Ecological Modelling 457 (2021) 109680

Contents lists available at ScienceDirect

Ecological Modelling

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

Bridging mechanistic conceptual models and statistical species distribution models of riverine fish

Bogdan Caradima [a,b,*], Andreas Scheidegger [a], Jakob Brodersen [c], Nele Schuwirth [a,b]

[a] Eawag: Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, Dübendorf 8600, Switzerland
[b] ETH Zurich, Institute of Biogeochemistry and Pollutant Dynamics, Universitätstrasse 16, Zürich 8092, Switzerland
[c] Eawag: Swiss Federal Institute of Aquatic Science and Technology, Seestrasse 79, Kastanienbaum 6047, Switzerland

A B S T R A C T

Statistical species distribution models (SDMs) are widely used to quantify how taxa respond to environmental conditions and to predict their distribution. However, the application of SDMs to freshwater fish taxa is complicated by the active dispersal of fish taxa through river networks, and the species- and habitat-dependent observation process (i.e., the sampling method and effort) required to accurately sample their distributions. Many studies have applied presence-absence models (PAMs) to fish taxa, while more recent studies have proposed zero-inflated models (ZIMs) to account for count observations with many zeroes. However, relatively few studies have incorporated the observation process into the model structure, which would facilitate the combination of data from various monitoring programs that differ in their observation process. In this study, we use conceptual models to identify potentially dominant natural and anthropogenic environmental conditions with a direct, mechanistic effect on the distributions of freshwater fish taxa in Switzerland, a region with a large range of environmental conditions, from alpine streams that are mainly affected by hydromorphological alterations to lowland streams in densely populated areas with intensive agricultural land use. Moreover, numerous barriers impede fish migration along the entire river network. Using combined data from two fish monitoring programs in Switzerland, we applied an exhaustive cross-validation procedure to select a set of environmental variables with the highest (out-of-sample) predictive performance for the PAM and ZIM for fish density (individuals/m^2) of the seven most prevalent fish taxa (*Salmo spp., Cottus spp., Squalius spp., Barbatula spp., Barbus spp., Phoxinus spp., Gobio spp.*). We used these variables to develop a PAM and ZIM for each taxon that accounts for differences in sampling methods and sampling effort. We quantified the quality of fit during calibration using all samples and predictive performance during 5-fold cross-validation of each model.

Results show that stream temperature and stream morphology within the accessible habitat commonly appear among the best predictive presence-absence models for multiple taxa. Spatial variables that account for migration barriers and quantify morphological conditions within the accessible habitat were selected for 6 out of 7 taxa. The selected PAMs performed well for all taxa with an intermediate prevalence (10–40%), with an explanatory power (D^2) of between 0.32 - 0.37 during calibration using all samples and only minor decreases in explanatory power during cross-validation ($D^2 = 0.34 – 0.44$). As expected, the PAM for the highly prevalent *Salmo spp*. (91%) failed to predict the few absence data points. By contrast, the ZIM model performed best for *Salmo spp.*, with a standardized likelihood ratio of 1.56. For all other taxa besides *Barbus spp.* the ZIM models also had likelihood ratios above one, indicating a better predictive performance than the null model. We hope this study stimulates the development and application of fish species distribution models based on prior knowledge of causally linked environmental variables and incorporating observation errors to improve their predictive performance. This can facilitate learning from biomonitoring data to support management.

1. Introduction

Freshwater ecosystems such as lakes and rivers are rich in biodiversity, however over-exploitation places freshwater ecosystems at greater risk of habitat destruction and degradation than their terrestrial and marine counterparts (Dudgeon et al., 2006; Vörösmarty et al., 2010). Within the European Union, nearly half (47%) of lakes and rivers failed to achieve a good ecological status in 2015 (as defined by the EU Water Framework Directive using indicators of the quality of biological communities), with many freshwater species either increasingly threatened or endangered, including mammals, birds, fish, insects and other invertebrate communities.

Fish species in Europe are of particular concern, with 37% of the 547 native species listed as threatened or endangered, and 17% of species populations in decline (Brooks and Freyhof, 2011). The main anthropogenic threats to Europe's freshwater fish species are the destruction, degradation, and fragmentation of habitat due to the channelization of natural river courses, pollution from intensive agriculture and urban areas, the construction of dams, and water abstraction (Gozlan et al., 2019). Studies have emphasized the need for additional biomonitoring

* Corresponding author.
E-mail addresses: b.caradima@gmail.com (B. Caradima), andreas.scheidegger@eawag.ch (A. Scheidegger), jakob.brodersen@eawag.ch (J. Brodersen), nele.schuwirth@eawag.ch (N. Schuwirth).

https://doi.org/10.1016/j.ecolmodel.2021.109680
Received 14 January 2021; Received in revised form 17 June 2021; Accepted 22 July 2021
Available online 17 August 2021

efforts to more accurately characterize the geographic distribution and population trends of freshwater fish species to (1) improve future risk assessments, (2) identify underlying environmental drivers of species' distributions, and (3) better inform stream management (Gozlan et al., 2019; Vörösmarty et al., 2010).

Accurately assessing the total population of a fish species in a given river length can be difficult due to the active dispersal capabilities of fish (e.g., relative to benthic macroinvertebrates) and the observation error inherent to imperfect fish sampling methods. Electrofishing is perhaps the most common fishing method for stream biomonitoring, and involves passing a high-voltage current through a length of river to stun fish for easy collection and analysis. Additional sampling efforts may increase the proportion of fish caught (i.e., reducing observation error), including the placement of nets at the start and end points of the fished area to prevent individual fish from escaping prior to sampling, and performing multiple electrofishing "passes" or "rounds" to accurately quantify the number of individuals in a reach. Multiple fishing rounds can provide particularly useful data for fisheries management, including population estimates that can be derived based on the number of fish captured in successive fishing rounds and on prior knowledge of the probability of capturing an individual of a given species (e.g., Carle and Strub, 1978).

In obtaining presence-absence and abundance (i.e., count) observations from biomonitoring data, studies over the past two decades have applied statistical species distribution models (SDMs) to quantify how observed distributions of species respond to a range of natural and anthropogenic environmental conditions. Techniques in statistical and machine learning have been applied, ranging from classical statistical models such as generalized linear models (GLMs) and generalized additive models (GAMs) (e.g., Fukushima et al., 2007; Olden and Jackson, 2002) to machine learning algorithms such as random forests and boosted regression trees (Chee and Elith, 2012; Maloney et al., 2013).

Environmental conditions used as explanatory variables in fish SDMs include topography (e.g., elevation; Bond et al., 2011), soil and geology (e.g., Maloney et al., 2012), climatic variability (e.g., temperature, precipitation; Creque et al., 2005; McNyset, 2005), hydrological regime (e.g., flow velocity, discharge; Bond et al., 2011; McNyset, 2005), and habitat quality (e.g., substrates, hiding spots; Creque et al., 2005). Anthropogenic impacts due to land use (e.g., agriculture, urban areas), impaired river morphology, and (to a lesser extent) habitat fragmentation due to barriers (e.g., distance to barriers such as dams) (Radinger et al., 2017, 2019; Rolls et al., 2014) have been incorporated into fish SDMs to quantify the effect of explanatory variables at multiple spatial scales (Chee and Elith, 2012; Peterson et al., 2011). The importance of including environmental conditions that have a direct, mechanistic effect on species distributions has been emphasized as a means to improve model interpretability and to improve understanding of the mechanisms underlying species distributions (Austin, 2002). This principle is difficult to implement in practice due to limited data availability, but is especially important when the models are intended to inform stream management.

Moreover, the presence-absence and abundance (i.e., count) observations of fish species distributions often exhibit statistical properties that pose additional challenges when applying classical statistical SDMs. Studies applying SDMs in wider ecological contexts often assume species presence-absence or count observations to follow a binomial or Poisson distribution, respectively. However, count distributions of fish may exhibit overdispersion (i.e., the variance is greater than the mean) due to a patchy distribution of individuals. The Poisson distribution assumes that its mean and variance are equal, which yields biased parameter estimates when used to model overdispersed counts (Martin et al., 2005; Zuur et al., 2009). Typically, the negative binomial distribution is used to account for overdispersion (e.g., Warton, 2005), but may still be inadequate for explaining overdispersed counts with excess zeroes (Potts and Elith, 2006).

To model counts of species exhibiting overdispersion and many zeroes, zero-inflated models[1] (ZIMs) have been proposed that effectively combine a binary outcome model (i.e., with a Bernoulli distribution) that predicts excess zeroes with a model for count data (i.e., following a Poisson or negative binomial distribution) (Martin et al., 2005; Wenger and Freeman, 2008; Zuur et al., 2009). ZIMs have been applied to marine fish species for some time (e.g., Stefánsson, 1996), often with the aim of better informing fish stock assessments and fisheries management (Cosandey-Godin et al., 2014; Thorson et al., 2016; Walsh and Brodziak, 2015).

More recently, ZIMs have been applied to freshwater fish, with initial studies comparing the performance of classical statistical techniques with their zero-inflated counterparts (Lewin et al., 2010; Vaudor et al., 2011). Increasingly sophisticated applications of ZIMs have been proposed to model freshwater fish distributions, including models that include species observations with multiple size classes (Kanno et al., 2012), account for non-linear responses of species to environmental conditions (Arab et al., 2012), and quantify environmental conditions at multiple spatial scales (Stewart-Koster et al., 2013). Hierarchical model structures have also been proposed to account for spatial autocorrelation among environmental conditions due to the hierarchical structure of the river network (Boone et al., 2012). However, while studies applying ZIMs to marine fish observations have incorporated the observation process (i.e., the equipment used and sampling effort) into their proposed models, there are relatively few studies that propose similar models for freshwater species (Wildhaber et al., 2012).

In this study, our overarching aim is to develop conceptual models of fish autecology and use them to develop statistical models of fish species distributions (i.e., fish SDMs). More concretely, we develop conceptual models to summarize our prior knowledge of dominant natural and anthropogenic environmental conditions that have a direct, mechanistic effect on freshwater fish species. We then propose two distinct statistical models to predict the occurrence of different fish taxa, namely a presence-absence model (PAM), and a zero-inflated model (ZIM) to predict fish density (i.e., fish count per unit area fished), respectively. Although we use the conceptual models to identify environmental conditions that can potentially be included as explanatory variables in the statistical models, we ultimately select explanatory variables among these that provide the best predictive performance for each species. In pursuing the overarching aim of this study, we address the following research questions:

1 To what extent can we predict the occurrence and density of freshwater fish in a region with a large range of environmental conditions, from densely populated areas with intensive agriculture to alpine regions (based on currently available data from different monitoring programs and based on different fish sampling methods)?
2 Which of the available environmental variables are most important for predicting occurrence and density of the most common fish taxa and do they reflect prior knowledge?

In developing the PAMs and ZIMs for multiple fish taxa, we combine several innovations. We include the observation process in the structure of the statistical models to combine observational data collected using different sampling methods (i.e., with qualitative, semi-quantitative, and quantitative fishing). In classical approaches (e.g., Carle and Strub, 1978), the "true" density of fish is estimated from repeated samplings. These estimates are then used to calibrate a model for the "true" density. In our approach, the measurement process is included in the model, which predicts the "true" density as internal state and provides the estimated fish caught as additional model output. This makes it possible to consider all sources of uncertainty during the calibration

[1] Although early ecological applications of zero-inflated models are highlighted, zero-inflated distributions were first proposed by Aitchison (1955) and later applied by Lambert (1992).

process. The same applies for occurrence in the PAM. In addition, we quantify habitat fragmentation (due to natural and anthropogenic barriers in the river network) and influence factors that act at different spatial scales (e.g., reach and accessible area). With this study, we explore what we can learn from model-based analysis of currently available biomonitoring data and identify limitations and gaps to support river management.

