Effects of site selection and taxonomic resolution on the inference of stream invertebrate responses to environmental conditions

Bogdan Caradima[1,2,3], Peter Reichert[1,2,4], Nele Schuwirth[1,2,5]

[1]Eawag: Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, 8600, Dübendorf, Switzerland

[2] Eidgenössische Technische Hochschule Zurich, Institute of Biogeochemistry and Pollutant Dynamics, Universitätstrasse 16, 8092 Zürich, Switzerland

E-mail addresses: [3]bogdan.caradima@eawag.ch; [4]peter.reichert@eawag.ch; [5]nele.schuwirth@eawag.ch

Abstract: Key decisions in the design of biomonitoring programs include taxonomic resolution, geographic extent, and site selection, each of which can affect our ability to infer human impacts on biodiversity from biomonitoring data. These decisions are constrained by monitoring goals and budget limitations, which may require trade-offs between them. In this study, we use species distribution models (SDMs) to assess the effects of key decisions in biomonitoring design on our ability to infer the effects of natural and anthropogenic environmental conditions on the occurrence of benthic macroinvertebrates in streams. We compared four datasets that differ in their site-selection strategy, geographic extent, and taxonomic resolution using data from Swiss federal and cantonal biomonitoring programs. We used individual SDMs with 3-fold cross-validation to identify the environmental variables that best predict the probability of occurrence for taxa across the datasets. We then used a hierarchical multi-species distribution model (hmSDM) to identify how key aspects of biomonitoring design influence the relative importance of the selected explanatory (predictor) variables in the model as well as the model's predictive performance. The relative importance of the explanatory variables in the hmSDMs was lowest for the dataset with a grid-based site selection approach and family-level resolution. An increase in predictive performance was achieved by either using a species-level taxonomic resolution for Ephemeroptera, Plecoptera, and Trichoptera or by combining different biomonitoring programs at the family level to increase the number of sites and improve the coverage of environmental conditions. Selecting monitoring sites to provide a good coverage of environmental conditions, while also targeting sites with rare combinations of environmental conditions, could further improve biomonitoring program data. Models based on finer taxonomic resolution revealed that widespread families consist of species and genera with different and stronger responses to environmental conditions. However, many families include species that are too rare to allow

2

inference of significant responses to environmental conditions. We show that hmSDMs of stream invertebrate communities can contribute to the selection of specific taxa for identification at a finer taxonomic resolution. This strategy could facilitate the standardization and combination of multiple biomonitoring datasets and improve the identifiability of stream invertebrate responses to environmental conditions in biomonitoring programs.

Key words: benthic macroinvertebrates, indicator species, stream biomonitoring, biomonitoring design, multiple stressors, hierarchical species distribution models

1 Introduction

Since the Industrial Revolution, human demands on freshwater resources have increased dramatically on a global scale, leading to the widespread modification, fragmentation, and destruction of freshwater habitats and ecosystems (Strayer and Dudgeon 2010). In response, many countries have established long-term freshwater biomonitoring programs of algae, fish, and benthic macroinvertebrate assemblages to provide biological assessments of stream ecosystem health. Benthic macroinvertebrates have proven to be particularly suitable for biomonitoring as they are widespread, easy to collect, and have a relatively high taxonomic and functional diversity, with taxa exhibiting different preferences for habitat conditions and different responses to multiple stressors. Previous studies have identified macroinvertebrate responses to natural and anthropogenic environmental conditions (Townsend et al. 2008, Leps et al. 2015), including hydro-morphology (e.g., modified river-bed substrates or channelization; see Feld et al. 2014), hydrological conditions (e.g., flow velocity, temperature, conductivity), impaired water quality (e.g., high nutrient and fine sediment concentrations; see Wagenhoff et al. 2011, Piggott et al. 2015, Elbrecht et al. 2016), and other habitat characteristics under human management (e.g., riparian vegetation). Pollution from urban areas (Walsh and Webb 2016), wastewater treatment plants (Bunzel et al. 2013, Eggen et al. 2014, Stamm et al. 2016), and agricultural land use (including nutrients, fine sediment, and insecticides; see Liess et al. 2012, Liess et al. 2012, Beketov et al. 2013, Bunzel et al. 2014) affect habitat quality and invertebrate community composition at varying spatial scales (Sponseller et al. 2001, Domisch et al. 2015).

The ability to quantify stream invertebrate responses to natural and anthropogenic environmental conditions largely depends on several key decisions during the design of a stream biomonitoring program (Cao and Hawkins 2011). These decisions must consider monitoring objectives including, among others, 1) selecting a suitable taxonomic resolution (i.e., taxonomic sufficiency; see Lenat and Resh 2001) to identify taxa (e.g., all taxa at family level or to the lowest practical level), 2) the site-selection strategy (e.g., based on a regular grid or targeting sites with potential impacts), and 3) the geographic extent of the biomonitoring area.

Identifying taxa at a finer taxonomic resolution often reveals that taxa of the same family can respond differently to environmental conditions. Genera or species within a family may have

more varied geographic distributions and specific ecological preferences (i.e., the environmental tolerances that define the niche of a taxon) that differ from the family as a whole (Bailey et al. 2001, Jones 2008). However, identifying stream invertebrates at a finer taxonomic resolution is costlier and requires specialized taxonomic expertise, additional labor, and quality-control procedures (Haase et al. 2004, Marshall et al. 2006, Jones 2008). In addition, some research has found that higher taxonomic resolution is not always necessary for understanding the effects of particular environmental conditions on stream invertebrates. In the case of one biotic index to assess pesticide pollution, the SPEAR$_{pesticides}$ index (Liess and Ohe 2005), previous analyses have shown that using family-level instead of species-level data does not lead to a significantly lower ability of the index to predict pesticide toxicity as quantified by chemical analyses (Beketov et al. 2009).

The selection of sampling locations is usually constrained by the spatial extent of the biomonitoring area (e.g., limited by political or administrative boundaries) and dependent on the goals of the program, such as assessing biodiversity or identifying the impacts of specific stressors (e.g., impaired water quality caused by intensive agriculture) or management measures (e.g., river restoration) on aquatic organisms. Applying a site-selection strategy that matches the program goals is crucial because biased sampling of specific environmental conditions can affect the quantification and inferred relative importance of natural and anthropogenic impacts driving invertebrate assemblage composition (Thuiller et al. 2004, Kuemmerlen et al. 2016). For example, a grid-based site-selection strategy covering a wide range of environmental conditions may be suitable to obtain a representative overview of the diversity of a region but may undersample impaired sites, such as sites immediately downstream of wastewater treatment plants.

In general, effects of human impacts on organisms can be quantified in two ways. The first method is a reference condition approach wherein the researcher calibrates a model with data from (near-) natural reference sites, applies the model to predict the natural assemblage at impaired sites, and compares the predictions with observations. The discrepancy between expected and observed probability of occurrence is used as a measure for the impairment. In the second method, the researcher uses a calibrated model that considers both natural and anthropogenically modified environmental conditions as explanatory variables (predictors) from

natural and altered sites and analyzes the sensitivity of the predictions to the environmental conditions. The most prominent implementation of the first approach is the River Invertebrate Prediction and Classification System method developed in the UK in the late 1980s (see overview by Wright et al. 2000). Since its development, the technique has been refined and applied to other countries (e.g., Hawkins and Yuan 2016, Rääpysjärvi et al. 2016).

The second approach to quantifying how macroinvertebrate taxa respond to natural and anthropogenically modified environmental conditions is the use of statistical species distribution models (SDMs). Increasingly sophisticated SDMs have been applied to stream invertebrate assemblages. Typical approaches to applying an SDM include fitting an individual model to each taxon in the community (often with identical explanatory variables for each model) to yield "stacked" individual SDMs (iSDMs; Guisan and Rahbek 2011) or combining multiple SDMs for one taxon in an ensemble-based approach (Thuiller et al. 2009). However, taxa identified at a higher taxonomic resolution often include families with rare genera and species (i.e., with a low prevalence and insufficient presence data points), and the responses of rare taxa are difficult to identify with individual models (Elith et al. 2006, Hernandez et al. 2006, Hernandez et al. 2006, Sor et al. 2017). More recent studies have applied hierarchical multi-species distribution models (hmSDM) in which the responses of multiple taxa to environmental conditions are jointly inferred (Latimer et al. 2009, Ovaskainen and Soininen 2011). hmSDMs are often implemented as extensions of the generalized linear model (GLM; Guisan et al. 2002) and quantify the responses of all individual taxa and the overall assemblage as an aggregation of the taxon-specific responses in the model. For example, Ovaskainen and Soininen (2011) extended stacked iSDMs (in this case, GLMs) into an hmSDM by assuming that the parameters quantifying the responses of all taxa to each environmental variable in the model are distributed according to an overarching community distribution (i.e., a community response), effectively leading to a hierarchical GLM. Multi-species distribution models have been extended into joint-species distribution models (jSDMs) that parameterize the residual correlations among taxa (i.e., the covariance among taxa that is not explained by the predictors) to account for biotic interactions or joint responses to unknown factors (Clark et al. 2014, Warton et al. 2015).

In this paper, we follow the SDM approach for the following reasons: 1) in highly impacted regions, few reference sites with near-natural conditions can be found; 2) the approach

allows us to directly extract information about the impact of different stressors; 3) it allows us to infer responses of different taxa to natural and anthropogenic stressors; and 4) if the model fits well, it will allow us to predict expected macroinvertebrate assemblages under natural conditions as well as the impact of different stressors.

The overarching goal of this study is to explore how key aspects of monitoring design affect the inference of stream invertebrate responses to natural and anthropogenic environmental conditions. Based on a previous investigation (Caradima et al. 2019), we selected and applied an hmSDM to four different datasets of stream invertebrate assemblages in Switzerland that differ in taxonomic resolution, site-selection strategy, and geographic extent to address the following two research questions:

1. How do various aspects of biomonitoring design (taxonomic resolution, site-selection strategy, geographic extent) affect our ability to explain and predict the occurrence of stream invertebrates and quantify the relative importance of explanatory variables in the model?

2. To what extent can a model-based analysis support the identification of taxa for which a finer taxonomic resolution would improve our ability to infer anthropogenic impacts on freshwater ecosystems, and how does this ability depend on the prevalence of these taxa?

Based on these research questions we derive recommendations for practice on how to further improve monitoring design and management of macroinvertebrates in streams.

2   Method

2.1   Study area

Located in central Europe, Switzerland has a wide variation in topography, a temperate climate, and abundant freshwater resources. The flat-to-hilly terrain of the Swiss Plateau contains the largest urban and agricultural areas and is bordered by the sub-alpine Jura Mountains to the northwest and the Swiss Alps to the south. The Swiss Alps are significant headwater sources for four major European river basins, including the Rhine catchment encompassing most of Switzerland, the Rhone in the southwest, the Ticino River in the south, and the Inn to the east.

Historically, excessive nutrient inputs from urban areas and agriculture, extensive channelization of lowland rivers (for land reclamation and flood control), and the construction of hydroelectric dams have led to habitat losses, impaired ecosystem functioning, and declines in biodiversity (Hering et al. 2012). Since the early 1970s, significant advances in environmental regulation and investments in wastewater treatment plants have greatly improved the ecological integrity of Swiss rivers. Current challenges in Swiss water management include the conservation of biodiversity and native species, particularly through impact mitigation of constructed dams, restoring the natural morphology of channelized rivers, and reducing in-stream micropollutant concentrations (e.g., pharmaceutical substances, pesticides) from wastewater discharge and agricultural runoff (Hering et al. 2012, Eggen et al. 2014).

