POLITECNICO DI TORINO
Repository ISTITUZIONALE

Modeling Benthic Animals in Space and Time Using Bayesian Point Process With Cross Validation: The Case of Holoturians

*Original*
Modeling Benthic Animals in Space and Time Using Bayesian Point Process With Cross Validation: The Case of Holoturians / Poggio, D., Mario Sangiovanni, G., Mastrantonio, G., Jona Lasinio, G., Casoli, E., Moro, S., Ventura, D.. - In: ENVIRONMETRICS. - ISSN 1180-4009. - ELETTRONICO. - 37:3(2026), pp. 1-11. [10.1002/env.70096]

*Availability:*
This version is available at: 11583/3010181 since: 2026-04-22T12:30:31Z

*Publisher:*
Wiley

*Published*
DOI:10.1002/env.70096

*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)

27 November 2026

Check for updates

*Environmetrics*

WILEY

RESEARCH ARTICLE OPEN ACCESS

# Modeling Benthic Animals in Space and Time Using Bayesian Point Process With Cross Validation: The Case of Holoturians

Daniele Poggio[1] | Gian Mario Sangiovanni[2] | Gianluca Mastrantonio[1] | Giovanna Jona Lasinio[2] | Edoardo Casoli[3] | Stefano Moro[4] | Daniele Ventura[3]

[1]Department of Mathematical Sciences 'G.L. Lagrange', Polytechnic of Turin, Torino, Italy | [2]Department of Statistical Sciences, University of Rome 'La Sapienza', Rome, Italy | [3]Department of Environmental Biology, University of Rome 'La Sapienza', Rome, Italy | [4]Department of Integrative Marine Ecology (EMI), Stazione Zoologica Anton Dohrn, Rome, Italy

**Correspondence:** Daniele Poggio (daniele.poggio@polito.it) | Gian Mario Sangiovanni (gianmario.sangiovanni@uniroma1.it)

**Received:** 31 May 2025 | **Revised:** 24 February 2026 | **Accepted:** 1 April 2026

**Keywords:** holothurian | INLA | Log-Gaussian Cox process | spatio-temporal point processes

## ABSTRACT

Holothurian populations in the Mediterranean are relatively understudied, with limited knowledge of their spatial distribution, habitat preferences, and ecological dynamics, making their monitoring a key challenge for ecosystem assessment and sustainable management. However, species distribution modeling is often complicated by the presence-only nature of the data and heterogeneous sampling designs. This study develops a spatio-temporal framework based on Log-Gaussian Cox Processes to analyze Holothurians' positions collected across nine survey campaigns conducted from 2022 to 2024 near Giglio Island, Italy. The surveys combined high-resolution photogrammetry with diver-based visual censuses, leading to varying detection probabilities across habitats, especially within *Posidonia oceanica* meadows. We adopt a model with a shared spatial Gaussian process component to accommodate this complexity, accounting for habitat structure, environmental covariates, and temporal variability. Model estimation is performed using Integrated Nested Laplace Approximation. We evaluate the predictive performances of alternative model specifications through a novel k-fold cross-validation strategy for point processes, using the Continuous Ranked Probability Score. Results highlight the influence of habitat-type covariates, strong variability across campaigns, and a locally structured spatial field capturing residual spatial heterogeneity. Our approach provides a flexible and computationally efficient framework for integrating heterogeneous presence-only data in marine ecology and comparing the predictive ability of alternative models.

## 1 | Introduction

Sea cucumbers (Holothuroidea) are ecologically significant marine invertebrates that contribute fundamentally to nutrient cycling, sediment bioturbation across various marine ecosystems, functioning across a range of marine habitats, from shallow seagrass beds to deep-sea ecosystems (Lopez and Levinton 1987; Purcell et al. 2016; Schneider et al. 2013). Their feeding and bioturbation activities enhance sediment oxygenation and organic matter turnover, rendering them "ecosystem engineers" in many coastal systems. Despite their ecological importance and increasing commercial exploitation (González-Wangüemert et al. 2014; Hamel et al. 2022), Mediterranean holothurian populations remain understudied, with significant knowledge gaps regarding their distribution patterns, habitat preferences, and ecological dynamics (Pasquini et al. 2022; Rakaj and Fianchini 2024). This gap is

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Environmetrics* published by John Wiley & Sons Ltd.

striking given that Mediterranean sea cucumber fisheries have expanded in recent decades, often relying on rudimentary catch records and opportunistic surveys.

Modeling sea cucumber distributions—and marine ecological patterns more broadly—faces two persistent challenges: the heterogeneity of environmental covariates and the presence-only nature of observational data (Warton and Shepherd 2010). Environmental variables are typically provided as GIS rasters or spatial lattices, requiring careful preprocessing and alignment before statistical analysis (Simpson et al. 2016). At the same time, records of sedentary benthic invertebrates lack absence data due to their nature, introducing complex detection biases and uneven sampling effort across surveys (Martino et al. 2021; Sicacha-Parada et al. 2021; Yuan et al. 2017). Failing to address these issues can lead to biased inference and suboptimal predictive performance.

In this study, we integrate presence data from nine survey campaigns conducted between 2022 and 2024 at Punta Del Lazzaretto (Giglio Island). The 2023 campaign, carried out within a *Posidonia oceanica* (*P. oceanica*) meadow, employed a distinct sampling technique, increasing the heterogeneity of the data set. To account for these differences, we model the observations from each campaign as realizations of a Log-Gaussian Cox process (LGCP), a flexible class of space–time point process models widely used in ecological applications (Brix and Diggle 2001; Serra et al. 2014). Similar modeling strategies have recently been applied to holothurian data in comparable contexts (Mastrantonio et al. 2024; Ventura et al. 2025).

Our contributions are fourfold. First, this work represents a comprehensive, spatiotemporal LGCP analysis of Mediterranean sea cucumbers that integrates multiple heterogeneous survey protocols. Second, we extend the point-process cross-validation strategy of Cronie et al. (2024) to spatially bounded subsets in a spatio-temporal setup, enabling localized assessment of predictive residuals. Third, we quantify the influence of *P. oceanica* and other environmental drivers on relative intensity, producing high-resolution maps that can guide habitat conservation and fishery management. Finally, we demonstrate how cross-validation for point processes enables rigorous comparison of alternative models and sampling designs, offering a robust alternative to the Deviance Information Criterion (DIC) by facilitating model comparison in a predictive rather than purely explanatory framework.

The remainder of this paper is structured as follows. In Section 2, we describe the study area, outline the methodology, and highlight differences among the survey campaigns. We also detail the data modeling strategy, and the procedure used for model comparison. Section 3 presents the preprocessing of environmental covariates and the results of the cross-validation analysis across different models, followed by an in-depth evaluation of the best-performing model. Finally, Section 4 is dedicated to the discussion.

## 2 | Materials and Methods

### 2.1 | Study Area

The study site is located at Punta Del Lazzaretto (N: 42.364867°; E: 10.920210°) on the northeastern coast of Giglio Island in the Tyrrhenian Sea, Italy (Figure 1). It encompasses approximately 5500 square meters of infralittoral seabed, with depths ranging from 8 to 27 m, characterized by a complex seabed that has undergone significant human-induced disturbances. This area gained prominence after the Costa Concordia shipwreck incident (Casoli et al. 2017) since the wreck and associated salvage operations created a shadowing effect that adversely affected the native *P. oceanica* Delile meadows, resulting in a marked decline in their coverage (Mancini et al. 2019; Toniolo et al. 2018). After the removal of the wreck in July 2014 and the completion of seabed remediation in April 2018, a five-year restoration project began in 2019, aiming at large-scale transplantation of *P. oceanica* (Mancini et al. 2022). Currently, the research area exhibits a diverse mosaic of marine infralittoral communities, featuring sandy substrates interspersed with granitic formations, such as boulders and pebbles, which are covered by photophilic algal communities. The site also displays both natural *P. oceanica* patches and a transplanted meadow, colonizing discontinuous patches of dead *P. oceanica* rhizomes with trapped sediments defined as dead "matte" and sandy areas.