2. Material and methods

2.1. Observational data

We combined data from the Swiss National Surface Water Quality Monitoring Program (NAWA) (Kunz et al., 2016) and the Progetto Fiumi program (Brodersen and Seehausen, 2014) (Table 1). Both programs aim to describe fish community structure in Swiss streams and rivers through quantitative, semi-quantitative or qualitative electrofishing. Sites in the Progetto Fiumi program were chosen to include all Swiss drainages and the entire altitudinal gradient, which are habituated by fish (203–2297 m.a.s.l.). During quantitative sampling, block nets were installed at the start and end points of the fished reach and two or three rounds of fishing were done, with all fish caught included in the sample data and thus allowing for population estimates of individual taxa. Semi-quantitative fishing was similar to quantitative fishing efforts, but without the use of block nets and a single round of fishing. During qualitative sampling, one fishing round was done along a reach without the use of nets. In qualitative sampling, the reach was not entirely fished and not all captured fish were included in the data, but all captured taxa were recorded.

In total, 55 species were recorded. As the taxonomic status of several of the most important riverine fish in Central Europe is currently being revised (e.g., Kottelat and Freyhof, 2007; Lucek et al., 2018; Palandačić et al., 2017) and species-specific field records are therefore often unreliable, we aggregated all species at the genus level (see SI section 1.1 for further explanations).

Based on the combined Progetto Fiumi and NAWA data, we derived the abundance (i.e., count) and presence-absence observations of fish taxa in each sample. Due to the difficulty of modeling rare taxa (Guisan et al., 2006; Potts and Elith, 2006; Sor et al., 2017), we selected seven taxa that occur in 10% or more of all samples (Table 2) for model development.

2.2. Conceptual model

Based on literature sources and consultations with biologists on the autecology of the selected fish taxa, we developed conceptual models that show the current knowledge about dominant natural and anthropogenic environmental conditions that drive the distribution of fish taxa throughout their major life stages (see Fig. 1 for *Salmo spp.* and SI section 1.4 for similar conceptual models for additional taxa). The main goal of these conceptual models is to inspire the development of statistical models based on causally linked explanatory variables to the degree possible (Schuwirth et al., 2019). We included environmental conditions and processes in the conceptual models regardless of their data availability (e.g., fish stocking, impacts of angling on fish populations, predation), while acknowledging that numerous additional factors for which there is little or no available data could also be included (e.g., prevalence of parasites and proliferative kidney disease, effects of hydropeaking). However, the lack of data for specific environmental conditions (e.g., water quality variables such as fine sediment loading) does not exclude the possibility of indirectly quantifying their effect on the distribution of fish taxa (e.g., by including agricultural land use indicators as explanatory variables in our statistical models).

Table 1
Number of sites and samples in the combined Progetto Fiumi and NAWA datasets.

Program	Sites	Samples	Months	Year
Progetto Fiumi	249	249	August – November	2013 – 2017
NAWA	69	106	April – November	2012 – 2013, 2015
TOTAL	318	355		

Note: The number of sites and samples included in the model for a taxon depends on the availability of environmental data during model selection, and in turn affects the prevalence of a taxon in the datasets.

Table 2
Prevalence and total abundance of fish genera selected (≥10% prevalence) in all samples of the combined Progetto Fiumi and NAWA datasets.

Latin Name	Common Name	Total abundance	Prevalence (%)
Salmo spp.	Brown trout	20,434	91.8
Cottus spp.	European bullhead	11,451	43.9
Squalius spp.	Chub	12,288	23.7
Barbatula spp.	Stone loach	9,533	20.8
Barbus spp.	Barbel	21,354	18.9
Phoxinus spp.	Minnow	16,965	18.6
Gobio spp.	Gudgeon	1,976	12.1

2.3. Model definition

In this section, we introduce two statistical models: first, a presence/absence model that takes observation errors into account, and second a model that predicts the fish density (i.e., the number of individuals per square meter) at a site but is calibrated on the observed counts (i.e., abundance) of a given fish taxon. Both models include an observation process (based on the fishing method) in the model structure. For all model definitions the following indices are used:

$$\text{Sites}: i \in \{1, \ldots, I\}$$

$$\text{Time of sampling at site } i: t_i \in \{1, \ldots, T_i\}$$

$$\text{Explanatory variables}: k \in \{1, \ldots, K\}$$

The time point of sampling t_i at a site is needed to represent the 11% of sites that are repeatedly fished. However, for simplicity we omit this index when defining the presence-absence and zero-inflated models.

2.3.1. Presence-absence model

The PAM consists of two parts: a generalized linear model that predicts the probability of occurrence of a given fish taxon based on environmental conditions, and an observation model that describes the probability of observing the taxon based on its probability of occurrence, the number of fishing rounds, and the probability of catching and correctly identifying an individual that is present. The model structure is visualized as a network in Fig. 2.

The presence or absence of a fish taxon at site i is encoded by the variable Y_i, which equals one if the taxon is present or zero if the taxon is absent. We describe the presence-absence observations with a conventional logistic regression

$$Y_i \sim \text{Bernoulli}(\pi_i) \tag{1}$$

where the probability of occurrence π_i is given by

$$\pi_i = \left(1 + \exp\left(-\alpha - \sum_{k=1}^{K} x_{ik}\beta_k\right)\right)^{-1}, \tag{2}$$

with the intercept α, selected environmental conditions as explanatory variables x_{ik}, and coefficients β_k that quantify the taxon-specific responses to the environmental conditions.

It is a common approach to derive observations of Y_i directly from count data. However, counting fish is prone to errors due to the incomplete sampling of all individuals present and potential misidenti-

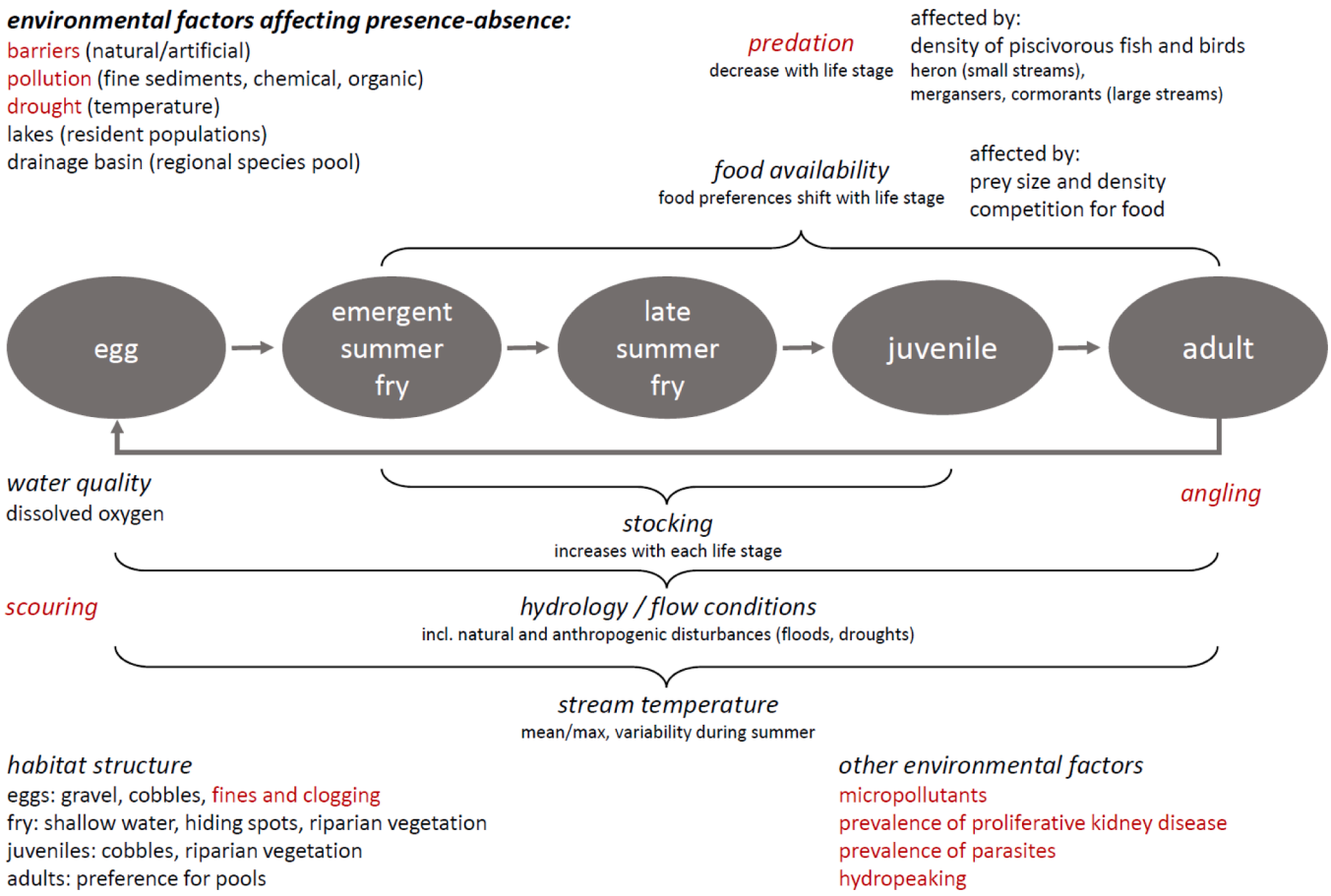

Fig. 1. Conceptual model identifying dominant natural and anthropogenic environmental conditions affecting *Salmo spp*. throughout their life stages based on expert knowledge and Borsuk et al. (2006). Environmental conditions with expected negative effects are shown in red color.

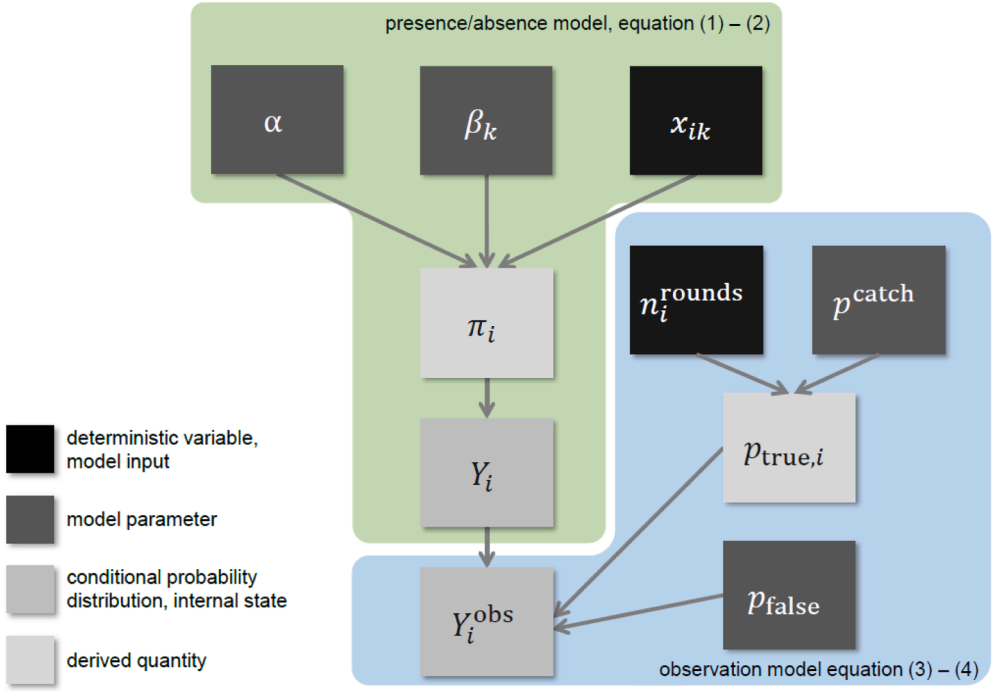

Fig. 2. Network representation of the conditional probability distributions in the presence-absence model. See text for the explanation of the variables.

fication of taxa. To represent this in the model, we make a distinction between the presence-absence observations Y_i^{obs} which are based on the count data and the "true" presence-absence of a taxon Y_i. An observation model links the two variables:

$$P\left(Y_i^{obs}\middle|Y_i\right) \sim \text{Bernoulli}(\omega_i), \tag{3}$$

$$\omega_i = \begin{cases} p_{true,i}, & \text{if } Y_i = 1 \\ p_{false,i}, & \text{if } Y_i = 0 \end{cases}$$

The probability to catch at least one individual at site i is represented by the true positive probability $p_{true,i}$. This probability is derived from a taxon-specific probability p^{catch} of catching an individual and the number of fishing rounds n_i^{rounds}:

$$\begin{aligned} p_{true,i} &= \sum_N P\left(Y_i^{obs}=1\middle|p^{catch}, n_i^{rounds}, N\right)P(N) \\ &= \sum_N 1-\left(1-p^{catch}\right)^{n_i^{rounds}\cdot N} P(N) \end{aligned} \tag{4}$$

This derivation requires the distribution of the number of fish $P(N)$, which was approximated with the empirical distribution of the count data across all sites.

