Author(s): Ilaria Pia, Elina Numminen, Lari Veneranta, Jarno Vanhatalo

Title: **Spatially Explicit Model to Disentangle Effects of Environment on Annual Fish Reproduction**

Year: 2025

Version: Published version

Please cite the original version:

Pia, I., Numminen, E., Veneranta, L. and Vanhatalo, J. (2025), Spatially Explicit Model to Disentangle Effects of Environment on Annual Fish Reproduction. Environmetrics, 36: e2894. https://doi.org/10.1002/env.2894

Environmetrics

WILEY

RESEARCH ARTICLE OPEN ACCESS

Spatially Explicit Model to Disentangle Effects of Environment on Annual Fish Reproduction

Ilaria Pia[1] | Elina Numminen[1] | Lari Veneranta[2] | Jarno Vanhatalo[1,3]

[1]Department of Mathematics and Statistics, Faculty of Science, University of Helsinki, Helsinki, Finland | [2]Natural Resources Institute Finland, Vaasa, Finland | [3]Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

Correspondence: Jarno Vanhatalo (jarno.vanhatalo@helsinki.fi)

Received: 12 October 2023 | **Revised:** 15 December 2024 | **Accepted:** 19 December 2024

Funding: This work has received funding from the Doctoral Programme in Mathematics and Statistics of the University of Helsinki Academy of Finland (grant 317255), Jane & Aatos Erkko Foundation (IP, JV, EN), the European Union's Horizon Europe research and innovation programme under grant agreement No. 101081642 (JV, IP), and from the Doctoral Program in Mathematics and Statistics of the University of Helsinki (IP). In addition, JV acknowledges funding from the European Union (ERC Consolidator Grant BEFPREDICT, 101087409).

Keywords: Beverton–Holt model | change-of-support model | fisheries management | Ricker model | species distribution model | zero-inflation

ABSTRACT

Population growth models are essential tools for natural resources management and conservation since they provide understanding on factors affecting renewal of natural animal populations. However, we still do not properly understand how the processes underlying reproduction of natural animal populations are affected by the environment at the spatial scale at which reproduction actually happens. A particular challenge for analyzing these processes is that observations from different life cycle stages are often collected at different spatial scales, and there is a lack of statistical methods to link local and spatially aggregated information. We address this challenge by developing spatially explicit population growth models for annually reproducing fish. Our approach integrates mechanistic Ricker and Beverton–Holt population growth models with a zero-inflated species distribution model and utilizes the hierarchical Bayesian approach to estimate the model parameters from data with varying spatial support: local scale count data on offspring and environment, and areal data from commercial fisheries informing about a spawning stock size. We show, both theoretically and empirically, that our models are identifiable and have good inferential performance. As a proof of concept application, we used the proposed models to analyze the drivers of whitefish *Coregonus laveratus* (L.) s.l.) reproduction along the Finnish coast of the Gulf of Bothnia in the Baltic Sea. The results show that the proposed model provides novel understanding beyond what would be attainable with earlier methods. The distributions of the reproduction areas, spawner density, and maximum proliferation rate were strongly dependent on local environmental conditions, but the effects and the relative importance of the covariates varied between these processes. The proposed models can be extended to other systems and organisms and enable ecologists to extract a better understanding of processes driving animal reproduction.

1 | Introduction

Population growth models predict the rate of recruitment of new individuals to a population at a given population size (Quinn and Deriso 1999; Munch, Giron Nava, and Sugihara 2018). This is essential information for the management of wild animal populations since it aids implementation of sustainable exploitation rates for living natural resources, such as fish and game, and conservation actions for rare species. Furthermore, in the era of rapidly progressing global change, knowledge of the effects of the

environment on population growth is mandatory for proactive and efficient adaptation to these changes (Pörtner and Peck 2010; Crozier and Hutchings 2014). For example, reproductive success of many fish species is sensitive to environmental conditions on their spawning sites so that environmental change in these sites may have a direct effect on the renewal of fish populations (Heikinheimo, Pekcan-Hekim, and Raitaniemi 2014; Kallasvuo, Vanhatalo, and Veneranta 2017; Kraufvelin et al. 2018; Macura et al. 2019).

Traditionally, population growth models, such as the Ricker and the Beverton–Holt models, are derived for a population as a whole and parameterized with population census, such as mark-recapture or catch and effort data (Quinn and Deriso 1999; Myers 2001; Pulkkinen and Mäntyniemi 2013). This implies that the estimated population growth parameters, such as the proliferation rate, are estimated without acknowledging the spatial variation in population distribution (i.e., species' local abundance) and the environmental conditions, nor the ways these variations influence the reproduction potential over the entire distribution area of a population. While environmental covariates have been demonstrated to have an effect on population growth parameters (see e.g. Rahikainen et al. 2017; Weigel et al. 2021), previous studies have shown this without accounting for spatial heterogeneity in environmental conditions or population distribution. Since environmental conditions, the predicted changes in them, and the distribution of individuals may vary strongly in space, compressing the inherently spatially structured environmental information to a population-level value may not provide comprehensive understanding on the effects of the environment on population reproduction capability.

On the other hand, effects of the environment on local abundance of a species are commonly studied with species distribution models (SDMs; Elith and Leathwick 2009), which relate local species observations to environmental covariates from the observation sites. SDMs are based on the theory of species environmental niches (Peterson et al. 2011), which states that a species exists in a given landscape if its population growth rate in those conditions is positive (Hutchinson 1978). They are routinely used for natural resources management and conservation since they can provide valuable information on essential habitats of species (Kallasvuo, Vanhatalo, and Veneranta 2017; Kraufvelin et al. 2018; Franklin 2023). SDMs do not model the processes that underlie the realized population growth rate, such as proliferation rate or density-dependent survival, explicitly, but are mere statistical descriptions of the relationship between environment and species occurrence or abundance. For this reason, their utility for analyzing population viability in the face of environmental change is limited.

Currently, there is a lack of rigorous statistical methods to analyze how the processes underlying reproduction of natural animal populations are affected by the environment at the spatial scale at which reproduction actually happens. In this work, we address this challenge by integrating mechanistic population growth models with a zero-inflated SDM to generate a spatially explicit reproduction model for annually spawning fish. Furthermore, we propose a hierarchical Bayesian approach to estimate the parameters of this model from data with varying spatial support: (1) local-scale count data on offspring, and (2) areal catch-effort data on a spawning stock.

As a result, we arrive at a model where the number of fish larvae at a given location is proportional to the size of the local reproductive population (spawners), maximum fecundity of spawners, and the survival of offspring from birth to recruitment at a given location (Brännström and Sumpter 2005). We further divide the survival of offspring into density-dependent and -independent survival, so that population growth levels off at high spawner densities. The parameters of the reproductive model vary spatially as functions of local environmental covariates and species traits (size and fecundity), allowing inference on the effects of the environment on different components of the reproduction process at a local scale.

Since our approach provides a spatially explicit description of the mechanisms of the reproduction process, it provides novel ecological insight on the distribution of fish reproduction areas beyond what is possible with the current methods based on species distribution modeling, such as estimates of the distribution of spawners and local scale variation in egg and larvae survival. To demonstrate these, we apply our model to the whitefish (*Coregonus laveratus*) population along the Finnish coastal area of the Gulf of Bothnia, located at the northern Baltic Sea (Figure 1A). We use extensive data on whitefish larvae abundance, maps of environmental covariates, and commercial fisheries data to parameterize our spatially explicit annual population growth model. We then demonstrate how our model can be used to analyze the environmental drivers of whitefish reproduction and to predict population level effects of the environment as an integral of the local scale effects over the distribution area of a species. We also discuss and provide examples of future application areas beyond our case study.

2 | Motivating Case Study

2.1 | Gulf of Bothnia Whitefish

Our case study area comprises the Finnish coast of the Gulf of Bothnia, a brackish water basin in the northern Baltic Sea (Figure 1A). Gulf of Bothnia is inhabited by two ecotypes of whitefish: migratory anadromous ecotype, and the sea-spawning ecotype (*Coregonus laveratus* (L.) s.l.; Lehtonen 1981; Himberg, Vasemägi, et al. 2015). Both whitefish ecotypes are important for fishery and have high economic and social impact (Verliin et al. 2011; Kallio-Nyberg et al. 2019; OSF 2020). In this work, we study the sea-spawning ecotype and refer to it shortly as whitefish. Whitefish spawn in the littoral zone in October and November, and the size of the spawning whitefish increases towards the south (Figure 1D). Whitefish larvae hatch from eggs soon after ice break up in April and May (Figure 1D; Veneranta et al. 2013). Because whitefish eggs develop and overwinter in the nearshore bottom, their survival and growth are presumably significantly affected by the environmental conditions, such as the amount of fast ice, eutrophication, and sedimentation, in the reproduction area (Freeberg, Taylor, and Brown 1990; Müller 1992; Veneranta Hudd, and Vanhatalo, 2013; Veneranta et al.2013; Peltonen and Weigel 2022).