### 2.2 | Data Collection

Nine sampling campaigns were carried out in 2022, 2023, and 2024. In 2022, five campaigns were conducted, and in 2023 and 2024, only two campaigns were carried out. In 2022, data were collected using the high-resolution structure from motion (SFM) photogrammetric technique in order to monitor the progress of the expansion of restoration initiatives in the study area (Ventura et al. 2022) and to assess the structure and composition of benthic communities (Ventura et al. 2025). This technique, while effective for broader spatial mapping, may have led to an underestimation of the sea cucumber abundance within natural *P. oceanica* meadows, due to limitations in detecting individuals hidden within dense seagrass.

To improve the accuracy in estimating the population density within the *P. oceanica* habitat, two additional campaigns were conducted in 2023. In each campaign of 2023, three SCUBA divers performed the sampling of sea cucumbers in the natural meadow surrounding the edge of the study area through underwater visual census (UVC). Each diver employed a rope to establish a sequence of linear transects ranging from 50 to 200 m in length, contingent upon the size of the meadow. They then conducted a count of specimens within a 1-m width, measuring 50 cm from the rope on either side. To identify all sea cucumber specimens, divers manually displaced the seagrass leaves surrounding the rope and documented the estimated locations on a waterproof plasticized map of the study area, which was created using orthophoto mosaics based on SFM obtained previously. When the *P. oceanica* patches exhibited a polygonal shape,



1099095x, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.70096 by CochraneItalia, Wiley Online Library on [22/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

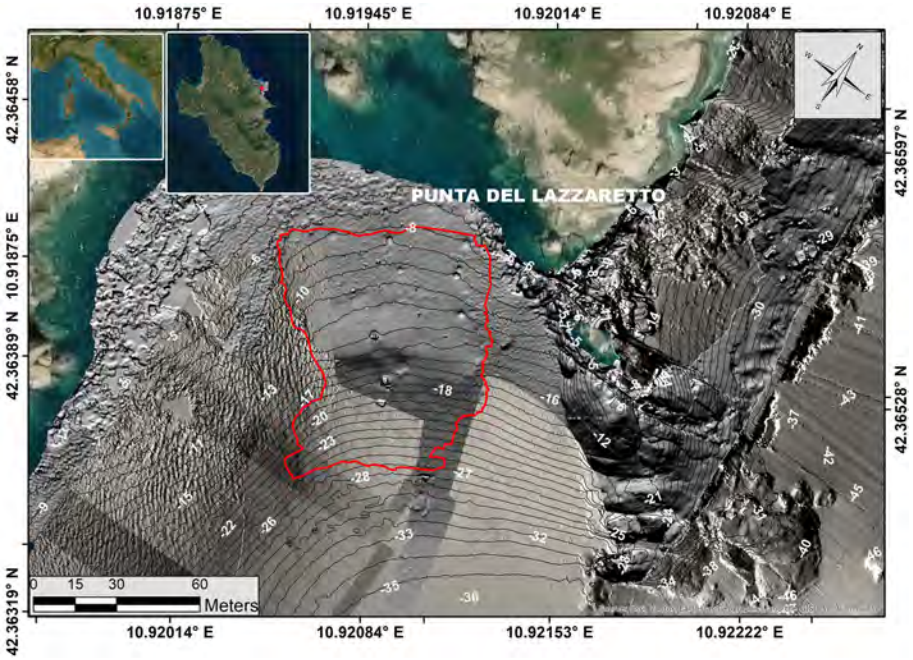



**FIGURE 1** | Map of the study area. The surveyed area mapped by high spatial resolution SFM photogrammetry is represented by the red polygon. Shaded relief (hillshade) derived MBES-based bathymetry and contour lines at 1 m were also reported to represent the seabed morphology. The map utilizes the coordinate reference System EPSG: 32632–WGS 84/UTM Zone 32 N Transverse Mercator and was generated using ESRI ArcMap (version 10.6) software.

particularly in the deeper sections of the study area, the circle search pattern technique was implemented to enhance the manual identification of sea cucumbers. This method involved a diver swimming in concentric circles, progressively enlarging the radius around the center of the patch. A graduated rope was employed to ensure precise control over the distance of each circular trajectory from the center.

In 2024, the two campaigns were carried out collecting data using both SFM and UVC, to get estimations of the abundance of sea cucumbers in all the study area, including in the natural *P. oceanica* meadow. The integration of manual and photogrammetric data facilitated the development of a unified analytical framework, combining the strengths of both data collection methods.

## 2.3 | Data Modeling

We envision the Holoturian positions $\mathbf{U}_t \in \mathcal{D}$, where $t = 1, \ldots, 9$ denotes the campaign where the data has been recorded, as a realization of a Log-Gaussian Cox Process (LGCP) (Møller et al. 1998). The spatial domain $\mathcal{D} \subset \mathbb{R}^2$, where the point process is defined, is the union of the *P. oceanica* habitat $\mathcal{D}_1 \subset \mathbb{R}^2$, while the remaining habitat types are indicated as $\mathcal{D}_2 \subset \mathbb{R}^2$. The two domains are disjoint, that is, $\mathcal{D}_1 \cap \mathcal{D}_2 = \emptyset$, and their union covers the entire study area: $\mathcal{D}_1 \cup \mathcal{D}_2 = \mathcal{D}$. Let $\lambda_t(\mathbf{s})$ be the intensity function of the LGCP at time $t = 1, \ldots, 9$, and spatial location $\mathbf{s} = (s_1, s_2)^T \in \mathcal{D}$. Each time point $t$ corresponds to a specific sampling campaign indexed as $C_{y.i}$, where $y \in \{1,2,3\}$ denotes the year and $i$ the campaign number within that year. Specifically, $C_{1.1}$ to $C_{1.5}$ represent the five campaigns conducted in 2022, $C_{2.1}$ and $C_{2.2}$ those from 2022, $C_{3.1}$ and $C_{3.2}$ the two campaigns from 2023.

The log-intensity $\log(\lambda_t(\mathbf{s}))$ is modeled as a linear combination of parameters $\mu_t$ accounting for changes in the total number of observed individuals across time (campaign effects), a common intercept across seasons $\mu_0$, a fixed term $\beta$ to model the effects of spatial covariates $\mathbf{x}(\mathbf{s})^T = (x_1(\mathbf{s}), \ldots, x_p(\mathbf{s}))^T$, an additional term $\gamma$ to take into account the effect of the habitat *P. oceanica* versus the others. Spatial dependence is modeled through a Gaussian process $w(\mathbf{s})$ (GP), which captures residual spatial structure beyond the observed covariates. Following previous applications in similar ecological contexts (Mastrantonio et al. 2024; Ventura et al. 2025), a shared GP is adopted to describe the underlying spatio-temporal variation. For the campaign effects, we define a second hierarchical level where $\mu_t$ is distributed as a Gaussian distribution with 0 mean and variance $\tau^2$. The model is defined as follows:

$$
\begin{aligned}
\mathbf{U}_t \mid \lambda_t(\mathbf{s}) &\sim \mathrm{PP}\left(\lambda_t(\mathbf{s})\right) \\
\log\left(\lambda_t(\mathbf{s})\right) &= \mu_0 + \mu_t + \gamma z(\mathbf{s}) + \mathbf{x}(\mathbf{s})^T \boldsymbol{\beta} + w(\mathbf{s}), \\
w(\mathbf{s}) &\sim \mathrm{GP}\left(0, C\left(\cdot; \sigma^2, \rho\right)\right), \\
\mu_t &\sim N\left(0, \tau^2\right).
\end{aligned}
$$

where $\sigma^2$ and $\rho$ are the variance and the range of the Matèrn covariance function, while $\alpha = 2$ (Lindgren et al. 2011). In this work, $\mathbf{s} \in D_2$ for $t \in \{1,2,3,4,5\}$, $\mathbf{s} \in D_1$ if $t \in \{6,7\}$ and $\mathbf{s} \in \mathcal{D}$ if $t \in \{8,9\}$. We adopt this modeling framework for several reasons. First, the term $\mu_t$ allows us to flexibly account for seasonal or campaign-specific differences in the total abundance, without imposing a fixed trend or requiring homogeneous effort across time. This is crucial in ecological studies where survey efforts and environmental conditions vary over campaigns. Second, the inclusion of the binary habitat indicator $z(\mathbf{s})$ and its



