$ as a thinned realization of the underlying ecological process for each campaign t. Specifically, each true individual located at $\mathbf{s} \in \mathcal{D}$ is independently detected with a spatially varying detection probability $p_t(\mathbf{s}) \in [0, 1]$, producing an observed thinned process (Dorazio 2014). Under this framework, $\mathbf{U}_t^*$ is modeled as a realization of a thinned LGCP (Møller et al. 1998):

$$\begin{aligned} \mathbf{U}_t^* \mid \lambda_t^*(\mathbf{s}) &\sim \mathrm{PP}(\lambda_t^*(\mathbf{s})) \\ \lambda_t^*(\mathbf{s}) &= \lambda_t^{\mathrm{pot}}(\mathbf{s})\, p_t(\mathbf{s}) \\ \log(\lambda_t^{\mathrm{pot}}(\mathbf{s})) &= \mu_t + \mathbf{x}(\mathbf{s})^\top \boldsymbol{\beta} + w_l(\mathbf{s}) \\ w_l(\mathbf{s}) &\sim \mathrm{GP}\big(0, C(\cdot;\sigma_l^2, \rho_l)\big), \quad l = 1, 2. \end{aligned} \tag{1}$$

Here, $\lambda_t^{\mathrm{pot}}(\mathbf{s})$ denotes the potential intensity, representing the latent ecological distribution of sea cucumbers in the absence of detection bias (Chakraborty et al. 2011; Warton and Shepherd 2010). The parameter μ_t denotes a campaign-specific intercept, $\mathbf{x}(\mathbf{s})$ is a vector of spatial covariates, and $w_l(\mathbf{s})$ is a Gaussian Process. The index l equals 1 for the first three campaigns and 2 for the remaining ones. This specification reflects the assumption that campaigns conducted under comparable survey conditions share a similar residual spatial dependence structure. In particular, the last two campaigns were carried out under more similar environmental and operational settings, whereas the earlier campaigns were conducted under a different set of conditions. Differences in overall abundance across campaigns are instead captured by the campaign-specific intercepts μ_t. The observed intensity $\lambda_t^*(\mathbf{s})$ thus represents a filtered version of this potential intensity, modulated by the spatially varying detection probability $p_t(\mathbf{s})$. Several formulations of $p_t(\mathbf{s})$ are possible, including detection functions inspired by distance sampling theory (Martino et al. 2021; Yuan et al. 2017). In this study, we adopt a half-normal detection function:

$$p_t(\mathbf{s}) = \exp\left(-\frac{(z_t(\mathbf{s}))^2}{2\tau^2}\right), \tag{2}$$

where $z_t(\mathbf{s})$ is a covariate affecting detectability and τ is a scale parameter controlling the rate of detection decay. This formulation assumes maximum detection when the covariate equals zero, with detection probability decreasing as the covariate deviates from this reference value. A key advantage of this framework is its flexibility in accommodating multiple sources of detection heterogeneity. The detection probability can be extended by defining $p_t(\mathbf{s})$ as a product of independent half-normal components:

$$p_t(\mathbf{s}) = \prod_{k=1}^{K} \exp\left(-\frac{(z_{t,k}(\mathbf{s}))^2}{2\tau_k^2}\right), \tag{3}$$

where K represents the number of covariates employed. Crucially, identifiability issues arise because both the latent ecological intensity and the detection process contribute multiplicatively to the observed point pattern, making it impossible to uniquely attribute variation in the data to one component or the other without additional structure. To address this issue, as discussed in Fithian and Hastie (2012), we specify the ecological covariates $\mathbf{x}(\mathbf{s})$ and the detection covariates $z_t(\mathbf{s})$ so that they represent distinct mechanisms and avoid linear dependence, thereby reducing potential confounding between the ecological and observation processes. Furthermore, we do not include an intercept term in the thinning function. Such a term would not be identifiable, as it would be confounded with the overall scale of the latent

intensity, leading to a non-unique decomposition of the observed process. Detection covariates were transformed prior to model fitting so that the half-normal detection function attains its maximum at $z_t(\mathbf{s}) = 0$, corresponding to baseline detection conditions. When variables were not naturally expressed relative to such a baseline, appropriate scaling (min-max normalization) and, where needed, monotone transformations (including complements) were applied to ensure a consistent interpretation of detection effects across campaigns. Finally, detection covariates were also checked to avoid strong pairwise correlations, thereby limiting multicollinearity within the detection component. Critically, the detection mechanism functions as a spatial filter that preserves the underlying ecological point pattern structure while selectively revealing individuals as a function of covariates.

It is important to specify that $\mathbf{U}_t$ and $\mathbf{U}_t^*$ are used in separate model-fitting procedures. The manually annotated data $\mathbf{U}_t$ serve to fit a baseline model representing the "true" ecological state, whereas the automated detections $\mathbf{U}_t^*$ are used to fit the imperfect detection framework. Furthermore, the ability of the model to mitigate under-detection relies on operating conditions under which the thinning representation remains informative. In particular, reliable recovery of the potential intensity requires that the level of thinning is not extreme, so that sufficient spatial signal is retained in the detected pattern, and the detection probability is driven by covariates that meaningfully capture the main sources of detectability variation.