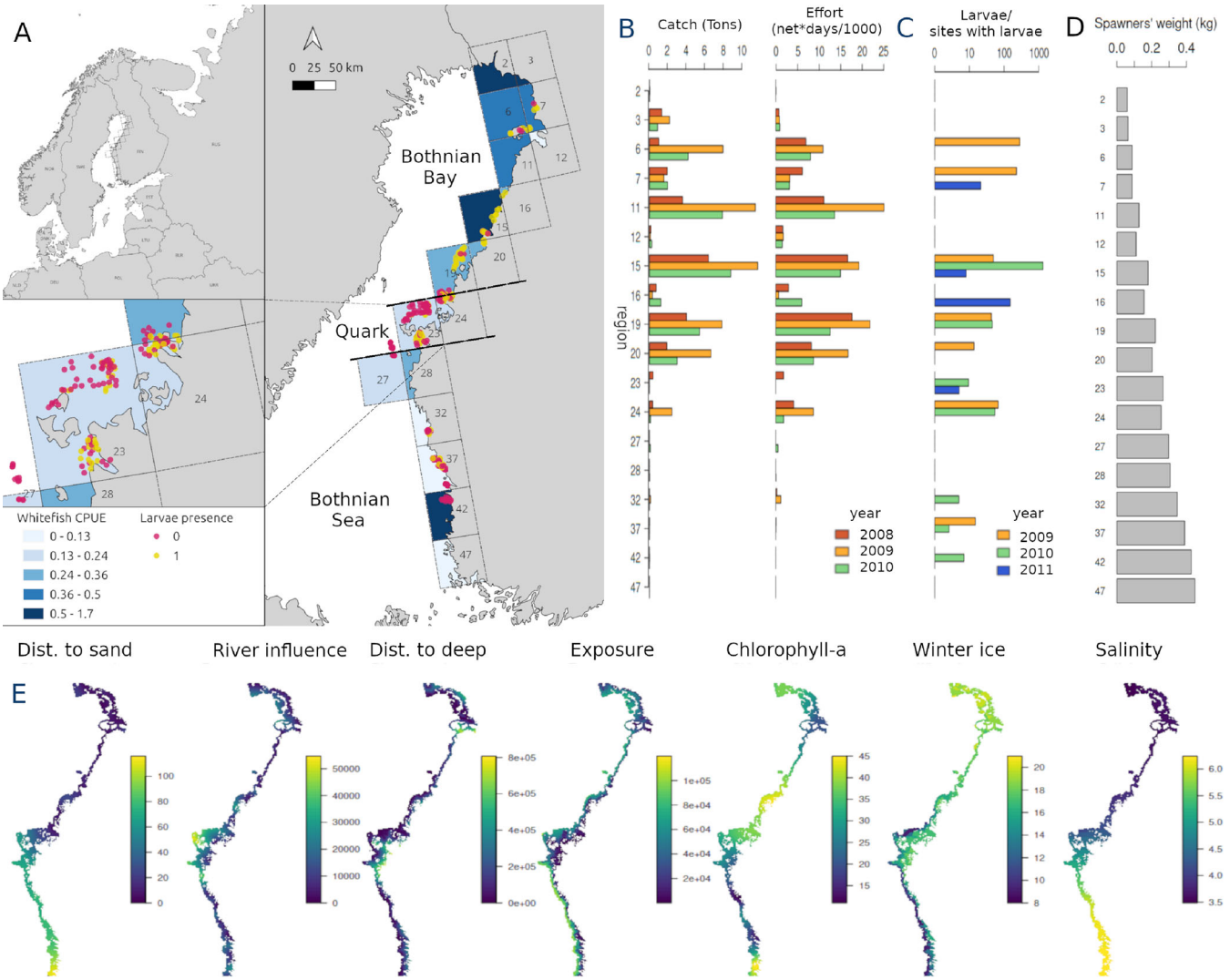

FIGURE 1 | (A) The study area, Gulf of Bothnia, with information on whitefish larvae sampling sites and 18 statistical rectangles of the International Council for the Exploration of the Seas (ICES). The sampling sites' colors indicate presence (yellow) or absence (red) of larvae at the site. The ICES rectangles' colors represent the average sea-spawning whitefish catch per unit effort (CPUE) over the years 2008–2010. The zoom into rectangles 23 and 24 provides an example of the roughness of the shoreline and on the spatial distribution of the sampling locations therein. (B) summary of the spawning time whitefish fisheries catch and effort statistics over the statistical rectangles. (C) The average number of whitefish larvae in those sampling sites that contained larvae. (D) Whitefish spawners average weight (in kg) in each statistical rectang. (E) The maps of the environmental covariates used in the analyses (see Table 1 for their explanation).

2.2 | Whitefish Larvae Data

In 2009–2011, whitefish larvae abundance data were collected at 328 sampling sites over the Finnish coast of the Gulf of Bothnia (see Figure 1A). Sampling sites were randomized over the area using a stratified sampling scheme, and each site was visited approximately one week after ice break-up, when larvae had hatched (Veneranta, Hudd, and Vanhatalo 2013). Each larval observation contained the information on the number of larvae caught with a beach seine and the volume of a sample (on average 14 m^2). Approximately 50% of all whitefish larvae observations were zero (Figure 1A) and the counts were heavily right-skewed with a median at zero, 75% of data below 8, and the maximum of 3000 (Figure A.2 in the Supporting Information). The mean of all larval counts was 42, and the variance was 35,000 and the mean larval counts at sites with larval presence varied considerably from south to north (Figure 1C).

2.3 | Whitefish Fisheries Data

Finnish commercial, coastal fisheries target whitefish using gillnets in the vicinity of the spawning sites during their spawning season in October–November (Kallio-Nyberg et al. 2019), thus producing information on spawning stock size. We extracted data from these fisheries over the study area from the fisheries monitoring database of the Natural Resources Institute Finland (OSF 2020). These data comprise fishing effort (net-days; the cumulative sum of days nets have been in the water) and whitefish catch (kilograms) from 18 statistical rectangles of the International Council for the Exploration of the Seas (ICES) for

years 2008–2010 (Figure 1B). Each rectangle is approximately of size 50 km × 50 km but varies in water area (see Figure 1A for a zoomed in example). Moreover, samples collected as a part of EU Data Collection Framework (DCF) from caught fish also provide data on the size (kg; Figure 1E) of the spawning whitefish and the egg production of spawners (Section 1 in the Supporting Information). Since the sea-spawning whitefish remain smaller at maturity than the anadromous ecotype (Kallio-Nyberg et al. 2019), we used only data from gillnets of mesh sizes under 36 mm, which have been shown to filter out possible anadromous ecotypes from the catch data (Kallio-Nyberg et al. 2019).

2.4 | Environmental Covariates

The geological, climatic, and eutrophication conditions in the Gulf of Bothnia vary considerably following south-north and open water-coastline gradients (Håkansson, Alenius, and Brydsten, 1996; Lundberg, Jakobsson, and Bonsdorff 2009; Leppäranta and Myrberg 2009, see also Figure 1E). Numerous islands, bays, and peninsulas cause fine-scale variation in exposure to wind and wave action, water depth, the influence of freshwater discharge from rivers, and bottom type throughout the area. Presumably, sandy bottom types are the most important areas for whitefish spawning and average water temperature, salinity, and winter ice cover, which change considerably from south to north (Veneranta, Vanhatalo, and Urho 2016; Haapala and Leppäranta 1997; Leppäranta 2022), can affect whitefish reproduction success as well (Veneranta, Hudd, and Vanhatalo 2013). Ice covers the coastal areas of the Gulf of Bothnia typically from October–November to April–May with 1–3 month difference in the length of the annual ice cover period between south and north. Here we used the last ice cover date in spring as a proxy for the length of the ice cover period. Salinity ranges from limnic waters (< 1 psu) in the innermost reaches in the north, and in the estuaries to 6–7 psu in the south (Voipio 1981; Håkansson, Alenius, and Brydsten, 1996; Helsinki Commission 2002; Kankaanpää et al. 2022). The south and inner coastal regions of the Gulf of Bothnia are also more heavily impacted by eutrophication than the north and open water areas (Veneranta, Hudd, and Vanhatalo 2013). We summarized the environmental conditions with seven covariates (Figure 1E, and Table 1) collected from several raster maps of 300 m resolution spanning the whole coastal region of the Gulf of Bothnia. We grouped the covariates into three classes: Climatic covariates, including length of winter ice cover period and sea surface salinity; Geological covariates, including distance to sandy bottom, river influence (river discharge scaled by the distance to river mouth), distance to deep (> 10 m) water, and exposure to wind and wave (see Veneranta, Hudd, and Vanhatalo 2013); and Eutrophication covariate, consisting of the chlorophyll-a concentration based on HELCOM assessments (see Andersen et al. 2011).

3 | Spatially Explicit Annual Population Growth Models

To build our model, we first discretized the study area according to the 300 m resolution lattice grid formed by the environmental covariates (see Section 2.4). We then formalized the distribution of the spawners at the grid cell locations as a two-step process: a process explaining whether a site is suitable for spawning or not (Figure 2A) and a process explaining the distribution of spawners between the suitable sites (Figure 2B). Third, we built mechanistic reproduction models for the number of fish larvae (Figure 2E) at the grid cell locations conditional on the spawner density, maximum proliferation rate (egg production multiplied by density-independent survival of eggs to larvae; (Figure 2C)) and density-dependent processes affecting the mortality of eggs from spawning to larvae (e.g., competition for spawning sites, density-dependent predation of eggs and larvae, and resource competition between larvae; Figure 2D) at them. For last, we built the observation models for the larvae data (Figure 2F) and the fisheries data (Figure 2G).

3.1 | Sites Suitability for Spawning

We denote by $x_{i,t} \in \mathbb{R}^p$ the environmental covariates and by $s_i \in \mathbb{R}^2$ the spatial coordinates of the ith grid cell at year t. A Bernoulli process $z_{i,t}$ sets grid cells either suitable ($z_{i,t} = 1$) or not suitable ($z_{i,t} = 0$) for spawning with probability

$$\theta_{i,t} = \text{logit}^{-1}(\overline{\alpha} + \overline{\beta}^{\top} x_{i,t}) \quad (1)$$

Here, $\overline{\alpha}$ and $\overline{\beta}$ are the intercept and weight parameters of the logit-linear model.

TABLE 1 | Environmental covariates used in the study, available as thematic raster maps.

	Covariate	Description
Geological covariates	Distance to sand	An index value of distance to the nearest sandy shore distance along water multiplied by the surface area of each individual shallow (< 1 m) sandy shore
	Exposure	Average wind/wave exposure over all directions
	River influence	River outflow weighted average distance to river mouths (m)
	Distance to deep	Distance (m) to 20 m deep water
Eutrophication covariate	Chlorophyll-a	Chlorophyll-a concentration as a proxy of phytoplankton biomass
Climatic covariates	Salinity	Mean salinity (psu) in May–June 2009
	Winter ice	Last week of ice (calendar week) in winter 2009–10

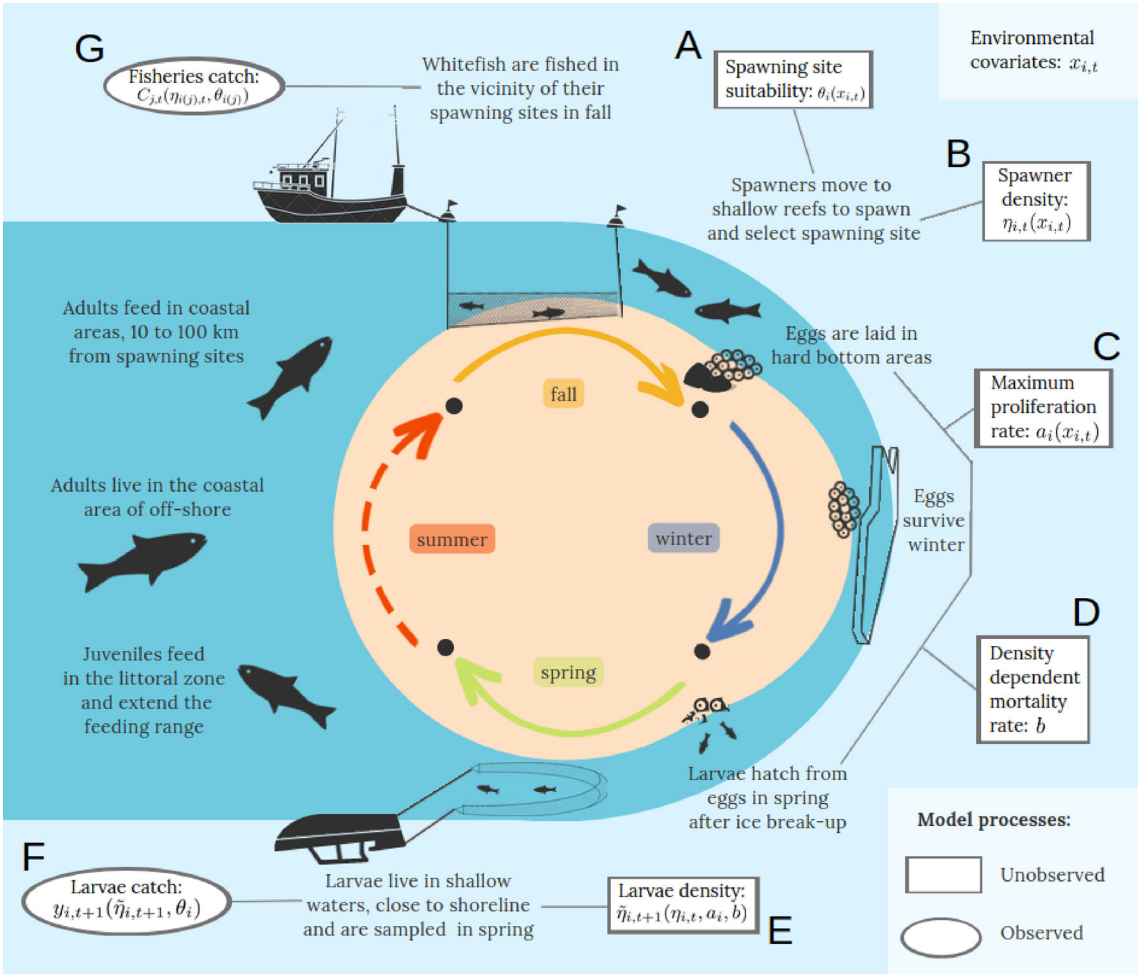

FIGURE 2 | The main phases (A–G) of the whitefish life cycle during the year show how the whitefish moves through different coastal areas according to the season. The model described in Section 3 captures the different processes behind the stages of the whitefish life cycle. The observed and unobserved model components are reported, respectively, in oval and rectangular white frames and linked to the corresponding whitefish life stages.

