A hierarchical Bayesian model to monitor pelagic larvae in response to environmental changes

Alessia Granata[1,5] · Antonino Abbruzzo[2] · Bernardo Patti[3] · Angela Cuttitta[4] · Marco Torri[4]

Abstract

European anchovies and round sardinella play a crucial role, both ecological and commercial, in the Mediterranean Sea. In this paper, we investigate the distribution of their larval stages by analyzing a dataset collected over time (1998–2016) and spaced along the area of the Strait of Sicily. Environmental factors are also integrated. We employ a hierarchical spatio-temporal Bayesian model and approximate the spatial field by a Gaussian Markov Random Field to reduce the computation effort using the Stochastic Partial Differential Equation method. Furthermore, the Integrated Nested Laplace Approximation is used for the posterior distributions of model parameters. Moreover, we propose an index that enables the temporal evaluation of species abundance by using an abundance aggregation within a spatially confined area. This index is derived through Monte Carlo sampling from the approximate posterior distribution of the fitted models. Model results suggest a strong relationship between sea currents' directions and the distribution of larval European anchovies. For round sardinella, the analysis indicates increased sensitivity to warmer ocean conditions. The index suggests no clear overall trend over the years.

1 Introduction

Ichthyoplankton abundance and distribution serve as valuable indicators of the condition and well-being of a marine ecosystem. Fish larval stages are affected by high natural mortality rates, including the impact of predation, resulting in a predominance of eggs and early-stage larvae in ichthyoplankton samples. This

---

Handling Editor: Luiz Duczmal.

Extended author information available on the last page of the article

phenomenon allows us to measure the reproductive capacity of fish species such as anchovies and sardines, offering an index of their relative population size (Allen et al. 2006). Several factors affect the composition and abundance of larval fish communities (Boehlert and Mundy 1993; Patti et al. 2022). An important factor is the spawning strategy adopted by the adult populations, which are linked to topographical characteristics and hydrographical, chemical and biological conditions (Basilone et al. 2006; Patti et al. 2020). Moreover, oceanographic structures, such as currents, fronts or eddies, govern the formation or disruption of the larval assemblages and are responsible for environmental variations and consequently the survival rate of larvae (Bakun 2006; Quinci et al. 2022). Ichthyoplankton is a crucial component of marine food webs, providing a valuable food source for many predators (Hilborn et al. 2003). The availability and abundance of ichthyoplankton affect the recruitment success of the adult parent populations and directly influence the growth and survival of predator populations, ultimately shaping the structure and functioning of marine ecosystems. Thus, ichthyoplankton surveys contribute to our understanding of marine ecosystems. Such surveys can generate an assessment of the standing stock biomass (Ingram Jr et al. 2017). Monitoring ichthyoplankton associated with these fish species allows for increases or decreases in adult fish stocks to be detected more quickly and sensitively than directly monitoring the adults. Furthermore, it is generally easier and cheaper to monitor trends in egg and larval populations than trends in adult fish populations (Matarese 2003).

European anchovies (*Engraulis encrasicolus*, Linnaeus, 1758) and round sardinella (*Sardinella aurita*, Valenciennes, 1847) are small pelagic fish belonging respectively to the families Engraulidae and Dorosomatidae. They are among the most important fisheries resources in many regions of the Mediterranean Sea. From the IREPA (Istituto di Ricerche Economiche per la Pesca e l'Acquacoltura) data of 2009, it emerged that in Italy, the fishing of *Engraulis encrasicolus* represents on average about 26% of the total catch. The *Sardinella aurita* appears to be a vital commercial fish resource, especially for the North African countries bordering the Mediterranean. Furthermore, since the 1990s, its exploitation has continuously increased (National Statistical Service of Hellas, 1990–2002) due to its involvement in food preservation methods and as bait for the profitable tuna and swordfish fishing activities (Tsikliras and Antonopoulou 2006). The monitoring programs of these species have highlighted very pronounced inter-annual biomass fluctuations (Cergole et al. 2002; Patti et al. 2004, 2020). The causes of these oscillations can be multiple and linked to anthropic factors, such as the high fishing effort, and natural factors (Torri et al. 2018). In particular, the biological and environmental dynamics that influence the survival of the first life stages of these species and the subsequent recruitment may be of fundamental importance in determining the annual declines and increases of the adult stock (Cuttitta et al. 2006; Patti et al. 2020; Torri et al. 2023). Furthermore, the knowledge of the reproductive biology of these species gives insights into the impact of fishing efforts on the spawning fraction of the adult populations to avoid the risk of overfishing. The study of the ichthyoplankton phases and their relationship with the environment and other organisms is therefore

of primary importance for providing the necessary information supporting the correct exploitation of fisheries resources.

The last decade has seen a significant increase in the number of studies focusing on ichthyoplankton communities. These surveys are widely recognized as essential tools for understanding the trophic dynamics and variations of commercially valuable fish populations. Consequently, the results of these studies are invaluable for developing stock assessments and fisheries management plans (Boeing and Duffy-Anderson 2008). This can be attributed to the significant impact of habitat conditions on fish survival during the early stages of their life, from eggs to juveniles. These conditions strongly influence the success of recruitment and, consequently, the size of the adult population (Bakun 2006; Basilone et al. 2013; Patti et al. 2020). The study of ichthyoplankton significantly impacts economic fisheries policies, providing valuable information for effective management and sustainable utilization of fisheries resources. Policymakers can assess the effectiveness of conservation measures and implement sustainable fisheries management practices by monitoring ichthyoplankton populations.

The paper's main objective is to estimate the spatio-temporal distribution of the larvae of European anchovy and round sardinella in the Strait of Sicily during the summer periods from 1998 to 2016. This involves utilizing available data on ichthyoplankton populations, environmental variables, and relevant factors to establish statistical models that can estimate the abundance of larvae in a given area and/or time. The analysis aims to identify the key variables influencing larval abundance, such as temperature, salinity, currents and others, and incorporate them into a predictive model. By examining the relationships between larval abundance and environmental factors, the analysis provides insights into the drivers of larval population dynamics and potential variations over time. This information can contribute to a better understanding of the factors influencing larval survival, dispersal, and recruitment into adult populations. In addition to predicting larval abundance, the analysis aims to develop an abundance index. This index is a quantitative measure of larval population size, allowing for comparisons across different periods. The development of an abundance index provides a standardized metric that can be used to track changes in larval abundance over time and assess the relative health and productivity of fish stocks. By establishing an abundance index, the analysis aims to provide a practical tool for monitoring larval populations and useful information supporting the scientific advice for the sustainable management of two crucial small pelagic fisheries resources. Hierarchical spatio-temporal Bayesian models are utilized to estimate the spatio-temporal distribution of the larvae and develop an abundance index. These models offer a robust and comprehensive framework for studying the spatio-temporal distribution of larvae, and enable an explicit stochastic framework to account for the underlying dependence between observations (Campbell et al. 2017). Biological data are often characterized by a spatio-temporal structure, as species biomass and availability continuously change in space over time (Zhou et al. 2019). Therefore, spatial and temporal correlation must be considered during the modelling process because observations of species in geographically close locations are subject to similar life habits and environmental characteristics. Typically, ichthyoplankton sampling data are records of a specific vessel at a specific time and place. In addition,

these data often show an excess of zero counts, which can arise due to various factors such as species rarity, localized distributions or challenges in detection methods. This phenomenon, known as zero inflation, poses additional challenges to traditional modelling techniques (Zuur et al. 2009; Wenger and Freeman 2008; Agarwal et al. 2002). Spatial models using hierarchical approaches are known to work well for this type of nested data (Izquierdo et al. 2022; Lezama-Ochoa et al. 2020; Izquierdo et al. 2021; Cavieres and Nicolis 2018). Several authors have applied Hierarchical spatio-temporal Bayesian models using the Integrated Nested Laplace Approximation (INLA) (Rue et al. 2009) to standardize species abundance indices (Cao et al. 2011). Hierarchical spatio-temporal Bayesian models have an advantage over standard models (e.g. GLM or GAM) as they account for spatio-temporal autocorrelation through spatially structured random effects and autoregressive terms, thus reducing the uncertainty of estimated abundance indices (Zhou et al. 2019). Furthermore, it is essential to highlight that the hierarchical spatio-temporal Bayesian models also allow for the inclusion of smoothed (non-linear) terms for environmental covariates (e.g. sea surface temperature, chlorophyll-a, bathymetry, etc.), which may be crucial in explaining the spatio-temporal distribution and abundance (Muñoz et al. 2013).

2 Materials

The data used in this analysis were provided by the National Research Council of Italy (CNR), and it is a collection of detailed information on European anchovy and round sardinella larvae obtained from sampling over time and space along the area of the Strait of Sicily in summer surveys from 1998 to 2016. The study area is the Strait of Sicily (southern Sicilian coast, see Fig. 1), covering a surface of about 25000 km^2. It is a relatively narrow waterway that separates the island of Sicily (Italy) from the coast of North Africa, connecting the Tyrrhenian Sea in the north with the Mediterranean Sea in the south. With its strategic location, the Strait plays a significant role in maritime trade and transportation between Europe, North Africa, and the Middle East.

Oceanographic data and ichthyoplanktonic samples were collected in nineteen annual oceanographic surveys carried out during the summer period on board of the R/V " Urania" (1998–2014) and the R/V " Minerva Uno" (2015–2016).

Sampling, in the 19 summer surveys carried out from 1998 to 2016, was based on a systematic station grid of 6×6 nautical miles on the continental shelf (bottom depth ≤ 200 m) and a grid of 12×12 nautical miles for the off-shore areas with a bottom depth greater than 200 m (Fig. 2). Stations were sampled twenty-four hours a day to minimize bias in the catch of fish species during the larval stage, which typically may show relatively large diel vertical migration patterns. Plankton samples were collected using a bongo net (Bongo40, with a 40 cm opening) towed from the straight side of the ship at a speed of 2 knots. The Bongo 40 drops are oblique, carried out from the surface to 100 m depth or 5 ms from the bottom in shallower stations, and equipped with a 200 μm mesh size net. A General Oceanics 'flowmeter' is mounted in each mouth's centre to measure the filtered water volume (Patti et al.
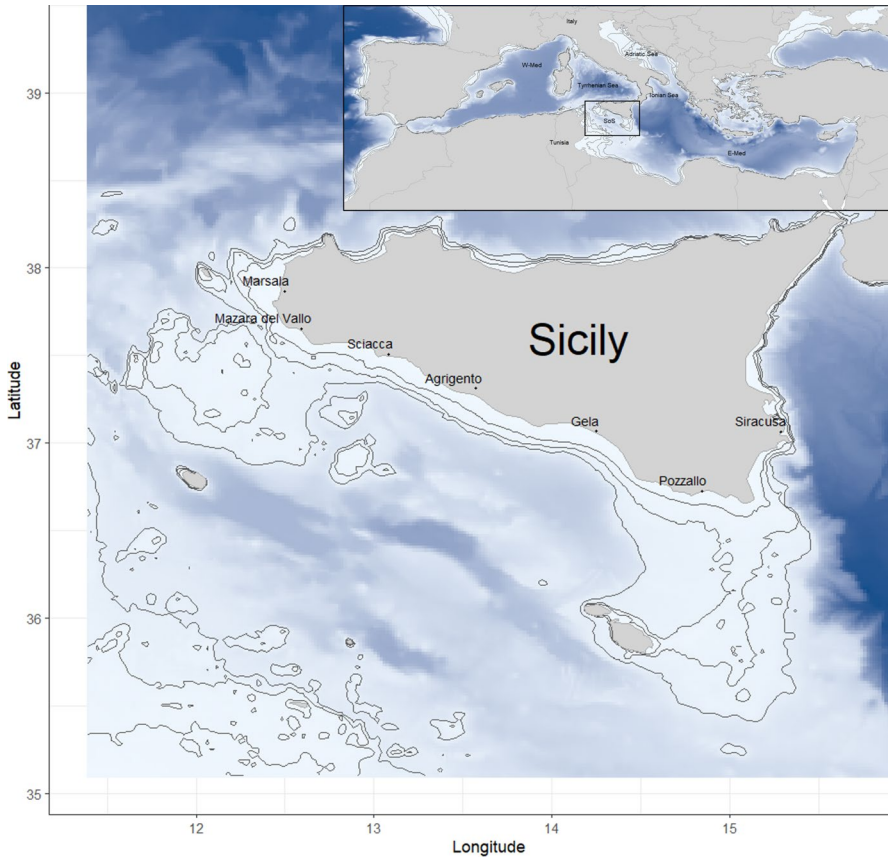
Fig. 1 Map of the Mediterranean Sea (upper-right panel) and the Strait of Sicily showing the study area. The bathymetry is reported in the scale of blue, where the darker the blue, the more deep the sea