2.5 Model estimation

Model estimation was performed using the R-INLA package (Rue et al. 2009), via the *inlabru* interface (Bachl et al. 2019). INLA provides a computationally efficient framework for approximate Bayesian inference in Latent Gaussian Models, characterized by Gaussian latent fields governed by a limited set of hyperparameters, and non-Gaussian likelihoods. The approach relies on deterministic Laplace approximations to estimate posterior marginals, offering substantial computational advantages over traditional Markov Chain Monte Carlo (MCMC) methods.

Weakly informative Gaussian priors were assigned to the intercept parameters μ_t and regression coefficients $\boldsymbol{\beta}$, both centered at zero with precision 0.001. For the spatial Gaussian processes, penalized complexity (PC) priors (Simpson et al. 2017) were specified for the marginal standard deviation σ_l and range ρ_l (expressed in meters). The study domain covers approximately 5500 m^2 ($\approx 70 \times 80\ m$), so we set $\mathbb{P}(\rho_l < 50) = 0.5$ and $\mathbb{P}(\sigma_l > 0.5) = 0.01$ for both spatial components, reflecting the spatial scale and variability expected within the study domain. Finally, the scale parameters of the detection functions were modeled through a non-informative log-Gaussian prior to mitigate numerical instabilities observed during estimation.

2.6 Model comparison

Model selection in a Bayesian context requires evaluation criteria that appropriately balance predictive accuracy and model complexity. Although traditional measures such as the Deviance Information Criterion (DIC) are widely used,

they exhibit well-documented limitations, particularly for hierarchical or spatially structured models (Gelfand and Schliep 2018; Leininger and Gelfand 2017). We therefore adopted a predictive evaluation framework grounded in the theory of innovation processes for spatial point patterns (Baddeley et al. 2005), complemented by proper scoring rules such as the Continuous Ranked Probability Score (CRPS) (Matheson and Winkler 1976). To account for spatial heterogeneity and temporal variation, the study region was partitioned into bounded subdomains $B_1, \ldots, B_G$. This spatial blocking enables localized residual analysis, providing more detailed model diagnostics across heterogeneous habitats. For each campaign t, let $\mathbf{O}_t \mid \lambda_t$ denote a non-homogeneous Poisson process with intensity $\lambda_t(\mathbf{o})$ defined over the spatial domain $\mathcal{D}$. A realization $\mathbf{O}_t$ can be partitioned into $\mathbf{o}_t^1, \ldots, \mathbf{o}_t^G$, corresponding to observations within each subregion. Given an innovation function $h_t(\cdot)$, the innovation process in the g-th subset at time t is defined as

$$R_h^t(B_g) = \sum_{\mathbf{o} \in \mathbf{o}_t^g} h_t(\mathbf{o}) - \int_{B_g} h(\mathbf{o})\, \lambda_t(\mathbf{o})\, d\mathbf{o}. \tag{4}$$

Setting $h_t(\mathbf{o}) = 1$ yields the raw residuals:

$$R_{\text{raw}}^t(B_g) = N(B_g) - \int_{B_g} \lambda(\mathbf{o})\, d\mathbf{o}, \tag{5}$$

where $N(B_g)$ denotes the observed number of points within B_g. Raw residuals measure the difference between observed and expected point counts but may be dominated by regions of high intensity, thereby masking potential model deficiencies in sparser areas. To mitigate this imbalance, we computed Pearson residuals by setting $h_t(\mathbf{o}) = 1/\sqrt{\lambda_t(\mathbf{o})}$:

$$R_{\text{pearson}}^t(B_g) = \sum_{\mathbf{u} \in \mathbf{u}_t^g} \frac{1}{\sqrt{\lambda_t(\mathbf{u})}} - \int_{B_g} \sqrt{\lambda_t(\mathbf{u})}\, d\mathbf{u}. \tag{6}$$

This normalization stabilizes the variance by weighting each observation inversely to its expected intensity, reducing heteroscedasticity and allowing for balanced residual diagnostics across regions of varying density. Integrals of the intensity function were approximated using Monte Carlo quadrature (Berman and Turner 1992). Let $\{c_i \in \mathcal{D}\}_{i=1}^N$ denote quadrature nodes and $\{w_i\}_{i=1}^N$ their associated weights. The integral of the intensity over $\mathcal{D}$ was then approximated as

$$\int_{\mathcal{D}} \lambda_t(\mathbf{o})\, do \approx \sum_{i=1}^{N} w_i\, \lambda_t(c_i), \tag{7}$$

and this computation was repeated across M posterior samples to characterize the uncertainty in residual-based diagnostics. To complement residual analysis, we employed the CRPS as a proper scoring rule for evaluating predictive distributions. Under correct model specification, residuals should have a zero expectation