3.2 | Spawner Density

Assuming that spawning happens only in suitable grid cells, we modeled the density of (female) spawning whitefish with

$$\eta_{i,t} = \begin{cases} \exp(\alpha + \beta^\top x_{i,t} + \delta_{j_i,t}) & \text{if } z_{i,t} = 1 \\ 0 & \text{if } z_{i,t} = 0 \end{cases} \quad (2)$$

where $\delta_{j_i,t}$ is a random effect of year t in the statistical rectangle (hereafter shortly region) where the ith grid cell is located at. The random effects account for variation in spawning stock size not attributable to environmental covariates (e.g., differences in sizes of year classes), and they were modeled as independent over years but spatially correlated between the J regions (within a year). We modeled them with a multivariate Gaussian distribution, $\delta_t = [\delta_{1,t}, \ldots, \delta_{J,t}]^\top \sim N(0, \Sigma_\delta)$, where $[\Sigma_\delta]_{j,j'} = \sigma_\delta^2 \exp(-D_{j,j'}/l)$, with variance parameter σ_δ^2, range parameter l, and euclidean distance between the centers of the regions $D_{j,j'}$.

3.3 | Larval Production Models

The number of eggs produced by fish (i.e., fecundity) is strongly dependent on spawning individuals' size (Figure A.1). The survival of eggs to larvae stage, on the other hand, can depend on environmental conditions (density-independent mortality) and density-dependent processes such as the saturation of eggs to spawning site, which increases risk for predation by other fish and invertebrates, and resource competition between newly hatched larvae (density-dependent mortality). Hence, we defined the proliferation rate of female fish at a given location as a product of these three terms. We denote by w_j the average weight (kg) of spawners in region j and model the *maximum proliferation rate*, $a_{i,t}$, (i.e., the product of fecundity and density-independent survival) in an ith grid cell as

$$\log a_{i,t} = \bar{a} + \bar{b}^\top x_{i,t} + f(w_{j(i)}) \quad (3)$$

where $\bar{a}$ is the maximum proliferation rate of (one kg) spawners on average environmental conditions, $w_{j(i)}$ is the average weight (kg) of spawners in the region j where grid-cell i is located, $f(w_{j(i)})$ describes the log relative change in maximum proliferation rate as a function of fish weight, and $\bar{b}^\top x_{i,t}$ describes the log-relative effect of covariates on the density-independent survival of eggs to larvae.

Maximum proliferation rate represents the proliferation rate in the absence of any density-dependent effects and is an essential

parameter in all population growth models. However, different mechanistic assumptions concerning the distribution of spawners over a reproduction site, spawning site competition, and mortality of newly hatched larvae in the presence of competition lead to different functional forms for density-dependent mortality of eggs (Geritz and Kisdi, 2004; Brännström and Sumpter 2005). Here, we considered larval production models with density-independent survival and two alternatives for density-dependent survival: the Ricker and the Beverton–Holt models (Quinn and Deriso 1999). These two models implicitly assume individuals compete for the same resources, leading to different outcomes according to the type of competition. They are commonly used in population dynamics modeling and fisheries stock assessments, where the resource could be represented by, for example, a suitable spot to lay eggs.

3.3.1 | Density-Independent Survival Model

The density-independent larval production model is simply

$$\tilde{\eta}_{i,t+1} = a_{i,t}\eta_{i,t} \tag{4}$$

where $a_{i,t}$ and $\eta_{i,t}$ are the maximum proliferation rate and spawner density, as described above, while $\tilde{\eta}_{i,t+1}$ is the expected density of larvae next spring. Here, the assumption is that there is no competition between the spawners or larvae, so that the expected density of larvae is only linearly proportional to spawner density and proliferation rate. This is also the limiting form of the Ricker and Beverton–Holt models at small population sizes.

3.3.2 | Ricker Model

Scramble competition happens when there is an equal distribution of resources between individuals, so that competition with neighbors leads to a rapid decrease of reproductive success when the shared resources become insufficient for survival. In the presence of scramble competition such that each resource site can maintain exactly one single individual and uniform, random, distribution of spawners within one cell, the expected density of larvae next spring $\tilde{\eta}_{i,t+1}$ follows the Ricker model (Brännström and Sumpter 2005)

$$\tilde{\eta}_{i,t+1} = a_{i,t}\eta_{i,t}e^{-b_i\eta_{i,t}} \tag{5}$$

where $a_{i,t}$ and $\eta_{i,t}$ are again the maximum proliferation rate and spawner density, and b_i explains the density-dependent mortality of whitefish eggs from spawning until the time of larvae sampling. The latter includes the effect of egg saturation at spawning sites, potentially higher predation rates at higher densities, and the early-stage larvae mortality through resource competition. Therefore, we defined the density-dependent mortality at a location to be proportional to the density of eggs, so that $b_i = b \times e^{f(w_{j(i)})}$, where b is a density-dependent survival parameter common across locations and $e^{f(w_{j(i)})}$ is the fish weight-dependent relative fecundity of spawners at the ith location as in (3). The Ricker model is a unimodal function of spawner density so that the expected number of larvae first increases and, after peaking at an optimal spawner density, decreases with increasing spawner density.

3.3.3 | Beverton–Holt Model

Alternatively, contest competition happens when, in case of a shortage of a resource, only successful individuals get the needed resource to reproduce, and defeated ones do not get any. Assuming contest competition and clustered (Negative-Binomial) distribution of spawners, the expected larval production follows the Beverton–Holt model (Brännström and Sumpter 2005)

$$\tilde{\eta}_{i,t+1} = \frac{a_{i,t}\eta_{i,t}}{1 + b_i\eta_{i,t}} \tag{6}$$

where $a_{i,t}$ and $\eta_{i,t}$ are again the maximum proliferation rate and spawner density, and b_i explains the density-dependent mortality of whitefish eggs from spawning until the time of larvae sampling. We defined parameter b_i similarly as in the Ricker model. The Beverton–Holt model is a monotonically increasing and asymptotically saturating function of spawner density so that the expected number of larvae first increases and at the limit of infinite spawner density, it reaches a limiting maximum value.

3.4 | Observation Models

3.4.1 | Larval Observations

Given a grid cell is suitable for spawning, we modeled the (conditional) distribution for the number of larvae caught in larvae sampling, $y_{i,t+1}$, with a Negative-Binomial distribution with mean $\tilde{\eta}_{i,t+1}V_i$, where V_i is the volume of the sampled water. Then, by marginalizing over the suitability indicators, $z_{i,t}$, we arrived at a zero-inflated Negative-Binomial model for the number of sampled larvae:

$$P(y_{i,t+1}|\tilde{\eta}_{i,t+1}, \theta_{i,t}, V_i, r) = \begin{cases} (1-\theta_{i,t}) + \theta_{i,t} \times \mathrm{NB}(y_{i,t+1}|\tilde{\eta}_{i,t+1}V_i, r), & \text{if } y_{i,t+1} = 0 \\ \theta_{i,t} \times \mathrm{NB}(y_{i,t+1}|\tilde{\eta}_{i,t+1}V_i, r), & \text{if } y_{i,t+1} > 0 \end{cases} \tag{7}$$

where r is an overdispersion parameter. We used a Negative Binomial distribution since, due to the high variability in the larvae count data (Section 2.2), it clearly outperformed the Poisson distribution in our preliminary tests. By denoting vectors of all the larval observations as well as density and suitability parameters with y, η, $\tilde{\eta}$, and θ, the joint model for all larval observations can be written as:

$$P(y|\tilde{\eta}, \theta, V, r) = \prod_{t=1}^{T}\prod_{j=1}^{J}\prod_{i\in S_j} P(y_{i,t}|\tilde{\eta}_{i,t+1}, \theta_t, V_i, r) \tag{8}$$

where T is the total number of study years, S_j is the set of grid cells within region j, and $V = [V_1, \ldots, V_n]^\top$ is the vector of volumes of sampled water.

3.4.2 | Fisheries Catch and Effort Observations

Our fisheries data is collected from spawning time fisheries, during which fishermen are known to concentrate fishing effort in areas where they believe whitefish spawn. We will denote by $\eta^{\mathrm{F}}_{i,t}$ fishermen's expectation for the spawner density at grid cell

i in year t. Naturally, each fisherman has their own expectation, so $\eta^{F}_{i,t}$ can be interpreted as the mean of all subjective expectations. Because fishermen target the spawning sites, we assumed they allocate the total fishing effort in region j in year t, $E_{j,t}$, proportionally to $\eta^{F}_{i,t}$, so that the fishing effort in grid cell i is $E_{i,t} \propto E_{j,t} \times \eta^{F}_{i,t}$. Fishermen typically have long experience with their fishing region, and on average, they know rather well how fish distribute therein (Kaurila et al. 2022). Hence, we assumed that on average their beliefs about the spawner density follow the spawning suitability of the cells, so that $\eta^{F}_{i,t} \propto \theta_{i,t}$. This implies that the model for the fishing effort at a grid cell is

$$E_{i,t} = E_{j,t} \frac{\theta_{i,t}}{\sum_{i' \in S_j} \theta_{i',t}} \tag{9}$$

where S_j is the set of indexes i that belong to region j.

To derive a model for fisheries catch statistics, we denote by $N_{i,t}$ the number of spawning whitefish before fishing and with $C_{i,t}$ the number of spawners caught by fishing in grid cell i. We then assumed the event that a fish survives a unit fishing effort in a day is independent from the events that the fish survives in all the other fishing days and hence used an exponential distribution, with catch rate parameter λ, to model the probability that a fish at a grid cell i survives a unit fishing effort (i.e., one net day) applied to that cell. Hence, the probability that a fish survives $E_{i,t}$ net-days is $P(\text{survived } E_{i,t}|\lambda) = e^{-\lambda E_{i,t}}$ and the expected number of fish caught by fishing in grid cell i is $E[C_{i,t}] = (1 - e^{-E_{i,t}\lambda})N_{i,t}$. We can then write a system of two simultaneous equations

$$\begin{cases} \theta_{i,t}\eta_{i,t}A_i = e^{-E_{i,t}\lambda}N_{i,t} \\ E[C_{i,t}] = (1 - e^{-E_{i,t}\lambda})N_{i,t} \end{cases} \tag{10}$$

where A_i is the area of a lattice grid cell (300 m × 300 m here, at each grid cell). Hence, the left and right-hand sides of the first equation give the expected number of spawners that survive fishing in the ith grid cell. By solving the system of equations for the expected catch, we can write $E[C_{i,t}] = (e^{E_{i,t}\lambda} - 1)A_i\theta_{i,t}\eta_{i,t}$ and by summing up the contribution of all grid cells in a region, we obtain an estimate for the expected fisheries catch in that: $E[C_{j,t}] = \sum_{i \in S_j}(e^{E_{i,t}\lambda} - 1)A_i\theta_{i,t}\eta_{i,t}$. Since fisheries statistics are reported in kilograms (Section 3), we wrote the observation model for the total catches in region j as

$$\begin{aligned} C^{(kg)}_{j,t}|\eta_{j,t}, \theta_{j,t}, E_{j,t}, \lambda, \sigma^2_\epsilon &\sim \text{log-}t_4 \\ &\times \left(\log\left(\frac{\omega_j}{\phi} \sum_{i \in S_j} (e^{E_{j,t}\lambda \frac{\theta_{i,t}}{\sum_{i' \in S_j}\theta_{i',t}}} - 1)A_i\theta_{i,t}\eta_{i,t} \right), \sigma^2_\epsilon \right) \end{aligned} \tag{11}$$

where $C^{(kg)}_{j,t}$ is the reported fisheries catch (kg), $\eta_{j,t}$ and $\theta_{j,t}$ are, respectively, the vector of spawner density and suitability in each grid cell of region j, ϕ is the fraction of female whitefish in the spawning population (fixed to 1/2 according to the fisheries statistics data), and σ^2_ϵ is the scale parameter of the observation model. Log-Gaussian distribution would be a more common choice to model positive, skewed quantities, such as our fisheries catch data. However, the uncertainty in our catch data might vary between the regions, and they can also include outliers. Hence, we wanted to introduce heteroscedasticity into the model. This can be attained with the Student's-t distribution, since it is a scale mixture of Gaussian distributions. Moreover, a Student's-t distribution is also robust for outliers.