1099095x, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.70096 by CochraneItalia, Wiley Online Library on [22/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

associated parameter $\gamma$ enables the model to adjust for heterogeneous sampling intensity and ecological characteristics specific to the *P. oceanica* habitat, especially considering its enhanced exploration during certain campaigns. This additive structure for the log intensity leads to a multiplicative decomposition of the intensity function.

$$\lambda_t(s) = \exp\big(\mu_0 + \mathbf{x}(\mathbf{s})^T\boldsymbol{\beta} + w(\mathbf{s})\big) \cdot \exp\big(\mu_t\big) \cdot \exp(\gamma z(s)), \quad s \in D \quad (1)$$

In the formulation of Equation (1), $\exp\big(\mu_t\big)$ accounts for variation across sampling times. In contrast, $\exp(\gamma z(\mathbf{s}))$ corrects the intensity for differing sampling intensities within the *P. oceanica* region ($D_1$), so it can be interpreted as a measure of the sampling effort. This factorization thus provides an effective means of correcting the overall intensity to incorporate these influences.

## 2.4 | Model Estimation

For the estimations of some alternative models we used INLA (Rue et al. 2009) R-INLA, and in particular the wrapper inlabru (Bachl et al. 2019). INLA is an algorithm for approximate Bayesian inference in Latent Gaussian Models (LGM), where the latent field is Gaussian and governed by a small number of hyperparameters, and the response variables can be non-Gaussian. It relies on the Laplace approximation to efficiently compute posterior distributions.

INLA supports spatial models where the GP is approximated by a Gaussian Markov Random Field using the SPDE approach (Lindgren et al. 2011), which substantially reduces the estimation time. The core idea is to construct a finite-dimensional representation of the spatial Gaussian process over a mesh defined on the spatial domain, and to approximate the LGCP likelihood in a way that makes it tractable for INLA-based inference (Simpson et al. 2016).

In terms of prior distributions, for the intercept $\mu_0$ and the regression coefficients $\boldsymbol{\beta}$, we employ non-informative Gaussian priors centered at 0 and with precision 0.001. For the parameters $\sigma^2$ and $\rho$ of the spatial Gaussian process, we choose PC priors (Simpson et al. 2017). Specifically, we set $\mathbb{P}(\rho < 50) = 0.5$ and $\mathbb{P}(\sigma > 0.5) = 0.01$. Those choices are consistent with the spatial extent of the study area. We set for the inverse of $\tau^2$, that is, precision, as a Log-Gamma prior with shape and rate fixed to 1.0 and 0.01, respectively.

## 2.5 | Model Comparison

To effectively compare the performance of alternative models, it is essential to employ evaluation metrics that account for both model complexity and prediction capability. While the DIC is a commonly used tool for model evaluation in Bayesian settings, it has notable limitations (McElreath 2018). Specifically, the DIC tends to favor overfitted models and lacks consistency in model identification, especially in hierarchical settings. Given these concerns, we opted to asses models within the predictive framework (Gelfand and Schliep 2018; Leininger and Gelfand 2017). In particular, we employ a k-fold cross-validation procedure for point processes (Cronie et al. 2024), allowing us to assess the out-of-sample predictive performance. As our primary measure of predictive accuracy, we use raw residuals (Baddeley et al. 2005). To enhance the spatial resolution and adaptability to temporal settings, we extend this approach by partitioning the study region into bounded spatial subsets $B_1^t, \ldots, B_{G_t}^t$ at each time point $t$, enabling a localized approximation of the residuals. This partitioning is motivated by the spatial structure of our data: the first five sampling times are defined over $D_2$, the next two over $D_1$, and the final two over the entire region $D$. As a result, the same partition is used for the 2022 sampling campaigns, a different one for the 2023 sampling campaigns, and yet another for the remaining two campaigns. For each time $t$, suppose that $\mathbf{U}_t \mid \lambda_t$ is a Non-Homogeneous Poisson Process with intensity function $\lambda_t$ defined over the corresponding spatial domain. A realization of this process is denoted by $\mathbf{u}_t$, which can be partitioned into $\mathbf{u}_t^1, \ldots, \mathbf{u}_t^{G_t}$, representing the observed points within each of the bounded subsets. The raw residual in the $g$-th subset at time $t$ is then defined as

$$R_t^g = \left\|\mathbf{u}_t^g\right\| - \int_{B_g^t} \lambda_t(\mathbf{s})d\mathbf{s},$$

where $N_t^g = \left\|\mathbf{u}_t^g\right\|$ represents the observed number of points in the $g$-th bounded subset at time $t$. This residual measures the discrepancy between observed and expected counts, and its expected value is 0 (Baddeley et al. 2005). The idea behind Cross-Validation for point processes is to construct a marked point process by assigning, at each observed location $\mathbf{u}$, a mark $m(\mathbf{u})$ drawn from a $k$-dimensional Multinomial distribution with equal probability $p = K^{-1}$, where $K$ is the number of employed folds. For each fold $k$, we define the training and testing sets as $\mathbf{u}_{k,t}^{train} = \left\{u \in \mathbf{u}_t : m(u) \neq k\right\}$ and $\mathbf{u}_{k,t}^{val} = \left\{u \in \mathbf{u}_t : m(u) = k\right\} = \mathbf{u}_t \backslash \mathbf{u}_{k,t}^{train}$, respectively. These are still LGCP with intensities:

$$\lambda_t^{train}(\mathbf{s}) = \frac{K-1}{K}\lambda_t(\mathbf{s}), \quad \lambda_t^{val}(\mathbf{s}) = \frac{1}{K-1}\lambda_t^{train}(\mathbf{s}). \quad (2)$$

By using the intensities in (2) and partitioning the validation set into $G_t$ subsets $\mathbf{u}_{k,t}^{val,g}$ for each time $t$, we can easily compute the raw residuals for the validation sets

$$R_{k,t}^{val,g} = \left\|\mathbf{u}_{k,t}^{val,g}\right\| - \int_{B_g^t} \lambda_t^{val}(\mathbf{s})d\mathbf{s} = \left\|\mathbf{u}_{k,t}^{val,g}\right\| - \frac{1}{K-1}\int_{B_g^t} \lambda_t^{train}(\mathbf{s})d\mathbf{s}.$$

An estimator of $R_{k,t}^{val,g}$ is what should be used to assess the model performance, since, if the model is correctly specified, it is expected to have a value close to 0. Because the closed-form expression of the residual posterior distribution is generally unavailable, we use Monte Carlo simulations with a sample size $A$ to approximate it. Evaluating the spatial integral is often the most challenging aspect of this procedure. Since the different domains are in $R^2$, the integral can be approximated using a quadrature formula based on a discretization of the spatial domain (Berman and Turner 1992):

$$\int_D \lambda_t^{train}(s)ds \approx \sum_{q=1}^{Q} \alpha_q \lambda_t\big(v_q\big), \quad (3)$$

where $\mathcal{V} = \left\{v_q \in D\right\}_{q=1}^{Q}$ are the quadrature nodes and $\mathcal{W} = \left\{\alpha_q\right\}_{q=1}^{Q}$ are the corresponding weights. At the end of this process, we obtain a $A \times K \times G_t$ tensor, where $A$ is the number of