2013). All the samples used were from the same side and cod-end collector of the Bongo.

Samples were immediately fixed after collection and preserved in a 10% buffered-formaldehyde (and/or 70% alcohol) and sea-water solution for further sorting in the laboratory by stereomicroscope. Larvae of European anchovy and round sardinella were selected from the rest of the plankton and identified according to Whitehead et al. (1988). The larval counts are used as a measure of abundance.

Environmental factors were selected to investigate their influence on the spatio-temporal distribution of European anchovy and round sardinella larvae. Continuous vertical profiles of environmental down-cast data were acquired by a SBE 11 plus CTD multiparameter probe at all Bongo40 plankton stations to characterize the physical properties of the water column. They were quality-checked and processed according to the Mediterranean and Ocean Data Base instructions using the Sea-Bird Scientific Seasoft V2 software. From the available environmental information
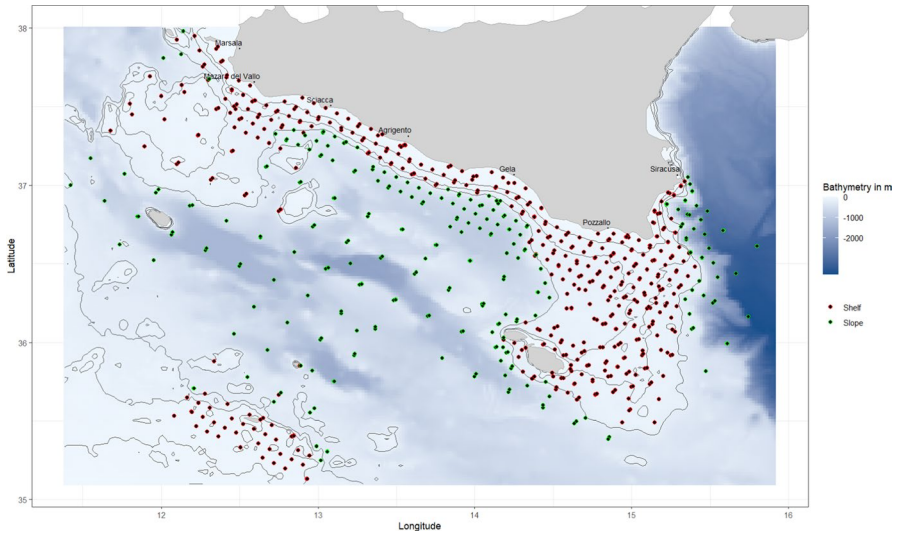
Fig. 2 Map of sampling station locations in the 19 summer surveys carried out from 1998 to 2016, represented with rounds. Stations with a bottom depth ≤ 200 m (" Shelf") are represented in red, and stations with a bottom depth > 200 m (" Slope") are visualized in green. The bathymetry is reported in the scale of blue, where the darker the blue, the more deep the sea

collected in CTD casts, the variables used in this study were "*Bottom Depth*", "*Temperature*" and "*Salinity*". Most of the larvae in the Mediterranean Sea during the summer are mainly concentrated in the upper mixed layer (Sabatés et al. 2007); in this study, the average values of the measurements recorded in the first 10 ms of the water column were used, as they can be considered representative of the surface conditions. Then, the Mixed Layer Depth (MLD, in m) is an additional parameter used in the data analysis. Its value was derived from each CTD profile using the algorithm based on water density calculation described in Kara et al. (2000), as it is an essential factor in defining the potential spawning habitat (Planque et al. 2007). The impact of mesoscale oceanographic features such as upwelling, cold filaments and fronts on the spatial distribution of ichthyoplankton (Torri et al. 2018; Patti et al. 2020) was also considered. Oceanographic structures can influence the distribution of chemical and physical properties of the water column (Placenti et al. 2022), potentially affecting larval survival and development (Falco et al. 2020; Cuttitta et al. 2022; Torri et al. 2021). Therefore, the surface circulation characteristics were assessed using satellite-based Absolute Dynamic Topography (ADT, in cm) data (daily data; spatial resolution: 0.125×0.125 degree) and the derived u and v components of geostrophic currents provided by Copernicus Marine Environment Monitoring Service (CMEMS, http://marine.copernicus.eu/). ADT data represent critical oceanographic features such as mesoscale eddies and meanders (Pujol and Larnicol 2005). These can influence primary production, act as physical barriers to larval distribution, or be responsible for offshore dispersal. Furthermore, the absolute Geostrophic Current Speed (GCS, in cm s^{-1}) was derived from the zonal (u) and meridional (v) components of the surface current and used as an additional potential

predictor in the subsequent modelling approach. Finally, information on the sea surface chlorophyll-*a* concentration (Chl-*a*, in mg m^{-3}) was also used, as Chl-*a* is a good proxy for primary productivity (Joint and Groom 2000) and indirectly can help to represent favourable feeding conditions for larvae. For this purpose, high-resolution (1 × 1 km) daily satellite data available for download from CMEMS was used. For all satellite information, available data was extracted for each plankton station included in the analysis based on the spatial and temporal location of the associated sampling hauls.

3 Methods

In marine biology research, comprehending the dynamics of fish populations and ecological processes is a significant goal. The spatial distribution of larval fish plays a crucial role in this understanding, as it directly impacts fish populations and ecological interactions. The analysis of ecological data often shows an excess of zero counts and is complex due to non-linear relationships between environmental variables and species abundance. To address these challenges, the analysis approach adopted for this study is based on hierarchical spatio-temporal Bayesian models.

3.1 Hierarchical Bayesian model

To estimate the abundance of larvae of European anchovy and round sardinella in the Strait of Sicily, we define a hierarchical spatio-temporal Bayesian model. Let $\{Y_{it}, s_i \in \mathcal{D}, t \in \mathcal{T}\}$, with $\mathcal{D} \in \mathbb{R}^2$ and $\mathcal{T} \in \mathcal{N}^+$, be a stochastic process representing the number of larvae at location s_i and time t. We use the index i to denote the generic spatial point s_i to simplify the notation. A high proportion of zero values characterizes this stochastic process, suggesting a hurdle model. This model should accommodate zero and nonzero values as an integrated process of two levels. The first level presents a latent binary component Z_{it} that generates zeros and ones, i.e.

$$Z_{it} \sim Ber(\pi_{it}),$$

where π_{it} is the probability of observing one. Hence,

$$z_{it} = \begin{cases} 1 & \text{with prob } \pi_{it}, \\ 0 & \text{with prob } 1 - \pi_{it}. \end{cases}$$

The second level assumes a Poisson distribution for the conditional stochastic process $Y_{it} \mid (Z_{it} = 1)$, i.e.

$$Y_{it} \mid (Z_{it} = 1) \sim Pois(\lambda_{it}).$$

We then use a Generalized Additive Model (GAM) to relate the explanatory variables with $f_1(\pi_{it})$ and $f_2(\lambda_{it})$, where f_1 is the logit and f_2 is the log function.

Thus, we can define a hierarchical spatio-temporal Bayesian model as:

$$Z_{it} \sim Ber(\pi_{it}) \qquad \text{logit}\,(\pi_{it}) = \alpha_0' + \sum_{j=1}^{J} \beta_j' q_{itj}' + \sum_{k=1}^{K} f'(x_{itk}') + w_{it}',$$

$$Y_{it} \mid (Z_{it} = 1) \sim Pois(\lambda_{it}), \qquad \log(E_{it}\lambda_{it}) = \eta_{it} = \alpha_0 + \sum_{j=1}^{J} \beta_j q_{itj} + \sum_{k=1}^{K} f(x_{itk}) + w_{it},$$

where $E_{it} = (depthgauge_{it}/filteredwatervolume_{it})$ is included as an offset, α_0' and α_0 are the intercepts, β' and β are the regression coefficients of the covariates $\mathbf{Q}$ and $\mathbf{Q}'$ (where $\mathbf{Q}$ and $\mathbf{Q}'$ can be the same set), the $f'(\cdot)$ and $f(\cdot)$ allow fitting of any possible non-linear relationship (of the environmental variables $\mathbf{X}$ and $\mathbf{X}'$), as first-order or second-order random walk processes (RW1 or RW2). All the environmental variable values were aggregated in six time knots (i.e. group increments based on quantiles). Finally w_{it}' and $w_{it} = \delta w_{it}'$ represent the spatio-temporal structure of the model, where δ is the scaling parameter for w_{it}', which is the spatio-temporal structure considered for the presence part of the model and is also shared for the abundance part (Rue et al. 2009).

The hierarchical spatio-temporal Bayesian model can also be modified in at least two directions. Firstly, one could share the complete linear predictor of the Poisson process with the linear predictor of the Bernoulli distribution. This sharing would imply that the higher the number of captures, the higher the probability of presence, and vice versa, the lower the number of captures, the lower the likelihood of presence (Krainski et al. 2018). On the other hand, it is also possible to share several elements of the latent field with a different scaling coefficient for each one (Paradinas et al. 2017; Krainski et al. 2018).

In this work, we first compared several models that have no covariates but only share a spatial-temporal component and differ from each other with respect to the temporal component. Specifically, three different temporal dependencies were considered. The first model ($Mod0_{iid}$) assumes that the spatial random effects are independent of time. The second model ($Mod0_{ar1}$) assumes that the latent process changes in time with a first-order autoregressive process (AR1). The third model ($Mod0_{rw1}$) assumes that the latent process changes in time according to a random walk of order 1 (RW1).