2.2 Biomonitoring design and data

We combined data from different monitoring programs into four datasets that differ in taxonomic resolution, site-selection strategy, and geographic extent to investigate how various aspects of biomonitoring design affect our ability to explain and predict the occurrence of stream invertebrates. We derived presence-absence data from semi-quantitative abundance observations of stream invertebrates throughout Switzerland from the Center for Swiss Cartography of Fauna Macroinvertebrate Database (MIDAT, © CSCF Neuenburg) and prepared the datasets based on Swiss federal and cantonal monitoring programs. All monitoring programs apply the same multi-habitat sampling strategy (Stucki 2010). The first dataset is based on the Biodiversity Monitoring Switzerland program (BDM; BDM Coordination Office 2009), a federal program designed to monitor terrestrial and freshwater biodiversity in Switzerland. This dataset resolves the Ephemeroptera (mayflies), Plecoptera (stoneflies), and Trichoptera (caddisflies) (EPT) orders to species and genus level and most other taxa to family level, with a few exceptions at coarser levels. It includes 580 samplings at 492 sites from 2011 to 2015. The site-selection strategy of BDM is based on the Swiss topographic grid (Fig. 1), with 49% of sites located in small headwater streams (i.e., second order streams; Fig. 2). Throughout this study, we abbreviate this dataset as SGs (Switzerland-wide dataset with Grid sampling of species as the finest taxonomic level).
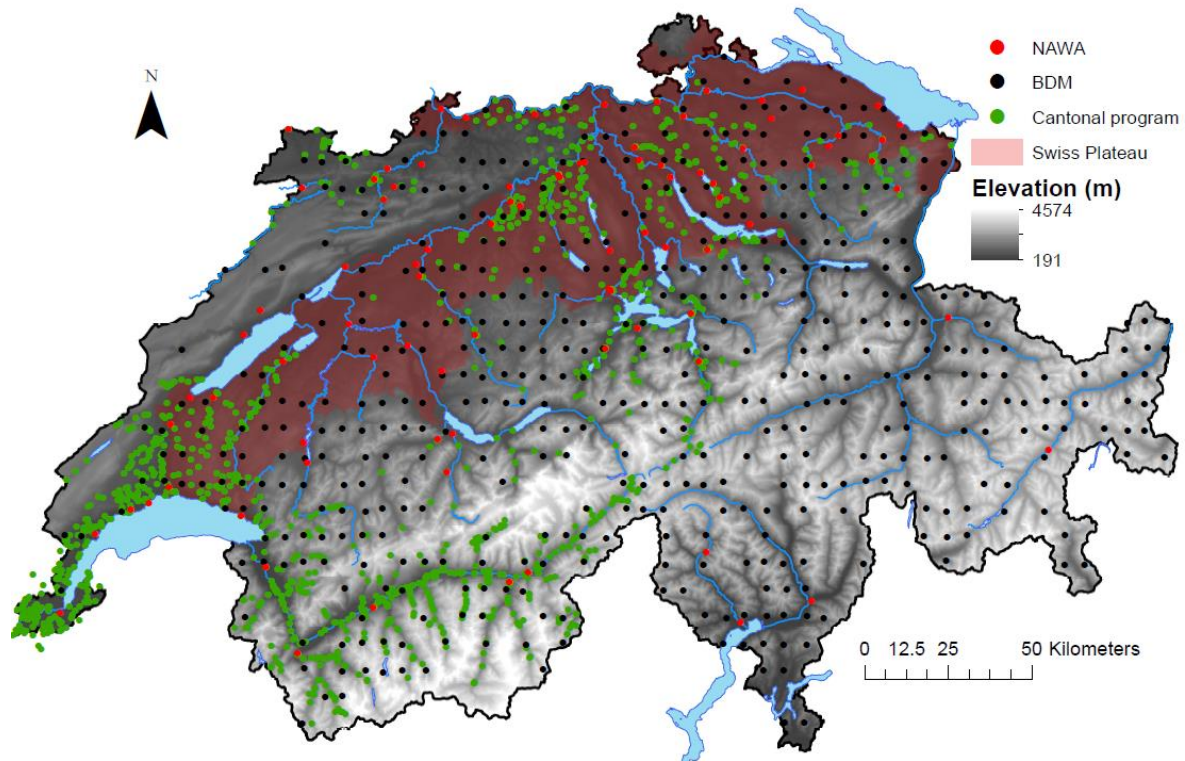
Fig. 1. Geographic distribution of federal and cantonal biomonitoring sites in Switzerland, including the National Surface Water Quality Monitoring Programme (NAWA) and Biodiversity Monitoring (BDM) program.
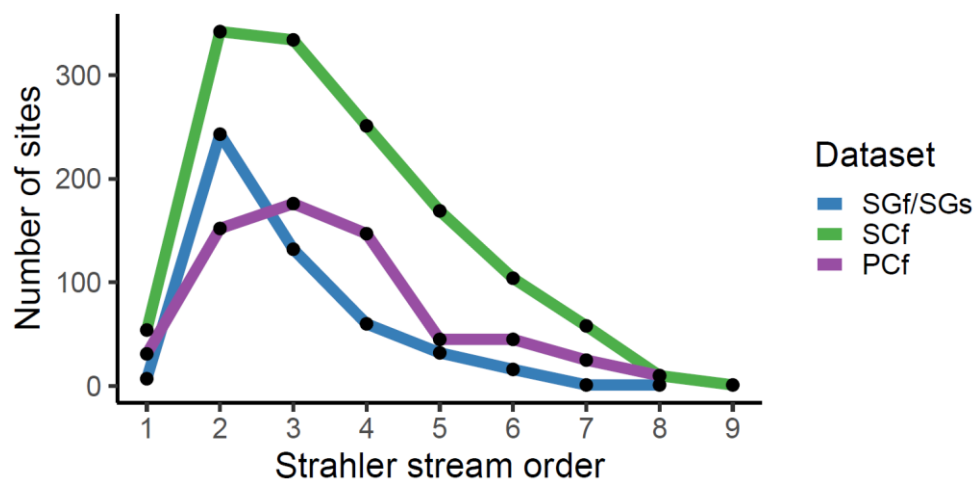

Fig. 2. Frequency of Strahler stream orders of sites in the Switzerland Grid families (SGf), Switzerland Grid species (SGs), Switzerland Combined families (SGf), and Plateau Combined families (PCf) datasets.

We prepared a second data set from this monitoring program by pooling the EPT species and genera at the family level to analyze the effects of taxonomic resolution on our ability to infer responses to environmental conditions (Table 1). We abbreviate this dataset as SGf (Switzerland-wide dataset with Grid sampling of family as finest taxonomic level). As a result, the SGf and SGs datasets share the same sites and sample size but differ greatly in taxonomic diversity (Table 1, Appendix S1 Fig. S9).

Table 1. Sample size and number of taxa in the datasets.

Dataset	In-text abbreviation	Sites	Samples	Taxa
Switzerland Grid families	SGf	493	581	103
Switzerland Grid species	SGs	493	581	245
Switzerland Combined families	SCf	1330	1802	124
Plateau Combined families	PCf	637	950	117

The third dataset included data from the National Surface Water Quality Monitoring Programme (NAWA; BAFU 2013, Kunz et al. 2016), which uses a different method for selecting stream invertebrate monitoring sites. We used this data to investigate the effect of site-selection strategy on modeling success and inference of environmental variables. NAWA selects only sites with catchment areas >25 km^2 to provide a representative overview of water quality in midsized to larger streams in Switzerland, covering different river types and all cantons (Fig. 1). As a result, mostly unimpaired sites with near-natural conditions are underrepresented in the NAWA program. In contrast, the cantonal programs (including Vaud, Aargau, Zürich, Bern, Geneva, St Gallen, Valais, Jura, Appenzell Ausserrhoden, Uri, Schwyz, Obwalden, Nidwalden, and Luzern) have various site-selection strategies and often include sites that are assumed to be impaired (e.g., downstream of wastewater treatment plants or with intensive agriculture; Fig. 1). We aggregated the data to the family level to derive a dataset that combines BDM, NAWA, and the cantonal monitoring programs. We chose family level because federal and cantonal programs use different approaches to select taxa to identify at the genus or species level (e.g., a species identified in Zürich may not be identified in the neighboring canton of St Gallen). Abbreviated as

SCf (Switzerland Combined families), this dataset includes a greater number of sites in midsized to larger streams than the SGf and SGs datasets (Figs 1, 2).

For our fourth dataset, we selected a subset of the SCf sites to analyze how the geographic extent of the monitoring data used for calibration affects model performance and the quantification of anthropogenic impacts (Fig. 1, Table 1). We abbreviate this dataset as PCf (Plateau Combined families).

2.3 Model definition

Here, we define the individual models that we used to select the most important explanatory variables to predict the composition of stream invertebrate assemblages. We then extended the individual models into a hmSDM (as defined in greater detail by Caradima et al. 2019) that we then applied to analyze how key aspects of monitoring design affect the inference of stream invertebrate responses to natural and anthropogenic environmental conditions. To define the models, we use the following indices: sites: $i \in \{1, ..., I\}$; sampling time at each site: $t_i \in \{1, ..., T_i\}$; taxa: $j \in \{1, ..., J\}$; and explanatory variables: $k \in \{1, ..., K\}$.

Given a presence-absence observation with the response $y_{it_ij} = 1$ if a taxon is observed as present and $y_{it_ij} = 0$ if the taxon is observed as absent, we used a generalized linear model (GLM) with a logistic link function to model the probability of occurrence for an individual taxon $P(Y_{it_ij} = 1 \mid \mathbf{x}, \boldsymbol{\alpha}^{\text{taxa}}, \boldsymbol{\beta}^{\text{taxa}})$. The probability of occurrence $P(Y_{it_ij} = 1 \mid \mathbf{x}, \boldsymbol{\alpha}^{\text{taxa}}, \boldsymbol{\beta}^{\text{taxa}})$ given the selected explanatory variables $\mathbf{x}$ and the taxon-specific parameters $\boldsymbol{\alpha}^{\text{taxa}}$ and $\boldsymbol{\beta}^{\text{taxa}}$ is equal to the link function evaluated at the linear predictor z, with the linear predictor equal to the linear combination of the explanatory variables and parameters:

$$z_{it_ij} = \alpha_j^{\text{taxa}} + \sum_{k=1}^{K} x_{it_ik}\beta_{kj}^{\text{taxa}} \text{ (Eq. 1), and}$$

$$P(Y_{it_ij} = 1 \mid \mathbf{x}, \boldsymbol{\alpha}^{\text{taxa}}, \boldsymbol{\beta}^{\text{taxa}}) = \frac{1}{1+e^{-z_{it_ij}}} \text{ (Eq. 2).}$$

Note that in our model, the predicted probability of occurrence for a given taxon j at site i at sample time t_i is dependent on specific explanatory variables that vary based on the sample (see Appendix S1 Table S1). The observations y_{it_ij} are assumed to be Bernoulli distributed with probabilities:

$$P\left(y_{it_ij} \mid \mathbf{x}, \boldsymbol{\alpha}^{\text{taxa}}, \boldsymbol{\beta}^{\text{taxa}}\right) = \begin{cases} P\left(Y_{it_ij} = 1 \mid \mathbf{x}, \boldsymbol{\alpha}^{\text{taxa}}, \boldsymbol{\beta}^{\text{taxa}}\right) & \text{if } y_{it_ij} = 1 \\ 1 - P\left(Y_{it_ij} = 1 \mid \mathbf{x}, \boldsymbol{\alpha}^{\text{taxa}}, \boldsymbol{\beta}^{\text{taxa}}\right) & \text{if } y_{it_ij} = 0 \end{cases} \quad \text{(Eq. 3)}.$$

When fitting an iSDM to each taxon, the joint probability of the observations given the model structure, inputs, and parameters is:

$$P(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\alpha}^{\text{taxa}}, \boldsymbol{\beta}^{\text{taxa}}) = \prod_{i=1}^{I} \prod_{t_i=1}^{T_i} \prod_{j=1}^{J} P\left(y_{it_ij} \mid \mathbf{x}, \boldsymbol{\alpha}^{\text{taxa}}, \boldsymbol{\beta}^{\text{taxa}}\right) \quad \text{(Eq. 4)}$$