There is also a (typically low) probability that a taxon is erroneously observed due to misidentification. This is modeled by the false positive probability p_{false}. The two probabilities p^{catch} and p_{false} are based on expert judgment and not inferred from the data.

2.3.2. Fish density model

The ZIM predicts the number of observed fish based on a zero-inflated distribution for the fish density and an observation model that accounts for the fished area, the number of fishing rounds, and the catch probability (Fig. 3).

For a given taxon, the fish density ρ_i (i.e., the number of fish per square meter) at a site i is modeled by two components: i) a linear combination of the environmental explanatory variables x_{ik}^{count} with parameters α^{count} and β_k^{count} (i.e., the count component; Zeileis et al., 2008) and ii) a normally distributed, site-specific random effect $\in_i$. This results in:

$$\rho_i = \exp\left(\alpha^{count} + \sum_{k=1}^{K} x_{ik}^{count}\beta_k^{count} + \in_i\right), \tag{5}$$

with $\in_i \sim N(0,\sigma)$.

The site effect $\in_i$ can be interpreted as a residual term, and would be included to quantify the intrinsic uncertainty of the fish density at each site due to environmental conditions not included in the model or due to biotic interactions. Despite this appealing interpretation, we eventually decided to omit the site effects because of the difficulty of defining appropriate informative priors, which would be needed to avoid identifiability problems during parameter estimation (see Discussion).

2.3.3. Including zero-inflation

With the exception of *Salmo spp.*, other taxa include observations with a large proportion of zero counts (SI Fig. 1) — more than what can be explained by the selected environmental variables for fish density. These excess zeros can arise if the taxon is absent at the time of sampling a site or the limited area fished relative to the spatial and temporal scale of the species movements (Martin et al., 2005). Alternatively, the habitat conditions at a site may be suitable for the ecological preferences of a taxon but is otherwise inaccessible due to habitat fragmentation, including physical barriers within the river network (e.g., dams or weirs). To model these excess zero densities, a Bernoulli distributed random variable is introduced

$$\pi_i^{zero} \sim \text{Bernoulli}\left(\text{logit}\left(\alpha^{zero} + \sum_{k=1}^{K} x_{ik}^{zero}\beta_k^{zero}\right)\right). \tag{6}$$

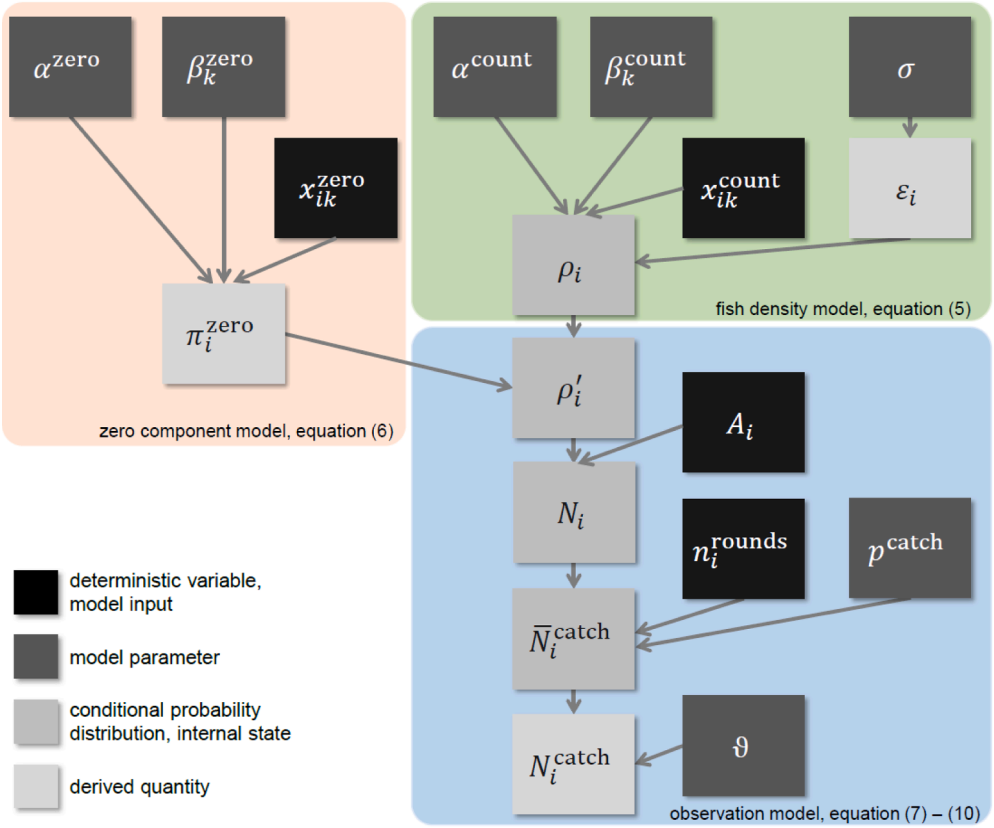

Fig. 3. Network representation of conditional probability distributions in the zero-inflated count model. See text for the explanation of the variables.

If π_i^{zero} equals one, the zero-inflated fish density $\rho_i^{'}$ at a site is zero:

$$\rho_i^{'} = \left(1 - \pi_i^{zero}\right)\rho_i \ . \quad (7)$$

This zero component is similar to a logistic regression applied to presence-absence observations, i.e. a generalized linear model with a logistic link function. Note that although we distinguish between environmental variables for the count and zero components in our model structure, an explanatory variable may be included in both components.

2.3.4. The observation process

To link the fish density with the observations, the model includes an observation process. The mean number of fish N_i that can be potentially caught at a site is the product of the zero-inflated fish density $\rho_i^{'}$ and the fished area A_i (m^2):

$$N_i = \rho_i^{'} A_i \quad (8)$$

However, the number of fish that are actually caught is influenced by the number of sampling rounds n_i^{rounds} and the taxon-specific probability p^{catch} of catching an individual. The expected total number of individuals to catch in a sample is calculated as

$$\overline{N}_i^{catch} = N_i \cdot \left(1 - \left(1 - p^{catch}\right)^{n_i^{rounds}}\right). \quad (9)$$

The randomness of the catch procedure is often modeled with a Poisson or negative binomial distribution. Because we cannot exclude the possibility of overdispersion (Martin et al., 2005; Zuur et al., 2009), the observations are assumed to follow a negative binomial distribution

$$N_i^{catch} \sim NB\left(\overline{N}_i^{catch},\ \vartheta_i\right), \quad (10)$$

with a mean of $\overline{N}_i^{catch}$ and the dispersion parameter ϑ_i. The dispersion parameter ϑ_i is given by $\overline{N}_i^{catch}/(\phi - 1)$, where ϕ is a parameter inferred from the data. Because of the zero-inflation, the modelled distribution of the number of fish caught at each site can be bimodal as illustrated in Fig. 4. At some sites, a qualitative fishing method was applied which provides only presence-absence data. Absence observations correspond to $N_i^{catch} = 0$ and presence observations correspond to $N_i^{catch} > 0$. Hence, qualitative presence observations contribute to the likelihood function with $P(N_i^{catch} > 0) = 1 - P(N_i^{catch} = 0)$.

2.4. Model performance

We evaluated the performance of the presence-absence and zero-inflated model for each taxon by quantifying (a) the quality of fit of each model during calibration using all samples in the data and (b) the predictive performance of each model during k-fold cross-validation. The quality of fit and predictive performance of the models is quantified with the metrics defined below.

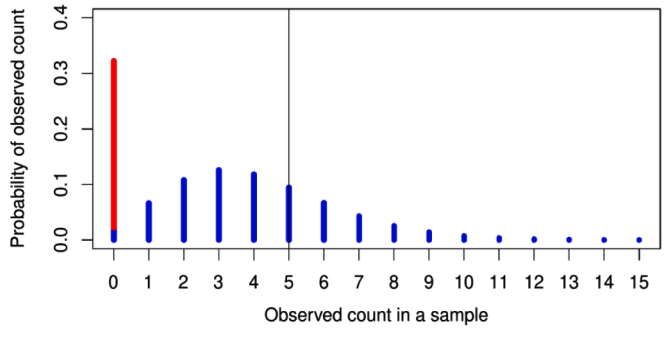

Fig. 4. A synthetic example of a ZIM prediction for the number of fish caught at a sampled site. A ZIM for a taxon predicts the probability of any potential (observed) fish count. The blue lines indicate model predictions arising from the count component for an observed count, the red line indicates the contribution of the zero component to the predicted probability of observing zero individuals, and the black vertical line indicates the actual observed fish count.

2.4.1. Quality of fit

Assuming independent observations, the likelihood of the presence-absence model is

$$L^{PAM} = \prod_{i=1}^{I} P\left(Y_i^{obs} \middle| \alpha, \boldsymbol{\beta}\right),$$

and the likelihood of the zero-inflated model

$$L^{ZIM} = \prod_{i=1}^{I} P\left(N_i^{catch} \middle| \boldsymbol{\Theta}\right) \ ,$$

where $\boldsymbol{\Theta}$ represents all ZIM parameters (α^{zero}, $\boldsymbol{\beta}^{zero}$, α^{count}, $\boldsymbol{\beta}^{count}$, p^{catch}, ϑ).

Because the number of observations I may vary across taxa, we standardize the likelihoods

$$\overline{L} = L^{1/I} \ .$$

The quality of fit of a model is quantified by the ratio of the standardized likelihood of the proposed model (i.e., including the environmental variables and their respective parameters) and the likelihood of the simpler null model (i.e., a model including only the intercepts α^{zero} and α^{count}):

$$L^{ratio} = \frac{\overline{L}^{proposed}}{\overline{L}^{null}}$$

This quantity expresses how much more likely an observation is on average based on the proposed model compared to the null model. Alternatively, the quality of fit of the PAM for each taxon can be quantified using the standardized deviance of the model predictions from the presence-absence observations

$$\overline{d} = -2 * log\overline{L} \quad .$$

The explanatory power of the environmental variables selected as model inputs in the presence-absence model for a taxon can be assessed using the D^2 statistic (similar in interpretation to the R^2 of a linear model assuming a normally distributed response; see Guisan and Zimmermann, 2000). The D^2 is calculated as the fraction of null model (i.e., a model with only an intercept α) deviance that is reduced by the proposed model deviance:

$$D^2 = \frac{\overline{d}^{null} - \overline{d}^{proposed}}{\overline{d}^{null}}$$

We did not use the deviance or D^2 for the ZIMs, because it is not clear how to specify the saturated model (Millar, 2011). Instead, we quantify the quality of fit of the ZIMs using the log-likelihood ($\log L$) of the proposed model and the likelihood ratios L^{ratio}.