In the Supplementary material section, Table 6 shows the results of the first two models considering as response variable the European anchovy larvae abundance, while Table 7 for the round sardinella larvae. These models are evaluated through different goodness-of-fit indices. Tables do not show the values relating to the model based on a first-order random walk error process (RW1) because this model encounters difficulties in achieving convergence.

After model comparison, we assume that the spatial random effects are independent of time such that $w_{it} = \xi_{it}$, where ξ_{it} is a zero-mean Gaussian Process (GP), assumed to be temporally independent and characterized by the following spatio-temporal covariance function:

$$Cov(\xi_{it}, \xi_{ju}) = \begin{cases} 0 & \text{if } t \neq u \\ \frac{\sigma_u^2}{\Gamma(v)2^{v-1}}(k||\mathbf{s}_i - \mathbf{s}_j||)^v K_v(k||\mathbf{s}_i - \mathbf{s}_j||) & \text{if } t = u, \end{cases} \tag{1}$$

for $i \neq j$, where (1) is the Matérn spatial covariance function, σ_u^2 is the marginal variance of the process, $v > 0$ is the smoothing parameter, $k > 0$ is a scale parameter, $||\mathbf{s}_i - \mathbf{s}_j||$ is the Euclidean distance between $\mathbf{s}_i$ and $\mathbf{s}_j$ and K_v is the modified Bessel function of second kind and order $v > 0$. This choice is further grounded in the recognition that the inter-annual spatial variability in the distribution of larvae of small pelagic species is typically exceptionally high, as it is driven by highly variable factors such as the spawning stock biomass, the selection of spawning areas and the mortality of the early life stages.

Denote the parameter vector as $\boldsymbol{\theta} = (\alpha_0, \alpha_0', \boldsymbol{\beta}, \boldsymbol{\beta}', \boldsymbol{f}, \boldsymbol{f}', w')$ and the hyperparameter vector as $\boldsymbol{\psi} = (\tau, k, \tau_f, \tau_f', \delta)$. The model just described is a three-level hierarchical Bayesian model with a latent Bernoulli structure, and a Gaussian Random Field (GRF) can represent the joint distribution of the model parameters. Then, the GRF, a continuous spatial process, is approximated by a spatial process with a discrete index (i.e. a Gaussian Markov Random Field (GMRF)). Thus, thanks to the sparsity of the precision matrix of such a GMRF, which is induced by the conditional independence structure of the process, appropriate computation techniques for sparse matrices can be used. Finally, the prior parameter of the hyperparameters should be specified for a full Bayesian inference. Since we have no prior information on the hyperparameters, a vague zero-mean Gaussian prior distribution ($N(0, \tau = 0, 001)$ is used for the parameters $\alpha, \alpha', \boldsymbol{\beta}, \boldsymbol{\beta}', \delta$. Furthermore, we compare different models by considering three priors for the log of precision parameters τ_f and τ_f'. Specifically, we considered a *logGamma*, a flat distribution, and PC-prior with parameters $u = 1$ and $\alpha = 0.01$. The model shows variations in the estimates of the precision parameters τ_f and τ_f' for which the prior distribution was changed, highlighting sensitivity to the choice of priors. Conversely, other parameters remain stable to changes in priors. Finally, we opted for the *logGamma* distribution as it exhibits lower LCPO values, which are relevant to the model's predictive capabilities. The detailed results of this analysis are provided in the Supplementary material in Tables 8, 9 and 10.

To summarize, for our model, a *logGamma*$(1, 10^{-5})$ is assumed on the logarithm of the precision parameters τ_f and τ_f' of environmental variables $\mathbf{X}$ and $\mathbf{X}'$, and a $\mathcal{N}(0, 1)$ is assumed on the logarithm of the precision τ and the logarithm of k. Hence, the posterior marginal distributions for each component of $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ can be efficiently estimated using INLA (for further details, see Rue et al. (2009)), and the computations can be conducted using the `INLA` package in R software.

3.2 Model selection

As in classical statistics, the Bayesian approach also requires indices to compare the goodness of fit of several models, which may differ in the prior distributions of the various parameters or the explanatory variables included in the model. Two commonly employed criteria are the Deviance Information Criterion (DIC) and the

Conditional Predictive Ordinate (CPO). The DIC, proposed by Spiegelhalter et al. (1998), represents the generalization of the Akaike Information Criterion (AIC) into the Bayesian domain. It comprises two components: the posterior expectation of the deviance, which measures the goodness of fit of the model, and the effective number of parameters, which captures the complexity of the model. Lower values of DIC indicate superior model performance. The CPO, introduced by Pettit (1990), on the other hand evaluates predictive ability by assessing how well the model predicts individual observations. Similarly to cross-validation, it quantifies the probability of observing a particular data point given the rest of the data. Empirical approximations enable its calculation without repeatedly fitting the model. The mean logarithmic CPO, suggested by Roos and Held (2011), summarizes all the information from individual CPO values. Lower values of the LCPO, i.e. the mean logarithm of the CPO (Gneiting and Raftery 2007) indicate a better model.

3.3 A Bayesian approach to construct an abundance index

This section illustrates the approach used to calculate the larval abundance index. The abundance index of larval fish is a quantitative measure used to assess the abundance, or relative density, of larval fish in a particular aquatic environment or during a specific time. This index is a crucial tool in fisheries science, marine ecology, and environmental monitoring as it provides insights into the early life stages of fish populations and the overall health of marine ecosystems. The most straightforward measure of population status is an estimate or index of the abundance at any given time. A series of estimates through time can be used to evaluate the population's trend (i.e., increasing, declining, stable) (Etienne et al. 2010).

The process of generating the abundance index involves the following steps:

1. Selection of a suitable spatio-temporal model for analyzing larval abundance;
2. Estimation of the parameters' posterior distributions $p(\theta \mid y)$;
3. Monte Carlo sampling from the approximate distribution of the fitted model $\theta^{(m)} \sim p(\theta \mid y)$, obtaining n samples and corresponding functions;
4. Collection of the index for the linear predictor corresponding to the stack data of the prediction scenario;
5. Extraction of the corresponding elements of the latent field of each sample;
6. Projection of estimates to the centroid of an equally spaced grid in the area of interest;
7. Drawn $y^{\star}$ from the posterior distribution $y^{\star(m)} \sim p(y^{\star} \mid y)$. This is performed by randomly drawing from the sampling distribution with the parameter draw plugged in;
8. Summation of randomly drawn values $y^{\star}$ across grid cells to generate an annual abundance index;
9. Repeat the process across the posterior samples;
10. Computation of posterior credible intervals for the abundance of each species in each year.
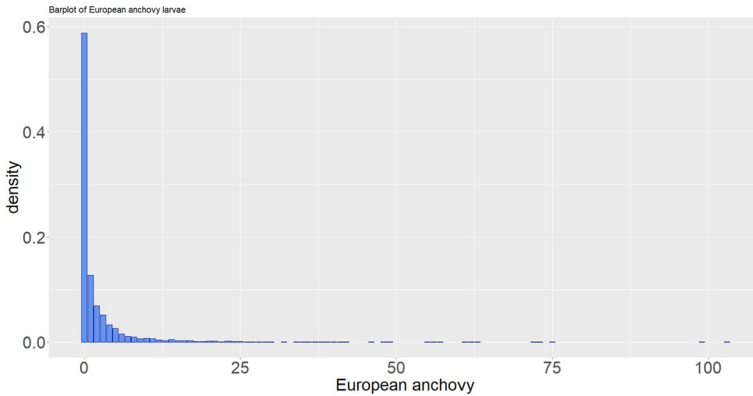
Fig. 3 Barplot of the European anchovy larvae
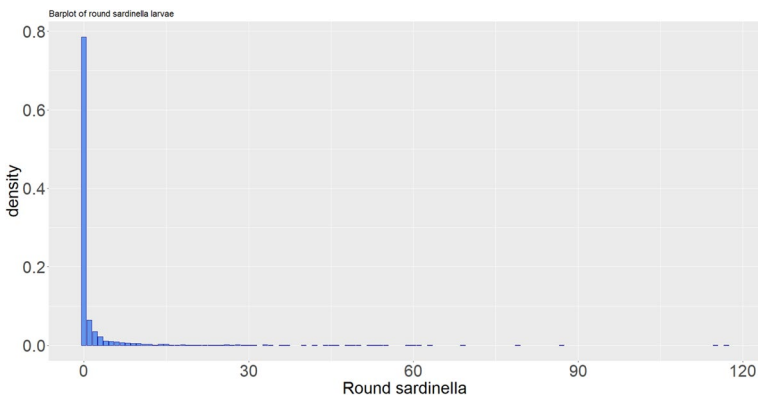

Fig. 4 Barplot of the round sardinella larvae

4 Results

Here, we present the results of the analysis of spatio-temporal models based on a Bayesian approach implemented through the Integrated Nested Laplace Approximation (INLA). The tools presented previously are used to analyze the distribution of European anchovy and round sardinella larvae in the Strait of Sicily during the study period, to understand their interactions with the surrounding environmental variables and, above all, to calculate the larval abundance index for both species. Since only a few observations of fish larvae were recorded in 1999, they were excluded from the analyzed time series (1998–2016). Note that a data cleaning operation was performed to remove any potential duplicate records, ensuring accurate data processing. Subsequently, an exploratory data analysis was conducted. This latter analysis suggested an excess of zeros for both species, with a value of 59% for anchovies and 79% for sardines as we can see in Figs. 3 and 4.

After the exploratory data analysis, several models were compared in terms of the DIC as the criterion for goodness of fit, and the leave-one-out cross-validated CPO score computed by the LCPO as a predictive quality measure. For both these measures, the smaller the score, the better the model. In addition, we worked closely with the biologists we collaborated with to determine the most appropriate covariates to include in the models.

4.1 European anchovy larvae

The presence and abundance of European anchovy are influenced by environmental factors, including chlorophyll-a and u and v-components of currents. Several models were compared; the most significant are reported in Table 1.

These models are similar in terms of both DIC and LCPO. However, we carry out the analysis on Model 3 in Table 1 which, for the random variable Y (representing the abundance of the species), incorporates the linear effects for the u and v components of the currents and their interaction, as well as chlorophyll-a. Only the intercept and the random component w' are considered for the dichotomous random variable Z, representing the presence or absence of the species. We also explored the influence of covariates on the presence probability as well. However, the available covariates did not show statistically significant effects on the presence probability.

$$
\begin{cases}
\text{logit}(\hat{\pi}_{it}) = \hat{\alpha}'_0 + w'_{it}, \\
\log(E_{it}\hat{\lambda}_{it}) = \hat{\alpha}_0 + \hat{\beta}_1 v_{it} + \hat{\beta}_2 u_{it} + \hat{\beta}_3 u_{it} * v_{it} + \hat{\beta}_4 chl_{it} + w_{it}.
\end{cases}
\tag{2}
$$

Figure 5 displays the graphical representation of the selected model (2) using a Directed Acyclic Graph (DAG). Table 2 presents the summary statistics of the parameters' posterior distributions.