3.4.3 | Full Likelihood and the Identifiability of the Model Parameters

Since the larval observations and fisheries catch observations are conditionally independent given the model parameters, the full likelihood of our model factorizes as

$$\begin{aligned} p(y, C^{(kg)}|\cdot) = &\prod_{t=1}^{T}\prod_{j=1}^{J} \overbrace{p(C^{(kg)}_{j,t}|\eta_j, \theta_j, \lambda, \sigma^2_\epsilon)}^{\text{Eq. (11)}} \\ &\times \prod_{i \in S_j} \overbrace{\left[p(y_{i,t+1}|\eta_{i,t}, \theta_{i,t}, a_{i,t}, b, r)\right]}^{\text{Eq. (7)}} \end{aligned} \tag{12}$$

where $C^{(kg)} = [C^{(kg)}_1, \ldots, C^{(kg)}_J]^T$ with $C_j = [C_{j,t}]_{t=1,\ldots,T}$, $\eta_{j,t} = [\eta_{i,t}]_{i \in S_j}$ and $\theta_{j,t} = [\theta_{i,t}]_{i \in S_j}$. Hence, the full likelihood function is a combination of point-wise (larvae observations) and areal (fisheries statistics) likelihood terms. As such, our model is also an example of a spatiotemporal change-of-support model (Gelfand, Zhu, and Carlin 2001).

Even though the full model is parameter rich it is identifiable for all key parameters:

Theorem 1. (**overview**). *The model (12) is fully identifiable for parameters r and σ, suitability function $\theta(x)$, larval density function $\tilde{\eta}(x)$, and the mean of the fisheries catch in (11). The model (12) is also identifiable for parameters β and $\bar{b}$ as well as the function $f(\omega)$, implying that it is identifiable, up to a constant of proportionality, for the maximum proliferation rate $a(x)$ and the spawner density $\eta(x)$. Moreover, if b, λ, or either of the scale parameters $\bar{a}$ or α is fixed, the model is identifiable for all the rest of the parameters and functions.*

A formal version of this theorem and its proof are given in Section 2 in the Supporting Information. The theorem implies that if we give an informative prior for b, λ, $\bar{a}$, or α, the inference for all parameters is well behaved, at least in the limit of large data.

3.5 | Prior Distributions

The general theory on environmental niche predicts that species responses to environmental covariates are either monotonic or bell-shaped (Hutchinson 1978). To account for this, we included in our model the second-order terms of all environmental covariates and restricted the covariate weights of the second-order terms to be non-positive. In the whitefish case study, since the full likelihood is not fully identifiable to all parameters (Section 3.4.3 and Section 2 in the Supporting Information), we elicited informative prior distributions for those model parameters for which relevant information was available from existing literature or data. For the other parameters, we used either weakly informative or shrinkage priors. All the priors are summarized in Table 2, and their derivation is described in detail in Section 1 in the Supporting Information.

TABLE 2 | Prior distributions of model parameters with standardized covariates (zero mean and standard deviation one). For detailed justifications and alterations for restricted covariate effects, see Section 1 in the Supporting Information.

Parameter	Prior	Parameter	Prior
σ_δ	Student − $t(4, 0, 0.1)$	τ_k	Cauchy$(0, \sigma_\tau)$
$\log(l)$	N(4.2, 1)	σ_τ	Student − $t(4, 0, 1)$
σ_ϵ	Student − $t(4, 0, 1)$	γ_{kg}	Gamma(0.5, 1)
α	N(0, 4)	ψ_{kgp}	InvGamma(0.5, 1)
$\overline{\alpha}$	N(0, 1)	$\bar{a}$	N(6.67, 1)
$\overline{b}$	N$(0, \tau_1^2\gamma_{1g}^2\psi_{1gp}^2)$	b	Student − $t(4, 0, 400)$
β	N$(0, \tau_2^2\gamma_{2g}^2\psi_{2gp}^2)$	λ	Gamma(2, 10)
$\overline{\beta}$	N$(0, \tau_3^2\gamma_{3g}^2\psi_{3gp}^2)$	r	Gamma(1, 0.1)

4 | Posterior Inference and Predictions

All the models considered in this study were implemented, and the posterior inference was done using Markov chain Monte Carlo (MCMC) using the RStan software (Stan Development Team 2023). In all experiments, we ran four independent Markov chains per model and checked the convergence of the chains visually and using the Rhat statistics (Gelman et al. 2014). We run the Markov chains long enough so that the bulk and tail effective sample size for model parameters (estimated alongside Rhat statistics; Gelman et al. 2014; Vehtari et al. 2021) were at minimum 400 samples. After this we used the models to make posterior (predictive) analyses on whitefish larvae production.

5 | Experiments

5.1 | Tests With Simulated Data

To test the behavior of our model and to assess whether the model parameters are identifiable from the type of data available to us, we did a simulation study, where we generated data under different scenarios, which match fully (scenario 1) or partially (scenarios 2 and 3) our model assumptions and used the model described in Section 3 for posterior inference. We considered three simulation settings: (1) a scenario, in which the suitability, the spawner density and the maximum proliferation rate were all independently affected by environmental covariates (the absolute value of the realized correlation between these parameters over the simulation area was less than 0.1); (2) a scenario, where only the suitability and the maximum proliferation rate were (independently) affected by environmental covariates and the spawner density was constant over the area; and (3) a scenario, where the suitability, the spawner density and the maximum proliferation rate were all affected by environmental covariates, and there was correlation between the latter two (the realized correlation between the spawner density and the maximum proliferation rate over the simulation area was over 0.8). We considered the third scenario to test the effect of correlation between the proliferation rate and the spawner density, as they were presumably the hardest processes to infer due to their weak identifiability (see Section 3.4.3). In all scenarios we used three covariates, generated so that the first covariate was independent of the others and followed a spatial gradient, while the second and third covariates were highly correlated. Details on the implementation of the simulation study are provided in the Section 3.1 in the Supporting Information.

We assessed the parameter inference and their identifiability by calculating the coverage of the central 80%, 90%, and 95% posterior credible intervals for the true model parameters and covariate effects; that is, the proportion of simulations where the true parameter value was included within these intervals. Since we were interested in the covariate effects, not the fixed effect parameters only, we calculated the proportion of simulations where the true rate of change along a covariate was within the 80%, 90%, or 95% credible interval over 90% of the range of covariate values in the simulated data. We defined the rate of change along a covariate as the (first) derivative of the covariate effect since it is comparable to the fixed effects (i.e., β parameters) in a linear model. To test for the predictive skill of the model in the above three scenarios, we compared the predicted and simulated latent Gaussian variables and calculated the correlation between them over the lattice grid cells.

In all three scenarios of the simulation study, the coverage of the credible intervals matched well the theoretical optimal coverages of a calibrated model because most of them contained the true parameter value in corresponding proportion of the simulations (Figure 3A), indicating good inference performance of the model and parameter identifiability. The best posterior coverage was attained for the intercept parameters $(\overline{\alpha}, \alpha, \bar{a})$, density-dependent mortality parameter (b), overdispersion parameter (r) and length-scale of the spawner density random effect (l). This is reasonable, since for the intercept parameters we used informative priors (Section 2.3 in the Supporting Information), and the latter three parameters were the ones for which the prior distributions were the least informative, so that the likelihood had presumably stronger influence than the prior on their posterior. The posterior coverage was somewhat smaller than the optimal coverage for the fisheries catch rate (λ), fisheries observation error variance (σ_ϵ), the variance parameter of the spawner density random effects σ_δ, and the covariate effects. These results are reasonable as well, since all these parameters and covariate effects had shrinkage priors that pull their posterior distributions towards zero. All predicted processes matched the simulated ones well (Figure 3B). There was a high correlation between the true simulated and the posterior predictive mean of the processes in all other cases except the spawner density for the second scenario, indicating that the model was able to learn and predict the simulated spatial patterns well. The lack of correlation between the true and inferred log spawner density in the second scenario was reasonable since in that scenario the true value was constant, implying zero variability in the process, and thus in expectation zero correlation between the truth and predicted process (Figure 3C).