2.4.2. Predictive performance

We quantified the predictive performance of the PAM and ZIM for each taxon by k-fold cross-validation. In k-fold cross-validation, we randomly partition the full dataset into k subsets, calibrate the proposed and null PAM or ZIM to every combination of $k - 1$ folds, and obtain model predictions from the independent subsample k. The predictive performance of the model over the independent subsamples is then quantified by the total log-likelihood (of the k testing data sets) divided by the number of observations. We chose $k = 5$ to obtain sufficiently large sample sizes during prediction on the independent data (not used to calibrate the model) that ensure more robust estimates of predictive performance.

In addition, for the PAMs we calculated the average explanatory

power (D^2) over the k folds during calibration and prediction on independent data (in calculating the mean explanatory power during prediction, the null model is calibrated using the independent subsample). The performance of the ZIMs during cross-validation was furthermore quantified with the standardized likelihood ratio L^{ratio} for testing and training data over all folds.

2.5. Variable selection

2.5.1. Preparation of potential variables

Based on the mechanistic conceptual models we developed for each fish taxon, we selected potential variables for which data was available or which we could indirectly derive from existing data. An inspection of the Pearson correlations (r) between the potential variables revealed no strong correlations (i.e., $|r| < 0.7$; see section 1.2 in the SI for the pairwise correlations among potential variable for all samples), with the exception of the stream width and depth variability below (see Table 3).

As potential explanatory variables, we included the habitat conditions (e.g., substrate, flow velocity) at each site (Table 3). In addition, we performed extensive spatial analyses of the river network to quantify the accessible habitat area available from each site at multiple spatial scales (1–10 km at 1 km intervals, however due to strong collinearity we used a maximum distance of 2 km for variable selection) as well as the accessibility of nearby lakes, by taking into account natural and artificial barriers as impassable obstacles in our analyses. We obtained estimates of the maximum morning stream temperature in summer (i.e. annual morning maximum) in each accessible reach from a simple linear model (see Table 3 for details), and calculated the mean within the spatially accessible habitat area. We included a quadratic term ($Temp^2$) to model taxa that prefer intermediate stream temperatures and respond negatively to both low and high temperatures.

Similar to our spatial analysis of stream temperature, we used available data from stream morphology assessments (BAFU, 2006) throughout the river network to quantify the mean morphological state of reaches accessible from the sampled sites. The spatial network analysis was implemented in ESRI's ArcGIS 10.7, with additional post-processing performed using the *sf* package (Pebesma et al., 2020) in the R statistical computing environment ver. 3.6 (R Core Team, 2020).

2.5.2. Selecting variables based on predictive performance

To select a set of environmental variables from Table 3 that maximize the predictive performance of the presence-absence and zero-inflated model for each taxon, we performed an exhaustive search procedure that combines a best subsets regression analysis (James et al., 2013) with 5-fold cross-validation. During this procedure, we constructed models containing all possible combinations of p parameters (while excluding models with Pearson correlations $|r| > 0.7$ from further analysis) and applied 5-fold cross-validation to each model for each taxon. The exhaustive search was applied to PAMs with between 1 and 10 potential variables and to ZIMs with between 1 and 4 potential variables. Testing the predictive performance of the ZIMs was limited to 1–4 potential variables due to the count and zero components in the model structure leading to a large number of combinations of variables in the model that quickly became computationally intractable (despite the use of parallelized processing) with additional potential variables. For the variable selection, we used models that do not explicitly account for the observation process to make use of standard R packages that have a short run-time (see next section) during maximum-likelihood estimation. We identified models with the highest predictive performance based on the total log-likelihood during independent predictions over the k-folds, and verified that parameter estimates of the top three performing models were consistently positive or negative over the five folds.

Table 3
Environmental conditions identified as potential explanatory variables for the presence-absence model and zero-inflated model of each taxon (including the untransformed minimum, mean, and maximum values).

Abbreviations (units)	Description
HabitatArea (km^2)	The total accessible habitat area of reaches within a maximum network distance (2 km) of the site, considering natural and anthropogenic barriers $\geq$ 50 cm high as impassable. Channel widths for accessible lengths of each reach were obtained from morphological assessments or based on the mean channel width of reaches in the same stream order (min: 0, mean: 0.08, max: 0.47 km^2).
Temp (°C)	Maximum morning stream temperature in summer (i.e. annual morning maximum) within the total accessible habitat (see HabitatArea above), predicted from a linear model ($n = 58$) (BAFU, 2016) based on catchment area and mean catchment elevation (swisstopo, 2015), see (Vermeiren et al., 2020) (min: 9.0, mean: 18.1, max: 25.1 °C)
Gravel, Pool (%)	Proportion of fished area consisting of gravel (Gravel, min: 0, mean: 13.2, max: 60); proportion of fished area categorized as pool habitat based on flow regime assessment (Pool, min: 0, mean: 13.1, max: 98%)
FV (m/s)	Mean annual stream flow velocity estimated from channel slope s, channel width W, simulated mean annual discharge Q, and Manning's coefficient n (Cowan, 1956; Pfaundler and Schönenberger, U, 2013; swisstopo, 2016): $FV = (\sqrt{s}/n)^{0.6}(Q/W)^n$ (min: 0.03, mean: 1.05, max: 4.4 m/s)
Farm (%), Urban (%), Forest (%)	Proportions of arable land (Farm, min: 0, mean: 6.4, max: 48.4%), urban and transport-related land use (Urban, min: 0, mean: 5.1, max: 53.2%), and forest cover (Forest, min: 0, mean: 23.7, max: 75.9%) in the catchment area (swisstopo, 2016)
NearLake (km^{-1})	River network distance to the nearest accessible lake within a maximum distance of 25 km, with natural and anthropogenic barriers $\geq$ 50 cm high assumed to be impassable. The data is rescaled as $1/(x+0.25)$ with x being the distance in km (min: 0.04, mean: 0.43, max: 4.0 km^{-1})
LUD (CE/km^2)	Livestock unit density: cattle equivalent (CE) units of livestock per square kilometer of catchment area (BFS, 2008; swisstopo, 2016) (min: 0, mean: 24.3, max: 135.1 CE/km^2)
HidingSpots (%)	Proportion of fished area with overhead cover (Progetto Fiumi) or hiding spots (NAWA) (min: 0, mean: 21.6, max: 100%)
WidthVar, DepthVar (ordinal)	Mean channel width variability (WidthVar, min: 1, mean: 2.2, max: 3) and mean depth variability (DepthVar, min: 1, mean: 2.4, max: 3) (BAFU, 2006) of accessible reaches (see HabitatArea for definition of accessible reaches). For both variables, each reach was classified based on field assessments as 1 = none, 2 = limited, 3 = high.

Note: The environmental variables above do not vary with repeated samplings t_i at site i, due to limitations in data availability. For example, stream temperature had to be estimated from spatial variables, i.e. catchment size and mean catchment elevation. Each variable was centered by their mean and normalized by their standard deviation ($x_k = (x_{ik} - \overline{x}_k)/\sigma_k$) to reduce correlations among the marginal posterior parameter distributions and thereby improve parameter inference, and to maintain parameter estimates relative to the units of the environmental variables. NAWA habitat data for the variables Gravel, Pool, and HidingSpots was combined with equivalent variables in Progetto Fiumi by assuming the ordinal values in NAWA data represent the following proportions of the fished area: none (0%), low (10%), recurrent (30%), recurrent/frequently (45%), frequently (60%).

2.6. Model implementation and parameter inference

Throughout most of the model development process, we used the R statistical computing environment ver. 3.6 (R Core Team, 2020). During variable selection, the parameters α, β_j in the presence-absence models were identified by maximum likelihood estimation with an iterative

Table 4
Prior distribution of model parameters, including their means (μ) and standard deviations (σ).

PAM	**Distribution**	μ	σ
α, β	normal	0	5
p^{catch}	delta	0.65	–
p_{true}	delta	0.01	–
ZIM	**Distribution**	μ	σ
α, β	normal	0	5
p^{catch}	truncated normal [0,1]	0.65	0.05
ϕ	truncated normal [1,∞]	1	2

weighted least squares algorithm in the *glm* function in R and using the null model parameters as starting values. If the iterative weighted least squares algorithm produced parameter estimates with a proposed model deviance greater than the null model deviance, we applied a more robust optimization method with the *optim* function in R to identify the maximum likelihood solution (as in Nelder and Mead 1965). For the zero-inflated models, we used the zero-inflated negative binomial model implemented in the *zeroinfl* function of the *pscl* package (Zeileis et al., 2008) in R. The parameters $\boldsymbol{\Theta}$ were identified by maximum likelihood estimation using the quasi-Newtonian Broyden--Fletcher-Goldfarb-Shanno algorithm implemented in the *optim* function in R (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970).

Based on the results of our variable selection, we implemented the selected PAMs and ZIMs (including the observation process) in Stan (Carpenter et al., 2017) and accessed it through the R package *rstan* (Stan Development Team, 2018). The joint posterior probability distributions of the model parameters were sampled by doing Bayesian inference with an adaptive No-U-Turn Hamiltonian Markov chain Monte Carlo algorithm (Brooks et al., 2011; Duane et al., 1987). The prior distributions of the parameters are provided in Table 4. We used rather wide priors for the parameters α, β, ϕ to reflect for our limited prior knowledge.

3. Results

3.1. Variable selection

The variable selection procedure provided insight into statistical associations between the occurrence and abundance of taxa and environmental variables. During the variable selection procedure of the presence-absence models, for all taxa the top three models in terms of predictive performance each included stream temperature within the accessible habitat (see Table 5). The models for *Cottus spp. and Phoxinus spp.* include a quadratic term that leads to a slight curvature of the response curve within the relevant temperature range covered in the data (see Section 3.3 and SI Fig. 14). Fish taxa exhibited consistently positive responses to additional spatially-explicit variables such as the accessible habitat area (HabitatArea) and the mean width variability within the accessible habitat area (WidthVar), while two taxa exhibited a negative response to the mean depth variability within the accessible habitat area (DepthVar).

Contrary to our expectation based on the conceptual models, three taxa responded negatively to overhead shelter in fished area (HidingSpots) and four taxa responded negatively to the proportion of forest in the catchment. With the exception of *Barbatula spp.*, the taxon-specific responses to agricultural (Farm and LUD) and urban land use were

Table 5
Variables in the three top-ranked presence-absence models based on predictive performance (quantified by the log-likelihood during testing). Individual variables are coloured based on whether the maximum likelihood estimates of the β parameters were positive (blue), negative (red), or inconsistent (black; i.e., positive and negative, depending on the fold) during calibration on the five training datasets. The results for *Salmo spp.* are not shown, because even the top models did not perform better than the null model.

Taxon	Model Rank	Environmental variables	$log\bar{L}_{\text{test}}$
Cottus spp.	1	Temp Temp^2 Farm LUD HabitatArea NearLake	-0.455
Cottus spp.	2	Temp Temp^2 FV Farm LUD HabitatArea NearLake	-0.455
Cottus spp.	3	Temp Temp^2 FV Farm LUD HabitatArea	-0.456
Squalius spp.	1	Temp FV DepthVar Pool Forest HidingSpots	-0.318
Squalius spp.	2	Temp FV DepthVar Pool HidingSpots	-0.318
Squalius spp.	3	Temp FV DepthVar Pool Forest Urban HidingSpots	-0.318
Barbatula spp.	1	Temp Farm HidingSpots HabitatArea	-0.280
Barbatula spp.	2	Temp FV Farm HidingSpots HabitatArea	-0.280
Barbatula spp.	3	Temp FV Pool Farm HidingSpots HabitatArea	-0.280
Barbus spp.	1	Temp WidthVar DepthVar Gravel Pool Forest Farm Urban HidingSpots	-0.299
Barbus spp.	2	Temp WidthVar DepthVar Pool Forest Farm Urban HidingSpots HabitatArea	-0.300
Barbus spp.	3	Temp WidthVar DepthVar Gravel Pool Forest Farm Urban HidingSpots HabitatArea	-0.300
Phoxinus spp.	1	Temp Temp^2 WidthVar DepthVar Forest LUD	-0.254
Phoxinus spp.	2	Temp WidthVar DepthVar Forest LUD	-0.255
Phoxinus spp.	3	Temp Temp^2 DepthVar LUD	-0.255
Gobio spp.	1	Temp Forest	-0.194
Gobio spp.	2	Temp Forest Urban	-0.196
Gobio spp.	3	Temp Urban	-0.196

Table 6
Environmental variables included in the three top-ranked zero-inflated models based on predictive performance during 5-fold cross-validation. Individual variables are colored based on whether the maximum likelihood estimates of the β parameters of the count component were positive (blue), or negative (red) during calibration on each of the folds. Note that negative β parameters in the zero component have a positive effect on the predicted abundance and vice versa, therefore the color coding for the zero component is inverted.