Model (2) highlights that the currents (represented by the main effects u and v and the interaction term $u * v$) and chlorophyll-a have a significant effect on the abundance of European anchovy larvae. In particular, the analysis highlights the importance of the direction of the currents. The interaction between the v and u components of geostrophic ocean currents has a notable impact. We can say that

Table 1 Model selection process for the European anchovy larvae (*Engraulis encrasicolus*) in the Strait of Sicily in 1998 and from 2000 to 2016

Model ID	Random variable	Linear predictor	DIC	LCPO
1	Z_{it}	$\alpha'_0 + w'$	3214.624	0.532
	$Y_{it} \mid Z_{it} = 1$	$\alpha_0 + \beta_1 v + w$	5791.886	15.754
2	Z_{it}	$\alpha'_0 + w'$	3214.771	0.532
	$Y_{it} \mid Z_{it} = 1$	$\alpha_0 + \beta_1 u + \beta_2 v + \beta_3 u * v + w$	5791.281	15.697
3	Z_{it}	$\alpha'_0 + w'$	3217.253	0.532
	$Y_{it} \mid Z_{it} = 1$	$\alpha_0 + \beta_1 u + \beta_2 v + \beta_3 u * v + \beta_4 chl + w$	5789.400	15.676

α_0 and α'_0 are the intercepts, w' and w are spatio-temporal structures, $v = v$ component of the currents, $u = u$ component of the currents and chl = chlorophyll-a
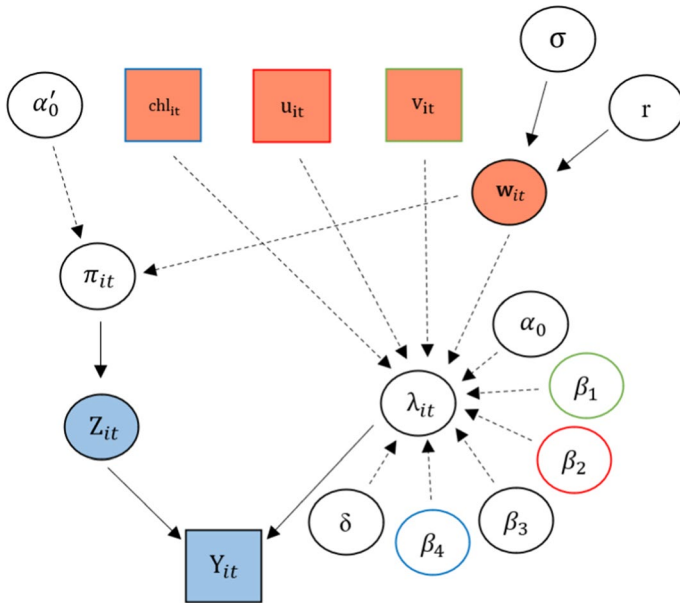
Fig. 5 Graphical structure of model (2) by a Directed Acyclic Graph (DAG). Squares denote the observed quantities, and circles denote the latent variables. A probability distribution characterizes each parameter. The arrows connecting the nodes represent stochastic (solid) and deterministic (dashed) dependencies

Table 2 Posterior estimates of the parameters (mean, standard deviation (sd) and quantiles) for the two-part model (2)

Summary statistics	logit $(\hat{\pi})$			$\log(E\hat{\lambda})$					
	$\hat{\alpha}'_0$	spatial sd	range	$\hat{\alpha}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\delta}$
mean	-0.749	4.235	0.357	0.155	-1.178	0.695	8.147	-2.128	0.627
sd	0.095	1.092	1.039	0.096	0.522	0.466	3.662	0.797	0.026
2.5%	-0.939	3.556	0.330	-0.035	-2.203	-0.219	0.960	-3.692	0.575
50%	-0.747	4.235	0.357	0.155	-1.178	0.696	8.148	-2.128	0.627
97.5%	-0.565	5.044	0.385	0.340	-0.153	1.608	15.325	-0.565	0.678
mode	-0.745	4.235	0.358	0.157	-1.179	0.696	8.151	-2.128	0.627

the abundance of European anchovy larvae increases when moving from north to south and west to east. Concerning chlorophyll-a, the mean posterior fixed effect on the abundance shows higher numbers of anchovies as the chlorophyll-*a* decreases. These results, occasionally counterintuitive when considering chlorophyll-*a* as a proxy for food availability, align with prior studies focused on the larval dispersion of this species in the Strait of Sicily (Lafuente et al. 2002; Cuttitta et al. 2006; Torri et al. 2018; Patti et al. 2020). These authors have underscored the role of physical forces, such as currents, gyres, and fronts, in concentrating larvae born along the

Sicilian coast in a larval retention area positioned to the southeast, precisely in the Capo Passero region. In contrast to the spawning areas situated in the northwestern coastal zone and characterized by elevated chlorophyll-*a* values associated with the upwelling phenomenon, the Capo Passero area typically exhibits a relatively lower concentration of food, although characterized by higher larval densities compared to the other areas (Russo et al. 2021; Torri et al. 2023). The model thus encapsulates a noteworthy pattern concerning the distribution of this species in the study area, namely the presence of larval concentration areas characterized by lower food availability compared to other egg deposition areas. The fundamental role of currents is further accentuated by the significant effect of the *u* and *v* components of the currents. The δ parameter is significantly different from zero, meaning that the absence/presence and abundance share the same spatial pattern. Concerning the spatial random effects, when setting the α parameter of the Matérn function to 2, the range can be defined as $r = \sqrt{8}/k$. The mean posterior value for the spatial effect range is 0.357°, while the standard deviation is 4.235. This indicates that the correlation becomes almost null at approximately 0.357° (around 42 km).

4.2 Round sardinella larvae

The presence and abundance of *Sardinella aurita* are influenced by several environmental factors, including salinity, temperature, bottom depth and *v*-component of currents. Again, several models were tested, and the most promising are reported in Table 3.

These models are similar in terms of both DIC and LCPO. Nevertheless, we carry out the analysis on Model 4 of Table 3, which incorporates linear effects for the *v* component of the currents, temperature and salinity for the random variable *Z* (presence/absence of the species). About the random variable *Y* (number of *Sardinella aurita*), a linear effect is considered for the *v* component of currents and a smoothed effect (RW2—random walk of second order) for the temperature. For salinity and

Table 3 Model selection for the round sardinella larvae (*Sardinella aurita*) in the Strait of Sicily in 1998 and from 2000 to 2016

Model ID	Random variable	Linear predictor	DIC	LCPO
1	Z_{it}	$\alpha'_0 + \beta'_1 temp + w'$	2533.100	0.405
	$Y_{it} \mid Z_{it} = 1$	$\alpha_0 + f_{s2}(temp) + w$	3317.661	18.451
2	Z_{it}	$\alpha'_0 + \beta'_1 temp + \beta'_2 sal + w'$	2536.992	0.403
	$Y_{it} \mid Z_{it} = 1$	$\alpha_0 + f_{s1}(temp) + f_{s2}(sal) + w$	3293.980	17.848
3	Z_{it}	$\alpha'_0 + \beta'_1 temp + \beta'_2 sal + \beta'_3 v + w'$	2537.365	0.404
	$Y_{it} \mid Z_{it} = 1$	$\alpha_0 + f_{s2}(temp) + f_{s1} sal + v_l + w$	3275.947	17.600
4	Z_{it}	$\alpha'_0 + \beta'_1 temp + \beta'_2 sal + \beta'_3 v + w'$	2535.325	0.403
	$Y_{it} \mid Z_{it} = 1$	$\alpha_0 + f_{s2}(temp) + f_{s1}(sal) + \beta_1 v + f_{s1}(B) + w$	3274.247	17.241

α_0 and α'_0 are the intercepts, w' and w are spatio-temporal structures, *temp* = temperature, *sal* = Salinity, v = v component of the currents and B = bottom depth. Subscript letters represent modelling effects with smoothed terms *s*1 (*RW1* random walk of first order), and *s*2 = smoothed (*RW2* random walk of second order)
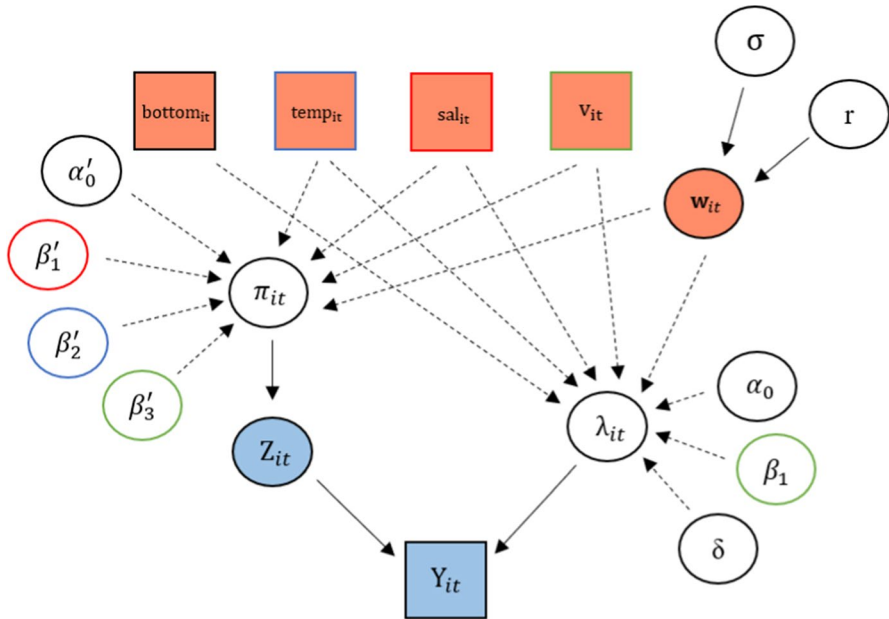
Fig. 6 Graphical structure of model (3) using a Directed Acyclic Graph (DAG)

Table 4 Posterior estimates of the parameters of Model (3) (mean, standard deviation (sd) and quantiles) for the random variable Z (presence/absence of a species)

Summary statistics	logit $(\hat{\pi})$					
	$\hat{\alpha}'_0$	$\hat{\beta}'_1$	$\hat{\beta}'_2$	$\hat{\beta}'_3$	spatial sd	Range
Mean	− 24.045	0.428	0.232	− 3.781	3.349	0.361
sd	6.583	0.182	0.054	0.920	1.185	1.063
2.5%	− 36.909	0.068	0.128	− 5.601	2.416	0.322
50%	− 24.063	0.428	0.231	− 3.776	3.337	0.361
97.5%	− 11.078	0.784	0.339	− 1.989	4.717	0.409
Mode	− 24.100	0.430	0.231	− 3.766	3.303	0.359

bottom depth, the model incorporates an RW1, i.e. a random walk of first order. We represent the selected model in (3) and via a DAG in Fig. 6.

$$\begin{cases} \text{logit}\,(\hat{\pi}_{it}) = \hat{\alpha}'_0 + \hat{\beta}'_1 sal_{it} + \hat{\beta}'_2 temp_{it} + \hat{\beta}'_3 v_{it} + w'_{it}, \\ \log(E_{it}\hat{\lambda}_{it}) = \hat{\alpha}_0 + \hat{\beta}_1 v_{it} + f(bottom_{it}) + f(sal_{it}) + f(temp_{it}) + w_{it}. \end{cases} \quad (3)$$

Posterior estimates (mean, standard deviation (sd) and quantiles) of the parameters are reported in Tables 4 and 5. The former are estimates for the random

Table 5 Posterior estimates of the parameters of Model (3) (mean, standard deviation (sd) and quantiles) for the random variable Y (abundance of species)

Summary statistics	$\log(E\hat{\lambda})$					
	$\hat{\alpha}_0$	$\hat{\beta}_1$	$\hat{\tau}_{bottom}$	$\hat{\tau}_{sal}$	$\hat{\tau}_{temp}$	$\hat{\delta}$
mean	− 0.852	− 3.153	378.563	131.377	516.288	0.826
sd	0.186	0.787	174.223	29.904	116.431	0.052
2.5%	− 1.190	− 4.702	144.097	82.279	324.753	0.722
50%	− 0.863	− 3.151	344.412	128.100	503.646	0.826
97.5%	− 0.464	− 1.613	811.692	199.446	781.010	0.929
mode	− 0.893	− 3.147	284.378	121.787	479.285	0.827

variable presence/absence (Z), and the latter are estimates for the random variable number of *Sardinella aurita*.