The hierarchical multi-species model assumes that the taxon-specific responses (i.e., the parameters α_j^{taxa} and β_{kj}^{taxa}) are distributed and constrained according to their respective community responses (i.e., the community parameter distributions α^{comm} and β_k^{comm}, which are not jointly conditioned upon a hyper-hyper prior) (Ovaskainen and Soininen 2011). The taxon-specific parameters are then inferred jointly with the community parameters from the data. The joint probability distribution of the model parameters and outcomes is given by the product of the marginal and conditional probability densities and is formulated as a probability network (Fig. 3, Eq. 5).
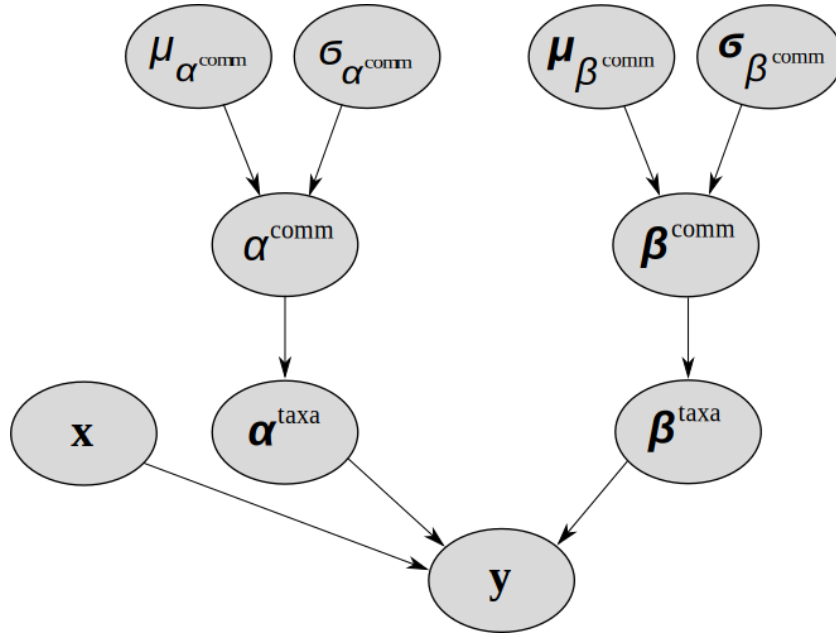

Fig. 3. Network representation of conditional probability distributions in the hierarchical multi-species model. The nodes $\mu_{\alpha}\text{comm}, \sigma_{\alpha}\text{comm}$ and $\boldsymbol{\mu}_{\boldsymbol{\beta}}\text{comm}, \boldsymbol{\sigma}_{\boldsymbol{\beta}}\text{comm}$ parameterize the means

and standard deviations of their respective community distributions α^{comm} and $\boldsymbol{\beta}^{\text{comm}}$. The nodes $\boldsymbol{\alpha}^{\text{comm}}$, $\boldsymbol{\beta}^{\text{comm}}$ are taxon-specific parameters (i.e., responses) constrained by and jointly inferred with their respective community distributions α^{comm}, β^{comm} from the model inputs (i.e., explanatory variables) $\mathbf{x}$ and model output $\mathbf{y}$. Variables and parameters in bold are vectors or matrices.

Assuming the community parameters α^{comm} and β_k^{comm} are normally distributed and each are informed by respective means and standard deviations $\mu_{\alpha^{\text{comm}}}$, $\sigma_{\alpha^{\text{comm}}}$ and $\mu_{\beta_k^{\text{comm}}}$, $\sigma_{\beta_k^{\text{comm}}}$, the joint probability distribution of the model is given as:

$$P\left(\mathbf{y}, \boldsymbol{\alpha}^{\text{taxa}}, \boldsymbol{\beta}^{\text{taxa}}, \mu_{\alpha^{\text{comm}}}, \sigma_{\alpha^{\text{comm}}}, \boldsymbol{\mu}_{\boldsymbol{\beta}^{\text{comm}}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}^{\text{comm}}} \mid \mathbf{x}\right) =$$
$$P(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\alpha}^{\text{taxa}}, \boldsymbol{\beta}^{\text{taxa}}) \times \sum_{j=1}^{J} \left(f_{\text{N}}(\alpha_j \mid \mu_{\alpha^{\text{comm}}}, \sigma_{\alpha^{\text{comm}}}) \times \sum_{k=1}^{K} f_{\text{N}}\left(\beta_{kj} \mid \mu_{\beta_k^{\text{comm}}}, \sigma_{\beta_k^{\text{comm}}}\right)\right) \text{(Eq. 5)},$$
$$\times f\left(\mu_{\alpha^{\text{comm}}}, \sigma_{\alpha^{\text{comm}}}, \boldsymbol{\mu}_{\boldsymbol{\beta}^{\text{comm}}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}^{\text{comm}}}\right)$$

where f_{N} is the density of the normal distribution and f is the density of the joint prior distribution of the means and standard deviations of the community parameters.

We estimated the parameters of the hmSDM by Bayesian inference and used uniform priors for the means and standard deviations of the community parameter distributions. We sampled from the joint posterior probability distribution of the model parameters by doing Bayesian inference with a Hamiltonian Monte Carlo (HMC; see Duane et al. 1987, Brooks et al. 2011) algorithm implemented in Stan (Carpenter et al. 2017) and accessed through the package *rstan* (Stan Development Team 2018) in R (version 3.5, R Core Team, 2018).

2.4 Model performance

The quality of fit, explanatory power, and predictive performance for each taxon in the individual and hierarchical multi-species models is based on the deviance (a statistic equivalent to the residual sum of squares under the assumption of normality) given as

$$d_j = -2 \sum_{i=1}^{I} \sum_{t_i=1}^{T_i} \log\left(P\left(y_{it_ij} \mid \mathbf{x}, \boldsymbol{\alpha}^{\text{taxa}}, \boldsymbol{\beta}^{\text{taxa}}\right)\right) \quad \text{(Eq. 6)},$$

where d_j is the taxon-specific deviance over all samples based on the probabilities of occurrence predicted by the proposed or null model. In the null model for a taxon, the taxon-specific intercept α_j^{taxa} corresponds to the prevalence of the taxon and the model parameters β_{kj}^{taxa} are set

to zero. Dividing the deviance by the number of samples for a taxon yields a standardized deviance that enables comparisons of quality of fit and predictive performance among taxa with different sample sizes.

When calibrating the hmSDM to each dataset, we measured the explanatory power of the selected variables for each taxon with the D^2 statistic (similar in interpretation to the R^2 of a linear model assuming a normally distributed response) to assess the proportion of null model deviance that is reduced by the proposed model (i.e., by including explanatory variables; Guisan and Zimmermann 2000):

$$D_j^2 = \frac{d_j^{\text{null}} - d_j^{\text{proposed}}}{d_j^{\text{null}}} \text{ (Eq. 7).}$$

The predictive performance for each taxon in each of the four models was assessed with *k*-fold cross-validation. Samples were randomly divided into $k = 3$ subsamples of nearly equal size. Each model was calibrated with all combinations of $k - 1$ subsamples and used to predict the probability of occurrence for each taxon in the independent subsample. We chose $k = 3$ because of the computationally intensive inference of the hmSDM and to avoid deteriorating the predicted probabilities of occurrence in the independent subsample. We then obtained the mean standardized deviance over k subsamples for each taxon as a measure of predictive performance. We quantified the geographic distribution of the quality of fit and predictive uncertainty for all individual taxa in each model by propagating a thinned subsample of the posterior distribution through the model and obtaining the 5th and 95th quantiles of the marginal posterior probabilities of occurrence for comparison with the observations at each sampling site.

2.5   Selection of explanatory variables

Following methods used by Caradima et al. (2019), we applied individual models to identify a set of explanatory variables that provides the best predictive performance for multiple datasets of stream invertebrate taxa. We identified potential explanatory variables for the occurrence of taxa based on *a priori* expert knowledge and existing literature about potential causal mechanisms (Table 2; see Appendix S1 Tables S1, S2 for details).

Table 2. Potential explanatory variables for the composition of stream invertebrate communities in Switzerland (selected variables for the model with the highest predictive performance are in bold and detailed definitions for all variables are found in Appendix S1 Tables S1, S2). The mean of each explanatory variable was subtracted at each site ($x_k = x_{ik} - \bar{x}_{ik}$) to reduce correlations among the marginal posterior parameter distributions. Temp (temperature) and FV (flow velocity) include a quadratic transformation to identify taxa with midrange preferences for temperature and flow velocity (i.e., Temp^2, FV^2).

Explanatory variable	Abbreviation (units)
Proportion of riparian agriculture (A) or forest (F) cover within a buffer distance (e.g., 10 m) along the upstream riparian zone in the catchment	**A10m**, A100m, A1km, F10m, F100m, F1km (%)
Land use index for agricultural insecticides (insecticide application rate): fractions of cropland in the catchment weighted by the average number of insecticide spray treatments for each respective cropland (mean annual spray treatments × fraction of cropland)	**IAR** (unitless)
Livestock unit density: cattle equivalent (CE) units of livestock per square kilometer of catchment area	**LUD** (CE/km^2)
Proportion of urban and transport-related land cover within the catchment	**Urban** (%)
Proportion of forest-river intersection within 150 m of the site and within the subcatchment of the site	**bFRI** (%)
Forest-river intersection: proportion of upstream river length in the catchment intersected by forest cover.	**FRI** (%)
Stream flow velocity estimated from spatial data (channel width, channel slope, mean annual simulated discharge, and Manning's friction coefficient)	**FV** (m/s)
Evaluation of stream channel width variability (0 = bad, 1 = very good)	**WV** (unitless)
Predicted mean maximum morning water temperature during the summer based on a linear model with catchment area and mean catchment elevation as inputs	**Temp** (^{o}C)

15

Explanatory variable	Abbreviation (units)
Evaluation of stream bed modification (0 = bad, 1 = very good)	BM (unitless)
Evaluation of overall stream morphology (0 = bad, 1 = very good)	Morph (unitless)
Total agricultural (A) or forest (F) land cover in the catchment inversely weighted by the Euclidean distance to the site	A.EDO, F.EDO (%)
Toxic units (TU) of modelled micropollutant concentrations for *Daphnia magna* (TU.Dm) and *Chironomidae* (TU.Cr)	TU.Dm, TU.Cr
Urban index: sum of specific urban land uses weighted by stormwater discharge coefficients	UI (%)
Discharge of treated wastewater as a proportion of stream discharge	WW (%)
The fraction of residual stream discharge from upstream hydroelectric dams	HP (unitless)

In an exhaustive search procedure similar to a best subsets approach (James et al. 2013), we constructed models containing all possible combinations of *p* potential explanatory variables and applied 3-fold cross-validation to each model for all taxa. The number of potential models to cross-validate was constrained by excluding models with collinear variables (i.e., Pearson correlation >0.6 or <-0.6; see Appendix S1 Figs S1–S3) and selecting models with stream temperature (Temp), the land use index for agricultural insecticides (IAR), and flow velocity (FV), which were previously shown to be key explanatory variables (Caradima et al. 2019). We also included quadratic terms for variables for which we can expect a parabolic shape of the response to the linear predictor (Eq. 1). The quadratic terms allow us to model taxa that prefer intermediate levels of an explanatory variable (e.g., temperature) and respond negatively to both low and high values.

We calculated the predictive performance of each model as the mean standardized deviance over all taxa. We did the exhaustive search procedure separately for each family-level dataset to explore the predictive performance of additional explanatory variables in a dataset with a larger sample size (Table 1). For the SGf and PCf data sets, we included models with 5 to 13 parameters, and for the SCf dataset we included 5 to 15 parameters. We selected a final compromise model based on the lowest mean standardized deviance across the family-level datasets.