Taxon	Model Rank	Environmental variables (count component | zero component)	$log\bar{L}_{test}$
Salmo spp.	1	$Temp^2$ Urban HidingSpots HabitatArea | -	-4.999
Salmo spp.	2	Urban HidingSpots HabitatArea NearLake | -	-5.003
Salmo spp.	3	Urban HidingSpots HabitatArea | WidthVar	-5.004
Cottus spp.	1	Forest Urban | LUD HabitatArea	-2.478
Cottus spp.	2	Gravel Forest Urban | Temp	-2.491
Cottus spp.	3	Forest | Temp Farm HabitatArea	-2.495
Squalius spp.	1	DepthVar Farm | Temp HidingSpots	-1.565
Squalius spp.	2	DepthVar | Temp Pool HidingSpots	-1.566
Squalius spp.	3	DepthVar Farm HabitatArea | Temp	-1.569
Barbatula spp.	1	Temp | Gravel Farm HabitatArea	-1.410
Barbatula spp.	2	Temp HidingSpots | Farm HabitatArea	-1.419
Barbatula spp.	3	Temp | Gravel Pool HabitatArea	-1.432
Barbus spp.	1	DepthVar Forest | Temp Urban	-1.266
Barbus spp.	2	Temp DepthVar Gravel | HabitatArea	-1.271
Barbus spp.	3	$Temp^2$ | FV DepthVar Forest	-1.273
Phoxinus spp.	1	Urban | Temp FV HidingSpots	-1.283
Phoxinus spp.	2	LUD | Temp FV HidingSpots	-1.284
Phoxinus spp.	3	Farm | Temp FV HidingSpots	-1.284
Gobio spp.	1	- | Temp WidthVar Gravel NearLake	-0.757
Gobio spp.	2	Temp | $Temp^2$ Gravel NearLake	-0.761
Gobio spp.	3	Temp HidingSpots | $Temp^2$ NearLake	-0.763

negative.

The top-performing zero-inflated models of *Salmo spp.* include a quadratic term for temperature and negative responses to accessible habitat areas, while also revealing consistent positive responses to the proportion of urban land use in the catchment and to the proportion of hiding spots in the fished areas (Table 6). In contrast, the ZIM models of several other fish taxa include a positive response of the predicted counts to the accessible habitat area and/or a negative response to overhead shelter (HidingSpots).

Table 7
Performance of the presence-absence model for each taxon, including the quality of fit during model calibration using all data and performance during 5-fold cross-validation based on the average D^2 to illustrate the explanatory power and the likelihood ratio L^{ratio} for the testing data over all folds. The prevalence (Prev) refers to the data set used for calibration with the omission of samples with missing environmental variables and can therefore differ from Table 2.

	calibration				cross-validation	
Taxon	Prev (%)	$\bar{d}$	D^2_{calib}	L^{ratio}_{calib}	D^2_{test}	L^{ratio}_{test}
Salmo spp.	95	0.70	0.02	1.01	0.004	1.00
Cottus spp.	44	0.84	0.37	1.29	0.34	1.26
Squalius spp.	24	0.62	0.43	1.27	0.39	1.24
Barbatula spp.	22	0.56	0.47	1.28	0.44	1.26
Phoxinus spp.	19	0.49	0.49	1.26	0.44	1.24
Barbus spp.	18	0.50	0.47	1.25	0.37	1.19
Gobio spp.	12	0.38	0.47	1.18	0.43	1.18

The results of the selected ZIMs for *Barbus spp.* are not discussed here, due to poor model performance compared to the null-model (see Section 3.2 Model Performance for results and discussion).

3.2. Model performance

The selected presence-absence models show a similarly good quality of fit for all taxa (except for the widespread *Salmo spp.*) due to the high explanatory power of the selected environmental variables in each model (Table 7). Moreover, the performance of the selected models is quite good during cross-validation, with only minor decreases in mean explanatory power D^2 and standardized likelihood ratios L^{ratio} during predictions on out-of-sample testing data.

For the taxa with intermediate prevalence, the explanatory power of the selected environmental variables in the PAM lead to models that predict probabilities of occurrence that closely matched the observations, and were similar during calibration (training) and prediction (testing) of 5-fold cross-validation (Fig. 5).

The geographic distribution of the predicted probabilities of occurrence of the presence-absence models can be used to identify specific sites or regions where the model predictions are consistent with or diverge from the observations (see Fig. 6 for an example for *Cottus spp.* and section 1.6 in the SI for other taxa).

The zero-inflated models generally performed better than the null model for all taxa except *Barbus spp.*, as indicated by the likelihood ratios above one for testing data during cross-validation Table 8.

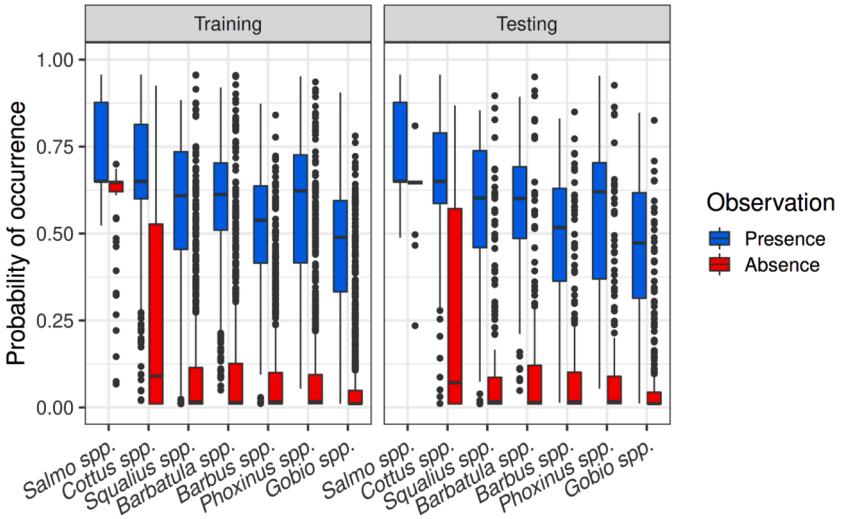

Fig. 5. Predicted probability of occurrence versus observations during 5-fold cross-validation of the presence-absence model for each taxon.

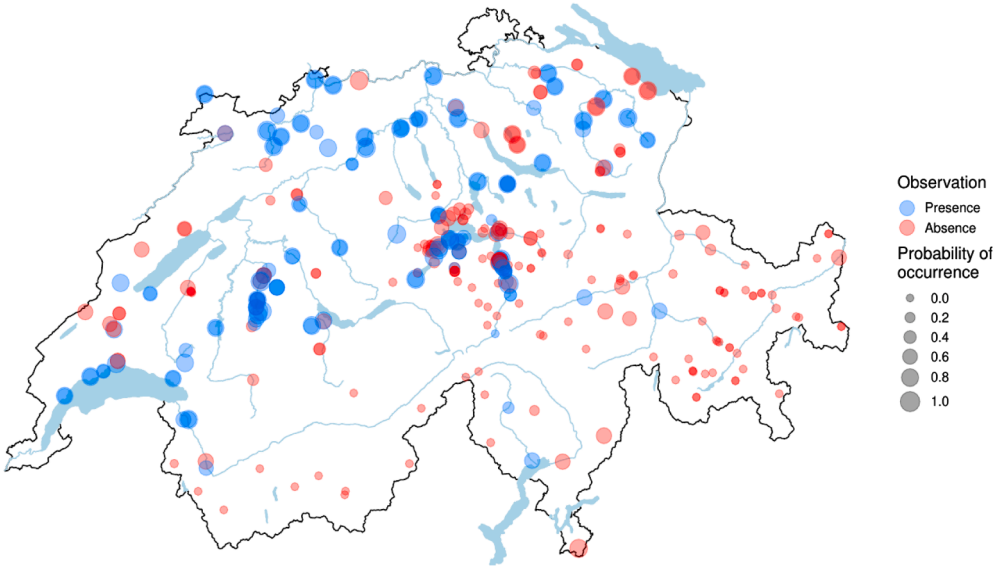

Fig. 6. Geographic distribution of the modeled probability of occurrence and observations in the presence-absence model for *Cottus spp.* Observations are indicated by color and the modelled probability of occurrence by the size of the dots (see legend). A good quality of fit is indicated by large blue dots and small red dots.

Table 8
Performance of the zero-inflated model for density of fish taxa, including the quality of fit during model calibration using all data (with the log-likelihood of the proposed model given as log$L$) and likelihood ratios L^{ratio} during calibration and 5-fold cross-validation. The prevalence refers to the data set used for calibration with the ommission of samples with missing environmental variables and can therefore differ from Table 2. The column I indicates the number of data points.

	calibration to all data			cross-validation	
Taxon	Prev (%)	I	log$L$	$L^{\text{ratio}}_{\text{calib}}$	$L^{\text{ratio}}_{\text{test}}$
Salmo spp.	94	301	−1363	1.64	1.56
Cottus spp.	44	353	−920	1.35	1.27
Squalius spp.	24	271	−488	1.56	1.22
Barbatula spp.	22	303	−512	1.37	1.40
Phoxinus spp.	19	301	−582	1.25	1.24
Barbus spp.	18	321	−608	1.16	0.78
Gobio spp.	15	280	−296	1.19	1.14

3.3. Taxon-specific responses

The maximum posterior parameter estimates during calibration of the PAMs for each taxon show taxon-specific responses during calibration using all samples (Table 9). All taxa responded positively to mean maximum morning summer stream temperature within the total accessible habitat (Temp) in the relevant range covered in the data set (9–25 °C) (see Figure 14 in the SI for the responses of the taxa that include a quadratic term, i.e. *Cottus spp.*, and *Phoxinus spp.*).

The maximum posterior parameters of the ZIM models for all taxa for which the model predicted better than the null model are provided in Table 10. In addition, the marginal posterior parameter distributions of the count and zero components illustrate the uncertainty in the parameter estimates (Fig. 7 for *Salmo spp.* and SI Figures 15–19 for other taxa). For example, for *Salmo spp.* the marginal posterior estimates do not overlap with zero, indicating significant responses to the selected environmental conditions. This taxon has no variables selected for the zero component. The count component shows the strongest response to the accessible habitat area, followed by hiding spots, temperature

Table 9
Maximum posterior parameter estimates of the presence-absence models during calibration using all samples. Positive values (blue) indicate an increase of the probability of occurrence with the corresponding environmental condition and negative values (red) a decrease. Black numbers indicate that the sign was not consistent across the 5-folds during cross-validation. The larger the value the stronger is the response. Results of *Salmo spp.* are not shown, because the model did not predict better than the null model.