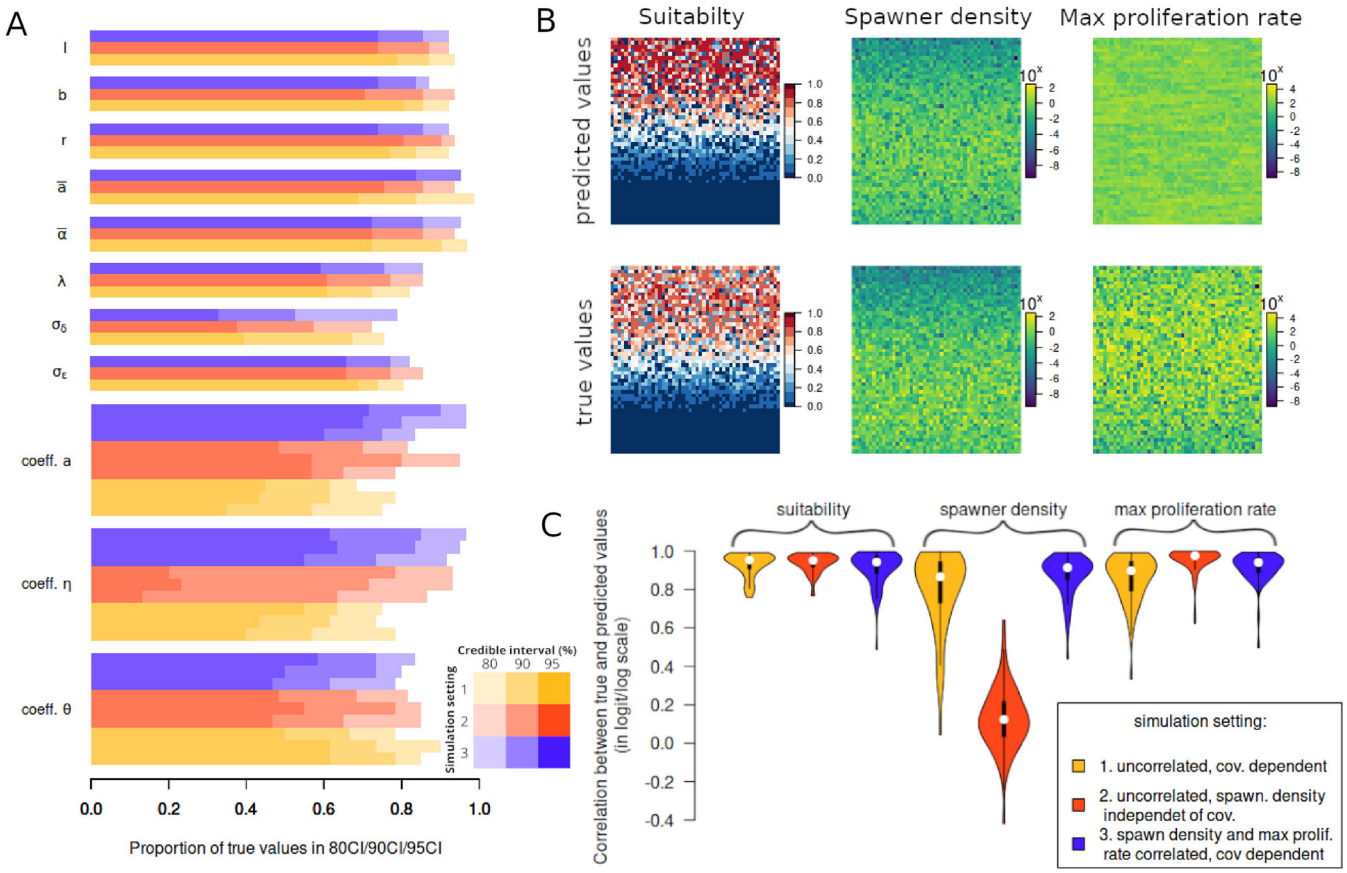

FIGURE 3 | (A) The proportion of simulations where posterior distribution includes the true value of a parameter in the 80%, 90%, and 95% credible interval for each model parameter and rate of change along the covariates. (B) True and predicted surface for suitability, log spawner density, and log maximum proliferation rate of one of the simulations, for scenario 1, where all processes depend on covariates. (C) The distribution of the correlation coefficients of predicted vs. true values of logit suitability (θ), log spawner density (η), and log maximum proliferation rate (a) over the three different simulation settings: (1) no correlation between processes and all processes depend on covariates, (2) no correlation between processes and spawner density is independent of covariate, and (3) all processes depend on covariates, and spawner density and maximum proliferation rate are highly correlated.

5.2 | Analysis of the Gulf of Bothnia Whitefish Spawning Area Data

5.2.1 | Model Assessment and Comparison

In the whitefish case study, we first compared and assessed the inference and predictive performance among the three different population growth models: the density-independent, Ricker, and Beverton–Holt models (Section 3.3). Second, as hurdle models are a common approach for modeling zero-inflated data, we formulated the proposed spatially explicit population growth models (Section 3) also through the hurdle approach for comparison. Under the hurdle formulation, the likelihood for larvae count observations becomes

$$P(y_{i,t+1}|\tilde{\eta}_{i,t+1},\theta_{i,t},V_i,r)=\begin{cases}(1-\theta_{i,t}), & \text{if } y_{i,t+1}=0\\ \theta_{i,t}\times\frac{\mathrm{NB}(y_{i,t+1}|\tilde{\eta}_{i,t+1}V_i,r)}{1-\mathrm{NB}(0|\tilde{\eta}_{i,t+1}V_i,r)}, & \text{if } y_{i,t+1}>0\end{cases} \tag{13}$$

where $\tilde{\eta}$ follows one of the larval production models described in Section 3.3. Third, to gain understanding on how our proposed approach differs from traditional species distribution modeling approaches, we also implemented SDMs, in which we modeled the larvae counts with either a zero-inflated or hurdle Negative-Binomial model (7). In the SDMs, the probability parameter θ followed the same logit-linear model as in (1) and $\log\tilde{\eta}=\bar{a}+\bar{b}^{\top}x+\log w_{j(i)}+\delta_{j,t}$, where δ is a region-specific random effect that follows the same Gaussian distribution as in our population growth model (see Equation 2). Hence, we compared in total of eight models: zero-inflated and hurdle versions of the density-independent, Ricker, and Beverton–Holt population growth models, as well as the standard SDM.

To assess the models' fit to the data, we conducted posterior predictive model assessment by comparing the empirical distribution of the data to posterior predictive distributions (Vehtari, Gelman, and Gabry 2016; Gabry et al. 2019). All the models performed practically equally well in the posterior predictive checks, and none of them showed any signs of significant deviations between the observed and predicted larval or fisheries data. Then, to compare the models' predictive performance, we calculated posterior predictive comparison measures both for an interpolation task and, in the case of larvae data, an extrapolation task. For interpolation, we used pareto smoothed importance sampling (PSIS) leave-one-out cross-validation (LOO-CV; Vehtari et al. 2024) posterior predictive distributions and calculated the

log pointwise predictive density (lppd; Gelman et al. 2014) and the scaled continuous ranked probability score (scrps Bolin and Wallin 2023). The lppd and scrps are proper scoring rules to measure the goodness of the probabilistic predictions so that scrps is less sensitive to potential outlying observations than lppd. For extrapolation tests, we run three-fold cross-validation (3-fold-CV), by partitioning the larvae count data between the three GoB sea areas: the Bothnian Sea, the Quark, and the Bothnia Bay (Figure 1). These three sea areas are known to be different in their geological, eutrophication, and climatic conditions (Veneranta, Hudd, and Vanhatalo 2013) so that this data split mimics extrapolation along covariates (Roberts et al. 2017). We used all fisheries data and larvae data from two of these three sea areas for training and predicted for the larvae data at the third, test, area. This way we tested whether fisheries data can provide information about the larval production areas. In the population growth models, we can split the total lppd and crps into contributions from the larval data and the fisheries data, such as

$$\text{lppd}_{\text{total}} = \overbrace{\sum_{j=1}^{J}\sum_{i\in S_j} \log \text{E}_{\text{post}}\left[p(y_{i,t+1}|\cdot)\right]}^{\text{lppd}_{\text{larvae}}} + \overbrace{\sum_{j=1}^{J}\sum_{t=1}^{T} \log \text{E}_{\text{post}}\left[p(C_{j,t}^{(kg)}|\cdot)\right]}^{\text{lppd}_{\text{fisheries}}}$$

To assess models' predictive performance for each data and in combination, we examined the larvae, fisheries, and total for lppd and scrps in all comparisons.

In interpolation tests, all the spatially explicit population growth models were practically equally good (Table 3). When comparing zero-inflated and hurdle formulations of the SDMs and the spatially explicit population growth models in larvae count interpolations, the hurdle models had better lppd scores, but according to scrps scores, all the models were practically equally good. Similar performance between the models in interpolating larvae counts is reasonable since all the models are flexible so that they can efficiently interpolate over covariates and spatial space. However, when extrapolating larvae counts, the spatially explicit Ricker population growth models outperformed the SDMs (extrapolation tests were not run for the Beverton–Holt and density-independent models because of computational time constraints; Table 3). This shows that the spatially explicit population growth models were able to extract useful information for larvae prediction from the fisheries data and indicates that these models were also able to extract information on the relative roles of suitability, spawner density, and proliferation rate on larvae density.

In the above comparisons, we compared models' marginal predictive distributions for both zero and positive observations. When comparing their Beta zero-inflated model to a hurdle model Tang et al. (2023), argued that as the concentration of the predictive distribution depends on the number of sources of zeros that are being modeled, it is fairer to compare hurdle and zero-inflated models only in their ability to predict the positive counts. Hence, to better understand the differences between the spatially explicit zero-inflated and hurdle models, we more carefully examined their point-wise lppds. With the Ricker formulation, both models predicted equally well larval absences ($y = 0$), hurdle model predicted better low larval counts ($y < 30$), and the zero-inflated model predicted better large larval counts ($y > 30$). Moreover, the posterior predictive point estimates (median) for larval counts were equally good across all spatially explicit population growth models when measured by their root mean squared error. The lppd and scrps scores measure simultaneously both the predictive point estimate and uncertainty so that they prefer less uncertain predictions when the predictive point estimates are equal (this phenomenon is stronger for lppd than (s)crps Gneiting, Balabdaoui, and Raftery 2007). Hence, the results indicate that, due to two processes for zeros, the predictions of the zero-inflated models contain more uncertainty than the corresponding predictions of hurdle models.

TABLE 3 | Posterior predictive model comparison scores in interpolation (LOO-CV) and extrapolation for larvae data (3-fold CV): log predictive density (lppd), and scaled continuous ranked probability score (scrps). We denote zero-inflated models with ZI.

				Spatially explicit population growth models					
		SDMs		Ricker		Beverton–Holt		Density-independent	
		ZI	Hurdle	ZI	Hurdle	ZI	Hurdle	ZI	Hurdle
LOO-CV									
lppd	larvae	−866	−846	−869	−847	−870	−847	−869	−847
	catch	NA	NA	−54	−55	−54	−54	−54	−51
	tot	NA	NA	−923	−902	−925	−901	−923	−898
scrps	larvae	−2.70	−2.74	−2.81	−2.71	−2.76	−2.71	−2.74	−2.70
	catch	NA	NA	−0.99	−0.98	−0.99	−0.92	−0.99	−0.96
	tot	NA	NA	−3.80	−3.69	−3.75	−3.63	−3.73	−3.66
3-fold-CV									
lppd	larvae	−970	−942	−944	−922				
scrps	larvae	−5.16	−5.31	−4.31	−4.33				

In summary, differences in predictive performances among the compared models were small but indicated a preference for the spatially explicit population growth models. The predictive comparisons did not conclusively rank the zero-inflated (7) and hurdle (13) formulations for the spatially explicit population growth models. Hence, the choice of the model for final inference should also be informed by the models' ecological realism. This supports the choice of a zero-inflated, spatially explicit population growth model (7). Zero-inflated models are in general more realistic for ecological data than hurdle models (Martin et al. 2005; Blasco-Moreno et al. 2019) and in our case study this is especially true. Even though we used the same formulation (and notation) for the model parameters under the hurdle (13) and the zero-inflated (7) formulation of the spatially explicit population growth model, their interpretation changes between the models. Under the hurdle formulation, the parameter θ could be interpreted as suitability only if there was no egg or larvae mortality and larval observation probability was one. This is clearly not the case since, even though survival of coregonid eggs from fertilization in fall to hatching in spring can vary substantially, it is reported to be generally approximately 2%–7% (Freeberg, Taylor, and Brown 1990). Hence, the parameter θ in (13) is only a technical probability of presence summarizing all processes that might lead to absence. Moreover, in the hurdle model (13) the expected number of observed larvae is $\tilde{\eta}_{i,t+1} V_i \theta_i (1 - (\frac{r}{r + V_i \tilde{\eta}_{i,t+1}})^r)^{-1}$. Hence, we cannot interpret η as the expected spawner density nor a as the maximum proliferation rate even though they are related to them. In the hurdle model, $\tilde{\eta}_{i,t+1}$ is a good proxy of the expected larvae density only when the mass of the Negative-Binomial at zero is small (Feng 2021). However, a practical benefit from using the hurdle model is that they are typically better identifiable than zero-inflated models with finite data and, thus, give sharper predictions (Blasco-Moreno et al. 2019; Tang et al. 2022, see also Section 2 in the Supporting Information). Since our main interest is in inference, we chose the Ricker zero-inflated, spatially explicit annual growth model for further examination of the whitefish data, to be presented in the next sections. However, we compared the inferential results from this model to the corresponding hurdle model and SDM to shed light on how the model choice matters for the inference.