1099095x, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.70096 by CochraneItalia, Wiley Online Library on [22/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Monte Carlo replicates and each column contains the residual samples corresponding to each $k$-fold given a subset. If we repeat this procedure for all the $T$ sampling times, we obtain $T$ tensors. Furthermore, as our objective is to assess the predictive performance of $L$ competing models, denoted by $\mathcal{M}_1, \ldots, \mathcal{M}_L$, we obtain $L \times T$ different tensors. To have a fully Bayesian evaluation of the model performance, we can use proper scoring rules, which provide summary measures to evaluate probabilistic forecasts by assigning a numerical score (Gneiting and Raftery 2007). In particular, we opted for the Continuous Ranked Probability Score (CRPS) (Matheson and Winkler 1976), measuring the mean square error between the predicted and empirical cumulative distribution functions. In a general setting where we want to predict the value of an observation $y$, it can be expressed as

$$\text{CRPS}(F, y) = \int_{\mathbb{R}} (F(x) - 1\{x \geq y\})^2 dx$$

where $F(\cdot)$ is the predictive distribution of $y$ and 1 is the Heaviside step-function. In our application, given that we assumed 0 value for the residual mean, we are particularly interested in the case when $y = 0$. Furthermore, indicating with $F^{val,g}_{R_{k,t}}$ the predictive distribution of the raw residuals, the CRPS is then

$$\text{CRPS}\left(F^{val,g}_{R_{k,t}}, 0\right) = \int_{\mathbb{R}} \left(F^{val}_{R_{k,t}}(x) - 1\{x \geq 0\}\right)^2 dx \qquad (4)$$

The integral in Equation (4) cannot be computed analytically, but it can be approximated (Jordan et al. 2019; Krüger et al. 2021) as

$$\text{CRPS}\left(F^{val,g}_{R_{k,t}}, 0\right) \approx \frac{1}{A}\sum_{a=1}^{A} \mid r^{g,a}_{k,t} - 0 \mid - \frac{1}{2A^2}\sum_{a=1}^{A}\sum_{l=1}^{A} \mid r^{g,a}_{k,t} - r^{g,l}_{k,t} \mid$$

where $r^{g,a}_{k,t}$ are samples from the distribution $F^{val,g}_{R_{k,t}}$, which can be easily obtained from the posterior marginal distributions of the fitted model by the inlabru wrapper. To apply Equation (3) for each posterior sample, it is necessary to draw the value of $\lambda_t$ from its posterior distribution for all the quadrature points. From a computational standpoint, the computation of residual samples across time points $t$ can be parallelized without loss of generality, significantly improving the efficiency. Lastly, since we employ cross-validation, we compute a more robust estimate of the CRPS by averaging over all folds:

$$\text{CRPS}^{g,t}_{CV} = \frac{1}{K}\sum_{k=1}^{K} \text{CRPS}\left(F^{val,g}_{R_{k,t}}, 0\right)$$

resulting in $G \times T$ CRPS values. In the context of such scores, constructing a local map of residuals/CRPS values becomes a viable solution. This map can offer valuable insights into the model's capacity. Nonetheless, it is possible to aggregate these scores over time and partitions to obtain a single summary score. This then allows for comparisons to be made across the $L$ candidate models. In conclusion, the model with the lower CRPS score is the most effective of the ones examined.

## 3 | Results

### 3.1 | Data Integration

We first report the outcomes of the data integration and filtering procedure, which establish the unified analytical framework used in the subsequent analyses. Figure 2b illustrates the resulting distribution of observations across campaigns. Unsurprisingly, the 2022 campaigns contributed the highest number of observations, due to the extensive spatial coverage of the SFM approach. However, nearly 90% of these observations were located outside *P. oceanica* meadows, reflecting a potential habitat bias inherent to the initial sampling design. After a critical

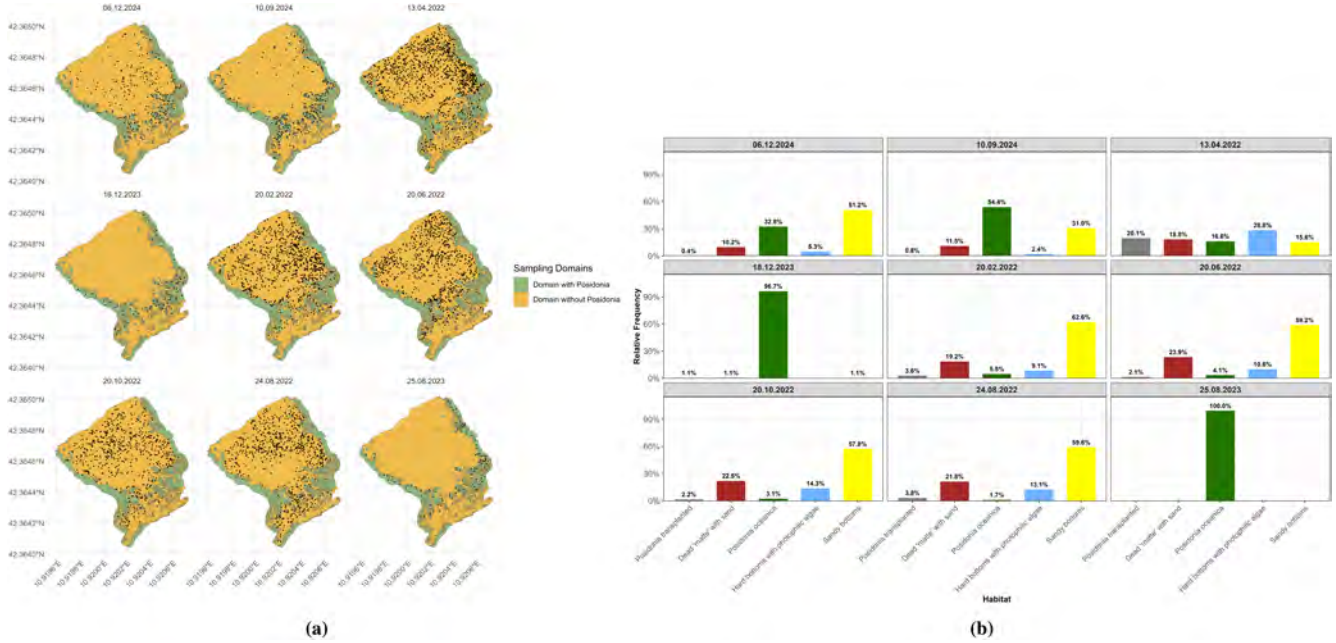



**FIGURE 2** | Spatial distribution of specimens (a) and relative frequencies of observed sea cucumbers within different habitats (b) for the nine sampling campaigns.