The taxon-specific parameters $\alpha_j^{\text{taxa}}, \beta_{kj}^{\text{taxa}}$ in the individual models were identified by maximum likelihood estimation. We obtained parameter estimates with an iterative weighted least squares algorithm in the glm function in R and used 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 (Eq. 6), 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). Because of the difficulty in identifying the parameters of individual models for rare taxa through maximum likelihood estimation, taxa occurring in <10% of the samples were excluded from model selection.

2.6 Relative importance of explanatory variables

With the hmSDM calibrated to each dataset, we calculated the relative importance of the selected explanatory variables for each taxon as the individual contribution of each explanatory variable to the linear predictor (Eq. 1) at each sample (z_{it_ikj}, excluding the taxon-specific intercept α_j^{taxa}):

$$z_{it_ikj} = x_{it_ik}\beta_{kj} \quad \text{(Eq. 8)},$$

where x_{ik} are the explanatory variables in the dataset with the widest range over the study area (i.e., SCf), and β_{kj}^{taxa} is the maximum of the posterior taxon-specific parameter distribution inferred for each dataset and explanatory variable. The linear predictor for an explanatory variable with a quadratic transformation (e.g., Temp^2) is:

$$z_{it_ikj} = x_{it_ik}\beta_{kj}^{\text{taxa}} + x_{it_ik}^2\beta_{kj}^{\text{quad}} \quad \text{(Eq. 9)},$$

where β_{kj}^{quad} is the taxon-specific parameter for a quadratic explanatory variable. The range of the linear predictor (z_{kj}^{range}) is then the overall contribution of an explanatory variable to the occurrence probabilities of a taxon over all samples:

$$z_{kj}^{\text{range}} = \max(z_{it_ikj}) - \min(z_{it_ikj}) \quad \text{(Eq. 10)}.$$

Given that not all taxa will exhibit a good quality of fit, using the variability of the linear predictor z_{kj}^{range} to quantify the relative importance of the explanatory variables may produce biased results.

2.7 Effect of taxonomic resolution

We sorted the EPT families according to their prevalence and then compared the explanatory power and predictive performance of the model for the different taxonomic levels to assess the effect of taxonomic resolution on model performance for each taxon. Furthermore, we compared the geographic distribution of the observations and the model output for taxa that belong to the same family. The effect of taxonomic resolution on our ability to infer responses of EPT taxa to environmental conditions was compared directly using the full samples of the

marginal posterior distributions of the taxon-specific parameters β_{kj}^{taxa} for all families and their respective genera and species.

3 Results

Effects of monitoring design on model performance and relative importance of explanatory variables

Selection of explanatory variables   Based on the exhaustive search procedure, we selected explanatory variables with the highest mean predictive performance among all four data sets. The top ten models have a similar standardized deviance and had a generally consistent set of explanatory variables, including multiple agricultural and forest-related explanatory variables (Table 3). Therefore, the choice among those top performing models is rather arbitrary. A robust choice is to select variables that occurred in most of the top performing models. This is the case for the model with the lowest mean predictive error among the family-level datasets, which we selected to compare model performance among the datasets. Based on the mean standardized deviance, the compromise model ranked 126th for SCf, 47th for PCf, and 1st for SGf.

Table 3. Top 10 models with the lowest mean standardized deviance during prediction over the family-level datasets. We provide three digits to illustrate the similarity of the values for these models. This does not imply that these differences are significant. For complete definitions of all explanatory variables, see Appendix S1 Tables S1, S2.

	Mean standardized deviance			
Model	SCf	PCf	SGf	Mean
A10m IAR LUD Urban bFRI FRI FV WV Temp Temp^2	0.847	0.940	0.871	0.886
A10m IAR LUD Urban FRI FV WV WW Temp Temp^2	0.847	0.936	0.882	0.888
A10m IAR LUD Urban FRI FV WV Temp Temp^2	0.847	0.936	0.882	0.889
A10m IAR LUD Urban F.EDO bFRI FRI FV WV Temp Temp^2	0.845	0.938	0.884	0.889
A10m IAR LUD Urban bFRI FRI FV WV WW Temp Temp^2	0.846	0.940	0.881	0.889
A10m IAR LUD Urban F.EDO FRI FV WV Temp Temp^2	0.846	0.934	0.887	0.889
A10m IAR LUD Urban F.EDO FRI FV WV WW Temp	0.849	0.935	0.885	0.890

19

A10m IAR LUD Urban F.EDO bFRI FRI FV Temp Temp^2	0.851	0.939	0.881	0.890
A10m IAR LUD Urban bFRI FRI FV WV BM Temp Temp^2	0.846	0.942	0.883	0.890
A10m IAR LUD Urban FRI FV WV Temp	0.852	0.938	0.880	0.890

The distributions of the selected explanatory variables in each dataset reflect the different site-selection strategies of the underlying monitoring programs (Fig. 4; see Appendix S1 Figs S4–S6 for pairwise comparisons and Appendix S11 for the geographic distributions of the values of the selected explanatory variables). Variables related to urban land use (Urban) and agricultural land use (e.g., riparian agriculture, agricultural insecticides, [A10m], insecticide land use index [IAR], livestock unit densities) [LUD]) were lowest in the grid-based SGf dataset, higher in the combined SCf dataset, and highest in the Swiss Plateau PCf dataset. In SGf, sites had higher flow velocity (FV), colder stream temperatures (Temp), and a larger range of riparian forest cover (FRI), which is consistent with a greater proportion of sites with low stream order (Fig. 2). By contrast, the PCf dataset included a higher proportion of sites with slower flow velocities and higher stream temperatures, and 20% of sites featured an impaired morphological evaluation of the channel width variability (WV; compared to 17% in SCf and 10% in SGf) (BAFU 2006).

20
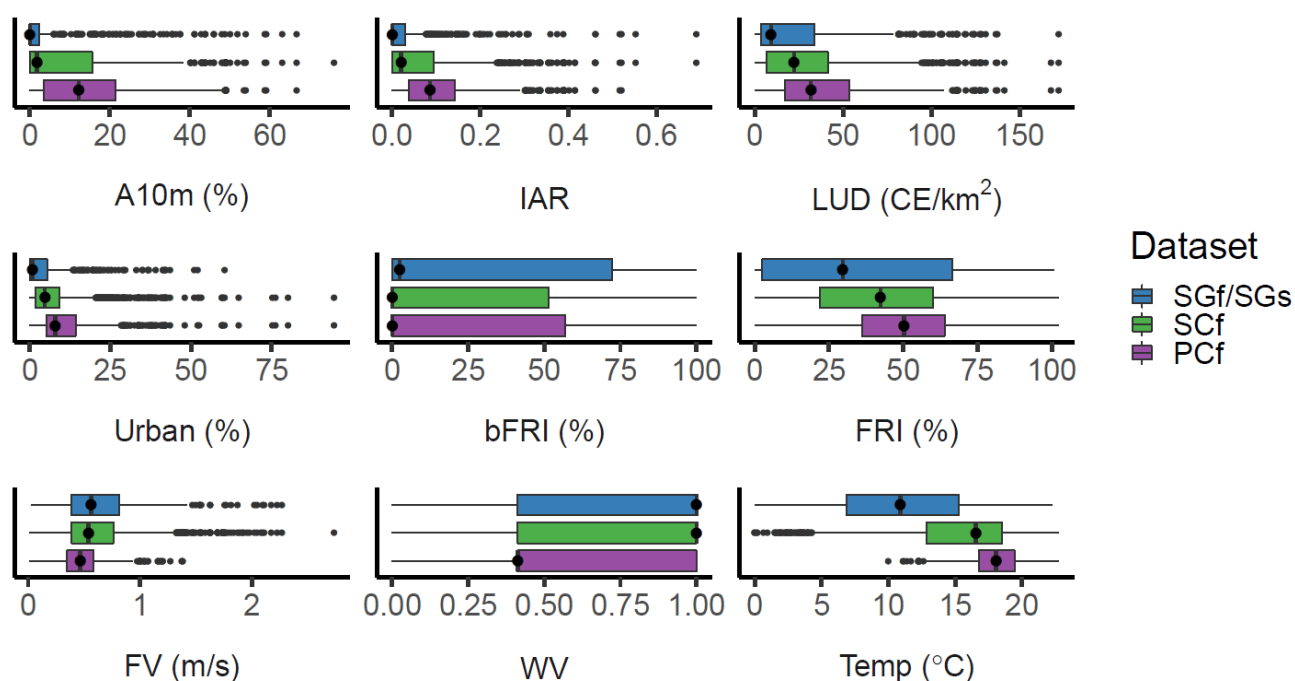
Fig. 4. Boxplots showing the distribution of the selected explanatory variables in each dataset, including variables related to riparian agriculture (A10m), agricultural insecticides (IAR), livestock unit density (LUD), urban land use (Urban), riparian forest cover at various spatial scales (bFRI, FRI), flow velocity (FV), channel width variability (WV), and stream temperature (Temp) (see Table 2 for complete definitions). Box plots show the quartiles (colored box) and whiskers (black line) extending to the most extreme data point (outliers are defined as >1.5× the interquartile range), and black points indicate the position of the median of skewed distributions (see Appendix S11 for maps of the selected explanatory variables).

The range of the linear predictor over all samples and for each taxon and explanatory variable (z_{kj}^{range}; Fig. 5) indicates the relative importance of a given explanatory variable in the model. Stream temperature emerged as a dominant explanatory variable in all model versions, particularly in the model for the dataset with a grid-based site selection and a fine taxonomic resolution (SGs) (Fig. 5). The land use index for agricultural insecticides (IAR), urban land use (Urban), and flow velocity (FV) are of secondary importance in the models for all datasets. Variables related to riparian agriculture (A10m) and livestock unit densities (LUD) have a lower but consistent impact in the models for all datasets. In all models, riparian forest cover within the entire catchment (FRI) is more important than the riparian forest cover at the site. Riparian forest cover at the site and the width variability of the stream channel are among the least important explanatory variables for all datasets.
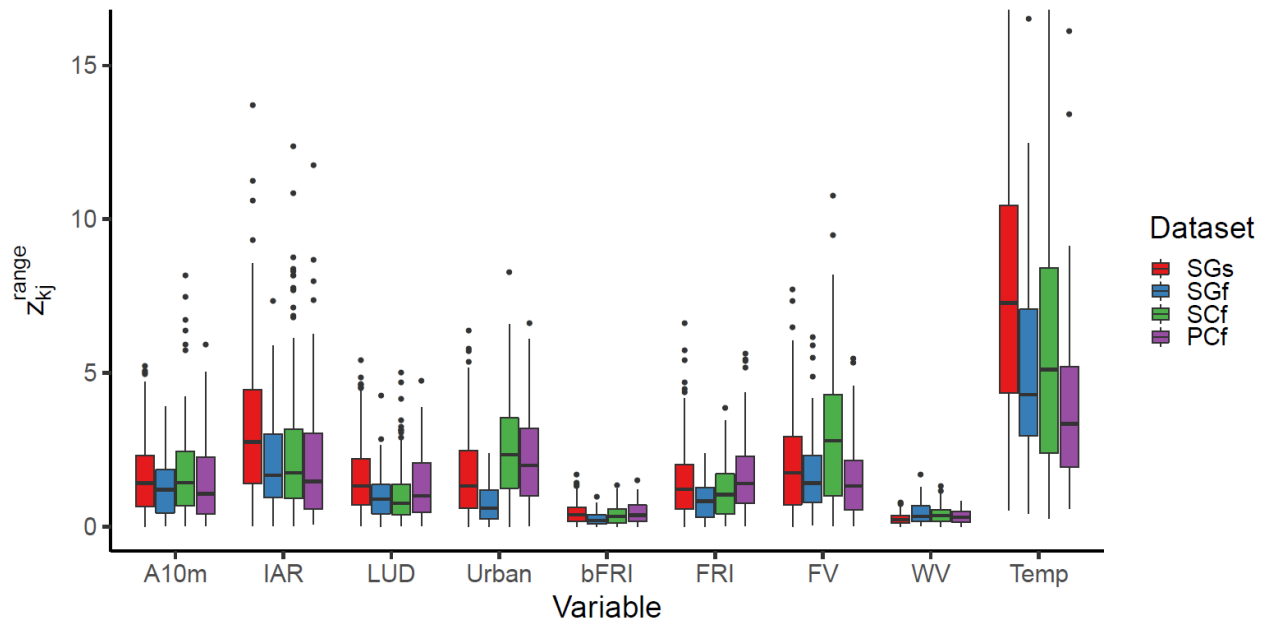
Fig. 5. Boxplots showing the relative importance of the explanatory variables in the model for each dataset quantified as the distribution of the range of the linear predictor $\left(z_{kj}^{\text{range}}\right)$ per taxon with the explanatory variables for the dataset with the widest range of environmental conditions, SCf (to ensure comparable z_{kj}^{range} values between models). The box shows the interquartile range and the black lines extending to outliers, which are defined as data points $>1.5\times$ the interquartile range. Note that the maxima of the z_{kj}^{range} distributions for stream temperature in SGs and SCf are 22 and 23, respectively.