5.2.2 | Parameter Posteriors and Covariate Effects

Most of the parameters of the spatially explicit (zero-inflated) Ricker population growth model were identified well, as indicated by clear differences between the posterior and prior distributions (see Figure 4A and Figure A.3 and Table A.1 in the Supporting Information), the only exceptions being the intercept parameters ($\overline{\alpha}$, $\bar{a}$, and α). The overdispersion parameter of the negative-Binomial distribution (r) was smaller than one, indicating large local variability in the larval density. Since also the posterior for the density-dependent mortality parameter b was updated from its prior, the data were informative on density-dependent effects (Figure A.3 in the Supporting Information). To our slight surprise, also the fisheries catch rate parameter (λ) was identified well (posterior mean 0.0058 and 95% CI [0.0012,0.0246] [net days]$^{-1}$) even though catch rate parameters are typically weakly identifiable in fisheries models. The obtained catch rate estimate was also reasonable. The fishing gear considered in this study targets adult, approximately 3+ year old, whitefish. The demographic survey samples of whitefish from the Gulf of Bothnia have included individuals from all age classes up to 14-year-old fishes (unpublished data of the Natural Resources Institute Finland), indicating that yearly fishing mortality must be less than approximately 20%, which corresponds to $(1 - 0.2)^{12} = 0.07$ probability to survive fishing from age 2 to 14. The fishing effort in the most heavily fished ICES statistical rectangles (11, 15, 19, and 20; water areas ranging from 218 to 676 km^2) ranged from 15×10^3 to 25×10^3 net days (Figure 1). If these fishing efforts were spread equally across the ICES statistical rectangles, 0.2 fishing mortality would correspond to catch rates from 0.02 to 0.1. However, since spawning time fishing concentrates near shores, the catch rate is likely much lower than these crude estimates, supporting our estimate.

To study the effects of environment on spawning suitability, spawner density, and survival of eggs, we drew samples from the posterior of the conditional (centered) covariate response curves (see Vanhatalo, Foster, and Hosack 2021). We also assessed the relative importance of different environmental covariates on suitability, spawner density, and survival of eggs by partitioning the total variance of the respective Gaussian latent variables over the whole study area between the contributions from the covariates (Schulz, Saastamoinen, and Vanhatalo 2024). The three processes, suitability, spawner density, and maximum proliferation rate, responded to environmental covariates somewhat differently. Most of the variability in sites' suitability for spawning was explained by distance to sandy shore, river influence (negative effect), distance to deep (bell-shaped effect), and winter ice coverage (positive impact; Figure 4). For spawner density, most of the variability was explained by distance to deep, exposure to wind, and chlorophyll-a concentration, all having unimodal (bell-shaped) response curves with optimal values within the covariate range. Most of the variability in maximum proliferation rate was explained by distance to sand and salinity (negative effects).

The above-summarized posterior distributions differed from the posterior distributions from the hurdle formulation of the spatially explicit Ricker population growth model (Figure A.4), demonstrating that it matters whether one uses a zero-inflated or hurdle model. Under the hurdle formulation, the density-dependent mortality parameter (b) did not identify (posterior distribution was equal to the prior distribution), the intercept of the logit occurrence probability ($\overline{\alpha}$) was estimated to be smaller than in the zero-inflated formulation (which is reasonable since the hurdle model explains all zeros by the Bernoulli process), and the fisheries catch rate was estimated to be approximately 10 times bigger than under the zero-inflated formulation (being at the upper end of the crude estimate for realistic level outlined above). Also, some of the response curves differed between the zero-inflated and hurdle formulations of the spatially explicit population growth models (Figure A.2 in the Supporting Information).

5.2.3 | Spatial Predictions

To study the spatial patterns in the processes underlying whitefish larvae production, we used the spatially explicit (zero-inflated) Ricker population growth model to predict the

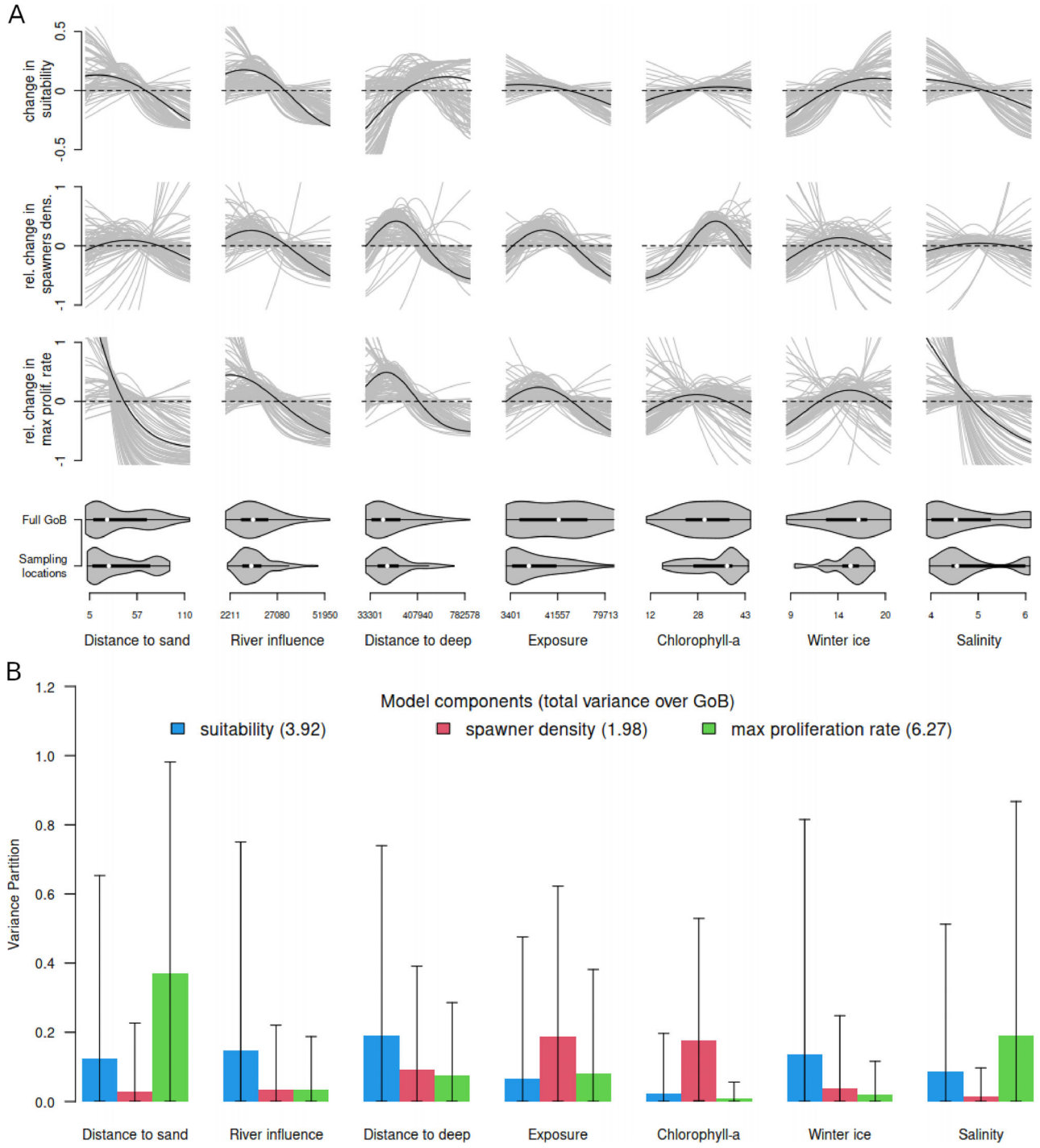

FIGURE 4 | (A) The change in spawning suitability and the relative change in spawner density and egg survival along the environmental covariates according to the spatially explicit Ricker population growth model (posterior samples of the curves with grey lines and the mean of them with a black line). The responses are visualized over the range of the covariate values over the whole study region (upper violin plot). The lower violin plots show the distribution of covariate values in the training data. (B) The proportion of variance of the logit suitability, log spawner density, and log maximum proliferation rate over the Gulf of Bothnia (GoB) explained by the environmental covariates (mean and 95% credible interval).

(average over study years) suitability, spawner density, proliferation rate, and the larvae density over the study area. In broad spatial scale, suitability, spawner density, and proliferation rate followed similar patterns so that all of them are relatively higher in the north than in the south (Figure 5A), the difference between north and south being most distinct for suitability (Figure 5B). However, at a local scale, all these processes showed distinct patterns. For example, in the south (Bothnian Sea), the model predicted significant concentrations of spawners but low maximum proliferation rates, whereas both of them were high in the north, indicating that whitefish reproduction success relative to population density is worse in the south than in the

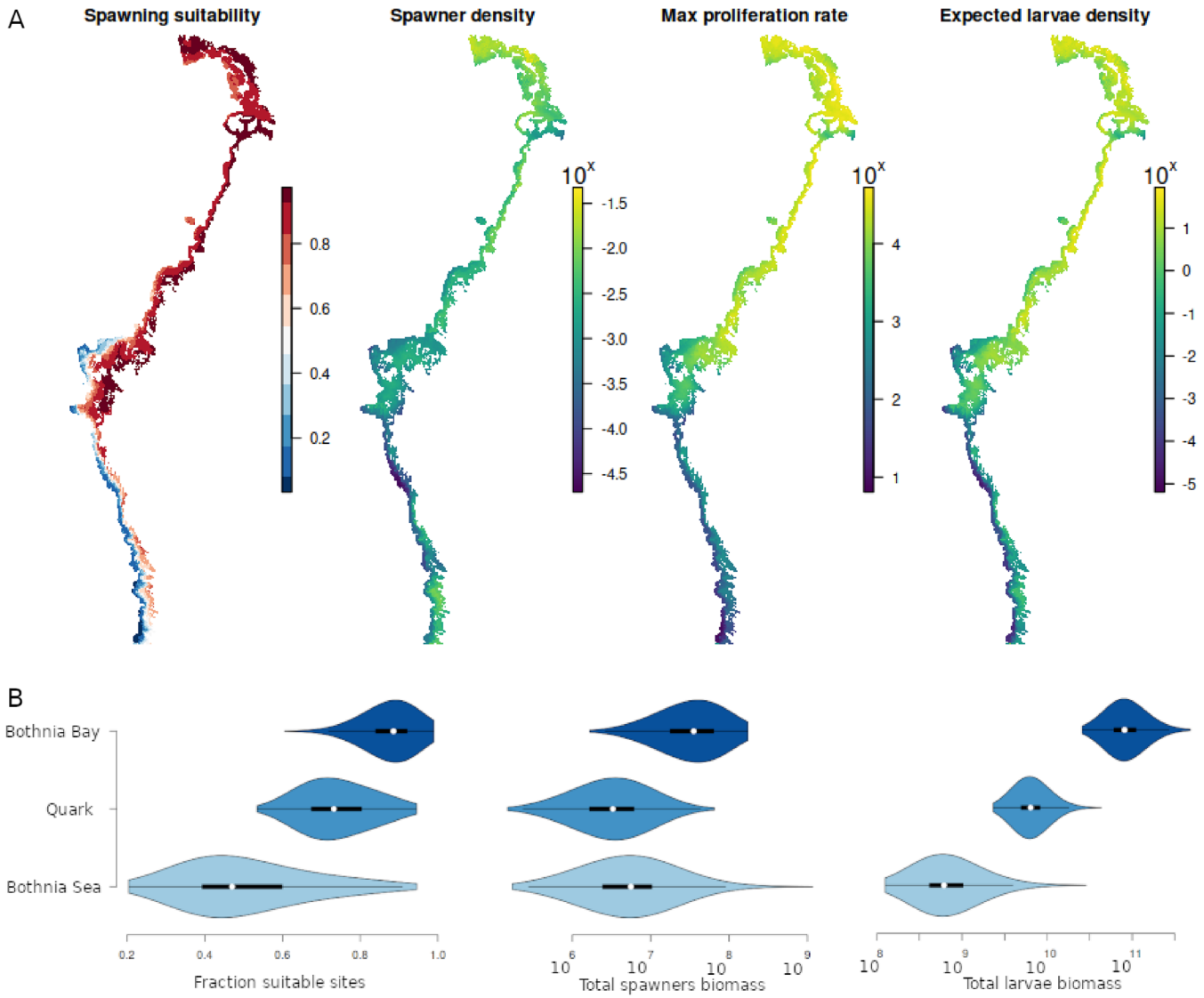

FIGURE 5 | (A) Predictive maps of the posterior median of suitability, spawner density, proliferation rate, and expected larvae density along the Finnish coastal area of the Gulf of Bothnia (GoB). (B) Posterior distribution for the fraction of suitable sites, the total spawner biomass, and the total larval production in the three GoB subregions: the Bothnian Bay, the Quark, and the Bothnian Sea (see Figure 1A).