Environmental variables, such as *temperature*, *salinity* and the *currents* (*v* component of geostrophic ocean currents), have significant effects on the probability of absence/presence of sardinella. The parameter $\log(\lambda_{it})$ is linearly related to the *v* component. In contrast, it results in an RW1 relationship with the variable *bottom* and *salinity*, and in an RW2 relationship with the *temperature*. It can be seen that the parameter "mean posterior fixed effect of *salinity* on the absence/presence of sardinella' larvae" indicates that the higher the salt concentration in the water, the greater the probability of observing a round sardinella. Furthermore, *salinity* has a non-linear influence on sardine abundance, with abundance non-linearly increasing as salinity increases (Fig. 7 (middle)). This result can be explained by considering the concentration of round sardinella larvae in the southeastern zone (Capo Passero area), typically characterized by relatively higher salinity values. In particular, the non-linearity of the relationship, i.e. a greater effect for salinities above 37.6, highlights the fundamental role of the salinity front in controlling larval dispersion dynamics in this area. Indeed, this oceanographic structure arises from the encounter of the Atlantic Ionian Stream (AIS), typically less saline, with the water of the Ionian Sea and is therefore associated with a concentration of larvae originating from both the Strait of Sicily and the northern Ionian coastal zone (Torri et al. 2023). Concerning the *temperature*, the mean posterior fixed effect on the probability of absence/presence shows a higher probability of observing sardines as the temperature increases with a probability equal to 0.55. This is because the *Sardinella aurita* is a thermophilic species and is, therefore, more likely to be found in warmer areas (Ben-Tuvia 1960). Furthermore, it has a non-linear influence on the abundance. The *v* component, which represents the vertical direction of currents, has a negative linear effect on the probability of observing a sardine and the abundance. For a unit increase in the *v* component, the probability of observing a round sardinella is 0.02. As the *v* component decreases, moving from north to south increases the probability and abundance. The analysis also highlights the importance of geographical location. The presence and abundance of round sardinella are more significant near the coast and decrease moving away from it. This is supported by the bottom
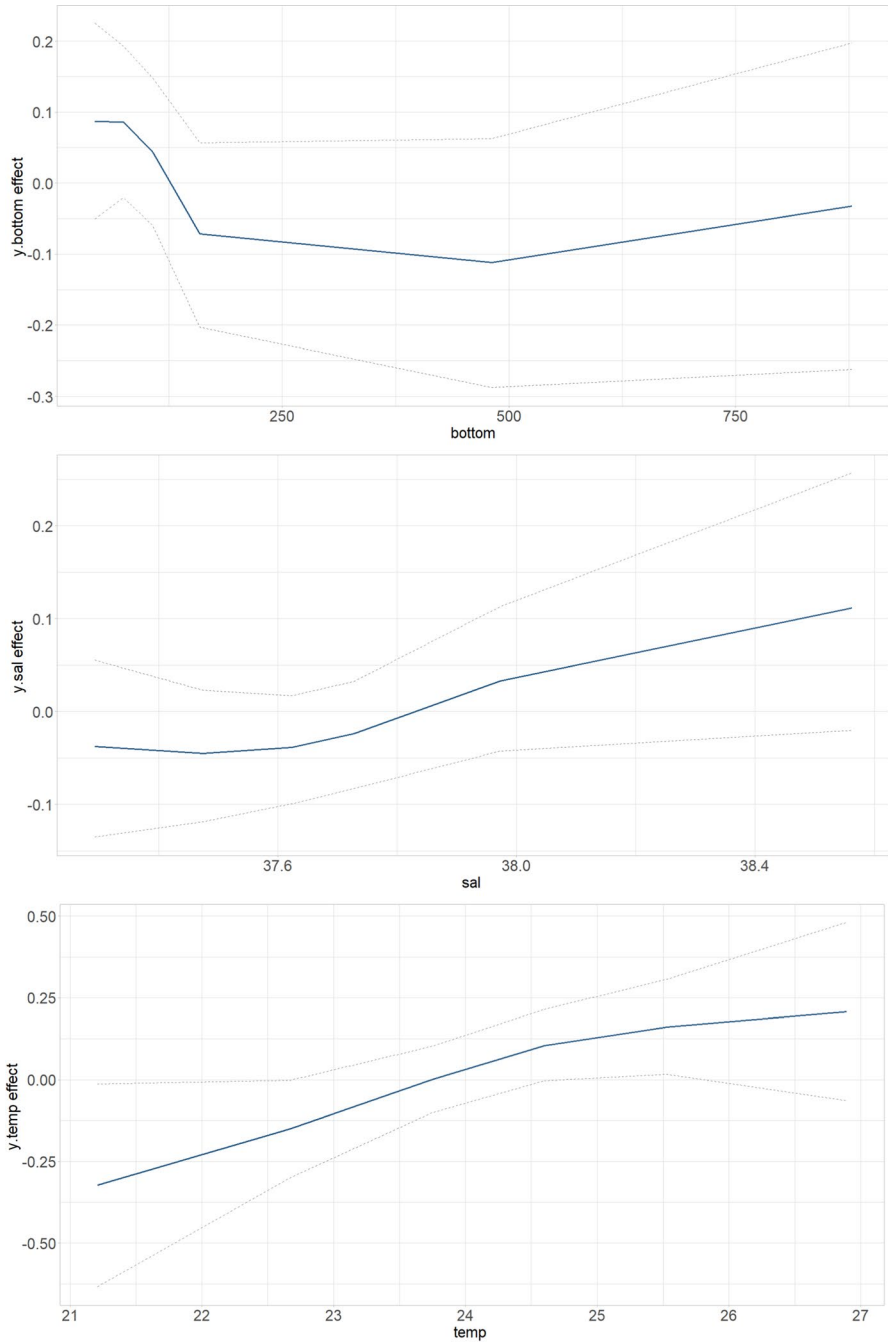
Fig. 7 Marginal smoothed (RW1) effects of bottom (top) and salinity (middle) and smoothed effect (RW2) of temperature (down) for the abundance (Y) of the best model. The dotted lines represent the approximate 95% credibility interval

depth, which influences abundance non-linearly. As we can see from Fig. 7 (top), the effect of the bottom depth is more significant near the coast and decreases up to a depth of about 200 ms, then it remains almost constant to greater depths. This highlights the difference between the continental shelf and the offshore areas. Again, the absence/presence and abundance share the same spatial pattern as the δ parameter is significantly different from zero. Concerning the spatial random effects, the mean posterior value for the spatial effect range is 0.361°, while the standard deviation is 3.349. This indicates that the correlation becomes almost null at a distance of around 40 km.

4.3 Model prediction

Once we have obtained the posterior distributions, we can use them to make predictions for unsampled locations. The predictive distribution is obtained by integrating over the parameter' posterior distribution and the likelihood. It reflects our updated beliefs about future observations based on the observed data and the model. Since the sampling process partially covered the area under study, we decided to focus on a restricted area domain, represented in the blue polygon area in Fig. 8. In other words, the domain of the posterior mean maps is a polygon comprising the minimum area consistently sampled over time.

Predictions were based on daily satellite data of surface temperature, Chlorophyll-*a*, Salinity and direction of currents retrieved from the CMEMS server
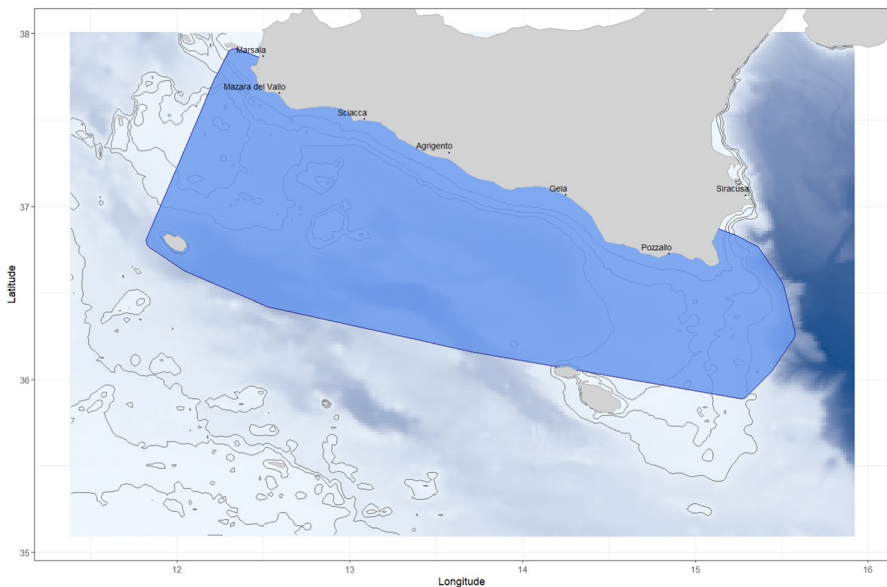

Fig. 8 The blue part of the figure comprises a polygon representing the most sampled area over time. This part was used for making predictions and building the abundance index
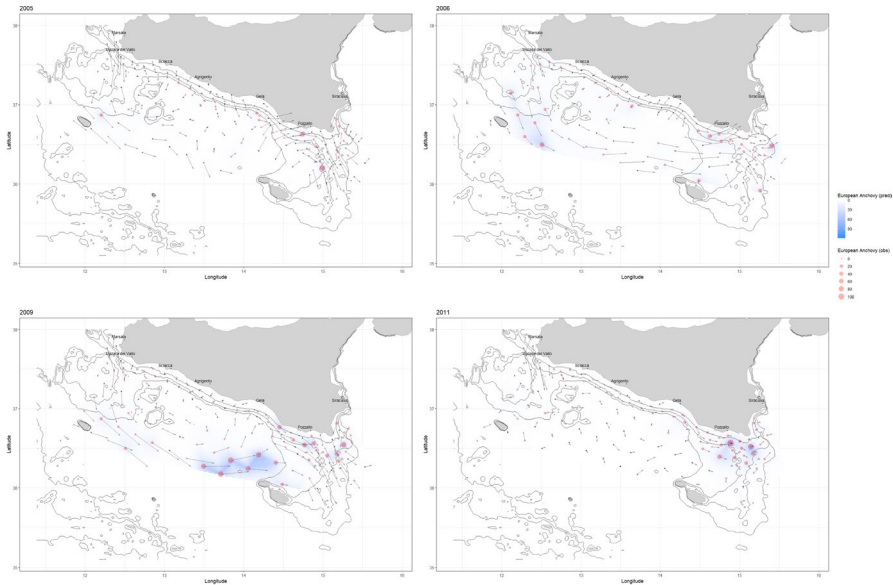
Fig. 9 Map of the predicted values of the European anchovy larvae abundance in 2005, 2006, 2009 and 2011. Red points represent the number of anchovy larvae in the sampling stations. The estimated number of larvae is also reported in a continuous spatial pattern by applying a white-blue colormap, indicating the estimated number of larvae per 4 km^2

(https://www.marine.copernicus.eu/) and projected in an equally spaced 2.5 km grid within the area of interest.