Effects of monitoring design on model performance   We found that model performance was affected by monitoring design decisions including taxonomic resolution, site-selection strategy, and geographic extent. The hmSDMs achieved a similar performance, based on the standardized deviance, for calibration to the whole data set and for prediction during 3-fold cross-validation (Appendix S1 Figs S10, S11). When the model was calibrated to each dataset with all samples, the model for the dataset with a grid-based site selection and a finer taxonomic resolution (SGs) achieved a higher quality of fit (i.e., a lower standardized deviance) and explanatory power (i.e., a higher D^2) than any of the models for datasets with a family-level resolution (Fig. 6; see Appendix S1 Fig. S8 for similar results during cross-validation). Among the family-level

datasets, the quality of fit of the model was highest when applied to the dataset with a better coverage of environmental influence factors and increased sample size (SCf). However, the model applied to the dataset with sites limited to the more densely populated Swiss Plateau and family-level resolution (PCf) yielded the lowest quality of fit and explanatory power. These results indicate that a finer taxonomic resolution as well as a better coverage of environmental influence factors can both increase our ability to infer human impacts on stream invertebrates from biomonitoring data.
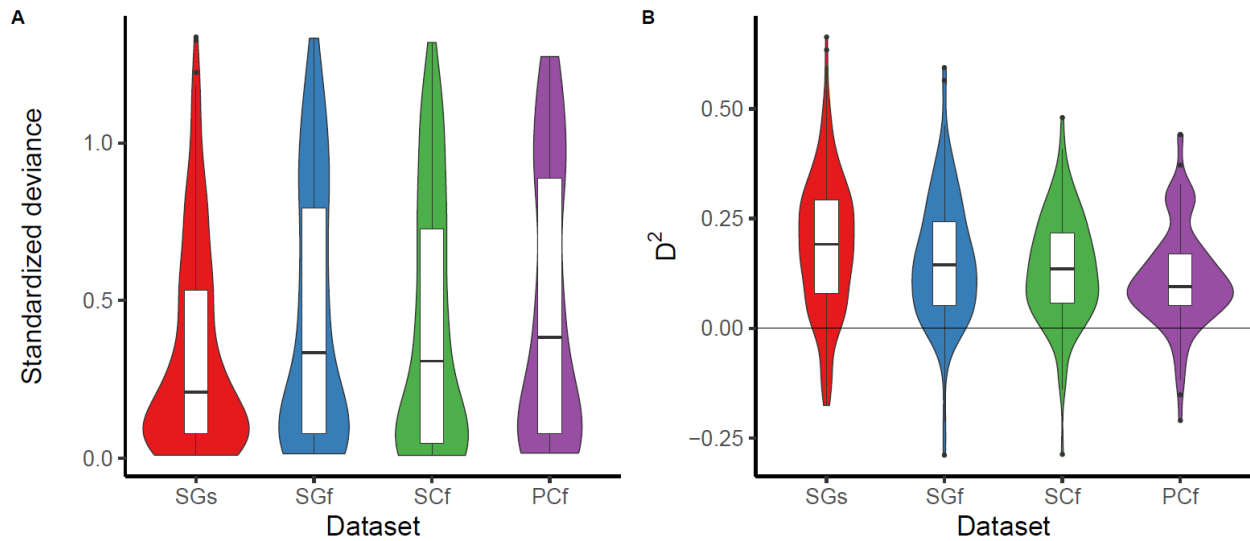
Fig. 6. Quality of fit (A) indicated by the standardized deviance (lower is better) and explanatory power (B), indicated by the D^2 (higher is better) of the selected variables for all taxa during calibration of the hierarchical multi-species distribution model to all samples in each dataset. (Note that the model can yield negative D^2 values for specific taxa because of the overarching community parameters constraining the taxon-specific responses; see Fig. 3). Box plots show the quartiles (white box) and whiskers (black line) extending to the most extreme data point (outliers are defined as greater than 1.5 times the interquartile range), and violin plots (colored) show the density distribution.

Effects of monitoring design on the relative importance of explanatory variables in the model Our model results showed that taxonomic resolution, site-selection strategy, and geographic extent had an influence on the relative importance of the explanatory variables (Fig.

5). Within grid-based site selection datasets, the relative importance of all explanatory variables (with the exception of the channel width variability) increased when moving from family-level (SGf) to genera and species resolution (SGs), particularly for anthropogenic variables such as riparian agriculture (A10), agricultural insecticides (IAR), livestock unit densities (LUD), and urban land use (Urban).

Contrary to our expectations, limiting the sites of the combined family-level dataset (SCf) to the more densely populated Swiss Plateau (PCf) did not increase the relative importance of anthropogenic explanatory variables (e.g., A10m, IAR, LUD, Urban) that mainly occur in the Swiss Plateau (Fig. 5). Note that this model also had a lower quality of fit and explanatory power than the other models (Fig. 6). Combining all biomonitoring programs at the family-level in the SCf dataset led to a relative importance of explanatory variables in the corresponding model that is similar to the model applied to grid-based data with a better taxonomic resolution (SGs), with the exception of stream temperature.

A comparison of the parameter estimates reflects the responses of taxa to the explanatory variables in the different model versions. The parameters generally had a consistent positive or negative response to the environmental variables in all models and had some differences in the absolute value for some variables (see Appendix S1 Figs. S12–S14). For example, family responses in SCf and SGf were similar for riparian agriculture (A10m) and agricultural insecticides (IAR), but differed in their responses to urban land use (Urban) and riparian forest cover at the site (bFRI) (Appendix S1 Fig. S12). When comparing responses of families in SCf with families observed at sites limited to the Swiss Plateau (PCf), the taxon-specific responses were generally similar but differed for flow velocity and riparian forest cover at the site, with minor differences for stream temperature (Appendix S1 Figs. S13, S14).

Effects of taxonomic resolution on explanatory power and inference of responses

Comparing EPT taxa in the models calibrated to the SGf and SGs datasets shows the effects of taxonomic resolution on the explanatory power of, and taxon-specific responses to, the explanatory variables. The model calibrated to the SGf dataset had a higher predictive performance (i.e., a lower deviance) for widespread families (e.g., Baetidae, Nemouridae, Limnephilidae) than for many of their genera and species when calibrated to the SGs dataset

(Fig. 7B). The distributions of very widespread families can be accurately predicted without explanatory variables (i.e., the null model has a reasonably good performance by predicting a high probability of occurrence everywhere) but have poor explanatory power (D^2) of the selected explanatory variables (Fig. 7A). However, multiple widespread families, such as Baetidae, Nemouridae, Heptageniidae, and Rhyacophilidae, include genera or species for which the proposed model had a higher explanatory power (Fig. 7A) than their respective family.
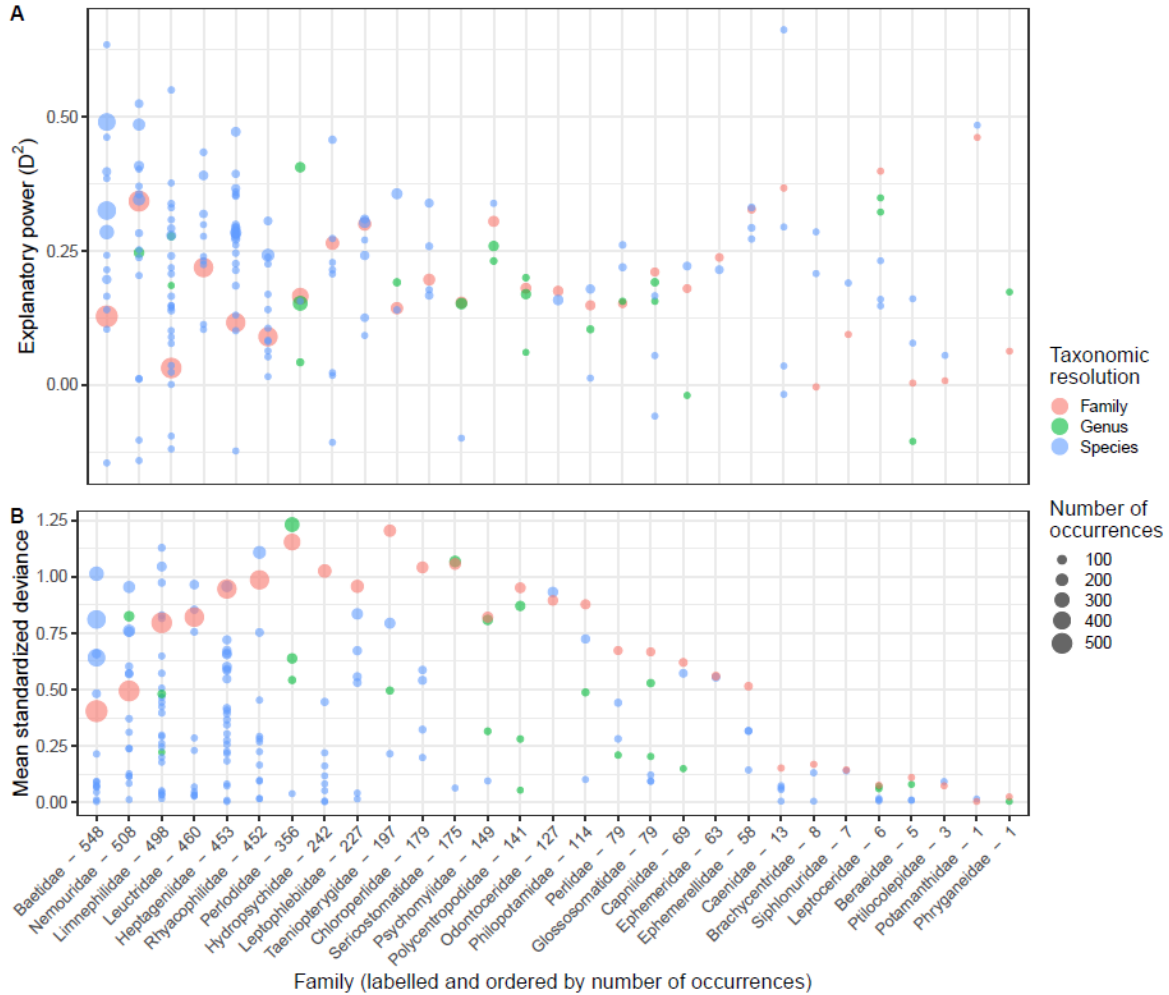
Fig. 7. Explanatory power (i.e., D^2 during model calibration) (A) and predictive performance (B) based on taxonomic resolution of Ephemeroptera, Plecoptera, and Trichoptera taxa in SGf and SGs, with predictive performance quantified as the mean standardized deviance during 3-fold cross-validation. A lower mean standardized deviance indicates a better

predictive performance. Numbers on the x-axis indicate the number of occurrence data points in the SGf dataset.