	Cottus spp.	*Squalius spp.*	*Barbatula spp.*	*Barbus spp.*	*Phoxinus spp.*	*Gobio spp.*
α_j	0.9	-4.3	-4.2	-5.5	-5.1	-7.5
β Temp	5.4	5.8	4.4	9.2	5.8	6.5
β Temp^2	-0.7	-	-	-	0.6	-
β FV	-	0.2	-	-	-	-
β WidthVar	-	-	-	1.7	0.7	-
β DepthVar	-	0.0	-	-1.1	-0.6	-
β Gravel	-	-	-	-0.3	-	-
β Pool	-	0.9	-	0.6	-	-
β Farm	-1.1	-	0.7	-1.4	-	-
β Forest	-	-0.5	-	-3.0	-0.6	-1.0
β Urban	-	-	-	-2.1	-	-
β LUD	-0.9	-	-	-	-1.0	-
β HidingSpots	-	-1.1	-0.4	-1.5	-	-
β HabitatArea	4.0	-	1.3	-	-	-
β NearLake	-0.2	-	-	-	-	-

squared and urban areas in the catchment. The quadratic temperature term leads to a preference for moderate (mean maximum morning summer) temperatures with an optimum at 18.3 °C (see Figure 14 in the SI).

4. Discussion

In this study, we set out to model and predict the occurrence and density observations of riverine fish using natural and anthropogenic environmental conditions in a region with a high variation in topography. The sites are diverse including alpine streams with mainly hydromorphological alterations and lowland rivers in densely populated areas with intensive agriculture. To achieve this aim, we developed mechanistic conceptual models of fish autecology to summarize dominant environmental variables that could be included in a statistical species distribution model. We then developed a presence-absence model and a zero-inflated count model for each fish taxon, incorporating the observation process (i.e., fishing method) and sampling effort (fished area) into the structure of each model. This enabled us to integrate data from different programs and account for the uncertainty of the observation process during calibration.

4.1. Overall performance

In selecting environmental variables to include in the PAM and ZIM for each taxon based on a simplified model, our results show that we can explain and predict the occurrences of riverine fish with an intermediate prevalence using a presence-absence model rather well. By contrast, the PAM performs poorly for the highly prevalent *Salmo spp.* (prevalence of 95%). The dependence of the PAM performance on prevalence is consistent with previous findings (Sor et al., 2017), demonstrating that it is very difficult to predict absence observations for taxa that occur almost everywhere or presence observations for taxa that are very rare with a statistical model that relies on the information content in the data. For this reason, rare taxa are usually excluded from statistical species distribution models (Sor et al., 2017).

The ZIM predicted reasonably well compared to the null model for most taxa (with a likelihood ratio for testing data between 1.1 and 1.6, indicating that the likelihood is 1.1 to 1.6 times larger for the proposed model than for the null model on average for each data point). *Barbus spp.* was the only taxon with a standardized likelihood ratio below one, indicating that the ZIM model predicts worse than the null model due to overfitting.

The performance of the ZIM may be further improved by additional environmental variables that were unavailable for this study, by increasing the number of data points with non-zero counts for taxa with low prevalence, and/or by extending the variable selection procedure, which was restricted to 1–4 environmental variables due to computational limits (but see next sections).

Initially, we tested the inclusion of random site effects in the ZIM to account for uncertainties in the fish densities in addition to uncertainty by patchiness and the observation process (Fig. 3). However, the absence of prior information about the variance of the site effects and the variance of the negative binomial distribution for counts led to difficulties identifying these two sources of error due to the low information content in the data. We therefore omitted the random site effects from the ZIM.

4.2. Taxon-specific responses in the models

The top-ranked PAMs for all taxa included stream summer morning temperature, with consistently positive parameter estimates during cross-validation. The PAMs for *Squalius spp.*, *Barbus spp.*, *Phoxinus spp.*, and *Gobio spp.* showed clear positive responses to stream temperature that are consistent with existing knowledge of habitat preferences among Cyprinids, while the positive response of *Cottus spp.* in the PAM contrasts with the expected optimum temperature of 10–11 °C (Fig. SI 3). The inferred optimum summer morning temperature for *Salmo spp.* around 18 °C (Fig. SI 14) in the ZIM is roughly in agreement with prior knowledge (Jonsson and Jonsson, 2009). Additional spatially-explicit habitat conditions appeared frequently in the top-ranked models, including the mean channel width variability and depth variability

Table 10
Maximum posterior parameter estimates of the ZIM for all taxa with good model performance during calibration using all samples. Parameter estimates are colored according to positive (blue) or negative (red) effects on predicted counts, see Table 9. Note that positive β parameter values of the zero component increase the probability of zeros and therefore have a negative effect on modelled probability for presence (and vice versa), therefore the color coding is inverted.

	Salmo spp.	*Cottus spp.*	*Squalius spp.*	*Barbatula spp*	*Phoxinus spp.*	*Gobio spp.*
zero component						
α_j^{zero}	-3.9	-0.9	3.0	-0.2	4.2	5.8
β Temp	-	-	-3.6	-	-4.3	-5.6
β FV	-	-	-	-	-0.4	-
β WidthVar	-	-	-	-	-	0.3
β Gravel	-	-	-	-0.04	-	0.2
β Farm	-	-	-	-0.8	-	-
β LUD	-	-1.1	-	-	-	-
β HidingSpots	-	-	0.5	-	0.5	-
β HabitatArea	-	-0.9	-	-0.6	-	-
β NearLake	-	-	-	-	-	-0.9
count component						
α_j^{count}	-3.0	-4.5	-4.0	-7.5	-3.2	-4.9
β Temp	-	-	-	3.4	-	-
β Temp^2	-0.4	-	-	-	-	-
β DepthVar	-	-	-0.6	-	-	-
β Forest	-	1.3	-	-	-	-
β Farm	-	-	0.5	-	-	-
β Urban	0.3	0.5	-	-	0.3	-
β HidingSpots	0.5	-	-	-	-	-
β HabitatArea	-1.0	-	-	-	-	-
additional parameters						
p^{catch}	0.62	0.61	0.57	0.66	0.62	0.64
φ	27	28	26	25	33	19

within the accessible habitat. The negative responses of *Barbus spp.* and *Phoxinus spp.* to depth variability and concurrent positive responses to width variability should be interpreted with care because the correlation between these two variables was close to the threshold of exclusion (Pearson correlation coefficient of 0.7, see section 1.3 in the SI). The unexpected negative responses among many taxa to hiding spots in the fished area in the PAMs suggests that habitats with undercut banks may provide more shelter for piscivorous fish, which can negatively impact e.g. cyprinid fish, through increased predation pressure (e.g. Walser et al., 1999). This is further supported by the positive response of *Salmo spp.* abundance to hiding spots in the ZIM. The total habitat area, which is determined by the presence (or rather absence) of migration barriers, had a positive association with the presence and abundance of *Cottus spp.* and *Barbatula spp.* and a negative association with *Salmo spp.* abundance. Since *Salmo spp.* is subject to stocking in Switzerland (Borsuk et al., 2006), it is plausible that it is more likely to maintain populations in fragmented habitats as compared to other species and that local extinctions of competitor species may lead to higher densities (Holmen et al., 2003; Keeley, 2001). Previous studies also showed that habitat fragmentation can alter fish community structure in streams (e.g. Perkin and Gido, 2012). The models indicated positive associations between urban land use in the catchment and *Salmo spp.*, *Cottus spp.*, *Phoxinus spp.* abundance and negative associations with the presence of *Barbus spp.* Agricultural land use indicators had different effects on taxa presence and abundance in the models, while most taxa showed a negative association with the proportion of forest cover in the catchment. For example, the livestock unit densities had a negative association with *Phoxinus spp.* and opposing effects on *Cottus spp.* in the PAM and the ZIM. The proportion of arable land (Farm) had a negative association with *Cottus spp.* and *Barbus spp.* presence, and positive associations with *Squalius spp.* and *Barbatula spp.*, the latter of which is expected to be tolerant to organic pollution (SI Fig. 5).

Although potential mechanistic pathways that can lead to these responses are known, their relative importance is uncertain. For example, the negative responses of specific fish taxa to arable land use may be attributable to impaired water quality (e.g. organic matter inputs leading to oxygen depletion), stream morphology (e.g., clogging with fine sediments), or altered stream hydrology, while some taxa may profit from higher stream productivity due to nutrient inputs. To disentangle these effects, more detailed water quality data for the fish monitoring sites would be needed.

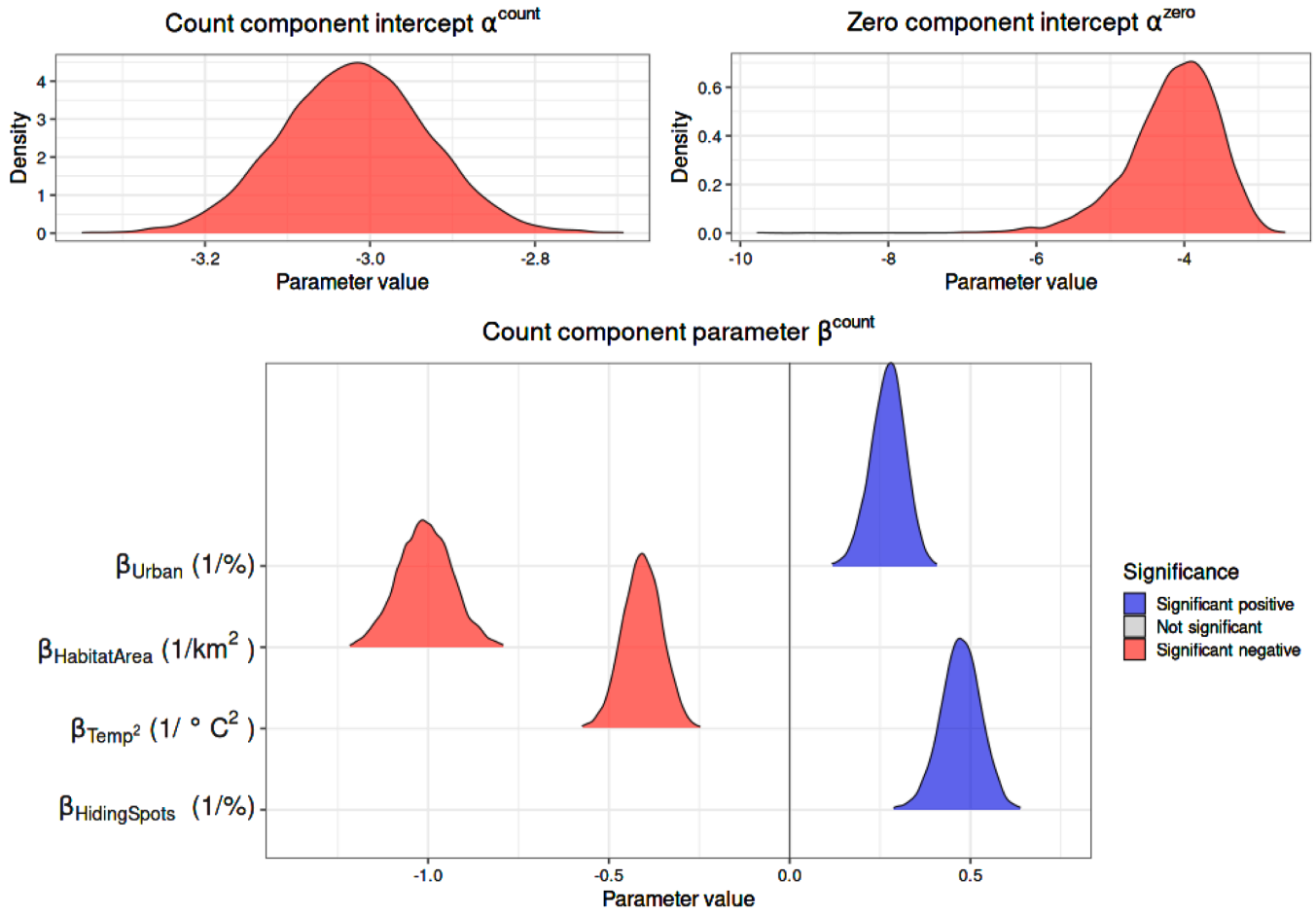

Fig. 7. Marginal posterior parameter distributions for the count and zero components of the zero-inflated model when calibrated using all samples for *Salmo spp.* The width of the marginal posterior parameter distributions illustrates the uncertainty of the parameter estimates.