1099095x, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.70096 by CochraneItalia, Wiley Online Library on [22/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 1** | Sea cucumber observations stratified by season and habitat, excluding records within *P. oceanica* for 2022 surveys and excluding records outside *P. oceanica* for 2023 surveys.

| Sampling date | P. transplanted | Dead 'matte' | *P. oceanica* | Hard bottom | Sandy bottom | Total |
|---|---|---|---|---|---|---|
| 20 Feb 2022 | 35 | 189 | 0 | 90 | 616 | 930 |
| 13 Apr 2022 | 302 | 282 | 0 | 433 | 235 | 1252 |
| 20 Jun 2022 | 20 | 227 | 0 | 101 | 562 | 910 |
| 24 Aug 2022 | 30 | 171 | 0 | 103 | 468 | 772 |
| 20 Oct 2022 | 15 | 151 | 0 | 96 | 388 | 650 |
| 25 Aug 2023 | 0 | 0 | 110 | 0 | 0 | 110 |
| 18 Dec 2023 | 0 | 0 | 88 | 0 | 0 | 88 |
| 10 Sep 2024 | 2 | 29 | 137 | 6 | 78 | 252 |
| 06 Dec 2024 | 1 | 25 | 80 | 13 | 125 | 244 |
| Total | 405 | 1074 | 415 | 842 | 2472 | 5208 |

quality assessment of the data, observations within *P. oceanica* from 2022 were excluded due to unreliability, primarily stemming from visibility limitations. Similarly, observations outside of *P. oceanica* from 2023 were also discarded, as they lacked sufficient observational rigor. This careful filtering process resulted in a clean and well-separated dataset, both spatially and temporally, as summarized in Table 1 and Figure 2a. This refined dataset not only enhances the robustness of our model inference but also ensures that habitat comparisons are not confounded by sampling methodological inconsistencies across years.

In terms of covariates, we used the same data considered in Ventura et al. (2025) and Mastrantonio et al. (2024), with a resolution of $0.21 \times 0.24$ $m$. We considered continuous spatial covariates such as slope, depth, and rugosity, as well as a categorical covariate representing the benthic habitat type. Depth, slope, and rugosity were derived from high-resolution Digital Surface Models (DSMs). Due to the extremely high resolution of the imagery, to extract the benthic habitat cover types (Fallati et al. 2024), an object-based image analysis (OBIA) approach was employed to facilitate the classification of orthophoto mosaics. The classification process involved segmenting the images into homogeneous regions based on spectral and geometric features, followed by a supervised classification that assigned each segment to one of five benthic habitat classes: sandy bottoms, hard bottoms with photophilic algae, dead matte with sand, natural *P. oceanica* meadow, and transplanted *P. oceanica*. Rugosity was excluded from the analysis due to its high correlation with slope (Figure 3d). As the quantitative covariates violated the assumption of normality, differences across habitat types were assessed using the Kruskal–Wallis test. For both slope and depth, the test revealed statistically significant differences among habitat classes at the 0.05 significance level, leading to the rejection of the null hypothesis of homogeneity. A visual comparison of the covariates across habitats is presented in Figure 3. These environmental variables were aggregated over a regular $1 \times 1$ $m$ grid using zonal average due to computational efficiency.

### 3.2 | Model Evaluation

In this subsection, we present the results of the proposed model applied to the motivating dataset. We tested 240 models, each of them with a different combination of the following covariates: habitat types, seabed slope, depth, and spatial coordinates. These covariates were selected based on prior ecological knowledge and after exploratory data analysis (Section 2.2). All model combinations were tested under the constraint that *P. oceanica* was always included as a covariate. Due to identifiability constraints, not all habitat types could be included simultaneously as separate indicators. Given the relatively low number of observations in the most recent sampling campaigns, we opted for a five-fold cross-validation scheme to ensure that each fold retained sufficient data for reliable estimation. The group assignments for the folds were generated once and kept fixed throughout all model evaluations to ensure consistency and comparability of the results. The choice of the number of folds in cross-validation can be adapted based on the available data, fewer folds may be preferable in cases of limited observations to avoid overly small training and test sets. However, it is important to consider the computational burden: increasing the number of folds significantly increases the number of model fittings, especially when repeated across multiple replicates and models, as in our study. We set the number of Monte Carlo replicates used for integral approximation to 1000, and divided each sampling domain into 324 bounded subsets using an $18 \times 18$ regular grid. This resolution provided a sufficiently detailed representation of the spatial domain while remaining computationally tractable. The grid resolution can be adjusted based on the spatial extent and resolution of the data.

For each model, we computed residuals and the CRPS. Table 2 reports the top 10 models ranked by CRPS. In the Supporting Information, you can also find a comparison with a simpler model. The first observation is that the CRPS values differ only in the fourth or fifth decimal place, suggesting that all top models perform similarly in terms of predictive accuracy. Given this similarity, we prioritized models where all covariates had 95% credible intervals that did not include zero and whose effects were ecologically interpretable. When using DIC for model selection instead, the model rankings changed substantially. This discrepancy raises questions about the reliability of DIC for model comparison in this context.

Based on these criteria, we selected the model that includes dead matte and P. transplanted as habitat covariates, alongside



1099095x, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.70096 by CochraneItalia, Wiley Online Library on [22/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

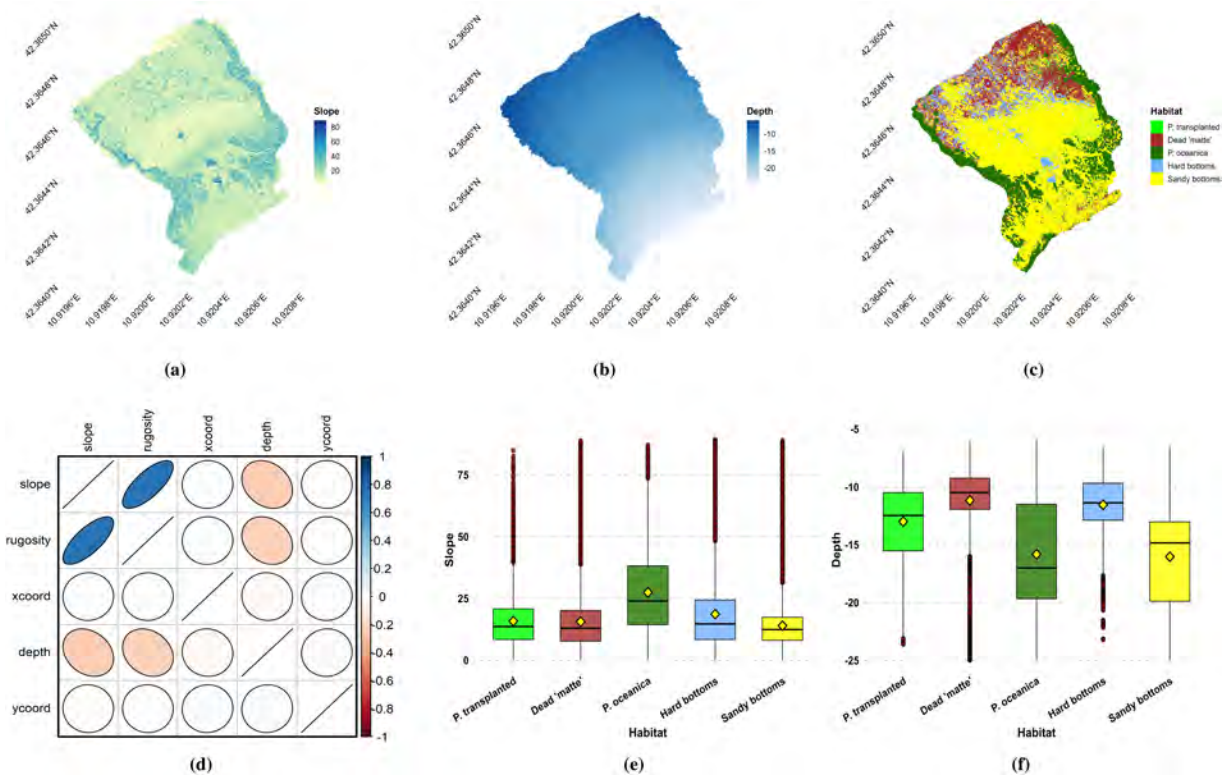



**FIGURE 3** | Spatial distribution of the variables (a) slope, (b) depth, and (c) habitat. The correlation plot of all the covariates is also presented (d). The conditional distribution of slope (e) and depth (d) is presented for the five distinct habitats. The yellow rhomboids represent the value of the covariate mean given a specific habitat.