The maps with the most significant results are shown below, while the complete maps are reported in Supplementary Fig 15.

Figure 9 shows the predicted values of the European anchovy larvae abundance in 2005, 2006, 2009 and 2011. There was a high concentration of European anchovies in the southeastern area. In particular, 2006 and 2009 show a substantial increase in the abundance of anchovy larvae in the offshore area (at depths greater than 200 ms), in line with the geostrophic ocean current direction, suggesting a significant advection far from the coastal spawning areas. It is worth noting that there are some years in which the predictions are underestimated, for example in the summer surveys of 2005 and 2011 in the south-east area between Capo Passero and Malta. However, from the analysis of the maps, it emerges that the model predictions are very close to the observed values. The overall trend of the predictions for the observed data demonstrates the model's effectiveness in capturing the patterns underlying the abundance, thus confirming the validity of the hypotheses and methodologies adopted in the analysis.

The maps in Fig. 10 show the predicted values of the round sardinella larvae abundance in 2004, 2005, 2009 and 2014. There is a noticeable trend towards a greater concentration of sardines in the south-eastern area, including places such as Pozzallo and Capo Passero. Interestingly, in 2009 there was a significant abundance of larvae in the offshore area, in line with the geostrophic ocean current
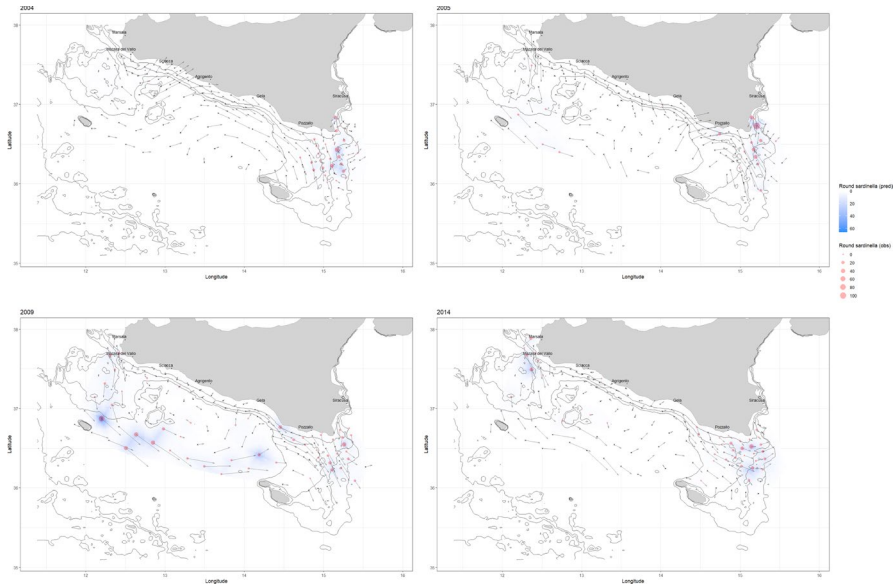
Fig. 10 Map of the predicted values of the round sardinella larvae abundance in 2004, 2005, 2009 and 2014. Red points represent the number of round sardinella larvae. The estimated number of larvae is also reported in a continuous spatial pattern by applying a white-blue colormap, indicating the estimated number of larvae per 4 km^2

direction. Furthermore, 2014 showed a greater concentration in the north-western area (Mazara del Vallo), which had shown lower quantities in previous years. However, we can see some areas where the predictions underestimate the real observed value of sardine larvae, such as in 2004 and 2005 in the south-eastern area near Capo Passero.

Next, we evaluate the model's predictive capability by showing the mean predicted abundance within the defined polygon and the mean observed values of European anchovies (Fig. 11) and round sardinella larvae (Fig. 12).

The results reveal a similarity between the observed and predicted mean values for anchovy larvae. Figure 11 shows the trend of these values for the surveys carried out in 1998 and between 2000 and 2016. A considerable disparity can be observed, particularly evident in 2011. Furthermore, 2009 appears to be the year in which the mean number of European anchovy larvae is higher.

Examination of the round sardinella data again reveals a similarity between the observed and predicted mean values. Despite this consistency, the confidence interval for sardines shows a slightly wider range. In this case, the most pronounced distinction occurred in 2005, as shown in Fig. 12, but it was not as sharp as that for anchovies in 2011. Noteworthy, however, is the remarkably wide confidence interval during 2009. The notable abundance of sardinella larvae distinguishes this specific year.
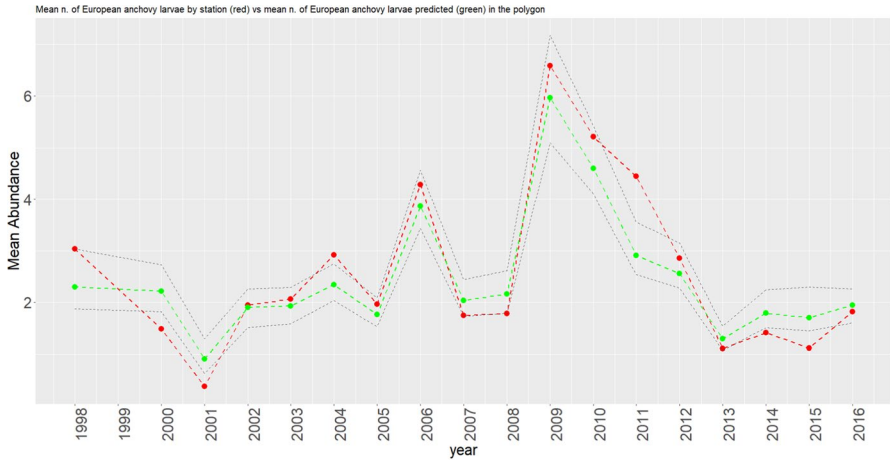
Fig. 11 Comparison between the mean posterior predicted abundance (green) of European anchovies larvae within the defined polygon and mean observed values (red) for the summer surveys in 1998 and from 2000 to 2016. The predicted abundance values are expressed as the mean number of larvae per 4 km^2. The dotted lines represent the approximate 95% credibility interval
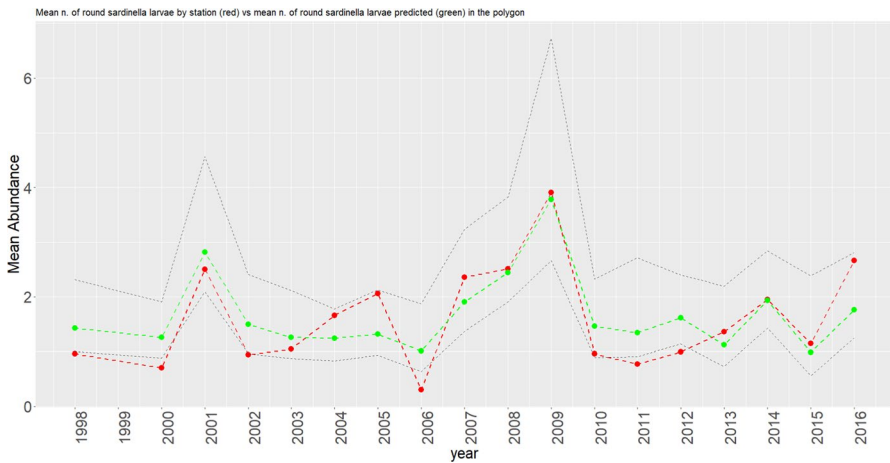

Fig. 12 Comparison between the mean posterior predicted abundance (green) of round sardinella larvae within the defined polygon and mean observed values (red) for the summer surveys in 1998 and from 2000 to 2016. The predicted abundance values are expressed as the mean number of larvae per 4 Km^2. The dotted lines represent the approximate 95% credibility interval

4.4 Abundance index analysis

This section presents the larval abundance index obtained after estimating the two models in (2) and (3), respectively. It allows a global perception of the abundance of the larvae of the two species in the Strait of Sicily in the summer seasons of 1998 and between 2000 and 2016.
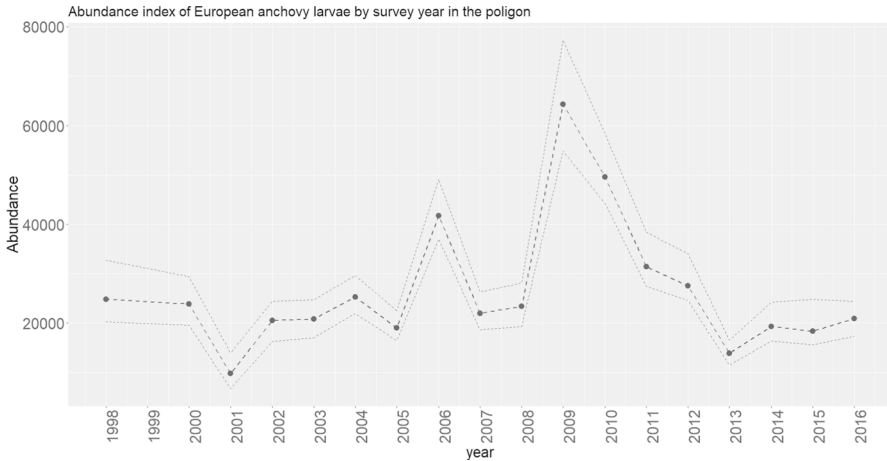
Fig. 13 Abundance index of European anchovy larvae by survey year in the polygon. The y-axis represents the number of European anchovy larvae. The dotted lines represent the approximate 95% credibility interval
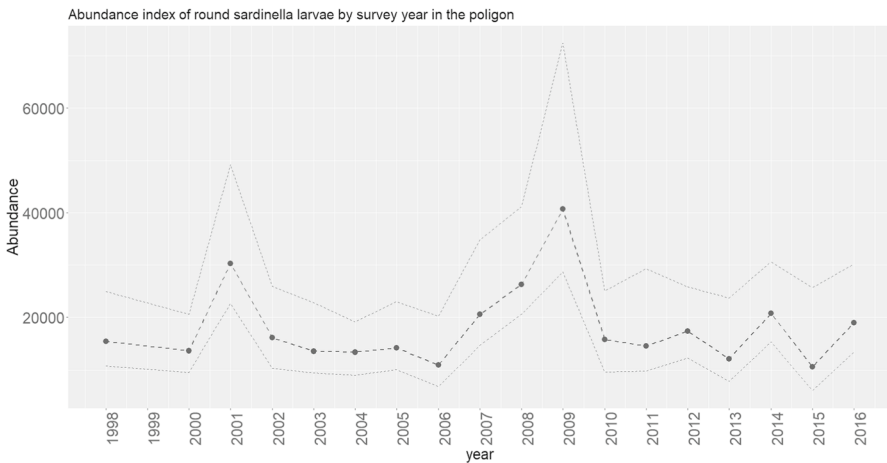

Fig. 14 Abundance index of round sardinella larvae by survey year in the polygon. The y-axis represents the number of round sardinella larvae. The dotted lines represent the approximate 95% credibility interval

Figure 13 shows the abundance index of European anchovy larvae. The annual fluctuations emerge, with a maximum peak in 2009, characterized by values close to 60000 larvae, followed by 2010. In contrast, 2001 shows the lowest predicted number of European anchovy larvae. The overall analysis does not reveal a clear trend, motivating the selection of a model that considers the spatial component independent of time.