(Baddeley et al. 2005). Denoting by $F^g_{R_{\text{raw},t}}$ and $F^g_{R_{\text{pearson},t}}$ the predictive distributions of raw and Pearson residuals, respectively, the CRPS is defined as

$$\text{CRPS}(F^g_{R_{\text{raw},t}}, 0) = \int_{\mathbb{R}} (F^g_{R_{\text{raw},t}}(x) - \mathbf{1}\{x \geq 0\})^2 \, dx, \tag{8}$$

$$\text{CRPS}(F^g_{R_{\text{pearson},t}}, 0) = \int_{\mathbb{R}} (F^g_{R_{\text{pearson},t}}(x) - \mathbf{1}\{x \geq 0\})^2 \, dx. \tag{9}$$

Conceptually, this predictive framework addresses the challenge of identifying where and why a spatial model may be failing across a complex landscape. By partitioning the study area into blocks and applying innovation processes, we can pinpoint specific regions where the model over (or under) -predicts. The CRPS offers distinct advantages over information criteria such as DIC or BIC, as it assesses the entire predictive distribution rather than a single summary statistic. This makes it particularly appropriate for spatial point process models with hierarchical structures, where predictive calibration and uncertainty quantification are of primary interest. A lower CRPS indicates a model that is better calibrated, meaning its predictions are centered on the observed value.

3 Results

Because missed detections cannot be directly observed, their treatment requires explicit assumptions regarding their dependence on observable covariates. We hypothesize that the probability of missed detections is not uniformly distributed across space but varies systematically according to three factors. First, we expect missed detections to be more frequent in areas of high local density, where individuals may overlap or occlude one another. To capture this effect, we define the local frequency (f_r) as the number of predicted points within a circle of radius r centered on each detection. The choice of the spatial scale r was guided by the empirical spatial structure of the data. To select an appropriate radius, we examined the radius at which most detections have at least one neighbor and found that $r = 1$ meter ensures approximately 85% coverage. This radius therefore represents the smallest spatial scale at which crowding effects are consistently observable, providing a pragmatic balance between capturing local interactions and avoiding overly sparse neighborhoods. Second, we anticipate that the detection probability is related to the confidence score (CS) assigned by the neural network detector, which ranges from 0 to 1. Individuals with lower confidence scores are more likely to represent marginal detections in regions of poor algorithm performance. Third, we expect a relationship with object size, approximated by the diagonal length (DL) of each bounding box, also ranging from 0 to 1. Smaller individuals are inherently more difficult to detect. To incorporate these covariates into the half-normal detection function, we applied transformations so that each variable acts as a monotonic modifier of detectability. All covariates were

first rescaled to the unit interval using min-max normalization. We then used the complement to one for the confidence score and bounding box diagonal, and the reciprocal transformation for local frequency, so that larger transformed values consistently correspond to lower detectability. Preliminary exploratory analyses (Sangiovanni et al. 2025) indicated that detection errors were approximately evenly distributed across habitat categories, suggesting that habitat type does not systematically affect detectability and therefore was not included in the thinning function to avoid unnecessary model complexity.

Across all models, the potential intensity function was specified using the same set of covariates, namely the percentage cover of *P. Oceanica*, the slope (standardized) of the seafloor and a Gaussian process component to capture residual spatial variation. Model 1, fitted to the manually annotated locations, is included as a benchmark representing the closest available proxy to the true point pattern. Its role is therefore to provide a reference for assessing the bias induced by missed detections and to evaluate the extent to which the thinning approach can recover the underlying ecological signal. Model 2 was fitted to the detected locations without correction, while Model 3 comprises a set of models incorporating the thinning mechanism applied to the detected subset. These included specifications using f_r, CS, and DL individually, as well as their combinations. Residuals were computed on a localized 18×18 grid superimposed on the study area, following the procedure described in Subsection 2.6. Table 2 reports CRPS values for the baseline models and the best-performing thinning models, using both Pearson and raw residuals. Lower CRPS values indicate better predictive performance and should therefore be interpreted as improvements in model fit.

Table 2 Comparison of methods across five campaigns using raw residuals (lower table) and Pearson residuals (upper table)

Method	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	Average
Model 1	4.356	4.441	4.536	3.369	3.250	3.990
Model 2	5.867	6.929	5.912	6.246	8.475	6.694
Model 3 diag	5.558	6.444	5.551	5.998	7.948	6.303
Model 3 freq	5.486	6.383	5.480	5.937	7.867	6.230
Model 3 conf	5.484	6.376	5.505	5.914	7.791	6.214
Model 3 conf × freq	5.420	6.313	5.438	5.864	7.696	**6.150**
Method	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	Average
Model 1	1.615	1.905	1.839	1.152	0.998	1.500
Model 2	1.964	2.685	2.101	1.803	1.911	2.092
Model 3 diag	1.887	2.540	2.021	1.749	1.837	2.010
Model 3 freq	1.865	2.515	1.997	1.740	1.825	1.999
Model 3 conf	1.864	2.516	2.004	1.734	1.816	1.992
Model 3 conf × freq	1.852	2.493	1.987	1.724	1.802	**1.979**