north. The maximum proliferation rate also expresses a higher distinction between the nearshore and the pelagic areas than the spawner density (Figure 5A). Our spawner density predictions ($10^{-2} \ldots 10^{-1}$ fish/m^2; Figure 5) are the first spatially explicit, and quantitative estimates for the Baltic Sea whitefish. These density estimates correspond to approximately 20–300 kg of fish per hectare, which are realistic values in spawning aggregations. Estimates on the biomass densities of whitefish in lakes are approximately 5–30 kg per hectare (Rask and Arvola 1985; Linløkken 1995) but, at spawning time, whitefishes typically aggregate near spawning sites (Semenchenko and Smeshlivaya 2021), which results in at least an order of magnitude increase in whitefish catch-per-unit effort in spawning time aggregations (Veneranta and Harjunpää 2021) compared to other times (OSF 2020). As expected, the predicted larvae density was a compromise between all three processes (Figure 5A). The total number of larvae was predicted to be highest in the Bothnian Bay and the lowest in the Bothnian Sea (Figure 5B). The hurdle formulation of the spatially explicit Ricker population growth model (13) differed from the above summarized results most clearly in its larval occurrence predictions, which showed more pronounced differences across the region than the suitability predictions of the zero-inflated formulation (compare Figure 5 and Figure A.10 in the Supporting Information). The hurdle formulation also predicted somewhat lower spawner densities than the zero-inflated formulation. The zero-inflated SDM predicted similar occurrence across the region as the spatially explicit (zero-inflated) Ricker population growth model, but the SDM's prediction for the larvae density conditional on suitability was different from both the spawner density and the maximum proliferation rate predictions. This is reasonable since the spatially explicit (zero-inflated) Ricker population growth model and the zero-inflated SDM had exactly the same functional formulation for the suitability parameter (θ), whereas the former split the conditional larvae density into spawner density and maximum proliferation rate. The marginal larvae density prediction was similar among all compared spatially explicit population growth models

and SMDs, which is reasonable when reflected with the predictive comparison results (Section 5.2.1).

6 | Discussion and Conclusions

In this work, we developed a spatially explicit population growth model to study the effects of the environment on annual reproduction of animals. Our model combines methodological elements from the population dynamics modeling (e.g., the Ricker model), species distribution modeling, and data integration approaches. We followed the general hierarchical Bayesian approach to data integration in spatiotemporal models, wherein the true underlying processes are modeled with latent models to which the complementary data sources (larvae data and fisheries data herein) are linked with likelihood functions (e.g., Gelfand, Zhu, and Carlin 2001; Berrocal, Gelfand, and Holland 2009; Vanhatalo et al. 2016; Miller et al. 2019; Vanhatalo, Hartmann, and Veneranta 2020; Isaac et al. 2020). As a result, we arrived at a model that allows inference and predictions on the effects of environment on suitability, spawner density, and the survival of eggs and larvae in a spatially explicit manner. The model could also be used to study the effect of spatially explicit fisheries management on larval production, but we did not explore this option here.

Our results show that the proposed model is identifiable for its key parameters both theoretically (large data limit, Theorem 1) and empirically (finite data, Section 5.1). These results held also when the three processes shared exactly the same covariates. In previous works, to improve identifiability, the Bernoulli and the count processes of zero-inflated models have typically been given distinct sets of covariates (see, e.g., Tang et al. 2022). To avoid the covariate selection step, we instead used shrinkage priors that pulled covariate effects towards zero (Section 1 in the Supporting Information) and, thus, decreased confounding between the suitability, spawner density, and maximum proliferation rate processes. The likelihood of the proposed model is not fully identifiable for all scale parameters, though, and it might be weakly identifiable with a given dataset. Hence, we advise using informative and shrinkage priors where possible and provided an example of constructing them by our case study (Section 1 in the Supporting Information).

Many studies in ecology have stressed the importance of accounting for different potential sources of animal absence and exemplified how the assumptions underlying hurdle models are rarely ecologically realistic (Martin et al. 2005; Blasco-Moreno et al. 2019). Our results strengthen this message. In our whitefish case study, the inference results from the spatially explicit (zero-inflated) Ricker population growth model were clearly different from the inference results when using its hurdle reformulation. These differences have also practical implications. For example, when selecting conservation actions, it can make a difference whether the southern Gulf of Bothnia is estimated to be unsuitable for spawning, as predicted by the hurdle formulation (Figure A.10 in the Supporting Information), or suitable for spawning but suffering from low larval survival, as predicted by the proposed (zero-inflated) model (Figure 5).

Our case study also indicates that care is needed in applying predictive model comparison methods to select a model when the focus is on inference. Even though the proposed spatially explicit population growth model did not clearly improve LOO-CV predictive scores for larvae counts over traditional SDMs (Table 3), by separating the suitability, spawner density, and maximum proliferation rate processes in a spatially explicit manner, it provided new information on processes underlying whitefish reproduction that is not attainable with SDMs. This leads also to more realistic extrapolation predictions when doing, for example, climate change scenario predictions. Moreover, the proposed spatially explicit Ricker population growth model provided better extrapolation predictions for larvae counts than the SDM. This result aligns well with earlier studies that have shown that accounting for preferential sampling (i.e., sampling that depends upon the biomass of the target species) associated with commercial data in fishery reduces bias in fish abundance and biomass estimates (Pennino et al. 2019; Alglave et al. 2022).

One technical aspect of our model that deserves a comment is the spatial random effects – or the lack of them from the site's suitability and maximum proliferation rate processes. Including spatiotemporal random effects for all three processes would have been computationally infeasible since the study area comprised almost 10^5 lattice grid cells. Earlier studies on this larval data have also shown that, after accounting for the covariates, the spatial structure in the data is negligible (Vanhatalo, Veneranta, and Hudd 2012; Vanhatalo, Hartmann, and Veneranta 2020). Hence, the spatiotemporal random effect for the spawner density was mainly motivated by possible annual variation in spawning stock size and was encoded in a computationally feasible manner over the 18 ICES statistical rectangles. Extending the model to allow spatial random effects also for sites' suitability and the maximum proliferation rate is, thus, left for the future development.

The fundamental motivation behind our work was to generate a method that can be used to analyze how the different processes underlying animal reproduction are affected by the local scale variation in environmental conditions and population distribution. Understanding these local-scale processes is important for sustainable development since conservation decisions and natural resources management actions typically take place at a local or regional scale. The model developed in this work tackles this challenge in the context of annually reproducing fish. However, a similar model could be applied in other contexts as well. Informative priors can be elicited from the literature for many other species, and the observation models can be modified to fit other types of observations and spatial supports using the principles outlined here. For example, our fisheries catch data and their associated model (11) could be changed to mark-recapture data and model (Royle 2004), distance sampling data and point process model (Yuan et al. 2017) or census survey data and their model (Schulz, Vanhatalo, and Saastamoinen 2020). Hence, the principles underlying our approach are generic and can be extended for other systems.

Acknowledgments

This work has received funding from the Doctoral Programme in Mathematics and Statistics of the University of Helsinki Academy of Finland (grant 317255), Jane & Aatos Erkko Foundation (IP, JV, EN), the European Union's Horizon Europe research and innovation programme under

grant agreement No. 101081642 (JV, IP), and from the Doctoral Program in Mathematics and Statistics of the University of Helsinki (IP). In addition, JV acknowledges funding from the European Union (ERC Consolidator Grant BEFPREDICT, 101087409).

Data Availability Statement

The data that supports the findings of this study are available in the Supporting Information of this article.

References

Alglave, B., E. Rivot, M.-P. Etienne, M. Woillez, J. T. Thorson, and Y. Vermard. 2022. "Combining Scientific Survey and Commercial Catch Data to Map Fish Distribution." *ICES Journal of Marine Science* 79, no. 4: 1133–1149.

Andersen, J., P. Axe, H. Backer, et al. 2011. "Getting the Measure of Eutrophication in the Baltic Sea: Towards Improved Assessment Principles and Methods." *Biogeochemistry* 106, no. 2: 137–156.

Berrocal, V., A. E. Gelfand, and D. M. Holland. 2009. "A Spatio-Temporal Downscaler for Output From Numerical Models." *Journal of Agricultural, Biological and Environmental Statistics* 15, no. 2: 176–197.

Blasco-Moreno, A., M. Pérez-Casany, P. Puig, M. Morante Moret, and E. Castells. 2019. "What Does a Zero Mean? Understanding False, Random and Structural Zeros in Ecology." *Methods in Ecology and Evolution* 10: 949–959.

Bolin, D., and J. Wallin. 2023. "Local Scale Invariance and Robustness of Proper Scoring Rules." *Statistical Science* 38, no. 1: 140–159.

Brännström, Å., and D. J. T. Sumpter. 2005. "The Role of Competition and Clustering in Population Dynamics." *Proceedings of the Royal Society B* 272, no. 1576: 2065–2072.

Crozier, L. G., and J. A. Hutchings. 2014. "Plastic and Evolutionary Responses to Climate Change in Fish." *Evolutionary Applications* 7, no. 1: 68–87.

Elith, J., and J. R. Leathwick. 2009. "Species Distribution Models: Ecological Explanation and Prediction Across Space and Time." *Annual Review of Ecology, Evolution, and Systematics* 40: 677–697.

Feng, C. 2021. "A Comparison of Zero-Inflated and Hurdle Models for Modeling Zero-Inflated Count Data." *Journal of Statistical Distributions and Applications* 8: 8.

Franklin, J. 2023. "Species Distribution Modelling Supports the Study of Past, Present and Future Biogeographies." *Journal of Biogeography* 00: 1–13.

Freeberg, M. H., W. W. Taylor, and R. W. Brown. 1990. "Effect of Egg and Larval Survival on Year-Class Strength of Lake Whitefish in Grand Traverse Bay, Lake Michigan." *Transactions of the American Fisheries Society* 119, no. 1: 92–100.

Gabry, J., D. Simpson, A. Vehtari, M. Betancourt, and A. Gelman. 2019. "Visualization in Bayesian Workflow." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 182, no. 2: 389–402.

Gelfand, A. E., L. Zhu, and B. P. Carlin. 2001. "On the Change of Support Problem for Spatio-Temporal Data." *Biostatistics* 2, no. 1: 31–45.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2014. *Bayesian Data Analysis*. third ed. Boca Raton, FL: CRC Press, Taylor and Francis Group.

Geritz, S., and E. Kisdi 2004. "On the Mechanistic Underpinning of Discrete-Time Population Models With Complex Dynamics." *Journal of Theoretical Biology* 228, no. 2: 261–269.

Gneiting, T., F. Balabdaoui, and A. E. Raftery. 2007. "Probabilistic Forecasts, Calibration and Sharpness." *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 69, no. 2: 243–268.

Haapala, J., and M. Leppäranta. 1997. "The Baltic Sea Ice Season in Changing Climate." *Boreal Environment Research* 2: 93–108.

Håkansson, B., P. Alenius, and L. Brydsten 1996. "Physical Environment in the Gulf of Bothnia." *Ambio Special Report* 8: 5–12.