**TABLE 2** | Covariates included in the top 10 models ranked by CRPS. Ticks denote inclusion in model fitting computation, while crosses indicate exclusion.

| Depth | Slope | Xcoord | Ycoord | P. transplanted | Dead 'matte' | Hard bottom | Sandy bottom | CRPS |
|---|---|---|---|---|---|---|---|---|
| ✗ | ✗ | ✗ | ✗ | ✓ | ✓ | ✗ | ✓ | 0.454320 |
| ✗ | ✗ | ✗ | ✗ | ✓ | ✓ | ✗ | ✗ | 0.454329 |
| ✗ | ✗ | ✓ | ✗ | ✓ | ✓ | ✗ | ✗ | 0.454373 |
| ✗ | ✗ | ✓ | ✓ | ✗ | ✓ | ✗ | ✓ | 0.454388 |
| ✗ | ✗ | ✓ | ✓ | ✓ | ✗ | ✗ | ✗ | 0.454389 |
| ✗ | ✗ | ✗ | ✓ | ✓ | ✓ | ✗ | ✓ | 0.454400 |
| ✗ | ✗ | ✗ | ✗ | ✗ | ✓ | ✗ | ✓ | 0.454411 |
| ✗ | ✗ | ✗ | ✓ | ✓ | ✓ | ✗ | ✗ | 0.454416 |
| ✗ | ✗ | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | 0.454420 |
| ✗ | ✗ | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | 0.454451 |

*Note:* It is worth noticing that the inclusion of a covariate in the model does not imply that the credible interval of the related effect does not include the zero point.

*P. oceanica*. Notably, sandy bottoms appears in about half of the best-performing models, but its 95% credible interval consistently includes zero. Neither depth nor slope were retained in the list of the best models. Figure 4a shows the residual mean for the selected model across the nine different sampling campaigns. There does not appear to be a clear spatial pattern in the residuals, in fact the model does not systematically overestimate or underestimate in specific areas.

Table 3 reports the posterior summary statistics for the intercept, regression coefficients, campaign-specific effects, their precision, and the hyperparameters of the spatial GP. The intercept is significantly negative, indicating a low baseline intensity. The habitat covariates show distinct effects: *P. oceanica* has a consistently negative coefficient, suggesting reduced expected counts in this habitat relative to the others, possibly due to differential detectability of Sea Cucumbers in this particular habitat. Dead matte also has a small negative effect, while P. transplanted exhibits a positive and well-identified influence on the intensity. The campaign-specific effects reflect considerable variability over time. Campaigns in 2021 generally show positive



1099095x, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.70096 by CochraneItalia, Wiley Online Library on [22/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

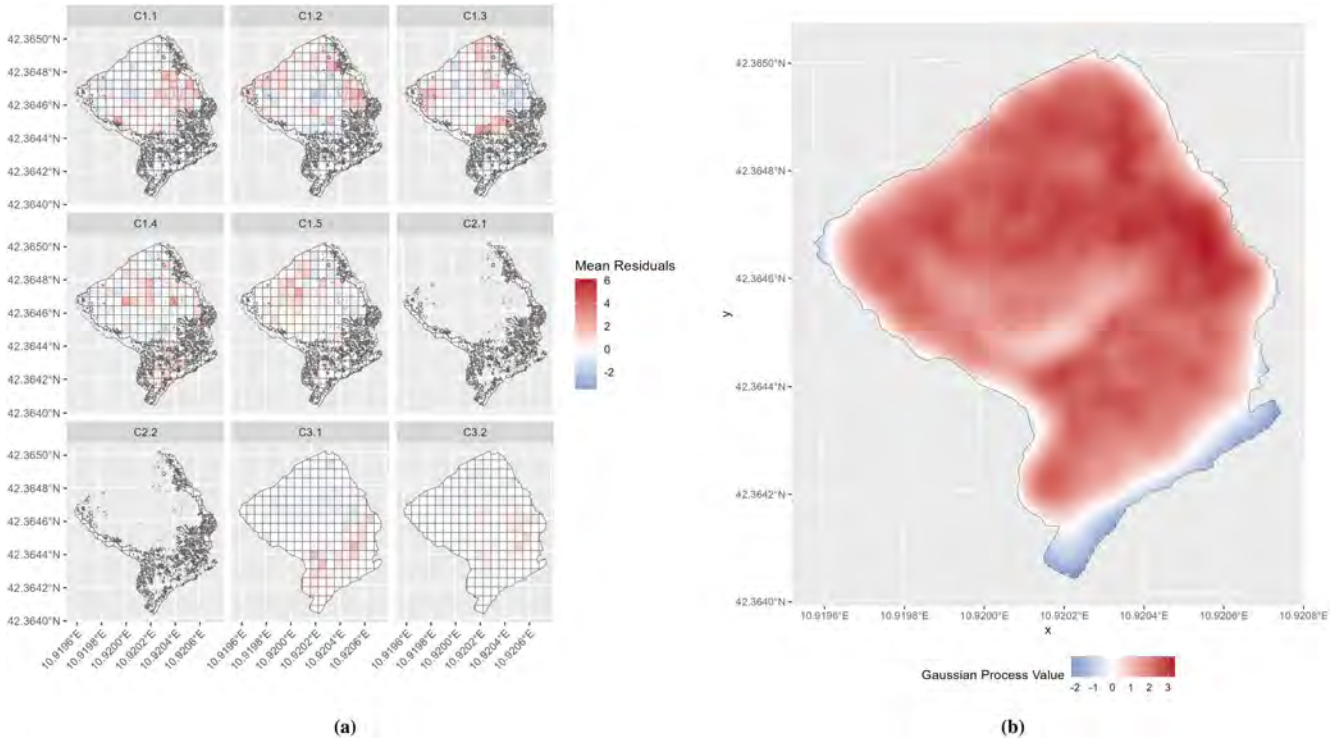



**FIGURE 4** | Average value of the residuals for the chosen model across the different sampling times and the different defined partitions (a) and Posterior mean for the shared GP (b).

**TABLE 3** | Posterior summaries (mean, standard deviation, quantiles) for the intercept, regression coefficients, campaign-specific effects, the precision for the latter, and the parameters associated with the GP of the selected model using the k-fold cross-validation procedure.

| Parameter | Mean | SD | 2.5% | Median | 97.5% |
|---|---|---|---|---|---|
| Intercept | −4.718 | 0.075 | −4.864 | −4.718 | −4.572 |
| *P. oceanica* | −0.388 | 0.071 | −0.528 | −0.388 | −0.248 |
| Dead matte | −0.134 | 0.054 | −0.239 | −0.134 | −0.028 |
| P. transplanted | 0.308 | 0.100 | 0.113 | 0.308 | 0.504 |
| Random effect | | | | | |
| C1.1 | 0.819 | 0.035 | 0.749 | 0.819 | 0.888 |
| C1.2 | 1.116 | 0.032 | 1.053 | 1.116 | 1.179 |
| C1.3 | 0.797 | 0.036 | 0.727 | 0.797 | 0.867 |
| C1.4 | 0.633 | 0.038 | 0.559 | 0.633 | 0.707 |
| C1.5 | 0.461 | 0.040 | 0.382 | 0.461 | 0.539 |
| C2.1 | −1.303 | 0.085 | −1.471 | −1.303 | −1.136 |
| C2.2 | −1.521 | 0.095 | −1.706 | −1.521 | −1.335 |
| C3.1 | −0.485 | 0.059 | −0.600 | −0.485 | −0.369 |
| C3.2 | −0.517 | 0.060 | −0.634 | −0.517 | −0.400 |
| Hyperparameters | | | | | |
| Precision for random effect | 1.33 | 0.611 | 0.474 | 1.22 | 2.83 |
| Range for GP | 56.72 | 7.684 | 43.406 | 56.11 | 73.60 |
| Std. for GP | 2.05 | 0.259 | 1.598 | 2.03 | 2.61 |

*Note:* Recalling Section 2 *P. oceanica* effect is indicated as $\gamma$ in 2.3.