Model 1 is the model fitted on the manually annotated locations, while Model 2 is the model fitted on the detected locations without thinning and Model 3 the model fitted on the detected locations with a different thinning mechanism

Several patterns emerge from this comparison. First, Model 2 shows substantially worse performance than Model 1, with average Pearson residuals of 6.694 compared to 3.990 and raw residual of 2.092 compared to 1.500. This confirms that ignoring missed detections in the ground truth introduces considerable bias in spatial intensity estimation. Second, all thinning models systematically improve performance relative to Model 2. The improvement is consistent across all five temporal campaigns, suggesting that the thinning mechanism successfully corrects for spatially structured missed detections. Third, among the tested thinning specifications, the model incorporating both local frequency and confidence score through a product of half-normal functions (Model 3 conf $\times$ freq) achieves the best performance. This suggests that missed detections are most strongly associated with local crowding and marginal detections, as indicated by low confidence scores. However, none of the thinning models fully recovers the performance of Model 1. Nevertheless, the substantial improvement over Model 2 demonstrates that the thinning framework provides a practical method for reducing bias when ground truth data contain missed detections.

To assess whether the thinning approach recovers the underlying ecological relationships, we compared the posterior distributions of the intensity function parameters across Model 1, Model 2, and the best thinning model. Figure 4 displays the posterior means and 95% credible intervals for the seasonal intercepts, the environmental covariates and the scale parameters of the half-normal. For the seasonal intercepts, Model 3 shows closer agreement with Model 1 than does Model 2. This suggests that accounting for missed detections through thinning partially recovers the true temporal variation in population intensity. For the environmental covariates, the posterior distributions are similar across Model 2 and Model 3. This indicates that while missed detections bias the absolute intensity

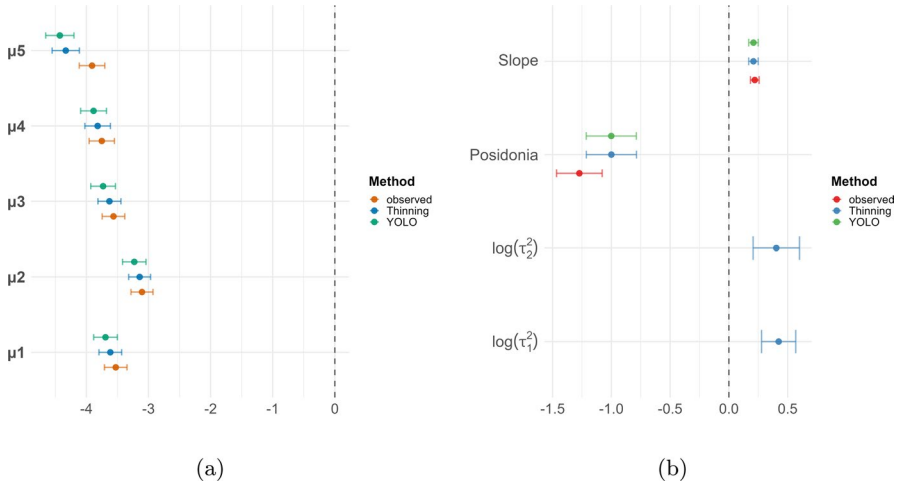

Fig. 4 Posterior distributions of the parameters representing mean and 95% credible intervals involved in the model fitted on the manually annotated locations, the model fitted on the detected locations without and with (best model) thinning

estimates (captured by the intercepts), they have less impact on the relative effects of environmental covariates.

Overall, the thinning correction mainly alters the magnitude of estimated intensities while leaving the spatial allocation of ecological effects largely unchanged. The posterior distributions of the thinning scale parameters indicate moderate thinning, with comparable magnitudes suggesting that local crowding and detection confidence contribute approximately equally to the spatial pattern of missed detections.

4 Discussion

4.1 Thinning function and alternative formulations

The chosen thinning function provides a robust and tractable mechanism for correcting missed detections. Our framework is fundamentally grounded on the premise that the data are a representative sample of the true population. This conservative stance underpins our decision to avoid complex measurement error models (Chakraborty and Gelfand 2010), as our data do not exhibit the spatial shifts or misalignments that would justify such additional complexity. Alternative formulations were considered. A "pogit" specification (Dorazio 2014; Winkelmann and Zimmermann 1993) was explored, but identifiability constraints (Fithian and Hastie 2012) and preliminary results indicated excessive thinning and poorer fit, while adding a Gaussian process to represent near-absence (Guttorp et al. 2026) proved weakly identified without auxiliary data to inform this latent field (Jones-Todd et al. 2018). Finally, non-parametric thinning approaches (Arima et al. 2024) are conceptually appealing but currently require further development for continuous spatial point process settings and, in our view, cannot yet be naturally embedded within a latent Gaussian modeling framework. Ultimately, the simpler parametric specification is preferable for real-world ecological data. It remains identifiable under standard survey conditions and provides a reliable correction for missed detections.