Heikinheimo, O., Z. Pekcan-Hekim, and J. Raitaniemi. 2014. "Spawning Stock-Recruitment Relationship in Pikeperch Sander Lucioperca (l.) in the Baltic Sea, With Temperature as an Environmental Effect." *Fisheries Research* 155: 1–9.

Helsinki Commission. 2002. "Environment of the Baltic Sea Area 1994-1998. Baltic Sea Environment Proceedings No. 82A."

Himberg, M., M. von Numers, A. Vasemägi, et al. 2015. "Gill Raker Counting for Approximating the Ratio of River- and Sea-Spawning Whitefish, Coregonus Lavaretus (Actinopterygii: Salmoniformes: Salmonidae) in the Gulf of Bothnia, Baltic Sea." *Acta Ichthyologica et Piscatoria* 45: 125–131.

Hutchinson, G. E. 1978. *An Introduction to Population Ecology*. London, England: Yale University Press, Ltd.

Isaac, N. J., M. A. Jarzyna, P. Keil, et al. 2020. "Data Integration for Large-Scale Models of Species Distributions." *Trends in Ecology & Evolution* 35, no. 1: 56–67.

Kallasvuo, M., J. Vanhatalo, and L. Veneranta. 2017. "Modeling the Spatial Distribution of Larval Fish Abundance Provides Essential Information for Management." *Canadian Journal of Fisheries and Aquatic Sciences* 74, no. 5: 636–649.

Kallio-Nyberg, I., L. Veneranta, I. Saloniemi, E. Jokikokko, and A. Leskelä. 2019. "Different Growth Trends of Whitefish (Coregonus Lavaretus) Forms in the Northern Baltic Sea." *Journal of Applied Ichthyology* 35, no. 3: 683–691.

Kankaanpää, H., P. Alenius, P. Kotilainen, and P. Roiha. 2022. "Decreased Surface and Bottom Salinity and Elevated Bottom Temperature in the Northern Baltic Sea Over the Past Six Decades." *Science of the Total Environment* 859: 160241.

Kaurila, K., S. Kuningas, A. Lappalainen, and J. Vanhatalo. 2022. "Species Distribution Modeling With Expert Elicitation and Bayesian Calibration." arXiv:2206.08817.

Kraufvelin, P., Z. Pekcan-Hekim, U. Bergström, et al. 2018. "Essential Coastal Habitats for Fish in the Baltic Sea." *Estuarine, Coastal and Shelf Science* 204: 14–30.

Lehtonen, H. 1981. "Biology and Stock Assessments of Coregonids by the Baltic Coast of Finland." *Finnish Fisheries Research* 3: 31–83.

Leppäranta, M. 2022. "History and Future of Snow and Sea Ice in the Baltic Sea." https://doi.org/10.1093/acrefore/9780190228620.013.891.

Leppäranta, M., and K. Myrberg. 2009. *Physical Oceanography of the Baltic Sea*. Berlin, Heidelberg: Springer.

Linløkken, A. 1995. "Monitoring Pelagic Whitefish (*Coregonus Lavaretus*) and Vendace (*Coregonus Albula*) in a Hydroelectric Reservoir Using Hydroacoustics." *Regulated Rivers: Research & Management* 10: 315–328.

Lundberg, C., B.-M. Jakobsson, and E. Bonsdorff. 2009. "The Spreading of Eutrophication in the Eastern Coast of the Gulf of Bothnia, Northern Baltic Sea – An Analysis in Time and Space." *Estuarine, Coastal and Shelf Science* 82, no. 1: 152–160.

Macura, B., P. Byström, L. Airoldi, B. K. Eriksson, L. Rudstam, and J. G. Støttrup. 2019. "Impact of Structural Habitat Modifications in Coastal Temperate Systems on Fish Recruitment: A Systematic Review." *Environmental Evidence* 8, no. 1: 14.

Martin, T., B. Wintle, J. Rhodes, et al. 2005. "Zero Tolerance Ecology: Improving Ecological Inference by Modelling the Source of Zero Observations." *Ecology Letters* 8: 1235–1246.

Miller, D. A. W., K. Pacifici, J. S. Sanderlin, and B. J. Reich. 2019. "The Recent Past and Promising Future for Data Integration Methods to Estimate Species' Distributions." *Methods in Ecology and Evolution* 10, no. 1: 22–37.

Müller, R. 1992. "Trophic State and Its Implications for Natural Reproduction of Salmonid Fish." *Hydrobiologia* 243-244: 261–268.

Munch, S., J. Giron Nava, and G. Sugihara. 2018. "Nonlinear Dynamics and Noise in Fisheries Recruitment: A Global Meta-Analysis." *Fish and Fisheries* 19, no. 6: 964–973.

Myers, R. 2001. "Stock and Recruitment: Generalizations About Maximum Reproductive Rate, Density Dependence, and Variability Using Meta-Analytic Approaches." *ICES Journal of Marine Science - ICES J MAR SCI* 58: 937–951.

OSF. 2020. *Official Statistics Finland. Commercial Marine Fishery*. Helsinki: Natural Resources Institute Finland.

Peltonen, H., and B. Weigel. 2022. "Responses of Coastal Fishery Resources to Rapid Environmental Changes." *Journal of Fish Biology* 101, no. 3: 686–698.

Pennino, M. G., I. Paradinas, J. B. Illian, et al. 2019. "Accounting for Preferential Sampling in Species Distribution Models." *Ecology and Evolution* 9, no. 1: 653–663.

Peterson, A. T., J. Soberón, R. G. Pearson, et al. 2011. *Ecological Niches and Geographic Distributions (MPB-49)*. Princeton, New Jersey: Princeton University Press.

Pörtner, H. O., and M. A. Peck. 2010. "Climate Change Effects on Fishes and Fisheries: Towards a Cause-and-Effect Understanding." *Journal of Fish Biology* 77, no. 8: 1745–1779.

Pulkkinen, H., and S. Mäntyniemi. 2013. "Maximum Survival of Eggs as the Key Parameter of Stock–Recruit Meta-Analysis: Accounting for Parameter and Structural Uncertainty." *Canadian Journal of Fisheries and Aquatic Sciences* 70, no. 4: 527–533.

Quinn, T. J., and R. B. Deriso. 1999. *Quantitative Fish Dynamics*. New York, New York: Oxford University Press.

Rahikainen, M., K.-M. Hoviniemi, S. Mäntyniemi, et al. 2017. "Impacts of Eutrophication and Oil Spills on the Gulf of Finland Herring Stock." *Canadian Journal of Fisheries and Aquatic Sciences* 74, no. 8: 1218–1232.

Rask, M., and L. Arvola. 1985. "The Biomass and Production of Pike, Perch and Whitefish in Two Small Lakes in Southern Finland." *Annales Zoologici Fennici* 22: 129–136.

Roberts, D. R., V. Bahn, S. Ciuti, et al. 2017. "Cross-Validation Strategies for Data With Temporal, Spatial, Hierarchical, or Phylogenetic Structure." *Ecography* 40, no. 8: 913–929.

Royle, J. A. 2004. "N-Mixture Models for Estimating Population Size From Spatially Replicated Counts." *Biometrics* 60, no. 1: 108–115.

Schulz, T., M. Saastamoinen, and J. Vanhatalo. 2024. "Model-Based Variance Partitioning for Statistical Ecology." *Ecological Monographs*: 10–1101.

Schulz, T., J. Vanhatalo, and M. Saastamoinen. 2020. "Long-Term Demographic Surveys Reveal a Consistent Relationship Between Average Occupancy and Abundance Within Local Populations of a Butterfly Metapopulation." *Ecography* 43, no. 2: 306–317.

Semenchenko, S., and N. V. Smeshlivaya. 2021. "Spawning Behaviour of Whitefishes (Coregonidae)." *Annales Zoologici Fennici* 58: 129–140.

Stan Development Team. 2023. "RStan: The R Interface to Stan." *R Package Version* 2, no. 26: 23.

Tang, B., H. A. Frye, A. E. Gelfand, and J. A. Silander. 2023. "Zero-Inflated Beta Distribution Regression Modeling." *Journal of Agricultural, Biological and Environmental Statistics* 28, no. 1: 1–21.

Vanhatalo, J., S. D. Foster, and G. R. Hosack. 2021. "Spatiotemporal Clustering Using Gaussian Processes Embedded in a Mixture Model." *Environmetrics* 32, no. 7: e2681.

Vanhatalo, J., M. Hartmann, and L. Veneranta. 2020. "Additive Multivariate Gaussian Processes for Joint Species Distribution Modeling With Heterogeneous Data." *Bayesian Analysis* 15, no. 2: 415–447.

Vanhatalo, J., A. J. Hobday, L. R. Little, and C. M. Spillman. 2016. "Downscaling and Extrapolating Dynamic Seasonal Marine Forecasts for Coastal Ocean Users." *Ocean Modelling* 100: 20–30.

Vanhatalo, J., L. Veneranta, and R. Hudd. 2012. "Species Distribution Modeling With Gaussian Processes: A Case Study With the Youngest Stages of Sea Spawning Whitefish (Coregonus Lavaretus l. s.l.) Larvae." *Ecological Modelling* 228: 49–58.

Vehtari, A., A. Gelman, and J. Gabry. 2016. "Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and WAIC." *Statistics and Computing* 27, no. 5: 1413–1432.

Vehtari, A., A. Gelman, D. Simpson, B. Carpenter, and P.-C. Bürkner. 2021. "Rank-Normalization, Folding, and Localization: An Improved $\hat{R}$ for Assessing Convergence OF MCMC (With Discussion)." *Bayesian Analysis* 16, no. 2: 667–718.

Vehtari, A., D. Simpson, A. Gelman, Y. Yao, and J. Gabry. 2024. "Pareto Smoothed Importance Sampling." *Journal of Machine Learning Research* 25: 1–57.

Veneranta, L., and H. Harjunpää. 2021. *Merenkurkun Merikutuisen Siian Istutustuotto Ja Syönnösalueet. Luonnonvara- Ja Biotalouden Tutkimus 59/2021*, 35. Helsinki: Natural Resources Institute Finland.

Veneranta, L., R. Hudd, and J. Vanhatalo. 2013. "Reproduction Areas of Sea-Spawning Coregonids Reflect the Environment in Shallow Coastal Waters." *Marine Ecology Progress Series* 477: 231–250.

Veneranta, L., L. Urho, J. Koho, and R. Hudd. 2013. "Spawning and Hatching Temperatures of Whitefish (Coregonus Lavaretus (l.)) in the Northern Baltic Sea." *Advances in Limnology* 64: 39–55.

Veneranta, L., J. Vanhatalo, and L. Urho. 2016. "Detailed Temperature Mapping-Warming Characterizes Archipelago Zones." *Estuarine, Coastal and Shelf Science* 182: 123–135.

Verliin, A., L. Saks, R. Svirgsden, et al. 2011. "Whitefish (Coregonus lavaretus (L.)) Landings in the Baltic Sea During the Past 100 years: Combining Official Datasets and Grey Literature." *Advances in Limnology* 64: 133–152.

Voipio, A. 1981. *The Baltic Sea*, edited by A. Voipio. New York: Elsevier Scientific Pub. Co. Distributors for the U.S. and Canada, Elsevier/North-Holland Amsterdam, New York.

Weigel, B., J. Mäkinen, M. Kallasvuo, and J. Vanhatalo. 2021. "Exposing Changing Phenology of Fish Larvae by Modeling Climate Effects on Temporal Early Life-Stage Shifts." *Marine Ecology Progress Series* 666: 135–148.

Yuan, Y., F. E. Bachl, F. Lindgren, et al. 2017. "Point Process Models for Satio-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.