4.3. Conceptual and statistical models

The discrepancies between the environmental conditions included in our conceptual and statistical models may arise for two major reasons. First, our conceptual models are more mechanistic than our statistical models because (i) we included environmental influence factors in the conceptual models that have a direct, mechanistic effect on the selected taxa and regardless of data availability, and (ii) for specific taxa such as *Salmo spp.*, *Squalius spp.*, *Barbus spp.*, *Barbatula spp.*, the conceptual models are structured according to major fish growth stages. Despite these differences, we acknowledge that our conceptual models may be incomplete or contain a high degree of uncertainty due to limited knowledge of the effects of specific natural environmental influence factors (e.g., prevalence of disease and parasites) and anthropogenic influences (e.g., stocking and angling) on the development and distribution of specific taxa.

Second, not all environmental variables included in the conceptual model could be included in the statistical model due to the limited availability of data (e.g., water quality parameters). We used the conceptual models to identify potential environmental variables, and selected variables to include in the statistical models (i.e., the PAM and ZIM) for each taxon based on the data availability and predictive performance of candidate models. Differences between the conceptual and statistical model (in terms of the selected variables and taxon-specific responses) may indicate that our prior knowledge in the conceptual model is incorrect or that the explanatory variables are not accurate enough, especially those that were estimated from other factors or are only indirectly linked to stressors, such as land use. For the ZIM, we cannot exclude the possibility that the variable selection procedure was not comprehensive enough, because we tested only models with 1–4 explanatory variables, due to computational limitations. However, given that even the relatively simple selected ZIMs led to overfitting for *Barbus spp.*, this is (at least for *Barbus spp.*) unlikely to be the cause of the poor predictive performance.

4.4. Improving the zero-inflated model

Additional analysis of the ZIM is necessary to further improve predictive performance and to identify more intuitive methods of illustrating model performance. The standardized likelihood ratios of the zero-inflated count models for the testing data during cross-validation below one for *Barbus spp.* indicate an overfitting of the model, despite the low number of parameters included, and a failure of the model to predict the counts for independent data points. The skewed distribution of observed counts with a long tailing (see Observed Counts in the SI) already indicates that it will be difficult to predict the few observations with high counts with a rather simple statistical model based on environmental conditions, especially if the number of available data points with observations above zero are low. Increasing the sample size of the observational data to increase the information content relating fish densities to environmental conditions may improve model performance. In the future, it might be worth trying to model biomass instead of counts. Fish biomass can be expected to have a more even distribution, because observations of high numbers can be caused by a large number of juveniles with low biomass.

We found that many studies applying ZIMs to fish use information-

theoretic statistics (such as BIC and DIC) to evaluate and select a model (e.g., Arab et al., 2012). Instead of penalizing the likelihood of a candidate model based on increasing model complexity (i.e., adding environmental variables to the model), we opted to quantify the performance of the ZIMs using the likelihood during calibration using all samples and 5-fold cross-validation. While the standardized likelihood ratio indicates how much more likely the observations are in the proposed model relative to the null model, it would be valuable to also visualize the results by comparing observations and model predictions. However, since the zero-inflated densities can have a bi-modal distribution (see Fig. 4), it is not straightforward to visualize and summarize model predictions (e.g. based on quantiles) (but see SI Fig. 20 for an attempt to visualize the distributions of the predicted vs. observed counts).

In conclusion, the use of conceptual models for the presence/absence and densities of fish taxa contributed to the pre-selection of environmental variables and the development of statistical models that reflect plausible cause-and-effect relationships. We were able to develop presence-absence models for taxa with intermediate prevalence and a ZIM for the most prevalent taxa that have a reasonable predictive performance, incorporating spatially-explicit environmental variables and making use of monitoring data with different sampling methods and levels of sampling effort to account for the observation process.

We have shown that statistical models can indicate potential positive or negative effects of environmental factors on the occurrence and abundance of common riverine fish taxa. Many of these factors are currently subject to management actions or influenced by global changes. Examples for such management actions are the removal of barriers to fish migration (O'Hanley et al., 2013), the morphological restoration of rivers (Haase et al., 2013), and changes in agricultural practices to reduce pesticide and nutrient inputs (Acero Triana et al., 2021). The models developed in this study are far from able to accurately predict the effect of a specific management action on a local fish community, which would additionally require considerations of the colonization potential, biotic interactions and less common taxa. Still, this study may help to adjust expectations towards management actions under changing environmental conditions. For example, according to our results, increasing stream temperatures can be expected to increase the occurrence of most common fish species, while the removal of barriers and increased morphological variability could lead to a species turn over. With this attempt for "mechanistically inspired" statistical models, we hope to encourage the use of increasingly available fish biomonitoring data to learn about fish responses to multiple stressors in a changing environment.

Credit author statement

BC: Bogdan Caradima
AS: Andreas Scheidegger
JB: Jakob Brodersen
NS: Nele Schuwirth

BC lead the data preparation, statistical analysis, and writing. AS contributed to the statistical methodology, model development, and model concept diagrams. JB contributed to data documentation and curation, fish taxonomy, and to the writing with biological interpretations of results. NS contributed to the writing, statistical analysis, and study methodology and design.

Declaration of Competing Interest

The authors have no conflicts of interest to declare in this study.

Acknowledgments

We thank the Swiss Federal Office the Environment for funding, contributing the NAWA monitoring data, stimulating discussions during the project and support, especially Bänz Lundsgaard Hansen, Yael Schindler Wildhaber and Gregor Thomas. We acknowledge Lena Ninck from Fischwerk for answering questions about the NAWA data, and Johannes Hellmann and his team for collecting and managing the Progetto Fiumi data. Furthermore, we thank Rosi Siber for GIS support on environmental data, and Karin Ghilardi and Ruth Scheidegger for support with data preparation. We thank Peter Reichert, Jukka Jokela and Florian Hartig for stimulating discussions.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ecolmodel.2021.109680.

References

Acero Triana, J.S., Chu, M.L., Stein, J.A, 2021. Assessing the impacts of agricultural conservation practices on freshwater biodiversity under changing climate. Ecol. Modell. 453, 109604 https://doi.org/10.1016/j.ecolmodel.2021.109604.

Aitchison, J., 1955. On the distribution of a positive random variable having a discrete probability mass at the origin*. J. Am. Statist. Assoc. 50 (271), 901–908. https://doi.org/10.1080/01621459.1955.10501976.

Arab, A., Holan, S.H., Wikle, C.K., Wildhaber, M.L., 2012. Semiparametric bivariate zero-inflated Poisson models with application to studies of abundance for multiple species. Environmetrics 23 (2), 183–196. https://doi.org/10.1002/env.1142.

Austin, M.P., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecol. Modell. 157 (2–3), 101–118. https://doi.org/10.1016/S0304-3800(02)00205-3.

BAFU, 2006. Ökomorphologie Stufe S (systembezogen): Methoden zur Untersuchung und Beurteilung der Fliessgewässer Gemäss Dem Modul-Stufen-Konzept. Bundesamt für Umwelt BAFU. http://www.modul-stufen-konzept.ch/download/oekom_stufe_s_d.pdf.

BAFU, 2016. Auswirkungen Des Hitzesommers 2003 Auf Die Gewässer (SRU-369-D). Bundesamt für Umwelt BAFU, Bern. https://www.bafu.admin.ch/bafu/en/home/topics/water/state/water–monitoring-networks/monitoring-network–water-temperature.html.

BFS, 2008. Landwirtschaftliche Betriebszählung /Census of Farming. Bundesamt für Statistik BFS.

Bond, N., Thomson, J., Reich, P., Stein, J., 2011. Using species distribution models to infer potential climate change-induced range shifts of freshwater fish in south-eastern Australia. Mar. Freshwater Res. 62 (9), 1043–1061.

Boone, E.L., Stewart-Koster, B., Kennard, M.J., 2012. A hierarchical zero-inflated Poisson regression model for stream fish distribution and abundance. Environmetrics 23 (3), 207–218. https://doi.org/10.1002/env.1145.

Borsuk, M.E., Reichert, P., Peter, A., Schager, E., Burkhardt-Holm, P., 2006. Assessing the decline of brown trout (Salmo trutta) in Swiss rivers using a Bayesian probability network. Ecol. Modell. 192 (1), 224–244. https://doi.org/10.1016/j.ecolmodel.2005.07.006.

Brodersen, J., Seehausen, O., 2014. Why evolutionary biologists should get seriously involved in ecological monitoring and applied biodiversity assessment programs. Evol. Appl. 7 (9), 968–983. https://doi.org/10.1111/eva.12215.

Brooks, E.G.E., Freyhof, J., 2011. European Red List of Freshwater Fishes. Publications Office of the European Communities. https://doi.org/10.2779/85903.

Brooks, S., Gelman, A., Jones, G., Meng, X.-.L., 2011. Handbook of Markov Chain Monte Carlo. CRC Press.

Broyden, C.G., 1970. The convergence of a class of double-rank minimization algorithms2. The new algorithm. IMA J. Appl. Math. 6 (3), 222–231. https://doi.org/10.1093/imamat/6.3.222.

Carle, F.L., Strub, M.R., 1978. A new method for estimating population size from removal data. Biometrics 34 (4), 621–630. https://doi.org/10.2307/2530381. JSTOR.

Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A., 2017. Stan: a Probabilistic Programming Language. J. Stat. Softw. 76 (1), 1–32. https://doi.org/10.18637/jss.v076.i01.

Chee, Y.E., Elith, J., 2012. Spatial data for modelling and management of freshwater ecosystems. Int. J. Geogr. Inf. Sci. 26 (11), 2123–2140. https://doi.org/10.1080/13658816.2012.717628.

Cosandey-Godin, A., Krainski, E.T., Worm, B., Flemming, J.M., 2014. Applying Bayesian spatiotemporal models to fisheries bycatch in the Canadian Arctic. Can. J. Fish. Aquat.Sci. 72 (2), 186–197. https://doi.org/10.1139/cjfas-2014-0159.

Cowan, W., 1956. Estimating hydraulic roughness coefficients. Agric. Eng. 473–475.

Creque, S.M., Rutherford, E.S., Zorn, T.G., 2005. Use of GIS-derived landscape-scale habitat features to explain spatial patterns of fish density in Michigan rivers. North Am. J. Fisheries Manag. 25 (4), 1411–1425. https://doi.org/10.1577/M04-121.1.

Duane, S., Kennedy, A.D., Pendleton, B.J., Roweth, D., 1987. Hybrid Monte Carlo. Phys. Lett. B 195 (2), 216–222. https://doi.org/10.1016/0370-2693(87)91197-X.

Dudgeon, D., Arthington, A.H., Gessner, M.O., Kawabata, Z.-.I., Knowler, D.J., Lévêque, C., Naiman, R.J., Prieur-Richard, A.-.H., Soto, D., Stiassny, M.L.J., Sullivan, C.A., 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. Biol. Rev. 81 (2), 163–182. https://doi.org/10.1017/S1464793105006950.