4.2 Limitations and identifiability

A critical practical consideration is the inherent asymmetry in handling detection errors. Our methodology is designed to be robust against ghost points by prioritizing a conservative detection threshold, yet this architectural choice renders the system highly sensitive to missed detections. The framework operates reliably when ghost points are rare and missed detections follow patterns explained by environmental covariates. Conversely, the model reaches a functional failure point if missed detection rates exceed a critical threshold or if the occurrence mechanisms are stochastic and decoupled from observable data. Furthermore, identifying which environmental variables are most relevant for explaining detection failures a priori remains a formidable challenge, suggesting that the model's reliability is ultimately contingent on the availability of informative, spatially-indexed covariates to explain the detection gap. To specifically address the limitations linked to error asymmetry, latent variable

formulations, akin to those used in degraded point pattern analysis (Chakraborty et al. 2011), may provide a future pathway for incorporating unobserved events.

4.3 Uncertainty quantification and model selection

The methodology also carries significant implications for uncertainty and model selection, representing a departure from traditional approaches that combine detection and spatial processes in a single modeling stage. Separating detection uncertainty from spatial modeling allows for a clearer delineation between observational error and underlying ecological structure. However, because the outputs of the object detector are incorporated as deterministic inputs, uncertainty is not formally propagated across stages; instead, variability is only partially captured through the LGCP in the second step. Future work could leverage conformal prediction (Vovk et al. 2022; Deliu and Liseo 2025) or related methods to define more informative priors on detection function parameter and improve uncertainty quantification. From a model selection perspective, standard cross-validation (CV) (Cronie et al. 2024; Poggio et al. 2026) is inherently problematic for thinned point processes. Since the true point pattern is unobserved, evaluating predictions against detection-based data introduces circularity. In this particular case, practitioners should therefore prioritize residual-based diagnostics on the training data. While not a complete solution to the model selection challenge, it provides a more pragmatic and theoretically grounded means of comparing competing specifications in imperfectly detected systems.

4.4 Ecological relevance

From an ecological perspective, accurately identifying the spatial distribution of benthic populations is essential for understanding resilience and informing management. Ignoring detection bias in these systems can lead to significantly flawed conclusions regarding population health and habitat suitability (Royle and Link 2006; Monk 2014). In this work, sea urchins are mentioned only as an illustrative example of a potential future application of the proposed framework and were not analyzed here. Previous work based on in-field visual census data (Addis et al. 2012) has shown that sea urchin populations often display strong small-scale autocorrelation and patchy distribution patterns. Consequently, a spatial model's capacity to estimate detection probability while concurrently recovering the actual underlying spatial intensity and assessing its relationship with environmental covariates can serve as a significant resource for monitoring benthic communities. The increasing reliance on automated image-based surveys and underwater photogrammetry amplifies this challenge. Despite this, recent efforts to map the densities of sea urchins (Piazza et al. 2019; Sastraantara et al. 2024), macroalgae canopy, and the coverage of other benthic species (Spyksma et al. 2022) have alleviated the constraints of monitoring budgets and timelines by delivering cost-effective, high-quality data across significantly expanded temporal and spatial scales. Nevertheless, several challenges persist in the rapid assessment of spatial patterns. Indeed, the integration of underwater photogrammetry with unmanned

platforms, such as unmanned aerial vehicles (UAV), for shallow water applications (Sugara et al. 2025) frequently results in extensive imagery datasets that must be analyzed using deep learning-based detectors, which may inadvertently introduce bias in detection.

In summary, our approach balances tractability, interpretability and methodological rigor, while acknowledging inherent limitations. The obtained results demonstrate that missed detections in manual annotations can substantially bias spatial intensity estimates when using deep learning-based object detection systems. While the correction does not fully eliminate the bias introduced by imperfect manual annotations, it represents a considerable improvement over naive approaches that ignore detection errors in training data. From an applied perspective, this translates into more robust monitoring signals, improving the reliability of trend estimation and spatial comparisons in ecological surveys. The practical value of this work lies in its ability to provide more credible assessments of restoration efforts and management interventions. At the same time, it supports the deployment of scalable, image-based benthic monitoring workflows without relying on unrealistic perfect-detection assumptions. Our formulation represents a pragmatic compromise tailored to the available data, with several extensions left for future work.

5 Supplementary Information

The Supplementary Information provides the posterior mean of the thinning mechanism and a comparison of the proposed approach with alternative methods under a different specification of the thinning function.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10651-026-00732-7.

Acknowledgments The authors would like to thank Daniele Poggio for some stimulating conversation on the topic. They are also grateful to the two anonymous reviewers for their constructive comments and valuable suggestions, which helped improve the quality of the manuscript.