1099095x, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.70096 by CochraneItalia, Wiley Online Library on [22/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

values, suggesting higher total abundances relative to the overall mean. In contrast, campaigns from 2022 have strongly negative effects, highlighting a notable drop in the observed intensity. Campaigns from 2023 fall in between, with moderately negative effects. These random effects underscore the importance of adjusting for campaign-specific variation in total counts, consistent with the ecological understanding of fluctuations in detection or abundance across seasons and years. The posterior mean of the GP, shown in Figure 4b, reveals a clear spatial structure. The GP captures the spatial variation in the intensity of occurrences that is not explained by the covariates or by the campaign-specific effects. The spatial field shows higher values concentrated in the northern and northeastern parts of the study area, indicating regions where the observed point intensity is systematically higher than what would be expected based solely on the included covariates. This could be due to local habitat features not being explicitly modeled. Conversely, the southern and southeastern edge of the domain exhibits lower GP values. These areas are characterized by limited ecological visibility during field campaigns. Furthermore, the posterior estimate of the range parameter implies that on average, spatial correlation decays significantly beyond approximately 56 m. It suggests that the residual variation is locally structured but not globally persistent across the entire domain.

## 4 | Discussion

Sustainable management of sea cucumber populations, particularly in the Mediterranean, requires detailed knowledge of species-habitat associations and the consequences of their depletion for ecosystem functioning. *Holothuria tubulosa* and related species serve key roles in benthic ecosystems as deposit feeders and bioturbators, influencing sedimentary processes and nutrient cycling (Purcell et al. 2016). Despite this, management frameworks often lack robust data, and exploitation continues in many areas, frequently under unregulated or poorly documented conditions (Rakaj and Fianchini 2024). Our study addresses this gap by combining high-resolution spatial modeling with multiple sampling protocols, enabling fine-scale inference on distributional patterns. In contrast to traditional census methods, which are logistically demanding, our approach integrates photogrammetric surveys with diver-based transect data, accounting for the variability in detection related to habitat structure. This integration is crucial in environments dominated by dense seagrass, where detection is known to be impaired due to visual occlusion and topographic complexity. The model results highlight distinct habitat effects: *P. oceanica* exhibits a consistently negative association with sea cucumber counts, likely due to both lower detectability and potentially less favorable habitat conditions. Dead matte also shows a negative but weaker effect, suggesting it may offer suboptimal habitat or support lower densities. In contrast, P. transplanted areas display a clear positive association with sea cucumber intensity, indicating that these restored habitats may support favorable environmental conditions or increased visibility, thus enhancing detectability and/or actual presence.

In this study, we implemented the k-fold cross-validation procedure for point processes as theoretically developed by Cronie et al. (2024), and extended it to spatio-temporal point process models. To the best of our knowledge, this represents the first practical implementation of this method in a spatio-temporal setting. This is a significant contribution that allows for model comparison in a predictive framework, providing a robust alternative to the traditional DIC. We evaluated a total of 240 models, encompassing various combinations of continuous and categorical spatial covariates. The cross-validation framework enabled a systematic comparison based on the CRPS, but every suitable proper score can be used. Notably, in our case study, this procedure yielded a markedly different ranking of models compared to the DIC, highlighting that DIC may not always be the most appropriate choice for model selection in certain contexts. This observation aligns with findings by Leininger and Gelfand (2017), who emphasized that model comparisons should be conducted in a predictive space, as traditional criteria like DIC may not adequately capture a model's predictive performance.

Most of the analyses were conducted using standard computational resources, thanks to the efficiency of the INLA algorithm, as implemented in the R-INLA and inlabru packages. To further accelerate computations during the cross-validation phase, we leveraged the Terastat 2.0 infrastructure (Bompiani et al. 2020). Specifically, we employed 4 cores with 2 GB of RAM each, achieving an average runtime of approximately 15 min per model, which is highly efficient given the complexity of the underlying operations. It is important to note that the cross-validation procedure is method-agnostic and can, in principle, be applied within any inferential framework.

Despite the methodological strengths of our framework, the application to the sea cucumber data revealed only marginal differences in the predictive performance across the tested models. Two factors may explain this outcome. First, the study area may lack sufficient variation in key environmental covariates, limiting the model's ability to distinguish meaningful patterns in species abundance. Second, the supervised classification process used to generate habitat labels from imagery may have introduced a systematic bias, potentially obscuring ecological signals and reducing the discriminative power of the models. In summary, our work demonstrates how integrating spatio-temporal modeling with multiple survey modalities can improve ecological inference for benthic species. By explicitly accounting for habitat structure and survey heterogeneity, this approach contributes both methodologically and practically to ecosystem-based fishery assessments and habitat restoration monitoring.


### Acknowledgments

This research was partially supported by the Pure Ocean Fund ('3DR-4-Seac' research grant) and PRIN 2022 (Project: LAGO-ON, grant number 20224LYJM5), which contributed to the acquisition of some equipment used for underwater field sampling. Additionally, Daniele Poggio's participation was made possible through the PNRR-NGEU project, funded by the Ministry of University and Research (MUR) under DM 630/2024. The authors would like to thank Prof. Alessio Pollice for his valuable discussions and insightful suggestions. The authors would also like to thank the anonymous reviewers for their valuable suggestions.

### Funding

Ventura Daniele reports financial support related to this work. This study was funded in part by the Pure Ocean Fund (research grant 3DR-4-Seac) and by the PRIN 2022 project LAGO-ON (Grant No. 20224LYJM5).




1099095x, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.70096 by CochraneItalia, Wiley Online Library on [22/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Daniele Poggio was supported by the PNRR-NGEU initiative funded by the Ministero dell'Università e della Ricerca under DM 630/2024. All other authors declare no known financial interests or personal relationships that could be perceived to have influenced the content of this work.

### Conflicts of Interest

The authors declare no conflicts of interest.

### Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

### References

Bachl, F. E., F. Lindgren, D. L. Borchers, and J. B. Illian. 2019. "Inlabru: An R Package for Bayesian Spatial Modelling From Ecological Survey Data." *Methods in Ecology and Evolution* 10: 760–766. https://doi.org/10.1111/2041-210X.13168.

Baddeley, A., R. Turner, J. Møller, and M. Hazelton. 2005. "Residual Analysis for Spatial Point Processes (With Discussion)." *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 67, no. 5: 617–666.

Berman, M., and T. R. Turner. 1992. "Approximating Point Process Likelihoods With Glim." *Journal of the Royal Statistical Society: Series C: Applied Statistics* 41, no. 1: 31–38.

Bompiani, E., U. F. Petrillo, G. Jona Lasinio, and F. Palini. 2020. "High-Performance Computing With Terastat." In 2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech). https://doi.org/10.1109/DASC-PICom-CBDCom-CyberSciTech49142.2020.00092.

Brix, A., and P. J. Diggle. 2001. "Spatiotemporal Prediction for Log-Gaussian Cox Processes." *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 63, no. 4: 823–841.

Casoli, E., D. Ventura, L. Cutroneo, et al. 2017. "Assessment of the Impact of Salvaging the Costa Concordia Wreck on the Deep Coralligenous Habitats." *Ecological Indicators* 80: 124–134.

Cronie, O., M. Moradi, and C. A. Biscio. 2024. "A Cross-Validation-Based Statistical Theory for Point Processes." *Biometrika* 111, no. 2: 625–641.