These differences in model performance are generally due to genera and species exhibiting stronger and different responses to the explanatory variables than their respective families. For example, Baetidae is a widespread mayfly family (Fig. 8A) that was present in 94% of the samples. The model has a low explanatory power ($D^2 = 0.13$) for Baetidae, which shows only a slight positive response to stream temperature and a negative response to riparian agriculture (Fig. 9). Specific baetid species, such as *Baetis alpinus*, *B. rhodani*, and *B. muticus*, differ in observed geographic distributions, which can be captured reasonably well by the model (Fig. 8). In the model, these species show stronger and more variable responses to explanatory variables than the family, which is illustrated by their taxon-specific parameters (Fig. 9) and higher D^2 values (Fig. 8). (See Appendix S2 for parameter plots for all taxa in SGf and SGs, Appendices S3–S6 for complete maps for all taxa and all datasets, and Appendices S7–S10 for plots showing response curves of the predicted probabilities of occurrence and observations vs the explanatory variables for all taxa in each dataset).

Additionally, *B. alpinus* is the only selected taxon of this family that responds negatively to the insecticide land use index (IAR) and positively to flow velocity, whereas *B. rhodani* and *B. muticus* respond negatively to flow velocity. Furthermore, many of the selected Baetis species show stronger and different responses to stream temperature (Temp and Temp^2) in the hmSDM than the Baetidae family. For example, *B. alpinus* has decreasing probabilities of occurrence at temperatures greater than ~15°C (Appendix S8, see p5), *B. rhodani* has lower occurrence probabilities at temperatures below ~7°C (Appendix S8, see p7), and *B. muticus* appears to prefer intermediate temperatures indicated by higher occurrence probabilities between ~10 and 17°C (Appendix S8, p13).

The fact that multiple widespread families contain species with a wider range of responses demonstrates that identifying specific taxa at finer taxonomic resolution can reveal additional information. However, many families include rare taxa for which we cannot infer significant responses (β parameter distributions largely overlap with zero, see Appendix S2 for

complete results on the effects of taxonomic resolution on the responses of EPT taxa in SGf and SGs datasets).
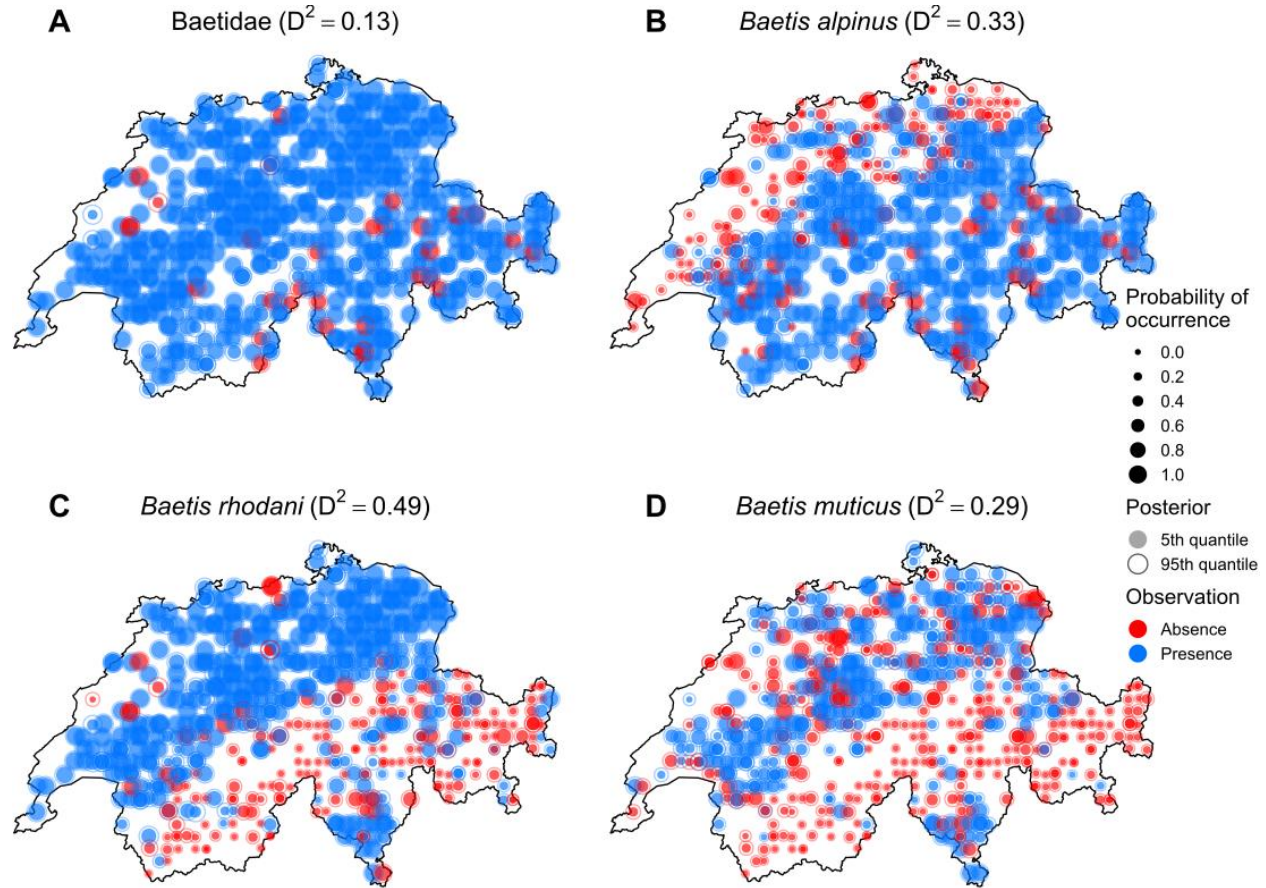

Fig. 8. Geographic distribution of the calibrated probabilities of occurrence (indicated by the point size) compared with presence-absence observations (blue or red points, respectively) of the family Baetidae (A) and selected Baetis species (B, C, and D). The size and color of the points indicate observations with a good (i.e., large blue and small red points) or bad (i.e., large red and small blue points) quality of fit. To visualize predictive uncertainty, the point fill shows the 5th (filled) and 95th (unfilled) quantiles of the marginal posterior probabilities of occurrence, with large differences between the filled and unfilled points indicating a high predictive uncertainty.
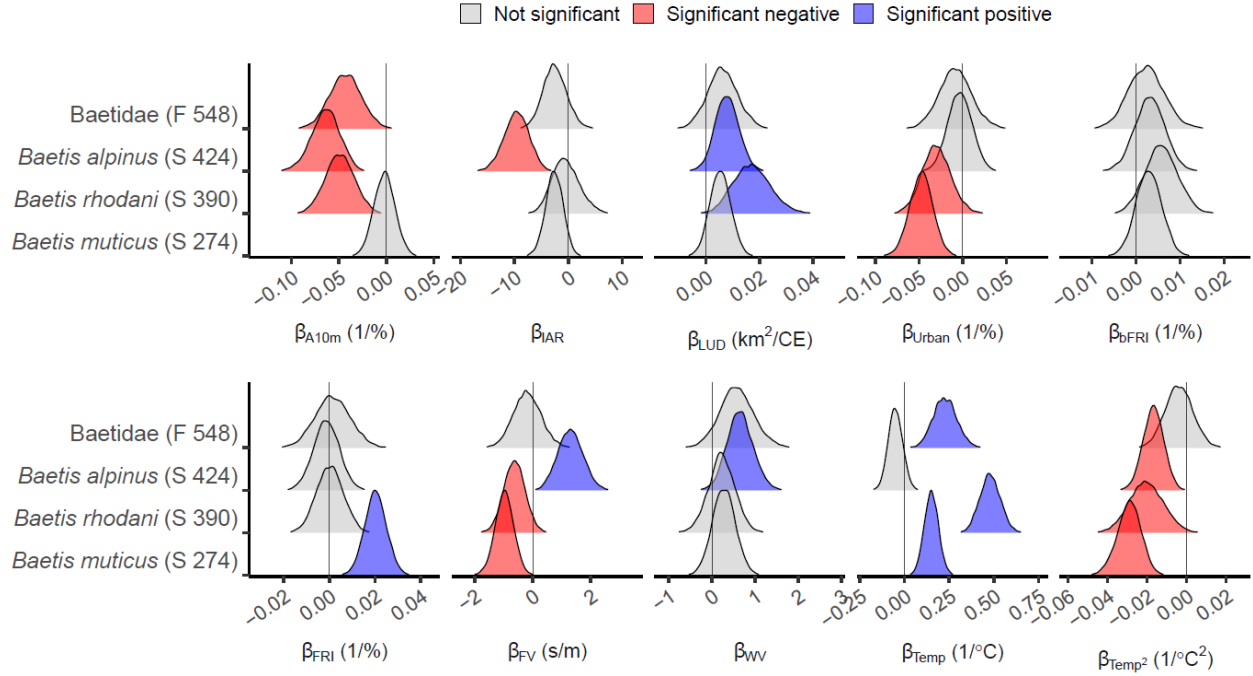
Fig. 9. Taxon specific responses to explanatory variables (i.e., the marginal posterior distributions of the taxon-specific parameters β_{kj}^{taxa}) in the SCf and SCs model, respectively, for the Baetidae family (F) and selected species (S) in the genus *Baetis*. The number of occurrences of each taxon is provided in brackets after the taxon names. Responses are colored according to significant positive (blue), negative (red), and non-significant (gray) responses based on the 5th and 95th quantiles of the marginal posterior distributions of the taxon-specific parameters β_{kj}^{taxa}.

4 Discussion

The main objective of this study is to identify how key aspects of biomonitoring design affect the inference of taxa responses to anthropogenic and natural environmental conditions. Because resources for biomonitoring are limited, monitoring design should be optimized to best fulfill monitoring objectives. Identifying the response of taxa to human impacts is one important objective, and can be supported with the application of SDMs. However, in the past, research using SDMs has often focused on the prediction of species occurrences (e.g., Elith and Leathwick 2009) with less focus on improving understanding about underlying mechanisms. Moreover, the selection of explanatory variables in SDMs often largely depends on data availability and is less often based on the understanding of causal mechanisms that drive species distributions (Dormann 2007, Elith and Leathwick 2009). However, if the pre-selection of explanatory variables accounts for potential mechanistic links, correlative species distribution models can support the identification of important mechanisms (Buckley et al. 2010).

Effects of monitoring design on model performance and relative importance of explanatory variables

In our application of SDMs, we identified two predominantly natural and seven anthropogenic environmental variables as the best predictor variables for macroinvertebrate occurrences in Switzerland across different biomonitoring datasets. Stream temperature emerged as the most important explanatory variable in all datasets, indicating potential effects of climate change on the future distributions of sensitive species. Anthropogenic variables representing catchment-wide properties, such as the land use index for the application of agricultural insecticides (IAR) and urban land use (Urban) have a greater effect on stream invertebrate communities (particularly at finer taxonomic resolutions) than riparian agriculture (A10m). The presence of drainage systems may contribute to the importance of agriculture beyond the riparian zone (e.g., Doppler et al. 2012) and biocides from urban sources seem to be relevant as well (Wittmer et al. 2010, 2011). Data on drainage networks of intensively cultivated farmland would be needed to better resolve the spatial scale at which they affect stream invertebrate assemblages. Although the land use index for agricultural insecticides used as a variable (IAR) in this study is a proxy designed to represent pollution caused by agricultural insecticides, it cannot be excluded

that it more generally indicates other types of agricultural pollution (e.g., organic matter inputs, fine sediment runoff) that are known to adversely affect stream invertebrates (Liess et al. 2012, Wagenhoff et al. 2012, Baumgartner and Robinson 2017). Studies selecting variables to include in SDMs should consider that multiple stressors potentially underlie a single proxy variable. Our general finding that many taxa responded to multiple anthropogenic explanatory variables highlights the need for a coordination of management measures that aim to tackle multiple stressors, such as river restoration or water quality improvements, to increase their success with regard to biological responses.