Author contributions Conceptualization: [G.M.S, G.J.L]; Methodology: [G.M.S, G.J.L, G.M]; Fundings: [D.V]; Formal analysis and investigation: [G.M.S, G.M]; Writing—original draft preparation: [G.M.S, A.P]; Writing—review and editing: [All]; Supervision: [G.J.L, G.M, A.P].

Funding Open access funding provided by Università degli Studi di Roma La Sapienza within the CRUI-CARE Agreement. The corresponding author received funding from Sapienza University of Rome under the "Progetti per Avvio alla Ricerca - Tipo 1" scheme (Grant No. AR125199C28815FC), for the project entitled "Integration of Object Detection and Point Process Models for Marine Monitoring".

Data availability Data for this paper is available on https://github.com/justgimmi/Integrating-Deep-Learning-and-Spatial-Statistics-in-Marine-Ecosystem-Monitoring.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Aszkowski P, Ptak B, Kraft M, Pieczyński D, Drapikowski P (2023) Deepness: deep neural remote sensing plugin for QGIS. SoftwareX 23:101495

Arima S, Polettini S, Pasculli G, Gesualdo L, Pesce F, Procaccini D-A (2024) A bayesian nonparametric approach to correct for underreporting in count data. Biostatistics 25(3):904–918

Addis P, Secci M, Angioni A, Cau A (2012) Spatial distribution patterns and population structure of the sea urchin *Paracentrotus lividus* (Echinodermata: Echinoidea), in the coastal fishery of Western Sardinia: a geostatistical analysis. Sci Mar 76(4):733–740

Alexandrova S, Tatlock Z, Cakmak M (2015) Roboflow: A flow-based visual programming language for mobile manipulation tasks. In: 2015 IEEE International Conference on Robotics and Automation (ICRA), 5537–5544 . IEEE

Beijbom O, Edmunds PJ, Kline DI, Mitchell BG, Kriegman D (2012) Automated annotation of coral reef survey images. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, 1170–1177 . IEEE

Bachl FE, Lindgren F, Borchers DL, Illian JB (2019) Inlabru: an R package for bayesian spatial modelling from ecological survey data. Methods Ecol Evol 10(6):760–766

Berman M, Turner TR (1992) Approximating point process likelihoods with GLIM. J Roy Stat Soc: Ser C (Appl Stat) 41(1):31–38

Baddeley A, Turner R, Møller J, Hazelton M (2005) Residual analysis for spatial point processes (with discussion). J R Stat Soc Ser B Stat Methodol 67(5):617–666

Chakraborty A, Gelfand AE (2010) Analyzing spatial point patterns subject to measurement error. Bayesian Analysis. 5:97

Chakraborty A, Gelfand AE, Wilson AM, Latimer AM, Silander JA (2011) Point pattern modelling for degraded presence-only data over large regions. J R Stat Soc Ser C Appl Stat 60(5):757–776

Cronie O, Moradi M, Biscio CA (2024) A cross-validation-based statistical theory for point processes. Biometrika 111(2):625–641

Ciriminna L, Signa G, Cilluffo G, Rakaj A, Vizzini S (2024) Aquaculture of emerging species in North-Eastern Atlantic and Mediterranean Sea: a systematic review on sea cucumber farming and potential development. Front Mar Sci 11:1381836

Casoli E, Ventura D, Cutroneo L, Capello M, Jona-Lasinio G, Rinaldi R, Criscoli A, Belluscio A, Ardizzone G (2017) Assessment of the impact of salvaging the Costa Concordia wreck on the deep coralligenous habitats. Ecol Ind 80:124–134

De Iaco S, Hristopulos D, Lin G (2022) Special issue: geostatistics and machine learning. Math Geosci 54:459–465

Deliu N, Liseo B (2025) The interplay between bayesian inference and conformal prediction. In Objective Bayes Methodology Conference. 9

Dorazio RM (2014) Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. Glob Ecol Biogeogr 23(12):1472–1484

Descombes X, Zerubia J (2002) Marked point process in image analysis. IEEE Signal Process Mag 19(5):77–84

Fithian W, Hastie T (2012) Finite-sample equivalence in statistical models for presence-only data. Ann Appl Stat 7(4):1917

Fallati L, Panieri G, Argentino C, Varzi AG, Bünz S, Savini A (2024) Combining rov-based acoustic data and underwater photogrammetry to characterize hakon mosby mud volcano (barents sea) cold seep systems. In: EGU General Assembly Conference Abstracts, 10620

Goodwin M, Halvorsen KT, Jiao L, Knausgård KM, Martin AH, Moyano M, Oomen RA, Rasmussen JH, Sørdalen TK, Thorbjørnsen SH (2022) Unlocking the potential of deep learning for marine ecology: overview, applications, and outlook. ICES J Mar Sci 79(2):319–336

Guttorp P, Illian J, Kostensalo J, Kuronen M, Myllymäki M, Särkkä A, Thorarinsdottir TL (2026) What you see is not what is there: mechanisms, models and methods for point pattern deviations. Stat Sci 41(1):143–162