Figure 14 shows the round sardinella larvae's abundance index. Once again, the peak occurred in 2009, with about 40000 larvae, followed by 2001. Interestingly, in

2001, there was an increase in round sardinella compared to European anchovies, while 2006 and 2010 are distinguished by the opposite trend, with a greater presence of anchovies and a decrease in sardines. Furthermore, the global analysis does not reveal a clear trend even for sardine larvae.

5 Conclusions

The work carried out in this study addressed the analysis of spatio-temporal data on European anchovy and round sardinella larvae in the Strait of Sicily from 1998 to 2016. Through a rigorous approach to data analysis and exploring different models, relevant results have emerged that contribute to our understanding of the distribution and abundance of the species under study. This concluding section summarizes the main findings, critically addresses some limitations of these applications, and proposes suggestions for future research in this ever-evolving field of study.

One of this work's main results was identifying an optimal model for estimating the presence and abundance of the two fish species under consideration. Hurdle models have been considered as a high percentage of zero values characterizes this data type. Thus, two-part models were particularly suitable for capturing the spatio-temporal distribution. They incorporate linear effects and smooth (non-linear) terms for environmental covariates, including temperature, salinity, chlorophyll-a, and currents' u and v components.

Another remarkable result emerged from the analysis of the models, which highlighted the significant role of the geostrophic ocean current components in influencing the presence and abundance of the larvae. In particular, the linear effect of the v components of the currents has proved crucial in explaining the presence and abundance of round sardinella larvae. As far as European anchovies are concerned, however, it emerged that the interaction of the currents' u and v components influences the abundance. Since the first stages of fish larvae have relatively poor swimming abilities, their horizontal distribution is strongly linked to the drifting by ocean surface current, which can also regulate connectivity with neighbouring regions (Patti et al. 2018; Falcini et al. 2020). This emphasizes the role of oceanic conditions in influencing the presence and abundance of these species. Interestingly, similar results have been found in studies focused on larval dispersal of other species occurring in the same area (Torri et al. 2021; Russo et al. 2022), highlighting the importance of the hydrodynamics in a key area for the thermohaline circulation of the Mediterranean Sea, where exchange of water masses between western and eastern basins take place (Placenti et al. 2022).

Furthermore, the analysis showed that temperature plays a significant role in explaining the abundance of *Sardinella aurita*. In particular, the model emphasizes higher larval counts in correspondence to relatively warmer areas. This finding is in line with the biology of this species which, in the Mediterranean Sea, shows the maturation of gonads at temperatures above 23 degrees (Palomera and Sabatés 1990) and allows us to recognize, for the first time on a long time series, Capo Passero as the most important spawning area in this region. Considering our results, the presence of an upwelling system that led to colder water in the north-west area could

underlie the lower larval densities observed in this zone. Temperature is thus a critical environmental factor, suggesting that climate variations can significantly affect the larval spatio-temporal distribution of this fish species.

Salinity was another relevant variable in the model for round sardinella species, with a significant positive linear effect on the presence of this species. For the abundance, a more complex relationship was considered, with a first-order random walk (RW1) effect that accounts for the role of a thermohaline front in shaping the spatial distribution of fish larvae in the study area (Torri et al. 2023).

Moreover, the analysis highlighted the importance of the geographical position. Bottom depth was included in the model, but results indicated that it is not significant for the presence but for the abundance of round sardinella with a first-order random walk (RW1) relationship. The abundance of this species is more important near the coast and decreases offshore. This highlights the difference between shelf and continental slope areas.

As concerns the European anchovy (*Engraulis encrasicolus*) the relevant variables included in the model were the interaction between the u and v components of the currents and chlorophyll-a concentration, which showed a significant negative linear effect on the abundance of this species. Thus, our model evidences an essential feature of the early life stages that, being passively advected by physical forcings, are not necessarily retained in the chl-enriched areas, affecting the mortality rates of the larval stages and the recruitment. The results of this research have important implications for marine ecology and the sustainable management of marine ecosystems. In this framework, our models can be used as predictive tools to monitor trends in fish species abundance in response to changing environmental conditions.

The prediction capacity of these models is quite good even though the model underestimates the observed values for the round sardinella in some areas. What emerged from the analysis of the maps of the predicted values of the abundance is that the distribution of European anchovy larvae is mainly distributed along the entire coast and is strongly related to the direction of the surface currents. In contrast, round sardinella larvae are mainly concentrated in the southeastern area of the Strait of Sicily, where warmer conditions are prevalent.

The analysis of the abundance indices obtained for the two species is of particular interest. For both species, significant variations were detected over the years. These highly variable trends reflect, on the one hand, the variability of the environmental conditions that can regulate the spawning as well as the natural mortality of the early life stages. On the other hand, they are influenced by the fluctuations of the spawning stock biomass that typically characterize short-living species such as small pelagics. In this context, this index represents a valuable tool for the sustainable management of fishery resources in this region. Its implementation could provide crucial information supporting fisheries management policies based on solid scientific evidence and insights into the processes governing the population dynamics of these important pelagic fish.

A criticality of this work is linked to the temporal complexity of the data: as seen in the initial models in which different temporal and spatial specifications were compared, the different spatial specifications do not excessively affect the computational

cost and the difficulty of converging the results. Instead, it seems that the temporal part influences the efficiency and effectiveness of the various models. In this context, the high number of observations leads to a consistent increase in the number of effective parameters leading to approximation problems in INLA and increasing its computational cost.

Looking to the future, there are further exciting research perspectives to explore. One possible direction concerns the analysis of the eggs of these species. Eggs represent a crucial stage in the life cycle of fish, and understanding their spatio-temporal dynamics could provide a valuable framework for providing insights into the reproductive ecology of European anchovy and round sardinella populations in the Strait of Sicily. Furthermore, testing the interactions between these different fish species and how they affect species abundance could be interesting. This could contribute to a more comprehensive understanding of ecological dynamics in complex marine ecosystems.

References

Agarwal DK, Gelfand AE, Citron-Pousty S (2002) Zero-inflated models with application to spatial count data. Environ Ecol stat 9(4):341–355. https://doi.org/10.1023/a:1020910605990

Allen LG, Pondella DJ, Horn MH (2006) The ecology of marine fishes: California and adjacent waters. Univ of California Press, Oakland

Bakun A (2006) Fronts and eddies as key structures in the habitat of marine fish larvae: opportunity, adaptive response and competitive advantage. Sci Marina 70(S2):105–122. https://doi.org/10.3989/scimar.2006.70s2105

Basilone G, Guisande C, Patti B et al (2006) Effect of habitat conditions on reproduction of the European anchovy (Engraulis encrasicolus) in the strait of sicily. Fish Oceanogr 15(4):271–280. https://doi.org/10.1111/j.1365-2419.2005.00391.x

Basilone G, Bonanno A, Patti B et al (2013) Spawning site selection by European anchovy (Engraulis encrasicolus) in relation to oceanographic conditions in the strait of sicily. Fish Oceanograph 22(4):309–323. https://doi.org/10.1111/fog.12024

Ben-Tuvia A (1960) Synopsis of biological data on/sardinella aurita/of the Mediterranean sea and other waters. FAO Fish Biol Synopsis 14:287–312

Boehlert G, Mundy B (1993) Ichthyoplankton assemblages at seamounts and oceanic islands. Bull Marine Sci 53(2):336–361

Boeing WJ, Duffy-Anderson JT (2008) Ichthyoplankton dynamics and biodiversity in the gulf of alaska: responses to environmental change. Ecol Indic 8(3):292–302. https://doi.org/10.1016/j.ecolind.2007.03.002

Campbell RA, Zhou S, Hoyle SD et al (2017) Developing innovative approaches to improve CPUE standardisation for Australia's multispecies pelagic longline fisheries. Fisheries Research and Development Corporation, Canberra

Cao J, Chen X, Chen Y et al (2011) Generalized linear Bayesian models for standardizing CPUE: an application to a squid-jigging fishery in the northwest Pacific Ocean. Sci Marina 75(4):679–689. https://doi.org/10.3989/scimar.2011.75n4679

Cavieres J, Nicolis O (2018) Using a spatio-temporal bayesian approach to estimate the relative abundance index of yellow squat lobster (Cervimunida johni) offchile. Fish Res 208:97–104. https://doi.org/10.1016/j.fishres.2018.07.002

Cergole MC, Saccardo SA, Rossi-Wongtschowski CL (2002) Fluctuations in the spawning stock biomass and recruitment of the Brazilian sardine (Sardinella brasiliensis) 1977–1997. Revista Brasileira de Oceanografia 50:13–26. https://doi.org/10.1590/s1413-77392002000100002

Cuttitta A, Guisande C, Riveiro I et al (2006) Factors structuring reproductive habitat suitability of *Engraulis encrasicolus* in the south coast of sicily. J Fish Biol 68(1):264–275. https://doi.org/10.1111/j.0022-1112.2006.00888.x

Cuttitta A, Patti B, Musco M et al (2022) Inferring population structure from early life stage: the case of the European anchovy in the sicilian and maltese shelves. Water 14(9):1427. https://doi.org/10.3390/w14091427

Etienne MP, Obradovich S, Yamanaka L, et al (2010) Extracting abundance indices from longline surveys: method to account for hook competition and unbaited hooks. arXiv preprint arXiv:1005.0892 https://doi.org/10.48550/ARXIV.1005.0892

Falcini F, Corrado R, Torri M et al (2020) Seascape connectivity of European anchovy in the central mediterranean sea revealed by weighted lagrangian backtracking and bio-energetic modelling. Sci Rep 10(1):1–13. https://doi.org/10.1038/s41598-020-75680-8