In applying hmSDMs to all four datasets of stream invertebrate assemblages that differ in taxonomic resolution, site-selection strategy, and geographic extent, we found that our ability to quantify the effects of selected natural and anthropogenic environmental conditions on these communities is strongly interdependent on these key decisions in biomonitoring design. Comparing the models for two datasets that differ in taxonomic resolution (for EPT taxa) but share the same sites revealed that identifying families at a finer taxonomic resolution (genus and species) led to improvements in both the predictive performance of the model and explanatory power of the selected variables. As expected, the relative importance of the explanatory variables also increased with finer taxonomic resolution. However, in practice, increasing taxonomic resolution also increases the cost per sample. Budget constraints may therefore limit the number of samples that can be collected, potentially resulting in tradeoffs between taxonomic resolution and other monitoring design choices, such as the selection of sites and the range of environmental conditions included in data collection.

The grid-based site-selection strategy in the SGf dataset led to a low coverage of sites with high levels of anthropogenic impacts, such as urban areas in the catchment, compared to the combined SCf data set. This imbalance may explain the lower relative importance of urban land use in the SGf model compared to the SCf model. These results indicate that a stratified site-selection strategy with a factorial design could improve the inference of taxa responses to anthropogenic impacts. Combining biomonitoring programs to increase sample size and improve the coverage of influence factors allowed us to infer stronger taxon-specific responses, and thus improve the quality of fit and predictive performance of the model (Fig. 6). This outcome implies that a harmonization of sampling methods among multiple biomonitoring programs

30

would be beneficial because it would allow a joint analysis of monitoring programs that are designed for different purposes and therefore may apply different site-selection strategies (e.g., grid-based site selection vs targeting impaired sites). However, in our study the combination of data from different biomonitoring programs required aggregating observations to the family-level because the different monitoring programs did not resolve the same taxa to genus or species level, which reduced the model's predictive performance and explanatory power. Despite this limitation, the explanatory variables in the model applied to the SCf dataset had a relative importance similar to the model for SGs.

Contrary to our expectation, limiting the geographic extent of sites to the more densely populated areas did not increase the relative importance of anthropogenic explanatory variables (Fig. 5). The limited range of natural environmental conditions (temperature, flow velocity) within the Swiss Plateau may have caused its poorer model performance. This outcome indicates that it may be advantageous to include larger datasets covering not only larger ranges of environmental conditions, but also a better coverage of different combinations of explanatory variables instead of focusing on a specific area with more homogenous environmental conditions.

Effects of taxonomic resolution on explanatory power and inference of responses

It would be beneficial for different programs to agree on the taxa that are resolved to genus or species level to avoid a loss in taxonomic resolution when combining different monitoring programs, as is done in several other countries (e.g., Haase et al. 2004). This decision should be based on practical considerations (e.g., taxon-specific effort and error rates) but also based on the gain in information that can be achieved with a finer resolution (e.g., inferring more differentiated responses of taxa to environmental conditions). Therefore, we investigated if a model-based analysis can support the identification of taxa for which a finer taxonomic resolution is particularly useful in identifying human impacts and how this ability depends on the prevalence of a given taxon. By comparing the explanatory power and parameter distributions for the different taxonomic levels, we found that genera and species with an intermediate prevalence that belonged to widespread families showed more differentiated and stronger responses to explanatory variables than their respective families. These taxa would be

particularly useful to improve bioassessment methods that are currently based on family level, as is the case in Switzerland (BAFU 2019). However, a majority of families in our study area (e.g., Rhyacophilidae, Hydropsychidae) showed no clear differentiation in their responses at finer resolutions, mainly because of their low prevalence. This lack of differentiation due to rarity would likely remain consistent, regardless of the explanatory variables or modeling approach chosen (Sor et al. 2017, Caradima et al. 2019).

A particularly interesting outcome of our study is the variable response of species within families to agricultural insecticides. Most of the ten most widespread EPT families responded negatively to agricultural insecticides (see Appendix S2), which supports independent findings that family-level identification of stream invertebrates is generally sufficient to identify significant responses to agricultural insecticides (Beketov et al. 2009). However, Baetidae, Limnephilidae, and Leptophlebiidae were exceptions to this trend. These families had an insignificant response at the family level but included 1 to 3 species that responded negatively to agricultural insecticides, which points to the benefit of a finer resolution for selected families when assessing the impacts of insecticide use.

Ensuring consistently low rates of observation error with traditional methods of identifying stream invertebrates (i.e., relying on morphological characteristics of individuals) is difficult for specific taxa and increases program costs. In the future, DNA-based methods may serve as complementary tools to traditional identification methods to increase the taxonomic resolution of biomonitoring data (Hering et al. 2018, Pawlowski et al. 2018). DNA identification may also allow for identification of previously unknown taxa and may help to develop new biotic indices based on genetic characteristics (Gibson et al. 2015, Elbrecht et al. 2017). Our model-based approach could be used to compare the information gained from DNA-based vs traditional identification methods. As demonstrated in this study, the model could be used to identify which specific taxa provide the greatest increase in information at a finer resolution for specific areas of interest.

Conclusions

We successfully applied hierarchical multi-species distribution models (hmSDMs) to datasets of benthic macroinvertebrate assemblages in Swiss rivers to assess the effects of key

biomonitoring design criteria on the inference of stream invertebrate responses to natural and anthropogenic environmental conditions. Site-selection strategy and taxonomic resolution both affected the predictive performance of the models and the strength of inferred responses of taxa. The degree to which taxonomic resolution affected the predictive and explanatory power of the models depended on how rare or widespread the taxa were and on how differently species from the same family responded to environmental conditions. Based on the findings of this study, we draw the following conclusions regarding practical implications for biomonitoring and management.

Biomonitoring

1. Systematic site selection (ideally with a factorial design) is an important aspect of the design of biomonitoring programs that aims to disentangle the effects of different natural and anthropogenic environmental conditions. The extent to which the effects of different explanatory variables can be disentangled with statistical methods depends on the coverage and correlation of the environmental conditions in the data. Additional sites with rare combinations of important influence factors should be targeted to improve the coverage of explanatory variables in existing monitoring programs. In the Swiss context, sites with low temperature and impaired water quality would be most valuable.

2. A finer taxonomic resolution can greatly improve the identifiability of responses to explanatory variables of genera and species if they have a sufficient prevalence for statistical analyses and show different and stronger responses than their respective families. These improvements are particularly evident among genera and species with an intermediate prevalence, as the responses of very widespread or very rare taxa are difficult to identify. Harmonizing the taxonomic resolution across different monitoring programs by agreeing on an operational taxa list that specifies the taxonomic level of identification for each taxon would avoid a loss of information in joint analyses and could increase cost efficiency.

3. The more biomonitoring data are available, the stronger are the conclusions that can be drawn from statistical analyses and the greater the potential to use the results for the development of new stressor-specific macroinvertebrate indices. A central database to store

data from different programs, as implemented in Switzerland, facilitates analyses across different monitoring programs and can improve data management. We recommend the inclusion of additional environmental information collected during sampling in such a data base.

Management

1. As many macroinvertebrate taxa in the model respond to multiple stressors, a coordination of management programs regarding different stressors seems important to increase program effectiveness regarding the response of aquatic organisms. For example, coordination between morphological restoration and water quality management programs could improve outcomes compared to the application of these programs at different locations.

2. Model results can facilitate the selection of taxa for the development of biological indices to assess the impact of management measures on organisms. Taxa that appear to be sensitive to flow velocity and width variability can be expected to respond to morphological restoration measures. Likewise, taxa with sensitive responses to indicators of agricultural activities (proportion of riparian agriculture, land use index for the application of agricultural insecticides, livestock unit density) can be expected to respond to measures to improve agricultural practice. Taxa with sensitive responses to the proportion of urban area in the catchment will potentially respond to measures to improve urban wastewater management.

3. Because temperature was identified as the most important influence factor, climate change can be expected to change the distribution of species and decrease the habitat availability for cold water species. This potential impact should be taken into account when assessing management effects over longer time scales, for example by implementing a before-after control-impact monitoring design. The model could facilitate the distinction of temperature effects from effects caused by other environmental conditions.

4. The prohibition of pesticide applications only within the riparian zone of streams may not be sufficient to greatly reduce adverse impacts on invertebrate communities, as indicated by the greater importance of catchment-wide vs riparian zone indicators of agricultural land use.

We conclude that a finer taxonomic identification of carefully selected taxa combined with a strategic choice of monitoring sites to improve the coverage of environmental conditions could increase the information gain achievable at given monitoring costs. We also conclude that hierarchical multi-species distribution models can support the optimization of such a monitoring design.

Acknowledgements

References

BAFU (Das Bundesamt für Umwelt). 2006. Ökomorphologie Stufe S (systembezogen): Methoden zur Untersuchung und Beurteilung der Fliessgewässer gemäss dem Modul-Stufen-Konzept. Bundesamt für Umwelt, CH-3003 Bern, Switzerland.

BAFU (Das Bundesamt für Umwelt). 2013. Nationale Beobachtung Oberflächengewässerqualität: Konzept Fliessgewässer. Bundesamt für Umwelt, Bern, Switzerland.

BAFU (Das Bundesamt für Umwelt). 2019. Methoden zur Untersuchung und Beurteilung von Fliessgewässern. Makrozoobenthos – Stufe F. 1. Bundesamt für Umwelt, Bern, Switzerland.

Bailey, R. C., R. H. Norris, and T. B. Reynoldson. 2001. Taxonomic resolution of benthic macroinvertebrate communities in bioassessments. Journal of the North American Benthological Society 20:280–286.

Baumgartner, S. D., and C. T. Robinson. 2017. Changes in macroinvertebrate trophic structure along a land-use gradient within a lowland stream network. Aquatic Sciences 79:407–418.

BDM Coordination Office (Biodiversity Monitoring in Switzerland). 2009. Biodiversity Monitoring in Switzerland (BDM) Coordination Office 2009: The state of biodiversity in Switzerland. Overview of the findings of Biodiversity Monitoring Switzerland as of May 2009. Federal Office for the Environment, Bern, Switzerland.

Beketov, M. A., K. Foit, R. B. Schäfer, C. A. Schriever, A. Sacchi, E. Capri, J. Biggs, C. Wells, and M. Liess. 2009. SPEAR indicates pesticide effects in streams—Comparative use of species- and family-level biomonitoring data. Environmental Pollution 157:1841–1848.

Beketov, M. A., B. J. Kefford, R. B. Schäfer, and M. Liess. 2013. Pesticides reduce regional biodiversity of stream invertebrates. Proceedings of the National Academy of Sciences 110:11039–11043.

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

Buckley, L. B., M. C. Urban, M. J. Angilletta, L. G. Crozier, L. J. Rissler, and M. W. Sears. 2010. Can mechanism inform species' distribution models? Ecology Letters 13:1041–1054.

Bunzel, K., M. Kattwinkel, and M. Liess. 2013. Effects of organic pollutants from wastewater treatment plants on aquatic invertebrate communities. Water Research 47:597–606.

Bunzel, K., M. Liess, and M. Kattwinkel. 2014. Landscape parameters driving aquatic pesticide exposure and effects. Environmental Pollution 186:90–97.

Cao, Y., and C. P. Hawkins. 2011. The comparability of bioassessments: A review of conceptual and methodological issues. Journal of the North American Benthological Society 30:680–701.