Gelfand AE, Schliep EM (2018) Bayesian inference and computing for spatial point patterns. In: NSF-CBMS Regional Conference Series in Probability and Statistics, 10:125. JSTOR

Jones-Todd CM, Swallow B, Illian JB, Toms M (2018) A spatiotemporal multispecies model of a semicontinuous response. J R Stat Soc C Appl Stat 67(3):705–722

Khanam R, Hussain M (2024) Yolov11: An overview of the key architectural enhancements. arXiv preprint arXiv:2410.17725

Leininger TJ, Gelfand AE (2017) Bayesian inference and model assessment for spatial point patterns using posterior predictive samples. Bayesian Anal 12(1):1–30

Lewy P, Kristensen K (2009) Modelling the distribution of fish accounting for spatial correlation and overdispersion. Can J Fish Aquat Sci 66(10):1809–1820

Lopez GR, Levinton JS (1987) Ecology of deposit-feeding animals in marine sediments. Q Rev Biol 62(3):235–260

Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL (2014) Microsoft coco: Common objects in context. European Conference on Computer Vision. Springer, Berlin, pp 740–755

Mahmood A, Bennamoun M, An S, Sohel F, Boussaid F, Hovey R, Kendrick G, Fisher R (2016) Automatic annotation of coral reefs using deep learning. In: Oceans 2016 mts/IEEE Monterey, 1–5 . IEEE

Mateu J, Jalilian A (2022) Spatial point processes and neural networks: a convenient couple. Spat Stat 50:100644

MacKenzie DI, Nichols JD, Lachman GB, Droege S, Royle AJ, Langtimm CA (2002) Estimating site occupancy rates when detection probabilities are less than one. Ecology 83(8):2248–2255

Monk J (2014) How long should we ignore imperfect detection of species in the marine environment when modelling their distribution? Fish Fish 15(2):352–358

Mabon J, Ortner M, Zerubia J (2022) Point process and cnn for small object detection in satellite images. In: SPIE, Image and Signal Processing for Remote Sensing XXVIII

Mabon J, Ortner M, Zerubia J (2023) Learning point process models for vehicles detection using cnns in satellite images. In: 2023 17th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), 86–92. IEEE

Mabon J, Ortner M, Zerubia J (2024) Learning and scoring point process models for object detection in satellite images. In: 2024 32nd European Signal Processing Conference (EUSIPCO), 1771–1775 . IEEE

Martino S, Pace DS, Moro S, Casoli E, Ventura D, Frachea A, Silvestri M, Arcangeli A, Giacomini G, Ardizzone G et al (2021) Integration of presence-only data from several sources: a case study on dolphins' spatial distribution. Ecography 44(10):1533–1543

Møller J, Syversveen AR, Waagepetersen RP (1998) Log Gaussian Cox processes. Scand J Stat 25(3):451–482

Mancini G, Ventura D, Casoli E, Belluscio A, Ardizzone G (2022) Transplantation on a *Posidonia oceanica* meadow to facilitate its recovery after the Concordia shipwrecking. Mar Pollut Bull 179:113683

Mastrantonio G, Ventura D, Casoli E, Rakaj A, Jona Lasinio G, Poggio D, Vitiello C, Calculli C (2024) Species distribution models with masking: The case of holothurians in a posidonia rich area. Scientific Meeting of the Italian Statistical Society. Springer, Berlin, pp 531–536

Matheson JE, Winkler RL (1976) Scoring rules for continuous probability distributions. Manage Sci 22(10):1087–1096

Piazza P, Cummings V, Guzzi A, Hawes I, Lohrer A, Marini S, Marriott P, Menna F, Nocerino E, Peirano A et al (2019) Underwater photogrammetry in Antarctica: long-term observations in benthic ecosystems and legacy data rescue. Polar Biol 42(6):1061–1079

Purcell SW, Conand C, Uthicke S, Byrne M (2016) Ecological roles of exploited sea cucumbers. Oceanography and Marine Biology. CRC Press, USA, pp 375–394

Pham TT, Hamid Rezatofighi S, Reid I, Chin T-J (2016) Efficient point process inference for large-scale object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2837–2845

Panunzi G, Moro S, Marques I, Martino S, Colloca F, Ferretti F, Jona Lasinio G (2025) Estimating the spatial distribution of the white shark in the Mediterranean Sea via an integrated species distribution model accounting for physical barriers. Environmetrics 36(1):2876

Poggio D, Sangiovanni GM, Mastrantonio G et al (2026) Modeling benthic animals in space and time using Bayesian point process with cross validation: the case of holoturians. Environmetrics 37(3):70096. https://doi.org/10.1002/env.70096

Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 779–788

Royle JA, Link WA (2006) Generalized site occupancy models allowing for false positive and false negative errors. Ecology 87(4):835–841

Rue H, Martino S, Chopin N (2009) Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. Journal of the Royal Statistical Society Series B (Statistical Methodology) 71(2):319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x (https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2008.00700.x)