Fletcher, R., 1970. A new approach to variable metric algorithms. Comput. J. 13 (3), 317–322.
Fukushima, M., Kameyama, S., Kaneko, M., Nakao, K., Ashley Steel, E., 2007. Modelling the effects of dams on freshwater fish distributions in Hokkaido. Japan. Freshwater Biol. 52 (8), 1511–1524. https://doi.org/10.1111/j.1365-2427.2007.01783.x.
Goldfarb, D., 1970. A family of variable metric updates derived by variational means. Math. Comput. 24 (109), 23–26.
Gozlan, R.E., Karimov, B.K., Zadereev, E., Kuznetsova, D., Brucet, S., 2019. Status, trends, and future dynamics of freshwater ecosystems in Europe and Central Asia. Inland Waters 9 (1), 78–94. https://doi.org/10.1080/20442041.2018.1510271.
Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N.G., Lehmann, A., Zimmermann, N.E., 2006. Using niche-based models to improve the sampling of rare species. Conserv. Biol. 20 (2), 501–511. https://doi.org/10.1111/j.1523-1739.2006.00354.x.
Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Modell. 135 (2–3), 147–186. https://doi.org/10.1016/S0304-3800 (00)00354-9.
Haase, P., Hering, D., Jähnig, S.C., Lorenz, A.W., Sundermann, A., 2013. The impact of hydromorphological restoration on river ecological status: a comparison of fish, benthic invertebrates, and macrophytes. Hydrobiologia 704 (1), 475–488. https://doi.org/10.1007/s10750-012-1255-1.
Holmen, J., Olsen, E.M., Vøllestad, L.A., 2003. Interspecific competition between stream-dwelling brown trout and Alpine bullhead. J. Fish Biol. 62 (6), 1312–1325. https://doi.org/10.1046/j.1095-8649.2003.00112.x.
James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning with Applications in R (Vol. 112). Springer.
Jonsson, B., Jonsson, N., 2009. A review of the likely effects of climate change on anadromous Atlantic salmon Salmo salar and brown trout Salmo trutta, with particular reference to water temperature and flow. J. Fish Biol. 75 (10), 2381–2447. https://doi.org/10.1111/j.1095-8649.2009.02380.x.
Kanno, Y., Vokoun, J.C., Holsinger, K.E., Letcher, B.H., 2012. Estimating size-specific brook trout abundance in continuously sampled headwater streams using Bayesian mixed models with zero inflation and overdispersion. Ecol. Freshw. Fish 21 (3), 404–419. https://doi.org/10.1111/j.1600-0633.2012.00560.x.
Keeley, E.R., 2001. Demographic responses to food and space competition by juvenile steelhead trout. Ecology 82 (5), 1247–1259.doi:10.1890/0012-9658 (2001)082 [1247:DRTFAS]2.0.CO;2
Kottelat, M., Freyhof, J., 2007. Handbook of European freshwater Fishes. Publications Kottelat. http://agris.fao.org/agris-search/search.do?recordID=US201300126031.
Kunz, M., Wildhaber, Y.S., Dietzel, A., Wittmer, I., Leib, V., 2016. Ergebnisse Der Nationalen Beobachtung Oberflächengewässerqualität (NAWA) 2011–2014 (Umwelt-Zustand Nr. *1620: 87 S)*. Bundesamt für Umwelt BAFU, Bern. www.bafu.admin.ch/uz-1620-d.
Lambert, D., 1992. Zero-inflated poisson regression, with an application to defects in manufacturing. Technometrics 34 (1), 1–14. https://doi.org/10.1080/00401706.1992.10485228.
Lewin, W.-C., Freyhof, J., Huckstorf, V., Mehner, T., Wolter, C., 2010. When no catches matter: coping with zeros in environmental assessments. Ecol. Indic. 10 (3), 572–583. https://doi.org/10.1016/j.ecolind.2009.09.006.
Lucek, K., Keller, I., Nolte, A.W., Seehausen, O., 2018. Distinct colonization waves underlie the diversification of the freshwater sculpin (Cottus gobio) in the Central European Alpine region. J. Evol. Biol. 31 (9), 1254–1267. https://doi.org/10.1111/jeb.13339.
Maloney, K.O., Schmid, M., Weller, D.E., 2012. Applying additive modelling and gradient boosting to assess the effects of watershed and reach characteristics on riverine assemblages. Methods Ecol. Evol. 3 (1), 116–128. https://doi.org/10.1111/j.2041-210X.2011.00124.x.
Maloney, K.O., Weller, D.E., Michaelson, D.E., Ciccotto, P.J., 2013. Species distribution models of freshwater stream fishes in Maryland and their implications for management. Environ. Model. Assess. 18 (1), 1–12. https://doi.org/10.1007/s10666-012-9325-3.
Martin, T.G., Wintle, B.A., Rhodes, J.R., Kuhnert, P.M., Field, S.A., Low-Choy, S.J., Tyre, A.J., Possingham, H.P., 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. Ecol. Lett. 8 (11), 1235–1246. https://doi.org/10.1111/j.1461-0248.2005.00826.x.
McNyset, K.M., 2005. Use of ecological niche modelling to predict distributions of freshwater fish species in Kansas. Ecol. Freshw. Fish 14 (3), 243–255. https://doi.org/10.1111/j.1600-0633.2005.00101.x.
Millar, R., 2011. Maximum Likelihood Estimation and Inference: With Examples in R, SAS and ADMB. John Wiley & Sons, Ltd.
Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. Comput J. 7 (4), 308–313. https://doi.org/10.1093/comjnl/7.4.308.
O'Hanley, J.R., Wright, J., Diebel, M., Fedora, M.A., Soucy, C.L., 2013. Restoring stream habitat connectivity: a proposed method for prioritizing the removal of resident fish passage barriers. J. Environ. Manage. 125, 19–27.
Olden, J.D., Jackson, D.A., 2002. A comparison of statistical approaches for modelling fish species distributions. Freshw. Biol. 47 (10), 1976–1995. https://doi.org/10.1046/j.1365-2427.2002.00945.x.
Palandačić, A., Naseka, A., Ramler, D., Ahnelt, H., 2017. Contrasting morphology with molecular data: an approach to revision of species complexes based on the example of European Phoxinus (Cyprinidae). BMC Evol. Biol. 17 (1), 184. https://doi.org/10.1186/s12862-017-1032-x.
Pebesma, E., Bivand, R., Racine, E., Sumner, M., Cook, I., Keitt, T., Lovelace, R., Wickham, H., Ooms, J., Müller, K., Pedersen, T.L., & Baston, D. (2020). sf: simple features for R (0.9-3) [computer software]. https://CRAN.R-project.org/package=sf.
Perkin, J.S., Gido, K.B., 2012. Fragmentation alters stream fish community structure in dendritic ecological networks. Ecol. Appl. 22 (8), 2176–2187.
Peterson, E.E., Sheldon, F., Darnell, R., Bunn, S.E., Harch, B.D., 2011. A comparison of spatially explicit landscape representation methods and their relationship to stream condition: spatially explicit landscape representation methods. Freshw. Biol. 56 (3), 590–610. https://doi.org/10.1111/j.1365-2427.2010.02507.x.
Pfaundler, M., Schönenberger, U., 2013. Datensatz MQ-GWN-CH, Modellierte Mittlere Natürliche Abflüsse für das Gewässernetz der Schweiz. Bundesamt für Umwelt BAFU.
Potts, J.M., Elith, J., 2006. Comparing species abundance models. Ecol. Modell. 199 (2), 153–163. https://doi.org/10.1016/j.ecolmodel.2006.05.025.
R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org/.
Radinger, J., Alcaraz-Hernández, J.D., García-Berthou, E., 2019. Environmental filtering governs the spatial distribution of alien fishes in a large, human-impacted Mediterranean river. Divers. Distrib. 0 (0) https://doi.org/10.1111/ddi.12895.
Radinger, J., Essl, F., Hölker, F., Horký, P., Slavík, O., Wolter, C., 2017. The future distribution of river fish: the complex interplay of climate and land use changes, species dispersal and movement barriers. Global Change Biol. 23 (11), 4970–4986. https://doi.org/10.1111/gcb.13760.
Rolls, R.J., Stewart-Koster, B., Ellison, T., Faggotter, S., Roberts, D.T., 2014. Multiple factors determine the effect of anthropogenic barriers to connectivity on riverine fish. Biodivers. Conserv. 23 (9), 2201–2220. https://doi.org/10.1007/s10531-014-0715-5.
Schuwirth, N., Borgwardt, F., Domisch, S., Friedrichs, M., Kattwinkel, M., Kneis, D., Kuemmerlen, M., Langhans, S.D., Martínez-López, J., Vermeiren, P., 2019. How to make ecological models useful for environmental management. Ecol. Modell. 411, 108784 https://doi.org/10.1016/j.ecolmodel.2019.108784.
Shanno, D.F., 1970. Conditioning of quasi-Newton methods for function minimization. Math. Comput. 24 (111), 647–656. https://doi.org/10.1090/S0025-5718-1970-0274029-X.
Sor, R., Park, Y.-.S., Boets, P., Goethals, P.L.M., Lek, S., 2017. Effects of species prevalence on the performance of predictive models. Ecol. Modell. 354, 11–19. https://doi.org/10.1016/j.ecolmodel.2017.03.006.
Stan Development Team. (2018). RStan: the R interface to Stan. https://mc-stan.org/users/interfaces/rstan.
Stefánsson, G., 1996. Analysis of groundfish survey abundance data: combining the GLM and delta approaches. ICES J. Mar. Sci. 53 (3), 577–588. https://doi.org/10.1006/jmsc.1996.0079.
Stewart-Koster, B., Boone, E.L., Kennard, M.J., Sheldon, F., Bunn, S.E., Olden, J.D., 2013. Incorporating ecological principles into statistical models for the prediction of species' distribution and abundance. Ecography 36 (3), 342–353. https://doi.org/10.1111/j.1600-0587.2012.07764.x.
swisstopo, 2015. SwissALTI3D (Art. 30 Geo IV). swisstopo. https://shop.swisstopo.admin.ch/en/products/height_models/alti3D.
swisstopo, 2016. SwissTLM3D (Art. 30 Geo IV). swisstopo. https://shop.swisstopo.admin.ch/en/products/landscape/tlm3D.
Thorson, J.T., Ianelli, J.N., Larsen, E.A., Ries, L., Scheuerell, M.D., Szuwalski, C., Zipkin, E.F., 2016. Joint dynamic species distribution models: a tool for community ordination and spatio-temporal monitoring: joint dynamic species distribution models. Global Ecol. Biogeogr. 25 (9), 1144–1158. https://doi.org/10.1111/geb.12464.
Vaudor, L., Lamouroux, N., Olivier, J.-.M., 2011. Comparing distribution models for small samples of overdispersed counts of freshwater fish. Acta Oecol. 37 (3), 170–178. https://doi.org/10.1016/j.actao.2011.01.010.
Vermeiren, P., Reichert, P., Schuwirth, N., 2020. Integrating uncertain prior knowledge regarding ecological preferences into multi-species distribution models: effects of model complexity on predictive performance. Ecol. Modell. 420, 108956 https://doi.org/10.1016/j.ecolmodel.2020.108956.
Vörösmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S.E., Sullivan, C.A., Liermann, C.R., Davies, P.M., 2010. Global threats to human water security and river biodiversity. Nature 467 (7315), 555–561. https://doi.org/10.1038/nature09440.
Walser, C.A., Belk, M.C., Shiozawa, D.K., 1999. Habitat use of leatherside chub (Gila Copei) in the presence of predatory brown trout (Salmo trutta). Great Basin Nat. 59 (3), 272–277.
Walsh, W.A., Brodziak, J., 2015. Billfish CPUE standardization in the Hawaii longline fishery: model selection and multimodel inference. Fish. Res. 166, 151–162. https://doi.org/10.1016/j.fishres.2014.07.015.
Warton, D.I., 2005. Many zeros does not mean zero inflation: comparing the goodness-of-fit of parametric models to multivariate abundance data. Environmetrics 16 (3), 275–289. https://doi.org/10.1002/env.702.
Wenger, S.J., Freeman, M.C., 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.
Wildhaber, M.L., Gladish, D.W., Arab, A., 2012. Distribution and habitat use of the Missouri river and lower Yellowstone river benthic fishes from 1996 to 1998: a baseline for fish community recovery. River Res. Appl. 28 (10), 1780–1803. https://doi.org/10.1002/rra.1559.
Zeileis, A., Kleiber, C., Jackman, S., 2008. Regression models for count data in R. J. Stat. Softw. 27 (1), 1–25. https://doi.org/10.18637/jss.v027.i08.
Zuur, A., Ieno, E.N., Walker, N., Saveliev, A.A., Smith, G.M., 2009. Mixed Effects Models and Extensions in Ecology with R. Springer Science & Business Media.