Caradima, B., N. Schuwirth, and P. Reichert. 2019. From individual to joint species distribution models: A comparison of model complexity and predictive performance. Journal of Biogeography 46:2260–2274.

Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. 2017. Stan: A probabilistic programming language. Journal of Statistical Software 76:1–32.

Clark, J. S., A. E. Gelfand, C. W. Woodall, and K. Zhu. 2014. More than the sum of the parts: Forest climate response from joint species distribution models. Ecological Applications 24:990–999.

Domisch, S., S. C. Jähnig, J. P. Simaika, M. Kuemmerlen, and S. Stoll. 2015. Application of species distribution models in stream ecosystems: The challenges of spatial and temporal scale, environmental predictors and species occurrence data. Fundamental and Applied Limnology: Archiv für Hydrobiologie 186:45–61.

Doppler, T., L. Camenzuli, G. Hirzel, M. Krauss, A. Lück, and C. Stamm. 2012. Spatial variability of herbicide mobilisation and transport at catchment scale: Insights from a field experiment. Hydrology and Earth System Sciences 16:1947–1967.

Dormann, C. F. 2007. Promising the future? Global change projections of species distributions. Basic and Applied Ecology 8:387–397.

Duane, S., A. D. Kennedy, B. J. Pendleton, and D. Roweth. 1987. Hybrid Monte Carlo. Physics Letters B 195:216–222.

Eggen, R. I. L., J. Hollender, A. Joss, M. Schärer, and C. Stamm. 2014. Reducing the discharge of micropollutants in the aquatic environment: the benefits of upgrading wastewater treatment plants. Environmental Science & Technology 48:7683–7689.

Elbrecht V., A. J. Beermann, G. Goessler, J. Neumann, R. Tollrian, R. Wagner, A. Wlecklik, J. J. Piggott, C. D. Matthaei, and L. Florian. 2016. Multiple stressor effects on stream invertebrates: A mesocosm experiment manipulating nutrients, fine sediment and flow velocity. Freshwater Biology 61:362–375.

Elbrecht, V., E. E. Vamos, K. Meissner, J. Aroviita, and F. Leese. 2017. Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. Methods in Ecology and Evolution 8:1265–1275.

Elith, J., G. Catherine, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. M. Overton, A. Townsend, Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberón, S. Williams, M. S. Wisz, and N. E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129–151.

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.

Feld, C. K., F. de Bello, and S. Dolédec. 2014. Biodiversity of traits and species both show weak responses to hydromorphological alteration in lowland river macroinvertebrates. Freshwater Biology 59:233–248.

Gibson, J. F., S. Shokralla, C. Curry, D. J. Baird, W. A. Monk, I. King, and M. Hajibabaei. 2015. Large-scale biomonitoring of remote and threatened ecosystems via high-throughput sequencing. PLoS ONE 10:e0138432.

Guisan, A., O. Broennimann, R. Engler, M. Vust, N. G. Yoccoz, A. Lehmann, and N. E. Zimmermann. 2006. Using niche-based models to improve the sampling of rare species. Conservation Biology 20:501–511.

Guisan, A., T. C. Edwards, and T. Hastie. 2002. Generalized linear and generalized additive models in studies of species distributions: Setting the scene. Ecological Modelling 157:89–100.

Guisan, A., and C. Rahbek. 2011. SESAM—A new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. Journal of Biogeography 38:1433–1444.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135:147–186.

Haase, P., S. Lohse, S. Pauls, K. Schindehütte, A. Sundermann, P. Rolauffs, and D. Hering. 2004. Assessing streams in Germany with benthic invertebrates: Development of a practical standardised protocol for macroinvertebrate sampling and sorting. Limnologica: Ecology and Management of Inland Waters 34:349–365.

Hawkins, C. P., and L. L. Yuan. 2016. Multitaxon distribution models reveal severe alteration in the regional biodiversity of freshwater invertebrates. Freshwater Science 35:1365–1376.

Hering, D., A. Borja, J. I. Jones, D. Pont, P. Boets, A. Bouchez, K. Bruce, S. Drakare, B. Hänfling, M. Kahlert, F. Leese, K. Meissner, P. Mergen, Y. Reyjol, P. Segurado, A. Vogler, and M. Kelly. 2018. Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive. Water Research 138:192–205.

Hering, J. G., E. Hoehn, A. Klinke, M. Maurer, A. Peter, P. Reichert, C. Robinson, K. Schirmer, M. Schirmer, C. Stamm, and B. Wehrli. 2012. Moving targets, long-lived infrastructure, and increasing needs for integration and adaptation in water management: An illustration from Switzerland. Environmental Science & Technology 46:112–118.

Hernandez, P. A., C. H. Graham, L. L. Master, and D. L. Albert. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29:773–785.

James, G., D. Witten, T. Hastie, and R. Tibshirani. 2013. An introduction to statistical learning with applications in R. Springer, New York, New York.

Jones, F. C. 2008. Taxonomic sufficiency: The influence of taxonomic resolution on freshwater bioassessments using benthic macroinvertebrates. Environmental Reviews 16:45–69.

Kuemmerlen, M., S. Stoll, A. Sundermann, and P. Haase. 2016. Long-term monitoring data meet freshwater species distribution models: Lessons from an LTER-site. Ecological Indicators 65:122–132.

Kunz, M., Y. S. Wildhaber, A. Dietzel, I. Wittmer, and V. Leib. 2016. Ergebnisse der Nationalen Beobachtung Oberflächengewässerqualität (NAWA) 2011–2014. Bundesamt für Umwelt, Bern, Switzerland.

Latimer, A. M., S. Banerjee, H. Sang Jr, E. S. Mosher, and J. A. Silander Jr. 2009. Hierarchical models facilitate spatial analysis of large data sets: A case study on invasive plant species in the northeastern United States. Ecology Letters 12:144–154.

Lenat, D. R., and V. H. Resh. 2001. Taxonomy and stream ecology—The benefits of genus- and species-level identifications. Journal of the North American Benthological Society 20:287–298.

Leps, M., J. D. Tonkin, V. Dahm, P. Haase, and A. Sundermann. 2015. Disentangling environmental drivers of benthic invertebrate assemblages: The role of spatial scale and riverscape heterogeneity in a multiple stressor environment. Science of The Total Environment 536:546–556.

Liess, A., A. Le Gros, A. Wagenhoff, C. R. Townsend, and C. D. Matthaei. 2012. Landuse intensity in stream catchments affects the benthic food web: Consequences for nutrient supply, periphyton C:nutrient ratios, and invertebrate richness and abundance. Freshwater Science 31:813–824.

Liess, M., and P. C. V. D. Ohe. 2005. Analyzing effects of pesticides on invertebrate communities in streams. Environmental Toxicology and Chemistry 24:954–965.

Marshall, J. C., A. L. Steward, and B. D. Harch. 2006. Taxonomic resolution and quantification of freshwater macroinvertebrate samples from an Australian dryland river: The benefits and costs of using species abundance data. Hydrobiologia 572:171–194.

Nelder, J. A., and R. Mead. 1965. A simplex method for function minimization. The Computer Journal 7:308–313.

Ovaskainen, O., and J. Soininen. 2011. Making more out of sparse data: Hierarchical modeling of species communities. Ecology 92:289–295.

Pawlowski, J., M. Kelly-Quinn, F. Altermatt, L. Apothéloz-Perret-Gentil, P. Beja, A. Boggero, A. Borja, A. Bouchez, T. Cordier, I. Domaizon, M. J. Feio, A. F. Filipe, R. Fornaroli, W. Graf, J. Herder, B. van der Hoorn, J. Iwan Jones, M. Sagova-Mareckova, C. Moritz, J. Barquín, J. J. Piggott, M. Pinna, F. Rimet, B. Rinkevich, C. Sousa-Santos, V. Specchia, R. Trobajo, V. Vasselon, S. Vitecek, J. Zimmerman, A. Weigand, F. Leese, and M. Kahlert. 2018. The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. Science of The Total Environment 637–638:1295–1310.

Piggott, J. J., C. R. Townsend, and C. D. Matthaei. 2015. Climate warming and agricultural stressors interact to determine stream macroinvertebrate community dynamics. Global Change Biology 21:1887–1906.

Rääpysjärvi, J., H. Hämäläinen, and J. Aroviita. 2016. Macrophytes in boreal streams: Characterizing and predicting native occurrence and abundance to assess human impact. Ecological Indicators 64:309–318.

Sor, R., Y.-S. Park, P. Boets, P. L. M. Goethals, and S. Lek. 2017. Effects of species prevalence on the performance of predictive models. Ecological Modelling 354:11–19.

Sponseller, R. A., E. F. Benfield, and H. M. Valett. 2001. Relationships between land use, spatial scale and stream macroinvertebrate communities. Freshwater Biology 46:1409–1424.

Stamm, C., K. Räsänen, F. J. Burdon, F. Altermatt, J. Jokela, A. Joss, M. Ackermann, and R. I. L. Eggen. 2016. Unravelling the impacts of micropollutants in aquatic ecosystems: Interdisciplinary studies at the interface of large-scale ecology. Pages 183–223 *in* A. J. Dumbrell, R. L. Kordas, and G. Woodward (editors). Advances in Ecological Research. Academic Press, Cambridge, Massachusetts.

Strayer, D. L., and D. Dudgeon. 2010. Freshwater biodiversity conservation: Recent progress and future challenges. Journal of the North American Benthological Society 29:344–358.

Stucki, P. 2010. Methoden zur Untersuchung und Beurteilung der Fliessgewässer: Makrozoobenthos Stufe F. Bundesamt für Umwelt, Bern Umwelt-Vollzug 1026:61.

Thuiller, W., L. Brotons, M. B. Araújo, and S. Lavorel. 2004. Effects of restricting environmental range of data to project current and future species distributions. Ecography 27:165–172.

Thuiller, W., B. Lafourcade, R. Engler, and M. B. Araújo. 2009. BIOMOD—A platform for ensemble forecasting of species distributions. Ecography 32:369–373.

Townsend, C. R., S. S. Uhlmann, and C. D. Matthaei. 2008. Individual and combined responses of stream ecosystems to multiple stressors. Journal of Applied Ecology 45:1810–1819.

Wagenhoff, A., C. R. Townsend, and C. D. Matthaei. 2012. Macroinvertebrate responses along broad stressor gradients of deposited fine sediment and dissolved nutrients: A stream mesocosm experiment. Journal of Applied Ecology 49:892–902.

Wagenhoff, A., C. R. Townsend, N. Phillips, and C. D. Matthaei. 2011. Subsidy-stress and multiple-stressor effects along gradients of deposited fine sediment and dissolved nutrients in a regional set of streams and rivers: Sediment and nutrients in streams. Freshwater Biology 56:1916–1936.

Walsh, C. J., and J. A. Webb. 2016. Interactive effects of urban stormwater drainage, land clearance, and flow regime on stream macroinvertebrate assemblages across a large metropolitan region. Freshwater Science 35:324–339.

Warton, D. I., F. G. Blanchet, R. B. O'Hara, O. Ovaskainen, S. Taskinen, S. C. Walker, and F. K. C. Hui. 2015. So many variables: Joint modeling in community ecology. Trends in Ecology & Evolution 30:766–779.

Wittmer, I. K., H.-P. Bader, R. Scheidegger, H. Singer, A. Lück, I. Hanke, C. Carlsson, and C. Stamm. 2010. Significance of urban and agricultural land use for biocide and pesticide dynamics in surface waters. Water Research 44:2850–2862.

Wittmer, I. K., R. Scheidegger, H.-P. Bader, H. Singer, and C. Stamm. 2011. Loss rates of urban biocides can exceed those of agricultural pesticides. Science of The Total Environment 409:920–932.

Wright, J. F., D. W. Sutcliffe, and M. T. Furse. 2000. Assessing the biological quality of fresh waters—RIVPACS and other techniques. Freshwater Biological Association, Ambleside, Cumbria, UK.