Robinson GS (1977) Edge detection by compass gradient masks. Comput Graphics Image Process 6(5):492–501

Sastraantara MS, Agus SB, Susilo SB (2024) Mapping the distribution of sea urchin (echinoidea) and benthic habitat using drone in the waters of lancang island. In: BIO Web of Conferences, vol. 106, 04006. EDP Sciences

Sangiovanni GM, Jona Lasinio G, Poggio D, Mastrantonio G, Pollice A, Rakaj A, Mohan M, Moro S, Casoli E, Ventura D (2025) Automated object detection based on YOLOv11 for monitoring benthic population dynamics: a new approach combining photogrammetry and open-source GIS tools applied to sea cucumbers. Accepted for Publication in *Remote Sensing in Ecology and Conservation*

Spyksma AJ, Miller KI, Shears NT (2022) Diver-generated photomosaics as a tool for monitoring temperate rocky reef ecosystems. Front Mar Sci 9:953191

Simpson D, Rue H, Riebler A, Martins TG, Sørbye SH (2017) Penalising model component complexity: a principled, practical approach to constructing priors. Stat Sci 32(1):1–28

Schneider K, Silverman J, Kravitz B, Rivlin T, Schneider-Mor A, Barbosa S, Byrne M, Caldeira K (2013) Inorganic carbon turnover caused by digestion of carbonate sands and metabolic activity of holothurians. Estuar Coast Shelf Sci 133:217–223

Sugara A, Sari YP, Sangadji MS, Rudiastuti AW, Pianto TA, Siregar VP, Agus SB, Sutrisno D (2025) Detection of sea urchins (diadema setosum) colony distribution on uav imagery data using fuzzy logic approach. Kuwait Journal of Science 52(2):100389

Ünel FÖ, Özkalayci BO, Çiğla C (2019) The power of tiling for small object detection. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 582–591 . IEEE

Ventura D, Castoro L, Mancini G, Casoli E, Pace DS, Belluscio A, Ardizzone G (2022) High spatial resolution underwater data for mapping seagrass transplantation: A powerful tool for visualization and analysis. Data Brief 40:107735

Vovk V, Gammerman A, Shafer G (2022) Algorithmic learning in a random world. Springer Books

Ventura D, Rakaj A, Lasinio GJ, Sangiovanni GM, Poggio D, Mastrantonio G, Pollice A, Grasso G, Casoli E, Mancini G et al (2025) Detecting habitat preferences and monitoring population dynamics of sea cucumbers in coastal ecosystems through underwater photogrammetry. J Environ Manage 377:124589

Weinstein BG (2018) A computer vision for animal ecology. J Anim Ecol 87(3):533–545

Warton DI, Renner IW, Ramp D (2013) Model-based control of observer bias for the analysis of presence-only data in ecology. PLoS ONE 8(11):79168

Warton DI, Shepherd LC (2010) Poisson point process models solve the" pseudo-absence problem" for presence-only data in ecology. Ann Appl Stat. https://doi.org/10.1214/10-aoas331

Winkelmann R, Zimmermann KF (1993) Poisson-logistic Regression. Volkswirtschaftl. Fakultät d. Ludwig-Maximilians-Univ. München

Yuan Y, Bachl F, Lindgren F, Borchers DL, Illian J, Buckland S, Rue H, Gerrodette T (2017) Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. Ann Appl Stat 11(4):2270–2297

Zoph B, Cubuk ED, Ghiasi G, Lin T-Y, Shlens J, Le QV (2020) Learning data augmentation strategies for object detection. European Conference on Computer Vision. Springer, Berlin, pp 566–583
Zhai X, Wei H, He Y, Shang Y, Liu C (2022) Underwater sea cucumber identification based on improved yolov5. Appl Sci 12(18):9105

Authors and Affiliations

Gian Mario Sangiovanni[1] · Gianluca Mastrantonio[2] · Alessio Pollice[3] · Daniele Ventura[4] · Giovanna Jona Lasinio[1]

✉ Gian Mario Sangiovanni
gianmario.sangiovanni@uniroma1.it

Gianluca Mastrantonio
gianluca.mastrantonio@polito.it

Alessio Pollice
alessio.pollice@uniba.it

Daniele Ventura
daniele.ventura@uniroma1.it

Giovanna Jona Lasinio
giovanna.jonalasinio@uniroma1.it

[1] Department of Statistical Sciences, University of Rome 'La Sapienza', P.le Aldo Moro 5, Rome 00185, Italy

[2] Department of Mathematical Sciences 'G.L.Lagrange', Polytechnic of Turin, Corso Duca degli Abruzzi 24, Torino 10129, Italy

[3] Department of Economics and Finance, Universita degli Studi di Bari Aldo Moro, Largo Abbazia Santa Scolastica, Bari 70124, Italy

[4] Department of Environmental Biology, University of Rome 'La Sapienza', P.le Aldo Moro 5, Roma 00185, Italy