Falco F, Barra M, Wu G et al (2020) *Engraulis encrasicolus* larvae from two different environmental spawning areas of the central mediterranean sea: first data on amino acid profiles and biochemical evaluations. Eur Zool J 87(1):580–590. https://doi.org/10.1080/24750263.2020.1823493

Gneiting T, Raftery AE (2007) Strictly proper scoring rules, prediction, and estimation. J Am Stat Assoc 102(477):359–378. https://doi.org/10.1198/016214506000001437

Hilborn R, Quinn TP, Schindler DE et al (2003) Biocomplexity and fisheries sustainability. Proc Natil Acad Sci 100(11):6564–6568. https://doi.org/10.1073/pnas.1037274100

Ingram GW Jr, Alvarez-Berastegui D, Reglero P et al (2017) Incorporation of habitat information in the development of indices of larval bluefin tuna (*Thunnus thynnus*) in the western mediterranean sea (2001–2005 and 2012–2013). Deep Sea Res Part II: Top Stud Oceanograph 140:203–211. https://doi.org/10.1016/j.dsr2.2017.03.012

Izquierdo F, Paradinas I, Cerviño S et al (2021) Spatio-temporal assessment of the european hake (*Merluccius merluccius*) recruits in the northern iberian peninsula. Front Marine Sci 8:614675. https://doi.org/10.3389/fmars.2021.614675

Izquierdo F, Menezes R, Wise L et al (2022) Bayesian spatio-temporal cpue standardization: case study of European *Sardine* (*Sardina pilchardus*) along the western coast of portugal. Fish Manag Ecol 29(5):670–680. https://doi.org/10.1111/fme.12556

Joint I, Groom SB (2000) Estimation of phytoplankton production from space: current status and future potential of satellite remote sensing. J Exp Marine Biol Ecol 250(1–2):233–255. https://doi.org/10.1016/s0022-0981(00)00199-4

Kara AB, Rochford PA, Hurlburt HE (2000) An optimal definition for ocean mixed layer depth. J Geophys Res: Oceans 105(C7):16803–16821. https://doi.org/10.1029/2000jc900072

Krainski E, Gómez-Rubio V, Bakka H et al (2018) Advanced spatial modeling with stochastic partial differential equations using R and INLA. Chapman Hall/CRC. https://doi.org/10.1201/9780429031892

Lafuente JG, Garcia A, Mazzola S et al (2002) Hydrographic phenomena influencing early life stages of the sicilian channel anchovy. Fish Oceanograph 11(1):31–44. https://doi.org/10.1046/j.1365-2419.2002.00186.x

Lezama-Ochoa N, Pennino MG, Hall MA et al (2020) Using a bayesian modelling approach (inla-spde) to predict the occurrence of the spinetail devil ray (*Mobular mobular*). Sci Rep 10(1):18822. https://doi.org/10.1038/s41598-020-73879-3

Matarese AC (2003) Atlas of abundance and distribution patterns of ichthyoplankton from the northeast Pacific Ocean and Bering Sea ecosystems based on research conducted by the Alaska Fisheries Science Center (1972–1996), vol 1. US Department of Commerce, National Oceanic and Atmospheric Administration, Washington D.C.

Muñoz F, Pennino MG, Conesa D et al (2013) Estimation and prediction of the spatial occurrence of fish species using Bayesian latent Gaussian models. Stoch Environ Res Risk Assess 27:1171–1180. https://doi.org/10.1007/s00477-012-0652-3

Palomera I, Sabatés A (1990) Co-occurrence of *Engraulis encrasicolus* and *Sardinella aurita* eggs and larvae in the northwestern Mediterranean. Sci Marina 54:61–67

Paradinas I, Conesa D, López-Quílez A et al (2017) Spatio-temporal model structures with shared components for semi-continuous species distribution modelling. Spat Stat 22:434–450. https://doi.org/10.1016/j.spasta.2017.08.001

Patti B, Bonanno A, Basilone G et al (2004) Interannual fluctuations in acoustic biomass estimates and in landings of small pelagic fish populations in relation to hydrology in the strait of sicily. Chem Ecol 20(5):365–375. https://doi.org/10.1080/02757540410001727972

Patti C, Cuttitta A, Musco M et al (2013) Rapporto tecnico sulle attività di campagna oceanografica "BANSIC 2013". Tech. rep, IAMC-CNR Capo Granitola

Patti B, Zarrad R, Jarboui O et al (2018) Anchovy (*Engraulis encrasicolus*) early life stages in the central mediterranean sea: connectivity issues emerging among adjacent sub-areas across the strait of sicily. Hydrobiologia 821:25–40. https://doi.org/10.1007/s10750-017-3253-9

Patti B, Torri M, Cuttitta A (2020) General surface circulation controls the interannual fluctuations of anchovy stock biomass in the central mediterranean sea. Sci Rep 10(1):1554. https://doi.org/10.1038/s41598-020-58028-0

Patti B, Torri M, Cuttitta A (2022) Interannual summer biodiversity changes in ichthyoplankton assemblages of the strait of sicily (central mediterranean) over the period 2001–2016. Front Marine Sci 9:960929. https://doi.org/10.3389/fmars.2022.960929

Pettit L (1990) The conditional predictive ordinate for the normal distribution. J Royal Stat Soc: Ser B (Methodol) 52(1):175–184. https://doi.org/10.1111/j.2517-6161.1990.tb01780.x

Placenti F, Torri M, Pessini F et al (2022) Hydrological and biogeochemical patterns in the sicily channel: new insights from the last decade (2010–2020). Front Marine Sci 9:733540. https://doi.org/10.3389/fmars.2022.733540

Planque B, Bellier E, Lazure P (2007) Modelling potential spawning habitat of sardine (*Sardina pilchardus*) and anchovy (*Engraulis encrasicolus*) in the bay of biscay. Fish Oceanograph 16(1):16–30. https://doi.org/10.1111/j.1365-2419.2006.00411.x

Pujol MI, Larnicol G (2005) Mediterranean sea eddy kinetic energy variability from 11 years of altimetric data. J Marine Syst 58(3–4):121–142. https://doi.org/10.1016/j.jmarsys.2005.07.005

Quinci EM, Torri M, Cuttitta A et al (2022) Predicting potential spawning habitat by ensemble species distribution models: the case study of European anchovy (*Engraulis encrasicolus*) in the strait of sicily. Water 14(9):1400. https://doi.org/10.3390/w14091400

Roos M, Held L (2011) Sensitivity analysis in Bayesian generalized linear mixed models for binary data. Bayesian Anal 6(2):259–278. https://doi.org/10.1214/11-ba609

Rue HSM, Chopin N (2009) Approximate Bayesian inference for latent gaussian models by using integrated nested laplace approximations. J Royal Stat Soc: Ser b (Stat Methodol) 71(2):319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Russo S, Torri M, Patti B et al (2021) Unveiling the relationship between sea surface hydrographic patterns and tuna larval distribution in the central mediterranean sea. Front Marine Sci 8:708775. https://doi.org/10.3389/fmars.2021.708775

Russo S, Torri M, Patti B et al (2022) Environmental conditions along tuna larval dispersion: insights on the spawning habitat and impact on their development stages. Water 14(10):1568. https://doi.org/10.3390/w14101568

Sabatés A, Olivar MP, Salat J et al (2007) Physical and biological processes controlling the distribution of fish larvae in the NW mediterranean. Prog Oceanograph 74(2–3):355–376. https://doi.org/10.1016/j.pocean.2007.04.017

Spiegelhalter DJ, Best NG, Carlin BP et al (1998) Bayesian deviance, the effective number of parameters, and the comparison of arbitrarily complex models. Tech. rep, Citeseer

Torri M, Corrado R, Falcini F et al (2018) Planktonic stages of small pelagic fishes (Sardinella aurita and Engraulis encrasicolus) in the central mediterranean sea: the key role of physical forcings and

implications for fisheries management. Prog Oceanograph 162:25–39. https://doi.org/10.1016/j.pocean.2018.02.009

Torri M, Pappalardo AM, Ferrito V et al (2021) Signals from the deep-sea: genetic structure, morphometric analysis, and ecological implications of Cyclothone braueri (Pisces, *Gonostomatidae*) early life stages in the central mediterranean sea. Marine Environ Res 169:105379. https://doi.org/10.1016/j.marenvres.2021.105379

Torri M, Russo S, Falcini F et al (2023) Coupling lagrangian simulation models and remote sensing to explore the environmental effect on larval growth rate: the Mediterranean case study of round sardinella (Sardinella aurita) early life stages. Front Marine Sci 9:1065514. https://doi.org/10.3389/fmars.2022.1065514

Tsikliras AC, Antonopoulou E (2006) Reproductive biology of round sardinella (*Sardinella aurita*) in northeastern mediterranean. Sci Marina 70(2):281–290. https://doi.org/10.3989/scimar.2006.70n2281

Wenger SJ, Freeman MC (2008) Estimating species occurrence, abundance, and detection probability using zero-inflated distributions. Ecology 89(10):2953–2959. https://doi.org/10.1890/07-1127.1

Whitehead PJP, Nelson GJ, Wongratana T (1988) Clupeoidi fishes of the World (suborder *Clupeoidei*): an annotated and illustrated catalogue of the herrings, sardines, pilchards, sprats, shads, anchovies and wolf-herrings. Food and AgricultureOrg, Rome

Zhou S, Campbell RA, Hoyle SD (2019) Catch per unit effort standardization using spatio-temporal models for Australia's Eastern Tuna and Billfish Fishery. ICES J Marine Sci 76(6):1489–1504. https://doi.org/10.1093/icesjms/fsz034

Zuur AF, Ieno EN, Walker NJ et al (2009) Zero-truncated and zero-inflated models for count data. Springer, New York, pp 261–293

Authors and Affiliations

Alessia Granata[1,5] · Antonino Abbruzzo[2] · Bernardo Patti[3] · Angela Cuttitta[4] · Marco Torri[4]

✉ Antonino Abbruzzo
  antonino.abbruzzo@unipa.it

✉ Angela Cuttitta
  angela.cuttitta@cnr.it

  Alessia Granata
  alessia.granata@unipa.it

  Bernardo Patti
  bernardo.patti@ias.cnr.it

  Marco Torri
  marco.torri@cnr.it

[1] Department of Earth and Marine Sciences, University of Palermo, Via Archirafi 22, 90123 Palermo, Italy

[2] Department of Economics, Business and Statistics, University of Palermo, Viale delle Scienze, ed. 13, 90146 Palermo, Italy

[3] National Research Council (CNR-IAS), Institute for the Study of Anthropic Impacts and Sustainability in the Marine Environment, Lungomare Cristoforo Colombo 4521, 90149 Palermo, Italy

[4] National Research Council (CNR-ISMed), Institute for Studies on the Mediterranean, Via Filippo Parlatore, 65, 90145 Palermo, Italy

[5] NBFC, National Biodiversity Future Center, Piazza Marina 61, 90133 Palermo, Italy