Fallati, L., G. Panieri, C. Argentino, A. G. Varzi, S. Bünz, and A. Savini. 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).

Gelfand, A. E., and E. M. Schliep. 2018. "Bayesian Inference and Computing for Spatial Point Patterns." In *Nsf-Cbms Regional Conference Series in Probability and Statistics*, vol. 10, 1–125. Institute of Mathematical Statistics and the American Statistical Association.

Gneiting, T., and A. E. Raftery. 2007. "Strictly Proper Scoring Rules, Prediction, and Estimation." *Journal of the American Statistical Association* 102, no. 477: 359–378.

González-Wangüemert, M., M. Aydin, and C. Conand. 2014. "Assessment of Sea Cucumber Populations From the Aegean Sea (Turkey): First Insights to Sustainable Management of New Fisheries." *Ocean and Coastal Management* 92: 87–94.

Hamel, J.-F., I. Eeckhaut, C. Conand, J. Sun, G. Caulier, and A. Mercier. 2022. "Global Knowledge on the Commercial Sea Cucumber Holothuria Scabra." *Advances in Marine Biology* 91: 1–286.

Jordan, A., F. Krüger, and S. Lerch. 2019. "Evaluating Probabilistic Forecasts With Scoringrules." *Journal of Statistical Software* 90: 1–37.

Krüger, F., S. Lerch, T. Thorarinsdottir, and T. Gneiting. 2021. "Predictive Inference Based on Markov Chain Monte Carlo Output." *International Statistical Review* 89, no. 2: 274–301.

Leininger, T. J., and A. E. Gelfand. 2017. "Bayesian Inference and Model Assessment for Spatial Point Patterns Using Posterior Predictive Samples." *Bayesian Analysis* 12, no. 1: 1–30.

Lindgren, F., H. Rue, and J. Lindström. 2011. "An Explicit Link Between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach." *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 73, no. 4: 423–498. https://doi.org/10.1111/j.1467-9868.2011.00777.x.

Lopez, G. R., and J. S. Levinton. 1987. "Ecology of Deposit-Feeding Animals in Marine Sediments." *Quarterly Review of Biology* 62, no. 3: 235–260.

Mancini, G., E. Casoli, D. Ventura, et al. 2019. "Impact of the Costa Concordia Shipwreck on a *Posidonia oceanica* Meadow: A Multi-Scale Assessment From a Population to a Landscape Level." *Marine Pollution Bulletin* 148: 168–181.

Mancini, G., D. Ventura, E. Casoli, A. Belluscio, and G. Ardizzone. 2022. "Transplantation on a *Posidonia oceanica* Meadow to Facilitate Its Recovery After the Concordia Shipwrecking." *Marine Pollution Bulletin* 179: 113683.

Martino, S., D. S. Pace, S. Moro, E. Casoli, D. Ventura, and A. Frachea. 2021. "Integration of Presence-Only Data From Several Sources: A Case Study on Dolphins' Spatial Distribution." *Ecography* 44, no. 10: 1533–1543.

Mastrantonio, G., D. Ventura, E. Casoli, et al. 2024. "Species Distribution Models With Masking: The Case of Holothurians in a Posidonia Rich Area." In Scientific Meeting of the Italian Statistical Society. 531–536.

Matheson, J. E., and R. L. Winkler. 1976. "Scoring Rules for Continuous Probability Distributions." *Management Science* 22, no. 10: 1087–1096.

McElreath, R. 2018. *Statistical Rethinking: A Bayesian Course With Examples in R and Stan*. Chapman and Hall/CRC.

Møller, J., A. R. Syversveen, and R. P. Waagepetersen. 1998. "Log Gaussian Cox Processes." *Scandinavian Journal of Statistics* 25, no. 3: 451–482.

Pasquini, V., C. Porcu, M. F. Marongiu, M. C. Follesa, A. A. Giglioli, and P. Addis. 2022. "New Insights Upon the Reproductive Biology of the Sea Cucumber Holothuria Tubulosa (Echinodermata, Holothuroidea) in the Mediterranean: Implications for Management and Domestication." *Frontiers in Marine Science* 9: 1029147.

Purcell, S. W., C. Conand, S. Uthicke, and M. Byrne. 2016. "Ecological Roles of Exploited Sea Cucumbers." In *Oceanography and Marine Biology*, 375–394. CRC press.

Rakaj, A., and A. Fianchini. 2024. "Mediterranean Sea Cucumbers – Biology, Ecology, and Exploitation." In *The World of Sea Cucumbers*, 753–773. Elsevier.

Rue, H., S. Martino, and N. Chopin. 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, no. 2: 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x.

Schneider, K., J. Silverman, B. Kravitz, et al. 2013. "Inorganic Carbon Turnover Caused by Digestion of Carbonate Sands and Metabolic Activity of Holothurians." *Estuarine, Coastal and Shelf Science* 133: 217–223.

Serra, L., M. Saez, J. Mateu, et al. 2014. "Spatio-Temporal Log-Gaussian Cox Processes for Modelling Wildfire Occurrence: The Case of Catalonia, 1994–2008." *Environmental and Ecological Statistics* 21: 531–563.

Sicacha-Parada, J., I. Steinsland, B. Cretois, and J. Borgelt. 2021. "Accounting for Spatial Varying Sampling Effort due to Accessibility in Citizen Science Data: A Case Study of Moose in Norway." *Spatial Statistics* 42: 100446.



1099095x, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.70096 by CochraneItalia, Wiley Online Library on [22/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Simpson, D., J. B. Illian, F. Lindgren, S. H. Sørbye, and H. Rue. 2016. "Going Off Grid: Computationally Efficient Inference for Log-Gaussian Cox Processes." *Biometrika* 103, no. 1: 49–70.

Simpson, D., H. Rue, A. Riebler, T. G. Martins, and S. H. Sørbye. 2017. "Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors." *Statistical Science* 32, no. 1: 1–28.

Toniolo, C., A. Di Sotto, S. Di Giacomo, et al. 2018. "Seagrass *Posidonia oceanica* (L.) Delile as a Marine Biomarker: A Metabolomic and Toxicological Analysis." *Ecosphere* 9, no. 3: e02054.

Ventura, D., L. Castoro, G. Mancini, et al. 2022. "High Spatial Resolution Underwater Data for Mapping Seagrass Transplantation: A Powerful Tool for Visualization and Analysis." *Data in Brief* 40: 107735.

Ventura, D., A. Rakaj, G. J. Lasinio, G. M. Sangiovanni, D. Poggio, and G. Mastrantonio. 2025. "Detecting Habitat Preferences and Monitoring Population Dynamics of Sea Cucumbers in Coastal Ecosystems Through Underwater Photogrammetry." *Journal of Environmental Management* 377: 124589.

Warton, D. I., and L. C. Shepherd. 2010. "Poisson Point Process Models Solve the Pseudo-Absence Problem for Presence-Only Data in Ecology." *Annals of Applied Statistics* 4: 1383–1402.

Yuan, Y., F. E. Bachl, F. Lindgren, et al. 2017. "Point Process Models for Spatio-Temporal Distance Sampling Data From a Large-Scale Survey of Blue Whales." *Annals of Applied Statistics* 11, no. 4: 2270–2297.

## Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Figure S1:** CRPS for each cell and each season using Poisson Regression with GAM for Depth and spatial Gaussian effect. **Figure S2:** Posterior mean residuals for each cell and each season using Poisson Regression with GAM for Depth and spatial Gaussian effect. **Figure S3:** Posterior mean and quantile for Depth using Poisson Regression with GAM for Depth and spatial Gaussian effect. **Figure S4:** Posterior IC 95% for season effect using Poisson Regression with GAM for Depth and spatial Gaussian effect.



1099095x, 2026, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/env.70096 by CochraneItalia, Wiley Online Library on [22/04/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License